

Continuous Monitoring of the Ambient Factors via ε -Smooth Support Vector Regression [†]

Yan-Ru Jhuo ¹, Chi-Yu Chen ¹, Yu-Hsuan Yang ², Hsing-Chuan Hsieh ^{3,*} and Yuh-Jye Lee ^{1,4}

¹ Department of Applied Mathematics, National Chiao Tung University, Hsinchu 300, Taiwan; eve8362000@gmail.com (Y.-R.J.); chris85712@gmail.com (C.-Y.C.); yuhjye@math.nctu.edu.tw (Y.-J.L.)

² Institute of Data Science and Engineering, National Chiao Tung University, Hsinchu 300, Taiwan; sb0953330882@gmail.com

³ Institute of Statistics, National Chiao Tung University, Hsinchu 300, Taiwan

⁴ Research Center for Information Technology Innovation, Academia Sinica, Taipei 11529, Taiwan

* Correspondence: connie0915549431@hotmail.com; Tel.: +886-917-877-145

[†] Presented at the 13th International Conference on Ubiquitous Computing and Ambient Intelligence UCAmI 2019, Toledo, Spain, 2–5 December 2019.

Published: 21 November 2019



Abstract: Thanks to the advances of the Internet of Things (IoTs), more and more wireless sensor networks applications have been realized. One of the fundamental but crucial applications is the continuous monitoring of environmental factors including temperature, humidity, illumination, etc. We develop a nonlinear regression model which takes spatial and temporal information into account to construct a globally three-dimensional heat map for a closed space based on very sparse sensor deployment. However, fitting the whole-space heat map with a very limited number of sensor observations gives a very poor estimation when we use a nonlinear model. We call it the coverage hole problem. We utilize the uniform experimental design which is well known in industrial statistics to allocate the synthetic sensors. We estimate those synthetic sensor readings on the basis of linear model locally. We then apply ε -SSVR, a nonlinear support vector regression model to fit the globally three-dimensional heat map by combining real sensor and synthetic sensor readings. The numerical results demonstrate our proposed model can enhance the accuracy significantly.

Keywords: smooth support vector regression; coverage hole problem; inverse distance weighting; uniform design

1. Introduction

Thanks to the advances of the Internet of Things (IoTs), wireless sensing networks (WSNs) are widely used in many fields such as medical, military, industry, agriculture, intelligent transportation, etc. [1–5]. Specifically, the application of WSN continuous monitoring of environmental factors (such as temperature, humidity, etc.) for the entire region in the domain of greenhouse cultivation has been gradually noticed [6,7]. The dynamic responses of equipment (e.g., irrigation system, ventilation system, etc.) to the monitoring results enable farmers to grow high-quality and high-yielding crops.

We would like to develop a nonlinear regression model taking spatial and temporal information into account to construct a globally three-dimensional (3-D) heat map (e.g., temperature in our experiment) for a closed space based on sensor deployment. In reality, the larger cost of sensor deployment has limited the amount of sensors being able to be installed in the greenhouse, resulting in coverage hole problems [8,9]. What is worse, fitting the whole-space heat map with very limited number of sensor observations will give a very poor estimation when we use a nonlinear model. In order to solve this problem, the most widely used method is to insert a value at a position where

there is no sensor (i.e., to found synthetic sensors). The traditional method includes Inverse Distance Weighting (IDW) and Ordinary Kriging (OK) [10,11]. However, the interpolated values of the above methods will fall between the minimum and maximum reading values, which might cause a great monitoring error especially in a variable environment.

As a result, here we introduce the Uniform Experimental Design (UD) method [12] to decide the positions of synthetic sensors in the greenhouse space. For interpolating the virtual sensing points, we develop a method based on IDW and OK to deal with the stratospheric characteristics of environmental factors in greenhouse. We then apply the data from the real as well as the synthetic sensors to a nonlinear ε -Smooth Support Vector Regression model (ε -SSVR or SSVR for short) [13–15], in order to construct a globally 3-D heat map of each environmental factor for the greenhouse. The ultimate goal is to control the related equipment according to monitoring results, as shown in Figure 1.

The main advantage of ε -SSVR model is the capability to adjust the objective ε -insensitive loss function based on real and synthetic sensors, where ε -insensitive loss function indicates that only the absolute error between the estimated and the observed values of an instance larger than ε value will be accounted for. Since the readings of synthetic sensors are interpolated, which may reflect higher measurement errors than those of real sensors, we give real sensors with smaller ε , while synthetic sensors with larger one.

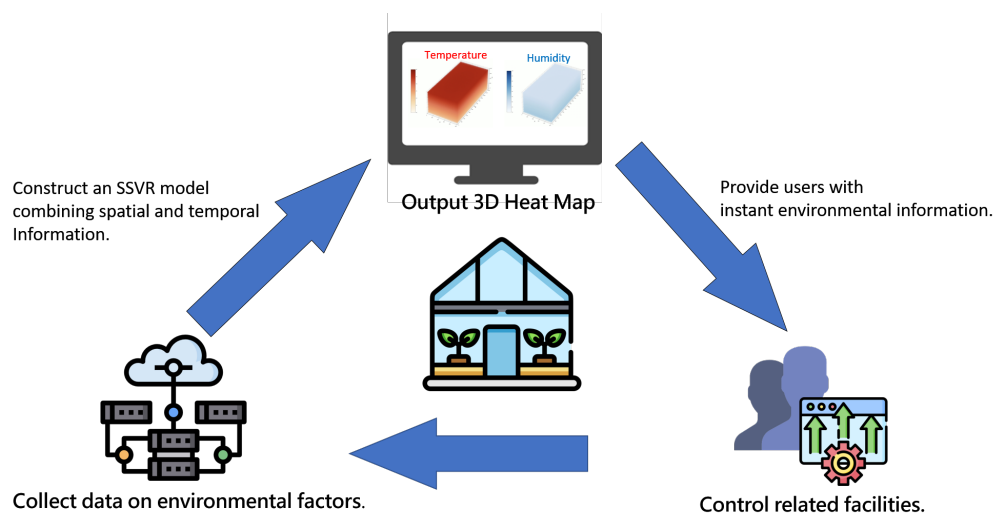


Figure 1. Continuous Monitoring and Decision Process.

2. Related Work

The related research topics for our experiment are divided into three categories: (1) coverage hole problem, (2) interpolation method for globally 3-D heat map, and (3) monitoring model.

The coverage hole problem is common for sensor deployment and is discussed in many studies of WSNs. To deal with such problem, we have to first define a coverage model for the sensor. Boudaren et al. [16] mentioned three kinds of coverage models. Deterministic coverage models classify a point to be either certainly within a specific region (i.e., covered by a sensor) or not. Probabilistic models represent the probability of a point being covered. The longer the distance between the point and the sensor, the less likely it is for the point to be covered. At last, Evidential sensing models are designed to handle the uncertainties by taking into account the reliability of the sensors. Note that the coverage models do not have to be in a circular shape. Some coverage hole problems are due to abnormalities in the values recovered by the sensors.

Since it is impossible to arrange full sensing points in space, and the sensors in most cases are sparse, the interpolation method becomes very important for those positions without sensors. The study of Li and Heap [17] divides the interpolation methods into three categories and briefly describes more than 40 spatial interpolation methods. They also discuss the factors that affect the

performance [18]; there are a total of 25 classifications of common methods according to characteristics of the interpolation models. Finally, a decision tree that selects the appropriate method from these 25 methods is developed. The study of Bargaoui and Chebbi [19] uses a cross-validation scheme to compare the accuracy of two kriging methods (2-D, 3-D). The 3-D kriging method derived from the rainfall data leads to a significantly lower prediction error than the classic 2-D kriging method. In the study of Zhang et al. [20], rainfall data from the agro-ecological monitoring network was assessed, and four spatial interpolation methods were used to generate a map of the total monthly rainfall in Sri Lanka.

In regards to monitoring models, the study of Danielsen et al. [21] integrates various monitoring methods to monitor natural resources for the habitats of the species in order to assist conservation biologists in identifying and establishing monitoring systems that are most relevant to specific environmental protection. Hu et al. [22] use a support vector regression model to monitor air pollution and obtain results that are more accurate than those from artificial neural network models. Obviously, monitoring is widely used for different purposes. The research of Shen et al. [23] is related to our goal, which monitors the 2-D plane of greenhouse and has proposed a distributed anomaly detection mechanism to report the abnormal information. Based on the data from the greenhouse of the Martens Greenhouse Research Foundation in western Finland [7], it is possible to detect local differences caused by various disturbances in greenhouse climate. In our research, we are going to combine either spatial and/or temporal features to build a monitoring model and output continuous 3-D heat maps.

3. Coverage Hole Problem

3.1. Virtual Sensors via Uniform Design Sampling

To solve the coverage hole problem (Figure 2), here we introduce a method called “uniform design (UD)” which is popular in the industrial experimental design [24]. In ANOVA-based experiment design model, there are s factors with q levels for each, resulting in a total of q^s experimental combinations for testing the main and interaction effects among all factors. Obviously, the number of experiments increases exponentially with the number of factors as well as levels. Therefore, the objective is to utilize a few experiments instead of all while maintaining a good parameter estimation.

UD can effectively reduce the number of experiments (say, from q^s to n) with better design flexibility and uniformity, which reduces time consumption effectively in industrial experiments and has a good estimate of nonlinear models [12]. The n experimental combination is symbolized as $U_n(q^s)$ in the uniform design table referred to [25].

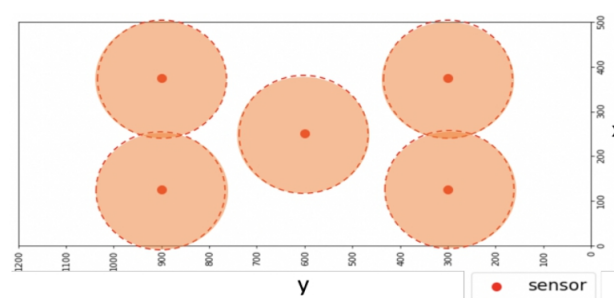


Figure 2. Coverage Hole Problem. The orange circle is the sensing range of the sensor. It shows that the range of each sensor is limited, so it cannot cover the entire plane.

In our research, we use UD in the following two applications: (1) positioning of the synthetic sensors and (2) parameter tuning for our monitoring model (i.e., ϵ -SSVR) instead of the grid search.

In application of positioning the synthetic sensors in 2-D space, the factors represent the length and width of the plane (i.e., $s = 2$), and each factor has $q = 30$ levels (hence we can position $n = 30$

synthetic sensors on the plane). We then can refer to UD $U_{30}30^2$ for the coordinate of each synthetic sensor. An illustration of all synthetic sensors is shown in Figure 3.

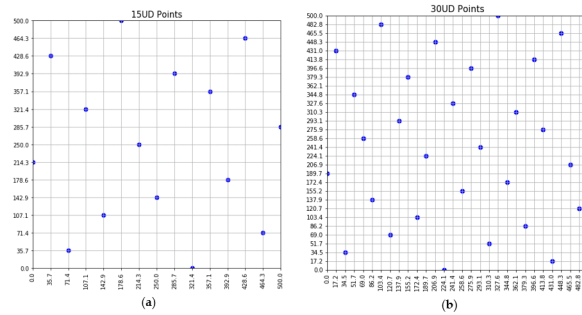


Figure 3. Uniform design showing the positions of synthetic sensors with: (a) $n = 15$; (b) $n = 30$ points.

3.2. Interpolation Method

After positioning the synthetic sensors, the next important task is to estimate the readings of each sensor at any time point, which is an interpolation problem. The commonly used local linear models for interpolation are IDW and OK, which can be referred to [10,11,26,27]. By “local” it means only the nearest k neighboring sensor readings are used for calculation. In our experiment, we consistently set $k = 3$ for all interpolation.

In addition, we notice empirically that the readings (e.g., temperature in our case) in the vertical direction change more drastically than those in the horizontal direction. More specifically, the temperature reading becomes higher along with greater height; while temperature readings from the same horizontal plane have subtle changes. This causes issues when using ordinary OK and IDW in our research domain, such as over-estimation for synthetic sensor values on smaller height (z -axis) and under-estimation on larger height, showing the unreasonable influence of z -axis. We thus develop a two-step new interpolation method, that is, the variants of IDW and OK (called zIDW and zOK).

Operationally, we treat the information from xy -plane differently from z -axis. More specifically, given any synthetic sensor point at position (x_0, y_0, z_0) , we first project it vertically to all non-redundant planes where real sensors locate (i.e., $L_r = \{(x_0, y_0, z_r) | r \in \{1, \dots, m_0\}\}$, $m_0 \geq 2$, which means the number of non-redundant planes). We then locally interpolate each projected location in L_r either by IDW or OK based solely on their horizontal real sensors and hence derive a vertical line of interpolated values along (x_0, y_0) . At last, we can apply these m_0 interpolations to again interpolate the target synthetic sensor at (x_0, y_0, z_0) linearly or nonlinearly (according to the scatter of interpolations of L_r). Notice that if $m_0 = 2$, only linear interpolation model can be used as the case in our experiment.

4. Monitoring Model via ε -SSVR

4.1. ε -Smooth Support Vector Regression Model

Let $S = \{(x_i, y_i) | x_i \in R^n, y_i \in R\}_{i=1}^m$ be a training dataset. In traditional ordinary regression, the objective is to find a hyperplane $f(x)$ with the least-squared method so as to best fit the dataset S . SVR can be regarded as a generalization from ordinary regression, which adds a tolerance parameter ε to discard small errors when fitting the training points. As shown in Figure 4, the loss function of each instance $i \in 1, \dots, m$ given ε parameter can be written as:

$$|\xi_i|_\varepsilon = \max(0, |f(x_i) - y_i| - \varepsilon) \quad (1)$$

That is, if the absolute discrepancy between predicted value and the observed value is less than a given ε value (i.e., $|f(x_i) - y_i| < \varepsilon$), then the loss value of such instance is regarded as 0.

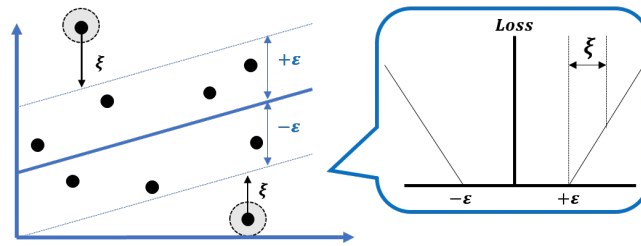


Figure 4. ε -insensitive loss function.

The objective of linear SVR is to find a hyperplane $f(x) = w^T x + b$ of which the constrained minimization problem is generally formulated as follows:

$$\begin{aligned} \min_{(w,b,\xi,\xi^*) \in R^{n+1+2m}} & \frac{1}{2} \|w\|_2^2 + C \sum_{i=1}^m (\xi_i + \xi_i^*) \\ \text{subject to} & \begin{cases} y_i - (w^T x_i + b) \leq \varepsilon + \xi_i^* \\ -y_i + (w^T x_i + b) \leq \varepsilon + \xi_i \\ \xi_i, \xi_i^* \geq 0 \end{cases} \quad i = 1, 2, \dots, m \end{aligned} \quad (2)$$

where $\|w\|_2^2$ is a regularization term applied for improving the generalization ability, C is a positive constant that represents the tradeoff between the flatness of the model and fitting errors of the function $f(x)$, η is called a slack variable, and ε is the insensitive zone.

In linear ε -SSVR [13,14], the above Equation (2) can be reformulated into an unconstrained minimization problem is as the following [13]:

$$\min_{(w,b) \in R^{n+1}} \frac{1}{2} (\|w\|_2^2 + b^2) + \frac{C}{2} \sum_{i=1}^m |w^T x_i + b - y_i|_\varepsilon^2 \quad (3)$$

Furthermore, to make the objective function twice differentiable, which can be quickly solved by Newton's method, Lee et al. [13] used the smooth p -function (i.e., p_ε^2) tuned by a smoothing parameter β to replace the error term ($|x_i^T w + b - y_i|_\varepsilon^2$) of SVR (i.e., β controls how much the error term be approximated by p -function). Thus, the minimization problem is then reformulated into the following (refer to [13] for more detail):

$$\min_{(w,b) \in R^{n+1}} \frac{1}{2} (\|w\|_2^2 + b^2) + \frac{C}{2} \sum_{i=1}^m p_\varepsilon^2(w^T x_i + b - y_i, \beta) \quad (4)$$

Notice that synthetic sensors have a larger tolerance values ε_v than that of real sensors ε_r in our experiment, which shows:

$$\min_{(w,b) \in R^{n+1}} \frac{1}{2} (\|w\|_2^2 + b^2) + \frac{C}{2} \left(\sum_{i \in I_v} p_{\varepsilon_v}^2(w^T x_i + b - y_i, \beta) + \sum_{i \in I_r} p_{\varepsilon_r}^2(w^T x_i + b - y_i, \beta) \right) \quad (5)$$

where I_v and I_r are the index sets of synthetic and real sensors, respectively, s. t. $|I_v| + |I_r| = m$.

4.2. Nonlinear ε -Smooth Support Vector Regression

We can extend linear case to nonlinear case through the dual formulation of minimization problem [13]. In order to achieve this, we employ a kernel function $K(x, z) = \langle \phi(x), \phi(z) \rangle$ which represents the dot product of two training points in a higher dimensional space. In general, there are many forms of kernels, and the Gaussian kernel (also known as radical basis function (RBF) kernel) is one of the commonly used as defined in Equation (6), which is also used in our experiment:

$$K(x_i, x_j) = e^{-\gamma \|x_i - x_j\|^2} \quad (6)$$

where γ is a tuning parameter determining how far the influence of squared Euclidean distances between the two feature vectors reaches.

4.3. Spatial and Temporal Features

Usually, the sensor information includes ID, location, measurement (i.e., temperature in our case), and timestamp. Thus, to predict the temperature at any unmeasured location at timestamp t (Y^t), we incorporate spatial (S) as well as temporal (T) features, and introduce four ϵ -SSVR related models.

For spatial model (SSVR (S)), features include 3-D sensor location (i.e., $\mathbf{x}_i = (Loc.X, Loc.Y, Loc.Z)$ for each sensor case i), of which the training data is illustrated in Table 1 with, say, $m = 10$ sensors at timestamp t . For temporal model (SSVR (T)), features include the past h values of temperature (i.e., $\mathbf{x}_i = (y_i^{t-1}, y_i^{t-2}, \dots, y_i^{t-h})$), of which the training data is illustrated in Table 2 with $h = 3$ (which is also the value we set for temporal model in our experiment). In the same manner, the spatial and temporal model (SSVR (S&T)) includes aforementioned features (i.e., $\mathbf{x}_i = (Loc.X, Loc.Y, Loc.Z, Y_i^{t-1}, Y_i^{t-2}, \dots, Y_i^{t-h})$, of which the training data can be formed by concatenating Tables 1 and 2. Notice that if we incorporate synthetic sensors for model training, it follows the exact rationale. The only difference is that the training data size is expanded by containing cases of both real and synthetic sensors (e.g., $m = 40$ with 10 real sensors and 30 synthetic sensors).

Table 1. The form of training data for spatial model at timestamp t .

Sensor ID	Loc.X	Loc.Y	Loc.Z	Temperature (hY^t)
1	375	300	150	28.29
2	125	300	150	29.93
\vdots	\vdots	\vdots	\vdots	\vdots
9	375	900	350	34.03
10	125	900	350	34.16

Table 2. The form of training data for temporal model at timestamp t .

Sensor ID	Y^{t-3}	Y^{t-2}	Y^{t-1}	Temperature (Y^t)
1	27.77	27.66	28.28	28.29
2	28.12	28.38	29.31	29.93
\vdots	\vdots	\vdots	\vdots	\vdots
9	33.79	34.45	33.94	34.03
10	34.70	34.78	34.29	34.16

Beyond these models, we also combine SSVR (S) and SSVR (T) models to strengthen our prediction (named SSVR (S + T) for convenience); this is to deal with the scenario that, when the temperature has a sharp rise or falls along the time, the temporal model will be overestimated or underestimated. We then design the model as below:

$$s_0^{S+T}(t) = \begin{cases} \alpha s_0^S(t) + (1 - \alpha) s_0^T(t), & |s_0^S(t) - s_0^T(t)| \leq \eta \\ s_0^S(t), & |s_0^S(t) - s_0^T(t)| > \eta \end{cases} \quad (7)$$

where $s_0^S(t)$ and $s_0^T(t)$ represent the temperatures estimated by the spatial and temporal models, respectively, at timestamp t ; α is a weighting parameter $\in [0, 1]$; η is to limit the difference of temperatures from the spatial and temporal models. That is to say, if the difference is larger than η , then we only consider the estimated values from spatial model; otherwise, we derive a weighted average. $s_0^{S+T}(t)$ is thus the estimated temperature from a combination of the spatial and temporal models.

4.4. Architecture

The continuous monitoring architecture to build a globally 3-D heat map at any timestamp t is shown in Figure 5. There are two steps. The first step is to reduce the coverage hole problem due to sparsity of the sensor deployment. We thus apply UD to generate multiple synthetic sensors deployed in 3-D space, of which the readings are interpolated by local linear model (e.g, IDW, OK, ...). The second step is then to apply the ε -SSVR models (as introduced previously) to fit the globally monitoring model by combining real and synthetic sensor readings. At last, the output is a consecutive 3-D heat map for any timestamp t .

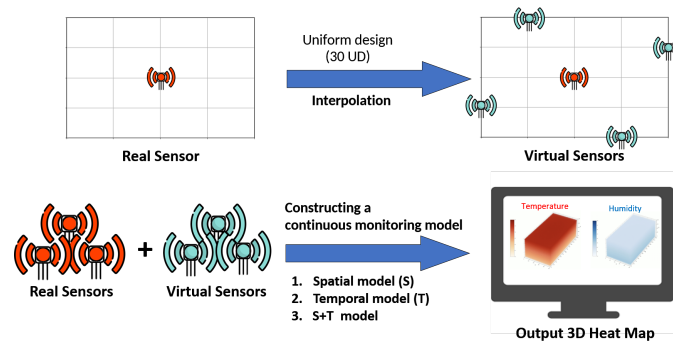


Figure 5. Continuous Monitoring Architecture.

5. Experiments

5.1. Data Description

Our smart greenhouse field is in Wufeng District, Taichung—a rectangular parallelepiped with (length (X), width (Y), height (Z)) = (500, 1200, 400) (scale: cm), which could monitor the growth conditions of cherry tomatoes. In smart greenhouse, there are 2 horizontal layers (i.e., Loc.Z = 150 and 350 cm), each having 5 E-type thermometers deployed (red points in Figure 6). The frequency of each sensor data is one reading per ten minutes. The sensors are coded from 1 to 5 on Loc.Z = 150, and 6 to 10 on Loc.Z = 350. For either layer of Z, the corresponding (Loc.X, Loc.Y) = { (375, 300), (125, 300), (250, 600), (375, 900), (125, 900)}.

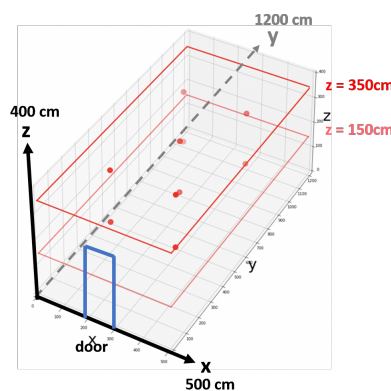


Figure 6. Smart Green House Architecture.

5.2. Experimental Process and Parameter Tuning

We evaluate our nonlinearly extended models—SSVR(S), SSVR(T), SSVR(S&T), and SSVR(S+T) under the condition either with or without synthetic sensors, to compare the marginal effect of synthetic sensors; we also compare our method with OK and IDW as baseline models. To validate our methods, the leave-one-out (LOO) strategy is used to each of real sensors where we can calculate our error rate

by comparing the actual value and the predicted value. More specifically, for i -th real sensor being tested ($i = 1, \dots, 10$) at time t , the training data would be composed of the rest of sensor information (either with or without synthetic sensors) at time t . We can thus estimate the values of i -th real sensor for any time t ($t = 1, \dots, T$). As a result, for each trial of real sensor, the evaluation metrics can be calculated by Mean Absolute Error (MAE) and Root Mean Square Error (RMSE):

$$MAE_i = \frac{1}{T} \sum_{t=1}^T |y_i^t - \hat{y}_i^t| \quad (8)$$

$$RMSE_i = \sqrt{\frac{1}{T} \sum_{t=1}^T (y_i^t - \hat{y}_i^t)^2} \quad (9)$$

where y_i^t and \hat{y}_i^t represent the actual and predicted values of real sensor i at time t , respectively.

In regard to parameter tuning for each of our nonlinear models, we use the aforesaid LOO strategy on the whole sample, and the set of parameters with the lowest averaged MAE as well as averaged RMSE across all trials of real sensors is determined as the best solution (refer to Table 3). Except for SSVR (S&T) model, we incorporate 30 synthetic sensors by UD method (i.e., $(U_{30}30^3)$) during parameter tuning process. The reason for SSVR (S&T) model will be explained at the latter section. Notice that the error tolerance ε_v of synthetic sensors is larger than that of the real sensors (ε_r), thus the parameter values tuned for the former are mostly larger than that of the latter (i.e., $[0.1, 2]$ v.s. $[0.01, 0.2]$), and the results of mean and standard deviation (STD) of MAE and RMSE with the optimum parameter setting are displayed in the following section, which can be regarded as the best-case scenario for each model incorporating synthetic sensors.

Table 3. Parameter setting.

Parameter	Tuning Range	Spatial (S + T)		
		Spatial (S)	Temporal (T)	Spatial (S&T)
C	[50, 400]	387.93	134.48	267.24
γ	$[10^{-5}, 0.05]$	0.0207	0.000177	0.00121
ε_r	[0.01, 0.2]	0.0493	0.0427	0.0755
ε_v	[0.1, 2]	0.0782	0.2068	-
α	[0, 1]	0.52		-
η	[0, 1]	0.27		-

6. Results

Our experimental results are presented in two ways: (i) numerical results and (ii) output of 3-D heat maps. The results of such experiments can indicate whether there is an improvement in model accuracy and a solution to the coverage hole problem.

6.1. Numerical Results

Table 4 shows the results under the existence of coverage hole problem (i.e., without synthetic sensors), which are derived from traditional spatial models such as IDW and OK, as well as regression-based nonlinear ε -SSVR models introduced previously. It can be found that the accuracy of the ε -SSVR models is better than that of traditional ones. Among ε -SSVR models, SSVR (S + T) has the best performance, while SSVR (T) the worst.

Table 4. Results under the coverage hole setting (without synthetic sensors).

Method	MAE	MAE (STD)	RMSE	RMSE (STD)
IDW	0.5990	0.0791	1.0920	0.0740
OK	0.5943	0.0834	1.0874	0.0787
SSVR (S)	0.3458	0.1378	0.8163	0.1588
SSVR (T)	0.5892	0.0835	1.0827	0.0783
SSVR (S&T)	0.5733	0.1058	1.0664	0.0973
SSVR (S + T)	0.3095	0.1519	0.7656	0.1812

Next we show the results of adding synthetic sensors in Table 5. Notice that the SSVR(S&T) model is not included since it does not improve much from SSVR (S) and SSVR (T) models (thus there is no tuning parameter ε_v). It is found that no matter which interpolation method (zIDW or zOK) we used, the accuracy was improved by reducing MAE and RMSE by about 0.03 to 0.3, indicating a good marginal effect of incorporating synthetic sensors. Again, SSVR(S + T) has the best performance (with zIDW + SSVR (S + T) slightly better than zOK + SSVR (S + T)), while SSVR (T) the worst.

Table 5. Results with synthetic sensors.

Method	MAE	MAE (STD)	RMSE	RMSE (STD)
zIDW + SSVR (S)	0.1364	0.0373	0.5167	0.0760
zIDW + SSVR (T)	0.5512	0.1808	1.0337	0.1838
zIDW +SSVR (S+T)	0.1268	0.0309	0.4997	0.0633
zOK + SSVR (S)	0.1286	0.0333	0.5028	0.0666
zOK + SSVR (T)	0.5540	0.1817	1.0365	0.1839
zOK + SSVR (S+T)	0.1195	0.0258	0.4862	0.0524

In both conditions above, spatial information contributes more than temporal information, which uses only $h = 3$ past values as features. An interesting observation is that, after adding synthetic sensors, the STD of MAE and RMSE from SSVR(T) model increase significantly, whereas that from SSVR(S) decreases also significantly, although they all improve MAE and RMSE in average. That means the spatial interpolation methods diverge from the predictions of the temporal model (SSVR(T)) but narrow them down for the spatial model (SSVR(S)). Accordingly, it is suggested that an interpolation method using temporal clues be introduced for improving the temporal model.

6.2. Output of 3-D Heat Maps

We also examine the results by heat maps. Here we compare the heat maps from four models—SSVR (S), SSVR (S + T), zIDW + SSVR (S), and zIDW + SSVR (S + T) illustrated in Figure 7. The lighter the color is, the lower the temperature becomes; and it is obvious for models without synthetic sensors (i.e., without zIDW interpolation) that they have mis-predicted values at the eight corners of the greenhouse (like a darker football crammed in a light-colored container), while models with synthetic sensors have less of such issues. On the whole, adding synthetic sensors improves model performance, and zIDW + SSVR (S + T) model has more realistic results than other models.

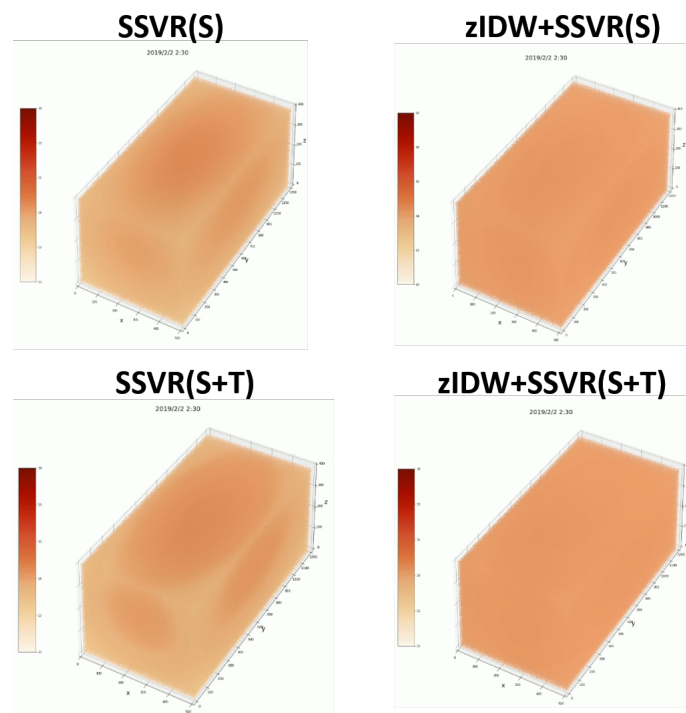


Figure 7. Heat maps comparison with four continuous monitoring models of temperature.

7. Conclusions and Future Work

In this thesis, we can see from the experimental results that using the ϵ -SSVR model to construct a global monitoring model has better performance than the traditional interpolation method IDW and OK. Moreover, the coverage hole problem can be solved by incorporating synthetic sensors through the following procedure: (i) adding virtual points in space by using uniform design, (ii) applying interpolation method (e.g., zIDW or zOK in our case) for synthetic data generation, (iii) combining real and synthetic sensor data (spatial and temporal) for ϵ -SSVR model buildup. The above effectively improve the estimation accuracy of continuous monitoring model.

It is worth noting that in our experiment, spatial clues contribute more to the model performance than temporal clues of only $h = 3$ past points. The possible suggestion is to add more past points (i.e., greater h) to the temporal model in future work. Furthermore, the interpolation method including temporal clues may also be considered to improve the quality of the data.

While the current model verification uses LOO strategy to each sensor, it is preferred, for the future experiment, to launch an automatic drone equipped with sensors to patrol the greenhouse field, from which the collected data (i.e., testing set) can be directed used for model validation. The drone can collect more thorough ambient data than those of the fixed-point sensors, thus providing more complete evaluation of the models.

In future applications, the 3-D monitoring model of ambient factors (e.g., temperature, humidity, etc.) can be integrated into the automated control system of the greenhouse; for heat management, as an instance, it is crucial when and where to open the windows or fans under current estimation of heat map. Furthermore, to improve the estimation of greenhouse environment, we can also combine the meteorological data from a local micro-weather station [28,29] such as wind speed, solar radiation [30], and so on. Other factors might include ambient conditions, for example, the state of ventilation equipment (windows or fans on or off), endothermic processes from crop transpiration, and soil evaporation.

In addition, greenhouse environmental factors are mostly modeled and monitored by the computational fluid dynamics (CFD) field. However, in the modeling process of CFD, the setting of environmental parameters is often an important task. As a result, we can also apply ϵ -SSVR global

monitoring model to help decide parameter values such as greenhouse transparency, air thermal conductivity and aerodynamic viscosity for the fluid mechanics.

Author Contributions: Conceptualization, Y.-J.L.; methodology, Y.-J.L., Y.-R.J., C.-Y.C., and Y.-H.Y.; software and formal analysis, Y.-R.J., C.-Y.C., and Y.-H.Y.; writing—original draft preparation, Y.-R.J.; writing—review and editing, H.-C.H.; supervision, Y.-J.L. and H.-C.H.

Funding: This research was funded by the Ministry of Science and Technology, via grant number MOST 107-2321-B-055-002.

Acknowledgments: We are very grateful to the Taiwan Agricultural Research Institute Council of Agriculture, Executive Yao for providing us with the experimental field in Taichung and helping considerably with setting up the sensor.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

ϵ -SSVR (or SSVR)	ϵ -Smooth Support Vector Regression
UD	Uniform Design
IDW	Inverse Distance Weighting
OK	Ordinary Kriging

References

- Chong, C.Y.; Kumar, S.P. Sensor networks: Evolution, opportunities, and challenges. *Proc. IEEE* **2003**, *91*, 1247–1256.
- Ramson, S.J.; Moni, D.J. Applications of wireless sensor networks—A survey. In Proceedings of the 2017 International Conference on Innovations in Electrical, Electronics, Instrumentation and Media Technology (ICEEIMT) Coimbatore, India, 3–4 February 2017; pp. 325–329, doi:10.1109/ICEEIMT.2017.8116858.
- Ko, J.; Lu, C.; Srivastava, M.B.; Stankovic, J.A.; Terzis, A.; Welsh, M. Wireless sensor networks for healthcare. *Proc. IEEE* **2010**, *8*, 1947–1960, doi:10.1109/JPROC.2010.2065210.
- Römer, K.; Mattern, F. The design space of wireless sensor networks. *IEEE Wirel. Commun.* **2004**, *11*, 54–61.
- Akyildiz, I.F.; Su, W.; Sankarasubramaniam, Y.; Cayirci, E. A survey on sensor networks. *IEEE Commun. Mag.* **2002**, *40*, 102–116.
- Abbasi, A.Z.; Islam, N.; Shaikh, Z.A. A review of wireless sensors and networks' applications in agriculture. *Comput. Stand. Interfaces* **2014**, *36*, 263–270, doi:10.1016/j.csi.2011.03.004.
- Ahonen, T.; Virrankoski, R.; Elmusrati, M. Greenhouse Monitoring with Wireless Sensor Network. In Proceedings of the 2008 IEEE/ASME International Conference on Mechatronic and Embedded Systems and Applications, Beijing, China, 12–15 October 2008; pp. 403–408.
- Ahmed, N.; Kanhere, S.S.; Jha, S. The holes problem in wireless sensor networks: A survey. *ACM SIGMOBILE Mob. Comput. Commun. Rev.* **2005**, *9*, 4–8.
- Wang, B.; Wang, W.; Srinivasan, V.; Chua, K.C. Information coverage for wireless sensor networks. *IEEE Commun. Lett.* **2005**, *9*, 967–969, doi:10.1109/LCOMM.2005.11002.
- Zimmerman, D.; Pavlik, C.; Ruggles, A.; Armstrong, M.P. An Experimental Comparison of Ordinary and Universal Kriging and Inverse Distance Weighting. *Math. Geol.* **1999**, *31*, 375–390.
- Caruso, C.; Quarta, F. Interpolation methods comparison. *Comput. Math. Appl.* **1998**, *35*, 109–126.
- Fang, K.T.; Lin, D.K.; Winker, P.; Zhang, Y. Uniform Design: Theory and Application. *Technometrics* **2000**, *42*, 237–248, doi:10.1080/00401706.2000.10486045.
- Lee, Y.J.; Hsieh, W.F.; Huang, C.M. ϵ -ssvr: A smooth support vector machine for ϵ -insensitive regression. *IEEE Trans. Knowl. Data Eng.* **2005**, *5*, 678–685, doi:10.1109/TKDE.2005.77.
- Gu, W.; Chen, W.P.; Ko, C.H.; Lee, Y.J.; Chen, J.S. Two smooth support vector machines for ϵ -insensitive regression. *Comput. Optim. Appl.* **2018**, *70*, 171–199.

15. Smola, A.J.; Schölkopf, B. A tutorial on support vector regression. *Stat. Comput.* **2004**, *14*, 199–222, doi:10.1023/B:STCO.0000035301.49549.88.
16. Boudaren, M.E.Y.; Senouci, M.R.; Senouci, M.A.; Mellouk, A. New trends in sensor coverage modeling and related techniques: A brief synthesis. In Proceedings of the 2014 International Conference on Smart Communications in Network Technologies (SaCoNeT), Vilanova i la Geltru, Spain, 18–20 June 2014; pp. 1–6.
17. Li, J.; Heap, A.D. *A Review of Spatial Interpolation Methods for Environmental Scientists*; Geoscience Australia: Canberra, Australia, 2008; pp. 137–145.
18. Li, J.; Heap, A.D. Spatial interpolation methods applied in the environmental sciences: A review. *Environ. Model. Softw.* **2014**, *53*, 173–189, doi:10.1016/J.ENVSOFT.2013.12.008.
19. Bargaoui, Z.K.; Chebbi, A. Comparison of two kriging interpolation methods applied to spatiotemporal rainfall. *J. Hydrol.* **2009**, *365*, 56–73, doi:10.1016/j.jhydrol.2008.11.025.
20. Zhang, R.P.; Jing, G.; Liang, T.G.; Feng, Q.S.; Aimaiti, Y. Comparing interpolation techniques for annual temperature mapping across Xinjiang region. *IOP Conf. Ser. Earth Environ. Sci.* **2016**, *46*, 1–6, doi:10.1088/1755-1315/46/1/012028.
21. Danielsen, F.; Burgess, N.D.; Balmford, A.; Donald, P.F.; Funder, M.; Jones, J.P.; Alviola, P.; Balet, D.S.; Blomley, T.; Brashares, J.; et al. Local participation in natural resource monitoring: A characterization of approaches. *Conserv. Biol.* **2009**, *23*, 31–42, doi:10.1111/j.1523-1739.2008.01063.x.
22. Hu, K.; Sivaraman, V.; Bhargubanda, H.; Kang, S.; Rahman, A. SVR based dense air pollution estimation model using static and wireless sensor network. *IEEE Sens.* **2016**, doi:10.1109/ICSENS.2016.7808827.
23. Shen, Y.C.; Chiang, A.; Yeh, Y.R.; Lee, Y.J. Continuous monitoring and distributed anomaly detection for ambient factors. In Proceedings of the 2014 IEEE International Conference on Internet of Things (iThings), and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom), Taipei, Taiwan, 1–3 September 2014; pp. 31–38, doi:10.1109/iThings.2014.14.
24. Huang, C.M.; Lee, Y.J.; Lin, D.K.; Huang, S.Y. Model selection for support vector machines via uniform design. *Comput. Stat. Data Anal.* **2007**, *52*, 335–346, doi:10.1016/J.CSDA.2007.02.013.
25. Hong Kong Baptist University. The Uniform Design. Available online: <http://www.math.hkbu.edu.hk/UniformDesign/> (accessed on 2 October 2019).
26. Oliver, M.A.; Webster, R. A tutorial guide to geostatistics: Computing and modelling variograms and kriging. *Catena* **2014**, *113*, 56–69, doi:10.1016/J.CATENA.2013.09.006.
27. Chiles, J.P.; Delfiner, P. *Geostatistics: Modeling Spatial Uncertainty*, 2nd ed.; John Wiley & Sons: Hoboken, NJ, USA, 2012; ISBN 978-047-018-315-1.
28. Zhang, Y.; Hamm, N.A.; Meratnia, N.; Stein, A.; Van De Voort, M.; Havinga, P.J. Statistics-based outlier detection for wireless sensor networks. *Int. J. Geogr. Inf. Sci.* **2012**, *26*, 1373–1392, doi:10.1080/13658816.2012.654493.
29. Liu, C.Y.; Chuang, C.L.; Chen, C.P.; Chang, W.Y.; Shieh, J.C.; Lin, C.H.; Tseng, C.L.; Jiang, J.A. Development of an embedded system-based gateway for environmental monitoring using wireless sensor network technology. In Proceedings of the 2011 Fifth International Conference on Sensing Technology (ICST), Palmerston North, New Zealand, 28 November–1 December 2011; pp. 544–548, doi:10.1109/ICST.2011.6137039.
30. Mesmoudi, K.; Soudani, A.; Bournet, P.E. Determination of the inside air temperature of a greenhouse with tomato crop under hot and arid climates. *J. Appl. Sci. Environ. Manag.* **2010**, *5*, 117–129.

