On designs for fields and designs for computers

Hervé Monod, *INRA, Unité MIA de Jouy-en-Josas*

- Introduction: a short story
- Designs for field experiments
- Designs for computer simulations
- Not a conclusion

Notations

INRA : Institut National de la Recherche Agronomique

MIA : Mathématique et Informatique Appliquées

Jouy-en-Josas : small city 20 km south-west of Paris

Notations

INRA : Institut National de la Recherche Agronomique

MIA : Mathématique et Informatique Appliquées

Jouy-en-Josas : small city 20 km south-west of Paris

RAB : my “fairy godmother” in statistics

Introduction : a short history

Period 1 : Designs for field (and lab) experiments

1987 : first encounter with RAB at Rothamsted Experimental Station

1991 : my PhD and recruitment at INRA

$1991 - 1997$: regular mutual visits, 8 joint papers

$1997 - 2003$: applications to variety trials

Other members of the “French connection”:

André Kobilinsky, Jean-Marc Azaïs, Pierre Druilhet, Olivier David

Period 2 : Designs for computer experiments (sensitivity analysis)

$2001-$: collaborations on crop and gene flow modelling

$2003-$ increasing reliance on “complex models” at INRA

2004 : SAMO (Sensitivity Analysis of Model Output), Santa Fe

$\Longrightarrow$ need for more statistical tools to better assess model behaviour

Designs for fields

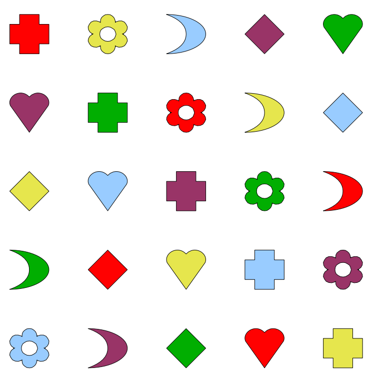
or . . . manipulating a bit of Rosemary's toolbox

- combinatorics and algebra
- factorial structures
- permutation groups
- real-life motivation

Definitions

Definition 1 ***Mutually Orthogonal Latin squares*** *of order n (= 5):*

Definition 2 ***Fractional factorial design*** *of* ***resolution*** *3*
or ***Orthogonal Array*** *of* ***strength*** *2*



F_1 (Row)	F_2 (Column)	F_3 (Color)	F_4 (Symbol)
1	1	1	1
1	2	2	2
1	3	3	3
1	4	4	4
1	5	5	5
2	1	4	5
2	2	5	1
2	3	1	2
...	...	...	...

Definition 3 *a* ***field***

Example 1 : Factorial structure for non-factorial designs

Pseudofactors useful to design and analyse

- square lattices
- lattice squares
- rectangular lattices
- some partially balanced incomplete block designs
- ...

Monod and Bailey, 1992

Extension: affine-resolvable designs

Problem: resolvable design for 12 treatments in blocks of size 6 ?

Solution: by using a fractional factorial design for 2-level factors

Resolution III factorial in 12 runs (Orthogonal Array)

Placket and Burman, 1948

Factor	Run (treatment)											
	1	2	3	4	5	6	7	8	9	10	11	12
F_1	1	1	2	1	1	1	2	2	2	1	2	2
F_2	2	1	1	2	1	1	1	2	2	2	1	2
F_3	1	2	1	1	2	1	1	1	2	2	2	2
F_4	2	1	2	1	1	2	1	1	1	2	2	2
$\ldots$												
F_{11}	1	2	1	1	1	2	2	2	1	2	1	2

Resolution III factorial in 12 runs (Orthogonal Array)

Factor	Run (treatment)											
	1	2	3	4	5	6	7	8	9	10	11	12
F_1	1	1	2	1	1	1	2	2	2	1	2	2
F_2	2	1	1	2	1	1	1	2	2	2	1	2
F_3	1	2	1	1	2	1	1	1	2	2	2	2
F_4	2	1	2	1	1	2	1	1	1	2	2	2
...												
F_{11}	1	2	1	1	1	2	2	2	1	2	1	2

Optimal resolvable design for 12 treatments

Replicate	Block 1						Block 2					
1	1	2	4	5	6	10	3	7	8	9	11	12
2	2	3	5	6	7	11	1	4	8	9	10	12
3	1	3	4	6	7	8	2	5	9	10	11	12
4	2	4	5	7	8	9	1	3	6	10	11	12

Model: $Y = X\tau + Z\beta + \varepsilon$

Information matrix: $M = X'\mathrm{pr}^{\perp}(Z)X$

Properties: Schur-optimality (w.r.t. the eigenvalues of M)

Bailey, Monod and Morgan, 1995

Example 2 : Factorial structure and neighbour-balance

Problem: intercropping experiment

Each plot : 1 tree species (factor 1) associated with 1 crop species (factor 2)

Possible *interferences* from trees on neighbour plots

Can associations be neighbour-balanced with respect to tree varieties?

One solution (among others): mutually orthogonal Latin squares number $=$ tree species (4)

colour $=$ crop species (2)

0	1	2	3	0	1	2	3
1	0	3	2	2	3	0	1
2	3	0	1	3	2	1	0
3	2	1	0	1	0	3	2

$\Longrightarrow$

3	0	0	1	1	2	2	3	3	0
1	1	2	0	3	3	0	2	1	1
0	2	3	3	2	0	1	1	0	2
2	3	1	2	0	1	3	0	2	3

$\Longrightarrow$ *Tree-balanced designs*

Model: $Y = Z\beta + X\tau + L\lambda + R\rho + \varepsilon$

Properties:

- orthogonality and optimality
- valid randomization

Monod and Bailey, 1992; Monod and Bailey, 1993

Example 3 : semi-Latin squares

Complete blocks by rows and by columns for 8 treatments?

0	1	2	3
1	0	3	2
2	3	0	1
3	2	1	0

0	1	2	3
2	3	0	1
3	2	1	0
1	0	3	2

$\Longrightarrow$

0 0	1 1	2 2	3 3
1 2	0 3	3 0	2 1
2 3	3 2	0 1	1 0
3 1	2 0	1 3	0 2

Mut. Orthog. Latin squares $\Longrightarrow$ *optimal doubly resolvable designs*

Bailey, 1988, 1992;

Bailey and Monod, 2001 (semi-Latin rectangles)

Applications to real field experiments

Control of inter-plot competition

- workshop on competition in variety trials

$\Longrightarrow$ no neighbour-balanced designs but . . .

adoption of systematic guard rows around harvested rows (maize, cereals, oilseed rape)

Control of between-plot heterogeneity

- increasing use of incomplete block designs in variety trials

$\Longrightarrow$ adoption of semi-Latin squares (maize, cereals, oilseed rape)

ITCF semi-Latin square with 32 varieties of winter barley

	Column 1							
L1	20	26	28	27	4	5	32	10
L2	6	2	23	15	7	24	19	30
L3	9	29	17	1	8	12	16	25
L4	11	31	13	22	21	18	3	14

	Column 2							
L1	31	14	1	8	6	22	24	19
L2	13	3	10	27	4	17	12	16
L3	30	20	7	26	11	32	18	21
L4	5	2	23	15	28	29	25	9

	Column 3							
L1	13	17	16	21	29	7	23	11
L2	28	9	25	8	18	20	31	22
L3	27	14	6	5	15	4	3	2
L4	19	12	32	1	10	26	24	30

	Column 4							
L1	18	3	2	25	9	30	15	12
L2	14	32	5	21	26	11	1	29
L3	31	23	13	22	28	19	24	10
L4	16	6	8	27	4	20	17	(7)

Analysis semi-L. sq.	Term	df	Sum Squares	Mean Square	F
	Variety	31	2850.1	91.9	8.4
	Row	3	2949.4	983.1	89.5
	Column	3	556.6	185.5	16.9
	Sub-block	9	853.8	94.9	8.6
	Residual	80	879.0	11.0	

Designs for sensitivity analysis

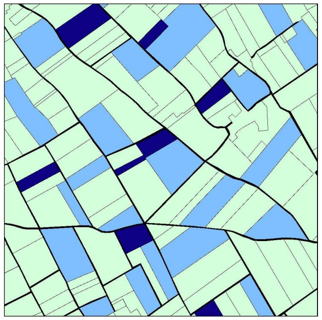
Saltelli, Chang and Scott 2000;

Santner, Williams and Notz 2003;

Fang, Li and Sudjianto 2006

Example: modelling gene flow over a landscape

Landscape and maize allocation



+

meteo daily data

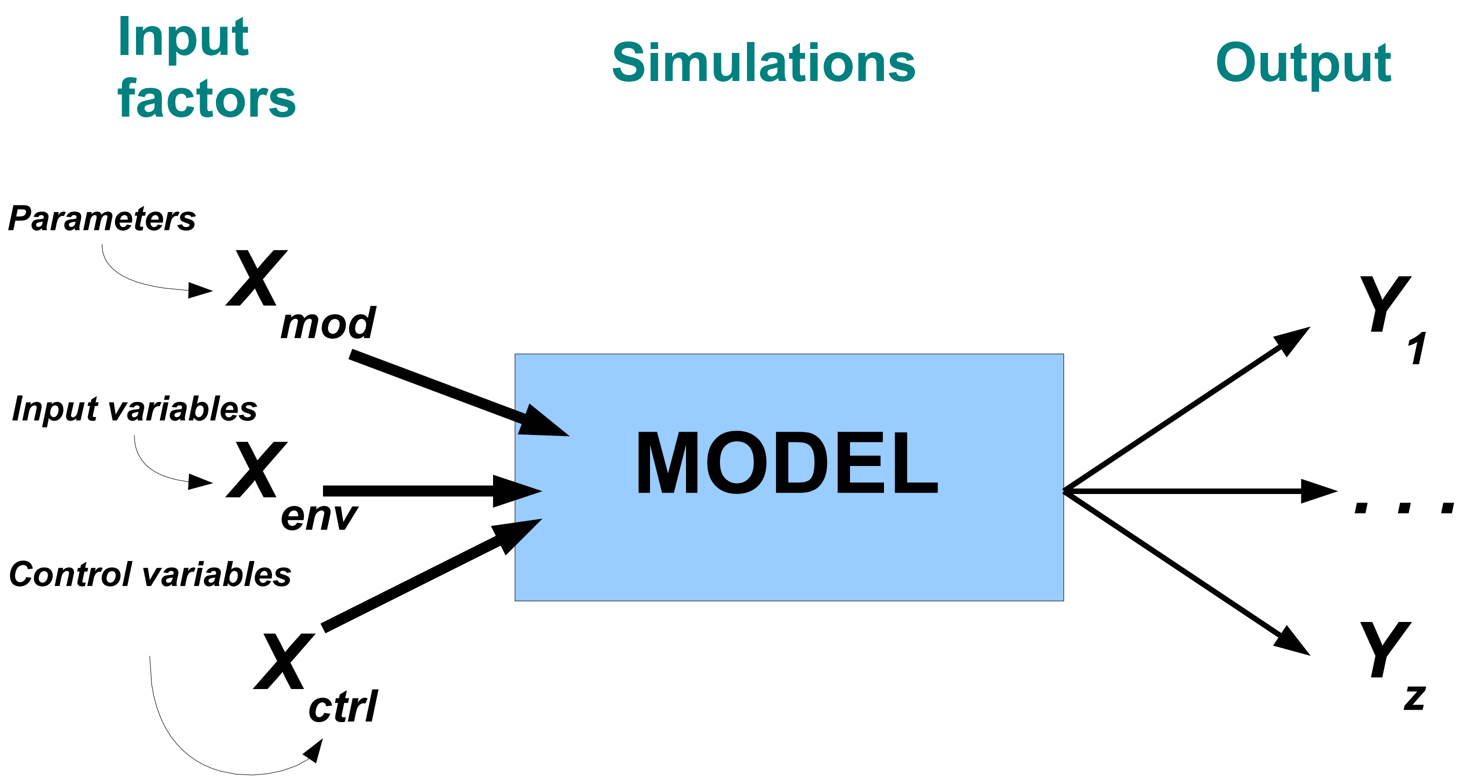
+

variety characteristics

+

many parameters

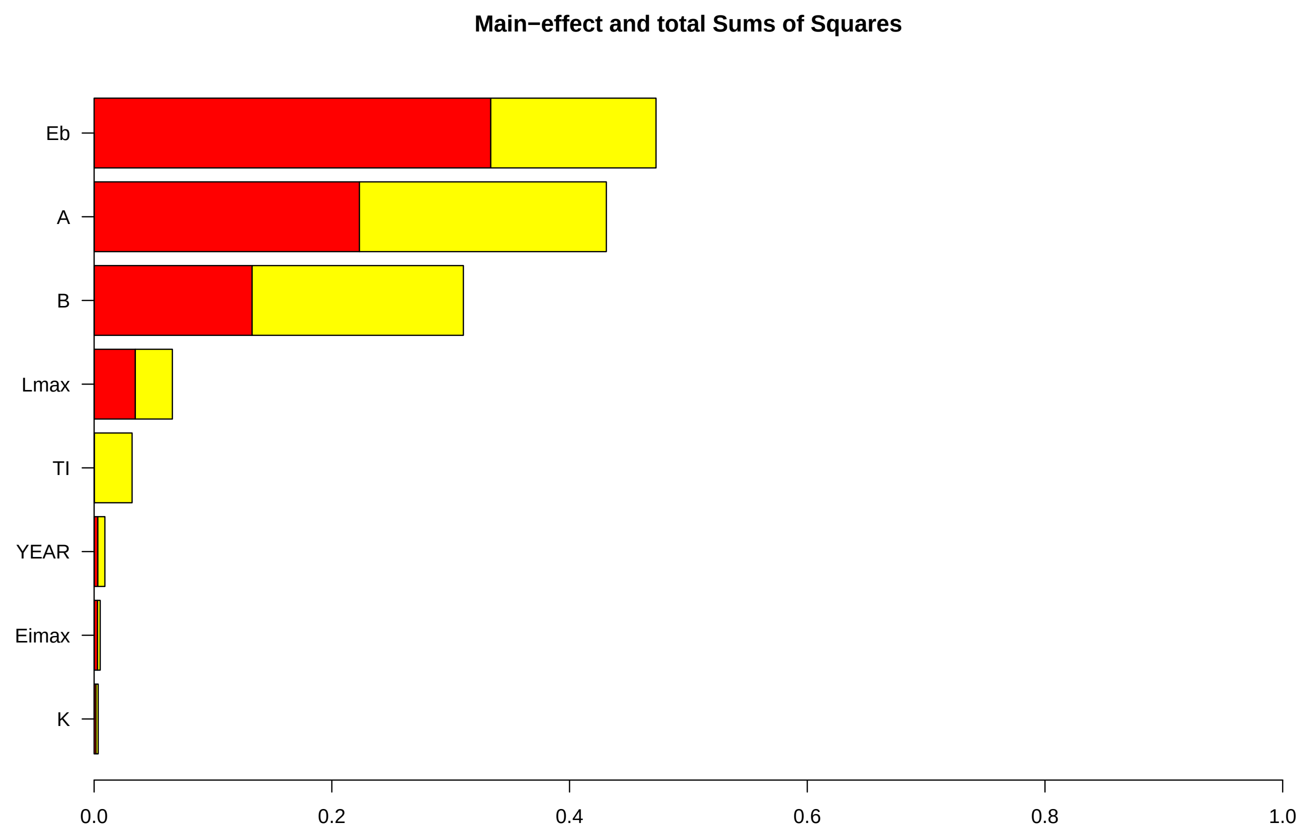
$\Longrightarrow$ MAPOD model $\Longrightarrow$ expected outcrossing rates


Input factors
Simulations
Output
Parameters
X mod
Input variables
X env
Control variables
X ctrl
MODEL
Y 1
. . .
Y z

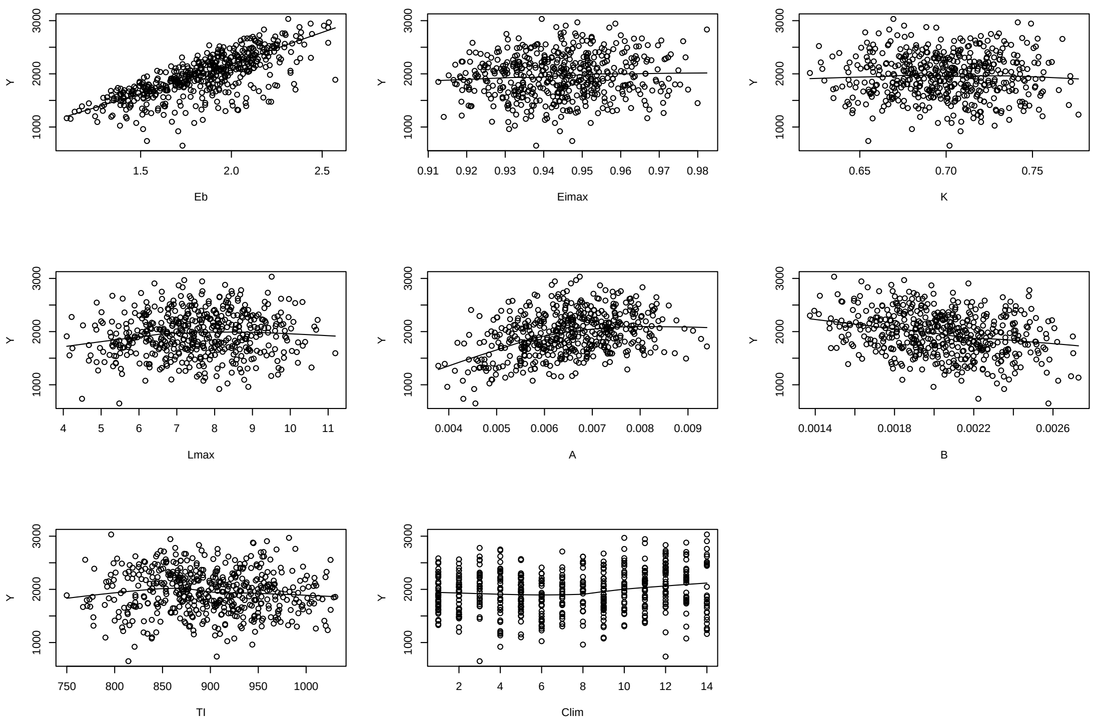
Main objectives of sensitivity analysis

- identify the most influent input factors
- quantify the contribution of each input factor to output variability
- better understand/check model behaviour

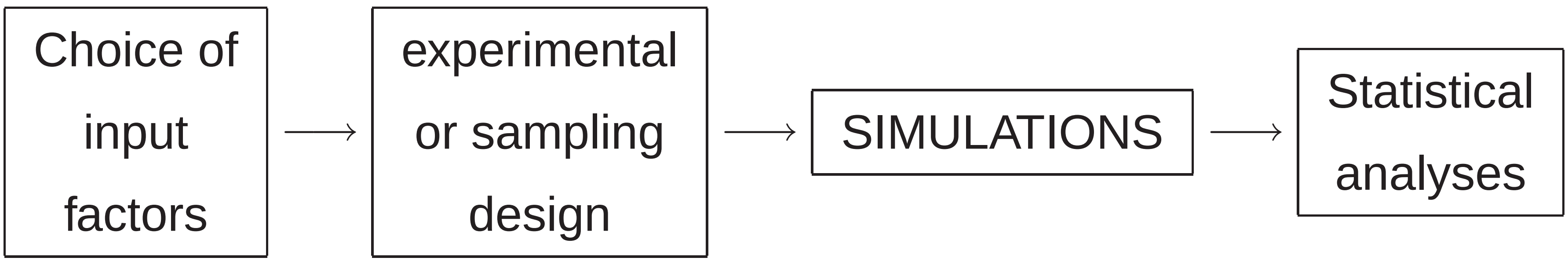
Example: Calculation of sensitivity indices



Example: Analysis and representation of output variability



Main steps of sensitivity analysis



Comparison with classical designs

Similarities

- temptation for one-factor-at-a-time
- analysis of variance approach
- *sparsity principle*
  - high-order interactions $\longrightarrow$ smaller (negligible) effects
  - few factors have a non-negligible effect

Differences

- large diversity of factor and possible run numbers
- no need for replication and blocking (if model is deterministic)
- *sparsity principle*
  - emphasis on dense exploration of the intervals of variation

Variance-based approaches

Sobol decomposition of the model function $\hat{Y} = f(x)$:

Input domain : $\Omega^n = \{ \quad x \quad | \quad 0 \leq x_i \leq 1 \quad ; \quad i = 1...n \quad \}$

$$f(x_1, .., x_s) \quad = \quad f_0 + \sum_{i=1}^{n} f_i(x_i) + \sum_{0 \leq i,j \leq n} f_{i,j}(x_i, x_j) + ... + f_{1,..,n}(x_1, .., x_n),$$

with

$$f_0 \quad = \quad \int_{\Omega^n} f(x)\, dx$$

$$f_1(x_1) \quad = \quad \int_{\Omega^{n-1}} f(x_1, x_2, \ldots, x_n)\, dx_2 \ldots dx_n - f_0$$

etc.

Orthogonality $\Rightarrow$ Variance decomposition of $\hat{Y}$

$$D = \int_{\Omega^n} f^2(x)\,dx - f_0^2.$$

$$D_1 = \int_0^1 f_1^2(x_1)\,dx_1$$

etc.

Sensitivity indices on Y:

$$S_{i_1,\ldots,i_k} = \frac{D_{i_1,\ldots,i_k}}{D} =$$

$$TS_i = 1 - \frac{D_{\sim i}}{D}.$$

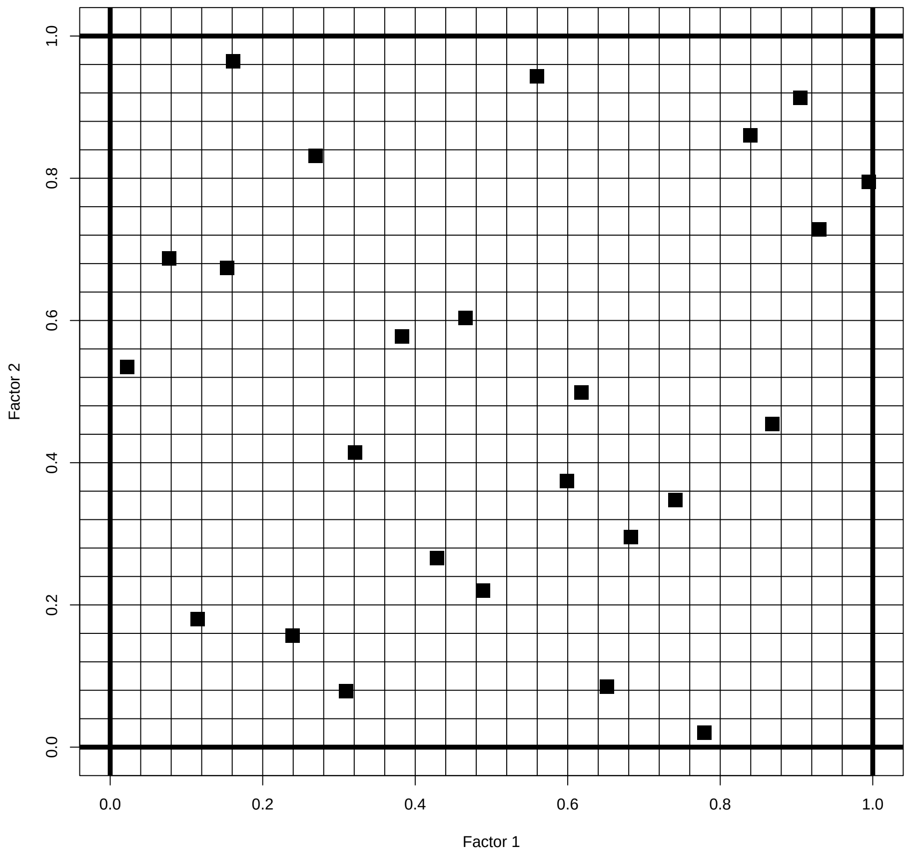
Design issues

or . . . trying to manipulate a toolbox without Rosemary!

- “factorial design + anova” useful but also
- *Latin hypercube* (McKay et al 1979)
- good lattices (different meaning)
- “quasi-replications” (Morris designs, “winding stairs”)
- random sampling and scrambling
- + quasi-Monte Carlo methods, uniform designs, optimality, discrepancy, . . .

Latin hypercube

Latin hypercube of order $s = 25$ for n factors :
one sample in each sub-interval of size $1/s$



Advantages :

- good exploration of each input factor

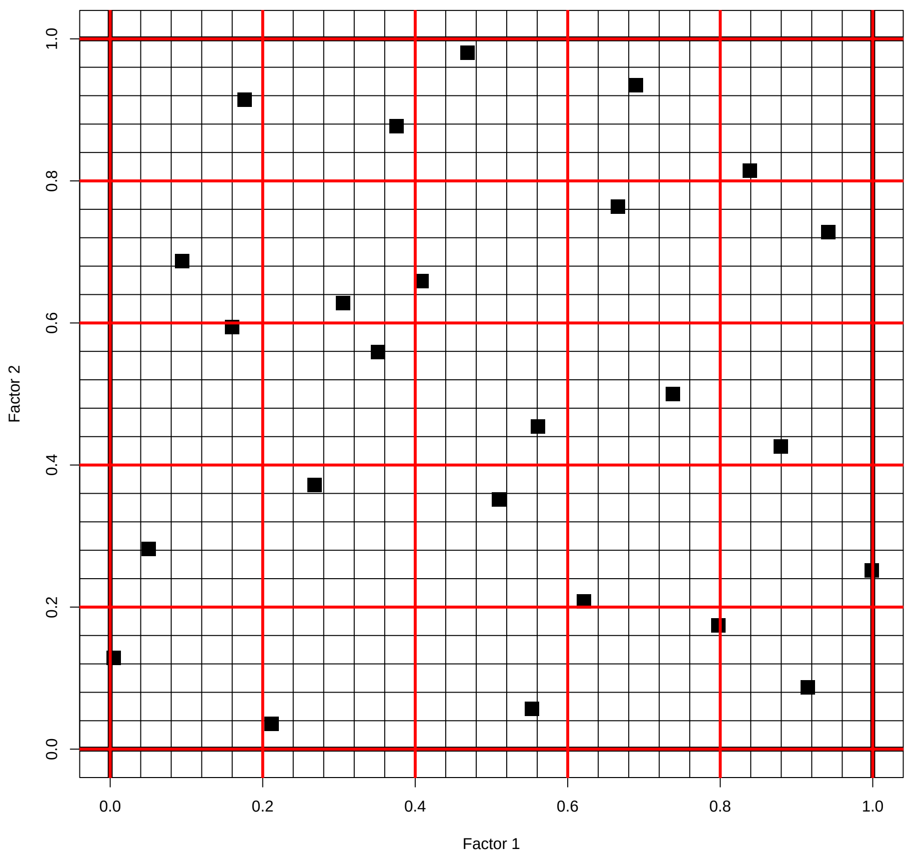
Drawback :

- no control on the marginals of order > 1

More structured Latin hypercubes

Motivations: better control of marginals of order > 1

Example 1: OA-based Latin hypercubes of order 5 (Tang, 1993)

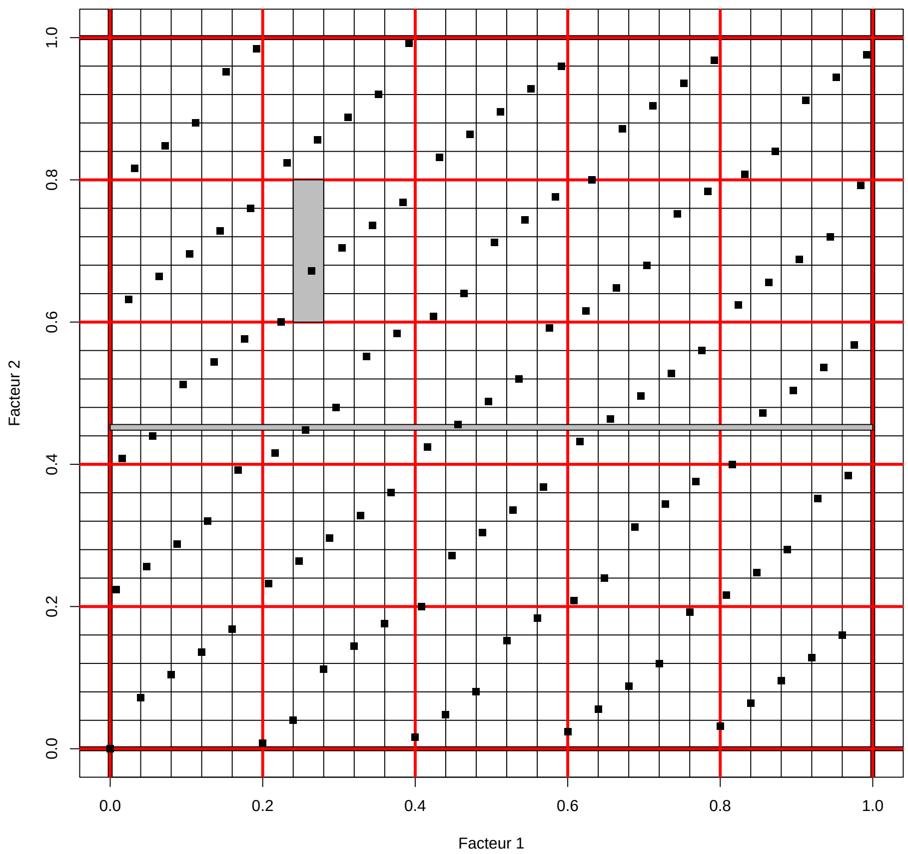


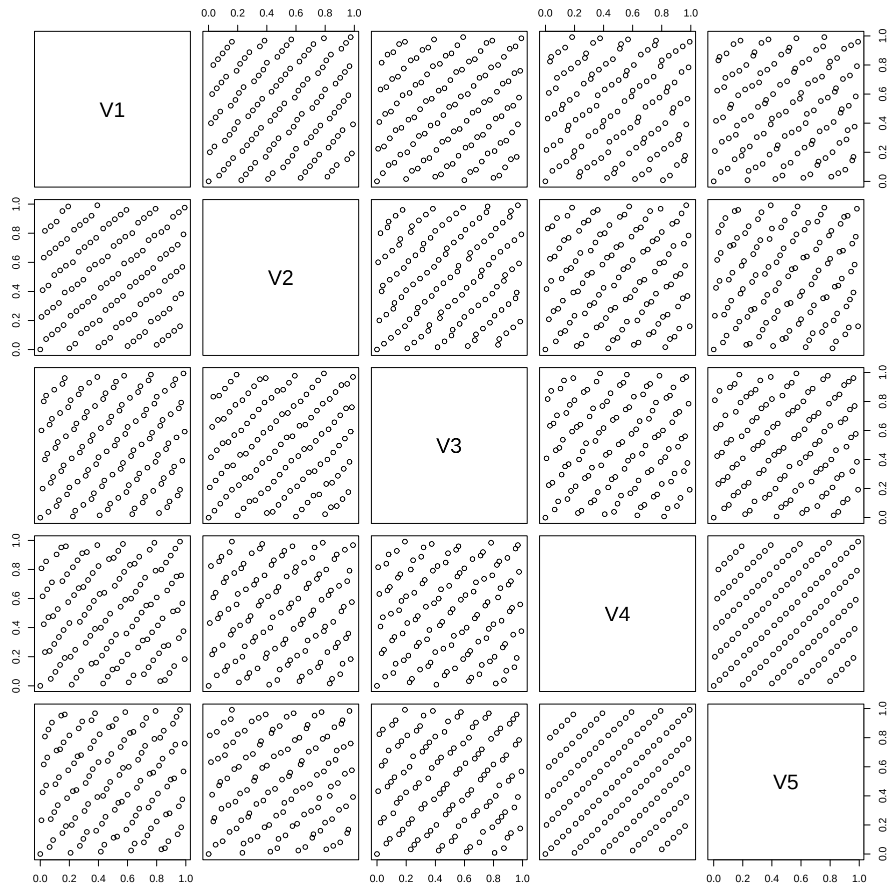
Properties:

one sample in each sub-interval of size $1/s$

one sample in each $1/s \times 1/s$ square, for any two factors

Example 2: $(0, m, s)$-net (Niederreiter, 1992); scrambled version (Owen, 1995)





Properties:

one sample in each sub-interval of size $(1/s)^m$

one sample in each $1/s \times \ldots \times 1/s$ hypercube, for any m factors

one sample in each hyperrectangle of volume $(1/s)^m$, for any $k \leq m$ factors

Applications to complex models

Just starting

- working groups at INRA (Mexico)
- much motivation from modellers
- Groupement De Recherche with wide area of applications (MascotNum)