Model validation with multivariate responses via factor analysis

Danqing Wu

The 28th Research Institute of China Electornics Technology Group Corporation

Email: wudanqing86@163.com

Abstract. Computational models have been widely applied to simulate complex physical phenomena because of rapid development of computers’ abilities and inexecutable conditions of the experiments. Validating the effectiveness of computational models is then essential and significant. Existing validation metrics focusing on a single response may cause confusing results when they are applied for multiple responses simultaneously. The target of this work is to validate computational models with multiple responses via the factor analysis method by considering both the uncertainty and correlation of the multiple responses. Factor analysis aims to explain the correlated multi-dimensional variables with fewer common factors and the latter then can describe the main information the former containing. Briefly, factor analysis has a favorable ascendency——dimensionality reduction, which makes multiple responses validation easy to implement. The proposed method is based on the factor analysis and the concept of “area metric” for validating a single response, which avoids comparing the joint distribution of the computational models and that of the experimental observations. It is applicable for validating multiple responses both at a single validation site and at multiple validation sites. Both numerical example and engineering example are employed to demonstrate the rationality and necessity of the proposed methodology.

1. Introduction

Limited by the huge cost of physical experiments and complex engineering environment, also with the rapid development of computers’ abilities, computational models are increasingly established to simulate the system behavior and support decision in many research fields, such as engineering design, performance estimation, and risk analysis[1]-[3]. The increased dependence on applying computational simulation models makes it necessary and important to find an approach to measure the accuracy and adequacy of a computational model. Model validation is a major procedure to assess the degree of the computational models matching the physical observations. Generally, model validation is defined as the process of determining the degree to which a model is an accurate representation of the real world [1]-[5]. Model validation provides substantiation of how accurately the computational model simulates the physical observations. Various validation metrics have been proposed to measure the mismatch between the computational simulations and the physical observations. The existing model validation methods can be generally classified into four categories [6]: classical hypothesis testing [7, 8], Bayes factor [9, 10], frequentist’s metric [11, 12], and area metric [13].

The Classical hypothesis testing cannot assess the quantitative accuracy of a model as it mainly focuses on determining which of the two alternative propositions is correct. The Bayes factor approach is primarily interested in evaluating the probability (i.e., the belief) that the model is correct by incorporating an analyst’s prior belief of model validity. Though the Bayes factor approach has some
advantages over the classical hypothesis testing method, it still cannot provide quantitative accuracy of the model with a “yes” or “no” answer [13]. The Frequentist’s metrics can assess the adequacy of the given model for an application of interest quantitatively. However, it only considers the central tendencies or other specific behaviors of data and predictions rather than their entire distributions, as it is based on comparing means or other summary statistics (e.g. the maximum values) of the experimental data and model predictions. With this consideration, Ferson et al [13] proposed an area metric method. The area metric assesses the accuracy of the given model by comparing the difference between the cumulative distribution function (CDF) of the model prediction and the empirical CDF of experimental data. It is popular and widely used due to its favorable characteristics [14]. According to its definition, Ferson’s area metric compares the marginal distribution of the model predictions with that of the experimental data, thus it is only suitable for validation of models with single response or with uncorrelated multiple responses [15]. However, the engineers are always interested in correlated multiple responses. For instances, multiple response quantities, like stress, strain, displacement etc, often are predicted simultaneously from the same experiment at a single location. The different quantities are different functions of the same inputs, thus there is no doubt that they are correlated. In addition, the interested model responses from the same experiment may change with location (in space or time coordinates). In this case, there is a strong correlation between any two pairs of response quantities from the same experiment. As a result, these responses quantities make the experiment as a multivariate response problem. Although the classical hypothesis testing and the Bayes factor have been extended for validation of models with multivariate responses [16, 17], these methods cannot avoid their short comings in the univariate case.

Li et al [15] extended the concept of “area metric” and “u-pooling method” with the multivariate probability integral transformation (PIT) theorem. The proposed method can assess the accuracy of the computational models with correlated multivariate responses. The extending metrics are separately PIT area metric for a single validation site and t-pooling metric for multiple validation sites. The PIT area metric and t-pooling metric consider both the uncertainties and correlations of the multivariate responses in the process of model validation, thus it makes the validation results creditable. However, they require estimating the joint cumulative distribution function (CDF) of model responses to transform the multivariate experimental observations, which is often very difficult for high dimensional response space.

In this paper, a new method is proposed for the validation assessment of models with multiple correlated responses based on the factor analysis (FA) and the idea of “area metric”. The FA method describes the correlated variables with some common factors and special factors. Then the fewer common factors can be used to represent the initial high dimension variables. It aims to find several common factors which can represent complete information of a whole system [18]. Then the area metric can be applied to each common factor to establish the corresponding validation metric value. The total model validation metric is established by aggregating these validation metric values of common factors. With the proposed FA method, model validation with multiple responses can be simplified into a validation for the common factors. Compared with the PIT area metric, FA method is more feasible in practical, as the latter avoids the joint CDF estimation of the multiple responses.

The remainder of the paper is organized as follows: Section 2 briefly reviews the existing area metric and u-pooling technique of the model validation. Section 3 describes the theory of the proposed FA method. An illustrative mathematical example is used to show the advantages of the proposed method by comparing with the existing ones in Sections 4. The proposed FA method is also applied to an engineering example where the experimental data is assumed to be sparse in Section 5. Finally, the conclusions come in section 6.

2. A review of the area metric and u-pooling metric

Area metric is widely used because of its desirable features: objectiveness, autocephaly and with uncertainties. It uses entire distributions of the computational prediction data and the physical observations to assess the mismatch. $F^x$ denotes the CDF of the response $y$ predicted by the
computational model at a single validation site, and $F^e$ is defined as the empirical observation distribution. The area surrounded by $F^m$ and $F^e$ can be estimated by equation (1):

$$d(F^m, F^e) = \int_{-\infty}^{\infty} |F^m(y) - F^e(y)|\,dy$$  

(1)

When the predictions and experimental data are adequate, it is easy to distinguish that a smaller area metric indicates a better match between the model predictions and the experimental observations at the specified validation site compared with a larger one. Either a not so good computation model or a lack of data can be a causation of a larger area metric.

Equation (2) illustrates the agreement of the predictions and the observations at a single validation site. It does not work well when the distributions of predictions and observations are collected at multiple validation sites. To solve the problem, Ferson et al [13] proposed the u-pooling metric. The u-pooling metric pools all physical observations over the intended prediction domain at multiple validation sites into a single aggregated metric by transforming the data at different validations into a universal probability scale through the corresponding predictive distributions. At each single validation site $z_j$, the $i$-th physical observation can be translated into a scalar of $u_{ji}$ by utilizing prediction distribution $F^m_{z_j}(\cdot)$ of model data. The translation can be achieved by:

$$u_{ji} = F^m_{z_j}(y^e_{zi}(i))$$  

(2)

Through the u-pooling process, the entire mismatch between the observations and predictions at different validation sites is summarized into a single aggregated area metric, which provides a global assessment for multiple validation sites.

Taking three validation sites as an example, figure 1 shows the process of u-pooling. Figure 1 (b) provides an illustration of calculating the values $u_{ji}$ for two observations at every single validation site $z_1$, $z_2$ and $z_3$. The three groups of observations are respectively $(y^e_{zi}(1), y^e_{zi}(2))$, $(y^e_{zi}(1), y^e_{zi}(2))$, and $(y^e_{zi}(1), y^e_{zi}(2))$. Similar to the area metric, when the predictions and experimental data are sufficient, a larger value of u-pooling metric would indicate stronger evidence for disagreement between the model predictions and the experimental measurements over the intended prediction domain (i.e. at the multiple validation sites). As both the standard uniform distribution and an empirical distribution function of the transformed u-values are constrained to the unit square, figure 1 (a) illustrates the u-pooling metric which has a domain of $[0, 0.5]$. According to the domain, a zero-value metric represents a best computation model while a 0.5-value metric represents a worst one.

![Figure 1](image.png)

**Figure 1.** Illustration of the u-pooling method (a) Area metric of u-values and the standard uniform distribution (b) u-values at multiple validation sites.
3. The proposed method applied in model validation

3.1. An introduction of the factor analysis

Factor analysis is a virtual method for dimensionality reduction in the field of multivariate statistical analysis. It serves as an important tool for pattern recognition [19] and it has popular application in many fields. FA has been used successfully for feature selection of color guard members [20], stock market prediction [21], delay analysis [22], and development of a CAD tool for diagnosis of the Alzheimer’s Disease [23].

FA was first proposed in the psychology area to explain people’s mental ability with several common factors. Generally, FA aims to explain correlated multi-dimensional variables with fewer common factors. It makes the structure of correlations among measured variables understood by estimating the pattern of relations between the common factors and each of the measured variables. In another word, it is a statistical method that finds the common latent factors for the variables using the correlation coefficients. FA can be divided into two sorts: the exploratory factor analysis (EFA) and the confirmatory factor analysis (CFA). When the researchers are unaware of the common latent factors, and they attempt to identify the latent factors, the EFA is then carried out to find out the common latent factors. The CFA is used to examine whether the studied data agree with the theoretical framework, while the theoretical framework of the latent factors are known. In the process of model validation, the engineers focus on the form of different computational models other than the theoretical framework, hence they have no knowledge about the theoretical framework, and as a result the EFA is applied in this work.

Figure 2 depicts the sketch map of FA. Correspondingly the mathematic model of FA can be expressed as equation (3):

\[
\begin{align*}
Y_i &= a_{i1}F_1 + a_{i2}F_2 + \ldots + a_{im}F_m + e_i \\
Y_2 &= a_{21}F_1 + a_{22}F_2 + \ldots + a_{2m}F_m + e_2 \\
&\vdots \\
Y_m &= a_{m1}F_1 + a_{m2}F_2 + \ldots + a_{mm}F_m + e_m \\
\end{align*}
\]

where \( F_1, F_2, \ldots, F_m \) are the common factors, and they can be expressed as linear combinations of all initial variables. \( e_i \) \((i = 1, 2, \ldots, m)\) are the specific factors which describe the unique parts of each
initial variable. $a_{ij}$ is the factor loading. It is the correlation coefficient between $Y_i$ and $F_j$, which depicts the dependency degree $Y_i$ has on $F_j$.

Several methods can be used to find out the common factors, for instance, the principal component analysis method [24], the maximum likelihood method [25], the minres method and so on. Compared with how to execute factor analysis, the authors care more about how to apply the FA method into model validation, hence the FA process is done by the R language which helps the authors to decide the common factors with the tolerant minres method.

When there are more than one common factors for the measured variables, there exists an infinite number of alternative orientations of the factors in multidimensional space that will explain the data equally well. This indicates that EFA models with more than one factor do not have a unique solution. Therefore, a researcher must select a single solution from the infinite number of equally fitting solutions [26]. Thurstone [27] suggested that factors should be rotated in multidimensional space to satisfy the simple structure principle. Two sorts of factor rotations have been developed, i.e., the orthogonal rotation and oblique rotation. Orthogonal rotations constrain the common factors to be independent from each other. Among the several developed orthogonal rotations, varimax rotation [28] maximizes the variance of squared factor loadings across variables. According to its peculiarity, the varimax rotation has universally been considered as the best orthogonal rotation and is the most popularly used one in practical application [26]. In contrast to orthogonal rotations, oblique rotations permit correlations among factors. To make the relation between different factors concise and perspicuous, the varimax rotation is employed in this work.

After the rotation, we can focus on the factor scores which can describe the initial variables with common factors as shown in equation (4).

\[
\begin{align*}
F_1 &= s_{11} Y_1 + s_{12} Y_2 + \ldots + s_{1m} Y_m \\
F_2 &= s_{21} Y_1 + s_{22} Y_2 + \ldots + s_{2m} Y_m \\
&\vdots \\
F_n &= s_{n1} Y_1 + s_{n2} Y_2 + \ldots + s_{nm} Y_m
\end{align*}
\]

where $S = [s_{ij}]$ $(i = 1, \ldots, n, j = 1, \ldots, m)$ is the factor score matrix. Thus, the common factors are sufficient to represent the whole variability of the original variables, which can reduce the number of dimensions, while retaining as much as possible of the data’s variation. Instead of investigating multidimensional original variables, the first few common factors containing the majority of the data’s variation are explored. The visualization and statistical analysis of these new variables, the common factors, can help to find similarities and differences between samples. These considerations motivate the use of FA for validation assessment of model with multivariate responses.

3.2. New validation method for model with multivariate responses based on FA

The multiple responses may be predicted at a single location, or single or multiple correlated model responses maybe collected at different locations (in different space or time coordinates). In both cases, the dimensionality of the multiple responses may be very high, and a strong correlation exists between any a pair of responses. Validation of model in these cases requires considering both the correlations and uncertainties of the multiple responses. Obviously, for the high dimensionality of these multiple responses in practice, direct comparison of the model and experiment may be impossible. Whereas, not each response in the multiple responses has significant contribution to the whole uncertainty of responses, and there is usually some degree of information redundancy among the responses. If the main information can be reserved and the redundant information eliminated in the validating assessment process, the problem can be simplified. FA provides an excellent tool for extracting the main information of the multiple variables. Therefore, FA is combined with the idea of area metric to assess the predictive capability of the models with multivariate responses in this work.
It is reasonable if a model validation metric can consider uncertainty in the computation data and the physical observations. Several types of uncertainty in engineering computation models and physical experiments have been classified in the work of Kennedy and O'Hagan [29], i.e., parameter uncertainty (e.g. boundary condition of the partial differential equation), model inadequacy, residual variability, the experimental uncertainty in the form of measurement error, systematic error, random error and so on [30]. Among them, input data uncertainty and the model parameter uncertainty are the main concerns of the model validation, which are also considered in this paper. The variables in a model can be sorted in three kinds: input variables $x$, model parameters $\theta$ and site variables $z$. $x$ and $\theta$ can be either deterministic or nondeterministic. $z$ are deterministic space or time domain variables which represent the sites where to validate the model. The Model-related uncertainty of input variables is studied in this work. The physical experiment in this work can be described as $y = f(x, z)$, and correspondingly the computational models is formed as $y_m = f(x, z, \theta)$, where $x$ are nondeterministic while $\theta$ and $z$ are deterministic.

According to our consideration with both the correlations and uncertainties of the responses in this work, if the data of a physical experiment are collected at $d$ validation sites with $p$ responses at each site, the output should be regarded as a $m$ ($m = dp$) dimensional responses. The main procedure of the proposed method is illustrated in figure 3 and explained as follows:

![Figure 3. The procedure of model validation with the FA method.](image)

1) Data collection: collect the observations with a number of $Nm$ and calculate the responses of the candidate model with a number of $Mm$ ($M > N$), where $N$ is the data mount of the each dimensional response’s observations and $M$ is the data mount of the each dimensional response’s candidate model.

2) Factor analysis: Use the factor analysis method for the candidate model and find out appropriate common factors to reduce the dimension (the R language is relied for in this work). Assume there are
$n$ common factors needed to be considered, then the dimension of the problem decreases from $m$ to $n$ ($m>n$). The factor matrix $B$ then can be estimated.

$$F^n = By^n$$  (5)

3) Data transformation: The physical observations can be transformed with the factor matrix $B$, and correspondingly the common factors of the observations $F^c$ can be obtained as:

$$F^c = By^c$$  (6)

4) Area metric for each common factor: Estimate the CDF of the factors of the predictions, i.e., $C_i^w(F)$ and the empirical CDF of the factors of the observations, i.e. $S_i^c(F)$. Then compare the difference between $C_i^w(F)$ and $S_i^c(F)$ by the area metric, i.e.

$$d_k(C,S) = \int_{-\infty}^{\infty} |C_i^w(F) - S_i^c(F)| dF, \quad k = 1,2,\ldots,n$$  (7)

5) Aggregate area metric for all common factors: Estimate the overall validation metric for multivariate response by aggregating these validating metrics of each factor, that is

$$d(C,S) = \sum_{k=1}^{n} d_k(C,S)$$  (8)

According to the theory and procedure of the proposed method, the FA method simplifies validation with multiple responses through dimension reduction. Compared with the PIT area metric for multiple responses, it avoids estimating the joint CDF of the model responses which is a tough task for high-dimensional problem. As the aggregate area metric may obscure the differences between the observations and the candidate models, the area metrics of each common factor in equation (7) are also shown for reference.

4. A Numerical example

In this section, a numerical example is discussed with the proposed FA method. Furthermore, the PIT area metric and the proposed FA method are compared based on the example to check the effectiveness of the latter.

The experimental observations in this section are generated using the following response:

$$y_i^c = (x_i + \theta_1 x_2)^{1.5} + \varepsilon_i$$  
$$y_i^c = 2x_i + \theta_2 x_2 + \varepsilon_i$$  
$$y_i^c = \theta_3 x_i - 2x_2 + \varepsilon_i$$  
$$y_i^c = \theta_4 x_1 x_2 + \varepsilon_i$$  (9)

where $\theta_i$ ($i=1,2,3,4$) are model parameters with $\theta_1 = \theta_3 = \theta_4 = 1$, $\theta_2 = 3$. The measurement error vector $\varepsilon_i$ ($i=1,2,3,4$) all follow a zero mean Gaussian distribution $N(0,1)$. $x=(x_1,x_2)$ are uncontrollable input variables which follow uniform distribution, with $x_1 \sim U(0,1)$ and $x_2 \sim U(1,2)$. Three candidate computational models are given as:

$$\begin{align*}
\text{model 1:} & \\
& y_1^m = (x_1 + \theta_1 x_2)^{1.5} \quad (\theta_1 = 1) \\
& y_2^m = 2x_1 + \theta_2 x_2 \quad (\theta_2 = 3) \\
& y_3^m = \theta_3 x_1 - 2x_2 \quad (\theta_3 = 1) \\
& y_4^m = \theta_4 x_1 x_2 \quad (\theta_4 = 1)
\end{align*}$$  (10)
Figure 4. Area metric based on the FA method for three models. (a) Area metric of model 1, (b) area metric of model 2, and (c) area metric of model 3.
To make a comparison between the PIT area metric and the proposed FA method, both validation results are estimated here. There are 500 observations and \(10^4\) samples of candidate models. Table 1 shows the results of the PIT area metric.

**Table 1.** The validation results of the PIT area metric.

| Models  | Model 1 | Model 2 | Model 3 |
|---------|---------|---------|---------|
| Validation results | 0.0513  | 0.0318  | 0.0848  |

According to the data in table 1, the PIT area metric does not show satisfying results of the candidate models. Model 1 is supposed to be the best model to describe the observations, followed by model 2, and model 3 is regarded as the worst one by judging according to model descriptions. Nevertheless, the data in table 1 depict a completely opposite conclusion with a judgment that model 2 is the best model, while model 1 behaves worse than model 2. It can be explained that the joint CDF of high dimension responses is hard to estimate correctly. When the proposed method is applied, different validation results appear as shown in figure 4 and table 2.

**Table 2.** Aggregate area metric of three models.

| models  | Model 1 | Model 2 | Model 3 |
|---------|---------|---------|---------|
| Aggregate area metric \(d(C,S)\) | 14.9634 | 30.7519 | 42.924  |

According to figure 4 and table 2, some conclusions can be drawn. Firstly, 2 common factors can describe the total variation of the initial multiple responses with the FA method, which signifies that the initial 4-dimensional problem can be simplified into a 2-dimensional one. Secondly, in terms of every single factor, values of the area metric based on the FA methods of the three candidate models depict a satisfying validation results, that is, model 1 is the best model to describe the observations, followed by model 2, and model 3 is the worst one. Thirdly, from the standpoint of the aggregative area metric, the same conclusion can be drawn as a single factor is considered. Generally, the comparison shows that the PIT area metric is inapplicable of differentiating models with high-dimensional multiple responses, while in these cases the proposed method is rational and effective.

5. **An Engineering example: an infectious disease model validation**

The transformation mechanism of an infectious disease has become a study hotspot in decades as it is vital to human’s daily life. Generally, there are plenty of factors which affect the transformation, for instance, the number of the patients, the number of the susceptible people, the infectious rate, and the cure rate and so on. In addition, the immigratory people and out-migration ones, latent period and the geographical environment should be considered. It is impossible to model the infectious disease if all the factors are taken into account. The differential equation is a prior choice as it can construct an appropriate model with principal factors reserved and subordinate factors ignored according to some reasonable assumptions. The assumptions can be described as follows:

1) The total population of the study area is a constant value, without regard to the immigratory people and out-migration ones, or the birth and natural death during the study period. 2) The transformation mode is contact infection, thus people won’t be infected unless they get in touch with patients. 3) People will turn into the latent ones after infected and during which period they are noncommunicable. 4) The patients will be isolated once they are discovered and they are also noncommunicable after isolation. 5) The recovery ones are immune and noncommunicable.

On the basis of the assumptions, people in the study area are classified into five sorts.

1) The Susceptible crowd. They have not been infected and do not have immunity. They may be infected as they are in general surroundings.

2) The Latent crowd. They have been infected and are in the latent period. The Susceptible ones cannot be infected by them.

3) The Infective but undiscovered crowd. They have become patients and have not been discovered, hence the Susceptible ones can be infected by them.
4) The Infective and discovered crowd. They have become patients and have been isolated, thereby the Susceptible ones cannot be infected by them.
5) The Recovery or death crowd. They are immune and have no influence in others.

Figure 5 depicts the sketch map of the transformation mechanism of an infectious disease.

\[
\begin{align*}
S(t) & \xrightarrow{\sigma I_s S} E(t) \\
& \xrightarrow{gE} I_a(t) \\
& \xrightarrow{zI_a} I_d(t) \\
& \xrightarrow{cI_d} R(t)
\end{align*}
\]

**Figure 5.** The sketch map of the transformation mechanism of an infectious disease.

In accordance to figure 5, \( \sigma \) is the ratio which describes the susceptible people infected by the patients in a day. \( g \) is the daily morbidity of the latent crowd, and \( \mu \) is the daily self-recovery rate of the latent ones. \( z \) is the daily isolation rate of patients, and \( c \) is the immunization rate. \( S(t) \) is the susceptible crowd, \( E(t) \) represents the Latent crowd, \( I_a(t) \) and \( I_d(t) \) separately describe the Infective but undiscovered crowd and the Infective and discovered crowd, and \( R(t) \) denotes the Recovery or death crowd.

A model of differential equations about the infectious disease is constructed based on its transformation mechanism, as expressed in equation (13).

\[
\begin{align*}
\frac{dS}{dt} &= -\sigma I_s S \\
\frac{dE}{dt} &= \sigma I_s S - gE - \mu E \\
\frac{dI_a}{dt} &= gE - zI_a \\
\frac{dI_d}{dt} &= zI_a - cI_d \\
\frac{dR}{dt} &= cI_d + \mu E
\end{align*}
\]

where \( S + E + I_a + I_d + R = N_r \), and the values of all the crowds are nonnegative integers, moreover we have \( 0 \leq g, z, c \leq 1 \).

\( \sigma \) can be estimated as a product of daily contact rate \( \lambda \) and the infection rate \( q \), that is \( \sigma = \lambda q \). A city with several millions population is considered in this work, and based on that a person gets in touch with about several hundred persons. According to the definition of probability, \( q \) is in a range of \( [0, 1] \), therefore \( \sigma \) should be in the order of magnitudes \( 10^{-3} \). The latent period is regarded as a uniform distribution in \( [a, b] \), which means that people infected may turn into patients or become self-healing in any other day during the period, and thus \( g = \frac{1}{b-a} \) can be obtained. \( \mu \) can be estimated once \( g \) and the corresponding latent days are decided. On the basis of the historical data released by hygiene departments, the latent days \( a \) is assumed as 2 and \( b \) is 12 in this work. The distributions of the input variables of the infectious disease model are listed in table 3.
Table 3. The distributions of the variables of the infectious disease model.

| variables                              | Types of Variables | Distribution style | Distribution parameters                  |
|----------------------------------------|--------------------|--------------------|------------------------------------------|
| Infectious rate $\sigma$              | Model parameter    | Normal             | $\sigma \sim N(3.82 \times 10^{-5},(3 \times 10^{-4})^2)$ |
| daily morbidity $g$                    | Model parameter    | Normal             | $g \sim N(10^{-3},(10^{-4})^2)$          |
| daily self-recovery rate $\mu$         | Model parameter    | Normal             | $\mu = 0.1 - g$                           |
| immunization rate $c$                  | Model parameter    | Normal             | $c \sim N(0.014,(0.001)^2)$              |
| daily isolation rate $z$               | Model parameter    | Uniform            | $z \sim U(0.4,0.8)$                      |
| Initial value of the Latent crowd $E(0)$ | Input              | Uniform            | $E(0) \sim U(0,100)$                     |
| Initial value of the Infective but undiscovered crowd $I_u(0)$ | Input              | Uniform            | $I_u(0) \sim U(1,40)$                     |
| Initial value of the Infective and discovered crowd $I_d(0)$ | Input              | Uniform            | $I_d(0) \sim U(1,20)$                     |
| Initial value of the susceptible crowd $S(0)$ | Input              | Uniform            | $S(0) = N_p - E(0) - I_u(0) - I_d(0)$    |

In addition, the initial value of the Recovery or death crowd is assumed as 0, that is, $R(0) = 0$.

City A with a total population of 6.7 million is taken as an example, namely $N_p = 6.7 \times 10^8$. The two candidate models are constructed as follows, candidate model 1 is supposed exactly the model as the one for the experimental data source. Candidate model 2 is a model with two model parameters changed, that is, daily morbidity $g \sim N(10^{-4},(10^{-4})^2)$, and correspondingly $\mu = 0.1 - g$.

Actually, not all the 5 responses appeal to the researchers, compared with the latent crowd and the infective and undiscovered crowd, the rest responses have more practical meaning and thereby they are regarded as model responses here. To investigate the spreading of the infectious disease, the responses at different dates in the first three months since the outbreak of the infectious disease need to be studied, hence the semimonthly data of responses are considered. Briefly, $t = (15,30,45,60,75,90)$ are the various validation sites to be studied. According to the idea mentioned in this paper, the problem is an 18-dimensional problem, apparently, it is impossible to compare the joint distribution of the computational and experimental responses and the FA method shows its magnitude facility.

The validation results of two models are shown in figure 6. There are 100 observations and $10^4$ samples of candidate models.

According to figure 6, model 1 shows a better accordance with the observations compared with model 2 in both single factor and the aggregative one. Furthermore, the huge difference in the area metrics’ values of two models manifests the changed model parameters — the daily morbidity $g$ and daily self-recovery rate $\mu$ have great influence in the responses, which gives the scientists a guideline to control or study the spreading of the infectious disease.
Figure 6. Area metric based on the FA method for two models (a) Area metric of model 1, and (b) area metric of model 2.

6. Conclusions
A novel method based on a combination of factor analysis and the area metric is proposed for model validation with multiple correlated responses. To take the correlations among different responses at various validation sites into consideration, every single response at each validation site is regarded as a one-dimensional response, thus model with multiple responses at a single validation site or various sites can be unified into multiple responses. FA is a popular method in the field of multivariate statistical analysis which explains correlated variables with fewer factors. As a result, validation of models with multiple correlated high dimensional responses can be decomposed into a series of model validations in independent one-dimensional space, which then can be easily estimated by the existing area metric method.

The results of several examples demonstrate the rationality of the proposed method. Furthermore, with a comparison with the PIT method, the proposed method shows its advantages. It does not require estimating the joint CDF of the multiple responses which is an essential step in the PIT method, and the proposed method is lower computational cost and less time consuming compared with the PIT method. In addition, with FA, the proposed method also avoids directly comparing the joint distribution of the computational and experimental responses, which makes it especially suitable for validation of models with high dimensional responses.
References

[1] Abdesssemed M, Kenai S and Bali A 2015 *CONSTRUCTION AND BUILDING MATERIALS* **94** 547-554

[2] Park J J 2016 *INTERNATIONAL JOURNAL OF HEAT AND MASS TRANSFER* **93** 491-499

[3] He W, Duan Y L, Deng L W and Zhou W G 2016 *JOURNAL OF PERFORMANCE OF CONSTRUCTED FACILITIES* **30 (1)** C4015003

[4] Oberkampf W L, Sindir M and Conlisk A 1998 Guide for the verification and validation of computational fluid dynamics simulations *American Institute of Aeronautics and Astronautics*

[5] Oberkampf W L, Trucano T G and Hirsch C 2004 *Applied Mechanics Reviews* **57** 345-384

[6] Liu Y, Chen W, Arendt P and Huang H-Z 2011 *Journal of Heat Transfer-Transactions of the ASME* **133** 071005

[7] Sornette D, Davis A B, Ide K, Vixie K R, Pisarenko V and Kamm J R 2007 *Theory and Applications. Proc. Natl. Acad. Sci. U.S.A* **104(16)** 6562-6567

[8] Buranathiti T, Cao J, Chen W, Bagdasaryan L and Xia Z C 2006 *ASME J. Manuf. Sci. Eng.* **128(2)** 588–597

[9] Rebbas R and Mahadevan S 2006 *AIJA J.* **44(10)** 2376-2384

[10] Mahadevan S and Rebbas R 2005 *Reliab. Eng. Syst. Saf.* **87(1)** 223-232

[11] Oberkampf W L and Barone M F 2006 *J. Comput. Phys.* **217(1)** 5-36

[12] Oberkampf W L and Trucan T G 2002 *Prog. Aerosp. Sci.* **38(2)** 209-272

[13] Ferson S, Oberkampf W L and Ginzburg L 2008 *Computer Methods in Applied Mechanics and Engineering* **197** 2408-2430

[14] Roy C J and Oberkampf W L 2011 *Computer Methods in Applied Mechanics and Engineering* **200** 2131-2144

[15] Li W, Chen W, Jiang Z, Lu Z Z and Liu Y 2014 *Reliability Engineering & System Safety* **127** 1-11

[16] Rebbas R and Mahadevan S 2006 *Reliability Engineering & System Safety* **91** 861-871

[17] Jiang X and Mahadevan S 2008 *Journal of Applied Statistics* **35** 49-65

[18] Wang H and Kuo C Y 2005 *Computers and Mathematics with Applications* **48** 1765-1778

[19] Pal N R 1998 Soft computing for feature analysis *Fuzzy Sets and Systems* **103** 201-221

[20] Hsia T C, Hsu Y L and Jen H L 2009 *Expert Systems with Applications* **36** 2719-2726

[21] Anish C M, Majhi B and Tonde H S *In Proc. of international conference on communication and computing, ICC-2014-Bangalore* 18-25

[22] Kim H, Soibelman L and Grobler F 2008 Factor selection for delay analysis using knowledge discovery in databases *Automation in Construction* **17** 550-560

[23] Martinez-Murcia F J, Gorriz J M, Ramirez J, Puntone C G and Salas-Gonzalez D 2012 Computer aided diagnosis tool for Alzheimer’s disease based on Mann–Whitney–Wilcoxon U-test for the Alzheimer’s Disease Neuroimaging Initiative *Expert Systems with Applications* **39** 9676-9685

[24] Harman H H 1976 Modern factor analysis (3 rded) (Chicago: University of Chicago Press)

[25] Joreskog K G and Lawley D N 1968 New methods in maximum likelihood factor analysis *British Journal of Mathematical and Statistical Psychology* **21** 85-96

[26] Fabrigar L R, Wegener D T, MacCallum R C and Strahan E J 1999 Evaluating the use of exploratory factor analysis in psychological research *Psychological Methods* **4(3)** 272-299

[27] Thurstone L L 1947 Multiple factor analysis (Chicago: University of Chicago Press)

[28] Kaiser H F 1958 The varimax criterion for analytic rotation in factor analysis *Psychometrika* **23** 187-200

[29] Kennedy M C and O’Hagan A 2001 Bayesian calibration of computer models *Journal of the Royal Statistical Society Series B-Statistical Methodology* **63(3)** 425-464

[30] Wu D Q, Lu Z Z, Wang Y P and Cheng L 2015 Model validation and calibration based on component functions of model output *Reliability Engineering and System Safety* **140** 59-70