Annals of Nuclear Energy xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Annals of Nuclear Energy

journal homepage: www.elsevier.com/locate/anucene

Linear regression and sensitivity analysis in nuclear reactor design

Akansha Kumar *, Pavel V. Tsvetkov, Ryan G. McClarren

Department of Nuclear Engineering, Texas A&M University 3133, TAMU, College Station, TX 77843, United States

ARTICLE INFO

Article history:
Received 25 March 2015
Received in revised form 26 June 2015
Accepted 27 June 2015
Available online xxxx

Keywords:
Predictive
Optimization
Reactor
Regression
Sensitivity
Fast

ABSTRACT

The paper presents a general strategy applicable for sensitivity analysis (SA), and uncertainity quantification analysis (UA) of parameters related to a nuclear reactor design. This work also validates the use of linear regression (LR) for predictive analysis in a nuclear reactor design. The analysis helps to determine the parameters on which a LR model can be fit for predictive analysis. For those parameters, a regression surface is created based on trial data and predictions are made using this surface. A general strategy of SA to determine and identify the influential parameters those affect the operation of the reactor is mentioned. Identification of design parameters and validation of linearity assumption for the application of LR of reactor design based on a set of tests is performed. The testing methods used to determine the behavior of the parameters can be used as a general strategy for UA, and SA of nuclear reactor models, and thermal hydraulics calculations. A design of a gas cooled fast breeder reactor (GCFBR), with thermal–hydraulics, and energy transfer has been used for the demonstration of this method. MCNP6 is used to simulate the GCFBR design, and perform the necessary criticality calculations. Java is used to build and run input samples, and to extract data from the output files of MCNP6, and R is used to perform regression analysis and other multivariate variance, and analysis of the collinearity of data.

1. Introduction

In this section the background, motivation, and literature review of this work has been presented. This section also presents a brief introduction to the current work.

1.1. Background and motivation

Nuclear reactors are complex systems having numerous parameters those affect the system independently, as well as in an interactive fashion. In addition, the system has an inherent uncertainty in the model structure. The interaction between the parameters are usually of a higher order, hence modeling those uncertainties, is a difficult problem to solve. These parameters also have a non-linear effect on the system behavior. Software codes have been developed based on stochastic as well as deterministic methods to model these systems, however with an increase in complexity due to a need for heterogeneity, and accuracy, the codes are expensive to use. Therefore, it is important to understand the behavior of these uncertainties in the model to perform an acceptable predictive analysis. Uncertainty quantification analysis (UA), and sensitivity analysis (SA) provide a very accurate understanding of the uncertainties in the system, which is imperative for predictive analysis. UA may be used to assess the variability i.e. imprecision in the predictions in the output parameter that is due to the uncertainty in estimating the values of the input parameters. This is due to the propagation of uncertainties in the system. However, SA deals with identifying the input parameters those are important in contributing to the imprecisions in the prediction of the output parameter. In simple words, SA is a means to quantify how sensitive a parameter is to the behavior of the system. UA and SA of complex models are usually conducted by executing the model many times with a set of random samples while varying the parameter inputs (Saltelli et al., 2000). Varying all the input parameters simultaneously allows analysis of the model response to individual input parameters and their interaction. In addition, this analysis helps in understanding the robustness of the system, and behavior of parameter relationship such as, non-linearity, independence, and association.

Linear regression (LR) analysis (LRA) is the art of fitting straight lines to a pattern of data. In a LR model, the output parameter is predicted from the set of input parameters using a linear equation. The output parameter can be inherently and non-linear function of the input parameters, however, a transformation to this non-linear relationship to achieve linearity is called LR. Assume that $\hat{Y} = \alpha X^2$, is a non-linear relationship. If $Q = X^2, \hat{Y}$ can be transformed and

* Corresponding author.
E-mail addresses: akansha.tamu@gmail.com (A. Kumar), tsvetkov@tamu.edu (P.V. Tsvetkov), rgm@tamu.edu (R.G. McClarren).

http://dx.doi.org/10.1016/j.anucene.2015.06.037
0306-4549/© 2015 Elsevier Ltd. All rights reserved.

written as $\hat{Y} = \alpha Q$, which is actually a linear representation of the non-linear equation $\hat{Y} = \alpha X^2$. This transformation has underlying assumptions, and a detailed analysis of the validity of LR in the system needs to be addressed. Nuclear reactor designs are non-linear systems, but with extensive transformations, a linear model can be fit to certain relationships between the parameters. In LR, a regression surface or a linear fit is built based on a set of trial data. The trial data is obtained from experiments or from simulations. The regression surface is used to predict the output parameters for a set of test input parameters (test data). In addition, a neutron transport and depletion code like MCNP (X-5 Monte Carlo Team, 2008), SERPENT take a significantly long execution time if higher accuracy of results is desired. And, in a global optimization technique like GA (Kumar and Tsvetkov (2015)), simulated annealing, particle swarm optimization the code needs to be run for a significant number of input samples before a desired solution is obtained. Hence, regression methods could be useful to emulate the physics solver in the code, and preform a predictive analysis for test data. However, this is possible only if there is a detailed analysis of the accuracy of the predictions given by the regression method. Therefore, the validity of LR in terms of linearity, normality, association, independence, and other methods needs to be done.

1.2. Current work

SA is performed on a specific set of explanatory parameters, and the behavior of the predictors is analyzed. Trial input set is built using samples obtained from the Latin Hypercube Sampling (LHS) method. Analysis has been done to determine the sensitivity of input parameters on the output parameters. It is to be noted that the output parameters defined in this work are the quantities of interest (QOI). An LR fit is used to perform predictive analysis. The regression fit is obtained from a set of trial data. Trial output parameters are obtained when MCNP6 is run on a predetermined sample of input set obtained from LHS method. The combination of the trial input set, and trial output parameters form the trial data. The trial data is used to build the regression fit for the parameters. The regression fit is applied on the test data to perform predictive analysis in a satisfactory confidence level. The accuracy of the predictions depend on the size of samples used as trial data. In addition, analysis has been done to determine the measure of importance, measure of association, test of normality, test of linearity, distribution statistics, and homoscedasticity of the parameters on the regression surface to determine the validity of LR.

The model of a gas cooled fast breeder reactor (GCFBR) design (Kumar et al., 2014) is used to perform the SA, and LR analysis. The system consists of a, heterogeneous neutronics model for criticality, and flux calculations, basic thermal hydraulics calculation for transfer of heat from the fuel pin to the coolant, and energy transfer using the Brayton's cycle. Following physics based modules have been determined to perform the analysis. The first module, is the fuel pin cell module that determines the infinite neutron multiplication factor, K_∞ (KINF), fraction of fission due to intermediate energy neutrons, $T_{i,f}$ (INTERFF), and fraction of fission due to fast energy neutrons, $T_{f,f}$ (FASTFF) based on a specific radius of the fuel element, r_F (RADIUS), and enrichment of U-233 in $(U-Th)O_2$ fuel (ENRICH). Therefore, the input parameters defined in this module are RADIUS, and ENRICH, and the output parameters are KINF, INTERFF, and FASTFF. The second module, is the design of the whole core of the GCFBR reactor. The output parameters in this module are the radial power peaking factor, $F_{PF,rad}$ (RADPF), axial power peaking factor, $F_{PF,ax}$ (AXPF), and effective neutron multiplication factor, K_{eff} (KEFF). This module has the same input parameters as defined in the first module. The third module, is a basic thermal–hydraulics and heat transfer module, where a hot channel analysis is performed to analyze the heat transfer across the fuel pin cell i.e. the flow of heat from fuel pin to the coolant wherein, the input parameters are the core inlet temperature, T_{in} (TIN), and, flow rate of the coolant, W (W) to determine the core outlet temperature, T_{out} (TOUT), and pressure drop, ΔP (DELTAP) across the flow channel. In this module the input parameters are, TIN, W, and ENRICH, and the output parameters are TOUT, and DELTAP. The fourth module is the energy transfer module that performs a basic Brayton's cycle calculation to determine the thermal efficiency (EFF). The input parameters defined in this module are TOUT, TIN, and the out parameter is EFF. A detailed description of the parameters of the mentioned modules is presented in a future section.

It is very important to understand the validity of LR in a non-linear system like the reactor power system. Particularly, when it is a steady state calculation with no depletion, based on

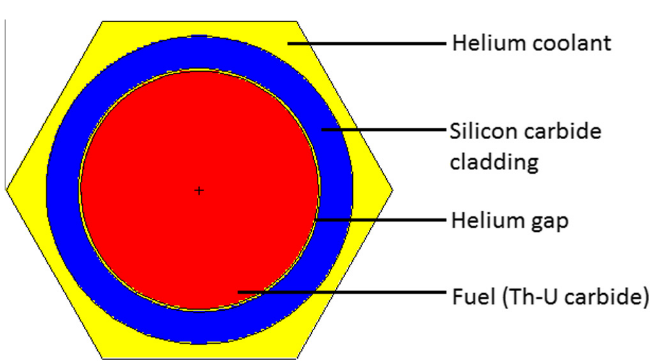


Fig. 1. Single fuel pin cell with reflective boundary conditions.

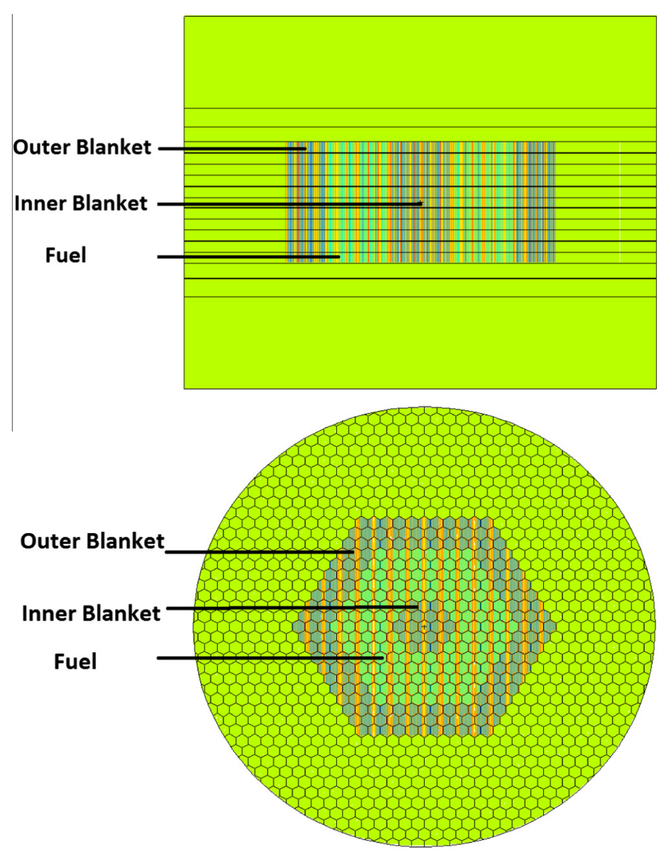


Fig. 2. Axial and radial view of the whole core.

Please cite this article in press as: Kumar, A., et al. Linear regression and sensitivity analysis in nuclear reactor design. Ann. Nucl. Energy (2015), http://dx.doi.org/10.1016/j.anucene.2015.06.037

data a linear fit is applicable and could give a valid prediction. Another thing is that linear models are applicable where the perturbations are small because any system can be represented linearly for small changes. Also, a linear model can have terms that are parametrically nonlinear i.e. x^2 or $x1 * x2$ terms. These types of terms could be added to the model where appropriate. And, in traditional safety and reactor parameter analysis, a linear model is usually assumed. Within small perturbation limits, the reactor behaves linearly and a fundamental mode for criticality can be assumed. Therefore, even though nuclear reactors are non-linear systems, within the standard traditional evaluation approach (reactivity coefficients, etc.) only linear component of their behavior is considered. This is the case in this work, second and higher non-linear terms assumed to be negligible within the range of perturbations around nominal operation mode.

1.3. Past work and literature review

LR has been applied in a wide range of engineering problems for sensitivity analysis, and optimization. Few non-nuclear applications include water resource (He et al., 2011), disease transmission analysis (Blower and Dowlatabadi, 1994), environmental models design and analysis (Hamby, 1994), hydrology (Mccuen, 1973), and water quality (Beck, 1987). However, the applications of LR in nuclear reactor design, particularly optimization has not been significantly explored. In nuclear engineering domain, sensitivity analysis have had significant progress in fuel cycle, waste management (Saltelli et al., 1993; Helton, 1993), risk analysis, and probabilistic safety assessment (Saltelli and Homma, 1992). However, SA (Homma and Saltelli, 1996), and linearization studies of reactor design has not had significant progress in the research community. This work focuses on the validity of linearization of a nuclear reactor design, by perform a detailed sensitivity analysis and linearity tests. In this work the, safety aspects of a nuclear reactor, material and structural details, control elements including the control rods, fuel cycle,optimization methods, and economics of the nuclear reactor core design has been ignored. The method and theoretical background of how to build the transformation to model the non-linear relationships in a reactor design with a linear model is not the focus of the paper. Rather, this paper validates a LR given by R, and shows its applicability in reactor design.

The remainder of this paper is organized as follows. First, the detailed design of the neutronics and thermal hydraulics system is presented. This is just a vehicle to demonstrate the application of SA, and LR in reactor design. Next, the sampling method is presented, and a detailed description and analysis of the tests for linearity is shown. Then, the results from R has been presented with a detailed discussion. Finally, the paper ends with a summary and a discussion on the prospects for future work.

2. Reactor design and thermal–hydraulics analysis

In this section, the motivation, assumptions and implementation of each of the modules introduced in Section 1 is presented.

2.1. Fuel pin cell

A single fuel pin cell analysis with reflective conditions on all external boundaries is done to determine the behavior of the parameters: $k_\infty, T_{f,f}$. In other words, $T_{f,f}$ is the fraction of the fission events caused by the neutrons above the fast neutron energy of 100 keV. To obtain a controlled nuclear fission chain reaction with breeding capability the following objectives need to be met: higher value for $T_{f,f}$, and $k_\infty > 1$. The parameters analyzed in the fuel pin cell module are, the radius of the fuel pin in the fuel element (r_f), and the enrichment of U-233 in Th-U fuel (Kumar et al., 2014). It is to be noted that, $T_{f,f}$ depends on the $\frac{p}{D}$ ratio. In this case a constant value for the pitch has been used, hence, the diameter varies due to varying $\frac{p}{D}$ ratio. Following figure (Fig. 1) shows a graphical view of the single fuel pin cell used for the analysis. The density of the materials is assumed to be constant, even though in reality, density depends on the temperature. In addition, a constant value for the thickness of the gap (t_G), and the thickness of the cladding (t_C) has been assumed.

2.2. Whole core

For the whole core analysis, the configuration of the core of an existing gas cooled fast breeder reactor design (Kumar et al., 2014) has been used. However, for simplicity, the control rods, and the axial blankets have been ignored in the design. A detailed description of other components are presented in the paper (Kumar et al., 2014). The core consists of an array of hexagonal assemblies wherein, an assembly consists of an array of fuel elements in a hexagonal lattice. The assemblies with the fuel elements having fissile material are called as the "fuel" assemblies, and those with fertile material are called as the "blanket" assemblies. The fuel assemblies have fuel elements having a mixture of Th-232 and U-233. The blanket assemblies have light water reactor used fuel. The core consists of internal and external blanket assemblies. The parameters analyzed in this module are $k_{\text{eff}}, F_{\text{PF,rad}}$, and $F_{\text{PF,ax}}$. The total mass of the fuel (fissile + fertile) material is kept constant. Therefore, when the radius of the fuel element is varied, the height of the core (L_{FC}) is affected, and due to having a constant density, the total mass of the fuel remains constant. For a safe operation in terms of preventing meltdown of the fuel rod, peaking factors play an important role. The peaking factor is the ratio between maximum local energy depositions to the average energy deposition in the reactor core. It has been assumed that the external blanket is not part of the reactor core while calculating the radial peaking factors. In the whole core analysis a constant value for the, power, and positioning of blankets has been assumed. Following figure (Fig. 2) presents a radial and an axial view of the whole core.

2.3. Thermal hydraulics and heat transfer

A standard thermodynamic analysis of the transfer of heat across the fuel gap, cladding and bulk coolant has been performed to the core design. However, for completeness a simplified model is implemented, and a rigorous thermal hydraulics and energy transfer analysis has not been performed. In a single phase coolant heat transfer domain, the pressure drop across the length of the active core is the sum of the pressure drop due to friction, form and elevation, given by,

$$\Delta P = \Delta P_{\text{friction}} + \Delta P_{\text{form}} + \Delta P_{\text{elevation}}. \tag{1}$$

For simplicity, the total pressure drop is assumed to be from friction only hence, $\Delta P_{\text{form}} = 0$, and $\Delta P_{\text{elevation}} = 0$. Therefore, the primary loop pressure drop (ΔP) is given by a simplified equation,

$$\Delta P = \left(\frac{\rho_{\text{FC}} \cdot V_{\text{FC}}^2}{2}\right) \cdot \left[f_{\text{Darcy-Weisbac}} \cdot \frac{L_{\text{FC}}}{D_{\text{FC}}}\right], \tag{2}$$

where

$$V_{\text{FC}} = \frac{W}{N_{\text{FC}} \cdot \rho FC \cdot A_{\text{FC}}}, \tag{3}$$

where ρ_{FC} is the density of the coolant, V_{FC} is the velocity of the coolant, $f_{\text{Darcy-Weisbac}}$ is the Darcy–Weisbach constant with a value of 0.016, D_{FC} is the diameter of the fuel element, N_{FC} is the number of fuel elements, and A_{FC} is the flow area of the coolant. The temperature of the coolant at core outlet (T_{out}) is given by,

$$T_{out} = T_{in} + \frac{Q}{W \cdot C_p}, \tag{4}$$

where Q is the total thermal power of the reactor core, and C_p is the specific heat capacity of the coolant. The pumping power of coolant (P_{pump}) is given by,

$$P_{pump} = \frac{\Delta P \cdot A_{FC} \cdot V_{FC}}{\eta_{pump}}, \tag{5}$$

where η_{pump} is the pump efficiency. For a safe operation, and to ensure fuel material structural integrity, it is important to compute the maximum radial, and axial fuel temperature. There is a temperature variation radially on the fuel element due to the presence of heterogeneous components: fuel, gap, and clad. The governing equations with assumptions for simplicity are given by,

$$\Delta T_b = \frac{q'_{peak}}{2 \cdot \pi \cdot L_{FC} \cdot (r_F + t_G + t_C)}, \tag{6a}$$

$$\Delta T_C = \frac{q'_{peak}}{2 \cdot \pi k_C} \cdot \ln\left(\frac{r_F + t_G + t_C}{r_F + t_G}\right), \tag{6b}$$

$$\Delta T_G = \frac{q'_{peak}}{2 \cdot \pi k_G} \cdot \ln\left(\frac{r_F + t_G}{r_F}\right), \tag{6c}$$

$$\Delta T_F = \frac{q'_{peak}}{2 \cdot \pi k_F}, \tag{6d}$$

where ΔT_F is the temperature drop across the fuel, ΔT_G is the temperature drop across the gap, ΔT_C is the temperature drop across the cladding, ΔT_b is the temperature drop across the bulk coolant, q'_{peak} is the peak linear heat generation rate i.e. linear heat generation rate multiplied by the radial and axial peaking factors, k_C is the thermal conductivity coefficient of the cladding, k_G is the thermal conductivity coefficient of the gap, and k_F is the thermal conductivity coefficient of the fuel. The output parameter is ΔP, and the input parameters are T_{in}, and W.

2.4. Energy conversion

Brayton cycle is used to analyze energy conversion of Helium. Efficiency of energy conversion from heat to electricity is important in the economics point of view. A higher efficiency is always desired. For simplicity, a simple model of the Brayton cycle with no regeneration, and reheating is used. The optimum pressure ratio ($r_{p,opt}$) is given by,

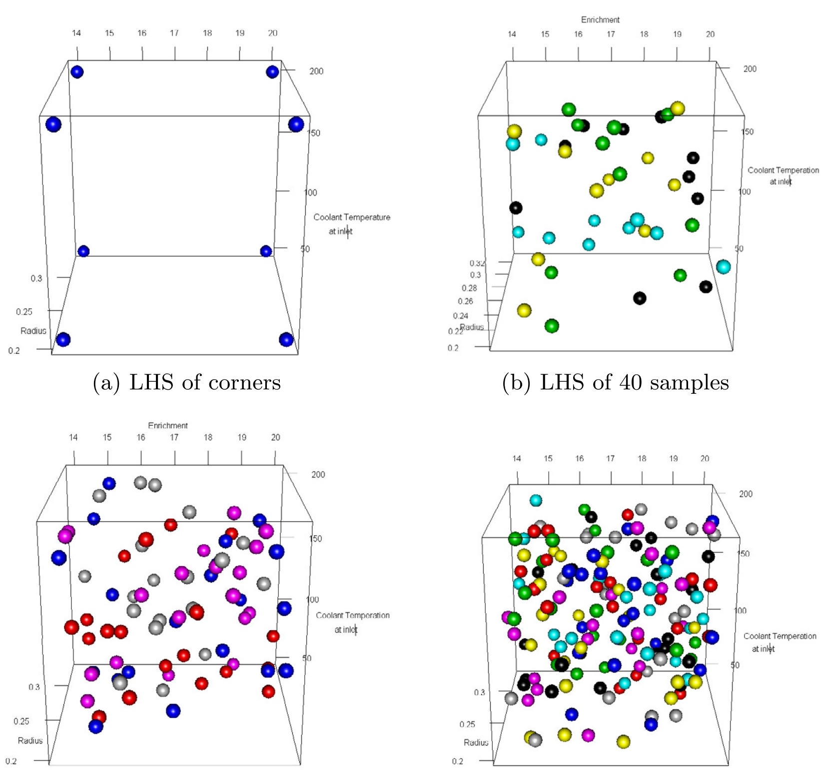
$$r_{p,opt} = \left(\frac{T_3}{T_1}\right)^{\frac{\gamma}{\gamma-1}}, \tag{7}$$

where, T_1 is the temperature of coolant at core inlet, T_3 is the temperature of coolant at core outlet, and γ is the heat capacity ratio. The amount of work done by the turbine per unit mass flow rate of the coolant ($\dot{W}_T$) is given by,

$$\dot{W}_T = \eta_T \cdot C_p \cdot T_3 \cdot \left[1 - \frac{1}{r_p^{\frac{\gamma-1}{\gamma}}}\right], \tag{8}$$

where η_T is the turbine efficiency, and C_p is the specific heat capacity of the coolant. The amount of work done by the compressor per unit mass flow rate of the coolant ($\dot{W}_{CP}$) is given by,

$$\dot{W}_{CP} = \frac{C_p \cdot T_1}{\eta_{CP}}\left[r_p^{\frac{\gamma-1}{\gamma}} - 1\right], \tag{9}$$



(a) LHS of corners (b) LHS of 40 samples

(c) LHS of 80 samples (d) LHS of 160 samples

Fig. 3. Sampling for determination of trial data for sensitivity analysis.

Please cite this article in press as: Kumar, A., et al. Linear regression and sensitivity analysis in nuclear reactor design. Ann. Nucl. Energy (2015), http://dx.doi.org/10.1016/j.anucene.2015.06.037

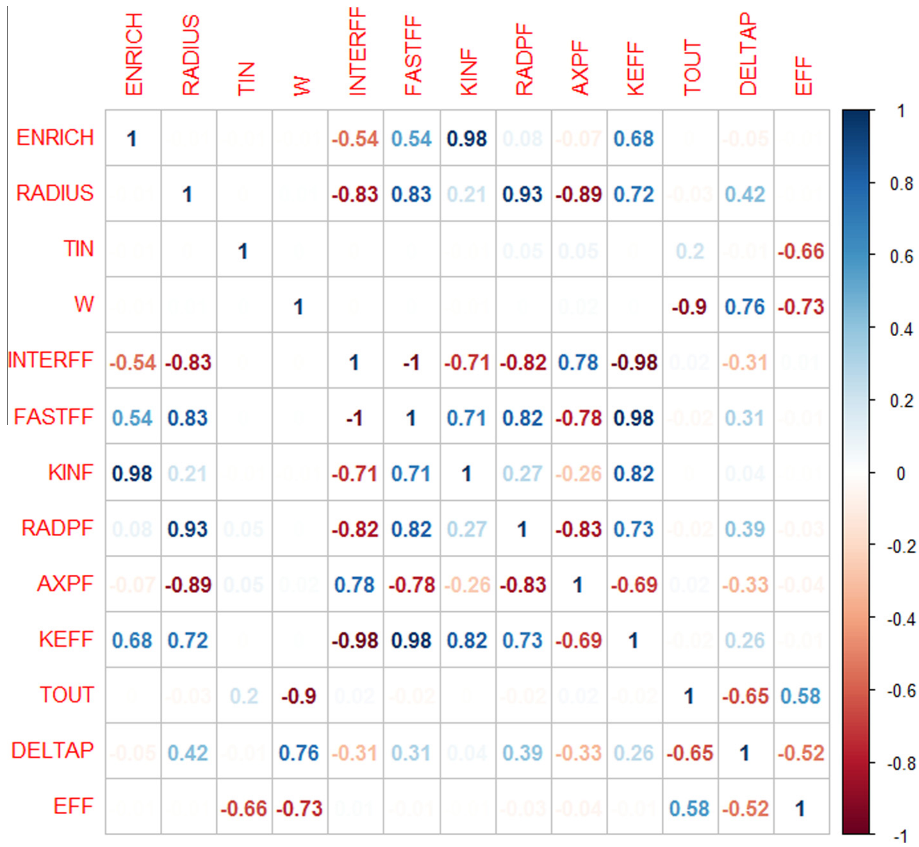


Fig. 4. Correlation plot with their coefficients.

where η_{CP} is the compressor efficiency. Therefore, the maximum amount of work done ($\dot{W}_{max}$) is given by,

$$\dot{W}_{max} = C_p \cdot T_1 \left[\frac{T_3}{T_1} - r_p^{\frac{\gamma-1}{\gamma}} \right]. \tag{10}$$

Thermal efficiency (η_{eff}) is given by,

$$\eta_{eff} = \frac{\dot{W}_T - \dot{W}_{CP}}{\dot{W}_{max}}. \tag{11}$$

The output parameter is η_{eff}, and the input parameters are, T_{in}, and T_{out}.

3. Linear regression

In this section a detailed understanding of LR is presented. In addition, all the tests performed for the validity of LR is discussed and analyzed. The coupled neutronics, and thermal–hydraulics system is explained in Section 2 can be represented in a simple functional form (Helton and Davis, 2003) as,

$$Y = f(X), \tag{12}$$

where, $X = [x_1, x_2, \ldots, x_K]$ is a vector of K input parameters, and $Y = [y_1, y_2, \ldots, y_L]$ is a vector of L output parameters. X is also called as predictors, independent variables, explanatory variables, control variables, and regressors in the research community. Similarly, Y is also called as response variables, dependent variables, explained variables, predictor variables, and regressands in several articles. Regression is a method of modeling a relationship between y_i, and X. When the relationship is done using a linear predictor function, it is called as linear regression (LR). Assuming a system is linear, Eq. (12) is represented as,

Table 1
Statistical distribution of test data.

Parameter	Mean	Stdev	Skew	Kurtosis
ENRICH	17.00	1.74	0.00	−1.22
RADIUS	0.26	0.04	0.00	−1.22
TIN	125.00	43.44	0.00	−1.22
W	60.00	23.17	0.00	−1.22
INTERFF	50.09	2.73	0.21	−0.94
FASTFF	49.91	2.73	−0.21	−0.94
KINF	1.47	0.07	−0.13	−1.17
RADPF	2.87	0.10	0.15	−0.85
AXPF	1.09	0.03	0.10	−0.72
KEFF	1.11	0.09	−0.11	−0.94
TOUT	512.62	192.23	1.05	0.36
DELTAP	82.24	70.83	1.96	6.16
EFF	49.58	10.44	0.23	−0.62

$$Y = X\beta + \epsilon, \tag{13}$$

where, β is the vector of regression coefficients, and ϵ is a vector of model error. Eq. (13) in an expanded form would look like,

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_K x_{iK} + \epsilon_i, \tag{14}$$

for an observation i. β, and ϵ are estimated using standard methods. Let the estimated coefficients be defined as $\hat{\beta}$, the fitted response $\hat{y}_i$ is given by,

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \cdots + \hat{\beta}_K x_{iK}, \tag{15}$$

the coefficient of determination R^2 is given by,

$$R^2 = \frac{\sum_{i=1}^{N} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{N} (y_i - \bar{y})^2}, \tag{16}$$

Please cite this article in press as: Kumar, A., et al. Linear regression and sensitivity analysis in nuclear reactor design. Ann. Nucl. Energy (2015), http://dx.doi.org/10.1016/j.anucene.2015.06.037

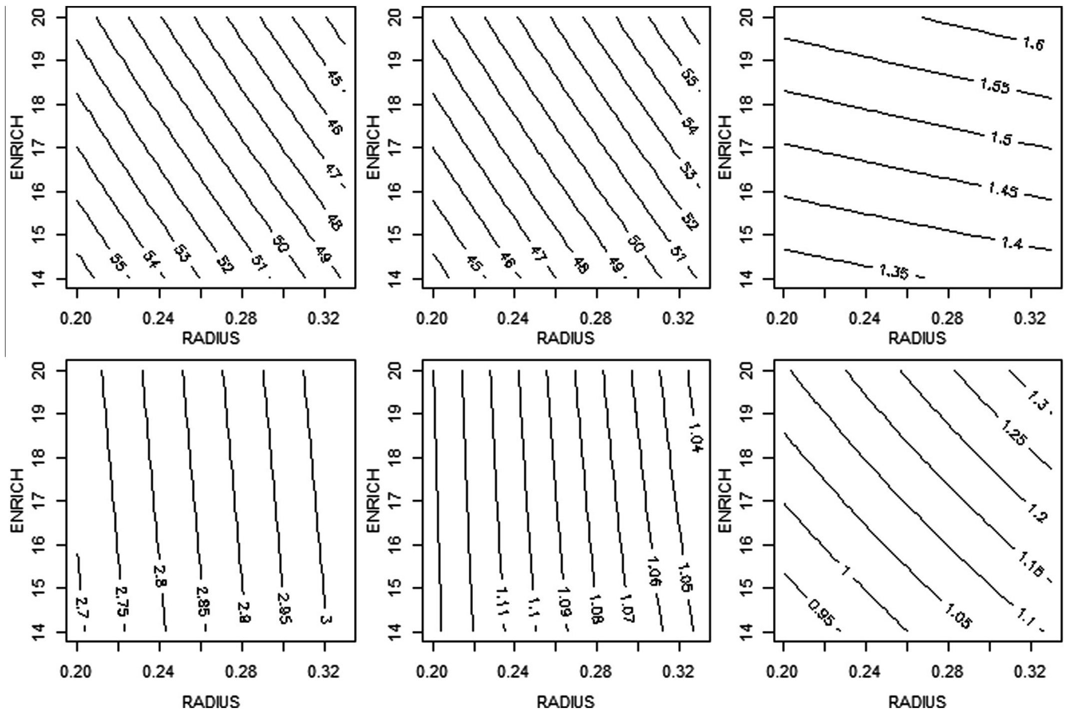


Fig. 5. ENRICH vs RADIUS contour plots for all the six predictors (clockwise from top left corner: INTERFF, FASTFF, KINF, RADPF, AXPF, KEFF).

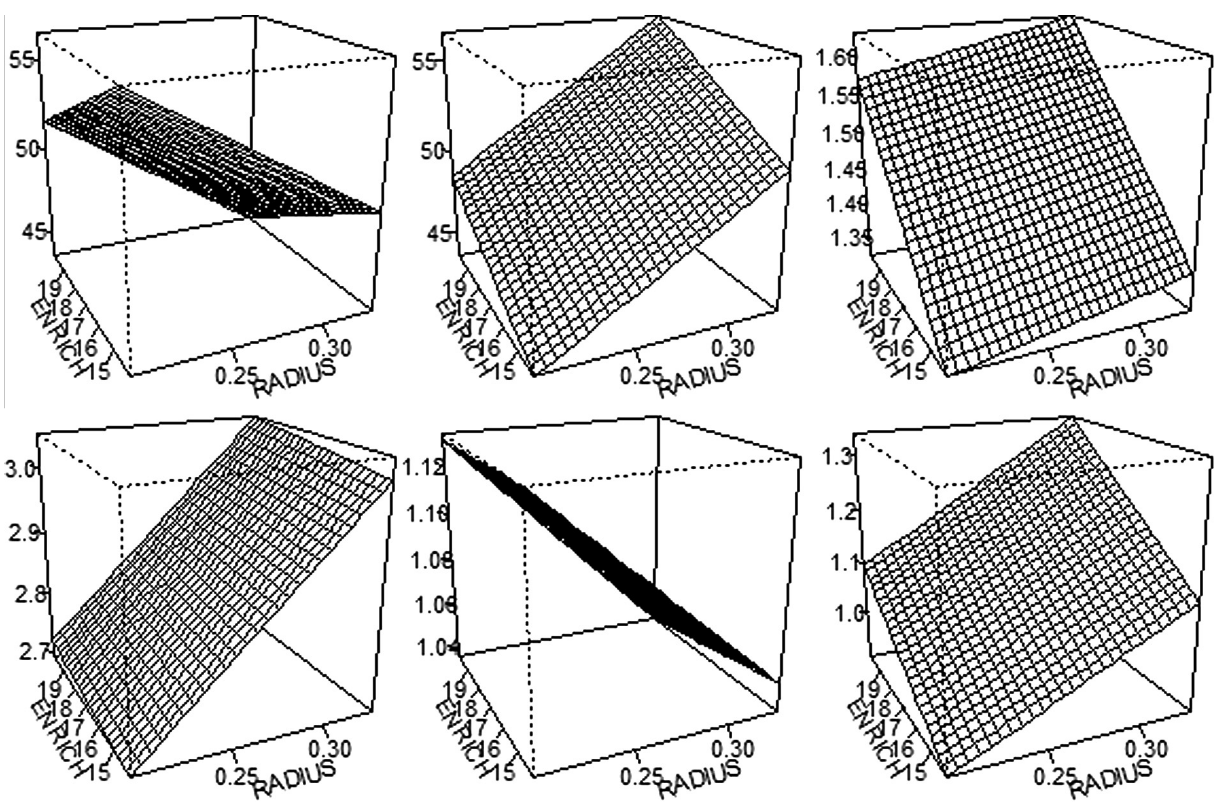


Fig. 6. ENRICH vs RADIUS surface plots for all the six predictors (clockwise from top left corner: INTERFF, FASTFF, KINF, RADPF, AXPF, KEFF).

where, $\bar{y}$ is the mean of all the observations, y_i. R^2 is a measure of the proportion of variation in y explained by the K input parameters. In this work, R (R Core Team, 2014) is used to determine β, R^2 and ϵ. For a system to have the relationship between the input and output parameters linear, following factors need to be met: linearity, independence, homoscedasticity, and normality.

Please cite this article in press as: Kumar, A., et al. Linear regression and sensitivity analysis in nuclear reactor design. Ann. Nucl. Energy (2015), http://dx.doi.org/10.1016/j.anucene.2015.06.037

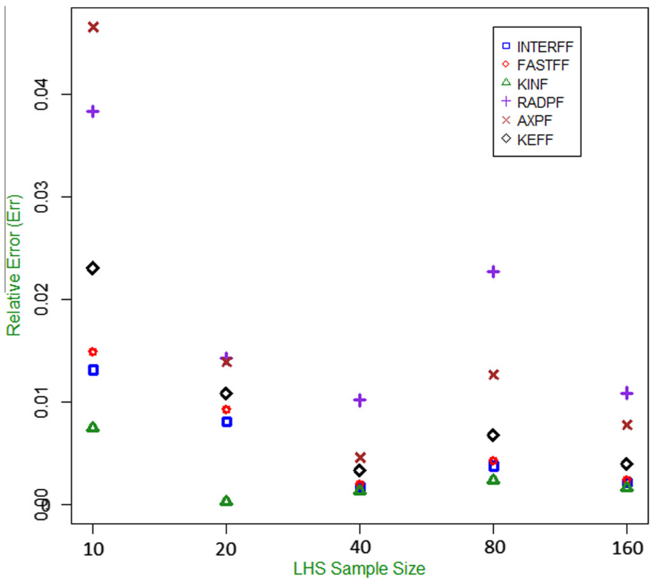


Fig. 7. Sample size vs relative error.

Table 2
Test data.

Parameter (P)	Value	Units
ENRICH	15.5	wt.%
RADIUS	0.2325	cm
TIN	87.5	°C
W	80	$\frac{kgs}{sec}$

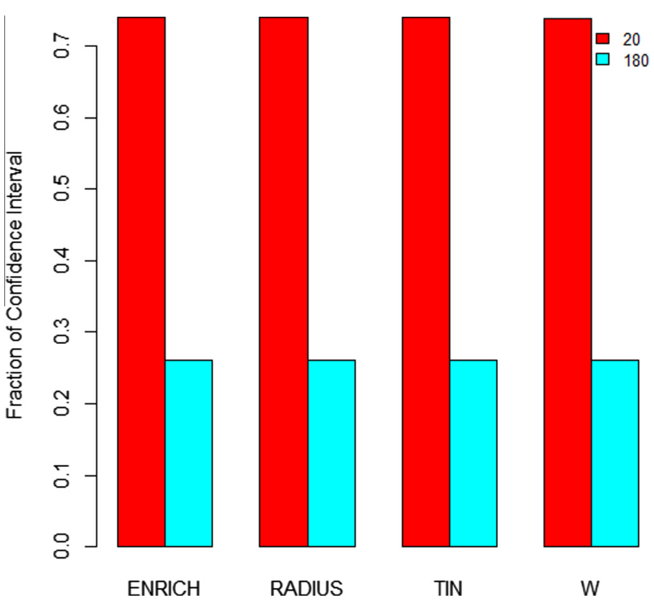


Fig. 8. Fraction of confidence interval vs sample size.

3.1. Linearity tests

Linearity behavior is validated if the expected value of the output parameter can be represented as a straight-line function of each input parameter, holding the other parameters fixed. The slope of the line should also have no dependence on the other parameters. In addition, the effect of the input parameters on the expected value of the output parameters should be additive.

3.1.1. Measure of linearity

A standard assumption for linearity is that the error should be zero everywhere, and any deviation from this assumption tends the functional form of the relationship between the input and the output parameters towards non-linearity. *Component* + *Residual* (partial-residual) plots are a good source of detection of non-linearity in the system. Partial-residual plots are formed by plotting the partial residuals (e_{ij}) against the input parameter (x_{ij}) where,

$$e_{ij} = e_i + b_i + x_{ij}, \tag{17}$$

where, e_i is the residual of the system, and b_i is the regression coefficient. For reference, a non-parametric regression smooth line, and the least squares line of data is presented in the partial residuals plot.

3.1.2. Measure of normality

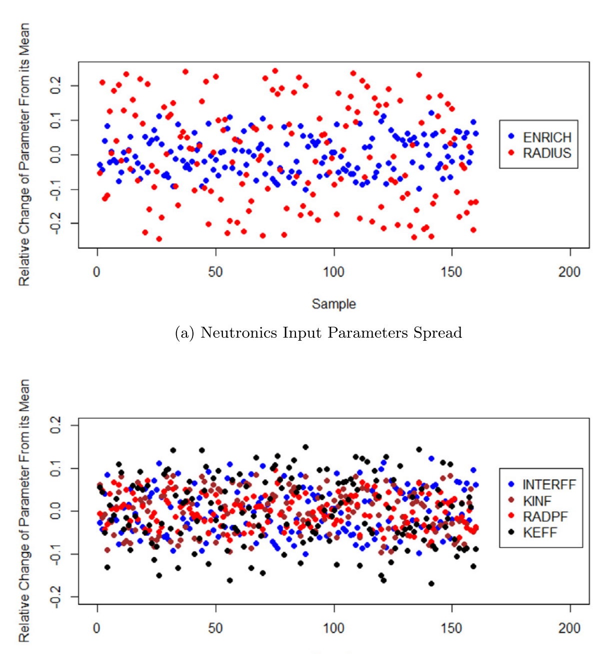
Test of normality is done by plotting the studentized residuals (Fox, 2002) against a t distribution with $n - p - 1$ degrees of freedom, where n is the sample size and p is the number of regression parameters. Studentized residuals plots are obtained using the mean-shift outlier model (Fox, 2002),

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_K x_{iK} + \gamma d_i + \epsilon_i, \tag{18}$$

where, d is a dummy parameter that takes a value 1 for a particular observation, i and 0 for all other observations, $\gamma \neq 0$ signifies that the conditional expectation of observation, i differs from the other

Table 3
Uncertainty quantification of input parameters.

Input parameter	Mean value	Range	δP in %	Units
ENRICH	17	14 – 20	17.64	wt.%
RADIUS	0.275	0.2 – 0.33	27.27	cm
TIN	125	50 – 200	60	°C
W	60	20 – 100	66.67	$\frac{kgs}{sec}$

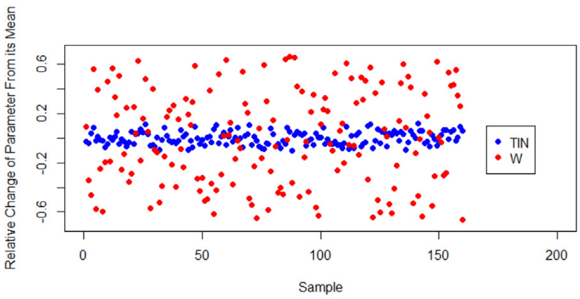


(a) Neutronics Input Parameters Spread

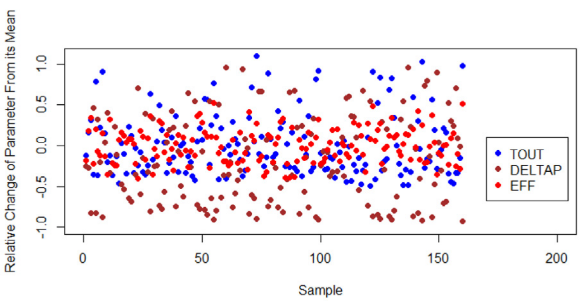
(b) Neutronics Output Parameters Spread

Fig. 9. Spread of the neutronics parameters.

Please cite this article in press as: Kumar, A., et al. Linear regression and sensitivity analysis in nuclear reactor design. Ann. Nucl. Energy (2015), http://dx.doi.org/10.1016/j.anucene.2015.06.037



(a) Thermal-hydraulics and Heat Transfer Input Parameters Spread



(b) Thermal-hydraulics and Heat Transfer Output Parameters Spread

Fig. 10. Spread of the thermal–hydraulics parameters.

observations. Then, t statistic test is determined to test the null hypothesis that the observation i does not differ from the rest of the observations. The t statistic is the studentized residual for the observation i. A studentized residual is the quotient resulting from the division of a residual by an estimate of its standard deviation. The t statistic is obtained for all the K observations, and then a quantile-comparison plot (Atkinson, 1985) with $n-p-1$ degrees of freedom of the studentized residuals, with a 95% confidence envelope is generated. Plotting is done using the qqPlot() function in R. A strong linear relationship is observed if the errors are normally distributed. Any deviation from the normal distribution shows a deviation from normality for that linear relationship.

3.1.3. Outliers

Outliers are observations that are not predicted well by the linear model due to presence of a significant residual. These observations have a greater influence on the regression coefficients as compared to other observations. These points are said to have a "high leverage" (Fox, 2002). A measure of leverage is done using the "hat values". Higher hat value for a point, means the model is overestimating/underestimating the output parameter for that point. The hat values (h_i) are obtained from a relationship between the fitted values ($\hat{y}_i$), and the observed response (y_i) given by,

$$\hat{y}_i = \sum_{j=1}^{N} h_{i,j} y_j, \tag{19}$$

where, $h_{i,j}$ is the weight attached to y_j in determining $\hat{y}_i$. The hat value h_i is given by,

$$h_i = \sum_{j=1}^{N} h_{i,j}^2. \tag{20}$$

In simple words, h_i summarizes the weights (Fox, 2002) associated with y_i, in the determination of all of the fitted values. A reference line is drawn at $2\bar{h}, 3\bar{h}$ for a detailed analysis, where $\bar{h}$ is the average hat value given by,

$$\bar{h} = \frac{K+1}{N}. \tag{21}$$

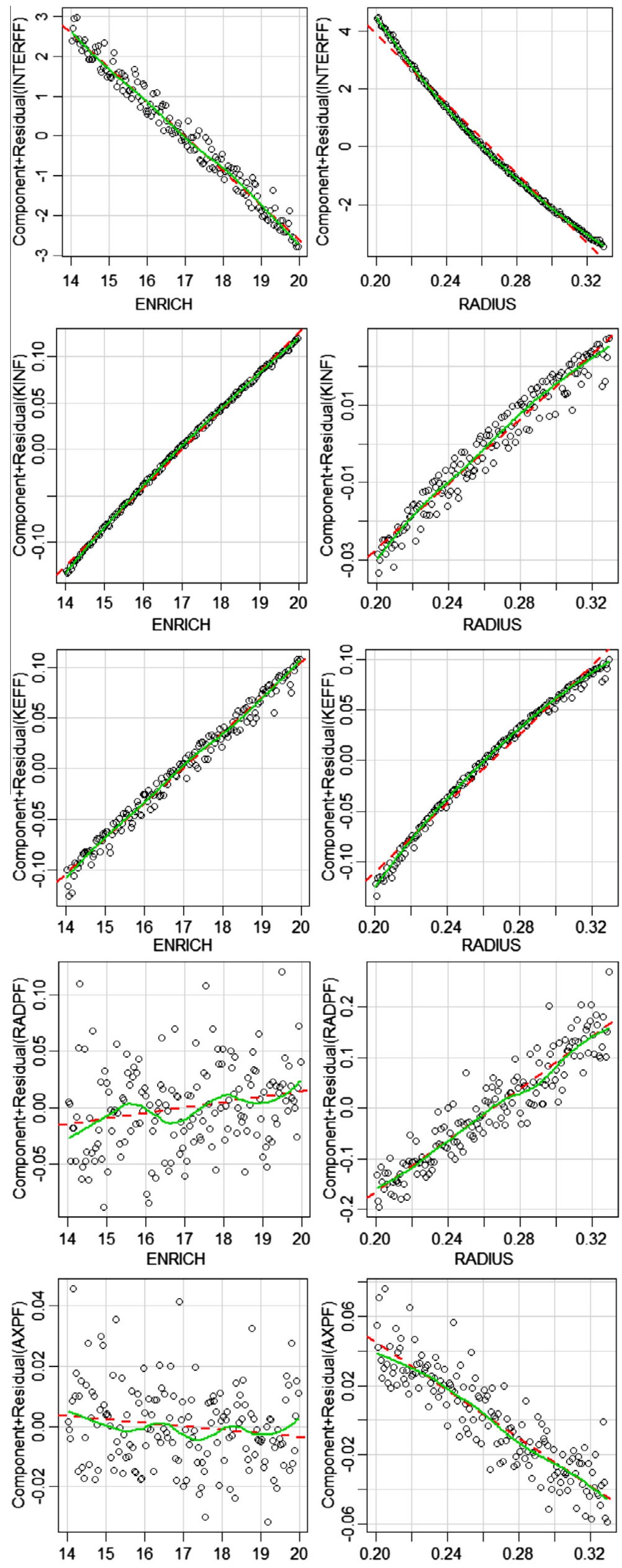
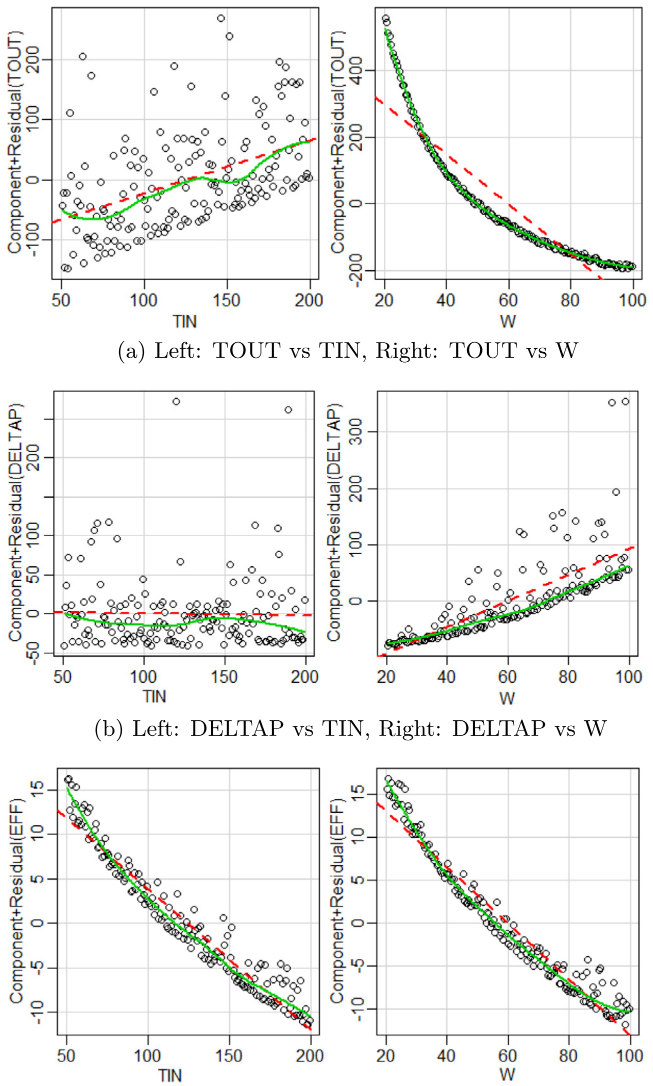


Fig. 11. Measure of linearity of the neutronics parameters.

3.1.4. Measure of importance

For the validity of LR in a complex system, the measure of the relative importance of predictors in a linear model is necessary. Relative importances are determined using the *relaimpo* package



(a) Left: TOUT vs TIN, Right: TOUT vs W

(b) Left: DELTAP vs TIN, Right: DELTAP vs W

(c) Left: EFF vs TIN, Right: EFF vs W

Fig. 12. Measure of linearity.

in R (Gromping, 2006). The functions used from this package provide bootstrap confidence intervals for relative importances. In Fig. 16, the confidence level indication of two input parameters versus all the four output parameters, of *lmg* (Lindeman et al., 1980; Johnson and Lebreton, 2004), *last*, *first*, and *pratt* (Pratt, 1987; Liu et al., 2014) metrics (Gromping, 2006) are analyzed. In *lmg*, the R^2 contribution averaged over orderings among the input parameters is analyzed. A detailed description of how *lmg* is computed is given in Gromping (2006, 2007). In the *first* metric (Gromping, 2006; Johnson and Lebreton, 2004), the R^2 values from all the K input parameters are compared with one input parameter. The R^2 values are identical to the squared correlations of the input parameters with the response. If the input parameters are correlated, the sum of these individual R^2 values would be larger than the overall R^2 value of the model. In the *last* metric (Gromping, 2006; Darlington, 1968), the increase in R^2 due to the inclusion on the input parameter as the last of the K parameters is analyzed. If the input parameters are correlated, the sum of these individual R^2 values would be smaller than the overall R^2 value of the model. In *pratt*, the product of the standardized coefficient and the correlation is analyzed.

3.1.5. Measure of association

In a multivariate regression analysis, it is useful to understand the strength of relationships between the predictors. The vcd package (Meyer et al., 2006) has been used to determine the Phi-Coefficient, Contingency Coefficient, and Cramer's V (Cramer, 1999) values. In simple words, the measure of association tells whether the relationship in the sample is likely to really exist in the general population or how likely it is to be an accident from sampling error. The N observations are divided into discrete n cells, with an assumption that values would occur in each cell in equal frequency. The expected frequency F_i for any cell is given by,

$$F_i = \frac{N}{n}. \tag{22}$$

The Pearson's chi-squared test-statistic (χ^2) is given by,

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - F_i)^2}{F_i}, \tag{23}$$

where, O_i is the observed frequency. The phi-coefficient ϕ is given by,

$$\phi = \sqrt{\frac{\chi^2}{N}}. \tag{24}$$

The contingency-coefficient C is given by,

$$C = \sqrt{\frac{\chi^2}{N + \chi^2}}, \tag{25}$$

and the Cramer's V value is given by,

$$V = \sqrt{\frac{\chi^2}{N(k-1)}}. \tag{26}$$

The statistic ϕ, C, and V have determined using the vcd package (Meyer et al., 2006) in R.

3.1.6. Homoscedasticity

Homoscedasticity, literally means "same variance " is a necessary condition for the validity of LR. Homoscedasticity is the property in which the error term in the relationship between the output variables and the input parameters is the same across all values of the input parameters. In other words, there is a constant variance of errors versus the observations, and any input parameter. One of the ways, by which homoscedasticity behavior is analyzed is, by plotting the absolute studentized residual against the fitted values. Then, a best fit line is drawn to the scattered data. The points form a random horizontal band around the best fit line. For non-linearity i.e. violation of the linearity assumption, the best fit line would not be horizontal. Homoscedasticity test is done using the *car* package (Fox and Weisberg, 2011) in R.

4. Results and discussion

In this section the results obtained from simulations and LRA have been presented and relevant discussions have been made. A detailed discussion on the theoretical background of the tests is given in Section 3.

4.1. Latin Hypercube Sampling

LHS (Iman et al., 1980) is used to perform UA and SA of the parameters described in previous sections. LHS has had severe acceptance in the scientific community, and apparently it is one of the most efficient sampling techniques as compared to other random sampling techniques (McKay et al., 1979; Helton and Davis, 2003; Xu and George, 2008). LHS stratifies the range of each variable into N disjoint intervals of equal probability and then one random value is drawn from each interval. Therefore, each parameter would have N random values. For a system with K parameters a $N \times K$ matrix is formed given by X,

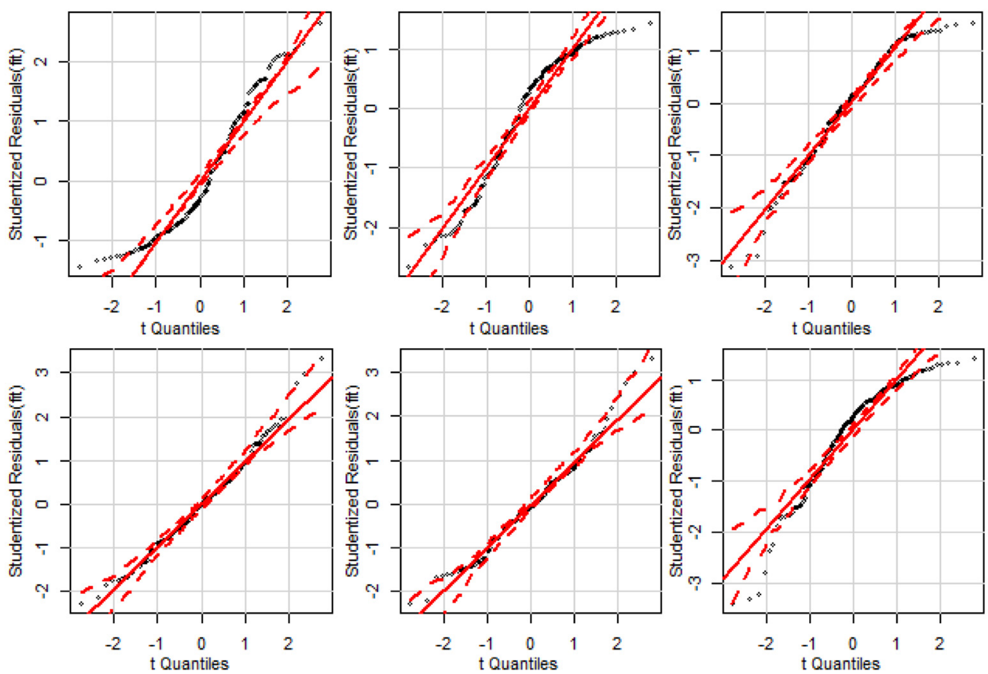
$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1K} \\ x_{21} & x_{22} & \dots & x_{2K} \\ \cdot & & \cdot & \cdot \\ \cdot & & \cdot & \cdot \\ x_{N1} & x_{N2} & \dots & x_{NK} \end{bmatrix} \tag{27}$$

where, each row corresponds to a sample for K parameters, and each column corresponds to the N samples for a specific parameter.

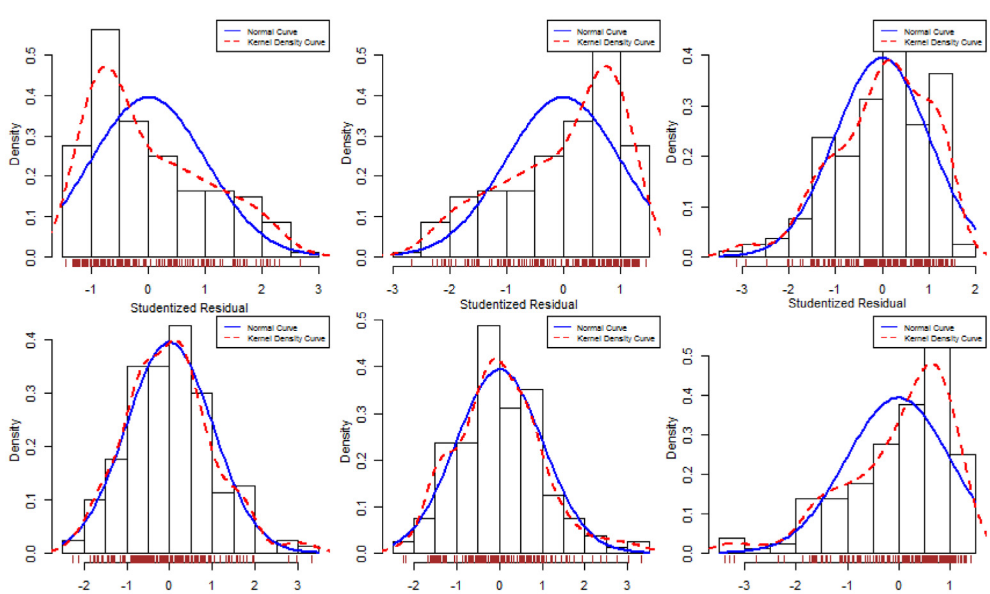
This set of input parameters is run by the computational model, that solves the equation, Eq. (13) to generate,

$$Y = \begin{bmatrix} y_{11} & y_{12} & \dots & y_{1L} \\ y_{21} & y_{22} & \dots & y_{2L} \\ \cdot & & \cdot & \cdot \\ \cdot & & \cdot & \cdot \\ y_{N1} & y_{N2} & \dots & y_{NL} \end{bmatrix} \tag{28}$$

where, Y is the set of output parameters for the specific set of input parameters, X, and L is the number of output variables. (see Fig. 3)

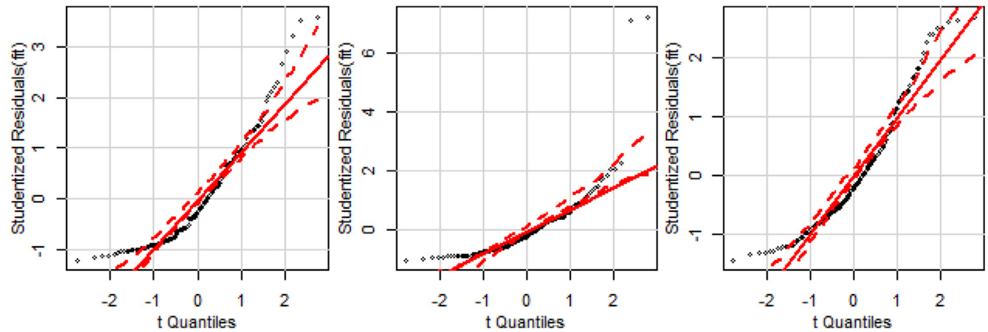


(a) QQ Plots of Neutronic Parameters.

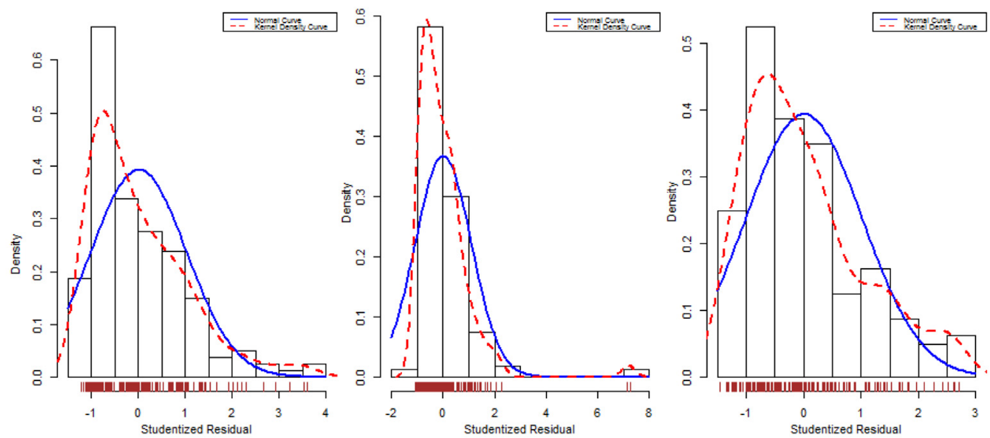


(b) Histogram of Distribution of Neutronic Parameters.

Fig. 13. Test of normality (clockwise from top left: INTERFF, FASTFF, KINF, RADPF, AXPF, KEFF).

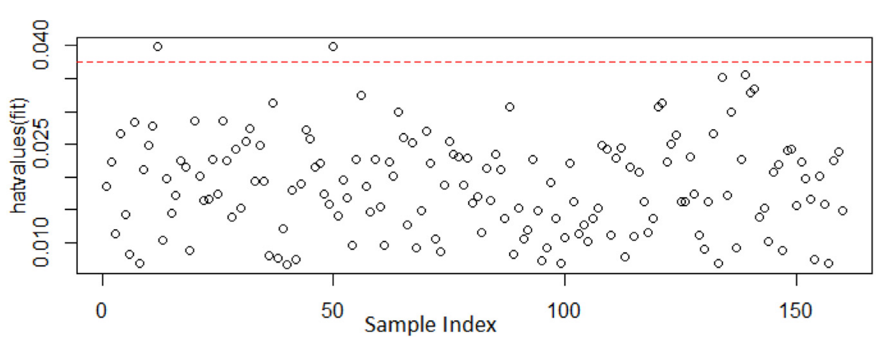


(a) QQ Plots of Thermal-hydraulics and Heat Transfer Parameters.

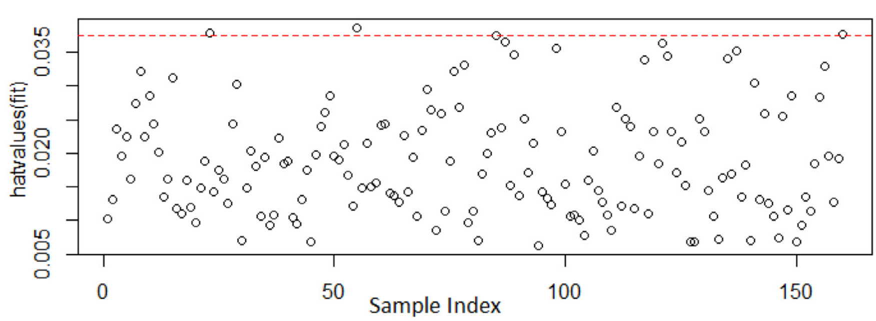


(b) Histogram of Distribution of Thermal-hydraulics and Heat Transfer Parameters.

Fig. 14. Test of normality (clockwise from top left: TOUT, DELTAP, EFF).



(a) Neutronics(KEFF vs ENRICH + RADIUS)



(b) Thermal-Hydraulics(EFF vs TIN + W)

Fig. 15. Outliers.

4.2. Correlation

Correlation between the parameters defined in Section 1 is presented in Fig. 4. This is built using the *stats*, and *corrplot* packages (Wei, 2013; R Core Team, 2014) in R. This figure can be thought of as a two dimensional matrix with each cell represented as a squared colored block. And, every cell correlates the parameters defined by the specific row and column name. Correlation values range from 0 to 1. The correlation value is printed in the cell with a sign. The sign represents if a set of parameters are positively correlated or negatively correlated. Higher the magnitude of the value, higher is the sensitivity of the output parameter to the input parameter. The figure, Fig. 4 helps in understanding the set of input parameters those affect each output parameter. It can be observed that the neutronics output parameters, INTERFF, FASTFF, KINF, RADPF, AXPF, and KEFF hold a value 0 with respect to TIN, and W. It can be inferred that the neutronics parameters are not affected due to changes in the thermal–hydraulics input parameters, and physically this is what is observed. From Eqs. (2)–(6d) it can be observed that DELTAP is related to RADIUS, and this correlation is observed in Fig. 4. INTERFF, and FASTFF are more sensitive to RADIUS as compared to ENRICH due to neutron thermalization. However, KINF shows an opposite trend. RADPF and AXPF are significantly affected due to changes in RADIUS as compared to ENRICH.

4.3. Data distribution

Analysis of the statistical details from the distribution of output parameters obtained from the system based on the input parameters present a basis linearity check on the behavior of the parameters. Statistical parameters such as skewness and excess kurtosis are necessary to understand the linearity relationship of the parameters. Table 1 presents the distribution statistics of all the parameters after the training data is obtained using the input samples obtained from LHS. Ideally, a normal distribution has skewness and excess kurtosis of 0, however from the table, it can be inferred that TOUT, and DELTAP have a deviation from normality in terms of skewness, and KINF, and DELTAP deviate from normality due to kurtosis. In addition, TOUT, and DELTAP can also be inferred to have a deviation from the linearity relationship. For most of the parameters the skewness is approximately symmetric, except for TOUT, and DELTAP. The thermal–hydraulics implementation is a lumped implementation of the Navier Stoke's equation. The equations have a significant number of inherent assumptions and are based on a non-fidelity model. Therefore, it a non normal distribution of TOUT, and DELTAP is not surprising.

4.4. Regression

A linear regression fit is made on the trial data using the *car* package (Fox and Weisberg, 2011) in R. Fig. 5 presents the ENRICH vs RADIUS contour plots for the regression fit of all the output parameters. The regression surface is shown in Fig. 6. Accuracy of the predictions depends on the number of samples used in trial data to build the regression fit. Let E_{rel} be the relative error for a parameter given by,

$$E_{\mathrm{rel}} = \frac{P_{\mathrm{pred}} - P_{\mathrm{actual}}}{P_{\mathrm{actual}}}, \tag{29}$$

where, P_{pred} is the predicted value of the output parameter from the regression fit, P_{actual} is the actual value from the simulations. Fig. 7 presents the E_{rel} versus sample size for all the neutronics output parameters. With an increase in the sample size, the error, and the confidence level of the predictions are affected. The test data for this analysis is given by Table 2. The analysis is done by predicting the output parameters for the test input data using a regression surface built with different sample sizes of the trial data. Larger the sample size, larger would be the confidence level. This phenomena is presented in Fig. 8. However, minor deviations from this phenomenon is observed, particularly for RADPF, and AXPF, which is due to the outliers in sampled data. In addition, RADPF and AXPF are obtained from the energy deposition tally in MCNP which strongly depends on the heterogeneity of the system. Due to using a simplified model with not reflectors and control elements, this behavior of TOUT, and DELTAP is valid.

4.5. Uncertainty quantification

In Table 3, the mean and the maximum deviation from the mean for all the input parameters used in this study is presented. These values are calculated from the LHS samples. In Fig. 9(a) the spread of the values of the neutronics input parameters in LHS samples is shown.

4.5.1. Neutronics

In Fig. 9, the spread of the neutronics output parameters based on the trial data is presented. In Fig. 10(a) the spread of the values of the thermal–hydraulics input parameters in LHS samples is shown.

4.5.2. Thermal Hydraulics and Heat Transfer

In Fig. 10(b), the spread of the thermal–hydraulics output parameters based on the trial data is presented.

4.6. Measure of linearity

The measure of linearity for the relationship between the output parameters INTERFF, KINF, KEFF, RADPF, AXPF, TOUT, DELTAP, and EFF, and the input parameters ENRICH, RADIU, TIN, and W is determined and the results are analyzed.

4.6.1. Neutronics

In Fig. 11 it can be observed that INTERFF, KINF, and KEFF have a linear relationship with ENRICH, and RADIUS. RADPF, and AXPF have a more linear relationship with RADIUS as compared to ENRICH. This behavior is very surprising, as the peaking factors depend on the energy deposition due to fission rate, and with increase in ENRICH the fission rate increases. A proportional relationship between the peaking factors and RADIUS is expected due to tha fact that, increase in the RADIUS decreases moderation, hence hardens the spectrum that enhances fission rate. Therefore the functional form of relationships RADPF vs ENRICH, and AXPF vs ENRICH show a non-linearity relationship. Linearity analysis has been performed using the *car* package (Fox and Weisberg, 2011) in R.

4.6.2. Thermal–hydraulics and energy transfer

From Fig. 12, it can be inferred that EFF has a linear relation with W (Fig. 12(c)) as compared to DELTAP (Fig. 12(b)), and TOUT (Fig. 12(a)). With a change is W, TOUT is affected, hence this varies the EFF. Hence, a linear relationship between EFF and W is expected. However, TOUT behaves most non-linearly with respect to TIN and W. This can be attributed to the fact that a lumped model for thermal–hydraulics is not very accurate.

4.7. Test of normality

Normality test is done using the QQ plots, and analyzing the distribution of residuals of all the output parameters.

4.7.1. Neutronics

Fig. 13(b) presents the distribution of residuals for all the neutronics output parameters. It can be observed that KINF, RADPF, AXPF, and KEFF have a desired normal distribution of residuals as compared to INTERFF, and FASTFF.

4.7.2. Thermal–hydraulics and heat transfer

Fig. 14(b) presents the distribution of residuals for all the thermal–hydraulics output parameters. It can be observed that EFF has a desired normal distribution of residuals as compared to TOUT, and DELTAP. This follows the same argument, that due to using a less-accurate lumped thermal–hydraulics model, TOUT, and DELTAP show a non-linear trend.

4.8. Outliers

Fig. 15 presents the scatter plot of the hat values (h_i) corresponding to all the samples for a regression surface of KEFF, and EFF. The red dotted line is the reference line at $2\bar{h}$. The circles above the red dotted line are the outliers. It can be observed that there are very few outliers in Fig. 15 which is expected from simulations. However, a different result would be observed if the model is an experiment, or is part of a human survey, or social data. The scatter plot of the hat values has been generated using the *car* package (Fox and Weisberg, 2011) in R.

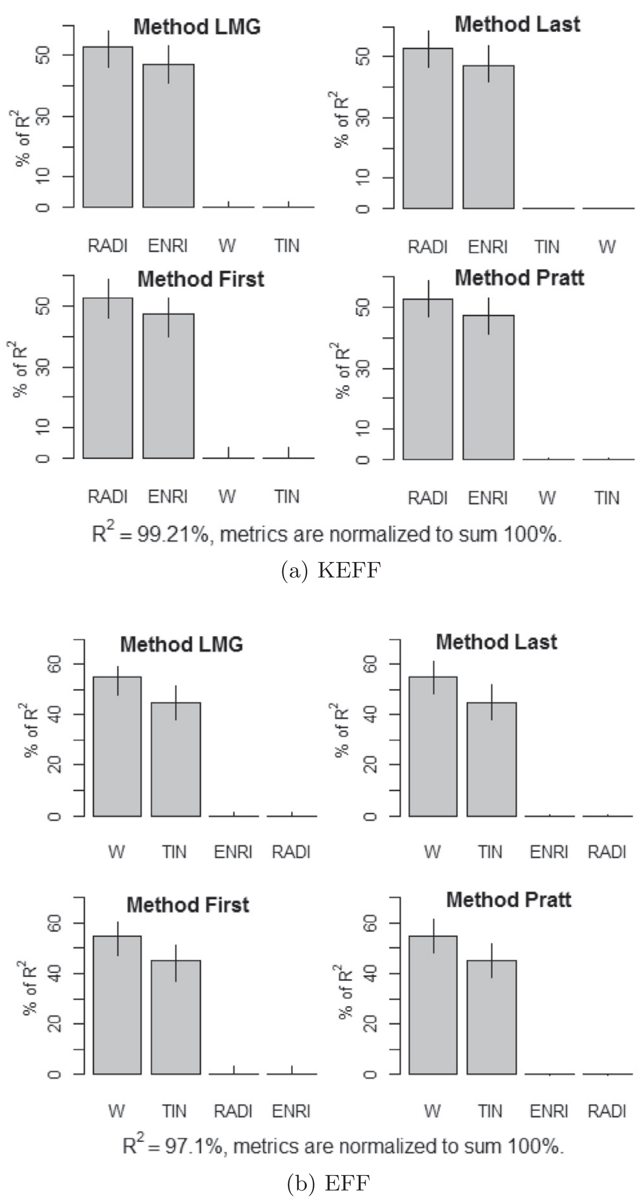


Fig. 16. Relative importances with 95% bootstrap confidence intervals.

4.9. Measure of importance

Fig. 16 shows the measure of the relative importance of the input parameters in the model. In the figure, the confidence level indication of two input parameters versus all the four output parameters, of *lmg*, *last*, *first*, and *pratt* metrics are presented. In Fig. 16(a), the importance of input parameters with respect to KEFF is shown, and it can be observed that TIN, and W have a zero importance on the KEFF of the model. and this is what is expected from the system. Because, the KEFF do not depend on the TIN, and W. Similarly from Fig. 16(b), it can be observed that the importance of RADIUS and ENRICH is zero on the EFF of the energy transfer system in the model.

4.10. Measure of association

The measure of association test is done by determining the Phi-Coefficient, Contingency Coefficient, and the Cramers V values. These values make us infer about the strength of the relationship among the predictors. These values also present a strength of interaction between the predictors. In the Table 4, the values for the three coefficients for the measure of association are high, hence there is a strong relationship between the predictor variables.

4.11. Homoscedasticity

Fig. 17 presents a scatter plot of the absolute standardized residuals versus the fitted values. From the figure it can be inferred that all the residuals have a very insignificant variance, hence an LR model can be used for the system.

5. Summary and future work

The linearity assumption for the implementation of LR in a nuclear reactor design, has been verified using a detailed sensitivity analysis (SA) of the parameters, and implementation of standard mathematical tests including, normality, independence, homoscedasticity, and association. Optimization of a nuclear reactor core using global optimization methods like GA, simulated annealing (SA) etc. necessitate a significant number of calls to the design simulator, which is very expensive to implement. However, with LR this call to the simulator is significantly reduced. With an assumption of linearity, a regression surface is built based on a set of trial data obtained from a series of MCNP runs. This surface behaves as an emulator of a nuclear reactor design, and predictive analysis is done using the emulator. Hence, instead of the simulator, the optimization methods calls the emulator. The research proved that certain parameters, such as effective and infinite neutron multiplication factor, fast fission factor, pressure drop across the core etc. have shown acceptance of linearity, however

Table 4
Measures of association of the predictor variables.

Coefficient	Value
Phi-coefficient	12.61
Contingency coefficient	0.997
Cramers V	1

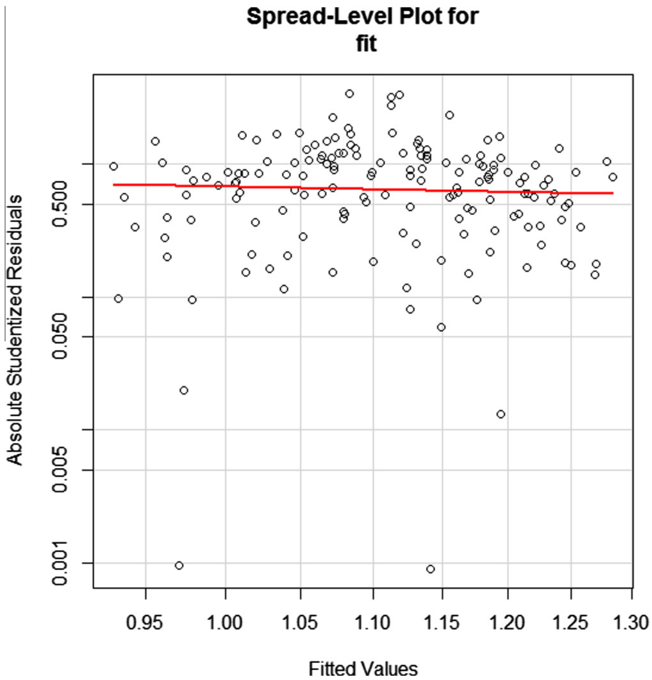


Fig. 17. Spread-level plot for analysis of the constant error variance.

parameters like power peaking factors, thermal efficiency etc. have shown a minor deviation from linearity that can be ignored. There were no strong indications that LR will be un-successful in modeling the reactor design. There is prospects for the application of LR in fuel cycle analysis, waste management, fuel shuffling, and most of other optimization problems in nuclear engineering. In addition, LR also helps in SA, and UA of parameters in many multivariate complex systems. Validity of LR in the design of a nuclear reactor modeled using a deterministic code would also be interesting to analyze and a comparative analysis between the stochastic as well as the deterministic model is useful in the research community.

References

Atkinson, 1985. Plots, Transformations, and Regression: An Introduction to Graphical Methods of Diagnostic Regression Analysis. Clarendon Press, Oxford.

Beck, M.B., 1987. Water quality modeling: a review of the analysis of uncertainty. Water Resour. Res. 23 (8), 1393–1442.

Blower, S.M., Dowlatabadi, H., 1994. Sensitivity and uncertainty analysis of complex models of disease transmission: an HIV model, as an example. Int. Statist. Rev. 62 (2), pp. 229-143.

Cramer, 1999. Mathematical Methods of Statistics, vol. 9. Princeton university press.

Darlington, 1968. Multiple regression in psychological research and practice. Psychol. Bull. 69, 161–182.

Fox, 2002. An R and S-Plus Companion to Applied Regression. Sage Publications, Thousand Oaks.

Fox, Weisberg, 2011. An R Companion to Applied Regression, Sage, Thousand Oaks CA <http://socserv.socsci.mcmaster.ca/jfox/Books/Companion>.

Gromping, 2006. Relative importance for linear regression in R: the package relaimpo. J. Stat. Softw. 17 (1), 1–27.

Gromping, 2007. Estimators of Relative Importance in Linear Regression Based on Variance Decomposition. Am. Stat. 61 (2).

Hamby, D.M., 1994. A review of techniques for parameter sensitivity analysis of environmental models. Environ. Monit. Assess. 32, 135–154.

He, Minxue, Hogue, Terri S., Franz, Kristie J., Margulis, Steven A., Vrugt, Jasper A., 2011. Characterizing parameter sensitivity and uncertainty for a snow model across hydroclimatic regimes. Adv. Water Resour. 34, 114–127.

Helton, Jon C, 1993. Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal. Reliab. Eng. Syst. Saf. 42 (2,3), 327–367.

Helton, Davis, 2003. Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. Reliab. Eng. Syst. Saf. 81, 2369.

Homma, Toshimitsu, Saltelli, Andrea, 1996. Importance measures in global sensitivity analysis of nonlinear models. Reliab. Eng. Syst. Saf. 52 (1), 1–17.

Iman, R.L., Davenport, J.M., Zeigler, 1980. Latin Hypercube Sampling (A Program User's Guide), Technical Report SAND79-1473, Sandia Laboratories, Albuquerque.

Johnson, Lebreton, 2004. History and use of relative importance indices in organizational research. Organ. Res. Methods 7, 238–257.

Kumar, Akansha, Tsvetkov, Pavel V., 2015. A new approach to nuclear reactor design optimization using genetic algorithms and regression analysis. Ann. Nucl. Energy 85, 27–35.

Kumar, Akansha, Tsvetkov, Pavel V., Chirayath, Sunil S., 2014. Analysis of a sustainable gas cooled fast breeder reactor concept. Ann. Nucl. Energy 69, 252–259.

Lindeman, Merenda, Gold, 1980. History and use of relative importance indices in organizational research. Introduction to Bivariate and Multivariate Analysis. Scott, Foresman, Glenview, IL.

Liu, Zumbo, Wu, 2014. Relative importance of predictors in multilevel modeling. J. Mod. Appl. Stat. Methods 13 (1), 2–22.

Mccuen, Richard H., 1973. The role of sensitivity analysis in hydrologic modeling. J. Hydrol. 18, 37–53.

McKay, M.D., Beckman, R.J., Conover, W.J., 1979. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics 21, 23945.

Meyer, Zeileis, Hornik, 2006. The strucplot framework: visualizing multi-way contingency tables with vcd. J. Stat. Softw. 17 (3), 1–48 <http://www.jstatsoft.org/v17/i03/>.

Pratt, 1987. Dividing the indivisible: using simple symmetry to partition variance explained. In: Proceedings of the Second International Conference in Statistics. University of Tampere, Tampere, Finland.

R Core Team, 2014. R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria <http://www.R-project.org/>.

Saltelli, A., Homma, T., 1992. Sensitivity analysis for model output: performance of black box techniques on three international benchmark exercises. Comput. Stat. Data Anal. 13 (1), 73–94.

Saltelli, A., Andres, T.H., Homma, T., 1993. Sensitivity analysis of model output: An investigation of new techniques. Comput. Stat. Data Anal. 15 (2), 211–238.

Saltelli, Chan, Scott, 2000. Sensitivity Analysis Wiley Series in Probability and Statistics. Wiley, New York.

Wei, 2013. corrplot: visualization of a correlation matrix R package version 0.73. <http://CRAN.R-project.org/package=corrplot>.

X-5 Monte Carlo Team, 2008. MCNP – A General Monte Carlo N-Particle Transport Code. Version 5, User Manual, Los Alamos National Laboratory, 416.

Xu, Chonggang, George, Zdzislaw Gertner, 2008. Uncertainty and sensitivity analysis for models with correlated parameters. Reliab. Eng. Syst. Saf. 93, 1563–1573.