Statistical Characterization of Environmental Hot Spots and Hot Moments and Applications in Groundwater Hydrology

Jiancong Chen[1], Bhavna Arora[2], Alberto Bellin[3] and Yoram Rubin[1]

[1]Department of Civil and Environmental Engineering, University of California, Berkeley, California, USA
[2]Energy Geosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California, USA
[3]Department of Civil, Environmental and Mechanical Engineering, University of Trento, Italy

Correspondence to: Yoram Rubin (rubin@ce.berkeley.edu)

Abstract

Environmental hot spots and hot moments (HSHMs) represent rare locations and events that exert disproportionate influence over the environment. While several mechanistic models have been used to characterize HSHMs behavior at specific sites, a critical missing component of research on HSHMs has been the development of clear, conventional statistical models. In this paper, we introduced a novel stochastic framework for analyzing HSHMs and the uncertainties. This framework can easily incorporate heterogeneous features in the spatiotemporal domain and can offer inexpensive solutions for testing future scenarios. The proposed approach utilizes indicator random variables (RVs) to construct a statistical model for HSHMs. The HSHMs indicator RVs are comprised of spatial and temporal components, which can be used to represent the unique characteristics of HSHMs. We identified three categories of HSHMs and demonstrated how our statistical framework are adjusted for each category. The three categories are (1) HSHMs defined only by spatial (static) components, (2) HSHMs defined by both spatial and temporal (dynamic) components, and (3) HSHMs defined by multiple dynamic components. The representation of an HSHM through its spatial and temporal components allows researchers to relate the HSHM's uncertainty to the uncertainty of its components. We illustrated the proposed statistical framework through several HSHM case studies covering a variety of surface, subsurface, and coupled systems.

1 Introduction

Environmental hot spots and hot moments (HSHMs) were originally defined as rare locations or events that support or induce disproportionately high activity levels (e.g., chemical reaction rates) compared to surrounding areas or preceding times (McClain et al., 2003). Vidon et al. (2010) further classified HSHMs into either transport-driven or biogeochemically-driven HSHMs, based on the mechanisms causing the HSHMs. Bernhardt et al. (2017) derived the concept of ecological control points (CPs) related to HSHMs, defining CPs as areas of the landscape that exert a disproportionate influence on the biogeochemical behavior of an ecosystem under study. These definitions have mainly focused on HSHMs related to elevated biogeochemical activities triggered by hydrological or biogeochemical processes, or a confluence of both processes. The concept of HSHMs is also used in climate science, where it is related to elevated greenhouse gas emissions or specific locations that are subject to extreme natural hazards (e.g., sea-level rise, floods, hurricanes, or earthquakes) caused by climate change (Arora et al., 2021; Shrestha and Wang, 2018). Further, Henri et al. (2015) related HSHMs to locations experiencing elevated environmental risks and developed the

incremental lifetime cancer risk (ILCR) model to quantify the effects of hot spots on human health. In the present study, we provide a unified treatment of both positive and negative impacts of HSHMs, which allows us to present an integrative analytical framework for understanding and modeling HSHMs in various fields.

40   Various approaches have been developed to better quantify HSHMs dynamics, including numerical modeling, empirical modeling and data-based approaches with statistics. For example, Dwivedi et al. (2017) developed a 3-D high-resolution numerical model to investigate whether organic-carbon-rich and chemically-reduced sediments located within the riparian zone act as denitrification hot spots. Their study demonstrated a significantly higher potential (~70%) of the naturally reduced zones (NRZs) to remove nitrate than the non-NRZ locations. Arora

45   et al. (2016b) used a 2-D transect model and showed that temperature fluctuations constituted carbon hot moments in a contaminated floodplain aquifer that resulted in a 170% increase in annual groundwater carbon fluxes. Gu et al. (2012) developed a Monte Carlo reactive transport approach and discovered how denitrification HSHMs are triggered by river stage fluctuations. Abbott et al. (2016) developed the HotDam framework that combines the HSHM concept and Darmköhler number with multiple tracers to advance our understanding of ecohydrology. Statistical concepts

50   have also been used to identify HSHMs through simple comparison to average; substantial percentage of total flux; outlier in distribution of data; statistically significant difference between or among landscape elements and contribution to flux/total area or time (Bernhardt et al. (2017) and references therein). Wavelet and entropy-based approaches have also been used to identify non-uniform regions and times and consequently HSHMs (Arora et al., 2013, 2019a). However, most of these quantitative methods are derived based on site-specific data, which severely

55   limits the transferability of these approaches. In contrast, a unified HSHM approach offers multiple advantages. First, a unified strategy based on commonly-used parameters for a given HSHM would allow modelers to create probability priors that could be used for prediction of said HSHM at unsampled- or poorly-sampled sites (Li et al., 2018). Second, such a standardized approach for modeling HSHMs could be beneficial to developing and implementing monitoring standards and regulations for environmentally-sensitive HSHMs. Last, but not the least, a unified approach can be

60   used together with mechanistic models to capture uncertainty and heterogeneity for HSHMs in environmentally-relevant applications.

        Successful characterization of HSHMs through deterministic physically-based models or purely statistical approaches relies on experts' knowledge of a site, intensive field characterization, and possibly continuous field sampling to provide the data to develop and validate these approaches. Understandably, intensive site characterization

65   and long-term sampling can be quite challenging due to the associated costs and efforts. In this regard, having access to a stochastic approach that could improve predictions through built-in model updating (i.e., Bayesian) capabilities could prove to be an advantage.

        Stochastic concepts and models have been widely applied in hydrology and hydrogeology for addressing situations subject to uncertainty, including but not limited to modeling flow and contaminant transport, quantifying

70   subsurface heterogeneity and the associated uncertainties, developing strategies for site characterization, and providing informative priors for ungauged watersheds. Bayesian approaches were found particularly useful, especially through concepts such as conditioning and updating. In this paper, we aim to bring the experience gained in hydrology and hydrogeology into HSHM modeling.

An important characteristic of HSHMs is that they occupy a limited portion of the investigated domain and may be active for a limited amount of time since they are activated when the control variable exceeds a given threshold. Physical and geochemical heterogeneities, and the impossibility to fully characterize them, render the deterministic identification of HSHMs a vanishing objective. To address this hurdle, we propose to cast the problem into a probabilistic framework by seeking the probability of HSHM occurrence at a given position and time. For a given time and/or space intervals and for a-priori specified criteria, an HSHM occurrence could be viewed as a binary variable where the ensemble mean is the probability of occurrence. Indicator statistics have previously been applied to model flow and transport phenomena in groundwater (Rubin and Journel, 1991), where indicators were used to model the spatial distribution in a sand-shale formation. Wilson and Rubin (2002) and Bellin and Rubin (2004) used indicator statistics to characterize aquifer heterogeneity. These studies suggest that representation of a system's structure through indicator formulation holds the potential of taking informed decisions, for example concerning remediation actions, under incomplete site characterization.

Based on the mechanisms that trigger HSHMs, we identified three categories of HSHMs: (1) those triggered only by spatial (static) contributors, (2) those triggered by both spatial (static) and temporal (dynamic) contributors, and (3) those triggered by multiple dynamic contributors. Applications of the proposed indicator formulation to a diverse range of HSHM situations are presented to illustrate the generality of our proposed approach. The remainder of the paper is structured as follows. Section 2 outlines the proposed statistical framework for predicting HSHMs. In section 3, various reported cases from previous HSHM studies are presented using the framework of our proposed approach, intended to demonstrate its generality. In section 4, we present an HSHM application in groundwater hydrology and show how the HSHM uncertainty relates to the spatial variability of the hydraulic conductivity. Advantages and limitations of our approach are discussed in section 5.

2 Methodology and statistical formulation of HSHMs

Herein, we present a probabilistic formulation of hot spots and hot moments, which considers the HSHM occurrence as a binary event, expressed through indicator statistics embedded with the HSHM underlying physics. Section 2.1 summarizes the indicator formulation of HSHMs. Based on the contributors to HSHMs, we classified HSHMs into three different types, and we demonstrate how indicators are constructed for each type of HSHM in sections 2.2, 2.3 and 2.4, respectively. Section 2.5 focuses on the linkages between indicators and Bayesian concepts. Case studies for each class of HSHMs are provided in Section 3.

2.1 Indicator formulation of HSHMs

HSHMs represent intervals in space and/or time characterized by hydrobiogeochemical activity rates or fluxes that differ significantly from the background conditions, thus exerting disproportionate influences over an ecosystem's dynamics. We define $\boldsymbol{\Omega}^*$ as the volume (subdomain) within which the hot spot may be located and t^* the time at which the hot moment occurs. An indicator random variable, $I_{HSHM}(\boldsymbol{\Omega}^*, t^*)$, is used to identify whether HSHM occurs at $(\boldsymbol{\Omega}^*, t^*)$ or not. If user-defined critical conditions needed to trigger an HSHM are met at $(\boldsymbol{\Omega}^*, t^*)$, then $I_{HSHM}(\boldsymbol{\Omega}^*, t^*) = 1$, and it is equal to zero otherwise. What makes the indicator a random variable is the uncertainty in the spatial and temporal distribution of the quantities triggering the HSHM event in real-life applications. Following

110 the original definition by McClain et al. (2003), in our method, $I_{HSHM}(\boldsymbol{\Omega}^*, t^*)$ can take the value of 0 or 1, depending on whether suitable thresholds are exceeded or not as follows:

$$I_{HSHM}(\boldsymbol{\Omega}^*, t^*) = \begin{cases} 1, if\ C(\boldsymbol{x}, t^*) > C_{th};\ \boldsymbol{x} \subseteq \boldsymbol{\Omega}^* \\ 0, \qquad\quad otherwise \end{cases}, or$$

$$I_{HSHM}(\boldsymbol{\Omega}^*, t^*) = \begin{cases} 1, if\ R(\boldsymbol{x}, t^*) > R_{th};\ \boldsymbol{x} \subseteq \boldsymbol{\Omega}^* \\ 0, \qquad\quad otherwise \end{cases}, \qquad\qquad (1)$$

where $C(\boldsymbol{x}, t^*)\ and\ R(\boldsymbol{x}, t^*)$ are the concentration and reaction rate at the position $\boldsymbol{x}$ and time t^*, respectively.

115 $C_{th}\ and\ R_{th}$ represent the concentration and reaction rate thresholds, respectively, which identify whether HSHM is triggered or not. Defining indicators with concentration or reaction rate depends on the target of HSHM. The threshold values can also be based on regulatory limits or defined by the user.

    The critical values of C_{th} and/or R_{th} are keys to an effective application of the above framework and should be determined based on the specific scenario under investigation. For example, in the case of contaminants that are

120 associated with significant environmental or health risks (e.g., nuclear waste or a cancerous substance), $C_{th} = 0$ or $R_{th} = 0$ can be used so that the HSHM will be triggered as soon as there is the presence of such contaminants and relevant chemical reactions. As an alternative, a limit in the total accumulated mass or fluxes within hot spots may also be set, such as suggested by EPA (USEPA, 2001), but in this case the definition (1) of the indicators should be modified. For water quality parameters, $C_{th} = MCL$ or $R_{th} = R^*$ can be assigned, where MCL represents the maximum

125 concentration limit for a specific solute whereas R^* could represent a critical reaction rate. The critical thresholds can be determined based on statistics, such as percentiles and extremes as defined by regulations or analytical studies. Alternatively, C_{th} and R_{th} could also be chosen based on the experts' domain knowledge or from well-documented studies at similar environments. Through the flexibility to adopt different choices for activation thresholds, our approach could allow users to compare relevant indicator models and assess their applicability by testing how different

130 thresholds would influence the probability of the HSHM to occur and assessing said probabilities against risk tolerance and/or regulations.

    Following the definition of $I_{HSHM}(\boldsymbol{\Omega}^*, t^*)$ as a binary random variable (Eq. 1), we propose to model it with a Bernoulli distribution, such as $I_{HSHM}(\boldsymbol{\Omega}^*, t^*) \sim Bernoulli(< I_{HSHM}(\boldsymbol{\Omega}^*, t^*) >)$, where $<.>$ is the ensemble averaging operator. An important characteristic of the Bernoulli distribution is that all the statistical moments of

135 $I_{HSHM}(\boldsymbol{\Omega}^*, t^*)$ can be expressed as a function of the ensemble mean $< I_{HSHM}(\boldsymbol{\Omega}^*, t^*) >$. For example, the variance is given by $var\big(I_{HSHM}(\boldsymbol{\Omega}^*, t^*)\big) = < I_{HSHM}(\boldsymbol{\Omega}^*, t^*) > \cdot (1 - < I_{HSHM}(\boldsymbol{\Omega}^*, t^*) >)$. Thus, being able to fold the HSHM physics into an indicator formulation, a simplified approach is presented through Eq. (1).

    Case-based formulation of the Bernoulli distribution of $I_{HSHM}(\boldsymbol{\Omega}^*, t^*)$ requires the incorporation of the mechanisms that govern the development and occurrence of HSHMs into the indicator model. To facilitate this

140 undertaking, we propose to decompose $I_{HSHM}(\boldsymbol{\Omega}^*, t^*)$ into a Type-A (static) indicator random variable, $I_s(\boldsymbol{\Omega}^*)$, and a Type-B (dynamic) indicator random variable, $I_d(\boldsymbol{\Omega}^*, t^*)$. Definitions of the Type-A and Type-B indicators are provided herein:

- **Type-A (Static) Contributors.** This category covers discrete spatial elements (and their associated critical states) that could trigger an HSHM once they come into contact with Type-B contributors (see discussion

145   below). Critical states are the range of values needed to trigger an HSHM (either in standalone mode or when coupled with Type-B contributors).

- **Type-B (Dynamic) Contributors**. This category covers dynamic variables (and their associated critical states) that could trigger an HSHM once they come into contact with Type-A contributors. This category includes, for example, mass transport variables. It also includes changes in local hydrological and
150   environmental conditions (e.g., water table fluctuations). The displacement of solutes in the subsurface (trajectories and travel times) from below- and above-ground processes are prime examples of Type-B contributors.

As an example, naturally reduced sediments (Type-A contributor) occurring next to the river corridor at the Rifle site were identified as carbon export hot spots (Arora et al., 2016a; Wainwright et al., 2015). Studies showed
155   that these hot spots were triggered when temperature conditions (Type-B contributor) varied in the subsurface, resulting in a 170% increase in groundwater carbon export from the floodplain site to the river (Arora et al., 2016b). In another example, topographic features, such as the backslope of the lower montane hillslope (Type-A contributor) within the East River Watershed (Hubbard et al., 2018), were considered denitrification hot spots, which can have a significant impact on the watershed-scale nitrogen loss pathway. These hot spots were often triggered by spring
160   snowmelt and storm events (Type-B contributor).

Both indicators of the type-A and type-B contributors assume a value of either 0 or 1. If one of these indicators takes a value of 1, it can be viewed as an HSHM contributor. However, for an HSHM to occur, both indicators must have a value of 1 at the same location and time. This idea can be expressed as follows:

$$P(I_{HSHM}(\boldsymbol{\Omega}^*, t^*) = 1) = P(I_s(\boldsymbol{\Omega}^*) = 1, I_d(\boldsymbol{\Omega}^*, t^*) = 1)$$

165
$$= P(I_s(\boldsymbol{\Omega}^*) = 1) \cdot P(I_d(\boldsymbol{\Omega}^*, t^*) = 1 | I_s(\boldsymbol{\Omega}^*) = 1)$$

$$= P(I_d(\boldsymbol{\Omega}^*, t^*) = 1) \cdot P(I_s(\boldsymbol{\Omega}^*) = 1 | I_d(\boldsymbol{\Omega}^*, t^*) = 1). \tag{2}$$

Based on the mechanisms of HSHMs, we can distinguish between three different HSHM categories as discussed below. These categories can be used to guide the application of the above statistical framework in a variety of complex HSHM scenarios, and they can also be used to develop analytical or numerical solutions for both type-A
170   and type-B contributors.
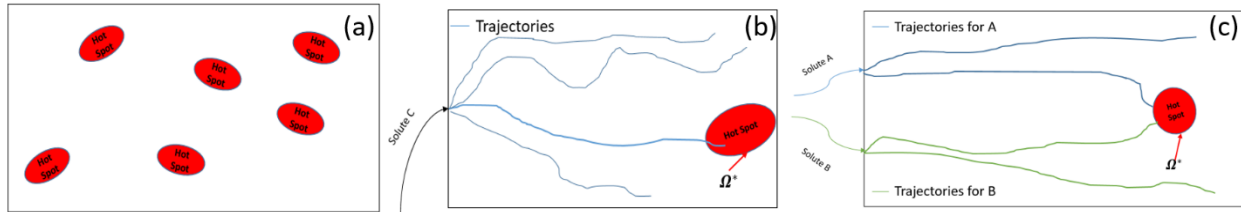
Figure 1. Identified categories of HSHMs. Panel (a) presents HSHMs resulting from Type-A (static) indicator only; panel (b) presents HSHMs resulting from coupled action (static + dynamic) and panel (c) presents HSHMs resulting from multiple (two) dynamic indicators.

2.2 HSHMs induced by type-A (static) indicators

In this section, we consider HSHMs that are defined by static indicators only (Figure 1a). This list can include zones of high, persistent concentration and reactivity that are due to the subsurface or the ecosystem's unique hydrological and biogeochemical properties. For example, the accumulation of contaminants in the subsurface (e.g., the high nuclide concentration in the subsurface at the Hanford site) could lead to the evolution of persistent, high reactivity zones. An aquifer's reactivity is another example that could distinguish certain regions with high reactivity compared to surrounding areas (Loschko et al., 2016). Such high reactivity spots (hereafter denoted as $\boldsymbol{\Omega}^*$) can be characterized by static indicator RVs due to the persistence of high concentration or reactivity. The static indicators are defined as follows:

$$I_{HSHM}(\boldsymbol{\Omega}^*) = I_s(\boldsymbol{\Omega}^*) = \begin{cases} 1, & if\ Z(\boldsymbol{\Omega}^*) \subseteq Z_s^* \\ 0, & otherwise \end{cases}, \tag{3}$$

where Z_s^* represents the conditions needed to trigger a hot spot at $\boldsymbol{\Omega}^*$, and $Z(\boldsymbol{\Omega}^*)$ represents the corresponding local conditions at $\boldsymbol{\Omega}^*$. Notice that $\boldsymbol{\Omega}^*$ is a volume centered at a selected position of the domain where the probability of developing HSHM is evaluated.

2.3 HSHMs induced by type-A (static) and type-B (dynamic) indicators

HSHMs can also result from dynamic processes encountering specific local conditions at $\boldsymbol{\Omega}^*$ (Figure 1b). This is the situation described by Eq. (2), where the type-A indicators are determined first, and then used jointly with the type-B indicators for complete HSHM characterization. For example, Bundt et al. (2001) concluded that locations ($\boldsymbol{\Omega}^*$) interested by preferential flow paths are possible biological hot spots for soil microbial activities. Meanwhile, dynamic factors, such as snowmelt or rainfall infiltration control contaminant transport via the preferential flow paths and thus, constitute the hot moment component. Additional case studies are presented in Section 3.

For HSHMs induced by both type-A and type-B indicators, the static locations are selected first, based on their HSHM-related properties. After this, we can focus on characterizing the HSHM dynamics as they relate to the relevant locations. A selected location, $\boldsymbol{\Omega}^*$, could become an HSHM site based on characteristics defined through the following type-A and type-B indicators, respectively:

$$I_s(\boldsymbol{\Omega}^*) = \begin{cases} 1, & if\ Z_s(\boldsymbol{\Omega}^*) \subseteq Z_s^* \\ 0, & otherwise \end{cases}, \tag{4}$$

$$I_d(\boldsymbol{\Omega}^*, t^*) = \begin{cases} 1, & if\ Z_d(\boldsymbol{\Omega}^*, t^*) \subseteq Z_d^* \\ 0, & otherwise \end{cases}, \tag{5}$$

$$I_{HSHM}(\boldsymbol{\Omega}^*, t^*) = \begin{cases} 1, & if\ Z_s(\boldsymbol{\Omega}^*) \subseteq Z_s^*, and\ Z_d(\boldsymbol{\Omega}^*, t^*) \subseteq Z_d^* \\ 0, & otherwise \end{cases}, \tag{6}$$

where Z_d^* represents the critical conditions needed to trigger the hot moment, whereas $Z_d(\boldsymbol{\Omega}^*, t^*)$ represents the corresponding, critical-state local conditions at t^* and $\boldsymbol{\Omega}^*$. The statistical model of $I_{HSHM}(\boldsymbol{\Omega}^*, t^*)$ can be expressed using the statistical models of I_s and I_d, as shown in Eq. (2).

2.4 HSHMs induced by multiple type-B (dynamic) indicators

A confluence of dynamic processes could result in the formation of a HSHM (Figure 1c). Unlike the previous scenarios where static locations can be determined through known characteristics provided by geophysical or other

types of data, HSHMs can also emerge due to a confluence of dynamic processes. For example, Gu et al. (2012) analyzed how streamflow fluctuations could trigger a nitrogen HSHM. In their example, the dynamics of the streamflow and groundwater controlled the transport and mixing of the chemical reactants, thus triggering the occurrences of HSHMs. For this case, the static locations of $\boldsymbol{\Omega}^*$ are determined by the confluence of multiple dynamic processes, not being restricted by a set of local conditions. In this case, only type-B indicators need to be modeled.

We can consider the case where an HSHM is predicated on m dynamic processes, d_j, where $I_{d,j}(\boldsymbol{\Omega}^*, t^*)$ is the dynamic for each dynamic d_j at $\boldsymbol{\Omega}^*$ and time t^*. The hot spot location $\boldsymbol{\Omega}^*$ is determined by the confluence of all dynamic processes at time t^*. These dynamic processes are not necessarily independent. Therefore, generally, the statistical model for the comprehensive dynamic indicator (which covers all dynamic contributors) assumes the following form:

$$P[I_d(\boldsymbol{\Omega}^*, t^*) = 1] = P[I_{d,1}(\boldsymbol{\Omega}^*, t^*) = 1, \dots, I_{d,m}(\boldsymbol{\Omega}^*, t^*) = 1]. \tag{7}$$

In situations where the various dynamic contributors can be viewed as independent (e.g., Destouni and Cvetkovic, 1991)—i.e., where the reactants travel via different paths—then, assuming independence, we can state that

$$P(I_d(\boldsymbol{\Omega}^*, t^*) = 1] = \prod_{j=1}^{m} P[I_{d,j}(\boldsymbol{\Omega}^*, t^*) = 1]. \tag{8}$$

Here, the mean of the dynamic indicator becomes

$$< I_d(\boldsymbol{\Omega}^*, t^*) > = \prod_{j=1}^{m} < I_{d,j}(\boldsymbol{\Omega}^*, t^*) >. \tag{9}$$

If $\boldsymbol{\Omega}^*$ is a hot spot, then Eq. (9) also defines $< I_{HSHM}(\boldsymbol{\Omega}^*, t^*) >$. However, if $\boldsymbol{\Omega}^*$ is not a hot spot, then we need to resort to coupled statistical modeling, as suggested by Eq. (2).

2.5 Additional advantages of stochastic formalism

The statistical framework provides several benefits. Unified formulations of HSHMs through indicators provide us a platform to evaluate alternative HSHM models thoroughly and objectively. For example, the Akaike information criteria (AIC, Akaike, 1974) and Bayesian information criteria (Schwarz, 1978) can be used to rank between alternative indicator formulations and evaluate their ability to explain HSHM observations. Smaller AIC and BIC values indicate more information preserved in a given indicator HSHM model and implies better model quality than other indicator models. On the other hand, if larger AIC and BIC values are observed, important processes for HSHMs are likely missing indicating the necessity of increasing site characterization and refinement of conceptual models.

In addition, informative priors constructed from similar HSHM sites (Cucchi et al., 2019; Li et al., 2018) could advance early stage planning for HSHM site investigation. Knowledge from studies at similar HSHM sites can be summarized into prior distributions, which can account for variabilities within and between sites. For example, Cucchi et al. (2019) demonstrated how the distribution of hydraulic parameters at unknown target sites can be predicted using information from hydrologically similar sites with existing tool packages such as exPrior. Goal-oriented site characterization also becomes feasible with informative priors; for example, Li et al. (2018) demonstrated the usefulness of informative priors in reducing model uncertainty and potential risks for estimating groundwater

drawdown at Mintang tunnel in China. Therefore, through integration with statistical concepts, unified formulations of HSHMs enable us to integrate Bayesian concepts to obtain combined and less risky estimations of HSHMs at new sites, which can help us gain better understanding of the underlying mechanism.

3 Examples of the statistical formulation of HSHMs with case studies

In this section, we selected numerous examples from published research to present how our approach can be used to derive statistical representations for the HSHMs investigated in these studies. We grouped these studies into three categories based on the similarities of their underlying HSHM mechanisms, as described in section 2. Section 3.1 demonstrate the formulation of static only HSHM; section 3.2 present the case with static and dynamic triggered HSHMs and section 3.3 summarize the steps to construct multiple dynamic indicators for HSHMs. Table 1 presents a summary of these cases, where environmental risk levels as well as impacts on ecosystem were also included.

Reference	HS Location	Category	Seasonality	Environmental Risk	Causes	Impact	Static Mechanism	Dynamic Mechanism	HSHM Action	Metrics for threshold	Equation(s)
Examples of static only mechanisms											
Wainwright et al. (2015)	Naturally reducing zone	Subsurface	--	Short-term low risk; long-term high risk	Anthropogenic + Natural	Negative	Mineralogical and lithological differences	--	Vanadium, uranium, metallic minerals	Concentration	(3)
Sassen et al. (2012)	Reactive facies	Subsurface	--	Short-term low risk; long-term high risk	Anthropogenic + Natural	Negative	Lithological differences	--	Uranium and other isotopes	Concentration	(3)
Examples of static + dynamic mechanism											
Andrews et al. (2011)	Shale hill	Subsurface + Surface	Snowmelt and fall flushing periods	Low risk	Natural	Neutral	South-facing concave hillslopes	Snowmelt and fall flushing periods	Organic carbon	Concentration	(4) – (6)
Henri et al. (2015)	Preferential flow path	Subsurface	--	High risk	Anthropogenic	Negative	Subsurface heterogeneity	Contaminant transport and travel time distribution	Chlorinated compounds	Concentration	(4) – (6)
Duncan et al. (2013)	Microtopography	Surface	Unimportant	High risk	Natural	Positive	Riparian hollows	Transport and retention of reactants	Nitrogen	Concentration or reaction rate	(4) – (6)
Arora et al. (2016)	Naturally reducing zone-induced transport	Subsurface	Temperature and water table fluctuation	Low risk	Anthropogenic + Natural	Neutral	Naturally reduced zones	Temperature and water table fluctuation	Carbon fluxes	Concentration or reaction rate	(4) – (6)
Examples of multiple dynamic mechanisms											
Hill et al. (2000)	Riparian zone	Subsurface	--	High risk	Natural	Positive	Interfaces in the riparian zone	Supply of electron donor and acceptor from flow transport	Nitrogen and carbon	Concentration or reaction rate	(7) – (9)
Mitchell et al. (2008)	Peatlands	Subsurface + Surface	Summer periods	High risk	Natural	Negative	Upland-peatland interfaces induced by flow	Interactions between upland and peatland flow	Methylmercury	Concentration	(7) – (9)
Frei et al. (2012)	Microtopography	Surface	--	Neutral	Natural	Neutral	Flowpaths induced by microtopography	Biogeochemical evolution along flow paths	Organic matter and nitrogen	Concentration or reaction rate	(7) – (9)
Gu et al. (2012)	Mixing zones	Subsurface + Surface	River discharge + Water table fluctuation	High risk	Natural	Positive	Mixing zones caused by river stages	Interaction between surface water and groundwater	Nitrogen	Concentration or reaction rate	(7) – (9)

Table 1. Example cases considered in this study for constructing the statistical formulation of HSHM.

3.1 HSHMs triggered by static contributors only

255   In this section, we use Wainwright et al. (2015) as an example to illustrate our process to construct $I_{HSHM}(\boldsymbol{\Omega}^*, t^*)$ following Eq. (3), where an HSHM is triggered by static contributors only (section 2.1). NRZs within floodplain environments at the Rifle site are considered biogeochemical hot spots because they represent elevated concentrations of uranium, organic matter, and geochemically reduced minerals and they have been found to contribute to significant carbon fluxes to the atmosphere and to local rivers (Arora et al., 2016). Due to its

260   characteristics, we considered the spatial distribution of an NRZ to be a static-mechanism-based hot spot. Wainwright et al. (2015) used geophysical data (e.g., induced polarization) to map the distribution of an NRZ at the subsurface level. They found that the phase shift (ϕ) from the induced polarization data of the NRZ was within $[4.5, 5]mrad$, compared to non-NRZ locations at $\phi \subseteq [1, 3.5]mrad$. Thus, ϕ can be used to construct the static indicator with a critical condition of $[4.5, 5]mrad$. Therefore,

265
$$I_s(\boldsymbol{\Omega}^*) = \begin{cases} 1, & if \ Z_\phi(\boldsymbol{\Omega}^*) \subseteq [4.5, 5] \ mrad \\ 0, & otherwise \end{cases}. \tag{10}$$

Other static attributes, including but not limited to elevation, hydraulic conductivity, and resistivity, can also be used to define the critical conditions to construct the static indicator for hot spots through Bayesian conditioning.

3.2 HSHMs occurring when dynamic contributors coincide at locations defined by static contributors

The second case we present here utilizes Eq. (4)–(6), where HSHMs are triggered when dynamic contributors

270   coincide at hot spots determined by static contributors. Here, we present the case investigated by Duncan et al. (2013), where riparian hollows representing less than 1% of the total catchment area contributed to more than 99% of the total denitrification within the watershed. In their study, the denitrification rates peaked during the base flow (midsummer) period, when the riparian hollows were partially oxygenated and the hydrologic fluxes were at a minimum. The site was considered to have low inorganic N availability, and thus, nitrate was supplied via nitrification. The highest rates

275   of denitrification were therefore tied to nitrification and the partially aerated conditions.

The static indicator needs to be constructed based on the micro-topographical features within the riparian zone. Specifically, the topographic wetness index (TWI) (Beven and Kirkby, 1979; Sørensen et al., 2006) was used in Duncan et al. (2013) to delineate the riparian hollows from other riparian locations. Terrain analysis indicated a TWI threshold value of 6.0 and 8.0 for riparian hollows under wet and dry conditions, respectively, whereas 4.8 and smaller

280   TWI values corresponded to other riparian locations (e.g., hummocks). Thus, the static indicator can be constructed using the TWI values within the riparian zone to determine the hot spot locations—the hollows. Hence,

$$I_s(\boldsymbol{\Omega}^*) = \begin{cases} 1, & if \ Z_{TWI}(\boldsymbol{\Omega}^*) > 6 \ (wet \ condition) \ or \ 8 \ (dry \ condition) \\ 0, & otherwise \end{cases}. \tag{11}$$

Multiple dynamic processes control the denitrification rate at the riparian hollows. As examined by Duncan et al. (2013), a partially aerated condition ($C_{O_2} > 5\%$) is needed to support nitrification, which supplies the nitrate for

285   denitrification. As quiescent, non-storm periods during base flow favor the coupled nitrification-denitrification mechanism, this is another key process that needs to be represented by a dynamic indicator. Although Duncan et al.

(2013) did not mention specific concentration ranges for nitrogen species, the major components, such as organic N, should be available. Therefore, we can construct the dynamic indicators as follows:

$$P[I_d(\boldsymbol{\Omega^*}, t^*) = 1] = P[I_{d,O_2}(\boldsymbol{\Omega^*}, t^*) = 1, I_{d,Hydro}(\boldsymbol{\Omega^*}, t^*) = 1, I_{d,N}(\boldsymbol{\Omega^*}, t^*) = 1], \qquad (12)$$

where $I_{d,Hydro}(\boldsymbol{\Omega^*}, t^*)$ is the dynamic indicator representing the streamflow stages; this will be 1 if the base flow conditions are met. Additionally, here, $I_{d,N}(\boldsymbol{\Omega^*}, t^*)$ is the dynamic indicator for the transport of the nitrogen species in the subsurface that support the coupled nitrification-denitrification mechanism.

$$I_{d,O_2}(\boldsymbol{\Omega^*}, t^*) = \begin{cases} 1, & if\ C_{O_2}(\boldsymbol{\Omega^*}, t^*) > 5\% \\ 0, & otherwise \end{cases},$$

$$I_{d,Hydro}(\boldsymbol{\Omega^*}, t^*) = \begin{cases} 1, & if\ t^* \subseteq base\ flow\ periods \\ 0, & otherwise \end{cases}, \qquad (13)$$

$$I_{d,N}(\boldsymbol{\Omega^*}, t^*) = \begin{cases} 1, & if\ C_N(\boldsymbol{\Omega^*}, t^*) > 0 \\ 0, & otherwise \end{cases}.$$

It is noted that these dynamic processes are not statistically independent. Usually, when one condition is met (e.g., base flow conditions), other conditions may consistently be satisfied (e.g., the transport of nitrogen in riparian hollows). Alternatively, numerical modeling approaches can be used to construct the dynamic indicators based on the critical conditions at riparian hollows ($\boldsymbol{\Omega^*}$), where we could directly target N_2 fluxes using a Monte Carlo approach. The statistical formulation used here is constructed specifically for the mechanisms described by Duncan et al. (2013). Thus, the detailed threshold limits could change under other denitrification HSHMs cases, such as the case presented in Hill et al. (2000), who focus on desert landscapes, or the one by Harms and Grimm (2008), where the monsoon season is influential for the nitrogen transport. Nonetheless, the general formulation of HSHMs using indicators is still applicable.

3.3 HSHMs occurring when multiple dynamic processes converge in space

HSHMs can also be triggered by the confluence of multiple dynamic processes that lead to the convergence of complementary reactants at $\boldsymbol{\Omega^*}$. Complementary reactants can be mobilized and transported via different hydrologic flowpaths. They can converge at hot spot locations and trigger hot moments during the mixing. Following the statistical framework developed in this study, Eqs. (7) to (9) are suitable for this condition. In order to illustrate how the dynamic indicators are constructed, we consider here the case reported by Gu et al. (2012), where high biogeochemical activity was observed at the interface of groundwater and surface water during the stream stage fluctuations, which resulted in significant in-stream denitrification and NO_3^- removal.

In their study, hot spots form around the near-stream riparian subsurface during river stage fluctuations, where active biogeochemical reaction (e.g., denitrification) requires both O_2 depletion and the simultaneous presence of NO_3^- and the dissolved organic carbon (DOC). Specifically, the spatiotemporal distribution of denitrification hot spots coincides with an O_2 depletion zone along the DOC infiltration flowpaths. In order to determine the mixing of groundwater and surface water during stage fluctuations, Gu et al. (2012) defined bank storage volume $V(t)$ and maximum bank storage volume V_{max}. The flood hydrograph was subdivided into the rising limbs, recession limbs and return flow, the latter representing the slow restitution of part of the water that infiltrated during the previous stages. Considering the different dynamics of these components, they observed that the largest infiltration rate occurred prior

to the maximum stage rise, while $V_{max} = 5m^3m^{-1}$ (critical condition) occurred in the recession limb of the flood event. Instead, maximum return flow occurred toward the end of the recession curve before stream hydrograph stabilizes. Maximum NO_3^- rate removal occurred when return flow phase was almost complete and then decreased until the depletion of NO_3^-. Through statistical analysis, they found that V_{max}, viewed as an integrated index for hydrological exchange, could explain 64% of the variation in the NO_3^- removal. Thus, V_{max} can be used as the critical state to determine whether or not the hyporheic dynamics is significant to enhance relevant biogeochemical processes. In order for the hot moments to be significant, the stream-riparian zone should also be microbially active. Based on these conditions, the dynamic indicators can be constructed as follows:

$$P[I_d(\boldsymbol{\Omega}^*, t^*) = 1] = P\left[I_{d,Hydro}(\boldsymbol{\Omega}^*, t^*) = 1, I_{d,Chem}(\boldsymbol{\Omega}^*, t^*) = 1 \right], \tag{14}$$

where $I_{d,Hydro}(\boldsymbol{\Omega}^*, t^*)$ represents the dynamic process induced by the hydrologic conditions (e.g., stage fluctuation), and $I_{d,Chem}(\boldsymbol{\Omega}^*, t^*)$ represents the dynamic process controlled by the transport and accumulation of chemical reactants. Based on the critical values or ranges, we formulate the indicators as follows:

$$I_{d,Hydro}(\boldsymbol{\Omega}^*, t^*) = \begin{cases} 1, & Z_{V_{max}}(\boldsymbol{\Omega}^*, t^*) \geq 5m^3m^{-1} \\ 0, & otherwise \end{cases},$$

$$I_{d,Chem}(\boldsymbol{\Omega}^*, t^*) = \begin{cases} 1, & if\ C_{O_2}(\boldsymbol{\Omega}^*, t^*)\ is\ small\ and\ C_{NO_3^-}(\boldsymbol{\Omega}^*, t^*) > 0\ and\ C_{DOC}(\boldsymbol{\Omega}^*, t^*) > 0 \\ 0, & otherwise \end{cases}. \tag{15}$$

Typically, because of the complexity of the processes, no analytical solutions are available for formulating the indicators. However, Monte Carlo simulations can be useful in constructing such indicators. For this case, an HSHM at any given location and time $(\boldsymbol{\Omega}^*, t^*)$ will only be triggered when all of the conditions are met and the ensemble mean of the indicator assumes the following form:

$$< I_d(\boldsymbol{\Omega}^*, t^*) > = \frac{1}{N}\sum_{i=1}^{N} I_{d,i}(\boldsymbol{\Omega}^*, t^*), \tag{16}$$

where $I_{d,i}(\boldsymbol{\Omega}^*, t^*)$ is the value that the indicator assumes in the ith realization and N is the total number of realizations.

Overall, our choices of the three studies should not limit the generalizability of the indicator statistics approach for deriving statistical formulations for HSHM applications. The critical conditions chosen to construct the indicators are determined solely on the findings from these selected studies, and they will vary under different scenarios.

4 HSHM applications in groundwater hydrology

This section focuses on HSHMs in the subsurface for demonstration of linking HSHM models with the contributing physical processes, such as the migration of groundwater carrying reducing substrates, nuclear waste transport within the subsurface, the accumulation and transport of dense non-aqueous phase liquid (DNAPL) and other biogeochemical processes. Some current modeling approaches that focus on subsurface HSHMs assume simplified hydrologic structures (e.g., homogeneous and isotropic domains) in quantifying the fate and transport of solutes in the subsurface. However, such assumptions neglect the effect of the heterogeneity in the subsurface, potentially missing localized HSHMs arising as the combined effect of heterogeneity in physical and geochemical properties, and do not allow to assess uncertainties in the HSHM occurrences.

This section therefore focuses on HSHMs taking place in the subsurface, with a particular emphasis on the
355   role of spatial variability of the hydrologic parameters. Section 4.1 illustrates the potential of subsurface heterogeneity
for triggering and timing of HSHMs. In section 4.2, we develop closed-form analytical solutions for HSHM
probability. In doing this, we demonstrate the linking between our indicator model and the physics of the HSHMs in
the subsurface. In section 4.3 we demonstrate applications under various conditions of spatial variability.

4.1 Importance of spatial variability in the subsurface

360   The heterogeneous structure of hydraulic conductivity leads to significant variability in the transport of
solutes in the subsurface, which couples with heterogeneous geochemical properties leading to a spatially varying
reactivity (Arora et al., 2019b; Loschko et al., 2016; Sassen et al., 2012; Wainwright et al., 2015). Figure 2
demonstrates the uncertainty associated with HSHMs by looking at the flow fields in two-dimensional log-hydraulic
conductivity ($Y = ln(K)$) fields with streamlines resulting from a uniform mean head gradient, left to right. The three
365   panels differ in terms of the variance, σ_Y^2, of the log-conductivity. The covariance function used for generating the
fields is exponential and isotropic. σ_Y^2 is shown to have a profound impact upon the conductivity field. As the variance
increases, regions of high and low log-conductivity emerge, creating preferential flow paths bypassing the low
conductivity zones as shown by particle trajectories. At smaller variance (i.e., $\sigma_Y^2 = 0.1$), particles mainly travel along
the mean flow direction with very limited departure from the mean trajectory, which are the straight lines connecting
370   the left and right boundaries. In this situation, the arrival times of solute particles to a critical location (i.e., $\boldsymbol{\Omega}^*$, where
for example geochemical conditions are favorable to certain types of reactions to occur) are predictable. With large
variances (i.e., $\sigma_Y^2 = 2$), the streamlines assume a very irregular, hard-to-predict geometry, and we can observe the
emergence of flow channels, where particles can move fast, next to stagnant flow regions. Arrival times become more
uncertain, because the exact geometry of the streamlines is hard to predict unless the Y field is known deterministically.
375   However, since this is never the case, in another equally likely realization of the Y field, the situation may be different,
resulting in significant uncertainties in predicting the particle travel times. Thus, spatial variability of log-conductivity
is a major uncertainty-inducing factor, and by extension, obviating the need for stochastic modeling of HSHMs in
situations where the associated processes and attributes are subject to uncertainty. In the following sections, we will
present illustrative examples to analyze how subsurface spatial variability influences $< I_{HSHM}(\boldsymbol{\Omega}^*, t^*) >$, including
380   variance and anisotropy ratio of the log-conductivity.
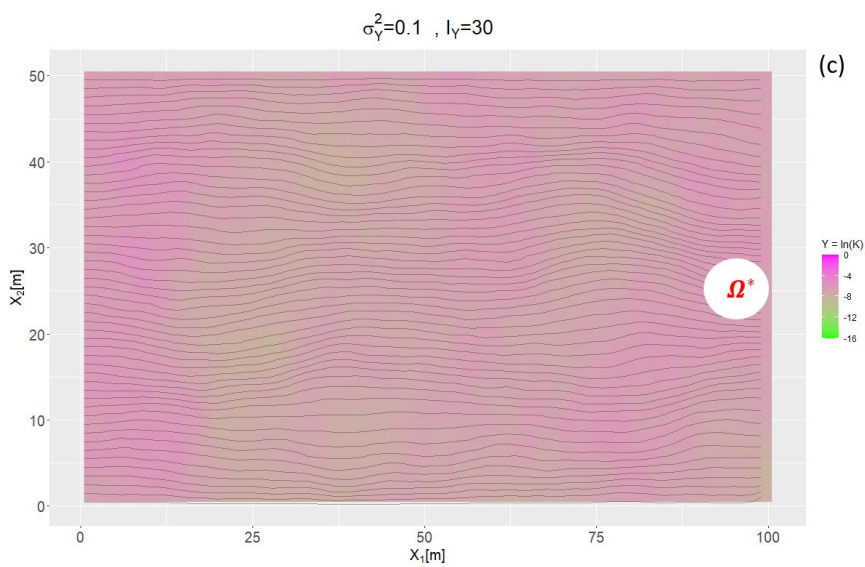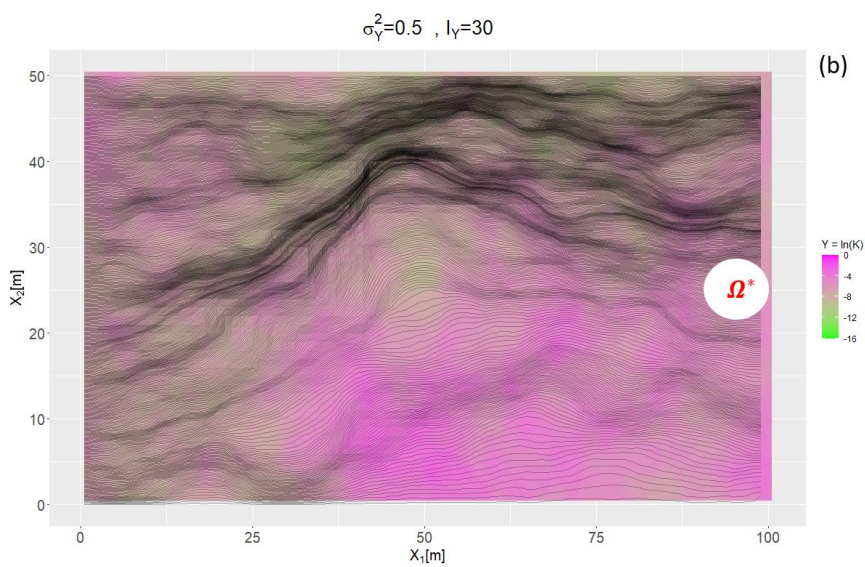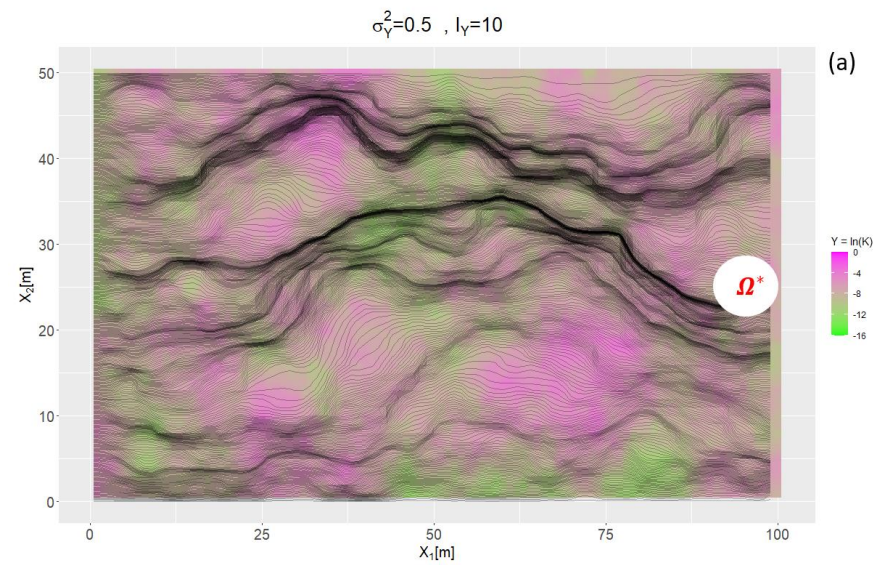
Figure 2. Illustrative example of a heterogeneous log-hydraulic conductivity field and solute particle transport. Black lines represent simulated particle travel paths. A left to right hydraulic gradient of 0.1 is applied. Mean of log-conductivity is set at -3. Note color scales for log-conductivity are consistent in all three panels.

4.2 Illustrative example and indicator formulation

In this section we illustrate the proposed indicator approach by means of synthetic case studies developed by using methods of stochastic hydrogeology. The choice of the synthetic case studies does not limit our approaches to broader applications where stochastic modeling with Monte Carlo simulations are applicable. Figure 3 displays the configuration of this case example. Consider the case of an instantaneous point source release of a target compound at the location x_0 and time t_0. HSHMs are triggered at any $(\boldsymbol{\Omega}^*, t^*)$ if the solute is present. Consider the hot spot $(\boldsymbol{\Omega}^*)$ to be confined within the following volume: $w_1 \leq x_1 \leq w_1'; w_2 \leq x_2 \leq w_2'; w_3 \leq x_3 \leq w_3'$, the dynamic indicator is therefore defined as follows:

$$I_d(\boldsymbol{\Omega}^*, t^*) = \begin{cases} 1, & if\ X(t^*) \subseteq \boldsymbol{\Omega}^*\ at\ t^* \\ 0, & otherwise \end{cases}. \tag{17}$$

The injected solute can be modelled in a Lagrangian framework as a particle moving according to the velocity field without changing its volume. The latter is the consequence of neglecting pore scale dispersion. The expected value of this dynamic indicator at t^* is therefore:

$$< I_d(\boldsymbol{\Omega}^*, t^*) >= \int_{\Omega^*} f_{X(t^*)}(\boldsymbol{a}|\boldsymbol{x}_0, t_0) d\boldsymbol{a}, \tag{18}$$

where $f_{X(t^*)}(\boldsymbol{a}|\boldsymbol{x}_0, t_0)$ is the probability distribution function (pdf) of the particle's trajectory at time t^* (Dagan and Nguyen, 1989; Rubin, 2003). If we also assume steady, uniform in the average flow with mild heterogeneity of the log hydraulic conductivity field with Gaussian displacement pdf—then we can compute $< I_{HSHM}(\boldsymbol{\Omega}^*, t^*) >$ analytically using the following equation:

$$P(I_{HSHM}(\boldsymbol{\Omega}^*, t^*) = 1) =< I_{HSHM}(\boldsymbol{\Omega}^*, t^*) >=< I_s(\boldsymbol{\Omega}^*) > < (I_d(\boldsymbol{\Omega}^*, t^*) >$$

$$= prob(I_d(\boldsymbol{\Omega}^*, t^*) = 1) = prob\{X(t^*) \subseteq \boldsymbol{\Omega}^*\}$$

$$= \prod_{i=1}^{m} \int_{w_i}^{w_i'} f_{X_i(t^*)}(a_i|x_0, t_0) da_i = \int_{w_1}^{w_1'} f_{X_1(t^*)}(a_1|x_0, t_0) da_1 \int_{w_2}^{w_2'} f_{X_2(t^*)}(a_2|x_0, t_0) da_2 \int_{w_3}^{w_3'} f_{X_3(t^*)}(a_3|x_0, t_0) da_3$$

$$= \frac{1}{(2\pi)^{\frac{3}{2}}\sqrt{X_{11}(t^*)X_{22}(t^*)X_{33}(t^*)}} \int_{w_1}^{w_1'} \exp\left[-\frac{1}{2}\frac{(a_1 - Ut^*)^2}{X_{11}(t^*)}\right] da_1$$

$$\cdot \int_{w_2}^{w_2'} \exp\left[-\frac{1}{2}\frac{a_2^2}{X_{22}(t^*)}\right] da_2 \int_{w_3}^{w_3'} \exp\left[-\frac{1}{2}\frac{a_3^2}{X_{33}(t^*)}\right] da_3. \tag{19}$$

which can be integrated to yield:

$$P(I_{HSHM}(\boldsymbol{\Omega}^*, t^*) = 1) =< I_{HSHM}(\boldsymbol{\Omega}^*, t^*) >= \frac{1}{8}\left[\text{erfc}\left(\frac{w_1 - Ut^*}{\sqrt{2X_{11}(t^*)}}\right) - \text{erfc}\left(\frac{w_1' - Ut^*}{\sqrt{2X_{11}(t^*)}}\right)\right]$$

$$\cdot \left[\text{erfc}\left(\frac{w_2}{\sqrt{2X_{22}(t^*)}}\right) - \text{erfc}\left(\frac{w_2'}{\sqrt{2X_{22}(t^*)}}\right)\right]\left[\text{erfc}\left(\frac{w_3}{\sqrt{2X_{33}(t^*)}}\right) - \text{erfc}\left(\frac{w_3'}{\sqrt{2X_{33}(t^*)}}\right)\right]. \tag{20}$$

For simplicity, but without lack of generality, in Eq. (20) we assumed $x_0 = (0,0,0)$. The displacement variances $X_{ii}, i = 1, 2, 3$ depend on the spatial distribution of the hydraulic conductivity in the subsurface. Eqs. (A4) to (A6) present the displacement variances for an axisymmetric exponential covariance function of the log-conductivity (A3) are given in the Appendix A. In Eqs. from (A7) to (A19), we have provided derivations of indicator formulations for other HSHMs scenarios, including indicator formulation for complex concentration thresholds and indicator formulation for hot moment durations. Notice that in obtaining Eq. (20) we postulated ergodicity, which in practical terms reflects the actual situation of an instantaneous injection into a source zone with transverse dimension much larger than the integral scale of the hydraulic conductivity (Dagan, 1990), such that the ensemble mean is representative of the effects of the actual, but unknown, distribution of hydraulic conductivity.
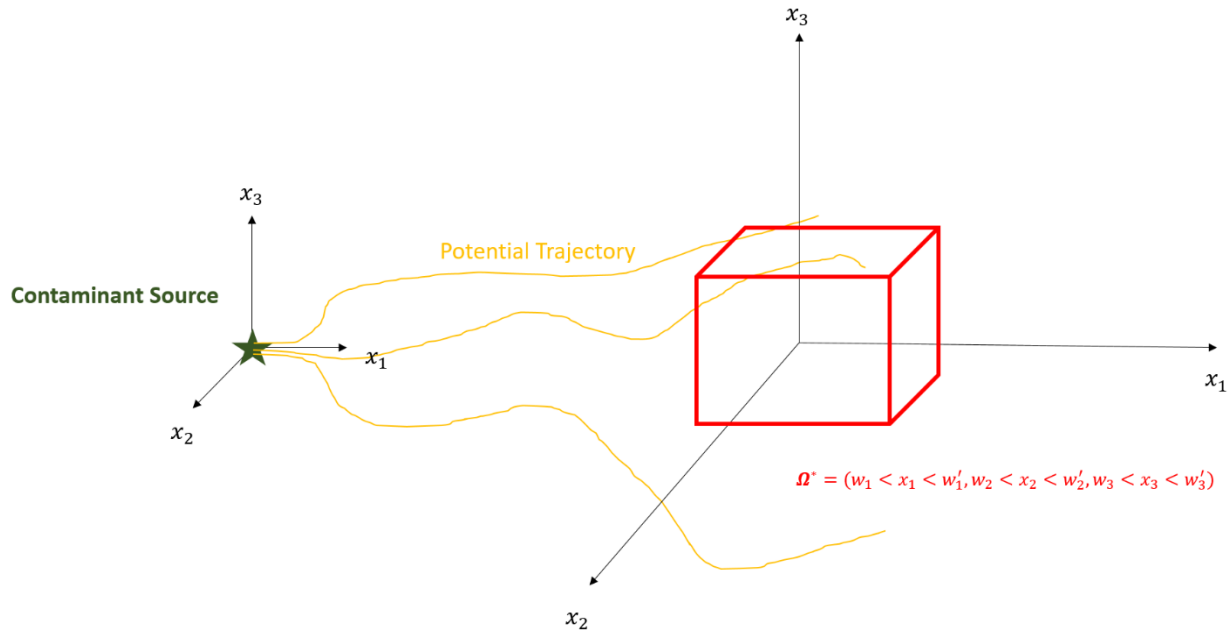
Figure 3. Configuration of the synthetic case study. x_1, x_2 and x_3 represent the longitudinal, transverse and vertical directions, respectively. $w_1, w_1', w_2, w_2', w_3$ and w_3' are the coordinates that set up the volume of $\boldsymbol{\Omega}^*$

4.3 Probability of HSHM occurrence controlled by subsurface heterogeneity

In the following sections, we present the results from the case study described in section 4.3. Specifically, in section 4.3.1 and 4.3.2, we explore how heterogeneity of log-hydraulic conductivity influences the probability of HSHM occurrences. To make results as general as possible, lengths are made dimensionless with respect to the integral scales (I_{Yh} in the two horizontal directions and I_{Yv} in the vertical one) and time with respect to the following advective time scale: I_{Yh}/U, where U is the mean velocity). In the following, we explore the effect of the remaining parameters, i.e. the anisotropy ratio $e = \frac{I_{YV}}{I_{YH}}$ and the variance of the log-conductivity σ_Y^2, on the emergence of HSHM. We placed $\boldsymbol{\Omega}^*$ along the mean trajectory at $(21I_{YH}, 0, 0)$ with dimensions as $(2I_{YH}, 2I_{YH}, 2I_{YV})$. The dimensions of the hot spot are therefore of two integral scales in the three coordinate directions (x_1, x_2, x_3) and is placed at a dimensionless distance of 21 from the point source.

4.3.1 Dependence of $P(I_{HSHM}(\boldsymbol{\Omega}^*, \tau) = 1)$ on variance in the spatial correlation structure of the log-conductivity
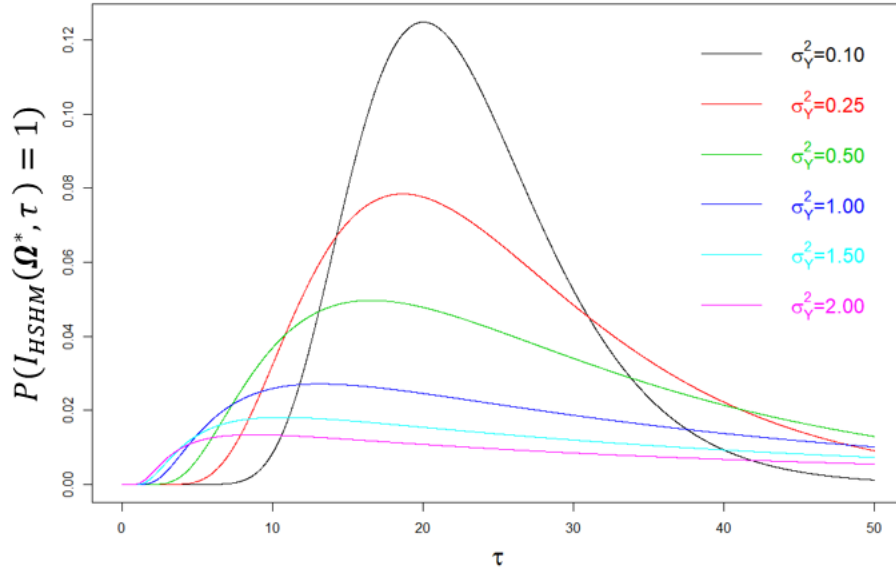
Figure 4. Dependence of $P(I_{HSHM}(\boldsymbol{\Omega}^*, \tau) = 1)$ on time (τ) and on the level of spatial variability of the log-hydraulic conductivity (σ_Y^2)

Figure 4 shows $P(I_{HSHM}(\boldsymbol{\Omega}^*, \tau) = 1)$, which is the propability that a HSHM is triggered at $\boldsymbol{\Omega}^*$ and the dimensionless time $\tau = \frac{tU}{I_{Yh}}$, for a few values of σ_Y^2. Here, $I_{HSHM}(\boldsymbol{\Omega}^*, \tau) = 1$ represents the situation in which a HSHM is triggered. At early time (e.g., $\tau < 5$), larger probability $P(I_{HSHM}(\boldsymbol{\Omega}^*, \tau) = 1)$ is observed with increase in σ_Y^2. At intermediate time, i.e., at times comparable with the mean travel time $\tau = 21$, $P(I_{HSHM}(\boldsymbol{\Omega}^*, \tau) = 1)$ is inversely proportional to σ_Y^2. At late time (e.g., $\tau > 40$), the largest $P(I_{HSHM}(\boldsymbol{\Omega}^*, \tau) = 1)$ occurs at intermediate σ_Y^2. We observe that σ_Y^2 regulates the timing of the peak in $P(I_{HSHM}(\boldsymbol{\Omega}^*, \tau) = 1)$, which is located in the proximity of the mean travel time, $\tau = 21$, for weak heterogeneity, and shifts towards earlier times as σ_Y^2 increases. From the practical perspective, Figure 4 shows the probability of developing a HSHM at the identified position $\boldsymbol{\Omega}^*$ at the given time τ.

These effects relate to the relationship between travel times (from the source to $\boldsymbol{\Omega}^*$) and σ_Y^2. The key point to note is that σ_Y^2 controls the spread of the travel time around its mean value. A larger σ_Y^2 enhances channeling effects (Fiori and Jankovic, 2012; Moreno and Tsang, 1994, also in Figure 2), which in turn enable earlier arrival times. But at the same time, it also leads to the emergence low-conductivity zones with low velocity or stagnant groundwater. The solute tends to bypass low hydraulic conductivity zones, as shown by the streamlines depicted in Figure 2, however, the small amount of solute that actually penetrates these zones by slow advection and diffusion gets trapped for long time before being released and this results in an extended tailing with low concentration, which increases the probability of observing a HSHM at later times. Thus, with an increase in σ_Y^2, we notice an increase in the probability to observe both increasingly earlier and increasingly delayed arrival times, which widens the probability distribution. On the contrary at small variance, particles deviate little from the ensemble mean trajectory, because of the small contrast in conductivity between high and low conductivity zones. This results in small particle spreading and travel

times that differ only slightly from the mean travel time ($\tau = 21$), and a probability distribution less spread around the mean, where the peak is observed.

In summary, hydraulic conductivity contrast between low and high conductive lithofacies increases with σ_Y^2 leading to the emergence of organized high conductivity pathways sneaking through surrounding low conductivity zones with the latter acting as "trapping" elements. This causes the emergence of both early and late arrival times, with the consequent larger probability of triggering HSHMs at early and later times, with respect to the case of low heterogeneity. Early arrival times are controlled by the connected high conductivity pathways and the late arrival times are influenced by the low conductivity zones, which act as low-release reservoirs for solutes.

4.3.2 Dependence of $P(I_{HSHM}(\Omega^*, \tau) = 1)$ on on anisotropy in the spatial correlation structure of the log-hydraulic conductivity
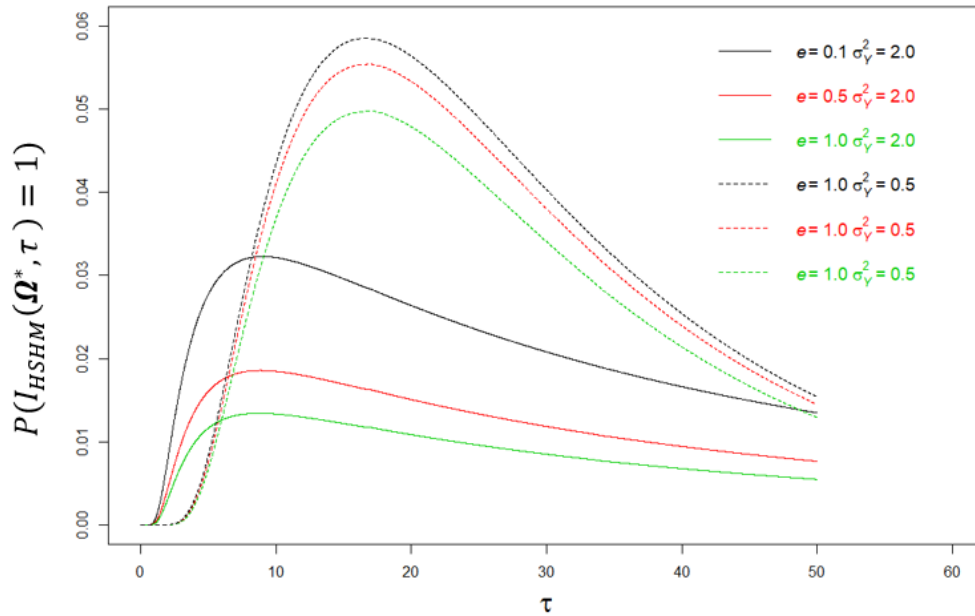
Figure 5. Dependence of $P(I_{HSHM}(\Omega^*, \tau) = 1)$ on (τ) and on anisotropy ratio e. Solid and dashed lines represent, respectively, the probabilities for large ($\sigma_Y^2 = 2.0$) and small ($\sigma_Y^2 = 0.5$) variance of the log-hydraulic conductivity.

The discussion here (accompanying Figure 5) focuses on the impact of the anisotropy ratio in the correlation structure (e, defined above) on the probability of triggering HSHMs. The anisotropy ratio, e, provides an indication about the persistence of the log-conductivity (Y) in the principal directions. The spatial correlation model used here for demonstration is that of axis-symmetry, which is common to sedimentary formations (Dagan, 1989; Rubin, 2003), with e providing the ratio between the persistence of Y in the vertical (x_3) direction, represented by I_{YV}, and the ones on the horizontal plane ($x_1 - x_2$), represented by I_{YH}. In unconsolidated sedimentary formations, I_{YV} is typically smaller than I_{YH} by as much as one order of magnitude, due to the different time scales of the depositional process taking place in the horizontal and vertical directions, which leads to thin and elongated lithofacies and consequently to a much smaller persistence of Y values in normal to the horizontal plane (Miall, 1985, 1988; Ritzi et al., 2004).

Figure 5 compares $P(I_{HSHM}(\Omega^*, \tau) = 1)$ between formations defined by different anisotropy ratios and different σ_Y^2. It shows that we have two factors to consider when explaining the differences in $P(I_{HSHM}(\Omega^*, \tau) = 1)$.

First factor, as discussed earlier, is the widening of the probability distribution (direct consequence of the widening of the travel times distribution) due to increase in σ_Y^2. With larger variance, we observe higher probabilities for departure of the travel times away from the average. The anisotropy ratio e adds a compounding factor. To understand its effect, we should recall the analyses of lateral displacement variances of solute particles moving in heterogeneous formations (cf., Dagan, 1989, and Eq. A4 to A6 here), showing that smaller e leads to smaller lateral (both vertical and horizontal) displacement variances, implying smaller probabilities for lateral departures from the mean flow trajectory. Smaller e limits lateral spreads, and increase the probability of particle to enter $\boldsymbol{\Omega}^*$, sooner or later, and to trigger HSHM. The effect could also be viewed as a channeling effect of sorts: smaller e implies Y blocks of small aspect ratio (i.e., long but thin elements), which provide fast tracks for particles when defined by high Y values, while blocking lateral spreads when defined by low Y values.

In additional notes: first, $\boldsymbol{\Omega}^*$ is known in the present analysis is located downstream from the source, along with the mean trajectory of the solute displacement. We expect different results in situations where $\boldsymbol{\Omega}^*$ is positioned at an offset with respect to the mean flow direction, or when its position is unknown. In both cases we expect a reduction of the probability of triggering a HSHM. Relevant of our discussion is that the proposed probabilistic framework can address the case of unknown position for $\boldsymbol{\Omega}^*$ as well. Second, we note that the analytical models used to compute the displacement statistics are formally limited to small variance of the log-conductivity ($\sigma_Y^2 < 1$), although they are shown to provide good approximations for large variances (Bellin et al., 1992; Salandin and Fiorotto, 1998). Third, the stochastic formulation provides the theoretical and computational formalism for conditioning the probabilities on in-situ measurements (Ezzedine and Rubin, 1996; Rubin and Dagan, 1992) as well as on information borrowed from similar sites (Li et al., 2018; Cucchi et al., 2019).

5 Discussion and Summary

In this study, we developed a general stochastic framework for characterizing the spatiotemporal distribution of environmental Hot Spots Hot Moments (HSHMs). The stochastic formulation is built around the following principles:

- The HSHMs are defined as random variables and the goal is to derive their stochastic distribution in terms of the relevant processes and attributes.
- HSHMs processes cover the dynamic components of the HSHMs. An example could be the transport of solutes and reactants. HSHMs attributes refer to the static components of the HSHMs, e.g., in situations related to the nitrogen cycle, attributes could represent pyrite concentration or naturally-reducing zones. HSHMs could be defined through the confluence of a variety of contributors, both static and dynamic.
- The processes and attributes are modeled as stochastic processes and random variables, respectively, based on the underlying physics.
- The static contributors are modeled stochastically using geostatistical space random functions.
- The dynamic contributors are modeled stochastically using probability distribution functions derived from the underlying mathematical-physical models.

- Several HSHMs categories are defined, based on the contributing factors, as follows: HSHMs defined by dynamic contributors only, HSHMs defined by static contributors, and most commonly, HSHMs requiring the coupling of static and dynamic contributors. The HSHMs stochastic formulations are expressed in terms of the stochastic formulations of the relevant contributors.

- We provided a detailed review of multiple HSHMs and showed how they relate to our definitions.

The framework we proposed in this study is advantageous in that it allows to calculate the uncertainty associated with HSHMs based on the uncertainty associated with its contributors. Additionally, it provides a formalism, well established by Bayesian theory, for conditioning the HSHM probabilities on *in-situ* measurements as well as on information borrowed from geologically and otherwise similar sites.

We demonstrated our proposed approach through applications in the area of subsurface transport and hydrogeology, focusing on the impacts of subsurface heterogeneity on HSHMs. We analyzed, quantitatively, how subsurface heterogeneity of the conductivity field control the HSHM statistics, for example, the time expected for the probability of the HSHM to occur to reach a-priori set thresholds or time to peak probability.

Lastly, as mentioned both here and in previous studies, statistical methods for quantifying the occurrences of HSHMs and the associated uncertainties are needed to advance our understanding of the mechanisms that cause HSHMs, as well as to enhance our ability to predict HSHMs and manage their consequences.

Acknowledgments

We gratefully acknowledge the Jane Lewis Fellowship Committee of the University of California, Berkeley, and the Earth and Environmental Science Division at Lawrence Berkeley National Laboratory for their generous support through fellowships awarded to the first author. The second author's work is supported by the Office of Science, Office of Advanced Scientific Computing, as part of the project "Deduce: Distributed Dynamic Data Analytics Infrastructure for Collaborative Environments" under contract no. DE-AC02-05CH11231. The third author acknowledges funding from the Italian Ministry of Education, University and Research (MIUR) in the frame of the Departments of Excellence Initiative 2018—2022 granted to DICAM of the University of Trento. As this study focused on theory development, data were not used, nor created for this research. All the information needed for evaluating and checking the results is provided in the paper.

Appendix

A1. Equations for the displacement pdf

Assuming steady, uniform in the average flow, with mild heterogeneity of the log-hydraulic conductivity field with Gaussian displacement, the pdf of the longitudinal (x_1, Figure 3) displacement of a solute particle starting at time $t_0 = 0$, at $\boldsymbol{x}_0 =(0,0,0)$ is given by the following equation (Dagan and Nguyen, 1989; Dagan and Rubin, 1992):

$$f_{X_1(t^*)}(x_1) = \frac{1}{\sqrt{2\pi X_{11}(t^*)}} \exp\left[-\frac{1}{2}\frac{(x_1 - Ut^*)^2}{X_{11}(t^*)}\right]. \tag{A1}$$

Additionally, the displacement pdf in the transverse directions (x_2 and x_3) is given by

$$f_{X_i(t^*)}(x_i) = \frac{1}{\sqrt{2\pi X_{ii}(t^*)}} \exp\left[-\frac{1}{2}\frac{x_i^2}{X_{ii}(t^*)}\right], i = 2,3. \tag{A2}$$

A2. Equations for displacement variances under anisotropic conditions

Dagan (1984) developed a solution of the displacement variances $X_{ii}, i = 1,2,3$ for an exponential and axisymmetric log-conductivity covariance function:

$$C_Y(\boldsymbol{r}) = \langle (Y(\boldsymbol{x}) - \langle Y \rangle)\, (Y(\boldsymbol{x} + \boldsymbol{r}) - \langle Y \rangle) \rangle = \sigma_Y^2 \exp\left[-\sqrt{\frac{r_1^2 + r_2^2}{I_{Yh}^2} + \frac{r_3^2}{I_{Yv}^2}}\right], \tag{A3}$$

$$X_{11} = \sigma_Y^2 I_Y^2 \{2t^* + 2[\exp(-t^*) - 1] + 8e \int_0^\infty [\bar{J}_0(Kt^*) - 1]$$

$$\cdot \left[\frac{1}{(1 + K^2 - e^2 K^2)^2} - \frac{eK}{(1 + K^2 - e^2 K^2)^2 (1 + K^2)^{0.5}} - \frac{eK}{2(1 + K^2 - e^2 K^2)(1 + K^2)^{1.5}}\right] dK$$

$$-2e \int_0^\infty \left[\bar{J}_0(Kt^*) - \frac{\bar{J}_1(Kt^*)}{Kt^*} - \frac{1}{2}\right] \cdot \left[\frac{e^3 K^3 (e^2 K^2 - 5 - 5K^2)}{(e^2 K^2 - 1 - K^2)^3 (1 + K^2)^{1.5}} + \frac{1 + K^2 - 5e^2 K^2}{(1 + K^2 - e^2 K^2)^3}\right] dK\}, \tag{A4}$$

$$X_{22} = -2e\sigma_Y^2 I_Y^2$$

$$\cdot \int_0^\infty \left[\frac{\bar{J}_1(Kt^*)}{t^*} - \frac{K}{2}\right]\left[\frac{e^3 K^2 (e^2 K^2 - 5K^2 - 5)}{(e^2 K^2 - 1 - K^2)^3 (1 + K^2)^{1.5}} + \frac{1 + K^2 - 5e^2 K^2}{K(1 + K^2 - e^2 K^2)}\right] dK, \tag{A5}$$

$$X_{33} = -4e\sigma_Y^2 I_Y^2 \int_0^\infty [\bar{J}_0(Kt^*) - 1]$$

$$\cdot \left\{\frac{1}{(e^2 K^2 - 1 - K^2)^2}\left[\frac{1}{2} + \frac{2e^2 K^2}{1 + K^2 - e^2 K^2} + \frac{eK(e^2 K^2 + 3 + 3K^2)}{2(e^2 K^2 - 1 - K^2)(1 + K^2)^{0.5}}\right]\right\} dK. \tag{A6}$$

where $\boldsymbol{r}$ is the two-point separation distance and $\langle Y \rangle$ the ensemble mean of the log-conductivity $Y = \ln K$. $\bar{J}_0$ and $\bar{J}_1$ are, respectively, the zero and first order of the first kind Bessel functions.

A3. Equations for displacement variances under isotropic conditions

Dagan (1984) provided analytical solutions for longitudinal and transverse displacement variances. This is a special case for the anisotropic case with $e = 1$. The solutions are as follows:

$$X_{11} = \sigma_Y^2 I_Y^2 \left\{2t^* - 2 \cdot \left[\frac{8}{3} - \frac{4}{t^*} + \frac{8}{t^{*3}} - \frac{8}{t^{*2}}\left(1 + \frac{1}{t^*}\right)\exp(-t^*)\right]\right\}. \tag{A7}$$

$$X_{22} = X_{33} = 2\sigma_Y^2 I_Y^2 \left[\frac{1}{3} - \frac{1}{t^*} + \frac{4}{t^{*3}} - \left(\frac{4}{t^{*3}} + \frac{4}{t^{*2}} + \frac{1}{t^*}\right)\exp(-t^*)\right]. \tag{A8}$$

A4. Indicator formulation for complex concentration thresholds

When considering local dispersion, or in case of a reactive tracer, the condition that the particle is inside the volume $\boldsymbol{\Omega}^*$ does not suffice to define the dynamic indicator and a concentration threshold C_{th} should be introduced:

$$I_d(\boldsymbol{\Omega}^*, t^*) = \begin{cases} 1, & if\ \boldsymbol{X}(t^*; \boldsymbol{x}_0, t_0) \subseteq \boldsymbol{\Omega}^*\ and\ C(\boldsymbol{X}, t^*) > C_{th} \\ 0, & otherwise \end{cases}. \tag{A9}$$

In the absence of local dispersion and for a reactive solute decaying at a (spatially) constant rate k, the ensemble mean assumes the following expression (Cvetkovic and Shapiro, 1990):

$$P(I_d(\boldsymbol{\Omega}^*, t^*) = 1) = <I_d(\boldsymbol{\Omega}^*, t^*)> = \left\{1 - H\left[t^* - \frac{1}{k}\ln(\frac{C_0}{C_{th}})\right]\right\} \int_{\Omega^*} f_{X(t^*)}(\boldsymbol{a}|\boldsymbol{x}_0, t_0) d\boldsymbol{a}, \tag{A10}$$

580   where C_0 is the initial concentration and $H[\cdot]$ is the Heaviside step function. The ensemble mean (Eq. A10) is the product of the probability that the particle assumes a concentration larger than the threshold at t^* (given that reaction rate k is constant, this probability is either 0 or 1) and the probability that at the same time t^* the particle is within the hot spot $\boldsymbol{\Omega}^*$. In other words, Eq. (A10) expresses the fact that a particle inside $\boldsymbol{\Omega}^*$ contributes to the hot moment only if its concentration is greater than the threshold, and this occurs for $t^* < \frac{1}{k}\ln(\frac{C_0}{C_{th}})$. Eq. (A10) can be generalized to the

585   cases of instantaneous injection into a source of volume V_0, as discussed before for the non-reactive case. For other complex situations, such as that in which k is spatially variable and complex reaction networks, the ensemble mean of the indicators can be addressed by Eq. (16) in a Monte Carlo framework.

A5. Indicator formulation for hot moment durations

As hot moments can persist over short time periods, estimating the corresponding probabilities for any given

590   time interval becomes also very important. The probability that the hot moment persists over the interval $[t_1, t_2]$ at $\boldsymbol{\Omega}^*$ can be formally computed as follows:

$$<I_d(\boldsymbol{\Omega}^*, t_1, t_2)> = P(t_1, \boldsymbol{\Omega}^*)P(t_2|t_1, \boldsymbol{\Omega}^*), \tag{A11}$$

where $P(t_1, \boldsymbol{\Omega}^*)$ is the probability that the particle is inside $\boldsymbol{\Omega}^*$ at time $t^* = t_1$ and $P(t_2|t_1, \boldsymbol{\Omega}^*)$ is the probability that the particle is still inside $\boldsymbol{\Omega}^*$ at time $t^* = t_2$, provided that at time t_1, it was also inside $\boldsymbol{\Omega}^*$. If the particle exits $\boldsymbol{\Omega}^*$

595   during interval $[t_1, t_2]$, this time interval will not be qualified as hot moment; and thus the probability computation needs to ensure the particle stays within $\boldsymbol{\Omega}^*$ during the entire time interval.

Under the First-Order Approximation (FOA) (see e.g., Dagan, 1989; Gelhar 1993; Rubin, 2003), the pdf of the particle displacement is normal with mean $<\boldsymbol{X}(t^*; \boldsymbol{x}_0, t_0)>$ and auto-covariance tensor of the residual displacements $\boldsymbol{X}'(t^*) = \boldsymbol{X}(t^*) - \langle\boldsymbol{X}(t^*)\rangle$ defined by $\boldsymbol{X}_{ij}(t^*; \boldsymbol{x}_0, t_0) = \langle\boldsymbol{X}_i'(t^*; \boldsymbol{x}_0, t_0)\boldsymbol{X}_j'(t^*; \boldsymbol{x}_0, t_0)\rangle$, $i, j = 1, 2, 3$.

600   For simplicity in the following, we assume $\boldsymbol{x}_0 = 0$ and $t_0 = 0$. Under these assumptions,

$$<I_d(\boldsymbol{\Omega}^*, t_1, t_2)> = \int_{\Omega^*} \int_{\Omega^*} f_{X(t_1)}(\boldsymbol{a})f_{X(t_2)}^c(\boldsymbol{b}|\boldsymbol{X}(t_1) = \boldsymbol{a}) d\boldsymbol{b} \, d\boldsymbol{a}, \tag{A12}$$

where the conditional pdf $f_{X(t_2)}^c(\boldsymbol{b}|\boldsymbol{X}(t_1) = \boldsymbol{a})$ is multi-normally distributed with conditional mean and variance tensor given by

$$\langle\boldsymbol{X}(t_2)| \boldsymbol{X}(t_1) = \boldsymbol{a}\rangle = <\boldsymbol{X}(t_2)>$$

605    $$+Cov[\boldsymbol{X}'(t_2), \boldsymbol{X}'(t_1)] \cdot Var[\boldsymbol{X}'(t_1)]^{-1} \cdot (\boldsymbol{a} - <\boldsymbol{X}(t_1)>), \tag{A13}$$

and

$$\boldsymbol{\sigma}(t_1, t_2) = Var[\boldsymbol{X}'(t_2)] - Cov[\boldsymbol{X}'(t_2), \boldsymbol{X}'(t_1)] \cdot Var[\boldsymbol{X}'(t_1)]^{-1} \cdot Cov[\boldsymbol{X}'(t_1), \boldsymbol{X}'(t_2)], \tag{A14}$$

respectively, which further yields the following,

$$f_{X(t_2)}^c(\boldsymbol{b}|\boldsymbol{X}(t_1) = \boldsymbol{a})$$

$$= \exp\left[-\frac{1}{2}\ [\boldsymbol{b} - \langle \boldsymbol{X}(t_2)|\ \boldsymbol{X}(t_1) = \boldsymbol{a})\rangle]^T \cdot \boldsymbol{\sigma}(t_1, t_2)^{-1} \cdot [\boldsymbol{b} - \langle \boldsymbol{X}(t_2)|\ \boldsymbol{X}(t_1) = \boldsymbol{a})\rangle]\right]$$

$$\cdot \{8\ \pi^3\ \cdot |\boldsymbol{\sigma}(t_1, t_2)|\}^{-\frac{1}{2}}, \tag{A15}$$

where $|\cdot|$ indicates the determinant, exp is the exponential function and the exponent T indicates the transpose of the vector.

In Eq. (A13) and (A14), $\boldsymbol{X}'(t^*) = \boldsymbol{X}(t^*) - \langle \boldsymbol{X}(t^*)\rangle$ stands for the departure of the particle's displacement with respect to the ensemble mean trajectory, and $Var[\boldsymbol{X}]^{-1}$ is the auto-covariance tensor of the residual displacement whose elements are defined above. Similarly, $Cov[\boldsymbol{X}'(t_1), \boldsymbol{X}'(t_2)]$ is the covariance tensor of residual displacement which elements are: $\boldsymbol{X}_{ij}(t_1, t_2; \boldsymbol{x_0}, t_0) = \langle X_i'(t_1)\ X_j'(t_2)\rangle, i, j = 1, 2, 3$. Note that in the general three-dimensional case $\langle \boldsymbol{X}(t_2)|\ \boldsymbol{X}(t_1) = \boldsymbol{a})\rangle$ is a three-dimensional vector and $\boldsymbol{\sigma}(t_1, t_2)$ is a 3×3 second-order tensor.

For $t_2 \to t_1$, $f^c_{X(t_2)}[\boldsymbol{b}|\boldsymbol{X}(t_1) = \boldsymbol{a}] \to \delta(\boldsymbol{b} - \boldsymbol{a})$, where $\delta(\cdot)$ is the Dirac Delta, such that $P(t_2|t_1, \boldsymbol{\Omega}^*) \to 1$. On the other hand, for $t_2 \gg t_1$, $Cov[\boldsymbol{X}'(t_1), \boldsymbol{X}'(t_2)] \to 0$ and $P(t_2|t_1, \boldsymbol{\Omega}^*) \to P(t_2, \boldsymbol{\Omega}^*)$ the marginal probability that the particle is within $\boldsymbol{\Omega}^*$ at time $t^* = t_2$. Eq. (A12) to (A15) are obtained under the FOA approximation and assuming that the particle can enter $\boldsymbol{\Omega}^*$ only once. Such assumption is needed to obtain analytical solutions and is reasonable for situations with small to mild subsurface heterogeneity (e.g., $\sigma_Y^2 \leq 1.6$), such as the cases presented in Bellin et al. (1992, 1994); Cvetkovic et al. (1992). In particular, FOA assumes small heterogeneity and under this assumption the particle trajectory deviates slightly from its ensemble mean, which is directed along the regional hydraulic head gradient. For a regular volume $\boldsymbol{\Omega}^*$, this reduces the probability of the particle entering more than once the hot spot. This probability reduces further if in horizontal and vertical transverse directions $\boldsymbol{\Omega}^*$ is much larger than the respective integral scales, because the probability of observing negative longitudinal velocity components (i.e., along the mean flow field) is much smaller than in the transverse directions (Bellin et al., 1992) and vanishes as formation heterogeneity reduces.

If the hotspot $\boldsymbol{\Omega}^*$ is the volume confined between two planes at $x_1 - \frac{l_1}{2}$ and $x_1 + \frac{l_1}{2}$, with the other two dimensions much larger than the transverse horizontal and vertical integral scales: $l_2 \gg I_h$, $l_3 \gg I_v$, Eq. (A13) simplifies to:

$$< I_d(\boldsymbol{\Omega}^*, t_1, t_2) >= \int_{x_1-\frac{l_1}{2}}^{x_1+\frac{l_1}{2}} \int_{x_1-\frac{l_1}{2}}^{x_1+\frac{l_1}{2}} f_{X_1(t^*)}(a_1)f^c_{X_1(t^*)}(\ b_1|X_1(t_1) = a_1)\ db_1\ da_1, \tag{A16}$$

where X_1 is the longitudinal component of the particle's trajectory and $f^c_{X_1(t^*)}$ is its conditional pdf, which is normal with conditional mean and variance given by

$$\mu[a_1] = \langle X_1(t_2)|\ X_1(t_1) = a_1)\rangle = < X_1(t_2) > + \frac{X_{11}(t_1, t_2)}{X_{11}(t_1)}(a_1 - < X_1(t_1) >), \tag{A17}$$

and

$$\sigma^2(t_1, t_2) = X_{11}(t_2) - \frac{X_{11}(t_1, t_2)^2}{X_{11}(t_1)}, \tag{A18}$$

respectively. Consequently, $f_{X^c(t^*)}$ in Eq. (A16) assumes the following form:

$$f^c_{X_1(t^*)}(b_1|X_1(t_1) = a_1) = \frac{1}{\sqrt{2\ \pi\ \sigma(t_1, t_2)}}\exp\left[-\frac{1}{2}(b_1 - \mu[a_1])^2\ \sigma(t_1, t_2)^{-1}\right]. \tag{A19}$$

Substituting Eq. (A16) into Eq. (A19) allows us to compute $< I_d(\boldsymbol{\Omega}^*, t_1, t_2) >$. For situations where the FOA assumptions are not valid (e.g., large heterogeneity), Monte Carlo simulation framework is still applicable as alternative approach to construct the dynamics indicators (see Eq. 16).

645 **References**

Abbott, B. W., Baranov, V., Mendoza-Lera, C., Nikolakopoulou, M., Harjung, A., Kolbe, T., Balasubramanian, M. N., Vaessen, T. N., Ciocca, F., Campeau, A., Wallin, M. B., Romeijn, P., Antonelli, M., Gonçalves, J., Datry, T., Laverman, A. M., de Dreuzy, J. R., Hannah, D. M., Krause, S., Oldham, C. and Pinay, G.: Using multi-tracer inference to move beyond single-catchment ecohydrology, Earth-Science Rev., 160, 19–42,

650 doi:10.1016/j.earscirev.2016.06.014, 2016.

Akaike, H.: A New Look at the Statistical Model Identification, IEEE Trans. Automat. Contr., doi:10.1109/TAC.1974.1100705, 1974.

Andrews, D. M., Lin, H., Zhu, Q., Jin, L. and Brantley, S. L.: Hot spots and hot moments of dissolved organic carbon export and soil organic carbon storage in the shale hills catchment, Vadose Zo. J., 10(3), 943–954,

655 doi:10.2136/vzj2010.0149, 2011.

Arora, B., Mohanty, B. P., McGuire, J. T. and Cozzarelli, I. M.: Temporal dynamics of biogeochemical processes at the Norman Landfill site, Water Resour. Res., 49(10), doi:10.1002/wrcr.20484, 2013.

Arora, B., Dwivedi, D., Hubbard, S. S., Steefel, C. I. and Williams, K. H.: Identifying geochemical hot moments and their controls on a contaminated river floodplain system using wavelet and entropy approaches, Environ. Model.

660 Softw., 85, 27–41, doi:10.1016/j.envsoft.2016.08.005, 2016a.

Arora, B., Spycher, N. F., Steefel, C. I., Molins, S., Bill, M., Conrad, M. E., Dong, W., Faybishenko, B., Tokunaga, T. K., Wan, J., Williams, K. H. and Yabusaki, S. B.: Influence of hydrological, biogeochemical and temperature transients on subsurface carbon fluxes in a flood plain environment, Biogeochemistry, 127(2–3), 367–396, doi:10.1007/s10533-016-0186-8, 2016b.

665 Arora, B., Wainwright, H. M., Dwivedi, D., Vaughn, L. J. S., Curtis, J. B., Torn, M. S., Dafflon, B. and Hubbard, S. S.: Evaluating temporal controls on greenhouse gas (GHG) fluxes in an Arctic tundra environment: An entropy-based approach, Sci. Total Environ., doi:10.1016/j.scitotenv.2018.08.251, 2019a.

Arora, B., Dwivedi, D., Faybishenko, B., Jana, R. B. and Wainwright, H. M.: Understanding and Predicting Vadose Zone Processes, Rev. Mineral. Geochemistry, 85(1), doi:10.2138/rmg.2019.85.10, 2019b.

670 Arora, B., Briggs, M. A., Zarnetske, J., Stegen, J., Gomez-Valez, J., Dwivedi, D. and Steefel, C.: Hot Spots and Hot Moments in the Critical Zone: Identification of and Incorporation into Reactive Transport Models, edited by A. Wymore, W. Yang, W. Silver, B. McDowell, and J. Chorover, Biogeochemistry of the Critical Zone, Springer-Nature, Accepted., 2021.

Bellin, A. and Rubin, Y.: On the use of peak concentration arrival times for the inference of hydrogeological

675 parameters, Water Resour. Res., doi:10.1029/2003WR002179, 2004.

Bellin, A., Salandin, P. and Rinaldo, A.: Simulation of dispersion in heterogeneous porous formations: Statistics, first-order theories, convergence of computations, Water Resour. Res., doi:10.1029/92WR00578, 1992.

Bellin, A., Rubin, Y. and Rinaldo, A.: Eulerian-Lagrangian approach for modeling of flow and transport in

heterogeneous geological formations, Water Resour. Res., 30(11), 2913–2924, doi:10.1029/94WR01489, 1994.

680   Bernhardt, E. S., Blaszczak, J. R., Ficken, C. D., Fork, M. L., Kaiser, K. E. and Seybold, E. C.: Control Points in Ecosystems: Moving Beyond the Hot Spot Hot Moment Concept, Ecosystems, 20(4), 665–682, doi:10.1007/s10021-016-0103-y, 2017.

Beven, K. J. and Kirkby, M. J.: A physically based, variable contributing area model of basin hydrology, Hydrol. Sci. Bull., doi:10.1080/02626667909491834, 1979.

685   Bundt, M., Widmer, F., Pesaro, M., Zeyer, J. and Blaser, P.: Preferential flow paths: Biological "hot spots" in soils, Soil Biol. Biochem., 33(6), 729–738, doi:10.1016/S0038-0717(00)00218-2, 2001.

Copty, N., Rubin, Y. and Mavko, G.: Geophysical-hydrological identification of field permeabilities through Bayesian updating, Water Resour. Res., 29(8), 2813–2825, doi:10.1029/93WR00745, 1993.

Cucchi, K., Heße, F., Kawa, N., Wang, C. and Rubin, Y.: Ex-situ priors: A Bayesian hierarchical framework for

690   defining informative prior distributions in hydrogeology, Adv. Water Resour., doi:10.1016/j.advwatres.2019.02.003, 2019.

Cvetkovic, V., Shapiro, A. M. and Dagan, G.: A solute flux approach to transport in heterogeneous formations: 2. Uncertainty analysis, Water Resour. Res., 28(5), 1377–1388, doi:10.1029/91WR03085, 1992.

Cvetkovic, V. D. and Shapiro, A. M.: Mass arrival of sorptive solute in heterogeneous porous media, Water Resour.

695   Res., 26(9), 2057–2067, doi:10.1029/WR026i009p02057, 1990.

Dagan, G.: Solute transport in heterogeneous porous formations, J. Fluid Mech., 145(2), 151–177, doi:10.1017/S0022112084002858, 1984.

Dagan, G.: Flow and Transport in Porous Formations., Springer Verlag, Berlin., 1989.

Dagan, G.: Transport in heterogeneous porous formations: Spatial moments, ergodicity, and effective dispersion,

700   Water Resour. Res., 26(6), doi:10.1029/WR026i006p01281, 1990.

Dagan, G. and Nguyen, V.: A comparison of travel time and concentration approaches to modeling transport by groundwater, J. Contam. Hydrol., 4(1), 79–91, doi:10.1016/0169-7722(89)90027-2, 1989.

Dagan, G. and Rubin, Y.: Conditional estimation of solute travel time in heterogeneous formations: Impact of transmissivity measurements, Water Resour. Res., 28(4), 1033–1040 [online] Available from:

705   http://www.agu.org/pubs/crossref/1992.../91WR02759.shtml, 1992.

Destouni, G. and Cvetkovic, V.: Field scale mass arrival of sorptive solute into the groundwater, Water Resour. Res., 27(6), 1315–1325, doi:10.1029/91WR00182, 1991.

Duncan, J. M., Groffman, P. M. and Band, L. E.: Towards closing the watershed nitrogen budget: Spatial and temporal scaling of denitrification, J. Geophys. Res. Biogeosciences, 118(3), 1105–1119, doi:10.1002/jgrg.20090,

710   2013a.

Duncan, J. M., Groffman, P. M. and Band, L. E.: Towards closing the watershed nitrogen budget: Spatial and temporal scaling of denitrification, J. Geophys. Res. Biogeosciences, 118(3), 1105–1119, doi:10.1002/jgrg.20090, 2013b.

Dwivedi, D.: Hot Spots and Hot Moments of Nitrogen in a Riparian Corridor, Water Resour. Res., 1–43,

715   doi:10.1002/2017WR022346, 2017.

Ezzedine, S. and Rubin, Y.: A geostatistical approach to the conditional estimation of spatially distributed solute concentration and notes on the use of tracer data in the inverse problem, Water Resour. Res., doi:10.1029/95WR02285, 1996.

Fiori, A. and Jankovic, I.: On Preferential Flow, Channeling and Connectivity in Heterogeneous Porous Formations,
720   Math. Geosci., 44(2), 133–145, doi:10.1007/s11004-011-9365-2, 2012.

Frei, S., Knorr, K. H., Peiffer, S. and Fleckenstein, J. H.: Surface micro-topography causes hot spots of biogeochemical activity in wetland systems: A virtual modeling experiment, J. Geophys. Res. Biogeosciences, 117(4), 1–18, doi:10.1029/2012JG002012, 2012.

Gu, C., Anderson, W. and Maggi, F.: Riparian biogeochemical hot moments induced by stream fluctuations, Water
725   Resour. Res., 48(9), 1–17, doi:10.1029/2011WR011720, 2012.

Harms, T. K. and Grimm, N. B.: Hot spots and hot moments of carbon and nitrogen dynamics in a semiarid riparian zone, J. Geophys. Res. Biogeosciences, 113(1), 1–14, doi:10.1029/2007JG000588, 2008.

Henri, C. V., Fernàndez-Garcia, D. and de Barros, F. P. J.: Probabilistic human health risk assessment of degradation-related chemical mixtures in heterogeneous aquifers: Risk statistics, hot spots, and preferential
730   channels, Water Resour. Res., 51(6), 4086–4108, doi:10.1002/2014WR016717, 2015.

Hill, A. R., Devito, K. J. and Campagnolo, S.: Subsurface Denitrification in a Forest Riparian Zone : Interactions between Hydrology and Supplies of Nitrate and Organic Carbon Author (s): Alan R . Hill , Kevin J . Devito , S . Campagnolo and K . Sanmugadas Published by : Springer Stable URL : http:/, , 51(2), 193–223, 2000.

Hubbard, S. S., Rubin, Y. and Majer, E.: Ground-penetrating-radar-assisted saturation and permeability estimation
735   in bimodal systems, Water Resour. Res., 33(5), 971–990, doi:10.1029/96WR03979, 1997.

Hubbard, S. S., Williams, K. H., Agarwal, D., Banfield, J., Beller, H., Bouskill, N., Brodie, E., Carroll, R., Dafflon, B., Dwivedi, D., Falco, N., Faybishenko, B., Maxwell, R., Nico, P., Steefel, C., Steltzer, H., Tokunaga, T., Tran, P. A., Wainwright, H. and Varadharajan, C.: The East River, Colorado, Watershed: A Mountainous Community Testbed for Improving Predictive Understanding of Multiscale Hydrological–Biogeochemical Dynamics, Vadose
740   Zo. J., 17(1), 0, doi:10.2136/vzj2018.03.0061, 2018.

Li, X., Li, Y., Chang, C. F., Tan, B., Chen, Z., Sege, J., Wang, C. and Rubin, Y.: Stochastic, goal-oriented rapid impact modeling of uncertainty and environmental impacts in poorly-sampled sites using ex-situ priors, Adv. Water Resour., doi:10.1016/j.advwatres.2017.11.008, 2018.

Loschko, M., Woehling, T., Rudolph, D. L. and Cirpka, O. A.: Cumulative relative reactivity: A concept for
745   modeling aquifer-scale reactive transport, Water Resour. Res., 52(10), 8117–8137, doi:10.1002/2016WR019080, 2016.

Maxwell, R. M., Kastenberg, W. E. and Rubin, Y.: A methodology to integrate site characterization information into groundwater-driven health risk assessment, Water Resour. Res., doi:10.1029/1999WR900103, 1999.

McClain, M. E., Boyer, E. W., Dent, C. L., Gergel, S. E., Grimm, N. B., Groffman, P. M., Hart, S. C., Harvey, J. W.,
750   Johnston, C. A., Mayorga, E., McDowell, W. H. and Pinay, G.: Biogeochemical Hot Spots and Hot Moments at the Interface of Terrestrial and Aquatic Ecosystems, Ecosystems, 6(4), 301–312, doi:10.1007/s10021-003-0161-9, 2003.

Miall, A. D.: Architectural-element analysis: A new method of facies analysis applied to fluvial deposits, Earth Sci.

Rev., doi:10.1016/0012-8252(85)90001-7, 1985.

Miall, A. D.: Facies Architecture in Clastic Sedimentary Basins., 1988.

755 Mitchell, C. P. J., Branfireun, B. A. and Kolka, R. K.: Spatial characteristics of net methylmercury production hot spots in peatlands, Environ. Sci. Technol., 42(4), 1010–1016, doi:10.1021/es0704986, 2008.

Moreno, L. and Tsang, C. -F: Flow channeling in strongly heterogeneous porous media: A numerical study, Water Resour. Res., doi:10.1029/93WR02978, 1994.

Ritzi, R. W., Dai, Z., Dominic, D. F. and Rubin, Y. N.: Spatial correlation of permeability in cross-stratified

760 sediment with hierarchical architecture, Water Resour. Res., doi:10.1029/2003WR002420, 2004.

Rubin, Y.: Prediction of tracer plume migration in disordered porous media by the method of conditional probabilities, Water Resour. Res., doi:10.1029/91WR00094, 1991.

Rubin, Y.: Applied Stochastic Hydrogeology, Oxford University Press, Oxford, UK., 2003.

Rubin, Y. and Dagan, G.: Conditional estimation of solute travel time in heterogeneous formations: Impact of

765 transmissivity measurements, Water Resour. Res., doi:10.1029/91WR02759, 1992.

Rubin, Y. and Journel, A. G.: Simulation of non-Gaussian space random functions for modeling transport in groundwater, Water Resour. Res., doi:10.1029/91WR00838, 1991.

Rubin, Y., Mavko, G. and Harris, J.: Mapping permeability in heterogeneous aquifers using hydrologic and seismic data, Water Resour. Res., doi:10.1029/92WR00154, 1992.

770 Salandin, P. and Fiorotto, V.: Solute transport in highly heterogeneous aquifers, Water Resour. Res., 34(5), doi:10.1029/98WR00219, 1998.

Sassen, D. S., Hubbard, S. S., Bea, S. A., Chen, J., Spycher, N. and Denham, M. E.: Reactive facies: An approach for parameterizing field-scale reactive transport models using geophysical methods, Water Resour. Res., 48(10), doi:10.1029/2011WR011047, 2012.

775 Schwarz, G.: Estimating the Dimension of a Model, Ann. Stat., doi:10.1214/aos/1176344136, 1978.

Shrestha, N. K. and Wang, J.: Current and future hot-spots and hot-moments of nitrous oxide emission in a cold climate river basin, Environ. Pollut., 239, 648–660, doi:10.1016/j.envpol.2018.04.068, 2018.

Sørensen, R., Zinko, U. and Seibert, J.: On the calculation of the topographic wetness index: Evaluation of different methods based on field observations, Hydrol. Earth Syst. Sci., doi:10.5194/hess-10-101-2006, 2006.

780 USEPA: Risk Assessment Guidance for Superfund., 2001.

Vidon, P., Allan, C., Burns, D., Duval, T. P., Gurwick, N., Inamdar, S., Lowrance, R., Okay, J., Scott, D. and Sebestyen, S.: Hot spots and hot moments in riparian zones: Potential for improved water quality management, J. Am. Water Resour. Assoc., 46(2), 278–298, doi:10.1111/j.1752-1688.2010.00420.x, 2010.

Wainwright, H. M., Orozco, A. F., Bucker, M., Dafflon, B., Chen, J., Hubbard, S. S. and Williams, K. H.:

785 Hierarchical Bayesian method for mapping biogeochemical hot spots using induced polarization imaging, Water Resour. Res., 51, 9127–9140, doi:10.1002/2014WR016259, 2015.

Wilson, A. and Rubin, Y.: Characterization of aquifer heterogeneity using indicator variables for solute concentrations, Water Resour. Res., 38(12), 19-1-19–12, doi:10.1029/2000wr000116, 2002.