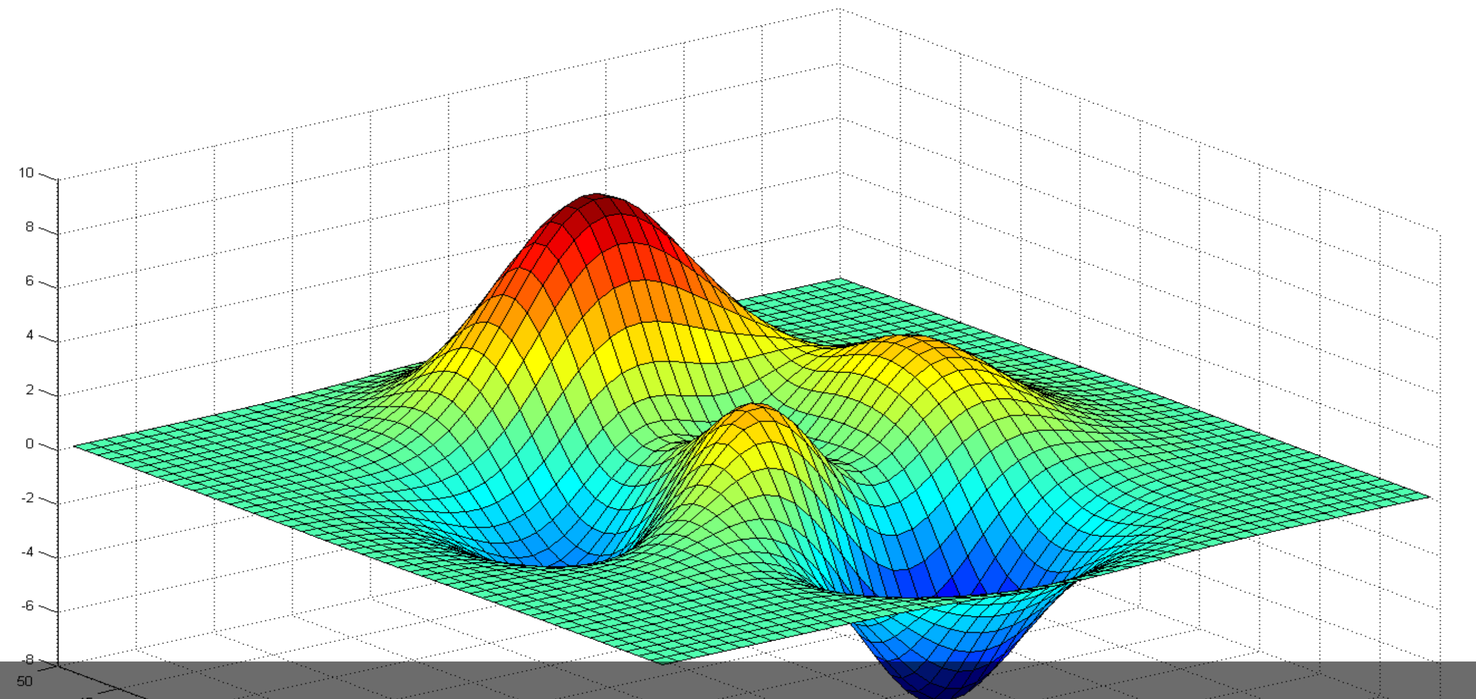


Space-Time Kernels

Dr. Jiaqiu Wang, Dr. Tao Cheng James Haworth

University College London

Joint International Conference on Theory, Data Handling and Modelling in GeoSpatial Information Science, Hong Kong, 26th-28th May 2010

STANDARD
Spatio-Temporal Analysis of Network Data and Route Dynamics

Methods for modelling spatio-temporal data

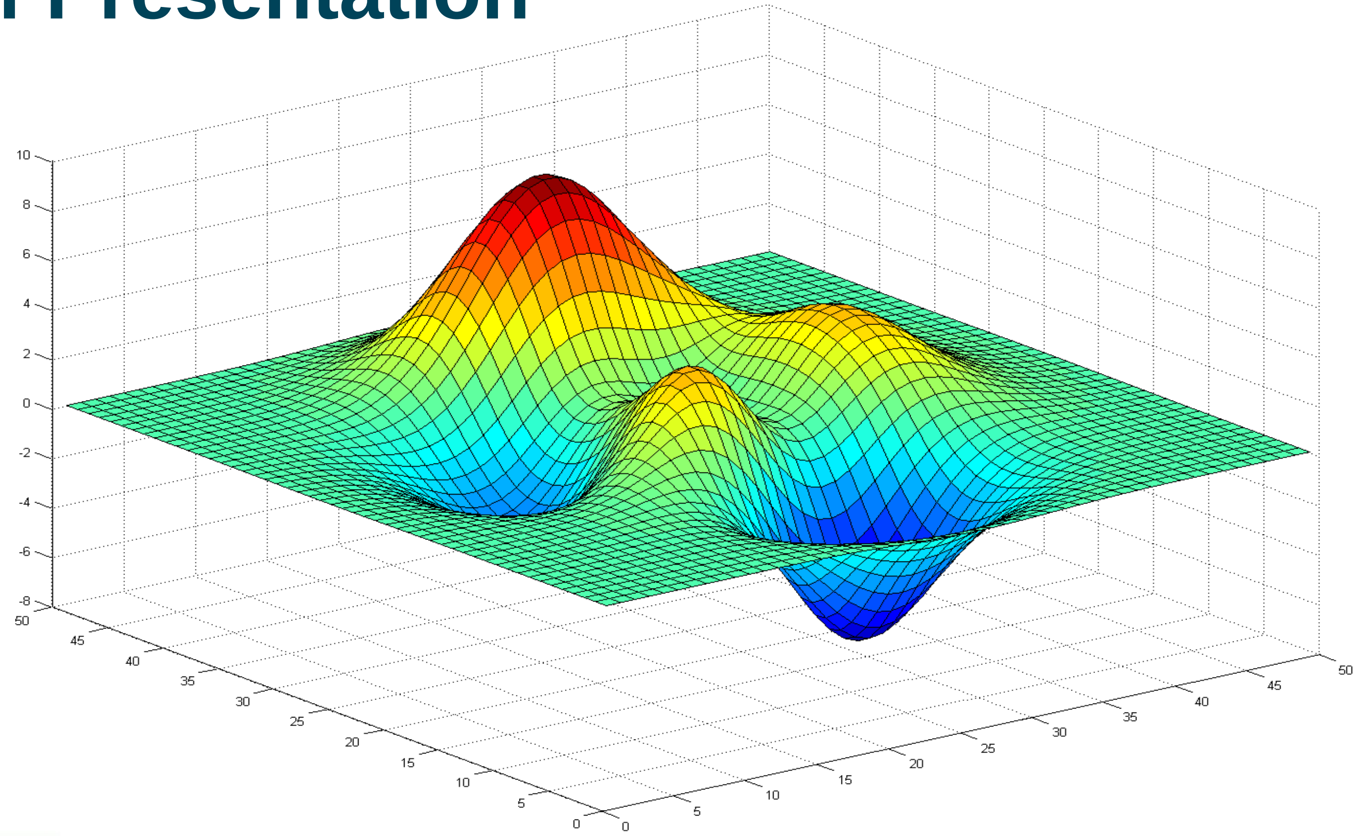
Griffith (2010) describes basic approaches

1. Multiple ARIMA Models
2. STAR Models
3. 3D Geostatistical Models
4. Panel data models with fixed and random effects
5. Spatial Filter Models

6. Kernel Methods and Machine Learning

Griffith, D. A. 2010. Modeling spatio-temporal relationships: retrospect and prospect. *Journal of Geographical Systems.* 12(2). pp . 111-123

Outline of Presentation

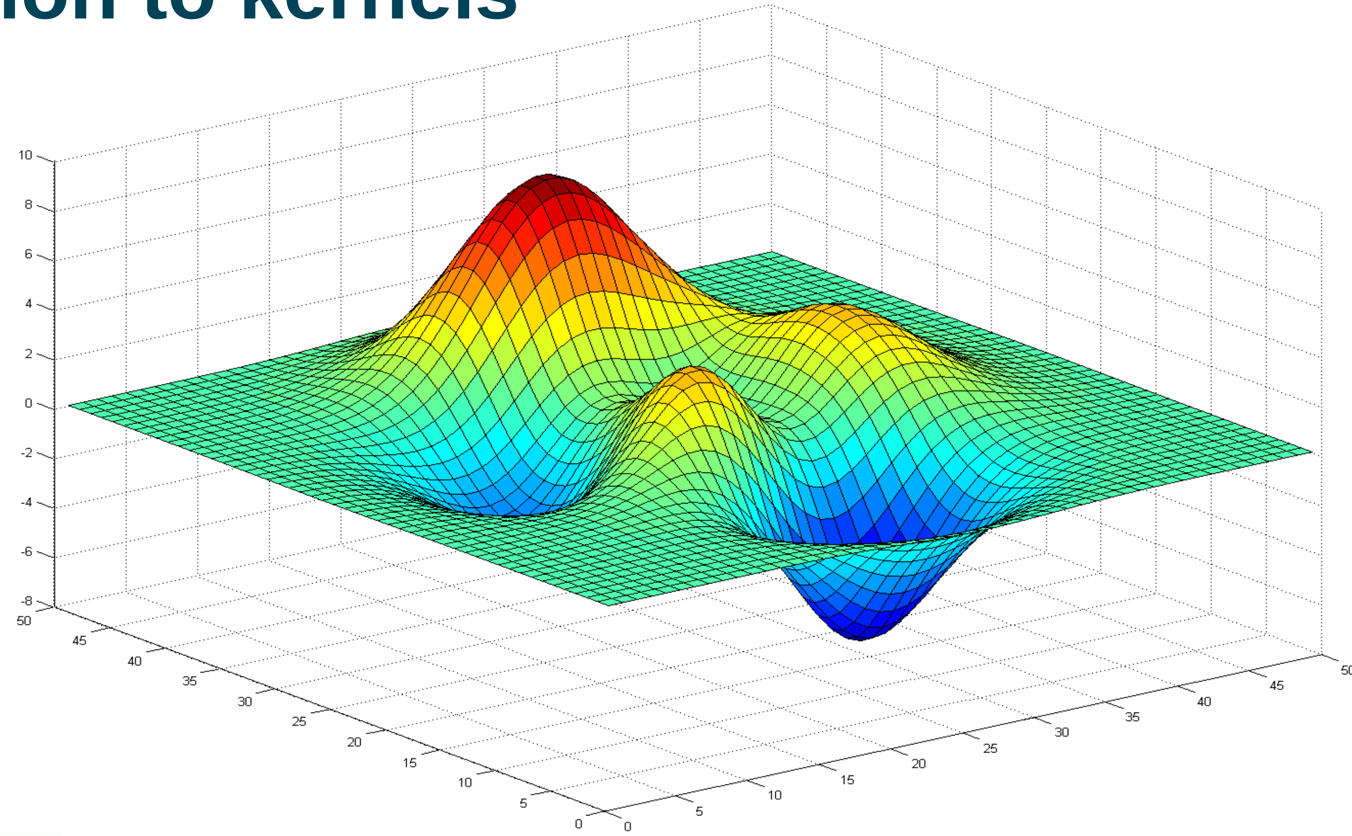


STANDARD
Spatio-Temporal Analysis of Network Data and Route Dynamics

Outline

- Introduction to kernels
  - What is a kernel?
  - Kernel Methods
  - Types of kernels
- Review of kernels in space-time analysis
  - Kernels in spatial analysis
  - Kernels in temporal analysis
  - Kernels in space-time analysis – **space time kernel**
- Application of STK to temperature prediction
- Conclusion and Discussion

Introduction to kernels



Introduction to kernels

- The problem
  - Machine Learning and Statistical Algorithms are well developed for the linear case.
  - Real world data is often complex and non-linear.
  - However, many of these algorithms depend on dot products between two vectors.

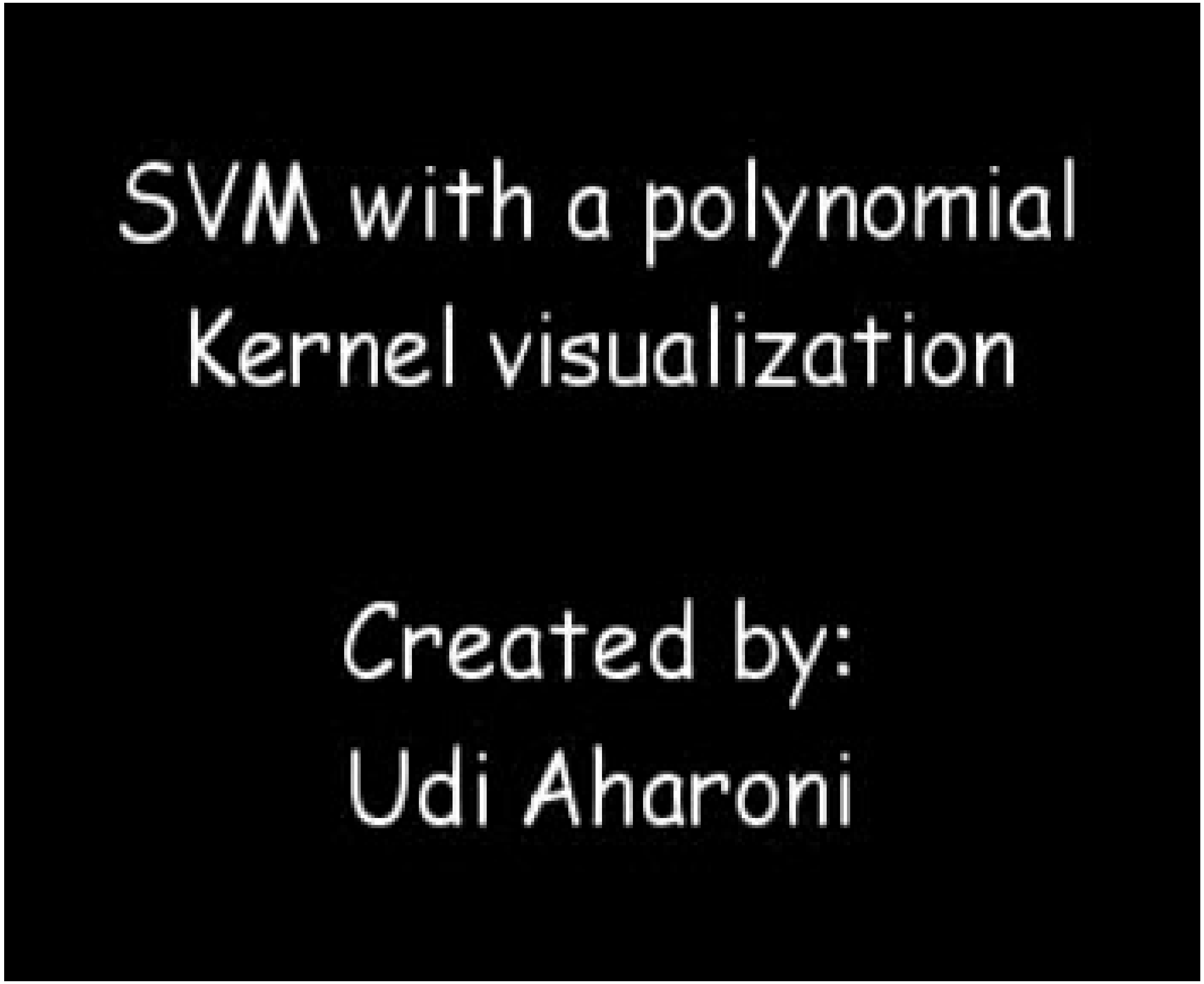
Introduction to kernels

- **Mercer’s Theorem**: Any continuous, symmetric, positive semi-definite function *K(x, y)* that can be expressed as a dot product in a high dimensional space is a **kernel**.

$$K(x,y) = \varphi(x).\varphi(y)$$

- Transforms linear algorithms.

- Data are mapped to high dimensional feature space where a linear solution can be found.



URL: http://www.youtube.com/watch?v=3liCbRZPrZA

Types of kernel – General purpose kernels

- **Linear:**

$$K(x,y) = X^T Y$$

- **Polynomial:**

$$K(x,y) = \left\{ ((X^T Y) + 1)^d \middle| d > 0 \right\}$$

- **Gaussian RBF**:

$$K(x,y) = \left\{ \exp\left(-\frac{||x-y||^2}{\sigma^2} \right) \middle| \sigma > 0 \right\}$$

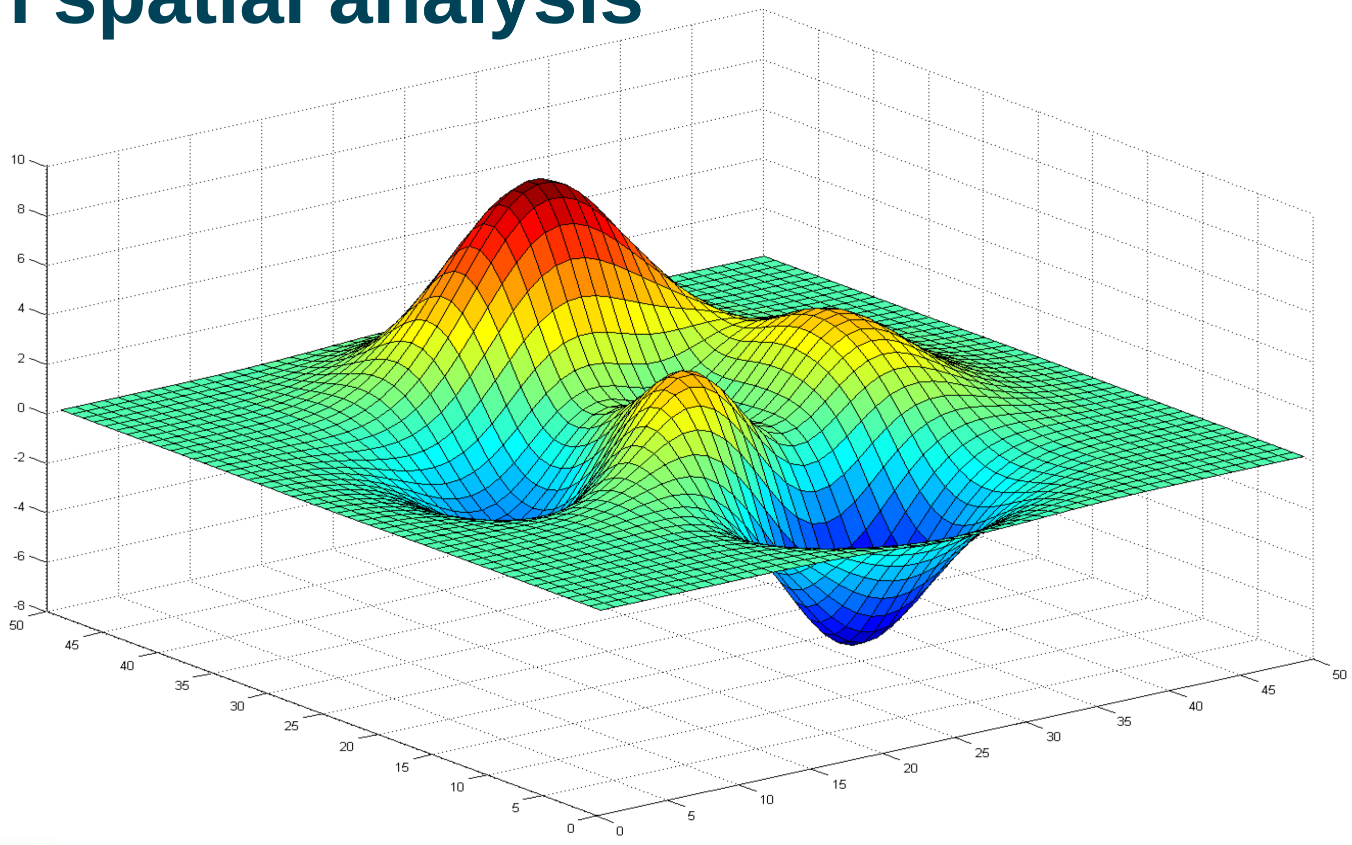
Custom Kernels

- Kernels can be designed to suit specific problem domains
  - Examples:
    - “Bag of words” kernel
    - Tree kernels
    - Graph kernels

Kernel Methods

- Support Vector Machines
- Kernel Principal/Independent Components Analysis
- Canonical Correlation Analysis
- Spectral Clustering
- Gaussian Processes
- Fisher's linear discriminant analysis
- Ridge regression
- Many More…

Kernels in spatial analysis

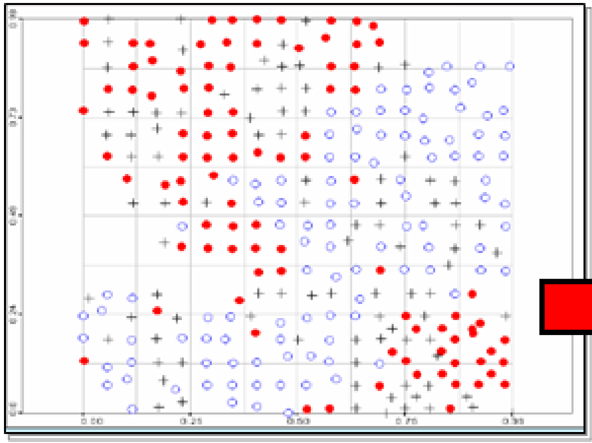


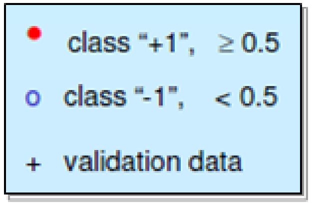
STANDARD
Spatio-Temporal Analysis of Network Data
and Route Dynamics

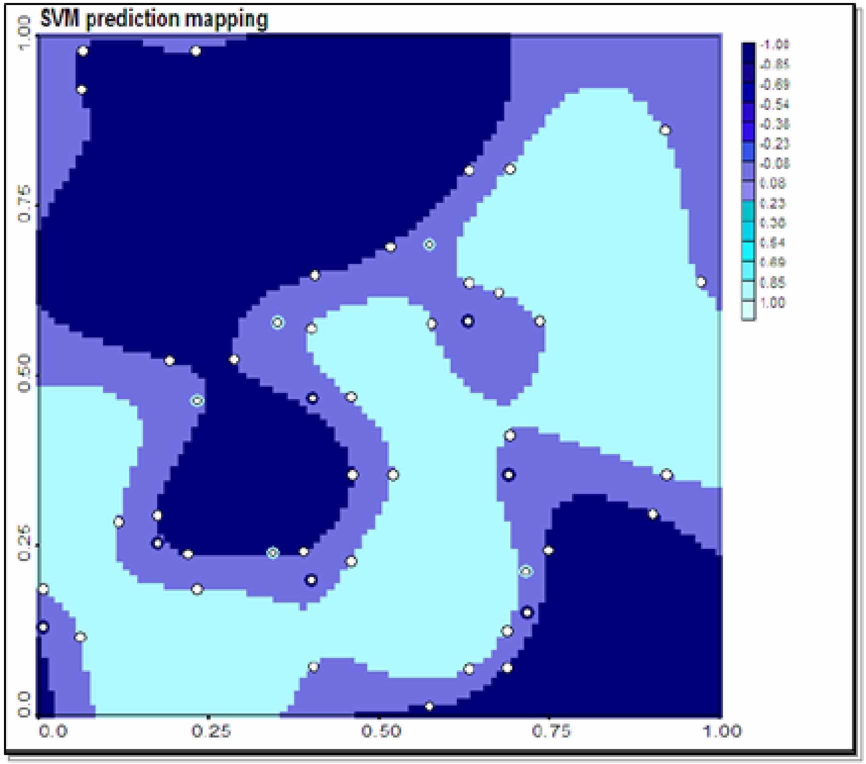
Kernels in spatial analysis– Spatial Interpolation

- Kanevski et al (2007) use SVM with Gaussian kernels for reservoir porosity mapping.
- Kernel parameter σ (bandwidth) dictates spatial influence of training samples.

Optimal Solution



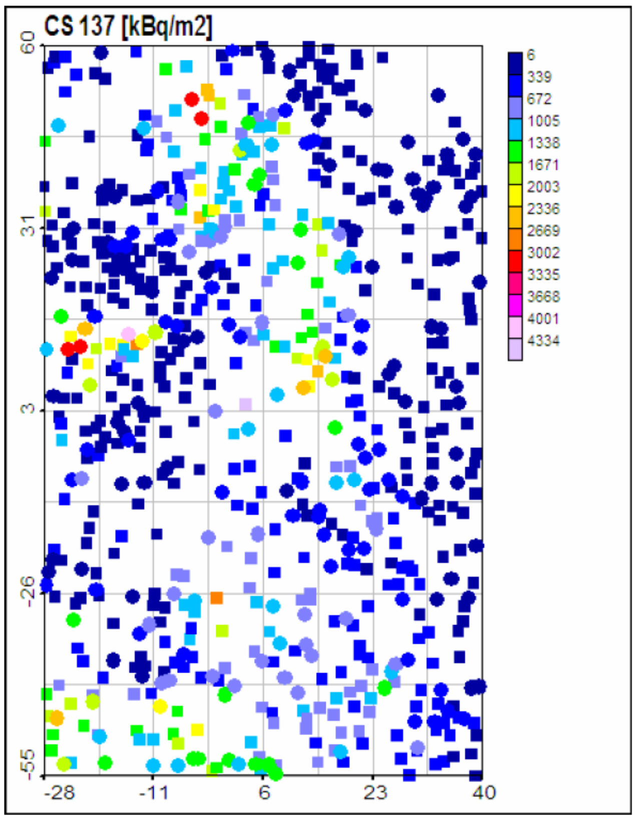




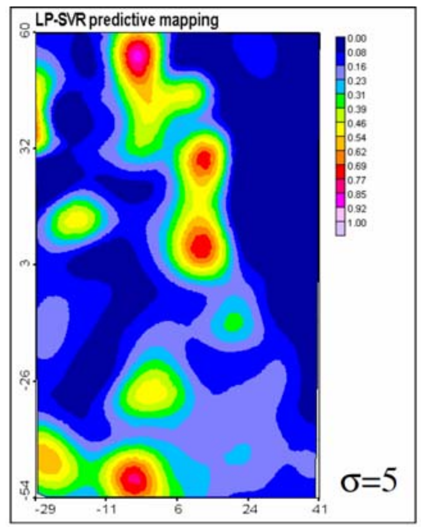
Kanevski M., Pozdnoukhov A., and Timonin V. Machine Learning Algorithms for Environmental Spatial Data. Theory and Case Studies. EPFL Press, Lausanne, 2007. A book with software on machine learning algorithms, 450 pages.

Kernels in spatial analysis– Spatial Regression

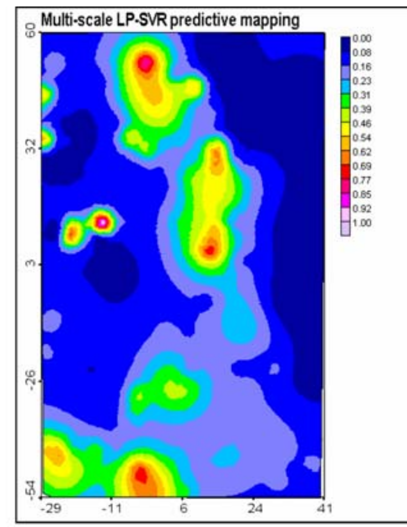
- What about local variations in data?
- Custom multi-scale kernels can be employed to improve results.



Input data



Single-scaled prediction
Testing RMSE 0.14

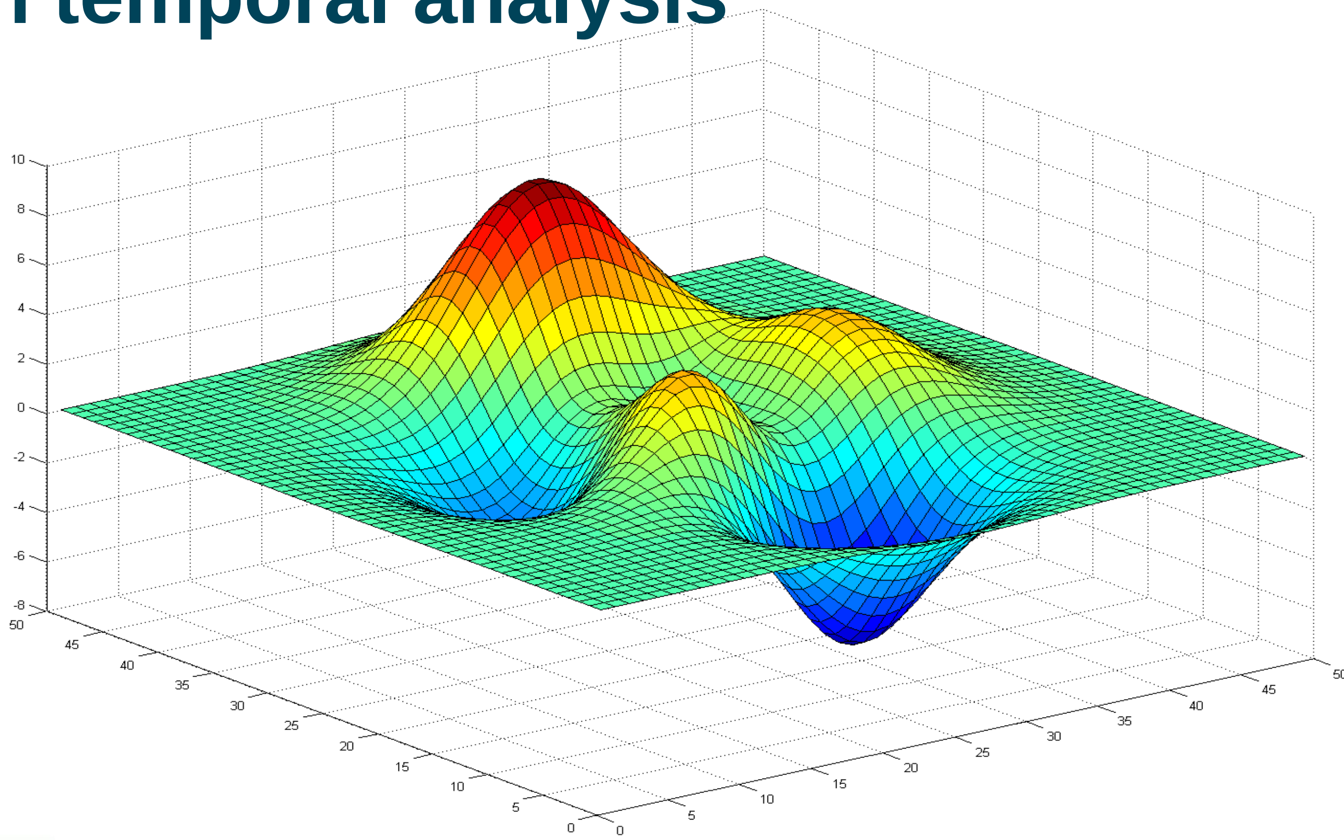


Double-scaled prediction
Testing RMSE 0.12

The multi-scale kernel captures large and small scale variations

Pozdnoukhov, A. And Kanevski, M.(2007). Multi-scale Support Vector Algorithms for Hot Spot Detection and Modelling. *Stochastic Environmental Research and Risk Assessment.*22(5), pp.647-660

Kernels in temporal analysis



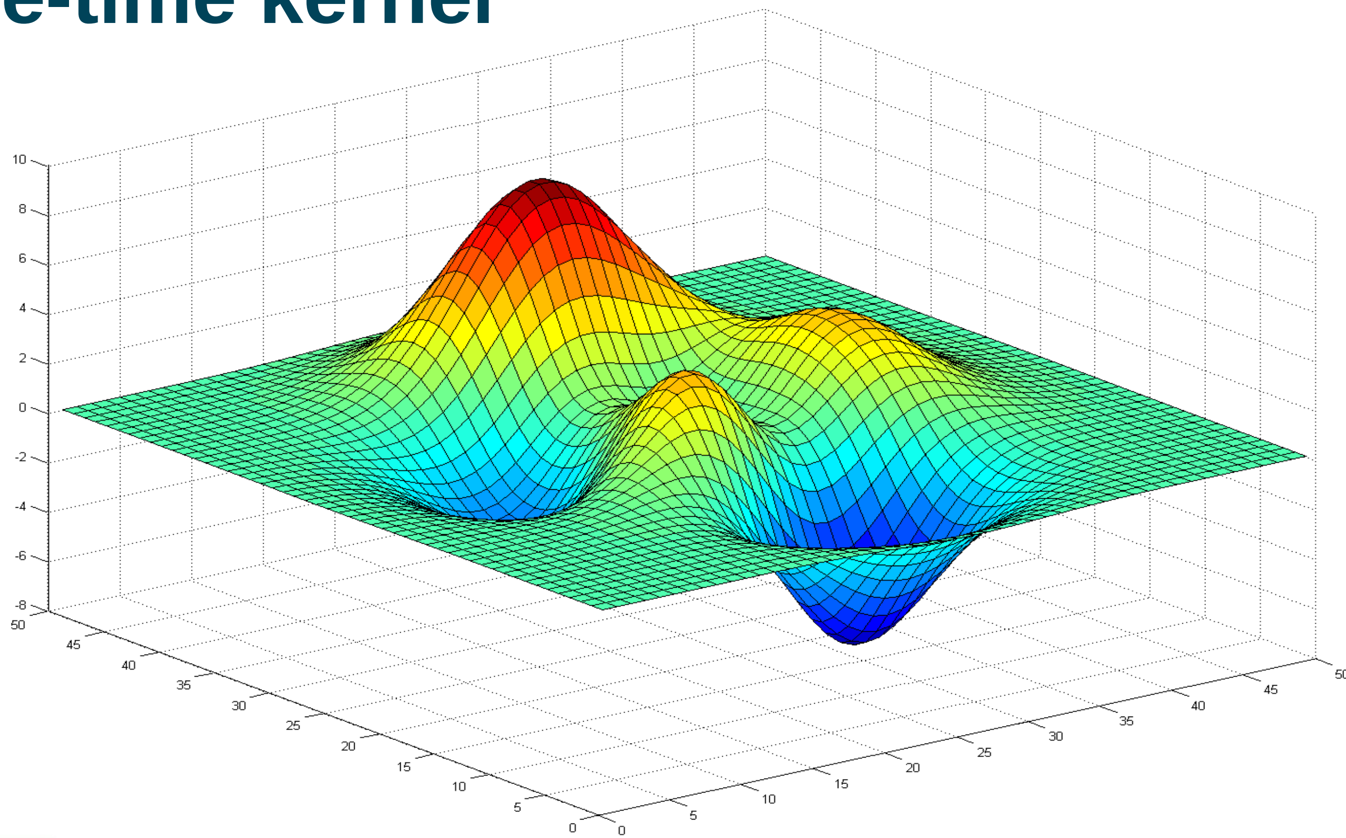
STANDARD
Spatio-Temporal Analysis of Network Data and Route Dynamics

Kernels in temporal analysis

- Kernel methods are widely used in time series forecasting
  - Travel time prediction
  - Financial time series
  - Environmental time series
- Rüeping (2001) reviews kernels in time series analysis
  - RBF kernel – good general purpose
  - Fourier kernel – periodic component
  - Polynomial kernel – non-periodic series

Ruping, S. 2001. SVM Kernels for Time Series Analysis. In Klinkenberg, R. , Ruping, S., Fick, A., Henze, N., Herzog, C., Molitor, R., and Schroder, O. (eds), *LLWA 01 - Tagungsband der GI-Workshop-Woche Lernen - Lehren - Wissen - Adaptivitat*, pp.43-50, Dortmund, Germany.

The Space-time kernel



The Space-Time Kernel

Problem

- Can a kernel be designed to deal with space-time series data that is:
  - Non-stationary?
  - Heterogeneous?
  - Multi-scale?

The Space-Time Kernel

- We have seen that different kernels can be applied to different problems
  - Gaussian kernels for spatial problems
  - Fourier, Polynomial kernels for temporal applications

- According to the theory of convolution kernels (Haussler, 1999), the product of two kernels is also a kernel.

The Space-time kernel

Weighting parameter

Gaussian (RBF) kernel

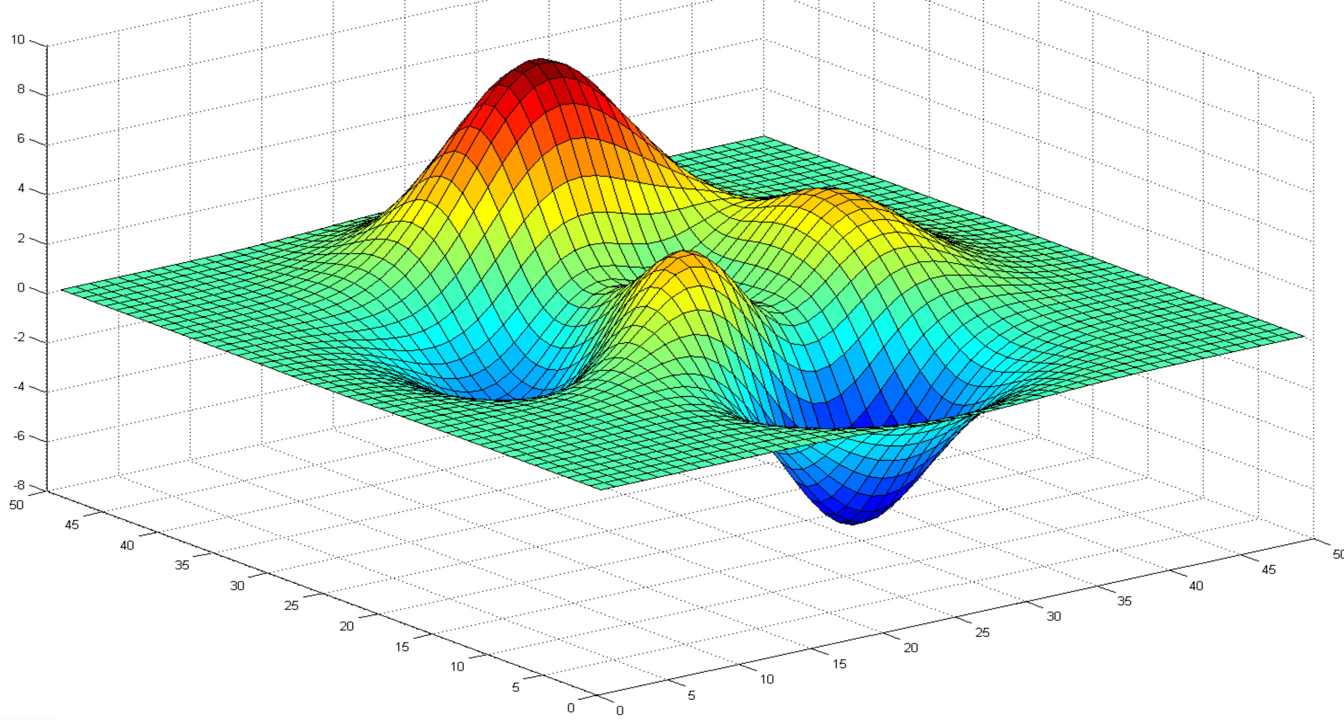
Fourier Kernel

$$K_{ST}(x,y) = \left\{ \left(\exp\left(-\frac{\|x-y\|^2}{\sigma^2} \right) \right) \cdot \left(\begin{array}{c} \alpha \cdot \dfrac{1-q^2}{2(1-2q\cos(x-y+q^2)} \\ +(1-\alpha) \cdot \left((x \cdot y)+1\right)^d \end{array} \right) \middle| \sigma > 0; 0 \le \alpha \le 1; d > 0; 0 < q < 1 \right\}$$

Polynomial Kernel

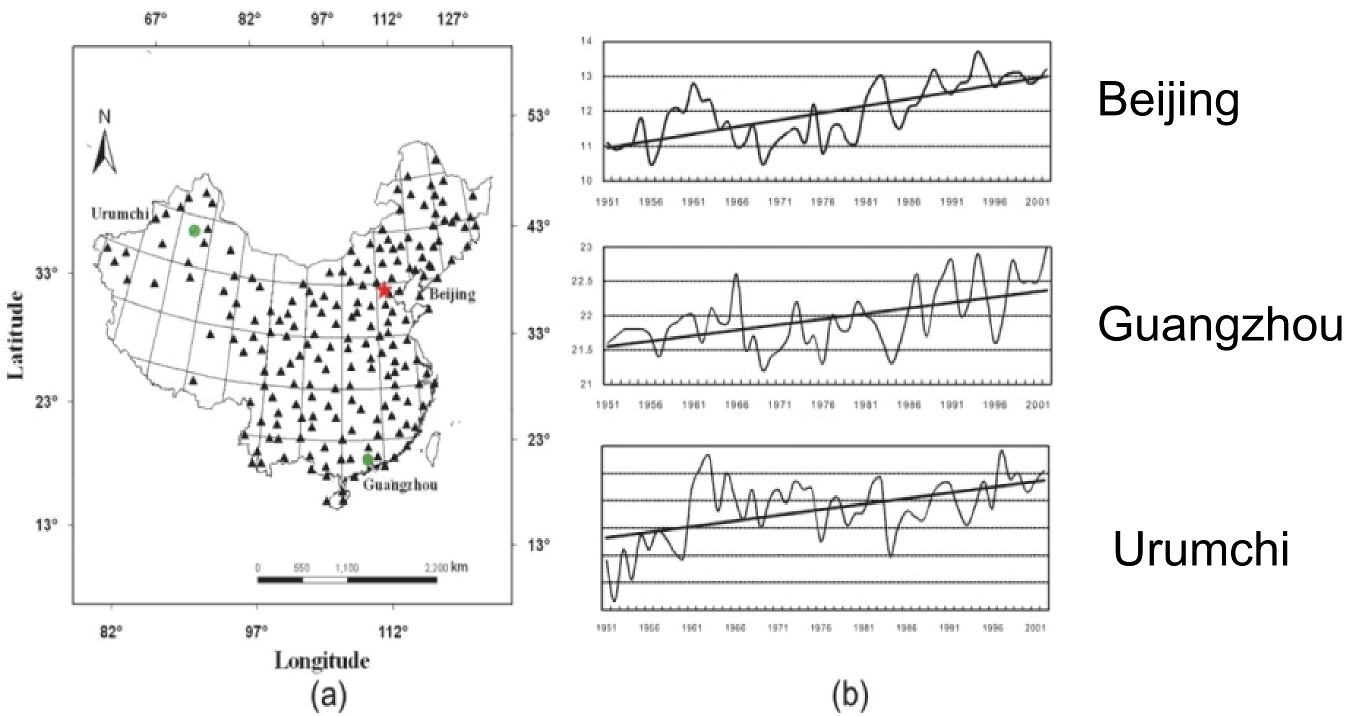
STANDARD
Spatio-Temporal Analysis of Network Data and Route Dynamics

Case Study – Annual average temperature prediction using Support Vector Regression with Space-time kernel



Case Study - Data

- Data – Georeferenced historical annual average temprature time series from 137 meteorological stations in China between 1951 and 1992.
- All stations are fitted, we consider results from three.



T. Cheng a, J.Q. Wang b, X. Li b, W. Zhang.2008. A HYBRID APPROACH TO MODEL NONSTATIONARY SPACE-TIME SERIES. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences.* Vol. XXXVII. Part B2. Beijing 2008

Case Study – Experimental Procedure

1. Variography is used to determine the spatial neighbourhood of each station.

2. Data is split into training (42 years) and testing (10 years) sets.

3. Each station is fitted using SVR based on its previous values and those of its spatial neighbours.

4. Once the model is trained, multi-step forecasting is used to predict the next ten years.

5. The results are compared with pure time series SVR and standard SVR.

STANDARD
Spatio-Temporal Analysis of Network Data and Route Dynamics

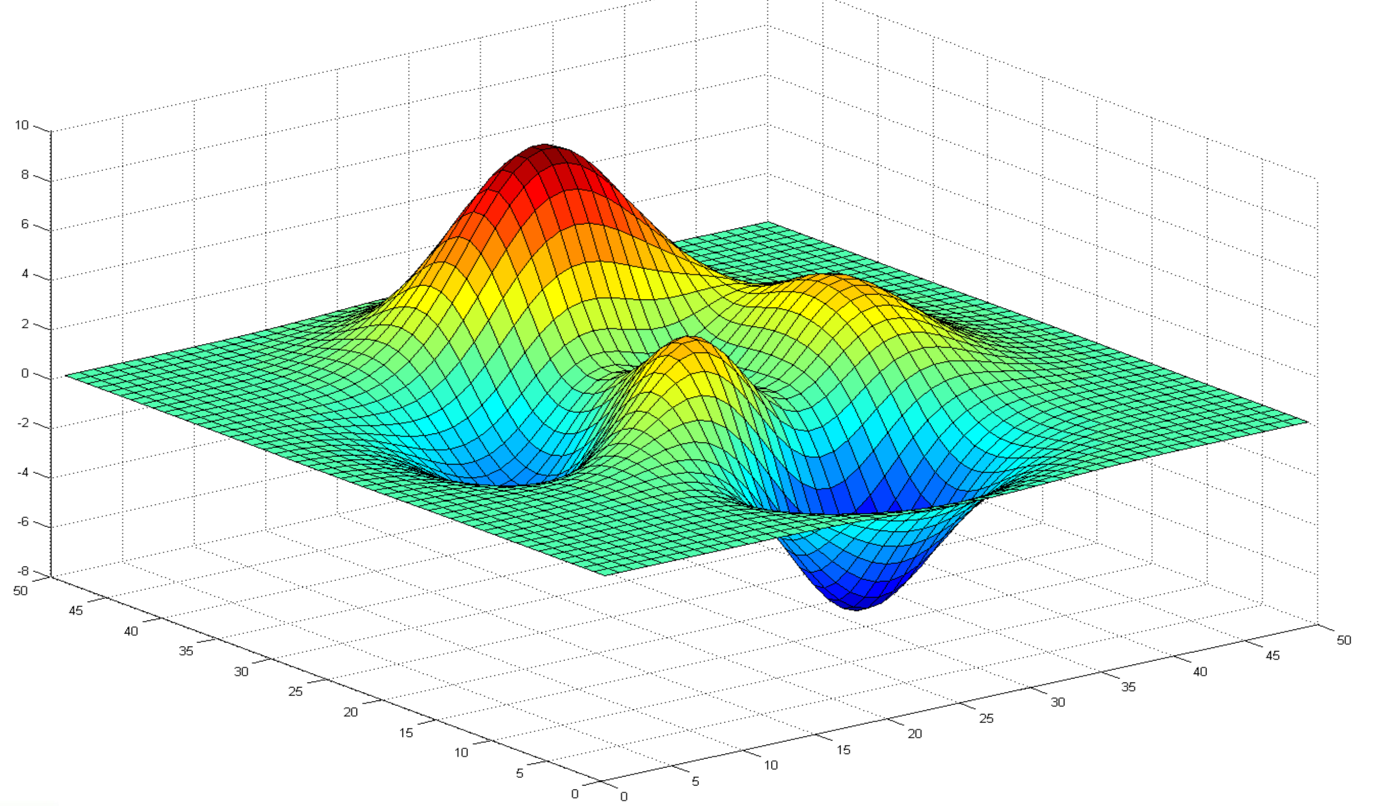
Case Study - Results

	Plain SVR	Time series SVR	SVR-STK
Fitted (1951-1992)			
RMSE			
Beijing	0.981	0.462	0.209
Guangzhou	0.910	0.314	0.084
Urumchi	1.173	0.853	0.306
Forecasting (1993-2002)			
RMSE			
Beijing	0.802	0.316	0.403
Guangzhou	0.813	0.418	0.387
Urumchi	0.837	0.551	0.541

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y_i - Y_i^*}{Y_i}\right|^2}$$

Table 1. Accuracy (RMSE) measures for three meteorological stations Beijing, Guangzhou and Urumchi in 52 years

Conclusions and future directions



STANDARD

Spatio-Temporal Analysis of Network Data and Route Dynamics

Conclusions and future directions

- The results are promising but…
- The space-time series is very short and has no periodic component.
- The model produced is still a global model.
- SVR is just one implementation of the space-time kernel.

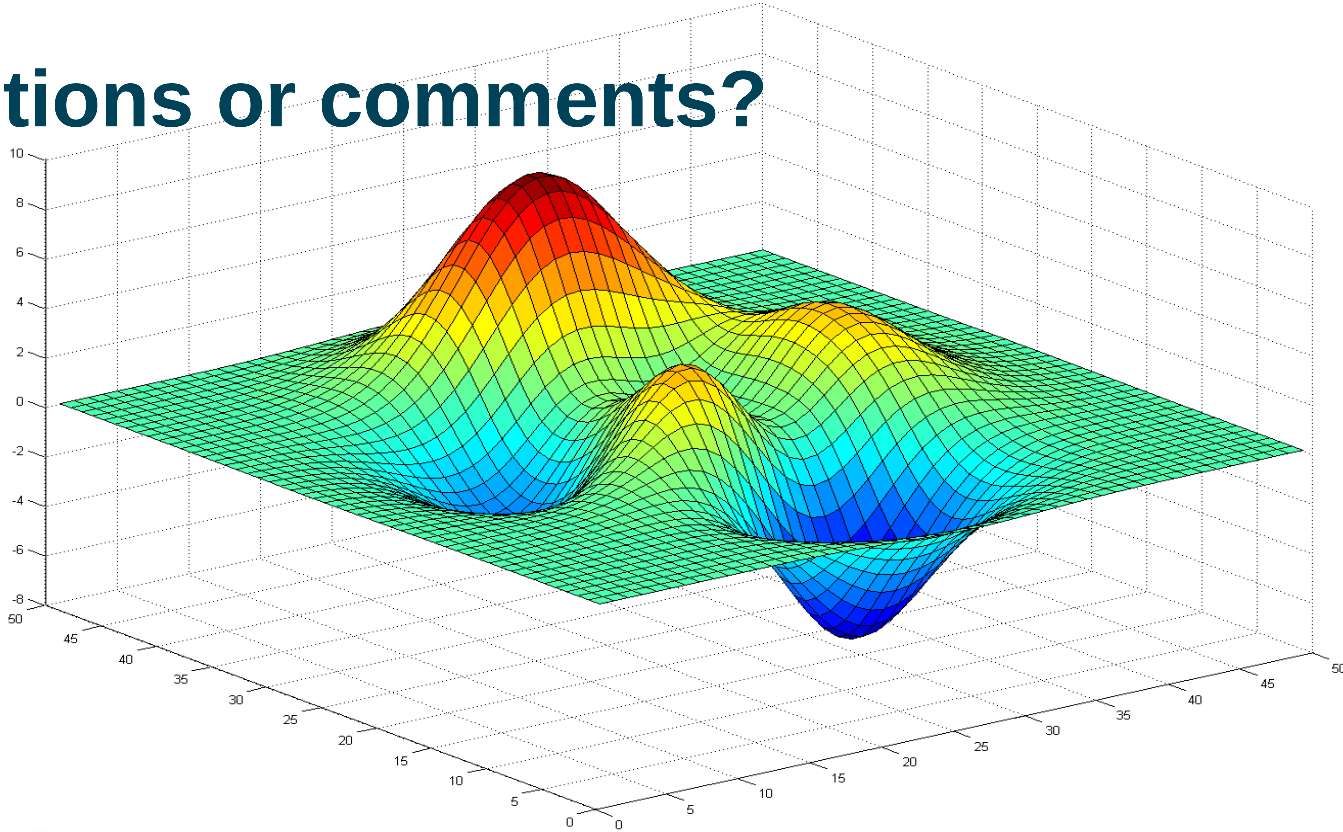
Some Major References

1. Aizerman, M., Braverman, E., and Rozonoer, L., 1964. Theoretical foundations of the potential function method in pattern recognition learning. Automation and Remote Control 25: 821-837.
2. Cheng, T., Wang, J.Q., 2009. Accommodating spatial associations in DRNN for space–time analysis. Computers, Environment and Urban Systems, 33(6), 409-418.
3. Haussler, D. 1999. Convolutional kernels on discrete structures. Technical ReportUCSC-CRL-99-10, Computer Science Dept., UC Santa Cruz.
4. Nussbaumer, H. J., 1982. Fast fourier transform and convolution algorithms. Springer, Berlin.
5. Pozdnoukhov, A., and Kanevski, M., 2008. GeoKernels: modeling of spatial data on GeoManifolds. In M. Verleysen, editor, ESANN 2008: European Symposium on Artificial Neural Networks – Advances in Computational Intelligence and Learning, Bruges, Belgium, 23-25, April.
6. Ralaivola, L., and d'Alché-Buc F., 2004. Dynamical modeling with kernels for nonlinear time series prediction. Advances in neural information processing systems, 16, pp. 129 - 136.
7. Vapnik, V., 1995. The nature of statistical learning theory. New York, Springer-Verlag.
8. Weston J., 1999 Extensions to the Support Vector Method. Ph.D. Thesis, Royal Holloway University of London

Thank you for listening!

Visit our website: www.standard.cege.ucl.ac.uk

Any questions or comments?

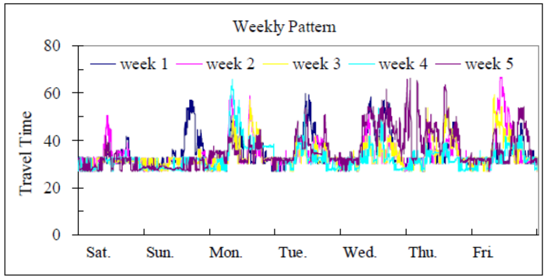


STANDARD
Spatio-Temporal Analysis of Network Data and Route Dynamics

Kernels in time analysis– time series forecasting

SVR for travel time prediction

- SVR is used to predict travel times over three distances on Taiwan highways



Input data – Travel time on 45km stretch of highway; 178 and 350km stretches are also used.

Experimental setup

- Time series SVR incorporates the principles of ***time series modelling***.
- 28 days are used for training, 7 for testing
- The ***one-step ahead method*** is used for training and testing of the data.
- An ***embedding dimension*** of 5 is chosen to train the model.

Wu, C.H., Wei, C.C, Su, D.C, Chang, M.H., Ho, J.M. (2003).Travel Time Prediction with Support Vector Regression. IEEE. http://www.csie.nuk.edu.tw/~wuch/publications/2003-itsc-svr.pdf

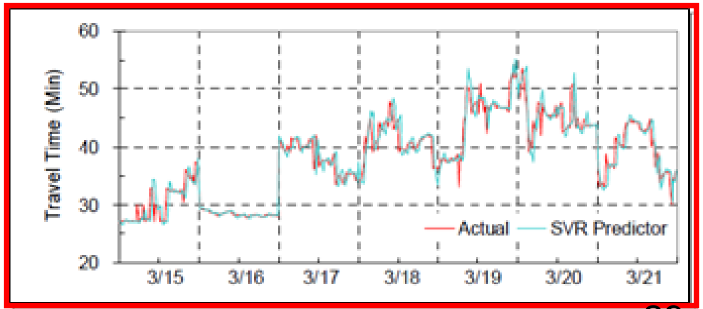
Performance Comparison

- A comparison with conventional techniques reveals significant improvements in forecasting ability.

RME	Current-time Predictor	Historical-mean Predictor	SVR Predictor
45 km	9.29%	12.52%	3.91%
161 km	3.88%	5.01%	1.71%
350 km	2.85%	2.56%	0.96%

RMSE	Current-time Predictor	Historical-mean Predictor	SVR Predictor
45 km	28.75%	16.20%	6.79%
161 km	9.98%	6.66%	2.57%
350 km	5.49%	3.42%	1.33%

Forecasting errors – SVR versus conventional techniques



Forecasting results – real versus predicted

Application of STK