Outlier Detection Using the Outlier Probability for Robust Linear Regression

Ka-Veng Yuen[1,2], He-Qing Mu[1]
[1] University of Macau, Macao, China
[2] Corresponding author: Ka-Veng Yuen, e-mail: kvyuen@umac.mo

Abstract

Outlier detection is an important problem in statistics. In this paper, we introduce a novel concept of outlier probability for outlier detection and robust linear regression. First, the Mahalanobis distance is utilized to identify the leverage points. By excluding the leverage points, the maximum trimmed likelihood estimation will be used to obtain an initial set of regular data points while the remaining and the leverage points are included in the initial suspicious data set. Then, each suspicious data point and the combinations are evaluated by a novel outlier probability that depends not only on the residuals but also the size of the data set. Incorporating the data size is important as it controls the probability of the existence of a data point (or a batch or data points) exceeding a given value of the normalized residual. This outlier probability is robust because it incorporates also the posterior uncertainty quantified using the Bayesian approach. Then, the data points with outlier probability below 0.5 will be transferred to the set of regular data points. Iteration is continued until all suspicious data points are associated with outlier probability over 0.5. Finally, robust regression can be conducted by considering only the set of regular points. Therefore, it is expected that the parametric identification results are robust to outliers. In contrast to other existing outlier detection criteria that require some subjective threshold (e.g., normalized residual larger than 2.5), the outlier probability criterion of 0.5 is objective. Finally, a challenging application will be presented and comparison to other well-known methods will be given.

Keywords: Bayesian inference, leverage, inverse problem, outlier, robust analysis, system identification

1. Introduction

Outlier is a major concern in statistics, science and engineering. In practice, it is commonly encountered that some data points deviate substantially from the model output. Presence of outliers indicates irregularities of data and/or weakness of the model/theory (Beckman 1997). On one hand, outliers may be due to extraordinarily large measurement error, e.g., human error, instrument failure or unknown environmental disturbances. On the other hand, they may also be due to imperfection of theory, i.e., outliers are observations generated by unmodeled mechanism (Hawkins 1980). Due to the fact that ordinary least squares (OLS) method is highly sensitive to outliers, a number of robust estimators have been developed (Huber 1981; Rousseeuw 1984). In addition, a class of outlier detection methods is based on Bayesian inference (Box and Tiao 1968; Chaloner and Brant 1988), which is a method widely applied to various research disciplines (Beck and Katafygiotis 1998; Yuen 2010; Yuen and Kuok 2011). In Yuen and Mu (2012), the concept of outlier probability was introduced by considering the probability to obtain a data point of its residual level, given the size of the data set. In this paper, this concept will be enhanced to evaluate the data points not only as individuals but also as a set. This outlier probability will be used for outlier detection robust parametric identification in linear regression problems.

2. Bayesian inference for linear regression

Consider a linear regression model with N data points:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (1)$$

where $\boldsymbol{Y} \in \mathbb{R}^N$ is the output vector; $\boldsymbol{X} \in \mathbb{R}^{N \times N_{\boldsymbol{\beta}}}$ is the design/input matrix; $\boldsymbol{\beta} \in \mathbb{R}^{N_{\boldsymbol{\beta}}}$ is regression parameter vector; and $\boldsymbol{\varepsilon} \in \mathbb{R}^N$ is the residual vector following normal distribution $\mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_N)$. The likelihood function is:

$$p(D|\boldsymbol{\beta}, \sigma^2) = (2\pi\sigma^2)^{-N/2} \exp[-(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^T (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})/(2\sigma^2)] \quad (2)$$

The prior probability density function (PDF) of the parameters is considered as:

$$p(\boldsymbol{\beta}, \sigma^2) = p(\boldsymbol{\beta}) p(\sigma^2) \quad (3)$$

where $p(\boldsymbol{\beta})$ is the independent uniform distribution with sufficiently wide range to cover important regions of the likelihood function and $p(\sigma^2)$ is the inverse Gamma distribution $IG(a_0, b_0)$ with shape parameter a_0 and scaling parameter b_0.

By the Bayes' theorem, the posterior probability density function (PDF) is given by:

$$p(\boldsymbol{\beta}, \sigma^2|D) = p(D)^{-1} p(\boldsymbol{\beta}, \sigma^2) p(D|\boldsymbol{\beta}, \sigma^2) \quad (4)$$

where $p(D)^{-1}$ is the normalizing constant. Referring to Yuen and Mu (2011), the optimal value of the regression parameter vector and the prediction-error variance are:

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{Y};\ \hat{\sigma}^2 = [2b_0 + (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^T (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})]/[2(a_0 + 1) + N] \quad (5)$$

The conditional posterior PDF $p(\boldsymbol{\beta}|D, \sigma^2)$ follows the normal distribution $\mathcal{N}\left(\widehat{\boldsymbol{\beta}}, \sigma^2 (\boldsymbol{X}^T \boldsymbol{X})^{-1}\right)$. The marginal posterior PDF $p(\sigma^2|D)$ follows $IG\left(\hat{a}, \hat{b}\right)$, where the updated shape $\hat{a}$ and scaling parameter $\hat{b}$ are given by:

$$\hat{a} = (N - N_{\boldsymbol{\beta}})/2 + a_0;\ \hat{b} = (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^T (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})/2 + b_0 \quad (6)$$

3. Probabilistic robust outlier detection

3.1 Initial regular dataset

Since parametric identification results may be drastically affected by outliers and the probability model of outliers is unavailable in practice, the proposed approach attempts to identify the uncertain parameters by excluding the outliers. It starts from selecting a reliable subset (referred to as the initial regular dataset D^R) of the entire dataset.

Since parametric identification is highly sensitive to leverage points, the initial regular dataset should exclude all leverage points. The set of leverage points L can be obtained by utilizing the Mahalanobis distance. After L is determined, the dataset $D_{-L} = D \backslash L$ will be used to obtain the maximum trimmed likelihood estimation (MTLE). Furthermore, the number of data points in D_{-L} is denoted by N_{-L}.

Use $D^H = \{D_i = (\boldsymbol{X_i}, y_i), D_i \in D_{-L}, i \in H\}$ to denote a sub-dataset with h data points from D_{-L}, where H is an index set with h distinct elements from the set $\{1,2,\dots,N_{-L}\}$. Furthermore, $\boldsymbol{X}^H$ is the design/input submatrix to take the corresponding rows of $\boldsymbol{X}$ and $\boldsymbol{Y}^H$ takes the corresponding model outputs in $\boldsymbol{Y}$. In contrast to Eq. (2), the trimmed likelihood function can be defined as (Hadi and Luceno 1999):

$$p(D^H|\boldsymbol{\beta}, \sigma^2) = (2\pi\sigma^2)^{-h/2} \exp[-(\boldsymbol{Y}^H - \boldsymbol{X}^H\boldsymbol{\beta})^T (\boldsymbol{Y}^H - \boldsymbol{X}^H\boldsymbol{\beta})/(2\sigma^2)] \quad (7)$$

For a given value of h, the MTLE $\boldsymbol{\beta}_{MTLE}$ and the MTLE sub-dataset D^H_{MTLE} can be obtained by:

$$\{\boldsymbol{\beta}_{MTLE}, D^H_{MTLE}\} = \text{argmax}_{\boldsymbol{\beta}, \sigma^2, D^H} p(D^H|\boldsymbol{\beta}, \sigma^2) \quad (8)$$

where $\boldsymbol{\beta}_{MTLE}$ and D^H_{MTLE} can be efficiently calculated by the algorithm in (Rousseeuw and Van Driessen 2006). Here, it is suggested to use a conservatively small value of h (e.g., 70% of the total number of data points N) so that any suspicious data points will be excluded from the MTLE sub-dataset D^H_{MTLE}. Then, the initial regular dataset is given by $D^R = D^H_{MTLE}$, and the initial suspicious dataset is given by $D^S = D \backslash D^R$.

3.2 Probability of outlier

Since a conservatively small value of h is chosen, it is expected that some regular

data points are included in the initial suspicious dataset. Therefore, in order to enhance the efficiency of the identification, each suspicious data point will be re-evaluated through its probability of outlier. For a data point with outlier probability below 0.5, it will be reclassified as a regular data point. Otherwise, it will remain in the suspicious dataset. Consider a suspicious data point $D_k^S = (\boldsymbol{X}_k^S, y_k^S) \in D^S$ and its residual $\varepsilon_k^S = y_k^S - \boldsymbol{X}_k^S\boldsymbol{\beta}$. Since the probability model for the regular measurement noise is normal, the probability of a data point falling outside the interval $\left(-\left|\varepsilon_k^S\right|, \left|\varepsilon_k^S\right|\right)$ is:

$$q = 2\Phi\left(-\left|\varepsilon_k^S\right|/\sigma\right) \tag{9}$$

where $\Phi(\cdot)$ is the cumulative distribution function (CDF) of the standard normal random variable. Let η_k^S denote the number of regular data points with absolute normalized residual larger than or equal to $\left|\varepsilon_k^S\right|$. Therefore, the conditional probability of outlier for D_k^S can be defined as the probability that not more than η_k^S samples fall outside the interval $\left(-\left|\varepsilon_k^S\right|, \left|\varepsilon_k^S\right|\right)$ among the N standard normal samples:

$$P_o\left(D_k^S|\varepsilon_k^S, \sigma^2\right) = \sum_{\tau=0}^{\eta_k^S} C(N,\tau)\, q^\tau (1-q)^{N-\tau} \tag{10}$$

where $C(N,\tau) = N!/((N-\tau)!\,\tau!)$ is a binomial coefficient. Note that in Yuen and Mu (2012) only the term with $\tau = 0$ was considered. The additional terms included in Eq. (10) are used to account also for the combinations of different data points. For example, given a particular size of the data set and a particular largest absolute normalized residual, the outlier probability is 0.4. However, with the same size of data set and the same absolute normalized residual, the outlier probability should be larger than 0.4 when there are more than one data point with this residual.

Considering the uncertainty in ε_k^S and σ^2, the probability of outlier for a suspicious data point D_k^S is readily obtained:

$$P_o\left(D_k^S|D^R\right) = \int_0^\infty \int_{-\infty}^\infty P_o\left(D_k^S|\varepsilon_k^S, \sigma^2\right) p\left(\varepsilon_k^S, \sigma^2|D^R\right) d\boldsymbol{\beta}\, d\sigma^2 \tag{11}$$

Since there is no closed-form solution for this integral, a Monte Carlo simulation (MCS) procedure is proposed to compute the probability of outlier. To simulate samples of $\varepsilon_k^{S(j)}, \sigma^{2(j)}$, one can utilize the following relationship:

$$p\left(\varepsilon_k^S, \sigma^2|D^R\right) = p\left(\varepsilon_k^S|, \sigma^2, D^R\right) p(\sigma^2|D^R) \tag{12}$$

Given $\varepsilon_k^S = y_k^S - \boldsymbol{X}_k^S\boldsymbol{\beta}$, the conditional PDF of the residual $p\left(\varepsilon_k^S|, \sigma^2, D^R\right)$ is normal:

$$p\left(\varepsilon_k^S|\sigma^2, D^R\right) = \mathcal{N}\left(\hat{\varepsilon}_k^S, Q_k^S\right) \tag{13}$$

where the mean $\hat{\varepsilon}_k^S$ and the variance Q_k^S are given by:

$$\hat{\varepsilon}_k^S = y_k^S - \boldsymbol{X}_k^S\widehat{\boldsymbol{\beta}}^{\boldsymbol{R}};\ Q_k^S = \sigma^2\boldsymbol{X}_k^S[(\boldsymbol{X}^R)^T\boldsymbol{X}^R]^{-1}(\boldsymbol{X}_k^S)^T \tag{14}$$

Recall that the marginal posterior PDF $p(\sigma^2|D^R)$ follows the inverse Gamma distribution $IG\left(\hat{a}, \hat{b}\right)$. Because some of regular data points with relatively large absolute residuals are not in D^R, σ^2 is underestimated. To resolve this bias problem, σ^2 is simulated according to the modified posterior PDF $IG\left(\tilde{a}, \tilde{b}\right)$ with the modified shape parameter $\tilde{a}$ and modified scaling parameter $\tilde{b}$ as

$$\tilde{a} = (N - N_{\boldsymbol{\beta}})/2 + a_0;\ \tilde{b} = (\bar{a}+1)\tilde{\sigma}^2 \tag{15}$$

where $\tilde{\sigma}$ is the consistent estimator for the standard deviation (Hampel 1974):

$$\tilde{\sigma} = 1.4826 \times \mathrm{med}_k\ (|\hat{\varepsilon}_k^R|) \tag{16}$$

where $\mathrm{med}(\cdot)$ is the sample median; $\hat{\varepsilon}_k^R$ is the residual of the *k*-th regular data points in D^R; 1.4826 is the corresponding consistent factor for the sample median. Finally, the outlier probability in Eq. (11) can be evaluated by MCS with the Gibbs sampler:

$$P_o\left(D_k^S|D^R\right) = E_{p(\boldsymbol{\beta},\sigma^2|D^R)} P_o\left(D_k^S|\varepsilon_k^S, \sigma^2\right) = \frac{1}{N_s}\sum_{j=1}^{N_s} P_o\left(D_k^S|\varepsilon_k^{S(j)}, \sigma^{2(j)}\right) \tag{17}$$

where N_s is the number of samples to be simulated; $\sigma^{2(j)}$ follows $IG\left(\tilde{a}, \tilde{b}\right)$ in Eq. (15); and $\varepsilon_k^{S(j)}$ follows $p\left(\varepsilon_k^S|\sigma^2, D^R\right)$ in Eq. (13).

3.3 Summary of the proposed method

1. (a) Identify the set of leverage points L.
   (b) Use $D\backslash L$ to determine the initial regular dataset, i.e. $D^R = D^H_{MTLE}$.
   (c) Obtain the initial suspicious dataset $D^S = D\backslash D^R$.
2. Based on D^R, calculate $\widehat{\boldsymbol{\beta}}^R$ by using Eq. (5).
3. Compute the residuals of the data points in D^S and D^R. They are denoted as $\boldsymbol{\varepsilon}^S$ and $\boldsymbol{\varepsilon}^R$, respectively. Rearrange $D^S = \{D^S_k : k = 1, \dots, K\}$ such that $|\varepsilon^S_1| > |\varepsilon^S_2| > \cdots > |\varepsilon^S_K|$, where K is the number of points in D^S.
4. Count the number of regular points with absolute normalized residuals larger than $|\varepsilon^S_k|$.
5. Compute $P_o(D^S_k|D^R)$ by using Eq. (17).
6. If $P_o(D^S_k|D^R) \geq 0.5$, move this point to D^R. Otherwise, keep it in D^S.
7. Update D^R and D^S, repeat Step 2 to 6 until $P_o(D^S_k|D^R) \geq 0.5$ is satisfied for all suspicious data points in D^S. The final updated parameters and their posterior PDFs are based on the final regular dataset D^R_f. All data points in the final suspicious dataset D^S_f are considered possible outliers.

4. Illustrative examples

In this section, the proposed method is compared with the ordinary least-squares (OLS) method and two well-known robust estimators, namely the Huber estimator (HE) (Huber 1981) and the least trimmed squares (LTS) (Rousseeuw 1984). The well-known outlier criterion $\varepsilon/\sigma > 2.5$ is used for these three methods while the proposed method detects the outliers by the procedure described in Section 3.3.

Recall that there are two types of outliers. The first type of outliers occurs due to extraordinarily large measurement noise while the second type of outliers is due to imperfection of theory (modeling error). To generate the dataset with both types of outliers, two sub-datasets are generated, namely the contaminated dataset $D^{(1)}$ and the disordered dataset $D^{(2)}$. The contaminated dataset $D^{(1)}$ contains regular data points and outliers of the first type. In order to generate this set, a weighted mixture distribution of error for a data point in $D^{(1)}$ is used:

$$p(\bar{e}) = (1-\rho)\cdot \mathcal{N}(\bar{e}|0,\bar{\sigma}^2) + \rho \cdot f(\bar{e}) \tag{18}$$

where the normal distribution $\mathcal{N}(\bar{e}|0,\bar{\sigma}^2)$ is used for the regular data points, and $f(\bar{e})$ is used for the outliers of the first type. In this example, $f(\bar{e})$ is chosen as the mixture of triangular distributions:

$$f(\bar{e}) = 0.5T(\bar{e}|-5\bar{\sigma}, -4\bar{\sigma}, -3\bar{\sigma}) + 0.5T(\bar{e}|3\bar{\sigma}, 4\bar{\sigma}, 5\bar{\sigma}) \tag{19}$$

where $T(\bar{e}|l', m', r')$ is the triangular distribution with lower limit l', upper limit r' and mode m'. The variable ρ in Eq. (18) is the contaminated level parameter, which can be interpreted as the probability of a data point in $D^{(1)}$ being an outlier. It is worth noting that the value of ρ and the probability distributions in Eq. (18) and (19) are assumed unknown in the identification process and they are used only for the purpose of data generation. On the other hand, the data points in the contaminated dataset $D^{(2)}$ will be generated with substantial bias. Then, the entire dataset $D = D^{(1)} \cup D^{(2)}$ contains both types of outliers.

A seven-parameter regression model is considered with both types of outliers. The contaminated dataset $D^{(1)} = \{D^{(1)}_i, i = 1, \dots, 83\}$ follows

$$\begin{aligned} y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 \log(x_{2i}) + \beta_3 \sin(x_{3i}) + \beta_4 x_{1i} \cdot \log(x_{2i}) + \\ + \beta_5 x_{1i} \cdot \sin(x_{3i}) + \beta_6 \log(x_{2i}) \cdot \sin(x_{3i}) + \bar{e}_i \end{aligned} \tag{20}$$

where the error $\bar{e}_i$ follows the weighted mixture distribution in Eq. (18) with $\rho = 0.1$; the prediction-error level is 40%; the actual values of the parameters are $\beta_0 = 10$, $\beta_1 = -1$, $\beta_2 = 1$, $\beta_3 = -1$, $\beta_4 = 1$, $\beta_5 = -1$, $\beta_6 = 1$, $\beta_7 = -1$; and all the input variables follow uniform distribution: $x_{1i} \sim U(-5,5)$, $x_{2i} \sim U(1,30)$, $x_{3i} \sim U(0,4\pi)$.

The disordered dataset $D^{(2)} = \{D_i^{(2)}, i = 84{,}85 \ldots{,}100\}$ is generated from:

$$(x_{1i}, x_{2i}, x_{3i}, y_i)^T \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \text{ for } i = 84{,}85 \ldots{,}100 \quad (21)$$

where the mean vector is $\boldsymbol{\mu} = [30{,}100{,}10\pi, -100]^{\boldsymbol{T}}$ and the covariance matrix is $\boldsymbol{\Sigma} = 15\boldsymbol{I}_4$. The data points in $D^{(2)}$ dramatically deviate from the model output described by Eq. (20). This type of outliers is more critical than the first type because they are subjected to substantial bias.

To compare the performance of the proposed method with the three aforementioned methods, five hundred independent sets of data are generated. Two indicators are used to quantify the performance of outlier detection: (a) masking denotes the case that an outlier is not being detected; and (b) swamping denotes the case that a regular data point is mistakenly identified as an outlier. The masking percentage is the average of the masking number divided by the actual number of outliers while the swamping percentage is the average of the swamping number divided by the actual number of regular data points. It is desirable to obtain low values of both indicators. Table 1 shows the outlier detection results by different methods. Only the proposed method achieves simultaneously low values of the masking percentage (1.5%) and swamping percentage (2.1%). This confirms the capability of the proposed method for outlier detection in the presence of large-error leverage points.

Table 2 shows the parametric identification statistics of the five hundred simulation runs using different methods. The 2nd, 4th, 6th and 8th columns show the sample average of the corresponding parameters identified by different methods. The 3rd, 5th, 7th and 9th columns show the corresponding root-mean-square errors between the actual and identified values of the uncertain parameters. The proposed method provides the best performance in the sense that its average parameters are the closest to the actual value and the corresponding root-mean-square errors are the smallest.

Table 1. Outlier detection results of different methods

Method	Masking %	Swamping %
OLS	96.7%	0.1%
HE	90.8%	1.0%
LTS	25.6%	0.0%
Proposed method	1.5%	2.1%

5. Conclusion

A novel probabilistic method was proposed for outlier detection and robust parametric identification. First, not only the optimal values of the regression parameters and residuals but also the associated uncertainties are taken into account for outlier detection. Second, the size of the dataset is incorporated because it is one of the key factors that determine the probability to obtain a data point with large residual. Third, the proposed method requires no information on the outlier distribution model. Fourth, instead of definite determination of the outlierness of a data point, the proposed approach provides the probability of outlier. The proposed method was confirmed, through the example, to be capable for robust parametric identification and outlier detection for linear regression problem.

Table 2. Parametric identification results of different methods

Method	β_0	β_0^{RMS}	β_1	β_1^{RMS}	β_2	β_2^{RMS}	β_3	β_3^{RMS}
Actual	10	-	-1	-	1	-	-1	-
OLS	14.66	7.04	6.05	7.12	-1.04	2.81	-3.22	7.47
HE	14.86	7.30	6.47	7.55	-1.03	2.88	-2.56	7.17
LTS	9.93	1.53	-1.03	0.55	1.03	0.57	-0.95	2.10
Proposed method	9.97	1.34	-1.07	0.53	1.01	0.49	-1.01	2.02
Method	β_4	β_4^{RMS}	β_5	β_5^{RMS}	β_6	β_6^{RMS}	$\bar{\sigma}$	$\bar{\sigma}^{RMS}$
Actual	1	-	-1	-	1	-	2.20	-
OLS	-2.09	3.10	-0.23	0.80	1.92	2.80	9.21	7.07
HE	-2.18	3.19	-0.22	0.81	1.69	2.66	9.07	6.92
LTS	1.01	0.20	-0.99	0.19	0.98	0.80	4.28	2.24
Proposed method	1.02	0.19	-1.00	0.17	1.01	0.75	2.05	0.38

Acknowledgment

The generous support by the Statistics and Census Service (DSEC) of the Macao SAR government is gratefully acknowledged.

References

Beck, J. L. and Katafygiotis, L.S. (1998) "Updating Models and Their Uncertainties. I: Bayesian Statistical Framework," *Journal of Engineering Mechanics*, 124, 455-461.

Beckman, R.J. and Cook, R.D. (1983) "Outlier.....s," *Technometrics*, 25, 119-149.

Box, G.E.P. and Tiao, G. C. (1968) "A Bayesian Approach to Some Outlier Problems," *Biometrika*, 55, 119-129.

Chaloner, K. and Brant, R. (1988) "A Bayesian Approach to Outlier Detection and Residual Analysis," *Biometrika*, 75, 651-659.

Hadi, A.S. and Luceno, A. (1997) "Maximum Trimmed Likelihood Estimators: a Unified Approach, Examples and Algorithms," *Computational Statistics & Data Analysis*, 25, 251–72.

Hampel, F.R. (1974). "The Influence Curve and its Role in Robust Estimation," *Journal of the American Statistical Association, 69*, 383-393.

Hawkins, D.M. (1980) *Identification of Outliers*, Chapman and Hall, London.

Huber, P. J. (1981) *Robust Statistics*, John Wiley & Sons, Inc, New York.

Rousseeuw, P.J. (1984) "Least Median of Squares Regression," *Journal of the American Statistical Association*, 79, 871-880.

Rousseeuw, P. and van Driessen, K. (2006) "Computing LTS Regression for Large Data Sets," *Data Mining and Knowledge Discovery*, 12, 29-45.

Yuen, K.V. (2010) *Bayesian Methods for Structural Dynamics and Civil Engineering*, John Wiley & Sons, New York.

Yuen, K.V. and Kuok, S.C. (2011). "Bayesian Methods for Updating Dynamic Models," *Applied Mechanics Reviews*, 64(1), 010802-1 -- 010802-18.

Yuen, K.V. and Mu, H.Q. (2011) "Peak Ground Acceleration Estimation by Linear and Nonlinear Models with Reduced Order Monte Carlo Simulation," *Computer-Aided Civil and Infrastructure Engineering*, 26, 30-47.

Yuen, K.V. and Mu, H.Q. (2012) "A Novel Probabilistic Method for Robust Parametric Identification and Outlier Detection," *Probabilistic Engineering Mechanics*, 30, 48-59.