Extending Morris Method: identification of the interaction graph using cycle-equitabe designs

Jean-Marc Fédou, Maria João Torres Dolores Rendas

To appear in the *Journal of Statistical Computation and Simulation*
Vol. 00, No. 00, Month 20XX, 1–25

PAPER

Extending Morris Method: identification of the interaction graph using cycle-equitabe designs (Special Issue: SAMO-MASCOT)

J.-M. Fédou, M.-J. Rendas*

Laboratoire I3S, CNRS-UNS, 2000 rte de Lucioles, Sophia Antipolis, France

{fedou}{rendas}@i3s.unice.fr

+334929427{48}{14}

(*Received 00 Month 20XX; final version received 00 Month 20XX*)

The paper presents designs that allow detection of mixed effects when performing preliminary screening of the inputs of a scalar function of d input factors, in the spirit of Morris' Elementary Effects approach. We introduce the class of (d, c)-cycle equitable designs as those that enable computation of exactly c second order effects on all possible pairs of input factors. Using these designs, we propose a fast Mixed Effects screening method, that enables efficient identification of the interaction graph of the input variables. Design definition is formally supported on the establishment of an isometry between sub-graphs of the unit cube Q_d equipped of the Manhattan metric, and a set of polynomials in $(X_1, \ldots, X_d)$ on which a convenient inner product is defined. In the paper we present systems of equations that recursively define these (d, c)-cycle equitable designs for generic values of $c \geq 1$, from which direct algorithmic implementations are derived. Application cases are presented, illustrating the application of the proposed designs to the estimation of the interaction graph of specific functions.

Keywords: Sensitivity analysis; interaction graph; Morris method; elementary effects; clustered designs; second-order effects

AMS Subject Classification: 62K99, 05B30, 05C75, 05C90

1. Introduction

In this paper we present designs that support an extension of the Morris' method [1] that enables the estimation of the graph of interactions between pairs of input factors of a function. Being based on empirical statistics of mixed effects, much in the same manner as Morris' method is based on empirical statistics of elementary effects, our method inherits its appealing numerical complexity. Moreover, the interaction graph built by our method is more informative than the previously proposed FANOVA graph [12], possibly leading to simpler models, as our numerical examples show.

1.1 *Sensitivity analysis*

In sensitivity analysis, one wishes to characterise the dependency of an unknown function $f : \mathcal{A} \subset \mathbb{R}^d \to \mathbb{R}$ on each of its d input factors. In general, we know nothing about the function $f(\cdot)$, but can evaluate it at chosen locations $\xi \in \mathcal{A}$. Most commonly interest is

*Corresponding author

on partitioning the factors of $f(\cdot)$ into those that have no impact on the function value (class $\mathcal{C}_0$), those that have a linear effect (class $\mathcal{C}_1$) and those that are non-linear or have interactions with other input factors (class $\mathcal{C}_2$). As noted in [9], fast screening is often done in the context of factor fixing, with the goal of restricting subsequent analysis of $f(\cdot)$ to the smaller set $\mathcal{C}_2$. This is the context we address.

1.2 *Elementary Effects method*

Morris scheme for One At a Time (OAT) designs for sensitivity analysis [1] is widely used for rapid identification of the groups of important (further classified into linear or mixed/non-linear) and unimportant inputs of a multivariate function $f(x), x \in \mathcal{A} \subset \mathbb{R}^d$, and is particularly relevant for models whose execution is computationally expensive and time consuming [2]. Although Morris' elementary effects do not have the nice theoretical properties of Sobol indexes, the fact that they can be efficiently computed makes them a valuable tool for preliminary sensitivity analysis, as some comparative studies (see e.g. [9]) as well as their wide adoption confirm.

Morris method [1] implements statistical tests over a set of finite differences along each principal direction i of a function's input space, $d_i(\xi)$, computed at a set of points $\{\xi_n^i\}_{n=1}^r$ of the input domain:

$$d_i(\xi) = \frac{1}{\Delta}\left[f(\xi + \Delta e_i) - f(\xi)\right], \quad \xi \in \mathcal{A},\ i = 1, \ldots, d \ , \tag{1}$$

where $\Delta > 0$ is an increment, and e_i is the vector with components $e_{i_j} = \delta_{ij}, j = 1, \ldots, d$. Let (μ_i, σ_i^2) be empirical estimates of the mean $\overline{d}_i$ and variance s_i^2 of d_i obtained from the set $\{d_i(\xi_n^i)\}_{n=1}^r$.

Input factors are classified as $(\mathcal{C}_0)$ negligible, $(\mathcal{C}_1)$ linear, or $(\mathcal{C}_2)$ non-linear/interaction if: $(\mathcal{C}_0)$ their mean and variance are both close to zero, $(\mathcal{C}_1)$ the mean is non-zero, but variance is small, or $(\mathcal{C}_2)$ variance is large. The revised version of Morris method in [3] replaces μ_i by $\mu_i^\star$, the empirical mean of $|d_i(\xi)|$, improving the robustness for oscillating functions, with derivatives of alternating sign.

If points $\{\xi_n^i\}_{n=1,i=1}^{r,d}$ are chosen completely at random, the sensitivity analysis of a function of d variables requires a total of $2dr$ evaluations of $f(\cdot)$. The basic Morris scheme is a One-At-a-Time (OAT) method that increases efficiency with respect to random sampling by using most evaluations of $f(\cdot)$ twice. It relies on empirical moment estimates using r samples of $\{d_i(\cdot)\}_{i=1}^d$ computed along r randomly oriented paths T_{d+1} along which each one of the d coordinates is changed at a time, see Figure 1. The total number of evaluations of $f(\cdot)$ is $r(d+1)$, linear in the number of input factors, irrespective of the number of levels that each input factor can assume.

Morris clustered designs, see Section 5 in [1], improve on the efficiency of these OAT designs by using each value of f in the computation of more than two elementary differences. The simple paths T_{d+1} are replaced by denser graphs that enable determination of $m \geq 1$ elementary differences along each direction. Surprisingly, they seem to have attracted much less interest than the standard ($m = 1$) version, possibly partially justified by the lack of a constructive method for finding them, and also, presumably, due to concerns about the impact of the induced residual correlation amongst the resulting set of elementary effects, see [4, 8]. The former difficulty has been addressed in [8], where we introduced the concept of edge equitable designs and presented procedures for their recursive construction for arbitrary values of k and $m \leq 2^{d-1}$, while a thorough study of the latter issue has not yet been, as far as we know, published in the literature.

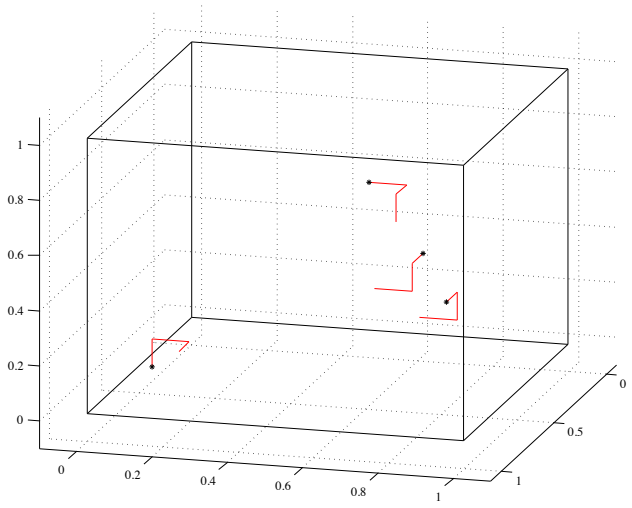


Figure 1. OAT method: a complete set of d elementary effects is computed along a trajectory contained in a scaled and translated version of Q_d.

1.3 *The New Morris Method*

In [6] the authors extend Morris' original method, which was restricted to analysis of first-order partial derivatives, presentig designs that enable efficient computation of two-factor interaction effects, of the form X_iX_j. Their detection of terms in the product of two input factors X_i and X_j is based on a numerical approximation of the second-order cross derivatives at points $\xi \in \{\xi_n^{ij}\}_{n=1}^r$ by the mixed effects $dd_{ij}(\xi)$:

$$\frac{\partial f(\xi)}{\partial X_i \partial X_j} \simeq dd_{ij}(\xi) = \frac{f(\xi + e_i\Delta + e_j\Delta) - f(\xi + e_i\Delta) - f(\xi + e_j\Delta) + f(\xi)}{\Delta^2} \ . \quad (2)$$

where for the sake of simplicity a regular grid spacing Δ has been assumed. For a design to enable assessment of all $d(d-1)/2$ possible interactions, it is necessary that it contains the corresponding 4-tuples $\{\xi, \xi + e_i\Delta, \xi + e_j\Delta, \xi + e_i\Delta + e_j\Delta\}$. We remark that these 4-tuples define a 4-cycle in $\xi + \Delta Q_d$, where Q_d is the unit-cube of dimension d.

The designs used by the New Morris Method are a slight modification of block designs on graphs, i.e., decompositions of a given graph G into copies of a basic graph S (blocks) where each edge of G appears in the same number of blocks. Rephrasing the problem of finding designs that can compute all the $\binom{d}{2}$ second-order cross derivatives in this graph theoretical framework, corresponds to taking $G = K_d$, the completely connected graph of size d, each node $i \in \{1, \ldots, d\}$ representing the presence of and edge oriented along

X_i. By constraining their designs to be "multiple trajectories", which are composition of several OAT (Morris' type) trajectories, the authors rephrase their problem as a graph decomposition problem: OAT trajectories are easily seen to correspond to identifying S with the graph of simple paths, and thus the overall goal is to decompose K_d into the smallest possible number of edge-disjoint paths. The problem is recognised as a particular instance of the handcuffed designs problem where each edge appears only once and the length of the paths is equal to d, and the authors rely on existing results from graph theory to establish existence of the multiple trajectories required to compute the complete set of second order differences, and provide rules for their construction.

Since the block design guarantees only the existence of consecutive edges aligned over all pairs of directions (i, j), of the generic form $\{x, x+e_i\Delta, x, x+e_i\Delta+e_j\Delta\}$, and second-order differences require the existence of the complete 4-cycles in the unit hyper-cube Q_d, points $(x+e_j\Delta)$ must be added to the block design to enable computation of the total set of first order differences required around each point. With an analysis based on a folded representation of the hypercube (in fact only edge orientation is explicitly represented, edges implicitly indicating paths of length 2), the overall topological properties of the resulting design in Q_d cannot be properly observed or controlled. In particular, the actual set of edges of the resulting subgraph of Q_d is difficult to describe.

1.4 *FANOVA graph*

Recently, some authors [12] have proposed the use of pair-wise total variation indexes to build a graph describing the pair-wise interactions (the FANOVA graph) of the input factors of a function. The authors show that subsequent modelling benefits by exploitation of the structure of this graph. In particular, the FANOVA graph is intended as a tool supporting structural kernel design in the context of non-parametric modelling using Gaussian processes (kriging), whose complexity and performance can improve if the kernel's structure closely reflects the clique structure of the FANOVA graph. Being based on Sobol indices, the set of total variation indices which the FANOVA graph describes are computed by Monte Carlo techniques, and thus they inherit the complexity of Sobol indices.

1.5 *Our contributions*

We present generic families of graphs that allow computation of exactly $c \geq 1$ mixed effects in all pairs of directions of $\mathbb{R}^d$, in the vicinity of a given point of $\mathbb{R}^d$, allowing the extension of Morris method to the analysis of cross second-order derivatives, and thus the detection of mixed terms, of the form X_iX_j, in the function $f : \mathbb{R}^d \to \mathbb{R}$ under analysis. We call the graphs that possess this property (d, c)-cycle equitables.

Figure 2 presents an illustration of this class of graphs, showing an example of a $(5, 1)$-cycle equitable graph, i.e., that has exactly one cycle of size 4 involving all 10 possible pairs of the 5 input factors. In this Figure, as in the rest of the paper, subgraphs of Q_d are represented using edge labels/colours to indicate the direction along which they are aligned (there will be d different labels/colours in subgraphs of Q_d).

In Section 3, Theorems 3.2, 3.3/3.4 and 3.6, we give recursive definitions of families of cycle-equitable designs for $c = 1$, $c = 2$ and $c \geq 3$, respectively. To our knowledge, this is the first time that a constructive and generic procedure to generate cycle-equitable graphs is presented in the literature.

The work is based on the polynomial representation of subgraphs of the hypercube and an associated algebra introduced in Section 2. As in [8], our construction is recursive, (d, c)-cycle equitable graphs H_c^d being defined as composition of a $(d-1, c)$-cycle equitable

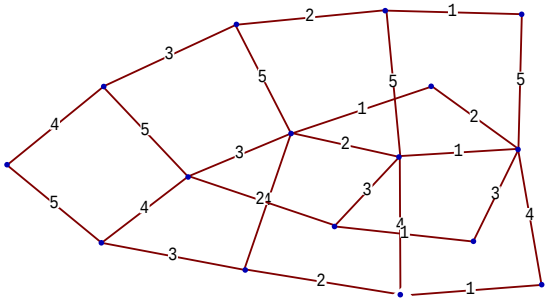


Figure 2. A $(5,1)$-cycle equitable subgraph of Q_5.

graph with a "lifted" version of a subgraph G_c^{d-1} of H_c^{d-1}, which in terms of their polynomial representations is written as

$$H_c^d = H_c^{d-1} + X_d G_c^{d-1}, \qquad G_c^{d-1} \subset H_c^{d-1} \ .$$

The basic idea is illustrated in Figure 3, for $d = 4$ and $c = 1$. From this generic recursion, necessary and sufficient conditions for cycle-equitability of H_c^d are derived, supporting the proposed algorithms for design definition.

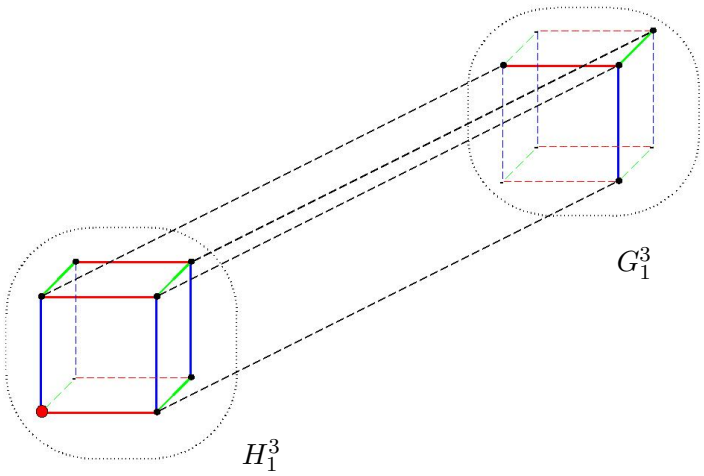


Figure 3. Recursive construction of a $(4,1)$-cycle equitable subgraph by composition of a $(3,1)$-cycle equitable graph (H_1^3) and a lifted version of $G_3^3 \subset H_1^3$. The dashed edges are along X_4.

Our designs are related to those supporting the "New Morris method" presented in [6, 7]. The work proposed here improves the work of these references in several directions: *(i)* our solution is more general, since we handle values of $c \geq 1$, that is, we define designs that enable computation of more than one mixed effect for all the $\binom{d}{2}$ pairs of input factors; *(ii)* our solution is more efficient: even if the authors' solution is optimal – in terms of design size– inside the constrained class of "multiple trajectories", restriction to this class compromises the attainable efficiency, and the $(d,1)$-cycle equitable designs defined in Theorem 3.2 are smaller; *(iii)* equitability: our designs are provably edge and cycle equitable, which is not the case of the designs in [6, 7]; *(iv)* recursivity: even if we do not

present here these results, the fact that our designs are recursively defined in d, enables analytical determination of the sets of points that contribute to each elementary or mixed effects, while such expressions do not seem to be easy to obtain for the handcuffed designs of [6].

As shown in Section 5, this extended Morris method enables the estimation of an annotated interaction graph, that characterizes the interactions of all possible pairs of input factors, being an efficient alternative to the FANOVA graphs defined in [12]. In fact, as the examples presented in Section 5 show, exploitation of the additional information in our interaction graphs can lead to simpler and potentially better models (as well as more efficiently identified) for the function under study.

2. Polynomial graph representation and algebra

2.1 *Polynomial representation of subgraphs of Q_d*

Definition 2.1 Subgraphs of the unit hypercube $Q_d = \{0,1\}^d$, are graphs whose vertices belong to Q_d, two elements being joined by an edge if and only if they differ in exactly one coordinate. ■

Note that with the definition above a subgraph of Q_d is uniquely defined by its node set. For example, the graph corresponding to the node set $S = \{(0,0),(1,1)\}$ has an empty edge set, while the graph $T = \{(0,0),(0,1)\}$ has a single edge. We will adhere to the definition usual in graph theory, the size of a graph being equal to its number of nodes.

Definition 2.2 Edge coloring of Q_d. We define an edge-coloring of Q_d by saying that the edge joining nodes s and s' has color i when they differ in the i-th coordinate ($s_i \neq s'_i$). ■

Note that from the definition of subgraphs of Q_d, the two nodes linked by an edge of color i have $s_j = s'_j$ for all $j \neq i$. The graph T in the example above has a single edge of color 2, as the second coordinate changes.

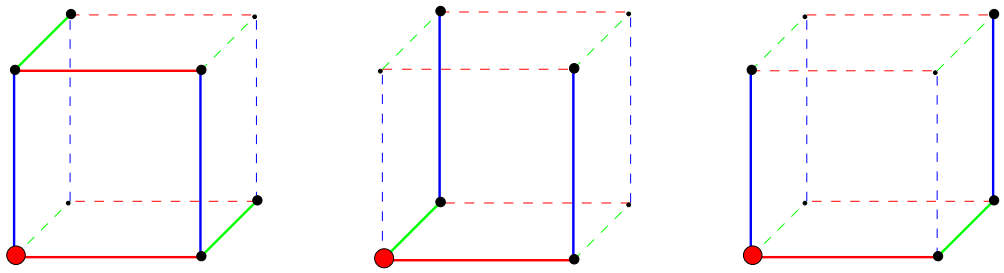
We associate to each $s \in Q_d$ a monomial $\mathcal{P}_s$ in the ring $\mathbb{R}[X_1,\ldots,X_d]$ of the polynomials over the variables $X_1,\ldots,X_d$:

$$s = \{s_1,\ldots,s_d\} \longrightarrow \mathcal{P}_s(X_1,\ldots,X_d) = X_1^{s_1}\ldots X_d^{s_d} \ .$$

Definition 2.3 Polynomial associated with a subgraph of Q_d.
The subgraph induced by a set $S \subset Q_d$ is the polynomial $\mathcal{P}_S = \sum_{s\in S}\mathcal{P}_s$. The empty set is represented by the zero polynomial. ■

For instance, the graph (a) in Figure 4 has polynomial representation $1 + X_1 + X_2 + X_1X_2 + X_2X_3 + X_1X_3$ (the origin of Q_3 is indicated by the large red node, and the following colour code is used for the edges: red $\leftrightarrow X_1$; blue $\leftrightarrow X_2$ and green $\leftrightarrow X_3$.

The set of the polynomials representing simple subgraphs of Q_d will be denoted by K_d, and corresponds to the polynomials of degree at most 1 (square-free polynomials) in each variable having coefficients in $\{0,1\}$.



(a) Graph A, $(3,2)$ edge-equitable.

(b) Graph B, not edge-equitable.

(c) Graph C, not edge-equitable.

Figure 4. Origin is indicated by the large red nodes. Edge colours indicate their direction: red $\leftrightarrow X_1$, blue $\leftrightarrow X_2$, green $\leftrightarrow X_3$. Graph C is the reflection of graph B along X_1.

2.2 *Scalar product in K_d*

The set K_d can be embedded in the algebra $\mathbb{R}[X_1,\ldots,X_d]/\{X_i^2{=}1, i=1\ldots d\}$ of the quotient polynomial ring induced by the ideal generated by the set $\{X_i^2-1\}_{i=1}^d$. This algebra is a vector space for which the set of monomials can be taken as a natural basis. By defining a scalar product such that this basis is orthogonal, we endow K_d with a structure that has several interesting properties in term of the underlying subgraphs of Q_d.

Definition 2.4 We define the scalar product between monomials $\mathcal{P}_s, \mathcal{P}_{s'} \in K_d$ as[1]

$$\langle \mathcal{P}_s, \mathcal{P}_{s'} \rangle = 1_{s=s'} \ ,$$

and extend it naturally to the entire K_d by bilinearity

$$\langle \mathcal{P}_S, \mathcal{P}_{S'} \rangle = \sum_{s\in S, s\in S'} \langle \mathcal{P}_s, \mathcal{P}_{s'} \rangle \, , \qquad \mathcal{P}_S, \mathcal{P}_{S'} \in K_d \ .$$

■

Addition coincides with the usual polynomial addition, corresponding to union of the corresponding graphs. Note that this may yield graphs that are not simple (multiple nodes). If G and H are two subgraphs of Q_d having polynomial representations $\mathcal{P}_G$ and $\mathcal{P}_H$ respectively, then $\mathcal{P}_G \oplus \mathcal{P}_H = \mathcal{P}_{G\cup H}$. Remark that $\oplus$ is the usual sum of polynomials using the binary *or* operation as addition in the field of coefficients.

Multiplication, however, must be defined under the identification $X_i^2 = 1, i = 1,\ldots,d$, such that the resulting polynomials correspond to subgraphs of Q_d.

The following lemmas can be easily demonstrated.

LEMMA 2.5 *The scalar product of two subgraphs of Q_d, S_1 and S_2, is equal to the size of their intersection: $\langle \mathcal{P}_{S_1}, \mathcal{P}_{S_2} \rangle = |S_1 \cap S_2|$. In particular, $\langle \mathcal{P}_S, \mathcal{P}_S \rangle = |S|$.* ■

[1]Notation 1_a denotes the Kroenecker symbol, begin equal to one if proposition a is true and zero otherwise.

LEMMA 2.6 *Let $s \in Q_d$ and $S \subset Q_d$. The subgraph S' defined by $\mathcal{P}_{S'} = \mathcal{P}_s\mathcal{P}_S$ is the reflection of S along the directions present in s. In particular, $X_i\mathcal{P}_S$ corresponds to the mirror of S along direction i.* ■

LEMMA 2.7 *The scalar product is invariant with respect to multiplication by a polynomial: for all $s \in Q_d$, $S, S' \subset Q_d$ $< \mathcal{P}_s\mathcal{P}_S, \mathcal{P}_s\mathcal{P}_{S'} >=< \mathcal{P}_S, \mathcal{P}_{S'} >$.* ■

Using Lemmas 2.6 and 2.5 the following is immediate.

LEMMA 2.8 *The number m_i of edges of $S \subset Q_d$ having color i satisfies*

$$\langle \mathcal{P}_S, X_i\mathcal{P}_S \rangle = 2m_i, \qquad i \in \{1, \ldots, d\} \ . \tag{3}$$

■

In [8] we defined the class of (d, m)-edge equitable designs, see Definition 2.9 below. These graphs correspond to the clustered designs introduced in [1] and allow the computation of m elementary effects (1) along each of the directions of the input space.

Definition 2.9 Let $G \subset Q_d$. We say that G is a (d, m)-edge equitable graph if and only if G has exactly m edges in every coordinate $i \in \{1, \ldots, d\}$. ■

The graphs in Figure 4 illustrate the Lemmas above and Definition 2.9. Graph A is identified with the polynomial $\mathcal{P}_A = 1 + X_1 + X_2 + X_1X_2 + X_1X_3 + X_2X_3$. It is a $(3, 2)$-edge equitable graph since

$$\langle \mathcal{P}_A, X_1\mathcal{P}_A \rangle = \langle \mathcal{P}_A, X_2\mathcal{P}_A \rangle = \langle \mathcal{P}_A, X_3\mathcal{P}_A \rangle = 4 \ .$$

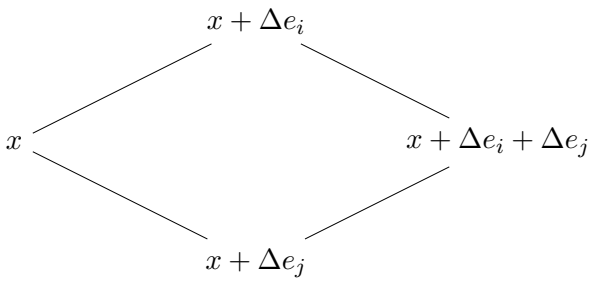
Graphs B and C are the reflection of each other along the red (X_1) direction, since $\mathcal{P}_C = 1 + X_1 + X_2 + X_1X_3 + X_1X_2X_3 = X_1\mathcal{P}_B, (\mathrm{mod}\, X_i^2 = 1)$. Note that the number of edges of each colour is the same for these two graphs. They are not edge-equitable since

$$\langle \mathcal{P}_B, X_i\mathcal{P}_B \rangle = \langle \mathcal{P}_C, X_i\mathcal{P}_C \rangle = 2, \ i = 1, 3, \qquad \langle \mathcal{P}_B, X_2\mathcal{P}_B \rangle = \langle \mathcal{P}_C, X_2\mathcal{P}_C \rangle = 4 \ .$$

Using the polynomial representation introduced in this section, we presented in [8] algorithms for recursive construction of edge equitable designs for arbitrary values of d and m and studied their complexity.

3. Cycle-equitable graphs

As seen in section 1.3, the computation of a mixed effect along directions (i, j) requires that the design graph contains a 4-cycle, as in the diagram



In the previous section we remembered the definition of (d, m)-edge equitable graphs, that allow the computation of a given number m of all d elementary effects. In this section we address the problem of finding subgraphs of Q_d that have exactly c 4-cycles in all pairs of input factors, allowing the computation of exactly $c \geq 1$ mixed effects (2).

Definition 3.1 Let $G \subset Q_d$. We say that G is a (d, c)-cycle equitable graph if and only if G has exactly c 4-cycles in every pair of coordinates $(i, j) \in \{1, \ldots, d\}^2,\ i \neq j$.

We reserve the notation H_c^d for (d, c)-cycle equitable graphs. Using the polynomial algebra introduced above, we can rephrase the previous definition by saying that G is (d, c)-cycle equitable if and only if

$$|\mathcal{P}_G \cap X_i\mathcal{P}_G \cap X_j\mathcal{P}_G \cap X_iX_j\mathcal{P}_G| = 4c, \qquad i \neq j \in \{1, 2, \ldots, d\}^2 \ .$$

To simplify the presentation we sometimes will, as in the equation above, identify a subset S of Q_d and its corresponding polynomial $\mathcal{P}_S$, using for instance expressions such as X_iG where G is a graph, as a shorthand notation for $X_i\mathcal{P}_G$. Naturally, all operations should be understood with respect to the representation for which they are defined: products and sums will always involve the corresponding polynomials, while unions and intersections are meant to be applied to sets and induced hypercube subgraphs.

Our construction of cycle-equitable graphs is recursive. Starting with a suitable initialisation (for a value of d that depends on c, as we will see below), we set

$$H_c^{d+1} = H_c^d + X_{d+1}G_c^d, \qquad \text{where} \qquad G_c^d \subset H_c^d \ . \tag{4}$$

Figure 5 illustrates the recursion for $c = 1$ and $d = 2, \ldots, 5$.

For H_c^{d+1} defined by (4) to be a $(d+1, c)$-cycle equitable graph whenever H_c^d is (d, c)-cycle equitable, G_c^d must satisfy the following conditions:

Cond0. $G_c^d \subset H_c^d$. This condition is already explicit in (4). Its assumption is justified by the desire to maintain the size of the graphs small (requiring the addition of a minimal number of nodes). It is repeated here to emphasise that it is assumed to hold in the discussion of the two conditions below.

Cond1. G_c^d must be (d, c)-edge equitable, i.e. have exactly c edges along each X_j, $j \in \{1, \ldots, d\}$. Together with $G_c^d \subset H_c^d$ this implies that c new 4-cycles in all pairs (X_{d+1}, X_j), $j \in \{1, \ldots, d\}$ will be induced by composition (4). In terms of the polynomial representation this condition is expressed as

$$\left\langle \mathcal{P}_{G_c^d}, X_i\mathcal{P}_{G_c^d} \right\rangle = 2c, \qquad \forall i \in \{1, \ldots, d\} \ . \tag{5}$$

Cond2 G_c^d must have no 4-cycles, otherwise H_c^{d+1} would have more than c cycles in some pairs of coordinates, and thus would no longer be a $(d+1, c)$-cycle equitable subgraph.

In the following subsections we provide expressions for graphs G_c^d that satisfy these conditions, considering separately the cases of $c \leq 2$ and $c \geq 3$. Before, we introduce the following compact notations that will help us simplifying our presentation.

- $X_i^j = \prod_{i \in I} X_i$, where $I = \{\min(i,j), \cdots, \max(i,j)\}$.
- Let $\mathcal{P}_i^j = \sum_{k=i}^{j} X_i^k$. Note that $\mathcal{P}_0^j$ is well defined if we make the convention $X_0 = 1$.
- If $\pi = (\pi(1) \cdots \pi(d))$ is a permutation of $\{1, 2, \ldots, d\}$, the polynomial $\mathcal{P}_\pi$ associated to π is defined by

$$\mathcal{P}_\pi = \mathcal{P}_\pi(X_1, \ldots, X_d) = \sum_{i=0}^{d-1} X_{\pi(1)} \cdots X_{\pi(i)} \ .$$

- For a permutation σ of $\{1, \ldots, d\}$ we denote by $\sigma^{(k)}$ the permutation of $\{1, \ldots, k\}$ corresponding to the subsequence of σ obtained by eliminating all components strictly greater than k. For instance, if $\sigma = (2, 5, 1, 4, 6, 3)$, then $\sigma^{(4)} = (2, 1, 4, 3)$.

3.1 *Complete families, $c \leq 2$*

3.1.1 $c = 1$

THEOREM 3.2 *Let $d \geq 2$. The graphs H_c^d with polynomials recursively defined by (4) initialised with $H_1^2 = Q_2$, and*

$$G_1^d = \mathcal{P}_0^d = 1 + \sum_{i=1}^{d} X_1^i \tag{6}$$

are $(d,1)$-cycle equitable and (d,d)-edge equitable subgraphs of Q_d. ■

Proof
We only prove cycle-equitability, edge-equitability being easily established directly using equation (3).
We need to verify that (**Cond0**) G_1^d is a subgraph of H_1^d, that (**Cond1**) it a single edge along each direction and (**Cond2**) it has no 4-cycles. The proof is by induction. H_1^2 is trivially $(2,1)$-cycle equitable, consisting of a single 4-cycle in (X_1, X_2). Note that (4) implies that

$$H_1^d \supset X_d G_1^{d-1}, \qquad \text{and} \qquad d > d' \Rightarrow H_1^d \supset H_1^{d'} \ , \tag{7}$$

and that by (6) the graphs G_1^d satisfy the recurrence:

$$G_1^d = G_1^{d-1} + X_d X_1^{d-1} \ .$$

Assume that $G_1^{d-1} \subset H_1^{d-1}$. Then

$$G_1^d = G_1^{d-1} + X_d X_1^{d-1} \overset{(a)}{\subset} H_1^{d-1} \cup X_d X_1^{d-1} \overset{(b)}{\subset} H_1^d \cup X_d X_1^{d-1} \overset{(c)}{=} H_1^d \ ,$$

concluding the proof of **Cond0**. Above, *(a)* follows from the assumption, in *(b)* we used (7), and (c) follows from (7) and the fact that $X_1^{d-1} \in G_1^{d-1}$.

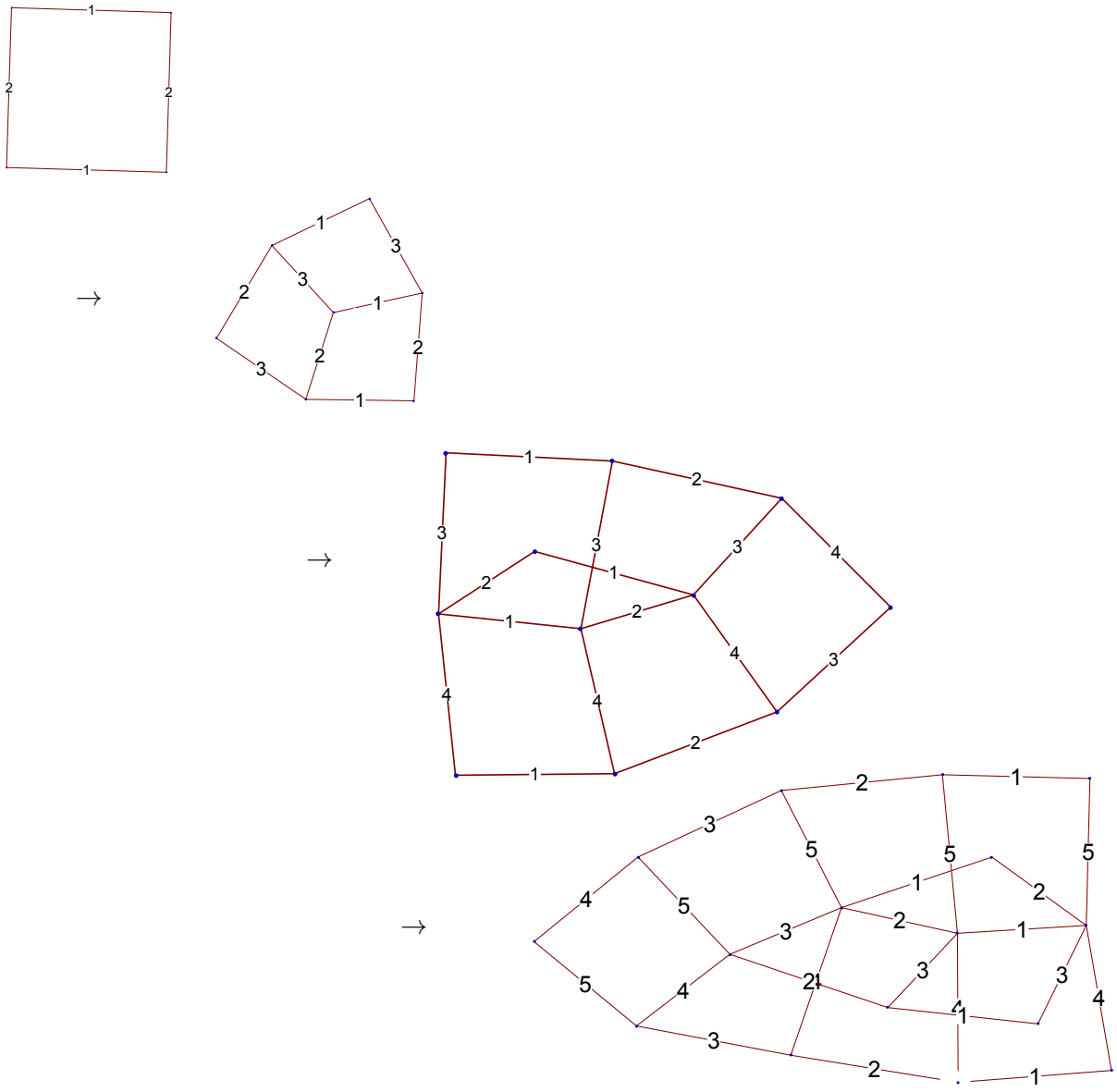


Figure 5. GenerationX of $(d, 1)$-cycle equitable graphs, $d = 2, 3, 4, 5$.

Cond1 is easily verified by using Lemma 3 and checking that

$$\left\langle G_1^d, X_k G_1^d \right\rangle = 2 \ ,$$

which follows directly from noting that

$$X_k \sum_{j=1}^{d} X_1^j = \sum_{j=1}^{k-2} X_1^j X_k + X_1^{k-1} X_k + X_1^k X_k + \sum_{j=k+1}^{d} X_1^{k-1} X_j^d$$
$$= \sum_{j=1}^{k-2} X_1^j X_k + X_1^k + X_1^{k-1} + \sum_{j=k+1}^{d} X_1^{k-1} X_j^d \ ,$$

which has only two terms of G_1^d. Since a cycle must contain at least two edges along the same direction, **Cond2** is also verified (G_1^d has no 4-cycles). ■

Figure 5 illustrates the recursive construction of H_1^d. We can see that at each step d one cycle in the new coordinate (X_d) and each of the "old" ones ($X_1, \ldots, X_{d-1}$) is created by adding one edge of each colour parallel to edges already in H_1^{d-1} (this set of edges, G_1^{d-1}, is a $(d-1)$ OAT trajectory in H_1^{d-1}).

We remark that similarly to the work in [6] our construction for $c = 1$ is based on the linear graphs that are actually OAT paths in a set of coordinates. However, the two constructions are remarkably different: while the designs of the New Morris Method are directly defined as the composition of a set of paths of length $d+1$ in a given set of coordinates ($d/2$ paths in all d factors if d is pair) to which an extra set of nodes must be added to guarantee the existence of complete 4-cycles, our construction recursively adds a OAT path of increasing size (the graphs $X_{d+1}G_1^d$) to smaller cycle-equitable graphs, and the target number of 4-cycles are directly enforced by the recursive construction. Moreover, the graphs H_1^d are also (d,d)-edge-equitable, which cannot be guaranteed for the construction in [6].

3.1.2 $c = 2$

THEOREM 3.3 *Let $d \geq 3$. The designs H_2^d with polynomials recursively generated by equation (4) initialised with $H_2^3 = Q_3$ and with*

$$G_2^d = \mathcal{P}_{e_d} \oplus \mathcal{P}_{(d,d-1,\ldots,1)} \ . \tag{8}$$

are $(d,2)$-cycle equitable and $(d, 2(d-1))$-edge equitable subgraphs of Q_d. ■

We do not detail the proof of this Theorem, that is a special case of Theorem 3.4 given below.

Consider a target dimension $d^\star$. A more flexible construction of $H_2^{d^\star}$ than the one above is provided by Theorem 3.4 below. Before presenting it, we introduce some notation. Denote by $e_{d^\star} = (1, 2, \ldots, d^\star)$ the identity permutation and let $\sigma_{(*21)}$ be a permutation of size $d^\star$ such that $\sigma_{(*21)}(d^\star) = 1$ and $\sigma_{(*21)}(d^\star - 1) = 2$. As defined before, let $\sigma_{(*21)}^{(d)}$, with $d < d^\star$, be the permutation obtained by eliminating all entries of $\sigma_{(*21)}$ that are greater than d.

THEOREM 3.4 *Let $d \geq 3$ and $\sigma_{(*21)}$ a permutation of $\{1, \ldots, d^\star\}$ as above. The graphs $H_2^{d^\star}$ with polynomials recursively generated by equation (4) initialised with $H_2^3 = Q_3$ and with*

$$G_2^d = \mathcal{P}_{e_d} + \mathcal{P}_{\sigma_{(*21)}^{(d)}}, \qquad d < d^\star \ , \tag{9}$$

are $(d,2)$-cycle equitable and $(d, 2(d-1))$-edge equitable subgraphs of Q_d. ■

Proof
Again, we need to check conditions **Cond0**—**Cond2**.
Cond0 Demonstration is again by induction.
For $d^\star = 3$, the only possible permutation is $\sigma_{(*21)} = (3, 2, 1)$. It can easily be verified that $G_2^d = \mathcal{P}_{e_d} \oplus \mathcal{P}^{(d)}_{\sigma_{(*21)}} = \mathcal{P}_{e_{d-1}} \oplus \mathcal{P}^{(d)}_{\sigma_{(*21)}}$, since the term X_1^d is also in $\mathcal{P}_\sigma$. We know that $\mathcal{P}_{e_{d-1}} \subset G_2^{d-1} \subset H_2^{d-1} \subset H_2^d$. It remains to show that $\mathcal{P}^{(d)}_{\sigma_{(*21)}} \subset H_2^d$. Consider all monomials $s \in \mathcal{P}^{(d)}_{\sigma_{(*21)}}$.

- If s does not contain X_d, then by construction $s \in \mathcal{P}_{\sigma^{(k)}}$ for some $k < d$ and thus $s \in G^k \subset H_2^k \subset H_2^d$
- If s contains X_d, then $s \in X_d \mathcal{P}_{\sigma^{(d-1)}} \subset H_2^d$.

We can thus conclude that $\mathcal{P}^{(d)}_{\sigma_{(*21)}} \subset H_2^d$, and thus that $G_2^d \subset H_2^d$.

Cond1 G_2^d contains at least two edges along each direction, one belonging to $\mathcal{P}_{e_d}$, and another to $\mathcal{P}_\sigma$. It can be easily checked that polynomials $\mathcal{P}_{e_d}$ and $\mathcal{P}_\sigma$ have only two common monomials which are 1 and X_1^d. If a monomial $s \in \mathcal{P}_{e_d}$ would create an additional edge along X_i with a monomial $s' \in \mathcal{P}_\sigma$, then s and s' should differ only in X_i, and in particular, s and s' should both contain X_1X_2, which can only happen if $s = s' = X_1^d$, since $\sigma(d) = 1$ and $\sigma(d-1) = 2$. We can thus conclude that no extra edges are created, and thus that G_2^d is a edge-equitable graph.

Cond2 This property is a consequence of the previous one. Indeed, the edges of G_2^d are exactly the union of those of $\mathcal{P}_{e_d}$ and those of $\mathcal{P}_\sigma$, and not more. G_2^d is a simple cycle of length $2d$ that contains no smaller cycles. ■

3.2 *Incomplete families, $c \geq 3$*

For $c \geq 3$ we only have a construction for $d \geq c + 1$. We start by stating a lemma that provides the initialisation of our recursive construction.

LEMMA 3.5 *The graph with polynomial defined by*

$$H_c^{c+1} = 1 + \sum_{i=1}^{c+1} X_i + \sum_{i,j=1,j>i}^{c+1} X_iX_j + \sum_{i,j,k=1,k>j>i}^{c+1} X_iX_jX_k$$

has exactly c 4-cycles in every pair of the $c+1$ coordinates. ■

Proof
Every cycle in coordinates (X_i, X_j) corresponds to the presence in H_c^{c+1} of four terms of the form $q\,(1 + X_i + X_j + X_iX_j))$, for some $q \in H_c^{c+1}$. We must show that we can find these four terms for c distinct monomials q not containing X_i and X_j. Given the complete symmetry of H_c^{c+1}, we consider $i = 1$ and $j = 2$. Since H_c^{c+1} contains all 1st, 2nd and 3rd order monomials, we can make the following c choices:

$$q = 1, \text{ and } q = X_j, \quad j = 3, \ldots, c+1 \ .$$

Note that no other choices are possible, since there are only $c - 1$ 3rd order terms. This concludes the proof that there are indeed c cycles in all pairs of coordinates (X_i, X_j) in H_c^{c+1}. ■

THEOREM 3.6 *Let $c \geq 3$ and $d \geq c+1$. The graphs H_c^d with polynomials defined by (4) initialised with H_c^{c+1} given in Lemma 3.5 and with*

$$G_c^d = 1 + \left(\sum_{i=1}^{c} X_i\right)\left(1 + X_{c+1} + \cdots + X_{c+1}^i + \cdots + X_{c+1}^d\right) + X_{c+1}^d \sum_{i,j=1, j\neq i}^{c} X_i X_j \ , \quad (10)$$

are (d,c)-cycle and $(d,d+c)$-edge equitable. ■

Proof
The two lemmas below state that the three conditions on G_c^d for H_c^d to be (d,c)-cycle equitable are verified. ■

LEMMA 3.7 *For $d \geq c+1$, the polynomial G_c^d given by (10) has exactly c edges of each colour* (**Cond1**)*, and no 4-cycles* (**Cond2**)*.* ■

Proof
Decompose

$$G_c^d = S^d + T^d \ ,$$

where

$$S^d = 1 + \left(\sum_{i=1}^{c} X_i\right)\left(1 + \sum_{j=c+1}^{d} X_{c+1}^j\right), \qquad T^d = X_{c+1}^d \sum_{i,j=1, j\neq i}^{c} X_i X_j \ .$$

Remark that S^d can be written as

$$S^d = \sum_{i=1}^{c} S_i^d, \qquad S_i^d = 1 + X_i \mathcal{P}_{(c+1\cdots d)} \ ,$$

and thus it has a single edge along coordinates $X_1, \ldots, X_c$, and exactly c edges along coordinates $X_{c+1}, \ldots, X_d$.

Consider $s = X_{c+1}^d X_i X_j \in T^d$. Monomial s has only two neighbours in S^d: $s_i = X_i X_{c+1}^d \in S_i^d$ and $s_j = X_j X_{c+1}^d \in S_j^d$, inducing edges along coordinates X_j and X_i, respectively.

Since for each i there are $c-1$ distinct values of $j \in \{1, \ldots, c\} \setminus \{i\}$, we can conclude that the number of edges along direction X_i linking S^d to T^d is $c-1$, proving that G_c^d has exactly c edges along all directions.

The proof that H_c^d has no 4-cycles is immediate from the fact that all polynomials S_i^d have $d-c+1 > 2$ terms for $d > c+1$ and thus the cycles completed with the nodes of T^d cannot have length smaller than 4+2=6.

LEMMA 3.8 *For $d > c+1$, $G_c^d \subset H_c^d$* (**Cond0**)*.* ■

First note that $G_c^{c+1} \subset H_c^{c+1}$ by construction, since H_c^{c+1} has all monomials up to

order 3, and G_c^{c+1} has no monomials of order greater than 3. Since we can write

$$S^d = S^{d-1} + X_d \sum_{i=1}^{c} X_i X_{c+1}^{d-1}, \qquad T^d = X_d T^{d-1} \ ,$$

multiplication by X_d in composition (4) creates all terms in G_c^d that were not in G_c^{d-1}, and the lemma is demonstrated.

4. Size of designs

To take into account the number of mixed effects computed by a design, the definition of economy presented in [1] must be slightly modified.

Definition 4.1 Let H_c^d be a (d, c)-cycle equitable and (d, m)-edge equitable design. Its economy is

$$\chi = \frac{|\{\text{elementary effects}\} \cup \{\text{mixed effects}\}|}{|H_c^d|} = \frac{md + c\binom{d}{2}}{|H_c^d|}$$

Explicit formulas for the designs sizes follow easily from their recursive definition:

$$|H_1^d| = |H_1^{d-1}| + d, \qquad d > 2, \qquad |H_1^2| = 4 \ ;$$
$$|H_2^d| = |H_2^{d-1}| + 2d + 1, \qquad d > 3, \qquad |H_2^3| = 8 \ ;$$
$$|H_c^d| = |H_c^{d-1}| + \frac{1}{6}(6 + 5d + d^3), \qquad d \geq c+1, \quad |H_c^{c+1}| = 1 + c + \binom{d}{2} + \binom{d}{3}, c \geq 3 \ ,$$

yielding

$$|H_1^d| = 1 + \frac{1}{2}d(d+1) \sim \frac{d^2}{2} \ ; \tag{11}$$
$$|H_2^d| = 2 + d(d-1) \sim d^2 \ ; \tag{12}$$
$$|H_c^d| = \frac{1}{6}\left(6 - c + c^3 + 6d - 3c^2 d + 3cd^2\right) \sim \frac{cd^2}{2}, \qquad c \geq 3, d > c+1 \ . \tag{13}$$

The complexity of our designs is $|H_c^d| \sim \frac{cd^2}{2}$, for all values of c. Remark the increased efficiency of our designs compared to [6], whose size is of order d^2 for $c = 1$, for which our designs are twice as small.

5. Annotated interaction graph

We present in this section application of the cycle-equitable designs of Section 3 to screening analysis of a function $f : \mathcal{A} \subset \mathbb{R}^d \to \mathbb{R}$. We assume a modelling framework, i.e., where the ultimate goal is to produce a model of f from a finite set of carefully chosen evaluations of f.

As we recalled in section 1.2, the elementary effects method partitions the d input factors $X_1, \ldots, X_d$ into three classes, $\mathcal{C}_0$, $\mathcal{C}_1$ and $\mathcal{C}_2$, of negligible, linear and non/linear

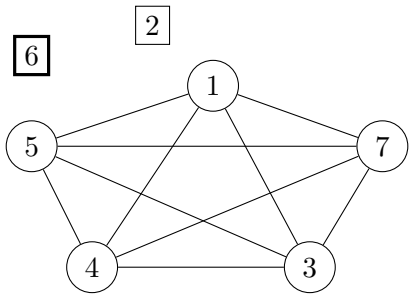


Figure 6. Interaction graph resulting of Morris method. Note the completely connected graph $K_{\mathcal{C}_2}$ connecting nodes in $\mathcal{C}_2 = \{1, 3, 4, 5, 7\}$.

mixed input factors, respectively. Once this partitioning is known, f can be modelled as the sum of a parametric (linear) and a non-parametric component:

$$f(X_1, \ldots, X_d) \simeq \sum_{X_i \in \mathcal{C}_1} a_i X_i + h\left(\{X_i\}_{X_i \in \mathcal{C}_2}\right) \ ,$$

where $h(\cdot)$ is some non-parametric function to be determined. We can expect that $|\mathcal{C}_2| \ll d$, resulting in a considerable reduction of the complexity of subsequent non-parametric model identification when compared to direct non-parametric modelling on all d input factors.

Morris method can be used to build a first interaction graph, where all negligible and linear factors are isolated nodes, while a completely connected graph $K_{\mathcal{C}_2}$ links the nodes in $\mathcal{C}_2$. An example is given in Figure 6. In the representation, bold square nodes are linear nodes (class $\mathcal{C}_1$) identified by the Morris method, and thin squares indicate negligible inputs, i.e., in class $\mathcal{C}_0$, while the elements of $\mathcal{C}_2$ are circles.

Using the cycle-equitable designs presented in this paper we can implement a second screening step, restricted to the set of inputs in classe $\mathcal{C}_2$, that will further refine the structure of $K_{\mathcal{C}_2}$.

While the original Morris method enables classification of the input factors $X_i, i = 1, \ldots, d$, the second screening step that we propose will enable classification of the edges $\mathcal{E}_{\mathcal{C}_2}$ of $K_{\mathcal{C}_2}$:

$$\mathcal{E}_{\mathcal{C}_2} = \mathcal{E}_0 \cup \mathcal{E}_1 \cup \mathcal{E}_2 \ .$$

Classification of an edge $(X_i, X_j) \in \mathcal{E}_{\mathcal{C}_2}$ in one of the classes is done by using the empirical mean and variance of a set of mixed effects $\{dd_{ij}(\xi_n), \xi_n \in \mathcal{A}\}, i \neq j$, see equation (2):

$$\begin{aligned}
&\text{If } \mu_{ij} \simeq 0, \text{ and } \mathrm{Var}(dd_{ij}) \simeq 0 \quad &&\Rightarrow \text{ no interaction between } (i,j) &&\Leftrightarrow (X_i, X_j) \in \mathcal{E}_0 \\
&\text{If } \mu_{ij} > 0, \text{ and } \mathrm{Var}(dd_{ij}) \simeq 0 \quad &&\Rightarrow \text{ bilinear interaction between } (i,j) &&\Leftrightarrow (X_i, X_j) \in \mathcal{E}_1 \\
&\text{If } \mathrm{Var}(dd_{ij}) > 0 \quad &&\Rightarrow \text{ more complex dependencies} &&\Leftrightarrow (X_i, X_j) \in \mathcal{E}_2
\end{aligned}$$

Above, μ_{ij} and $\mathrm{Var}(dd_{ij})$ are the empirical mean and variance of the mixed effects dd_{ij}.

A new graph $K^a_{\mathcal{C}_2}$ can be defined using this classification by *(i)* removing all edges in class $\mathcal{E}_0$, and *(ii)* annotating the remaining edges either as either $\mathcal{E}_1$ or $\mathcal{E}_2$. Remark that if $(X_i, X_j) \in \mathcal{E}_0, \forall j \neq i$, then X_i is an isolated node of $K^a_{\mathcal{C}_2}$, indicating a non-interacting

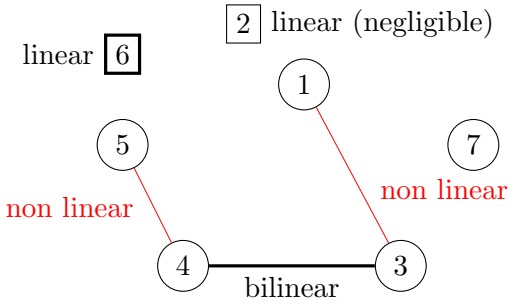


Figure 7. Annotated interaction graph $K^a_{\mathcal{C}_2}$ built using the mixed effects. $\mathcal{E}_1$: black edges, $\mathcal{E}_2$: red edges.

non-linear term. The annotated graph $K^a_{\mathcal{C}_2}$ will in general no longer be completely connected, revealing a model structure that can be identified more efficiently.

We present the concept using a slightly modified version of example b in [12], that we designate by MRCK function:

$$\begin{aligned} f(X_1,\ldots,X_7) =& \alpha X_6 + \exp(-\nu X_7) + (\delta_0 + \delta_3 X_3 + \delta_4 X_4)^2 \\ &+ \cos(\beta_0 + \beta_1 X_1 + \beta_3 X_3) + \sin(\gamma_0 + \gamma_4 X_4 + \gamma_5 X_5) \quad . \end{aligned} \tag{14}$$

Morris method will identify input X_2 as negligible ($\mathcal{C}_0$), input X_6 as linear ($\mathcal{C}_1$), all others being assigned to $\mathcal{C}_2$, resulting in the graph in Figure 6.

By analysis of the mixed effects in $\mathcal{C}_2$ a number of edges can be removed, resulting in the identification of the isolated non-linear input factor X_7. Moreover, it will show that there is no interaction between X_1 and $\{X_4, X_5\}$ and between X_3 and X_5. Additionally, the classification of the remaining edges indicates that the interaction between X_3 and X_4 is bilinear. The resulting graph is shown in Figure 7, where bilinear edges (class $\mathcal{E}_1$) are shown in black and non-linear/higher-order interactions (class $\mathcal{E}_2$) are shown in red.

Graph $K^a_{\mathcal{C}_2}$ can be considered as an annotated version of the FANOVA graph $G = (V, E)$ defined in [12], that enables direct visualisation of the structure of interaction between input variables of a function. The vertex set coincides with the set of input factors: $V = \{X_1, \ldots, X_d\}$ and the set of edges $E = \{e_{ij} = (X_i, X_j),\ X_i, X_j \in V\}$ indicates presence of interaction (any form of interaction) between the corresponding input variables. Past work on the identification of the FANOVA graph and on its use for modelling has stressed the role of its maximal cliques, which enable, in the context of ordinary Kriging modelling, the use of additive kernels defined over lower dimensional subspaces of the function's domain, significantly decreasing the complexity of model building while improving at the same time the prediction accuracy of the derived models [12].

Figure 8 shows the FANOVA graph as defined in [12] for the example under study. For this example the set of maximal cliques of the original FANOVA graph is $\{\{X_2\}, \{X_6\}, \{X_7\}, \{X_1, X_3, X_4, X_5\}\}$ leading to a model of the form

$$\begin{aligned} f(X_1,\ldots,X_7) &\simeq h_1(X_2) + h_2(X_6) + h_3(X_7) + h_4(X_1, X_3, X_4, X_5) \\ &\simeq h_2(X_6) + h_3(X_7) + h_4(X_1, X_3, X_4, X_5) \quad . \end{aligned}$$

The models $h_i(\cdot), i = 1, \ldots, 4$ in the expression above can be identified by kriging. Using the information in the annotated interaction graph in Figure 7, we can further infer the presence of a series of parametric terms, and the set of maximal cliques over which

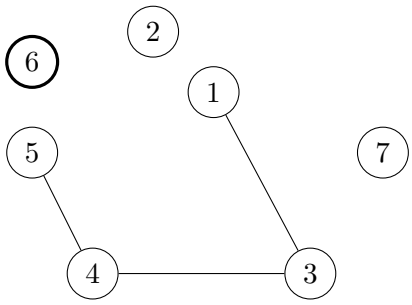


Figure 8. FANOVA graph.

non-parametric models must be defined needs only to cover edges of class $\mathcal{E}_2$, yielding:

$$f(X_1,\ldots,X_6) \simeq aX_6 + bX_3X_4 + h_1'(X_1,X_3) + h_2'(X_4,X_5) + h_3'(X_7) \ .$$

Again, functions $h_i'(\cdot)$ can be built using kriging. The final model is the superposition of two parametric components (on $\{X_3,X_4\}$ and on $\{X_6\}$) and three non-parametric components (on $\{X_1,X_3\}$, $\{X_4,X_5\}$ and $\{X_7\}$). The reduced dimension of each of the non-parametric models, which are now always smaller than 2, significantly decreases the complexity of the subsequent model identification step with respect to the structure implied by the original FANOVA graph, which requires model identification in a space of dimension 4.

Analysis of the mixed effects can, as in this example, lead to a finer characterisation of the dependency on different input factors. Note, however, that without observation of the second-order derivatives $\partial^2 f/\partial X_i^2$ we cannot rule out the presence of non-linear terms in each isolated input factor in $\mathcal{C}_2$. The non-parametric components must thus cover all nodes of $K^a_{\mathcal{C}_2}$, as it is the case in the example shown above.

5.1 *Morris function*

We consider now the function used in the original Morris' paper [1]:

$$f(x) = \beta_0 + \sum_{i=1}^{20}\beta_i w_i + \sum_{i<j}^{20}\beta_{ij}w_iw_j + \sum_{i<j<l}^{5}\beta_{ijl}w_iw_jw_l + \sum_{i<j<l<s}^{4}\beta_{ijls}w_iw_jw_lw_s \ ,$$

where $w_i = 2X_i - 1, i \in \{1,2,4,6,8,\ldots,20\}$, $w_i = 2.2X_i/(X_i+0.1) - 1, i \in \{3,5,7\}$, and the following values are used for the important terms:

$$\begin{array}{llll} \beta_i = 20, & i \in \{1,\ldots,10\}, & \beta_{ij} = -15, & i,j \in \{1,\ldots,6\} \\ \beta_{ijl} = -10, & i,j,l \in \{1,\ldots,5\}, & \beta_{ijls} = 5, & i,j,l,s \in \{1,\ldots,4\}. \end{array}$$

The remaining 1^{st} and 2^{nd} order coefficients are independent realisations of a standard normal distribution, $\beta_i \sim \mathcal{N}(0,1),\ i \notin \{1,\ldots,10\}, \beta_{ij} \sim \mathcal{N}(0,1),\ i,j \notin \{1,\ldots,6\}$. The relevant classes of input factors are

$$\mathcal{C}_0 = \{11,\ldots,20\}, \qquad \mathcal{C}_1 = \{8,9,10\}, \qquad \mathcal{C}_2 = \{1,\ldots,7\} \ .$$

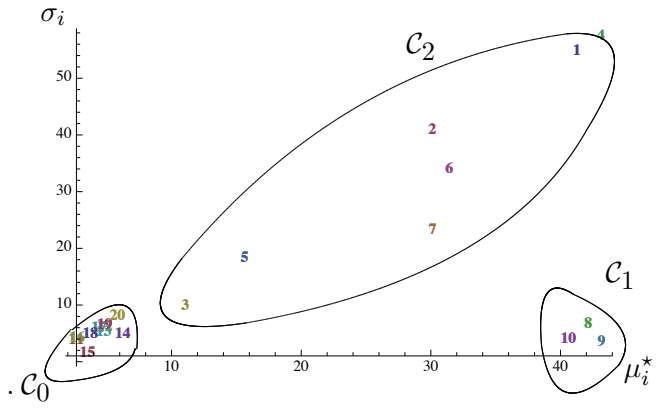


Figure 9. Morris statistics ($m = 4$) for $X_i, i = 1, \ldots, 20$. Position of the label i indicates the observed $(\mu_i^\star, \sigma_i)$.

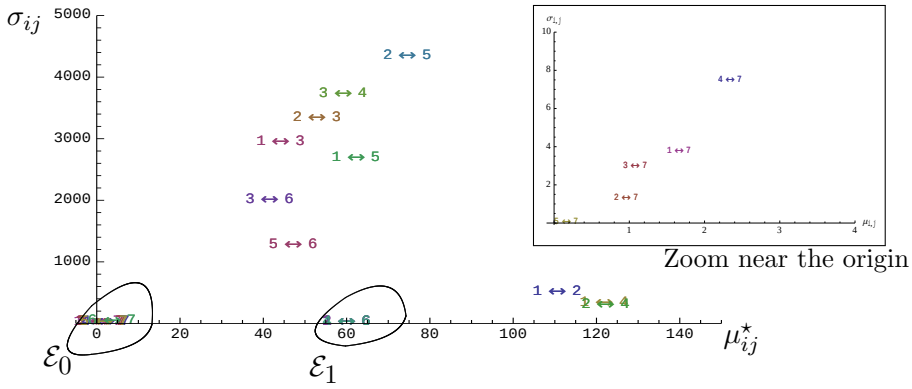


Figure 10. Mixed effects for Morris function, $c = 4$, $r = 3$.

Note that X_7 is a purely non-linear term, while X_6 interacts bilinearly with input factors $\{X_1, X_2, X_4\}$.

We performed factor screening using $r = 3$ instantiations of the $(20, 4)$-edge equitable designs presented in [8], yielding a total number of derivatives per direction equal to 12. The values of $\{(\mu_i, \sigma_i)\}_{i=1}^{20}$ obtained are shown in Figure 9.

Analysis of this Figure reveals that the factors are correctly assigned to classes. We thus concentrated on study of the mixed effects for elements of the class $\mathcal{C}_2$. The mean and variance of the set of mixed effects computed using $r = 30$ random replications of $(7, 3)$-cycle equitable designs inside the domain of f is shown in Figure 10. Random variants of the designs were obtained by considering random permutations of the coordinates. A zoom of the region around the origin is shown in the inset of Figure 10.

We can see that the method correctly detects X_7 as non-linear input factor (all interactions with other factors being negligible) as well as the bilinear terms $X_i X_6, i \in \{1, 2, 4\}$, who appear super-imposed in the plot since their variance is close to zero. The resulting annotated interaction graph is shown in Figure 11.

The maximal cliques are $\{1, 2, 3, 4, 5\}, \{3, 5, 6\}, \{7\}$, leading to a model with the structure

$$f(X_1, \ldots, X_{20}) \simeq \sum_{i \in \{8,9,10\}} \alpha_i X_i + \sum_{i \in \{1,2,4\}} \beta_i X_i X_6 + \epsilon_1(X_3, X_5, X_6) + \epsilon_2(X_1, X_2, X_3, X_4, X_5) + \epsilon_3(X_7) \ .$$

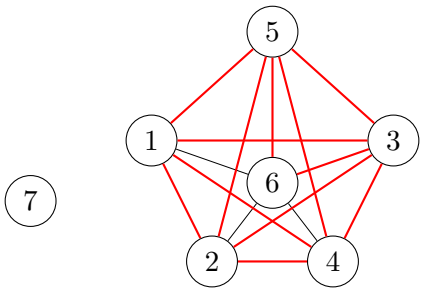


Figure 11. Annotated interaction graph of non-linear factors of Morris function (red edges indicate nonlinear interaction).

Note that the FANOVA graph proposed in [12], with no information on the bilinear terms involving X_6, would have maximal cliques $\{1,2,3,4,5,6\},\{7\}$, implying a subsequent modelisation step of higher complexity.

Finally, we remark that the method in [12], that resorts to Monte Carlo techniques to compute the total interaction indices relative to all $\binom{d}{2}$ pairs of input variables, requires a much larger number of evaluations of f than the 1140 that were used to compute the graph in Figure 11 using our mixed effects method, showing that the method proposed in this paper, based on use of our cycle-equitable designs, is an efficient alternative to the analysis proposed in [12].

5.2 *Moon high-dimensional function*

The function considered next is a particular realisation of the stochastic polynomials model used in [10] and documented in [11]. It is a nonlinear function defined in $x \in [0,1]^{20}$ with many terms, including linear, quadratic and mixed terms, with only four active effects, which all appear through mixed or quadratic (non-linear) effects:

$$f(x) = [x^T \ 1] A \begin{bmatrix} x \\ 1 \end{bmatrix} \ ,$$

where matrix A is given in [11], and the dependency on the four active inputs X_1, X_7, X_{12} and X_{19} is

$$f(x) = -19.71 X_1 X_{18} + 23.72 X_1 X_{19} - 13.34 X_{19}^2 + 28.99 X_7 X_{12} \ . \tag{15}$$

The presence of 3 bilinear terms makes this function a good candidate to illustrate the ability of our designs to screen purely non-linear (X_{19}^2) and bilinear terms. Figure 5.2 displays the elementary effects, computed using 30 replications of $(20,4)$-edge equitable designs as presented in [8].

The set of non-linear/mixed input factors has been correctly identified as $\mathcal{C}_2 = \{1,7,12,18,19\}$. The analysis of the mixed effects (30 randomised versions of $(6,3)$-cycle equitable graphs were used) displayed in Figure 13, shows that all variances are very small, indicating that no mixed terms or non-linear transformations of the input factors seem to be present, i.e. $\mathcal{E}_2 = \emptyset$, and leads to the annotated interaction graph shown in Figure 14. There are only three interaction (bilinear, class $\mathcal{E}_1$) terms, to which must be added a set of non-linear components along each input factor in $\mathcal{C}_2$:

$$f(X_1, X_7, X_{12}, X_{18}, X_{19}) \simeq a_1 X_1 X_{18} + a_2 X_1 X_{19} + a_3 X_7 X_{12} + \sum_{X_i \in \mathcal{C}_2} \epsilon_i(X_i) \ .$$

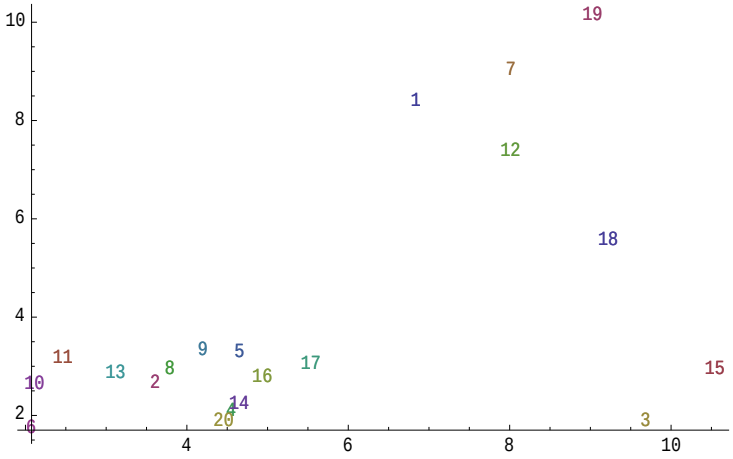


Figure 12. Elementary effects of Moon function, eq. (15).

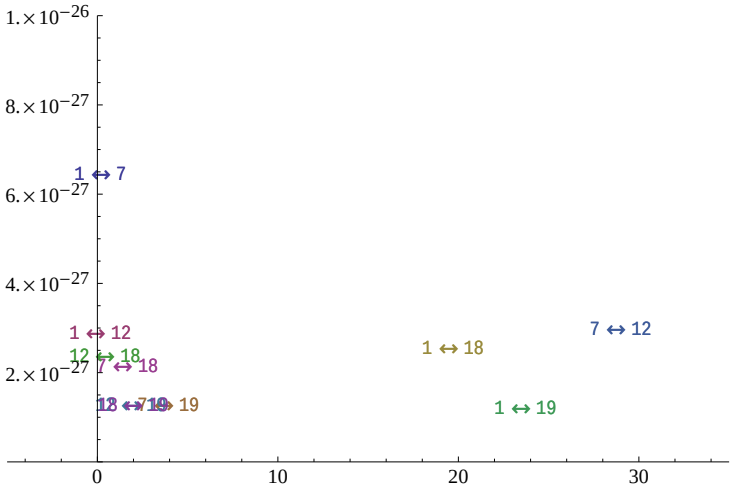


Figure 13. Mixed effects of Moon function, eq. (15).

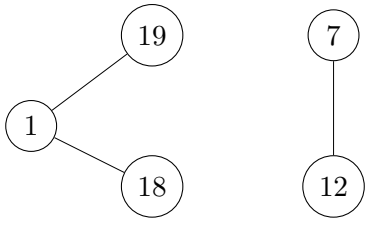


Figure 14. Annotated interaction graph of non-linear factors of Moon's function (black edges indicate bilinear dependency).

This function reveals the impact that observation of second-order derivatives may have. Would we have the ability to efficiently detect quadratic variation in isolated input factors, we would easily conclude that a simple parametric model is sufficient to express the main variability of this function.

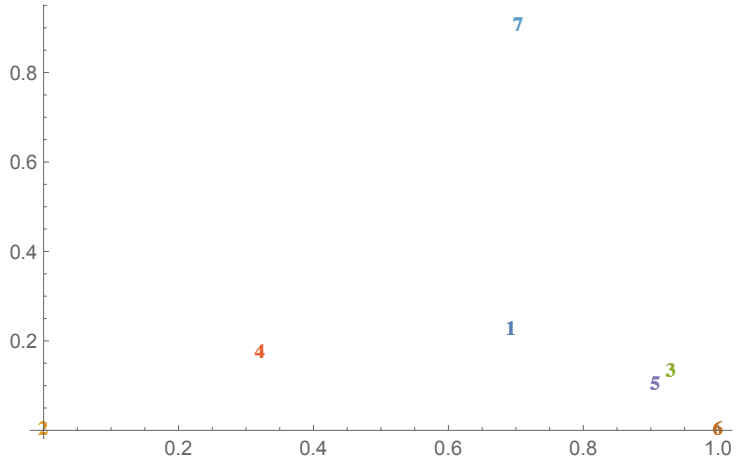


Figure 15. Elementary effects of MRCK function, eq. (14).

5.3 *MRCK function [12]*

We present here the results of the numerical study of the function defined in the introduction of this section in eq. (14), with the following values of the parameters:

$$f(X_1, \dots X_7) = X_6 + \exp(-4X_7) + (0.5 + 0.35X_3 - 0.6X_4)^2 + \cos(0.8 + 1.1X_1 - X_3) - \sin(0.5 - 0.9X_4 - X_5) \quad .$$

Figure 15 shows the elementary effects from which input factors have been classified as negligible ($\mathcal{C}_0 = \{X_2\}$), linear ($\mathcal{C}_1 = \{X_6\}$) or "other" ($\mathcal{C}_2 = \{X_1, X_3, X_4, X_5, X_7\}$). The elementary effects have been computed using 30 replications of $(7, 2)$-edge equitable designs, see [8] for construction of these designs. The mean and variance of the set of mixed effects computed using 30 $(5, 3)$-cycle equitable designs along the coordinates in $\mathcal{C}_2$ are shown in Figure 16, leading to the annotated interaction graph shown in Figure 7. Note that since all other mixed effects cluster near the origin, we omitted their labels in Figure 16. We remark that not only our annotated interaction graph has the same topology as the FANOVA graph, but detailed comparison of the means of the mixed effects also shows that derived input / interaction ranking is consistent with the one obtained in [12]. That similar conclusions can be drawn from a much smaller number of evaluations (a total of 1140) of f (the method in [12] resorts to Monte Carlo methods to estimate "projected" Sobol indexes), clearly indicates the potential of the analysis proposed here.

6. Conclusions

The paper proposes an extension of Morris' screening method that enables the determination of an annotated interaction graph – similar to the FANOVA graph proposed in [12] – of an unknown multi-variate function, with a limited number of function evaluations (quadratic on the number of non-linear/interaction input factors). The work presented here concentrates on identifying designs that can efficiently support evaluation of the mixed effects on which the method is based. The class of cycle equitable subgraphs of the hypercube, parameterized by the number of mixed effects that are computed for each pair of input variables, is introduced as a formalisation of the properties that these designs should present, and algorithms are given for computation of the associated graphs. As far

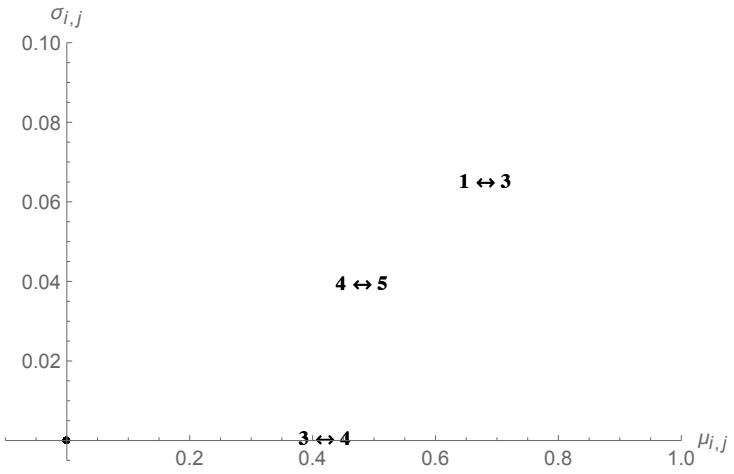


Figure 16. Mixed effects of MRCK function, eq. (14).

as we know, this is the first time this class of subgraphs of the hypercube is formally defined and studied. The algorithms are recursive, incrementally building graphs on higher dimensional hypercubes, and their derivation is based on a polynomial representation of subgraphs of the unit hypercube equipped of a convenient scalar product.

The work presented here is motivated by the desire to extend Morris Elementary Effects screening method [1] to the study of the interactions between the input factors of a function, and improves on previously published contributions in this direction [6] by proposing more generic and efficient designs. We illustrate the use of the cycle-equitable graphs proposed in a series of examples, demonstrating that this extension of Morris' method enables estimation of an annotated interaction graph of a function with a complexity much lower than related proposed approaches [12]. In fact, our method inherits the computational advantages of the Morris method, but extends its exploitation beyond isolated classification of each input factor, to the analysis of their interaction. We believe it can be an efficient preliminary analysis tool in the context of model building, whenever function evaluations are expensive: not only it indicates the relevant factors and interactions, but, and probably more importantly, the ability to explicitly indicate the edges of the interaction graph that can be modelled as simple bilinear (and thus parametric) terms. This can contribute to significantly decrease the complexity and improve the stability [12] of subsequent non-parametric (most commonly kriging) modelling steps, which can be confined to lower dimensional subspaces of the input space.

We expect to continue the work presented here by addressing a number of questions that remain open. Some concern the class of cycle-equitable graphs itself. What are the minimal elements of the cycle-equitable class we defined? How far are our recursive solutions from optimality? Subgraphs of the hypercube have been extensively studied, and a number of interesting classes have been defined (median graphs, mesh graphs, regular partial cubes, etc.). What is the relation of the graphs studied in this paper to these other classes, if any?

Other issues concern improvements of the work presented here: can we complete the spectrum of values of input dimension d and cycle multiplicity c, and define a solution for $d \leq c$ when $c \geq 3$? (We remark that this issue is more of a formal than practical importance, since in practice one would rarely use values of $c \geq 5$.) One extension that seems to us particularly interesting concerns the definition of equitable graphs on finite lattices, $\{0, 1, \ldots, k\}^d$ rather than on the hypercube $\{0, 1\}^d$, i.e., of designs that could be used to observe the behaviour of higher order derivatives (e.g., computation of $\partial^2 f / \partial X_i^2$ requires three points aligned along input X_i).

We concentrated here on the definition of the designs for computation of the mixed effects, once Morris' elementary effects method has signalled the input factors entering non-linearly and/or in interaction with other variables. In particular, we did not assess how to constraint the two designs (the one used for Morris' method and the one used for studying mixed effects) such that the evaluations of f in the first step are a subset of those indicated by the cycle-equitable design. Since our cycle-equitable designs are also edge-equitable, we are sure that we can chose the former as a subset of the latter, and only complete (or create) the cycles along the variables indicated as relevant by the analysis of the elementary effects. This question will be addressed in a future paper. Previous studies have suggested that the discrimination of the original Morris test on the first and second moments of the set of sampled mixed effects may be affected when clustered designs are used, affecting the ability to correctly classify the input factors. In our opinion, assessment of this issue requires a precise formulation of the decision problem underlying Morris method (what exactly means that a factor is "approximately linear" or "approximately negligible"?), and it is our intention to address it in a future publication.

Acknowledgements

This work was partially funded by the ANR through project DESIRE (*Designs for Spatial Random Fields*), contract ANR 1-IS01-0001. Project web-site : www.i3s.unice.fr/desire. The authors acknowledge fruitful discussions with other colleagues during the SAMO 2013 meeting, and thank the reviewers for their careful reading and comments, that helped to considerably improve the original manuscript.

References

[1] Max D. Morris, "Factorial Sampling Plans for Preliminary Computational Experiments", *Technometrics.*, Vol. 32 (2), pp. 161–174, 1991.

[2] F. Campolongo, A. Saltelli, "Design of experiments", in: A. Saltelli, K. Chan, M. Scott (Eds.), *Sensitivity Analysis*, *Probability and Statistics Series*, John Wiley, pp. 5163, 2000.

[3] F. Campolongo, J. Cariboni, A. Saltelli, "An effective screening design for sensitivity analysis of large models", *Environmental Modelling & Software*, Elsevier, Vol. 22 (10), pp. 1509–1518, 2007.

[4] A. Saltelli, P. Annoni, I. Azzini, F. Campolongo, M. Ratto, S. Tarantola, "Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index", *Comput. Phys. Commun.*, Vol. 181, pp. 259–270, 2010.

[5] F. Campolongo, A. Saltelli, J. Cariboni, "From screening to quantitative sensitivity analysis. A unified approach", *Comput. Phys. Commun.*, Vol. 182, pp. 978–988, 2011.

[6] F. Campolongo, R.D. Braddock, "The use of graph theory in the sensitivity analysis of the model output: a second order screening method," *Reliability Engineering & System Safety*, Vol. 64 (1), pp. 1–12, 1999.

[7] R.A. Cropp, R.D. Braddock, "The New Morris Method: an efficient second-order screening method," *Reliability Engineering & System Safety*, Elsevier, Vol 78 (1), pp. 77–83, 2002.

[8] J.-M. Fédou, M.-J. Rendas, "Equitable (d, m)-edge designs for Sensitivity Analysis," (http://arxiv.org/pdf/1307.1996.pdf)

[9] A. Saltelli, M. Ratto, S. Tarantola, F. Campolongo, (2006). "Sensitivity Analysis Practices: Strategies for Model Based Inference." *Reliability Engineering & System Safety* **91**, 1109–1125.

[10] H. Moon, A. M. Dean, and T. J. Santner. "Two-stage sensitivity-based group screening in computer experiments." *Technometrics*, **54**(4), 376-387, 2012.

[11] http://www.sfu.ca/∼ssurjano/moon10hd.html (as of August 2014).

[12] T. Muehlenstaedt, O. Roustant, L. Carraro, S. Kuhnt, "Data-driven Kriging models based on FANOVA-decomposition." *Statistics and Computing*, May 2012, Vol. 22, Issue 3, pp 723-738.