Mining Markov Network Surrogates for Value-Added Optimisation

Alexander E.I. Brownlee
Computing Science and Mathematics
University of Stirling
FK9 4LA Scotland UK
sbr@cs.stir.ac.uk

ABSTRACT

Surrogate fitness functions are a popular technique for speeding up metaheuristics, replacing calls to a costly fitness function with calls to a cheap model. However, surrogates also represent an explicit model of the fitness function, which can be exploited beyond approximating solution fitness. This paper proposes that mining surrogate fitness models can yield useful additional information on the problem to the decision maker, adding value to the optimisation process. An existing fitness model based on Markov networks is presented and applied to the optimisation of glazing on a building façade. Analysis of the model reveals how its parameters point towards the global optima of the problem after only part of the optimisation run, and reveals useful properties like the relative sensitivities of the problem variables.

Keywords

metaheuristics; surrogates; fitness approximation; decision making

1. INTRODUCTION

Surrogate fitness functions [8, 18, 19] are a useful tool for improving the efficiency of metaheuristic search. They have gained popularity in recent years, primarily as a means of achieving speed up: a computationally cheap surrogate can take the place of a costly fitness function such as a long-running simulation or a human-in-the-loop evaluation (e.g. [7, 8, 19, 21, 30]).

Surrogates are typically constructed by training a model either prior to, or in parallel with, the optimization run. A little-used additional benefit of a surrogate is that it represents an explicit model of the problem. Given that the initial motivation for using the surrogate was to improve the speed of the search, this model is effectively obtained "for free"; at least in terms of additional CPU time required. This explicit model can subsequently be mined (similar in principle to Regression Analysis [12]) to enrich the feedback on the problem that is provided to the end user, and support enhanced decision making. It can be seen as one of several tools that enable a philosophy of *value-added optimisation*: rather than simply offering an optimal solution or solutions to the decision maker, offering deeper insights to the decision maker. These insights can include the sensitivity of objectives to the decision variables, possible interactions between decision variables, or more qualitative feedback on the optimal regions of the search space. Alternative approaches to achieving value-added optimisation include systematic analysis of the relationships between variables and objectives [6] and using solutions arising from the search process to seed classic sensitivity analysis [34].

Mining of models in metaheuristics has already been demonstrated in the context of Estimation of Distribution Algorithms (EDAs) [15, 20, 23]. EDAs construct a probabilistic model that reflects the distribution of highly fit solutions in the population, and sample this distribution to yield new solutions that have a high probability of having high fitness values. Some existing work has shown how useful information can be extracted from such probabilistic models [16,17,26,27]. It has been observed [25] that in some real-world problems, the information extracted from the EDA evolution could be as important as the optimisation results.

Naturally, how useful this information will be is highly problem-dependent. It also depends on the nature of the surrogate model: black box approaches like artificial neural networks may prove harder to mine than transparent methods like functions evolved by genetic programming [24].

This paper revisits the Markov network Fitness Model (MFM), a probabilistic model of fitness for bit string encoded problems originally developed as part of the EDA, DEUM [28, 29]. Subsequently, the MFM was demonstrated as a more general model of fitness functions expressed in terms of their Walsh functions [5]: this was exploited as a surrogate fitness function [3, 7]. The relationship between the parameters of the MFM and the global optima for a given problem can be exploited to yield useful information about the fitness function. This can be provided alongside the global optima found during the search, adding value to the optimisation process for the decision maker.

We begin in Section 2 by making the necessary definitions and by summarising the MFM. In Section 3, we note how the model has been mined in the context of other applications. In Sections 4 and 5 we describe a civil engineering optimisation problem and some previously published optimisation results. In Section 6 we then present some new

GECCO'16 Companion, July 20-24, 2016, Denver, CO, USA

DOI: http://dx.doi.org/10.1145/2908961.2931711

results that demonstrate the mining of the MFM for this problem to support decision making. Finally, in Section 7 we draw some conclusions.

2. MARKOV NETWORK FITNESS MODEL

We begin with a recap of the Markov Network Fitness Model. Let $\Omega = \{0,1\}^n$ be the search space (that is, bit string encoded solutions). $f(x) \Rightarrow \mathbb{R}$ is the fitness function and $X = (X_1, \ldots, X_n)$ is the variable vector. $X_i = x_i$ denotes that variable X_i has value x_i, and $x = x_1 \ldots x_n$ denotes a joint configuration of X.

A *neighbourhood structure* μ is a relation on the variables $\{X_1, \ldots, X_n\}$. For each X_i, $\mu(X_i)$ is a subset of $\{X_1, \ldots, X_n\}$, called the neighbourhood of X_i, satisfying:

$$\begin{cases} X_i \notin \mu(X_i) & \forall i \\ X_i \in \mu(X_j) \Leftrightarrow X_j \in \mu(X_i) & \forall i,j \end{cases} \tag{1}$$

The neighbourhood structures model the linkage between variables.

A joint probability distribution on X is denoted $p(X)$. Similarly, $p(x)$ denotes the probability $p(X = x)$ and $p(x_i)$ denotes the probability $p(X_i = x_i)$.

A *Markov Random Field* (MRF) [22] consists of a set of random variables X, a neighbourhood structure μ, and a joint probability distribution $p(X)$. A defining property of an MRF is that the distribution of a particular variable depends only on its neighbours.

Potential functions $V_K(x)$ for each clique K (set of mutually neighbouring variables) given a configuration x are defined as follows:

$$\text{For } K = \varnothing \qquad V_\varnothing(x) \equiv 1 \; \forall \; x \tag{2}$$

$$\text{For } K = \{X_i\} \qquad V_i(x) = \begin{cases} 1 & x_i = 1 \\ -1 & x_i = 0 \end{cases} \tag{3}$$

$$\text{For } K \subseteq \{X_1, \ldots, X_n\}, \quad |K| \geqslant 2, \quad V_K(x) = \prod_{X_i \in K} V_i(x) \tag{4}$$

We define an energy function as a weighted sum of clique potentials:

$$U(x) = \sum_K \alpha_K V_K(x) \tag{5}$$

The Hammersley-Clifford Theorem (see [1]) states that the probability distribution of a MRF factorises as a Gibbs distribution:

$$p(X) = \frac{e^{-U(X)/T}}{Z} \tag{6}$$

Here, T is a temperature coefficient and Z is the normalising constant (never explicitly computed in practice) $Z = \sum_{y \in \Omega} e^{-U(y)/T}$. For this paper, T is constant, with a value of 1. The probability distribution is completely determined by the neighbourhood structure and its associated clique potential parameters α_K. The set of clique potential parameters Ψ will now be referred to as the parameters of the MRF.

Given a MRF, we can construct a graph G from the neighbourhood structure. The nodes of G correspond to the variables in the set X. We add an edge to G between two nodes if and only if the corresponding variables are neighbours. The neighbourhood structure can either be inferred from data or supplied using domain-specific knowledge: in the problems we will study it is supplied based on existing knowledge of the problem.

We define a Markov network model of a set of solutions to be a pair (G, Ψ) where G is a linkage structure and Ψ is the parameter set of the associated MRF learned from the solutions. The key idea of the MFM is identifying the Gibbs distribution of the Markov network with the mass distribution of fitness estimated from the population:

$$p(x) = \frac{e^{-U(x)/T}}{Z} = \frac{f(x)}{\sum_{y \in \Omega} f(y)} \tag{7}$$

Sampling this distribution will generate high fitness individuals with high probability. This distribution can be estimated by identifying corresponding terms for each solution in the expressions forming the right-hand equality in (7). This gives, for each solution $x = x_1, \ldots, x_n$, a negative log relationship between the fitness function and the MRF:

$$-lnf(x) = U(x) = \sum_K \alpha_K V_K(x) \tag{8}$$

The clique potential functions correspond to the well-known Walsh Transform [2] which has been widely used in the analysis of fitness functions in binary spaces [2,13,14]. These are a set of rectangular waveforms taking the values +1 and -1 which can represent any bit string encoded fitness function (similar to the use of Fourier transforms for representing analogue waveforms). With a large enough population of solutions and their fitnesses, (8) yields a system of equations in the parameters. The parameters can be estimated by solving this using a least-squares approximation. (It has been observed [31] that this stage can be seen as a linear regression problem).

With the parameters specified, (8) becomes a model of the fitness function in terms of the parameters. We call this the *Markov Fitness Model* (MFM) of f, and we can use this model to predict the fitness $f(x)$ for solutions. Further background to the MFM can be found in [5]. Several publications, including [29], describe how this is used in the DEUM EDA.

3. MINING THE MFM

We now consider how the MFM may be further exploited. [4, 5] also explored in more detail how the α values of G can be mined to yield insights into the original fitness function and, in particular, the region around the global optima. In short, the process is as follows. Equation (8) specifies a negative log relationship between energy and fitness in the MFM. This means that minimising energy is equivalent to maximising fitness. For a univariate term, $V_i(x)$ (i.e. corresponding to a single x_i), if $\alpha_i > 0$, setting $x_i = 0$ will minimise energy and thus maximise fitness. Conversely, if $\alpha_i < 0$, setting $x_i = 1$ will maximise fitness. For terms with two variables $V_{i,j}(x)$, then having $\alpha_{i,j} > 0$ requires that x_i != x_j to minimise energy and thus maximise fitness. Having $\alpha_{i,j} < 0$ requires that $x_i == x_j$ to maximise fitness. So, the

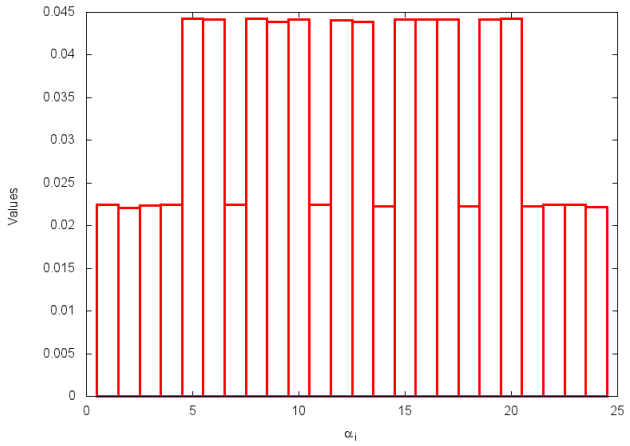

Figure 1: Pairwise coefficients for checkerboard

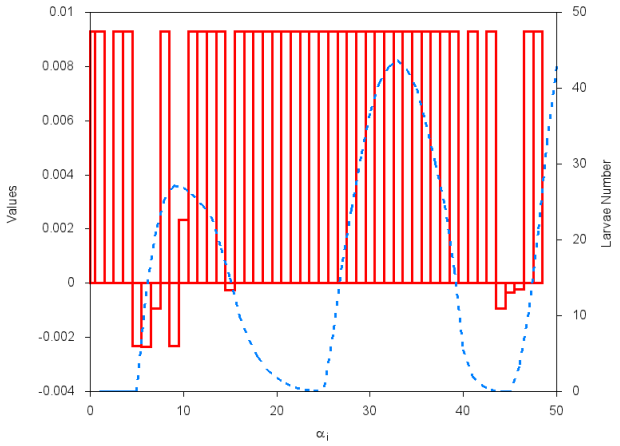

Figure 2: Univariate coefficients for the biocontrol problem and larval lifecycle

signs of the α_K point towards the values taken by variables in the globally optimal solutions. The magnitude indicates the sensitivity of $f(x)$ to the values taken by each clique.

Two examples explored in [4, 5] were the toy benchmark problem Checkerboard, and a biocontrol problem. We now replicate part of those results for illustration. With Checkerboard, the goal is to maximise the number of cells with oppositely-valued neighbours when the bit string is laid out in a grid. The structure used for the MFM was the set of all univariate terms (i.e. one for each variable x_i), and pairwise terms corresponding to the lattice structure of the checkerboard (i.e. one for each neighbouring pair x_i,x_j). The univariate terms were all around zero, implying that each cell could either be 0 or 1 in the optima. The pairwise coefficients given from the MFM trained on solutions for the 25-bit variant of this problem are given in Figure 1. All the coefficients are positive, implying that neighbouring cells should take opposite values. In addition, the coefficients for several of the pairs have a magnitude twice that of the others: subsequent analysis revealed that these correspond to the pairs of cells in the centre of the checkerboard.

The bio-control problem has the objective of minimising the growth of insect larvae on mushrooms by choosing the optimal times to spray the crop with nematode worms. Solutions are encoded as bit strings: 50 bits representing times at which the bio-control spray is applied or not applied. The MFM structure in this case comprised univariate terms, and pairwise terms representing a chain, i.e. $V_{i,i+1}(x)$ for $1 < i < n-1$. The univariate coefficients (each corresponding to one bit) for the MFM applied to this problem are given in Figure 2. Most are positive, indicating that no spray should be applied at that point. However, a few are negative, indicating points when the spray should be applied. These coincide with growth points in the life cycle of the pest insect larvae being targeted (represented by the blue dotted line).

For both of these problems, the MFM coefficients have a clear relationship with the underlying problem, giving pointers towards the optimal solutions. Indication is also given of the sensitivity of the objective to particular variables or variable interactions. Furthermore, for both problems, the MFM was generated using only a few hundred randomly generated solutions. Further analysis, results and explanation for these problems is given in [4, 5].

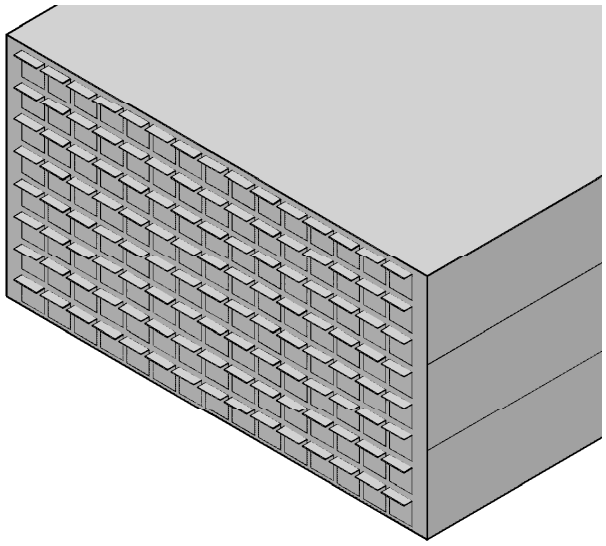

Figure 3: Fully glazed building façade

4. CASE STUDY: CELLULAR WINDOWS

The problem forming the focus of our case study has previously been presented in [9, 32, 33]. We seek to optimise the size, shape and position of windows placed on a building façade; the goal is a design which minimises energy use and capital construction cost.

The building in this study is based on an atrium of a commercial building located in Chicago, USA. The atrium is 15m wide, 15m long and 8.1m high with only the southern façade exposed to the external environment. The other three sides of the atrium are connected to interior spaces that are controlled to have the same thermal conditions as the atrium. The roof, internal partition walls and the external façade have a light-weight construction; the floor is

constructed from uninsulated concrete; and the window cells have a double-glazed construction. The external wall is divided into 120 cells which may be glazed, in a grid 15 wide and 8 high. Figure 3 shows the fully glazed building.

4.1 Objectives

4.1.1 Energy

The first objective is to minimise the energy use of the building. This is the unweighted sum total of the energy used by heating, cooling and lighting systems over a specified period in a particular set of environmental circumstances. This is relatively complex as the energy consumption for the systems varies with the glazing in different ways [33]. Electric lighting demand is reduced by incoming sunlight. In contrast, at different times of day and in different months, solar gain can increase cooling energy demand and decrease heating energy demand. Furthermore, heat losses through the glass at night have the opposite effect.

These figures are computed by the EnergyPlus building simulation package [10] and the process is explained in more detail in [32]. We have chosen EnergyPlus as it is a freely available simulation in common use among the building design community. EnergyPlus can be run to simulate the building's performance over an entire year, using a publicly available weather data set for the location. For this problem, a single run of the simulation takes around 1-2 minutes on a a reasonably fast (Intel i7) CPU, giving the original motivation for the use of surrogates to speed up the optimisation.

4.1.2 Cost

The second objective is the minimisation of the construction cost for the building given the specified window configuration. This is a straightforward linear function of the number of windows n_w and does not involve the simulation software. The total cost c is defined in equation (9).

$$c_w = 112(120 - n_w) + 350n_w \tag{9}$$

4.2 Variables and Encoding

The problem naturally lends itself to a binary representation. The wall is divided into 120 cells in a 15 x 8 grid. Each cell may be glazed or unglazed. This translates into a 120 variable bit string, where a bit is true for a glazed cell and false for an unglazed one. Previous papers [9, 32, 33] have considered several constraints and adding shades to the windows: for simplicity we omit these from the current work.

5. OPTIMISATION RESULTS

Previous publications [9, 33] have presented comparisons and analysis of results from several multi-objective evolutionary algorithms applied to this problem. The focus of the present paper is on mining a surrogate model of the problem, rather than on the optimisation process, so for convenience we replicate the best set of results from [33]. The attainment curve in Figure 4 represents the Pareto-optimal solutions from the combined final populations of 32 repeat runs of NSGA-II [11]. The specific choice of algorithm is unimportant for this work and could be substituted by another that makes use of fitness to drive the search; however NSGA-II was found to perform well for this problem in a previous study [9]. The algorithm used binary tournament

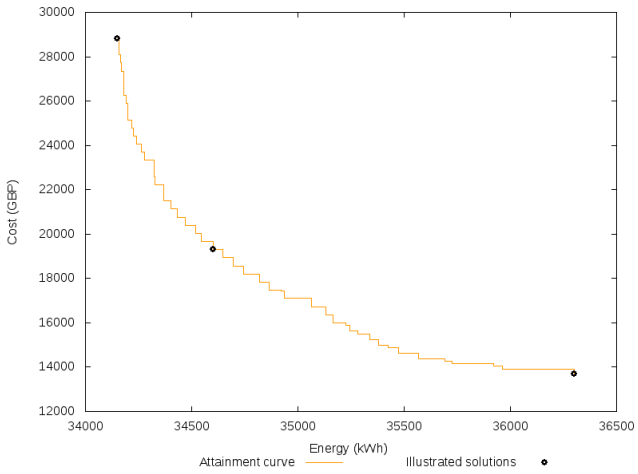

Figure 4: Best attainment curve from multi-objective optimisation run.

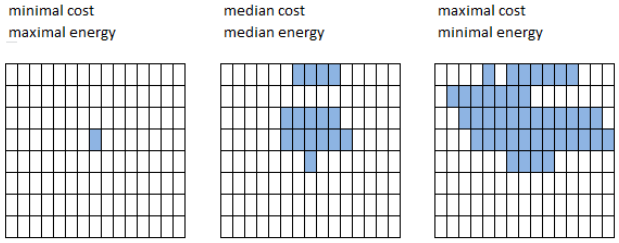

Figure 5: Minimal, median and maximal cost solutions from the best attainment curve. White cells are unglazed, blue cells are glazed.

selection; 100% crossover rate using uniform crossover; single bit-flip mutation for each new solution; population size 30 and a stopping criterion of 5000 unique evaluations.

The minimal, median and maximal cost solutions in this approximated Pareto front are illustrated in Figure 5, minus the shading overhangs that were present in the original paper. There is a substantial range of capital costs in the solutions, reflecting the extra expense of glazing over that of unglazed wall. The range of energy consumptions for the solutions is more modest, but is still around 6% of the maximal energy consumption, representing considerable savings in emissions and energy costs over the life of the building.

The approximated Pareto-optimal trade-off and the specific designs in each solution are already of great value to a decision maker. However, there are some limits to this. For example, it would appear that as a result of the algorithm's randomness, it has missed the lowest cost solution (that is, zero glazing). It has also produced slightly odd shapes of glass on the higher-cost solutions. It would be helpful for the decision maker to know what the impact might be of making small aesthetic changes to the Pareto-optimal solutions. Essentially, we would like to *add value* to the optimisation process, by providing further information on the problem. In [33], two approaches were taken to adding value.

Firstly, the analysis considered the full Pareto-optimal set, and presented heat maps (Figure 6) showing the frequency that each cell was glazed within the approximated Pareto-

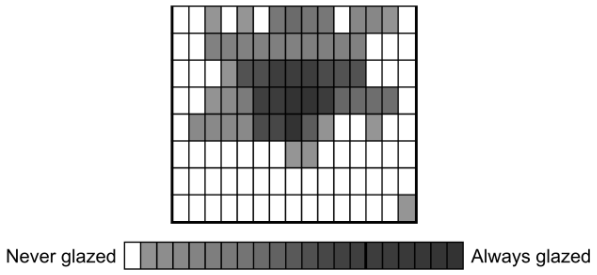

Figure 6: Frequency that each cell was glazed within the approximated Pareto optimal set.

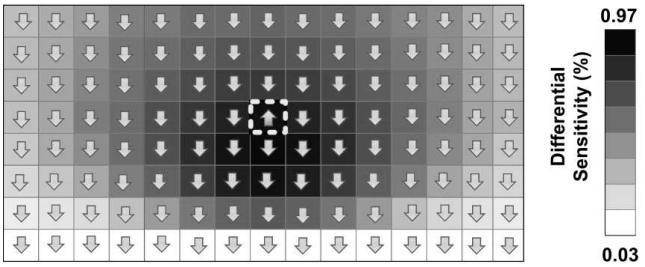

Figure 7: Local sensitivities around the minimal cost solution. Arrows indicate the direction of change in energy consumption caused by mutating that bit in the solution and the shading shows the magnitude of the change.

optimal solutions. This shows the glazed cells that are common to all or most of the optimal solutions (it is unlikely that these can be changed without impacting on optimality). It was also cheap to compute as it is a simple summation of values within the existing approximated Pareto-optimal solutions. However, it is unable to show how the individual cells impact on the objectives separately.

Secondly, the analysis considered local sensitivity. For selected solutions, each cell was flipped from glazed to unglazed (or vice-versa), and the change in energy use determined. This is illustrated in Figure 7. The local sensitivities help to identify cells that were glazed or unglazed as a result of noise coming from using a stochastic algorithm, and those which could be changed without impacting negatively on the objectives. However, the approach has the disadvantage that it requires further runs of the building performance simulation.

Both pieces of analysis were useful in their own right, but could be supplemented with further information.

6. MINING THE SURROGATE

We will now consider how the surrogate can be mined to add value to the optimisation results, beyond that provided in the analysis replicated in the previous section. We consider both objectives to allow for benchmarking of the approach. Note that, in practice for this problem, the surrogate would likely only be used for the energy objective as the cost objective is very fast to compute.

For these experiments, the surrogate model of fitness was constructed in parallel with the optimisation run. This allows for direct comparisons between the additional information provided by the surrogate and that arising from the original optimisation. Work to consider the speed up provided by using this surrogate in place of some building performance simulations is ongoing. In the present work, solutions evaluated as part of the NSGA-II run were used as training data for the MFM. This means that currently, solutions are only passed from the algorithm to the surrogate, but nothing is returned in the other direction.

The structure for the MFM (the neighbourhoods for each x_i) was fixed. Two sets of experiments were performed using different structures for the MFM.

6.1 Lattice structure

Initially, a lattice structure was adopted for the MFM. This was based on the intuition that applying glazing to one cell would impact on whether glazing should be applied to cells next to it. While this is only true for the energy objective, the lattice structure was also considered for cost to allow for comparison. Included in the MFM were all 120 univariate V_K (that is, one term for each of the glazed/unglazed cells). Also included were the 240 pairwise V_K representing neighbouring cells on the façade. This means that there are 361 parameters in the model (including the constant term), and according to [5], a training population of around 1.1x this value should be used to obtain a good model.

Thus, the fittest (that is, lowest energy or cost) 400 of the first 1000 solutions visited by NSGA-II in each run were used as the training data for a least squares fit to estimate the α_K in the MFM (as per equation (8)). Two MFMs were formed for each repeat run of NSGA-II: one for the energy objective and one for the cost objective. These models have a strong predictive capability: this was demonstrated by using each model to predict the objective values for 400 randomly generated solutions. The r^2 values comparing the predicted objective values with the true objective values coming from the simulation were 0.982 for energy and 0.997 for cost.

The mean and standard deviation for each coefficient α_K in the MFM was then calculated over all the energy MFMs and over all the cost MFMs. These values are plotted in Figures 8 and 9. The jump in the value at α number 120 coincides with the change from univariate α_Ks to the pairwise α_Ks. It is immediately apparent that for both energy and cost objectives, the pairwise α_K values are all near zero. This means that they have little to no influence on either objective: it would seem that only the univariate α_K have any influence on the objectives (having non-zero values in the MFMs for both objectives) and our intuition on the appropriate structure was incorrect.

6.2 Univariate structure

A second set of experiments repeated the process in the first, using a univariate structure for both MFMs. That is, only the 120 α_Ks for each of the variables were included in the model. The model was trained on the fittest 140 of the 400 solutions arising from the first few generations of the optimisation run. r^2 for the models on 400 randomly generated solutions was 0.993 for energy and 0.998 for cost.

Figures 10 and 11 give the mean α_Ks for the energy and cost MFMs respectively. The mean values for these coefficients have also been rendered in Figures 12 and 13 as a grid so where the coefficient's location corresponds to the cell that it represents on the façade. In the latter two figures, the cells have been coloured to show each coefficient's value relative to those of the others: high values being blue,

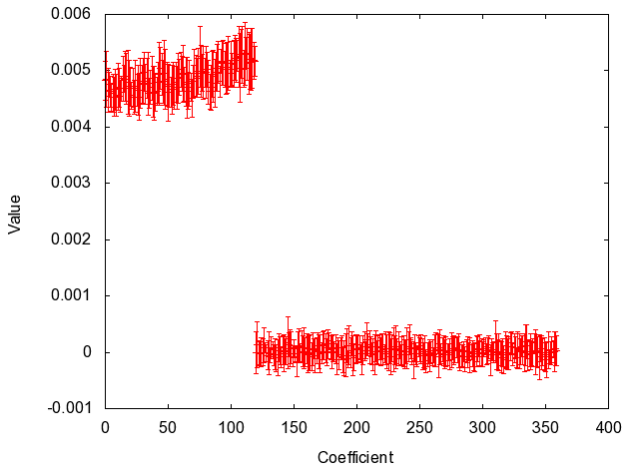

Figure 8: The mean coefficient (α_K) values for the energy MFMs. 1-120 are the univariate α_K, and 121-360 are the pairwise α_K between neighbouring cells on the façade. Error bars represent one standard deviation.

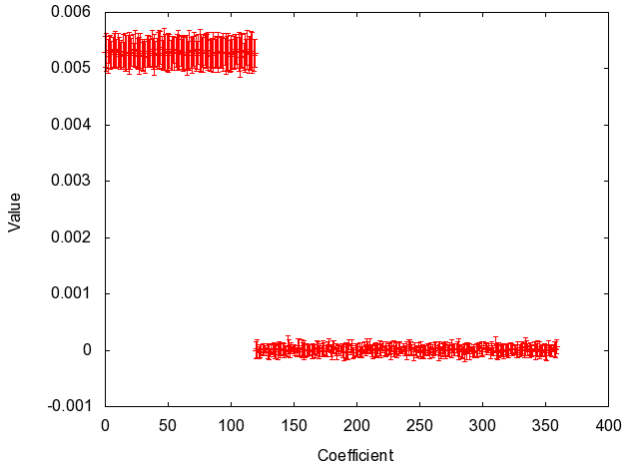

Figure 9: The mean coefficient (α_K) values for the cost MFMs. 1-120 are the univariate α_K, and 121-360 are the pairwise α_K between neighbouring cells on the façade. Error bars represent one standard deviation.

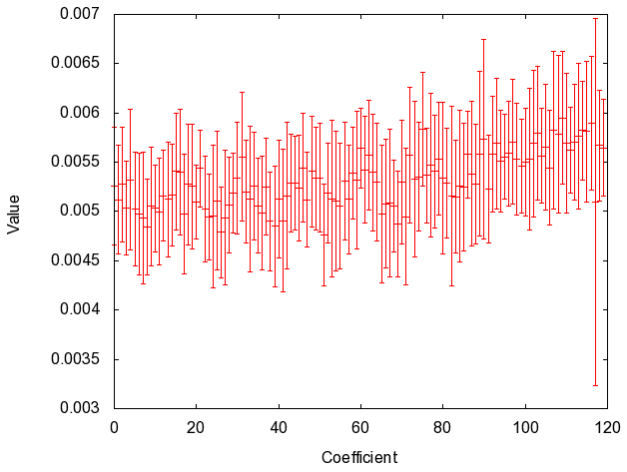

Figure 10: The mean coefficient (α_K) values for the energy MFMs. 1-120 are the univariate α_K, starting with that corresponding to the top-left cell on the façade, and working along each row to the bottom right. Error bars represent one standard deviation.

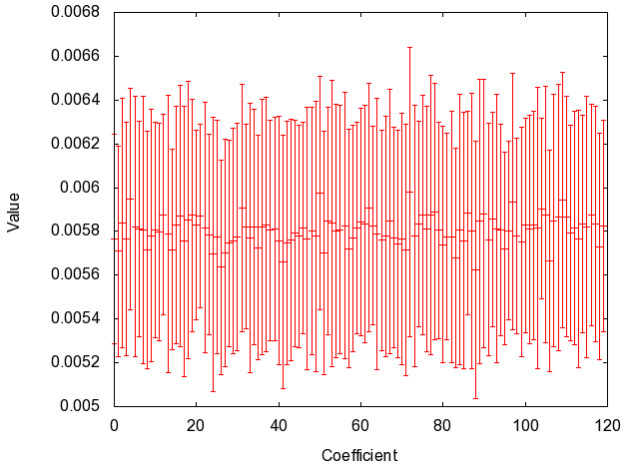

Figure 11: The mean coefficient (α_K) values for the cost MFMs. 1-120 are the univariate α_K, starting with that corresponding to the top-left cell on the façade, and working along each row to the bottom right. Error bars represent one standard deviation.

through white to low values being red. Recall (from the end of Section 3) that a positive coefficient α_i suggests that the global optimum should have $x_i == 0$, and a negative α_i suggests that the global optimum should have $x_i == 1$. In this case, all the α_k in the model are positive.

For the cost objective, the magnitudes of the α_k are highly similar, indicating that optimal solutions should be unglazed and that the individual cells make equal contributions to the cost (i.e. the objective is equally sensitive to all cells). This matches with the problem definition, whereby an equal cost is associated with each cell in the façade.

For the energy objective, there is a clear (though small) bias towards the lower and outer edges of the façade. This can also be seen in the higher values to the right of Figure 12. This suggests that cells in those regions should not be glazed, and any glazing that is present should be concentrated in the upper centre (biased slightly to the right / East). This matches the patterns seen in the analysis of the Pareto-optimal front (Figure 6 and the local sensitivity analysis (Figure 7). However, there are some benefits of mining this information from the surrogate model in addition to (or in place of) the other analyses. The bias towards the centre is identified specifically as driven by the energy objective not the cost objective (in contrast to simple analysis of the variables in the Pareto-optimal solutions, where trends could be driven by either objective). The analysis is not simply rooted in the final population from the simulation, but models representing solutions spanning several

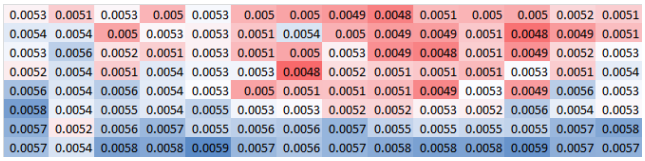

Figure 12: Mean coefficient values for the energy MFM, arranged to match the locations of cells on the façade. Blue cells have high values, white medium, and red have low values.

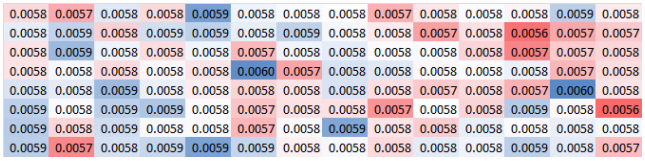

Figure 13: Mean coefficient values for the cost MFM.

generations. No additional evaluations are required as the model was constructed from solutions already evaluated during the optimisation (in contrast to the local sensitivity analysis). It is also concordant with the real-world problem: in practice we would expect that any glazing should be central and high up the façade to allow the maximum penetration of daylight into the atrium with less glazed area, balancing heat gain, heat losses, and lighting needs (although it is less clear specifically how much glass there should be or precisely where it should be placed, thus motivating the use of optimisation). A slight bias to the right (East) also catches more of the early morning sunlight.

It is important to note that this information in the model has come from 400 evaluations that were performed anyway as part of the search process. The MFM is a surrogate for the fitness function, and can be used to reduce the further number of simulations required. Consequently, the qualitative information that the MFM provides about the relationship between the problem variables and the objectives comes with little to no extra cost in terms of CPU time.

7. CONCLUSION

Value-added optimisation is a philosophy of presenting a decision maker with more than just optimal solutions for a problem. This can be an indication of the optimality of the solutions, relationships between problem variables, sensitivity of the objectives to the variables or simply greater insight into the underlying problem. This can have several benefits:

- Knowing the sensitive variables, the decision maker can adjust the returned solutions to reflect factors not considered by the optimisation (e.g. aesthetics), aware of the likely impact on optimality. For example, the optimal solutions for the glazing problem have quite odd window shapes: these could be reshaped for more visual appeal within the region suggested by the model.
- The model may indicate where a metaheuristic has not fully converged on the global optimum.
- If the returned solutions match the conclusions drawn from the model, the decision maker can have added confidence in the optimality of the results.
- If the results are counter-intuitive, they may suggest where potential problems are located in the model used to calculate the objectives.
- The model may point towards the globally optimal solutions long before the algorithm has converged on those solutions. For example, with the glazing problem, analysis of the model suggested the overall glazing shape based on observations from the first 1000 solutions of a 5000 solution run. With long-running fitness evaluations, this could allow for early analysis of what the optima are likely to look like, which is particularly helpful if they indicate flaws in the model that might necessitate a restart.

This paper has applied a surrogate, the Markov network fitness model (MFM), to the problem of placing glazing on a building façade. We have presented some analysis of the model coefficients for this problem, showing how mining the MFM can be used to add value to the results of the optimisation run. As this model is constructed as part of the optimisation run, the additional information that it contains can be provided for little - if any - additional computational cost. This mirrors earlier work [4, 5] on several benchmark functions showing how the MFM coefficients point towards the global optima for different fitness functions.

Obviously, in order to generalise this to a much wider range of problems, considerably more work needs done to establish a framework for analysis of the coefficients in a systematic way. It would also be interesting to consider how the MFM, or other surrogate models, could be applied to problems with encodings other than bit strings, and mined in the same way. What this work has done is set out the possibility that surrogate models can be used to supplement the optimisation process, enriching the information available to the decision maker.

Acknowledgment

A. Brownlee was part-funded in this work by UK EPSRC grants EP/J017515/1 (DAASE) and EP/N002849/1 (FAIME). Data generated during this work (fitnesses and model parameters) is available at http://hdl.handle.net/11667/74.

8. REFERENCES

[1] Julian Besag. Spatial interaction and the statistical analysis of lattice systems. *Jnl. of the Royal Stat. Soc. Series B (Methodological)*, 36(2):192–236, 1974.

[2] A.D. Bethke. *Genetic Algorithms as Function Optimizers*. PhD thesis, University of Mitchigan, Ann Arbor, MI, 1980.

[3] A. E. I. Brownlee, O. Regnier-Coudert, J. A. W. McCall, and S. Massie. Using a Markov network as a surrogate fitness function in a genetic algorithm. In *Proc. IEEE CEC*, pages 4525–4532, Barcelona, 2010.

[4] A. E. I. Brownlee, Y. Wu, J. A. W. McCall, P. M. Godley, D. E. Cairns, and J. Cowie. Optimisation and fitness modelling of bio-control in mushroom farming using a Markov network EDA. In *Proc. GECCO*, pages 465–466, Atlanta, GA, USA, 2008.

[5] A.E.I. Brownlee, J.A.W. McCall, and Q. Zhang. Fitness Modeling With Markov Networks. *IEEE T. Evolut. Comput.*, 17(6):862–879, 2013.

[6] Alexander Brownlee and Jonathan Wright. Solution analysis in multi-objective optimization. In *Proc.*

Building Simul. and Optim. Conf., pages 317–324, Loughborough, UK, 2012. IBPSA-England.

[7] Alexander E. I. Brownlee, Olivier Regnier-Coudert, John A. W. McCall, Stewart Massie, and Stefan Stulajter. An application of a GA with Markov network surrogate to feature selection. *Int. J. Syst. Sci.*, 44(11):2039–2056, 2013.

[8] Alexander E.I. Brownlee and Jonathan A. Wright. Constrained, mixed-integer and multi-objective optimisation of building designs by NSGA-II with fitness approximation. *Applied Soft Computing*, 33:114–126, 2015.

[9] Alexander E.I. Brownlee, Jonathan A. Wright, and Monjur M. Mourshed. A multi-objective window optimisation problem. In *Proc. GECCO*, GECCO '11, pages 89–90, New York, NY, USA, 2011. ACM.

[10] D.B. Crawley, L.K. Lawrie, F.C. Winkelmann, W.F. Buhl, Y.J. Huang, C.O. Pedersen, R.K. Strand, R.J. Liesen, D.E. Fisherm, M.J. Witte, and J. Glazer. EnergyPlus: creating a new generation building energy simulation program. *Energ. Buildings*, 33(4):319–331, 2001.

[11] Kalyanmoy Deb, Amrit Pratap, Sameer Agarwal, and T. Meyarivan. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE T. Evolut. Comput.*, 6(2):182–197, 2002.

[12] Rudolf J. Freund, William J. Wilson, and Ping Sa. *Regression Analysis: Statistical Modeling of a Response Variable*. Academic Press, 2006.

[13] D. Goldberg. Genetic Algorithms and Walsh Functions: Part I, A Gentle Introduction. *Complex Systems*, 3(2):129–152, 1989.

[14] D. Goldberg. Genetic Algorithms and Walsh Functions: Part II, Deception and its Analysis. *Complex Systems*, 3(2):153–171, 1989.

[15] Mark Hauschild and Martin Pelikan. An introduction and survey of estimation of distribution algorithms. *Swarm Evol. Comput.*, 1(3):111 – 128, 2011.

[16] Mark Hauschild, Martin Pelikan, Kumara Sastry, and Claudio Lima. Analyzing probabilistic models in hierarchical BOA. *IEEE T. Evolut. Comput.*, 13(6):1199–1217, December 2009.

[17] Mark W. Hauschild, Martin Pelikan, Kumara Sastry, and David E. Goldberg. Using previous models to bias structural learning in the hierarchical BOA. In *Proc. of the 10th annual conf. on Genetic and evolutionary computation*, GECCO '08, pages 415–422, New York, NY, USA, 2008. ACM.

[18] Yaochu Jin. Surrogate-assisted evolutionary computation: Recent advances and future challenges. *Swarm Evol. Comput.*, 1(2):61–70, 2011.

[19] Yaochu Jin, Markus Olhofer, and Bernhard Sendhoff. A framework for evolutionary optimization with approximate fitness functions. *IEEE T. Evolut. Comput.*, 6(5):481–494, Oct 2002.

[20] P. Larrañaga and J. A. Lozano. *Estimation of Distribution Algorithms: A New Tool for Evolutionary Computation*. Kluwer, Boston, 2002.

[21] Rui Li, M.T.M. Emmerich, J. Eggermont, E.G.P. Bovenkamp, T. Back, J. Dijkstra, and J.H.C. Reiber. Metamodel-assisted mixed integer evolution strategies and their application to intravascular ultrasound image analysis. In *Proc. IEEE WCCI*, pages 2764–2771, 2008.

[22] S. Z. Li. *Markov random field modeling in computer vision*. Springer-Verlag, 1995.

[23] J. A. Lozano, P. Larrañaga, I. Inza, and E. Bengoetxea. *Towards a New Evolutionary Computation: Advances on Estimation of Distribution Algorithms (Studies in Fuzziness and Soft Computing)*. Springer-Verlag, 2006.

[24] Glen D Rodriguez Rafael and Carlos Javier Solano Salinas. Empirical study of surrogate models for black box optimizations obtained using symbolic regression via genetic programming. In *Proc. GECCO Companion*, pages 185–186. ACM, 2011.

[25] Roberto Santana. Estimation of distribution algorithms: from available implementations to potential developments. In *Proc. GECCO*, pages 679–686, New York, NY, USA, 2011. ACM.

[26] Roberto Santana, Concha Bielza, Jose A. Lozano, and Pedro Larrañaga. Mining probabilistic models learned by EDAs in the optimization of multi-objective problems. In *Proc. of the 11th Annual Conf. on Genetic and Evolutionary Comp. (GECCO 2009)*, pages 445–452, New York, NY, USA, 2009. ACM.

[27] Roberto Santana, Carlos Echegoyen, Alexander Mendiburu, Concha Bielza, Jose A. Lozano, Pedro Larrañaga, Rubén Armañanzas, and Siddartha Shakya. Mateda-2.0 : A MATLAB package for the implementation and analysis of estimation of distribution algorithms. *Journal of Statistical Software*, 35(7), 2010.

[28] S. K. Shakya, A. E. I. Brownlee, J. A. W. McCall, F. Fournier, and G. Owusu. A fully multivariate DEUM algorithm. In *Proc. IEEE CEC*, pages 479–486. IEEE Press, 2009.

[29] Siddhartha Shakya, John McCall, and Alexander Brownlee. DEUM - distribution estimation using Markov networks. In Siddhartha Shakya and Roberto Santana, editors, *Markov Networks in Evolutionary Comp.*, pages 55–71. Springer, 2012.

[30] Es Tresidder, Yi Zhang, and Alexander I. J. Forrester. Optimisation of low-energy building design using surrogate models. In *Proc. of the Building Simulation 2011 Conf., Sydney, Australia*, pages 1012–1016, 2011.

[31] Gabriele Valentini, Luigi Malagò, and Matteo Matteucci. Optimization by ℓ1-Constrained Markov fitness modelling. In *Proc. LION (LNCS 7219)*, pages 250–264. Springer, 2012.

[32] Jonathan Wright and Monjur Mourshed. Geometric optimization of fenestration. In *Proc. IBPSA Building Simulation*, pages 920–927, Glasgow, Scotland, 2009.

[33] Jonathan A. Wright, Alexander E.I. Brownlee, Monjur M. Mourshed, and Mengchao Wang. Multi-objective optimization of cellular fenestration by an evolutionary algorithm. *J. Build. Perform. Sim.*, 7(1):33–51, 2014.

[34] Jonathan A Wright, Mengchao Wang, Alexander EI Brownlee, and Richard A Buswell. Variable convergence in evolutionary optimization and its relationship to sensitivity analysis. In *Proc. IBSPA BSO 2012*, pages 102–109. Loughborough University© IBPSA-England, 2012.