Statistical Models and Methods for Computer Experiments

Olivier ROUSTANT

Ecole des Mines de St-Etienne

Habilitation à Diriger des Recherches

8th November 2011
Outline

Foreword

1. Computer Experiments: Industrial context & Mathematical background

2. Contributions: Metamodelling, Design and Software

3. Perspectives
Foreword
Some of my recent research deal with large data sets:
- Databases of atmospheric pollutants
  [Collab. with A. Pascaud, PhD student at the ENM-Douai]
- Databases of an information system
  [Co-Supervision of M. Lutz’ PhD, ST-MicroElectronics]

On the other hand, I have been studying time-consuming computer codes ➔ few data

For timing reasons, I will focus today on the 2nd topic, called computer experiments
Part I. CE: Industrial context & Mathematical background
Complex phenomena and metamodeling

Vehicle inputs

CAR DESIGN STAGE

TEST STAGE

Simulator outputs

Metamodel outputs

Reality outputs
Industrial context

- **Time-consuming computer codes**
  - car crash-test simulator, thermal hydraulic code in nuclear plants, oil production simulator, etc.

\[ x_1 \rightarrow y_1 \]
\[ x_2 \rightarrow y_2 \]
\[ \ldots \]
\[ x_d \rightarrow y_k \]

- xi’s: **input** variables — yj’s: the **output** variables
- Many possible configurations for the variables: often **uncertain**, **quantitative** / qualitative, sometimes spatio-temporal, nested...
Industrial context

- **Frequent Asked Questions**
  - **Optimization** (of the outputs)
    - Ex: Minimize the vehicle mass, subject to crash-test constraints

- **Risk assessment** (for uncertain inputs)
  - **Uncertainty propagation**: probability that $y_j > T$? Quantiles?
  - **Sensitivity analysis (SA)**: which proportion of $y_j$'s variability can be explained by $x_i$?
Mathematical background

- The idea is to build a metamodell, *computationally efficient*, from a few data obtained with the costly simulator.
Mathematical background

- How to build the metamodel?
  *Interpolation or approximation problem*

- How to choose the design points?
  *Related theory: design of experiments*

- *Can we trust the metamodel and how can we use it to answer the questions of engineers?*
Mathematical background

- Metamodel building: the **probabilistic framework**
  - Interpolation is done by conditioning a Gaussian Process (GP)

*Keywords: GP regression, Kriging model*
Mathematical background

- Main advantages of probabilistic metamodels:
  - **Uncertainty** quantification
  - **Flexibility** w.r.t. the addition of new points
  - **Customizable**, thanks to the trend and the covariance kernel

\[
K(x,x') = \text{cov}(Z(x), Z(x'))
\]

Smoothness of the sample paths of a stationary process depending on the kernel smoothness at 0
Mathematical background

- **Metamodel building: the functional framework**
  - Interpolation and approximation problems are solved in the setting of Reproducing Kernel Hilbert Spaces (RKHS), by regularization.

- The probabilistic and functional frameworks are not fully equivalent, but *translations* are possible via the Loève representation theorem (Cf. Appendix II).

\[
\phi : \mathcal{H}_K \rightarrow \mathcal{L}(Z) \\
K(x, .) \rightarrow Z_x
\]

\[
\langle K(x, .), K(y, .) \rangle = K(x, y) = \langle Z_x, Z_y \rangle
\]

- In both frameworks, *kernels play a key role.*
When industrials meet mathematicians

- The DICE (Deep Inside Computer Experiments) project
  A 3 years project gathering 5 industrial partners (EDF, IRSN, ONERA, Renault, TOTAL) and 4 academic partners (EMSE, Univ Aix-Marseille, Univ. Grenoble, Univ. Orsay)

- 3 PhD thesis completed + 2 initiated at the end of the project:
  - J. Franco (TOTAL), on Design of computer experiments
  - D. Ginsbourger (Univ. Berne), on Kriging and Kriging-based optimization
  - V. Picheny (Postdoc. CERFACS), on Metamodeling and reliability
  - B. Gauthier (Assistant Univ. St-Etienne), on RKHS
  - N. Durrande (Postdoc. Univ. Sheffield), on Kernels and dimension reduction
Part 2
Contributions

Selected Works
Contributions – Metamodels
An introductive case study

- Context: Supervision of J. Joucla’ Master internship at IRSN
  - IRSN is providing evaluations for Nuclear Safety
  - IRSN wanted to develop an expertise on metamodeling

- The problem: simulation of an accident in a nuclear plant
  - One functional output: temporal temperature curve
    - Only the curve maximum is considered -> scalar output
  - 27 inputs, with a given distribution for each

- The aim:
  - To investigate Kriging metamodeling
  - Final problem (not considered here): use Kriging for quantile estimation in a functional framework.
An introductive case study

- **Kernel choice**
  - Marginal simulations show different levels of “smoothness” depending on the inputs
  - The *Power-Exponential kernel* is chosen
    \[
    k(x - y) = \sigma^2 \exp \left( - \sum \left| \frac{x_i - y_i}{\theta_i} \right|^{p_i} \right)
    \]
  - The “smoothness” depends on \( p_j \) in \( ]0, 2] \)
    - Estimations: \( p_{11} \approx 1.23; \quad p_8 = 2 \)
  - **Remark:** The jumps are not modeled
An introductive case study

- Variable selection and estimation
  - Forward screening (alg. of Welch, Buck, Sacks, Wynn, Mitchell, and Morris)
- Post-treatment: Sensitivity analysis
  - To sort the variables hierarchically & Discard non-influent variables
  - To visualize the results
An introductive case study

- Acceptable results
  - Better than the usual 2\text{nd} order polynomial

- Several issues remain
  - How to model the jumps?
  - Shouldn’t we add $x_8$ and $x_{20}$ as part of the trend?
  - Can we re-use the MatLab code for another study?
    - Answer: No, because we have not paid enough attention to the code!
    - Solution? Coming soon…!
Additive kernels

- Additive Kriging [at least: Plate, 1999]
  - Adapt the idea of Additive Models to Kriging
    \[ Z(x) = Z_1(x_1) + \ldots + Z_d(x_d) \]
  - Resulting kernels, for independent processes:
    \[ k = k_1 \oplus \ldots \oplus k_d \]
- The aim: To deal with the curse of dimensionality

- Our contribution [Co-Supervision of N. Durrande’ PhD]
  - Theory: Equivalence between kernel & sample paths additivity
  - Empiric: Investigation of a relaxation algorithm for inference
Block-additive kernels

- The idea [Collab. with PhD std. T. Muehlenstaedt and J. Fruth]
  - To identify groups of variables that have no interaction together
  - To use the interactions graph to define block-additive kernels

- New mathematical tools
  - **Total interactions**
    - Involves the inputs sets containing both $x_i$ and $x_j$
    \[
    S_{\{i,j\}}^{TI} = \sum_{J \supseteq \{i,j\}} S_J
    \]
  - **FANOVA graph**
    - Vertices: input variables – Edges: weighted by the total interactions
Block-additive kernels

- Illustration of the idea relevance on the Ishigami function
  \[ f(x) = \sin(x_1) + A\sin^2(x_2) + B(x_3)^4\sin(x_1) = f_2(x_2) + f_{1,3}(x_1, x_3) \]
Block-additive kernels

Illustration of the blocks identification on a 6D function ("b")

\[ f(x) = \cos([l,x_1,x_2,x_3]a') + \sin([l,x_4,x_5,x_6]b') + \tan([l,x_3,x_4]c') \]

\[ f(x) = f_{1,2,3}(x_1,x_2,x_3) + f_{4,5,6}(x_4,x_5,x_6) + f_{3,4}(x_3,x_4) \]

\[ Z(x) = Z_{1,2,3}(x_1,x_2,x_3) + Z_{4,5,6}(x_4,x_5,x_6) + Z_{3,4}(x_3,x_4) \]

\[ k(h) = k_{1,2,3}(h_1,h_2,h_3) + k_{4,5,6}(h_4,h_5,h_6) + k_{3,4}(h_3,h_4) \]
Block-additive kernels

- Graph thresholding issue
  - Sensitivity of the method accuracy to the graph threshold value
Kernels for Kriging mean SA

- **Motivation:**
  - To perform a sensitivity analysis (independent inputs) of the proxy
  - To avoid the curse of recursion

- **The idea [Co-Supervision of N. Durrande’ PhD]**
  - Adapt the FANOVA kernels,
    \[ k = (1 + k_1) \otimes \ldots \otimes (1 + k_d) \]
  - based on the fact that the FANOVA decomposition of
    \[ f = (1 + f_1) \otimes \ldots \otimes (1 + f_d) \]
  - where the \( f_i \)'s are zero-mean functions, is obtained directly by expanding the product (Sobol, 1993)
Kernels for Kriging mean SA

- Solution with the functional interpretation
  - Start from the 1d- RKHS $H_i$ with kernel $k_i$
  - Build the RKHS of zero-mean functions in $H_i$, by considering the linear form $L_i$: $h \rightarrow \int h(t) d\nu_i(t)$. Its kernel is:
    \[
    k_{i,0}(x, y) = k_i(x, y) - \frac{\int k_i(x, s) d\nu_i(s) \times \int k_i(y, t) d\nu_i(t)}{\iint k_i(s, t) d\nu_i(s) d\nu_i(t)}
    \]
  - Use the modified FANOVA kernel
    \[
    k = (1 + k_{1,0}) \otimes ... \otimes (1 + k_{d,0})
    \]

- Remark
  - The zero-mean functions are not orthogonal to 1 in $H_i$, but orthogonal to the representer of $L_i$: $R_i(.) = \int K(., t) d\nu_i(t)$
Contributions – Designs
Selection of an initial design

- The radial scanning statistic (RSS)
- Automatic defects detection in 2D or 3D subspaces
- Visualization of defects
- Underlying mathematics:
  - law of a sum of uniforms, GOF test for uniformity based on spacings

If we use this design with a deterministic simulator depending only on $x_2-x_7$, we lose 80% of the information!
Selection of an initial design

- **Context:** first investigation of a deterministic code
- **Two objectives, and the current practice:**
  - To catch the code complexity
    - *space-filling designs (SFDs)*
  - To avoid losing information by dimension reduction
    - *space-fillingness should be stable by projection onto margins*

- **Our contribution [Collaboration with J. Franco, PhD stud.]:**
  - Dimension reduction techniques involve variables of the form $b'x$
    - *space-fillingness should be stable by projection onto oblique straight lines*
Selection of an initial design

- Application of the RSS to design selection

<table>
<thead>
<tr>
<th>Design type</th>
<th>Statistic value $^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uniform</td>
<td>0.039 (0.003)</td>
</tr>
<tr>
<td>Maximin Latin hypercube</td>
<td>0.048</td>
</tr>
<tr>
<td>Audze-Eglais Latin hypercube</td>
<td>0.037</td>
</tr>
<tr>
<td>Halton sequence</td>
<td>0.244</td>
</tr>
<tr>
<td>Faure sequence</td>
<td>0.161</td>
</tr>
<tr>
<td>Sobol sequence</td>
<td>0.101</td>
</tr>
<tr>
<td>Sobol sequence, with Owen scrambling</td>
<td>0.041 (0.006)</td>
</tr>
<tr>
<td>Sobol sequence, with Faure-Tezuka scrambling</td>
<td>0.088 (0.010)</td>
</tr>
<tr>
<td>Sobol sequence, with Owen + Faure-Tezuka scrambling</td>
<td>0.041 (0.006)</td>
</tr>
<tr>
<td>Strauss</td>
<td>0.040 (0.004)</td>
</tr>
</tbody>
</table>
Adaptive designs for risk assessment

- In frequent situations, the *global* accuracy of metamodels is not required
  - Example: Evaluation of the probability of failure $P(g(x) > T)$
    A good accuracy is required for $g(x) \approx T$

- Our contribution [Co-Supervision of V. Picheny’s PhD]
  - Adaptation of the IMSE criterion with suited weights
  - Implementation of an adaptive design strategy
Adaptive designs for risk assessment

- The static criterion. For a given point \( x \), and initial design \( X \):
  - With Kriging, we have a stochastic process model \( Y(x) \)
  - Use its density to weight the prediction error \( \text{MSE}(x) = s_K^2(x) \)
    - Large weight when the probability (density) that \( Y(x) = T \) is large

\[
\text{IMSE}_{T|X} = \int s_K^2(x|X) f_Y(x|Y(X) = y(T)) \, d\mu(x)
\]

\( \text{MSE}_{T}(x) \)
Adaptive designs for risk assessment

\[ \text{MSE}_T(x) \]

\[ \text{MSE}(x) \]

**True process and Kriging predictor**

**Original and weighted prediction variance**

- True process
- Training points
- Kriging expectation
- New Point

**Statistical models and methods for CE**
Adaptive designs for risk assessment

The dynamic criterion

\[
\text{IMSE}_{T|X}(x_{\text{new}}) = \int s_K^2(x|X, x_{\text{new}}) f_Y(x)|Y(X)\equiv Y(T) d\mu(x)
\]

Does not depend on \(Y(x_{\text{new}})\)

Illustration of the strategy, starting from 3 points: 0, 1/2, 1
Contributions – Software
Software for data analysis

- The need
  - To apply the applied mathematics on industrial case studies
  - To investigate the proposed methodologies
  - To re-use our [own!] codes 1 year later (hopefully more)…

- The software form
  - R language:
    - Freeware - Easy to use - Huge choice of updated libraries (packages)
  - User-friendly software prototypes
    - Trade-off between professional quality (unwanted) and un-re-usable codes
Software for data analysis

- The packages and their authors
  - A collective work: Supervisors [really], (former) PhD students and… some brave industrial partners!
    - **DiceDesign**: J. Franco, D. Dupuy, O. Roustant
    - **DiceKriging**: O. Roustant, D. Ginsbourger, Y. Deville
    - **DiceOptim**: D. Ginsbourger, O. Roustant
    - **DiceEval**: D. Dupuy, C. Helbert
    - **DiceView**: Y. Richet, Y. Deville, C. Chevalier
    - **KrigInv**: V. Picheny, D. Ginsbourger
    - **fanovaGraph**: J. Fruth, T. Muehlenstaedt, O. Roustant
    - **AKM**: N. Durrande

Statistical models and methods for CE
Software for data analysis

- The Dice packages (Feb. and March 2010) and their satellites

**DiceDesign**
*Design creation and evaluation*

**DiceKriging**
*Creation, Simulation, Estimation, and Prediction of Kriging models*

**DiceOptim**
*Kriging-Based optimization*

**DiceEval**
*Validation of statistical models*

**DiceView**
*Section views of Kriging predictions*

**KrigInv**
*Kriging-Based inversion*

**fanovaGraph (forthcoming)**
*Kriging with block-additive kernels*

**AKM (in preparation)**
*Kriging with additive kernels*
Software for data analysis

- DiceOptim: Kriging-Based optimization
  - Illustration of the adaptive constant liar strategy for 10 processors
  
  **Start:** 9 points (triangles) – Estimate a Kriging model.
  **1st stage:** 10 points simultaneously (red circles) – Reestimate.
  **2nd stage:** 10 new points simult. (violet circles) – Reestimate.
  ...

Statistical models and methods for CE
Software with data analysis

- Some comments about implementation [ongoing work with D. Ginsbourger (initiated during his PhD), and Y. Deville]

- Leading idea
  - The code should be as close as possible as the underlying maths
    - Example: Operations on kernels.

Illustration with isotropic kernels

\[ k_{\text{iso}}(x, y; \theta) = k(x, y; \theta, \ldots, \theta) \]

\[ \frac{\partial k_{\text{iso}}}{\partial \theta}(x, y; \theta) = \sum_{i=1}^{p} \frac{\partial k}{\partial \theta_i}(x, y; \theta, \ldots, \theta) \]

Unwanted solution: to create a new program \( k_{\text{iso}} \) for each new kernel \( k \)

Implemented solution: to have the same code for any basis kernel \( k \)

Tool: object-oriented programming

41  Statistical models and methods for CE
Part 3
Conclusions and perspectives
The results at a glance

- An answer to several practical issues
  - Kriging-Based optimization
  - Kriging-Based inversion
  - Model error for SA (not presented here)
  - A suite of R packages

- Development of the underlying mathematical tools
  - Designs
    - Selection of SFDs – Robustness to model error (not presented here)
  - Customized kernels
    - Dimension reduction with (block-)additive kernels
    - Sensitivity analysis with suited ANOVA kernels
General perspectives

- To extend the scope of the Kriging-Based methods
  - Actual scope of our contributions
    - Output: 1 scalar output
    - Inputs: d scalar inputs ($1 \leq d \leq 30$), quantitative
    - Stationary phenomena

- The needs
  - Spatio-temporal inputs / outputs
  - Several outputs
  - Also categorical inputs, possibly nested
  - $d \geq 30$
  - Several simulators for the same real problem
  - …
A fact: The kernels are underexploited

- In practice:
  - The class of tensor-product kernels is used the most

- In theory:
  - (Block-)Additive kernels for *dimension reduction*
  - FANOVA kernels for *sensitivity analysis*
  - Convolution kernels for *non stationarity*
  - Scaled-transformed kernels for *non stationarity*
  - Kernels for *qualitative variables*…
  - Kernels for *spatio-temporal variables*…
What’s missing & Directions to widen new kernel classes

- To adapt the methodologies to the kernel structures
  - Inference, designs, applications
  - Potential gains
    - Ex: Additive kernels should also reduce dimension in optimization

- To extend the softwares to new kernels
  - Several classes of kernels should live together
    - Object-oriented programming required
  - Challenge: To keep the software controllable
    - Collaborations with experts in computer science
Thank you for your attention!
Supplementary slides
Supplementary slides

- DiceView: 2D (3D) section views of the Kriging curve (surface) and Kriging prediction intervals (surfaces) at a site.