Numerical Methods

Fall 24-25

Project

Student Name : Laith Ismail Ahmed Hijazi
Student ID : 64220039
Email : laith.hijazi@std.medipol.edu.tr

1. Abstract

Weather forecasts are essential for decision-making and planning. This project develops a user-friendly graphical user interface (GUI) to provide accurate weather predictions tailored for Istanbul Medipol University and its surroundings. Using historical weather data, a Long Short-Term Memory (LSTM) neural network is trained to predict patterns over time. Missing data is filled with Lagrange interpolation, and cubic spline interpolation is applied for smooth trend visualization. An API integrates real-time data, allowing users to view current conditions for any city, alongside localized predictions for the university area. This system aims to deliver precise and actionable insights to support campus planning and activities.

2. Introduction

Weather forecasting plays a crucial role in modern life, influencing decisions in sectors such as agriculture, transportation, and urban planning. With advancements in technology, the demand for local weather predictions has grown because they provide more precise insights tailored to specific areas. While large-scale forecasting systems offer general trends, creating accurate and localized solutions remains challenging, especially in regions with dynamic and diverse climates like Istanbul.

To address these challenges, the project develops a weather prediction system specifically tailored for Istanbul Medipol University and its surroundings. The focus is on leveraging historical weather data to train a Long Short-Term Memory (LSTM) neural network. This approach is chosen due to LSTM's ability to learn complex temporal patterns and dependencies, which are essential for accurate weather forecasting. Complementary data preprocessing techniques, including Lagrange interpolation for handling missing data and cubic spline interpolation for smooth trend visualization, ensure the reliability and clarity of predictions. By integrating real-time weather data via an API, the system combines up-to-date conditions with localized forecasting, all presented through an intuitive graphical user interface (GUI).

3. *Problem*

Most weather forecasting systems are designed to provide general predictions for large regions, which often fail to capture the specific conditions of smaller local areas. In a city like Istanbul, where weather can change dramatically within short distances, these broad forecasts may not accurately represent the conditions in specific locations, such as Istanbul Medipol University and its surroundings.

For institutions like the university, precise and localized weather predictions are crucial for effective campus management, event planning, and daily operations. However, the lack of continuous and accurate historical weather data poses a significant challenge. Gaps in data or incomplete records make it difficult to build reliable forecasting models, leading to less accuracy.

4. Proposed Solution

To address this challenge, this project proposes a system designed to predict weather conditions specifically for the area surrounding Istanbul Medipol University. The system integrates historical weather data, advanced interpolation techniques, and machine learning models to generate reliable predictions.

However, the core of the solution involves training a Long Short-Term Memory (LSTM) neural network using historical data to forecast temperature, humidity, and pressure. To overcome the challenge of incomplete datasets, missing values are addressed using Lagrange interpolation, ensuring the continuity of the training data. Additionally, cubic spline interpolation is used to create smooth trend curves, enhancing the presentation of weather data.

5. Used Methodology

This project followed a systematic approach to develop a localized weather forecasting system tailored to Istanbul Medipol University. The methodology consisted of the following phases:

5.1 Data Collection

The weather data for this project, covering the period from 1/1/2019 to 30/12/2024, was collected from trusted online sources. The dataset contained various meteorological variables, but only temperature, humidity, and pressure were selected for analysis, as they were key to the project's objectives. Once the data was acquired, it was cleaned by removing rows with all missing values and it was re-indexed to ensure consistency and prepare it for further analysis.

5.2 Data Preprocessing

The collected weather data was preprocessed to handle missing values in the temperature, pressure, and humidity columns. First, the number and indices of missing values were identified. To fill these gaps, backward filling was used as an initial step, providing temporary estimates based on subsequent values. A more precise approach, Lagrange interpolation, was then applied nearby data points to fill missing values in a way that preserved continuity and reflected realistic trends. Special handling was implemented for edge cases, such as missing values at the start or end of the dataset. These steps ensured data continuity and prepared the dataset for further analysis.

5.3 Trend Visualization

The trend visualization process used the interpolated data for temperature, pressure, and humidity to create smooth and continuous curves, providing a clear representation of historical trends from 1/1/2019 to 30/12/2024. A customized solution for natural cubic spline interpolation was implemented to fit these curves accurately. This process relied on the get_natural_cubic_spline_model function, which combines natural cubic spline transformations with linear regression in a pipeline. This combination allowed for the prediction of new values based on input (x) and output (y) data ensuring the fitted curves were linear outside the range of specified knots (control points) while remaining smooth and differentiable within the range.

The implementation defined two classes: AbstractSpline, which served as the base class for spline transformations, and NaturalCubicSpline, which applied the natural cubic spline transformation to the input data. Knots, essential for controlling the flexibility of the spline, were either specified manually or generated automatically. This approach was adapted from a solution shared on Stack Overflow, where the original cubic spline function was modified and extended to support natural cubic splines with added functionality for flexibility and usability.

The trends were visualized through plots, segmented into early years (2019–2021) and later years (2022–2024), allowing for a detailed examination of weather patterns. These visualizations provided meaningful insights into past weather trends, and enhanced the interpretability of the data for subsequent analysis.

5.4 Model Development

A Long Short-Term Memory (LSTM) neural network was developed to predict weather conditions using the interpolated data. The data was split into two parts: 70% for training, while the remaining 30% was used for testing to evaluate the model's performance.

To prepare the data for the model, all features (temperature, pressure, and humidity) were scaled to values between 0 and 1 using a MinMaxScaler ensuring consistent and efficient model training. A time-series generator was created using the training data to provide input-output pairs for the model, with one previous time step used for prediction.

The LSTM architecture consisted of a single layer with 50 units and ReLU activation, followed by a dense layer that produce three outputs, one for each weather feature. The model was compiled using the Adam optimizer, with mean squared error as the loss function to ensure smooth accurate predictions.

The model was trained for 50 epochs using the time-series generator enabling it to learn sequential patterns and relationships in the data. After training, the model was saved to a file for future use. This LSTM design was chosen for its effectiveness in analyzing sequential data, making it an ideal approach for predicting localized weather trends.

5.5 Real-Time Data Integration

Real-time weather data was integrated into the system using the OpenWeatherMap API, enabling the application to fetch current weather conditions for any specified city such as temperature, humidity, pressure, and weather icons, which are processed and displayed in the application. The fetched temperature data, originally in Kelvin, is converted to Celsius. The integration ensures users can access up-to-date weather information alongside localized predictions.

To predict weather conditions near Istanbul Medipol University, the most recent data from the processed data is used as the starting point. The trained LSTM model takes this input and generates predictions for the current day and subsequent days. These predictions, including average temperature, pressure, and humidity, are inversely transformed from their normalized scale back to their original values using the MinMaxScaler.

5.6 Development of the Graphical User Interface (GUI)

The weather application's GUI was designed to be simple and user-friendly, making it accessible to everyone. Built with Tkinter and enhanced with ttkbootstrap, it lets users enter a city name to retrieve real-time weather data, including temperature, pressure, humidity, and a weather icon from the OpenWeatherMap API.

The interface has sections to show current weather and buttons to predict weather near Istanbul Medipol University or for the next three days.

The GUI includes features to handle errors, such as showing messages if a city is not found or if there is an issue loading the weather icon. This ensures the system works smoothly and provides a better user experience.

5.7 Validation and Testing

The system was tested to ensure accuracy and usability. The LSTM model was evaluated on the testing dataset, using metrics like Mean Squared Error (MSE) to check the accuracy of its predictions.

Additionally, the functionality of the GUI was tested for usability to ensure a seamless user experience. Various scenarios such as entering incorrect city and API errors were simulated to verify that the system handled them correctly by displaying appropriate error messages.

Localized weather predictions near Istanbul Medipol University and future forecasts were validated to ensure they matched real-world trends. This testing process confirmed the system’s reliability, accuracy, and ease of use.

6. Background and Theory

6.1 Lagrange Interpolation

The Lagrange Interpolation function explicitly follows the mathematical formula:

$$\boldsymbol{P(x) = \sum_{i=0}^{n} y_i \prod_{\substack{j=0 \\ j\neq i}}^{n} (x - x_j)/x_i - x_j}$$

This method estimates missing values P(x) using known data points (x_i, y_i). It interpolates missing weather data in temperature, pressure, and humidity. Special cases (e.g., at the edges of the dataset) are handled avoid accessing out-of-range indices. This ensures continuity and prepares the dataset for LSTM model training.

6.2 Cubic Spline Interpolation

It is used in the code to create smooth and continuous curves for weather data trends. It follows the formula:

$$\boldsymbol{S_i(x) = a_i + b_i(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^3}$$

Knots, which control the curve's flexibility, are either defined manually or generated automatically. Smoothness is maintained across all intervals, and natural boundary conditions ensure the second derivatives at the edges are zero

$$S_1''(x_0) = 0 \ and \ S_n''(x_0) = 0$$

The implementation combines spline transformation with linear regression to simplify training and prediction. This method effectively smooths noisy data, providing clear and continuous visualizations of temperature, pressure, and humidity trends over time. It highlights long-term weather patterns while minimizing abrupt transitions, making it ideal for trend analysis.

7. Difficulties

The development of this weather forecasting system involved addressing several significant challenges, which shaped its design and implementation. These challenges are summarized as follows:

7.1 Data Availability and Completeness

Obtaining high-quality historical weather data specific to Istanbul Medipol University was difficult due to gaps in key parameters such as temperature, humidity, and pressure. To fill these gaps, Lagrange interpolation was applied, ensuring data continuity but adding complexity to the preprocessing process.

7.2 Model Training and Optimization

Developing and optimizing the LSTM neural network required careful experimentation with the network structure and hyperparameters. Striking a balance between accuracy and efficiency involved iterative adjustments to factors like batch size, and number of LSTM units, which was time-intensive and computationally demanding.

7.3 User Interface Design

Creating a GUI that was functional and easy to use presented its own challenges. The interface needed to clearly display real-time updates, daily predictions, and future forecasts without overwhelming users. Achieving a visually appealing, organized layout with smooth integration of all features required multiple design iterations and attention to detail.

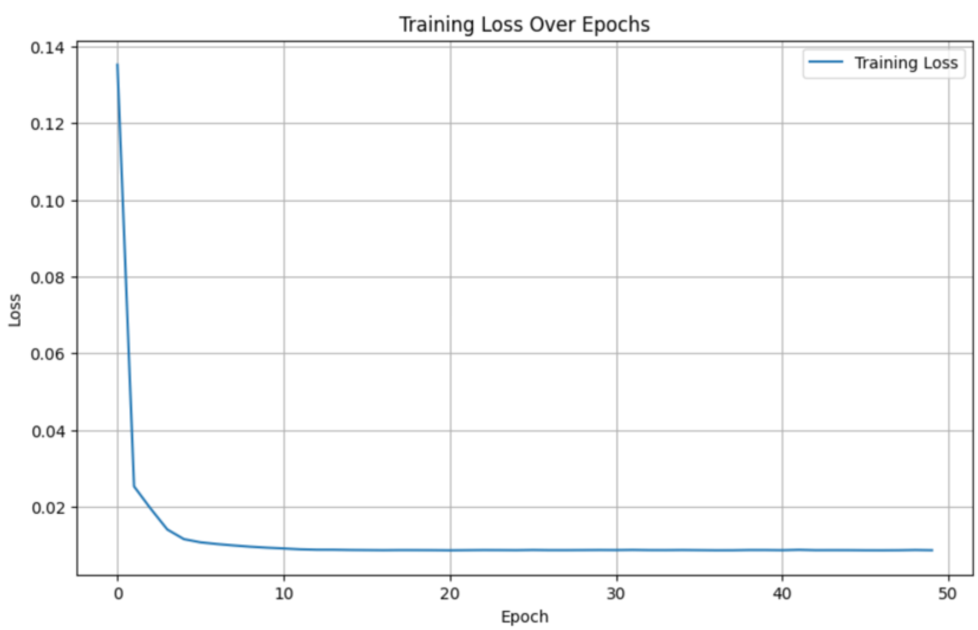
8. Results and Discussions

The developed weather forecasting system successfully delivered localized predictions for Istanbul Medipol University while providing real-time weather updates for user-specified cities. This section outlines the key results, insights, and comparisons to initial expectations.

8.1 Model Performance

The LSTM model was trained for 50 epochs, and its performance was evaluated using the Mean Squared Error (MSE) metric. The model achieved a low MSE of 0.0085 on the testing dataset, showing its ability to accurately predict weather patterns.

The training loss curve shows a rapid decrease during the initial epochs, indicating that the model was learning well and adjusting its weights appropriately. After around 10 epochs, the loss stabilized, showing that the model had successfully captured the patterns in the data without overfitting. This outcome highlights the stability of the training process and the model's suitability for the task.



8.2 Real-Time Data Integration

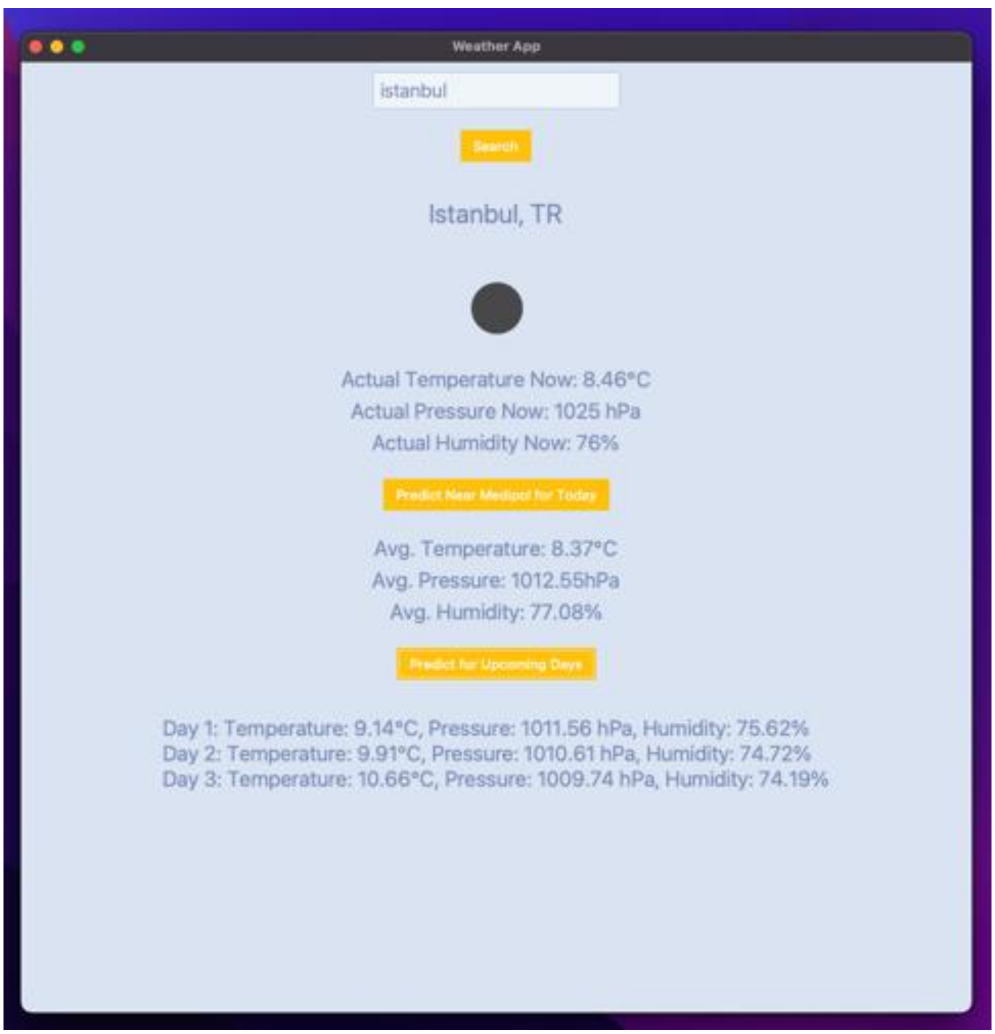
The system successfully used the OpenWeatherMap API to fetch real-time weather data like temperature, humidity, and pressure, which were processed and shown in the application for any city. While the real-time data provided instant updates for the entered city, the LSTM model predicted the average weather conditions (temperature, pressure, and humidity) for the area near Istanbul Medipol University of the current day and forecasting the next three days. This setup ensured users received both up-to-date weather conditions and accurate daily predictions in a simple and reliable way.

8.3 Graphical User Interface (GUI)

The GUI functioned as expected, providing users with a simple and accessible way to view weather information. It enabled users to:

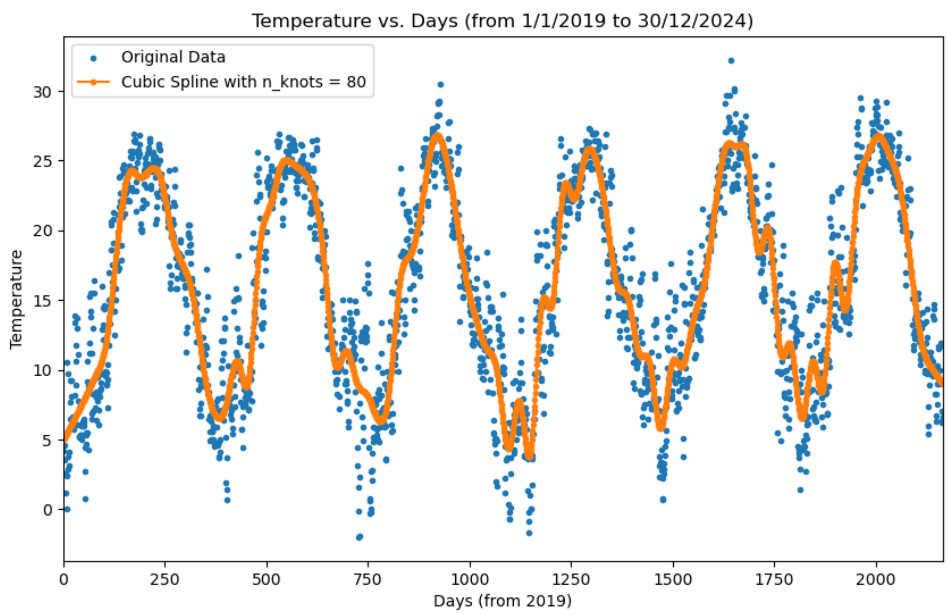
- Retrieve real-time weather data for any selected city.
- Access the model's predictions for average weather conditions near Istanbul Medipol University for the current day.
- View forecasts for the next three days specific to the university area.

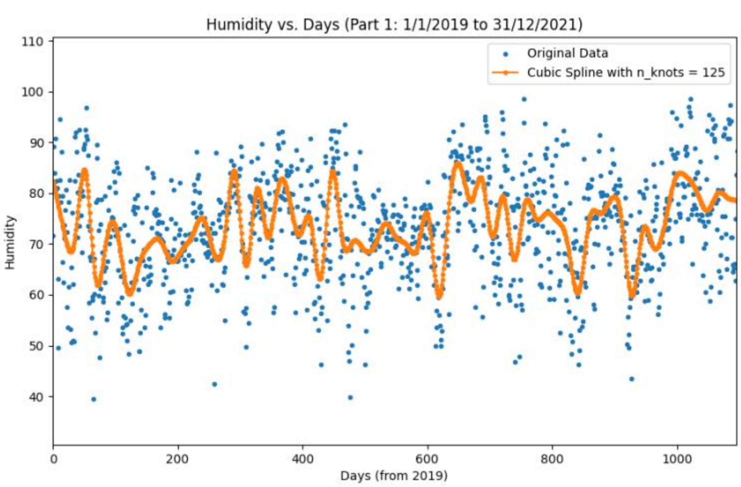
The interface operated as intended, successfully integrating live data and model forecasts into a user-friendly platform.

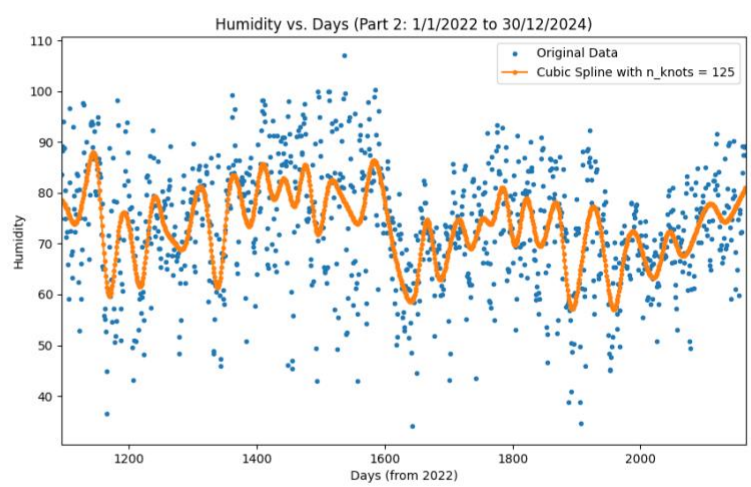


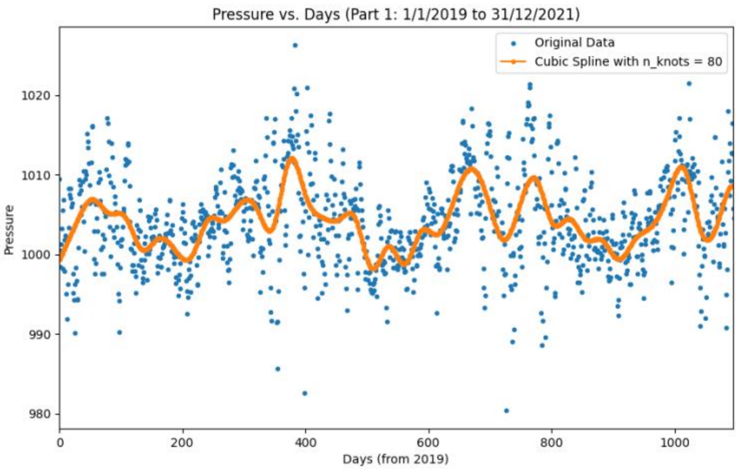
8.4 Visual Output

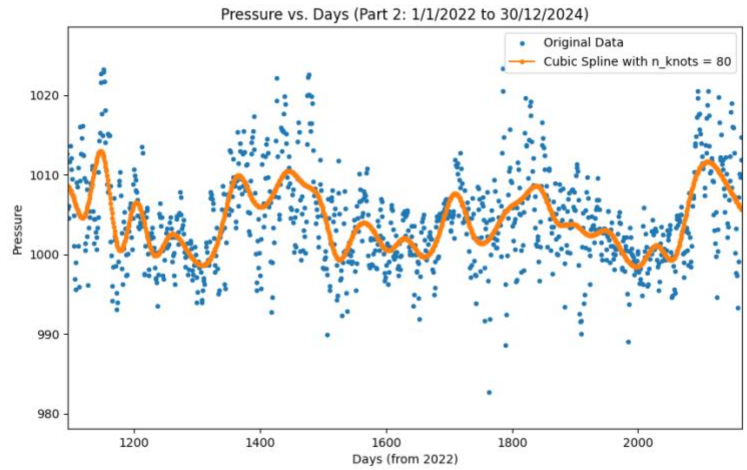
The system produced clear and smooth trend visualizations of historical weather data using cubic spline interpolation. These visual outputs effectively highlighted changes in temperature, pressure, and humidity.











8.5 Achieved and Expected Outcomes

Aspect	***Expected***	***Achieved***
Prediction Accuracy	Capture trends and provide accurate forecasts.	Successfully forecast trends with minor discrepancies.
Real-Time Data Integration	Seamless API integration for current conditions.	Fully integrated and operational with real-time updates.
GUI Functionality	User-friendly and intuitive design.	Delivered a seamless user experience.
Weather Trends	Smooth and visually interpretable curves	Achieved with cubic spline interpolation.

Discussion

The system shows a good balance between accuracy and ease of use, providing reliable forecasts and real-time updates through a simple interface. While there were minor errors in some predictions, they do not greatly affect the system's overall performance. These small issues highlight the need for further improvements, especially in handling complex weather patterns. The successful integration of real-time data and forecasts proves the system's practical value, making it a helpful tool for everyday use. This provides a strong base for future enhancements and developments.

9. Possible Improvements

9.1 Expanded Dataset

Including additional variables, such as wind speed, precipitation, and visibility, could enhance the model's accuracy in generating detailed forecasts. Expanding the historical weather dataset to cover a longer time period (more years) would also improve the precision of predictions.

9.2 Refined Data Handling Techniques

While Lagrange interpolation worked well for filling missing values, exploring advanced methods like multivariate interpolation could provide more accurate estimates, ensuring greater data consistency and further optimize the model’s performance.

9.3 Dynamic Model Updates

Integrating live data streams for continuous model retraining would allow the system to adapt to real-time weather changes. This dynamic approach would make forecasts more responsive and accurate in varying conditions.