Preprints of the 21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

Fast Probabilistic Uncertainty Quantification and Sensitivity Analysis of a Mars Life Support System Model ⋆

Georgios Makrygiorgos * **Soumyajit Sen Gupta** **
Amor A. Menezes ** **Ali Mesbah** *

* *Department of Chemical and Biomolecular Engineering, University of California, Berkeley, CA 94720 USA (gmakr,mesbah@berkeley.edu)*
** *Department of Mechanical and Aerospace Engineering, University of Florida, Gainesville, FL 32611 USA (s.sengupta,amormenezes@ufl.edu)*

Abstract: Mars life support system models consist of numerous mission-critical, interrelated, and scenario-specific parameters. The large size and involved nature of these models make them computationally expensive, with parameters that are subject to several sources of uncertainty. Accurately characterizing these uncertainties and their impact on overall model predictions is crucial for decision-support and mission optimization. This paper focuses on uncertainty quantification of a model of a space crop cultivation system, which is one of several systems that are required on a long-duration manned Mars mission. The model performs constrained optimization of the equivalent system mass (ESM) metric, which augments shipped mass costs with those of pressurized volume, demanded power, thermal control, and needed crew time. This paper uses surrogate modeling for fast quantification of the effect of probabilistic uncertainty in mission-critical parameters of semi-empirical equations that describe crop growth and equipment operation. This work shows sparse polynomial chaos-Kriging (PCK) yields a computationally cheap-to-evaluate surrogate for the minimum ESM that accounts for probabilistic uncertainty in 86 model parameters. This surrogate model accelerates a global sensitivity analysis that elucidates which crop growth and equipment operation parameters are critical to mission outcome variability. The PCK surrogate model realizes a 100-fold computational speed gain in the estimation of the probability distribution of the minimum ESM.

Keywords: Mixed integer linear programming; uncertainty quantification; global sensitivity analysis; sparse polynomial chaos; Kriging; space exploration and transportation

1. INTRODUCTION

Crew life support is the most critical element of long, deep space, manned Mars missions (Drake, 2009). Life support systems manage astronaut air, water, and waste, and include their food, habitation, power, thermal control, and biomass production (Anderson et al., 2018). Necessary system technologies and components may be shipped from Earth at mission start to ensure guaranteed availability. But shipping all items at their required level for a full mission duration is uneconomical since the mass and cost of launch fuel and support structure directly depend on launch payload. Thus, there are newer, alternative ways to use available local resources *in situ* for life support (Menezes et al., 2015; Do et al., 2016). Food demand for a two-and-a-half year Mars mission can be met by shipping all food, producing it on site with novel space biomanufacturing (Menezes et al., 2015) or with traditional crop cultivation, or some combination of these options. Shipping food requires freezers and negatively affects food freshness over the mission horizon. Producing food on site with local resources ensures that food is fresh, but requires a (possibly expensive) shipping of agricultural cultivation setups, lighting systems, and related accessories.

To systematically analyze alternative options for food production in deep space missions, this paper presents a mixed integer linear programming (MILP) model that optimizes a food production life support system with the objective of minimizing *equivalent system mass* (ESM) (Levri et al., 2003). ESM is a metric that accounts for mission costs, such as pressurization, power, thermal control, and crew time (Anderson et al., 2018). Here, ESM provides the cost of generating one or more crop yields in the food production life support system. Crop yield is obtained by incorporating the modified energy cascade (MEC) model (Anderson et al., 2018) in the MILP formulation. The MEC model is semi-empirical, and depends on crop growth chambers and supporting device functionality.

The resultant MILP model has numerous uncertain variables and parameters. For example, the yield of food crops over a cultivation period is a function of the varying

⋆ This material is based upon work supported by the National Aeronautics and Space Administration (NASA) under grant number NNX17AJ31G. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NASA.

availability of photosynthetic light and carbon dioxide, as well as the characteristics of the cultivar. Further, currently-proposed technologies to provide light to the cultivation chamber, to generate power, and to control temperature are of unknown design and efficiency 15 years ahead of an intended mission. These considerations necessitate accurate quantification of the impact of uncertainty in the MILP model's minimization of ESM to inform future decisions about the food production life support system architecture.

This paper presents a computationally efficient framework for uncertainty quantification (UQ) of the food production life support system under probabilistic model parameter uncertainty. UQ for this system is computationally challenging due to the relatively large computation time of solving the MILP model per uncertainty realization, as well as the large number of uncertain model parameters (86 uncertain parameters) that can render conventional sample-based UQ methods prohibitive. In this paper, we employ a surrogate modeling approach to derive a cheap-to-evaluate model for predicting the minimum ESM, which is the main quantity of interest from the MILP model. Polynomial chaos (Xiu and Karniadakis, 2002) has been widely used for forward and inverse UQ tasks in the analysis and optimization of various engineering applications (Paulson et al., 2019; Paulson and Mesbah, 2019). However, a major challenge in polynomial chaos arises from handling a large number of uncertain inputs. To overcome this challenge, we use sparse polynomial chaos expansions (PCEs) (Deman et al., 2016), where the expansion is sparsely truncated. Sparsity is further reinforced by using the least-angle-regression (Blatman and Sudret, 2013) algorithm for coefficient estimation, retaining only a small number of non-zero terms in the expansion. The sparse PCEs are then combined with Kriging, referred to as polynomial chaos-Kriging (PCK) (Schobi et al., 2015), to increase accuracy and to quantify the variance of the surrogate model predictions.

We demonstrate the application of PCK surrogate modeling for fast, sample-based estimation of the probability distribution of the minimum ESM, as well as for global sensitivity analysis (GSA) with respect to the 86 uncertain parameters of the MILP model. The GSA is particularly useful for elucidating the impact of various uncertain model and mission-critical parameters on the minimum ESM, toward insights on performance optimization of the food production life support system. We show that the PCK surrogate model leads to approximately 100-fold savings in computational cost of probability distribution estimation when compared to direct sampling of the MILP model.

2. LIFE SUPPORT MODEL DESCRIPTION

We restrict decision variables in our MILP model that specify equipment quantities to the non-negative integers, and we take lighting source selection variables to be binary. We assume that the remaining decision variables are continuous and non-negative.

2.1 Model Formulation

We start with a mass balance for all crops in index set $\mathfrak{C}$ that are to be cultivated over mission duration T

$$A_{i+1,j} = A_{i,j} + C_{i,j} - D_j, \quad \forall j \in \mathfrak{C}, i \in [1, T], \quad (1)$$

where $A_{i,j}$ is the available quantity of crop j at the start of day i, $C_{i,j}$ is the amount of edible food crop j harvested at the start of day i, and food demand D_j is a fixed model parameter that depends on human metabolism (Anderson et al., 2018) and the fractions of different food j in the diet, which comprise decision variables. These fractions must sum to one, and are constrained by a minimum requirement.

The mass balance (1) is valid from day 2 to the last day T for all crops. On day 1 of the mission, the amount of available food is the same as the amount shipped S_j, also a decision variable. S_j can be consumed until freshness expiry, which is crop-specific. The total food shipped determines the number of freezers N^f to be shipped, an integer variable, with each being of a fixed, predefined capacity. Food demand D_j must be met daily. Accordingly,

$$A_{i,j} \geqslant D_j, \quad \forall j \in \mathfrak{C}, i \in [1, T].$$

For each crop j, the harvest $C_{i,j}$ at the end of cultivation period τ_j is directly related to the number of agricultural cultivation receptacles (ACRs) $N_{i,j}^{ACR_u}$ that grow the crop on day i, their fixed predefined cultivation area α^{ACR}, and the areal yield of edible biomass Y_j. Hence,

$$C_{i,j} = N_{i-\tau_j,j}^{ACR_u} \alpha^{ACR} Y_j, \quad \forall j \in \mathfrak{C}, i \in [\tau_j + 1, T]. \quad (2)$$

In (2), the cultivation period τ_j for crop j is the crop-specific maturity period. The yield of edible biomass Y_j is also crop-specific, and is adapted from the MEC model for hydroponic growth systems (Anderson et al., 2018). This yield is a function of environmental factors like photosynthetic photon flux (PPF), carbon dioxide (CO_2) concentration, and crop characteristics that include photon absorption capacity and carbon use efficiency. Among these parameters, canopy quantum yield (CQY) and time until canopy closure are determined using empirical functions of PPF and CO_2 concentration.

We compute the required amount of fertilizer by multiplying the edible and inedible biomass amounts with their respective compositions of essential nitrogen, phosphorus, and potassium. These quantities then give us the number of fertilizer container tanks N^t that must be shipped. The total number of ACRs that must be shipped N_j^{ACR} is the maximum number of ACRs in use at any time. Hence,

$$N_j^{ACR} \geqslant \sum_{m=\max[1, i-\tau_j+1]}^{i} N_{m,j}^{ACR_u}, \quad \forall j \in \mathfrak{C}, i \in [1, T - \tau_j].$$

We also optimize the choice of lighting source $k \in \mathfrak{L}$, which is captured by the binary variable B_k

$$\sum_{k \in \mathfrak{L}} B_k = 1.$$

The number of ACRs that grow crop j and are connected to lighting source k are defined by $\tilde{N}_{j,k}^{ACR}$, so that

$$\sum_{k \in \mathfrak{L}} \tilde{N}_{j,k}^{ACR} = N_j^{ACR}, \quad \forall j \in \mathfrak{C}.$$

The lighting source decision B_k is assumed to be the same for all crops. Instead of multiplying N_j^{ACR} by B_k, we use a traditional big-M constraint to avoid non-convexity

$$\sum_{j \in \mathfrak{C}} \tilde{N}_{j,k}^{ACR} \leqslant B_k M, \quad \forall k \in \mathfrak{L}.$$

We choose this big-M value judiciously, as too large values result in numerical instability, while too small values exclude a segment of the feasible space.

Mass M_k^{light} and power P_k^{light} for lighting system k are specified by the connection to the number of ACRs growing crop j, and the thermal control demand T_k^{light} is assumed to be equal to P_k^{light} (Anderson et al., 2018). Crops are illuminated for the duration of their photoperiods. Let γ_k be the PPF generated per unit power-input for a light source k, and η_k be its light source efficiency. Then,

$$P_k^{light} = \sum_{j \in \mathfrak{C}} \tilde{N}_{j,k}^{ACR} \alpha^{ACR} \frac{PPF_j}{\gamma_k \eta_k}, \quad \forall k \in \mathfrak{L}.$$

Let $\bar{M}_k^{light}$ be the specific mass of light source k corresponding to unit power output. Then,

$$M_k^{light} = \sum_{j \in \mathfrak{C}} \tilde{N}_{j,k}^{ACR} \alpha^{ACR} \frac{PPF_j}{\gamma_k} \bar{M}_k^{light}, \quad \forall k \in \mathfrak{L},$$

where PPF_j is the range of crop-specific allowable values in the MEC model.

We aim to minimize the total ESM of the shipment that is required for the food production system

$$z_1 = ESM^{fo} + ESM^{fr} + ESM^{ACR} + ESM^{li} + ESM^{fe}. \quad (3)$$

Here, z_1 is the summation of the ESMs for shipping food ESM^{fo}, freezers ESM^{fr}, ACRs ESM^{ACR}, lighting apparatus ESM^{li}, and fertilizer ESM^{fe}. We obtain ESM^{fo} by multiplying the shipped food quantity S_j with a packing factor from Anderson et al. (2018). ESM^{fr} and ESM^{ACR} are the weighted sums of the mass, volume, power, and thermal demand of the total number of freezers N^f and ACRs N_j^{ACR} to ship, respectively, with ESM weighting coefficients as defined in Anderson et al. (2018). This is the same for ESM^{li} except for the volume penalty, since we assume that the lighting system is embedded in the ACRs. Likewise, ESM^{fe} has cost weights on the amount of fertilizer, as well as the mass and volume of the fertilizer container tanks to ship. A secondary objective function z_2

$$z_2 = \sum_{j \in \mathfrak{C}} \sum_{i=1}^{T} A_{i,j}$$

is used to ensure food freshness by minimizing its total available quantity over the mission.

2.2 Motivation for uncertainty analysis

The above MILP model has numerous crop-specific and empirical parameters, such as crop maturity period, photoperiod, carbon use efficiency, incident PPF, and CQY, among others. The predictions of crop growth dynamics are dependent on these parameters, and are thus subject to uncertainty. The model also has scenario-dependent parameters, such as lighting source efficiency, CO_2 concentration, PPF generation per unit of power, capacity values of ACRs, fertilizer tanks and freezers. Parameter variations can occur with changes in environment conditions, crop physiology differences in a batch, or failures in the system. Moreover, the ESM coefficients in (3) depend on the choice of power generation and thermal control technologies, and most of these values are estimates. Therefore, the substantial uncertainty in model parameters necessitates a systematic investigation of impacts of these parameters on optimal decision variables.

3. SURROGATE MODELS FOR UNCERTAINTY QUANTIFICATION AND GLOBAL SENSITIVITY ANALYSIS

Consider a computationally expensive model $\mathcal{M}$ that contains M uncertain parameters $\boldsymbol{\xi} = \{\xi_1, \xi_2, ..., \xi_M\}$. Parameter uncertainty is represented by a random vector $\boldsymbol{\Xi} \in \mathbb{R}^M$ with some known joint probability distribution function $\boldsymbol{\Xi} \sim f_{\boldsymbol{\Xi}}$. Here, we briefly present the procedure for performing sensitivity analysis based on Sobol' indices (Sobol, 2001).

Consider some quantity of interest (QoI) Y predicted by the model $\mathcal{M}$. We represent the QoI in the form of the analysis of variance (ANOVA)

$$Y = \mathcal{M}(\boldsymbol{\Xi}) = \mathcal{M}_0 + \sum_{\boldsymbol{v} \neq \emptyset} \mathcal{M}_{\boldsymbol{v}}(\boldsymbol{\Xi}_{\boldsymbol{v}}),$$

where $\mathcal{M}_0$ is constant and equal to the expected value of the response $\mathcal{M}(\boldsymbol{\Xi})$ over the uncertain input space. The notation $\boldsymbol{v}$ is used to denote subsets of the input space, so that $\boldsymbol{v} = \{i_1, ..., i_s\} \subset \{1, ..., M\}$. Thus, $\boldsymbol{\Xi}_{\boldsymbol{v}}$ represents the corresponding subvector of $\boldsymbol{\Xi}$. The uniqueness of the decomposition is guaranteed if the integrals of the summands over their variables satisfy

$$\int_{\mathcal{D}_{\boldsymbol{\Xi}_k}} \mathcal{M}_{\boldsymbol{v}}(\boldsymbol{\xi}_{\boldsymbol{v}}) f_{\Xi_k}(\xi_k) d\xi_k = 0, \quad k \in \boldsymbol{v}, \quad (4)$$

where $\mathcal{D}_{\boldsymbol{\Xi}_k}$ denotes the support, and f_{Ξ_k} is the marginal of the uncertainty subspace. Moreover, the following orthogonality condition holds

$$\mathbb{E}\left[\mathcal{M}_{\boldsymbol{v}}(\boldsymbol{\Xi}_{\boldsymbol{v}}) \mathcal{M}_{\boldsymbol{u}}(\boldsymbol{\Xi}_{\boldsymbol{u}})\right] = 0, \quad \boldsymbol{v} \neq \boldsymbol{u}. \quad (5)$$

We represent the total variance of Y as D. Properties (4)–(5) allow decomposition of the variance of Y

$$D = \sum_{\boldsymbol{v} \neq \emptyset} D_{\boldsymbol{v}},$$

where $D_{\boldsymbol{v}}$ denotes the partial variance given by

$$D_{\boldsymbol{v}} = \mathbb{E}\left[\mathcal{M}_{\boldsymbol{v}}^2(\boldsymbol{\Xi}_{\boldsymbol{v}})\right].$$

This leads to the definition of the Sobol' index $S_{\boldsymbol{v}}$, which denotes the amount of variance of Y attributed to perturbations of uncertainties that define subvector $\boldsymbol{\Xi}_{\boldsymbol{v}}$

$$S_{\boldsymbol{v}} = \frac{D_{\boldsymbol{v}}}{D}.$$

The total sensitivity indices S_i^T that account for the main effects and interactions among the parameters for a single parameter ξ_i are defined as

$$S_i^T = \sum_{I_i} \frac{D_{\boldsymbol{v}}}{D}, \quad I_i = \{\boldsymbol{v} \supset i\}.$$

Therefore, the total sensitivity index for a given parameter accounts for all the subsets that contain this parameter.

There are several approaches for accurately estimating the total Sobol' indices. The computational bottleneck, however, arises from the fact that the partial variances do not admit a closed-form solution in the general case of nonlinear systems, since they require recursive computations of

high-dimensional integrals. Thus, sample-based techniques, such as Monte Carlo sampling, can be utilized. A sample-based approach is costly for the application considered in this work, since a MILP problem must be solved for each uncertainty sample. We address this computational challenge using surrogate modeling. We seek to learn an approximation $\hat{\mathcal{M}}$ of the model $\mathcal{M}: \mathbb{R}^M \rightarrow \mathbb{R}$ by utilizing training data obtained from limited evaluations of the system model $\mathcal{M}$. The evaluation of the resulting surrogate model $\hat{\mathcal{M}}$ per sample should be significantly faster than $\mathcal{M}$; hence, accelerating sample-based UQ.

The experimental design for surrogate model training consists of N_{ed} samples $\widetilde{\boldsymbol{\Xi}} = \{\boldsymbol{\xi_1}, \boldsymbol{\xi_2}, ..., \boldsymbol{\xi_{N_{ed}}}\}$ with corresponding responses $\widetilde{\boldsymbol{Y}} = \{\mathcal{M}(\boldsymbol{\xi_1}), \mathcal{M}(\boldsymbol{\xi_2}), ..., \mathcal{M}(\boldsymbol{\xi_{N_{ed}}})\}$. The quantity and quality of the training data directly affects the approximation accuracy. A useful metric for evaluating the goodness of the surrogate model approximation is the so-called "leave-one-out" cross validation error ϵ_{LOO}

$$\epsilon_{LOO} = \frac{\sum_{i=1}^{N_{ed}} \left(\mathcal{M}(\boldsymbol{\xi}^i) - \hat{\mathcal{M}}^{\setminus i}(\boldsymbol{\xi}^i)\right)^2}{\sum_{i=1}^{N_{ed}} \left(\mathcal{M}(\boldsymbol{\xi}^i) - \mu_{\widetilde{\boldsymbol{Y}}}\right)^2},$$

where $\mu_{\widetilde{\boldsymbol{Y}}}$ is the empirical mean of the training data and $\hat{\mathcal{M}}^{\setminus i}$ denotes the surrogate model trained without taking the i^{th} data point into account.

In this work, polynomial chaos expansions (PCE) and Kriging are used for surrogate modeling. According to polynomial chaos theory (Xiu and Karniadakis, 2002), a finite-variance QoI denoted by Y can be expressed by an infinite expansion

$$Y = \sum_{\boldsymbol{a} \in \mathbb{N}^M} y_{\boldsymbol{a}} \Psi_{\boldsymbol{a}}(\boldsymbol{\Xi}), \tag{6}$$

where the basis functions $\Psi_{\boldsymbol{\alpha}}(\boldsymbol{\Xi})$ are multivariate polynomials, orthonormal with respect to $f_{\boldsymbol{\Xi}}$. That is, the condition

$$\mathbb{E}\{\Psi_{\boldsymbol{\alpha}}(\boldsymbol{\Xi})\Psi_{\boldsymbol{\beta}}(\boldsymbol{\Xi})\} = \int_{\mathcal{D}_{\boldsymbol{\Xi}}} \Psi_{\boldsymbol{\alpha}}(\boldsymbol{\xi})\Psi_{\boldsymbol{\beta}}(\boldsymbol{\xi}) f_{\boldsymbol{\Xi}} d\boldsymbol{\xi} = \delta_{\alpha\beta}, \quad \forall \boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{N}^M$$

holds over $\mathcal{D}_{\boldsymbol{\Xi}}$, which is the support of the joint distribution $\boldsymbol{\Xi}$, with $\delta_{\alpha\beta}$ denoting the Kronecker delta. The orthogonal polynomials can be defined based on probability distribution of uncertainties using the Wiener-Askey scheme (Xiu and Karniadakis, 2002), or can be directly constructed from data (Paulson et al., 2017).

The PCE (6) must be truncated up to a finite order described by the multi-index $\boldsymbol{a} \in \mathcal{A}$, where $\mathcal{A} \subset \mathbb{N}^M$ represents the set of multi-indices kept in the truncated expansion. The truncation scheme aims to limit the infinite expansion to a series of maximum order p, so that $\mathcal{A}^{M,p} = \{\boldsymbol{a} : |\boldsymbol{a}| \leq p\}$. Thus, the cardinality of $\mathcal{A}$ is equal to $P = \binom{M+p}{p}$. Further truncation of the expansion is introduced by employing the q-norm scheme, so that the set of multi-indices is given by

$$\mathcal{A}^{M,p,q} = \{\boldsymbol{a} \in \mathcal{A}^{M,p} : ||\boldsymbol{a}||_q \leq p\},$$
$$\left(||\boldsymbol{a}||_q = \sum_{i=1}^{M} a_i^q\right)^{\frac{1}{q}}.$$

Thus, the QoI can be approximated by the truncated PCE

$$\hat{Y} = \mathcal{M}^{PC}(\boldsymbol{\Xi}) = \sum_{\boldsymbol{a} \in \mathcal{A}} y_{\boldsymbol{a}} \Psi_{\boldsymbol{a}}(\boldsymbol{\Xi}). \tag{7}$$

Once a polynomial basis set is constructed using the q-norm scheme, the coefficients of the expansion (7) must be estimated. We take a sparse regression approach (Deman et al., 2016) to coefficient estimation, as non-intrusive approaches are more versatile and can treat the model $\mathcal{M}$ as a "black-box." We solve a regularized regression problem of the form

$$\hat{\boldsymbol{y}} = \text{argmin}_{y \in \mathbb{R}^P} \left(\mathbb{E}\left[\mathcal{M}(\boldsymbol{\Xi}) - \left(\boldsymbol{y}^T \Psi(\boldsymbol{\Xi})\right)^2\right] + \lambda ||\boldsymbol{y}||_1\right), \tag{8}$$

where the regularization term $||\boldsymbol{y}||_1 = \sum_{\boldsymbol{a} \in \mathcal{A}} \boldsymbol{y_a}$ results in a sparse approximation, since low order solutions are favored. Solving (8) requires estimating the regularization parameter $\lambda \geq 0$ that affects the number of terms with non-zero coefficient in (7). The problem of finding the regularization parameter can be efficiently solved using the least-angle-regression (LAR) algorithm (Efron et al., 2004). Although the q-norm scheme can reduce the cardinality of the basis set, we may still have to estimate a much larger number of coefficients than the available experimental design samples. LAR avoids this issue by using $\min(P, N_{ed} - 1)$ regressors.

In this work, we combine polynomial chaos with Kriging (Rasmussen, C., Williams, 2006). Polynomial chaos-Kriging (PCK) (Schobi et al., 2015) enhances the global approximation capability of PCEs by the local accuracy of Kriging (i.e., Gaussian process regression). PCK approximates the predictions of model $\mathcal{M}$ as a realization of a Gaussian process, so that the surrogate model takes the form

$$\hat{Y} = \mathcal{M}^{PCK}(\boldsymbol{\xi}) = \boldsymbol{\beta}^T \boldsymbol{f}(\boldsymbol{\xi}) + \sigma^2 Z(\boldsymbol{\xi}). \tag{9}$$

The first term in the PCK model (9) describes the trend of the Gaussian process, which is given by (7)

$$\boldsymbol{\beta}^T \boldsymbol{f}(\boldsymbol{\xi}) = \sum_{\boldsymbol{a} \in \mathcal{A}} y_{\boldsymbol{a}} \Psi_{\boldsymbol{a}}(\boldsymbol{\xi}). \tag{10}$$

The zero-mean Gaussian process $Z(\boldsymbol{\xi})$ in (9) is completely determined by the so-called kernel function that defines a pairwise correlation between input samples based on their distance, defined as $R(\boldsymbol{\xi}, \boldsymbol{\xi'}) = R(|\boldsymbol{\xi} - \boldsymbol{\xi'}|; \boldsymbol{\theta})$; see Makrygiorgos et al. (2020) for further details. The hyperparameters $\boldsymbol{\theta}$ of the kernel function must be estimated along with the aforementioned unknowns that characterize the Gaussian process. $\boldsymbol{\beta}$ and σ are explicit functions of the hyperparameters $\boldsymbol{\theta}$ and the training data. These parameters can be estimated via maximum likelihood estimation (Marrel et al., 2008). Here, the PCK surrogate model (9)-(10) is trained using the sequential algorithm in Schobi et al. (2015). For every uncertainty sample, the prediction of the PCK model for the QoI Y consists of a mean value $\mu_{\hat{Y}}$ and a standard deviation $\sigma_{\hat{Y}}$.

4. RESULTS AND DISCUSSION

We apply the MILP model of Section 2 to cultivate rice, wheat, potato, and lettuce to meet daily food demands for a Mars mission that lasts 220 days. The MILP model has 2,665 continuous, 882 integer, and 4 binary variables. A single run of the MILP model takes approximately one minute when solved with the Python interface of GUROBI solver 8.1.0 on a machine with Intel Xeon(R) CPU @ 3.50 GHz with 31.3 GB RAM. We consider 86 uncertain parameters that follow independent uniform distributions.

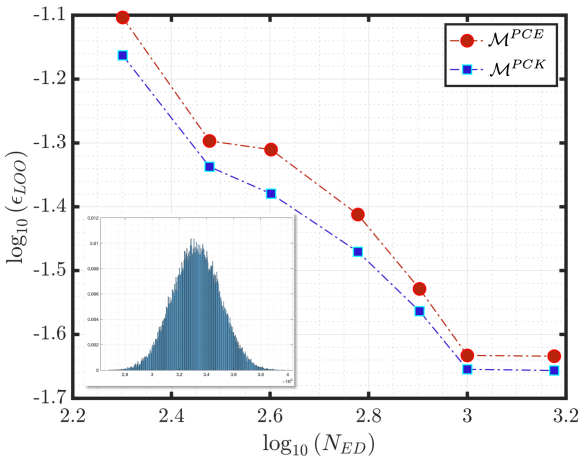

Fig. 1. Leave-one-out cross validation error ϵ_{LOO} of the sparse PCE (red) and PCK (blue) surrogate models for different sizes N_{ED} of experimental design used for surrogate model training. The inset shows the probability distribution of the minimum ESM, estimated using the PCK surrogate model trained with 1,000 samples, on a validation set.

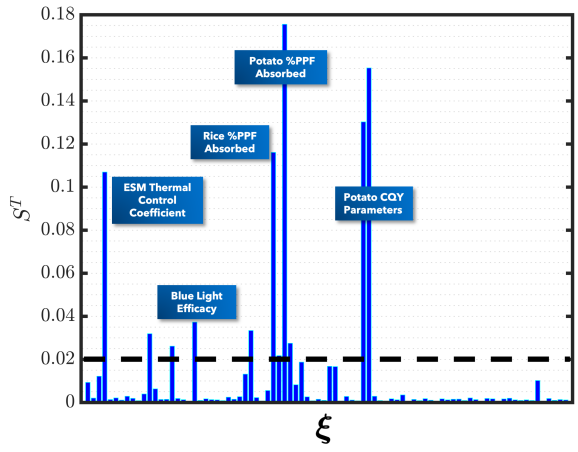

Fig. 2. Total Sobol' indices that quantify the sensitivity of the minimum ESM to the 86 uncertain MILP model parameters. The black dashed line is S^T_{min} threshold.

4.1 Building the Surrogate Models

Surrogate models were constructed using the UQLab toolbox in Matlab. Constructing PCEs with a moderate basis order of 4 or 5 for the 86 uncertain parameters leads to an expansion with $P \approx \mathcal{O}(10^6)$ terms, which is intractable. The order-adaptive LAR algorithm adopted in this work for building the sparse PCE and PCK surrogate models allows for optimal basis selection with limited training data. For the PCK surrogate model, we used the Matern 3-2 kernel. We constructed surrogate models for the minimum ESM using different experimental design sizes. Fig. 1 shows the estimated ϵ_{LOO} for each surrogate model as a function of experimental design size N_{ED}. We observe that ϵ_{LOO} decreases rapidly after adding about 300 experimental design points, reaching a plateau thereafter.

Fig. 1 suggests that sparse PCE and PCK surrogate models yield comparable approximation accuracy. However, the PCK surrogate model provides confidence intervals for its predictions, which is useful in applications where quantifying prediction uncertainty is important. After constructing an accurate surrogate model for fast estimation of the minimum ESM, we performed Monte Carlo sampling to estimate the probability distribution of the minimum ESM. The estimated probability distribution is inset in Fig. 1. The probability distribution was constructed using 100,000 samples with the PCK surrogate model, with computations that took less than 15 s. An identical number of MILP model evaluations would have taken about 10^5 minutes. Thus, the computational speed gains of the surrogate model are approximately 100-fold after accounting for the time it took to build the model.

4.2 Global Sensitivity Analysis

The PCK surrogate model was used to perform global sensitivity analysis of the minimum ESM. Fig. 2 shows sample-based estimates of the Sobol' indices that quantify the sensitivity of the minimum ESM to the 86 uncertain model parameters. As is typical of systems with high uncertainty dimension, not all uncertain inputs have a major effect on model responses (Deman et al., 2016). Here, we set a threshold of $S^T_{min} = 0.02$, so that uncertain parameters with sensitivity indices below S^T_{min} are deemed insignificant. Hence, a total of 12 of the 86 parameters have a significant impact on the minimum ESM. We observe that the minimum ESM is most sensitive to the following parameters in descending order: the fraction of PPF absorbed by potato, two of the several empirical parameters that determine edible potato yield, the fraction of PPF absorbed by rice, the ESM coefficient for thermal control, the blue light efficacy, the minimum rice demand, the ACR thermal control demand, several of rice and wheat empirical parameters, and the PPF incident on lettuce.

The fraction of PPF absorbed by each of the crops significantly impacts the minimum ESM because this parameter directly affects biomass yield and lighting requirements. A greater absorbed PPF increases biomass yield, while decreasing lighting needs. Blue light parameters are unsurprisingly critical. In fact, it was the chosen light alternative among the four available options for most of the training data. Potato yield parameters are dominant, since its maturity time is the highest of all four crops. Thus, a change in potato yield has substantial impact on potato shipments, compared to similar impacts for other crops.

The Sobol' indices in Fig. 2 allow us to evaluate the relative impact of the uncertain parameters on system design decisions. The Sobol' indices also suggest where further priorities in mission planning and engineering of the life support system must be placed. For instance, the results indicate that future space plant experimental effort should emphasize better estimation of the absorbed PPF fraction, and of the empirical coefficients related to crop growth equations. Additionally, the ACR thermal control parameters should be estimated accurately due to their predominance in computed ESM.

4.3 Estimating the Distribution of Decision Variables

The solution of the MILP model comprises optimal values for the decision variables of the life support system, such as the amount of food or nitrogen that must be shipped from

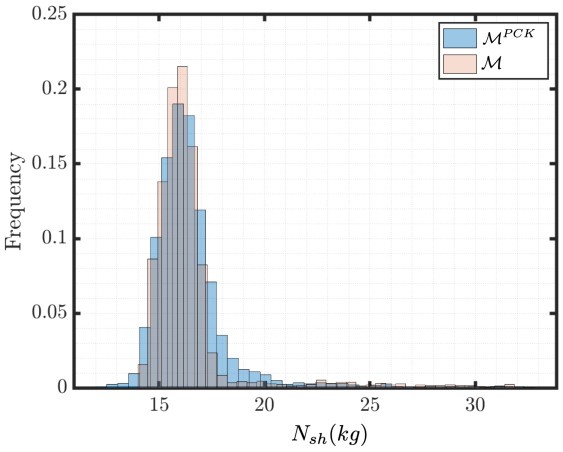

Fig. 3. Estimated distribution of the optimal amount of nitrogen, N_{sh}, shipped from Earth. The pink distribution is estimated from the response of the MILP model, whereas the blue distribution is estimated by the PCK surrogate model.

Earth. Fig. 3 shows the estimated probability distribution for the optimal amount of shipped nitrogen, N_{sh}. This distribution exhibits a wide variability that encompasses high values of nitrogen demand due to the nature of the decision-support problem. The consideration of integer and binary variables within the MILP model leads to multiple scenarios that yield disparate optimal values for the decision variables, resulting in a wide range of scenarios to analyze for mission design. The global sensitivity analysis provides a first step towards a better estimation of model parameters to mitigate uncertainty in the mission design by reducing the number of mission scenarios to consider.

5. CONCLUSIONS

We demonstrate the usefulness of surrogate modeling for accelerating forward uncertainty quantification tasks on a MILP model that optimizes food production for life support during a Mars exploration mission. We present a sparse polynomial chaos-Kriging surrogate model for this MILP model that has 86 uncertain parameters. The computational time needed to evaluate 100,000 samples of uncertain parameters using the surrogate model is about 15 seconds, approximately $2.5 \times 10^{-4}\%$ of the time needed to perform the same number of MILP model evaluations. After accounting for the time required to generate data for surrogate model training and the training procedure itself, the computational speed gains of the surrogate model are over 100-fold. This speed-up in computations enables global sensitivity analysis, which shows that only 12 out of 86 uncertain parameters in the MILP model drive variability in a quantity of interest. Eliminating uncertainty in these parameters is crucial, given their drastic impact on key decision variables in Mars mission planning.

REFERENCES

Anderson, M.S., Ewert, M.K., and Keener, J.F. (2018). Life Support Baseline Values and Assumptions Document (BVAD). Technical Report NASA/TP-2015218570/REV1, NASA.

Blatman, G. and Sudret, B. (2013). Sparse polynomial chaos expansions of vector-valued response quantities. In *Proceedings of the 11th International Conference on Structure Safety and Reliability*. New York.

Deman, G., Konakli, K., Sudret, B., Kerrou, J., Perrochet, P., and Benabderrahmane, H. (2016). Using sparse polynomial chaos expansions for the global sensitivity analysis of groundwater lifetime expectancy in a multi-layered hydrogeological model. *Reliability Engineering and System Safety*, 147, 156–169.

Do, S., Owens, A., Ho, K., Schreiner, S., and de Weck, O. (2016). An independent assessment of the technical feasibility of the Mars One mission plan – Updated analysis. *Acta Astronautica*, 120, 192–228.

Drake, B.G. (2009). Human Exploration of Mars Design Reference Architecture (DRA) 5.0. Technical Report NASA/SP2009566, NASA.

Efron, B., Hastie, T., Johnstone, I., Tibshirani, R., Ishwaran, H., Knight, K., Loubes, J.M., Massart, P., Madigan, D., Ridgeway, G., Rosset, S., Zhu, J.I., Stine, R.A., Turlach, B.A., Weisberg, S., Johnstone, I., and Tibshirani, R. (2004). Least angle regression. *Annals of Statistics*, 32(2), 407–499.

Levri, J., Fisher, J.W., Jones, H.W., Drysdale, A.E., Ewert, M.K., Hanford, A.J., Hogan, J.A., Joshi, J., Vaccari, D.A., et al. (2003). Advanced Life Support Equivalent System Mass Guidelines Document. Technical Report NASA/TM-2003-212278, NASA.

Makrygiorgos, G., Maggioni, G.M., and Mesbah, A. (2020). Surrogate modeling for fast uncertainty quantification: Application to 2D population balance models. *Computers & Chemical Engineering*, 138, 106814.

Marrel, A., Iooss, B., Van Dorpe, F., and Volkova, E. (2008). An efficient methodology for modeling complex computer codes with Gaussian processes. *Computational Statistics and Data Analysis*, 52(10), 4731–4744.

Menezes, A.A., Cumbers, J., Hogan, J.A., and Arkin, A.P. (2015). Towards synthetic biological approaches to resource utilization on space missions. *Journal of the Royal Society Interface*, 12, 20140715.

Paulson, J.A., Buehler, E.A., and Mesbah, A. (2017). Arbitrary polynomial chaos for uncertainty propagation of correlated random variables in dynamic systems. *IFAC-PapersOnLine*, 50, 3548–3553.

Paulson, J.A., Martin-Casas, M., and Mesbah, A. (2019). Fast uncertainty quantification for dynamic flux balance analysis using non-smooth polynomial chaos expansions. *PLOS Computational Biology*, 15(8), e1007308.

Paulson, J.A. and Mesbah, A. (2019). An efficient method for stochastic optimal control with joint chance constraints for nonlinear systems. *International Journal of Robust and Nonlinear Control*, 29(15), 5017–5037.

Rasmussen, C., Williams, K. (2006). *Gaussian Processes for Machine Learning*. MIT Press.

Schobi, R., Sudret, B., and Wiart, J. (2015). Polynomial-chaos-based Kriging. *International Journal for Uncertainty Quantification*, 5(2), 171–193.

Sobol, I.M. (2001). Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation*, 55, 271–280.

Xiu, D. and Karniadakis, G.E. (2002). The Wiener-Askey polynomial chaos for stochastic differential equations. *SIAM Journal on Scientific Computing*, 24(2), 619–644.