Development of Energy Simulation Models from Smart Meter Data using Inverse Modelling and Genetic Algorithms

Daniel Costola[1], Ana Paula Melo[2] and Loïc Jacob[1]
[1]Department of Mechanical and Aerospace Engineering, University of Strathclyde, Glasgow, G1 1XJ, UK
[2]Laboratory for Energy Efficiency in Buildings, Federal University of Santa Catarina, Brazil

Abstract

To combat anthropogenic climate change the energy use of buildings must be reduced significantly within the coming decades. There is a pressing need for cheap and accurate baseline building energy models to inform owners on opportunities for energy efficiency improvements. The increasing penetration of smart meters in existing buildings provides a wealth of data that can be leveraged with dynamic simulation models to achieve energy savings. This paper describes a method to generate physically driven dynamic simulation models from metered data using inverse modelling techniques. The application of inverse modelling methods to generate baseline building energy models using a physically driven software is novel in the field, with potential for cheap prediction of the impacts of energy efficiency upgrades in stock modelling. The tools used are the dynamic simulation software ESP-r and the genetic algorithm from MATLAB's Global Optimisation Toolbox. Building data and gas meter readings for the year of 2013 provided by the University of Strathclyde were used to drive the modelling of an office building in the campus. Results show that the optimised models have an energy consumption with 5% difference to the metered data provided. Convergence can be achieved within a reasonable number of generations and with a population size that is not prohibitively large. However, adding variables significantly increases the computation time required for convergence. Further work could explore the limitations of the method when applied to complex models with higher number of variables.

Introduction

Humanity's use of fossil fuels to produce energy is a key driver of climate change (IPCC, 2014). Approximately 40% of all energy in the EU is consumed in buildings (EC, 2010). To combat climate change, legislation has been introduced to increase the energy efficiency of new and existing buildings at an EU and national (UK) level (EC, 2010; DCLG, 2012). For energy targets in the built environment to be met most of the savings will have to come from retrofits as up to 87% of domestic and many non-domestic buildings that will be standing in 2050 have already been built (Boardman, 2007). There is therefore a need for building energy models that can predict the likely energy effects of efficiency measures on existing buildings.

Building energy modelling

The current baseline energy modelling methods in the UK, the Standard Assessment Procedure (England) and National Calculation Method (Scotland), are limited in their ability to provide detailed feedback and accurate predictions on energy performance in buildings due to their broad assumptions (Kelly et al., 2012). Dynamic simulation models perform better but can be expensive and time consuming to develop.

Broadly, dynamic simulation modelling methods are either based on first principles (based on physical equations and detailed building information) or data-driven (using statistical methods to identify patterns in building performance data in relation to building characteristics). Typically, first-principle models, e.g. ESP-r (ESRU, 2016) and EnergyPlus (US DOE, 2015) are more accurate in their predictions and more transparent in their workings but can be prohibitively expensive and time consuming to develop, along with requiring a high level of detail about the building in question. As such, they are typically only used late in the design process, addressing a limited set of design alternatives rather exploring multiple complex design or retrofit/control scenarios in the building operation phase (Hensen, 2011). Data-driven models are cheaper and quicker to build but can be opaque in their workings. They are often used post-retrofit to identify the effect of Energy Conservation Measures (ECMs). The performance of data-driven models is sensitive to the quality and availability of building data (Foucquier et al., 2013; Harish and Kumar, 2016; Zhang et al., 2015; Zhao and Magoulès, 2012).

In this work, we explore the idea of using inverse modelling techniques based on optimization to simplify the development of building models in first-principle based software (e.g. ESP-r). This could potentially make use of the capabilities of data-driven approaches to facilitate the development of detailed, accurate models based in the laws of physics. This approach is particularly suitable for the modelling of large estates or for building stock modelling of entire cities for energy policy purposes.

https://doi.org/10.26868/25222708.2017.701

Genetic Algorithms

Genetic Algorithms (GAs) have been used to solve discontinuous, in-differentiable complex problems that do not lend themselves to traditional linear methods of solving. They have been shown to find “better solutions with less function evaluations than simulated annealing” (Houck et al., 1995).

GAs produce a population of individuals across the solution space. An individual is made up of a number of variables, or chromosomes, that represent the parameters being optimised by the algorithm. The initial population (the first generation) is produced randomly in order to get an even spread of possible solutions. Each individual is then assigned a fitness value based on an objective fitness function that the user defines. The fitter the individual, the greater the chance that they will produce offspring, in keeping with the “survival of the fittest” concept in natural evolution on which GAs are based. The probabilistic selection of parents (as opposed to a straight ranking) ensure that there is always a non-zero chance that an unfit individual will pass on their genes. This prevents an “elitist strategy” emerging, where the fittest individuals of a population are passed through continuously to successive generations, resulting in a population lacking diversity (Chipperfield et al., 1994).

The next generation is then created based on the fitter individuals of the initial generation. Elements such as genetic crossover between parents and mutation chances change the chromosomes of the offspring so that subsequent generations are more diverse (Figure 1). These stochastic elements, along with the probabilistic individual selection of parents, result in a non-zero chance that every part of the solution space will be explored.

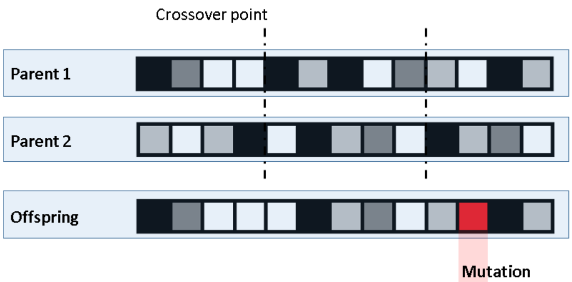

Figure 1: Individuals breeding with crossover and mutation. Each square represents a chromosome

Through successive generations, the individuals in a population become fitter and fitter (by passing on the chromosomes of fit individuals) until they converge. This convergence represents the solution produced by the GA. If there are a range of solutions possible (as is commonly the case in multi-objective optimisation) the GA will produce a range of optimal solutions along the Pareto front, the line of minimum fitness value at which the fitness of the individual relative to one objective cannot be improved without harming the fitness relative to another objective.

Because of the stochastic nature of GAs successive runs will not always produce the same results. It is therefore necessary to run GA optimisation multiple times to ensure that the best solution can be found.

In building performance evaluation, GAs have been used to optimise simplified transient energy models in order to minimise the difference between the model and measured data (Mihail–Bogdan et al., 2016).

GAs explore a solution space in a parallel way as opposed to the more common linear solvers. This, combined with its somewhat random nature in generating populations, means that GAs are much more effective at finding global solutions to complex, non-linear problems. They are much less likely to get “stuck” in local minima. This makes them well suited for solving the complex problem that is generating a model for building energy performance evaluation.

Furthermore, GAs can be applied to any problem type as long as possible solutions can be ranked according to some objective function (Holland, 1992). Again, this is well suited to building simulation models as the fitness of individuals can be assessed based on how the model compares with real data.

GAs have been used to optimise a variety of energy systems in the built environment ranging from distributed urban energy systems, (Ooka and Komamura, 2007) and district waste heat recovery system components (Kayo and Ooka, 2009) to building level analysis of HVAC and building envelope potential (Palonen et al., 2009), multi-objective optimisation of refurbishment potentials (Pernodet et al., 2009) and analysis of natural ventilation potential (Wang and Malkawi, 2015). GAs are usually applied to calibrate or optimize an existing building model that was carefully developed by a professional with expertise in the field. This paper explores the possibility of using GA to drive the modelling process, leveraging the large amount of information provided by smart-meters, as further explained in the following sections.

Smart meters

In the UK there are national targets for every building, domestic and non-domestic alike, to have a smart meter installed by 2020 (BEIS and OFGEM, 2013a; BEIS and OFGEM, 2013b; DECC 2013). This will provide a wealth of high-resolution energy data that can be used in inverse modelling and optimisation methods to develop baseline energy models for existing buildings.

Inverse modelling

Inverse modelling methods are used across a broad range of domains, from weather modelling to determining soil composition (Adam and Branda, 2016; Le Bourgeois et al., 2016).

Inverse modelling methods have two key advantages over traditional physical models (Zhang et al., 2015):

1. They do not require detailed building information (typically just the building form, weather and metered data)
2. They are much less costly and time consuming to develop.

Despite these advantages, inverse modelling techniques are more vulnerable to data quality and availability. Inverse models must be trained with metered data in order to develop a robust relationship between inputs (building data) and outputs (energy consumption, internal temperatures etc.). If the data available doesn't cover a sufficiently long or variable time span, the data is not of a fine enough resolution or the data measurements are inaccurate the model developed will suffer in accuracy.

A common example of an inverse modelling method is Artificial Neural Networks (ANNs). They are used to model non-linear processes, making them better suited to capture the transient nature of aspects of building energy consumption such as occupancy casual heat gains. ANNs are a system of nodes with connections that have an input layer, hidden layers and an output layer (Figure 2). They are a popular choice for short term forecasting of energy loads (Karatasou et al., 2006).

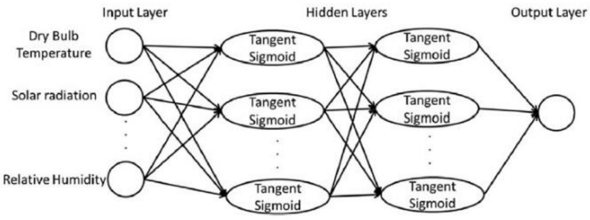

Figure 2: Artificial Neural Network model to estimate building energy use (Zhang et al., 2015)

The key aspect of ANNs that make them a typical example of inverse models is the matching of parameters to data by mathematical or statistical methods. Because the relationships between the inputs and outputs are determined algorithmically, the addition of new parameters to the system is problematic. As such they are not used to predict the effects of ECMs, although they can determine how effective they have been after the event (requiring re-training of the algorithm).

Summary

Figures 3 and 4 demonstrate the difference between the traditional development of physically driven models and the proposed inverse modelling method. Although the inverse model is less accurate than the traditional model it has scope to be automated, with less need for expert modellers and a lower cost of model development. Such cost reductions can potentially make dynamic simulation an affordable solution for stakeholders who otherwise would not consider it.

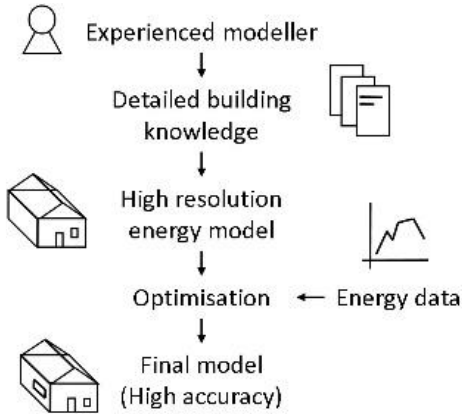

Figure 3: Workflows for traditional modelling

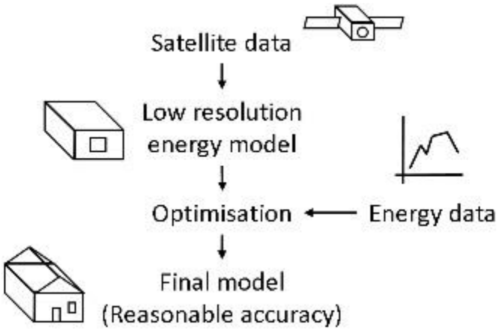

Figure 4: Workflows for inverse modelling proposed in this paper

Through a review of the literature it is apparent that there is a need for a cheap and simple to use tool that can be used to generate baseline energy models for buildings. These models can benefit from inverse modelling and optimisation techniques to apply the data being generated by increasing penetration of smart meters. GAs can be used with dynamic simulation software to achieve this as they are well suited to solving non-linear problems such as building energy simulation.

Despite the range of examples in the literature of GAs used for energy saving analysis in the built environment, the current research is focussed on GAs used to answer complex design questions i.e. to aid in the design of a system that has not yet been built. There is little research into using GAs to create baseline models of existing buildings from metered data.

If government smart meter targets are met, every building in the UK will have high-resolution data by 2020. By providing a cheap and easy (or easier) way of generating baseline models that is more accurate than current standardised methods (such as the SAP and NCM) the potential of smart meter data driving energy efficiency in the built environment can be tapped. Once baseline models can be established, energy saving potentials and their effects can be investigated.

This paper describes a methodology for using smart meter data and genetic algorithms to develop physically driven baseline energy models for existing buildings.

Experiments

A base case model was established on which to test the proposed methodology (described below). Initially, model parameters were randomized then optimized against the base case model (Case 1). An optimization match of 100% is theoretically possible. Comparing the results of the ideal optimized case model to the base case model tested the strength of the proposed methodology, allowing for any uncertainties in metered data or model complexity to be factored out.

Following Case 1, the model was optimized against metered data to determine whether it converged with sensible input parameters (Case 2). This was the true test of the methodology as it would be deployed.

Figure 5 demonstrates the base case establishment, Case 1 and Case 2 described in testing the optimization.

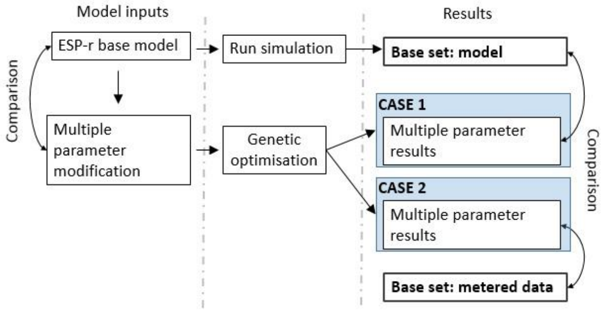

Figure 5: Schematic representation of the methodology adopted in this paper: establishment of a base case, optimisation against base case (Case 1), optimisation against metered data (Case 2)

Building model and data

The Graham Hills building on the University of Strathclyde campus was used in this study (Figure 6). The building was selected for the following reasons: it is fairly homogeneous in zoning, allowing for the building to be reasonably approximated by a simplified single-zone, single floor model; building data including metered data, floor drawings and Energy Performance Certificate information was readily available from the University of Strathclyde Estates Department (AECOM, 2009; University of Strathclyde, 2015a; University of Strathclyde, 2015b). Benchmark values for casual gains (CIBSE, 2006) were used in conjunction with the building data to complete the base model.

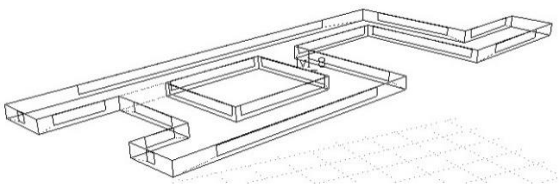

Figure 6: Model of Graham Hills building

A simplified model was used. A more complex model was not necessary for this investigation. A more complex model would have required significantly more time to develop and optimise when the concept had not yet been proven to be worth pursuing.

The base model was built using information from satellite and aerial images (Figures 7, 8 and 9) publicly available (e.g. Google Maps, Bing Maps), combined with information on windows-to-wall ratio (WWR) from images of the façade, also publicly available. This process was not automated in this research, but available APIs from service providers can be used by future works to support automatic extraction (Google, n.d.) of key building features used in this research, such as building shape, orientation, number of floors, WWR, façade and roof absorptivity and likely finishing materials.

The information described above was used to develop a single zone model, representing one floor of the building. No previous knowledge on construction materials, HVAC performance, setpoints, schedules was available and likely values were initially adopted as placeholders.

The model was developed using the software ESP-r. HVAC was not modelled explicitly, relying on ideal controls with infinity capacity. Warm-up period was set to 5 days.

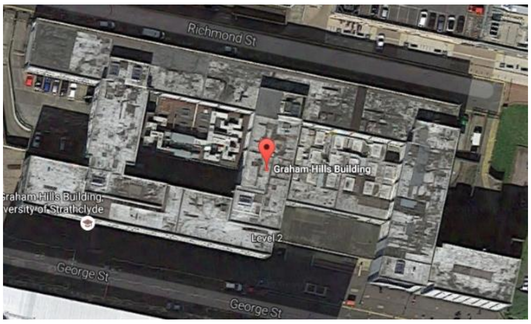

Figure 7: Graham Hills Building satellite view

Figure 8: Graham Hills Building aerial view

Figure 9: Graham Hills Building from George Street

Variables optimised

There is a vast literature on optimization that can be leveraged in studies using inverse modelling. In this study, we focused our attention on variables less explored in previous studies, such as the variation during the day of air change rates and different sorts of casual gains. Material thicknesses were also included in the optimization, as U-values have a major impact in heating dominated climates, such as the Glaswegian climate. A total of 34 variables were optimised in the model (summarised in Table 1).

Table 1: Model variables optimised with GA

CATEGORY	SUB CATEGORIES	NO. VARIABLES
Material thickness	Wall (3 layers), glazing (single layer), floor slab (2 layers)	6
Air change rate	4 day types	5
Casual gains	Occupants, IT, lighting	23
Total		34

Occupant gains are consistent within day types (weekdays, weekends etc. but varied in four blocks throughout the day i.e. night, morning, lunchtime and afternoon. They were calibrated hourly.

Air change rates are separated into two blocks throughout the day: unoccupied (night) and occupied (day).

Measured data

Gas consumption data with 30 minute resolution was used drive the optimization. Measurements performed over a 7 day period in the year 2013 were adopted (Figure 10). The data available was aggregated for the entire building, and energy consumption at the floor modelled in ESP-r was estimated by dividing the total energy consumption by the building floor area (i.e. consumption was assumed equal in all floors).

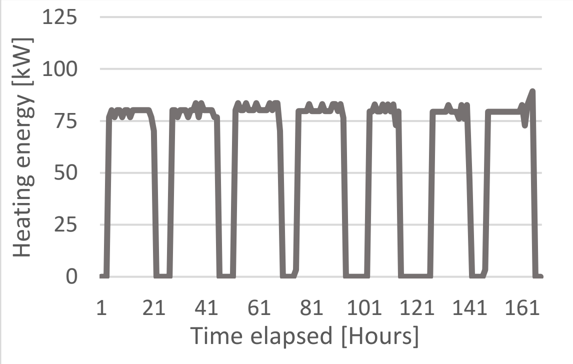

Figure 10: Metered data from Graham Hills Building adjusted for single floor square meterage and boiler efficiency

Genetic Algorithm parameters

GA was done using the optimization package of MATLAB. MATLAB was also used to automate changes in ESP-r input files, making use of scripting tools available in ESP-r.

The GA parameters are outlined in Table 2, below.

Table 2: GA input parameters

PARAMETER	VALUE	EFFECT
nvars	34	Population individuals will have 34 elements
Upper bound	*1x34 array*	Upper limits of variables corresponding to “nvars” elements
Lower bound	*1x34 array*	Lower limits of variables corresponding to “nvars” elements
Generations	100	Maximum number of generations
PopulationSize	510	Population size of (“nvars”*15) as recommended in the GA guidance
TolFun	10	Monitor convergence and stop optimisation
StallGenLimit	5	Number of generations over which “TolFun” is measured

Genetic Algorithm fitness function

The fitness function inputs are the 34 variables being optimised. These variables are read into the ESP-r input text files then a simulation is run. The hourly energy consumption of the model is extracted to the MATLAB domain to be compared with the base results set (base model results and metered data for Case 1 and Case 2 respectively). The model fitness is calculated by summing the square root difference between the model and base results each hour (Equation (1)). This calculation method prevents any model overconsumption being cancelled by underconsumption.

$$Fitness = \sum_{n=hours}^{0} \sqrt{\left(E_{n_{base}} - E_{n_{new}}\right)^2} \quad (1)$$

Discussion and results analysis

The results and discussion will be presented in order then general conclusions presented below.

Case 1: optimisation to base model results

Table 3 shows results from several Case 1 GA runs. A fitter individual does not necessarily perform well in terms of overall energy consumption compared to the base case. Comparing run C301 with run C305, the former has a better fitness but over 10 times worse overall energy performance compared to the base model.

Table 3: Select results for Case 1

RUN	FITNESS	%ENERGY FROM BASE CASE
C301	2571	5.36%
C302	3106	3.63%
C303	3299	1.64%
C304	3317	6.32%
C305	3338	0.78%
C306	3577	2.73%

Table 4 displays select variables from 2 optimisation runs and compares them with the base case. Note the significantly higher air changes accompanied by increases in air changes and occupant gains for both optimised models. Interestingly, the fitter optimised model (C301) has deviated further from the base model in most variables, consistent with the sample shown here.

Table 4: select input variables for Case 1 after optimisation compared with base model

RUN	WALL [m]	AIR CHANGES	INTERNAL GAINS
Base model	0.30	3.0	11200
C301	0.51	5.1	25000
C306	0.35	4.8	24000

Figure 11 shows the convergence of the C301 GA optimisation run, typical of the optimisations carried out. The run terminates early after plateauing in mean fitness. This settling on a local minimum is indicative of the complexity of the problem.

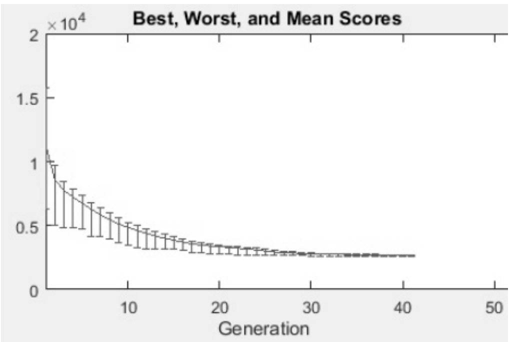

Figure 11: Case 1 GA optimisation population best, worst and mean fitness for each generation for run C301

Case 1 discussion

The performance of the GA in optimising a model to an ideal case is satisfactory. Despite a fitness of 0 being achievable the best fitness achieved was 2571. This is likely due to the complexity of the optimisation problem and could likely be overcome with a larger population or by adjusting the GA crossover and mutation probabilities.

There appears to be a trade off in optimised variables in order to achieve a satisfactory energy consumption. A high air change rate is compensated with higher wall thickness and casual gains to give a similar energy result as the base model. Multiple objective optimisation could remedy this by adding a dimension to the solution space (e.g. zone temperature or humidity), thus narrowing the field of possible solutions. Also, the upper and lower bounds were kept deliberately high to test the GA's capabilities. These could be narrowed to limit unrealistic variable values and energy trade-offs.

The most problematic element exposed by the optimisation is the fact that the heat energy from casual gains is the same in the model to the specific heat gains from the heating system. If the model is consuming too much energy the casual gains can be increased to offset this. Obviously, this is undesirable as the model can become unrealistic with very large casual gains. However, an advantage of using inverse modelling techniques to develop a physically driven model is that these unrealistic values are easily identified. They can be limited by tightening variable bounds.

Case 2: optimisation to metered data results

Table 5 shows the performance of selected runs for Case 2. Similar to Case 1, a good fitness does not translate to a good overall energy performance. Note also the very poor performance, in terms of both fitness and overall energy, of the base model when compared to the optimised models. Note also the relatively close spread of fitness compared to Case 1.

Table 5: select results for Case 2

RUN	FITNESS	%ENERGY FROM METERED
Base model	9982	-82.28%
C401	5759	-3.66%
C402	5889	-6.60%
C403	6050	-2.81%
C404	6194	-2.23%
C405	6437	9.72%

Figure 12 shows a daily profile where areas with good agreement can be identified, as well as points where simulation can still be improved.

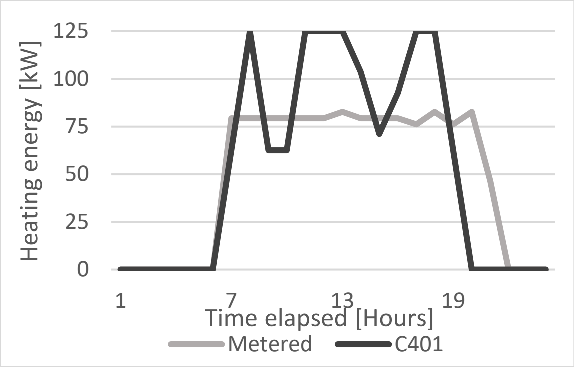

Figure 12: Metered data and C401 optimised model results for Saturday (top) and typical weekday (bottom)

Table 4 shows that the deviation of variables from the base model appear less severe in Case 2 than in Case 1, despite the fact that the optimised models have a much better fitness than the base model. In Case 2 the fitter model (C401) is closer to the base case than the less fit model (C405) in terms of the model input variables.

Table 6: select input variables for Case 2 after optimisation compared with base model

RUN	WALL [m]	AIR CHANGES	OCC. GAINS
Base model	0.30	3.0	11200
C401	0.26	3.2	20000
C405	0.25	5.5	38789

Case 2 discussion

The most significant result of Case 2 is the great improvement of both model fitness and energy consumption of the optimised models compared to the base model. This demonstrates that using the GA to generate a baseline model is more effective than using a best guess and benchmark values.

The rate of convergence is slower for Case 2 than for Case 1. A fitness value of 0 is impossible for Case 2 as the model resolution is not fine enough to match the data. Also, an averaged climate over 20 years was used for the model. If model complexity was increased, a matching climate file for Glasgow in 2013 was used and some GA parameters were adjusted (such as mutation and crossover as discussed above) then this could likely be improved upon. Multi-objective optimisation could also accelerate optimisation convergence. However, these adjustments were beyond the scope of this investigation.

The variables optimised are more realistic in Case 2 than they are in Case 1, although there does appear to be some variable trade off occurring as discussed in Case 1. Occupant gains are higher in Case 2 but can be due to a higher occupancy in the first week of February (in the middle of the Spring term) than anticipated in the base model.

The model has difficulty matching the data hour to hour, especially on the weekends. The key reasons for this are outlined below.

Metered data: the data supplied was for the whole building so was adjusted to represent a single floor within it. Sub metering would enable greater model and optimisation accuracy. Sensors taking readings for lighting levels, internal temperature or humidity would enable multi-objective optimisation.

Data adjustments: the data provided was scaled to represent a single floor of the building. This necessitated making fairly broad assumptions which are inherently sources of error. Accurate zonal sub-metering could aid in narrowing these error margins.

Model simplicity: the model used represents an entire floor of a large office-style building as a single zone. If the base model was more complex a closer fit to metered data would be more achievable.

Model climate: as mentioned above, a normalised model climate over 20 years was used in the model. For a better optimisation convergence, climate data matching the metered data year should be used.

Conclusion

This paper has demonstrated that GAs can be used to generate baseline energy models with better fitness and energy performance than a model built using benchmark assumptions. The GA consistently produced a model with overall energy consumption within 5% of the metered data.

The most pressing issue of the method is the observed trade-off between model variables to match the results set. This can be countered by using more realistic variable bounds and introducing multiple optimisation objectives.

An obstacle to the implementation of this method is the computation time required. Some Case 2 optimisations took up to 18 hours to complete. However, the processors used were old and advances in computing, along with the option of using cloud-based servers, will more than likely solve this issue.

By using GAs through MATLAB scripts, the method can be easily adapted and expanded to include very high numbers of variables and multiple objectives.

It would appear from the research displayed here that an inverse modelling approach to generate physically driven dynamic simulation models is possible.

In conclusion, the use of GAs to develop baseline energy models appears promising. Further work is required to ascertain whether it is more advantageous than existing inverse modelling methods such as Artificial Neural Networks or other statistical regression models. The major advantage would appear to be the greater model transparency offered by the resulting physically driven model, enabling more accurate predictions of the likely effects of energy conservation measures.

Applications

The authors see two principle applications of the proposed method, depending on the desired results of the optimisation:

Firstly, where high resolution metered data is available in conjunction with sensor data, multi-objective optimisation could be used to develop accurate baseline energy models with a high level of model transparency.

Secondly, where metered data alone is available along with limited building data, simplified energy models could be developed (such as the model developed here) quickly and cheaply for use in stock modelling.

Acknowledgements

A special thanks to Dean Drobot and David Charles at the University of Strathclyde Estates Department for all the help and information they provided, without which this work would not have been possible.

References

Adam, L. and Branda, M. (2016). Sparse optimization for inverse problems in atmospheric modelling. *Environmental Modelling & Software, 79*, 256-266.

AECOM (2009). Energy Performance Certificate for Graham Hills Building.

BEIS & OFGEM. (2013a). Smart meters: a guide [Online]. www.gov.uk: GOV.UK. Available: https://www.gov.uk/guidance/smart-meters-how-they-work [Accessed 16 Aug 2016].

BEIS & OFGEM. (2013b). Smart meters: information for industry and other stakeholders [Online]. www.gov.uk: GOV.UK. Available: https://www.gov.uk/guidance/smart-meters-information-for-industry-and-other-stakeholders [Accessed 16 Aug 2016].

Boardman, B. (2007). Home Truths: a Low-Carbon Strategy to Reduce UK Housing Emissions by 80% by 2050. Oxford, UK: University of Oxford.

Chipperfield, A., Fleming, P., Pohlheim, H. & Fonseca, C. (1994). Genetic Algorithm Toolbox For Use with MATLAB. User's Guide version 1.2. *Department of Automatic Control and Systems Engineering*, University of Sheffield.

CIBSE (2006). CIBSE Guide A: Environmental Design, London, The Chartered Institution of Building Services Engineers.

DCLG (2012). Recast of the Energy Performance of Buildings Directive. www.gov.uk.

DECC (2013). Smart Metering Implementation Programme: Smart Metering for Non-Domestic Customers. www.gov.uk: GOV.UK.

EC (2010). DIRECTIVE 2010/31/EU OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL: On the energy performance of buildings. *Official Journal of the European Union.*

ESRU (2016). ESP-r [computer program]. http://www.esru.strath.ac.uk/Programs/ESP-r.htm

Foucquier, A., Robert, S., Suard, F., Stéphan, L. & Jay, A. (2013). State of the art in building modelling and energy performances prediction: A review. *Renewable and Sustainable Energy Reviews, 23*, 272-288.

Google (n.d.) accessed in December 2016: https://developers.google.com/maps/documentation/javascript/

Harish, V. S. K. V. & Kumar, A. (2016). A review on modeling and simulation of building energy systems. *Renewable and Sustainable Energy Reviews, 56*, 1272-1292.

Hensen, J. L. (2011) Building performance simulation for sustainable building design and operation. *Proceedings of the 60th Anniversary Environmental Engineering Department, 2011* Czech Technical University, Prague, Czech Republic. 365.

Holland, J. H. (1992). Adaptation in natural and artificial systems [internet resource] an introductory analysis with applications to biology, control, and artificial intelligence, Cambridge, Mass., MIT Press.

Houck, C. R., Joines, J. & Kay, M. G. (1995). A genetic algorithm for function optimization: a Matlab implementation. NCSU-IE TR, 95.

IPCC (2014) Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]. IPCC, Geneva, Switzerland, 151 pp..

Kayo, G. & Ooka, R. Application of multi-objective genetic algorithm for optimal design method of distributed energy system. Eleventh International IBPSA Conference, 2009 Glasgow, Scotland. IBPSA.

Kelly, S., Crawford-Brown, D. & Pollitt, M. G. 2012. Building performance evaluation and certification in the UK: is SAP fit for purpose? Renewable and Sustainable Energy Reviews.

Le Bourgeois, O., Bouvier, C., Brunet, P. & Ayral, P.A. 2016. Inverse modelling of soil water content to estimate the hydraulic properties of a shallow soil and the associated weathered bedrock. Journal of Hydrology, 541, 116-126.

Mihail–Bogdan, C., Constantin, I. & Horia, N. 2016. The Influence of Genetic Algorithm Parameters Over the Efficiency of the Energy Consumption

Estimation in a Low–energy Building. Energy Procedia, 85, 99-108.

Ooka, R. & Komamura, K. Optimal design method for building & urban energy systems using genetic algorithms. Building Simulation 2007, 2007 Beijing, China. IBPSA.

Palonen, M., Hasan, A. & Siren, K. A Genetic Algorithm for Optimisation of Building Envelope and HVAC System Parameters. Eleventh International IBPSA Conference, 2009 Glasgow, Scotland. IBPSA.

Pernodet, F., Lahmidi, H. & Michel, P. Use of Genetic Algorithms for Multicriteria Optimisation of Building Refurbishment. Eleventh International IBPSA Conference, 2009 Glasgow, Scotland. IBPSA.

US DOE 2015. EnergyPlus [computer program]. http://apps1.eere.energy.gov/buildings/energyplus/

University of Strathclyde 2015a. Campus Key Performance Indicators.

University of Strathclyde 2015b. Graham Hills Building gas meter readings 2013-2015.

Wang, B. & Malkawi, A. Genetic algorithm based building form optimisation study for natural building ventilation. 14th Conference of International Building Performance Simulation Association, 2015 Hyderabad, India.

Zhang, Y., O'neill, Z., Dong, B. & Augenbroe, G. 2015. Comparisons of inverse modeling approaches for predicting building energy performance. Building and Environment, 86, 177-190.

Zhao, H.-X. & Magoulès, F. 2012. A review on the prediction of building energy consumption. Renewable and Sustainable Energy Reviews, 16, 3586-3592.