Sparse Interpolation via ε-Smooth Support Vector Regression and Uniform Design

李 育 杰
Data Science and Machine Intelligence Lab
國立交通大學應用數學系

中央氣象局
September 05, 2016

IoT: An Expected View in (near)Future

Data Deluge: 資料的洪荒之力

Potential of IoT

- The IoT has the potential to connect 26 billion Things to the Internet by 2020, in contrast to 7.3 billion units of PCs/notebooks/smartphones (Peter Middleton, research director, Gartner Inc.)
- Examples: wearable wristbands, home devices (ICS), transportation (ITS), smart cities and industry 4.0

Data from IoT

IoT provides a channel for smart sensing and continuously monitoring the interesting targets:

- Data generated by things: to monitor devices, machines or infrastructure such as energy meters, elevators, airplane engines, bridges. This data can be used for predictive analytics, to repair or replace these items before they break
- Data about things: to ***monitor natural phenomena*** such as meteorological patterns, underground pressure during oil extraction, or patient vital statistics during recovery from a medical procedure.

Data from IoT (cont'd)

- Sensor/IoT data is one of the major data resources now and IoT data analytics has become a paradigm in the Big Data era
  - Volume: Vast amount of sensors collecting data continuously
  - Velocity: Data coming in minutes, seconds and even in microseconds
  - Variety: Nominal/numerical types, text/multimedia, images, audio, and video, etc.
  - Veracity: High noises, inconsistency and incompleteness

Applying Sensing Technology

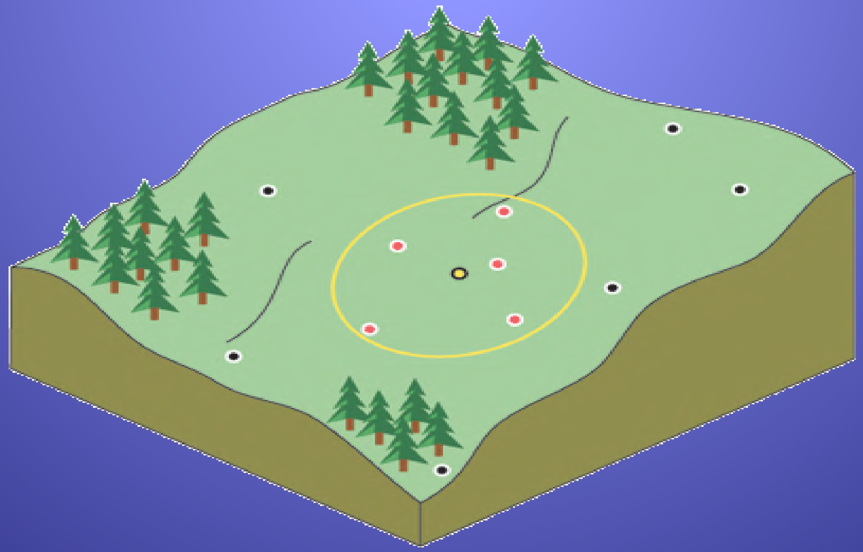
- Monitoring physical phenomena is an important application domain for wireless sensor networks.
- Continuous Monitoring is useful in a wide range of applications such as
  - Sensing
  - Monitoring in health conditions
  - Environmental monitoring
  - Sensing spectrum in cognitive networks

Reference by Field Estimation in Wireless Sensor Networks Using Distributed Kriging (Hern´andez-Pe˜naloza,2010)

Continuous monitoring and detection with model adaptation

Motivation

- A precise continuous monitoring systems is often impractical due to restrictions in sensor placement and availability.
- Discrete number of sensors – continuous variable of interest
- Sensors aren't always deployed uniformly



Interpolation methods

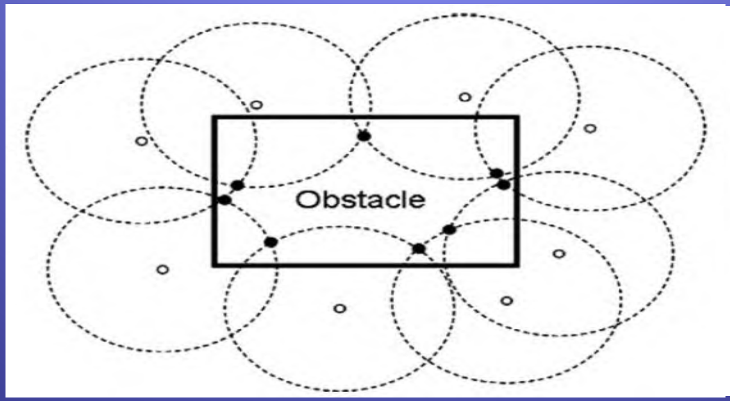
- Traditional methods include Ordinary Kriging (OK) and Inverse Distance Weighting (IDW)
- IDW weight formula: $w_i = \frac{1/d_i^2}{\sum_{j=1}^{n} 1/d_j^2}, \sum_{i=1}^{n} w_i = 1$ n = Sensor amount
- OK weight formula: $\sum_{j=1}^{n} \lambda_j cov(d_{ij}) = cov(d_{i0})$

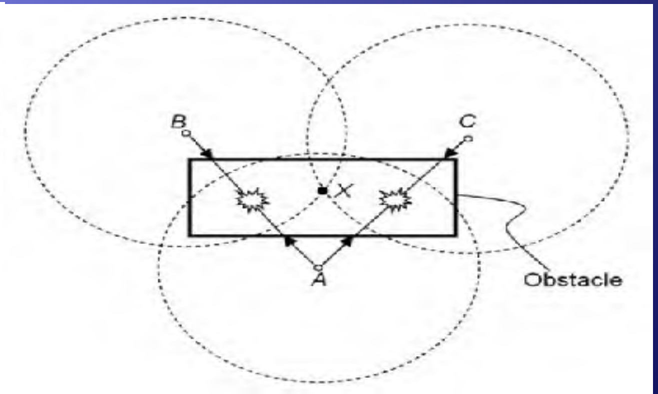
Calculate a weighted sum of measurements from surrounding sensors to interpolate a surface over the region.

Coverage Holes Problem

- The quality of any interpolation produced will suffer if sensors are too sparsely deployed.

1) Unavoidable obstacles such as walls and geographic formations.
2) Node failures caused by power depletion.
3) Environmental factors (heat, vibration, failure of electronic components or software bugs)
4) The vast scale and/or inherent hostility of the monitored area, e.g. ocean buoys.



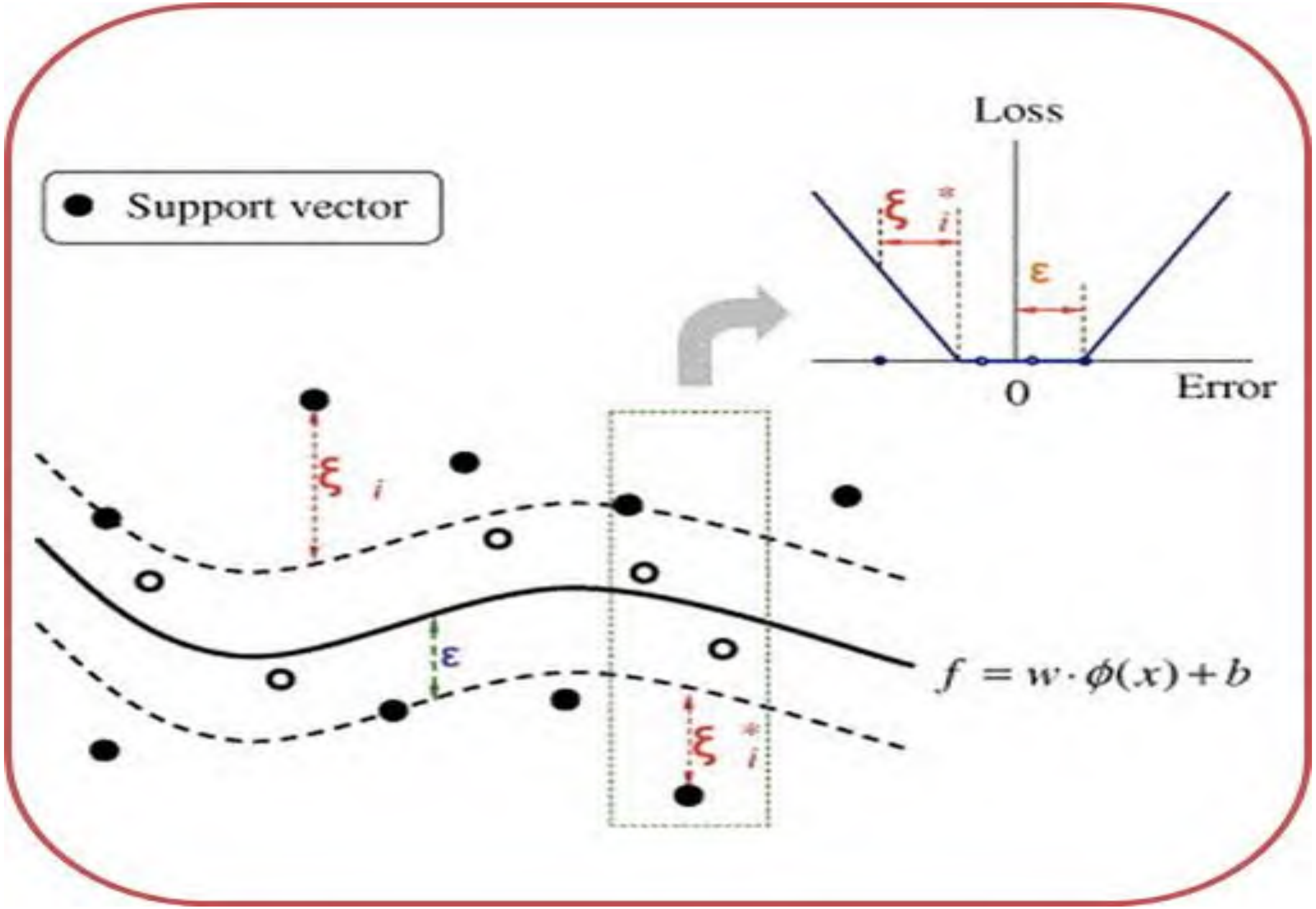


Reference by Obstacle detection and estimation in wireless sensor networks(Wang,2013)

Main Idea

1. Use a local linear interpolation to create "artificial sensors", *scattered uniformly across the region (Uniform Design)*
2. With the values of all the sensors, artificial and real, perform a global regression to find the surface for the whole region
3. Set *different tolerances* for the *real and artificial sensors (ε-insensitive smooth support vector regression)*

ε-insensitive SVR



ε-insensitive SVR

Consider:

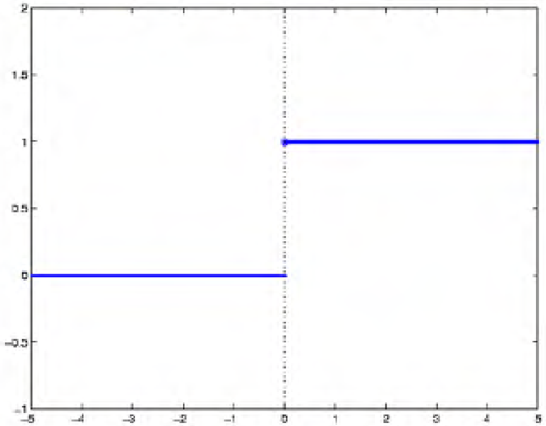
$$\min_{(w,b,\xi)\in\mathbb{R}^{n+1+m}} \frac{1}{2}\|w\|_2^2 + C\mathbf{1}^\top|\xi|_\epsilon$$

where $|\xi|_\epsilon \in \mathbb{R}^m, (|\xi|_\epsilon)_i = \max(0, |A_i w + b - y_i| - \epsilon)$

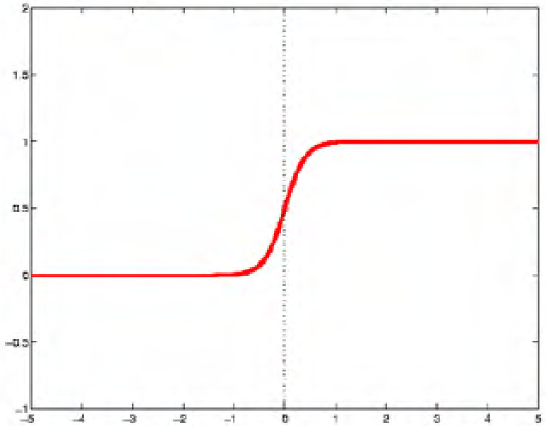
- C is the weight parameter that balances the training error and model complexity

Smooth Functions

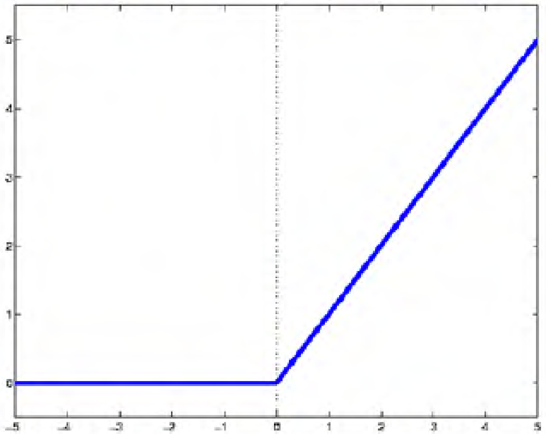
Step function x_*



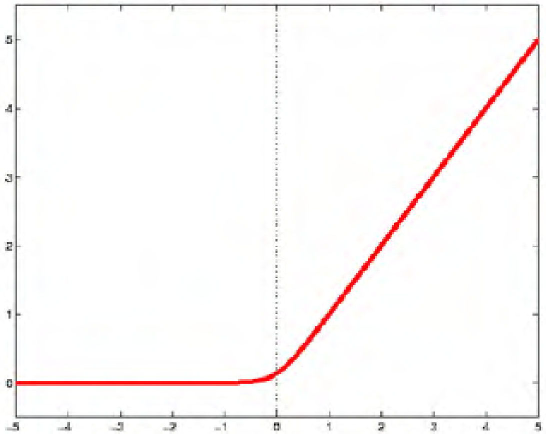
Sigmoid function: $\frac{1}{(1+e^{-5x})}$



Plus function: x_+

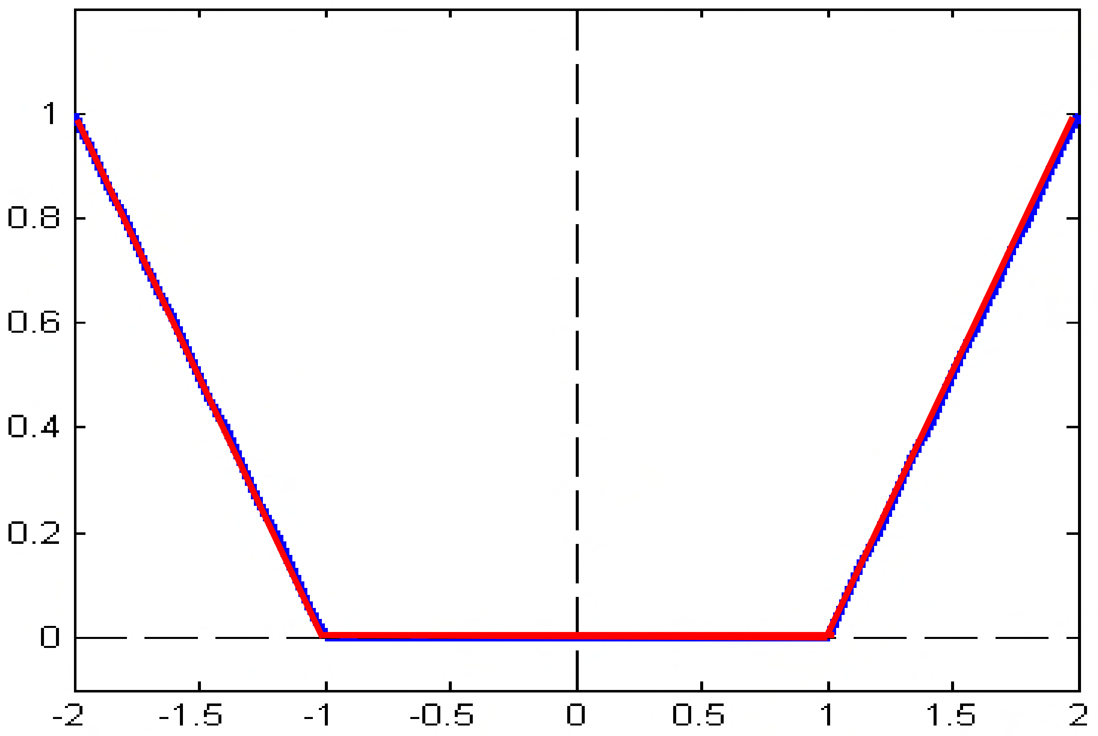


p-function: $p(x, 5)$

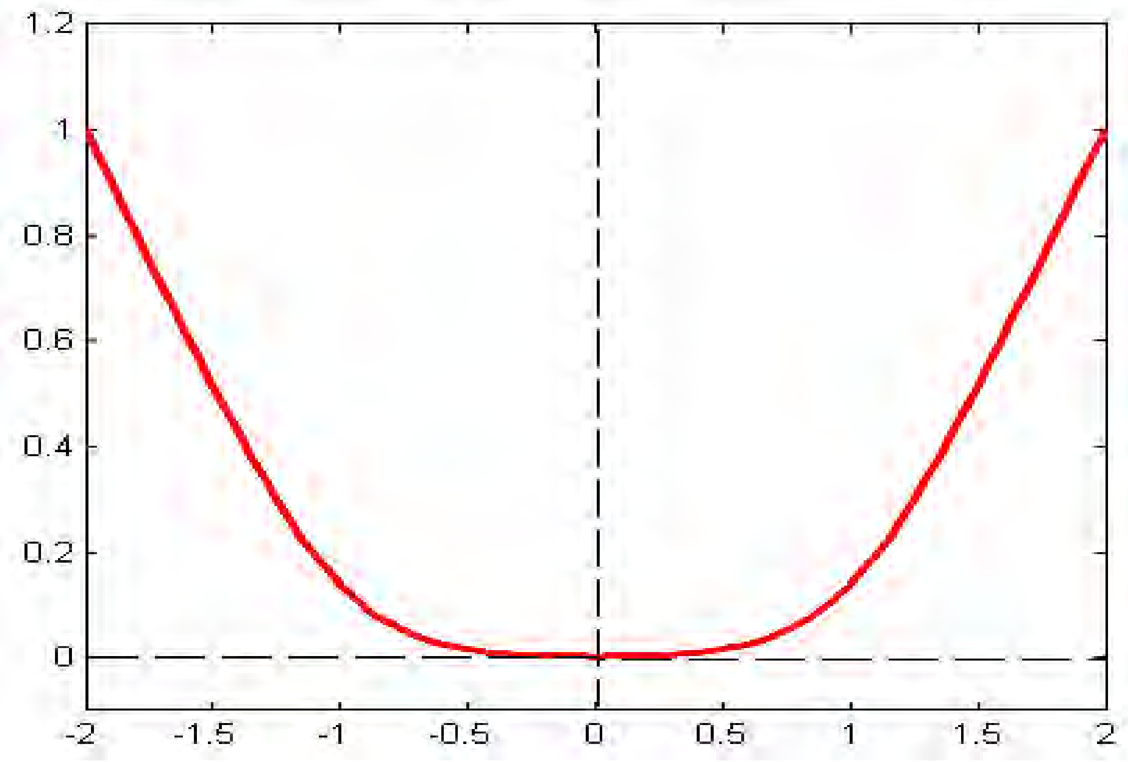


Smooth the ε-Insensitive Function

$$p_{\varepsilon}(x,\alpha) = p(x-\varepsilon,\alpha) + p(-x-\varepsilon,\alpha)$$



$\varepsilon = 1$

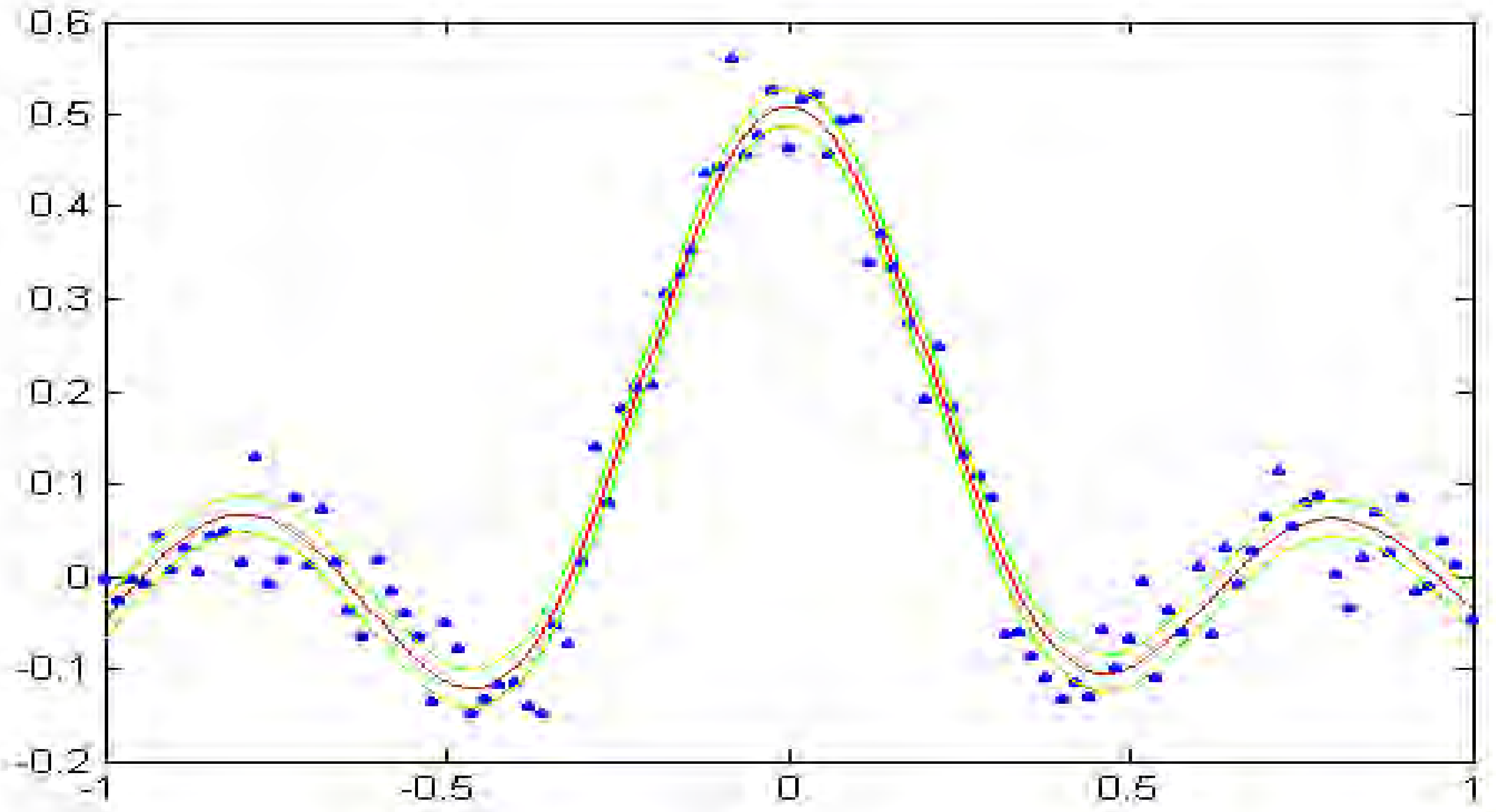


$\varepsilon = 1,\ \alpha = 5$

101 Data Points in $R \times R$

Nonlinear SSVR with Kernel: $\exp^{-\mu \|x^i - x^j\|_2^2}$

$$f(x) = 0.5 * \sin c(\frac{10}{\pi}) + noise$$



$x \in [-1,1]$, 101 points

Noise: mean=0

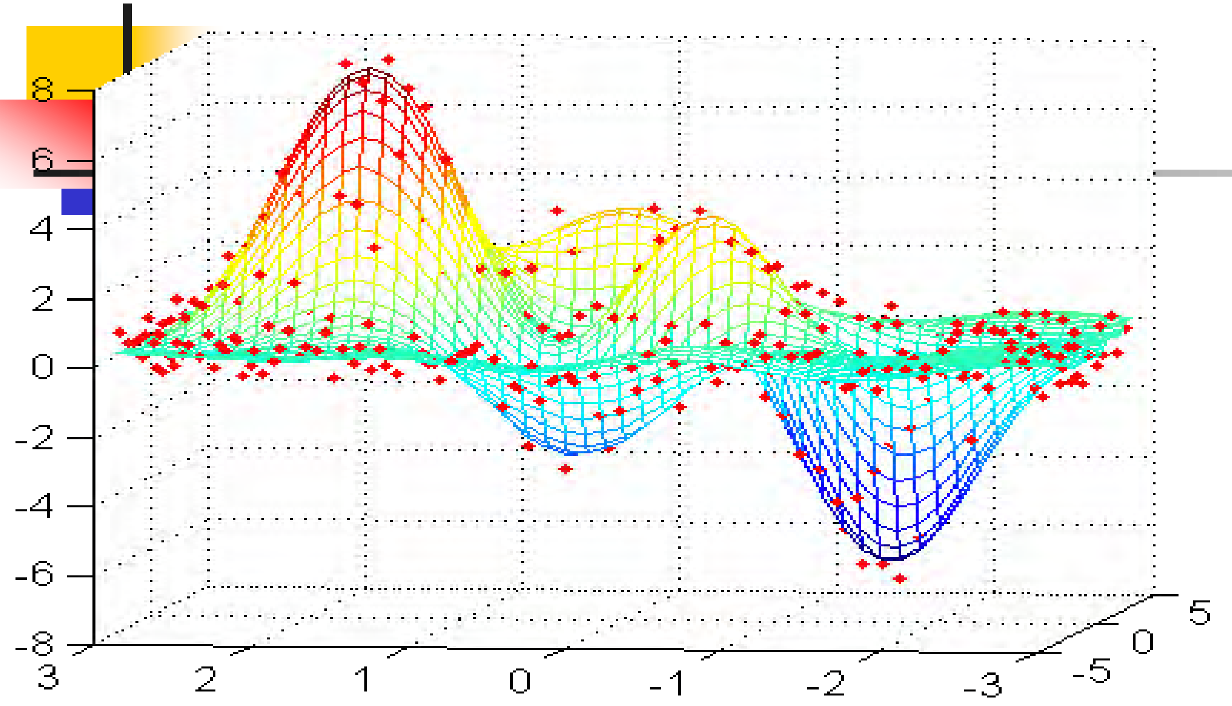
$\sigma = 0.04$

Parameter:

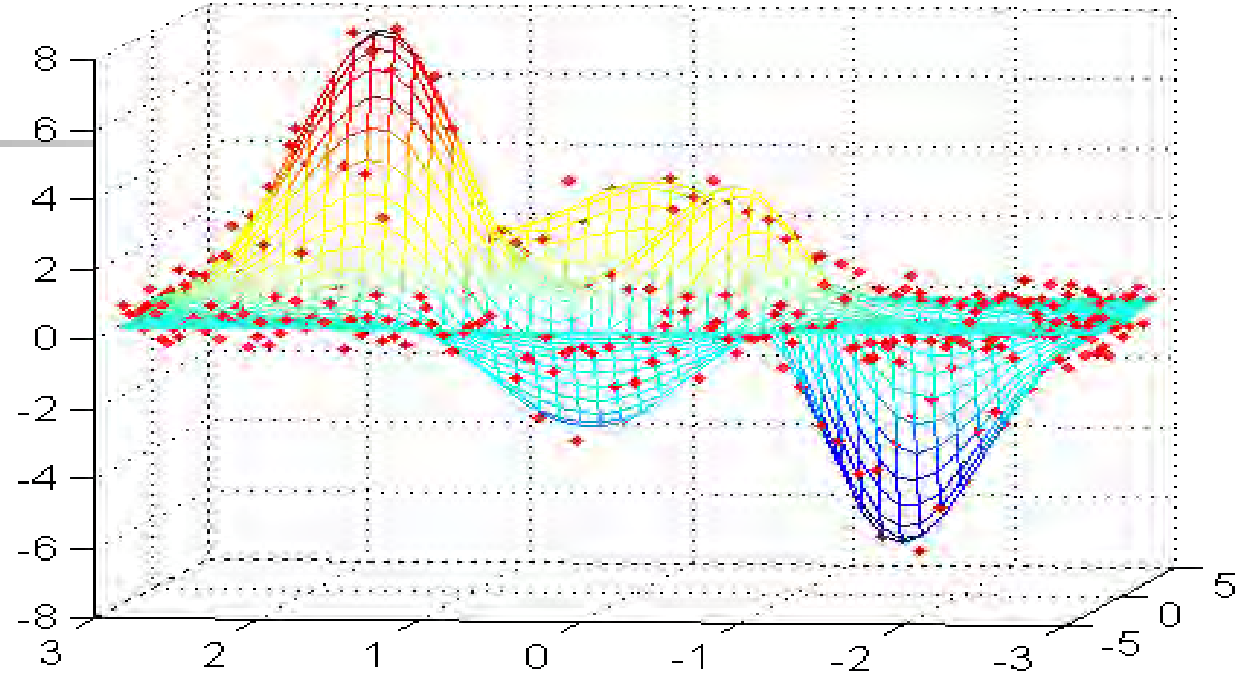
$\nu = 50, \mu = 5, \varepsilon = 0.02$

Training time : 0.3 sec.

481 Data Points in $R^2 \times R$



Estimated Function



Original Function

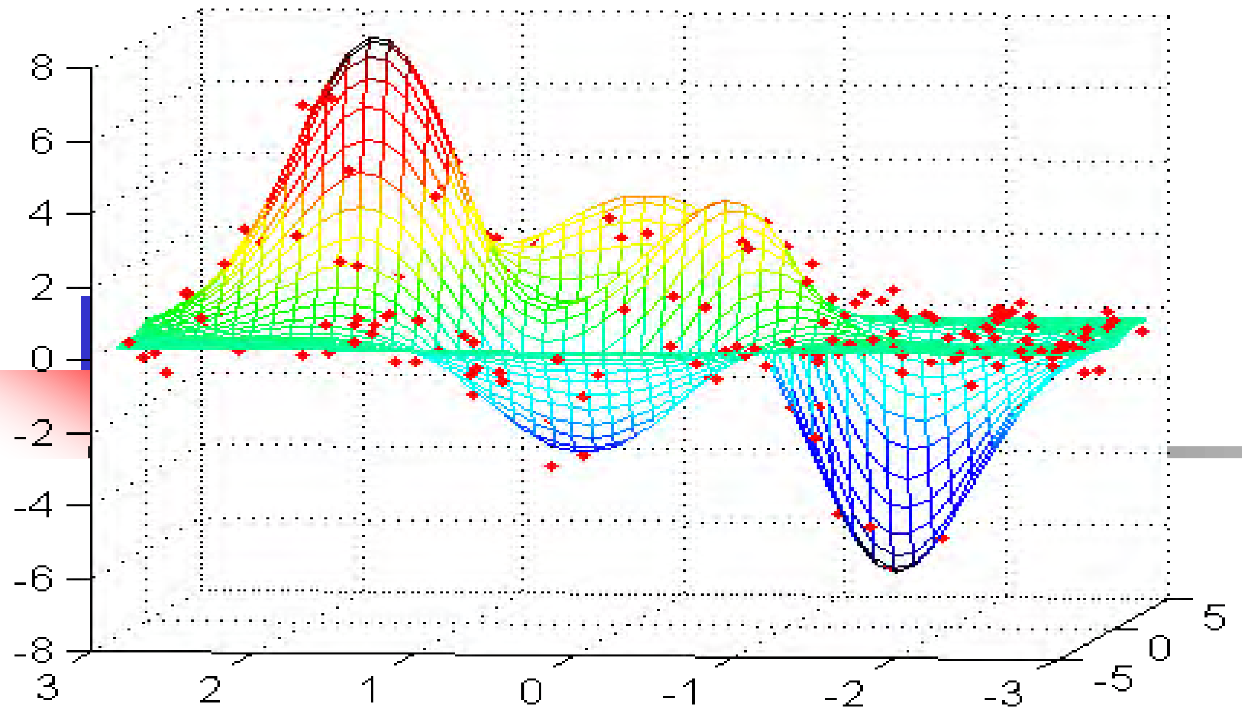
Noise : mean=0 , $\sigma = 0.4$

Parameter : $\nu = 50, \mu = 1, \varepsilon = 0.5$

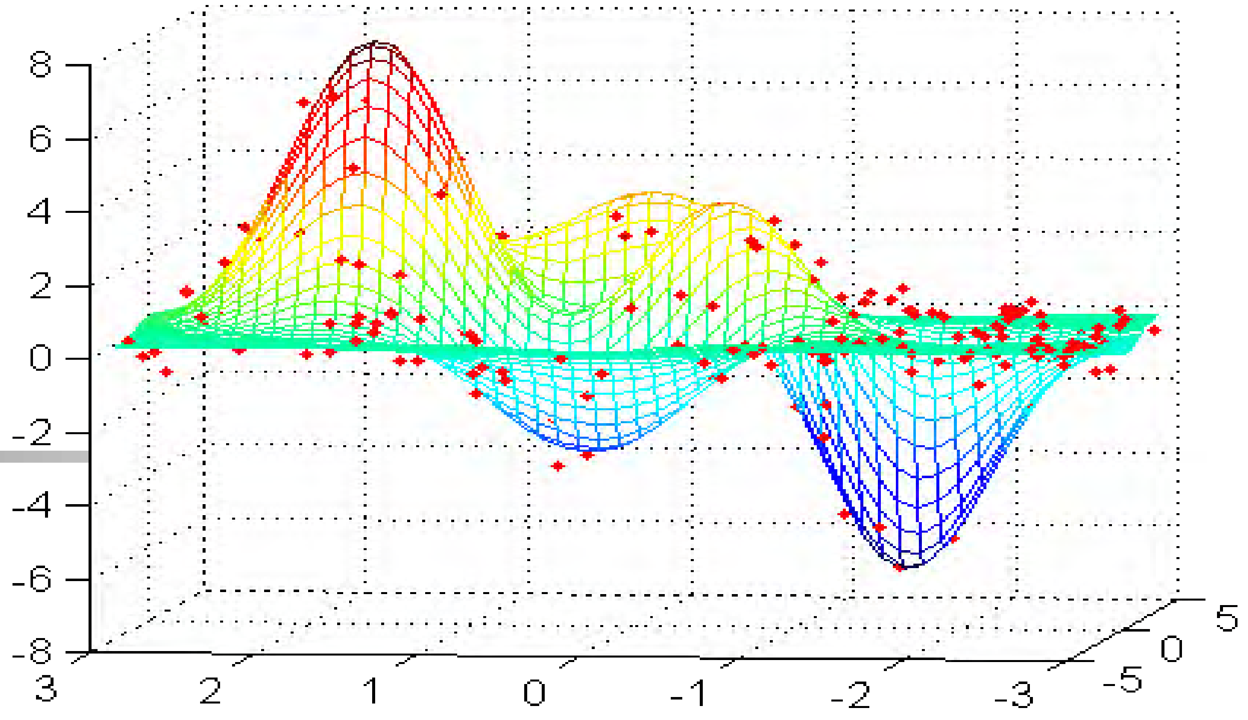
Mean Absolute Error (MAE) of 49x49 mesh points : 0.1761

Training time : 9.61 sec.

Using Reduced Kernel: $K(A, \overline{A'}) \in R^{28900 \times 300}$



Estimated Function



Original Function

Noise : mean=0 , $\sigma = 0.4$

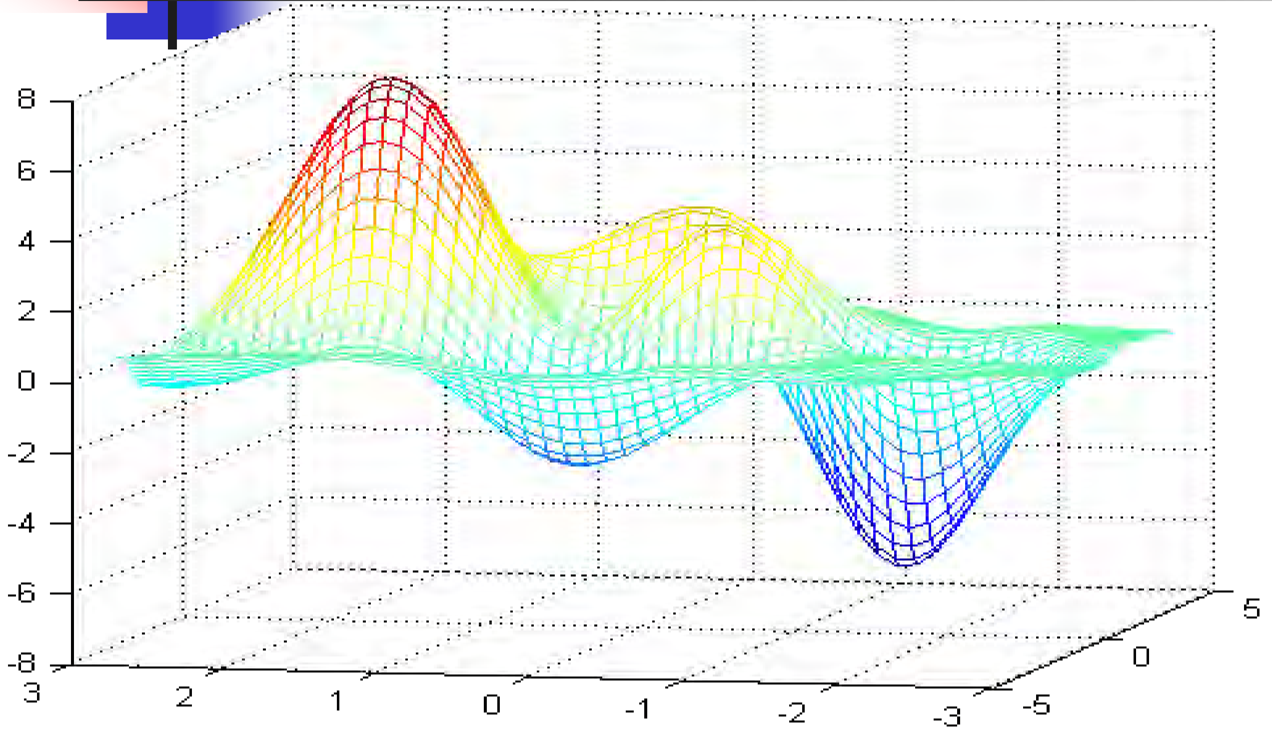
Parameter : $\nu = 50, \mu = 1, \varepsilon = 0.5$

MAE of 49x49 mesh points : 0.0513

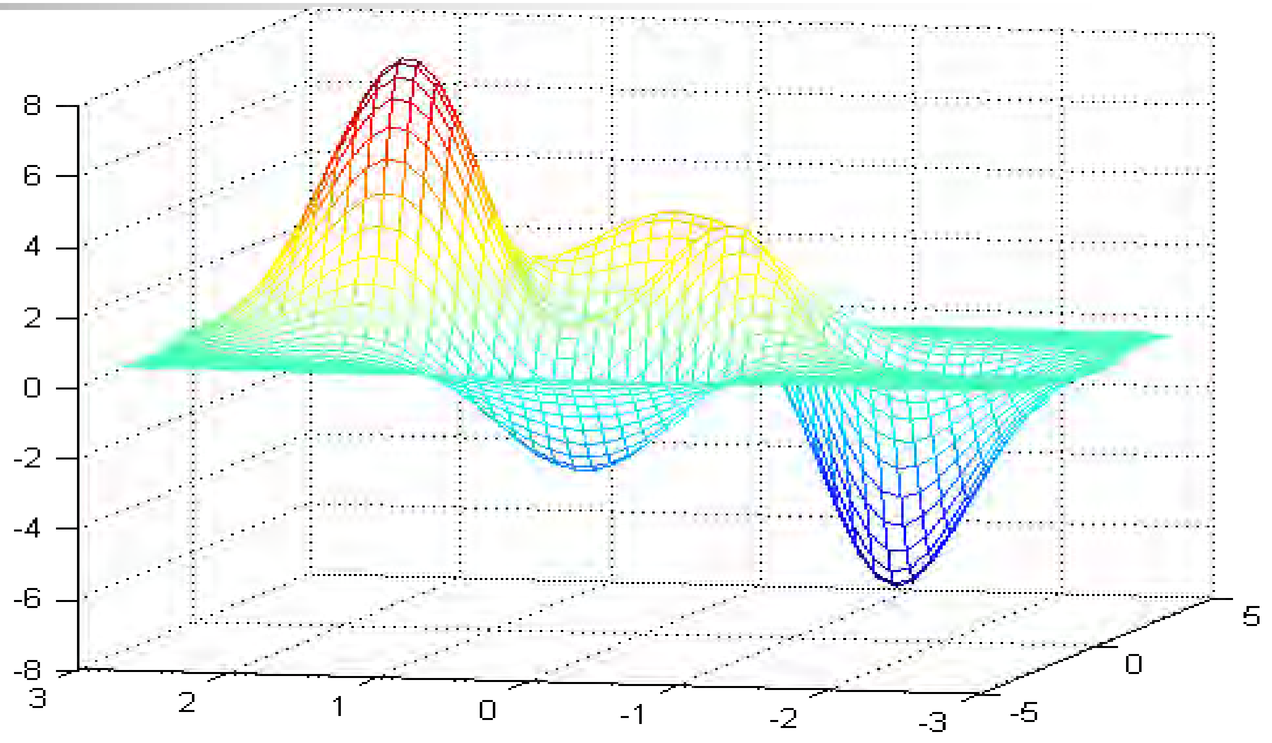
Training time : 108 sec.

Using SAME 300 Random Points Out of 28900

$$K(\overline{A}, \overline{A}') \in R^{300 \times 300}$$



Estimated Function



Original Function

Noise : mean=0 , $\sigma = 0.4$

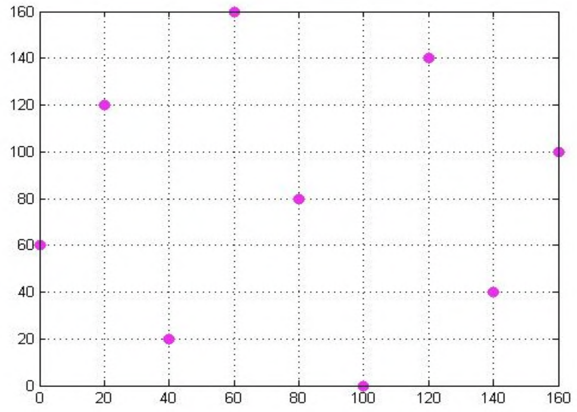
Parameter : $\nu = 50, \mu = 1, \varepsilon = 0.5$

MAE of 49x49 mesh points : 0.2529

Uniform Design

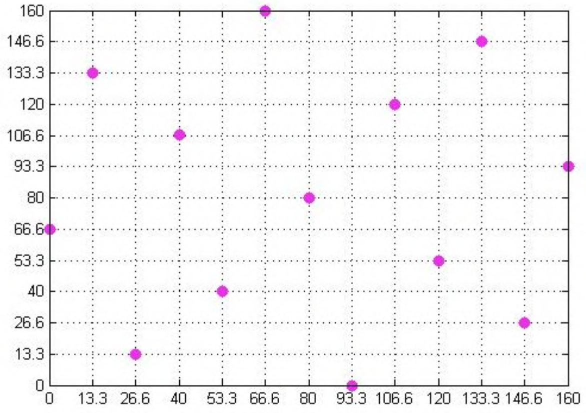
9-point UD Pattern

5	5
1	4
7	8
2	7
3	2
9	6
8	3
6	1
4	9



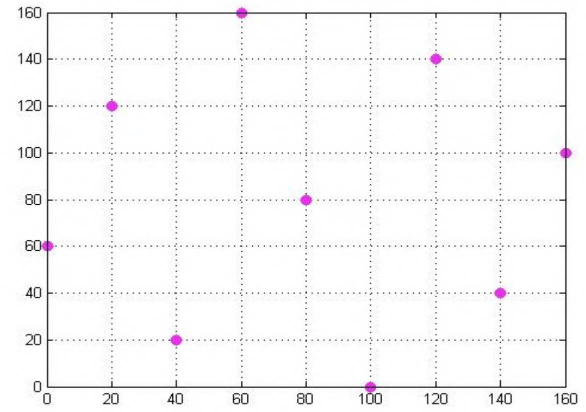
13-point UD Pattern

5	4
12	3
2	11
9	10
7	7
6	13
3	2
11	12
13	8
10	5
1	6
4	9
8	1

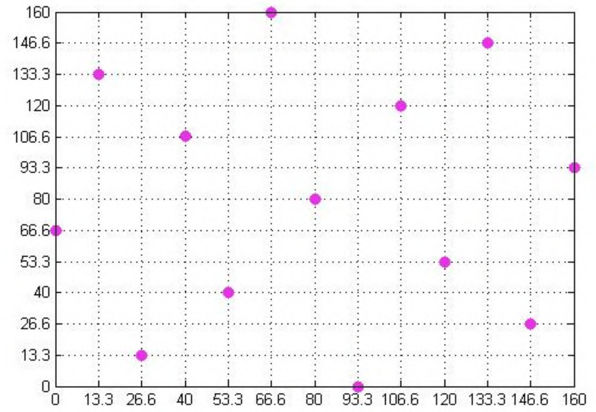


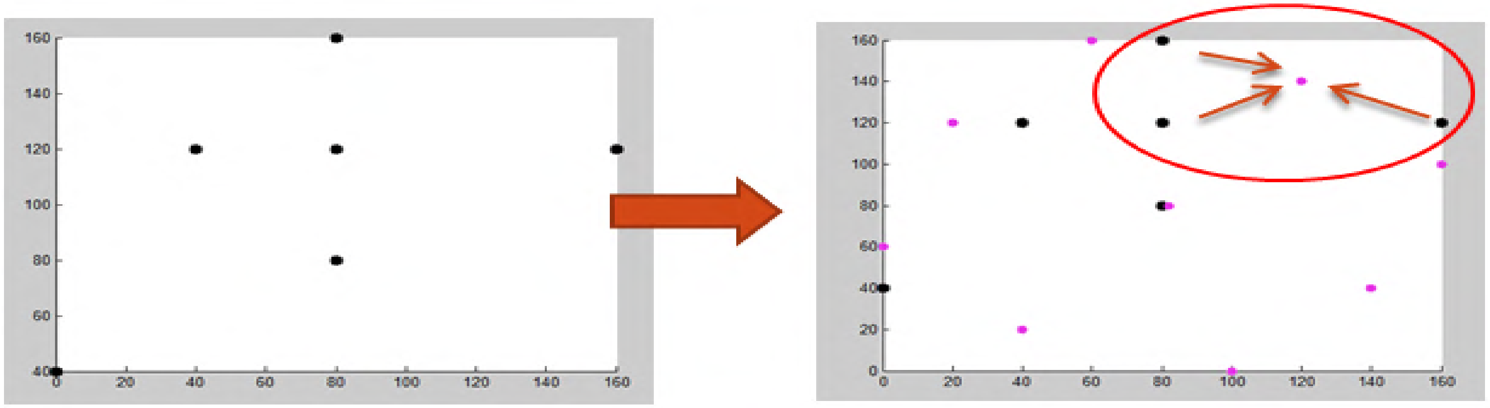
Synthesizing Sensor Readings via Uniform Design Sampling

9-point UD Pattern

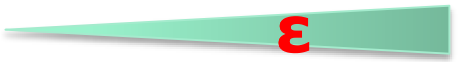


13-point UD Pattern





- Use interpolation methods for estimating readings at UD points.
- Use different **ε** for real and synthetic readings.



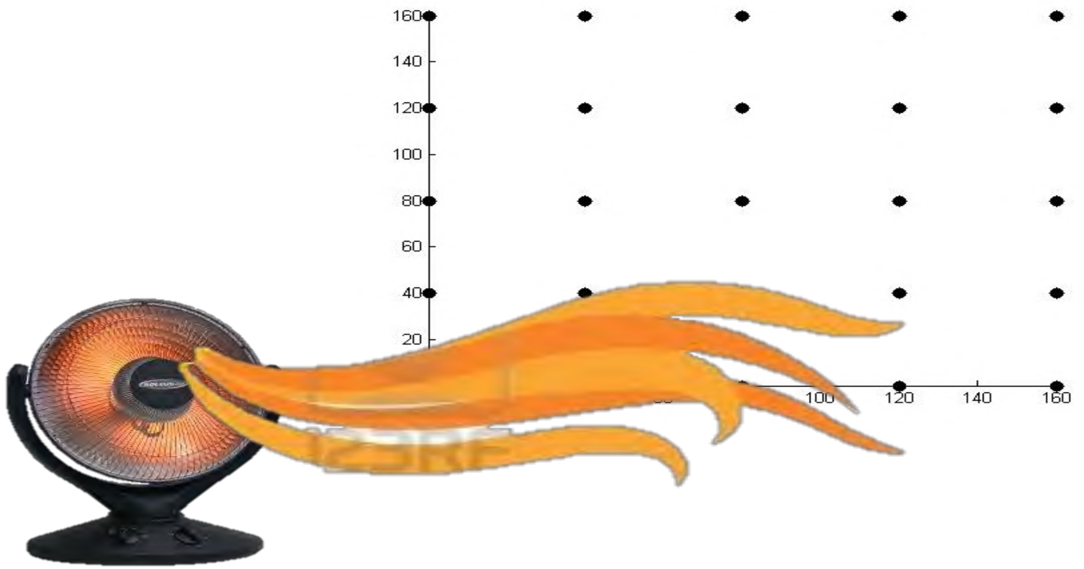
real UD

Global Regression: Combine Real Sensor Reading and Synthetic Estimations

- Apply the k-nearest neighbors information to interpolate the synthetic sensor values (linear regression, IDW or OK)
- Utilized nonlinear ε-SSVR to do global regression
  - Use different ε values in the ε-insensitive function
  - Synthetic sensor should use a bigger ε

Smart Agriculture demo

NodeID	LocationX	LocationY	Temperature	Humidity	Time
1	0	0	27.4	60.2	00:38:27
2	40	40	23.7	62.4	00:38:27
3	80	40	23.5	62.7	00:38:27
4	0	120	23.7	62.5	00:38:27



Number of data points: 25
Map(Environment) Range:[160x160cm]
Time Stamp: 20 round
Time Interval: 1 minute
Scenario: Simulate anomalous temperatures

Auto-Regressive and Moving Average Model

- Store past T time-stamped temperature data (Assume T =3)

NodeID	LocationX	LocationY	Temperature	Time
1	0	0	27.4	00:38:27
1	0	0	27.9	00:39:27
1	0	0	30.1	00:40:27
1	0	0	31.4	00:41:27

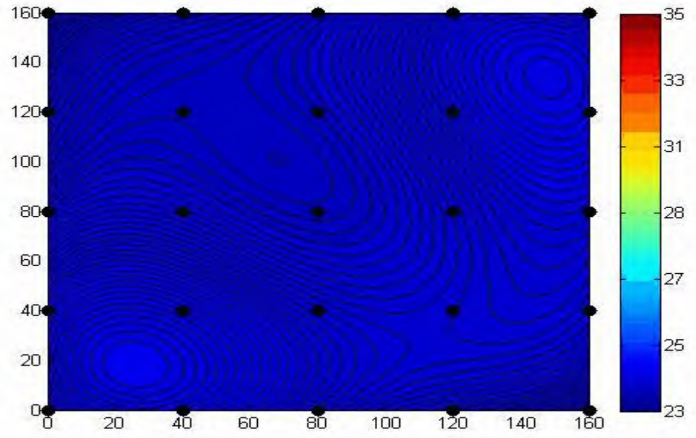
The current time

New Features	
Feature1	0
Feature2	0
Feature3	30.1
Feature4	27.9
Feature5	27.4

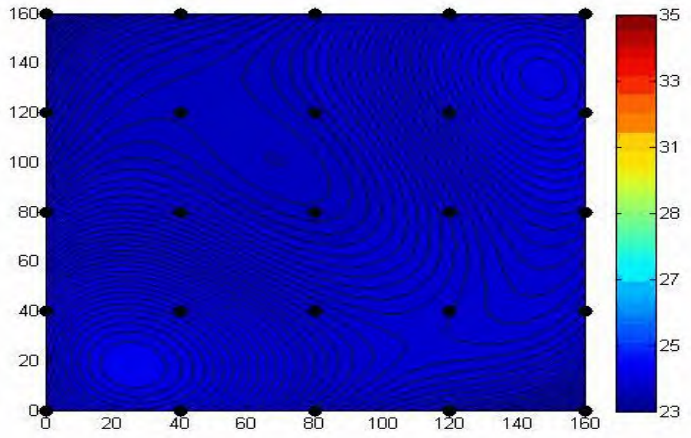
- Use stored temperature readings as new additional features

Visualization: Interpolation under uniformly distributed sensors

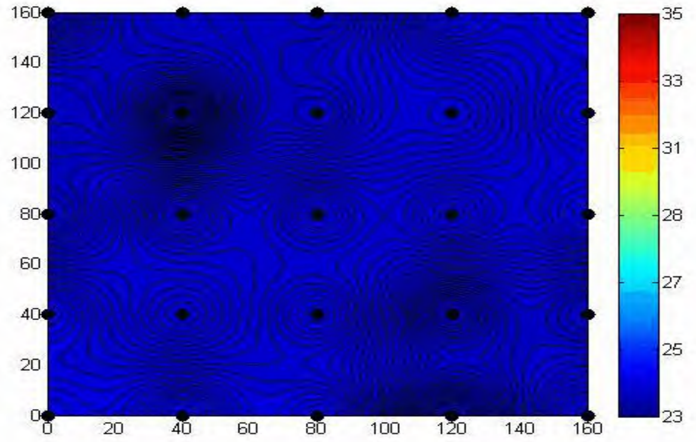
ε-SSVR



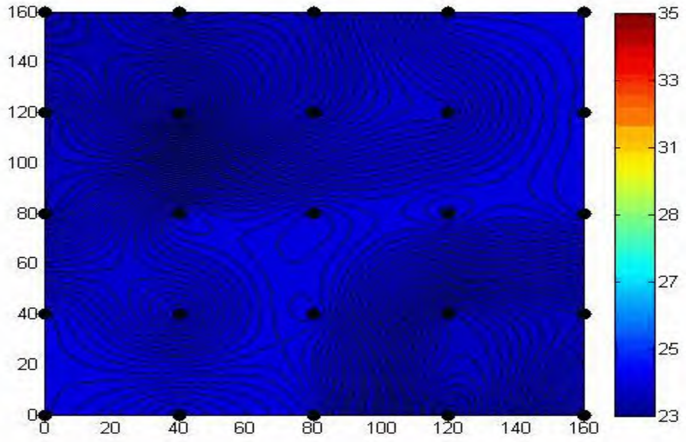
ε-SSVR(Spatial + Temporal(T=5))



IDW

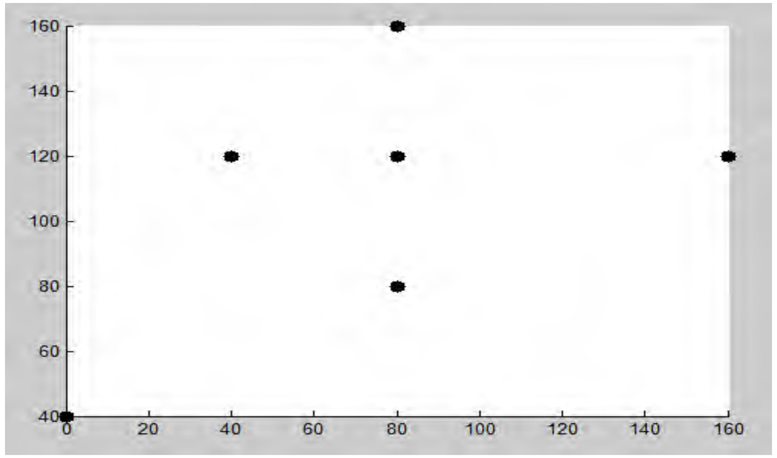


Ordinary Kriging

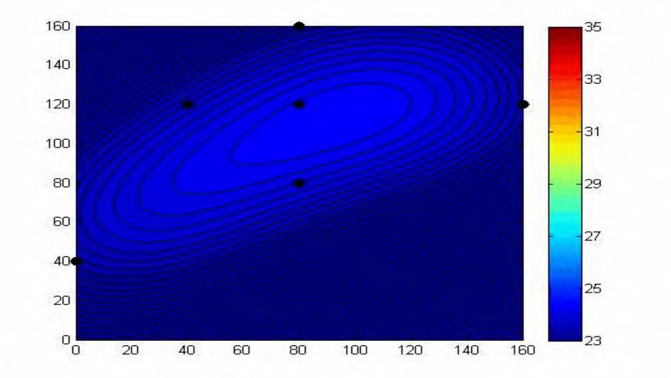


Sparse Coverage Experiment

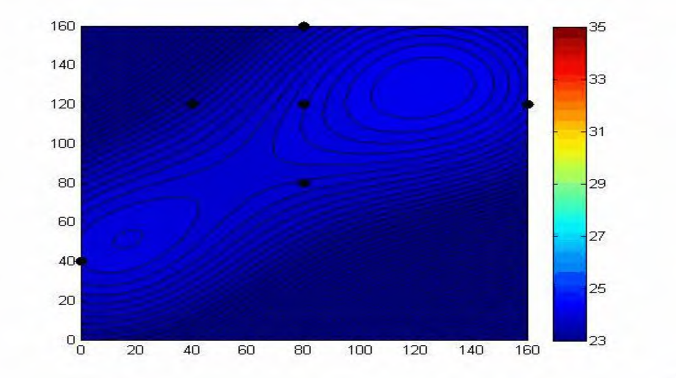
We randomly remove 19 nodes to be a validation set



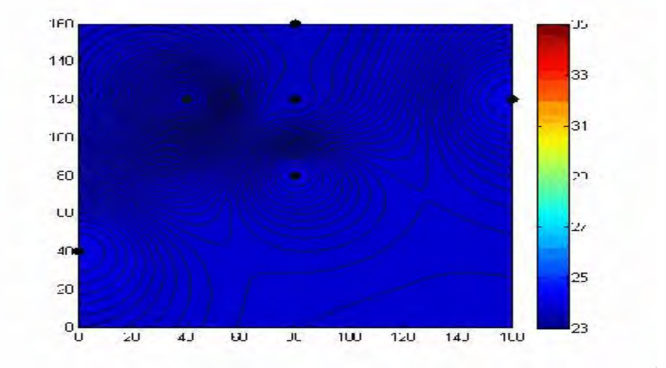
SSVR



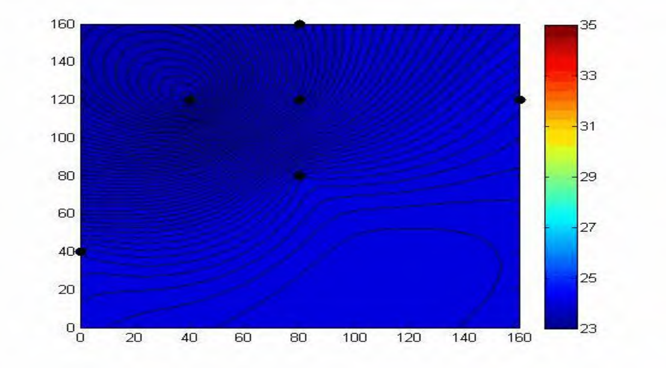
SSVR(Spatial + Temporal)



IDW



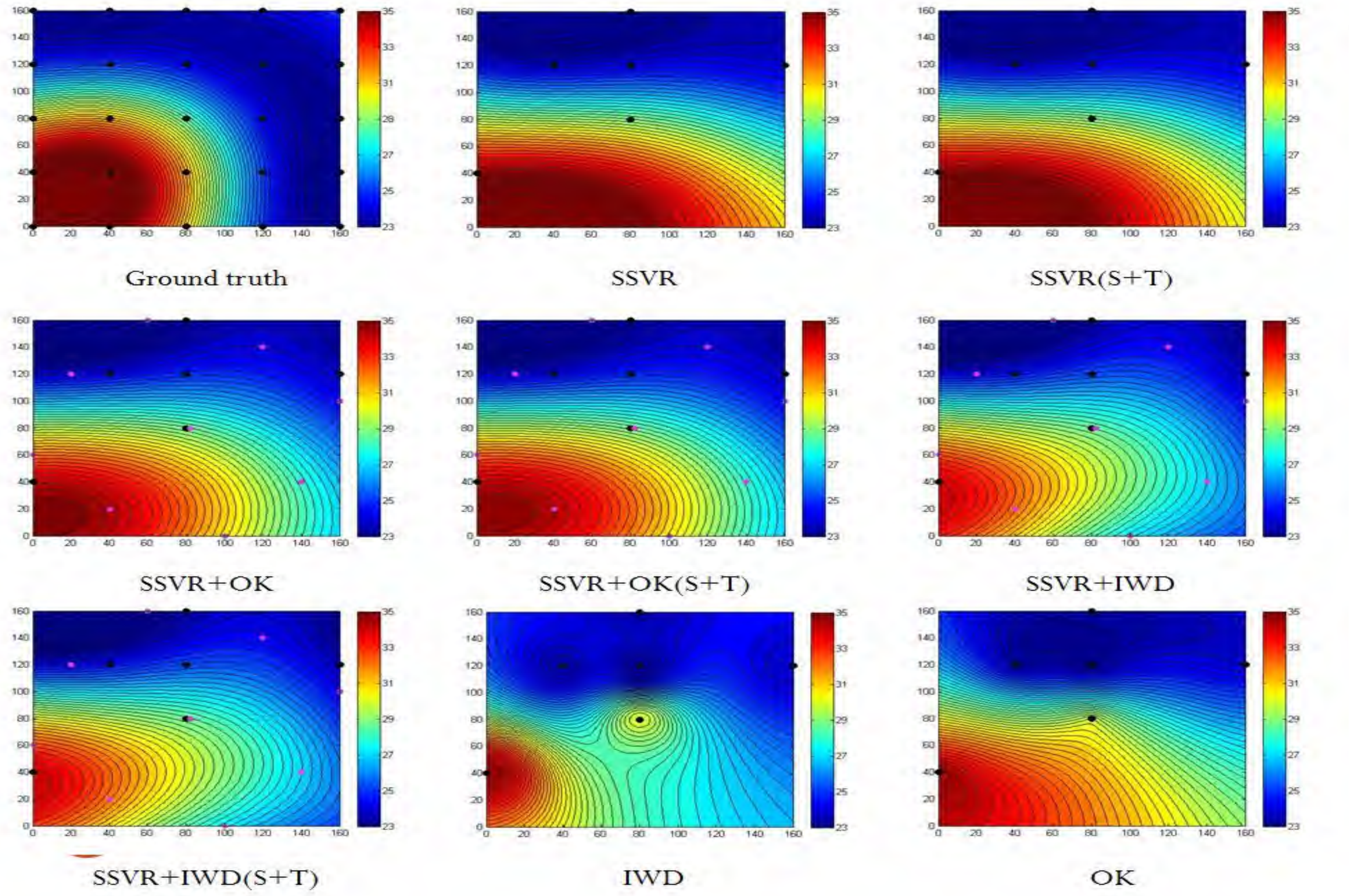
Ordinary Kriging



Results

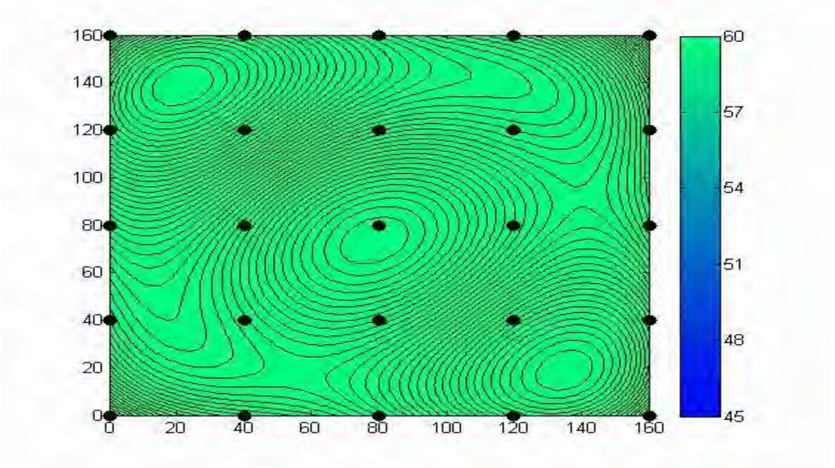
	MAE	RMSE	CPU sec.
ε-SSVR	1.9359	2.3923	0.0926
ε-SSVR(S+T)	1.9336	2.3894	0.4399
ε-SSVR+OK	1.5632	1.983	0.1902
ε-SSVR(S+T)+OK	1.5577	1.9791	0.5429
ε-SSVR+IDW	1.5613	1.9902	0.1424
ε-SSVR(S+T)+IDW	1.5575	1.9861	0.441
OK	1.6656	2.0998	57.8343
IDW	1.8652	2.3135	0.0126

Visualization of Sparse Coverage Experiment

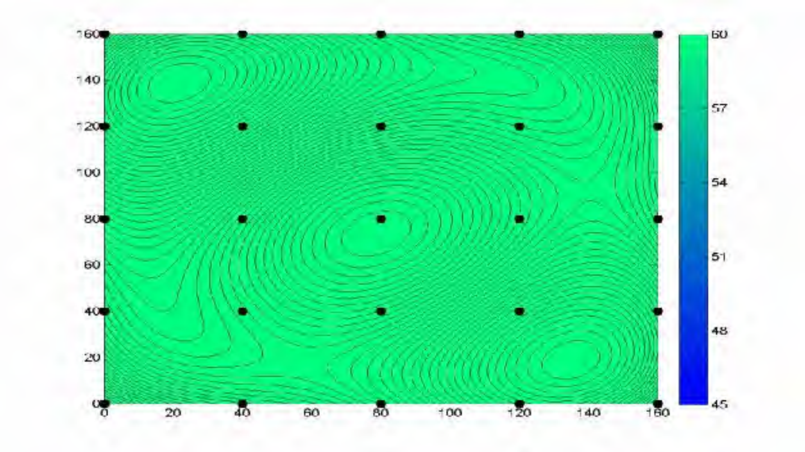


Visualization (humidity)

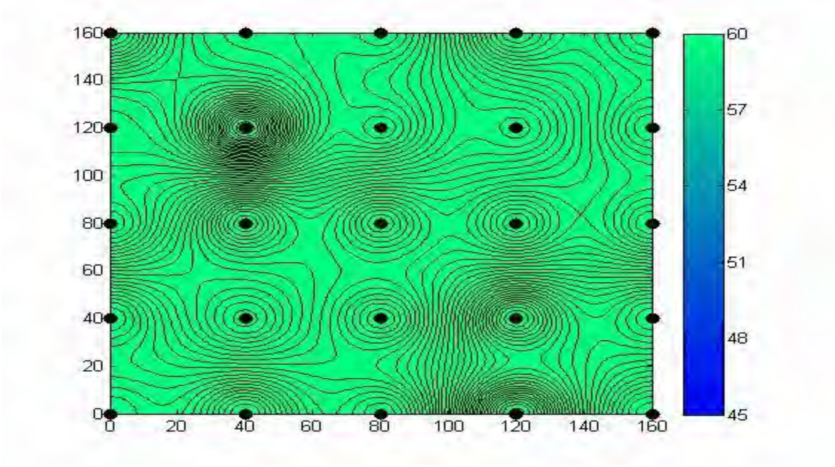
ε-SSVR



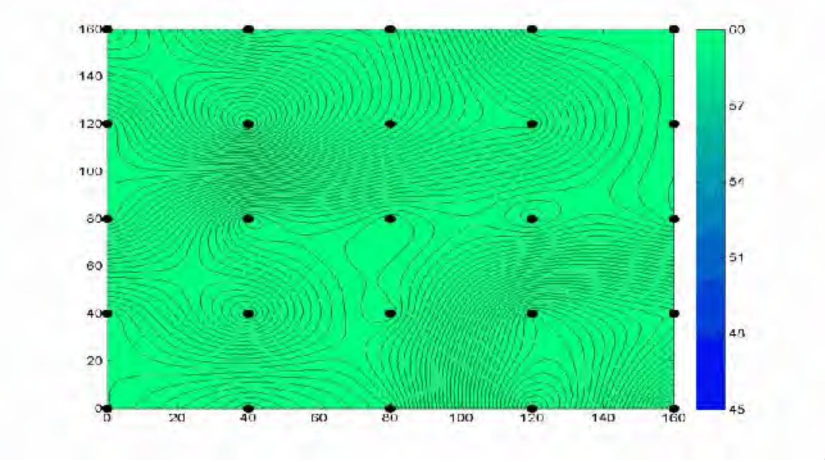
SSVR(Spatial + Temporal(T=5))



IWD



Ordinary Kriging



Definition of Anomaly (1/3)

- One possible definition of anomaly
  - An outlier is an observation that deviates so much from other observations as to arouse suspicion that it was generated by a different mechanism (by Hawkins).

- Michael Jordan is an outlier because of a well-known quotation by Charles Barkley: “I am the best basketball player in the earth, Jordan? He is an alien”.

Definition of Anomaly (2/3)

- Our proposed definition of Anomaly
  - An outlier is an observation that enormously affects model when we add or remove it from the entire dataset.

- Wilt Chamberlain is an outlier on account of his responsibility for several rule changes in basketball.
- In order to diminish his dominance, the basketball authorities set some rules including widening the lane, as well as changes to rules regarding inbounding the ball and shooting free throws.

Definition of Anomaly (3/3)

- Conventional anomaly detection approach:
  - Distance-based
  - Density-based
- Based on our definition, we can have
  - The perturbation of the principal component *with or without* an individual instance. E.g., online oversampling PCA
  - The data compression rate of *with or without* a certain portion of data. E.g., Kolmogorov complexity
- All computation has to be done very quickly
- Have to be able to evaluate the change of this *with or without* effect

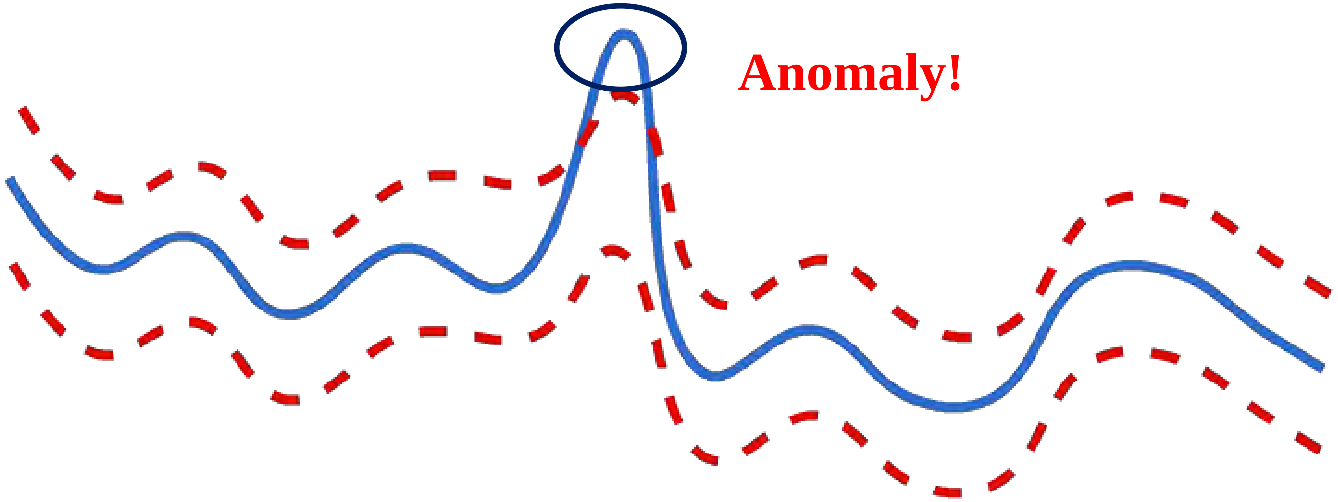
Finding a Needle in a Haystack

- Anomalous behaviors are rare events
- However, when they do occur, their consequences can be quite dramatic and quite often in a negative sense
- We aim to develop an unsupervised online anomaly detection mechanism
  - Can deal with stream data
  - A self-learning front-end model
    - Online learning
  - An accurate backend model
  - Cooperation between above models

"Mining needle in a haystack.
So much hay and so little time"

Front-end detection via Dynamic Range Checking (Single Sensor)

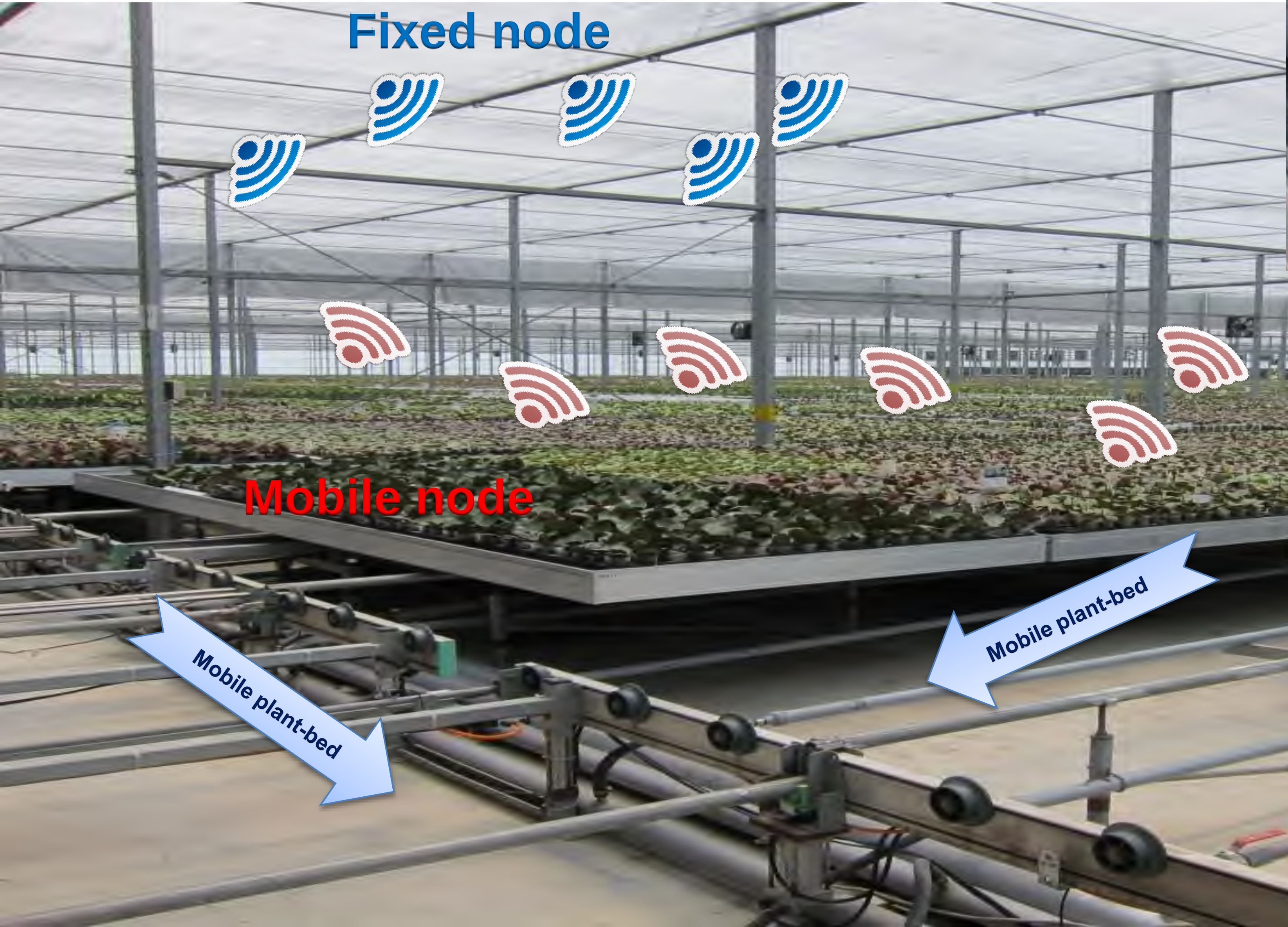
- Data read at close intervals are expected to have similar distributions
- Aggregate past data by storing mean and standard deviation
- Online update values as new readings arrive.



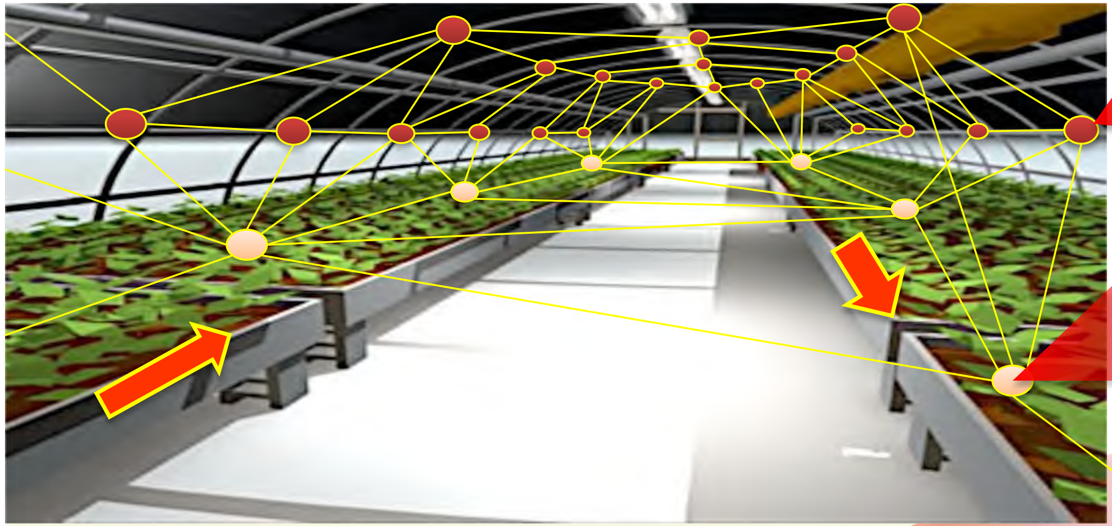
Anomaly occurs outside the range

Events Detection in Greenhouse

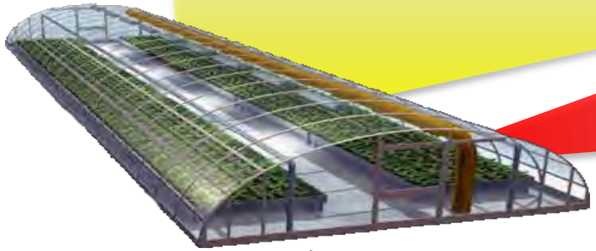
Green House



Project overview



M2M Networking (Fixed+ Mobile Nodes)



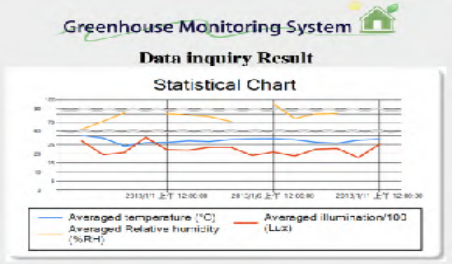
Automatic Greenhouse

52 Fixed Nodes

68 Mobile Nodes

Gateway

Real-time data inquiry interface



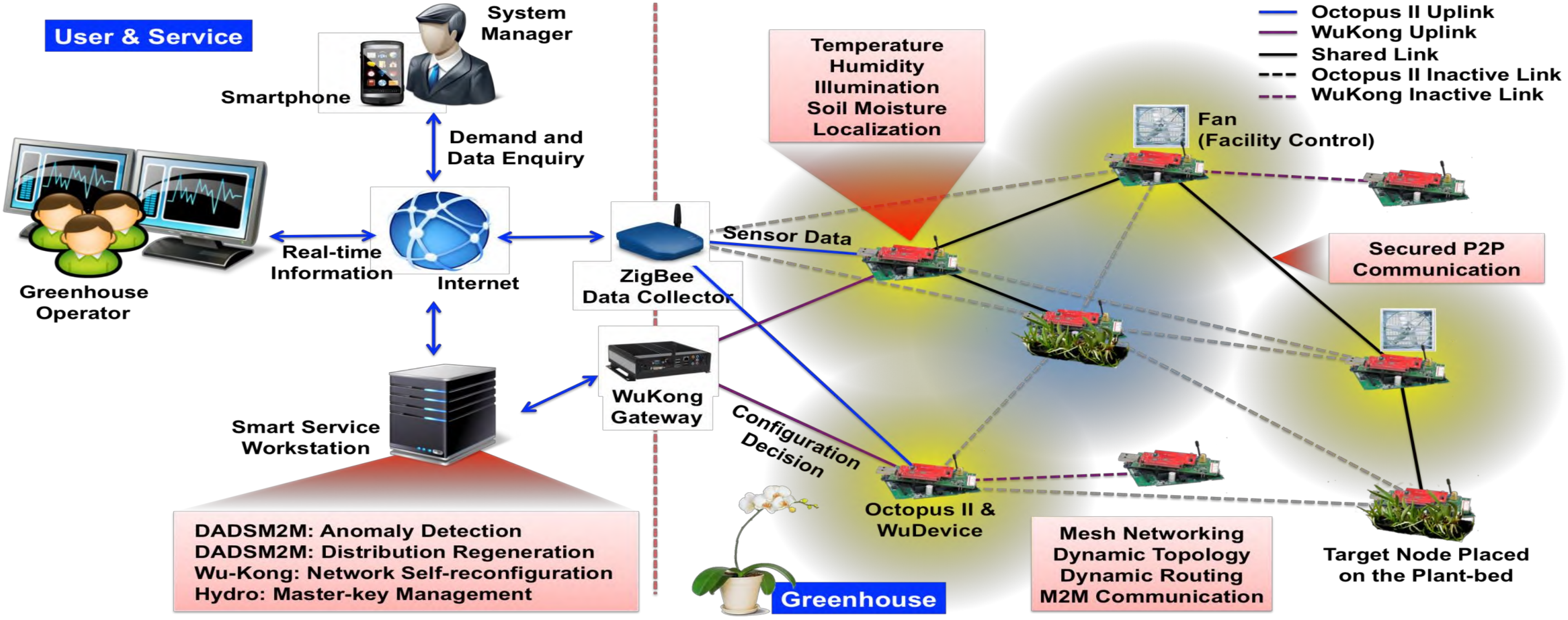
Visualization interface

Objectives

- **Scalability (unlimited number of sensor nodes in a single PAN)**
- **Robustness (dynamic topology, routing and localization)**
- **Heterogeneous (ZigBee + WiFi + different devices)**
- **Smart services (lighting, irrigation and inspection)**

System Architecture

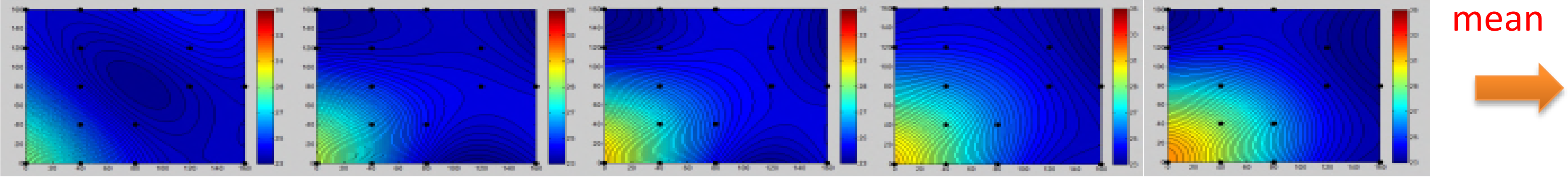
(Projects in NTU CCC Center, 2012-2015)



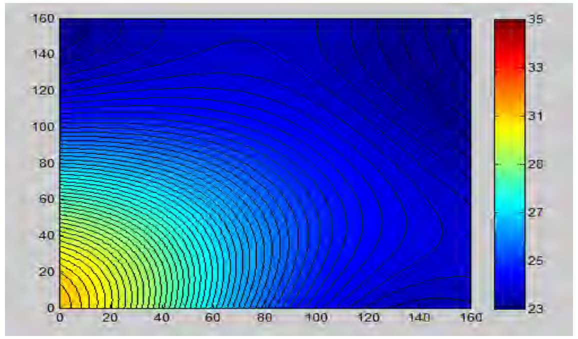
Back-end Detection via Continuous Monitoring

We keep past T readings map and take the average

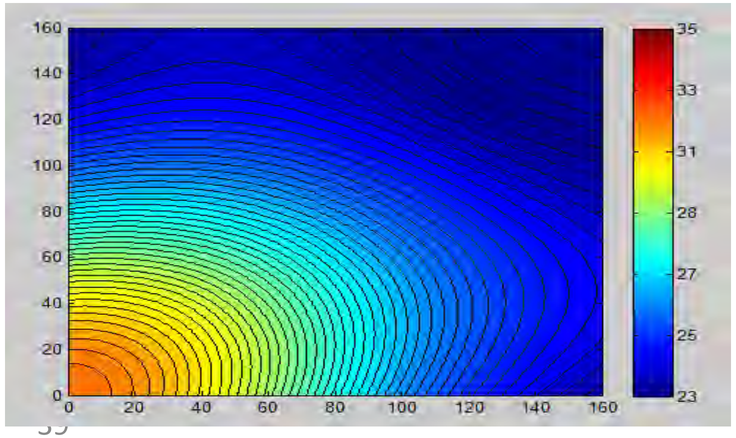
If T = 5



mean

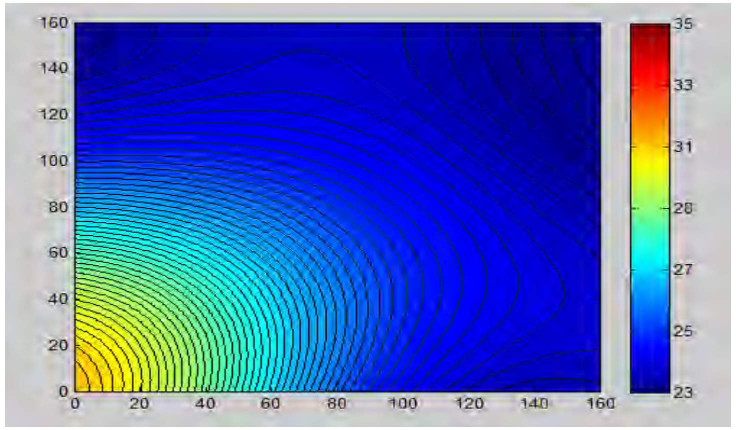


New map



—

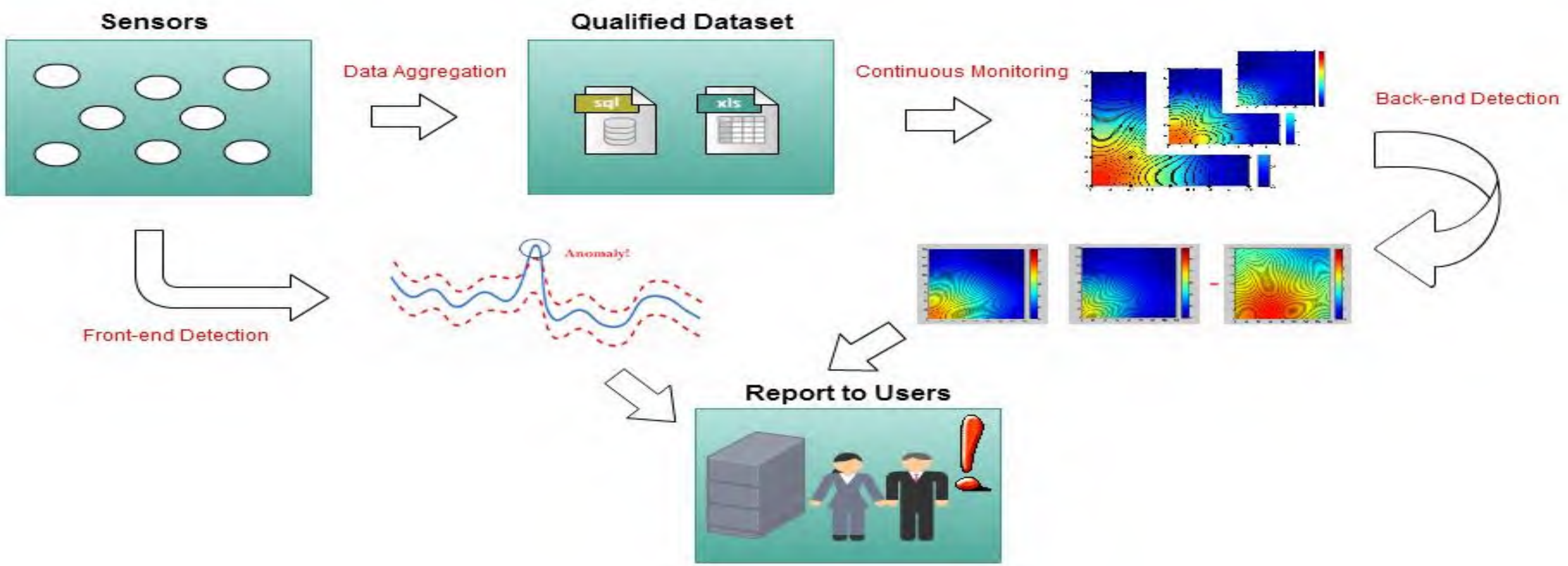
T readings average



=

Difference

Proposed Anomaly Detection Architecture



Questions?

Thank You!