# Uncertainty in Surrogate Models

July 14, 2016

Handing Wang
handing.wang@surrey.ac.uk

Department of Computer Science
University of Surrey
UK

# Outline



Infill Sampling Criteria

Benefits of Uncertain Solutions to Surrogate Models

Uncertainty Measurement

Uncertainty in Ensemble Surrogates

# On-line SAEA



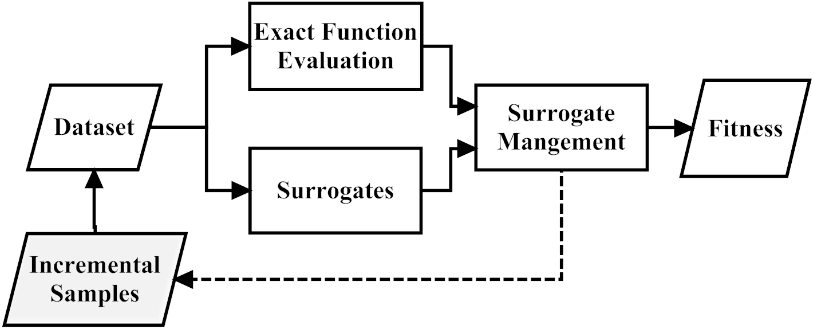



Figure: Surrogate management in on-line SAEA.

# Infill Sampling Criteria



## Three types

- The optimum of the surrogate model can improve the exploitation ability of the surrogate model.
- The most uncertain points of the surrogate model can help improve the exploration ability by searching the unexplored regions.
- Combination of the former two types, ExI and LCB for instance.

# An Example



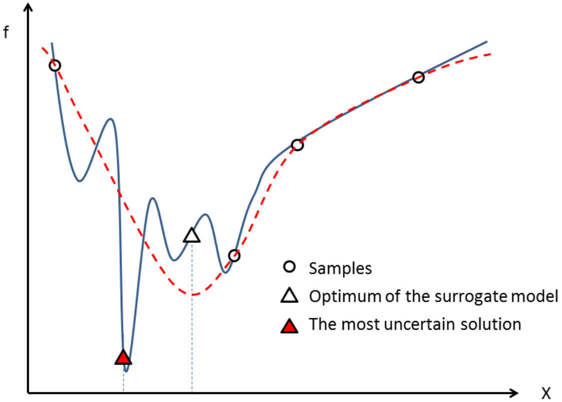



Figure: Surrogate model based on initial samples.

# An Example: Update by the optimum



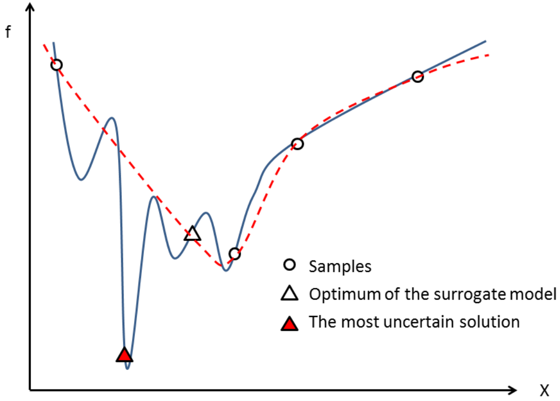



Figure: Surrogate model based on initial samples and the optimum of the surrogate model.

# An Example: Update by the most uncertain solution



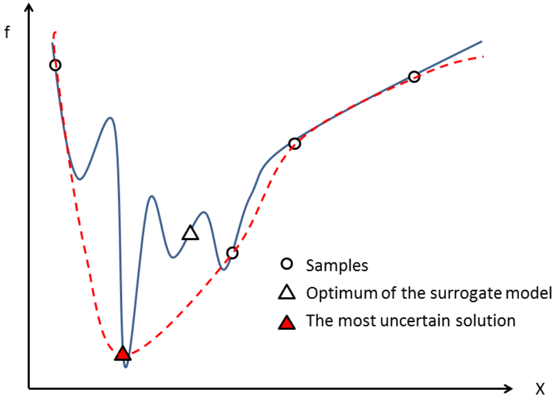



Figure: Surrogate model based on initial samples and the most uncertain solution.

# Benefits of Uncertain Solutions



## Error analysis

- $y = f(\mathbf{x})$
- $\hat{y} = \hat{f}(\mathbf{x})$

# Benefits of Uncertain Solutions



## Error analysis

- $y = f(\mathbf{x})$
- $\hat{y} = \hat{f}(\mathbf{x})$
- $f(\mathbf{x}) = \hat{f}(\mathbf{x}) + \varepsilon(\mathbf{x})$

# Benefits of Uncertain Solutions



## Error analysis

- $y = f(\mathbf{x})$
- $\hat{y} = \hat{f}(\mathbf{x})$
- $f(\mathbf{x}) = \hat{f}(\mathbf{x}) + \varepsilon(\mathbf{x})$
- $f(\mathbf{x}) - f(\mathbf{x}^*) > \varepsilon(\mathbf{x}) - \varepsilon(\mathbf{x}^*)$

# Benefits of Uncertain Solutions



## Error analysis

- $y = f(\mathbf{x})$
- $\hat{y} = \hat{f}(\mathbf{x})$
- $f(\mathbf{x}) = \hat{f}(\mathbf{x}) + \varepsilon(\mathbf{x})$
- $f(\mathbf{x}) - f(\mathbf{x}^*) > \varepsilon(\mathbf{x}) - \varepsilon(\mathbf{x}^*)$
- $\varepsilon(\mathbf{x}) - \varepsilon(\mathbf{x}^*)$ of any solution $\mathbf{x}$ to the optimum $\mathbf{x}^*$ should not be larger than the differences of their exact FE values, or the searching on $\hat{f}(\mathbf{x})$ will be led to other solutions rather than $\mathbf{x}^*$.

# Benefits of Uncertain Solutions



## Error analysis

- $y = f(\mathbf{x})$
- $\hat{y} = \hat{f}(\mathbf{x})$
- $f(\mathbf{x}) = \hat{f}(\mathbf{x}) + \varepsilon(\mathbf{x})$
- $f(\mathbf{x}) - f(\mathbf{x}^*) > \varepsilon(\mathbf{x}) - \varepsilon(\mathbf{x}^*)$
- $\varepsilon(\mathbf{x}) - \varepsilon(\mathbf{x}^*)$ of any solution $\mathbf{x}$ to the optimum $\mathbf{x}^*$ should not be larger than the differences of their exact FE values, or the searching on $\hat{f}(\mathbf{x})$ will be led to other solutions rather than $\mathbf{x}^*$.

## Benefits

- Exploration for $\mathbf{x}^*$.
- Reduction for $\varepsilon(\mathbf{x})$.

# Uncertainty Measurement



## Sample Uncertainty

- Variance in the Kriging model: It can provide a confidence level of the predictions but lose performance when the decision space is of high dimensions.
- Distance-based metrics: They evaluate the distance to the existing sample points but have problems on the computational cost and indexing peroformance in the high-dimensional space.

## Model Uncertainty

- Ensemble: The differences of the errors of different models, so multiple diverse models are required.

# Weakness of the Kriging model

The average std of the variance of 10000 random samples on the Kriging model built by 11$d$ samples is shown below.

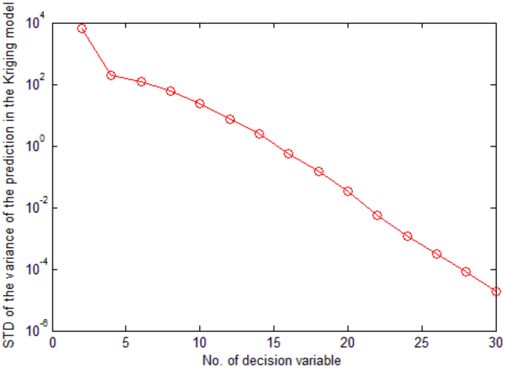



Figure: STD of the variance of the prediction in the Kriging model on the Rastrigin problem.

# Weakness of the Kriging model



## Disadvantages of the Kriging model

- Computational complexity
- Less-effective confidence level of the predictions for high-dimensional problems

## Reason

- Calculation of hyper-parameter and correlation
- Correlation function loses performance in the high-dimensional space

# An Alternative Option for Uncertainty Measurement



Using an ensemble consisting of a large number of computationally very efficient models that might provide useful uncertainty information similar to the Kriging model.

- Train data: 11*d* samples
- Test data: 1000 samples
- Model: Kriging and random forest (50 CARTs)

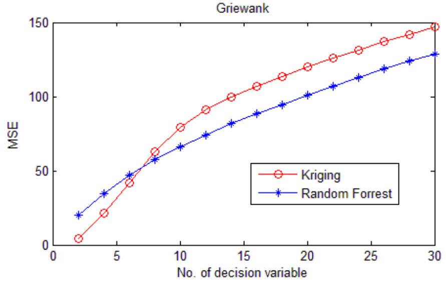



Figure: MSE of the Kriging model and random forrest on the Griewank problem.

# Experiment Settings



- Infill sampling criteria: LCB ($f_{lcb}(x) = \hat{f}(x) - ws(x)$)
- Model: Kriging and random forest (500 CARTs)
- Independent runs: 30
- Test problems: with 10 and 30 decision variables

# Results:Ellipsoid



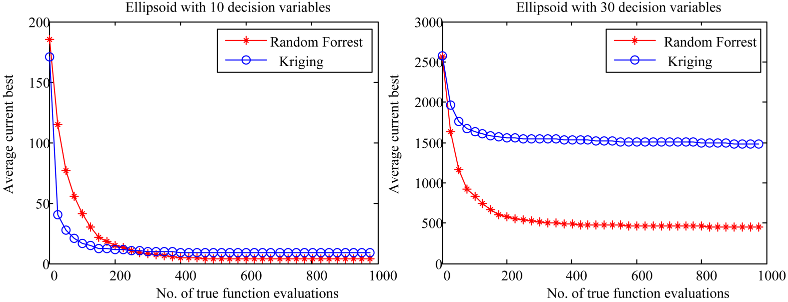
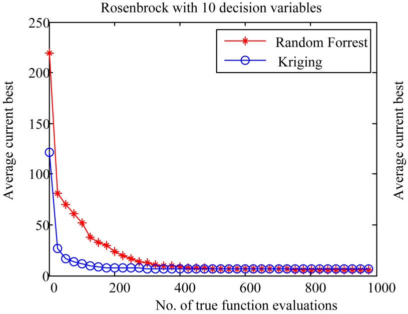



Figure: Convergence curve of the Kriging model and random forrest on the Ellipsoid problem.

# Results:Rosenbrock



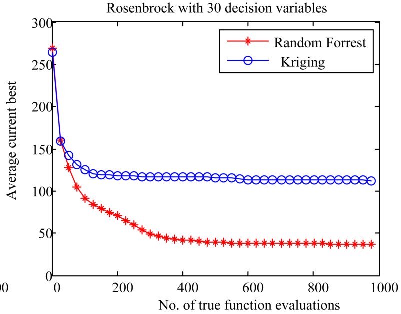




Figure: Convergence curve of the Kriging model and random forrest on the Rosenbrock problem.

# Results:Griewank

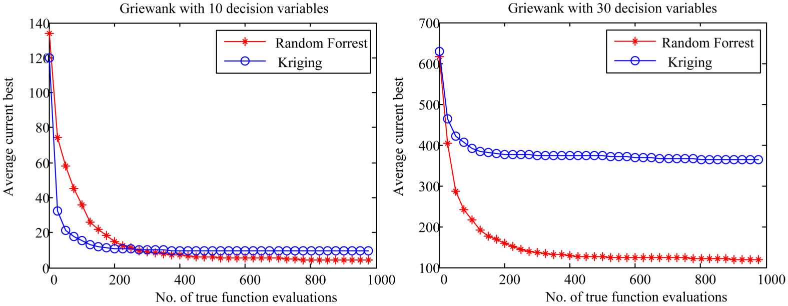



Figure: Convergence curve of the Kriging model and random forrest on the Griewank problem.

# Results:Rastrigin



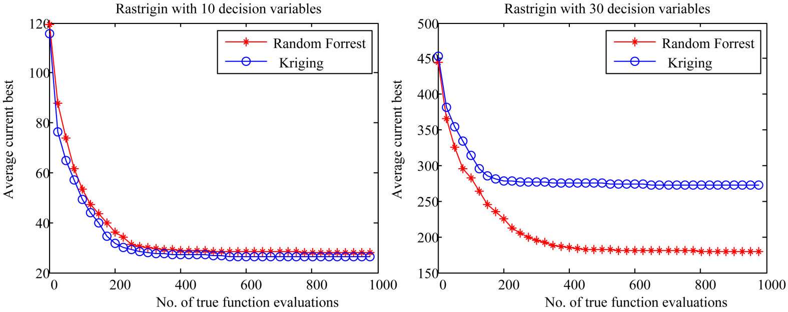
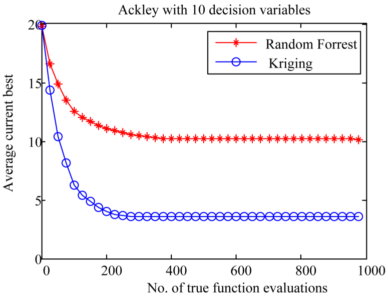



Figure: Convergence curve of the Kriging model and random forrest on the Rastrigin problem.

# Results:Ackley





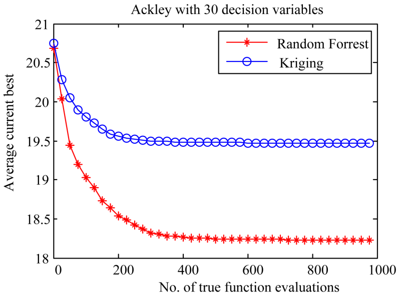




Figure: Convergence curve of the Kriging model and random forrest on the Ackley problem.

# Conclusions



- Model uncertainty is comparable with sample uncertainty for low-dimensional problems.
- Model uncertainty is significantly better than sample uncertainty for high-dimensional problems.
- It needs a large number of models to obtain the information of uncertainty.

Thank you!