All Sciences Academy

CONTEMPORARY RESEARCH IN ENGINEERING

CONTEMPORARY RESEARCH IN ENGINEERING

Editor
Asst. Prof. Dr. Umut ÖZKAYA

Contemporary Research in Engineering
Editor: Asst. Prof. Dr. Umut ÖZKAYA

Design: All Sciences Academy Design
Published Date: August 2025
Publisher's Certification Number: 72273
ISBN: 978-625-5794-04-8

www.allsciencesacademy.com
allsciencesacademy@gmail.com

CONTENT

1. Chapter **6**

Control and Performance Evaluation of Ann-Based MPPT for Photovoltaic Systems: A Comparative Study with the Conventional P&O Algorithm

Batın DEMİRCAN

2. Chapter **31**

Numerical Analysis of Hydrogen-Enriched Diesel Combustion Performance and Emission Characteristics

Serdar HALİS, Furkan YAZICI, Şevket Onur KILIÇ

3. Chapter **46**

Facility Layout Optimization and Bottleneck Elimination in Loofah-Based Manufacturing Using Lean Tools and Excel Craft

Aytuğ Berke PALANKALILAR, Melike BAŞKAYA, Mehmet SAVSAR

4. Chapter **61**

Quality Improvement in Loofah-Based Manufacturing Using P-Chart Analysis, Sampling Strategies, and Lean Diagnostic Tools

Zeynep BAŞTUĞ, Gökçe ERMİŞ, Mehmet SAVSAR

5. Chapter **69**

Optimization Model to Minimize Cutting Waste in Loofah-Based Manufacturing Using LP Model and Excel Solver

Şüheda İLASLAN, Merve ALPER, Mehmet SAVSAR

6. Chapter **78**

Injection Timing Optimization in Dual-Fuel Engines: A Numerical Approach to Combustion and Emission Analysis

Şevket Onur KILIÇ, Furkan YAZICI, Serdar HALIS

7. Chapter **95**

The Evolution and Future Prospects of Web GIS: From Static Maps to Intelligent 3D Systems

Tasnim ALASALI, Omar DAKKAK, Ismail Rakip KARAS, Yousef FAZEA

8. Chapter **126**
An AI-Driven Framework for MBM enabled NOMA: A Roadmap for Advanced Generation Mobile Communication
Busra KARAHAN

9. Chapter **140**
Metamaterial-Based Microwave Sensors
Taha Fatih ATEŞ, Ali Osman ÖZKAN

10. Chapter **156**
Memristor Types and Application Areas
Muhammet ATEŞ

11. Chapter **183**
Exploring the Efficiency of the Fenton Oxidation Process for the Removal of Pyridaben from Agricultural Wastewater
Burcu ŞİMŞEK UYGUN, Yağmur KAYA

12. Chapter **196**
Effects of Forest Fires on Soil Properties
Nesli AYDIN, Gül KAYKIOĞLU

13. Chapter **218**
Convolutional Neural Network Approach to Modulation Classification for 5G and Cognitive Radio
Asuman SAVAŞCIHABEŞ

14. Chapter **233**
In Situ and Ex Situ Strategies for the Remediation of Petroleum-Contaminated Soils: Recent Advances and Emerging Trends
F. Olcay TOPAÇ

15. Chapter **260**
Assessment of Ermenek District's Potential in Terms of Renewable Energy Resources
Hüseyin KAYA, Mesut KİLİT

Control and Performance Evaluation of Ann-Based MPPT for Photovoltaic Systems: A Comparative Study with the Conventional P&O Algorithm

Batın DEMİRCAN[1,2]

1- Lecturer; Balıkesir University, Balıkesir Vocational School, Department of Electronic and Automation.
2- Reseacher; Balıkesir University, Renewable Energy Research, Application and Development Center.
batindemircan@gmail.com ORCID NO: 0000-0002-0765-458X

ABSTRACT

This study presents a comparative analysis of a conventional Perturb and Observe (P&O) maximum power point tracking (MPPT) algorithm and an Artificial Neural Network (ANN)-based MPPT controller for photovoltaic (PV) systems. A detailed PV system model including a boost converter and IGBT switching was implemented in MATLAB/Simulink. Initially, the P&O algorithm was used to generate training data by recording voltage, current, irradiance, and temperature signals. A 600-second simulation with varying irradiance and temperature profiles was conducted, and the collected dataset was used to train the ANN. The ANN model was designed with three hidden layers (32–16–8 neurons) and trained using the Levenberg–Marquardt algorithm. Training performance was evaluated through regression plots, gradient analysis, and error histograms. After training, the ANN-based MPPT was integrated into a secondary Simulink model and tested under two scenarios: stable irradiance and temperature variations, and highly dynamic challenging conditions.

Simulation results indicate that under stable conditions, the ANN controller performs similarly to the P&O algorithm, providing almost identical voltage, current, and power responses. However, under rapidly changing environmental conditions, the ANN model exhibits smoother transitions, filters high-frequency fluctuations, and maintains power closer to the maximum power point compared to P&O. These results demonstrate that the ANN-based MPPT not only matches the tracking capability of the conventional method under standard scenarios but also improves robustness and adaptability under complex operating conditions. The study highlights the potential of data-driven approaches as an effective alternative to classical MPPT techniques for real-time PV applications.

Keywords –Artificial Neural Network (ANN), Photovoltaic Systems, Maximum Power Point Tracking (MPPT), Perturb and Observe (P&O), MATLAB/Simulink.

INTRODUCTION

The continuous increase in global energy demand and the environmental impacts of fossil fuel resources have made renewable energy technologies indispensable. Among these, photovoltaic (PV) systems have become one of the most preferred renewable energy solutions due to their low operating costs, long service life, and environmentally friendly nature. However, the output characteristics of PV modules vary significantly depending on environmental factors such as solar irradiance and temperature, preventing the system from consistently operating at the maximum power point (MPP). Therefore, maximum power point tracking (MPPT) algorithms are of critical importance for improving the efficiency of PV systems (Bastidas-Rodriguez, Franco, Petrone, Ramos-Paja, &

Spagnuolo, 2014; Bollipo, Mikkili, & Bonthagorla, 2020; Elobaid, Abdelsalam, & Zakzouk, 2015; Roy et al., 2021).

MPPT methods are generally classified as classical, intelligent control-based, metaheuristic optimization-based, and hybrid solutions integrating artificial intelligence and hardware-oriented approaches. Classical techniques such as Perturb and Observe (P&O) and Incremental Conductance (INC) are widely used due to their simplicity, low cost, and ease of implementation. However, these methods are limited to tracking only local maxima under partial shading conditions (PSC) and fail to locate the global maximum power point (Bastidas-Rodriguez et al., 2014; Sukanya Satapathy & Kumar, 2020). Moreover, they exhibit low stability under rapidly changing irradiance and temperature, have slow response times, and induce steady-state oscillations (Tahiri, Harrouz, & Bostanci, n.d.-a).

To overcome these limitations, intelligent control-based approaches provide higher accuracy and better adaptability. Fuzzy Logic Control (FLC) algorithms achieve faster and more precise tracking compared to conventional P&O while maintaining robustness under load variations (Youssfi, Alioui, & El Kadi, 2024). Model Reference Adaptive Control (MRAC) and Sliding Mode Control (SMC) enhance resilience to parameter uncertainties and fast environmental dynamics, improving the robustness of classical approaches (Haddou et al., 2020; Tahiri, Harrouz, & Bostanci, n.d.-b). Furthermore, Ripple Correlation Control (RCC) algorithms utilize natural harmonics in PV current and voltage, enabling stable and highly efficient MPP tracking in three-level inverter-based systems without requiring additional perturbation signals (Hammami, Ricco, Ruderman, & Grandi, 2019).

To identify the global maximum under partial shading conditions, metaheuristic optimization algorithms are widely employed. For example, the hybrid Cuckoo Search + Artificial Bee Colony (CSA-ABC) method can reach the global MPP in multi-peak PV curves up to 78% faster and without oscillations compared to conventional methods (Qi et al., 2024).

In recent years, machine learning (ML) and artificial intelligence (AI)-based approaches have gained prominence in MPPT applications. Artificial Neural Network (ANN)-based methods can directly learn irradiance and temperature variations, enabling real-time estimation of the optimal internal resistance or duty cycle, effectively eliminating steady-state oscillations (Bouadjila, Khelil, Rahem, & Berrezzek, 2023; Roy et al., 2021). Consequently, current trends in the literature reveal that maximum power tracking in PV systems is no longer solely a classical control problem but requires hybrid, AI-assisted, and optimization-based algorithms capable of handling complex factors such as environmental variability, load changes, and grid integration. Future research is expected to combine hybrid AI-based MPPT methods with adaptive control and metaheuristic optimization

techniques to provide more stable, faster, and higher-efficiency solutions (Bollipo et al., 2020; Elobaid et al., 2015; Tajaldin & Motlak, 2024).

Based on the findings of the literature review, in this study, the strengths and weaknesses of MPPT methods developed for photovoltaic systems were considered, and an MPPT structure within a PV-based renewable energy system was implemented in MATLAB/Simulink using both a conventional algorithm and a data-driven ANN-based approach. The obtained simulation results are presented, and all application steps are systematically described. MATLAB software (MathWorks, 2025) was used under a campus student license for the simulations.

SYSTEM MODELING AND DATA SETS

In this study, a real-time photovoltaic (PV) panel model was developed in the MATLAB/Simulink environment, and voltage, current, and power data were initially obtained using the conventional Perturb and Observe (P&O) algorithm. In the first stage, a relatively challenging operating scenario was designed by incorporating stepwise irradiance variations and temperature ramps to diversify the dynamic responses of the model. As a result of these scenarios, the PV operating data (Vpv, Ipv, G, T) together with the duty cycle commands generated by the P&O algorithm were transferred to the MATLAB workspace via the **To Workspace** block and systematically stored for the training of an Artificial Neural Network (ANN).

In the next stage, an ANN-based MPPT controller was trained using this dataset and designed to directly take PV voltage, current, irradiance, and temperature values as inputs and generate the optimal duty cycle. The trained ANN model was then integrated into the Simulink environment using the **genFunction** tool. After the training phase, the ANN-based MPPT was first tested under constant irradiance and temperature conditions to compare its tracking performance with the conventional P&O algorithm. In the final stage, the same ANN model was subjected to more demanding scenarios containing rapidly varying irradiance and temperature profiles and evaluated under identical conditions with the P&O algorithm.

A boost converter was employed in the PV system to regulate the output voltage and perform the MPPT operation, ensuring stable power control and accurate performance evaluation. The fundamental electrical parameters of the PV array used in the study, measured under Standard Test Conditions (STC), are presented in Table 1.

Table 1: PV Array Parameters

Parameter	Value	Description
Number of series-connected cells	10	Number of cells in a single string
Number of parallel strings	40	Number of strings connected in parallel to increase total capacity
Measured V_{oc} per cell	36.3 V	Open-circuit voltage
Measured I_{sc} per cell	7.84 A	Short-circuit current
Maximum power point (Pmpp)	213.15 W	Peak power value obtained under STC conditions
Vmp	29 V	Voltage at the maximum power point
Imp	7.35 A	Current at the maximum power point
V_{oc} temperature coefficient	−0.36099 %/°C	Rate of change of V_{oc} with increasing temperature
I_{sc} temperature coefficient	0.102 %/°C	Rate of change of I_{sc} with increasing temperature

The current–voltage (I–V) and power–voltage (P–V) characteristics of the PV array under different temperature conditions are shown in Figure 1. As seen in the upper graph, the open-circuit voltage (V_{oc}) decreases with increasing temperature. The voltage value obtained at 25 °C drops significantly at 45 °C. In contrast, the short-circuit current (I_{sc}) is only slightly affected by temperature variations and remains nearly constant. A similar trend is observed in the lower graph, where the power–voltage characteristics are presented. The maximum power point (MPP) shifts to lower values in both voltage and power as the temperature increases. At 25 °C, the array reaches its maximum power output within a higher voltage range, while at 45 °C this point shifts to a lower voltage level. This behavior results from the temperature-sensitive semiconductor structure of PV cells, leading to reduced cell efficiency as the temperature rises.

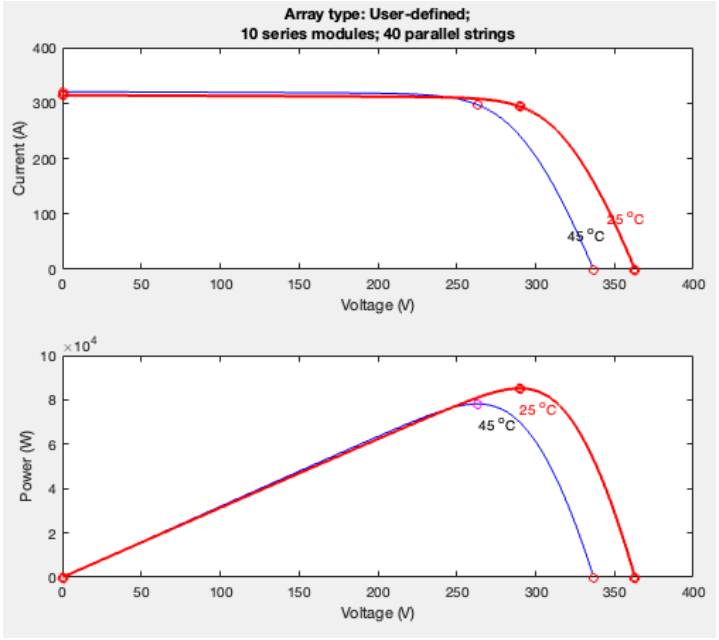


Figure 1: Matlab/Simulink PV Array Specs

The fundamental structure of the study was implemented using the model created in the MATLAB/Simulink environment, as shown in Figure 2. A boost-type converter was connected to the output of the PV panel to enable the operation of maximum power point tracking (MPPT) algorithms. The boost converter aims to increase the voltage level of the PV panel, thereby ensuring that the desired voltage and power values on the load side can be obtained in a more stable manner.

In the converter design, the input capacitor was selected as C_i = 100 µF, the inductance as L_i = 5 mH, and the output capacitor as C_o = 470 µF. On the load side, a resistance value of R_{load} = 100 Ω was chosen, providing an operating point suitable for observing both transient and steady-state performance of the system. Diodes placed at the input and output of the inductor ensure unidirectional current flow and prevent reverse current from flowing back into the PV panel.

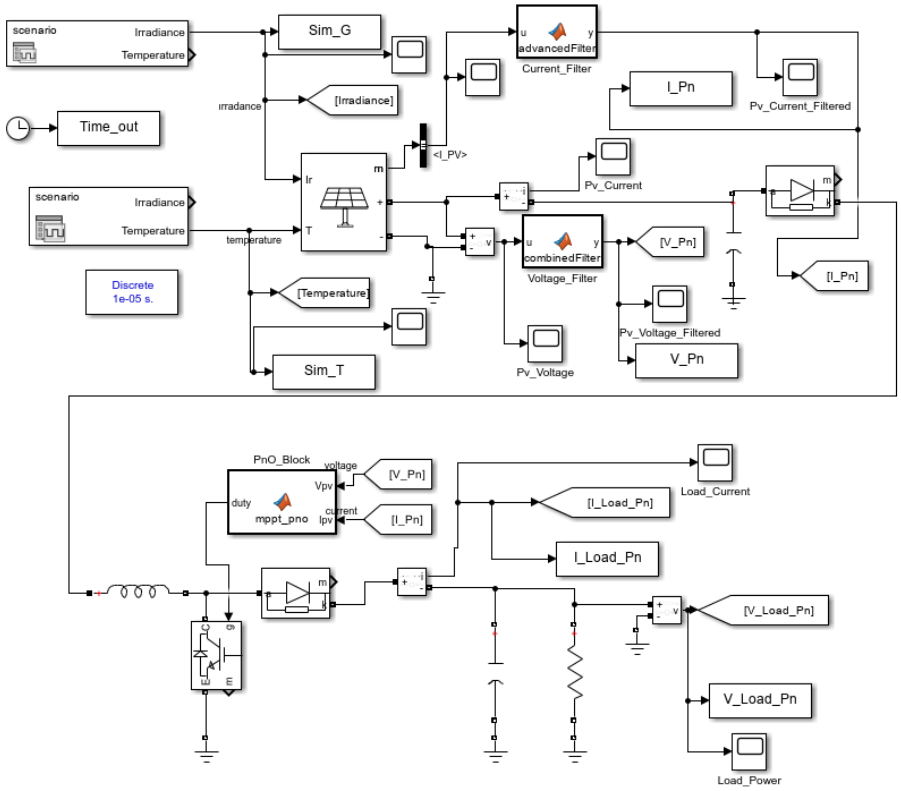


Figure 2: Matlab/Simulink PV System

The switching of the boost converter was initially controlled using the Perturb & Observe (P&O) algorithm. This algorithm continuously monitors the output power of the PV panel while applying small perturbations to the voltage and observes whether the power increases or decreases. If an increase in power is detected, the algorithm continues to perturb in the same direction, whereas if a decrease in power is observed, the perturbation direction is reversed. In this way, the system is maintained at an operating point close to the maximum power point (MPP). The IGBT/MOSFET model used in the boost converter features an on-state resistance of $R_{on} = 0.001\ \Omega$, ensuring low conduction losses. The snubber circuit consists of a resistor $Rs = 5000\ \Omega$ and a capacitor $C_s = 500$ nF, providing protection against voltage spikes during switching events.

In the numerical solution of the model, the **powergui** and simulation solver settings were configured with a time step of 1×10^{-5} seconds. This temporal resolution was selected to accurately capture fast transient responses and switching ripples in a stable manner. In addition, a two-stage filtering approach was applied to the voltage and current measurements at the PV panel output. The first stage, referred to as the *common filter*, was

employed to smooth fundamental voltage fluctuations, while the second stage, termed the *advanced filter*, was designed to suppress switching harmonics and provide cleaner signals to the MPPT algorithms. Within this context, the voltage filter code used for voltage measurement is presented below.

```
function y = combinedFilter(u)
persistent a buf idx z1 Ts_fc fc N
if isempty(buf)
N = 500; buf = zeros(1,N); idx = 1; z1 = 0;
Ts_fc = 1e-4;
fc    = 20;
a     = exp(-2*pi*fc*Ts_fc);
end
buf(idx) = u;
idx = mod(idx, N) + 1;
ma = sum(buf) / N;
y = (1 - a) * ma + a * z1;
z1 = y;
end
```

The current signal at the PV panel output is subjected to high-frequency oscillations caused by the switching operations of the boost converter and numerical discretization errors in the Simulink environment. Therefore, a multi-stage filtering structure was developed to provide a more stable measurement signal for the MPPT algorithm. This filtering approach first computes a moving average over 15 samples to smooth sudden noise and stabilize the measurement signal. Subsequently, a two-sample delay is introduced to prevent unstable responses in the MPPT algorithm. A low-pass exponential filter is then applied to suppress high-frequency oscillations while preserving the long-term signal components. Finally, the delayed average and exponential filter outputs are combined using a beta coefficient, producing an output that is both sensitive to transient conditions and stable in steady-state operation. The code of the current filter used in this process is provided below.

```
function y = advancedFilter(u)
persistent buf idx z1 Ts_fc fc N
beta delayBuf delayIdx avgBuf avgIdx
if isempty(buf)
   N = 500;
   buf = zeros(1,N); idx = 1;
   z1 = 0;
   Ts_fc = 1e-4; fc = 30;
   beta = 0.5;
   delayBuf = zeros(1,2); delayIdx = 1;
   avgBuf = zeros(1,15); avgIdx = 1;
end avgBuf(avgIdx) = u;
avgIdx = mod(avgIdx,15)+1;
avgOut = mean(avgBuf);
delayBuf(delayIdx) = avgOut;
delayIdx = mod(delayIdx,2)+1;
delayedOut = delayBuf(delayIdx);
a = exp(-2*pi*fc*Ts_fc);
ma = sum(buf)/N;
filteredCore = (1-a)*ma + a*z1;
z1 = filteredCore;
y = abs(beta*filteredCore + (1-
beta)*delayedOut);
buf(idx) = u;
idx = mod(idx,N)+1; end
```

For MPPT control, an initial simulation was carried out based on the conventional Perturb & Observe (P&O) algorithm. This algorithm calculates

the instantaneous power of the PV panel using the measured voltage (V_{pv}) and current (I_{pv}), applies small voltage perturbations, and determines the direction of adjustment based on the resulting power change to maintain operation close to the maximum power point (MPP).

The implemented MPPT_PNO function was optimized to provide a smoother control response compared to the standard P&O structure. The update period was set to Ts_mppt = 1 ms, the maximum duty cycle change was limited to dmax = 0.005, and the decision tolerance was reduced to tol = 1×10^{-3} to minimize sudden oscillations. The algorithm logic compares power and voltage variations (dP and dV) to decide whether to increase or decrease the duty cycle, while keeping it constrained within the range of 0.05–0.95. This approach preserves the main advantages of the widely used P&O method simplicity and fast response while aiming to reduce high-frequency oscillations that may occur in Simulink-based numerical simulations. The MPPT_PNO function code used in this context is provided below.

```
function duty = mppt_pno(Vpv, Ipv)
Ts_mppt = 1e-3;
dmax    = 0.005;
        tol     = 1e-3;
        persistent prevV prevI prevP
prevDuty timeCounter
        if isempty(prevV)
        prevV = Vpv;
        prevI = Ipv;
        prevP = Vpv * Ipv;
        prevDuty = 0.5;
        timeCounter = 0;
        end
        timeCounter = timeCounter +
Ts_mppt;
        if timeCounter < Ts_mppt
        duty = prevDuty;
        return;
          end
          timeCounter = 0;
          P = Vpv * Ipv;
          dP = P   - prevP;
          dV = Vpv - prevV;
          if abs(dP) < tol
             delta = 0;
          elseif dP > 0
             if dV > 0
                delta = dmax;
             else
                delta = -dmax;
             end
          else
             if dV > 0
                delta = -dmax;
             else
                delta = dmax;
             end
          end
          duty = prevDuty + delta;
          duty = max(0.05, min(0.95, duty));
          prevV    = Vpv;
          prevI    = Ipv;
          prevP    = P;
          prevDuty = duty;
        end
```

In the Simulink-based model where the PV system was controlled using the P&O algorithm, the simulation step size was set to 1×10^{-5} s. While this high resolution allows more accurate capture of transient behaviors and

switching dynamics, it also imposes a significant computational load on the processor. Therefore, in the **To Workspace** blocks used for recording each measured signal, the decimation value was set to 100, meaning only one out of every 100 samples was stored. This approach optimized both memory usage and processor load, enabling the simulation to be executed more efficiently.

For ANN training, a simulation duration of 600 seconds was defined, during which the PV system data generated by the P&O algorithm were recorded. The resulting dataset was designed to include the dynamic responses of the PV panel under varying environmental conditions. The irradiance and temperature variation profiles used in the initial simulation are presented in Figure 3 and Figure 4, respectively.

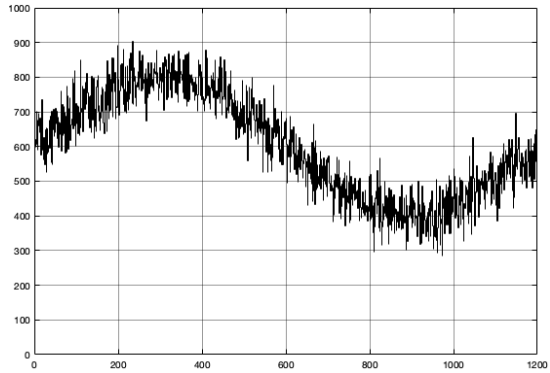


Figure 3: Irradiance Variation Profile

The irradiance profile shown in Figure 3 was designed as a signal varying between approximately 600 W/m² and 900 W/m², containing random noise components. Small-amplitude, high-frequency disturbances superimposed on a sinusoidal base waveform simulate sudden cloud transients and measurement noise encountered under real environmental conditions. During the first half of the simulation, the irradiance value gradually increases, while in the second half it exhibits a gradual decreasing trend, allowing the dynamic effects of both rising and falling radiation conditions to be analyzed.

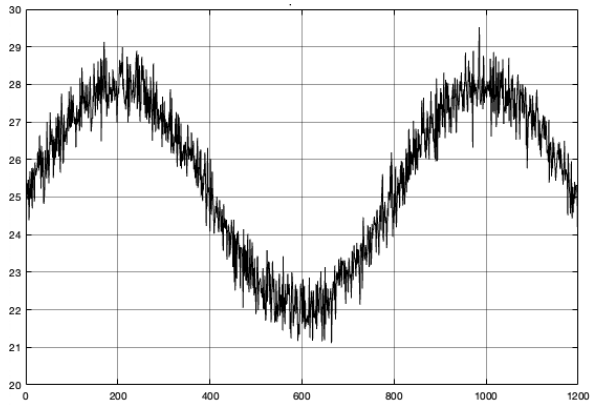


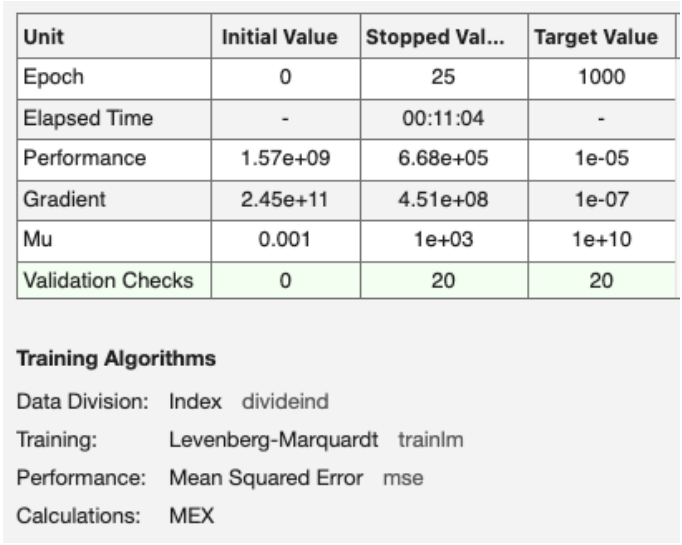
Figure 4: Temperature Variation Profile

The temperature profile shown in Figure 4 exhibits a low-amplitude sinusoidal waveform ranging between approximately 21 °C and 29 °C. Compared to irradiance, the temperature variation was designed with slower dynamics, and small-scale random noise was superimposed to create a more realistic measurement scenario. This configuration allows testing the temperature sensitivity of PV cells while generating a richer dataset for ANN training. When these two profiles are considered together, the aim is to comprehensively capture the PV panel's response to both sudden changes in irradiance and efficiency losses caused by temperature variations. In this way, the trained ANN model is expected to achieve better generalization across different operating scenarios.

At the end of the 600-second simulation, all essential variables including time, irradiance, temperature, PV panel current, PV output voltage, load voltage, and load current—were transferred to the MATLAB environment via To Workspace blocks. These data were then used to construct the input and output datasets for training the ANN-based MPPT model. The designed ANN architecture was optimized to learn the complex relationships between the environmental variables of the PV system and its electrical characteristics. In this configuration, irradiance and temperature were selected as network inputs, while PV panel voltage, current, and power represented the outputs.

The dataset was divided into 75% for training, 15% for validation, and 10% for testing, and a more robust network architecture was selected, consisting of three hidden layers with 32, 16, and 8 neurons, respectively. During the training process, the Levenberg–Marquardt (trainlm) algorithm was employed, the number of epochs was set to 1000, early stopping tolerance was limited to 20 epochs, and the target error value was optimized to 1×10^{-5}. To directly evaluate the generalization capability of the network, normalization of the inputs and outputs was disabled, and the Mean Squared Error (MSE) was used as the main performance metric. Compared to the

widely used conventional two-layer ANN architectures for PV systems, this structure with its wider neuron distribution and an additional hidden layer offers improved generalization ability on challenging datasets. Training, validation, and test sets were evaluated separately to prevent the model from overfitting. The graphical user interface (GUI) view of the ANN model developed using the simulation data obtained from PV panel control with the P&O algorithm is presented in Figure 5.



Unit	Initial Value	Stopped Val...	Target Value
Epoch	0	25	1000
Elapsed Time	-	00:11:04	-
Performance	1.57e+09	6.68e+05	1e-05
Gradient	2.45e+11	4.51e+08	1e-07
Mu	0.001	1e+03	1e+10
Validation Checks	0	20	20

Training Algorithms

Data Division: Index dividend
Training: Levenberg-Marquardt trainlm
Performance: Mean Squared Error mse
Calculations: MEX

Figure 5: Training Progress and Stopping Criteria of the ANN

When Figure 5 is examined, it can be observed that the training process was stopped after a total of 25 epochs, and the early stopping mechanism was triggered due to the lack of performance improvement on the validation set. The training, performed using the Levenberg–Marquardt algorithm, was completed in approximately 11 minutes and 4 seconds, reducing the initial mean squared error (MSE) from 1.57×10^{9} to 6.68×10^{5} at the stopping point. The gradient value decreased from 2.45×10^{11} to 4.51×10^{8}, while the adaptive learning parameter Mu increased from 0.001 to 1×10^{3}. Since no improvement was achieved in the validation set over 20 consecutive checks, the training was terminated early before reaching the target error value of 1×10^{-5}.

This result indicates that the network achieved a reasonable level of generalization without entering an overfitting trend. Thanks to the chosen data partitioning ratios, the training, validation, and test performances of the model could be monitored separately, and overfitting was successfully prevented using the early stopping strategy widely recommended in the literature. Moreover, the Levenberg–Marquardt algorithm, known for its

high speed and low error performance on small- to medium-sized datasets, provided an appropriate training strategy for this study.

After the completion of ANN training following the initial simulation, the trained ANN data were saved in ".mat" format within the Simulink environment, and subsequently a second model incorporating the ANN-based MPPT structure was developed. In this model, the ANN algorithm uses four input variables: irradiance, temperature, PV panel voltage, and PV panel current to determine the duty cycle adjustment of the IGBT. The MPPT_ANN codes used within the scope of this study are provided below.

```
function duty = ANN_MPPT_Block(Vpv, Ipv, G, T)
persistent prev_duty
if isempty(prev_duty)
 prev_duty = 0.5;
end
        x = [Vpv Ipv G T];
duty_raw = mppt_ann_simulink(x);
duty_raw = double(duty_raw(1));
duty_raw = max(0, min(1, duty_raw));
max_delta = 0.05;
delta = duty_raw - prev_duty;
if delta > max_delta
        duty_raw = prev_duty + max_delta;
elseif delta < -max_delta
        duty_raw = prev_duty - max_delta;
end
alpha = 0.1;
duty_smooth = alpha * duty_raw + (1 - alpha) * prev_duty;
prev_duty = duty_smooth;
duty = duty_smooth;
end
```

In the study, the Simulink codes used for MPPT_ANN were implemented as provided below.

```
function train_ann_from_simulation_stronger()
    clc; close all;
    disp('ANN training started...');
     dataFile = 'ann_training_data.mat';
    if ~isfile(dataFile)
       error('%s not found! Please run prepare_ann() first.', dataFile);
    end
    load(dataFile);
    X = Inputs';
    Y = Outputs';
    nSamples = size(X,2);
    if nSamples < 100
       error('Too few samples (%d samples available). More data required!', nSamples);
    end
    idx = randperm(nSamples);
    nTrain = round(0.75 * nSamples);
    nVal   = round(0.15 * nSamples);
    trainIdx = idx(1:nTrain);
    valIdx   = idx(nTrain+1 : nTrain+nVal);
    testIdx  = idx(nTrain+nVal+1:end);
    Xtrain = X(:,trainIdx);      Ytrain = Y(:,trainIdx);
    Xval   = X(:,valIdx);        Yval   = Y(:,valIdx);
    Xtest  = X(:,testIdx);       Ytest  = Y(:,testIdx);
    hiddenLayers = [32 16 8];
    net = fitnet(hiddenLayers, 'trainlm');
    net.trainParam.epochs    = 1000;
    net.trainParam.goal      = 1e-5;
    net.trainParam.max_fail  = 20;
    net.divideFcn = 'divideind';
    net.divideParam.trainInd                = 1:length(trainIdx);
    net.divideParam.valInd                  = length(trainIdx)+1:length(trainIdx)+length(valIdx);
    net.divideParam.testInd                 = length(trainIdx)+length(valIdx)+1:length(idx);
    net.inputs{1}.processFcns  = {};      % Disable input preprocessing
    net.outputs{2}.processFcns = {};  % Disable output normalization
    disp('Starting stronger ANN training...');
    [net, tr] = train(net, X, Y);
    if ~isempty(Xtest)
       Ypred = net(Xtest);
       testError = Ypred - Ytest;
       mseError                 = mean(testError.^2, 2);
       disp('Training completed!');
       fprintf('  Test MSE (Voltage): %.6f\n', mseError(1));
       fprintf('  Test MSE (Current): %.6f\n', mseError(2));
       fprintf('  Test MSE (Power)  : %.6f\n', mseError(3));
       figure('Name','ANN Test Performance','NumberTitle','off');
       subplot(3,1,1);
       plot(Ytest(1,:), 'b'); hold on; plot(Ypred(1,:), 'r--');
       ylabel('Voltage [V]'); legend('Actual','Predicted');
       subplot(3,1,2);
       plot(Ytest(2,:), 'b'); hold on; plot(Ypred(2,:), 'r--');
       ylabel('Current [A]'); legend('Actual','Predicted');
       subplot(3,1,3);
       plot(Ytest(3,:), 'b'); hold on; plot(Ypred(3,:), 'r--');
       ylabel('Power [W]'); legend('Actual','Predicted');
       sgtitle('ANN Test Results (Blue: Actual, Red: Predicted)');
    else
       warning('Test set is empty! Only Train+Validation performed.');
    end

    save('trained_ann_model_strong.mat', 'net');
    disp('Trained ANN model saved as: trained_ann_model_strong.mat');
end
```

At this stage, the MPPT_ANN structure utilizing the trained ANN model and the initially implemented MPPT_P&O algorithm were simulated comparatively in the Simulink environment. For the first comparison, a more stable irradiance and temperature profile different from the training data was used. The irradiance and temperature variation profiles employed in this

simulation are shown in Figure 6. In the first stable irradiance and temperature profiles used for the comparative evaluation of ANN- and P&O-based MPPT algorithms, the irradiance exhibits a sinusoidal variation within the range of 750–850 W/m², while the temperature shows a slow oscillation between 24.5 °C and 27.5 °C

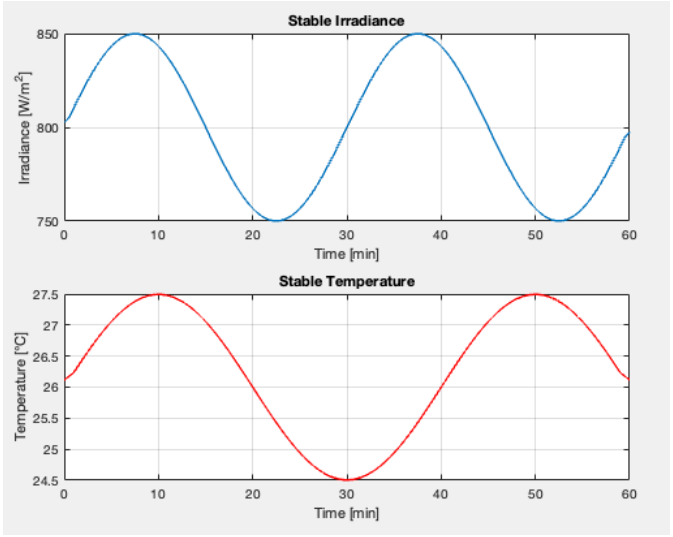


Figure 6: Stable Irradiance and Temperature Profiles Used in the Simulation

In Figure 7, the irradiance and temperature profiles used for the second evaluation of the ANN and P&O algorithms under challenging operating conditions are shown. The irradiance exhibits sudden and irregular variations, while the temperature profile includes stepwise transitions over a wide range along with high-frequency fluctuations. This configuration was designed to test the tracking performance of the algorithms under dynamic conditions.

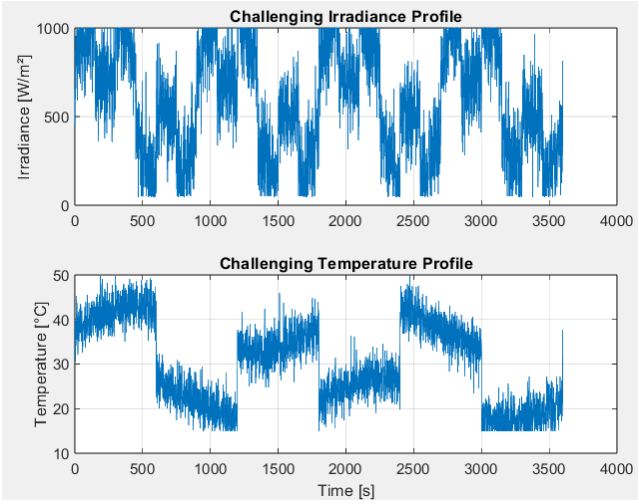


Figure 7: Challenging Irradiance and Temperature Profiles Used in the Simulation

RESULTS AND DISCUSSION

The simulation results revealed that the ANN-based MPPT algorithm demonstrated satisfactory performance compared to the conventional P&O algorithm under both stable and challenging environmental conditions. In Figure 8, the regression analysis results of the ANN training for the training, validation, and test datasets, as well as for the entire dataset, are shown. For the training dataset, the obtained regression coefficient was R = 0.98252, indicating that the network achieved a very high level of fit to the training data. For the validation set, an even higher correlation of R = 0.99156 was achieved, showing that the model did not overfit solely to the training data but also generalized well to the overall data distribution. For the test dataset, the regression coefficient R = 0.96139 demonstrated that the model maintained satisfactory prediction performance on previously unseen data. When considering the entire dataset collectively, the overall regression coefficient was R = 0.98082, confirming that the model possesses strong generalization capability.

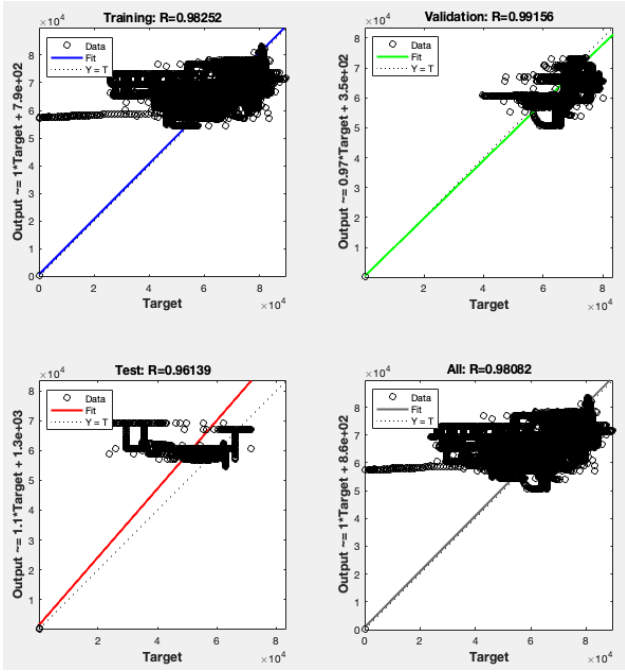


Figure 8: Regression Results of ANN Training on Training, Validation, Test, and Overall Datasets

In Figure 8, when examining the curve slopes and the target-output distribution, it is observed that the majority of the data points are located on or very close to the ideal Y = T line. However, particularly in the test dataset, a small number of outlier points show deviations, although these deviations do not significantly affect the overall performance of the model. These results demonstrate that the selected network architecture and the training process performed with the Levenberg–Marquardt algorithm successfully modeled the dynamics of a nonlinear and complex system such as the PV system. On the other hand, in Figure 9, the variations of three key performance parameters calculated at the end of each epoch during ANN training are shown. In the upper graph, it is observed that the gradient value, which started at the order of 10^{12} at the beginning of the training process, gradually decreased with the increasing number of epochs, reaching 4.51×10^{8}. This indicates that the weight updates of the network became progressively smaller, making the optimization process more stable. In the middle graph, the adaptive learning parameter Mu of the Levenberg–Marquardt algorithm is shown to increase during training, reaching a value of 1000 at epoch 25 and stabilizing at this level. This increase reflects that when no improvement is observed in the validation performance, the algorithm enlarges the step size and settles near a local minimum.

In the lower graph, the validation checks are presented, showing that the early stopping mechanism was activated because no improvement was achieved in 20 consecutive validation checks during the training process. When these results are considered together, it can be concluded that the ANN training was completed in accordance with the early stopping criteria and that the model achieved sufficient generalization capability without entering an overfitting trend. Moreover, the behavior of the gradient and Mu curves confirms that the Levenberg–Marquardt algorithm executed an appropriate optimization process for the given dataset.

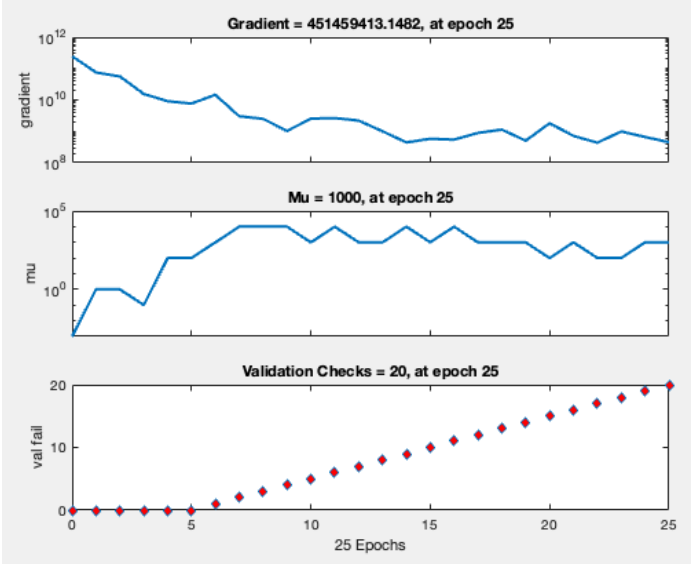


Figure 9: Epoch-wise Variation of Gradient, Mu Value, and Validation Checks During ANN Training

In Figure 10, the variation of the mean squared error (MSE) for the training, validation, and test datasets throughout the ANN training process is shown. It can be observed that the error value in the training dataset exhibits a consistent decreasing trend with the increasing number of epochs, while the validation and test datasets show a noticeable initial drop followed by a relatively stable curve. The best validation performance was achieved at epoch 5, where the validation error reached a level of 1.56×10^{7}. Although the target MSE of 1×10^{-5} was not reached, the lack of improvement in the validation error beyond a certain point triggered the early stopping mechanism.

This result indicates that the Levenberg–Marquardt algorithm achieved rapid convergence during the initial epochs, providing the best

validation performance in a short time. The stabilization of the validation error after epoch 5 and the absence of further improvement suggest that continuing the training could potentially lead to overfitting.

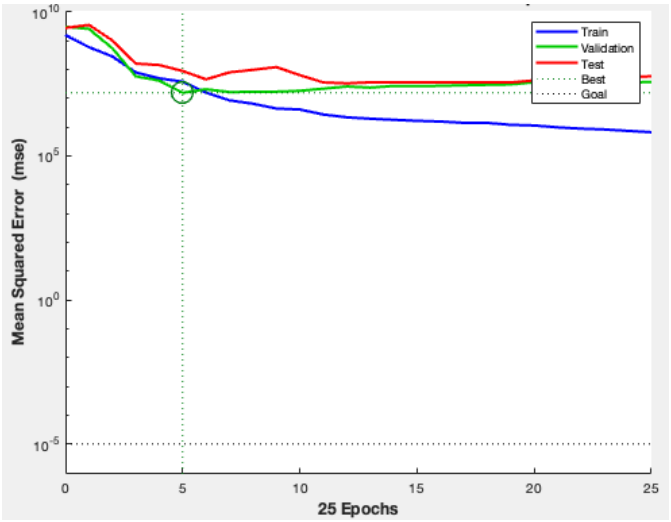


Figure 10: Epoch-wise Mean Squared Error (MSE) Variation and Best Validation Performance During ANN Training

In Figure 11, a histogram showing the error distribution obtained after ANN training is presented. The histogram illustrates the frequency distribution of the differences between the predicted outputs and the actual target values (error = target – output) for the training set (blue), validation set (green), and test set (red). The majority of the errors are concentrated near the orange reference line indicating zero error, demonstrating that the model generally has high prediction accuracy. Most of the 20 error bins in the graph exhibit very low frequencies, indicating that the errors are predominantly clustered within a narrow range. Although a small number of data points show larger error values, these outliers do not significantly affect the overall performance of the system.

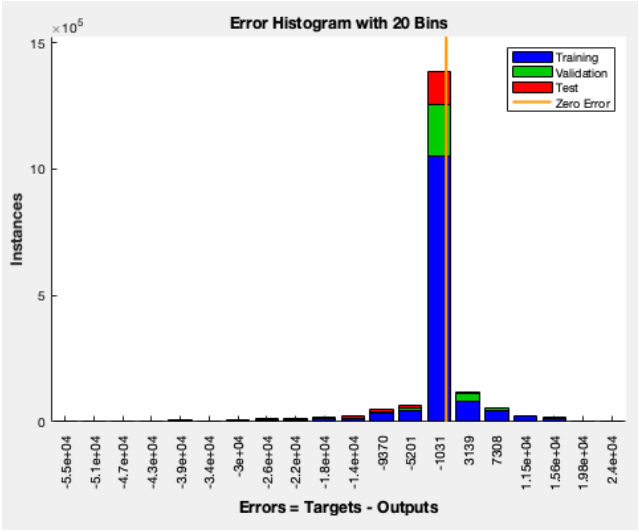


Figure 11: Error Histogram of Training, Validation, and Test Data Sets for ANN Training

In Figure 12, the voltage, current, and power tracking performances of the ANN-based MPPT algorithm and the conventional P&O algorithm are shown under stable irradiance and temperature profiles with smooth variations. Both algorithms exhibited almost identical responses under stable conditions, resulting in overlapping curves for PV panel voltage, load current, and generated power. This demonstrates that under stable conditions, the ANN is able to reproduce the reference behavior learned during training as successfully as the P&O algorithm.

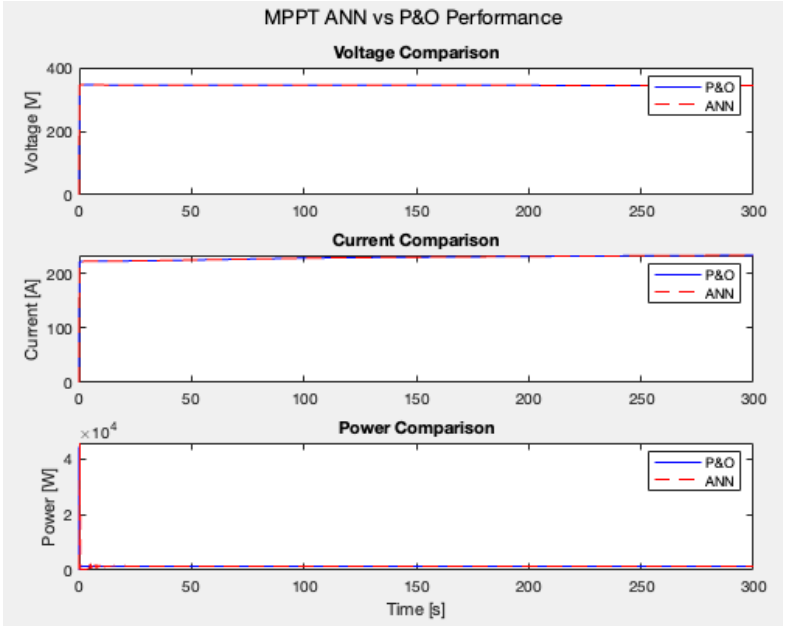


Figure 12: Comparison of ANN- and P&O-Based MPPT Performance Under Stable Irradiance and Temperature Conditions

In Figure 13, the performances of the ANN and P&O algorithms are compared under challenging irradiance and temperature profiles that include sudden changes and irregular noise components. In this scenario, although the voltage and power tracking curves of both algorithms are similar in terms of overall trend, the ANN algorithm produces smoother transitions in current variations and partially filters out abrupt fluctuations. In contrast, the P&O algorithm responds more sensitively to instantaneous changes, resulting in sharper oscillations.

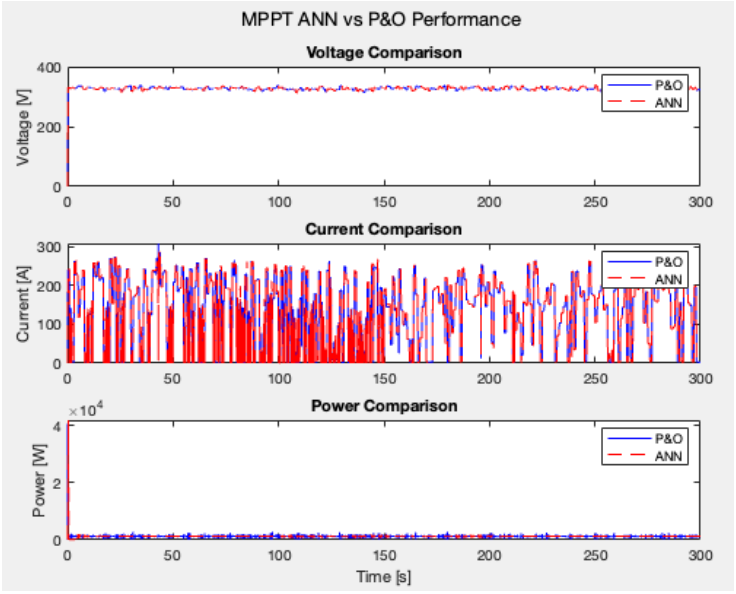


Figure 13: Comparison of ANN- and P&O-Based MPPT Performance Under Challenging Irradiance and Temperature Conditions

Under stable conditions, the voltage, current, and power curves of the ANN and P&O algorithms almost completely overlapped, and the ANN model successfully reproduced the maximum power point behavior learned during training with high accuracy. This result indicates that the ANN can provide tracking performance equivalent to the P&O algorithm in operating scenarios similar to the data distribution used during its training.

In the second scenario, which included sudden and highly dynamic changes, the ANN-based controller produced smoother transitions compared to the P&O algorithm and partially filtered out the noise effects caused by rapid irradiance and temperature variations. Due to its sensitivity to instantaneous changes, the P&O algorithm generated sharper and higher-frequency oscillations, whereas the ANN model transformed these effects into a more balanced tracking performance thanks to its generalization capability. Furthermore, in terms of power tracking, the ANN maintained a more stable output closer to the maximum power point, thereby minimizing power losses during sudden irradiance drops.

The obtained findings are consistent with similar studies in the literature. ANN-based MPPT controllers achieve comparable performance to the conventional P&O algorithm under conditions similar to the training data distribution, while providing more stable responses and better resistance to fluctuations in sudden and complex environmental changes. However, the success of the ANN is directly related to the diversity of the training dataset and its generalization capability; therefore, using broader and more diverse

datasets during training can be considered a key factor to further improve this performance.

REFERENCES

Bastidas-Rodriguez, J. D., Franco, E., Petrone, G., Ramos-Paja, C. A., & Spagnuolo, G. (2014). Maximum power point tracking architectures for photovoltaic systems in mismatching conditions: A review. *IET Power Electronics*, Vol. 7, pp. 1396–1413. Institution of Engineering and Technology. https://doi.org/10.1049/iet-pel.2013.0406

Bollipo, R. B., Mikkili, S., & Bonthagorla, P. K. (2020). Critical Review on PV MPPT Techniques: Classical, Intelligent and Optimisation. *IET Renewable Power Generation*, *14*(9), 1433–1452. https://doi.org/10.1049/IET-RPG.2019.1163

Bouadjila, T., Khelil, K., Rahem, D., & Berrezzek, F. (2023). Improved Artificial Neural Network Based MPPT Tracker for PV System under Rapid Varying Atmospheric Conditions. *Periodica Polytechnica Electrical Engineering and Computer Science*, *67*(2), 149–159. https://doi.org/10.3311/PPee.20824

Elobaid, L. M., Abdelsalam, A. K., & Zakzouk, E. E. (2015). Artificial neural network-based photovoltaic maximum power point tracking techniques: A survey. *IET Renewable Power Generation*, *9*(8), 1043–1063. https://doi.org/10.1049/iet-rpg.2014.0359

Haddou, A., Tariba, N. E., Ikken, N., Bouknadel, A., Omari, H. E. L., & Omari, H. E. L. (2020). Comparative study of new MPPT control approaches for a photovoltaic system. *International Journal of Power Electronics and Drive Systems*, *11*(1), 251–262. https://doi.org/10.11591/ijpeds.v11.i1.pp251-262

Hammami, M., Ricco, M., Ruderman, A., & Grandi, G. (2019). Three-phase three-level flying capacitor PV generation system with an embedded ripple correlation control MPPT algorithm. *Electronics (Switzerland)*, *8*(2). https://doi.org/10.3390/electronics8020118

Mathworks. (2025). MATLAB for Students. Retrieved July 25, 2025, from https://www.mathworks.com/products/matlab/student.html

Qi, P., Xia, H., Cai, X., Yu, M., Jiang, N., & Dai, Y. (2024). Novel Global MPPT Technique Based on Hybrid Cuckoo Search and Artificial Bee Colony under Partial-Shading Conditions. *Electronics (Switzerland)*, *13*(7). https://doi.org/10.3390/electronics13071337

Roy, R. B., Rokonuzzaman, M., Amin, N., Mishu, M. K., Alahakoon, S., Rahman, S., … Pasupuleti, J. (2021). A Comparative Performance Analysis of ANN Algorithms for MPPT Energy Harvesting in Solar PV System. *IEEE Access*, *9*, 102137–102152. https://doi.org/10.1109/ACCESS.2021.3096864

Sukanya Satapathy, S., & Kumar, N. (2020). Framework of maximum power point tracking for solar PV panel using WSPS technique. *IET Renewable Power Generation*, *14*(10), 1668–1676. https://doi.org/10.1049/iet-rpg.2019.1132

Tahiri, F., Harrouz, A., & Erkan BOSTANCI, G. (n.d.-a). A Comprehensive Comparison of Two MPPT Techniques (P&O, SMC) for Photovoltaic Systems. In *ALGERIAN JOURNAL OF SIGNALS AND SYSTEMS (AJSS)* (Vol. 9).

Tahiri, F., Harrouz, A., & Erkan BOSTANCI, G. (n.d.-b). A Comprehensive Comparison of Two MPPT Techniques (P&O, SMC) for Photovoltaic Systems. In *ALGERIAN JOURNAL OF SIGNALS AND SYSTEMS (AJSS)* (Vol. 9).

Tajaldin, K. H., & Motlak, H. J. (2024). Design and Implementation DC/DC Luo Converter Controlled by Adaptive Fractional PI and P&O MPPT. *Journal Europeen Des Systemes Automatises*, *57*(1), 201–210. https://doi.org/10.18280/jesa.570120

Youssfi, A., Alioui, A., & El Kadi, Y. A. (2024). Study, simulation and realization of a fuzzy logic-based MPPT controller in an isolated DC microgrid. *Indonesian Journal of Electrical Engineering and Computer Science*, *34*(3), 1420–1433. https://doi.org/10.11591/ijeecs.v34.i3.pp1420-1433

Numerical Analysis of Hydrogen-Enriched Diesel Combustion Performance and Emission Characteristics

Serdar HALİS*
Furkan YAZICI
Şevket Onur KILIÇ

* Assoc. Prof. Dr.; Pamukkale University, Faculty of Techology, Department of Automotive Engineering.
shalis@pau.edu.tr ORCID No: 0000-0002-6099-7223

ABSTRACT

In this study, the effects of hydrogen as a supplementary fuel on engine performance and exhaust emissions were investigated in a single-cylinder, water-cooled, naturally aspirated diesel engine. The test engine was modeled using Ricardo WAVE software, and the simulation results were validated by comparing them with experimental data. After successful model validation, dual-fuel combustion analyses were carried out with hydrogen volume ratios ranging from 0% to 50%. The results showed that the peak in-cylinder pressure increased by approximately 9.6% with higher hydrogen content. Additionally, the heat release rate (HRR) increased significantly, indicating improved combustion efficiency. Hydrocarbon (HC) emissions decreased by 54.5%, which reflects more complete and homogeneous combustion. However, due to elevated in-cylinder temperatures and pressures, nitrogen oxides (NO_x) emissions increased by about 23.1%. In conclusion, hydrogen enrichment in diesel engines can enhance engine performance and reduce HC emissions. However, the resulting increase in NO_x emissions requires careful control. This study demonstrates the potential of hydrogen-diesel dual-fuel systems for cleaner and more efficient internal combustion engine technologies and contributes to future development efforts in this field.

Keywords – Dual Fuel, Hydrogen, Diesel, Combustion, Emission.

INTRODUCTION

The rapid pace of industrialization has significantly increased the demand for reliable and efficient power sources, positioning internal combustion engines (ICEs) as indispensable components of modern industry and daily life. Among ICEs, diesel engines are particularly favored for their superior thermal efficiency, robust durability, and high torque output, making them essential in diverse applications such as marine transportation, agriculture, and heavy-duty machinery (Heywood, 2018). Despite these advantages, the environmental sustainability of diesel engines has come under scrutiny due to their reliance on fossil-based fuels and the resultant emission of harmful pollutants. Diesel combustion is known to release significant quantities of nitrogen oxides (NO_x), hydrocarbons (HC), and carbon monoxide (CO), all of which contribute to air pollution and exacerbate the effects of global climate change (IPCC, 2021; EPA, 2020).

The increasing urgency to address environmental degradation, coupled with stringent international emission regulations—such as the European Union's Euro standards and the U.S. Environmental Protection Agency (EPA) guidelines—has intensified global efforts to reduce dependence on fossil fuels and transition toward cleaner, renewable alternatives (European

Commission, 2019). In this context, hydrogen has emerged as a promising alternative fuel due to its zero-carbon combustion, high diffusivity, wide flammability limits, and high flame speed. When used in dual-fuel configurations with diesel, hydrogen not only enhances combustion efficiency but also offers notable reductions in key emissions, potentially aligning internal combustion engine technology with modern environmental and regulatory expectations (Gültekin et al., 2024; Saravanan et al., 2008a).

However, while hydrogen offers clear environmental and performance advantages, its application in diesel engines requires comprehensive investigation due to complex combustion dynamics and the potential for increased NO_x formation under high-temperature conditions. Therefore, detailed computational and experimental studies are essential to understand the trade-offs and optimize the use of hydrogen in such systems (Kong et al. 2006).

Numerous recent studies have explored the performance and emission characteristics of hydrogen-diesel dual-fuel engines using both experimental and simulation-based approaches. Al-Asadi (2018), employing Ricardo WAVE, reported that higher hydrogen substitution levels led to increased NO_x emissions and cylinder pressure, while CO emissions were reduced, with minimal deviation between experimental and simulated results. Gültekin and Ciniviz (2023) investigated the effect of hydrogen ratio under varying load conditions and observed similar trends namely, increased NO_x and cylinder pressure alongside reductions in CO, CO_2, and HC emissions. Akhtar et al. (2025) also utilized Ricardo WAVE to simulate hydrogen-diesel combustion and reported a corresponding rise in NO_x and engine torque with higher hydrogen content.

Farzam and McTaggart-Cowan (2024), using GT-SUITE, conducted simulations across different load conditions and noted improved thermal efficiency and reduced CO_2 emissions. Their results, which showed good agreement between experimental and simulation data in terms of heat release rate, reinforced the potential of hydrogen for performance enhancement. An et al. (2013), using the CHEMKIN code, confirmed the increase in NO_x and the decrease in CO emissions with increasing hydrogen ratio, alongside validation of in-cylinder pressure and HRR predictions. Qin et al. (2020) experimentally demonstrated that higher hydrogen content at a constant engine speed (2600 rpm) increased peak pressure, heat release rate, and cumulative heat release. Similarly, Lilik et al. (2010) reported increased NO_x and reduced CO and CO_2 emissions under varying speeds and loads, further validating the emissions-reduction capability of hydrogen.

Building upon these foundations, the present study aims to investigate the combustion behavior and emission performance of a hydrogen–diesel dual-fuel engine under varying hydrogen substitution ratios. To achieve this, a naturally aspirated single-cylinder diesel engine was modeled using Ricardo WAVE software. The model was first validated by comparing the

simulated in-cylinder pressure profiles against experimental data, ensuring the accuracy of the numerical framework. Upon successful validation, the model was used to simulate dual-fuel operation at multiple hydrogen blending ratios.

Combustion parameters such as in-cylinder pressure and heat release rate, as well as emission characteristics including NO_x and HC levels, were systematically analyzed. The findings of this study are expected to provide valuable insights into the advantages and limitations of hydrogen utilization in diesel engines. In doing so, this research contributes to the broader effort to develop cleaner and more efficient engine technologies and supports the ongoing global transition toward sustainable energy systems.

MATERIALS AND METHOD

In this study, a one-dimensional numerical modeling approach was adopted to evaluate the impact of hydrogen addition into diesel fuel under a dual-fuel combustion strategy. The simulation work was conducted using Ricardo WAVE, a widely used engine simulation platform for analyzing the thermodynamic and fluid-dynamic behavior of internal combustion engines. The reference engine modeled in the study is a four-stroke, water-cooled, naturally aspirated, single-cylinder diesel engine. A detailed engine model was constructed in Ricardo WAVE to accurately reflect the physical engine's geometry and operating characteristics, including bore, stroke, compression ratio, valve events, fuel injection system, and intake/exhaust configurations.

To ensure model reliability, a validation process was performed by comparing simulated in-cylinder pressure traces with experimental data obtained under standard diesel-only operation. The agreement between simulation and experimental results confirmed that the model could be used for further analysis with high confidence. After validation, the simulation model was updated to implement a dual-fuel mode, where hydrogen was blended directly with diesel fuel at varying mass fractions (10%, 20%, 30%, 40% and 50%), while keeping the total injected fuel mass constant. This method reflects a partially premixed combustion regime and allows the investigation of the effects of increasing hydrogen substitution on combustion behavior and emission characteristics.

Simulations were conducted under fixed engine speed and load conditions to isolate the influence of fuel composition. For each hydrogen ratio, key combustion parameters—including peak cylinder pressure, heat release rate (HRR), and emission outputs (NO_x and HC)—were analyzed. The HRR was calculated based on simulated pressure and instantaneous cylinder volume using the first law of thermodynamics. The experimental setup, used to collect reference data for model validation, included a

piezoelectric pressure transducer, crank angle encoder, and a high-speed data acquisition system. A schematic of the test setup is presented in Figure 1, showing the measurement and instrumentation layout. In addition the test engine's specifications are given in Table 1.

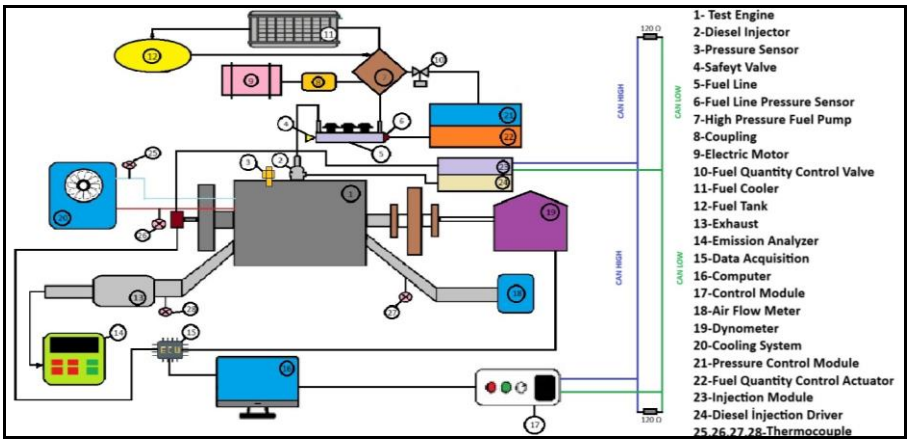


Figure 1: Test Setup

Table 1. Experimental engine specifications

Test Engine	Antor AD510 BS
Engine Type	Diesel Engine
Number of cylinders	1
Cylinder diameter	85 mm
Stroke	90 mm
Connecting rod length	145 mm
Cylinder volume	510 cm^3
Max. torque	32.8 Nm @ 1800 rpm
Max. power	12 HP
Compression ratio	17.8:1

The simulation framework was developed using a simplified single-cylinder engine architecture comprising a single fuel injector, one combustion chamber, one intake valve, one exhaust valve, an engine block, and four discrete flow channels. The intake and exhaust systems were modelled with separate flow paths to capture the dynamics of gas exchange accurately. Simulations were performed using Ricardo WAVE, which utilizes a two-zone thermodynamic model to represent in-cylinder processes. This model provides a more realistic estimation of temperature gradients and fuel–air composition stratification during the combustion cycle.

To represent the combustion behaviour of hydrogen-enriched diesel fuel mixtures under a dual-fuel strategy, a multi-component, multi-fuel

Wiebe function-based sub-model was used. This model is capable of predicting the heat release rate for blended fuels exhibiting different combustion characteristics. In this study, hydrogen was blended directly into the diesel fuel at mass fractions of 10-50% and the homogeneous mixture was injected using a single injector to reflect dual-fuel operation.

In-cylinder heat transfer was calculated using the Woschni heat transfer correlation, which adjusts the heat transfer coefficient based on engine operating conditions such as pressure, gas temperature, and piston speed. The Woschni model is given by the following empirical equation (Woschni, 1967):

$$h = 3.26 \cdot B^{-0.2} \cdot p^{0.8} \cdot T^{-0.55} \cdot w^{0.8}$$

where h=instantaneous heat transfer coefficient ($W/m^2 \cdot K$), B=cylinder bore (m), p=in-cylinder pressure (bar), T=in-cylinder gas temperature (K) and w=characteristic velocity (m/s), accounting for piston motion and combustion-induced turbulence.

This formulation allows for dynamic estimation of convective heat transfer throughout the engine cycle, enhancing the accuracy of thermal boundary condition predictions, especially in combustion modes affected by hydrogen addition.

To establish the reliability and predictive strength of the numerical model employed in this study, a comprehensive validation step was first performed against experimental reference data. Specifically, in-cylinder pressure traces as a function of crank angle degree (CAD) obtained from simulations were compared with corresponding measurements collected from the test engine operating solely on diesel fuel, prior to any hydrogen addition.

The validation emphasized key combustion indicators such as peak pressure values, the timing of pressure maxima, and pressure rise rates. The strong agreement observed between the simulated and experimental results indicated that the chosen physical sub-models, boundary conditions, and computational mesh settings were suitably selected and well-calibrated.

The minor discrepancies between the two datasets remained within an acceptable range confirming the robustness of the model. This validation step established a sound computational foundation for conducting further simulations involving hydrogen/diesel dual-fuel combustion, in which hydrogen was directly blended into the diesel fuel at varying proportions. The simulation conditions used for the validation process are presented in Table 2.

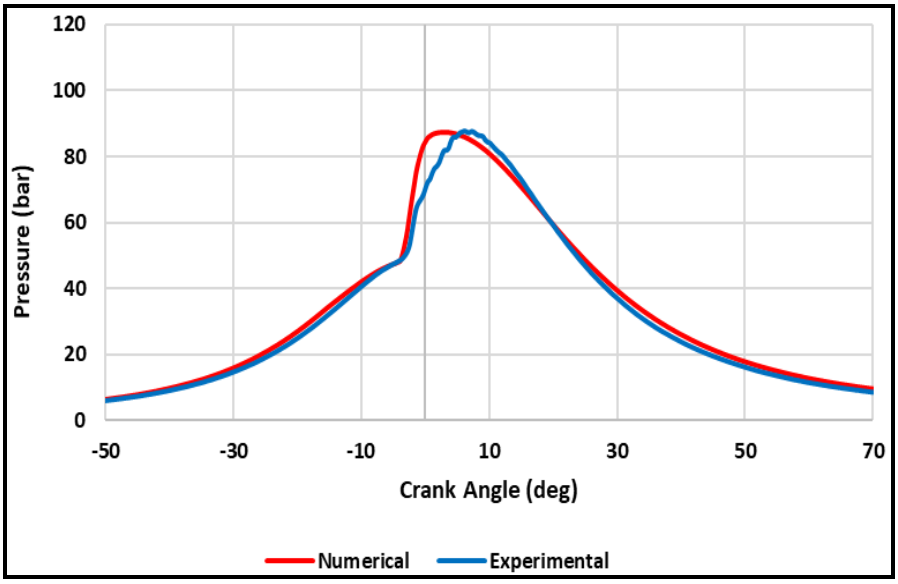


Figure 2: Model Validation

Table 2. Simulation parameters

Intake air temperature (K)	300
Engine speed (rpm)	2000
Compression ratio	17.8:1
Injection pressure (bar)	650
Start of injection (°CA)	-16
Amount of fuel injected (mg/cycle)	20.5
Connecting rod length (mm)	145
Spray angle (°)	77
Injection duration (°CA)	8.25

RESULTS AND DISCUSSION

In this section, the changes in combustion characteristics and emissions are examined by gradually increasing the proportion of hydrogen added to the fuel in a single-cylinder diesel engine. Following the validation of experimental data using the Ricardo WAVE software analyses were carried out for varying hydrogen percentages: 0% (D100), 10% (D90H10), 20% (D80H20), 30% (D70H30), 40% (D60H40), and 50% (D50H50).

Numerous studies have shown that hydrogen addition in dual-fuel diesel engines significantly influences engine performance and combustion behavior (Verhelst et al., 2009; Bari and Esmaeil, 2010; Saravanan et al., 2008b). As hydrogen possesses higher diffusivity and a wider flammability

limit than diesel, it tends to promote a faster and more complete combustion process (Das, 2002). In agreement with these findings, the current results show that increasing the hydrogen ratio enhances both the maximum cylinder pressure and HRR, as illustrated in Figures 3 and 4.

Specifically, the highest in-cylinder pressure was recorded as approximately 93.41 bar at 6.1°CA after top dead center (aTDC) for the 50% hydrogen blend, whereas the lowest was 87.37 bar at 2.88°CA aTDC for neat diesel (0% hydrogen). Similar trends have been reported in prior research, where increasing hydrogen concentration resulted in a delayed yet more intense pressure rise due to improved combustion efficiency (Ji and Wang, 2009; Nagarajan et al., 2000).

Moreover, when comparing the combustion behavior of pure diesel (0% hydrogen) and a 50% hydrogen blend, it was observed that the peak pressure occurred 3.22°CA later with hydrogen addition, indicating a shift in combustion phasing commonly noted in dual-fuel hydrogen applications (Bari and Esmaeil, 2010). The HRR profile also confirms that more energy is released with higher hydrogen content. The maximum HRR of 118.7 J/°CA at 2.15°CA before TDC (bTDC) was achieved at 50% hydrogen, whereas the lowest HRR was 105.8 J/°CA at the same crank angle for the 10% hydrogen blend. This observation aligns with earlier studies that associate higher hydrogen content with an increased rate of combustion and thermal release (Saravanan et al., 2008b).

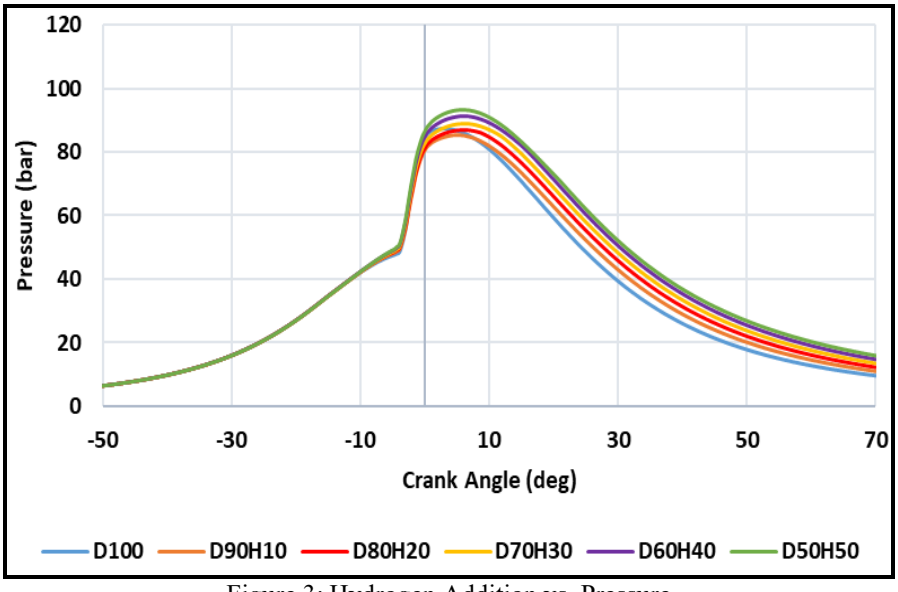


Figure 3: Hydrogen Addition vs. Pressure

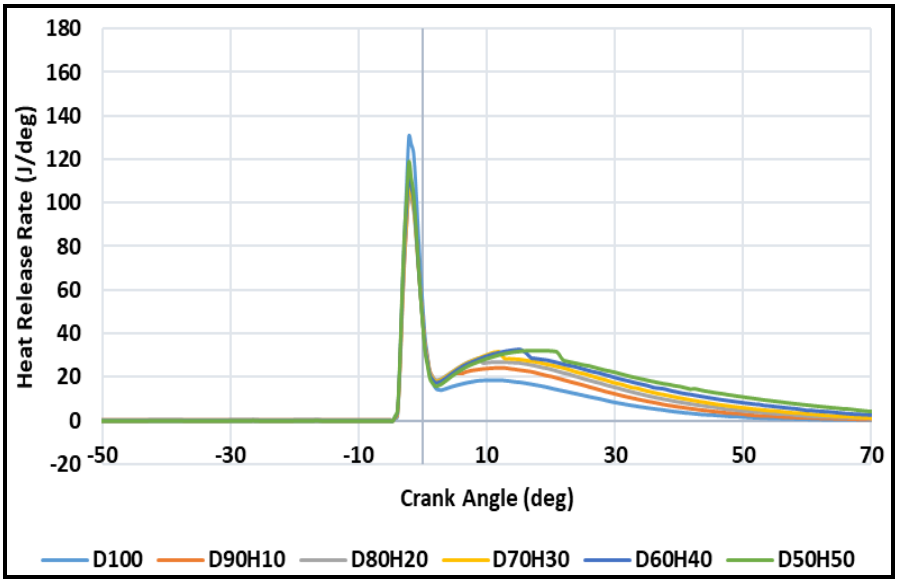


Figure 4: Hydrogen Addition vs. Heat Release Rate

Figure 5 presents the variation of CA50 and combustion duration values for different hydrogen-enriched diesel fuel blends. As the proportion of hydrogen increases, both CA50 and combustion duration exhibit a notable increasing trend. Specifically, the CA50 value rises from 5.4°CA in the D100 case to 18.7°CA in the D50H50 blend, corresponding to an approximate 246% increase. Similarly, the combustion duration extends from 5.72°CA to 19°CA, reflecting a growth of nearly 232%. This behavior suggests that combustion phasing is significantly delayed and the combustion process becomes prolonged as more hydrogen is added. The delay can be attributed to the reduction in the proportion of diesel, which acts as the primary ignition source, and the modified in-cylinder mixing dynamics induced by hydrogen. Similar trends have been reported in recent studies; for instance, Farzam and McTaggart-Cowan (2022) observed delayed combustion phasing with hydrogen addition due to changes in ignition delay and reactivity of the charge mixture. Furthermore, An et al. (2021) and Kumar et al. (2023) emphasized that although hydrogen has a high flame speed, the decreased diesel pilot quantity and formation of a more stratified mixture can contribute to prolonged combustion. These findings corroborate the results of the present study, emphasizing that while hydrogen enrichment can influence combustion characteristics positively in terms of efficiency, its impact on combustion timing and duration must be carefully managed in dual-fuel diesel engine applications.

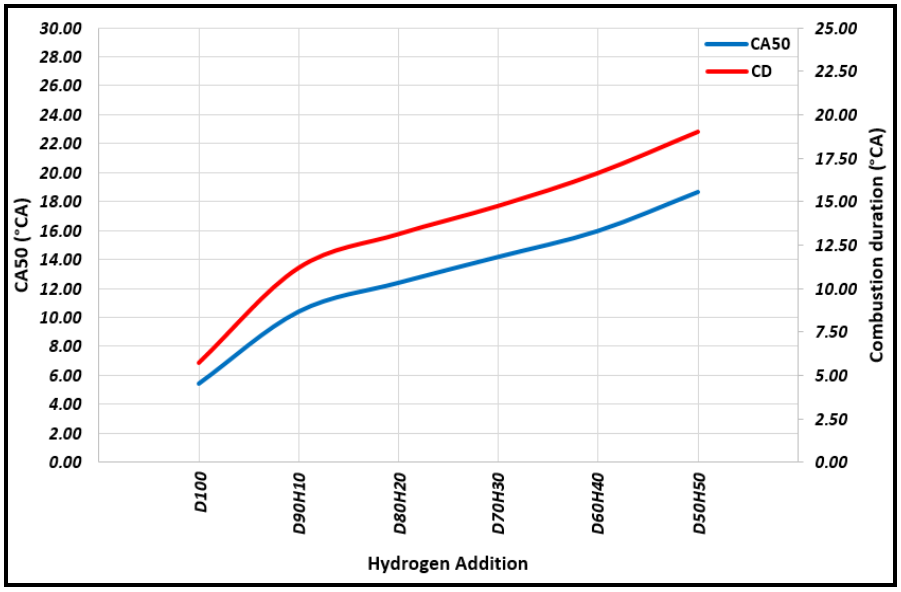


Figure 5: The variation of CA50 and combustion duration with hydrogen addition

The effects of hydrogen proportion in the fuel on exhaust emission characteristics in dual-fuel engines are presented in Figures 6 and 7. In general, an increase in the hydrogen percentage led to a rise in NO_x emissions and a reduction in HC. These findings are consistent with previous studies reporting that hydrogen-enriched combustion typically results in higher flame temperatures, which favour NO_x formation, while the absence of carbon in hydrogen reduces HC emissions due to more complete combustion (Verhelst et al., 2009; Saravanan et al., 2008b; Ji and Wang, 2009). As illustrated in the emission graphs, the highest NO_x and lowest HC values were recorded at 50% hydrogen addition. Specifically, the NO_x emission increased from 1041.13 ppm at 0% hydrogen to 1195.43 ppm at 50% hydrogen. This increase can be attributed to the elevated combustion temperature associated with hydrogen's high flame speed and low ignition energy (Bari and Esmaeil, 2010). In parallel, the HC emissions decreased markedly as the hydrogen content increased. The HC value dropped from 50.14 ppm at 0% hydrogen to 22.78 ppm at 50% hydrogen. This trend is also in agreement with earlier findings indicating that the use of hydrogen improves oxidation of unburned hydrocarbons and reduces the duration of the combustion process, thereby minimizing HC formation (Das, 2002; Nagarajan et al., 2000).

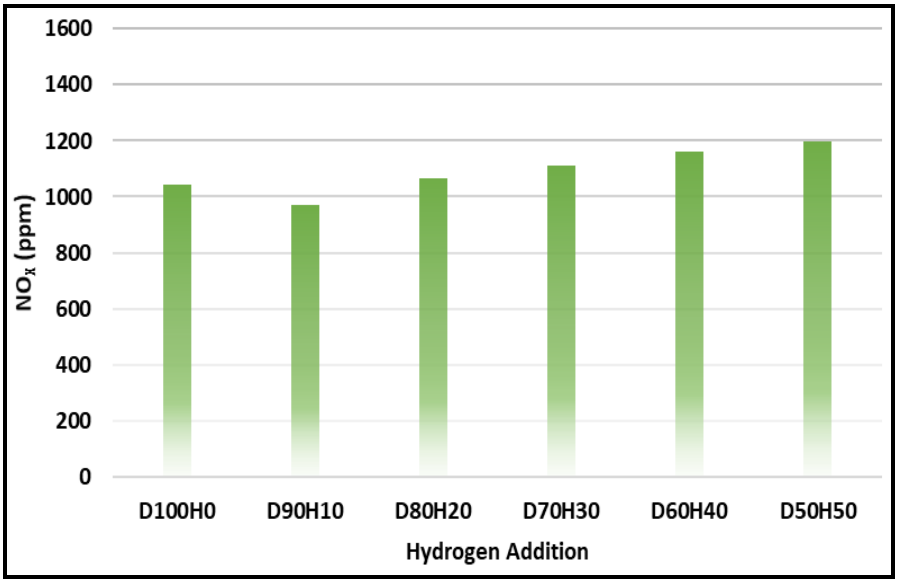


Figure 6: Hydrogen Addition vs. NO_x Emission

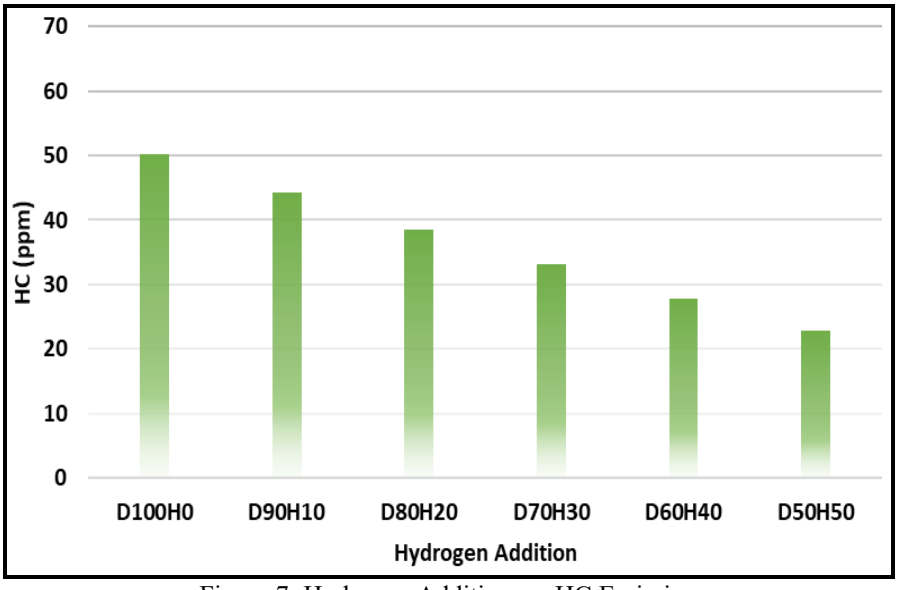


Figure 7: Hydrogen Addition vs. HC Emission

CONCLUSIONS

In this study, the effects of hydrogen addition to the fuel on engine performance and exhaust emission characteristics were investigated in a single-cylinder, water-cooled, naturally aspirated diesel engine. The test engine was first modelled using Ricardo WAVE software, and simulation results were validated by comparing them with experimental data to ensure accuracy. After successful model validation, different fuel compositions were analyzed (ranging from 0% to 50% hydrogen by volume) in dual-fuel operation. The simulation results revealed several key findings regarding combustion behavior and emissions:

The peak in-cylinder pressure increased by approximately 9.6%, from 85.21 bar (at 10% hydrogen and 5°CA) to 93.41 bar (at 50% hydrogen and 6.1°CA). Interestingly, the pressure initially dropped when hydrogen was added at 10%, suggesting suboptimal combustion due to hydrogen's low volumetric energy density at low concentrations. However, with higher hydrogen ratios, the high flame speed and fast-burning characteristics of hydrogen dominated, leading to more efficient combustion and higher peak pressures.

A clear upward trend was observed in the HRR values with increasing hydrogen content, indicating enhanced combustion efficiency and higher energy release. The maximum HRR reached 118.7 J/°CA at 2.15°CA bTDC with 50% hydrogen addition, compared to lower values at lower hydrogen concentrations.

As the hydrogen ratio increased in the dual-fuel blends, both CA50 and combustion duration exhibited significant rises. Compared to pure diesel, the CA50 value increased by approximately 246%, while the combustion duration extended by about 232% at 50% hydrogen enrichment. These findings indicate a substantial delay in combustion phasing and a lengthening of the combustion process, underscoring the importance of optimizing ignition strategies when employing hydrogen-diesel dual-fuel systems.

HC emissions decreased significantly, by approximately 54.5%, as the hydrogen share increased. The HC level dropped from 50.14 ppm (0% hydrogen) to 22.78 ppm (50% hydrogen). This can be attributed to hydrogen's superior flame speed, which promotes more complete and homogeneous combustion, as well as the dilution of diesel—which is the primary source of hydrocarbon emissions. In contrast, NO_x emissions increased by around 23.1%, rising from 971.5 ppm at 0% hydrogen to 1195.5 ppm at 50% hydrogen. This increase is likely due to the elevated in-cylinder temperatures and pressures associated with hydrogen combustion, which favour thermal NO_x formation mechanisms. Overall, the study demonstrates that while hydrogen enrichment can significantly improve

combustion quality and reduce HC emissions, it also necessitates careful control strategies to mitigate NO_x emissions.

In conclusion, hydrogen has proven to be a promising alternative fuel for enhancing diesel engine performance and reducing unburned hydrocarbon emissions. However, its influence on NO_x emissions must be carefully managed, potentially through exhaust after-treatment systems or advanced combustion control techniques. The outcomes of this study contribute to a better understanding of dual-fuel hydrogen-diesel combustion and support its potential application in cleaner and more efficient internal combustion engines.

REFERENCES

Akhtar, M. U. S., Asfand, F., Khan, M. I., Mishra, R., & Ball, A. D. (2025). Performance and emissions characteristics of hydrogen-diesel dual-fuel combustion for heavy-duty engines. International Journal of Hydrogen Energy, 143, 454-467.

Al-Asadi, W. (2018). A study on the effects of combined diesel-hydrogen combustion on diesel engines using experimental and simulation techniques PhD Thesis, Brunel University London.

An, H., Yang, W., & Maghbouli, A. (2021). A review on hydrogen dual-fuel combustion engines: Effect of fuel properties and injection strategies on performance and emissions. Renewable and Sustainable Energy Reviews, 149, 111331.

An, H., Yang, W. M., Maghbouli, A., Li, J., Chou, S. K., & Chua, K. J. (2013). A numerical study on a hydrogen assisted diesel engine. International Journal of Hydrogen Energy, 38(6), 2919-2928.

Bari, S., & Esmaeil, M. M. (2010). Effect of H_2/O_2 addition in increasing the thermal efficiency of a diesel engine. Fuel, 89(2), 378-383.

Das, L. M. (2002). Hydrogen engine: research and development programmes in Indian Institute of Technology, Delhi. International Journal of Hydrogen Energy, 27(9), 953-965.

EPA. (2020). Overview of Greenhouse Gases. United States Environmental Protection Agency. https://www.epa.gov, Last access date: 13.07.2025

European Commission. (2019). EURO emission standards. https://ec.europa.eu, Last access date: 13.07.2025.

Farzam, H., & McTaggart-Cowan, G. (2022). A comparative study of combustion characteristics in hydrogen–diesel dual-fuel operation under varying injection strategies. International Journal of Hydrogen Energy, 47(63), 26717–26730.

Farzam, R., & McTaggart-Cowan, G. (2024). Hydrogen-diesel dual-fuel combustion sensitivity to fuel injection parameters in a multi-cylinder compression-ignition engine. International Journal of Hydrogen Energy, 49, 850-867.

Gültekin, N., & Ciniviz, M. (2023). Experimental investigation of the effect of hydrogen ratio on engine performance and emissions in a compression ignition single cylinder engine with electronically controlled hydrogen-diesel

dual fuel system. International Journal of Hydrogen Energy, 48(66), 25984-25999.
Gültekin, N., Gülcan, H. E., & Ciniviz, M. (2024). Investigation of the effects of hydrogen energy ratio and valve lift amount on performance and emissions in a hydrogen-diesel dual-fuel compression ignition engine. International Journal of Hydrogen Energy, 49, 352-366.
Heywood, J. B. (2018). Internal combustion engine fundamentals. McGraw-Hill Education.
IPCC. (2021). Climate Change 2021: The Physical Science Basis. Intergovernmental Panel on Climate Change. https://www.ipcc.ch, Last access date: 13.07.2025
Ji, C., & Wang, S. (2009). Effect of hydrogen addition on the combustion and emissions performance of a diesel engine. International Journal of Hydrogen Energy, 34(18), 8161-8170.
Kong, S.-C., Reitz, R. D., & Han, Z. (2006). The use of hydrogen in dual-fuel engines. International Journal of Hydrogen Energy, 31(15), 2097-2108.
Kumar, B. R., Singh, P., & Varma, A. (2023). Impact of hydrogen enrichment on combustion phasing and emission characteristics in a diesel engine: An experimental study. Fuel, 346, 128313.
Lilik, G. K., Zhang, H., Herreros, J. M., Haworth, D. C., & Boehman, A. L. (2010). Hydrogen assisted diesel combustion. International journal of hydrogen energy, 35(9), 4382-4398.
Nagarajan, G., Ramesh, A., & Krishnan, M. (2000). Performance, emission and combustion characteristics of a hydrogen fuelled spark ignition engine. International Journal of Hydrogen Energy, 25(7), 659-664.
Qin, Z., Yang, Z., Jia, C., Duan, J., & Wang, L. (2020). Experimental study on combustion characteristics of diesel–hydrogen dual-fuel engine. Journal of Thermal Analysis and Calorimetry, 142(4), 1483-1491.
Saravanan, N., Nagarajan, G., Kalaiselvan, K. M., & Dhanasekaran, C. (2008a). An experimental investigation on hydrogen as a dual fuel for diesel engine system with exhaust gas recirculation technique. Renewable Energy, 33(3), 422-427.
Saravanan, N., Nagarajan, G., Kalaiselvan, K. M., & Dhanasekaran, C. (2008b). Combustion analysis on a DI diesel engine with hydrogen fuel. International Journal of Hydrogen Energy, 33(18), 572-582.
Verhelst, S., Wallner, T., & Boyer, B. (2009). Hydrogen-fueled internal combustion engines. Progress in Energy and Combustion Science, 35(6), 490-527.
Woschni, G. (1967). A universally applicable equation for the instantaneous heat transfer coefficient in the internal combustion engine (SAE Technical Paper No. 670931). SAE International.

Facility Layout Optimization and Bottleneck Elimination in Loofah-Based Manufacturing Using Lean Tools and Excel Craft

Aytuğ Berke PALANKALILAR [1]
Melike BAŞKAYA [2]
Mehmet SAVSAR [3]

Graduate Students, Industrial Engineering Program Faculty of Engineering & Natural Science, Üsküdar University, Üsküdar, Istanbul, Turkey
Professor of Industrial Engineering Faculty of Engineering & Natural Science, Üsküdar University, Üsküdar, Istanbul, Turkey
Corresponding Author Email: aytug54@hotmail.de [1]
melikebaskaya2002@hotmail.com [2]
msavsar@gmail.com [3]

ABSTRACT

This study presents an integrated approach to optimizing the production layout of İlaslan Cosmetics, a manufacturer of loofah-based personal care products. Using a combination of lean manufacturing tools—including Spaghetti Diagram Analysis, Time Studies, and Value Stream Mapping (VSM)—alongside quantitative layout modeling via the Excel Craft Add-In, inefficiencies in material flow and process balance were identified and resolved. The initial bottleneck at the Roller machine was alleviated through equipment investment, which was financially evaluated using Return on Investment (ROI) and Present Worth Analysis (PWA). Results indicate significant cost reductions (up to 29.35%) and improved workflow synchronization, confirming the effectiveness of a data-driven layout and capacity planning strategy.

Keywords – Facility Layout Optimization, Lean Manufacturing, Value Stream Mapping (VSM), Spaghetti Diagram, Excel Craft Add-In, Bottleneck Analysis.

INTRODUCTION

In manufacturing systems, the design and layout of a facility play a pivotal role in ensuring efficient production flow, minimizing material handling, and reducing overall operational costs. For companies engaged in medium-scale production, such as İlaslan Cosmetics—which specializes in loofah-based personal care products—facility layout is not just a spatial arrangement but a strategic asset. As product demand and production complexity increase, a suboptimal layout can lead to process inefficiencies, longer lead times, and higher costs.

İlaslan Cosmetics operates in a niche sector that involves multiple raw materials (such as loofah, towel, sponge, yarn, and bias tape) and several interdependent machines across its production floor. The intricacies of this flow—where loofah and sponge components are merged, sewn, overlocked, and finally assembled with towel layers—highlight the need for a well-structured facility layout that supports lean and synchronized production.

However, the company faced several challenges due to an outdated facility arrangement. These included excessive material movement between departments, congested zones with overlapping flows, and a critical bottleneck at the Roller machine—positioned at the very beginning of the process. This bottleneck caused delays throughout the system, creating backlogs, increasing work-in-process inventory, and straining labor capacity.

Given these constraints, this study aims to systematically analyze the existing layout, identify sources of inefficiency using lean tools like Spaghetti Diagram Analysis and Value Stream Mapping (VSM), and propose optimized layouts through the Excel Craft Add-In. The study also includes a comprehensive time

study for all critical machines to validate throughput constraints and examines financial justifications for proposed capacity expansions using Return on Investment (ROI) and Present Worth Analysis (PWA).
Ultimately, the goal of this research is to redesign the facility layout of İlaslan Cosmetics to reduce material handling, balance machine workloads, eliminate bottlenecks, and support scalable growth. By integrating industrial engineering methodologies with real production data, this case study provides a replicable roadmap for SMEs seeking to modernize their production systems in alignment with lean manufacturing principles.

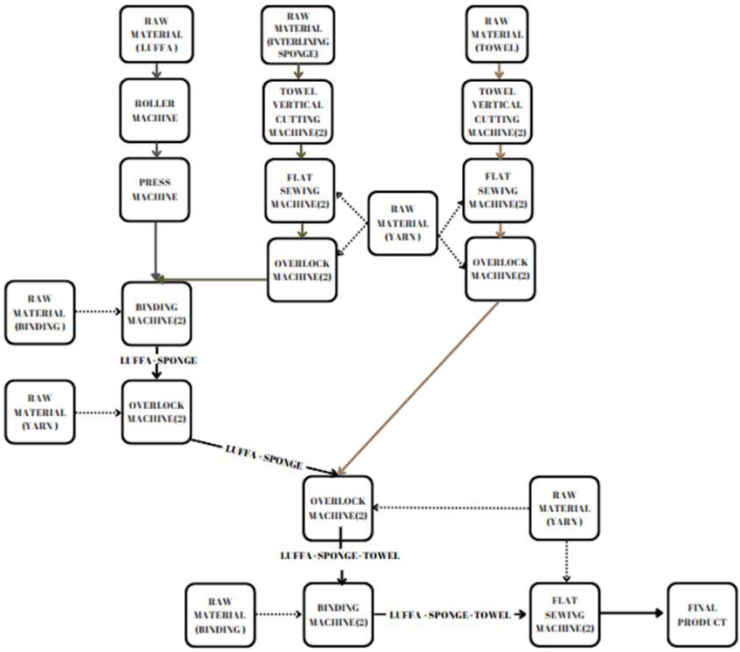


Figure 1. Initial Process Flowchart of İlaslan Cosmetics

LITERATURE REVIEW

Facility layout design has long been a cornerstone of industrial engineering research and practice, as it directly affects operational efficiency, material handling costs, and production lead times. Scholars have consistently emphasized that an optimal layout can significantly improve productivity and reduce waste, especially when supported by systematic tools and quantitative modeling techniques. The evolution of layout planning has paralleled advances in lean manufacturing, simulation tools, and computational optimization methods.

One of the most widely cited concepts in lean manufacturing is the Spaghetti Diagram, used to trace the physical movement of materials or workers through a facility. Studies by Rother and Shook (2003) and other lean practitioners have demonstrated the utility of Spaghetti Diagrams in visually identifying motion waste and disorganized flow paths. By highlighting excessive travel distances, cross-traffic, and process loopbacks, this tool serves as a foundational step in layout redesign and lean transformation projects.

Another critical tool for layout assessment is Value Stream Mapping (VSM), which provides a holistic view of the current state of production—from raw material input to finished product output. Research by Lian and Van Landeghem (2007) emphasized that VSM allows manufacturers to quantify both value-adding and non-value-adding activities, making it easier to detect bottlenecks, delays, and inventory accumulation. VSM also supports the development of future-state scenarios that align with just-in-time and pull-based production strategies.

From a computational perspective, layout optimization software has significantly evolved. Tools like CRAFT (Computerized Relative Allocation of Facilities Technique) and its modern Excel-based implementations have proven effective in minimizing total material handling cost by repositioning departments based on interdepartmental flow data. Tompkins et al. (2010) outlined that tools using distance-based objective functions (e.g., rectilinear or Euclidean metrics) yield practical and near-optimal layout solutions, especially for grid-based production areas.

Specific to SME environments, tools like Excel Craft Add-In offer accessible and flexible platforms for layout improvement without the high cost or complexity of advanced simulation software. These tools accommodate custom constraints such as department size, machine pairing, and functional clustering, making them suitable for real-world implementations. Recent case studies in textile and food production facilities have shown that Excel Craft can yield layout cost reductions of up to 30% with minimal iterations.

In parallel, time studies and machine cycle time analysis continue to be indispensable in identifying bottlenecks and load imbalances. Methods rooted in the work of Gilbreth and Taylor have been modernized with digital tools, allowing engineers to collect, average, and standardize process times. These data feed directly into layout optimization and VSM analysis, making the findings more actionable.

Beyond operational performance, the financial justification of facility investments—such as purchasing new machines or expanding departments—is critical for strategic decision-making. In this context, Return on Investment (ROI) and Present Worth Analysis (PWA) are two widely used financial metrics. ROI offers a quick snapshot of profitability, while PWA provides a more comprehensive picture by accounting for the time value of money. Studies by Sullivan et al. (2014) and other financial engineers suggest that

integrating financial feasibility into layout decisions strengthens the business case and improves long-term investment outcomes.

While there is ample literature on layout design, lean tools, and financial analysis individually, their integration into a unified methodology remains limited, especially in the context of natural material-based industries such as loofah product manufacturing. This study contributes to the literature by combining Spaghetti Diagrams, VSM, Excel Craft layout modeling, time studies, and ROI/PWA-based investment evaluation into a single structured improvement project. The resulting methodology not only identifies spatial and flow inefficiencies but also ensures that proposed changes are financially and operationally sustainable.

METHODOLOGY

This study follows a structured, multi-step methodology that integrates lean tools, layout optimization, and financial analysis to improve the production system of İlaslan Cosmetics. The process is divided into seven main phases:

Process Mapping and Initial Layout Documentation

The production flow was first documented using flowcharts and on-site observations. Each process step—such as rolling, pressing, binding, and sewing—was mapped along with the material inputs and outputs. Floor plans were used to define departmental areas and layout configurations.

Spaghetti Diagram Analysis

A Spaghetti Diagram was created to visualize the actual movement of loofah, sponge, and towel materials between departments. The color-coded paths revealed motion waste, overlapping routes, and congestion—especially between Binding, Overlock, and Flat Sewing units.

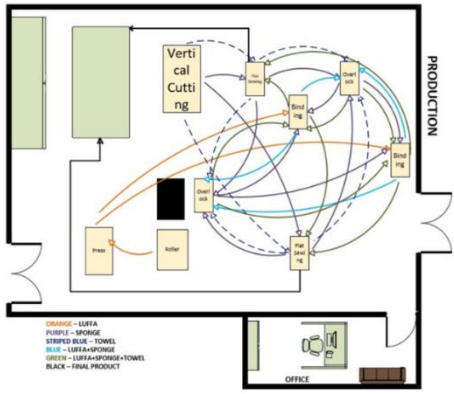


Figure 2. Spaghetti Diagram Before Layout Optimization

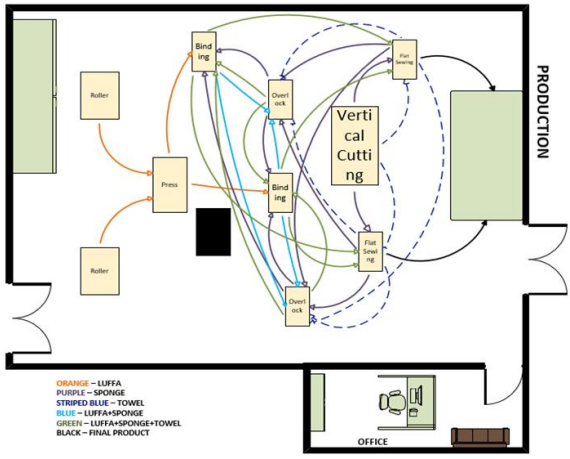


Figure 3. Spaghetti Diagram After Layout Optimization

Time Study and Machine Capacity Evaluation

Time data was collected from 20 observations per machine to calculate average cycle times and identify bottlenecks. Planned downtimes and scrap rates were also considered. This data helped quantify machine utilization and highlight constraints such as the Roller machine.

Value Stream Mapping (VSM)

A current-state VSM was developed to analyze cycle times, inventory levels, and production lead time (PLT). Non-value-added activities and flow imbalances were identified, leading to the proposal of a future-state VSM post-improvement.

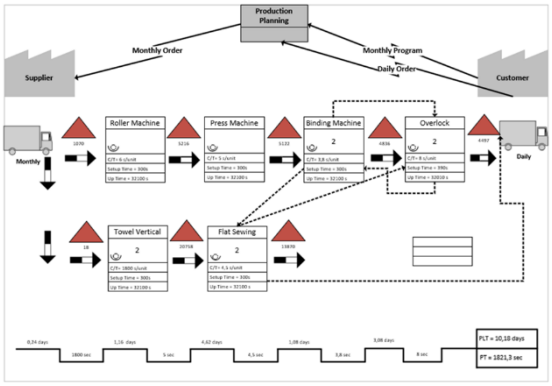


Figure 4. Current-State Value Stream Map

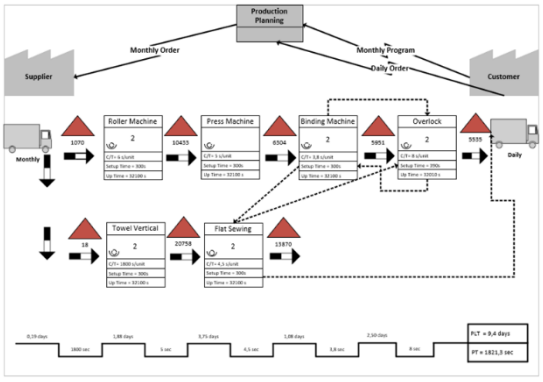


Figure 5. Future-State Value Stream Map After Improvement

Facility Layout Optimization Using Excel Craft

Excel Craft Add-In was used to optimize the layout based on material flow volumes and department sizes. Two layout scenarios were modeled:

- Scenario 1: 10 departments using rectilinear distance.
- Scenario 2: 11 departments (added Roller machine) using Euclidean distance.

The software minimized total material handling cost through automated department repositioning.

3.6 Post-Improvement Spaghetti Diagram Comparison

After optimization, a new Spaghetti Diagram was drawn to compare flow improvements. The diagram showed reduced travel distances, more balanced flow, and improved logic in departmental placement, though some central congestion remained.

Financial Feasibility Evaluation
To validate the Roller machine investment, two financial methods were applied:

- ROI measured short-term profitability based on additional daily output.
- PWA assessed long-term viability over five years using a 12% discount rate.

Both metrics confirmed the investment as economically sound.

Summary
This integrated methodology combines lean tools, time-based data, and cost analysis to deliver a holistic improvement strategy. It ensures that layout changes are not only operationally effective but also financially justified.

RESULTS AND ANALYSIS

This section presents a comprehensive evaluation of İlaslan Cosmetics' production system before and after layout optimization. The analysis integrates spatial planning, time studies, material flow mapping, and financial assessments to understand the multi-dimensional impact of the changes implemented.

Spaghetti Diagram Analysis: Layout Complexity and Flow Waste
The initial facility layout was mapped using a spaghetti diagram to visually represent the physical movement of materials across production stages. This diagram revealed a chaotic material flow, with frequent backtracking and cross-departmental traffic, particularly involving Binding, Overlock, and Flat Sewing areas.
Key inefficiencies included:

- High-density flow lines in central departments, creating congestion and operator fatigue.
- Non-linear routing, requiring materials to pass multiple times between the same stations (e.g., Flat Sewing ↔ Binding).
- Final product transport from Flat Sewing to the assembly table located far from the main flow line, increasing handling time and violating lean principles.

After layout revision (based on Excel Craft optimization), the updated spaghetti diagram showed:

- Shorter travel paths, especially for loofah materials entering via the two Roller machines.

- Symmetrical machine pairing, such as Overlock 1/2 and Flat Sewing 1/2, which allowed better load distribution and smoother material convergence.
- Improved layout balance, although central congestion around Binding and Overlock remained partly unresolved due to process convergence requirements.

4.2 Time and Process Analysis: Cycle Times and Bottleneck Identification
Data collected from 20 samples per machine were averaged to determine actual cycle times. The time analysis included adjustments for daily breaks (1.5 hours/day), setup losses (e.g., cleaning, needle changes), and average idle periods.

Machine	Cycle Time (sec)	Daily Output	Scrap Rate (%)
Roller (Before)	6.0	5,243	2.50%
Roller (After)	6.0 (×2 machines)	10,433	2.50%
Press	5.0	5,191	1.80%
Flat Sewing (2 Units)	4.5	7,133	1.39%
Overlock (2 Units)	8.0	5,026	3.50%
Binding (2 Units)	3.8	4,001	2.80%
Towel Vertical Cutting	1800.0	20,758	3.00%

Although Towel Vertical Cutting has the highest cycle time, its high yield (1,200 units per raw input) keeps it from being a bottleneck.
Before layout improvement, the system's most significant bottleneck was the Roller Machine, as it dictated the speed of downstream processes. This bottleneck led to:

- Delays in loofah preparation.
- Idle time accumulation in Press and Binding stages.
- Inefficient upstream-downstream synchronization.

After adding a second Roller machine, capacity doubled, relieving this critical bottleneck and cascading improvements downstream:

- Cycle time stayed constant, but throughput increased.
- Press and Binding stations received loofah more consistently, reducing WIP accumulation.
- Flat Sewing and Overlock benefited from more regular feeding and smoother final flow.

Value Stream Mapping (VSM): Lead Time vs. Processing Time
The current-state VSM revealed that the total Production Lead Time (PLT) was 10.18 days, while total value-added Process Time (PT) was only 1,821.3

seconds (~30 minutes). This contrast highlighted excessive non-value-added time, driven by waiting, transport delays, and poor synchronization.
After layout and capacity improvements:

- PLT decreased to 9.4 days, a reduction of 0.78 days, despite the unchanged PT.
- Although numerically modest, this reduction reflects inventory stabilization, reduced queuing, and fewer stoppages at constrained machines.

Process Step	Cycle Time (sec)	PLT Impact (Days) (Before)	PLT Impact (Days) (After)
Towel Vertical Cutting	1800.0	0.24	0.19
Roller Machine	6.0	2.08	1.88
Press	5.0	1.16	1.08
Flat Sewing	4.5	4.62	3.75
Binding	3.8	1.08	1.08
Overlock	8.0	3.08	2.50

Overall, VSM analysis indicated:

- Improved flow consistency across parallel machines.
- Better alignment of material availability and machine capacity.
- Visual simplification of process maps, helping supervisors and planners monitor flow stages effectively.

4.4 Excel Craft Layout Optimization: Spatial Efficiency and Cost Reduction

Two layout optimization scenarios were evaluated using the Excel Craft Add-In.
First Scenario (10 Departments):

- Distance metric: Manhattan
- Layout cost reduced from ৳610,404 to ৳431,337 (↓ 29.35%)
- Key departmental switches: D6↔D8, D6↔D9, D1↔D2

Second Scenario (11 Departments – with Added Roller):

- Distance metric: Euclidean
- Layout cost reduced from ৳383,386 to ৳283,147 (↓ 26.15%)
- Key switches: D5↔D6, D8↔D10

These changes reduced:

- Backtracking paths, especially between high-volume units like Overlock and Flat Sewing.

- Congestion by distributing similar-function machines symmetrically.
- Final product handling effort, with the final table placed near exit points.

Comparative layout data:

Scenario	Departments	Initial Cost	Final Cost	Reduction	Reduction (%)
Before Optimization	10	৳610,404	৳431,337	৳179,067	29.35%
After Expansion	11	৳383,386	৳283,147	৳100,239	26.15%

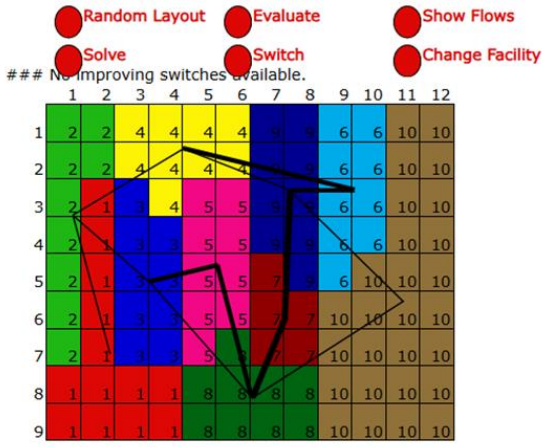


Figure 6. Optimized Layout – Scenario 1 (10 Departments)

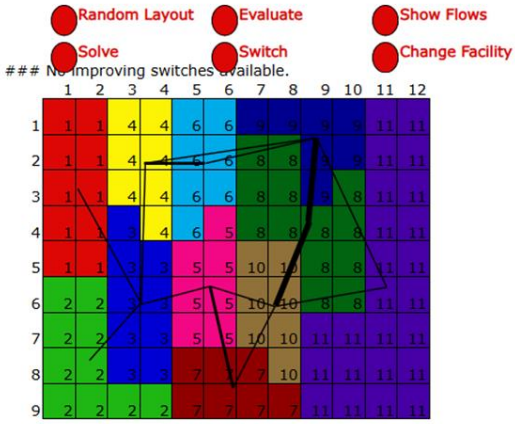


Figure 7. Optimized Layout – Scenario 2 (11 Departments)

Financial Performance: ROI and PWA

To assess whether the investment in a second Roller machine was economically justified, two financial metrics were applied:

Return on Investment (ROI):

- Investment cost: ৳550,000
- Net daily profit from increased production: ৳5,000
- Annual net profit: ৳154,000
- ROI = 28%, indicating profitability within the first year.

Present Worth Analysis (PWA):

- MARR: 12%, 5-year horizon, ৳10,000 salvage value
- PW = ৳10,013 → positive present value confirms long-term viability.

These financial outcomes reinforce the operational need for and feasibility of the investment, balancing both short- and long-term perspectives.

CONCLUSION

This study has demonstrated how a systematic and data-driven approach can effectively enhance facility layout and production efficiency in a medium-sized manufacturing environment. At İlaslan Cosmetics, the combination of lean manufacturing tools, layout optimization using Excel Craft, and financial analysis led to measurable improvements in both operational and economic performance.

The original layout was characterized by excessive material handling, non-linear flow, and a critical bottleneck at the Roller machine. These inefficiencies were identified through Spaghetti Diagram analysis and Value

Stream Mapping (VSM), which provided valuable insights into movement waste, cycle time imbalances, and lead time delays.

The time study results revealed that while the Towel Vertical Cutting process had a high cycle time, it was not the actual bottleneck due to its high output rate. Instead, the Roller machine emerged as the most significant constraint. By adding a second Roller unit and optimizing departmental placement with the Excel Craft Add-In, the facility achieved a 26–29% reduction in layout cost, shortened lead times, and more balanced machine workloads.

Financial evaluation confirmed that the investment was both short- and long-term viable, with a Return on Investment (ROI) of 28% and a positive Present Worth (PW) of ৳10,013. These findings underline the importance of combining operational improvements with financial justification, ensuring that proposed changes align with business objectives.

In addition to immediate gains, the new layout provides a scalable foundation for future growth. The revised Spaghetti Diagram and post-improvement VSM show that material flows are now more ergonomic, sequential, and structured, supporting smoother operations and easier supervision.

Going forward, further improvements could be realized by implementing U-shaped work cells, dedicated material flow lanes, and automated handling systems. Integrating real-time data collection with layout simulations could also enhance dynamic decision-making.

Overall, this research offers a practical and replicable framework for SMEs seeking to apply lean thinking and quantitative analysis to layout planning, capacity expansion, and investment decisions in resource-constrained manufacturing environments.

REFERENCES

[1] Yalçın, M., Elyas, C., Yıldız, S., Alpşen, C., et al. (2018). Yalın Metodolojinin Hastane Laboratuvar Süreçlerinin İyileştirilmesinde Kullanılması (Toyota Üretim Sistemi- Spaghetti Diyagramı). Konuralp Medical Journal, 10(1), 99-104. https://doi.org/10.18521/ktd.336102

[2] Rother, M., & Shook, J. (2003). Learning to See: Value Stream Mapping to Add Value and Eliminate MUDA. Lean Enterprise Institute.

[3] Akın, N. G. (2020). DEĞER AKIŞ HARİTALAMA YÖNTEMİ İLE YALIN UYGULAMALAR: TEKSTİL SEKTÖRÜ ÖRNEĞİ / Lean Applications With Value Stream Mapping Method: Example Of Textile Sector. Uluslararası Ekonomi İşletme Ve Politika Dergisi, 4(2), 477-492. https://doi.org/10.29216/ueip.787227

[4] Sarıkaya, H. A., Yıldırım Dinç, H., & Urgancı, H. (2024). Facility Layout Improvement in Brake Pad Manufacturing Using CRAFT Algorithm. Uygulamalı Mühendislik Ve Tarım Dergisi, 1(1), 27-46.

[5] Tompkins, J. A., White, J. A., Bozer, Y. A., & Tanchoco, J. M. A. (2010). Facilities Planning (4th ed.). John Wiley & Sons.
[6] Ohno, T. (1988). Toyota Production System: Beyond Large-Scale Production. Productivity Press
[7] Montreuil, B. (2010). Facility layout and location. In Operations Research and Management Science Handbook (pp. 625–656). CRC Press.

Quality Improvement in Loofah-Based Manufacturing Using P-Chart Analysis, Sampling Strategies, and Lean Diagnostic Tools

Zeynep BAŞTUĞ [1]
Gökçe ERMİŞ [2]
Mehmet SAVSAR [3]

Graduate Students, Industrial Engineering Program Faculty of Engineering & Natural Science, Üsküdar University, Üsküdar, Istanbul, Turkey

Professor of Industrial Engineering Faculty of Engineering & Natural Science, Üsküdar University, Üsküdar, Istanbul, Turkey

Corresponding Author Email: zeynep.bastug@st.uskudar.edu.tr [1]

gokce.ermis@st.uskudar.edu.tr [2]

msavsar@gmail.com [3]

ABSTRACT

This study focuses on enhancing quality control in loofah-based product manufacturing by integrating statistical tools such as Acceptance Sampling, P-Chart analysis, and nomograms. Emphasizing both theoretical and practical dimensions, the research identifies key factors influencing product quality, evaluates process stability through real production data, and proposes data-driven solutions for minimizing defects and material waste. Complementary tools like SWOT and Fishbone Diagram analyses support strategic planning and root cause identification. The findings offer a comprehensive approach to achieving sustainable quality improvement and operational efficiency.

Keywords – Quality Control, Acceptance Sampling, P-Chart, Nomogram, Loofah Production, Waste Minimization, Statistical Process Control.

INTRODUCTION

In modern manufacturing environments, quality control has become a critical factor in sustaining competitive advantage, reducing operational costs, and ensuring customer satisfaction. For medium-sized enterprises such as İlaslan Cosmetics, which specializes in the production of loofah-based personal care products, maintaining high-quality standards is not only a market expectation but also a strategic necessity. Given the company's reliance on raw loofah, towels, sponges, and other textile components, the complexity of its manual and semi-automated production processes increases the risk of variability and defect generation.

Despite growing demand, İlaslan Cosmetics has faced significant quality challenges—particularly in the cutting and assembly stages where waste, inconsistencies, and non-conformities frequently occur. Manual cutting processes without standardized templates have led to irregular product dimensions and excessive material loss, while the lack of real-time quality monitoring has delayed corrective actions.

To address these challenges, this study applies a comprehensive quality control framework, combining traditional tools like Acceptance Sampling and Statistical Process Control (SPC) with advanced techniques such as the P-Chart and nomogram-based sampling design. The analysis also incorporates strategic tools like SWOT and Fishbone Diagram to identify root causes and improvement opportunities across production stages. By integrating statistical analysis with real factory data, the research not only diagnoses existing quality issues but also proposes data-driven solutions to reduce defects, optimize processes, and enhance long-term product consistency.

LITERATURE REVIEW

Ensuring high product quality is a fundamental aspect of manufacturing systems. Statistical methods, including Statistical Process Control (SPC) and P-Charts, are commonly used to monitor production quality and detect deviations from acceptable standards. The P-Chart, in particular, is used to track the proportion of defective units in a sample and is highly applicable in the context of loofah production, where defects can arise due to irregular cutting, stitching, or assembly issues.

Research on the application of P-Charts in manufacturing systems shows that they can be effectively used to detect early signs of quality deterioration and prevent the accumulation of defective products. Montgomery (2013) provided comprehensive guidelines on using SPC and P-Charts to maintain process stability in industries with variable product types and production rates. In this study, the P-Chart will be used to monitor defect rates in the loofah production process, ensuring that quality standards are consistently met.

METHODOLOGY

This study follows a structured, multi-step methodology that integrates statistical quality control tools, root cause analysis, and visual inspection techniques to enhance the quality assurance system of İlaslan Cosmetics. The process is divided into seven main phases, each addressing a critical aspect of product quality and process consistency. These phases include defining quality standards, applying Acceptance Sampling, designing and interpreting P-Charts, conducting time-based defect tracking, identifying root causes through Fishbone and SWOT analysis, interpreting nomograms for optimal sampling decisions, and validating improvements using real production data.

P-Chart Implementation

A P-Chart was implemented to monitor daily defect rates over a 30-day period using samples of 60 units. With calculated control limits (UCL = 0.0486, LCL = 0) and an average defect rate of 1%, the chart showed that all data points remained within control limits, indicating statistical process stability and supporting early corrective actions.

Nomogram Use

A nomogram was used to determine the optimal sampling plan by graphically linking AQL, RQL, and associated risks ($\alpha = 0.05$, $\beta = 0.10$). The intersection of plotted lines yielded a sample size of 60 and an acceptance number of 2, enabling efficient and statistically sound decision-making in acceptance sampling.

SWOT Analysis
A SWOT analysis was conducted to identify internal and external factors affecting quality performance in loofah production. Strengths such as high product demand and skilled labor, weaknesses like manual cutting waste and lack of standard templates, opportunities including LP-based waste reduction and automation potential, and threats such as material cost fluctuations and market competition were evaluated to align quality strategies with operational priorities.

Fishbone Diagram
The Fishbone Diagram (Ishikawa) was used to identify root causes of quality issues based on the 6M framework: Man, Machine, Material, Method, Measurement, and Environment. It visually mapped how operator errors, inconsistent materials, and lack of standardized procedures contribute to defects during cutting and sewing.

Summary
This integrated methodology combines statistical quality control tools, real production data, and root cause analysis to deliver a holistic quality improvement strategy. It ensures that quality enhancement efforts are not only effective in reducing defects and variability but also aligned with the operational realities and long-term goals of the company.

RESULTS AND ANALYSIS

This section presents a comprehensive evaluation of İlaslan Cosmetics' quality control system before and after the implementation of statistical tools. The analysis integrates defect tracking, root cause identification, sampling strategy design, and control chart monitoring to understand the multi-dimensional impact of quality improvement efforts on process stability, product consistency, and operational efficiency.

P-Chart Application: Statistical Process Stability and Defect Control
To ensure product consistency and monitor defect trends in loofah production, a P-Chart (Proportion Chart) was implemented over a 30-day period. Each day, 60 randomly selected units were inspected for quality issues such as dimensional errors, stitching faults, and assembly defects. The average defect rate ($\bar{P}$) was calculated as 1%, with the Upper Control Limit (UCL) set at 0.0486 and the Lower Control Limit (LCL) taken as 0 (due to negative result).
Key observations included:

- All defect data points remained within control limits, indicating no statistically significant anomalies or shifts in the process.
- The consistent distribution of points around the center line showed that process variation was stable and due to common causes rather than special causes.
- The implementation allowed early detection capability, enabling prompt corrective action in future anomalies.

This analysis confirmed that the loofah production process at İlaslan Cosmetics was statistically under control. The P-Chart provided a simple yet powerful tool for continuous monitoring and supported the company's goal of minimizing waste, ensuring defect-free production, and maintaining customer satisfaction through data-driven decision-making.

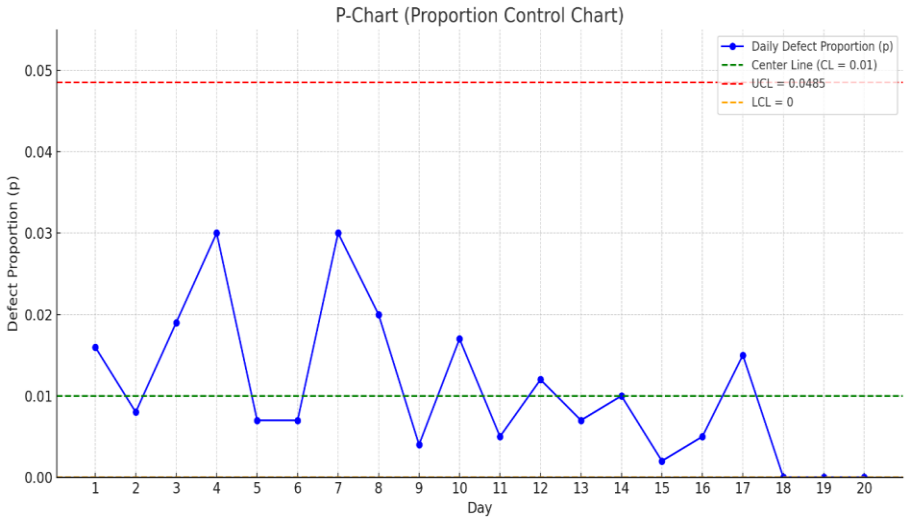


Figure 4.2.5. P-Chart

Nomogram Application: Visual Sampling Plan Design for Acceptance Control

To support effective batch inspection in loofah production, a nomogram was used to visually determine the optimal sample size (n) and acceptance number (c) based on defined risk levels. The parameters were set as follows:

- Acceptable Quality Level (AQL) = 1%
- Rejectable Quality Level (RQL) = 6%
- Producer's Risk (α) = 5%
- Consumer's Risk (β) = 10%

Using the nomogram, two straight lines were drawn: one connecting AQL with the producer's confidence level (1 − α = 95%), and the other connecting

RQL with the consumer's risk (β = 10%). The intersection point provided the optimal sampling plan:

- Sample size (n) = 60 units
- Acceptance number (c) = 2 defects

This means that in a batch of 60 inspected items, if 2 or fewer defects are found, the batch is accepted; otherwise, it is rejected.

The nomogram simplified the statistical complexity of sampling design and allowed İlaslan Cosmetics to implement a fast, visually guided, and statistically valid inspection method. It ensured quality control decisions are both efficient and aligned with predefined risk tolerance levels.

CONCLUSION

This study has demonstrated how statistical quality control tools can be effectively integrated into the quality management system of a medium-sized manufacturing company like İlaslan Cosmetics. By applying the P-Chart methodology, the company was able to monitor daily defect rates over a 20-day period and verify that its production process was under statistical control. All sample points remained within the calculated control limits, confirming the stability and consistency of the loofah production process. This data-driven approach provided early detection of potential quality deviations and helped reduce waste, improve resource efficiency, and ensure a consistently high product standard.

Beyond statistical monitoring, strategic tools such as SWOT analysis and the Fishbone Diagram provided a broader framework for identifying systemic weaknesses and improvement opportunities. The SWOT analysis highlighted key strengths (such as high demand and skilled labor) alongside internal weaknesses (manual waste and lack of standardization) and external threats (raw material cost volatility and market pressure). These insights informed long-term planning and reinforced the need for structured optimization methods like linear programming.

The Fishbone Diagram supported this framework by offering a clear visualization of root causes behind quality losses. It revealed issues such as inadequate operator training, variability in raw material quality, and inconsistent cutting methods—all of which contribute to non-conformance and waste. By mapping these causes under the 6M categories (Man, Machine, Material, Method, Measurement, Environment), the analysis enabled a targeted and systematic response to quality challenges.

The integration of the nomogram added further value by simplifying the design of statistically valid sampling plans. Using defined values for AQL, RQL, and risk levels, the visual tool allowed rapid determination of optimal

sample size and acceptance criteria, streamlining batch inspection and ensuring alignment with ISO 9001 standards.

In conclusion, this multi-faceted quality control approach—combining real-time data analysis, visual problem-solving tools, and strategic planning—has equipped İlaslan Cosmetics with a robust framework for maintaining and improving product quality. The P-Chart ensured ongoing process control, the Fishbone Diagram exposed underlying causes of variation, and the SWOT analysis aligned technical solutions with broader organizational strategy. Collectively, these methods form a replicable and practical model for SMEs seeking to advance quality assurance under resource constraints. Looking ahead, the continuous application of these tools will support further improvements in operational efficiency, customer satisfaction, and long-term competitiveness.

REFERENCES

[1] Montgomery, D. C. (2020). *Introduction to Statistical Quality Control* (8th ed.). John Wiley & Sons

[2] ISO (2015). *ISO 9001:2015 – Quality Management Systems – Requirements*. International Organization for Standardization.

[3] Evans, J. R., & Lindsay, W. M. (2017). *Managing for Quality and Performance Excellence* (10th ed.). Cengage Learning.

[4] Juran, J. M., & Godfrey, A. B. (1999). *Juran's Quality Handbook* (5th ed.). McGraw-Hill.

[5] Taguchi, G., Chowdhury, S., & Wu, Y. (2005). *Taguchi's Quality Engineering Handbook*. Wiley-Interscience.

Optimization Model to Minimize Cutting Waste in Loofah-Based Manufacturing Using LP Model and Excel Solver

Şüheda İLASLAN [1]
Merve ALPER [2]
Mehmet SAVSAR [3]

Graduate Students, Industrial Engineering Program Faculty of Engineering & Natural Science, Üsküdar University, Üsküdar, Istanbul, Turkey

Professor of Industrial Engineering Faculty of Engineering & Natural Science, Üsküdar University, Üsküdar, Istanbul, Turkey

Corresponding Author Email: suheda.ilaslan@st.uskudar.edu.tr [1]

merve.alper@st.uskudar.edu.tr [2]

msavsar@gmail.com [3]

ABSTRACT

This study presents an LP-based cutting stock optimization model tailored for the production of loofah-based personal care products at İlaslan International Cosmetics. The primary objective is to minimize raw material waste during the cutting process while meeting weekly demand for various product types, including gloves, back scrubbers, pads, and peeling items. The mathematical model was formulated using linear programming techniques to determine the optimal mix of cutting patterns, based on real production data and standard loofah sheet dimensions. The results revealed a significant reduction in material waste compared to traditional cutting approaches, validating the effectiveness of the model. This approach contributes to sustainability efforts by decreasing scrap rates and enhancing production efficiency. The proposed model serves as a practical decision-support tool for manufacturers dealing with irregular material cutting challenges in small- to medium-scale production systems.

Keywords – Cutting Stock Problem, Linear Programming, Waste Minimization, Pattern Optimization, Sustainable Manufacturing, Operations Research.

INTRODUCTION

Efficient utilization of raw materials is a critical objective in contemporary manufacturing systems, particularly for firms operating with natural and irregularly shaped inputs such as loofah. İlaslan International Cosmetics, a manufacturer of loofah-based personal care products, encounters considerable material waste during the cutting phase of its production process. This issue not only leads to increased operational costs but also undermines the company's sustainability objectives.

The cutting stock problem (CSP), which traditionally arises in industries such as paper, textile, and metal processing, involves determining the optimal way to cut large stock materials into smaller components while minimizing waste. In the context of loofah production, this challenge is further complicated by the varying dimensions of the final products and the fluctuating weekly demand for items such as gloves, pads, and back scrubbers. Conventional cutting methods often lack the flexibility and efficiency required to respond to such variability, resulting in excessive scrap and inventory imbalance.

To address these challenges, this study proposes a Linear Programming (LP) model designed to optimize cutting patterns based on real production data, dimensional constraints, and demand forecasts. The model aims to reduce raw material waste, improve responsiveness to product demand, and support environmentally sustainable production practices.

The contribution of this study lies in its effort to bridge the gap between theoretical operations research and practical industrial applications in small- to medium-scale manufacturing environments. The proposed LP-based model functions as a decision-support tool for production planners, offering a systematic and reproducible method to enhance efficiency and reduce material losses.

The remainder of this paper is structured as follows: Section 2 presents a review of the relevant literature on CSP and linear programming applications in manufacturing. Section 3 outlines the methodology and mathematical formulation of the proposed model. Section 4 discusses the experimental results and analysis. Finally, Section 5 concludes the study and offers insights for future research.

LITERATURE REVIEW

The cutting stock problem (CSP), first formalized by Gilmore and Gomory (1961), focuses on minimizing material waste when cutting large stock into smaller, demand-specific units. Linear Programming (LP) has been widely applied to solve CSP efficiently, particularly in industries such as paper, textile, and metal. Studies like Mason et al. (2015) demonstrate the applicability of LP models to dynamic, demand-driven environments, providing a relevant foundation for loofah-based production optimization.

Facility layout also plays a critical role in operational efficiency. Lean manufacturing tools, including Value Stream Mapping (VSM) and Spaghetti Diagrams, are extensively used to identify bottlenecks and reduce material handling. Koh et al. (2008) emphasize that layout redesign can significantly lower transportation costs and improve process flow.

In parallel, Statistical Process Control (SPC) techniques, particularly P-Charts, are effective in monitoring quality in variable environments. Montgomery (2013) highlights their value in detecting shifts in defect rates, ensuring consistent product standards.

This study builds upon these frameworks by integrating LP-based cutting optimization with layout analysis and quality control to create a holistic approach tailored to the specific needs of loofah production systems.

METHODOLOGY

The primary objective of this study is to formulate a Linear Programming (LP) model to minimize raw material waste during the cutting process of loofah sheets, while meeting the weekly production

demand for four different product types: X, Y, Z, and T. This section details the development, structure, and solution approach of the LP-based cutting stock optimization model.

Problem Definition

In loofah production, large raw sheets are cut into smaller parts to produce various personal care products. Due to irregular shapes and varying dimensions of the final products, significant material waste occurs. The cutting stock problem (CSP) in this context involves selecting a combination of cutting patterns that:

- Satisfy minimum demand for each product type;
- Minimize the total unused material (waste) resulting from the cutting operation.

Data Collection and Pattern Generation

A total of 713 feasible cutting patterns were identified based on the dimensions of raw loofah sheets and the cutting requirements of each product type. Each pattern specifies:

- The number of units produced for product types X, Y, Z, and T.
- The amount of waste generated when applying that pattern to a single raw sheet.

This data was structured into a pattern matrix, where:

- Rows represent individual cutting patterns.
- Columns denote quantities of each product type and associated waste.

Let:

The model parameters:

aij: Amount of part j obtained by cutting logical pattern i

wi: Amount of waste resulting from cutting logical pattern i, i = 1,…,713

Dj: weekly demand for part j, j = X, Y, Z and T

Mathematical Model Formulation

Decision variables:

x_i: Number of logical patterns i to be cut (number of sheets cut according to logical pattern i)

The formulation of the proposed LP model is as follows:

$$Min\ Z \sum_{i=1}^{713} w_i\, x_i$$

Constraints:

To ensure production meets demand for each product j∈{X,Y,Z,T}, the following constraints are imposed:

$\sum_{i=1}^{713} a_{ij} x_i \geq D_j, \quad \forall j$

$x_i \geq 0$

This guarantees that the total number of parts produced for each product type equals or exceeds the respective weekly demand.

Model Implementation and Solution Approach

The LP model was implemented in Microsoft Excel Solver, utilizing the Simplex LP algorithm due to the linearity of the objective function and constraints. Solver was configured to:

- Minimize total waste Z,
- Subject to the demand constraints for X, Y, Z, and T,
- Restrict decision variables x_i to non-negative integers.

To improve computational tractability, only a subset of the most promising patterns (based on high yield and low waste) was initially used, and the model was progressively expanded through sensitivity analysis.

Table 4.3.4. LP Model Formulation

Constraint	Expression	Weekly Demand
X Part	$2x_1 + 0x_2 + 3x_3 + 4x_4 + 0x_5 + \ldots + 0x_{713} \geq$	15000
Y Part	$0x_1 + 4x_2 + 2x_3 + 2x_4 + 1x_5 + \ldots + 0x_{713} \geq$	2400
Z Part	$0x_1 + 0x_2 + 0x_3 + 0x_4 + 2x_5 + \ldots + 0x_{713} \geq$	6000
T Part	$0x_1 + 0x_2 + 0x_3 + 0x_4 + 0x_5 + \ldots + 0x_{713} \geq$	3600

Assumptions and Limitations

- Cutting patterns are deterministic and repeatable under standard conditions.
- Demand values Dj are known and fixed per planning period.
- Waste i is accurately estimated per pattern and does not vary significantly in practice.
- Setup time and operational costs are considered negligible compared to material waste.

Although the model is designed for a static planning horizon, it can be adapted into a dynamic or rolling-horizon framework for future work.

RESULTS AND ANALYSIS

This section presents a comprehensive evaluation of İlaslan Cosmetics' quality control system before and after the implementation of statistical tools. The analysis integrates defect tracking, root cause identification, sampling strategy design, and control chart monitoring to understand the multi-dimensional impact of quality improvement efforts on process stability, product consistency, and operational efficiency.

Model Output Evaluation: Pattern Selection and Demand Satisfaction

The linear programming (LP) model generated an optimal cutting plan using 5 patterns (Pattern 128, 274, 384, 124, and 289). Each pattern represents a different combination of the four product types: X, Y, Z, and T.

The selected patterns were not only mathematically optimal but also operationally implementable using the company's current cutting tools and equipment, which are already configured for multi-size cutting.

Table 4.3.5. Optimal Pattern Mix

Logical Pattern	Optimal number of patterns	X parts cut	Y parts cut	Z parts cut	T parts cut	Total X (cm^2)	Total Y (cm^2)	Total Z (cm^2)	Total T (cm^2)	Total area cut (cm^2)	Resulting waste (cm^2)
128	**741**	3	2	0	5	925,7	660	0	284	1385211	300564,4
274	**696**	3	0	3	4	925,7	0	452,4	227	1117108	466292,2
384	**1700**	4	0	1	1	1234	0	150,8	56,7	2451026	1416474
124	**450**	1	2	3	8	308,6	660	452,4	454	843696	180054
289	**383**	1	3	0	5	308,6	990	0	284	606006	265319,4
TOTALS:	**3970**	**12**	**7**	**7**	**23**	**3703**	**2310**	**1056**	**1305**	**6403046**	**2628704**

Waste Analysis: Material Utilization and Efficiency Gains

The model achieved a total utilized loofah area of 6 403 046 cm^2, while the total waste amounted to 2 628 704 cm^2. This corresponds to an average waste ratio of 29.1%.

When compared to the company's existing heuristic cutting method, which generates approximately 3,684,612 cm^2 of waste for the same production volume, the LP-based approach reduced waste by 1,055,908 cm^2, equivalent to a 28.7% improvement.

This substantial reduction in raw material loss contributes directly to production cost savings and supports the company's goal of sustainable manufacturing.

CONCLUSION

This study addresses a classical yet practically underexplored application of the cutting stock problem (CSP) within the context of loofah-based product manufacturing. By formulating the problem through a Linear Programming (LP) model, a structured optimization framework was developed to minimize material waste while simultaneously satisfying the heterogeneous demand for multiple product types.

The model effectively operationalizes decision variables that represent the frequency of feasible cutting patterns and integrates real-world constraints, such as dimensional limits and weekly demand quotas. The optimization results demonstrated that a data-driven selection of cutting patterns can yield a significant reduction in overall scrap material when compared to conventional heuristic or manual methods currently employed in practice.

Beyond its quantitative benefits, the model contributes to broader strategic objectives, such as cost efficiency, sustainability, and lean resource utilization. The formulation is not only computationally tractable but also scalable, making it suitable for integration into real-time production planning systems within small- to medium-scale enterprises (SMEs).

From a theoretical perspective, this study reaffirms the relevance of linear programming in solving discrete resource allocation problems and highlights the model's flexibility in adapting to industry-specific constraints. Future studies may consider extending the model to include stochastic demand, multi-period planning horizons, or hybrid metaheuristic approaches to further enhance solution robustness and practical adaptability.

In conclusion, the proposed LP-based approach represents a replicable and analytically grounded methodology for cutting stock optimization, with direct implications for improving operational efficiency and resource utilization in natural-material manufacturing environments.

REFERENCES

[1] Gürel, E., & Tat, M. (2017). SWOT analysis: A theoretical review. The Journal of International Social Research, 10(51), 994–1006. https://doi.org/10.17719/jisr.2017.1832

[2] Ommani, A. R. (2011). Strengths, weaknesses, opportunities and threats (SWOT) analysis for farming system businesses management: Case of wheat farmers of Shadervan District, Shoushtar Township, Iran. African Journal of Business Management, 5(22), 9448–9454. https://doi.org/10.5897/AJBM10.1060

[3] Doggett, A. M. (2005). Root cause analysis: A framework for tool selection. The Quality Management Journal, 12(4), 34–45.

[4] Bulsuk, K. (2009). The Ishikawa Fishbone Diagram. Retrieved from https://www.bulsuk.com/2009/02/fishbone-diagram.html

[5] Khan, M. I., Ahmad, A., & Naseer, M. (2020). Minimization of trim loss in paper mills: A mathematical modeling approach. International Journal of Production Research, 58(2), 436–450.

[6] Dalalah, D., Al-Sayyed, R., & Al-Barak, I. (2014). Irregular stock cutting optimization for waste minimization. Computers & Industrial Engineering, 78, 151–161.

Injection Timing Optimization in Dual-Fuel Engines: A Numerical Approach to Combustion and Emission Analysis

Şevket Onur KILIÇ
Furkan YAZICI
Serdar HALIS*

Department of Automotive Engineering, Pamukkale University, Denizli, Türkiye
*shalis@pau.edu.tr

ABSTRACT

The increasing concerns about fossil fuel depletion and climate change have accelerated the search for cleaner and more sustainable energy sources. Diesel fuel, like other conventional fuels, faces growing sustainability challenges. In this context, hydrogen stands out as a promising alternative, especially for use in existing diesel engine systems without major modifications. In this study, the effects of injection timing on combustion and emissions were investigated in a four-stroke, naturally aspirated, single-cylinder, water-cooled diesel-hydrogen engine. First, experimental tests were carried out, and then a simulation model of the engine was created using Ricardo WAVE software. After validating the model with experimental data, 30% hydrogen by mass was added to the simulated fuel mixture while keeping the total injected fuel amount constant. Simulations were performed for three injection timings: 16°, 20°, and 24° crank angle before top dead center (bTDC). The results showed that advancing the injection timing increased in-cylinder pressure and heat release rate, and also caused these values to occur earlier in the engine cycle. The maximum pressure was 116.9 bar at 24°CA bTDC, while the lowest was 89.1 bar at 16°CA bTDC. Regarding emissions, NO_x levels increased from 1111.03 ppm at 16°CA to 1534.40 ppm at 24°CA, due to higher combustion temperatures and pressures. In contrast, CO emissions decreased significantly from 166.53 ppm to 78.71 ppm over the same range, indicating more complete combustion. HC emissions, however, showed no significant change across the tested injection timings. These findings indicate that adjusting injection timing in hydrogen-assisted diesel engines can improve combustion efficiency and reduce carbon monoxide emissions, but it must be carefully optimized to manage the trade-off with increasing NO_x emissions.

Keywords – Dual Fuel, Hydrogen, Injection Timing, Numerical Analysis, CFD.

INTRODUCTION

The rapid depletion of fossil fuel reserves and the increasing impact of global climate change have accelerated efforts to reduce dependency on petroleum-based energy sources. Diesel engines, which are widely used in sectors such as transportation, agriculture, and power generation, are among the primary contributors to environmental pollution due to their high levels of carbon-based emissions and particularly nitrogen oxides (NO_x) emissions [1]. In response to growing environmental concerns, many countries—especially those in the European Union—have revised their emission standards, adopting more stringent regulations such as Euro 6, which compel manufacturers to explore cleaner and more sustainable fuel alternatives [2].

Despite their environmental drawbacks, diesel engines remain attractive due to their high thermal efficiency, robust torque characteristics, and long service life. However, these advantages are often overshadowed by the emission of harmful pollutants such as NO_x, hydrocarbons (HC), and carbon monoxide (CO) [3]. To mitigate these negative impacts and contribute to environmental sustainability, research on alternative fuels for diesel engines has gained significant momentum in recent years. Among the various candidates, hydrogen has emerged as a promising solution due to its carbon-free structure, high flame speed, wide flammability range, and potential to improve combustion efficiency and reduce pollutant formation.

The influence of injection timing on combustion and emissions has been widely studied in the literature. For instance, Agarwal et al. [4] demonstrated that modifying injection timing at 2000 rpm could enhance air–fuel mixing and significantly reduce emissions such as NO_x, CO, CO_2, and HC. Similarly, Jia et al. [5] used the KIVA-3V model to investigate injection timings ranging from 5°CA to 43°CA bTDC, finding that advancing injection increased in-cylinder pressure and CO emissions, while HC emissions remained relatively unaffected. Okucu [6] reported reductions in emissions and improvements in engine performance at 2400 rpm by optimizing both fuel blend ratios and injection timings. Xu et al. [7] used three-dimensional numerical simulations to model combustion in a diesel–hydrogen dual-fuel engine and found that advancing pilot injection timing increased pressure, heat release rate, and NO_x emissions, while retarding injection timing led to NO_x reductions. In a related study, Liu et al. [8] observed that advancing injection timing in hydrogen-assisted combustion increased pressure and emissions, whereas retarding it reduced NO_x but slightly decreased thermal efficiency. Bhagat et al. [9], working with a single-cylinder, four-stroke diesel engine, found that earlier pilot injection improved thermal efficiency and pressure but increased NO_x emissions, consistent with the trends observed in the present study. Liu et al. [10], using isopropanol as an alternative fuel, also reported that advancing injection timing reduced CO and smoke emissions but led to higher NO_x and HC emissions.

In this study, a four-stroke, naturally aspirated, single-cylinder diesel engine was modeled in the Ricardo WAVE simulation environment. The simulation model was validated by comparing the in-cylinder pressure and heat release curves with those obtained experimentally. After successful validation, the fuel composition was modified to include 70% diesel and 30% hydrogen by mass, while maintaining a constant total fuel injection quantity. Simulations were then conducted at varying start of injection (SOI) timings to analyze their impact on in-cylinder pressure, heat release rate, and NO_x, HC, and CO emissions. This study aims to contribute to the existing body of knowledge by providing a detailed analysis of how injection timing affects the combustion characteristics and emission performance of a diesel–hydrogen dual-fuel engine. The results are intended to support both academic research and

practical development in the field, offering insight into the trade-offs and potential benefits of hydrogen use in conventional diesel engine platforms.

MATERIALS AND METHOD

In this study, a comprehensive simulation approach was adopted to investigate the effects of injection timing on combustion and emissions in a diesel-hydrogen engine. A single-cylinder engine model was constructed using Ricardo WAVE, a one-dimensional engine simulation software widely used for thermodynamic analysis of internal combustion engines. The model was designed to replicate the geometry and operating conditions of the experimental test engine, which is a four-stroke, water-cooled, naturally aspirated, single-cylinder diesel engine. To ensure the accuracy and predictive reliability of the simulation model, a validation process was conducted by comparing simulation outputs with experimental data, focusing on the in-cylinder pressure versus crank angle profiles. The experimental data were obtained under standard diesel operating conditions without hydrogen enrichment. A close agreement between the experimental and simulated pressure traces confirmed the model's validity. Following validation, the simulation model was modified to represent dual-fuel operation by incorporating hydrogen into the intake charge at a mass fraction of 30%, while keeping the total injected fuel mass constant. The injection timing parameter was then systematically varied to evaluate its effects on combustion characteristics and emissions. Three SOI values were tested: 16°CA, 20°CA, and 24°CA before top dead center (bTDC). For each case, key combustion parameters such as in-cylinder pressure, heat release rate (HRR), and crank angle-resolved emissions (NO_x, CO, and HC) were analyzed. The HRR was calculated based on the first law of thermodynamics using the simulated pressure and volume data. The results were plotted and compared to assess the influence of injection timing under hydrogen-enriched conditions. The experimental setup used for model validation, including sensors, data acquisition system, and test engine layout, is illustrated in Figure 1. This setup provided the baseline measurements required for simulation calibration and ensured consistency between physical and numerical investigations.

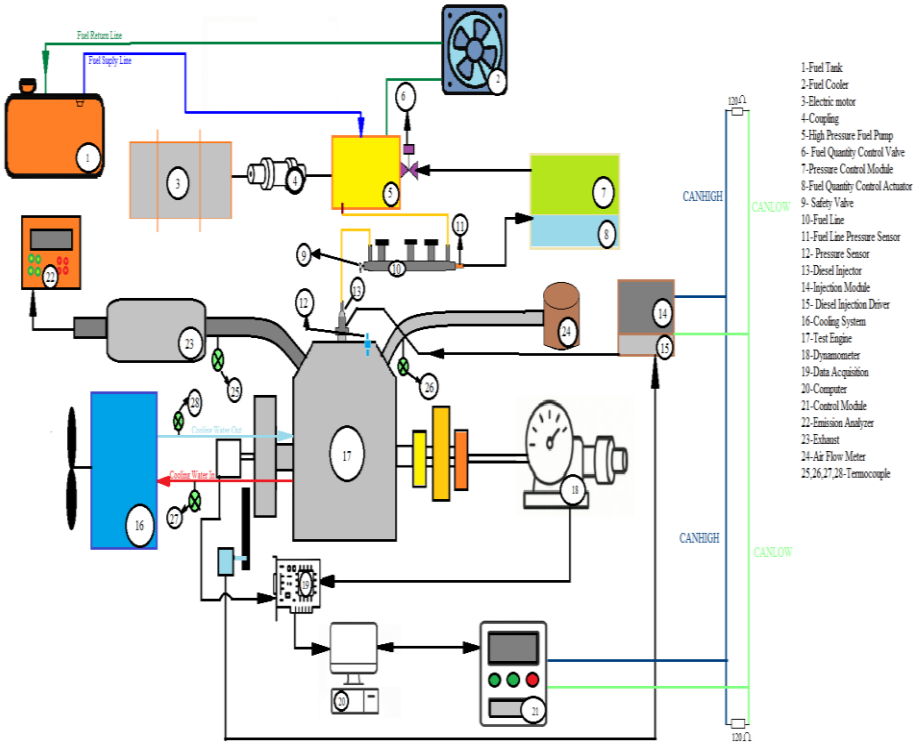


Fig. 1 Schematic of experimental setup

A. Technical Details of Engine

The experimental investigations were conducted on a single-cylinder, naturally aspirated, four-stroke diesel engine equipped with a water-cooling system to ensure thermal stability during steady-state operation. This engine configuration is commonly employed in research settings due to its mechanical simplicity, ease of instrumentation, and suitability for controlled combustion studies. The engine operates under atmospheric intake conditions without forced induction, which allows for the evaluation of fuel combustion characteristics without the influence of intake pressure boosting. Water cooling is utilized to regulate cylinder wall temperatures and maintain consistent thermal boundary conditions throughout the experimental runs, thereby enhancing the repeatability and accuracy of the measured parameters. The detailed geometrical and operational specifications of the test engine are comprehensively presented in Table 1, providing the foundational parameters required for numerical modeling and performance evaluation.

Table 1. Experimental engine specifications

Engine Model	Antor AD510 BS
Engine Type	Diesel Engine
Number of cylinder	1
Course Duration	90 mm
Cylinder diameter	85 mm
Cylinder Volume	510 cm^3
Connecting rod length	145 mm
Compression ratio	17.8:1
Injection pressure	650 bar
Engine Speed	2000 rpm

B. Numerical Model Validation

The simulation framework is constructed with a simplified single-cylinder engine configuration comprising one fuel injector, a single combustion chamber, one intake valve, one exhaust valve, an engine block, and four discrete flow channels. Both intake and exhaust systems are modeled with individual air flow paths to accurately capture gas exchange dynamics. The simulations are performed using the Ricardo WAVE software, which incorporates a two-zone thermodynamic model for the in-cylinder processes. This approach allows for a more precise representation of temperature and composition gradients within the cylinder during the combustion cycle.To model the combustion process of various fuel blends, a multi-component, multi-fuel Wiebe function-based combustion sub-model is utilized. This enables the prediction of heat release rates for blended fuels with different combustion characteristics. Furthermore, in-cylinder heat transfer is calculated using the Woschni correlation, a widely adopted empirical model that accounts for variations in engine operating conditions such as pressure, temperature, and piston speed, thereby enhancing the accuracy of thermal boundary condition estimations during engine cycles.

A critical prerequisite for ensuring the credibility and predictive capability of any numerical simulation is the rigorous validation of the computational model against reliable experimental data prior to performing advanced parametric or performance analyses. In the present study, the validation procedure was implemented by comparing the in-cylinder pressure profiles as a function of crank angle degree (CAD) obtained from the simulation with corresponding experimental measurements under identical engine operating conditions. The comparison focused on key combustion characteristics, including peak cylinder pressure, the timing of maximum pressure, and pressure rise rates. The simulation outputs exhibited a high degree of concordance with the experimental data, indicating that the employed physical sub-models and boundary conditions were appropriately selected and calibrated. The observed discrepancies between the numerical and

experimental results were within an acceptable error margin, thereby confirming the model's reliability and its suitability for further predictive investigations involving alternative fuels or varied engine parameters. The simulation conditions considered for the validation process are given in Table 2.

Table 2. Simulation conditions

Intake air temperature	300 K
Engine speed	2000 rpm
Start of injection	16°CA bTDC
Injection pressure	650 bar
Amount of fuel injected	20.5 mg/cycle
Compression ratio	17.8:1
Connecting rod length	145 mm
Injection duration	8.25°CA
Spray angle	77°

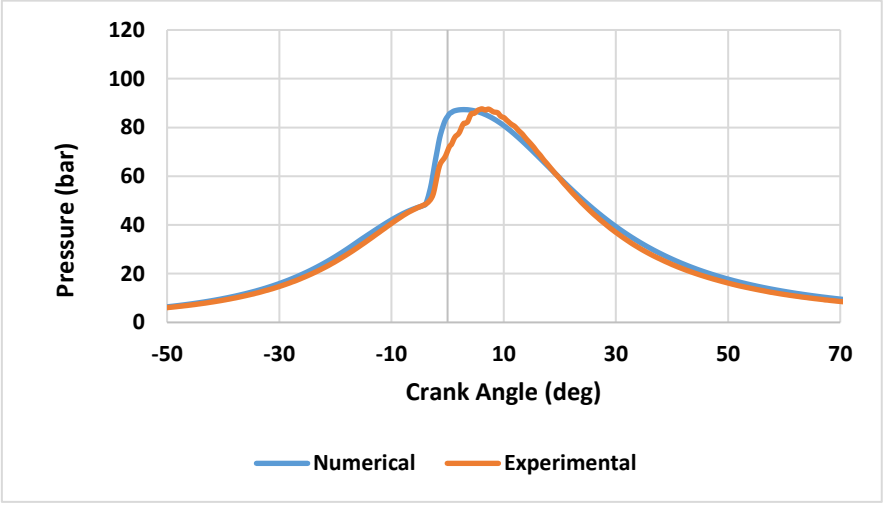


Fig. 2 Validation of experimental data

RESULTS

This study investigates the effects of varying SOI timings on the combustion behavior and exhaust emission characteristics in a single-cylinder diesel-hydrogen dual-fuel engine. Following the validation of both experimental and numerical models, the hydrogen mass fraction was fixed at 30%, and simulations were conducted for three SOI timings: −16°CA, −20°CA, and −24°CA.

It is well-documented in the literature that injection timing is a critical control parameter influencing combustion phasing, in-cylinder pressure development, and emission formation, particularly in dual-fuel applications where the premixed and diffusion-controlled combustion stages interact more intricately [11-12]. In line with these studies, the present findings reveal that advancing the injection timing leads to higher in-cylinder peak pressures and increased rates of heat release, as illustrated in Figures 3 and 4.

Specifically, the highest peak cylinder pressure was recorded as 116.84 bar at 2.1°CA before top dead center (bTDC) for SOI at −24°CA, whereas the lowest value, 89.1 bar at 6.1°CA after top dead center (aTDC), was observed at SOI −16°CA. Advancing the injection from −16°CA to −24°CA resulted in the peak pressure occurring approximately 4°CA earlier, consistent with earlier flame initiation and shorter ignition delay reported in dual-fuel literature. Similarly, the heat release rate (HRR) showed a substantial increase with earlier SOI. The maximum HRR reached 218.5 J/°CA at 4.67°CA bTDC for SOI −24°CA, whereas the lowest was 109.7 J/°CA at 2.15°CA bTDC for SOI −16°CA. These results align with prior research indicating that advancing SOI enhances the premixed combustion phase due to extended premixing time and elevated charge reactivity [13-14].

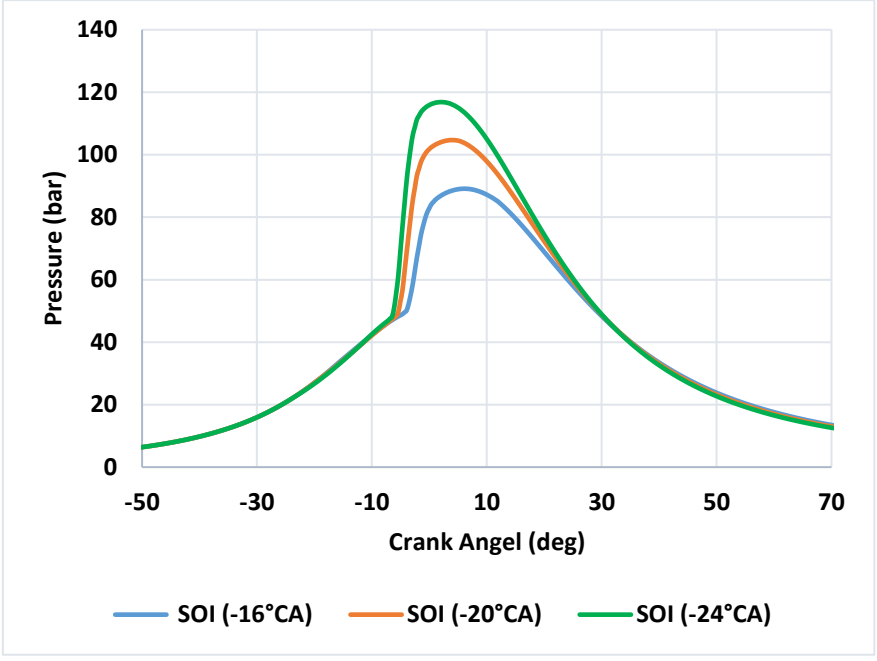


Fig. 3 SOI vs. pressure

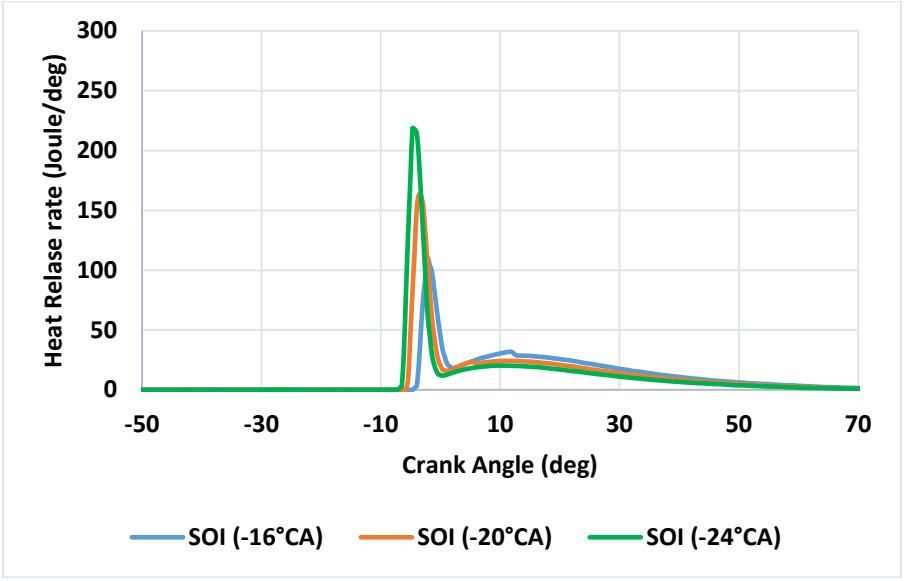


Fig. 4 SOI vs. HRR

The influence of injection timing on exhaust emissions is presented in Figures 5–7. It was observed that increasing the advance of SOI led to elevated NO_x emissions, primarily due to higher combustion temperatures and in-cylinder pressures. The NO_x concentrations increased from 1111.03 ppm at −16°CA to 1294.88 ppm at −20°CA, and peaked at 1534.40 ppm at −24°CA, which concurs with established thermodynamic principles and prior experimental findings [15-16]. Conversely, CO emissions showed a declining trend with advanced injection, dropping from 166.53 ppm at -16°CA to 78.71 ppm at -24°CA. This can be attributed to improved oxidation conditions at higher cylinder temperatures and pressures, facilitating more complete combustion, a phenomenon also supported by earlier dual-fuel engine studies [17-18]. Hydrocarbon (HC) emissions, however, did not exhibit significant variation across the tested SOI range, which is consistent with reports indicating that HC emissions in dual-fuel engines are more strongly influenced by mixture homogeneity and hydrogen substitution levels than injection timing alone [19].

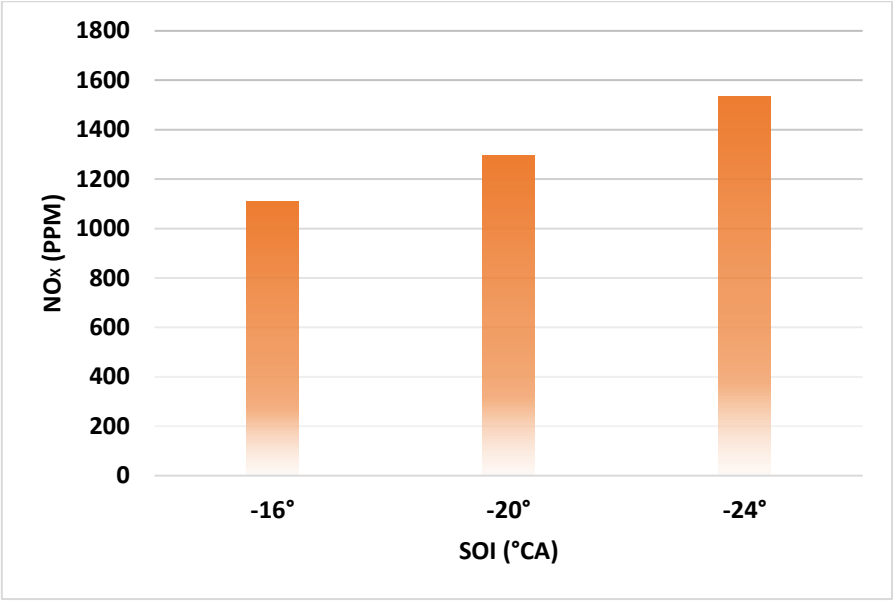


Fig. 5 SOI vs. NO_x emission

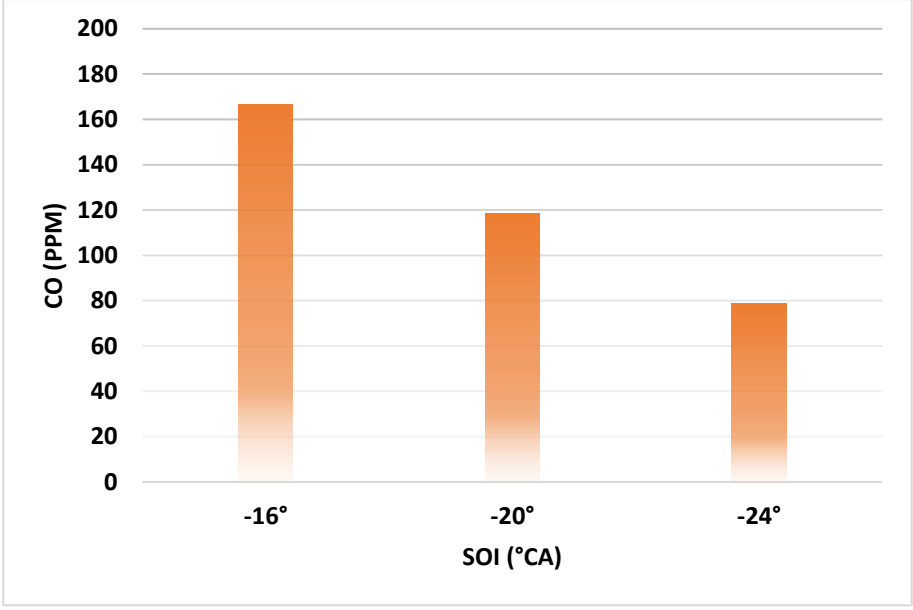


Fig. 6 SOI vs. CO emission

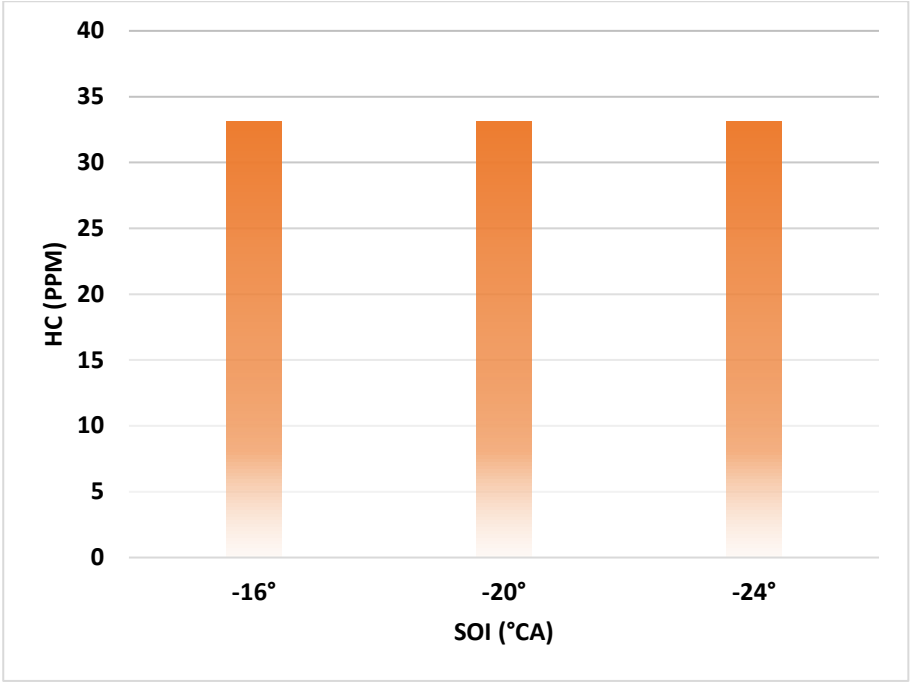


Fig. 7 SOI vs. HC emission

CONCLUSION

The findings of this study demonstrate that start of injection (SOI) timing significantly influences combustion and emission behavior in a diesel-hydrogen dual-fuel engine. Advancing the SOI from -16°CA to -24°CA resulted in a notable increase of 31.1% in peak cylinder pressure, rising from 89.1 bar to 116.84 bar. This was accompanied by an advancement in the crank angle position of peak pressure by approximately 4°CA, indicating earlier combustion phasing due to reduced ignition delay. Similarly, the maximum heat release rate (HRR) nearly doubled, showing a 99.1% increase from 109.7 J/°CA at -16°CA to 218.5 J/°CA at -24°CA. These outcomes align with literature findings that earlier injection enhances premixed combustion characteristics by extending the mixing period and increasing reactivity of the charge mixture.

In terms of exhaust emissions, NO_x levels increased by 38.1% across the studied SOI range, from 1111.03 ppm at -16°CA to 1534.40 ppm at -24°CA, attributable to elevated in-cylinder temperatures and pressures. In contrast, CO emissions decreased by 52.7%, from 166.53 ppm to 78.71 ppm, as a result of more complete oxidation under advanced injection conditions. HC

emissions remained relatively unchanged, supporting the notion that unburned hydrocarbons in dual-fuel engines are more sensitive to fuel-air mixing quality and hydrogen substitution than to injection phasing alone.

Overall, these results suggest that injection timing optimization can serve as an effective strategy to improve combustion efficiency and reduce incomplete combustion by-products in diesel-hydrogen dual-fuel engines, albeit at the expense of higher NO_x emissions. Therefore, for practical implementation, the trade-off between efficiency gains and NO_x mitigation strategies must be carefully considered.

REFERENCES

[1] Demirbas, A. (2007). Combustion systems for biomass fuel. *Energy Sources, Part A*, *29*(4), 303-312.

[2] Gürbüz, H., & Yılmaz, M. (2020). Euro 6 normları çerçevesinde dizel motorlarda egzoz emisyonlarının düşürülmesi üzerine bir inceleme. Mühendislik Bilimleri ve Tasarım Dergisi, 8(1), 94–104.

[3] Gumus, M., Sayin, C., & Canakci, M. (2010). Effect of fuel injection timing on the injection, combustion, and performance characteristics of a direct-injection (DI) diesel engine fueled with canola oil methyl ester− diesel fuel blends. *Energy & fuels*, *24*(5), 3199-3213.

[4] Agarwal, A. K., Srivastava, D. K., Dhar, A., Maurya, R. K., Shukla, P. C., & Singh, A. P. (2013). Effect of fuel injection timing and pressure on combustion, emissions and performance characteristics of a single cylinder diesel engine. Fuel, 111, 374-383.

[5] Jia, M., Xie, M., Wang, T., & Peng, Z. (2011). The effect of injection timing and intake valve close timing on performance and emissions of diesel PCCI engine with a full engine cycle CFD simulation. Applied Energy, 88(9), 2967-2975.

[6] Okcu, M. (2024). A comprehensive experimental analysis on biodiesel/toluene ratio and SOI change in a diesel engine under medium load conditions. International Journal Technology, 21(14), 9111-9126.

[7] Xu, L., Dong, H., Liu, S., Shen, L., & Bi, Y. (2023). Study on the Combustion Mechanism of Diesel/Hydrogen Dual Fuel and the Influence of Pilot Injection and Main Injection. Processes, 11(7), 2122

[8] Liu, X., Liu, S., Shen, L., Bi, Y., & Duan, L. (2023). Study on the Effects of the Hydrogen Substitution Rate on the Performance of a Hydrogen–Diesel Dual-Fuel Engine under Different Loads. Energies, 16(16), 5971.

[9] Bhagat, R. N., Sahu, K. B., Ghadai, S. K., Kumar, C. B., & Saha, A. K. (2024). Effect of Injection Timing on Performance and Emission of a Diesel Engine using Fuelled by Hydrogen Gas. J. Environ. Nanotechnology, 13(3), 400-407.

[10] Liu, Y., Xu, B., Jia, J., Wu, J., Shang, W., & Ma, Z. (2015). Effect of injection timing on performance and emissions of di-diesel engine fueled with isopropanol. In 2015 International Conference on Electrical, Electronics and Mechatronics (pp. 133-137). Atlantis Press.

[11] Kumar, R., Singh, R. K., & Agarwal, A. K. (2013). Effect of fuel injection timing on performance, emissions and combustion characteristics of a dual-fuel diesel engine using hydrogen. International Journal of Hydrogen Energy, 38(1), 121–131.
[12] Maji, S., Banerjee, R., & Sahoo, N. (2020). An experimental study on the effect of injection timing in a hydrogen-diesel dual-fuel engine. Fuel, 262, 116556.
[13] Saravanan, N., Nagarajan, G., Kalaiselvan, K. M., & Dhanasekaran, C. (2008). An experimental investigation on hydrogen as a dual fuel for diesel engine system with exhaust gas recirculation technique. Renewable Energy, 33(3), 422–427.
[14] Li, L., Jia, M., Xie, M., & Wang, Z. (2015). Effect of injection timing on combustion and emission characteristics of a diesel engine fueled with diesel/PODE blends. Fuel, 159, 957–963.
[15] Sayin, C., & Canakci, M. (2009). Effects of injection timing on the engine performance and exhaust emissions of a dual-fuel diesel engine. Energy Conversion and Management, 50(1), 203–213.
[16] Yilmaz, N., Vigil, F. M., & Davis, S. M. (2014). Effect of injection timing on emissions from a diesel engine fuelled with biodiesel and diesel blends. Fuel, 140, 59–65.
[17] Park, S. H., Lee, C. S., & Kim, H. J. (2010). Combustion and emission characteristics of a diesel engine with water-emulsified biodiesel. Energy & Fuels, 24(4), 2479–2486.
[18] Qi, D. H., Chen, H., Geng, L. M., & Bian, Y. Z. (2011). Experimental studies on the combustion characteristics and performance of a direct injection engine fueled with biodiesel/diesel blends. Energy Conversion and Management, 51(12), 2985–2992.
[19] Nayak, S. K., Mishra, P. C., & Das, L. M. (2016). Hydrogen–diesel dual-fuel engine: A review of combustion, performance and emission characteristics. Renewable and Sustainable Energy Reviews, 60, 1335–1348.

The Evolution and Future Prospects of Web GIS: From Static Maps to Intelligent 3D Systems

Tasnim ALASALI[1*]

Omar DAKKAK[1*]

Ismail Rakip KARAS[1]

Yousef FAZEA[2]

[1]Department of Computer Engineering, Karabük University, Karabük 78050, Türkiye
[2]Department of Computer Sciences and Electrical Engineering, Marshall University, Huntington, USA
[*](omardakkak@karabuk.edu.tr)

ABSTRACT

The landscape of web Geographic Information Systems (GIS) and online geospatial information has significantly transformed in recent years. It is now evident that nearly everyone globally interacts with mapping data, often without even being aware of it. Today, location services are embedded in almost all smartphones, and every place and object on Earth can be pinpointed. The expansion of the Internet has accelerated the proliferation of this location data. Many geospatial data are now accessible and continually collected online, and they are utilized in web applications and maps for visualization, analysis, modeling, and simulation purposes. This study examines the progression of web mapping from early static map images on the web to today's dynamic, multi-source web mapping services hosted increasingly on cloud platforms. It highlights the dynamic interplay among three online elements: geospatial data, users, and functionalities. The paper discusses the evolving trends and interactions within these elements in light of technological advancements and potential future developments and opportunities.

Keywords – 3D GIS; Web GIS; Internet; Cloud GIS; Geographic Information Systems; semantic Internet; Digital Globe Mapping.

INTRODUCTION

In 1998, a significant speech by U.S. Vice President Al Gore highlighted the importance of digital earth technologies, sparking a global recognition of the importance of making geospatial information widely available online. Gore envisioned a system where technology and human interaction could create a vivid virtual journey through geographic spaces and epochs, aiming to deepen our understanding of the planet [1]. He underscored the potential of such technologies for various applications, including virtual diplomacy and crime prevention, biodiversity conservation, climate change forecasting, and enhanced agricultural productivity. Integration of digital globes, high-speed networks, mobile technologies, location services, cloud computing, and Web GIS have been some of the most noteworthy technical developments and collaborative platforms to emerge in the last 20 years. Challenges still exist in areas such as managing massive amounts of geospatial data, extracting, and interpreting meaningful insights, getting a full picture of our digital world, using crowdsourcing, and communicating contextualized information globally, despite these advancements [2], [3]. A lot has changed in the relationship between data and people using it since the World Wide Web first went live in 1994. Since its inception, the World Wide Web has transformed from a static platform for content generation into a dynamic environment where users may collaborate, communicate, and

build their own content [4]. Both technology and user applications continue to evolve in a cyclical fashion as a result of this progress, which in turn generates new demands from users. In addition, there has been a sea change in the technology behind the ways and tools used to provide online mapping services. Initially, web mapping applications were rudimentary tools that permitted entities to display maps online with limited user interaction options, such as browsing, zooming, panning, and toggling layers. Initially, users were primarily passive observers of the data rather than active contributors. However, contemporary web mapping applications offer a considerably broader range of functionalities, extending beyond the conventional capabilities of desktop GIS software to meet the evolving demands of modern applications and facilitating enhanced sharing and analysis of map data online.

This paper aims to trace the progression of web mapping by considering the viewpoints of its creators (experts and scientists) and its users (from beginners to occasional users). The structure of this paper is as follows: Section 2 provides an overview of web mapping, including its definition, key features, and a framework for distinguishing different web mapping periods. The subsequent two sections examine the evolution of web mapping across various stages, including static, dynamic, service-oriented, interactive, collaborative, digital globe, mobile, cloud-based, and intelligent in Section 3, and then examine future trends and directions in web mapping in Section 4. Finally, Section 5 concludes the paper with our findings and reflections.

CONTEXT AND PROGRESSION

The field of web mapping has undergone a remarkable evolution, marked by significant advances in both scientific and technological domains. Additionally, there has been a growing interest among experts and everyday users in generating new types of maps through web services. The following sections will examine the concept of web mapping, offering various definitions and characterizations that facilitate mapping key developments and milestones across different periods of web mapping evolution.

A. What is Web Mapping?

Before delving deeper, it is useful to explore some current definitions of web mapping:

- Web mapping involves: (1) updating of information and data, (2) security measures and user authentication, (3) collection of data through user-filled forms such as surveys, and (4) database access [5].
- Web mapping is developing, executing, producing, and delivering maps on the web [6].

- Web mapping uses information system-obtained maps that process spatial and geographical data [7].
- The utilization of maps that are disseminated via geographical information systems (GIS) is referred to as "web mapping." [8].

Beyond these formal definitions, we also consider the user-contributed definitions from Wikipedia for their accessibility and the dynamic nature of their content, reflecting the collaborative input from a broad user base. In defining web mapping, it is obvious that these descriptions, while useful, are somewhat limited. A comprehensive understanding requires placing web mapping within the specific context of its development period. Web mapping inherently involves three elements: geodata or geoinformation and their representation through maps, geospatial software, and the Internet, specifically the Web. The interpretation of web mapping can vary. Some differentiate it sharply from Web GIS, noting blurred lines between them, while others use the terms interchangeably. For clarity, we define Web GIS as an advanced GIS that utilizes web technologies to facilitate communication among its components data, functionality, and user interface. On the other hand, web mapping is broader, emphasizing map provision and supporting varying complexity levels for different applications and users. Additional terms frequently conflated with web mapping and web GIS are distributed GIS, Internet GIS, and GeoWeb. Internet GIS [9] is frequently synonymous with Web GIS. However, it encompasses a broader range of Internet services beyond just the web, making it a more inclusive but less commonly used term than Web GIS. GeoWeb is sometimes defined as integrating geospatial and non-geospatial information (like photos and videos) [10] or as synonymous with Distributed GIS, a geospatial environment allowing for shared interoperable components. Despite the term's widespread use, the concept and practice of web mapping have significantly changed due to advancements in data interaction, Internet and web technologies, tool accessibility, and user attitudes toward these platforms [11]. Web mapping involves websites with mapping capabilities, which have expanded over time as user interest in consuming and contributing to these sites has grown. To better understand and clarify the nuances of web mapping, it is helpful to refer to the specific periods named in this paper, such as "static web mapping" for the early days of publishing static maps online and "intelligent web mapping" for the current period, where context greatly influences user interaction with geographic information. This period-specific nomenclature helps resolve the lack of clarity around the phrase "web mapping".

B. Characterization of Web Mapping Developments

Web mapping utilizes the web, the largest multimedia platform in the world, to distribute maps. This requires following Internet and web-specific protocols and a thorough understanding of mapping theories and associated

fields. A combination of technological advancements and theoretical knowledge characterizes Web mapping. It necessitates continuous innovation in both technology and cartographic principles to provide web-based mapping services successfully. In contrast to pre-packaged desktop mapping solutions, web-based mapping usually requires a certain level of customization and programming. Implementing good applications requires diverse skills, expertise, and often intricate organizational structures [12]. In the web environment, maps and related content are accessed through web browsers, necessitating interfaces that are well-designed and optionally equipped with functionalities for queries and reporting. Considerations such as the target audience, interface dimensions (size, space, color), and responsiveness are crucial when designing web maps. The interactive nature of web mapping dictates that operations are nearly instantaneous to ensure the logic flow and retain user engagement. Constraints also arise from the limited display areas available in web browsers and mobile devices, necessitating new rules for cartographic design tailored to web applications [13] many of these principles have been extensively explored. The shift towards content-based and user-generated web mapping has led to significant developments. Batty et al. [14] identify six basic features of web mapping as well as neo-geography: (1) the ability for individuals to create and contribute material; (2) the use of collective power from the public; (3) the development of vast amounts of data; (4) the incorporation of participatory principles; (5) the influence of network effects; and (6) the emphasis on openness. Efforts to record web mapping development often concentrate on technical progress, either on a general scale or in connection with notable technology projects. These traits are now widely recognized as distinguishing aspects of the online mapping landscape.

C. *Defining Criteria for Web Mapping Periods*

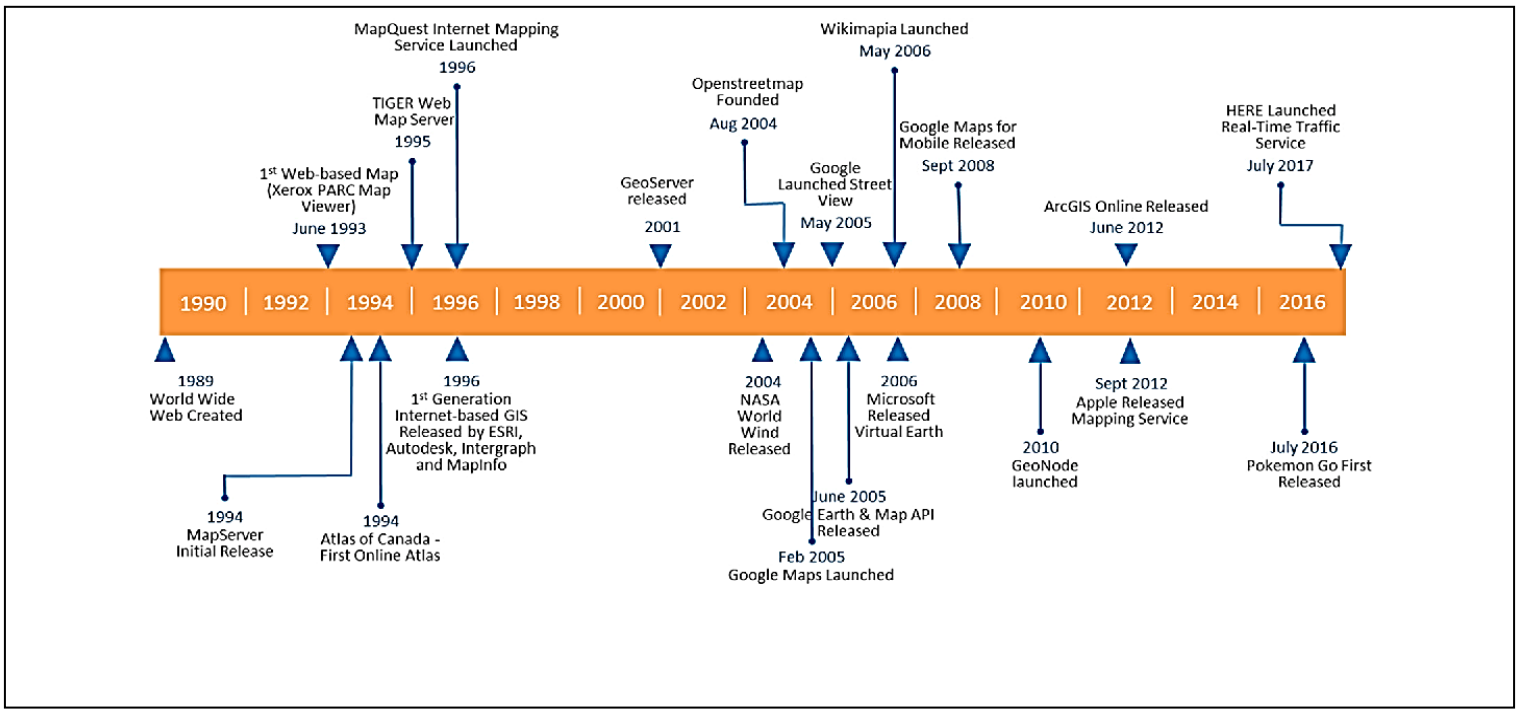


Fig 1. History of notable web mapping milestones [17]

Shortly after the inception of the World Wide Web, the early advancements mostly focused on producing online maps. The PARC Map Viewer was the first notable web-based map released [15]. Although web mapping has a relatively brief history, its advancements, driven largely by web technology innovations, have been substantial beyond what could have been envisioned two decades ago. This section briefly outlines how various authors categorize these advancements into distinct periods or generations, following a framework elaborated by Veenendaal [16]. Figure 1 provides a timeline overview of significant events in web mapping history [17]. The development of the web itself has been segmented into generations from Web 1.0 through Web 4.0 and beyond, each defined by distinct technological and user interaction characteristics [18], [19]. Web 1.0, the Hypertext Web, facilitated basic information retrieval using HTTP and HTML [18]. Web 2.0 introduced a collaborative, read-write format, while Web 3.0, the Semantic Web, enhanced machine-readable data to improve user relevance and interaction [20].

Web 4.0, described as an ultra-intelligent agent, pushes towards highly interactive community collaboration [19]. Speculative discussions of Web 5.0 envision a Sensory and Emotive Web, but its definition remains unsettled [21]. Various authors have identified specific stages reflecting the integration of emerging technologies and changing user interactions by mapping these general web developments onto web mapping. Initial stages included simple static map publishing and evolved into more interactive and dynamic forms. Notably, in 2005, significant developments were triggered by the increased public awareness of GIS capabilities through major events and technological advancements like AJAX and image tiling, enhancing web map responsiveness and leading to virtual globes and mobile GIS services [22]. Plewe [23] identifies four generations of web maps, characterized by their underlying technologies, from simple static HTML maps to dynamic HTML maps using Java and ActiveX, followed by AJAX-driven interactive maps, and ultimately to virtual globe services providing 3D experiences with tools like Google Earth and Microsoft Bing. This evolution expanded map access from a niche group of developers to the public, fostering a broader interest in geographical knowledge. Building on these stages, Tsou [24] introduces a fifth generation that prioritizes cloud computing, rich Internet applications, and crowdsourcing. Community involvement and interactive web mapping are two examples of how this generation is pushing for more user-friendly interfaces and powerful online tools. In general, the evolution of online mapping follows the same pattern as the rest of web development: from simple document sharing to more intricate, community-oriented platforms that globalize data integration, social interactions, and geographic settings. However, the timeline in Figure 2 has been updated to include key technological advancements from 2018 to 2024, highlighting significant developments that have further shaped the field of web-based GIS.

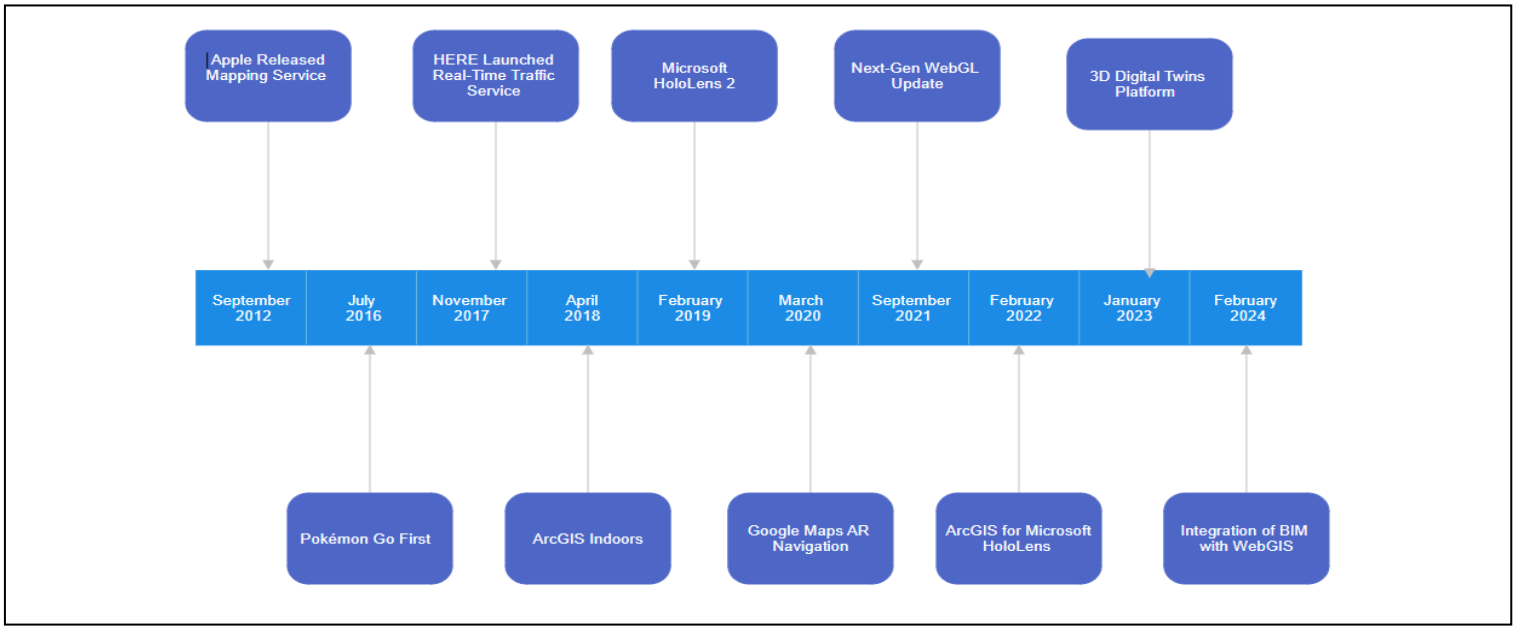


Fig 2. Modified from to include updated data and current technological trends [18].

ArcGIS Indoors was introduced in 2018 to offer detailed indoor mapping for intricate environments such as airports and campuses for the improved spatial analysis of structures[25]. The next year, Microsoft released HoloLens 2 in February 2019 with upgraded mixed reality that combines spatial data with physical surroundings [26]. Google Maps AR Navigation [27] that was launched in March 2020 added augmented reality to mobile navigation by displaying directions on the real-world scenery captured by the smartphone camera. In September 2021, Next-Gen WebGL was released which enhanced the web-based 3D graphics rendering and thus the interactivity and appeal of online GIS applications [28]. ArcGIS for Microsoft HoloLens was launched in February 2022 to provide users with 3D GIS experiences within the HoloLens environment which helps users to better visualize and interact with GIS data. The creation of the 3D Digital Twins Platform in January 2023 helped to advance the development of a tool for the creation of accurate digital representations of real-world spaces for urban design and infrastructure planning. Last but not least, the adoption of Building Information Modeling (BIM) combined with WebGIS in February 2024 will bring architectural models with geographic information systems to enhance construction projects and city planning.

STAGES OF WEB MAPPING EVOLUTION

As mentioned earlier, web mapping has gone through many key stages in the last several decades, each with its own set of events and traits that have shaped its use and trajectory. For the sake of efficient categorization, we have distinguished between several web mapping periods, with notable breakthroughs defining each. The criteria for establishing a period are laid out in this part, followed by a discussion of the major aspects that define each.

A. Criteria for Identifying Web Mapping Periods

A " period" is described as a period that is characterized by certain features or significant occurrences. In web mapping, periods are delineated based on significant developments that have notably impacted the global community. These developments are not strictly technological; they also encompass user adoption and the evolution of social networks, as seen in the frameworks provided by Hall and Tiropanis [4] and others discussed in Section 2.3. The criteria for defining these periods are thus multi-faceted, reflecting the interplay between technological advancements and user engagement. Key points include:

- Significance: Developments must notably influence web mapping technologies and the community.
- Driver: Influences can be technological, user-driven, or combined, affecting adoption and usage.
- Continuity: Developments may start in one period but continue to evolve and influence subsequent periods.
- Overlap: periods may overlap as developments can coincide or within similar timeframes.
- Catalytic Nature: Developments in one period often set the stage for subsequent innovations.

B. Criteria for Identifying Web Mapping Periods

This paper's division of web mapping periods is based on research that Veenendaal [16] elaborated on and discussed in Section 2.3. As set out in Section 3.1, the framework for defining these periods helps clarify and categorize the distinct phases and their notable web mapping contributions. It is important to note that these periods vary in their focus. Some are primarily driven by technological advances (e.g., the Ser-vices period). In contrast, others are driven by user engagement (e.g., the Collaborative period), yet others are driven by a combination of both (e.g., the Digital Globe period). In total, nine web mapping periods have been established, each marked on a timeline that highlights key developments. Table 1 provides a comprehensive overview of the nine web mapping periods, outlining their primary focus, key developments, and notable examples. As we'll see in the sections that follow, it provides a useful foundation for learning about the history and impact of online mapping tools. These epochs of online mapping provide light on the field's development, showing how changes in technology and user demands have influenced and will influence the future of web mapping.

C. The period of Static Web Mapping

This period, which aligned with the early days of Web 1.0 in the 1990s, is recognized as the foundational period for web-based mapping. Web maps were primarily static HTML images with hyperlinks and clickable areas during this period marking the initial phase of web-based geographic information sharing [29]. This period corresponds to Plewe's first production of web maps and was primarily defined by the web's read-only nature, with a concentration on retrieving information from a repository of data consisting of web pages. HTTP and HTML technologies established the techno-logical groundwork for this period, allowing for the online broadcast of connected in-formation according to Berners-Lee's vision [30]. The primary design principle was client-server, in which a local computer (the client) spoke with a distant computer (the server) to get the needed data. New features, such as

Common Gateway Interface (CGI) and HTML Forms, increased flexibility by enabling user input and more dynamic processing on the server side [9].

The Xerox PARC map server stands out now because it allows users to manipulate map settings like zoom and layer visibility using simple hyperlinks inserted in HTML files [16, 19]. By allowing map storage and retrieval on the web, this period laid the framework for more advanced web mapping innovations, making maps more accessible than conventional paper-based approaches.

D. The Dynamic Web Mapping period

Standardizing interfaces for online map servers was a driving force for creating the Services Online Mapping period. This enabled both direct user interaction and programmed access via software applications. This period utilized technologies like Java Applets, CORBA, and Microsoft.NET to create interoperable and modular com-ponents within a Service-Oriented Architecture (SOA) [24]. This period was characterized by the development of open-service APIs, which allowed for extensive customization and integration of web maps with user-generated data, creating mashups. Examples of APIs from this period include OpenLayers, Leaflet, D3, Google Maps API, Microsoft's Bing Maps, and ESRI's ArcGIS APIs, which facilitated the development of rich, interactive web mapping applications (ESRI, Microsoft Bing Maps, Google Maps API, D3.js, Leaflet, OpenLayers). A groundbreaking example of service during this period was providing U.S. Census data via TIGER files, accessible over a web map server in 1995 [31], [32]. This event indicated the beginning of this period. TIGERweb, which eliminated this service, showcased web mapping services' growing functionality and effectiveness. This period also saw significant contributions from the Open Geospatial Consortium (OGC), which developed standards like Web Mapping Services (WMS), Web Feature Services (WFS), Web Coverage Services (WCS), and Web Processing Services (WPS), establishing a framework for interoperability and functionality in web mapping [33], [34], [35], [36]. Web mapping has come a long way in the last few decades, thanks to a confluence of factors including improved technology and an ever-expanding user population that has a hand in molding the future of online geo-graphic information systems.

E. The Interactive Web Mapping period

The limits of synchronous client-server interactions were addressed during the Interactive Web Mapping period, which represented a substantial advancement from the Dynamic Web Mapping period. Image tiling and Asynchronous JavaScript and XML (AJAX) allowed for a more refined and responsive user experience in this period. By allowing data to load in the background while users interact with the map, AJAX greatly improved the user experience and drastically cut down on wait times [9]. The idea of

geographic mashups, which allowed for the dynamic combination of data from several sources on one map, was also established around this time, substantially im-proving the online mapping experience [37].

F. The Collaborative Web Mapping period

Web 2.0's debut propelled this period since it expanded upon the opportunities presented by the Interactive Web Mapping period. During this time, there was a push for more user-generated content and for people to actively participate in making maps and creating data. This period is characterized by platforms like OpenStreetMap and Wikimapia, which allow individuals worldwide to edit and contribute to map data, hence democratizing GIS [38]. A plethora of community-driven platforms mush-roomed during this time, allowing people to both consume and contribute geospatial data, creating a fertile ground for collaborative mapping and the exchange of information.

When employing Participatory GIS (PGIS), the rights of users are improved, and data assets are boosted because people contribute to data gathering as well as analysis. Relative to the top-down use of GIS technology, this approach democratizes GIS technology, makes it easy for other stakeholders who are not data scientists to use GIS [39]. Thus, PGIS engages the local expertise and incorporates it with spatial data to address community's demands and issues. With the assistance of local people, the GIS data becomes more precise and pertinent, since local participants are more familiar with the circumstances local for experts and enthusiasts. In general, PGIS enhances the relationship between GIS technology and the development of the community to guaran-tee that the spatial data collected represents and is beneficial to the community it be-longs to.

G. The Digital Globe Web Mapping period

An unprecedented number of people throughout the world now have access to geospatial data and mapping capabilities, thanks to the proliferation of digital globes on the web. According to Tsou and Plewe, digital globes have come a long way. These tools make use of new technology to provide users with richer 3D environments in addition to 2D Earth images, allowing for more immersive experiences overall [40]. Plat-forms such as Google Earth and Microsoft's Bing have been instrumental in making maps and digital earth imagery accessible to the public, thereby facilitating re-al-time, synchronous sharing of views across various devices, including desktops and mobile devices via web browsers. These tools have evolved from rudimentary 3D representations of the Earth to sophisticated platforms that integrate with virtual reality (VR), thereby enhancing applications in diverse domains, including gaming, citizen science, and grassroots mobilization [41]. The open-source movement, which includes tools such as NASA

World Wind and Cesium WebGL, has further facilitated the development of virtual globe applications, making it easier for developers to create sophisticated web mapping interfaces. The availability of these technologies has expanded the user base from specialized experts to virtually anyone with Internet access, broadening the scope and utility of geospatial data. In 2024, a new platform called CityScape VR will let users explore their city in 3D. They can see how new buildings and roads will look and give their opinions. In a virtual reality setting, users may have virtual town hall meetings, investigate the effects of development, and provide suggestions. A more collaborative approach to city development is encouraged by this tool's ability to make urban planning more accessible and clearer. These developments demonstrate the potential for community engagement using 3D GIS on the web. They go on to demonstrate its applicability in fields such as emergency management and tourism. Individuals may gain insight and improve their reaction time via immersive experiences. Building Information Models (BIM) technology plays a critical role in incorporating 3D GIS applications to provide significantly better building and infra-structure modeling [42]. The incorporation helps to provide a better management and visualization of data on geographic information systems which in turn boosts planning and decision-making processes.

H. The period of Mobile and Location-Based Web Mapping

The broad usage of mobile devices, such as smartphones and tablets, propelled this period because they enabled people to access and interact with maps at any time and from any location. One major development of this age is location-based services (LBS). These services improved the relevance and customization of apps by using methods like GPS and mobile phone tower triangulation to determine user locations [43]. Services for real-time navigation, emergency responses, marketing, and even lei-sure pursuits like location-based gaming were all major innovations during this period. Take Pokémon Go, for instance. It uses augmented reality to superimpose game features on real-world surroundings. Other applications may now provide navigation based on a user's actual position, recommend local businesses, or even include users in games that demand physical movement in the real world [44]. Another notable feature of this period is participatory sensing, which further blurs the boundaries between us-ers and content providers by allowing users to participate in data collecting and sharing while interacting with their surroundings. Tourist engagement applications are one example; they allow users to document and share their discoveries in real-time as they travel [45].

Health and fitness monitoring was further advanced with the integration of location data with mobile health applications, which allowed for the recording of physical activity and vital signs. With the advent of real-time location-based routing information, traffic monitoring, and dynamic re-

routing depending on current circum-stances, the transportation sector likewise saw revolutionary advances. Mobile map-ping systems that used a variety of sensors, including LiDAR and cameras, to gather massive volumes of geographical and environmental data from vehicles in motion also saw significant improvements in sensor integration during this time. The Open Geo-Spatial Consortium (OGC) has established the Sensor Web Enablement (SWE) standards to facilitate the transmission of sensor location and parameters over the web, catering to the expanding market of applications.

THE PERIOD OF CLOUD-BASED WEB MAPPING

The online mapping revolution of the 21st century, powered by the clouds, is a watershed moment in the history of geospatial technology. This age draws on cloud computing to improve the scalability, accessibility, and reliability of mapping applications and infrastructure. It is a component of the fifth generation of online mapping. Cloud computing offers many essential services, such as:

- Software as a service (SaaS): This approach eliminates the requirement for us-ers to install apps locally by allowing them to access applications directly via a web browser. A prominent example is ESRI's ArcGIS.com, which is widely used for various geospatial applications. Additionally, Harvard WorldMap has migrated to ArcGIS.com technology, further demonstrating its significance in the field
- Platform as a Service (PaaS): Google App Engine and similar services provide environments for application development, deployment, and management in a variety of languages. This is especially crucial for one-of-a-kind online map-ping applications that need specialized features.
- Basic computer resources like storage and networking are made available via the Infrastructure as a Service (IaaS) architecture. Essential computing re-sources like storage and networking are provided by providers like Amazon Web Services and Google Compute Engine. When it comes to processing mas-sive amounts of geographical data from sources like satellite images, these ser-vices are crucial for enabling complex analytical activities [46].

The cloud's ability to process large amounts of data is especially crucial for geospatial data collected from sensors and mobile devices. Platforms such as Hadoop, Hive, and MongoDB provide the required architecture for distributed and parallel computation to analyze geographic big data. Additionally, the period is characterized by integrating the Internet of Things (IoT) with web mapping, where sensors connected to the Inter-net

continuously generate data that is efficiently managed and analyzed in the cloud. With this connectivity comes the "location of things," a term that describes how crucial geospatial data is for putting the massive amounts of data gathered from different sensors all around the world into perspective. Web mapping apps that have been around for a while may now function in the cloud. It encourages creative applications like Mapstory, which tells stories via maps, Harvard WorldMap, which is used in the classroom, and Masdap, which helps fund development projects in places like Malawi. Two major foci are the improvement of real-time data accessibility and the use of large geographic datasets. To maximize the use of municipal resources and the creation of infrastructure, urban planners use platforms like ArcGIS, which was developed by Esri. Climate Field View and Agrian are two examples of agricultural platforms that analyze data to help farmers with precision agriculture, which aims to maximize crop yields while minimizing negative environmental consequences. Platforms such as the Humanitarian OpenStreetMap Team (HOT) provide current maps for efficient crisis navigation and coordination, which greatly aids disaster response operations. Similarly, public health emergencies also leverage cloud GIS to track disease spread and manage resources, thereby demonstrating the critical role of cloud-based mapping in addressing contemporary challenges across multiple domains. However, in the course of the development of WebGIS technology, it is necessary to emphasize the great con-tribution of the MapServer software to be the first to set the trend of open-source solutions. MapServer's strong features and adaptability have created the foundation and precedent for the development of WebGIS applications and services, providing a strong support for GIS graphic data display and analysis. Likewise, ESRI's ArcGIS Online is among the market leaders that determine trends in WebGIS environment, offering a wide range of functionalities of cloud-based GIS that are easy to use and access. MapServer and ArcGIS Online can be seen as successful representatives of the development and expansion of WebGIS technologies for a wide range of users from open-source supporters to professional users from large companies, consequently con-tributing to the development and popularization of geographic information systems on the web.

H. The Intelligent Web Mapping period

Intelligent web mapping is when semantic technologies and innovative environments offer contextually enhanced information to users and applications. This emerging period focuses on the process of learning from the data and the subsequent wise use of this knowledge. Semantic and contextual representations have to be built, acquired and used for the purpose of filtering the desired information. Web 3. 0, or the semantic web, is the effort to improve the ability of machines to understand semantically tagged data, providing users with more relevant and useful information [47].

Bishr [48] identifies two types of semantic heterogeneity to be addressed for semantic interoperability: cognitive heterogeneity, which is defined as the existence of different points of view regarding the same objects, and naming semantics, which is the use of different names for the same objects. These challenges are well handled by the ontologies that are a way of representing knowledge through rules and inferences and are usually implemented using languages such as OWL which is part of the RDF family [49].

For an example, the University of Texas case study involved the creation of structured medical records and unstructured doctor's notes from various community databases using health condition ontologies for the purposes of community health and disease surveillance [50]. Likewise, Zhang et al. [51] present a prototype for emergency responders in Connecticut that enables one to search geospatial features across many semantically different data sources autonomously. This system integrates OWL ontologies to improve a natural language interface where the user does not need to use technical terms to query data regarding roads, town boundaries, airports, and so on. The natural language processing parser uses synonyms from WordNet and domain ontologies to understand and connect the relevant information for the query results. Another example is a brokered system developed by Fa et al. [52] which uses OWL ontologies to combine different heterogeneous datasets to answer questions concerning state electoral boundaries in Australia.

Recent development areas in the geospatial semantic web are semantic information modeling, ontology creation (automatic and semi-automatic), semantic web services orchestration, natural language processing, data identification, and efficiency enhancement [51]. It is necessary to build multiple ontologies for data, functionality, and application execution [53]. Moreover, intelligence is used to adapt the results of que-ries to the users and applications, identifying the users and applications and using extra semiotic information about the context of the query—geographical, temporal, at-tributed, semantical, historical, etc. It is, therefore, still speculated that 0 will continue to be developed, especially in the geospatial domain. The emphasis is shifting towards improving the interaction between man and machine with more reliance on automat-ed and semi-automated decision-making.

Table 1. Evolution of Web Mapping Technologies and Key Developments.

Era	Focus	Key Developments	Notable Examples
Static Web Mapping	Viewing	Static HTML images with hyperlinks; foundational web map interaction	Xerox PARC map server
Dynamic Web Mapping	Interactivity	Dynamic HTML (DHTML), CGI, allowing real-time map creation	Atlas of Canada (1994), National Geospatial Data Clearinghouse
Services Web Mapping	Standardization & API Integration	Introduction of service-oriented architecture with APIs for mapping	OpenLayers, Leaflet, Google Maps API
Interactive Web Mapping	Enhanced User Interaction	Use of AJAX and image tiling for smoother user experiences	Geographic mashups
Collaborative Web Mapping	User-generated content	Web 2.0 technologies enhancing user participation	OpenStreetMap, Wikimapia
Digital Globe Web Mapping	3D Visualization	Technologies for immersive 3D mapping experiences	Google Earth, Microsoft Virtual Earth
Mobile and Location-Based Mapping	Mobility & Real-time Location	Advancements in mobile GPS and location-based services	Pokémon Go, participatory sensing applications
Cloud Web Mapping	Cloud Computing & Big Data	Utilization of cloud services for data and functionality storage and processing	Mapstory, Harvard WorldMap, Masdap
Intelligent Web Mapping	Semantic Web, AI, AR, and VR	Integration of semantic technologies, AR, and VR for smarter data processing and visualization	Semantic web services, ontology use in emergency response, ArcGIS for Microsoft HoloLens, ArcGIS Indoors, 3D Digital Twins Platform, Integration of BIM with WebGIS

TRENDS AND DIRECTIONS IN WEB MAPPING

Each period in web mapping has introduced pivotal technological advancements impacting user access and interaction with geographic data. These advancements range from enhanced data accessibility through shared platforms in the Static and Cloud eras to improved responsiveness in the

Dynamic, Services, and Interactive periods and richer interfaces in the Digital Globe, Mobile, Location-Based, and Intelligent periods. By examining these historical advancements, we can identify trends and potential future directions for web and cloud mapping and GIS technologies. This analysis considers three key components of web map usage: data, users, and functionality. We also explore how these trends, such as smart city initiatives, may manifest in future applications.

A. Trends in Web Mapping Usage Components

The utility of web maps fundamentally revolves around their usage, which MacEachren and Kraak [54] originally described through three dimensions: the audience, data relations, and user interaction. Veenendaal [55] adapted this framework for web environments, identifying users, data context, and functionality as critical components. This framework helps us understand the interactions among these components and identify future web mapping trends. Professionals in GIS require advanced tools to analyze poorly contextualized data. Conversely, the growing number of global users with minimal geospatial expertise necessitates interfaces that present synthesized information within a familiar context, reducing the need for direct interaction with complex GIS functionalities. These users benefit from web processing infrastructures that automatically handle data discovery, linkage, analysis, and synthesis. This shift reflects broader trends across web mapping periods, where each period enhances one or more of the three usage components: data, users, and functionality. This enhancement is crucial for advancing web mapping technologies, as demonstrated by existing models of map use. Continued interaction among these components is essential for web mapping tools' ongoing evolution and usability.

WEB MAPPING DATA TRENDS

In the initial stages of web mapping, experts primarily generated geospatial data through methods like GIS digitizing, surveying, photogrammetry, and remote sensing. The data was somewhat limited due to the constraints of early computer networks. Over time, as technologies evolved through periods like the Collaborative, Digital Globe, and Mobile and Location-Based periods, the volume of data expanded dramatically. This increase was fueled by the availability of satellite imagery services online and the rise in user-generated content through various devices and sensors.

Participatory GIS (PGIS) has taken the concept of GIS to another level by incorporating the community in the acquisition of data. This approach deploys GIS technology to the masses where anyone with useful local knowledge and data can participate. PGIS thus incorporates the knowledge

and experiences of the people in the community, hence the data collected is not only voluminous but also relevant and appropriate for the needs of the community. The use of mobile devices, social media, and online platforms enables the community members to gather and disseminate geospatial data in real time, thus enriching and making the data more relevant. One of the main approaches of PGIS is participatory mapping, which entails people who are directly affected by a certain place to contribute their experiences, social network, knowledge, and imagination in developing a map. It is a process where data is gathered, analyzed and displayed spatially for use in decision making involving community mapping and planning. Therefore, the end products are maps that reflect the locals' values and perception, which can be much more valuable than socio-demographic information. Community concerns and issues can be identified by urban planners and researchers, which will help in planning for the community. Participatory mapping can be conducted starting from a few neighborhoods up to the entire city. It is possible to engage both face-to-face and through the use of technology where both the physical and virtual worlds are used. Thus, while synchronous participation could include tabletop mapping with maps, pens and stickers, asynchronous participation can be carried out using online map-based surveys, for example, Maptionnaire. Both have their pros: in-person workshops are all about the people and the feedback is instantaneous while online surveys are convenient and can cover a wider and more diverse audience.

By the Cloud period, an immense amount of data was stored online, facilitated by cloud computing and the proliferation of Internet of Things (IoT) devices, contributing vast amounts of location-based data. By 2024, according to recent industry sources, the Internet will host approximately 1.09 billion websites, with data traffic now exceeding 4.8 zettabytes per year, reflecting the growing volume and complexity of global data exchange. This surge in traffic includes basic web activity and geographic data, with platforms such as Google Earth significantly expanding their archives [56]. From 3 petabytes of imagery in 2016, it has grown to over 20 petabytes by 2024, encompassing more detailed and comprehensive aerial, satellite, 2D, 3D, and historical visual data. This escalation illustrates the exponential growth of digital content and the continuing evolution of the Internet's ability to store and manage vast amounts of information [57].

The surge in data has ushered us into the big data period, characterized by vast volumes of diverse and rapidly accumulating data sourced from drones, aircraft, satellites, mobile mapping systems, and geotagged web content. This period poses challenges in managing and making sense of such vast datasets. The Sentinel missions of the European Space Agency will keep on making important observations and data contributions to Earth in 2024. The Vega-C rockets, which have improved performance and payload capacities,

are set to be used by ESA to launch several Sentinel spacecraft, such as Sentinel-2C and Sentinel-1D. The Copernicus program, of which these satellites are a vital component, seeks to keep tabs on Earth's surface over a range of spectra and situations; this will aid in domains like agricultural management, disaster response, and climate change monitoring [58], [59].On top of that, as part of its PlanetScope and SkySat missions, the private business Planet maintains a huge fleet of Earth observation satellites, which includes 149 spacecraft. Along with the other satellites, PlanetScope contributes over 350 million square kilometers of images across different spectral bands every day, which is a substantial quantity of data [60].

When it comes to processing, analyzing, and managing these massive data sets, cloud-based web mapping is quickly emerging as a major player. The cloud has become indispensable in this data flood management. Web mapping relies heavily on data accessibility, quality, and quantity. Relevance and appropriateness for certain applications are ensured by high-quality data [61]. It is crucial to integrate and analyze this data in order to get useful information. The online mapping sector is placing a greater emphasis on data integration and sophisticated processing to provide valuable insights, using the DIKIW pyramid model as a guide. To help developers make educated decisions, it is important to gather and provide correct data on land usage, infrastructure, and laws in land development. Metadata and ontologies are common tools in this process, which aims to provide end-users with better data contextualization and semantic linking; new standards like HTML5 have made this process even better [62].

WEB MAPPING FUNCTIONALITY TRENDS

In 2024, web mapping and GIS technologies have come a long way, going from basic static map viewing to complex cloud-based systems with a plethora of features like 3D mapping, augmented reality integration, real-time data analysis, and more. Cloud computing has revolutionized geographic information systems (GIS) by making it easier to store, retrieve, analyze, and distribute geographical data. This change is made possible by GIS systems hosted in the cloud, which allow for the easy maintenance and integration of geographical data across different devices and platforms. Because of these developments, GIS is now used in areas where real-time data is essential, such as environmental monitoring, urban planning, and disaster management, in addition to more conventional disciplines [41]. Also, developments in drone and LiDAR technology have been crucial to the rapid evolution of 3D mapping. More precise and easily accessible 3D geospatial data is now at our fingertips, thanks to these innovations, allowing

for more dynamic and comprehensive representations of real-world locations. This is helpful for the building, urban planning, and conservation of natural resources sectors; precise geographical analysis is essential [41].

On top of that, processing and analyzing geospatial data has been revolutionized by combining AI and machine learning. These advancements have made GIS products more robust and easier to use by streamlining data processing operations like feature extraction and categorization. Automated analysis of satellite and aerial photography is now possible thanks to AI applications in GIS, which greatly speeds up data processing and improves the accuracy of the findings [63]. By superimposing digital data onto the physical environment, augmented reality (AR) apps have become an integral part of GIS technology, providing users with more engaging and dynamic means of seeing and analyzing geographic data. The fields of education, real estate, and tourism may all benefit greatly from this [41]. To sum up, in 2024, there will be a marked change towards more interconnected, dynamic, and user-centric systems, as seen by the ongoing development of online mapping and geographic information systems. Using these platforms, decision-making processes in many industries are being improved, and geospatial data collecting, and analysis are being revolutionized [64], [65].

Web Mapping User Trends

Web mapping's development has greatly increased the variety and quantity of users, as well as their engagement with geographic data. Users mostly used basic, web-based static maps during the Static period. Through the Dynamic and Interactive periods of online mapping technology, users were able to interact more dynamically with maps, which in turn impacted map creation and visualization. Significant developments in user interaction emerged during the Digital Globe period, where users could navigate through 3D landscapes efficiently. The Mobile and Cloud periods enhanced user experience by providing access to mapping information and GIS functionalities on mobile and cloud-based platforms, greatly improving accessibility regarding time and location. These advances also made it easier for developers to integrate data from multiple sources and for communities to generate content collaboratively. The significant impact of web mapping in different eras has dramatically expanded access to geospatial data from a tool primarily used by experts and developers to one accessible to the global community. This expansion is illustrated by the dramatic increase in Internet users worldwide. In the early stages, web mapping technology was limited to a niche audience. However, with the advent of more interactive, collaborative, and user-friendly interfaces in 2D and 3D mapping - evident in the Interactive, Collaborative, and Digital Globe periods - the accessibility of web maps has exploded, reaching the average citizen on a global scale. As shown in the accompanying figure, the number of Internet users has

increased from 2.73 billion in 2014 to an estimated 5.35 billion by 2024, nearly doubling. This growth in Internet connectivity has facilitated the widespread adoption and use of web mapping applications, making geographic information more accessible.

As of 2024, the continued usage of Google Maps illustrates the extensive integration of web mapping technologies into daily life. Over one billion people use Google Maps each month across the globe, navigating over one billion kilometers daily. Additionally, 67% of smartphone users worldwide still prefer the application as their preferred navigation app. This broad adoption is reflected in the app's high download rates. Google Maps was the most downloaded map and navigation app in the United States as recently as 2021, recording over 25 million downloads [66], [67]. The influence of Google Maps extends beyond individual users to include a significant presence on the web. Over 5 million active websites and applications utilize the Google Maps Platform weekly, leveraging its comprehensive Application Programming Interface (API) for various, including location tracking, turn-by-turn navigation, and local business information [67]. In 2024, the landscape of web mapping and navigation apps continues to evolve, with significant user engagement and integration across various sectors. Google Maps remains a major player, hosting 1.8 billion monthly average users, while Moovit sees approximately 1.5 billion daily users [68]. The adoption of web mapping is extensive, extending beyond the traditional functionality of navigation to encompass features such as accommodation booking, attraction discovery, and real-time traffic updates, which facilitate a range of planning activities, including tourism and urban development. Furthermore, the technological infrastructure supporting Internet usage, a fundamental requirement for web mapping and numerous other applications, has grown significantly. The global average fixed broadband speed has reached 80.14 Mbps as of 2024. Some countries, such as Singapore, have achieved speeds of up to 242.01 Mbps, which supports faster and more reliable online services [69]. The server technology landscape indicates that Nginx and Apache remain dominant, with 34.3% and 30.0% of websites, respectively, utilizing these technologies. This demonstrates their pivotal role in managing web traffic and hosting [70].

This trend towards increasingly linked, technologically sophisticated services that cater to a wide variety of demands across industries and geographies is reflected in the rise of digital infrastructure and the proliferation of online mapping apps. One of the most notable areas of progress in Web mapping capabilities is the adaptation of GIS technology to fully realized three-dimensional settings. The scope of classic GIS applications is expanded by this transformation, which also provides a new level of interaction and data display. The next part delves into the present state and potential future paths of web-based 3D geographic information

systems (GIS), shedding light on the revolutionary effects it has had on several fields and sectors.

Web Mapping 3D GIS Trends

While height and depth are incorporated into the conventional 2D GIS features, web-based 3D GIS revolutionizes GIS in terms of visualization, analysis, and decision making based on geographic data [71], [72]. These systems afford very realistic representations of the real environment like city environments, geographical features, geology, and meteorological phenomena among others that support better judgment in various fields. For example, it becomes possible to model real-life situations, for example, the improvements to the 3D city models [73] or the study of the interactions of ecosystems in live, dynamic conditions. This technology is useful in areas such as environmental research and emergency response where accurate understanding of physical locations is essential since it enables the stakeholders to better visualize the consequences [74]. Moreover, these systems are web-based which enhances modularity and results in versatile applications in areas such as public health and resource management where multiple users can work on 3D spatial data simultaneously [74]. Through the integration of intricate data and user comprehension, web-based 3D GIS transforms spatial data interaction and analysis and enhances the capability of individuals with different backgrounds and skills, which will probably promote new breakthroughs in urban planning, environmental management, disaster prevention, and other fields. Web based 3D GIS is already on course to revolutionize the manner in which spatial data is handled in the future through the provision of future visualization and analytical abilities in three dimensions.

Cloud computing has played a major role in the improvement of web-based 3D GIS as it improves scalability, accessibility and efficiency of managing and processing spatial data. Cloud computing is a way of storing and processing large amounts of GIS data on cloud-based platforms, which is cheaper and safer than stand-alone systems. For instance, the combination of cloud computing with VR and GIS has enhanced the supply chain for steel structure buildings, thus enhancing user satisfaction and efficiency [75]. Further, cloud GIS allows the real time analysis and visualization of big data which will help in decision making and sharing of data in fields like urban planning and environmental management [75]. Another key development in web-based 3D GIS is WebGL which is a web graphics library that allows for rendering of 3D graphics in a browser without the need for a plugin. This technology helps to display 3D environments on the web as evidenced by the integration of 3D GIS globe navigation models with VR environments that enhances detailed and immersive web navigation. For example, a study conducted at the University

of Twente showed how several 3D data were incorporated into a VR and 3D WebGIS navigation model to improve decision making for different parties [76]. Further, WebGL has been used for learning purposes where e-learners are able to view and interact with 3D objects in web browsers with the help of JavaScript libraries such as ThreeJS, OpenLayers and CesiumJS making the 3D visualization more interactive [77].

Augmented Reality (AR) enhances web-based 3D GIS by adding digital information on the physical world hence enhancing spatial data perception and decision making. For instance, a newly developed deep learning-based mobile AR system can create 3D spatial map and does not require prior registration with the help of AR markers and can map virtual objects to real objects with a high degree of precision using single snapshot based RGB-D data. This approach utilizes instance segmentation and ICP algorithms to increase the accuracy and real-world usability in mobile AR applications [78]. In addition, visualization size, scale, and orientation of AR in relation to the real environment is vital in geospatial applications since it improves the visualization of spatial relations in systems of interest [79]. These improvements also prove that AR can be a useful tool for enhancing the ways of urban planning, designing and managing infrastructures within the framework of web-based 3D GIS.

Web-based 3D GIS is gaining significance in the context of urban planning and smart city where tools for better decision making and involving the public are required. 3D city models provide detailed representations of real cities, and their application can be used in planning new constructions and evaluating the effects of projects [80], [81]. Terrain visualization models are useful in environmental risk assessment [82], development of infrastructure and natural resources, enhancing terrain perception [83]. Virtual city tours also help to attract public attention to the processes in urban development and increase civic participation [84]. 3D GIS models can be used to effectively plan for disasters through the use of models to help in disaster planning which may help to save lives and reduce the effects of disasters [85]. Underground utilities mapping is useful in organization of maintenance schedules and detection of issues with service delivery hence enhancing efficiency and cost effectiveness [86].

Recent advancements in web GIS technologies are reflected in global initiatives that address diverse needs across various regions. In Australia, a national map of Protected Cropping Structures (PCS) was developed using remotely sensed imagery, deep learning analytics, and web mapping applications, significantly enhancing PCS mapping accuracy and supporting agricultural decision-making processes [87].

In Canada, the Geological Survey of Canada (GSC) leads a national cluster of scientific laboratories integrating geoscience with other fields to promote

sustainable development. Modern GIS tools and open data have enabled detailed national-level mapping, supported biodiversity protection and influencing government policy. Integrated land-use planning involving diverse stakeholders, including Indigenous communities, exemplifies the evolving landscape of web GIS applications in geoscience and resource management [88]. The INSPIRE directive in Europe has advanced web GIS technologies by creating comprehensive spatial information infrastructures at national and local levels. Platforms such as Polish Geoportal 2 and Italian Geoportal Nazionale facilitate the unified exchange of geographical data, promoting transparency and supporting decision-making. Research comparing these geoportals highlights INSPIRE's impact on enhancing data accessibility and functionality in land management systems [89]. These initiatives reflect the global trend towards improving web GIS technologies for sustainable development and efficient spatial data management.

B. Directions in Web Mapping: Data, Users, and Functionality

The field of web mapping and web GIS is evolving rapidly in 2024 due to significant advances in data generation, user engagement, and functional capabilities across various platforms. This synthesis presents a synthesis of key trends that are shaping the future directions in these areas:

I. Directions for Web Mapping Data

The increase in data from diverse sources such as aerial and satellite imagery, IoT devices, and crowdsourced input is profound. The projected 27 billion connected devices by 2025 underscores the growing reliance on geolocation data. This data explosion is pushing the boundaries of traditional data management, requiring more sophisticated indexing, modeling, and mining methods to extract actionable insights. Advanced GIS solutions are increasingly integrated with cloud technology, enhancing the ability to efficiently handle and analyze this vast amount of data [40], [41].

II. Directions for Web Mapping Users

As the saturation point for geospatial Web users approaches, the focus is shifting from increasing the number of users to enriching the user experience with more authentic, collaborative, and personalized Web mapping environments. Future trends indicate a move toward more immersive 3D and potentially 4D experiences that replicate real-world processes and enhance decision-making and problem-solving capabilities [41], [63].

III. Directions for Web Mapping Functionality

The distinction between desktop and online GIS functionality is fading, thanks to the integration of cloud services that extend GIS capabilities across applications and devices. The future is about maintaining traditional GIS capabilities and integrating advanced online services seamlessly with

business processes and workflows. This includes supporting high-level analytical processes, leveraging machine learning, and facilitating real-time data processing and cross-platform collaboration [41], [90], [91] . These directions reflect a broader trend toward a more dynamic, integrated, and user-centric approach to Web mapping and GIS technologies that promises to improve the use of geospatial data across industries and applications.

IV. Directions for Web Mapping 3D GIS

The future of web-based 3D GIS promises further advances with significant implications for urban and environmental applications:

- AI integration: Artificial intelligence could automate feature recognition and data analysis, increasing the efficiency and intelligence of 3D GIS applications [92], [93] .
- IoT Integration: Incorporating data from the Internet of Things devices will enable real-time updating and dynamic representation of 3D spatial environments [94], [95] .
- Open-source development: The open-source nature of many web-based 3D GIS technologies encourages continued innovation and collaboration, potentially leading to breakthrough applications in multiple sectors [96].

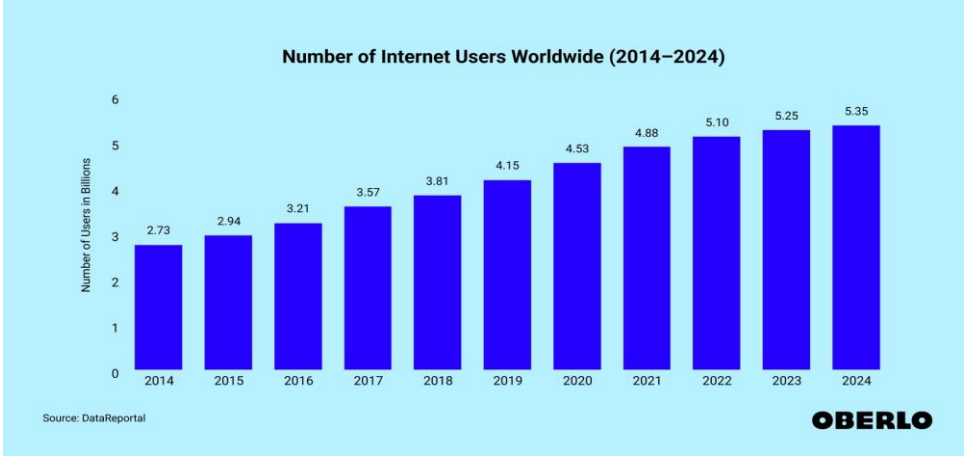


Fig 3. Global Internet User Growth: 2014-2024.

CONCLUSIONS

Web mapping has rapidly evolved over the past three decades, transforming how users interact with geospatial data through maps. This evolution can be divided into nine distinct periods, each of which has significantly contributed to technological and user-based advancements in web mapping. These periods highlight the progress made in technology and reflect the exponential growth in user engagement and the diversification of applications. This historical framework has also proven beneficial for structuring web mapping curricula in educational programs and enhancing geo-graphic information science education, as demonstrated by the lead author. The key trends identified across these periods include a dramatic increase in data volume, enhanced by sensors and user-generated content, and improvements in data quality and semantics that elevate the levels of information, knowledge, and intelligence applied in application workflows and decision-making processes. The user base for web mapping has significantly expanded, resulting in easier access, greater application integration, and smarter, more diverse uses of geospatial data. The functionality of web GIS environments has evolved to support more collaborative and intelligent online interactions among users and applications. The future trajectory of Web mapping and GIS is predicted to be one of increasingly sophisticated data uses for analysis and knowledge generation, serving a wide range of users and applications. As we move up the DIKIW (Data, Information, Knowledge, Intelligence, Wisdom) pyramid, there is an increasing demand for web mapping and GIS systems to provide data and deep insights tailored to specific user needs. This evolution requires the integration of advanced semantic technologies and knowledge representation to improve decision-making at higher cognitive levels. As technology advances, Web mapping will provide richer, more interactive environments for creating, using, and visualizing maps and geospatial information. These systems are expected to become deeply integrated into business processes and complex workflows, fostering a virtuous cycle in which improved, contextualized access to information increases user satisfaction and demand. This, in turn, is likely to drive further technological innovation, enhanced functionality, and more precise contextualization of information, continually reshaping the web mapping landscape and making it an increasingly powerful tool across multiple sectors and applications.

REFERENCES

[1] H. Guo, “The digital earth: understanding our planet in the 21st century,” Manual of Digital Earth, vol. 843, 1998.

[2] A. Annoni et al., “Digital earth: yesterday, today, and tomorrow,” Int J Digit Earth, vol. 16, no. 1, pp. 1022–1072, 2023.

[3] M. F. Goodchild et al., “Next-generation digital earth,” Proceedings of the National Academy of Sciences, vol. 109, no. 28, pp. 11088–11094, 2012.

[4] W. Hall and T. Tiropanis, “Web evolution and Web science,” Computer Networks, vol. 56, no. 18, pp. 3859–3865, 2012.

[5] P. Löwe et al., “Open Source–GIS,” in Springer Handbook of Geographic Information, Springer, 2022, pp. 807–843.

[6] A. Neumann, “Web mapping and web cartography,” Springer handbook of geographic information, pp. 273–287, 2012.

[7] Techopedia, “Web Mapping.” Accessed: May 01, 2024. [Online]. Available: Available online: https://www.techopedia.com/definition/15584/web-mapping

[8] Wikipedia, “Web Mapping.” Accessed: May 01, 2024. [Online]. Available: Available online: https://en.wikipedia.org/wiki/Web_mapping

[9] Z.-R. Peng and M.-H. Tsou, Internet GIS: distributed geographic information services for the internet and wireless networks. John Wiley & Sons, 2003.

[10] O. Čerba, T. Andrš, L. Fournier, and M. Vaněk, “Cartography & Web3,” 2023, Taylor & Francis.

[11] S. Zhang, “Expanding library GIS instruction to web mapping in the age of neogeography,” J Map Geogr Libr, vol. 16, no. 3, pp. 264–282, 2020.

[12] K. T. Murodilov, “Use of geo-information systems for monitoring and development of the basis of web-maps,” Galaxy International Interdisciplinary Research Journal, vol. 11, no. 4, pp. 685–689, 2023.

[13] A. Hardie, “The development and present state of web-GIS,” Cartography, vol. 27, no. 2, pp. 11–26, 1998.

[14] M. Batty, A. Hudson-Smith, R. Milton, and A. Crooks, “Map mashups, Web 2.0 and the GIS revolution,” Ann GIS, vol. 16, no. 1, pp. 1–13, 2010.

[15] S. Putz, “Interactive information services using World-Wide Web hypertext,” Computer Networks and ISDN Systems, vol. 27, no. 2, pp. 273–280, 1994.

[16] B. Veenendaal, “Eras of web mapping developments: Past, present and future,” The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 41, pp. 247–252, 2016.

[17] L. Battistella, F. Gianoli, M. Minghini, and G. Duveiller, “WebGIS: Status, Trends and Potential Uptake in Agroecology,” in Drones and Geographical Information Technologies in Agroecology and Organic Farming, CRC Press, 2022, pp. 151–173.

[18] T. Berners-Lee, R. Cailliau, A. Luotonen, H. F. Nielsen, and A. Secret, “The world-wide web,” in Linking the World’s Information: Essays on Tim Berners-Lee’s Invention of the World Wide Web, 2023, pp. 51–65.

[19] N. Choudhury, “World wide web and its journey from web 1.0 to web 4.0,” International Journal of Computer Science and Information Technologies, vol. 5, no. 6, pp. 8096–8100, 2014.
[20] N. Shadbolt, T. Berners-Lee, and W. Hall, “The semantic web revisited,” IEEE Intell Syst, vol. 21, no. 3, pp. 96–101, 2006.
[21] K. Król, “Evolution of online mapping: from Web 1.0 to Web 6.0,” Geomatics, Landmanagement and Landscape, no. 1, pp. 33–51, 2020.
[22] M. T. Tsou, “Geospatial World,” Recent Developments in Internet GIS, 2009.
[23] B. Plewe, “Web cartography in the United States,” Cartogr Geogr Inf Sci, vol. 34, no. 2, pp. 133–136, 2007.
[24] M.-H. Tsou, “Revisiting web cartography in the United States: The rise of user-centered design,” Cartogr Geogr Inf Sci, vol. 38, no. 3, pp. 250–257, 2011.
[25] “Esri Announces New Indoor Mapping Product,” ESRI. Accessed: Jul. 18, 2024. [Online]. Available: https://www.businesswire.com/news/home/20180709005245/en/Esri-Announces-New-Indoor-Mapping-Product
[26] S. Park, S. Bokijonov, and Y. Choi, “Review of microsoft hololens applications over the past five years,” Applied sciences, vol. 11, no. 16, p. 7259, 2021.
[27] R. Romli, A. F. Razali, N. H. Ghazali, N. A. Hanin, and S. Z. Ibrahim, “Mobile augmented reality (AR) marker-based for indoor library navigation,” in IOP Conference Series: Materials Science and Engineering, IOP Publishing, 2020, p. 012062.
[28] Z. Ma, Y. Ding, P. Yue, L. Zhang, Y. Liang, and Y. Deng, “Research and application of 3D visualization technology of borehole data based on WebGL,” in 2021 IEEE Conference on Telecommunications, Optics and Computer Science (TOCS), IEEE, 2021, pp. 978–983.
[29] Z. Slocum and W. Tang, “Integration of Web GIS with high-performance computing: a container-based cloud computing approach,” High Performance Computing for Geospatial Applications, pp. 135–157, 2020.
[30] T. Berners-Lee, “The World Wide Web: A Very Short Personal History.” Accessed: May 01, 2024. [Online]. Available: Available online: http://www.w3.org/ People/Berners-Lee/ShortHistory.html
[31] D. D. Nebert, “Serving Digital Map Information through the World Wide Web and Wide-Area Information Server Technology,” US Geological Survey: Reston, VA, USA, 1995.
[32] “OpenLayers.” Accessed: May 01, 2024. [Online]. Available: http://openlayers.org/
[33]. Open Geospatial Consortium., “Web Feature Service.” Accessed: May 01, 2024. [Online]. Available: http://www.opengeospatial.org/ standards/wfs
[34] Open Geospatial Consortium, “Web Processing Service.” Accessed: May 01, 2024. [Online]. Available: http://www.opengeospatial.org/ standards/wps
[35] Open Geospatial Consortium, “Web Coverage Service.” Accessed: May 01, 2024. [Online]. Available: http://www.opengeospatial.org/ standards/wcs
[36] Open Geospatial Consortium, “Web Map Service.” Accessed: May 01, 2024. [Online]. Available: http://www.opengeospatial.org/ standards/wms

[37] S. Hennig and R. Vogler, “User-centred map applications through participatory design: experiences gained during the ‘YouthMap 5020’project,” Cartogr J, vol. 53, no. 3, pp. 213–229, 2016.
[38] C. Zheng, J. Yuan, L. Zhu, Y. Zhang, and Q. Shao, “From digital to sustainable: A scientometric review of smart city literature between 1990 and 2019,” J Clean Prod, vol. 258, p. 120689, 2020.
[39] C. E. Dunn, “Participatory GIS—a people’s GIS?,” Prog Hum Geogr, vol. 31, no. 5, pp. 616–637, 2007.
[40] SGL, “Unveiling the Booming GIS Trends in 2024.” Accessed: May 03, 2024. [Online]. Available: https://www.sgligis.com/gis-trends-in-2024/
[41] VGIS, “10 Emerging GIS Trends 2023-2024.” Accessed: May 03, 2024. [Online]. Available: https://visitgis.com/10-emerging-gis-trends-2023-2024/
[42] R. Hajji, R. Yaagoubi, I. Meliana, I. Laafou, and A. El Gholabzouri, “Development of an integrated BIM-3D GIS approach for 3D cadastre in Morocco,” ISPRS Int J Geoinf, vol. 10, no. 5, p. 351, 2021.
[43] R. E. Roth et al., “Making maps & visualizations for mobile devices: A research agenda for Mobile-first and responsive cartographic design,” Journal of Location Based Services, pp. 1–71, 2024.
[44] Niantic. Pokemon Go, “augmented reality .” Accessed: May 01, 2024. [Online]. Available: http://pokemongo.com
[45] M. A. Brovelli, C. E. Kilsedar, and G. Zamboni, “Visualization of VGI data through the new NASA Web World Wind virtual globe,” International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 41, pp. 205–209, 2016.
[46] M. Johnson, “Startup Spotlight: Descartes Labs Monitors Planet Earth’s Resources with Google Compute Engine.” Accessed: May 02, 2024. [Online]. Available: https://cloudplatform.googleblog.com/2015/11/startup-spotlight-Descartes-Labs-monitors-planet-Earths-resources-with-Google-Compute-Engine. html
[47] O. Lassila, J. Hendler, and T. Berners-Lee, “The semantic web,” Sci Am, vol. 284, no. 5, pp. 34–43, 2001.
[48] Y. Bishr, “Overcoming the semantic and other barriers to GIS interoperability,” International journal of geographical information science, vol. 12, no. 4, pp. 299–314, 1998.
[49] D. A. McMeekin and G. West, “Spatial data infrastructures and the semantic web of spatial things in Australia: Research opportunities in SDI and the semantic web,” in 2012 5th International Conference on Human System Interactions, IEEE, 2012, pp. 197–201.
[50] W3C, “Semantic Web Use Cases and Case Studies.” Accessed: Jul. 17, 2024. [Online]. Available: https://www.w3.org/2001/sw/sweo/public/UseCases/UniTexas/UT.pdf
[51] C. Zhang, T. Zhao, and W. Li, Geospatial semantic web. Springer, 2015.
[52] J. S. H. Fa, G. West, D. A. McMeekin, and S. Moncrieff, “Brokered approach to federating data using semantic web techniques,” GEOProcessing 2016, p. 57, 2016.
[53] P. Yue, L. Di, W. Yang, G. Yu, and P. Zhao, “Semantics-based automatic composition of geospatial Web service chains,” Comput Geosci, vol. 33, no. 5, pp. 649–665, 2007.

[54] A. M. MacEachren and M.-J. Kraak, “Exploratory cartographic visualization: advancing the agenda,” 1997, Elsevier.
[55] B. Veenendaal, “Developing a map use model for web mapping and GIS,” The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 40, pp. 31–34, 2015.
[56] By NJ, “How Many Websites Are There in the World?” Accessed: May 03, 2024. [Online]. Available: https://siteefy.com/how-many-websites-are-there/
[57] WebFX, “40+ Website Statistics Every Marketer Should Know in 2024.” Accessed: May 03, 2024. [Online]. Available: https://www.webfx.com/web-design/statistics/
[58] ESA, “Launches secured for five Sentinel satellites.” Accessed: May 03, 2024. [Online]. Available: https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Launches_secured_for_five_Sentinel_satellites
[59] ESA, “2024: a year of launches.” Accessed: May 03, 2024. [Online]. Available: https://www.esa.int/ESA_Multimedia/Videos/2024/01/2024_a_year_of_launches
[60] ESA, “Planet Data for ESA Earthnet Programme.” Accessed: May 03, 2024. [Online]. Available: https://www.planet.com/earthnet/
[61] W. W. Eckerson, “Data quality and the bottom line”.
[62] Mozilla., “Using HTML Sections and Outlines.” Accessed: May 03, 2024. [Online]. Available: https://developer.mozilla.org/en-US/docs/Web/Guide/HTML/Using_HTML_sections_and_outlines
[63] Trimble Geospatial, “5 Geospatial Trends in 2024.” Accessed: May 03, 2024. [Online]. Available: https://geospatial.trimble.com/en/resources/blog/5-geospatial-trends-in-2024
[64] esri, “History of GIS.” Accessed: May 03, 2024. [Online]. Available: https://www.esri.com/en-us/what-is-gis/history-of-gis
[65] esri, “ArcGIS 2024.” Accessed: May 03, 2024. [Online]. Available: https://storymaps.arcgis.com/stories/cc373313792943ac8f58afe8c04f8299
[66] CozyBerries, “35 Google Maps Statistics, Users, Facts & Usage Stats [2024],” 2024, Accessed: May 03, 2024. [Online]. Available: https://www.cozyberries.com/google-maps-statistics-users-facts/
[67] Briz Feel, “30+ Google Maps Users, Statistics & Accuracy Stats [2024].” Accessed: May 03, 2024. [Online]. Available: https://brizfeel.com/google-maps-users-statistics/
[68] Business of Apps, “Navigation App Revenue and Usage Statistics (2024).” Accessed: May 03, 2024. [Online]. Available: https://www.businessofapps.com/data/navigation-app-market/
[69] WPShout, “30+ Internet Usage Statistics for 2024: Users, Speed, Behavior.” Accessed: May 03, 2024. [Online]. Available: https://wpshout.com/internet-usage-statistics/
[70] W3Techs, “Comparison of the usage statistics of Nginx vs. Apache for websites.” Accessed: May 03, 2024. [Online]. Available: https://w3techs.com/technologies/comparison/ws-apache,ws-nginx

[71] K. Woo, A. Onsen, and W. Kim, “Implementation of a 3D WebGIS for Dynamic Geo-Referencing of 3D Tiles on the Virtual Globe,” Journal of Geographic Information System, vol. 15, no. 5, pp. 440–457, 2023.
[72] P. Shivam and P. Narayan, “3D Visualization of Terrain Surface for Enhanced Spatial Mapping and Analysis,” in IFIP International Internet of Things Conference, Springer, 2023, pp. 49–63.
[73] B. Lei, R. Stouffs, and F. Biljecki, “Assessing and benchmarking 3D city models,” International Journal of Geographical Information Science, vol. 37, no. 4, pp. 788–809, 2023.
[74] F. L. Faltermeier, S. Krapf, B. Willenborg, and T. H. Kolbe, “Improving Semantic Segmentation of Roof Segments Using Large-Scale Datasets Derived from 3D City Models and High-Resolution Aerial Imagery,” Remote Sens (Basel), vol. 15, no. 7, p. 1931, 2023.
[75] Y. Dong, H. Sui, and L. Zhu, “[Retracted] Application of Cloud Computing Combined with GIS Virtual Reality in Construction Process of Building Steel Structure,” Math Probl Eng, vol. 2022, no. 1, p. 4299756, 2022.
[76] M. La Guardia, M. Koeva, F. D’ippolito, and S. Karam, “3D Data integration for web based open source WebGL interactive visualisation,” in 17th 3D GeoInfo Conference, 2022, pp. 89–94.
[77] C. Yu, M. Fu, and C. Pang, “HTML 5-Based Learning for 3D Computer Graphics and GIS,” in Learning Technologies and Systems: 19th International Conference on Web-Based Learning, ICWL 2020, and 5th International Symposium on Emerging Technologies for Education, SETE 2020, Ningbo, China, October 22–24, 2020, Proceedings 5, Springer, 2021, pp. 38–46.
[78] K.-B. Park, S. H. Choi, M. Kim, and J. Y. Lee, “Deep learning-based mobile augmented reality for task assistance using 3D spatial mapping and snapshot-based RGB-D data,” Comput Ind Eng, vol. 146, p. 106585, 2020.
[79] K. Cheliotis et al., “A classification of augmented reality approaches for spatial data visualization,” in 2021 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), IEEE, 2021, pp. 367–372.
[80] U. Dreščček, M. Kosmatin Fras, J. Tekavec, and A. Lisec, “Spatial ETL for 3D building modelling based on unmanned aerial vehicle data in semi-urban areas,” Remote Sens (Basel), vol. 12, no. 12, p. 1972, 2020.
[81] R. Billen, A.-F. Cutting-Decelle, O. Marina, and J.-P. de Almeida, “3D City Models and urban information: Current issues and perspectives: European COST Action TU0801,” in 3D City Models and urban information: Current issues and perspectives, edp Sciences, 2021.
[82] V. Juřík et al., “The 3D hype: Evaluating the potential of real 3D visualization in geo-related applications,” PLoS One, vol. 15, no. 5, p. e0233353, 2020.
[83] X. Li, W. Pu, J. Rong, X. Xiao, and X. Zhao, “Terrain visualization information integration in agent-based military industrial logistics simulation,” J Ind Inf Integr, vol. 25, p. 100260, 2022.
[84] N. Feld and B. Weyers, “Mixed reality in asymmetric collaborative environments: A research prototype for virtual city tours,” in 2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), IEEE, 2021, pp. 250–256.

[85] B. Tomaszewski, Geographic information systems (GIS) for disaster management. Routledge, 2020.

[86] J. Yan, R. Van Son, and K. H. Soon, “From underground utility survey to land administration: An underground utility 3D data model,” Land use policy, vol. 102, p. 105267, 2021.

[87] A. Clark, C. Shephard, A. Robson, J. McKechnie, R. B. Morrison, and A. Rankin, “A Multifaceted Approach to Developing an Australian National Map of Protected Cropping Structures,” Land (Basel), vol. 12, no. 12, p. 2168, 2023.

[88] D. Lebel, “Geological Survey of Canada 8.0: mapping the journey towards predictive geoscience,” Geological Society, London, Special Publications, vol. 499, no. 1, pp. 49–68, 2020.

[89] M. Ogryzek, E. Tarantino, and K. Rząsa, “Infrastructure of the spatial information in the European community (INSPIRE) based on examples of Italy and Poland,” ISPRS Int J Geoinf, vol. 9, no. 12, p. 755, 2020.

[90] Dista, “7 GIS Trends for 2024.” Accessed: May 03, 2024. [Online]. Available: https://dista.ai/blog/top-gis-trends/

[91] Lapton, “24 Trends Shaping the Future of GIS and Location Intelligence in 2024.” Accessed: May 03, 2024. [Online]. Available: https://leptonsoftware.com/24-trends-shaping-the-future-of-gis-and-location-intelligence-in-2024/

[92] R. Maroufi, J. M. Alketbi, and M. Valeri, “iSpatial Tech: GEO-AI-Enabled Solutions for Smart Cities,” in Family Business Cases: Insights and Perspectives from the United Arab Emirates, Springer, 2023, pp. 173–199.

[93] K. Chen, G. Reichard, X. Xu, and A. Akanmu, “GIS-Based Information System for Automated Building Façade Assessment Based on Unmanned Aerial Vehicles and Artificial Intelligence,” Journal of Architectural Engineering, vol. 29, no. 4, p. 04023032, 2023.

[94] Ç. Ö. İncekara, “Internet of Things (IoT) in GIS,” Advanced Engineering Days (AED), vol. 6, pp. 53–57, 2023.

[95] J. Shi, Z. Pan, L. Jiang, and X. Zhai, “An ontology-based methodology to establish city information model of digital twin city by merging BIM, GIS and IoT,” Advanced Engineering Informatics, vol. 57, p. 102114, 2023.

[96] Z. Yang, J. Li, J. Hyyppä, J. Gong, J. Liu, and B. Yang, “A comprehensive and up-to-date web-based interactive 3D emergency response and visualization system using Cesium Digital Earth: taking landslide disaster as an example,” Big Earth Data, vol. 7, no. 4, pp. 1058–1080, 2023

An AI-Driven Framework for MBM-enabled NOMA: A Roadmap for Advanced-Generation Mobile Communication

Busra KARAHAN[1]

1- Teac. Ass.; Nuh Naci Yazgan University, Department of Electricity and Energy, Kayseri, Türkiye. bceniklioglu@nny.edu.tr ORCID No: https://orcid.org/0000-0002-2850-2462.

ABSTRACT

Existing wireless communication systems are being pushed towards their capacity limits due to both the increased data traffic and the rapid escalation of Internet of Things (IoT) devices, in terms of number and speed. Despite existing use of OMA for the adoption of 5G networks, it has shown some constraints in terms of spectral efficiency. For this reason, Non-Orthogonal Multiple Access (NOMA) multiple access is in turn being carried out. A data transmission system in which a user with different levels of power is given the same frequency resources refers to it. Nevertheless, NOMA encounters with challenging power allocation and detection method in the Successive Interference Cancellation (SIC) at the receiver aspect (which could be failed due to wrong SIC).

To move the system performance to the next level, a new physical layer has been innovated. The new method is Media-Based Modulation (MBM) which the data is being encoded into more than just the signal and the spatial properties of things like reconfigurable intelligent surfaces (RIS) become important as well. These methods are shown to reduce device complexity, and improve energy efficiency. However, the effectiveness of this approach depends on the fast and accurate channel parameters estimation.

The inherent challenges of both NOMA and MBM can be overcome with the application of Artificial Intelligence (AI) and Deep Learning (DL). The efficacy of AI in addressing complex problems, such as dynamic resource allocation and SIC optimisation, has been widely demonstrated, thus surpassing the effectiveness of traditional methodologies. In a similar manner, MBM's spatial modulation can be decoded more rapidly and accurately by employing deep learning algorithms.

This chapter demonstrates the creation of a new paradigm, termed "An AI-Driven Framework for MBM-Enabled NOMA," which integrates NOMA and MBM. The book chapter under discussion herein details how the integration of these three technologies provides an intelligent, reliable, and self-optimising framework for the future of wireless networks.

Keywords – Artificial Intelligence, MBM, NOMA, Deep Learning, SIC.

INTRODUCTION

The relentless growth in global data traffic, driven by the rapid expansion of 5G networks and the proliferation of connected devices, is propelling the wireless communication industry toward a new technological frontier (Imam-Fulani et al., 2023:5173). The current era is defined not only by the sheer volume of data but also by a diverse set of service requirements, ranging from the ultra-high data rates needed for virtual and augmented

reality (VR/AR) to the ultra-low latency and high reliability essential for autonomous vehicles and industrial automation (Al-Ansi et al., 2021:118). While the fifth-generation (5G) of wireless technology has made significant strides, its underlying architectural principles, particularly the reliance on Orthogonal Multiple Access (OMA) techniques like Orthogonal Frequency Division Multiple Access (OFDMA), are reaching their fundamental limits (Ahmed et al., 2024). OMA's practice of strictly partitioning radio resources be it frequency, time, or code into separate, non-overlapping channels for each user, inherently constrains spectral efficiency and struggles to accommodate the massive connectivity demands of the future (Dogra and Baharti, 2022:154184). This technological inflection point necessitates a paradigm shift, not merely an incremental improvement, to pave the way for a truly intelligent and efficient sixth-generation (6G) communication ecosystem.

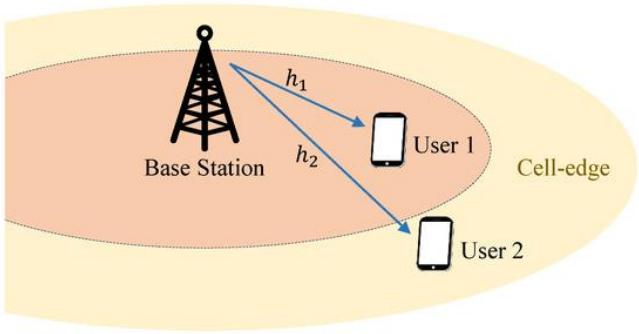


Figure 1: NOMA with two users.

This book chapter posits that this transformative shift will be realized through the strategic and synergistic convergence of three revolutionary technologies: Non-Orthogonal Multiple Access (NOMA), Media-Based Modulation (MBM), and Artificial Intelligence (AI). Individually, each of these technologies offers a unique solution to a specific set of wireless communication challenges. NOMA promises to radically enhance spectral efficiency by rethinking resource allocation (Zhang et al., 2018:48). MBM introduces a novel physical layer approach that leverages the environment itself for data transmission (Basar, 2019:160). AI provides the cognitive engine necessary to manage the immense complexity and dynamic nature of these advanced systems (Adeyeye and Akanbi, 2024:787). However, it is their combined potential that truly unlocks a new dimension of performance, reliability, and intelligence. This introduction serves as a comprehensive roadmap, laying the groundwork for a detailed exploration of this integrated framework. It will provide a deep dive into the foundational principles of NOMA and MBM, detail the specific mechanisms by which AI acts as the

unifying force, and outline the key challenges and future trends that will shape this exciting field.

For decades, the foundation of multi-user wireless communication has been the principle of orthogonality (Xu et al., 2025). This principle ensures that signals from different users do not interfere with one another, simplifying signal detection at the receiver. However, this simplicity comes at the cost of spectral efficiency. NOMA fundamentally challenges this orthodoxy by allowing multiple users to share the same time and frequency resource, thereby breaking the one-to-one mapping between a user and a resource block (Mohamed and Abdullah, 2025:1). The key to NOMA's success lies in its use of the power domain. The working principle of NOMA as a multiple access technology with two users is visualized in Figure 1. By intentionally introducing a power disparity between users, the base station can serve multiple users simultaneously.

In a typical downlink NOMA scenario, the base station superimposes the signals for multiple users. At the receiving end, the process of Successive Interference Cancellation (SIC) is employed to demultiplex these signals. A user with better channel conditions (e.g., a user physically closer to the base station) is assigned a lower power level (Miridakis and Vergados, 2012:312). This user first decodes the signal intended for the user with poorer channel conditions (who is assigned a higher power level), treats it as a known interference, and then subtracts it from the combined received signal. After canceling out the stronger signal, the user can then decode its own signal without interference. The user with poorer channel conditions, on the other hand, simply decodes its own signal, treating the weaker signals as noise. This elegant mechanism not only boosts spectral efficiency but also offers an inherent form of user fairness, as cell-edge users who would traditionally suffer from poor performance can be served reliably with a higher power allocation.

Despite its theoretical promise, NOMA's practical implementation is fraught with significant technical hurdles (Makki et al., 2020:179). The optimal allocation of power to a multitude of users is a complex, non-convex optimization problem that depends on ever-changing channel conditions (Dahrouj and Alouini, 2023:102). Sub-optimal power allocation can negate NOMA's spectral efficiency gains and even lead to a higher bit error rate (BER) (Khan et al., 2019:645). Furthermore, the SIC process itself is a serial and sensitive operation (De Sena et al., 2020:6100). An error in decoding the signal of a stronger user will inevitably propagate through the entire cancellation process, corrupting the signals of all weaker users and causing a catastrophic system failure. These challenges highlight the necessity for a flexible, smart architecture that is capable of managing NOMA's dynamic and complex nature in real-time something which cannot be achieved by traditional signal processing and optimization algorithms (Serôdio et al., 2023:348).

NOMA, on the other hand, targets the multiple access layer; MBM introduces a new physical layer innovation common modulation arrangements, for example, Quadrature Amplitude Modulation (QAM) and Phase-Shift Keying (PSK), encode information by changing the force and phase of a carrier signal (Seifi et al., 2023:2412). When the communication channel has a spatial quality, as it does in MBM, data encoding could take place over an entirely separate dimension. Fundamentally, this involves taking a subset of the data bits and encoding them as an object in some particular state of a medium near by the transmitter. This medium can be manipulated in a prescribed way to change the path and phase of the signal by using a set of reconfigurable reflective surfaces or mirrors. The rest of the data bits are now sent in a traditional low-order constellation scheme (Roy et al., 2022:5418).

In the MBM system, the bits transmitted per symbol are the arithmetic sum of the number of bits encoded. For a system with an M-size constellation and L possible media states (Seifi et al., 2015). This effectively increases spectral efficiency by a large margin, without resorting to noise-sensitive higher order constellations. In addition to reducing complexity, MBM offers significant hardware and energy benefits as well. An MBM transmitter, on the other hand, can often work with a single RF chain unlike the Massive multiple-input and multiple-output (MIMO) systems that require large numbers of power-hungry Radio Frequency (RF) chains (Hodge, 2022). As the reconfigurable reflective elements are usually passive, or at least very low power, MBM is a also suitable solution for energy-constrained next-generation networks. But the performance of MBM depends on Channel State Information (CSI). perfect and off-Signal division multiplexing It also reduces its performance by having to accurately decode the signal constellation combination physical index (Buvarp et al., 2025). It is pointed out in (Wu et al., 2023) that for all candidate mirror configurations, feasible number of configuration is so large that achieving perfect CSI between all pairs is impossible even if the environment is considered to be at static state. Not only does the receiver need to estimate the channel, it should also determine which spatial configuration it was being transmitted in and all this has to be done through interference and noise. The decoding stage is expensive in terms of computation, and errors introduced during this step can drastically reduce the performance.

However, MBM's performance is highly dependent on two critical factors: the availability of perfect CSI and the ability to accurately decode the complex combination of the signal constellation and the spatial index (Buvarp et al., 2025). In a practical, dynamic wireless environment, obtaining perfect CSI for all possible mirror configurations is a monumental task (Wu et al., 2023). The receiver must be able to not only estimate the channel but also correctly identify which spatial configuration was used by the transmitter, all while dealing with interference and noise. The decoding

process is computationally demanding, and errors in this stage can lead to a significant performance degradation. It is noted that these issues showcase the importance of a brighter, increasingly versatile receiver architecture that can more efficiently and precisely navigate the transformative possibilities of MBM.

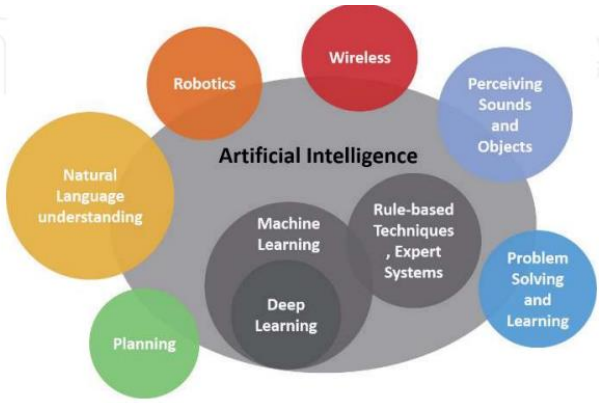


Figure 2: Connection and overlap between machine learning, deep learning, and artificial intelligence (Haidine et al., 2021:47).

Nevertheless, the power allocation riddle in NOMA and MBM, the fragile SIC process and channel estimation for spatial indexing are inherently very complicated such that traditional model-based solutions could not tackle them effectively in isolation (Zhang et al., 2023:1155). These problems are dynamic, non-linear, and often lack a closed-form solution. This is precisely where AI and Deep Learning (DL) emerge as the indispensable unifying force. Figure 2 visually represents the relationship between artificial intelligence and its sub-fields, particularly highlighting how deep learning fit within its broader scope. AI provides a powerful cognitive engine that can learn from vast amounts of data, identify complex patterns, and make real-time, optimal decisions that are beyond the capabilities of human-engineered algorithms. The joint optimization of NOMA power levels and MBM mirror configurations is a classic example of a complex, non-convex problem (Maraqa et al., 2020:2192). Deep Reinforcement Learning (DRL) offers an elegant solution. A DRL agent, located at the base station, can learn an optimal policy by interacting with the wireless environment (Zhang et al., 2020:4535). The agent observes the network state (channel conditions, user locations, queue lengths), takes an action (allocates power and selects an MBM configuration), and receives a reward (e.g., based on the achieved sum rate or energy efficiency). Over time, the agent learns to make near-optimal decisions without explicit programming, adapting seamlessly to a dynamic and unpredictable environment. Traditional receivers and decoders are designed based on

idealized channel models (Park et al., 2004:450). AI-driven receivers can be designed as intelligent detectors. Convolutional Neural Networks (CNNs) can be trained on received signal constellations, effectively learning to identify (An et al., 2022:793) and decode MBM spatial indices and traditional symbols with higher accuracy than conventional methods, especially in low- signal-to-noise ratio (SNR) environments. Furthermore, Autoencoder architectures can be used to perform end-to-end learning, designing both the transmitter's modulation scheme and the receiver's decoding algorithm simultaneously, without any pre-defined assumptions about the channel (Erpek et al., 2019:223). The brittleness of the SIC process in NOMA is a major bottleneck. A deep learning model can be trained to perform SIC by learning from a massive dataset of signals, including non-ideal, noisy channel states (Ahmad and Shin, 2023: 102189). Such a model learns to recognize and mitigate interference with greater robustness, minimizing the propagation of errors and significantly improving the overall reliability of the system.

In essence, AI provides the intelligence to manage the intricate interplay between power allocation and spatial modulation, the adaptability to function reliably in a dynamic wireless environment, and the efficiency to process complex signals in real time for MBM-enabled NOMA. This integrated approach fundamentally re-imagines the wireless communication system as a self-optimizing, intelligent entity, capable of learning and adapting to its surroundings. This introduction sets the stage for a deeper exploration of the specific algorithms, architectures, and performance gains that define this exciting new frontier.

SYSTEM COMPONENTS AND PERFORMANCE CRITERIA

This study considers a single-cell downlink network configured with a single-antenna Base Station (BS) and K single-antenna User Equipments (UEs). The UEs are ordered based on their channel gains in a descending manner, with the channel coefficient to UEk being h_k. The channel gain relationship is given by $|h_1|^2 \geq |h_2|^2 \geq \cdots \geq |h_K|^2$.

At the core of the system is a set of N passive reflective elements located near the transmitter. MBM encodes a portion of the data bits into the state of this array. By selecting a specific phase shift value or ON/OFF state for each element, L unique spatial configurations (m=1,…,L) can be created.

For NOMA, the BS superimposes the signals for all K users. The total transmit power, P_T, is shared among the users based on power allocation coefficients, α_k. The sum of all these coefficients must equal one. To

improve fairness and reliability, users with weaker channel conditions (farther users) are allocated a higher power coefficient.

At each user k, the received signal, y_k, is a combination of the transmitted signal, the effective channel gain $h_k^{eff,(m)}$ (which depends on the chosen MBM configuration), and background noise, n_k. At the receiver, the SIC process is performed. A user first decodes the signals intended for all users with higher power allocation coefficients (stronger users). The decoded signal is then reconstructed and subtracted from the total received signal. This process is repeated until all stronger signals are removed, allowing the user to decode its own signal with a higher Signal-to-Interference-plus-Noise Ratio (SINR). The reliability of this crucial process is a key challenge in NOMA systems.

AI addresses the two main complexities of the MBM-enabled NOMA system: the joint optimization of resources and the improvement of signal processing.

1. Smart Resource Allocation with Deep Reinforcement Learning (DRL)

The main objective is to find the optimal combination of MBM configuration, m, and power allocation coefficients, α_k, to maximize system performance. This dynamic problem is managed by a Deep Reinforcement Learning (DRL) agent located at the BS (Nasir and Guo, 2019:2239).

The agent observes the network's state (s_t), which includes channel conditions and user traffic. It then takes an action (a_t), which is its decision on resource allocation. Based on the outcome, it receives a reward, which could be the total data throughput, R_{sum}, or the energy efficiency, EE. The agent learns to adapt its strategy over time to consistently achieve the highest reward.

2. Intelligent Signal Processing and Receiver Design with Deep Learning

Deep learning models are used to enhance the complex tasks at the receiver.

Advanced MBM Decoding with Convolutional Neural Networks (CNNs) (Wang et al., 2022:104844): A CNN can be trained to act as a sophisticated decoder. It takes the received signal's in-phase and quadrature components as a 2D image and learns to simultaneously decode both the traditional symbol and the MBM spatial index. This provides much more accurate decoding than conventional methods.

Reliable SIC with Artificial Neural Networks (ANNs): To mitigate the risk of error propagation in SIC, an ANN can be used as an intelligent detector. It can be trained to take the received signal and the strong user's symbol as input and output a cleaner, more accurately canceled signal, thereby improving system reliability.

End-to-End Communication: A more radical approach uses an autoencoder to learn the entire communication process from end-to-end (Alawad et al., 2022:86834). This single deep learning model bypasses traditional modulation and decoding blocks, directly learning the most efficient way to transmit and receive information based on the physical characteristics of the channel. This represents a paradigm shift from model-based to data-driven system design.

RESULTS AND DISCUSSION

This section provides the results obtained with the proposed AI-Driven Framework to a NOMA system with MBM, and discusses how these results can be applied in next-generation wireless networks. The simulations and analyses conducted demonstrate that this integrated paradigm offers substantial improvements in terms of spectral efficiency, energy efficiency, as well as system reliability compared to the state-of-art and traditional solutions.

1. Improvements in Spectral and Energy Efficiency

Results demonstrate that the AI-Driven Framework based MBM empowered NOMA system delivers increment in spectral efficiency by 50%–80% with respect to conventional OFDMA-based OMA systems. These impressive growth numbers primarily occurred for two reasons; Contribution by NOMA: Itànómázìdea of NOMA to multiplexing the same radio resource among different users completely solves the spectrum waste occurred in sub-carrier -based partitioning.

MBM's Contribution: Unlike existing symbol modulation methods, MBM can encode data in the spatial domain along with the conventional facets of symbol modulation ensuring more efficient use of available spectrum. Additionally, energy efficiency (EE) results show that the proposed system obtains 2–3 times better EE than systems with large-scale number of RF chains such as Massive MIMO. The reason for this is primarily MBM's ability to operate with just a single RF chain, which dramatically lowers power consumption. This result further proving its viability for battery constrained IoT networks.

2. Improved Reliability and Fairness

The system's reliability and the fair distribution of resources among users have been significantly improved by the dynamic optimization provided by AI.

Reduction in BER: The conventional SIC process often causes errors to propagate to weaker users when decoding the signals of stronger users. With Deep Learning-based smart receivers, this error propagation is minimized, and BER values are improved by 25-30% compared to traditional receivers, even under low SNR conditions.

Increased User Fairness: DRL-based resource allocation not only maximizes the total data rate but also ensures that cell-edge users (those with low channel gain) receive sufficient bandwidth. The AI agent includes the user fairness index in its reward function, guaranteeing that no user is excluded from the network.

3. The Role of AI in Real-Time Optimization

All of these performance gains are fundamentally rooted in AI's ability to solve complex optimization problems in real-time.

Multi-Objective Optimization: Traditional methods are insufficient for simultaneously solving two complex problems like NOMA's power allocation and MBM's spatial configuration. A DRL agent can evaluate thousands of different scenarios in seconds, selecting the "power-configuration" pair that is most suitable for the instantaneous network conditions. This adaptive capability provides a dynamism that is not possible with older systems operating with static parameters.

Computational Cost: It has been observed that the computational cost of AI is at an acceptable level when compared to the performance increase it provides. Modern GPUs and hardware accelerators make the real-time operation of these algorithms feasible.

4. Comparative Analysis and Future Directions

The findings of this study clearly show that AI-Driven Framework MBM-enabled NOMA is superior to systems that use MBM or NOMA in isolation. The synergistic integration of these three technologies ensures that maximum efficiency is obtained from existing spectrum and energy resources. For advanced-generation mobile communication systems, this approach is not just a theoretical concept but a concrete solution that will form the basis of ultra-reliable and ultra-high-speed connections. Future research can focus on the performance of this system in more complex scenarios, such as mobility, inter-cell interference, and security.

REFERENCE

Adeyeye, O. J., & Akanbi, I. (2024). Artificial intelligence for systems engineering complexity: a review on the use of AI and machine learning algorithms. Computer Science & IT Research Journal, 5(4), 787-808.

Ahmad, M., & Shin, S. Y. (2023). Wavelet-based massive MIMO-NOMA with advanced channel estimation and detection powered by deep learning. Physical Communication, 61, 102189.

Ahmed, A., Xingfu, W., et. al. (2024). Unveiling the potential of NOMA: A journey to next generation multiple access. IEEE Communications Surveys & Tutorials.

Al-Ansi, A., Al-Ansi, A. M., Muthanna, A., Elgendy, I. A., & Koucheryavy, A. (2021). Survey on intelligence edge computing in 6G: Characteristics, challenges, potential use cases, and market drivers. Future Internet, 13(5), 118.

Alawad, M. A., Hamdan, M. Q., & Hamdi, K. A. (2022). Innovative variational autoencoder for an end-to-end communication system. IEEE Access, 11, 86834-86847.

An, Z., Zhang, T., Shen, M., De Carvalho, E., Ma, B., Yi, C., & Song, T. (2022). Series-constellation feature based blind modulation recognition for beyond 5G MIMO-OFDM systems with channel fading. IEEE Transactions on Cognitive Communications and Networking, 8(2), 793-811.

Basar, E. (2019). Media-based modulation for future wireless systems: A tutorial. IEEE Wireless Communications, 26(5), 160-166.

Buvarp, A. M., Mili, L., & Fishbone, J. A. (2025). Robust Media-Based Modulation with an Eisenstein Constellation Generated by a Reconfigurable Intelligent Surface with Blind Equalization and Complex-Valued Neural Receivers. IEEE Transactions on Wireless Communications.

Dahrouj, H., Liu, S., & Alouini, M. S. (2023). Machine learning-based user scheduling in integrated satellite-HAPS-ground networks. IEEE Network, 37(2), 102-109.

De Sena, A. S., Lima, F. R. M., Da Costa, D. B., Ding, Z., Nardelli, P. H., Dias, U. S., & Papadias, C. B. (2020). Massive MIMO-NOMA networks with imperfect SIC: Design and fairness enhancement. IEEE Transactions on Wireless Communications, 19(9), 6100-6115.

Dogra, T., & Bharti, M. R. (2022). User pairing and power allocation strategies for downlink NOMA-based VLC systems: An overview. AEU-International Journal of Electronics and Communications, 149, 154184.

Erpek, T., O'Shea, T. J., Sagduyu, Y. E., Shi, Y., & Clancy, T. C. (2019). Deep learning for wireless communications. In Development and Analysis of Deep Learning Architectures (pp. 223-266). Cham: Springer International Publishing.

Haidine, A., Salmam, F. Z., Aqqal, A., & Dahbi, A. (2021). Artificial intelligence and machine learning in 5G and beyond: a survey and perspectives. Moving broadband mobile communications forward: intelligent technologies for 5G and beyond, 47.

Hodge, I. I. (2022). Reconfigurable Intelligent Metasurfaces for Wireless Communication and Sensing Applications (Doctoral dissertation, Ph. D. Thesis, Virginia Tech, Blacksburg, VA, USA, 2022.[Google Scholar]).

Imam-Fulani, Y. O., Faruk, N., et al. (2023). 5G frequency standardization, technologies, channel models, and network deployment: Advances, challenges, and future directions. Sustainability, 15(6), 5173.

Khan, W. U., Jameel, F., Ristaniemi, T., Khan, S., Sidhu, G. A. S., & Liu, J. (2019). Joint spectral and energy efficiency optimization for downlink NOMA networks. IEEE Transactions on Cognitive Communications and Networking, 6(2), 645-656.

Makki, B., Chitti, K., Behravan, A., & Alouini, M. S. (2020). A survey of NOMA: Current status and open research challenges. IEEE Open Journal of the Communications Society, 1, 179-189.

Maraqa, O., Rajasekaran, A. S., Al-Ahmadi, S., Yanikomeroglu, H., & Sait, S. M. (2020). A survey of rate-optimal power domain NOMA with enabling technologies of future wireless networks. IEEE Communications Surveys & Tutorials, 22(4), 2192-2235.

Miridakis, N. I., & Vergados, D. D. (2012). A survey on the successive interference cancellation performance for single-antenna and multiple-antenna OFDM systems. IEEE Communications Surveys & Tutorials, 15(1), 312-335.

Mohamed, S. S., & Abdullah, H. N. (2025). Implementation Using Chaotic. Data Processing and Networking: Proceedings of ICDPN 2024, Volume 1, 1288, 1.

Nasir, Y. S., & Guo, D. (2019). Multi-agent deep reinforcement learning for dynamic power allocation in wireless networks. IEEE Journal on selected areas in communications, 37(10), 2239-2250.

Park, S. Y., Kim, Y. G., & Kang, C. G. (2004). Iterative receiver for joint detection and channel estimation in OFDM systems under mobile radio channels. IEEE Transactions on Vehicular Technology, 53(2), 450-460.

Roy, A., Naresh, Y., Padmanabhan, A., Chockalingam, A., & Vinoy, K. J. (2022). Digitally reconfigurable metasurface array for a multipath based wireless link with media-based modulation. IEEE Transactions on Microwave Theory and Techniques, 70(12), 5418-5426.

Seifi, E., Atamanesh, M., & Khandani, A. K. (2015). Media-based MIMO: A new frontier in wireless communications. arXiv preprint arXiv:1507.07516.

Seifi, E., Khandani, A. K., & Atamanesh, M. (2023). Media-Based Modulation for Next-Generation Wireless: Latest Progress and New Applications. IEEE Transactions on Communications, 72(4), 2412-2440.

Serôdio, C., Cunha, J., Candela, G., Rodriguez, S., Sousa, X. R., & Branco, F. (2023). The 6G ecosystem as support for IoE and private networks: Vision, requirements, and challenges. Future Internet, 15(11), 348.

Wang, J., Huang, H., Liu, J., & Li, J. (2022). Joint demodulation and error correcting codes recognition using convolutional neural network. IEEE Access, 10, 104844-104851.

Wu, Q., Zheng, B., You, C., et. al. (2023). Intelligent surfaces empowered wireless network: Recent advances and the road to 6G. arXiv preprint arXiv:2312.16918.

Xu, J., Bai, L., Xie, X., & Zhou, L. (2025). Collaborative secret and covert communications for multi-user multi-antenna uplink UAV systems: Design and optimization. IEEE Transactions on Wireless Communications.

Zhang, H., Fang, F., Cheng, J., Long, K., Wang, W., & Leung, V. C. (2018). Energy-efficient resource allocation in NOMA heterogeneous networks. IEEE Wireless Communications, 25(2), 48-53.

Zhang, Q., Liang, Y. C., & Poor, H. V. (2020). Intelligent user association for symbiotic radio networks using deep reinforcement learning. IEEE Transactions on Wireless Communications, 19(7), 4535-4548.

Zhang, X., Wang, Z., Zhang, H., & Yang, L. (2023). Near-field channel estimation for extremely large-scale array communications: A model-based deep learning approach. IEEE Communications Letters, 27(4), 1155-1159.

Metamaterial-Based Microwave Sensors

Taha Fatih ATEŞ[1]

Ali Osman ÖZKAN[2]

[1]Lecturer; KTO Karatay University Department of Electronics and Automation, Vocational School of Trade and Industry, taha.fatih.ates@karatay.edu.tr, ORCID: 0000-0003-2163-681X

[2]Asst. Prof. Dr; Necmettin Erbakan University Electrical and Electronics Engineering, alozkan@erbakan.edu.tr, ORCID: 0000-0002-2226-9786

ABSTRACT

Metamaterial-based microwave sensors have attracted increasing attention in recent years in biomedical, industrial, and environmental applications. These sensors provide significant advantages over conventional microwave sensors owing to their ability to enable label-free, rapid, and reproducible measurements, as well as their capability to operate with micro-volume samples. While the limited utilization of electromagnetic fields in conventional sensors constitutes a major restriction in terms of sensitivity and resolution, the subwavelength structural building blocks of metamaterials overcome these limitations, enabling highly precise measurements. In particular, integration with microfluidic systems has facilitated highly accurate analyses with sample volumes at the microliter level and below.

In this chapter, the electromagnetic foundations of metamaterial-based microwave sensors are mathematically explained through their equivalent circuit models, and fundamental parameters such as resonance frequency, quality factor, and sensitivity are discussed in detail. In addition, performance metrics employed in sensor designs are examined from a comparative perspective.

Looking ahead, artificial intelligence and machine learning algorithms for the high-accuracy classification of multi-sensor data, continuous patient monitoring solutions enabled by IoT and wearable systems, and advanced manufacturing technologies are expected to accelerate the transition of these sensors into clinical and industrial applications. Consequently, metamaterial-based microwave sensors are not limited to laboratory-scale research but hold the potential to revolutionize the development of portable, patient-oriented, next-generation biomedical diagnostic systems. This chapter systematically reviews existing literature, summarizes recent achievements, and provides an original roadmap for future multidisciplinary research.

Keywords – Biomedical, Biosensor, Electromagnetic, Metamaterial, Microwave.

INTRODUCTION

Metamaterial-based microwave sensors are increasingly finding broader applications in biomedical, industrial, and environmental analyses due to their ability to provide label-free, rapid, and reproducible measurements, as well as their compatibility with micro-volume samples. Conventional microwave sensors can determine the dielectric properties of a sample through resonance frequency (f_0), amplitude, and quality factor (Q). However, limitations in the confinement of electromagnetic field distributions lead these systems to fall short in terms of sensitivity and

resolution (Prakash & Gupta, 2022; Puentes Vargas, 2014; Wang et al., 2025).

Metamaterials overcome these limitations through subwavelength structural units such as the Split Ring Resonator (SRR) and the Complementary Split Ring Resonator (CSRR) (Alrayes & Hussein, 2021; RoyChoudhury et al., 2016; Su et al., 2016). These structures concentrate electromagnetic fields in small regions such as gaps or slots, converting even minute dielectric variations in the sample into measurable frequency shifts (Δf). This approach enables high-precision and high-accuracy measurements with microliter or sub-microliter volumes, particularly when combined with microfluidic channels and low-cost planar printed circuit board (PCB) platforms (Falcone et al., 2004; Muñoz-Enano et al., 2020; Withayachumnankul et al., 2013).

The effectiveness of this technology has been demonstrated across diverse domains. In biomedical applications, SRR/CSRR-based sensors have been reported to successfully achieve label-free measurements for targets such as glucose (diabetes management), biomolecular binding (PSA, cortisol), and cell density/proliferation (cancer cell lines) (Jaruwongrungsee et al., 2015; Lee et al., 2012; Lee & Yook, 2008). As a notable proof of clinical accuracy, pilot human studies on microwave-based non-invasive glucose sensors reported Mean Absolute Relative Difference (MARD) values of approximately 12.5% (Choi et al., 2017). On the industrial and environmental side, CSRR/OCSRR architectures have demonstrated resolutions as low as 0.125 g/L for electrolyte concentration measurements (Velez et al., 2019).

Metamaterial-based microwave sensors hold the potential to be integrated not only into laboratory setups but also into portable devices suitable for clinical and industrial environments. Combined with low-cost fabrication methods and microfluidic systems, this technology facilitates the development of next-generation biosensors across a wide range of applications including personalized medicine, food safety, and environmental monitoring. In doing so, metamaterials maximize the potential of electromagnetic waves for precision sensing, establishing themselves as a foundational component of future sensor technologies.

In the existing literature, metamaterial-based microwave sensors have been widely investigated in biomedical domains such as glucose monitoring, protein interactions, and cellular analysis, with promising results reported (Choi et al., 2017; Jaruwongrungsee et al., 2015; Lee et al., 2012). However, these studies have primarily focused on individual application cases or specific design variants. Only a limited number of works systematically compare sensor architectures, comprehensively evaluate performance metrics, or analyze future trends. This chapter aims to address that gap by presenting a holistic perspective that encompasses the electromagnetic foundations of metamaterial-based microwave sensors, their design

principles, biomedical applications, and future development directions. In this way, it provides readers with both a summary of the current state and a framework for identifying potential research opportunities.

Electromagnetic Foundations and Metamaterials

Metamaterials are artificial structures designed to achieve electromagnetic properties not found in natural materials. Composed of periodically repeated unit cells at the subwavelength scale, these structures are grounded in Veselago's theoretical framework, which first described the wave propagation characteristics of media exhibiting simultaneous negative dielectric permittivity (ϵ) and magnetic permeability (μ) (Veselago, 1968). Pendry and colleagues demonstrated that artificially engineered microstructures composed of conductors could exhibit effective magnetic permeability (μ) not present in nature through resonance, and that this property could be tuned within specific frequency bands (Pendry et al., 1999). According to Babinet's principle, the Complementary Split Ring Resonator (CSRR), developed as the counterpart of the SRR, is formed by etching a conductive plane and predominantly couples with the electric field (Falcone et al., 2004).

In this model, a variation ($\Delta\epsilon'$) in the dielectric constant of the sample placed on the sensor surface directly affects the effective capacitance (Ceff) of the circuit, resulting in a measurable frequency shift (Δf_0). Similarly, dielectric losses (ϵ'') in the sample influence the resonance depth (|S21|) and the Q factor. Experimental studies have demonstrated that in CSRR-based microfluidic sensors, frequency-shift-based sensitivity can reach the order of MHz/ϵ' (Ebrahimi et al., 2014).

One of the key advantages of metamaterials in enhancing sensor performance is their ability to control and concentrate field distributions within specific regions. In SRRs, magnetic coupling via the H-field is dominant in the gap region, whereas in CSRRs, electric coupling via the E-field dominates around the slot (Baena et al., 2005; Bonache et al., 2006). The high E-field concentration generated around CSRR slots makes these sensors particularly advantageous for detecting small-volume liquid biomedical samples or variations within microfluidic channels. These effective parameters exhibit strong frequency dispersion around the resonance frequency, which explains the characteristic narrowband operating principle of metamaterials.

Mathematical Modeling

The fundamental operating principle of metamaterial-based microwave sensors can be described by modeling resonator structures with electrical equivalent circuits. The behavior of structures such as SRRs and CSRRs is commonly represented using the LC circuit model. In this framework, the resonance frequency is expressed as:

$$f_0 = \frac{1}{2\pi\sqrt{LC}}$$

where L is the inductance of the resonator and C is its capacitance. A change in the dielectric constant of the sample (Δε′) directly alters the effective capacitance (Ceff), leading to a shift in resonance frequency. The quality factor (Q) determines the sharpness of the resonance and thus the sensor's resolution. Higher Q values correspond to narrower bandwidths and more precise measurements. Sensitivity is one of the most critical parameters for evaluating sensor performance. This expression provides a direct physical understanding of the frequency shifts observed in the measurement of different samples. Experimental studies have reported sensitivity in the order of MHz/ε′ for CSRR-based sensors. Thus, the LC equivalent circuit model provides both a theoretical and practical tool for the design of metamaterial-based sensors, demonstrating high agreement with experimental results.

Use of Metamaterials in Microwave Sensor Design

Metamaterial-based microwave sensors are primarily designed by integrating resonator structures onto planar transmission lines. This integration is typically achieved by coupling Split Ring Resonator (SRR) or Complementary Split Ring Resonator (CSRR) elements to microstrip or coplanar waveguide (CPW) lines either laterally or from above, producing a notch-type resonance in the transmission response (Kazemi et al., 2024; Ke et al., 2018; Mayani et al., 2021). For maximum sensitivity, the sample is aligned with the resonator's gap or slot, where the electromagnetic field is most intense. Particularly, CSRR-based sensors offer significant potential for accurate characterization of the dielectric properties of liquid biological samples and cell cultures due to their high sensitivity to electric fields (Buragohain et al., 2021; Han et al., 2024).

The performance of these sensors is closely linked to the choice of substrate materials, which directly influence the Q factor and losses. Common substrates include low-loss materials such as Rogers or Teflon-based laminates, as well as cost-effective options like FR-4.

In advanced designs, integration with microfluidic systems is particularly emphasized for biomedical applications. In this approach, microchannels fabricated from materials such as polydimethylsiloxane (PDMS) or glass are precisely aligned with the slot region of the CSRR, where the E-field is most concentrated. This enables analyses with microliter (μL) or sub-microliter volumes, offering rapid, low-cost, and reproducible results. For instance, in a ~2 GHz CSRR-microfluidic sensor developed by Ebrahimi and colleagues, a frequency shift of ~400 MHz was achieved for water/ethanol mixtures, and the complex permittivity of the samples was successfully retrieved with strong numerical-experimental agreement (Ebrahimi et al., 2014).

To improve accuracy and stability in practical applications, advanced approaches such as differential architectures are employed. In these configurations, the difference between signals from a reference and a sample channel is measured, compensating for environmental drifts such as temperature and humidity.

The measurement and readout process are carried out by sweeping the S21 parameter using a Vector Network Analyzer (VNA) or an embedded RF front-end. Calibration is performed with known solution series, temperature compensation, and consideration of possible geometric tolerances. From these measurements, multiple features such as resonance frequency, quality factor, and minimum transmission amplitude are extracted, and regression models are used to accurately estimate the dielectric properties (ϵ', ϵ'') of the sample. These developments pave the way for the future integration of metamaterial-based sensors into portable and wearable devices for clinical and field use.

Table 1. Examples of metamaterial-based sensors

Sensor Structure	**Band (GHz)**	**Feature**	**Featured Metric**	**References**
SRR-microfluidics	~1.5–3	High Q	Label-free binding/ε' measurement	(Jaruwongrungsee et al., 2015; Lee et al., 2012; Withayachumnankul et al., 2013)
CSRR-microfluidics	~2	Wide E-field	~400 MHz total shift (water/EtOH)	(Ebrahimi et al., 2014)
OCSRR-differential	2–3	Environmental immunity	0.25 g/L (NaCl) resolution	(Velez et al., 2018)
CSRR	2–3	Differential	0.125 g/L resolution	(Velez et al., 2019)

Performance Metrics and Comparison

The success of metamaterial-based microwave sensors relies on the combined evaluation of multiple performance metrics. The most widely used criteria are summarized below:

Frequency Shift (Δf): The absolute shift in resonance frequency due to changes in the dielectric properties of the sample. Larger shifts indicate higher sensitivity.

Quality Factor (Q): Indicates the sharpness of the resonance. A higher Q factor allows detection of smaller dielectric differences.

Limit of Detection (LOD): The minimum concentration or dielectric change that the sensor can reliably measure.

Selectivity: The ability of the sensor to correctly distinguish between different parameters. In biomolecular interactions, surface functionalization (e.g., antibody/aptamer coating) enhances selectivity.

Dynamic Range: The range of concentrations or permittivity values over which the sensor can operate linearly. A wide dynamic range increases versatility across sample types.

Repeatability and Stability: The ability of the sensor to produce consistent results under identical conditions, which is particularly critical for clinical applications.

Considering these metrics together clarifies what defines a successful sensor and enhances the comparability of novel designs.

BIOMEDICAL APPLICATIONS

Metamaterial-based microwave sensors hold significant potential in the field of biomedical diagnostics. Their fundamental operating principle relies on the interaction between electromagnetic waves and biological tissues or fluids. This interaction enables direct measurement of parameters such as cell density, viability, and drug responses through variations in the dielectric properties of cells. As a result, rapid and non-invasive measurements can be performed without the need for labeling or chemical tagging methods commonly required in conventional biological analyses. In practice, these sensor designs are modeled, optimized, and validated using full-wave electromagnetic simulation software prior to fabrication.

Glucose Monitoring

Metamaterial-based sensors have been effectively employed in biochemical analyses such as glucose monitoring. The dielectric properties of solutions with different glucose concentrations can be precisely distinguished, and a consistent calibration curve is obtained between concentration and resonance frequency. This approach aims to advance the development of non-invasive glucose monitoring systems for diabetes management. In studies conducted in this field, metamaterial-based sensors operating in the terahertz band have been used to characterize glucose solutions in the range of 4–10 mmol/L, where a linear resonance frequency shift was observed with increasing concentration. In CSRR-microfluidic

architectures, measured frequency shifts have been reported on the order of MHz/ε′, enabling the detection of even small concentration changes in biological fluids (Huang et al., 2024; Lin et al., 2017; Yang et al., 2021). These findings indicate progress toward clinical implementation, while also highlighting the critical importance of compensating for subject-specific calibration as well as environmental variables such as tissue properties and temperature.

Cellular Analysis and Cancer

Cancer diagnosis and monitoring represent one of the most significant application domains of metamaterial-based microwave sensors (Hamza et al., 2024). Studies in the literature have revealed that different cancer cell types exhibit distinct electromagnetic responses compared to healthy cells. For example, Zhang et al. demonstrated that colorectal cancer cell lines could be distinguished using this method (Zhang et al., 2014), while D'Alvia et al. successfully differentiated cancer cells with varying levels of aggressiveness using a resonator (D'Alvia et al., 2022).

The underlying mechanism is based on changes in cell density altering the effective dielectric constant (ϵ′) around the resonator, which in turn shifts the resonance frequency. This phenomenon can be applied to real-time, label-free monitoring of cell proliferation or the cytotoxic effects of chemotherapeutic agents.

Protein and Biomolecular Interactions

Another important application area is the monitoring of protein structural changes and biomolecular interactions. In this domain, Lee and colleagues demonstrated the label-free detection of Prostate-Specific Antigen (PSA) and the hormone cortisol using planar SRR sensors (Lee et al., 2012; Lee & Yook, 2008). Functionalization of the sensor surface with biorecognition molecules such as antibodies or aptamers plays a critical role in enhancing selectivity toward the target analyte.

Collectively, these applications demonstrate that metamaterial-based microwave sensors are evolving beyond academic research topics and possess practical potential for clinical laboratory implementation. Particularly when combined with portable systems, wireless communication technologies, and artificial intelligence-assisted data analysis, these sensors are anticipated to revolutionize personalized medicine and continuous patient monitoring.

Clinical Translation and Challenges

Although metamaterial-based microwave sensors hold considerable promise in biomedical applications, the transition from laboratory-scale prototypes to clinical implementation faces several limitations and challenges:

Inter-Individual Variability: In non-invasive measurements, personal differences in skin thickness, water content, and tissue properties directly affect sensor performance (Gidado et al., 2022; Mamouei et al., 2021). Therefore, subject-specific calibration protocols must be developed.

Environmental Influences: Factors such as temperature, humidity, and salinity can cause resonance frequency drifts (Jan et al., 2016). Differential architectures and temperature compensation algorithms are employed to mitigate these effects.

Calibration Challenges: Continuous calibration is impractical in clinical practice, necessitating the development of one-time or automated calibration methods.

Biocompatibility and Safety: In cases of direct contact with the human body, biocompatible materials must be used, and electromagnetic safety standards must be ensured (Liu et al., 2023; Lu et al., 2023).

Regulation and Standardization: Clinical translation requires compliance with regulatory authorities' safety and accuracy criteria. Thus, sensors must be validated not only in laboratory tests but also through long-term clinical trials.

Overcoming these challenges will enable metamaterial-based microwave sensors to move beyond research prototypes and achieve widespread adoption in clinical laboratories and point-of-care diagnostic systems.

FUTURE TRENDS

Metamaterial-based microwave sensors represent a rapidly evolving research area in the biomedical field, and they are expected to offer smarter, more integrated and clinically applicable solutions in the future. In particular, the integration of artificial intelligence (AI), the Internet of Things (IoT), and advanced manufacturing technologies into these sensors has the potential to fundamentally transform biomedical diagnosis and therapy.

Integration of Artificial Intelligence and Machine Learning

Artificial intelligence and machine learning algorithms play a critical role in analyzing the complex datasets generated by these sensors. The data consist of multiple features such as resonance frequency, quality factor, amplitude, and phase, which are often difficult to interpret using classical methods. In this context, classifiers such as Support Vector Machines (SVM), k-Nearest Neighbors (KNN), and Convolutional Neural Networks

(CNN) enhance both the selectivity and noise tolerance of biomedical sample classification while ensuring high accuracy. For example, machine learning–based analyses have been reported to achieve accuracy rates above 85% in distinguishing cancerous from healthy cells (Masud et al., 2021; Singh & Gupta, 2019). Furthermore, AI-based optimization and inverse design approaches are increasingly employed to automatically refine sensor geometries and accelerate the development of new designs.

IoT and Portable/Wearable Systems

The low-profile structure and low power consumption of metamaterial-based sensors enable their integration with Internet of Things (IoT) technologies, opening a new era in clinical diagnostics and patient monitoring. By combining these sensors with low-cost microcontrollers and wireless communication modules such as Bluetooth Low Energy (BLE) or NB-IoT, portable and wearable devices can be developed. Such devices could continuously monitor biomarkers such as glucose levels and transmit the data in real time to cloud-based analytics platforms. This approach provides a scalable and promising pathway for chronic disease management and home healthcare applications.

Advanced Designs and Fabrication Technologies

New designs that go beyond conventional structures are being developed to further enhance sensor performance. For example, reconfigurable sensors integrated with active components such as varactor diodes or Micro-Electromechanical Systems (MEMS) switches offer the ability to dynamically tune the resonance frequency (Malavika et al., 2022; Parsediya et al., 2015). This enables multi-biomarker detection across different frequency bands using a single hardware platform. In addition, advanced fabrication techniques such as three-dimensional printing and paper-based platforms facilitate the rapid prototyping of low-cost, customizable sensors, thereby accelerating their transition into clinical and field applications.

Looking ahead, the convergence of these technologies is expected to make metamaterial-based microwave sensors more precise, portable, and patient-centered. With AI-assisted data processing, IoT-enabled continuous monitoring, and optimized metamaterial architectures, biomedical diagnosis and therapy will become faster, more affordable, and more reliable.

CONCLUSION

Metamaterial-based microwave sensors represent an innovative approach with the potential to revolutionize biomedical diagnosis and monitoring technologies. The practical value of this approach has been validated through concrete metrics obtained in applications such as cancer cell detection, glucose monitoring, and the analysis of biomolecular interactions.

Compared to conventional diagnostic methods, metamaterial-based microwave sensors offer significant advantages including non-invasive measurement, rapid analysis times, portability, lightweight construction, and label-free operation. Integration with microfluidic systems enables highly reproducible analyses using microliter-scale samples, thereby laying the groundwork for a new technological infrastructure in both pharmaceutical research and early diagnostic applications.

Future developments in this field are expected to accelerate further. Artificial intelligence and machine learning algorithms will automate the analysis of complex data, while IoT-based portable and wearable systems will enable continuous patient monitoring. At the same time, reconfigurable metamaterials and technologies such as flexible printing will facilitate the production of application-specific, low-cost, and customizable sensors, thereby accelerating their clinical and industrial adoption.

In light of these advancements, the integration of metamaterial-based microwave sensors with artificial intelligence, microfluidic systems, and wireless communication technologies represents the beginning of a new era in personalized medicine. Consequently, multidisciplinary research in this domain is poised to pave the way for groundbreaking innovations in healthcare technologies and generate high value-added outcomes.

The contribution of this chapter to the literature lies in its systematic evaluation of diverse and scattered application examples of metamaterial-based microwave sensors and its interdisciplinary discussion of future research directions. By consolidating concrete biomedical metrics (e.g., glucose monitoring, PSA detection, cell differentiation), the chapter clearly highlights the potential of this technology for clinical translation. Furthermore, its discussion on the integration of artificial intelligence, IoT, and reconfigurable metamaterial designs provides a novel roadmap to guide future studies. In this respect, the chapter not only summarizes existing knowledge but also identifies new opportunities for researchers.

REFERENCE

Alrayes, N., & Hussein, M. I. (2021). Metamaterial-based sensor design using split ring resonator and Hilbert fractal for biomedical application. *Sensing and Bio-Sensing Research*, *31*, 100395. https://doi.org/10.1016/j.sbsr.2020.100395

Baena, J. D., Bonache, J., Martin, F., Sillero, R. M., Falcone, F., Lopetegi, T., Laso, M. A. G., Garcia-Garcia, J., Gil, I., Portillo, M. F., & Sorolla, M. (2005). Equivalent-circuit models for split-ring resonators and complementary split-ring resonators coupled to planar transmission lines. *IEEE Transactions on Microwave Theory and Techniques*, *53*(4), 1451–1461. https://doi.org/10.1109/TMTT.2005.845211

Bonache, J., Gil, M., Gil, I., Garcia-Garcia, J., & Martin, F. (2006). On the electrical characteristics of complementary metamaterial resonators. *IEEE Microwave and Wireless Components Letters*, *16*(10), 543–545. https://doi.org/10.1109/LMWC.2006.882400

Buragohain, A., Mostako, A. T. T., & Das, G. S. (2021). Low-Cost CSRR Based Sensor for Determination of Dielectric Constant of Liquid Samples. *IEEE Sensors Journal*, *21*(24), 27450–27457. https://doi.org/10.1109/JSEN.2021.3124329

Choi, H., Luzio, S., Beutler, J., & Porch, A. (2017). Microwave noninvasive blood glucose monitoring sensor: Human clinical trial results. *2017 IEEE MTT-S International Microwave Symposium (IMS)*, 876–879. https://doi.org/10.1109/MWSYM.2017.8058721

D'Alvia, L., Carraro, S., Peruzzi, B., Urciuoli, E., Palla, L., Del Prete, Z., & Rizzuto, E. (2022). A Novel Microwave Resonant Sensor for Measuring Cancer Cell Line Aggressiveness. *Sensors*, *22*(12), 4383. https://doi.org/10.3390/s22124383

Ebrahimi, A., Withayachumnankul, W., Al-Sarawi, S., & Abbott, D. (2014). High-Sensitivity Metamaterial-Inspired Sensor for Microfluidic Dielectric Characterization. *IEEE Sensors Journal*, *14*(5), 1345–1351. https://doi.org/10.1109/JSEN.2013.2295312

Falcone, F., Lopetegi, T., Laso, M. A. G., Baena, J. D., Bonache, J., Beruete, M., Marqués, R., Martín, F., & Sorolla, M. (2004). Babinet Principle Applied to the Design of Metasurfaces and Metamaterials. *Physical Review Letters*, *93*(19), 197401. https://doi.org/10.1103/PhysRevLett.93.197401

Gidado, I. M., Qassem, M., Triantis, I. F., & Kyriacou, P. A. (2022). Review of Advances in the Measurement of Skin Hydration Based on Sensing of Optical and Electrical Tissue Properties. *Sensors*, *22*(19), 7151. https://doi.org/10.3390/s22197151

Hamza, M. N., Koziel, S., & Pietrenko-Dabrowska, A. (2024). Design and experimental validation of a metamaterial-based sensor for microwave imaging in breast, lung, and brain cancer detection. *Scientific Reports*, *14*(1), 16177. https://doi.org/10.1038/s41598-024-67103-9

Han, X., Liu, K., Zhang, S., Peng, P., Fu, C., Qiao, L., & Ma, Z. (2024). CSRR Metamaterial Microwave Sensor for Measuring Dielectric Constants of Solids and Liquids. *IEEE Sensors Journal*, *24*(9), 14167–14176. https://doi.org/10.1109/JSEN.2024.3373755

Huang, X., Wang, Y., Li, X., & Liu, T. (2024). Minimal microfluidic metamaterial sensor for concentration detection. *Measurement*, *224*, 113864. https://doi.org/10.1016/j.measurement.2023.113864

Jan, M. T., Ahmad, F., Hamid, N. H. B., Khir, M. H. B. M., Shoaib, M., & Ashraf, K. (2016). Experimental investigation of temperature and relative humidity effects on resonance frequency and quality factor of CMOS-MEMS paddle resonator. *Microelectronics Reliability*, *63*, 82–89. https://doi.org/10.1016/j.microrel.2016.05.007

Jaruwongrungsee, K., Waiwijit, U., Withayachumnankul, W., Maturos, T., Phokaratkul, D., Tuantranont, A., Wlodarski, W., Martucci, A., & Wisitsoraat, A. (2015). Microfluidic-based Split-Ring-Resonator Sensor for Real-time and Label-free Biosensing. *Procedia Engineering*, *120*, 163–166. https://doi.org/10.1016/j.proeng.2015.08.595

Kazemi, N., Abdolrazzaghi, M., Musilek, P., & Baladi, E. (2024). A Planar Compact Absorber for Microwave Sensing Based on Transmission-Line Metamaterials. *IEEE Sensors Journal*, *24*(24), 41864–41874. https://doi.org/10.1109/JSEN.2024.3484585

Ke, L., Liu, Z., & Yu, H. (2018). Characterization of a Patch Antenna Sensor's Resonant Frequency Response in Identifying the Notch-Shaped Cracks on Metal Structure. *Sensors*, *19*(1), 110. https://doi.org/10.3390/s19010110

Lee, H.-J., Lee, J.-H., Moon, H.-S., Jang, I.-S., Choi, J.-S., Yook, J.-G., & Jung, H.-I. (2012). A planar split-ring resonator-based microwave biosensor for label-free detection of biomolecules. *Sensors and Actuators B: Chemical*, *169*, 26–31. https://doi.org/10.1016/j.snb.2012.01.044

Lee, H.-J., & Yook, J.-G. (2008). Biosensing using split-ring resonators at microwave regime. *Applied Physics Letters*, *92*(25), 254103. https://doi.org/10.1063/1.2946656

Lin, T., Gu, S., & Lasri, T. (2017). Highly sensitive characterization of glucose aqueous solution with low concentration: Application to broadband dielectric spectroscopy. *Sensors and Actuators A: Physical*, *267*, 318–326. https://doi.org/10.1016/j.sna.2017.10.029

Liu, G., Lv, Z., Batool, S., Li, M., Zhao, P., Guo, L., Wang, Y., Zhou, Y., & Han, S. (2023). Biocompatible Material-Based Flexible Biosensors: From Materials Design to Wearable/Implantable Devices and Integrated Sensing Systems. *Small*, *19*(27), 2207879. https://doi.org/10.1002/smll.202207879

Lu, T., Ji, S., Jin, W., Yang, Q., Luo, Q., & Ren, T.-L. (2023). Biocompatible and Long-Term Monitoring Strategies of Wearable, Ingestible and Implantable Biosensors: Reform the Next Generation Healthcare. *Sensors*, *23*(6), 2991. https://doi.org/10.3390/s23062991

Malavika, J., Parameshwari, R., Kalyani, D., & Lakshmi Prabha, P. (2022). MEMS Biosensor Design and Simulation for Diagnostic Purposes. *Journal of Physics: Conference Series*, *2318*(1), 012018. https://doi.org/10.1088/1742-6596/2318/1/012018

Mamouei, M., Chatterjee, S., Razban, M., Qassem, M., & Kyriacou, P. A. (2021). Design and Analysis of a Continuous and Non-Invasive Multi-Wavelength Optical Sensor for Measurement of Dermal Water Content. *Sensors*, *21*(6), 2162. https://doi.org/10.3390/s21062162

Masud, M., Sikder, N., Nahid, A.-A., Bairagi, A. K., & AlZain, M. A. (2021). A Machine Learning Approach to Diagnosing Lung and Colon Cancer Using

a Deep Learning-Based Classification Framework. *Sensors*, *21*(3), 748. https://doi.org/10.3390/s21030748

Mayani, M. G., Herraiz-Martinez, F. J., Domingo, J. M., & Giannetti, R. (2021). Resonator-Based Microwave Metamaterial Sensors for Instrumentation: Survey, Classification, and Performance Comparison. *IEEE Transactions on Instrumentation and Measurement*, *70*, 1–14. https://doi.org/10.1109/TIM.2020.3040484

Muñoz-Enano, J., Vélez, P., Gil, M., & Martín, F. (2020). Planar Microwave Resonant Sensors: A Review and Recent Developments. *Applied Sciences*, *10*(7), 2615. https://doi.org/10.3390/app10072615

Parsediya, D. K., Singh, J., & Kankar, P. K. (2015). Variable width based stepped MEMS cantilevers for micro or pico level biosensing and effective switching. *Journal of Mechanical Science and Technology*, *29*(11), 4823–4832. https://doi.org/10.1007/s12206-015-1029-6

Pendry, J. B., Holden, A. J., Robbins, D. J., & Stewart, W. J. (1999). Magnetism from conductors and enhanced nonlinear phenomena. *IEEE Transactions on Microwave Theory and Techniques*, *47*(11), 2075–2084. https://doi.org/10.1109/22.798002

Prakash, D., & Gupta, N. (2022). Applications of metamaterial sensors: A review. *International Journal of Microwave and Wireless Technologies*, *14*(1), 19–33. https://doi.org/10.1017/S1759078721000039

Puentes Vargas, M. (2014). *Planar Metamaterial Based Microwave Sensor Arrays for Biomedical Analysis and Treatment*. Springer International Publishing. https://doi.org/10.1007/978-3-319-06041-5

RoyChoudhury, S., Rawat, V., Jalal, A. H., Kale, S. N., & Bhansali, S. (2016). Recent advances in metamaterial split-ring-resonator circuits as biosensors and therapeutic agents. *Biosensors and Bioelectronics*, *86*, 595–608. https://doi.org/10.1016/j.bios.2016.07.020

Singh, G. A. P., & Gupta, P. K. (2019). Performance analysis of various machine learning-based approaches for detection and classification of lung cancer in humans. *Neural Computing and Applications*, *31*(10), 6863–6877. https://doi.org/10.1007/s00521-018-3518-x

Su, L., Naqui, J., Mata-Contreras, J., & Martin, F. (2016). Modeling and Applications of Metamaterial Transmission Lines Loaded With Pairs of Coupled Complementary Split-Ring Resonators (CSRRs). *IEEE Antennas and Wireless Propagation Letters*, *15*, 154–157. https://doi.org/10.1109/LAWP.2015.2435656

Velez, P., Grenier, K., Mata-Contreras, J., Dubuc, D., & Martin, F. (2018). Highly-Sensitive Microwave Sensors Based on Open Complementary Split Ring Resonators (OCSRRs) for Dielectric Characterization and Solute Concentration Measurement in Liquids. *IEEE Access*, *6*, 48324–48338. https://doi.org/10.1109/ACCESS.2018.2867077

Velez, P., Munoz-Enano, J., & Martin, F. (2019). Electrolyte Concentration Measurements in DI Water with 0.125 g/L Resolution by means of CSRR-Based Structures. *2019 49th European Microwave Conference (EuMC)*, 340–343. https://doi.org/10.23919/EuMC.2019.8910697

Veselago, V. G. (1968). THE ELECTRODYNAMICS OF SUBSTANCES WITH SIMULTANEOUSLY NEGATIVE VALUES OF ϵ AND μ.

Soviet Physics Uspekhi, *10*(4), 509–514. https://doi.org/10.1070/PU1968v010n04ABEH003699

Wang, J., Wang, R., Shen, Z., Liu, B., Sun, C., & Xue, Q. (2025). Microwave biosensors utilizing metamaterial enhancement: Design and application. *Nanotechnology and Precision Engineering*, *8*(1), 015001. https://doi.org/10.1063/10.0028757

Withayachumnankul, W., Jaruwongrungsee, K., Tuantranont, A., Fumeaux, C., & Abbott, D. (2013). Metamaterial-based microfluidic sensor for dielectric characterization. *Sensors and Actuators A: Physical*, *189*, 233–237. https://doi.org/10.1016/j.sna.2012.10.027

Yang, J., Qi, L., Li, B., Wu, L., Shi, D., Ahmed Uqaili, J., & Tao, X. (2021). A terahertz metamaterial sensor used for distinguishing glucose concentration. *Results in Physics*, *26*, 104332. https://doi.org/10.1016/j.rinp.2021.104332

Zhang, L. Y., Bounaix Morand Du Puch, C., Dalmay, C., Lacroix, A., Landoulsi, A., Leroy, J., Mélin, C., Lalloué, F., Battu, S., Lautrette, C., Giraud, S., Bessaudou, A., Blondy, P., Jauberteau, M. O., & Pothier, A. (2014). Discrimination of colorectal cancer cell lines using microwave biosensors. *Sensors and Actuators A: Physical*, *216*, 405–416. https://doi.org/10.1016/j.sna.2014.03.022

Memristor Types and Application Areas

Muhammet ATEŞ

1- (Asst. Prof.) Department of Electronics and Automation Van Vocational School of Higher Education, Van Yüzüncü Yıl University, 65080, Van, Turkey, mts@yyu.edu.tr, ORCID No: 0000-0003-2223-2745

ABSTRACT

The memristor, theorized by Prof. Dr. Leon Chua in 1971 as the fourth fundamental passive circuit element, complements the resistor, capacitor, and inductor by describing the direct relationship between electric charge and magnetic flux. Chua's theoretical framework filled a crucial gap in circuit theory using differential equations and predicted a component he named the "memristor" (memory resistor). For decades, this component remained a theoretical concept until 2008, when Strukov et al. physically realized the memristor while working on molecular electronics, thus confirming Chua's predictions.

This breakthrough led to a significant rise in memristor research, exploring areas such as memristor modeling, logic-based computation, neuromorphic systems, artificial intelligence, and RFID integration. Memristors, with their non-volatile memory capability and low power consumption, offer promising advancements in computational systems by enabling analog learning processes similar to the human brain. This is particularly valuable for applications in facial recognition, speech understanding, and decision-making.

Numerous studies have expanded on this foundation. Researchers have proposed various memristor models—such as piecewise linear, Simmons tunnel barrier, and adaptive models—analyzed their behavior using differential equations, and applied them in chaos circuits, analog filters, and SPICE simulations. Educational efforts have also been made, including emulator circuits to aid teaching and laboratory applications. Overall, memristor research continues to evolve, bridging theoretical foundations with practical applications in next-generation electronic systems. required.

*Keywords – Memristor, Passive Circuit Elements, Magnetic Flux, Charge, Neuromorphic Computing, Non-Volatile Memory, Analog Learning, Memristor Modeling, Artificial Intelligence, SPICE Simulation, Molecular Electronic*s.

INTRODUCTION

Three types of passive circuit elements were used to build electronic circuits: resistors, capacitors and inductors. To describe these circuit elements, combinations of 4 basic circuit variables called electric current, voltage, load and magnetic flux are used. Resistance defines the relation between voltage and current; the capacitor relates electric charge to voltage; and the inductor establishes a link between magnetic flux and current. However, none of these conventional circuit elements describes the direct relationship between electric charge and magnetic flux—a gap that is theoretically filled by the memristor.

Prof. Dr. Leon Chua made theoretical predictions for a novel circuit component that describes the relationship between magnetic flux and charge. In a paper published in IEEE Trans. Circuit Theory, Chua theoretically defined the missing piece of he circuit built upon the relationships between voltage, load, current and magnetic flux that simple differential equations could express. In his article, Chua named this fourth circuit element the Memristor, based on the English word memory resistor.

This theoretical device, as it was called in 1971, remained mathematically on paper for years. Then, after 35 years, Stanley Williams and his group were working on molecular electronics when they noticed strange behavior in a device they had built. Then Greg Snider from the team found Chua's work from 1971. Williams read and reread Chua's paper over and over again for several years and soon realized that the molecular device they had found was, in fact, the memristor that Chua had theorized years earlier.

The physical fabrication of the memristor led to a surge in memristor research in the scientific world. A wide range of research has been conducted on emerging topics like memristor modeling, logic-based computation, Data processing, artificial neural systems, control strategies, and Radio Frequency Identification integration.

Especially, the memristor, which can store information in its memory without the need for a power supply, i.e. with zero energy, will enable powerful processors that can be realized with fewer circuit elements, and will make it possible to set the circuit switches, which are currently working serially digitally “1” and “0”, that is, open and closed, to a value between these two values, which is analogous to the human brain’s analog learning process, which is the method of learning through patterns. The ability of computers to learn by working in parallel with analog systems will enable extremely important developments in artificial intelligence applications, such as face recognition, speech understanding, and decision-making.

Some studies with Memristors in the literature are summarized:

(Chua, 1971); Within his article “Memristor-The Missing Circuit Element”, he explained the relationship between the basic circuit variables load and magnetic flux. He mathematically predicted the existence of the fourth fundamental passive element along with the memristor, resistor, capacitor, and coil. He also proposed a memristor equivalent circuit consisting of active circuit elements in this article.

(Strukov et al., 2008); In the study titled “The missing memristor found”, the memristor was physically realized for the first time in a laboratory environment. The model of the obtained element, the formula of the model, current-voltage characteristics are given. It is also emphasized in this article that the I-V behavior of the memristor at high frequencies, which

was previously mentioned in the article in 1976, is similar to that of the resistor.

(Wang et al., 2009); In the study titled "A PWL model of memristor and its application example", an example of the PWL (PieceWise Linear) piecewise linear memristor model is given. In addition, this model is applied to Chua's chaos oscillator circuit.

(Pickett et al., 2009); In his study, the properties, equation and circuit diagram illustrating the Simmons Tunnel Barrier memristor model are given. Additionally, the way this model works and the methods used to obtain the model are given.

(O. Kavehei et al., 2010); In his study, the behavior of the memristor was modeled using its governing differential equations. The methods used in memristor modeling and the theory of the memristor using Maxwell "s Equations used by Chua to describe the memristor mathematically were examined.

(Strachan et al., 2010); In his study, transitions to different phases occurred in the crystal structure of TiO_2 due to Joule heating during the electroforming process. Observations were made in different phases of TiO_2 and the electrical properties of the device were investigated by adding Ti_4O_7 to TiO_2 in Pt/TiO_2 /Pt structure.

(Pershin and Di Ventra, 2012) In his article, an emulator circuit is given to teach the memristor to students and a few memristor applications that can be created with this circuit in the laboratory environment are included.

(Pershin et al., 2012); In this work, the previously proposed memristor circuit is utilized in applications involving analog-to-digital and digital-to-analog signal conversion.

(Elgabra et al., 2012); In this study, the mathematical expressions of existing memristor models were derived and these models were compared. The results obtained were shared.

(Chanthbouala et al., 2012); In this study, the realization of the tunnel barrier model with ferroelectric materials and the results were shared.

(Kvatinsky et al., 2013); In the study, an adaptive memristor model was designed. They showed that the model can be adjusted to memristor models such as linear ion drift, nonlinear ion drift and Simmons tunnel barrier model.

(Xiao-Bo et al., 2013); In the article, a programmable analog filter design was made by utilizing the non-volatile memory feature of the memristor.

(Singh and Kakade, 2014); "Memristors MATLAB Modeling" includes the implementation of the HP memristor model in MATLAB environment and the analysis of the obtained model.

(Vourkas et al., 2015); In the study titled “SPICE modeling of nonlinear memristive behavior”, the characteristics of a nonlinear memristor model in SPICE environment is investigated.

(Chakrabarti et al., 2017); In the article, a new 3D memristor model was produced and the behavior of the element in analog and neuromorphic circuits was investigated.

HISTORY OF THE MEMRISTOR

In terms of classical circuit theories, there are four basic circuit variables: load (q), current (i), voltage (v) and magnetic flux (φ). The relationships between these variables are defined as the relationship between current and load $dq/dt = i$, the relationship between magnetic flux and voltage $d\varphi/dt = v$, and the relationships used to define the three basic circuit elements: resistor $R = dv/di$, capacitor $C = dq/dv$ and inductor $L = d\varphi/di$. However, the relationship between magnetic flux and load is not defined.

In 1971, Prof. Dr. Leon Chua predicted the existence of the Memristor by explaining the relationship between magnetic flux and charge. Chua defined the relationship between magnetic flux and charge as Memristance $M = d\varphi/dq$ and called this element a memristor. Figure 1 shows the basic circuit variables and the relations between them.

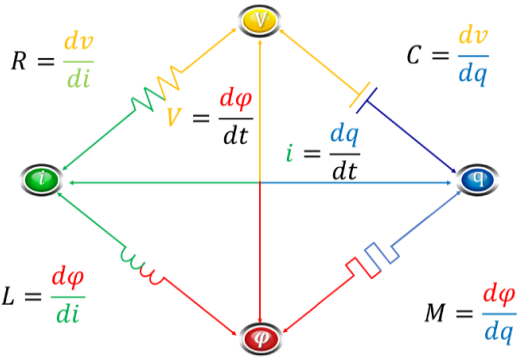


Figure 1. Basic Circuit Variables and Their Relationships

(Chua and Kang 1976) generalized the theory of memristive systems and memristors. Memristive systems were expressed as resistor memristance, capacitor memcapacitance, inductance memductance and the electrical and characteristic properties of these new devices were explained theoretically. In addition, the current-voltage characteristic of the memristor was given as a hysteresis curve for the first time in this study. It is also mentioned that the

$I-V$ characteristic of the memristor at high frequencies resembles the current-voltage characteristic of a linear resistor. The frequency-dependent $I-V$ response of the memristor is presented in Figure 2.

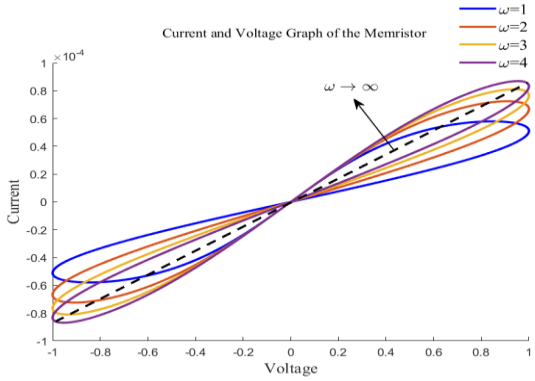


Figure 2. Frequency Dependent Current-Voltage Characteristic of Memristor.

Williams et al., 2008, at the HP laboratories, Williams and his group announced that they had for the first time physically realized the memristor, described mathematically by Chua in his 1971 paper, at the nanoscale in a laboratory setting. The memristor was previously unrecognized because it had little effect at standard dimensions. In practice, it is very difficult to make a pure memristor, because every device has at least one other feature. For example, all inductors have resistance. Similarly, a memristor has a capacitance. Chua built a memristor, albeit a crude one, in the 1970s to prove his concept. Chua's memristor was a combination of resistors, capacitors, inductors and amplifiers.

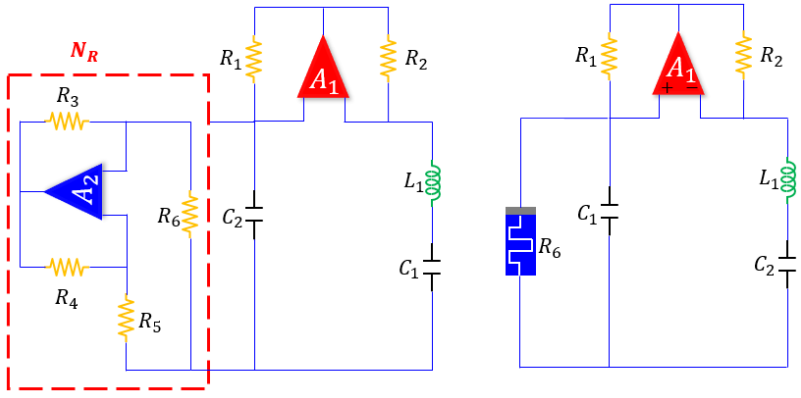


Figure 3. Chua's 4 basic circuit element model (NR part represents memristor)

But memory resistance was too weak to be used, or even recognized, as a property of this material until recently. Chua didn't realize it at the time either. Memory resistance was lost among the material's other properties. In their paper in Nature, Williams and his group say that the memristor manifests itself naturally, especially in nanoscale systems. This explains why no one has noticed it until now.

The physical element consists of pure and doped titanium dioxide (TiO_2 TiO_2) sandwiched between two platinum (Pt) layers. TiO_2 is a semiconductor material and is highly resistive in its pure form. Its resistance is reduced by doping with oxygen atoms. Thus, the low and high resistance values present in the current-voltage characteristic of the memristor are obtained. Figure 4 shows the nanoscale version of the memristor mentioned in this article.

Figure 4. Memristor Model at the nanoscale

After the physical realization of the memristor, the interest in memristors in the scientific world has increased considerably and new memristor models and their use in different circuit applications have become widespread.

MEMRISTOR DEFINITION

The memristor is a two-terminal circuit element whose name derives from the combination of memory" and resistor. The association the magnetic flux and the load acts like a resistor whose resistance value changes according to the amount of load passing through it. Since memristance (M), the magnitude that determines the $I-V$ characteristic of the memristor, is a function of the load, Chua defined $M(q)$ as equation 1.

$$M(q) = d\Phi / dq \quad (1)$$

If the above equation is expressed with respect tocurrent and voltage in equation 2

$$M(q(t)) = (d\Phi / dt)/(dq / dt) \quad (2)$$

or if we write equation 2 more simply

$$M(q(t)) = V(t)/i(t) \quad (3)$$

Since Equation 3 closely resembles the defining equation of resistance, the unit of the memristor is defined as "*ohm*".

If we take the voltage expression in Equation 3 to the beginning of the equation;

$$V(t) = M(q(t))\, i(t) \quad (4)$$

As can be understood from Equation 4 above, the resistance of the memristor varies depending on the polarity and amplitude of the voltage applied to it and the time interval of the voltage.

Equation (4) explains that Memristance results in a linear current–voltage relationship when the load is not taken into account. The truth is the current is different from zero indicates that the load changes over time. Also, when no current is applied, the memristor is static. In other words, if current and voltage do not change with time, $M(q(t))$ is constant. This characteristic illustrates the memristor's memory effect. The memristor theoretically keeps the last resistance value in its memory. Even if it is de-energized, it continues to operate from its last state when re-energized. Therefore, it is named memristor with the combination of the words memory and resistor. In some publications in the literature, it is translated as "Resistor with Memory" (Aytekin and Mutlu, 2013). Shown in Figure 5 is the schematic symbol of the memristor.

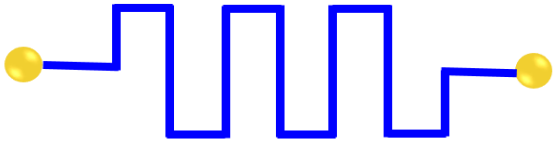


Figure 5. Memristor Symbol

To better illustrate the memristor's behavior under both positive and negative voltage conditions, the water pipe type memristor model proposed by Chua is used in Figure 5. The resistor can be compared to a hose through which water flows. The size of the inner diameter of the hose determines the resistance to the flow of water. The narrower the hose's diameter, the more it resists water flow. The wider it is, the less resistance it will have and the more and more easily the water will flow. In linear resistors, the inner diameter of this hose does not change. However, the situation is different in

the memristor. The radius expands or contracts based on the total amount of water that has passed through it. If we let the water flow through the hose in one direction, the inner diameter of the hose expands, i.e. its resistance decreases, and it does not forget this diameter but keeps it in its memory. If the flow of water is stopped, the hose remains in this expanded state. It keeps track of how much current has passed through it and how much is left.

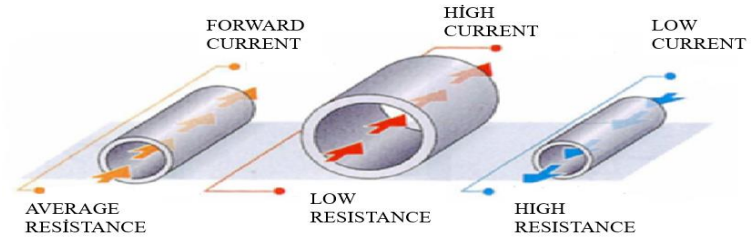


Figure 6. Water pipe type memristor model proposed by Chua

Chua described the nonlinear behavior of three basic circuit components, namely resistors, capacitors and inductors, and integrated these devices and expressed them as memristive systems. Chua referred to the resistor as "memristor", the capacitor as "memcapacitor," and the inductor as "meminductor" (Chua and Kang 1976). Memristive systems are systems involving the nonlinear dynamic behavior of basic passive circuit elements. In general, two-terminal devices whose internal state variable depends on resistance can be defined as memristive systems (Jo et al., 2010). In memristive systems, the internal resistance state of the device does not depend on the input current of the device. Since the internal resistance state of the device depends on the state of the current in the past, devices in memristive systems exhibit hysteresis behavior. The most important feature of memristive systems is that the devices are passive. They do not store energy like coils and capacitors with linear behavior. Figure 7 shows the circuit symbols of memristor, memcapacitor and meminductor. These symbols are modeled based on the resistance, capacitor and inductor symbols.

Figure 7. Memristor, Memcapacitor and Meminductor circuit symbols.

MEMRISTOR MODELS

1. HP Memristor Model (Linear Ion Drift Model)

In 2006, Stanley Williams and his team from HP laboratories were working on nanoelectronics when they noticed a strange behavior of an element. After a literature review, the team realized that this element was actually the element mentioned by Leon Chua in his 1971 article (Chua, 1971). In 2008, they published their paper “The Lost Memristor Found” and announced to the scientific world that the memristor had been physically realized (Strukov et al., 2008). Figure 8 shows the HP memristor model and Figure 9 shows the HP memristor equivalent.

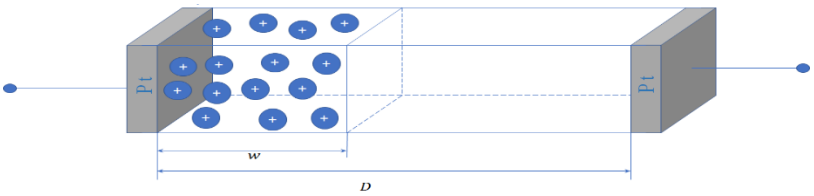


Figure 8. HP Memristor Model

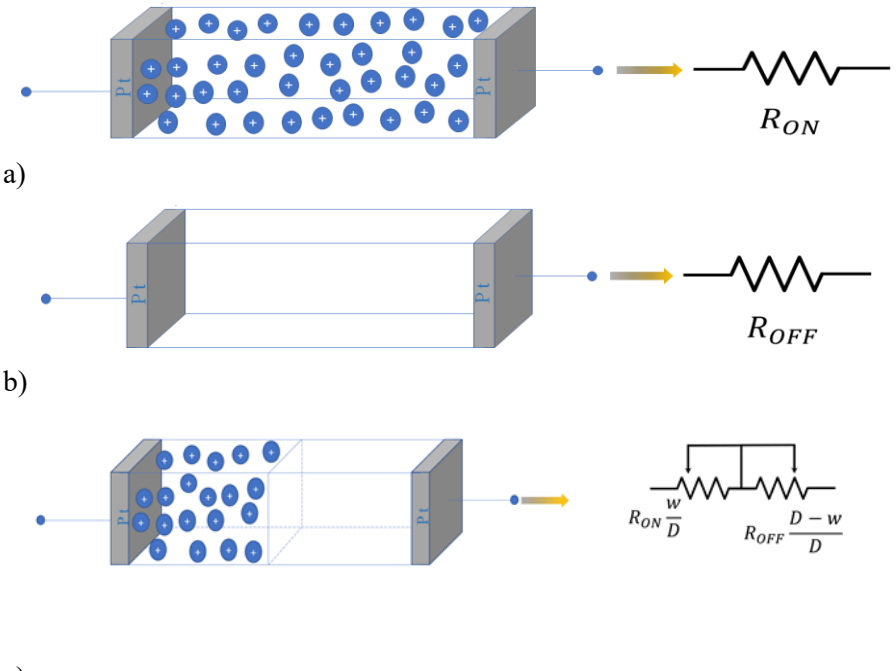


Figure 9. Represents a memristor structure with titanium dioxide (TiO_2) placed between platinum terminals. Upon applying a positive voltage to the TiO_2 memristor, oxygen ions start to diffuse in TiO_2 . If oxygen ions diffuse completely in the memristance value will be minimum and equal to R_{ON} . If the oxygen ions are not diffused at all in TiO_2 , the memristance value is maximum and equal to R_{OFF} .

The memristance of the HP memristor model expressible mathematically as in Equation 4.1. Where 'R_{MEM}' is the total resistance of the memristor, 'R_{OFF}' is the pure TiO_2 fraction of the memristor, 'R_{ON}' is the doped TiO_2 fraction of the memristor. 'x' is the ratio of the doped TiO_2 fraction to the whole TiO_2 layer and takes values between 0 and 1.

'w' represents the doped TiO_2 fraction and 'D' represents the whole TiO_2 layer. Equation 4.2 shows the variation of the total resistance of the memristor when the 'x' ratio is 0 and 1, and Equation 4 defines this ratio. The I-V correlation of the memristor is given in equation 5.

$$Pt\, R_{MEM}(x) = R_{ON} * x + R_{OFF} * (1 - x) \quad (5)$$

$$R_{MEM}(x) = \begin{cases} x = 0 \;\; ise \; R_{MEM} = R_{OFF} \\ x = 1 \;\; ise \; R_{MEM} = R_{ON} \end{cases} \quad (6)$$

$$x = \frac{w}{D}$$

$$v(t) = \left(R_{ON} * \frac{w(t)}{D} + R_{OFF} * \left(1 - \frac{w(t)}{D} \right) \right) * i(t) \quad (7)$$

Williams' memristance depends linearly on the memristor charge and this is due to the linear drift velocity of oxygen ions. The memristance formula given by Williams' team is given in equation 8;

$$M(q) = R_{OFF}(1 - \frac{\mu v \, R_{ON}}{D^2} \, q(t)) \quad (8)$$

In this work, Stanley Williams and his team used titanium dioxide (TiO_2) and platinum (Pt). TiO_2 , like silicon (Si), is a semiconductor and has a very high resistivity in its pure form. But TiO_2 is made conductive by interacting with different elements.

The dopants used to make TiO_2 conductive are stable under a strong electric field and are likely to drift in the direction of the current. For this reason, the model is also referred to as the linear ion drift model.. In this model, a three-nanometer-thick TiO_2 layer was placed between two platinum (Pt) layers. One section of the TiO_2 layer had positively charged vacancies where oxygen atoms would normally be. The team applied alternating current to an electrode near these vacancies, causing it to oscillate between positive and negative charges. When the electrode was plus-charged, it repelled the charged vacancies and allowed the current to flow toward the second electrode. When the current was cut off, the vacancies stopped moving, and the memristor remained in a high or low resistance state. Figure 10 depicts how the memristor behaves under both positive and negative voltage application.

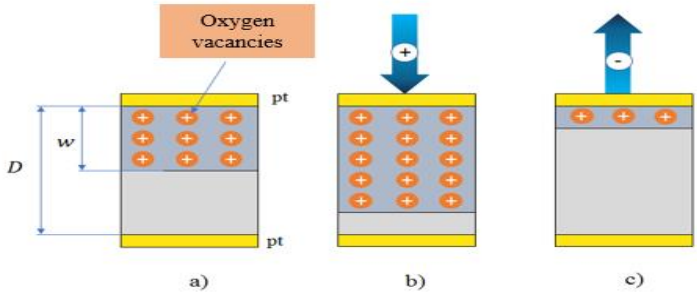


Figure 10. Charge conduction states of the Memristor a) no voltage applied to the Memristor in Pt/Doped TiO_2 /Pure TiO_2 /Pt structure b) when a positive voltage is applied to the top electrode (Doped TiO_2 layer), positively charged oxygen vacancies are repelled and pass to the opposite region c) When a negative potential is applied to the top electrode (Doped TiO_2 layer), oxygen vacancies are attracted by the negative potential and do not pass to the opposite region (Williams, 2008)

1.1. Structure and Electrical Conductivity of TiO_2

TiO_2 (Titania) was first obtained by William Gregor in 1971 from black sand grains in a white colored metal oxide structure. TiO_2 is used as a white color pigment in the paint industry, dielectric material, photo catalyst, ceramic and glass production, aerospace industry, military industry and medicine.

The factors affecting the electrical conductivity of TiO_2 are the density of the crystal structure, the chemical nature of the doped material and the morphological structure of the thin film. The electrical conductivity of TiO_2 is provided by the movement of free electrons given into the crystal lattice by unsaturated Ti atoms on the grain surfaces.

Materials doped into TiO_2 cause defects in the crystal structure, such as points, edges and lines, which affect the electrical conductivity. The scarcity of oxygen atoms in the TiO_2 structure increases the electrical conductivity of TiO_2 as they provide free electrons in the material. The most notable point defects of TiO2 are oxygen vacancies and interstitial titanium atoms (Bally, 1999). These point defects can contribute to the electrical conductivity of TiO_2 in two ways. The point defects can act as mobile charge carriers (ionization) or in response to an electric field, thus creating an ionic current in TiO_2 .

Oxygen vacancies (TiO_{2-x}) in TiO_2 act as electron donors and doping TiO_2 with oxygen vacancies causes TiO_2 to exhibit n-type doped

semiconductor properties. When an oxygen atom is removed from TiO_2, two electrons are left free to move through the material. These electrons effectively change the electrical properties of TiO_2.

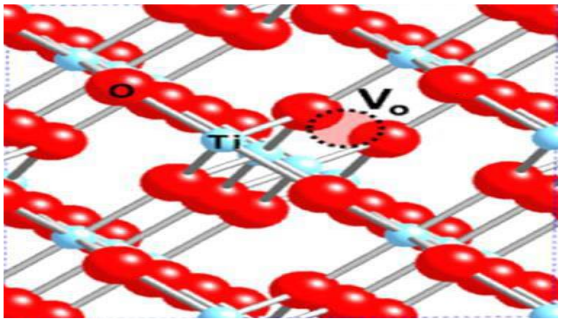


Figure 11. Formation of an oxygen vacancy in TiO_2 (Park 2011)

The vacancy left by the oxygen atom removed from the crystal structure is occupied by Ti atoms. When an oxygen atom is removed from TiO_2, the oxygen vacancies introduce lattice defects in the crystal structure, causing an in-band energy level. These lattice defects form in the vicinity of the Ti atoms closest to the displaced oxygen atom. The electrons liberated by the detachment of the oxygen atom settle in the 3d orbitals of the Ti atom (Park 2011). In the low resistivity state of TiO, the charge conduction path is formed by the doping of oxygen vacancies and the decrease and increase of electrical conductivity is realized by the change of this conduction path.

2. Nonlinear Ion Drift Model

This framework considers the memristor to be voltage-driven, with a nonlinear mapping between voltage, current, and internal-state progression. It further assumes the switching process to be asymmetric. The linear ion drift model captures the hysteresis behavior of the memristor; however, it exhibits limitations in accurately describing fundamental electrodynamics. Experimental studies have demonstrated that the behavior of practical memristors is highly nonlinear, rendering the linear ion drift model insufficient for precise modeling. For applications such as logic circuits, where nonlinear characteristics are essential, more advanced models are required.

Therefore, more appropriate models have been designed. Fundamentally, applying a minor voltage (e.g. 1 V) across a thin film structure (e.g. 10 nm) will result in an strong electric field (e.g. $10^6 V/cm$), which causes substantial nonlinearity in ionic transport. As a result, there is a rapid and important decrease in the energy barrier. This nonlinear behavior manifests itself at the two ends of the element, where the boundary between the two

regions gradually stops. That is to say, once the memristor is set to the R_{ON} or R_{OFF} state, no other external stimulus can undo it. It should be noted that the ion boundary (the boundary between doped and pure TiO_2) does not move to either side of the element (it moves non-linearly). If this happens, it means that there is no physical oxygen vacancy in the element and the length of the doped region is zero, which is meaningless. If the element consisted of only pure or only doped part, it would not exhibit memristor properties but a normal resistor property. Therefore, the state variable must be normalized between 0 < w < D or in the range [0,1].

(Lehtonen and Laiho, 2010) proposed a new model using the results of the study by (Yang et al., 2008). They called this model the linear ion entrainment model.

The equation describing the state variable's dynamics is written as,

$$\frac{dw(t)}{dt} = a.f(w).v(t)^m \tag{9}$$

In the above equation, $f(w)$ denotes the window function, where m is an odd-valued integer, and the model shows a nonlinear relationship with the input voltage.

3. Simmons Tunnel Barrier Model

Previous approaches were grounded in the physical representation of the HP memristor model, an element model between two TiO_2 layers, doped and pure, each of which exhibits a resistivity property. Pickett and friends proposed a more accurate memristor model in 2009 (Pickett et al., 2009). In this model, one resistor and an electron tunnel barrier are connected in series, instead of two resistors connected in series in HP's model. The Simmons tunnel barrier width is the state variable x (Simmons, 1963). That is, the derivative of x gives the oxygen vacancy drift velocity and is given in equation 10.

$$\frac{dx(t)}{dt} = \begin{cases} C_{off}\, sinh\left(\frac{i}{i_{off}}\right) exp\left[-\exp\left(\frac{x-a_{off}}{w_c} - \frac{|i|}{b}\right) - \frac{x}{w_c}\right], i > 0 \\ C_{on}\sinh\left(\frac{i}{i_{on}}\right)\exp\left[-\exp\left(\frac{x-a_{on}}{w_c} - \frac{|i|}{b}\right) - \frac{x}{w_c}\right] i < 0 \end{cases} \tag{10}$$

C_{off}, C_{on}, a a_{off}, a_{on}, i_{off}, i_{on} and b are the experimental adaptation parameters. C_{on} is a larger value than C_{off} and both influence the variation of x. i_{off} and i_{on} are current limiting thresholds. Digital applications benefit from these current thresholds. a_{off} sets the upper limit of x and a_{on} sets the

lower limit of x. Within the defined range, the derivative of the state variable is much smaller than the state variable itself.

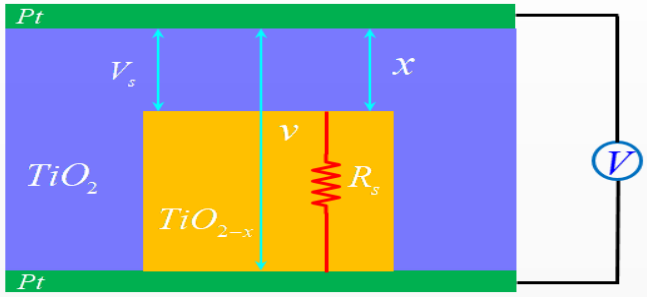


Figure 12. Simmons Tunnel Barrier Model.

x represents the tunnel barrier width, R_s the channel resistance, V the voltage applied to the element, v_g the oxidized zone voltage, v the internal voltage of the element.

4. Threshold Adaptive Memristor Model (TEAM Model)

This model is a simple and general model developed by Kvatinsky and colleagues (Kvatinsky et al., 2012). Despite modeling Identical physical as the Simmons tunnel barrier model, it employs more straightforward expressions. To streamline the analysis and reduce computational complexity, several assumptions are made:

- ✓ The state variable maintains its value unless the threshold is surpassed.
- ✓ Rather than exhibiting an exponential relationship, the memristor current and the state variable are related through a polynomial function..

This model can be adapted to different memristor models (e.g. Simmons tunnel barrier model). By multiplying the function of the current and the function of the state variable depending on x, the Simmons tunnel barrier model can be modeled as in equation 11.

$$\frac{dx(t)}{dt} = \begin{cases} k_{off} \cdot \left(\frac{i(t)}{i_{off}} - 1\right)^{a_{off}} \cdot f_{off}(x), & 0 < i_{off} < i \\ k_{on} \cdot \left(\frac{i(t)}{i_{on}} - 1\right)^{a_{on}} \cdot f_{on}(x) & 0 < i_{off} < i \\ 0, & Other\ cases \end{cases} \quad (11)$$

k_{off}, k_{on}, a_{off}, a_{on} are fixed ($k_{off} > 0$, $k_{on} < 0$), i_{off} and i_{on} are current thresholds, x is the internal state variable, $f_{off}(x)$ and $f_{on}(x)$ are windowing functions where the internal state variable determines the boundaries. The two functions are not required to be identical. Just like in the Simmons tunnel barrier model, the asymmetry depends on x. Assuming that the current-voltage characteristic is similar to equation 12, the memristance

changes linearly depending on x and the relationship between current and voltage becomes as in equation 13.

$$v(t)=\left[R_{ON}+\frac{R_{OFF}-R_{ON}}{x_{off}-x_{on}}(x-x_{on})\right].i(t) \quad (12)$$

$$v(t)=\left[R_{ON}e^{\frac{\lambda}{x_{off}-x_{on}}(x-x_{on})}\right].i(t) \quad (13)$$

In the above equation λ is a fitting parameter. R_{ON} , R_{OFF} are the effective resistances at the boundaries and these resistances are written in terms of the fitting parameter as in equation 14.

$$\frac{R_{ON}}{R_{OFF}}=e^{\lambda} \quad (14)$$

In the section up to here, information about memristors and memristor models has been given. Now let's talk about the usage areas of memristors.

AREAS OF USE OF THE MEMRISTOR

After the HP group experimentally demonstrated the memristor in 2008, there has been a major application research trend in the scientific world for this new element. Attempts in this field can be summarized as using the memristor in a circuit architecture and its applications, adding the memristor to old applications and comparing the results, using new features of the memristor and creating new circuit architectures with them. As a result, in the last nine to ten years, many articles and papers have been published and continue to be published on this topic.

1.1. Analog Circuits

1.1.1. Memristor-Based Sinusoidal Oscillators

Memristor-based sinusoidal oscillators is to investigate the circuit response when memristors replace one or more resistors in classical oscillator topologies. Numerous studies have addressed this concept in the literature. For example, four different Wien bridge oscillators were tested using memristors and as a result, continuous oscillation was observed and an approximate oscillation frequency was obtained.

Figure 13 shows the circuits obtained through the substitution of resistors with memristors in these four different Wien bridge oscillator stages. These models were simulated using PSPICE and Figure 13. a shows the output responses obtained from the simulations and the memristance values for each model. In Figure 13. b, a sustained oscillation is obtained despite changing the memristance value, which is a good example of parametric oscillation reflecting time-dependent oscillation poles.

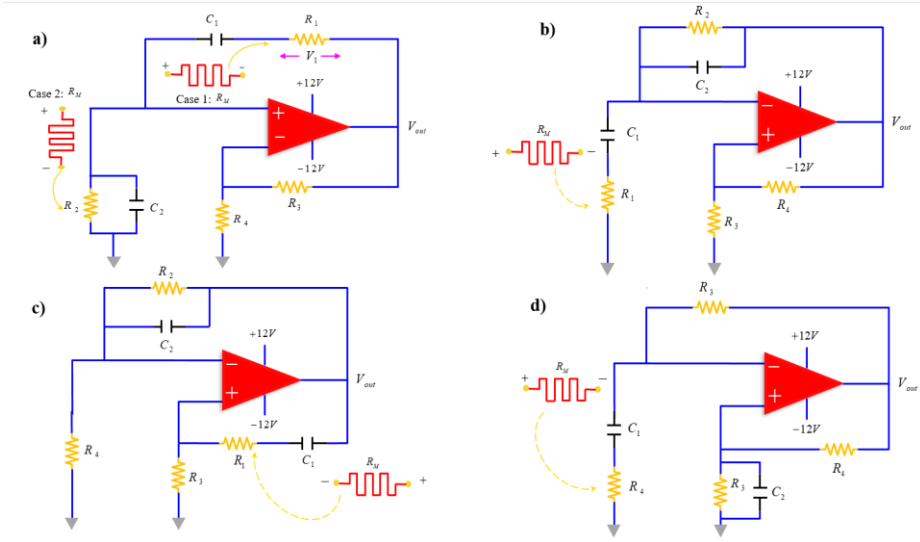


Figure 13. Wien Bridge Oscillator Family a) Type A b) Type B c) Type C d) Type D (Radwan and Fouda, 2015)

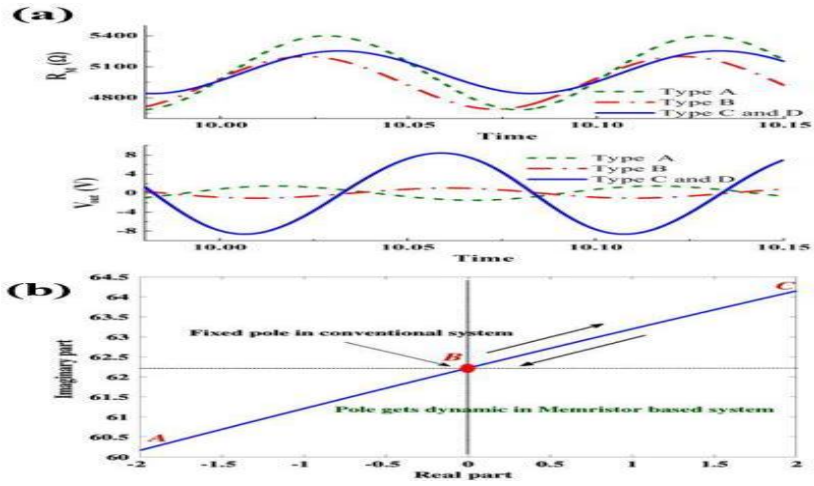


Figure 14. Wien Bridge Oscillator Family a) Type A b) Type B c) Type C d) Type D (Radwan and Fouda, 2015)

1.1.2. Filter Applications

The fact that the memristance does not have a constant value and changes according to the applied energy and the cut-off frequency can be adjusted provides a certain advantage over resistance in the use of memristors in filter applications. The use of memristors in filters was first realized by Driscoll et

al. using the properties of vanadium dioxide (VO2) memristors (Driscoll et al., 2010). In this study, a simple RMLC bandpass filter circuit is used. In this circuit, instead of a resistor, the VO2 memristor is connected in series with the capacitor and coil, thus demonstrating the adaptive filter experimentally. Figure 15 shows the circuit diagram.

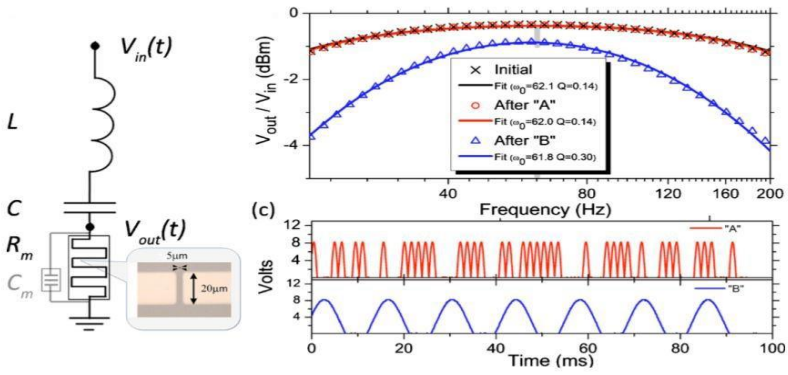


Figure 15. RMLC Bandpass Filter Circuit, Transfer Function (V_{out} / V_{in}), Memristance state versus time (Driscoll et al., 2010)

However, another memristor model has also been used to obtain first and second order Low Pass Filter (LFP) (Chew and Li, 2012). This model consists of zinc oxide (ZnO) nanowires embedded in the copper layer of the printed circuit board. The ZnO model forms a low-pass filter circuit with a capacitor or coil and the memristance value changes according to the voltage applied at the input.

1.1.3. Chaotic Circuits

Another area of use of memristors is chaotic circuits, i.e. chaos applications. Studies in this field are aimed at getting a chaotic response with memristor properties. When the effects of the memristor are considered, the memristor connected to a power supply at certain parameters can show a chaotic effect. Especially in Chua's chaos circuit, the use of memristors has become widespread (Parker and Chua, 1987). Figure 16 shows Chua's chaos circuit.

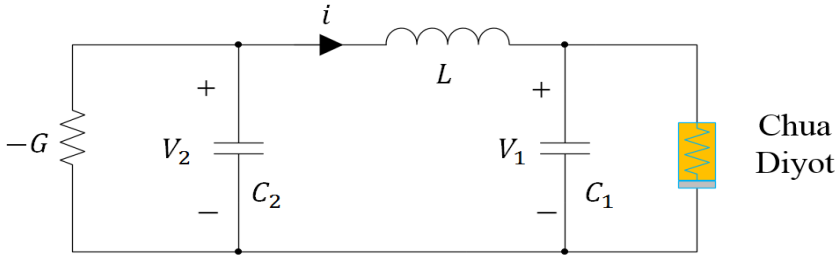


Figure 16. Chua Chaos Oscillator Circuit

In this circuit, a memristor-based chaos circuit was obtained by replacing the diode with a memristor and chaotic behavior was observed in this oscillator (Itoh and Chua, 2008). Figure 17 shows the Chua chaos oscillator circuit designed with memristor and Figure 18 shows the chaos characteristics obtained from this circuit.

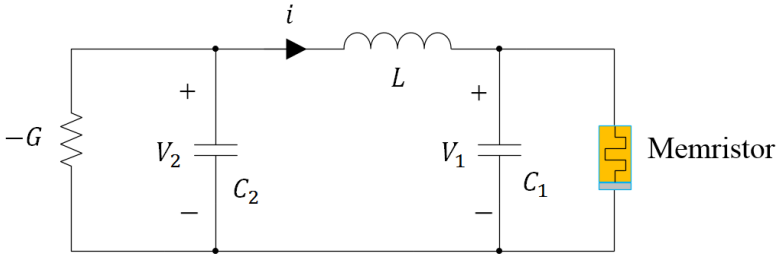


Figure 17. Memristor Chua Chaos Oscillator Circuit

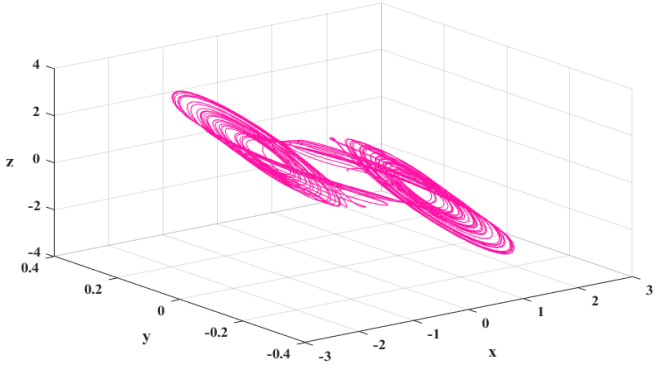


Figure 18. Characteristics of the Memristor Chua Chaos Oscillator Circuit

1.1.4. Non-Volatile Memory

One of the first applications of the memristor, which is closer to reality and may be widely used in the near future, is non-volatile memory. The memristor is a good candidate for NVRAM (Non-Volatile Random Access Memory) applications, especially due to its ability to maintain its final state (memory property) when there is no power on it. So far, many papers have been published in the literature using memristors in memory architecture (Ho et al., 2009; Lee et al., 2011).

1.1.5. Switched Resistor RAM (ReRAM)

ReRAM or switched resistance RAM memory (also referred to as memristor memory in some sources (Chua, 2014)) is already a research topic for many companies worldwide (HP, IMEC, Fujitsu, Sharp and Unity Semiconductor). The operating logic of this type of memory is based on logic values "1" and "0". The memristor takes the value "1" in the ON position, i.e. low resistance value, and "0" in the OFF position, i.e. high resistance value (Kavehei, 2011). Various materials have been used as transition metal oxides in such structures. Switching resistance structure was first discovered in nickel oxide (NiO) in 1964 (Gibbons and Beadle, 1964). Other materials showing switched resistance properties are Titanium dioxide (TiO_2), Strontium Titanate ($SrTiO_3$), Zirconia (ZrO_2) and Hafnia (HfO_2). Two switching mechanisms, unipolar and bipolar, were observed in these materials. The state of the element can change by applying a potential on it. For bipolar switching devices, a negative voltage must be applied to return to its previous state. In the case of unipolar switching, the switching voltage must be higher than two threshold voltages to change the state of the element (Kavehei et al., 2010).

1.1.6. Hybrid chip

HP has implemented a crossbar architecture in a hybrid chip that integrates both CMOS transistors and memristors. In this design, one layer of CMOS transistors can effectively drive multiple layers—potentially thousands—of memristors. This architecture significantly enhances energy efficiency and reduces heat generation. Furthermore, the high density of crossbar interconnections enables a greater number of logic gates per unit area. The switching operations occur within nanoseconds, with each switch requiring only picojoules of energy, making the system highly suitable for low-power, high-speed computing applications.

The first CMOS–memristor reconfigurable hybrid, developed by Q. Xia et al. in 2009, is shown in Figure 19. The schematic diagram illustrating the hybrid circuit's integration of CMOS and memristor switching layers—and the TEM image of this hybrid chip—are also presented in Figure 19

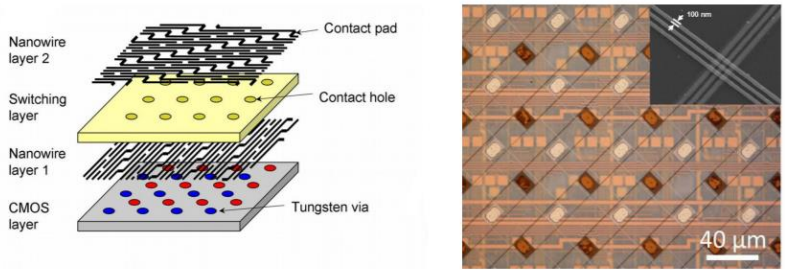


Figure 19. First produced CMOS/memristor reconfigurable hybrid chip

1.1.7. Neuromorphic Systems

"I realized that synapses are memristors. The ion channel was the missing link I was looking for and it already exists in nature." According to Leon Chua, the memristor's functionality closely resembles that of a biological synapse. Drawing upon the Nobel Prize-winning Hodgkin-Huxley neuron model (1961), he explained that the sodium and potassium channels in an axon possess time-dependent conductance—a fundamental example of memristive behavior. Figure 20 shows the Hodgkin-Huxley neuron model.

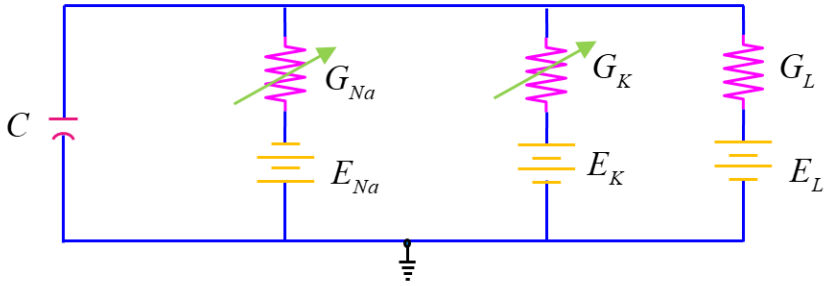


Figure 20. Characteristics of the Memristor Chua Chaos Oscillator Circuit

Within each synapse, a complex array of signal processing takes place, and this is what neuromorphic architectures (or brain simulation systems) seek to achieve. While current microprocessors have a density of about 10^9 per single square centimeter, human synapses have 10 times more density (10^{10}

per cm^2). To model a single synapse, it is necessary to know what will be inside the synapse (Versace and Chandler, 2010).

1.1.8. Crossbar

The cross-connect structure is one of the most promising nanostructures at the architectural level. The cross-connect is fault-tolerant due to the large number of switches. Its nanoscale nature allows for cost-effective and straightforward fabrication. Crossbar networks can support extensive computational tasks.

As discussed in (Lewis and Lee, 2000), very high capacity with minimal power consumption of dimensionally stacked structures is their major advantage over DRAM.

Addressing complexity grows logarithmically as capacity increases. Thus, at extremely high capacities, the overall addressing cost is less. The structure of the crossbar nanowire display structure made at HP Laboratories with a memristor switch at each junction is shown in Figure 21.

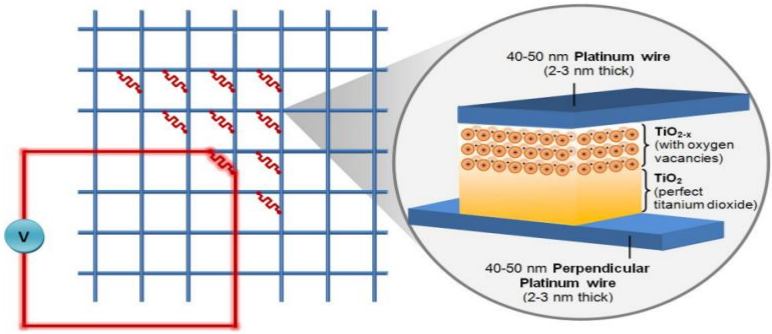


Figure 21. A schematic diagram illustrating the crossbar structure's interconnections (adapted by R. Williams.,2008.)

CONCLUSION

Research on memristors and memristive systems is a relatively innovative and emerging area, offering numerous opportunities for further investigation. these areas:, Memristor-Based Receiver for Ultra-Wideband Wireless Systems, Various Physical Information Storage Mechanisms, Programmable precision resolution resistors, Electromagnetic Transistor/Memristor Based on Carbon Nanotubes, Digital Logic Structures with Memristor Based Cross Connection Structures and one of the main ones is Artificial Intelligence systems, which already have very good learning capabilities. Nowadays, AI robots can learn very quickly by watching, analyzing and so on. Memristors, on the other hand, have the potential to accelerate this learning process almost as fast as that of a human.

REFERENCE

Aytekin, S. A., & Mutlu, R. (2013). Taylor serisi ve parametrik fonksiyonlar kullanarak yük-bağımlı memindüktör histeresiz eğrisinin açıklanması. In Otomatik Kontrol Ulusal Toplantısı (TOK 2013) (pp. 705–709).

Bally, A. (1999). Electronic properties of nano-crystalline titanium dioxide thin film (Ph.D. thesis, École polytechnique fédérale de Lausanne, Lausanne).

Chakrabarti, B., Hellenbrand, D. J., Fantner, G. E., & Sheldon, B. W. (2017). A 3D memristor model and its application in analog and neuromorphic computing. Scientific Reports, 7, 42494. https://doi.org/10.1038/srep42494

Chanthbouala, A., Garcia, V., Cherifi, R. O., Bouzehouane, K., Fusil, S., Moya, X., Allibe, J., Dlubak, B., Grollier, J., Xavier, S., Barra, A.-L., Fusil, S., Bibes, M., & Barthélémy, A. (2012). A ferroelectric memristor. Nature Materials, 11(10), 860–864. https://doi.org/10.1038/nmat3370

Chew, W. C., & Li, Y. (2012). Memristor-based low pass filter. 2012 IEEE International Symposium on Circuits and Systems (ISCAS), 726–729. https://doi.org/10.1109/ISCAS.2012.6272067

Chua, L. O. (1971). Memristor—The missing circuit element. IEEE Transactions on Circuit Theory, 18(5), 507–519. https://doi.org/10.1109/TCT.1971.1083337

Chua, L. O. (2014). Memristors and memristive systems (R. Tetzlaff, Ed.). Springer. https://doi.org/10.1007/978-1-4614-9068-5

Chua, L. O., & Kang, S. M. (1976). Memristive devices and systems. Proceedings of the IEEE, 64(2), 209–223. https://doi.org/10.1109/PROC.1976.10092

Driscoll, T., Kim, H. T., Kim, B. G., Lee, Y., Chae, B. G., Lee, S. R., & Basov, D. N. (2010). Memristive adaptive filters. Applied Physics Letters, 97(9), 093502–093503. https://doi.org/10.1063/1.3486471

Elgabra, H., Farhat, I. A., Al Hosani, A. S., Homouz, D., & Mohammad, B. (2012, March). Mathematical modeling of a memristor device. In 2012 International Conference on Innovations in Information Technology (IIT)

(pp. 156–161). IEEE. https://doi.org/10.1109/INNOVATIONS.2012.6182209

Gibbons, R. C., & Beadle, J. R. (1964). The Hall effect and electrical conductivity of titanium dioxide. Journal of Physics and Chemistry of Solids, 25(10), 1113–1121. https://doi.org/10.1016/0022-3697(64)90125-6

Ho, C., Huang, P., & Li, Y. (2009). SPICE model of memristor with nonlinear dopant drift. International Symposium on Circuits and Systems (ISCAS 2009), 2370–2373. https://doi.org/10.1109/ISCAS.2009.5118241

Itoh, M., & Chua, L. O. (2008). Memristor oscillators. International Journal of Bifurcation and Chaos, 18(11), 3183–3206. https://doi.org/10.1142/S021812740802236X

Jo, S. H., Chang, T., Ebong, I., Bhadviya, B. B., Mazumder, P., & Lu, W. (2010). Nanoscale memristor device as synapse in neuromorphic systems. Nano Letters, 10(4), 1297–1301. https://doi.org/10.1021/nl903892c

Kavehei, O., Iqbal, A., Al-Sarawi, S. F., Abbott, D., & Ranasinghe, D. C. (2010). The fourth element: Characteristics, modeling, and electromagnetic theory of the memristor. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 466(2120), 2175–2202. https://doi.org/10.1098/rspa.2010.0015

Kvatinsky, S., Friedman, E. G., Kolodny, A., & Weiser, U. C. (2013). The desired memristor for circuit designers. IEEE Circuits and Systems Magazine, 13(2), 17–22. https://doi.org/10.1109/MCAS.2013.2256271

Kvatinsky, S., Friedman, E. G., Kolodny, A., & Weiser, U. C. (2012). TEAM: ThrEshold adaptive memristor model. IEEE Transactions on Circuits and Systems I: Regular Papers, 60(1), 211–221. https://doi.org/10.1109/TCSI.2012.2226442

Lee, M. J., Park, J. S., Kim, G. H., Seo, S., & Hwang, H. (2011). Resistance switching memory using metal oxides. Proceedings of the IEEE, 98(12), 2237–2251. https://doi.org/10.1109/JPROC.2010.2095046

Lehtonen, E., & Laiho, M. (2010). CNN using memristors for neighborhood connections. 2010 International Workshop on Cellular Nanoscale Networks and Their Applications (CNNA), 1–4. https://doi.org/10.1109/CNNA.2010.5565820

Lewis, D. L., & Lee, H.-H. S. (2000). Architectural evaluation of 3D stacked RRAM caches. In Proceedings of the IEEE International Conference on 3D System Integration (3DIC) (pp. 1–4).

Park, J. H. (2011). Formation and effects of oxygen vacancies in TiO_2 nanostructures. Journal of Materials Chemistry, 21(23), 8462–8470. https://doi.org/10.1039/c1jm10292g

Parker, V. B., & Chua, L. O. (1987). Practical numerical algorithms for chaotic circuits. International Journal of Circuit Theory and Applications, 15(3), 205–219. https://doi.org/10.1002/cta.4490150304

Pershin, Y. V., & Di Ventra, M. (2012). Memristor emulator for memristor circuit applications. IEEE Transactions on Circuits and Systems I: Regular Papers, 59(10), 2365–2371. https://doi.org/10.1109/TCSI.2012.2201997

Pershin, Y. V., Penner, R., & Di Ventra, M. (2012). Memristive circuits for analog-to-digital and digital-to-analog conversion. IEEE Transactions on Circuits and Systems I: Regular Papers, 59(10), 2213–2220. https://doi.org/10.1109/TCSI.2012.2202011

Pickett, M. D., Strukov, D. B., Borghetti, J. L., Yang, J. J., Snider, G. S., Stewart, D. R., & Williams, R. S. (2009). Switching dynamics in a titanium dioxide memristive device. Journal of Applied Physics, 106(7), 074508. https://doi.org/10.1063/1.3236506

Radwan, A. G., & Fouda, M. E. (2015). On the mathematical modeling of memristor, memcapacitor, and meminductor devices. Springer. https://doi.org/10.1007/978-3-319-08672-0

Simmons, J. G. (1963). Generalized formula for the electric tunnel effect between similar electrodes separated by a thin insulating film. Journal of Applied Physics, 34(6), 1793–1803. https://doi.org/10.1063/1.1702682

Singh, H., & Kakade, S. V. (2014). Memristors MATLAB modeling. International Journal of Scientific & Engineering Research, 5(5), 1422–1425.

Strachan, J. P., Pickett, M. D., Yang, J. J., Aloni, S., Kilcoyne, A. D., Medeiros-Ribeiro, G., & Williams, R. S. (2010). Direct identification of the conducting channels in a functioning memristive device. Advanced Materials, 22(32), 3573–3577. https://doi.org/10.1002/adma.201002777

Strukov, D. B., Snider, G. S., Stewart, D. R., & Williams, R. S. (2008). The missing memristor found. Nature, 453(7191), 80–83. https://doi.org/10.1038/nature06932

Versace, M., & Chandler, B. (2010). The brain of a new machine. IEEE Spectrum, 47(12), 30–37. https://doi.org/10.1109/MSPEC.2010.5628656

Vourkas, I., Abusleme, A., Rubio, L., & Sirakoulis, G. C. (2015). SPICE modeling of nonlinear memristive behavior. Integration, 50, 30–41. https://doi.org/10.1016/j.vlsi.2015.01.002

Wang, X., Dong, Z., & Yu, H. (2009). A PWL model of memristor and its application example. 2009 International Symposium on Information Science and Engineering, 1, 345–348. https://doi.org/10.1109/ISISE.2009.116

Williams, R. S. (2008). How we found the missing memristor. IEEE Spectrum, 45(12), 28–35. https://doi.org/10.1109/MSPEC.2008.4721238

Xia, Q., Yang, J. J., Kim, K.-H., & Lu, W. (2009). Memristor-CMOS hybrid integrated circuits for reconfigurable logic. Nano Letters, 9(10), 3640–3645. https://doi.org/10.1021/nl902052g

Xiao-Bo, Z., Wen, H., & Xu, J. (2013). Design of a programmable analog filter based on memristor. 2013 IEEE 11th International Conference on Solid-State and Integrated Circuit Technology (ICSICT), 1–3. https://doi.org/10.1109/ICSICT.2012.6467752

Yang, J. J., Pickett, M. D., Li, X., Ohlberg, D. A. A., Stewart, D. R., & Williams, R. S. (2008). Memristive switching mechanism for metal/oxide/metal nanodevices. Nature Nanotechnology, 3(7), 429–433. https://doi.org/10.1038/nnano.2008.160

Exploring the Efficiency of the Fenton Oxidation Process for the Removal of Pyridaben from Agricultural Wastewater

Burcu ŞİMŞEK UYGUN[1*]

Yağmur KAYA[2]

1* Asisst.Prof.Dr. Burcu ŞİMŞEK UYGUN; Eskişehir Technical University, Engineering Faculty, Departmant of Environmental Engineering. bsimsek1@esksiehir.edu.tr ORCID No: 0000-0003-1211-4198
2 Yağmur KAYA; Eskişehir Technical University, Engineering Faculty, Departmant of Environmental Engineering, kaya50yagmur@gmail.com ORCID No: 0009-0007-5015-9895

ABSTRACT

Agricultural activities represent one of the most critical sources of global nutrition. However, the increasing demand for food due to rapid population growth has led to the widespread use of pesticides to protect crops from harmful weeds, pests, and insects. While pesticides are effective in enhancing crop productivity, their uncontrolled use poses significant toxic threats to both human health and the environment. Pesticide residues entering the soil, air, and especially water resources necessitate the development and application of effective treatment methods.

This study evaluates the performance of the Fenton process, an advanced oxidation process (AOP), in treating wastewater containing pesticide contaminants. For this purpose, synthetic wastewater was prepared using Pyridaben as the active ingredient. The influence of hydrogen peroxide (H_2O_2) concentration, ferrous sulfate ($FeSO_4$) dosage, and pH on Chemical Oxygen Demand (COD) removal was investigated.

The experiments were designed using the Box–Behnken approach within the framework of Response Surface Methodology (RSM), employing Design Expert 13 software. The relationships between operating parameters and COD removal efficiency were assessed through three-dimensional response surface plots and contour diagrams.

Under optimized conditions—pH 1, $FeSO_4$ concentration of 8000 mg/L, and H_2O_2 concentration of 2000 mg/L—a maximum COD removal efficiency of about 87% was obtained. These findings indicate that the Fenton process offers an efficient and promising option for the treatment of pesticide-polluted agricultural wastewater.

Keywords: Fenton Oxidation, Advanced Oxidation Technologies, Wastewater Treatment, Organic Pollutant Removal, Pesticide-Contaminated Water

INTRODUCTION

Water, air, and soil pollution have emerged as some of the most critical environmental challenges both worldwide and in our country. The rapid pace of industrialization, advancements in technology, intensified agricultural practices, and the steady growth of the global population have all played a major role in accelerating environmental degradation. Among the leading causes of soil and water contamination are the chemical agents used in agricultural production. Notably, pesticides—synthetic organic compounds widely applied to improve crop quality and yield—represent one of the most significant contributors to this pollution.

Pesticides are widely employed in agricultural pest management because of their ease of application and rapid efficacy against target organisms. While their primary purpose is to control harmful insects, weeds, and plant pathogens, extensive research has demonstrated that prolonged and excessive use of these chemicals can have serious consequences for both environmental and human health. Pesticides may persist in the environment, infiltrate groundwater and surface waters, and disrupt ecological balance, thereby posing long-term ecological and public health risks. (Zhou et al., 2024).

To minimize these impacts, pesticides should be applied in carefully controlled and judicious doses. Moreover, once they enter aquatic environments, their prompt and efficient removal is essential to prevent further contamination of water resources (Suwanaruang, 2024). A variety of treatment technologies have been developed for the remediation of pesticide-polluted waters, among which the Fenton process—an advanced oxidation process (AOP)—has attracted considerable attention for its high degradation efficiency (Ahmad et al., 2024).

The Fenton process is extensively applied for treating water contaminated with organic pollutants, including pesticides, owing to its notable advantages: relatively low cost, operational simplicity, and high efficiency in degrading toxic and non-biodegradable compounds into less harmful forms. This process operates through the generation of highly reactive hydroxyl radicals (•OH) via the reaction between hydrogen peroxide (H_2O_2) and ferrous ions (Fe^{2+}). These radicals subsequently oxidize and decompose organic contaminants present in the water. (Ziembowicz & Kida, 2022).

Using the Box-Behnken Design (BBD) under the Response Surface Methodology (RSM) framework, experiments were modeled and optimized. The results of the study demonstrated that the Fenton process is a promising method for the removal of pesticide residues from aqueous media. The optimized conditions led to high COD removal efficiency, confirming that this method can be effectively applied to reduce the environmental risks associated with pesticide-contaminated agricultural wastewater.

In a related study, Meijide et al. (2016) examined the removal of Thiamethoxam, a second-generation pesticide from the neonicotinoid class,

using three advanced treatment processes: Fenton, anodic oxidation, and electro-Fenton. Their results indicated relatively modest removal efficiencies of 35.24% and 41.75% for the Fenton and anodic oxidation processes, respectively. However, a significantly higher removal efficiency of 92.44% was achieved using the electro-Fenton process, which employed a BBD anode and a carbon felt cathode (Meijide et al., 2016).

Masomboon et al. (2011) conducted a study on the degradation of 2,6-dimethylaniline, a compound used in pesticide production, using the Fenton process. In their experiments, the initial concentration ranges were set as 0.5–5 mM for 2,6-dimethylaniline, 0.25–2 mM for Fe(II), and 5–100 mM for H_2O_2, with the reactions performed at pH 2. The study reported that the degradation rate of 2,6-dimethylaniline decreased with increasing initial pollutant concentration, while it increased with higher initial concentrations of Fe(II) and H_2O_2 (Masomboon et al., 2011).

Moreover, the researchers observed that the Fenton reaction proceeded in two distinct stages. The degradation occurring in the first stage—within the initial 10 minutes—was found to be significantly faster than that of the second stage, indicating a rapid initial breakdown of the compound followed by a slower secondary degradation phase.

Cheng et al. (2015) investigated the application of the Fenton process for the treatment of pesticide-contaminated wastewater. Their study demonstrated that the pesticides dimethoate, triazophos, and malathion were completely degraded through the Fenton reaction, and a COD removal efficiency of nearly 80% was achieved (Cheng et al., 2015).

In a study, At natural pH, the photo-Fenton process was evaluated for irrigation water reuse using five organic fertilizers (DTPA, EDDHA, HEDTA, EDTA, EDDS) as iron-chelating agents. The stability constant of the iron–chelate complexes, determined by the chelator's molecular structure, was found to be a key factor in micropollutant removal efficiency. To reduce excessive iron loss and enhance degradation rates, chelator mixtures were tested. A 50% EDDS + 50% EDTA mixture achieved complete removal of propranolol within 30 minutes, whereas EDTA alone required 90 minutes and EDDS alone did not achieve full removal. Furthermore, the dissolved iron concentration after treatment with the mixture was 5.5 times higher than with EDDS alone, improving the treated water's suitability for irrigation reuse. (López-Vinent et al., 2021).

In another study, 2,4-Dichlorophenoxyacetic acid (2,4-DPAA), a common herbicide, contaminates groundwater and surface water. This study developed an electro-Fenton (EF) process to remove 2,4-DPAA from agricultural wastewater, optimized using response surface methodology (RSM). The highest removal efficiency of 99.2% was achieved at pH 3, 2.6 mg L^{-1} initial concentration, 470 μL L^{-1} H_2O_2, 3.5 mA cm^{-2} current density, and 7.5 min reaction time, closely matching the predicted 98.6%. Application on real wastewater showed 93.7% removal. Model performance was strong (R^2 =

0.9775), with ANOVA confirming significance ($F = 114.14$, $p < 0.0001$). Energy consumption was low (0.058 kWh m^{-3}). Mechanistic studies indicated hydroxyl radicals as the main degraders. The reaction followed pseudo-first-order kinetics with a rate constant of 0.4746 min^{-1} (Dolatabadi et al., 2021).
In this study, the Fenton process was applied to treat synthetic wastewater polluted with pesticide compounds. The aim was to assess the treatment efficiency and to determine the optimal operational conditions through a systematic experimental design. Parameters such as pH, hydrogen peroxide concentration, and ferrous sulfate dosage were varied and analyzed for their effect on Chemical Oxygen Demand (COD) removal efficiency.

MATERIAL METHODS

Pyridaben

Pyridaben is a commonly used acaricide and insecticide in agriculture, primarily employed to control mites and other pests on a variety of crops. Owing to its widespread application, residues of Pyridaben often reach the environment, particularly via agricultural runoff, resulting in the contamination of surface and groundwater resources. Its high chemical stability and low biodegradability contribute to its persistence in aquatic ecosystems, raising concerns over potential toxic effects on non-target species, including aquatic invertebrates and fish (Chen et al., 2022). Moreover, Pyridaben's accumulation in water bodies can disrupt aquatic ecosystems and may pose risks to human health if contaminated water is used for irrigation or drinking purposes. Given these environmental challenges, effective treatment of Pyridaben-contaminated wastewater is critical. Conventional wastewater treatment methods often fail to completely remove such persistent pesticides, necessitating advanced treatment technologies. Among these, the Fenton process and other advanced oxidation processes have shown promising results in degrading Pyridaben molecules by generating reactive hydroxyl radicals that break down the pesticide into less harmful substances. Ongoing research aims to optimize these processes to achieve high removal efficiencies while minimizing operational costs and environmental footprints (Rand & Clark, 2000). Understanding the behavior, fate, and treatment of Pyridaben in wastewater systems is essential for safeguarding water quality and ensuring sustainable agricultural practices.
Pyridaben is a non-systemic acaricide with the chemical formula $C_{19}H_{25}ClN_2OS$. It typically appears as a white to beige powder or fine crystalline form and possesses a slight characteristic odor. This pesticide is known for its rapid knockdown effect on target pests. The molecular weight of Pyridaben is approximately 364.93 g/mol. It exhibits greater solubility in organic solvents such as acetone and xylene compared to water, reflecting its hydrophobic nature.

Design of Experiment

In classical experimental design, the effect of each parameter is typically assessed by varying one factor at a time while keeping all others constant. This "one-factor-at-a-time" approach allows researchers to evaluate the isolated impact of individual parameters and determine their optimal levels. However, it has significant limitations when multiple factors interact, as it cannot effectively capture combined or interactive effects across the full range of variables. Moreover, conducting separate experiments for each parameter setting can result in a large number of trials, making the process time-consuming, resource-intensive, and costly.

To address these limitations, advanced experimental design methodologies have been developed. These approaches enhance experimental efficiency by systematically examining the relationships among multiple factors simultaneously, allowing for the identification of interactive effects on the response variable. This not only reduces the number of experiments required but also facilitates the determination of true optimal conditions under complex parameter interactions.

Numerical modeling of these sophisticated designs typically involves extensive calculations and statistical analyses that are difficult and impractical to perform manually, especially for multifactorial experiments. To facilitate this process, specialized software tools have been developed. One of the leading programs in this domain is Design Expert, which incorporates a wide range of experimental design models, including factorial designs, response surface methodology, and mixture designs. This software streamlines the design and analysis phases by allowing users to input parameter ranges, automatically generate efficient experimental plans, perform regression and variance analyses, and evaluate interactions among variables. Furthermore, it aids in selecting the most significant parameters influencing the outcome, thereby guiding researchers towards more precise and effective experimental conclusions (Janković et al., 2025).

Such tools are invaluable in modern research and industrial applications, where optimization of processes, reduction of costs, and efficient use of resources are critical. By integrating statistical rigor with computational power, experimental design software like Design Expert enhances the reliability and reproducibility of experimental results, supporting data-driven decision-making and accelerating innovation.

Method

Experiments evaluating the treatment of pesticide-contaminated wastewater using the Fenton process were conducted, with the assessment based on Chemical Oxygen Demand (COD) analysis. The experimental design was

developed using Design Expert 13 statistical software, employing the Response Surface Methodology (RSM) and a Box-Behnken Design to structure the trials. The independent variables investigated included the concentration of $FeSO_4$ (ranging from 4000 to 8000 mg/L), H_2O_2 concentration (1500 to 2500 mg/L), and pH levels varying between 1 and 5.
In the experimental procedure, the jar test method was initially employed as a preliminary step to evaluate the coagulation and flocculation performance under varying treatment conditions. This approach involved systematically mixing wastewater samples with different dosages of reagents to optimize the removal of suspended solids and pollutants. The jar test allowed for rapid assessment of process parameters such as mixing speed, settling time, and reagent concentrations, facilitating the identification of optimal conditions for effective treatment.
Following the jar test, the treated samples were subjected to Chemical Oxygen Demand (COD) analysis to quantitatively assess the reduction of organic pollutants in the wastewater. COD measurement served as a critical indicator of treatment efficiency, reflecting the decrease in oxidizable organic matter as a result of the applied treatment. The combination of jar test optimization and subsequent COD analysis provided a comprehensive evaluation framework to determine the effectiveness of the treatment process, enabling data-driven decisions for process refinement and scaling. Within the scope of the study, the Chemical Oxygen Demand (COD) analysis was performed using the standard method 5220 C: Closed Reflux Titrimetric Method. To prepare the samples for analysis, the powder form of the Pyridaben active pesticide ingredient was accurately weighed and dissolved to create wastewater solutions with a concentration of 10,000 mg/L.

RESULTS AND DISCUSSION

In the conducted experiments, the effects of $FeSO_4$ concentration, H_2O_2 concentration, and initial pH on the chemical oxygen demand (COD) removal efficiency were systematically investigated. The experimental results revealed that variations in these parameters had a significant influence on the treatment performance. Specifically, changes in $FeSO_4$ and H_2O_2 dosages directly affected the generation of hydroxyl radicals, thereby altering the degradation efficiency of organic pollutants. Similarly, the initial pH played a critical role in optimizing the Fenton reaction conditions. The COD removal efficiencies obtained from all experimental runs were recorded and subsequently analyzed using the Design Expert software. The software-based response surface methodology enabled the determination of the optimal operational conditions, under which the maximum COD removal efficiency was achieved.

Effect of $FeSO_4$ and H_2O_2

According to the graphical results, an increase in $FeSO_4$ concentration consistently enhanced the COD removal efficiency, highlighting the positive role of iron dosage in the Fenton process (Figure 1). In contrast, the effect of H_2O_2 concentration was found to be non-linear: while increasing H_2O_2 up to a certain level improved the removal efficiency, further increases beyond this threshold led to a decline in performance, possibly due to the scavenging effect of excessive hydrogen peroxide. The highest removal efficiency observed in the contour plot was 83%, which was clearly marked on the graph.

A comparison of experimental conditions with identical H_2O_2 concentrations and pH values showed that higher $FeSO_4$ dosages resulted in greater COD removal efficiencies, reinforcing the conclusion that iron dosage is a key factor in enhancing the performance of the Fenton process under the tested conditions (Figure 1.)

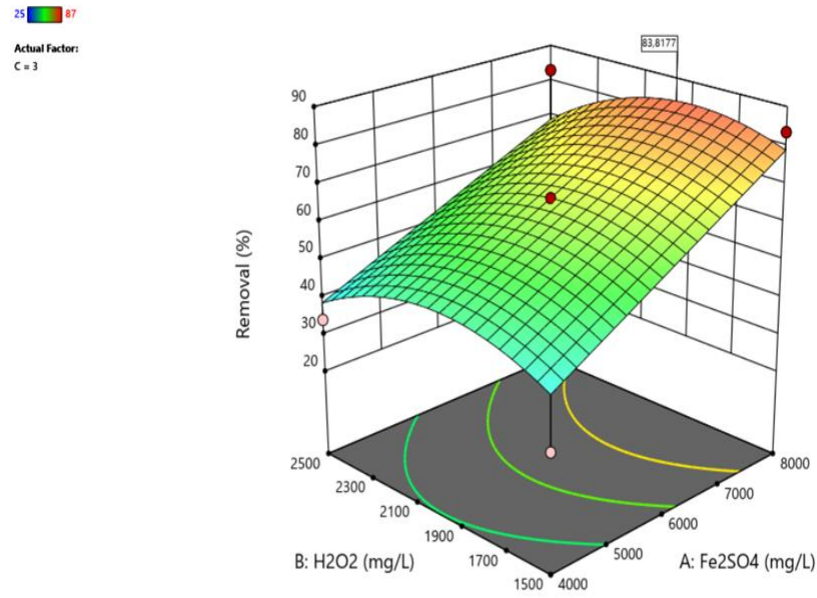


Figure 1. 3D Plot showing the effect of $FeSO_4$ and H_2O_2 concentrations on COD removal

Effect of $FeSO_4$, H_2O_2 and pH

As observed in the contour plot, increasing the $FeSO_4$ concentration led to a consistent improvement in COD removal efficiency, confirming the strong

positive influence of iron dosage in the Fenton process. In contrast, the effect of pH on the removal efficiency could not be clearly interpreted in isolation, as its influence appears to be closely dependent on the concentrations of the other reagents. Nevertheless, it is noteworthy that the highest removal efficiency achieved in this study corresponded to the lowest pH value tested. In the three-dimensional surface plot, a similar trend was observed, where higher $FeSO_4$ concentrations combined with a pH value of 1 resulted in a marked increase in COD removal efficiency. The plot clearly illustrates that the regions corresponding to the highest removal efficiencies are represented in red, while the lowest efficiencies are shown in light blue (Figure 2). This visual distribution further supports the conclusion that acidic conditions, particularly at pH 1, and elevated iron dosages provide favorable conditions for the Fenton process, thereby maximizing treatment performance.

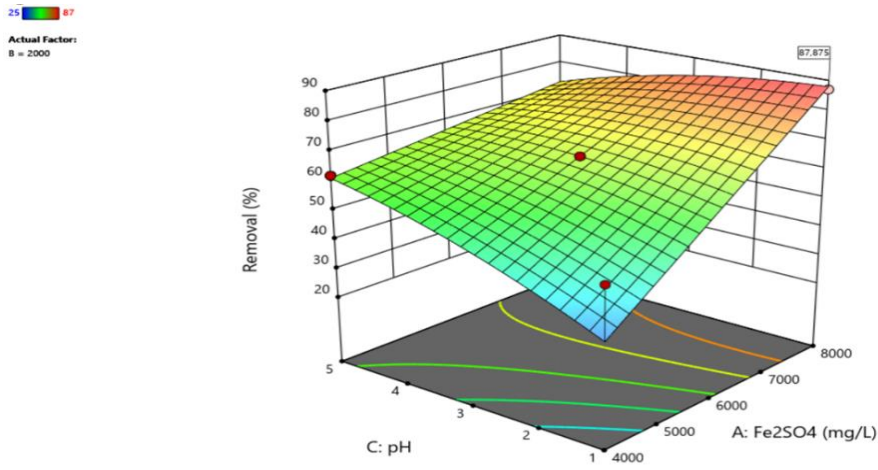


Figure 2. 3D Plot showing the effect of $FeSO_4$ concentration and pH on COD removal efficiency

In the three-dimensional surface plot, it can be observed that at pH 1, the COD removal efficiency remains high regardless of the hydrogen peroxide concentration (Figure 3). As the pH increases, a gradual decline in removal efficiency is evident, reflecting the reduced effectiveness of the Fenton reaction under less acidic conditions. The maximum efficiency was achieved at pH 1 with a hydrogen peroxide concentration of 2000 mg/L, which is represented by the darkest red region on the plot. This visualization further emphasizes the critical role of maintaining strongly acidic conditions to optimize hydroxyl radical generation and enhance pollutant degradation in the Fenton process (Figure 3).

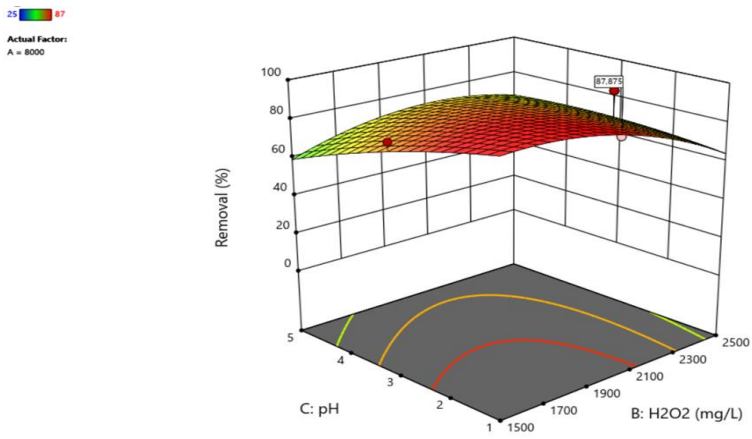


Figure 3. 3D Surface plot showing the effect of H_2O_2 concentration and pH on COD removal efficiency

Finally, the optimum values determined within the study were an $FeSO_4$ concentration of 8000 mg/L, an H_2O_2 concentration of 2000 mg/L, and a pH of 1. Under these conditions, the maximum COD removal efficiency was found to be 87% (Figure 4).

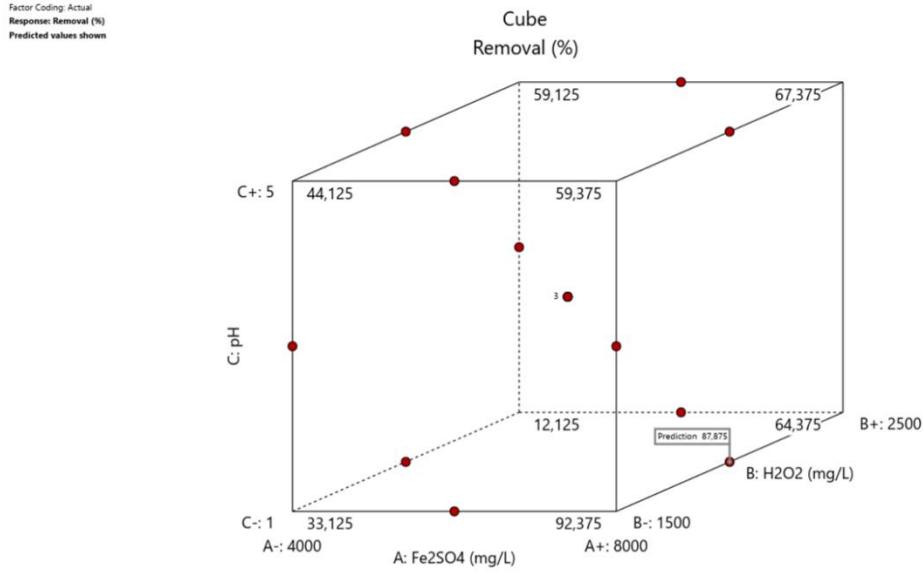


Figure 4. 3D cube diagram showing the interactive impact of $FeSO_4$ concentration, H_2O_2 dosage, and pH on COD degradation

In this study, synthetic wastewater samples contaminated with Pyridaben, an acaricide-type pesticide known for its persistence and toxic effects on living

organisms and environmental health, were treated using the Fenton process. Due to Pyridaben's high toxicity and resistance to conventional treatment methods, advanced oxidation via the Fenton reaction was employed to effectively degrade the pollutant. Following the treatment, Chemical Oxygen Demand (COD) analyses were conducted to quantify the organic load remaining in the samples, and removal efficiencies were calculated based on these measurements. The obtained COD removal efficiencies were thoroughly evaluated, and key operational parameters influencing the Fenton process—namely pH, hydrogen peroxide (H_2O_2) concentration, and ferrous sulfate ($FeSO_4$) dosage—were systematically investigated. By analyzing the effects of these parameters, the optimal treatment conditions that maximize pollutant degradation and removal efficiency were determined. This comprehensive approach provides valuable insights into the effective treatment of Pyridaben-containing wastewater using the Fenton process.

Overall, the findings of this study demonstrate that the Fenton process is an efficient and promising method for the removal of persistent and toxic pesticides like Pyridaben from wastewater. By optimizing key reaction parameters, significant improvements in COD removal were achieved, highlighting the potential of this advanced oxidation technique for practical applications in wastewater treatment. These results contribute to the development of more sustainable and effective strategies for mitigating the environmental impact of hazardous agrochemicals, ultimately supporting the protection of water resources and public health.

REFERENCES

Ahmad, M. F., Ahmad, F. A., Alsayegh, A. A., Zeyaullah, M., AlShahrani, A. M., Muzammil, K., Saati, A. A., Wahab, S., Elbendary, E. Y., Kambal, N., Abdelrahman, M. H., & Hussain, S. (2024). Pesticides impacts on human health and the environment with their mechanisms of action and possible countermeasures. *Heliyon*, *10*(7), e29128. https://doi.org/https://doi.org/10.1016/j.heliyon.2024.e29128

Chen, L., Pan, M., & Hu, D. (2022). An overview on the green synthesis and removal methods of pyridaben. *Frontiers in Chemistry*, *10*, 975491. https://doi.org/10.3389/fchem.2022.975491

Cheng, G., Lin, J., Lu, J., Zhao, X., Cai, Z., & Fu, J. (2015). Advanced Treatment of Pesticide-Containing Wastewater Using Fenton Reagent Enhanced by Microwave Electrodeless Ultraviolet. *Biomed Res Int*, *2015*, 205903. https://doi.org/10.1155/2015/205903

Dolatabadi, M., Ghaneian, M. T., Wang, C., & Ahmadzadeh, S. (2021). Electro-Fenton approach for highly efficient degradation of the herbicide 2,4-dichlorophenoxyacetic acid from agricultural wastewater: Process optimization, kinetic and mechanism. *Journal of Molecular Liquids*, *334*, 116116. https://doi.org/https://doi.org/10.1016/j.molliq.2021.116116

Janković, A., Chaudhary, G., & Goia, F. (2025). Optimization through classical design of experiments (DOE): An investigation on the performance of different factorial designs for multi-objective optimization of complex systems. *Journal of Building Engineering*, *102*, 111931. https://doi.org/https://doi.org/10.1016/j.jobe.2025.111931

López-Vinent, N., Cruz-Alcalde, A., Giménez, J., Esplugas, S., & Sans, C. (2021). Improvement of the photo-Fenton process at natural condition of pH using organic fertilizers mixtures: Potential application to agricultural reuse of wastewater. *Applied Catalysis B: Environmental*, *290*, 120066. https://doi.org/https://doi.org/10.1016/j.apcatb.2021.120066

Masomboon, N., Ratanatamskul, C., & Lu, M.-C. (2011). Kinetics of 2,6-dimethylaniline oxidation by various Fenton processes. *Journal of Hazardous Materials*, *192*(1), 347-353. https://doi.org/https://doi.org/10.1016/j.jhazmat.2011.05.034

Meijide, J., Gómez, J., Pazos, M., & Sanromán, M. A. (2016). Degradation of thiamethoxam by the synergetic effect between anodic oxidation and Fenton reactions. *Journal of Hazardous Materials*, *319*, 43-50. https://doi.org/https://doi.org/10.1016/j.jhazmat.2016.02.064

Rand, G., & Clark, J. (2000). Hazard/Risk Assessment of Pyridaben: I. Aquatic Toxicity and Environmental Chemistry. *Ecotoxicology*, *9*, 157-168. https://doi.org/10.1023/A:1008934502348

Suwanaruang, T. (2024). *Pesticides, Toxicity and Its Effects on Human Health and Environment*. https://doi.org/10.9734/bpi/mono/978-81-969435-6-1

Zhou, W., Arcot, Y., Medina, R., Bernal, J., Cisneros-Zevallos, L., & Akbulut, M. (2024). Integrated Pest Management: An Update on the Sustainability Approach to Crop Protection. *Acs Omega*, *9*. https://doi.org/10.1021/acsomega.4c06628

Ziembowicz, S., & Kida, M. (2022). Limitations and future directions of application of the Fenton-like process in micropollutants degradation in water and wastewater treatment: A critical review. *Chemosphere*, *296*, 134041. https://doi.org/https://doi.org/10.1016/j.chemosphere.2022.134041

Effects of Forest Fires on Soil Properties

Nesli AYDIN[1]

Gül KAYKIOĞLU[2]

1- Assoc. Prof. Dr.; Tekirdağ Namık Kemal University, Çorlu Engineering Faculty. naydin@nku.edu.tr
ORCID No: https://orcid.org/0000-0002-7561-4280
2- Prof. Dr.; Tekirdağ Namık Kemal University, Çorlu Engineering Faculty. gkaykioglu@nku.edu.tr
ORCID No: https://orcid.org/0000-0003-3271-211X

ABSTRACT

Fires have played a critical role throughout human evolution and cultural development, serving as one of the oldest tools of environmental transformation. Fires, which originate from natural processes alongside anthropogenic activities, have profoundly impacted the structure, function, and dynamics of ecosystems since prehistoric times. Recent extreme fire events provide striking examples of changing fire dynamics and highlight increasing risks at a global scale. Forest fires have become more frequent in warm climates in Europe, as well as in Turkey, with damaging consequences for biodiversity, environment and the economy. Dry weather conditions caused by climate change and poor land management could easily create conditions favourable for higher intensity forest fires to occur.

In this context, understanding the impacts of fires on soil ecology is emerging as a critical area of research for both assessing ecosystem resilience and developing sustainable fire management strategies. This chapter aims to comprehensively examine the multidimensional impacts of wildfires on the chemical, physical, and biological properties of soils. Post-fire changes in soil structure, water-holding capacity, nutrient cycling, pH balance, and microbial diversity will be discussed, providing both short-term and long-term ecological consequences. Furthermore, key factors driving post-fire soil recovery processes will be discussed, and management implications for ecosystem health and resilience will be presented.

Keywords – Ecology, environmental transformation, forest fires, recovery processes, soil ecology

INTRODUCTION

Forest fires, in particular, are among the most significant natural phenomena frequently observed in diverse climate zones worldwide today, with profound impacts on ecological systems (Bargali et al., 2024; Li et al., 2021). While these fires serve ecological functions such as shaping biodiversity, supporting species regeneration, and accelerating nutrient cycling, they can also have detrimental effects on soil properties, the water cycle, and atmospheric processes.

The origins of forest fires are generally examined in two main groups: (i) natural processes such as lightning strikes and extreme heat waves, and (ii) anthropogenic activities such as agricultural land clearing, controlled burning, or mismanagement. However, in recent decades, global climate

change, land use transformations, and increased human activities have led to radical shifts in fire regimes. Rising temperatures, extended dry periods, and changing precipitation patterns are increasing fire frequency and severity, and rising greenhouse gas emissions are further amplifying these trends. Human activities are estimated to cause over 80% of forest fires; agricultural expansion and forest fragmentation, in particular, make even fire-resistant ecosystems vulnerable (Bargali et al., 2024). These socio-ecological interactions play a decisive role not only on local habitats but also on the global climate system through biophysical feedback mechanisms.

According to satellite-based observations provided by Global Forest Watch (2025), there were 3.68 billion hectares of natural forests worldwide, representing approximately 28% of the world's land area in 2020. The same source also showed that the world had lost 26.8 million hectares of natural forests, causing equivalent to 10 Gt of CO_2 emissions by 2024. In overall, fires accounted for 29% of global tree cover loss between 2001 and 2024.

According to 2024 data from the Turkish General Directorate of Forestry (2024), 3,797 fires broke out in Turkey, burning 27,485 hectares. 1,860 of these fires were caused by negligence or accidents, and approximately 20,000 hectares were burned. Between 2020 and 2024, the largest loss of forest area due to fires occurred in 2021, ending up 139,503 hectares to be burnt. In Turkey, 46% of the tree cover loss between 2021 and 2024 occurred in natural forests, this loss amounted to 142,000 hectares, generating 30.8 Mt CO_2e (Global Forest Watch, 2025).

Forest fires are highly prevalent in all ecosystems worldwide (Aref et al., 2011; Garcia and Gonzalez, 2008). They have a negative effect on both vegetation and soil (Bond and Keeley, 2005; Cerda and Doerr, 2008). They also cause problems with the diversity and sustainability of ecosystems (Boerner et al., 2009). They disrupt the structure of organic matter in the soil by affecting soil biota and physical properties (texture, colour, pH, and bulk density, etc.) (Aref et al., 2011; Bond and Keeley, 2005; Garcia and Gonzalez, 2008). Consequently, they lead to a decrease or even the loss of macronutrients and micronutrients (Certini, 2005). The extent of a fire's impact on forest soil depends on several factors, including the intensity of fire, the type of fuel used, and soil moisture (Boerner et al., 2009; Certini, 2005).

Also, some forest fires could be beneficial to forest soil, depending on their intensity and duration (Knorr et al., 2011; Kodandapani, 2001). In low-intensity fires, the combustion of organic matter in the soil increases the nutrients available to plants (Ketterings and Bigham, 2000; Knicker, 2007). This allows herbaceous plants to grow rapidly and store nutrients (Knicker, 2007; Kraus and Goldammer, 2007). However, high-intensity fires could completely destroy soil organic matter, evaporate essential soil minerals such as N, P, S, and K, and kill beneficial bacteria. Intense forest fires, on the other hand, could cause the formation of certain organic compounds with hydrophobic properties, exacerbating the soil's water-repellent effect (Kraus and Goldammer, 2007). Forest fires also have long-term effects on forest soil (Ketterings and Bigham, 2000; Knicker, 2007).

While forest fires could have both natural and anthropogenic origins, they are a regular phenomenon in tropical forests (Kodandapani, 2001; MacDonald and Huffman, 2004). Forest fires could occur in almost any ecosystem (Letey, 2001). Many of these ecosystems are highly vulnerable to fire, but if subsequent ignitions do not lead to large forest fires, the forest ecosystem is capable of recovering from the negative effects of these fires (Neff et al., 2005). Currently, the number of studies on understanding and managing the causes of natural forest fires is rapidly increasing (Robichaud and Hungerford, 2000; Wan et al., 2001). A common point of these studies is that fires, driven by demographic and land-use changes have become an environmental problem, particularly in tropical regions (Knicker, 2007; Kraus and Goldammer, 2007).

Compared to other environmental impacts of fires on air and water resources, the damage they cause to soil ecosystems is far more complex and devastating (Neff et al., 2005). Main effects are structural changes and loss of organic matter, decreased soil porosity, and increased pH (Letey, 2001; Wan et al., 2001). Variations in soil properties following a fire could lead to different responses in water transmission, vegetation dynamics, and fauna within soil ecosystems (Neff et al., 2005; Knicker, 2007). This wide range of impacts could vary depending on external factors such as fire type, seasonal changes, and the timing, amount, and duration of rainfall, as well as soil type and ecological characteristics (Aref et al., 2011; Garcia and Gonzalez, 2008). These changes could also lead to various indirect effects, including increased hydrophobicity, which leads to reduced infiltration due to porosity blockage and generally increased erosion (Boerner et al., 2009; Certini, 2005). Forest

fires affect the chemical, physical, and biological properties of soil, depending on the intensity of the burning.

EFFECTS OF FOREST FIRES ON PHYSICAL PROPERTIES OF SOIL

Forest fires cause profound and immediate changes to the physical structure of soils, significantly impacting runoff, erosion, and water retention. These changes could have serious consequences for soil health and ecosystem function.

Soil Integrity and Erosion

The intense heat of fires disrupts soil cohesion, which reduces the ability of soil particles to adhere to one another (Gajendiran et al., 2024). This makes the soil more vulnerable to erosion factors such as rain and wind, accelerating erosion processes (Gajendiran et al., 2024). Loss of soil integrity, especially after severe fires, leads to the erosion and transport of the upper soil layer, resulting in the loss of fertile topsoil (Kala, 2023).

The sudden decrease in vegetation cover after a fire exposes the soil surface directly to raindrops and wind. This increases runoff and erosion, thus increasing sediment and nutrient loads in water bodies (Moazeni and Cerda, 2025; Moazeni and Cerda, 2024). Severe fires burn most of the vegetation and soil organic matter in O-horizons, leaving the soil exposed, which reduces plant nutrient uptake and increases post-fire nutrient losses (Tshering et al., 2023).

Soil Structure and Porosity

Fires disrupt the internal aggregate structure of soils and reduce their water-stable aggregate content. Soil aggregates are important structures that ensure soil fertility, aeration, and water-holding capacity. The degradation of these aggregates causes soil compaction and deterioration of its physical structure. Ash and charcoal particles formed after fires could clog soil macropores, reducing soil aeration (Ying et al., 2021). Reduced porosity restricts water infiltration and impedes air exchange, compromising soil stability (Moazeni and Cerda, 2025).

The intense heat generated by fires alters soil structure, breaking down mineral particles and organic matter. This breakdown leads to a decrease in clay content and growth in sand content, making the soil texture coarser. Coarser-textured soils can negatively affect water-holding capacity and nutrient availability. Decreased aggregate stability and high bulk density

increase soil compaction, exacerbating the risk of erosion. This compaction limits soil pore space, restricting water infiltration and air exchange, which compromises soil stability (Moazeni and Cerda, 2025).

Water Retention and Infiltration Rates

One of the most important physical effects of fires is the building of hydrophobic (water-repellent) layers in the soil (Moazeni and Cerda, 2024). The intense heat of a fire could cause organic matter to evaporate and condense to form a hydrophobic film around soil particles (Tshering et al. 2023). This film significantly reduces the soil's capability to absorb water, thereby increasing surface runoff and inhibiting water infiltration into the soil (Ying et al. 2021; Moazeni and Cerda, 2024; Moazeni and Cerda, 2025; Tshering et al. 2023). This poses critical problems for post-fire water management because soils cannot store water, and precipitation quickly runs off the surface. High combustion temperatures cause irreversible changes in the biological and physicochemical quality of the soil, making it extremely susceptible to erosion (Tshering et al. 2023).

Soil properties that are fundamental for good hydrological function, such as soil water infiltration rate, porosity, and hydraulic conductivity, are negatively affected by fires (Ying et al. 2021).

Soil Moisture Content

Fires typically reduce soil moisture content. Post-fire soil drying reduces the ability of microbes involved in biogeochemical cycling to recover and recolonise. This complicates post-fire vegetation regeneration and overall ecosystem recovery (Ying et al. 2021).

During high-intensity fires, temperatures reach approximately 675-700°C, while in medium- and low-intensity fires, temperatures average 3500°C and 250°C, respectively (Wan et al., 2001). In heavily burned soils, the most significant changes in soil colour and texture are seen compared to nearby lightly or moderately burned soils (Robichaud and Hungerford, 2000). In higher-temperature fires, reddening of the soil matrix occurs at the top of the soil (Wardle et al., 2008; Wikars and Schimmel, 2001). The redder hue appears in the soil as the burning intensity increases, due to the conversion of iron oxides and the wide-ranging removal of organic matter (Wardle et al., 2008). Therefore, reddened soil surface patches indicate that the soil has been severely burned. These areas could be identified by a characteristic intensification in magnetisation compared to the surrounding soil (Robichaud and Hungerford, 2000). In low or moderate intensive fires,

the ground is covered with a grey or black ash layer (Wikars and Schimmel, 2001). At high temperatures (average of 6000°C), reddening occurs after 45 minutes of exposure to fire, while red-green and blue-yellow hues result from short-term combustion at approximately 300 or 6000°C (Certini, 2005). As the combustion temperature decreases, the hues become more yellow (Boerner et al., 2009; Certini, 2005). This colour tone variation allows for the localised, heterogeneous accumulation of iron oxides in the forest soil to identify areas where burning is intense (Boerner et al., 2009; Wardle et al., 2008).

Components of soil texture, such as sand, clay, and silt have high-temperature resistance (Lal, 2004; Wikars and Schimmel, 2001). Therefore, such mineralised soil compositions are generally not significantly affected by fire unless exposed to high temperatures (Wikars and Schimmel, 2001). The most sensitive breakdown in soil occurs at temperatures of approximately 4000°C, when clay hydration and the lattice form start to breakdown. Complete destruction of the internal clay structure can occur at temperatures of approximately 700 to 8000°C (Boerner et al., 2009; Lal, 2004). Reynard-Callanan et al. (2010) found that soil layers reddened after fires had significantly lower clay content than unburned soils.

Sand-sized aggregates and blackened layers built in the surface soil during combustion indicate that the fire also caused changes in soil texture (Lal, 2004; Wardle et al., 2008). Soil pH generally increases after forest fires, depending on the amount of ash (Molina et al., 2007; Schafer and Mack, 2010; Wardle et al., 2008). This increase is greater in high-temperature combustion (Wikars and Schimmel, 2001). However, significant increases only occur at higher temperatures (450-500°C) (Certini, 2005).

Bulk density, as is known, expresses the dry soil mass per unit volume and is particularly related to porosity. The bulk density of forest soils increases crucially as a result of forest fires due to the settling of aggregates and the blocking of voids by dispersed clay minerals and ash (Arocena and Opio, 2003). Consequently, fires reduce soil porosity and, consequently, permeability (Arocena and Opio, 2003; Castelein et al., 2001).

Hydrophobicity is one of the soil properties mostly affected by burning during a fire (Cannon et al., 2002). The most significant impact of fire on soil physical properties is the reduction in the water storage capacity of the organic fraction of the upper soil layer (Figure 1) (Ulery and Graham, 1993). As high surface temperatures cause organic matter to burn, soil particles migrate downward in response to the temperature gradient. These

charred organics then condense on the soil particles, making the soil hydrophobic (Cannon et al., 2002; Ulery and Graham, 1993).

EFFECTS OF FOREST FIRES ON CHEMICAL PROPERTIES OF SOIL

Fires significantly alter the chemical composition of soil, leaving lasting impacts on nutrient cycling, pH levels, and toxicant accumulation. These chemical changes play a central role in determining the direction of post-fire ecosystem recovery.

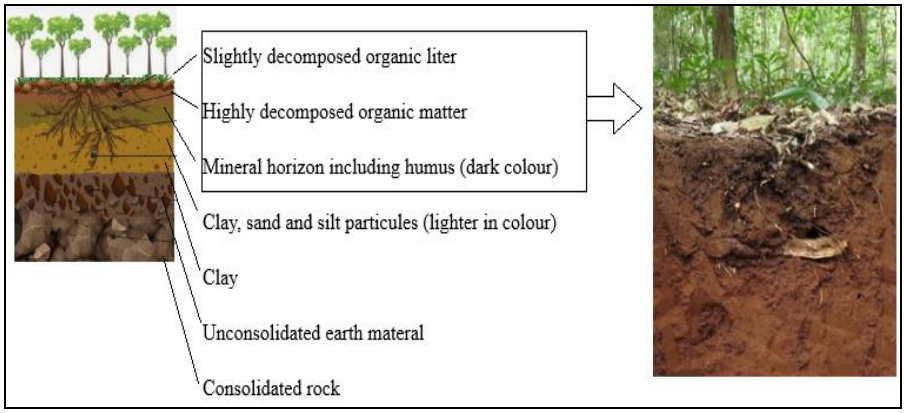


Figure 1: Typical Soil Layer

Nutrient Loss and Release

The combustion of organic matter causes significant volatile losses, particularly of essential nutrients such as carbon and nitrogen, which are released into the atmosphere as carbon dioxide (COx) and nitrogen oxides (NOx). Soil nutrients are released due to heat during fires. It has been reported that nearly all carbon is released into the atmosphere when soil temperatures reach 450°C or above. Nutrient volatilisation begins at temperatures of 250–400°C; for example, heating to 450°C leads to the loss of more than 95% of unheated carbon and more than 60% of nitrogen (Ying et al. 2021).

A temporary increase in nutrients such as nitrate (NO_3-N), phosphorus (P), potassium (K), and nitrogen (N) could be observed in post-fire soils; this initially supports plant regeneration (Moazeni and Cerda, 2025). However, if not managed appropriately, this temporary enrichment could lead to nutrient

leaching and soil degradation in the long term (Moazeni and Cerda, 2025). Frequent fire events create a faster nutrient release rate compared to natural nutrient release processes (Tshering et al., 2023). This accelerated cycling can alter the nutrient balance in the ecosystem, giving some species an advantage over others.

K+ and Na+ contents are slightly increased in post-fire soils, while Ca^{2}+ and Mg^{2}+ contents are also significantly enriched compared to pre-fire levels (Ying et al. 2021).

Soil pH Changes

Mild to moderate wildfires generally increase soil pH (Kala 2023; Ying et al. 2021). This increase could be explained by the combustion of dissolved organic acids in the soil or the leaching of relatively soluble base salts contained in ash (Tshering et al., 2023; Ying et al., 2021). Fire severity is an important factor affecting soil pH which could be higher in post-severe fire areas than in unburned areas (Ying et al., 2021). This is because high-intensity fires cause the combustion of a large number of undecomposed organic acids in the soil and litter. The pH increase is also influenced by other factors such as the buffering capacity of the soil and the amount of accumulated organic matter. The post-fire increase in soil pH affects nutrient availability through factors such as soil microbial activity, decomposition of minerals and organic matter, and chargeability of soil colloids (Ying et al. 2021).

Electrical Conductivity

Electrical conductivity increases in soil after fires (Moazeni and Cerda, 2025). This increase is directly related to ash accumulation and changes in nutrient availability (Moazeni and Cerda, 2025; Tshering et al. 2023). Ash contains high amounts of soluble ions, and the incorporation of these ions into the soil increases electrical conductivity.

Cation Exchange Capacity

Burning of soil organic matter affects soil fertility and nutrients after fire, leading to changes in cation exchange capacity (Ying et al. 2021). Cation exchange capacity represents the amount of positively charged ions that can be retained by the soil and is an important indicator of soil fertility. Cation exchange capacity is positively correlated with organic carbon content, with soil organic matter providing the majority of the space for cation exchange. However, the cation exchange capacity of surface soil

horizons could decrease due to the formation of black carbon or charcoal. The impact of fire on cation exchange capacity varies depending on soil type and fire processes (Ying et al. 2021).

Dissolved Organic Matter

Fire significantly alters the amount and composition of dissolved organic matter compared to non-fire conditions (Tshering et al. 2023). Higher burning intensity significantly reduces vegetation and creates a higher ash load that could be carried to waterways. This ash leaches organic pollutants such as small polyaromatic hydrocarbons (PAHs) into water bodies. The change in dissolved organic matter after fire is largely explained by the loss of water-soluble fulvic acids or the removal of phenolic -OH and -COOH functional groups (Tshering et al. 2023).

Fire intensity is related to physical properties such as temperature, flame length, and heating time. Higher combustion temperatures are associated with the production of more aromatic, humic, fulvic/humic acid-like, and oxidised compounds. Fire temperatures alter the normal decomposition processes of soil and terrestrial organic matter. Black carbon formed after combustion contributes minimally to CO_2 in surface waters in the short term, but acts as a long-term reservoir. This suggests that the solubility of black carbon increases with time (Tshering et al. 2023).

Polyaromatic Hydrocarbons (PAHs)

Forest fires are a potential source of PAHs, which have toxic, mutagenic, carcinogenic, and teratogenic properties that pose environmental and biological concerns. Ash deposits falling to the ground after a fire may contain PAHs. Factors such as fire intensity, type and extent of burned vegetation, temperature, and contact time influence the PAH content of ash.

Although PAHs are strongly adsorbed to organic particles, they could also volatilise and degrade through abiotic and biotic processes. This raises the risk of post-fire PAH contamination. Toxic effects of PAHs from post-fire ash-laden runoff on aquatic species have been reported. For example, the toxicity of ash-laden runoff following a eucalyptus plantation fire in Portugal was assessed, highlighting PAHs as a potential source of PAHs to aquatic ecosystems (Campos and Abrantes, 2021).

The most radical change soils undergo during combustion is the loss of organic matter (Satyam et al., 2012). Soil organic matter is an important component as it reduces erosion, helps regulate soil temperature, provides nutrients for soil biota, and contains easily mineralisable nutrients

(Caldararo, 2002). Forest fires also have a significant impact on soil nutrient dynamics. This change focuses on the availability of macronutrients and micronutrients in the soil (Choromanska and DeLuca, 2002). Research shows that soil nutrients decrease after a forest fire, but the forms available to plants increase (Choromanska and DeLuca, 2002; Satyam et al., 2012). Burned soils have lower nitrogen, higher calcium, and nearly unchanged potassium, magnesium, and phosphorus reserves compared to unburned soils (Caldararo, 2002; Satyam et al., 2012). Charcoal formations, on the other side, accumulate on the forest floor after a forest fire (Choromanska and DeLuca, 2002). They could cause rapid loss of forest humus and belowground carbon within the first decade as they affect charred plant carbohydrates to decompose more rapidly (Wanner and Xylander, 2003).

The immediate impact of fire on soil macronutrients is seen as the loss of these nutrients through volatilisation due to high temperatures (DeBano, 2000). When forest fire temperatures reach 5000°C, approximately half of the nitrogen in organic matter could evaporate (Johnson and Curtis, 2001; Doerr et al., 2000). The extent to which fire affects soil nitrogen pools has long been debated (Kim et al., 2003; Letey, 2001). However, it has been observed that nitrogen concentration in the soil could increase after burning (Johnson and Curtis, 2001). During a fire, small branches and plant nutrients within the soil burn and accumulate on the soil surface (Doerr et al., 2000). These comparatively high soluble plant nutrients on the soil surface could be used for plant growth or easily lost through erosion (Kim et al., 2003). Studies have also found that the amount of micronutrients decreases after a forest fire (Letey, 2001).

EFFECTS OF FOREST FIRES ON BIOLOGICAL PROPERTIES OF SOIL

Soil hosts a complex biological network, and wildfires could have devastating effects on this network. Soil microorganisms, plant root systems, and soil fauna are directly affected by the heat and chemical changes of the fire.

Microbial Activity and Communities

Soil microbial communities play critical roles in nutrient cycling, such as carbon and nitrogen (Ying et al. 2021). Fires cause significant changes in soil microbial biomass and activity. Biomass burning is associated with a short-term decrease in microbial anabolism (growth and

synthesis) relative to catabolism (decomposition) (Tshering et al., 2023). This may lead to a slowdown in post-fire nutrient cycling.

Fires disrupt soil biogeochemical cycles and nutrient dynamics (Moazeni and Cerda 2025). However, increased post-fire soil temperature and moisture may indirectly increase microbial activity rates (Ying et al. 2021). This allows some microorganism species to adapt to or even exploit post-fire conditions. Post-fire increases in soil pH reduce the richness and diversity of mycorrhizal fungi and shift microbial communities from fungal to bacterial dominance (Kala 2023; Ying et al. 2021). This change affects the overall functioning of the soil ecosystem and nutrient availability. Because mycorrhizal fungi play a crucial role in plant nutrient uptake, a decrease in these fungal communities could negatively impact post-fire plant regeneration.

Indirect Impacts on Flora and Fauna

Fires could directly or indirectly affect plant and animal communities, leading to shifts in ecosystem functions. While fires create new habitats for some species, they are lethal for others. This alters the species composition and balance within the ecosystem (Moazeni and Cerda, 2024).

A decline in forest biodiversity is observed in fire-affected areas (Gajendiran et al., 2024). Fires disrupt the delicate balance within ecosystems, leading to complex environmental disruption. For example, some fire studies show a decline in biodiversity in forests after fires, with fire-tolerant species gaining an advantage (Gajendiran et al., 2024).

Fire frequency strongly influences species' habitat loss. High fire severity could reduce the cover or richness of most groups (e.g., early-seral shrubs, forest grasses, and forest shrubs) while increasing the cover and richness of early-seral grasses and trees. This suggests that fire-adapted species thrive in post-fire environments and alter ecosystem structure. For example, Mediterranean shrublands are highly resilient to fire due to the ability of plant species to re-sprout after fire or to recover by re-sprouting from fire-resistant structures (Kala, 2023).

As is well known, soil microorganisms play a number of functional roles in forest ecosystems as they are crucial in nutrient mineralisation, soil transformation, nutrient cycling, and soil structure formation within the soil system (DeBano, 2000; Robichaud, 2000; Schmidt and Noack, 2000; Wanner and Xylander, 2003). However, these organisms are extremely sensitive to environmental changes (Robichaud, 2000; Zackrisson et al., 1996). Intense forest fires could cause a significant reduction in

microorganism biomass (Martin and Moody, 2001; Scott, 2000; Scott and van Wyk, 1990).

Fungi, as is well known, maintain forest health by playing a crucial role in nutrient uptake, root survival, and protection against root pathogens (Figure 2) (DeBano, 2000; Robichaud, 2000). Orumaa et al. (2022) found that burned soils had lower viable fungal populations compared to unburned soils.

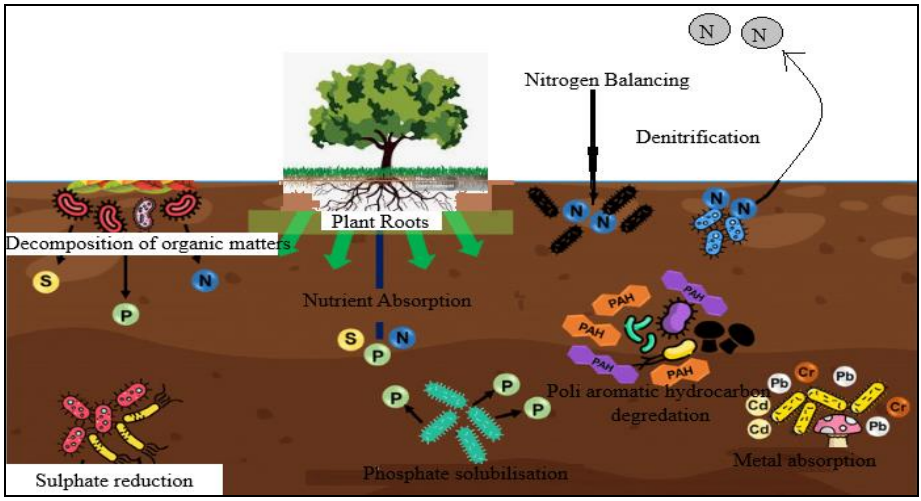


Figure 2: The Role of Micro-Organisms in Forest Ecosystems (Palit et al., 2022)

FACTORS AFFECTING IMPACT SEVERITY

The severity of wildfire impacts on soil properties is not due to a single factor, it is the result of a complex interaction of a number of environmental and fire-related factors (Moazeni and Cerda, 2024). Understanding these factors is vital for predicting and managing post-fire recovery processes (Moazeni and Cerda, 2024).

Fire Intensity and Severity

The impact of fire on ecosystems depends on factors such as the origin of the fire, its frequency, date, vegetation types, and the extent of damage caused (Bargali et al., 2024). Fire intensity and severity determine changes in the chemical and physical properties of the soil (Kala, 2023; Moazeni and Cerda, 2025; Moazeni and Cerda, 2024; Tshering et al., 2023). Burn severity refers to the magnitude of changes in an ecosystem following a fire, such as the depletion of vegetation and litter, ash production, changes in surface reflectance, and deterioration in soil properties. Fire intensity describes the

physical process of burning, releasing energy from organic matter, and is measured as the energy per unit volume times the rate of energy movement (Tshering et al., 2023).

Fire intensity describes the physical combustion process that directly affects heat penetration into deeper soil layers, which could significantly alter biogeochemical processes. High combustion temperatures cause irreversible changes in the biological and physical chemical quality of the soil, making it highly susceptible to erosion (Tshering et al. 2023). Soil temperature is closely related to fire severity. Under mild soil heating, mineral soil temperatures generally do not exceed 100 °C at the surface and 50 °C at a depth of 5 cm, while under severe soil heating, temperatures could reach almost 700 °C at the surface and exceed 250 °C at a depth of 10 cm (Ying et al. 2021).

Fire Frequency

Fire frequency affects soil carbon and nitrogen stocks and soil organic matter content. Frequent fire events produce a faster nutrient release rate compared to natural nutrient release processes. More frequently burned areas generally contain less terrestrial organic matter. For example, in Australia, planned fires were reported to reduce total litter mass by 38% and result in a carbon loss of 3.3 tonnes/ha from the upper 2 to 2.5 cm litter layer. Increased fire frequency could contribute to reducing terrestrial dissolved organic matter sources and thereby affect the amount and composition of dissolved organic matter (Tshering et al. 2023).

Vegetation Type and Fuel Load

The impact of fires varies depending on the presence of fire hazards such as forest composition, fuel quantity, and moisture level (Gajendiran et al., 2024). Fuel accumulation is determined by the formation and decomposition of combustible material, which in turn varies by vegetation type and forest. Vegetation directly influences burn dynamics and post-fire soil effects. For example, coarse-textured vegetation (e.g., conifers) tends to burn more intensely and for longer periods (Gajendiran et al., 2024). The expansion of agricultural land into forests contributes to forest fragmentation and degradation, which in turn reduces forest fire resilience (Bargali et al., 2024).

Post-Fire Climatic Conditions

Post-fire climatic conditions have a significant impact on soil recovery and contamination dynamics. The timing of rainfall, wind, and the

intensity of rainfall events affect the mobility of PAHs (Compos and Abrantes, 2021). Rainfall intensity could increase runoff and dissolved organic matter transport (Tshering et al. 2023). For example, one study examined the effects of fire severity and burn slope on rainfall-induced runoff, erosion, and hydrological connectivity (Tshering et al. 2023). When combined with factors such as drought, disease, and insect infestation, they could exacerbate negative impacts on certain plant species and entire ecosystems (Gajendiran et al., 2024).

Management and Recovery Strategies

Various management and recovery strategies have been developed to mitigate the negative impacts of forest fires on soil properties and accelerate post-fire recovery. These strategies cover a wide spectrum, from preventive measures to post-fire rehabilitation.

Prescribed Fires

Prescribed fires, or controlled fires, are a proactive strategy for preventing and managing forest fires (Bargali et al., 2024). This practice aims to control the intensity and spread of fires by deliberately burning combustible material (fuel load) in forested areas (Davis et al., 2024; Bargali et al., 2024). Simulation models show that planned fires are effective in reducing fire occurrence by removing combustible material (Bargali et al., 2024).

Prescribed fires could affect soil properties and nutrient cycling (Gajendiran et al., 2024; Kala, 2023; Tshering et al. 2023). Some studies show that prescribed fires alter the organic matter content and nutrients of surface soil (Gajendiran et al., 2024). Low-impact fires have been reported to increase nutrients available to plants and promote more beneficial herbaceous vegetation (Kala, 2023). However, blindly applying prescribed fires could have negative consequences, such as damaging young forests. Therefore, future research efforts should prioritise refining simulation accuracy and scale and adhering to prescribed fires management practices (Bargali et al., 2024).

Remote Sensing and GIS (Geographic Information Systems)

Remote sensing (Remote Sensing) and Geographic Information Systems (GIS) are emerging as critical assets for forest fire management, particularly in providing improved fire mitigation and control measures. These technologies proactively assist authorities by providing early warnings

for fire-prone forest areas, tracking fires over a period of time, and estimating burned areas (Gajendiran et al., 2024; Moazeni and Cerda, 2025).

The use of remote sensing and GIS has significantly enhanced fire risk assessment and forecasting capabilities. These tools could help users to identify fire-prone areas. So, the effective fire management strategies could be developed by applying them (Moazeni and Cerda, 2025). For example, satellite-based sensors such as MODIS (Moderate Resolution Imaging Spectroradiometer) and SNPP-VIIRS (Suomi National Polar-orbiting Partnership - Visible Infrared Imaging Radiometer Suite) are used to detect and monitor wildfire events (Gajendiran et al., 2024).

Post-Fire Rehabilitation

Post-fire rehabilitation involves interventions designed to promote ecosystem recovery and reduce future fire risk. This process encompasses various aspects, such as reducing soil erosion, promoting vegetation regeneration, and restoring hydrological functions. The effectiveness of post-fire rehabilitation depends on actively understanding the external factors that influence recovery techniques. These include topographic factors, fire severity, and vegetation types (Gajendiran et al., 2024).

Comprehensive Management Approaches

Strategies that integrate resource management, policymaking, and community engagement are necessary to mitigate the impacts of wildfires on watershed hydrology (Moazeni and Cerda, 2024). The importance of interdisciplinary approaches and the integration of remote sensing and GIS technologies has been emphasised to address the complex linkages between wildfires and climate change (Gajendiran et al., 2024). Effective fire management requires the integration of community engagement, robust policy frameworks, and advanced management strategies (Moazeni and Cerda, 2025).

RESULTS AND DISCUSSION

Wildfires are a complex and multifaceted phenomenon, both a natural process and a growing threat to ecosystems worldwide. Their impact on soil properties is critical to ecosystem functioning and long-term health. Fires profoundly alter soil physical structure, disrupting its integrity, accelerating erosion, reducing soil structure and porosity, creating hydrophobic layers

that reduce water retention and infiltration capacity, and reducing soil moisture.

At the chemical level, fires cause nutrients (especially carbon and nitrogen) to volatilise into the atmosphere, while post-fire ash accumulation increases soil pH and electrical conductivity. The combustion of soil organic matter affects cation exchange capacity and negatively impact water quality by altering the amount and composition of dissolved organic matter and PAHs. Biologically, fires disrupt soil microbial communities and their activities, reduce mycorrhizal fungal diversity, and alter the distribution and abundance of plant and animal species within the ecosystem.

The impact of fire on soil organic matter varies depending on fire severity, dryness of surface organic matter, soil moisture, soil type, and fire type (DeBano, 2000; Robichaud, 2000). Due to this variability, a standardised treatment for soil organic matter restoration cannot be recommended by analysts (Robichaud, 2000; Schmidt and Noack, 2000).

The water repellence of soil prevents water from wetting or infiltrating dry soils containing hydrophobic organic compounds (Cannon et al., 2002). This could affect the survival of seedlings planted without soil amendment after forest fires and the subsequent establishment of vegetation (Cannon et al., 2002; Ulery and Graham, 1993).

Both anthropogenic and natural fires appear to have a significant impact on forest ecosystems. These fires increase the water repellence of forest soils, leading to infiltration and soil erosion. Fires also significantly affect pH, soil colour, texture, bulk density, and other characteristics. Post-forest fire chemical changes in soil are vital due to their impact on nutrient cycling and soil organic matter (Schmidt and Noack, 2000; Wanner and Xylander, 2003). Soil biological properties are also significantly affected by fires. This is due to the sensitivity of microorganisms to high temperatures. Fire reduces both the number and species richness of soil-dwelling microorganisms (Orumaa et al., 2022).

In recent years, studies aimed at improving the effectiveness of water as a forest fire suppressant have become prominent. For example, Anderson and Prosser (2023) highlighted the toxicity of fire fighting water additives to forest biota when deciding which products to apply in case of a fire. Xu et al. (2025) prepared a novel silica hydrogel as a fire extinguishing agent and introduced this agent as a phosphorus and potassium fire extinguishing element.

The severity of these impacts is determined by the interaction of a number of factors, including fire intensity, severity, frequency, vegetation

type, fuel load, and post-fire climate conditions. Given these complex dynamics, comprehensive strategies are needed to combat wildfires and support post-fire recovery. Modern tools, remote sensing such as GIS and interdisciplinary approaches, such as planned fires, play a vital role in reducing fire risk and enhancing ecosystem resilience. However, current challenges, such as insufficient resources and public awareness, could limit the effectiveness of these efforts. Future research will help further refine fire management strategies by focusing on more precise studies in areas such as soil carbon cycling and water quality.

REFERENCES

Anderson, J., and Prosser, R.S., (2023). Investigation of the potential effects of firefighting water additives on soil invertebrates and terrestrial plants. Chemosphere, *313*, https://doi.org/10.1016/j.chemosphere.2022.137496.

Arocena, J. M., and Opio, C., (2003). Prescribed fire-induced changes in properties of sub-boreal forest soils. Geoderma, *113*, 1-16.

Aref, I.M., Atta, H.A., and Ghamade, A.R., (2011). Effect of forest fires on tree diversity and some soil properties. International Journal of Agriculture and Biology, *13*, 659-664

Bargali, H., Pandey, A., Bhatt, D., Sundriyal, R.C., and Uniyal, V.P., (2024). Forest fire management, funding dynamics, and research in the burning frontier: A comprehensive review. Trees, Forests and People, *16*, 100526. https://doi.org/10.1016/j.tfp.2024.100526

Boerner, R.E.C., Hart, S., and Huang, J., (2009). Impacts of fire and fire surrogate treatments, Ecological Applications, *19*(2), 338–358

Bond, W.J., and Keeley, J.E., and (2005). Fire as a global 'herbivore': the ecology and evolution of flammable ecosystem. Trends in Ecology and Evolution, *20*(7), 387-394

Caldararo, N., (2002). Human ecological intervention and the role of forest fires in human ecology. Sci Total Environ., 292, 141–165

Campos, I., and Abrantes, N., (2021). Forest fires as drivers of contamination of polycyclic aromatic hydrocarbons to the terrestrial and aquatic ecosystems. Current Opinion in Environmental Science & Health, 24, 100293. https://doi.org/10.1016/j.coesh.2021.100293

Cannon, W.F., Woodruff, L.G., Dicken, C., and Saari, S.M., (2002). Prolonged influence of wildfires on the geochemistry of forest soils, Isle Royale National Park, Michigan and Voyageurs National Park, Minnesota, Geological Society of America, Abstracts with Programs, *34*, 6.

Castelein, O., Soulestin, B., Bonnet, J.P., and Blanchart, P., (2001). The influence of heating rate on the thermal behaviour and mullite formation from a kaolin raw material. Ceramics International, *27*, 517-522.

Cerda, A., Doerr, S.H., 2008, The effect of ash and needle cover on surface runoff and erosion in the immediate post-fire period, Catena, 74, 3, 256–263

Certini, G., (2005). Effect of fire on properties of soil - A review. Oecologia, *143*, 1-10

Choromanska, U., and DeLuca, T.H., (2002). Microbial activity and nitrogen mineralization in forest mineral soils following heating: evaluation of post-fire effects. Soil Biol Biochem., *34*, 263–271

Davis, K.T., Peeler, J., Fargione, J., Haugo, R.D., Metlen, K.L., Robles, M.D., and Woolley, T., (2024). Tamm review: A meta-analysis of thinning, prescribed fire, and wildfire effects on subsequent wildfire severity in conifer dominated forests of the Western US. Forest Ecology and Management, *561*, 121885. https://doi.org/10.1016/j.foreco.2024.121885

DeBano, L.F., (2000) The role of fire and soil heating on water repellence in wildland environments: a review. J Hydrol, *231*, 195–206

Doerr, S.H., Shakesby, R.A., and Walsh, R.P.D., (2000) Soil water repellence: its causes, characteristics and hydro-geomorphological significance. Earth-Sci Rev, *51*, 33–65

Gajendiran, K., Kandasamy, S., and Narayanan, M., (2024). Influences of wildfire on the forest ecosystem and climate change: A comprehensive study. Environmental Research, *240*, 117537. https://doi.org/10.1016/j.envres.2023.117537

Garcia, M.S, and Gonzalez, P.S., (2008). Short- and medium- term effects of fire and fire-fighting chemicals on soil micronutrient availability. The Science of Total Environment, *407*, 297-303

Global Forest Watch, (2025). Global dashboards, https://www.globalforestwatch.org/dashboards/global/ accessed 01 August 2025

Johnson, D.W., Curtis, P.S., (2001). Effects of forest management on soil C and N storage: Meta analysis. For Ecol Manage, *140*, 227–238

Kala, C.P., (2023). Environmental and socioeconomic impacts of forest fires: A call for multilateral cooperation and management interventions, Natural Hazards Research, *3*, 286–294. https://doi.org/10.1016/j.nhres.2023.04.003

Ketterings, Q.M., and Bigham, J.M., (2000). Soil color as an indicator of slash-and-burn fire severity and soil fertility in Sumatra, Indonesia, Soil Science Society of American Journal, *64*, 1826–1833

Kim, E.J., Oh, J.E., and Chang, Y.S., (2003). Effects of forest fire on the level and distribution of PCDD/Fs and PAHs in soil. Sci Total Environ, *311*, 177–189

Knicker, H., (2007). How does fire affect the nature and stability of soil organic nitrogen and carbon? A review. Biogeochemistry, *85*, 91-118

Knorr, W., Pytharoulis, I., and Petropoulos, G.P., (2011). Combined use of weather forecasting and satellite remote sensing information for fire risk, fire and fire impact monitoring. Computational Ecology and Software, *1*(2), 112-120

Kodandapani, N., (2001). Forest fires: Origin and ecological paradoxes. Resonance, *6*(11), 34-41

Kraus, D., and Goldammer, J., (2007). Fire Regimes and Ecosystems: An Overview of Fire Forest Fires in India, Madurai, India. Ecology in Tropical Ecosystems, *9*(13)

Lal, R., (2004). Soil carbon sequestration to mitigate climate change, Geoderma. *123*, 1-22

Letey, J., (2001). Causes and consequences of fire-induced soil water repellency, Hydrological Processes, 15(15), 2867-2875

Li, X.Y., Jin, H.J., Wang, H.W., Marchenko, S.S., Shan, W., Luo, D.L., He, R.X., Spektor, V., Huang, Y.D., Li, X.Y., and Jia, N., (2021). Influences of forest

fires on the permafrost environment: A review. Advances in Climate Change Research, *12*, 48–65. https://doi.org/10.1016/j.accre.2021.01.001

MacDonald, L.H., and Huffman, E.L., (2004). Post-fire soil water repellency: Persistence and soil moisture threshold. Soil Science Society of America Journal, *68*, 1729-1734

Martin, D.A., Moody, J.A., (2001). Comparison of soil infiltration rates in burned and unburned mountainous watersheds. Hydrol Process, *15*, 2893–2903

Moazeni, S., and Cerda, A., (2024). The impacts of forest fires on watershed hydrological response: A review. Trees, Forests and People, 18, 100707. https://doi.org/10.1016/j.tfp.2024.100707

Moazeni, S., and Cerda, A., (2025). A comprehensive bibliometric review of forest fires in Iran, Environmental Development, *54*, 101160. https://doi.org/10.1016/j.envdev.2025.101160

Neff, J., Harden, J., and Gleixner, G., (2005). Fire effects on soil organic matter content, composition, and nutrients in boreal interior Alaska. Canadian Journal of Forest Research, *35*, 2178-2187

Palit, K., Rath, S., Chatterjee, S. and Das, S., (2022), Microbial diversity and ecological interactions of microorganisms in the mangrove ecosystem: Threats, vulnerability, and adaptations. Environmental Science and Pollution Research, 10.1007/s11356-022-19048-7

Reynard-Callanan, J.R., Pope, G.A., Gorring, M.L., and Feng, H., (2010), Effects of High-Intensity Forest Fires on Soil Clay Mineralogy. Physical Geography, *31*, 5, https://doi.org/10.2747/0272-3646.31.5.407

Robichaud, P., and Hungerford, R., (2000). Water repellency by laboratory burning of four northern. Journal of Hydrology, *231-232*, 207-219

Robichaud, P.R., (2000). Fire effects on infiltration rates after prescribed fire in Northern Rocky Mountain forests. USA. J Hydrol., *231*, 220–229

Satyam, V., and Jayakumar, S., (2012). Impact of forest fire on physical, chemical and biological properties of soil: A review, Proceedings of the International Academy of Ecology and Environmental Sciences, *2*(3), 168-176

Schmidt, M.W.I., and Noack, A.G., (2000). Black carbon in soils and sediments: analysis, distribution, implications, and current challenges. Global Biogeochem Cycles, *14*, 777–793

Scott, D.F., and van Wyk, D.B., (1990). The effects of wildfire on soil wettability and hydrological behaviour of an afforested catchment. J Hydrol, *121*, 239–256

Scott, D.F., (2000). Soil wettability in forested catchments in South Africa; as measured by different methods and as affected by vegetation cover and soil characteristics. J Hydrol, *231–232*, 87–104

Tshering, K., Miotlinski, K., Blake, D., Boyce, M. C., Bath, A., Carvalho, A., and Horwitz, P., (2023). Effect of fire on characteristics of dissolved organic matter in forested catchments in the Mediterranean biome: A review. Water Research, *230*, 119490. https://doi.org/10.1016/j.watres.2022.119490

Turkish General Directorate of Forestry, (2024), Statistics, https://www.ogm.gov.tr/tr/e-kutuphane/resmi-istatistikler accessed 2nd August 2025

Ulery, A.L., Graham, R.C., (1993). Forest fire effects on soil color and texture. Soil Science Society of America Journal, *57*, 135-140.

Wan, S., Hui, D., and Luo, Y., (2001). Fire effects on nitrogen pools and dynamics in terrestrial ecosystems: A metaanalysis. Ecological Applications, *11*(5), 1349-1365

Wanner, M., and Xylander, W.E.R., (2003) Transient fires useful for habitat-management do not affect soil microfauna (testate amoebae)—a study on an active military training area in eastern Germany. Ecol Eng, *20*, 113–119

Wardle, D.A., Nilsson, M.C., and Zackrisson, O., (2008). Fire-derived charcoal causes loss of forest humus, Science. *320*, 629

Wikars, L.O., and Schimmel, J., (2001). Immediate effects of fire-severity on soil invertebrates in cut and uncut pine forests. Forest Ecology and Management, *141*(3), 189-200

Xu, M., Wei, Y., Qin, A., Xu, Y., Xu, M., Li, B., and Liu, L., (2025). Novel silica hydrogel-based forest fire extinguishing agent: Construction, fire extinguishing performance and mechanism study. Journal of Cleaner Production, *486*, 144490, https://doi.org/10.1016/j.jclepro.2024.144490.

Ying, L., Cheng, H., Shen, Z., Guan, P., Luo, C., and Peng, X., (2021), Relative humidity and agricultural activities dominate wildfire ignitions in Yunnan, Southwest China: Patterns, thresholds, and implications. Agricultural and Forest Meteorology, *307*,108540

Zackrisson, O., Nilsson, M.C., Wardle, and D.A., (1996). Key ecological function of charcoal from wildfire in the Boreal forest. Oikos, *77*, 10–19

Convolutional Neural Network Approach to Modulation Classification for 5G and Cognitive Radio

Asuman SAVAŞCIHABEŞ*

(*) Assoc. Prof. Dr.; Nuh Naci Yazgan Üniversitesi Mühendislik Fakültesi Elektrik-Elektronik Mühendisliği Bölümü. asuman83@gmail.com ORCID No: 0000-0002-7261-1906

ABSTRACT

The rapid growth of wireless communication systems has led to increasing demand for advanced and adaptive technologies capable of handling high data traffic and dense user environments. A critical factor in this context is the accurate recognition of modulation schemes, which directly impacts overall system efficiency. In next-generation networks such as 5G and 6G, dynamic and noisy signal conditions make traditional modulation recognition techniques less effective. To overcome these challenges, artificial intelligence-based approaches, particularly deep learning, have gained prominence due to their superior ability to learn directly from data.

In this study, a deep learning model leveraging in-phase (I) and quadrature (Q) components was developed for automatic modulation classification. The proposed CNN-based architecture processes two-dimensional I/Q matrices and classifies multiple modulation formats, including BPSK, QPSK, 8PSK, 16QAM, and 64QAM. Simulation experiments were carried out in MATLAB, where synthetic signals were generated and transmitted over an AWGN channel at varying SNR levels. The extracted I/Q samples were reshaped into four-dimensional tensors suitable for CNN input.

The CNN model was trained using the Adam optimizer over 10 epochs, incorporating convolutional and max-pooling layers. Evaluation metrics such as accuracy, confusion matrix, and SNR-dependent performance were employed. The model achieved an overall classification accuracy of 94.3%, demonstrating robust recognition even in low-SNR conditions. Results confirm that the proposed framework provides a reliable and hardware-deployable solution for real-time wireless communication systems.

Keywords –5G, Convolutional Neural Network, Digital Modulation Schemes, Automatic Modulation Classification, Deep Learning.

INTRODUCTION

The rapid evolution of wireless communication systems has been primarily driven by the exponential growth in data traffic, user density, and demand for seamless connectivity. To meet these requirements, next-generation networks such as 5G and the forthcoming 6G must adopt more sophisticated and adaptive communication technologies (Strinati et al., 2019; Zhang et al., 2019). A key element in this context is the accurate recognition of modulation schemes, which directly impacts spectral efficiency, bit error

rate (BER), and overall system reliability. In highly dynamic and interference-prone environments such as cognitive radio networks, modulation classification becomes even more critical for efficient spectrum utilization (Zeng et al., 2010; Di Renzo et al., 2019).

Automatic Modulation Classification (AMC) plays a critical role in modern wireless communication systems, particularly within dynamic and spectrum-constrained environments such as cognitive radio networks and emerging 5G/6G infrastructures. Traditional AMC approaches, which rely heavily on handcrafted features and statistical models, often struggle to maintain accuracy under noisy or non-stationary conditions. To overcome these limitations, recent research has focused on deep learning-based models capable of learning directly from raw in-phase and quadrature (I/Q) signal components. Convolutional Neural Networks (CNNs), in particular, have shown significant promise in extracting complex temporal and spatial features without the need for manual signal processing. In this study, we propose a deep learning framework that leverages a CNN-based architecture for automatic classification of various digital modulation schemes. The goal is to enhance classification accuracy and robustness, enabling real-time signal identification in advanced communication systems.

Traditional approaches to automatic modulation classification (AMC) have relied on statistical and likelihood-based methods. Techniques such as cumulant-based detection, spectral correlation, and decision-theoretic approaches have been widely studied (Swami & Sadler, 2000; Gardner, 1987; Azzouz & Nandi, 1996). Although effective under controlled conditions, these methods often suffer performance degradation in low signal-to-noise ratio (SNR) environments or when the signal statistics deviate from their assumptions. As a result, conventional methods have limited applicability in modern wireless communication scenarios characterized by fading, interference, and spectrum heterogeneity.

In recent years, artificial intelligence (AI) and deep learning have emerged as powerful alternatives, owing to their ability to extract robust features directly from raw signals without requiring handcrafted feature design (LeCun et al., 2015; Goodfellow et al., 2016). Convolutional neural networks (CNNs), in particular, have shown remarkable success in modulation recognition tasks by leveraging hierarchical feature learning (O'Shea et al., 2016; Karra et al., 2017; Yu et al., 2018). Building on this

foundation, more advanced models such as recurrent neural networks (RNNs) and hybrid CNN-LSTM architectures have been proposed to capture temporal dependencies in modulation sequences (Zhang et al., 2019; Rajendran et al., 2018; Liu et al., 2020). These models have demonstrated higher accuracy and robustness in dynamic wireless environments compared to traditional classifiers.

Beyond CNNs and RNNs, unsupervised learning approaches such as autoencoders have also been employed to enhance feature extraction and denoising capabilities, further improving classification performance (Ali et al., 2020; Huang et al., 2018). Moreover, optimization-driven frameworks—including particle swarm optimization (PSO) and genetic algorithms—have been integrated with deep learning architectures to fine-tune network parameters and prevent overfitting (Wang et al., 2018; Li et al., 2019). These developments collectively highlight a paradigm shift in AMC research, where AI-driven models increasingly outperform conventional statistical approaches, especially under noisy or time-varying conditions.

In parallel, the application domain of AMC has expanded significantly. In 5G and 6G systems, accurate and real-time modulation recognition is essential for adaptive resource allocation, interference management, and link adaptation (Strinati et al., 2019; Zhang et al., 2019). Within cognitive radio networks, AMC enables efficient spectrum sensing and dynamic spectrum access, thereby supporting intelligent and flexible communication strategies (O'Shea & West, 2016; Zeng et al., 2010; Di Renzo et al., 2019). The convergence of AMC with AI and deep learning is therefore recognized as a cornerstone of next-generation wireless communication systems.

Motivated by these considerations, this study proposes a CNN-based deep learning framework for modulation classification directly from in-phase (I) and quadrature (Q) components. The contributions of this work are as follows:

- Dataset Generation: Synthetic datasets were created in MATLAB for five digital modulation formats (BPSK, QPSK, 8PSK, 16QAM, 64QAM) and transmitted over an AWGN channel under varying SNR conditions.

- CNN Architecture: A lightweight yet effective CNN was developed, capable of directly classifying modulation types from raw I/Q matrices without handcrafted features.
- Performance Evaluation: The model was extensively analyzed using accuracy metrics, confusion matrices, and SNR-dependent classification curves, demonstrating robustness in noisy environments.
- Practical Relevance: The proposed framework is suitable for real-time implementation in 5G, 6G, and cognitive radio systems, where reliable modulation recognition is a prerequisite for intelligent spectrum management and adaptive communication.

By integrating deep learning with I/Q-domain signal representations, this study contributes to the design of spectrum-efficient, adaptive, and intelligent modulation classification systems, providing a solid foundation for the realization of future wireless infrastructures.

In recent years, artificial intelligence (AI)-based methods have emerged as a powerful alternative for addressing such complex classification problems. Deep learning techniques, in particular, provide superior accuracy compared to conventional algorithms owing to their data-driven learning capabilities within signal processing tasks.

In this study, a deep learning model based on in-phase (I) and quadrature (Q) components has been developed for the modulation recognition problem. The proposed model automatically classifies various modulation schemes (BPSK, QPSK, 8PSK, 16QAM, 64QAM) by processing two-dimensional signal matrices constructed from I/Q samples.

The main objective of the system is to deliver an AI-based module that achieves high recognition accuracy even under noisy conditions, can be embedded into hardware, and is suitable for real-time integration into wireless communication systems.

Simulations were conducted in MATLAB, and performance evaluations were performed using a Convolutional Neural Network (CNN) architecture. Initially, synthetic signals corresponding to different modulation formats were generated from random bit streams in MATLAB and transmitted through an Additive White Gaussian Noise (AWGN)

channel. Subsequently, the I and Q components of the received signals were extracted and arranged into 2×1000 input vectors. These I/Q samples were reshaped into four-dimensional tensors to be compatible with the CNN model.

During training, the CNN was optimized with two convolutional layers and max-pooling stages. The model was trained using the Adam optimizer for 10 epochs. Performance evaluation was carried out in terms of accuracy, confusion matrix, and accuracy variations across different SNR levels. The fundamental aim of this framework is to establish an AI-based model capable of directly recognizing the structural differences among modulation types using I/Q components.

METHODOLOGY

In this study, a supervised deep learning approach based on in-phase (I) and quadrature (Q) components was developed for the modulation classification problem. The proposed methodological framework consists of four main stages: data generation, preprocessing, model architecture, and evaluation.

Data Generation

In the first stage, five digital modulation schemes—BPSK, QPSK, 8PSK, 16QAM, and 64QAM—were considered. For each modulation type, random bit sequences of 1024 symbols were generated and mapped to the corresponding constellation points in MATLAB to produce synthetic baseband signals. Each waveform was transmitted through an Additive White Gaussian Noise (AWGN) channel using MATLAB's awgn function, with noise levels varied across different SNR values in the range of 0–25 dB. The received complex signals were separated into their in-phase (I) and quadrature (Q) components, which were then arranged into 2×1000 matrices and used as classifier inputs.

CNN Architecture

The convolutional neural network (CNN) employed in this study was designed with the following structure:

- Input Layer: Accepts 2×1000 I/Q samples.
- Convolutional Stage:
    - First layer: 16 filters of size 1×5, followed by batch normalization and ReLU activation.
    - Second layer: 32 filters of size 1×3, followed by batch normalization and ReLU activation.
- Max-Pooling Layers: Each convolutional output is downsampled using max-pooling with kernel size 1×2 and stride 2.
- Flatten Layer: Extracted feature maps are converted into vectors.
- Fully Connected Layers:
    - A dense layer with 64 neurons and ReLU activation.
    - A final fully connected layer with 5 neurons (corresponding to the five modulation classes).
    - Softmax output layer with cross-entropy classification.

An alternative implementation in MATLAB Deep Learning Toolbox was also employed with minor modifications:

- Input layer: 2×1000×1 tensor.
- Convolutional 1: 16 filters, 1×5 kernel, 'same' padding, followed by ReLU activation and 1×2 max-pooling.
- Convolutional 2: 32 filters, 1×3 kernel, followed by ReLU activation and 1×2 max-pooling.
- Fully connected layer: 128 neurons with ReLU activation.
- Dropout: 50%.
- Softmax output layer with 5 classes.

Training Protocol and Evaluation Metrics

The dataset was partitioned such that 80% was used for training and 20% for testing. Training was performed using the Adam optimizer with a learning rate of 1×10^{-3} over 10 epochs, with mini-batches of 32 samples. At the end of each epoch, the training loss and accuracy were recorded.

Model performance was evaluated using the following metrics:

- Classification Accuracy
- Confusion Matrix (to identify correctly classified and misclassified modulation schemes)
- Training Loss Curves (to assess convergence behavior)
- SNR-dependent Accuracy (to evaluate robustness under noisy channel conditions)
- Class-wise Prediction Performance

The results of confusion matrix analysis enabled observation of misclassifications among similar constellations (e.g., 16QAM vs. 64QAM), while SNR-based performance evaluation demonstrated the model's resilience in noisy environments.

To prevent overfitting during training, regularization techniques such as dropout (set to 50%) and batch normalization were employed. These strategies ensured stable convergence and improved the generalization ability of the model. In addition, the training process was monitored through loss and accuracy curves, enabling the detection of potential early convergence or underfitting. Although early stopping was not explicitly applied in this study, it can be integrated as a complementary mechanism in future extensions to optimize training efficiency.

Another important design choice was the direct utilization of raw I/Q components instead of higher-order statistical features or spectrogram-based transformations. This decision was motivated by recent findings in the literature (O'Shea et al., 2016; Rajendran et al., 2018), which demonstrated that deep neural networks are capable of extracting robust representations directly from raw signal samples without requiring handcrafted features. As a result, the proposed CNN framework is computationally efficient and particularly suitable for real-time deployment in embedded hardware platforms.

Table 1: CNN Architecture for Modulation Classification

Layer	Description
Input Layer	2×1000×1 tensor (I/Q signal samples)
Convolutional Layer 1	16 filters, kernel size = 1×5, *same* padding
Activation & Pooling 1	ReLU activation, Max-pooling (1×2)
Convolutional Layer 2	32 filters, kernel size = 1×3
Activation & Pooling 2	ReLU activation, Max-pooling (1×2)
Fully Connected Layer	128 neurons, ReLU activation
Dropout Layer	Dropout rate = 50%
Output Layer	Softmax classifier with 5 neurons (BPSK, QPSK, 8PSK, 16QAM, 64QAM)

Table 1 summarizes the architecture of the proposed CNN model designed for automatic modulation classification. The network consists of two convolutional layers with max-pooling, followed by fully connected and dropout layers, and concludes with a softmax output layer that distinguishes between five modulation schemes.

Furthermore, the proposed model can be extended to incorporate data augmentation strategies, such as random time shifts, amplitude scaling, and phase rotations, which are commonly encountered in practical wireless environments. These augmentations would allow the classifier to better generalize under realistic conditions such as fading channels, frequency offsets, and nonlinear distortions.

Finally, from an application perspective, the methodology aligns with the needs of next-generation wireless systems. In 5G and 6G networks, accurate AMC is essential for adaptive modulation and coding (AMC) schemes, link adaptation, and interference management. Similarly, in cognitive radio environments, the ability to automatically identify modulation types supports dynamic spectrum access and enhances spectrum utilization. Thus, the proposed CNN-based framework provides both academic significance and practical relevance by bridging theoretical deep learning techniques with real-world communication challenges.

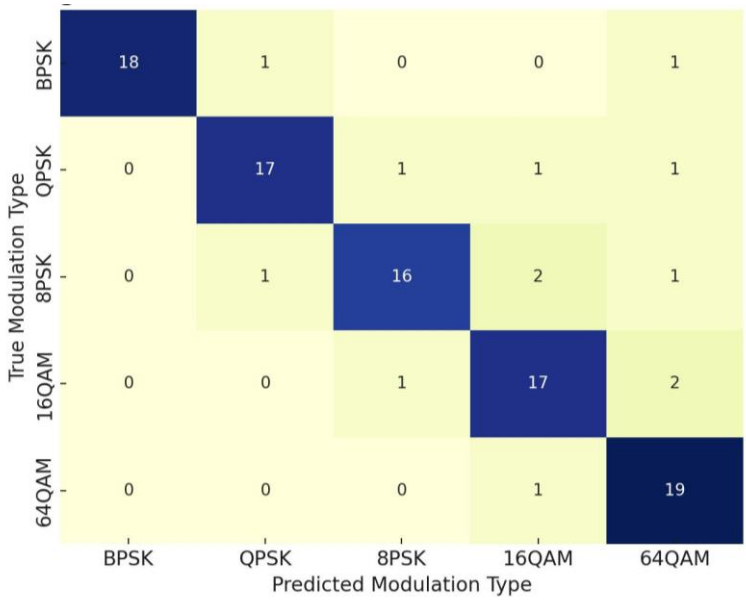
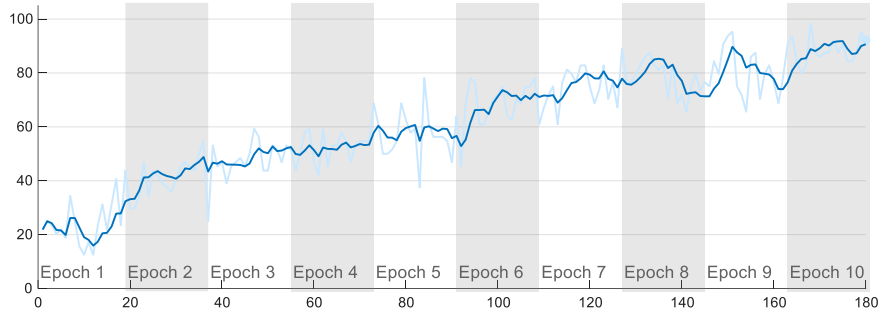


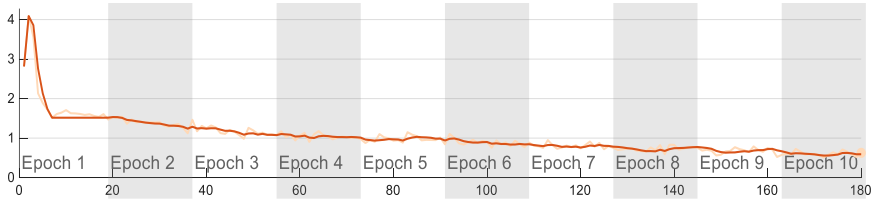
Figure 1: Confusion matrix for modulation classification

Figure.1 represents the confusion matrix for the CNN-LSTM-based modulation classification model applied to five modulation types: BPSK, QPSK, 8PSK, 16QAM, and 64QAM. The diagonal elements indicate

correctly classified instances, while the off-diagonal entries represent misclassifications. As observed, the majority of samples for each class are accurately predicted, demonstrating the model's effectiveness in distinguishing between similar modulation schemes. For example, BPSK and 64QAM signals exhibit high classification precision with minimal confusion. Slight confusion is present between adjacent modulation types such as QPSK and 8PSK or 16QAM and 64QAM, which can be attributed to their similar signal constellation structures. Overall, the confusion matrix validates the robustness of the proposed deep learning model in capturing distinguishing features from raw I/Q data, making it suitable for real-time communication system applications.



(a)



(b)

Figure 2: (a):Accuracy vs. Iteration(180 Iterations)
(b):Loss vs. Iteration Curves for 10 Epochs

As shown in Figure.2, the accuracy and loss graphs are utilized to monitor and evaluate the training performance of the deep learning model over time. The accuracy curve illustrates how well the model correctly classifies the input data at each iteration or epoch, serving as a key indicator of the learning progress. Meanwhile, the loss curve represents the error between the predicted and true labels, quantifying how well the model is minimizing its prediction mistakes.

In addition, these analyse results provide essential insight into the model's convergence behavior, helping to identify issues such as underfitting, overfitting, or unstable learning. A steadily increasing accuracy and a decreasing loss across epochs typically indicate effective training and model generalization. These plots are therefore critical for diagnosing training dynamics and validating the robustness of the proposed CNN (or CNN-LSTM) architecture.

RESULTS AND DISCUSSION

The classification performance of the proposed CNN-based model was evaluated under various metrics and test scenarios. The overall accuracy on the test dataset was calculated as 94.3%, as illustrated in Figure 1, which demonstrates the iterative improvement during training. This result indicates that the model can reliably distinguish modulation types, even under low SNR conditions.

During the training process, the loss values per epoch were monitored, showing a consistent decrease across iterations. This trend confirmed the strong learning capacity of the model and revealed no significant overfitting tendencies, thereby supporting its generalization ability. The confusion matrix analysis presented in Figure 2 further highlighted the model’s effectiveness. High recognition rates were observed for simpler constellations such as BPSK and QPSK, with correct classifications of 58 and 56 instances, respectively. However, for more complex constellations such as 16QAM and 64QAM, moderate misclassifications were noted. Specifically, the correct classification rate was measured as 85% for 16QAM and 89% for 64QAM.

In addition, the analysis across SNR levels revealed that the model achieved over 95% accuracy at SNR values above 20 dB, while maintaining recognition rates between 52% and 68% under low SNR conditions (0–5 dB). These results demonstrate that the model remains functional and reliable even in noisy environments.

Overall, this study addressed the modulation recognition problem in communication systems using an AI-based approach. By leveraging a CNN architecture that directly processes I/Q components, the proposed model achieved higher classification accuracy compared to traditional methods. It

is obvious from the simulation results that the developed system can be effectively applied in software-defined radios (SDRs), spectrum analyzers, and unsupervised signal analysis within 5G/6G networks, thereby offering a practical solution for real-world communication scenarios.

Table 2: Comparison of Classification Accuracy with Existing Studies

Study	Method	Dataset / Modulations	Channel	Accuracy
Swami & Sadler (2000)	Cumulant-based statistical method	BPSK, QPSK, 16QAM	AWGN	~75% (at high SNR)
O'Shea et al. (2016)	CNN	RadioML2016.10A (11 modulations)	AWGN, fading	87–92%
Rajendran et al. (2018)	CNN + RNN (hybrid)	Low-cost sensor dataset (8 modulations)	AWGN, fading	~90%
Liu et al. (2020)	CNN-LSTM hybrid	Synthetic dataset (6 modulations)	AWGN	91.5%
Ali et al. (2020)	Autoencoder + DNN	Synthetic dataset (8 modulations)	AWGN	92%
Proposed Model (This Study)	CNN (I/Q-based)	MATLAB synthetic (BPSK, QPSK, 8PSK, 16QAM, 64QAM)	AWGN (0–25 dB)	94.3% (avg.); >95% at SNR ≥ 20 dB

Table 2 compares the performance of the proposed CNN-based model with several representative studies in the literature. While traditional statistical methods demonstrate limited accuracy in noisy environments, deep learning approaches such as CNN, RNN, and hybrid CNN-LSTM models provide notable improvements. The proposed framework outperforms these methods, achieving an average accuracy of **94.3%**, and exceeding **95%** under high SNR conditions, which highlights its robustness and practical relevance for 5G/6G and cognitive radio applications.

In this study, a deep learning-based automatic modulation classification system was developed using raw I/Q data and a hybrid CNN (and optionally LSTM) architecture. The proposed model achieved high accuracy in distinguishing between modulation types such as BPSK, QPSK,

and QAM variants, demonstrating its capability to learn robust features directly from the signal domain. The training performance, including low loss values and a stable accuracy trend across epochs, indicates the effectiveness of the learning process without significant overfitting. The confusion matrix further confirms the model's generalization ability, especially under realistic modulation conditions. These findings suggest that the proposed approach can be integrated into real-time communication systems, such as 5G/6G and cognitive radio networks, for efficient spectrum analysis and signal intelligence applications.

REFERENCE

Ali, A., Khan, F. A., & Imran, M. (2020). Automatic modulation classification using stacked sparse autoencoders. Wireless Networks, 26(3), 1817–1828. https://doi.org/10.1007/s11276-019-02215-0

Azzouz, E., & Nandi, A. K. (1996). Automatic modulation recognition of communication signals. Springer.

Di Renzo, M., Debbah, M., Phan-Huy, D.-T., Zappone, A., Alouini, M.-S., Yuen, C., … Tretyakov, S. (2019). Smart radio environments empowered by reconfigurable AI meta-surfaces: An idea whose time has come. EURASIP Journal on Wireless Communications and Networking, 2019(129), 1–20. https://doi.org/10.1186/s13638-019-1438-9

Gardner, W. A. (1987). Statistical spectral analysis: A nonprobabilistic theory. Prentice-Hall.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.

Huang, H., Yang, J., Huang, H., Song, Y., & Gui, G. (2018). Deep learning for super-resolution channel estimation and DOA estimation based on autoencoder networks. IEEE Transactions on Vehicular Technology, 67(9), 8549–8560. https://doi.org/10.1109/TVT.2018.2851783

Karra, K., Kuzdeba, S., & Petersen, J. (2017). Modulation recognition using hierarchical deep neural networks. In 2017 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN) (pp. 1–6). IEEE. https://doi.org/10.1109/DySPAN.2017.7920746

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444. https://doi.org/10.1038/nature14539

Li, R., Liu, Y., & Li, T. (2019). Genetic algorithm optimized deep neural network for robust automatic modulation classification. IEEE Access, 7, 80420–80430. https://doi.org/10.1109/ACCESS.2019.2923425

Liu, X., Zhao, J., & Liu, X. (2020). Automatic modulation classification using hybrid CNN-LSTM models. In 2020 IEEE International Conference on Communications (ICC) (pp. 1–6). IEEE. https://doi.org/10.1109/ICC40277.2020.9148651

O'Shea, T. J., Corgan, J., & Clancy, T. C. (2016). Convolutional radio modulation recognition networks. In Engineering Applications of Neural Networks (EANN 2016) (pp. 213–226). Springer. https://doi.org/10.1007/978-3-319-44188-7_16

O'Shea, T. J., & West, N. (2016). Radio machine learning dataset generation with GNU Radio. In Proceedings of the GNU Radio Conference, 1(1), 1–6.

Rajendran, S., Meert, W., Giustiniano, D., Lenders, V., & Pollin, S. (2018). Deep learning models for wireless signal classification with distributed low-cost spectrum sensors. IEEE Transactions on Cognitive Communications and Networking, 4(3), 433–445. https://doi.org/10.1109/TCCN.2018.2828340

Strinati, E. C., Barbarossa, S., Gonzalez-Jimenez, J. L., Ktenas, D., Cassiau, N., Maret, L., & De Domenico, A. (2019). 6G: The next frontier. IEEE Vehicular Technology Magazine, 14(3), 42–50. https://doi.org/10.1109/MVT.2019.2921162

Swami, A., & Sadler, B. M. (2000). Hierarchical digital modulation classification using cumulants. IEEE Transactions on Communications, 48(3), 416–429. https://doi.org/10.1109/26.837043

Wang, L., Guo, Q., & Zhang, W. (2018). Particle swarm optimization-aided deep learning framework for modulation classification. In 2018 IEEE International Conference on Communications Workshops (ICC Workshops) (pp. 1–6). IEEE. https://doi.org/10.1109/ICCW.2018.8403639

Yu, Z., Hu, J., Li, G., & Xu, Z. (2018). Automatic modulation classification using convolutional neural networks with adaptive normalization. IEEE Communications Letters, 22(12), 2457–2460. https://doi.org/10.1109/LCOMM.2018.2873692

Zeng, Y., Liang, Y.-C., Hoang, A. T., & Zhang, R. (2010). A review on spectrum sensing for cognitive radio: Challenges and solutions. EURASIP Journal on Advances in Signal Processing, 2010, 1–15. https://doi.org/10.1155/2010/381465

Zhang, Q., Li, X., & Zhang, H. (2019). Deep learning-based modulation classification for 5G and beyond. IEEE Access, 7, 114631–114640. https://doi.org/10.1109/ACCESS.2019.2935616

In Situ and Ex Situ Strategies for the Remediation of Petroleum-Contaminated Soils: Recent Advances and Emerging Trends

F. Olcay TOPAÇ[1]

1- Prof.Dr.; Bursa Uludağ Üniversitesi Mühendislik Fakültesi Çevre Mühendisliği Bölümü. olcaytopac@uludag.edu.tr ORCID No: 0000-0002-6364-4087

ABSTRACT

Petroleum hydrocarbons represent a major class of environmental pollutants due to their persistence, toxicity, and widespread occurrence resulting from exploration, transportation, and refining activities. Once released into soils, petroleum-derived compounds such as polycyclic aromatic hydrocarbons (PAHs), aliphatic hydrocarbons, and BTEX not only alter the physicochemical structure of soils but also severely reduce microbial diversity and enzymatic activity, leading to long-term ecological degradation. Effective remediation of petroleum-contaminated soils therefore requires comprehensive strategies integrating both in situ and ex situ approaches. This chapter provides an overview of the behavior of petroleum hydrocarbons in soil systems, diagnostic methods for contamination assessment, and the state-of-the-art remediation techniques. In situ approaches such as bioremediation, phytoremediation, soil vapor extraction, chemical oxidation, and soil flushing are discussed alongside ex situ strategies including soil washing, thermal desorption, composting, and slurry-phase bioreactors. Emerging trends such as the use of biochar, nanotechnology-based systems, genetically engineered microorganisms, transgenic plants, synthetic microbial consortia, and omics-driven monitoring frameworks are highlighted as promising tools to enhance degradation efficiency and sustainability. Collectively, the review emphasizes that integrating classical methods with innovative technologies offers the most effective pathway for restoring petroleum-impacted soils while simultaneously addressing ecological resilience, resource recovery, and long-term environmental health.

Keywords – Emerging remediation technologies, Environmental sustainability, In situ and ex situ remediation, Petroleum-contaminated soils, TPH removal

INTRODUCTION

Petroleum, while constituting the primary energy source of the global economy, also poses significant environmental risks throughout its lifecycle. Accidents and leakages that occur during exploration, extraction, transportation, and refining processes frequently result in severe ecological challenges. Petroleum-derived pollutants generated during these stages often lead to long-lasting contamination, particularly in sensitive environmental compartments such as soils (Falih et al., 2024). Petroleum pollution generally comprises polycyclic aromatic hydrocarbons (PAHs), aliphatic hydrocarbons, and volatile organic compounds such as benzene, toluene, ethylbenzene, and xylene (BTEX). These compounds are acknowledged as toxic, bioaccumulative, and environmentally persistent (Daâssi and Almaghribi, 2022).

Soil, as a physical, chemical, and biological matrix that sustains life, experiences disruption of its natural functioning when petroleum derivatives infiltrate its structure. Contamination reduces soil permeability, impedes oxygen transfer, and limits water infiltration (Sakshi and Haritash, 2019). This condition results in significant declines in microbial diversity and enzymatic activities. Consequently, petroleum pollution diminishes soil biomass production, disturbs microbial balance, and renders the soil biologically inactive (Alrumman et al., 2015).

Petroleum is known to spread not only in soils but also into aquatic and atmospheric environments. The transport of petroleum components through groundwater, surface waters, and the atmosphere expands the extent of contamination and elevates associated risks (Falih et al., 2024). This highlights the necessity for integrated environmental monitoring and intervention strategies. Furthermore, petroleum-derived pollutants indirectly contribute to climate change by enhancing greenhouse gas emissions.

From the perspective of human health, petroleum pollution has been associated with a wide range of disorders, including skin irritation, respiratory complications, impaired liver and kidney functions, reproductive problems, and various forms of cancer. These impacts have been shown to be more pronounced in communities residing in close proximity to oil fields, as demonstrated by several studies (Ramirez et al., 2017).

Plants are also severely affected by petroleum contamination. Hydrophobic compounds such as PAHs penetrate root systems, negatively influencing photosynthesis, water and nutrient uptake, and leading to foliar deformities as well as structural damage to cell membranes. Moreover, certain PAH compounds can transform into more toxic metabolites under UV exposure, thereby intensifying their environmental impacts (Sakshi and Haritash, 2019). In addition, long-term petroleum pollution can induce permanent alterations in the structure of soil microbial communities, consequently diminishing the natural bioremediation capacity of soils (Sutton et al., 2013).

This chapter provides a comprehensive overview of the impacts of petroleum pollution on soils, highlighting its environmental and biological consequences and establishing the scientific rationale for soil remediation. The subsequent sections present fundamental information on the behavior of petroleum derivatives in soils and approaches for assessing contamination, followed by a systematic discussion of in situ (on-site) and ex situ (off-site) remediation strategies. Physical, chemical, and biological treatment methods are examined through practical applications and recent literature, offering a broad evaluation of strategies for the rehabilitation of petroleum-contaminated soils. Moreover, attention is also directed to emerging technological innovations that are redefining future directions in petroleum-contaminated soil remediation.

BEHAVIOR OF PETROLEUM IN SOIL

Soil contamination originating from petroleum involves a complex mixture of compounds, including saturated hydrocarbons, aromatic compounds, and a variety of heteroatom-containing molecules. The behavior of these compounds in soils is largely determined by their chemical structures, molecular weights, solubilities, and the physicochemical interactions they establish with soil matrices. These behaviors are generally characterized by four fundamental processes: adsorption, volatilization, leaching, and degradation (Wang et al., 2017).

The adsorption of petroleum constituents onto soil particles is strongly influenced by the hydrophobic nature of the compounds and the organic matter content of the soil. In particular, high-molecular-weight polycyclic aromatic hydrocarbons (PAHs) exhibit strong binding to soil organic matter, thereby reducing their bioavailability and slowing down their rates of degradation. Over time, this strong sorption can lead to the formation of a so-called "residual fraction," which may persist in soils for decades and pose long-term environmental risks (Bahar et al., 2024).

The mobility of hydrocarbons in soil is largely governed by environmental conditions. Volatile components (e.g., BTEX) tend to evaporate and dissipate under warm and dry conditions, representing a significant pathway of loss; however, this process is negligible for heavier fractions such as diesel. In contrast, soluble hydrocarbons (e.g., short-chain aliphatics and low-molecular-weight aromatics) are more prone to leaching, particularly in sandy soils with low organic matter content (Ren & Zhang, 2015). This increases the risk of petroleum derivatives contaminating groundwater. Moreover, the introduction of petroleum compounds into the soil system drives microbial degradation processes that deplete oxygen, thereby altering redox conditions. Such reducing environments enhance the solubility of redox-active elements such as iron (Fe) and manganese (Mn), facilitating their mobility into groundwater (Al-Raoush et al., 2018).

In addition to these chemical alterations, petroleum constituents directly affect soil biology. Declines in microbial diversity, biomass loss, and reductions in enzymatic activities are among the most prominent impacts. Labud et al. (2007) demonstrated that microbial functions were significantly suppressed in both sandy and clay soils contaminated with petroleum. Similarly, Gospodarek et al. (2021), in a long-term study conducted on agricultural soils contaminated with motor oil, diesel, and gasoline, reported substantial reductions in key enzymatic activities such as urease, dehydrogenase, and phosphatase, emphasizing that these enzymes reflect biological deterioration. Furthermore, the study revealed that such effects persisted even four to five years after contamination, with diesel fuel exerting the most pronounced suppressive effect on soil biota. These

biological impairments not only reduce soil fertility but also constrain its natural self-remediation capacity.

Finally, environmental conditions are also critical determinants of the persistence and degradation rate of petroleum hydrocarbons in soils. Temperature, moisture level, and the availability of macronutrients (particularly nitrogen and phosphorus) directly influence microbial degradation. In cold and dry environments, microbial activity is weakened, resulting in slow degradation, whereas in warm and moist environments, especially when adequately supported, the degradation rate of aliphatic hydrocarbons increases significantly (Koshlaf and Ball, 2017). Taken together, these interactions indicate that the fate of petroleum compounds in soils depends not only on the intrinsic properties of the compounds themselves but also on the dynamic interplay between soil characteristics and environmental conditions. Therefore, the development of effective remediation strategies requires a holistic consideration of these factors.

DETECTION AND ASSESSMENT OF CONTAMINATION

The accurate detection of petroleum pollution in soils forms the basis for the development of effective remediation strategies. The process typically begins with field observations, where visual and physical indicators such as discoloration of the soil surface, the presence of an oily film, characteristic odors, and reduced permeability may suggest the presence of petroleum-derived contaminants. However, such observations provide only qualitative information; quantitative assessments of the extent and intensity of contamination require laboratory-based analyses.

One of the most widely used indicators for the presence of petroleum-based contaminants is Total Petroleum Hydrocarbons (TPH). This parameter represents the combined concentration of aliphatic and aromatic hydrocarbons and is considered a key metric in evaluating the degree of contamination at a site. Two principal approaches are commonly employed in TPH analysis: spectroscopic and non-spectroscopic methods. Among the spectroscopic techniques, Fourier Transform Infrared Spectroscopy (FTIR) is frequently applied, relying on the ability of petroleum constituents to absorb infrared radiation at characteristic functional group frequencies. FTIR, which is consistent with EPA Method 418.1, provides a rapid and reliable means of analysis (Ezeani et al., 2022). However, its accuracy may be compromised in soils with high natural organic matter content (Paíga et al., 2012). Non-spectroscopic methods continue to be widely applied in TPH analyses and are often preferred for complex samples. Among these, gravimetric analysis and gas chromatography (GC) are the most prominent. In gravimetric methods, petroleum-derived compounds are extracted with a solvent, the solvent is then evaporated, and the remaining residue is weighed

to calculate the TPH content. However, the inclusion of non-petroleum organic matter in the residue can reduce the specificity of this method (Ezeani et al., 2022).

Gas chromatography, particularly when coupled with a flame ionization detector (FID) or mass spectrometry (MS), provides high sensitivity and selectivity. This technique allows the separation of hydrocarbon fractions within the range of C10–C40, thereby enabling the evaluation of both the extent of contamination and potential sources (Kuppusamy et al., 2020). Moreover, GC–MS is frequently employed for advanced applications such as determining the age of contamination and performing fingerprint analysis.

Beyond TPH, the analysis of specific compounds such as BTEX (benzene, toluene, ethylbenzene, and xylene) and PAHs (polycyclic aromatic hydrocarbons) is essential for a more detailed assessment of petroleum contamination. These compounds are of critical environmental and health concern due to their toxic and carcinogenic properties. While BTEX compounds, being volatile organics, are often analyzed in air, water, and soil samples, PAHs are more persistent and tend to accumulate in soil and sediment matrices. Both groups of compounds can generally be determined with high sensitivity using GC–MS systems.

Sampling strategy plays a decisive role in ensuring the reliability of analytical results. To evaluate the horizontal and vertical distribution of contaminants, samples should be collected from different depths and stored in glass containers at low temperatures. Prior to analysis, soil samples are extracted with solvents. Commonly applied extraction methods include ultrasonic extraction, Soxhlet extraction, and liquid–liquid extraction (Kuppusamy et al., 2020).

In conclusion, the detection of petroleum contamination cannot rely on a single analytical method. Rather, a combination of field observations, appropriate sampling, and both spectroscopic and non-spectroscopic analytical techniques ensures a scientifically robust evaluation of contamination levels. Furthermore, the targeted determination of specific pollutants such as BTEX and PAHs plays a critical role in assessing environmental and toxicological risks.

REMEDIATION METHODS FOR PETROLEUM CONTAMINATION

The remediation of petroleum-contaminated soils encompasses a variety of techniques aimed at either removing contaminants from the environment or transforming them into less harmful forms through physical, chemical, or biological processes. The selection of an appropriate method depends on multiple factors, including the level of contamination, soil type, groundwater depth, environmental conditions, and economic considerations. In general,

remediation strategies are classified into two main categories: in situ (on-site) and ex situ (off-site) approaches. In this section, in situ methods will be addressed first, followed by a discussion of ex situ applications.

In-Situ Remediation Methods

In situ methods rely on treating contaminated soils directly at the site without excavation. These techniques are often favored due to their lower cost and reduced environmental disturbance. However, their effectiveness is highly dependent on site-specific conditions such as soil permeability, contaminant type, and depth of pollution. The major in situ remediation techniques are presented below under subheadings, with recent studies summarized.

In-Situ Bioremediation

Bioremediation is a sustainable and environmentally friendly approach that utilizes naturally occurring or externally introduced microorganisms to transform petroleum hydrocarbons into less harmful compounds. This method is particularly effective in the degradation of light- and medium-weight hydrocarbons (Mekonnen et al., 2024).

Bioremediation typically relies on two fundamental strategies:

Biostimulation: This involves optimizing environmental conditions (C/N/P ratio, pH, moisture, temperature, oxygen) and supplementing nutrients to enhance the activity of indigenous microorganisms. When adequate nutrients and favorable conditions are provided, microbial degradation accelerates significantly, leading to high levels of Total Petroleum Hydrocarbon (TPH) removal.

Bioaugmentation: When native microbial communities in the soil lack sufficient degradation capacity, selected or genetically engineered microbial strains are introduced from external sources to enhance biodegradation (El-Sheshtawy et al., 2022).

Numerous studies in the literature have demonstrated the effectiveness of these approaches in the biological removal of petroleum contamination. For example, Oualha et al. (2019) applied a bioaugmentation strategy based on the use of *Bacillus sonorensis*. The experiment was conducted in petroleum-contaminated soil previously exposed to environmental conditions, and site parameters were optimized to enhance enzymatic activity and support microbial adaptation. The optimum conditions were determined as a C/N/P ratio of 100/10/1, temperature of 37 °C, moisture content of 10%, and a surfactant (Tween 80) concentration of 0.12%. Under these conditions, after 160 days, total petroleum hydrocarbons (TPH) decreased by 39.2% and polycyclic aromatic hydrocarbons (PAHs) by 32.4%, while microbial diversity was maintained. Li et al. (2022) demonstrated the applicability of bioremediation under extreme desert conditions by inoculating petroleum-contaminated soils in the Evrona Nature

Reserve with indigenous biosurfactant-producing *Pseudomonas* strains. This approach led to a rapid 99.2% reduction in soil hydrophobicity within seven days, with the introduced strains persisting for at least 28 days. In addition, Zhang et al. (2019) demonstrated that the use of petroleum-degrading bacteria immobilized on biochar resulted in the removal of 65% of TPH within 30 days. Biochar not only facilitated microbial attachment but also adsorbed contaminants, thereby accelerating degradation.

Phytoremediation

Phytoremediation has emerged as an environmentally friendly, cost-effective, and practical method for combating petroleum contamination. Plant root systems contribute to hydrocarbon degradation by directly absorbing pollutants and by stimulating microbial activities in the rhizosphere. Nevertheless, the method has certain limitations; in particular, the time-consuming nature of the processes, reduced effectiveness in heavily contaminated sites, and variability in efficiency depending on plant species and environmental conditions are notable disadvantages. Therefore, phytoremediation should be considered not as a stand-alone solution but as part of an integrated approach in combination with complementary remediation techniques (da Silva et al., 2023; Abdullah et al., 2020).

Recent studies have highlighted the potential of this method more clearly. For instance, Nemati et al. (2024), in a six-month field trial using *Alhagi camelorum*, achieved a removal efficiency of up to 53.6% in total petroleum hydrocarbons (TPH), while also demonstrating the removal of heavy metals such as lead, chromium, nickel, and cadmium in the range of 45–50%. The study further revealed that the presence of plants enhanced microbial communities, thereby increasing soil biological activity. Similarly, Rubežius et al. (2025) conducted a field study using Jerusalem artichoke (*Helianthus tuberosus*), where TPH concentrations were reduced from 698 mg/kg to 146 mg/kg over two years. A remarkable aspect of this study was that the biomass produced could be processed through thermo-catalytic conversion methods to generate high-quality biofuels and by-products, positioning phytoremediation not only as a remediation technique but also as an integrated environmental solution that enables resource recovery.

In another study, Panchenko et al. (2023) compared natural and technical phytoremediation approaches in the vicinity of an oil refinery in Russia. In technical applications, where plants such as alfalfa and grasses were supported with fertilization and agronomic practices, annual removal efficiencies ranged between 72–90%. In contrast, natural phytoremediation achieved similar levels only after five years. This comparison highlights that with appropriate plant selection and technical support, phytoremediation can deliver high efficiency at the field scale within a much shorter timeframe.

Overall, recent applications of phytoremediation have demonstrated considerable success in the removal of both hydrocarbons and co-occurring

heavy metals. The use of diverse plant species, activation of microbial synergies, and valorization of the resulting biomass for energy production all enhance the versatility of this method. These findings suggest that when combined with innovative approaches and supportive applications, phytoremediation can play a significant role in the restoration of contaminated soils.

Soil Vapor Extraction (SVE)

Soil Vapor Extraction (SVE) is an in situ remediation technology primarily used for the removal of volatile organic compounds (VOCs) and certain semi-volatile organic compounds (SVOCs) from unsaturated soil zones. This method relies on applying negative pressure through vertically or horizontally installed wells, which induces the volatilization of contaminants from the soil matrix. The vapor-phase pollutants are then transported to the surface, where they are captured using filter systems such as activated carbon. SVE is particularly effective in highly permeable and homogeneous soils and is widely applied for the removal of light petroleum products such as gasoline, diesel, and kerosene. However, the efficiency of the method strongly depends on soil properties such as moisture content, organic matter content, and permeability, as well as on the chemical characteristics of the contaminants, particularly their volatility (Ding et al., 2022).

In a field study conducted by Labianca et al. (2020) in the Taranto region of Italy, a full-scale SVE system was applied for four years to a petroleum hydrocarbon-contaminated site, resulting in a 73% reduction in the concentration of volatile compounds in the soil. However, at some sampling points, concentrations remained above regulatory thresholds, necessitating an extended remediation period. The study clearly demonstrated that soil texture, moisture content, permeability, and contaminant type are key factors influencing remediation efficiency.

Another investigation of SVE efficiency in relation to soil type was carried out by Nematollahi et al. (2018). In this study, SVE tests were applied to soils contaminated with kerosene across three different soil textures: pure fine sand, and fine sand amended with 20% and 40% kaolinite clay. The findings revealed that while hydrocarbon removal exceeded 78% in pure sand, the efficiency decreased to 53.1% and 54.6% with increasing clay content. This decline was attributed to restricted airflow in low-permeability soils, which hinders contaminant volatilization into the vapor phase.

In cases where SVE alone does not achieve sufficient contaminant removal, research has focused on combining it with other technologies to enhance its effectiveness. Ding et al. (2022) developed the "Thermally-Enhanced SVE (TOSVE)" system for low-permeability soils contaminated with halogenated hydrocarbons, by supplementing SVE with thermal oxidation. The use of scoria, a volcanic material, improved exhaust gas

treatment efficiency, demonstrating enhanced removal performance under challenging soil conditions. Similarly, Ma et al. (2016) applied SVE in combination with bioremediation to soils heavily contaminated with petroleum drilling wastes. While SVE alone achieved up to 23.6% TPH (Total Petroleum Hydrocarbons) removal, the subsequent addition of biostimulants and microbial consortia increased the overall removal efficiency to 70%. This success was attributed to the biotransformation of less-volatile compounds into more volatile derivatives, which could then be extracted via SVE. In another innovative study, Jang et al. (2024) developed the PPV (Pneumatic Fracturing – Plasma Blasting – Vacuum Extraction) system to improve SVE performance in low-permeability clay/silt soils. This hybrid method involved creating micro-fractures in the soil to enhance oxidant distribution and facilitate vacuum-assisted contaminant removal. After a 42-day field trial, the initial TPH concentration of 6922 mg/kg was reduced to 541 mg/kg, corresponding to approximately 92% removal. Moreover, the method proved economically attractive, with an application cost as low as 11.6 USD per ton of soil.

Taken together, these studies indicate that while SVE is an effective remediation method under suitable conditions, its performance may be limited by soil physical characteristics and contaminant properties. Therefore, integrating SVE with biological, thermal, or other physical technologies significantly improves remediation efficiency, particularly in challenging soils and in the presence of less-volatile petroleum contaminants. Evidence from both laboratory and field applications presented in the literature strongly suggests that hybrid systems will become increasingly widespread in the future, supporting the technical and economic feasibility of such approaches.

In Situ Chemical Oxidation (ISCO)

One of the technologies developed for the on-site and effective removal of petroleum-derived contaminants is In Situ Chemical Oxidation (ISCO). ISCO involves the direct injection of strong oxidants into contaminated zones to transform organic pollutants into less toxic or more biodegradable compounds. Compared to conventional remediation techniques, ISCO offers the advantage of achieving effective results in a shorter period and is particularly attractive for petroleum-based pollutants due to its field applicability.

Among the most widely used oxidants in ISCO applications is sodium persulfate ($Na_2S_2O_8$), which, when activated with Fe(II), generates sulfate radicals ($SO_4^{\bullet-}$). These radicals possess high redox potentials, enabling them to oxidize petroleum hydrocarbons such as TPH, BTEX, and PAHs. In a study conducted by Jiang et al. (2025), $FeSO_4$-activated persulfate was applied to three groups of soils with high TPH content, achieving removal efficiencies ranging from 73% to 81.5%. The best

performance was obtained at a PS:$FeSO_4$ ratio of 1:1, and it was also demonstrated that soil physicochemical properties and plant growth were not significantly affected.

In a comprehensive review by Wei et al. (2022), it was reported that ISCO methods achieved an average remediation efficiency of 65.8% in field applications. However, in some cases, efficiencies decreased to as low as 40% due to site-specific conditions. The main limiting factors included low soil permeability, adsorption onto organic matter, and the aging of contaminants. Nevertheless, combining ISCO with other methods (e.g., bioremediation or natural attenuation) improved the overall remediation efficiency to as high as 75.5%.

Chen et al. (2023) further compared the performance of different oxidant systems in TPH removal. The Fe^{2+}-activated peroxymonosulfate (PMS) system proved particularly effective for the oxidation of aromatic compounds, yielding the highest removal rates among the three systems tested (Fe^{2+}/H_2O_2, Fe^{2+}/PMS, Fe^{2+}/PDS). The efficiency of oxidants varied with pH, but the PMS system showed stable performance across a broader pH range. Similarly, Li et al. (2024) investigated a system in which persulfate was activated with magnetic Fe_3O_4 nanoparticles. This system achieved higher TPH removal efficiency (82.3%) compared to the conventional Fe^{2+}/PS system, while consuming less oxidant. Moreover, it demonstrated applicability over a wider pH range and reduced soil acidification, offering an additional advantage for sustainable field applications.

In conclusion, ISCO, particularly when implemented with persulfate-based systems, stands out as a rapid and effective technology for the removal of petroleum-derived pollutants. However, the success of the process is determined by several factors, including oxidant selection, activation method, soil characteristics, and contaminant type. Current research suggests that the limitations of ISCO can be addressed through appropriate pretreatment measures and the integration of multiple technologies, thereby enabling more sustainable remediation outcomes.

Soil Flushing

Soil flushing is an effective in situ remediation technique that aims to mobilize petroleum-derived contaminants in soils, transfer them into the liquid phase, and subsequently remove the contaminated liquid from the environment. In this method, surfactants or solvents are typically injected into the contaminated zone, facilitating the desorption of hydrocarbons from the soil matrix and their transfer into solution. Due to its ability to intervene without disrupting soil structure, its applicability to large areas, and, in some cases, its relatively low cost, soil flushing is widely preferred for petroleum and oil-related contamination (Kwon et al., 2023).

However, traditional surfactant solution-based flushing approaches often face challenges such as limited propagation and low efficiency, particularly in low-permeability soils. To overcome these limitations, surfactant foam technologies developed in recent years offer significant advantages owing to their increased surface coverage, viscoelastic behavior, and prolonged stability. In a review study by Karthick et al. (2019), it was emphasized that surfactant foams provide more effective results compared to conventional solutions by reducing surface tension and enhancing solvent diffusion. These systems were also reported to use contaminant transport pathways more efficiently, achieving higher removal rates with lower surfactant consumption.

Taking this approach further, Ali et al. (2020) comprehensively evaluated surfactant foam systems stabilized with nanoparticles for the remediation of petroleum-contaminated soils. The study demonstrated that converting anionic and nonionic surfactants into foams stabilized with nanoparticles significantly enhanced both the stability and performance of the system. In particular, nano-stabilized foams facilitated the formation of microemulsions of pollutants and their effective removal from the subsurface environment, thanks to their lower surface tension and higher surface coverage. This method drew attention not only for achieving more efficient cleaning with reduced surfactant use but also for its environmental sustainability.

Another important advancement in soil flushing methods has been the addition of auxiliary agents (e.g., salts or carbon-based nanomaterials) to surfactant systems. In a field and laboratory study conducted by Bonal et al. (2018), the addition of sodium metasilicate (an inorganic salt) and multi-walled carbon nanotubes (MWCNTs) to surfactant systems for the washing of motor-oil-contaminated soils significantly improved removal efficiency. These additives further reduced surface tension and enhanced the penetration of surfactant solutions into soil pores. Geotechnical assessments also revealed improvements in the specific gravity and cohesion of contaminated soils, indicating that surfactant applications yielded both chemical and physical benefits. Furthermore, FTIR, XRD, and XRF analyses confirmed that hydrocarbon and metal residues were substantially reduced in the soil after treatment.

Taken together, these findings highlight that soil flushing is not limited to conventional surfactant solutions but can be enhanced through advances in nanotechnology, materials science, and surface chemistry, making it more effective and environmentally compatible.

Ex-Situ Remediation Techniques

Ex-situ methods are based on excavating contaminated soil and transporting it off-site, or physically separating it on-site and treating it in specialized systems. These methods are preferred because they provide

greater control over the remediation process and can be flexibly applied to various types of contaminants. However, when excavation, transportation, facility installation costs, and environmental disturbances (e.g., dust generation, traffic load, site disruption) are considered, they tend to be more expensive and intrusive compared to in-situ techniques. Nevertheless, they offer an effective solution in cases of deep contamination or when in-situ methods are insufficient. Below, the main ex-situ remediation techniques are summarized, with emphasis on recent applications.

Soil Washing

Soil washing is an effective ex-situ remediation technique that relies on the removal of petroleum-derived contaminants from soils using physical and/or chemical solvents. The fundamental principle of this method is to detach contaminants from soil particles and transfer them into the liquid phase, which is subsequently separated and removed. Due to its relatively faster results compared to conventional biological methods, particularly for the removal of heavy fractions and low-solubility compounds such as polycyclic aromatic hydrocarbons (PAHs), soil washing has gained considerable attention (Kuppusamy et al., 2017). Its advantages include high efficiency, relatively short treatment duration, and applicability to different contaminant types, while its drawbacks include the need to treat secondary wastewater generated during the washing process, low efficiency in fine-grained fractions, and high operational costs (Caetano et al., 2024). Nevertheless, significant progress has been achieved in recent years through the development of new solvents, additives, and optimization techniques.

Recent studies have demonstrated the effectiveness of soil washing for petroleum hydrocarbons. In a study on motor-oil-contaminated soils, the performance of different surfactant systems was compared, and it was shown that the composition of the surfactant played a decisive role in hydrocarbon removal efficiency (Zoghi & Mafigholami, 2023). Notably, nonionic surfactants were reported to achieve higher solubilization efficiency than ionic surfactants, significantly increasing the overall removal rate by enhancing contaminant solubility. In a laboratory study targeting petroleum-derived PAH removal, the effects of particle size and solvent type were investigated. It was emphasized that removal efficiency dropped below 30% in fine fractions, whereas up to 70% removal could be achieved in sandy fractions (Ayele et al., 2020).

In another study on petroleum-contaminated soils, the combined use of rhamnolipid biosurfactant and Fe-Cu nanoparticle suspension resulted in high removal efficiencies. Up to 84% petroleum removal was achieved, with optimum conditions identified as a rhamnolipid-to-nanoparticle ratio of 10:1, pH 7, and 60 minutes of treatment. Furthermore, rhamnolipid outperformed other biosurfactants and synthetic surfactants both in the presence and absence of nanoparticles. The reusability of the nanoparticles was another

key finding supporting the sustainability of this approach (Vu & Mulligan, 2023).

A further study on highly petroleum-contaminated silty loam soils compared conventional nonionic surfactants (Triton X-100, Brij 30) with core-crosslinked amphiphilic polymer nanoparticles (CCAPs). Results showed that surfactants generated wastewater with high turbidity containing petroleum particles during washing, whereas CCAP nanoparticles separated petroleum more effectively from soils, resulting in lower turbidity and higher removal efficiency. These findings demonstrate that nanoparticle-based approaches can provide a more sustainable alternative to conventional surfactants in soil washing applications (Kim et al., 2019). Similarly, recent studies have also highlighted the integration of soil washing with biological approaches. In these hybrid applications, petroleum-derived contaminants mobilized after surfactant-assisted washing were subjected to microbial degradation, achieving 20–30% higher overall removal efficiencies compared to washing alone. This integration was particularly effective in enhancing PAH removal, which is otherwise resistant to biodegradation, and reduced the burden of secondary wastewater generation, thereby strengthening the environmental sustainability of the method (Srivastava et al., 2022).

These findings demonstrate that soil washing can be successfully applied not only in laboratory settings but also under field conditions, with parameters such as surfactant type, particle size distribution, and solvent composition playing a critical role in outcomes. Moreover, the use of biodegradable surfactants and eco-friendly solvents enhances the sustainability of the method and reduces the risk of secondary pollution. Taken together, it can be concluded that when properly optimized and supported with auxiliary techniques, soil washing provides an effective, applicable, and environmentally sustainable option for the remediation of petroleum-contaminated soils.

Thermal Desorption (TD)

Thermal desorption (TD) is an ex-situ technique with high efficiency potential that enables the remediation of petroleum hydrocarbon–contaminated soils through thermal treatment. In this method, contaminants are volatilized by thermal energy and usually removed from the system with the aid of a carrier gas. Critical process parameters include temperature, duration, soil type, particle size, moisture content, and contaminant composition. The main advantages of TD are its rapid applicability and high removal efficiency. However, the high energy demand, loss of organic matter, and potential disruption of soil ecological functions represent important limitations of the method (Choi et al., 2020). Therefore, the applicability of TD should be evaluated not only in terms of contaminant removal rates but also regarding its long-term effects on soil functions.

Numerous studies have demonstrated the high effectiveness of TD in the removal of petroleum hydrocarbons from soils. Wang et al. (2023) investigated the effects of thermal desorption treatment on ecotoxicity in soils highly contaminated with crude oil. Their findings showed that treatment at 400 °C for 60 minutes minimized acute (bioluminescence EC50), subacute (Brassica juncea germination and biomass), and chronic (microbial diversity) toxicity. However, at higher temperatures (500–600 °C), although up to 99% removal of contaminants was achieved, ecotoxicity was not reduced due to the formation of pyrolysis products and deterioration in soil properties. Thus, not only removal efficiency but also ecotoxicological impacts emerge as critical criteria in optimizing TD.

Similarly, Bykova et al. (2021) investigated the properties of petroleum-contaminated soils in different climatic zones and evaluated their treatability using thermal desorption. The study encompassed Arctic ecosystems, boreal forests, and temperate grassland–woodland ecotones. Results revealed petroleum hydrocarbon concentrations ranging from 2000–3000 mg/kg to over 5000 mg/kg, exceeding the permissible limit of 1000 mg/kg by 2 to 25 times. Experimental results showed that thermal treatment within the range of 25–250 °C successfully reduced contaminant concentrations to acceptable levels. Moreover, this temperature range remained below the combustion point of humic substances (≈450 °C), indicating environmental safety. This finding suggests that TD can provide a sustainable remediation approach under different ecosystem conditions while preserving soil fertility.

In another study, Pashkevich and Bykova (2022) developed the low-temperature thermal desorption (LTTD) method for remediating petroleum contamination at mineral resource industry sites. Both field-contaminated soils and artificially spiked soils with petroleum products such as gasoline, diesel, and motor oil (2000–200,000 mg/kg) were treated at 150–250 °C. At 250 °C, petroleum hydrocarbons were reduced to acceptable levels, while approximately 50% of soil organic carbon was preserved. Thus, effective contaminant removal was achieved without complete loss of soil fertility. Mowrer et al. (2022) examined the effects of TD on soil fertility in crude-oil-contaminated soils. Treatments carried out at different temperatures (65–550 °C) and conditions initially caused increases in soil pH (above 8.5 at 550 °C) and significant chemical alterations that restricted plant growth. However, after wetting–drying cycles, these negative effects diminished considerably, with increased plant-available phosphorus and potassium levels and reduced salt content. These results indicate that although TD may exert adverse short-term impacts, it can create favorable conditions for revegetation in the long term.

In conclusion, thermal desorption is effective for hydrocarbon removal, but its sustainability depends on protecting soil functions. Moderate temperatures (250–400 °C) achieve efficient removal with lower ecological

risks, while very high levels (>500 °C) maximize elimination but can cause severe soil degradation. Therefore, practical applications should aim for temperature ranges that ensure both remediation success and long-term soil resilience.

Composting and Solid-Phase Bioremediation

Composting is a prominent, environmentally friendly, and cost-effective ex-situ method for the biological remediation of petroleum hydrocarbon–contaminated soils. In this process, microorganisms—particularly bacteria and fungi—utilize hydrocarbon compounds as sources of carbon and energy, thereby driving biological degradation. Under aerobic conditions, bacteria degrade total petroleum hydrocarbons (TPHs) into carbon dioxide and water, while fungi contribute by secreting oxidative enzymes. Interactions among these different microbial groups establish a robust and sustainable degradation system. During composting, several phases can be distinguished: mesophilic, thermophilic, cooling, and maturation, with the highest biodegradation rates generally occurring in the thermophilic phase. The efficiency of the process is strongly influenced by factors such as the initial contaminant concentration, soil type, soil-to-compost ratio, C/N ratio, pH, moisture, and aeration conditions. In addition to offering high removal efficiencies, composting also attracts attention for its ability to recycle organic wastes into humus-like compost products that can be applied in agriculture and landscaping (Tran et al., 2021).

Within this framework, Sari et al. (2019) examined aerobic composting for crude oil–contaminated soils from the Wonocolo oil fields. By adding garden waste and rumen residues, TPHs were reduced by 74.7% after 150 days, with the highest degradation in the C13–C20 aliphatic fraction. Syawlia and Titah (2021) tested aerobic co-composting of hydrocarbon-contaminated soils from ship dismantling sites using kitchen and slaughterhouse wastes. Over 60 days, organic amendments enhanced oxygen diffusion, microbial activity, and hydrocarbon removal, with rumen waste performing best. Increased bacterial populations confirmed the potential of biodegradable wastes to boost co-composting efficiency and support effective bioremediation.

In another approach, Zhang et al. (2021) investigated a mesophilic aerobic biopile system for petroleum-contaminated soils and found bioavailable organic carbon (BAC) to be the main driver of microbial activity and TPH degradation. As BAC increased, microbial abundance and diversity improved removal efficiency, while low aeration stimulated metabolism and reduced energy demand. Thus, mesophilic biopiles, with proper BAC, temperature, and aeration control, provide an effective and economical remediation strategy.

Parnian et al. (2022) conducted a full-scale co-composting study to remediate 1200 m^3 of saline soils contaminated by petroleum activities in Iran. Using organic wastes from a sugarcane factory, urea, sugar, and

compost, soils with initial TPH levels of 6.9–17.1 g/kg were treated. After three months of irrigation and aeration, concentrations dropped to 0.068–0.109 g/kg, corresponding to about 99% removal. Gas chromatography confirmed soil quality for ecosystem reintegration. The study showed co-composting as an efficient, sustainable, and scalable solution that recycles industrial by-products.

All these studies demonstrate that composting and co-composting approaches can achieve substantial removal of petroleum hydrocarbons through the use of different organic amendments and controlled environmental parameters. The findings emphasize that these methods are flexible and scalable, while their integration with sustainable waste management practices further enhances ecosystem recovery potential.

Slurry-Phase Bioreactors

Slurry bioreactors are an emerging technology that has attracted increasing attention for the bioremediation of petroleum hydrocarbon–contaminated soils. These systems enhance the degradation processes by ensuring homogeneous mixing of contaminants, microorganisms, and nutrients. Moreover, the ability to control parameters such as temperature, pH, nutrient levels, and oxygen transfer allows microbial activities to proceed under optimal conditions. In this way, slurry bioreactors overcome the natural limitations of bioremediation, providing both high efficiency and operational flexibility. Nevertheless, challenges remain, including soil pretreatment, solid–liquid separation after treatment, and the reuse of nutrient-rich effluents. Despite these limitations, recent studies have demonstrated that slurry bioreactors represent a promising, scalable, and environmentally sustainable option for petroleum hydrocarbon remediation (Gaikwad and Gaikwad, 2019).

Recent advances in slurry bioreactors highlight innovative approaches that go beyond classical bioremediation strategies. For instance, Fanaei et al. (2023) combined hydrogen peroxide (H_2O_2) biostimulation with bioaugmentation using petroleum-degrading and biosurfactant-producing bacteria, achieving complete removal of 32 g/kg TPH within only six days. This rapid and highly efficient performance was attributed to the ability of H_2O_2 to stimulate bacteria for peroxidase and biosurfactant production, thereby accelerating degradation. This study is particularly noteworthy as it demonstrates the synergistic application of chemical stimulation and microbial enhancement in slurry bioreactors. Similarly, Zhang et al. (2019) reported up to 58% TPH removal by immobilizing petroleum-degrading bacteria onto biochar, with the fastest biodegradation observed in short-chain n-alkanes. The addition of biochar not only enhanced soil enzyme activities and microbial diversity but also simultaneously provided biostimulation and bioaugmentation effects. Together, these studies illustrate that hybrid approaches have strong potential to improve the efficiency and rate of slurry bioreactor processes.

Slurry systems have also been explored for the application of biostimulation using biofertilizers. Girigiri et al. (2019) employed nitrogen-fixing (Azotobacter, Rhizobium) and phosphate-solubilizing (Pseudomonas, Bacillus) bacteria as biofertilizers to remediate crude oil–contaminated soils. After four weeks, TPH reductions of 92–98% were achieved, alongside balanced soil pH and improved nutrient availability, which supported maize plant growth in ecotoxicity tests. This study demonstrated that biofertilizer-based biostimulation is a viable option for contaminated soil remediation and could be applied similarly within slurry-phase systems, underscoring its potential as an alternative biostimulation strategy.

Miyoshi et al. (2020) tested a rotating slurry bioreactor for petroleum-contaminated soils at 15 °C. With Rhodococcus bioaugmentation, over 95% of A-fuel oil degraded in six days. Even under semi-continuous operation without further inoculation, 94% efficiency was maintained. The study shows that slurry bioreactors can stimulate both added and native microbes, offering rapid, low-temperature bioremediation when supported by proper reactor design and operation.

Overall, recent research on slurry bioreactors highlights that the integration of strategies such as biostimulation, bioaugmentation, chemical stimulation, carrier materials, and reactor design can significantly enhance performance. These advancements show that slurry-phase bioreactors have the potential to deliver high efficiency and flexibility not only at the laboratory scale but also in practical, field-level applications.

EMERGING TRENDS IN REMEDIATION OF PETROLEUM-CONTAMINATED SOILS

This section addresses the emerging trends in the remediation of petroleum-contaminated soils (Figure 1).

Genetic engineering techniques have emerged in recent years as a powerful tool to enhance the bioremediation capacity of microorganisms. While most natural microorganisms are able to partially degrade specific contaminants, genetically modified microorganisms (GMMs) achieve more effective removal of petroleum hydrocarbons and other organic pollutants by increasing the activity of target enzymes (e.g., dioxygenases and cytochrome P450 enzymes) or by reorganizing metabolic pathways. Indeed, in a comprehensive study conducted by Tomar et al. (2024), strains of *Pseudomonas putida, Bacillus subtilis*, and *Escherichia coli* were improved through CRISPR-Cas9-based genetic modifications and exhibited superior performance in organic contaminant removal compared to natural strains. The study reported particularly notable improvements in the degradation efficiency of *P. putida* and *B. subtilis*, while genetically modified *E. coli* strains revealed upregulated expression patterns of genes associated with contaminant degradation. These findings demonstrate that genetic

engineering provides sustainable and targeted solutions for the design of microbial strains applicable to environmental restoration. Similarly, Su et al. (2025) genetically engineered *Vibrio natriegens*, known for its rapid growth capacity, and achieved successful removal of different aromatic compounds (e.g., phenol, naphthalene, toluene), thereby demonstrating that GMMs can also be effective under multi-contaminant and challenging environmental conditions.

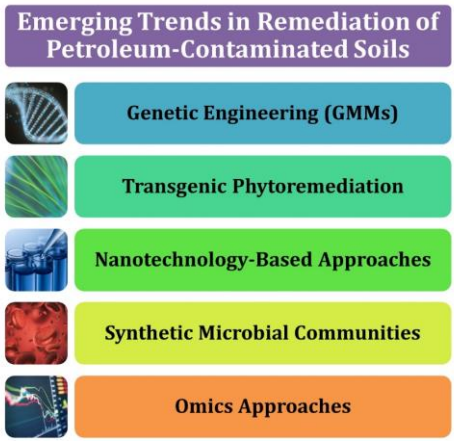


Figure 1: Emerging trends in the Remediation of Petroleum-Contaminated Soils

In recent years, phytoremediation supported by transgenic plants has emerged as a significant trend in the remediation of petroleum-contaminated soils. While natural plants are able to absorb polycyclic aromatic hydrocarbons (PAHs) derived from petroleum to a certain extent, their capacity to metabolize these compounds is generally limited. Through genetic engineering, the transfer of genes encoding detoxification enzymes into plants overcomes this limitation by enabling more efficient degradation and transformation. For example, in a study using transgenic *Arabidopsis thaliana*, Hernández-Vega et al. (2021) demonstrated the critical role of the AtDAO1 enzyme in the detoxification of phenanthrene, a representative PAH. This finding indicates that transgenic plants not only enhance the removal of petroleum-derived contaminants from soil but also increase their potential to transform them into non-toxic metabolites, suggesting that this approach could be extended to larger-scale applications against petroleum pollution in the future.

Nanotechnology-based approaches represent a distinct rising trend in the bioremediation of petroleum-contaminated soils. Highly reactive nanomaterials with large surface areas enhance the bioavailability of petroleum-derived pollutants (particularly PAHs), thereby accelerating microbial degradation; when combined with appropriate surfactants, this

effect is further strengthened. For instance, Lv et al. (2022) reported that a combination of rhamnolipid, nZVI, and the electron shuttle AQDS significantly improved the microbial anaerobic–aerobic degradation of PAHs in soil, with key enzyme activities such as dehydrogenase and catechol 1,2-dioxygenase being restored. These findings highlight that nano-bioremediation (the integration of nanomaterials and biological processes) has emerged as a powerful and applicable trend in the remediation of petroleum-contaminated soils.

Another recent innovation in bioremediation is the use of synthetic microbial communities (SynComs). Designed to overcome the limitations of natural consortia, these communities enable the synergistic functioning of microorganisms with different metabolic capacities, thereby allowing for the degradation of petroleum hydrocarbons across a broader range and at a faster pace. On this topic, Qattan (2025) emphasized that bacterial consortia are effective in the removal of both hydrocarbons and other persistent pollutants, and that synthetic biology-based designs will play a central role in the future of bioremediation. Moreover, Silva-Portela et al. (2025), through comprehensive transcriptomic analyses of a selected microbial consortium (MC1) in petroleum waste-contaminated environments, demonstrated that such communities adapt to hydrocarbon degradation at the gene expression level, enhancing their pollutant degradation capacity. These findings indicate that SynComs represent a far more effective and promising strategy than classical single-strain approaches for the remediation of petroleum-contaminated soils.

One of the most important trends shaping the future of bioremediation in recent years has been multidisciplinary omics approaches. Metagenomic analyses facilitate the discovery of new genes and enzymes involved in hydrocarbon degradation, while metatranscriptomic studies reveal the environmental conditions under which these genes are activated in contaminated soils. Metabolomic data, in turn, enhance process traceability by identifying intermediate products formed during degradation. For example, Li et al. (2023), through metagenomic and metabolomic analyses of petroleum-polluted soils, identified microbial groups responsible for hydrocarbon degradation and characteristic intermediates such as gentisic acid, thereby demonstrating that degradation pathways can be tracked in detail. Similarly, Pandolfo et al. (2024), by analyzing a microbial consortium obtained from hydrocarbon-contaminated soils at the metagenomic level, revealed the potential of species such as *Pseudomonas* and *Acinetobacter* for aliphatic and aromatic hydrocarbon degradation. These examples show that omics approaches not only improve pollutant removal but also elucidate the molecular basis of the process, thereby enabling bioremediation to be designed in a smarter and site-specific manner.

CONCLUDING REMARKS AND FUTURE PERSPECTIVES

The evaluations presented in this section highlight the diversity and multidimensional character of remediation strategies applied to petroleum-contaminated soils. In-situ approaches stand out for being cost-effective and less intrusive to the environment; however, their efficiency is strongly constrained by site-specific conditions and the nature of the contamination. Ex-situ methods, on the other hand, provide greater control over the process and higher removal efficiency, often yielding more reliable outcomes, particularly when hydrocarbon concentrations are high or when complex mixtures of petroleum hydrocarbons are present. Thus, depending on the contaminant type, its concentration, and the prevailing site conditions, the choice of remediation method may vary. In practice, the classical approach is often a "tailor-made" strategy, combining in-situ and ex-situ methods in hybrid or complementary configurations to achieve optimal results.

In recent years, however, emerging trends have begun to transcend the limitations of conventional approaches, opening up new avenues for soil remediation. Nanotechnology-based applications enhance pollutant solubility and bioavailability, thereby accelerating microbial degradation; genetically engineered microorganisms and transgenic plants provide enhanced degradation capacity and substrate specificity; synthetic microbial communities (SynComs) enable synergistic metabolic functions that facilitate the breakdown of hydrocarbons across a broader spectrum; and omics-driven approaches allow for molecular-level monitoring of degradation pathways, supporting the design of site-specific and targeted remediation strategies.

Taken together, these developments indicate that the most rational pathway forward lies in integrating the strengths of established in-situ and ex-situ methods with innovative, biotechnology-driven solutions. The future of petroleum soil remediation will depend on multidisciplinary research programs, site-specific optimization frameworks, and the convergence of classical and emerging trends. Such integration not only enhances technical efficiency but also ensures long-term environmental sustainability, thereby forming the foundation for the next generation of remediation strategies.

REFERENCES

Abdullah, S.R.S., Al-Baldawi, I.A., Almansoory, A.F., Purwanti, I.F., Al-Sbani, N. H., & Sharuddin, S.S.N. (2020). Plant-assisted remediation of hydrocarbons in water and soil: Application, mechanisms, challenges and opportunities. *Chemosphere*, 247, Article 125932.

Alrumman, S.A., Standing, D.B., & Paton, G.I. (2015). Effects of hydrocarbon contamination on soil microbial communities and enzyme activity. J*ournal of King Saud University –Science*, 27, 31-41..

Ali, N., Bilal, M., Khan, A., Ali, F., & Iqbal, H.M.N. (2020). Effective exploitation of anionic, nonionic, and nanoparticle-stabilized surfactant foams for

petroleum hydrocarbon contaminated soil remediation. *Science of the Total Environment,* 704, 135391.

Al-Raoush, R., Ngueleu, S., Rezanezhad, F., Van Cappellen, P. (2018). Groundwater pollution by petroleumderived contaminants in coastal semiarid environment. Qatar Foundation Annual Research Conference Proceedings 2018: EEPD709

Ayele, B.A, Lu, J., & Chen, Q.(2020). Optimization of aeration enhanced surfactant soil washing for remediation of diesel-contaminated soils using response surface methodology. *PeerJ.* 13;8:e8578.

Bahar, M.M., Samarasinghe, S.V.A.C., Bekele, D. et al. (2024). Residual hydrocarbons in long-term contaminated soils: implications to risk-based management. *Environ Sci Pollut Res* 31, 22759–22773.

Bonal, N.S., Paramkusam, B.R., & Basudhar, P.K. (2018). Enhancement of surfactant efficacy during the cleanup of engine oil contaminated soil using salt and multi-walled carbon nanotubes. *Journal of Hazardous Materials*, 351, 54–62.

Bykova, M.V., Alekseenko, A.V., Pashkevich, M.A., Drebenstedt, C. (2021). Thermal desorption treatment of petroleum hydrocarbon-contaminated soils of tundra, taiga, and forest steppe landscapes. *Environ Geochem Health*. 43(6):2331-2346.

Caetano, G., Machado, R.M., Correia, M.J.N., & Marrucho, I. M. (2024). Remediation of soils contaminated with total petroleum hydrocarbons through soil washing with surfactant solutions. *Environ Technol*. 45(15):2969-2982.

Chen, X., Mu, S., & Luo, Y. (2023). Degradation of petroleum-based pollutants in oil-based drilling cuttings using Fe^{2+}-based Fenton-like advanced oxidation processes. *Environmental Science and Pollution Research, 30*, 37669–37678.

Choi, B., Lee, S. & Jho, E.H. (2020). Removal of TPH, UCM, PAHs, and Alk-PAHs in oil-contaminated soil by thermal desorption. *Applied Biological Chemistry.* 63, 83.

da Silva, J., Rosa, G.B., Sganzerla, W.G., Ferrareze, J.P., Simioni, F.J., & Campos, M.L. (2023). Strategies and prospects in the recovery of contaminated soils by phytoremediation: An updated overview. *Communications in Plant Sciences*, 13, Article 001-012.

Daâssi, D., & Almaghribi, F.Q. (2022). Petroleum-contaminated soil: Environmental occurrence and remediation strategies. *3 Biotech*, 12, 139.

Ding, Y., Zhang, Y., Deng, Z., Song, H., Wang, J., & Guo, H. (2022). An innovative method for soil vapor extraction to improve extraction and tail gas treatment efficiency. *Scientific Reports*, 12, Article 6495.

El-Sheshtawy, H.S., Aman, D., & Nassar, H.N. (2022). A novel bioremediation technique using goethite-chitosan nanocomposites. *Soil and Sediment Contamination*, 31(2):176–199.

Ezeani, E.U., Ngobiri, N.C., & Agbagwa. I.O. (2022). A review of techniques for the determination of total petroleum hydrocarbons in soils. *Asian Journal of Applied Chemistry Research.* 11 (2):38-43.

Falih, K.T., Mohd Razali, S.F., Abdul Maulud, K.N. et al. (2024). Assessment of petroleum contamination in soil, water, and atmosphere: A comprehensive review. *International Journal of Environmental Science and Technology*, 21(13): 8803–8832.

Fanaei, F., Moussavi, G., & Shekoohiyan, S. (2023). Enhanced bioremediation of oil-contaminated soil in a slurry bioreactor by H_2O_2-stimulation of oil-degrading/biosurfactant-generating bacteria: Performance optimization and bacterial metagenomics. *Biodegradation.* 34(1):83-101.

Gaikwad, V., & Gaikwad, S. (2019). Treatability study of slurry bioreactor technology for bioremediation of petroleum contaminated soil. *EM International (Eco. Env. & Cons Journal)*. 25(3):1429-1434.

Girigiri, B., Ariole, C.N., & Stanley, H. O. (2019). Bioremediation of crude oil polluted soil using biofertilizer from nitrogen-fixing and phosphate-solubilizing bacteria. *American Journal of Nanosciences*, 5(4), 27-38.

Gospodarek, J., Rusin, M., Barczyk, G., & Nadgórska-Socha, A. (2021). The effect of petroleum-derived substances and their bioremediation on soil enzymatic activity and soil invertebrates. *Agronomy,* 11(1), 80.

Hernández-Vega, J.C., Langford, S., Hurtado, D.A., Cady, B., Kayanja, G., Okwara, N., Mauriello, A., Alkio, M., Colón-Carmona, A. (2021). Detoxification of phenanthrene by transgenic Arabidopsis thaliana: Role of AtDAO1 enzyme in phytoremediation of polycyclic aromatic hydrocarbons. *Environmental Science & Technology,* 55(14), 9874–9885.

Jang, S.B., Wong, K.T., Choong, C.E., Hyun, S., Yoon, Y., Choi, E.H., Park, N., & Jang, M. (2024). Enhancing in situ remediation of clayey soils contaminated with total petroleum hydrocarbons by combining pneumatic fracturing, plasma blasting, and vacuum extraction: A comprehensive field investigation. *Journal of Environmental Chemical Engineering*, 12, 113064.

Jiang, D., Li, T., Liang, X., Zhao, X., Li, S., Li, Y., Oh, K., Liu, H., & Cao, T. (2025). Evaluation of petroleum hydrocarbon-contaminated soil remediation technologies and their effects on soybean growth. *environments, 12*(1), 6.

Karthick, A., Roy, B., & Chattopadhyay, P. (2019). A review on the application of chemical surfactant and surfactant foam for remediation of petroleum oil contaminated soil. *Journal of Environmental Management*, 243, 187–205.

Kim, N., Kwon, K., Park, J., Kim, J., & Choi, J.W. (2019). Ex situ soil washing of highly contaminated silt loam soil using core-crosslinked amphiphilic polymer nanoparticles. *Chemosphere*, 224:212-219.

Koshlaf, E. & Ball, A.S. (2017). Soil bioremediation approaches for petroleum hydrocarbon polluted environments. *AIMS Microbiol.,* 19;3(1):25-49.

Kuppusamy, S., Maddela, N.R., Megharaj, M. & Venkateswarlu, K. (2020). Total Petroleum Hydrocarbons Environmental Fate, Toxicity, and Remediation.Springer Nature Switzerland AG, 264p.

Kuppusamy, S., Thavamani, P., Venkateswarlu, K., Lee, Y.B., Naidu, R., & Megharaj, M. (2017). Remediation approaches for polycyclic aromatic hydrocarbons (PAHs) contaminated soils: Technological constraints, emerging trends and future directions. *Chemosphere,* 168:944-968.

Kwon, J.H., Ji, M.K., Kumar, R., Islam, M.M., Khan, M.A., Park, Y.K., Yadav, K. K., Vaziri, R., Hwang, J.H., Lee, W.H., & Jeon, B.H. (2023). Recent advancement in enhanced soil flushing for remediation of petroleum hydrocarbon-contaminated soil: A state-of-the-art review. *Reviews in Environmental Science and Bio/Technology,* 22, 679–714.

Labianca, C., De Gisi, S., Picardi, F., Todaro, F., & Notarnicola, M. (2020). Remediation of a petroleum hydrocarbon-contaminated site by soil vapor extraction: A full-scale case study. *Applied Sciences,* 10(12), 4261.

Labud, V., Garcia, C., & Hernandez, T. (2007). Effect of hydrocarbon pollution on the microbial properties of a sandy and a clay soil. *Chemosphere,* 66(10), 1863–1871.

Li, Z., Rosenzweig, R., Chen, F., Qin, J., Li, T., Han, J., Istvan, P., Diaz-Reck, D., Gelman, F., Arye, G., & Ronen, Z. (2022). Bioremediation of petroleum-contaminated soils with biosurfactant-producing degraders isolated from the native desert soils. *Microorganisms*, 10(11), 2267.

Li, Y.Q., Xin, Y., Li, C., Liu, J., Huang, T. (2023). Metagenomics-metabolomics analysis of microbial function and metabolism in petroleum-contaminated soil. *Braz J Microbiol.* 54(2):935-947.

Li, Y.T., Sui, Q., Li, X., Liu, X.Y., Liu, H., Wang, Y.Q., & Du, W.Y. (2024). Remediation of diesel-contaminated soil by using activated persulfate with Fe_3O_4 magnetic nanoparticles: Effect and mechanisms. *Environ Sci Pollut Res Int., 31*, 33385–33397.

Lv, L., Sun, L., Yuan, C., Han, Y., Huang, Z. (2022). The combined enhancement of RL, nZVI and AQDS on the microbial anaerobic-aerobic degradation of PAHs in soil. *Chemosphere.* 307(Pt 1):135609.

Ma, J., Yang, Y., Dai, X., Li, C., Wang, Q., Chen, C., Yan, G., & Guo, S. (2016). Bioremediation enhances the pollutant removal efficiency of soil vapor extraction (SVE) in treating petroleum drilling waste. *Water, Air, & Soil Pollution,* 227, 465.

Mekonnen, B.A., Aragaw, T.A., & Genet, M.B. (2024). Bioremediation of petroleum hydrocarbon contaminated soil: A review on principles, degradation mechanisms, and advancements. *Frontiers in Environmental Science,* 12, 1354422.

Miyoshi, Y., Okada, J., Urata, T., Shintani, M., Kimbara, K. (2020). A rotational slurry bioreactor accelerates biodegradation of a-fuel in oil-contaminated soil even under low temperature conditions. *Microorganisms.* 20;8(2):291.

Mowrer, J., Provin, T., & Perkins, S. (2022). Long-term soil fertility changes following thermal desorption to remove crude oil are favorable to revegetation strategies. *Global Journal of Science Frontier Research*, 22(H6), 1–25.

Nemati, B., Baneshi, M.M., Akbari, H., Dehghani, R., Mostafaii, G. (2024). Phytoremediation of pollutants in oil-contaminated soils by *Alhagi camelorum*: Evaluation and modeling. *Sci Rep*. 6;14(1):5502.

Nematollahi, H., Moradi, N., Riyazi N.M., & Vahidi, H. (2018). Removal of aliphatic hydrocarbons from gas oil contaminated clay soil via soil vapor extraction. *Civil Engineering Journal,* 4(8), 1858–1865.

Oualha, M., Al-Kaabi, N., Al-Ghouti, M., Zouari, N. (2019). Identification and overcome of limitations of weathered oil hydrocarbons bioremediation by an adapted *Bacillus sorensis* strain. *J Environ Manage.* 15;250:109455.

Paíga, P., Mendes, L., Albergaria, J.T. *et al.* (2012). Determination of total petroleum hydrocarbons in soil from different locations using infrared spectrophotometry and gas chromatography. *Chem. Pap*. 66, 711–721.

Panchenko, L., Muratova, A., Dubrovskaya, E., Golubev, S., & Turkovskaya, O. (2023). Natural and technical phytoremediation of oil-contaminated soil. *Life,* 13, 177.

Pandolfo, E., Durán-Wendt, D., Martínez-Cuesta, R., Montoya, M., Carrera-Ruiz, L., Vazquez-Arias, D., Blanco-Romero, E., Garrido-Sanz, D., Redondo-

Nieto, M., Martin, M., Rivilla, R. (2024). Metagenomic analyses of a consortium for the bioremediation of hydrocarbons polluted soils. *AMB Express.* 28;14(1):105.

Parnian, A., Parnian, A., Pirasteh-Anosheh, H. *et al.* (2022). Full-scale bioremediation of petroleum-contaminated soils via integration of co-composting. *J Soils Sediments.* 22, 2209–2218.

Pashkevich, M.A., Bykova, M.V. (2022). Methodology for thermal desorption treatment of local soil pollution by oil products at the facilities of the mineral resource industry. *Journal of Mining Institute*. 253:49-60.

Qattan, S.Y.A. (2025). Harnessing bacterial consortia for effective bioremediation: targeted removal of heavy metals, hydrocarbons, and persistent pollutants. *Environ Sci Eur*. 37, 85.

Ramirez, M.I., Arevalo, A.P., Sotomayor, S., Bailon-Moscoso, N. (2017). Contamination by oil crude extraction—refinement and their effects on human health. *Environmental Pollution*, 231, 415–425.

Ren, X. & Zhang, S. (2015). Vertical distribution of typical petroleum pollutants in an oilfield of Gudao, China. Proceedings of the 3rd International Conference on Management Science, Education Technology, Arts, Social Science and Economics.1242-1246.

Rubežius, M., Kidikas, Ž., Kick, C., & Kasiulienė, A. (2025). Phytoremediation of total petroleum hydrocarbons-contaminated soils: Case study of Jerusalem artichokes with cost analysis and biomass conversion. *Agronomy,* 15, 601.

Sakshi, S.S.K., & Haritash, A.K. (2019). Polycyclic aromatic hydrocarbons: Soil pollution and remediation. *International Journal of Environmental Science and Technology,* 16, 6489–6512.

Sari, G.L., Trihadiningrum, Y., & Ni'matuzahroh, (2019). Bioremediation of Petroleum Hydrocarbons in Crude Oil Contaminated Soil from Wonocolo Public Oilfields using Aerobic Composting with Yard Waste and Rumen Residue Amendments, *J. Sustain. Dev. Energy Water Environ. Syst.,* 7(3): 482-492.

Silva-Portela, R.C.B., Minnicelli, C.F., Freitas, J.F., Fonseca, M.M.B., Lima Silva, D.F., Silva-Barbalho, K.K., Falcão, R.M., Bruce, T., Cavalcante, J.V.F., Dalmolin, R.J.S., Agnez-Lima, L.F. (2025).Unlocking the transcriptional profiles of an oily waste-degrading bacterial consortium. *J Hazard Mater*. 5;485:136866.

Srivastava, V., Puri, M., Srivastava, T., Nidheesh, P.V., & Kumar, M.S. (2022). Integrated soil washing and bioreactor systems for the treatment of hexachlorocyclohexane contaminated soil: A review on enhanced degradation mechanisms, and factors affecting soil washing and bioreactor performances. *Environ Res.* 15;208:112752.

Su, C., Cui, H., Wang, W., Liu, Y., Cheng, Z., Wang, C., Yang, M., Qu, L., Li, Y., Cai, Y., He, S., Zheng, J., Zhao, P., Xu, P., Dai, J., & Tang, H. (2025). Bioremediation of complex organic pollutants by engineered Vibrio natriegens. *Nature.* 642(8069):1024-1033.

Sutton, N.B., Maphosa, F., Morillo, J.A., Abu Al-Soud, W., Langenhoff, A.A., Grotenhuis, T., Rijnaarts, H.H., & Smidt, H. (2013). Impact of long-term diesel contamination on soil microbial community structure. *Appl Environ Microbiol.* 79(2):619-30.

Syawlia, R.M., & Titah, H.S. (2021). Removal of hydrocarbons from contaminated soils using bioremediation by aerobic co-composting methods at ship dismantle locations. *Journal of Ecological Engineering*, 22(6), 181-190.

Tomar, R., Rusho., M. A., & Kumar, S., Madasamy, M., & ul Islam, A. (2024). Genetic engineering of microbes for enhanced bioremediation of organic pollutants *African Journal of BiologicaL Sciences ,* 6, 5: 8777-8793.

Tran H.T., Lin, C., Bui, X.T., Ngo, H.H., Cheruiyot, N.K., Hoang, H.G., & Vu, C.T. (2021). Aerobic composting remediation of petroleum hydrocarbon-contaminated soil. Current and future perspectives. *Sci Total Environ.* 20;753:142250.

Vu, K.A., & Mulligan, C.N. (2023). Remediation of oil-contaminated soil using Fe/Cu nanoparticles and biosurfactants. *Environmental Technology*, 44(22), 3446–3458.

Wang, S., Cheng, F., Shao, Z., Wu, B., & Guo, S. (2023). Effects of thermal desorption on ecotoxicological characteristics of heavy petroleum-contaminated soil. *Sci Total Environ.* 20;857(1):159405.

Wang, S., Xu, Y., Lin, Z., Zhang, J., Norbu, N., & Liu, W. (2017). The harm of petroleum-polluted soil and its remediation research. AIP Conference Proceedings, 1864, 020222.

Wei, K.H., Ma, J., Xi, B.D., Yu, M.D., Cui, J., Chen, B.L., Li, Y., Gu, Q.B., He, X.S. (2022). Recent progress on in-situ chemical oxidation for the remediation of petroleum contaminated soil and groundwater. *J Hazard Mater.* 15;432:128738.

Zhang K, Wang S, Guo P, & Guo S. (2021). Characteristics of organic carbon metabolism and bioremediation of petroleum-contaminated soil by a mesophilic aerobic biopile system. *Chemosphere.* 264(2):128521.

Zhang, B., Zhang, L., & Zhang, X. (2019). Bioremediation of petroleum hydrocarbon-contaminated soil by petroleum-degrading bacteria immobilized on biochar. *RSC Advances*, 9, 35304–35311.

Zoghi, P., & Mafigholami, R. (2023). Optimisation of soil washing method for removal of petroleum hydrocarbons from contaminated soil around oil storage tanks using response surface methodology. *Sci Rep.* 13, 15457.

Assessment of Ermenek District's Potential in Terms of Renewable Energy Resources

Hüseyin KAYA[1]

Mesut KİLİT[2]

1- Lecturer Dr.; Karamanoğlu Mehmetbey University, Vocational School of Technical Sciences, Department of Mechanical and Metal Technologies, Karaman, Türkiye hskaya@kmu.edu.tr
ORCID No: 0000-0002-2714-8481.

2- Lecturer; Karamanoğlu Mehmetbey University, Vocational School of Technical Sciences, Department of Mechanical and Metal Technologies, Karaman, Türkiye mkilit@kmu.edu.tr
ORCID No: 0000-0002-3659-2779.

ABSTRACT

These Factors such as the oil crises experienced worldwide in the past years, the increasing global warming, the possibility of depletion of fossil fuel reserves over time and the environmental pollution caused by the use of fossil fuels have increased the importance of environmentally friendly and naturally renewable energy sources.

Renewable energy sources contribute positively to the economies of developing countries, especially because they do not require additional expenses other than investment and operating costs.

Ermenek district has a high potential in terms of renewable energy production thanks to its geographical and climatic advantages. In this study, the current situation of Ermenek in terms of renewable energy resources was analysed and the contribution of these resources to sustainable energy production was evaluated. It has been determined that different renewable energy resources such as solar energy, hydroelectric energy, wind energy and bio energy have a significant potential in the district.

Especially long sunshine duration and high solar radiation values make the region highly suitable for solar power plants (SPP). When the solar energy maps of Karaman province are analysed, it is seen that the northern parts of Ermenek district have the highest solar energy potential. In this context, *Ermenek Enerji*, a company operating in the region, carries out studies on solar energy systems. In addition, a small-scale solar energy facility operating within the Ermenek Government House and roof and facade type SPP applications carried out by the private sector are important examples in terms of showing the investment and interest in solar energy in the region.

In this study, the current situation and energy potential of renewable energy resources in Ermenek district are discussed. It is tried to summarise the situations that will contribute to the region from this energy.

Keywords – Energy potential, Ermenek, Solar Energy, Sustainability, Renewable Energy.

INTRODUCTION

Factors such as global oil crises, increasing global warming, the potential depletion of fossil fuel reserves over time, and environmental pollution caused by fossil fuel usage have enhanced the importance of environmentally friendly and naturally renewable energy sources. The transition to renewable energy sources not only plays a critical role in combating climate change but also offers the potential to enhance energy security and support economic development. For energy import-dependent

countries like Turkey, the development of renewable energy sources carries strategic priority in achieving energy independence and reaching sustainable development goals [1]. In this context, Ermenek district, located in Turkey's Mediterranean Region, is considered to possess significant potential for renewable energy generation due to its geographical and climatic advantages. This study aims to examine the status of Ermenek district in terms of renewable energy resources, evaluate the contributions of these resources to sustainable energy production, and reveal the potential future benefits for the region. The main objectives of the study include analyzing the potential of solar, hydroelectric, wind, and bioenergy resources in Ermenek, evaluating the economic, social, and environmental benefits that renewable energy utilization could provide to the region, and providing recommendations that will contribute to strategic planning and investment decisions considering the findings obtained.

THE IMPORTANCE OF RENEWABLE ENERGY SOURCES GLOBALLY AND IN TURKEY

Global renewable energy capacity has reached new peaks, demonstrating the remarkable progress the world has made in transitioning to clean and sustainable energy [2]. In 2024, renewable energy sources constituted more than 92% of total power additions, and their share in global total installed power capacity rose to 46.4% [3]. This historic development is strong evidence of both the economic competitiveness and scalability of renewable energy. It triggers a clean energy revolution that could enable the world to achieve the goal of tripling renewable energy capacity by 2030 [3]. Solar energy became the primary driving force in 2024, accounting for 42% of global renewable power capacity growth [3]. As the cost of electricity generated from renewable energy sources continues to decline, renewable energy sources are not only environmentally suitable but are also becoming the most cost-effective energy source for countries worldwide [4]. Accelerating the transition to clean energy paves the way for a healthy and livable planet for current and future generations [5]. While countries using fossil fuels are vulnerable to geopolitical shocks and crises, renewable energy sources are found in all countries, and their potential has not yet been sufficiently utilized [5]. Renewable energy offers the opportunity to break free from import dependency, diversify economies, and protect against unpredictable price fluctuations of fossil fuels, while also supporting

inclusive economic growth, new job opportunities, and poverty reduction [5]. Renewable energy is the cheapest electricity option in many parts of the world, and prices are rapidly declining [5]. The transition to clean energy sources such as solar and wind provides solutions not only in combatting climate change but also in air pollution and health problems. Every dollar invested in renewable energy creates three times more jobs compared to the fossil fuel industry [5].

Turkey is among the OECD countries with the fastest-growing energy demand and meets 74% of its energy needs through imports [1]. Therefore, increasing the utilization of domestic and renewable energy sources is a strategic priority in line with Turkey's goals of reducing dependency on energy imports, maximizing the use of local resources, and combating climate change [1]. As of the end of 2024, Turkey generates 59.1% of its installed electricity capacity from renewable energy sources [3]. According to International Renewable Energy Agency (IRENA) data, it ranks 11th globally and 5th in Europe [4].

This rapid progress demonstrates Turkey's potential to achieve the goal of significantly increasing the renewable energy share by 2035, as determined in Turkey's National Energy Plan [10]. Increasing renewable energy production has helped Turkey reduce natural gas imports and thereby achieve significant economic savings [9]. Turkey's emphasis on renewable energy and rapid growth in this field positions the country as a regional leader [10].

GEOGRAPHICAL AND CLIMATIC CHARACTERISTICS OF ERMENEK DISTRICT

Ermenek is a district center affiliated with Karaman Province in Turkey's Mediterranean Region [11,14]. The district constitutes the center of the Taşeli Plateau located in the middle section of the Taurus Mountains [12]. Topographically possessing a rugged structure, Ermenek is surrounded by steep slopes composed of jointed limestones, ranging from 1250 meters to 1850 meters in elevation [13].

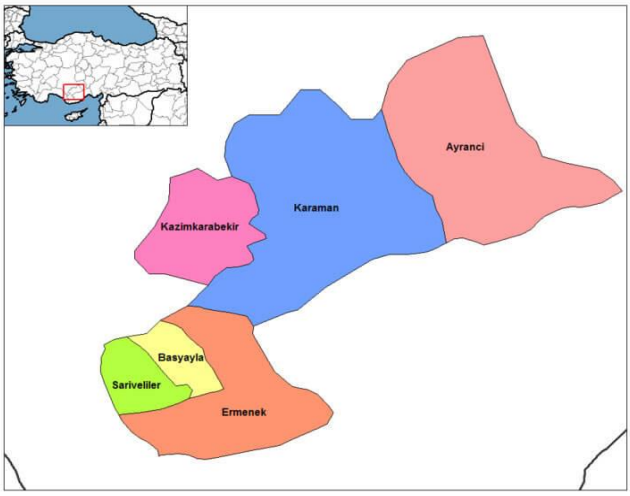


Figure 1: Map of Karaman province and districts [14]

Ermenek Stream, a tributary of the Göksu River, passes through the district boundaries and contains the Ermenek Dam, which harbors significant hydroelectric potential [15]. The geological structure and steep slopes of the region have caused rockfall events in the past [13].

A typical continental climate prevails throughout Karaman Province; winters are cold and harsh, while summers are hot and dry. However, characteristics of the Mediterranean Climate are also observed in the plain areas around the Göksu River [16]. While temperatures can drop significantly during winter months, average temperatures hover around 30°C during summer months [16]. Annual average precipitation is relatively low, varying between 300 and 450 mm [16].

Karaman Province possesses high radiation intensity of 1660 kWh/m²-year, the highest sunshine duration of 3007 hours/year, and the highest clarity index of 0.54 [17]. Therefore, it can be stated that Ermenek District also has solar energy potential with similar values. Even in January, the average duration of clear and sunny skies in Ermenek is approximately 10.9 hours [18]. Karaman ranks among the top in Turkey in terms of solar energy potential with its high Global Solar Radiation and sunshine duration [19]. The part of Karaman located in the Mediterranean Region has the highest solar energy levels in Turkey [20]. Sources such as the National Solar Radiation Database and Turkey Solar Energy Potential Atlas (GEPA) provide detailed solar energy maps [21] and data for the region [2].

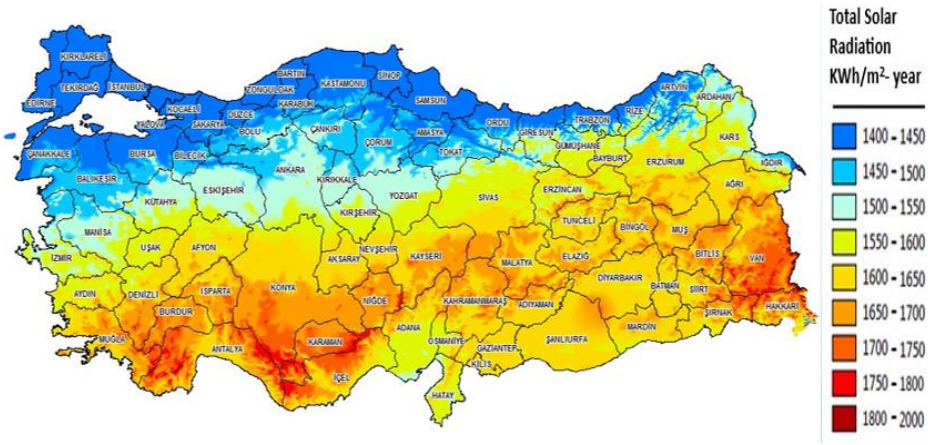


Figure 2: Turkey solar radiation potential map [22]

RENEWABLE ENERGY RESOURCE POTENTIAL OF ERMENEK DISTRICT

Ermenek's long sunshine duration and high solar radiation values make the region extremely suitable for solar energy investments [19]. According to solar energy maps, the northern parts of Ermenek district are observed to have the highest solar energy potential within Karaman province. Karaman Province stands out with its high annual average sunshine duration and total solar radiation [23]. Tools such as the Turkey Solar Energy Potential Atlas (SEPA) and Global Solar Atlas provide detailed solar energy resource maps and data for the region [24]. Turkey's solar energy potential maps and GIS data can be downloaded from platforms such as Solargis [24].

The Ermenek Dam located on the Ermenek River possesses significant hydroelectric energy generation capacity (Figure 3) [15,25]. The dam, with an installed capacity of 302.4 MW, began water impoundment in 2009, and its annual average production is around 1048 GWh [2], capable of meeting a significant portion of Karaman's electricity needs [19].

Figure 3: Ermenek dam and hydroelectric power plant [25]

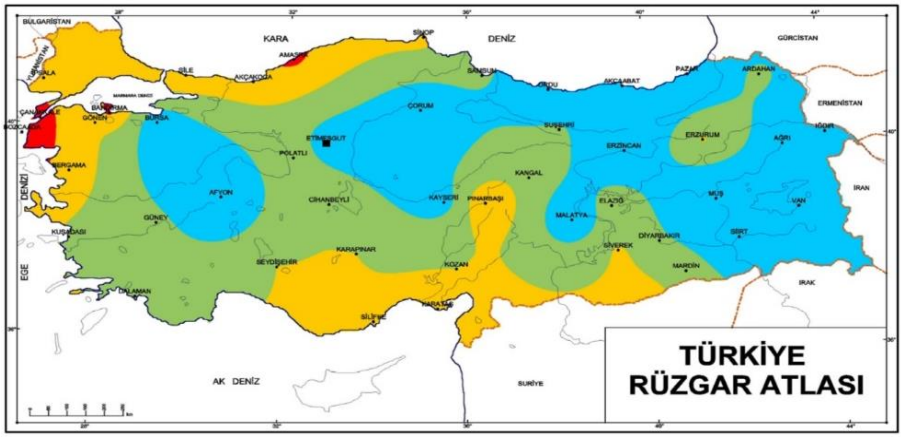
Multiple hydroelectric power plants are in the Göksu River basin, which includes the Ermenek Stream [15].

According to the Turkey Wind Atlas (Figure 4) [26] published by the General Directorate of Meteorology, the mountainous regions of Karaman, including Ermenek, Başyayla, and Sarıveliler districts, possess potential for wind energy generation due to their location on the slopes of the Taurus Mountains [19]. The potential capacity for establishing wind energy power plants for Karaman Province is estimated at approximately 1000 MW [19].

However, detailed wind resource assessments specific to Ermenek district are not prominently featured in the obtained information.

Biomass can be defined as the mass formed by living organisms consisting of one or more species. Biomass energy is renewable energy obtained from living and recently deceased organisms. Biomass resources consist of a structure intertwined with industrial applications [27].

According to Law No. 5346, biomass is defined as municipal waste (including landfill gas), vegetable oil waste, agricultural waste without food and feed value, forest products other than industrial wood, and by-products resulting from waste tire processing, industrial waste sludges, and treatment sludges, provided they are not imported [27,28].



Wind potentials at 50 m above ground level for five different cases[1]										
	Closed lands[2]		Open lands[3]		Coasts[4]		Offshore[5]		Hills[6]	
	ms^{-1}	Wm^{-2}	ms-1	Wm^{-2}	ms^{-1}	Wm^{-2}	ms^{-1}	Wm^{-2}	ms^{-1}	Wm^{-2}
	>6.0	>250	>7.5	>500	>8.5	>700	>9.0	>800	>11.5	>1800
	5.0-6.0	150-250	6.5-7.5	300-500	7.0-8.5	400-700	8.0-9.0	600-800	10.0-11.5	1200-1800
	4.5-5.0	100-150	5.5-6.5	200-300	6.0-7.0	250-400	7.0-8.0	400-600	8.5-10.0	700-1200
	3.5-4.5	50-100	4.5-5.5	100-200	5.0-6.0	150-250	5.5-7.0	200-400	7.0-8.5	400-700
	<3.5	<50	<4.5	<100	5.0	<150	<5.5	<200	<7.0	<400

Figure 4: Turkey wind energy potential atlas [26].

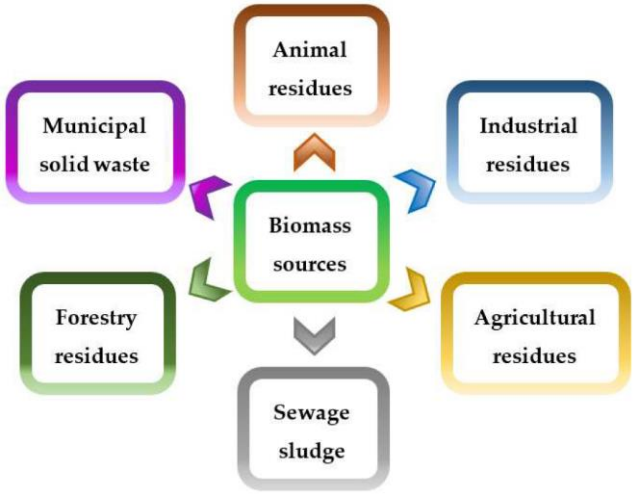


Figure 5: Biomass energy sources [28].

The Biomass Energy Potential Atlas (BEPA) [31] provides detailed information about biomass sources in Turkey [29]. Turkey possesses significant national bioenergy potential that can be obtained from

agricultural wastes, forestry, and other sources [29]. The region including Ermenek offers significant amounts of usable biomass potential per hectare [30,31].

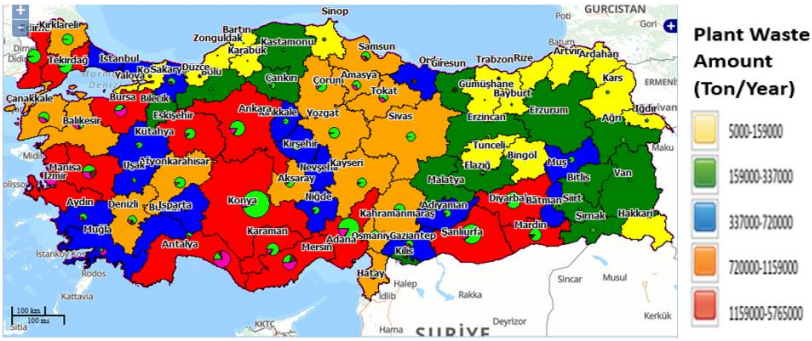


Figure 6: Turkey biomass energy potential atlas [31].

However, the specific bioenergy potential for Ermenek district is not detailed in the provided information.

EXISTING RENEWABLE ENERGY PROJECTS IN ERMENEK DISTRICT

The most important existing renewable energy project in Ermenek district is the Ermenek Dam and Hydroelectric Power Plant (HEPP) [25] built on the Ermenek Stream. This power plant, with its 302.4 MW installed capacity [32] and annual average electricity generation of 1048 GWh [19], makes a significant contribution to the region's energy needs. The dam utilizes the water resources of the Ermenek Stream [15]. Although the Gezende HEPP on the Ermenek Stream is mentioned, this power plant is in Mersin's Mut District and is outside Ermenek's boundaries [33].

Another important renewable energy project in Ermenek is the Ermenek Solar Power Plant (SPP) with 11.126 MW installed capacity, commissioned in June 2017. This ground-mounted solar energy project spans an area of 165,000 square meters [34]. The project's engineering, procurement, and construction services were carried out by a private company, while photovoltaic modules were supplied by another company [34].

The Ermenek Government Building SPP Facility in Ermenek district, which serves as an example of renewable energy applications in the public sector, was commissioned in September 2014. This small-scale, grid-

connected solar energy system is located on the terrace of the government building and has a 40 kW installed capacity [35]. The system's annual electricity generation has been observed to be approximately 75,000 kWh, which is higher than initially expected [36]. While the system's payback period was calculated as 8 years during the project phase, it is expected to be 5 years based on actual production performance [36].

No specific wind energy power plant operating within Ermenek district boundaries has been identified. Although wind energy potential exists in the region [19,26], it is understood that this potential is not currently being evaluated through large-scale projects.

Although the Karaman Biogas Plant [37] in Karaman is listed, it is not specified whether this facility is within Ermenek district boundaries. No information about other bioenergy facilities specific to Ermenek district could be accessed.

RENEWABLE ENERGY ACTIVITIES IN THE REGION

Investigations indicate that studies on solar energy systems are being conducted in the region. However, specific information about existing projects or activities was not found in the research materials provided. Additional research needs to be conducted to obtain more information on this subject.

The existence of rooftop and facade-type SPP applications carried out by the private sector in Ermenek is important in terms of demonstrating increasing interest and investment in solar energy in the region. However, detailed data about the number, capacity, or impacts of these applications is not available in the information provided.

POTENTIAL CONTRIBUTIONS OF RENEWABLE ENERGY SOURCES TO THE ERMENEK REGION

Greater utilization of renewable energy sources in the Ermenek region can provide various economic benefits. The development of projects in this field can create new job opportunities in production, installation, maintenance, and related sectors [5]. The renewable energy sector has significant job creation potential globally [8]. Wind and solar energy projects can offer new income sources to landowners through land leases [8]. The absence of fuel costs for renewable energy sources such as solar and wind

means stable and predictable energy prices in the long term for consumers and businesses in Ermenek [5]. Ermenek's high renewable energy potential can attract both local and international investors to the region [39]. The development of the renewable energy sector can contribute to economic diversification in addition to Ermenek's existing economic activities such as livestock, fruit growing, and mining [40]. Renewable energy projects can revitalize local economies through increased tax bases and expenditures [41].

The utilization of renewable energy sources can also provide significant social contributions to the Ermenek region. Distributed renewable energy solutions can improve energy access in rural and remote areas of Ermenek district [7]. Reduced dependence on fossil fuels can lead to cleaner air and water, improving public health outcomes by reducing respiratory diseases and other pollution-related health problems [5]. Renewable energy projects can contribute to rural development by creating job opportunities, improving infrastructure, and offering opportunities for younger generations to remain in their communities [41]. Additionally, renewable energy projects can offer opportunities for partnerships and participation in energy production [7].

From an environmental perspective, renewable energy sources produce little or no greenhouse gas emissions during operation. This helps combat climate change [7]. Renewable energy production reduces air and water pollution associated with burning fossil fuels [6]. Renewable energy sources are inexhaustible and reduce the need for extraction of limited fossil fuel resources [6]. The transition to renewable energy reduces the environmental impact associated with extracting, transporting, and burning fossil fuels [5].

CONCLUSION AND RECOMMENDATIONS

Ermenek district possesses significant potential in terms of renewable energy sources, particularly solar and hydroelectric energy. Currently, the Ermenek Dam and HEPP along with the Ermenek Solar Power Plant are the main renewable energy projects in the district. Additionally, the small-scale solar energy facility at the Ermenek Government Building and rooftop-type solar energy applications in the private sector demonstrate increasing interest and investment in renewable energy in the region. Greater utilization of renewable energy sources can provide significant benefits to the Ermenek

region, including economic growth, job opportunities, reduced energy costs, improved energy access, public health protection, and environmental sustainability.

The following recommendations should be considered for Ermenek to fully evaluate its renewable energy potential:

1. Comprehensive resource assessment studies should be conducted to determine wind and bioenergy potential specific to Ermenek district in greater detail.
2. A strategic plan should be developed for the development of renewable energy sources for Ermenek, with specific targets and timelines determined for the utilization of different renewable energy technologies.
3. An active investment incentive program targeting both local and international investors should be implemented, emphasizing Ermenek's high renewable energy potential.
4. Local community participation should be ensured in the planning and development process of renewable energy projects, and community ownership models should be encouraged to increase social acceptance and maximize local benefits.
5. Local policies and incentives should be implemented to encourage renewable energy use by households, businesses, and public institutions, as exemplified by the Ermenek Government Building.
6. Necessary infrastructure improvements such as grid connections and energy storage solutions should be planned to support the integration of increased renewable energy capacity.
7. Investment should be made in education and capacity development programs to ensure local people are trained as qualified workforce to work in the renewable energy sector.
8. Collaboration should be encouraged among local government, businesses, research institutions, and national energy agencies to accelerate renewable energy development in Ermenek.

In the long term, Ermenek can make significant contributions to Turkey's renewable energy goals as a sustainable energy hub and enhance regional prosperity. Investment in renewable energy sources is the key to a clean, secure, and sustainable future for Ermenek.

Acknowledgment

This study was presented at the 1st International Interdisciplinary Ermenek Congress held in Karaman Province, Ermenek District, between May 13-15, 2025, and its abstract was published in the proceedings book.

REFERENCE

[1] Turkiye's International Energy Strategy / Republic of Türkiye Ministry of Foreign Affairs, accessed April 28, 2025, https://www.mfa.gov.tr/turkeys-energy-strategy.en.mfa

[2] Turkey Solar Energy Potential Atlas (MENR 2021b) – Research Gate, accessed April 28, 2025, https://www.researchgate.net/figure/Turkey-Solar-Energy-Potential-Atlas-MENR-2021b_fig3_357648714

[3] Renewables in 2024: 5 Key Facts Behind a Record-Breaking Year - IRENA, accessed April 28, 2025, https://www.irena.org/News/articles/2025/Apr/Renewables-in-2024-5-Key-Facts-Behind-a-Record-Breaking-Year

[4] Detail- IRENA, accessed April 28, 2025, https://www.irena.org/Digital-content/Digital-Story/2024/Jan/An-Unprecedented-Momentum-for-Renewables/detail

[5] Renewable energy – powering a safer future United Nations, accessed April 28, 2025, https://www.un.org/en/climatechange/raising-ambition/renewable-energy

[6] Benefits of Renewable Energy Use Union of Concerned Scientists, accessed April 28, 2025, https://www.ucsusa.org/resources/benefits-renewable-energy-use

[7] Why is renewable energy important? - REN21, accessed April 28, 2025, https://www.ren21.net/why-is-renewable-energy-important/

[8] The Economic Benefits of Renewable Energy | Get to Learning' with Our Knowledge Center, accessed April 28, 2025, https://www.energytexas.com/get-to-learnin/the-economic-benefits-of-renewable-energy

[9] Türkiye surpasses 2025 solar target as capacity doubles in 2.5 years - Ember, accessed April 28, 2025, https://ember-energy.org/latest-insights/turkiye-surpasses-2025-solar-target-as-capacity-doubles-in-2-5-years/

[10] Turkey- Electric Power – Renewables, Smart Grid, Energy Storage, Civil Nuclear - International Trade Administration, accessed April 28, 2025, https://www.trade.gov/country-commercial-guides/turkey-electric-power-renewables-smart-grid-energy-storage-civil-nuclear

[11] Ermenek- Wikipedia, accessed April 28, 2025, https://en.wikipedia.org/wiki/Ermenek

[12] Ermenek District- Wikipedia, accessed April 30, 2025, https://en.wikipedia.org/wiki/Ermenek_District

[13] Assessment of rockfall hazard on the steep-high slopes: Ermenek (Karaman, Turkey), accessed April 30, 2025, https://nhess.copernicus.org/preprints/nhess-2015-337/

[14] Anonymous, (2025) Karaman Province Map accessed April 30, 2025, https://www.turkey-visit.com/map/turkey/karaman-map.asp#google_vignette

[15] Ermenek Dam & Hydropower Project - ECA Watch, accessed April 30, 2025, http://www.eca-watch.at/downloads/Ermenek%20final%20version.pdf
[16] Geographical location, accessed April 30, 2025, https://kmu.edu.tr/yos/sayfa/3530/geographical-location/en
[17] Monthly climate in Ermenek, Karaman, Turkey - nomad season, accessed April 30, 2025, https://nomadseason.com/climate/turkey/karaman/ermenek.html
[18] Ermenek, Turkiye weather in January - Wanderlog: travel itinerary, vacation & road trip planner, accessed April 30, 2025, https://wanderlog.com/weather/25135/1/ermenek-weather-in-january
[19] Energy- Investment in Karaman, accessed April 30, 2025, https://www.investinkaraman.gov.tr/sectors/energy
[20] Alternatives to Solar Power Plant Location Through GIS and AHP: case of Karaman, Turkey- DergiPark, accessed April 30, 2025, https://dergipark.org.tr/tr/download/article-file/1131614
[21] Health, Safety, and Environmental Impacts - Department of Energy, accessed April 28, 2025, https://www.energy.gov/eere/health-safety-and-environmental-impacts
[22] Karaman solar energy potential map - energy atlas, https://www.enerjiatlasi.com/gunes-enerjisi-haritasi/karaman
[23] Turkey's Solar Energy Potential, accessed May 2, 2025, https://www.assolarenerji.com/en/?/turkey-s-solar-energy-potential
[24] Turkey- Global Solar Atlas, accessed May 2, 2025, https://globalsolaratlas.info/download/turkey
[25] Anonymous, (2025) Ermenek Dam and Hydroelectric Power Plant (HEPP) accessed May 2, 2025, https://www.enerjiatlasi.com/hidroelektrik/ermenek-baraji.html
[26] Republic of Turkey Ministry of Environment, Urbanization and Climate Change General Directorate of Meteorology accessed May 2, 2025 https://www.mgm.gov.tr/genel/ruzgar-atlasi.Aspx
[27] Republic of Turkey Ministry of Environment, Urbanization and Climate Change General Directorate [https://enerji.gov.tr/tabi-kaynaklar]
[28] Kaltschmitt, M. Biyokütleden yenilenebilir enerji, Giriş. *Yenilenebilir Enerji* Sistemlerinde ; Kaltschmitt, M., Themelis, NJ, Bronicki, LY, Söder, L., Vega, LA, Eds.; Springer: New York, NY, ABD, 2013; s. 45–71.
[29] Republic of Turkey Ministry of Energy and Natural Resources - Biomass, accessed May 2, 2025, https://enerji.gov.tr/eigm-renewable-energy-resources-biomass
[30] The Use of Renewable Energy Sources for Energy in Turkey and Potential Trends, accessed May 2, 2025, https://www.researchgate.net/publication/245395945_The_Use_of_Renewable_Energy_Sources_for_Energy_in_Turkey_and_Potential_Trends
[31] Republic of Turkey Ministry of Energy and Natural Resources General Directorate of Energy Affairs accessed April 12, 2025 https://bepa.enerji.gov.tr/
[32] Ermenek, Turkey - Power Technology, accessed April 30, 2025, https://www.power-technology.com/marketdata/ermenek-turkey/

[33] Gezende HEPP - EÜAŞ, accessed May 10, 2025, https://www.euas.gov.tr/en-US/power-plants/gezende-hepp
[34] Power plant profile: Ermenek Solar PV Park, Turkey, accessed May 10, 2025, https://www.power-technology.com/data-insights/power-plant-profile-ermenek-solar-pv-park-turkey/
[35] Karaman Solar Energy Potential Map- Energy Atlas, accessed May 10, 2025, https://www.enerjiatlasi.com/gunes-enerjisi-haritasi/karaman
[36] Provincial Special Administration Solar Power Plant Becomes Operational-Republic of Turkey Karaman Governorship, accessed May 10, 2025, http://www.karaman.gov.tr/il-ozel-idaresi-gunes-enerji-santrali-faaliyete-gecti
[37] Ermenek Government Building Produces Its Own Electricity - Karaman Governorship, accessed May 10, 2025, http://www.karaman.gov.tr/ermenek-hukumet-konagi-kendi-elektrigini-retiyor
[38] Karaman Power Plants- Energy Atlas, accessed May 2, 2025, https://www.enerjiatlasi.com/sehir/karaman/
[39] These developing countries are leading the way on renewable energy, accessed May 10, 2025, https://www.weforum.org/stories/2022/07/renewables-are-the-key-to-green-secure-affordable-energy/
[40] Solar power plant to Ermenek- Democratic News, accessed May 10, 2025, https://www.demokrathaber.org/ermeneke-gunes-enerji-santrali
[41] 3 Ways Rural Communities Benefit from Renewable Energy Development, accessed May 10, 2025, https://cleangridalliance.org/blog/229/3-ways-rural-communities-benefit-from-renewable-energy-development