FORECASTING ALGORITHM BASED ON TEMPERATURE ERROR PREDICTION USING KALMAN FILTER FOR MANAGEMENT SYSTEM DEVELOPMENT

N. Bogdanovs, R. Belinskis, V. Bistrovs, E. Petersons, A. Ipatovs

Department of Transport Electronics and Telematics, Faculty of Electronics and Telecommunications, Riga Technical University, LV-1658, Riga, LATVIA

The study offers a new method of collection and processing of meteorological data from the meteorological service based on observations and correction of numerical weather forecast errors using a new prediction algorithm. This algorithm vastly increases the accuracy of the short-term forecast of outdoor air temperature, which is subject to uncertainty due to the stochastic nature of atmospheric processes. Processing of temperature data using Kalman filter provides the decrease in predicted temperature errors. The main setup methods of Kalman filter have been examined. The article also describes the implementation of accuracy improving algorithm of predicted temperature using Python.

Keywords: Algorithm, forecast, Kalman filter, prediction, temperature.

1. INTRODUCTION

Temperature forecast data are of importance in different fields, such as agriculture, tourism, hydrology, and construction for heating, ventilation, and air conditioning system operation in the buildings.

Many meteorological services provide information about forecast temperatures for exact place or region. However, this does not mean that temperature forecast is available for each point on the Earth. In some cases, the temperature information is used for the nearest available place.

The main problem from the perspective of meteorological data consists in their
inherent uncertainty due to the stochastic nature of atmospheric processes, imperfect knowledge of initial conditions of weather model and modelling errors. To solve this problem, the authors of the study propose using Kalman filter (KF), which will gather the data from the meteorology service. Temperature information is based on the outputs of special weather models or weather data numeric analysis. It is impossible to forecast the temperature values that would be equal to the observed ones using weather models. Therefore, there is a certain level of deviation from the observed temperature. This will be demonstrated in Section II. Section III will give possible analytical presentation of temperature forecast errors. Section IV defines KF based algorithm for temperature data processing. The results of temperature data processing will be discussed in Section V.

2. RELATED WORKS

Perera discussed the energy saving potential in residential buildings in Norway using the developed mathematical model [1]. Research did not use Kalman filtering, but the author’s method obtained a good approximation to model parameters.

One of the widely used techniques for improving numerical weather forecast is model output statistics (MOS). It works well but cannot account for every local effect and does not correct systematic errors from numeric weather prediction (NWP) model [2]-[4].

Another technique for improving weather forecast is based on regression equation development. This algorithm is called PERFECT PROG. The drawback of perfect program is its inability to correct biases of forecast data [5]-[7].

Kalman filter based algorithm is naturally designed to eliminate systematic errors, and it easy integrates forecast data based on NWP model, which considerably simplifies the core of the algorithm. The last leads to a decrease in the data processing time.

Despite a large number of valuable research papers describing different types of weather forecast algorithm, there is no weather forecast system that could be integrated into existing buildings using predictions and ensuring corrections in air temperature forecast within next three hours.

3. TEMPERATURE FORECAST ERRORS

The temperature forecasts for Riga, Latvia, were provided by Eurometeo or Gismeteo. Local temperatures measurements were obtained from meteorological station Riga (University).

Figure 1 presents daily evolution of the measured and forecast (Eurometeo) temperatures in January 2018 for Riga. The difference between the observed and forecast temperatures is well observable. This difference can be interpreted as a temperature forecast error. Distribution of this forecast error is shown in Fig. 2.

Figure 3 presents distribution of the temperature forecast errors for March. The bias and the standard deviation of temperature forecast errors for January, February, and March are shown in Table 1. The fore-
cast errors values are changing randomly within a certain range of temperatures. In addition, there is non-zero bias for error values.

Fig. 1. The temperature observations and forecasts (www.eurometeo.ru) for Riga from 8 to 15 January 2018.

Fig. 2. The distribution of the temperature forecasting error for Riga.

Kalman filter (KF) can be suggested to process forecast temperature data for decreasing a systematic forecast error [1]–[5]. As an output of temperature model (temperature prediction) is available from various meteorological services, there is no need to develop temperature model for KF. This simplifies the algorithm development process and increases the data processing speed. The measured temperatures are available in time series data for the needed period. Thus, the advances in weather forecast processes allow considering the KF algorithm for data processing from another perspective and obtaining valuable results.
Table 1. Statistical Parameters of the Temperature Forecasting Error

| Errors                  | January | February | March  |
|-------------------------|---------|----------|--------|
| Bias of the forecast error, °C | 0.7520  | 0.5960   | 0.6859 |
| Std. deviation of the forecast error, °C | 3.2512  | 1.9958   | 1.4479 |

4. ANALYTICAL DEFINITION OF THE FORECAST ERROR

It was shown in [1], [4] that the forecasting error can be represented in function of the forecast temperatures:

$$
\varepsilon = \alpha_0 + \alpha_1 T_m + \alpha_2 T_m^2 + \cdots,
$$

where $\alpha_n$ – corresponding regression coefficients, $\varepsilon$ – a temperature forecast error.

The advantages of Eq. (1) are possible forecast error non-linearity and integration of such a measurement model in the updating stage of linear KF. The regression coefficients are the system states of KF.

5. KALMAN FILTER DEFINITION

The Kalman filter is an estimation algorithm that provides optimal estimates for the system state. The algorithm uses information about deterministic and statistical properties of the system and direct and indirect measurements.

The basic elements of the KF algorithm are system state vector, $X_k$; error covariance matrix, $P_k$; system noise covariance matrix, $Q_k$; system model, $\Phi_k$; measurement vector, $Z_k$; measurement noise covariance matrix, $R_k$; and measurement model, $H_k$.

The system noise covariance matrix describes uncertainty of the true values of the system states. The error covariance matrix describes uncertainty of the state...
estimates. The system model describes evolution of the states and error covariance matrix during discrete time step, k, of KF operating. In the present case, there is no knowledge how system states are changing over the time; therefore, the state transition matrix is a simple identity matrix. The system noise covariance matrix will be a diagonal matrix, as it is assumed that there is no correlation between system noises of the states. The values of such a diagonal covariance matrix can be increased to account for possible unknown dynamics of the analysed system. The measurement vector consists of only one element – difference between locally observed or measured temperature and temperature forecast. If the quality of the data measurement is unknown, the diagonal elements of the measurement noise covariance matrix can be increased. Normally, this matrix describes statistical properties of the temperature sensor. In the present case, this property is unknown; therefore, the values were defined empirically. The measurement model describes relation between the system state and the measurement vector. Measurement vector is defined as difference (forecast error) between forecast temperature $T_{mo}$ and observed temperature $T_{me}$. The polynomial of n degree describes the relation between the forecast error and forecast temperature using regression coefficients $\alpha$ [1].

The KF c algorithm has two steps – prediction and updating (or correction). During the prediction step, the algorithm estimates the error covariance matrix and system states. As the state transition matrix is an identity matrix, the predicted states for next time step are equal to the corrected states from the previous time step [7], [8]. During updating steps, the system states are corrected using the observed temperature information. The error covariance matrix is updated as well. The set of the equations of KF is given below:

**Updating:**

$$
\begin{align*}
    K_k &= P_k (\text{predicted}) H_k^T (H_k P_k (\text{predicted}) H_k^T + R_k)^{-1}, \\
    \hat{X}_k &= \hat{X}_k (\text{predicted}) + K_k (Z_k - H_k \hat{X}_k (\text{predicted})), \\
    P_k &= (I - K_k H_k) P_k (\text{predicted}),
\end{align*}
$$

**Predicting:**

$$
\begin{align*}
    \hat{X}_{k+1} (\text{predicted}) &= \Phi_k \hat{X}_k, \\
    P_{k+1} (\text{predicted}) &= \Phi_k P_k \Phi_k^T + Q_k,
\end{align*}
$$

where $K_k$ – Kalman gain matrix, $P_k$ – state uncertainty covariance matrix, $H_k$ – measurement sensitivity matrix, $R_k$ – measurement noise covariance matrix, $\Phi_k$ – state transition matrix, $Q_k$ – system noise covariance matrix, $\hat{X}_k$ – state estimates.

The corresponding state vector is defined as $X_k = [\alpha_{0,k} \alpha_{1,k} \alpha_{2,k}]$, where $\alpha$ – a regression coefficient. The degree of polynomial equals two. It was observed that when using higher degrees (three and more) of polynomials, considerable estimation error deviation appears. A similar conclusion was obtained in [1]. Initial values for regression coefficients were set to zero.

Taking into account Eq. (1), the measurement sensitivity matrix was defined as:

$$
H_k = [T_{mo,k} T_{mo,k}^2]
$$

The diagonal values for covariance matrices were defined during data processing trials by KF in order to achieve decrease bias and standard deviation for the temperature forecast error. The corresponding diagonal elements of the covariance matrices were defined within a value range from 0.4 to 1.4 for the measurement noise and from 0.01 to 0.1 for the system noise.

Algorithm steps are shown in Fig. 4. Three-hourly (at 2:00, 5:00, 8:00, 11:00, 14:00, 17:00, 20:00, 23:00) temperature forecast (external data) for Riga was provided by Eurometeo (www.eurometeo.ru) or Gismeteo (www.gismeteo.ru). The corresponding local temperature measure-
ments were obtained from the meteorological station in Riga for the same hours as the forecast. These data were processed by KF and the regression coefficients, \( \alpha_0, \alpha_1, \alpha_2 \), were evaluated as a result. The time of the processing step for prediction and updating stages was equal to 3 hours. Correction error (\( \varepsilon \)) was calculated by the algorithm and added to the temperature forecast value for the next time step. Then, cubic interpolation for corrected temperature data series was made in order to obtain temperature forecast for each hour (output data).

The functionality of this algorithm was also checked when the observed temperature was not available for 1–5 days. During this situation, the algorithm continued to provide forecast at an acceptable quality level (better than the temperature forecast provided by the model of the considered meteorological services).

![Algorithm steps](image)

**Fig. 4.** Algorithm steps.

### 6. TEMPERATURE DATA PROCESSING RESULTS

Results of data processing, using algorithm described in Section IV, will be discussed in this section. The temperature data were analysed for three months in order to calculate bias and standard deviation.

The evaluation of statistical parameters of corrected forecast error is given in Table 2.

|               | January  |   | February |   | March  |   |
|---------------|----------|---|----------|---|--------|---|
|               | model    | KF| model    | KF| model  | KF|
| Bias, °C      | 0.752    | 0.025 | 0.5960   | 0.0007 | 0.686  | 0.002 |
| Std. dev., °C | 3.251    | 2.599 | 1.9958   | 1.3057 | 1.448  | 1.221 |

Results in Table 2 show a decrease in systematic bias for corrected forecast temperature. Standard deviation slightly decreased as well. The observed decrease in the bias error is more than 100 times comparing with uncorrected forecast. Standard deviation decreases in average by 20%.

Figure 5 shows the distribution of temperature error after meteorological data processing by KF.

![Figure 5](image)
Figure 5 shows that error values (difference between the corrected and measured temperature) are concentrated around zero and the shape of histogram is very similar to the normal one. Corrected values by KF follow the measured values more exactly comparing with direct output of temperature model (forecast values). Error distribution (for February 2018) between forecast and measured values is shown in Fig. 2.

Additional experiment was conducted for forecast temperature in summer time. The forecast temperatures using model were taken from the meteorological service gismeteo.ru.

Figure 6. The distribution of forecast error (www.gismeteo.ru) for Riga from 11 to 19 July 2018.
The temperatures were forecast for the period of 11–19 July 2018 for Riga (Latvia). During this period, the heat wave reached the region of Riga. The distribution of temperature difference between the forecast and measured temperature is shown in Fig. 6. There are significant error components in the range of 1 to 3 degree. This indicates that it was quite complicated to forecast precisely temperature dynamics during this period. One of the reasons can be that the air was warm itself and there was no need for sun to spend energy to maintain it warm. However, sun energy can be spent to increase air temperature at once. Therefore, temperature highly depends (more than usually) on the level of cloudiness. The forecast for cloudiness was quite unstable during this period and in some cases was stable only for next 8 hours.

The results of temperature data (the forecast and measurement data) processing are shown in Fig. 7. The temperature data processing algorithm gives the corrected forecast for next three hours based on the observed temperature for the present moment and temperature dynamics from the past period. This temperature dynamics is reflected by values of regression coefficients. Other mode of temperature data processing by KF was checked, when data processing algorithm worked a certain period with availability of the observed temperature data and the algorithm estimated regression coefficients, then the measured temperature was not available and algorithm used the predicted temperature value instead. The results of such processing are shown in Fig. 8.
The measured temperature was available from 11 July, 00-00 till 12 July, 00-00, and temperature was not available from 12-07, 01-00 till 19-07, 00-00. The correction of the forecast was still working but not so good compared with the situation when the observed temperatures were available, but still better than an initial forecast. Figure 8 shows that a forecasting error increases with time. This is logical as KF adjustment became worse, because the measured data were not available for correction.

Additional data processing (by KF) with the use of “past” temperature statistics allows improving precision of temperature forecast. This is especially true when the meteorological service uses an improper model of the temperature forecast for a local area of interest. KF uses real measurement data to provide optimal estimates of the system states [9], [10]. This means that the algorithm can take into account temperature changes due to the natural factors (altitude, type of local area, etc.), and changes in temperatures due to human factors which are difficult to model, but this information can be easily integrated into KF measurements. For example, the temperature sensor (needed for heating system adjustment in the room) is fixed somewhere near the window of the house. In wintertime, the window can be open and local temperature sensor will indicate that it is warmer outside and the heating system will provide less heating, as KF output will react on the obtained sensor measurement by increasing the forecast temperature. This is an advantageous behaviour of such an algorithm.

7. PRACTICAL IMPLEMENTATION OF THE WEATHER PREDICTION ALGORITHM

Great part of general energy consumed by buildings increased interest in optimizing building operation and promoted the development of new approaches to climate
Control in premises. The present study uses the algorithm providing correction of air temperature forecast within the next three hours [11], [12].

Figure 9 shows the implementation of the advanced algorithm of temperature prediction using Python.

Control system developed by Python can regulate the curve of building heating based on the external temperature forecast. The idea of the developed management logic consists in adjusting the heating curve depending on a future state [13]-[15]. The research describes the shifted profile of external temperature, and the changed curve of heating is calculated according to a set of parameter settings for the whole heating season. Weather forecast data are based on Eurometeo (www.eurometeo.ru) or Gismeteo (www.gismeteo.ru) databases considering the typical meteorological year. There were a number of tests performed to select the best period for the shift, and, finally, the period of one-hour standby was used. The results during the modelling were obtained in Matlab & Python [16], [17].

We managed to integrate it into the building old automatic equipment using OPC (Open Platform Communications) standard. We used Matlab machine learning toolbox and weather forecast. OPC is the interoperability standard for secure and reliable exchange of data in an industrial automation area [17]. By analysing next day weather forecast and the database with
data from wireless temperature sensors, a self-learning algorithm was created that allowed adjusting automatic equipment of the building.

This program analysis of Matlab helps define energy efficiency of buildings according to the European standard EN 15603. This standard sets rules for energy consumption assessment in the building and methods, which should be used to determine nominal parameters of building engineering systems.

8. CONCLUSIONS

The article proposed a new method of gathering and processing the meteorological data from the meteorological service. This algorithm can significantly increase the accuracy of short-term forecast of external air temperature. The considered algorithm provides temperature forecast correction for the next three hours. The temperature data processing by KF provides bias (decrease 100 times) and standard deviation (decrease by 20 %) decreasing forecasted temperature error. The design of Kalman filter can be simplified owing to data availability from different temperature forecast models. Therefore, it can provide the best temperature for enzymatic detection to ensure the accuracy of results. Such an algorithm takes into account unmodelled temperature deviations based on a series of measured temperatures. It was shown that useful temperature correction could be provided when the observed temperature was not available. One of the possible improvements of the analysed algorithm can be automatic tuning implementation.

9. FUTURE RESEARCH

In future, it is planned to address all the above-mentioned problems within the framework of research by providing the following solutions:

Several weathers forecast models used at the prediction stage of KF could potentially decrease a temperature forecast error. Introducing adaptive tuning of KF noise matrix coefficients makes set up of data processing faster. As weather is a nonlinear dynamic system, another type of KF – unscented Kalman filter – can be applied for better tackling with system nonlinearity.

Next research of our control algorithm will include human occupancy in the building. New algorithm using machine learning and building new data will upgrade our model results.

REFERENCES

1. Perera, D.W., & Skeie, N. (2016). Comparison of Space Heating Energy Consumption of Residential Buildings Based on Traditional and Model-Based Techniques. Buildings, 7(2), 27.

2. Galanis, G., Louka, P., Katsafados, P., Pytharoulis, I., & Kallos, G. (2006). Applications of Kalman Filters Based on Non-Linear Functions to Numerical Weather Predictions. Ann. Geophys, 24, 2451–2460.
3. Anadranistakis, M., Lagouvardos, K., Kotroni, V., & Skouras, K. (2002). Combination of Kalman Filter and an Empirical Method for the Correction of Near-Surface Temperature Forecasts: Application over Greece. *Geophysical Research Letters, 29* (16), 231–234.

4. Libonati, R., Trigo, I., & DaCamara, C. (2008). Correction of 2 m-Temperature Forecasts Using Kalman Filtering Technique. *Atmospheric Research, 87*, 183–197.

5. Homleid, M. (1995). Diurnal Corrections of Short-Term Surface Temperature Forecasts Using the Kalman Filter. *Weather and Forecasting, 10* (4), 689–707.

6. Brunet, N., Verret, R., & Yacowar, N. (1988). An Objective Comparison of Model Output Statistics and “Perfect Prog Systems” in Producing Numerical Weather Element Forecasts. *Weather and Forecasting, 3* (4), 273–281.

7. Bo, P. (2004). A Statistical Method for Forecasting Extreme Daily Temperatures Using ECMWF 2-m Temperatures and Ground Station Measurements. *Meteorol, 245–251.*

8. Chai, B., Tushar, W., Hassan, N. U., Yuen, C., & Yang, Z. (2016). Managing energy consumption in buildings through offline and online control of HVAC systems. In *IEEE Region 10 Conference (TENCON),* (pp. 3368–3373), 22–25 November 2016, Singapore.

9. Hadwan, H., & Reddy, P. (2016). Smart Home Control by Using Raspberry Pi & Arduino. *International Journal of Advanced Research in Computer and Communication Engineering, 5* (4), 283–288.

10. Lynggaard, P. (2014). *Artificial Intelligence and Internet of Things in a “Smart Home” Context: A Distributed System Architecture.* PhD Thesis. Press / Media.

11. Gomez Ortega, L., Han, L., Whittacker, N., & Bowring, N. (2015). A machine-learning based approach to model user occupancy and activity patterns for energy saving in buildings. In *Science and Information Conference,* (pp. 474–482), 10–12 June 2015, London, UK.

12. Zhang, Y., & Handy, I. (2007). Short-Term Prediction of Weather Parameters Using Online Weather Forecasts. *Building Simulation 2007*, 1411 –1417.

13. Mohinder, S., Grewal, & Angus, P. (2008). *Kalman filtering: Theory and practice using MATLAB* (3rd ed.). New Jersey: John Wiley & Sons.

14. Rolando, D., Madani, H., Braida, G., Tomasetig, R., & Mohammadi, Z. (2017). Heat pump system control: The potential improvement based on perfect prediction of weather forecast and user occupancy. In *12th IEA Heat Pump Conference 2017,* (pp. 1–9), 15–18 May 2017, Rotterdam, Netherlands.

15. Bogdanovs, N., Bistrovs, V., Ipatovs, A., & Beļinskis, R. (2018). Weather prediction algorithm based on historical data using Kalman filter. In *2018 Advances in Wireless and Optical Communications (RTUWO),* (pp. 94–99), 15–16 November 2018, Riga, Latvia.

16. Siroky, J., Oldewurtel, F., Cigler, J., & Privara, S. (2011). Experimental Analysis of Model Predictive Control for an Energy Efficient Building Heating System. *Applied Energy, 88* (9), 3079 –3087.

17. Brown, R.G., & Hwang, P.Y.C. (1997). *Introduction to random signals and applied Kalman filtering.* Wiley.