Comparing and generating Latin Hypercube designs in Kriging models

Giovanni Pistone, Grazia Vicario

Politecnico di Torino

Department of Mathematics
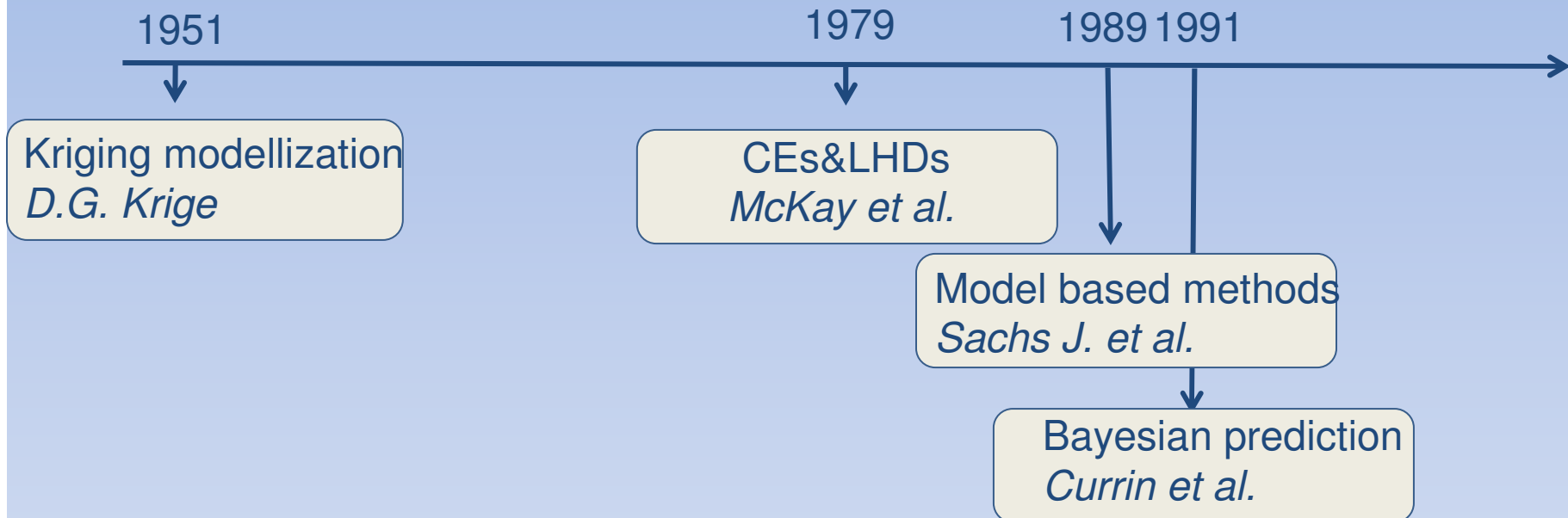
Corso Duca degli Abruzzi, 24 – 10129 Torino, Italy

Outline

✓ Background: Kriging models and Latin Hypercube Designs

✓ Introduction

✓ Ordinary kriging on a lattice: the correlation function

✓ Ordinary kriging on a lattice: the output prediction

✓ Classes of Latin Hypercube designs on lattices

✓ Results and conclusions

Background

Official starts

Standard modern book references:

M.J. Sasena. *Flexibility and Efficiency Enhacements for Costrained Global Design Optinmization with Kriging Approximations*. PhD Thesis University of Michigan, 2002.

T.J. Santner, B.J. Williams, and W.I. Notz. *The design and analysis of computer experin* Springer Series in Statistics. Springer-Verlag, New York, 2003.

K-T. Fang, R. Li, and A. Sudjianto. *Design and modeling for computer experiments. Con Science and Data Analysis Series*. Chapman & Hall/CRC, Boca Raton, FL, 2006

Introduction

DoE: a protocol for designing **physical experiments** in order to achieve valid, correct and unprejudiced inferences

And for **Computer Experiments**??

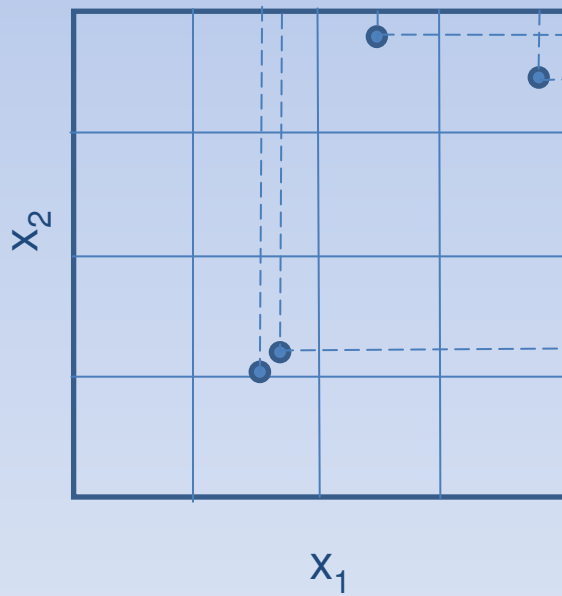Ideal design strategy: to uniformly spread the points across the experimental region ⟹ **space-filling designs**

Designs based
on sampling methods
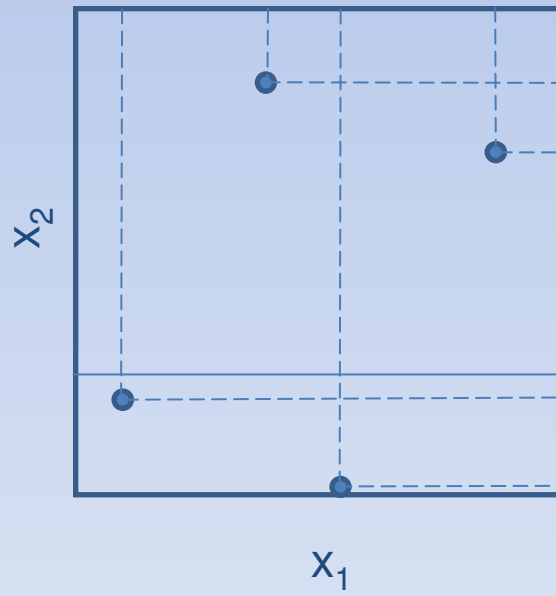
Designs based
on measures of distance

Designs based
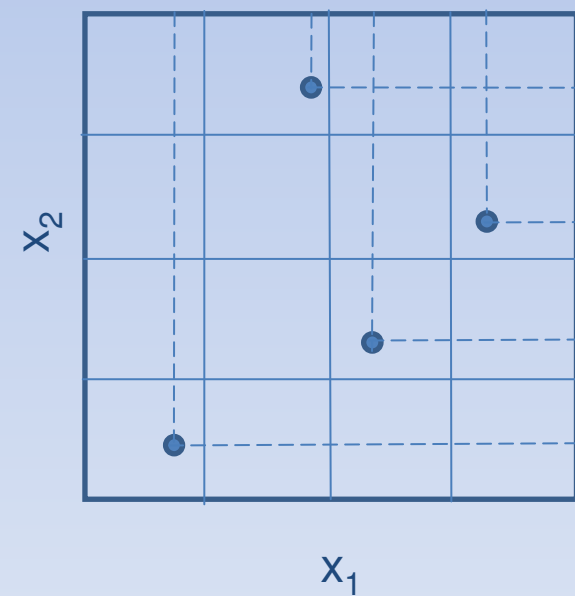on the uniform distribution

Introduction

Random sampling

Stratified sampling

Latin-hypercube
sampling

Introduction

A problem of interest: the design of experiment, i.e. the choice of a training set with good performances when evaluated with respect to a statistical index (e.g. Mean Squared Prediction Error, MSPE)

❑ How to build a predictor

❑ How to evaluate the efficiency of the prediction

❑ How to choose the points of the design

Ordinary kriging on a lattice: the correlation function

The underlying model is a parametric model of Gaussian type:

$$Y(\mathbf{x}) = f'(\mathbf{x})\beta + Z(\mathbf{x})$$

$f'(\mathbf{x})$: known regression function

β: unknown regression coefficients

$Z(\mathbf{x})$: Gaussian random field with zero mean and stationary covariance over a design space $\mathcal{X}_d \subset \mathbb{R}_d$, $\operatorname{cov}\left[Z(\mathbf{x}_i), Z(\mathbf{x}_j)\right] = \sigma_Z^2 R(\mathbf{x}_i - \mathbf{x}_j)$ $\qquad \sigma_Z^2$ where is

the field variance, R is the Stationary Correlation Function (SCF) depending only on the displacement vector $\mathbf{h}$:

$$R(\mathbf{x}_i - \mathbf{x}_j) = R(\mathbf{h})$$

$$R(\mathbf{h}) = R(-\mathbf{h})$$

$$R(\mathbf{0}) \equiv \operatorname{var}[Z(\mathbf{x})] = \sigma_Z^2 > 0$$

If the space of locations is a lattice, the model is an algebraic statistical model

Ordinary kriging on a lattice: the correlation function

Choice of the correlation function: Exponential Correlation Function

$$R(h;\theta) = \prod_{s=1}^{d} \exp\left\{-\theta_s |h_s|^p\right\} = \exp\left\{-\sum_{s=1}^{d} \theta_s |h_s|^p\right\}$$

θ_s, $s = 1, 2, \ldots, d$, are positive scale parameters

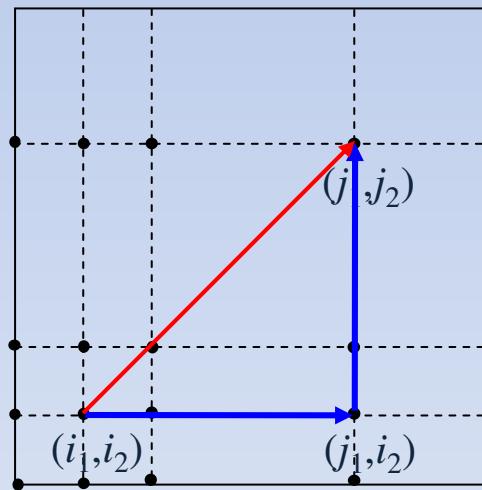p between 0 and 2

Assumptions in this paper:

- $\theta_s = \theta$, $\forall s = 1, 2, \ldots, d$: the correlation depends only on the distance $|\mathbf{h}|$ between any pair of points $\mathbf{x}$ and $\mathbf{x+h}$
- $p = 1$
- $\sigma_Y^2 = 1$

Assumptions:

□ the Gaussian field is defined on a regular rectangular lattice $\mathcal{X}_d = \{1, \dots, l\}^d$
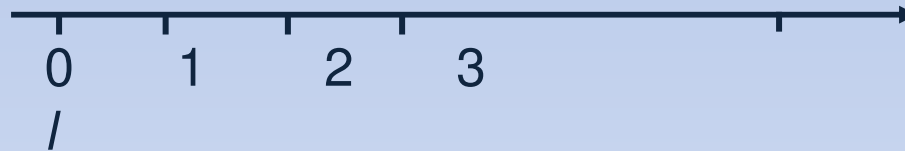
□ Manhattan distance: $|\mathbf{x} - \mathbf{y}| = \displaystyle\sum_{s=1}^{d} |x_s - y_s|$

Ordinary kriging on a lattice: the correlation function

$$
D_1 = \begin{pmatrix}
0 & 1 & 2 & . & . & l-1 \\
1 & 0 & 1 & 2 & . & . \\
2 & 1 & 0 & . & . & . \\
. & . & . & . & . & 2 \\
. & . & . & . & . & 1 \\
l-1 & . & . & . & . & 0
\end{pmatrix}
\qquad
\Gamma_1 = \begin{pmatrix}
1 & t & t^2 & . & . & t^{l-1} \\
t & 1 & t & t^2 & . & . \\
t^2 & t & 1 & . & . & . \\
. & . & . & . & . & t^2 \\
. & . & . & . & 1 & t \\
t^{l-1} & . & . & . & t & 1
\end{pmatrix}
\qquad
t = \exp(-\theta)
$$

One single factor

$$\xrightarrow{\hspace{6cm}}$$

$$
\begin{array}{cccc}
0 & 1 & 2 & 3
\end{array}
$$

l

$$
D_d = \begin{pmatrix}
\mathbf{D_{d-1}} & \mathbf{D_{d-1}+1} & \mathbf{D_{d-1}+2} & . & . & \mathbf{D_{d-1}+l-1} \\
\mathbf{D_{d-1}+1} & \mathbf{D_{d-1}} & \mathbf{D_{d-1}+1} & \mathbf{D_{d-1}+2} & . & . \\
\mathbf{D_{d-1}+2} & \mathbf{D_{d-1}+1} & \mathbf{D_{d-1}} & . & . & . \\
. & . & . & . & . & \mathbf{D_{d-1}+2} \\
. & . & . & . & . & \mathbf{D_{d-1}+1} \\
\mathbf{D_{d-1}+l-1} & . & . & . & . & \mathbf{D_{d-1}}
\end{pmatrix}
\qquad
\Gamma_d = \Gamma_{d-1} \otimes \Gamma_1
$$

d factors

Ordinary kriging on a lattice: the output prediction

Kriging is a linear method of spatial interpolation: the random variable $Y(x_0)$ is
predicted with a linear (affine) combination of observed random variables $Y(\mathbf{x}_1)$,
..., $Y(\mathbf{x}_n)$ in the *training set*:

$$\hat{Y}(\mathbf{x}_0) = a_0 \mathbf{x}_n + \sum_{i=1}^{n} a_i Y(\mathbf{x}_i)$$

The weights in the l.c. are evaluated according a statistical model on the joint
distribution of $(Y_0', \mathbf{Y}')', \beta, \sigma_Z^2 \Sigma$:

$$\Sigma = \begin{pmatrix} 1 & \mathbf{r_0} \\ \mathbf{r_0} & \mathbf{R} \end{pmatrix}$$

Ordinary Kriging model $Y(\mathbf{x}) = \beta + Z(\mathbf{x})$

The kriging model can be considered an empirical bayesian approach to
computer experiments

Assume β and Γ unknown

A Linear Predictor LP $\hat{Y}(\mathbf{x}_0) = a_0 + \sum_{i=1}^{n} a_i Y(\mathbf{x}_i)$ is unbiased iff:

$$\beta \equiv \mathbb{E}\left[\hat{Y}(\mathbf{x}_0)\right] = a_0 + \sum_{i=1}^{n} a_i \beta \implies a_0 = 0 \text{ and } \sum_{i=1}^{n} a_i = 1$$

and it is the Best (BLUP) if it minimizes the Mean Squared Prediction Error (MS

$$\text{MSPE}\left[\hat{Y}_0\right] = 1 - \mathbf{r}_0' \mathbf{R}^{-1} \mathbf{r}_0 + \mathbf{c}_0' \left(\mathbf{u}_n' \mathbf{R}^{-1} \mathbf{u}_n\right)^{-1} \mathbf{c}_0$$

$$\mathbf{c}_0 = 1 - \mathbf{u}_n' \mathbf{R}^{-1} \mathbf{r}_0$$

The unknown value of the correlation is estimated from the set of the training points and plugged in into the formula of the estimator

Classes of Latin Hypercube designs on lattices

Step 1

Permutations of the l integers (number of the levels) and construction of the matrix $l \times (l!)^{d-1}$ containing all the LH designs with d factors.

Example: the possible 24 LH designs relative to $d = 2$ factors each one with $l = 4$ levels

LH	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Training points	11	11	11	14	14	11	11	11	13	13	13	14	14	13	13	13	12	12	12	14	14	12	12	12
	22	22	24	21	21	24	23	23	21	21	24	23	23	24	22	22	23	23	24	22	22	24	21	21
	33	34	32	32	33	33	34	32	32	34	31	31	32	32	34	31	31	34	33	33	31	31	34	33
	44	43	43	43	42	42	42	44	44	42	42	42	41	41	41	44	44	41	41	41	43	43	43	44

Classes of Latin Hypercube designs on lattices

Step 2

Construction of the distance matrix between any pair of points in the lattice

Step 3

Implementation of the Kronecker product between any pair of matrices, so the computing of the covariance matrix between any pair of points of the lattice is available

Example: covariance sub-matrix of the 11th LH design (lattice points (1,3), (2,4), (3,1) and (4,2))

$$\begin{pmatrix} 1 & t^2 & t^4 & t^4 \\ t^2 & 1 & t^4 & t^4 \\ t^4 & t^4 & 1 & t^2 \\ t^4 & t^4 & t^2 & 1 \end{pmatrix} \qquad t = \exp(-\theta)$$

Classes of Latin Hypercube designs on lattices

Step 4

Computation of the statistical index chosen for the comparison: Total Mean Squared Prediction Error (TMSPE), Entropy, the Minimax Distance and Maxmin Distance, ...

Step 5

Clustering of the LHs according to the same value of the index in the previous step. Both the TMSPE and the determinant of the covariance matrix are rational function of the parameter t. The rational functions are exactly computed with a symbolic software. Designs with the same function are in the same cluster.

For computing the predictor variances in closed form:
CoCoA (Computations in Commutative Algebra), see http://cocoa.dima.unige.it

Other computations related with the exponential model for covariances: software R, see http://www.R-project.org/

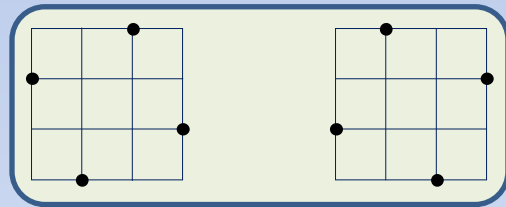# Classes of Latin Hypercube designs on lattices
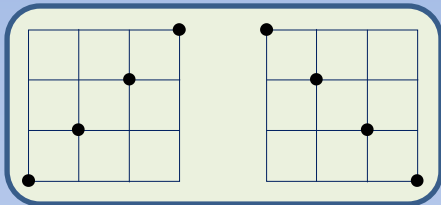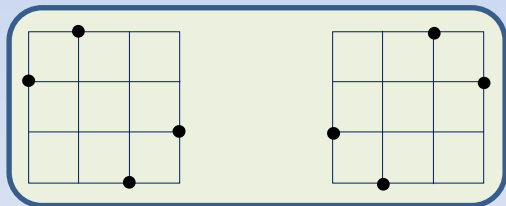
Class 1
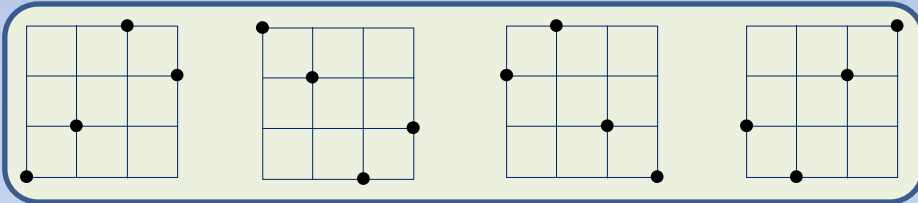
Class 2
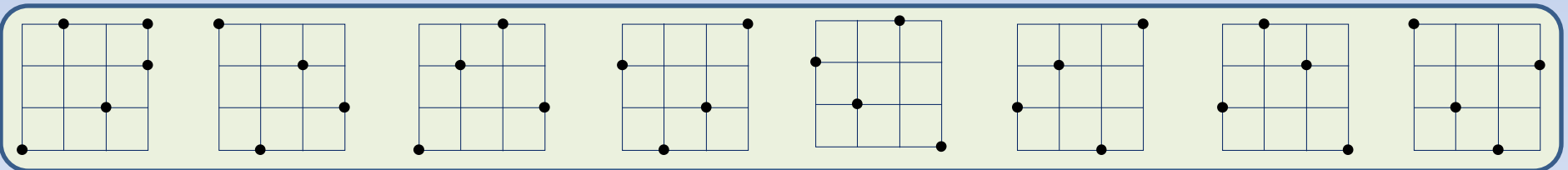
Class 3
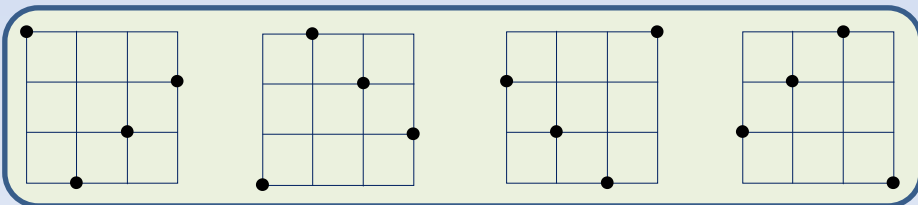
Class 4

Classes of Latin Hypercube designs on lattices
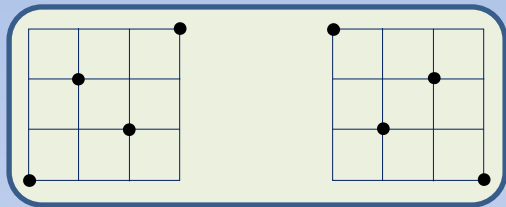
Class 5

Class 6

Class 7

Results and conclusions

Comments

✓ Class 6 is the best one (it consists of U-design according B. Tang (1993).
   These designs are also tilted 2^2.

✓ Classes 3,4,7 are essentially equivalent and worse than class 6

✓ Classes 4 and 5 are essentially equivalent to the cyclic designs (Bates
   et al. (1996), very recommended for Fourier regression models

✓ Class 2 is second worse

✓ Class 1 and 4 consist of regular fractions 4^{2-1}

✓ Class 3 contains regular fractions 2^{4-2} (pseudofactors)

✓ An LH design is an orthogonal array with strength 1 and vice versa
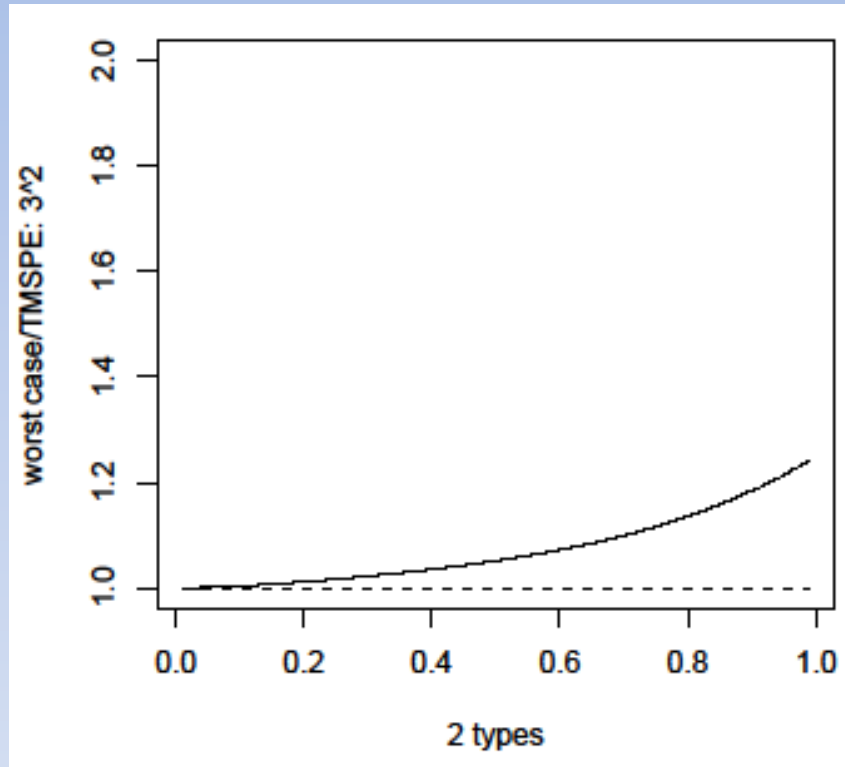
Results and conclusions

I=3 levels, d= 2 variables

✓ Each cluster has a different performance with respect to the TMSPE criteri
✓ The worst case are the two diagonals LHDs (dashed line)

Results and conclusions
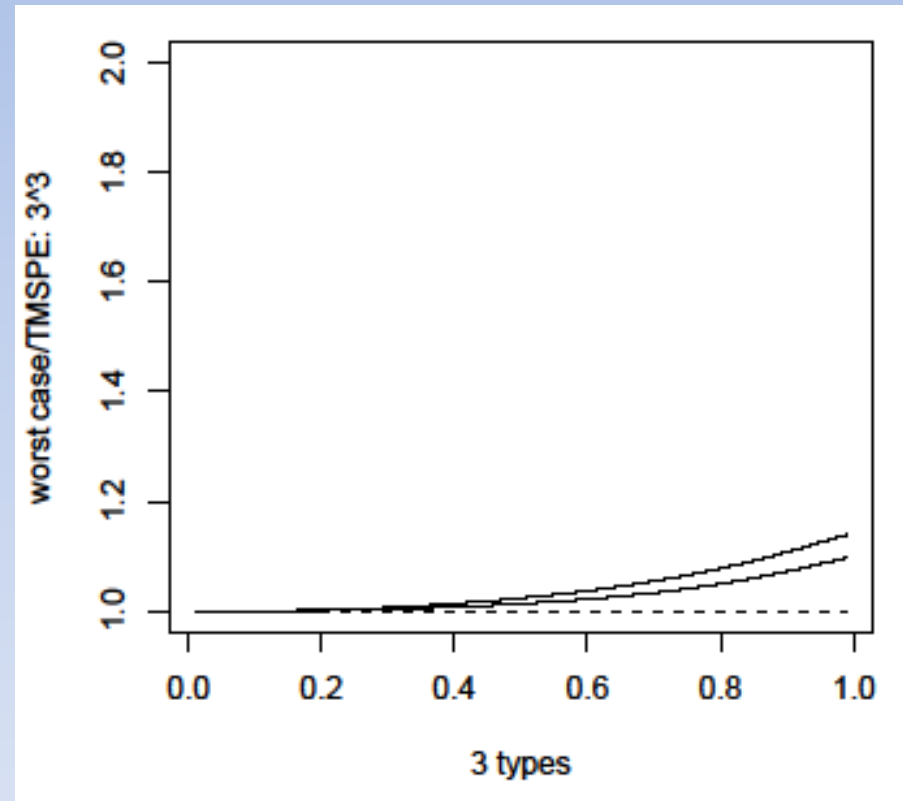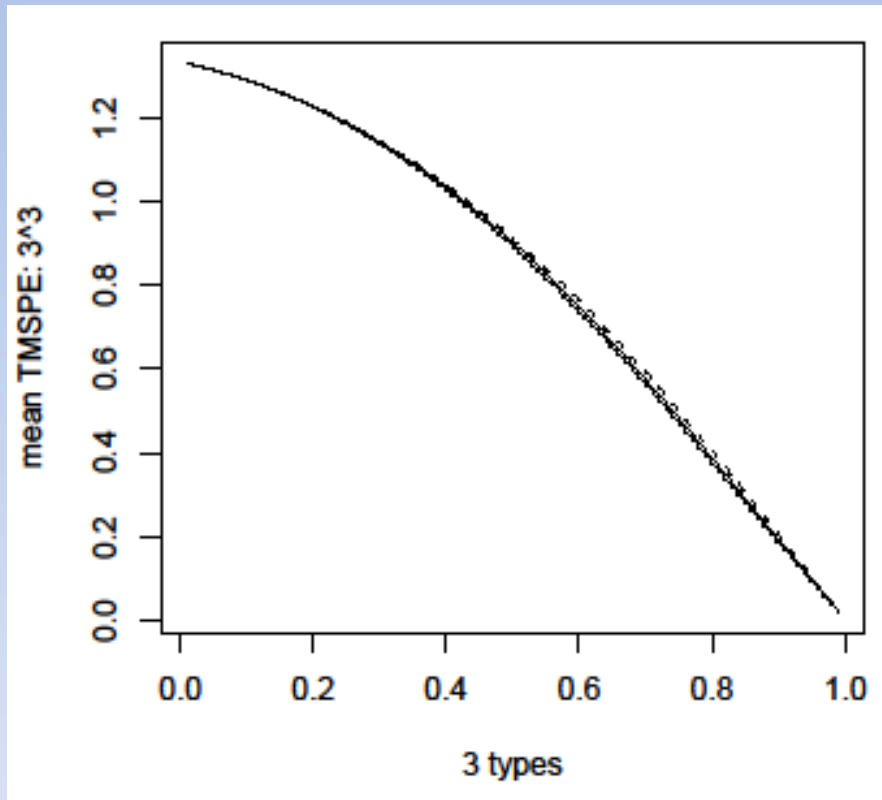
I=3 levels, d= 2 variables

The speed of convergence near $\theta = 0$ ($t = 1$) is very different!!!

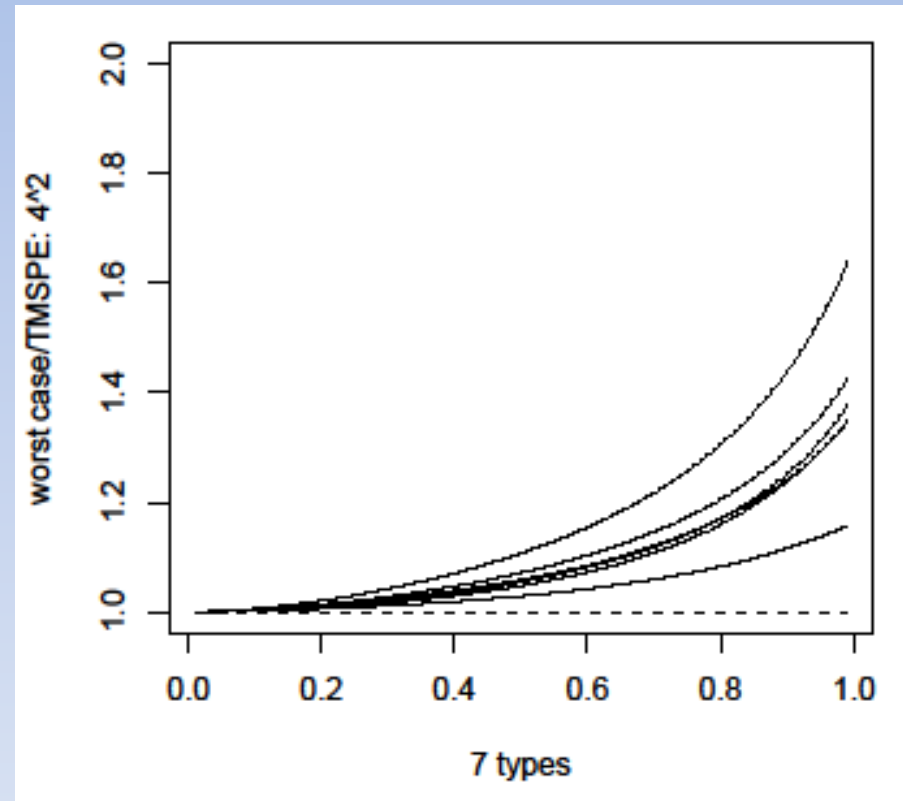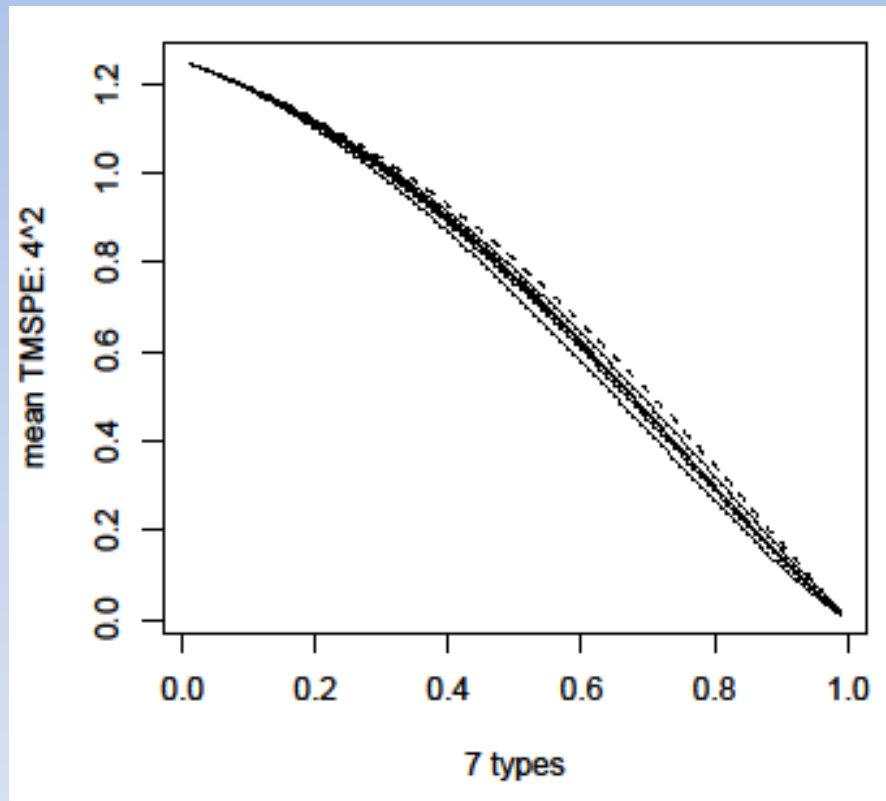*The formal computation allows a precise evaluation of the behavior near $t = 1$*

Results and conclusions

l=3 levels, d= 3 variables
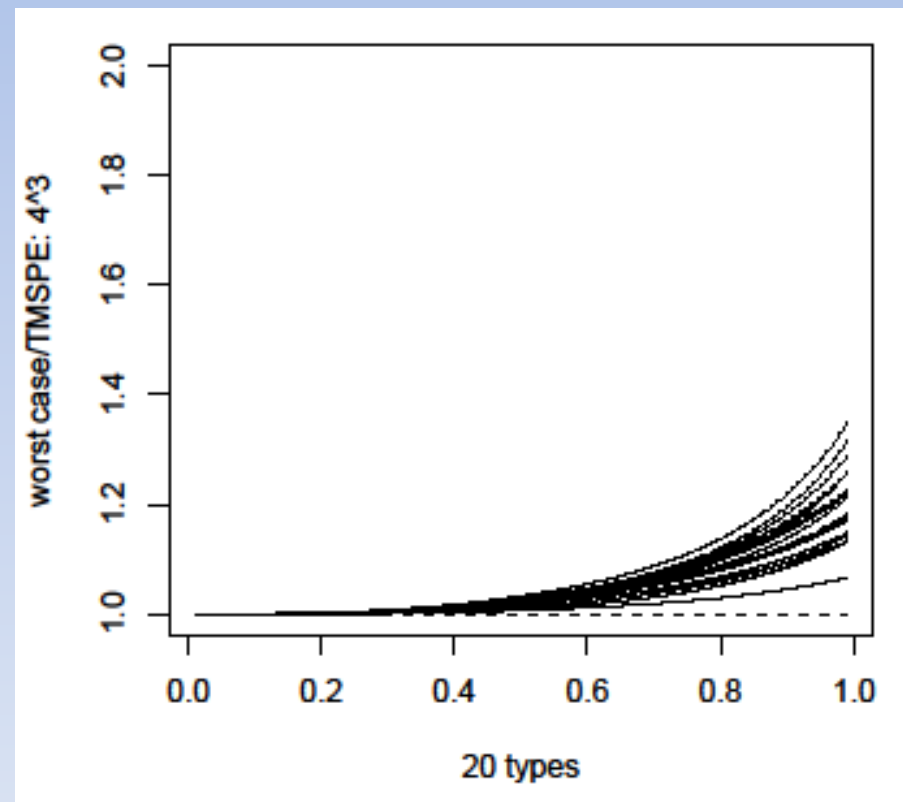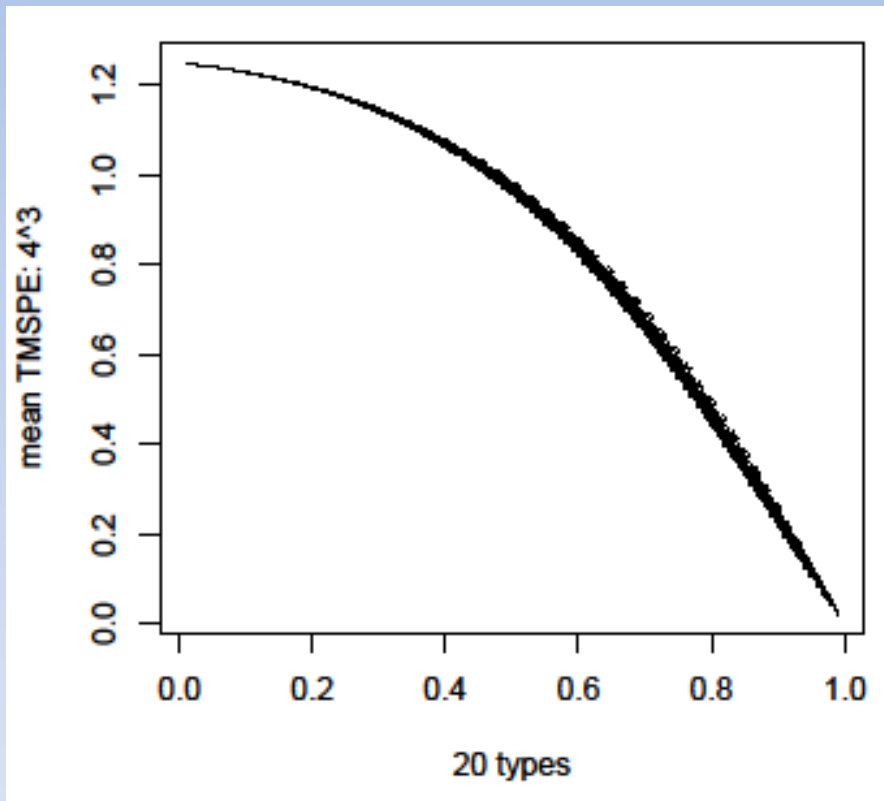
Results and conclusions

l=4 levels, d= 2 variables

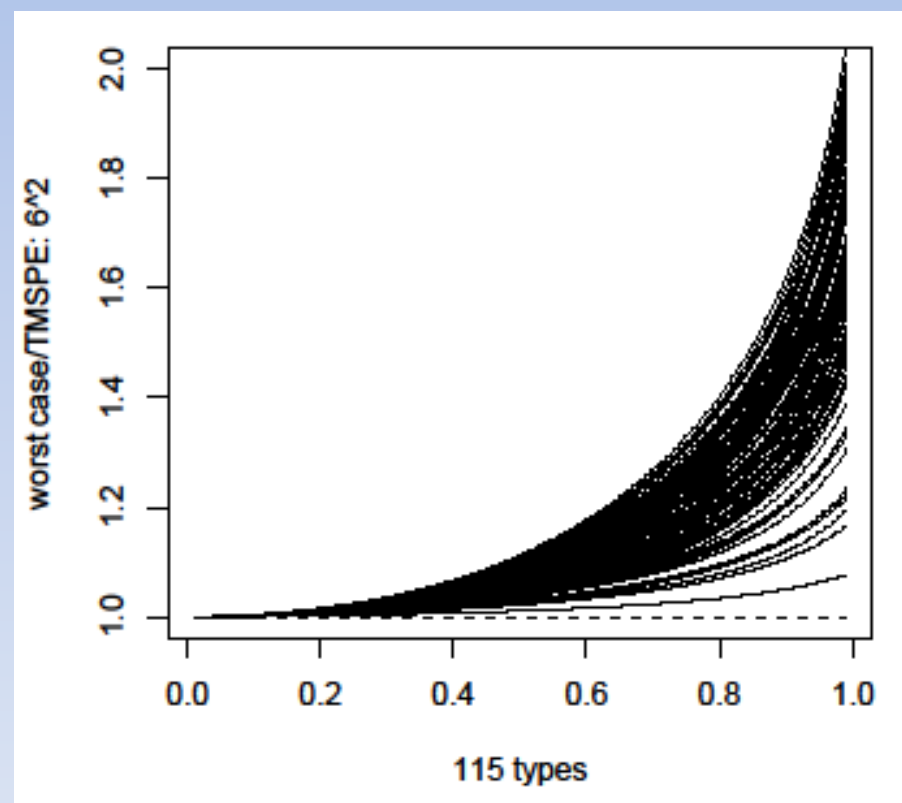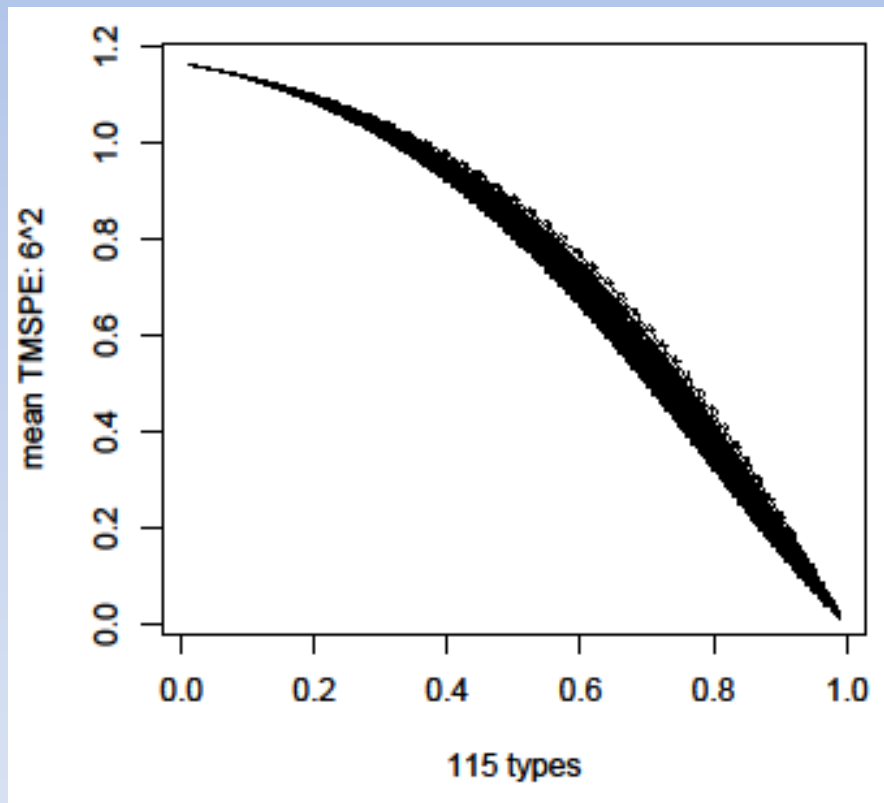For a given number of factors, the difference increases with the number of levels!!!

Results and conclusions

l=4 levels, d= 3 variables

Results and conclusions

l=6 levels, d= 2 variables

References

Butler, N.A. (2001). Optimal and Orthogonal Latin Hypercube Designs for Computer Experiments.*Biometrika* **88**: 847-857.

Currin, C., Mitchell T.J., Morris, M.D. and Ylvisaker, D. (1991). Bayesian prediction of deterministic functions, with applications to the design and analysis of computer experiments. *Journal of the American Statistical Association* **86:** 953-963.

Fang, K.-T., Li R., Sudjianto, A. (2006). Design and Modelling for Computer Experiments. Chapman&Hall Boca Raton, FL.

Fang, K.-T., Lin D.K., Winker, P., Zhang, Y. (2000). Uniform Design: Theory and Applications. *Technometrics* **42**: 237–248.

Krige, D.G. (1951). A statistical approach to some mine valuations and allied problems at the Witwatersrand. Master's thesis of the University of Witwatersrand.

McKay, M.D. Conover, W.J., and Beckman, R.J. (1979). A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code". *Technometrics* **21**: 239–245.

Park, J.-S. (1994) Optimal latin-hypercube designs for computer experiments. *J. Statist. Plann. Inference* **43:** 381-402.

Pistone G., Vicario G. (2009). *Design for Computer Experiments: comparing and generating designs in Kriging models*, In: Erto P., Statistics for Innovation. (pp. 91-102). Springer-Verlag Italia. ISBN: 978-88-470-0815.

Sachs J., Welch W.J., Mitchell T.J., Wynn H.P. (1989). Design and analysis of computer experiments. *Statistical Science*, **4**: 409-423.

References

Santner, T. J., Williams, B. J., and Notz, W. I. (2003). The Design and Analysis of Computer Experiments. Springer-Verlag, New York.

Sasena, M. J. (2002). Flexibility and Efficiency Enhacements for Costrained Global Design Optinmization with Kriging Approximations, PhD Thesis University of Michigan.

Ye, Q., Li, W. and Sudjianto, A. (2000). Algorithmic construction of optimal symmetric latin hypercube designs, *J. Statist. Plann. Inference*, **90**: 145-159

Welch W. J., Buck, R.J., Sacks J., Wynn H. P., Mitchell T. J. and Morris, M.D. (1992). Sreening, predicting, computer experiments. *Technometrics* **34**: 15–25.