POLITECNICO DI MILANO

Scuola di Ingegneria Industriale e dell'Informazione

Corso di Laurea Magistrale in
Ingegneria Energetica – Energy Engineering

Information Theory Approach to Groundwater Flow Characterization

Relatore: Prof. Alberto GUADAGNINI

Correlatore: Ing. Aronne DELL'OCA

Tesi di Laurea di:

Matteo RENDINA Matr. 884805

Anno Accademico 2018 - 2019

Politecnico di Milano:
www.polimi.it

Scuola di Ingegneria Industriale e dell'Informazione:
www.ingindinf.polimi.it

Contents

List of Figures

List of Tables

Abstract

In groundwater flow problems the complexity of the models and the importance of having reliable tools for their interpretation made uncertainty quantification an essential part of modelling itself. At the same time Upscaling and Downscaling techniques helped modellers bridging across different resolution scales to tackle scale inconsistency, while often increasing even more the uncertainty. Information Theory (IT) provides powerful tools to quantify information contained in a model, to investigate information loss during Upscaling and to analyze the behaviour of the spatial structure of a field at different resolution scales. While also other studies used these IT metrics in groundwater modelling, this work innovatively coupled them with Monte Carlo simulation method to analyze Upscaling quality and effects in randomly generated hydraulic conductivity fields. Nevertheless, the method presented here could be used with a broad range of Upscaling techniques to asses their effectiveness.

Key words: Groundwater Flow, Uncertainty Quantification, Upscaling, Information Theory, Monte Carlo Method.

Sommario

Nei problemi di flusso delle acque sotterranee la complessità dei modelli e l'importanza di avere strumenti affidabili per la loro interpretazione, ha reso la quantificazione dell'incertezza una parte essenziale della modellazione stessa. Allo stesso tempo, le tecniche di Upscaling e di Downscaling hanno aiutato gli studiosi a muoversi tra diverse scale di risoluzione e ad ovviare a problemi che rendevano i loro modelli inconsistenti, tuttavia incrementando spesso l'incertezza. La Teoria dell'Informazione, o Information Theory (IT), fornisce ottimi strumenti per quantificare l'informazione contenuta in un modello, monitorare l'informazione persa durante il processo di Upscaling ed analizzare l'evoluzione della struttura spaziale di un campo a diverse scale di risoluzione. Già altri studi hanno usato alcune di queste metriche nella modellazione di problemi relativi alle acque sotterranee, ma l'innovazione di questo lavoro risiede nell'unirle con il metodo Monte Carlo per analizzare la qualità e gli effetti dell'Upscaling su dei campi generati casualmente. I metodi presentati rimangono comunque validi per studiare una vasta gamma di problemi legati all'Upscaling.

Parole chiave: Flusso di Acque Sotterranee, Quantificazione dell'Incertezza, Upscaling, Teoria dell'Informazione, Metodo Monte Carlo.

Extended Abstract

0.1 Introduction

Groundwater system is a complex and open system, which is affected by natural conditions and human activities. Natural hydrological processes is conceptualized through relatively simple flow governing equations in groundwater models. Moreover, observation data is always limited in field hydrogeological conditions. Therefore, the predictive results of groundwater simulation often deviate from true values, as a result of the uncertainty of groundwater numerical simulation [**1**]. Addressing uncertainty is an indispensable part of prediction. Groundwater management faces uncertainty on many fronts: in understanding the behaviour of the groundwater system, in anticipating possible future climatic, economic or geopolitical conditions, in prioritising objectives; all of them combining to add ambiguity in the evaluation of management options. Focusing on the first, it is apparent that scientific research has achieved relative success in reducing this uncertainty, culminating in the ability to approximate the behaviour of a groundwater system using a "model". There are, however, limits to the ability of science. Far from being all known, there will always be recognised and unrecognised unknowns this means that a model will always be a simplification of reality, and the predictions it makes will always be uncertain [**2**]. Moreover, often the scale at which transport and flow phenomena in the porous media are best described could be different from the scale at which measurements are available, but also different from the scale required for management decisions [**3**]. In this cases, although it causes the loss of information, it is necessary to use Upscaling or Downscaling techniques to bridge the gap between scales. It goes without saying that, whichever method is used, some information loss in Upscaling is inevitable. One of the most applied modelling tools for analysing a system under conditions of parameter uncertainty is the Monte Carlo simulation. In this technique, we construct a large number of realizations of the considered domain, say with respect to a property like hydraulic conductivity. Each realization

is investigated, yielding a forecast, and the collective behaviour of all forecasts is then analysed, providing the probabilistic information needed for making management decisions under conditions of uncertainty. When it comes to quantify uncertainty, Information Theory (IT) [**4**] provides powerful tools, as entropy and mutual information, to evaluate information of random variables and to study their transmission across different models. We note that information as proposed by Shannon and uncertainty are equivalent concepts, as gaining information about something reduces at the same time the uncertainty related to it. The scope of this study is to use these IT metrics to qualify Upscaling technique as proposed in [**16**]. So far, only few efforts have been made to quantify how information loss occurs when moving to coarser scales [**5**], while IT metrics had already been applied to groundwater problems, we cite [**6**],[**7**] in this sense, while [**8**],[**9**] focused on the study of hydrological time-series data. The paper is organized as follow: *Section 0.2* contains a brief overview of IT concepts used to quantify information content at each scale and information transfer between scales while in *Section 0.3* we explain how we built our model and the Upscaling technique. Results for case $\sigma_Y^2 = 0.5$ are presented in *Sections 0.4,5,6*, which contain, respectively the entropy calculations of different scales, the information partitioning between scales and the evolution of spatial correlation during Upscaling. Finally, we analyze the results in *Section 0.7*, while we refer to the Appendix for the results relative to case $\sigma_Y^2 = 2$.

0.2 Theoretical background

This study aimed at uncertainty quantification for flow problems in porous media. Its governing equations at Darcy's scale, for a chemically inactive, viscous, Newtonian fluid with constant temperature are:

$$\begin{cases} \varphi \mathbf{V} = -\frac{\mathbf{k}}{\mu} \left(\nabla P + \rho \mathbf{g} \right) \\ \nabla \cdot \mathbf{V} = 0 \end{cases} \tag{1}$$

where we could divide $\mathbf{V}$ into its horizontal and vertical components, for our notation, respectively, V_y and V_x. Shannon [**4**] provided mathematical tools for information quantification, which he has defined as "the resolution of uncertainty". He stated that the amount of surprise related to an outcome value of a random variable is a measure of information, as the surprise increases as the probability of the outcome decrease. Given a random variable X, which has a distribution $p\left(X\right)$, the average amount of

surprise (i.e. information) contained into it, is called the entropy of $p(X)$, it represented as $H(X)$ and it has the unit of measure of *bits*. Entropy is computed as follow:

$$H(X) = -\sum_{i=1}^{N} p(X_i) \log_2 p(X_i) \tag{2}$$

Where i is the bin number, N is the number of populated bins of an histogram, and $p(X_i)$ is the proportion of data falling into the i-th populated bin. Given two random variables X, Y that are defined on a probability space, the joint probability distribution $p(X_i, Y_j)$ for X, Y is a probability distribution that gives the probability that each of X, Y falls in any particular range or discrete set of values specified for that variable. The corresponding joint entropy of X, Y is then computed as:

$$H(X,Y) = -\sum_{i=1}^{N}\sum_{j=1}^{M} p(X_i, Y_j) \log_2 p(X_i, Y_j) \tag{3}$$

Where N is the number of different values of X and M is the number of different values of Y, while $p(X_i, Y_i)$ is the joint probability of the X_i and Y_i values. In the case in which these two variables are somehow correlated, whatever is the nature of the relationship, it is possible that they shared an amount of information. This portion of information that the observation of a variable provides about the other variable is called mutual information and it can be computed as:

$$I(X,Y) = \sum_{i=1}^{N}\sum_{j=1}^{M} p(X_i, Y_j) \log_2 \frac{p(X_i, Y_j)}{p(X_i)\, p(Y_j)} \tag{4}$$

When three variables are involved, it is possible to compute the information that two variables, let's say X and Y, provide about the other one, which is Z in this case:

$$I(X,Y;Z) = \sum_{i=1}^{N}\sum_{j=1}^{M}\sum_{k=1}^{K} p(X_i, Y_j, Z_k) \log_2 \frac{p(X_i, Y_j, Z_k)}{p(X_i, Y_j)\, p(Z_k)} \tag{5}$$

Where $p(X_i, Y_j, Z_k)$ is the multivariate joint probability of the X_i, Y_i and Z_i values. Recent research on information partitioning **[8]** has enabled more precise classification of the nature of multivariate shared information. Information partitioning categorizes shared information quantities between multiple source variables and a target variable as either synergistic, unique,

or redundant. Redundant information (R) is the information that multiple sources provide to a target such that they overlap in their information content. Unique information (U) from a source refers to the information it shares with a target that is not redundant with information provided by another source. Synergistic information (S) refers to the information that two sources provide to a target only jointly. We can compute:

$$I\left(X,Y;Z\right)=U_1\left(X;Z\right)+U_2\left(Y;Z\right)+R+S \tag{6}$$

Where U_1, U_2, R and S are non-negative quantities. $U_1\left(X;Z\right)$ and $U_2\left(Y;Z\right)$ are, respectively, the information that sources X and Y share with the target Z; R is interpreted as overlapping shared information; S is a cooperative provision of shared information that is possible to gain only if X and Y are considered jointly. These metrics are useful to quantify the amount of information contained in a variable and the information that two or three variables share together; this powerful tool has been used to analyze entropies and information partitioning of reference and Upscaled fields for variables hydraulic conductivity (or better, as explained in *Section 0.3*, Y) and velocity fields (V_x and V_y). The studied variables (Y, V_x and V_y) may have a spatial correlation within the field, that could be linear or not; again IT metrics are useful to detect and quantify these correlations and study their evolution during the Upscaling process. For a variable X, sampling all its possible couples of values at (i) one location and (ii) a location which is distant of a given lag, three coefficients, ρ, R and U could be computed as **[10]**,**[11]**,**[7]**:

$$\rho\left(X,X\left(lag\right)\right)=\frac{Cov\left(X\left(x\right),X\left(x+lag\right)\right)}{\sigma_{X(x)}\sigma_{X(x+lag)}} \tag{7}$$

$$R\left(X,X\left(lag\right)\right)=\left\{1-exp[-2I\left(X\left(x\right),X\left(x+lag\right)\right)]\right\}^{1/2} \tag{8}$$

$$U\left(X,X\left(lag\right)\right)=2\frac{I\left(X\left(x\right),X\left(x+lag\right)\right)}{H\left(X\left(x\right)\right)H\left(X\left(x+lag\right)\right)} \tag{9}$$

being $Cov\left(X\left(x\right),X\left(x+lag\right)\right)$ the covariance, $\sigma_{X(x)}$ and $\sigma_{X(x+lag)}$ the standard deviations of $X\left(x\right)$ and $X\left(x+lag\right)$, $I\left(X\left(x\right),X\left(x+lag\right)\right)$ the mutual information between two points of a couple and $H\left(X\left(x\right)\right)$ and $H\left(X\left(x+lag\right)\right)$ the entropies of the two points. Coefficient U (or uncertainty coefficient) lies between 0 and 1: when the uncertainty coefficient is zero, it means that $X\left(x\right)$ and $X\left(x+lag\right)$ are not dependent on each other; if its value is unitary, the knowledge of $X\left(x\right)$ is able to completely predict $X\left(x+lag\right)$, and the opposite is also true**[7]**. The Bravis-Pearson index, ρ,

is known as the linear correlation coefficient or Pearson correlation coefficient. It is a measure of the linear dependence of two random variables. if $|\rho| = 1$, a perfect linear relationship exists between $X(x)$ and $X(x + lag)$ and the variables are fully correlated; instead, if $\rho = 0$, the variables are not correlated[**7**]. R takes values in the range [0,1]. R is zero if $X(x)$ and $X(x + lag)$ are independent, and is unity if there is an exact linear or nonlinear relationship between $X(x)$ and $X(x + lag)$ [**23**]. Taking the spatial average, over all the possible pairs of spatial locations in the domain, we could study the nature of the spatial dependence, investigating the effect of Upscaling over the latter for fields Y, V_x and V_y.

0.3 Problem setting

Monte Carlo method requires the generation of random fields, in which the hydraulic conductivity K is modelled as an isotropic randomly generated field with imposed statistical features. The domain is a $2D$ confined aquifer of side $L = 600$ $[m]$ with constant thickness on which it is imposed a uniform grid with $n_x = n_y = 600$ squared elements. The hydraulic conductivity is modelled as:

$$K = K_g e^{Y(x,y)} \tag{10}$$

where K_g is a typical value of limestone hydraulic conductivity and $Y(x, y)$ is a zero-mean second-order stationary random process characterized by a truncated power law variogram (TPV) with correlation lengths $l_x = l_y = 8$ $[m]$ and an isotropic covariance function:

$$C(h) = \gamma_G^2(h, \lambda_u) - \gamma_G^2(h, \lambda_l) \tag{11}$$

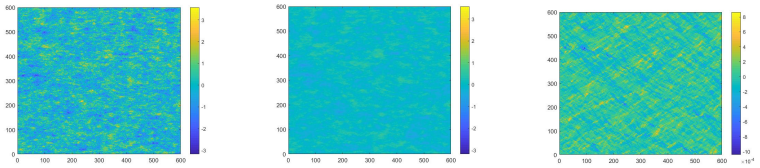
where, for $m = l, u$:

$$\gamma_G^2(h, \lambda_m) = \sigma_Y^2(\lambda_m)\, \rho(h/\lambda_m) \tag{12}$$

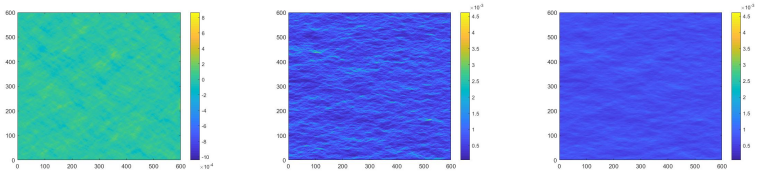
$$\sigma_Y^2(\lambda_m) = A\frac{\lambda_m^{2H}}{2H} \tag{13}$$

$$\rho(h/\lambda_m) = e^{-\frac{h}{\lambda_m}} - \left(\frac{h}{\lambda_m}\right)^{2H} \Gamma(1 - 2H, h/\lambda_m) \tag{14}$$

being h the distance (lag), H the Hurst coefficient (0.333 in our model), Γ the gamma function, A the variance, λ_u the characteristic scale associated with the upper frequency cut-off and λ_l the characteristic scale associated with the lower frequency cut-off. The model studied here is based on



(a) Y at η_0 **(b)** Y at η_{16} **(c)** V_x at η_0



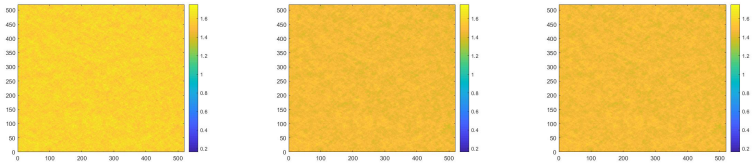
(d) V_x at η_{16} **(e)** V_y at η_0 **(f)** V_y at η_{16}

Figure 1: Generated fields for scales η_0 and η_{16} ($\sigma_Y^2 = 0.5$ case).

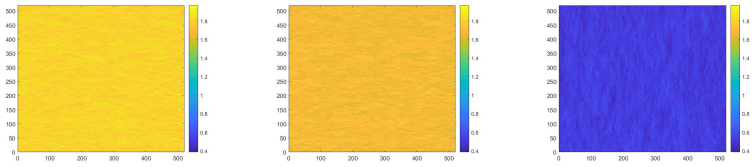
Upscaling by changing the characteristic scale λ_l, as in **[16]** starting from value 1 $[m]$ (reference scale) to values 2,4,8,16 $[m]$. Those fields will be called, from now on, $\eta_{0,2,4,8,16}$. 1000 Monte Carlo realization were generated for every scale, first for a weakly heterogeneous ($\sigma_Y^2 = 0.5$) case, then for a more heterogeneous ($\sigma_Y^2 = 2$) one. Flow problem was solved with a Finite Element Method software (FEM) for each of this realization, prescribing a fixed head on the left side of the domain and a fixed inflow flux on the right side, while top and bottom borders where characterized by a no-flow condition (impermeable boundaries). The impact of boundary conditions have been eliminated by cutting-off the domain along each side of 5 correlation scales (i.e. by 40 $[m]$), producing a reshaped squared domain of side $L = 520$ $[m]$. Figure 1 shows the generated Y field and the V_x and V_y fields for η_0 and η_{16} for the $\sigma_Y^2 = 0.5$ case; it can be observed that the average flow direction is along the horizontal plane (y axis in our model).

0.4 Entropy fields

The entropy of the random variables Y, V_x and V_y has been computed for each point of all the resolution scales. Entropy values in a point of the domain, since they are computed by averaging across the Monte Carlo realizations, actually represent a measure of heterogeneity; in other words low values of entropy for a variable are detected where the likelihood that the variable occurs is high. In fact, prescribed boundary conditions, such



(a) V_x at η_0 **(b)** V_x at η_4 **(c)** V_x at η_{16}



(d) V_y at η_0 **(e)** V_y at η_4 **(f)** V_y at η_{16}

Figure 2: Reshaped fields entropies of V_x and V_y for scales η_0, η_4 and η_{16} ($\sigma^2_Y = 0.5$ case).

Y	η_0	η_2	η_4	η_8	η_{16}
H_m	2.483	2.410	2.277	1.997	0.9590
H_{min}	2.332	2.253	2.117	1.837	0.804
H_{max}	2.643	2.553	2.429	2.129	1.114
% loss	0	2.93	8.30	19.55	61.37

Table 1: Entropy of Y for each scale (case $\sigma^2_Y = 0.5$).

as fixed inflow flux across one side of the domain, resulted in values of entropy close to zero for velocity in that area. An important observation is that this border effect increased when we Upscaled; if we conceptually think of Upscaling as a sort of average, it becomes clear that, moving on coarser scales, this effect influences a growing area. This means as well that we expect the entropy of a field to decrease with Upscaling, as it decreases the variability. Figure 2 reports the entropy of reshaped fields V_x and V_y for scales η_0, η_4 and η_{16} of the $\sigma^2_Y = 0.5$ case. Tables 1,2 and 3 show instead a brief sum up of the results, reporting also the loss of entropy in relative terms (with respect to η_0) for Y, V_x and V_y.

0.5 Information partitioning

Considering the possible triplets formed by the finest scale field as target variable and all the possible couples formed by other scales as source

V_x	η_0	η_2	η_4	η_8	η_{16}
H_m	1.581	1.549	1.466	1.251	0.303
H_{min}	1.412	1.373	1.296	1.072	0.163
H_{max}	1.753	1.714	1.633	1.396	0.444
% loss	0	6.01	7.23	20.87	80.85

Table 2: Entropy of V_x for each scale (case $\sigma_Y^2 = 0.5$).

V_y	η_0	η_2	η_4	η_8	η_{16}
H_m	1.806	1.770	1.686	1.455	0.529
H_{min}	1.622	1.612	1.512	1.253	0.388
H_{max}	1.969	1.928	1.844	1.608	0.661
% loss	0	1.91	6.55	19.38	70.70

Table 3: Entropy of V_y for each scale (case $\sigma_Y^2 = 0.5$).

variables, we studied the evolution of information within Upscaling. Two kind of representation has been made: Venn diagrams in Figure 3 (case $\sigma_Y^2 = 0.5$) reports the information partitioning for triplets η_0-η_2-η_4, η_0-η_2-η_8, η_0-η_4-η_{16} and η_0-η_8-η_{16}, where every circle's area is proportional to the average entropy of the field it represents and the intersection between circles depicts the mutual information between those variables. Instead, pie diagrams in Figure 4 (case $\sigma_Y^2 = 0.5$) shows the shared information partitioning components, normalized with respect to the multivariate mutual information, for fields Y, V_x and V_y. Finally all the results are summed up in Tables 4,5 and 6.

As we previously stated, during Upscaling the total information contained in a model decreases in absolute value (the circles area decrease while moving on coarser scales), moreover we observe that the models

Y	η_{0-2-4}	η_{0-2-8}	η_{0-2-16}	η_{0-4-8}	η_{0-4-16}	η_{0-8-16}
$I(x_{s1}, x_{s2}; x_{tar})$	1.954	1.959	1.941	1.551	1.516	1.168
$U_{x_{s1}}$	0.432	0.805	1.430	0.381	1.003	0.642
$U_{x_{s2}}$	0.005	0.010	0.003	0.013	0.004	0.011
R	1.501	1.128	0.503	1.125	0.503	0.496
S	0.016	0.016	0.005	0.032	0.006	0.019

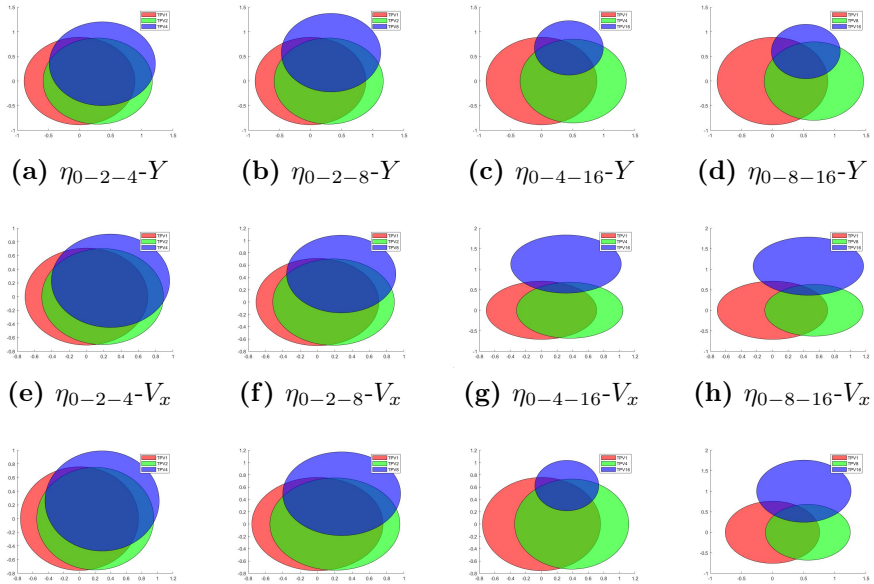
Table 4: Trivariate information of Y for each triplet (case $\sigma_Y^2 = 0.5$).

V_x	η_{0-2-4}	η_{0-2-8}	η_{0-2-16}	η_{0-4-8}	η_{0-4-16}	η_{0-8-16}
$I(x_{s1}, x_{s2}; x_{tar})$	1.306	1.309	1.305	1.035	1.031	0.732
$U_{x_{s1}}$	0.279	0.591	1.168	0.313	0.892	0.588
$U_{x_{s2}}$	0.001	0.002	0.002	0.003	0.004	0.010
R	1.023	0.711	0.134	0.711	0.132	0.126
S	0.004	0.004	0.002	0.003	0.003	0.008

Table 5: Trivariate information of V_x for each triplet (case $\sigma_Y^2 = 0.5$).

V_y	η_{0-2-4}	η_{0-2-8}	η_{0-2-16}	η_{0-4-8}	η_{0-4-16}	η_{0-8-16}
$I(x_{s1}, x_{s2}; x_{tar})$	1.490	1.493	1.487	1.187	1.178	0.850
$U_{x_{s1}}$	0.312	0.660	1.208	0.349	0.898	0.561
$U_{x_{s2}}$	0.001	0.004	0.001	0.004	0.002	0.010
R	1.171	0.824	0.275	0.823	0.274	0.267
S	0.005	0.006	0.002	0.011	0.003	0.013

Table 6: Trivariate information of V_y for each triplet (case $\sigma_Y^2 = 0.5$).

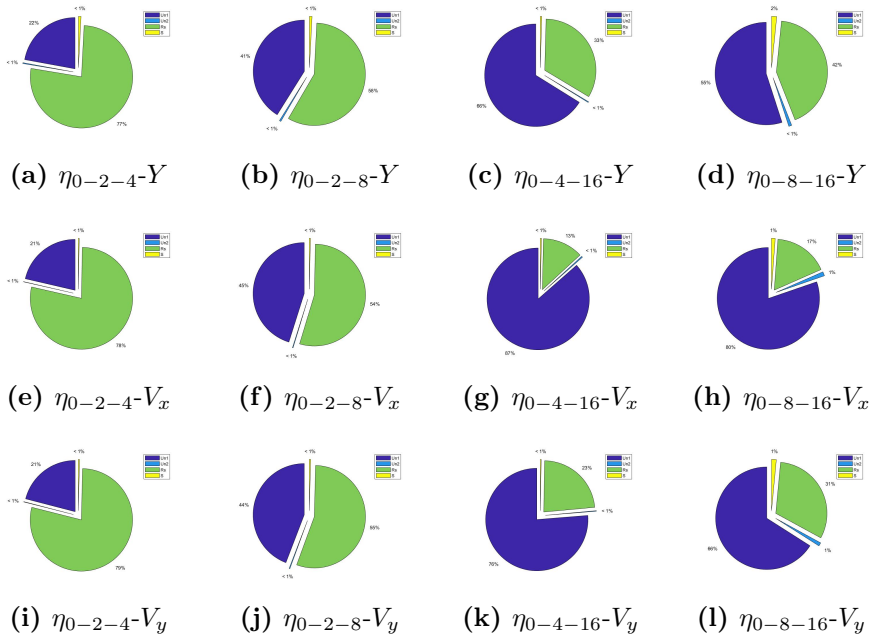


(a) η_{0-2-4}-Y **(b)** η_{0-2-8}-Y **(c)** η_{0-4-16}-Y **(d)** η_{0-8-16}-Y

(e) η_{0-2-4}-V_x **(f)** η_{0-2-8}-V_x **(g)** η_{0-4-16}-V_x **(h)** η_{0-8-16}-V_x

(i) η_{0-2-4}-V_y **(j)** η_{0-2-8}-V_y **(k)** η_{0-4-16}-V_y **(l)** η_{0-8-16}-V_y

Figure 3: Venn diagram representations of entropy and mutual information for Y, V_x and V_y for different triplets (case $\sigma_Y^2 = 0.5$). Red circle represents the target field, while green and blue ones represent, respectively, the more fine and the coarser source fields.



(a) η_{0-2-4}-Y **(b)** η_{0-2-8}-Y **(c)** η_{0-4-16}-Y **(d)** η_{0-8-16}-Y

(e) η_{0-2-4}-V_x **(f)** η_{0-2-8}-V_x **(g)** η_{0-4-16}-V_x **(h)** η_{0-8-16}-V_x

(i) η_{0-2-4}-V_y **(j)** η_{0-2-8}-V_y **(k)** η_{0-4-16}-V_y **(l)** η_{0-8-16}-V_y

Figure 4: Trivariate information partitioning through pie diagrams for Y, V_x and V_y for different triplets (case $\sigma_Y^2 = 0.5$).

become less representative of the reference one (the intersection area decreases more and more when we Upscale). We can see that Venn's diagrams are not made up of concentric circles, more specifically we note that also coarser scales contains some original information in their model. This can be explained considering that, when we Upscale, we are not simply eliminating the extreme values, but we are calculating new values that will tend to concentrate around the initial average values, but they will also be numerically different from the initial ones. Pie diagrams described above, may be used as well to monitor the transmission of information during Upscaling; analyzing the metrics represented by them, we see how some information is "lost" by the shifting of the shared information from the redundant component to the unique component of the more fine source.

0.6 Spatial correlation evolution

Eq. 7,8 and 9 presented above, were applied for pairs of the same variable (i.e.,Y or V_x or V_y) sampled at *lags* ranging from 0 to 32 correlation scales l_y (along the main flow direction). Two different tools are used to depict the results obtained: a graphical one, in which all the collected couples of

variables Y, V_x and V_y are collected into scatter plots and an analytical one, that is the analysis of coefficients ρ,R,U versus the lag. Figure 5 reports scatter plots at $lag = 1\ l_y$, $lag = 6\ l_y$, $lag = 32\ l_y$, for scales η_0,η_{16} of variables Y, V_x and V_y for the weakly heterogeneous case. A comparation of the behaviour of ρ,R,U versus lag is presented in Figure 6 for the case $\sigma_Y^2 = 0.5$.

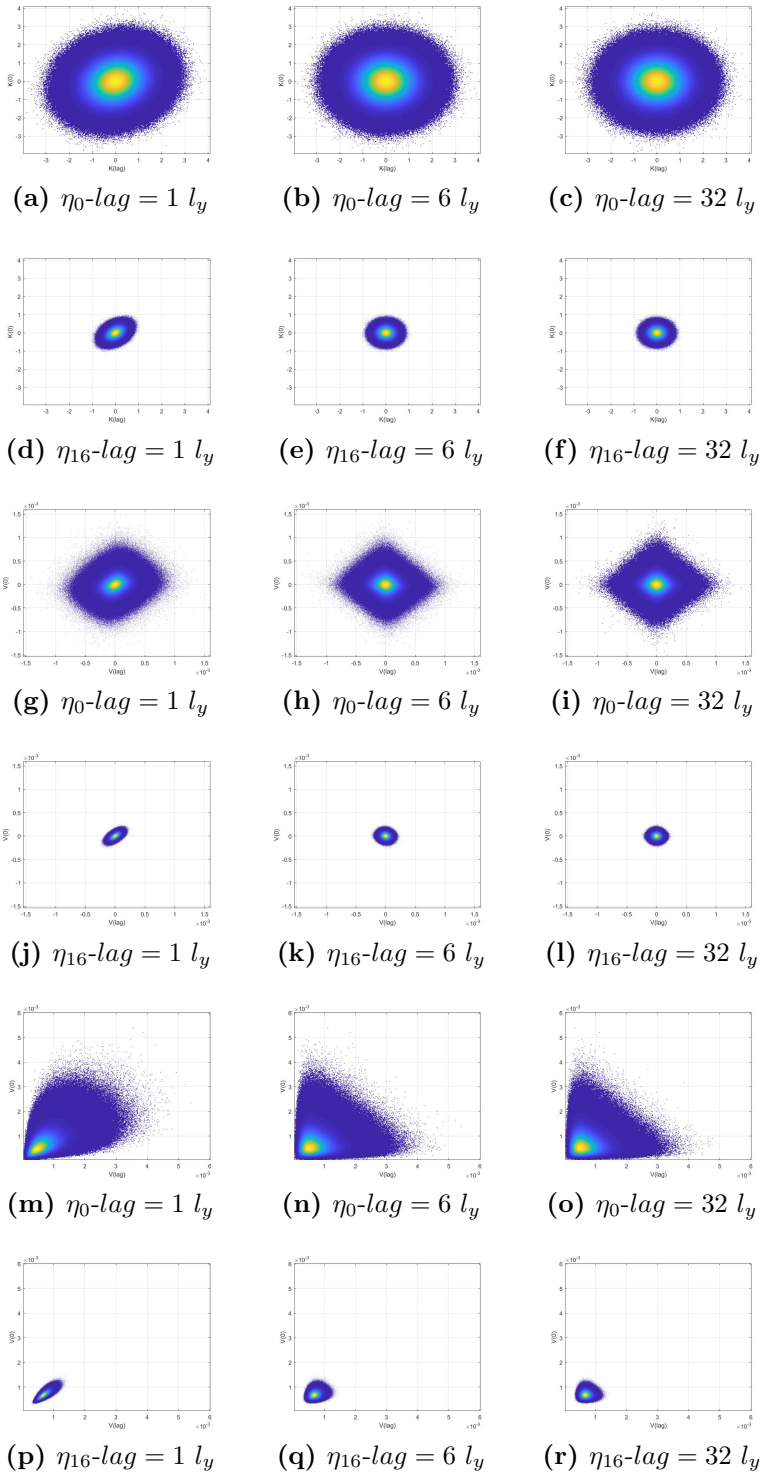
Observing these plots we note the effect of homogenization that Upscaling has, this could be seen by the concentration of the couples of points in more and more narrow regions moving from more fine scales to coarse ones. By analyzing the coefficients ρ, R and U we could state that Upscaling has an effect of linearization on the variables, nevertheless the magnitude of this effect seems also to be limited and not true for high $lags$.

0.7 Conclusions and future applications

This study aimed at quantifying information contained in a model, investigating loss of information and quality of a model generated by means of Upscaling. A new approach has been proposed coupling Monte Carlo method with IT theory tools to analyze the effects of Upscaling on the Y, V_x and V_y fields. These three variables have been studied for two different cases ($\sigma_Y^2 = 0.5$ and $\sigma_Y^2 = 2$), all of them leading to similar results. This work could be divided into three main parts:

1. Quantification of information at a given resolution scale: as we defined it, entropy, is not a measure of heterogeneity of a field, but rather an indicator of the presence of likely or unlikely to occur data (low and high entropy, respectively) in a specific point of the grid. However when averaged on the entire field, entropy indicates how variable is that field and consequently, the average amount of surprise embedded in it. This latter definition let us use entropy to quantify the amount of information included in a model, giving us the possibility to study its behaviour throughout the Upscaling process. Both the $\sigma_Y^2 = 0.5$ and the $\sigma_Y^2 = 2$ cases evidenced that entropy, and so the amount of information of a model, decrease with Upscaling. We propose to refer to relative entropies (compared with reference case) as it will be less significative to consider absolute values of entropy for our purposes.

2. Behaviour of information during Upscaling: while entropy let us quantify the amount of information of different scales, we still did not know enough about the evolution of information during the Upscaling



(a) η_0-$lag = 1\ l_y$ **(b)** η_0-$lag = 6\ l_y$ **(c)** η_0-$lag = 32\ l_y$

(d) η_{16}-$lag = 1\ l_y$ **(e)** η_{16}-$lag = 6\ l_y$ **(f)** η_{16}-$lag = 32\ l_y$

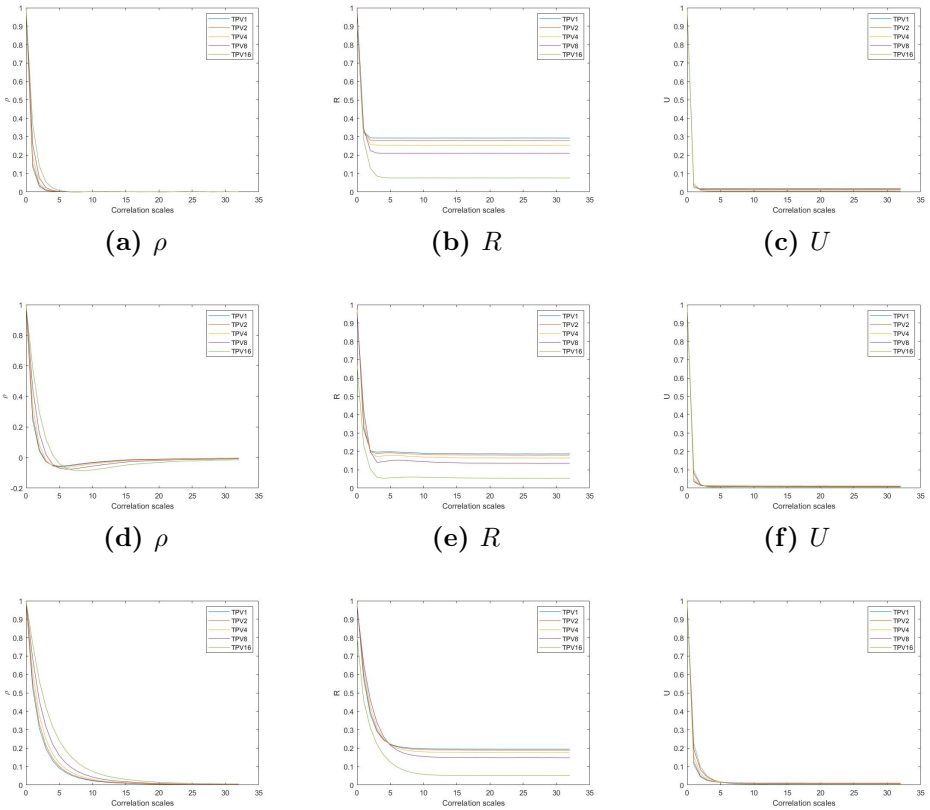
(g) η_0-$lag = 1\ l_y$ **(h)** η_0-$lag = 6\ l_y$ **(i)** η_0-$lag = 32\ l_y$

(j) η_{16}-$lag = 1\ l_y$ **(k)** η_{16}-$lag = 6\ l_y$ **(l)** η_{16}-$lag = 32\ l_y$

(m) η_0-$lag = 1\ l_y$ **(n)** η_0-$lag = 6\ l_y$ **(o)** η_0-$lag = 32\ l_y$

(p) η_{16}-$lag = 1\ l_y$ **(q)** η_{16}-$lag = 6\ l_y$ **(r)** η_{16}-$lag = 32\ l_y$

Figure 5: Scatter plots of Y ((a)-(f)), V_x ((g)-(l)) and V_y ((m)-(r)) (case $\sigma_Y^2 = 0.5$).



(a) ρ **(b)** R **(c)** U

(d) ρ **(e)** R **(f)** U

(g) ρ **(h)** R **(i)** U

Figure 6: ρ,R and U coefficients for variables Y ((a)-(c)), V_x ((d)-(f)) and V_y ((g)-(i)) (case $\sigma_Y^2 = 0.5$). Different colors refer to results associated with different Upscaled fields.

process. Given that the magnitude of information reduces when we Upscale, it is not trivial to understand how much do two different models have in common and how much information is transmitted from one to another. By using the proposed IT metrics we could answer to those questions and give some tools to decide when an Upscaled model is too different from the reference one (and from the finer scale ones). This might facilitate managerial decisions for aquifer characterization, as it helps understanding precisely what happens to information during the Upscaling process. As an example of the application of these metrics, it could be observed that couples η_{2-4} and η_{2-8} share the same amount of information with target variable η_0; using couple η_{2-8} to characterize η_0 is equivalent to do it by using η_{2-4}.

3. Variation in the spatial correlation: lastly we investigated the nature of the relations between different point of the same field, for variables Y, V_x and V_y. It has been showed that Upscaling alters slightly the structure of the fields, by increasing the linear correlation, but not in a significant way, as coefficient U denotes still a low spatial correlation between variables.

When we analyzed the strongly heterogeneous fields we could see that the entropy of the Y field remained almost the same with respect to the weakly heterogeneous ones; even though the variance increased, the borders of the binning became more wide and, as the number of bins did not change, this led to similar results with respect to the weakly heterogeneous case. This was not true for velocity fields, as we could notice an appreciable reduction of entropy for both fields. Moreover, with respect to the $\sigma_Y^2 = 0.5$ case, the relative reduction of entropy during Upscaling increased in this case, in particular for variable V_y (the mean direction of the flow): for this case we observe that we lost more information relevant to flow characterization when we Upscaled the field. Again, other small differences were not significant nor qualitatively different from the previous case. All the results were obtained with a fixed-binning technique, in order to get probability functions from discrete variables, choosing to divide all the data in 15 bins as suggested by [**15**]; we assert that this is a critical part of our work, since it does not exist a rigorous binning procedure for this case. We suggest that an improvement to this limitation could be the use of Kernel Density Estimation (KDE), in order to obviate the use of discrete variables. This could be then a starting point for future applications; while future studies could be also based on real data, instead of synthetic ones, as those provided by [**18**]; this would allow to develop an

IT-based Upscaling technique starting from real data at different resolution scales. Other possible future studies could analyze the Downscaling process, which is much more obscure by now than Upscaling, or simply use this same approach to study transport and multi-component reactive transport problems instead of flow problems like the present study.

Chapter 1

Introduction

Groundwater system is a complex and open system, which is affected by natural conditions and human activities. Natural hydrological processes is conceptualized through relatively simple flow governing equations in groundwater models. Moreover, observation data is always limited in field hydrogeological conditions. Therefore, the predictive results of groundwater simulation often deviate from true values, as a result of the uncertainty of groundwater numerical simulation [**1**]. Addressing uncertainty is an indispensable part of prediction. Groundwater management faces uncertainty on many fronts, in understanding the behaviour of the groundwater system, in anticipating possible future climatic, economic or geopolitical conditions, in prioritising objectives, all of them combining to add ambiguity in the evaluation of management options. Focusing on the first, it is apparent that scientific research has achieved relative success in reducing this uncertainty, culminating in the ability to approximate the behaviour of a groundwater system using a "model". There are, however, limits to the ability of science. Far from being all known, there will always be recognised and unrecognised unknowns this means that a model will always be a simplification of reality, and the predictions it makes will always be uncertain [**2**]. Moreover, often the scale at which transport and flow phenomena in the porous media are best described could be different from the scale at which measurements are available, but also different from the scale required for management decisions [**3**]. In this cases, although it causes the loss of information, it is necessary to use Upscaling or Downscaling techniques to bridge the gap between scales. It goes without saying that, whichever method is used, some information loss in Upscaling is inevitable. One of the most applied modelling tools for analysing a system under conditions of parameter uncertainty is the Monte Carlo simulation. In this technique, we construct a large number of

realizations of the considered domain, say with respect to a property like hydraulic conductivity. Each realization is investigated, yielding a forecast, and the collective behaviour of all forecasts is then analysed, providing the probabilistic information needed for making management decisions under conditions of uncertainty. When it comes to quantify uncertainty Information Theory (IT) [**4**] provides powerful tools, as entropy and mutual information, to evaluate information of random variables and to study their transmission across different models. We note that information as proposed by Shannon and uncertainty are equivalent concepts, as gaining information about something reduces at the same time the uncertainty related to it. The scope of this study is to use these IT metrics to qualify Upscaling technique as proposed in [**16**]. So far, only few efforts have been made to quantify how information loss occurs when moving to coarser scales [**5**], while IT metrics had already been applied to groundwater problems, we cite [**6**],[**7**] in this sense, while [**8**],[**9**] focused on the study of hydrological time-series data. After recalling the governing equation of the flow problem related to the studied scales, entropy in Chapter 2, mutual information and other IT metrics will be presented. Chapter 3 will set-up the problem, explaining the procedure and parameters of the Monte Carlo simulation. Results are presented in Chapter 4, which contains, the entropy calculations of different scales, the information partitioning between scales and the evolution of spatial correlation during Upscaling. Finally, we will analyze the results in Chapter 5. All the results refer to the weakly heterogeneous case ($\sigma_Y^2 = 0.5$), while we refer to the Appendix for the strongly heterogeneous case ($\sigma_Y^2 = 2$).

Chapter 2

Theoretical background

2.1 Modelling of groundwater flows

Groundwater is the water present beneath Earth's surface in soil pore spaces and in the fractures of rock formations. A unit of rock, or an unconsolidated deposit, is called an aquifer when it can yield a usable quantity of water. More specifically, an aquifer is a porous medium domain that contains water (i.e. the entire interconnected void space is filled with water) and that allows water to move through it under ordinary field conditions. The study of the flow through a porous medium is involved in all models of groundwater systems. The biggest problem is that we are not able to know in detail the porous structure of the rock formations and the spatial distribution of their properties. However, since it is neither feasible nor required to model the detailed flow inside the pore space, we shall discuss how the flow could be modelled without information both on the spatial distribution of properties and on the details of the pore space geometry. Uncertainty in groundwater modelling is caused, mostly, by the heterogeneity in aquifer properties, primarily, hydraulic conductivity and porosity. In fact, we can take measurements at certain locations, but they cannot be used to describe with absolute certainty the surrounding areas. In most practical cases, there are never enough data to describe the spatial distributions of these properties in sufficient detail, and interpolation is used to fill in missing data. One way to deal with uncertainty associated with a phenomenon, or a process, is to envision it as a random process, also called a stochastic process. Therefore, the basic idea is not to treat the considered phenomenon (flow or transport) in a deterministic way, but to interpret it in a stochastic way. In practice, we construct a large number of realizations of the considered domain, say with respect to a property like

hydraulic conductivity. Each realization is investigated, yielding a forecast. The collective behaviour of all forecasts is then analysed, providing the probabilistic information needed for making management decisions under conditions of uncertainty. This requires the generating of a large number of realizations of the parameter field and it is done by a random field generation algorithms.

2.1.1 Flow equation

To start, we consider chemically inactive, viscous, Newtonian, of a constant temperature fluid. Two different approaches could be used:

- Eulerian: a fixed control volume is considered in a space frame of reference;
- Lagrangian: an individual fluid parcel is controlled as it moves through space and time.

We will use the former approach, along with conservation of mass and Navier-Stokes equation to derive the flow equation. The mass of the infinitesimal element $d\Omega$ is dM, which could be calculated as:

$$dM = \rho d\Omega \tag{2.1}$$

We assume the absence of sources and sinks terms. Since the control mass M is conserved, recalling Reynolds transport theorem:

$$\frac{dM}{dt} = \int_{\Omega} \frac{d\rho}{dt} d\Omega + \int_{\Gamma} \rho \mathbf{V} \cdot \mathbf{n} d\Gamma \tag{2.2}$$

Which leads us to (using Gauss-Divergence theorem):

$$\frac{d\rho}{dt} + \nabla \cdot (\rho \mathbf{V}) = 0 \tag{2.3}$$

According to **[17]**, we assume relatively incompressible fluids in the pore scale domain, where changes in pressure, hence density, are small compared to the overall pressure, the density can be approximated as constant and so eq. 2.3 can be rewritten as $\nabla \cdot \mathbf{V} = 0$. The second equation we will implement is the conservation of momentum, given by the Navier-Stokes equation:

$$\rho \left(\frac{\delta \mathbf{V}}{dt} + \mathbf{V} \cdot \nabla \mathbf{V} \right) = -\nabla P + \mu \nabla^2 \mathbf{V} \tag{2.4}$$

where P indicates pressure $[MLt^{-2}]$ and μ the viscosity $[MLt^{-1}]$. According to [**17**] and [**21**], the flow field changes slowly over time, and it is reasonable to neglect any time dependence. We introduce Reynold's Number Re, ratio of inertial to viscous forces:

$$Re = \frac{\rho v_c L_c}{\mu} \tag{2.5}$$

where v_c and L_c are the characteristic velocity and length. Roughly speaking, for a sand field and water fluid, typical values of the quantities, involved in eq. 2.5, are $L_c = 10^{-5} \div 10^{-4}$ $[m]$ $\rho = 10^3$ $[kg/m^3]$, $\mu = 10^3$ $[Pa \cdot s]$, $v_c = 10$ $[m/day]$, as consequence, assumption of laminar regime is so reasonable (see [**17**]). It is possible to average the Navier-Stokes equation and derive a linear relation between volumetric flow rate and pressure gradient, known as Darcy's law, as follows:

$$\mathbf{Q} = -\frac{\mathbf{k}}{\mu} A \left(\nabla P - \rho \mathbf{g}\right) \tag{2.6}$$

where Q is the volumetric flow rate $[L^3 t^{-1}]$, $\mathbf{k}$ is the permeability $[L^2]$, A is the cross sectional area $[L^2]$, P is the pressure, ρ is the density $[ML^{-3}]$, g is gravitational acceleration $[Lt^{-2}]$. In case of isotropic and homogeneous media, permeability is reduced to a scalar quantity. In our case, we assume to deal with an heterogeneous system, which is characterized in terms of spatial distribution of permeability. We consider a permeability $\mathbf{k} = k\mathbf{I}$, with $\mathbf{I}$ the identity matrix and random process k described in Chapter 3. It is important to remind the flow occurs only in the pore space, therefore the effective area has to take into account the porosity φ. This is particularly significant in order to calculate the actual velocity, because only a fraction of the total formation volume is available.

$$\mathbf{V} = \frac{\mathbf{q}}{\varphi} \tag{2.7}$$

where we could divide $\mathbf{V}$ into its horizontal and vertical components, for our notation, respectively, V_y and V_x. Putting together Darcy's law with the continuity equation we finally obtain a description of the fluid flow, that could be summed up by:

$$\begin{cases} \varphi \mathbf{V} = -\frac{\mathbf{k}}{\mu} \left(\nabla P + \rho \mathbf{g}\right) \\ \nabla \cdot \mathbf{V} = 0 \end{cases} \tag{2.8}$$

2.1.2 Monte Carlo Analysis

The Monte Carlo method assumes that the modelled phenomena can be represented by a deterministic mathematical model, with known coefficient values. Consider, for example, a single-phase flow in a two-dimensional isotropic aquifer having the spatial structure of the field velocity driven by the previous formulations. Given the knowledge of the domain geometry, the initial conditions, the boundary conditions and the spatial distribution of the hydraulic conductivity, we can solve eq. 2.8 to yield a unique prediction (solution) of future head values and velocities. However, we cannot face the problem in this way because we are uncertain about the input information, in fact we are not able to determine the hydraulic conductivity field in a deterministic way. The Monte Carlo method deals this kind of uncertainty as a probability issue. In practice, instead of attempting to obtain the uncertain, or missing information needed as input, it uses the available data to produce the statistical characteristics (e.g., mean, standard deviation, covariance) of the parameters associated with the considered flow domain (the hydraulic conductivity), and then creates a large number of realizations, each of which is a possible manifestation of the unknown reality. A large number of simulations is conducted in this way, each making use (as input) of one of the realizations of the spatial distribution of the model parameters. Each of the produced outputs contains detailed information on the distributions of the sought variables. In this way, instead of a single deterministic prediction, obtained by solving the given mathematical model with known parameters, we obtain many solutions, one for each realization of the parameter field. From them, we obtain the statistical characterization of the solution. By applying a probabilistic (or statistical) analysis to these many equally likely to occur outcomes, we can provide quantitative, albeit probabilistic, answers to questions like, what is the probability that, at a certain location, the flow velocity is lower than a certain target? The Monte Carlo procedure seems simple and straightforward despite that it has a high computational cost and we need to know the probability distribution of the parameter of interest (hydraulic conductivity) to generate the many realizations of its spatial distribution. The procedure for the stochastic analysis described above is the following. The generation of random realizations, required in the Monte Carlo simulation, calls for the generation of a sequence of random numbers. Although such sequence can be obtained, for example, by throwing a dice repeatedly, we must use a computerized pseudo random number generator. This is based on a mathematical algorithm, programmed for a computer, that can generate a seemingly random sequence of numbers with a certain

precision. Actually, the process is only pseudo-random, because we need a seed number to generate a sequence of random numbers by means of a computer and it is possible that the same sequence of numbers will be generated every time if we do not change the seed number. Using an algebraic transformation, meaning replacing one variable by another, defined by a functional relation, this sequence of random numbers, can be mapped onto a Gaussian probability distribution. Armed with a random number generator and a probability distribution, we can now generate the random fields (of parameters, such as hydraulic conductivity) needed as input for the Monte Carlo simulations of the considered mathematical model. We divide the domain of interest into a number of small cells, each assumed to be homogeneous. Selecting one cell to start from, we can randomly (pseudo-randomly) assign to it a parameter value. We then move to the next cell and assign to it another random hydraulic conductivity value. We continue this process, until the entire transmissivity field is defined. Summarising, to obtain a stochastic field of hydraulic conductivity values, we start by inserting a seed number into a random number generator. The random number generator then produces a sequence of pseudo-random numbers taking into account the prescribed spatial covariance function in order to accommodate the spatial correlation of the aquifer properties. Based on the assumed pdf (for example, log-normal distribution), and a provided mean and standard deviation, these numbers are then mapped into the sought hydraulic conductivity values. The values that are assigned to the various cells are not random but follow the properties of the probability distribution specified. Thus, the values at neighbouring cells should be conditionally generated, on the basis of known information concerning the covariance. Generally, there are two types of random parameter fields that can be generated: conditional and unconditional. The conditional (or constrained) random parameter field must satisfy the requirement that its values at sampled points should be exactly equal to those actually measured, or observed, there. These measured values are true values (known by measurements); hence, the generated realizations should conform to this constraint, as in our case. In an unconditional (i.e., unconstrained) random parameter field generation instead, the observed values are ignored, or better, they are unknown.

2.2 Upscaling

In groundwater flow problems scale inconsistency is very common. The scale at which transport and flow phenomena in the porous media are best

described is usually very different from (larger than) the scale at which measurements are available, but also very different from (smaller than) the scale required for management decisions **[3]**. When focusing at the modelling of groundwater flow and transport the following spatial scales are often distinguished:

The pore scale ($10^{-6} \div 10^{-2}$ $[m]$): the scale at which flow and transport through porous media is described in terms of forces and mass fluxes within the fluid phase and the solid phase and between these phases. Groundwater flow for instance is described by the Navier-Stokes equations.

The core scale ($10^{-1} \div 10^{0}$ $[m]$): the scale at which flux and transport are described in terms of continuity equations and simplified flux equations such as Darcy's law and Fick's law. This is exactly the scale at which measurements of hydraulic properties are performed on samples from drilling cores;

The model block scale ($10^{1} \div 10^{2}$ $[m]$): the scale of blocks or elements of numerical flow and transport models;

The local scale ($10^{2} \div 10^{3}$ $[m]$): the scale at which groundwater flow and -transport is considered as three-dimensional. Examples of local scale groundwater problems are pollution and remediation studies around waste sites and the assessment of travel time distributions in protection areas around drinking water wells;

The regional scale (horizontal dimensions $10^{3} \div 10^{5}$ $[m]$): the scale at which the subsoil is divided into permeable layers (aquifers) and less permeable layers (aquitards). The pore scale is usually not considered in practical groundwater modelling studies. Instead, one directly starts with the simplified core scale equations and the representative parameters are measured directly on sediment cores. These equations are then used to describe local scale and regional scale groundwater problems. However, hydraulic properties such as hydraulic conductivity and dispersivity, that are measured on sediment cores, cannot be used to describe flow and transport at larger scales. The reason for this is that hydraulic properties usually exhibit a large spatial heterogeneity. The techniques we use to bridge the gap between these scale discrepancies are called Upscaling and Downscaling.

2.2.1 Upscaling of hydraulic conductivity

The Upscaling of aquifers properties for flow simulation is one of the most important steps in the workflow for building predictive models. It is basically the process through which we scale-up properties defined at a fine-grid system, like hydraulic conductivity, to equivalent properties defined at a coarse-grid system in such a way that the two systems act as similarly as possible. For example, the equivalent hydraulic conductivity K_{eq} of a homogeneous medium (upscaled scale) is derived from the hydraulic conductivities of an equivalent heterogeneous medium (reference scale) that, for the same boundary conditions, would give the same flux. It goes without saying that, whichever method is used, some information loss in Upscaling is inevitable.

2.2.2 The Upscaling problem

Despite its importance, Upscaling is not a straightforward process because we need to bridge the gap between the scale's discrepancies and, at the same time, we must retain the geological realism to effectively represent fluid flow in the reservoir. The common problem of Upscaling methods is that they tend to smear out the spatially continuous extremes, such as shale barriers and open fractures. Two strategies are possible for decreasing the information loss due to Upscaling: one is to decrease the extent of Upscaling, while the other to minimize the information loss in the Upscaling procedure. However, there is not a well-established methodology to measure the quality of Upscaling routines. So, how to qualify the Upscaling results, that is, whether an Upscaled results provides a good or bad approximation, is one of the outstanding problems remaining in this research field [**20**]. According to [**20**], the main method that could be used to assess the quality of Upscaling are the following: The simulation results of the Upscaled coarse model and the original fine model can be compared. If they both have the same flow performances, the Upscaling results can be considered to be a perfect representation of the original domain. In practice this qualifying method is directly applicable, but it may be time and money consuming because the simulation at the finest grid are difficult to obtain. The simulation results can be compared with performance parameters of the reservoir, such as well pressures, cumulative produced oil, water breakthrough time or saturation at

specific locations. The limitation of this approach is that usually these parameters are not available at the finest grid scale or is too expensive to obtain them. These are the simplest methods, nowadays others have been proposed. For example [**5**] studies the conservation of spatial autocorrelation to assess information loss through the Upscaling process, while [**6**] have used concept of information theory to develop an IT-based Upscaling technique.

2.3 Information Theory

In 1948, Claude Shannon published a paper called *A Mathematical Theory of Communication* [**4**], which heralded a transformation in our understanding of the concept of information. Before Shannon's paper, information had been viewed as a kind of poorly defined miasmic fluid, but after his work it became clear that information is a well-defined and, above all, measurable quantity [**19**]. The importance of this theory is that it provides a mathematical definition of information, which allows to quantify precisely the information contained in discrete random variables.

2.3.1 Information and entropy

Information is the resolution of uncertainty [**4**]. Shannon not only gave theoretical notions about the meaning of information, but above all he has provided mathematical tools for its quantification. To be useful, a mathematical formulation must have a minimal set of properties which are known as "Shannon's desiderata":

Continuity : The amount of information associated with an outcome increases or decreases continuously as the probability of that outcome changes.

Symmetry : The amount of information associated with a sequence of outcomes does not depend on the order in which those outcomes occur.

Maximal Value : The amount of information associated with a set of outcomes cannot be increased if those outcomes are already equally probable.

Additive : The information associated with a set of outcomes is obtained by adding the information of individual outcomes.

Information is inextricably related to probability, perhaps it is useful to consider jointly the concept of surprise and probability to better understand what is meant by information. For example, suppose we are given a coin, and we are told that it lands heads up 90% of the time. When this coin is flipped, we expect it to land heads up, so when it does so, we are less surprised than when it lands tails up. In practice, the more improbable a particular outcome is, the more surprised we are to observe it. One way to express this might be to define the amount of surprise of an outcome value X of a random variable to be 1 divided by the probability of X or $1/p\,(X)$, so that the amount of surprise associated with the outcome value X increases as the probability of X decreases. However, in order to satisfy the additivity condition above, Shannon showed that it is better to define surprise as the logarithm of $1/p\,(X)$. This is known as the Shannon information of X and it is also called surprisal because it reflects the amount of surprise when that outcome is observed. If we use logarithms to the base 2 then the Shannon information of a particular outcome is measured in *bits*:

$$i\,(X) = \log_2 \frac{1}{p\,(X)} \tag{2.9}$$

where i stands for Shannon information. A general rule for logarithms states that:

$$\log_2 \frac{1}{p\,(X)} = -\log_2 p\,(X) \tag{2.10}$$

So that the previous equation can be written as:

$$i\,(X) = -\log_2 p\,(X) \tag{2.11}$$

In essence, Shannon information is a measure of surprise and, higher is the surprise of an outcome, higher will be its information. We must not confuse bits and binary digits because they are different types of entities. Even though the word bit is derived from binary digit, there is a subtle, but vital, difference between them. A binary digit is the value of a binary variable, where this value can be either a 0 or a 1, but a binary digit is not information per se. In contrast, a bit is a definite amount of information. It is clear from previous equations that to quantify the amount of surprise (i.e. information) of the outcome of a random variable, we need to know the probability of the possible outcomes which collectively define the probability distribution $p\,(X)$ of the random variable X. However, we are not

usually interested in the surprise of one outcome of a random variable, but we would like to know how much surprise (i.e. information), on average, is associated with the entire set of possible values. That is, we would like to know the average surprise defined by the probability distribution of a random variable. For this purpose Shannon defined the concept of entropy, which lies at the core of information theory and allows to quantify average information. The average surprise of a variable X which has a distribution $p(X)$ is called the entropy of $p(X)$, this average surprise is represented as $H(X)$ and it has the unit of measure of *bits* (when the logarithm has base equal to 2). For convenience, we often speak of the entropy of the variable X, even though, strictly speaking, entropy refers to the distribution $p(X)$ of X. Entropy is computed as follow:

$$H(X) = -\sum_{i=1}^{N} p(X_i) \log_2 p(X_i) \tag{2.12}$$

Where i is the bin number, N is the number of populated bins of an histogram, and $p(X_i)$ is the proportion of data falling into the i-th populated bin, subjected to the condition $\sum_{i=1}^{N} p(X_i) = 1$. Previously, we have said that entropy quantifies the average information of a random variable, but entropy can be also interpreted as a measure of uncertainty. Conceptually, when we gain information about something, its uncertainty is reduced, so information and entropy are two sides of the same coin. Average information shares the same definition as entropy, but whether we call a given quantity information or entropy usually depends on whether it is being given to us or taken away. For example, a variable may have high entropy, so our initial uncertainty about the value of that variable is large and is, by definition, exactly equal to its entropy. If we are then told the value of that variable then, on average, we have been given an amount of information equal to the uncertainty (entropy) we initially had about its value. Thus, receiving an amount of information is equivalent to having exactly the same amount of entropy (uncertainty) taken away.

2.3.2 Joint entropy and mutual information

The concepts of information and entropy of one random variable can be extended to the case of interacting random variables. In this case

we do not refer to the probability of one single random variable, but we must consider the joint probability of multiple variables. More specifically, given two random variables X, Y that are defined on a probability space, the joint probability distribution for X, Y is a probability distribution that gives the probability that each of X, Y falls in any particular range or discrete set of values specified for that variable. In the case of only two random variables, this is called a bivariate distribution, but the concept generalizes to any number of random variables, giving a multivariate distribution. The entropy of a joint distribution is a straightforward generalisation of the entropy of a single variable:

$$H(X,Y) = -\sum_{i=1}^{N}\sum_{j=1}^{M} p(X_i, Y_j) \log_2 p(X_i, Y_j) \tag{2.13}$$

Where N is the number of different values of X and M is the number of different values of Y, while $p(X_i, Y_i)$ is the joint probability of the X_i and Y_i values. In the case in which these two variables are somehow correlated, whatever is the nature of the relationship, it is possible that they shared an amount of information. This portion of information that the observation of a variable provides about the other variable is called mutual information and it can be computed as:

$$I(X,Y) = \sum_{i=1}^{N}\sum_{j=1}^{M} p(X_i, Y_j) \log_2 \frac{p(X_i, Y_j)}{p(X_i)\, p(Y_j)} \tag{2.14}$$

When the variable X,Y are correlated, their joint entropy is not given by the some of the entropy of X and the one of Y, because some information is shared between them. The joint entropy of correlated variables can be obtained in this way:

$$H(X,Y) = H(X,) + H(Y,) - I(X,Y) \tag{2.15}$$

While, for independent variables, it holds:

$$H(X,Y) = H(X,) + H(Y,) \tag{2.16}$$

Historically, entropy and mutual information have been represented by means of Venn's diagrams.

The graphical representation is straightforward: the area of circles is proportional the entropy of variables, while the intersection between circles represents the mutual information. This is also useful to

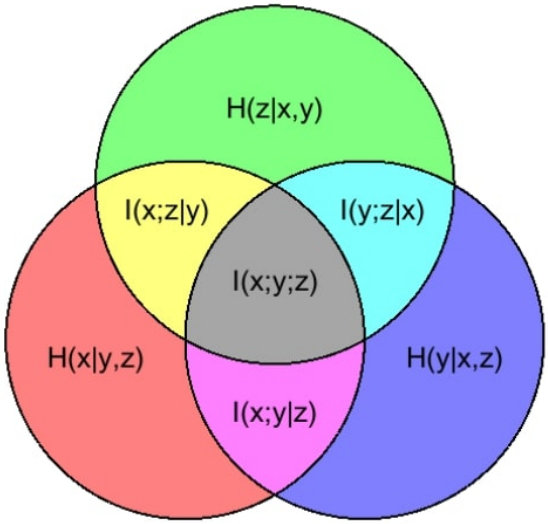


Figure 2.1: Trivariate information partitioning

understand the meaning of entropy and mutual information when three variables are considered. In this case the joint entropy is the overall area given by the figure formed by the three circles, thus it is lower than the entropy we would obtain by summing each entropy individually. This because, as mentioned above, some information is shared by the variables. When three variables are involved, it is possible to compute the information that two variables, let's say X and Y, provide about the other one, which is Z in this case:

$$I(X,Y;Z) = \sum_{i=1}^{N}\sum_{j=1}^{M}\sum_{k=1}^{K} p(X_i,Y_j,Z_k)\log_2\frac{p(X_i,Y_j,Z_k)}{p(X_i,Y_j)\,p(Z_k)} \quad (2.17)$$

Where $p(X_i,Y_j,Z_k)$ is the multivariate joint probability of the X_i, Y_i and Z_i values. Moreover, we can also compute the information $I(X;Y;Z)$ in common between all the three variables, which is represented graphically by the central area (grey area):

$$I(X;Y;Z) = I(X;Z) + I(Y;Z) - I(X,Y;Z) \quad (2.18)$$

2.3.3 Information partitioning

In a complex natural system, we expect that many source variables influence the behaviour of a target variable. Consider two source

variable X,Y and a target one Z. The multivariate mutual information $I(X, Y; Z)$ quantify the total information that sources provided to the target, but it does not separate unique influences or indicates how the sources may be acting jointly. Moreover, if we consider the relationships between the target and a single source, we would therefore be neglecting the presence and the effects of the other source variables. However, this is not a negligible aspect. For example, a variable may be synchronized with another variable that is driving the target, or many source variables may be jointly driving the target. More specifically, when two sources inform a target, the target may receive information uniquely from each source, redundantly from both sources, synergistically from both sources, or in some combination of these. These problems highlight the need to understand the nature of multivariate mutual information $I(X, Y; Z)$, which can reveal much about the function or process that physically links the two sources with the target, in addition to any relationship between sources. Recent research on information partitioning [**8**] has enabled more precise classification of the nature of multivariate shared information. Information partitioning categorizes shared information quantities between multiple source variables and a target variable as either synergistic, unique, or redundant. Redundant information (R) is the information that multiple sources provide to a target such that they overlap in their information content. Unique information (U) from a source refers to the information it shares with a target that is not redundant with information provided by another source. Synergistic information (S) refers to the information that two sources provide to a target only jointly. We can compute:

$$I(X, Y; Z) = U_1(X; Z) + U_2(Y; Z) + R + S \tag{2.19}$$

Where U_1, U_2, R and S are non-negative quantities. $U_1(X; Z)$ and $U_2(Y; Z)$ are, respectively, the information that sources X and Y share with the target Z; R is interpreted as overlapping shared information; S is a cooperative provision of shared information that is possible to gain only if X and Y are considered jointly. The individual mutual information between each source and the target can be decomposed as:

$$I(X; Z) = U_1(X; Z) + R \tag{2.20}$$

$$I(Y; Z) = U_2(Y; Z) + R \tag{2.21}$$

If one source provides a larger amount of information than another (i.e., $I(X;Z) > I(Y;Z)$), as reflected through a higher reduction in uncertainty of the target, it will have a higher uniqueness (U), indicating its dominant influence as an individual source. A high synergy (S) indicates that two sources provide information jointly, since both sources must be known together to reduce target uncertainty. In contrast redundancy (R) indicates an overlap in information due to lagged synchronization or correlation between sources, and the extent to which either source reduces the same target uncertainty. Since mutual information quantities in previous equations are directly computable, one of the four components S, U_1, U_2, and R must be obtained independently to solve this underdetermined system. To address this, we perform information partitioning using a recently developed approach where redundancy, R, is obtained based on scaling by source dependency $I(X;Y)$, such that independent sources are minimally redundant (maximally unique) and highly dependent sources are assumed to be maximally redundant (minimally unique) [**9**]. The redundancy can be computed as follows [**8**]:

$$R = R_{min} + I_s \left(R_{MMI} - R_{min} \right) \tag{2.22}$$

Where:

$$R_{MMI} = min[I(X;Z), I(Y;Z)] \tag{2.23}$$

$$R_{min} = max\left(0, -I(X;Y;Z)\right) \tag{2.24}$$

With:

$$I_s = \frac{I(X;Y)}{min[H(X), H(Y)]} \tag{2.25}$$

R_{MMI} denotes the minimum mutual information between sources and target and it represents the upper limit for redundancy while Is represent the source dependency. R_{min} denotes the lower limit for redundancy, but its interpretation is more subtle. In the case where the sum of the two mutual information is greater than $I(X,Y;Z)$, then $I(X;Y;Z)$ will be positive. In this case, some of the information about Z provided by knowing X is also provided by knowing Y, causing their sum to be greater than the information about Z from knowing both together. That is to say, there is a redundancy in the information about Z provided by the X and Y variables. In the case where the sum of the mutual information is less than $I(X,Y;Z)$, the multivariate mutual information $I(X;Y;Z)$ will be negative. In this case, knowing both X and Y together provides more information

about Y than the sum of the information yielded by knowing either one alone. That is to say, there is a synergy in the information about Z provided by the X and Y variables. Rmin therefore provides a minimum bound for R: for cases where $I(X;Y;Z) = S - R$ is negative, indicating that $R > S$, any chosen metric for R must be greater than or equal to the positive value of $R - S$ in order to be S non-negative.

2.4 Spatial correlation

One variable might exhibit a certain spatial correlation throughout the field (depending on the lag) and this correlation might be linear or non-linear. There are now two ways to visualize this kind of relations: one is graphical and the other analytical. A graphical method to detect correlations is by simply plotting into scatter plots, for a given scale at a given lag all the possible couples of values formed by a point and another point distant lag from the former. For a variable X, sampling all its possible couples at (i) one location and (ii) a location which is distant of a given lag (along the main flow direction) three coefficients, ρ, R and U could be computed as **[10]**,**[11]**,**[7]**: **[10]**,**[11]**,**[7]**:

$$\rho(X, X(lag)) = \frac{Cov(X(x), X(x+lag))}{\sigma_{X(x)}\sigma_{X(x+lag)}} \tag{2.26}$$

$$R(X, X(lag)) = \{1 - exp[-2I(X(x), X(x+lag))]\}^{1/2} \tag{2.27}$$

$$U(X, X(lag)) = 2\frac{I(X(x), X(x+lag))}{H(X(x))\,H(X(x+lag))} \tag{2.28}$$

being $Cov(X(x), X(x+lag))$ the covariance, $\sigma_{X(x)}$ and $\sigma_{X(x+lag)}$ the standard deviations of $X(x)$ and $X(x+lag)$, $I(X(x), X(x+lag))$ the mutual information between two points of a couple and $H(X(x))$ and $H(X(x+lag))$ the entropies of the two points. Coefficient U (or uncertainty coefficient) lies between 0 and 1: when the uncertainty coefficient is zero, it means that $X(x)$ and $X(x+lag)$ are not dependent on each other; if its value is unitary, the knowledge of $X(x)$ is able to completely predict $X(x+lag)$, and the opposite is also true **[7]**. The Bravis-Pearson index, ρ, is known as the linear correlation coefficient or Pearson correlation coefficient. It is a measure of the linear dependence of two random variables. if $|\rho| = 1$,

a perfect linear relationship exists between $X(x)$ and $X(x+lag)$ and the variables are fully correlated; instead, if $\rho = 0$, the variables are not correlated [**7**]. R takes values in the range [0,1]. R is zero if $X(x)$ and $X(x+lag)$ are independent, and is unity if there is an exact linear or nonlinear relationship between $X(x)$ and $X(x+lag)$ [**23**]. Taking the spatial average, over all the possible pairs of spatial locations in the domain, we could study the nature of the spatial dependence, investigating the effect of Upscaling over the latter for Y, V_x and V_y.

Chapter 3

Methodology

In this chapter we present the setting of the problem and the methodology used to address it. It will include informations about the software used and the MATLAB scripts implemented and it will be organized, into four parts:

- Problem setting
- Entropy calculations
- Information partitioning
- Spatial correlation

The first part will regard all the software settings and characteristic that have been used in order the create and Upscale the aquifer model with the Monte Carlo method. Secondly we will explain in details the methodology used for the estimation of probability function for variables Y, V_x and V_y and its application for entropy calculations. Thirdly we will adapt the IT metrics introduced in Chapter 2 to our problem and present the main features of the method implemented. Finally we will introduce the methods used to evaluate linear and non-linear correlations among our variables.

3.1 Problem setting

As stated in Chapter 2, the uncertainty that characterizes groundwater flow could be tackled by using Monte Carlo method. This requires the generation of random fields, in which the hydraulic conductivity K is modelled as an isotropic randomly generated field with

imposed statistical features. The domain is a $2D$ confined aquifer of side $L = 600$ $[m]$ with constant thickness on which it is imposed a uniform grid with $n_x = n_y = 600$ squared elements. The hydraulic conductivity is modelled as:

$$K = K_g e^{Y(x,y)} \tag{3.1}$$

where K_g is a typical value of limestone hydraulic conductivity and $Y(x, y)$ is a zero-mean second-order stationary random process characterized by a truncated power law variogram (TPV) with correlation lengths $l_x = l_y = 8[m]$ and an isotropic covariance function:

$$C(h) = \gamma_G^2(h, \lambda_u) - \gamma_G^2(h, \lambda_l) \tag{3.2}$$

where, for $m = l, u$:

$$\gamma_G^2(h, \lambda_m) = \sigma_Y^2(\lambda_m)\,\rho(h/\lambda_m) \tag{3.3}$$

$$\sigma_Y^2(\lambda_m) = A\frac{\lambda_m^{2H}}{2H} \tag{3.4}$$

$$\rho(h/\lambda_m) = e^{-\frac{h}{\lambda_m}} - \left(\frac{h}{\lambda_m}\right)^{2H} \Gamma(1 - 2H, h/\lambda_m) \tag{3.5}$$

being h the distance (lag), H the Hurst coefficient (0.333 in our model), Γ the gamma function, A the variance, λ_u the characteristic scale associated with the upper frequency cut-off and λ_l characteristic the scale associated with the lower frequency cut-off. These particular kind of relationship has been proved to render the expressions for integral scale and variance dependent on domain size in a manner consistent with observation **[16]** by filtering out (truncating) high and low-frequency cut-offs. The characteristic scales of the lowest and highest frequency modes (cut-offs) are related, respectively, to domain and sample (support) size. The model studied here is based on Upscaling by changing the characteristic scale λ_l starting from value 1 $[m]$ (reference scale) to 2,4,8,16 $[m]$ that will be called, from now on, $\eta_{0,2,4,8,16}$. 1000 Monte Carlo realization were generated for every scale, first for a weakly heterogeneous ($\sigma_Y^2 = 0.5$) case, then for a more heterogeneous ($\sigma_Y^2 = 2$) one, with the help of a software produced by the DICA department of *Politecnico di Milano*, called RF GEN. After noting that the horizontal direction coincides with the y axis and the vertical direction with the x axis, we solved the flow problem (as presented in Chapter 2) for each of this realization,

prescribing a fixed head on the left side of the domain and a fixed inflow flux on the right side, while top and bottom borders where characterized by a no-flow condition (impermeable boundaries). We used *FreeFem++*, a free and open source software to solve flow equations using the Finite Element Method (FEM). The results, along with Y fields, constituted the data set of our analysis.

3.2 Entropy calculations

Shannon's entropy has been already presented in Chapter 2, nevertheless to compute it we had implicitly considered that we were given the probability functions of our study variables. This is clearly not the case, so we had to find a way to obtain a probability function starting from our discrete random variables. We decided to go for the fixed-binning technique, that is to use a fixed number of bins and compute for every point in the grid the probability to stay within a specific bin. There are several studies that face the problem of density estimation and optimal data-based histograms, we cite as a reference **[12]**,**[13]**,**[14]**,**[15]**. However, there is not a univocal procedure about binning. In this work we decided to use a constant number of bins equal to 15, as suggested by **[15]** when the observations of a variable are between five hundred and a thousand, and we also maintained fixed the width of the bins. The extreme borders of our bins will be, for every variable at every scale, the maximum and minimum value of that variable across all scales; as Upscaled fields are obtained from the reference scale, these values will be located in it. We then computed entropy of a variable for all the points in the domain across all the Monte Carlo representations. As a result a grid of values of entropies was generated, with low values of entropy in a point indicating low uncertainty about the variable and high values of entropy indicating high uncertainty about the considered variable. The imposed boundary condition generated values of entropy close to zero along the borders, with this effect growing in size as we Upscale (this phenomenon will be explained later in Chapter 4 While an averaged value of entropy has been considered for our studies, we decided to cut-off the borders by eliminating a portion of external frame of the domain (5 l_y or 40 $[m]$ per each side) finally obtaining a $2D$ domain of side $L = 520$ $[m]$.

3.3 Information partitioning

The IT metrics used in this part have been introduced in Chapter 2, in which we considered the possible triplets formed by the finest scale field as target variable and all the possible couples formed by other scales as source variables. All the metrics were computed as described in Chapter 2 and for the probability function estimation the bins used were the same as those described in the previous section. The method used in the scripts written to compute them is actually the same of the one used to compute entropies. We will propose a graphic way to visualize concepts related to IT metrics using Venn diagrams, used to present results of this part in Chapter 4.

3.4 Spatial correlation

For this part the coefficients used were those introduced in Chapter 2; a MATLAB script was implemented, to compute them for every scale and for every variable, collecting the results obtained, and plotting them in the same graph. The couple of points (both used for coefficient calculations and for scatter plots) were taken collecting all the possible couples with an increasing lag between each other along the mean flux direction (y). Being the correlation length of Y along y axis $l_y = 8$ $[m]$, $lags$ considered could range from 1 to 32 l_y, while the initial point had to stay within the left half of the domain.

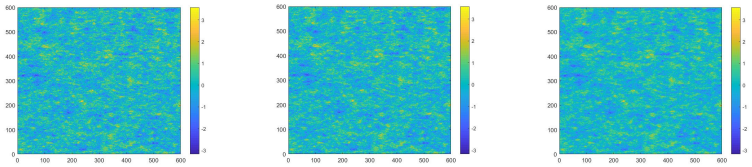
Chapter 4

Results

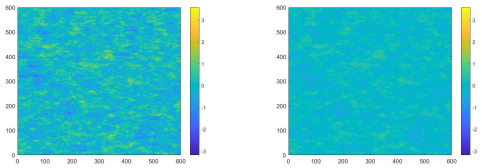
In the following chapter the results of our simulation are presented. This chapter is divided into two parts: in the first we will discuss the main features of the weakly heterogeneous field ($\sigma_Y^2 = 0.5$), that are, some representation of the fields produced by solving the flow problem for every field, the entropy of variables Y, V_x and V_y, computed with the equations presented in Chapter 1 and method explained in Chapter 2 the information partitioning of the aforementioned variables, with the metrics proposed by [**8**] and finally we will present a study of the transformation of the spatial correlation of Y, V_x and V_y. In the second part we will briefly discuss the same results for the strongly heterogeneous field ($\sigma_Y^2 = 2$), underlining the possible differences with the former case, while we decided to report the complete results in the Appendix. All the data mining and calculations were performed on MATLAB, using scripts written specifically for this study.

4.1 Y and velocity fields

While we easily plotted the Y field by simply reshaping the output file of the *RF GEN* software, we needed to rearrange the velocity fields given by solving the flow problem with *FreeFem++* by transforming the triangular mesh values into a rectangular grid that we use also later on to compute all the other statistics. From now on, we will refer to the finest scale as η_0, while the coarser fields will be called η_2, η_4, η_8, η_{16}. Instead V is computed as the log_{10} of the velocity of the fluid. Figure 4.1 reports the Y field for the Monte Carlo representation



(a) η_0 **(b)** η_2 **(c)** η_4



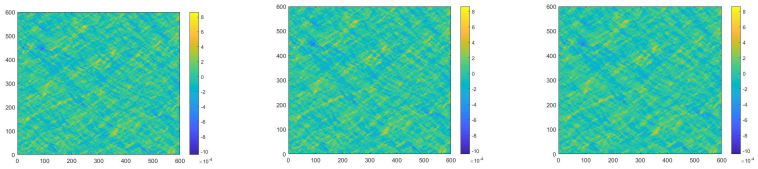
(d) η_8 **(e)** η_{16}

Figure 4.1: Y fields for each different resolution scale (case $\sigma_Y^2 = 0.5$).

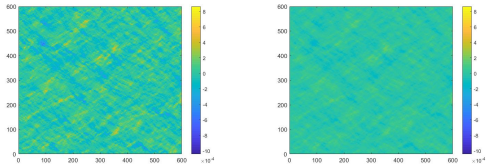
number 1 for every scale, while Figures 4.2 and 4.3 represent the V_x and V_y fields, respectively, for the same representations. Lastly, velocity V is showed in Figure 4.4. Firstly we can observe the structure of the conductivity (its log_{10}) and relate it to the velocity field and notice the presence of some preferential channels, this can be seen especially from the plots of V_y. All the variables studied tells us that upscaling reduced heterogeneity by "smoothing out" the fields, and this can be understood clearly when observing the figures. As it can be noted, the peak values smooths out and the standard deviation decrease while we Upscale and the field becomes more homogeneous; this can be explained considering Upscaling as an operation in which the less likely to occur values are cut off.

4.2 Entropy

In this section we will firstly compute the entropies for Y and velocity fields, being aware that border effects could affect our data. As explained in Chapters 1 and 2, entropy is a measure of uncertainty; this means, in our case, that a high entropy is related to a high uncertainty about our model. Roughly speaking, the higher the entropy, the higher the variability of the field (and the information related to it). This means that we expect the entropy (and so the information) of a field to decrease with Upscaling, with the decrease

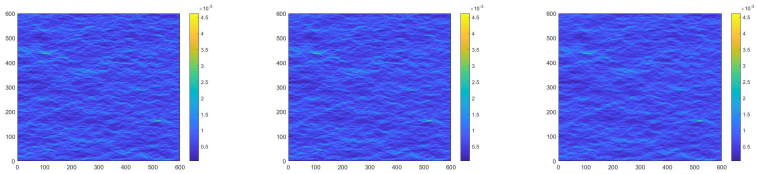


(a) η_0 **(b)** η_2 **(c)** η_4

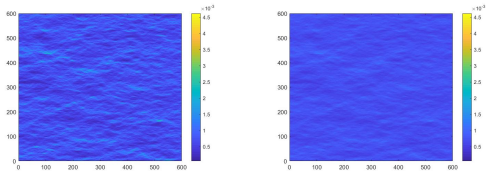


(d) η_8 **(e)** η_{16}

Figure 4.2: V_x fields for each different resolution scale (case $\sigma_Y^2 = 0.5$).

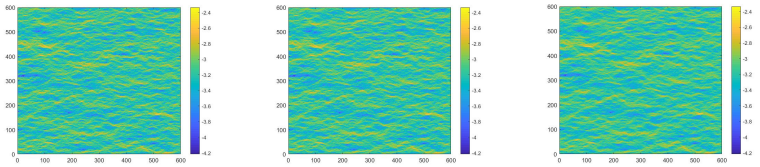


(a) η_0 **(b)** η_2 **(c)** η_4

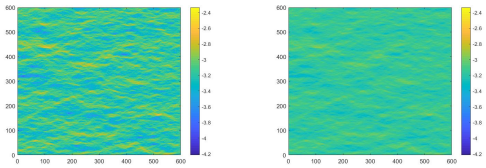


(d) η_8 **(e)** η_{16}

Figure 4.3: V_y fields for each different resolution scale (case $\sigma_Y^2 = 0.5$).



(a) η_0 **(b)** η_2 **(c)** η_4



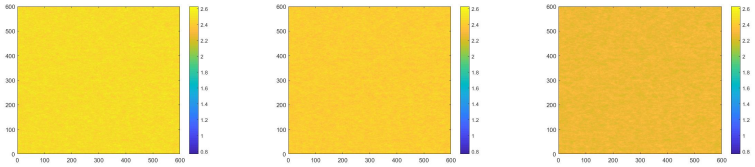
(d) η_8 **(e)** η_{16}

Figure 4.4: V fields for each different resolution scale (case $\sigma_Y^2 = 0.5$).

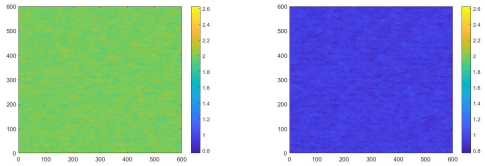
of variability as proven from the figures in the subsequent section.

4.2.1 Entropy fields

The graphical results are now shown and later discussed in this paragraph. The entropy of the random variables Y and x,y velocity components has been computed for each point of all the resolution scales. Figure 4.5 reports the entropy of the Y field for every scale, while Figures 4.6 and 4.7 represent the V_x and V_y fields, respectively, for all scales. We notice that all the variables share one important feature, the decrease of entropy with Upscaling, as expected before. The entropies of the velocity fields gives us a clear picture of the border effects: from the entropy of V_y it can be observed that the prescribed inflow flux (please remember that the horizontal direction correspond to axes y in our plots) on the right side of the grid brings the entropy of this variable to zero in that area. This result is not unexpected: if a variable is constrained by some conditions, its value will be no more random but instead fixed (and easy to predict), this will bring its variability (and so its entropy) to zero. In the same way we observe this phenomenon for V_x on the top and bottom side and on the right side of the grid. Another important observation is that the border effect increases when we Upscale; if we conceptually think of Upscaling as a sort of average, it becomes clear that, moving on coarser scales, this effect influences a growing area.

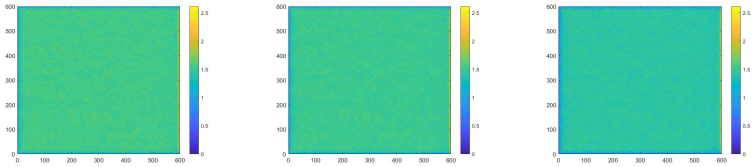


(a) η_0 **(b)** η_2 **(c)** η_4

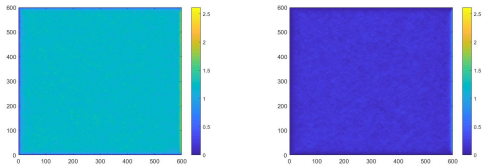


(d) η_8 **(e)** η_{16}

Figure 4.5: Entropy of Y for each scale (case $\sigma_Y^2 = 0.5$).

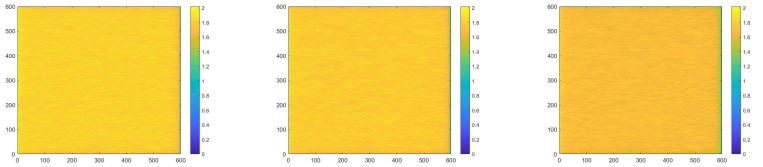


(a) η_0 **(b)** η_2 **(c)** η_4

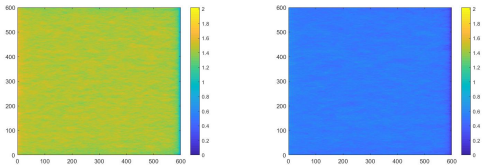


(d) η_8 **(e)** η_{16}

Figure 4.6: Entropy of V_x for each scale (case $\sigma_Y^2 = 0.5$).



(a) η_0 **(b)** η_2 **(c)** η_4



(d) η_8 **(e)** η_{16}

Figure 4.7: Entropy of V_y for each scale (case $\sigma_Y^2 = 0.5$).

Y	η_0	η_2	η_4	η_8	η_{16}
H_m	2.483	2.410	2.277	1.997	0.9590
H_{min}	2.332	2.253	2.117	1.837	0.804
H_{max}	2.643	2.553	2.429	2.129	1.114
% loss	0	2.93	8.30	19.55	61.37

Table 4.1: Entropy of Y for each scale (case $\sigma_Y^2 = 0.5$).

4.2.2 Reshaped fields

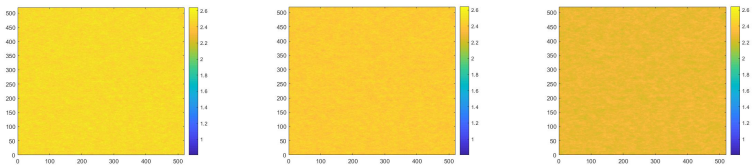
As previously anticipated in Chapter 3, and recalled in previous section a border effect affected our data, and we decided to address this by eliminating a portion of external frame of the domain (5 l_y or 40 $[m]$ per each side). The same calculations done in previous section were repeated, leading to slightly higher entropies as expected. The results are qualitatively the same, despite the change made to correct the border effect; results for V_x and V_y are summed up in Tables 4.1,4.2 and 4.3 (which contain also values referred to Y) and represented in Figures 4.9 and 4.10. Minimum and maximum values of entropy throughout the field are reported, while a unique value is then computed by taking an average over the entire grid and taken as a reference for the scale. A quantification of entropy loss in relative terms (from the reference case η_0) is also proposed for a fast comparation of different scales.

V_x	η_0	η_2	η_4	η_8	η_{16}
H_m	1.581	1.549	1.466	1.251	0.303
H_{min}	1.412	1.373	1.296	1.072	0.163
H_{max}	1.753	1.714	1.633	1.396	0.444
% loss	0	6.01	7.23	20.87	80.85

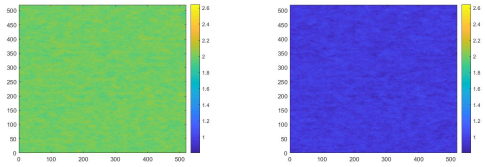
Table 4.2: Entropy of V_x for each scale (case $\sigma_Y^2 = 0.5$).

V_y	η_0	η_2	η_4	η_8	η_{16}
H_m	1.806	1.770	1.686	1.455	0.529
H_{min}	1.622	1.612	1.512	1.253	0.388
H_{max}	1.969	1.928	1.844	1.608	0.661
% loss	0	1.91	6.55	19.38	70.70

Table 4.3: Entropy of V_y for each scale (case $\sigma_Y^2 = 0.5$).

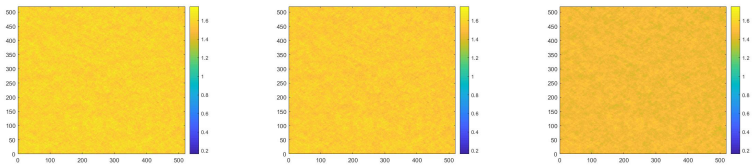


(a) η_0 **(b)** η_2 **(c)** η_4

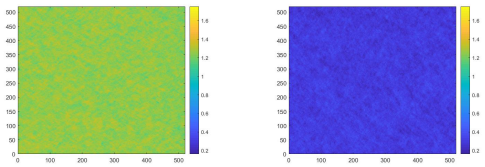


(d) η_8 **(e)** η_{16}

Figure 4.8: Reshaped fields entropy of Y for each scale (case $\sigma_Y^2 = 0.5$).

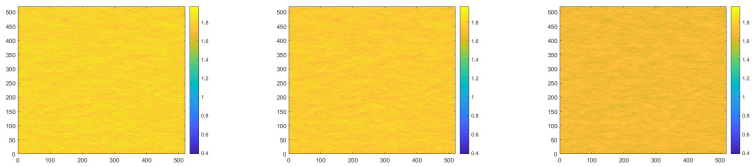


(a) η_0 **(b)** η_2 **(c)** η_4

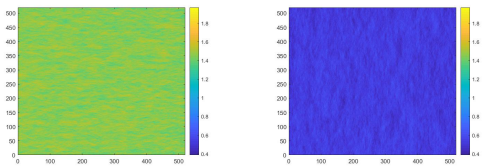


(d) η_8 **(e)** η_{16}

Figure 4.9: Reshaped fields entropy of V_x for each scale (case $\sigma_Y^2 = 0.5$).



(a) η_0 **(b)** η_2 **(c)** η_4



(d) η_8 **(e)** η_{16}

Figure 4.10: Reshaped fields entropy of V_y for each scale (case $\sigma_Y^2 = 0.5$).

It is important to notice that the reduce in entropy, and therefore in uncertainty, does not come from acquiring new information, but rather from an alteration of the nature of the data. This means that we are not more confident about the model, but rather that the coarser scale models contain less information than the finest one. As a result the decrease of entropy can be interpreted as loss of the level of detail contained in the reference scale.

4.3 Information Partitioning

While previous section quantified in terms of entropy the information loss, giving us a magnitude of the loss of quality, we might be interested in understanding how the information evolved from one scale to another. Since the other scales (those different from the finest one) could be seen as somehow derived from the latter, one could argue that the coarser scale fields contain the same kind of information of the finest one, but reduced of a certain percentage (as we computed previously. We are going to see that this is not true, for now it would be sufficient to note that during Upscaling new information is generated, and although it has been derived from the "old" one, it is different from that. Information partitioning metrics, as stated by [**22**] and [**8**], is particularly useful when we want to study how information is shared between scales. Given a target variable (x_{tar}) and two source variables (x_1, x_2), those two variables may provide information to the target variable in many different ways. In our case the source variables will be two coarse-scale fields (from $\eta_{2,4,8,16}$) and the target variable will be the fine-scale field (η_0). This study will be done separately for Y, V_x and V_y and all the possible triplets will be investigated. While in [**8**] the source variables were physically related to the target, the coarse-scale fields are generated with an Upscaling process and consequently they are being generated from it. The innovative part of this study is to apply these metrics to evaluate an Upscaling mechanism; for this reason it has no meaning to think of the shared information as an influence of source variables on target one. Instead we may want to understand if knowing jointly two "low-resolution" fields gives us more information than knowing only the reference one. If the partitioning returns as a result high synergic and unique components it could be more useful to consider two scales rather than one, while a dominant redundant component

Y	η_{0-2-4}	η_{0-2-8}	η_{0-2-16}	η_{0-4-8}	η_{0-4-16}	η_{0-8-16}
$I\left(x_{s1}, x_{s2}; x_{tar}\right)$	1.954	1.959	1.941	1.551	1.516	1.168
$U_{x_{s1}}$	0.432	0.805	1.430	0.381	1.003	0.642
$U_{x_{s2}}$	0.005	0.010	0.003	0.013	0.004	0.011
R	1.501	1.128	0.503	1.125	0.503	0.496
S	0.016	0.016	0.005	0.032	0.006	0.019

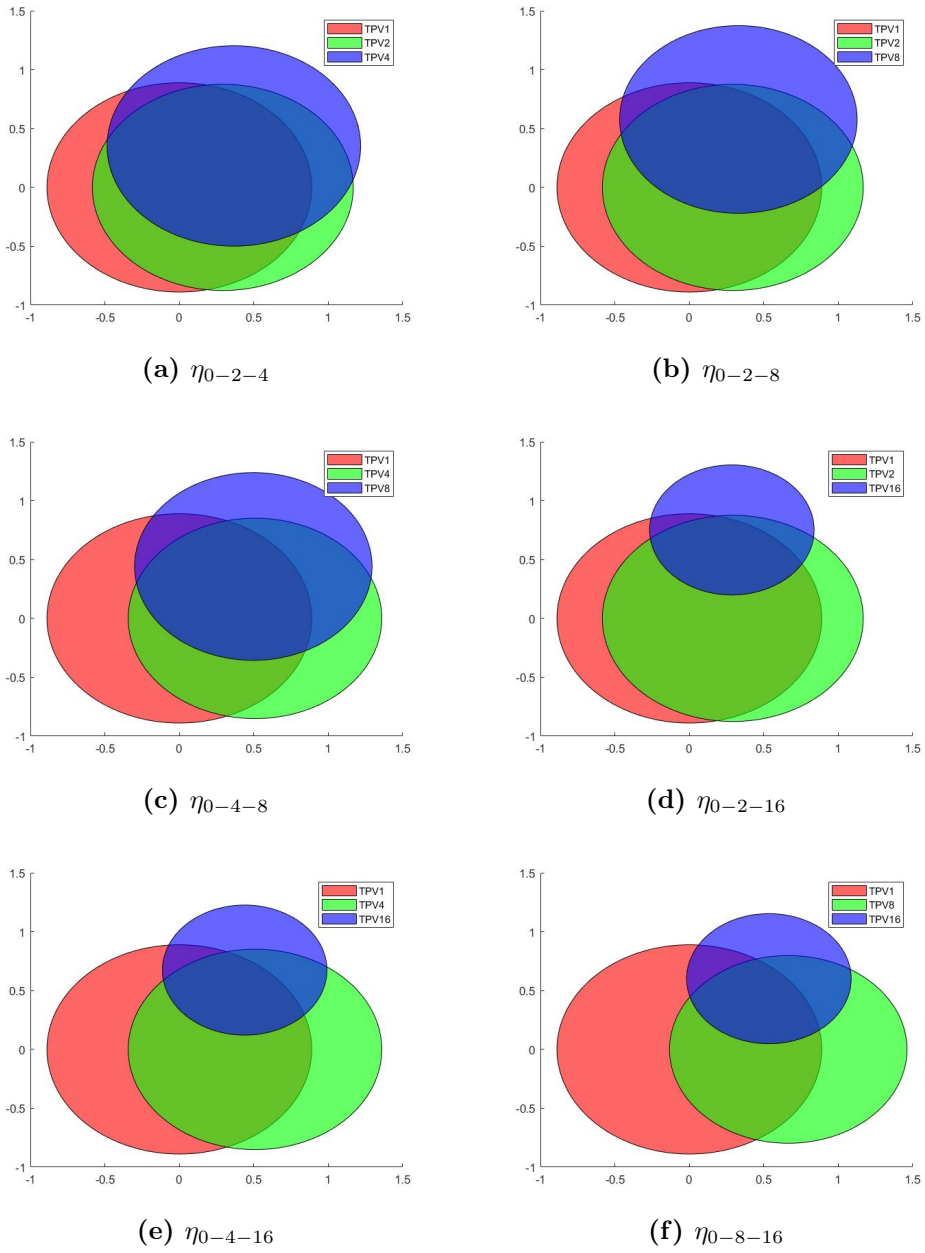
Table 4.4: Trivariate information of Y different triplets (case $\sigma_Y^2 = 0.5$).

would suggest the opposite. Studying the information partitioning of our triplets will help us to evaluate the quality of our coarse-scale models, in terms of ability to represent the reference model.

4.3.1 Information partitioning Y

Triplets η_0-η_2-η_4, η_0-η_2-η_8, η_0-η_2-η_{16}, η_0-η_4-η_8, η_0-η_4-η_{16}, η_0-η_8-η_{16}, η_2-η_4-η_8 and η_4-η_8-η_{16} are now considered. Please note that from now on the finest scale model will be called as x_{tar} and the other two scales completing the triplets will be x_{s1} and x_{s2}. Results are reported in Table 4.4.

We can immediately see that the information shared between sources and target decrease while we Upscale. The multivariate mutual information $I\left(x_{s1}, x_{s2}; x_{tar}\right)$, decreases with the Upscaling extent and this confirms what already said regarding the loss of information during the upscaling process. Before further discussions could be useful to use more immediate graphic tools to represent the partitioning: Figure 4.11 reports Venn diagrams to represent the shared information among different scales, where every circle's area is proportional to the average entropy of the field it represents and the intersection between circles depicts the mutual information between those variables. Red circle represents the target field, while green and blue ones represent, respectively, the more fine and the coarser source fields. Figure 4.12 uses pie diagrams to give an idea of the information partitioning of the information shared between sources and variables by reporting its components for every triplet considered; for these diagrams the partitioning components are normalized with respect to the multivariate mutual information. Only the cases with η_0 as target variable are reported.

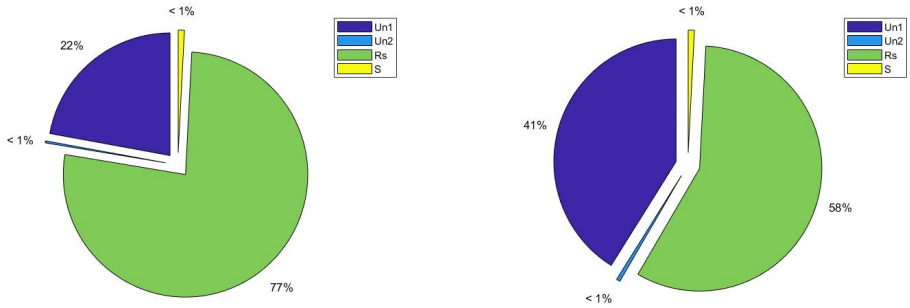


(a) η_{0-2-4} **(b)** η_{0-2-8}

(c) η_{0-4-8} **(d)** η_{0-2-16}

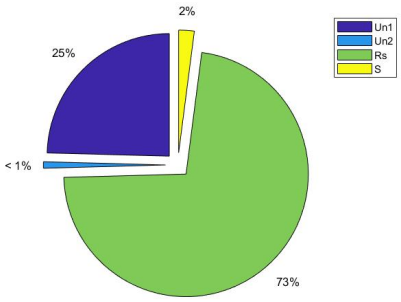
(e) η_{0-4-16} **(f)** η_{0-8-16}

Figure 4.11: Venn diagram representations of entropy and mutual information for variable Y for different triplets (case $\sigma_Y^2 = 0.5$).

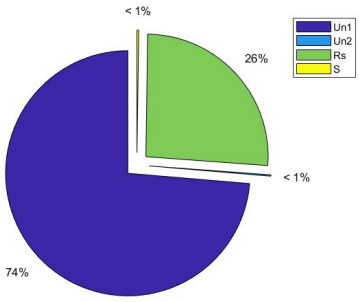


(a) η_{0-2-4}

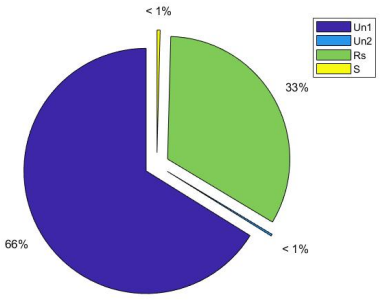
(b) η_{0-2-8}



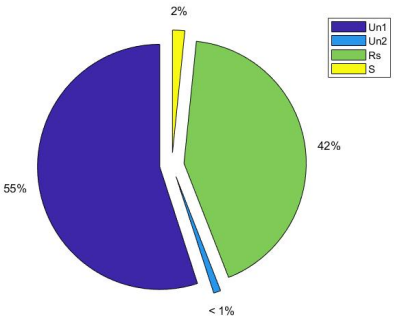
(c) η_{0-4-8}



(d) η_{0-2-16}



(e) η_{0-4-16}



(f) η_{0-8-16}

Figure 4.12: Trivariate information partitioning through pie diagrams for Y for different triplets (case $\sigma_Y^2 = 0.5$).

We note that Venn's diagrams are not made up of concentric circles. As said before, when we Upscale, we are not simply eliminating the extreme values, but we are calculating new values which represent a more homogeneous nature than the heterogeneous one of the initial situation. Therefore, the data will tend to concentrate around the initial average values, but they will also be numerically different from the initial ones. This is clear in the diagrams, where we can see that the information of Upscaled fields is not only that transmitted by the more fine scales (overlapping area of the circles), but there is also some new information, values generated from the reference fields but actually different from them. From pie diagrams we can observe the evolution of the partitioning during Upscaling: the synergic component and the unique component of the Upscaled variables from the reference scale are almost always around 1%, while we note a shifting from redundant component to unique component (of the scale "closer" to reference). When redundant component is high we can say that the two coarse models are similar, this is true for triplets η_0-η_2-η_4, η_0-η_4-η_8 and to a lesser extent for η_0-η_2-η_8. Instead, when a unique component tends to be dominant it means that the other model becomes not so representative of the reference one, as it happens for scale η_{16} in every triplet, including η_0-η_8-η_{16}; this suggests us that scale η_{16} might be too far from the reference scale η_0. Recalling what has been previously stated, during Upscaling the total information contained in a model decrease in absolute value (the circles area decrease while moving on coarser scales), and the models become less representative of the reference one (the intersection area, which represents the shared information between two models, decreases more and more when we Upscale), while still maintaining some characteristics of it. Moreover some new information is generated, this is represented by the circles area of the coarser scales which do not overlap with those of the more fine scales; this new information, although generated from previous models, it is not describe of them.

4.3.2 Information partitioning V_x, V_y

The qualitative behaviour shown for variable Y is confirmed when we analyze the velocity fields, for this reason we propose a summary in Tables 4.5 and 4.6 and we decided to report the Venn and pie diagrams in the Appendix. From all cases we could observe, as an

V_x	η_{0-2-4}	η_{0-2-8}	η_{0-2-16}	η_{0-4-8}	η_{0-4-16}	η_{0-8-16}
$I(x_{s1}, x_{s2}; x_{tar})$	1.306	1.309	1.305	1.035	1.031	0.732
$U_{x_{s1}}$	0.279	0.591	1.168	0.313	0.892	0.588
$U_{x_{s2}}$	0.001	0.002	0.002	0.003	0.004	0.010
R	1.023	0.711	0.134	0.711	0.132	0.126
S	0.004	0.004	0.002	0.003	0.003	0.008

Table 4.5: Trivariate information of V_x for different triplets (case $\sigma_Y^2 = 0.5$).

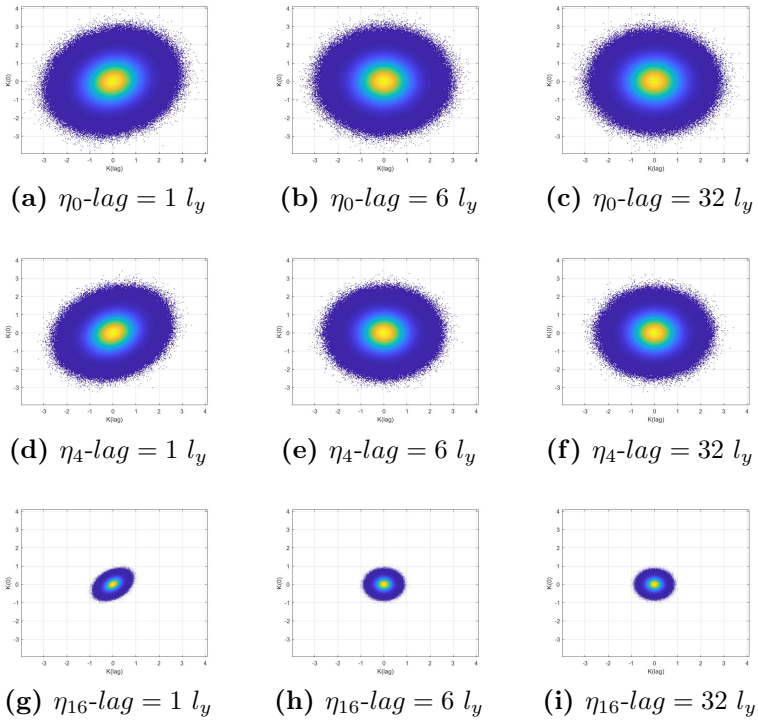
V_y	η_{0-2-4}	η_{0-2-8}	η_{0-2-16}	η_{0-4-8}	η_{0-4-16}	η_{0-8-16}
$I(x_{s1}, x_{s2}; x_{tar})$	1.490	1.493	1.487	1.187	1.178	0.850
$U_{x_{s1}}$	0.312	0.660	1.208	0.349	0.898	0.561
$U_{x_{s2}}$	0.001	0.004	0.001	0.004	0.002	0.010
R	1.171	0.824	0.275	0.823	0.274	0.267
S	0.005	0.006	0.002	0.011	0.003	0.013

Table 4.6: Trivariate information of V_y for different triplets (case $\sigma_Y^2 = 0.5$).

example of possible applications of these metrics, that couples η_{2-4} and η_{2-8} share the same amount of information with target variable η_0; use couple η_{2-8} to characterize η_0 is equivalent to do it by using η_{2-4}.

4.4 Spatial correlation

As a last part of this work we now present a study of the spatial correlation of variables Y, V_x and V_y, as anticipated in Chapter 3, to see if the structure of the fields are preserved trough Upscaling (and if so, how). We have previously cut-off our domain, that now is a square of side 520 $[m]$, so, recalling that the correlation scale of field Y is $l_y = 8$ $[m]$, we have 65 correlation scales in every direction. We consider only 32 correlation scales in the mean flow direction (i.e. y direction). Firstly, scatter plots between pairs of the same variable (i.e.,Y or V_x or V_y) sampled at (i) one location and (ii) a location which is distant of a given lag (along the main flow direction) are presented. Note that we scroll over all the location and we consider 32 $lags$: Figure 4.13 reports scatter plots for variable Y only at scales η_0,η_4,η_{16} and for $lag = 1, 6, 32$ l_y, Figure 4.14 does the same for V_x



(a) η_0-$lag = 1\ l_y$ **(b)** η_0-$lag = 6\ l_y$ **(c)** η_0-$lag = 32\ l_y$

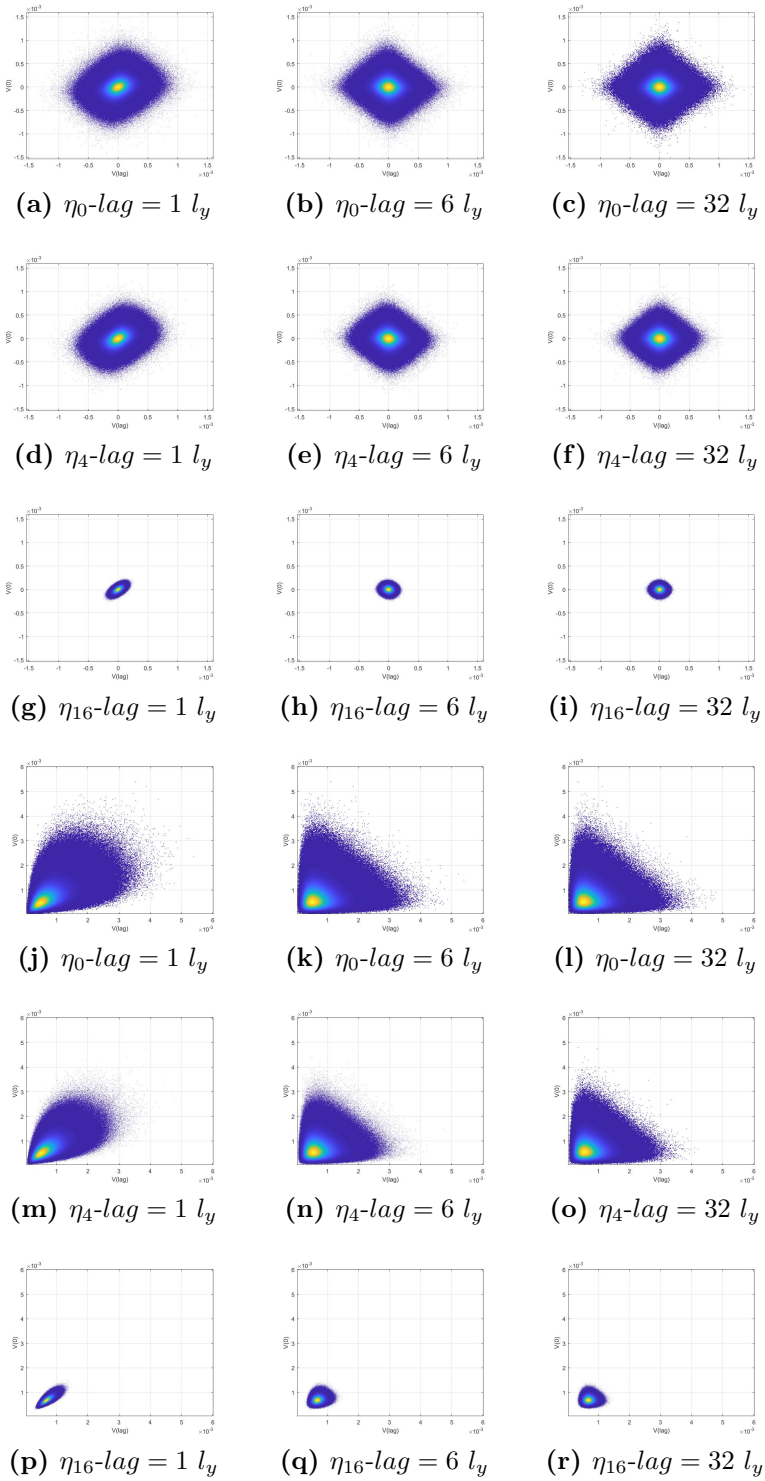
(d) η_4-$lag = 1\ l_y$ **(e)** η_4-$lag = 6\ l_y$ **(f)** η_4-$lag = 32\ l_y$

(g) η_{16}-$lag = 1\ l_y$ **(h)** η_{16}-$lag = 6\ l_y$ **(i)** η_{16}-$lag = 32\ l_y$

Figure 4.13: Scatter plots of Y (case $\sigma_Y^2 = 0.5$).

and V_y. Not all the fields and not all the *lags* have been reported in the plots: we decided to present here only the cases for maximum and minimum *lag* and one case at the lag that maximizes the difference between R and ρ. Moreover, at a $lag = 6\ l_y$ not only is maximized the difference between R and ρ, but the behaviour of these variables tends to be constant. For these reasons, the rest of the scatter plots are presented in the Appendix. If a variable at a given lag has a high linear correlation the pair of values will lay on a straight line with unitary slope, or, in other words, points with high (low) Y are surrounded (at distance equal to *lag*) only by points with high (low) Y. As the non-linear correlation grows this line tends to transform into a figure of uniform or circular shape: given let's say, as input a high vale of Y we will no more find only "high Y" points at a distance equal to *lag* from the input.

Secondly we recall from Chapter 3 coefficients ρ, R and U, for a



(a) η_0-$lag = 1\ l_y$ **(b)** η_0-$lag = 6\ l_y$ **(c)** η_0-$lag = 32\ l_y$

(d) η_4-$lag = 1\ l_y$ **(e)** η_4-$lag = 6\ l_y$ **(f)** η_4-$lag = 32\ l_y$

(g) η_{16}-$lag = 1\ l_y$ **(h)** η_{16}-$lag = 6\ l_y$ **(i)** η_{16}-$lag = 32\ l_y$

(j) η_0-$lag = 1\ l_y$ **(k)** η_0-$lag = 6\ l_y$ **(l)** η_0-$lag = 32\ l_y$

(m) η_4-$lag = 1\ l_y$ **(n)** η_4-$lag = 6\ l_y$ **(o)** η_4-$lag = 32\ l_y$

(p) η_{16}-$lag = 1\ l_y$ **(q)** η_{16}-$lag = 6\ l_y$ **(r)** η_{16}-$lag = 32\ l_y$

Figure 4.14: Scatter plots of V_x ((a)-(i)) and V_y ((j)-(r)) (case $\sigma_Y^2 = 0.5$).

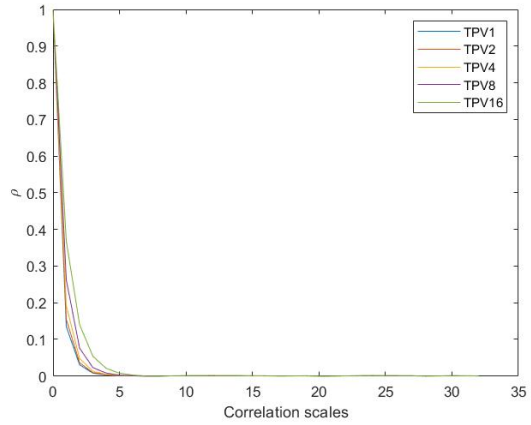
variable, say X as **[10]**,**[11]**,**[7]**:

$$\rho\left(X, X\left(lag\right)\right) = \frac{Cov\left(X\left(x\right), X\left(x+lag\right)\right)}{\sigma_{X(x)}\sigma_{X(x+lag)}} \tag{4.1}$$

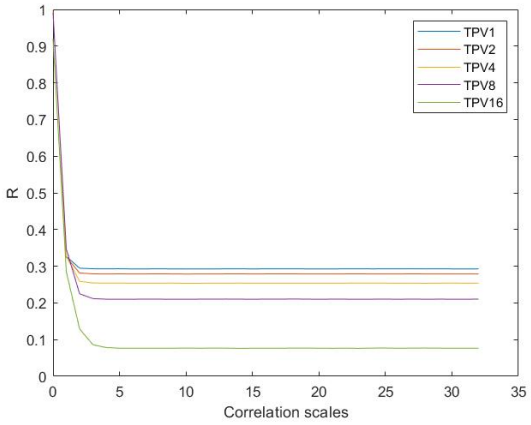
$$R\left(X, X\left(lag\right)\right) = \left\{1 - exp[-2I\left(X\left(x\right), X\left(x+lag\right)\right)]\right\}^{1/2} \tag{4.2}$$

$$U\left(X, X\left(lag\right)\right) = 2\frac{I\left(X\left(x\right), X\left(x+lag\right)\right)}{H\left(X\left(x\right)\right)H\left(X\left(x+lag\right)\right)} \tag{4.3}$$

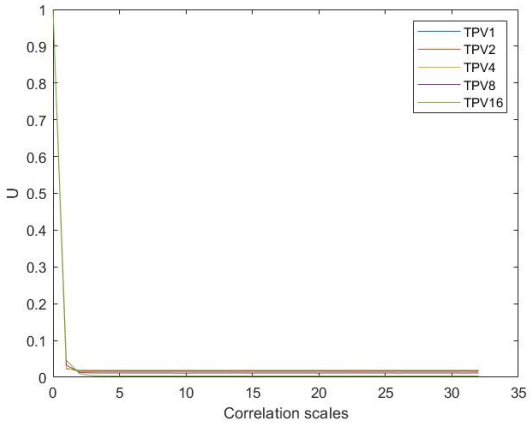
being $Cov\left(X\left(x\right), X\left(x+lag\right)\right)$ the covariance, $\sigma_{X(x)}$ and $\sigma_{X(x+lag)}$ the standard deviations of $X\left(x\right)$ and $X\left(x+lag\right)$, $I\left(X\left(x\right), X\left(x+lag\right)\right)$ the mutual information between two points of a couple and $H\left(X\left(x\right)\right)$ and $H\left(X\left(x+lag\right)\right)$ their entropies. Note that while ρ represents a linear correlation, coefficient R tells if two variables are correlated, linearly or not. It is interesting then to compare the evolution of these two coefficients at increasing lag and Upscaling, to understand if and when a variable becomes linearly or non-linearly correlated. Finally, recalling from **[23]**, we observe that typical values of R and $|\rho|$ of 0.6-0.7 mark and strong association (i.e. linear for rho), while values of 0.2-0.3 marks a weak association (linear or not) between two variables. Figures 4.15, 4.16, 4.17 report, respectively, the behaviour of the coefficients for variables Y, V_x and V_y. Two consideration can be made by observing those plots: firstly Upscaling has the effect of homogenization, as stated also in previous sections, and this could be seen by the concentration of the couples of points in more and more narrow regions moving from more fine scales to coarse ones. The second effect is less intuitive and needs some comment: we note that Upscaling has also the effect of linearization, but this is true only at low $lags$, while it becomes null when distance between points increase. This becomes particularly evident, and true for all the variables, when we analyze the behaviour of ρ: it grows with Upscaling at low $lags$ (less than 4 correlation scales) while it rapidly goes to zero for every scale at high distances. The negative values of ρ for variable V_x can be explained considering that, in some parts of the grid, the speed of the fluid in the x direction inverts its direction, showing then negative values. As expected U fastly goes to zero after few correlation scales, as its numerator becomes low rapidly. This tells us that, in any case, the points at different $lags$ have low relation between each other. About the R coefficient we observe an opposite trend from ρ: it decreases when Upscaling increases, this confirms that we are amplifying linear correlation and decreasing



(a) ρ

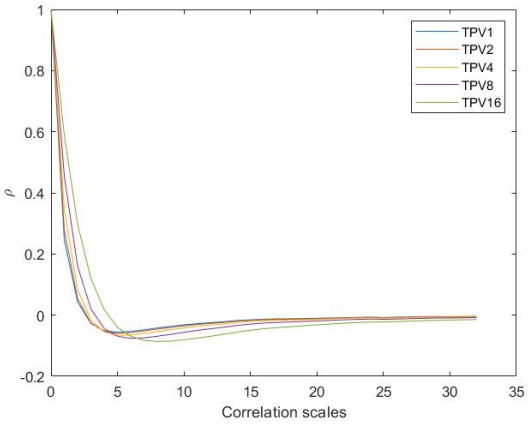


(b) R

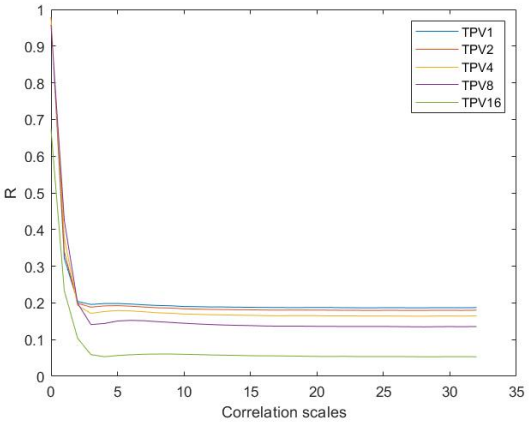


(c) U

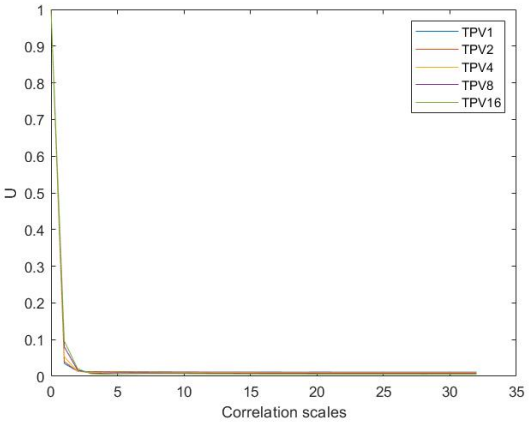
Figure 4.15: ρ,R and U coefficients for variable Y (case $\sigma_Y^2 = 0.5$). Different colors refer to results associated with different Upscaled fields.



(a) ρ

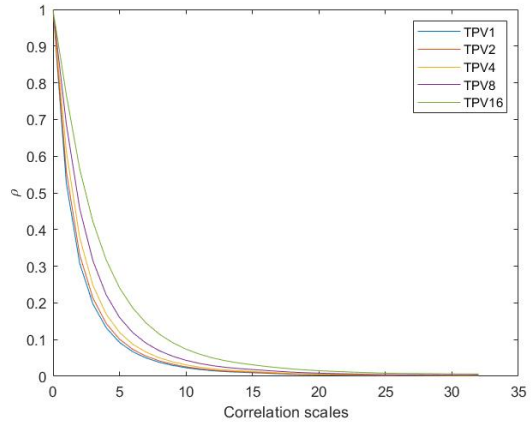


(b) R

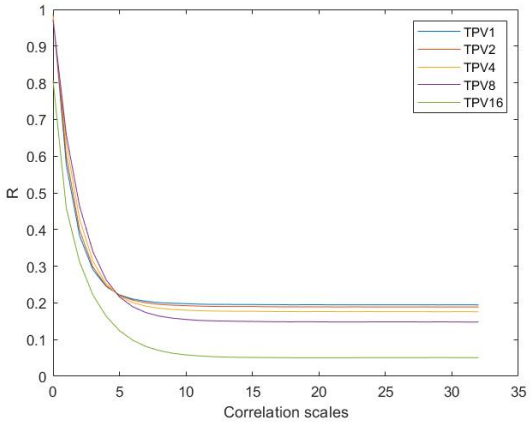


(c) U

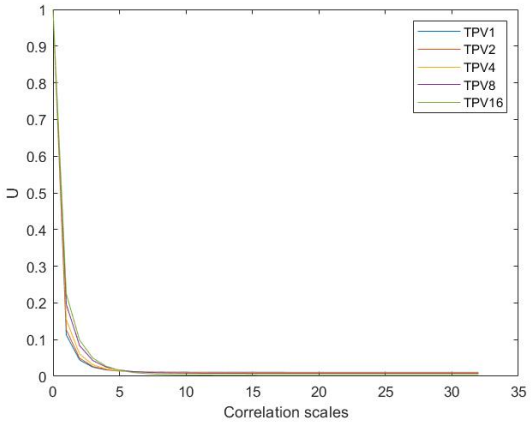
Figure 4.16: ρ,R and U coefficients for variable V_x (case $\sigma_Y^2 = 0.5$). Different colors refer to results associated with different Upscaled fields.



(a) ρ



(b) R



(c) U

Figure 4.17: ρ,R and U coefficients for variable V_y (case $\sigma_Y^2 = 0.5$). Different colors refer to results associated with different Upscaled fields.

non-linear one when we move to coarser scale fields. In some points, at high $lags$, when rho coefficient goes to zero, R is not null; although this means that a non-linear relation is presented at those $lags$, the magnitude of R is not high enough to suggest a strong non-linear component, as said initially.

4.5 Strongly heterogeneous field

All the analyses done previously in this chapter have been repeated for a strongly heterogeneous field. The results obtained do not differ substantially from those of the less heterogeneous field; although slightly different values were obtained, the order of magnitude and the qualitative behaviour of the it metrics analyzed are almost the same. Nevertheless we could see that the entropy of the Y field remained almost the same, even though the variance increased, the borders of the binning became more wide and, as the number of bins did not change, this led to similar results with respect to the weakly heterogeneous case. This is not true for velocity fields, as we could notice an appreciable reduction of entropy for both fields. Moreover, with respect to the $\sigma^2_Y = 0.5$ case, the relative reduction of entropy during Upscaling increased in this case, in particular for variable V_y (the mean direction of the flow): for this case we observe that we lost more information relevant to flow characterization when we Upscaled the field. Moreover a small difference from lower variance case has been observed for the spatial correlation: linear correlation goes faster to zero with the distance (among two points of the couple), this follows from the nature of the field itself. For the complete results please refer to the Appendix.

Chapter 5

Conclusions

This master Thesis aimed at quantifying information contained in a model, investigating loss of information and quality of a model generated by means of Upscaling. A new approach has been proposed coupling Monte Carlo method with IT theory tools to analyze the effects of Upscaling on the Y, V_x and V_y fields. These three variables have been studied for two different cases ($\sigma_Y^2 = 0.5$ and $\sigma_Y^2 = 2$), all of them leading to similar results. The initial part of the work involved generating 1000 Monte Carlo representations for every one of five models, each with different scales; this was done firstly for a low heterogeneity case and then for a more heterogeneous one. Flow problem, with proper border conditions, was then solved for the porous media using a finite element method (FEM) software (i.e. *FreeFem++*). As an output two velocity fields for each representation were generated, one along x direction and one along y, while the mean flow direction has been recognized to be direction y. At the end of this preparation part, we started processing and analyzing the generated data. This is the main part of the work and it could be divided into three parts:

1. Quantification of information at a given resolution scale: as we defined it, entropy, is not a measure of heterogeneity of a field, but rather an indicator of the presence of likely or unlikely to occur data (low and high entropy, respectively) in a specific point of the grid. However when averaged on the entire field, entropy indicates how variable is that field and consequently, the average amount of surprise embedded in it. This latter definition let us use entropy to quantify the amount of information included

in a model, giving us the possibility to study its behaviour throughout the Upscaling process. Both the $\sigma_Y^2 = 0.5$ and the $\sigma_Y^2 = 2$ cases evidenced that entropy, and so the amount of information of a model, decrease with Upscaling. We propose to refer to relative entropies (compared with reference case) as it will be less significative to consider absolute values of entropy for our purposes.

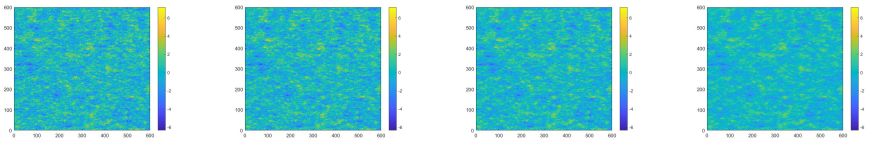
2. Behaviour of information during Upscaling: while entropy let us quantify the amount of information of different scales, we still did not know enough about the evolution of information during the Upscaling process. Given that the magnitude of information reduces when we Upscale, it is not trivial to understand how much do two different models have in common and how much information is transmitted from one to another. By using the proposed IT metrics we could answer to those questions and give some tools to decide when an Upscaled model is too different from the reference one (and from the finer scale ones). This might facilitate managerial decisions for aquifer characterization, as it helps understanding precisely what happens to information during the Upscaling process. As an example of the application of these metrics, it could be observed that couples η_{2-4} and η_{2-8} share the same amount of information with target variable η_0; using couple η_{2-8} to characterize η_0 is equivalent to do it by using η_{2-4}.
3. Variation in the spatial correlation: lastly we investigated the nature of the relations between different point of the same field, for variables Y, V_x and V_y. It has been showed that Upscaling alters slightly the structure of the fields, by increasing the linear correlation, but not in a significant way, as coefficient U denotes still a low spatial correlation between variables.

When we analyzed the strongly heterogeneous fields we could see that the entropy of the Y field remained almost the same with respect to the weakly heterogeneous ones; even though the variance increased, the borders of the binning became more wide and, as the number of bins did not change, this led to similar results with respect to the weakly heterogeneous case. This was not true for velocity fields, as we could notice an appreciable reduction of entropy for both fields. Moreover, with respect to the $\sigma_Y^2 = 0.5$ case, the relative reduction of entropy during Upscaling increased in this case, in particular for variable V_y (the mean direction of the flow): for

this case we observe that we lost more information relevant to flow characterization when we Upscaled the field. Again, other small differences were not significant nor qualitatively different from the previous case. All the results were obtained with a fixed-binning technique in order to get probability functions from discrete variables, choosing to divide all the data in 15 bins as suggested by [**15**]; we assert that this is a critical part of our work, since it does not exist a rigorous binning procedure for this case. We suggest that an improvement to this limitation could be the use of Kernel Density Estimation (KDE), in order to obviate the use of discrete variables. This could be then a starting point for future applications; while future studies could be also based on real data, instead of synthetic ones, as those provided by [**18**]; this would allow to develop an IT-based Upscaling technique starting from real data at different resolution scales. Other possible future studies could analyze the Downscaling process, which is much more obscure by now than Upscaling, or simply use this same approach to study transport and multi-component reactive transport problems instead of flow problems like the present study.

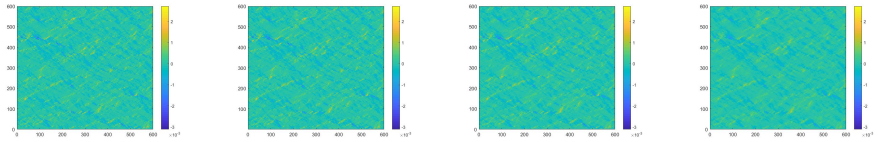
Appendix

Here are presented all the missing results omitted in the results chapter. They are mainly the $\sigma^2_Y = 2$ "counterpart" of the results presented for the $\sigma^2_Y = 0.5$ case. Results for entropies and information partitioning are summed up in tables, while all the fields, entropy grids, Venn and pie diagrams of trivariate information and domain conservation IT metrics are represented in figures, each one labelled to be distinguished. Please note that for the last part scales η_2 and η_8 are showed also for the weakly heterogeneous field as they haven't been presented in the main body of this work. Following all the results, and this concludes this work. For the comments related to this results please refer to Chapter 4.



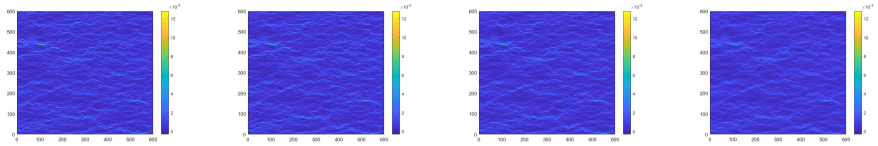
(a) η_0 **(b)** η_2 **(c)** η_4 **(d)** η_8

Figure 5.1: Y fields for each different resolution scale (case $\sigma^2_Y = 2$).



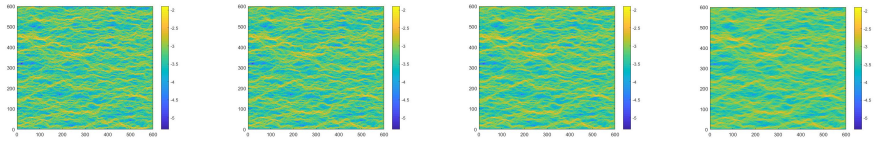
(a) η_0 **(b)** η_2 **(c)** η_4 **(d)** η_8

Figure 5.2: V_x fields for each different resolution scale (case $\sigma_Y^2 = 2$).



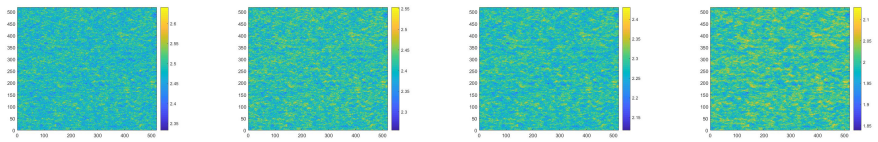
(a) η_0 **(b)** η_2 **(c)** η_4 **(d)** η_8

Figure 5.3: V_y fields for each different resolution scale (case $\sigma_Y^2 = 2$).



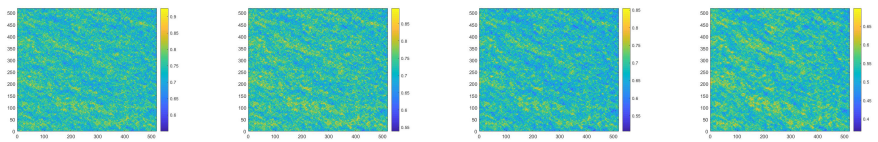
(a) η_0 **(b)** η_2 **(c)** η_4 **(d)** η_8

Figure 5.4: V fields for each different resolution scale (case $\sigma_Y^2 = 2$).



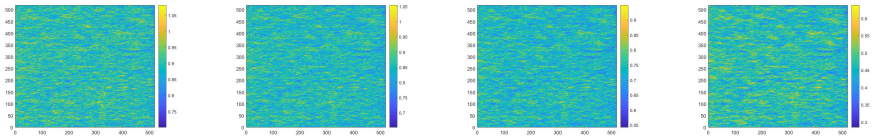
(a) η_0 **(b)** η_2 **(c)** η_4 **(d)** η_8

Figure 5.5: Reshaped fields entropy of Y for each scale (case $\sigma_Y^2 = 2$).



(a) η_0 **(b)** η_2 **(c)** η_4 **(d)** η_8

Figure 5.6: Reshaped fields entropy of V_x for each scale (case $\sigma_Y^2 = 2$).



(a) η_0 **(b)** η_2 **(c)** η_4 **(d)** η_8

Figure 5.7: Reshaped fields entropy of V_y for each scale (case $\sigma_Y^2 = 2$).

Y	η_0	η_2	η_4	η_8
H_m	2.483	2.411	2.277	1.998
H_{min}	2.333	2.255	2.117	1.834
H_{max}	2.641	2.554	2.430	2.129
% loss	0	2.90	8.30	19.53

Table 5.1: Entropy of Y for each scale (case $\sigma_Y^2 = 2$).

V_x	η_0	η_2	η_4	η_8
H_m	0.742	0.725	0.677	0.539
H_{min}	0.550	0.538	0.506	0.367
H_{max}	0.926	0.895	0.855	0.699
% loss	0	2.29	8.76	27.36

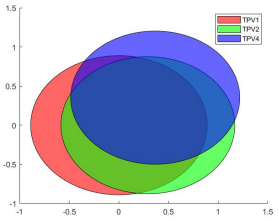
Table 5.2: Entropy of V_x for each scale (case $\sigma_Y^2 = 2$).

V_y	η_0	η_2	η_4	η_8
H_m	0.896	0.850	0.734	0.465
H_{min}	0.700	0.652	0.542	0.285
H_{max}	1.082	1.055	0.950	0.639
% loss	0	5.13	18.08	48.10

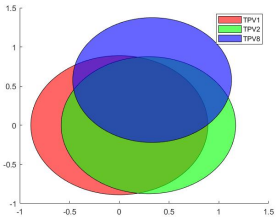
Table 5.3: Entropy of V_y for each scale (case $\sigma_Y^2 = 2$).

Y	η_{0-2-4}	η_{0-2-8}	η_{0-4-8}
$I(x_{s1}, x_{s2}; x_{tar})$	1.955	1.956	1.551
$U_{x_{s1}}$	0.433	0.805	0.381
$U_{x_{s2}}$	0.005	0.010	0.013
R	1.501	1.128	1.125
S	0.016	0.016	0.032

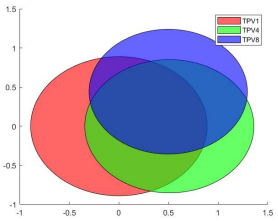
Table 5.4: Trivariate information of Y for each scale (case $\sigma_Y^2 = 2$).



(a) η_{0-2-4}

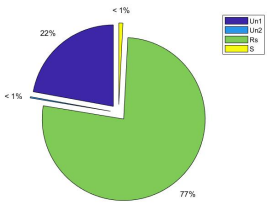


(b) η_{0-2-8}

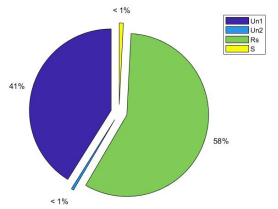


(c) η_{0-4-8}

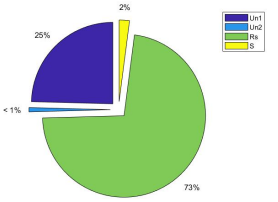
Figure 5.8: Venn diagram representations of entropy and mutual information for Y for different triplets (case $\sigma_Y^2 = 2$).



(a) η_{0-2-4}

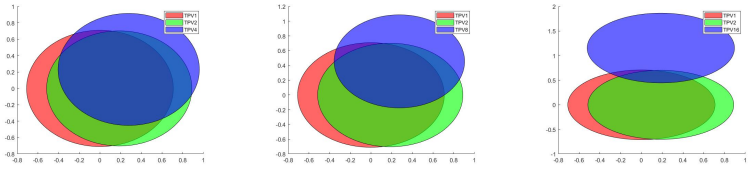


(b) η_{0-2-8}

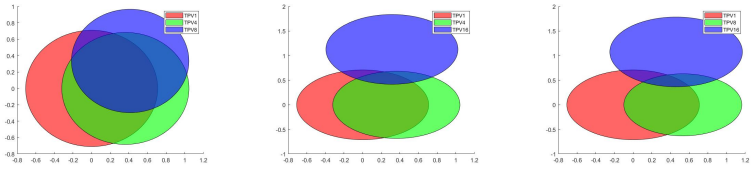


(c) η_{0-4-8}

Figure 5.9: Trivariate information partitioning through pie diagrams for Y for different triplets (case $\sigma_Y^2 = 2$).

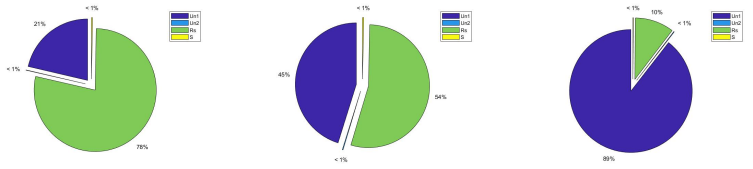


(a) η_{0-2-4} (b) η_{0-2-8} (c) η_{0-2-16}

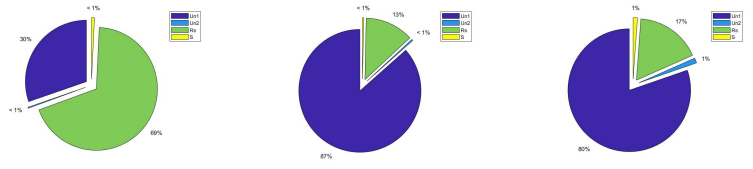


(d) η_{0-4-8} (e) η_{0-4-16} (f) η_{0-8-16}

Figure 5.10: Venn diagram representations of entropy and mutual information for V_x for missing triplets (case $\sigma_Y^2 = 0.5$).

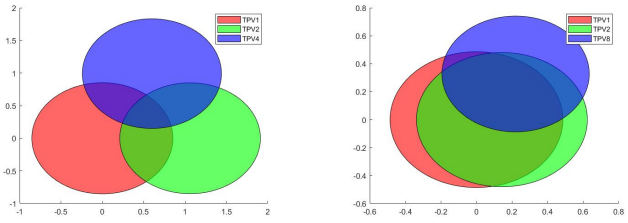


(a) η_{0-2-4} (b) η_{0-2-8} (c) η_{0-2-16}

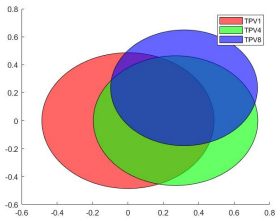


(d) η_{0-4-8} (e) η_{0-4-16} (f) η_{0-8-16}

Figure 5.11: Trivariate information partitioning through pie diagrams for V_x for missing triplets (case $\sigma_Y^2 = 0.5$).

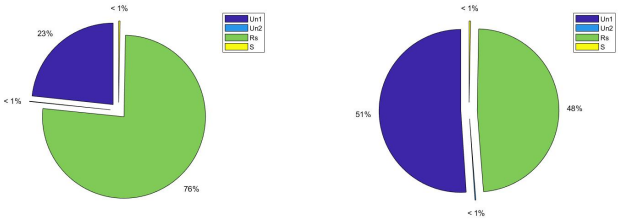


(a) η_{0-2-4} **(b)** η_{0-2-8}

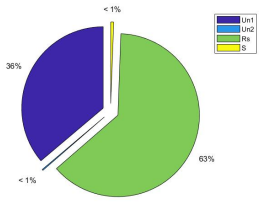


(c) η_{0-4-8}

Figure 5.12: Venn diagram representations of entropy and mutual information for V_x for different triplets (case $\sigma_Y^2 = 2$).

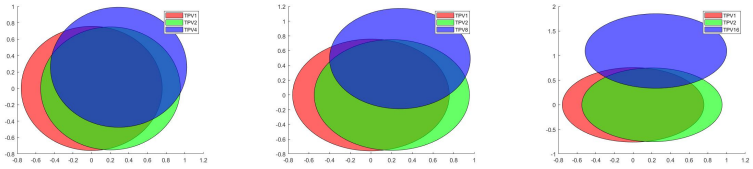


(a) η_{0-2-4} **(b)** η_{0-2-8}

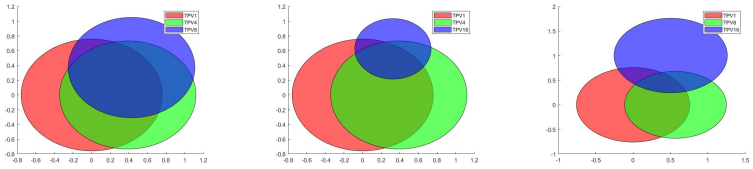


(c) η_{0-4-8}

Figure 5.13: Trivariate information partitioning through pie diagrams for V_x for different triplets (case $\sigma_Y^2 = 2$).

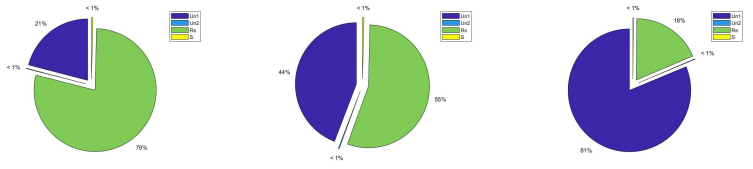


(a) η_{0-2-4} **(b)** η_{0-2-8} **(c)** η_{0-2-16}

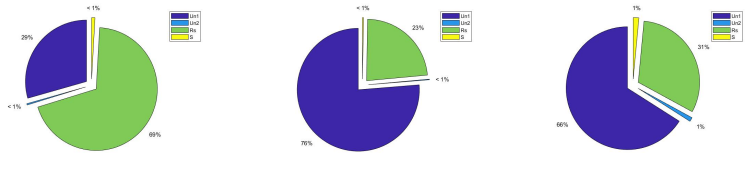


(d) η_{0-4-8} **(e)** η_{0-4-16} **(f)** η_{0-8-16}

Figure 5.14: Venn diagram representations of entropy and mutual information for V_y for missing triplets (case $\sigma_Y^2 = 0.5$).

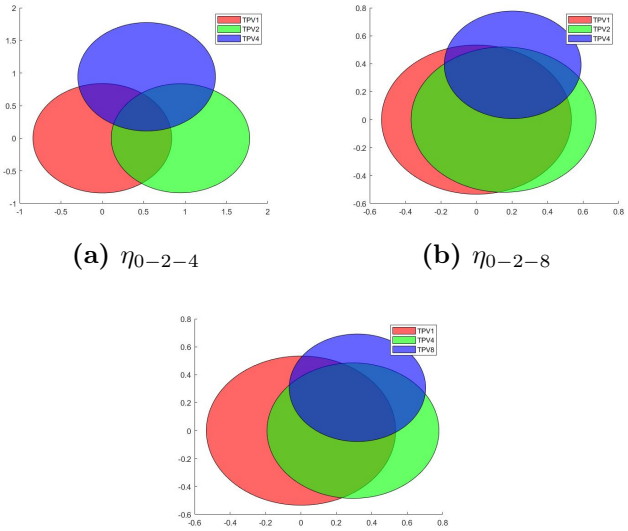


(a) η_{0-2-4} **(b)** η_{0-2-8} **(c)** η_{0-2-16}



(d) η_{0-4-8} **(e)** η_{0-4-16} **(f)** η_{0-8-16}

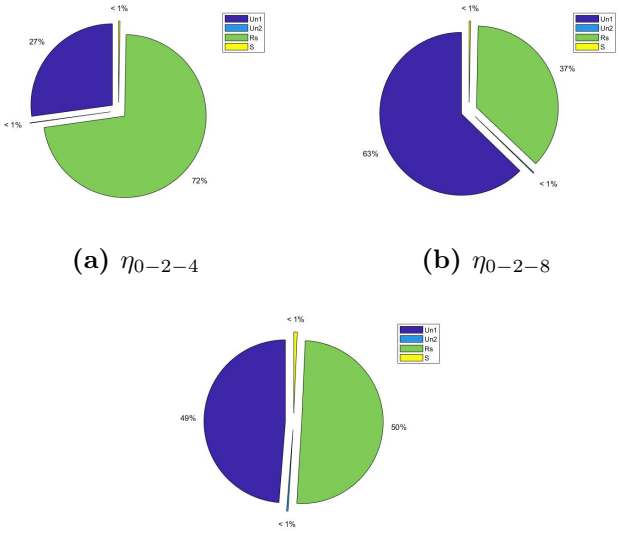
Figure 5.15: Trivariate information partitioning through pie diagrams for V_y for missing triplets (case $\sigma_Y^2 = 0.5$).



(a) η_{0-2-4} **(b)** η_{0-2-8}

(c) η_{0-4-8}

Figure 5.16: Venn diagram representations of entropy and mutual information for V_y for different triplets (case $\sigma_Y^2 = 2$).



(a) η_{0-2-4} **(b)** η_{0-2-8}

(c) η_{0-4-8}

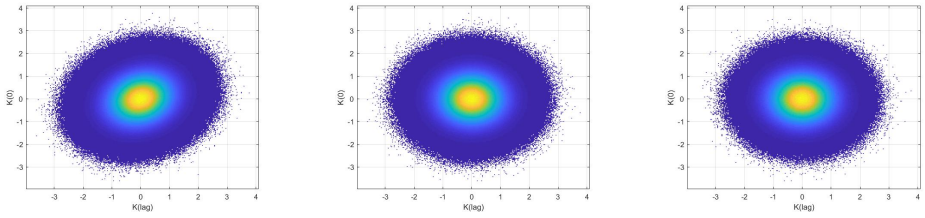
Figure 5.17: Trivariate information partitioning through pie diagrams for V_y for different triplets (case $\sigma_Y^2 = 2$).

V_x	η_{0-2-4}	η_{0-2-8}	η_{0-4-8}
$I\left(x_{s1}, x_{s2}; x_{tar}\right)$	0.598	0.599	0.460
$U_{x_{s1}}$	0.139	0.305	0.167
$U_{x_{s2}}$	$4e-4$	0.001	0.001
R	0.456	0.290	0.290
S	0.002	0.002	0.003

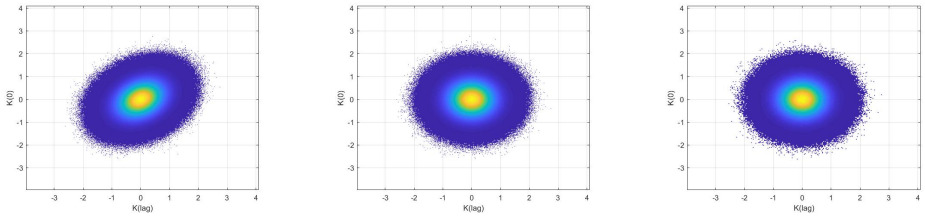
Table 5.5: Trivariate information of V_x for each scale (case $\sigma_Y^2 = 2$).

V_y	η_{0-2-4}	η_{0-2-8}	η_{0-4-8}
$I\left(x_{s1}, x_{s2}; x_{tar}\right)$	0.714	0.716	0.524
$U_{x_{s1}}$	0.194	0.449	0.255
$U_{x_{s2}}$	$5e-4$	0.001	0.002
R	0.518	0.263	0.263
S	0.002	0.002	0.004

Table 5.6: Trivariate information of V_y for each scale (case $\sigma_Y^2 = 2$).

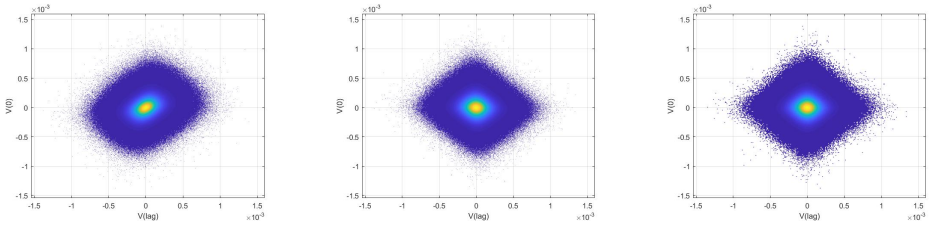


(a) η_2-$lag = 1\ l_y$ **(b)** η_2-$lag = 6\ l_y$ **(c)** η_2-$lag = 32\ l_y$

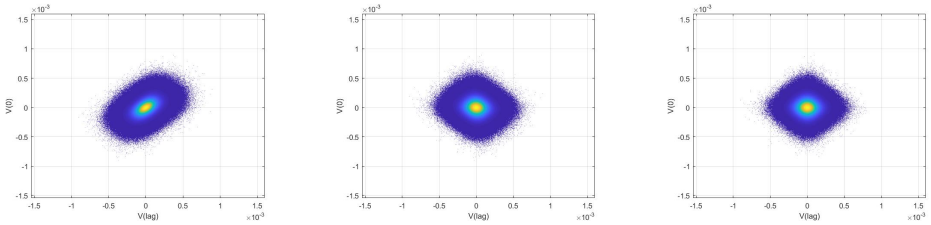


(d) η_8-$lag = 1\ l_y$ **(e)** η_8-$lag = 6\ l_y$ **(f)** η_8-$lag = 32\ l_y$

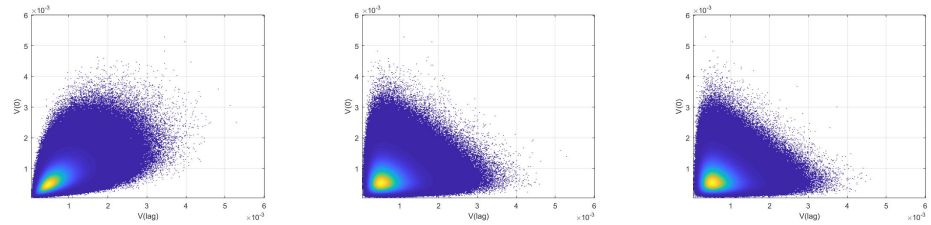
Figure 5.18: Missing scatter plots of Y (case $\sigma_Y^2 = 0.5$).



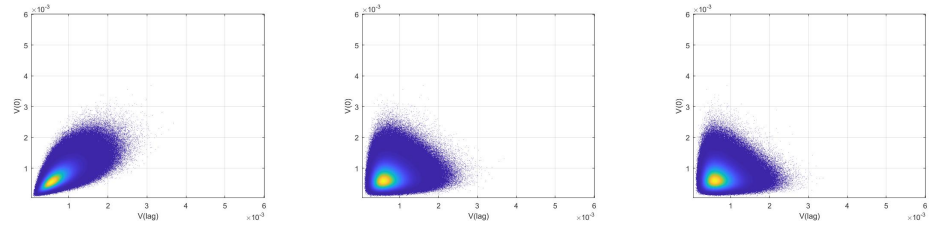
(a) η_2-$lag = 1\ l_y$ **(b)** η_2-$lag = 6\ l_y$ **(c)** η_2-$lag = 32\ l_y$



(d) η_8-$lag = 1\ l_y$ **(e)** η_8-$lag = 6\ l_y$ **(f)** η_8-$lag = 32\ l_y$

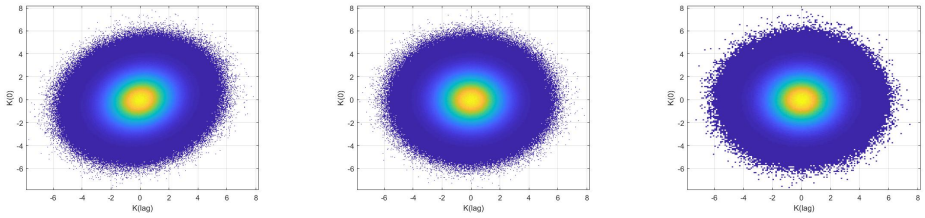


(g) η_2-$lag = 1\ l_y$ **(h)** η_2-$lag = 6\ l_y$ **(i)** η_2-$lag = 32\ l_y$

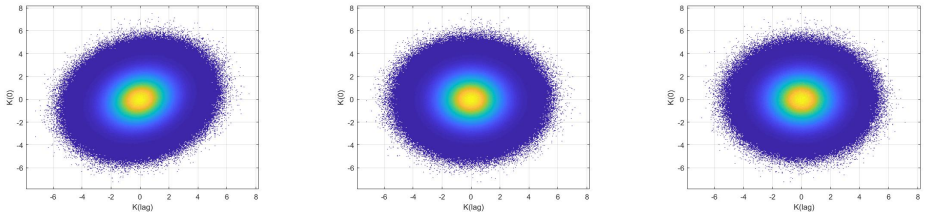


(j) η_8-$lag = 1\ l_y$ **(k)** η_8-$lag = 6\ l_y$ **(l)** η_8-$lag = 32\ l_y$

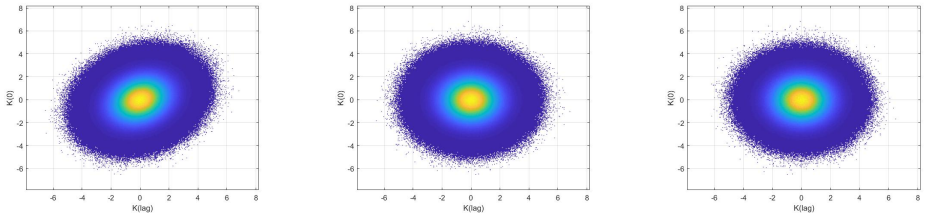
Figure 5.19: Missing scatter plots of V_x ((a)-(f)) and V_y ((g)-(l)) (case $\sigma_Y^2 = 0.5$).



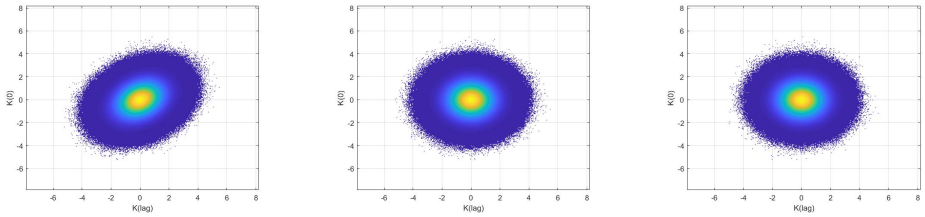
(a) η_0-$lag = 1\ l_y$ **(b)** η_0-$lag = 6\ l_y$ **(c)** η_0-$lag = 32\ l_y$



(d) η_2-$lag = 1\ l_y$ **(e)** η_2-$lag = 6\ l_y$ **(f)** η_2-$lag = 32\ l_y$

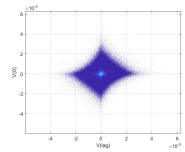


(g) η_4-$lag = 1\ l_y$ **(h)** η_4-$lag = 6\ l_y$ **(i)** η_4-$lag = 32\ l_y$

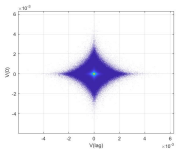


(j) η_8-$lag = 1\ l_y$ **(k)** η_8-$lag = 6\ l_y$ **(l)** η_8-$lag = 32\ l_y$

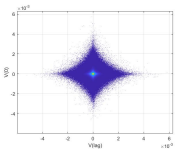
Figure 5.20: Scatter plots of Y (case $\sigma_Y^2 = 2$).



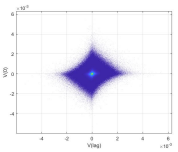
(a) η_0-$lag = 1\ l_y$



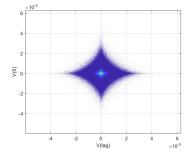
(b) η_0-$lag = 6\ l_y$



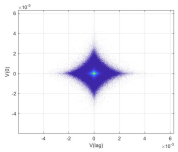
(c) η_0-$lag = 32\ l_y$



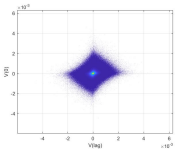
(d) η_2-$lag = 1\ l_y$



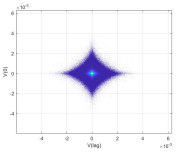
(e) η_2-$lag = 6\ l_y$



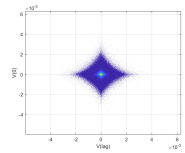
(f) η_2-$lag = 32\ l_y$



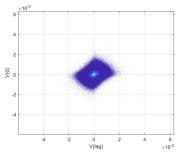
(g) η_4-$lag = 1\ l_y$



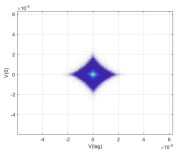
(h) η_4-$lag = 6\ l_y$



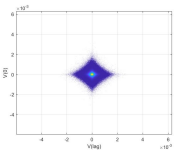
(i) η_4-$lag = 32\ l_y$



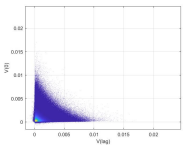
(j) η_8-$lag = 1\ l_y$



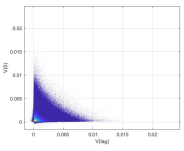
(k) η_8-$lag = 6\ l_y$



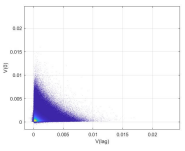
(l) η_8-$lag = 32\ l_y$



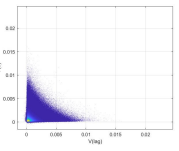
(m) η_0-$lag = 1\ l_y$



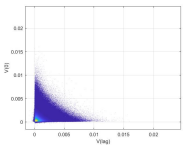
(n) η_0-$lag = 6\ l_y$



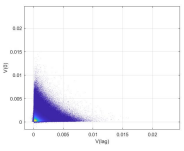
(o) η_0-$lag = 32\ l_y$



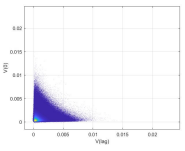
(p) η_2-$lag = 1\ l_y$



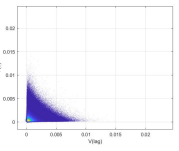
(q) η_2-$lag = 6\ l_y$



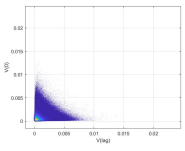
(r) η_2-$lag = 32\ l_y$



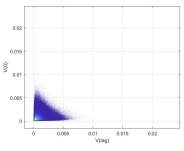
(s) η_4-$lag = 1\ l_y$



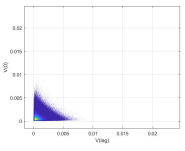
(t) η_4-$lag = 6\ l_y$



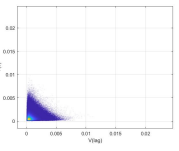
(u) η_4-$lag = 32\ l_y$



(v) η_8-$lag = 1\ l_y$

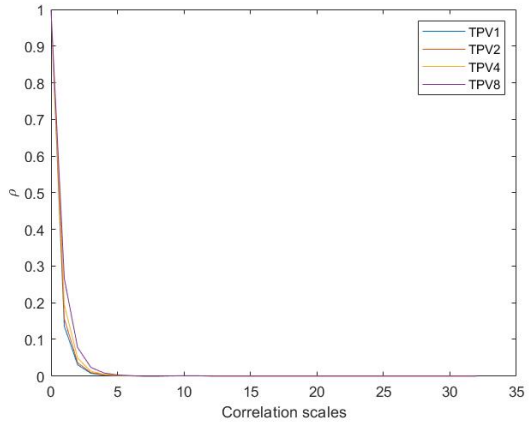


(w) η_8-$lag = 6\ l_y$

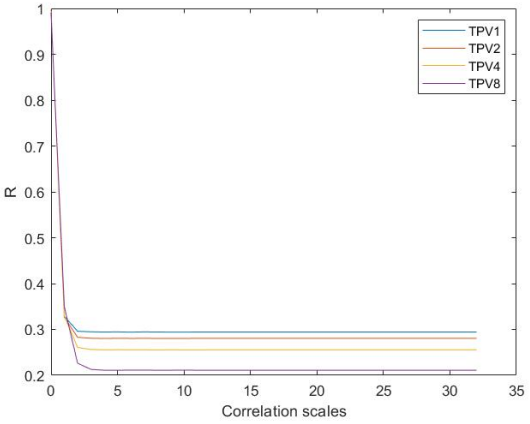


(x) η_8-$lag = 32\ l_y$

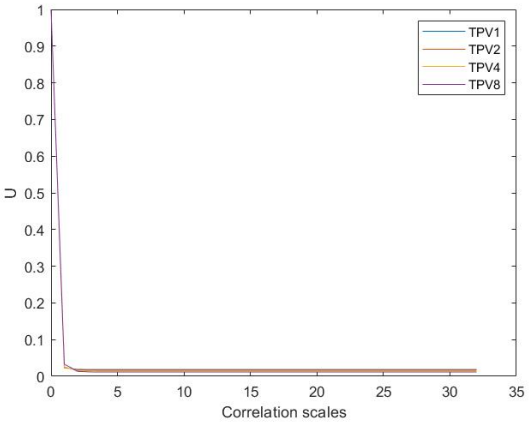
Figure 5.21: Scatter plots of V_x ((a)-(l)) and V_y ((m)-(x)) (case $\sigma_Y^2 = 2$).



(a) ρ

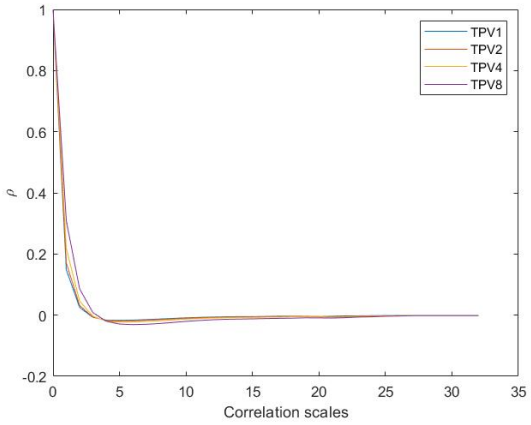


(b) R

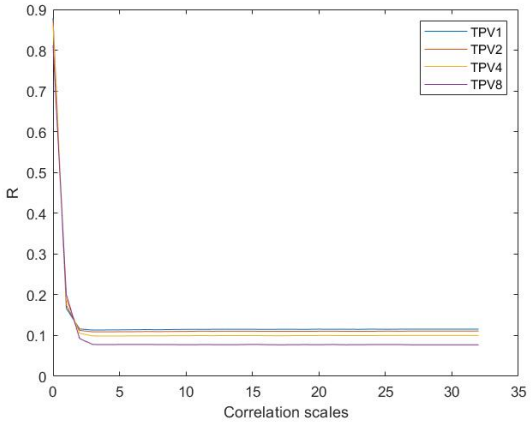


(c) U

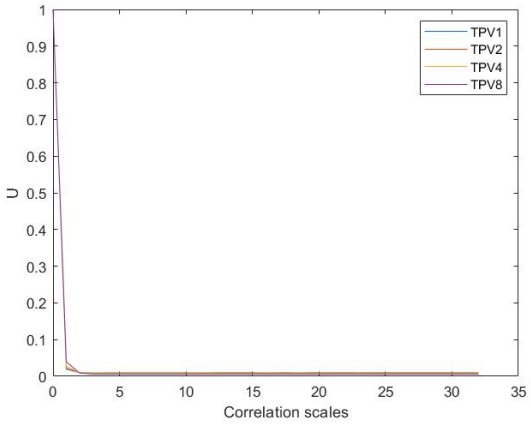
Figure 5.22: ρ,R and U coefficients for variable Y (case $\sigma_Y^2 = 2$). Different colors refer to results associated with different Upscaled fields.



(a) ρ

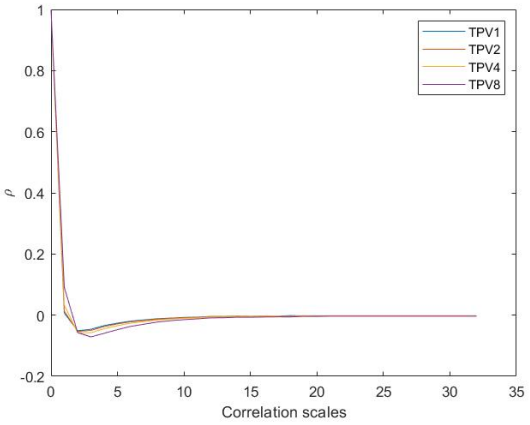


(b) R

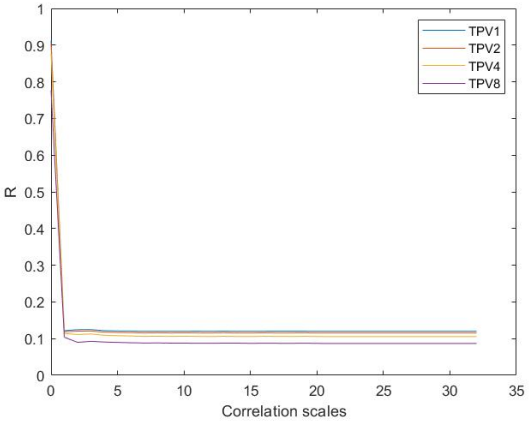


(c) U

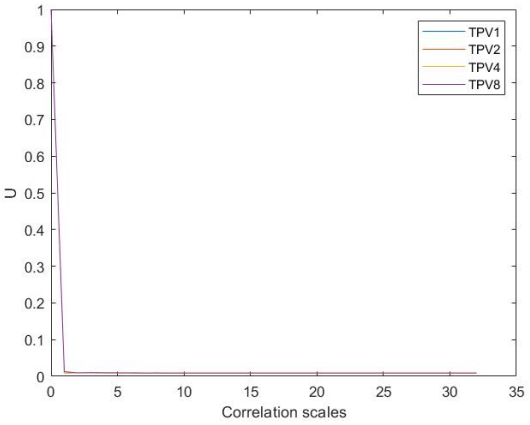
Figure 5.23: ρ,R and U coefficients for variable V_x (case $\sigma_Y^2 = 2$). Different colors refer to results associated with different Upscaled fields.



(a) ρ



(b) R



(c) U

Figure 5.24: ρ,R and U coefficients for variable V_y (case $\sigma_Y^2 = 2$). Different colors refer to results associated with different Upscaled fields.

Bibliography

[1] Ji C. Wu, Xian K. Zeng. *Review of the uncertainty analysis of groundwater simulation.* Chinese Science Bulletin, 2013.

[2] Joseph H. A. Guillaume, M. Ejaz Qureshi, Anthony J. Jakeman. *A structured analysis of uncertainty surrounding modeled impacts of groundwater-extraction rules.* Hydrogeology Journal, 2012.

[3] M.F.P. Bierkens, Jaco van der Gaast. *Upscaling hydraulic conductivity: Theory and examples from geohydrological studies.* Nutrient Cycling in Agroecosystems, 1998.

[4] C.E. Shannon. *A Mathematical Theory of Communication.* The Bell System Technical Journal, 1948.

[5] M. Zen, S. Candiago, U. Schirpke, L.E. Vigl. *Upscaling ecosystem service maps to administrative levels: Beyond scale mismatches.* Science of The Total Environment, 2019.

[6] Francesca Boso, Daniel M. Tartakovsky. *Information-Theoretic Approach to Bidirectional Scaling.* Water Resources Research, 2018.

[7] Ilaria Butera, Luca Vallivero, Luca Ridolfi. *Mutual information analysis to approach nonlinearity in groundwater stochastic fields.* Stochastic Environmental Research and Risk Assessment, 2018.

[8] Allison E. Goodwell, Praveen Kumar. *Temporal information partitioning: Characterizing synergy, uniqueness, and redundancy in interacting environmental variables.* Water Resources Research, 2017.

[9] Allison E. Goodwell, Praveen Kumar, Aaron W. Fellows, Gerald N. Flerchinger. *Dynamic process connectivity explains ecohydrologic responses to rainfall pulses and drought.* PNAS, 2018.

[10] H. Theil. *Statistical Decomposition Analysis.* North-Holland Publishing Co, 1972.

[11] Clive Granger, Jin-Lung Lin. *Using the mutual information coefficient to identify lags in nonlinear models*. Journal of Time Series Analysis, 1994.

[12] D.W. Scott. *On Optimal and Data-Based Histograms*. Biometrika, 1979.

[13] B.W. Silverman. *TDensity estimation for statistics and data analysis*. Monographs on Statistics and Applied Probability, 1986.

[14] Sri Purwani, Julita Nahar, Carole Twining. *Analyzing bin-width effect on the computed entropy*. AIP Conference Proceedings, 2017.

[15] Benjamin L. Ruddell, Praveen Kumar. *Ecohydrologic process networks: 1. Identification*. Water Resources Research, 2009.

[16] Vittorio Di Federico, Shlomo P. Neuman. *Scaling of random fields by means of truncated power variograms and associated spectra*. Surface Water and Climate, 1997.

[17] Martin J. Blunt. *Reservoir Engineering*. Notes for students, 2010.

[18] Vincent C. Tidwell, John L. Wilson. *Upscaling experiments conducted on a block of volcanic tuff: Results for a bimodal permeability distribution*. Water Resources Research, 1999.

[19] James V. Stone. *Information Theory: A Tutorial Introduction*. Sebtel Press, 2015.

[20] Dasheng Qi, Tim Hesketh. *An Analysis of Upscaling Techniques for Reservoir Simulation*. Petroleum Science and Technology, 2007.

[21] J Bear, Y Bachmat. *Macroscopic modelling of transport phenomena in porous media. 2: Applications to mass, momentum and energy transport*. Transport in Porous Media, 1986.

[22] P.L. Williams, R.D. Beer. *Nonnegative Decomposition of Multivariate Information*. arXiv, 2010.

[23] Srikanta Mishra, Neil Deeds, Greg Ruskauff. *Global Sensitivity Analysis Techniques for Probabilistic Ground Water Modeling*. Groundwater, 2009.