

PAPER • OPEN ACCESS

Development of a Sparse Polynomial Chaos Expansions Method for Parameter Uncertainty Analysis

To cite this article: C X Wang *et al* 2020 *IOP Conf. Ser.: Earth Environ. Sci.* **435** 012011

View the [article online](#) for updates and enhancements.

You may also like

- [Fast and accurate sensitivity analysis of IMPT treatment plans using Polynomial Chaos Expansion](#)

Zoltán Perkó, Sebastian R van der Voort, Steven van de Water *et al.*

- [Uncertainty Quantification and Global Sensitivity Analysis of Batteries: Application to a Lead-Acid Battery](#)

Hojat Dehghandorost, Vahid Esfahanian and Farzin Chaychizadeh

- [Polynomial chaos expansion for uncertainty analysis and global sensitivity analysis](#)

Ming Chen, Xinhua Zhang, Kechun Shen *et al.*





The
Electrochemical
Society

Advancing solid state &
electrochemical science & technology

DISCOVER
how sustainability
intersects with
electrochemistry & solid
state science research

Development of a Sparse Polynomial Chaos Expansions Method for Parameter Uncertainty Analysis

C X Wang^{1*}, J Liu², Y P Li^{2,3,4}, J Zhao¹ and X M Kong¹

¹College of Fundamental Research, Beijing Polytechnic, Beijing 100176, China

²School of Environmental Science and Engineering, Xiamen University of Technology, Xiamen 361024, China

³School of Environmental, Beijing Normal University, Beijing 100875, China

⁴Research Scientist, Environmental Systems Engineering Program, Faculty of Engineering and Applied Science, University of Regina, Regina, Sask. S4S 0A2

*Corresponding Author E-mail: chunxiao20083366@163.com

Abstract. Incorporating uncertainty assessment into hydrological simulation is of vital significance for providing valuable information for conserving and restoring the ecology environment in arid and semiarid regions. In this study, a sparse polynomial chaos expansions method was developed to quantify hydrological model parameter uncertainties on model performance in Kaidu river basin, China. A four dimension two order polynomial chaos expansions model was built and the effect of four parameters were quantified based on the coefficients of the polynomial chaos expansions model. Results indicated that precipitation in summer has more significant influence on model output than that in other seasons. High Sobol sensitivity indices values (0.22 in spring, 0.17 in summer, 0.21 in autumn and 0.29) for the interaction of precipitation and maximum capacity for fast store demonstrate that they are the major factors affecting runoff generation. These results can help reveal the flow processes and provide valuable information for water resources management.

1. Introduction

With the widespread application of digital elevation models, geographical information system and remote sensing, hydrological models have been widely used to provide information for catchment management with information on the interaction of water, energy and vegetation processes distributed over space and time in a way that cannot be done through field experiment and direct observation [1][2]. Nevertheless, there are a variety of uncertainties involved in hydrological processes. Such uncertainties would influence model performance due to randomness characteristic of precipitation, temperature, infiltration and so on. Model representations of real-world hydrological systems are complicated with a variety of factors, including inadequate conceptualizations of physical processes, errors related to spatial and temporal scales and derivation of model-parameter values directly from basin traits [3]. Parameters obtained from calibration are also affected by several factors such as correlations among parameters, sensitivity or insensitivity in parameters and statistical features of model residuals. Therefore, evaluating the effect of parameter uncertainties on model output would provide more information for hydrological forecasting and related catchment management.

A number of efforts were made in developing more effective methods for reflecting parameter uncertainty in hydrological modelling and their effects on model performance [4]. Among them, the



Content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/3.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

parameter uncertainty assessment methods widely used are mainly based on Bayesian approaches. These methods can be roughly classified into formal Bayesian methods using an explicit statistical error model and a Markov Chain Monte Carlo sampling procedure [5][6] and informal Bayesian approaches, based on the Generalized Likelihood Uncertainty Estimation method [7]. Although stochastic analysis method is capable of handling uncertainties with known probability distributions, it cannot ensure a sufficient precision of the statistics inferred from the retained solutions unless the sampling of the parameter space is dense enough [8].

Therefore, a sparse polynomial chaos expansions method integrating of the semi-distributed land use based runoff processes (SLURP) model, polynomial chaos expansions, sparse grid collection and Sobol sensitivity indices, would be developed to quantify parameter uncertainties on model performance. SLURP model is capable of simulating the interaction among overland, river and groundwater flows. A polynomial chaos expansions model would be to build a surrogate model for SLURP. Sparse grid collection would be used to determine the sampling dates of different parameters. Finally, the Sobol sensitivity indices would be calculated based on the coefficient of the polynomial chaos expansions model to quantify the contribution of different parameters and theirs to the total to discover the effect of different parameters.

2. Methodology

2.1 Polynomial chaos expansions

The generalized polynomial chaos expansion model can be written in the form:

$$y = a_0 + \sum_{i_1=1}^n a_{i_1} \Gamma_1(\xi_{i_1}) + \sum_{i_1=1}^n \sum_{i_2=1}^{i_1} a_{i_1, i_2} \Gamma_2(\xi_{i_1}, \xi_{i_2}) + \sum_{i_1=1}^n \sum_{i_2=1}^{i_1} \sum_{i_3=1}^{i_2} a_{i_1, i_2, i_3} \Gamma_3(\xi_{i_1}, \xi_{i_2}, \xi_{i_3}) + \dots \quad (1)$$

Where y is the output and $\Gamma_p(\xi_{i_1}, \xi_{i_2}, \dots, \xi_{i_p})$ denotes the orthogonal Hermite polynomials of order p . In order to perform the global sensitivity of the random variables, We assume that the components of $\xi_{i_1}, \xi_{i_2}, \dots, \xi_{i_p}$ are multi-dimensional independent random variables. The polynomial of order m can be defined as [9]:

$$\Gamma_m(\xi_{i_1}, \dots, \xi_{i_n}) = (-1)^m e^{\frac{1}{2} \xi^T \xi} \frac{\partial^m}{\partial \xi_{i_1} \dots \partial \xi_{i_p}} e^{-\frac{1}{2} \xi^T \xi} \quad (2)$$

Where ξ is the vector of n random variables. Then, the second and third order Hermite polynomials with two random variables are $\{1, \xi_1, \xi_2, \xi_1^2 - 1, \xi_1 \xi_2, \xi_2^2 - 1\}$ and $\{1, \xi_1, \xi_2, \xi_1^2 - 1, \xi_1 \xi_2, \xi_2^2 - 1, \xi_1^3 - 3\xi_1, \xi_1^2 \xi_2 - \xi_2, \xi_2^2 \xi_1 - \xi_1, \xi_2^3 - 3\xi_2\}$. The series n -dimensional polynomial chaos would be truncated to a finite number of terms for real computational.

2.2 Sparse grid collection method

The quadrature method takes a tensor product of the univariate leading to an exponential dependence of the number of points of the dimension n . In order to alleviate this ‘‘Curse of dimensionality’’ sparse grid quadrature method was proposed based on tensor products of hierarchical difference sets ($\Delta_k = U_m - U_{m-1}$ with $U_0 = 0$). The Smolyak interpolation operator can be given as follows [10]:

$$A(L, N) f = \sum_{N < |k|_1 \leq L + d - 1} (-1)^{N + L - |k|_1 - 1} \binom{N-1}{|k|_1 - L} \left(\bigotimes_{i=1}^n U_{k_i} \right) f \quad (3)$$

Based on the collection points for one-dimension integral, the collection points of multi-dimension integral can be obtained as follows:

$$H(w, N) = \bigcup_{i \in Y(w, N)} \{x^{i_1} \otimes \dots \otimes x^{i_N}\} \quad (4)$$

In detail, the collection points for $N=3$ and $L=2$ are $(0, 0, 0)$, $(\pm 1, 0, 0)$, $(0, \pm 1, 0)$, $(0, 0, \pm 1)$, $(\pm 1.73, 0, 0)$, $(0, \pm 1.73, 0)$, $(0, 0, \pm 1.73)$, $(\pm 1, \pm 1, 0)$, $(0, \pm 1, \pm 1)$ and $(\pm 1, 0, \pm 1)$.

2.3 Sobol sensitivity indices

Sobol sensitivity indices describing the contribution that the set of variables have on to the total variance can be formulated as $S_u = v_u / v_T$, where v_u is the variance of variables u . v_T is the total variance. To calculate the sensitivity indices from the polynomial chaos coefficients, the truncated polynomial chaos expansion model is rewritten as follows:

$$y = M_0 + \sum_{i=1}^N M_i(x_i) + \sum_{1 \leq i < j = N} M_{i,j}(\xi_i, \xi_j) + \dots + M_{1,\dots,N}(\xi_1, \dots, \xi_N) \quad (5)$$

$$M_{i_1, \dots, i_d}(x_{i_1}, \dots, x_{i_d}) = \sum_{t \in I_{i_1, \dots, i_d}} a_t \Gamma_t(\xi_{i_1}, \dots, \xi_{i_d}) \quad (6)$$

Based on the orthogonality of different random variables :

$$E[M_{i_1, \dots, i_d}(x_{i_1}, \dots, x_{i_d}) M_{j_1, \dots, j_t}(x_{j_1}, \dots, x_{j_t})] = 0, \quad \{i_1, \dots, i_d \neq j_1, \dots, j_t\} \quad (7)$$

$$\text{Var}[y(x)] = \sum_{i=1}^N D_i + \sum_{1 \leq i < j \leq N} D_{ij} + \dots + D_{1,2,\dots,N} \quad (8)$$

Thus the total variance and the variance of different variables can be determined and the Sobol sensitivity indices can be calculated as follows:

$$S_{i_1, \dots, i_d} = D_{i_1, \dots, i_d} / D = \sum_{s \in I_{i_1, \dots, i_d}} a_s^2 / D \quad (9)$$

Where S_i is the main effect of variable i and S_{ij} is the cooperative effect of variable i and j .

3. Case study

Kaidu River Basin is located in North China with an area of 18,827 km² and an average elevation of 3100 m [11]. It plays an important role in protecting the ecological environment and green corridor of the lower reaches of the Tarim River [12]. The low rainfall and high temperature in this area bring notable drought periods and the regional water resources system degraded due to more and more frequently human activities. As a sparsely populated region, precise data in Kaidu River Basin is hard to be obtained due to the lack of regular monitoring. Moreover, uncertainties associated with the temporal and spatial variations in hydrological processes (e.g. precipitation, topography and evaporation) may bring errors in hydrological simulation and influence model performance. Thus, it is necessary to develop an effective uncertain quantification method for hydrological simulation to provide detailed information on water resource management and environmental protection.

The meteorological data and daily stream flow records, from 1996 to 2001, are obtained from the Bayanbulak meteorological station and the Dashankou hydrological station, respectively. The meteorological inputs for each ASA in SLURP model are derived using a weighted Thiessen polygon method with a lapse rate of 0.75°C and 1% per 100 m for temperature and precipitation. River network and boundaries of watershed are obtained based on the topography map, with a resolution of 100 × 100 m, using the topographic parameterization package.

4. Results and Discussion

The performance of SLURP model was estimated using Nash-Sutcliffe efficiencies (NSE) and determination coefficient (R^2). NSE values for calibration (2001-2005) and verification (2006-2010) are 0.693 and 0.673, respectively. R^2 values for calibration and verification are 0.850 and 0.826, respectively. The results not only indicate a good performance of SLURP for hydrological simulation in Kaidu river basin, but also lay a good foundation for further revealing the factors affecting hydrological processes from rainfall to streamflow in the study area. Then, a four dimension two order polynomial chaos expansions model based on sparse grid collection is formulated and the Sobol sensitivity indices for four sensitive parameters of SLURP, i.e., precipitation factor (P1), maximum capacity for fast store (P2), retention constant for fast store (P3) and retention constant for

slow (P4), are quantified to discovering the effect of parameters and their interactions on model performance.

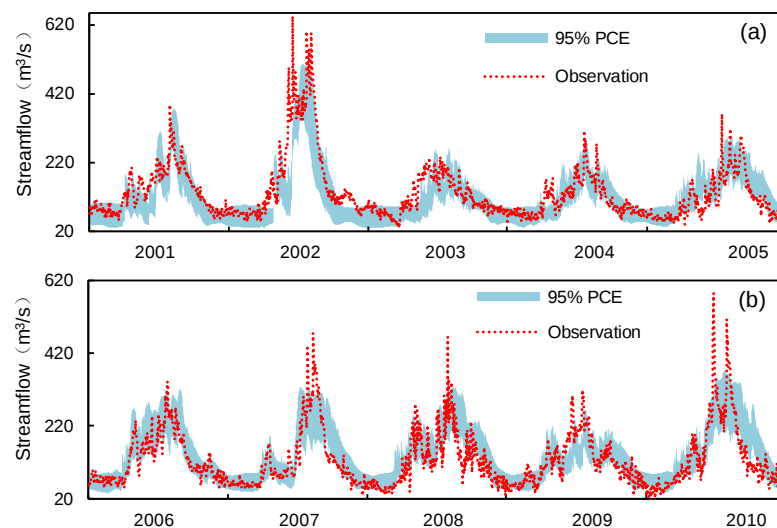


Figure 1. Daily streamflow [(a) Calibration (b) Verification]

The estimated uncertainty bounds of daily streamflow associated with 95% confidence interval of polynomial chaos expansions for the four parameters are presented in Figure 1. The results indicate a good coverage of the observed stream with 26% observation points out of the confidence interval. The results indicate that the four dimension two order polynomial chaos expansions model can be used as a surrogate model for hydrological simulation. On the other hand, about 9% observation points are lower than the lower bound of the confidence interval and 17% observation points are higher than the upper bound of the confidence interval, indicating a better simulation of low flow than that of peak flow. The underestimation of peak flow may be attributed to the potential observation errors.

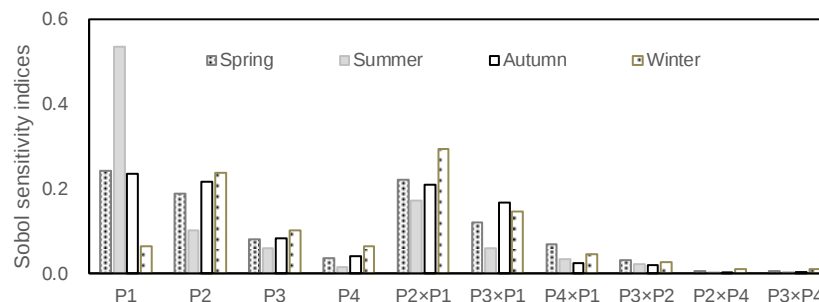


Figure 2. Sobol sensitivity indices in different seasons

Sobol sensitivity indices of different parameters and their interactions in different seasons (spring from March to May, summer from June to August, autumn from September to November and winter from December to the next February) are calculated based on the coefficients of polynomial chaos expansions (Figure 2). In summer, the indices of P1 is 0.535, which is higher than that of the other parameters. It is indicated that precipitation in summer is the major factor that affect runoff generation. On the contrary, the indices of P1 in winter is only 0.065, which is lower than that of the other parameters. These can be attributed to the low precipitation, as well as the freezing water temperature forming a thick layer of ice and snow on the land. The indices of P4 is 0.04 in spring, 0.01 in summer, 0.04 in autumn and 0.06 in winter, which are the lowest among the four parameters. It demonstrated that the effect of retention constant for slow, which is associated with the residence time of water in the saturated zone, is not significant. The indices of P2 are 0.19, 0.22 and 0.24 in spring, autumn and winter, which is higher than that (0.1) in summer. It is indicated that water holding in soil have higher

effect in dry season than in rainy season. The indices of $P2 \times P1$ are 0.22 in spring, 0.17 in summer, 0.21 in autumn and 0.29 in winter, which is higher than the value of the other interactions. It is because that excess infiltration is the main runoff mechanism and precipitation and maximum capacity for fast store are the major factors affecting runoff generation.

5. Conclusion

In this study, a sparse polynomial chaos expansions method was developed to quantify parameter uncertainties on model performance. The hydrological process from rainfall to runoff was simulated using SLURP model with consideration of canopy interception, snowpack, aerated soil storage and groundwater. The Sobol sensitivity indices was calculated based on the polynomial chaos expansions model and the sparse grid collection. On the other hand, uniform distribution of model parameters was used as the sampling distribution for sparse grid collection. In real world, multiple uncertain parameters may be interrelated to each other, leading to uncertain parameters distributions. Thus, other distributions (e.g., Gaussian) may be introduced to enhance the applicability of the developed sparse polynomial chaos expansions model.

References

- [1] Lenhart T, Fohrer, N, Frede H G. Effects of land use changes on the nutrient balance in mesoscale catchments. *Physics and Chemistry of the Earth, Parts A/B/C*, 2003, **28(33-36)**: 1301-1309.
- [2] Vincendon B, Ducrocq V, Saulnier G M, et al. Benefit of coupling the ISBA land surface model with a TOPMODEL hydrological model version dedicated to Mediterranean flash-floods. *Journal of Hydrology*, 2010, **394(1-2)**: 256-266.
- [3] Guerrero J L, Westerberg I K, Halldin S, Lundin L C, Xu C Y. Exploring the hydrological robustness of model-parameter values with alpha shapes. *Water Resources Research*, 2013, **49(10)**: 6700-6715.
- [4] Mahiny A S, Clarke K C. Simulating Hydrologic Impacts of Urban Growth Using SLEUTH, Multi Criteria Evaluation and Runoff Modeling. *Journal of Environmental Informatics*, 2013, **22(1)**: 27-38.
- [5] Jeremiah E, Sisson S A, Sharma A, Marshall L. Efficient hydrological model parameter optimization with Sequential Monte Carlo sampling. *Environmental Modelling & Software*, 2012, **38**: 283-295.
- [6] Kottegoda N T, Natale L, Raiteri E. Monte Carlo Simulation of rainfall hyetographs for analysis and design. *Journal of Hydrology*, 2014, **519**, A: 1-11.
- [7] Mortier S T, Hoey S V, Cierkens K, et al. A GLUE uncertainty analysis of a drying model of pharmaceutical granules. *European journal of pharmaceuticals and biopharmaceutics: official journal of Arbeitsgemeinschaft für Pharmazeutische Verfahrenstechnik*, 2013, **85(3)**: 984-995.
- [8] Blasone R S, Vrugt J A, Madsen H, et al.. Generalized likelihood uncertainty estimation (GLUE) using adaptive Markov Chain Monte Carlo sampling. *Advances in Water Resources*, 2008, **31(4)**: 630-648.
- [9] Garcia-Cabrejo O, Valocchi A. Global Sensitivity Analysis for multivariate output using Polynomial Chaos Expansion. *Reliability Engineering and System Safety*, 2014, 126:25-36.
- [10] Cooling C M, Ayres D A F, Prinja A K, et al. Uncertainty and global sensitivity analysis of neutron survival and extinction probabilities using polynomial chaos. *Annals of Nuclear Energy*, 2016, **88**: 158-17.
- [11] Kalra A, Li L, Li X, Ahmad S. Improving streamflow forecast lead time using oceanic-atmospheric oscillations for Kaidu River Basin, Xinjiang, China. *Journal of Hydrologic Engineering*, 2013, **18(8)**: 1031-1040.
- [12] Li Y P, Huang G H, Nie S L. Planning water resources management systems using a fuzzy-boundary interval-stochastic programming method. *Advances in Water Resources*, 2010, **33(9)**: 1105-1117.