Iterative Multi-Task Learning on Spatial Time Series Data with Applications to Improvement of Performance Prediction and Monitoring for Solar Panels

Tahasin Shireen
ITERATIVE MULTI-TASK LEARNING ON SPATIAL TIME SERIES DATA
WITH APPLICATIONS TO IMPROVEMENT OF PERFORMANCE PREDICTION AND
MONITORING FOR SOLAR PANELS

By

TAHASIN SHIREEN

A Thesis submitted to the
Department of Industrial and Manufacturing Engineering
in partial fulfillment of the
requirements for the degree of
Master of Science

2016

Copyright © 2016 Tahasin Shireen. All Rights Reserved.
Tahasin Shireen defended this thesis on April 12, 2016.
The members of the supervisory committee were:

Hui Wang
Professor Directing Thesis

Mei Zhang
Committee Member

Abhishek K. Shrivastava
Committee Member

The Graduate School has verified and approved the above-named committee members, and certifies that the thesis has been approved in accordance with university requirements.
To my parents, who are my inspiration
ACKNOWLEDGMENTS

I would like to express my most sincere gratitude to my advisor Dr. Hui Wang for his continuous support and guidance during the research work. His encouragement and subject insights motivated me to face the challenges of my research study. I would also like to thank my other committee members, Dr. Mei Zhang and Dr. Abhishek K. Shrivastava for their invaluable suggestions and feedback on my research.

I am also grateful to Mr. Frank Allen, Ms. Judy Gardner, Mr. Jim Horne, Mr. John Taylor, and Ms. Stephanie Salters for their helpful assistance and cooperation. I am thankful to all faculty members, staffs and students in Department of Industrial and Manufacturing Engineering and High-Performance Materials Institute for their support during my graduate study.

Finally, I would like to thank my parents, my sister and my friends for their continuous encouragement and support throughout my life. This thesis paper is an outcome of their confidence and belief in me.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of Tables</td>
<td>vii</td>
</tr>
<tr>
<td>List of Figures</td>
<td>viii</td>
</tr>
<tr>
<td>Abstract</td>
<td>x</td>
</tr>
<tr>
<td>1 Introduction</td>
<td>1</td>
</tr>
<tr>
<td>2 Literature Review</td>
<td>4</td>
</tr>
<tr>
<td>3 Research Gaps and Objectives</td>
<td>8</td>
</tr>
<tr>
<td>3.1 A Summary of Research Gaps</td>
<td>8</td>
</tr>
<tr>
<td>3.2 Research Objectives</td>
<td>8</td>
</tr>
<tr>
<td>4 Iterative Multi-task Learning on Spatial Time Series</td>
<td>10</td>
</tr>
<tr>
<td>4.1 Methodology</td>
<td>10</td>
</tr>
<tr>
<td>4.1.1 A review of multi-task learning for Gaussian process</td>
<td>10</td>
</tr>
<tr>
<td>4.1.2 Proposed multi-task learning Gaussian process time series algorithm (MTL-GP-TS)</td>
<td>11</td>
</tr>
<tr>
<td>4.2 Case Study</td>
<td>12</td>
</tr>
<tr>
<td>4.2.1 Preparation of solar panel dataset</td>
<td>14</td>
</tr>
<tr>
<td>4.2.2 Prediction of solar panel performance using MTL-GP-TS</td>
<td>14</td>
</tr>
<tr>
<td>4.2.3 Results</td>
<td>18</td>
</tr>
<tr>
<td>4.3 Discussion</td>
<td>18</td>
</tr>
<tr>
<td>4.3.1 Application of MTL-GP-TS in maintenance of solar panels</td>
<td>18</td>
</tr>
<tr>
<td>4.3.2 Solar panel prognosis when historical data are insufficient</td>
<td>20</td>
</tr>
<tr>
<td>4.3.3 Hyper parameter selection for MTL-GP-TS algorithm</td>
<td>22</td>
</tr>
<tr>
<td>4.3.4 Effect of incorporating all available data sources in MTL-GP-TS algorithm</td>
<td>22</td>
</tr>
<tr>
<td>5 Weather Factors Consideration on MTL-GP-TS for Spatial Time Series</td>
<td>24</td>
</tr>
<tr>
<td>5.1 Motivation</td>
<td>24</td>
</tr>
<tr>
<td>5.2 Possible Approaches for MTL-GP-TS Method with Weather Consideration</td>
<td>26</td>
</tr>
<tr>
<td>5.3 Case Study</td>
<td>27</td>
</tr>
<tr>
<td>5.3.1 Description of solar panel dataset</td>
<td>28</td>
</tr>
<tr>
<td>5.3.2 Solar panel performance prediction by MTL-GP-TS with weather consideration</td>
<td>29</td>
</tr>
<tr>
<td>5.3.3 A summary of results</td>
<td>30</td>
</tr>
<tr>
<td>5.4 Discussion</td>
<td>31</td>
</tr>
<tr>
<td>5.4.1 Performance monitoring of solar panels with weather consideration</td>
<td>31</td>
</tr>
<tr>
<td>5.4.2 Improvement of MTL-GP-TS with weather consideration</td>
<td>34</td>
</tr>
<tr>
<td>6 Conclusion</td>
<td>37</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

4.1 RMSE-forecast by STL and MTL-GP-TS ........................................ 18
4.2 RMSE-forecast of H school with different number of data sources .......... 22
5.1 RMSE-forecast by STL and MTL-GP-TS with weather consideration ............ 30
# LIST OF FIGURES

1.1 Components of time series (Decomposition). .................................................. 1
1.2 Plot of solar panel time series observations with missing values. ......................... 2
2.1 Representation of the idea of multi-task learning. ............................................. 6
4.1 Flowchart of MTL-GP-TS algorithm. ................................................................. 13
4.2 Schematic representation of MTL-GP-TS method. ............................................. 15
4.3 From top, MTL-GP-TS learned trend plots after removing noise for N, J and H schools respectively. .......................................................... 16
4.4 Comparison of RMSE-forecast of MTL-GP-TS and STL for N, J and H solar panel data sources. .......................................................... 17
4.5 Percentage improvement in RMSE-forecast by MTL-GP-TS for N, J and H solar panel data sources. .......................................................... 17
4.6 Performance monitoring for school H solar panels with forecasted trend ............... 19
4.7 Response surface plots of hyper parameters with RMSE ................................... 21
4.8 Results of best subset selection method for school H as target data source. ........ 23
5.1 Schematic diagram of predicting time series observations from predictors with temporal relation .......................................................... 25
5.2 Diagram of MTL-GP-TS with weather factors consideration using PV-Weather data integration model. .......................................................... 28
5.3 Comparison of RMSE-forecast of MTL-GP-TS and STL for C and K solar panel data sources while considering weather factors. ......................... 31
5.4 Percentage improvement in RMSE-forecast by MTL-GP-TS for C and K solar panel data sources while considering weather factors. ......................... 32
5.5 Performance monitoring for school K solar panels when weather factors are considered 33
5.6 Comparison of RMSE-forecast of MTL-GP-TS and STL on dataset for C and K solar panel data sources without weather factors consideration (non seasonal time series). 34
5.7 Comparison of %RMSE-forecast improvement of degradation trend while considering (seasonal time series) and ignoring weather factors (non seasonal time series) respectively for C and K solar panel data sources. ......................... 35
7.1 (A) Growth of nanowire: ZnO nanowires grown on Si(100) substrate. (B) Au catalyzed silicone nanowire growth (VLS mechanism)
ABSTRACT

Health condition monitoring and failure detection play a crucial role in estimating the performance of solar panels such as degradation trend over time and occurrence of failures. Monitoring and detecting significant degradation can help solar panel owners establish as-needed maintenance strategies on a timely manner. But in some occasions, degradation trend estimation becomes challenging due to limited availability of training data such as many missing observations in time series over a large time span and a lack of history of failure records that are sufficient to establish statistical models. To fill the gap, this thesis proposes a new approach of iterative multi-task learning of Gaussian process in time series data (MTL-GP-TS) by sharing information among similar-but-not-identical datasets from multiple solar panel locations. The proposed MTL-GP-TS model learns unobserved or missing values in a particular time series dataset to forecast the future trend with autoregressive integrated moving average (ARIMA) model, resulting in substantial improvement of forecast over conventional time series modeling approaches. Moreover, the estimated degradation trend with proposed MTL-GP-TS method has the potential to improve the monitoring of significant performance degradation compared with the conventional time series model. This thesis also studies the effect of temporal dependent weather factors on the solar panel performance by integrating a covariate with the MTL-GP-TS algorithm. A case study has demonstrated that inclusion of weather factors into the monitoring of degradation with PV-Weather data integration model can significantly improve the solar panel performance prediction.
CHAPTER 1
INTRODUCTION

Monitoring of performance degradation trend is important for the evaluation of life-cycle assessment and efficiency of solar panels. Degradation in solar panels can occur due to variety of reasons like infiltration of contaminants, soiling, migration of metal ions through solar cells, deterioration of anti reflecting, etc. An accurate estimation of degradation trend can help the solar panels owners and the manufacturers perform as-needed preventive maintenance before any failure occurrence. The degradation trend estimation and/or occurrence of failures monitoring are usually conducted based on a time series dataset, which is collected at a specific time interval. The components of time series are shown in Fig. 1.1, where our interest of study are “Trend” and “Irregular components”.

Such time series models for solar panel data require sufficient training data to establish adequate statistical confidence. However, one of the major challenges in estimation of degradation trend is the limited availability of training data. Also, it has been observed that in some occasions, observations over certain timespan could be missing for a variety of reasons. For example, data collection may be interrupted due to technical problems, data measurement equipment failure, erroneous measurement, weather factors, etc. First, there may be a gap in time series data collection by frequent disruptions from unavoidable external factors. In such cases, the missing values with limited training data in time series can result in poor estimation of future forecasts, causing misdetection or false alarms of potential failures in solar panels. Second, the failure records of solar panels may not always exist, making it a grand challenge to forecast when certain failure occurs for the first time. Therefore, it is essential to develop an efficient way of monitoring significant
performance degradation of solar panels when a limited amount of training data is available to fit the time series model. The main problem highlighted in this thesis is limited availability of training data to fit time series as shown in Fig.1.2 with an example of solar panels.

This thesis envisions an opportunity of improving the estimation of learning missing observations for better prediction of trend forecast by complementing information from other similar-but-non-identical solar panel data and environment. If a particular solar panel time series lacks sufficient dataset, one can gather information from other solar panel time series data that share similarity with that particular solar panel of interest. In this thesis, the main study of interest is improvement of degradation (slow or gradual) trend monitoring based on limited time series observations. Also if the study of interest is a particular failure mode like corrosion occurrences in solar panels over a period which occur infrequently, one can learn better about future corrosion occurrences by fusing information from other available time series data sources. Another study of interest is the effect of weather factors on the performance degradation of solar panels over years. Weather factors like solar irradiance, wind speed, rainfall, humidity can affect the performance of solar panel output by certain extent as the weather changes with seasons. Incorporation of the weather factor as covariate
can potentially help further improve the monitoring of solar panel performance degradation based on spatial time series data.
CHAPTER 2

LITERATURE REVIEW

This section focuses on several forecasting techniques in solar panel data and the conventional methods of dealing with missing observations in time series modeling. Based on the current existing approaches, future scope of improvement is also discussed at the later part of this section.

In recent years there have been a noticeable amount of works to forecast time series data in solar panels and solar radiation. A comparison between different time series forecasting techniques like ARIMA [1], transfer function, neural networks [2], hybrid models [3] and regression in logs, was carried out in [4] for predicting solar radiation at high resolution. In [5] artificial neural networks with multi-layer perceptron is used to forecast the solar daily radiation in time series dataset using transfer function of hidden layers. Using the MLP, Paoli et al. learned the prediction with a fixed number of $x$ past values considered as inputs and outputs generated as the predicted future time series observations. A moving time window technique referred as “sliding window technique”, is used to select $Y$ times $x$ training inputs. Among the non parametric forecasting approaches, Almeida et al. presented a non parametric model of AC power forecast with quantile regression forests as machine learning tools in [6]. Quantile regression forests keeps the values of all observations in every leaf of random forests, not just their weighted mean and then assesses the conditional distribution based on this information which enables the construction of prediction intervals. The earlier works of forecast with non parametric approach are also found in Mandal et al. [7], Pedro and Coimbra [8] and Zamo et al. [9]. In [7], Mandal et al. used a combination of wavelet transform and neural network to forecast one hour ahead PV output. Another approach of hybrid method using support vector machine with firefly algorithm (FFA) was introduced by Olatomiwa et al. [10] to forecast the solar radiation prediction.

The most conventional forecast techniques studied in earlier years include ARIMA [11], K-nearest neighbors [12], Markov chains [13] and Bayesian inference [14]. In ARIMA models, auto regression is combined with moving average after differencing to make the dataset stationary. ARIMA $(a, b, d)$ models are described as a non seasonal model where $a$ is the order of auto regressive part, $b$
is the degree of first differencing involved and \( d \) is order of the moving average part [11]. In Bayesian Inference method, observations are taken as inputs to update the probability that a hypothesis may be true for that case [14]. Regarding Markov chains forecasting technique, Logofet et al. explained in [13] that a Markov process is a stochastic process with the Markov property. Markov property can be explained by taking future observations independent of past values while the present observations are already given. These previous works of forecasting related to solar panels highlight the importance of forecasting the degradation trend in solar panel time series dataset. Selection of right forecasting technique among the available forecasting methods can significantly improve the learning of degradation trend of a time series.

The main challenge of forecasting methods is how to fit a time series model in order to forecast future values when the dataset contains many missing observations with limited training data. Previously, the missing observations in time series dataset was dealt with by several methods like interpolation [15], AR predictor [16], auto regressive conditionally heteroscedastic [17], auto regressive method with similarity matrix [18], Kalman filtering [19, 20], neural networks [21] and multiple imputation model [22]. In [15], a long time series with missing observations was divided into multiple subseries to find out the ARIMA models for each subseries while interpolating each subseries with the identified model and then reestimating the parameters. Multiple imputation models were mostly used in political science dataset where missing values are present in multiple variables. It is mostly used where missing values are present in multivariate dataset. A strong assumption of MAR (missing at random) in dataset was held for the multiple imputation model algorithm [22], which means missing values in the dataset has conditional dependency on the observed variables. In general cases, the occurrence of missing values in time series data is dealt by interpolation methods which can be considered as single task learning (STL) technique for this thesis. But on some particular scenarios, when availability of training data over a long time period is limited in time series, interpolation technique may not be an efficient way to learn missing values to capture the true values in trend estimation.

A new opportunity exists to improve the learning of the unobserved values by sharing knowledge or information available in each task, such as transferring knowledge between similar tasks with multi-task learning. Multi-task learning is a machine learning framework aiming to improve performance through the learning of multiple tasks simultaneously while sharing the information...
of each task [23] as shown in Fig. 2.1. In multi task learning, the term ‘task’ defines learning and predicting each observation with multiple datasets. The sharing information among similar types of tasks is the key idea in multi-task learning GP model to learn the missing information in individual tasks. Joint learning of all tasks with multi-task learning allows sharing information between different tasks to significantly reduce problems like overfitting and unstable search due to sparse training data while improving performance compared to single task models or no transfer case learning [24]. In order to share information among tasks, the similarity or commonality among the tasks can be assumed by expert knowledge or explained by a high-level empirical structure. Relatedness among tasks is vital in multi-task learning [25] as unrelated or dissimilar tasks can be detrimental for effective learning, aiding negative transfer of information [26]. To exploit the idea of task relatedness, Gaussian process models have been applied to symmetric multi-task learning scenarios based on joint priors for functions underlying the tasks for learning a specific task while incorporating knowledge learned through other similar tasks [26]. In [27], a self measuring similar-
ity model was introduced to measure the degree of relatedness among tasks for multi-task learning with a covariance matrix by the responses themselves without requiring any input or task specific feature.

In hierarchical Bayesian modeling, parameters are assumed to be drawn from a common hyperprior distribution in order to jointly learn particular task that regularize each other while sharing information [28]. The underlying structure of tasks in MTL has also been studied with multi-task sparse learning in [29]. Further, the parametric representation of multi-task Gaussian process can be found in the study of [23] that uses a shared covariance function to infer the parameters of the model. Later Bonilla et al. proposed another approach of multi-task learning GP process with a free form covariance matrix to represent the inter task dependency while avoiding the requirement of large training data [30]. In [31], a new approach of transfer learning is introduced by sharing information from two sources, information specific to each task and common information of all tasks in IC matrix. Using Gaussian graphical models, task relatedness is represented by estimating IC matrix for samples taken from wishart distribution under Bayesian framework, focusing on graphical modeling of networks with multiple tasks [31]. With the benefit of improved learning of a target task by sharing information among similar tasks over single task learning, the scope of multi-task learning for prediction has been previously explored in [32–35] for classification problems, in [36] and [37] with features and in [38] with informative vector machines for learning parameters when enough observations were not available in target dataset but observations were available in similar related datasets. The study of temporal relation among inputs of MTL has also been evaluated in [39] for predicting the Alzheimer’s disease progression. Lastly, forecasting of spatiotemporal events with multi task learning has been studied in [40] while considering dynamic patterns of features and heterogeneity in geographical location for improving forecast by increasing sample size for each location.
CHAPTER 3

RESEARCH GAPS AND OBJECTIVES

3.1 A Summary of Research Gaps

Based on the literature review, two research gaps have been identified. These are:

- **Application gap**: A lack of research or method that uses multi-task learning in time series for improving the prediction of performance monitoring of solar panels based on the opportunity of multiple similar-but-non-identical solar panel data and multi-task learning. One challenge is that significant performance degradation events such as failure induced by corrosion or soiling do not frequently occur though they generate significant impact. As such, historical data may not be sufficient to make prediction with sufficient statistical confidence. There is a gap to address the problem of insufficient training observations for performing trend monitoring with MTL approach in spatial time series.

- **Theoretical gap**: A lack of approach considering temporal dependency in time series data for multi-task learning algorithm. Current MTL method was mostly developed for Guassian spatial process. However, the time series data may not necessarily satisfy the GP assumptions. An improved MTL tailored for time series data is highly desired.

3.2 Research Objectives

This thesis proposes a cost-effective approach to addressing the issue of improving prediction of missing observations in time series dataset in an iterative manner by sharing common information among similarly related spatial time series datasets. To our knowledge, the proposed approach has not been explored before to estimate future degradation trend of solar panel time series dataset. In this research, a multi-task learning Gaussian process time series (MTL-GP-TS) algorithm is introduced to learn the missing observations in time series dataset by updating the trend at each iteration. The implementation of MTL GP method in spatiotemporal dataset may improve prediction of missing values in time series data by an iterative approach as there is a possibility of using multiple similar time series datasets from other locations that share some common information about the dataset of interest from spatiotemporal perspective. Based on two case studies with
solar panel data from Hawaiian schools, we will demonstrate the proposed MTL-GP-TS method to learn missing observations in time series dataset and thereby predict and monitor the performance of solar panels. The research objectives of the thesis can be summarized as:

- Development of an iterative multi-task learning Gaussian process in time series (MTL-GP-TS) algorithm: Chapter 4 will develop an iterative MTL-GP-TS algorithm for learning missing observations in spatial time series dataset by joint learning of similar other time series datasets.

- Monitoring of future performance of solar panels based on MTL-GP-TS: The effectiveness of developed MTL-GP-TS algorithm is analyzed in Chapter 4 by monitoring future performance of solar panels with the proposed algorithm for the target time series with limited training data.

- Improvement of degradation monitoring considering weather factors on MTL-GP-TS: Chapter 5 will improve trend estimation by incorporating different weather factors into MTL-GP-TS for monitoring degradation while sharing similarity between multiple time series datasets.
CHAPTER 4
ITERATIVE MULTI-TASK LEARNING ON SPATIAL TIME SERIES

4.1 Methodology

The parametric representation of MTL-GP uses a shared covariance function to infer the model parameters [23]. In hierarchical Bayesian framework, it is assumed that the model parameters of each individual dataset are drawn from a common hyper prior distribution to learn individual models [28]. For creating the common hyper prior distribution, a kernel is learned from the available existing information that reflects all the information from multiple tasks. This method has the computational advantages in parameter estimation and flexibility for incorporating time series data into the MTL algorithm. In this thesis, we adopt and improve this method through establishing a MTL-GP-TS learning algorithm for spatiotemporal data of solar panels. Section 4.1.1 first reviews the MTL algorithm for GP.

4.1.1 A review of multi-task learning for Gaussian process

The purpose of multi-task learning is to learn $m$ related functions $g_l$ where $l = (1,2,...,m)$ from dataset $D_l$. Define $D_l= (P_l, q_l)$ where $P_l$ is the training data and $q_l$ is the predicted values by MTL-GP model and assume that $n_l$ is number of training points of function $g_l$. Among the different sets of training data points for each function task, we select only the distinct $n$ points in dataset $D_l$.

In MTL-GP model, the mean $\mu_g$ and covariance function $K$ of $g_l$ values, follow a normal inverse Wishart distribution associated with a positive definite base kernel $\kappa$. The base kernel $\kappa$ describes the properties of shared Wishart prior [14], capturing the effect between each individual point and the neighboring points to estimate the missing observations in dataset of $D_l$. Let $P$ be a set of data points where $\forall p, p' \in P$, then the positive base kernel is $\kappa(p, p') = \langle p, p' \rangle$. In inductive setting of GP model, the uniqueness of mean $\mu_g$ of each location $g_l$ and covariance function $K$ is utilized to represent the distribution of hyper priors. Each function $g_l$ has a unique $\gamma \in R^n$, where
function $g_l = \kappa_{\gamma_l}, \mu_{\gamma_l} \in \mathbb{R}^n, C_{\gamma_l} \in \mathbb{R}^{nxn}$ and $\gamma \sim \mathcal{N}(\mu_{\gamma_l}, C_{\gamma_l})$. The unique $\mu_{\gamma_l}$ and $C_{\gamma_l}$ jointly follow normal inverse Wishart distribution with scale matrix $\kappa^{-1}$ [28]. The parameters $\mu_{\gamma_l}$ and $C_{\gamma_l}$ can be estimated from sampling from the hyper prior distribution once and thus for each function $g_l$, $\gamma_l \sim \mathcal{N}(\mu_{\gamma_l}, C_{\gamma_l})$, where $g_l$ be values of $g_l$ on $P$ data points set where $\cup P_l \subseteq P$ and error term in prediction of test points, $\epsilon \sim \mathcal{N}(0, \sigma^2)$. The parameters $\Psi = (\mu_{\gamma_l}, C_{\gamma_l}, \sigma^2)$ can be learned by an Expectation-Maximization (EM) algorithm [41]. The estimated inductive function is shown below in Eq.(4.1.1).

$$g_l(p) = \sum_{i=1}^{n} \gamma_i \kappa(p_i, p) \tag{4.1.1}$$

It is assumed that prior estimates in the MTL-GP model can be characterized by a Gaussian process with mean and covariance function following a normal-inverse Wishart distribution.

### 4.1.2 Proposed multi-task learning Gaussian process time series algorithm (MTL-GP-TS)

A framework of MTL-GP model for time series data is introduced to learn missing observations. We first decompose the time series data into three parts including trend, seasonal component (if applicable), and irregular/random component. After detrending and deseasonalizing a time series dataset, the remaining component left to study, is the irregular or random component. The irregular component of time series is usually assumed to follow a Gaussian process. As such, MTL-GP method can be employed to transfer information among irregular components from similar time series datasets. The prediction through such information sharing is anticipated to be improved when training data from each source is limited. However, learning of the irregular components is also influenced by the way trend and seasonal components are decomposed. Thus, joint consideration of the trend estimation and MTL for irregular components poses a new challenge. To fill the gap, this thesis develops an iterative MTL-GP-TS algorithm to learn the missing observations by updating the trend of time series dataset $^1$. The trend estimated at each iteration of MTL-GP-TS reflects the temporal order between observations that is ensured by taking the irregular components inputs for MTL following the exact order of the time series observations. The final trend (denoted as $T_f(t)$) learned at last iteration via MTL-GP-TS algorithm is then fitted into ARIMA models to forecast the future observations or the degradation trend. The procedure is summarized in Fig. 4.1 for non

---

$^1$This work of iterative MTL-GP-TS algorithm on spatial time series data has been submitted to Solar Energy, an Elsevier Journal, in 2016.
seasonal time series. Similar method is also applicable to seasonal time series including seasonal components.

**MTL-GP-TS algorithm for irregular component of non seasonal time series dataset:**

1. Start with initial time series data $Z_0(t)$ of size $(n \times 1)$ including $M$ missing observations and $P$ observed training data, where $M \cup P \subseteq Z_0(t)$

2. Apply interpolation on $Z_0(t)$ and generate updated time series data $Z(t)$

3. Set $i = 0$ for initialization

4. Generate irregular component, $\epsilon_i(t)$ by time series decomposition

5. Estimate $\epsilon_{m_i}(t)$ with MTL-GP algorithm

6. Update trend by subtracting $\epsilon_{m_i}(t)$ from $Z_0(t)$ and then apply interpolation to learn the updated trend, $T_u(t) = T_{i+1}(t)$

7. Check if $|\Delta \text{RMSE}_{i+1}|\% = \left| \frac{\text{RMSE}_{i+1} - \text{RMSE}_i}{\text{RMSE}_i} \right| < \delta$ where RMSE is root-mean-square error and $\delta$ is assumed a very small or negligible value. If no, go to next iteration by updating iteration sequence, $i = i + 1$ and again start with step 4. If yes, get the final trend, $T_f(t)$.

ARIMA models are then be fitted to the final estimated trend $T_f(t)$ via MTL-GP-TS to forecast future performance or likelihood of failures. It should be noted that noise may occur in estimated trend due to reading error of data measurement instruments or fluctuation in weather factors. Therefore, the noise should be removed first using smooth moving average (SMA) method.

### 4.2 Case Study

MTL-GP-TS is demonstrated with a case study based on solar panel dataset consisting of solar panel power output (PV, unit: kilowatts) and plane of array solar irradiance (POA, unit: kilowatts per meter square) from four schools in Hawaii (data source available at \(^2\)). In addition, solar panel degradation data were simulated using a benchmark study of different solar panel manufacturers \(^3\) \(^4\) \(^5\). The school locations are selected as four data sources based on the information that solar panel power output and plane of array solar irradiance are significantly impacted by environmental factors such as temperature, humidity, and air pressure.

Start

Initial raw data, $Z_0(t)$

Interpolation to fill up missing value

Interpolated data, $Z(t)$

Decomposition

Irregular component, $\epsilon_i(t) = Z(t) - \text{Trend, } T_i(t)$

Multi-task learning

MTL irregular component = $\epsilon_{m_i}(t)$

Trend update = $Z_0(t) - \epsilon_{m_i}(t)$

Interpolation to fill up missing value

Updated trend = $T_{i+1}(t)$

$|\Delta RMSE_{i+1}|% < \delta$?

No

Update iteration sequence: $i = i + 1$

Yes

Final trend, $T_f(t) = T_{i+1}(t)$

End

Figure 4.1: Flowchart of MTL-GP-TS algorithm.
panels are manufactured by the same manufacturer (i.e., Sunpower). The simulated data using the benchmark study is the fifth data source.

4.2.1 Preparation of solar panel dataset

The solar panel data were simulated over a ten-year timespan based on the data collected from Jan’2008 to Dec’2011. The data sources are from Nanakuli High and Intermediate school (N), Jarrett Middle school (J), Highlands Intermediate school (H), Waianae High school (W) and “Sunpower” manufacturers (S). All data sources show strong positive correlation between the PV and POA value. Thus, the ratio of PV over POA (i.e., $PV/POA$) is used as time series observations to diminish the effect of seasonality in the dataset. As a result, the PV-POA ratio data can be considered as non-seasonal datasets.

The dataset contains training, testing, and validation sets. Random selection of training data and test data is generated with cross validation. The generated dataset has total 480 time points with weekly interval over 10 years. The training and testing set contain 77 data points and 307 data points, respectively. The test data points are considered as missing observations in time series. The validation set includes 96 data points with weekly time interval from Years 9 and 10 for each solar panel data source. The representation of time series is arranged in a way where each observation $Z_{ij}$ in time series data belongs to $i_{th}$ month $j_{th}$ week where $i = (1, 2, ..., k)$ and $j = (1, 2, 3, 4)$.

4.2.2 Prediction of solar panel performance using MTL-GP-TS

In MTL-GP-TS algorithm, the missing observations are learned by sharing information among multiple time series if the datasets are from a similar source. For example, as shown in Fig. 4.2, three solar panel time series data can be learned jointly using MTL-GP-TS by assuming solar panels are of the same type from the same manufacturer. According to [28], the base kernel matrix shares the information among the available training data from each task and EM algorithm estimates the parameters for each solar panel location. In this study, EM algorithm reached convergence within 5000 maximum iterations starting with the initial values of hyper parameters.

The proposed MTL-GP-TS method was compared with traditional time series modeling based on one single data source (i.e., STL). The degradation trend were estimated using ARIMA models once the missing observations are learned by the two methods (MTL-GP-TS and STL). The degradation trend for next two years (Years 9-10) is forecasted for the solar panel of interest (target).
The residuals from predicted values follow a normal distribution, ensuring appropriate selection of ARIMA models. The autocorrelation (ACF) plot further shows that there is no correlation between the residuals, indicating that the selected ARIMA models yield the best forecast predictions with the lowest RMSEs. The improvement in forecasting future trend values of solar panels were also measured by comparing RMSE values of both MTL-GP-TS and STL method. This RMSE is specified as “RMSE-forecast” in the remainder of the thesis paper.
Figure 4.3: From top, MTL-GP-TS learned trend plots after removing noise for N, J and H schools respectively.
Figure 4.4: Comparison of RMSE-forecast of MTL-GP-TS and STL for N, J and H solar panel data sources.

Figure 4.5: Percentage improvement in RMSE-forecast by MTL-GP-TS for N, J and H solar panel data sources.
### 4.2.3 Results

Table 4.1: RMSE-forecast by STL and MTL-GP-TS

<table>
<thead>
<tr>
<th>Solar panel data source</th>
<th>RMSE-STL</th>
<th>RMSE-MTL-GP-TS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nanakuli</td>
<td>0.11</td>
<td>0.09</td>
</tr>
<tr>
<td>Jarrett</td>
<td>0.19</td>
<td>0.07</td>
</tr>
<tr>
<td>Highlands</td>
<td>0.45</td>
<td>0.15</td>
</tr>
</tbody>
</table>

The RMSE-forecast values presented in Table 4.1 for three target solar panels compares the outcome from both STL and MTL-GP-TS method. The target solar panels are from N, J and H schools. The MTL-GP-TS trend value after removing noise for the target solar panels are shown in Fig. 4.3. For all target solar panels, MTL-GP-TS demonstrated substantial improvement in reducing RMSE of forecasted values as shown in Fig. 4.4. The RMSE-forecast improvements by MTL-GP-TS are 18.18% for solar panels in N school, 63.16% for solar panels in J school and 66.67% for solar panels in H school as shown in Fig. 4.5 while comparing with STL. The improvement in RMSE-forecast by MTL-GP-TS indicates that MTL can outperform traditional STL method in forecasting degradation trend when time series dataset contains missing observations.

### 4.3 Discussion

#### 4.3.1 Application of MTL-GP-TS in maintenance of solar panels

Improved forecast predictions of the degradation trend by MTL-GP-TS can help identify the possible failure occurrences in solar panels. Proper monitoring or identification of potential failure occurrences can warn the owner or manufacturer beforehand to take preventive maintenance if the estimated degradation trend falls below a predetermined threshold value. Without losing generality, the threshold value for significant degradation monitoring can be determined by converting the percentage degradation of a solar panel over its life cycle under ideal operation conditions, to the output value of solar panel after certain years of usage. In this case study, the threshold value of \((PV/POA)\) is assumed to be 1 to check for potential failure occurrences in three target schools.
Figure 4.6: Performance monitoring for school H solar panels with forecasted trend. The plot shows trend forecast estimates of solar panels of H school with prediction intervals by MTL-GP-TS and STL for Years 9 and 10 with 7 days interval. The solid line represents point estimates of solar panels by MTL-GP-TS. The dash and dot combined line represents point estimates of solar panels by STL. The long dash and dot combined line represents true trend value. The dot and dash lines represent bound of trend forecasts by MTL-GP-TS and STL respectively. The circle shows that only the lower bound of MTL-GP-TS can monitor significant performance degradation of solar panels on 3rd week of April in Year 9.

Fig. 4.6 shows the degradation trend forecast of H school by the proposed MTL-GP-TS and STL methods for forecasted period during Years 9 and 10. Degradation trend estimated by MTL-GP-TS can effectively monitor performance to anticipate potential failure occurrences for all target solar panel data sources (N, J and H school). By contrast, STL failed to observe significant degradation in one of the target solar panel data sources (H school). The degradation trend was estimated in the forecasted period by both MTL-GP-TS and STL with 90% prediction interval. If the lower bound of the prediction intervals falls below the predetermined threshold value, the estimated trend may indicate a significant performance degradation leading to potential failure occurrence in future. In
In this case study, we checked for probable significant degradation on the 3rd week of April in Year 9 during the forecasted time period. Our findings can be summarized as follows, i.e.,

- Only the lower bound of MTL-GP-TS fell below the threshold value in H school solar panels on 3rd week of April in Year 9, implying a warning for potential failure occurrence at that time or near future.
- The estimated trend with STL showed frequent fluctuation in the forecasted values, jeopardizing the monitoring capability.
- The bandwidth of the prediction intervals for MTL-GP-TS was relatively constant over time whereas that of STL expand drastically, rendering STL less effective for future forecasts. This happens as the inputs of STL trend for fitting into ARIMA shows significant peaks which results in large prediction interval bounds while MTL-GP-TS trend inputs show very less fluctuation or peaks resulting in narrow prediction interval bounds.
- The true trend of H school from the original data was closely aligned with the point estimates of MTL-GP-TS trend and also within the prediction interval. On the contrary, point estimates of the STL trend were above the true trend with a wide prediction interval.

### 4.3.2 Solar panel prognosis when historical data are insufficient

With proper prediction of failures (prognosis capability) by MTL-GP-TS, solar panel owners can take preventive measure to avoid any interruption in normal operation. In some occasions when target data source lacks historical records of a particular failure mode (i.e., corrosion in solar panel) to forecast future failure occurrences, available information on failure mode from other data sources can be transferred via MTL-GP-TS to learn about the likelihood of certain failure occurrence in the target. Specifically, we can first segment the time into a number of periods so that the frequency of failure occurrences can be collected. The failure occurrence at certain time segment \((ts = 1, 2, 3... \) can be represented by an indicator variable \(X\) in nominal scale where \(X=1\) indicates the failure occurrence and 0 otherwise. The failure mode of target data source can be jointly learned with MTL and this newly learned information can then be used as the predictors for forecasting any possible failure warnings. A logistic regression [42] presented in Eq.(4.3.1) can be employed to predict the probability of certain failure \(\text{p}(X)\), thus helping the solar panel owners decide if preventive maintenance is required or not to reduce the chances of future corrosion occurrences.

\[
\log \left[ \frac{\text{p}(X; ts)}{1 - \text{p}(X; ts)} \right] = \beta_0 + \beta_1 Z_{ts-1} + ... + \beta_p Z_{ts-p} - \theta_1 \epsilon_{ts-1} - ... - \theta_q \epsilon_{ts-q}, \quad (4.3.1)
\]
Figure 4.7: Response surface plots of hyper parameters with RMSE. Starting from top, plot 1 shows RMSE results with hyper parameter $\tau$ and $\pi$ when $\sigma^2$ is fixed. Plot 2 shows RMSE results with hyper parameter $\tau$ and $\sigma^2$ when $\pi$ is fixed. Plot 3 shows RMSE results with hyper parameter $\pi$ and $\sigma^2$ when $\tau$ is fixed.
where $Z_i$ is the predictor variable such as $PV/POA$, weather conditions, and/or humidity, and $\epsilon_i$ is the forecast error within time segment $ts$.

4.3.3 Hyper parameter selection for MTL-GP-TS algorithm

The improvement by MTL-GP-TS method also depends on the best selection of hyper parameters in MTL algorithm. In this case study, each hyper parameter is assumed to take a number of discrete values within certain range. The hyper parameter values for $\tau$, $\pi$ and $\sigma^2$ 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. It has been observed that the smaller $\tau$, $\pi$ and $\sigma^2$ values could yield prediction results with lower RMSEs. For the target solar panel data sources, the prediction results with the lowest RMSE were found when $\tau=20$, $0.0001< \pi <0.1$ and $\sigma^2=1$. Fig. 4.7 shows the response surface plots of hyper parameters with RMSE. For the generation of response surface plots, one hyper parameter was always kept constant at a certain value while exploring lowest RMSE within the range of other two hyper parameters.

4.3.4 Effect of incorporating all available data sources in MTL-GP-TS algorithm

In MTL, it is not always necessary that inclusion of more data sources will improve prediction results [43, 44]. Increasing data sources may have very slight or negligible improvement on RMSE-forecast for the given solar panel data sources. For instance, inclusion of all five solar panel data sources into MTL-GP-TS algorithm to predict unobserved values in target solar panel data sources resulted in larger or same RMSE-forecast than that estimated by including only four or three data sources. Therefore, task selection is necessary to identify the potential data sources prior to multi-task learning.

Table 4.2: RMSE-forecast of H school with different number of data sources

<table>
<thead>
<tr>
<th>No. of data sources in MTL-GP-TS</th>
<th>RMSE-forecast of H school</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.15</td>
</tr>
<tr>
<td>4</td>
<td>0.14</td>
</tr>
<tr>
<td>3</td>
<td>0.14</td>
</tr>
</tbody>
</table>
A best subset selection method [45] was performed to identify the potential data sources to be included in MTL-GP-TS algorithm. As an example shown in Fig. 4.8, for target data source H school, the best subset selection method chose only two (N and J schools) among other available data sources (N, J, W and S) with both validation and cross validation approaches, which show that inclusion of only two data sources - N and J schools yield the lowest test error. To demonstrate the findings from the best subset selection, MTL-GP-TS joint learning of H school using only N and J schools’ solar panels was compared with joint learning of all available data sources by comparing RMSE-forecast values for Years 9 and 10. The RMSE-forecast values of H school when using five, four and three other data sources are shown in Table: 4.2. It can be seen that MTL-GP-TS using only three data sources (N, J and H schools) yield lower RMSE-forecast 0.14.

Figure 4.8: Results of best subset selection method for school H as target data source.
CHAPTER 5
WEATHER FACTORS CONSIDERATION ON MTL-GP-TS FOR SPATIAL TIME SERIES

The performance of solar panels is also affected by weather conditions. Thus, such weather factors are anticipated to exhibit some degree of correlation to the solar panel power output and can be considered as a co-variate (secondary variable) in addition to the primary variable (output power of solar panel PV). Both the primary variable and secondary variables (weather factors) are time series observations for respective solar panels. In this thesis, the solar panel power output (PV) is considered as the primary variables to monitor degradation or performance of solar panels. This chapter further considers the correlation between primary and secondary variables in time series modeling for solar panel data.

5.1 Motivation

When primary variable is not sufficient enough to make the predictions of missing time series observations, there is a possibility that the secondary variables can provide useful information about that particular missing observations if any secondary variable is present for that corresponding time point. In many spatial statistical studies, learning of primary variable by considering secondary variables into model showed improved results when secondary variable was abundant or easily available than primary variable. The target to improve prediction of degradation trend for solar panels is the prime reason behind considering weather factors into the proposed MTL-GP-TS algorithm with both primary and secondary variables. Dataset with both space and time dependency scenario can be dealt by kriging techniques like Simple kriging, Ordinary kriging and Dynamical Spatio-temporal models [46]. For our research study, we focused on PV-Weather data integration model to fit the temporal dependency under a spatial statistical framework. A schematic diagram of predicting target variable with predictors is shown in Fig. 5.1 where each variable represents time series observations for a particular solar panel.
One important consideration regarding this scenario is the under sampled situation or missing observations in time series dataset of primary variable PV. Limited availability of PV observations in time series can be considered as similar to under sampling scenario. It is also considered in our study that secondary variables are sufficiently available compared to the primary variable. The main challenge under such situation is how to incorporate both the under sampled primary variable and sufficiently available secondary variables as inputs for predicting response variable while maintaining the temporal dependency between observations. This can be written as,

\[ Y(t) = f(Y(t_1), X_1(t_1), \ldots, X_n(t_1)) + \epsilon \]  

(5.1.1)

Here, \( Y(t) \) is the response of solar panel power output PV at any time \( t \) and \( X(t) \) is the secondary response variables. Function of \( f \) represents the temporal correlation between the primary variable
PV and the temporal cross correlation between PV and other \( n \) secondary variables introduced as weather factors. Here \( \epsilon \) is referred as error term. An assumption is made here that each predicted value reflects the temporal relation of the corresponding time point with all other time points of that particular time series.

In order to define Eq. 5.1.1, a number of spatiotemporal statistical models [46, 47] can be applied. Due to the complex structure of spatiotemporal model, we proposed studying the temporal relation between observations under the framework of spatial statistics models. Among spatial models, co-kriging [48] can be apt for predicting primary variabile PV while considering both primary variable (under sampled) and secondary variable for modeling. Kriging [49] method may perform average in this case as the primary variable is not sufficiently available to make predictions. Besides PV-Weather data integration model, generalized additive models [50] can also be used to predict primary variable PV where weather factors (secondary variables) can be defined by polynomial regression and the primary variable PV can be expressed as Gaussian process term with kriging method.

5.2 Possible Approaches for MTL-GP-TS Method with Weather Consideration

Among several approaches, one possible approach for PV-Weather data integration model can be incorporation of weather factors into the prediction of temporal dependent PV observations, \( X(t) \) by decomposing a model into global and local effect with co-kriging. The reason of selecting co-kriging method for including weather factors, is to improve prediction results by fusing information from both limited amount of primary variable and sufficiently available secondary variables. The idea is to learn the missing observations of primary variable \( Y(t) \) initially from other weather factors available in each solar panel time series dataset using co-kriging. The predictions of \( Y(t) \) can be illustrated as;

\[
Y(t) = \mu(t) + \nu(t) + \xi(t)
\]  

(5.2.1)

where the mean \( \mu(t) \) can be modeled as a deterministic function of the correlated primary and secondary variables, covariance \( \nu(t) \) is defined as temporal process of the residuals from temporal dependent variables with mean zero and variance \( \omega^2 \psi(\lambda) \), where \( \omega^2 \) is marginal variance. The \( \psi(\lambda) \) can be described as the correlation function of exponential form, \( exp(-\lambda||t_i - t_j||) \) where \( \lambda \) is a
scalar parameter and \( ||t_i - t_j|| \) represents the euclidean distance between two distinct time points \( t_i \) and \( t_j \) respectively. The error \( \xi(t) \) is normally distributed with mean zero and variance \( \sigma^2 \).

The assumptions made for this model are briefly stated below,

- The global variation is related to seasonal variation among variables over time and the local variation is related to measurement errors while collecting data.
- The primary and secondary variables are assumed positively correlated as the change in each variable is affected by changes in weather factors.
- When both primary and secondary variable are available at a certain time point, they are assumed to be co-located.

Another approach of incorporating weather factors into the MTL-GP-TS study can be proposed by a parametric representation. Here, the parameters will capture the weights or cross correlation between variables via co-kriging method. The prediction equation for PV observations \( Y(t) \) at any time \( t \) can be illustrated as;

\[
Y(t) = \sum_{a=1}^{T} \sum_{b=1}^{K} \lambda_{ab} X(a)
\]  

(5.2.2)

where \( \lambda \) represents weights to minimize variance estimation error of \( K \) variables. \( X \) represents the predictor variables at any time point \( a \), where \( a = 1, 2, \ldots, T - 1, T \) and \( t \in N(T) \).

In this thesis, we followed the second approach to incorporate weather factors into MTL-GP-TS for monitoring significant performance degradation in solar panel time series. The response variable \( Y \) is predicted for each location via co-kriging. The predicted \( \hat{Y} \) values by co-kriging for a particular location is taken as inputs for MTL-GP-TS to learn the degradation trend \( \hat{Y}_{mtl} \) by fusing information from all co-kriged task locations. The procedure of MTL-GP-TS with PV-Weather data integration model is briefly summarized in Fig. 5.2.

### 5.3 Case Study

MTL-GP-TS with weather consideration is demonstrated with a case study based on solar panel dataset of solar panel power output (PV, unit: kilowatts) with three other variables considered as weather factors. The weather factors are plane of array solar irradiance (POA, unit: kilowatts per meter square), wind speed (WS, unit: meters per second) and humidity (HD, unit: percentage)
\[ Y_0 = \text{Primary response variable at location } g \] (under sampled)

\[ X_i = \text{Secondary variables at location } g \]
\[ i = 1, 2, \ldots, n \text{ and } g = 1, 2, \ldots, m \]

\[ \hat{Y} = \text{Estimated primary variable at location } g \]

\[ \hat{Y}_{mtl} = \text{Estimated trend at location } g \]
while sharing information between all task locations

Figure 5.2: Diagram of MTL-GP-TS with weather factors consideration using PV-Weather data integration model.

from two schools in Hawaii (data source available at \(^1\)). The school locations are selected as two data sources based on the information that solar panels are manufactured by the same manufacturer (i.e., Ascension) and from same geographical location (i.e., Hawaii). While incorporating weather factors to estimate the degradation trend of solar panels, PV is considered as primary variable and the weather factors are considered as secondary variables. For each solar panel source, it is also considered that primary variable PV is not available at all time points or under sampled whereas the secondary variables are sufficiently available.

5.3.1 Description of solar panel dataset

The solar panel data were selected over a ten-year timespan based on the data collected from Jan’2001 to Dec’2010. The data sources are from Castle High school (C) and Kahuku Intermediate and High school (K). Among the secondary variables, only POA and WS show significant correlation.

\(^1\)http://www.hawaiianelectric.com/heco/_hidden_Hidden/EducationAndConsumer/Sun-Power-for-Schools?cpsextcurrchannel=1
with primary variable PV. HD was eliminated from the dataset for further study due to lack of positive correlation with primary variable PV. The results of “variable selection” method also shows that inclusion of only POA and WS is sufficient to predict the primary variable PV with lowest test error. Consideration of weather factors into the data set with secondary variables results in a seasonal time series dataset for each solar data source. As a reason, for MTL-GP-TS algorithm with weather consideration, a decomposed time series will have three components: trend, seasonal and irregular components.

The observations for each variable are collected at 12 p.m of the day. The interval between data collection is weekly. For each data sources, the dataset contains training, testing, and validation sets for the primary variable PV. For preparing the under sampled primary variable PV observations, random selection of PV training data and test data is done with cross validation. The secondary variables (POA and WS) contain only the training set. The aggregated dataset with both primary and secondary variables, has total 480 time points with weekly interval over 10 years starting from 2001 (considered as Year 1). The PV training and testing set contain 77 data points and 307 data points, respectively. Here, the PV test data points are considered as missing observations in time series. The PV validation set includes 96 data points with weekly time interval from Years 9 and 10 for each solar panel data source.

5.3.2 Solar panel performance prediction by MTL-GP-TS with weather consideration

The missing observations in PV primary variable are predicted by co-kriging with the assumption of having temporal dependency among each predicted value. The temporal correlation among primary variable (PV) ans cross correlation between primary variable (PV) and secondary variables (POA and WS) are measured by covariance. Predicted primary variable (PV) for each solar data source, are then taken as time series training inputs for learning degradation trend with MTL-GP-TS algorithm. Similar to proposed MTL-GP-TS algorithm, MTL-GP-TS with weather consideration approach also learns the missing observations by sharing information among multiple time series if the datasets are from a similar source. For this particular case study, the decomposition of time series generated an additional component considered as seasonal component along with trend and irregular component. Following the previous case study described in Chapter 4, EM algorithm reached convergence within 5000 maximum iterations starting with the initial values
of hyper parameters, also for this case study with weather considerations.

The proposed MTL-GP-TS with weather consideration was compared with traditional time series modeling based on one single data source (i.e., STL with weather consideration). The degradation trend were estimated using seasonal ARIMA models once the missing observations are learned by the two methods (MTL-GP-TS and STL). The degradation trend is then forecasted for next two years (Years 9-10) for the solar panels of interest (target solar panels - C and K school). The residuals from forecasted values follow a normal distribution indicating the choice of ARIMA models is appropriate for the target time series. The autocorrelation (ACF) plot for the respective solar data sources further ensures that there is no significant correlation between the residuals, emphasizing the fact that the selected ARIMA models yield the best forecast predictions with the lowest RMSEs. The improvement in forecasting future trend values of solar panels were also measured by comparing RMSE-forecast values of both MTL-GP-TS and STL method when weather factors are considered.

### 5.3.3 A summary of results

<table>
<thead>
<tr>
<th>Solar panel data source</th>
<th>RMSE-STL</th>
<th>RMSE-MTL-GP-TS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Castle</td>
<td>0.23</td>
<td><strong>0.09</strong></td>
</tr>
<tr>
<td>Kahuku</td>
<td>0.5</td>
<td><strong>0.18</strong></td>
</tr>
</tbody>
</table>

The RMSE-forecast values presented in Table: 5.1 for two target solar panels data sources (C and K), compares the outcome from both STL and MTL-GP-TS method with weather consideration. For all target solar panels, MTL-GP-TS demonstrated substantial improvement in reducing RMSE of forecasted values as shown in Fig. 5.3. The RMSE-forecast improvements by MTL-GP-TS with weather consideratio are 60.87 % for solar panels in C school and 64% for solar panels in K school as shown in Fig. 5.4 while comparing with STL. The significant improvement in RMSE-forecast by MTL-GP-TS with weather consideration indicates that MTL-GP-TS can outperform traditional STL method in forecasting degradation trend when time series dataset contains limited
5.4 Discussion

5.4.1 Performance monitoring of solar panels with weather consideration

This section focuses mainly on the performance monitoring of above mentioned school C and K when additional information of weather factors included into MTL-GP-TS algorithm. After incorporation of weather factors for predicting PV via co-kriging, both MTL-GP-TS and STL method were performed to forecast future trend estimates for degradation monitoring of solar panels. Improved monitoring of significant degradation by MTL-GP-TS with weather consideration can alert the owner about any upcoming breakdown or potential failure occurrences. Preventive maintenance actions can be undertaken in such circumstances if the estimated degradation trend falls below a
predetermined threshold value. The threshold value for significant performance degradation monitoring is set by converting the percentage degradation of a solar panel over its life cycle under ideal operation conditions, to the output value of solar panel after several years of usage. While considering inclusion of weather factors in performance monitoring of PV, the threshold value of PV is assumed to be 1.1 to monitor significant degradation in two target schools (C and K). For weather consideration in time series for learning degradation trend, MTL-GP-TS can effectively monitor significant degradation in both target school whereas STL fails to monitor in one of the schools (school K).

Fig. 5.5 shows the degradation trend forecast of PV values for K school by the proposed MTL-GP-TS and STL methods with weather consideration for forecasted period during Years 9 and 10. The degradation trend was estimated in the forecasted period by both MTL-GP-TS and STL with 90% prediction interval. If the lower bound of the prediction intervals falls below the predetermined
threshold value, the estimated trend may indicate significant degradation in performance. In this section, we checked for significant degradation on the 3rd week of February in Year 9 during the forecasted time period. The findings are briefly discussed as follows, i.e.,

- Only the lower bound of MTL-GP-TS fell below the threshold value in K school solar panels on 3rd week of February in Year 9, indicating of potential failure occurrence at that time or near future.

- The STL trend showed recurrent fluctuation in the forecasted PV trend values losing the prospect of monitoring capability.

- The bandwidth of the prediction intervals for MTL-GP-TS was relatively narrow and con-
stant over time compared with STL prediction intervals that expand widely, making STL less effective for future forecasts. The STL trend inputs contain significant peaks in fitted value resulting in wider prediction interval bounds whereas MTL-GP-TS results in narrow prediction interval bounds due to less significant peak in fitted value.

- The true PV trend values of K school from the original dataset was closely aligned with the point estimates of MTL-GP-TS trend and also within the prediction interval. By contrast, the point estimates of the STL trend were far above the true trend with a wide prediction interval.

5.4.2 Improvement of MTL-GP-TS with weather consideration

![Figure 5.6: Comparison of RMSE-forecast of MTL-GP-TS and STL on dataset for C and K solar panel data sources without weather factors consideration (non seasonal time series).](image)

In order to measure the RMSE-forecast improvement of degradation trend using seasonal and non seasonal dataset, the case study of Castle High school and Kahuku Intermediate and High school is also studied by the proposed approach in Chapter 4 that ignores weather consideration.
For preparing a non-seasonal dataset, the ratio of $PV/POA$ is taken as inputs for MTL-GP-TS algorithm. The ratio of $PV/POA$ eliminates the effect of seasonal component from the time series dataset. The RMSE-forecast results for non seasonal dataset is shown in Fig. 5.6. On the other hand, the seasonal dataset is prepared while incorporating weather factors as described in 5.2.1 section. The comparison between RMSE-forecast improvement of degradation trend estimation while considering and ignoring weather factors respectively is shown in Fig.5.7.

![RMSE-forecast of solar panel data](image)

**Figure 5.7:** Comparison of %RMSE-forecast improvement of degradation trend while considering (seasonal time series) and ignoring weather factors (non seasonal time series) respectively for C and K solar panel data sources.

From Fig. 5.7, it can be seen that degradation trend estimation by MTL-GP-TS with weather consideration (seasonal time series) outperforms the MTL-GP-TS without weather consideration (non seasonal time series) for the same solar panel data sources (C and K school). The %RMSE-forecast improvement by MTL-GP-TS without weather consideration is 7.5% and 7.69% for school C and K, respectively. The percentage RMSE-forecast improvement by MTL-GP-TS with weather
consideration is remarkably larger than MTL-GP-TS without weather consideration by a higher percentage indicating that the inclusion of weather factors (POA and WS) into learning of degradation trend for C and K school solar panels, yields better result than when ignoring weather factors in MTL-GP-TS algorithm.
CHAPTER 6

CONCLUSION

This thesis develops a novel method based on multi-task learning for improving prediction of degradation in spatial time series data (MTL-GP-TS) from solar panels while overcoming the challenge of time series prediction based on limited data observations, by sharing common information from other similar-but-not-identical time series. In addition, the proposed approach generalizes MTL for Gaussian spatial processes to irregular time series with missing observations. We can conclude from the findings of the case study that MTL-GP-TS model can efficiently learn the unobserved values in time series dataset, resulting in improved future forecasts for the selected solar panel data sources. The estimated future degradation trend from MTL-GP-TS has the potential to improve monitoring of significant performance degradation for warning about potential failure occurrences in future. In addition, we find that the performance of the proposed MTL-GP-TS depends on the selection of solar panel data sources that could be necessary to ensure the prediction improvement. As RMSE improvement in trend monitoring by MTL-GP-TS is closely dependent on the selection of data sources, it always may not be attainable for each type of data source. Lastly, it is also observed that the proposed MTL-GP-TS method can significantly improve prediction of degradation trend when considering weather factors for initial estimation of missing observations of time series rather than performing interpolation.

The major advantage brought by the proposed method is the improved prediction of solar panel degradation and potential occurrence of defects which do not frequently occur without increasing measurement cost for data collection. The outcome of the proposed method will help establish preventive maintenance that is anticipated to outperform the current strategy based on fixed inspection interval to perform necessary repairs and inspections. Moreover, the methodology is not limited to the prediction of degradation trend of solar panel in time series. The method can be applicable to health condition monitoring for a wide range of applications such as construction structures and mechanical/electronic devices.
CHAPTER 7
FUTURE WORK

The proposed MTL-GP-TS method can be extended to time series dataset of different types of devices other than solar panels where performance monitoring can be improved by fusing information from multiple similar sources. For example, the effectiveness of the MTL-GP-TS method can be observed for the monitoring of nanowire growth [51] (nanowire length and diameter) over the time. Fig. 7.1 (A) shows an electron microscopic image of nanowire - Zinc Oxide nanowire grown in Silicone (100) substrate [52] and Fig. 7.1 (B) shows Au catalyzed silicone nanowire growth by VLS (Vapor-liquid-solid) mechanism (Image source available at 1) [53]. Specifically, data collection from each individual nanomanufacturing process could be time-consuming and insufficient. Fusion of the data from multiple similar nano processes provides an opportunity of improving the prediction of spatiotemporal evolution in nanostructures.

Figure 7.1: (A) Growth of nanowire: ZnO nanowires grown on Si(100) substrate. (B) Au catalyzed silicone nanowire growth (VLS mechanism).

1http://pan-group.engr.uga.edu/Research-Catalyst.html
Another future research can be consideration of both qualitative and quantitative variables in MTL-GP-TS model to monitor the performance. In this thesis, while only quantitative variables are considered to estimate the engineering process of solar panels, qualitative variables may also be considered in future to evaluate the performance of MTL-GP-TS algorithm for degradation monitoring. Generalized linear model or logistic regression can be possible choices for incorporating both quantitative and qualitative variable into MTL-GP-TS.
BIBLIOGRAPHY


[43] Anastasia Pentina, Viktoriia Sharmanska, Christoph H Lampert, and Am Campus. Curriculum Learning of Multiple Tasks.


BIOGRAPHICAL SKETCH

Tahasin Shireen was born in Bangladesh. She earned her Bachelor of Science degree in Industrial and Production Engineering from the Department of Industrial and Production Engineering at Bangladesh University of Engineering and Technology, Bangladesh in 2013. Under the Supervision of her advisor Dr. Hui Wang, Tahasin Shireen will receive her Master of Science degree in Industrial Engineering with honor, from Florida State University.