Water Resources Research

RESEARCH ARTICLE
10.1029/2019WR025844

Distribution-Based Global Sensitivity Analysis in Hydrology

Valentina Ciriello[1], Ilaria Lauriola[1], and Daniel M. Tartakovsky[2]

[1]Dipartimento di Ingegneria Civile, Ambientale e dei Materiali (DICAM), Università di Bologna, Bologna, Italy, [2]Department of Energy Resources Engineering, Stanford University, Stanford, CA, USA

Key Points:
- Surrogate models based on polynomial chaos accelerate moment-independent GSA
- Our algorithm is used to identify impact of temperature variability on evapotranspiration
- Our spatial analysis of GSA metrics identifies possible climate change pathways

Correspondence to:
D. M. Tartakovsky,
tartakovsky@stanford.edu

Citation:
Ciriello, V., Lauriola, I., & Tartakovsky, D. M. (2019). Distribution-based global sensitivity analysis in hydrology. *Water Resources Research*, *55*, 8708–8720. https://doi.org/10.1029/2019WR025844

Received 21 JUN 2019
Accepted 3 SEP 2019
Accepted article online 9 SEP 2019
Published online 8 NOV 2019

Abstract Global sensitivity analysis (GSA) is routinely used in academic setting to quantify the influence of input variability and uncertainty on predictions of a quantity of interest. Practical applications of GSA are hampered by its high computational cost, which arises from the need to run large (e.g., groundwater) models multiple times, and by its reliance on the analysis of variance, which formally requires input parameters to be uncorrelated. The former difficulty can be alleviated by replacing expensive models with inexpensive (e.g., polynomial) surrogates, while adoption of distribution-based (rather than variance-based) metrics can, in principle, overcome the latter but at significantly increased computational cost. To make use of distribution-based GSA feasible for regional-scale models with a large number of degrees of freedom, we supplement it with a surrogate model built with polynomial chaos expansions with analytically updated coefficients. We demonstrate the computational efficiency of our algorithm on a case study dealing with evaluation of the effects of temperature variability on annual evapotranspiration at the regional scale.

1. Introduction

Incomplete knowledge about natural environments affects our ability to model many physical, chemical, and biological processes of interest, giving rise to so-called epistemic uncertainty. Uncertainty affects both geomaterial properties (e.g., porosity and permeability) and mathematical representations of relevant phenomena (e.g., the choice between local and nonlocal models of subsurface solute transport). Probabilistic approaches have long been used to deal with these issues and to provide predictions with an appropriate/desired level of confidence (e.g., Jurado et al., 2012; Tartakovsky, 2013, and the references therein).

Sensitivity analysis plays an important role in assessing the impact of parametric uncertainty on the robustness of model predictions. Local sensitivity analysis starts with the premise that model calibration gives a well-defined point in the parameter space and then investigates how small deviations from that point (typically, one parameter at a time) affect predictions of a quantity of interest (QoI). In contrast, global sensitivity analysis (GSA) makes no assumption about either uniqueness or optimality of a particular set of parameters obtained via calibration. Instead, it adopts the probabilistic framework by assigning to each parameter a confidence interval equipped with a probabilistic model (e.g., treating hydraulic conductivity as a lognormal random field) and then evaluating the impact of uncertainty in parameter values on uncertainty in predictions of the QoI. Consequently, it estimates the degree to which each parameter and interactions between different parameters contribute to uncertainty in (typically, the variance of) the model predictions (Saltelli et al., 2000, and reference therein). GSA enables one to identify parameters, which should be measured to maximally reduce the predictive uncertainty. Despite of its onerous computational cost, GSA represents a necessary component of a rigorous risk assessment and provides relevant information to support calibration and selection of hydrologic models (Borgonovo et al., 2017; Ciriello, Di Federico, et al., 2013; Ciriello, Guadagnini, et al., 2013, 2015).

A recent review of the progress in GSA practice can be found in Ferretti et al. (2016). Its increasing popularity is linked to the use of surrogate models, which provide an easy-to-compute approximation to computationally demanding high-fidelity models. One could argue that without such surrogates, GSA would be unfeasible for most, if not all, hydrologic modeling of practical significance (e.g., Ciriello, Di Federico, et al., 2013, 2017). Polynomial chaos expansions (PCEs) (Ghanem & Spanos, 1991; Wiener, 1938) are widely used to construct surrogate models of subsurface processes (e.g., Ashraf et al., 2013; Deman et al., 2015; Marrel et al.,

2015). PCEs are often deployed to estimate variance-based sensitivity metrics, such as the Sobol' indices, (Assouline et al., 2017; Di Fusco et al., 2018).

Variance-based GSA is strictly applicable to uncorrelated random inputs. This renders it inapplicable, except as an uncontrollable approximation, to hydrologic systems, whose parameters are often correlated random fields and are cross-correlated with each other. Such problems call for the use of moment-independent GSA techniques, including distribution-based GSA (Borgonovo, 2007), a PCE-based method of Caniou and Sudret (2011), and Bayesian nets (Um et al., 2019). Yet hydrologic applications of moment-independent GSA accelerated by PCE-based surrogates remain scarce (e.g., Rajabi et al., 2015).

In section 2 we present a distribution-based GSA, which is accelerated by the efficient construction of PCE-based surrogates with analytically updated coefficients. To be specific, we use the Borgonovo δ index (Borgonovo, 2007) as a metric for the distribution-based GSA. Section 3 contains the an application of our approach to the problem of identification of the impact of temperature variability on potential evapotranspiration in the Emilia-Romagna Region (Italy) over a 30-year period, from 1971 to 2000. Main findings and conclusions drawn from our study are summarized in section 4.

2. Materials and Methods

A brief overview of PCE is provided in section 2.1. It is followed by presentation of our approach to analytic update of PCE coefficients (section 2.2) and its use within a distribution-based GSA framework (section 2.3).

2.1. Generalized PCE

Let h denote a QoI computed with a model $f(\cdot)$, whose input is a set of M_{par} model parameters $\mathbf{p} = \{p_1, \ldots, p_{M_{\text{par}}}\}$. Parametric uncertainty is dealt with by treating $f(\mathbf{p})$ as a model with random inputs (or "stochastic" model). The parameters might be, and typically are, cross-correlated, for example, if $\mathbf{p}$ represents uncertain (random) values of hydraulic conductivity $K(\mathbf{x})$ in elements or nodes of a discretized simulation domain. In that case, singular-value decomposition techniques, such as a truncated Karhunen-Loève transformation, are used to approximate $\mathbf{p}$ with a set of M ($M \ll M_{\text{par}}$) mutually uncorrelated identically distributed random variables $\boldsymbol{\xi}=\{\xi_1,\ldots,\xi_M\}$. Alternatively, if $\mathbf{p}=\{p_1,\ldots,p_M\}$ represents $M_{\text{par}}=M$ correlated random variables (rather than random fields), then the Rosenblatt transform (e.g., section 4.1 in Um et al., 2019) maps $\mathbf{p}$ onto a set of random variables $\boldsymbol{\xi}=\{\xi_1,\ldots,\xi_M\}$ that are independent and identically distributed on the interval (0,1).

If the QoI h has finite variance, σ_h^2, then the function/model $h=f(\boldsymbol{\xi})$ can be approximated via a PCE (e.g., Ghanem & Spanos, 1991; Wiener, 1938; Xiu & Karniadakis, 2002)

$$\hat{h} = \sum_{\mathbf{a}\in\mathbb{N}^M} s_{\mathbf{a}} \Psi_{\mathbf{a}}(\xi). \tag{1}$$

Here multi-indices $\mathbf{a} = \{a_1, \ldots, a_M\} \in \mathbb{N}^M$ are associated with a multivariate polynomial $\Psi_{\mathbf{a}}(\mathbf{x})$ of degree $|\mathbf{a}| = \sum_{i=1}^{M} a_i$, which is given by

$$\Psi_{\mathbf{a}}(\xi) = \prod_{i=1}^{M} \Gamma_{a_i}^{(i)}(\xi_i). \tag{2}$$

The polynomials $\Gamma_k^{(i)}$ (with $k\in\mathbb{N}$) are orthogonal, $\mathbb{E}[\Gamma_{k_1}^{(i)}, \Gamma_{k_2}^{(i)}] = \alpha_i \delta_{k_1 k_2}$ with α_i being the normalizing constants and $\delta_{k_1 k_2}$ denoting the Kronecker delta function. The polynomial type is selected to match a probability density function (PDF) of the input variables ξ_i (Xiu & Karniadakis, 2002).

In (1), the coefficients $s_{\mathbf{a}}$ are deterministic coordinates of the spectral decomposition, and the multivariate polynomials $\Psi_{\mathbf{a}}(\boldsymbol{\xi})$ constitute an orthonormal basis with respect to the joint PDF of $\boldsymbol{\xi}$, that is, $E[\Psi_{\mathbf{a}}(\xi), \Psi_{\mathbf{b}}(\xi)] = \delta_{\mathbf{ab}}$ (Xiu & Karniadakis, 2002). The PCE (1) can be rewritten as

$$\hat{h} = \sum_{j=0}^{P-1} s_j \Psi_j(\xi), \qquad P = \frac{(M+p)!}{M!p!}, \tag{3}$$

where p is the highest degree of the expansion, that is, $|\mathbf{a}| \le p$ for all $\mathbf{a}\in\mathbb{N}^M$. Table 1 provides an example of the terms involved in a second-order Legendre PCE with p=2 and M=3 uniformly distributed uncertain input parameters.

Table 1
Terms of the Legendre PCE of Order p=2, for M=3 Uncertain Input Parameters Uniformly Distributed on [−1,1]

a	$\|\mathbf{a}\|$	j	$\Psi_{\mathbf{a}}(\boldsymbol{\xi})$
(000)	0	0	1
(100)	1	1	ξ_1
(010)	1	2	ξ_2
(001)	1	3	ξ_3
(200)	2	4	$(3\xi_1^2-1)/2$
(110)	2	5	$\xi_1\xi_2$
(101)	2	6	$\xi_1\xi_3$
(020)	2	7	$(3\xi_2^2-1)/2$
(011)	2	8	$\xi_2\xi_3$
(002)	2	9	$(3\xi_3^2-1)/2$

Note. PCE = polynomial chaos expansion.

Among nonintrusive methods for the computation of PCE coefficients, s_j can be selected to minimize the variance of a suitable residual, $\varepsilon = |\hat{h}-h|$ (Sudret, 2008). An optimal set of regression points in the (random) parameter space is determined on the basis of the arguments adopted for integral estimation through Gaussian quadrature. To this end, the probabilistic collocation method employs the roots of the polynomial of one order higher than p, to assure proper sampling of the region associated with the largest probability according to the PDF of the input parameters $\boldsymbol{\xi}$.

2.2. Analytic Update of PCE Coefficients

If one of the input parameters, for example, ξ_i, is set to a given value, $\overline{\xi}_i$, then the approximation of the response surface provided by the PCE (3) changes to

$$\hat{h} = \sum_{j=0}^{P'-1} q_j \Psi_j(\xi_{-i}), \qquad P' = \frac{(M-1+p)!}{(M-1)!p!}, \tag{4}$$

where $\boldsymbol{\xi}_{-i}$ is the set of the input parameters excluding ξ_i, i.e., $\boldsymbol{\xi}_{-i}=\{\xi_1,\ldots,\xi_{i-1},\xi_{i+1},\ldots,\xi_M\}$. Computation of the coefficients q_j requires a given number of evaluations of $f(\boldsymbol{\xi})$, as described in the preceding paragraph. If a single solve of the model $f(\cdot)$ is expensive, this computation might become prohibitively expensive.

Instead, we propose to update the PCE coefficients s_j, once the value of the ith input ξ is set to $\overline{\xi}_i$, without resorting to the additional full model runs. Specifically, we compute s_0', the updated leading coefficient s_0 in (3), as

$$s_0' = s_0 + \sum_{\mathbf{a}\in\mathbb{N}^M | a_i\neq 0} s_{\mathbf{a}} \Psi_{\mathbf{a}}(\xi_{-i}, \overline{\xi}_i). \tag{5a}$$

All the other coefficients $s_{\mathbf{a}}$ are updated according to

$$s'_{\mathbf{a}} = \begin{cases} 0 & \text{if } a_i \neq 0 \quad (5b) \\ s_{\mathbf{a}} + \sum_{\mathbf{b}\in\mathbb{N}^M | b_i\neq 0, b_{k\neq i}=a_k} s_{\mathbf{b}} \Gamma^{(i)}_{b_i}(\boldsymbol{\xi}_{-i}, \overline{\xi}_i) & \text{if } a_i = 0. \quad (5c) \end{cases}$$

For illustrative purposes, Table 2 shows the implementation of this update strategy for the example reported in Table 1 when ξ_1 is fixed and set to $\xi_1 = \overline{\xi}_1$.

Table 2
Automatic Update of the PCE Coefficients in Table 1 When ξ_1 is Fixed and Set to $\xi_1 = \overline{\xi}_1$

a	Eq.	s_j'
(000)	(5a)	$s_0 + s_1\overline{\xi}_1 + s_4(3\overline{\xi}_1^2-1)/2$
(100)	(5b)	0
(010)	(5c)	$s_2 + s_5\overline{\xi}_1$
(001)	(5c)	$s_3 + s_6\overline{\xi}_1$
(200)	(5b)	0
(110)	(5b)	0
(101)	(5b)	0
(020)	(5c)	s_7
(011)	(5c)	s_8
(002)	(5c)	s_9

Note. PCE = polynomial chaos expansion.

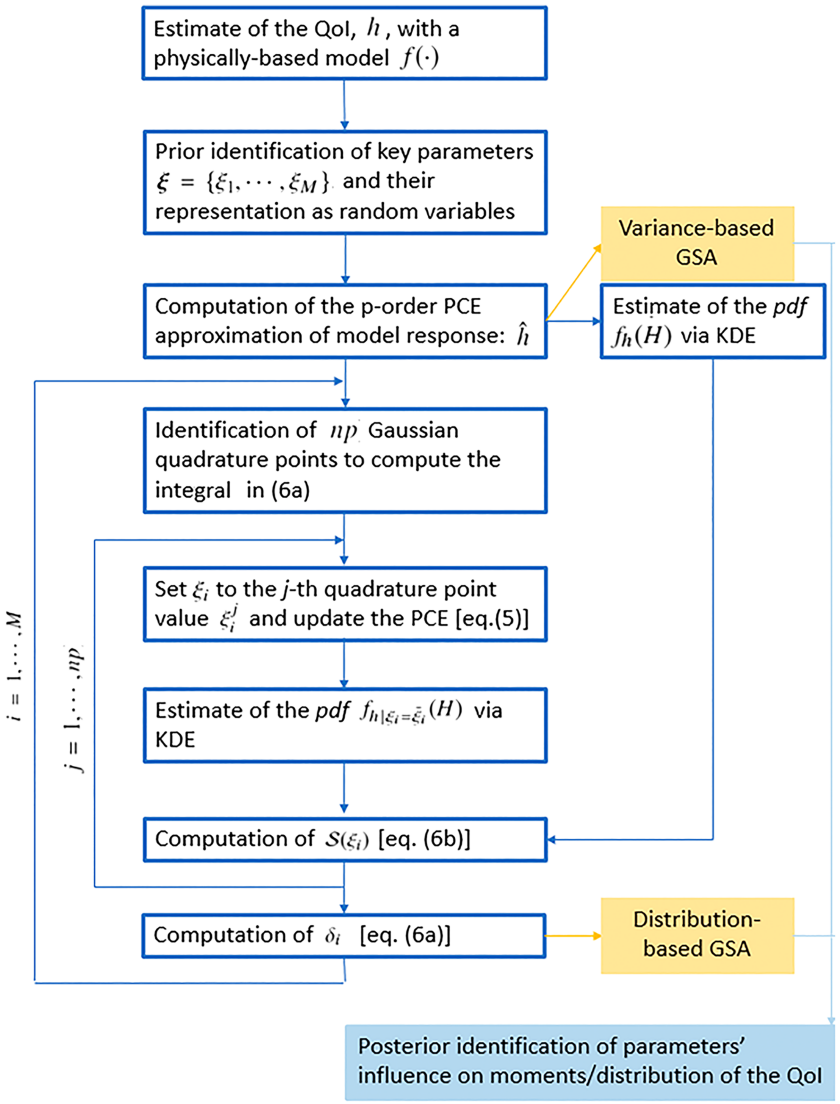

Figure 1. Workflow of the proposed approach for distribution-based global sensitivity analysis (GSA). PCE = polynomial chaos expansion; KDE = kerned density estimator.

2.3. Algorithm for Distribution-Based GSA

As mentioned in section 1, PCE is routinely used to accelerate variance-based GSA. For example, Sobol' metrics (Sobol', 1993, 2001) can be expressed analytically in terms of PCE coefficients, drastically reducing the computational cost of GSA (Sudret, 2008). The Sobol' indices do not require a model to be linear, and enable one to explore the full range of input uncertainty and interactions among input parameters. Nevertheless, variance-based approaches ignore possible correlations among inputs, and provide only a moment-dependent description of uncertainty.

Distribution-based methods overcome these limitations by analyzing the influence of parametric uncertainty on the entire PDF of the QoI h, $f_h(H)$, rather than on its variance σ_h^2. One of the metrics used in distribution-based GSA is the Borgonovo importance measure (Borgonovo, 2007) of the ith parameter ξ_i with PDF $f_{\xi_i}(\Xi_i)$,

$$\delta_i = \frac{1}{2}\mathbb{E}[\mathcal{S}(\xi_i)] = \int_{D_{\xi_i}} \mathcal{S}(\Xi_i) f_{\xi_i}(\Xi_i)\, \mathrm{d}\Xi_i, \tag{6a}$$

where D_{ξ_i} is the domain of definition of the random variable ξ_i and $\mathcal{S}(X_i)$ is the Borgonovo shift, defined as a discrepancy between the PDF $f_h(H)$ and the conditional PDF $f_{h|\xi_i=\bar{\xi}_i}(H)$:

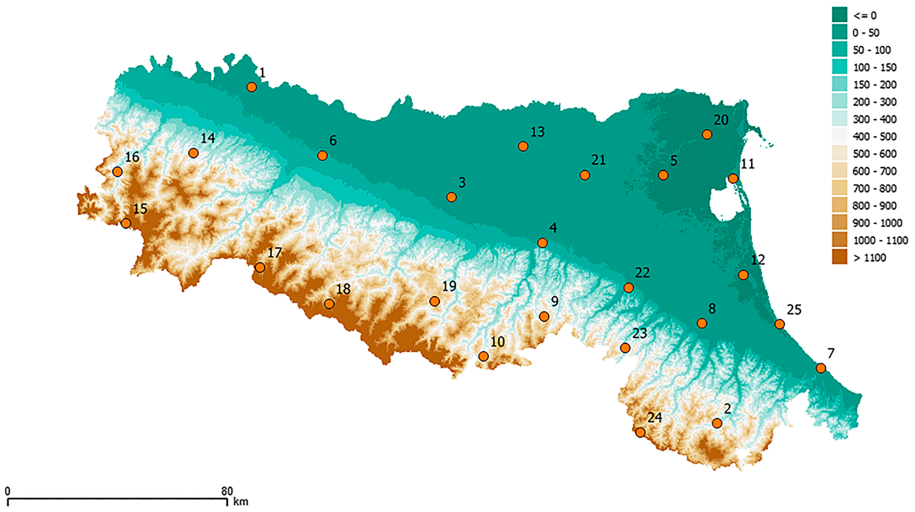

Figure 2. Locations of the meteorological stations in Emilia-Romagna region of Italy.

$$\mathcal{S}(\xi_i) = \int_{D_h} |f_h(H) - f_{h|\xi_i=\bar{\xi}_i}(H)| \mathrm{dH}. \tag{6b}$$

By definition, the metric δ_i of a given parameter ξ_i takes values within the range [0,1] and satisfies the condition $\delta_{1,2,\ldots,M}=1$.

To make the computation of the Borgonovo metric feasible, we combine it with the PCE with analytically updated coefficients (5). Figure 1 shows the workflow of our algorithm, with its main steps grouped into the following five categories:

1. computation of the PCE surrogate $\hat{h}$ in (3), obtained over the range of variability of all the uncertain parameters;
2. estimation of the PDF of $\hat{h}$, $f_h(H)$, by means of a kerned density estimator;
3. computation of the integrals in (6a) by means of the Gaussian quadrature with np points for each parameter ξ_i;
4. estimation of $\hat{h}(\cdot, \xi_i = \xi_i^j)$ (where j=1,...,np) by applying the analytic update of PCE coefficients every time the parameter ξ_i is fixed to its quadrature point value; and
5. computation of δ_i and identification of the parameter's influence on $f_h(H)$.

The only significant computational time is required to construct the PCE surrogate $\hat{h}$. All subsequent steps are either analytic postprocessing or efficient computational schemes (e.g., Gaussian quadrature and kernel density estimation) that bear a virtually negligible computational cost, when applied to the PCE approximation.

3. Application

Evapotranspiration is one of the key hydrological phenomena, whose dependence on several interconnected factors makes its quantitive predictions notoriously uncertain. Variance-based GSAs have been used to investigate the relative importance of uncertainty in some of these factors on predictions of evapotranspiration (e.g., Guo et al., 2017). By construction, such analyses rely on a dubious assumptions of the input parameters being uncorrelated. The moment-independent GSA described above obviates the need for this oversimplification; we use this approach to investigate the influence of temperature variability on potential evapotranspiration at the regional scale. We also demonstrate how performing GSA in time allows one to identify both spatial and temporal evapotranspiration patterns. The computational efficiency of our method enables this analysis, which otherwise might not be feasible given its significant cost.

3.1. Study Area and Data Collection

To demonstrate the applicability of our GSA approach to real-world problems, we consider Emilia-Romagna region of Italy, which is instrumented with 25 meteorological stations covering the entire region (Figure 2). The names and coordinates of the selected stations are collated in Table 3.

Table 3
Characteristics of the Meteorological Stations

k	Name	Alt (m slm)	Lat (°)	Long (°)	$ET_{0_k}^{ann}$ (mm)	$P_{0_k}^{ann}$ (mm)	r_k
1	Villanova sull'Arda	40	45.03	10.01	911.50	818.46	1.11
2	Sarsina	247	43.89	12.10	860.31	1,024.56	0.84
3	Modena urbana	73	44.66	10.92	881.30	636.17	1.39
4	Bologna urbana	78	44.5	11.33	981.78	729.05	1.35
5	Ostellato	0	44.71	11.89	937.43	619.86	1.51
6	Parma urbana	79	44.8	10.33	968.61	809.87	1.20
7	Rimini urbana	16	44.06	12.57	858.76	706.58	1.22
8	Forli urbana	51	44.22	12.04	907.33	752.87	1.21
9	Loiano	741	44.26	11.33	806.40	978.61	0.82
10	Porretta terme	352	44.15	10.98	860.28	1,436.43	0.60
11	Idrovora di Guagnino	1	44.69	12.21	889.21	634.17	1.40
12	Classe	2	44.37	12.24	940.68	663.65	1.42
13	Finale Emilia	16	44.82	11.25	950.54	638.14	1.49
14	Castellana Groppo	434	44.81	9.73	836.25	969.91	0.86
15	Boschi d'Aveto diga	616	44.59	9.42	682.47	1,504.52	0.45
16	Bobbio	270	44.76	9.38	884.38	925.68	0.96
17	Bosco centrale	902	44.44	10.03	738.71	1,816.92	0.41
18	Ligonchio centrale	900	44.32	10.34	662.80	1,772.44	0.37
19	Pavullo	678	44.32	10.83	870.18	898.98	0.97
20	Codigoro	2	44.84	12.10	927.45	624.95	1.48
21	Malalbergo	12	44.72	11.53	1,005.84	628.00	1.60
22	Imola	42	44.35	11.71	961.25	783.63	1.23
23	San Cassiano	230	44.15	11.69	894.77	934.97	0.96
24	Campigna	1,068	43.87	11.75	581.45	1,688.08	0.34
25	Cesenatico	2	44.21	12.40	932.71	721.54	1.29

Note. A Station's reference number k and name; altitude, latitude, and longitude; annual potential evapotranspiration $ET_{0_k}^{ann}$ and precipitation $P_{0_k}^{ann}$, averaged over the 30-year period 1971–2000; and their ratio $r_k = ET_{0_k}^{ann}/P_{0_k}^{ann}$.

We use minimum (T_{min}) and maximum (T_{max}) temperature measurements collected at these stations over a 30-year period, from 1971 to 2000. Based on these data, we compute the daily averages $T_{av}=(T_{max}+T_{min})/2$ and ranges $\Delta T=T_{max}-T_{min}$, as well as their monthly averages and standard deviations. This procedure yields random quantities $T_{av_{ijk}}$ and ΔT_{ijk}, where i=1,...,30 indicates the year; j=1,...,12 the month; and k=1,...,25 the station (Table 3).

Potential evapotranspiration $ET_{0_{ijk}}$ is computed following Hargreaves and Samani (1982) as

$$ET_{0_{ijk}} = 0.0135 R_{s_{ijk}} (T_{av_{ijk}} + 17.8), \tag{7a}$$

where the solar radiation $R_{s_{ijk}}$ is estimated from

$$R_{s_{ijk}} = R_{a_{ijk}} K_T \sqrt{\Delta T_{ijk}} \tag{7b}$$

with R_a denoting the extraterrestrial radiation and K_T being an empirical coefficient whose value typically varies between 0.16 (interior regions) and 0.19 (coastal regions). Table 3 contains the averages, over the time period between 1971 and 2000, of the annual potential evapotranspiration, $ET_{0_k}^{ann}$, precipitation, $P_{0_k}^{ann}$, and their ratio, $r_k = ET_{0_k}^{ann}/P_{0_k}^{ann}$. Appendix A provides further details on trends in seasonal and annual evapotranspiration at the selected locations over the reference period.

3.2. Uncertainty Propagation

We model $T_{av_{ijk}}$ and ΔT_{ijk} as lognormal random variables, whose moments (mean and variance) are estimated from the 30-year dataset as described in the previous section. The goal is to analyze the effect of their annual variability on the annual potential evapotranspiration at each station, that is, on $ET_{0_{ik}}^{ann}$. In order to reduce the random dimension of the problem, we consolidate the ET data into four new random variables for each station and year: cumulative potential ET in winter, $ET_{0_{ik}}^{win}$; spring, $ET_{0_{ik}}^{spr}$; summer, $ET_{0_{ik}}^{sum}$; and

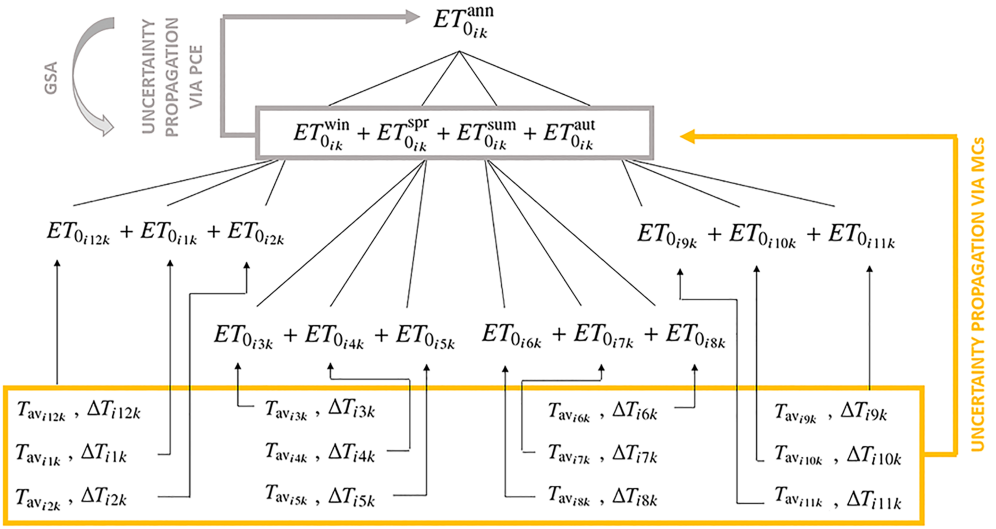

Figure 3. Workflow for data analysis and uncertainty quantification for the problem of estimation the impact of uncertainty in temperature measurements on the predictions of annual potential evapotranspiration. GSA = global sensitivity analysis. PCE = polynomial chaos expansion.

autumn, ET^{aut}_{0ik}. Monte Carlo simulations are used to propagate uncertainty in the inputs ($T_{av_{ijk}}$ and ΔT_{ijk}) toward outputs (these four new variables). The model $f(\cdot)$ in this case is given explicitly by (7a). The histogram of the Monte Carlo simulation results (not shown here) reveals the four output variables to be approximately lognormal. The workflow for data analysis and uncertainty quantification is shown in Figure 3.

3.3. Distribution-Based GSA

We apply our GSA algorithm (see the workflow in Figure 1) to the explicit model

$$ET^{ann}_{0ik} = ET^{win}_{0ik} + ET^{spr}_{0ik} + ET^{sum}_{0ik} + ET^{aut}_{0ik} \tag{8}$$

in order to evaluate the relative impact of variability of evapotranspiration in the different seasons on its overall annual value. For each year in the range 1971–2000, we compute the Borgonovo importance metric for the four inputs, δET^{win}_{0ik}, δET^{spr}_{0ik}, δET^{sum}_{0ik}, and δET^{aut}_{0ik}. Averages of these indices over the 30 years are denoted by $\delta ET^{\cdot}_{0_k}$, and averages over the 25 meteorological stations by $\delta ET^{\cdot}_{0_i}$.

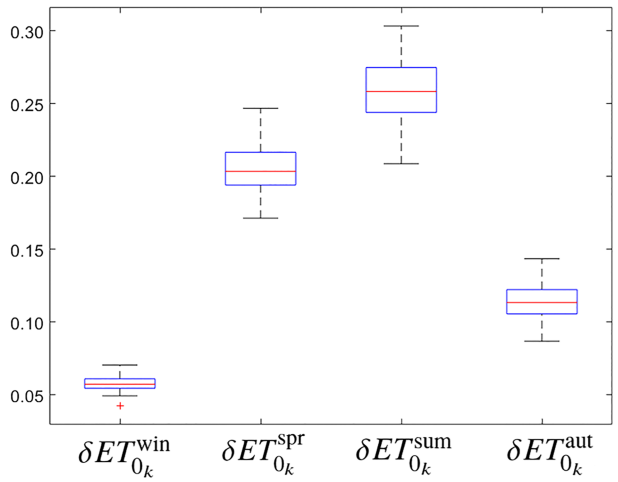

Figure 4. Boxplot of the variation in the Borgonovo indices $\delta ET^{win}_{0_k}$, $\delta ET^{spr}_{0_k}$, $\delta ET^{sum}_{0_k}$, and $\delta ET^{aut}_{0_k}$. The red line in each box indicates the median, the blue lines indicate the upper/lower quartiles, and the black lines the 10th/90th percentiles.

Figures 4 and 5 exhibit boxplots of the variation in $\delta ET^{win}_{0_k}$, $\delta ET^{spr}_{0_k}$, $\delta ET^{sum}_{0_k}$, and $\delta ET^{aut}_{0_k}$ and their spatial distribution, respectively. Conversely, Figure 6 depicts trends of δET^{win}_{0ik}, δET^{spr}_{0ik}, δET^{sum}_{0ik}, and δET^{aut}_{0ik} at the different stations (gray lines), together with the average trends given by $\delta ET^{win}_{0_i}$, $\delta ET^{spr}_{0_i}$, $\delta ET^{sum}_{0_i}$, and $\delta ET^{aut}_{0_i}$ (red lines). Table 4 reports the number of stations characterized by positive/negative trends of the time series of the δ indices, $\delta ET^{\cdot}_{0_{i,k}}$, with a 0.05 confidence level, based on the Mann-Kendall tau test (Kendall, 1938). A spatial description of these results is plotted in Figure 7.

The index $\delta ET^{win}_{0_k}$ has the average 0.056 over the stations and the standard deviation 0.003 (Figure 4). It is much smaller than the other three indices at each station as shown in Figure 5. This signifies that the variability/uncertainty in the winter potential ET plays a minor role in annual-scale predictions throughout Emilia-Romagna region. The Mann-Kendall tau test identifies a positive trend in δET^{win}_{0ik} with the confidence level ρ<0.05 at 11 stations

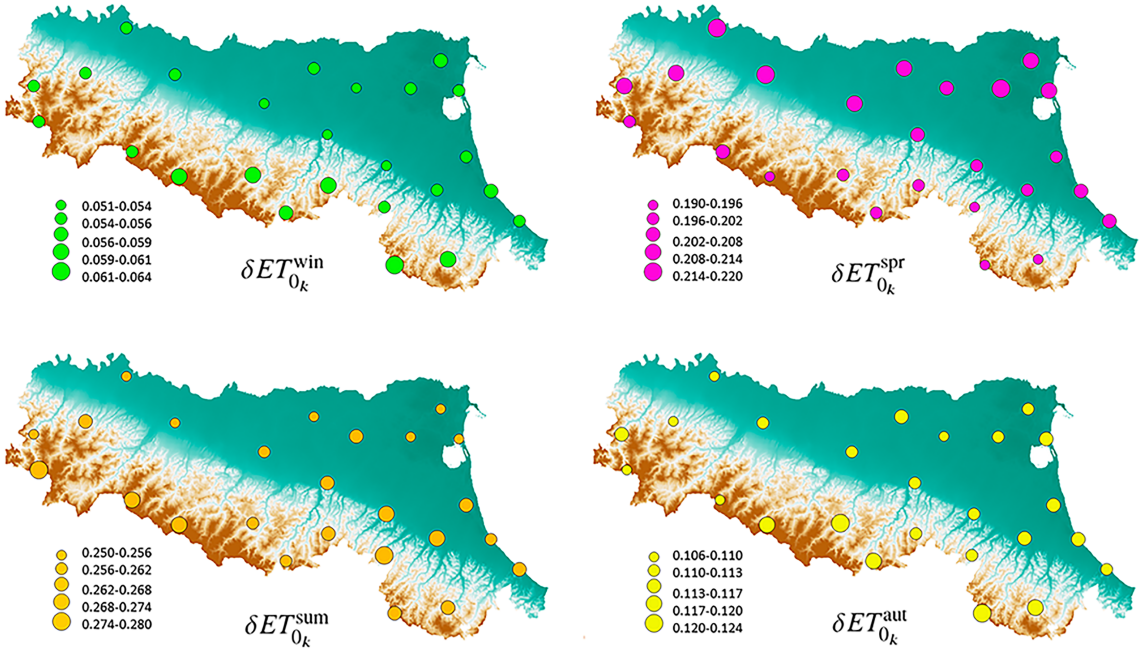

Figure 5. Values of the Borgonovo indices $\delta ET_{0_k}^{win}$, $\delta ET_{0_k}^{spr}$, $\delta ET_{0_k}^{sum}$, and $\delta ET_{0_k}^{aut}$ at the 25 meteorological stations.

(Table 4; Figure 7). This suggests a slight increase in the relative impact of the variability/uncertainty in winter evapotranspiration with time as shown in Figure 6.

The second smallest index is $\delta ET_{0_k}^{aut}$ (Figure 4), with the average 0.11 and standard deviation of 0.0048, indicating the relative unimportance of the variability/uncertainty in the autumn potential ET. No significant positive or negative trends are associated with $\delta ET_{0_{ik}}^{aut}$ according to the Mann-Kendall tau test with the confidence level ρ<0.05 (Table 4; Figure 7) and as represented in Figure 6.

Variability/uncertainty in the springtime evapotranspiration is more relevant, as indicated by average value of $\delta ET_{0_k}^{spr}$ equal to 0.20, with a standard deviation 0.0085 (Figure 4). Variability/uncertainty in the summertime evapotranspiration dominates predictive uncertainty in estimates of the annual ET at all stations throughout the region, with the index $\delta ET_{0_k}^{sum}$ having the mean 0.26 and standard deviation of 0.0078

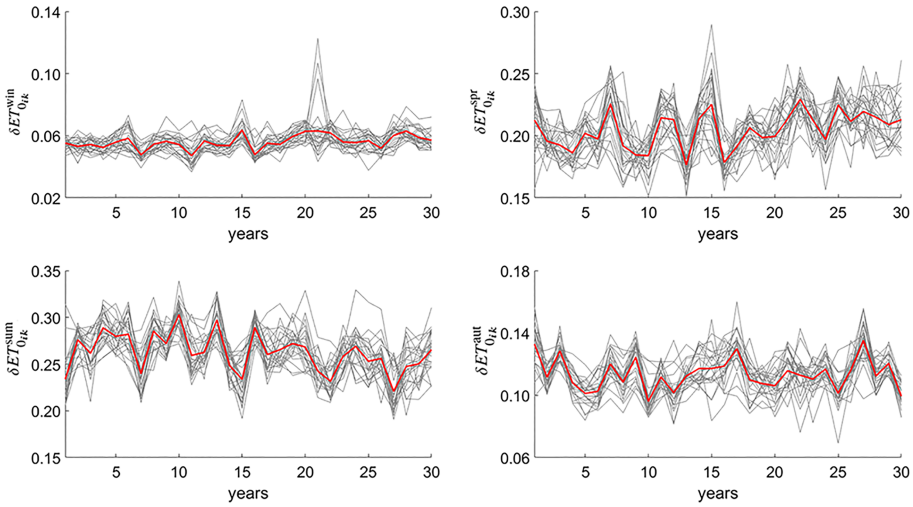

Figure 6. Values of $\delta ET_{0_{ik}}^{win}$, $\delta ET_{0_{ik}}^{spr}$, $\delta ET_{0_{ik}}^{sum}$, and $\delta ET_{0_{ik}}^{aut}$ assumed over the 30 years at each station (gray lines); red lines in the subplots represent the average trends of $\delta ET_{0_i}^{win}$, $\delta ET_{0_i}^{spr}$, $\delta ET_{0_i}^{sum}$, and $\delta ET_{0_i}^{aut}$.

Table 4
Trends in the Borgonovo Indices $\delta ET_{0_{ik}}^{win}$, $\delta ET_{0_{ik}}^{spr}$, $\delta ET_{0_{ik}}^{sum}$, and $\delta ET_{0_{ik}}^{aut}$

	$\delta ET_{0_{ik}}^{win}$	$\delta ET_{0_{ik}}^{spr}$	$\delta ET_{0_{ik}}^{sum}$	$\delta ET_{0_{ik}}^{aut}$
Number of stations:				
Positive (ρ<0.05)	11	11	0	1
Positive n.s.	11	12	3	13
Negative n.s.	3	2	12	11
Negative (ρ<0.05)	0	0	10	0

Note. The trends are identified with the Mann-Kendall tau test. The abbreviation "n.s." refers to "not significant" and denotes a level of significance ≥0.05.

(Figure 4). Trends of indices $\delta ET_{0_{ik}}^{spr}$ and $\delta ET_{0_{ik}}^{sum}$ are opposite as shown in Figure 6. While $\delta ET_{0_{ik}}^{spr}$ exhibits a clear positive trend associated with the confidence level ρ<0.05 at 11 stations, $\delta ET_{0_{ik}}^{sum}$ has a negative trend at 10 stations (Table 4). These stations are located along the coastline (East boundary) and along *via Aemilia*, which separates the northwestern and southeastern parts of the region; $\delta ET_{0_{ik}}^{sum}$ shows no negative trend at only three stations. At the stations in the coastal area, $\delta ET_{0_{ik}}^{spr}$ exhibits a positive trend with the confidence level ρ<0.05; the same is detected at some stations in the mountain area (Figure 7).

A positive trend in the average potential evapotranspiration during the 30-year period is observed at each season, and, in particular, $ET_{0_i}^{sum}$ has the highest rate of increase (Appendix A). At the same time, our results reveal that the variability/uncertainty in annual evapotranspiration throughout the region is most sensitive to the variability/uncertainty in summer temperature. The trend in $\delta ET_{0_{ik}}^{sum}$ shows that the importance of the variability/uncertainty in summer evapotranspiration decreases, being overtaken by the importance of the variability in spring temperature.

This finding indicates that analysis of the time averages of a variable (e.g., seasonal) ET and its sensitivity indices provides only a partial picture of the phenomenon. Analysis of temporal trends yields information about the rates of change, which can have opposite signs for the variable and its sensitivity indices so that the increase in the mean of the variable can be accompanied by the decrease in the impact of its variability. Furthermore, a spatial analysis of trends in global sensitivity metrics facilitates the identification of possible

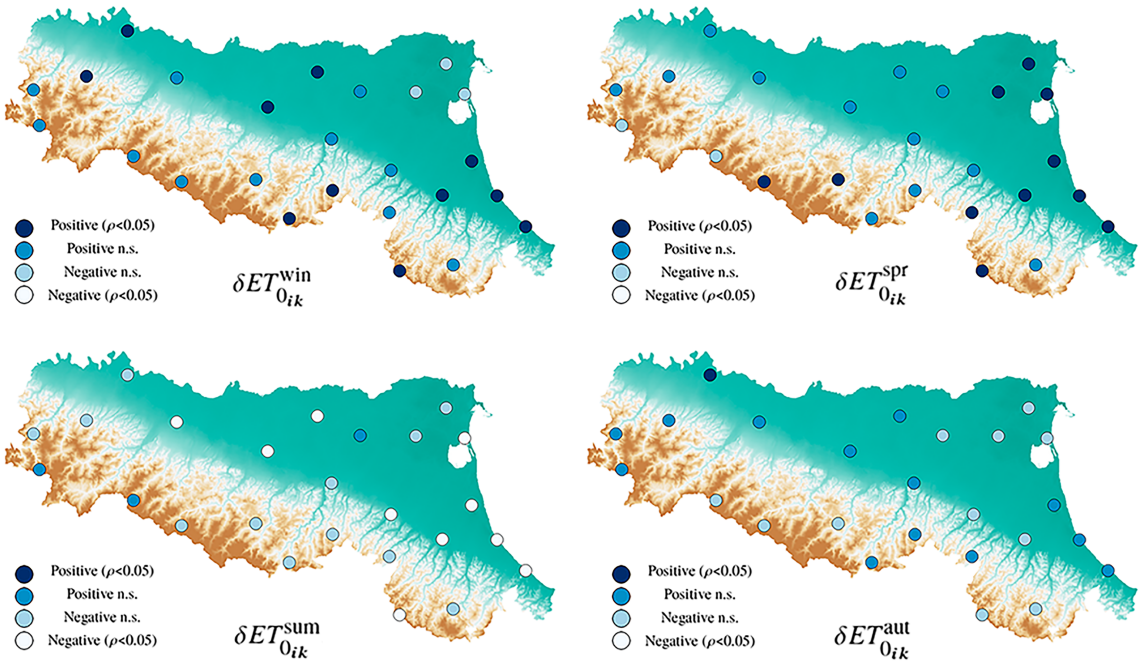

Figure 7. Meteorological stations with positive/negative trends in the Borgonovo indices $\delta ET_{0_{ik}}^{win}$, $\delta ET_{0_{ik}}^{spr}$, $\delta ET_{0_{ik}}^{sum}$, and $\delta ET_{0_{ik}}^{aut}$. The trend analysis is based on the Mann-Kendall tau test, whose results are reported in Table 4.

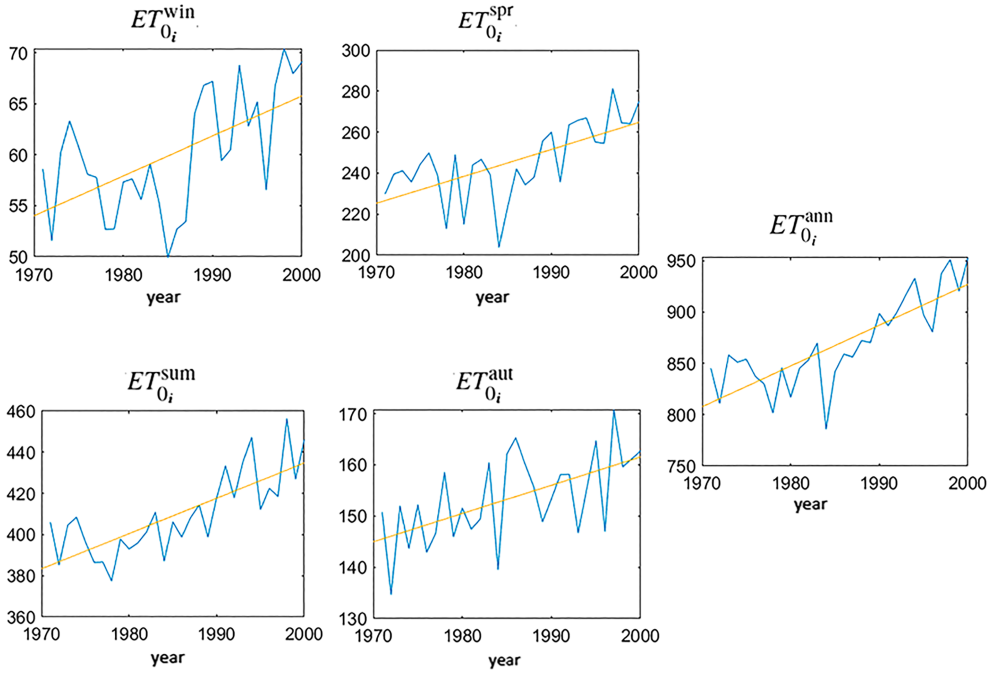

Figure A1. Seasonal and annual potential evapotranspiration (in mm), averaged over the 25 meteorological stations, during the 30-year period from 1971 to 2000.

climate change pathways at the scale of interest. This analysis can be readily extended to larger spatial scales and to longer time horizons, thus becoming relevant to studies of the impact of climate change on evapotranspiration.

This demonstration of our distribution-based GSA required the Bergonovo sensitivity metrics to be computed 750 times (at each station, at each year). This relatively low number of model runs highlights the effectiveness of the proposed method, making it useful in other contexts in which traditional GSA approaches would be associated with an unfordable computational effort. A comparison with variance-based metrics (Sobol' indices) is provided in Appendix B. A good agreement between these two GSA approaches is to be expected in our application, since it deals with a model that is both linear and additive.

4. Conclusions

We presented a new PCE-based algorithm to accelerate distribution-based GSA, which allows one to handle cross-correlations between model input parameters. The algorithm is based on the analytical update of PCE coefficients in order to avoid the derivation of multiple PCE surrogates when parameters are set to given values. The latter occurs when a numerical scheme (e.g., Guassian quadrature) is applied to compute the integrals defining moment-independent metrics. Consequently, the computational time associated with the analysis is largely determined by the construction of the primary PCE surrogate for the full random parameter space. All the subsequent steps are analytic postprocessing or involve numerical schemes applied to the polynomial surrogates and, hence, have a negligible computational cost. The reduced computational cost makes it possible to conduct distribution-based GSA of realistic, time-consuming simulations. This is critical to properly understand the influence of cross-correlated uncertain parameters, which are allowed to vary simultaneously.

We demonstrated our approach by computing the Borgonovo importance indices (a sensitivity metric for models correlated inputs) for a regional-scale analysis of sensitivity of potential evapotranspiration to temperature variability at 25 meteorological stations located throughout Emilia-Romagna region of Italy. Using the data collected at these stations over a 30-year period, we estimated the statistical characteristics of the inputs and carried out the distribution-based GSA. The latter requires 750 solves of the model (at

Table A1
Trends in Seasonal and Annual Potential Evapotranspiration During 1971–2000 Identified With the Mann-Kendall Tau Test and the Rate of Increase (in mm/Decade) in These Quantities Averaged Over the 25 Meteorological Stations

	Winter	Spring	Summer	Autumn	Annual
Number of stations:					
Positive (ρ<0.05)	18	23	19	12	21
Positive n.s.	7	2	6	12	4
Negative n.s.	0	0	0	1	0
Negative (ρ<0.05)	0	0	0	0	0
Variation (mm/decade):	3.79	12.71	16.56	5.34	38.41

Note. n.s. = not significant.

each station, for each year), which is significantly fewer than one would need for traditional GSA approaches. Hence, the proposed algorithm extends the range of analysis that can be performed on complex systems and enhances the interpretation of key physical processes.

Appendix A: Data Analysis

This appendix is devoted to the analysis of trends in seasonal and annual evapotranspiration. Figure A1 shows averages over the 25 stations of $ET_{0_i}^{\mathrm{win}}$, $ET_{0_i}^{\mathrm{spr}}$, $ET_{0_i}^{\mathrm{sum}}$, and $ET_{0_i}^{\mathrm{aut}}$ and their corresponding annual values, for years from 1971 to 2000. A positive trend in potential evapotranspiration at the selected locations is visible in each season, especially in summer and spring. This is supported by the results collated in Table A1, which

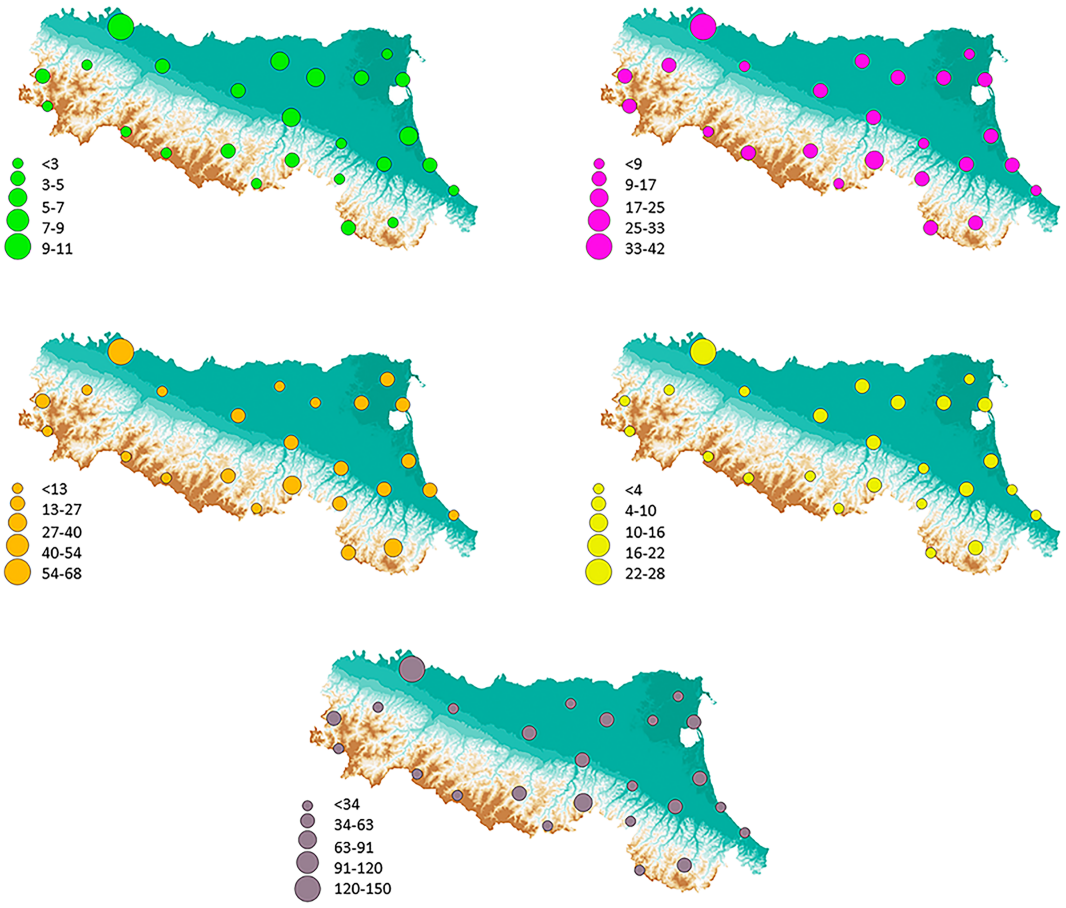

Figure A2. Rate of increase (in mm/decade) in potential evapotranspiration at the selected locations throughout Emilia-Romagna region, from 1971 to 2000.

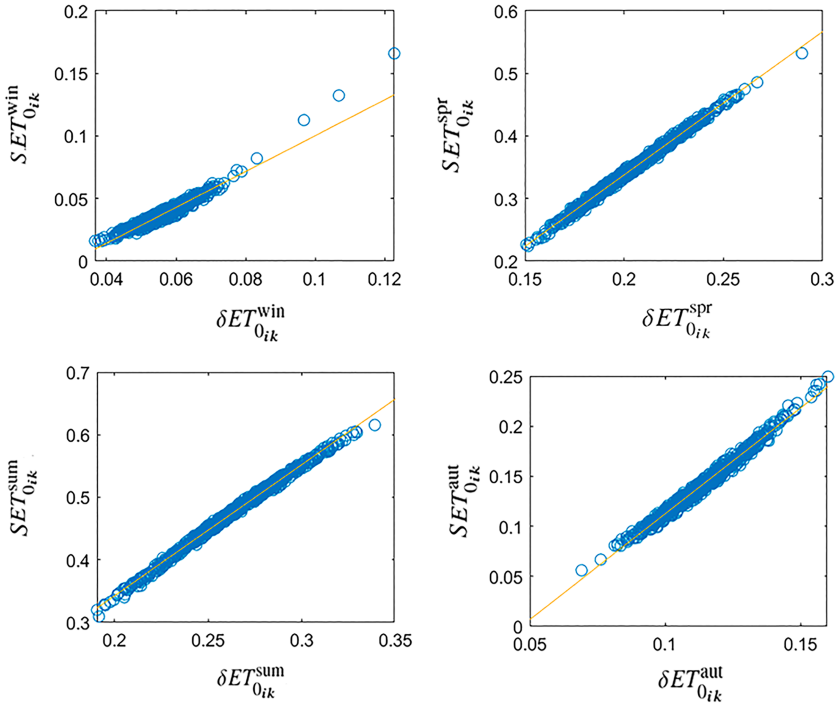

Figure B1. Comparison of values of the distribution-based (Borgonovo) indices δ and the variance-based (principal Sobol') indices S for the four random inputs.

reports the number of stations characterized by a positive/negative trend with the 0.05 confidence level, as determined by the Mann-Kendall tau test. The majority of the stations reveal a clear positive trend (with ρ<0.05). None of the stations shows a negative trend with the same confidence level. Table A1 also presents the rate of increase (in mm/decade) of the seasonal and annual values of potential evapotranspiration averaged over the 25 stations. Figure A2 disaggregates these values in order to show their spatial variability within the region.

Appendix B: Comparison With Variance-Based GSA

This appendix is devoted to the comparison between distribution-based and variance-based GSA results for the same model. The former approach computes the Bogonovo importance indices, while the latter relies on the principal Sobol' indices, S (e.g., Sudret, 2008; Ciriello et al., 2015, and the references therein). As expected, given the linearity of the ET model, Figure B1 shows a good agreement between the variance-based and distribution-based results for the selected case study. It also confirms the order of relevancy among the random inputs.

Acknowledgments
VC acknowledges support from University of Bologna RFO (Ricerca Fondamentale Orientata) 2015-2016 and ALMAIDEA Grant 2017. IL acknowledges support from University of Bologna Marco Polo Grant 2017. DMT was supported in part by U.S. Department of Energy under award number DE-SC0019130. Data used to produce the results of this manuscript are open and available from the corresponding author upon request. There are no data sharing issues since all of the numerical information is provided in the figures produced by solving the equations in the paper.

References

Ashraf, M., Oladyshkin, S., & Nowak, W. (2013). Geological storage of CO2: Application, feasibility and efficiency of global sensitivity analysis and risk assessment using the arbitrary polynomial chaos. *International Journal of Greenhouse Gas Control*, *19*(2), 704–719.

Assouline, S., Ciriello, V., & Tartakovsky, D. M. (2017). Estimation of intrinsic length scales of flow in unsaturated porous media. *Water Resources Research*, *53*, 9980–9987. https://doi.org/10.1002/2017WR021629

Borgonovo, E. (2007). A new uncertainty importance measure. *Reliability Engineering & System Safety*, *92*(6), 771–784.

Borgonovo, E., Lu, X., Plischke, E., Rakovec, O., & Hill, M. C. (2017). Making the most out of a hydrological model data set: Sensitivity analyses to open the model black-box. *Water Resources Research*, *53*, 7933–7950. https://doi.org/10.1002/2017WR020767

Caniou, Y., & Sudret, B. (2011). Distribution-based global sensitivity analysis in case of correlated input parameters using polynomial chaos expansions. In *Proc. 11th Int. Conf. on Applications of Stat. and Prob. in Civil Engineering (ICASP11)*.

Ciriello, V., Di Federico, V., Riva, M., Cadini, F., De Sanctis, J., Zio, E., & Guadagnini, A. (2013). Polynomial chaos expansion for global sensitivity analysis applied to a model of radionuclide migration in a randomly heterogeneous aquifer. *Stochastic Environmental Research and Risk Assessment*, *27*, 945–954.

Ciriello, V., Edery, Y., Guadagnini, A., & Berkowitz, B. (2015). Multimodel framework for characterization of transport in porous media. *Water Resources Research*, *51*, 3384–3402. https://doi.org/10.1002/2015WR017047

Ciriello, V., Guadagnini, A., Di Federico, V., Edery, Y., & Berkowitz, B. (2013). Comparative analysis of formulations for conservative transport in porous media through sensitivity-based parameter calibration. *Water Resources Research*, *49*, 5206–5220. https://doi.org/10.1002/wrcr.20395

Ciriello, V., Lauriola, I., Bonvicini, S., Cozzani, V., Federico, V. D., & Tartakovsky, D. M. (2017). Impact of hydrogeological uncertainty on estimation of environmental risks posed by hydrocarbon transportation networks. *Water Resources Research*, *53*, 8686–8697. https://doi.org/10.1002/2017WR021368

Deman, G., Kerrou, J., Benabderrahmane, H., & Perrochet, P. (2015). Sensitivity analysis of groundwater lifetime expectancy to hydro-dispersive parameters: The case of ANDRA Meuse/Haute-Marne site. *Reliability Engineering & System Safety*, *134*, 276–286.

Di Fusco, E, Lauriola, I., Verdone, R., Di Federico, V., & Ciriello, V. (2018). Impact of uncertainty in soil texture parameters on estimation of soil moisture through radio waves transmission. *Advances in Water Resources*, *122*, 131–138.

Ferretti, F., Saltelli, A., & Tarantola, S. (2016). Trends in sensitivity analysis practice in the last decade. *Science of the Total Environment*, *568*, 666–670.

Ghanem, R. G., & Spanos, P. D. (1991). *Stochastic finite elements: A spectral approach*. Berlin: Springer.

Guo, D., Westra, S., & Maier, H. R. (2017). Sensitivity of potential evapotranspiration to changes in climate variables for different Australian climatic zones. *Hydrology and Earth System Sciences*, *21*(4), 2107–2126.

Hargreaves, G. H., & Samani, Z. A. (1982). Estimating potential evapotranspiration. *Journal of Irrigation and Drainage Engineering*, *108*, 223–230.

Jurado, A., Gaspari, F. D., Vilarrasa, V., Bolster, D., Sanchez-Vila, X., Fernandez-Garcia, D., & Tartakovsky, D. M. (2012). Probabilistic analysis of groundwater-related risks at subsurface excavation sites. *Engineering Geology*, *125*, 35–44. https://doi.org/10.1016/j.enggeo.2011.10.015

Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika*, *30*(1-2), 81–93.

Marrel, A., Perot, N., & Mottet, C. (2015). Development of a surrogate model and sensitivity analysis for spatio-temporal numerical simulators. *Stochastic Environmental Research and Risk Assessment*, *29*(3), 959–974.

Rajabi, M. M., Ataie-Ashtiani, B., & Simmons, C. T. (2015). Polynomial chaos expansions for uncertainty propagation and moment independent sensitivity analysis of seawater intrusion simulations. *Journal of Hydrology*, *520*, 101–122.

Saltelli, A., Tarantola, S., & Campolongo, F. (2000). Sensitivity analysis as an ingredient of modeling. *Statistical Science*, *15*(4), 377–395.

Sobol′, I. M. (1993). Sensitivity estimates for nonlinear mathematical models. *Mathematical Modeling and Computation*, *1*, 407–414.

Sobol′, I. M. (2001). Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematical Modeling and Computation*, *55*, 271–280.

Sudret, B. (2008). Global sensitivity analysis using polynomial chaos expansions. *Reliability Engineering & System Safety*, *93*, 964–979.

Tartakovsky, D. M. (2013). Assessment and management of risk in subsurface hydrology: A review and perspective. *Advances in Water Resources*, *51*, 247–260.

Um, K., Hall, E. J., Katsoulakis, M. A., & Tartakovsky, D. M. (2019). Causality and Bayesian Network PDEs for multiscale representations of porous media. *Journal of Computational Physics*, *394*, 658–678. https://doi.org/10.1016/j.jcp.2019.06.007

Wiener, N. (1938). The homogeneous chaos. *American Journal of Mathematics*, *60*, 897–936.

Xiu, D., & Karniadakis, G. E. (2002). The Wiener-Askey polynomial chaos for stochastic differential equations. *Journal of Scientific Computing*, *24*(2), 619–644.