Applied Energy 212 (2018) 654–662

Contents lists available at ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apenergy

Iterative multi-task learning for time-series modeling of solar panel PV outputs

Tahasin Shireen[a], Chenhui Shao[b], Hui Wang[a,*], Jingjing Li[c], Xi Zhang[d], Mingyang Li[e]

[a] Department of Industrial and Manufacturing Engineering, Florida State University, 2525 Pottsdamer St., Tallahassee, FL 32310, USA
[b] Department of Mechanical Science and Engineering, University of Illinois at Urbana-Champaign, 1206 W. Green St., Urbana, IL 61801, USA
[c] Department of Industrial and Manufacturing Engineering, Pennsylvania State University, 310 Leonhard Building, University Park, PA 16802, USA
[d] Department of Industrial Engineering and Management, Peking University, 298 Chengfu Rd., Beijing 100871, China
[e] Department of Industrial and Management Systems Engineering, University of South Florida, Tampa, FL 33620, USA

HIGHLIGHTS

- An improved method for solar panel PV generation prediction is developed.
- Fusion of PV data from similar solar panels can significantly improve PV prediction.
- The multi-task learning algorithm is improved and generalized for time-series data.
- Systematic discussions and guidelines for implementation of the method are provided.

ARTICLE INFO

Keywords:
Multi-task learning
Time series
Solar panels
Prediction
Forecasting

ABSTRACT

Time-series modeling of PV output for solar panels can help solar panel owners understand the power systems' time-varying behavior and be prepared for the load demand. The time-series forecast/prediction can become challenging due to many missing observations or a lack of historical records that are not sufficient to establish statistical models. Increasing PV measurement frequency over a longer period increases the cost in the detection of the PV fluctuation. This paper proposes an efficient approach to iterative multi-task learning for time series (MTL-GP-TS) that improves prediction of the PV output without increasing measurement efforts by sharing the information among PV data from multiple similar solar panels. The proposed iterative MTL-GP-TS model learns/imputes unobserved or missing values in a dataset of time series associated with the solar panel of interest to predict the PV trend. Additionally, the method improves and generalizes the traditional multi-task learning for Gaussian Process to the learning of both global trend and local irregular components in time series. A real-world case study demonstrated that the proposed method could result in substantial improvement of predictions over conventional approaches. The paper also discusses the selection of parameters and data sources when implementing the proposed algorithm.

1. Introduction

Photovoltaic (PV) output is one of the most critical performance indicators for solar panels. The PV forecast/prediction can help solar panel owners be prepared for load demand and efficiently supply the solar energy as a complimentary source for power grids, without overestimating or underestimating the capabilities of solar panels. Additionally, the temporal prediction can help detect potential abnormal PV variations in a timely manner. Modeling of the PV time series usually employs statistical methods to progressively predict the PV output over time in near future based on the trend as reflected in historical data, which is collected at a specified time interval. The time-series model requires sufficient training data to establish adequate statistical confidence.

As reviewed in [1], there has been a noticeable amount of research on forecasting time-series data in solar panels and solar radiation including ARMA (Auto-regressive and Moving Average), ARIMA (autoregressive integrated moving average) [2]. Time-series modeling also considers the impact of environmental factors on the PV outputs based on ARMAX (autoregressive-moving-average model with exogenous inputs) [3,4]. In [5], artificial neural networks with multi-layer perceptron were developed to forecast the daily solar radiation in time-series

* Corresponding author.
E-mail address: hwang10@fsu.edu (H. Wang).

https://doi.org/10.1016/j.apenergy.2017.12.058
Received 30 July 2017; Received in revised form 5 December 2017; Accepted 9 December 2017
Available online 22 December 2017
0306-2619/

dataset by using transfer function of hidden layers. Almeida et al. [6] presented another forecasting approach with a nonparametric model of AC power forecast with quantile regression forests as machine learning tools. The research of the forecast with nonparametric approaches are found in [7,8] to forecast short-term PV outputs. A hybrid method using support vector machine with firefly algorithm was introduced by [9] to forecast the solar radiation. Time series modeling has also been applied to other green energy generation such as wind power forecasting [10–12].

One of the major **challenges** in the time-series modeling of PV output for a solar panel is missing data or data gaps due to a variety of reasons. Data collection may be interrupted due to technical problems such as data measurement equipment failure, erroneous recordings, weather factors. The problem of missing historical data is particularly significant for a newly deployed solar panel system. As a result, there exist substantial gaps in data collection for time-series modeling and the missing values in time series can cause a poor prediction of PV output. Increasing the measurement frequency of PV data or collecting more measurements over a longer period adds to the measurement cost. Therefore, it is essential to develop an efficient way to improve the performance prediction of the solar panel trend when there are many missing-value gaps in the training data for statistical modeling.

The effects of missing observations on the predicted result have been studied by various research. The missing observations in time series dataset can be dealt with by the following two approaches, i.e.,

1. Interpolation and extrapolation methods. Example approaches include linear interpolation, AR predictor [13], autoregressive conditionally heteroscedastic [14], Kalman filtering [15], neural networks [16] and multiple imputation models discussed in [17]. Techniques were also developed to reduce sampling rate or time resolution for PV time-series modeling [18]. This line of research can help smooth the prediction data, but nevertheless does not necessarily improve the prediction/forecast accuracy due to the limited useful information added to the time-series modeling.
2. Incorporation of environmental factors. The problem of predicting the time series response $PV(t)$ can be mitigated by fusing the effect of historical environmental data such as temperature, wind speed, and humidity in ARMAX, artificial intelligent (AI) model including ANN (artificial neural network) [1,19,20], wavelet-coupled support vector machine model [21], and analog ensemble method [22]. The historical PV data such as $PV(t-1), PV(t-2), PV(t-3)$...in these models can be de-seasoned by deducting the effect of solar irradiance $IR(t)$ via $PV(t-k)/IR(t-k), k = 1,2,3$.... It can be seen that the historical PV responses are one of the most significant variables that affect the forecasting/prediction performance based on the de-seasoned data. As such, improving the imputation of historical PV data is highly beneficial for time series modeling.

Time-series modeling based on these methods is considered as single task learning (STL) because the learning is accomplished by using the data from one single source, i.e., the solar panel of interest. Due to the limitation of information sources, such STL offers limited improvement in learning missing values to capture the true PV trend when there are many missing values.

A new opportunity emerges in improving the learning of the unobserved values by sharing the knowledge or information available from other data collected under similar conditions. For example, when a solar panel may not have sufficient data to estimate its performance such as a new PV system, the history of an old solar panel of the same type under similar deployment conditions can potentially provide cross inference about this panel. To enable such a knowledge transfer, a machine learning framework called multi-task learning (MTL) has been developed, which learns multiple tasks simultaneously with the key idea of sharing the information of each task [23]. The term “task” refers to the learning and prediction of certain target performance based on data. Joint learning for all tasks using MTL can significantly mitigate problems such as overfitting and unstable search due to sparse training data while improving performance compared to STL methods [24]. Relatedness among tasks is vital in MTL [25] as unrelated or dissimilar tasks can be detrimental to effective learning resulting in a negative transfer of information [26]. By utilizing the task similarity, the MTL algorithm has been applied to a Gaussian process (GP) for spatial data based on joint priors in [26–30]. In the past, the scope of the MTL has also been explored in [31,32] for classification problems, in [33] for learning using task-specific features and in [34] for learning with the informative vector machines by learning parameters based on the idea of task relatedness. The scope of MTL is yet to be explored in time series features for forecasting.

Based on the review, we identified two **research gaps**, i.e.,

1. The traditional STL has demonstrated its effectiveness in short-term PV prediction but has its limitations when dealing with a significant amount of missing values.
2. The traditional MTL-GP method has its limitations in time-series modeling for PV data. We conclude that a joint learning method using multiple similar-but-not-identical datasets has the potential in improving learning/imputation of missing values and prediction of solar panel performance. However, the prior research on MTL for GP mostly dealt with spatial data without considering both local temporal dependence and global trend. To the best of our knowledge, there exists little research that explores the values of the MTL-GP method and improves this method in time-series modeling for PV data.

To overcome the challenge, this paper proposes an approach to improving prediction of solar panel performance by sharing common information among similarly related time series datasets at multiple locations. An iterative multi-task learning Gaussian process time series (MTL-GP-TS) algorithm is developed to learn the missing observations in time series dataset without increasing measurement efforts. The term ”iterative” refers to the implementation of MTL procedures in an iterative manner to gradually update the trend and local component in the model. Specifically, an initial value will be proposed for the trend, and multi-task learning will be implemented to estimate the local components. The MTL-learned local components will be used to update the trend. The updating procedure continues until the estimation error converges. We demonstrate the proposed iterative MTL-GP-TS method using a real-world case study based on the data collected from four Hawaiian schools. The paper also discusses two practical issues when generalizing the applications of the proposed algorithm.

The paper consists of the following sections. Section 2 describes the iterative MTL-GP-TS algorithm to learn unobserved values from the time series dataset of multiple solar panels. Section 3 presents a real-world case study based on the PV data from four Hawaiian schools demonstrating the iterative MTL-GP-TS method. A discussion on implementing the algorithm is given in Section 4. Section 5 summarizes the paper.

2. Development of the iterative MTL-GP-TS algorithm

The MTL algorithm for GP is first reviewed in Section 2.1 following [27,28], and the iterative MTL-GP-TS algorithm is proposed in Section 2.2. It should be noted that we use the term ”estimation” in this paper as the process of finding an approximated value for the model parameters and output. We also use the term ”learning” as the process of gaining insights through making predictions or estimation on data. The two terms are used interchangeably to avoid word repetition in this paper. The term ”prediction” is used for statistical assessment of the values in final outputs that have not been measured or may happen in the future.

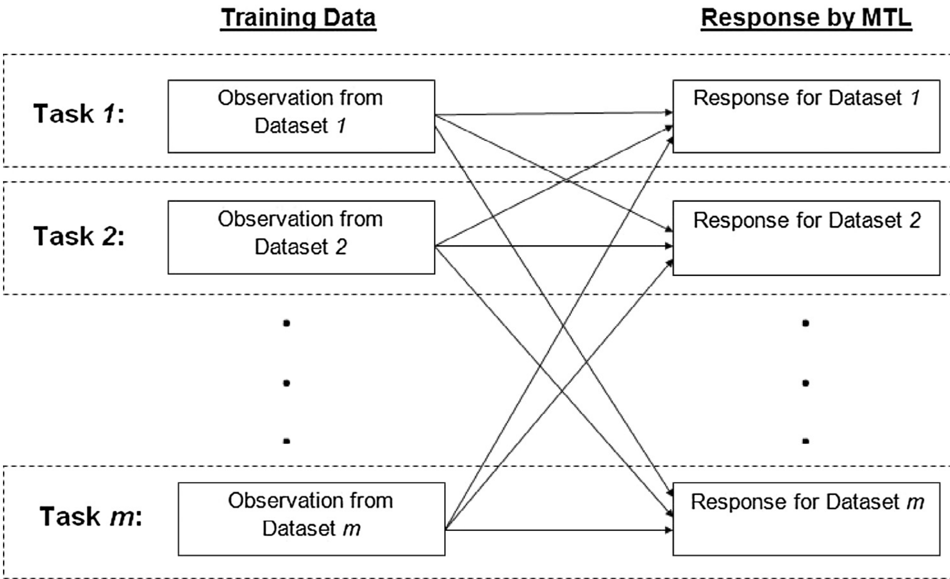

Fig. 1. The framework of multi-task learning.

2.1. A review of multi-task learning for Gaussian process

MTL formulation for Gaussian Processes. Assume that the data given as a set of responses $\epsilon^1(p_1)$, …, $\epsilon^1(p_N)$,$\epsilon^2(p_1)$, …, $\epsilon^2(p_N)$,…$\epsilon^M(p_1)$, …, $\epsilon^M(p_N)$, for learning M tasks at different inputs p_1,$p2$,...pN, where $\epsilon^l(p_i)$ is the response for the lth task given the ith input p_i. The problem of multi-task learning is to predict an observed response given an input p for a certain task based on the information from all tasks (Fig. 1). By [27,28], the response is assumed to be generated from a function of input p, which follows a multivariate Gaussian distribution with a covariance function $\kappa(p_i,p_j)$ between the responses at input p_i and p_j (defined as a Gaussian process). Yu et al. [27] used standard equations for prediction using the Gaussian process to estimate the prediction at any input p by $\hat{\epsilon}^l(p) = \sum_i \gamma_i^l \kappa(p,p_i)$, which estimates the response at p as a linear combination of kernel values $\kappa(p_i,p_j)$. The coefficients, γ_i^l, are assumed to follow a normal distribution $\gamma_i^l \sim N(\mu_\gamma,C_\gamma)$. The distribution parameters are assumed to jointly follow a normal-inverse-Wishart distribution as $(\mu_\gamma,C_\gamma) \sim \mathcal{N}\left(\mu_\gamma \,\middle|\, 0,\frac{1}{\pi}C_\gamma\right)\mathcal{IW}(C_\gamma\,|\tau,\kappa^{-1})$, and π and τ are parameters in the distributions. Given the hyper prior distribution of μ_γ,C_γ, Yu et al. [27] further proposed a data prediction model under the MTL framework as follows

1. Parameters μ_γ and C_γ are generated (or statistically sampled) for once based on the normal-inverse-Wishart distribution
2. At a certain input p,

$$\epsilon^l(p) = \sum_i \gamma_i^l \kappa(p,p_i) + e, \tag{2.1}$$

where $e \sim N(0,\sigma^2)$.

The estimates of parameters μ_γ,C_γ,σ and γ_i^l can be learned by an Expectation-Maximization (EM) algorithm. The details about the EM algorithm is reviewed in Appendix B.

MTL for solar panel PV forecasting. When the MTL model is applied to the PV data forecasting for solar panels, $\epsilon^l(p_i)$ can represent PV-related output variable for the solar panel at the location of interest. Index p becomes a time index, and we replace it with a notation t for MTL of time-series data in the remainder of this paper. The PV-related output variable can be chosen as PV, PV regularized by solar irradiance, or a local component in the PV data (see details in Section 2.2). As such, the MTL problem aims to estimate the PV-related output of the solar panel at the location of interest (target) by aggregating the similarity between the solar panel's PV output at a time point i (including those solar panels at different locations) and the target. The similarity between different solar panels are captured by the kernel function κ and its strength is reflected by coefficient γ_i^l.

The limitation of the traditional MTL-GP method is that the approach lacks the capability of capturing the temporal trend in data, which are critical for temporal data modeling. Specifically, the traditional MTL-GP method only predicts the spatial data in a local model (by which data can only be estimated from those surrounding it, such as a Gaussian Process model) by borrowing the information from other spatial datasets. For the challenge in this paper, MTL needs to estimate the parameters for both local model and a trend, which is separated from the local model. However, the traditional MTL-GP cannot distinguish the trend from the local model. Additionally, the requirement of Gaussian-distribution assumption in the MTL-GP method is not mostly appropriate for PV data with significant trending patterns. This paper will enhance the MTL-GP method to characterize both global trend and local variations in time series data for solar panels.

2.2. Improved method: iterative multi-task learning Gaussian process for time series (MTL-GP-TS)

The time series data can be decomposed into a trend T (including seasonal component) and an irregular/random component ϵ, i.e., $Z(t) = T(t) + \epsilon(t)$. The irregular or random component can usually be assumed to follow a Gaussian process plus noise. The reasons for choosing the Gaussian process are that (1) The local irregular component in time series usually exhibit strong temporal correlation or local dependency. The Gaussian process is a local modeling method, which predicts the outcome based on the value surrounding it by using the local dependency as learned from data; and (2) When the trend can be correctly captured by the model, the left-over residual should be a correlated stochastic process. The multivariate normal (Gaussian) distribution is a very common model to characterize the statistical attributes of the residual stochastic process. As such, the MTL-GP method is suitable to transfer the information among irregular components from similar time series datasets to the dataset of interest (target). However, learning accuracy of the irregular components using MTL-GP is greatly influenced by the way how the trend and irregular components are decomposed. Thus, joint consideration of the trend estimation and learning of the irregular components poses a new challenge.

To fill the gap, this paper develops an iterative MTL-GP-TS algorithm that simultaneously estimates the trend and irregular component through information transfer among datasets. The estimation procedure

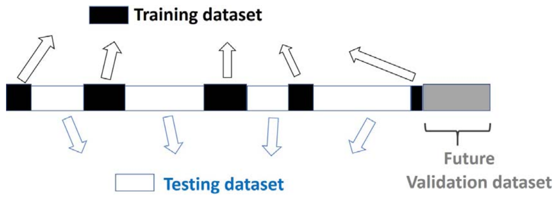

Fig. 2. An illustration of training, testing, and validation datasets. Data are randomly selected and removed from the original data to simulate the "missing values." These data create a testing dataset (white bar). The proposed iterative MTL-GP-TS algorithm is applied to the training datasets (black bar) and thereby estimate the missing values corresponding to the time period of the testing dataset. The validation dataset (grey bar) is obtained from a future period after training and testing datasets. The training and estimated testing data are jointly used to fit an ARIMA time series model to estimate the future values corresponding to the validation dataset. The estimation for the future period will be compared with the values in the validation dataset to evaluate forecasting performance.

is described as follows.

Iterative MTL-GP-TS algorithm:

1. Start with initial time series data $Z_0^l(t)$ of size $(n\times 1)$, where $l = 1,2,3\ldots$indicates data source or task, i.e., different solar panels;
2. Clean the raw data to generate sorted data $Z_s^l(t)$ by removing outliers and replacing them using the estimates based on a K-Nearest Neighbors (KNN) method. Details about the KNN method are reviewed in Appendix C;
3. Randomly select data points from $Z_s^l(t)$ during a time period to create the training dataset and the rest data during this period form the testing dataset to simulate "missing observations." The validation dataset is formed using those observations after the period to validate the forecasted values for future data. The separation of the datasets are summarized in Fig. 2;
4. Provide an initial estimate of testing datasets (that simulate "missing observations"), $Z^l(t)$, by a linear interpolation based on the training dataset;
5. Set $i = 0$ for the iteration index and $T_0^l(t) = 0\ \forall\ t$ for initialization;
6. Estimate irregular component, $\epsilon_i^l(t)$, which is assumed to follow Gaussian process, by a decomposition, i.e., $\epsilon_i^l(t) = Z^l(t) - T_i^l(t)$;
7. Estimate the irregular component using MTL-GP algorithm given $\epsilon_i^l(t), l = 1,2,3\ldots$from all data sources, denoted as $\epsilon_{iM}^l(t)$. The MTL procedure includes (1) setting the initial values for μ_γ, C_γ, and γ, which will be used for the first iteration of EM and (2) using the EM algorithm (Appendix A) to estimate μ_γ, C_γ, and γ.
8. Estimate the output value by $\widehat{Z}_i^l(t) = T_i^l(t) + \epsilon_{iM}^l$;
9. When iteration $i > 1$, check if $|\triangle RMSE_i|\% = |\frac{RMSE_i - RMSE_{i-1}}{RMSE_{i-1}}|\% < \delta$, where $RMSE_i$ is the root-mean-square error calculated by comparing the predicted output $\widehat{Z}_i^l(t)$ with true value $Z_0^l(t)$ in the raw training dataset at the ith iteration and δ assumes a very small value. If yes, the result converges, the final trend $T_f^l(t)$ is obtained, and we can exit the iteration. If no or when iteration index $i \le 1$, update the trend by subtracting $\epsilon_{iM}^l(t)$ from $Z^l(t)$ and then apply interpolation to learn $T_{i+1}^l(t)$;
10. Go to next iteration by $i \leftarrow i + 1$ and start with step 4;
11. Conduct parameter selection based on the steps above to identify the appropriate hyperparameters in the model (Discussed in Section 4.1);
12. Conduct data source selection based on the steps above to find the datasets to be included for the learning (Discussed in Section 4.2).

The flowchart of the procedure is summarized in Fig. 3. A time series model such as ARIMA can be fitted to the final estimated trend $T_f^l(t)$ via the iterative MTL-GP-TS to predict future performance. It should be noted that smooth moving average (SMA) method can be applied for the trend estimation to mitigate the effect of noise induced by measurement and/or weather factors.

The rationale of the proposed algorithm is as follows. The iterative method is used to deal with a challenge on simultaneous estimation of both trend and local irregular components in the time series data. In this problem formulation, the local irregular components can be learned through multi-task learning (MTL) based on the similarities with the solar panels at other locations. The iterative approach is a logical way to gradually learn about the two terms by first tentatively estimating the local component through MTL based on initial values and then updating the trend and local component until convergence after several iterations.

Remark: The determination of the necessary data points or sample size sufficient for the multi-task learning should be judged by how much the algorithm can improve the prediction accuracy (RMSE). As the sample size increases, the prediction performance of multi-task learning and STL both improve and RMSE improvement by the iterative multi-task learning compared with the traditional STL is expected to increase first and then approximate asymptotically to a stable constant when the sample size grows large. Therefore, the amount of RMSE improvement becomes bounded and the benefits brought by multi-task learning thus become limited when the sample size is relatively large.

3. A case study on four Hawaiian schools

The iterative MTL-GP-TS method is demonstrated using a case study based on solar panel datasets consisting of photovoltaic output (PV, unit: kilowatts) and plane of array (POA, unit: kilowatts per square meter) irradiance from four schools in Hawaii.[1] The data were collected using "Schott Solar SunTrack Local" data acquisition system. The data from four school locations are selected as the input to the iterative MTL-GP-TS algorithm based on the fact that these solar panels are supplied by the same manufacturer (i.e., SunPower SPR-200-BLK PV modules) and expected to exhibit similar variation patterns that enable the knowledge transfer across solar panels.

Remark: The distance between school locations does not play a dominant role that impacts the prediction results. The key to the prediction is if the data at different locations may exhibit similar variation patterns. In Section 4.1, this paper provides a data source selection method that allows for judgment on whether or not a location can be included in the prediction.

The data sources are from Nanakuli High and Intermediate school (N), Jarrett Middle school (J), Highlands Intermediate school (H), and Waianae High school (W). All data sources show a strong positive correlation between the PV and POA value. Such variation patterns induced by POA changes could potentially mask the abnormal patterns that are associated with the intrinsic problems in solar panels, which users expect to detect. The removal of the POA induced effect in data can better expose the variation patterns due to potential problems in the solar panels. Thus, the ratio of PV over POA (i.e., PV/POA) is used as time series observations to suppress the effect of POA. After removing the POA induced variation, the data still exhibit cyclic patterns for time series modeling. The solar panel PV data for "H school" is to be forecast, and the rest of schools (N, J, W) supply complementary data for multi-task learning.

The data are recorded daily and collected from Jan'2008 to April'2012 for the four schools (N, J, H, and W). The scope of the study is to predict the PV performance of solar panels over days and months. Although this study deals with a relatively long-term forecasting, the methodology is applicable for short-term PV forecasting as well. The data from Jan'2008 to April'2012 contained some inaccurate measurements and missing observations for some months. Outliers induced by inaccurate measurements are removed using Tukey's approach while specifying reasonable cleaning parameter that ensures no data point is

[1] The data source is available at http://www.hawaiianelectric.com/heco/_hidden_Hidden/EducationAndConsumer/Sun-Power-for-Schools?cpsextcurrchannel=1.

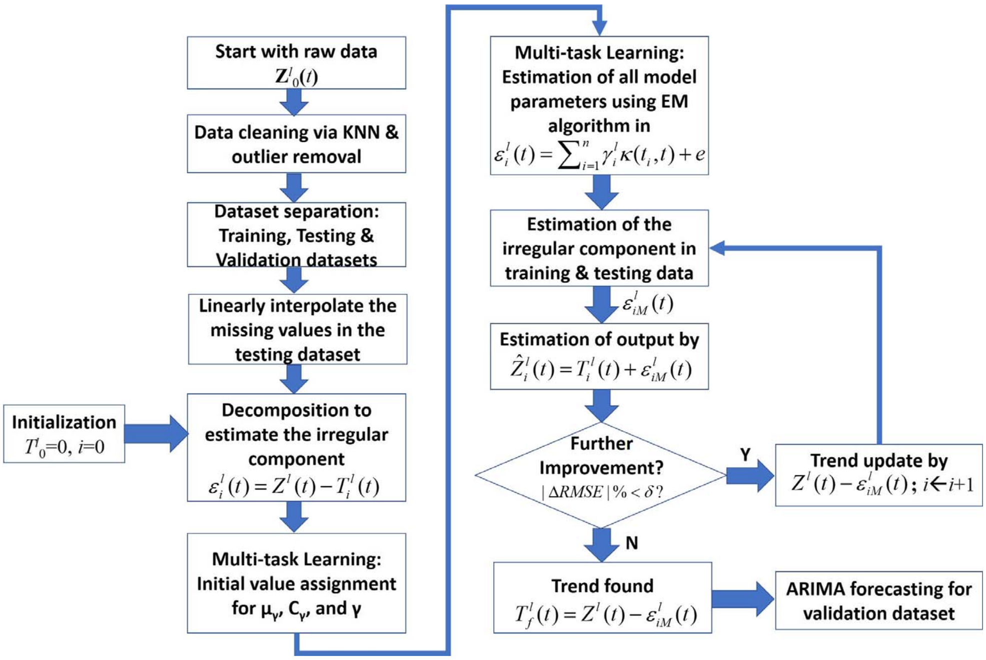

Fig. 3. Flowchart of the iterative MTL-GP-TS algorithm.

too large or too small compared to the rest of data points. Afterwards, a KNN method is used to replace values of those outliers in raw datasets. In this study, a 2-NN approach is used to generate missing observations in datasets by searching for values in two neighboring datasets.

The performance of the model is evaluated by testing the difference between predictions and true values in the testing and validation datasets. Specifically, the data are separated into training, testing, and validation data sets. The training dataset is used to estimate the ARIMA model based on the iterative MTL-GP-TS algorithm. The testing dataset includes some observations within the timespan of training dataset but excluded from the training set in order to simulate missing values. Its purpose is to test if the proposed algorithm can correctly impute the data within the timeframe of the training dataset. The validation dataset refers to those observations that happen after the timeframe of the

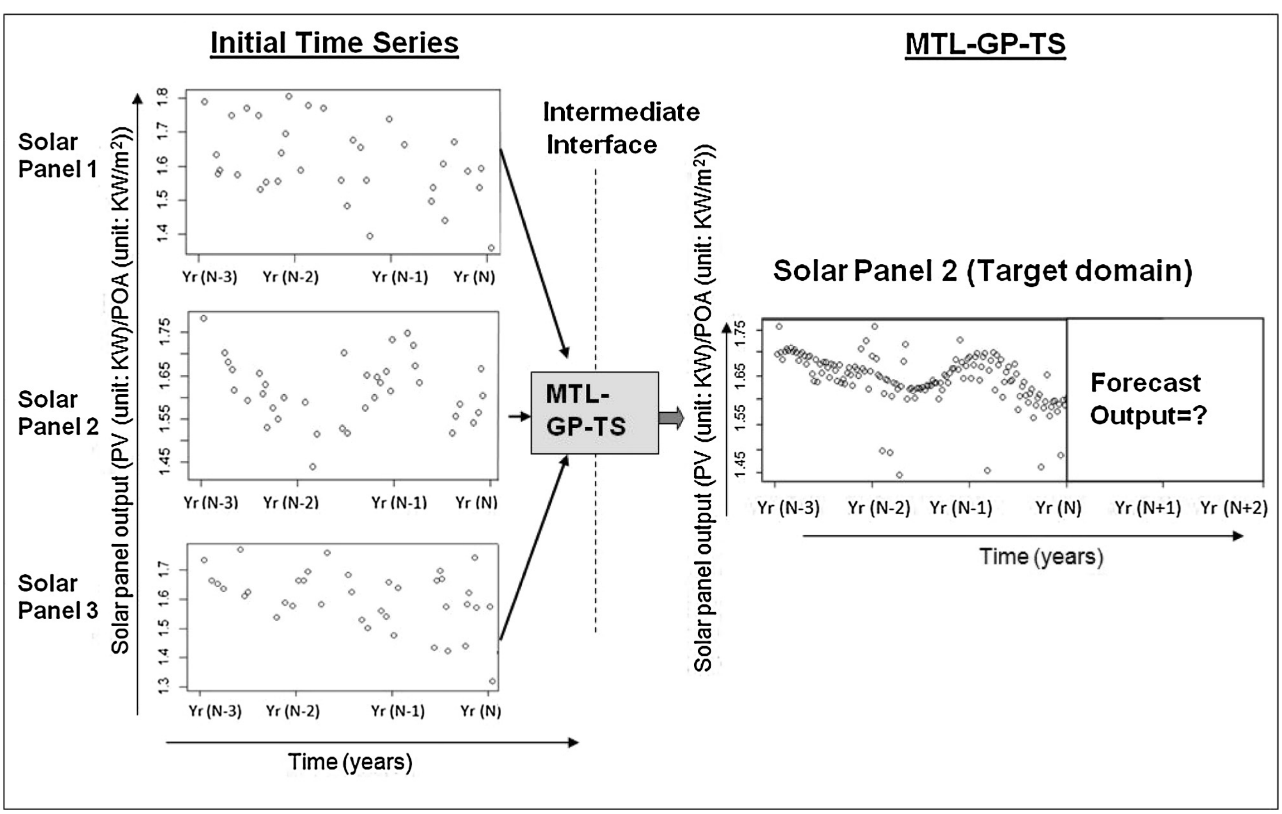

Fig. 4. A schematic example for iterative MTL-GP-TS learning. Left side: Initial time series (data point labeled as a circle) of solar panel data source-1, data source-2 and data source-3 starting during Years 1 to 3. Right side: MTL-GP-TS predictions for H school solar panel data source-2 from Years 1 to 3. Jointly learned MTL-GP-TS time series observations for solar panel data source-2 from Years 1 to 3 can be used to further predict the PV outputs in Years 4 and 5. Note: data are not fully displayed due to space limitation.

training data. The purpose of the validation set is to test if the algorithm can forecast the future observations after the training data. In this study, the dataset has a weekly time interval, where 44 data points are training data, and 144 data points are testing (missing) data points. The test data points are treated as missing observations in time series. It can be seen that 75% of values are considered as missing data points and excluded from the training data. The validation dataset contains 17 data points.

In the iterative MTL-GP-TS algorithm, the missing observations were learned by sharing the information among multiple time series if the datasets are from similar sources. For example, as shown in Fig. 4, three solar panel time series data can be learned jointly using the iterative MTL-GP-TS. The forecasting is achieved based on ARIMA, the most general class of models for time series prediction, which also excels in capturing the periodical cycles in the PV data [1]. It also deals with non-stationarity in the data by taking initial differencing steps. More detailed explanation of ARIMA is provided in Appendix B.

Fig. 5 shows the predicted trend along with confidence interval by using MTL-GP-TS after two iterations and STL ARIMA model based on the dataset with missing data, respectively. The results are compared with the trend values fitted to full data (as called *true trend*). It can be seen that the trend predicted by MTL-GP-TS is significantly closer to the true trend and the confidence interval is significantly narrower. The confidence interval for STL (bounded by green curves in Fig. 5) is wider than that of iterative MTL-GP-TS (bounded by red curves) due to a large standard deviation in interpolated sample dataset (missing values learned by interpolation). Also, it is observed that both iterative MTL-GP-TS and STL trends are mostly above the true trend during these months. Such a pattern was caused by the training dataset chosen lacks data points that can capture a significant trend drop for both iterative MTL-GP-TS and STL.

The RMSE of the predicted output (denoted as RMSE-predict) values between the STL and iterative MTL-GP-TS methods for H school are presented in Table 1. It is noticed that the RMSE-predict are improved by 25% after two iterations by the proposed method for solar panels in H school compared with STL, and the iterative MTL-GP-TS demonstrates substantial improvement in reducing RMSE of predicted values for performance prediction of trend estimation. The improvement in RMSE-predict indicates that the iterative MTL-GP-TS can outperform the traditional STL time-series modeling in predicting trend when time series dataset contains a significant amount of missing observations. Table 1 also shows the values for MAE (Mean Absolute Error) and MAPE (Mean Absolute Percentage Error). All these metrics demonstrate a significant improvement compared with STL. Fig. 6 shows a residual plot for the prediction based on the iterative MTL-GP-TS algorithm, and the residual mostly exhibits a relatively random pattern with a constant mean and stable variation. Still, some patterns can be observed, and it could be further mitigated by adopting a GARCH (generalized autoregressive conditional heteroskedasticity) model, which refines the error terms and can potentially make the residual plot more random. Table 2 further provides a common approach to the cross-validation for time series data based a walk forward method, by which the obtained model is used to forecast the observations at k steps ahead, where $k = 1,2,3....$ It can be seen that the prediction RMSE based on the imputation using the iterative MTL-GP-TS algorithm is similar to that in Table 1 and is very stable for multi-step forecasting.

Table 1
Prediction performance measurement metrics using the iterative MTL-GP-TS and STL based on H school data.

Metrics	MTL-GP-TS	STL	% Improvement
RMSE	0.38	0.5	24
MAE	0.31	0.38	18.42
MAPE	40.47	53.97	25

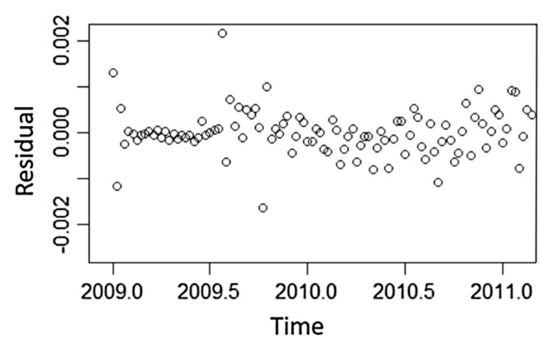

Fig. 6. Residuals of fitted ARIMA model with MTL-GP-TS learned trend.

4. Discussion

The applications of the proposed MTL-GP-TS algorithm in this paper can be generalized for various PV data. This section discusses two practical issues when implementing the algorithm for generic PV data including hyper-parameter and data source selections.

4.1. Hyper-parameter selection for the iterative MTL-GP-TS algorithm

The improvement by the iterative MTL-GP-TS method also depends on the appropriate selection of hyper-parameters in the iterative MTL-GP-TS algorithm. In this section, we show the effect of hyper-parameter

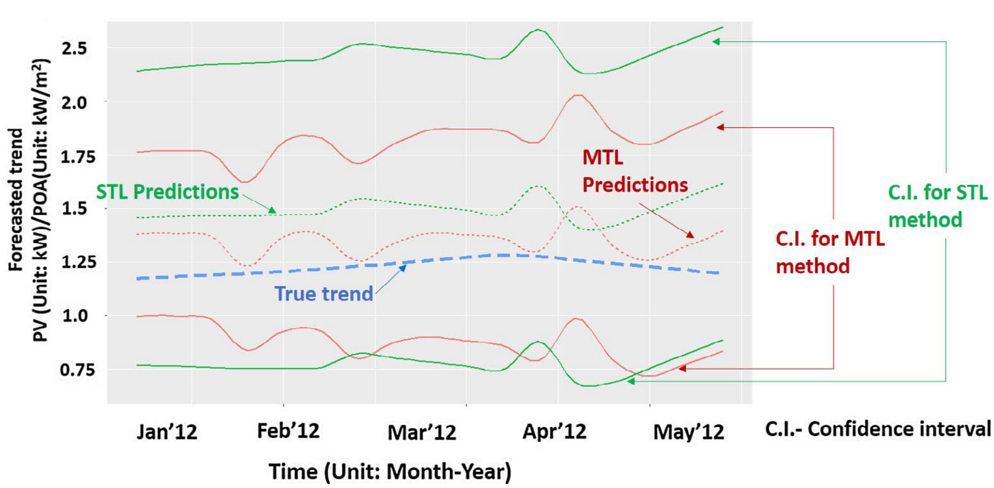

Fig. 5. Residuals of fitted ARIMA model with MTL-GP-TS learned trend. The MTL-predicted trend is closer to the true trend and MTL generates a narrower confidence interval (bounded by red curves) compared with that by STL method (bounded by green curves).

Table 2
Cross validation (Walk forward Validation) of the fitted ARIMA model for the MTL-GP-TS trend.

Step	RMSE
1-step	0.2301343
2-step	0.2338434
4-step	0.2369701
5-step	0.236674
10-step	0.2233119

selection based on a case study, in which the real data from the four Hawaiian schools along with a reference dataset from solar panel manufacturer, Sunpower (available in[2,3,4]), were employed to generate/simulate long-term data. Specifically, six years of simulated PV data were generated beyond 2012 over a ten-year timespan in total by applying time series models fitted to the real data from the four Hawaiian schools. The expanded dataset is separated into training, testing, and validation sets. The generated dataset has in total 480 time points over ten years with a weekly time interval. The training and testing sets contain 77 data points and 307 data points, respectively. Thus, about 80% data (testing data) are considered as missing and excluded from training dataset. The validation dataset includes 96 data points with a weekly interval in Years 9 and 10 for each school.

For hyperparameter selection, the RMSE values are computed for multiple candidate combinations of hyperparameter values for validation dataset and compared to select the combination that achieves the minimum RMSE. Each hyper parameter is assumed to take a number of discrete values within a certain range. The hyper-parameter values for τ,π and σ^2 were explored within [0.0001, 0.1, 1, 10, 20], [0.0001, 0.1, 1, 10, 20] and [1, 250, 500, 750, 1000], respectively. All 125 exhaustive combinations were tested to identify the appropriate parameter settings by comparing RMSE values. Due to the page limitation, Fig. 7 partially shows the response surface subplots of hyper parameters vs. RMSE, where one hyper parameter in each subplot was always kept constant at a certain value while searching for the lowest RMSE within the range of two other hyper parameters. It has been observed that the smaller τ,π and σ^2 values could yield prediction results with lower RMSE's. For the solar panel of interest (H school), the prediction results with the lowest RMSE were found when $\tau = 20$, $0.0001 < \pi < 0.1$ and $\sigma^2 = 1$.

After hyper-parameter selection, the RMSE-predict value for H school is presented in Table 3, which compares the results between the STL and iterative MTL-GP-TS methods. By using the selected hyper parameters, the RMSE-predict improvement by the iterative MTL-GP-TS is 66.67% for solar panels in H school compared with using traditional STL ARIMA method.

4.2. Selection of data sources in the iterative MTL-GP-TS algorithm

In the MTL, it is not always true that the inclusion of more data sources can improve the prediction [35,36]. Adding more data sources may generate very slight improvement or even negative impact on the prediction RMSE. For instance, the inclusion of all data sources into the iterative MTL-GP-TS algorithm to predict unobserved values in the solar panel of interest (H school) might result in a larger or similar RMSE-predict compared with that estimated by including only fewer data sources. Therefore, data source selection could be necessary to identify useful data sources prior to multi-task learning.

A best subset selection method can be performed to identify the potential data sources to be included in the iterative MTL-GP-TS algorithm. Specifically, the best subset selection method tests different subsets of the available data sources (N, J, W, S, and Sunpower's reference data) as inputs to perform the iterative MTL-GP-TS, and the subset that yields the lowest RMSE for the testing data will be identified as the most appropriate data sources to be included.

Based on the same dataset in Section 4.1, the results learned by the iterative MTL-GP-TS algorithm for H school using the solar panel data from different subsets of data sources were compared. Table 4 shows the RMSE-predict values of the predicted PV data in H school in Years 9

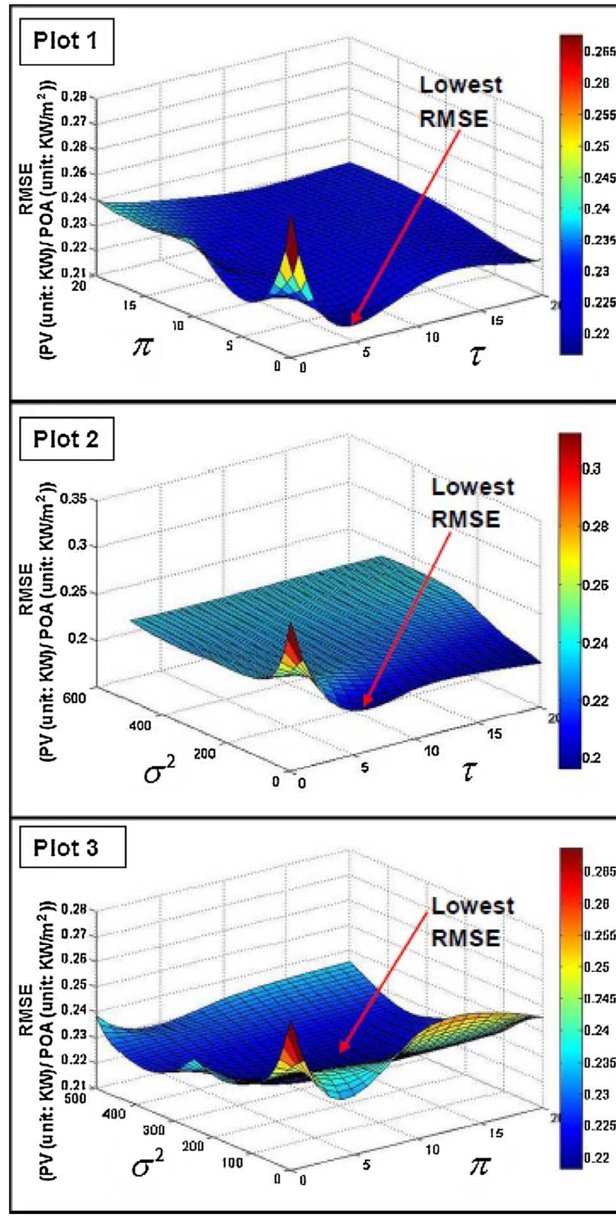

Fig. 7. Response surface plots of hyper parameters with RMSE. Starting from top, plot 1 shows RMSE results vs. hyper parameters τ and π when σ^2 is fixed. Plot 2 shows RMSE results vs. hyper parameters τ and σ^2 when π is fixed. Plot 3 shows RMSE results vs. hyper parameters π and σ^2 when τ is fixed.

Table 3
RMSE-predict using the iterative MTL-GP-TS vs. STL for H school after hyper-parameter selection.

RMSE-MTL-GP-TS	RMSE-STL	%RMSE improvement
0.15	0.45	66.67

[2] http://us.sunpower.com/sites/sunpower/files/media-library/white-papers/wp-sunpower-module-degradation-rate.pdf.
[3] http://energyinformative.org/lifespan-solar-panels/.
[4] http://us.sunpower.com/utility-scale-solar-power-plants/oasis-power-plant/.

Table 4
RMSE-predict of H school with different number of data sources.

No. of data sources in MTL-GP-TS	RMSE-predict of H school
5 (N, J, H, W, Sunpower)	**0.15**
4 (N, J, H, W)	**0.14**
3 (N, J, H)	**0.14**

and 10 when using five, four and three data sources. By combining the results in Tables 3 and 4, it can be seen that the inclusion of three data sources (N, J, and H schools) from four schools can yield a substantial improvement in PV value prediction. However, the inclusion of all schools and manufacturer (Sunpower) reference data into the learning does not improve the prediction. In fact, the RMSE-predict even slightly increases. Thus, N, J, and H schools are sufficient for the iterative MTL-GP-TS algorithm.

5. Conclusion

This paper identifies and explores the value of diversely recorded PV data for solar panels of similar types over a living or working area. Specifically, a new data fusion method based on multi-task learning is proposed for improving time series data imputation and prediction of PV output of solar panels without increasing measurement efforts. The method overcomes the challenge in time series data modeling when a significant amount of historical data are missing or unavailable by sharing information from other similar-but-not-identical time series. In addition, the proposed MTL-GP-TS algorithm generalizes the existing MTL method for Gaussian spatial processes to the learning of both global trend and local irregular components in time series data in an iterative manner. The algorithm was applied to predict PV data over time for solar panels deployed in four Hawaiian schools. The real-world case study demonstrated that the iterative MTL-GP-TS method could efficiently learn the unobserved values in time series dataset, resulting in improved predictions of PV time series for solar panels. Parameter selection was conducted to identify appropriate hyper parameters. It has also been found that the performance of the proposed algorithm depends on the selection of solar panel data sources and therefore data source selection is necessary. The outcome of this research can help on-time preparations for load demand and assessment of solar panel conditions. The method can also be applicable to health condition prediction and missing data imputation for a wide range of applications such as construction structures and mechanical/electronic devices. The future study involves (1) consideration of weather condition into the proposed algorithm and (2) time series model refinement to better incorporate the data volatility by considering the variation in the error term such a Garch model.

Acknowledgment

This research was conducted at the High-Performance Materials Institute at Florida State University. It has been partially supported by a grant from National Science Foundation (USA) CMMI-1434411 and 1744131. The authors would like to thank Sun Power for Schools Program hosted by Hawaiian Electric Companies, State of Hawaii Department of Education and Members of the community for the solar panel installation and data collection. Assistance given by Mr. Steve Luckett from Hawaiian Electric Co Inc. was also greatly appreciated.

Appendix A. EM Algorithm

This section reviews the Expectation-Maximization (EM) algorithm used to estimate the parameters in the model $\boldsymbol{\Theta} = \{\boldsymbol{\mu}_\gamma, \mathbf{C}_\gamma, \sigma^2\}$. The EM algorithm involves an iterative procedure that alternates between an expectation step, and a maximization step. Details about the EM algorithm has been documented in [27].

- *Expectation, E-step*: This step estimates the expectation and covariance of $\boldsymbol{\gamma}, l = 1,\ldots,m$, given the initial parameters of $\boldsymbol{\Theta}$.

$$\hat{\boldsymbol{\gamma}}^l = \left(\frac{1}{\sigma^2}\boldsymbol{\kappa}_l^{\mathrm{T}}\boldsymbol{\kappa}_l + \mathbf{C}_\gamma^{-1}\right)^{-1}\left(\frac{1}{\sigma^2}\boldsymbol{\kappa}_l^{\mathrm{T}} \in_l + \mathbf{C}_\gamma^{-1}\boldsymbol{\mu}_\gamma\right), \tag{A.1}$$

$$\mathbf{C}_{\gamma^l} = \left(\frac{1}{\sigma^2}\boldsymbol{\kappa}_l^{\mathrm{T}}\boldsymbol{\kappa}_l + \mathbf{C}_\gamma^{-1}\right)^{-1}, \tag{A.2}$$

where $\boldsymbol{\kappa}_l \in \mathbb{R}^{n_l \times n}$ is the kernel function $\kappa(\cdot,\cdot)$ that estimates the similarity between $\mathbf{p}_l$ and $\mathbf{p}$, where $\mathbf{p}_l$ is the vector of data points for task l.
- *Maximization, M-step*: This step optimizes the parameters $\boldsymbol{\Theta} = \{\boldsymbol{\mu}_\gamma, \mathbf{C}_\gamma, \sigma^2\}$ based on the E step. The parameter estimation for $\boldsymbol{\Theta}$ is given by [27] as follows,

$$\boldsymbol{\mu}_\gamma = \frac{1}{\pi + m}\sum_{l=1}^{m} \hat{\boldsymbol{\gamma}}^l, \tag{A.3}$$

$$\mathbf{C}_\gamma = \frac{1}{\tau + m} \times \left\{\pi\boldsymbol{\mu}_\gamma\boldsymbol{\mu}_\gamma^{\mathrm{T}} + \tau\boldsymbol{\kappa}^{-1} + \sum_{l=1}^{m} \mathbf{C}_{\gamma^l} + \sum_{l=1}^{m} [\hat{\boldsymbol{\gamma}}^l - \boldsymbol{\mu}_\gamma][\hat{\boldsymbol{\gamma}}^l - \boldsymbol{\mu}_\gamma]^{\mathrm{T}}\right\}, \tag{A.4}$$

$$\sigma^2 = \frac{1}{\sum_{l=1}^{m} n_l}\sum_{l=1}^{m} \|\in_l - \boldsymbol{\kappa}\hat{\boldsymbol{\gamma}}_l\|^2 + \mathrm{tr}[\boldsymbol{\kappa}_l\mathbf{C}_{\gamma^l}\boldsymbol{\kappa}_l^{\mathrm{T}}], \tag{A.5}$$

where $\mathrm{tr}(\cdot)$ is the trace operator.

Appendix B. ARIMA model

This section briefly reviews the ARIMA (auto-regressive integrated moving average) model. An ARIMA model takes initial differencing steps on data to eliminate the non-stationarity and generalize the time series model for non-stationary data. The math representation for an ARIMA (p,d,q) is given by [1]

$$\left(1-\sum_{i=1}^{p}\phi_i L^i\right)(1-L)^d\epsilon(t)-\left(1+\sum_{i=1}^{q}\theta_i L^i\right)e(t)=0 \tag{A.6}$$

where the notations $\epsilon(t)$ and $e(t)$ are the same as in Eq. (2.1); L is a time lag operator, the parameters p and q represent the time lags involved in the autoregressive and moving average terms, respectively; and the parameter d reflects the order of differencing performed on the data. These parameters are usually designated before fitting the ARIMA model parameters ϕ_i and θ_i to data.

Appendix C. KNN method

This section briefly reviews the K-nearest Neighbors (KNN) method used in the paper to replace the values of outliers in the training and testing datasets. The KNN method is a non-parametric approach to approximate values locally based on the surrounding data [37]. The approximation can be made by assigning weight w_i to the contributions of the neighbors t_{N_i} so that nearest neighbors contribute more to the average than the more distant data points as follows,

$$Z(t)=\sum_i w_i Z_(t_{N_i}), \tag{A.7}$$

where weight w_i is usually set to be inversely proportional to the distance d between t and t_{N_i}, i.e., $w_i \propto 1/d_i(t,t_{N_i})$ and $\sum_i w_i = 1$. The approximation adopted in this paper is essentially a generalization of linear interpolation.

References

[1] Wan C, Zhao J, Song Y, Xu Z, Lin J, Hu Z. Photovoltaic and solar power forecasting for smart grid energy management. CSEE J Power Energy Syst 2015;1(4):38–46.
[2] Reikard Gordon. Predicting solar radiation at high resolutions: a comparison of time series forecasts. Sol Energy 2009;83(3):342–9.
[3] Li Yanting, Su Yan, Shu Lianjie. An ARMAX model for forecasting the power output of a grid connected photovoltaic system. Renew Energy 2014;66:78–89.
[4] Yang Chen, Thatte Anupam A, Xie Le. Multitime-scale data-driven spatio-temporal forecast of photovoltaic generation. IEEE Trans Sustain Energy 2015;6(1):104–12.
[5] Paoli Christophe, Voyant Cyril, Muselli Marc, Nivet Marie-Laure. Forecasting of preprocessed daily solar radiation time series using neural networks. Sol Energy 2010;84(12):2146–60.
[6] Almeida Marcelo Pinho, Perpiñán Oscar, Narvarte Luis. PV power forecast using a nonparametric PV model. Sol Energy 2015;115:354–68.
[7] Mandal Paras, Madhira Surya Teja Swarroop, Haque Ashraf Ul, Meng Julian, Pineda Ricardo L. Forecasting power output of solar photovoltaic system using wavelet transform and artificial intelligence techniques. Procedia Comput Sci 2012;12(915):332–7.
[8] Pedro Hugo TC, Coimbra Carlos FM. Assessment of forecasting techniques for solar power production with no exogenous inputs. Sol Energy 2012;86(7):2017–28.
[9] Olatomiwa Lanre, Mekhilef Saad, Shamshirband Shahaboddin, Mohammadi Kasra, Petković Dalibor, Sudheer Ch. A support vector machine firefly algorithm based model for global solar radiation prediction. Sol Energy 2015;115:632–44.
[10] Liu Hui, Tian Hong qi, Li Yan fei. Comparison of two new ARIMA-ANN and ARIMA-Kalman hybrid methods for wind speed prediction. Appl Energy 2012;98:415–24.
[11] Liu Hui, Tian Hong Qi, Pan Di Fu, Li Yan Fei. Forecasting models for wind speed using wavelet, wavelet packet, time series and Artificial Neural Networks. Appl Energy 2013;107:191–208.
[12] Zhao Yongning, Ye Lin, Li Zhi, Song Xuri, Lang Yansheng, Su Jian. A novel bidirectional mechanism based on time series model for wind power forecasting. Appl Energy 2016;177:793–803.
[13] Anava Oren, Hazan Elad, Zeevi Assaf. Online time series prediction with missing data. In: Proceedings of the 32nd international conference on machine learning, vol. 37; 2015.
[14] Abraham Bovas, Thavaneswaran A. A nonlinear time series model and estimation of missing observations. Ann Inst Stat Math 1991;43(3):493–504.
[15] Sinopoli Bruno, Schenato Luca, Franceschetti Massimo, Poolla Kameshwar, Jordan Michael I, Sastry Shankar S. Kalman filtering with intermittent observations. IEEE Trans Autom Control 2004;49(9):1453–64.
[16] Mohamed Shakir, Marwala Tshilidzi. Neural network based techniques for estimating missing data in databases; October, 2015.
[17] Honaker James, King Gary. What to do about missing values in time series cross-section data. Am J Polit Sci 2010;54(2):561–81.
[18] Pfenninger Stefan. Dealing with multiple decades of hourly wind and PV time series in energy models: a comparison of methods to reduce time resolution and the planning implications of inter-annual variability. Appl Energy 2017;197:1–13.
[19] Voyant Cyril, Darras Christophe, Muselli Marc, Paoli Christophe, Nivet Marie Laure, Poggi Philippe. Bayesian rules and stochastic models for high accuracy prediction of solar radiation. Appl Energy 2014;114:218–26.
[20] Hussain Sajid, Al-Alili Ali. A new approach for model validation in solar radiation using wavelet, phase and frequency coherence analysis. Appl Energy 2016;164:639–49.
[21] Deo Ravinesh C, Wen Xiaohu, Qi Feng. A wavelet-coupled support vector machine model for forecasting global incident solar radiation using limited meteorological dataset. Appl Energy 2016;168:568–93.
[22] Alessandrini S, Delle Monache L, Sperati S, Cervone G. An analog ensemble for short-term probabilistic solar power forecast. Appl Energy 2015;157:95–110.
[23] Caruana Rich, Mitchell Tom, Pomerleau Dean, Dietterich Tom, State Oregon. Multitask learning. Mach Learn Spec Issue Inductive Transf 1997;28:41–75.
[24] Xu Zhao, Kersting Kristian. Multi-task learning with task relations. In: 2011 IEEE 11th international conference on data mining. IEEE; 2011. p. 884–93.
[25] Ben-David S, Schuller Reba. Exploiting task relatedness for multiple task learning. Learn Theory Kernel Mach 2003:567–80.
[26] Leen Gayle, Peltonen Jaakko, Kaski Samuel. Focused multi-task learning using Gaussian processes. Mach Learn Knowl Discov 2011:1–16.
[27] Yu Kai, Tresp Volker, Schwaighofer Anton. Learning Gaussian processes from multiple tasks. In: Proceedings of the international conference on machine learning (ICML), vol. 22; 2005. p. 1012–9.
[28] Bonilla Edwin V, Chai Kian M, Williams Christopher. Multi-task Gaussian process prediction. Adv Neural Inform Process Syst 2007:153–60.
[29] Hayashi Kohei, Takenouchi Takashi, Tomioka Ryota, Kashima Hisashi. Self-measuring similarity for multi-task Gaussian process. Trans Jpn Soc Artif Intell 2012;27(3):103–10.
[30] Huang Shuai, Li Jing, Chen Kewei, Wu Teresa, Ye Jieping, Wu Xia, et al. A transfer learning approach for network modeling. IIE Trans Qual Reliab Eng 2012;44(11):915–31.
[31] Skolidis Grigorios, Hansen Katja, Sanguinetti Guido, Rupp Matthias. Multi-task learning for pKa prediction. J Comput-Aided Mol Des 2012;26(7):883–95.
[32] Skolidis Grigorios, Sanguinetti Guido. Bayesian multitask classification with Gaussian process priors. IEEE Trans Neural Netw 2011;22(12 PART 1):2011–21.
[33] Bonilla Edwin V, Agakov Felix V, Williams Christopher KI. Kernel multi-task learning using task-specific features. In: Proceedings of the eleventh international workshop on artificial intelligence and statistics, vol. 2; 2007. p. 43–50.
[34] Lawrence Neil D, Platt John. Learning to learn with the informative vector machine. In: Twenty-first international conference on machine learning ICML 04; 2004. p. 512–9.
[35] Pentina Anastasia, Sharmanska Viktoriia, Lampert Christoph H. Curriculum learning of multiple tasks. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2015. p. 5492–500.
[36] Kang Zhuoliang, Grauman Kristen, Sha Fei. Learning with whom to share in multi-task feature learning. In: Proceedings of the 28th international conference on machine learning (ICML-11); 2011. p. 521–8.
[37] Altman NS. An introduction to kernel and nearest-neighbor nonparametric regression. Am Stat 1992;46(3):175–85.