AAS/AIAA 18-296

STATISTICAL APPROACHES TO INCREASE EFFICIENCY OF LARGE-SCALE MONTE-CARLO SIMULATIONS

David Shteinman,[*] Thibaud Teil,[†] Scott Dorrington, [‡] Huan Lin, [§]

Thomas Dixon, [**] Hanspeter Schaub,[††] John Carrico,[‡‡] and Lisa Policastri[§§]

Numerical astrodynamics simulations are regularly characterized by a large input space and complex, nonlinear input-output relationships. Monte Carlo runs of these simulations are typically time-consuming and numerically costly. In this paper the Design and Analysis of Computer Experiments (DACE) approach is adapted to astrodynamics simulations to improve runtimes and increase information gain. Two case studies are presented: a satellite detumbling simulation using the BASILISK software, and orbit trajectory simulations in the IBEX-extended mission. The space-filling and meta-modelling techniques of DACE are shown to provide significant improvements for astrodynamics simulations in speed of sensitivity analysis, determination of outliers and identifying extreme output cases not found by standard simulation and sampling methods.

INTRODUCTION

Many numerical astrodynamics analyses are characterized by a large input space with dispersions on those inputs. They also require numerical integration to propagate orbital trajectories, as well as the spacecraft attitude and actuator states forward in time. Often, Monte Carlo simulations are used, where each sample point is propagated numerically. These features all contribute to long Monte Carlo simulation times. Furthermore, the underlying input-output relationships are nonlinear with many variables interacting with one another. Hence, it is difficult to study the behavior in simulation of output responses as a function of the inputs - as that requires testing of a wide range of input values. Using traditional methods of varying one factor at a time and re-running the whole simulation for each instance is excessively time consuming. In addition, varying one factor at a time means the end user of the simulation's results cannot be certain they have captured the full range of possible input values.

The aim of this paper is to adapt a method for astrodynamics simulations from industrial statistics and empirical modeling, to achieve the following outcomes:

1) Significantly reduce the run time of large-scale Monte Carlo simulations;

2) Ensure the simulation covers a wider range of values/worst case scenarios for significantly less runs than required under standard Monte-Carlo methods;

[*] Industrial Sciences Group, Sydney, Australia
[†] Autonomous Vehicles Systems (AVS) Laboratory, University of Colorado, Boulder
[‡] Industrial Sciences Group, Sydney, Australia
[§] Industrial Sciences Group, Sydney, Australia
[**] Industrial Sciences Group, Sydney, Australia
[††] Autonomous Vehicles Systems (AVS) Laboratory, University of Colorado, Boulder
[‡‡‡‡] Space Exploration Engineering Inc.
[§§] Space Exploration Engineering Inc.

3) Increase the efficiency of Sensitivity Analysis and Optimization by using a fast/computationally cheap approximate model of the simulation, thus avoiding the need to re-run the simulation to test the effect of alternate input values on the output.

To achieve outcomes 1-3, we adapt the techniques of *Design & Analysis of Computer Experiments Techniques (DACE)* to astrodynamics simulation and illustrate with two Case Studies.

DESIGN & ANALYSIS OF COMPUTER EXPERIMENTS TECHNIQUES (DACE)

DACE is an adaptation for simulations of the DOE approach (*Design of Experiments*) used in physical experiments on industrial processes. DOE is a statistical method of experimentation that *varies all inputs in a simulation simultaneously* (rather than one factor at a time) and achieves the following:

- Determines the critical inputs (those with the largest effect on outputs of interest)
- Quantifies the input/output relationships in an analytical form within the experimental range
- Shows interactions between inputs.

The method of DACE (see Reference 1) has been used extensively in the automotive and other manufacturing sectors, but so far as we know, not in astrodynamics. DACE is used to augment the limited number of runs of a simulation by fitting an approximate statistical model – a surrogate or "Meta-Model" – based on a set of limited observation data acquired by running the simulation at carefully selected design points, generated from a "Space Filling" design. The Meta-model is easier and faster to run than the simulation and is used to *predict the simulation performance at unobserved input values.* The approximate Meta-model is much simpler than the true one: it approximates the original model as closely as possible and is computationally cheap to compute**.** The Meta-model may be used for sensitivity analysis, optimization and prediction *without needing to re-run the simulation.*

This paper investigates the applicability of DACE to two astrodynamics simulations that use Monte Carlo runs: The first is aimed at predicting the results of an attitude control simulation using the Basilisk program developed by the *Autonomous Vehicle Simulation (AVS) Laboratory at University of Colorado at Boulder*.* The second is on using the sampling aspects of DACE in simulations of trajectory design for the Interstellar Boundary Explorer (IBEX) spacecraft - to seek out performance results not previously observed.

In both case studies, we use a five-step process to adapt DACE to astrodynamics simulations (see Reference 2)**:** Step 1) Factor Screening; 2) Design the Simulation "Experiment**"** using a space -filling design; 3) Build a Meta-model; 4) Validating/Checking the Meta-model; 5) Using the Meta-model for Sensitivity Analysis. This process is demonstrated in Figure 1.

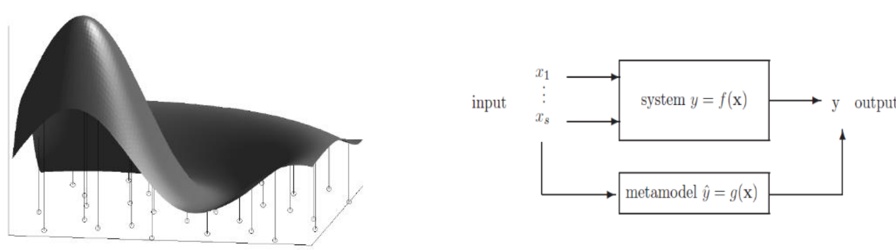

Figure 1. Schematic showing concept of a Space-Filling Design and Meta-model as used in Design & Analysis of Computer Experiments Techniques

The efficiency gain using DACE is highest when repeatedly running a large Monte-Carlo is a time-consuming process and unfeasible when large number of scenarios are to be tested. Also, DACE is highly effective in cases where it is required to quantify the effect of uncertainties and errors in inputs on some measure of performance.

* http://hanspeterschaub/bskMain.html

SAMPLING METHODS

Monte Carlo (MC) sampling is commonly used to obtain a sample space in a computer simulation. However, as pointed out in Reference 3, while MC sampling is simple to use, a set of MC samples will often leave *large regions of the design space unexplored*. This occurs due to the random and independent nature of the sample sites produced by a random number generator.

To address this problem, this paper tests a range of "Space Filling" designs to generate sampling schemes that capture the maximum information between the input-output relationships. The simulation and modeling performance are compared using five sampling methods:

i. **Random Sampling** - each input parameter is drawn randomly from the distribution.

ii. **Hypercube** - the 2^n vertices of an *n*-dimensional hypercube. Hypercube sampling only samples from the vertices of the resulting *n*-dimensional hypercube and thus only considers data that are "rare". Hence it only focuses on the extreme values/events and ignores the rest of the design space. Sampling from the vertices only can be very fast and minimize the experimental runs. However, if the underlying interest lies in the centre of the design space then sampling method such as hypercube will produce samples of limited value.

iii. **Latin Hypercube Sampling (LHS**) - LHS is a design that is space-filling in every dimension of the input space that if collapsed in any one or more dimensions, would not result in duplicate test points. LHS has the chief advantage of spreading its runs out over the entire design space. The other main features of LHS are described in detail by References 3-5. In brief: Given an equal number of samples, an LHS estimate of the mean will have less error than the mean value obtained by Monte Carlo sampling. Also, LHS is superior to simply creating a grid over the entire design space as, due to the *Sparsity of Effects principle*, few of the many inputs in a model will prove to be statistically significant. When a computer model exemplifies effect sparsity, variation in the response *Y* will be significant only when those few significant inputs are varied. If the significant inputs are kept constant and other inputs varied, the output *Y* will vary by a negligible amount (Reference 4, p. 14).

iv. **Maximin sampling**: Draws a Latin Hypercube Sample from a set of uniform distributions for use in creating a Latin Hypercube Design. This function attempts to optimize the sample by maximizing the minimum distance between design points (maximin criteria).

v. **Maximum Projection Sampling**: Draws a Latin Hypercube Sample from a set of uniform distributions for use in creating a Latin Hypercube Design. This function attempts to optimize the sample by the maximum projection (MaxPro) criterion: the average reciprocal product of squared one-dimensional distances should be maximised (Reference 5).

Figure 2a and 2b show examples of sample points generated from these methods over a 2D input space.

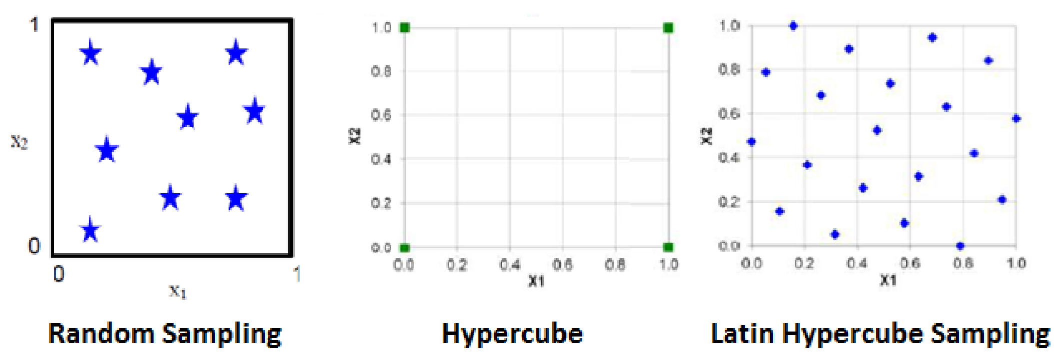

Fig. 2a Comparison of Space-Filling Designs in 2-Dimesnional Space

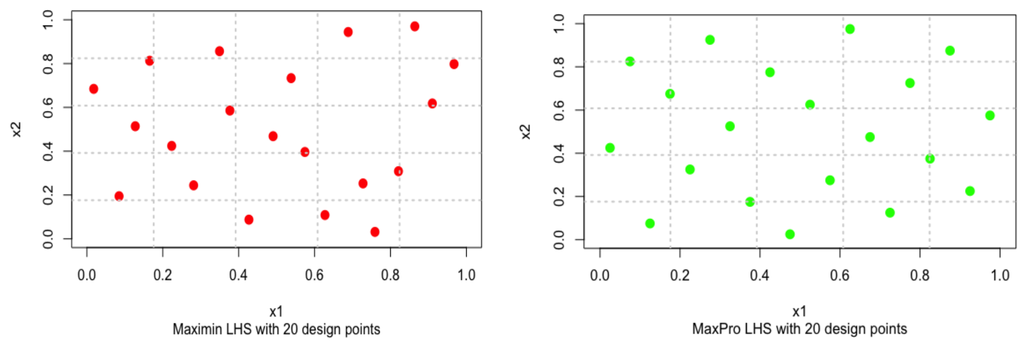

Figure 2b. Comparison of Modified LHS Space-Filling Designs in 2-Dimesnional Space (References 3,4,5)

SURROGATE OR "META-MODELS"

The objective of building a Meta-model is to construct an accurate, simple and easy to run model that approximates the true model of the simulation process.

The initial runs of a sampling design are the "Screening" stage (Reference 6) where, due to the *Principle of Sparsity of Effects*, only a few input variables $X_1, .., X_k$ are identified as important for determining the target variable Y of interest. Each run of the simulation model for specific values $(X_1^*, .., X_k^*)$ will generate a model value $Y^* = m(X_1^*, .., X_k^*)$, where $m(\cdot)$ is an unknown function.

The Meta-model is determined by the following two steps:

1. Sampling from a Space Filling Design: Generate a set of input values $(X_1^1, .., X_k^1), ..., (X_1^n, .., X_k^n)$ which is representative of the whole input space and compute the corresponding values $Y^1 = m(X_1^1, .., X_k^1), ..., Y^n = m(X_1^n, .., X_k^n)$.

2. Model Development: On the basis of the computed pairs of function values and input values Y^i, $(X_1^i, .., X_k^i)$ approximate the unknown function $Y = m(X_1, .., X_k)$.

For Step 2, four main types of Meta-modeling techniques are compared and contrasted:

1. **Parametric models e.g. Second-order Polynomials:** These models employ a polynomial basis $x_1^{r_1} \dots x_s^{r_s}$ where $r_1, \dots, r_s$ are non-negative integers. The number of polynomial basis functions dramatically increases with the number of input variables and the degree of polynomial. (Reference 1, p.27). For our Meta-modelling purposes, we note an issue with Polynomials is that for modelling large and complex simulations there may be many local minimums/maximums within the design space. This requires a high-degree polynomial to approximate the true model (Reference 1, p.27). In such cases collinearity between inputs becomes a serious problem.

2. **Radial Basis Functions:** Radial basis function (RBF) methods are techniques for exact interpolation of data in multi-dimensional space (Reference 7 and Reference 1, p.177-179). The RBF maps the inputs to the outputs using a linear combination of the basis functions. We tested an RBF interpolation with cubic and thin plate basis functions.

3. **Random Forests:** Random forests have no formal distributional assumptions (hence a "non-parametric" regression), and automatically performs screening/variable selection. This is very useful in astrodynamics simulations where there a large number of input variables and need to reduce the dimensionality to the critical inputs with most significant effect on the output response. Random forests can also handle highly non-linear interactions.

4. **Kriging:** Also known as *Gaussian Process Regression*, Kriging is a non-parametric method of interpolation where the interpolated values are modelled by a Gaussian process governed by prior covariances. The general goal of Kriging is to predict the value of an underlying random function Y=Y(x) at any random location of interest x0. The main idea of Kriging is that near sample observed points should get more weight in the prediction to improve the estimate. Therefore, Kriging relies on the knowledge of some kind of spatial distance structure of the sample points, which is in turned modelled via the second-order properties, i.e. the prior covariance matrix. Kriging uses a form of Bayesian inference that produces both a deterministic prediction and a standard error that can be used to quantify confidence intervals.

The alternative modelling methods are assessed in terms of making satisfactory predictions at untried points. The chosen Meta-model is verified by using the *Root Mean Square Error* (RMSE) of prediction at untried points, which represents the departure of the Meta-model from the true model. The smaller the RMSE value, the better the Meta-model.

CASE STUDY 1: ATTITUDE CONTROL SIMULATION

The first case study applies the methods of DACE to an attitude control simulation of a tumbling spacecraft using the Basilisk software package developed by the *Autonomous Vehicle Simulation (AVS) Laboratory at University of Colorado at Boulder*. For a given initial state (attitude and rotation rate) and spacecraft design parameters, the software simulates a sun-pointing feedback control problem to determine the motor torques controls of the three orthogonal Reaction Wheels (RWs) to achieve a final target state, with zero rotation rate.

The goal of the simulation is to assure that, given a range of initial states and uncertainties of real-life missions, a specific spacecraft design can achieve the de-tumbling maneuver while fulfilling a set of mission requirements. For this scenario the following mission requirements are assumed for an acceptable maneuver:

- The attitude will enter and remain within 1º of the goal frame within 5 minutes
- The RW speeds will not exceed 3000 RPM during the manoeuvre.

Inputs & Dispersions

The model contains 29 inputs, describing the initial state of the spacecraft, the initial rotation rates of the reaction wheels, and a set of spacecraft parameters. Each input has dispersions with either uniform or normal distributions (detailed in Table 1).

Table 1. Dispersions on the 29 inputs.

Input Factor	Description	Distribution	Parameters	Units
Inertial Attitude (x,y,z)	Initial attitude of spacecraft (Modified Rodrigues Parameters with respect to inertial frame)	UNIFORM	[0,2π] (on each component)	rad
Inertial Rotation rate (x,y,z)	Initial rotation vector in inertial frame.	NORMAL	Mean 0, std dev 0.25 deg/s (on each)	deg/s
Mass	Mass of the Spacecraft	UNIFORM	Bounds [712:5; 787:5], mean 750	kg
Centre of Mass Offset (x,y,z)	Position of Centre of mass	NORMAL	Mean [0,0,1], std dev [0.0017,0.0017,0.0033]	m
Inertia Tensor	3x3 diagonal inertia tensor	NORMAL	Mean diag(900,800,600), std dev diag(0.1,0.1,0.1)	kg.m^2

Inertia Tensor Rotation Angle (x,y,z)	Rotation of inertia tensor by 3 Euler angles	NORMAL	Mean 0, std dev 0.1 deg (on each)	deg
RW1 axes	Spin axis of reaction wheel 1	NORMAL	Mean [1,0,0], std dev [0.0033,0.0017,0.0017]	-
RW2 axes	Spin axis of reaction wheel 2	NORMAL	Mean [0,1,0], std dev [0.0017,0.0033,0.0017]	-
RW3 axes	Spin axis of reaction wheel 3	NORMAL	Mean [0,0,1], std dev [0.0017,0.0017,0.0033]	-
RW speeds	Initial rotation speed of the 3 RWs	UNIFORM	RW1: [95,105], RW2: [190,210], RW3: [285,315]	RPM
Voltage to Torque Gain	Gain in commanded voltage to actual torque of RWs	UNIFORM	Bounds: [0.019,0.021], Mean 0.020	-

Outputs

The performance of the de-tumbling maneuver is determined by five output parameters: the rotation rates of the three RWs, and the times taken for the attitude and rotation rate to settle to their targeted states. The main output of interest is *attitude settle time*.

Sampling

To cover the input space, 1000 sample points are produced using the LHS sampling method with dispersions on the 29 parameters. For each sample point, time-series data is produced for the five output parameters over the duration of the de-tumbling maneuver, from which the maximum values were determined. Figure 3 shows a time-series plot of the attitude error (from the targeted attitude). The attitude settle time is determined from the time at which the attitude error levels off to near-zero.

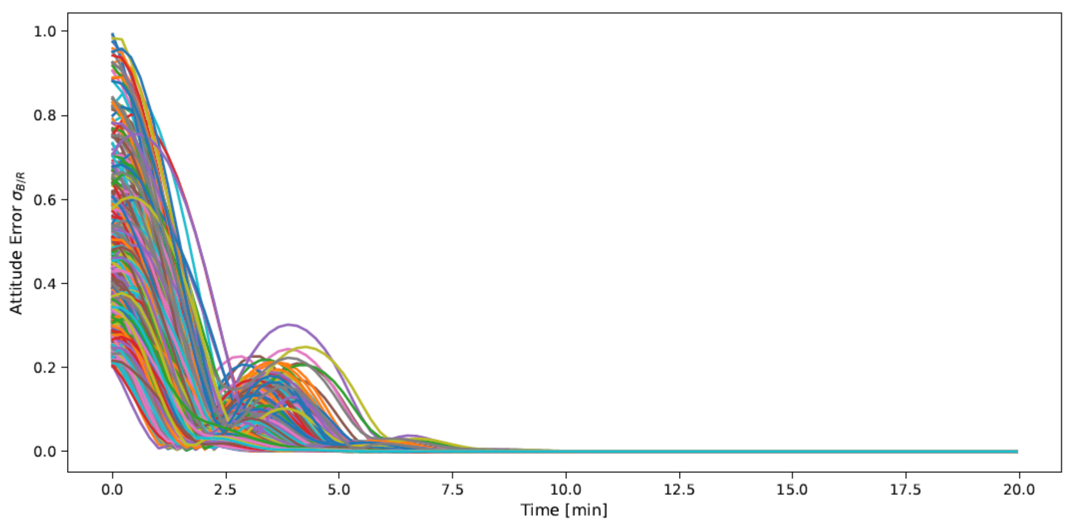

Figure 3. Time-series data of the spacecraft attitude for the 1000 sample points.

Initial Meta-Models

The inputs and outputs from these 1000 points are used to produce a Meta-model that is used to predict the outputs of any untested point in the input space. Several types of models are tested: non-parametric (Radial Basis Functions, Random Forests and Kriging) and parametric models (e. g Linear Regression, Quadratic Polynomials).

Random Forest (RF) is an effective machine learning model for predictive analytics. It was initially chosen due to its ability to simultaneously perform automatic variable selection on the large number of inputs (29) in the attitude control simulation. As RF is a non-parametric method it does not assume any distributions, making it ideal for a large and complex simulations (as in astrodynamics) characterized by a large input space, and an unknown input/output relationship that is expected to be highly non-linear.

The first RF model used was a *full model*, generated from all 29 input variables. The relative importance of each of the 29 input variables can be displayed using a Variable Importance plot (shown in Figure 4), where an output variable (in this case attitude settle time) is regressed against each input variable. From this figure, the initial attitude, rotation state and RW speeds were found to be the most significant contributors to the predicted attitude settle time. These results were used to decrease the dimensionality of subsequent models.

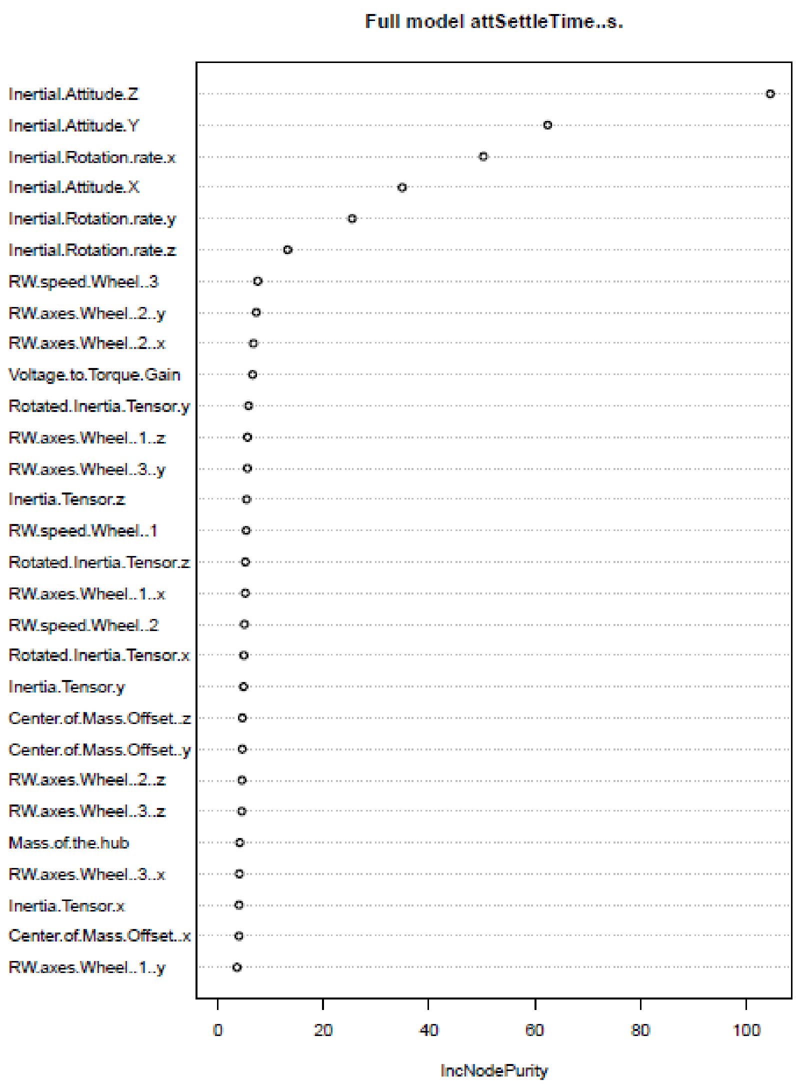

Figure 4. Variable selection plot showing the significance of each input

Two additional random forest models are generated using 6 and 7 inputs: the 6 most significant variables (Figure 4), and the top 6 plus Hub Mass, respectively. The addition of the mass parameter to the modelling is based on its apparent importance to simulation results, according to co-authors' advice.

To test the predictive power of each method, 1000 points are sampled for each input space, and produced a Meta-model. A '*leave-p*-out' *cross-validation* (where p=100) is used to build each meta-model: 900 randomly selected pairs of inputs and outputs are included in the training set to build each meta-model; then predict the remaining 100 outputs in the test set and calculate their residuals (compared to the actual results generated from the simulation), along with the RMSE of the model.

Figure 5 shows a plot of the actual vs the predicted *attitude settle time* for the 100 test points for the three Random Forest models, plus a Quadratic Polynomial model. From the plots, there is a noticeable spread in the predicted attitude settle time. Table 2 below shows the RMSE for 4 Meta-models and a Kriging polynomial model. The "best" model (lowest RMSE) was the Random Forest model of 6 inputs.

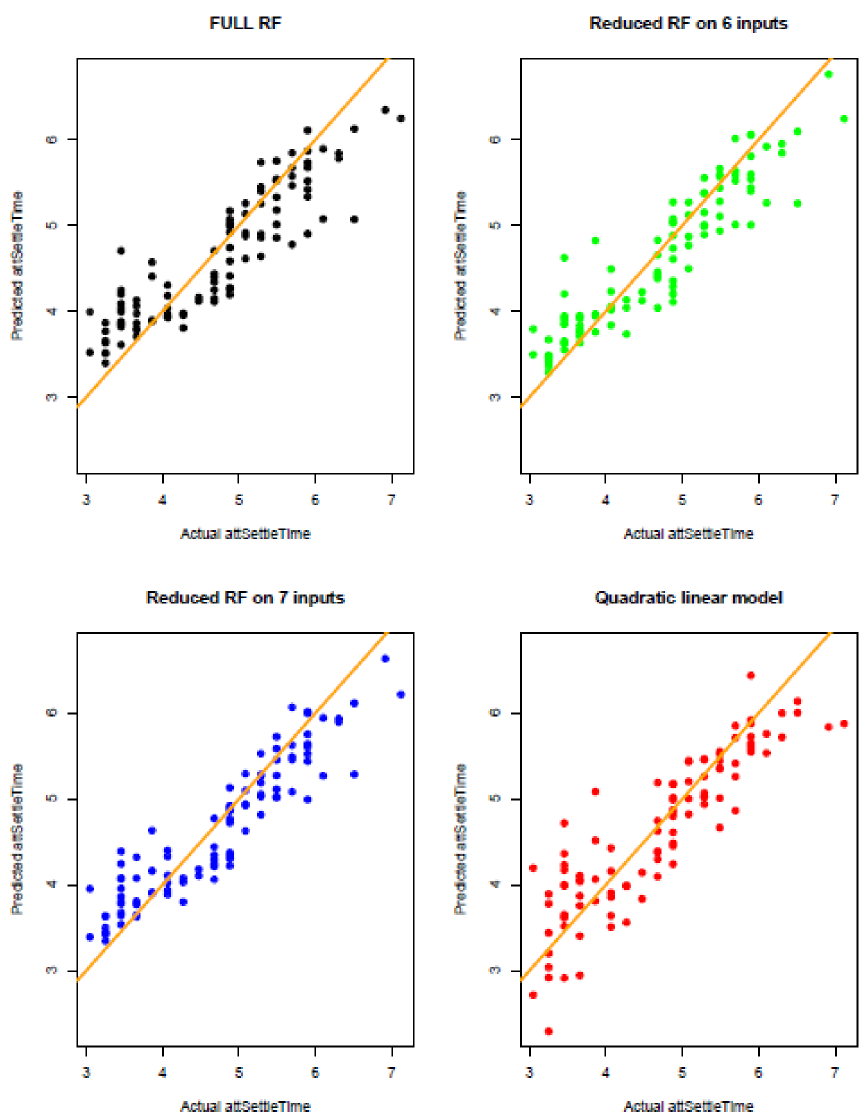

Figure 5. Predicted vs. Actual outputs for various Meta-Models

Table 2. RMSE for various Meta-Models

MODEL TYPE	RMSE
Quadratic linear model based on 7 inputs	0.4685
Kriging full model	0.384
RF full model	0.357
RF reduced 6 inputs model	0.3178
RF reduced 7 inputs model	0.3186

Alternate Sampling Plans & Sample Size

In order to improve our predictive capabilities the effect of different input sampling schemes (LHS, MaxPro and Maximin) on the accuracy of subsequent meta models at a large sample of 10,000 runs is explored. The dimensionality of each meta model is increased to 10 input parameters, to ensure symmetry across all three spatial dimensions – accounting for **all** initial attitudes, rotation rates and reaction wheel speeds (total of 9 inputs), as well as the mass of the spacecraft. RMSEs across all five output dimensions for Random Forest (RF), Radial Basis Function (RBF) and Kriging models using each of these three sampling plans are presented below in Table 3.

Table 3. RMSE for alternate meta models and sampling plans.

MODEL TYPE	Sampling TYPE	RMSE
10 Input RF	Latin Hypercube Sampling	0.075
	Maximin Sampling	0.066
	MaxPro Sampling	0.118
RBF	Latin Hypercube Sampling	0.585
	Maximin Sampling	0.655
	MaxPro Sampling	0.774
Kriging	Latin Hypercube Sampling	0.211
	Maximin Sampling	0.205
	MaxPro Sampling	0.200

Comparing the results of the models using a Maximin LHS sampling method, both RF and Kriging lead to better (lower RMSE) results for *attitude Settle Time* and *Rate of Settle time*. Random Forest models gives the lowest RMSE when the sampling scheme is based on Maximin LHS

Exploratory Data Analysis

In addition to producing a meta model, exploratory data analysis was performed on the results produced from the original 1000 sample inputs to identify output values that exceeded the mission requirements. Figure 6 shows a histogram and box and whiskers plot of the actual attitude and rate settle times. The results show that 95% of the sample points had attitude settle times within the 5-minute constraint.

These results suggest that the maneuver does not meet a 3-sigma success rate per the requirements. During mission design and testing, this analysis provides critical insight on the bounding cases and on the sensitivity to input parameters. This leads to a better understanding of the spacecraft's dynamics and capabilities. In this specific scenario, the simulation would first be validated, and the inputs would be examined to ensure realistic values for this maneuver. If the success rate remains low, the design of the maneuver or the spacecraft could be modified. The requirements could also be put into question as slightly relaxing the required attitude settle time would yield positive results.

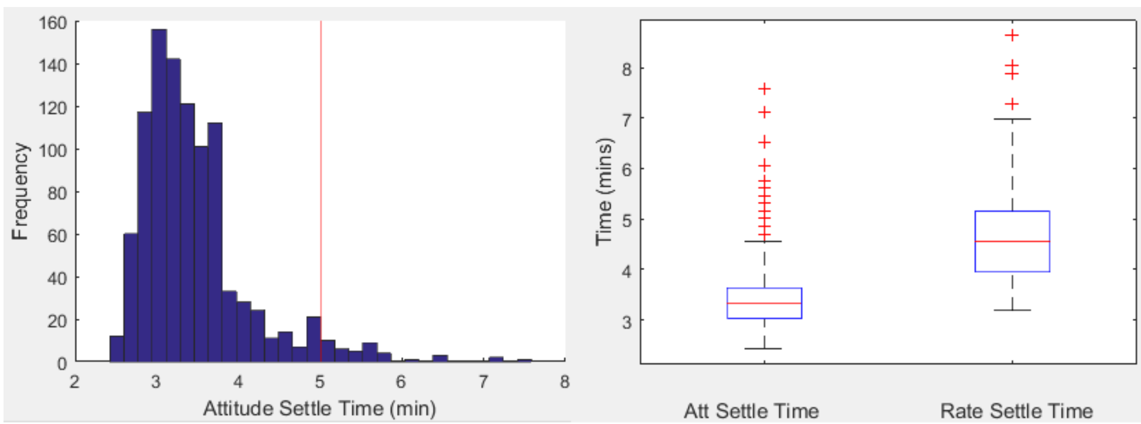

Figure 6. Histogram and Box and Whiskers plot of attitude settle time for 1000 sample points

Using the Meta-Models for Sensitivity Analysis

Further testing of the RF Meta-model tested input values expected to lead to extreme output results that may exceed mission requirements. This Sensitivity Analysis will be used as a low cost /rapid development of the operating window of the entire system under a wide range of worst case scenarios.

In particular, the Meta-model is used to search regions in the input space that exceed the constraints. By analyzing the outliers in the data –regions may be discovered around the outlier points, where all points within that region will exceed the bounds.

Scenario 1: Prediction Using Extreme Input Values.

The analysis of outliers described above are limited to the output values produced from running the finite sample inputs. While the increased 10,000-point set produced additional outliers, the computation time increased significantly (by approximately 8 hours). Running additional sample points to further explore regions of interest would lead to long run times. The Meta-model produced may be used to rapidly explore the input space, without needing to re-run the simulation.

Using the RF Meta-model, a Sensitivity Analysis is performed to test the range of outputs *within a wider input space* to see if the mission requirements are still met. Sets of points were generated to test extreme input cases that may be encountered during a mission. Two sets of extreme points were tested in the initial analysis:

A. Initial speeds of 500 RPM on all 3 RWs, as well as a +3-sigma variations on the remaining parameters.

B. A Hypercube with ±3-sigma variation on the initial rotation rate, and +3-sigma variation on the remaining inputs (a total of $2^3 = 8$ points)

The attitude settle time of **Set A** is found to be 6.166 minutes, exceeding the 5-minute threshold. This value is found to be in the upper range of the results from the 1000 tested, with only 3 points exceeded this value. Despite the increased initial speeds of the three RWs, the maximum speeds are all found to be below the 3000 RPM threshold (the maximum was found to be 1628 RPM on RW2).

The results from **Set B** show that none of the points exceeded the mission requirements on attitude settle time or RW speeds. This suggests that input regions that lead to exceeding the mission requirement are not determined by maximum values on each of the input parameters but are instead determined by *specific combinations of values* that may be non-deterministic. This confirms our expectation that *the Hypercube sampling method may not be sufficient* to ensure that none of the inputs lead to outputs that exceed the mission requirements.

Scenario 2 Predict Specific Regions of Output Space

Besides predicting the entire output space, we use the best performing Meta-models to predict specific regions of interest of the output space. Our co-authors suggested predicting specific ranges of *attitude settle time and the reaction wheel (RW) speeds at settling*. Figure 7 shows histograms of the actual output values for attitude settle time, and the settle speed of the RW 1 , for all 10,000 points. The attitude settle time shows a right-skewed distribution with a Mean of 3.4479 min, and standard deviation 0.6836 min. The settle speed for the first reaction wheel shows a Gaussian distribution with mean 80.32 RPM and standard deviation 454.204 RPM.

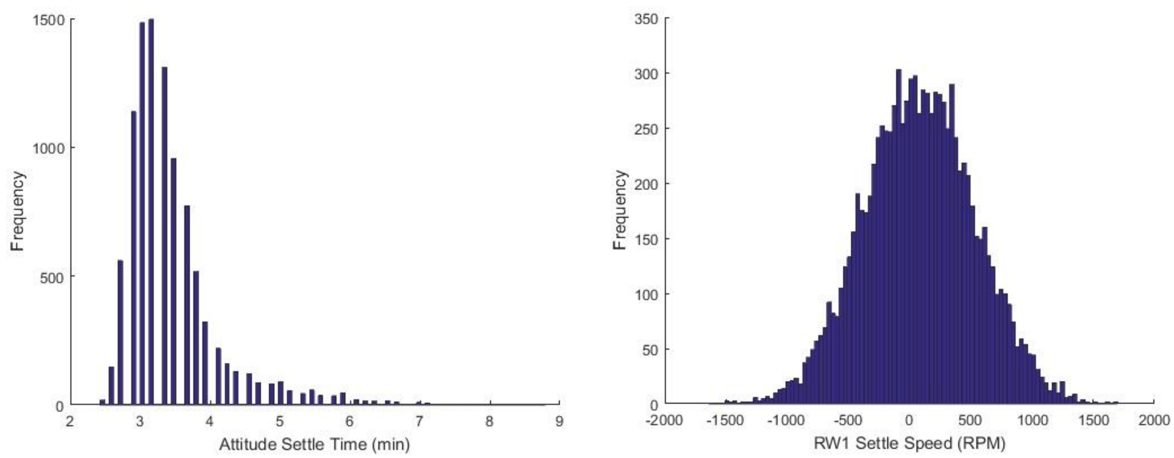

Fig. 7 Histogram of the actual attitude settle time and settle speed of RW1 for 10,000 points

Predicting Ranges of Attitude Settling Time

As described above, Meta-models are generated from a set of 1000 points, with actual outputs from the simulation. To test the accuracy of the predicted attitude settle time, three sets of test points are generated across different regions of the attitude settle time from the 10,000 point set.

1. Low: 500 points randomly sampled from the 6152 points with actual attitude settle time attST < 3.4479
2. Medium: 500 points randomly sampled from the 3368 points with actual attitude settle time: 3.4479 < attST < 5
3. High: All 480 points with actual attST > 5 min

The various Meta-models are used to predict the attitude settle times of these ranges of output. For each point, the residual is calculated as the difference between the predicted and actual value generated from the simulation. The predictive accuracy of each meta-model in each of the three regions is compared using the RMSE of the residuals. The results for the best- performing model are summarized in Table 2 below. In the majority of cases the Random Forest method performed better than Radial Basis Functions and Kriging (as measured by RMSE of the residuals).

Attitude Settle Time Regions	Meta-Model	Residuals		
		Mean	STD	RMSE(Mins)
Low (2.5-3.5 mins)	Random Forest B	-0.1081	0.1460	0.18
Medium(3.5-5mins)	Random Forest A	0.0862	0.3113	0.32
High (>5mins)	Random Forest A	0.9127	0.4969	1.04

Table 2 Summary of Meta Model Performance: Predicting Ranges of Attitude Settling Time

For Low attitude settle times the model slightly under-predicts the attitude settle times. The accuracy in this region is to within one standard deviation of 0.146 minutes (an accuracy of 2.92 % with respect to the 5 minute threshold). For Medium attitude settle times, the model slightly over-predicts the attitude settle times. The accuracy is to within one standard deviation of 0.3155 mins (an accuracy of 6.31%). For High attitude settle times, the model over predicts the attitude settle time. The standard deviation is still fairly small (an accuracy of 11.4 %), however the mean is far away from zero (0.9127 minutes). Figure 8 shows a histogram of residuals for the low and medium attitude settle time regions.

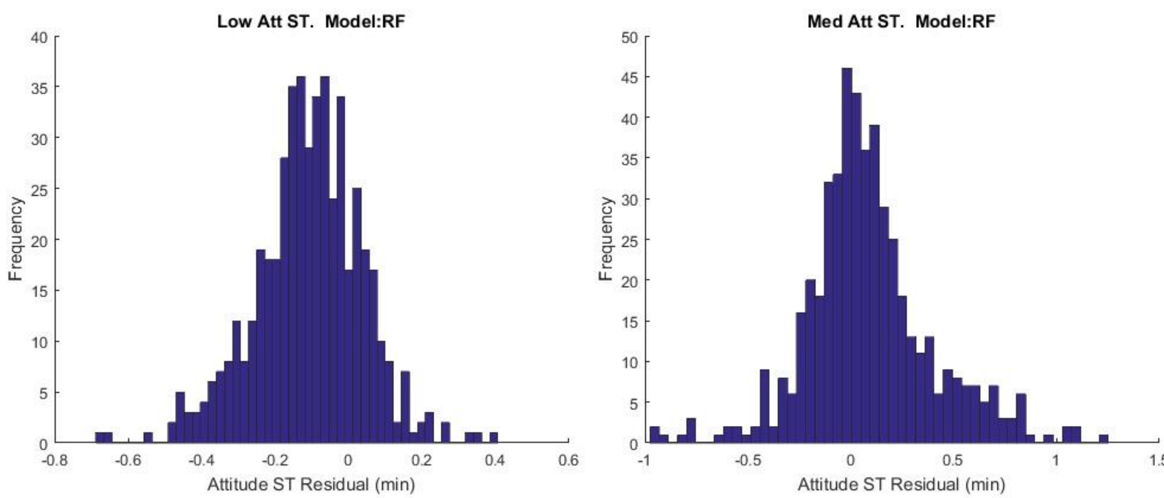

Fig. 8 Histogram of Residuals for the Low and Medium Attitude Settle Time regions.

Predicting Ranges of Reaction Wheel Speeds At Settling

The *Reaction Wheel settled speed* is important to understand the dynamics and momentum buildup after the maneuver has ended. If a maneuver causes the RWs to exceed the limit for a small duration, that is typically satisfactory. However , if the wheels settle at a large wheel speed, causing a large new RW-cluster momentum vector, it makes the next maneuver much harder to manage.

As above, Meta-models were generated from a set of 1000 points, with simulated outputs. The accuracy of these models to predict *Reaction Wheel Speeds At Settling* was tested on a set of points drawn from a set of 10,000 points run through the simulator. To test the accuracy of the predicted *Reaction Wheel Speeds At Settling*, two sets of test points were generated across different regions of the attitude settle time.

i. Negative: RW1 Settle Speed <= 0 RPM
ii. Positive : RW1 Settle Speed >= 0 RPM

The results for the best- performing model are summarized in Table 3 below. In the majority of cases the Kriging model performed better than Radial Basis Functions and Random Forests (as measured by RMSE of the residuals). Figure 9 shows a histogram of residuals for the two RW1 settle speed regions.

RW Settling Speeds	**Meta-Model**	**Residuals for RW1**		
		Mean	**STD**	**RMSE(RPM)**
RW1 < 0	Kriging	-14.2	94.9	95.99
RW1 > 0	Kriging	3.1	57.77	57.85

Table 3 Summary of Meta Model Performance: Predicting Ranges of RW Settling Speeds

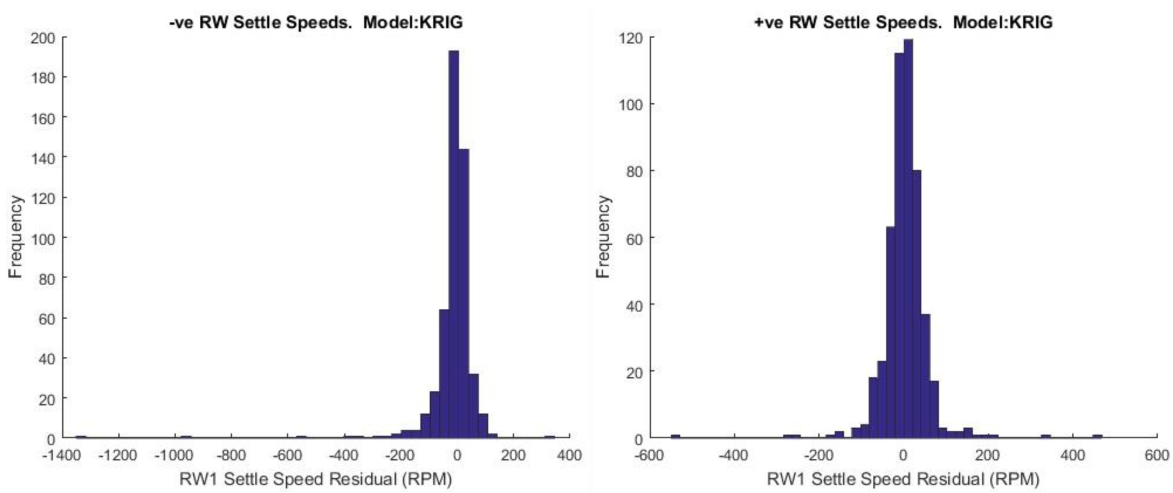

Fig.9 Histogram of Residuals for Reaction Wheel Speeds at Settling.

The analytical form of the Kriging model is:

$$RW1\ Settle\ Speed = 7878.716 + 18.829283x_1 - 67.13862x_2 + 25.7522x_3 + 1492.338x_4 + 597.177x_5 - 328.0331x_6 - 12.28614x_7 - 201.581x_8 + 31.88362x_9 + 24.13904x_{10} - 2.399211x_1^2 + 5.830943x_2^2 - 2.30998x_3^2 - 29.17384x_4^2 + 123.6164x_5^2 - 6.55749x_6^2 + 0.008051773x_7^2 + 1.025763x_8^2 - 0.07848589x_9^2 - 0.04069705x_{10}^2$$

where x_1 to x_{10} are the 10 inputs of the model (Inertial attitude x-y-z, Inertial rotation rate x-y-z, spacecraft mass, and initial speeds of the three RWs).

Other Analysis Performed

The results presented in this paper are limited to analyses deemed most relevant by the users of the attitude control simulation (our co-authors at University of Colorado). There were several other analyses performed, worth briefly noting. One of these used the meta-model to predict the region *surrounding outliers* to try to gauge the size of the causes of outlier region in input space. Two outliers of Reaction Wheel (RW) speeds > 3000 RPM were identified. A 10-dimensional grid of test points was generated in the input space surrounding these points. The results show that the predicted RW speed of the outlier did not exceed the 3000 RPM limit, and as a result, neither did the surrounding points. These outliers were found in the 10,000 sample points. It is expected that the 1000 points sampled did not have enough resolution to find this "hill" region of the output , and so the meta-model smoothed over the region.

For further research into outlier region prediction we suggest adding additional sample points in regions of the input space that are expected to produce outliers, to give a better resolution in these regions.

CASE STUDY 2: MISSION ANALYSIS FOR IBEX

The Interstellar Boundary Explorer (IBEX) mission is a NASA mission that was launched in 2008 with a goal of mapping the boundary of the solar system (Reference 8). At the conclusion of the primary mission, IBEX performed a series of maneuvers to change its orbit for an extended mission. Prior to launch, the flight dynamics team for the IBEX mission performed a covariance analysis on a number of mission segments including the pre-launch trajectory design, post-launch orbit verification, post-launch ascension planning, and long-term orbit evolution. The purpose of the analysis was to ensure that given uncertainties in the initial state vector and uncertainties in the magnitudes of each maneuvers, the resulting trajectories would meet a set of mission requirements. The main mission objectives were that the orbit's radius of perigee does not drop below 2.3 Earth radii, and that the eclipse time in Earth's shadow does not exceed 4 hours.

For each of these mission scenarios, a set of perturbed initial states were generated using two sampling methods: First, a traditional Monte Carlo simulation with 1000 random draws from the 3-sigma covariance ellipse, and secondly, the 3-sigma vertices of a Hypercube with 128 points. Each sample point was propagated using AGI's Systems Tool Kit (STK), producing time-series data of the perigee radius, eclipse/shadow time, and other trajectory parameters. This data was used to confirm the mission design would meet the trajectory constraints for a specified variance on the input parameters. The details of this analysis were presented in Reference 9.

Mission Scenario

In this case study, we repeat the same analysis as that presented in Reference 9, focusing on the extended mission of IBEX (Reference 10). Following its two-year primary mission, the spacecraft performed a series of orbit maintenance maneuvers (OMM) to place it into a highly elliptical orbit of period 9.1 days, resonant with the Moon's orbit.

The inputs to the simulation are the initial state of the spacecraft prior to the final OMM (three-dimensional position and velocity), and the magnitude of the delta-V of the final OMM. Perturbations are added to each of these 7 inputs to account for the uncertainty of the initial state, and the performance of the maneuver execution. The uncertainty in the initial state is described by a 6x6 covariance matrix, generated from orbit determination. The coordinates of the perturbed states are expressed in the frame of the principal axis of this covariance matrix, found through a Cholesky decomposition.

The uncertainty in the delta-V of the maneuver is normally distributed with mean a mean around 77 m/s with an uncertainty of about 3%. (The uncertainty is due to the uncertainty in the predicted tank pressure at the time of maneuver ignition.)

Application of Sampling Methods

The previous analyses (Reference 9) generated two sets of perturbed initial states using 1) Random sampling over 1000 points and 2) Hypercube (HC) sampling on a 7-dimensional Hypercube with vertices at ±3-

sigma in each input. Similar to Case Study #1 above, in this Case Study we have tested a range of space- filling designs:

- Latin Hypercube Sampling (LHS)
- Maximin
- MaxPro

Each of the sample points is propagated forward in STK for 100 orbits (approximately 2 years) producing time-series data of the perigee, eclipse times and other orbit parameters. The objective is to confirm that for any perturbation of the initial state, the trajectory will meet constraints on the minimum periapsis radius and maximum eclipse time.

Results and Analysis

Simulation outputs for three-sigma HyperCube (HC), Latin Hypercube Sampling (LHS), Maximin Sampling and MaxPro Sampling were assessed for the occurrence of outputs that strayed beyond the bounds of the mission (Reference 10). *The goal of this analysis is to identify cases (if any) where the three-sigma hypercube input had not mapped to the most extreme output cases.* Results for each output are grouped by orbit – as such, our analysis compares the extrema for each input case at every orbit separately.

The orbital properties examined were;

- Perigee Altitude
- Time of Perigee
- Maximum shadow duration
- Angle between Apogee and Helionose (the tip of the bow shock where the solar wind collides with the interstellar wind)

Perigee Altitude

For the extended IBEX mission (Reference 10) the principal perigee requirement was that the post-maneuver perigee remained above 2.3 earth radii for the duration of the mission. Comparison of the minimum simulated perigee at each orbit shows that this is maintained in all cases, and that there appears to be little difference in the output extrema for 3-sigma hypercube and space-filling input schemes – the 3-sigma approach adequately captured the output extremes for perigee. Overlaid minimum perigees are displayed below in Figure 10.

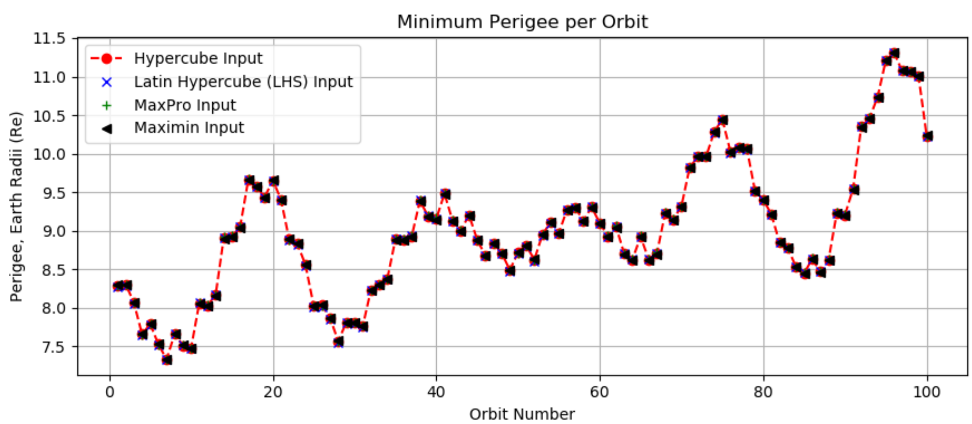

Figure 10. Minimum Perigee for differing input sampling regimes. No deviation between input schemes was observed.

Time of Perigee

In examining if hypercube sampling missed errant perigee times, the total range of possible occurrence times for each perigee were compared for the different input schemes (Figure 11). It was found that each input scheme generates perigee times with a periodic uncertainty, with sub 1-hour ranges at minima to ranges of up to 7 hours at maxima. At some of these uncertainty maxima, the comprehensive input schemes (LHS, MaxPro, Maximin) predict a larger range of possible times – up to 30 minutes more than the 3-sigma HC. Although relatively minor compared to the total range of uncertainty, this difference suggests that **wayward values were missed** by the hypercube sampling – the extreme input case (HC) did not effectively identify the most extreme possible perigee times.

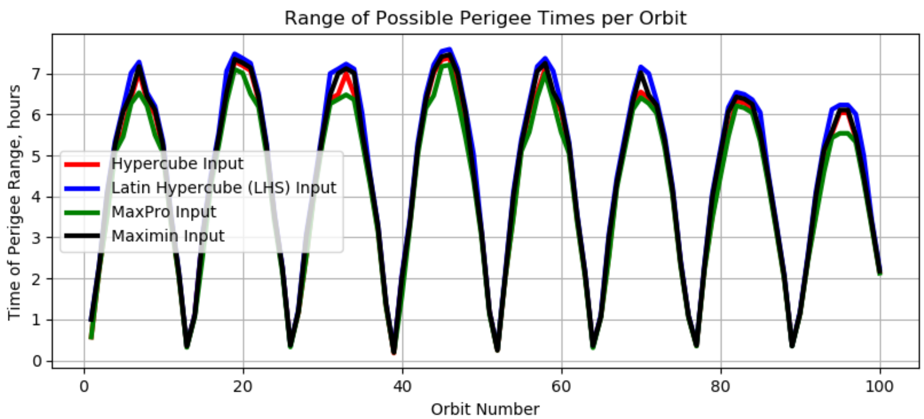

Figure 11. Range of Perigee Times, per perigee, across different sampling schemes. Minor differences are seen at the extremes.

Maximum Shadow/Eclipse Time

For both the prime and extended IBEX missions, constraints were placed on the largest continuous time the spacecraft could be in shadow. The concern with hypercube analysis was that the original 3-sigma input may underestimate shadow duration or miss shadows completely. To test this, the maximum possible shadow duration is compared per-orbit for our four input schemes. Firstly, instances where the same shadow was predicted across multiple orbits were identified – these outputs would skew the data by indicating 'missing' shadows in some cases, where they have simply shifted to an adjacent orbit for that particular input value. The spread of shadows over orbits can be seen in Figure 12 – for all input cases, the spread increases through time, as the system becomes more chaotic. Importantly, some comprehensive input schemes (LHS and MaxPro) show an increase in shadow-spread over the 3-Sigma HC case.

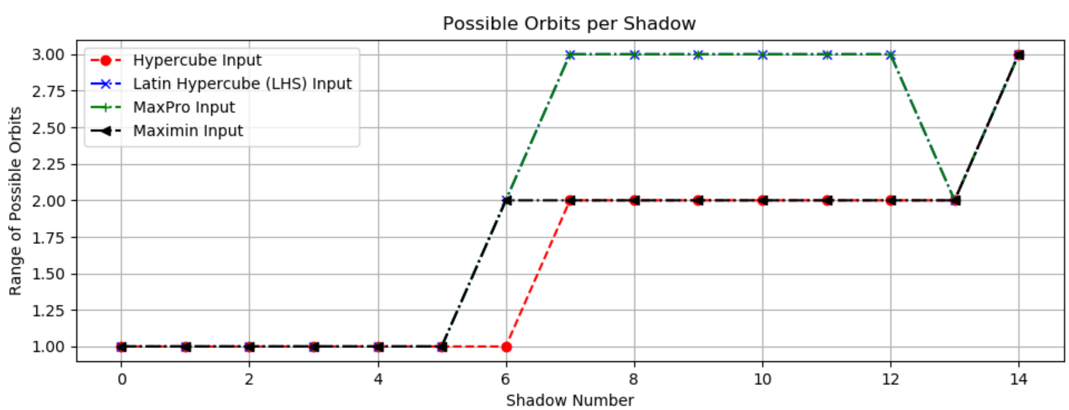

Figure 12. Number of possible orbits for each shadow instance. Different input schemes predicted specific shadows occurring over two or three orbits – this is the primary variation between the sampling schemes.

Secondly, the maximum duration of shadow was calculated at each shadow instance (normalized across orbits), for each input scheme. The maximum allowable shadow duration of the original IBEX mission (Reference 9) was four hours – this was used as a suitability benchmark for each predicted shadow. Figure 13 shows the maximum duration of each shadow per orbit, formatted to display only the 'central orbit' of each shadow (if the predictions spread across multiple orbits). Strong agreement is found between all input schemes in early timesteps, but a diverging behavior as the number of orbits increases. Overall, the comprehensive input scheme presents similar, if slightly higher, estimates for the maximum shadow duration at each orbit, with these differences increasing over time. Certain input schemes (Maximin) predict shadow-spreads that are 'centered' around different orbits – this is merely an artefact of the analysis and does not indicate a staggered output. Nevertheless, we can conclude that the HC input scheme was not completely sufficient in mapping the output space– without a space filling simulation regime, possible shadow-containing orbits would remain unidentified, as would the full range of possible shadow durations.

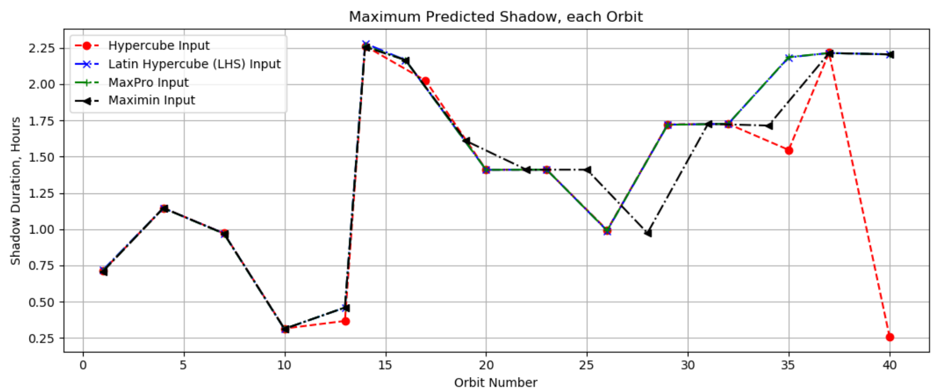

Figure 13. Maximum shadow time for all input schemes across a range of 45 orbits. Shadows with a predicted spread across multiple orbits are normalized to a single orbit for magnitude comparison.

Helionose Angle

For the extended IBEX mission (Reference 10), it was preferable to orient the orbit apogee as close as possible to the direction of the helionose, a supposed bow shock between solar wind and the surrounding interstellar medium. As with perigee, the concern was that three-sigma hyper cube analysis did not locate every extreme output – within the input space, some non-extreme combinations would yield a substantial deviation once propagated forward. However, as Figure 14 shows, HC, LHS and other input methods all yield identical results for max helionose angle; the 3-sigma input scheme adequately surveyed the full range of helionose output. Similar results were found for the minimum angle of helionose at each orbit (not shown).

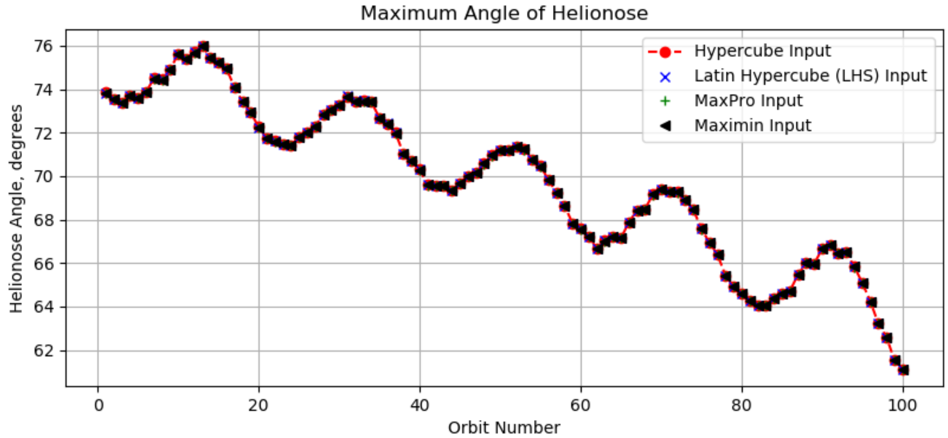

Figure 14. Maximum angle of Helionose for HyperCube and space filling input schemes. No deviation between input schemes was observed.

Analysis Conclusions - Case Study 2

The use of non-random space filling methods to generate simulation inputs allows a more comprehensive assessment of the state of the output space. For the IBEX extended mission, where identifying large scale deviations from normal output conditions is the goal of the simulation scheme, it was found that a three-sigma hypercube - where simulation inputs are simply combinations of the highest-expected-errors – provided adequate coverage of much of the output space, particularly angular and distance measurements. However, timing of orbit events and the existence of particular phenomena (eclipses/shadows) during each orbit showed non-negligible improvements in output range when a more comprehensive input scheme is used, suggesting that within the input space *there are cases which non-extreme combinations of inputs generate extreme outputs for chaotic/resonant orbits*. As such, a space filling input regime is a welcome improvement to any simulation experiment set, both to verify that the extreme-input cases do identify the outputs of concern, and to identify missed occurrences / underestimated magnitudes from an extremes-only system.

CONCLUSIONS

The adaption of DACE methods to astrodynamics simulations in these proof of concept experiments has yielded promising results that warrant further attention. The use of space-filling designs to identify data extremes provide economic advantage over Monte-Carlo sampling schemes and can identify anomalous outputs missed by a conventional extremes-only hypercube. Time-based outputs (Attitude Settle Time, Case Study 1 and Shadow Time, Case Study 2) are found to be particularly susceptible to extreme values from seemingly innocuous inputs. In addition, these space filling designs can validate that a prior hypercube or Monte-Carlo analysis has successfully explored the output space for less chaotic parameters.

Meta-modelling to reduce simulation complexity provides a low-cost for data exploration, sensitivity design and analysis. However, as shown in Case Study 1, the upfront costs of developing a meta model are steep – one must first identify the most significant inputs, decide on a sampling scheme and iterate through several meta models until one with suitable accuracy is developed. Nevertheless, meta-modelling can perform an important function in any simulation regime, particularly if the per-run costs of the simulation are high. Meta-models with a sufficiently low RMSE can be used to pre-estimate results for previously untested data points, reducing the need for additional simulation runs if the predicted outputs are below/above design thresholds.

One potential use for the Meta-model is in *screening regions* of the input space to give a quick indication of whether or not the inputs will produce outliers. If we know that the Meta-Model is accurate within a certain percentage (%), then any output point that we predict to be below the threshold less the error (i.e. Predicted output < Threshold – Error), we can be confident that the output for that input point will be within the bounds.

If the predicted output is close to the outlier (within the accuracy of the model), the input point may or may not be causing an outlier (it is in a grey area). If the predicted output is above the threshold plus the error of the model, we can be confident that this input point will lead to an exceedance of the bounds.

REFERENCES

1. T. Fang; R. Li; A. Sudjianto, *Design and Modelling for Computer Experiments* Chapman & Hall, 2006
2. D. Shteinman; Z. Lazarov, Y. Ting, Y (2017) "Use of Advanced Statistical Techniques for Mission Analysis: Case Study from A Google Lunar X Team (SpaceIL) *" AAS/AIAA Astrodynamics Specialist Conference*, August 2017, AAS/AIAA 17-664
3. A. Giunta; S. Wojtkiewicz; M. Eldred (2003)*"Overview of Modern Design of Experiments Methods for Computational Simulations" AIAA 2003-0649, 41st AIAA Aerospace Sciences Meeting*
4. J. Glanovsky (2013) "An Introduction To Space-Filling Designs*" ASQ Statistics Division Newsletter, V*ol.32, No. 3, pp.12-18
5. V. Joseph (2016**) "**Space-Filling Designs for Computer Experiments : A Review**"** *Quality Engineering; Vol.28, No. 1, pp. 28-35*
6. M.D. Mckay; R.J. Bekman, W.J. Conover (2000) "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output From a Computer Code", Technometrics, Vol.42:1, 55-61
7. B.Jones, C. Nachtsheim (2011) *"*A Class of Three-Level Designs for Definitive Screening in the Presence of Second-Order Effects*", Journal of Quality Technology, Vol.43, No. 1, January 2011, pp. 1-15*
8. P. Chandrashekarappa; R. Duvigneau (2007) " Radial Basis Functions and Kriging Metamodels for Aerodynamic Optimization", *INRIA Research Report 00137602v1*
9. McComas, et al., (2005) "The Interstellar Boundary Explorer (IBEX) mission*"*
10. J. Carrico, L. Policastri, R. Lebois & M. Loucks (2010) "Covariance Analysis and Operational Results for the Interstellar Boundary Explorer (IBEX)", *AIAA Astrodynamics Specialist Conference*
11. J. Carrico Jr., D. Dichmann, L.Policastri, et.al. "Lunar-Resonant Trajectory Design For The Interstellar Boundary Explorer (IBEX) Extended Mission," Paper No. AAS 11-454, *AAS/AIAA Astrodynamics Specialist Conference*, Girdwood, Alaska, USA, August 2011