Comparison of machine learning models in outdoor temperature sensing by commercial microwave link

Ondřej Pospíšil[1], Petr Musil[1] and Radek Fujdiak[1]

[1]Brno University of Technology, FEEC, Department of Telecommunications, Brno, Czech Republic

E-mail: xpospi89@vut.cz, xmusil56@vutbr.cz, fujdiak@vut.cz

Abstract—The main objective of this work is to focus on outdoor temperature prediction using machine learning based on parameters from commercial microwave links. This information can be used to refine the weather information at a given link location. Three machine learning models (random forest, linear regression, and lasso) are used for prediction using a combination of two datasets (ERA5 weather dataset and CML monitoring database dataset). The results were evaluated based on two evaluation metrics (R^2 and mean absolute error (MAE)). In this work, the ERA5 outdoor temperature was found to be correlated with the temperature of the microwave link unit, and results were obtained with an accuracy of 0.87144 based on the MAE metric. Thus, the results can fairly well predict actual outdoor temperatures in the microwave link area based on the microwave link unit temperature.

Keywords— microwave link, machine learning, random forest, linear regression, lasso

1. INTRODUCTION

Possibilities of weather observation based on opportunistic approach by using commercial microwave links (CML) operated within cellular networks have been a subject of research for many years [1]. Numerous studies were already published where they mainly targeted the estimation of rainfall derived from attenuation of a microwave signal within the CML [2, 3]. The studies have shown that commercial microwave links could provide vast possibilities of very accurate weather observations and also predictions even within very short time periods (e.g. tens of seconds).

Beside measuring rainfall, researchers in study [4] also focused on sensing the temperature through a network of devices, where temperature sensors are built-in and monitored for different reasons - usually for monitoring whether the inner temperature of the device is within its safe operational range. Study [5] collects data from smartphones from 8 different cities to determine the temperature in different periods of the year. Some microwave units of microwave links also provide a possibility of monitoring the inner temperature of the outdoor unit (ODU) placed usually on the roof of buildings where they are directly affected by the actual weather at the given location. In this article we use temperature data obtained from CML via SNMP (Simple network management protocol).

Obtained values of inner temperature of ODU is being affected by two groups of factors:

- **factors increasing temperature:** sunbeams falling directly on the surface of ODU, heat produced by the electronics of the ODU, actual modulation scheme and transmission power of the whole microwave link.
- **factors decreasing temperature:** air flow and wind of temperature lower than the ODU temperature, precipitation (rain and snowfall).

Factors influencing the device temperature tend to affect the microwave unit in a different way throughout the year. Warmer months of the year have predominant effect of temperature increase caused by high intensity of sun radiation that warms up the chassis of the unit. Strong storms with high rain intensity also might cause immediate drop of the device temperature due to the cooling effect of raindrops which spray over the microwave unit chassis. On the other hand in colder months the intensity of sun radiation is significantly lower than in the summer period of the year. Cold wind and precipitation lead to further decrease of the device temperature. The difference of temperature characteristics in a typical summer and winter week can be seen in Figure 1.

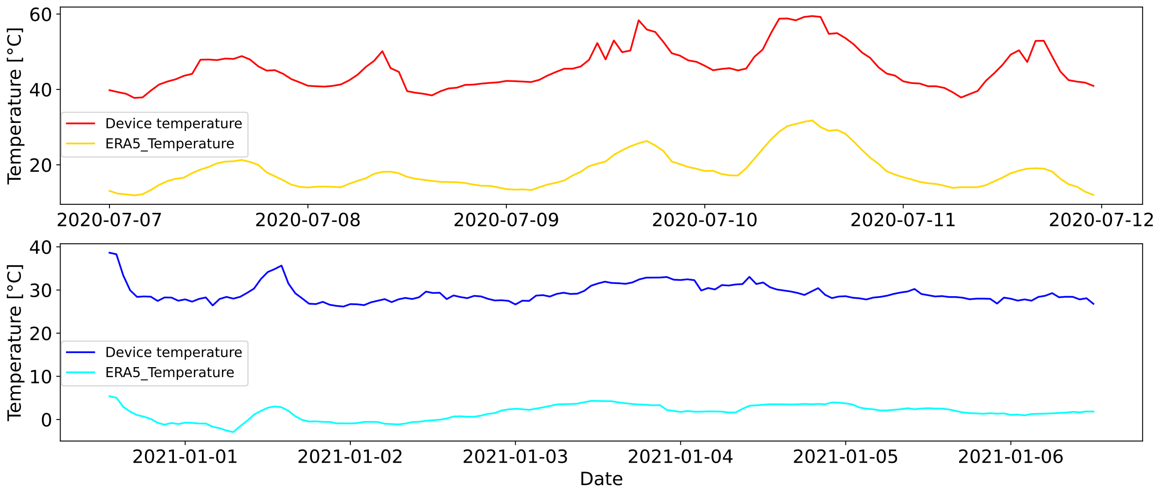

Figure 1: ERA5 temperature and device temperature in July and in January.

If all factors that influence the temperature of ODU were considered, we could estimate the actual outside temperature with a very high accuracy. Microwave links operators usually run networks consisting of large numbers of microwave units which could form a distributed network of sensors which could provide additional data for weather observation and monitoring that could improve accuraccy of predictions. If temperature readings from precise certified weather stations were taken as reference, network of microwave link units with their coordinates would form a map that could provide a detailed overview about detailed temperature changes with high accuracy. Temperature trends of this microwave unit network could be also used as an estimate of cloudiness, sun radiation, wind, etc., due to its increase/decrease in the temperature of ODU. In this article we utilize methods of machine learning to produce accurate estimates of the outdoor temperature.

2. METHODS

The analysis consists of exploring the possibilities of predicting weather patterns based on information from microwave links. A large amount of data is generated and stored from these microwave links and can be used to refine information about weather conditions at the locations where they are deployed. Based on these data, we attempted to predict values from the ERA5 weather model. We focused on prediction using regression. For this we used machine learning methods namely Random forest, Linear regression and Lasso. The algorithms were chosen because of the nature of the data in the dataset, which are based on time series. We have also chosen the three given methods because of their frequent use with good results in works dealing with weather forecasting.

2.1. Dataset

For purposes of this work we were using two separate datasets:

- **ERA5 weather dataset** is being covered by the project of the European Centre for Medium-Range Weather Forecasts (ECMWF) which provides hourly estimates of a large number of meteorological and climatological variables. The data cover the whole Earth surface on a 30 km grid in the various heights up to 80 km. They are obtained via the reanalysis process, which combines the observations data (e.g. measurements from satellites, weather stations, radars (where available) but also aircrafts, radiosondes, ships, etc.) with past short-range weather forecasts. The method used at ECMWF for the combination of these atmospheric data is called four-dimensional variational data assimilation (4D-Var). The temperature parameter is the temperature of the air at 2 meters above the surface of land, sea or in-land waters. It is calculated by interpolating between the lowest model level and the Earth's surface, taking account of the atmospheric conditions.

- **CML monitoring database dataset** consist of monitored parameters of each microwave link which is being performed by the commercial microwave link operator. Collection and analysis of such data helps to determine early degradation of microwave link parameters such as signal degradation, radio interference, length of outages, etc. Such data could be also used in weather observation and analysis as a secondary use-case. Every microwave link is being interrogated every 30 seconds via

SNMP and all the information is being stored in the server database. Such values are being stored for 3 months, data older then 3 months are being downsampled and averaged to 5 minutes.

The dataset was created from the values from one microwave link. We have different datasets for 79 individual microwave links. We performed tests on 5 links in parallel and we obtained values that showed small variation. A description of our results is demonstrated using data from a single microwave link. Thus, the dataset used for the regression has a total of 17 parameters. The parameters can be divided into three categories, namely the parameters from the microwave link, the parameters from the ERA5 model (information about the weather in the microwave link area) and the time parameters:

- **Parameters from the microwave links:** Signal, signal to noise ratio, modulation scheme and device temperature.
- **Parameters from ERA5:** Temperature, horizontal wind speed in m/s, vertical wind speed in m/s, cloud cover in the range 0-1 (0 = clear, 1 = completely cloudy), hourly precipitation in m, hourly snowfall in m.
- **Time parameters:** Unixtime, year, month, day, day of the week, day of the year, time.

The dataset contains information on individual microwave links and weather values from ERA5 in hourly intervals. The dimension of the dataset is 14317 rows and 17 columns (14317,17). The data was collected from the period 26.06.2020 - 14.02.2022. For training and testing, the dataset was divided into two sets with a ratio of 80% training data to 20% testing data. In order to achieve the objectives of the work, all the parameters from the microwave link were extracted. Based on these parameters, the ERA5 temperature (the actual temperature from the ERA5 model at a given location) was first predicted.

In Table I, the first three rows of the dataset can be seen. It contains the following parameters. The first two parameters are the index number (not counted as a separate column, it is only informative) and unixtime. Then there are six values from the ERA5 weather model. ERA5_Temp (outdoor temperature), ERA5_W_u (wind speed in horizontal vector), ERA5_W_v (wind speed in vertical vector both values in meters per second), ERA5_Clouds (cloud value), ERA5_Prec (precipitation value) and ERA5_S (snowfall value). There are also 4 parameters from the microwave link, namely signal (signal strength), quality (signal quality), Mod (modulation) and temp (device temperature). The time values are Year, M (month), D (day), dW (day of the week), dY (day of the year) and T (time).

Table I: Sample of the first lines of the dataset.

	Unixtime	ERA5_Temp	ERA5_W_u	ERA5_W_v	ERA5_Clouds	ERA5_Prec	ERA5_S	signal	quality	Mod	temp	Year	M	D	dW	dY	T
0	1593158400	20.164209	-2.927006	0.287172	0.966443	0.000158	0.0	-48.913	27.74	5.0	50.242	2020	6	26	4	178	8
1	1593162000	21.133203	-2.709867	-0.281766	0.973783	0.000092	0.0	-48.797	27.74	5.0	49.048	2020	6	26	4	178	9
2	1593165600	20.563867	-3.737828	0.562858	0.553122	0.000126	0.0	-48.080	28.38	5.0	46.196	2020	6	26	4	178	10

2.2. Approach

As part of the regression problem, we addressed the temperature prediction (ERA5 model) in the microwave link area based on the microwave link parameters. Here, the most important parameter for this prediction was the temperature of the microwave link device itself. For this prediction, we dropped all other information from the ERA5 model from the dataset so that it would not affect the prediction. Subsequently, we used the three machine learning models namely Random forest, Linear regression and Lasso. For training, testing and evaluation of the individual machine learning models we used the best known machine learning library scikit-learn. We validated the evaluation and procedure on five different independent microwave link datasets.

2.3. Evaluation metrics

We used two popular metrics for evaluating regression models, namely R squared (R^2) and Mean Absolute Error (MAE). R^2 or the coefficient of determination, means how much of the variance of the depending variable can be explained with the variance of the independent variable. Simplified R^2 compares models predictions to the mean of the targets. The values can range from negative infinity (meaning very bad results) to the positive value of 1 (the best possible result). The MAE is calculated as the average of the sum of the absolute differences between the predicted values and the actual values. That is, how far on average the model predictions are from the actual values.

3. RESULTS

3.1. Regression results

In this section, we present experimental results of the regression models used for temperature prediction (ERA5_Temperature). The results of each model can be seen in Table II. These models are namely Random forest, Linear regression and Lasso.

Table II: Results of regression models for forecasting.

	Train		Test	
Regression	R^2	MAE	R^2	MAE
Random forest	0.99708	0.32080	0.97960	0.87144
Linear regression	0.94773	1.51182	0.94401	1.55142
Lasso	0.93874	1.62440	0.93385	1.67967

From the results in the table, it can be seen that the most appropriate model for this problem is the random forest regressor. Its R^2 value for the test data is 0.97960 and the MAE value is 0.87144, that is, the predicted data differ on average by 0.87144. The prediction could be refined with more data. The other two models already achieve worse values and are quite inaccurate. The Lasso model predicts the worst values when the predicted values differ from the actual values on average by 1.6796.

Next, we plotted the prediction results of each model into individual graphs which can be seen in Figure 2. Due to the large number of samples, we plotted only the first 100 rows in the graphs for clarity. So on the x-axis is the row index and on the y-axis is the ERA5 Temperature. The actual temperature is marked as the blue square and the predicted temperature as the red circle.

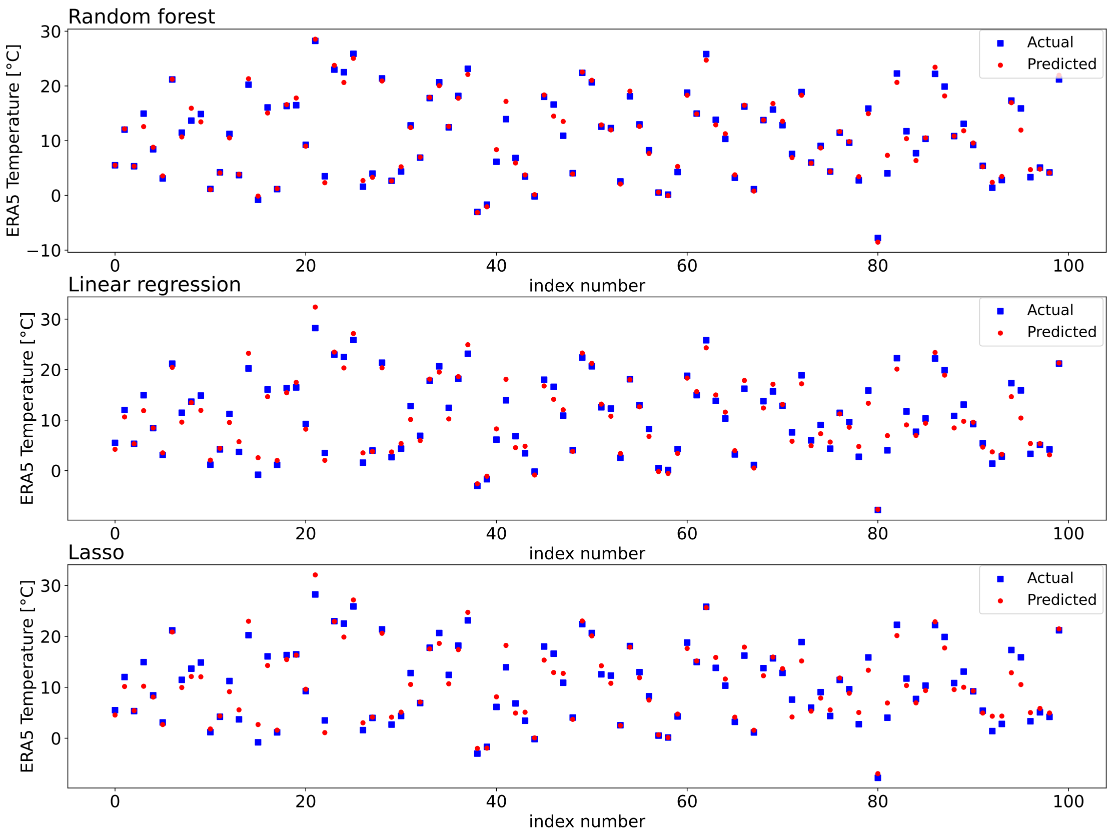

Figure 2: Actual and predicted values of the regression models.

According to the results it can be seen that the most accurate is the Random forest model, where the overlap of values is quite frequent, but there is a lot of deviation. On the other two models we can see already larger individual deviations of the prediction.

4. DISCUSSION

The main idea of the work was to investigate the possibilities of outdoor temperature prediction based on the values from commercially used microwave links, from which parameters such as signal strength, signal quality, modulation and device temperature can be obtained. As expected, the results show that the greatest influence on the prediction of the outdoor temperature for the models is the device's temperature. Thanks to this value, the outdoor temperature within a given microwave link can be predicted quite well after training. This information can be useful for adding information to weather models that may lack more precise data from a given area, and thus make this weather data more accurate. Of the three models chosen, the random forest had the best results. The reason may be that, in general, random forests provide better results and perform well on large data sets as they are able to handle missing data well by creating estimates for the missing data.

The limitations of our solution are mainly in testing fewer models and not using deep learning models. Furthermore, the hyperparameters of each model could also be better tuned in the future. The plan is to use other machine learning models, especially deep learning, which could make our results more accurate.

5. CONCLUSION

In this paper, we compared several machine learning methods for predicting ERA5 Temperature within our own dataset from a single microwave link. We validated the individual model values on four additional microwave link datasets, and the average difference for the MAE evaluation metric across the test results was as follows for each model: random forest – 0.03834, linear regression – 0.1283, and Lasso – 0.1562. From the point of view of the results, the random forest regressor seems to be the most appropriate model. Its accuracy was 0.97960 based on the evaluation metric R^2 and 0.87144 based on the MAE metric. With more weather values from the ERA5 weather model, we want to further explore the possibility of detecting other parameters such as cloud cover, precipitation, wind and snow. As mentioned in the text, the prediction of rainfall was not very successful for these models. Therefore, we would approach this problem by multiclass classification, where we would add several classes to the rain values based on the amount of rainfall. We would then use these values to predict how high the precipitation is based on the defined classes. In this way, other parameters (snowfall, wind, cloud cover, etc.) could also be approached. It would also be useful to compare deep learning methods with these results.

ACKNOWLEDGMENT

The described research is part of the grant project registered under no.TK02030013 and funded by the Technology Agency of the Czech Republic. This project was also supported by the company CBL Communication by light s.r.o. which provided datasets for this work. Authors would like to thank especially Mr. David Smékal for his willing cooperation.

REFERENCES

[1] H. Leijnse, R. Uijlenhoet, and J. N. M. Stricker, "Hydrometeorological application of a microwave link: 2. Precipitation", *Water resources research*, vol. 47, no. 4, 2007. doi 10.1029/2006WR004989.

[2] C. Chwala and H. Kunstmann, "Commercial microwave link networks for rainfall observation: Assessment of the current status and future challenges", WIREs Water, vol. 6, no. 2, 2019, doi: 10.1002/wat2.1337.

[3] M. Fencl, M. Dohnal, P. Valtr, M. Grabner, and V. Bares, "Atmospheric observations with E-band microwave links - challenges and opportunities", *Atmospheric measurement techniques*, vol. 13, no. 12, pp. 6559-6578, 2020. doi 10.5194/amt-13-6559-2020.

[4] D. Vos, A. M. Ste, M. J. Zander, A. Overeem, H. Leijnse, B. G. Heusinkveld, G. J. Steeneveld, and R. Uijlenhoet, "Hydrometeorological monitoring using opportunistic sensing networks in the Amsterdam metropolitan area", *Bulletin of the American Meteorological Society*, vol. 101, no. 2, pp. E167-E185, 2020. doi 10.1175/BAMS-D-19-0091.1.

[5] A. Overeem, J. C. R. Robinson, H. Leijnse, G. J. Steeneveld, B. K. P. Horn, and R. Uijlenhoet, "Crowdsourcing urban air temperatures from smartphone battery temperatures", *Geophysical research letters*, vol. 40, no. 15, pp. 4081-4085, 2013. doi 10.1002/grl.50786.