Machine Learning-Based Wet Refractivity Prediction Through GNSS Troposphere Tomography for Ensemble Troposphere Conditions Forecasting

Saeid Haji-Aghajany, Witold Rohm, Maciej Kryza, and Kamil Smolak

***Abstract*— This article introduces an innovative ensemble troposphere conditions forecasting method using wet refractivity within the context of Global Navigation Satellite System (GNSS) troposphere tomography. The current models lack coverage of diverse geographical locations and weather conditions, and they do not utilize high-spatial resolution tropospheric data to cover a large area. Moreover, their deterministic prediction mode may introduce high uncertainty into the results. This article leverages long short-term memory (LSTM) networks and genetic algorithms (GAs) to optimize hyperparameters, enabling the prediction of 3-D wet refractivity fields for ensemble forecasting under various weather conditions including rain bands sweeping in Poland and a storm in California, USA. A comparison of the 3-h predictions with Weather Research and Forecasting (WRF) model outputs at levels with a height lower than 3000 m shows root-mean-squared error (RMSE) values of 4.15 and 3.18 ppm for Poland and California, respectively. After utilizing the generative adversarial network (GAN) to produce realistic time series, ensemble forecasting is conducted. The model demonstrates exceptional accuracy in both regions, yielding an optimal threshold of 0.41, which shows a point at which the balance between true positive (TP) and true negative (TN) instances is optimized, achieving a sensitivity of 0.967 and a precision of 0.973 in Poland. Additionally, it achieves an optimal threshold of 0.52, yielding a sensitivity of 0.982 and a precision of 0.993 in California. The low false positive rate (FPR) of 0.027 in Poland and 0.011 in California underscore the adaptability and reliability of the model across diverse datasets.**

***Index Terms*— Ensemble forecasting, generative adversarial network (GAN), long short-term memory (LSTM), machine learning (ML), troposphere tomography.**

I. Introduction

THE presence of water vapor, which plays a crucial role in atmospheric phenomena, introduces significant measurement errors that profoundly affect Earth observation systems, particularly the Global Navigation Satellite System (GNSS). GNSS signals experience delays as they traverse Earth's troposphere, a phenomenon of significant importance in GNSS meteorology [1]. Variations in zenith tropospheric delay (ZTD) closely correlate with changes in pressure, temperature, and water vapor around GNSS stations [2]. ZTD comprises two main components: zenith hydrostatic delay (ZHD) and zenith wet delay (ZWD) [3]. ZWD relies on wet refractivity and presents challenges due to the uneven distribution of water vapor in both space and time. In addition to increasing the accuracy of real-time and near-real-time positioning modes [4], modeling of ZWD or other GNSS tropospheric products, like precipitable water vapor, is beneficial for forecasting of natural phenomena, including the subsidence [5], groundwater level reduction [6], flash floods [7], wind speed [8], cyclonic storm [9], rainfall [8], [10], [11], [12], [13], [14], and foehn winds [15].

The initial attempts to predict GNSS tropospheric products relied on mathematical methods, including the linear trend, decomposition, moving average, and exponential smoothing models [16], [17], [18]. However, as technology advanced, researchers shifted their focus to harnessing the potential of machine learning (ML) due to its remarkable capabilities. One prominent research avenue entails the creation of regional tropospheric models utilizing artificial neural networks (ANNs). Several other studies have employed a backpropagation NN (BPNN) explicitly designed for modeling regional tropospheric delay [19], [20]. A comprehensive analysis revealed that the radial basis function (RBF) algorithm outperformed others in modeling tropospheric delay with local-scale samples [21]. Some researchers employed the multilayer perceptron (MLP) method to capture variations in tropospheric delay time series [22]. In 2017, a ZTD prediction model was crafted utilizing GNSS data from Hong Kong. This model was created through a hybrid method, incorporating both BPNN and genetic algorithm (GA) [23]. After exploring recurrent NN (RNN) methods based on their ability for time-series prediction, some researchers focused on using different RNN methods like long short-term memory (LSTM) for ZTD prediction. A groundbreaking predictive model for ZTD was introduced using the LSTM network architecture [24], [25]. A study addressed the intricate spatiotemporal dynamics of Global Positioning System (GPS) stations in West Antarctica.

Manuscript received 21 January 2024; revised 19 April 2024 and 17 May 2024; accepted 18 June 2024. Date of publication 21 June 2024; date of current version 5 July 2024. The project is co-financed by the Polish National Agency for Academic Exchange under application no. BPN/ULM/2021/1/00263. The article processing charge (APC) is financed by Wrocław University of Environmental and Life Sciences. *(Corresponding author: Saeid Haji-Aghajany.)*

Saeid Haji-Aghajany, Witold Rohm, and Kamil Smolak are with the Institute of Geodesy and Geoinformatics, Wrocław University of Environmental and Life Sciences, 50-375 Wrocław, Poland (e-mail: saeid.haji-aghajany@upwr.edu.pl; Witold.rohm@upwr.edu.pl; kamil.smolak@upwr.edu.pl).

Maciej Kryza is with the Faculty of Earth Sciences and Environmental Management, University of Wrocław, 51-621 Wrocław, Poland (e-mail: maciej.kryza@uwr.edu.pl).

Digital Object Identifier 10.1109/TGRS.2024.3417487

This article approach involved BP and LSTM networks [26]. Enhancements were made in correcting ZTD values in the Antarctic region by incorporating RBF and LSTM models, effectively addressing both spatial and temporal factors [27]. In another research, ZTD was predicted at seven GNSS monitoring stations across China, leveraging static precise point positioning technology using an LSTM network [28]. GNSS tropospheric products have been employed in previous research efforts. However, one of the challenges which has not been solved before is the prediction of a high spatiotemporal local tropospheric map, along with the handling of challenging parameters such as wet refractivity. GNSS troposphere tomography is a potent technique for the reconstruction of the 3-D wet refractivity field, and therefore offers a promising solution for this challenge. It has shown its ability in different fields of study including increasing the accuracy of positioning [29], InSAR tropospheric correction [30], and downscaling of precipitation [31]. Various troposphere tomography methods have been developed in recent years, including voxel-based [32], [33], [34], [35], [36], function-based [37], [38], and node-based [39], [40] approaches, with efforts made by researchers to optimize these methods. Another challenge is that there has been a lack of utilization of more robust forecasting methods like ensemble forecasting to enhance the accuracy of weather predictions.

In this article, the primary focus is on predicting the 3-D local wet refractivity field using troposphere tomography in combination with the LSTM network and GA. This predictive model is employed to prepare input data for ensemble forecasting of both synoptic-scale and local-scale weather phenomena in Poland and California, USA, encompassing varying climate conditions and utilizing the wet refractivity field from the tomography technique. The process begins with the generation of extensive time-series data from troposphere tomography, which serves as the training data for LSTM-GA. The generative adversarial network (GAN) method is then employed to generate ensemble members for forecasting, ensuring the production of realistic time series. In this context, "realistic" implies that the generated time series closely mimics the specifications and behavior of actual time-series data. Subsequently, predictions are made using both types of time series—real data and the realistically generated synthetic data. The predicted wet refractivity values are then utilized as input data for ensemble forecasting of rain bands sweeping in Poland and a storm in California. Sections II–IV of this article delve into the theoretical foundations of the problem, introduce the study area and dataset, and thoroughly discuss the processing results.

II. Methodology

A. GNSS Troposphere Tomography

GNSS troposphere tomography furnishes a comprehensive 3-D depiction of wet refractivity, capitalizing on tropospheric delays derived from GNSS data in the line-of-sight direction [32].

1) Tropospheric Wet Delay: As GNSS signals traverse the troposphere, they undergo bending and delays influenced by a multitude of tropospheric factors. The ZTD is intimately intertwined with tropospheric parameters like pressure, temperature, and water vapor along the signal's path, and it can be extracted through the adept application of GNSS data processing techniques [41]. This division of ZTD into its pivotal constituents entails the ZHD and ZWD [42]. The precise determination of ZHD is achieved through the judicious utilization of the Saastamoinen model [41]

$$\mathrm{ZHD} = \frac{0.002277\, P_s}{(1 - 0.00266\cos(2\varphi) - 0.00000028\, H_s)} \tag{1}$$

where P_s represents the surface pressure in hPa, φ is the latitude, and H_s denotes the height of the station in meters. Subsequently, ZWD can be computed as the difference between ZTD and ZHD. In conjunction with ZTD, the results of GNSS processing encompass tropospheric horizontal gradients, offering insights into the nonuniform distribution of tropospheric gases. The ZWD values can be geometrically transformed into the respective directions of GNSS satellites to determine the slant wet delay (SWD). Subsequently, the estimation of SWD can be accomplished by employing the subsequent equation [32]

$$\mathrm{SWD} = \mathrm{mf}_{\mathrm{wet}}(\alpha) \times \mathrm{ZWD} + \cot(\alpha) \times \left(\left(G^W_{\mathrm{EW}} \times \sin \mathrm{Az}\right) + \left(G^W_{\mathrm{NS}} \times \cos \mathrm{Az}\right)\right) \tag{2}$$

where G^W_{EW} and G^W_{NS} represent the wet tropospheric horizontal gradients in the East–West and North–South directions, respectively. Moreover, α and Az correspond to the elevation angle and azimuth of the satellite, respectively, and $\mathrm{mf}_{\mathrm{wet}}$ characterizes the wet mapping function. In this study, the Vienna mapping function is used for $\mathrm{mf}_{\mathrm{wet}}$ [43]. The determination of hydrostatic gradients relied on a methodology described by Shoji [44], which leveraged the horizontal distribution of surface pressure. These gradients were then subtracted from the GNSS-derived gradient values, allowing us to focus exclusively on the wet components in (2). In prior studies, an established method for improving SWD values includes integrating postfit residuals derived from GNSS analysis. These residuals are used to account for signal delay factors that are not explicitly modeled [45], [46]. Nevertheless, alternative research suggests that the contribution of postfit residuals to the estimation process is relatively minor [47] and may potentially introduce inaccuracies associated with GNSS measurements, such as multipath interference [48]. In our investigation, we have made the deliberate choice to refrain from incorporating postfit residuals.

SWD serves as an input parameter for tomography, enabling the reconstruction of wet refractivity. In the context of this application, the fundamental tomography equation is articulated as shown in the following [32]:

$$\mathrm{SWD} = 10^{-6} \int_{\mathrm{Rec.}}^{\mathrm{Sat.}} N_w\, ds \tag{3}$$

where s denotes the ray's length, and N_w represents the wet refractivity value.

2) Three-Dimensional Wet Refractivity Reconstruction Through Tomographic Solution: In this article, voxel-based troposphere tomography is implemented. While (3) is well-suited for addressing continuous spatial challenges, voxel-based tomography operates within a discrete spatial framework, partitioning space into 3-D voxels. The associated equations governing tomographic observations are similarly discretized, as outlined in the work by Flores et al. [32]

$$\mathrm{SWD}^i = 10^{-6} \sum_{j=1}^{n} N_{wj} \Delta S_j^i \tag{4}$$

where i serves as the GNSS ray counter, j represents the number of voxels, ΔS_j^i denotes the length traveled by ray i within voxel j, and N_{wj} is the wet refractivity within voxel j. The matrix representation of this equation is provided as follows:

$$\mathrm{SWD} = A \cdot N_w \tag{5}$$

where A denotes the coefficient matrix based on the distances traveled by the GNSS rays in each voxel of the tomography model and N_w represents the vector of unknowns. The model resolution matrix serves as a distinctive feature of the coefficient matrix, offering insights into the resolution and geometry of the tomographic model [49]. Consequently, it plays a crucial role in determining the optimal resolution and geometry for the tomographic model [50]. The formulation of the model resolution matrix is outlined as follows:

$$R_m = A^{\dagger} A = V_q V_q^{\mathrm{T}} \tag{6}$$

where q represents the number of singular values, $A^{\dagger}$ denotes the generalized inverse, and matrix V represents the right singular vector of the coefficient matrix. The resolution matrix is an identity matrix when the model null space is trivial [50]. Specifically, if any of the diagonal elements of the resolution matrix are trivial, the corresponding parameters will be poorly resolved [51]. The resolution matrix captures the geometry and can be computed during the experiment to determine the optimal resolution for the tomographic model. Additionally, the horizontal gradient of tropospheric delay should be taken into account to prevent an unusual increase in the number of voxels.

B. ML for Time-Series Prediction

1) RNN: Mastering Time-Series Patterns and Dependencies: RNNs, distinguished in NN architectures, feature recurrent connections enabling them to maintain a dynamic hidden state as they process sequential data. This innate recurrence allows RNNs to capture temporal dependencies and intricate patterns, making them well-suited for tasks emphasizing context [52]. In time-series data modeling, RNNs excel at grasping underlying temporal dependencies, aiding in discerning evolving patterns and trends [53]. Their sequential processing adapts to data idiosyncrasies, reducing the need for extensive feature engineering. Despite demanding computational resources, RNNs' ability to parallelize computations across time steps enhances efficiency in predicting extended sequences. The recurrent hidden state functions as a memory mechanism, proving invaluable in tasks requiring historical context, such as weather trend forecasting [54].

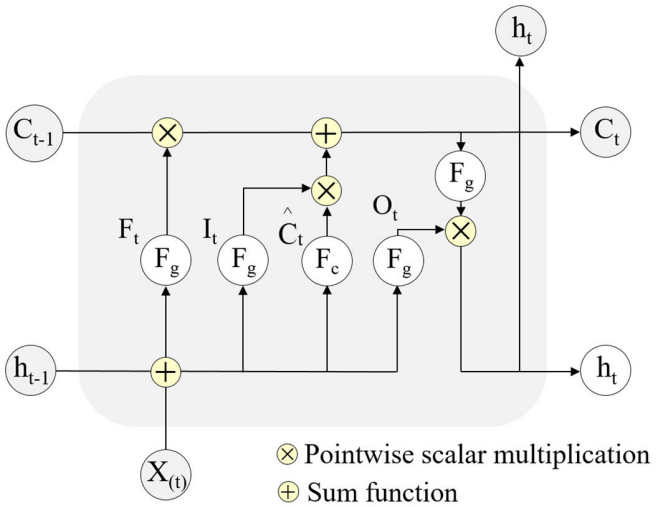


Fig. 1. Internal configuration of an LSTM unit.

2) LSTM: Overcoming the Vanishing Gradient Problem: LSTM, a specialized RNN architecture, tackles the vanishing gradient problem in traditional RNNs [55], [56]. This problem impedes learning long-term dependencies in sequences due to diminishing gradients during backward propagation. LSTM addresses this by incorporating a complex internal structure with three gates: input, forget, and output [5], [57]. The input gate controls what information should be stored in the cell state and regulates the flow of new data into the cell state. The forget gate determines what information from the previous cell state should be discarded, helping the LSTM "forget" irrelevant information from the past. The output gate decides what part of the cell state should be used to make predictions or provide outputs. These gates, in combination with the cell state, enable LSTM networks to effectively capture long-range dependencies in time-series data [5], [57]. Additionally, LSTM is inherently more parallelizable than traditional RNNs, enabling faster training and prediction times, which is especially beneficial when dealing with synoptic-scale time-series datasets.

A series of LSTM cells forms an LSTM NN, where time-series data samples represented as $h_0, h_1, \ldots, h_t$ are concurrently fed into different LSTM cells as inputs. The internal structure of an LSTM cell is depicted in Fig. 1, with the internal cell states denoted as C_t. Cell state information is updated, preserved, or deleted by the LSTM through the utilization of the forget gate (F_t), input gate (I_t), and output gate (O_t), respectively. At each time step t, the input sequence vector $X(t)$, the hidden layer output from the previous time step, h_{t-1}, and the cell state from the previous time step, C_{t-1}, are acted upon. Two primary outputs are generated by the LSTM cell: the LSTM hidden layer output, h_t, and the cell state, C_t [58].

The computation of the forget gate, input gate, and output gate is accomplished through the application of the following equations [59]:

$$I_t = F_g\left(W_I\left[h_{t-1}, X_t\right] + B_I\right) \tag{7}$$

$$F_t = F_g\left(W_F\left[h_{t-1}, X_t\right] + B_F\right) \tag{8}$$
$$O_t = F_g\left(W_O\left[h_{t-1}, X_t\right] + B_O\right). \tag{9}$$

In the provided context, W_I, W_F, and W_O represent the weight matrices for the input gate, forget gate, and output gate, respectively. Additionally, B_I, B_F, and B_O correspond to the bias terms for the input gate, forget gate, and output gate, respectively, while F_g represents the activation function for gating mechanisms. The sigmoid function outputs values between 0 and 1. It is particularly useful for modeling gating mechanisms in NNs. The present candidate cell state, denoted as $\hat{C}_t$, is determined using the following equation:

$$\hat{C}_t = F_c\left(W_C\left[h_{t-1}, X_t\right] + B_C\right). \tag{10}$$

The forget gate and the input gate ascertain the distribution of information between C_{t-1} and $\hat{C}_t$ within the current cell state C_t, respectively. Equation (11) calculates the state C_t, while (12) computes the current output of the hidden layer

$$C_t = F_t.C_{t-1} + I_t.\hat{C}_t \tag{11}$$
$$h_t = O_t.F_c(C_t). \tag{12}$$

The F_c function is employed in the computation of the current candidate cell state as an activation function.

In general, the hyperbolic tangent (tanh) function is selected for F_c because it exhibits an S-shaped curve similar to the sigmoid function, but it restricts input values to the range of -1 to 1. The tanh is frequently utilized to govern the update of the cell state due to its ability to accommodate both positive and negative values. The tanh exhibits heightened sensitivity to changes in input, which can be advantageous for capturing nonlinear relationships within data [59].

3) LSTM-GA: Hyperparameter Optimization: In time-series prediction using LSTM networks, key hyperparameters play a crucial role in shaping the model's architecture and influencing its ability to capture temporal dependencies. Optimizing these hyperparameters is essential for significantly improving predictive accuracy, especially in dealing with the intricate patterns and trends commonly observed in time-series data. The hyperparameters of LSTM and their specifications are visible in Table I.

Crucial hyperparameters for LSTM optimization include the number of neurons in hidden layers, LSTM layers, and regularization parameters like dropout rate [60]. Training-related parameters, such as learning rate and batch size, early stopping, activation functions, and optimizer, play integral roles, with poorly chosen values hindering convergence and training efficiency. However, certain parameters, like input/output layer neurons, input sequence length, loss function, and gradient clipping threshold, benefit less from extensive optimization, relying more on domain knowledge.

GAs systematically explore hyperparameter combinations for optimal solutions, enhancing prediction accuracy [61], [62]. It excels in distributed computing, concurrently exploring multiple solutions. GA's adaptability in adjusting search strategies reduces overall optimization time for LSTM parameters. Particularly useful for nondifferentiable functions,

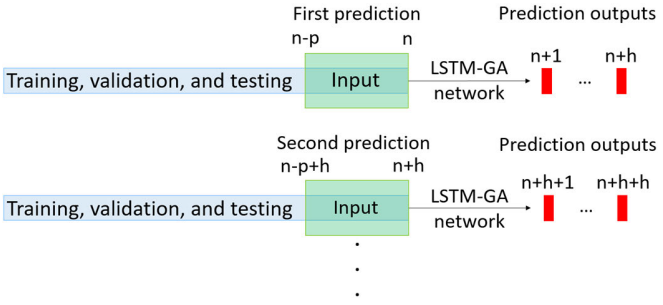


Fig. 2. Prediction algorithm used in this study.

GA, independent of gradients, is well-suited for optimizing complex time-series data [62]. The procedure commences with the initialization of a pool of potential solutions, each one corresponding to a different set of hyperparameters of the LSTM network. Then, each candidate solution undergoes an evaluation by training an LSTM model with the proper hyperparameters on the training data and testing on the separate validation data. Candidates are selected for the next generation by the way they are evaluated. Solutions demonstrating superior performance are favored, simulating the process of natural selection. Selected solutions then undergo crossover, where pairs of solutions are combined to produce offspring, mimicking genetic recombination. The offspring, as well as likely some of the initial solutions, are the next-generation population that replaces the previous one. The mentioned iterative process continues for a limited number of generations or until a specific termination condition is met, such as reaching the maximum number of iterations or meeting the convergence criteria [62].

After training and validating the LSTM-GA model, the prediction step is conducted at various time spans, including 1-, 2-, and 3-h intervals. This approach is implemented to assess the validity of the results in real-time and near real-time applications, providing insights into the practical utility of the predictions. The prediction diagram is visible in Fig. 2. In the first prediction, the model uses data from the previous p epochs to predict the next h epochs. In the next prediction, the model uses a combination of previously predicted and actual data for further prediction. The actual values from the $(n - p + h)$ to n epochs are used along with the values from $(n + 1)$ to $(n + h)$ epochs to predict the values for the next h epochs. Here, n represents the total number of epochs in the time-series dataset. In other words, once the predicted epochs were reached, the actual values of those epochs were used to predict the following epochs. The diagram outlines a two-stage prediction process. It should be noted that, to incorporate weather conditions from the last day and night before the prediction epoch, data from the previous 24 epochs preceding the prediction were used to predict subsequent epochs.

C. Ensemble Forecasting

1) Leveraging GAN for Improved Predictive Accuracy: Ensemble forecasting stands out as a powerful technique in the field of weather prediction and other scientific disciplines,

TABLE I
LSTM HYPERPARAMETERS AND THEIR ROLES AND RANGE

Hyperparameter	Role	Range
Number of neurons in input layer	Each neuron in the input layer corresponds to a feature or input variable (e.g., wet refractivity, longitude, latitude, height)	should match the number of features in dataset
Number of neurons in hidden layer	The number of LSTM units in hidden layers determines the model's capacity to capture short-term and long-term temporal dependencies in the data. More LSTM units can provide the network with greater expressive power to model complex patterns in the data. Too few LSTM units may result in underfitting, as the model might not be able to capture essential features in the sequences. Too many LSTM units may lead to overfitting if the model starts memorizing the training data instead of learning generalizable patterns	Can vary widely depending on the complexity of your prediction task and the amount of available training data
Number of neurons in output layer	The number of neurons in the output layer depends on the nature of your task.	typically have one neuron in the output layer in case of predicting single value
Number of LSTM layers	This hyperparameter determines the depth of the network. Each LSTM layer captures different levels of abstraction in the input sequence. Shallow networks may capture simple patterns, while deeper networks can capture increasingly complex and hierarchical patterns in the data. Increasing the number of LSTM layers can allow the network to learn more intricate and nuanced relationships within the sequences. However, adding too many layers can lead to overfitting if the network becomes too complex for the available data.	Can have a single layer or multiple layers (e.g., 1 to 4 layers) in a stacked LSTM architecture
Dropout rate	The dropout rate regularizes the model by preventing overfitting during training.	Commonly ranges from 0.0 (no dropout) to 0.5 (high dropout). Values outside this range are rare
Activation function	The activation function in an LSTM cell introduces non-linearity to the cell's computations, enabling the network to model complex patterns and relationships in the data (e.g., tanh, sigmoid, Rectified Linear Unit (ReLU))	Selection from a list of activation functions
Batch size	The batch size specifies the number of training examples used in each forward and backward pass during training.	Typically chosen based on available memory and computational resources. Common values include 16, 32, 64, or 128
Learning rate	The learning rate controls the step size during optimization to update model weights.	Usually between 0.001 and 0.1, but it can vary depending on the optimization algorithm and specific problem
Number of epochs	The number of epochs defines the number of times the entire training dataset is passed through the model during training.	Typically set based on convergence, often ranging from 10 to 100 or more.
Early stopping	The early stopping a form of regularization and a practical way to determine when to stop training the network. It keeps track of when the validation loss starts to increase or stagnate, indicating that the model is starting to overfit. Once the validation loss exceeds a predefined threshold or doesn't improve for a specified number of consecutive epochs (called the "patience"), the training is stopped early to prevent overfitting.	Configured with a patience parameter (e.g., 5 to 20 epochs)
Optimizer	The optimizer specifies the optimization algorithm used to update model weights during training (e.g., Adaptive Moment Estimation (Adam), Stochastic Gradient Descent (SGD), Root Mean Square Propagation (RMSprop).	Selection from a list of available optimizers
Input sequence length	This hyperparameter determines the length of input sequences or time steps processed by the LSTM.	Depends on the nature of the data and problem, usually determined by the dataset.
Loss function	The loss function defines the function used to calculate the error or loss during training.	Selection from available loss functions (e.g., mean squared error)
Gradient clipping	The Gradient clipping prevents the exploding gradient problem by setting a threshold for gradient values.	Threshold value (e.g., 1.0, 5.0) to limit gradient magnitudes

offering numerous advantages over traditional deterministic approaches [63]. In the deterministic approach, a single set of inputs is used to produce a single prediction. However, ensemble forecasting generates a diverse set of predictions, enhancing accuracy by capturing the inherent uncertainty and variability in complex systems. The need for diversity among ensemble members is critical, ensuring a more robust and informative forecast. The generation of ensemble members requires sophisticated methods, with one particularly valuable approach being the use of GAN. GAN, a class of ML models, is designed to generate realistic data by learning underlying patterns from a training dataset [64], [65]. The advantages of using GAN for ensemble member generation are manifold. First, GANs can learn and mimic the complex relationships within the data, enabling the creation of ensemble members that reflect the intricacies of the system under consideration. Second, GAN facilitates the generation of ensemble members with a level of diversity that might be challenging to achieve through traditional methods. GAN is composed of two interconnected NNs—a generator and a discriminator. At the heart of a GAN is the generator, an NN designed to create realistic data samples by capturing the underlying patterns within a given training dataset. In the context of ensemble forecasting, the generator is tasked with generating diverse and representative time series that encapsulate the complexities of atmospheric conditions. Through a process of iterative learning, the generator refines its ability to produce realistic data that closely resembles the characteristics of the training dataset. The discriminator, alongside the generator in a GAN, distinguishes between real and generated data. It continuously adapts as the generator refines its output, creating an adversarial process crucial for capturing intricate patterns in complex systems. In ensemble forecasting, this synergy enhances accuracy and robustness by creating diverse ensemble members mirroring observed conditions. In this article, a GAN is trained on a real time series to generate ten ensemble members encapsulating the diverse atmospheric conditions observed over time. Subsequently, the produced time series from the GAN are considered for training and validation in the LSTM model. Finally, there will be 11 sets of time series and 11 sets of predictions using the same optimal hyperparameters in LSTM. Ensemble forecasting can then be conducted by averaging the predicted values from different time series.

2) Binary Classification Metrics: Binary classification metrics are crucial in troposphere conditions forecasting, categorizing instances into two classes to predict specific weather events [66]. Key metrics include true positive (TP), true negative (TN), false positive (FP), and false negative (FN), used to calculate Precision, Sensitivity (Recall), and false positive rate (FPR). Precision measures the accuracy of positive predictions, Sensitivity captures the model's ability to identify positive instances, and FPR assesses specificity [67]. Optimal threshold selection, balancing classification errors, is pivotal for model effectiveness. It determines the decision boundary, influencing the model's real-world applicability and performance metrics. Adjusting the threshold allows customization based on specific needs, prioritizing sensitivity or precision. The threshold, determining when an instance is positive or negative, influences the tradeoff between evaluation metrics. In troposphere conditions forecasting, this optimization ensures accurate predictions aligned with application requirements. This article aims to balance TP and TN instances by identifying a threshold that minimizes the difference between them. This approach simplifies decision-making and is advantageous for classifiers that need to perform well on both positive and negative instances without favoring either class. Aligning with the intuitive notion of achieving balance, particularly in applications where both types of errors are equally undesirable, this method offers a robust approach in situations with unknown or challenging-to-quantify costs or consequences of FPs and FNs. Fig. 3 illustrates the flowchart representing the methodology employed in this study.

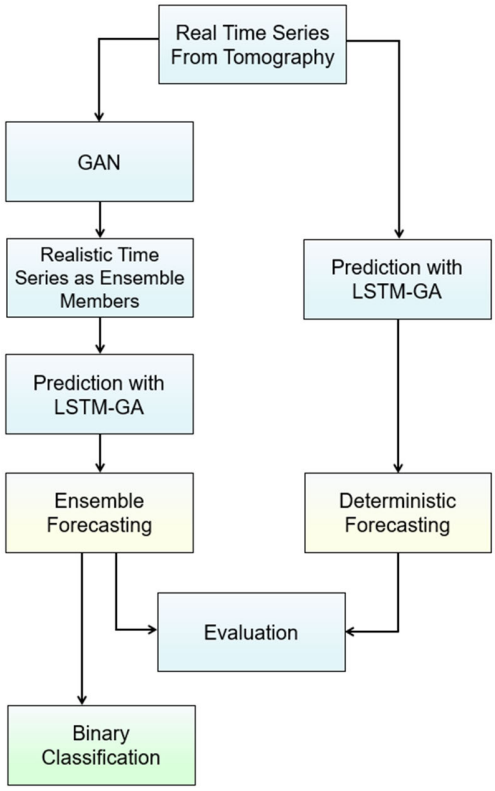


Fig. 3. Procedural steps followed throughout this article.

III. Study Area and DataSet

In order to gain a more complete assessment of the method, two study areas located on different continents with two different climate and weather conditions have been selected. The first area, located in Europe, is Poland with a moderate climate and the second one is California in North America with a Mediterranean climate. Meteorological reports from September 29 and 30, 2021, indicate the extension of sweeping rain bands over Poland, with a specific emphasis on its western and central regions. This occurrence serves as a valuable case study for evaluating the method proposed in this article for ensemble forecasting of both synoptic-scale and typical atmospheric phenomena. Moreover, in the last week of July

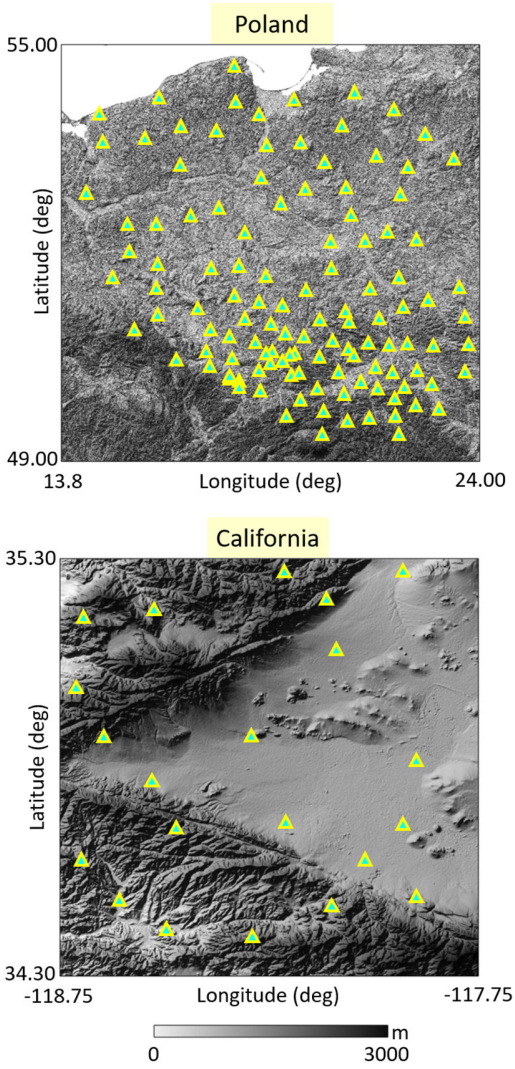


Fig. 4. Distribution of GNSS stations on the background of topography.

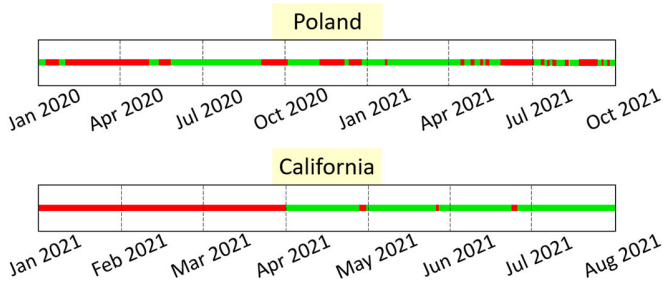


Fig. 5. Temporal distribution of processing days, represented by green lines.

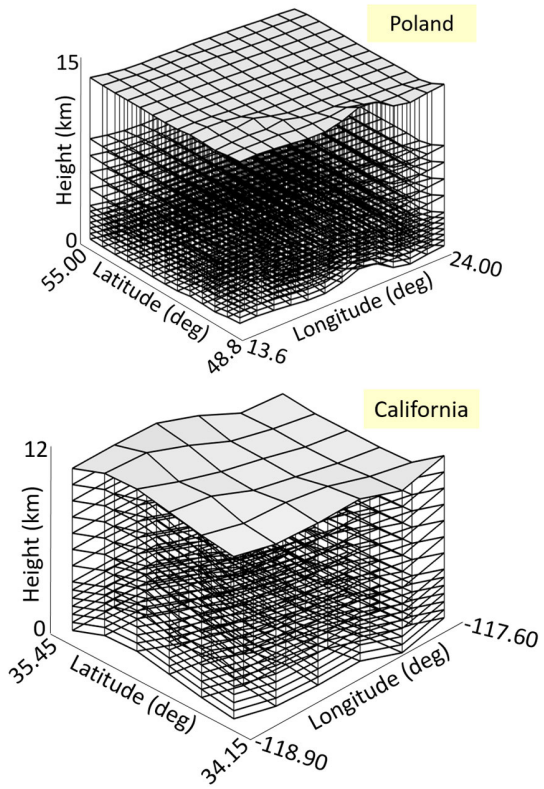


Fig. 6. Tomography skeletons for study areas.

2021, the onset of the monsoon season displayed significant activity characterized by a highly robust low-pressure system moving from East to West through the Southwestern region of the United States, particularly near California. This event led to a range of consequences, including damage from high winds in specific areas, heavy precipitation, and numerous instances of flash floods. Initially, this weather phenomenon exhibited a combination of thunderstorms and concentrated rainfall. Subsequently, it transitioned into widespread moderate rain, with certain localized areas experiencing even heavier precipitation. As a result, these events serve as suitable case studies for assessing the effectiveness of the proposed method in ensemble forecasting of local-scale atmospheric events.

In order to estimate SWD as input for troposphere tomography, a collection of observations from GNSS stations was compiled, encompassing a wide range of weather conditions. Fig. 4 shows the geographical location of the study area and the distribution of GNSS stations. In Poland, GNSS observations gathered on 300 days in 2020 and 2021, and in California, GNSS observations covered a span of 108 days. Fig. 5 depicts the temporal distribution of processed days in these two different areas. To thoroughly assess the tomographic results across the tomography models and validate the obtained predicted wet refractivity, the Weather Research and Forecasting (WRF) model was used. Table II shows the physical parameters used for running the WRF models. The physical configuration of the WRF model in Poland is similar to the one described by Kryza et al. [68] and Werner et al. [69].

IV. Processing and Results

A. Tomography Results and Validation

First, the GNSS observations were processed using Bernese 5.2 software [80]. In this processing, the ZTD and its horizontal gradients were estimated at 1-h resolution. The Vienna mapping function was employed to parameterize the mapping of tropospheric delays to the vertical direction. In the first case study, Poland, GNSS troposphere tomography was generally performed using the TOMO2 options developed

TABLE II
PHYSICS SCHEMES USED IN THE WRF CONFIGURATION

Specifications	Used scheme in Poland	Used scheme in California
Longwave radiation	Rapid Radiative Transfer Model for Global climate models (RTMG) [70]	RTMG [70]
Shortwave radiation	RTMG [70]	Dudhia [75]
Land surface	Noah [71]	Noah-MP [76]
Planetary boundary layer	Mellor-Yamada-Janic [72]	Mellor-Yamada and Nakanishi-Niino [77]
Microphysics	Morisson double-moment [73]	Thompson [78]
Cumulus parameterization	Grell 3D [74]	Kain-Fritsch [79]
Initial and boundary condition	National Centers for Environmental Prediction-Final Analysis (NCEP-FNL)	NCEP-FNL
Spatial resolution of inner domain	4×4 km	5×5 km
Temporal resolution	1 hour	1 hour

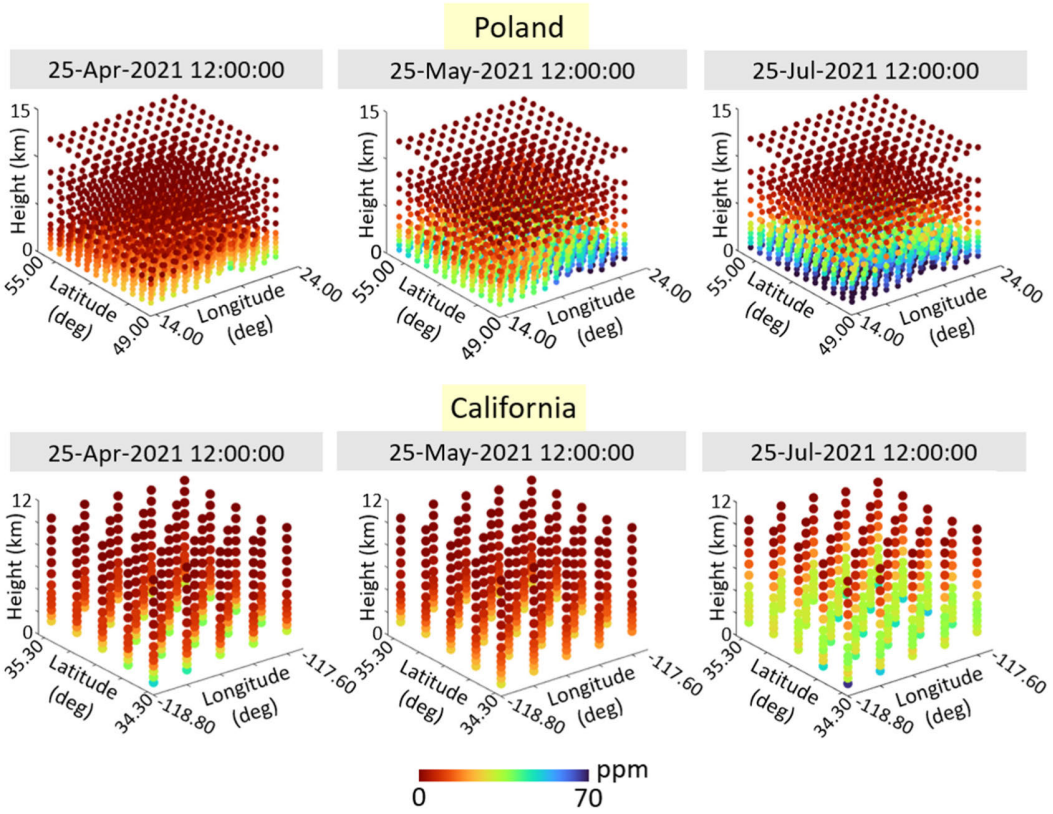


Fig. 7. Three-dimensional reconstructed wet refractivity in three-sample epochs.

at Wrocław University of Environmental and Life Sciences, Wrocław, Poland [33], [81], [82]. In the second case study, California, tomography was primarily based on the details mentioned in [29], [37], and [83]. It should be noted that the spatial resolution of the tomography models in the two case studies was determined based on the resolution matrix and tropospheric horizontal gradients [49]. In Poland, the tomography skeleton consisted of 12 × 12 × 11 voxels with a horizontal resolution of 0.8° in the x-direction and 0.5° in the y-direction across 11 irregularly distributed vertical layers with widths ranging from 500 to 5500 m. In California, the skeleton comprised 5 × 5 × 13 voxels with a horizontal resolution of 0.2° in both directions across 13 irregularly distributed vertical layers with widths ranging from 500 to 1000 m. Fig. 6 illustrates the tomography skeleton considered for the two study areas. Fig. 7 shows the examples of 3-D reconstructed wet refractivity in two areas during three different epochs. As is visible in Fig. 7, wet refractivity exhibits high variation in different months of the year and can be categorized into low, middle, and high values. As expected, the amount of wet refractivity in July is greater than that in May and April. After the wet refractivity was reconstructed,

TABLE III
STATISTICAL RESULTS OF VALIDATION OF TOMOGRAPHY RESULT WITH WRF MODEL OUTPUTS

Case study	Height (m)	Condition	RMSE (ppm)	MAE (ppm)	NSE
Poland	Height <= 3000	Rainy	3.51	2.83	0.82
		Non-Rainy	3.11	2.21	0.85
	3000 < Height <= 6000	Rainy	2.87	2.37	0.87
		Non-Rainy	2.32	1.68	0.88
	Height > 6000	Rainy	0.69	0.36	0.92
		Non-Rainy	0.41	0.23	0.90
California	Height <= 3000	Rainy	2.01	1.64	0.89
		Non-Rainy	1.88	1.29	0.91
	3000 < Height <= 6000	Rainy	1.31	1.04	0.91
		Non-Rainy	1.24	1.09	0.90
	Height > 6000	Rainy	0.38	0.21	0.87
		Non-Rainy	0.32	0.17	0.89

the need arose to validate the results that had been obtained. As mentioned previously, the obtained tomography results were evaluated using WRF model outputs. For this purpose, the negative values that are sometimes produced in the tomography solution at higher levels, where wet refractivity is very close to zero, were eliminated from the results. Apart from negative wet refractivity values, uncertainties in the tomographic reconstruction process can arise from factors such as measurement errors and model assumptions. According to the statistical analysis, negative wet refractivity values were found to comprise 1.89% and 1.13% of all data in Poland and California, respectively. Following the removal of negative values, outliers were identified through the utilization of a boxplot. The boxplot's whiskers were delineated in accordance with the interquartile range, specifically calculated as 1.5 times the difference between the first quartile and the third quartile. It should be noted that the first case study, Poland, is characterized by a significantly large area compared to the second one, and it exhibits a much higher average total precipitation. Poland experiences different weather conditions as well as varying amounts of precipitation. Therefore, to facilitate the identification of outliers and to account for the effect of precipitation on the variation of wet refractivity, the tomography days in Poland were categorized into rainy (daily total precipitation equal to or greater than 1 mm) and nonrainy days based on the total precipitation from WRF model outputs. In the case of California, the area is much smaller than that of Poland, and no distinct climate or weather conditions are observed in different parts of the tomography model. Additionally, the average precipitation in California is much lower than in Poland. Nevertheless, to achieve greater accuracy in outlier detection, the tomography days in California were also categorized into rainy and nonrainy days. To provide a more comprehensive analysis of the results, statistical parameters, including root-mean-squared error (RMSE), mean absolute error (MAE), and Nash–Sutcliffe efficiency (NSE), are presented in Table III [84]. For

TABLE IV
SEARCH RANGE FOR LSTM HYPERPARAMETERS

Hyperparameter	Range
Number of LSTM layers	1 to 5
Number of neurons	10 to 200
Dropout rate	0.2 to 0.5
Activation functions	tanh, sigmoid, and ReLU
Batch size	16, 32, 64, and 128
Learning rate	0.001 to 0.01
Early stopping parameters	Range for Patience: 5 to 20 Range for validation loss threshold: 0.001
Optimizer	Adam, RMSprop, and SGD

both case studies, Poland and California, the statistical results are presented separately for different height categories (Height $\leq$ 3000 m, 3000 $<$ Height $\leq$ 6000 m, and Height $>$ 6000 m) and for both rainy and nonrainy conditions. The results show variations in RMSE, MAE, and NSE across these categories, reflecting the differences in accuracy and model performance under varying conditions and altitudes. In both case studies, Poland and California, RMSE values consistently demonstrate a trend: they decrease with increasing altitude. RMSE values are generally higher during rainy conditions compared to nonrainy conditions in both case studies, reflecting the challenges associated with accurately modeling wet refractivity. MAE measures the average absolute differences between the tomography data and the WRF model outputs. Similar to RMSE, the MAE values of these differences both in Poland and California display a consistent pattern: they decrease with increasing altitude. Like RMSE, MAE values are generally higher during rainy conditions compared to nonrainy conditions in both case studies. NSE evaluates the relative accuracy of the tomography data compared to the WRF model outputs, with a value of 1 indicating a perfect match. Across all height categories and weather conditions, NSE values consistently indicate good agreement between the tomography data and the WRF model outputs in both case studies. In Poland, NSE values range from 0.82 to 0.92, while in California, they range from 0.87 to 0.91.

B. Optimal Hyperparameters

In this section, the prediction procedure is described. First, it is necessary to optimize LSTM hyperparameters. As mentioned before, the GA algorithm was used for this purpose. Based on the role of LSTM hyperparameters in time-series prediction, eight hyperparameters were optimized using GA. Table IV shows the considered range for selecting their optimal values. However, how to apply the LSTM-GA model to the 3-D tomography model influences the modeling of complex data dependencies with spatial variability. Using a single LSTM-GA model for the entire 3-D grid simplifies implementation and ensures computational efficiency, suitable for generalization but potentially oversimplifying irregular spatial distribution complexities. Alternatively, employing separate networks for each height level recognizes height-specific

TABLE V
OPTIMAL HYPERPARAMETERS FOR THE FIRST LSTM NETWORK

Hyperparameter	Optimal value for Poland	Optimal value for California
Number of LSTM layers	3	3
Number of neurons	105	90
Dropout rate	0.3	0.3
Activation functions	tanh	tanh
Batch size	32	64
Learning rate	0.005	0.006
Early stopping parameters	Patience: 10 Threshold: 0.001	Patience: 10 Threshold: 0.001
Optimizer	Adam	Adam

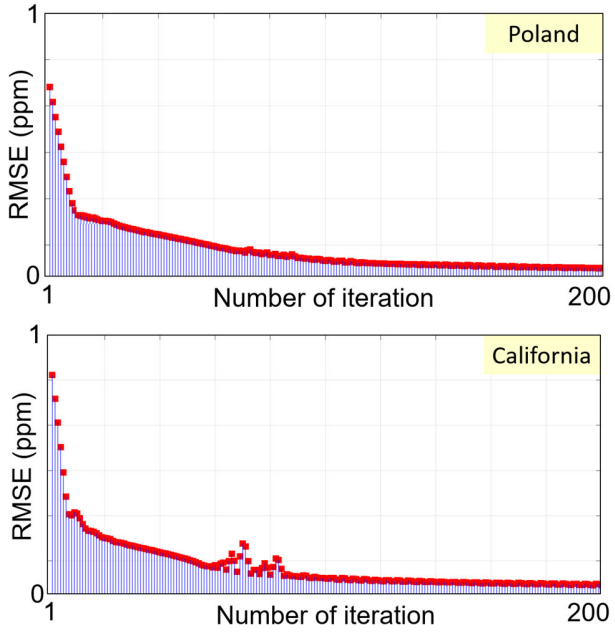


Fig. 8. Variations in RMSE during the training process using optimal hyperparameters.

TABLE VI
STATISTICAL PARAMETERS FOR COMPARING THE PREDICTED WET REFRACTIVITY WITH WRF MODEL OUTPUTS AND TOMOGRAPHY RESULTS

Case study	Height (m)	Time span	RMSE (ppm)		MAE (ppm)		NSE	
			WRF	Tomo	WRF	Tomo	WRF	Tomo
Poland	Height <= 3000	1-hour	2.58	2.48	2.19	2.15	0.87	0.88
		2-hour	3.63	3.39	2.58	2.89	0.84	0.88
		3-hour	4.15	3.56	3.81	3.41	0.79	0.84
	3000 < Height <= 6000	1-hour	2.21	2.10	1.78	1.79	0.85	0.89
		2-hour	2.37	2.15	2.02	2.08	0.83	0.87
		3-hour	2.48	2.25	2.03	2.10	0.78	0.82
	Height > 6000	1-hour	0.61	0.62	0.43	0.47	0.88	0.90
		2-hour	0.63	0.59	0.45	0.46	0.88	0.89
		3-hour	0.60	0.57	0.48	0.49	0.87	0.88
California	Height <= 3000	1-hour	2.29	2.19	1.84	1.74	0.89	0.89
		2-hour	2.41	2.74	2.32	2.09	0.81	0.85
		3-hour	3.18	2.91	2.91	2.21	0.78	0.81
	3000 < Height <= 6000	1-hour	2.12	1.61	1.41	1.15	0.74	0.88
		2-hour	2.17	1.78	1.63	1.39	0.80	0.86
		3-hour	2.28	2.14	1.72	1.63	0.84	0.83
	Height > 6000	1-hour	0.71	0.71	0.28	0.24	0.81	0.89
		2-hour	0.78	0.78	0.33	0.39	0.86	0.86
		3-hour	0.65	0.69	0.33	0.42	0.88	0.84

variations, crucial for addressing irregular spatial distribution but comes with computational costs and data limitations. Another approach involves individual LSTM-GA models for each grid point, accommodating localized patterns but demanding substantial computational resources and data volumes. This article employs the second scheme, utilizing common LSTM-GA networks for upper heights to reduce processing time, resulting in eight optimized and trained LSTM-GA networks for different vertical levels. Consequently, a total of eight LSTM-GA networks were optimized and trained. These eight LSTM-GA models consist of six networks for vertical levels from 0 to 6 km, one for heights between 6 and 9 km, and one for heights above 9 km. Table V displays the optimal hyperparameters for the first LSTM-GA model for tomography outputs between 0 and 1000 m, and Fig. 8 illustrates the RMSE throughout its training process.

C. Wet Refractivity Prediction Using LSTM-GA

Fig. 9 illustrates an example of predicted wet refractivity for sample epochs in two different areas. It should be noted that more than 95% of the time-series data were used for training and validation. The remaining data were reserved for testing. This allocation was essential to evaluate the predicted wet refractivity using validated data, considering both tomography results and WRF model outputs for this purpose. The prediction was conducted for three consecutive days, including September 28–30 in Poland and July 25–27 in California, with three different temporal resolutions, including 1-, 2-, and 3-h intervals.

The predicted values of wet refractivity were validated using WRF model outputs and real tomography results. Table VI indicates a comprehensive statistical validation of predicted wet refractivity. This validation was performed based on all

of the predicted wet refractivity values at the locations of the tomographic points at different vertical levels. Based on the statistical results, one noteworthy observation that stands out is the consistent and robust performance of the LSTM-GA model in short-term predictions, particularly within the 1-h time frame, when compared to its performance in longer term predictions spanning 2 and 3 h, across both case studies. In the case of Poland, when examining the RMSE for 1-h predictions, it is approximately 2.58 ppm, using the WRF model outputs as a reference point. However, this RMSE metric experiences an upward trajectory, reaching 3.63 ppm when extending the prediction horizon to 2 h.

Similarly, in California, a parallel increase in RMSE values from 2.29 ppm for 1-h predictions to 3.18 ppm for 3-h predictions is observable when compared to the WRF model outputs. These compelling trends paint a clear picture of the LSTM-GA model's proficiency in capturing and responding to rapid fluctuations in wet refractivity but also underscore its inherent challenge in maintaining precision over longer periods. This trend is further substantiated when examining the MAE values, which show a similar pattern of smaller errors for 1-h predictions and progressively larger errors for predictions covering longer time spans. Interestingly, the regional context plays a pivotal role in this narrative. In both case studies, the model's performance experiences a relative decline across various altitude ranges. For instance, at altitudes exceeding 6000 m in Poland, RMSE values for 1-h predictions drop as low as 0.61 ppm when compared with the WRF model outputs. In California, accuracy tends to decrease with increasing altitude, especially within the intermediate altitude range of 3000–6000 m. Regarding the NSE, a noteworthy observation emerges—short-term predictions (1 h) consistently yield higher values compared to predictions covering longer time spans. In the case of Poland, NSE values for 1-h predictions range from 0.85 to 0.88 when comparing the model with WRF data, while NSE values for 3-h predictions range from 0.78 to 0.87. A similar trend is discernible in California, where 1-h predictions consistently outperform 2- and 3-h predictions in terms of NSE values.

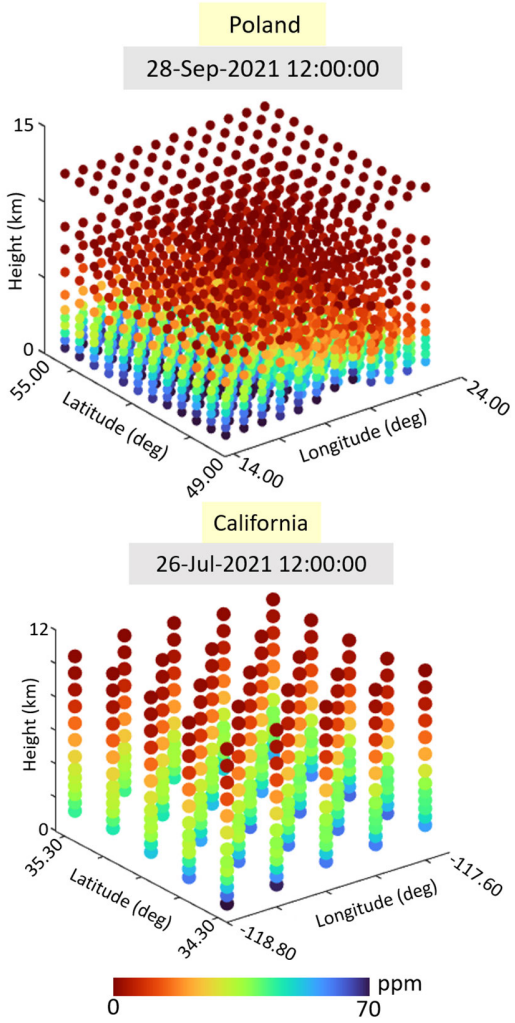


Fig. 9. Predicted 3-D wet refractivity in a sample epoch using LSTM-GA based on one-hour prediction.

D. Evaluation and Binary Classification

In this step, and in order to conduct ensemble forecasting, GAN was employed to generate realistic time series. Subsequently, 3-h predictions were made using the generated time series, employing the same LSTM structure and hyperparameters used for predictions based on real time series. In the next step, the ensemble prediction was formulated using ten realistic sets of predictions and one real set of predictions, employing the average method. The prediction started a few hours before the entrance of the phenomena in the two study areas. To evaluate the ability of ensemble forecasting, cross sections of the predicted model at different height levels were considered. Figs. 10 and 11 display samples of cross sections in the first layer of the model, selected based on the performance of ensemble forecasting compared to the deterministic forecasting method. In Fig. 10, it can be observed that ensemble forecasting performed better than the deterministic approach in the second hour of prediction. However, by hour 12, the results of the two methods were very close to each other and to the real data. By hour 42, the results were still close; however, the ensemble forecasting demonstrated greater success in reconstructing the pattern of the sweeping rain bands in the area. Analyzing the sample epochs' hours reveals the visibility of the arrival and departure of sweeping rain bands. It can be inferred that by hour 42, the occurrence is leaving the area. In Fig. 11, which is related to California, during the first hour of prediction, it can be seen that ensemble forecasting was not useful at all, and the results of ensemble and deterministic forecasting are almost the same. By hour 25, the deterministic method could not reconstruct the wet refractivity variations due to the storm properly, while ensemble forecasting demonstrated its effectiveness in reconstructing variations related to the storm in the area. In hour 39, when the storm is leaving the area, the results of the two techniques are close; however, the deterministic method worked slightly better than ensemble forecasting. To further assess the ability of ensemble forecasting, statistical parameters are presented in Table VII.

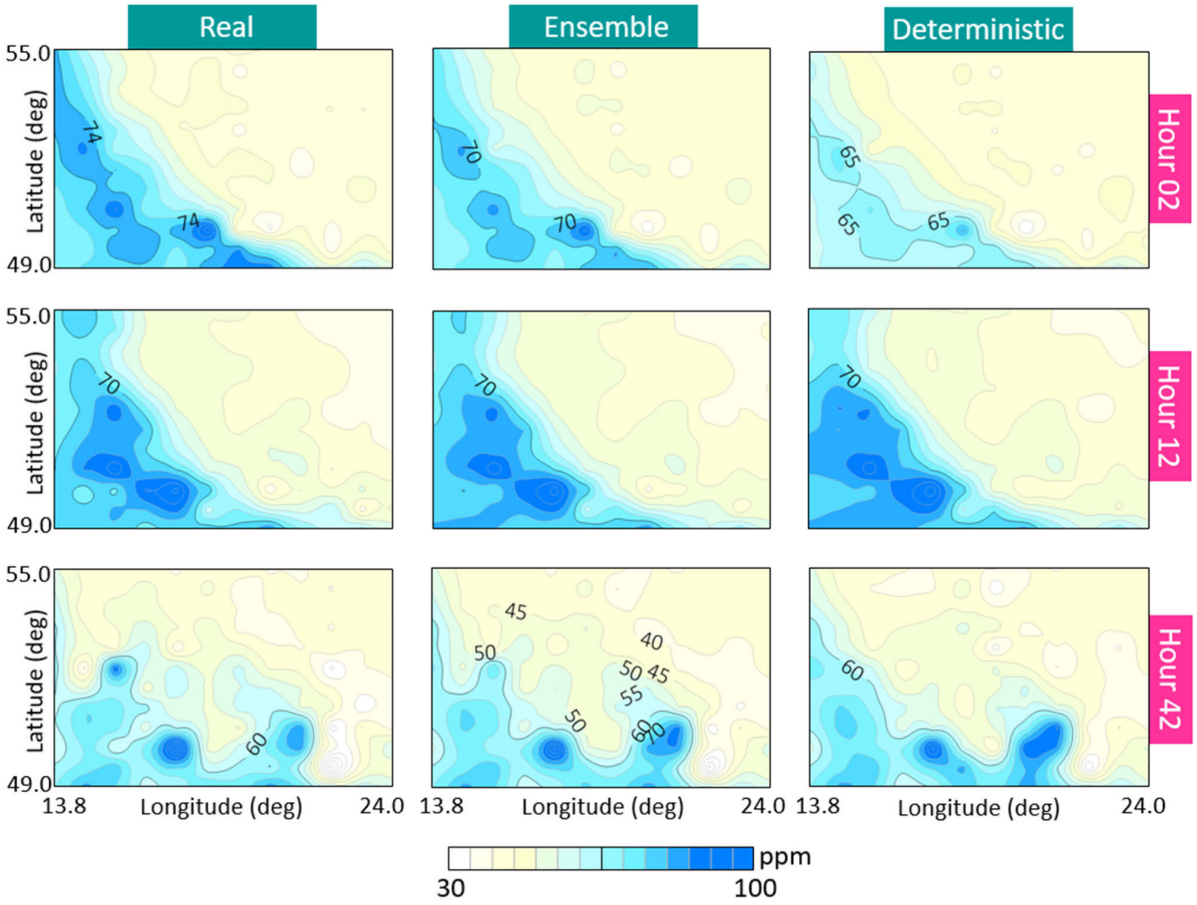


Fig. 10. Samples of obtained prediction against real tomography data in the first layer in Poland.

TABLE VII
STATISTICAL PARAMETERS FOR COMPARING ENSEMBLE AND DETERMINISTIC RESULTS WITH TOMOGRAPHY

Case study	Height (m)	RMSE (ppm)		MAE (ppm)		NSE	
		Ensem	Deter	Ensem	Deter	Ensem	Deter
Poland	Height <= 3000	2.68	3.56	2.39	3.41	0.89	0.84
	3000 < Height <= 6000	1.69	2.25	1.53	2.10	0.87	0.82
	Height > 6000	0.49	0.57	0.48	0.49	0.90	0.88
California	Height <= 3000	2.04	2.91	1.40	2.21	0.87	0.81
	3000 < Height <= 6000	1.43	2.14	0.98	1.63	0.91	0.83
	Height > 6000	0.61	0.69	0.38	0.42	0.90	0.84

In the case of Poland, at heights below 3000 m, the ensemble forecasting yields an RMSE of 2.68 ppm, significantly lower than the corresponding 3.56 ppm from the deterministic forecasting. This trend persists across different height levels, demonstrating the superiority of ensemble forecasting with RMSE values of 1.69 (versus 2.25 ppm) and 0.49 ppm (versus 0.57 ppm) for heights between 3000 and 6000 m and above 6000 m, respectively. Similar trends are observed in California, where the ensemble method consistently produces lower RMSE values compared to deterministic forecasting. For heights below 3000 m, the ensemble RMSE is 2.04 ppm, notably lower than the deterministic RMSE of 2.91 ppm. This pattern persists for heights between 3000 and 6000 m (ensemble RMSE: 1.43 ppm versus deterministic RMSE: 2.14 ppm) and heights above 6000 m (ensemble RMSE: 0.61 ppm versus deterministic RMSE: 0.69 ppm). These findings suggest that the ensemble forecasting approach, leveraging GAN-generated time series, enhances the accuracy of predictions across various height levels. The consistently lower RMSE values in the ensemble method compared to deterministic forecasting indicate its effectiveness in capturing and predicting weather condition in both Poland and California. After examining the statistical parameters of the ensemble and deterministic methods, the evaluation metrics for binary classification were reviewed. The first step involved finding the optimal threshold for the ensemble model. As depicted in the process of determining, the optimal threshold involved a meticulous analysis of various thresholds, assessing the disparities in the number of TP and TN instances. Fig. 12 illustrates this examination, showcasing how the differences in TP and TN instances

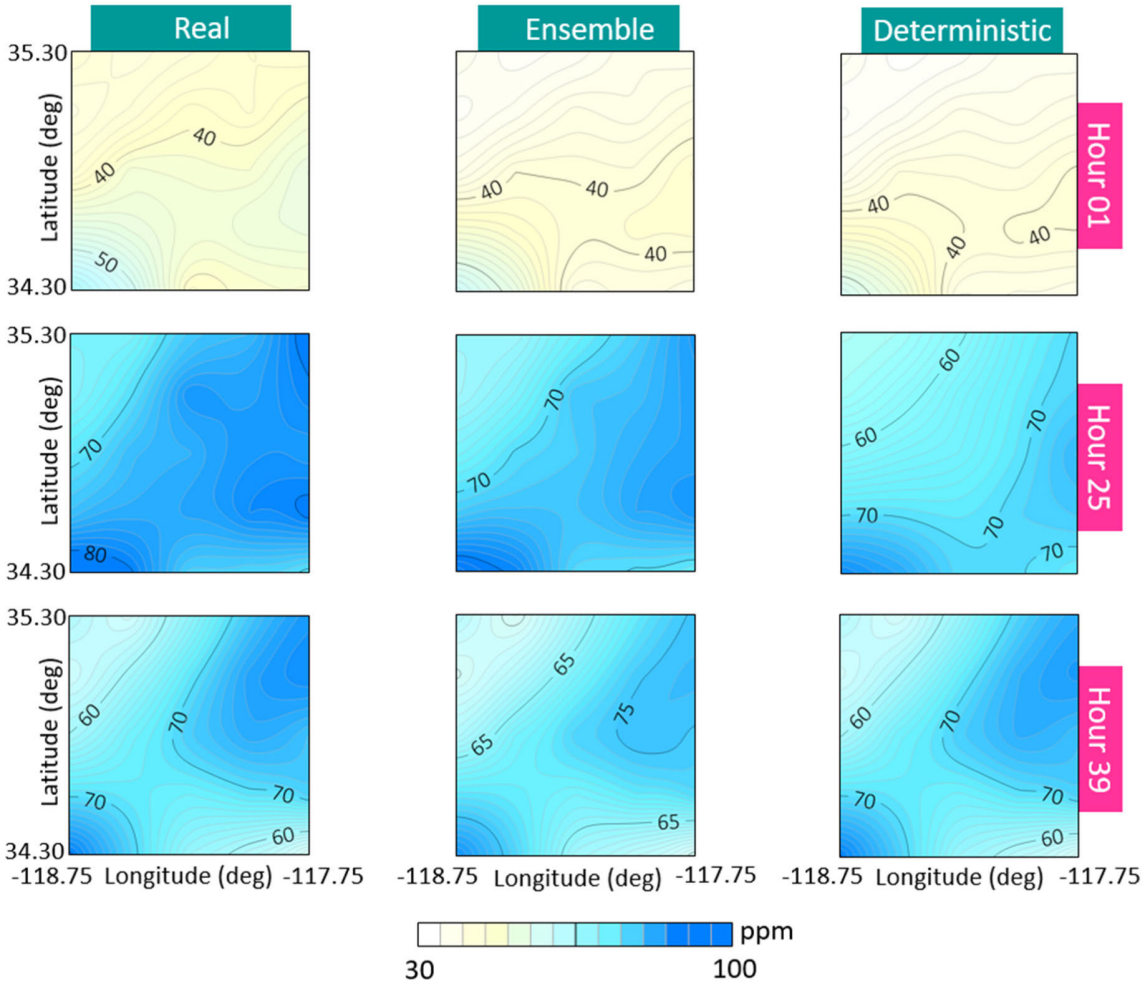


Fig. 11. Samples of obtained prediction against real tomography data in the first layer in California.

change across different threshold values. This detailed analysis culminated in the identification of distinct optimal thresholds for the two regions under consideration. For Poland, the optimal threshold was determined to be 0.41, signifying a point at which the balance between TP and TN instances is optimized. This threshold suggests a careful consideration of the tradeoff between capturing TPs and avoiding FPs, emphasizing a nuanced approach to classification. In the context of California, a different optimal threshold emerged at 0.52. This threshold indicates a tailored calibration to the characteristics of the California dataset, optimizing the balance between TP and TN instances in a manner specific to the area. After the optimal threshold was found, and to better assess the performance of the ensemble method, the tradeoff between FPR, sensitivity, and precision was plotted based on different threshold values in Fig. 13.

At varying threshold values, the FPR–sensitivity diagram, called the receiver operating characteristic (ROC) curve, demonstrates the inherent tradeoff between sensitivity and the FPR. At lower thresholds, sensitivity is high, implying effective identification of TPs, but this is accompanied by an increase in the FPR. As the threshold rises, the FPR decreases, signifying a more cautious approach in labeling positive instances but potentially missing some TPs. In contrast, the sensitivity–precision diagram reveals distinct trends. For lower thresholds, sensitivity tends to be higher, capturing a greater proportion of TPs. Table VIII shows the evaluation metrics for binary classification at the optimal threshold. In Poland, the model exhibited a high sensitivity of 0.967, indicating its ability to correctly identify a significant proportion of actual positive instances. The precision of 0.973 underscores the accuracy of positive predictions while maintaining a minimal FPR. The FPR of 0.027 reflected a low occurrence of misclassifying negative instances as positive. These results collectively suggest a robust performance in Poland, striking a favorable balance between correctly identifying positive instances and minimizing FPs. Similarly, in California, the ensemble model demonstrated even higher sensitivity at 0.982, emphasizing its superior capability to capture TPs. The precision of 0.993 highlighted the precision and accuracy of positive predictions, with an impressive reduction in FPs. The FPR of 0.011 further supported the model's proficiency in minimizing misclassifications of negative instances. These results indicate a strong performance in California, with an emphasis on both high sensitivity and precision. The ensemble model's consistent high performance in both

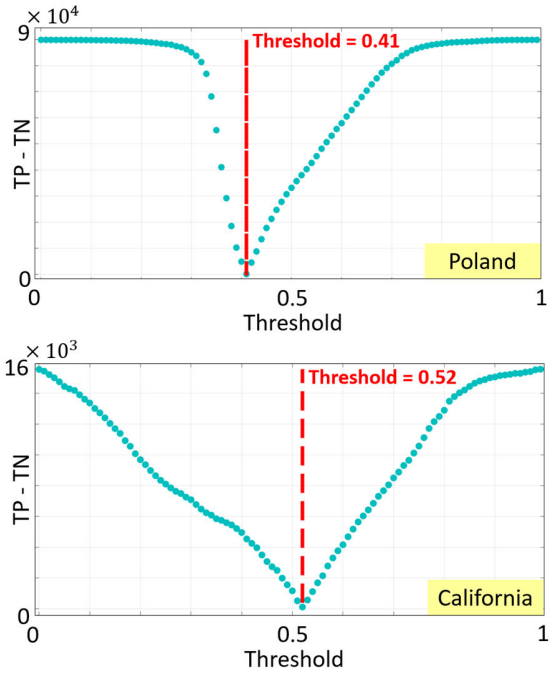


Fig. 12. Difference between the number of TP and TN at various thresholds.

TABLE VIII
EVALUATION METRICS FOR ENSEMBLE FORECASTING

Case study	Optimal threshold	Sensitivity	Precision	FPR
Poland	0.41	0.967	0.973	0.027
California	0.52	0.982	0.993	0.011

regions suggests its adaptability and reliability across diverse datasets.

V. DISCUSSION

The tomographic-based ensemble forecasting approach presented in this article represents a significant advancement in improving our understanding and prediction capabilities for a range of atmospheric phenomena such as sweeping rain bands and storms, with a specific focus on both synoptic-scale and local-scale events in two distinct areas exhibiting various climate conditions. The foundation of the ML-based troposphere conditions nowcasting model using GNSS products relies on tropospheric information extracted from the GNSS measurements. Previous studies have obtained this information from either one to three GNSS stations [10], [12], [13]. Alternatively, some studies gathered data from a network of GNSS stations in a specific area [7], [15]. One of the most crucial challenges in the proposed models by previous studies arises from the low spatial resolution and irregular spatial distribution of data, as well as the absence of values at height levels above the surface. To address this gap, we implemented tropospheric tomography techniques to generate 3-D time series of wet refractivity, serving as fundamental data for weather parameters prediction such as rainfall and storms. In prior research, we utilized

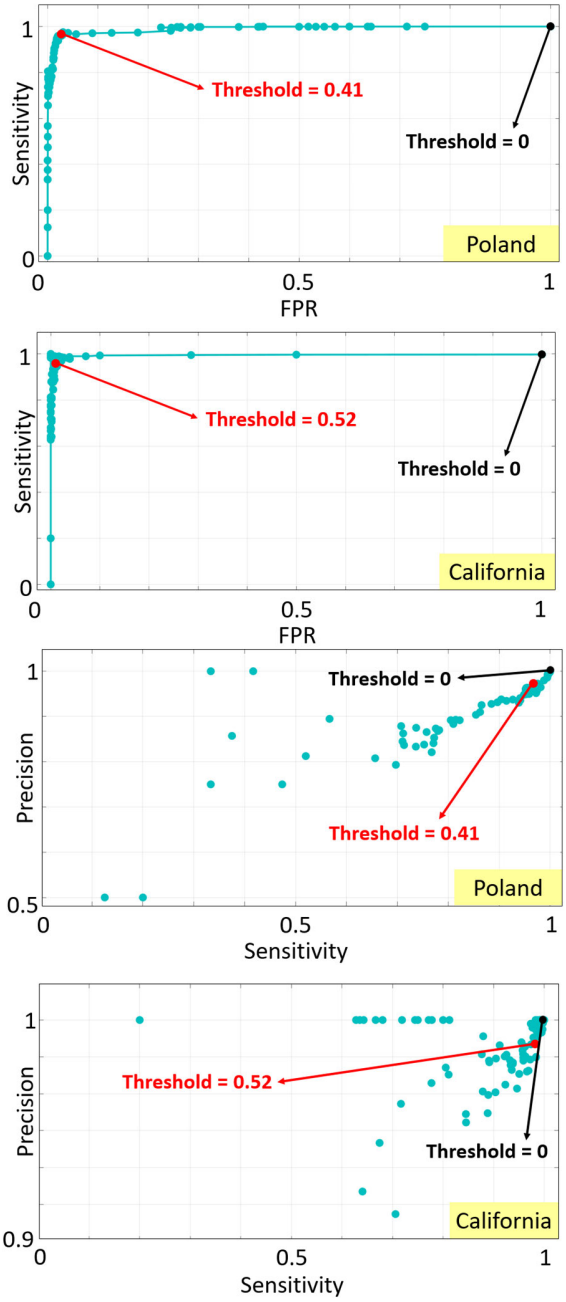


Fig. 13. Tradeoff between sensitivity, FPR, and precision at different threshold values.

troposphere tomography results to enhance the accuracy of GNSS positioning [29], [38]. Considering the significant capability of tropospheric tomography technique, in this article, we employed this method to assess its effectiveness in addressing issues related to the low spatial resolution and spatial irregularity of the weather prediction model, as well as having values at different vertical levels. The statistical results indicated that tomography is a promising method in the field of troposphere conditions forecasting, in addition to GNSS positioning. Table IX provides a comparison between the specifications of prior studies and this article. Another deficiency in previous ML-based weather forecasting models is the use of traditional ML methods that

TABLE IX
SPECIFICATIONS OF GNSS-BASED FORECASTING MODELS

Author	Study area	GNSS Products	ML Model	Forecasting Model	Phenomena
[10]	• Singapore	• Two stations	• BP-NN	• Deterministic	• Rainfall
[11]	• Lisbon	• One station	• Nonlinear AutoRegressive with eXogenous inputs (NARX)	• Deterministic	• Rainfall
[12]	• Tehran	• One station	• MLP • NARX	• Deterministic	• Rainfall
[13]	• Singapore	• Two stations	• Support Vector Machine (SVM)	• Deterministic	• Rainfall
[14]	• Yunmao	• Three stations	• SVM	• Deterministic	• Rainfall
[7]	• Israel	• A Network	• SVM • Random Forest (RF) • MLP	• Deterministic	• Flash flood
[15]	• Switzerland	• A Network	• Adaptive Boosting (AdaBoost) • Gradient Boosting (GB) • MLP • RF • Support Vector Classifier (SVC) • K-Nearest Neighbor (KNN)	• Deterministic	• Foehn winds
[9]	• India	• One station	• Rational Quadratic Gaussian Process Regression (RQ-GPR) • NN	• Deterministic	• Cyclonic storm
[8]	• Italy	• One station	• LSTM	• Deterministic	• Rain • Wind speed
This study	• Poland • California	• Troposphere tomography	• LSTM-GA	• Deterministic • Ensemble	• Rain bands • Storm

are not rigorous in capturing the temporal dependence of data. As shown in Table IX, only one study utilized LSTM, a powerful RNN method, for time-series prediction [8], while other studies relied on traditional methods. In this article, in addition to using the LSTM method, we optimized the hyperparameters using GA to ensure the best performance of the troposphere parameters prediction method. An additional challenge observed in prior research lies in the prevalent use of deterministic forecast methods, which grapple with uncertainties in initial conditions, resulting in significant deviations in predictions. The reliance on a deterministic forecast in normal forecasting meteorological models introduces inherent biases, often failing to accurately represent the intricacies of complex atmospheric processes. To fill this gap, we utilized a GAN-based ensemble forecasting method, showcasing its enhanced predictive abilities for atmospheric phenomena compared to the traditional deterministic forecasting approach. Additionally, the proposed model in this study, based on ensemble forecasting, provides the ability to deliver probability predictions, enhancing decision-making by offering a more nuanced understanding of the uncertainty associated with predictions. None of the studies that were looking into GNSS-based prediction ever used ensemble forecasting. Another challenge in the field is that an effective forecasting model should demonstrate proficiency across diverse regions and climates, accurately predicting various weather phenomena. To validate our model, we applied it to two distinct areas, with different climate conditions: Poland and California. In addition, a powerful forecasting model should be able to predict different kinds of atmospheric phenomena; that is why we tested our model with forecasting sweeping rain bands and storms. It should be noted that previous research focused on a specific phenomenon except a study done by Chkeir et al. [8] that focused on two phenomena, including rainfall and extreme wind speed; however, this investigation refers to a specific area with specific climate conditions. Importantly, as of authors' knowledge cutoff date, the mentioned gaps have not been explored in previous research.

In spite of these efforts, some limitations remain with the presented troposphere parameters forecasting models. The first problem is the lead time of forecasting models. The models can only forecast the troposphere conditions for a few hours. The challenge of extending this lead time to a longer forecasting horizon is related to a wider range of atmospheric factors and complex dynamic processes. The other limitation is due to the dynamics of the atmosphere and the interaction between meteorological parameters. The models currently being used may not consider the full interaction of the parameters such as temperature, humidity, and pressure, which are vital in getting the weather patterns and the occurrence of the phenomena. To disregard these interactions may result in biased forecasts, especially in regions and conditions where the dynamics of the atmosphere are complex. This study will be an initial step toward the artificial intelligence (AI) weather forecasting movement that is becoming more and more popular worldwide [85], [86], [87]. Detailed information on the critical analysis of current AI-based weather forecasting models can be found in [88].

Regarding the computational aspect of the proposed method, it is important to point out that LSTM networks require training with voluminous spatiotemporal data. Moreover, GA, being an iterative procedure, also adds to the computational cost since it conducts parameter search to find suitable configurations. Nevertheless, the improvement in hardware resources, including parallel computing architectures and graphics processing units (GPUs) which can speed up training and be used to find the best combination of hyperparameters, can be helpful. The proposed method and the obtained results in two different case studies and various phenomena exhibit promising scalability potential, enabling its application to larger datasets and broader geographical regions.

VI. Conclusion

This article introduced a novel ensemble forecasting approach that leveraged GNSS troposphere tomography, LSTM networks, and GA to predict 3-D wet refractivity fields. The method's robust performance was showcased through a thorough analysis of meteorological events, highlighting its effectiveness in predicting the movement of rain bands and storms, with specific case studies conducted in Poland and California. The focused exploration of the singular parameter, wet refractivity, within the LSTM-GA-based prediction model reflects a deliberate effort to enhance simplicity without compromising predictive accuracy. The model's adaptability to diverse climates, as evidenced by successful case studies in Poland and California, suggested broader applicability beyond the studied regions. The strategic use of separate networks for different height levels in the LSTM architecture struck a balance between efficiency and modeling accuracy, allowing the model to capture rapid fluctuations in wet refractivity and provide valuable insights for short-term predictions. The ensemble forecasting outperformed the deterministic method, emphasizing the utility of realistic time series generated by GAN. The evaluation of thresholds, indicating the point at which the balance between TP and TN instances is optimized, emphasizes the model's proficiency in distinguishing TPs and minimizing FPs. This is further underscored by the sensitivity and precision metrics, with customized threshold approaches tailored for each region. The low FPR in both Poland and California underscored the adaptability and reliability of the ensemble model across diverse datasets with varying climates. In summary, the ensemble forecasting approach introduced in this article not only advanced our understanding of atmospheric phenomena but also held promise for improving the accuracy and reliability of weather predictions in diverse and complex atmospheric contexts. Further advancements could include extending the lead time of forecasting models beyond a few hours and incorporating additional data sources, such as satellite imagery and ground-based observations, to capture a broader range of atmospheric factors and complex dynamic processes. Moreover, enhancing the models' understanding of the interactions between meteorological parameters could lead to a more accurate and reliable prediction of the tropospheric conditions.

Acknowledgment

All authors contributed to the study conception and design, material preparation, data collection, processing, and analysis. Conceptualization was done by Saeid Haji-Aghajany and Witold Rohm. Data were prepared by Witold Rohm and Maciej Kryza. Formal analysis was carried out by all authors. Processing was done by Saeid Haji-Aghajany. The original draft was written by Saeid Haji-Aghajany. Witold Rohm and Kamil Smolak provided review and editing. All authors have read and agreed to the published version of this article.

References

[1] H. Brenot et al., "Preliminary signs of the initiation of deep convection by GNSS," *Atmos. Chem. Phys.*, vol. 13, no. 11, pp. 5425–5449, Jun. 2013.

[2] E. Tunalı, "Water vapor monitoring with IGS RTS and GPT3/VMF3 functions over Turkey," *Adv. Space Res.*, vol. 69, no. 6, pp. 2376–2390, Mar. 2022.

[3] M. Bevis, S. Businger, T. A. Herring, C. Rocken, R. A. Anthes, and R. H. Ware, "GPS meteorology: Remote sensing of atmospheric water vapor using the global positioning system," *J. Geophys. Res., Atmos.*, vol. 97, no. D14, pp. 15787–15801, Oct. 1992.

[4] Y. Yao, X. Xu, C. Xu, W. Peng, and Y. Wan, "GGOS tropospheric delay forecast product performance evaluation and its application in real-time PPP," *J. Atmos. Solar-Terrestrial Phys.*, vol. 175, pp. 1–17, Oct. 2018.

[5] M. Tasan, Z. Ghorbaninasab, S. Haji-Aghajany, and A. Ghiasvand, "Leveraging GNSS tropospheric products for machine learning-based land subsidence prediction," *Earth Sci. Informat.*, vol. 16, no. 4, pp. 3039–3056, Dec. 2023.

[6] S. Haji-Aghajany, Y. Amerian, and A. Amiri-Simkooei, "Impact of climate change parameters on groundwater level: Implications for two subsidence regions in Iran using geodetic observations and artificial neural networks (ANN)," *Remote Sens.*, vol. 15, no. 6, p. 1555, Mar. 2023.

[7] S. Z. Ziv and Y. Reuveni, "Flash floods prediction using precipitable water vapor derived from GPS tropospheric path delays over the eastern Mediterranean," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5804017.

[8] S. Chkeir, A. Anesiadou, A. Mascitelli, and R. Biondi, "Nowcasting extreme rain and extreme wind speed with machine learning techniques applied to different input datasets," *Atmos. Res.*, vol. 282, Feb. 2023, Art. no. 106548.

[9] N. B. Jadala, M. Sridhar, D. Venkata Ratnam, and G. Dutta, "Assessment of machine learning techniques for prediction of integrated water vapor using meteorological data," *Vietnam J. Earth Sci.*, vol. 44, pp. 521–534, Jul. 2022, doi: 10.15625/2615-9783/17373.

[10] Y. Liu, Q. Zhao, W. Yao, X. Ma, Y. Yao, and L. Liu, "Short-term rainfall forecast model based on the improved BP-NN algorithm," *Sci. Rep.*, vol. 9, no. 1, p. 19751, Dec. 2019.

[11] P. Benevides, J. Catalao, and G. Nico, "Neural network approach to forecast hourly intense rainfall using GNSS precipitable water vapor and meteorological sensors," *Remote Sens.*, vol. 11, no. 8, p. 966, Apr. 2019, doi: 10.3390/rs11080966.

[12] A. S. Khaniani, H. Motieyan, and A. Mohammadi, "Rainfall forecast based on GPS PWV together with meteorological parameters using neural network models," *J. Atmos. Solar-Terrestrial Phys.*, vol. 214, Mar. 2021, Art. no. 105533.

[13] Y. Liu, Q. Zhao, Z. Li, Y. Yao, and X. Li, "GNSS-derived PWV and meteorological data for short-term rainfall forecast based on support vector machine," *Adv. Space Res.*, vol. 70, no. 4, pp. 992–1003, Aug. 2022.

[14] Z. Ma, G. Guo, M. Cai, X. Chen, W. Li, and L. Zhang, "A combined linear–nonlinear short-term rainfall forecast method using GNSS-derived PWV," *Atmosphere*, vol. 13, no. 9, p. 1381, Aug. 2022.

[15] M. Aichinger-Rosenberger, E. Brockmann, L. Crocetti, B. Soja, and G. Moeller, "Machine learning-based prediction of Alpine foehn events using GNSS troposphere products: First results for Altdorf, Switzerland," *Atmos. Meas. Techn.*, vol. 15, no. 19, pp. 5821–5839, Oct. 2022, doi: 10.5194/amt-15-5821-2022.

[16] J. S. H. Lo, "Estimation of tropospheric wet delay from GNSS measurements," Ph.D. dissertation, Curtin Univ. Technol., Perth, WA, Australia, 2011.

[17] A. El-Mowafy and J. Lo, "Prediction of troposphere wet delay," *J. Appl. Geodesy*, vol. 5, nos. 3–4, pp. 163–173, Jan. 2011.

[18] Y. Lai and D. A. Dzombak, "Use of the autoregressive integrated moving average (ARIMA) model to forecast near-term regional temperature and precipitation," *Weather Forecasting*, vol. 35, no. 3, pp. 959–976, Jun. 2020, doi: 10.1175/waf-d-19-0158.1.

[19] G. Xiao, J. Ou, G. Liu, and H. Zhang, "Construction of a regional precise tropospheric delay model based on improved BP neural network," *Chin. J. Geophys.*, vol. 61, no. 8, pp. 3139–3148, 2018.

[20] M. O. Selbesoglu, "Prediction of tropospheric wet delay by an artificial neural network model based on meteorological and GNSS data," *Eng. Sci. Technol., Int. J.*, vol. 23, no. 5, pp. 967–972, Oct. 2020.

[21] S. Li, T. Xu, Y. Xu, N. Jiang, and L. Bastos, "Forecasting GNSS zenith troposphere delay by improving GPT3 model with machine learning in antarctica," *Atmosphere*, vol. 13, no. 1, p. 78, 2022, doi: 10.3390/atmos13010078.

[22] E. Shehaj et al., "High-resolution tropospheric refractivity fields by combining machine learning and collocation methods to correct Earth observation data," *Acta Astronautica*, vol. 204, pp. 591–598, Mar. 2023.

[23] Y. Yang, T. Xu, and L. Ren, "A new regional tropospheric delay correction model based on BP neural network," in *Proc. Forum Cooperat. Positioning Service (CPGPS)*, 2017, pp. 96–100.

[24] Y. Shi, F. Wu, H. Zhu, and X. Han, "Prediction of tropospheric delay based on the LSTM model of Keras platform," *GNSS World China*, vol. 45, no. 6, pp. 115–122, 2020.

[25] H. Zhang, Y. Yao, C. Xu, W. Xu, and J. Shi, "Transformer-based global zenith tropospheric delay forecasting model," *Remote Sens.*, vol. 14, no. 14, p. 3335, Jul. 2022.

[26] Q. Zhang, F. Li, S. Zhang, and W. Li, "Modeling and forecasting the GPS zenith troposphere delay in West Antarctica based on different blind source separation methods and deep learning," *Sensors*, vol. 20, no. 8, p. 2343, Apr. 2020.

[27] S. Li, T. Xu, Y. Xu, N. Jiang, and L. Bastos, "Forecasting GNSS zenith troposphere delay by improving GPT3 model with machine learning in Antarctica," *Atmosphere*, vol. 13, no. 1, p. 78, Jan. 2022.

[28] H. Zhang et al., "A tropospheric zenith delay forecasting model based on a long short-term memory neural network and its impact on precise point positioning," *Remote Sens.*, vol. 14, no. 23, p. 5921, Nov. 2022.

[29] S. Haji-Aghajany, Y. Amerian, S. Verhagen, W. Rohm, and H. Schuh, "The effect of function-based and voxel-based tropospheric tomography techniques on the GNSS positioning accuracy," *J. Geodesy*, vol. 95, no. 7, p. 78, Jul. 2021.

[30] S. Maddahi, M. Tasan, and S. Haji-Aghajany, "Enhancing InSAR accuracy: Unveiling more accurate displacement fields through 3-D troposphere tomography," *J. Atmos. Solar-Terrestrial Phys.*, vol. 256, Mar. 2024, Art. no. 106207, doi: 10.1016/j.jastp.2024.106207.

[31] S. Haji-Aghajany, Y. Amerian, and A. Amiri-Simkooei, "Function-based troposphere tomography technique for optimal downscaling of precipitation," *Remote Sens.*, vol. 14, no. 11, p. 2548, May 2022.

[32] A. Flores, G. Ruffini, and A. Rius, "4D tropospheric tomography using GPS slant wet delays," *Ann. Geophys.*, vol. 18, pp. 223–234, Feb. 2000, doi: 10.1007/s00585-000-0223-7.

[33] W. Rohm and J. Bosy, "Local tomography troposphere model over mountains area," *Atmos. Res.*, vol. 93, no. 4, pp. 777–783, Aug. 2009.

[34] S. Haji-Aghajany and Y. Amerian, "Hybrid regularized GPS tropospheric sensing using 3-D ray tracing technique," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 10, pp. 1475–1479, Oct. 2018.

[35] S. Izanlou, S. Haji-Aghajany, and Y. Amerian, "Enhanced troposphere tomography: Integration of GNSS and remote sensing data with optimal vertical constraints," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 3701–3714, 2024, doi: 10.1109/JSTARS.2024.3354884.

[36] Z. Adavi, "Assessment of various processing schemes and solution strategies to improve the performance of GNSS tropospheric tomography," Ph.D. dissertation, Department Für Geodäsie und Geoinformation, Technische Universität Wien, Vienna, Austria, 2022.

[37] S. Haji-Aghajany, Y. Amerian, and S. Verhagen, "B-spline function-based approach for GPS tropospheric tomography," *GPS Solutions*, vol. 24, no. 3, p. 88, Jul. 2020.

[38] S. Haji-Aghajany, "Function-based troposphere water vapor tomography using GNSS observations," Ph.D. thesis, Fac. Geodesy Geomatics Eng., K. N. Toosi Univ. Technol., Tehran, Iran, 2021.

[39] D. Perler, A. Geiger, and F. Hurter, "4D GPS water vapor tomography: New parameterized approaches," *J. Geodesy*, vol. 85, no. 8, pp. 539–550, Aug. 2011.

[40] E. Trzcina, W. Rohm, and K. Smolak, "Parameterisation of the GNSS troposphere tomography domain with optimisation of the nodes' distribution," *J. Geodesy*, vol. 97, no. 1, Jan. 2023, Art. no. 2.

[41] F. Kleijer, "Troposphere modeling and filtering for precise GPS leveling," Ph.D. thesis, Civil Eng. Geosci., Delft Univ. Technol., Delft, The Netherlands, 2004.

[42] J. L. Davis, T. A. Herring, I. I. Shapiro, A. E. E. Rogers, and G. Elgered, "Geodesy by radio interferometry: Effects of atmospheric modeling errors on estimates of baseline length," *Radio Sci.*, vol. 20, no. 6, pp. 1593–1607, Nov. 1985.

[43] J. Boehm, J. Kouba, and H. Schuh, "Forecast Vienna mapping functions 1 for real-time analysis of space geodetic observations," *J. Geodesy*, vol. 83, no. 5, pp. 397–401, May 2009.

[44] Y. Shoji, "Retrieval of water vapor inhomogeneity using the Japanese nationwide GPS array and its potential for prediction of convective precipitation," *J. Meteorol. Soc. Japan. Ser. II*, vol. 91, no. 1, pp. 43–62, 2013.

[45] S. Haji-Aghajany and Y. Amerian, "An investigation of three dimensional ray tracing method efficiency in precise point positioning by tropospheric delay correction," *J. Earth Space Phys.*, vol. 44, pp. 39–52, Apr. 2018, doi: 10.22059/JESPHYS.2018.236885.1006913.

[46] T. Manning, W. Rohm, K. Zhang, F. Hurter, and C. Wang, "Determining the 4D dynamics of wet refractivity using GPS tomography in the Australian region," in *Earth on the Edge: Science for a Sustainable Planet* (International Association of Geodesy Symposia), vol. 139, C. Rizos and P. Willis, Eds. Berlin, Heidelberg: Springer, 2014, pp. 435–440, doi: 10.1007/978-3-642-37222-3_6.

[47] P. Hordyniec, J. Kapłon, W. Rohm, and M. Kryza, "Residuals of tropospheric delays from GNSS data and ray-tracing as a potential indicator of rain and clouds," *Remote Sens.*, vol. 10, no. 12, p. 1917, Nov. 2018.

[48] J. Böhm and H. Schuh, Eds., *Atmospheric Effects in Space Geodesy*, vol. 5. Berlin, Germany: Springer, 2013.

[49] S. H. Aghajany and Y. Amerian, "Three dimensional ray tracing technique for tropospheric water vapor tomography using GPS measurements," *J. Atmos. Solar-Terrestrial Phys.*, vol. 164, pp. 81–88, Nov. 2017.

[50] R. C. Aster, B. Borchers, and C. H. Thurber, *Parameter Estimation and Inverse Problems*. Amsterdam, The Netherlands: Elsevier, 2005.

[51] W. Menke, *Geophysical Data Analysis: Discrete Inverse Theory (Rev. Edn)* (International Geophysics Series), vol. 45. San Diego, CA, USA: Academic, 2012.

[52] Z. C. Lipton, J. Berkowitz, and C. Elkan, "A critical review of recurrent neural networks for sequence learning," 2015, *arXiv:1506.00019*.

[53] A. Graves, "Generating sequences with recurrent neural networks," 2013, *arXiv:1308.0850*.

[54] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," 2014, *arXiv:1412.3555*.

[55] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[56] A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional LSTM and other neural network architectures," *Neural Netw.*, vol. 18, nos. 5–6, pp. 602–610, Jul. 2005.

[57] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with LSTM," *Neural Comput.*, vol. 12, no. 10, pp. 2451–2471, Oct. 2000.

[58] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A search space Odyssey," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 10, pp. 2222–2232, Oct. 2017.

[59] K. Vijayaprabakaran and K. Sathiyamurthy, "Towards activation function search for long short-term model network: A differential evolution based approach," *J. King Saud Univ. Comput. Inf. Sci.*, vol. 34, no. 6, pp. 2637–2650, Jun. 2022.

[60] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.

[61] S. H. Aghajany, M. Pirooznia, M. R. Naeeni, and Y. Amerian, "Combination of artificial neural network and genetic algorithm to inverse source parameters of sefid-sang earthquake using InSAR technique and analytical model conjunction," *J. Earth Space Phys.*, vol. 45, no. 4, pp. 121–131, 2020.

[62] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*. Reading, MA, USA: Addison-Wesley, 1989.

[63] J. M. Lewis, "Roots of ensemble forecasting," *Monthly Weather Rev.*, vol. 133, no. 7, pp. 1865–1885, Jul. 2005.

[64] A. Bihlo, "A generative adversarial network approach to (ensemble) weather prediction," *Neural Netw.*, vol. 139, pp. 1–16, Jul. 2021.

[65] R. Brecht and A. Bihlo, "Computing the ensemble spread from deterministic weather predictions using conditional generative adversarial networks," *Geophys. Res. Lett.*, vol. 50, no. 2, Jan. 2023, Art. no. e2022GL101452, doi: 10.1029/2022GL101452.

[66] K. Doycheva, G. Horn, C. Koch, A. Schumann, and M. König, "Assessment and weighting of meteorological ensemble forecast members based on supervised machine learning with application to runoff simulations and flood warning," *Adv. Eng. Informat.*, vol. 33, pp. 427–439, Aug. 2017.

[67] J. Davis and M. Goadrich, "The relationship between precision-recall and ROC curves," in *Proc. 23rd Int. Conf. Mach. Learn. (ICML)*, 2006, pp. 233–240.

[68] M. Kryza, K. Wałaszek, H. Ojrzyńska, M. Szymanowski, M. Werner, and A. J. Dore, "High-resolution dynamical downscaling of ERA-interim using the WRF regional climate model for the area of Poland," *Pure Appl. Geophys.*, vol. 174, no. 2, pp. 511–526, 2017.

[69] M. Werner, M. Kryza, and P. Wind, "High resolution application of the EMEP MSC-W model over Eastern Europe—Analysis of the EMEP4PL results," *Atmos. Res.*, vol. 212, pp. 6–22, Nov. 2018.

[70] M. J. Iacono, J. S. Delamere, E. J. Mlawer, M. W. Shephard, S. A. Clough, and W. D. Collins, "Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models," *J. Geophys. Res., Atmos.*, vol. 113, no. D13, p. 13103, Jul. 2008.

[71] F. Chen and J. Dudhia, "Coupling an advanced land surface-hydrology model with the Penn state-NCAR MM5 modeling system," *Monthly Weather Rev.*, vol. 129, no. 3, pp. 569–585, 2001.

[72] Z. I. Janjic, "Nonsingular implementation of the Mellor–Yamada level 2.5 scheme in the NCEP Meso model," Nat. Centers Environ. Predict., College Park, MD, USA, Office Note Rep. 437, 2001.

[73] H. Morrison, J. A. Curry, and V. I. Khvorostyanov, "A new double-moment microphysics parameterization for application in cloud and climate models. Part I: Description," *J. Atmos. Sci.*, vol. 62, no. 6, pp. 1665–1677, Jun. 2005.

[74] G. A. Grell and D. Dévényi, "A generalized approach to parameterizing convection combining ensemble and data assimilation techniques," *Geophys. Res. Lett.*, vol. 29, no. 14, p. 1693, Jul. 2002.

[75] J. Dudhia, "Numerical study of convection observed during the winter monsoon experiment using a mesoscale two-dimensional model," *J. Atmos. Sci.*, vol. 46, no. 20, pp. 3077–3107, Oct. 1989.

[76] G.-Y. Niu et al., "The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements," *J. Geophys. Res.*, vol. 116, no. D12, 2011.

[77] M. Nakanishi and H. Niino, "An improved Mellor–Yamada level-3 model: Its numerical stability and application to a regional prediction of advection fog," *Boundary-Layer Meteorol.*, vol. 119, no. 2, pp. 397–407, May 2006.

[78] G. Thompson, P. R. Field, R. M. Rasmussen, and W. D. Hall, "Explicit forecasts of winter precipitation using an improved bulk microphysics scheme," *Monthly Weather Rev.*, vol. 136, no. 12, pp. 5095–5115, 2008.

[79] J. S. Kain and J. M. Fritsch, "A one-dimensional entraining/detraining plume model and its application in convective parameterization," *J. Atmos. Sci.*, vol. 47, no. 23, pp. 2784–2802, Dec. 1990.

[80] R. Dach, S. Lutz, P. Walser, and P. Fridez, *Bernese GNSS Software Version 5.2 User Manual*. Bern, Switzerland: Astronomical Institute, Univ. Bern, 2015, doi: 10.7892/boris.7229.

[81] W. Rohm and J. Bosy, "The verification of GNSS tropospheric tomography model in a mountainous area," *Adv. Space Res.*, vol. 47, no. 10, pp. 1721–1730, May 2011.

[82] W. Rohm, "The ground GNSS tomography—Unconstrained approach," *Adv. Space Res.*, vol. 51, no. 3, pp. 501–513, Feb. 2013.

[83] S. Haji-Aghajany, Y. Amerian, S. Verhagen, W. Rohm, and H. Ma, "An optimal troposphere tomography technique using the WRF model outputs and topography of the area," *Remote Sens.*, vol. 12, no. 9, p. 1442, May 2020.

[84] D. N. Moriasi, J. G. Arnold, M. W. Van Liew, R. L. Bingner, R. D. Harmel, and T. L. Veith, "Model evaluation guidelines for systematic quantification of accuracy in watershed simulations," *Trans. ASABE*, vol. 50, no. 3, pp. 885–900, 2007.

[85] P. Hewage, M. Trovati, E. Pereira, and A. Behera, "Deep learning-based effective fine-grained weather forecasting model," *Pattern Anal. Appl.*, vol. 24, no. 1, pp. 343–366, Feb. 2021, doi: 10.1007/s10044-020-00898-1.

[86] R. Keisler, "Forecasting global weather with graph neural networks," 2022, *arXiv:2202.07575*.

[87] R. Lam et al., "GraphCast: Learning skillful medium-range global weather forecasting," 2022, *arXiv:2212.12794*.

[88] S. Haji-Aghajany, W. Rohm, P. Lipinski, M. Kryza, 2024, "Beyond the horizon: A critical analysis of AI-based weather forecasting models," *ESS Open Arch.*, 2024, doi: 10.22541/essoar.171632600.06154714/v1.

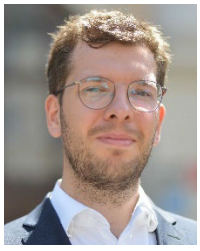
Saeid Haji-Aghajany received the Ph.D. degree in geodesy and geomatics engineering from the K. N. Toosi University of Technology, Tehran, Iran, in 2020.

He is currently a Research Fellow with Wrocław University of Environmental and Life Sciences (UPWr), Wrocław, Poland. His main research interests include Global Navigation Satellite System (GNSS) meteorology, GNSS remote sensing, and artificial intelligence (AI)-based troposphere modeling.

Witold Rohm received the Ph.D. degree from Wrocław University of Environmental and Life Sciences (UPWr), Wrocław, Poland, in 2011.

He is currently a Full Professor of remote sensing and the Head of the Institute of Geodesy and Geoinformatics, UPWr. His research interests include the integration of ground- and space-based Global Navigation Satellite System (GNSS) observations, weather data assimilation, as well as the use of artificial intelligence (AI) in geoinformatics.

Maciej Kryza is currently a Professor with the University of Wrocław, Wrocław, Poland. His research focuses on air pollution and meteorological modeling. He was responsible for the development of the air pollution forecasting system for SW Poland within the LIFE-APIS/PL Project and for Poland within the LIFE-Mapping Air Project. He is interested in applications of meteorological and chemical transport models, impacts of natural emission on air quality, and joint applications of Eulerian and machine learning models for air quality forecasting.



Kamil Smolak received the Ph.D. degree in geoinformatics from Wrocław University of Environmental and Life Sciences (UPWr), Wrocław, Poland, in 2022.

He is currently an Assistant Professor with UPWr. His research focuses on understanding the complexity of human and animal mobility as well as the development and application of machine learning and deep learning algorithms to spatiotemporally embedded phenomena.