This is the accepted version of the following article: Fonseca Junior et. al. "Regional forecasts of photovoltaic power generation according to different data availability scenarios: a study of four methods." Progress in Photovoltaics: Res. and App 23.10 (2015): 1203-1218, which has been published in final form at https://doi.org/10.1002/pip.2528.

Regional Forecasts of Photovoltaic Power Generation According to Different Data Availability Scenarios: A Study of 4 Methods

Joao Gari da Silva FONSECA Junior.[a*], Takashi OOZEKI[a], Hideaki OHTAKE[a], Takumi TAKASHIMA[a], Kazuhiko OGIMOTO[b]

[a] National Institute of Advanced Industrial Science and Technology, AIST Tsukuba Central 2, 305-8568 Tsukuba, Ibaraki, JAPAN,
* Telephone: +81 029-861-5551, Fax: +81 029-861-5829, E-mail: jon-fonseca@aist.go.jp, jfonsecajunior@yahoo.com.br
[b] Institute of Industrial Science (IIS), Collaborative Research Center for Energy Engineering (CEE), Tokyo University 153-8505 Komaba 4-6-1, Meguro-ku, Tokyo, JAPAN.

Abstract:
The development of methods to forecast PV power generation regionally is of utmost importance to support the spread of such power systems in current power grids. The objective of this study is to propose and to evaluate methods to forecast regional PV power one-day ahead of time and to compare their performances. Four forecast methods were regarded of which 2 are new ones proposed in this study. Together they characterize a set of forecast methods that can be applied in different scenarios regarding availability of data and infrastructure to make the forecasts. The forecast methods were based on the use of support vector regression and weather prediction data. Evaluations were done for 1 year of hourly forecasts using data of 273 PV systems installed in 2 adjacent regions in Japan, Kanto and Chubu. The results show the importance of selecting the proper forecast method regarding the region characteristics. For Chubu, the region with a variety of weather conditions, the forecast methods based on single systems' forecasts and the one based on stratified sampling provided the best results. In this case the best annual normalized RMSE and MAE were 0.25 kWh/kWh_{avg} and 0.15 kWh/kWh_{avg}. For Kanto, with homogeneous weather conditions, the 4 methods performed similarly. In this case, the lowest annual forecast errors were 0.33 kWh/kWh_{avg} for the normalized RMSE and 0.202 kWh/kWh_{avg} for the normalized MAE.

Keywords: photovoltaic systems, regional power generation, support vector regression, stratified sampling, principal component analysis

1. Introduction

With the current conditions of world-wide increases of CO_2 emissions and doubts regarding the safety of nuclear power, associated with their long term harmful effects in case of accidents, power systems based on renewable resources are expected to give a strong contribution in the attenuation efforts regarding greenhouse gases emissions. Moreover, the widespread use of such systems is seen as a way to decrease the world's dependence on fossil fuels, helping in the transition to more sustainable societies. In Japan, after the nuclear disaster in Fukushima on 2011, most of nuclear power plants were put in maintenance mode and their reactivation is being strongly opposed. As result, the country became highly dependent on fossil fuel based power generation [1], which is increasing costs and CO_2 emissions. Under these conditions, to increase the share that power systems based on renewable resources have in the total power generation mix in Japan became a pressing matter, causing the government to reformulate its policies in this area. One consequence of this reformulation was the enactment of a new feed-in tariff program for photovoltaic, PV, systems in 2012. Through this program it

This is the accepted version of the following article: Fonseca Junior et. al. "Regional forecasts of photovoltaic power generation according to different data availability scenarios: a study of four methods." Progress in Photovoltaics: Res. and App 23.10 (2015): 1203-1218, which has been published in final form at https://doi.org/10.1002/pip.2528.

became very attractive to install PV systems in Japan, resulting in more than 3.3 GW of new installed capacity from 2012 to May of 2013 [2].

Given such growth rates it is expected that soon PV power will have a meaningful share in the total power generation in Japan. Thus interesting challenges to power utilities regarding the operation and integration of these systems on current power grids will be created. In this regard, one important issue is the intermittence of PV power, which is strongly dependent on weather conditions. Sudden and unexpected changes in the weather can cause sharp variation in the PV power generation, and if such variations in large scale are not properly dealt with frequency regulation and power supply issues can arise. One way to deal with this problem is to forecast PV power so that counter-measures to keep the frequency stable and the balance of power supply and demand can be properly and timely prepared.

As forecasting of PV power is an important tool to support of the widespread use of PV systems, several researchers are actively looking for methods that can provide accurate forecasts as it can be seen in Paulescu et al[3]. Different methods were proposed to forecast PV power in different forecast horizons. In the case of a few minutes or hours ahead of time, there are works such as the one of Chow et al[4], who proposed to use neural networks to mimic the nonlinear relation between weather parameters and PV power on 10 minutes ahead forecasts. Bracale et al[5], proposed a Bayesian approach modifying an auto-regressive time-series model to take in account the relation between the clearness index with other meteorological variables on 15 minutes ahead forecasts. Takahashi and Mori[6], devised a method based on Generalized Radial Basis Function Network and deterministic annealing, which improved 30 minutes ahead forecasts of PV power near to 12% compared with a simpler approach. Pedro and Coimbra[7], compared several forecasting techniques that do not use exogenous inputs to forecast PV power of a 1 MW power plant up to 2 hours ahead of time, concluding that a neural network with its parameters set by a genetic algorithm was the best approach. Regarding forecasts one day ahead or more, there is also a variety of studies available. Jimenez et al[8] proposed the use of two numerical weather models and an artificial neural network to forecast PV power of plant aiming the determination of suitable hours to perform maintenance tasks. Bofinger and Heilscher[9] used model of output statistics with local weather stations data to improve one day and five days ahead forecasts of PV power in Germany using irradiance forecasts of the ECMWF. We, in previous studies also proposed a method for 1 day ahead forecasts using support vector regression and numerical weather prediction, evaluating the approach with 1 year of data for a 1 MW power plant[10]. Most of these forecast methods are based on the use of past measured or forecasted of weather related variables. Furthermore, they can directly use past PV power in the training data [11], or be based on forecasts of irradiance. In the latter case, PV power is obtained from irradiance forecasts using conversion or downscaling models [12],[13],[14]. Finally, regardless the forecast horizon and the way PV power forecasts are obtained, several techniques can be used, such as neural networks, support vector regression, fuzzy logic [15] and others.

In spite of the variety of forecast methods for the PV power generation most of them are tested and developed for single PV systems. Of the studies mentioned only a few were developed to tackle the problem on regional scale, such as the work of Lorenz et al [12] focused on Germany. In Japan the structure to provide the necessary information to make feasible the application of one of more of these methods in regional scale was not yet conceived, and it is not even clear which structure should be prepared. For example, it is not known if every PV system in a region should be monitored to provide better regional forecasts of PV power, or how the gains in accuracy that such approach provides compares with the costs it creates. Furthermore, due to privacy concerns, it may not even be possible to monitor the power generation of every PV system without changing current laws. In this case, to measure only solar irradiance or just a

This is the accepted version of the following article: Fonseca Junior et. al. "Regional forecasts of photovoltaic power generation according to different data availability scenarios: a study of four methods." Progress in Photovoltaics: Res. and App 23.10 (2015): 1203-1218, which has been published in final form at https://doi.org/10.1002/pip.2528.

sample of PV systems could be feasible options but they may imply forecasts with lower accuracy. For power utilities operating in a region or even smart grid operators these are important topics that should be considered when choosing a method to forecast PV power generation to assist in the operation of their grids. For policy makers these are also important questions that should be answered when discussing which structure regarding networks, information flow, data storage, monitoring, etc, should be conceived so that useful forecasts of PV power generation can be provided. Moreover, the issues regarding the impact of PV power in regional scale are not only relevant in Japan but in all places where PV power systems are expected to provide an important share of the required power supply.

Thus, the objective of this study is to propose and to evaluate different methods to obtain regional forecasts of PV power generation in different scenarios regarding availability of data. Two new forecast methods are proposed and compared with other 2 existent ones. Together the methods represent 4 alternatives that can be applied in different scenarios regarding the availability of data to make the forecasts. The basic input data used were based on weather forecast data and the methods used, in different ways, support vector regression as forecast technique. Each method differs from the other regarding how the input data are used and what data regarding the PV systems installed in a region are available for the forecasts. To evaluate the methods, they were used to forecast, one-day ahead of time, hourly PV power generation during 1 year in Japan, for 2 regions with different areas, installed PV power capacity, and weather conditions. Furthermore, a third region including the previous 2 was also regarded. In this way, the performance of the forecast methods in different conditions is also assessed. The different regions size, weather conditions, PV systems capacity and installation conditions, give the study general applicability, and the methods proposed can be adapted to other locations. Furthermore, it is expected that the results of this study will also provide to policy makers and stakeholders in the renewable energy sector useful information when deciding about data availability and data aggregation policies to support regional PV power generation forecasting.

2. Data Used in the Forecasts

The data used in the forecasts are described in this section. The input data are based on weather forecast information and they are described in subsection 2.1. The output of the forecast is the hourly regional PV power generation. The sets of PV systems used in the study to represent the regional forecasts and their respective locations are described in section 2.2. It should be noted that how and which data are used depends on the type of forecast method, described in section 3.

2.1 Input Data

Most of the input data used were based on the weather forecasts provided by the grid-point-values obtained from the meso-scale model, GPV-MSM, of the Japan Meteorological Agency[16]. The GPV-MSM weather forecast system uses non-hydrostatic meso-scale modeling to reproduce the atmospheric phenomena. The GPV-MSM data-set has a 3 dimensional domain, covering an area of 3600 x 2400 km surrounding Japan, Fig. 1, and reaching an altitude of 21.8 km. Furthermore, the vertical domain is divided in 50 layers modeling the atmosphere phenomena at different heights. The GPV-MSM contains forecasts with a geographical resolution of 25 km^2 (0.05^o x 0.0625^o). Regarding Initial conditions, the MSM is initialized with forecasts of a weather system with coarser resolution (400 km^2), but with a domain covering the entire planet, the global spectral model, GSM. Forecasts of several weather related variables are obtained with the MSM, and these variables are forecasted hourly 8 times a day with a maximum

This is the accepted version of the following article: Fonseca Junior et. al. "Regional forecasts of photovoltaic power generation according to different data availability scenarios: a study of four methods." Progress in Photovoltaics: Res. and App 23.10 (2015): 1203-1218, which has been published in final form at https://doi.org/10.1002/pip.2528.

forecast horizon of 15 or 33 hours ahead of time, according to the forecast time.

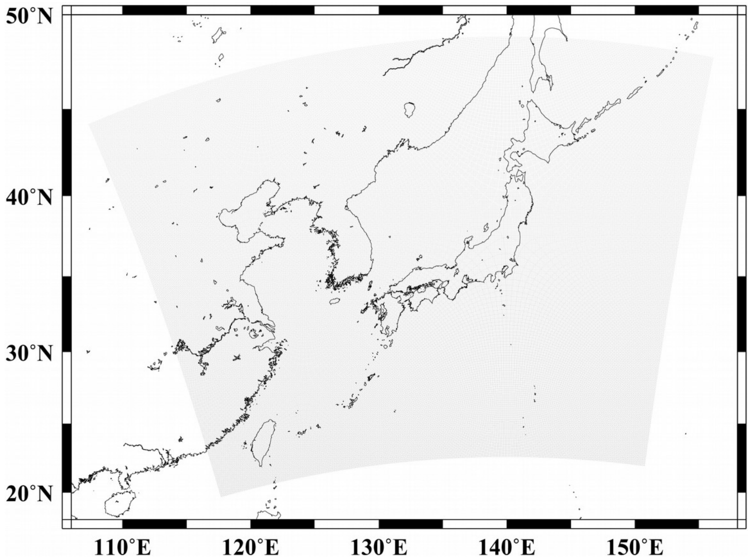


Figure 1 Computational domain of the MSM weather forecast system developed by the JMA.

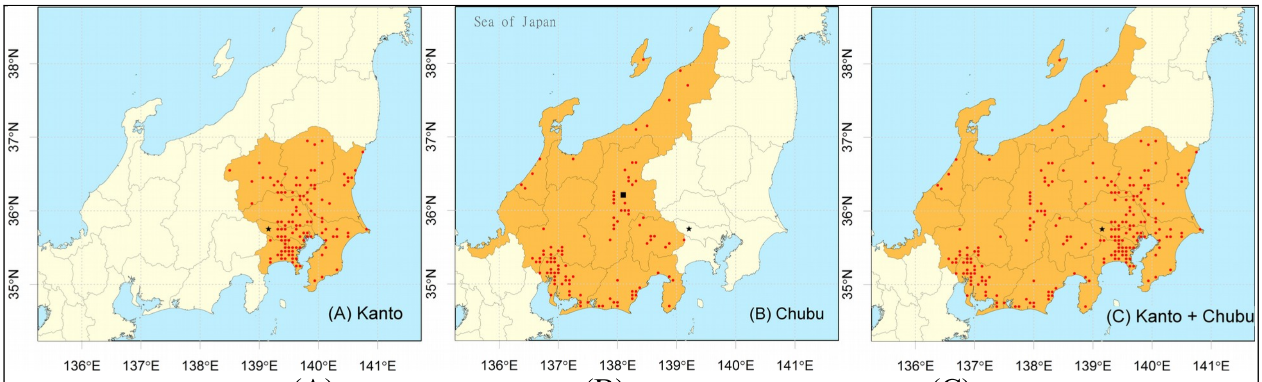
Surface and 2-dimensional data provided by the GPV-MSM were used in the forecasts. From the surface data, air temperature, and air relative humidity were used; from the 2 dimensional data cloudiness in three levels were used. The forecasts for next day were provided at 12h JST, characterizing a forecast horizon from 18h to 31h ahead of time. Besides the weather forecast data, the extraterrestrial horizontal insolation was also used as input data. This quantity is derived from the total amount of solar radiation reaching the planet before entering the atmosphere and it can be theoretically calculated for every location on the planet and hour of the forecasts [17].

This input data set was used in all forecast methods presented in section 3 excluding the persistence method. However, according to the forecast method the data set was partially used or used in different ways.

2.2 Output Data

The output data of the forecasts were the regional PV power generation for every hour of the period evaluated and every region regarded. Two regions were used in the study, the Kanto region and the Chubu region in Japan. To verify how the forecast methods perform with areas of different sizes, a third region considering Kanto and Chubu together was also regarded in the analyses. In Fig. 2 the three areas regarded and the location of their respective PV systems are presented.

This is the accepted version of the following article: Fonseca Junior et. al. "Regional forecasts of photovoltaic power generation according to different data availability scenarios: a study of four methods." Progress in Photovoltaics: Res. and App 23.10 (2015): 1203-1218, which has been published in final form at https://doi.org/10.1002/pip.2528.



(A) (B) (C)

Figure 2- Target areas (in orange) with their respective PV systems' locations (in red) regarded in the regional forecasts of PV power. * indicates the Tokyo administrative area. ▪Indicates Nagano Prefecture.

The Kanto region has generally a homogenous weather conditions with a high occurrence of clear-sky days on winter; the PV system set regarded was well distributed with some concentration in the Kanagawa prefecture in the south of Tokyo. On the other hand, Chubu region has distinct weather conditions with high mountains in its central area, the coast of the Sea of Japan with heavy snow fall and high occurrence of overcast days in winter, and the coast of pacific, with almost no snow in winter. Moreover, the PV systems regarded in Chubu were partially concentrated in the southwest and in Nagano prefecture in the central area, indicated in Fig. 2 by the ▪ symbol, with some examples in most of the prefectures of the region. As Fig. 2 shows, both regions have different distribution of PV systems, different sizes and different weather conditions providing good examples to evaluate the regional forecast methods.

In total data of 273 PV systems were used in the study, reaching more than 8.1MW of installed capacity. These systems are part of the Field Test project maintained by the New Energy and Industrial Technology Development Organization, NEDO, in Japan. The description of the PV systems per region is in Tab. 1 and Fig. 3.

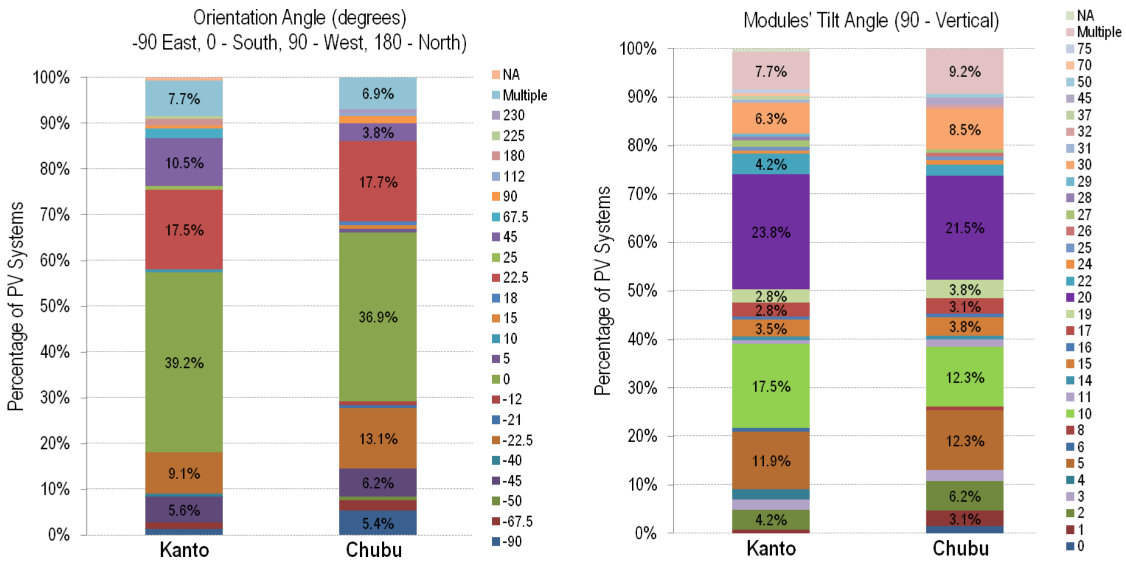
Table 1 – Regions' area and characteristic of their respective PV systems' sets.

		Region 1	Region 2	Region 3
Region name		Kanto	Chubu	Kanto +
Area		32 423.90 km^2	72 572. 34 km^2	104 996.24
Installed Capacity		4 409.02 kW	3 694.16 kW	8 103.18 kW
Number of PV systems		143 (100%)	130 (100%)	273 (100%)
Cell's type distribution	Amourphous-Si	5 (3.50%)	1 (0.77%)	6 (2.20%)
	Monocrystalline-Si	6 (4.20%)	0	6 (2.20%)
	Heterojunction with Intrinsic Thin	16 (11.19%)	14 (10.77%)	30 (10.99%)
	Polycrystalline-Si	116 (81.11%)	115 (88.46%)	231(84.61%)

In Table 1, the regional PV power output for every hour is regarded as the sum of the PV power generation of all 143 PV systems for Kanto, 130 PV systems for Chubu, and 273 PV systems for the region including Kanto and Chubu. Hourly forecasts were done for the 3 regions for one year, 2009. Furthermore, most of the PV systems were of the polycrystalline type; examples of other types were present in small quantities.

This is the accepted version of the following article: Fonseca Junior et. al. "Regional forecasts of photovoltaic power generation according to different data availability scenarios: a study of four methods." Progress in Photovoltaics: Res. and App 23.10 (2015): 1203-1218, which has been published in final form at https://doi.org/10.1002/pip.2528.

Regarding the orientation angle of the modules, more than 70% of them were facing south or near to south directions (Fig. 3A). Common tilt angles were 10º and 20º and near to 90% of the modules had tilt angles lower or equal to 30º (Fig. 3A), characterizing typical installation conditions in Japan.



(A) (B)

Figure 3- Modules' orientation angles (A) and tilt angles (B) of the PV systems set according to the region.

3. Methods to Forecast the Regional PV power

Four methods were evaluated in the forecast of regional PV power generation during any given hour. Each method reflects a scenario regarding the availability of data for the forecasts. Thus, regional forecasts can be obtained for different conditions, markets or regulations. The four methods were not only compared to each other but also with a reference approach based on persistence. The persistence method simply regards the hourly values of PV power generation measured in a day as the forecast values for the next day. This method is a naïve way to make a forecast and any other forecast method should provide results at least better than persistence to justify its use. Finally, Method 1 and Method 3 were proposed on previous studies. To complete the 4 scenarios analyses in this study we propose further 2 new forecast methods, Method 2 and Method 4, and present a comprehensive analysis of their performances.

3.1 Method 1 – Based on each PV Systems' Power Generation

In this method an hourly forecast for each PV system in a region is done. The regional value is then obtained adding the forecasted contributions of every PV system. To make a single PV system forecast we used a method based on the use of support vector regression [10].

As the method for forecasts of single PV systems was already presented, only a brief description of it will be given. The ν support vector regression was used as the forecast algorithm. In it the deviation that a forecast can have from a target answer is minimized; and the algorithm works expressing in a hyper dimensional space the problem of determining the relation between sets of input patterns and their targets. The problem is taken to a high dimension space with map functions and then restated with Lagrangian coefficients so that an optimization procedure can take place. In this way, the training stage with a support vector regression algorithm is treated as an optimization problem. Further information about support vector

This is the accepted version of the following article: Fonseca Junior et. al. "Regional forecasts of photovoltaic power generation according to different data availability scenarios: a study of four methods." Progress in Photovoltaics: Res. and App 23.10 (2015): 1203-1218, which has been published in final form at https://doi.org/10.1002/pip.2528.

regression can be found in literature [18], [19].

To use the support vector regression a set of configuration parameters in the optimization problem and a kernel function, which will perform the mapping, have to be chosen beforehand. To find suitable values for these parameters an approach based on ensemble set of models was used. In this approach ranges for the variation of the values of the configuration parameters are predefined and a forecast is done for each possible combination. After all the combinations generated a forecast, the median of the results is chosen as the algorithm forecast output. This procedure to set the support vector regression was previously proposed in the similar insolation forecast problem, yielding good results [20].

The kernel function used was the radial basis function described in Eq. 1. The parameter γ in Eq. 1 and the cost *C* and the ν parameters from the support vector regression optimization procedure, not showed here, varied in the range of values presented in Table 2. These ranges of variation were found based on trial and error experiments.

$$k(x_i, x_j) = e^{-\gamma \| x_i - x_j \|^2} \quad \text{..................................} \quad (1)$$

Table 2- Range of values for the configuration parameters of the support vector regression algorithm.

Configuration Parameter	Minimum Value	Maximum Value	Step Size
C	2^1	2^4	2^{X+1}
ν	0.5	0.9	0.2
γ	2^{-3}	2^0	2^{X+1}

In total 48 combinations of values for the configuration parameters were used. For each combination the algorithm is trained and yields hourly forecasts. Once all 48 forecasts were done their median, for every hour, was regarded as the output of the forecast method. This procedure was repeated for each day of forecasts. With this approach the configuration and training of the support vector regression are integrated and the use of spare data just to set the algorithm, with a grid search approach for example, becomes unnecessary as the only data used are the training data.

For each day of forecasts the algorithm is trained with the 60 days of data preceding such day. A day of forecasts for a single PV system is regarded as the period from 6h to 19h and hourly values of PV power generation are provided by the forecast method. As for the implementation of the support vector regression the port for R language of the LibSVM library[21] was used.

For a day of forecasts for each PV system, the data described in section 2.1 of the nearest grid-point to its location were used as input. Namely, to forecast the PV power generation of a system at a given hour, forecasts of air temperature, relative humidity, cloudiness in 3 levels and the horizontal extraterrestrial insolation for such hour were used with the values of temperature, relative humidity and extraterrestrial insolation of the preceding hour. Furthermore, each PV system had its power generation forecasted by a support vector regression algorithm trained with the PV system data as showed in Fig. 4. Finally, the regional value is easily calculated adding the forecasts of power of each PV system.

Some considerations should be done about this forecast method. First, it requires the monitoring of power generation of every PV system in a region. Depending on market

This is the accepted version of the following article: Fonseca Junior et. al. "Regional forecasts of photovoltaic power generation according to different data availability scenarios: a study of four methods." Progress in Photovoltaics: Res. and App 23.10 (2015): 1203-1218, which has been published in final form at https://doi.org/10.1002/pip.2528.

conditions, data aggregators or the PV systems' owners could do such monitoring although it may not be a feasible solution in some countries. Second, for thousands of PV systems installed in a region such method may not scale well depending on how the forecasts are done. If all data are centralized in one place and then individual forecasts are done, this method may require a significant computing power to yield solutions in short amount of time. However, if the forecasts are done in a decentralized way and just the result is aggregated, the scaling problem is attenuated.

In spite of these disadvantages, there is not any physical or technological impossibility to implement such method and if in some places individual PV systems' monitoring is already done, for maintenance of fault detection for example, it could be an option for the regional forecasting. The stand-alone performance of this method for 5 regions in Japan including Kanto and Chubu was studied in Fonseca et al [22].

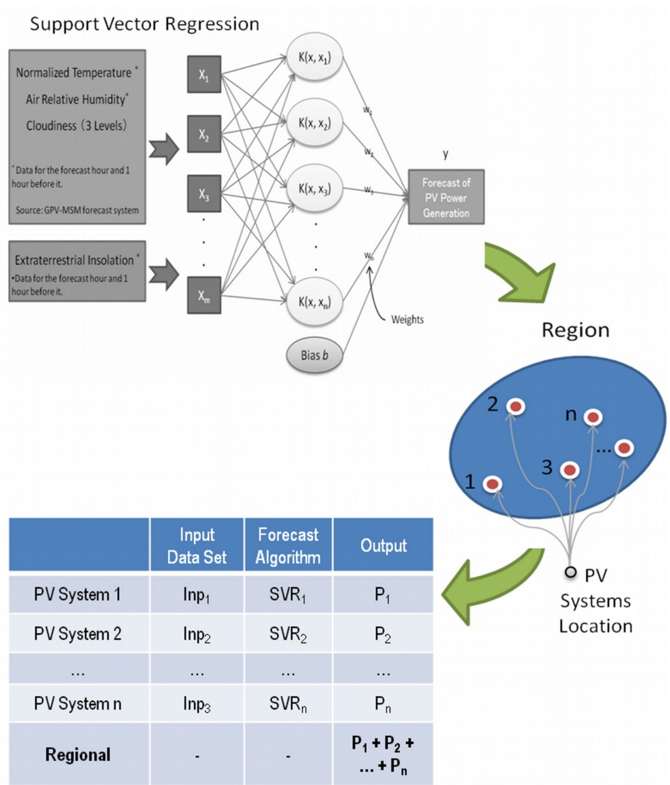


Figure 4 – Method 1 for regional forecasts (using each PV systems' power generation).

3.2 Method 2 – Based on Stratified Sampling

Another option to forecast regional PV power is to use sampling. This method is suitable for situations where it is not possible to monitor every single PV system in a region and where PV power generation is not measured regionally. In this case, it is possible to select a few PV systems through sampling, to make forecasts of their power generation using support vector regression as showed in section 3.1, and finally to calculate the regional yield from those forecasts. In order to that, a method based on stratified sampling regarding installed capacity and PV system location is proposed in this section.

Before describing the proposed method 2 issues about the sampling are discussed, the sampling size and the sampling procedure to calculate the regional forecast. To obtain a suitable sampling size it was assumed that, for every hour, the mean of the PV power generation of the systems in the samples, normalized by their capacity, follow a normal distribution. In this case the sampling size can be calculated according to Eq. 2, as showed in Nelson et al [23].

$$n = \left[\frac{t(\alpha/2;\infty)\, s}{d} \right]^2 \tag{2}$$

In Eq. 2 is *d* the error margin for the mean regional PV power provided by the sampling, *t* is the t-student distribution, *s* is the standard deviation of the samples value and *n* is the sample size. Using Eq. 2 the sampling size can be determined for a given confidence level as long as *s* and *d* can be estimated.

For the problem of the hourly regional PV power generation forecast, *s* and *d* vary according to the hour of day, season of year and weather conditions. Therefore, estimates for *s* and *d* were made so that constant values can be used regardless the hour of the day, and a reasonably low sampling size can be obtained. The value for *d* was assumed to be 0.05 kWh/kW_{cap}, which is approximately the annual root mean square error of regional forecasts of power reported in literature [24]. Based on knowledge of forecasts of PV systems in Japan, the value for *d* was assumed to be 0.15 kWh/kW_{cap}. The use of these values was preliminary assessed in a previous study [25], yielding good results. Using the proposed values and a confidence level of 95% the nominal sampling size achieved was 39 samples. Thus, if the sampling theory applies, the use of 39 samples should be enough to represent the regional PV power yield, regardless the total number of PV systems installed in a region.

Regarding the sampling procedure, it was used stratified sampling of PV systems according to their location. In stratified sampling the PV systems are separated in categories or strata and from each stratum a subset of PV systems is selected. The stratum regarded was the prefecture where the PV systems are located. The number of samples to be selected from each prefecture is decided according to the proportion of PV systems of such prefecture related to the total. Thus, if a prefecture contains 10% of the PV system of the set, 10% of samples will be retrieved from this prefecture. With stratified sampling the aim is to achieve a balanced set of samples that represent efficiently the total population. Applying stratified sampling to the 3 regions and rounding up the number of samples per stratum yielded a sampling size of 41 PV systems for Kanto and Chubu and 44 PV systems when Kanto and Chubu are regarded as one region.

Finally, after the samples of each stratum are selected and their corresponding PV power capacities are calculated a correction factor is applied to each stratum to account also for the weight of their capacity in the total installed capacity of the set. The correction factor is determined by Eq. (3).

$$f_i = \frac{\mathrm{Ct}_{\mathrm{pref},i}}{\mathrm{Ct}_{\mathrm{reg}}} \times \frac{\mathrm{Cs}_{\mathrm{tot}}}{\mathrm{Cs}_{\mathrm{pref},i}} \tag{3}$$

In Eq. (3) *f* is a proportionality correction factor. It is calculated based on total PV power capacity of a prefecture $Ct_{pref,\,i}$, the total PV power capacity of region Ct_{reg}, the PV power capacity of samples retrieved from such prefecture $Cs_{pref,\,i}$, and the total PV power capacity of the sample set Cs_{tot}. After *f* for a prefecture is found it is then applied to the forecasted PV power generation of the samples retrieved from such prefecture.

Once the sample set is selected and its PV power generation forecasted, P^{set}_{fcst} in kWh, normalized by the set capacity, Cs_{tot} in kW, is calculated and an upscaling procedure is done to calculate the regional yield, P^{reg}_{fcst} in kW, as showed in Eq. (4).

$$P_{\text{fcst}}^{\text{reg}} \approx \frac{P_{\text{fcst}}^{\text{set}}}{\text{Cs}_{\text{tot}}} \times \text{Ct}_{\text{reg}} \tag{4}$$

Using this sampling procedure it is possible to obtain the regional forecast of PV for regions of any size, any number of PV systems and capacity. In this study, the sampling procedure was used to make 1 year of hourly regional forecasts 100 times, each time with a different random sample, and the annual error distribution of these forecasts was calculated. The sampling the yielded the median annual forecast error was then selected to represent typical PV power regional forecasts that can be obtained with this method.

3.3 Method 3 – Based on Regional PV Power Generation

A third scenario proposed is one where individual power generation data of the PV systems installed in a region are not available, but the total regional yield of PV power is available. In this case, it is not necessary to make individual forecasts and an algorithm such as support vector regression can be trained and used to model directly the relation between regional PV power and weather data.

The question in this case is about which weather data to use for the forecasts. Average values characterizing the regional weather conditions could be used. Another idea is to use, within a region, all local weather data from the geographic locations near to the place where PV systems are installed or where they are expected to be installed. However, this approach, even for the few hundred PV systems regarded in this study, would generate a large number of input variables. Such condition associated with the fact that several of these input variables would be highly correlated, would cause difficult learning problems for the algorithm, resulting in forecasts with low accuracy.

To deal with these problems and to make method 3 feasible pre-processing of the weather data with principal component analysis, PCA, as showed in Fig.5 is done. With PCA a linear invertible transformation in a set with *n* correlated variables is performed to obtain a set of uncorrelated variables. After that, the variables of the new set are ordered according to the weight that their variance has in the total variance of the set. Finally, only the few first variables are selected so that most of the intrinsic information of the original set is retained. The theory behind PCA can be found in several references such as in Haykin [26].

The PCA transformation was done using eigenvalue decomposition of the covariance matrix of the input data. To find a suitable number of principal components to use as input data of the forecasts, a threshold for the cumulative variance of 90% of total variance of the transformed data set was used. Moreover, the procedure was applied sequentially to every forecast day, before training the forecast algorithm. PCA was applied using input data of the forecast day and of the previous 60 days to obtain a covariance matrix.

With method 3 regional forecasts can be done when there is only regional PV power generation data. Moreover, only one support vector regression is trained for a day of forecasts. However, the preprocessing step may require intensive computational resources if not properly conducted. For example, in large regions covered by a weather forecasts system with high resolution and thousands of PV systems installed, it may be computationally expensive and ineffective to perform PCA using all data. In this case, a limited number of points with weather forecasts data can be set so that they describe properly the main characteristics of the locations in the region where the PV systems are installed. Once this number is set PCA can then be easily applied. As the sample used involves at most 273 PV systems, this step was not necessary

This is the accepted version of the following article: Fonseca Junior et. al. "Regional forecasts of photovoltaic power generation according to different data availability scenarios: a study of four methods." Progress in Photovoltaics: Res. and App 23.10 (2015): 1203-1218, which has been published in final form at https://doi.org/10.1002/pip.2528.

and PCA was directly applied. The specific benefits of a PCA based forecast of regional PV power versus a no-PCA condition was assessed in Fonseca et al [27] and part of the results of this analysis for Kanto region will be presented in the results section for comparison purposes.

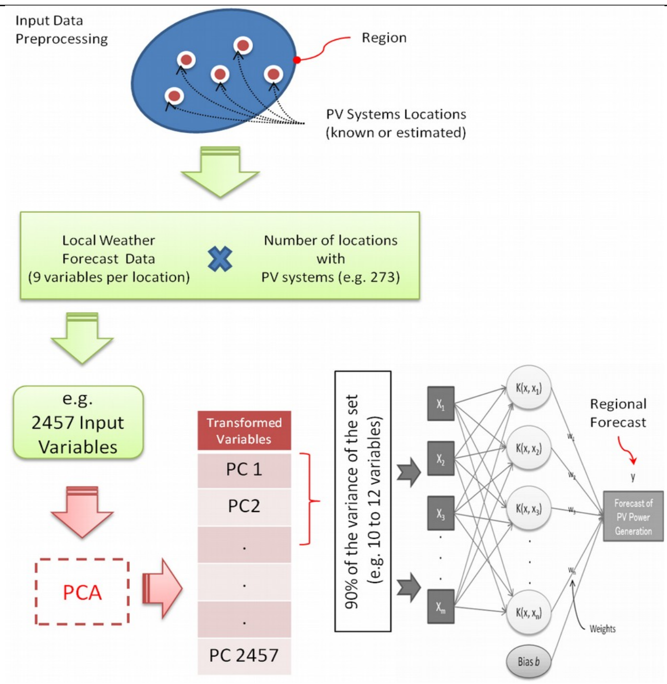


Figure 5 – Regional forecast method 3 (Using Regional PV Power Generation).

3.4 Method 4 – From Solar Irradiation Forecasts

To characterize a scenario where no PV power data are available a new method to forecast regional PV power was devised. This method is based on the assumption that PV power generation data are not available at all, either on local or regional scale. In this case, PV power generation forecasts can be obtained from solar irradiation forecasts. Even though PV power generation data are not required with this method, it is still necessary to know the installed capacity of PV systems in a region and their approximated distribution regarding their module tilt angles and orientations. If such data are available or can be estimated, the regional PV power generation, P^{reg}_{fcst} in kWh, can be calculated as showed in Eq. 5 and Eq. 6.

$$P_{\mathrm{fcst}}^{\mathrm{reg}} \approx \frac{P_{\mathrm{cap}}^{\mathrm{reg}}}{H_{\mathrm{ref}}} k \sum a_i H_{\mathrm{inc},i} \quad (5)$$

$$\sum a_i = 1 \quad (6)$$

In Eq. 5 the regional capacity is regarded as the total PV systems capacity, P^{reg}_{cap} in kW, rated at a reference incident solar irradiance, H_{ref}, equals to 1 kW/m^2. In the same equation k is the balance of system coefficient, and a is the weight given to the forecast of solar irradiation that reaches the PV modules with a given tilt angle and orientation, H_{inc} in kWh/m^2. For example, if 25% of the PV systems in the region have modules with a tilt angle of 10° and are facing south, the forecast of solar irradiation reaching these modules in any given hour is calculated and then multiplied by a weight a of 0.25 to yield the corresponding PV power generation of these modules. The procedure then is repeated for all categories of tilt angles and orientation of the PV systems in a region. With this approach the information about the distribution of tilt angles and orientation of the modules with the total rated capacity is enough to provide a forecast of regional PV power generation from the solar irradiation forecasts. Regarding the solar irradiation on tilted plane, H_{inc}, it was calculated from global horizontal solar irradiation forecasts using the models proposed by Perez et al[28] and Erbs et al[29].

It should be noted that a distribution of balance of system coefficients, *k*, can also be used if available. However, for simplification purposes *k* was kept constant in this study. The value for *k* and for the distribution of tilt angles and orientation are in Table 3.

Table 3- Values used for the distribution of PV systems in the 3 regions.

Orientation	South
Weight *a*	1
Tilt angle	10^0
k	0.8

It was assumed, for all regions, that all PV systems have modules with a tilt angle of 10^o, and that all of them are facing south. These assumptions are rough approximations of the real modules angle and orientation distributions of the PV systems used in the study, Fig. 3A and 3B. However, in a real case application it may be difficult to have precise information about the modules. Therefore, if method 4 is robust enough to work with rough estimates, its use can be better justified.

Regarding the forecasts of global horizontal solar irradiation, they can be obtained from meteorological service providers or directly forecasted. In this study, they were obtained from the forecasts of basic weather variables using a procedure similar to the one described in section 3.1 for individual PV power systems. An hourly forecast of global horizontal solar irradiation was done with support vector regression for every point of the grid of the GPV-MSM system covering the 3 regions analyzed. For each point in the grid, a support vector regression algorithm was trained with measured past and forecasted data of the nearest meteorological station. Once forecasts of global horizontal solar irradiation are done, the regional value is regarded as the mean value of the forecasts for all the point in the grid of the GPV-MSM covering the target region. This value is then converted to the equivalent amount reaching the PV modules, according to the tilt angle and orientations in Table 3, and then inserted in Eq. 5 with their respective weights *a*.

4. Error Parameters

Three error parameters were regarded, the normalized root mean squared error, *nRMSE* in kWh/kWh_{avg}, the normalized mean absolute error, *nMAE* in kWh/kWh_{avg}, and the skill score, *SS*. Only the period between 6h to 19h of each day were considered in the calculations. The *nRMSE* and *nMAE* are measures normalized by the average regional PV power generation during the evaluation period, 2009. They are expressed in Eq. 7 and Eq.8.

$$\mathrm{nRMSE} = \sqrt{1 N \sum_{j=1}^{N} \square} \quad (7)$$

This is the accepted version of the following article: Fonseca Junior et. al. "Regional forecasts of photovoltaic power generation according to different data availability scenarios: a study of four methods." Progress in Photovoltaics: Res. and App 23.10 (2015): 1203-1218, which has been published in final form at https://doi.org/10.1002/pip.2528.

$$\mathrm{nMAE}=1\,N\sum_{j=1}^{N}\frac{\left|P_{\mathrm{fcst},j}^{\mathrm{reg}}-P_{\mathrm{msd},j}^{\mathrm{reg}}\right|}{P_{\mathrm{msd,avg}}^{\mathrm{reg}}} \tag{8}$$

The skill score *SS* is based on a ratio of mean squared errors, and it is an indicative of the improvement achieved with the forecast methods when compared with a reference method. The reference used for the skill score calculations were the forecasts obtained with persistence as described in section 3.

$$\mathrm{ss}=1-\frac{1\,N\sum_{j=1}^{N}\left(P_{\mathrm{fcst},j}^{\mathrm{reg}}-P_{\mathrm{msd},j}^{\mathrm{reg}}\right)^2}{1\,N\sum_{j=1}^{N}\left(P_{\mathrm{ref},j}^{\mathrm{reg}}-P_{\mathrm{msd},j}^{\mathrm{reg}}\right)^2} \tag{9}$$

In each hour *j* in Eq 7, Eq. 8 and Eq. 9, P^{reg}_{msd} is the regional PV power measured, in kWh. In Eq. 7 and Eq.8 $P^{reg}_{msd,avg}$ is the annual mean PV power generation measured, in kWh. In Eq. 9 P^{reg}_{ref} is the regional PV power generation forecasted with persistence.

In recent studies the use of error measures based on the mean square errors as the main error parameter in the analysis of forecasts of weather related phenomena have been questioned[30] and even different metrics have been proposed[31]. Therefore, although the *nRMSE* and skill score will be presented most of the analyses will focus on the *nMAE*.

5. Results

In Fig. 6 are the annual skill scores obtained with the 4 methods to forecast regional PV power in the 3 regions studied. All forecast methods yielded high skill scores, which indicate that they provide forecasts significantly better than the forecasts that can be achieved with persistence.

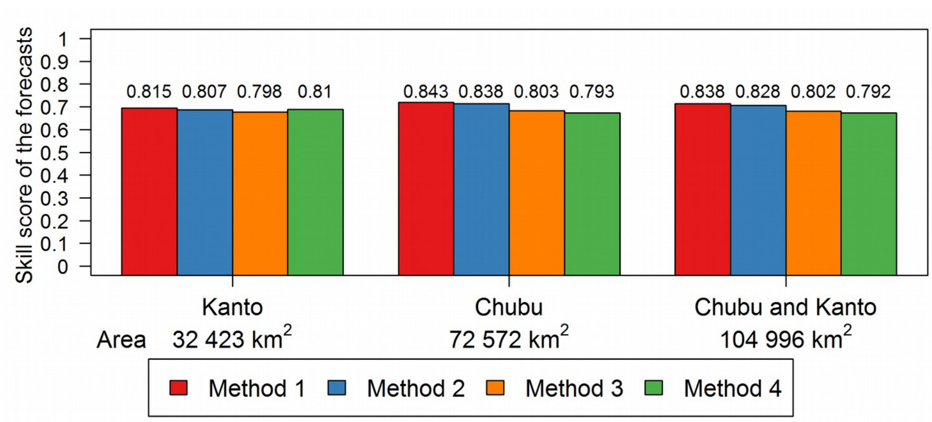


Figure 6. Annual skill score obtained with each regional PV power forecast method.

Comparing the 4 methods in the regions with different sizes, for Kanto they performed similarly and their forecasts had skill scores around 0.8. However, for Chubu and the region comprising Chubu and Kanto, Method 1 and Method 2, which are based on single systems' PV power forecasts and sampling, yielded better results than Method 3 and Method 4. Looking from the area size point of view, increasing the area from Kanto to Chubu improved the skill scores of the forecasts obtained with Method 1 and Method 2. For Method 3 and Method 4 the same did not happen as it can be noted in the percentage variation presented in Table 4.

Table 4- Variation of the skill score according to the area size and forecast method.

% Variation From Kanto to	Method 1	Method 2	Method 3	Method 4
Chubu	3.4%	3.8%	0.6%	-2.1%
Chubu and Kanto	2.8%	2.6%	0.5%	-2.2%

The percentage variations in Table 4 show that the skill scores obtained with Method 1 and Method 2 improved more than 3% going from Kanto region to Chubu region, but it did not reach higher values for the largest area (Chubu and Kanto). On the other hand, Method 3 was the most stable one yielding a maximum skill score percentage variation of 0.6% regardless the area. Finally, Method 4 did not scale up well as its related skill score became continuously lower with the increase of the area size.

The skill scores provided a comparison between the proposed forecast methods and persistence. To focus on the differences between the performances of the 4 methods, their annual *nRMSE* and *nMAE* in Fig. 7 and Fig. 8 are presented. The same error parameters calculated for the forecasts obtained with persistence are also presented.

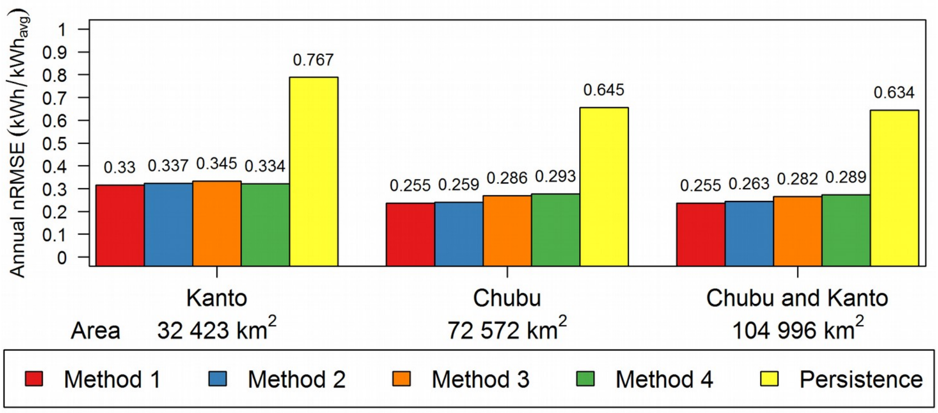


Figure 7 Annual *nRMSE* obtained with each regional PV power forecast method.

The *nRMSE* of the regional forecasts decreased with the increase of the region size, for Method 3 and Method 4. Nevertheless, the lowest values were reached with Method 1 and Method 2 at Chubu. Moreover, Method 3 and Method 4 were less sensitive to the area size than Method 1 and Method 2. These two methods provided a *nRMSE* 23% lower comparing Chubu with Kanto, whereas for Method 3 and Method 4 the variation was 17% and 12%.

This is the accepted version of the following article: Fonseca Junior et. al. "Regional forecasts of photovoltaic power generation according to different data availability scenarios: a study of four methods." Progress in Photovoltaics: Res. and App 23.10 (2015): 1203-1218, which has been published in final form at https://doi.org/10.1002/pip.2528.

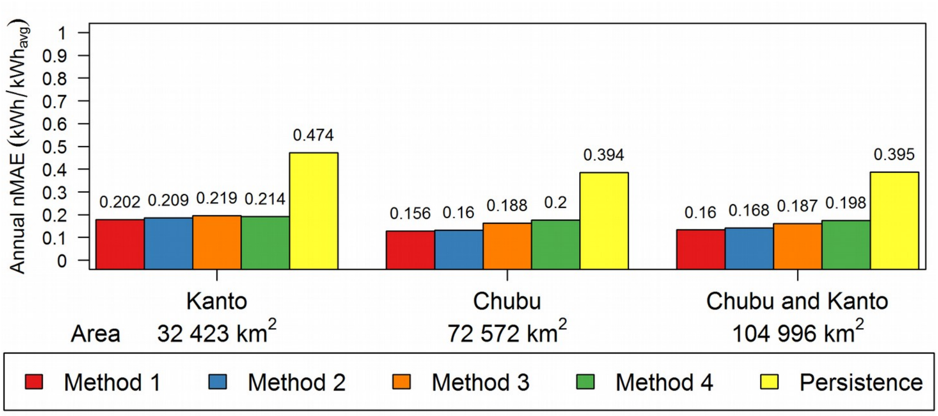


Figure 8. Annual *nMAE* obtained with each regional PV power forecast method.

The behavior of the *nMAE* was similar to the one observed with the *nRMSE*. In this case however, the weaker effect of the region size in the quality of the forecasts of Method 3 and Method 4 it is even clearer. Regardless the error parameter used or the region size, annually, Method 1 yielded the lowest errors. Method 2 provided forecast errors slightly higher than the ones obtained with Method 1, being the second best method.

The difference between the forecast errors of the forecast methods with the area size can also be noted in Table 5, where the percentage variation between the forecast method with the lowest and highest *nMAE* are showed. In this case the difference between the best and worst forecast method can reach almost 30% depending on the region size.

Table 5. Highest variation of the *nMAE* according to the area size and the 4 forecast methods.

	Kanto	Chubu	Kanto and Chubu
Method with lowest nMAE	Method 1	Method 1	Method 1
Method with highest nMAE	Method 3	Method 4	Method 4
Maximum % Variation*	6%	28%	23.7%

*Variation between the 4 methods (persistence not included) using the lowest forecast error of each region as reference.

Analyzing the annual results, a few trends are noted. First, Method 1, based on local PV power forecasts, had the best performance regardless the error parameter or area size. Method 2, based on sampling, approximated well the results obtained with Method 1 without requiring forecasts for all PV systems in a region. In spite the fact that method 1 had the best performance, its lowest annual errors were reached in Chubu, not the largest region regarded. Chubu, besides its large size, contains distinct weather conditions, providing a strong smoothing effect. For Kanto, with smaller area and homogeneous weather condition, the smoothing effect was weaker. Therefore, methods such as Method 1 and Method 2, which can account better for the smoothing effect than Method 3 and Method 4, presented a stronger performance in Chubu (an annual error reduction of 23% from Kanto to Chubu).

To verify seasonal variations, Fig. 9A, Fig. 9B and Fig. 9C contain the monthly *nMAE* obtained with Method 1 (values on the right axis) and the corresponding percentage variations (values on the left axis) that the other forecast methods presented compared with the values of Method 1. Looking at the monthly results, Method 1 also has the best performance most of the time. Only a few exceptions occurred such as in 3 months in Kanto, Fig. 9A, or 3 months in Chubu, Fig. 9(b).

Looking to the different regions, Kanto had its PV power forecasted with similar accuracy by all methods most of the time. Nevertheless, Method 4 presented a particularly poor performance on January. In Chubu the performance of method 4 was even poorer in other winter months and also in April. This behavior is somewhat unexpected if one considers that Method 4 provides regional PV power forecasts based on local forecasts of solar irradiation and thus it should also account for the smoothing effect, this time on the solar irradiation. However, one should also consider some factors that, on specific conditions, make the forecasts of solar irradiation to be decoupled from forecasts of PV power. First, snow accumulated on the PV panes does not affect the forecasts of solar irradiation, thus they will yield high overestimations on PV power forecasts not only on the day that snow falls but also on the following days. On the other hand if PV power is forecasted directly from PV panel past output, past few days of low PV power generation will enter in the training data and the forecast algorithm will consequently reduce its PV power generation for the following days. One should note that this characteristic does not act in favor of lower errors when snow is not accumulated in the panels, but generally it yielded lower errors than the error caused when snow is not regarded at all as in the case of solar radiation forecasts. The second factor is related with installation conditions and efficiencies of the PV panels. These conditions are only partially regarded in Method 4. Therefore, the smoothing effect will not be as high as the one provided by Method 1. For example, diverse shadow conditions of PV panels installed in different places will affect their power generation and this effect will enter in the training data of local PV power forecasts used by Method 1. The same does not happen with Method 4. The better expression by Method 1 of the smoothing effect in the regional PV power in Chubu can explain it better performance.

This is the accepted version of the following article: Fonseca Junior et. al. "Regional forecasts of photovoltaic power generation according to different data availability scenarios: a study of four methods." Progress in Photovoltaics: Res. and App 23.10 (2015): 1203-1218, which has been published in final form at https://doi.org/10.1002/pip.2528.

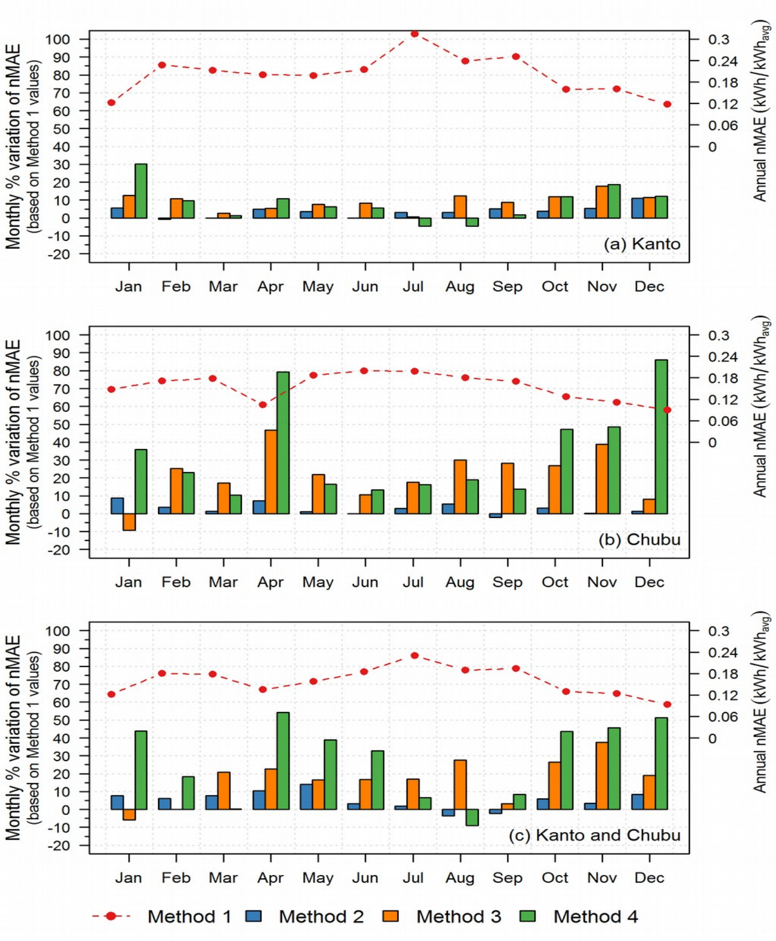


Figure 9. Monthly % variation of the *nMAE* of Methods 2, Method 3 and Method 4 compared with Method 1 for Kanto (A), Chubu (B), Kanto and Chubu (C).

Comparing the regional forecasts of Chubu and Kanto in Fig. 9A and Fig. 9B, it is also noted that for Kanto the effect of the rain season in summer months affected the accuracy of the forecast errors in July. For Chubu however with its larger area, different localized weather conditions and stronger smoothing effect, the rain season is not uniform throughout the region. Thus, stronger local forecast errors did not happened at the same time preventing the occurrence of the peak detected in July in Fig. 9A, and yielding monthly regional forecast error more stable than the ones of Kanto.

As a general trend, it is also noted that the regional forecast error for the area Chubu + Kanto was not significantly better the ones of Chubu. This results from the fact that there are more PV systems in Kanto than in Chubu in the set studied. Therefore joining the two areas caused more than half of the systems to be concentrated in one specific area. This concentration increased disproportionally the weight that the forecast errors for the PV systems in Kanto have on the total regional forecast error, preventing a better smoothing effect.

To further stress the difference between the performances of the forecast methods of regional PV power, Fig. 10 contains the error distributions for 1 year of regional forecasts with the 4 methods and for the 3 regions studied.

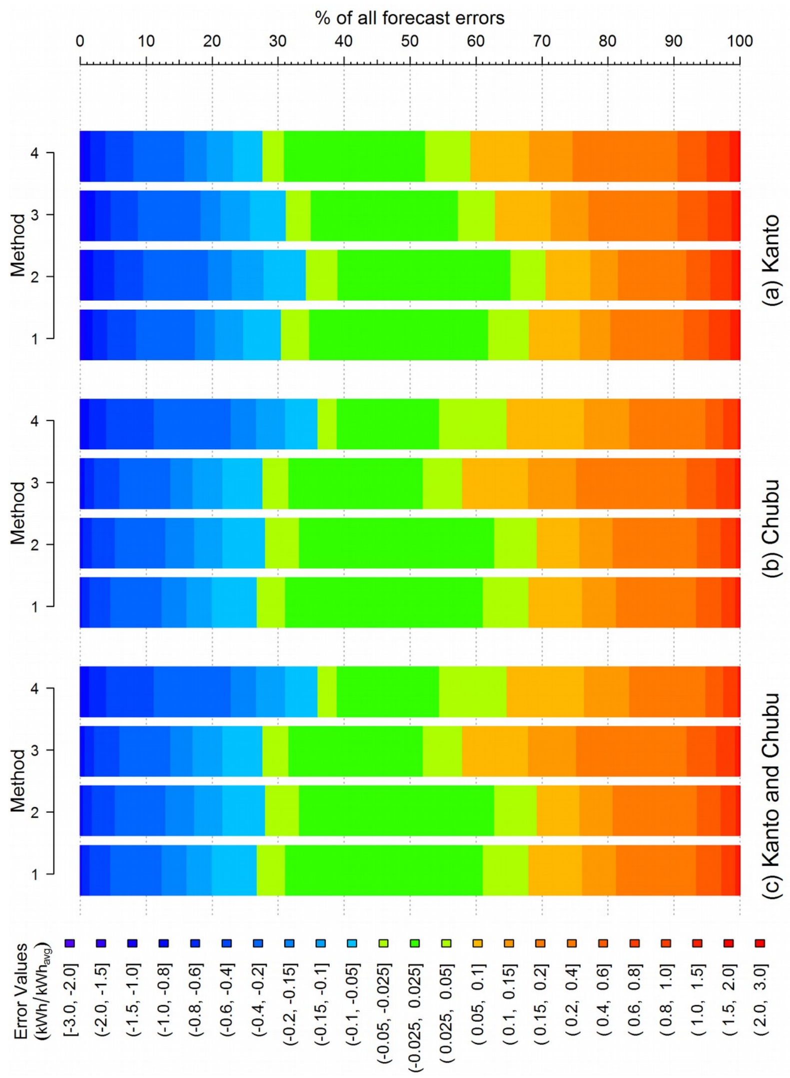


Figure 10 – Hourly error distribution in 3 regions and with 4 forecast methods (2009).

Although the previous results showed that for Kanto the annual error difference between the forecast methods is small, Fig. 10A shows that Method 1 and Method 2 provided forecasts with lowest biases. Method 3 and Method 4 often yielded overestimated forecasts, with positive bias. For Method 4, in the regions in Fig. 10B and Fig.10C the bias became stronger, whereas for Method 3 the bias increased but at a lower rate. Regarding the error values, with Method 1 and Method 2 around 40% of all regional forecast errors were between -0.05 kWh/kWh$_{avg}$ and 0.05 kWh/kWh$_{avg}$ and around 65% of all errors were between -0.15 kWh/kWh$_{avg}$ and 0.15 kWh/kWh$_{avg}$. Still on Method 1 and Method 2, overestimations between 0.15 kWh/kWh$_{avg}$ and 0.2 kWh/kWh$_{avg}$ comprised near to 10% of all total errors, being more common than underestimations in the same range of values.

Analyzing the performance of Method 4 on the different regions, it provided lower errors at specific ranges when going from Kanto to Chubu, Fig. 10A and Fig. 10B. Namely the errors between 0.15 kWh/kWh$_{avg}$ and 0.2 kWh/kWh$_{avg}$ corresponded to 15% of all errors in Kanto but only to 11% of the errors in Chubu. Furthermore, the share of the error range between 0.05 kWh/kWh$_{avg}$ and 0.10 kWh/kWh$_{avg}$ increased from 6% in Kanto to 9% in Chubu also indicating an improved performance and sensitiveness to the smoothing effect strength.

This is the accepted version of the following article: Fonseca Junior et. al. "Regional forecasts of photovoltaic power generation according to different data availability scenarios: a study of four methods." Progress in Photovoltaics: Res. and App 23.10 (2015): 1203-1218, which has been published in final form at https://doi.org/10.1002/pip.2528.

Nevertheless, the improvement noticed with Method 1 when going from Kanto to Chubu was stronger. The range of lower errors, between -0.025 to 0.025, increased from 27% to 30% of all errors going from Kanto to Chubu, whereas with Method 4 the share of the same range of errors actually decreased from 21% to 16% of all errors.

The last results presented are in Fig. 11A, Fig. 11B and Fig. 11C, and they contain the annual *nMAE* per hour of day with the different forecast methods and in the 3 regions studied. In Kanto Method 3 presented the highest errors at peak hours (11h to 13h). However, in the other 2 regions Method 4 had the worst performance. Where with Method 1, Method 2 and Method 3 the forecast errors in peak hours fall sharply when going from Kanto to Chubu, with Method 4 they remain almost constant causing the difference in the annual results presented in Fig. 8. Also, comparing Kanto with Chubu, the forecast errors in peak hours vary 25% with Method 1, reflecting the stronger smoothing effect that occurs in Chubu.

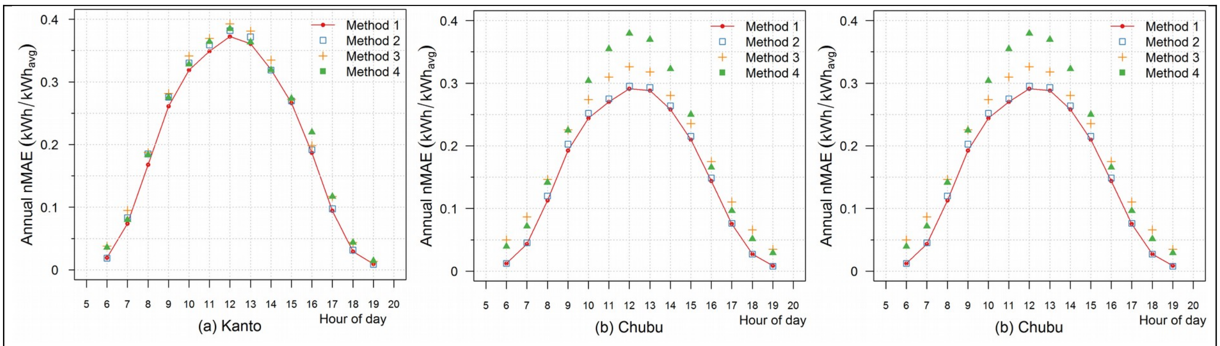


Figure 11. Annual *nMAE* per hour of day in the 3 regions with yielded by the 4 forecast methods of regional PV power.

6. Conclusions

The objective of this study was to propose and to evaluate different methods for one-day-ahead forecasts of regional PV power generation. Each forecast method assumed a different scenario regarding data available to make the forecasts. Two new methods to forecast regional PV power were proposed characterizing 4 scenarios regarding availability of data for the forecast. The methods were evaluated in 2 regions of different sizes, number of PV systems, and characteristics. A third area including both regions was also regarded. The results showed that for an area with similar weather conditions and small snow fall as Kanto, all methods can yield forecasts with the same level of accuracy. However, for Chubu, where a strong smoothing effect happens, Method 1, based on single PV systems forecasts, and Method 2, based on sampling, yielded the best results. Method 3 is based on past regional PV power data and preprocessing of weather variables. Therefore, it is not able to directly account for the error canceling process that occurs with the smoothing effect. Method 4, considered the smoothing effect, but the one that occurs on the forecasts of solar irradiation, presumably weaker than the one that occurs with regional PV power. Furthermore a series of assumptions have to be made, such as the solar irradiation in a tilt plane, the distribution of PV systems' modules tilt angles, orientation and balance of system coefficients. With such assumptions Method 4 was not able to incorporate well the smoothing effect that occurs in Chubu, yielding annual error values similar to the ones obtained in Kanto.

Based on the results, Method 1 or Method 2 were the best methods to provide regional forecasts of PV power. If monitoring of all single PV systems installed in a region is not possible the proposed forecast method based on stratified sampling, geographic information and installed

This is the accepted version of the following article: Fonseca Junior et. al. "Regional forecasts of photovoltaic power generation according to different data availability scenarios: a study of four methods." Progress in Photovoltaics: Res. and App 23.10 (2015): 1203-1218, which has been published in final form at https://doi.org/10.1002/pip.2528.

capacity, Method 2, is an valid option. Regarding the other new forecast method proposed, Method 4, it should not be completely discarded as an option. As showed by Lorenz et al[12], detailed representation of PV systems installed in a region associated with forecasts of irradiance are able to provide good forecasts of regional PV power generation. Besides, Method 4 may be the only option in places where monitoring of PV system power generation is not done. In this case better information regarding the distribution of systems' orientation and tilt angles, as well as the application of the method in smaller areas may yield better results. Finally, in the case of Method 3, a better application of preprocessing of input data of the forecasts also has the potential to yield better forecasts. Although, it should be difficult to obtain the same level of low regional forecast error achieved with the smoothing effect in large regions with different weather conditions, just improving the preprocessing of the input data.

Regarding the regions studied, Chubu is the largest geopolitical region in Japan and it was possible to forecast regional PV power forecast in it with accuracy 25% higher than the one achieved in Kanto. Annual error values for the regional forecasts were 0.15 kWh/kWh$_{avg}$ in the best case. Thus, this value should provide an indication of the lowest regional forecast errors that can currently be achieved in the geopolitical regions in Japan. Regardless the focus of the study on data of Japan, the new methods proposed can be easily applied to other locations, as long as there are similar input-output data.

In spite of the differences found between the accuracy of the forecast methods, each one has potential improvement points. Furthermore, one of the methods can be applied in a situation where another cannot be, providing options to stakeholders so that they can obtain regional forecasts of PV power according to their operation conditions. Additionally, improvements to yield better forecast error should be investigated as large regional errors are still occurring as showed in Fig. 9. Finally, the calculation of the reliability of the forecasts also provides important information about their quality and it will be addressed in future studies.

Acknowledgment

This work was supported by NEDO (New Energy and Industrial Development Organization, Japan) in the project Research and Development of PV Performance and Reliability Characterization Technologies, and by the Japan Science and Technology Agency (JST, CREST).

References

[1] U.S. Energy Information Administration (EIA), "Japan's fossil-fueled generation remains high because of continuing nuclear plant outages", Today in Energy, Available at: http://www.eia.gov/todayinenergy/detail.cfm?id=10391#. [Accessed: 10/Sep/2013].

[2] J. METI, "On the State of New Installations of Renewable Energy Systems", May 2013, in Japanese.

[3] M. Paulescu, E. Paulescu, P. Gravila, and V. Badescu, Weather Modeling and Forecasting of PV Systems Operation, 2013th ed. Springer, 2012.

[4] S. K. H. Chow, E. W. M. Lee, and D. H. W. Li, "Short-term prediction of photovoltaic energy generation by intelligent approach," Energy Build., vol. 55, pp. 660–667, Dec. 2012.

[5] A. Bracale, P. Caramia, U. De Martinis, and A. R. Di Fazio, "An Improved Bayesian-based Approach for Short Term Photovoltaic Power Forecasting in Smart Grids," in Proceedings of International Conference on Renewable Energies and Power Quality (ICREPQ 2012).

[6] Takahashi, Masato and Mori, Hiroyuki, "A Hybrid Intelligent System Approach to Forecasting of PV Generation Output," in Proceeding of the International Conference on Electrical Engineering 2012, Kanazawa, Japan, 2012, pp. 176–179.

This is the accepted version of the following article: Fonseca Junior et. al. "Regional forecasts of photovoltaic power generation according to different data availability scenarios: a study of four methods." Progress in Photovoltaics: Res. and App 23.10 (2015): 1203-1218, which has been published in final form at https://doi.org/10.1002/pip.2528.

[7] H. T. C. Pedro and C. F. M. Coimbra, “Assessment of forecasting techniques for solar power production with no exogenous inputs,” Sol. Energy, vol. 86, no. 7, pp. 2017–2028, Jul. 2012.
[8] L. A. Fernandez-Jimenez, A. Muñoz-Jimenez, A. Falces, M. Mendoza-Villena, E. Garcia-Garrido, P. M. Lara-Santillan, E. Zorzano-Alba, and P. J. Zorzano-Santamaria, “Short-term power forecasting system for photovoltaic plants,” Renew. Energy, vol. 44, pp. 311–317, Aug. 2012.
[9] S. Bofinger and G. Heilscher, “Solar Electricity Forecast - Approaches and First Results,” in Proceeding of the 21st European Photovoltaic Solar Energy Conference, Dresden, Germany, 2006.
[10] J. G. da S. Fonseca, T. Oozeki, T. Takashima, G. Koshimizu, Y. Uchida, and K. Ogimoto, “Use of support vector regression and numerically predicted cloudiness to forecast power output of a photovoltaic power plant in Kitakyushu, Japan,” Prog. Photovolt. Res. Appl., vol. 20, no. 7, pp. 874–882, 2012.
[11] P. Bacher, H. Madsen, and H. A. Nielsen, “Online short-term solar power forecasting,” Sol. Energy, vol. 83, no. 10, pp. 1772–1783, Oct. 2009.
[12] E. Lorenz, T. Scheidsteger, J. Hurka, D. Heinemann, and C. Kurz, “Regional PV power prediction for improved grid integration,” Prog. Photovolt. Res. Appl., Special Issue, Sep. 2010.
[13] A. Yona, T. Senjyu, A. Y. Saber, T. Funabashi, H. Sekine, and C. H. Kim, “Application of Neural Network to One-Day-Ahead 24 hours Generating Power Forecasting for Photovoltaic System,” in Intelligent Systems Applications to Power Systems, 2007. ISAP 2007. International Conference on, 2008, pp. 1–6.
[14] S. Pelland, G. Galanis, and G. Kallos, “Solar and photovoltaic forecasting through post-processing of the Global Environmental Multiscale numerical weather prediction model,” Prog. Photovolt. Res. Appl., 2011.
[15] J. Cao and X. Lin, “Application of the diagonal recurrent wavelet neural network to solar irradiation forecast assisted with fuzzy technique,” Eng. Appl. Artif. Intell., vol. 21, no. 8, pp. 1255–1263, Dec. 2008.
[16] K. Saito, T. Fujita, Y. Yamada, J. Ishida, Y. Kumagai, K. Aranami, S. Ohmori, R. Nagasawa, S. Kumagai, C. Muroi, T. Kato, H. Eito, and Y. Yamazaki, “The Operational JMA Nonhydrostatic Mesoscale Model,” Mon. Weather Rev., vol. 134, no. 4, pp. 1266–1298, Apr. 2006.
[17] M. Iqbal, Introduction to Solar Radiation. Academic Press Inc, 1984.
[18] V. Vapnik, The Nature of Statistical Learning Theory, 2nd ed. Springer, 1999.
[19] B. Schölkopf, P. Bartlett, A. Smola, and R. Williamson, “Support Vector Regression with Automatic Accuracy Control,” Proc. ICANN98 Perspect. Neural Comput., pp. 111–116, 1998.
[20] J. G. da S. Fonseca, O. Takashi, O. Hideaki, S. Ken-ichi, T. Takumi, and O. Kazuhiko, “Analysis of Different Techniques to Set Support Vector Regression to Forecast Insolation in Tsukuba, Japan,” J. Int. Counc. Electr. Eng., vol. 3, no. 2, pp. 121–128, 2013.
[21] C.-C. Chang and C.-J. Lin, “Training nu-support vector regression: theory and algorithms,” Neural Comput., vol. 14, no. 8, pp. 1959–1977, Aug. 2002.
[22] J. G. da S. Fonseca Jr., T. OOZEKI, H. Ohtake, K. Shimose, T. Takashima, and K. Ogimoto, “Characterizing the Regional Photovoltaic Power Forecast Error in Japan: A Study of 5 Regions,” IEEJ Trans. Power Energy, vol. 134, no. 6, 2014.
[23] P. R. Nelson, M. Coffin, and K. A. F. Copeland, Introductory statistics for engineering experimentation. Amsterdam; Boston: Academic Press, 2003.
[24] E. Lorenz, J. Hur, G. Karampela, D. Hei, H. G. Beyer, and M. Schneider, “Qualified

This is the accepted version of the following article: Fonseca Junior et. al. "Regional forecasts of photovoltaic power generation according to different data availability scenarios: a study of four methods." Progress in Photovoltaics: Res. and App 23.10 (2015): 1203-1218, which has been published in final form at https://doi.org/10.1002/pip.2528.

Forecast of Ensemble Power Production by Spatially Dispersed Grid-Connected PV Systems," in Proceedings of the 23rd European Photovoltaic Solar Energy Conference and Exhibition, Valenci, 2008, pp. 3285 – 3291.

[25] J. G. S. Fonseca Jr, T. OOZEKI, H. Ohtake, K. Shimose, T. Takashima, and K. OGIMOTO, "A Comparison of Sampling Techniques for the Regional Forecast of Photovoltaic Power in Japan," in Proc. 12th International Conference on Sustainable Energy Technologies, Hong kong, 2013, pp. 2074–2082.

[26] S. Haykin, Neural Networks and Learning Machines, 3rd ed. Pearson Education, 2008.

[27] J. G. da S. Fonseca Jr., T. Oozeki, H. Ohtake, K. Shimose, T. Takashima, and K. Ogimoto, "Regional Forecasts and Smoothing Effect of Photovoltaic Power Generation in Japan: An Approach with Principal Component Analysis," Renewable Energy, 2014, under review.

[28] R. Perez, P. Ineichen, R. Seals, J. Michalsky, and R. Stewart, "Modeling daylight availability and irradiance components from direct and global irradiance," Sol. Energy, vol. 44, no. 5, pp. 271–289, 1990.

[29] D. G. Erbs, S. A. Klein, and J. A. Duffie, "Estimation of the diffuse radiation fraction for hourly, daily and monthly-average global radiation," Sol. Energy, vol. 28, no. 4, pp. 293–302, 1982.

[30] C. J. Willmott and K. Matsuura, "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance," Clim. Res., vol. 30, no. 1, p. 79, 2005.

[31] R. Marquez and C. F. Coimbra, "A novel metric for evaluation of solar forecasting models," 2011.