Quantification of the information content of Darcy fluxes associated with hydraulic conductivity fields evaluated at diverse scales

Aronne Dell'Oca, Alberto Guadagnini, and Monica Riva

Department of Civil and Environmental Engineering, Politecnico di Milano, 20133, Milan, Italy;

Corresponding author: Aronne Dell'Oca (aronne.delloca@polimi.it)

Highlights

- Variations in the support/observation scale of hydraulic conductivity induce changes in the information content of the associated Darcy fluxes.
- Propagation of information about conductivity onto Darcy flux through the flow and mass balance equation can be quantified via Information Theory.
- Flux data at large scales provide more information about fluxes at small scales than what can be observed with reference to conductivities

Abstract

We rest on an Information Theory perspective and assess (*i*) the average information content and (*i*) the information shared between Darcy flux fields associated with fields of hydraulic conductivity (K) characterized by differing support (or measurement) scale. We treat hydraulic conductivity as a spatial random field, characterized by a given distribution and correlation structure. The latter is modeled through a truncated power law variogram (TPV), which explicitly takes into account a characteristic length scale of the support volume of K through a lower cutoff scale. We then frame our study in a numerical Monte Carlo context where groundwater flow is evaluated across a collection of realizations of hydraulic conductivity characterized by different values of the TPV parameters and subject to uniform in the mean flow. We quantify information through the Shannon entropy of the probability mass functions of the Darcy flux components as well as the mutual information and the multivariate mutual information respectively shared by pairs and triplets of Darcy flux components related to hydraulic conductivity fields evaluated at diverse scales and associated with various levels of heterogeneity. Partitioning of multivariate mutual information according to unique, redundant and synergetic contributions is also quantified. We found consistent trends (*i*) in the variation of the average information content with respect to the size of lower cutoff scale and (*ii*) in the way information is shared between pairs and triplets of Darcy flux components associated with diverse support scales of the underlying conductivities.

1.Introduction

Investigation, conceptualization and rendering of processes taking place within porous media are strongly influenced by and intimately tied to a variety of length scales. The latter are naturally the subject of various studies and analysis. For example, the size of the sampling window (or domain of investigation) is of relevance for studies conducted both at the pore (see e.g., Zhang et al., 2000; Puyguiraud et al., 2020) and at the laboratory or field scale (see e.g., Schad and Teutsch, 1994; Neuman, 1994; Neuman and Di Federico, 2003; Abidoye and Das, 2014). The strength of the spatial correlation (or degree of structural coherence) of the pore structure and/or of hydraulic properties (e.g., most notably hydraulic conductivity) is key to flow and solute transport phenomena (e.g., Porta et al., 2013, 2015; Siena et al., 2014, 2019; Meyer and Bijeljic, 2016; Wright et al., 2018; Comolli et al., 2019; Hyman et al., 2019 and references therein). Furthermore, the support scale (or data support) associated with porous media attributes is an important element to be taken into account for the analysis of a variety of processes taking place within porous media (e.g., Andersson et al., 1988; Neuman, 1995; Tidwell and Wilson, 1999a, b; Tartakovsky et al., 2004; Berkowitz et al. 2006; Tartakovsky et al., 2017; Icardi et al., 2019). It should also be noted that the relevance of a given length scale is typically tied to all length scales affecting the process under investigation (see e.g., Dagan, 1984; Chen et al., 2006; Meile and Tunkay, 2006; de Barros and Rubin, 2011; Dentz and de Barros, 2015; de Barros and Dentz, 2016; Moslehi et al., 2016; Moslehi and de Barros, 2017; Di Palma et al., 2017; de Barros, 2018; Wang et al., 2019).

In this broad context, our study is keyed to the analysis of the impact of the size of the support (or measurement) scale of hydraulic conductivity on the information content of the ensuing Darcy flux field. We do so by considering a stochastic approach within which we treat hydraulic conductivity as a spatial random field, characterized by a given distribution and correlation structure. We model the latter by explicitly taking into account the length scales associated with the size (*i*) of the domain where flow takes place and (*ii*) of the support volume (or data support) associated with available data. To do so we leverage on prior studies (see e.g., Di Federico and Neuman 1997, Neuman and Di Federico, 2003; Neuman et al., 2008) and model the spatial correlation structure of the natural logarithm of conductivity, Y', through a truncated power law variogram (TPV). The latter naturally arises upon considering geologic media as characterized by a continuum hierarchy of scales and is fully consistent with the documented behavior displayed by geostatistical parameters (i.e., variance and integral scale) inferred through typical analyses, which are seen to vary systematically with (*i*) the scale at which observations are taken and (*ii*) the length scale characterizing the domain of investigation. TPV models have the unique ability to capture these documented variations in terms of a few scaling parameters, which are tied to the length scales mentioned above. From a theoretical standpoint, a TPV is related to a view of Y' as a random fractal characterized by a power law variogram (PV). The latter has been shown (see Neuman and Di Federico, 2003 and references therein) to be constructed as an infinite hierarchy of second order stationary random fields, each characterized by a given (Exponential or Gaussian) variogram of characteristic length scale, λ. Truncation of such a hierarchy of random fields by way of a lower and upper cutoff scale (denoted as λ_l, and λ_u, respectively) yields a truncated version of the PV. In this context, λ_l is associated with the size of the support/measurement scale of log conductivity, whereas λ_u is linked to the characteristic length scale of the investigated domain. As such, the resulting (multiscale) TPV model can be used even in settings where the support/measurement scale and/or the structure of coarsening of observations are unknown a priori. These concepts have been used by Neuman et al. (2008) to develop and apply co-kriging equations enabling one to use information pertaining to a given support/measurement scale to predict the behavior of the system at a differing scale of interest.

Here, we focus on the lower cutoff scale, i.e., λ_l, and the way its value (which is associated with a given measurement/support scale for conductivity) can affect Darcy fluxes. As an example to frame the analysis, Fig.s 1a-d depict two-dimensional spatial distributions of log-conductivities related to four differing values of λ_l (hereafter, each of these is denoted by a subscript and they are ordered as $\lambda_l^1 < \lambda_l^2 < \lambda_l^3 < \lambda_l^4$; see also Section 2.2). Each of these exemplary fields is a random sample associated with the collection of fields generated according to the procedure described in Section 2.2 within the context of a numerical Monte Carlo framework. Fig.s 1e-h depict the logarithm of the module of the Darcy flux associated with the diverse values of λ_l. Inspection of Fig.s 1a-d reveals a decrease of the level of descriptive detail in the spatial field of Y' as λ_l increases. This is reflected by an increased degree of spatial uniformity of log-conductivity values. In other words, there is a loss of information about the spatial arrangement of log-conductivity values as λ_l increases (i.e., as its value shifts from λ_l^1 to λ_l^4). This pattern is then imprinted to the spatial distribution of Darcy flux magnitude (see Fig.s 1e-h), whose degree of spatial homogeneity is seen to increase with the value of λ_l. Fig.s 1i-m depict relative percentage differences between the module of Darcy flux related to λ_l^1 and its counterpart associated with increased values of λ_l (i.e., λ_l = λ_l^2, λ_l^3, λ_l^4). In general, as the difference in the size of the support scales increases there is an overall increase in the (absolute) values of these relative differences. It is noted that positive and negative values of the latter (see Fig.s i-m) tend to correspond to locations where values of the module of Darcy flux are low and high (see Fig.s f-h), respectively. This behavior is also reflected through the increased degree of spatial uniformity of the velocity field with increasing values of λ_l. One can also note that there are certain degrees of similarity between log-conductivity fields associated with the diverse values of λ_l, i.e., it is reasonable to assume that a given conductivity field can contain a certain amount of information about its counterparts characterized by differing values of λ_l. A similar behavior appears to be recognizable also in the fields of Darcy flux magnitude. In our study we aim at providing a quantitative description of these qualitative observations upon leveraging on elements of Information Theory (IT) (see e.g., Stone, 2015). Specifically, we aim at quantifying (*i*) how the (average) level of information about the hydraulic conductivity and about the Darcy flux field is affected by the support scale; and how information is shared between (*ii*) pairs and (*iii*) triplets of these investigated variables associated with diverse supports. We note that, even as we focus here on the impact that variations of the support/measurement scale of Y' can have on the ensuing Darcy flux field, the IT-based analysis illustrated in the following can (in principle) be extended and adapted to various contexts (including, e.g., upscaling/coarsening of flow and solute transport processes) to characterize the way information content (about some variables of interest) varies with characteristic length scale(s) and to quantify how information is shared between diverse variables of interest (e.g., solute concentrations rendered by a pore scale description of the system and its counterpart based on a continuum/Darcy scale model).

As compared to surface hydrology scenarios, Information Theory has been employed in a limited set of studies related to subsurface hydrology. Woodbury and Ulrych (1993, 1996, 2000) leverage on the principle of minimum relative entropy in the context of uncertainty propagation and inverse modeling. Kitanidis (1994) grounds the concept of dilution index, employed to describe chemical transport in heterogeneous aquifers, on the definition of entropy. In the context of subsurface systems analyses, Abellan and Noetinger (2010) define an optimal data acquisition strategy introducing an utility function grounded on the Kullback-Leibler divergence. Mishra et al. (2009) and Zeng et al. (2012) characterize global sensitivity of groundwater flow models through the concept of mutual information shared between pairs of model input(s) and output(s). Gotovac et al. (2010) employ the maximum entropy principle to characterize the probability distribution function of travel time of solute migrating through heterogeneous formations. Wellman and Regenaur-Lieb (2012) and Wellman (2013) characterize uncertainty affecting the geological structure of subsurface formations employing Shannon' entropy and mutual information concepts. Nowak and Guthke (2016) focus on the optimal experimental design to discriminate between various interpretative models to render the process of sorption of metals onto soils. Bianchi and Pedretti (2017, 2018) define a set of metrics, grounded on IT concepts, to characterize the heterogeneity of porous formations. Boso and Tartakovsky (2018) develop an IT-based framework to upscale/downscale the flow problem within heterogeneous porous systems. Butera et al. (2018) investigate the spatial dependence of observables related to flow and solute transport, focusing on the importance of non-linear effects. At the same time, IT has been used in a variety of studies associated with surface hydrology settings and covering a broad range of topics. For instance, IT elements have been employed to (*a*) quantify the quality of predictions of hydrological processes (e.g., Benedetti, 2010; Weijs et al., 2010), (*b*) design monitoring networks (e.g., Alfonso et al., 2010; Fahle et al., 2015), (*c*) assess data assimilation procedures (e.g., Nearing et al., 2013a, b), (*d*) benchmark model performance and model functioning (e.g., Nearing et al., 2016, 2018; Bennett et al., 2019; Ruddell et al., 2019), and (*e*) characterize the way information is shared between triplets of hydrological variables (e.g., Goodwell and Kumar, 2017). This marked variety of settings where IT concepts and evaluation tools are applied supports the flexibility of IT to provide insights in complex natural systems (Goodwell et al., 2020; Kumar and Gupta, 2020; Nearing et al., 2020; Perdigão et al., 2020, Weijs and Ruddell, 2020).

The rest of the work is organized as follows. In Section 2 we describe the groundwater flow set-up (Section 2.1), the geostatistical model (Section 2.2), and the IT-metrics (Section 2.3) we consider. Results of the analyses related to various degrees of system heterogeneity and lower cutoff scale characterizing hydraulic conductivities are illustrated in Section 3. Major conclusions are exposed in Section 4.

2.Methodology

2.1 Flow problem and set-up

We consider a two-dimensional domain of uniform side L = 300 (all quantities being here expressed in consistent units) where uniform (in the mean) groundwater flow takes place. The two spatial directions are identified by x (transverse to the mean flow direction) and y (aligned with the mean flow direction) (see Fig. 2). Steady-state Darcy-scale flow is governed by

$$\nabla \cdot \boldsymbol{q}(\boldsymbol{x}) = 0 ; \qquad \boldsymbol{q}(\boldsymbol{x}) = -\boldsymbol{K}(\boldsymbol{x}) \cdot \nabla h(\boldsymbol{x}) . \tag{1}$$

Here, $\boldsymbol{x}$ denotes the space coordinate vector, $\boldsymbol{q}(\boldsymbol{x})$ is Darcy flux vector (with components q_x and q_y), $h(\boldsymbol{x})$ is hydraulic head, and $\boldsymbol{K}(\boldsymbol{x})$ is hydraulic conductivity tensor. The latter is here taken as heterogenous and isotropic, i.e., $\boldsymbol{K}(\boldsymbol{x}) = K(\boldsymbol{x})\mathbf{I}$, $\mathbf{I}$ being the identity matrix. No-flow boundary conditions are imposed along the right and left boundaries, i.e., x = (0; L); a unit Darcy flux is set along the top boundary, i.e., $y = L$; hydraulic head is fixed at a constant (deterministically known) value, i.e., $h = h_{BC}$ across the bottom edge, i.e., at $y = 0$ (see Fig. 2).

Flow is solved numerically through a finite element approach (here, we rely on first order Raviart-Thomas elements), as coded within the FreeFEM++ environment (Hecht, 2012). The spatial domain is discretized through a structured and regular triangular mesh comprising 300 elements along each side of the domain (see Fig. 2).

2.2 Geostatistical modeling approach

Hydraulic conductivity is treated as a two-dimensional random field. The latter is described through a zero-mean random fluctuation, i.e.,

$$Y'(\boldsymbol{x}) = \ln\left(\frac{K(\boldsymbol{x})}{K_g}\right), \tag{2}$$

where K_g is the (spatially uniform) geometric mean of $K(\boldsymbol{x})$. We treat $Y'(\boldsymbol{x})$ as an unconditional multi-Gaussian random field, characterized by an isotropic truncated power law (TPV) variogram (see Neuman and Di Federico, 2003; Neuman et al., 2008 and references therein), i.e.,

$$\gamma_{Y'}^2\left(s;\lambda_l,\lambda_u\right)=\gamma^2(s;\lambda_u)-\gamma^2(s;\lambda_l)\,, \tag{3}$$

with

$$\gamma^2(s;\lambda_m)=\frac{A\lambda_m^{2H}}{2H}\left[1-\exp\left(-\frac{s}{\lambda_m}\right)+\left(\frac{s}{\lambda_m}\right)^{2H}\Gamma\left(1-2H,\frac{s}{\lambda_m}\right)\right]\qquad m=l,u\,. \tag{4}$$

Here, s is separation distance (or lag), A is a coefficient, $0<H\le 0.5$ is the Hurst exponent, $\Gamma(\cdot,\cdot)$ is the incomplete gamma function, λ_l and λ_u being termed as lower and upper cutoff scales, respectively. We recall that (*i*) λ_l and λ_u are associated with the measurement/support scale of hydraulic conductivity and with a characteristic length scale of the investigated domain, respectively and (*ii*) the TPV (3)-(4) is a weighted integral from λ_l to λ_u over a hierarchy of scales of exponential variograms associated with stationary random fields related to point support. Expression for the variance of Y' can be evaluated as (see Di Federico and Neuman, 1997)

$$\sigma_{Y'}^2\left(\lambda_l,\lambda_u\right)=\sigma_{Y'}^2(\lambda_u)-\sigma_{Y'}^2(\lambda_l)\,, \tag{5}$$

with

$$\sigma_{Y'}^2(\lambda_m)=\frac{A\left[\lambda_m\right]^{2H}}{2H}\qquad m=l,u\,, \tag{6}$$

while the integral scale associated with a TPV variogram reads

$$I_{Y'}\left(\lambda_l,\lambda_u\right)=\frac{2H}{1+2H}\frac{\lambda_u^{1+2H}-\lambda_l^{1+2H}}{\lambda_u^{2H}-\lambda_l^{2H}}\,, \tag{7}$$

Consistent with this modeling concept, the lower cutoff scale, λ_l, is here associated with the size of the support/measurement scale of hydraulic conductivity. Random spatial distributions of $Y'(\boldsymbol{x})$ are then generated by setting $H=0.15$ (corresponding to an antipersistent random field, where high and low values tend to alternate quite rapidly in space) and $\lambda_u=32$ (corresponding to $L/\lambda_u\approx$ 10). We explore the effect of various values of (*i*) the support/measurement scale associated with Y', as expressed in terms of the lower cutoff scale, i.e., λ_l = [1; 2; 4; 6] (corresponding to L/λ_l = [300; 150; 75; 50]) and (*ii*) the degree of system heterogeneity. The latter has been varied by considering three values of A, i.e., A = [0.082; 0.205; 0.41]. For conciseness of notation, we denote entries of λ_l = [1; 2; 4; 6] as λ_l^i with superscript i = 1, 2, 3, 4 and , in the following, we drop the dependence of the variance and of the integral scale on the upper cutoff scale, i.e., we write $\sigma_{Y'}^2(\lambda_l^i,\lambda_u)=\sigma_{Y'}^2(\lambda_l^i)$ and $I_{Y'}(\lambda_l^i,\lambda_u)=I_{Y'}(\lambda_l^i)$. The ensuing values of $\sigma_{Y'}^2(\lambda_l^i)$ and of $I_{Y'}(\lambda_l^i)$ are listed in Table 1. It is noted that, $\sigma_{Y'}^2(\lambda_l^i)$ decreases and $I_{Y'}(\lambda_l^i)$ increases as the lower cutoff scale of Y' increases. As an example, in Figure SM1 of the Supplementary Material we depict TPVs with A = 0.41 for various values of the lower cutoff scale. The pattern highlighted by these results is consistent the intuition that the degree of spatial variability of Y' tends to decrease as its support/measurement scale widens.

We note here that there is an extensive body of literature focusing on the upscaling of the flow problem (1) (see e.g., Renard and de Marsily, 1997; Wen and Gomez Hernandez, 1996; Chen et al., 2003; Noetinger et al., 2005; Sanchez-Vila et al., 2006; Boschan and Noetinger, 2012; Colecchio et al., 2020 and references therein). A common approach to this objective relies on applying an upscaling filter to the flow equation (1), thus viewing hydraulic head and conductivity as the sum of their upscaled counterparts (which are typically smoother in space than the reference values) and zero-mean fluctuations around these (see e.g., Dykaar and Kitanidis, 1992; Noetinger 2000; Attinger 2003; Eberhard et al., 2004). This typically leads to a flow equation for the upscaled head distribution which is associated with an increased level of complexity (including, e.g., the appearance of non-local integro-differential terms). Additional sets of equations are also required to solve the exact (upscaled) flow model. This issue is typically referred to as the closure problem. The (otherwise exact) upscaled equation can possibly be simplified (through, e.g., localization) leading to compact expressions for a so-called effective hydraulic conductivity distribution. We recall that, as the size of the coarsening (or upscaling) length scale increases, the ensuing effective hydraulic conductivity fields tend to be more spatially uniform. As the upscaling length increases, so does the level of spatial correlation of the resulting effective conductivity field, while its overall variance decreases. Such a pattern is similar to the effect that increasing the lower cutoff scale of the TPV has on the conductivity field.

It is also noted that in the context of the present study we aim at directly assessing the influence of variations in the support/observation scale of conductivity on the resulting Darcy flux fields as governed by (1) (i.e., we do not pursue an upscaling of the mathematical model describing the flow problem). In this sense, hydraulic conductivity drives a coarsening of the resolution in a manner consistent with typical measurement devices.

Table 1. Variance and integral scale of Y' for the values considered for coefficient A and the lower cutoff scales (λ_l^i) in (3)-(7).

		$\lambda_l^1 = 1$	$\lambda_l^2 = 2$	$\lambda_l^3 = 4$	$\lambda_l^4 = 6$
$A = 0.082$	$\sigma_{Y'}^2(\lambda_l^i)$	0.5	0.41	0.34	0.25
$A = 0.205$	$\sigma_{Y'}^2(\lambda_l^i)$	1.25	1.04	0.85	0.63
$A = 0.410$	$\sigma_{Y'}^2(\lambda_l^i)$	2.5	2.07	1.71	1.25
	$I_{Y'}(\lambda_l^i)$	11.3	12.7	14.8	18.1

We generate 1000 Monte Carlo samples for each combinations of A and λ_l values reported in table 1, which were sufficient to ensure convergence of the results illustrated in Section 3. Note that our collections of realizations corresponding to a given value of A and differing values of λ_l are designed upon preserving the initial seed number in the random generation process. This ensures that the n-th realizations of Y' included in the collection of Monte Carlo realizations with a given A and related to various λ_l differ solely as a result of the change in the lower cutoff scale of the underlying TPV. Hydraulic conductivity fields are generated at the centers of blocks forming a regular spatial grid composed of 300 square elements along each side of the domain. Numerical solution of the flow problem (1) is grounded on a grid composed by triangular elements (i.e., isosceles triangles) with a unit length base (see Fig. 2), each square block characterized by a given (generated) conductivity value being discretized by 2 triangles. This ensures that, for the smaller lower cutoff scale considered, approximately 11 grid elements are placed across each integral scale of the conductivity field, i.e., $I_{Y'}(\lambda_l^1)$ (see Table 1). This enables us to obtain a reliable representation of the heterogeneity of the hydraulic conductivity field, in terms of variogram, as well as a sufficiently accurate solution of the flow problem (1). As the lower cutoff scale increases, so does the number of mesh elements per integral scale (see Table 1), thus ensuring a satisfactory level of accuracy in the ensuing Darcy flux fields. As an additional test to assess the accuracy of our results with respect to the size of the numerical mesh employed, we evaluate the metrics introduced in Section 2.3 (and linked to the results illustrated in Section 3) considering a smaller domain (i.e., L = 150) and employing three numerical grids, characterized by an increased level of refinement obtained upon placing (*i*) 2, (*ii*) 8, and (*iii*) 18 triangles for each conductivity block, respectively. Considering, for example, the setting associated to the highest conductivity variance (i.e., A = 0.41, see Table 1) and results obtained with 100 Monte Carlo realizations, we find that the IT metrics we analyze (see Section 2.3 and 3) differ (at most) only by a few percentage points, the largest discrepancy being limited to about 6 - 7 % (for some metrics) when comparing grids (*i*) and (*iii*). The results of this analysis support the use of the selected numerical mesh as a compromise between numerical accuracy and computational cost.

Note that settings with $\lambda_l^1 = 1$ correspond here to test cases associated with the smallest value of the support available (as identified by the lower cutoff scale and corresponding to the size of the log-conductivity generation blocks). For this reason, in the following we quantify the reference level of system heterogeneity by $\sigma_{Y'}^2(\lambda_l^1)$, which corresponds to the fields characterized by the highest degree of descriptive details, $\sigma_{Y'}^2(\lambda_l^i)$ (with $i > 1$) being related to the (partial) knowledge of the (reference) heterogeneity one can capture by using measurement devices associated with increased support scales (λ_l^i).

2.3 Information Theory

The Shannon Entropy (Shannon, 1948) of the probability mass distribution of a discrete random variable X_T can be used to quantify the expected amount of information related to an event (or outcome) of X_T and is defined as

$$H(X_T) = \sum_{i=1}^{N} p_i \ln(p_i^{-1}), \tag{8}$$

where N is the number of bins employed to discretize the outcomes of X_T; p_i is the probability mass function, and $\ln(p_i^{-1})$ is the (so-called) information associated with the *i*-th bin. The quantity $\ln(p_i^{-1})$ in (8) corresponds to the degree of surprise for an outcome of X_T to be in the *i*-th bin, i.e., the higher (lower) the value of p_i, the lower (higher) the associated surprise for an outcome related to the *i*-th bin. Note that in this study we rely on the natural base for the logarithm in (8), thus leading to *nats* as unit of measure for Shannon entropy and for all of the IT metrics described in the following, other choices being fully compatible with our framework of analysis (e.g., logarithm in base two, leading to *bits*). The Shannon entropy can be interpreted as a metric quantifying the uncertainty associated with X_T, i.e., $H(X_T)$ is largest and equal to $\ln(N)$ in case p_i is uniform across all bins (i.e., $p_i = 1/N$), while vanishing when outcomes of X_T fall only within a single bin. In our study we identify the random quantity X_T with either (*i*) one of Darcy flux components or (*ii*) Y'. Thus, Shannon entropy can also be interpreted as a measure of the degree of spatial heterogeneity of a target quantity, i.e., in case samples of a given quantity collected across the physical domain of interest are associated with values that fall into one (or only a few) bin(s) (i.e., $H(X_T) \approx 0$), this can be viewed as a signature of spatial homogeneity. Fig. 3*a* depicts the concept of Shannon entropy relying on Venn diagrams, i.e., the radius of a circle in Fig. 3*a* is proportional to $H(X_T)$.

Given two random quantities, i.e., X_T and X_{S_1}, the information shared between them is quantified by the bivariate mutual information, i.e.,

$$I(X_{S_1};X_T)=\sum_{i=1}^{N}\sum_{j=1}^{M}p_{i,j}\ln\left(\frac{p_{i,j}}{p_i p_j}\right), \tag{9}$$

where N and M are the number of bins associated with X_T and X_{S_1}, respectively; p_i and p_j are the marginal probability mass function of X_T and X_{S_1}, respectively; and $p_{i,j}$ is the joint probability mass function of X_T and X_{S_1}. The bivariate mutual information (9) quantifies the average reduction of uncertainty (as defined through the Shannon entropy) about one variable stemming from knowledge on the other variable (see, e.g., Gong et al., 2013 and references therein). In this sense, the bivariate mutual information shared between two variables represents a reduction of uncertainty. For example, one can then see that $I(X_{S_1};X_T)$ (*a*) is null for two independent variables or (*b*) is equal to the entropy of either X_T or X_{S_1}, i.e. $H(X_T)=H(X_{S_1})=I(X_{S_1};X_T)$, if one variable fully explains the other one. Note that $I(X_{S_1};X_T)$ is symmetric, i.e., the amount of information that X_T shares with X_{S_1} is equal to that shared by the latter with the former. Fig. 3*b* depicts the concept of mutual information between X_T and X_{S_1} in terms of Venn diagrams, i.e., the overlapping region is proportional to $I(X_{S_1};X_T)$.

Considering a triplet of discrete random variables, it is possible to quantify the amount of information that two of these (hereafter identified as sources, i.e., X_{S_1} and X_{S_2}) provide to the third one (identified as target variable, i.e., X_T) through the multivariate mutual information, i.e.,

$$I(X_{S_1},X_{S_2};X_T)=\sum_{i=1}^{N}\sum_{j=1}^{M}\sum_{k=1}^{W}p_{i,j,k}\ln\left(\frac{p_{i,j,k}}{p_{i,j}p_k}\right). \tag{10}$$

Here, *N*, *M*, and *W* represent the number of bins associated with X_{S_1}, X_{S_2} and X_T, respectively; p_k is the marginal probability mass function of X_T; $p_{i,j}$ is the joint probability mass function of X_{S_1} and X_{S_2}; and $p_{i,j,k}$ is the joint probability mass function of X_{S_1}, X_{S_2}, and X_T. Here, we follow a typically employed notation according to which (*i*) the symbol ';' is used to demarcate the sets of variables that share information (e.g., (X_{S_1}) and (X_T) in (9) are the two sets sharing information; while (X_{S_1},X_{S_2}) and (X_T) are such sets in (10)) and (*ii*) symbol ',' is used as a separator in the list of variables belonging to the same set (e.g., (X_{S_1},X_{S_2}) in (10)). We note that the multivariate mutual information shared by the source variables with the target variable is also a measure of the average reduction in the uncertainty of the latter due to the simultaneous knowledge of the sources. For example, there is no multivariate mutual information shared between the sources and the target, i.e., $I(X_{S_1},X_{S_2};X_T)=0$, and there is no reduction in the uncertainty about X_T by the knowledge of both X_{S_1} and X_{S_2} when X_T is independent from both X_{S_1} and X_{S_2}. Otherwise, if the value of X_T is determined by only that of X_{S_1} and X_{S_2} the multivariate mutual information in (10) is equal to the Shannon entropy of X_T, i.e., $I(X_{S_1},X_{S_2};X_T)=H(X_T)$, so that the simultaneous knowledge of both sources allows determining the value of the target variable.

Considering partial information decomposition or information partitioning concepts (see e.g., Williams and Beer, 2010), $I(X_{S_1}, X_{S_2}; X_T)$ can be partitioned/decomposed into unique, redundant, and synergetic contributions, i.e.,

$$I(X_{S_1}, X_{S_2}; X_T) = U(X_{S_1}; X_T) + U(X_{S_2}; X_T) + R(X_{S_1}, X_{S_2}; X_T) + S(X_{S_1}, X_{S_2}; X_T)\,. \tag{11}$$

Here, $U(X_{S_1}; X_T)$ and $U(X_{S_2}; X_T)$ represent the information that is uniquely provided to the target X_T by X_{S_1} and X_{S_2}, respectively (i.e., the information $U(X_{S_1}; X_T)$ is uniquely provided to X_T by knowledge on X_{S_1} and cannot be provided by knowledge on X_{S_2}, a corresponding observation holding for $U(X_{S_2}; X_T)$); the redundant contribution $R(X_{S_1}, X_{S_2}; X_T)$ is the information that both sources provide to the target, the emergence of some redundancy of information being typically expected when the two sources are correlated; and the synergetic contribution $S(X_{S_1}, X_{S_2}; X_T)$ is the information about X_T that the simultaneous knowledge on X_{S_1} and X_{S_2} (possibly) brings in a synergic way. The synergetic contribution emerges when the information provided to the target by the '*whole*', i.e., considering X_{S_1} and X_{S_2} simultaneously, is larger than the information provided by considering the sum of the '*parts*', i.e., summing the information provided to the target by the X_{S_1} and X_{S_2} individually (see e.g., Griffith and Koch, 2014). Note that all components in (11) are positive (Williams and Beer, 2010). Fig. 3*c* illustrates these concepts upon relying on Venn diagrams.

The bivariate mutual information shared by X_T and each source is then seen as

$$\begin{aligned} I(X_{S_1}; X_T) &= U(X_{S_1}; X_T) + R(X_{S_1}, X_{S_2}; X_T) \\ I(X_{S_2}; X_T) &= U(X_{S_2}; X_T) + R(X_{S_1}, X_{S_2}; X_T) \end{aligned}, \tag{12}$$

where the information shared between each source and the target variable can be composed only by the corresponding unique and redundant contributions, as it is information shared with the target when the sources are taken separately (i.e., no synergy may emerge).

One can then define the interaction information, i.e.,

$$I(X_{S_1}; X_{S_2}; X_T) = I(X_{S_1}; X_T \mid X_{S_2}) - I(X_{S_1}; X_T) = I(X_{S_2}; X_T \mid X_{S_1}) - I(X_{S_2}; X_T)\,. \tag{13}$$

Here, $I(X_{S_i}; X_T \mid X_{S_j})$ is the bivariate mutual information shared by source X_{S_i} (i =1, 2) and the target, conditional to the knowledge of the other source X_{S_j} (j = 2, 1). We remark that the interaction information is the information shared by three sets of variables, each set being here demarcated through ';' according to our notation, and comprising a single variable, i.e., (X_{S_1}), (X_{S_2}), or (X_T). The quantity $I(X_{S_i}; X_T \mid X_{S_j})$ is evaluated through (10) and employing the conditional probability mass function for X_T. Williams and Beer (2010) show that

$$I(X_{S_1}; X_{S_2}; X_T) = S(X_{S_1}, X_{S_2}; X_T) - R(X_{S_1}, X_{S_2}; X_T)\,. \tag{14}$$

Thus, interaction information could be either negative, i.e., the amount of redundant information provided by the two sources overcomes the synergetic effects, or positive, i.e., the synergetic interactions due to the simultaneous knowledge on the two sources provides more information (to the target) than what is redundantly provided by the two sources.

Evaluation of all components in (11) requires an additional equation (further to (12)-(14)) and various strategies have been proposed to this end (e.g., Williams and Beer, 2010; Harder et al., 2013; Bertschinger et al., 2014; Griffith and Koch, 2014; Olbrich et al., 2015; Griffith and Ho, 2015). In this study we rest on the recent partitioning strategy proposed by Goodwell and Kumar (2017), in light of its ability to consider possible dependences between sources when evaluating the unique and redundant contributions. The rationale at the heart of this approach is that (*i*) even as the two sources are correlated, each can provide a unique contribution of information to the target, and (*ii*) redundancy should be lowest in case of independent sources. The redundant contribution is then evaluated as

$$R(X_{S_1}, X_{S_2}; X_T) = R_{\min}(X_{S_1}, X_{S_2}; X_T) + I_s\left[R_{MMI}(X_{S_1}, X_{S_2}; X_T) - R_{\min}(X_{S_1}, X_{S_2}; X_T)\right] \tag{15a}$$

with

$$\begin{aligned} &R_{\min}(X_{S_1}, X_{S_2}; X_T) = \max\left[0, -I(X_{S_1}; X_{S_2}; X_T)\right]; \\ &R_{MMI}(X_{S_1}, X_{S_2}; X_T) = \min\left[I(X_{S_2}; X_T), I(X_{S_1}; X_T)\right]; \\ &I_s = \frac{I(X_{S_1}; X_{S_2})}{\min\left[H(X_{S_1}), H(X_{S_2})\right]}. \end{aligned} \tag{15b}$$

Goodwell and Kumar (2017) propose (15) as a rescaled measure of redundancy whereas (*i*) $R_{\min}(X_{S_1}, X_{S_2}; X_T)$ is the lowest bound for redundancy, following the rationale that the minimum value of redundancy must at least be equal to $-I(X_{S_1}; X_{S_2}; X_T)$ in case $I(X_{S_1}; X_{S_2}; X_T) < 0$, also guaranteeing the positiveness of the synergy, see (14)) (*ii*) $R_{MMI}(X_{S_1}, X_{S_2}; X_T)$ is the upper bound for redundancy, according to the idea that the weakest source provides only redundant information; and (*iii*) I_s accounts for the degree of dependence between the sources, i.e., $I_s = 0$ and $R(X_{S_1}, X_{S_2}; X_T) = R_{\min}(X_{S_1}, X_{S_2}; X_T)$ for independent sources, while $I_s = 1$ and redundancy in (12) attains its upper limit value, $R_{MMI}(X_{S_1}, X_{S_2}; X_T)$, in case of a *complete* dependency between the sources. After redundancy is evaluated through (15), all remaining contributions in (11) can be determined with (12)-(14).

It is important to observe that, despite some additional complexities, exploring information partitioning yields valuable insights on the way information is shared among diverse variables. As a final remark, we note that the theoretical elements summarized here refer to discrete variables. While corresponding counterparts for continuous variables are available, these are associated with a less intuitive and immediate interpretation (e.g., entropy could be negative, see e.g., Stone, 2015).

3. Results

The results illustrated in this section are grounded on the analyses of the values of the Darcy flux components stemming from the collections of MC-based conductivity fields described in Section 2.2 and sampled within an inner region of the domain that is identified upon disregarding simulation results within strips of width $l_c = 50$ (i.e., for $\lambda_l = [1;\ 2;\ 4;\ 6]$, respectively) along each edge (see Fig. 2), to avoid effects of the imposed boundaries conditions (this has been verified upon increasing l_c from 0.0 to 50, with increments of 5, finding negligible discrepancies in the results after $l_c = 40$).

Evaluation of the Shannon entropy in (8) is obtained by pooling all samples of each target variable in a unique set, i.e., we quantify the average information content (i.e., Shannon entropy) associated with observations of a given variable collected within the region of the physical domain of interest identified as above. We follow the same rationale in the evaluation of the bivariate and multivariate mutual information, as well as its partitioning. For example, considering the bivariate mutual information shared by two variables that are associated with two distinct lower cutoff scales (e.g., $\lambda_l^1 = 1$ and $\lambda_l^2 = 2$), we define a two-column vector of samples in which each pair of entries correspond to: (*i*) the first variable of interest sampled at a given spatial location (e.g., $\boldsymbol{x} = (x^*; y^*)$) and in a specific MC realization (e.g., the 100th) within the set of realizations related to a given lower cutoff scale (e.g., $\lambda_l^1 = 1$) and (*ii*) the second variable of interest sampled at same spatial location (e.g., $\boldsymbol{x} = (x^*; y^*)$) considering the corresponding MC realization (e.g., the 100th) within the collection associated with the other lower cutoff scale (e.g., $\lambda_l^2 = 2$). Leveraging on this two-dimensional vector of samples, we evaluate the bivariate probability mass function in (9) and quantify the bivariate mutual information shared between two variables that have been collected over a given region of the physical domain of interest. A corresponding procedure is also considered for the evaluation of the multivariate mutual information (note that here the vector of samples is structured across three columns).

For each type of variable (i.e., either Darcy flux component or hydraulic conductivity) binning is performed upon discretizing through 100 bins of uniform width the range delimited by the lowest and largest value obtained considering the results associated with the lowest λ_l, i.e., λ_l^1 (which gives rise to the largest range of variation of the results; not shown). The same binning is then employed to discretize the samples of the corresponding variables associated with a diverse lower cutoff scale. This binning design facilitates the assessment of the impact of λ_l, in terms of the investigated IT metrics. An analysis on the robustness of the results with respect to the number of (*i*) bins (upon varying these from 40 to 200, according to regular increments of 10) and (*ii*) number of Monte Carlo realizations (which are varied from 100 to 1000, through regular increments of 100) yields only negligible variations of the results (with relative percentage errors of the order of a few percent, depending on the investigated quantity).

Note that, we introduce a superscript to clearly identify (when needed) the corresponding lower cutoff scale of the MC-based sample of a quantity (e.g., $q_y^{\lambda_l^i}$ is the longitudinal component of Darcy flux associated with Y' fields generated by setting the lower cutoff scale equal to λ_l^i).

Fig. 4*a* depicts the Shannon entropy of $q_y^{\lambda_l^i}$ for the set of four λ_l^i considered, and various degrees of system heterogeneity as quantified by $\sigma_{Y'}^2(\lambda_l^1)$ = 0.5 (red symbols), 1.25 (blue symbols), 2.5 (black symbols). Note that values of the Shannon entropies included in Fig. 4*a* are normalized by the Shannon entropy of $q_y^{\lambda_l^1}$, i.e., $H^*(q_y^{\lambda_l^i}) = H(q_y^{\lambda_l^i}) / H(q_y^{\lambda_l^1})$, to facilitate comparison among the results. Fig. 4*a* also depicts the corresponding (normalized) Shannon entropy for the log-conductivity fields $Y'^{\lambda_l^i}$, $H^*(Y'^{\lambda_l^i}) = H(Y'^{\lambda_l^i}) / H(Y'^{\lambda_l^1})$ (green symbols and dashed curve). One can see that $H^*(Y'^{\lambda_l^1})$ does not depend on $\sigma_{Y'}^2(\lambda_l^1)$. This behavior descends from the observation that (*i*) variations of $\sigma_{Y'}^2(\lambda_l^1)$ only contribute to rescale the observed lower and upper limits of the range of variability of $Y'^{\lambda_l^1}$ and (*ii*) a consistent binning rule (see details above) is employed for all values of $\sigma_{Y'}^2(\lambda_l^1)$. As a consequence, the discretized probability mass functions of $Y'^{\lambda_l^i}$ associated with differing values of $\sigma_{Y'}^2(\lambda_l^1)$ are similar for a given λ_l^i, their entropies being virtually indistinguishable. Inspection of Fig. 4*a* indicates that $H^*(q_y^{\lambda_l^i})$ decreases as λ_l^i increases, i.e., there is a homogenization of the values of

$q_y^{\lambda_l^i}$ as the lower cutoff scale of the TPV increases. This result is consistent with the decreasing trend of $H^*(Y'^{\lambda_l^i})$ with λ_l^i. Our results also show a generally similar qualitative pattern of the dependence of $H^*(q_y^{\lambda_l^i})$ and of $H^*(Y'^{\lambda_l^i})$ on λ_l^i for all values of $\sigma^2_{Y'}(\lambda_l^1)$. Otherwise, one can note that the rate at which $H^*(q_y^{\lambda_l^i})$ decreases with increasing lower cutoff tends to be more marked for the set of fields characterized by the largest $\sigma^2_{Y'}(\lambda_l^1)$ (i.e., $\sigma^2_{Y'}(\lambda_l^1)$ = 2.5) than for its low and moderate counterparts (i.e., $\sigma^2_{Y'}(\lambda_l^1)$ = 0.5 and 1.25). This suggests that the degree of spatial homogenization of the values of $q_y^{\lambda_l^i}$ tends to be enhanced as λ_l^i is increased in the presence of a high level of the reference system heterogeneity.

Despite the observed reduction with λ_l^i, values of $H^*(q_y^{\lambda_l^i})$ and $H^*(Y'^{\lambda_l^i})$ remain higher than 0.9 in all cases here analyzed, i.e., there is not a dramatic reduction of the average information in the $Y'^{\lambda_l^i}$ and $q_y^{\lambda_l^i}$ fields as λ_l^i increases (for the set of parameters here considered).

We then proceed to assess the way the information content of the $q_y^{\lambda_l^i}$ (with $i > 1$) fields relates to that of $q_y^{\lambda_l^1}$, which is the one corresponding to the richest degree of descriptive details. We do so by relying on Fig. 4*b* which depicts the bivariate mutual information shared between $q_y^{\lambda_l^1}$ and $q_y^{\lambda_l^i}$ normalized by the Shannon entropy of $q_y^{\lambda_l^1}$, i.e., $I^*(q_y^{\lambda_l^1};q_y^{\lambda_l^i}) = I(q_y^{\lambda_l^1};q_y^{\lambda_l^i}) / H(q_y^{\lambda_l^1})$, for the various degrees of system heterogeneity, i.e., associated with $\sigma^2_{Y'}(\lambda_l^1)$ = 0.5 (red symbols), 1.25 (blue symbols), 2.5 (black symbols). Corresponding results for $Y'^{\lambda_l^i}$ are also included (green symbols). Values of $I^*(q_y^{\lambda_{l,1}};q_y^{\lambda_{l,i}})$ represent the fraction of the information of the reference field, i.e., $q_y^{\lambda_l^1}$, that is also included in the fields related to larger lower cutoff scales. As such, one can interpret a given value of $I^*(q_y^{\lambda_l^1};q_y^{\lambda_l^i})$ as a measure of the level of *representativeness* of $q_y^{\lambda_l^i}$ with respect to $q_y^{\lambda_l^1}$. When $I^*(q_y^{\lambda_l^1};q_y^{\lambda_l^i}) = 1$, $q_y^{\lambda_l^i}$ is totally representative of $q_y^{\lambda_l^1}$, i.e., $q_y^{\lambda_l^i}$ contains all of the information about $q_y^{\lambda_l^1}$ (in terms of Venn diagrams, see Section 2.3 and Fig. 3*b*, this case would correspond to a scenario where the circle associated with $X_{S_1} = q_y^{\lambda_l^i}$ coincides with that of $X_T = q_y^{\lambda_l^1}$). From a physical standpoint, $I^*(q_y^{\lambda_l^1};q_y^{\lambda_l^i})$ can be interpreted as an average measure of the possibility to rend the variability of the values of $q_y^{\lambda_l^1}$ (across the set of Monte Carlo samples) through the variability of $q_y^{\lambda_l^i}$ (e.g., when $I^*(q_y^{\lambda_l^1};q_y^{\lambda_l^i}) = 0$, values of $q_y^{\lambda_l^i}$ in each realization are completely independent from those of $q_y^{\lambda_l^1}$, thus resulting in the lack of *representativeness*). For completeness, Fig. 4*b* also reports values of $1 - I^*(q_y^{\lambda_l^1};q_y^{\lambda_l^i})$ (right axis), which is a measure of the amount of information about $q_y^{\lambda_l^1}$ that is not captured by $q_y^{\lambda_l^i}$. Following the terminology of Gong et al. (2013) and Nearing et al. (2018), we refer to $1 - I^*(q_y^{\lambda_l^1};q_y^{\lambda_l^i})$ as a measure of *uncertainty* (in terms of Venn diagrams, there is *uncertainty* when there is a fraction of the circle associated with $X_T = q_y^{\lambda_l^1}$ that is not covered by that of $X_{S_1} = q_y^{\lambda_l^i}$, see Fig. 3*b*). From a physical standpoint, $1 - I^*(q_y^{\lambda_l^1};q_y^{\lambda_l^i})$ can be viewed as the average mismatch resulting from attempting to represent the variability of $q_y^{\lambda_l^1}$ (across the set of Monte Carlo realizations) through $q_y^{\lambda_l^i}$.

Fig. 4*b* suggests that the fields of $q_y^{\lambda_l^i}$ exhibit a continuously decreasing level of *representativeness* with respect to the reference field as λ_l^i increases, i.e., $I^*(q_y^{\lambda_l^1};q_y^{\lambda_l^i})$ decreases with λ_l^i. Obviously, this is mirrored by the corresponding increase of the *uncertainty* about $q_y^{\lambda_l^1}$ when the latter is approximated through $q_y^{\lambda_l^i}$, i.e., $1-I^*(q_y^{\lambda_l^1};q_y^{\lambda_l^i})$ increases with λ_l^i. Inspection of the results for the various values of $\sigma_{Y'}^2(\lambda_l^1)$ considered reveals a consistency in the trends of $I^*(q_y^{\lambda_l^1};q_y^{\lambda_l^i})$ (and of $1-I^*(q_y^{\lambda_l^1};q_y^{\lambda_l^i})$). A similar decreasing trend is detected for $I^*(Y'^{\lambda_l^1};Y'^{\lambda_l^i})$ (and for $1-I^*(Y'^{\lambda_l^1};Y'^{\lambda_l^i})$), whereas values of mutual information related to $Y'^{\lambda_l^i}$ are smaller than their counterparts related to $q_y^{\lambda_l^i}$. This result suggests that propagation of information about $Y'^{\lambda_l^i}$ onto $q_y^{\lambda_l^i}$ through (1) tends to enhance the possibility that knowledge of $q_y^{\lambda_l^i}$ (with i > 1) provides information about $q_y^{\lambda_l^1}$, as compared to what can be observed with reference to hydraulic conductivity.

We further note here that, due to the normalization employed, metrics such as $I^*(q_y^{\lambda_l^1};q_y^{\lambda_l^i})$ and $1-I^*(q_y^{\lambda_l^1};q_y^{\lambda_l^i})$ place emphasis on the field associated with the smallest lower cutoff scale, i.e., $q_y^{\lambda_l^1}$, in the sense that they assess how each of the various $q_y^{\lambda_l^i}$ (with i > 1) fields is informative (or not) with respect to $q_y^{\lambda_l^1}$. It is also of interest to quantify the fraction of information in $q_y^{\lambda_l^i}$ that is shared with $q_y^{\lambda_l^1}$. We do so by evaluating $I^{**}(q_y^{\lambda_l^1};q_y^{\lambda_l^i}) = I(q_y^{\lambda_l^1};q_y^{\lambda_l^i})/H(q_y^{\lambda_l^i})$ and we refer to $I^{**}(q_y^{\lambda_l^1};q_y^{\lambda_l^i})$ as a measure of *efficiency* of $q_y^{\lambda_l^i}$. When $I^{**}(q_y^{\lambda_l^1};q_y^{\lambda_l^i}) = 1$ a total *efficiency* is obtained, i.e., all the information in $q_y^{\lambda_l^i}$ is shared with $q_y^{\lambda_l^1}$. In terms of Venn diagrams (see Section 2.3 and Fig. 3*b*) this case would correspond to having the circle associated with $X_{S_1} = q_y^{\lambda_l^i}$ completely included in that of $X_T = q_y^{\lambda_l^1}$, the former being then smaller that the latter, i.e., $H(q_y^{\lambda_l^i}) \le H(q_y^{\lambda_l^1})$. We also evaluate $1-I^{**}(q_y^{\lambda_l^1};q_y^{\lambda_l^i})$, the latter being a measure of the fraction of information in $q_y^{\lambda_l^i}$ that is not pertinent to $q_y^{\lambda_l^1}$, and we refer to $1-I^{**}(q_y^{\lambda_l^1};q_y^{\lambda_l^i})$ as *inefficiency* (in terms of Venn diagrams there is *inefficiency* when the circle associated with $X_{S_1} = q_y^{\lambda_l^i}$ is not completely immersed within that of $X_T = q_y^{\lambda_l^1}$). Fig. 4*c* depicts $I^{**}(q_y^{\lambda_l^1};q_y^{\lambda_l^i})$ (left axis) and $1-I^{**}(q_y^{\lambda_l^1};q_y^{\lambda_l^i})$ (right axis) for all systems analyzed. For completeness, corresponding values associated with $Y'^{\lambda_l^i}$ are depicted (green symbols). Inspection of these results reveals that the *efficiency*, i.e., $I^{**}(q_y^{\lambda_l^1};q_y^{\lambda_l^i})$, of $q_y^{\lambda_l^i}$ to contribute to knowledge of $q_y^{\lambda_l^1}$ decreases with λ_l^i. At the same time, the *inefficiency* (i.e., $1-I^{**}(q_y^{\lambda_l^1};q_y^{\lambda_l^i})$) of the $q_y^{\lambda_l^i}$ fields increases with λ_l^i. The overall patterns displayed by these results is similar to what can be noted in Fig. 4*b* (i.e., general similarity in the decreasing trends depicted and values of mutual information related to $Y'^{\lambda_l^i}$ smaller than their counterparts linked to $q_y^{\lambda_l^i}$) and corresponding observations hold.

Joint inspection of Fig. 4*a* and of Fig.s 4*b-c* reveals that, even as there is a moderate decrease of the information content of $q_y^{\lambda_l^i}$ as the lower cutoff scale increases (see Fig. 4*a*), the decrease of the *representativeness* (i.e., $I^*(q_y^{\lambda_l^1};q_y^{\lambda_l^i})$) and the corresponding increase of the *uncertainty* (i.e., $1-I^*(q_y^{\lambda_l^1};q_y^{\lambda_l^i})$), as well as the decrease of the *efficiency* (i.e., $I^{**}(q_y^{\lambda_l^1};q_y^{\lambda_l^i})$) and the increase of *inefficiency* (i.e., $1-I^{**}(q_y^{\lambda_l^1};q_y^{\lambda_l^i})$) are quite marked. From a practical perspective, these findings suggest that: (*i*) if one is interested to the quantification of the degree of variability of $q_y^{\lambda_l^i}$ (as rendered through $H^*(q_y^{\lambda_l^i})$), replacing the results associated with λ_l^1 by way of those related to larger lower cutoff scales does not yield marked differences; (*ii*) otherwise, if it is relevant that local values (and not only the global information rendered in terms of entropy) of $q_y^{\lambda_l^i}$ are preserved as the lower cutoff scale increases (e.g., if the ensuing flow field are then employed for solute transport studies), one should be cautious in relying on conductivity fields associated with increased values of the lower cutoff scale, due to the marked decrease of the *representativeness* and of the *efficiency* observed with increasing such a scale. From a physical standpoint, these results suggest that, even as $q_y^{\lambda_l^1}$ and its various counterparts $q_y^{\lambda_l^i}$ are characterized by (approximately) the same level of variability (see e.g., Fig. 1e-h), there is a nonnegligible mismatch between their local values and such a mismatch increases with λ_l^i. These observations are consistent with Fig.s 1i-m, which document (absolute) values of relative percentage difference between Darcy flux modules that increase with λ_l^i.

The results illustrated above are focused on the analysis of the information content of $q_y^{\lambda_l^i}$ and on aspects about the way information is shared between pairs of $q_y^{\lambda_l^i}$ (or $Y'^{\lambda_l^i}$) related to differing values of λ_l. We consider now the analyses of triplets of $q_y^{\lambda_l^i}$, each associated with a given lower cutoff scale. In this context, relevant research and practical questions include 'How is information on $q_y^{\lambda_l^i}$ related to two support scales shared with corresponding fluxes associated with a third support? Would it be more beneficial to consider $q_y^{\lambda_l^i}$ related to two, rather than to just one, lower cutoff scales in terms of information shared with a third $q_y^{\lambda_l^i}$?'. Answers to these kinds of questions can be provided upon leveraging on the information partitioning framework detailed in Section 2.3.

Fig. 5 depicts the results of the information partitioning considering (*i*) the triplet $(q_y^{\lambda_l^{i+1}}, q_y^{\lambda_l^{i+2}}; q_y^{\lambda_l^i})$ (with $i = 1, 2$) and (*ii*) the triplet $(q_y^{\lambda_l^{i-2}}, q_y^{\lambda_l^{i-1}}; q_y^{\lambda_l^i})$ (with $i = 3, 4$). For ease of comparison between the results, we normalize the unique, synergetic and redundant contributions by the multivariate mutual information of the corresponding triplet, i.e., $U^*(q_y^{\lambda_l^{i+j}}, q_y^{\lambda_l^i}) = U(q_y^{\lambda_l^{i+j}}, q_y^{\lambda_l^i}) \,/\, I(q_y^{\lambda_l^{i+1}}, q_y^{\lambda_l^{i+2}}; q_y^{\lambda_l^i})$ with $j = 1, 2$; $R^*(q_y^{\lambda_l^{i+1}}, q_y^{\lambda_l^{i+2}}; q_y^{\lambda_l^i}) = R(q_y^{\lambda_l^{i+1}}, q_y^{\lambda_l^{i+2}}; q_y^{\lambda_l^i}) \,/\, I(q_y^{\lambda_l^{i+1}}, q_y^{\lambda_l^{i+2}}; q_y^{\lambda_l^i})$; and

$S^*(q_y^{\lambda_l^{i+1}}, q_y^{\lambda_l^{i+2}}; q_y^{\lambda_l^{i}}) = S(q_y^{\lambda_l^{i+1}}, q_y^{\lambda_l^{i+2}}; q_y^{\lambda_l^{i}}) \ / \ I(q_y^{\lambda_l^{i+1}}, q_y^{\lambda_l^{i+2}}; q_y^{\lambda_l^{i}})$. Fig. 5 depicts results for the cases related to $\sigma^2_{Y'}(\lambda_l^1) = 2.5$, similar results being observed for the other systems analyzed (see SM1).

Results of Fig. 5*a* suggest that, considering $q_y^{\lambda_l^{i+1}}$ and $q_y^{\lambda_l^{i+2}}$: (*i*) most of the information that these two variables provide to $q_y^{\lambda_l^{i}}$ is redundant; (*ii*) only the unique contribution associated with $q_y^{\lambda_l^{i+1}}$ is non-negligible; (*iii*) the unique contribution of $q_y^{\lambda_l^{i+2}}$ and (*iv*) the synergetic contribution of $q_y^{\lambda_l^{i+1}}$ and $q_y^{\lambda_l^{i+2}}$ are practically null. These results indicate that if one would represent $q_y^{\lambda_l^{i}}$ through its counterpart observed at a larger scale, i.e., $q_y^{\lambda_l^{i+2}}$, upon disregarding $q_y^{\lambda_l^{i+1}}$, there is a significant amount of information that is still retained (i.e., the redundant contribution), due to the level of similarity between $q_y^{\lambda_l^{i+2}}$ and $q_y^{\lambda_l^{i+1}}$ with $q_y^{\lambda_l^{i}}$ (see also Fig.s 1*e*-*h*). At the same time, there is also a non-negligible amount of information that will be lost (corresponding to the unique contribution associated with $q_y^{\lambda_l^{i+1}}$). From a physical standpoint, this result suggests that fluxes $q_y^{\lambda_l^{i+1}}$ are more similar to $q_y^{\lambda_l^{i}}$ than $q_y^{\lambda_l^{i+2}}$ (i.e., $U^*(q_y^{\lambda_l^{i+1}}, q_y^{\lambda_l^{i}}) > U^*(q_y^{\lambda_l^{i+2}}, q_y^{\lambda_l^{i}})$), this finding being also consistent with Fig.s 1*i*-*m*. It is then noted that the simultaneous knowledge on both $q_y^{\lambda_l^{i+2}}$ and $q_y^{\lambda_l^{i+1}}$ appears to have only a marginal benefit, following the observation that both the synergetic and the unique contribution associated with $q_y^{\lambda_l^{i+2}}$ are practically zero.

These sets of results are conducive to answer questions of the kind 'How much information is lost (or retained) when one would substitute $q_y^{\lambda_l^{i}}$ associated with the smallest support scale in a triplet with its counterparts associated with larger supports, i.e., $q_y^{\lambda_l^{i+1}}$ and $q_y^{\lambda_l^{i+2}}$?'. Otherwise, it is also of interest to focus on questions such as 'How does knowledge on $q_y^{\lambda_l^{i-2}}$ and $q_y^{\lambda_l^{i-1}}$, i.e., the longitudinal flow components linked to the smallest and intermediate supports in a triplet, provide information to $q_y^{\lambda_l^{i}}$, i.e., their counterpart associated with the largest support?'. Inspection of the results in Fig. 5*b* reveals that, when considering triplets of the kind $(q_y^{\lambda_l^{i-2}}, q_y^{\lambda_l^{i-1}}; q_y^{\lambda_l^{i}})$: (*i*) the largest contribution is the redundant information provided by $q_y^{\lambda_l^{i-2}}$ and $q_y^{\lambda_l^{i-1}}$ to $q_y^{\lambda_l^{i}}$; (*ii*) the unique contribution of $q_y^{\lambda_l^{i-1}}$ is non-negligible; while null values are observed for (*iii*) the unique contribution of $q_y^{\lambda_l^{i-2}}$ and (*iv*) the synergetic component associated with simultaneous knowledge of $q_y^{\lambda_l^{i-1}}$ and $q_y^{\lambda_l^{i-2}}$. This set of results indicates that it is convenient to focus on the variable associated with the intermediate lower cutoff scale (in a triplet), i.e., $q_y^{\lambda_l^{i-1}}$, to maximize the information content with respect to the variable related to the larger lower cutoff scale, i.e., $q_y^{\lambda_l^{i}}$. At the same time, the simultaneous knowledge on both $q_y^{\lambda_l^{i-1}}$ and $q_y^{\lambda_l^{i-2}}$ does not bring any additional benefit, given that $U^*(q_y^{\lambda_l^{i-2}}; q_y^{\lambda_l^{i}})$ and $S^*(q_y^{\lambda_l^{i-1}}, q_y^{\lambda_l^{i-2}}; q_y^{\lambda_l^{i}})$ vanish.

The analysis of information partitioning associated with triplets of the kind $(q_y^{\lambda_l^{i+1}}, q_y^{\lambda_l^{i+2}}; q_y^{\lambda_l^{i}})$ and $(q_y^{\lambda_l^{i-2}}, q_y^{\lambda_l^{i-1}}; q_y^{\lambda_l^{i}})$ has the potential to be employed to guide the selection of the optimal support/observation scale (according to some selected criteria involving the analyzed IT metrics, see e.g., Abellan and Noetinger, 2010) within the context of the design of sampling strategies in the presence of data associated with multiple scales.

The analysis for the transverse flux component, i.e., $q_x^{\lambda_l^i}$, which is patterned after the one performed for $q_y^{\lambda_{l,i}}$, provides very similar results to what illustrated above (see SM1).

4. Conclusions

We treat the logarithm of hydraulic conductivity as a random field characterized by a truncated power law variogram which involves the definition of a lower cutoff scale. The latter is linked to the size of the support/measurement scale of hydraulic conductivity. We then assess the impact that variations of the lower cutoff scale, i.e. of the log-conductivity support scale, have on (*i*) the information content of the Darcy flux components and (*ii*) the information shared by pairs and triplets of observations of Darcy flux components associated with differing supports. We investigate various degrees of system heterogeneity in a two-dimensional set-up under uniform in the mean flow. Our study leads to the following major conclusions:

1. An increase in the lower cutoff scale leads to a reduction of the Shannon entropy of the Darcy flux components. This is in line with the observed homogenization in the spatial distribution of Darcy flux as the support scale increases.
2. The bivariate mutual information shared by the Darcy flux components associated with the smallest support, λ_l^1 and larger supports, λ_l^i (i = 2, 3, 4), decreases in a regular fashion as λ_l^i increases regardless of the degree of system heterogeneity, once results are normalized by the Shannon entropy of the fields associated with λ_l^1. This result provides a quantification of how representative, with respect to the fields characterized by λ_l^1, are the counterparts associated with increased supports.
3. Trends which are similar to what observed above are documented for the normalized mutual information evaluated between pairs of log-conductivity values characterized by differing support scales. In this case, values of mutual information are smaller than their counterparts related to Darcy fluxes. This result suggests that propagation of information about conductivity onto Darcy flux through the flow and mass balance equation tends to enhance the possibility that flux observations at larger scales provide more information about (unknown) fluxes at smaller scales, as compared to what can be observed with reference to hydraulic conductivities.
4. Considering the Darcy flux components associated with (*i*) small (here denoted as target variable), (*ii*) intermediate and (*iii*) large support scales, evaluation of the information partitioning of the multivariate mutual information shared by such triplets of variables reveals that results associated with the intermediate and large scales provide mostly redundant information about the target variable while only the results at the intermediate scale provide a unique contribution. Correspondingly similar results are observed when considering triplets formed by the Darcy flux components associated with (*i*) large (denoted as target variable), (*ii*) small and (*iii*) intermediate support scales. Such a pattern is observed for all reference levels of system heterogeneity investigated.

The IT-based analysis detailed here can be readily employed in a variety of settings (e.g., solute transport scenarios where system properties are parameterized on a series of block-scale grids with diverse levels of refinement) where multi-scale (and/or multi-model) representations of a system are taken into consideration in order to assess: (*i*) the level of representativeness (and uncertainty) as well as the efficiency (and inefficiency) related to considering pairs of variables; and (*ii*) the way information is shared among triplets of variables of interest. The ensuing results could guide the selection of the most relevant system representation(s) or serve as a tool to quantify the uncertainty associated with a given representation, as compared to a reference one. In this view, the assessment of the information content (and corresponding uncertainty) that diverse system representations (e.g., diverse mathematical formulations or level of model parametrization) share with a set of available data (collected at either the pore, laboratory or field scale) will be the subject of a future study. These types of analysis could provide new insights in the context of model benchmarking and diagnostic for subsurface related analysis (see, e.g., Nearing et al., 2016, 2018, 2020; and Ruddell et al., 2019 in the context of surface hydrology).

Acknowledgements

The authors would like to thank the EU and MIUR for funding, in the frame of the collaborative international Consortium (WE-NEED) financed under the ERA-NET WaterWorks2014 Cofunded Call. This ERA-NET is an integral part of the 2015 Joint Activities developed by the Water Challenges for a Changing World Joint Programme Initiative (Water JPI). Prof. A. Guadagnini acknowledges funding from Région Grand-Est and Strasbourg-Eurométropole through the ‘Chair Gutenberg’.

References

Abidoye, L., Das, D. B., 2014. Scale dependent dynamic capillary pressure effect for two-phase flow in porous media. Adv. Water Resour. 74, 212-230. http://dx.doi.org/10.1016/j.advwatres.2014.09.009.

Abellan, A., Noetinger, B., 2010. Optimizing subsurface field data acquisition using information theory. Math. Geosc., 42(6), 603-630.

Alfonso, L., Lobbrecht, A., Price, R., 2010. Optimization of water level monitoring network in polder systems using information theory. Water Resour. Res., 46, W12553. https://doi.org/10.1029/2009WR008953.

Andersson, J. E., Ekman, L., Gustafsson, E., Nordqvist, R., Tiren, S., 1988. Hydraulic interference tests and tracer tests within the Bränd□an area, Finnsjon study site, the fracture zone project-Phase 3. Technical Report 89-12, Sweden Nuclear Fuel and Waste Management Company, Stockholm.

Attinger, S., 2003. Generalized coarse graining procedures for flow in porous media. Comput. Geosci., 7, 253-257.

Benedetti, R., 2010. Scoring rules for forecast verification. Month. Weath. Rev. 138(1), 203-211. https://doi.org/10.1175/2009MWR2945.1.

Bennett, A., Nijssen, B., Ou, G., Clark, M., Nearing, G. S., 2019. Quantifying process connectivity with transfer entropy in hydrologic models. Water Resour. Res. 55(6), 4613-4629. https://doi.org/10.1029/2018WR024555.

Bertschinger, N., Rauh, J., Olbrich, E., Jost, J., Ay, N., 2014. Quantifying unique information. Entropy 16(4), 2161-2183. doi:10.3390/e16042161.

Berkowitz, B., Cortis, A., Dentz, M., Scher, H., 2006. Modeling non-Fickian transport in geological formations as a continuous time random walk. Rev. Geophys. 44, RG2003. doi:10.1029/2005RG000178.

Bianchi, M., Pedretti, D., 2017. Geological entropy and solute transport in heterogeneous porous media. Water Resour. Res. 53, 4691-4708. doi:10.1002/2016WR020195.

Bianchi, M., Pedretti, D., 2018. An entrogram-based approach to describe spatial heterogeneity with applications to solute transport in porous media. Water Resour. Res. 54, 4432-4448. https://doi.org/10.1029/ 2018WR022827.

Boschan, A., Noetinger, B., 2012. Scale dependence of effective hydraulic conductivity distributions in 3D heterogeneous media: a numerical study. Transp. Porous Med., 94(1), 101-121.

Boso, F., Tartakovsky, D. M., 2018. Information-theoretic approach to bidirectional scaling. Water Resour. Res. 54, 4916-4928. https://doi.org/10.1029/2017WR021993.

Butera, I., Vallivero, L., Rodolfi, L., 2018. Mutual information analysis to approach nonlinearity in groundwater stochastic fields. Stoch. Environ. Res. Risk Assess. 32 (10), 2933-2942. https://doi.org/10.1007/s00477-018-1591-4, 2018.

Chen, J.-S., Liu, C.-W., Liang, C.-P., 2006. Evaluation of longitudinal and transverse dispersivities/distance ratios for tracer test in a radially convergent flow field with scale-dependent dispersion. Adv. Water Resour. 29, 887-898. doi:10.1016/j.advwatres.2005.08.001.

Chen, Y., Durlofsky, L. J., Gerritsen, M., Wen, X. H., 2003. A coupled local-global upscaling approach for simulating flow in highly heterogeneous formations. Adv. Water Resour. 26, 1041-1060. https://doi.org/10.1016/S0309-1708(03)00101-5

Comolli, A., Hakoun, V., Dentz, M., 2019. Mechanisms, upscaling and prediction of anomalous dispersion in heterogeneous porous media. Water Resour. Res. 55(10), 8197-8222, https://doi.org/10.1029/2019WR024919.

Colecchio, I., Boschan, A., Otero, A. D., Noetinger, B., 2020. On the multiscale characterization of effective hydraulic conductivity in random heterogeneous media: a historical survey and some new perspectives. Adv. Water Resour., 140, 103594. https://doi.org/10.1016/j.advwatres.2020.103594.

Dagan, G., 1984, Solute transport in heterogeneous porous formations. J. Fluid Mech. 145, 151–177.

de Barros, F. P. J., Rubin, Y., 2011. Modelling of block-scale macrodispersion as a random function. J. Fluid Mech. 676, 514-545. doi:10.1017/jfm.2011.65.

de Barros, F. P. J., Dentz, M., 2016. Pictures of blockscale transport: Effective versus ensemble dispersion and its uncertainty. Adv. Water Resour. 91, 11-22. https://doi.org/10.1016/j.advwatres.2016.03.004.

de Barros, F. P. J., 2018. Evaluating the combined effects of source zone mass release rates and aquifer heterogeneity on solute discharge uncertainty. Adv. Water Resour. 177, 140-150. https://doi.org/10.1016/j.advwatres.2018.05.010.

Dentz, M., de Barros, F. P. J., 2015. Mixing-scale dependent dispersion for transport in heterogeneous flows. J. Fluid Mech. 777, 178-195. https://doi.org/10.1017/jfm.2015.351.

Di Federico, V., Neuman, S. P., 1997. Scaling of random fields by means of truncated power variograms and associated spectra. Water Resour. Res. 33(5), 1075-1085.

Di Palma, P. R., Guyennon, N., Heße, F., Romano, E., 2017. Porous media flux sensitivity to pore-scale geostatistics: A bottom-up approach. Adv. Water Resour. 102, 99-110. http://dx.doi.org/10.1016/j.advwatres.2017.02.002.

Dykaar, B. B., Kitanidis, P. K., 1992. Determination of the effective hydraulic conductivity for heterogeneous porous media using a numerical spectral approach: 2. Results. Water Resour. Res., 28(4), 1167-1178. https://doi.org/10.1029/91WR03084.

Eberhard, J., Attinger, S., Wittum, G., 2004. Coarse graining for upscaling of flow in heterogeneous porous media. Multiscale Model. Simul., 2, 269 Fenton, G.A., Griffiths, D.V.: Statistics. https://doi.org/10.1137/030600497.

Fahle, M., Hohenbrink, T. L., Dietrich, O., Lischeid, G., 2015. Temporal variability of the optimal monitoring setup assessed using information theory. Water Resour. Res., 51, 7723–7743. https://doi.org/10.1002/2015WR017137.

Gong, W., Gupta, H. V., Yang, D., Sricharam, K., Hero II, A. O., 2013. Estimating epistemic and aleatory uncertainties during hydrologic modeling: An information theoretic approach. Water Res. Resear., 49, 2253-2273. doi:10.1002/wrcr.20161

Goodwell, A. E., Kumar, P., 2017. Temporal information partitioning: Characterizing synergy, uniqueness, and redundancy in interacting environmental variables. Water Resour. Res., 53, 5920-5942. https://doi.org/10.1002/2016WR020216.

Goodwell, A. E., Jiang, P., Ruddell, B. L., Kumar, P., 2020. Debates - Does information theory provide a new paradigm for Earth science? Causality, interaction, and feedback. Water Resour. Res., 56, e2019WR024940. https://doi.org/10.1029/2019WR024940.

Gotovac, H., Cvetkovic, V., Andrievic, R., 2010. Significance of higher moments for complete characterization of the travel time probability density function in heterogeneous porous media using the maximum entropy principle. Water Resour. Res. 46, W05502. https://doi.org/10.1029/2009WR008220.

Griffith, V., Koch, C., 2014. Quantifying synergistic mutual information. Guided Self-Organization: Inception, edited by M. Prokopenko, 159-190. Springer-Verlag Berlin Heidelberger, Berlin, Germany.

Griffith, V., Ho, T., 2015. Quantifying redundant information in predicting a target random variable. Entropy 17(7), 4644-4653. doi:10.3390/e17074644.

Hecht, F., 2012. New development in FreeFem++. J. Numer. Math. 20 (3-4), 251-265.

Harder, M., Salge, C., Polani, D., 2013. Bivariate measure of redundant information. Phys. Rev. E 87(1), 012130. doi:10.1103/PhysRevE.87.012130.

Hyman, J. D., Dentz, M., Hagberg, A., Kang, P. K., 2019. Linking structural and transport properties in three dimensional fracture networks. J. Geophys. Res.: Solid Earth 124,1185-1204. https://doi.org/10.1029/2018JB016553.

Icardi, M., Boccardo, G., Dentz, M., 2019. Upscaling flow and transport processes. In: Toschi, F., Sega, M. (eds) Flowing Matter. Soft and Biological Matter. Springer, Cham. https://doi.org/10.1007/978-3-030-23370-9_5.

Kitanidis, P. K., 1994. The concept of the dilution index. Water Resour. Res. 30(7), 2011-2016. https://doi.org/10.1029/94WR00762.

Kumar, P., Gupta, H. V., 2020. Debates - Does Information Theory provide a new paradigm for Earth science?. Water Resour. Res. 56, e2019WR026398. https://doi.org/10.1029/2019WR026398.

Mishra, S., Deeds, N., Ruskauff, G., 2009. Global sensitivity analysis techniques for probabilistic ground water modeling. Ground Water 47(5), 730-747. doi:10.1111/j.1745-6584.2009.00604.x.

Meyer, D. W., Bijeljic, B., 2016. Pore-scale dispersion: Bridging the gap between microscopic pore structure and the emerging macroscopic transport behavior. Phys. Rev. E 94, 013107. doi:10.1103/PhysRevE.94.013107.

Meile, C., Tunkay, K., 2006. Scale dependence of reaction rates in porous media. Adv. Water Resour. 29(1), 62-71. https://doi.org/10.1016/j.advwatres.2005.05.007.

Moslehi, M., de Barros, F. P. J., 2017. Uncertainty quantification of environmental performance metrics in heterogeneous aquifers with long-range correlations. J. Contam. Hydrol. 196, 21-29. https://doi.org/10.1016/j.jconhyd.2016.12.002.

Moslehi, M., de Barros, F. P. J., Ebrahimi, F., Sahimi, M., 2016. Upscaling of solute transport in disordered porous media by wavelet transformations. Adv. Water Resour. 96, 180-189. https://doi.org/10.1016/j.advwatres.2016.07.013.

Nearing, G. S., Gupta, H. V., Crow, W. T., 2013a. Information loss in approximately Bayesian estimation techniques: A comparison of generative and discriminative approaches to estimating agricultural productivity. J. Hydrol. 507, 163-173. https://doi.org/10.1016/j.jhydrol.2013.10.029.

Nearing, G. S., Gupta, H. V., Crow, W. T., Gong, W., 2013b. An approach to quantifying the efficiency of a Bayesian filter. Water Resour. Res. 49, 2164-2173. https://doi.org/10.1002/wrcr.20177.

Nearing, G. S., Mocko, D.S., Peters-Lidard, C. D., Kumar, S. V., Xia, Y., 2016. Benchmarking NLDAS-2 soil moisture and evapotranspiration to separate uncertainty contributions. J. Hydrometeor. 17, 745-759. doi:10.1175/JHM-D-15-0063.1.

Nearing, G. S., Ruddell, B. J., Clark, P. M., Nijssen, B., Peters-Lidard, C. D., 2018. Benchmarking and process diagnostic of land models, J. Hydrometeor. 19, 1835-1852. https://doi.org/10.1175/JHM-D-498 17-0209.1.

Nearing, G. S., Ruddell, B. L., Bennett, A. R., Prieto, C., Gupta, H. V., 2020. Does information theory provide a new paradigm for earth science? Hypothesis testing. Water Resour. Res. 56, e2019WR024918. https://doi.org/10.1029/2019WR024918.

Neuman, S. P., 1994. Generalized scaling of permeabilities: Validation and effect of support scale. Geophys. Res. Lett. 21(5), 349-352. https://doi.org/10.1029/94GL00308.

Neuman, S. P., 1995. On advective transport in fractal permeability and velocity fields. Water Resour. Res. 31(6), 1455-1460. https://doi.org/10.1029/95WR00426.

Neuman, S. P., Riva, M., Guadagnini, A., 2008. On the geostatistical characterization of hierarchical media. Water Resour. Res. 44, W02403. doi:10.1029/2007WR006228.

Neuman, S. P., Di Federico, V., 2003. Multifaceted nature of hydrogeologic scaling and its interpretation. Rev. Geophys. 41, 1014. doi:10.1029/2003RG000130.

Noetinger, B., 2000. Computing the effective permeability of log-normal permeability fields using renormalization methods. C.R. Acad. Sci. Sci Terre Des Planètes 331, 353-357. https://doi.org/10.1016/S1251-8050(00)01412-9.

Noetinger, B., Artus, V., Zargar, G., 2005. The future of stochastic and upscaling methods in hydrogeology. Hydrogeol. J. 13, 184-201. https://doi.org/10.1007/s10040-004-0427-0.

Nowak, W., Guthke, A., 2016. Entropy-based experimental design for optimal model discrimination in the geosciences. Entropy 18, 409. doi:10.3390/e18110409.

Olbrich, E., Bertschinger, N., Rauh, J., 2015. Information decomposition and synergy. Entropy 11, 3501-3517. doi:10.3390/e17053501.

Perdigão, R. A. P., Ehret, U., Knuth, K. H., Wang, J., 2020. Debates: Does information theory provide a new paradigm for Earth science? Emerging concepts and pathways of information physics. Water Resour. Res. 56, e2019WR025270. https://doi.org/10.1029/2019WR025270.

Puyguiraud, A., Gouze, P., Dentz, M., 2020. Is there a representative elementary volume for anomalous dispersion?. Transp. Porous Media 131, 767-778. https://doi.org/10.1007/s11242-019-01366-z.

Porta, G. M., Chaynikov, S., Thovert, J.-F., Riva, M., Guadagnini, A., Adler, P. M., 2013. Numerical investigation of pore and continuum scale formulations of bimolecular reactive transport in porous media. Adv. Water Resour. 62, 243-253. https://doi.org/10.1016/j.advwatres.2013.09.007.

Porta, G., Bijeljic, B., Blunt, M. J., Guadagnini, A., 2015. Continuum-scale characterization of solute transport based on pore-scale velocity distributions. Geophys. Res. Letters, 42(18), 7537-7545. doi:10.1002/2015GL065423.

Renard, P., de Marsily, G., 1997. Calculating equivalent permeability: a review. Adv. Water Resour., 20(5–6), 253-278.

Ruddell, B. J., Drewry, D. T., Nearing, G. S., 2019. Information theory for model diagnostics: structural error is indicated by trade-off between functional and predictive performance. Water Resour. Res. 55(8), 6534-6554. https://doi.org/10.1029/2018WR023692.

Sanchez-Vila, X., Guadagnini, A., Carrera, J., 2006. Representative hydraulic conductivities in saturated groundwater flow. Rev. Geophys. 44, RG3002. https://doi.org/10.1029/2005RG000169.

Schad, H., Teutsch, G., 1994. Effects of the investigation scale on pumping test results in heterogeneous porous aquifers. J. Hydrol. 159 (1-4), 61-77. https://doi.org/10.1016/0022-1694(94)90249-6.

Shannon, C., 1948. A mathematical theory of communication. Bell. Syst. Tech. J. 27(3), 531. doi:10.1002/j.1538-7305.1948.tb01338.x.

Siena, M., Guadagnini, A., Riva, M., Bijeljic, B., Nunes Pereira, J. P., Blunt, M. J., 2014. Statistical scaling of pore-scale Lagrangian velocities in natural porous media. Phys. Rev. E 90, 023013. doi:10.1103/PhysRevE.90.023013.

Siena, M., Iliev, O., Prill, T., Riva, M., Guadagnini, A., 2019. Identification of channeling in pore-scale flows. Geophys. Res. Lett. 46(6), 3270-3278. https://doi.org/10.1029/2018GL081697.

Stone, J. V., 2015. Information Theory: A Tutorial Introduction. Sebtel Press.

Tartakovsky, A. M., Meakin, P., Huang, H., 2004. Stochastic analysis of immiscible displacement of the fluidswith arbitrary viscosities and its dependence on support scale of hydrological data. Adv. Water Resour., 27, 1151-1166. doi:10.1016/j.advwatres.2004.09.003.

Tartakovsky, A. M., Panzeri, M., Tartakovsky, G. D., Guadagnini, A., 2017. Uncertainty quantification in scale-dependent models of flow in porous media. Water Resour. Res. 53, 9392-9401. doi:https://doi.org/10.1002/2017WR020905.

Tidwell, V. C., Wilson, J. L., 1999a. Permeability upscaling measured on a block of Berea Sandstone: Results and interpretation. Math. Geol. 31(7), 749-769. https://doi.org/10.1023/A:1007568632217.

Tidwell, V. C., Wilson, J. L., 1999b. Upscaling experiments conducted on a block of volcanic tuff: Results for a bimodal permeability distribution. Water Resour. Res. 35(11), 3375-3387. https://doi.org/10.1029/1999WR900161.

Wang, Y., Aryana, S. A., Allen, M. B., 2019. An extension of Darcy's law incorporating dynamic length scales. Adv. Water Resour. 129, 70-79. https://doi.org/10.1016/j.advwatres.2019.05.010

Wellman, F. J., Regenaur-Lieb, K., 2012. Uncertainties have a meaning: Information entropy as a quality measure for 3-D geological models. Tectonophys. 526-529, 207-216. doi:10.1016/j.tecto.2011.05.001.

Wellman, F. J., 2013. Information theory for correlation analysis and estimation of uncertainty reduction in maps and models. Entropy 15, 1464-1485. doi:10.3390/e15041464

Weijs, S. V., Ruddell, B. L., 2020. Debates: Does information theory provide a new paradigm for Earth science? Sharper predictions using Occam's digital razor. Water Resour. Res. 56(2), e2019WR026471. https://doi.org/10.1029/2019WR026471

Weijs, S. V., Schoups, G., van de Giesen, N., 2010. Why hydrological predictions should be evaluated using information theory. Hydrol. Earth Syst. Scienc. 14(12), 2545–2558. https://doi.org/10.5194/hess-14-2545-2010.

Wen, X. H., Gomez Hernandez, J. J., 1996. Upscaling hydraulic conductivities in heterogeneous media: an overview. J. Hydrol. 183(1-2), 9-32. https://doi.org/10.1016/S0022-1694(96)80030-8.

Williams, P. L., Beer, R. D., 2010. Nonnegative decomposition of multivariate information. CoRR. http://arxiv.org/abs/1004.2515.

Woodbury, A. D., Ulrych, T. J., 1993. Minimum relative entropy: forward probabilistic modeling. Water Resour. Res. 29(8), 2847-2860. https://doi.org/10.1029/93WR00923.

Woodbury, A. D., Ulrych, T. J., 1996. Minimum relative entropy inversion: theory and application to recovering the release history of a groundwater contaminant. Water Resour. Res. 32(9), 2671-2681. https://doi.org/10.1029/95WR03818.

Woodbury, A. D., Ulrych, T. J., 2000. A full-Bayesian approach to the groundwater inverse problem for steady state flow. Water Resour. Res. 36(8), 2081-2093. https://doi.org/10.1029/2000WR900086.

Wright, E. E., Sund, N. L., Richter, D. H., Porta, G. M., Bolster, D., 2018. Upscaling mixing in highly heterogeneous porous media via spatial markov model. Water, 11, 53. doi:10.3390/w11010053.

Zeng, X. K., Wan, D., Wu, J. C., 2012. Sensitivity analysis of the probability distribution of groundwater level series based on information entropy. Stoch. Environ. Res. Risk. Assess 26, 345-356. https://doi.org/10.1007/s00477-012-0556-2.

Zhang, D., Zhang, R., Chen, S., Soll, W. E., 2000. Pore scale study of flow in porous media: Scale dependency, REV, and statistical REV. Geophys. Res. Lett. 27(8), 1195-1198. doi:10.1029/1999GL011101.

Figures

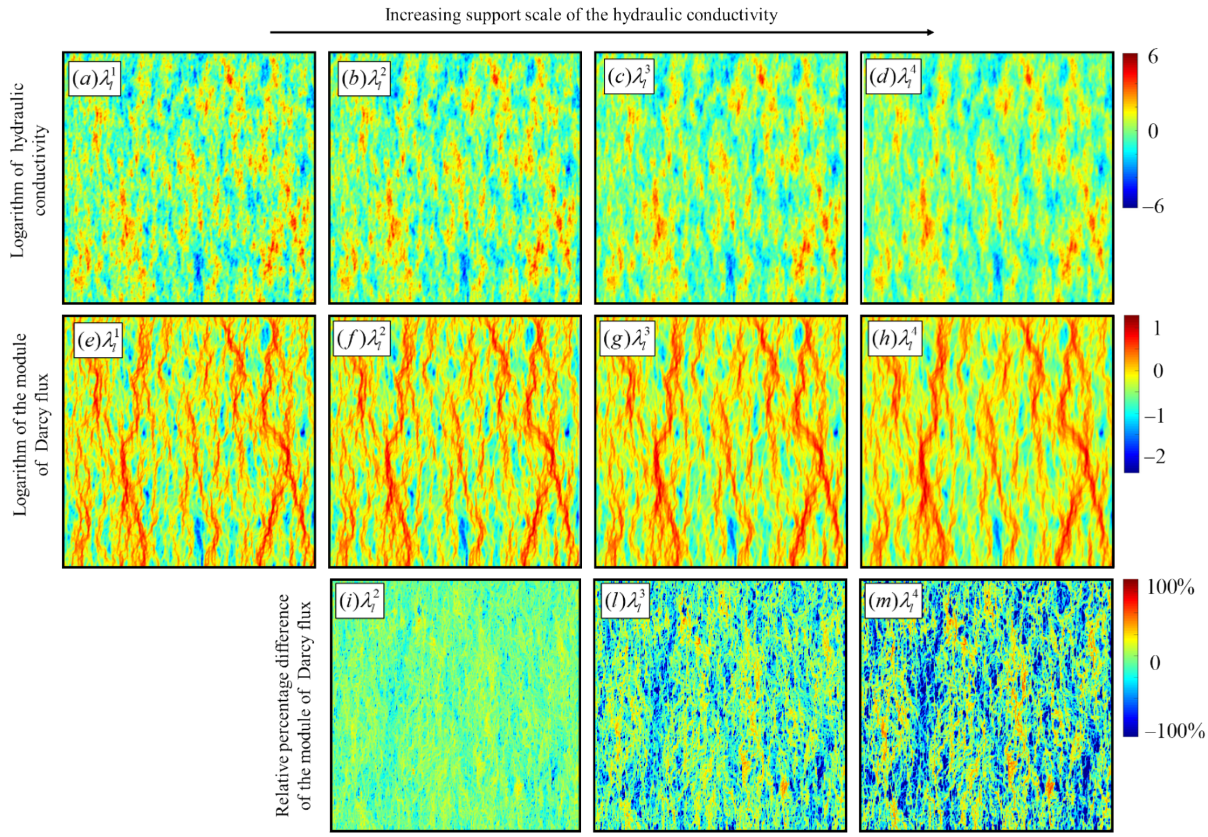

Fig. 1. Spatial distribution of the (natural) logarithm of hydraulic conductivity (first row) and associated distribution of the logarithm of Darcy flux module (second row), for increasing sizes of the lower cutoff scale λ_l (left to right), i.e., (*a*)-(*e*) λ_l^1, (*b*)-(*f*) λ_l^2, (*c*)-(*g*) λ_l^3 and (*d*)-(*h*) λ_l^4 (with $\lambda_l^1 < \lambda_l^2 < \lambda_l^3 < \lambda_l^4$, see also Sec. 3). Spatial distribution of relative percentage differences between the module of the Darcy flux related to the smallest log-conductivity support scale (i.e., λ_l^1) and its counterparts associated with (*i*) λ_l^2, (*l*) λ_l^3 and (*m*) λ_l^4.

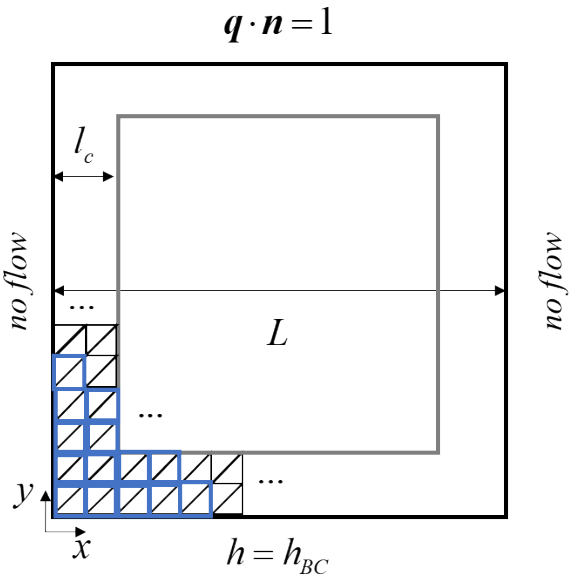

Fig. 2. Sketch of the two-dimensional square domain and boundary conditions associated with the flow problem in (1). The black triangular and blue square mesh correspond to the numerical discretization adopted for the solution of (1) and to the grid across which generation of hydraulic conductivity fields is performed, respectively. The square box (delimited by grey lines) corresponds to the inner spatial region where Darcy fluxes are sampled.

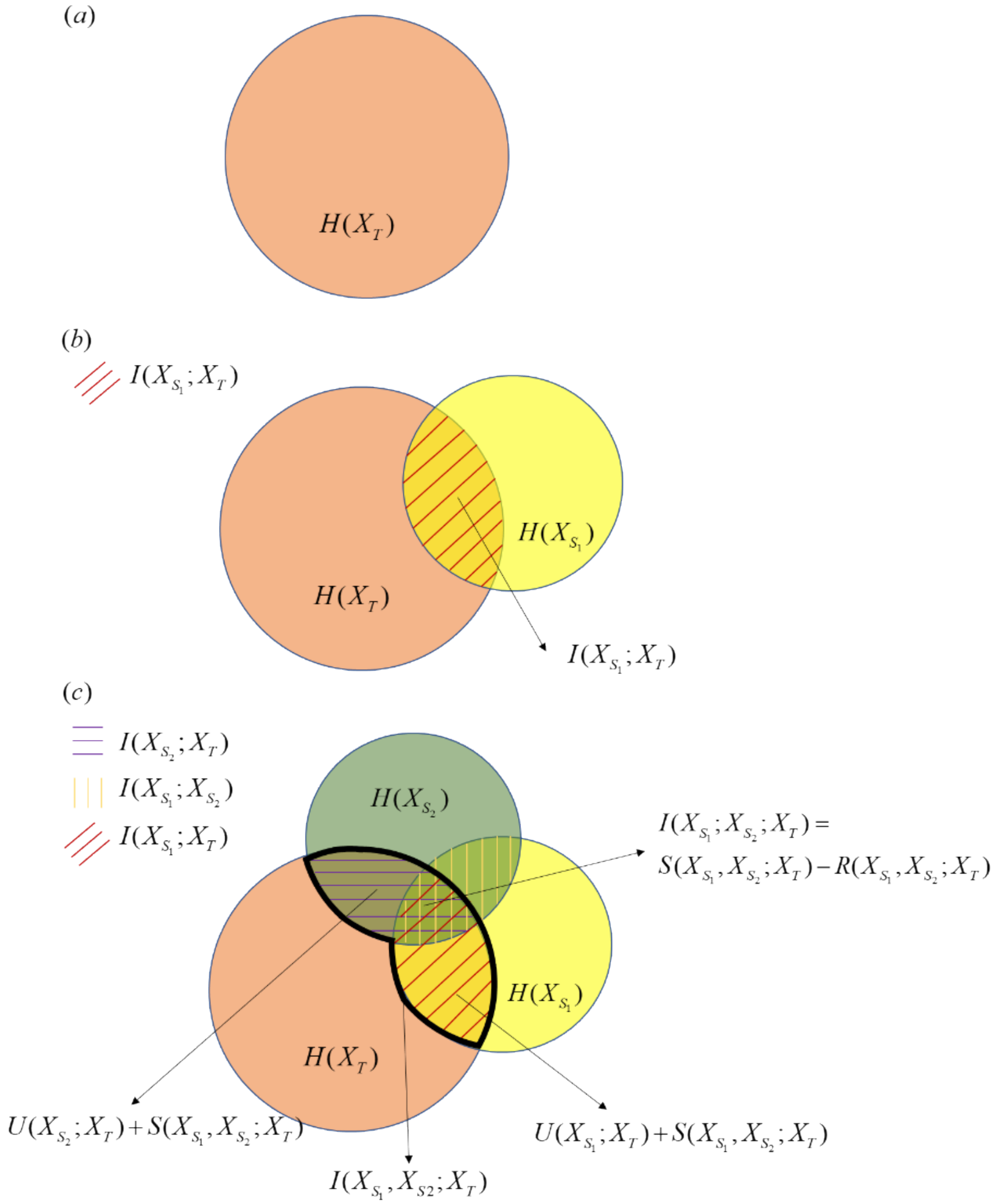

Fig. 3. Venn diagram depictions of the Information theory metrics described in Section 2.3 considering a target, i.e., X_T, and two source, i.e., X_{S_1} and X_{S_2}, variables. The size of each circle is proportional to the corresponding Shannon entropy (e.g., in (*a*) Shannon entropy for the target variable $H(X_T)$). The overlapping region in (*b*) reflects the amount of mutual information shared between pairs of variables, e.g., mutual information shared between X_T and X_{S_1}, i.e., $I(X_{S_1};X_T)$. The multivariate mutual information (i.e., $I(X_{S_1},X_{S_2};X_T)$) shared between the two sources and the target variable is the region demarcated by the black thick curve in (*c*), whereas the unique (i.e., $U(X_{S_1};X_T)$ and $U(X_{S_2};X_T)$), synergetic (i.e., $S(X_{S_1},X_{S_2};X_T)$), redundant (i.e., $R(X_{S_1},X_{S_2};X_T)$) contributions, and interaction information (i.e., $I(X_{S_1};X_{S_2};X_T)$) are highlighted.

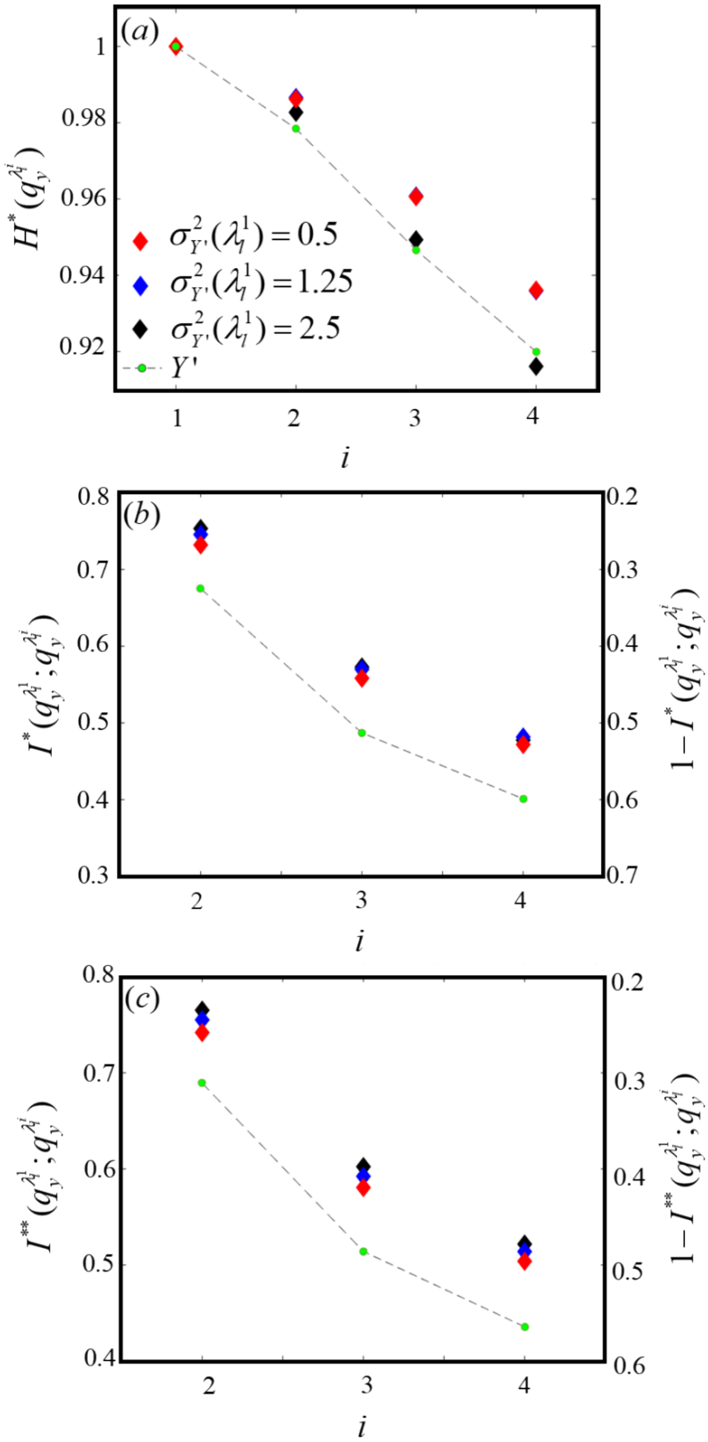

Fig. 4. (*a*) Normalized Shannon entropy, i.e., $H^*(q_y^{\lambda_l^i})$; (*b*) *representativeness*, i.e., $I^*(q_y^{\lambda_l^1};q_y^{\lambda_l^i})$ (left axis), and *uncertainty*, i.e., $1-I^*(q_y^{\lambda_l^1};q_y^{\lambda_l^i})$ (right axis); and (*c*) *efficiency*, i.e., $I^{**}(q_y^{\lambda_l^1};q_y^{\lambda_l^i})$ (left axis), and *inefficiency*, i.e., $1-I^{**}(q_y^{\lambda_l^1};q_y^{\lambda_l^i})$ (right axis), for the longitudinal Darcy flux component considering (*i*) four differing sizes of the lower cutoff scale, i.e., $\lambda_l^1 < \lambda_l^2 < \lambda_l^3 < \lambda_l^4$, and (*ii*) three degree of reference heterogeneity of the system, i.e., $\sigma^2_{Y'}(\lambda_l^1)$ = 0.5 (red symbols), 1.25 (blue symbols), 2.5 (black symbols). Corresponding results associated with the logarithm of hydraulic conductivity, i.e., Y', are also depicted (green symbols and dashed curves).

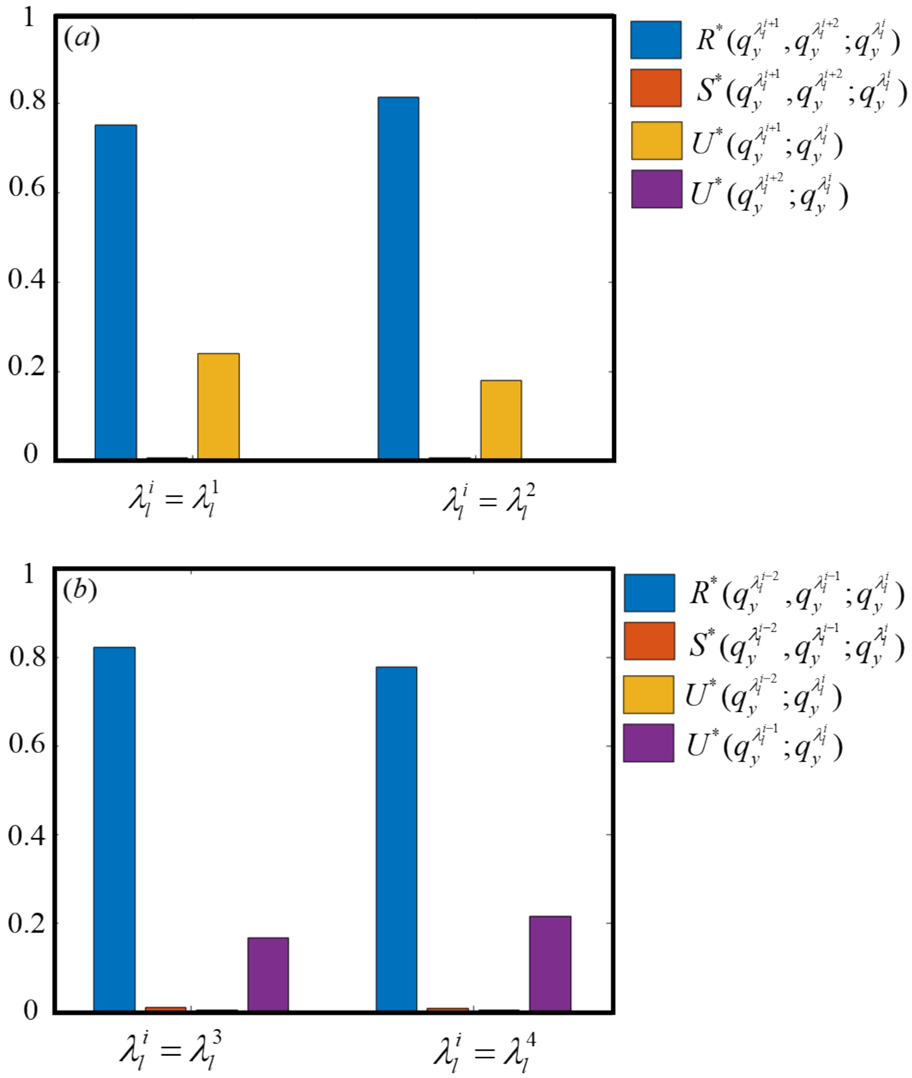

Fig. 5. Information partitioning of the multivariate mutual information considering the triplets (*a*) $(q_y^{\lambda_l^{i+1}}, q_y^{\lambda_l^{i+2}}; q_y^{\lambda_l^i})$ (with $i = 1, 2$) and (*b*) $(q_y^{\lambda_l^{i-2}}, q_y^{\lambda_l^{i-1}}; q_y^{\lambda_l^i})$ (with $i = 3, 4$). For ease of comparison between the results, we normalize the unique, synergetic and redundant contributions by the multivariate mutual information of the corresponding triplet. Results are depicted for the setting related to $\sigma_{Y'}^2(\lambda_l^1) = 2.5$.

Supplementary material: Quantification of the information content of Darcy fluxes associated with hydraulic conductivity fields evaluated at diverse scales

Aronne Dell'Oca, Alberto Guadagnini, and Monica Riva

Department of Civil and Environmental Engineering, Politecnico di Milano, 20133, Milan, Italy;

Corresponding author: Aronne Dell'Oca (aronne.delloca@polimi.it)

List of Figure as Supplementary Material:

- Figure SM1: Truncated power law variogram (TPV) considering $A = 0.41$ and $H = 0.15$, upper cutoff scale $\lambda_u = 32$ and diverse values of the lower cutoff scale $(\lambda_l^1, \lambda_l^2, \lambda_l^3, \lambda_l^4)$.
- Figure SM2: Information partitioning of the multivariate mutual information for the longitudinal Darcy flux component for the settings associated with $\sigma_{Y'}^2(\lambda_l^1)$ = 0.5, and 1.25.
- Figure SM3: Normalized Shannon entropy, *representativeness* (and *uncertainty*) and *efficiency* (and *inefficiency*) for the transverse component of Darcy flux.
- Figure SM4: Information partitioning of the multivariate mutual information for the transverse component of Darcy flux.

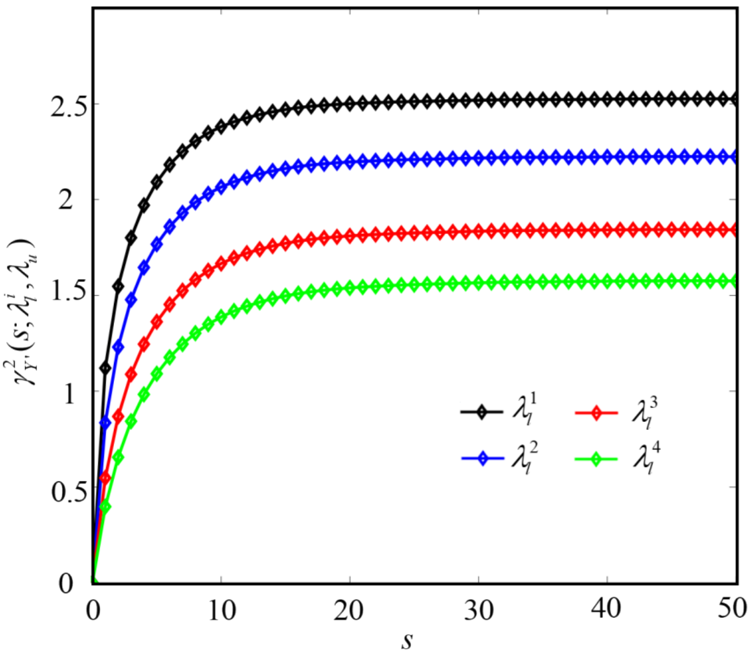

Fig. SM1. Truncated power law variogram, $\gamma^2_{Y'}(s;\lambda_l^i,\lambda_u)$ (see (3)-(4)), versus spatial lag, s, computed for four values of the lower cutoff scale ($\lambda_l^1 = 1$, $\lambda_l^2 = 2$, $\lambda_l^3 = 4$, $\lambda_l^4 = 6$) and setting A = 0.41, H = 0.15, $\lambda_u = 32$.

It is remarked that an increase in the lower cutoff scale of the variogram of Y' yields: (*i*) a decreased strength of the spatial variability of Y', as expressed in terms of the variogram sill (see (5)-(6) and Tab. 1); (*ii*) an increase in the spatial correlation of log-conductivity values, as expressed in terms of the integral scale (see (7) and Tab. 1). These elements contribute to a spatial homogenization of log-conductivity values as the lower cutoff scale increases.

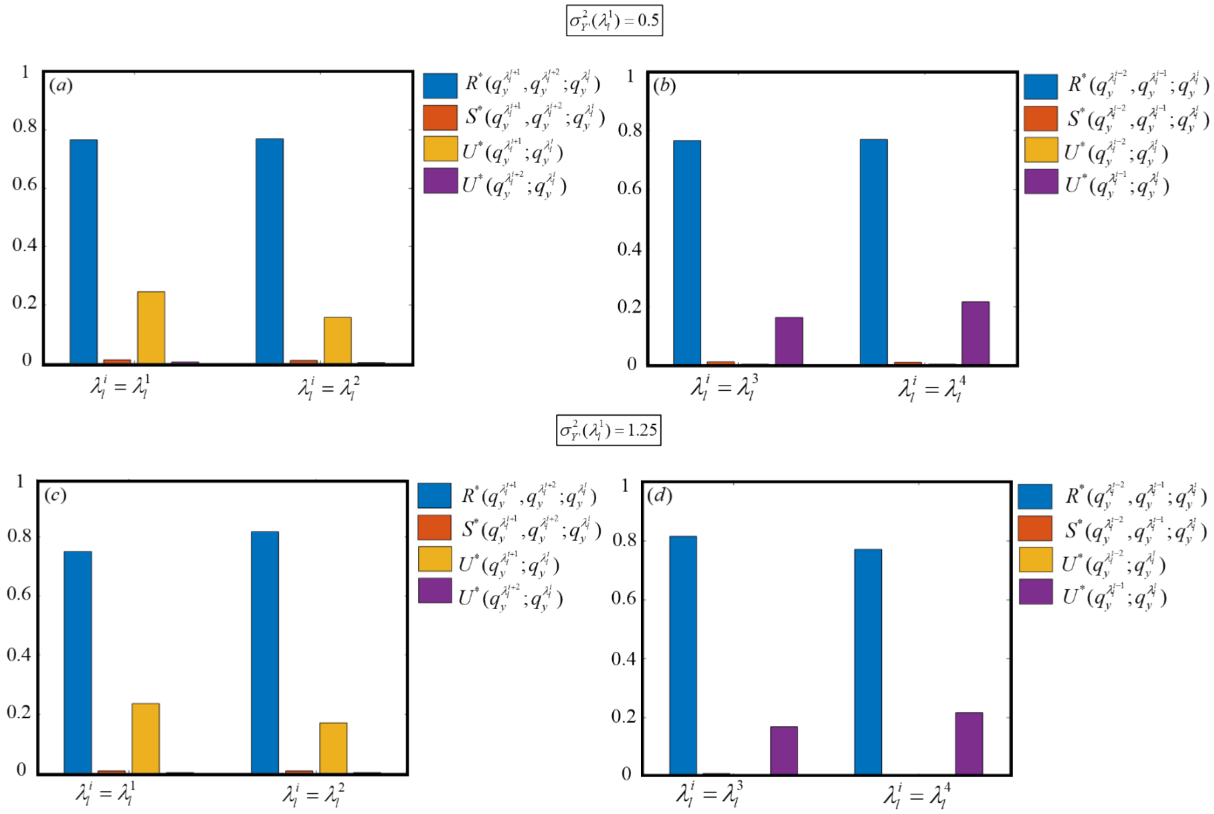

Fig. SM2. Information partitioning of the multivariate mutual information considering the triplets (*a*) $(q_y^{\lambda_l^{i+1}}, q_y^{\lambda_l^{i+2}}; q_y^{\lambda_l^i})$ (with $i = 1, 2$) and (*b*) $(q_y^{\lambda_l^{i-2}}, q_y^{\lambda_l^{i-1}}; q_y^{\lambda_l^i})$ (with $i = 3, 4$). For ease of comparison between the results, we normalize the unique, synergetic and redundant contributions by the multivariate mutual information of the corresponding triplet. Results are depicted for the settings related to (*a*)-(*b*) $\sigma^2_{Y'}(\lambda_l^1) = 0.5$, and (*c*)-(*d*) $\sigma^2_{Y'}(\lambda_l^1) = 1.25$.

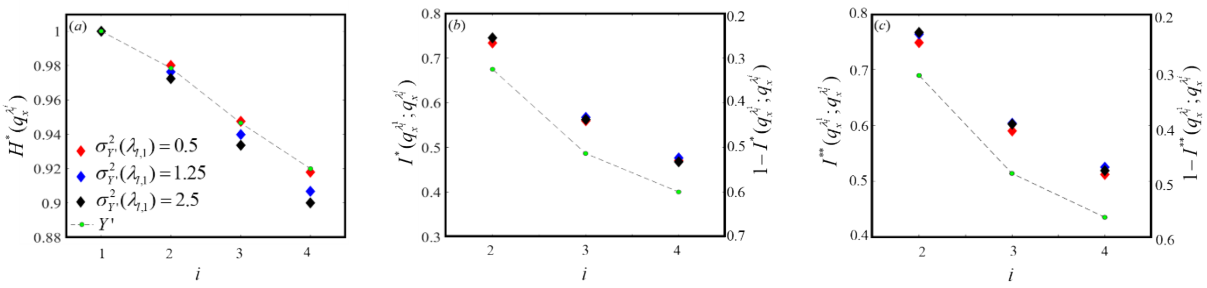

Fig. SM3. (*a*) Normalized Shannon entropy, i.e., $H^*(q_x^{\lambda_l^i})$; (*b*) *representativeness*, i.e., $I^*(q_x^{\lambda_l^1}; q_x^{\lambda_l^i})$ (left axis), and *uncertainty*, i.e., $1 - I^*(q_x^{\lambda_l^1}; q_x^{\lambda_l^i})$ (right axis); and (*c*) *efficiency*, i.e., $I^{**}(q_x^{\lambda_l^1}; q_x^{\lambda_l^i})$ (left axis), and *inefficiency*, i.e., $1 - I^{**}(q_x^{\lambda_l^1}; q_x^{\lambda_l^i})$ (right axis), for the longitudinal Darcy flux component considering (*i*) four differing sizes of the lower cutoff scale, i.e., $\lambda_l^1 < \lambda_l^2 < \lambda_l^3 < \lambda_l^4$,, and (*ii*) three degree of reference heterogeneity of the system, i.e., $\sigma^2_{Y'}(\lambda_l^1)$ = 0.5 (red symbols), 1.25 (blue symbols), 2.5 (black symbols). Corresponding results associated with the logarithm of hydraulic conductivity, i.e., Y', are also depicted (green symbols and dashed curves).

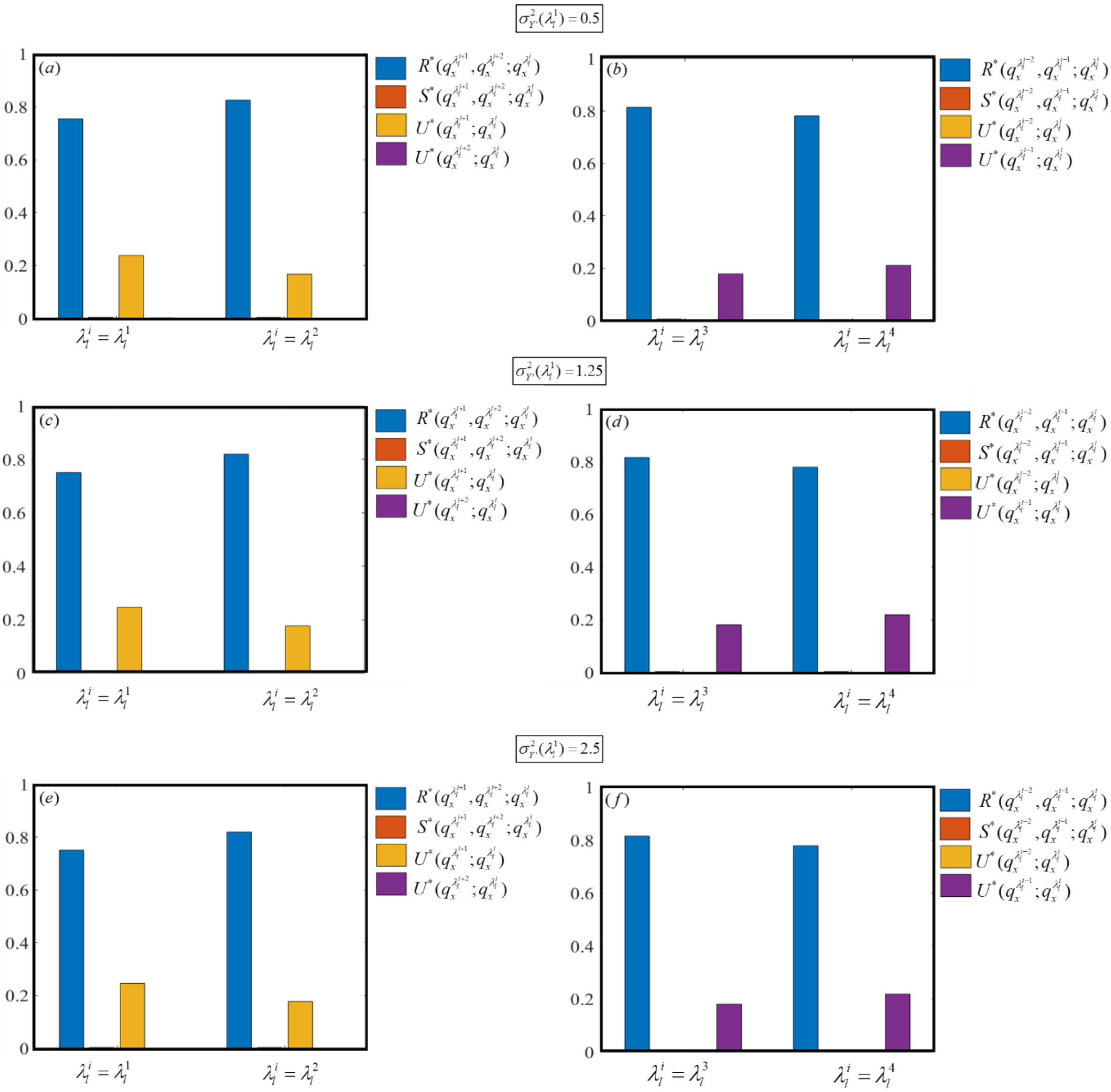

Fig. SM4. Information partitioning of the multivariate mutual information considering the triplets (*a*), (*c*), (*e*) $(q_x^{\lambda_l^{i+1}}, q_x^{\lambda_l^{i+2}}; q_x^{\lambda_l^i})$ (with $i = 1, 2$) and (*b*), (*d*), (*f*) $(q_x^{\lambda_l^{i-2}}, q_x^{\lambda_l^{i-1}}; q_x^{\lambda_l^i})$ (with $i = 3, 4$). For ease of comparison between the results, we normalize the unique, synergetic and redundant contributions by the multivariate mutual information of the corresponding triplet. Results are depicted for the settings related to (*a*)-(*b*) $\sigma_{Y'}^2(\lambda_l^1) = 0.5$, and (*c*)-(*d*) $\sigma_{Y'}^2(\lambda_l^1) = 1.25$, and (*e*)-(*f*) $\sigma_{Y'}^2(\lambda_l^1) = 2.5$.