IEEE Workshop on Statistical Signal Processing 2009

Grey-Box Radial Basis Function Modelling: The Art of Incorporating Prior Knowledge

Sheng Chen[†], **Xia Hong**[‡] **and Chris J. Harris**[†]

[†]School of Electronics and Computer Science
University of Southampton, Southampton SO17 1BJ, UK
E-mails: {sqc, cjh}@ecs.soton.ac.uk

[‡] School of Systems Engineering
University of Reading, Reading RG6 6AY, UK
E-mail: x.hong@reading.ac.uk

Motivations

- Standard radial basis function network is a **black-box** model
  - adopting black-box modelling approach is appropriate if no *a priori* information exists regarding underlying data generating mechanism
- If there are known **prior knowledge** concerning underlying process, they should be incorporated into model structure explicitly
- How to incorporate prior knowledge to form **grey-box** model is highly **problem dependent**, and is really an **art**
- Two types of prior information are considered
  - Underlying process exhibits known **symmetry** property
  - Underlying process obeys a set of **boundary value constraints**
- Existing learning algorithms can be applied to resulting grey-box models

RBF Data Modelling

- Give training data $\{\mathbf{x}(k), y(k)\}_{k=1}^{K}$ generated from **nonlinear** system

$$y(k) = f(\mathbf{x}(k)) + \epsilon(k)$$

  $f(\bullet)$ is **unknown**, and $\epsilon(k)$ represents observation noise

- **Radial basis function** model

$$\hat{y}(k) = \hat{f}(\mathbf{x}(k)) = \sum_{i=1}^{M} \theta_i p_i(k)$$

  with RBF **basis** $p_i(k) = \varphi\left(\|\mathbf{x}(k) - \mathbf{c}_i\|/\sigma\right)$ specified by RBF centre $\mathbf{c}_i$ and RBF variance σ^2

- **Black-box**, as every thing is learnt from data, which is inherently stochastic

- Efficient **orthogonal least squares** learning has been developed

Early Orthogonal Least Squares

- Orthogonal least squares methods and their application to non-linear system identification - S. Chen, S. A. Billings and W. Luo - International Journal of Control, 1989
  Google scholar citations: 467 ISI citations: 364 (July 2009)

- Orthogonal least squares learning algorithm for radial basis function networks - S. Chen, C. F. N. Cowan and P. M. Grant - IEEE Transactions on Neural Networks, 1991
  Google scholar citations: 1660 ISI citations: 1160 (July 2009)

- Simple and efficient, and capable of producing parsimonious models with good generalisation performance

- 20 year old, still popular with nonlinear data modelling practicians

More Recent Enhancements

- Recent enhancements to **orthogonal least squares** learning include
  - Local **regularisation** assisted OLS learning
  - Optimal **experiment design** enhanced OLS learning
  - OLS learning based on **leave-one-out** cross validation
- These **state-of-the-arts** bring further **benefits**
  - Enhance generalisation and sparseness
  - Improve model robustness and reduce parameter estimate variances
  - Select model terms by directly maximising generalisation capability
  - as well as fully automatic model selection

☞ In developing **grey-box** RBF models, these OLS statistical learning algorithms should readily be **applicable**

Symmetric RBF Network

- **Unknown** system $f(\bullet)$ possesses **odd symmetry** $f(-\mathbf{x}) = -f(\mathbf{x})$
  - ☞ e.g. from physics, underlying optimal discriminant function for BPSK digital signals has old symmetry
- **Radial basis function** model with **standard** node

$$p_i(k) = \varphi\left(\|\mathbf{x}(k) - \mathbf{c}_i\|/\sigma\right)$$

  - ☞ cannot guarantee to have odd symmetry
- **Symmetric RBF** model with symmetric RBF node

$$p_i(k) = \varphi\left(\|\mathbf{x}(k) - \mathbf{c}_i\|/\sigma\right) - \varphi\left(\|\mathbf{x}(k) + \mathbf{c}_i\|/\sigma\right)$$

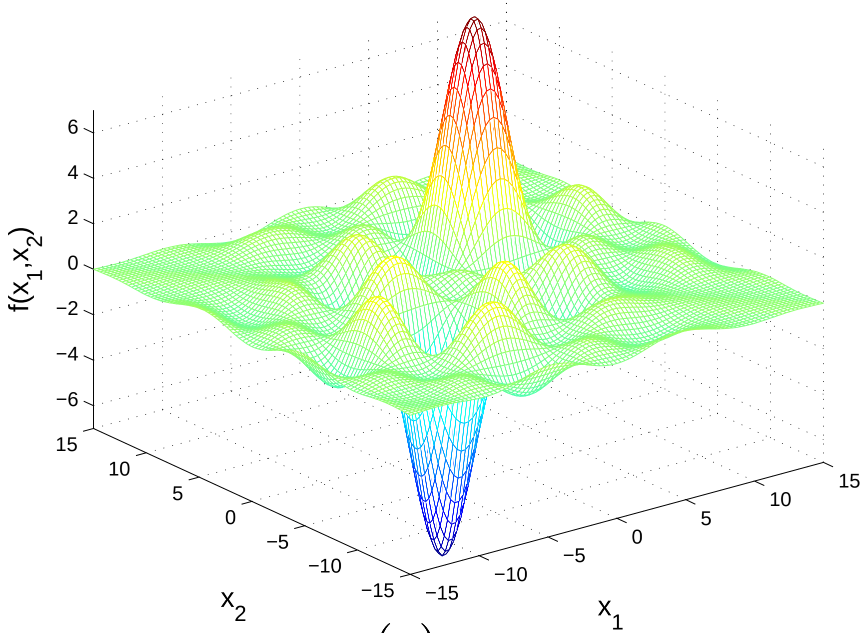
  - ☞ **guarantee** to obey same odd symmetry as underlying process
  - ☞ incorporate **prior information** naturally into model structure
  - ☞ all RBF learning methods **applicable** without any modification

Symmetric Function Modelling

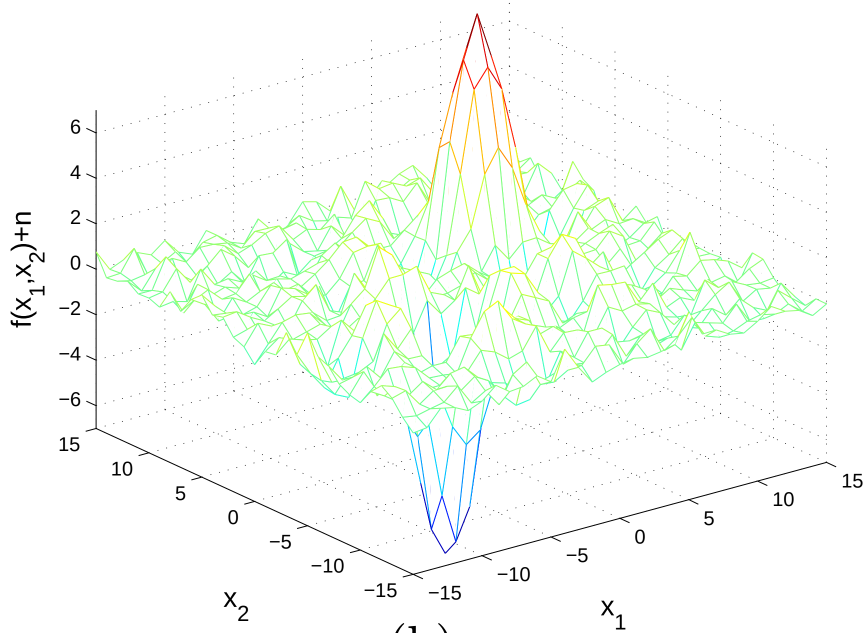
(a) Underlying function

$$f(x_1, x_2) = 10\left(\frac{\sin(x_1 - 5)\sin(x_2 - 5)}{(x_1 - 5)(x_2 - 5)} - \frac{\sin(x_1 + 5)\sin(x_2 + 5)}{(x_1 + 5)(x_2 + 5)}\right)$$

shown on the grid of 90601 points, and (b) 961 noisy training data points $y = f(x_1, x_2) + \epsilon$, where ϵ is Gaussian noise of zero mean and variance 0.16



(a)



(b)

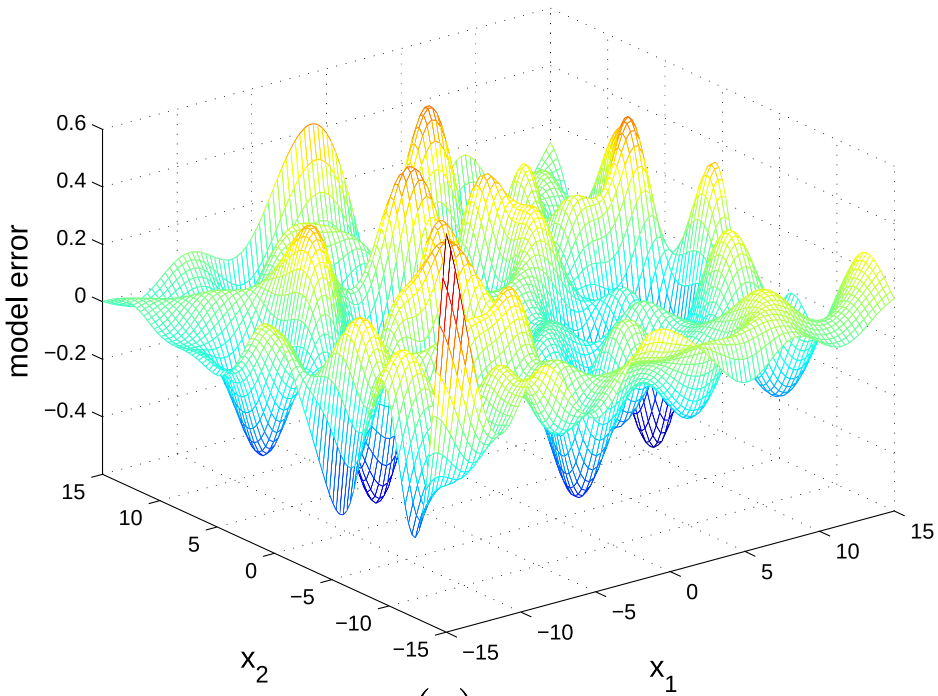
Symmetric Modelling Results

- Every training data used as a RBF centre with $M = K = 961$, RBF variance $\sigma^2 = 8.0$ was determined separately using cross validation
- Local regularisation assisted OLS algorithm with LOO MSE was used to automatically select sparse RBF / SRBF model
- **Mean square error** $\text{MSE} = E[(y - \hat{y})^2]$ was calculated over noisy training set and a separate noisy test set
- **Mean modelling error** was defined as $\text{MME} = E[(f(x_1, x_2) - \hat{y})^2]$ over grid of 90601 points noise-free $f(x_1, x_2)$

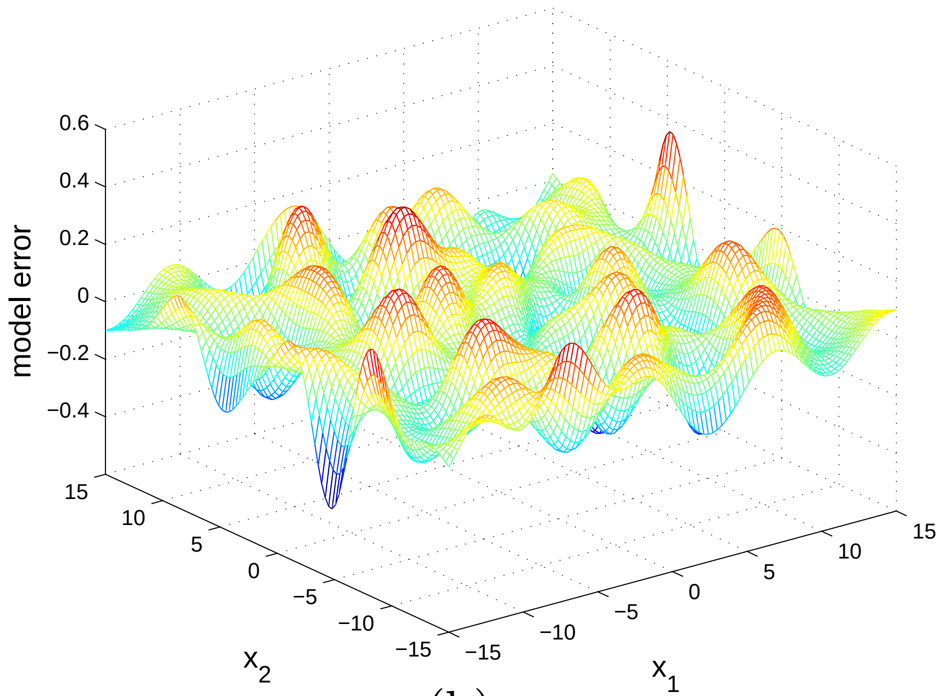
	model size	training MSE	test MSE	MME
RBF	**105**	**0.1543**	0.2047	**0.0294**
SRBF	**68**	**0.1566**	**0.1839**	**0.0093**

Symmetric Modelling (continue)

(a) modelling error $f(x_1, x_2) - \hat{f}(x_1, x_2)$ of standard RBF model, and (b) modelling error $f(x_1, x_2) - \hat{f}(x_1, x_2)$ of symmetric RBF model



(a)



(b)

Results Analysis

- By incorporating **prior information**, SRBF model offers significantly better **generalisation** performance
  - **Mean modelling error** is three times smaller than standard RBF
- OLS algorithm selecting $M^{'}$ model terms from K-term candidate set has **complexity**

$$C = \left(M^{'} + 1\right) \times K \times \mathcal{O}(K)$$

  - For SRBF, $M^{'} = 68$, while for standard RBF, $M^{'} = 105$ in this case
  - Thus, complexity of SRBF **model construction** is about half of complexity for constructing standard RBF model
- Computational requirements of a symmetric node is twice standard one
  - **Prediction** complexity of two models are similar

Boundary Value Constraints

- Underlying system satisfies a set of **boundary value constraints**

$$f(\mathbf{x}_j) = d_j, \ 1 \le j \le L$$

  - $\mathbf{x}_j$ and d_j, $1 \le j \le L$, are **known**
  - These BVCs may represent the fact that at some critical regions, there is a **complete knowledge** about system

- Any identified model $\hat{f}$ is required to **strictly** meet these BVCs

$$\hat{f}(\mathbf{x}_j) = d_j, \ 1 \le j \le L$$

  - RBF model with **standard** node $p_i(k) = \varphi\left(\|\mathbf{x}(k) - \mathbf{c}_i\|/\sigma\right)$ cannot meet these BVCs

- Using these BVCs as **constraints** dramatically complicates learning

  - Efficient state-of-the-art learning methods cannot be applied directly

Boundary Value Constraint RBF Network

- **Boundary value constraint**-RBF model takes the form

$$\hat{y}(k) = \hat{f}(\mathbf{x}(k)) = \sum_{i=1}^{M} p_i(\mathbf{x}(k))\theta_i + g(\mathbf{x}(k))$$

- with novel RBF **node structure**

$$p_i(\mathbf{x}) = h(\mathbf{x})\varphi(\|\mathbf{x} - \mathbf{c}_i\|/\sigma)$$

- **Geometric mean** of data sample $\mathbf{x}$ to BVCs $\mathbf{x}_j$, $1 \leq j \leq L$

$$h(\mathbf{x}) = \sqrt[L]{\prod_{j=1}^{L} \|\mathbf{x} - \mathbf{x}_j\|}$$

- Since $h(\mathbf{x}_j) = 0$ at any boundary point $\mathbf{x}_j$, node $p_i(\mathbf{x})$ has property of **zero forcing** at any $\mathbf{x}_j$

BVC-RBF (continue)

- **Offset function**

$$g(\mathbf{x}) = \sum_{j=1}^{L} \alpha_j e^{-\frac{\|\mathbf{x}-\mathbf{x}_j\|^2}{\tau}}$$

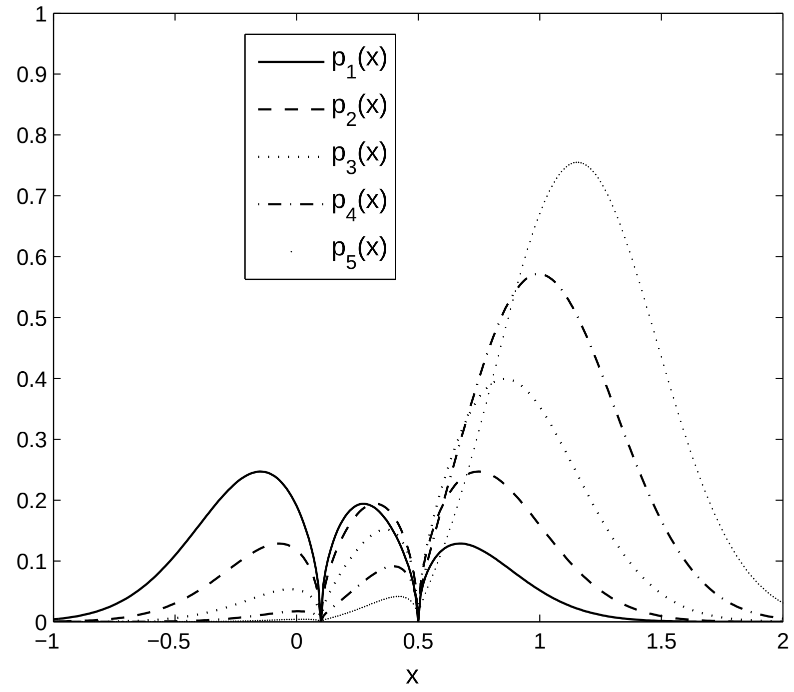
- with τ being a positive scalar, $\boldsymbol{\alpha} = [\alpha_1\ \alpha_2 \cdots \alpha_L]^T$ is obtained by solving linear equations $g(\mathbf{x}_j) = d_j$, $1 \le j \le L$, $\boldsymbol{\alpha} = \mathbf{G}^{-1}\mathbf{d}$, where $\mathbf{d} = [d_1\ d_2 \cdots d_L]^T$ and

$$\mathbf{G} = \begin{bmatrix} 1 & e^{-\frac{\|\mathbf{x}_1-\mathbf{x}_2\|^2}{\tau}} & \cdots & e^{-\frac{\|\mathbf{x}_1-\mathbf{x}_L\|^2}{\tau}} \\ e^{-\frac{\|\mathbf{x}_2-\mathbf{x}_1\|^2}{\tau}} & 1 & \ddots & e^{-\frac{\|\mathbf{x}_2-\mathbf{x}_L\|^2}{\tau}} \\ \vdots & \ddots & \ddots & \vdots \\ e^{-\frac{\|\mathbf{x}_L-\mathbf{x}_1\|^2}{\tau}} & e^{-\frac{\|\mathbf{x}_L-\mathbf{x}_2\|^2}{\tau}} & \cdots & 1 \end{bmatrix}$$

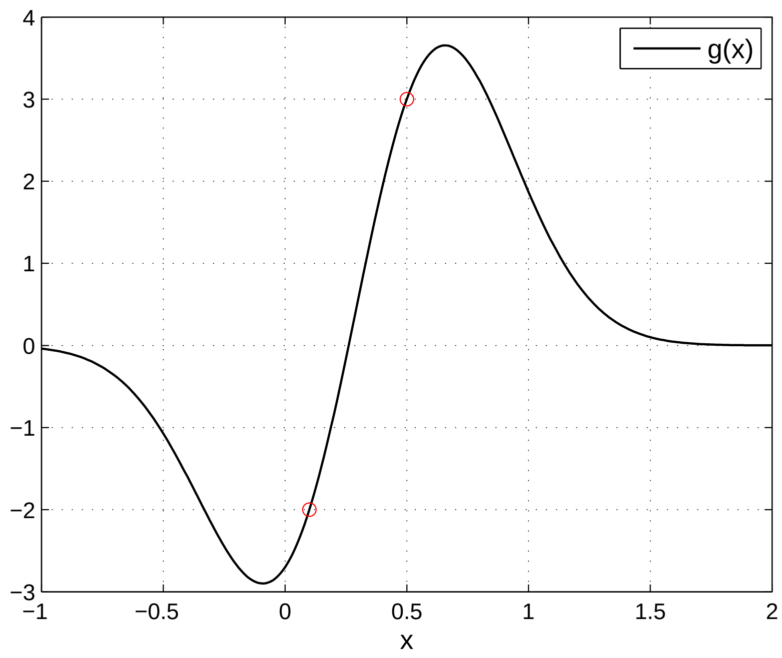
- Offset function $g(\mathbf{x})$ **passes** all predetermined **boundary values** $f(\mathbf{x}_j) = g(\mathbf{x}_j) = d_j$, $1 \le j \le L$, and it is completely determined by BVCs but does not contain any adjustable parameters dependent on D_K.

BVC-RBF Illustration

- One-dimensional function $f(x)$ with two BVCs: $f(0.1) = -2$, $f(0.5) = 3$
- Five RBFs with zero forcing at two boundary points (a), and offset passing function $g(x)$ (b)



(a)



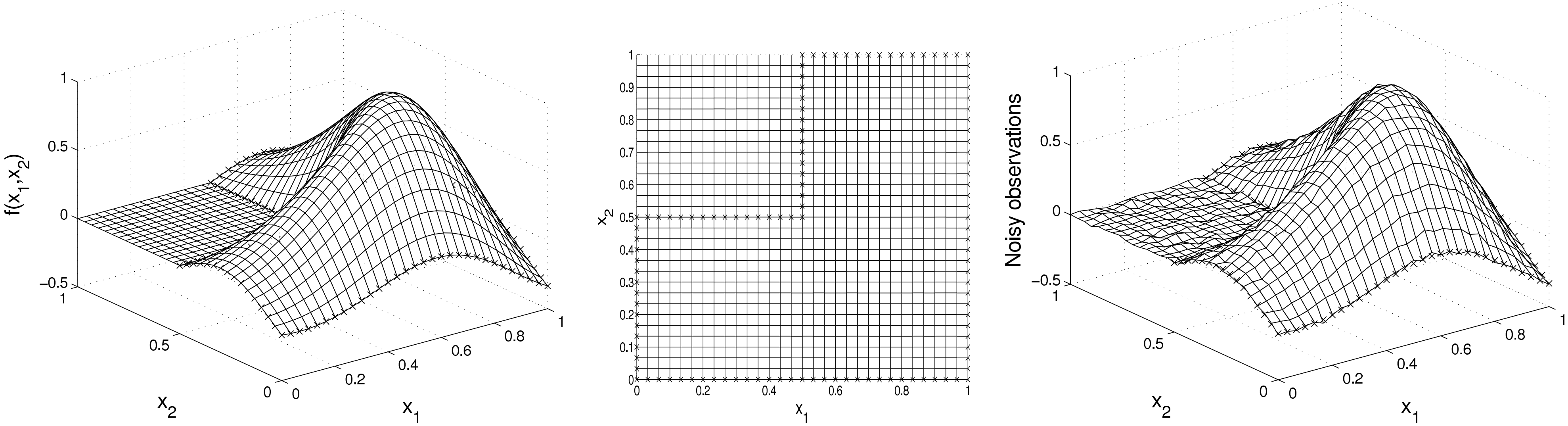
(b)

BVC-Function Modelling

(a) Underlying function $f(x_1, x_2)$ shown on grid of 961 points

(b) $L = 120$ BVCs given by coordinates marked as cross points

(c) 961 noisy training points, with Gaussian noise of zero mean and variance 0.01^2



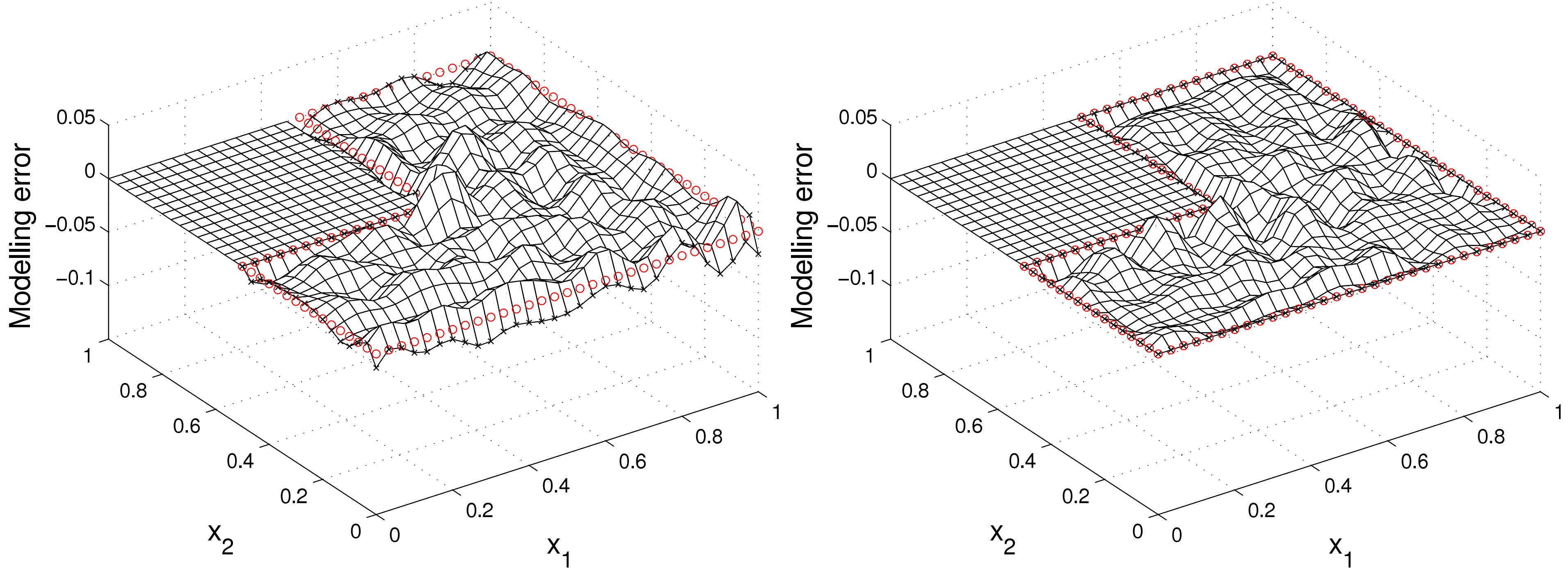
(a) (b) (c)

- OLS algorithm based on training MSE and D-optimality was used to automatically identify standard RBF and BVC-RBF models
- RBF variance $\sigma^2 = 0.01$ was determined by cross validation and $\tau = 0.04$

BVC-Function Modelling Results

	model size	training MSE (inside D_K)	MME (inside boundary)	MME (on boundary)
RBF	91	1.6894×10^{-4}	1.0229×10^{-4}	2.1249×10^{-4}
BVC-RBF	68	1.0736×10^{-4}	4.3787×10^{-5}	7.2598×10^{-11}

(a) Modelling error $f(x_1, x_2) - \hat{y}$ of standard RBF (a) and BVC-RBF (b)



(a) (b)

Summary

- Discuss **art** of using **prior knowledge** to form **grey-box** RBF model
- Two types of prior information have been considered
  - Underlying process exhibits known **symmetry** property
  - Underlying process obeys a set of **boundary value constraints**
- Novel **SRBF** model and **BVC-RBF** model have been proposed
  - Existing efficient state-of-the-arts RBF learning methods **readily applicable** without any modification
  - Result in **better generalisation** performance, **smaller model** size and **reduced complexity** in model construction