
Article:
Wate, P., Iglesias, M., Coors, V. et al. (1 more author) (2020) Framework for emulation and uncertainty quantification of a stochastic building performance simulator. Applied Energy, 258. 113759. ISSN 0306-2619

https://doi.org/10.1016/j.apenergy.2019.113759

Framework for emulation and uncertainty quantification of a stochastic building performance simulator

P.Wate[a,c], M.Iglesias[b], V.Coors[a], D.Robinson[d]

[a] *Centre for Geodesy and Applied Informatics, Stuttgart University of Applied Sciences, Germany*
[b] *School of Mathematical Sciences, The University of Nottingham, UK*
[c] *Department of Architecture and Built Environment,The University of Nottingham, UK*
[d] *Sheffield School of Architecture, University of Sheffield, UK*

Abstract

A good framework for the quantification and decomposition of uncertainties in dynamic building performance simulation should: (i) simulate the principle deterministic processes influencing heat flows and the stochastic perturbations to them, (ii) quantify and decompose the total uncertainty into its respective sources, and the interactions between them, and (iii) achieve this in a computationally efficient manner. In this paper we introduce a new framework which, for the first time, does just that. We present the detailed development of this framework for emulating the mean and the variance in the response of a stochastic building performance simulator (EnergyPlus co-simulated with a multi agent stochastic simulator called No-MASS), for heating and cooling load predictions. We demonstrate and evaluate the effectiveness of these emulators, applied to a monozone office building. With a range of 25-50 kWh/m^2, the epistemic uncertainty due to envelope parameters dominates over aleatory uncertainty relating to occupants' interactions, which ranges from 6-8 kWh/m^2, for heating loads. The converse is observed for cooling loads, which vary by just 3 kWh/m^2 for envelope parameters, compared with 8-22 kWh/m^2 for their aleatory counterparts. This is due to the larger stimuli provoking occupants' interactions. Sensitivity indices corroborate this result, with wall insulation thickness (0.97) and occupants' behaviours (0.83) having the highest impacts on heating and cooling load predictions respectively. This new emulator framework (including training and subsequent deployment) achieves a factor of c.30 reduction in the total computational budget, whilst overwhelmingly maintaining predictions within a 95% confidence interval, and successfully decomposing prediction uncertainties.

Keywords: Gaussian Process Emulator, building performance, stochasticity, uncertainty quantification and decomposition.

Email address: parag.wate@hft-stuttgart.de (P.Wate)

1. Introduction

Building performance simulation tools have evolved considerably in recent decades, with significant efforts having been invested to improve the scope and validity of their underlying algorithms and the usability of interfaces to these algorithms, to the extent that they have now entered into the mainstream. Indeed they are now commonly required to demonstrate compliance with national regulations, as required for example by the European Energy Performance of Buildings Directive [1]. Yet predictions from standard Building Performance Simulation tools - whether in relation to existing or proposed buildings - continue to deviate significantly from those that are observed. This is commonly referred to as the *performance gap*. This deviation is problematic for two main reasons. Firstly, it undermines confidence in the accuracy of energy use estimates, as required for example by energy performance certifications schemes. Secondly, it may negatively impact on the design decisions made, potentially leading to suboptimal design decisions, or design features that lack robustness.

The causes of the energy performance gap are many and complex [2, 3], including: (1) errors or omissions in the simulation of non-trivial phenomena [4] such as coupled heat and moisture flow and its dependence on changing thermo-physical properties or of occupants' stochastic interactions [5], (2) use of standardised external boundary conditions and building systems' control settings for thermostat and operation hours [6], (3) practical issues, such as the inability to fully predict a building's future functions during design, or to account for plausible future variations in function during the design process, (4) failure to meet specified insulation and airtightness standards due to poor workmanship, (5) unavailability of data regarding e.g. internal gain and plug loads that are representative of the building's post-occupancy operation [7].

Efforts to improve the robustness of BPS tools, for instance by developing 'reference simulations' as part of model calibration processes [8, 9], and through feedback from post-occupancy evaluation exercises [10], are ongoing. But a particularly promising avenue of current exploration is through the characterisation and propagation of aleatory (e.g. due to occupants' stochastic interactions) and epistemic (e.g. due to poorly or un- observed input parameters) uncertainties, so that these can be quantified and decomposed.

It is self evident that buildings' energy demands for heating and cooling are dependent upon the heat transfers across the building envelope: on conductive heat transfers across opaque and transparent surfaces; on radiative transfers through the transparent surfaces; on advective transfers through accidental (imperfections) or deliberate (windows) openings in the envelope. These latter pathways may also be influenced by occupants' interactions with windows and shading devices. Haldi and Robinson [11] have shown that these aleatory (behavioural) uncertainties in energy use can be equivalent, even greater, in magnitude to their epistemic (envelope) counterparts. They also argue [12] that as the performance of the envelope improves, thus better conserving energy, so the consequences of occupants' actions to regulate the envelope will be exaggerated; that it will be more and more important to accu-

rately model occupants' behaviours and the uncertainties arising from them, simultaneously with their epistemic counterparts. This paper introduces a new computationally efficient framework to do just that, accounting for uncertainties in the simulation of deterministic processes influencing heat flows in buildings, and for stochastic perturbations to them.

In this we also wish to accelerate the computation of uncertainties by substituting the expensive BP simulators, of both deterministic and stochastic processes, by inexpensive emulators / metamodels, trained on datasets from the simulators. The emulator that we have developed for this purpose is generic in nature; it can be trained using monitored real-world data to support future predictions for the case study in hand. Indeed, this would be a very useful future application (e.g. to support model predictive control applications).

1.1. Uncertainty quantification in the building simulation context

The relationship between building design / retrofit input parameters, $\mathbf{X} = (X_1, \ldots, X_p)$, (e.g. wall thickness, window transmittance) and the corresponding BPS outputs (e.g. annual heating and cooling demand), Y, can be expressed via $Y = f(\mathbf{X})$, where f is typically a deterministic function with values defined by running the simulator for a given choice of inputs $\mathbf{X}$. When the inputs are uncertain, only their probability density function, $p(\mathbf{X})$, may be available to us. Since the functional form of f is generally unknown, propagating uncertainty from BPS inputs through to the corresponding predictions/outputs is only feasible via Monte Carlo (MC) sampling as applied in [13] [14] for uncertainty and sensitivity analysis and demonstrated using an uncertainty analysis workbench [15]. More specifically, samples from $p(\mathbf{X})$ are mapped onto the corresponding outputs and then used to construct an empirical distribution that approximates the distribution of model predictions $p(Y)$. Even when optimal MC sampling techniques such as Latin Hypercube Sampling (LHS) [16] and quasi-random sequences [17] are used, accurate approximations of $p(Y)$ via sampling often requires thousands of samples (of the order of 10^3 to 10^5 simulation runs) depending upon model complexity and the number of input parameters under consideration. Furthermore, if we wish to decompose prediction uncertainty (i.e. to know which amongst k inputs are dominant) via a variance decomposition technique, the total computation cost is $N(2k + 2)$ simulations where N is the number of samples [18, 19]. For the simulation of complex phenomena, as is the case in building simulation, each individual run entails a non-trivial computational cost. Moreover, the number of uncertain inputs can be relatively large (e.g. 10^2). Comprehensive uncertainty quantification studies thus tend to be highly computationally intensive, resource and time consuming; at the limits, simply intractable.

To address the computational challenge of propagating uncertainties in building performance simulation, numerous statistical metamodeling techniques have been proposed in recent years for sensitivity analysis of building [20] and climatic [21] parameters, and to support more accurate prediction of building energy performance [22] [23]. The aim of metamodeling is to use outputs from simulations at carefully designed inputs to train a statistical

List of symbols used

$\mathbf{X} = (X_1, \ldots, X_p)$	vector of input random variables
Y	model output random variable
f	functional form
N	number of input samples
ω	uncontrollable variable representing aleatory uncertainty
F	function random variable
$\mathbf{X}_n$	realisation vector of input variables at n^{th} sample
Y_n	output response value for the n^{th} sample of inputs
ϵ	a gaussian noise
$m(\mathbf{X})$	mean function
$k(\mathbf{X}, \mathbf{X}')$	covariance function
l	characteristic length scale
σ_Y^2	signal variance
Q	predictivity coefficient or coefficient of determination
s_n^2	sample variance
J_n	number of realisations of S-BPS
$\mathcal{GP}$	Gaussian process
$f_d(\mathbf{X})$	mean response
$f_s(\mathbf{X}, \omega)$	stochastic response
$S_I(Y)$	individual effect sensitivity index value
$S_{T_i}(Y)$	total effect sensitivity index value of controllable variable
$S_{T_\omega}(Y)$	total effect sensitivity index value of uncontrollable variable
t	thickness of material (m)
λ	thermal conductivity of material (W/(mK)
C	specific heat capacity of material (J/(kgK)
ρ	density of material (kg/m^3)
U	thermal heat transfer coefficient (W/(m^{2}K))
Y_h	annual heating demand (kWh/m^2a)
Y_c	annual cooling demand (kWh/m^2a)

model that provides an accurate approximation of the simulator response at a lower computation cost than that of the simulator. Once a metamodel has been constructed, the BPS tool can be substituted by the emulator and employed to conduct Monte Carlo uncertainty propagation and variance-based sensitivity analysis in providing early stage building design guidance [24] [25] and to estimate the performance of passively designed buildings [26]. We note that metamodeling is also highly relevant for computationally intensive tasks relevant to BPS such as Bayesian calibration [27, 28, 29] and optimisation [30, 31]. However, existing approaches are restricted to the emulation and uncertainty quantification of deterministic phenomena, as is the standard case in building simulation, neglecting sources of aleatory uncertainty, for example that arise from stochastic phenomena such as occupants' presence and their interactions with the building envelope, and associated heat and mass transfers. Failing to incorporate these stochastic phenomena, we argue, undermines the utility of uncertainty quantification exercises, and efforts to reduce the performance gap.

However, coupling a framework for simulating occupants' stochastic behaviours with BPS tools can considerably increase the computational complexity in the simulation of buildings' energy performance; particularly if repeated simulations are required to adequately quantify these stochastic influences. For a fixed input $\mathbf{X}$, repeated runs of such a *stochastic* simulator will produce different outputs - yielding an ensemble of prediction values instead of a single fixed value, as with a conventional *deterministic* BPS *simulator*. These prediction ensembles now represent a probability distribution of the simulator output, associated with each input $\mathbf{X}$. A standard approach to describe the stochastic response (for each fixed input $\mathbf{X}$) is via

$$Y = f(\mathbf{X}, \omega) \tag{1}$$

where ω is a variable that we introduced to denote the intrinsic aleatory uncertainty in the simulator. This new variable can be thought as an uncontrollable parameter (or seed variable) encoded within the stochastic simulator. The distribution of this uncontrollable variable is often unknown and, consequently, standard metamodeling frameworks, based on sampling the (known) input space can no longer be applied in a straightforward fashion to emulate/infer the functional form of f.

The function f defines here a system of coupled differential algebraic equations concerning interactions between thermodynamic phenomena (such as heat conduction, convection and radiation) occurring inside the buildings [32]. A numerical solver or computer code such as the thermal building simulation program EnergyPlus employs numerical approximations to the functional form f to solve this system of equations [33]. However, as Wetter and Polak [34] point out, the numerical approximations to f result in discontinuity, nonlinearity and multimodality in the design parameters. As mentioned earlier, in order to reduce the computation cost of UQ studies, a metamodeling approach is highly advantageous. The chosen approach should then reliably approximate the underlying functional form as a precondition. Moreover, the approach should also consider stochastic perturbations due to occupants' behaviours

while approximating the functional response. Therefore, there is a clear need for a new metamodeling approach that: 1) estimates the functional form from training datasets whilst specifying the uncertainty in estimating it, and 2) accounts for stochastic perturbations, such as those arising from occupants' behaviours, impacting on the function response.

In this work we propose a new metamodeling approach to simulate the stochastic response of a BPS that simulates occupants' behaviour. In this, we follow the generic stochastic Kriging metamodeling approach of Ankenman et al. [35] which uses repetitions of the stochastic simulator, at a given design point $\mathbf{X}$, to characterise/approximate the input-depdendent mean response and variance of the stochastic building performance simulator, via Gaussian process regression (GPR). The proposed approach aims to provide GPR emulators to the following quantities

$$f_d(\mathbf{X}) = \mathbb{E}_\omega(Y|\mathbf{X}), \qquad V(\mathbf{X}) = \mathbb{E}_\omega(Y|\mathbf{X} - f_d(\mathbf{X}))^2, \tag{2}$$

where $\mathbb{E}_\omega$ denotes the expectation with respect to ω. By means of off-the-shelff GPR algorithms [36], we fit a GP to the variance $V(\mathbf{X})$ given the sample variance from repeated simulator runs computed at design/training points. We subsequently use predictions from the corresponding metamodel of $V(\mathbf{X})$ to fit a non-heteroscedastic GP to f_d.

We apply the proposed metamodeling approach to simulate the stochastic response from a hypothetical office building model. In this we show, for our particular choice of BPS settings in conjunction with a straightforward model of a monozone office building, that our GPR of $V(\mathbf{X})$ and $f_d(\mathbf{X})$ is both computationally efficient and provides an accurate emulation of the simulator response, in the sense that the stochastic predictions of the simulator fall within confidence intervals provided by the proposed framework. In this, we will also show that the pair of predicted GP emulator response surfaces plausibly capture the underlying behaviour of the detailed BPS program whilst representing the uncertainty bounds due to the emulators themselves; emulators that successfully propagate both the input parameter uncertainty and the stochastic uncertainty due to occupants' interactions, by reproducing the prediction probability distribution at test points, thereby quantifying the total uncertainty in energy predictions. Finally, we combine the proposed metamodeling approach for $V(\mathbf{X})$ and $f_d(\mathbf{X})$ with a variance-based decomposition of total uncertainty in predictions into its respective sources, computed in terms of sensitivity indices.

1.2. Paper Structure

In section 2 we introduce our approach to stochastic building performance simulation, through the co-simulation of deterministic and stochastic phenomena. In section 3, we discuss alternative candidate approaches that have been or could be applied to the emulation of building performance simulators, with a view to identifying the most promising candidate and how this can be extended for our purposes. In section 4, we present in detail our implementation of the emulation strategy introduced in section 1, going on to present our

related uncertainty quantification strategy in section 5. The combined methodology is then evaluated, using a case study which we describe in section 6, through a series of numerical experiments; the results of which we discuss in section 7. The paper closes with a brief summary and a discussion of how this new framework is being further developed and applied, for the more comprehensive quantification of uncertainties in building and urban performance simulation.

2. Stochastic Building Performance Simulation (S-BPS)

Significant progress has been made in recent years in the modelling of occupants' behaviours, employing a range of techniques. A large number of data-driven stochastic models have been developed to model occupants' presence (or activities and dependent presence in the case of residences) and dependent interactions with the building envelope and its systems. The underlying phenomena are typically represented as a Bernoulli process [predicting the probability that a particular state will be observed, such as a window being open], as a discrete-time random or Markov process [predicting the probability with which a state transition will take place], as a continuous-time random or survival time process [predicting the duration for which a particular state will survive], or hybrids of them [37]. Other techniques are needed to model phenomena for which data is not abundant. These include belief-desire-intention (BDI) and agent-learning frameworks, both applied to agent-based representations of occupants [38]. In the former case, agents evaluate their beliefs about the state of the environment in which they are immersed, comparing these with their desired states and effect plans intended to achieve their desires. In the latter case, agents progressively learn, through repeated simulations, behavioural interactions that maximise rewards. BDI rules are suited to relatively simple interactions, whereas agent learning is able to accommodate considerably more complex behaviours, though this can come with a penalty in terms of the number of computations needed to learn reward-optimal behaviours. Finally, and as noted by Robinson et al. [37], it is also desirable to represent social interactions amongst occupants' sharing spaces, emulating their negotiation processes and the outcomes (agreed interactions) arising from them. Chapman et al. [38] describe a vote casting and processing mechanism to achieve this in the case of data-driven models. Similar mechanisms are encodable in the case of BDI rules.

Several computational platforms have been developed to couple stochastic models of occupants' behaviours within BPS tools, including those of Langevin et al. [39], Hong et al. [40] and Chapman et al. [5]. Langevin et al's platform integrates rules (akin to BDI rules) that allow agents to modify their clothing and the use of windows, fans and heaters (and corresponding setpoints) to achieve their comfort desires. This is integrated with the EnergyPlus solver using the Building Controls Virtual Test Bed (BCVTB). Hong et al's platform is in principle more general in character, employing an XML schema to structure the modelling approaches to be employed for a particular application, populated with the parameters to be

employed by this structure, which is then solved using obFMU. The Functional Mockup Interface (FMI) co-simulation standard is employed to couple obFMU with EnergyPlus, with EnergyPlus acting as the master algorithm. As far as the authors are aware, only Chapman et al's platform (No-MASS) incorporates all of the above mechanisms; in common with obFMU, utilising FMI to co-simulate with EnergyPlus.

As noted in the introduction, the intention of this paper is to demonstrate a proof-of-concept - that we can effectively emulate a building performance simulator incorporating aleatory uncertainties arising from occupants' stochastic interactions, and to employ this simulator to quantify and decompose uncertainties. For this purpose we use EnergyPlus co-simulated with No-MASS. For this proof-of-concept study, we restrict our scope to uncertainties relating to the physical properties of the envelope of a simple monozone office and the stochastic behaviours of its occupants, limited to their presence and associated use of windows and shading devices. For presence, we employ the time-inhomogeneous Markov chain model of Page et al. [41]; for the use of windows we employ the hybrid discrete- and continuous- time model of Haldi and Robinson [42]; while for shading devices we employ a similar hybridisation, due to Haldi and Robinson [43].

For further information regarding behavioural models and associated modelling techniques, we refer the interested reader to the reviews of Dong *et. al.* [44], Wagner *et. al.* [45], Gunay *et. al.* [46] and Robinson *et. al.* [37].

3. Gaussian process metamodeling for deterministic BPS

Numerous metamodeling techniques have been employed to overcome the computational costs of running thousands of expensive BP simulations for building design space exploration [47, 25], optimization, [30, 31], model calibration [29, 48] and uncertainty quantification (UQ) [49, 26]. Techniques that have been used within (deterministic) BPS workflows include ordinary least squares regression (OLS) [50], multivariate adaptive regression splines (MARS) [51], random forest (RF) [52], Gaussian process regression (GPR) [53], support vector regression (SVR) [54] and artificial neural networks (ANN) [50]. For a thorough practical evaluation of these techniques, we refer the interested reader to : the works of Cheng and Cao [22] for the application of MARS, Yidil *et. al.* [55] and Amasyali and El-Gohary [56] for reviews of regression models, Østergård *et. al.* [57] and Wei *et. al.* [58] for comparative studies, and Lim and Zhai [59] for an analysis of the impact of metamodels on bayesian calibration.

While the literature above reveals that there exists no single emulator-based approach that is suitable for all purposes, numerous evaluations (see for example [59] and [57]), studying the effects of training set size, the nature of the simulator functions to be emulated and the number of inputs points to be represented, have shown that GPR is consistently able to emulate the underlying simulator response more accurately than competing techniques. Østergård et al. [57] compare the six mentioned metamodeling techniques applied to building performance simulations, judging these against performance indicators such as accuracy,

efficiency, robustness, simplicity and transparency. They conclude that GPR metamodels are the most accurate (with a coefficient of determination $R^2 > 0.99$) in approximating the nonlinear nature of BPS problems. Applications of GPR in the context of BPS include – (1) studies to optimise the design of glazing systems for office buildings [60]; real time optimization of cooling energy systems for efficient energy management in high-rise office buildings [61]; building design optimization [62] [63], (2) calibration and diagnosis of building energy simulation models for the formulation of energy conservation and model predictive control strategies [64], and (3) sensitivity analysis of energy performance predictions to weather and building envelope characteristics [21].

To the best of our knowledge, metamodeling approaches for BPS have only been applied where the simulator response is deterministic (for a particular choice of inputs). In this paper we address the case when BPS is coupled with a stochastic model for occupants' behaviour. As described in the introduction, standard metamodeling techniques (including GPR) cannot be directly applied to emulate the response of a stochastic BPS. The main challenge arises from the fact that the variance of the simulator (with respect to an uncontrollable variable), $V(\mathbf{X})$ is a function of input parameter $\mathbf{X}$ (see equation(2)) which is, in general, unknown. From the statistical literature we identify two possible routes to address this challenge. With the approach proposed in [65, 66] the unknown values of the variance $V(\mathbf{X})$, at the training points, can be treated as hidden variables within the GPR framework with non-heterosedastic variance. Alternatively, when repetitions of the computer code are available for each input, the stochastic Kriging (SK) method can be applied to emulate the variance (given the sample variance from repetitions) [67, 35, 68]. Once $V(\mathbf{X})$ has been approximated via an emulator, GPR can be applied to emulate the mean response surface using information from the GP emulator for the variance. There is good evidence to suggest that SK lends itself more naturally to variance based sensitivity analysis which, in turn, requires an explicit emulation of both the mean and variance [69]. An SK approach is proposed in Section 4 for the emulation for the stochastic response of BPS, for which we demonstrate its efficiency for the computation of sensitivity indices in Section 5. Since standard GPR is at the core of the SK method, we introduce in the following section the basic elements of standard GPR (for deterministic computer simuators) and refer the interested reader to [53] for further details.

3.1. Standard Gaussian Process (GP) Metamodeling for deterministic simulators

Let us assume that we have a deterministic computer simulator that depends on a vector of input parameters denoted by $\mathbf{X} = (X_1, \ldots, X_p)$, that we assume belong to a space of admissible inputs $\mathcal{X} \subset \mathbb{R}^p$. We further assume that the outputs of the simulator can be described via the evaluation of a deterministic function $F : \mathcal{X} \to \mathbb{R}$, which maps inputs $\mathbf{X} \in \mathcal{X}$ into the corresponding simulator response (output) $F(\mathbf{X})$. For simplicity, in what follows we consider the case of a univariate simulator response, but the results will be extended to the multivariate case.

As discussed earlier, when running the simulator (i.e. evaluating $F(\mathbf{X})$ for a given input $\mathbf{X}$) is a computationally costly exercise, propagating inputs uncertainty through the simulator can become unfeasible. The aim of metamodeling is to replace F with a function that emulates the simulator response at much lower computational cost. GPR provides for this, by inferring the underlying functional form of F, given observations of the evaluation of F at a carefully selected number of *design* or *training points* $\{\mathbf{X}_n\}_{n=1}^N$ in the input space $\mathcal{X}$. More specifically, we consider the training set defined by $\mathcal{D} = \{(\mathbf{X}_n, Y_n)\}_{n=1}^N$ where Y_n are obsevations of $F(\mathbf{X}_n)$. Following standard assumptions we consider an observational model of the form [53]

$$Y_n = F(\mathbf{X}_n) + \epsilon, \tag{3}$$

where $\epsilon \in \mathbb{R}^d$ is Gaussian noise with zero mean and covariance $\sigma^2 I$.

Before observing the data $\mathcal{D}$, the underlying Bayesian framework for GPR metamodeling requires the specification of a prior distribution (over a space of functions) for the function F that we wish to infer. In the standard framework, the prior on F is a Gaussian distribution of functions with mean (function) $m(\mathbf{X})$ and covariance $k(\mathbf{X}, \mathbf{X}')$ (we denote this by $F \sim \mathcal{GP}(m(\mathbf{X}), k(\mathbf{X}, \mathbf{X}'))$). For the sake of clarity, we consider here the case $m(\mathbf{X}) = 0$ for all $\mathbf{X} \in \mathcal{X}$ which corresponds to the case in which we have no prior knowledge of the simulator.

The covariance function $k(\mathbf{X}, \mathbf{X}')$, on other hand, characterizes the variability of the family of functions defined by the GP prior. A covariance that enables a wide class of functions to be characterised is the Mattern covariance, defined as [53] :

$$k(\mathbf{X}, \mathbf{X}') = \sigma_Y^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{||\mathbf{X} - \mathbf{X}'||}{l} \right)^{\nu} K_\nu \left(\frac{||\mathbf{X} - \mathbf{X}'||}{l} \right), \tag{4}$$

where Γ is the gamma function, l is the characteristic length scale, σ_Y^2 is the signal variance, K_ν is the modified Bessel function of the second kind of order ν, and $||\cdot||$ is the Euclidean norm (in $\mathbb{R}^p$). The parameter ν controls the regularity/smoothness of the samples. For the present work we consider ν to be fixed and so the Gaussian is fully characterised via the specification of the prior hyperparameters l, σ_Y that are contained in the single vector $\theta = (l, \sigma_Y^2)$.

In order to make predictions of F at a test location $\mathbf{X}^*$ (i.e. $F(\mathbf{X}^*)$), we consider the predictive distribution, $p(F(\mathbf{X}^*)|\mathbf{X}^*, \mathcal{D}, \theta)$, which from the standard GPR framework, is Gaussian with mean $\hat{F}(\mathbf{X}^*)$ and variance $\hat{K}(\mathbf{X}^*)$ defined by:

$$\hat{F}(\mathbf{X}^*) = K_\theta^T(\mathbf{X}^*)(K_\theta + \sigma^2 I)^{-1}\mathbf{Y} \tag{5}$$

$$\hat{K}(\mathbf{X}^*) = k_\theta(\mathbf{X}^*, \mathbf{X}^*) - K_\theta^T(\mathbf{X}^*)(K_\theta + \sigma^2 I)^{-1}K_\theta(\mathbf{X}^*), \tag{6}$$

where K_θ satisfies: $[K_\theta]_{i,j} = k_\theta(\mathbf{X}_i, \mathbf{X}_j)$, $K_\theta(\mathbf{X}^*) = (k_\theta(\mathbf{X}^*, \mathbf{X}_1), \ldots, k_\theta(\mathbf{X}^*, \mathbf{X}_N))^T$, and $\mathbf{Y} = (Y_1, \ldots, Y_N)$. We note that if θ and σ are known, expression (5) can be used to make

predictions of the simulator F, at testing point $\mathbf{X}^*$. A measure of the uncertainty in those predictions is provided by the variance (6). In practice, however, the (hyper)parameters θ and σ are unknown and hence these must be inferred from the training set $\mathcal{D}$. The standard practice for estimating these parameters is to maximise the posterior density $p(\theta, \sigma^2|\mathcal{D})$, which from Bayes' rule, satisfies the condition:

$$p(\theta, \sigma^2|\mathcal{D}) \propto p(\mathcal{D}|\theta, \sigma^2)p(\theta, \sigma^2) \tag{7}$$

where $p(\theta, \sigma^2)$ is the (hyper-)prior distribution on (θ, σ^2) and $p(\mathcal{D}|\theta, \sigma^2)$ is the marginal likelihood given by the following expression [53]:

$$p(\mathcal{D}|\theta, \sigma^2) = \frac{1}{2}\mathbf{Y}^T(K_\theta + \sigma^2\mathbf{I})^{-1}\mathbf{Y} + \frac{1}{2}\log|K_\theta + \sigma^2\mathbf{I}| + \frac{N}{2}\log 2\pi \tag{8}$$

The maximum a posteriori (MAP) estimate of the prior hyperparameters is:

$$(\hat{\theta}, \hat{\sigma}^2) = \arg\max_{\theta,\sigma^2} p(\theta, \sigma^2|\mathcal{D}) = \arg\min_{\theta,\sigma^2} \Big[-\log p(\mathcal{D}|\theta, \sigma^2) - \log p(\theta, \sigma^2) \Big] \tag{9}$$

Finally, predictions of $F(\mathbf{X}^*)$ together with a measure of their uncertainty can be used by employing $(\hat{\theta}, \hat{\sigma}^2)$ in expressions (5)-(6).

As stated earlier, expressions (5) -(6) are derived under the assumption of zero prior mean $m(\mathbf{X}) = 0$. It is worth mentioning, however, that it is often beneficial to incorporate a non-zero mean function. A common choice is to consider $m(\mathbf{X})$ a polynomial of some degree, with coefficients that can be inferred by including them in the setting of (7). We refer the reader to [53] for further details and the corresponding equations that generalise those in (5) -(6) for the case of non-zero prior mean.

Finally, we emphasize that the Gaussian assumptions in the GPR framework relate to the family of functions that approximate the functional form of the simulator (i.e. F). The GPR framework impose neither restrictions on the distribution of the input space $\mathcal{X}$ nor on the distribution of the simulated outputs $F(\mathcal{X})$. The GPR framework can be used to propagate uncertainties for an arbitrary distribution of inputs which may, in turn, yield a distribution of non-Gaussian outputs.

3.2. Selection of design points

Sampling strategies for the careful selection of design/training points are broadly classified into stratified sampling and quasi-random sequences. In contrast to pseudo-random sampling [17], these more sophisticated sampling techniques are designed to achieve uniformity in the design points' distribution in the input space, providing sufficient and well dispersed data/observations from which to construct a highly representative metamodel response surface.

Latin Hypercube Sampling (LHS) is a popular stratified sampling strategy, where each input X_i is divided into say m sub-intervals, having equal marginal probability $1/m$ [16]. A

sample is selected from each consecutive sub-interval to form a sequence of m points. These sequences of m points from each input are randomly matched to form a k-dimensional set of m sample points [70]. In contrast, quasi-random sequences follow space filling criteria of discrepancy and geometric deviation to generate a uniformly dispersed sequence of sample points [71].

In the context of this work and for future extentions of its scope, we adopt the standard practice of using LHS for GPR, which has an estimated error convergence rate that is independent of the dimensionality of the input space [72] and is almost as fast as quasi-random sequences for large samples (N) [70] [73].

3.3. Validation measures

Numerous validation measures can be employed to determine the goodness of fit of an emulator [31]. A common choice is a unit-less measure called the predictivity coefficient (or coefficient of determination) Q defined by:

$$Q(Y, \hat{Y}) = 1 - \frac{\sum_{i=1}^{N_t}(Y_i - \hat{F}(X_i^*))^2}{\sum_{i=1}^{N_t}(\bar{Y} - Y_i)^2} \tag{10}$$

where $Y = \{Y_i\}_{i=1}^{N_t}$ denotes the observed simulator outputs over a set of N_t testing points $\{X_i^*\}_{i=1}^{N_t}$, $\bar{Y}$ denotes their empirical mean, and $\hat{Y} = \{\hat{F}(X_i^*)\}_{i=1}^{N_t}$ are the GP metamodel predictions computed by (5).

Van Gelder et al [31] propose that a metamodel be constructed with as few training samples as possible, through a new strategy that is independent of the type of metamodel, by systematically adding samples in steps until the constructed metamodel meets defined validation criteria. If the evaluation measures indicate a poor fit (low Q and high error) then the number of sampling points, within a fixed computation budget, are increased and the model fit is re-analysed; the process being repeated until the goodness of fit criteria (high Q and low error) are satisfactory i.e. addition of more sampling points to the training datasets does not lead to any significant increase or change (say by more than 5%) in the predictivity coefficient. Since the desired accuracy is problem-dependent, the goodness of fit of the metamodel is evaluated based on the converged value of our chosen indicator of accuracy i.e. predictivity coefficient. We apply and demonstrate this new metamodel-type independent strategy in Section 7.1, to obtain mean and stochastic variance GP response surfaces by analysing their goodness of fit at test data points, and by observing the convergence in predictivity coefficient values over sampling sets of increasing size.

In addition, since the GP framework produces a statistical distribution of predicted outputs, it is important to quantify the uncertainty in the GP emulator response at validation/testing data points. To this end, we make use of a credible interval technique, as proposed in Bastos and O'Hagan [74]. At any given test point $\mathbf{X}^*$, the GP framework that

yields expressions (5) and (6) enables us to compute credible intervals defined by

$$C_\alpha(\mathbf{X}^*) = [\hat{F}(\mathbf{X}^*) - \tau\sqrt{\hat{K}(\mathbf{X}^*)}, \hat{F}(\mathbf{X}^*) + \tau\sqrt{\hat{K}(\mathbf{X}^*)}]) \tag{11}$$

where τ_α is the $(1-\frac{\alpha}{2})$ quantile of the standard Normal distribution and $\hat{F}(\mathbf{X}^*)$ and $\hat{K}(\mathbf{X}^*)$ are the values that we obtain from expressions (5) and (6), respectively. According to Bastos and O'Hagan [74], evaluating the statistical accuracy of the emulator response can be performed by computing the percentage of observations (obtained from a set of testing points) that fall within this interval. An accurate emulator response should yield credible intervals that contain a proportion of approximately $(1 - \frac{\alpha}{2})$ of those observations. In Section 7.1, we evaluate the uncertainty of the emulators trained in this study.

4. Proposed methodology for the emulation of S-BPS

In this section we propose a methodology to emulate the stochastic response of S-BPS described by (Eq. 2). Again we reiterate that while $\mathbf{X}$ is a vector of stochastic input parameters (e.g. wall thickness, window transmittance), these are assumed to have a well-characterised probability density function. In contrast, ω is a variable associated with the aleatoric nature of the S-BPS associated with occupants presence and their interactions with the envelope (e.g. use of windows and shading devices). We assume that we only have access to the distribution of ω via repetitions of the simulator at a training point. In this case, our aim is to construct metamodels for the mean response $f_d(\mathbf{X})$ and the variance, $V(\mathbf{X})$, of the stochastic component of the simulator. More precisely, we consider the log-transformed [1] variance $\log V(\mathbf{X})$ to ensure positivity of the estimates produced by the metamodel.

Figure 1 depicts the exchange of simulation variables between EnergyPlus and No-MASS, constituting our Stochastic Building Performance Simulator (S-BPS), while figure 2 illustrates the deployment of the S-BPS within our Monte Carlo uncertainty propagation scheme for the generation of training datasets with which to construct metamodels.

4.1. Gaussian process fitted to the log-variance $\log V(\mathbf{X})$.

We consider a set of training points $\{\mathbf{X}_n\}_{n=1}^N \in \mathcal{X}$ in the input space and denote by $Y_n = f(\mathbf{X}_n, \omega)$ [2] the random variable that arises from evaluating the stochastic simulator at each training point $\mathbf{X}_n$. For each $\mathbf{X}_n$ we generate a set of J_n realisations of the stochastic simulator. We denote these realisations by $\{Y_n^{(j)}\}_{j=1}^{J_n}$, and we assume that these are independent and identically distributed (iid) according to the law of Y_n. We compute the sample mean, $\overline{Y}_n$, and sample variance, s_n^2, of these realisations via the following expressions:

[1] Since variance is always positive, the log transformation of the sample variance avoids prediction of negative variance values.

[2] For the purposes of simplicity in notation, in this section we also use Y_n to denote the simulator response, so that Y_n is now a random variable since the function f is stochastic .

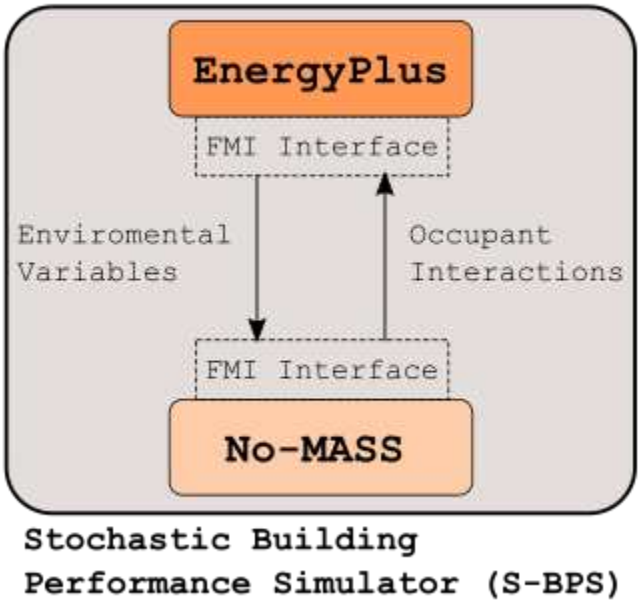


Figure 1: Schematic of EnergyPlus and No-MASS co-simulation platform [75]: the Stochastic Building Performance Simulator (S-BPS)

$$\overline{Y}_n = \frac{1}{J_n}\sum_{j=1}^{J_n} Y_n^{(j)} \quad \text{and} \quad s_n^2 = \frac{1}{(J_n-1)}\sum_{j=1}^{J_n}(Y_n^{(j)} - \overline{Y}_n)^2. \tag{12}$$

Following the work of [68, 67], we use the log of the sample variance $\{\log(s_n^2)\}_{n=1}^N$ computed at the training points to construct a GP metamodel for the unknown log-variance $\log(V(\mathbf{X}))$. In order to introduce an observational model of the form (3) suitable for GPR, we note that under Gaussian assumptions on the distribution of the observations $\{Y_n^{(j)}\}_{j=1}^{J_n}$, the log-transform introduces a bias β_n. More specifically, according to [68, 67], we know that:

$$\log V(\mathbf{X}_n) = \mathbb{E}(\log(s_n^2)) + \beta_n \tag{13}$$

where

$$\beta_n = -\Psi\left(\frac{J_n-1}{2}\right) + \log\left(\frac{J_n-1}{2}\right).$$

In the previous expression Ψ denotes the digamma function. Motivated by (13) we thus consider the following observational model for the true log-variance at the training points:

$$\log(s_n^2) + \beta_n = \log V(\mathbf{X}_n) + \eta_n \tag{14}$$

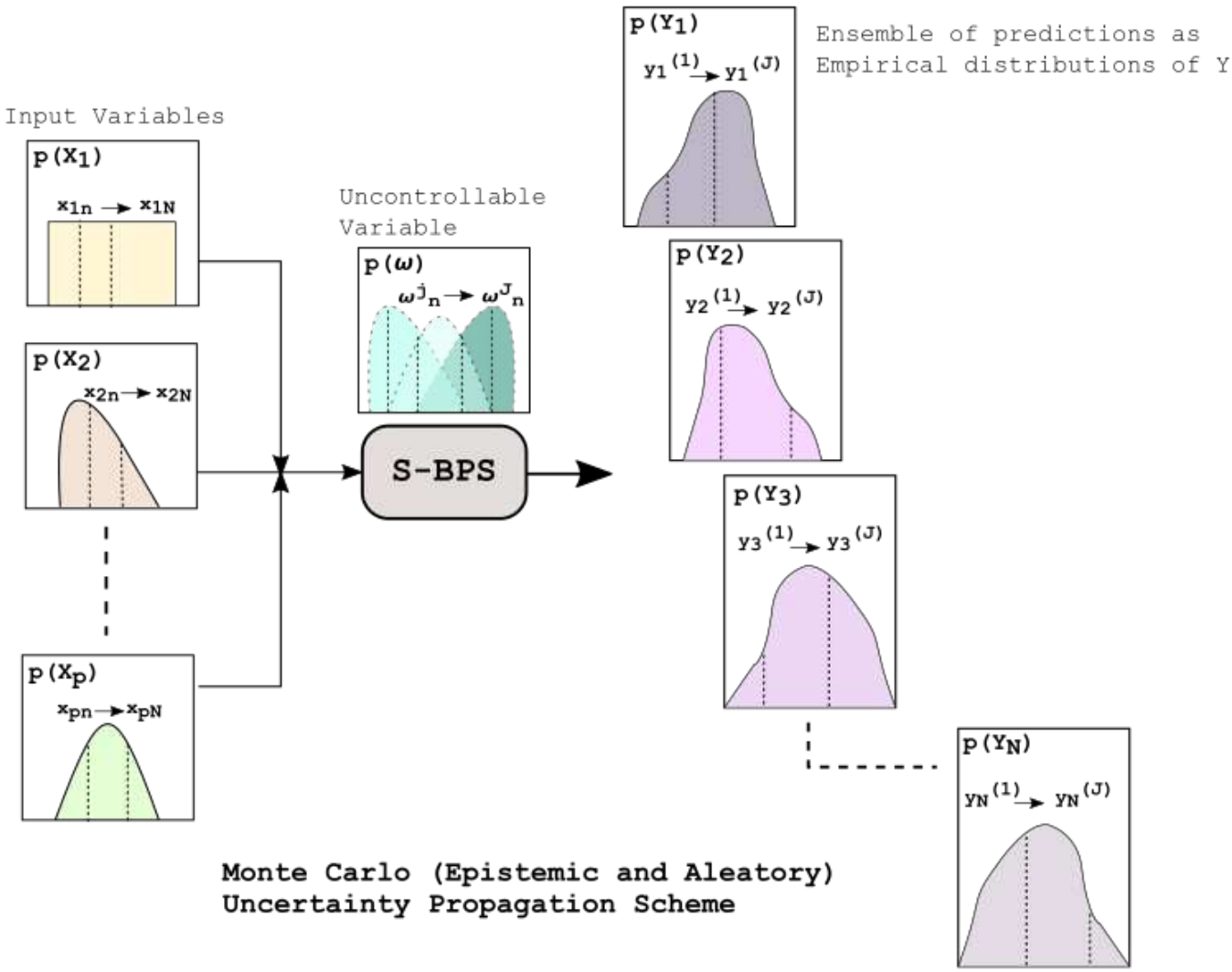


Figure 2: Monte Carlo Uncertainty Propagation Scheme for S-BPS

where $\eta_n \sim \mathcal{N}(0, \sigma^2)$. The training set for the metamodel $\log V(\mathbf{X})$ is thus defined by $\mathcal{D}_v = \{(\mathbf{X}_n, \log(s_n^2) + \beta_n)\}_{n=1}^N$. We use the GPR described in subsection 3.1 to fit a stationary Gaussian process $\log V(\mathbf{X})$, i.e. we assume a centred (i.e. zero mean) GP prior on $\log V(\mathbf{X})$:

$$\log V(\mathbf{X}) \sim \mathcal{GP}(0, k_{\theta_v}(\mathbf{X}, \mathbf{X}')) \tag{15}$$

with a covariance function defined by (4) and hyperparameters θ_v. From the GP framework we obtain mean predictions for $\widehat{\log V}(\mathbf{X}^*)$ and variance $\hat{K}_v(\mathbf{X}^*)$ defined via the following expressions:

$$\widehat{\log V}(\mathbf{X}^*) = K_{\theta_v}^T(\mathbf{X}^*)(K_{\theta_v} + \sigma^2\mathbf{I})^{-1}\boldsymbol{\xi} \tag{16}$$

$$\hat{K}_v(\mathbf{X}^*) = k_{\theta_v}(\mathbf{X}^*, \mathbf{X}^*) - K_{\theta_v}^T(\mathbf{X}^*)(K_{\theta_v} + \sigma^2\mathbf{I})^{-1}K_{\theta_v}(\mathbf{X}^*) \tag{17}$$

where K_{θ_v}, $K_{\theta_v}(\mathbf{X}^*)$ are defined analogous to those in (5)-(6), and where $\boldsymbol{\xi} = (\log(s_1^2) + \beta_1, \ldots, \log(s_N^2) + \beta_N)$. Hyperparameters θ_v and σ^2 are estimated by maximising the posterior density as described earlier. In this case the marginal likelihood is given by:

$$p(\mathcal{D}_v|\theta_v, \sigma^2) = \frac{1}{2}\boldsymbol{\xi}^T(K_{\theta_v} + \sigma^2\mathbf{I})^{-1}\boldsymbol{\xi} + \frac{1}{2}\log|K_{\theta_v} + \sigma^2\mathbf{I}| + \frac{N}{2}\log 2\pi \tag{18}$$

4.2. Gaussian process fitted to the mean response $f_d(\mathbf{X})$.

Let us note that if the realisations $\{Y_n^{(j)}\}_{j=1}^{J_n}$ are iid samples from a Gaussian, then from the Central Limit Theorem [76] we know that $\overline{Y}_n$ follows a Gaussian distribution with mean $f_d(\mathbf{X}_n)$ and variance $V(\mathbf{X}_n)/J_n$; we write this as follows:

$$\overline{Y}_n \sim \mathcal{N}(f_d(\mathbf{X}_n), V(\mathbf{X}_n)/J_n), \qquad n = 1, \ldots N \tag{19}$$

If the assumption of normality on $\{Y_n^{(j)}\}_{j=1}^{J_n}$ is not satisfied, expression (19) provides a good approximation for sufficiently large sample size J_n. Therefore, from (19) we consider the following observational model for $f_d(\mathbf{X})$:

$$\overline{Y}_n = f_d(\mathbf{X}_n) + \epsilon_n, \qquad n = 1, \ldots, N \tag{20}$$

where $\epsilon_n \sim \mathcal{N}(0, V(\mathbf{X}_n)/J_n)$. We now follow the approach proposed in [35] where the true (unknown variance) $V(\mathbf{X}_n)$ is replaced by the predictions of the metamodel constructed in (16). More specifically, we assume that:

$$\overline{Y}_n = f_d(\mathbf{X}_n) + \epsilon_n, \qquad \epsilon_n \sim \mathcal{N}\Big(0, \frac{1}{J_n}e^{\widehat{\log V}(\mathbf{X}_n)}\Big). \tag{21}$$

We note that (21) defines an observational model for $f_d(\mathbf{X})$ that can be used within a GPR framework with heteroscedastic (i.e. non-constant and input-dependent) variance [68, 67] to construct an emulator for $f_d(\mathbf{X})$. To this end, we define the training set $\mathcal{D}_Y = \{(\mathbf{X}_n, \overline{Y}_n)\}_{n=1}^N$. We further assume a centred GP prior for f_d. That is:

$$f_d(\mathbf{X}) \sim \mathcal{GP}(0, k_{\theta_y}(\mathbf{X}, \mathbf{X}')) \tag{22}$$

Under the GPR framework, predictions at testing points are then given via the mean $\hat{f}_d(\mathbf{X}^*)$ and variance $\hat{K}_y(\mathbf{X}^*)$ defined by:

$$\hat{f}_d(\mathbf{X}^*) = K_{\theta_y}^T(\mathbf{X}^*)(K_{\theta_y} + \Sigma_{\hat{\mathbf{V}}})^{-1}\overline{\mathbf{Y}} \quad (23)$$

$$\hat{K}_y(\mathbf{X}^*) = k_{\theta_y}(\mathbf{X}^*, \mathbf{X}^*) - K_{\theta_y}^T(\mathbf{X}^*)(K_{\theta_y} + \Sigma_{\hat{\mathbf{V}}})^{-1}K_{\theta_y}(\mathbf{X}^*), \quad (24)$$

where K_{θ_y} and $K_{\theta_y}(\mathbf{X}^*)$ are defined in a similar fashion to K_{θ_v} and $K_{\theta_v}(\mathbf{X}^*)$ above (i.e. with k_{θ_v} replaced by k_{θ_y}) and where $\Sigma_{\mathbf{V}} = \text{diag}(\frac{1}{J_1}e^{\widehat{\log V}(\mathbf{X}_1)}, \ldots, \frac{1}{J_N}e^{\widehat{\log V}(\mathbf{X}_N)})$ and $\overline{\mathbf{Y}} = (\overline{Y}(\mathbf{X}_1), \ldots, \overline{Y}(\mathbf{X}_N)$. Finally, we estimate the parameter θ_y from the pre-averaged marginal likelihood given by [68]:

$$p(\mathcal{D}_y|\theta_y) = \frac{1}{2}\overline{\mathbf{Y}}^T(K_{\theta_y} + \Sigma_{\hat{\mathbf{V}}}\mathbf{I})^{-1}\overline{\mathbf{Y}} + \frac{1}{2}\log|K_{\theta_y} + \Sigma_{\hat{\mathbf{V}}}| + \frac{N}{2}\log 2\pi \quad (25)$$

Similar to the discussion of the preceding subsection, predictions of $f_d(\mathbf{X}^*)$ can be computed via (23)-(24) with θ_y replaced by $\hat{\theta}_y$. The methodology is now summarised in Algorithm 1

Algorithm 1. *Algorithm to construct metamodels for* $f_d(\mathbf{X})$ *and* $\log V(\mathbf{X})$.

(1) Select/sample N *training points from the input space* $\{\mathbf{X}_n\}_{n=1}^N \subset \mathcal{X}$

for $n = 1, \ldots, N$ ***do*** *(e.g. via LHS)*

(1.1) Compute realisation J_n *of the stochastic simulator* $Y_n^{(j)} = F(\mathbf{X}_n, \omega^{(j)})$, $(j = 1, \ldots, J_n)$

(1.2) Compute the sample mean and variance $\overline{Y}_n$ *and* s_n^2

end for

(2) Use the training set $\mathcal{D}_v = \{(\mathbf{X}_n, \log(s_n^2) + \beta_n)\}_{n=1}^N$ *to fit a GP to* $\log V(\mathbf{X})$ *(given by (16) with* θ_v *and* σ^2 *computed for example via MAP estimation).*

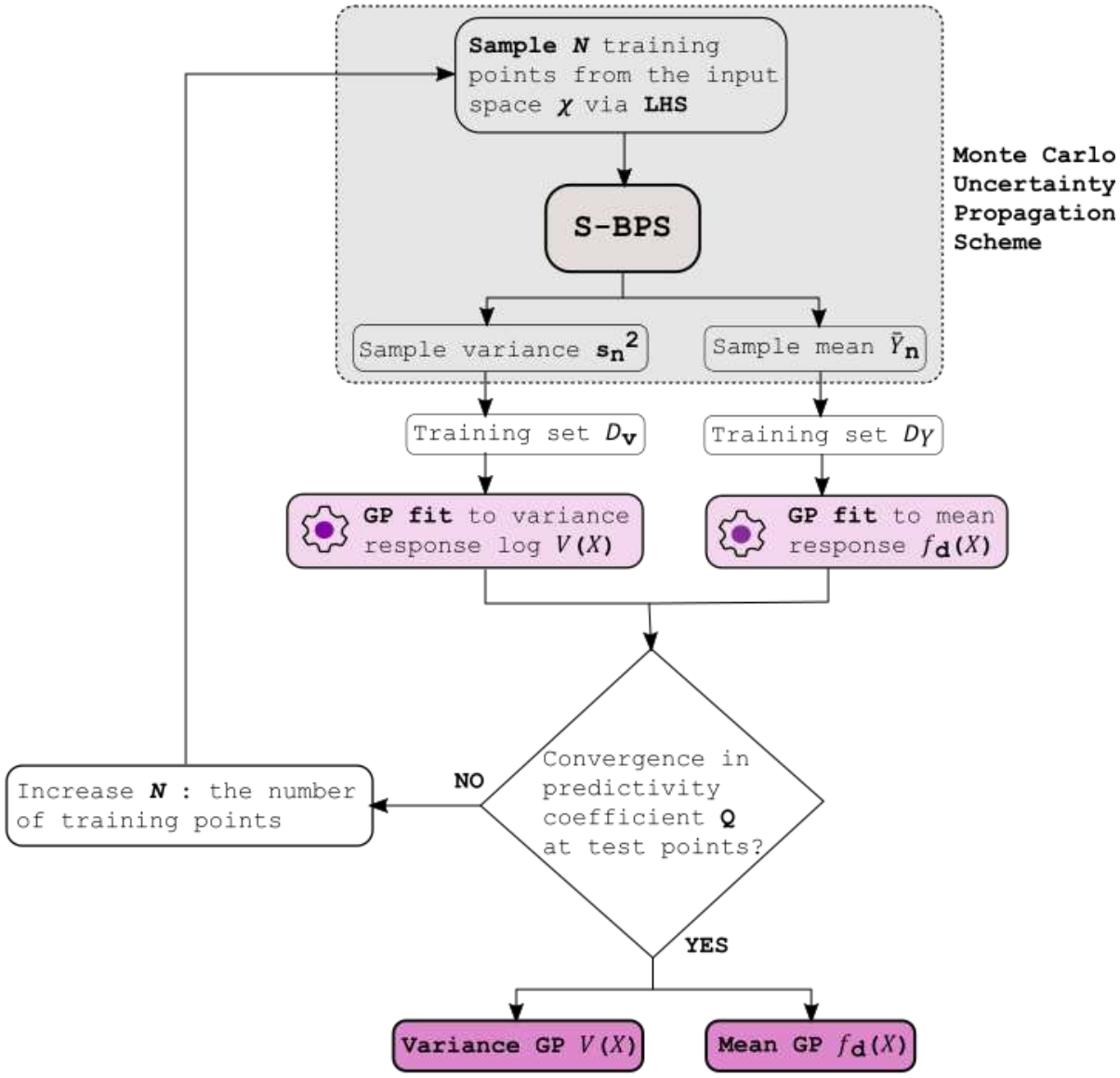
(3) Use the estimate from (2) and the training set $\mathcal{D}_Y = \{(\mathbf{X}_n, \overline{Y}_n)\}_{n=1}^N$ *to fit a GP with input-dependent variance to the mean response* $f_d(\mathbf{X})$ *(given by (23)-(24) and* θ_y *computed via MAP estimation).*

end

Figure 3 illustrates the methodology workflow, according to algorithm 1, for obtaining a validated pair of mean and variance GPs.

4.3. Computational aspects

For the selection of training points to construct the GPs involved in Algorithm 1 we use the LHS sampling strategy. To generate the GP metamodels for the log-variance and the mean response we use the *GPstuff* package for Gaussian process models developed in [36] and



Algorithm 1 : Construction of Gaussian Process (GP) Emulator to S-BPS

Figure 3: Flow diagram for Algorithm 1

available at `https://github.com/gpstuff-dev/gpstuff`. Note that the GP fitted to the mean response f_d requires a modification of GPstuff to incorporate non-constant variance given by (16), as well as the corresponding routines that compute the MAP estimator of hyperparameters.

For a given computational budget, in terms of the maximum number of simulation runs allowed, Algorithm 1 can be implemented with different combinations of N_d training/design vs N_r number of repetitions/samples. In a similar vein to a sequential search methodology, as proposed in [77], to optimise the specification of the input space and the number of stochastic simulation replicates, we specify a fixed total computational budget $C = N_d N_r$. Then, we conduct a number of numerical experiments corresponding to different combinations of the number of design points and associated replicates by systematically increasing the number of design points in steps, and by reducing the replicaties to respect our set computational budget. For each combination, we evaluate the goodness of fit of GP response surfaces using the predictivity coefficient and observe the set of these values for the considered computational budget. Depending upon the quality of GP fits, the initial computational budget is incrementally increased and the procedure for conducting numerical experiments for design point and replicate combinations is repeated for this new budget. In Section 7.1 we compare goodness of GP fits evaluated at test data for different combinations of design points and repetitions under obtained computational budget.

4.4. Emulation of the stochastic response

Constructing the metamodels in the previous sections enable us to capture uncertainty due to the stochasticity of the simulator at a given input point $\mathbf{X}$. This is relevant for the propagation of uncertainty through complex models [here through an inexpensive metamodel - a plausible approximation of our simulator] and its subsequent quantification under stochasticity. Note that a particular case of the stochastic simulator from (1) arises when the function f can be written in the following form:

$$Y = f(\mathbf{X}, \omega) = f_d(\mathbf{X}) + f_s(\mathbf{X}, \omega), \qquad f_s(\mathbf{X}, \omega) \sim \mathcal{N}(0, V(\mathbf{X})). \tag{26}$$

where, as before, $f_d(\mathbf{X})$ is the mean response. In this case $f_s(\mathbf{X}, \omega)$ is a Gaussian random variable with input-dependent variance. Under the modelling assumption given by (26), the mean response $f_d(\mathbf{X})$ and the variance $V(\mathbf{X})$ fully characterise the stochastic response at a given point $\mathbf{X}$ in the input space. Constructing metamodels according to the methodology described in subsections 4.1-4.2 allows us to simulate the response from the stochastic simulator. Indeed, simulations of the variable Y at at testing point $\mathbf{X}$ can be obtained by:

$$Y = f_d(\mathbf{X}) + \sqrt{V(\mathbf{X})}\xi, \qquad \xi \sim \mathcal{N}(0, 1). \tag{27}$$

Similarly, two-sided α-tailed credible intervals can be defined by:

$$\mathcal{CI}_\alpha(\mathbf{X}) = \left[f_d(\mathbf{X}) - \tau_\alpha \sqrt{V(\mathbf{X})}, f_d(\mathbf{X}) + \tau_\alpha \sqrt{V(\mathbf{X})} \right] \tag{28}$$

where, as before, τ_α is the $(1-\frac{\alpha}{2})$ quantile of the standard Normal distribution. The predictors from the metamodels for $f_d(\mathbf{X})$ and $V(\mathbf{X})$ can be used in expressions (27)-(28) for the efficient emulation of our stochastic simulator and to approximate credible intervals. In Section 7.3, we assess using (28) the capabilities of the emulator to capture the stochastic observations obtained from the stochastic building performance simulator.

5. Global sensitivity analysis through metamodeling

We now discuss how the proposed approach to quantify uncertainties in S-BPS can be combined with a Global Sensitivity Analysis (GSA) technique to decompose the total uncertainty in the model output into the respective sources of uncertainty in the model inputs [78]. We consider a variance based method, based on sensitivity indices that enable us to quantify the importance of a given uncertain input relative to others. These indices are the quantitative measures of knowing an impact of uncertain sources on the model predictions (so that those uncertainties can be subsequently addressed if possible).

5.1. The deterministic case

Let us consider again the case of a deterministic simulator $Y = F(\mathbf{X})$ discussed in subsection 3.1. We denote by I a set of indices from $\{1, \ldots, p\}$ with cardinality denoted by $|I|$. We use standard index notation for which X_I denotes a vector that contains the components of $\mathbf{X} = (X_1, \ldots, X_p)$ with indices in I. For example, if $I = \{1, 4\}$, then $X_{\{1,4\}} = (X_1, X_4)$. Furthermore, we denote by $\mathbf{X}_{\sim I}$ the vector with all components of $\mathbf{X}$ except those that belong to the set I. Using the total variance decomposition theorem, the unconditional variance of the model output Y can be expressed as [79]:

$$\mathbb{V}_{\mathbf{X}}[Y] = \sum_{i=1}^{p} \sum_{|I|} \mathcal{V}_I(Y), \tag{29}$$

where

$$\mathcal{V}_I(Y) = \mathbb{V}_{X_I}[\mathbb{E}_{\mathbf{X}_{\sim I}}(Y|X_I)]. \tag{30}$$

In (30) we use $\mathbb{E}_\kappa(Z)$ and $\mathbb{V}_\kappa(Z)$ to denote expectation and variance of Z with respect to κ. Let us note that when $I = \{i\}$ and $I = \{i, j\}$, for example, we have

$$\mathcal{V}_i(Y) = \mathbb{V}_{X_i}[\mathbb{E}_{\mathbf{X}_{\sim i}}(Y|X_i)] \tag{31}$$

and

$$\mathcal{V}_{i,j}(Y) = \mathbb{V}_{X_i,X_j}[\mathbb{E}_{\mathbf{X}_{\sim i,j}}(Y|X_i, X_j)] - \mathcal{V}_i(Y) - \mathcal{V}_j(Y), \tag{32}$$

respectively.

Sobol indices of $Y = F(\mathbf{X})$ with respect to controllable parameters is defined by [79]:

$$S_I(Y) = \frac{\mathcal{V}_I(Y)}{\mathbb{V}_{\mathbf{X}}(Y)}. \tag{33}$$

For $I = \{i\}$, S_i represents the main (first order) effect giving the measure of X_i on the model predictions Y. The second order effect S_{ij} expresses sensitivity of the model outcomes to the interaction between the variables X_i and X_j without the individual effects of X_i and X_j; higher order indices represent third, fourth, ... order interaction effects. From (29) it follows that the sum of these indices (all positive) is equal to one. The larger and close to one an index value is, the higher will be the importance of that input or the group of inputs that is linked to this index. Note that for a model with p inputs, the number of Sobol indices would amount to $2^p - 1$; leading to an intractable number of indices as p increases.

The total sensitivity of the model output variance to an input X_i is given by [79]:

$$S_{T_i}(Y) = \sum_{i \in \subset I} S_I \tag{34}$$

For instance, for a model with three input parameters, $S_{T_1} = S_1 + S_{12} + S_{13} + S_{123}$. In practice, when the model function F has multiple input parameters, only the Sobol indices for the main and total effect are computed.

5.2. The stochastic case

We follow the approach of [69] for GSA for our S-BPS of the form $Y = f(\mathbf{X}, \omega)$, in which variable ω is treated as an additional uncontrollable parameter. Using the total variance decomposition theorem, the unconditional variance of the model output Y can be expressed as:

$$\mathbb{V}_{\mathbf{X},\omega}[Y] = \mathcal{V}_\omega(Y) + \sum_{i=1}^{p} \sum_{|I|} [\mathcal{V}_I(Y) + \mathcal{V}_{I,\omega}(Y)], \tag{35}$$

where:

$$\mathcal{V}_\omega(Y) = \mathbb{V}_\omega[\mathbb{E}_{\mathbf{X}}(Y|\omega)], \qquad \mathcal{V}_I(Y) = \mathbb{V}_{X_I}[\mathbb{E}_{\mathbf{X}_{\sim I},\omega}(Y|X_I)] - \sum_{I' \subsetneq I} \mathcal{V}_{I'}(Y)), \tag{36}$$

$$\mathcal{V}_{I,\omega}(Y) = \mathbb{V}_{X_I,\omega}[\mathbb{E}_{\mathbf{X}_{\sim I}}(Y|X_I,\omega)] - \sum_{I' \subsetneq I} \mathcal{V}_{I',\omega}(Y) - \sum_{I' \subsetneq I} \mathcal{V}_{I'}(Y) - \mathcal{V}_\omega(Y) \tag{37}$$

Similar to the deterministic case, Sobol indices of $Y = f(\mathbf{X}, \omega)$ with respect to controllable parameters $\mathbf{X}$ can be defined by:

$$S_I(Y) = \frac{\mathcal{V}_I(Y)}{\mathbb{V}(Y)}, \qquad S_{T_i}(Y) = \sum_{i \in \subset I} S_I \tag{38}$$

Sensitivity indices that involve the uncontrollable parameter ω are given by:

$$S_{I,\omega}(Y) = \frac{\mathcal{V}_{I,\omega}(Y)}{\mathbb{V}(Y)}, \qquad S_{\omega}(Y) = \frac{\mathcal{V}_{\omega}(Y)}{\mathbb{V}(Y)}. \tag{39}$$

and the total sensitivity with respect to ω is:

$$S_{T_\omega}(Y) = \sum_{i=1}^{p} \sum_{|I|} S_{I,\omega} + S_{\omega}. \tag{40}$$

which from (29) can be written as:

$$S_{T_\omega}(Y) = \sum_{i=1}^{p} \sum_{|I|=i} S_{I,\omega} + S_{\omega} = \frac{1}{\mathbb{V}(Y)} \Big[\sum_{i=1}^{p} \sum_{|I|=i} \mathcal{V}_{I,\omega}(Y) + \mathcal{V}_{\omega}(Y) \Big] = 1 - \frac{1}{\mathbb{V}(Y)} \sum_{i=1}^{p} \sum_{|I|=i} \mathcal{V}_I(Y), \tag{41}$$

The method proposed in [69] then allows us to write (38) and (41) in terms of $f_d(\mathbf{X})$ and $V(\mathbf{X})$, for which we have constructed GP-based metamodels. Indeed, from (31), (2) and the assumption of independence between ω and X_i we have that:

$$\mathcal{V}_i(Y) = \mathbb{V}_{X_i} \mathbb{E}_{\mathbf{X}_{\sim i},\omega}(Y|X_i) = \mathbb{V}_{X_i}[\mathbb{E}_{\mathbf{X}_{\sim i}}[\mathbb{E}_{\omega}(Y|\mathbf{X})|X_i]] = \mathbb{V}_{X_i}[\mathbb{E}_{\mathbf{X}_{\sim i}}(f_d|X_i)] = \mathcal{V}_i(f_d(\mathbf{X})). \tag{42}$$

More generally, it can be shown that $\mathcal{V}_I(Y) = \mathcal{V}_I(f_d(\mathbf{X}))$. Thus, (38) becomes:

$$S_I(Y) = \frac{\mathcal{V}_I(Y)}{\mathbb{V}(Y)} = \frac{\mathcal{V}_I(f_d(\mathbf{X}))}{\mathbb{V}(Y)} = S_I(f_d(\mathbf{X})) \frac{\mathbb{V}_{\mathbf{X}}(f_d(\mathbf{X}))}{\mathbb{V}(Y)} \tag{43}$$

where $S_I(f_d(\mathbf{X})) = \frac{\mathcal{V}_I(f_d(\mathbf{X}))}{\mathbb{V}_{\mathbf{X}}(f_d(\mathbf{X}))}$ is the Sobol index of $f_d(\mathbf{X})$. In addition, from the total variance theorem and the definitions (2) we have:

$$\mathbb{V}[Y] = \mathbb{V}_{\mathbf{X}}[\mathbb{E}_{\omega}(Y|\mathbf{X})] + \mathbb{E}_{\mathbf{X}}[\mathbb{V}_{\omega}(Y|\mathbf{X})] = \mathbb{V}_{\mathbf{X}}[f_d(\mathbf{X})] + \mathbb{E}_{\mathbf{X}}[V(\mathbf{X})] \tag{44}$$

Substitution of (44) in (43) yields:

$$S_I(Y) = S_I(f_d(\mathbf{X})) \frac{\mathbb{V}_{\mathbf{X}}(f_d(\mathbf{X}))}{\mathbb{V}_{\mathbf{X}}[f_d(\mathbf{X})] + \mathbb{E}_{\mathbf{X}}[V(\mathbf{X})]} \tag{45}$$

Note that that the variance of $f_s(\mathbf{X})$ can now be decomposed into the respective contributions of its inputs $\mathbf{X}$:

$$\mathbb{V}_{\mathbf{X}}[f_d(\mathbf{X})] = \sum_{i=1}^{p} \sum_{|I|=i} \mathcal{V}_I(f_d(\mathbf{X})) = \sum_{i=1}^{p} \sum_{|I|=i} \mathcal{V}_I(Y) \tag{46}$$

Therefore, combining (46) with (41) and (44) we obtain:

$$S_{T_\omega}(Y) = 1 - \frac{1}{\mathbb{V}(Y)} \sum_{i=1}^{p} \sum_{|I|=i} [\mathcal{V}_I(Y)] = 1 - \frac{\mathbb{V}_{\mathbf{X}}[f_d(\mathbf{X})]}{\mathbb{V}(Y)} = \frac{\mathbb{E}_{\mathbf{X}}[V(\mathbf{X})]}{\mathbb{V}_{\mathbf{X}}[f_d(\mathbf{X})] + \mathbb{E}_{\mathbf{X}}[V(\mathbf{X})]}, \tag{47}$$

Expression (45) and (47) enable us to compute Sobol indices by using the approximation of $f_d(\mathbf{X})$ and $\log V(\mathbf{X})$ that we obtained from the GP metamodel constructed in subsections 4.1-4.2. In other words, we use the following approximations:

$$S_I(Y) \approx S_I(\hat{f}_d(\mathbf{X})) \frac{\mathbb{V}_{\mathbf{X}}(\hat{f}_d(\mathbf{X}))}{\mathbb{V}_{\mathbf{X}}[\hat{f}_d(\mathbf{X})] + \mathbb{E}_{\mathbf{X}}[\exp(\widehat{\log V}(\mathbf{X}))]}, \tag{48}$$

$$S_{T_\omega}(Y) \approx \frac{\mathbb{E}_{\mathbf{X}}[\exp(\widehat{\log V}(x))]}{\mathbb{V}_{\mathbf{X}}[\hat{f}_d(\mathbf{X})] + \mathbb{E}_{\mathbf{X}}[\exp(\widehat{\log V}(\mathbf{X}))]} \quad . \tag{49}$$

For the computation of sensitivity indices, accounting for both individual and total effects of our parameters, we use the *Sensitivity Analysis For Everybody (SAFE)* package developed in [80] and available at `https://www.safetoolbox.info`. After obtaining well trained and validated GP emulators for $f_d(\mathbf{X})$ and $\log V(\mathbf{X})$, we study the convergence of our sensitivity indices (48) and (49) by running this inexpensive substitute to the stochastic BPS for different sample sizes, and correspondingly perform uncertainty propagation and decomposition. Since the sensitivity indices (48) and (49) involve statistical quantities computed from samples, a convergence study is required to ensure that the enough samples are used to accurately approximate those sensitivity indices.

For the convergence of sensitivity indices we adopt the criterion proposed in [81], as the index value remains stable (or changes to a limited degree within the credible interval) when computed at multiple samples of increasing size. This convergence criterion uses the bootstrap technique to provide a measure of uncertainty in the value of the sensitivity index [82]. Since the sensitivity index value is an estimate obtained from the drawn samples, so that its true value cannot be known, it is advisable to account for uncertainty in estimates of sensitivity index due to the sampling variability. The application of a bootstrap technique, along with the chosen sampling scheme, allows for the estimation of an approximate probability distribution around the sensitivity index value, giving confidence intervals as lower and upper bounds for the measure of uncertainty in its estimate. This is particularly helpful when assessing the error in the index estimate at sampling sets of increasing size. In bootstrapping, the different resamples are randomly drawn with replacement from the original sample of the model input / output. To assess the convergence of the sensitivity indices, we observe the width of the 95% credible intervals of the index distribution obtained via bootstrapping. In particular, for each input and across different sample sizes, we observe:

$$CI_{indices} = (S_{ub} - S_{lb}) \tag{50}$$

where, S_{ub} and S_{lb} are the upper and lower bounds of the probability distribution of the sensitivity index of the input factor at the given sample size. When the value of the width of the confidence interval is close to or stabilizes near to zero or a small threshold value (say, 0.05), the sensitivity index is deemed to have converged. In Section 7.4, we present the

convergence of main and total effect sensitivity indices of the inputs involved in this study, across multiple sample sizes.

Algorithm 2 summarises the algorithm to estimate sensitivity indices for the stochastic simulator with respect to controllable variables.

Algorithm 2. *Algorithm to compute Sobol indices S_I and S_{T_ω}. Consider approximations $f_d(\mathbf{X}) \approx \hat{f}_d(\mathbf{X})$ and $V(\mathbf{X}) \approx \exp(\widehat{\log V}(\mathbf{X}))$ from the metamodels constructed from Algorithm 1.*

(1) Compute Sobol sensitivity indices for $\hat{f}_d(\mathbf{X})$ (i.e. $S_I(\hat{f}_d(\mathbf{X}))$) using standard techniques.

(2) Compute $\mathbb{V}_{\mathbf{X}}[\hat{f}_d(\mathbf{X})]$ via Monte Carlo sampling (i.e. Sample from the distribution of $\mathbf{X}$, evaluate $\hat{f}_d(\mathbf{X})$ via the GP predictor (23) and compute sample variance).

(3) Compute $\mathbb{E}_{\mathbf{X}}[\exp(\widehat{\log V}(\mathbf{X}))]$ via Monte Carlo sampling (i.e. Sample from the distribution of $\mathbf{X}$, evaluate $\exp(\widehat{\log V}(\mathbf{X}))$ via the GP predictor (16) and compute sample mean).

(4) Use (1)-(3) to compute $S_I(Y)$ and $S_{T_\omega}(Y)$ via expressions (48)-(49)

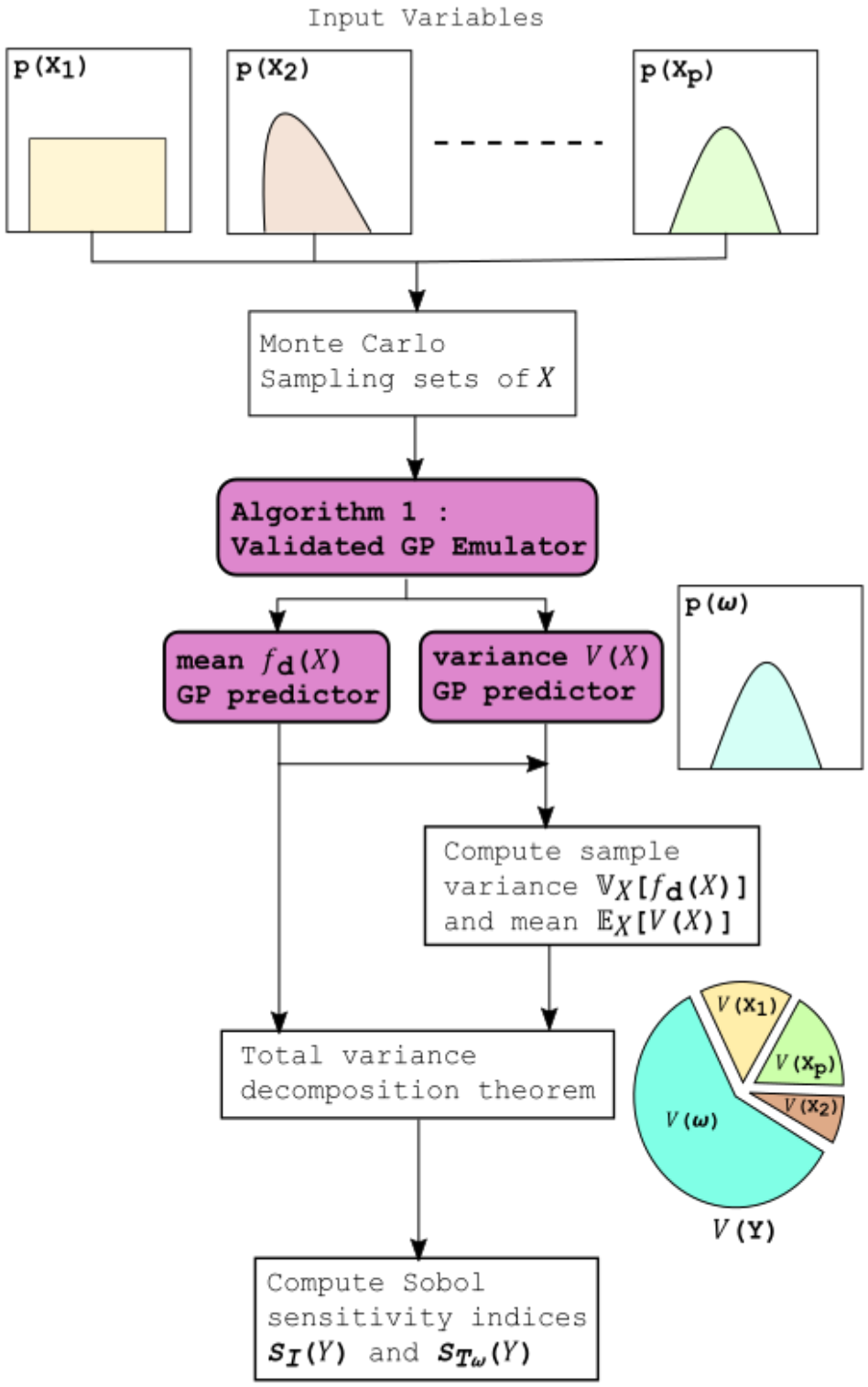
end

Figure 4 illustrates algorithm 2, employing validated metamodels following algorithm 1 for the efficient computation of sensitivity indices i.e. the quantification and decomposition of the total uncertainty into its respective sources.

6. Case Study

We analyse the individual and collective impacts of two deterministic phenomena: (d1) heat conduction through a wall construction, and (d2) incoming shortwave irradiation[3]; and through three stochastic phenomena: (s1) occupants' radiative and convective metabolic heat gains due to their presence, (s2) advective heat losses due to ventilation through the opening and closing of windows, and (s3) radiative heat gains due to the raising and lowering of shading devices. The impacts of variation in the deterministic phenomena are studied by assessing the impacts of variations in wall insulation thickness and window transmittance, in plausible ranges of their uncertainty. We have selected insulation thickness for convenience,

[3]Here, the incoming shortwave irradiation through a window is assumed to be a deterministic component because for the purpose of our study we are considering deterministic input weather excitations.



Algorithm 2 : Computation of Sobol sensitivity indices through metamodeling

Figure 4: Flow diagram for Algorithm 2

as this has a similar effect on heat transfer to variations in conductivity [4]. We realise that we could also have varied density and specific heat capacity for opaque materials, as well as emissivity and conductivity for transparent materials. However, as our primary aim here is merely to present the proof of principle of our framework and to compare the impact of deterministic phenomena against their stochastic counterparts, we believe that our pragmatic choices of parameters are sufficient for our purposes; that this peculiarity does not negatively impact on the usefulness of our evaluation exercise. For stochastic phenomena, we simply repeat simulations with the relevant models enabled, randomly seeding a random number generator for each run.

6.1. Experiment set up

As noted earlier (2), we generate our training data through a co-simulation approach, using the Functional Mockup Interface co-simulation standard, in which the dynamic BPS tool Energy Plus is the master algorithm, controlling the execution of the multi-agent stochastic simulation tool, No-MASS [5]. In common with Haldi and Robinson's 2011 study of the impact of stochastic models on simulated building performance [11], we employ a model of a simple hypothetical monozone (shoe-box) office building, also located in Geneva Switzerland ($46.25°N, 6.13°E$, elevation 416m), for which the corresponding Energy Plus weather data file was used. This hypothetical office building has dimensions 3.5m×4.5m, and a floor to ceiling height of 2.8m (Figure 5). The south facade of the building has a double glazed window occupying 40% of its surface area, located above the working plane height of 0.8m above the ground. The wall and window construction compositions used in this study are given in Table 1. The sensible internal heat gains are assumed to be a maximum of 400W, collectively accounting for metabolic heat gains due to occupants' presence and casual heat gains due to lighting and any electrical equipment in use whilst occupants are present.

The heating (and cooling) and ventilation temperature set-point (SP) schedules are set to 21°C (and 25°C) and 24°C during working hours, according to ANSI/ASHRAE/IES Standard 90.1 for a small office prototype building model, representing an ideal loads HVAC system operation.

The graphical user interface to Energy Plus, DesignBuilder 4.3, was used to generate the input data file (.idf) according to the above specifications.

6.2. Parameters studied

We consider two controllable parameters in the S-BPS model described earlier: insulation thickness X_1 and window transmittance X_2. The prediction outputs are annual heating Y_h and Y_c cooling demand per unit floor area. Our objective is to quantify the uncertainty in

[4] However, we do acknowledge that conductivity is more likely to vary as a result of manufacturing practices, and in response to variations in moisture content, than is thickness; but this latter remains useful for the purpose of demonstration.

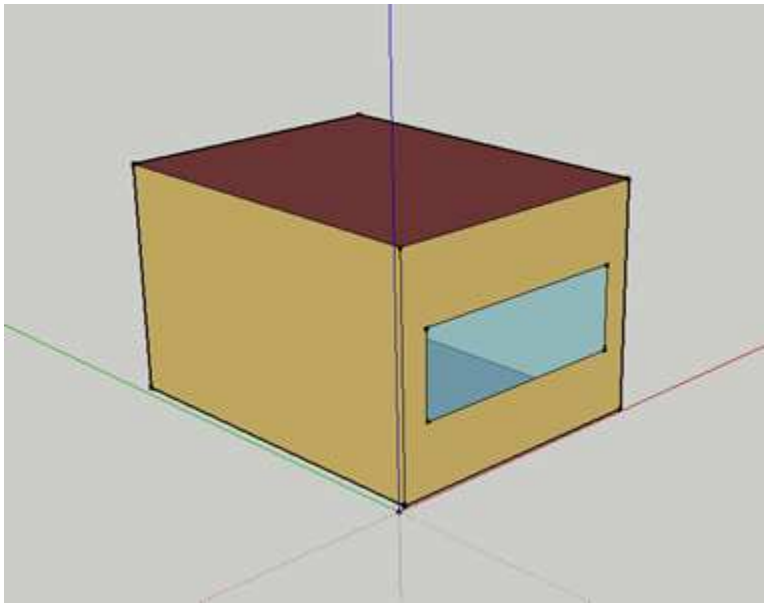


Figure 5: Building geometry of hypothetical office

Table 1: Thermophysical properties of construction layers : thickness, thermal conductivity, specific heat capacity, density and U-value

Surface	Material	t(m)	λ(W/(mK))	C(J/(kgK))	ρ(kg/m^3)	U(W/(m^{2}K))
Wall	Brick	0.1	0.84	800	1700	0.28
	Insulation	0.111	0.034	1400	35	
	Concrete	0.1	0.51	1000	1400	
	Plaster	0.013	0.4	1000	1000	
Ground	Foam	0.1327	0.04	1400	10	0.26
	Concrete	0.1	1.13	1000	2000	
	Screed	0.07	0.41	840	1200	
	Flooring	0.03	0.14	1200	650	
Roof	Asphalt	0.01	0.7	1000	2100	0.26
	Glass Wool	0.1445	0.04	840	12	
	No mass	Thermal	Resistance	= 0.18		
	Plasterboard	0.013	0.25	896	2800	
Window	Glass	0.003	0.9	750	2500	2
	Air	0.013	0.0262	1005	1.17	
	Glass	0.003	0.9	750	2500	

Y_h and Y_c given uncertainty in $\mathbf{X} = (X_1, X_2)$ as well as the (uncontrollable) uncertainty that arises from occupants' behaviour, encoded in the model described earlier. We assume that X_1 and X_2 are independent and characterised by uniform distributions, and that the values for these design variables are equally likely within the possible uncertain ranges specified [13] [83]. The assignment of a uniform distribution is desirable for the case of poorly defined design variables (e.g. during early design stage) where only an estimate of the lower and upper bounds may be available [84]. For the distribution of insulation thickness we use values between 0.06m and 0.16m to account for plausible variations in insulation thickness (e.g. from low to moderate thickness) within the office building's facades, entailing a corresponding variation in heat conduction through walls. For glazing solar transmittance, the distribution takes values between 0 and 1, effectively modulating the quantity and quality of glazing in the south façade, entailing a corresponding variation in transmitted shortwave radiation [5]. The solar transmittance is applied to the façade glazing ratio, where zero transmittance represents a combination of glazing utilisation (the proportion of the glazed surfaces that is taken up by frames) and transmission (more or less transparent materials) that effectively renders the glazing opaque, and a value of unity represents completely unobstructed glass of perfect transmission properties; with values tending to lie between these (unphysical) extremes. The formulation of the occupant behavioural models and their related parameter settings are adopted as described in Chapman [75].

To apply the proposed framework described in Sections 4- 5 we treat the annual heating and cooling demand Y_h and Y_c independently, and thus construct GP metamodels for each of these output variables. More specifically, we apply Algorithm 1 for each of the model outputs Y_h and Y_c, and subsequently use those GPs within Algorithm 2 to conduct sensitivity analysis for each output variable. Technical details relevant to the GP prior settings used within the proposed approach can be found in Appendix A.

7. Results and Discussion

7.1. Design points vs repetitions

We note from Algorithm 1 that the proposed metamodeling framework applied to the case study introduced above involves constructing GP emulators of both the mean response $f_d(X_1, X_2)$ and the log variance $V(X_1, X_2)$ for each of the S-BPS response outputs Y_h and Y_c. We also observe that the methodology requires us to select a number of design points (via LHS) within the input space for (X_1, X_2) that we introduced in subsection 6.2, and to conduct a number of repetitions of the model outputs at each of those design points. The accuracy of the proposed methodology relies on having a large enough sample of design points (so that

[5]Note that we have not considered in this work uncertainty in visible transmittance, and its possible intercorrelation with solar transmittance, as we are not modelling daylight responsive controls of artificial lights and associated heat gains.

Table 2: Comparison of quality of prediction Q of fitted GP response surfaces at test data given a fixed computation budget (of 3000 simulation runs) for training. Here the item entry 5×600 represents 600 simulation repetitions at 5 design points and so on.

Output target variable	Emulated quantity	Predictivity coefficient Q					
		5×600	10×300	15×200	20×150	25×120	30×100
Heating demand Y_h	$f_d(\mathbf{X})$	0.9367	0.9949	0.9997	0.9999	0.9999	0.9999
	$\log V(\mathbf{X})$	0.3179	0.9186	0.9159	0.9235	0.962	0.9483
Cooling demand Y_c	$f_d(\mathbf{X})$	0.9704	0.9858	0.9831	0.9782	0.9841	0.9719
	$\log V(\mathbf{X})$	0.0841	0.7765	0.932	0.8993	0.9184	0.9092

input uncertainty $\mathbf{X}$ is well characterised) and a sufficient number of repetitions that can accurately inform, at those design points, the mean and variance response (with respect to ω) of each model output. To arrive at the reasonable computational budget of 3000 simulator runs, we follow the sequential search approach outlined in 4.3 and observe the accuracy of the predictions of each GP involved in the proposed methodology using different combinations of design points and replicates.

Once all GPs have been constructed for each combination, we compute the predictivity coefficient Q and the CI validation measure discussed in subsection 3.3. The test data for this case study consists of simulation runs at 60 new sampling points in the input design space, each with 600 repetitions of the stochastic simulator. For the design point and replicate combinations having a budget of 3000 simulator runs to train the emulator, Table 2 reports the accuracy in terms of Q obtained at the test data points by the GP emulators of the mean and log-variance. The coefficient of predictivity saturates to a value of more than 0.92 when around 25 design points with 120 repetitions are available for training (Figure 6). Any further increase in the number of design points does not lead to significant increase in predictive accuracy. Table 3 reports the proportion of test observations that lie inside the 95% CI (according to Eq. 11)) predicted by the GP emulators. Also, for the training combination (25 × 120), the proportion of test observations within the emulator uncertainty interval is more than 80% without a loss in predictive accuracy. We note that the cooling demand log-variance GP displays less accuracy in the CI measure, arguably due to the large variation in non-constant variance exhibited by occupants' behaviours.

7.2. Gaussian Process emulation of the mean $f_d(\mathbf{X})$ and log-variance $\log V(\mathbf{X})$ of the S-BPS outputs

Based on the fixed computation budget study of prediction accuracy of GPs (Table 2 and 3), the combination 25 design / sample points with 120 repetitions of stochastic building performance simulator has been selected to be optimum in order to obtain a plausible pair of GPs predicting mean and log-variance of energy demands Y_h and Y_c. The GP regression surfaces for the computed energy demands' mean and log-variance are shown in Figure 7

Table 3: Comparison of emulator uncertainty (Eq. 11) of fitted GPs in terms of proportion of test points lying inside the uncertainty interval

Output target variable	Emulated quantity	Emulator uncertainty interval CI					
		5×600	10×300	15×200	20×150	25×120	30×100
Heating demand Y_h	$f_d(\mathbf{X})$	0.9999	0.95	0.9999	0.9999	0.9833	0.9999
	$\log V(\mathbf{X})$	0.9999	0.8167	0.6333	0.7	0.8667	0.7167
Cooling demand Y_c	$f_d(\mathbf{X})$	0.9999	0.9833	0.8667	0.7333	0.8167	0.75
	$\log V(\mathbf{X})$	0.6667	0.3833	0.6333	0.35	0.4333	0.4333

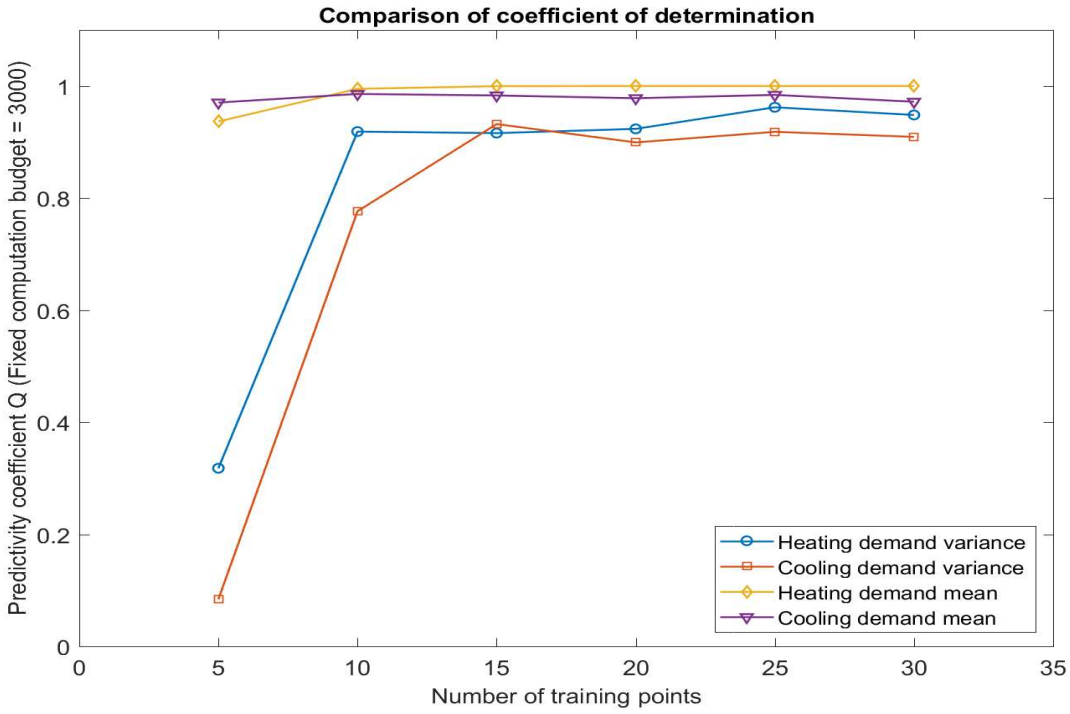


Figure 6: Comparison of prediction accuracies of mean and log-variance GPs

together with the training set used for the construction of each GP. Each GP response surface is an interpolation surface that predicts mean energy demand value and variance around it due to stochastic occupants' behaviours for any given test point within the uncertainty ranges of the input parameters.

The GP regression surfaces from Figure 7 allow us to understand the effect of input variables on predictions. These results show that an increase in insulation thickness lowers heating demand, due to the expected reduction in heat loss through building envelope, analogously to a lowering of thermal conductivity (see Figure 7 top-left). As variations in only solar (not visible) transmittance is considered, its direct impact on heating demand is somewhat dampened. Furthermore, with a relatively moderate availability of irradiation during the heating season, variations in transmittance have a relatively lower impact than those due to insulation thickness.

Conversely, during the cooling season, our results show the inverse trend; with variations in transmittance have a greater impact, owing to the increased magnitude of available solar irradiation. Transmittance and cooling demand now show positive correlation, as an increase in transmitted shortwave radiation entails a corresponding increase in cooling loads (Figure 7 bottom-left). Furthermore, the internal storage of heat can be exacerbated by envelope thermal insulation, which inhibits heat transfer to the outside whilst the indoor temperature is higher, so increasing cooling loads (as shown in Figure 7 bottom-left). These results are consistent with the underlying physics abstracted by the BPS. Furthermore, Figure 7 illustrates the heteroscedastic (i.e. non-constant) nature of the variance due to occupants' behaviours and it exhibits a positive correlation with energy demand i.e. the non constant variance is high at higher energy demands and vice-versa.

7.3. Uncertainty quantification using the GP emulator

The validation measures discussed in subsection 7.1 enable us to assess the accuracy of the predictive capabilities of each GP (for mean and log-variance). We now assess the accuracy of the proposed methodology to statistically reproduce uncertainty in the outputs of the S-BPS within the same input-space. To this end, we consider the credible intervals introduced in (28) for $\alpha = 0.05$ (i.e. 95% CI), where the mean and variance are replaced with those obtained via the GP emulators obtained in the previous subsection. We generate test data with 100 replicates of the S-BPS at a collection of 60 randomly selected input points. Some 95.2% and 96.08% of the simulator's random response values [at each test point] for heating and cooling demands lie within the 95% CI predicted by the pair of (mean and variance) GPs (Figure 8), suggesting that this new GP emulator can be used with confidence.

To further evaluate the effectiveness of the proposed GP-based metamodeling approach, we compare in Figures 9-10 the probability density functions (PDFs), at some specific test points, computed by Monte Carlo Simulation, the S-BPS outputs (in red) and outputs from the GP emulator (in blue); this latter by simulating stochastic observations with (27) using

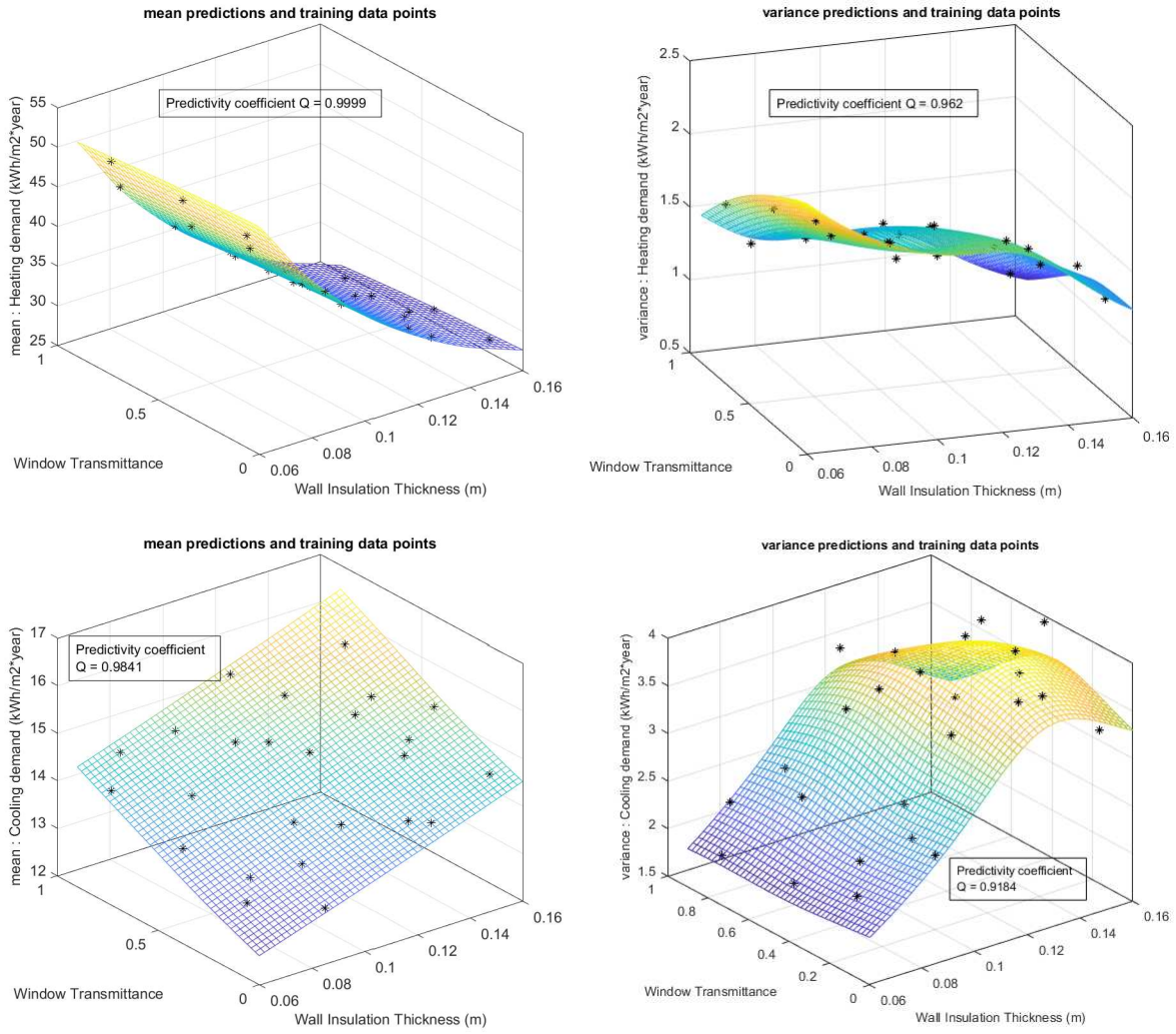


Figure 7: GP regression surface fitting to sampled mean (left) and log-variance (right) of heating demand (top) and cooling demand (bottom)

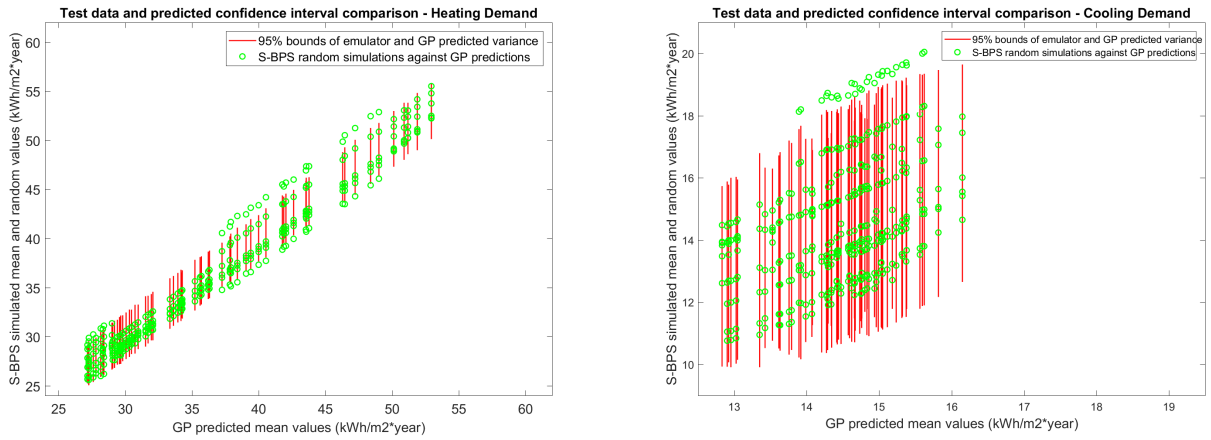


Figure 8: Comparison of S-BPS simulated random values within GP predicted credible intervals at test data for heating (a) and cooling (b) demand

GP emulators for $f_d(\mathbf{X})$ and $V(\mathbf{X})$). In Figure 11 we also compare cumulative distribution functions (CDFs) of predicted energy demands from the GP emulator with those of the S-BPS simulations, again obtained using Monte Carlo Simulation. A single CDF in these plots represent variation in prediction due to occupants' stochastic behaviour [aleatory uncertainty] and their spread represents uncertainty in predictions due to input parameters [epistemic uncertainty]. Figures 9-11 offer numerical evidence that the proposed framework enables us to accurately and robustly reproduce the underlying PDFs and CDFs obtained with the S-BPS, at a small fraction of the computational cost, since once the GP emulators have been constructed, S-BPS runs are no longer required for subsequent uncertainty analyses.

In the specific case of heating demand (the top row in Figure 11), the collective spread of CDFs is more than the spread of an individual one - illustrating that epistemic uncertainty due to inputs plays a more dominant role than the aleatory uncertainty due to occupants' behaviours. And in the case of cooling demand (the bottom row in Figure 11), we observe the opposite trend, the spread of an individual CDF exceeds the collective one, showing that the aleatory uncertainty is more dominant than its epistemic counterpart. This is expected, because the stochastic models used in this work (windows and blinds) are essentially biased towards more frequent user interactions during warmer weather (higher ambient temperatures and incident irradiances, these stimulating the opening of windows and the lowering of shading devices) when cooling loads are at their largest.

Finally, the quantitative comparison between the means of heating and cooling demand replicates presented in Figure 12 underline how accurate the GP emulator's predictions are.

Having demonstrated that the GP emulator is an acceptably accurate means for accelerating stochastic building performance simulations, we proceed now to decompose the total uncertainty in predictions into its component parts by computing sensitivity indices for our case study.

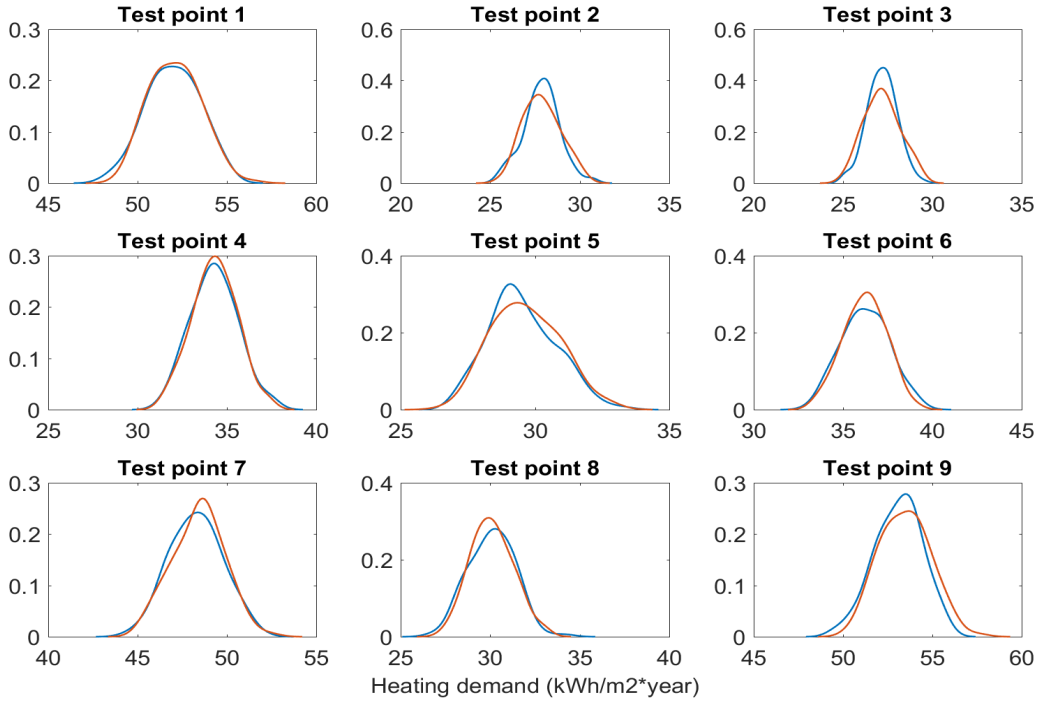


Figure 9: Uncertainty propagation comparison between GP predicted (blue) and S-BPS simulated PDFs (red) for heating demand

7.4. Uncertainty decomposition using the GP emulator : Sensitivity analysis

We now employ our GP-metamodeling framework in conjunction with the GSA approach discussed in section 5. More specifically, we decompose the total uncertainty in energy demand prediction using the individual and total effect indices of (48) and (49). The individual effect indices account for the impact of design variables 'insulation thickness' and 'window transmittance' and the total effect index includes that of occupants' behaviour. The index scale is between 0 and 1, with highly influential factors having values close to unity. Table 4 reports these indices, computed for both small (500) and large (20000) input sample sizes. In the case of heating demand, and as discussed earlier, epistemic uncertainty due to insulation thickness dominates the uncertainty in predictions, at an individual index value of 0.97. This result is in line with expectation. During heating season envelope heat losses (and thus impact of uncertainty in insulation thickness) dominates the energy balance, much more so than occupants' behaviour, since occupant are relatively less likely to open windows (to avoid excessive heat losses) and to lower shades (thus reducing useful transmitted solar heat gains) during the heating season. In contrast, for cooling demands, aleatory uncertainty due to occupants' behaviour is the most influential factor, at total index value of 0.83, as occupants are correspondingly more likely to interact with windows and shading devices, as explained in the previous section.

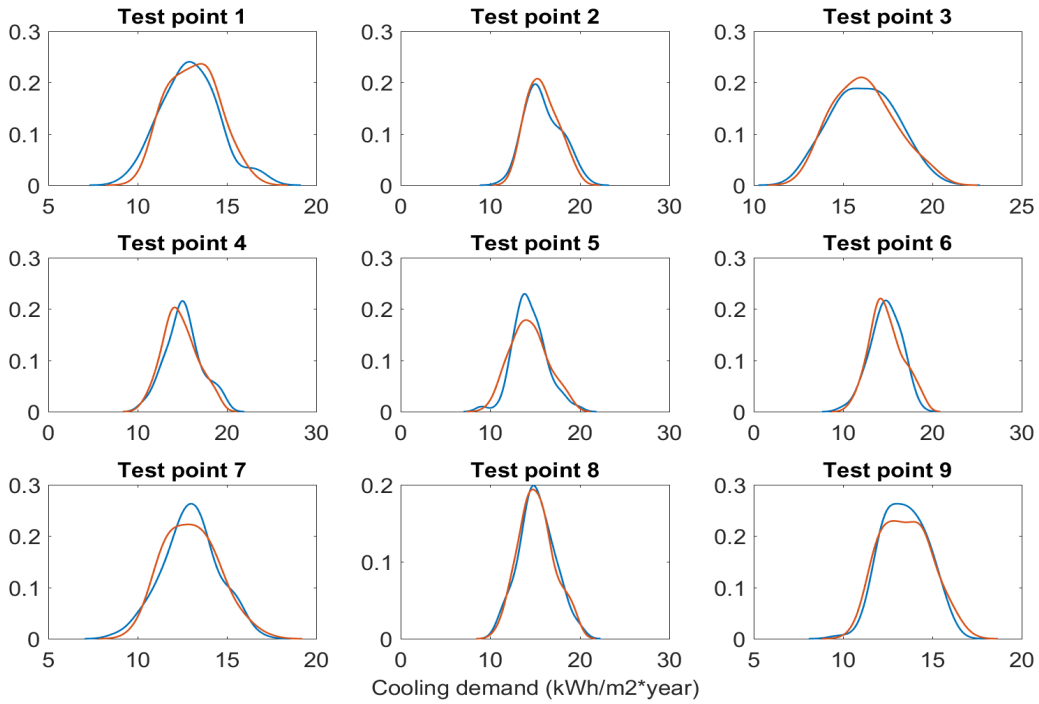


Figure 10: Uncertainty propagation comparison between GP predicted (blue) and S-BPS simulated PDFs (red) for cooling demand

Table 4: Sensitivity indices quantifying the impacts of thickness, transmittance and occupants' behaviour on the heating and cooling energy demand predictions. S_1 and S_2 are individual effects due to thickness and transmittance and S_{T_w} total effect due to occupants' behaviours

Output variable	Sample size	Sensitivity indices		
		S_1	S_2	S_{T_w}
Heating demand	Small (500)	**0.9709**	-0.0654	0.0266
	Large (20000)	**0.9715**	-0.0729	0.0266
Cooling demand	Small (500)	0.0752	0.1279	**0.829**
	Large (20000)	0.0527	0.1277	**0.8275**

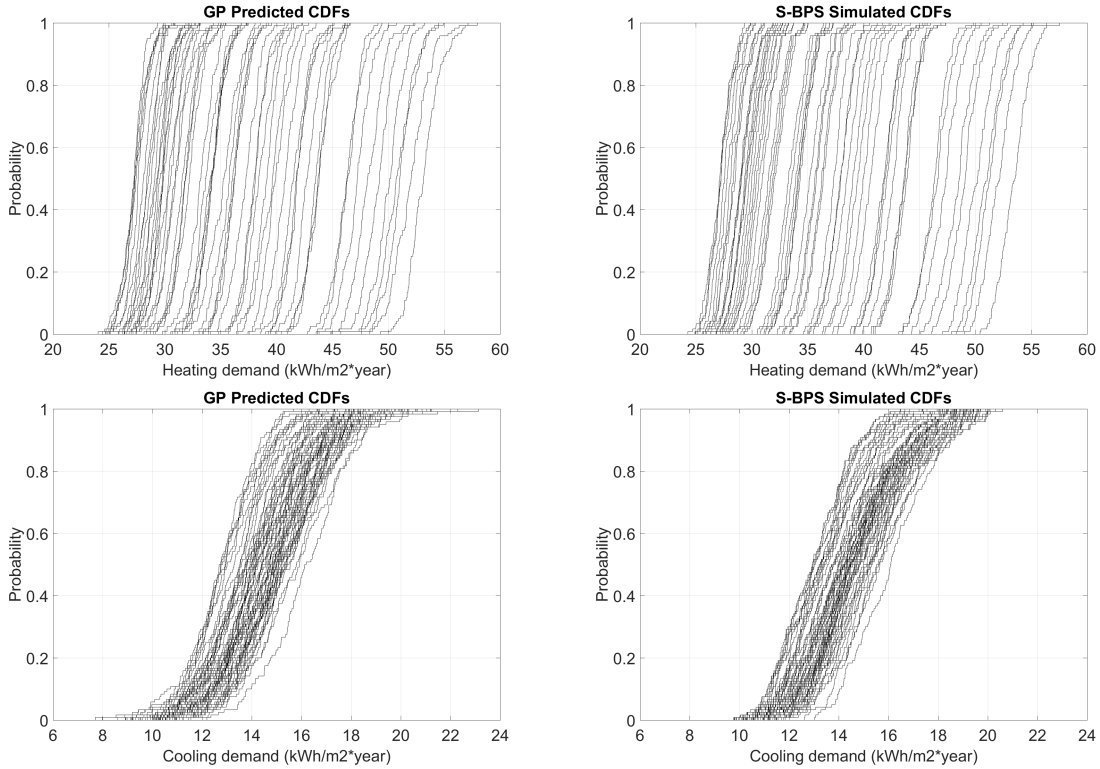


Figure 11: Comparison of GP predicted (left) and S-BPS simulated (right) CDFs for heating (top) and cooling (bottom) demand

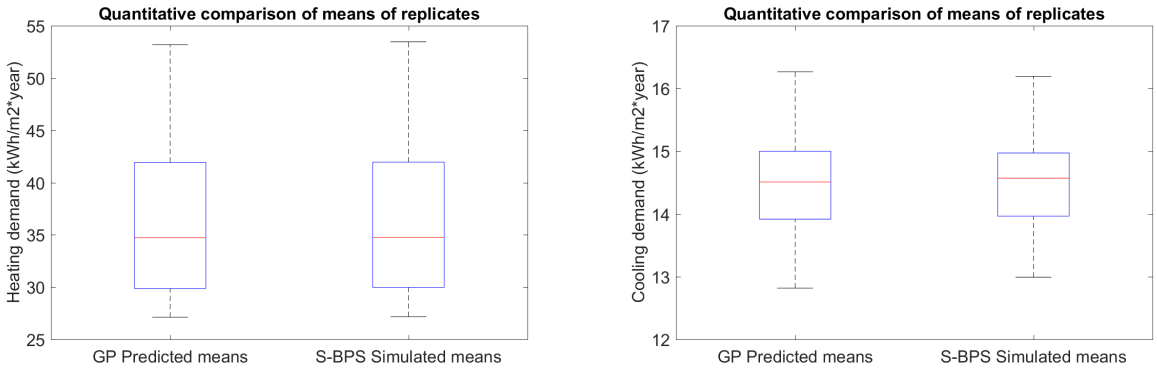


Figure 12: Comparison between S-BPS Simulated and GP predicted means of heating (left) and cooling (right) demand replicates in Figures 11

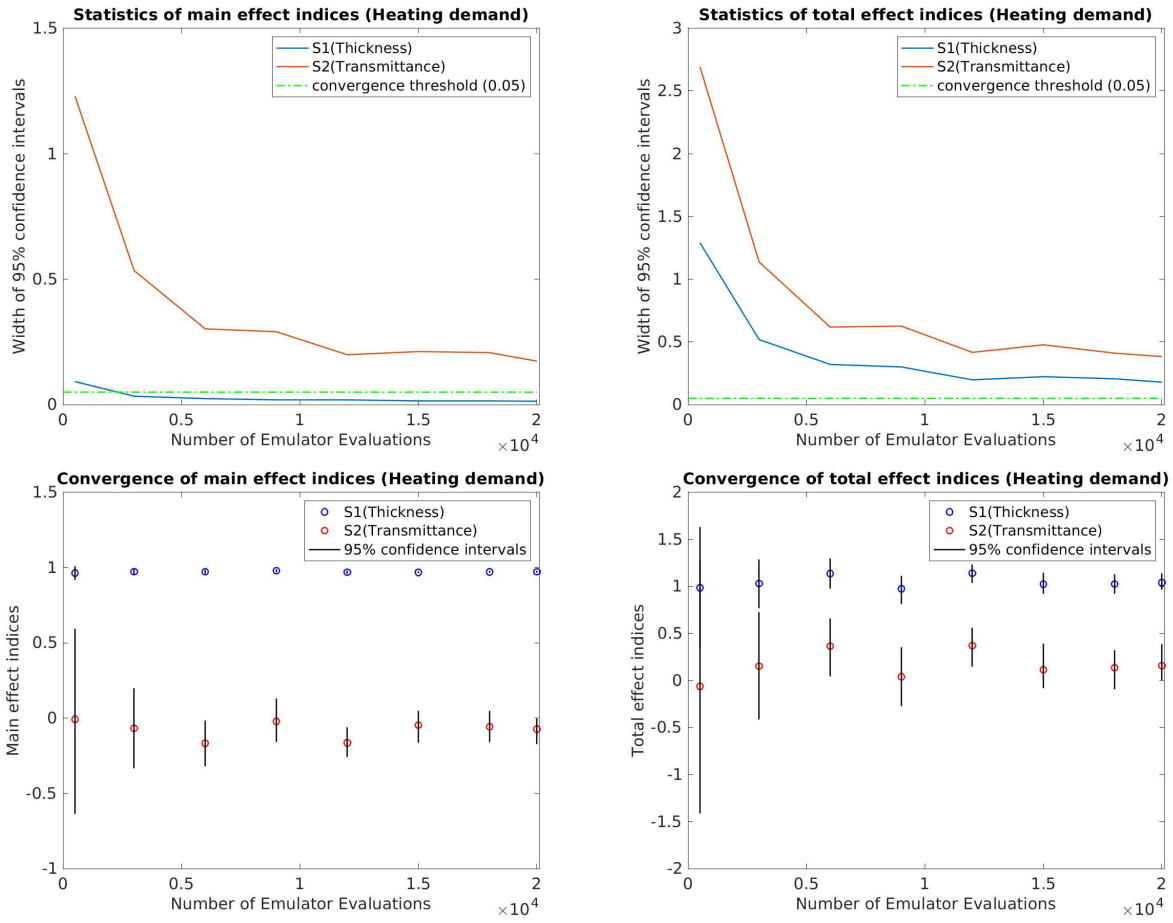


Figure 13: Convergence of main (left) and total (right) effect sensitivity indices of design variables for heating demand. The top row shows the convergence of the width of index credible interval and bottom row shows the actual credible interval.

Figures 13 and 14 (top rows) show the convergence of indices values for the emulator evaluations at different sample sizes, ranging from 500 to 20,000. We consider 0.05 as the convergence threshold for the width of the 95% credible interval $CI_{indices}$, below or close to which the index value is deemed to be stable. In the case of heating demand, for the main effect index of 'insulation thickness', the index value converges rapidly (at a small sample size of 1,000) below the threshold, while the other indices values approach the threshold (at values of around 0.2) at large sample sizes of 20,000. The bottom rows of the figures show the convergence of the actual credible intervals around indices values obtained via bootstrapping, demonstrating the trend for convergence at large sample sizes.

Now, the computational cost of a classical GSA framework is given by $N(k+2)$ [19] [70], where N and k are number of samples and inputs, respectively. For the present example with 2 input parameters and 20,000 samples (large sample sizes at which indices converge as shown), the number of simulator runs that we need in our proposed Emulation based Uncertainty and Sensitivity Analysis (EmUSA) framework is a small fraction (3,000 / 80,000 =

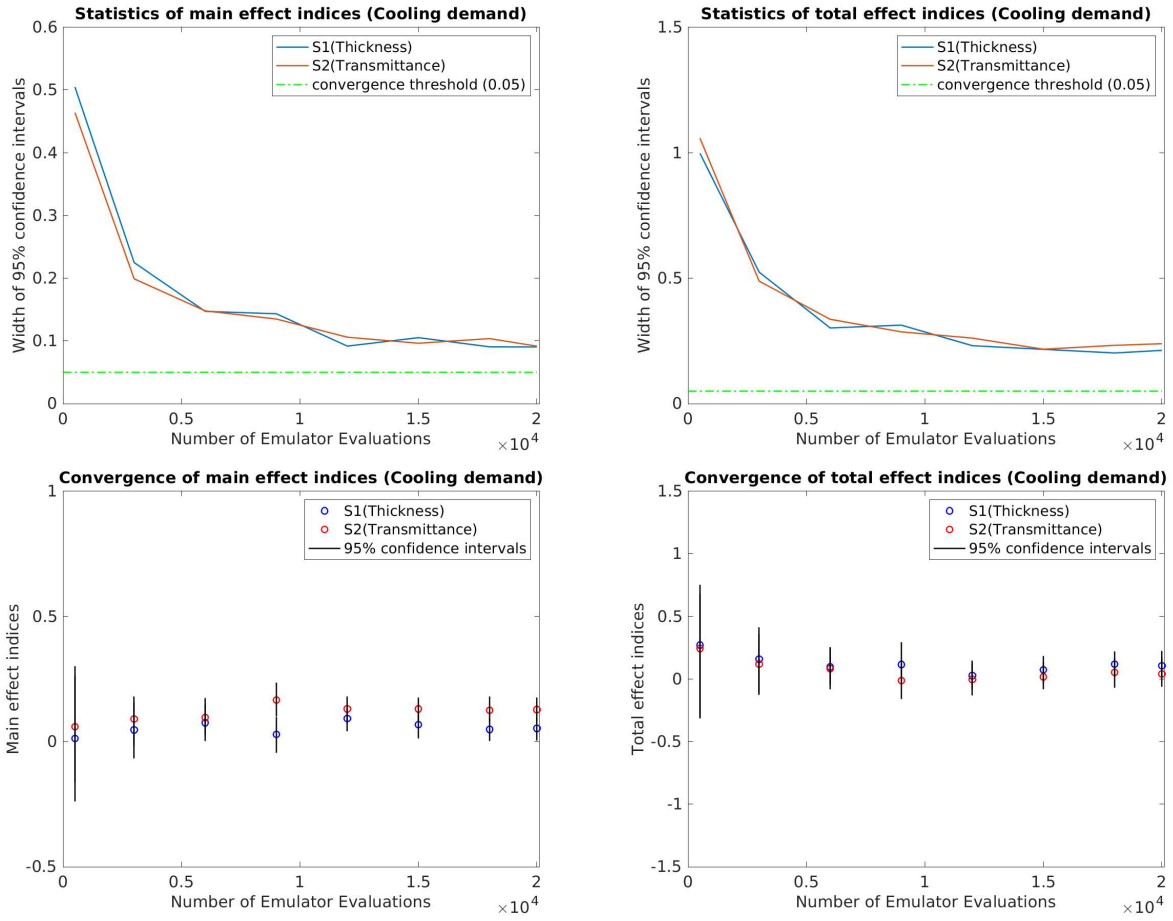


Figure 14: Convergence of main (left) and total (right) effect sensitivity indices of design variables for cooling demand. The top row shows the convergence of the width of index credible interval and bottom row shows the actual credible interval.

Table 5: Comparison of computational runs between Classical and EmUSA framework

	Classical	Emulator based	
		Training	Sensitivity Analysis
Simulator runs	80×10^3	3×10^3	-
Emulator runs	-	-	0.5 to 20×10^3

Table 6: Comparison of computational expense between Classical and EmUSA framework (One simulation run benchmark = 20 seconds)

	Classical	Emulator based	
		Training	Sensitivity Analysis
Simulator runs	18.5 days	16.67 hours	-
Emulator runs	-	-	15 to 20 minutes

0.0375) of that required by the standard approach (see Table 5), as the emulation framework only requires simulator runs to train GPs, which can then be employed to conduct sensitivity analyses at negligible computational cost. This represents a significant saving not only in terms of the number of simulator runs, but also of the computational time available to conduct of UQ study. Consequently, the computation time reduces to a manageable c.17 hours, in contrast to the infeasibly large number of days that are required by a classical approach (see Table 6). Moreover, these training runs can be readily distributed over a High Performance Computing cluster, thus benefiting from additional acceleration from computational hardware. In this way, it is possible to complete an entire UQ study using our proposed EmUSA framework in just a couple of hours.

However, for a whole multizone building, the conclusions drawn may differ from those of the current monozone office configuration; because this relatively small use case exaggerates to some extent the effects of envelope interactions and heat transfers (the effects of which would be relatively dissipated in a large multizone building). In this sense then, our small use case is a good choice for this proof-of-principle study.

8. Conclusions

The purpose of this study was to develop a computationally efficient emulator of a Stochastic Building Performance Simulator (EnergyPlus co-simulated with NoMASS), to support the quantification and decomposition of both epistemic and aleatory uncertainties. Our approach has been to estimate Gaussian Process Regression emulators for both the mean response and the variance of the stochastic component of the simulator, and to deploy these emulators to decompose the total uncertainty in predictions into its respective sources, computed in terms of sensitivity indices.

We have demonstrated the application of this new framework using a relatively simple

use case: a small monozone (shoebox) office building, computing both individual and total effect indices. The former accounts for the effect of design variables (e.g. insulation thickness or glazing transmittance), while the latter also accounts for stochastic variance arising from occupants' behaviours (e.g. opening windows and lowering shading devices). From these results we conclude that:

- The pair of fitted GP emulators successfully emulate the behaviour of the stochastic building performance simulator (S-BPS) at the chosen test design points, with predictivity coefficient $Q_2 > 0.95$, (save for a small number of moderate exceptions).
- The proposed GP Emulator-based UQ framework, requiring simulator runs only to fit the emulators, is highly computationally efficient; reducing the computational cost to a small fraction (0.0375) of its classical counterpart, as simulator runs are required only for training the emulator which subsequently computes at negligible cost.
- The variance-based uncertainty quantification and decomposition framework is an effective means for studying the combination of epistemic and aleatory uncertainties that arise from a S-BPS tool.
- The effects of uncertainties in insulation thickness on heating demand are more significant that those arising from stochasticity in occupants' behaviours; as these behaviours are relatively constrained during the heating season.
- Conversely, the effects of uncertainties arising from stochasticity in occupants' behaviours on cooling demand exceed those arising from uncertainties in effective glazing system transmittance; as these behaviours are less constrained - representing occupants' interactions with windows and shading devices to restore their thermal comfort, with corresponding impacts on cooling demand.

This work demonstrates, for the first time, the viability of an Emulation based Uncertainty and Sensitivity Analysis (EmUSA) framework for the computationally efficient quantification and decomposition of both epistemic (design variable) and aleatory (occupants' stochastic behaviours) uncertainties. So far this demonstration has been constrained in scope, in terms of the number of design variables and the range of occupants' behaviours investigated as well as the complexity of the adopted use case. We are currently extending the scope of applicability of this new framework, in terms of the design variables and occupants' behaviours that are addressed, and applying this to a more complex use case. We aim to report on the outcomes of this work in a future paper.

In terms of the wider applicability, the EmUSA framework supports, through a rigorous quantification and decomposition of uncertainties:

- Load and energy use prediction: More reliable load and energy use prediction, and the associated sizing of HVAC plant and renewable energy technologies.

- New building design: The more robust design of building envelopes and systems.
- Renovation design: The identification of parameters yielding the greatest potential performance improvements, identified through sensitivity analysis and the decomposition of uncertainties, and their corresponding design.
- Predictive control: (ongoing) Bayesian calibration of metamodels, using real measured training data, to support predictive control of building systems. In principle, this technique could also be applied to the control of larger scale heat and power grids.
- Stock modelling: metamodeling of archetypes describing regional or national building (e.g. housing, office) stocks, with which to test hypotheses for their decarbonisation, in support of building stock decarbonisation policy.

Acknowledgement

We gratefully acknowledge the European Commission for providing financial support during the conduct of research under the FP7-PEOPLE-2013 Marie Curie Initial Training Network "CI-NERGY" project with Grant Agreement Number 606851. The authors acknowledge support by the state of Baden-Württemberg through bwHPC. We are also grateful for access to the University of Nottingham High Performance Computing Facility.

Appendix A. Prior Informations used for GP emulation.

For the present work we we use a Matern covariance function (see equation 4) with parameter $\nu = 3/2$. Table A.7 shows the distributions used for the hyperparameters of these function. Table A.8 displayed the values for the hyperparameters in the mean and covariance functions. We use two distinct length scale parameter setting of covariance structure for respective two inputs - wall insulation thickness and window transmittance. For the log-variance emulators the uniform hyper-priors are used for the noise variance with small (0.001) value. The hyperpriors for the length scales and signal variances are log-Normal and uniform, respectively. We use a linear function for the mean prior mean functions for log-variance target variable. Unlike log-variance variables, we have fixed noise variances while fitting GP to mean variables as they are known following equation 21 from the predictions of GP fitted to log-variances. The length scales and signal variances both are assigned *loggaussian* priors. We use *constant*, *linear* and *squared* mean structures as prior mean functions for mean target variable.

References

[1] Joseph A Clarke, Cameron M Johnstone, Nicolas J Kelly, Paul A Strachan, and Paul Tuohy. The role of built environment energy efficiency in a sustainable UK energy economy. *Energy Policy*, 36(12):4605–4609, 2008.

Table A.7: GP prior distribution settings for hyperparameters

Function	Hyperparameter	GP target variable			
		mean		log-variance	
		Heating	Cooling	Heating	Cooling
prior covariance	signal variance	loggaussian	loggaussian	uniform	uniform
	noise variance	fixed	fixed	uniform	uniform
	length scale	loggaussian	loggaussian	loggaussian	loggaussian

Table A.8: GP prior value settings for hyperparameters

Hyperparameter	GP target variable			
	mean		log-variance	
	Heating	Cooling	Heating	Cooling
signal variance	0.1	0.1	0.01	0.1
length scale	[1/10, 1]	[1/10, 1/10]	[1/1000, 1/10]	[1/1000, 1/10]
prior mean	50	50	0.4	1
prior covariance	100	100	1	1

[2] Chris van Dronkelaar, Mark Dowson, E Burman, Catalina Spataru, and Dejan Mumovic. A review of the energy performance gap and its underlying causes in non-domestic buildings. *Frontiers in Mechanical Engineering*, 1:17, 2016.

[3] Ingrid Allard, Thomas Olofsson, and Gireesh Nair. Energy evaluation of residential buildings: Performance gap analysis incorporating uncertainties in the evaluation methods. *Building Simulation*, 11(4):725–737, 2018.

[4] Joseph Andrew Clarke and JLM Hensen. Integrated building performance simulation: Progress, prospects and requirements. *Building and Environment*, 91:294–306, 2015.

[5] Jacob Chapman, Peer-Olaf Siebers, and Darren Robinson. On the multi-agent stochastic simulation of occupants in buildings. *Journal of Building Performance Simulation*, 11(5):604–621, 2018.

[6] Esfand Burman, Dejan Mumovic, and Judit Kimpian. Towards measurement and verification of energy performance under the framework of the european directive for energy performance of buildings. *Energy*, 77:153–163, 2014.

[7] Pieter De Wilde. The gap between predicted and measured energy performance of buildings: A framework for investigation. *Automation in Construction*, 41:40–49, 2014.

[8] Jean Lebrun Philippe Andre, Bernard Georges, Vincent Lemort, and Ion Vladut Teodorese. From model validation to production of reference simulations: how to increase

reliability and applicability of building and hvac simulation models. *Building Services Engineering Research and Technology*, 29(1):61–72, 2008.

[9] Marcus Keane Paul Raftery and James O'Donnell. Calibrating whole building energy models: An evidence-based methodology. *Energy and Buildings*, 43(9):2356 – 2364, 2011.

[10] Anna Carolina Menezes, Andrew Cripps, Dino Bouchlaghem, and Richard Buswell. Predicted vs. actual energy performance of non-domestic buildings: Using post-occupancy evaluation data to reduce the performance gap. *Applied Energy*, 97:355 – 364, 2012.

[11] Frédéric Haldi and Darren Robinson. The impact of occupants' behaviour on building energy demand. *Journal of Building Performance Simulation*, 4(4):323–338, 2011.

[12] Darren Robinson and Frédéric Haldi. Modelling occupants' presence and behaviour–part i. *Journal of Building Performance Simulation*, 4:4:301–302, 2011.

[13] Houcem Eddine Mechri, Alfonso Capozzoli, and Vincenzo Corrado. Use of the anova approach for sensitive building energy design. *Applied Energy*, 87(10):3073–3083, 2010.

[14] Clara Spitz, Laurent Mora, Etienne Wurtz, and Arnaud Jay. Practical application of uncertainty analysis and sensitivity analysis on an experimental house. *Energy and Buildings*, 55:459–470, 2012.

[15] Benjamin D Lee, Yuming Sun, Godfried Augenbroe, and Christiaan JJ Paredis. Towards better prediction of building performance: a workbench to analyze uncertainty in building simulation. In *13th International Building Performance Simulation Association Conference, Chambéry, France*, 2013.

[16] Michael D McKay, Richard J Beckman, and William J Conover. Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2):239–245, 1979.

[17] Il'ya Meerovich Sobol'. On the distribution of points in a cube and the approximate evaluation of integrals. *Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki*, 7(4):784–802, 1967.

[18] Ilya M Sobol. Sensitivity estimates for nonlinear mathematical models. *Mathematical Modelling and Computational Experiments*, 1(4):407–414, 1993.

[19] Andrea Saltelli. Making best use of model evaluations to compute sensitivity indices. *Computer Physics Communications*, 145(2):280–297, 2002.

[20] A Ioannou and LCM Itard. Energy performance and comfort in residential buildings: Sensitivity for building parameters and occupancy. *Energy and Buildings*, 92:216–233, 2015.

[21] Parag Rastogi. *On the sensitivity of buildings to climate: the interaction of weather and building envelopes in determining future building energy consumption*. PhD thesis, Ecole Polytechnique Fédérale de Lausanne, 2016.

[22] Min-Yuan Cheng and Minh-Tu Cao. Accurately predicting building energy performance using evolutionary multivariate adaptive regression splines. *Applied Soft Computing*, 22:178–188, 2014.

[23] Athanasios Tsanas and Angeliki Xifara. Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools. *Energy and Buildings*, 49:560–567, 2012.

[24] Janelle S. Hygh, Joseph F. DeCarolis, David B. Hill, and S. Ranji Ranjithan. Multivariate regression as an energy assessment tool in early building design. *Building and Environment*, 57:165 – 175, 2012.

[25] Joshua Hester, Jeremy Gregory, and Randolph Kirchain. Sequential early-design guidance for residential single-family buildings using a probabilistic metamodel of energy consumption. *Energy and Buildings*, 134:202–211, 2017.

[26] Xi Chen, Hongxing Yang, and Ke Sun. Developing a meta-model for sensitivity analyses and prediction of building performance for passively designed high-rise residential buildings. *Applied energy*, 194:422–439, 2017.

[27] Matthew Riddle and Ralph T Muehleisen. A guide to bayesian calibration of building energy models. In *Building Simulation Conference*, 2014.

[28] Qi Li, Godfried Augenbroe, and Jason Brown. Assessment of linear emulators in lightweight bayesian calibration of dynamic building energy models for parameter estimation and performance prediction. *Energy and Buildings*, 124:194–202, 2016.

[29] Jun Yuan, Victor Nian, Bin Su, and Qun Meng. A simultaneous calibration and parameter ranking method for building energy models. *Applied Energy*, 206:657–666, 2017.

[30] Bryan Eisenhower, Zheng O'Neill, Satish Narayanan, Vladimir A Fonoberov, and Igor Mezić. A methodology for meta-model based optimization in building energy models. *Energy and Buildings*, 47:292–301, 2012.

[31] Liesje Van Gelder, Hans Janssen, and Staf Roels. Probabilistic design and analysis of building performances: methodology and application example. *Energy and Buildings*, 79:202–211, 2014.

[32] JA Clarke. *Energy Simulation in Building Design*. Butterworth-Heinemann, 2001.

[33] Drury B. Crawley, Linda K. Lawrie, Frederick C. Winkelmann, W.F. Buhl, Y.Joe Huang, Curtis O. Pedersen, Richard K. Strand, Richard J. Liesen, Daniel E. Fisher, Michael J. Witte, and Jason Glazer. Energyplus: creating a new-generation building energy simulation program. *Energy and Buildings*, 33(4):319 – 331, 2001.

[34] Michael Wetter and Elijah Polak. A convergent optimization method using pattern search algorithms with adaptive precision simulation. *Building services engineering research and technology*, 25(4):327–338, 2004.

[35] Bruce Ankenman, Barry L Nelson, and Jeremy Staum. Stochastic kriging for simulation metamodeling. *Operations research*, 58(2):371–382, 2010.

[36] Jarno Vanhatalo, Jaakko Riihimäki, Jouni Hartikainen, Pasi Jylänki, Ville Tolvanen, and Aki Vehtari. Gpstuff: Bayesian modeling with gaussian processes. *Journal of Machine Learning Research*, 14(Apr):1175–1179, 2013.

[37] Darren Robinson, URS Wilke, and Frédéric Haldi. Multi agent simulation of occupants' presence and behaviour. In *Proceedings of Building Simulation*, pages 2110–2117, 2011.

[38] Jacob Chapman, Peer-Olaf Siebers, and Darren Robinson. Data-scare behavioural modelling and the representation of social interactions in building simulation. *Energy*. under review.

[39] Jared Langevin, Jin Wen, and Patrick L Gurian. Including occupants in building performance simulation: integration of an agent-based occupant behavior algorithm with energy plus. In *ASHRAE/IBPSAUSA Buidling Simulation Conference. Atlanta, GA*, 2014.

[40] Tianzhen Hong, Hongsan Sun, Yixing Chen, Sarah C Taylor-Lange, and Da Yan. An occupant behavior modeling tool for co-simulation. *Energy and Buildings*, 117:272–281, 2016.

[41] Jessen Page, Darren Robinson, Nicolas Morel, and J-L Scartezzini. A generalised stochastic model for the simulation of occupant presence. *Energy and buildings*, 40(2):83–98, 2008.

[42] Frédéric Haldi and Darren Robinson. Interactions with window openings by office occupants. *Building and Environment*, 44(12):2378–2395, 2009.

[43] Frédéric Haldi and Darren Robinson. Adaptive actions on shading devices in response to local visual stimuli. *Journal of Building Performance Simulation*, 3(2):135–153, 2010.

[44] Bing Dong, Da Yan, Zhaoxuan Li, Yuan Jin, Xiaohang Feng, and Hannah Fontenot. Modeling occupancy and behavior for better building design and operation — A critical review. *Building Simulation*, 11(5):899–921, 2018.

[45] Andreas Wagner, William O'Brien, and Bing Dong. *Exploring Occupant Behavior in Buildings: Methods and Challenges*. Springer, 2017.

[46] H Burak Gunay, William O'Brien, and Ian Beausoleil-Morrison. A critical review of observation studies, modeling, and simulation of adaptive occupant behaviors in offices. *Building and Environment*, 70:31–47, 2013.

[47] Somayeh Asadi, Shideh Shams Amiri, and Mohammad Mottahedi. On the development of multi-linear regression analysis to assess energy consumption in the early stages of building design. *Energy and Buildings*, 85:246–255, 2014.

[48] Massimiliano Manfren, Niccolò Aste, and Reza Moshksar. Calibration and uncertainty analysis for computer models–a meta-model based approach for integrated building energy simulation. *Applied energy*, 103:627–641, 2013.

[49] Janelle S Hygh, Joseph F DeCarolis, David B Hill, and S Ranji Ranjithan. Multivariate regression as an energy assessment tool in early building design. *Building and Environment*, 57:165–175, 2012.

[50] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The elements of statistical learning data mining, inference, and prediction*. Springer Heidelberg, 2009.

[51] Jerome H Friedman. Multivariate adaptive regression splines. *The annals of statistics*, pages 1–67, 1991.

[52] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An introduction to statistical learning*, volume 112. Springer, 2013.

[53] Carl Edward Rasmussen and Christopher KI Williams. *Gaussian processes for machine learning*, volume 1. MIT press Cambridge, 2006.

[54] Vladimir Vapnik, Steven E Golowich, and Alex J Smola. Support vector method for function approximation, regression estimation and signal processing. In *Advances in neural information processing systems*, pages 281–287, 1997.

[55] B Yildiz, JI Bilbao, and AB Sproul. A review and analysis of regression and machine learning models on commercial building electricity load forecasting. *Renewable and Sustainable Energy Reviews*, 73:1104–1122, 2017.

[56] Kadir Amasyali and Nora M El-Gohary. A review of data-driven building energy consumption prediction studies. *Renewable and Sustainable Energy Reviews*, 81:1192–1205, 2018.

[57] Torben Østergård, Rasmus Lund Jensen, and Steffen Enersen Maagaard. A comparison of six metamodeling techniques applied to building performance simulations. *Applied Energy*, 211:89–103, 2018.

[58] Lai Wei, Wei Tian, Jian Zuo, Zhi-Yong Yang, YunLiang Liu, and Song Yang. Effects of building form on energy use for buildings in cold climate regions. *Procedia Engineering*, 146:182–189, 2016.

[59] Hyunwoo Lim and Zhiqiang John Zhai. Comprehensive evaluation of the influence of meta-models on bayesian calibration. *Energy and Buildings*, 155:66–75, 2017.

[60] Young-Jin Kim, Ki-Uhn Ahn, CS Park, and In-Han Kim. Gaussian emulator for stochastic optimal design of a double glazing system. In *Proceedings of the 13th IBPSA Conference, August*, pages 25–28, 2013.

[61] JE Yan, Young-Jin Kim, Ki-Uhn Ahn, and Cheol-Soo Park. Gaussian process emulator for optimal operation of a high rise office building. In *Proceedings of 13th International Building Performance Simulation Association Conference*, 2013.

[62] Siamak Safarzadegan Gilan and Bistra Dilkina. Sustainable building design: A challenge at the intersection of machine learning and design optimization. In *AAAI Workshop: Computational Sustainability*, 2015.

[63] Michael James Wood. *An exploration of building design and optimisation methods using Kriging meta-modelling*. PhD thesis, University of Exeter, 2016.

[64] Filippo Monari. *Sensitivity analysis and Bayesian calibration of building energy models*. PhD thesis, University of Strathclyde, 2016.

[65] Paul W. Goldberg, Christopher K. I. Williams, and Christopher Bishop. *Regression with Input-Dependent Noise: A Gaussian Process Treatment*, volume 10. MIT Press, advances in neural information processing systems edition, January 1998. Advances in Neural Information Processing Systems.

[66] Kristian Kersting, Christian Plagemann, Patrick Pfaff, and Wolfram Burgard. Most likely heteroscedastic gaussian process regression. In *Proceedings of the 24th International Conference on Machine Learning*, ICML '07, pages 393–400, New York, NY, USA, 2007. ACM.

[67] Wenjing Wang and Xi Chen. The effects of estimation of heteroscedasticity on stochastic kriging. In *Proceedings of the 2016 Winter Simulation Conference*, pages 326–337. IEEE Press, 2016.

[68] Alexis Boukouvalas and Dan Cornford. Learning heteroscedastic gaussian processes for complex datasets. *Technical report*, 2009.

[69] Amandine Marrel, Bertrand Iooss, Sébastien Da Veiga, and Mathieu Ribatet. Global sensitivity analysis of stochastic computer models with joint metamodels. *Statistics and Computing*, 22(3):833–847, 2012.

[70] Andrea Saltelli, Marco Ratto, Terry Andres, Francesca Campolongo, Jessica Cariboni, Debora Gatelli, Michaela Saisana, and Stefano Tarantola. *Global sensitivity analysis: the primer*. John Wiley & Sons, 2008.

[71] Lauwerens Kuipers and Harald Niederreiter. *Uniform distribution of sequences*. Courier Corporation, 2012.

[72] Russel E Caflisch. Monte carlo and quasi-monte carlo methods. *Acta numerica*, 7:1–49, 1998.

[73] William L Oberkampf and Christopher J Roy. *Verification and validation in scientific computing*. Cambridge University Press, 2010.

[74] Leonardo S Bastos and Anthony O'Hagan. Diagnostics for gaussian process emulators. *Technometrics*, 51(4):425–438, 2009.

[75] Jacob Chapman. *Multi-agent stochastic simulation of occupants in buildings*. PhD thesis, The University of Nottingham, 2017.

[76] David Williams. *Probability with Martingales*. Cambridge University Press, 1991.

[77] Ning Quan, Jun Yin, Szu Hui Ng, and Loo Hay Lee. Simulation optimization via kriging: a sequential search using expected improvement with computing budget constraints. *Iie Transactions*, 45(7):763–780, 2013.

[78] Andrea Saltelli, Stefano Tarantola, Francesca Campolongo, and Marco Ratto. *Sensitivity analysis in practice: a guide to assessing scientific models*. John Wiley & Sons, 2004.

[79] Toshimitsu Homma and Andrea Saltelli. Importance measures in global sensitivity analysis of nonlinear models. *Reliability Engineering & System Safety*, 52(1):1–17, 1996.

[80] Francesca Pianosi, Fanny Sarrazin, and Thorsten Wagener. A matlab toolbox for global sensitivity analysis. *Environmental Modelling & Software*, 70:80–85, 2015.

[81] Fanny Sarrazin, Francesca Pianosi, and Thorsten Wagener. Global sensitivity analysis of environmental models: convergence and validation. *Environmental Modelling & Software*, 79:135–152, 2016.

[82] GEB Archer, Andrea Saltelli, and IM Sobol. Sensitivity measures, anova-like techniques and the use of bootstrap. *Journal of Statistical Computation and Simulation*, 58(2):99–120, 1997.

[83] Wei Tian. A review of sensitivity analysis methods in building energy analysis. *Renewable and sustainable energy reviews*, 20:411–419, 2013.

[84] Iain Alexander Macdonald. *Quantifying the effects of uncertainty in building simulation*. PhD thesis, University of Strathclyde Glasgow, 2002.