Very short term irradiance forecasting using the lasso

Dazhi Yang[a,*], Zhen Ye[b], Li Hong Idris Lim[c], Zibo Dong[d]

[a]*Singapore Institute of Manufacturing Technology (SIMTech), Agency for Science, Technology and Research (A*STAR), 71 Nanyang Drive, Singapore 638075, Singapore*
[b]*Modules Division, REC Solar Pte Ltd., 20 Tuas South Avenue 14, Singapore 637312, Singapore*
[c]*Department of Electronic Systems, University of Glasgow (Singapore), 535 Clementi Road, Singapore 599489, Singapore*
[d]*Department of Electrical and Computer Engineering, National University of Singapore, 4 Engineering Drive 3, Singapore 117583, Singapore*

Abstract

We find an application of the lasso (least absolute shrinkage and selection operator) in sub-5-minute solar irradiance forecasting using a monitoring network. Lasso is a variable shrinkage and selection method for linear regression. In addition to the sum of squares error minimization, it considers the sum of ℓ_1-norms of the regression coefficients as penalty. This bias-variance trade-off very often leads to better predictions.

One second irradiance time series data are collected using a dense monitoring network in Oahu, Hawaii. As clouds propagate over the network, highly correlated lagged time series can be observed among station pairs. Lasso is used to automatically shrink and select the most appropriate lagged time series for regression. Since only lagged time series are used as predictors, the regression provides true out-of-sample forecasts. It is found that the proposed model outperforms univariate time series models and ordinary least squares regression significantly, especially when training data are few and predictors are many. Very short-term irradiance forecasting is useful in managing the variability within a central PV power plant.

Keywords: Lasso, Irradiance forecasting, Monitoring network, Parameter shrinkage

1. Introduction

Variability in solar irradiance reaching the ground is primarily caused by moving clouds. To accurately forecast the irradiance, cloud information must be directly or indirectly incorporated into the formulation. Due to the stochastic nature of the clouds, it is difficult to fully model their generation, propagation, and extinction using physical approaches. Statistical methods are therefore often used to extract cloud information from observations (e.g. Yang et al., 2015; Dong et al., 2014; Lonij et al., 2013).

We are particularly interested in very short term (sub-5-minute) irradiance forecasting as the clouds are relatively persistent during a short time frame. Unlike the forecasts with longer horizons where the results are essential for electricity grid operations, the very short term forecasts find their applications in large photovoltaics (PV) installations. Knowing the potential shading/unshading over a particular section of a PV system in advance may be advantageous to maximum power point tracking algorithms (Hohm and Ropp, 2000). Accurate sub-minute forecasts could also bring possibilities to better control of ramp-absorbing ultracapacitors (Mahamadou et al., 2011; Teleke et al., 2010).

Inman et al. (2013) reviewed the state-of-the-art methods for very short term irradiance forecasting. The methods involve using either sky cameras (Nguyen and Kleissl, 2014; Yang et al., 2014c; Quesada-Ruiz et al., 2014) or a sensor network (Lipperheide et al., 2015; Bosch and Kleissl, 2013; Bosch et al., 2013). All of

*Corresponding author at: Singapore Institute of Manufacturing Technology (SIMTech), Agency for Science, Technology and Research (A*STAR), 71 Nanyang Drive, Singapore 638075, Singapore; previously at: Solar Energy Research Institute of Singapore (SERIS), National University of Singapore. Tel.: +65 9159 0888.

Email address: yangdazhi.nus@gmail.com (Dazhi Yang)

these listed references aim at explicitly deriving the cloud motions and thus forecast the irradiance. Beside many assumptions, such as linear cloud edge, that have to be made, various types of error will be embedded in different phases of such methods, especially during the conversion from cloud condition to ground-level irradiance. It is therefore worth investigating the alternative methods where cloud information is considered indirectly.

Along-wind and cross-wind correlations observed between two irradiance time series have been studied intensively in the literature (e.g. Arias-Castro et al., 2014; Hinkelman, 2013; Lonij et al., 2013; Perez et al., 2012). If along-wind correlation between a pair of stations can be observed, we can use regression-based methods for forecasting. However, several problems have to be addressed before we describe our method:

- The discrepancy between the direction of a station pair and the direction of wind may result in a smaller correlation. How do we incorporate the strength of cross-correlation between monitoring sites into the forecasting model?
- When the wind speed changes from day to day or even within a day, the choices of lagged time series also need to be constantly updated. How do we then automatically select the most appropriate spatio-temporal neighbors for forecasting?
- When the correlation is unobserved, do we need to switch the spatio-temporal forecasting algorithm to a purely temporal algorithm in an ad hoc manner?

With these questions, we consider the lasso (least absolute shrinkage and selection operator) regression (Efron et al., 2004; Tibshirani, 2011, 1996). Lasso is a variable shrinkage and selection method for linear regression. In our application, the predictors (regressors) are the time series collected at the neighboring stations at various time lags (autocorrelated time series may also be used); the responses (regressands) are the time series collected at the forecast station. Some advantages of the lasso over the ordinary least squares regression, ridge regression and subset selection methods are discussed in section 2.

1.1. Data

Data from a dense grid of irradiance sensors located on Oahu Island, Hawaii, are used in this work. The network is installed by the National Renewable Energy Laboratory (NREL) in March 2010. It consists of 17 radiometers, as shown in Fig. 1. The sampling rate of these stations is 1 second. Previously, Hinkelman (2013) showed the possibility of observing highly correlated time series from this network; data from 13 days dominated by broken clouds were used in that study. We therefore use the data from the exact same days (Hinkelman, 2014) to study the predictive performance of such network configuration. The data are freely available at `http://www.nrel.gov/midc/oahu_archive/`.

Throughout the paper, the 1 second irradiance data will be averaged into various intervals to evaluate the forecasts with different forecast horizons. As high frequency data often have local maxima and minima caused by noise rather than cloud effects (Bosch and Kleissl, 2013), the smallest aggregation interval is 10 second. Prior to any forecasting, the global horizontal irradiance (GHI) time series from these 17 stations are first transformed into clearness index time series. Such transformation is commonly used in irradiance forecasting to stabilize the variance, i.e., to remove the diurnal trends in the GHI time series. We use the solar positioning algorithm developed by Reda and Andreas (2008) for extraterrestrial irradiance calculation. Finally, we include a zenith angle filter of $<80^\circ$.

1.2. Error metrics

All the forecasting models in this paper are built using the clearness index time series; the errors are evaluated using the GHI transformed back from the forecast clearness index. Two error metrics are used in this paper, namely, the normalized mean absolute error (nMAE) and the forecast skill (FS). The nMAE is

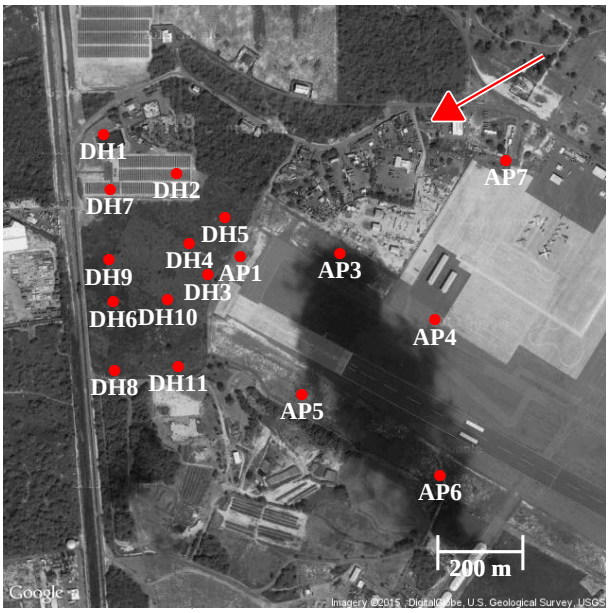

Figure 1: Layout of the 17 stations of the NREL Oahu network. The scale of the map is shown in the bottom right corner. The arrow in the top right corner shows the prevailing trade winds direction (60° from north). The average wind speed during the periods of analyses is 10 m/s. See (Arias-Castro et al., 2014; Hinkelman, 2013) for more details on the data.

given by:

$$\text{nMAE} = \frac{\frac{1}{n}\sum_{i=1}^{n}\left|\widehat{G}_i - G_i\right|}{\frac{1}{n}\sum_{i=1}^{n} G_i} \times 100\% \tag{1}$$

where G_i denotes the GHI measured at ith time step; $\widehat{G}_i$ denotes the forecast produced. The forecast skill (Chu et al., 2015) is given by:

$$\text{FS}(fh) = 1 - \frac{\text{nRMSE}(fh)}{\text{nRMSE}_{\text{p}}(fh)} \tag{2}$$

where fh denotes the forecast horizon; nRMSE_{p} and nRMSE are the normalized root mean square errors of the persistence model and the proposed model respectively. A persistence model assumes that the forecast is equal to the current observation; it is often used as a naive benchmark. The nRMSE is given by:

$$\text{nRMSE} = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\widehat{G}_i - G_i\right)^2}}{\frac{1}{n}\sum_{i=1}^{n} G_i} \times 100\% \tag{3}$$

The nMAE is a form of mean absolute error (MAE) while the forecast skill is a form of mean square error (MSE). MAE and MSE both measure the average magnitude of the errors and are frequently used in forecasting applications. MAE is a linear score which weights individual error equally. For the case of the MSE, the errors are squared before averaging; it gives higher weights to large errors. This indicates that the MSE is more useful when large errors are particularly undesirable, as in the case of solar power forecasting.

2. Method

Given data $(\boldsymbol{x}^i, y_i)$, $i = 1, \cdots, n$, where $\boldsymbol{x}^i = (x_{i1}, \cdots, x_{ip})^\top$ are the p predictor variables and y_i are the responses, the linear regression model has the form:

$$y_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} \tag{4}$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, \cdots, \beta_p)^\top$ is the regression parameter. The lasso estimate of $\boldsymbol{\beta}$ is defined by:

$$\begin{aligned} \widehat{\boldsymbol{\beta}} &= \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \left\{ \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 \right\}, \\ &\text{s.t.} \sum_{j=1}^{p} |\beta_j| \leq t \end{aligned} \tag{5}$$

where $t \geq 0$ is a tuning parameter which controls the amount of shrinkage. Eq. (5) is equivalent to the ℓ_1-penalized regression problem of finding:

$$\widehat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \left\{ \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right\} \tag{6}$$

where λ is a tuning parameter which regulates the strength of the penalty (Tibshirani, 1996).

Minimizing the sum of squares part of Eq. (6) gives the ordinary least squares (OLS) estimate. The OLS estimates have low bias but high variance. It can be shown that by removing predictors from the full least squares model, the variance of the estimated response is reduced with an increased bias as trade-off (Miller, 2002). Furthermore, the model accuracy can often be improved by shrinking or setting some coefficients to zero. Therefore, regression shrinkage and selection is a useful tool when the aim is to predict the response variable accurately.

Beside the lasso, ridge regression and subset selection (such as the stepwise regression) are also frequently used to improve OLS. Ridge regression penalizes the sum of squares using an ℓ_2-penalty, i.e., by changing the $|\beta_j|$ terms in Eqs. (5) and (6) to β_j^2. One of the advantages of the ridge regression is its stability, i.e., the ridge regression estimates are little affected by small changes in the regression inputs. Subset selection methods select a subset of predictors, the OLS is then used to estimate the regression coefficients of the predictors that are retained. The advantage of subset selection methods is its interpretability, i.e., the irrelevant predictors are excluded from the model. However, it is noted that the lasso model is interpretable like the subset selection method; it also has the stability of the ridge regression (Tibshirani, 1996). The above discussion can be inferred from a geometrical point of view; we refer the readers to (Hastie et al., 2009; Efron et al., 2004).

The most interesting property of the lasso comes from its ℓ_1-penalty, which often shrinks some regression coefficients to exactly zero. This property suits our application where down-wind stations should have minimal influence, if not no influence, on the measurements at up-wind stations. As uncorrelated predictors only contribute to the variance of the OLS estimates but not the accuracy, they therefore should not be included in the model. By selecting a proper t, the uncorrelated predictors can be truncated. The standard way to select t is k-fold cross validation (Efron et al., 2004); we use the k-fold cross validation (CV) in this work. Alternatively, information criteria can be used (see Zou et al., 2007; Tibshirani and Taylor, 2012). We use the implementation by Hastie and Efron (2013) for lasso computation. The library is implemented in the statistical software R (R Core Team, 2014); we list the key steps here:

1. Coefficients β_j are set as zeros at the start.

2. Find the predictor x_j with highest absolute correlation with y.
3. Increase the coefficient β_j in the direction of the sign of its correlation with y, until some other predictor x_k has as much correlation with r as x_j has. r is the residual, $r = y - \widehat{y}$.
4. Increase $(\beta_j,\ \beta_k)$ in their joint least squares direction, until some other predictor x_m has as much correlation with the residual r.
5. The algorithms stops when all the predictors are in the model.

For further understanding on the shrinkage and selection procedures of the lasso, we refer the readers to (Efron et al., 2004; Tibshirani, 2014).

Two families of time series models, namely, the exponential smoothing (ETS) family of models and the autoregressive integrated moving average (ARIMA) family of models are used to benchmark the lasso regression model. The implementations for these models follow our previous works (Dong et al., 2013; Yang et al., 2012). Systematic descriptions on the ETS model and the ARIMA model can be found in the books by Hyndman et al. (2008) and Box et al. (1994) respectively. The clearness persistence model (see Marquez and Coimbra, 2012, for definition) is included as the baseline model. The above benchmarks are univariate models, i.e., using data from a single station. Therefore, the OLS is used as a multivariate benchmarking model.

3. Results from a single day with a single forecast horizon

Throughout this section, only the 10 second averaged data from a single day, namely, 2010 July 31, is used. After applying the data filters described in section 1.1, 4133 data points are obtained for each station. A total of 5 case studies are presented in this section.

3.1. The effect of parameter shrinkage

To demonstrate the shrinkage effect of the lasso, we consider a forecasting example at station DH4. In this example, the first 2066 data points (50%) are used for fitting and the remaining data points are used for validation. Two stations, namely, DH5 (up-wind station) and DH6 (down-wind station), are used as the spatial neighbors of the forecast station. For each spatial neighbor, time series up to lag-3 are used. Together with the autocorrelated time series from DH4 itself, we have a lasso regression with design matrix $\boldsymbol{x}$ with $p = 9$ and $n = 2063$. Note that the first 3 training samples are used to produce the lagged time series, n for the design matrix is therefore reduced by 3.

Recall Eq. (5), the parameter t controls the amount of shrinkage. When t is large, the constraint on the ℓ_1-norms loses its effect. More specifically, when $t > \sum_j \widehat{\beta}_j^o$, where $\widehat{\beta}_j^o$ is the full OLS estimates, the lasso estimates $\widehat{\beta}_j$ is equal to $\widehat{\beta}_j^o$. To visualize the effect of t, we define the shrinkage factor:

$$s = t \Big/ \sum_{j=1}^{p} \left|\widehat{\beta}_j^o\right| \tag{7}$$

For any $0 < s \leq 1$, the corresponding solution of the lasso can be found. The solution path (the collection of all solutions for $0 < s \leq 1$) of the lasso can be found using the algorithm shown in section 2. Fig. 2 shows the solution path of the lasso. It can be seen that predictor DH5[1] (the lag-1 time series at station DH5) enters the model first. This agrees with the physical understanding, namely, the up-wind station has an effect on station DH4. On the other hand, predictor DH6[3] enters the model at a later stage with small coefficients throughout, indicating that this predictor has a small effect on the response variable. If we select s anywhere along the scale, the shrinkage effect is obvious. For example, if we choose $s = 0.35$, only two predictors, DH5[1] and DH4[1], will be used to predict the clearness index at DH4; their coefficients are given by the intersections of the solution path and the dashed vertical line, as shown in Fig. 2. Other predictors have zero coefficients.

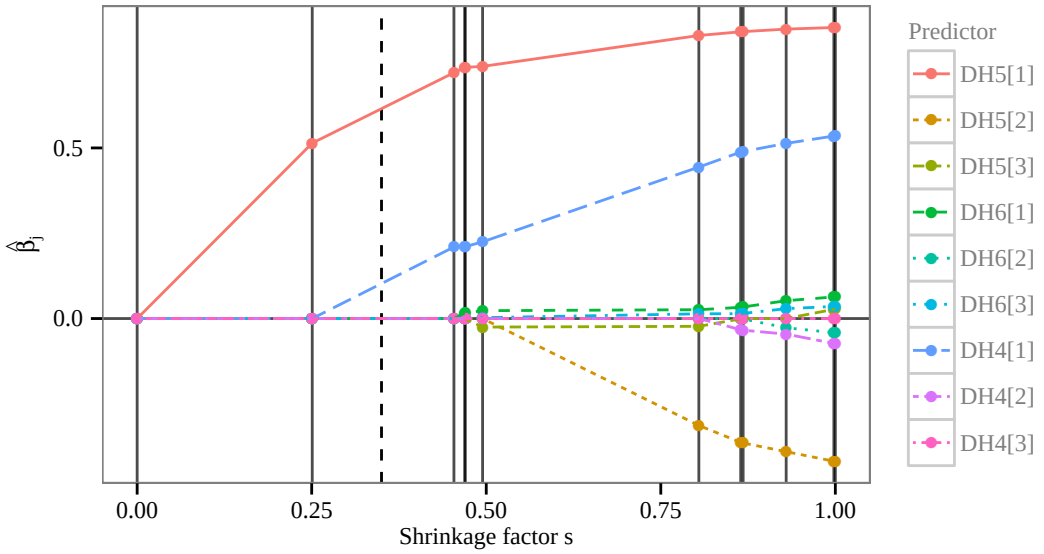

Figure 2: Estimates of lasso regression coefficients $\widehat{\beta}_j$, $j = 1, \cdots, 9$, for the case study in section 3.1. The covariates enter the regression model as s increases (indicated by the solid vertical lines). For $s = 1$, the lasso gives OLS estimates. The predictors are lagged time series, for example, DH5[1] corresponds to the lag-1 time series at station DH5. The response is the time series at station DH4. See text for the interpretation of the dashed vertical line.

3.2. Case study 1: forecast DH4 using 9 predictors

Once the lasso solution path is calculated and the best s value is determined by CV, we can fit the model using new predictor values (the remaining 50% of data). As all the predictors are lagged variables, the forecast of the response variable can be readily obtained. To benchmark the lasso, the persistence, ARIMA and ETS models are used. The ARIMA and ETS models use the same training length as the lasso; the forecasts are produced for the remaining data using the trained models. The persistence is only evaluated for the testing data. Beside the univariate (single-sensor) models, full OLS (the $s = 1$ case) is also used to benchmark the lasso. Suppose we have response vector $\boldsymbol{y}$ and design matrix $\boldsymbol{X}$, the OLS estimates are:

$$\widehat{\boldsymbol{\beta}}^o = \left(\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1} \boldsymbol{X}^\top \boldsymbol{y} \tag{8}$$

The nMAEs for the persistence, ETS, ARIMA, OLS and lasso models are 7.93%, 7.93%, 8.68%, 4.20%, 4.18%, respectively. The forecast skills (FS) for the ETS, ARIMA, OLS and lasso models are 0.00, 0.03, 0.49, 0.49 respectively. It is observed that the OLS and lasso produce significantly better forecasts than the univariate models. It is also found that the choice of error metric can lead to different conclusions on forecast results; the ARIMA model is worse than persistence in terms of nMAE but is better in terms of FS. We note that the predictive performance of the lasso is similar to the OLS in this example. However, as shown in later examples, when the number of predictors gets large and/or the training data are insufficient, the OLS produces unacceptable results.

3.3. Case study 2: forecast DH4 using various training data lengths

The above toy example assumes a fixed training length. If 50% of data are used for training in each day, it would not be acceptable for operational forecasting. Therefore, we investigate the effect of training length on forecast accuracy in this case study. Forecasts at DH4 using the lasso and OLS models with various training data lengths are evaluated. The results from the ETS and ARIMA models are omitted. Previously, the choice of 9 predictors is only made for simplicity, i.e., to make Fig. 2 readable. In this case study, a

full network model with $p = 170$ is considered, i.e., time series up to lag-10 from each of the 17 stations are used. The results are depicted in Table 1.

Table 1: The performance of the lasso and OLS models for various training data lengths (in % of total number of data points). 10 second averaged data from 2010 July 31 are used.

	nMAE [%]			forecast skill	
% train	pers	ols	lasso	ols	lasso
10	8.36	12.16	7.73	-0.06	0.39
20	8.65	8.89	4.39	0.23	0.54
30	8.67	5.59	4.16	0.49	0.57
40	8.20	5.05	3.99	0.51	0.56
50	7.98	4.89	3.95	0.51	0.56

It can be concluded from Table 1 that the lasso outperforms the OLS method for all training lengths. The accuracy of the OLS model reduces when the training data become fewer. Furthermore, even when the data are sufficient (such as the 50% case), the OLS still performs worse than the lasso. This is due to the large number of irrelevant predictors in the OLS model. On the other hand, the accuracy reduction in the lasso is marginal (except for the 10% case). These observations align with the discussions in Chapter 1 of the book by Miller (2002). In a regression problem with many predictors, when the data are few, the regression coefficients given by the OLS will be poorly determined, and the predictions also tend to be poor. Fig. 3 shows the distributions of regression coefficients fitted using the lasso and OLS for each training length. We note that Gaussian kernel is used in the plot for visualization. In each subplot, 170 red dots are plotted with some overlay, representing the 170 estimated regression coefficients. For the 10% case, the lasso identifies 4 predictors with larger coefficients while the remaining predictors have near zero coefficients. In contrast, the OLS produces a distribution of coefficients with a wide spread. As the length of the training data increases, the coefficients determined by the OLS become similar to those determined by the lasso. Consequently, the forecast errors of the OLS are comparable to the errors of the lasso for longer training lengths.

In addition to the analyses above, the time series plots and the scatter plots of the forecasts (the 20% case) are shown in Fig. 4. The top plot shows that the lasso improves the forecasts significantly. There is no time lag between the forecast and measured time series, which is otherwise unachievable using any univariate statistical method. The bottom plot of Fig. 4 shows that the lasso can reduce the variance of the forecast significantly, thus may result in narrower confidence intervals for interval-based forecasts.

We note that for operational forecasting, the problem of training length can be relaxed thanks to possible similarities in meteorological conditions. In other words, when the present day's meteorological conditions are similar to the conditions in some historical days, previously trained models can be readily applied to the present forecasts. Furthermore, we can adaptively update the model within a day when data become available. Such enhancements to the method are not discussed in this work.

3.4. Case study 3: forecast DH4 using various numbers of predictors

In this case study, we verify the effect of number of predictors on the lasso. Based on the results from case study 2, 20% of the data are used for training and the remaining data are used for testing. We can vary p by either adjusting the number of spatial neighbors or adjusting the number of lagged series (the temporal neighbors). The relationship is written as $p = (n_s + 1) \times n_t$, where n_s and n_t are numbers of spatial and temporal neighbors respectively. The $+1$ term indicates that the autocorrelated time series are used as predictors as well. The nMAE and FS of the lasso are shown in Fig. 5; the results from the benchmarking models are omitted.

The number of spatial neighbors in Fig. 5 are incremented sequentially following the increasing distance to station DH4, i.e., DH3 (nearest to DH4) is added first and AP6 (furthest from DH4) is added last. We note that $n_s < 3$ cases produce larger errors; to visualize the color contrast among other combinations, these cases are excluded from the plot. As the forecast results do not differ much from each other for different p, no obvious conclusion can be established regarding the best combination of n_s and n_t. In other words, once

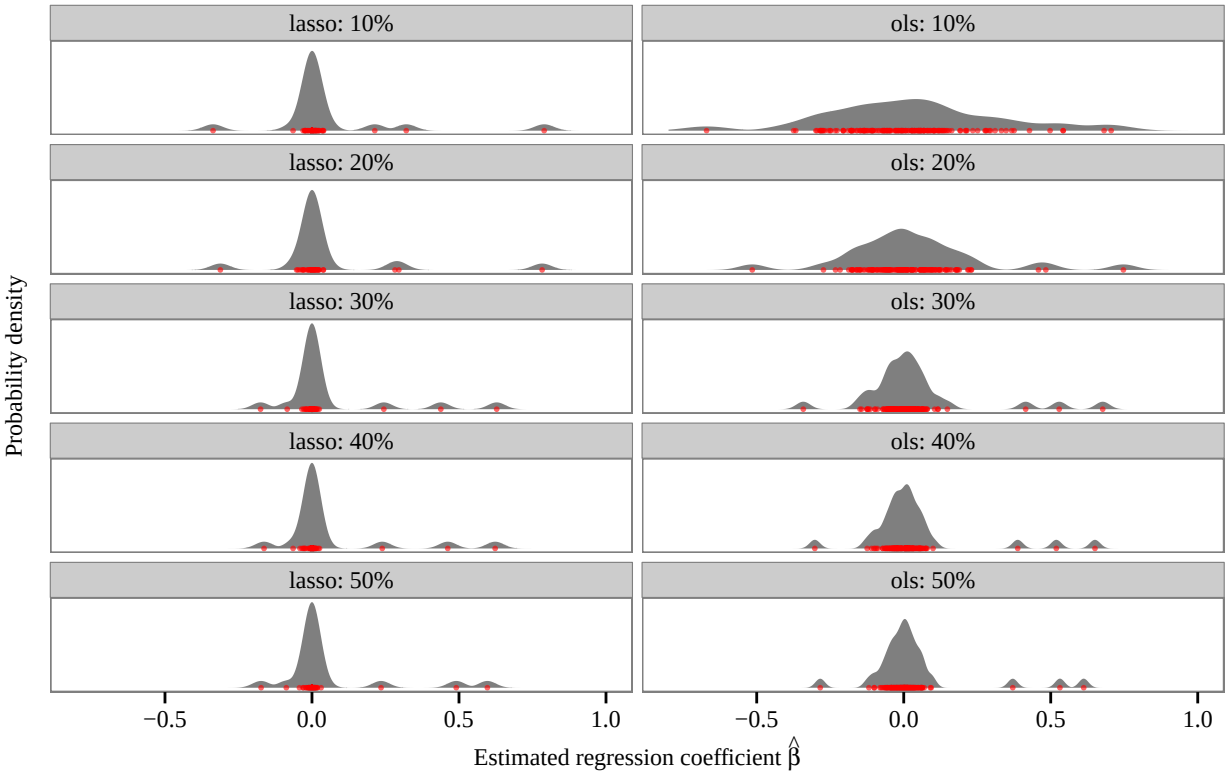

Figure 3: Distributions of regression coefficients using the lasso and OLS for various training data lengths (in % of total number of data points). In each subplot, there are 170 red dots (with some overlay) representing 170 regression coefficients, i.e., $p = 170$. Gaussian kernel is used for density estimation and visualization.

the most correlated predictors (DH5[1] in this case) are added into the lasso, the forecast does not improve substantially by adding in other predictors. It may also be noted that the distributions of nMAE and FS in Fig. 5 seem to possess certain geometrical patterns. At this stage, we hypothesize that these patterns are due to spatial and temporal frequency of the data. Further investigations may apply.

3.5. Case study 4: forecast all stations using preselected predictors (with known wind information)

In section 3.4, we showed that the irradiance measurements from the up-wind stations are essential to make good forecasts. It is therefore logical to select n_s and n_t based on the prior knowledge on wind speed and direction. We consider the wind speed u and timescale $\bar{t}$. Suppose Ω is the set of all up-wind stations to an arbitrary station $\boldsymbol{s}_0$, i.e., $\Omega = \{\boldsymbol{s}_j : \boldsymbol{s}_j \in \text{up-wind stations}\}$, then the following rules should apply:

$$n_s = \text{card}(\Omega) \tag{9}$$

$$n_t = \max\left(\zeta, \left\lceil \frac{\max(d_{0j})}{u\bar{t}} \right\rceil\right), \quad \boldsymbol{s}_j \in \Omega \tag{10}$$

where $\text{card}(\cdot)$ is the cardinality of the set; ζ is some positive integer which will be explained shortly; d_{0j} is the along-wind distance between the forecast station $\boldsymbol{s}_0$ and an up-wind station $\boldsymbol{s}_j$. Notation $\lceil\cdot\rceil$ is the ceiling operator; it is used because n_t can only take discrete values. We note that this preselection method is similar to the heuristic proposed by Yang et al. (2014a). It was shown that the preselection (shrinkage)

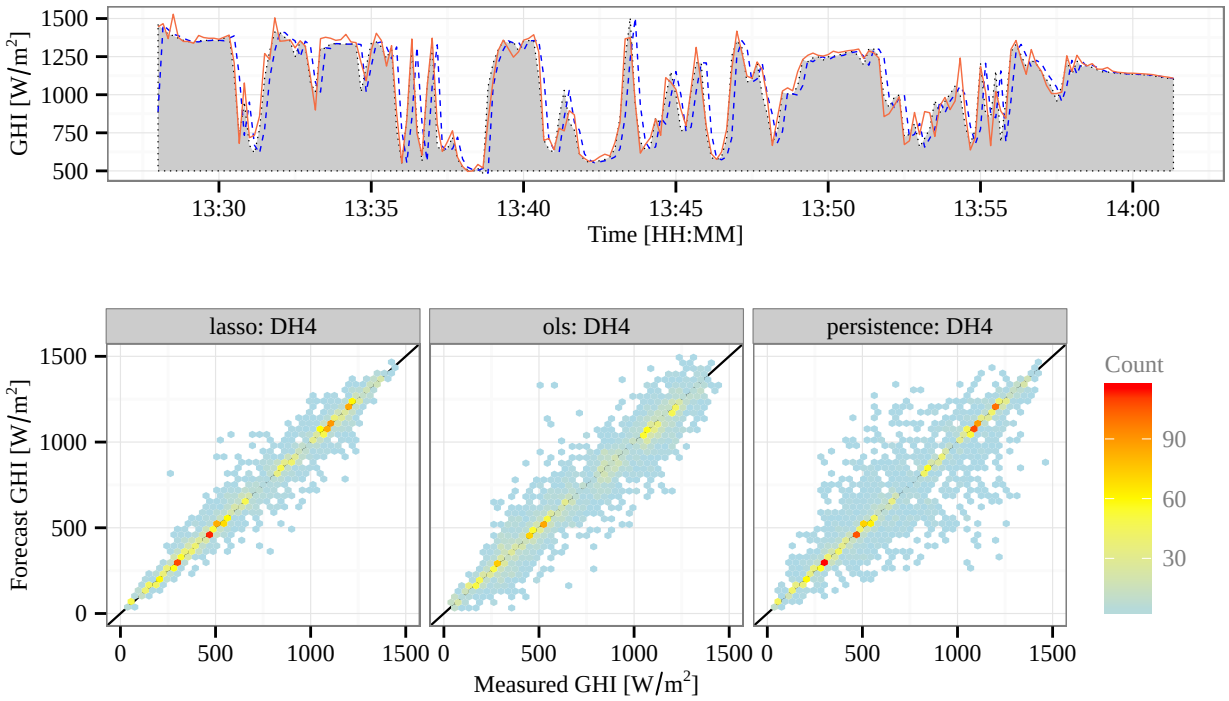

Figure 4: (top) Forecasts using the lasso (orange solid line) and measured GHI (black dotted line) at station DH4 for a period on 2010 July 31. 20% of the data are used for training, the remaining 80% are used for evaluation. Persistence model is represented by the blue dashed line. The utilization of the data from the neighboring stations improves the forecasts significantly. (bottom) Scatter plots of the same forecasts. Forecasts using the lasso have significant improvements over both the OLS and persistence models. Hexagon binning algorithm is used for visualization.

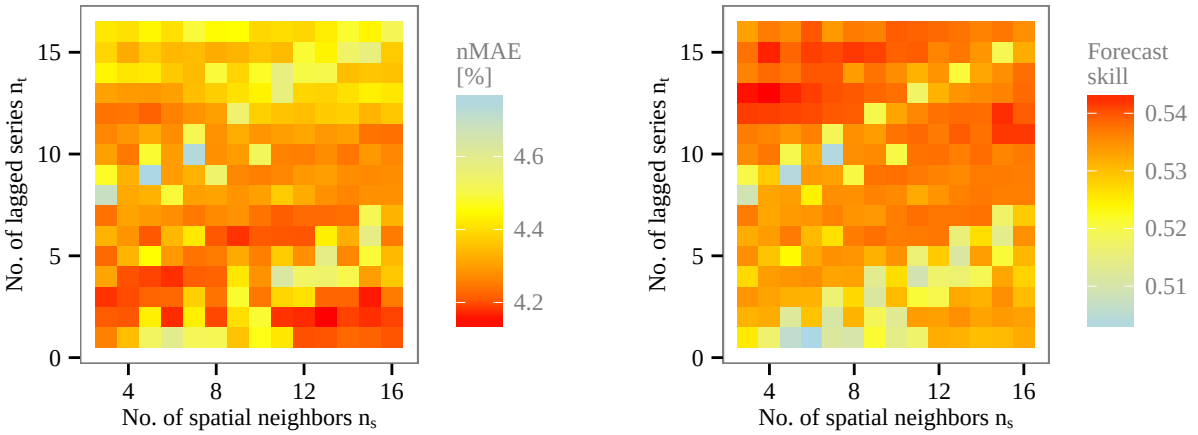

Figure 5: Lasso nMAE and forecast skill plotted against various combinations of n_s and n_t for station DH4, on 2010 July 31. 10 second averaged data are used. The spatial neighbors are added sequentially based on distance to DH4; the nearest station is added first.

can improve the forecast from the models which consider the full set of spatio-temporal neighbors (Yang et al., 2014a).

Eq. (9) ensures that all the up-wind stations are included as lasso predictors. Eq. (10) ensures the moving clouds have sufficient time to reach the forecast station. Moreover, when $n_t < \zeta$, we manually force it to be ζ, to include sufficient number of autocorrelated predictors. For example, for station AP7, it has no up-wind station, therefore its $n_s = 0$ and $n_t = \zeta$. It should be noted that when ζ is large enough, Eq. (10) loses its effect. We set $\zeta = 3$ for illustration. Hinkelman (2013) showed that the average inferred wind speed for the 13 selected days is 10 m/s, i.e., $u = 10$. Furthermore, in this section, $\bar{t} = 10$ s. Therefore, as another example, station DH8 should have $n_s = 16$ and $n_t = \lceil 1046/(10 \times 10) \rceil = 11$, where 1046 m is the along-wind distance between DH8 and AP7. With these assumptions, we use the lasso and other benchmarking models to predict the irradiance observed at all the stations on 2010 July 31. The training and testing data follows section 3.1, namely, 20% and 80% respectively. The nMAE and FS for various models are shown in Table 2.

Table 2: The nMAE [%] and FS of the forecasts for 2010 July 31. 10 second average data are used. The lasso implementations using Eqs. (9) and (10) are benchmarked using the persistence (pers), exponential smoothing state space (ets), the autoregressive integrated moving average (arima) and the ordinary least squares (ols) models.

	nMAE [%]					forecast skill			
station	pers	ets	arima	ols	lasso	ets	arima	ols	lasso
AP1	8.62	9.01	8.92	8.51	7.17	-0.02	0.00	0.17	0.25
AP3	9.00	10.15	10.17	9.03	8.51	-0.09	-0.05	0.12	0.15
AP4	9.16	11.20	10.03	10.36	10.15	-0.18	-0.03	-0.01	0.01
AP5	8.50	9.19	9.30	9.68	8.19	-0.04	-0.02	0.11	0.19
AP6	8.36	9.84	8.89	9.96	9.05	-0.12	-0.02	-0.04	0.01
AP7	9.85	11.48	11.08	10.20	10.15	-0.13	-0.04	-0.01	0.00
DH1	8.25	9.67	8.67	11.34	8.10	-0.13	-0.01	-0.07	0.12
DH2	8.70	9.62	9.83	10.93	9.33	-0.07	-0.04	-0.02	0.08
DH3	8.86	9.30	9.40	6.46	5.21	-0.03	-0.02	0.43	0.53
DH4	8.65	9.29	9.02	6.02	4.41	-0.04	0.00	0.44	0.54
DH5	8.55	9.36	8.86	10.67	9.16	-0.06	0.01	-0.01	0.09
DH6	8.59	9.10	9.33	8.53	5.89	-0.02	-0.03	0.25	0.44
DH7	8.44	9.31	10.20	6.56	4.92	-0.07	-0.07	0.40	0.51
DH8	8.49	9.29	9.06	9.29	5.75	-0.04	-0.01	0.20	0.44
DH9	8.79	9.51	9.07	8.55	6.46	-0.04	0.01	0.28	0.39
DH10	8.62	9.13	9.14	7.75	4.71	-0.03	0.00	0.34	0.56
DH11	8.42	8.94	8.58	12.98	7.61	-0.02	0.02	-0.15	0.20
average	8.70	9.61	9.39	9.22	7.34	-0.07	-0.02	0.14	0.27

Table 2 shows that the lasso outperforms the univariate models significantly; its improvement from the OLS model is also evident. This indicates that although the preselection using Eqs. (9) and (10) could reduce the number of potential predictors by excluding the down-wind stations, the lasso can further shrink and select the remaining predictors. In a later section of the paper, we show that the improvements made using the lasso from using the OLS are more significant for longer forecast horizons.

3.6. Case study 5: forecast all stations using 170 predictors (with unknown wind information)

The case study in section 3.5 considers the wind information. When wind information is unknown, sufficiently large n_s and n_t can be assumed based on expert view. For instance, we can assume the full network models used in case study 2, i.e., $n_s = 16$, $n_t = 10$ and $p = 170$. The nMAE and FS for the lasso and the OLS models are shown in Table 3. It is observed that the average nMAE and FS of the OLS models in Table 3 are much worse than those in Table 2. This is because that case study 5 contains more irrelevant predictors due to unknown wind information; the OLS models thus have larger variances. On the other hand, the performance of the lasso models in case study 5 is consistent with the earlier case study, indicating effective shrinkage and selection.

It is evident from the errors reported in Tables 2 and 3 that the lasso performs well at the center stations (e.g., DH3, DH4 and DH10) where suitable predictors can be found from lagged time series from the peripheral stations. However, for stations located at the boundary of the sensor grid, due to lack of

Table 3: Same as Table 2, except that the lasso assumes unknown wind information. The number of predictors $p = 170$ are used in both the lasso and OLS models.

	nMAE [%]		forecast skill	
station	ols	lasso	ols	lasso
AP1	12.22	7.85	-0.07	0.23
AP3	14.33	8.64	-0.23	0.17
AP4	13.77	9.62	-0.17	0.08
AP5	11.94	8.12	-0.05	0.20
AP6	14.28	9.38	-0.26	0.02
AP7	13.71	10.27	-0.14	0.02
DH1	14.05	8.31	-0.26	0.11
DH2	14.74	9.17	-0.29	0.12
DH3	9.29	5.41	0.22	0.52
DH4	8.89	4.39	0.23	0.54
DH5	13.37	9.28	-0.17	0.10
DH6	8.74	5.81	0.23	0.44
DH7	7.75	4.56	0.32	0.51
DH8	8.83	5.83	0.23	0.44
DH9	9.39	6.42	0.21	0.40
DH10	8.40	4.88	0.29	0.56
DH11	14.61	7.71	-0.27	0.20
average	11.67	7.39	-0.01	0.27

suitable predictors, the lasso may produce higher errors (see AP6 and AP7). We would like to note that if autocorrelated predictors are not used, the results will be worse. As we cannot "infinitely" expand the monitoring network so that an up-wind station can always be found, the best practice is thus to include the autocorrelated time series in the model. In this way, it allows the lasso to possibly reduce to an autoregressive model.

4. Results from all 13 days with various forecast horizons

In the previous section, performance of the lasso along with several benchmarking models is evaluated at a forecast horizon of 10 second for 2010 July 31. In this section, additional forecasting results are shown using data from all 13 selected days with various forecast horizons.

4.1. Case study 6: forecast all stations for all 13 days (with known wind information)

The configurations of this case study are identical to case study 4 in section 3.5, namely, using 10 second averaged data with a training length of 20%. For each day and for each station, autocorrelated time series are included in the lasso; n_s and n_t choices are made using Eqs. (9) and (10). Forecast skills of the lasso and OLS are shown in Table 4 with nMAE and the results of the univariate models omitted. By examining the average errors, it is observed that the lasso performs better than the OLS for all 13 selected days. We also observe that for most days, the boundary stations produce small FS due to lack of suitable spatial neighbors. Furthermore, it is found that for various days, the accuracies of the lasso can be very different. For examples, on September 7, many stations yield negative FS, whereas good forecasts are observed on July 31. We explain the results from a statistical point of view as follows.

Recall the lasso procedure shown in section 2, the correlations between the predictors and the residuals are considered at each step. We thus note that the performance of our lasso application depends on the spatio-temporal correlation structure of the clearness index. Similar to a purely spatial correlation structure, a spatio-temporal correlation structure describes not only the spatial cross-correlation (correlation between data collected at two sites), but also the temporal cross-correlation (correlation between lagged data collected at two sites). In matrix form, for n stations and m maximum lags, the empirical spatio-temporal correlation structure can be written as:

$$\mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma}_0 & \mathbf{\Sigma}_1 & \cdots & \mathbf{\Sigma}_{m-1} \end{pmatrix}^\top, \quad \in \mathbb{R}^{nm \times n} \tag{11}$$

Table 4: Evaluation of the forecast skill for the 13 days selected days expect for 2010 July 31. The lasso and OLS settings follow section 3.5. Bold numbers indicate better OLS forecasts.

	Jul 31		Aug 1–5										Aug 21	
station	ols	lasso	ols	lasso	ols	lasso	ols	lasso	ols	lasso	ols	lasso	ols	lasso
AP1	0.17	0.25	0.27	0.30	0.20	0.20	0.23	0.27	0.20	0.26	0.11	0.16	0.12	0.13
AP3	0.12	0.15	0.18	0.20	**0.20**	0.19	0.16	0.18	0.23	0.25	0.09	0.10	0.06	0.06
AP4	-0.01	0.01	0.06	0.07	0.07	0.07	0.07	0.08	0.07	0.07	0.05	0.06	0.08	0.08
AP5	0.11	0.19	0.21	0.25	0.17	0.25	0.25	0.28	0.14	0.18	-0.01	0.08	0.10	0.16
AP6	-0.04	0.01	-0.03	0.04	0.05	0.05	0.03	0.05	-0.02	0.02	0.04	0.07	0.03	0.06
AP7	-0.01	0.00	0.03	0.03	0.04	0.04	0.03	0.03	0.03	0.03	0.01	0.01	0.00	0.00
DH1	-0.07	0.12	-0.04	0.13	-0.03	0.12	0.00	0.10	-0.01	0.14	-0.14	0.08	0.00	0.10
DH2	-0.02	0.08	0.01	0.11	0.06	0.07	0.06	0.11	-0.01	0.09	-0.05	0.06	0.04	0.10
DH3	0.43	0.53	0.49	0.54	**0.46**	0.45	0.50	0.54	0.55	0.58	0.36	0.40	0.52	0.56
DH4	0.44	0.54	0.47	0.46	0.38	0.43	0.55	0.59	0.50	0.58	0.30	0.36	0.48	0.48
DH5	-0.01	0.09	0.05	0.11	0.09	0.10	0.10	0.13	0.15	0.16	0.01	0.07	0.05	0.07
DH6	0.25	0.44	0.27	0.45	0.29	0.44	0.42	0.50	0.42	0.53	0.09	0.23	0.09	0.14
DH7	0.40	0.51	0.37	0.55	0.36	0.46	0.45	0.51	0.34	0.47	0.13	0.25	**0.30**	0.29
DH8	0.20	0.44	0.09	0.47	0.21	0.50	0.31	0.43	0.29	0.41	0.00	0.20	0.07	0.18
DH9	0.28	0.39	0.26	0.35	0.22	0.37	0.35	0.43	0.36	0.42	0.03	0.15	0.12	0.11
DH10	0.34	0.56	0.47	0.54	0.40	0.46	0.52	0.57	0.46	0.54	0.18	0.26	0.52	0.53
DH11	-0.15	0.20	-0.11	0.19	-0.13	0.19	0.06	0.22	-0.03	0.18	-0.09	0.12	-0.05	0.09
average	0.14	0.26	0.18	0.28	0.18	0.26	0.24	0.29	0.22	0.29	0.06	0.16	0.15	0.19
	Aug 29		Sep 5–7						Sep 21		Oct 27		average	
station	ols	lasso	ols	lasso	ols	lasso	ols	lasso	ols	lasso	ols	lasso	ols	lasso
AP1	0.13	0.20	0.08	0.11	0.06	0.11	-0.09	-0.14	0.13	0.14	0.35	0.36	0.15	0.18
AP3	0.09	0.11	0.07	0.07	0.08	0.08	-0.01	-0.01	0.11	0.12	0.24	0.25	0.13	0.13
AP4	0.06	0.06	**0.08**	0.07	0.06	0.06	0.00	0.01	0.08	0.08	0.07	0.08	0.06	0.06
AP5	0.10	0.18	-0.05	0.13	-0.07	-0.03	-0.30	-0.11	-0.03	-0.05	0.29	0.32	0.07	0.14
AP6	0.00	0.04	-0.01	0.02	0.05	0.06	**0.01**	0.00	0.01	0.02	0.00	0.01	0.01	0.04
AP7	0.00	0.00	0.06	0.06	0.03	0.03	0.04	0.04	0.06	0.06	0.01	0.01	0.03	0.03
DH1	-0.02	0.15	-0.15	0.04	-0.13	-0.01	-0.84	-0.01	-0.10	0.00	-0.17	0.13	-0.13	0.08
DH2	0.04	0.12	-0.02	0.04	0.02	0.04	-0.26	-0.14	0.04	0.08	-0.01	0.07	-0.01	0.06
DH3	0.39	0.45	0.32	0.37	0.16	0.19	**0.06**	0.01	0.35	0.36	**0.56**	0.55	0.40	0.42
DH4	0.38	0.43	0.16	0.29	0.02	0.20	-0.14	0.15	0.23	0.27	0.49	0.56	0.33	0.41
DH5	0.07	0.12	0.05	0.08	-0.26	-0.19	-0.08	-0.03	0.04	0.05	0.17	0.22	0.03	0.08
DH6	0.38	0.48	-0.02	0.12	-0.61	-0.16	-0.53	-0.05	-0.20	-0.13	0.50	0.63	0.10	0.28
DH7	0.31	0.44	0.03	0.20	-0.08	0.01	-0.35	-0.31	0.05	0.06	0.57	0.64	0.22	0.31
DH8	0.30	0.40	-1.04	0.16	-0.71	-0.06	-1.13	-0.14	-0.17	0.14	0.23	0.55	-0.10	0.28
DH9	0.20	0.39	-0.14	0.14	-0.52	-0.05	-0.57	-0.14	-0.13	0.01	0.50	0.57	0.07	0.24
DH10	0.26	0.42	0.08	0.26	0.01	0.23	-0.50	-0.10	0.18	0.23	0.46	0.57	0.26	0.39
DH11	-0.01	0.16	-0.43	0.03	-0.51	-0.10	-1.58	0.02	-0.31	0.06	0.01	0.25	-0.26	0.12
average	0.16	0.24	-0.06	0.13	-0.14	0.02	-0.37	-0.06	0.02	0.09	0.25	0.34	0.08	0.19

where $\mathbf{\Sigma}_\tau$ represents the spatial submatrix at time lag τ:

$$\mathbf{\Sigma}_\tau = \begin{pmatrix} \Sigma_{11,\tau} & \Sigma_{12,\tau} & \cdots & \Sigma_{1n,\tau} \\ \Sigma_{21,\tau} & \Sigma_{22,\tau} & \cdots & \Sigma_{2n,\tau} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{n1,\tau} & \Sigma_{n2,\tau} & \cdots & \Sigma_{nn,\tau} \end{pmatrix}, \quad \in \mathbb{R}^{n\times n} \tag{12}$$

where $\Sigma_{ij,\tau}$ denotes the lag τ empirical correlation between stations i and j. We note that the correlation matrix $\mathbf{\Sigma}$ in Eq. (11) follows a kriging formulation (Yang et al., 2013b).

Following Eqs. (11) and (12), spatio-temporal correlation matrices for 2010 July 31 and September 7 are shown in Fig. 6, where m is set as 5 for illustration purpose. Each matrix therefore has dimension 85×17 with spatial submatrices stratified by gray dashed lines. To visualize the correlations, we arrange the stations $i = 1, \cdots, n$ in the along-wind direction, i.e., station-1 corresponds to AP7 while station-17 corresponds to DH8. By observing the spatial submatrices in the July 31 plot, directional property can be clearly identified as the correlations from the lower triangular regions (along-wind) are much stronger than the ones in the upper triangular regions (against-wind). However, the directional effects on September 7 are

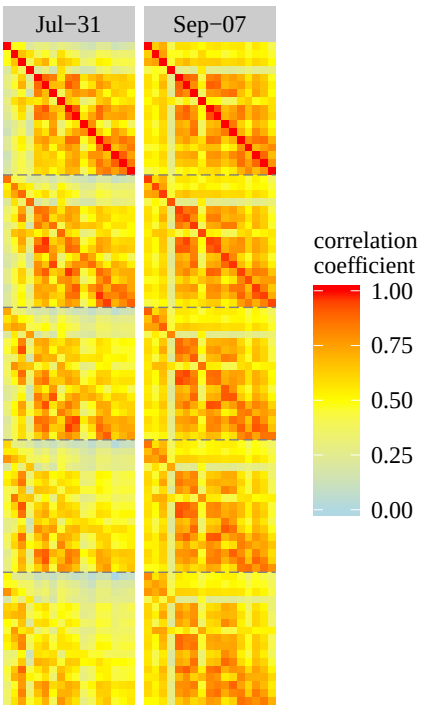

Figure 6: Spatio-temporal correlation structures on 2010 July 31 and September 07. Abscissa and ordinate of each subplot are the row and column index of the matrix $\mathbf{\Sigma}$ in Eq. (11) respectively, with $n = 17$ and $m = 5$. Asymmetric spatial submatrices can be seen on July 31 plot, indicating a strong evidence for along-wind correlation. Directional correlation is less obvious for the case of September 07.

present but are less significant. As mentioned earlier, the lasso selects a new predictor based on correlation in each step, the proximities of correlation between the residual and each candidate can make the selection degenerate. The forecasting results for September 7 are therefore worse than the results from other days.

4.2. Case study 7: forecast all stations with various forecast horizons (with known wind information)

The last case study in this paper evaluates performance of the lasso at various forecast horizons. Data from the 13 selected days are first averaged into 10, 20, 30, 40, 50, 60, 120, 180 and 300 second intervals. For each set of the averaged data, one-step-ahead forecasts using the lasso, OLS, ETS and ARIMA are performed. Case study 2 shows that the performance of the OLS improves substantially when the training data length increases. Furthermore, case studies 4 and 5 show that the performance of the OLS can be enhanced by preselecting the up-wind spatial neighbors. A pro-OLS formulation is considered in this case study, i.e., we assume the wind information is known and use 50% of the data for training.

The forecast skill of the lasso is plotted against the forecast horizon in Fig. 7. Forecast skills of the OLS, ETS and ARIMA models are shown using other types of lines. We can conclude from Fig. 7 that the lasso in general has a better performance than the benchmarking models for all forecast horizons. For sub-1-minute horizons, the lasso performs well over the univariate models. On the other hand, for $fh = 300$s, all the methods (except for the OLS) are comparable to the persistence model. Another observation which can be made is that the forecast skills of the leading (upwind) and the lagging (downwind) stations are significantly different at smaller forecast horizons. For the stations without any leading station, namely, AP4, AP6, AP7, DH1 and DH2, the lasso gives comparable results to the persistence model. On the other hand, even with the presence of only one leading station, as in the cases of AP3, AP5 and DH5, the lasso can effectively

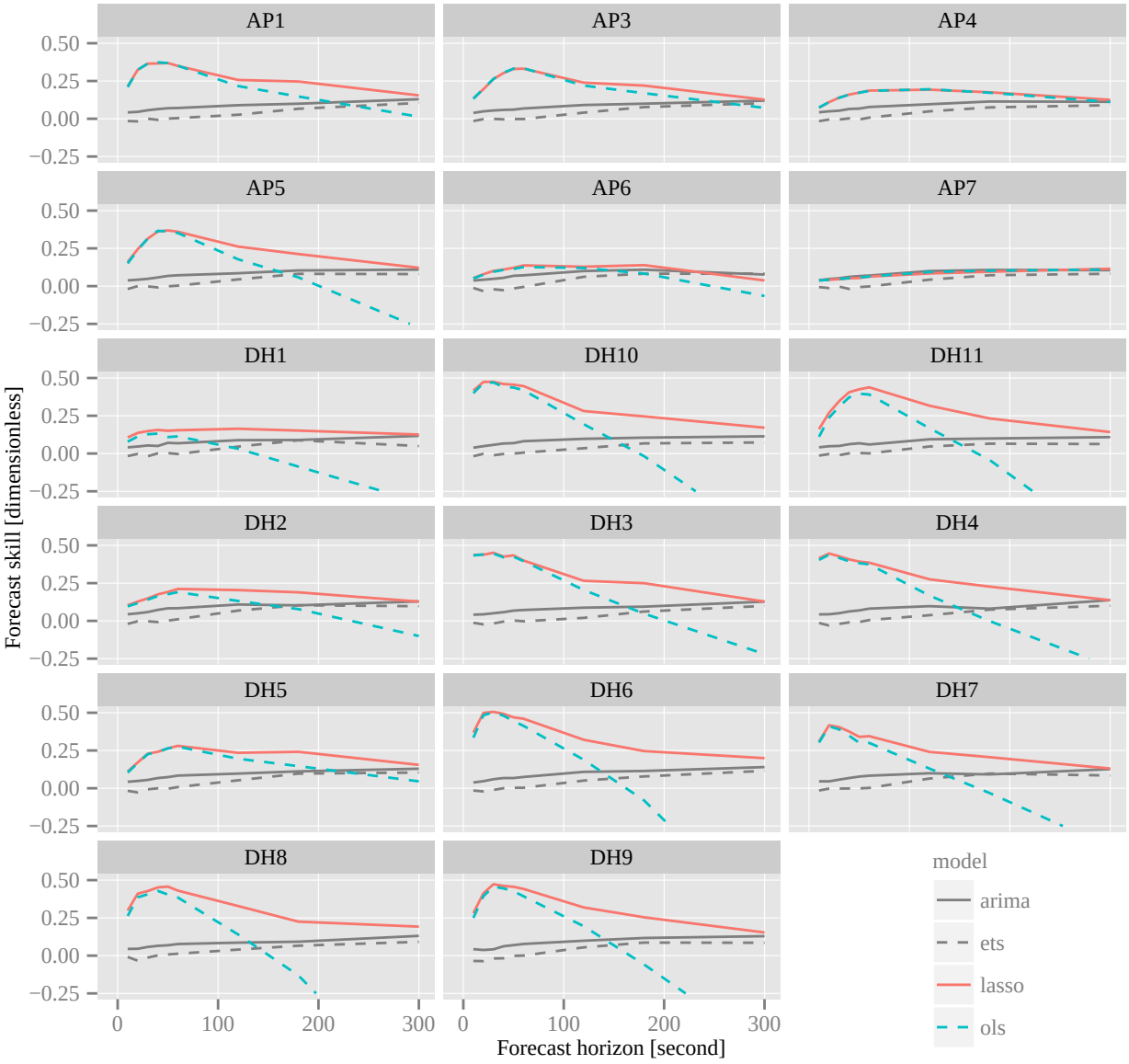

Figure 7: Average forecast skill of the lasso regression method during the 13 selected days at each station.

pick up the relevant predictors and show superiority. On the contrary, although a pro-OLS formulation is used here, OLS still performs badly for $fh > 60$. We note that when wind information is assumed to be unknown and/or fewer data are used for training, OLS is more likely to produce unacceptable results due to the degeneracies in the predictors (see Appendix A for more details).

5. Conclusions

A very short-term irradiance forecasting method is proposed. The lasso is used to shrink and select the spatio-temporal neighbors from lagged time series collected by a dense network of monitoring stations. Due

to the presence of highly correlated data from the along-wind station pairs, the forecast results improve significantly from persistence and other univariate time series methods. The lasso also outperforms the ordinary least squares model. The advantage of the lasso over OLS is more notable when the number of predictors in a regression model is large and/or training data are few.

The lasso method answers the earlier questions in section 1. As the lasso considers correlation intrinsically, magnitudes of the observed correlation are embedded in the lasso procedure. When wind speed and direction change from day to day or within a day, an adaptive model can be considered. In other words, the lasso is iteratively used to identify the most appropriate predictors for a given time period. When the correlation is unobserved at the boundary stations, autocorrelated time series from the station itself can be included as predictors. Such practice allows the lasso to behave similar to an autoregressive model; its performance is thus expected to be no worse than persistence and simple time series models.

Forecasting using the lasso requires a sensor network. The method herein described can be applied to networks with other spatial and temporal scales. However, the performance the lasso is limited by the observed spatio-temporal correlations. In a previous work by Yang et al. (2014a), the lasso was used to forecast the irradiance using a sparse network of 13 stations in Singapore, a 40×20 km island. Due to the low station density, thus low correlations among the stations, the performance of the lasso was shown to be suboptimal. This problem therefore brings the question on applicability of the lasso. Fortunately, as the ground-based irradiance sensing technologies advance, it would soon to be justifiable to install sensor networks with utility scale solar power plants. In addition, reference cells are often installed at plane of array to monitor the PV performance. These tilted data can also be utilized in forecasting by converting them to GHI using inverse transposition models, such as the ones shown in (Yang et al., 2014b, 2013a). Alternatively, in-plane clear sky models can be used to normalize the tilted reference cell measurements, and thus make them useful for forecasting (Lipperheide et al., 2015).

Appendix A. Additional forecasting results

We have shown that the OLS performs worse than the lasso at various forecast horizons in section 4.2. In this appendix, some additional forecast results are provided. Beside using the training length of 50%, other choices including 20%, 30% and 40% are explored with known wind information. Instead of using the forecast skill defined in Eq. (2), we consider a new forecast skill which may suit the situation better. The forecast skill of the lasso with respect to the OLS results is defined as:

$$\mathrm{FS}^*(fh;tl) = 1 - \frac{\mathrm{nRMSE}_{\mathrm{lasso}}(fh;tl)}{\mathrm{nRMSE}_{\mathrm{ols}}(fh;tl)} \tag{A.1}$$

where fh and tl denote the forecast horizon and training length respectively; $\mathrm{nRMSE}_{\mathrm{lasso}}$ and $\mathrm{nRMSE}_{\mathrm{ols}}$ are normalized root mean square errors of the lasso and OLS models. Following this definition, the FS^* values are plotted in Fig. A.8. It can be concluded that the lasso is superior to OLS at all forecast horizons and for all training lengths. It is evident that as the forecast horizon increases, the accuracies of the OLS models drop significantly due to fewer fitting samples (for the same training percentage, the $fh = 300$ cases have fewer training points than the $fh = 10$ cases). Consequently, the FS^* tends to increase as the forecast horizon increases.

Appendix B. Supplementary material

Supplementary data associated with this article can be found, in the online version, at *online URL placeholder*.

We provide the R code used to generate the results shown in Tables 2 and 3. Instead of providing the data (which would then violate the NREL data agreement), we provide the R code used to arrange the data, thus make the code readily executable once R is installed and the raw data are downloaded from the NREL website. The software installation information can be found at `http://www.r-project.org/`.

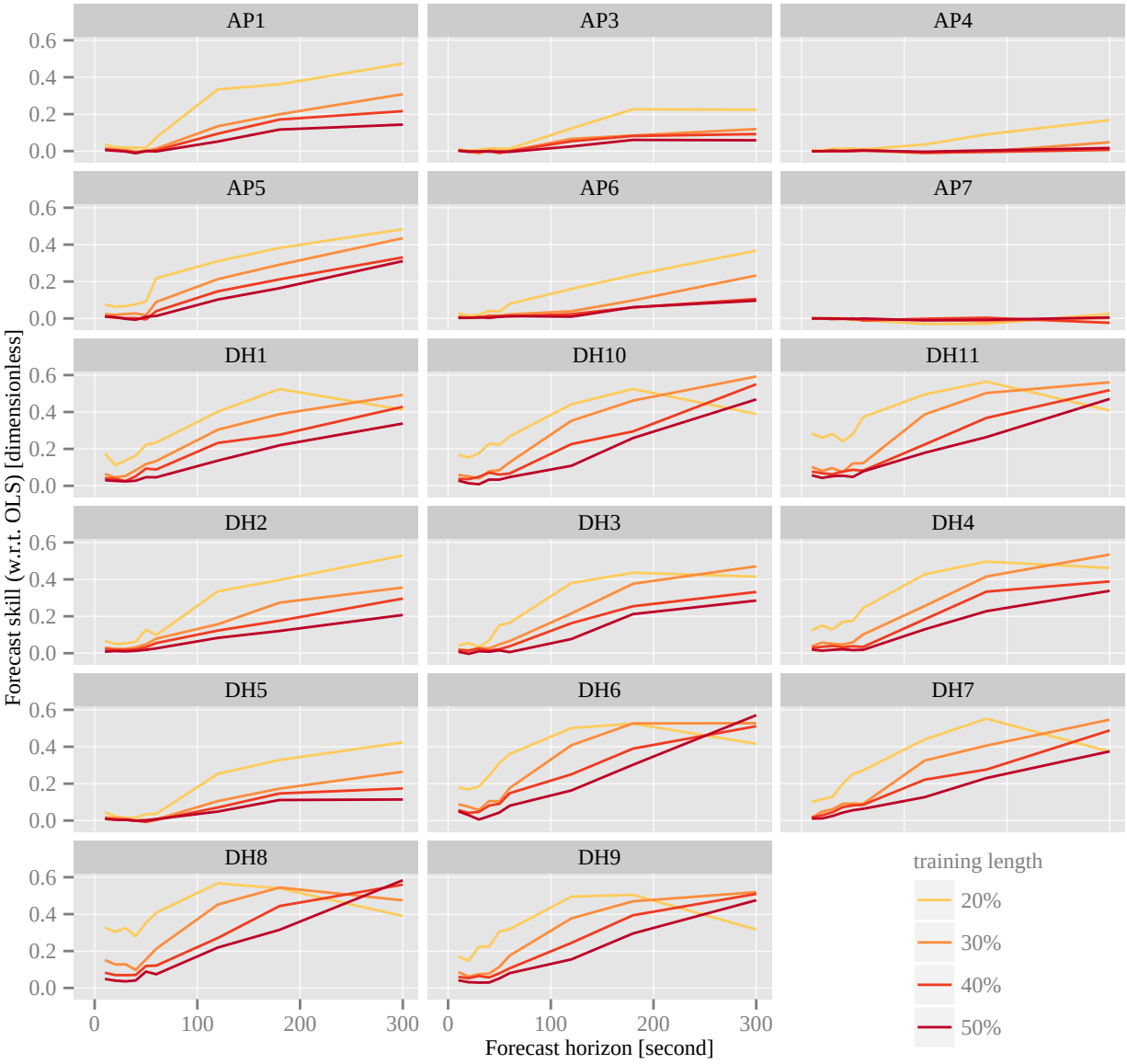

Figure A.8: Average forecast skill (with respect to the OLS results) of the lasso regression method during the 13 selected days at each station. Various training lengths are considered.

References

Arias-Castro, E., Kleissl, J., Lave, M., 2014. A poisson model for anisotropic solar ramp rate correlations. Solar Energy 101, 192 – 202. URL: http://www.sciencedirect.com/science/article/pii/S0038092X13005549, doi:http://dx.doi.org/10.1016/j.solener.2013.12.028.

Bosch, J., Kleissl, J., 2013. Cloud motion vectors from a network of ground sensors in a solar power plant. Solar Energy 95, 13 – 20. URL: http://www.sciencedirect.com/science/article/pii/S0038092X13002193, doi:http://dx.doi.org/10.1016/j.solener.2013.05.027.

Bosch, J., Zheng, Y., Kleissl, J., 2013. Deriving cloud velocity from an array of solar radiation measurements. Solar Energy 87, 196 – 203. URL: http://www.sciencedirect.com/science/article/pii/S0038092X12003854, doi:http://dx.doi.org/10.1016/j.solener.2012.10.020.

Box, G.E.P., Jenkins, G.M., Reinsel, G.C., 1994. Time Series Analysis: Forecasting and Control. Prentice Hall, Inc., Englewood Cliffs-New Jersey.

Chu, Y., Urquhart, B., Gohari, S.M., Pedro, H.T., Kleissl, J., Coimbra, C.F., 2015. Short-term reforecasting of power output from a 48 MWe solar PV plant. Solar Energy 112, 68 – 77. URL: http://www.sciencedirect.com/science/article/pii/S0038092X14005611, doi:http://dx.doi.org/10.1016/j.solener.2014.11.017.

Dong, Z., Yang, D., Reindl, T., Walsh, W.M., 2013. Short-term solar irradiance forecasting using exponential smoothing state space model. Energy 55, 1104 – 1113. URL: http://www.sciencedirect.com/science/article/pii/S0360544213003381, doi:http://dx.doi.org/10.1016/j.energy.2013.04.027.

Dong, Z., Yang, D., Reindl, T., Walsh, W.M., 2014. Satellite image analysis and a hybrid ESSS/ANN model to forecast solar irradiance in the tropics. Energy Conversion and Management 79, 66 – 73. URL: http://www.sciencedirect.com/science/article/pii/S0196890413007644, doi:http://dx.doi.org/10.1016/j.enconman.2013.11.043.

Efron, B., Hastie, T., Johnstone, I., Tibshirani, R., 2004. Least angle regression. The Annals of Statistics 32, 407–499. doi:10.1214/009053604000000067.

Hastie, T., Efron, B., 2013. lars: Least Angle Regression, Lasso and Forward Stagewise. URL: http://CRAN.R-project.org/package=lars. r package version 1.2.

Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning. Springer Series in Statistics, Springer New York. URL: http://link.springer.com.libproxy1.nus.edu.sg/chapter/10.1007/978-0-387-84858-7_3.

Hinkelman, L.M., 2013. Differences between along-wind and cross-wind solar irradiance variability on small spatial scales. Solar Energy 88, 192 – 203. URL: http://www.sciencedirect.com/science/article/pii/S0038092X12004021, doi:http://dx.doi.org/10.1016/j.solener.2012.11.011.

Hinkelman, L.M., 2014. Personal communication.

Hohm, D.P., Ropp, M., 2000. Comparative study of maximum power point tracking algorithms using an experimental, programmable, maximum power point tracking test bed, in: Photovoltaic Specialists Conference, 2000. Conference Record of the Twenty-Eighth IEEE, pp. 1699–1702. doi:10.1109/PVSC.2000.916230.

Hyndman, R.J., Koehler, A.B., Ord, J.K., Snyder, R.D., 2008. Forecasting with Exponential Smoothing. Springer, Deblik, Berlin, Germany.

Inman, R.H., Pedro, H.T., Coimbra, C.F., 2013. Solar forecasting methods for renewable energy integration. Progress in Energy and Combustion Science 39, 535 – 576. URL: http://www.sciencedirect.com/science/article/pii/S0360128513000294, doi:http://dx.doi.org/10.1016/j.pecs.2013.06.002.

Lipperheide, M., Bosch, J., Kleissl, J., 2015. Embedded nowcasting method using cloud speed persistence for a photovoltaic power plant. Solar Energy 112, 232 – 238. URL: http://www.sciencedirect.com/science/article/pii/S0038092X1400557X, doi:http://dx.doi.org/10.1016/j.solener.2014.11.013.

Lonij, V.P., Brooks, A.E., Cronin, A.D., Leuthold, M., Koch, K., 2013. Intra-hour forecasts of solar power production using measurements from a network of irradiance sensors. Solar Energy 97, 58 – 66. URL: http://www.sciencedirect.com/science/article/pii/S0038092X13003125, doi:http://dx.doi.org/10.1016/j.solener.2013.08.002.

Mahamadou, A.T., Mamadou, B.C., Brayima, D., Cristian, N., 2011. Ultracapacitors and batteries integration for power fluctuations mitigation in wind-PV-diesel hybrid system. International Journal of Renewable Energy Research 1, 86 – 95.

Marquez, R., Coimbra, C.F.M., 2012. Proposed metric for evaluation of solar forecasting models. Journal of Solar Energy Engineering 135, 011016–011016. URL: http://dx.doi.org/10.1115/1.4007496.

Miller, A., 2002. Subset Selection in Regression. C&H/CRC Monographs on Statistics & Applied Probability, Chapman and Hall/CRC. URL: http://dx.doi.org/10.1201/9781420035933.ch1.

Nguyen, D.A., Kleissl, J., 2014. Stereographic methods for cloud base height determination using two sky imagers. Solar Energy 107, 495 – 509. URL: http://www.sciencedirect.com/science/article/pii/S0038092X14002333, doi:http://dx.doi.org/10.1016/j.solener.2014.05.005.

Perez, R., Kivalov, S., Schlemmer, J., Jr., K.H., Hoff, T.E., 2012. Short-term irradiance variability: Preliminary estimation of station pair correlation as a function of distance. Solar Energy 86, 2170 – 2176. URL: http://www.sciencedirect.com/science/article/pii/S0038092X12000928, doi:http://dx.doi.org/10.1016/j.solener.2012.02.027. progress in Solar Energy 3.

Quesada-Ruiz, S., Chu, Y., Tovar-Pescador, J., Pedro, H., Coimbra, C., 2014. Cloud-tracking methodology for intra-hour {DNI} forecasting. Solar Energy 102, 267 – 275. URL: http://www.sciencedirect.com/science/article/pii/S0038092X14000486, doi:http://dx.doi.org/10.1016/j.solener.2014.01.030.

R Core Team, 2014. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. URL: http://www.R-project.org/.

Reda, I., Andreas, A., 2008. Solar Position Algorithm for Solar Radiation Applications. Technical Report TP-560-34302. National Renewable Energy Laboratory. Golden, CO.

Teleke, S., Baran, M., Bhattacharya, S., Huang, A., 2010. Rule-based control of battery energy storage for dispatching intermittent renewable sources. Sustainable Energy, IEEE Transactions on 1, 117–124. doi:10.1109/TSTE.2010.2061880.

Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B (Methodological) 58, 267–288. URL: http://www.jstor.org/stable/2346178.

Tibshirani, R., 2011. Regression shrinkage and selection via the lasso: a retrospective. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 73, 273–282. URL: http://dx.doi.org/10.1111/j.1467-9868.2011.00771.x, doi:10.1111/j.1467-9868.2011.00771.x.

Tibshirani, R., 2014. A simple explanation of the lasso and least angle regression. http://statweb.stanford.edu/~tibs/lasso.html. Accessed: 2014-12-07.

Tibshirani, R.J., Taylor, J., 2012. Degrees of freedom in lasso problems. Ann. Statist. 40, 1198–1232. URL: http://dx.doi.

org/10.1214/12-AOS1003, doi:10.1214/12-AOS1003.

Yang, D., Dong, Z., Nobre, A., Khoo, Y.S., Jirutitijaroen, P., Walsh, W.M., 2013a. Evaluation of transposition and decomposition models for converting global solar irradiance from tilted surface to horizontal in tropical regions. Solar Energy 97, 369 – 387. URL: http://www.sciencedirect.com/science/article/pii/S0038092X13003435, doi:http://dx.doi.org/10.1016/j.solener.2013.08.033.

Yang, D., Dong, Z., Reindl, T., Jirutitijaroen, P., Walsh, W.M., 2014a. Solar irradiance forecasting using spatio-temporal empirical kriging and vector autoregressive models with parameter shrinkage. Solar Energy 103, 550 – 562. URL: http://www.sciencedirect.com/science/article/pii/S0038092X14000425, doi:http://dx.doi.org/10.1016/j.solener.2014.01.024.

Yang, D., Gu, C., Dong, Z., Jirutitijaroen, P., Chen, N., Walsh, W.M., 2013b. Solar irradiance forecasting using spatial-temporal covariance structures and time-forward kriging. Renewable Energy 60, 235 – 245. URL: http://www.sciencedirect.com/science/article/pii/S0960148113002759, doi:http://dx.doi.org/10.1016/j.renene.2013.05.030.

Yang, D., Jirutitijaroen, P., Walsh, W.M., 2012. Hourly solar irradiance time series forecasting using cloud cover index. Solar Energy 86, 3531 – 3543. URL: http://www.sciencedirect.com/science/article/pii/S0038092X12003039, doi:http://dx.doi.org/10.1016/j.solener.2012.07.029.

Yang, D., Sharma, V., Ye, Z., Lim, L.I., Zhao, L., Aryaputera, A.W., 2015. Forecasting of global horizontal irradiance by exponential smoothing, using decompositions. Energy , to appear,doi:http://dx.doi.org/10.1016/j.energy.2014.11.082.

Yang, D., Ye, Z., Nobre, A.M., Du, H., Walsh, W.M., Lim, L.I., Reindl, T., 2014b. Bidirectional irradiance transposition based on the perez model. Solar Energy 110, 768 – 780. URL: http://www.sciencedirect.com/science/article/pii/S0038092X14004927, doi:http://dx.doi.org/10.1016/j.solener.2014.10.006.

Yang, H., Kurtz, B., Nguyen, D., Urquhart, B., Chow, C.W., Ghonima, M., Kleissl, J., 2014c. Solar irradiance forecasting using a ground-based sky imager developed at UC San Diego. Solar Energy 103, 502 – 524. URL: http://www.sciencedirect.com/science/article/pii/S0038092X14001327, doi:http://dx.doi.org/10.1016/j.solener.2014.02.044.

Zou, H., Hastie, T., Tibshirani, R., 2007. On the "degrees of freedom" of the lasso. Ann. Statist. 35, 2173–2192. URL: http://dx.doi.org/10.1214/009053607000000127, doi:10.1214/009053607000000127.