IOP Conference Series: Earth and Environmental Science

PAPER • OPEN ACCESS

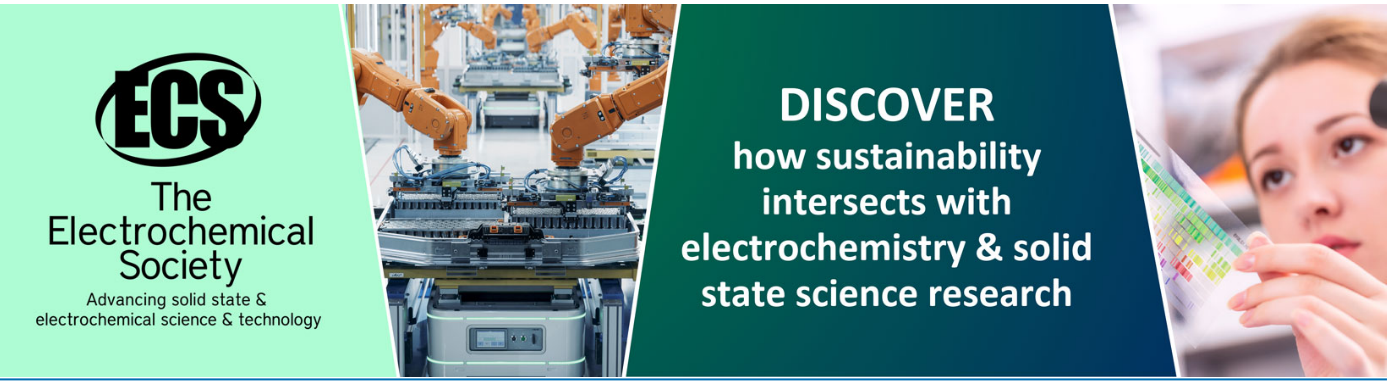
Construction of Forecast Model for Power Demand and PV Power Generation Using Tensor Product Spline Function

To cite this article: T Matsumoto and Y Yamada 2021 *IOP Conf. Ser.: Earth Environ. Sci.* **812** 012001

View the article online for updates and enhancements.

You may also like

- What drives historical and future changes in photovoltaic power production from the perspective of global warming?
Robert Constantin Scheele and Stephanie Fiedler
- The effect of particulate matter on solar photovoltaic power generation over the Republic of Korea
Junghoon Son, Sujong Jeong, Hayoung Park et al.
- A prediction model based on an adaptive genetic algorithm and optimized Elman neural network for PV power
Guanfu Wang, Difei Gao, Yu Jiang et al.



This content was downloaded from IP address 3.145.97.248 on 26/04/2024 at 12:57

IOP Conf. Series: Earth and Environmental Science **812** (2021) 012001 doi:10.1088/1755-1315/812/1/012001

Construction of Forecast Model for Power Demand and PV Power Generation Using Tensor Product Spline Function

T Matsumoto[1], Y Yamada[2]

[1] Socio-economic Research Center, Central Research Institute of Electric Power Industry, Japan, matsumoto3832@criepi.denken.or.jp, 1-6-1 Otemachi, Chiyoda-ku, Tokyo 100-8126, Japan
[2] Faculty of Business Sciences, University of Tsukuba, Tokyo, Japan, yuji@gssm.otsuka.tsukuba.ac.jp, 3-29-1 Otsuka, Bunkyo-ku, Tokyo 112-0012, Japan

Abstract. Forecasting power demand and photovoltaic (PV) power generation is indispensable for the economic operation of electric utility businesses. Herein, robust demand and PV power forecast models are proposed that enable electricity companies including new entrants to efficiently forecast from small sample size data by using only published weather forecasts. We further enhance the previously proposed forecast models by using the tensor product spline function to impose smoothing conditions, not only in the seasonal direction but also in the hour direction, thereby constructing effective models that can incorporate multiple explanatory variables. The empirical results of the estimated two-dimensional trends are consistent with the intuitive interpretation, and the validations of the out-of-sample forecast error obtained from the data of nine different areas ensure the high robustness of the proposed model.

1. Introduction

The Japanese electricity market was fully liberalised in 2016, and many retailers have entered the market. Subsequently, the trading volume of the Japan Electric Power Exchange (JEPX) has continued to increase significantly. Electricity utilities, including retailers and power producers, generally forecast the next day's power demand/generation volume on a daily basis, and procure/sell the forecasted volume in the day-ahead market (spot market in JEPX). The final volume of supply-demand gap (imbalance) incurred due to the forecast error is settled by an imbalance price (that is higher if in shortage imbalance, or lower if in surplus imbalance, than the spot price); hence, so the forecast error is directly linked to the financial loss. In addition, by forecasting the entire system imbalance and the imbalance price linked to it, the power procurement/sales operations can potentially become more strategic and profitable. Therefore, forecasting the demand for and photovoltaic (PV) power generation in a wide area also plays an important role in bringing economic benefits to the electricity business. This study proposes robust demand and PV power forecast models that enable electricity companies including new entrants to efficiently construct from small sample size data, for example, one or two years.
There are various previous studies dealing with PV and demand forecasting methods (see survey studies such as [1] for PV and [2] for demand), and especially in recent years, nonlinear prediction methods using Artificial Intelligence (AI) are the mainstream. However, most methods using AI have problems such as high computational load and difficulty in interpretation (e.g., see [3]). For this reason, this study aims for constructing models that are easy for practitioners to implement and understand.

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Published under licence by IOP Publishing Ltd

IOP Conf. Series: Earth and Environmental Science **812** (2021) 012001 doi:10.1088/1755-1315/812/1/012001

Previous studies have proposed forecast models with the same purpose, including for PV power generation [3] and for demand [4]. These both estimate the smooth yearly cyclical trends using the additive sum of smooth functions, namely generalised additive models (GAMs) [5] by separating the forecast target time series into each hourly clock time. In each case, however, only one type of explanatory variable is used for forecasting, which is either general weather conditions ([3] or temperature [4]. There is room for improvement in terms of the effective use of other available weather forecast data. Here, we consider enhancing these previous forecast models by using both temperature and general weather condition variables. To ensure robustness, we use the tensor product spline function to impose smoothing conditions, not only in the seasonal direction but also in the hour direction (as with the approach used for the hedge model in [6]), thereby constructing effective models that can accommodate the incorporation of multiple explanatory variables.
This paper is structured as follows: In Section 2, we construct each model for demand and PV power generation, and in Section 3, we provide the demonstration result using empirical data from the Japanese market. Our conclusions are presented in Section 4.

2. Model construction

For the model construction, we build forecast models for PV power generation and demand. The main variables used in these models are defined as follows:

- $V_{t,h}$[MW]: measured PV power generation volume at date t, hour h
- W_t[MW]: installed PV power capacity at date t, hour h
- $U_{t,\,h}$[MW/MW=1]: unit power generation at date t, hour h
- $D_{t,\,h}$[MW]: measured power demand at date t, hour h
- $I_{\cdot,t,h}$: dummy variables, which are 1 if the forecasted general weather condition at date t hour h is the same as the suffix's weather condition (or day of the week), or 0 otherwise
- $Tmax_t$, $Tmin_t$[°C]: previous day's maximum or minimum temperature forecast at date t
- $\epsilon_{tmax,t}$, $\epsilon_{tmin,t}$[°C]: maximum or minimum temperature forecast deviation at date t (observed temperature forecast minus its trend $f_\cdot(t)$)
- $u_\cdot(t,h)$: tensor product spline functions estimated by the GAM of the PV power forecast model
- $d_\cdot(t,h)$: tensor product spline functions estimated by the GAM of the demand forecast model
- $f_\cdot(t)$: univariate spline functions estimated by the GAM of the temperature trend model
- $\eta_{\cdot,\,t,\,h}$: residual terms with the average of 0

The procedure for calculating the forecast value of PV power generation is as follows:

1. Using the actual PV power generation capacity W_t (observed no more frequently than monthly), the daily increasing trend is estimated by a linear model. As a result, the daily forecast value of PV power generation capacity is obtained as $\widehat{W}_t$ (see Appendix A).
2. Obtain unit power generation $U_{t,\,h}$ by dividing the measured hourly PV power generation $V_{t,h}$ by $\widehat{W}_t$ (i.e., $U_{t,\,h} := V_{t,h}/\widehat{W}_t$).
3. For $U_{t,\,h}$, build a GAM with calendar information, general weather conditions forecast, and maximum/minimum temperature forecast as explanatory variables.
4. The out-of-sample forecasted value $\widehat{U}_{t,\,h}$ is obtained by substituting the explanatory variables observed into the estimated GAM forecast formula (predictor part of the GAM).
5. The out-of-sample forecasted value $\widehat{V}_{t,\,h}$ is obtained as the product of the forecasted PV power capacity $\widehat{W}_t$ and the forecasted unit power generation $\widehat{U}_{t,\,h}$.

In contrast, the forecasting procedure for power demand, which does not have a remarkable increasing trend like PV power generation, includes only 3 and 4 above (note that U should be replaced for D). A slight decreasing trend due to recent efforts to save energy is expected in power demand, and we will respond to it by directly incorporating a linear trend term into the GAM (described in Section 2.2).

In the following sub-sections, GAMs using the tensor product spline function are formulated for both forecast models.

2.1. Forecast model for PV power generation

To develop the PV power generation forecast model, we construct the following GAM for unit power generation $U_{t,\,h}$, obtained by dividing the observed PV power generation by the daily installed capacity predicted by the linear trend:

$$\begin{aligned} U_{t,\,h} = u_{sunny}(t,h)I_{sunny,t,h} + u_{cloudy}(t,h)I_{cloudy,t,h} + u_{rainy}(t,h)I_{rainy,t,h} \\ + u_{snowy}(t,h)I_{snowy,t,h} + u_{tmax}(t,h)\epsilon_{tmax,t} + u_{tmin}(t,h)\epsilon_{tmin,t} \\ + \eta_{u,\,t,\,h} \end{aligned} \tag{1}$$

where $u.(t,h)$ is the tensor product spline function estimated by the GAM (1), which is the two-dimensional time trend that smoothly connects in the direction of both the date t and hour h (see Appendix B for an overview of the tensor product spline function and the smoothing mechanism, and note that the "snowy" term is only included in the model for snowy areas[1]). The tensor product spline function imposes smoothing conditions in two orthogonal directions, and it is therefore, more robust than the previous studies that separately modelled target time series by each hour, which considered the smoothing condition only in the date direction. It is possible to incorporate multiple explanatory variables even with a small sample size.

Note that the maximum and minimum forecast deviations $\epsilon_{tmax,\,t}$ and $\epsilon_{tmin,\,t}$ are obtained by the following GAM:

$$Tmax_t = f_{tmax}(t) + \epsilon_{tmax,\,t}, Tmin_t = f_{tmin}(t) + \epsilon_{tmin,\,t}. \tag{2}$$

We estimate the yearly cyclical trends as the spline functions $f.(Seasonal_t)$ in (2) and $u.(Seasonal_t, h)$ in (1) using yearly cyclical dummy variables $Seasonal_t$ $(= 1, \dots, 365\ (or\ 366))$, whose allocation method is proposed in [7]. In this work, the starting point of the cyclical dummy variables is 1st January, and from 1–365 (366 for leap years) are allocated in order. To make the notation more concise, we denote $f.(Seasonal_t)$ and $u.(Seasonal_t, h)$ as $f.(t)$ and $u.(t,h)$.

2.2. Forecast model for demand

To develop the demand forecast model, we construct the following GAM:

$$\begin{aligned} D_{t,\,h} = d(t,h) + d_{saturday}(t,h)I_{saturday,t,h} + d_{sunday}(t,h)I_{sunday,t,h} \\ + d_{tmax}(t,h)\epsilon_{tmax,t} + d_{tmin}(t,h)\epsilon_{tmin,t} + d_{sunny}(t,h)I_{sunny,t,h} \\ + d_{rainy}(t,h)I_{rainy,t,h} + d_{snowy}I_{snowy,t,h} + \beta \times Period_t + \eta_{d,\,t,\,h} \end{aligned} \tag{3}$$

where $d.(t,h)$ and β are the tensor product spline function and fixed coefficient estimated by the GAM (3), respectively. In this model, the first term $d(t,h)$ represents the cloudy weekday's trend and β represents the increasing (decreasing if negative) trend per year (see Appendix B for the date dummy $Period_t$). Note that the "snowy" term is included only in snowy areas, and its coefficient is defined as the fixed value, considering robustness. In addition, $I_{sunny,t,h}$ and $\epsilon_{tmax,t,h}$ are set to 0 whenever hour

[1] In this study, the general weather conditions are roughly divided into the four categories: sunny, cloudy, rainy, and snowy. In practice, there are various weathers such as "Fog" and "Hail" depending on the area, but this paper classifies the former as rainy and the latter as snowy. The refinement of such classification is a future task.

h is between 0 and 6 because they are not expected to affect the demand in such a time zone.

3. Empirical analysis

In this section, we use empirical data to estimate smooth trends in each model using in-sample period data from 1st April 2016 to 31st December 2017, and we verify the forecast errors using out-of-sample data from 1st January 2018 to 31st December 2018 in nine different power areas.

3.1. Empirical result for PV power generation forecast

For the forecast of PV power generation, the following observed data are used:

- PV power generation volume $V_{t,h}$ [MW]: published by nine electricity power companies[2]
- PV power capacity W_t [MW]: month-end results published by the Ministry of Economy, Trade and Industry (METI)[3]
- Weather condition dummy $I_{\cdot,t,h}$, max (min) temperature $Tmax_t$ ($Tmin_t$) [°C]: forecast values announced by the Japan Meteorological Agency (JMA) on the previous morning (one major city in each of the nine areas)[4]

Figure 1 shows the estimated trends for the Tokyo area. It can be confirmed that power generation is greater in the order of sunny, cloudy, and rainy weather. The estimated trend of sunny weather declines in the summer, which reflects the knowledge that PV power generation decreases in efficiency during summer due to high temperatures. The significant decline in the rainy trend in winter is consistent with extreme darkening because the weather tends to change into sleet or snow. While the maximum temperature contributes to increasing power generation, the minimum temperature contributes to decreasing power generation, and this could be interpreted as under a fixed maximum temperature, the lower the minimum temperature (usually recorded around early dawn), the larger the solar radiation during the day should be (to raise the temperature). For this reason, there is a negative correlation between minimum temperature and PV power generation.

We conducted a forecast error analysis using the out-of-sample period data. Figure 2 summarises the forecast errors of PV power generation in each of the nine areas. The R-squared statistic does not show any noticeable decline in the out-of-sample period, which infers that the model has been built robustly (see e.g. [8] for out-of-sample R-squared statistic). The same applies for the mean absolute error (MAE). The slightly high MAE of Kansai was caused by the downwards bias in the capacity trend estimation because the introduction of PV power was more advanced than expected. Furthermore, the relatively high MAE of Hokkaido and Hokuriku is probably because the installed capacities are small in comparison with the other areas. Their ratios, compared to the nationwide amount, were 3.0% and 2.0%, respectively, at the end of December 2018. Compared with a model that does not incorporate temperature variables, which we constructed separately to evaluate the effects of temperature terms, the present model was found to have an R-squared value higher by 1.2% (2.5% for in-sample) and a lower MAE by 0.7% (1.7% for in-samples; note that these values were calculated as simple averages for each area). It was shown, therefore, that temperature information contributed to the further refinement of PV power generation forecasts.

[2] Downloaded from https://www.tepco.co.jp/forecast/html/area_data-j.html
[3] Downloaded from https://www.fit-portal.go.jp/PublicInfoSummary
[4] Downloaded from http://weather-transition.gger.jp/

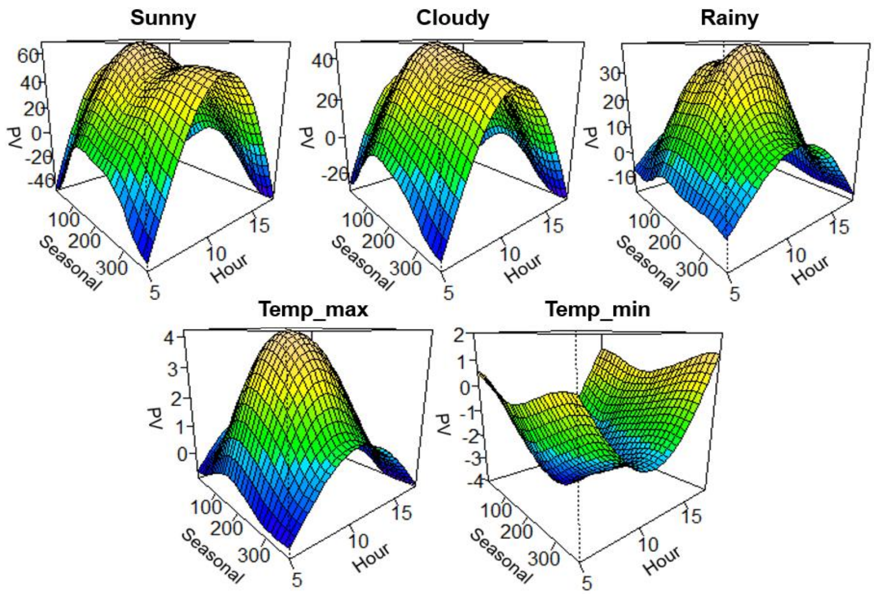


Figure 1. Estimated trends of PV power generation (example of Tokyo area).

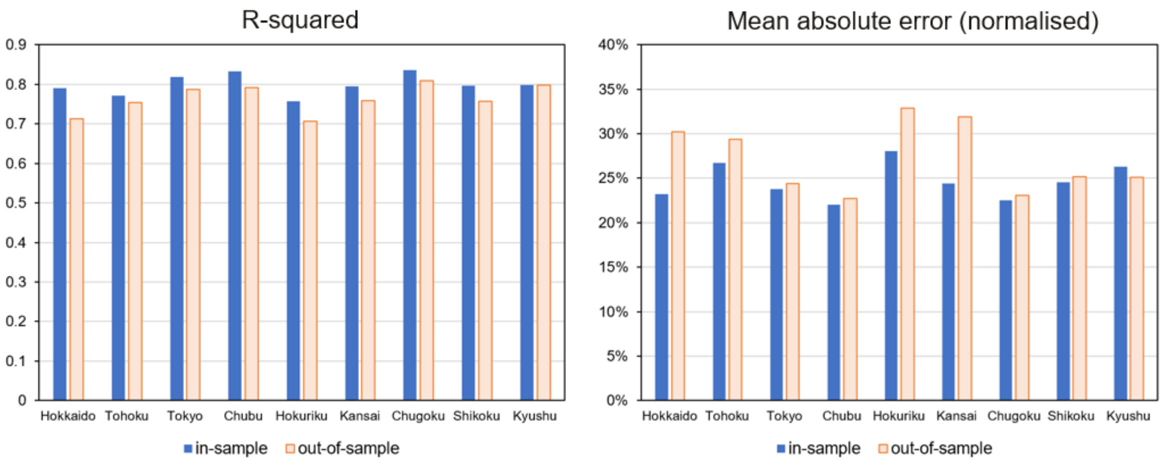


Figure 2. Evaluation of PV power generation forecast error (left: R-squared, right: MAE).

3.2. Empirical result for demand forecast

For the forecast of demand, we use the following observed data:

- Power demand $D_{t,\,h}$ [MW]: measured data published by Organization for Cross-regional Coordination of Transmission Operators (OCCTO)[5]
- Weather condition dummy $I_{\cdot,t,h}$ max (min) temperature $Tmax_t$ ($Tmin_t$) [°C]: forecast values announced by the JMA on the previous morning (one major city in each of nine areas)

Figure 3 illustrates the estimated trend of the Tokyo area. The trends by day type show that the effect of decreasing demand is strong in the order of Sunday (holidays) and Saturday. The estimated trends of the maximum and minimum temperatures, which correspond to the sensitivities of one-degree rise to demand, reflect the knowledge that, as with temperature, the demand rises in summer (positive sensitivity) and decreases in winter (negative sensitivity). Conversely, for the intraday time trend, the value of sensitivity for the maximum temperature becomes the highest at around 15:00, and the value

[5] Downloaded from http://occtonet.occto.or.jp/public/dfw/RP11/OCCTO/SD/LOGIN_login#

IOP Conf. Series: Earth and Environmental Science **812** (2021) 012001 doi:10.1088/1755-1315/812/1/012001

for the minimum temperature is attenuated as the time approaches the night (the forecast horizon is longer). When examining trends based on weather conditions, the sunny day trend has a similar shape to the trend of maximum temperature, but it differs in terms of the decrease in winter daytime. This may be due to self-generated PV power generation, meaning that sunny weather does not only directly decrease the demand, but also increases the self-generated PV power supply, which may result in further decreasing the system demand. The trend of rainy weather has an inverted U shape in summer, probably because the overnight cooling effect causes a decline in demand, whereas during daytime, humidity increases demand (for dehumidification). For the rainy trend in winter, the demand rises from midnight until the morning, perhaps because of freezing, and the sensitivity approaches 0 at the end of the day because the forecast horizon is longer. Note that it is natural that the trends are not at the same level (connected) at 0:00 and 23:00 because the forecast horizon is different.

Figure 4 summarises the forecast error for the out-of-sample period compared with the in-sample data. For both the R-squared and the MAE, there was no noticeable decline or increase in the out-of-sample period, and it can be concluded that the demand forecast model is sufficiently robust, as confirmed by the PV power forecast model.

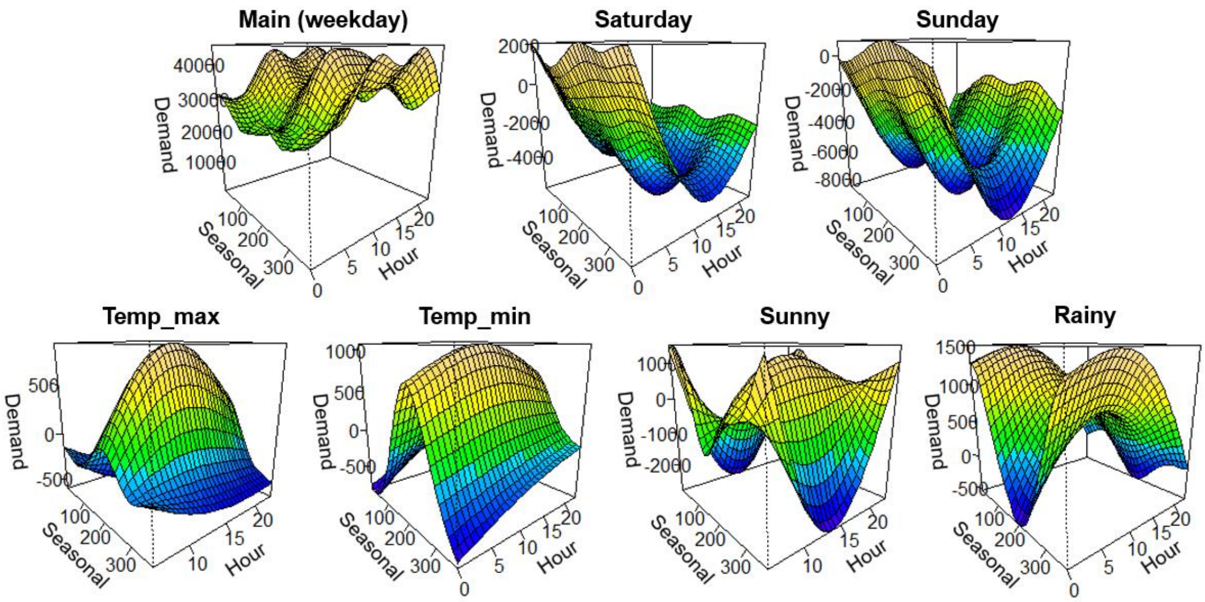


Figure 3. Estimated trend of demand (example of Tokyo area).

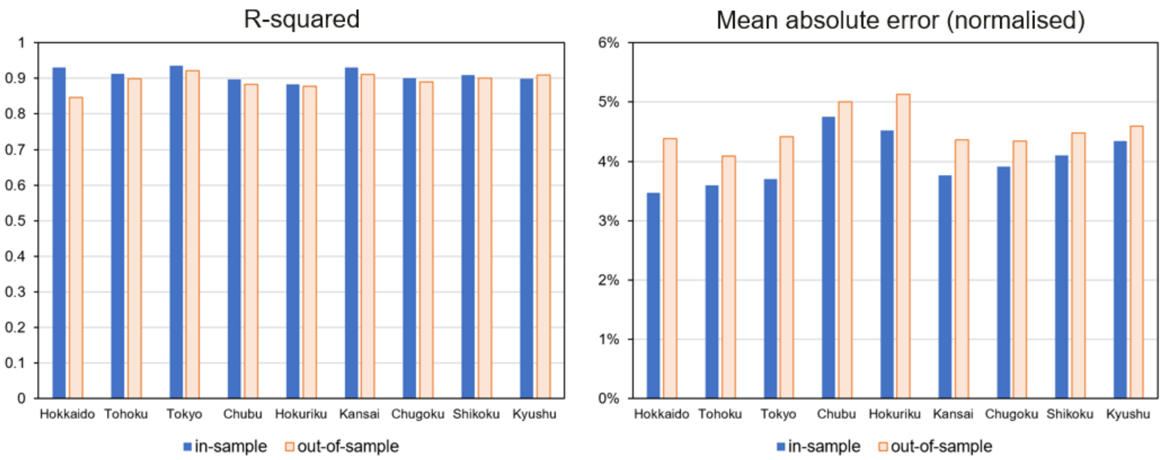


Figure 4. Evaluation of demand forecast error (left: R-squared, right: MAE).

4. Conclusion

In this study, we constructed forecast models for demand and PV power generation using two-dimensional tensor product spline functions. All the estimated two-dimensional trends, including variable coefficients of general weather condition data, were consistent with intuitive interpretation, and the validity of the model was visually demonstrated. Each model was verified using different data from nine power areas, demonstrating their robustness. Although the sample size used for model estimation was a little over a year and a half, the intertwined (non-linear) trends existing in the time direction, seasonal direction, and year direction were effectively estimated. Furthermore, the R-squared and MAE of out-of-sample did not show any remarkable deterioration compared to those of the in-sample for all of the nine area cases. This demonstrates the robustness of the proposed model.
In summary, our model has the following advantages:

- Only publicly available data are used, and therefore, most people can build a model
- The smoothing conditions are imposed in both hours and seasons, and so robustness is ensured, enabling the incorporation of multiple explanatory variables despite the small sample size
- The two-dimensional tensor product spline function is suitable for visualisation, highly interpretable, and easy for practitioners to manipulate when verifying or improving the model

The proposed model has advantages in terms of data availability, ease of implementation, robustness, applicability, and interpretability. It is expected to be used by new entrants, not only in the Japanese market, but also in many electric power markets.

Appendix A

The PV power capacity W_t is modelled by the following Ordinary Least Squares regression (OLS):

$$W_t = w_1 Period_t + w_2 + \eta_{w,\,t}. \tag{4}$$

where w_1 and w_2 are the coefficient and intercept, respectively, estimated by the OLS (4), $Period_t$ is the (annualised) daily dummy variable representing the number of years that have passed, and $\eta_{w,\,t}$ is the residual term. Using this equation, the forecast value of capacity $\widehat{W}_t$ can be obtained as follows:

$$\widehat{W}_t = w_1 Period_t + w_2. \tag{5}$$

Note that the original observed capacity W_t is monthly data with some missing values, but the forecast value $\widehat{W}_t$ can be obtained as daily granularity data because we use daily dummy $Period_t$.

Appendix B

The univariate smoothing spline function is estimated as the function h that minimises the penalised residual sum of squares (PRSS) given by

$$\text{PRSS} = \sum_{n=1}^{N} \{y_n - h(x_n)\}^2 + J(h), \quad \text{where} \;\; J(h) = \lambda \int \{h''(x)\}^2 \, dx. \tag{6}$$

In (6), the first term measures the approximation of the data, and the second term (penalty term) $J(h)$ adds penalties according to the magnitude of the curvature of the function. In this study, we construct the GAM using the R 3.6.1 package “mgcv” [9] to obtain the series of smoothing spline functions, where the smoothing parameter is calculated by general cross-validation criterion.
When estimating the two-dimensional tensor product spline function $h(x, z)$, the following penalty term $J_{te}(h)$ is included in PRSS that should be minimised [10]:

$$J_{te}(h) = \int_{x,z} \lambda_x \left(\frac{\partial^2 h}{\partial x^2}\right)^2 + \lambda_z \left(\frac{\partial^2 h}{\partial z^2}\right)^2 dxdz. \tag{7}$$

In this way, the tensor product spline function can incorporate the independent smoothing conditions for each variable (direction).

References

[1] Antonanzas J, Osorio N, Escobar R, Urraca R, Martinez-de-Pison F J and Antonanzas-Torres F 2016 Review of photovoltaic power forecasting *Solar Energy* **136** 78–111

[2] Suganthi L and Samuel A A 2012 Energy models for demand forecasting—A review *Renewable and Sustainable Energy Reviews* **16** (2) 1223–40

[3] Matsumoto T and Yamada Y 2019 Prediction method for solar power business based on forecasted general weather conditions and periodic trends by weather *Transac. Oper. Res. Soc. Japan* **62** 1–22 (in Japanese)

[4] Yamada Y, Kurahashi S and Yamaguchi N 2018 Demand forecasting model construction for electricity market trading simulation experiment *Abstracts of the 2018 Spring National Conference of the ORSJ* (in Japanese)

[5] Hastie T and Tibshirani R 1990 *Generalized additive models* (London: Chapman and Hall)

[6] Matsumoto T and Yamada Y 2018 Cross hedging using prediction error weather derivatives for loss of solar output prediction errors in electricity market *Asia-Pacific Financ. Mark.* **26** (2) 211–27

[7] Yamada Y, Makimoto N and Takashima R 2015 JEPX price predictions using GAMs and estimations of volume-price functions *JAFEE J.* **14** 8–39 (in Japanese)

[8] Campbell J Y and Thompson S B 2008 Predicting excess stock returns out of sample: Can anything beat the historical average? *Rev. Financ. Stud.* **21** (4) 1509–31

[9] Wood S N 2019 Package "mgcv" https://cran.r-project.org/web/packages/mgcv/mgcv.pdf (Accessed: 15th October 2020)

[10] Wood S N 2017 *Generalized additive models: an introduction with R*. (London: Chapman and Hall)

Acknowledgement

This work was supported by a Grant-in-Aid for Scientific Research (A) 16H01833, a Grant-in-Aid for Scientific Research (A) 20H00285, a Grant-in-Aid for Challenging Research (Exploratory) 19K22024, a Grant-in-Aid for Young Scientists 21K14374, and a Grant-in-Aid for Overseas Challenge Program for Young Researchers 201980226 from the Japan Society for the Promotion of Science (JSPS).