

Uncertainty Analysis on Electric Power Consumption

Oakyoung Han¹ and Jaehyouon Kim^{2,*}

¹University College, Sungkyunkwan University, Seoul, 03063, Korea

²Department of Computer Education, Sungkyunkwan University, Seoul, 03063, Korea

*Corresponding Author: Jaehyouon Kim. Email: jaekim@skku.edu

Received: 07 October 2020; Accepted: 15 February 2021

Abstract: The analysis of large time-series datasets has profoundly enhanced our ability to make accurate predictions in many fields. However, unpredictable phenomena, such as extreme weather events or the novel coronavirus 2019 (COVID-19) outbreak, can greatly limit the ability of time-series analyses to establish reliable patterns. The present work addresses this issue by applying uncertainty analysis using a probability distribution function, and applies the proposed scheme within a preliminary study involving the prediction of power consumption for a single hotel in Seoul, South Korea based on an analysis of 53,567 data items collected by the Korea Electric Power Corporation using robotic process automation. We first apply Facebook Prophet for conducting time-series analysis. The results demonstrate that the COVID-19 outbreak seriously compromised the reliability of the time-series analysis. Then, machine learning models are developed in the TensorFlow framework for conducting uncertainty analysis based on modeled relationships between electric power consumption and outdoor temperature. The benefits of the proposed uncertainty analysis for predicting the electricity consumption of the hotel building are demonstrated by comparing the results obtained when considering no uncertainty, aleatory uncertainty, epistemic uncertainty, and mixed aleatory and epistemic uncertainty. The minimum and maximum ranges of predicted electricity consumption are obtained when using mixed uncertainty. Accordingly, the application of uncertainty analysis using a probability distribution function greatly improved the predictive power of the analysis compared to time-series analysis.

Keywords: Machine learning; predictive modeling; time-series analysis; uncertainty analysis; COVID-19

1 Introduction

Prediction is a statement regarding what can be expected to occur in the future. Therefore, it suffers from uncertainty, and probabilistic and statistical tools involving big data, data science, and machine learning are necessary components of any scientific approach seeking to formalize the prediction process [1–7]. Despite the complexity of the process, accurate predictions are essential for supporting a wide range of human activities. For example, predictive modeling applied to the



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

coronavirus 2019 (COVID-19) outbreak can facilitate better patient care, such as by predicting intensive care unit requirements, evaluating patient survival potentials, and analyzing patient trajectories during treatment [8].

Time-series analysis is an essential aspect of the prediction process because many prediction problems have a time component. This process based on time-series analysis typically seeks to predict the future values of observed time-series data using a multivariate regression model with estimated and expected regression parameters [9]. A variety of models exist. The best-known class of models is autoregressive moving average (ARMA) and autoregressive integrated moving average (ARIMA) models for single time-series data. Multivariate ARIMA models and vector auto-regression models are also popular. However, a comparative study demonstrated that the Facebook Prophet algorithm has better prediction results than ARIMA [10]. Facebook Prophet is an open-source library that fits non-linear trends for time-series data having yearly, weekly, and daily seasonality, and also includes the impacts of non-seasonal events such as holidays.

Unfortunately, unpredictable phenomena, such as extreme weather events or the COVID-19 outbreak, can greatly limit the ability of time-series analyses to establish reliable patterns. This is because the same information can produce different outputs from the same model owing to the presence of uncertainty. However, introducing uncertainty into deterministic models is difficult.

Machine learning techniques and deep learning algorithms can make predictions by learning the inherent patterns within data, and therefore present new approaches to prediction by modeling relationships between variables in a deep and layered hierarchy. For example, long short-term memory has generated considerable attention, with applications in many disciplines [11–19]. These characteristics make machine learning an ideal solution to prediction problems that involve big datasets, large numbers of predictors, and different types and sources of data, including free-text notes [20–28]. Predictive modeling is classified mainly into theoretical modeling based on causality and the reduced-form approach based on correlation. However, the lack of any discernible causality in many prediction processes, and the increasing abundance of data and computational capability have facilitated the wide use of the reduced-form approach in predictive modeling research. In any case, predictive modeling requires a good understanding of the data and objective of the proposed application for data preprocessing, model generation, and evaluation [29].

Predictive modeling efforts must address two kinds of uncertainty, including aleatory and epistemic uncertainty, where aleatory uncertainty represents the inherent uncertainty in a completely random process, and epistemic uncertainty derives from ignorance or a lack complete information regarding the behavior of a process [30]. Many deep learning methods, including Bayesian and non-Bayesian methods, have been proposed to quantify predictive uncertainty. Here, Bayesian inference makes predictions using prior knowledge, and a probabilistic programming language based on this can express a model's randomness. Python, which supports a probabilistic programming language (PPL), is open-source, and it can generate scalable and efficient Bayesian machine learning models. This makes PPL ideal for modeling uncertainty. However, while a large-scale benchmark study of existing state-of-the-art deep learning strategies applied to classification problems and an investigation of the effect of dataset shift on accuracy and calibration has been conducted [31], no rigorous large-scale empirical comparison has yet been applied to these methods. Self-supervised learning represents another method by which model robustness to uncertainty can be improved. The predictive performance of this method exceeds the performance of fully supervised methods because it enhances out-of-distribution detection on difficult, near-distribution outliers [32].

The above discussion indicates that a probabilistic approach can transcend the limitations of time-series analysis, while contributing to the robustness of predictive models to uncertainty. The present work addresses this issue by applying uncertainty analysis using a probability distribution function, and applies the proposed scheme within a preliminary study involving the prediction of power consumption for a single hotel in Seoul, South Korea. This application of uncertainty analysis is quite topical because both increasing population and economic growth worldwide have greatly increased the share of the total energy consumption taken by commercial space, such that its prediction is increasingly important for the purpose of reducing energy consumption [33–36]. Moreover, data-driven approaches are the most advanced methods employed for electric energy consumption prediction (EECP) applications, which have been applied as a deep learning approach to intelligent power management systems, and plays an important role in national energy development policy [37–39].

First, time-series analysis is conducted based on 53,567 data items recorded from March 1, 2019 to September 8, 2020 for the hotel building over 558 days on the Korea Electric Power Corporation (KEPCO) website, which records electricity usage every 15 min using robotic process automation (RPA). We apply Facebook Prophet for the time-series analysis, and the results clearly demonstrate that the COVID-19 outbreak seriously compromised the reliability of the time-series analysis.

We then develop machine learning models in the TensorFlow framework for conducting uncertainty analyses based on modeled relationships between electric power consumption and outdoor temperature. We begin with a simple linear regression model, as the most basic machine learning algorithm, for predicting the electric power consumption of the hotel building with respect to outdoor temperature. In such a model, the value of one variable varies in proportion to that of another. Obtaining improved prediction requires a representation of the variation inherent to the underlying process, which is aleatory uncertainty. The remaining uncertainty in the prediction process involves known unknowns, which represents epistemic uncertainty due to a lack of knowledge [40]. A final method of uncertainty analysis involves the case of both known and unknown unknowns, which is mixed aleatory and epistemic uncertainty. The benefits of the proposed uncertainty analysis for predicting electric power consumption are demonstrated by comparing the results obtained when considering no uncertainty (i.e., the linear regression model), aleatory uncertainty, epistemic uncertainty, and mixed aleatory and epistemic uncertainty. The results indicate that the electricity consumption of the hotel cannot be precisely predicted using the linear regression model because the model ignores uncertainty. In contrast, the minimum and maximum ranges of the predicted power consumption are obtained when using mixed uncertainty.

2 Time-series Analysis

The electric power consumption data of the hotel building collected over the 558 days from March 1, 2019 to September 8, 2020 is presented as black dots in Fig. 1.

The prediction results of the electric power consumption obtained by Prophet from March 1, 2020 September 1, 2020 are given by the blue line, while the sky blue area corresponds to the upper and lower limits of the predictions. The data points circled in red and dark blue respectively represent the days with the highest and lowest electricity consumption in 2019 and 2020. The differences between the two highest points and the two lowest points are marked by red and dark blue dotted lines, respectively. Power consumption is seen to have decreased significantly after March 1, 2020 with the advent of the COVID-19 outbreak. This represents a change that could not be predicted. Accordingly, time-series analysis can no longer support predictive modeling.

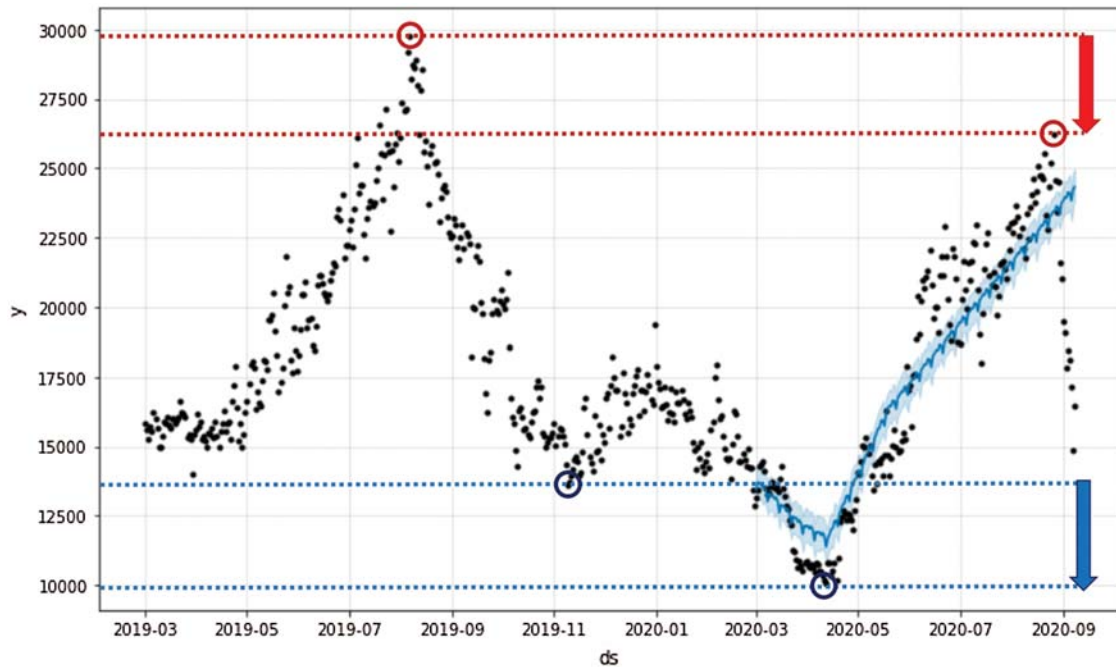


Figure 1: Predictive modeling of hotel electric power consumption for the year 2020

The Facebook Prophet time-series analysis decomposition results are presented in Fig. 2, which include the overall trend, and weekly, yearly, and daily variations. Here, the trend results represent the declining electric power demand observed in Fig. 1, which provided predictions that differed from the actual electric power usage. Accordingly, a general time-series analysis is not suitable under these conditions. Nevertheless, the weekly, yearly, and daily variations represent meaningful results. According to the weekly prediction in Fig. 2, electricity consumption on Tuesday and Friday is high, and it is relatively low on Thursday. The yearly analysis in Fig. 2 correctly predicts heat-related power consumption in August.

3 Methodology

3.1 Dataset

Daily recorded outdoor temperature data were used in conjunction with the electric power consumption data presented in Fig. 2. The quantile statistics of daily electricity consumption are listed in Tab. 1. We note that the maximum value is nearly twice the minimum.

3.2 Normalization

The electricity consumption data cannot be applied directly to a probability distribution function generated with respect to temperature without normalization. Therefore, we normalize the power consumption data as follows:

$$\text{Normalized dataset} = (\text{value} - \min(\text{dataset})) / (\max(\text{dataset}) - \min(\text{dataset})). \quad (1)$$

The normalized electricity consumption dataset employed for model training is plotted as the variable y on a scale of 0 to 1 with respect to the temperature T ($^{\circ}\text{C}$) in Fig. 3.

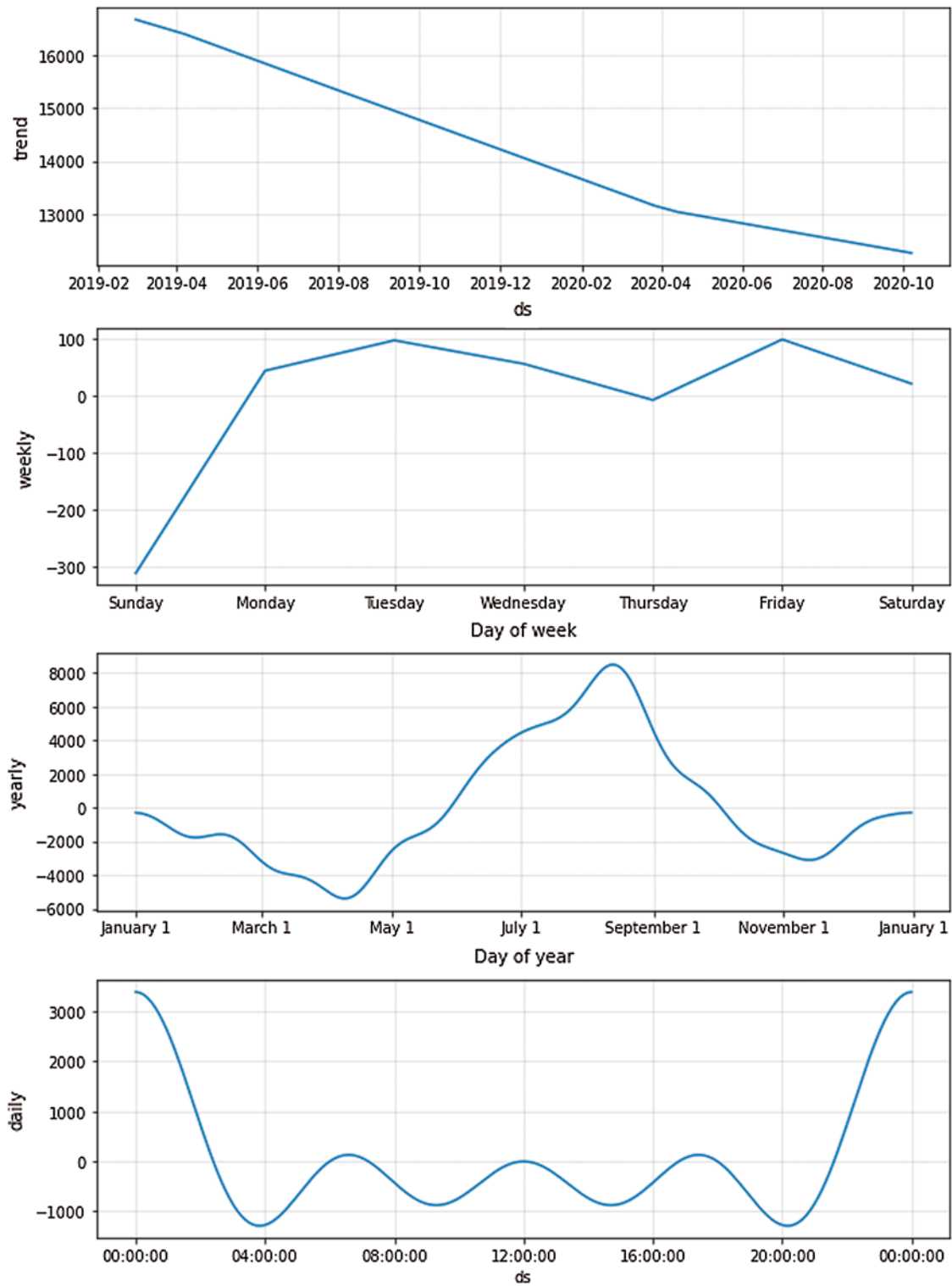


Figure 2: Facebook Prophet time-series analysis decomposition

Table 1: Quantile statistics of daily electricity consumption (unit: kW)

Minimum	14019.12
5th percentile	15224.736
25th percentile	16023.36
Median	19301.28
75th percentile	22804.32
95th percentile	26976.768
Maximum	29781.12

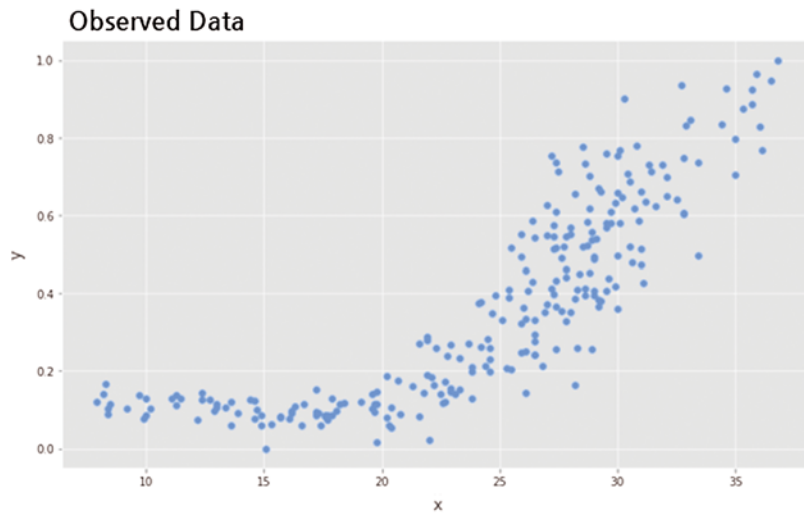


Figure 3: Normalized observed electricity consumption data with respect to temperature T

3.3 Prediction Methods

The four prediction methods are illustrated in Fig. 4. The mixed model was obtained by integrating the respective models accounting for aleatory and epistemic uncertainties, and the data distribution is modeled by adding a variation analysis layer to the previous model.

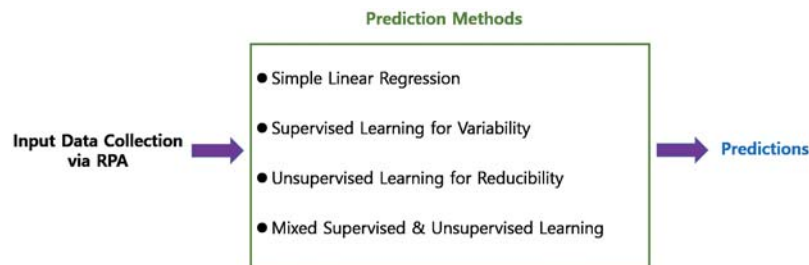


Figure 4: Prediction methods

The machine learning models were developed using TensorFlow probability (TFP) layers to manage the uncertainty inherent in regression predictions, and probabilistic layers in TFP with the Keras application programming interface (API), which is a high-level API for TensorFlow, to build the other models on that simple foundation.

The output of the linear regression model is a normal distribution with constant variance. The output of the second prediction model is a normal distribution whose mean and variance depend on the input. For the third prediction method, the posterior and prior were trained using the Keras layer, and the model was built by inference. For the last prediction method, the model was created and inferred. Then, several ensemble means and ensemble standard derivations were applied to obtain plots for various prediction lines indicative of the prediction results.

4 Results

4.1 No Uncertainty

The single line in Fig. 5 represents the overall trend of the predicted mean, and therefore does not account for uncertainty.

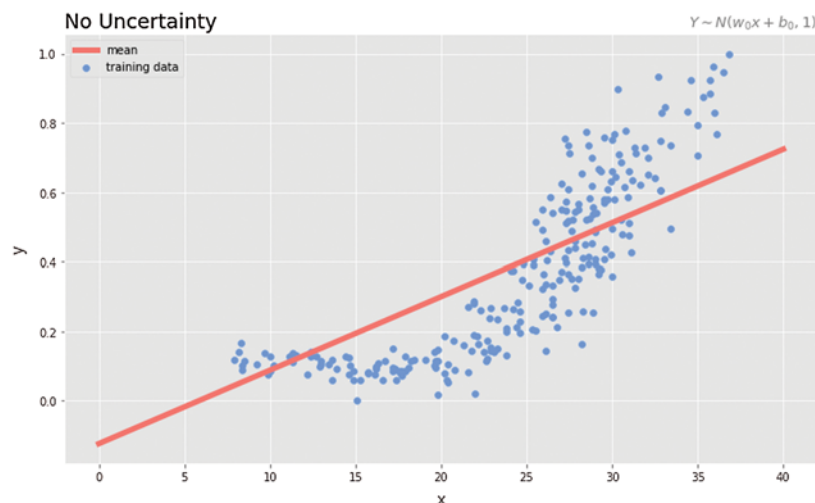


Figure 5: Overall trend of the predicted mean of the distribution

4.2 Aleatory Uncertainty

The results obtained by the supervised learning method accounting for aleatory uncertainty are presented in Fig. 6. Here, the overall trend in the predicted mean is plotted along with the standard deviation of the distribution. After training, the model provides meaningful predictions regarding the variability of y as a function of T , making it possible to produce a range of predictions indicative of aleatory uncertainty, rather than a simple line indicative of only the predicted mean value.

4.3 Epistemic Uncertainty

The results of the unsupervised learning method accounting for epistemic uncertainty are presented in Fig. 7. Here, the 20 red lines in the figure represent 20 guesses by the unsupervised model regarding the linear relationship between y and T , which are different each time because

the generated model resamples the data according to weighting imposed by the posterior distribution. Hence, we have presented 20 predictions to understand how the weighting imposed by the nobreakposterior distribution affects the final prediction. The epistemic uncertainty is reflected in the different slopes of the lines, which represent an increasing uncertainty in y with increasing T . Accordingly, accurate predictions are quite difficult to obtain without introducing prior knowledge.

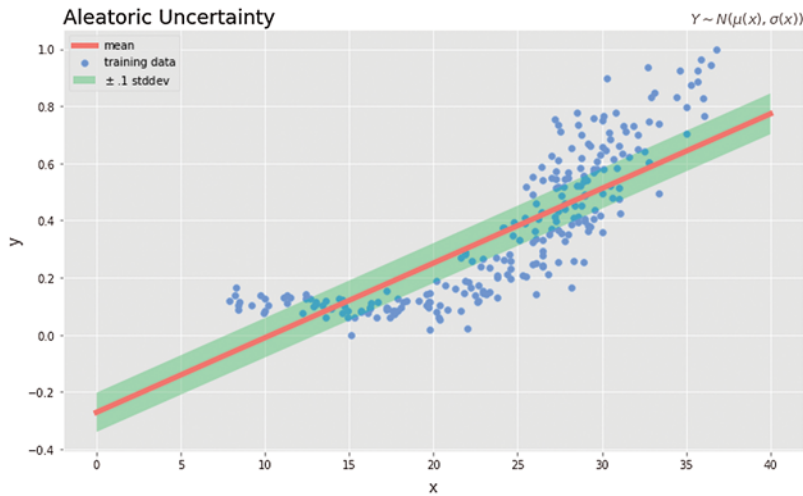


Figure 6: Variation inherent to the underlying prediction process

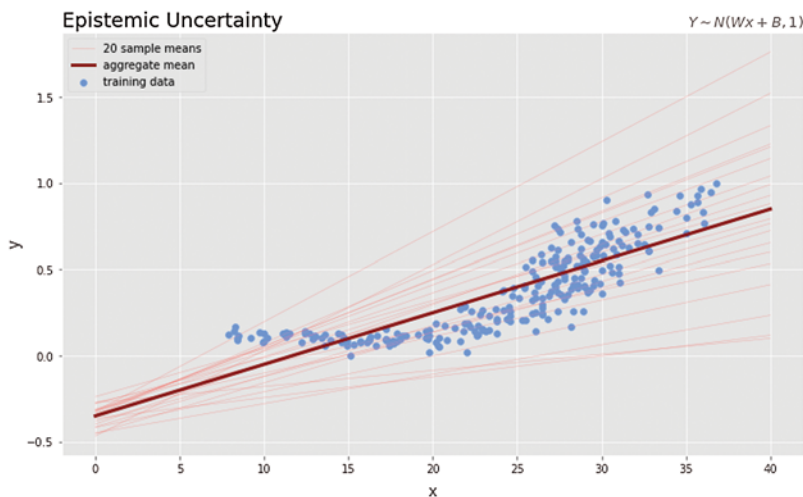


Figure 7: Ensemble means with overall mean for representing epistemic uncertainty

4.4 Mixed Aleatory and Epistemic Uncertainty

The results of the mixed supervised and unsupervised learning method that accounts for both aleatory and epistemic uncertainties are presented in Figs. 8 and 9, which represent the results obtained with three sample means and four sample means, respectively. The plots of the figures

on the left sides present predictions based on the slopes of the observed data over different x -axis data ranges, and the plots of the figures on the right sides reflect the application of the minimum and maximum values of the observed data.

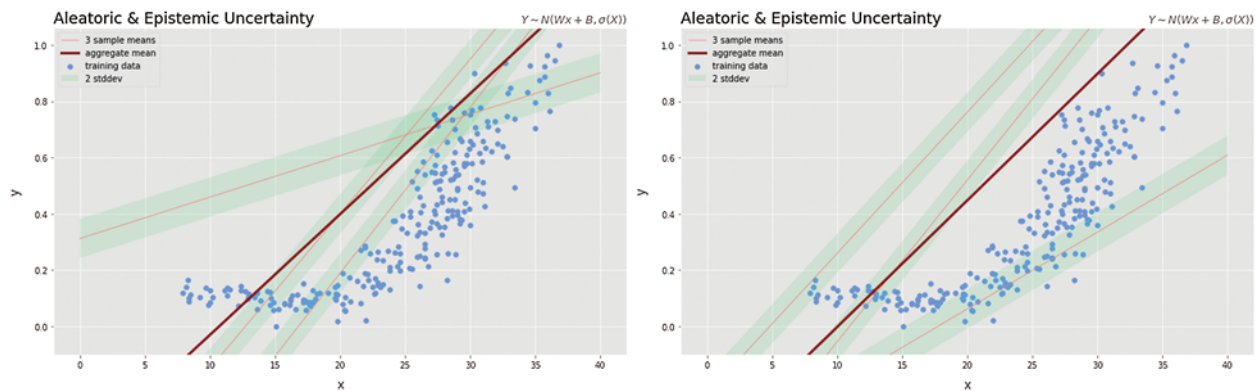


Figure 8: Ensemble models applying the variability of y as a function of x with three sample means

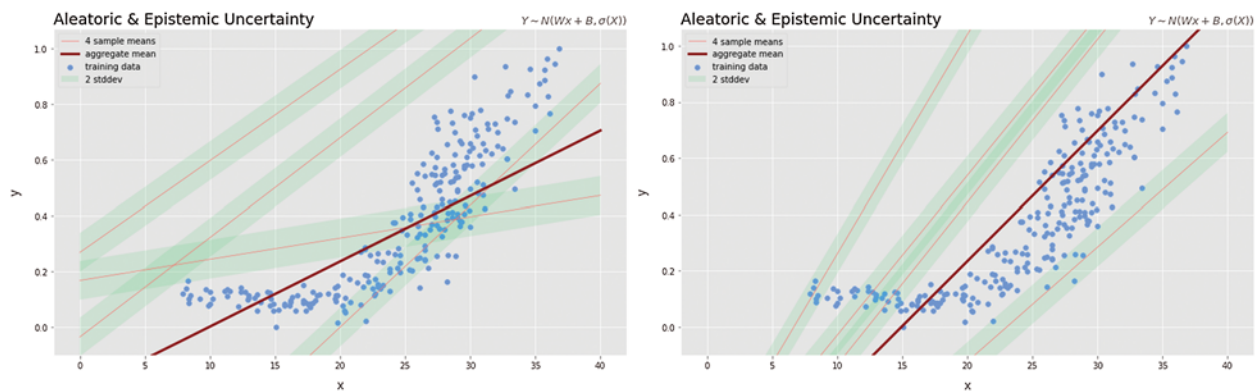


Figure 9: Ensemble models applying the variability of y as a function of x with four sample means

These results confirm the limitations of the simple linear regression analysis. Moreover, it can be seen that the results in Figs. 8 and 9 represent less uncertainty over that presented in Fig. 6 for the model considering only aleatory uncertainty and that presented in Fig. 8 for the model considering only epistemic uncertainty because they cover the wider area.

5 Conclusion

The present study demonstrated that a probabilistic approach can transcend the limitations of time-series analysis by applying uncertainty analysis using a probability distribution function to the prediction of power consumption for a single hotel in Seoul, South Korea. The results confirmed that the time-series analysis was insufficiently reliable due to uncertainty, such as the COVID-19 outbreak that could not be predicted. The application of models accounting

for aleatory uncertainty, epistemic uncertainty, and both aleatory and epistemic uncertainties demonstrated that electricity consumption can be predicted within a specific range according to the outdoor temperature. The results further demonstrated that probabilistic programming languages such as TensorFlow probability can provide a framework for accounting for aleatory and epistemic uncertainties, and can hasten the solution of complex probabilistic models. The output of algorithms trained on historical data can be applied to new data to make further predictions. This technology facilitates the analysis of big data, enabling the application of related research to real life and not just theoretical data.

Acknowledgement: The authors thank Dr. Joonsoo Jeong at Hansung University for his helpful advice. The authors are also grateful to the reviewers.

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors have no conflicts of interest to declare regarding the present study.

References

- [1] H. V. Roberts, "Probabilistic prediction," *Journal of the American Statistical Association*, vol. 60, no. 309, pp. 50–62, 1965.
- [2] J. Holtrop and G. G. J. Mennen, "A statistical power prediction method," *International Shipbuilding Progress*, vol. 25, no. 290, pp. 253–256, 1978.
- [3] J. Aitchison and I. R. Dunsmore, "Informative prediction," in *Statistical Prediction Analysis*. Cambridge, England: CUP Archive, pp. 68–87, 1980.
- [4] J. Holtrop and G. G. J. Mennen, "An approximate power prediction method," *International Shipbuilding Progress*, vol. 29, no. 335, pp. 166–170, 1982.
- [5] M. E. Miller, C. D. Langefeld, W. M. Tierney, S. L. Hui and C. J. McDonald, "Validation of probabilistic predictions," *Medical Decision Making*, vol. 13, no. 1, pp. 49–57, 1993.
- [6] J. Wiest, M. Höffken, U. Kreßel and K. Dietmayer, "Probabilistic trajectory prediction with gaussian mixture models," in *2012 IEEE Intelligent Vehicles Symp.*, Alcalá de Henares, Spain, pp. 141–146, 2012.
- [7] T. Gneiting and M. Katzfuss, "Probabilistic forecasting," *Annual Review of Statistics and Its Application*, vol. 1, no. 1, pp. 125–151, 2014.
- [8] J. P. Cohen, P. Morrison, L. Dao, K. Roth, T. Q. Duong *et al.*, "Covid-19 image data collection: Prospective predictions are the future," *arXiv Preprint*, vol. 2006.11988, pp. 1–25, 2020.
- [9] F. Stulajter, "Predictions of time series," in *Predictions in Time Series using Regression Models*. Berlin, Germany: Springer Science & Business Media, pp. 147–196, 2013.
- [10] C. Rotunâ, A. Cohal, I. Sandu and M. Dumitrache, "New tendencies in linear prediction of events," *Romanian Journal of Information Technology and Automatic Control*, vol. 29, no. 3, pp. 19–30, 2019.
- [11] S. Siami-Namini, N. Tavakoli and A. S. Namin, "A comparison of ARIMA and LSTM in forecasting time series," *2018 17th IEEE Int. Conf. on Machine Learning and Applications*, vol. 17, pp. 1394–1401, 2018.
- [12] J. Navarro-Moreno, "ARMA prediction of widely linear systems by using the innovations algorithm," *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 3061–3068, 2008.
- [13] I. Rojas, O. Valenzuela, F. Rojas, A. Guillen, L. J. Herrera *et al.*, "Soft-computing techniques and ARMA model for time series prediction," *Neurocomputing*, vol. 71, no. 4–6, pp. 519–537, 2008.
- [14] D. B. Cline and P. J. Brockwell, "Linear prediction of ARMA processes with infinite variance," *Stochastic Processes and Their Applications*, vol. 19, no. 2, pp. 281–296, 1985.
- [15] R. A. Davis and S. I. Resnick, "Basic properties and prediction of max-ARMA processes," *Advances in Applied Probability*, vol. 21, no. 4, pp. 781–803, 1989.

- [16] S. L. Ho, M. Xie and T. N. Goh, "A comparative study of neural network and Box-Jenkins ARIMA modeling in time series prediction," *Computers & Industrial Engineering*, vol. 42, no. 2–4, pp. 371–375, 2002.
- [17] R. Kohn and C. F. Ansley, "Estimation, prediction, and interpolation for ARIMA models with missing data," *Journal of the American Statistical Association*, vol. 81, no. 395, pp. 751–761, 1986.
- [18] P. L. Fernandez Jr and J. M. Co, "Improving the vector auto-regression technique for time-series link prediction by using support vector machine," *MATEC Web of Conf.*, vol. 56, no. 1008, pp. 1–5, 2016.
- [19] F. A. Gers, J. Schmidhuber and F. Cummins, "Learning to forget: Continual prediction with LSTM," *9th Int. Conf. on Artificial Neural Networks*, vol. 9, pp. 850–855, 1999.
- [20] N. Karunanithi, D. Whitley and Y. K. Malaiya, "Using neural networks in reliability prediction," *IEEE Software*, vol. 9, no. 4, pp. 53–59, 1992.
- [21] S. M. Weiss and N. Indurkha, "Rule-based machine learning methods for functional prediction," *Journal of Artificial Intelligence Research*, vol. 3, pp. 383–403, 1995.
- [22] C. Mair, G. Kadoda, M. Lefley, K. Phalp, C. Schofield *et al.*, "An investigation of machine learning based prediction systems," *Journal of Systems and Software*, vol. 53, no. 1, pp. 23–29, 2000.
- [23] D. L. Shrestha and D. P. Solomatine, "Machine learning approaches for estimation of prediction interval for the model output," *Neural Networks*, vol. 19, no. 2, pp. 225–235, 2006.
- [24] V. U. B. Challagulla, F. B. Bastani, I. L. Yen and R. A. Paul, "Empirical assessment of machine learning based software defect prediction techniques," *International Journal on Artificial Intelligence Tools*, vol. 17, no. 2, pp. 389–400, 2008.
- [25] A. Khosravi, S. Nahavandi, D. Creighton and A. F. Atiya, "Comprehensive review of neural network-based prediction intervals and new advances," *IEEE Transactions on Neural Networks*, vol. 22, no. 9, pp. 1341–1356, 2011.
- [26] A. Mackenzie, "The production of prediction: What does machine learning want?," *European Journal of Cultural Studies*, vol. 18, no. 4, 5, pp. 429–445, 2015.
- [27] G. S. Collins and K. G. Moons, "Reporting of artificial intelligence prediction models," *Lancet*, vol. 393, no. 1, pp. 1577–1579, 2019.
- [28] M. Kuhn and K. Johnson, *Applied Predictive Modeling*, vol. 26. New York: Springer, 2013.
- [29] M. Kuhn and K. Johnson, "A short tour of the predictive modeling process," in *Applied Predictive Modeling*, vol. 26. New York: Springer, pp. 36–64, 2013.
- [30] Y. H. Hou, Y. J. Li and X. Liang, "Mixed aleatory/epistemic uncertainty analysis and optimization for minimum EEDI hull form design," *Ocean Engineering*, vol. 172, no. 1, pp. 308–315, 2019.
- [31] O. Yaniv, F. Emily, R. Jie, N. Zachary, D. Sculley *et al.*, "Can you trust your model's uncertainty? Evaluating predictive uncertainty under dataset shift," *Advances in Neural Information Processing Systems*, vol. 32, pp. 13991–14002, 2019.
- [32] D. Hendrycks, M. Mazeika, S. Kadavath and D. Song, "Using self-supervised learning can improve model robustness and uncertainty," *Advances in Neural Information Processing Systems*, vol. 32, pp. 15663–15674, 2019.
- [33] C. Robinson, B. Dilkina, J. Hubbs, W. Zhang, S. Guhathakurta *et al.*, "Machine learning approaches for estimating commercial building energy consumption," *Applied Energy*, vol. 208, no. 9, pp. 889–904, 2017.
- [34] C. Li, Z. Ding, D. Zhao, J. Yi and G. Zhang, "Building energy consumption prediction: An extreme deep learning approach," *Energies*, vol. 10, no. 1525, pp. 1–20, 2017.
- [35] K. Amasyali and N. M. El-Gohary, "A review of data-driven building energy consumption prediction studies," *Renewable and Sustainable Energy Reviews*, vol. 81, pp. 1192–1205, 2018.
- [36] Y. Wei, X. Zhang, Y. Shi, L. Xia, S. Pan *et al.*, "A review of data-driven approaches for prediction and classification of building energy consumption," *Renewable and Sustainable Energy Reviews*, vol. 82, no. 3, pp. 1027–1047, 2018.
- [37] F. Wahid, L. H. Ismail, R. Ghazali and M. M. Aamir, "An efficient artificial intelligence hybrid approach for energy management in intelligent buildings," *Korean Society for Internet Information (KSII) Transactions on Internet & Information Systems*, vol. 13, no. 12, pp. 5904–5927, 2019.

- [38] H. Zhong, J. Wang, H. Jia, Y. Mu and S. Lv, "Vector field-based support vector regression for building energy consumption prediction," *Applied Energy*, vol. 242, no. 4, pp. 403–414, 2019.
- [39] T. Le, M. T. Vo, B. Vo, E. Hwang, S. Rho *et al.*, "Improving electric energy consumption prediction using CNN and Bi-LSTM," *Applied Sciences*, vol. 9, no. 20, pp. 4237, 2019.
- [40] E. Hofer, M. Kloos, B. Krzykacz-Hausmann, J. Peschke and M. Woltereck, "An approximate epistemic uncertainty analysis approach in the presence of epistemic and aleatory uncertainties," *Reliability Engineering & System Safety*, vol. 77, no. 3, pp. 229–238, 2002.