Atmosphere-Ocean

Taylor & Francis
Taylor & Francis Group

Weather Research and Forecasting (WRF) model and machine learning algorithms to improve marine fog predictions in areas offshore Atlantic Canada.

Journal:	*Atmosphere-Ocean*
Manuscript ID	Draft
Manuscript Type:	Applied Research / Recherche appliquée
Date Submitted by the Author:	n/a
Complete List of Authors:	Teeloku, Piyush; York University Faculty of Graduate Studies, Earth and Space Science and Engineering Chen, Zheqi; York University Taylor, Peter ; York University, Chen, Yongsheng; York University
Keywords:	WRF, NWP model, marine fog, machine learning, weather forecasting, North Atlantic, ERA5

Abstract:

Reduced visibility due to fog events poses serious challenges to transportation and public safety. Traditional numerical weather prediction (NWP) models face limitations in accurately forecasting fog due to complex microphysical processes. In this study, we propose a post-processing approach that combines the Weather Research and Forecasting (WRF) model forecasts with a machine learning (ML) classifier to distinguish between fog and no-fog conditions. Evaluated over an eleven-year period and tested with recent data, our method outperforms fog predictions based solely on liquid water content from the WRF model. According to the results obtained, this approach demonstrates the potential of ML techniques to enhance operational fog forecasting capabilities.

1. Introduction

The weather conditions for fog to occur can be quite mild, but when it happens, it can become quite disruptive due to the visibility drop, therefore aviation and marine industries are very interested in how fog will affect their operations (Flynn 2018). It is estimated to be the second most common cause of weather-related aviation accidents behind strong winds (National Center for Atmospheric Research, 2022). Beyond its impact on transportation, fog plays a crucial role in atmospheric and ecological processes (Koračin and Dorman, 2017). It carries water droplets containing ions, aerosols, and microorganisms, which influence coastal ecosystems' hydrological, thermodynamic, nutrient, and toxicological properties. In addition, fog regulates temperature and moisture conditions, shaping vegetation patterns and forest development in coastal regions.

There have been various Numerical Weather Prediction (NWP) approaches to modeling and predicting marine fog. Taylor et al. (2021) found that the amount of liquid water is too high in the Weather Research and Forecasting (WRF) model. They proposed a roughness length for fog droplets and allowed them to deposit to the surface through turbulence in the model. The modified model was run for North Atlantic during the summer 2018 on a domain extending from eastern Canada out beyond the Grand Banks and including Sable Island. They were able to have a better liquid water content (LWC) as well as visibility representation compared with observations. However, they did not guarantee that this modification improves the prediction of the onset and dissipation of fog. More recently, Gao et al. (2024) used another NWP approach

where they used microwave radiometer-retrieved cloud water path observations to improve numerical simulations of sea fog over the Yellow Sea in May 2018. They used an ensemble method with WRF and a Grid-point Statistical Interpolation/EnKF assimilation system. Then, they had more accurate initial conditions by retrieving the cloud water content from the satellite. By doing so, they were able to have improvements on sea fog forecasting. Their method worked very well with areas of high chance of fog but had an average false alarm rate of 0.463, meaning that they were overestimating the number of fog occurrences by roughly 46%. So currently NWP models are overestimating the LWC and doing a poor job at predicting fog occurrence.

Gultepe et al. (2024) applied machine learning techniques to nowcast marine fog visibility using observational data collected during the Fog and Turbulence Interactions in the Marine Atmosphere (FATIMA) campaign in July 2022. The campaign took place in the Grand Banks region and near Sable Island off the northeast coast of Canada, using measurements from the Research Vessel Atlantic Condor. Key meteorological variables were used alongside derived droplet number concentration to analyze fog characteristics. Nowcasting was performed for lead times of 30 and 60 minutes using support vector regression (SVR), least-squares boosting (LSB), and deep learning models, targeting visibility thresholds below 1 km and 10 km. Their results emphasize the value of ship-based observations and ML models in supporting short-term fog forecasting in complex marine environments.

Related work has demonstrated the effectiveness of machine learning for short-term fog and low-visibility forecasting. Castillo-Botón et al. (2022) investigated low-visibility event prediction using 23 months of data from the Mondoñedo weather station in Galicia, Spain. They applied both regression and classification approaches for 30-minute nowcasting of fog, demonstrating the utility of supervised learning techniques in both continuous and categorical forecasting contexts. Building on this, Peláez-Rodríguez et al. (2023) incorporated ERA5 reanalysis data alongside local station observations to improve fog-related visibility predictions at the same site. Their study explored both traditional ML models and evolutionary algorithms for short-term forecasts (1, 3, and 6 hours), reporting up to a 17% improvement in forecast accuracy. These findings support the value of integrating reanalysis data and advanced learning methods in visibility and fog prediction tasks.

Another sea fog ML approach was proposed by Chen et al. (2024) who developed a deep learning framework to improve the short-term prediction of sea fog regions in the Yellow Sea, with an emphasis on maritime navigational safety. The study introduced the Multivariable Sea Fog Forecast (MV-SFF) dataset, which contains 122 sea fog events from 2010 to 2020. Each event includes hourly geostationary satellite images and reanalysis-based meteorological variables, providing a rich data source for spatial fog forecasting. Using this dataset, the authors proposed the Rich-Element Aggregated (REA) model: a deep learning architecture designed to extract and integrate diverse meteorological and satellite-derived features across different times and spatial scales. The REA model includes a position-aware edge detection mechanism to accurately localize sea fog regions. A seven-hour ahead forecast starting from 09:16 local time demonstrated strong performance, with the REA model outperforming other state-of-the-art deep learning methods in both accuracy and consistency. This work highlights the potential of combining remote sensing with deep learning to enhance spatial sea fog forecasting capabilities.

In this study, we decided to take an alternative approach by using both WRF and machine learning to improve the hourly forecast of fog. Binary classification was chosen as the machine learning framework to post-process numerical output from the WRF model at St. John's Newfoundland and Labrador, and Yarmouth, Nova Scotia, due to its proven effectiveness in prior research. We used data from 2012 to 2023 since 2012 is the earliest time with available visibility reports at the weather stations and used only summer data from April to August for each year. This avoids complications with blowing snow/ice and we are looking mainly at advection fog in the coastal areas where the highest frequency of fog at these regions is during the summer.

The binary approach was used to identify fog events (typically defined by visibility below 1 km) while accounting for the inherent variability in visibility measurements. From other research, we observed that using ERA5 data for the meteorological features helped increase the performance of the ML approaches and compared to other research that used features to complement their ML models, we used a post-processing approach. Many existing studies have employed classification techniques to distinguish between fog and no-fog conditions and have reported reliable performance in identifying fog onset and duration. Classification is also particularly advantageous when the primary operational need is to issue timely alerts for fog events, rather than estimating precise visibility values. Our approach builds on this by integrating WRF model

forecasts with site-specific visibility observations to improve local fog prediction. Unlike many previous works that focus on very short-term forecasts (e.g., 30 to 60 minutes or up to 6 hour lead time), our method extends the forecast horizon to 24 hours and provides predictions at an hourly resolution, offering greater lead time for decision-making. This also enables us to identify when fog occurs during the day and potentially how long it can last compared to using only time windows of 3 hours, 6 hours and 12 hours. Additionally, by using a high resolution NWP model like WRF, our method benefits from better meteorological context, particularly in capturing key physical drivers of fog. While this approach does not resolve the full areal extent of fog, the combination of WRF-based physical guidance and site-specific ML classification offers an optimal balance between accuracy, interpretability, and operational utility for localized fog forecasting.

2. Data collection

2.1. Features

We chose features that are relevant to marine fog occurrence. Gultepe et al. (2024) tested ten different variables and performed a pairwise correlation analysis that was visualized on a heat map. So, we did the same process with a Pearson correlation analysis for 2 m temperature, 2 m relative humidity, 10 m U and V wind, land surface pressure, hour, day of year and month. Values of the features are from ECMWF's Reanalysis v5 (ERA5, Hersbach et al., 2023) dataset. ERA5 combines past observations with output from the ECMWF model to provide a consistent historical record of atmospheric, oceanic, and land variables from 1940 to the present. It has a resolution of 31 km and 137 vertical levels and is available every hour. For convenience, values are extracted using the WRF Preprocessing System (WPS) which is used for preparing input files for real-case WRF simulations. ERA5 does not contain 2 m relative humidity. Thus, it was calculated using an NCAR Command Language (NCL) function relhum_ttd with temperature and dew point from ERA5. Fig. 1 shows the domain settings used and the locations of the weather stations. The same domain setting is also used for running WRF for prediction.

Fig. 2 shows the heatmap of feature correlations with visibility for St. John's and Yarmouth, revealing distinct patterns between the two locations. According to Ratner (2009), for a feature to be relevant, the correlation should be greater than 0.3. This indicates a moderately positive or

negative linear relationship according to the fuzzy-firm linear rule. It is worth mentioning that the heat map will not capture any other types of relationship and so, some features may have a higher correlation than shown but we used the Pearson Correlation Coefficient (PCC) as a benchmark to proceed with the feature selection.

In Yarmouth, most features show moderate to strong correlation with visibility, particularly relative humidity, while the U component at 10 m is weakly correlated. In St. John's, correlations with temperature variables are generally weaker, though U and V wind components show moderate association. Relative humidity remains strongly correlated at both sites. Since we are dealing with time series data, time-based features are valuable to have and so we parsed the Time data into 2 variables: month and hour. By including these features, we were able to analyze the data based on different time components, such as looking at trends over specific hours of the day or different days. These features are useful in training ML models that need to capture time-based patterns, especially weather patterns that can last for hours, like fog.

From the heatmap in Fig. 2, we do not see much correlation with the visibility and so decided to look at the trends visually. From Fig. 3 and Fig.4, we see that the monthly trends for both locations vary. St. John's has the highest monthly occurrence in April to June, while Yarmouth has a higher occurrence in the later months, July and August. From this graph, we were able to identify that the month feature is important to have. Fig. 5 and Fig.6 show the line plots for fog occurrence count in a day for both locations. From both plots, we see the diurnal cycle influencing fog occurrence with daytime heating burning off most of the fog in the early mornings. The highest fog occurrence at both locations is around nighttime and very early mornings (till around 7 a.m for both local time). Therefore, the monthly fog occurrence and the hour at which fog happens are valuable to learn the behavior of fog, and by having those time-based features, the ML model can identify any useful traits to make better predictions.

Another important aspect is to use lagged features. Although in our models data at each hour is treated separately, lagged features are useful since they can capture historical dependencies and trends of the continuous meteorological conditions. Lagging a time series feature means shifting the current value forward in time by a desired amount of steps. For each of the

meteorological variables, its value one hour before the predicted time (lag 1) is also used as a feature. It is possible since our goal is to use the 24-hour forecast of the variables from WRF, the lag 1 variables also come from WRF forecasts. To summarize, the features include 5 meteorological variables at the forecast time (2 m temperature, 2 m relative humidity, 10 m U and V winds, and land surface pressure), 5 meteorological variables 1 hour before the forecast time and 2 variables of time in UTC (month and hour).

2.2. Target

Visibility data we used are at St. John's International Airport (47º37'07" N, 52º45'09" W with an elevation of 140.5 m) and at Yarmouth Airport (43º49'37"N, 66º05'17" W with an elevation of 42.9 m), collected by NAV CANADA (NAVCAN, 2017). These sites are influenced by marine air masses and complex coastal terrain, both of which are key contributors to fog formation. Both St. John's and Yarmouth are situated along the Atlantic coast of Canada, where the interaction of major ocean currents, the warm Gulf Stream and the cold Labrador Current, creates ideal conditions for sea advection fog. In this region, warm, humid air masses transported by the Gulf Stream often move over the colder waters carried southward by the Labrador Current (Bullock et al., 2016). This temperature contrast is particularly pronounced during the spring and early summer months, when sea surface temperatures remain low while atmospheric temperatures begin to rise. As a result, dense fog frequently develops offshore and is advected inland by prevailing winds, especially under stable marine layer conditions. The convergence of these ocean currents is therefore a critical driver of marine fog formation and plays a central role in the fog climatology of both coastal sites.

As shown in Fig. 7 and Fig. 8, NAVCAN has never reported a visibility of 1.4 km in 12 years (44064 hours of data), so it could be a clear border to distinguish between fog and clear. The visibility measurements are made by human observers looking at objects at a distance, which can be different for daytime and nighttime. At nighttime, background light will be an important factor. However, only fog or clear is of concern in this study and human observers should be reliable for distinguishing between the two. Therefore, it is decided that our "fog" class is when visibility is <= 1.2 km and "clear" class is when visibility is > 1.2 km, instead of the formal definition of 1 km. Tests using 1 km to distinguish between fog and clear have been done and they were less accurate. According to the reports, rain alone does not reduce the reported

visibility to below 1.2 km, and there are a lot of reports of rain combined with fog. Therefore, this study does not exclude any rain conditions.

One important aspect of classification is figuring out how the categorical labels are distributed. Google For Developers (2024) describes two types of datasets, balanced and imbalanced. For binary classification, a balanced dataset will have both labels nearly the same amount, while an imbalanced dataset will have one label that appears much more than the other. For both locations, the percentage of fog data is around 10-12% so it is a moderate case for imbalanced data. This enables us to understand our visibility data better and allows us to take measures such as using oversampling and undersampling techniques or balancing the class weights of our labels when preparing our data to run our model properly.

Before training the machine learning model, we first considered whether to build a single model for both locations or treat them separately. Given the geographic and environmental differences, with Yarmouth's lower elevation (42.9 m) compared to St. John's (140.5 m), we anticipated that oceanic and atmospheric influences could vary between the two sites. Heatmap analyses, along with observed differences in monthly and hourly weather patterns, supported this assumption. As a result, we chose to model the two locations separately, allowing each model to better capture the distinct weather dynamics of its respective region.

2.3. ML model

ETClassifier

Extremely Randomized Trees (ExtraTrees) is an ensemble method proposed by Geurts Pierre (2006) that constructs multiple decision or regression trees, using more randomness in its technique to improve robustness and computational efficiency. ExtraTrees combine predictions from multiple unpruned (not optimized) trees for classification or regression but introduces additional randomness in the splitting process. In ExtraTrees, both the feature and the cut-point for splitting at each node are chosen at random. Even if the split threshold is chosen at random, the ExtraTrees will still try to get the best score. The ExtraTrees model can be used for classification (ETClassifier). To optimize the ETClassifier model, we will use the RandomizedSearchCV function from Scikit-Learn (Pedregosa et al., 2011) with a TimeSeriesSplit cross-validation strategy tailored for time-series data. This approach ensures that the hyperparameter tuning respects the temporal order of the data, preventing data leakage

from future observations that influence the training process. This comprehensive randomized search, combined with cross validation based on time series, will allow us to identify the best combination of hyperparameters for the ETClassifier. The best hyperparameters for each location are given in Table 1.

LSTM

Hochreiter and Schmidhuber (1997) introduced the Long Short-Term Memory (LSTM) network, a type of Recurrent Neural Network (RNN) that uses gates (forget , input and output gates) to control information flow and manage long-term dependencies. For our bidirectional LSTM (biLSTM) model, input data is reshaped to [batch size, time steps, features], where time step is 1 and features total 14 (including lagged features). The model begins with an input layer followed by two BiLSTM layers with 64 and 32 units, respectively, both using tanh activation. Batch Normalization (Ioffe and Szegedy, 2015) and Dropout (0.2; Srivastava et al., 2014) are applied after each LSTM and dense layer to improve stability and reduce overfitting. Two dense hidden layers follow, using LeakyReLU activation to maintain gradient flow. The final output layer, from Singh et al. (2023), is a single sigmoid-activated neuron for binary classification. The model has 115,861 parameters (38,513 trainable), well-matched to our dataset size of 33,000 samples.

XGBoost

XGBoost (eXtreme Gradient Boosting, Chen and Guestrin, 2016) is a model based on decision trees. It uses gradient boosting where in each iteration a new tree is constructed to fix the errors of the previous tree (Hsieh 2023). Logistic regression is chosen as the loss function. According to the recommended calculation (hours of fog divided by hours of clear), scale_pos_weight is 7.97 for St. John’s and 5.75 for Yarmouth. Other hyperparameters are tuned by cross validation and are summarized in Table 2.

Therefore, we selected ExtraTrees as a baseline due to its ensemble of decision trees, which aggregates multiple weak learners to improve predictive performance. XGBoost, a related model, was also included for comparison as it employs gradient boosting to iteratively build a strong learner. In addition, we explored neural networks for fog classification to take advantage of their strength in capturing nonlinear relationships in the data. Specifically, we implemented a

bi-directional LSTM (biLSTM) model to account for long-term temporal dependencies inherent in time series weather data. This allowed us to compare traditional tree-based ensemble methods with deep learning approaches and identify the most effective model for fog prediction.

3. Result

To summarize, the dataset included data from March to August, 2012 to 2023, in which the features from ERA5 were downscaled using WPS, and the visibility reports were from the weather stations. Two locations are treated separately and have their own datasets. The dataset was split into three sets: training, validation and test sets, which were 64%, 16%, 20% of the dataset, respectively. We ensured with the training and validation sets that both models were generalizing well and not overfitting. Since we split the data considering the nature of the time component, the test set was approximately the year 2023. This means that we tested our trained models with the 2023 dataset for both locations and saw how they performed.

We used the recall and precision metrics to evaluate the test dataset. The recall looks at the proportion of true positives (TP) that are correctly classified by the model:

$$Recall = \frac{TP}{TP+FN}$$

For fog occurrence, recall is how many fog classes the model got correctly out of the total number of fog cases. False negative (FN) , in this case, is the amount of fog our model falsely labeled as ”clear”. False positive (FP) is the amount of clear incorrectly labeled as “fog” by our model and true negative (TN) is the amount of clear we got correctly. The next important metric is precision given by:

$$Precision = \frac{TP}{TP+FP}$$

Precision is the proportion of actual fog that our model got accurately from all the fog predictions it did. Using both metrics helps us to understand whether our model is just predicting a lot of fog and bumping up the performance or if it can differentiate between clear and fog well. We also used the F1-score, which is the harmonic mean of precision and recall, and offers a more balanced evaluation with one score. It ensures that both recall and precision are considered

together, making it particularly useful when dealing with imbalanced datasets where optimizing for just one metric can lead to misleading conclusions. The equation is given by:

$$F1\ score\ = 2\ \frac{precision \cdot recall}{precision + recall}\ =\ \frac{2TP}{2TP + FN + FP}$$

After training both models with the dataset collected at each location, we verified that it was generalizing well with the validation set. The next step was to see how our model did with the test data. Table 3 shows all the results we obtained from the three models, ExtraTreesClassifier, XGBoost and biLSTM, using a classification report showing the precision, recall, and F1-score for both clear and fog classes for St John's and Yarmouth for the 2023 dataset.

Results across all models are similar with F1-score for fog prediction being around 0.70 for both models. ETClassifier slightly outperformed the other models. This may be due to ETClassifier having a higher tolerance to noisy data such as weather data and it is using an ensemble of trees to reduce the variance in prediction. Therefore, for our case of fog prediction, the ETClassifier appears to be more suitable. ETClassifier is also a decision tree based model which can be more suitable for solving weather tasks which often relies on the condition or threshold of a meteorological feature. Now that we know that our ML models can predict fog, the next step was to see how they perform for forecasting.

There are many challenges in ML forecasting. Forecasting needs a lot of time series data availability that may contain missing values, have irregular sampling, or have abrupt changes due to external factors. Additionally, preparing the data is more complex due to more advanced feature engineering such as creating lag variables, rolling averages, or trend components. Therefore, instead of using the ML model for a time series forecast, we decided to approach the issue differently. The features we used for our ML models are the main meteorological variables and WRF does a good job of forecasting them, especially in summer with the largest error for wind components (Bughici et al. (2019) and Pappa et al. (2023)). We decided to use the forecast values of the features from WRF and then use our ML models to predict fog occurrences from them. To simulate a forecast environment, instead of analysis data, we used the forecast data from the Global Forecast System (GFS), which is available every three hours at a 0.25-degree resolution (National Centers for Environmental Prediction, 2015). WRF was run pseudo-operationally 36 hours long for every day in the study period, April to August in 2024, with identical physical parameterization settings. Meteorological variables in the last 24

hours of each run were extracted, with the first 12 hours used as a spin-up time for WRF. These forecasted variables were then processed using the same data preparation pipeline and fed into our trained ML models to evaluate their performance on new, unseen data. This approach allowed us to assess whether the models could be used operationally to forecast fog on an hourly basis for the next 24 hours.

The next crucial step is to compare their performance with other forecasting methods, specifically the WRF model, which is commonly used for operational marine fog prediction based on surface liquid water content. For this comparison, the cloud water mixing ratio at the lowest model level at approximately 3.6 meters was also extracted from the WRF runs mentioned above. To classify fog in a binary manner, we labeled any instance where the cloud mixing ratio was nonzero (indicating the presence of liquid water) as fog, and instances where it was zero as clear. This approach aligns with WRF's tendency to predict high liquid water content which always drops the visibility level in the fog zone. Table 4 shows the results for fog classification on the 2024 dataset for the post-processing ML models and the WRF model.

Below shows the confusion matrices for all the models at both locations in 2024 to have a better understanding of how many fog labels we were able to capture. The confusion matrix for each location will indicate the amount of clear we got correctly (TN) and the amount we mislabeled as fog (FP). It will also indicate the amount of fog we got correctly (TP) and the amount we mislabeled as clear (FN) and will be given in this format, $\begin{pmatrix} TN & FP \\ FN & TP \end{pmatrix}$. The first row is the total number of clear cases (TN + FP) and the second row is the total number of fog cases (FN + TP).

The confusion matrix for each model at St John's is given below:

$$ETClassifier = \begin{pmatrix} 2890 & 198 \\ 170 & 413 \end{pmatrix}$$

$$XGBoost = \begin{pmatrix} 2843 & 246 \\ 179 & 404 \end{pmatrix}$$

$$biLSTM = \begin{pmatrix} 2906 & 182 \\ 187 & 396 \end{pmatrix}$$

$$WRF\ LWC\ = \begin{pmatrix} 2864 & 225 \\ 223 & 360 \end{pmatrix}$$

The confusion matrix for each model at Yarmouth is given below:

$$ETClassifier = \begin{pmatrix} 2760 & 318 \\ 214 & 379 \end{pmatrix}$$

$$XGBoost = \begin{pmatrix} 2625 & 401 \\ 218 & 375 \end{pmatrix}$$

$$biLSTM = \begin{pmatrix} 2750 & 328 \\ 236 & 357 \end{pmatrix}$$

$$WRF\ LWC\ = \begin{pmatrix} 2669 & 410 \\ 236 & 357 \end{pmatrix}$$

For St. John's, ExtraTreesClassifier outperforms the WRF model across all classification metrics, achieving higher precision (0.68 vs. 0.62), recall (0.71 vs. 0.62), and F1-score (0.69 vs. 0.62), indicating a more balanced and effective detection of fog events. A similar performance advantage is seen at Yarmouth, where ETClassifier shows improvements in precision (0.54 vs. 0.47), recall (0.64 vs. 0.60), and F1-score (0.60 vs. 0.53). Higher recall values suggest that ETClassifier is particularly effective at reducing missed fog cases, which is critical for operational forecasting. As shown in matrices above, ETClassifier identified more fog events than WRF (413 vs. 360 for St. John's, and 379 vs. 357 for Yarmouth) while also producing fewer false positives and false negatives. WRF overestimated fog more frequently (222 vs. 198 false positives for St. John's; 410 vs. 318 for Yarmouth) and missed more actual fog cases than ETClassifier. These consistent improvements demonstrate the effectiveness of integrating machine learning with WRF for post-processing coastal fog forecasts. Our post-processing approach not only improves hourly fog prediction performance over a 24-hour horizon but also compares favorably with other ML-based classification studies, despite the differences in datasets and lead times. This suggests that tree-based ensemble methods like ETClassifier provide a reliable and efficient solution for operational fog forecasting when used alongside traditional numerical models.

To better understand the performance of our 24-hour fog forecast at an hourly resolution, we evaluated model skill over time by dividing the forecast horizon into four 6-hour bins. In many machine learning approaches, model performance tends to degrade as the forecast lead time increases due to accumulated uncertainty. However, in our case, we believed that the use of post-processing techniques would help stabilize accuracy across the forecast period. While we acknowledge that the underlying WRF model can introduce increasing error over time, we aimed to determine whether our post-processing could mitigate this effect.

Fig. 9 and Fig. 10 present the F1-score for fog detection in 6-hour bins for both St. John's and Yarmouth, based on forecasts from the 2024 dataset. In St. John's, the F1-scores remain relatively stable across all bins, with the final 18-23h bin surprisingly showing the highest performance. We have the number of fog events occurring at each bins on top which is showing for summer 2024. The first 12 hours had the highest fog counts compared to the last 12 hours. Since the times are in UTC, the first 12 hours represent late night to early mornings (21:30 to 09:30) which aligns with usual fog events. In Yarmouth, performance is slightly lower overall but also shows minimal variation across time bins, with no clear trend of degradation. The first 12 hours also represent nighttime to early mornings (21:00 to 09:00) and the contrast for fog events is more pronounced with only 71 fog events in the last 6-hour bin. This could be the reason why it has a slightly lower F1-score.

These results support our assumption that the post-processing approach helps maintain consistent fog classification skill across the 24-hour forecast window. Nevertheless, we note that the WRF model's intrinsic error growth with lead time was not explicitly assessed here, and future work should investigate the upper limits of this post-processing approach for longer forecast horizons.

4. Discussion

While a slight decrease in model performance is expected when evaluating on new, unseen data, we observed a more substantial decline in overall performance at Yarmouth compared to St. John's when tested on the 2024 dataset. This discrepancy may be attributed to differences in data quality between the two locations, particularly the forecasted input features derived from

WRF. For St John's, the F1-score was 0.73 on the 2023 dataset and 0.69 on the 2024 dataset using the ETClassifier model. The WRF LWC model had an F1-score of 0.62 at St John's. For the 2023 dataset, at Yarmouth, we got an F1-score of 0.70 but a steeper drop in performance for the 2024 dataset with an F1-score of 0.60.

We were able to improve on the WRF LWC model which had an F1-score of 0.53 but this suggests that the input data fed into the machine learning model may have already carried inherent limitations. If the forecasted features, such as temperature, relative humidity, or wind components, are poorly represented or less accurate in Yarmouth due to model bias, resolution, or local meteorological complexities, this could significantly hinder the model's ability to generalize. Therefore, the observed drop in ML model performance may not reflect a shortcoming of the classification model itself but rather a degradation in input feature quality, which propagates through the prediction pipeline.

Furthermore, we were able to see an improvement by using the ML models to predict fog instead of using traditional NWP models such as WRF in the section above. To make sure we got the best possible trained ML model, we had to ensure that we prepared the data correctly, optimized the model carefully, and tested it on different types of data. Besides, we implemented specific strategies to refine our approach, ensuring that our model was well-suited for the task and capable of delivering the most accurate fog predictions.

We evaluated which input data to use for our ML models between ERA5 reanalysis and GFS Final Analysis data. We used ETClassifier since it was the better performing model. We trained and did a comparison on the 2023 dataset. Since we were only able to collect data from 2016 to 2023 from GFS Final Analysis, compared to 2012 for ERA5, the performance when using features from the ERA5 data was better on all metrics. We also evaluated the importance of our lagged features, which showed slightly better or similar performance in terms of precision and F1-score but had higher recall at each location. Although the overall improvement was not drastic, incorporating a one-hour lag to the features helped the model to distinguish more accurately between fog and clear cases, which is crucial for operational forecasting.

One can argue that by having more data to train, the ML model may learn more about the patterns to classify fog or will generalize better. Therefore, we tested the case that combined both locations together and trained a single ETClassifier ML model. Data was prepared in the

same way as before: splitting the data while maintaining the time series, normalizing the features, using lagged features. After concatenating the datasets from both locations, we used OneHotEncoder from Scikit-Learn to transform the names of the locations into a binary format without any ordinality. After training the model, we tested it with the 2023 dataset for both St. John's and Yarmouth. When compared to when we trained the locations individually, we obtained an overall lower F1-score at each location.Therefore, we were able to confidently say that choosing a model for each location was the correct approach to get the best results as the model was able to learn the weather patterns of that location better, thus learning and generalizing better.

We have also tested the case of multi-classification. Between clear and fog (≤ 1.2 km), we added one more label of mist with visibility from 1.2 km to 5 km. A key advantage of multi-class classification is that if the model predicts mist instead of fog, it still serves as an early warning, bringing us closer to identifying fog conditions. Additionally, having mist as a separate category provides more nuanced predictions and better reflects real-world conditions. We prepared the data exactly as before and used RandomSearchCV to get the best hyperparameters for this task. However, since mist was severely under-represented in our dataset, our model could not predict "mist" classes at all and performed worse.

Regression is another supervised ML technique that models the relationship between a dependent variable (target) and one or more independent variables (features), similar to classification, but the model predicts continuous values instead of discrete labels (Sagar 2025). It is possible for forecasting visibility since we have visibility reports that range from 0 km to around 24 km. There are various types of regression, with the simplest being linear regression (Altman and Krzywinski, 2015), which assumes a linear relationship between the input variables and the target. We have tried two models, the ExtraTrees Regressor (ETRegressor) and Quantile Regression with Random Forests. We initially used ETRegressor as the ExtraTrees was the better performing model for classification but had to use a more robust model to get better results.

To evaluate the regression approach, we employed several standard metrics: Mean Absolute Error (MAE), which quantifies the average magnitude of errors between predicted and observed values; Mean Squared Error (MSE), which penalizes larger errors by averaging the squared differences; and the coefficient of determination (R^2), which measures how well the predicted

values align with the actual data. After training the model, we evaluated its performance on the 2023 test dataset for St. John's. While the model was occasionally able to capture low-visibility events (below 1.2 km), it frequently failed to detect sudden drops in visibility associated with fog formation or rapid increases during fog dissipation. These missed transitions highlighted the limitations of regression in identifying abrupt changes, which are critical for fog forecasting. Although the overall regression performance was consistent with previous studies on marine fog visibility, it became evident that binary classification was more effective at capturing the onset and dissipation of fog events.

5. Conclusion

This study explored the challenges of marine fog prediction by integrating physical and data-driven approaches. Initially, we used the WRF model to simulate marine fog for the Fog and Turbulence Interactions in the Marine Atmosphere (FATIMA) project. Recognizing the limitations of physical models in fog forecasting, we implemented a machine learning (ML) approach using binary classification to predict fog occurrence in coastal regions of Canada (St. John's and Yarmouth). We trained the models using ERA5 and labeled fog events using NAV Canada visibility reports (visibility ≤ 1.2 km). Among the models tested, ExtraTreesClassifier consistently outperformed WRF-based predictions, achieving an 11% improvement in fog classification performance based on F1-score. The binary classification approach proved most effective, aligning with previous studies such as Lam et al. (2023). Additionally, this approach enabled us to do 24-hour forecasts at hourly resolution which is an advancement over typical 1-7 hour lead times reported in earlier works.

Improvements could be made by incorporating alternative forecast datasets. Using ECMWF's Integrated Forecast System (IFS) or processing forecasted GFS data directly through WPS may enhance 24-hour prediction quality. While ERA5 and GFS perform well for surface variables, discrepancies with station observations (Velikou et al., 2022, Haiden et al., 2021) suggest that increasing observational input for training could further improve accuracy. Standalone AI forecasting systems also present a promising avenue. ECMWF's Artificial Intelligence Forecasting System (AIFS), operational since February 2025, provides high-resolution deterministic forecasts and could be used to drive post-processed ML fog predictions. Our

models, when coupled with such advanced inputs, may yield higher accuracy and faster outputs.

Additionally, nowcasting presents a promising opportunity to enhance fog prediction. This approach involves using lagged visibility observations as input features, enabling the model to learn from recent visibility trends and make more accurate short-term forecasts. Though not the focus of this study, preliminary results using visibility as a feature for short-term forecasts (1-6 hours) showed substantial gains, with an F1-score of around 0.85 for 1-hour lead times. Expanding on this approach could be beneficial for real-time applications. Finally, the broader context of climate change must be considered. Recent studies (e.g., Johnstone and Dawson, 2010, Taha and Abduljabbar, 2024) indicate a declining trend in marine fog frequency linked to rising sea surface temperatures and weakened air-sea temperature contrasts. For regions like St. John's and Yarmouth, shifts in wind patterns, ocean currents (e.g., the Labrador Current), and aerosol concentrations could further impact fog occurrence. These evolving dynamics emphasize the need for adaptive, ML-enhanced forecasting systems capable of responding to changing environmental conditions.

Disclaimer: The authors used Grammarly, an AI-assisted writing tool, to help improve grammar, clarity, and language flow during the manuscript preparation. No AI tools were used for content generation, data analysis, or interpretation of results.

References

Altman, N. and M. Krzywinski (2015). Simple linear regression. https://doi.org/10.1038/nmeth.3627.

Bughici, T., N. Lazarovitch, E. Fredj, and E. Tas (2019). Evaluation and bias correction in wrf model forecasting of precipitation and potential evapotranspiration. https://journals.ametsoc.org/view/journals/hydr/20/5/jhm-d-18-0160_1.xml. DOI: castill10.1175/JHM-D-18-0160.1.

Bullock, T., G. A. Isaac, J. Beale, and T. Hauser (2016, 10). Improvement of visibility and severe sea state forecasting on the grand banks of Newfoundland and Labrador. https://doi.org/10.4043/27406-MS.

Castillo-Botn, C., D. Casillas-Prez, C. Casanova-Mateo, S. Ghimire, E. CerroPrada, P. Gutierrez, R. Deo, and S. Salcedo-Sanz (2022). Machine learning regression and classification methods for fog events prediction. https://www.sciencedirect.com/science/article/pii/S0169809522001430. DOI: https://doi.org/10.1016/j.atmosres.2022.106157.

Chen, K., Y. Zhou, T. Ren, and X. Li (2024). Short-term sea fog area forecast: A new data set and deep learning approach. https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2024JH000230. DOI: https://doi.org/10.1029/2024JH000230.

Chen, T., Guestrin, Carlos (2016). XGBoost: A Scalable Tree Boosting System. Association for Computing Machinery. http://dx.doi.org/10.1145/2939672.2939785. DOI: 10.1145/2939672.2939785.

Flynn, C. (2018). The physics of fog. https://blog.metservice.com/Physics-of-Fog. [Online; accessed 25-December-2024].

Gao, X., X. Bao, S. Ma, Q. Chen, and B. Wang (2024). An online assimilation method to improve the numerical forecast of sea fog using microwave radiometer-retrieved cloud water path. https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2023JD040229. DOI: https://doi.org/10.1029/2023JD040229.

Google For Developers (2024). Datasets: Imbalanced datasets. https://developers.google.com/machine-learning/crash-course/overfitting/imbalanced-datasets. Accessed: 2024-10-07.

Geurts Pierre, Ernst Damien, W. L. (2006). Extremely randomized trees. https://doi.org/10.1007/s10994-006-6226-1. DOI: 10.1007/s10994-006-6226-1.

Gultepe, E., S. Wang, B. Blomquist, H. J. S. Fernando, O. P. Kreidl, D. J. Delene, and I. Gultepe (2024). Generative nowcasting of marine fog visibility in the grand banks area and sable island in canada. https://arxiv.org/abs/2402.06800.

Haiden, T., M. Janousek, F. Vitart, Z. Ben-Bouallegue, L. Ferranti, and F. Prates (2021, 09/2021). Evaluation of ECMWF forecasts, including the 2021 upgrade. https://www.ecmwf.int/node/20142. DOI: 10.21957/90pgicjk4.

Hersbach H, Bell B, Berrisford P, et al. The ERA5 global reanalysis. Q J R Meteorol Soc. 2020; 146: 1999–2049. https://doi.org/10.1002/qj.3803

Hsieh, W. W. (2023). Introduction to Environmental Data Science. Cambridge University Press

Hochreiter, S. and J. Schmidhuber (1997, November). Long short-term memory. https://doi.org/10.1162/neco.1997.9.8.1735. DOI: 10.1162/neco.1997.9.8.1735.

Ioffe, S. and C. Szegedy (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. http://arxiv.org/abs/1502.03167.

Johnstone, J. A. and T. E. Dawson (2010). Climatic context and ecological implications of summer fog decline in the coast redwood region. https://www.pnas.org/doi/abs/10.1073/pnas.0915062107. DOI: 10.1073/pnas.0915062107.

Koračin, D., C. E. Dorman, J. M. Lewis, J. G. Hudson, E. M. Wilcox, and A. Torregrosa (2014). Marine fog: A review. https://www.sciencedirect.com/science/article/ pii/S016980951300361X. DOI: https://doi.org/10.1016/j.atmosres.2013.12.012.

Lam, R., A. Sanchez-Gonzalez, M. Willson, P. W. Wirnsberger, M. Fortunato, F. Alet, S. Ravuri, T. Ewalds, Z. Eaton-Rosen, W. Hu, A. Merose, S. Hoyer, G. Holland, O. Vinyals, J. Stott, A. Pritzel, S. Mohamed, and P. Battaglia (2023). Graphcast: Learning skillful medium-range global weather forecasting. https://arxiv.org/abs/2212.12794.

National Center for Atmospheric Research (2022). Fog forecasting to avoid delays and accidents. https://ral.ucar.edu/pressroom/features/fog-forecasting-to-avoiddelays-accidents. [Online; accessed 25-December-2024].

National Centers for Environmental Prediction, National Weather Service, NOAA, U.S. Department of Commerce (2015a). Ncep gdas/fnl 0.25 degree global tropospheric analyses and forecast grids. https://rda.ucar.edu/datasets/dsd083003/.

NavCanada (2017). Aviation weather services guide for aviation users. https://www.navcanada.ca/en/aviation-weather-services-guide.pdf.

Pappa, A., E. Siouti, S. N. Pandis, and I. Kioutsioukis (2023). High-resolution wrf forecasts in the smartaq system: Evaluation of the meteorological forcing used for pmcamx predictions in an urban area. https://www.sciencedirect.com/science/article/pii/ S0169809523004386. DOI: https://doi.org/10.1016/j.atmosres.2023.107041.

Pedregosa, F. et al., Scikit-learn: Machine Learning in Python, JMLR 12, pp. 2825-2830, 2011.

Peláez-Rodrguez, C., J. Prez-Aracil, C. Casanova-Mateo, and S. Salcedo-Sanz (2023). Efficient prediction of fog-related low-visibility events with machine learning and evolutionary algorithms. Atmospheric Research 295, 106991.

Ratner, B. (2009). The correlation coefficient: Its values range between +1/1, or do they? Journal of Targeting, Measurement and Analysis for Marketing. DOI: https://doi.org/10.1057/jt.2009.5.

Sagar, S. (2025). Regression in machine learning. https://www.geeksforgeeks.org/regression-in-machine-learning/. [Online; accessed 19-February-2025]

Singh, Y., M. Saini, and Savita (2023). Impact and performance analysis of various activation functions for classification problems. DOI: 10.1109/InC457730.2023.10263129.

Srivastava, N., G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov (2014). Dropout: A simple way to prevent neural networks from overfitting. http://jmlr.org/papers/v15/srivastava14a.html.

Taha, N. W. and H. S. Abduljabbar (2024). Coastal fog, climate change, and the environment. https://www.e3s-conferences.org/articles/e3sconf/abs/2024/113/e3sconf_itese2024_02014/e3sconf_itese2024_02014.html. DOI: https://doi.org/10.1051/e3sconf/202458302014.

Taylor, P. A., Z. Chen, L. Cheng, S. Afsharian, W. Weng, G. A. Isaac, T. W. Bullock, and Y. Chen (2021). Surface deposition of marine fog and its treatment in the weather research and forecasting (wrf) model. https://acp.copernicus.org/articles/21/14687/2021/. DOI: 10.5194/acp-21-14687-2021.

Velikou, K., G. Lazoglou, K. Tolika, and C. Anagnostopoulou (2022). Reliability of the era5 in replicating mean and extreme temperatures across Europe. https://www.mdpi.com/2073-4441/14/4/543. DOI: 10.3390/w14040543.

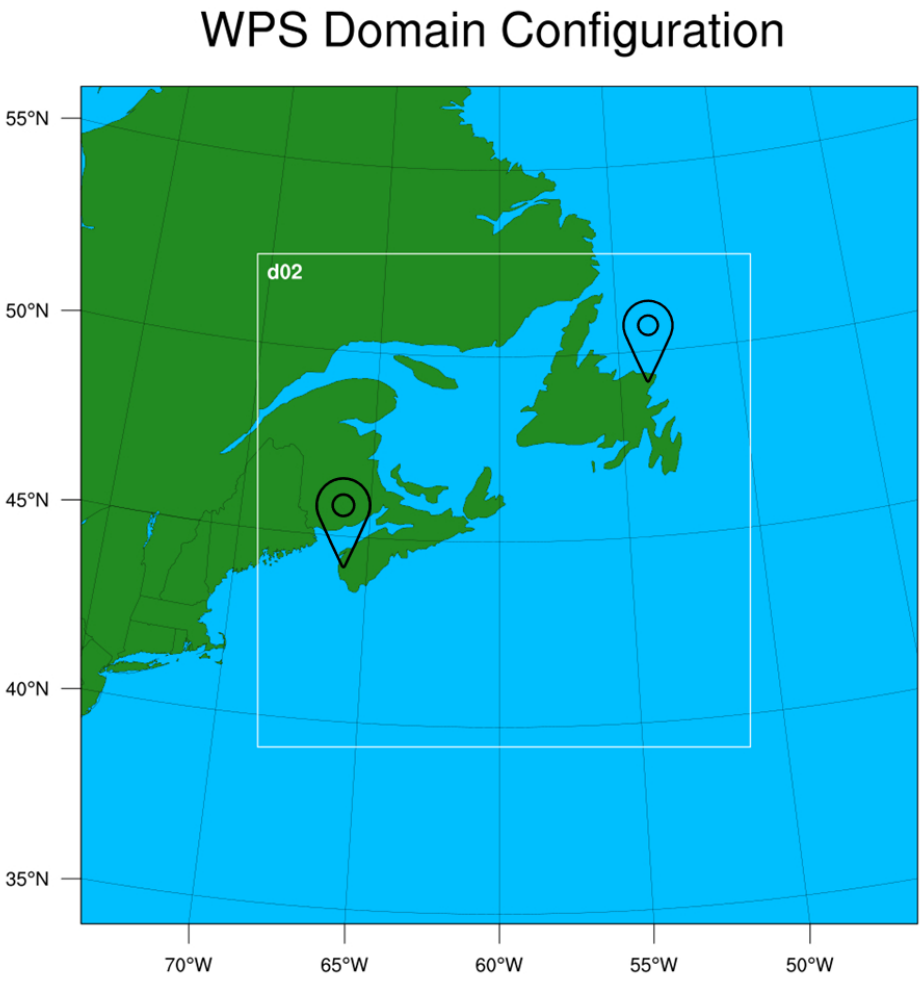

Figure 1: Two-nested WRF domain configuration with 27 km and 9 km resolutions with the locations of the weather stations pinned. The pin in the east is St. John's and the pin in the west is Yarmouth.

414x414mm (59 x 59 DPI)

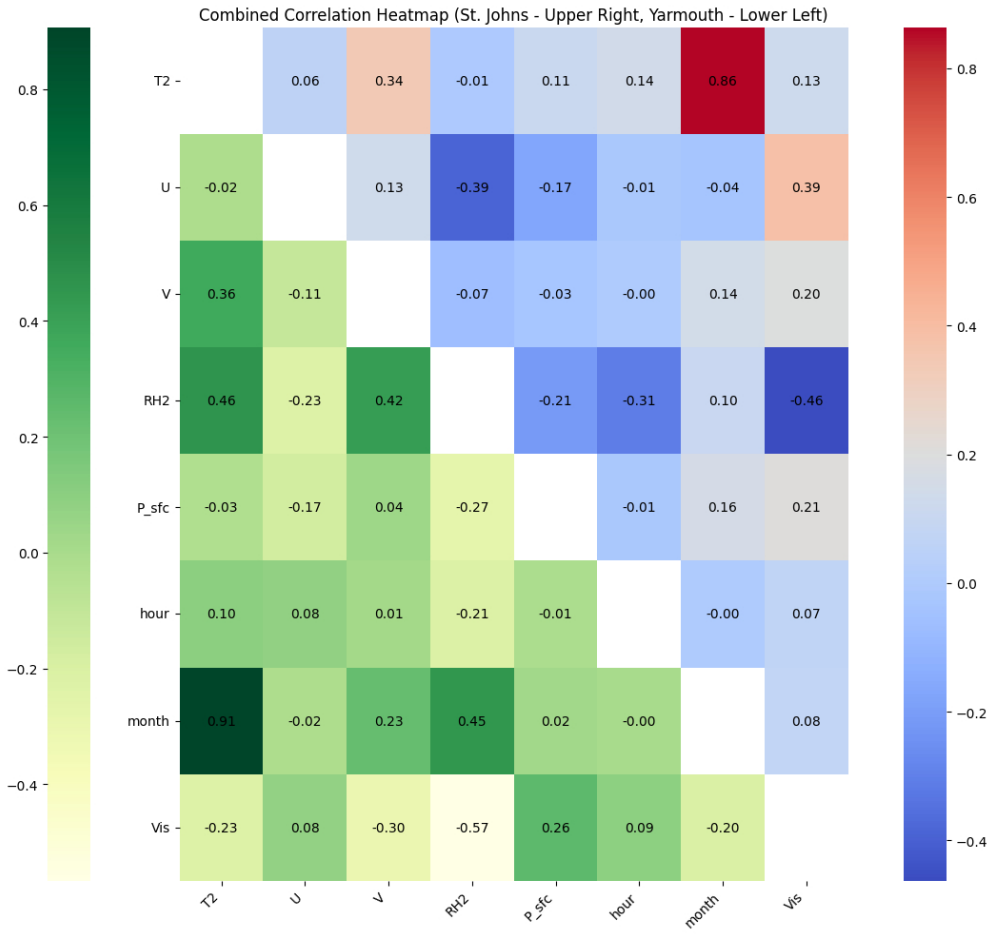

Figure 2: Correlation heatmap of the features relevant for fog prediction using the ERA5 data from 2012 to 2023 for St John's, Newfoundland and Labrador on the upper right in cool warm colors and Yarmouth, Nova Scotia on the lower left in green colors. Values in the diagonal are hidden because the correlation of a variable with itself is always equal to 1.

683x644mm (39 x 39 DPI)

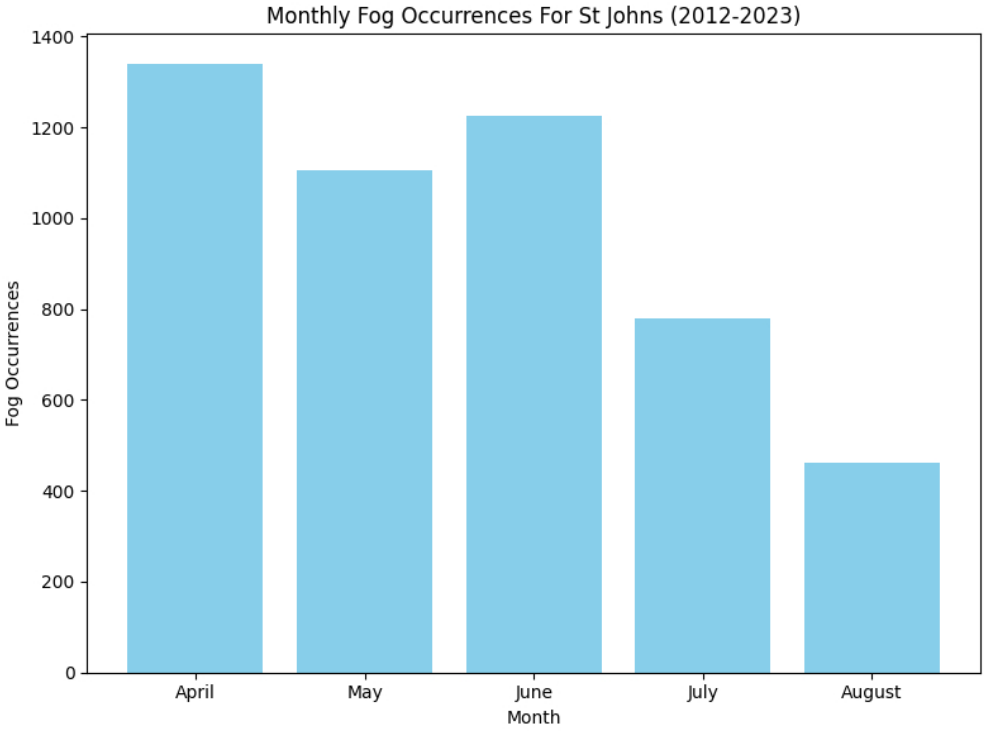

Figure 3: Hourly fog occurrence classified monthly for St John's, with the peak month being April.

278x208mm (72 x 72 DPI)

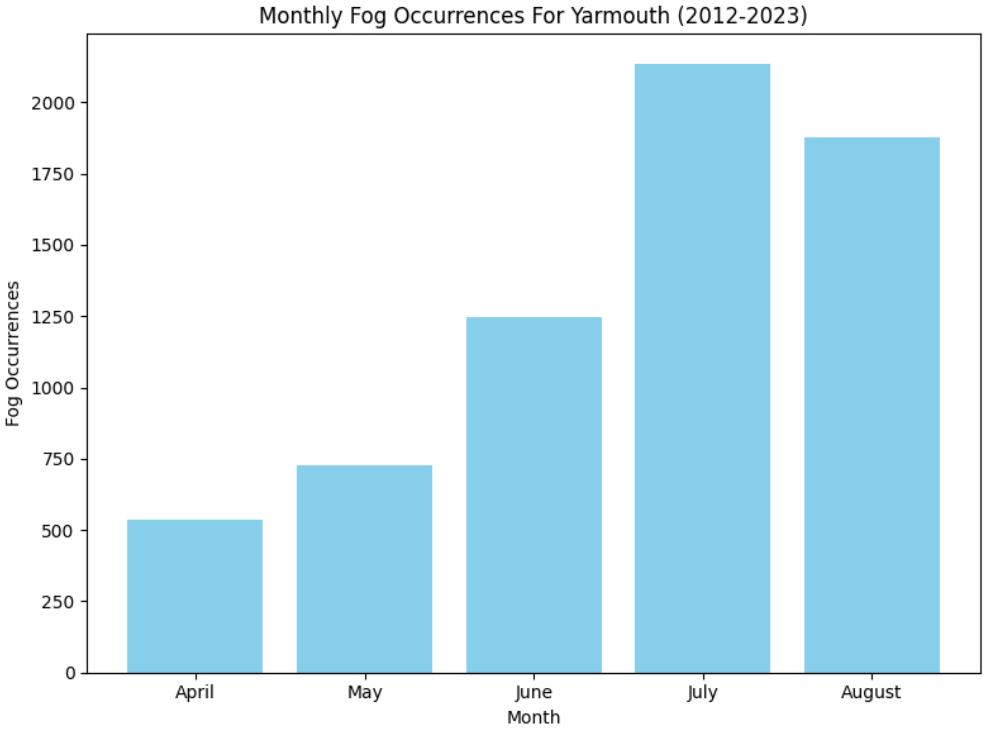

Figure 4: Hourly fog occurrence classified monthly for Yarmouth, with the peak month being July.

278x208mm (72 x 72 DPI)

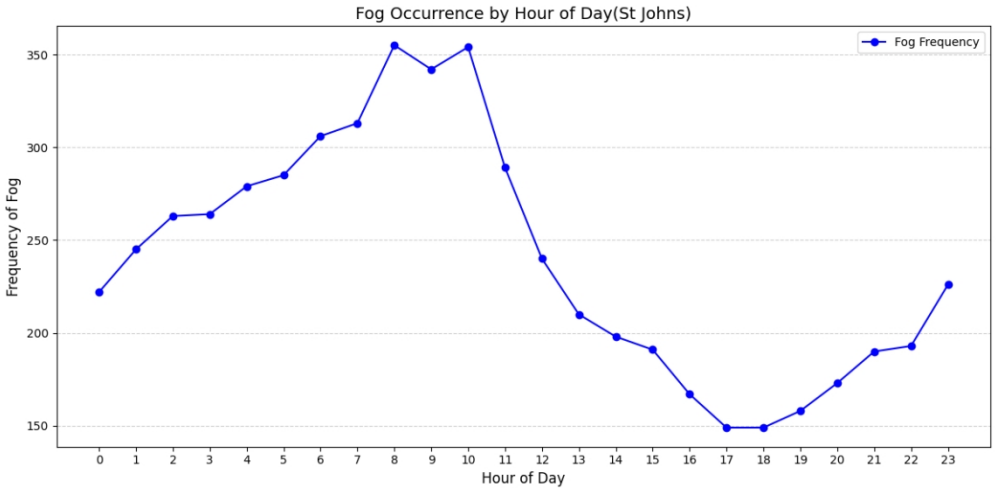

Figure 5: Line plots of hourly fog occurrences in a day during March to August, from 2012 to 2023 for St John's. Since the time is in UTC, this indicates higher fog during nighttime and early morning.

419x208mm (72 x 72 DPI)

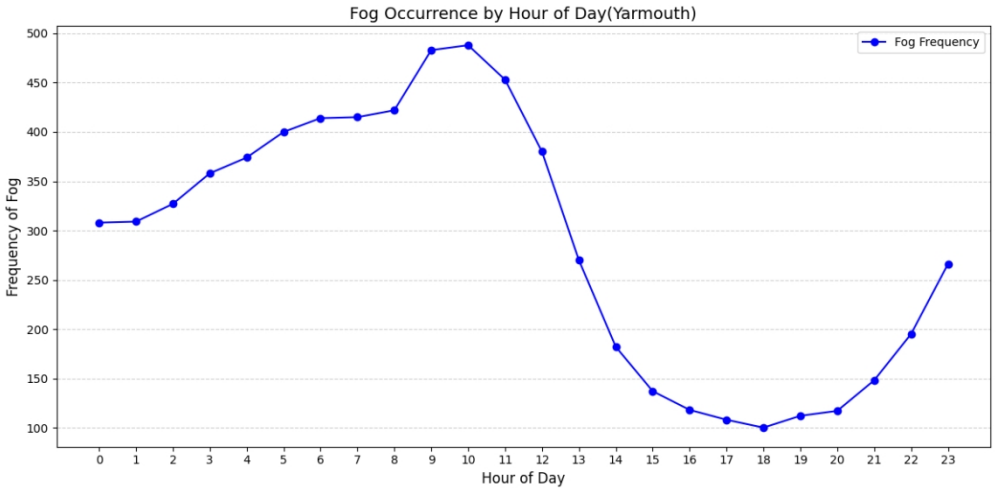

Figure 6: Line plots of hourly fog occurrences in a day during March to August, from 2012 to 2023 for Yarmouth. Similar to St John's, this indicates higher fog during nighttime and early morning.

419x208mm (72 x 72 DPI)

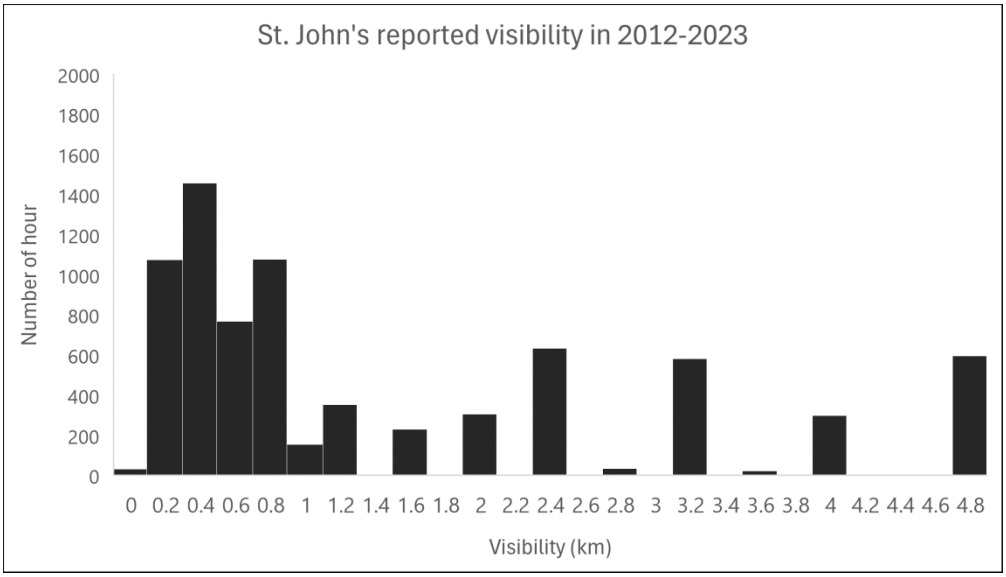

Figure 7: Histogram of reported visibility in St. John's in the year 2012-2023, from April to August each year. Only values lower than 5 km are shown.

412x234mm (130 x 130 DPI)

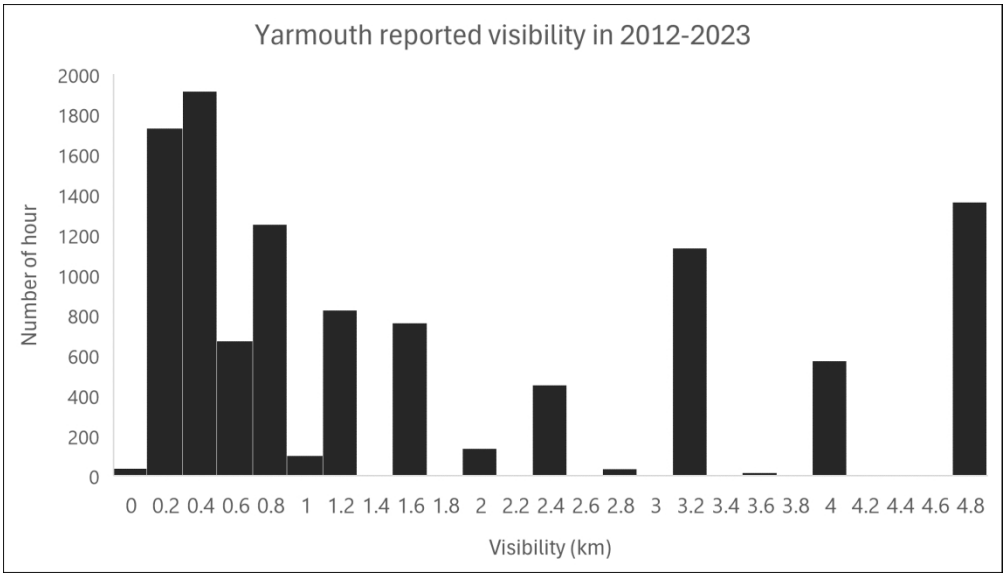

Figure 8: Histogram of reported visibility in Yarmouth, in the year 2012-2023, from April to August each year.

412x234mm (130 x 130 DPI)

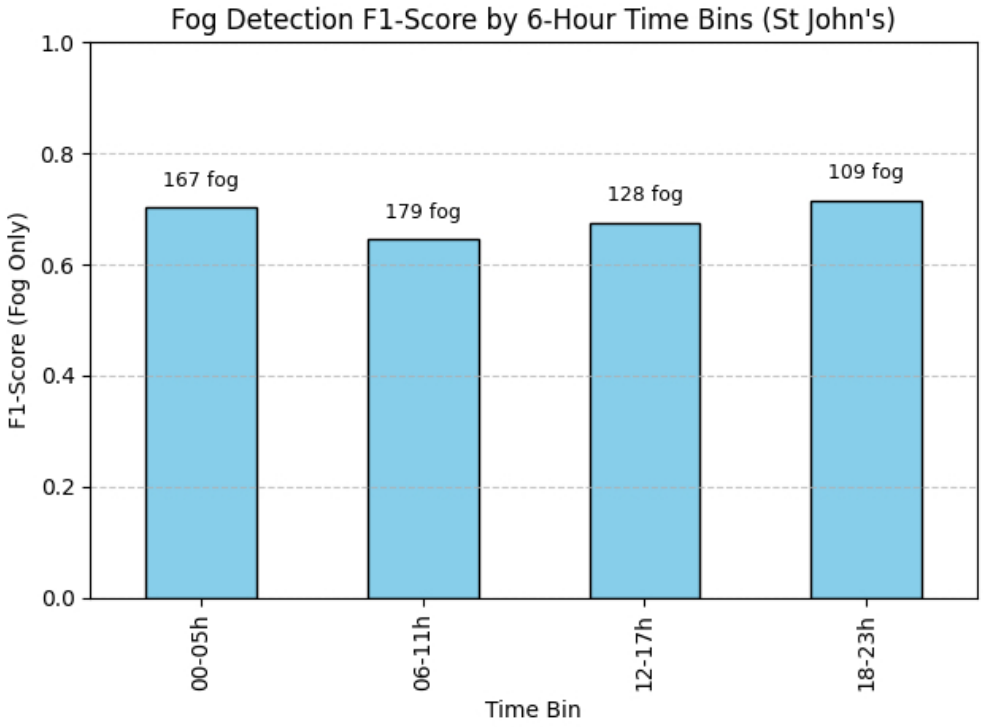

Figure 9: Hourly fog classification performance evaluated using F1-score in 6-hour time bins for St. John's, based on 24-hour forecasts from the 2024 dataset.

410x306mm (39 x 39 DPI)

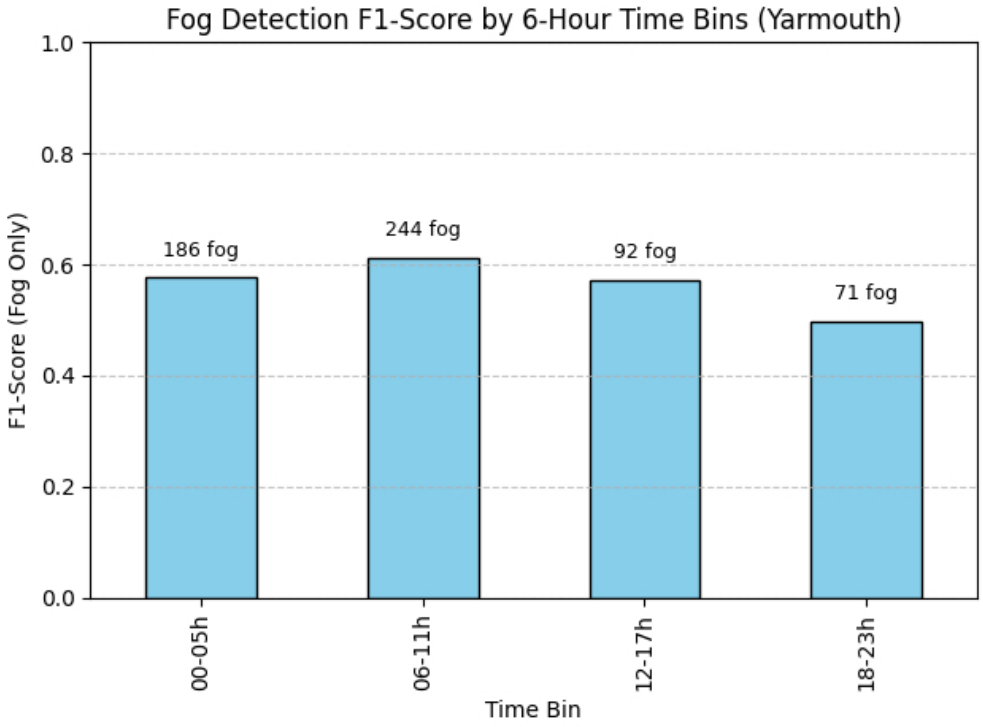

Figure 10: Hourly fog classification performance evaluated using F1-score in 6-hour time bins for Yarmouth, based on 24-hour forecasts from the 2024 dataset.

410x306mm (39 x 39 DPI)

Hyperparameters	St. John's	Yarmouth
bootstrap	False	False
class_weight	balanced	balanced_subsample
criterion	gini	entropy
max_depth	40	None
max_features	log2	sqrt
min_samples_leaf	2	1
min_samples_split	11	12
n_estimators	450	150

Table 1: Best parameters for the ETClassifier model for both locations using RandomSearchCV.

714x395mm (59 x 59 DPI)

Hyperparameters	St. John's	Yarmouth
max_depth	9	7
min_child_weight	7	11
subsample	0.9	1
lambda	4	4.4
alpha	2	0.3
gamma	2	1.5

Table 2: Best parameters for XGBoost for both locations using RandomSearchCV.

714x308mm (59 x 59 DPI)

Location	Class	Model	Precision	Recall	F1-score	Labels count
St John's	Clear	ETClassifier	0.96	0.94	0.95	3749
		XGBoost	0.94	0.95	0.95	
		biLSTM	0.96	0.94	0.95	
	Fog	ETClassifier	0.70	0.77	**0.73**	656
		XGBoost	0.75	0.72	0.73	
		biLSTM	0.70	0.75	0.72	
Yarmouth	Clear	ETClassifier	0.94	0.91	0.92	3540
		XGBoost	0.93	0.92	0.93	
		biLSTM	0.93	0.92	0.92	
	Fog	ETClassifier	0.66	0.74	**0.70**	835
		XGBoost	0.64	0.67	0.65	
		biLSTM	0.66	0.69	0.68	

Table 3: Classification Report for St John's and Yarmouth on the 2023 test data using ETClassifier, XGBoost and biLSTM.

194x170mm (150 x 150 DPI)

Location	Model	Precision	Recall	F1-score	Fog Count
St John's	ETClassifier	0.68	**0.71**	**0.69**	583
	XGBoost	0.62	0.69	0.66	
	biLSTM	**0.69**	0.68	0.68	
	WRF only	0.62	0.62	0.62	
Yarmouth	ETClassifier	**0.54**	**0.64**	**0.60**	593
	XGBoost	0.48	0.63	0.55	
	biLSTM	0.52	0.60	0.56	
	WRF only	0.47	0.60	0.53	

Table 4: Classification Report for St John's and Yarmouth on the 2024 forecast data using ETClassifier, XGBoost and biLSTM as a post-processing ML approach and WRF.

159x135mm (150 x 150 DPI)