Biosystems Engineering 239 (2024) 35–48

Contents lists available at ScienceDirect

Biosystems Engineering

journal homepage: www.elsevier.com/locate/issn/15375110

Research Paper

Prediction uncertainty of greenhouse electrical power and gas demand: Part 1, the role of parameter uncertainty

Henry James Payne *, Eldert van Henten, Simon van Mourik

Farm Technology Group, Wageningen University, Wageningen, the Netherlands

ARTICLE INFO

Keywords:
Greenhouse horticulture
Statistical uncertainty
Energy efficiency

ABSTRACT

Within the modern greenhouse horticultural sector energy usage is planned using mathematical models that simulate the greenhouse's future performance. These models contain parameters whose values can be inaccurate which create errors in model predictions. This reduces the effectiveness of energy management and planning done using these models. This study proposes and demonstrates an algorithm to quantify the impact of parameter errors on greenhouse gas and electric power prediction uncertainty. The proposed algorithm introduces a Polynomial Chaos Expansion as a method for the sensitivity analysis in the domain of greenhouse horticulture. Contrary to commonly used sensitivity analyses, this approach introduces the analysis of higher order interactions into the domain of greenhouse horticultural research. It was found that for both electric power and gas production the HPS lamp power rating was the most influential individual parameter. Moreover, this study found that for power demand the uncertainty in parameters relating to the lamp system were far more impactful than those related to the crop or greenhouse structure, with a respective coefficient of variation of 24 %, 5 % and 5 %. This study makes a notable and novel conclusion that for parameters related to the greenhouse structure, larger groups of parameters were responsible for prediction uncertainty through higher order interactions of second to sixth order. These results reinforce the importance of future greenhouse research considering the impact of higher order parameter interactions on prediction uncertainty using the algorithm proposed in this study.

1. Introduction

Growers in the modern greenhouse horticulture sector use computerised decision support systems to aid in the electrical power and gas buying processes. This is done by predicting the greenhouse's future electrical power and gas demand using a mathematical model of the greenhouse and forecasted weather data. However, these predictions are vulnerable to errors that are introduced through inaccuracies in the parameter values of the model.

The exact value of a parameter is often unknown. If we estimate the parameter value it is likely to contain some errors. When these parameters are used for model predictions their errors translate into errors in the predictions. As the values of these errors are unknown the true values of these parameters are uncertain, which leads to uncertainty in the predictions. In practice uncertainty in the predictions can result in the misprediction and mis-buying of power and gas, which can result in financial loss for the grower and unnecessary energy consumption.

This insight allows for the targeted improvement of model parameter and input data used in decision support tools. Any improvement in the accuracy of model parameters and input data would in turn create more accurate predictions of greenhouse electrical power and gas demand, which would lead to more efficient electrical power and gas buying by the grower. On a societal level this gained energy efficiency from greenhouse horticulture would cause a sizable decrease in the Dutch national electrical power demand, which in turn would result in less total electrical power generation, gas demand and a decrease in CO_2 emissions.

Previous research in greenhouse horticulture has included the impact of parametric uncertainty in greenhouse modelling. For example, López-Cruz, Martínez-Ruiz, Ruiz-Garcıá, and Gallardo (2020) and López-Cruz, Ruiz-García, Ramírez-Arias, and Vázquez-Peña (2013), performed a parametric uncertainty analysis on the uncertainty of a predicted greenhouse lettuce growth. Cooman and Schrevens (2006) performs a similar analysis but uses an individual Monte Carlo sample for each tomato crop model parameter. Cooman concluded that the light and CO_2 use efficiency of the crop are key parameters in propagating uncertainty into the predicted fruit dry weight. Schrevens, Jancsok, and Dieussaert (2008) assesses the impact of crop parameter uncertainty on the dehumidification and electrical power demand of a greenhouse and

* Corresponding author.
E-mail address: henry.payne@wur.nl (H.J. Payne).

https://doi.org/10.1016/j.biosystemseng.2024.01.006
Received 11 July 2023; Received in revised form 16 January 2024; Accepted 20 January 2024
Available online 13 February 2024

Nomenclature

c	Polynomial chaos expansion coefficients
CV	Coefficient of variance
d	Number of parameters
h	Hourly time instances
h_{max}	Final time index
h_0	Initial time index
i	Index of parameters being analysed
j	Index of parameters being sampled
M	Polynomial chaos expansion model
n	Index of parameter subsets
n_{max}	Number of parameter subsets
q	Latin Hypercube sampling index
q_{max}	Latin Hypercube sample size
s	Order of interaction
S	First order Sobol sensitivity indices
SI	One-by-one sensitivity indices
ST	Total order Sobol sensitivity indices
u	Uncontrolled weather input
u^R	Recorded weather
Y	Model output
α	Index of subset of polynomial chaos expansion indices
ε	Prediction error
ε^{RMS}	Prediction root mean square error from parameter variation
θ_l	Parameter lower bounds
θ	Model parameter
θ_u	Parameter upper bounds
$\overline{\theta}$	Nominal parameter vector
Ψ	Multivariate polynomials of the polynomial chaos expansion
φ	Subset of polynomial chaos expansion indices

concluded that the uncertainty in crop parameters had little effect on the uncertainty of the predicted power demand.

These studies have mainly focussed on the role of parametric uncertainty within the greenhouse crop model parameters. However, there is a knowledge gap pertaining to the effects of greenhouse climate and energy model parameters on greenhouse electrical power and gas demand prediction.

A few studies considered parametric uncertainty in the context of energy demand. For example, Golzar, Heeren, Hellweg, and Roshandel (2018) performed a sensitivity analysis on the climate setpoints and found a trade-off between crop yield and energy demand whereby large energy savings could be made but at the expense of a slightly lower crop production. Vanthoor, van Henten, Stanghellini, and de Visser (2011) performed a similar analysis on the effect of errors in the weather, greenhouse design parameters and set points on boiler energy demand and crop growth. This was done using a normalised derivative-based sensitivity index for the individual or first order effect of parameters and a meta-model-based approach for the combined effects of parameter pairs, also known as second order effects. Vanthoor's study highlighted the importance of glass PAR and FIR transmission properties as well as outdoor radiation levels for growth and energy predictions. However, in both of these cases the analysis considered the overall energy demand of the greenhouse and does not consider the impact on the constituent gas and electrical power demand that make up a greenhouse energy demand and only considered a system with a boiler. As a result, there is a clear opportunity to explore the impact of model parameter uncertainty on an operational level where gas and electrical power demand can be managed separately.

Although a number of studies have proposed methods to assess the impact of multiple sources of prediction uncertainty. These methods rely on sampling these sources which can become computationally intractable with a large number of parameters or data streams. This makes many of these methods unsuitable due to the large number of parameters associated with greenhouse models. To mitigate these issues previous studies have performed uncertainty analyses using a meta-model based approach to attribute the sources of uncertainty to a large number of parameters and the interactions between parameters (Blatman & Sudret, 2011). This was done using a polynomial chaos expansion (PCE) based meta-model, where the Sobol sensitivity indices of each parameter were analytically calculated from the coefficients of the meta-model (Mara & Becker, 2021; Sudret, 2008).

While previous parameter uncertainty analyses have been conducted in greenhouse horticulture as detailed above, in all these studies the number of analysed parameters were only a small fraction of the total number of parameters used within these models, leaving a gap for an algorithm that systematically considers a larger number of relevant parameters in an uncertainty analysis using a method such as PCE. Furthermore, the application of PCE would allow the assessment of the combined effects of groups of parameters. Additionally, while previous research has focused on predicting crop growth and energy demand, there is a gap in the literature for an uncertainty analysis that applies to the prediction of electrical power and gas demand.

To address the research gaps detailed above, this study proposes an algorithm inspired by the methodologies described in the aforementioned literature. This algorithm combines a Latin Hypercube sampling approach, parameter pre-screening that considers all model parameters, and a PCE-based sensitivity analysis to allow for an analysis of the variance in energy predictions. This PCE analysis is preferable as it efficiently attributes prediction variance to errors present in a large set of model parameter on both an individual parameter level and collectively via higher order sensitivity indices that are calculated from the PCE. The proposed uncertainty algorithms were applied on a Dutch tomato growing greenhouse use case to identify the comparative roles of different model parameters on the prediction of electrical power and gas usage.

2. Materials and methods

This study proposes and demonstrates an algorithm for the analysis of greenhouse power and gas demand prediction uncertainty that arises from parametric uncertainty. The following sections (2.1–2.2) describe the greenhouse model and weather data used to demonstrate the algorithm described in this study. The algorithm is described in section 2.3 and is then applied to three use cases, the results of which are described in section 3. It should be noted that for the remainder of this study electrical power will be referred to as power.

2.1. Greenhouse model

The greenhouse climate, tomato crop and energy model being used was Greenlight (Katzin, van Mourik, Kempkes, & van Henten, 2020), which is a calibrated, open source model. Greenlight is a dynamic differential equation-based model which emulates a tomato growing Venlo type greenhouse. The model receives input from weather data of the outside temperature, wind speed, radiation, vapour density and CO_2 concentration. The model predicts the greenhouse indoor climate states, which are the indoor air temperature, vapour concentration, ambient radiation and CO_2 concentration. In addition, Greenlight predicts the power and gas demand of the greenhouse and the growth of the tomato crop within the greenhouse. The model was parametrised for Bleiswijk in the Netherlands. This study used a rule based control scheme that is based on the current industry standard and was originally described by Vanthoor, Stanghellini, van Henten, and de Visser (2011). Due to the importance of the gas and power demand we have included a brief description of their corresponding equations. The power demand $Y_{\text{Power demand}}$ (W m^{-2}) was calculated as the product of the power rating of the HPS lamps ($\theta_{\text{lamp,max}}$) and the degree of actuation of the HPS lamps ($Y_{\text{Lamp actuation}}$), where 0 is no lighting and 1 is full lighting. Accordingly, $Y_{\text{Power demand}}$ is defined as

$$Y_{\text{Power demand}} = \theta_{\text{lamp,max}} * Y_{\text{Lamp actuation}} . \tag{1}$$

The formula for gas demand $Y_{\text{Gas demand}}$ ($\text{m}^3\ \text{s}^{-1}\ \text{m}^{-2}$) was defined as the amount of energy used by both the boiler ($Y_{\text{Boiler energy}}$) and CHP ($Y_{\text{CHP energy}}$) generator in watts per square meter, divided by the energy content per cube of gas $\theta_{\text{Gas Energy}}$ as defined by Vermeulen (2008, p. 185), where $\theta_{\text{Gas Energy}} = 32 \times 10^6$ (J m^{-3}). Accordingly

$$Y_{\text{Gas demand}} = \frac{1}{\theta_{\text{Gas Energy}}} * \left(Y_{\text{Boiler energy}} + Y_{\text{CHP energy}}\right). \tag{2}$$

2.2. Weather data

The recorded weather data used was taken from a weather recording station in Bleiswijk, the Netherlands from 2018 to 01-01 00:00 to 2019-01-01 00:00 at 5-min intervals. The recorded weather data variables are the outside temperature ($^\circ C$), wind speed (m s^{-1}), direct solar radiation (W m^2) and outside relative humidity (%). The outdoor CO_2 concentration for both the weather forecast and recordings was assumed to be constant at 410 (ppm). In addition, the cloudiness index (CI) was fixed to the average of the period (CI = 0.7) and the sky temperature and levels of diffuse radiation were estimated using the available climate variables and according to the respective methods proposed in (Luo et al., 2005) and (Orgill & Hollands, 1977). Any missing entries in the datasets were filled with the linearly interpolated values of the adjacent data points. For the purposes of demonstrating the algorithm presented in this paper in a way that is computationally tractable this study focussed on a simulation period of 2018-03-01 00:00:00 to 2018-03-15 00:00:00.

2.3. An algorithm to compute how parameter uncertainty propagates into prediction uncertainty

In this study the propagation of greenhouse model parameter uncertainty into greenhouse power and gas demand prediction uncertainty was investigated. This was done using an algorithm whose steps begin with a pre-selection of any parameters that are not relevant to this analysis. Then the distributions of each parameter that remains were defined. These parameters were then grouped into subsets of parameters that were linked by processes they are related to. For each of these subsets of parameters, sampled values were taken from each parameter distribution and were used to calculate the predicted greenhouse energy demand. This predicted energy demand was then compared with the prediction that was made with the nominal parameter values to calculate the prediction error that arose from sampling these parameters. This was then repeated until the full number of samples has been drawn. A PCE analysis was then performed using the sampled parameter values and the corresponding energy demand prediction error. These steps are then repeated for each subset of parameters.

Once this has been done the parameters that were found to be sensitive in each subset were used to form a new subset. This was done to investigate any combined effect that may exist between the most sensitive parameters of all of the parameter subsets. This new subset was then sampled and used for energy predictions and a PCE analysis in the same fashion as has been previously described. Having done this the final PCE will give a measure of contribution to energy demand prediction uncertainty from each parameter and combination of parameters. Crucially this algorithm proposes a structure way to consider all parameters within the model and arrive at a computationally tractable set of uncertainty indices.

The steps of a this algorithm are shown in Fig. 1 and applied in three use cases. In the first use case 3 subsets of model parameters and their influence on power demand prediction uncertainty is examined. In the second use case a subset of these model parameters is taken to examine its influence on greenhouse gas demand prediction uncertainty. In the third use case the two most sensitive parameters from the previous two use cases is taken and used to perform an analysis on both gas and power demand prediction uncertainty. The steps for the algorithm used in each of these use cases are described in detail in the following subsections and for a clear overview is displayed diagrammatically below in Fig. 1 in the form of a block diagram.

2.3.1. Model parameter distributions

The parameters were modelled as truncated normal distributions in which each parameter distribution (d) was defined as the product of a uniform distribution (p_1) and a normal distribution (p_2). The normal distribution (p_2) defined the statistical distribution of values for each parameter, and a uniform distribution (p_1) that sets limits to prevent extremely small and large sample values. As a result, each parameter had an associated distribution d with a mean (μ), standard deviation (σ) and an upper and lower limit (θ_l, θ_u) such that $d(\mu, \sigma, \theta_l, \theta_u) = p_1 p_2$ where:

$$p_1 \sim U[\theta_l, \theta_u]. \tag{3}$$

U was a uniform distribution with finite lower and upper bounds θ_l and θ_u that truncates a normal distribution p_2 which was defined as,

$$p_2 \sim N(\mu, \sigma). \tag{4}$$

N denotes a normal distribution with a mean set at the parameters nominal value $\overline{\theta}$ such that

$$\mu = \overline{\theta} \tag{5}$$

and 3 standard errors was set to 10 % of the mean, so the standard deviation is

$$\sigma = \frac{0.1\mu}{3}. \tag{6}$$

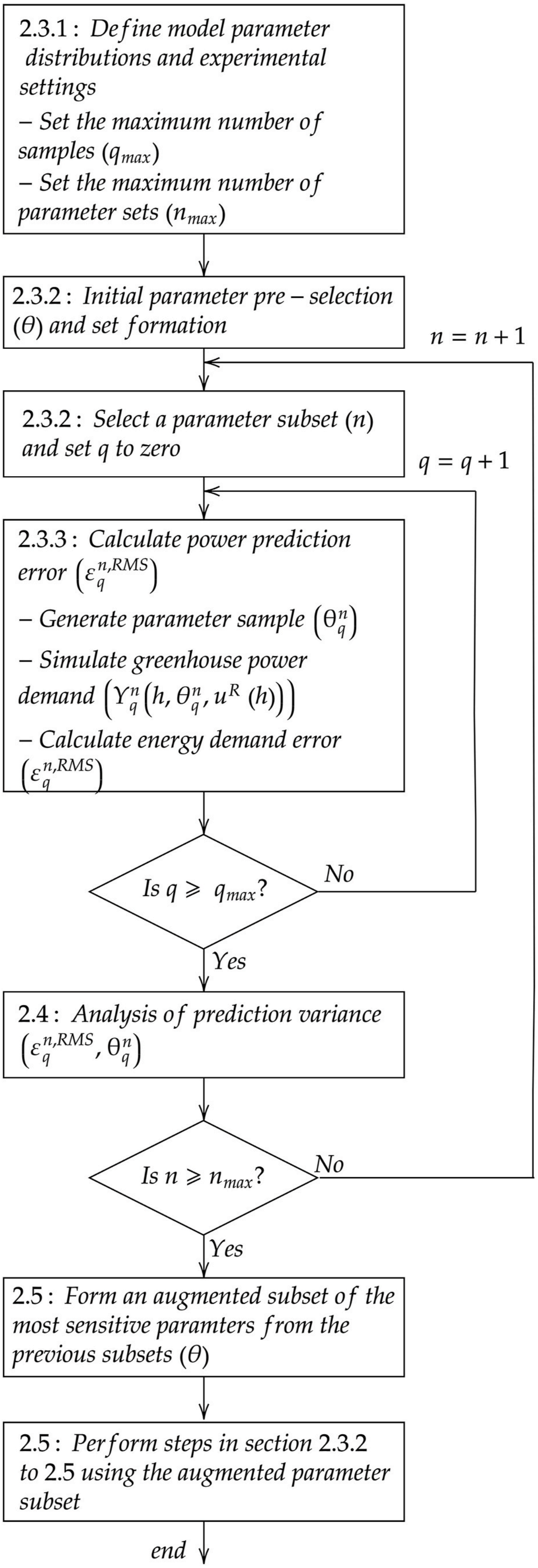
Each of the model parameter was assumed to be independently distributed from any other parameter.

2.3.2. Initial parameter pre-selection

The algorithm includes a pre-selection process from the full set of model parameters, then a metamodel based sensitivity analysis was done using a selected subset of parameters. A pre-selection process was done to reduce the number of relevant parameters as the application of uncertainty analysis methods on large models like Greenlight is computationally intensive. This is because sampling models with a large number of parameters require a large number of samples to cover all the possible combinations (Vazquez-Cruz et al., 2014). This pre-selection process was performed using a series of rules to exclude model parameters from the full sample set that were not relevant for this study. Parameters that met any one of the following criteria were discarded from the analysis:

- The model parameter is related to an unused section of the model.
- The model parameter does not contain uncertainty.
- The model parameter is a climate set point.
- The model parameter is not related to the process involved in the power generation, light physics or heating in the greenhouse.

Parameters relating to the greenhouse's power generation, light



(caption on next column)

Fig. 1. The steps taken in the model parameter uncertainty algorithm. Each block represents a step taken in the algorithm. Each step also includes the corresponding section in the text and the related variable assignation. This algorithm contains two loops. The first loop iterates through each subset of parameters (n) up to the number of parameter subsets per use case (n_{max}). The second loop iterates through the parameter sample (q) until the sample size (q_{max}) is reached.

physics and heating were selected to demarcate the study's scope and as the study only focusses on the prediction of power and gas demand. The key process contributing to power demand is the power demand of the lighting, so the processes that are related to the artificial lighting in the greenhouse were included. Furthermore, the process of indoor heating consumes a large amount of gas. For the use case analysing gas demand prediction uncertainty this study focusses on the parameters relating to the greenhouse heating system.

The parameters that were selected were then apportioned into four subsets where each subset is related to a specific operation that is simulated in the greenhouse model. This was done to highlight the sensitivity of different processes in the greenhouse as well as the parameters themselves and to reduce the computational intensity of the analysis by subdividing the parameters into relevant groups. The following subsets were used in this study:

1. Power demand and greenhouse structure related parameters
2. Power demand and HPS lamps related parameters
3. Power demand and crop related parameters
4. Gas demand and heating related parameters

These subsets of parameters were then assigned to a use case that focussed on either power or gas demand prediction. The first three parameter subsets are related to power demand prediction uncertainty and are used in the first use case. The fourth parameter subset is related to greenhouse gas demand and is used in the second use case. These two use cases are analysed and the parameters that were found to be sensitive were combined into a third use case that analyses the combined impact of these sensitive parameters on both gas and power demand prediction.

After this pre-selection process was completed, each subset of model parameters was included in an analysis of variance. This was done by drawing a Latin Hypercube (LH) sample (eq. (7)) from the selected parameters and using this to simulate the resulting prediction error when compared to predictions made with nominal parameter values. For simplicity the mean value of each parameter was set as its nominal value as in eq. (5). For the remainder of the study, steps in the algorithm that can be interchangeably applied to the analysis of both gas and power demand will be referred to using the collective term energy in place of either power or gas.

2.3.3. *Calculation of energy demand prediction error*

Initially a subset of parameters (θ_q^n) was selected using the index n where $n = 1..n_{max}$. This subset of parameters was then sampled, where q is the index of the sample, this sample was taken from the distributions D of each parameter where $D(j, \mu, \sigma, \theta_l, \theta_u)$ where j is the index of each parameter. These samples were taken using a Latin Hypercube sampling method, resulting in

$$\theta_q^n(j) \sim D(\mu(j), \sigma(j), \theta_l(j), \theta_u(j)) \text{ for } q = 1..q_{max}, n = 1..n_{max} \text{ and } j = 1..d, \tag{7}$$

where d is the number of parameters. Time was discretised to hourly time instances h with $h = h_0..h_{max}$ where h_0 is the starting index and h_{max} is the final index. The predicted energy demand of the greenhouse was defined as Y, where Y is a function of the initial time h_0, the time step of the simulation h, the parameter values of the model θ and the uncontrolled weather input to the model u. Each set of sampled parameter values θ^n and the recorded weather u^R was used to run the model and predict the energy demand $Y(h_0, h, \theta, u^R(h))$. The nominal predicted energy demand $Y(h_0, h, \overline{\theta}, u^R(h))$ was used as a base reference and was calculated using the nominal parameter values $\overline{\theta}$ (Table 3). Subsequently the prediction error ε_q^n and its root mean squared error were then calculated for each parameter sample q where:

$$\varepsilon_q^n\left(n, q, h_0, h, \theta_q^n, u^R(h)\right) = Y\left(h_0, h, \theta_q^n, u^R(h)\right) - Y\left(h_0, h, \overline{\theta}, u^R(h)\right) \text{ and } \tag{8}$$

$$\varepsilon_q^{n,RMS}\left(n, q, h_0, \theta_q^n, u^R\right) = \sqrt{\frac{1}{h_{max} - h_0 + 1} \sum_{h=h_0}^{h_{max}} \left(\varepsilon_q^n\left(n, q, h_0, h, \theta_q^n, u^R(h)\right)\right)^2}. \tag{9}$$

For each parameter subset n, θ^n is the set of samples for all the parameters in the subset and the corresponding set of root mean squared error ($\varepsilon^{n,RMS}$) are then used to perform an analysis of the prediction variance.

2.4. Analysis of prediction variance

The analysis of variance within this study was performed using a Polynomial Chaos Expansion based sensitivity analysis and a coefficient of variation-based uncertainty analysis which are detailed below.

2.4.1. Polynomial chaos expansion based global sensitivity analysis

Due to the large number of greenhouse model parameters, a sampling-based approach to a global sensitivity analysis would require a large number of samples to accurately assess the prediction variation resulting from all possible combinations of parameter values. For example, a uniform discretisation of a parameter space of N points per parameter would need N^P samples, with P being the number of parameters. Given this, if hypothetically $N = 10$ and $P = 10$ the number of samples required is 10^{10}. This requirement means that a sampling-based approach is computationally intensive and ultimately computational intractable.

To avoids these issues and conduct a sensitivity analysis on the large number of model parameters, a polynomial chaos expansion (PCE) based sensitivity analysis was conducted. A PCE is a form of meta-model and allows for greater computational efficiency when compared to a conventional sampling-based analysis. This is as parameter samples are used to calibrate the meta-model and the Sobol sensitivity indices can be calculated analytically from the coefficients of the meta-model (Sudret, 2008).

For the algorithm described in section 2.3, the variance in the energy demand prediction error was decomposed and attributed to the respective input parameters used in each parameter subset. Sobol 1st,2nd and total order indices (Archer, Saltelli, & Sobol, 1997; Sobol, 1993) were used as the metric of variance in the variance decomposition.

The PCE used in this study for each subset n was in the form of a deterministic model M and can be described as a polynomial. This model was calibrated using $\varepsilon^{n,RMS}$ and the corresponding parameters samples $\theta_\bullet^n$. Accordingly, this model described the relationship between a set of model parameters $\theta = [\theta^n(1) \ldots \theta^n(d)]$, and the approximated energy demand error $\widetilde{\varepsilon}^{n,RMS}$ such that

$$\widetilde{\varepsilon}^{n,RMS} = M^n(\theta). \tag{10}$$

This PCE is formed of a series of multivariate polynomials (Ψ) and coefficients (c) whose basis functions are based on Hermite polynomials. Each univariate component of the PCE is considered to be orthogonal each other. The PCE terms are described using the index i_s where $s = 1, \ldots, d$, and are used to group terms that represent every possible combination of parameters. The PCE can subsequently be described as a series of summations that collect the terms relating to the impact and interactions of model parameters θ. These summation terms describe the collection of PCE terms relating to the impacts of groups of parameters of different sizes such that,

$$M(\theta) = c_0 + \sum_{i=1}^{d} \sum_{\alpha \in \varphi_i} c_\alpha \Psi_\alpha(\theta_i) \tag{11}$$

$$+ \sum_{1 \le i_1 < i_2 \le d} \sum_{\alpha \in \varphi_{i_1, i_2}} c_\alpha \Psi_\alpha(\theta_{i_1}, \theta_{i_2}) + \ldots$$

$$+ \sum_{1 \le i_1 < \ldots < i_s \le d} \sum_{\alpha \in \varphi_{i_1, \ldots, i_s}} c_\alpha \Psi_\alpha(\theta_{i_1}, .., \theta_{i_s}) + \ldots$$

$$+ \sum_{\alpha \in \varphi_{1,2,\ldots,d}} c_\alpha \Psi_\alpha(\theta_1, .., \theta_d).$$

here $\alpha \in \mathbb{N}^d$ is a d-dimensional index representing the entire parameter space. The terms of the PCE are collected into groups using a subset φ_i which denoted a subset of terms related to parameters which are defined in its subscript i. This series of summations collect terms that relate to the mean (c_0) and the impact of groups of parameters, where the second term denotes the first order impact of parameters (i_1). The third term relates to the second order impact of any pair of parameters whose index is described using i_1 and i_2. The fourth term relates to the higher order impact of any number of parameters ($i_1, \ldots, i_s$) and the fifth term is the total order impact of all the parameters where the index is from 1 to d.

The polynomials used in this PCE were constructed using Hermite polynomials, whose coefficients (c) were calculated with the sparse-favouring least-square minimization least angle regression (LARS) method (Efron, Hastie, Johnstone, & Tibshirani, 2004). Using LARS was shown to be advantageous as it greatly improved overall computational efficiency by using an iterative method to only identify the PCE coefficients relating to the impactful parameters (Blatman & Sudret, 2011).

The method used in this study also included a degree-adaptive calculation of the order of the polynomials as part of the meta-model calibration process. The degree range is set from 1 to 10°. This method iteratively increased the degree of the PCE polynomials, assessed the a-posteriori cross-validation error using a leave-one-out error metric and selecting the degree of polynomial that has the lowest error. This study used a proposal range of one to ten degrees within which the optimal polynomial degree was found.

Table 2
Use cases and subsets of accepted parameters.

	Greenhouse power demand uncertainty use case			Greenhouse gas demand uncertainty use case
	Power demand and structure related parameters	Power demand and HPS lamps related parameters	Power demand and crop related parameters	Gas demand and heating related parameters
Count	11	11	4	10

Table 1
Pre-selection of accepted parameters.

	Full set	In active module	Uncertain constants	Not climate setpoint	Associated with power & Lighting	Associated with heating	Accepted
Count	242	222	223	210	43	20	36

Table 3
Definition of model parameter distributions.

Power demand and greenhouse structure related parameter subset					
Parameter name (θ)	Distribution range (θ_l, θ_u)	Distribution mean (μ)	Distribution standard error (σ)	Units	Mean value reference
Ratio of global radiation absorbed by the greenhouse construction ($\theta_{Rad,const}$).	$[0,\infty)$	0.1	3.3×10^{-3}	–	Vanthoor, Stanghellini, et al. (2011)
NIR reflection coefficient of the roof ($\theta_{NIR,ref,roof}$).	$[0,1]$	0.13	4.3×10^{-3}	–	Vanthoor, Stanghellini, et al. (2011)
PAR reflection coefficient of the roof ($\theta_{PAR,ref,roof}$).	$[0,1]$	0.13	4.3×10^{-3}	–	Vanthoor, Stanghellini, et al. (2011)
NIR transmission coefficient of the roof ($\theta_{NIR,trans,roof}$).	$[0,1]$	0.85	0.028	–	Vanthoor, Stanghellini, et al. (2011)
PAR transmission coefficient of the roof ($\theta_{PAR,trans,roof}$).	$[0,1]$	0.85	0.028	–	Vanthoor, Stanghellini, et al. (2011)
NIR reflection coefficient of thermal screen ($\theta_{NIR,ref,them}$).	$[0,1]$	0.35	0.012	–	Vanthoor, Stanghellini, et al. (2011)
PAR reflection coefficient of thermal screen ($\theta_{PAR,ref,them}$).	$[0,1]$	0.35	0.012	–	Vanthoor, Stanghellini, et al. (2011)
NIR transmission coefficient of thermal screen ($\theta_{NIR,trams,them}$).	$[0,1]$	0.6	0.02	–	Vanthoor, Stanghellini, et al. (2011)
PAR transmission coefficient of thermal screen ($\theta_{PAR,trams,them}$).	$[0,1]$	0.6	0.02	–	Vanthoor, Stanghellini, et al. (2011)
NIR reflection coefficient of the floor ($\theta_{NIR,ref,floor}$).	$[0,1]$	0.5	0.017	–	Vanthoor, Stanghellini, et al. (2011)
PAR reflection coefficient of the floor ($\theta_{PAR,ref,floor}$).	$[0,1]$	0.65	0.022	–	Vanthoor, Stanghellini, et al. (2011)

Power demand and HPS lamps related parameter subset					
Parameter name (θ)	Distribution range (θ_l, θ_u)	Distribution mean (μ)	Distribution standard error (σ)	Units	Mean value reference
Maximum intensity of lamps ($\theta_{lamp,max}$).	$[0,\infty)$	110	3.70	Wm^{-2}	Katzin et al. (2020)
Fraction of lamp input converted to PAR ($\theta_{PAR,frac,lamp}$).	$[0,1]$	0.37	0.012	–	Nelson and Bugbee (2014)
Fraction of lamp input converted to NIR ($\theta_{NIR,frac,lamp}$).	$[0,1]$	0.22	7.3×10^{-3}	–	Nelson and Bugbee (2015)
Transmissivity of lamp layer to PAR ($\theta_{PAR,trans,lamp}$).	$[0,1]$	0.98	0.033	–	de Zwart, Baeza, van Breugel, Mohammadkhani, and Janssen (2017)
Reflectivity of lamp layer to PAR ($\theta_{PAR,ref,lamp}$).	$[0,1]$	0	0.1	–	de Zwart et al. (2017)
Transmissivity of lamp layer to NIR ($\theta_{NIR,trans,lamp}$).	$[0,1]$	0.98	0.033	–	de Zwart et al. (2017)
Reflectivity of lamp layer to NIR ($\theta_{NIR,ref,lamp}$).	$[0,1]$	0	0.1	–	Katzin et al. (2020)
Lamp area ($\theta_{areas,lamp}$).	$[0,\infty)$	0.02	6.7×10^{-4}	m^2(lamp) m^{-2}(floor)	de Zwart et al. (2017)
Emissivity of topside of lamp ($\theta_{emis,top,lamp}$).	$[0,1]$	0.1	3.3×10^{-3}	–	Katzin et al. (2020)
Emissivity of bottom side of lamp ($\theta_{emis,bot,lamp}$).	$[0,1]$	0.9	0.03	–	Katzin et al. (2020)
Joules to micromole conversion of PAR output of lamp ($\theta_{PAR,con,lamp,}$).	$[0,\infty)$	4.9	0.16	μmol(PAR) J^{-1}	Nelson and Bugbee (2015)

Power demand and crop related parameter subset					
Parameter name (θ)	Distribution range (θ_l, θ_u)	Distribution mean (μ)	Distribution standard error (σ)	Units	Mean value reference
PAR extinction coefficient of the canopy ($\theta_{PAR,can}$).	$[0,1]$	0.7	0.023	–	Vanthoor, Stanghellini, et al. (2011)
PAR extinction coefficient of the canopy for light reflected from the floor ($\theta_{PAR,floor}$).	$[0,1]$	0.7	0.023	–	Vanthoor, Stanghellini, et al. (2011)
NIR extinction coefficient of the canopy ($\theta_{NIR,can}$).	$[0,1]$	0.27	0.0091	–	Vanthoor, Stanghellini, et al. (2011)

(continued on next page)

Table 3 (*continued*)

Power demand and crop related parameter subset					
Parameter name (θ)	Distribution range (θ_l, θ_u)	Distribution mean (μ)	Distribution standard error (σ)	Units	Mean value reference
Maximum capacity of the crop buffer $(\theta_{\text{buf,max}})$.	$[0, \infty)$	20×10^3	670	–	Vanthoor, Stanghellini, et al. (2011)

The Sobol sensitivity indices were derived from the coefficients of the meta-model (c_α) as described in (Sudret, 2008). The first order Sobol indices (S_i) for each parameter were defined as the fraction of prediction variance that can be attributed exclusively to a single parameter (θ_i) over the total variance in the PCE. The total variance of the PCE prediction was calculated as

$$Var[M(\theta)] = Var\left[\sum_{\alpha=0}^{d-1} c_\alpha \Psi_\alpha\right] = \sum_{\alpha=1}^{d-1} c_\alpha^2 E\left[\Psi_\alpha^2(\theta)\right]. \tag{12}$$

Subsequently the first order sensitivity (S_i) for each parameter was calculated as

$$S_i = \frac{Var[M(\theta_i)]}{Var[M(\theta)]} = \frac{\sum_{\alpha \in A_{\Theta_i}} c_\alpha^2 E\left[\Psi_\alpha^2(\theta_i)\right]}{Var[M(\theta)]}, A_{\theta_i} = \left\{\alpha \epsilon \mathbb{N}^d \middle| \alpha_k \neq 0 \Longleftrightarrow k \in \theta_i\right\} \tag{13}$$

in which the numerator of the first order Sobol indices was defined as the square sum of a subset (A_{θ_i}) of the non-zero terms which include coefficients (c_α) and bases (Ψ_α) relating exclusively to parameter θ_i. The total order Sobol indices (ST_i) were defined as

$$ST_i = \sum_{i \subset \{i_1, \ldots, i_s\}} S_{i_1, \ldots, i_s}. \tag{14}$$

where the total order sensitivity indices is the sum of a subset of sensitivity coefficients (S_i) that relate exclusively to parameter θ_i or having any interactions with θ_i at any order of basis function in the polynomial.

2.4.2. Coefficients of variation

In addition to the Sobol indices, the coefficients of variation (CV) were used to compare the variability of the energy prediction uncertainty created by each subset of parameters, where

$$CV\left(n, \varepsilon_{\cdot}^{n,RMS}\right) = \frac{\sigma\left(\varepsilon_{\cdot}^{n,RMS}\right)}{\mu\left(\varepsilon_{\cdot}^{n,RMS}\right)}. \tag{15}$$

2.5. Combined parameter subset

Following the analysis of all four parameter subsets, the two most sensitive parameters found in each subset were then combined into a new subset and analysed using the steps describes in section 2.3.3 and 2.4.1. For this subset of parameters both predicted electrical power and gas demand were considered.

3. Results

The algorithm proposed in this study is now demonstrated in three use cases using a model that describes a Dutch tomato producing greenhouse. The outcomes of the use cases below demonstrate in which areas uncertainty reduction can be most effectively applied to ensure accurate greenhouse power and gas demand prediction. According to the algorithm described in section 2.3, initially the full set of 242 parameters in Greenlight were reduced to 36 viable parameters (Table 1) using the pre-selection criteria. The full set of parameter descriptions can be found in (Katzin et al., 2020) and the full pre-selection process is detailed in the supplementary material. This document details how the pre-section criteria were applied to the complete dataset and how the subset of accepted parameters was reached.

The 36 parameter that were accepted in the pre-selection were then apportioned into four subsets and assigned to two use cases relating either to power or gas demand prediction uncertainty. The number of accepted parameters per subset and their associated use case can be seen in Table 2.

The following subsections address each use case in turn focussing first on the power demand uncertainty, then gas demand uncertainty and then the combined gas and power demand uncertainty.

3.1. Computational settings

For the application of the algorithm proposed in this study the following settings are used,

$n_{max} = 4, q_{max} = 1000$ and $h_{max} = 4032$.

3.2. Greenhouse power demand uncertainty use case

The parameters used for the greenhouse power demand uncertainty use case are described below in Table 3. This table details which parameter subset each parameter is assigned to, each parameter mean, standard error (as defined in section 2.3.1), range of possible values and literary reference.

In accordance with the algorithm set out in section 2.3 each subset of parameters had 1000 samples drawn. These sampled parameter values of each subset were used to predict the greenhouse power demand for the purpose of an uncertainty analysis. The results from each subset of parameters are described below.

3.2.1. Results for the power demand uncertainty analysis use case using the greenhouse structure related parameter subset

To assess the sensitivity of the predicted greenhouse power demand a PCE based sensitivity analysis was performed using the model parameter sample from the greenhouse structure related parameter subset and the corresponding greenhouse power demand prediction error. The predicted power demand error was found to have a CV of 5.1 %. The corresponding first order and total sensitivity indices are presented in Fig. 2.

Figure 2 shows that for the structure related parameter subset all parameters do impact the variation in the prediction via the total order indices. However, none of the included parameters are found to have a first order effect, meaning no single parameter was found to be

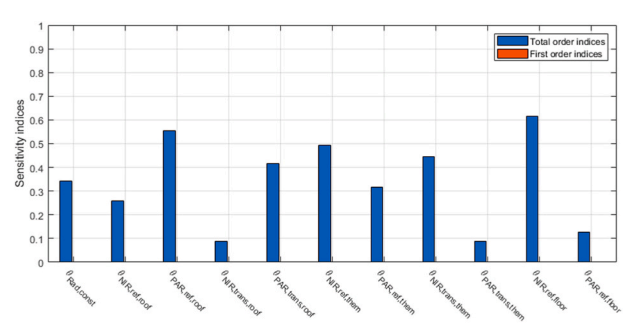


Fig. 2. First and total order sensitivity indices from the power demand uncertainty analysis use case using the greenhouse structure related parameter subset.

individually responsible for variation the predictions. Instead, this PCE analysis predicts that the variation in the predictions are only attributed to larger groups of parameters via higher order interactions of second to sixth order which implies errors amplify as they interact dynamically. The indices for these higher order effects can be seen in Table 4. This insight along with the low coefficient of variation indicate that while these parameters do produce a small amount of prediction variation no one parameters is notably impactful.

The two parameters with the greatest total order indices (Fig. 2) are $\theta_{NIR,ref,floor}$ and $\theta_{PAR,ref,roof}$. The most sensitive parameter $\theta_{NIR,ref,floor}$ is used to calculate how much radiation is reflected from the floor. Subsequently the amount of radiative energy absorbed by the floor was calculated and then how the floor temperature changes. This change in floor and crop canopy temperature influences the air temperature via latent heat exchange. The air temperature is used to control the lamps, which in turn affects the power demand of the greenhouse. The second most sensitive parameter $\theta_{PAR,ref,roof}$ is the reflection coefficient of the glass. $\theta_{PAR,ref,roof}$ determines how much radiation is being reflected and transmitted through the glass, cover and blackout screen. Then $\theta_{PAR,ref,roof}$ influences how much radiation reaches the thermal screen, top compartment, pipes and floor. As such these states describe how much heat from the sun is transferred to the aforementioned components and then to the indoor air. The air temperature is then used to control the lamps and therefore influences the power demand.

The analysis of this subset reveals that parametric uncertainty propagates into the indoor air temperature state through the absorption and transmission of radiative heat by the structure. The indoor temperature then influences the temperature-based lamp lighting rule set, which in turn affects the power demand. Overall, it can be concluded that uncertainty in the parameters related to the structure has a small net impact on prediction uncertainty. However, the design of the controller, particularly the air temperature-based rule, allows the propagation of uncertainty into greenhouse power demand prediction.

Table 4
Higher order sensitivity indices from the power demand uncertainty analysis use case using the greenhouse structure related parameter subset.

Parameter names (θ)	Second order indices
$\theta_{PAR,ref,roof} * \theta_{PAR,ref,them}$	0.12
$\theta_{NIR,ref,them} * \theta_{PAR,ref,floor}$	1.6×10^{-2}
	Third order indices
$\theta_{Rad,const} * \theta_{NIR,ref,them} * \theta_{PAR,ref,them}$	0.14
$\theta_{NIR,trans,roof} * \theta_{PAR,trams,them} * \theta_{NIR,trams,them}$	9.0×10^{-2}
	Fourth order indices
$\theta_{Rad,const} * \theta_{PAR,ref,roof} * \theta_{NIR,ref,them} * \theta_{NIR,ref,floor}$	0.14
	Fifth order indices
$\theta_{PAR,ref,roof} * \theta_{PAR,ref,floor} * \theta_{PAR,trans,roof} * \theta_{NIR,trams,them} * \theta_{NIR,ref,floor}$	0.24
	Sixth order indices
$\theta_{Rad,const} * \theta_{PAR,ref,roof} * \theta_{PAR,trans,roof} * \theta_{PAR,ref,them} * \theta_{NIR,trams,them} * \theta_{NIR,ref,floor}$	3.7×10^{-2}

3.2.2. *Results for the power demand uncertainty analysis use case using the HPS lamps related parameter subset*

Performing the analysis using the HPS lamps related parameter subset and the corresponding predicted power demand produced a PCE, the power demand predictions used to calculate this PCE had a CV of 24 %. The resulting first and total order sensitivities can be seen in Fig. 3.

For the HPS lamps related parameter subset the parameter relating to the maximum intensity of the HPS lamps ($\theta_{lamp,max}$) was by far the most impactful on power demand prediction uncertainty, accounting for nearly all of the variation in the predicted power demand. This result is understandable as the power demand of the greenhouse is almost entirely from operating the lamps, and by changing their power rating the total demand changes. In addition, the power rating of the lamps influences the amount of heat energy the lamps transfer into the greenhouse. As such $\theta_{lamp,max}$ also influences the air temperature and consequently the control dynamics. The parameter with the second largest impact was $\theta_{emis,top,lamp}$. This parameter is related to what fraction of the radiation from the lamps is emitted above the lamps. This radiation interacts with the greenhouse screen, cover and is transmitted into the sky, in doing so influencing lamp, screen and indoor air temperature. A change in air temperature influences the temperature-based rules controlling the lamps and subsequently the power demand.

While other parameters are found to have some impact, these impacts are small in comparison to the role of $\theta_{lamp,max}$. The PCE did predict a number of second order effects in which a pair of parameters was found to have a role in creating prediction uncertainty. The 5 largest second order indices can be seen in Table 5.

The second order sensitivities shown in Table 5 are all of small orders of magnitude when compared to the other sensitivity indices that range between 0 and 1, the largest of which is the combined influence of the $\theta_{lamp,max}$ and the emissivity of the top of the lamp $\theta_{emis,top,lamp}$. This combined effect is logical as the maximum intensity of the lamp influences the amount of radiation that can be transmitted upwards. These combined effects influence the air temperature and as previously described propagate into the control dynamics and the power demand. The remaining second order sensitivity indices reflect the combine impacts of the upper $\theta_{emis,top,lamp}$ and lower $\theta_{emis,bot,lamp}$ emissivity of the lamps and how the lamp radiation is transmitted and reflected from the cover.

3.2.3. *Results for the power demand uncertainty analysis use case using the crop related parameter subset*

The third parameter subset in the power demand use case was related to the parameters used in the crop model. The power demand predictions used for this PCE had a CV of 5.1 %. The first and total order sensitivity indices for this subset are displayed below in Fig. 4.

The analysis performed on the crop related parameter subset show

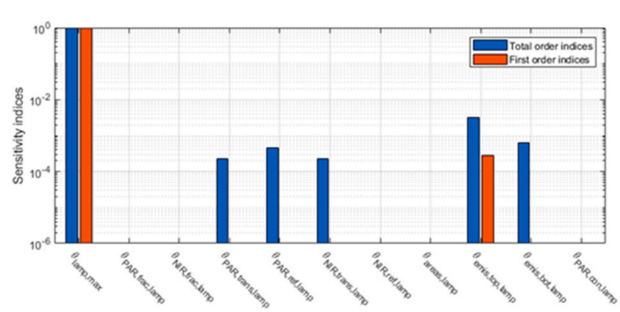


Fig. 3. First and total order sensitivity indices from the power demand uncertainty analysis use case using the HPS lamps related parameter subset on a logarithmic scale.

Table 5
Second order sensitivity indices from the power demand uncertainty analysis use case using the HPS lamps related parameter subset.

Parameter names (θ)	Second order indices
$\theta_{lamp,max} * \theta_{emis,top,lamp}$	2.4×10^{-3}
$\theta_{emis,top,lamp} * \theta_{emis,bot,lamp}$	3.8×10^{-4}
$\theta_{PAR,ref,lamp} * \theta_{emis,bot,lamp}$	2.6×10^{-4}
$\theta_{PAR,trans,lamp} * \theta_{NIR,trans,lamp}$	2.3×10^{-4}
$\theta_{PAR,ref,lamp} * \theta_{emis,top,lamp}$	2.0×10^{-4}

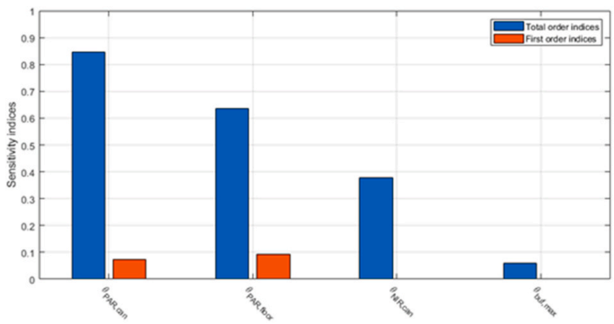


Fig. 4. First and total order sensitivity indices from the power demand uncertainty analysis use case using the crop related parameter subset.

that parameters in the crop model do influence the predicted power demand uncertainty and that the magnitude of this influence is comparatively small given the low value of the coefficient of variance. This outcome highlights the impact of the parameters relating to the PAR extinction coefficient of the crop's canopy from above ($\theta_{PAR,can}$) and below ($\theta_{PAR,floor}$). The parameter $\theta_{PAR,can}$ is used to calculate how much downwards lamp radiation is absorbed by the crop canopy. This is then used to calculate the amount of heat that the incoming radiation contributes to the air temperature through the canopy temperature, which as in the previous cases influences the control of the lamps. The parameter $\theta_{PAR,floor}$ is used to calculate how much lamp radiation is absorbed by the crop that is reflected from the floor and how this affects the temperature of the air. The analysis also provided 2 s order indices that are displayed below in Table 6.

The second order indices in Table 6 show that the largest second order interaction that the PCE defined was between $\boldsymbol{\theta_{PAR,can}}$ and $\boldsymbol{\theta_{PAR,floor}}$. These two parameters having a combined effect is logical as both influence the absorption of radiation by the crop canopy from above and below. The next second order sensitivity indices highlights the combined impact of the lamp PAR and NIR radiation that is absorbed by the crop canopy. These parameters both influence the amount of heat the lamps transmit to the indoor air and thereby influence the control of the lamps.

The parameter $\theta_{buf,max}$ was found to have a nonzero total order indices, this impact was caused by a small magnitude higher order interaction in which the parameter $\theta_{buf,max}$ was included. In Fig. 4 the total order indices are much greater than the first order indices. This is as the total order indices are a combination of the first and second order indices and as the second order effects are large the total order effects are far greater than the first order effects.

Table 6
Second order sensitivity indices from the power demand uncertainty analysis use case using the crop related parameter subset.

Parameter names (θ)	Second order indices
$\theta_{PAR,can} * \theta_{PAR,floor}$	0. 46
$\theta_{PAR,can} * \theta_{NIR,can}$	0. 29

3.3. Greenhouse gas demand uncertainty use case

The second use case in this study focusses on the prediction uncertainty in the prediction of gas demand arising from variations in parameters related to the heating system. The parameters that were selected as part of the pre-selection process are detailed in Table 7.

These parameters were sampled and used to calculate the gas demand. These samples and predictions were then used to calculate a PCE. The gas demand prediction errors were found to have a CV of 12 %. The first and total order sensitivity indices that were calculated from the PCE are displayed in Fig. 5.

The sensitivities displayed in Fig. 5 show that the gas demand prediction is sensitive to variations in the FIR emission coefficient of the heating pipes ($\theta_{epsPipe}$), the capacity of heating system (θ_{pBoil}) and heat exchange coefficient of lamp and the air ($\theta_{cHecLampAir}$). The parameters $\theta_{epsPipe}$ and θ_{pBoil} are related to the amount and efficiency of heat transferred from the boiler to the air temperature. By influencing the air temperature these parameters interact with the control dynamics as defined by the rules that control the boiler that are based on the air temperature. The same relationship is true for $\theta_{cHecLampAir}$ where the heat from the lamps influences the air temperature. The second order sensitivity indices found as part of the sensitivity analysis are displayed in Table 8.

The second order sensitivities are displayed in Fig. 5. The combined effect of $\theta_{epsPipe}$ and $\theta_{cHecLampAir}$ were found to have the greatest combined impact. These two parameters influence the temperature within the greenhouse by influencing the amount of heat energy that is transmitted into the greenhouse air from the lamps and hot water pipes.

3.4. Greenhouse combined gas and power demand uncertainty use case

The two most sensitive parameters from each of the previous two use cases were then taken and combined in an analysis of both gas and power demand prediction uncertainty, the selected parameters are described below in Table 9.

The parameters described in the Table above were sampled and used to calculate the power and gas demand. A PCE was subsequently fitted for the gas and power demand separately. In the case of the gas demand PCE which had a maximum polynomial degree of 5 and a final LOO error estimate of 4.7×10^{-3}. The gas demand prediction errors were found to have a CV of 18 %. The PCE generated using power demand predictions For the PCE made using the greenhouse power demand predictions which had a maximum polynomial degree of 9 and a final LOO error estimate of 2.9×10^{-4}. The gas demand prediction errors were found to have a CV of 18 % and the power demand prediction errors had a CV of 24 %. The first and total order sensitivity indices that were calculated from the gas and power PCE are displayed in Fig. 6.

Figure 6 shows us that the PCE for the power demand attributes almost all of the prediction variation to the uncertainty introduced via the parameter for the lamps power rating $\theta_{lamp,max}$. This corroborates the importance of $\theta_{lamp,max}$ that was highlighted in the sensitivity indices and large coefficient of variance found in the lamp parameter subset (section 3.1.2) and in the local sensitivity analysis in Appendix A. For the gas demand PCE, the parameter $\theta_{lamp,max}$ also had the greatest impact. This high sensitivity highlights the influence of the lamps power rating on the amount of heat that is transmitted to the air from the lamps. Which in turn influences the control of the boiler through the control dynamics and rules that references the air temperature. The second most sensitive parameter was $\theta_{PAR,trans,roof}$, this parameter it is used to calculate the amount of heat from the sun that is transmitted through the greenhouse glass and into the indoor air, which also affects the control of the boiler and CHP. A number of second order sensitivity indices were found for the PCE based on the gas demand which are displayed in Table 10.

In the case of the greenhouse gas demand the PCE does identify a number of second order interactions and these interactions are

Table 7
Definition of model parameter distributions for the gas demand and heating related parameter subset.

Parameter name (θ)	Distribution range (θ_l, θ_u)	Distribution mean (μ)	Distribution standard error (σ)	Units	Mean value reference
Ventilation discharge coefficient (θ_{cDgh})	$[0,1]$	0.75	0.0250	–	Vanthoor, Stanghellini, et al. (2011)
Greenhouse leakage coefficient ($\theta_{cLeakage}$)	$[0,1]$	1.0×10^{-4}	3.3×10^{-6}	–	Vanthoor, Stanghellini, et al. (2011)
Specific heat capacity of roof layer (θ_{cPRf})	$[0,\infty)$	0.84×10^{3}	28	$JK^{-1}kg^{-1}$	Vanthoor, Stanghellini, et al. (2011)
Thermal screen flux coefficient (θ_{kThScr})	$[0,1]$	0.05×10^{-3}	1.7×10^{-6}	$m^3m^{-2}K^{-\frac{2}{3}}s^{-1}$	Vanthoor, Stanghellini, et al. (2011)
FIR emission coefficient of the heating pipes ($\theta_{epsPipe}$)	$[0,1]$	0.88	0.029	–	Vanthoor, Stanghellini, et al. (2011)
Capacity of the heating system (θ_{pBoil})	$[0,\infty)$	2.1×10^{6}	7.0×10^{4}	W	Vermeulen (2016)
Heat capacity of lamp ($\theta_{capLamp}$)	$[0,\infty)$	100	3.3	$JK^{-1}m^{-2}$	Kusuma, Pattison, and Bugbee (2020)
Heat exchange coefficient of lamp ($\theta_{cHecLampAir}$)	$[0,1]$	0.09	3.0×10^{-3}	$Wm^{-2}K^{-1}$	Kusuma et al. (2020)

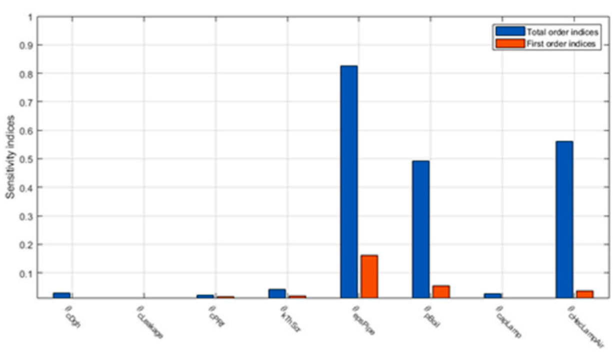


Fig. 5. First and total order sensitivity indices from the gas demand uncertainty analysis use case using the gas demand and heating related parameter subset.

Table 8
Second order sensitivity indices from the gas demand and heating related parameter subset.

Parameter names (θ)	Second order indices
$\theta_{epsPipe} * \theta_{cHecLampAir}$	0.25
$\theta_{epsPipe} * \theta_{pBoil}$	0.16
$\theta_{pBoil} * \theta_{cHecLampAir}$	0.015

comparatively small in magnitude. The largest of these combined effects identifies a combined impact from variation in $\theta_{PAR,ref,roof}$ and $\theta_{PAR,can}$. Both of these parameters influence the temperature within the greenhouse as they are used to calculate how much radiative heat enters the greenhouse respectively and is absorbed by the crop canopy. As such this combined sensitivity indices highlights the impact of the heat transferred from the ambient radiation to the air via the crop canopy.

In the case of power demand the PCE found no second order interactions but did find very small interactions at higher orders fourth and fifth order. This means that the PCE could not attribute variation in the prediction to any group of 2 or 3 parameters and that there is a high degree of interaction amongst larger groups of parameters that accounts for a small fraction of prediction variation. To corroborate the power demand uncertainty results a local one-by-one sensitivity analysis of Greenlight was performed and can be seen in Appendix A.

4. Discussion

This study proposed and demonstrated an algorithm to analyse the greenhouse energy prediction uncertainty arising from the combined and individual impact of errors in the model parameters using a global sensitivity analysis.

The analysis performed on the subset of parameters related to the greenhouse structure concluded that the most impactful parameters on power demand prediction were total order effects related to the amount of radiation that is reflected from the floor and transmitted through the greenhouse glass. This result was corroborated by Vanthoor, van Henten, et al. (2011), which also found that the transmissive properties of the glass to incoming PAR and NIR had the greatest impact on greenhouse energy demand. This study makes a clear departure from previous research by concluding that for the greenhouse structure subset of parameters, no first order effect was found and instead groups of parameters were responsible for the variation in the prediction. This collective impact stems from the compounding of uncertainties as multiple equations with uncertain parameters feed into each other and feedback propagation where parameters influence states in the model that are self-referential and iteratively create ever greater uncertainty. This insight offers a crucial new perspective from conventional wisdom in the field that has considered the impact of pairs of parameters to be sufficient. This study demonstrates that the impact of higher order interactions and groups of parameters is a central tool to understanding the causes of prediction uncertainty. Furthermore, this implies that the tuning of any single or pair or parameters will not necessarily reduce prediction uncertainty due to the high levels of interaction. As such this study highlights the opportunity and a method to consider higher order interactions in greenhouse parametric uncertainty analyses.

This study also examined the influence of crop parameters on greenhouse power demand prediction. The results from this subset highlighted the parameters relating to the amount of radiation absorbed from below and above the canopy to be the most influential factors on power demand prediction. This finding is corroborated by Schrevens et al. (2008), who found that parameters related to the light use efficiency of the crop also have the greatest impact on energy demand prediction. While this study and Schrevens et al. (2008) consider different greenhouses using different energy systems. Schrevens et al. (2008) concluded that the magnitude of power prediction uncertainty from crop parameters was 5.8 % and is comparable to the 5.1 % prediction uncertainty described in section 3.2.3. It should also be noted that while these studies provide some comparable insights to literature

Table 9
Definition of model parameter distributions.

Combined gas and power demand parameter subset					
Parameter name (θ)	Distribution range (θ_l, θ_u)	Distribution mean (μ)	Distribution standard error (σ)	Units	Reference
FIR emission coefficient of the heating pipes ($\theta_{epsPipe}$)	$[0, 1]$	0.88	0.029	–	Vanthoor, Stanghellini, et al. (2011)
Heat exchange coefficient of lamp ($\theta_{cHecLampAir}$)	$[0, 1]$	0.09	3.0×10^{-3}	$Wm^{-2}K^{-1}$	Kusuma et al. (2020)
PAR extinction coefficient of the canopy ($\theta_{PAR,can}$)	$[0, 1]$	0.7	0.023	–	Vanthoor, Stanghellini, et al. (2011)
PAR extinction coefficient of the canopy for light reflected from the floor ($\theta_{PAR,floor}$)	$[0, 1]$	0.7	0.023	–	Vanthoor, Stanghellini, et al. (2011)
Maximum intensity of lamps ($\theta_{lamp,max}$)	$[0, \infty)$	110	3.7	Wm^{-2}	Katzin et al. (2020)
Emissivity of topside of lamp ($\theta_{emis,top,lamp}$)	$[0, 1]$	0.1	3.3×10^{-3}	–	Katzin et al. (2020)
PAR transmission coefficient of the roof ($\theta_{PAR,trans,roof}$)	$[0, 1]$	0.85	0.028	–	Vanthoor, Stanghellini, et al. (2011)
PAR reflection coefficient of the floor ($\theta_{PAR,ref,floor}$)	$[0, 1]$	0.65	0.022	–	Vanthoor, Stanghellini, et al. (2011)

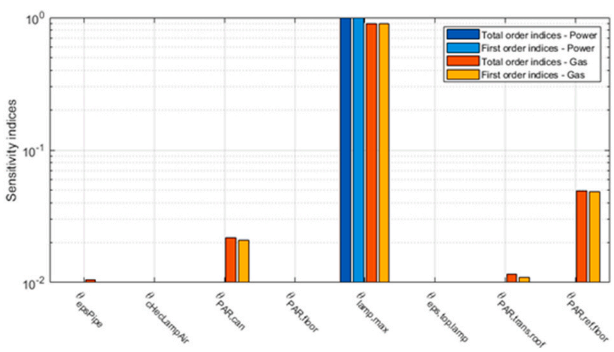


Fig. 6. First and total order sensitivity indices using the gas demand and heating related parameter subset. This figure is presented with a logarithmic scale.

Table 10
Second order sensitivity indices from the combined parameter subset on the uncertainty in gas demand.

Parameter names (θ)	Second order indices
$\theta_{PAR,can} * \theta_{PAR,ref,roof}$	3.7×10^{-4}
$\theta_{PAR,can} * \theta_{PAR,trans,floor}$	1.8×10^{-4}
$\theta_{lamp,max} * \theta_{PAR,trans,roof}$	1.1×10^{-4}
$\theta_{PAR,floor} * \theta_{PAR,ref,roof}$	1.1×10^{-4}

the degree to which any set of parameters influences the energy demand of the greenhouse is dependent on the design of the rule set used to control the greenhouse.

The analysis of the subset related to lamp parameters found that 99 % of the variation in greenhouse power demand prediction uncertainty could be attributed to a first order effect from the HPS lamp light intensity parameter ($\theta_{lamp,max}$). This result is understandable as the power demand of the greenhouse is almost entirely from operating the lamps and by changing their power rating the total demand changes. It should also be noted that $\theta_{lamp,max}$ also influences the air temperature through radiative heat exchange, which in turn influences the temperature-based rules that control the lamps themselves. In doing so $\theta_{lamp,max}$ has multiple routes of propagation throughout the model and influences a feedback loop between the air temperature and the lamp rules. The consistency of this result was corroborated by previous unpublished research done using the KASPRO model (de Zwart, 1996; Dieleman, Meinen, Marcelis, de Zwart, & van Henten, 2005) and using a local one-by-one sensitivity analysis of Greenlight in Appendix A. While we assume sufficiency for the other methods this does present an interesting avenue for future research whereby multiple methods are applied to the same sets of parameters proposed in this study. While this study did highlight the impact of lamp light intensity, other parameters were also found to be impactful through second order interactions as shown in Table 4. However, these effects were minor, meaning that no large improvement in prediction uncertainty could be made by tuning any one of the pairs highlighted in the second order indices.

A further subset of parameters was proposed to investigate the prediction uncertainty in gas demand. The analysis of this subset of parameters found that the parameters relating to the capacity of the boiler and the lamps to deliver heat to the greenhouse is key. Furthermore, this analysis found that the parametrisation of the greenhouse structure or air leakage was comparably unimportant for gas demand prediction.

The analysis of the combined subset found that the maximum intensity of the lamps ($\theta_{lamp,max}$) was the most sensitive parameter for power and gas demand prediction accounting for 99 and 90 % respectively. In the case of power demand this result was already indicated by the high sensitivity of $\theta_{lamp,max}$ and high coefficient of variation for the lamp parameter subset. For the gas demand prediction uncertainty, the CV for the combined parameter subset (18 %) is higher than the initial heating parameter set (12 %). This indicates that parameters that were added from the sets related to power demand had some impact of gas demand prediction uncertainty. Specifically, $\theta_{lamp,max}$ was found to be the most sensitive in the augmented set. The overall importance of $\theta_{lamp,max}$ stems directly from the air temperature-based control rules that operate the boiler and the lamps. This highlights the importance of the augmented subset and the need for a carefully designed selection criteria so that impactful parameters like $\theta_{lamp,max}$ are not overlooked.

The demonstration of the algorithm proposed in this study found that for power demand prediction uncertainty variation in crop and structurally related parameter caused a coefficient of variation of 5.1 % and 5.2 % respectively. Variation in the subset of parameters related to the HPS lamp lighting resulted in a coefficient of variation of 24 % for power demand prediction. This outcome shows that for the purposes of reducing power demand prediction uncertainty the accurate parametrisation of the lamp lighting system is more impactful than the greenhouse structure or crop.

A key conclusion that can be drawn from this study is that the greenhouse air temperature is a major contributor to uncertainty propagation in both gas and power prediction uncertainty. This route for uncertainty propagation is facilitated by the way the greenhouse controller is designed. Accordingly, an effective way to combat prediction uncertainty of greenhouse power and gas demand is to focus first on the attributes used in the rules that control the greenhouse lighting before addressing the accuracy of the internal light physics of the greenhouse. In a similar way Van Henten (2003) concluded that the economic optimisation of the greenhouse's operation was not sensitive to the internal climate dynamics of the greenhouse. The reason the parameterisation of the greenhouse's light physics does not have an effect on the control rules that respond to light levels is that the rules used to control the lighting in Katzin et al. (2020) do not consider the internal light physics of the greenhouse but instead respond to the ambient outdoor radiation.

Despite the benefits of the proposed algorithms, there is potential for improvement. For example, due to a lack of available information this study assumed that all the model parameters have standard error that are defined according to eq. (6). It may be that for some of the parameters the standard error may differ from this assumed value and may be known very precisely. Despite this limitation, the algorithm proposed in this study offer insight as to what processes in the model are vulnerable to uncertainty.

A potential limitation of this study's algorithms is that a PCE is a form of regression and as such has an associated error, this may marginally alter the sensitivities but not the algorithms main conclusions. While this study has addressed the impact of this error on the insight the algorithm produces using a validatory local sensitivity analysis (Appendix A). Future research may assess the impact of this PCE error via an analysis where an increasing sample size is used to assess the development of PCE error. Furthermore, there is an opportunity to conduct conventional sample based Sobol sensitivity indices for higher orders of interaction to validate the higher order insights gained from this study. It should also be noted that for all the analyses described in this study a number of factors with low total order sensitivities are given a value of zero for their first order sensitivity indices. This is an outcome of using the LARS algorithm (described in section 2.4.1) whose sparce-favouring method sets low correlation coefficients from the meta-model to zero to reduce the required computation.

A further limitation of this study is that the use of subsets to further subdivide the parameter population does preclude the analysis of the effect of interactions between all the parameters within different subsets. Interactions within the subsets are included in the augmented subset but only for the parameters that were initially found to be most sensitive. This design decision in the proposed algorithm does effectively focus the analysis on the most important factors but may also remove interactions between the subsets from parameters that initially were not found to be sensitive. This does open the opportunity of further analysis where all the preselected parameters might be repeatedly shuffled into new subset to explore the impact of all the possible combinations.

While previous studies have analysed the effect of parameter uncertainty in greenhouse (Cooman & Schrevens, 2006; López-Cruz et al., 2013; Schrevens et al., 2008) this study progresses the field by proposing an algorithm that systematically considers all the model parameters and ultimately selects and analyses the impact of the relevant parameters. Moreover, unlike previous studies (Cooman & Schrevens, 2004; López-Cruz, Rojano-Aguilar, Salazar-Moreno, & Ruiz-Garcia, 2012; Vanthoor, van Henten, et al., 2011) this study introduced the use of polynomial chaos expansions for uncertainty and sensitivity analyses in the field of greenhouse horticultural research. In doing so this study was able to attribute prediction uncertainty to individual and grouped uncertainty sources, ranging in size from 1 to 10 members, allowing for a more detailed and targeted analysis of larger groups of parameters.

This study introduces a new form of promising uncertainty analysis in the form of polynomial chaos expansions, and while it contains a number of outstanding challenges it also opens avenues for furthering the use of uncertainty analysis in the field of greenhouse horticultural research. Future research in this field might consider if the design of the greenhouse controller influences the sensitivity of the model parameters. This could be done by performing a PCE analysis on a greenhouse that is controlled by an optimal controller and rule based controller.

5. Conclusion

This study introduced an algorithm to investigate the role of parametric uncertainty on the prediction uncertainty of greenhouse gas and electrical power demand. This algorithm included a pre-screening process to reduce the number of relevant parameters considered in the analysis and a Polynomial Chaos Expansion based sensitivity analyses. The application of this algorithm concluded that parameters related to the greenhouse lighting were the greatest contributors to greenhouse power demand prediction uncertainty over crop and greenhouse structure related parameters. In particular, the power relating to the power rating of the lamps was found to be the single greatest contributor to both gas and power demand prediction uncertainty. In addition, this study made the notable finding that larger groups of model parameters were responsible for energy demand prediction uncertainty. This novel insight highlights the need for future research to consider the impact of higher order parameter interactions on prediction uncertainty using the algorithm proposed in this study.

Funding

This work is part of the research programme FlexCrop with the project number 647.003.006, which is (partly) financed by the Dutch Research Council (I) and the companies, AgroEnergy, Blue Radix, B-Mex, LTO Glaskracht Nederland, Letsgrow.com, WUR Greenhouse Horticulture and Delphy.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank David Katzin for his support in the use of the Greenlight greenhouse model and we would also like to thank Letsgrow.com and Solyco for the provision of the data used in this study.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.biosystemseng.2024.01.006.

Appendix A. Local one-by-one parametric uncertainty analysis

To corroborate and assess the consistency of the results found in the parametric PCE meta-model-based analysis (section 3.1) a one-by-one local sensitivity analysis was done. This uses the power prediction demand calculated using a parameter sample set in which each parameter is sampled 20 times individually. All the remaining parameters retained their nominal value, as defined in Table 3. The sensitivity of each parameter (i) was then defined as the fraction of variation in the power demand prediction RMS error made using a parameter set with variations in only one parameter ε_i^{RMS}, over the variation of perturbing all of the parameters simultaneously 1000 times, $\varepsilon_{\cdot}^{RMS}$. As such these sensitivity indices are defined as follows,

$$\mathrm{SI_i} = \frac{\mathrm{var}\left(\varepsilon_i^{RMS}\right)}{\mathrm{var}\left(\varepsilon_{\cdot}^{RMS}\right)}. \tag{A.1}$$

The resulting sensitivity indices are shown in the figure below.

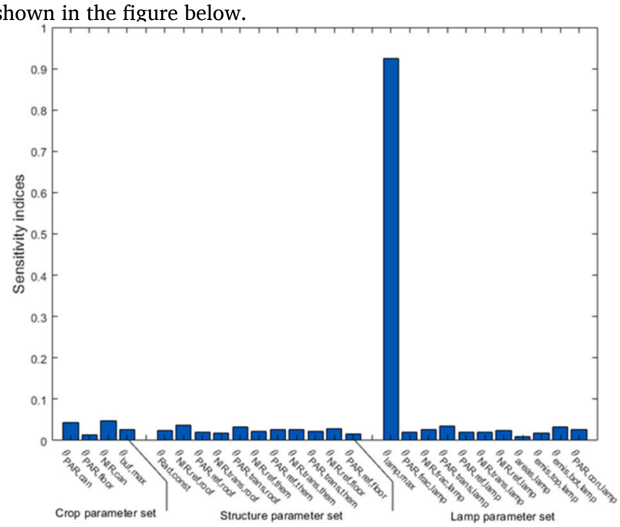


Fig. A.1. The one-by-one sensitivity indices (SI_i) of the greenhouse's power demand prediction to individual variation in the model parameters.

Figure A.1 shows that even though this analysis does not consider interaction and has a limited sample size the key outcomes mirror that of the total order indices of the PCE used in the study. These being that the lamp light intensity has the greatest contribution and that remaining parameters do have a limited impact.

References

Archer, G. E. B., Saltelli, A., & Sobol, I. M. (1997). Sensitivity measures, anova-like Techniques and the use of bootstrap. *Journal of Statistical Computation and Simulation, 58*(2), 99–120. https://doi.org/10.1080/00949659708811825

Blatman, G., & Sudret, B. (2011). Adaptive sparse polynomial chaos expansion based on least angle regression. *Journal of Computational Physics, 230*(6), 2345–2367. https://doi.org/10.1016/j.jcp.2010.12.021

Cooman, A., & Schrevens, E. (2004). Sensitivity analyses of TOMGRO output variables to variations in climate conditions. *Acta Horticulturae, 654*, 317–324. https://doi.org/10.17660/ActaHortic.2004.654.37

Cooman, A., & Schrevens, E. (2006). A Monte Carlo approach for estimating the uncertainty of predictions with the tomato plant growth model, Tomgro. *Biosystems Engineering, 94*(4), 517–524. https://doi.org/10.1016/j.biosystemseng.2006.05.005

de Zwart, H. F. (1996). *Analysis energy-saving options in greenhouse cultivation using a simulation model* (Ph.D Thesis). Landbouwuniversiteit Wageningen.

de Zwart, H. F., Baeza, E., van Breugel, B., Mohammadkhani, V., & Janssen, H. (2017). *De uitstralingmonitor*.

Dieleman, J. A., Meinen, E., Marcelis, L. F. M., de Zwart, H. F., & van Henten, E. J. (2005). Optimisation of CO2 and temperature in terms of crop growth and energy use. *Acta Horticulturae, 691*, 149–154. https://doi.org/10.17660/ActaHortic.2005.691.16

Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least angle regression. *Annals of Statistics, 32*(2), 407–499. https://doi.org/10.1214/009053604000000067

Golzar, F., Heeren, N., Hellweg, S., & Roshandel, R. (2018). A novel integrated framework to evaluate greenhouse energy demand and crop yield production. *Renewable and Sustainable Energy Reviews, 96*(November 2017), 487–501. https://doi.org/10.1016/j.rser.2018.06.046

Katzin, D., van Mourik, S., Kempkes, F., & van Henten, E. J. (2020). GreenLight – an open source model for greenhouses with supplemental lighting: Evaluation of heat requirements under LED and HPS lamps. *Biosystems Engineering, 194*, 61–81. https://doi.org/10.1016/j.biosystemseng.2020.03.010

Kusuma, P., Pattison, P. M., & Bugbee, B. (2020). From physics to fixtures to food: Current and potential LED efficacy. *Horticulture Research, 7*(1). https://doi.org/10.1038/s41438-020-0283-7

López-Cruz, I. L., Martínez-Ruiz, A., Ruiz-Garcıá, A., & Gallardo, M. (2020). Uncertainty analyses of the VegSyst model applied to greenhouse crops. *Acta Horticulturae, 1271*, 199–206. https://doi.org/10.17660/ActaHortic.2020.1271.28

López-Cruz, I. L., Rojano-Aguilar, A., Salazar-Moreno, R., & Ruiz-Garcia, A. (2012). Global sensitivity analysis of greenhouse crop models. *Acta Horticulturae, 952*, 103–110. https://doi.org/10.17660/ActaHortic.2012.952.11

López-Cruz, I. L., Ruiz-García, A., Ramírez-Arias, A., & Vázquez-Peña, M. (2013). Uncertainty analysis of a greenhouse lettuce crop model. *Revista Chapingo Serie Horticultura, 19*(1), 33–47. https://doi.org/10.5154/r.rchsh.2011.09.049

Luo, W., de Zwart, H. F., DaiI, J., Wang, X., Stanghellini, C., & Bu, C. (2005). Simulation of greenhouse management in the subtropics, Part I: Model validation and scenario study for the winter season. *Biosystems Engineering, 90*(3), 307–318. https://doi.org/10.1016/j.biosystemseng.2004.11.008

Mara, T. A., & Becker, W. E. (2021). Polynomial chaos expansion for sensitivity analysis of model output with dependent inputs. *Reliability Engineering & System Safety, 214*. https://doi.org/10.1016/j.ress.2021.107795

Nelson, J. A., & Bugbee, B. (2014). Economic analysis of greenhouse lighting: Light emitting diodes vs. high intensity discharge fixtures. *PLoS One, 9*(6). https://doi.org/10.1371/journal.pone.0099010

Nelson, J., & Bugbee, B. (2015). Analysis of environmental effects on leaf temperature under sunlight, high pressure sodium and light emitting diodes. *PLoS One, 10*(10). https://doi.org/10.1371/journal.pone.0138930

Orgill, J., & Hollands, K. (1977). Correlation equation for hourly diffuse radiation on a horizontal surface. *Solar Energy, 19*(4), 357–359. https://doi.org/10.1016/0038-092X(77)90006-8

Schrevens, E., Jancsok, P., & Dieussaert, K. (2008). Uncertainty on estimated predictions of energy demand for dehumidification in a closed tomato greenhouse. *Acta Horticulturae, 801*(PART 2), 1347–1354. https://doi.org/10.17660/ActaHortic.2008.801.165

Sobol, I. M. (1993). Sensitivity estimates for nonlinear mathematical models. *Math Modeling Computer Exp, 1*(4), 407–414.

Sudret, B. (2008). Global sensitivity analysis using polynomial chaos expansions. *Reliability Engineering & System Safety, 93*(7), 964–979. https://doi.org/10.1016/j.ress.2007.04.002

Van Henten, E. J. (2003). Sensitivity Analysis of an Optimal Control Problem in Greenhouse Climate Management. *Biosystems Engineering, 85*(3), 355–364. https://doi.org/10.1016/S1537-5110(03)00068-0.

Vanthoor, B., Stanghellini, C., van Henten, E. J., & de Visser, P. (2011). A methodology for model-based greenhouse design : Part 1 , a greenhouse climate model for a broad range of designs and climates. *Biosystems Engineering, 110*(4), 363–377. https://doi.org/10.1016/j.biosystemseng.2011.06.001

Vanthoor, B., van Henten, E. J., Stanghellini, C., & de Visser, P. H. B. B. (2011). A methodology for model-based greenhouse design: Part 3, sensitivity analysis of a combined greenhouse climate-crop yield model. *Biosystems Engineering, 110*(4), 396–412. https://doi.org/10.1016/j.biosystemseng.2011.08.006

Vazquez-Cruz, M. A., Guzman-Cruz, R., Lopez-Cruz, I. L., Cornejo-Perez, O., Torres-Pacheco, I., & Guevara-Gonzalez, R. G. (2014). Global sensitivity analysis by means of EFAST and Sobol' methods and calibration of reduced state-variable TOMGRO

model using genetic algorithms. *Computers and Electronics in Agriculture, 100*, 1–12. https://doi.org/10.1016/j.compag.2013.10.006
Vermeulen, P. C. M. (2008). *Kwantitatieve informatie voor de glastuinbouw 2008. Kengetallen Voor Groenten-Snijbloemen-Potplanten Teelten*. Rapport: Wageningen UR Glastuinbouw.
Vermeulen, P. C. M. (2016). Kwantitatieve informatie voor de glastuinbouw 2016-2017. In *Business unit glastuinbouw*. Wageningen University & Research.