Criteria for assessing space filling of a design, with emphasis on the stratification pattern

Ulrike Grömping

30 August 2024, COMPSTAT Gießen

BHT Berliner Hochschule für Technik

Table of contents

Thanks

Thanks to an anonymous reviewer for comments on

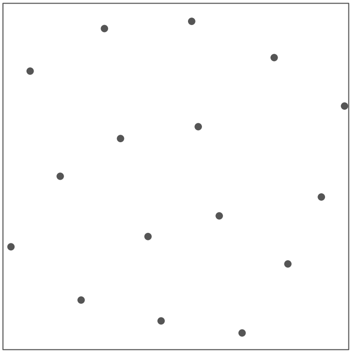
A new kid on the block: The stratification pattern for space-filling, with dimension by weight tables
which is available at
https://doi.org/10.1002/qre.3627

How does **design performance on various space-filling metrics** affect **prediction performance of the emulator model** ***in practice***?

Curse of dimensionality

16 points in 1D

16 points in 2D



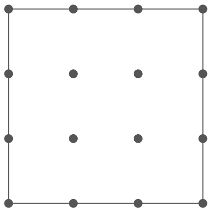
Computer experiments for quantitative variables:

- replication not needed (same input $\longrightarrow$ same output)
- easy to modify levels $\longrightarrow$ columns can have many levels
- large experimental space may need non-linear model, possibly non-parametric

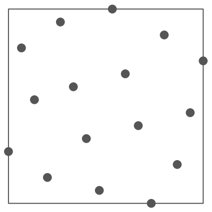
Latin hypercube design (LHD) in n runs: each column has levels 0 to $n-1$

Four 16 run designs in 2D

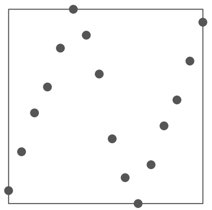
Orthogonal array (OA)



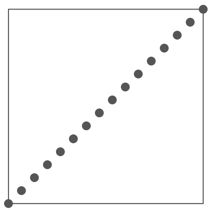
Optimized LHD



Less optimized LHD

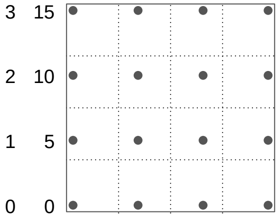


Completely aliased LHD

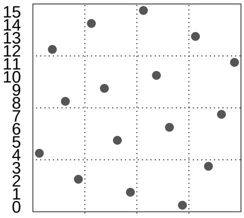


Level expansion / collapsing

Orthogonal array (OA)



Optimized LHD



LHD = expansion of OA (non-unique operation)

OA = collapsed version of LHD (unique operation)

Any expansion of the OA has a single point in each of the sixteen cells.

Table of contents

Designs with m columns in s^ℓ levels

For applicability of stratification pattern:

- each column has s^ℓ levels for some s and ℓ (e.g. 2^2 or 2^4),
- individual columns can be coarsened to $s^1, \ldots, s^{\ell-1}$ levels,
- one can consider
  - s to s^ℓ strata for individual columns (1D), e.g. 2, 4, 8, 16,
  - s^2 to $s^{2\ell}$ strata for pairs of columns (2D), e.g., 4, 8, 16, 32, 64, 128, 256,
  - … s^d to $s^{d\ell}$ strata for sets of $d < m$ columns (dD)
  - s^m to $s^{m\ell}$ strata involving all m columns (mD)
- Coarser stratifications and lower dimensions are considered more important.

Stratification pattern (Tian and Xu 2022)

Stratification pattern for given s and ℓ:

$$\left(S_1, S_2, \ldots, S_{m\ell}\right)$$

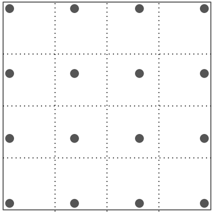
S_j measures imbalance from stratification into s^j strata
Perfect balance $\rightarrow S_j = 0$.

Consider, e.g., $16 = 2^4$ strata for the four example designs ($s = 2, j = 4$):

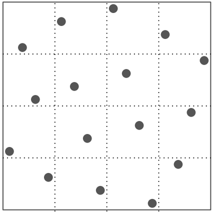
- 1D is perfectly balanced for all LHDs (and impossible for OA)
- 2D: 4×4 or 8×2 or 2×8 for all LHDs, 4×4 only for the OA

Stratification into $2^4 = 16$ strata

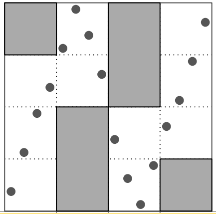
Orthogonal array (OA)



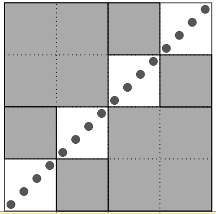
Optimized LHD



Less optimized LHD

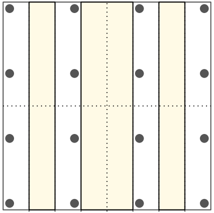


Completely aliased LHD

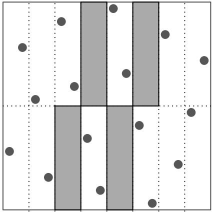


Stratification into $2^4 = 16$ strata

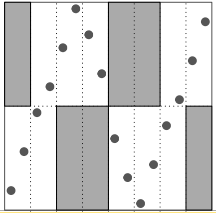
Orthogonal array (OA)



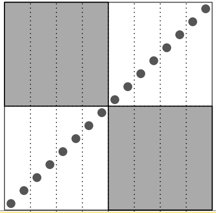
Optimized LHD



Less optimized LHD

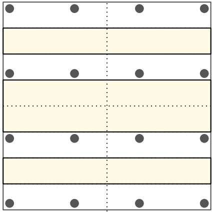


Completely aliased LHD

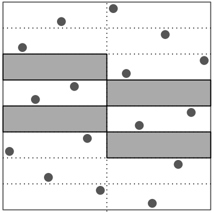


Stratification into $2^4 = 16$ strata

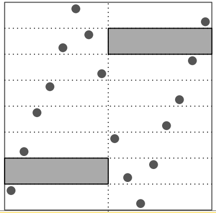
Orthogonal array (OA)



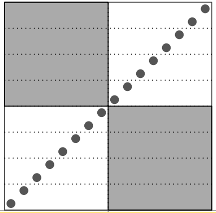
Optimized LHD



Less optimized LHD



Completely aliased LHD



S_j measures imbalances of stratifications into 2^j strata

	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	Sum
OA	0	0.0000	0.000	0.0000	.	.	.	.	0
LHD	.	.	.	.	.	.	.	.	.
opt.	0	**0.0000**	**0.000**	**1.0000**	**2.00**	**4**	**4**	**4**	15
less opt.	0	0.0625	0.625	0.5625	1.75	3	5	4	15
aliased	0	1.0000	0.000	2.0000	0.00	4	0	8	15

Best LHD worst LHD

Stratification aberration: better, if leftmost different pattern entry is smaller

- can only fairly compare designs with same numbers of levels
- possibly enhance by dimensional considerations

Refinement: Dimension by weight tables (Shi and Xu 2023)

LHD	dim	1	2	3	4	5	6	7	8
optimized	1D	0	0	0	0	.	.	.	.
.	2D	.	0	0	1	2	4	4	4
less opt.	1D	0	0	0	0	.	.	.	.
.	2D	.	0.0625	0.625	0.5625	1.75	3	5	4
aliased	1D	0	0	0	0	.	.	.	.
.	2D	.	1	0	2	0	4	0	8
	1D	2	4	8	16	.	.	.	.
	2D	.	2x2	2x4 4x2	2x8 4x4 8x2	2x16 4x8 8x4 16x2	4x16 8x8 16x4	8x16 16x8	16x16

Table of contents

Maximin distance and ϕ_p (based on Manhattan distance)

- **minimum distance** between pairs of points should be **maximal** (**maximin**), the larger the better (Johnson, Moore and Ylvisaker 1990)
- ϕ_p: minimizing ϕ_p serves as a norm-based emulation of maximin, the smaller the better (Morris and Mitchell 1995)

$$\phi_p(\mathbf{X}) = \left(\sum_{\{i,j\} \subset \{1,\ldots,n\}, i \neq j} d(\mathbf{x}^{(i)}, \mathbf{x}^{(j)})^{-p} \right)^{1/p}$$

Centered L_2 discrepancy (Hickernell 1998):
measures discrepancy from continuous uniform distribution on *m*-dimensional experimental hypercube,
in our example case on $[0, 15]^2$.

With z_{ij} referring to a design rescaled to $[0, 1]^m$, centered L_2 discrepancy is:

$$CD(\mathbf{X}) = \left(\frac{13}{12}\right)^m - \frac{2}{n}\sum_{i=1}^{n}\sum_{j=1}^{m}\left(1 + 0.5|z_{ij} - 0.5| - 0.5|z_{ij} - 0.5|^2\right) + \\ + \frac{1}{n^2}\sum_{i=1}^{n}\sum_{k=1}^{n}\prod_{j=1}^{m}\left(1 + 0.5|z_{ij} - 0.5| + 0.5|z_{kj} - 0.5| - 0.5|z_{ij} - z_{kj}|\right),$$

For a design **X** with *m* columns, the maximum projection criterion (Joseph, Gul and Ba 2015) considers **2D projections** and requires minimizing

$$\psi(\mathbf{X}) = \left(\frac{1}{\binom{n}{2}} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{1}{\prod_{k=1}^{m} (x_{ik} - x_{jk})^2} \right)^{1/m}.$$

Uniformity of projections

This criterion (Sun, Wang and Xu 2019) considers the average centered L_2 discrepancy over all pairs of columns, i.e., over all 2D projections.

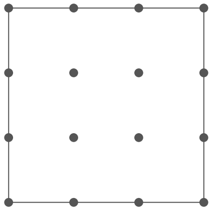
For the four above designs, it coincides with the centered L_2 discrepancy, because there is only one pair.

Design columns should be as uncorrelated as possible.

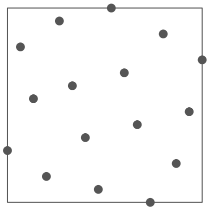
- **maximum absolute correlation** between columns (the smaller the better) or
- average absolute correlation or
- average squared correlation
- ...

or further demands, regarding correlation also with product columns (Ye 1998 design)

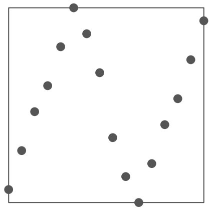
Orthogonal array (OA)



Optimized LHD



Less optimized LHD



Completely aliased LHD

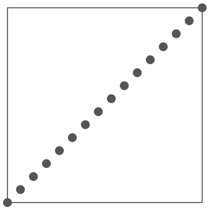


Table of all metrics

		small = good						
	min. dist.	ϕ_p	cent. discr.	max. proj.	S_2	S_3	S_4	max. abs. corr.
OA	5	0.247	0.1773	Inf	0.0000	0.000	0.000	0.000
OptLHD	**4**	**0.258**	**0.0460**	**0.111**	**0.0000**	**0.000**	**1.000**	**0.000**
LessOptLHD	3	0.376	0.0567	0.162	0.0625	0.625	0.562	0.138
AliasedLHD	2	0.599	0.1210	0.366	1.0000	0.000	2.000	1.000

Best LHD worst LHD

Table of contents

Six 64 run designs were tested:

- two LHDs from the 1998 Ye construction (one optimized for maximin distance), which have uncorrelated columns and even correlation zero between linear and bilinear columns
- two SOAs ("**S**trong" or "**S**tratified" **O**rthogonal **A**rrays) with eight-level columns that perform well on the stratification pattern
- expansions of the SOAs to LHDs, also called **G**eneral SOAs (GSOAs)

Metrics for the designs

	small = good							
	min. dist.	ϕ_p	cent. discr.	unif. proj.	max. proj.	S_3	S_4	max. abs. corr.
YeRandom	89	0.0131	0.159	0.0173	0.0265	1.88	20.4	**0.0000**
YeOpt	101	0.0121	0.162	0.0177	0.0153	1.88	19.8	**0.0000**
SOA_ST	108	**0.0110**	0.293	0.0820	Inf	**0.00**	**10.0**	**0.0000**
GSOA_ST	**120**	**0.0110**	**0.148**	**0.0157**	**0.0117**	**0.00**	**10.0**	0.0485
SOA_HT	99	0.0120	0.393	0.0880	Inf	0.00	47.0	0.3810
GSOA_HT	116	0.0116	0.244	0.0286	0.0190	0.00	47.0	0.3975

GSOA_ST is best for all criteria except maximum absolute correlation (but still acceptable).

Ye designs have optimal correlation properties, including further ones not captured in any of the metrics.

Dimension by weight tables for designs (weights go up to $9 \cdot 6 = 54$)

Design	Dimension	1	2	3	4	...	Sum
YeRandom	2D	.	0.00	1.88	7.81	...	2268
.	3D	.	.	0.00	6.56	...	328104
.	4D	.	.	.	6.00	...	31013766
YeOpt	2D	.	0.00	1.88	7.25	...	2268
.	3D	.	.	0.00	6.56	...	328104
.	4D	.	.	.	6.00	...	31013766
(G)SOA_ST	2D	.	**0**	**0**	**0**	...	2268
.	3D	.	.	**0**	**0**	...	328104
.	4D	.	.	.	10	...	31013766
(G)SOA_HT	2D	.	0	0	45	...	2268
.	3D	.	.	0	0	...	328104
.	4D	.	.	.	2	...	31013766

Simulation settings

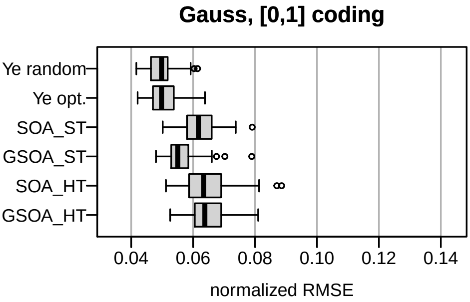
Borehole function (deterministic, 8 meaningful inputs on very different scales, 9th irrelevant input)

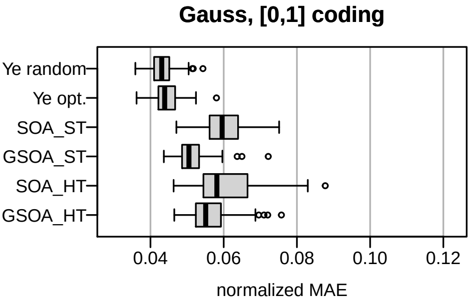
For emulator model:
log water flow modeled via kriging using function km from R package DiceKriging:

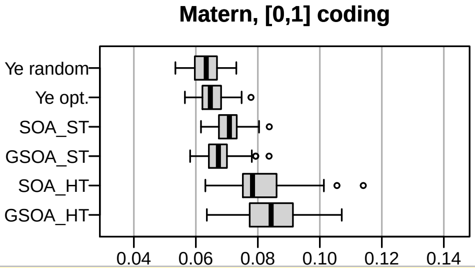
- optimizer BFGS
- up to 200 iterations with a nugget of 10^{-8}
- two alternative strategies of handling the huge differences between parameter ranges:
  - working on original scale with km's $\mathrm{parscale}$ argument
  - working on $[0, 1]^m$, adjusting the inputs in the borehole function
- Gauss or Matern covariance structure

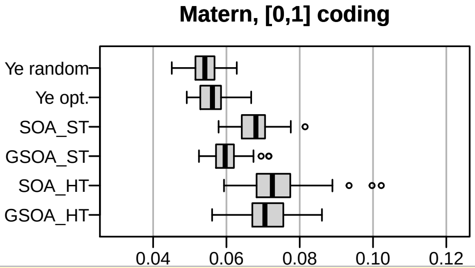
For each design under each of the four settings:
100 emulator models based on random allocations of variables to the design.

Results: Root MSE and MAE for [0,1] coding



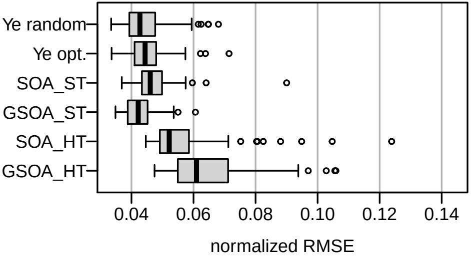




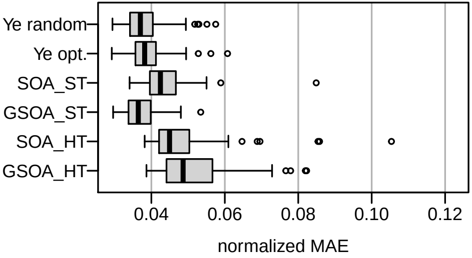


Results: Root MSE and MAE for original coding with parscale (better)

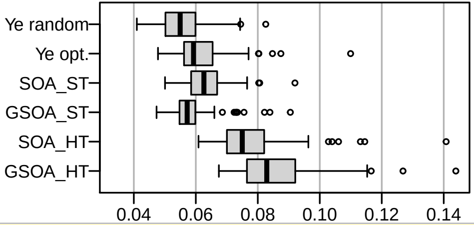
Gauss, original coding



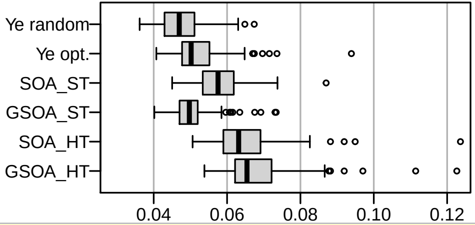
Gauss, original coding



Matern, original coding



Matern, original coding



Surprises from the simulation

- Ye designs better than GSOA_ST, especially for [0,1] coding
- GSOA_HT often worse than SOA_HT
- Random Ye design slightly better than the one with optimized maximin

Table of contents

GSOA_ST is best in terms of the metrics,
but outperformed by the Ye designs [0,1] coding

The good performance of the Ye designs is not sufficiently explained by the metrics considered - is the strong correlation performance key?

Such questions deserve more attention, with a view to guiding practitioners.

Difficulties: Prediction performance strongly depends on

- the benchmarking problem considered
- the modelling strategy (even relative performance within the same benchmarking problem)

It is not trivial to design meaningful comparison settings.

But it should be done!

Table of contents

References

Grömping, U. (2024). A new kid on the block: The stratification pattern for space-filling, with dimension by weight tables. *Quality and Reliability Engineering International* online early.

Hickernell F.J. (1998). A generalized discrepancy and quadrature error bound. *Mathematics of Computation* **67**, 299-322.

Johnson, M. E., Moore, L. M., and Ylvisaker, D. (1990), “Minimax and Maximin Distance Designs,” *JSPI* **26**, 131-148.

Joseph, V. R., Gul, E., and Ba, S. (2015) Maximum Projection Designs for Computer Experiments. *Biometrika* **102**, 371-380.

Morris, M. D. and Mitchell, T. J. (1995). Exploratory designs for computational experiments. *JSPI* **43**, 381-402.

Shi, C. and Xu, H. (2023). A Projection Space-Filling Criterion and Related Optimality Results. *JASA* (online early).

Sun, F., Wang, Y. and Xu, H. (2019). Uniform projection designs. *The Annals of Statistics* **47**, 641-661.

Surjanovic, S. and Bingham, D. (2013). Borehole function. http://www.sfu.ca/~ssurjano/borehole.html.

Tian, Y. and Xu, H. (2022). A minimum aberration-type criterion for selecting space-filling designs. *Biometrika* **109**, 489-501.

Wang, Y., Sun, F. and Xu, H. (2022). On Design Orthogonality, Maximin Distance, and Projection Uniformity for Computer Experiments. *JASA* **117**, 375-385.

Ye, K. Q. (1998), “Orthogonal Column Latin Hypercubes and Their Application in Computer Experiments,” *JASA* **93**, 1430–1439.