Paula Ceccon Ribeiro

Uncertainty Analysis of 2D Vector Fields through the Helmholtz-Hodge Decomposition

Tese de Doutorado

Thesis presented to the Programa de Pós-Graduação em Informática of the Departamento de Informática, PUC-Rio, as partial fulfillment of the requirements for the degree of Doutor em Ciências – Informática.

Advisor: Prof. Hélio Côrtes Vieira Lopes

Rio de Janeiro
December 2016

Paula Ceccon Ribeiro

Uncertainty Analysis of 2D Vector Fields through the Helmholtz-Hodge Decomposition

Thesis presented to the Programa de Pós-Graduação em Informática, of the Departamento de Informática do Centro Técnico Científico da PUC-Rio, as partial fulfillment of the requirements for the degree of Doutor.

Prof. Hélio Côrtes Vieira Lopes
Advisor
Departamento de Informática — PUC-Rio

Prof. Haroldo Fraga de Campos Velho
Instituto Nacional de Pesquisas Espaciais — INPE

Prof. Marcos de Oliveira Lage Ferreira
Instituto de Computação — UFF

Prof. Alex Laier Bordignon
Departamento de Geometria — UFF

Prof. Rubens Sampaio Filho
Departamento de Engenharia Mecânica — PUC-Rio

Prof. Roberta de Queiroz Lima
Departamento de Engenharia Mecânica — PUC-Rio

Prof. Sinesio Pesco
Departamento de Matemática — PUC-Rio

Prof. Waldemar Celes Filho
Departamento de Informática — PUC-Rio

Prof. Márcio da Silveira Carvalho
Coordinator of the Centro Técnico Científico da PUC-Rio

Rio de Janeiro, December 15th, 2016

Paula Ceccon Ribeiro

Bachelor's in Computer Science at the Federal Fluminense University (2011). Masters' in Informatics at the Pontifical Catholic University of Rio de Janeiro (2013), with emphasis in Game Development and Computer Graphics.

Bibliographic data

CDD: 004

Acknowledgments

First and foremost, I would like to thank my parents, Maria Elvira Ceccon Ribeiro and José de Almeida Ribeiro, and sister, Marcela Ceccon Ribeiro, for all the support that they gave me during this journey. They believed on me and have done the possible and impossible so I could achieve my goals.

I would like to thank my advisor, Hélio Côrtes Vieira Lopes, for all opportunities and patience. Under his supervision, I was able to explore new research areas and to spend an year at the New York University. I have learned more things than I though I was able to in short amount of time.

I am thankful to my friends, with whom I have shared so many thoughts. They were always there for me and we learned a lot together.

I would like to thank CNPq, FINEP and MICROSOFT for the financial support of this research. Additionally, I would like to thank INPE, Alex Almeida Fernandes and Haroldo Fraga de Campos Velho for the wind forecast data models used in this work.

At last but not least, I am glad that I have someone up there Who is always looking for me. I wouldn't have be able to go so far without His blessing.

Abstract

Ceccon, Paula; Lopes, Hélio (advisor). **Uncertainty Analysis of 2D Vector Fields through the Helmholtz-Hodge Decomposition**. Rio de Janeiro, 2016. 109p. D.Sc. Thesis — Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Vector field plays an essential role in a large range of scientific applications. They are commonly generated through computer simulations. Such simulations may be a costly process because they usually require high computational time. When researchers want to quantify the uncertainty in such kind of applications, usually an ensemble of vector fields realizations are generated, making the process much more expensive. The Helmholtz-Hodge Decomposition is a very useful instrument for vector field interpretation because it traditionally distinguishes conservative (rotational-free) components from mass-preserving (divergence-free) components. In this work, we are going to explore the applicability of such technique on the uncertainty analysis of 2-dimensional vector fields. First, we will present an approach of the use of the Helmholtz-Hodge Decomposition as a basic tool for the analysis of a vector field ensemble. Given a vector field ensemble $\mathcal{E}$, we firstly obtain the corresponding rotational-free, divergence-free and harmonic component ensembles by applying the Natural Helmholtz-Hodge Decomposition to each vector field in $\mathcal{E}$. With these ensembles in hand, our proposal not only quantifies, via a statistical analysis, how much each component ensemble is point-wisely correlated to the original vector field ensemble, but it also allows to investigate the uncertainty of rotational-free, divergence-free and harmonic components separately. Then, we propose two techniques that jointly with the Helmholtz-Hodge Decomposition stochastically generate vector fields from a single realization. Finally, we propose a method to synthesize vector fields from an ensemble, using both the Dimension Reduction and Inverse Projection techniques. We test the proposed methods with synthetic vector fields as well as with simulated vector fields.

Keywords

Helmholtz-Hodge Decomposition; Vector-Fields; Uncertainty Quantification; Stochastic Simulation; Vector Field Synthesis.

Resumo

Ceccon, Paula; Lopes, Hélio. **Análise de Incertezas em Campos Vetoriais 2D com o uso da Decomposição de Helmholtz-Hodge**. Rio de Janeiro, 2016. 109p. Tese de Doutorado — Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Campos vetoriais representam um papel principal em diversas aplicações científicas. Eles são comumente gerados via simulações computacionais. Essas simulações podem ser um processo custoso, dado que em muitas vezes elas requerem alto tempo computacional. Quando pesquisadores desejam quantificar a incerteza relacionada a esse tipo de aplicação, costuma-se gerar um conjunto de realizações de campos vetoriais, o que torna o processo ainda mais custoso. A Decomposição de Helmholtz-Hodge é uma ferramenta útil para a interpretação de campos vetoriais uma vez que ela distingue componentes conservativos (livre de rotação) de componentes que preservam massa (livre de divergente). No presente trabalho, vamos explorar a aplicabilidade de tal técnica na análise de incerteza de campos vetoriais 2D. Primeiramente, apresentaremos uma abordagem utilizando a Decomposição de Helmholtz-Hodge como uma ferramenta básica na análise de conjuntos de campos vetoriais. Dado um conjunto de campos vetoriais $\mathcal{E}$, obtemos os conjuntos formados pelos componentes livre de rotação, livre de divergente e harmônico, aplicando a Decomposição Natural de Helmholtz-Hodge em cada campo vetorial em $\mathcal{E}$. Com esses conjuntos em mãos, nossa proposta não somente quantifica, por meio de análise estatística, como cada componente é pontualmente correlacionado ao conjunto de campos vetoriais original, como também permite a investigação independente da incerteza relacionado aos campos livre de rotação, livre de divergente e harmônico. Em sequência, propomos duas técnicas que em conjunto com a Decomposição de Helmholtz-Hodge geram, de forma estocástica, campos vetoriais a partir de uma única realização. Por fim, propomos também um método para sintetizar campos vetoriais a partir de um conjunto, utilizando técnicas de Redução de Dimensionalidade e Projeção Inversa. Testamos os métodos propostos tanto em campos sintéticos quanto em campos numericamente simulados.

Palavras-chave

Decomposição de Helmholtz-Hodge; Campos Vetoriais; Quantificação de Incertezas; Simulação Estocástica; Síntese de Campos Vetoriais.

Contents

List of Figures

List of Tables

Where there's a will, there's a way.

English Proverb

1
Introduction

Many authors consider *uncertainty* as a metadata representing the lack of knowledge about a model (Viard et al., 2011). Such metadata includes several unknowns (Potter et al., 2013), as for example: errors, deviations, missing information, or confidence levels. The objective of *uncertainty analysis* is to quantify the uncertainties of the relevant variables of a model. Aligned with *visualization* techniques, this kind of analysis can provide valuable assistance for *decision-making* tasks. According to Ware (2013), visualization is important to decision making because it provides the ability to comprehend large amounts of data; it allows us to perceive emergent properties that were not anticipated; it enables problems with the data to become immediately apparent; and it facilitates not only an understanding of the features of the data in different scales but also the formation of a hypothesis.

Uncertainty Visualization studies methods to encode uncertainty information in combination with the primary data into different graphics primitives (which includes color, glyph and texture) so that the visual perception is not overloaded. This overloading problem is a big challenge for visualization researchers because there is a limited number of visual channels (such as spatial position, color, texture, and opacity) that can be used. As Potter et al. (2012) highlight, to move from quantified uncertainty to visualized uncertainty we often need to simplify the uncertainty data to make it fit into the available visual representation. That challenge promotes Uncertainty Visualization as one of the top research problems in visualization research (Mihai & Westermann, 2014; Johnson & Sanderson, 2003).

Vector fields, on the other hand, are ubiquitous in a large range of applications in *Scientific Computing* (Laidlaw et al., 2005; Forsberg et al., 2009), encompassing different problems such as fluid flow simulation, weather forecasting, the aerodynamic design of cars and aircrafts, and fingerprint matching. A very useful mathematical tool to distinguish conservative (rotational-free) components from mass-preserving (divergence-free) components is the Helmholtz-Hodge Decomposition (HHD) (Chorin, 1968), which defines a vector field as the sum of three L^2-orthogonal components: a rotational-free term, a divergence-free term and a harmonic term. This suitable decomposition al-

lows us to simplify the analysis of complex flows by studying the divergence-free and rotational-free related properties separately (Pascucci et al., 2014). In the 2D case, the rotational-free and divergence-free solutions are given in terms of scalar potentials, which are obtained by solving Poisson equations.

In Scientific Computing, most of the vector field data come from simulations, which may introduce uncertainty from different sources (Laccarino, 2009). To overcome this situation, multiple instances are often generated. This approach helps to predict and quantify the range of outcomes, allowing us to classify features according to their stability across instances because their locations are affected by uncertainty (Mihai & Westermann, 2014).

1.1
Motivation

In one hand, it is recognized that the Helmholtz-Hodge Decomposition simplifies the analysis of vector fields, once some important properties can be studied directly on the components (Bhatia et al., 2013). Indeed, in atmospheric dynamics this decomposition has a relevant role in meteorology (Holton & Hakim, 2012). For example, the higher energy air flows are described by wind rotational (divergence-free) component (Chen & Tribbia, 1981). Moreover, the simulated kinetic energy (related to the squared norm of the velocity vector) is evaluated through the analysis of the rotational and divergent wind components (Blažica et al., 2013). There are other applications, beyond wind dynamics, in which the study of the decomposed components matters, such as the analysis of turbulent flows by the use of Particle Image Velocimetry (PIV) (de Silva et al., 2013) and the study of complex flows using Particle Tracking Velocimetry (PTV) (Sadati et al., 2011), which shows to be very important for understanding and designing polymeric liquids. In applications like these, several realizations are usually simulated/acquired to compose a reliable ensemble of the flow behavior.

In the other hand, *Modeling a physical spatial/temporal phenomenon* is a very important task on several decision making applications (Beccali et al., 2003). To represent *uncertainty* is for sure a relevant step on this task not only because there is an incomplete understanding of the process itself, but also because it is difficult to restrict their physical parameters (Mariethoz & Caers, 2014).

The modeling methods of several physical phenomena under uncertainty have two building blocks: a deterministic and a stochastic. Deterministic models come up with physically-based simulated outcomes. Stochastic models try to provide realizations that somehow cover the uncertainty space and at the

same time mimic the physics (providing a certain level of realism) (Mariethoz & Caers, 2014).

1.2
Contributions

Firstly, we aim to explore the applicability of the Helmholtz-Hodge Decomposition on the uncertainty quantification of vector fields. To do so, we first explore the uncertainty analysis of vector fields ensembles using an approach based on such method. In our approach, from a given vector field ensemble $\mathcal{E}$, we firstly generate the corresponding rotational-free, divergence-free and harmonic component ensembles by applying the Natural Helmholtz-Hodge decomposition (Pascucci et al., 2014) to each vector field in $\mathcal{E}$. With these ensembles in hand, our proposal not only quantifies how much each component ensemble is point-wisely correlated to the original vector field ensemble (via statistical analysis), but also allows the uncertainty investigation of rotational-free, divergence-free, and harmonic components separately.

Secondly, because *vector fields* are very important in a variety set of decision making problems related to *Scientific Computing*, we present two new stochastic methods to generate 2D vector fields from a single realization, as well as one to synthesize vector fields from an ensemble. The stochastic methods are based on the concepts of *Bootstrapping* and *Interpolation*. The vector fields synthesis from an ensemble, on the other hand, makes use of *Multidimensional Reduction* and *Inverse Projection*. Applications that make use of vector fields include, for example: fluid flow simulation (Anderson & Wendt, 1995), analysis of MRI data for medical prognosis (Tong et al., 2003) and weather prediction (Luo et al., 2012), just to cite a few. The deterministic simulation of vector fields in such applications may require expensive numerical computations (Anderson & Wendt, 1995).

The stochastic generation of physically realistic vector fields realizations is a challenging task. Many algorithms for multivariate stochastic simulation are based on very complex probabilistic models (Popescu et al., 1998; Xiu, 2009; Lall et al., 2016) and generally they are not adequate to mimic physical phenomena such as wind, for example.

1.3
Dissertation Outline

The remainder of this document is organized as follows: Firstly, Chapter 2 describes some related works regarding the applicability of the HHD, as well as previous works in the fields of uncertainty quantification and visualization.

It also presents related works of the techniques used in this work, like the Bootstrap and the Multidimensional Projection. Then, Chapter 3 presents the definition and mathematical background related to the Helmholtz-Hodge Decomposition. After that, Chapter 4 presents an uncertainty quantification approach for 2D vector field ensembles, followed by two methodologies to stochastically generate 2D vector field realizations from a single sample (Chapter 5) and one approach to synthesize vector fields from an ensemble (Chapter 6). Finally, Chapter 8 presents some final remarks and future works.

All simulations and methods here presented were developed using Python 2.7[1] and the following libraries/packages:

- Pillow 3.2[2];
- Matplotlib 1.5.1[3];
- Numpy 1.11.1[4];
- Scikit-learn 0.18.0[5]
- Scipy 0.17.1[6].

1.4
Publications

The uncertainty quantification approach for 2D vector field ensembles that will be presented in Chapter 4, as well as its results, was previously published[7] in the *Computer & Graphics* Journal[8], Volume 55.

---

[1]http://www.python.org
[2]http://pypi.python.org/pypi/Pillow
[3]http://matplotlib.org
[4]http://www.numpy.org
[5]http://scikit-learn.org
[6]http://www.scipy.org/
[7]http://dx.doi.org/10.1016/j.cag.2016.01.001
[8]http://www.elsevier.com/locate/cag

2
Previous Works

This Chapter aims to describe some previous works related to the topics used in this Thesis.

2.1
Helmholtz-Hodge Decomposition

For a complete study regarding the Helmholtz-Hodge Decomposition, Bhatia et al. (2013) present a survey. Among the methods to obtain the HHD, the least squares finite element minimizes an energy function to obtain the rotational-free and divergence-free components (Polthier & Preuß, 2003; Tong et al., 2003). This technique can be applied to a piece-wise constant vector field, which is defined on the triangles of the mesh and results in potential functions defined on the vertices of the mesh. Another technique is smoothed particle hydrodynamics, which is commonly used to simulate fluid flows (Monaghan, 1992). With this approach, it is possible to approximate the differential operators and, consequently, the Poisson equation of the HHD (Petronetto et al., 2010). Finite Difference Methods can also be used to numerically approximate differential operators to compute the HHD components (M. & Michel, 2010). Other techniques include the use of Fourier (Hinkle et al., 2009) and Wavelet (Deriaz & Perrier, 2009) domains.

The HHD is unique in unbounded domains for flows vanishing at infinity (Pascucci et al., 2014). However, for closed domains its uniqueness is guaranteed by enforcing some boundary conditions. It is well known, though, that imposing boundary conditions may introduce artifacts that were not present in the original field, because the boundary conditions create a strong dependency between the vector-field components and the shape and orientation of the boundary. Pascucci et al. (2014) introduced a new method called the Natural HHD (NHHD), which computes, on a point-wise basis, an artifact-free HHD on vector fields with open boundaries or unknown boundary conditions. A concise description of this method is given in Chapter 3; this is the method adopted in this work to obtain the decomposition.

The HHD has been used in a large range of applications. Gao et al. (2010) used the HHD to detect singularities for fingerprint matching, in which singular

points play a major role. The authors also proposed a combination of this technique with the Poincare index to obtain better performance in detecting singular points. Their results show that the proposed method is robust to noise in fingerprint image analyses.

Wang & Deng (2014) applied the HHD in the field of complex ocean flow visualization and analysis for feature extraction. The authors affirmed that using the HHD facilitates the process of detecting features, because after the decomposition both the rotational-free and divergence-free components are considerably less sensitive to noise compared to the original field. They conclude that the proposed method is able to detect the majority of critical points, with some inaccuracy at the boundary caused by the boundary conditions adopted, which are a requirement of the HHD.

In addition to these examples of HHD applications, we may cite cardiac video analysis (Guo et al., 2006), hurricane eye tracking (Palit, 2005), and the aerodynamic design of cars and aircrafts (Tong et al., 2003).

2.2
Vector Field Uncertainty

Wittenbrink et al. (1996) proposed the use of glyphs to visualize vector field uncertainty, graphically representing not only the mean direction and length, but also the uncertainty in direction and magnitude. Lodha et al. (1996) presented a set of methods for visualizing uncertainty in streamlines of flow data generated by different integration algorithms. These methods include glyphs, envelopes, animations, priority sequencing, trace viewpoints, and rakes. For time-dependent vector field uncertainty visualization, we may cite the work of Botchen et al. (2005), which introduced two texture-based techniques.

An approach to visualize the global uncertainty of vector fields is presented by Otto et al. (2010). To create this visualization, the authors considered the uncertainty of a particle location transported by the flow, using particle density functions to describe the probability that a particle is at a certain location. They used the Monte Carlo approach to integrate probabilistic particle paths. After testing with synthetic and real data sets, they concluded that the amount of computational time required is still significant. Moreover, they mentioned that the proposed method produces fewer uncertain critical points than the true amount presented by the input field, and they stated that their technique could be interpreted as a topology simplification method.

Petz et al. (2012) presented a general framework for extracting probabilistic local features from uncertain vector fields under consideration of the spatial correlation structure. To perform this task, they defined probabilistic

equivalents to critical points in uncertain 2D and 3D vector fields and cores of swirling motion in 3D. Such equivalents are also computed by the Monte Carlo integration.

Pfaffelmoser et al. (2013) introduced a method to analyze and visualize the variability of gradients in uncertain 2D discrete scalar fields. They first derived confidence intervals for the strength of the derivative in any direction and orientation. To visualize the derivative strength, they proposed an scheme using color diffusion to simultaneously show the data values and gradient variations. They also used circular glyphs to convey the uncertainty in gradient orientation. Results showed that their approach is useful to analyze the stability of uncertain 2D scalar fields with respect to both local derivatives and feature orientation.

A study to analyze the stability of critical points in uncertain scalar fields is provided by Mihai & Westermann (2014). The authors derived confidence intervals based on the gradient of the scalar ensemble to identify possible locations of critical points. Using the Hessian matrix, they were also able to identify the tendency of the critical points to behave similarly to the maxima, minima or saddle near a specified location. Their results illustrate how the proposed methods emphasize the possible critical points in ensembles of uncertain scalar fields.

2.3
Visualization of Data Correlations

Jen et al. (2004) proposed a tool named *ImageSurfer*, which aims to visualize correlations between two scalar field volumes. Sauber et al. (2006) presented the *Multifield-Graphs*, which is a very interesting approach to visually understand the correlation of multiple scalar fields. Their proposal not only allows an easily identification of which field correlate to another, but also reveals the strength of their correlation. Pfaffelmoser & Westermann (2012) proposed interactive visualization tools to help to analyze correlation structures in uncertain 2D scalar fields, which are modeled by multivariate Gaussian distributions. They also presented a correlation clustering algorithm that groups points of the domain according to a criterion based on correlations. Another interesting work is the one produced by Pfaffelmoser & Westermann (2013), which introduced a glyph-based visualization model that uses correlation as an indicator of uncertainty in 3D scalar fields. Finally, we should mention again the work of Petz et al. (2012), which used local covariance structures to analyze the uncertainty in vector fields.

2.4
Stochastic Simulation

The stochastic generation of physically realistic vector fields realizations is a challenging task. In one side, many algorithms based on probabilistic models for multivariate stochastic simulation (Popescu et al., 1998; Xiu, 2009; Lall et al., 2016) are very complex mathematically speaking and generally they are not adequate to mimic physical phenomena such as wind, for example. In the other side, there are several geostatistical methods in the literature dedicated to the stochastic simulation of spatial physical phenomena (Lantuéjoul, 2013). Generally, they are applied to the generation of univariate continuous or categorical functions defined on a 2D or 3D grid. They usually propose a parametric model of uncertainty to formulate the lack of knowledge, and models based on variogram are the most traditional ones (Oliver & Webster, 2014). Alternatively, non-parametric approaches, such as the ones based on Multiple-Point Statistics (MPS), have received a lot of investigation in the last five years. These approaches generate realizations of a spatial phenomenon based on a training image, which implicitly describes the phenomenon's construction process (Mariethoz & Caers, 2014). These methods have a very strong connection with computer graphics' texture synthesis techniques (Mariethoz & Lefebvre, 2014), like Image Quilting (Efros & Freeman, 2001), for example.

2.5
Bootstrap

The *Bootstrap method* is a statistical method based on resampling with replacement. It is commonly applied to measure the accuracy of statistical estimators (Efron, 1979). In general, such accuracy could be defined in terms of bias, variance, confidence intervals, prediction error or some other dispersion measure. This technique has been applied to visual computing problems, such as: performance evaluation for computer vision systems (Cho et al., 1997), searching for radial basis function parameter (Liew et al., 2016), evaluation of the influence of hidden information on supervised learning problems (Wang & Deng, 2014) and edge detection (Fu et al., 2012), among others.

2.6
Dimensionality Reduction

Dimensionality Reduction techniques have been largely applied to different problems and research areas. For instance, Casaca et al. (2012) proposed the use of multidimensional projection for image colorization. Through an interactive tool, users can manipulate the projected data to improve clusters

and the colorization results. By the time of the study, the authors concluded that their technique outperformed existing techniques in terms of accuracy and flexibility.

As another example, we can cite its applicability on the visual analysis of social network. In (Martins et al., 2012), the authors describe each individual as an array of attributes, projecting them in a low dimensional space. Hence, they are able to visually encode individuals that are highly related to one another. They also conclude that the achieved results present better node distribution when compared to conventional force-based graph drawing.

Finally, other applications of dimensionality reduction include sentiment analysis (Kim & Lee, 2014), face recognition (Schroff et al., 2015) and texture and email classification (Liu & Fieguth, 2012; Gomez & Moens, 2012).

3
The Helmholtz-Hodge Decomposition

3.1
Vector Fields

A *scalar field* is a map $f : \mathbb{R}^n \to \mathbb{R}$ that assigns a real number to each point in $\mathbb{R}^n$. A *vector field* is a map $\mathbf{V} : \mathbb{R}^n \to \mathbb{R}^n$ that assigns a vector in $\mathbb{R}^n$ to each point in its domain. In this work, we will deal with 2-dimensional vector fields. Thus, here a vector field $\mathbf{V}(x, y)$ on $\mathbb{R}^2$ has two component scalar fields that will be named V_1 and V_2. In this chapter, we establish other notations and present some useful properties of vector fields.

– **Gradient**

The gradient of a scalar function $f : \mathbb{R}^2 \to \mathbb{R}$ is a function from $\mathbb{R}^2$ to $\mathbb{R}^2$ given by:

$$\nabla f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right) \tag{3-1}$$

where $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ correspond to the partial derivatives of f in relation to x and y, respectively.

– **Divergence**

The divergence of a vector field $\mathbf{V} : \mathbb{R}^2 \to \mathbb{R}^2$ is given by:

$$
\begin{aligned}
\nabla \cdot \mathbf{V} &= \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y} \right) \cdot (V_1, V_2) \\
&= \frac{\partial V_1}{\partial x} + \frac{\partial V_2}{\partial y}
\end{aligned}
\tag{3-2}
$$

The divergence of a vector field is a real-valued function measuring at each point in its domain the amount of flow generated in an infinitesimal region around it (Polthier & Preuß, 2003).

– **Curl**

For a 2-dimensional vector field, the curl of **V** is defined as:

$$\nabla \times \mathbf{V} = \left\langle \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y} \right), (V_1, V_2) \right\rangle$$

$$
\begin{aligned}
&= \left(\frac{\partial V_2}{\partial x} - \frac{\partial V_1}{\partial y} \right) \\
&= \nabla \cdot (V_2, -V_1) \\
&= \nabla \cdot J(V_1, V_2) \\
&= (\nabla \cdot J)\mathbf{V}
\end{aligned}
\tag{3-3}
$$

where J is defined as:

$$J(u, v) = (v, -u) \tag{3-4}$$

We now define some special cases of vector fields. They are:

– **Potential Vector Field**

A vector field **V** is called potential if there is a scalar field φ as:

$$\mathbf{V} = \nabla \varphi \tag{3-5}$$

The scalar field φ is called potential scalar field of **V**. When a vector field **V** is defined over a simply connected domain, then **V** is a potential vector field if and only if its curl is null, i.e.:

$$(\nabla \cdot J)\mathbf{V} = 0 \tag{3-6}$$

– **Solenoidal Vector Field**

A vector field **V** is called soleinoidal if it is defined by the curl of another vector field ψ, i.e.:

$$\mathbf{V} = (\nabla \cdot J)\psi \tag{3-7}$$

The field ψ is called potential vector of **V**. A vector field is solenoidal if and only if its divergence is null, i.e.:

$$\nabla \cdot \mathbf{V} = 0 \tag{3-8}$$

– **Harmonic Vector Field**

If a vector field is both potential and solenoidal, it is called an harmonic field. So, an harmonic field satisfies the following conditions:

$$\begin{aligned}
\nabla \cdot \mathbf{V} &= 0 \\
(\nabla \cdot J)\mathbf{V} &= 0
\end{aligned} \tag{3-9}$$

3.2
Helmholtz-Hodge Decomposition for 2D Vector Fields

The Helmholtz-Hodge Decomposition (HHD) (Chorin, 1968) states that any vector field $\mathbf{V}$, defined in a simply connected domain, can be described as a sum of three orthogonal components:

$$\mathbf{V} = \nabla \varphi + (\nabla \cdot J)\psi + \mathbf{h} \tag{3-10}$$

where $\nabla \varphi$ is the *rotational-free* term, $(\nabla \cdot J)\psi$ is the *divergence-free* term and $\mathbf{h}$ is the *harmonic* term (Figure 3.1).

3.1(a): $\mathbf{V}$ 3.1(b): $\nabla \varphi$

3.1(c): $(\nabla \cdot J)\psi$ 3.1(d): $\mathbf{h}$

0.08 0.16 0.24 0.32 0.40 0.48 0.56
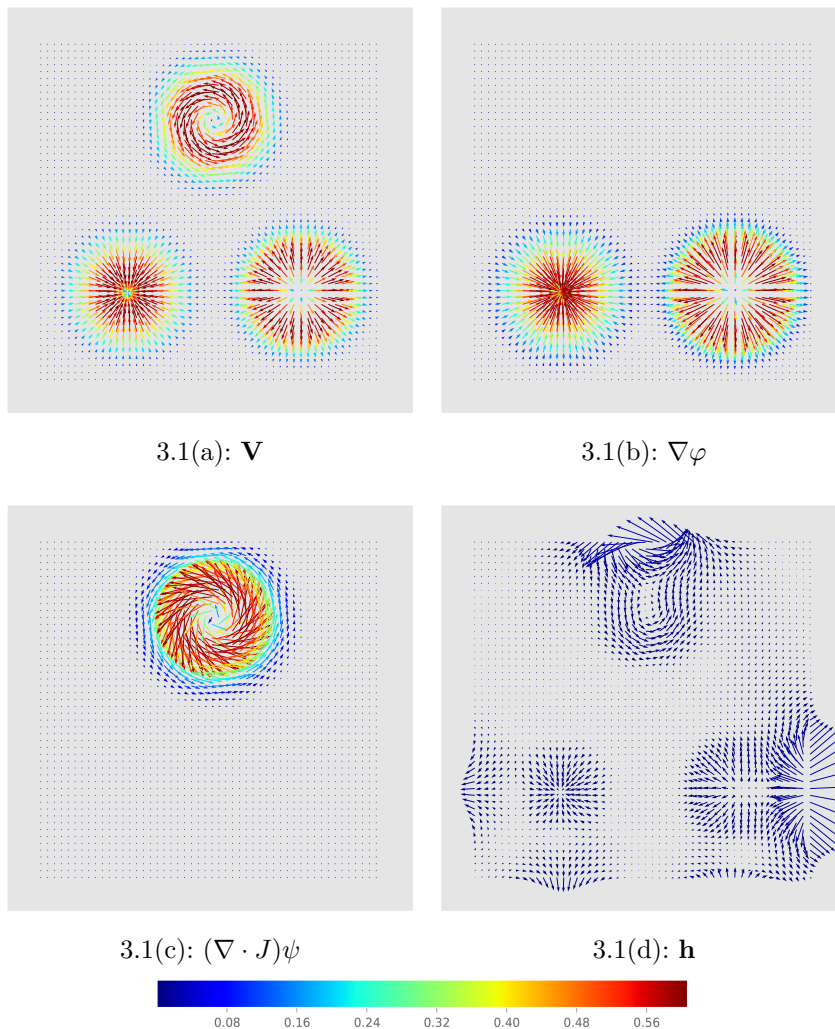
Figure 3.1: The HHD states that a vector field (a) is composed by a rotational-free (b), a divergence-free (c), and a harmonic component (d). The color bar represents the vector magnitudes.

Based on the explanation presented by do Carmo (2008), in the following it will be show how to derive the Helmholtz-Hodge Decomposition for 2-dimensional vector fields.

As presented before, it is known that:

$$
\begin{aligned}
(\nabla \cdot J)\nabla\varphi &= 0 \\
\nabla \cdot (\nabla \cdot J)\psi &= 0 \\
(\nabla \cdot J)\mathbf{h} &= 0 \\
\nabla \cdot \mathbf{h} &= 0
\end{aligned}
$$

Applying the divergence on both sides of Equation 3-10 results in:

$$
\begin{aligned}
\nabla \cdot \mathbf{V} &= \nabla \cdot (\nabla\varphi) + \nabla \cdot (J(\nabla\psi)) + \nabla \cdot \mathbf{h} \\
&= \nabla \cdot (\nabla\varphi) + 0 + 0 \\
&= \nabla \cdot (\nabla\varphi)
\end{aligned}
\tag{3-11}
$$

And, applying the curl in the same equation we have:

$$
\begin{aligned}
(\nabla \cdot J)\mathbf{V} &= (\nabla \cdot J)(\nabla\varphi) + (\nabla \cdot J)(J(\nabla\psi)) + (\nabla \cdot J) \cdot \mathbf{h} \\
&= \nabla \cdot J(\nabla\varphi) + \nabla \cdot J((J(\nabla\psi))) + \nabla \cdot J(h) \\
&= 0 + \nabla \cdot (-\nabla\psi) + 0 \\
&= -\Delta\psi
\end{aligned}
\tag{3-12}
$$

Now, one can find the divergence-free and rotational-free components of a 2-dimensional vector field $\mathbf{V}$ with the following linear equation system:

$$
\begin{cases}
\nabla \cdot \mathbf{V} &= \Delta\varphi \\
(\nabla \cdot J)\mathbf{V} &= -\Delta\psi
\end{cases}
\tag{3-13}
$$

where $\nabla\varphi$ is the potential component and $J(\nabla\psi)$ is the solenoidal component. The harmonic field $\mathbf{h}$ can be obtained using:

$$
\mathbf{h} = \mathbf{V} - \nabla\varphi - J(\nabla\psi).
$$

The HHD is unique on unbounded domains for flows vanishing at infinity. However, most practical cases deal with flows on bounded domains, where the HHD is not unique (Pascucci et al., 2014). Still accordingly to Pascucci et al. (2014), the fundamental reason for non-uniqueness in the HHD is that harmonic flows are both divergence-free and rotational-free. Then, one can add an arbitrary harmonic flow to any one of the three HHD components, and its negative to another, to obtain a different valid decomposition. As a result, any two valid HHDs differ in only how the harmonic flow is represented. Therefore, to obtain an unique solution for closed domains, some boundary conditions should be established. The normal-parallel (NP) boundary condition is the

most commonly used, which requires the divergence-free and the rotational-free components to be parallel and normal to the boundary, respectively:

$$\begin{cases} \nabla\varphi \times \mathbf{n} &= 0 \\ (\nabla \cdot J)\psi \cdot \mathbf{n} &= 0 \end{cases}, \tag{3-14}$$

where $\mathbf{n}$ represents the outward normal to the boundary.

Another possible boundary condition is to impose constant potentials on the boundary, which implies the rotational-free component normal to the boundary and the divergence-free tangent to it (Petronetto et al., 2010). However, such conditions may introduce artifacts that were not observed in the original field due to the imposed dependency between the vector field components and the shape and orientation of the boundary. To overcome this problem, Pascucci et al. (2014) proposed the Natural HHD (NHHD):

$$\mathbf{V}^* = \nabla\varphi^* + (\nabla \cdot J)\psi^* + \mathbf{h}^* \tag{3-15}$$

This technique separates the components by its influences, which can be internal or external. Here, $\nabla\varphi^*$ is the natural divergence and $(\nabla \cdot J)\psi^*$ is the natural rotational. They represent the components influenced by the divergence and rotational of $\mathbf{V}$ inside the domain. On the other hand, $\mathbf{h}^*$ is the natural harmonic, which is influenced only by the exterior of the domain.

Pascucci et al. (2014) use an integral kernel called the *Green's Function*, $G(\mathbf{x}, \mathbf{x}_{i,j})$, to solve the Poisson and Laplace equations. Such function is defined as the potential created at $\mathbf{x}_{i,j}$ due to an impulse source represented by the *Dirac Delta Function* located at $\mathbf{x}$:

$$\nabla^2 G(\mathbf{x}, \mathbf{x}_{i,j}) = \delta(\mathbf{x} - \mathbf{x}_{i,j})$$

For a single source located at $\mathbf{x}$ in an infinite domain, it is also called the *Free-Space Green's Function*, $G_\infty = G_\infty(\mathbf{x}, \mathbf{x_{i,j}})$, and it is given by:

$$G_\infty(\mathbf{x}, \mathbf{x}_{i,j}) = \frac{1}{2\pi} \log(\|\mathbf{x} - \mathbf{x}_{i,j}\|) \quad \mathbf{x}_{i,j}, \mathbf{x} \in \mathbb{R}^2$$

With this knowledge, the natural potentials can be computed as:

$$\begin{cases} \varphi^*(\mathbf{x}_{i,j}) &= \int_\Omega G_\infty \nabla \cdot \mathbf{V}(\mathbf{x}) & \mathbf{x}_{i,j}, \mathbf{x} \in \Omega \\ \psi^*(\mathbf{x}_{i,j}) &= - \int_\Omega G_\infty (\nabla \cdot J)\mathbf{V}(\mathbf{x}) & \mathbf{x}_{i,j}, \mathbf{x} \in \Omega \end{cases}, \tag{3-16}$$

where Ω is a bounded subset of $\mathbb{R}^n$ and $\mathbf{V} : \Omega \to \mathbb{R}^n$, for $n = 2$.

According to Pascucci et al. (2014), the only limitation of the NHHD is that it does not guarantee the components to be L_2-orthogonal. However, they consider that, for applications in visualization and analysis, orthogonality is

not required. Instead, they seek to preserve properties like flow topology, which can be attributed to the interior and exterior of the domain.

3.3
Implementation Details

Pascucci et al. (2014) also defined an approach to compute the integration over the domain. Considering C, E and I the set of vertices in $\mathbb{S}_{m,n}$ at the 4 corners, at the boundary (not including the corners) and the interior of $\mathbb{S}_{m,n}$ (where $\mathbb{S}$ is a grid of size $m \times n$), respectively, the integral of a function f over $\mathbb{S}_{m,n}$ can be obtained by applying the following integration method:

$$\int_{\mathbb{S}_{m,n}} f(x,y)\delta x \delta y \approx \frac{\Delta_x \Delta_y}{4} \left(\sum_{v \in \mathsf{C}} f_v + 2 \sum_{v \in \mathsf{E}} f_v + 4 \sum_{v \in \mathsf{I}} f_v \right)$$

where Δ_x and Δ_y represent the grid spacing in the x and y directions, respectively, and f_v is the value of f at vertex v.

In this work, to calculate the required partial derivatives on $\mathbb{S}_{m,n}$, we adopted the Finite Difference Methods. More specifically, we apply the central finite difference in the interior nodes of $\mathbb{S}_{m,n}$, using the kernels defined in Equations 3-17 and 3-18 for the x and y derivatives, in that order. Along the boundary of $\mathbb{S}_{m,n}$, we apply either forward finite difference or backward finite difference, according to the boundary point position. These derivatives are used to approximate the curl, the divergence and the gradient operators.

$$\frac{1}{2h} \cdot \begin{bmatrix} -1 & 0 & 1 \end{bmatrix}, \qquad (3\text{-}17) \qquad \qquad \frac{1}{2h} \cdot \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \qquad (3\text{-}18)$$

4
Uncertainty Analysis of Vector Field Ensembles

In this Chapter, we will present the application of the Helmholtz-Hodge Decomposition as a tool for the analysis of 2D vector field ensembles, described by Ribeiro et al. (2016). In summary, given a vector field ensemble $\mathcal{E}$, we first compute the corresponding rotational-free, divergence-free and harmonic component ensembles by using the Natural Helmholtz-Hodge Decomposition on each vector field in $\mathcal{E}$. With these ensembles in hand we are able not only to quantify, via a statistical analysis, how much each component ensemble is point-wisely correlated to the original vector field ensemble, but also investigate the uncertainty of the rotational-free, divergence-free and harmonic components separately.

4.1
Motivation

It is important to notice that the Helmholtz-Hodge Decomposition has been successfully applied to different fields (Petronetto et al., 2010; Palit, 2005; Wiebel, 2004; Polthier & Preuß, 2003).

Using the Helmholtz-Hodge Decomposition in the analysis of 2D vector field ensembles can improve the understanding of the flow kinematics because:

– It allows the identification of regions where a set of vector field realizations is dominated by an extensional flow (rotational equals to zero) or regions where it behaves like an incompressible flow (divergence equals to zero). In some applications like blow molding and fiber spinning, to verify that extensional flow is the dominant type of deformation is important (Huang et al., 2013).

– It facilitates the analysis of each decomposed component ensemble directly. This is central for the verification of theories, and for the understanding of non-linear rheological properties of sampled complex flow ensembles (Sagis & Fischer, 2014; van Oudheusden et al., 2007). In several studies the interest is mainly concentrated in the divergence-free derived ensemble (Sadati et al., 2011).

– It enables the study of each component ensemble uncertainty. The uncertainty analysis of vector field ensemble is significant for the complete

understanding of the flow behavior (Petz et al., 2012; Pfaffelmoser et al., 2013).

4.2
Uncertainty Model

Given a probability space $(\Omega, \mathcal{F}, P)$ and a discrete sampling of a 2-dimensional domain on a Cartesian grid structure $\mathbb{S}_{m,n} = \{\mathbf{x}_{i,j} \in \mathbb{R}^2 : 1 \leq i \leq m, 1 \leq j \leq n\}$, we model the data uncertainty by using a two-valued random field $\mathbf{V}$, i.e., a collection of two scalar-valued random variables, named $\mathbf{S}_1$ and $\mathbf{S}_2$, indexed by the elements in $\mathbb{S}_{m,n}$ (Adler, 1981).

Once we are applying the NHHD to the vector field, it is also necessary to define how we quantify the uncertainty of the potential scalar fields and their respective gradient fields.

Suppose that the underlying data uncertainty is represented by an ensemble $\mathcal{E} = \{\mathbf{V}_1, \ldots, \mathbf{V}_R\}$ of R realizations of 2-dimensional vector fields $\mathbf{V}_r$, with $r \in \{1, \ldots, R\}$, at all spatial points in $\mathbb{S}_{m,n}$. Also, suppose that for each 2-dimensional vector field realization $\mathbf{V}_r \in \mathcal{E}$, we have its NHHD at each $\mathbf{x}_{i,j} \in \mathbb{S}_{m,n}$:

$$\mathbf{V}_r(\mathbf{x}_{i,j}) = \nabla \varphi_r(\mathbf{x}_{i,j}) + (\nabla \cdot J)\psi_r(\mathbf{x}_{i,j}) + \mathbf{h}_r(\mathbf{x}_{i,j}) \tag{4-1}$$

After applying the NHHD on each element $\mathbf{V}_r \in \mathcal{E}$, we define two other scalar random field ensembles with R realizations at each $\mathbf{x}_{i,j} \in \mathbb{S}_{m,n}$. The first ensemble is the *rotational-free scalar potential ensemble* $\mathcal{C} = \{\varphi_1, \ldots, \varphi_R\}$ and the second ensemble is the *divergence-free scalar potential ensemble* $\mathcal{D} = \{\psi_1, \ldots, \psi_R\}$.

From the $\mathcal{C}$ and $\mathcal{D}$ scalar potential ensembles is then possible to derive the corresponding rotational-free and divergence-free vector fields ensembles. The *rotational-free vector field components ensemble*, denoted by $\mathcal{E}_{\mathcal{C}}$, is obtained by computing the gradient of the potential field ensemble $\mathcal{C}$ at each $\mathbf{x}_{i,j} \in \mathbb{S}_{m,n}$. The *divergence-free vector field component ensemble*, denoted by $\mathcal{E}_{\mathcal{D}}$, is obtained by calculating the gradient of the potential field ensemble $\mathcal{D}$ and by applying the J operator at each $\mathbf{x}_{i,j} \in \mathbb{S}_{m,n}$.

Given the divergence-free and rotational-free components, the harmonic component for each realization could be directly determined by the following equation:

$$\mathbf{h}_r = \mathbf{V}_r - \nabla \varphi_r - J(\nabla \psi_r) \tag{4-2}$$

Additionally, we can group these R realizations of the harmonic components to build the *harmonic vector field component ensemble*, $\mathcal{E}_{\mathcal{H}}$.

4.3
Uncertainty Quantification

According to Potter et al. (2013), two particular summary statistics stand out as the *de facto* characterization: mean and standard deviation. These statistics reduce uncertainty to an expected value and variation from that value and are particularly effective in expressing normally distributed data. Given that, to compute the uncertainty of the magnitude and orientation of a vector field ensemble $\mathcal{F} \in \{\mathcal{E}, \mathcal{E}_\mathcal{C}, \mathcal{E}_\mathcal{D}, \mathcal{E}_\mathcal{H}\}$, we calculate the standard deviation, which measures the amount of dispersion from the average value.

For instance, let us define, for each $\mathcal{F} \in \{\mathcal{E}, \mathcal{E}_\mathcal{C}, \mathcal{E}_\mathcal{D}, \mathcal{E}_\mathcal{H}\}$, the mean vector field as:

$$\mathbf{V}_\mu^\mathcal{F}(\mathbf{x}_{i,j}) = \frac{1}{R} \sum_{r=1}^{R} \mathbf{V}_r^\mathcal{F}(\mathbf{x}_{i,j}),$$

where $\mathbf{V}_r^\mathcal{F}$ is a vector in the ensemble $\mathcal{F}$.

To measure the uncertainty about the magnitude, we use the following standard statistical estimators (Wasserman, 2004) at each point $\mathbf{x}_{i,j} \in \mathbb{S}_{m,n}$:

$$\mu_{mag}^\mathcal{F}(\mathbf{x}_{i,j}) = \frac{1}{R} \sum_{r=1}^{R} ||\mathbf{V}_r^\mathcal{F}(\mathbf{x}_{i,j})||$$

and

$$\sigma_{mag}^\mathcal{F}(\mathbf{x}_{i,j}) = \sqrt{\frac{1}{R-1} \sum_{r=1}^{R} (||\mathbf{V}_r^\mathcal{F}(\mathbf{x}_{i,j})|| - \mu_{mag}^\mathcal{F}(\mathbf{x}_{i,j}))^2}$$

Finally, to quantify the uncertainty of the scalar field potential ensembles $\mathcal{C} = \{\varphi_1, \ldots, \varphi_R\}$ and $\mathcal{D} = \{\psi_1, \ldots, \psi_R\}$, we also compute the mean and standard deviations at each point $\mathbf{x}_{i,j} \in \mathbb{S}_{m,n}$:

$$\mu_\varphi(\mathbf{x}_{i,j}) = \frac{1}{R} \sum_{r=1}^{R} \varphi_r(\mathbf{x}_{i,j}),$$

$$\sigma_\varphi(\mathbf{x}_{i,j}) = \sqrt{\frac{1}{R-1} \sum_{r=1}^{R} (\varphi_r(\mathbf{x}_{i,j}) - \mu_\varphi(\mathbf{x}_{i,j}))^2},$$

and

$$\mu_\psi(\mathbf{x}_{i,j}) = \frac{1}{R} \sum_{r=1}^{R} \psi_r(\mathbf{x}_{i,j}),$$

$$\sigma_\psi(\mathbf{x}_{i,j}) = \sqrt{\frac{1}{R-1} \sum_{r=1}^{R} (\psi_r(\mathbf{x}_{i,j}) - \mu_\psi(\mathbf{x}_{i,j}))^2}$$

To measure the correlation between the vector field in the given ensemble $\mathcal{E}$ and the vector fields of each component ensemble derived from the NHHD, at each point $\mathbf{x}_{i,j} \in \mathbb{S}_{m,n}$, we adopted the *Spearman's Rank Correlation*.

The Spearman's Rank Correlation coefficient is a non-parametric correlation coefficient that measures the strength of association between two ranked random variables, i.e., the statistical dependence between them (Croux & Dehon, 2010). Its objective is to measure how well the relationship between two variables can be described by a monotonic function.

Given two samples of size n, say $A = \{A_i; i = 1, \ldots, n\}$ and $B = \{B_i; i = 1, \ldots, n\}$, the Spearman's Rank Correlation coefficient is calculated using the variables a_i and b_i, which represent the rank of A_i and B_i, respectively, according to a sorting criterion, as follows:

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}, \tag{4-3}$$

where $d_i = a_i - b_i$ is the difference between ranks.

With this knowledge, the correlation between the vector field and its components is calculated for each $\mathbf{x}_{i,j} \in \mathbb{S}_{m,n}$, between the original ensemble $\mathcal{E}$ and each component ensemble generated by the NHHD: $\mathcal{E}_\mathcal{C}, \mathcal{E}_\mathcal{D}, \mathcal{E}_\mathcal{H}$.

For the r-th vector $\mathbf{V}_r^\mathcal{F}(\mathbf{x}_{i,j})$ on the ensemble $\mathcal{F} \in \{\mathcal{E}, \mathcal{E}_\mathcal{C}, \mathcal{E}_\mathcal{D}, \mathcal{E}_\mathcal{H}\}$ at $\mathbf{x}_{i,j} \in \mathbb{S}_{m,n}$, we compute its similarity to the mean vector field of the original ensemble at that point, $\mathbf{V}_\mu^\mathcal{E}(\mathbf{x}_{i,j})$, according to the following formula:

$$sim_r^\mathcal{F}(\mathbf{x}_{i,j}) = \langle \mathbf{V}_r^\mathcal{F}(\mathbf{x}_{i,j}), \mathbf{V}_\mu^\mathcal{E}(\mathbf{x}_{i,j}) \rangle.$$

We have chosen this measure of similarity because it is simple and it comprehensively includes the magnitude and the orientation aspects of the two vector fields.

After performing this computation, for each vector $\mathbf{V}_r^\mathcal{F}(\mathbf{x}_{i,j})$, we sort $sim_r^\mathcal{F}(\mathbf{x}_{i,j})$ in ascending order for each ensemble in $\mathcal{F}$.

Let a_r be the rank position of the mean vector field of $\mathcal{E}$, and let b_r^G be the rank position of the r-th vector in $\mathcal{G} \in \{\mathcal{E}_\mathcal{C}, \mathcal{E}_\mathcal{D}, \mathcal{E}_\mathcal{H}\}$. In a next step, using Equation 4-3, we evaluate at each point $\mathbf{x}_{i,j}$ the Spearman's Rank Correlation of the mean vector field and the NHHD component ensemble $\mathcal{G}$, obtaining $\rho^G(\mathbf{x}_{i,j})$.

With the three correlation coefficients of the corresponding rotational-free, divergence-free and harmonic components in hand, we could classify at each point in the grid which component ensemble is more cor-

related to the original ensemble by just computing the maximum of $\{\rho^{\mathcal{E}_C}(\mathbf{x}_{i,j}), \rho^{\mathcal{E}_D}(\mathbf{x}_{i,j}), \rho^{\mathcal{E}_H}(\mathbf{x}_{i,j})\}$.

4.4
Results and Discussion

Tests using the proposed technique were carried out with two data ensembles: a synthetic ensemble and a simulated ensemble. In the following, we present the results obtained for each of these ensembles.

4.4.1
Synthetic Data Ensemble

Input Ensemble The synthetic data ensemble is obtained constructing a vector field ensemble from 100 randomly generated rotational-free and divergence-free potentials. A rotational-free realization is given by:

$$\varphi(x,y) = e^{-((x-x_0)^2 + (y-y_0{}^2)/2)} - e^{-((x-x_1{}^2) + (y-y_1{}^2)/2)},$$

where (x_0, y_0) and (x_1, y_1) are the coordinates of the source and the sink, respectively. The coordinates of these two points are obtained by adding a random vector (z_1, z_2), where z_i is a normal random variable with 0 mean and variance 1, to the points $(3, -3)$ and $(-3, -3)$, respectively.

The divergence-free potential realization, on the other hand, is given by:

$$\psi(x,y) = e^{-((x-x_2)^2 + (y-y_2{}^2)/2)},$$

where (x_2, y_2) is the coordinate of the vortex center, which are also obtained by adding a random vector (z_1, z_2), where z_i is a normal random variable with 0 mean and variance 1, to the point $(0, 3)$.

Given $R = 100$ realizations of each potential field, the vector field ensemble $\mathcal{E}$ is given by the summation of $\nabla\varphi$ and $J(\nabla\psi)$. We set the harmonic component as zero. The vector field's domain is the square $[-6, 6] \times [-6, 6]$ and it is discretized in a grid with size 100×100.

Results Henceforth, we will refer to the vector field ensemble as $\mathcal{E}$. With $\mathcal{E}$ in hand, we perform the NHHD to obtain the rotational-free and divergence-free scalar potentials ensembles $\mathcal{C}$ and $\mathcal{D}$. Then, through the NHHD we compute the rotational-free φ_r and divergence-free ψ_r potentials for each element $\mathbf{V}_r \in \mathcal{E}$. Figure 4.1 presents the maps of the means for the $\mathcal{C}$ and $\mathcal{D}$ scalar potential ensembles, i.e., the maps of $\mu_{\mathcal{C}}$ and $\mu_{\mathcal{D}}$. Figure 4.2, on the other hand, shows

their standard deviation. Through this figure, we can see that the uncertainty regarding the potential fields has a non-symmetric behavior close to the points defined as the main vector field's features (where the noise was added).

4.1(a): Rotational-free potential mean: $\mu_{\mathcal{C}}$

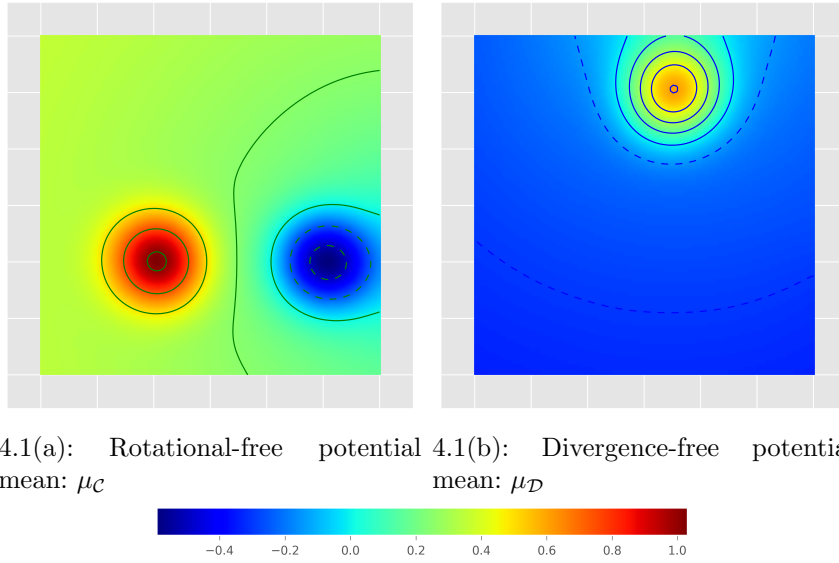4.1(b): Divergence-free potential mean: $\mu_{\mathcal{D}}$

Figure 4.1: Maps of the means of the rotational-free and divergence-free potential ensembles. Positive contours are represented by continuous lines and negative contours are represented by dashed lines.

4.2(a): Rotational-free potential standard deviation: $\sigma_{\mathcal{C}}$

4.2(b): Divergence-free potential standard deviation: $\sigma_{\mathcal{D}}$
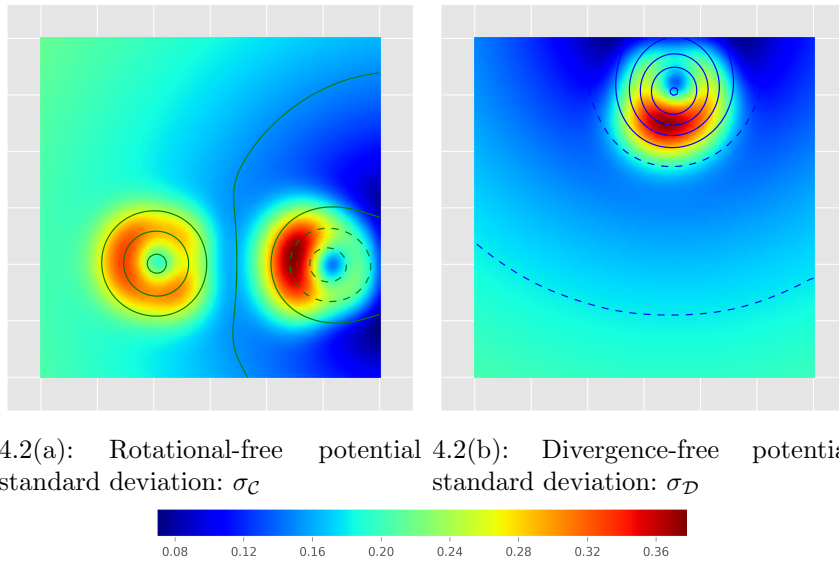
Figure 4.2: Standard deviation of the rotational-free and divergence-free potentials. Contours represent the mean potential fields. Positive contours are represented by continuous lines and negative contours are represented by dashed lines.

Figure 4.3 shows the mean vector fields for the synthetic ensemble $\mathcal{E}$ as well as for its corresponding rotational-free $\mathcal{E}_{\mathcal{C}}$, divergence-free $\mathcal{E}_{\mathcal{D}}$, and harmonic $\mathcal{E}_{\mathcal{H}}$ ensembles, in that other.

4.3(a): Vector field mean: $\mu_{\mathcal{E}}$ 4.3(b): Rotational-free mean: $\mu_{\mathcal{E}_c}$

4.3(c): Divergence-free mean: $\mu_{\mathcal{E}_\mathcal{D}}$ 4.3(d): Harmonic mean: $\mu_{\mathcal{E}_\mathcal{H}}$
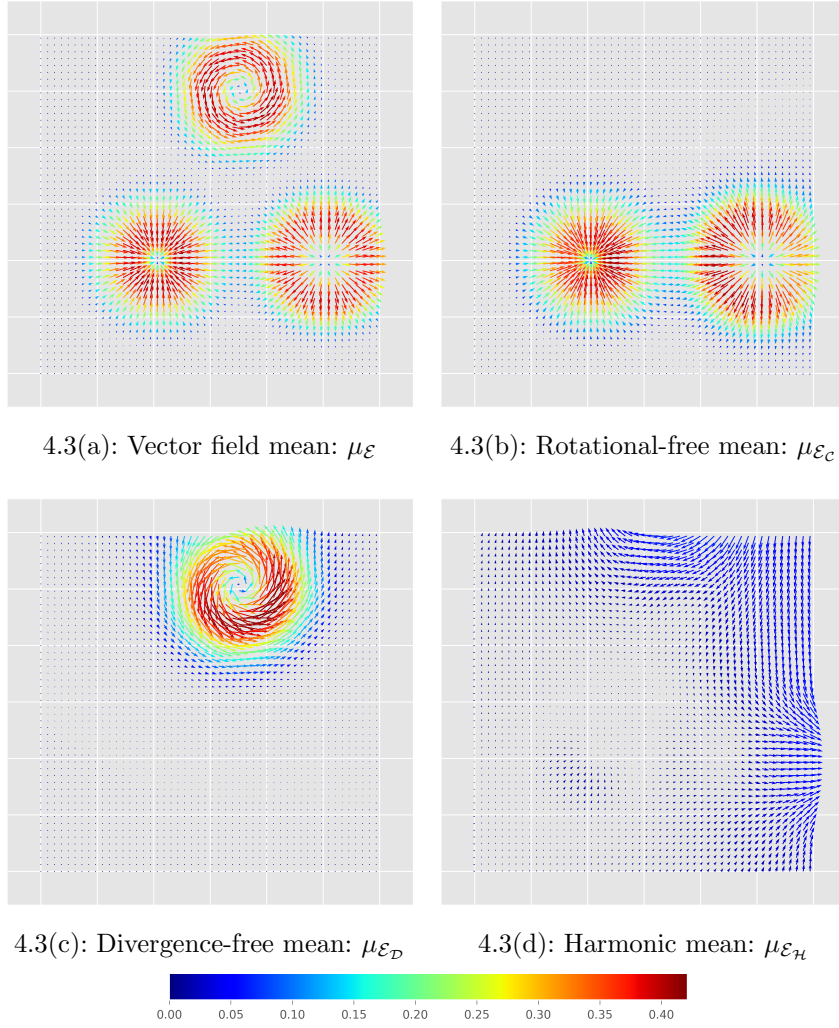
Figure 4.3: Mean vector fields obtained through the NHHD for the synthetic data ensemble $\mathcal{E}$. The color represents the vectors' magnitude, while the color scale is common for all NHHD components. However, the size of the rendered vectors is re-scaled on each subfigure for better visualization.

Figures 4.4 (a), (b) and (c) show the results of the Spearman's Rank correlation between each component ensemble and the original ensemble $\mathcal{E}$, computed as explained in Section 4.3. In Figure 4.4 (d), we show the maximum correlation classification. Color is used to represent which of the three components — rotational-free (blue), divergence-free (white) and harmonic (red) — has the maximum correlation value at each point. With this approach is possible to observe areas where each of the component ensembles is more correlated to the original ensemble.

Finally, Figure 4.5 shows the ensemble standard deviation regarding the magnitude.
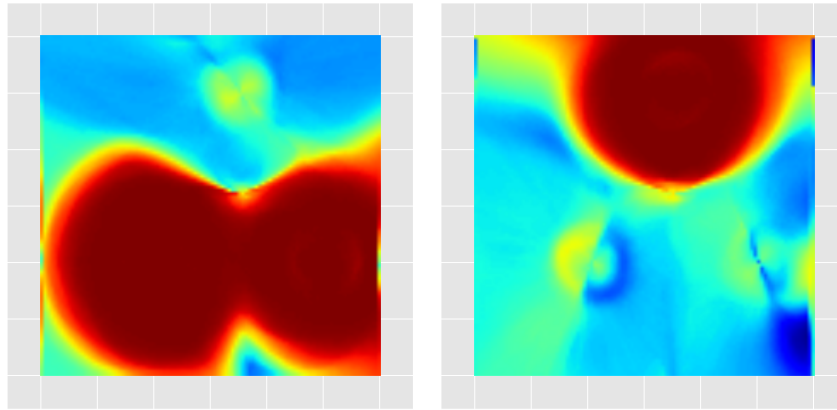
4.4(a): Rotational-free / vector-field correlation

4.4(b): Divergence-free / vector-field correlation
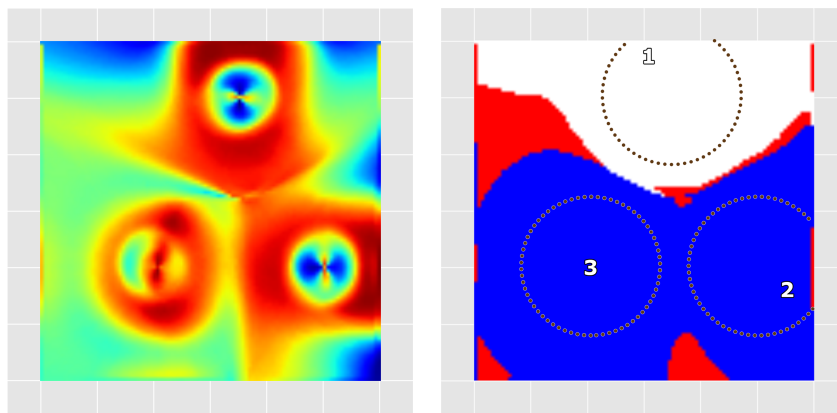
4.4(c): Harmonic / vector-field correlation

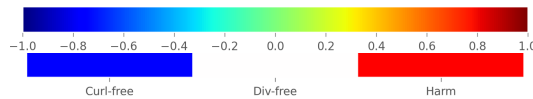4.4(d): Maximum correlation classification

Figure 4.4: Correlation between the vector field ensemble $\mathcal{E}$ and each of its NHHD component ensemble.

Analysis With he proposed approach, it is possible to analyze the contribution of each component — rotational-free, divergence-free and harmonic — to the total uncertainty of the ensemble magnitude (Figure 4.5). In this specific example, one can see that no significant overlap can be found in the magnitude uncertainty of the rotational-free and divergence-free components. However, if an overlap does exist, this technique may clarify how each component contributes to the uncertainty of a region. For instance, it is possible to perceive that the harmonic component uncertainty slightly overlaps both the divergence-free and rotational-free components uncertainty in two regions, at the top (1) and right (2) borders. Using the correlation information depicted in Figure 4.4, we can now observe that the divergence-free component is the major contributor
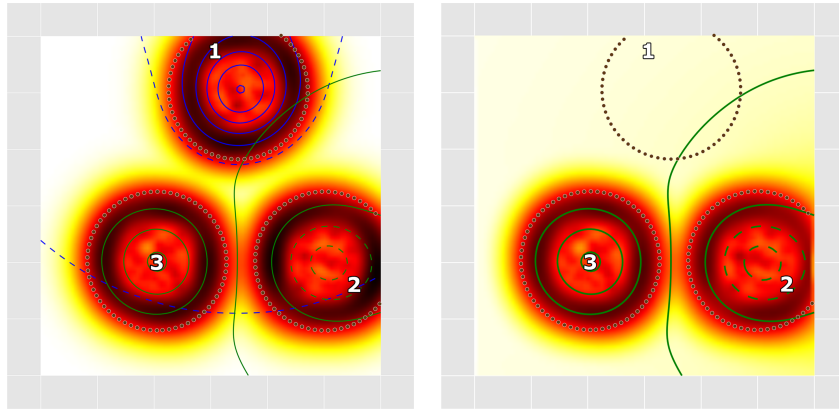
4.5(a): Original ensemble magnitude standard deviation

4.5(b): Rotational-free magnitude standard deviation
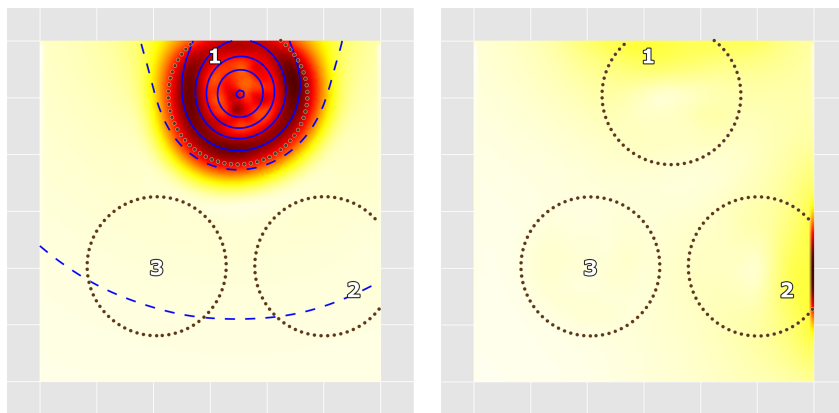
4.5(c): Divergence-free magnitude standard deviation

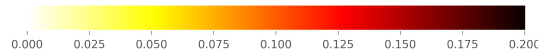4.5(d): Harmonic magnitude standard deviation

Figure 4.5: Standard deviation of the magnitude of the rotational-free, divergence-free and harmonic components.

to the uncertainty of region (1), as expected. The same occurs with region (3), with its uncertainty coming from the rotational-free component. As to region (2), though, we notice that, at the right border, the harmonic component contributes more than the rotational-free component.

4.4.2
Wind Forecast Ensemble

Input Ensemble This set is comprehended by seven multi-method wind forecast realizations $\mathcal{E}$, generated by the Instituto Nacional de Pesquisas Espaciais (INPE), Brazil. Each ensemble member defines a possible realization of a wind forecast for a region delimited by $35°48'S$ and $83°W$ as the minimum latitude and longitude coordinates (DMS), respectively, and by $6°12'N$ and

$25°48'W$ as the maximum latitude and longitude coordinates, in that order. The data is discretized in a 144×106 grid.

Results   We apply the NHHD to each realization of the weather forecast ensemble to obtain the rotational-free and divergence-free components of the vector fields: $\mathcal{C}$ and $\mathcal{D}$. Figure 4.6 illustrates the mean maps of the $\mathcal{C}$ and $\mathcal{D}$ potential data ensembles, i.e., the maps of $\mu_{\mathcal{C}}$ and $\mu_{\mathcal{D}}$. Figure 4.7 shows the corresponding standard deviation maps of the potential ensembles.

Figure 4.8 presents the mean vector field of the forecast ensemble $\mathcal{E}$ and the mean vector field of its corresponding rotational-free $\mathcal{E}_{\mathcal{C}}$, divergence-free $\mathcal{E}_{\mathcal{D}}$ and harmonic $\mathcal{E}_{\mathcal{H}}$ ensembles.

Figures 4.9 (a), (b), and (c) show the Spearman's correlation between the original vector field and the rotational-free, divergence-free and harmonic components, respectively. Figure 4.9 (d) shows the components which have a greater point-wise correlation with the vector field. In this figure, a blue, white and red points correspond to the rotational-free, divergence-free, and harmonic components, respectively.

Finally, Figure 4.10 illustrates the uncertainty related to the magnitude.

Analysis   As we have seen, through this approach it is possible to note how the uncertainty of each component contributes to the original field uncertainty. For instance, looking at Figure 4.10 we can state that the uncertainty of region (1) has a strong contribution from the rotational-free component. However, in this region we could also observe a small contribution from the divergence-free magnitude uncertainty. When analyzing Figure 4.9, it is possible to perceive that the core of this region is indeed formed by the rotational-free component. However, it is surrounded by the uncertainty that comes from the divergence-free component. Analyzing only Figure 4.10, we might also conclude that the magnitude uncertainty in region (2) has a strong contribution from the rotational-free component ensemble. With the proposed technique, one can verify that the uncertainty magnitude in this region is, indeed, defined by the rotational-free component uncertainty. Region (3), on the other hand, seems to have a strong contribution from the uncertainty of the divergence-free component. Nevertheless, analyzing the correlation between the original vector field and its components, we can perceive that this region, besides having its core related to the uncertainty of the divergence-free component, its also influenced by the harmonic and rotational-free components uncertainty. Something similar occurs with region (4). The uncertainty in this region seems to have a main contribution from the divergence-free component, with some
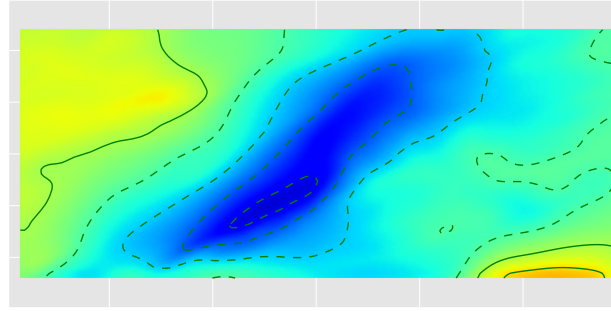
4.6(a): Rotational-free potential mean: $\mu_{\mathcal{C}}$
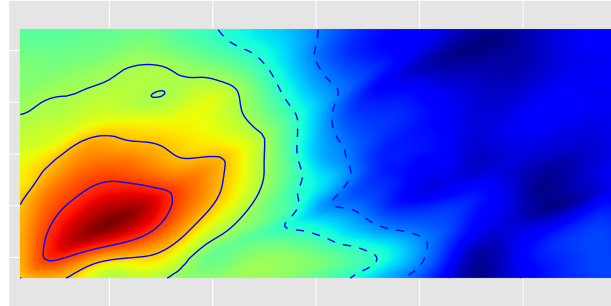
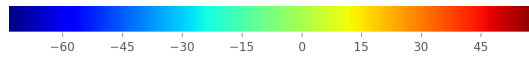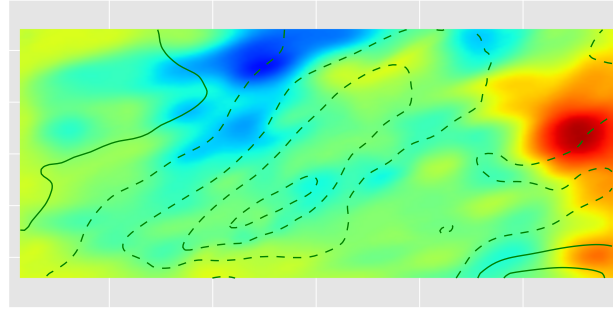4.6(b): Divergence-free potential mean: $\mu_{\mathcal{D}}$

Figure 4.6: Maps of the rotational-free and divergence-free potential ensembles mean. Positive contours are represented by continuous lines and negative contours are represented by dashed lines.

contribution coming also from the rotational-free magnitude's uncertainty. However, with the correlation data, we found that region (4) is mainly correlated to the rotational-free component. At last, notice that, from Figure 4.10, we can assume that uncertainty in the magnitude in region (5) has a strong contribution from the uncertainty of the divergence-free magnitude. In that case, Figure 4.9 shows that the core of these regions is, in fact, defined by the rotational-free component.

Unfortunately, for this data set there are no relevant meteorological discovery observed in the analysis. However, an important fact observed in the analysis of this example is that we could find regions in the ensemble domain where one component dominates the flow but another component contributes strongly to the uncertainty in the magnitude, which is related to kinetic energy.

According to an atmospheric data analyst opinion, in recent meteorological analysis, the simulated kinetic energy of the rotational and divergent wind components shows to be crucial for the wind forecasting analysis (Blažica et al., 2013). Moreover, this kind of analysis could be improved by taking into account the uncertainties because it is pivotal for observing the balance between

4.7(a): Rotational-free potential standard deviation: $\sigma_{\mathcal{C}}$
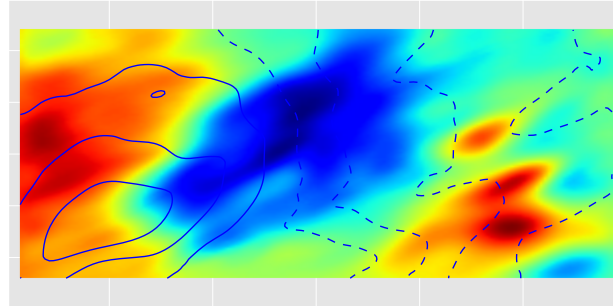
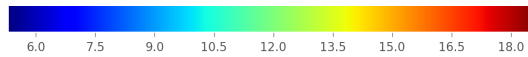4.7(b): Divergence-free potential standard deviation: $\sigma_{\mathcal{D}}$

Figure 4.7: Standard deviation of the rotational-free and divergence-free potentials.

these energies (Chen, 1980; Chen & Tribbia, 1981).

4.4.3
Final remarks

To better analyze the proposed approach, we performed another experiment. Its goal was to investigate the correlation analysis approach when the given ensemble is composed by pure divergence-free vector fields. To perform such an analysis, we consider as an input ensemble the derived divergence-free ensemble obtained from the application of the NHHD on the wind forecast data set. Figure 4.11 shows the mean vectors of the resulting NHHD decomposition. The result of the correlation analysis, illustrated in Figure 4.12, shows that this pure divergence-free vector field ensemble is mostly correlated to the divergence-free component, as expected. We can observe that there is a noisy region around the border and that there are some points in the interior dominated by the other components. The norms of the vectors at such points are very small, in their great majority. In terms of the vector magnitude uncertainty, we can observe in Figure 4.13 that the uncertainty on the derived divergence-free ensemble is quite similar to the uncertainty on the input ensemble. Although
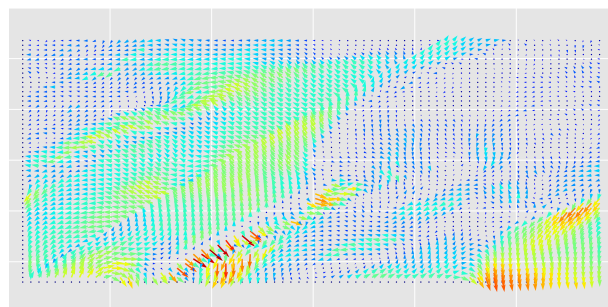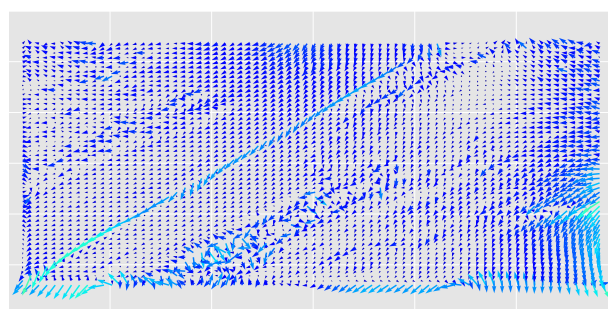
4.8(a): Vector Field mean: $\mu_{\mathcal{E}}$
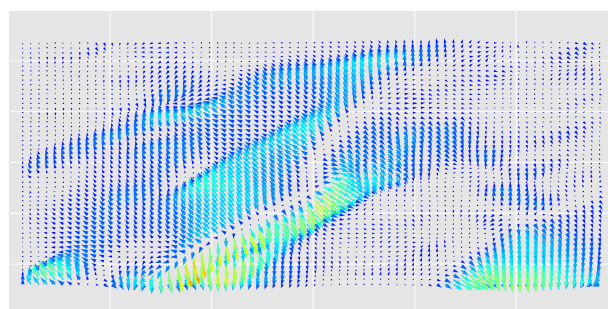
4.8(b): Rotational-free component mean: $\mu_{\mathcal{E}_{\mathcal{C}}}$
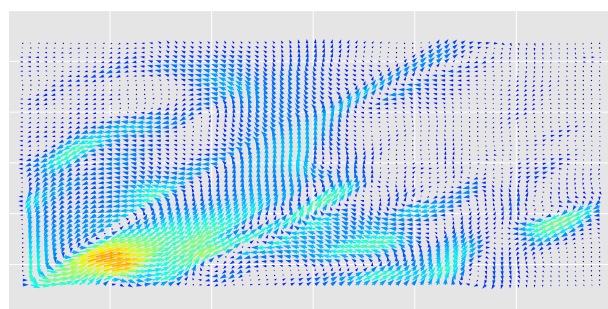
4.8(c): Divergence-free component mean: $\mu_{\mathcal{E}_{\mathcal{D}}}$

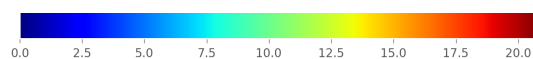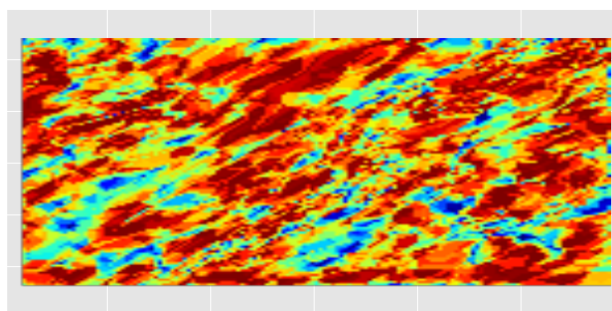4.8(d): Harmonic component mean: $\mu_{\mathcal{E}_{\mathcal{H}}}$

Figure 4.8: Mean vector fields obtained through NHHD for the wind forecast data ensemble. The color represents the vectors' magnitude, while the color scale is common for all NHHD components. The size of the rendered vectors is individually scaled in each subfigure for better visualization.

4.9(a): Rotational-free / vector-field correlation
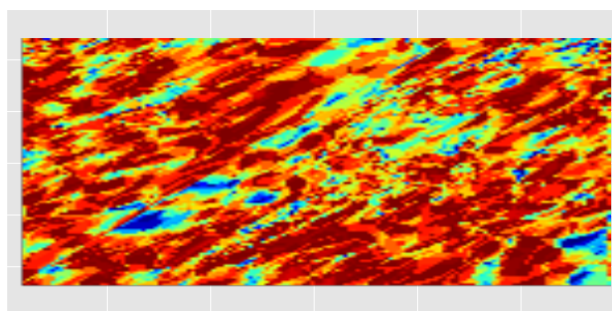
4.9(b): Divergence-free / vector-field correlation
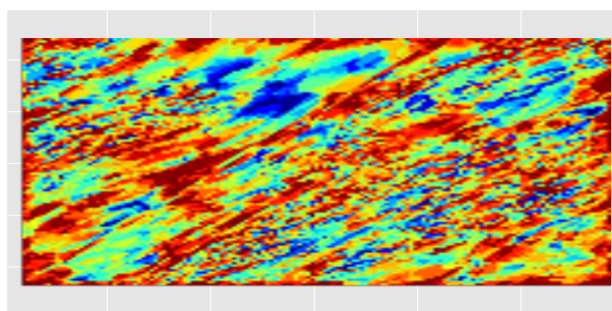
4.9(c): Harmonic / vector-field correlation
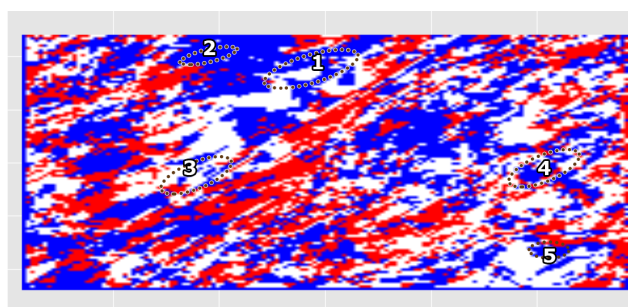
4.9(d): Maximum correlation

Figure 4.9: Correlation between the vector field and the magnitude of each of its components.
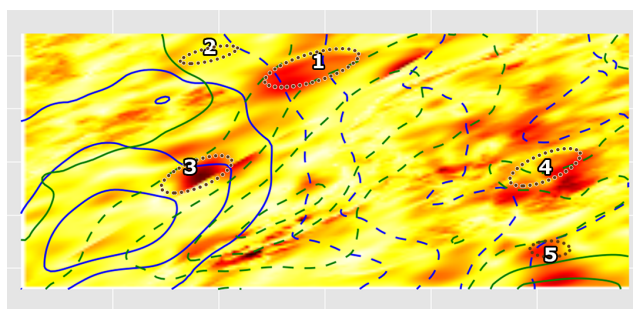
4.10(a): Vector field magnitude standard deviation
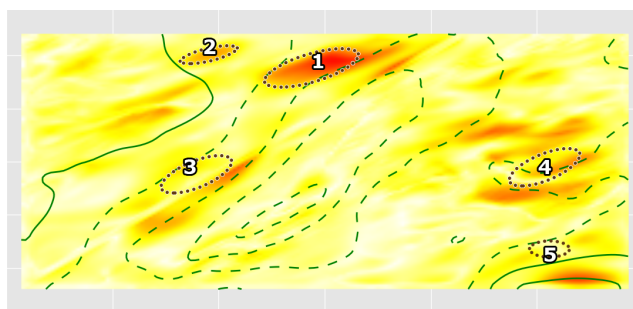
4.10(b): Rotational-free magnitude standard deviation
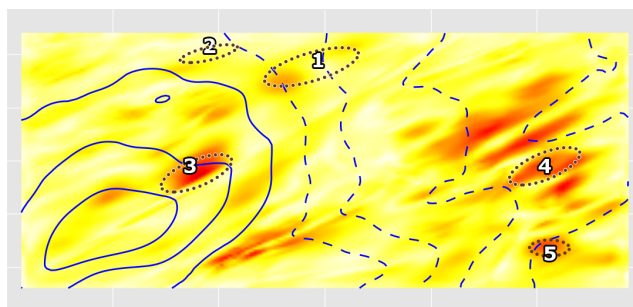
4.10(c): Divergence-free magnitude standard deviation

4.10(d): Harmonic magnitude standard deviation

Figure 4.10: Standard deviation of vector magnitude in the rotational-free, divergence-free, and harmonic component ensembles.

the harmonic and the rotational-free ensembles obtained by the NHHD present some noisy regions close to the boundary, as shown in Figure 4.11, they do not contribute significantly to the magnitude uncertainty of the original ensemble. Finally, we also would like to mention that for ensembles having quite similar elements, the proposed method generated component fields with low uncertainty because of the robustness of the NHHD method.
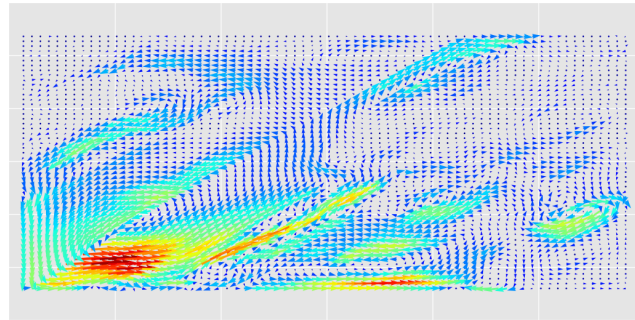
4.11(a): Vector Field mean: $\mu_{\mathcal{E}}$
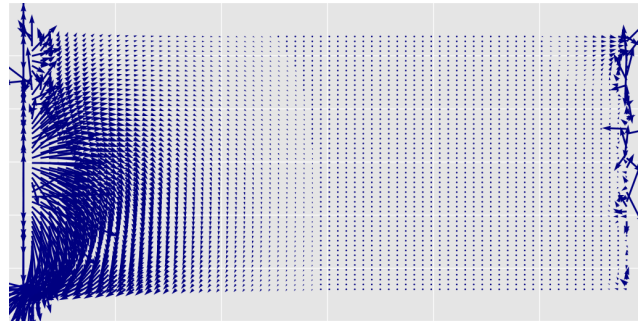
4.11(b): Rotational-free component mean: $\mu_{\mathcal{E}_C}$
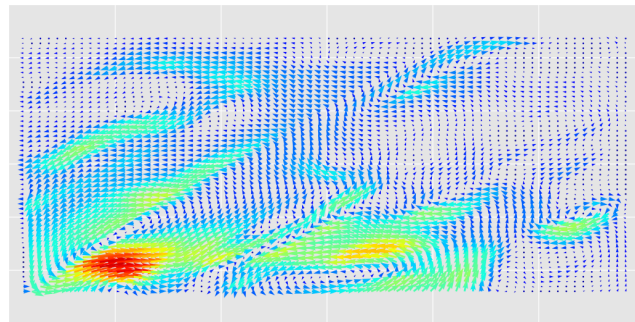
4.11(c): Divergence-free component mean: $\mu_{\mathcal{E}_D}$
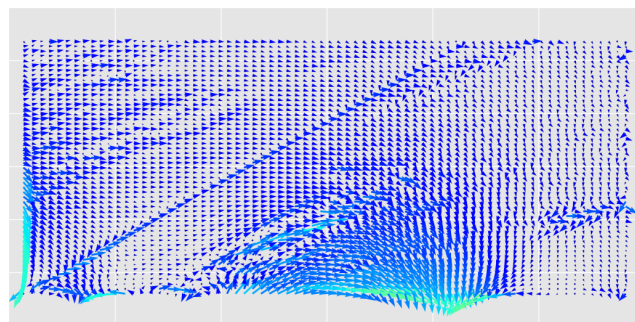
4.11(d): Harmonic component mean: $\mu_{\mathcal{E}_H}$
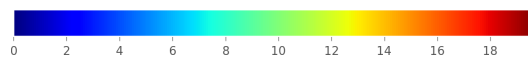
Figure 4.11: Mean vector fields obtained through NHHD for the divergence-free ensemble derived from the wind forecast data ensemble. The color represents the vectors magnitude, while the color scale is common for all NHHD components. The size of the rendered vectors is individually scaled in each subfigure for better visualization.
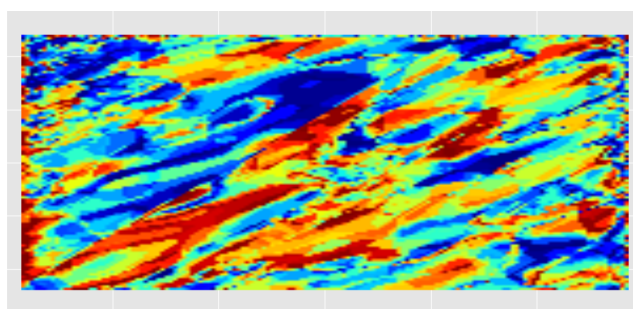
4.12(a): Rotational-free / vector-field correlation
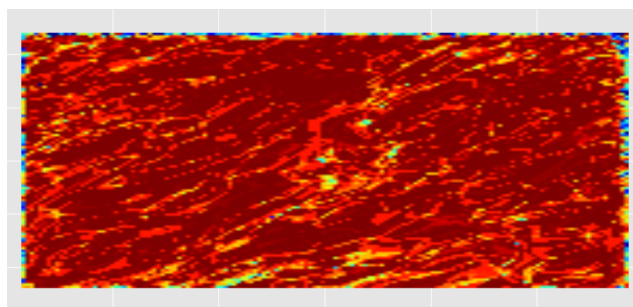
4.12(b): Divergence-free / vector-field correlation
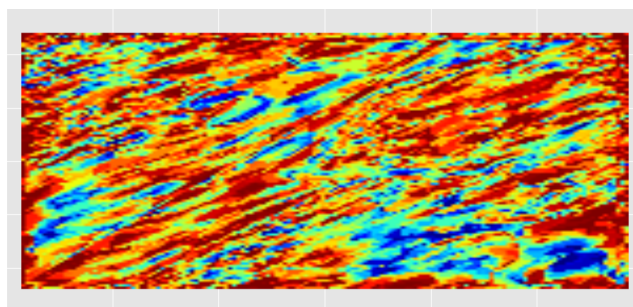
4.12(c): Harmonic / vector-field correlation
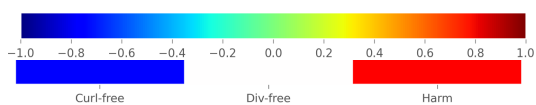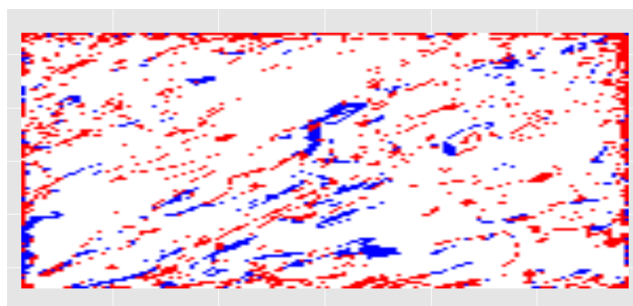

4.12(d): Maximum correlation
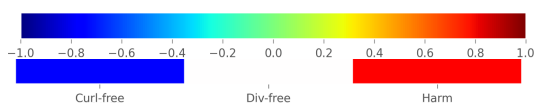
Figure 4.12: Correlation between the pure divergence-free vector field and each of its NHHD derived component ensemble.
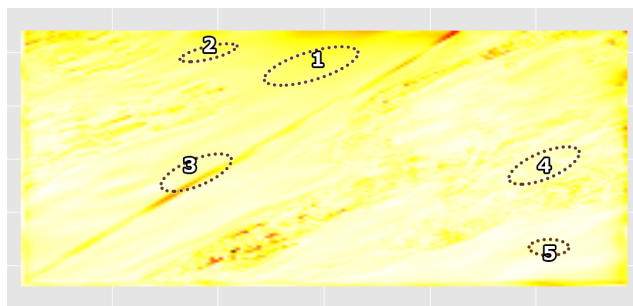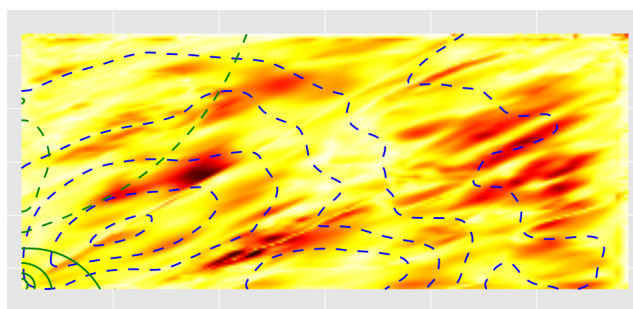
4.13(a): Vector field magnitude standard deviation
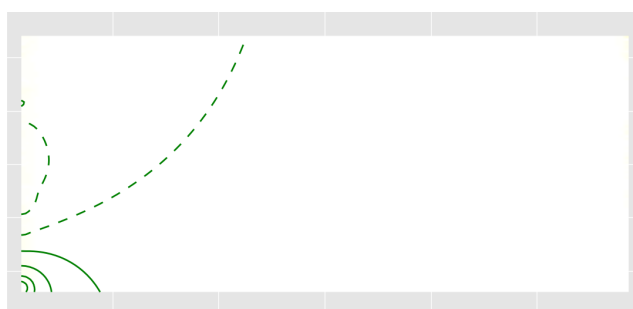
4.13(b): Rotational-free magnitude standard deviation
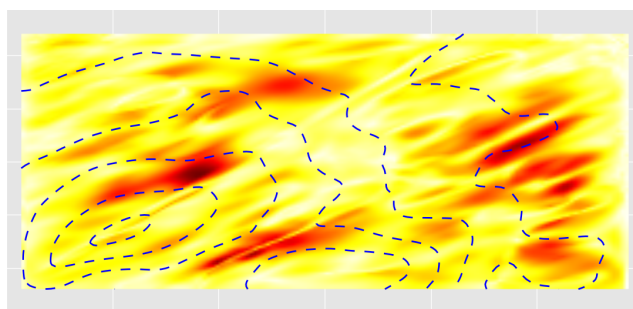
4.13(c): Divergence-free magnitude standard deviation
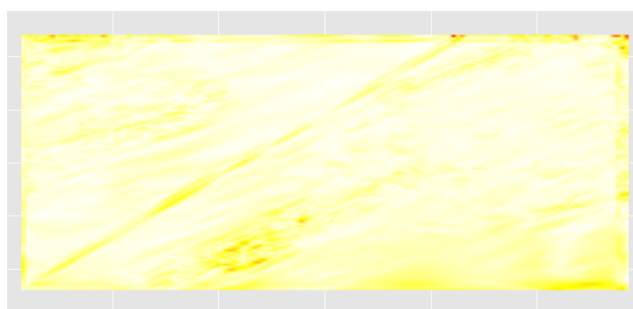
4.13(d): Harmonic magnitude standard deviation
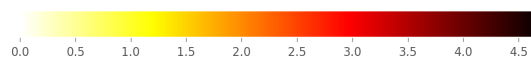
Figure 4.13: Standard deviation of vector magnitude in the derived rotational-free, divergence-free and harmonic component ensembles obtained from the NHHD of a pure divergence-free ensemble.

5
Stochastic Generation of Vector Fields

In this Chapter, we proceed with the application of the Helmholtz-Hodge Decomposition in the uncertainty analysis of vector fields. More specifically, we present a set of methods to stochastically generate new realizations of a vector field given an unique vector field realization.

5.1
Motivation

Vector fields comprehend an important mathematical object in a variety of problems related to *Scientific Computing* applications. These applications include, for example, fluid flow simulation (Anderson & Wendt, 1995), analysis of MRI data for medical prognosis (Tong et al., 2003) and weather prediction (Luo et al., 2012), just to cite a few. Often, those vector fields are obtained through computer simulations, which may require expensive numerical computations (Anderson & Wendt, 1995).

As previously described in Chapter 4.1, the Helmholtz Hodge Decomposition presents a large range of applicability in the analysis of vector fields. We believe that proposing a method to derive new realizations from a given vector field using the HHD can be helpful because:

– It can be less time consuming than the required techniques to simulate vector fields from scratch in some areas of Scientific Computing.

– Several realizations could provide the same set of benefits as the multi-method ensemble presented in Chapter 4.1.

– As we are going to show, the derived realizations provide some variability when compared to the given vector field, while not being far from it.

5.2
Method 1: Bootstrap Based Stochastic Simulation

In this Section we propose an algorithm to stochastically simulate vector field realizations based on a given gridded 2D vector field $\mathbf{V}$, which will from now on be called the *training data*. Such algorithm is based on the Helmholtz-Hodge Decomposition and on the non-parametric Bootstrap method (Efron, 1979). The proposed algorithm aims to physically mimic $\mathbf{V}$ and appropriately cover the space of uncertainty. More precisely, our algorithm first use the NHHD of $\mathbf{V}$ to obtain its rotational-free and divergence-free potentials components. With such potentials in hand, we perform a Bootstrap-like approach to generate R other realizations of these potentials. In the sequence, we differentiate them. Finally, we add the generated components to the original harmonic component to generate these R vector field realizations.

5.2.1
The Bootstrap Method

The *Bootstrap method* is a statistical method based on resampling with replacement. It is commonly applied to measure the accuracy of statistical estimators (Wasserman, 2004). In general, such accuracy could be defined in terms of bias, variance, confidence intervals, prediction error or some other measure.

The Bootstrap method is based on the notion of a *bootstrap sample*. To better understand it, let $\hat{F}$ be an empirical distribution, with probability $1/n$ on each of the n observed values x_i, with $i \in \{1, 2, \cdots, n\}$. Then, a bootstrap sample is defined to be a random sample of size n drawn from $\hat{F}$ with replacement, say $\mathbf{x}^* = (x_1^*, x_2^*, \cdots, x_n^*)$ (Efron, 1979). The star notation indicates that $\mathbf{x}^*$ is not the actual data set $\mathbf{x}$, but a randomized, or *resampled*, version of $\mathbf{x}$. More details about this technique can be found in (Wasserman, 2004).

With this concept in mind, assume that $T_n = g(x_1, x_2, \cdots x_n)$ is a statistic of the data set $\{x_1, \cdots, x_n\}$. To compute the variance of T_n, denoted by $\mathbb{V}_F(T_n)$, it would be necessary to know the distribution F of the data. Often, however, that is unknown. The Bootstrap technique estimates $\mathbb{V}_F(T_n)$ by the use of stochastic simulations, where the unknown distribution F is approximated by a distribution named $\hat{F}$. Then, an approximation of $\mathbb{V}_F(T_n)$ is computed as $\mathbb{V}_{\hat{F}}(T_n)$. Generating several bootstrap samples, it is now possible to approximate the distribution of T_n by evaluating $T_n^* = g(x_1^*, \cdots, x_n^*)$. Using this distribution, we can finally compute the variance $\mathbb{V}_{\hat{F}}(T_n)$ according to the following formula:

$$\mathbb{V}_{\hat{F}}(\mathrm{T}_n) = \frac{1}{B} \sum_{i=1}^{B} \left(\mathrm{T}_i^* - \frac{1}{B} \sum_{b=1}^{B} \mathrm{T}_{n,b}^* \right)^2, \qquad (5\text{-}1)$$

where T_i^*, $i = 1, \ldots, B$, represents the statistics computed at the ith bootstrap sample.

Figure 5.1 shows the Bootstrap sampling procedure. As can be seen, it takes a sample data as input and, as output, a new data obtained through resampling with replacement from the original sample is generated.
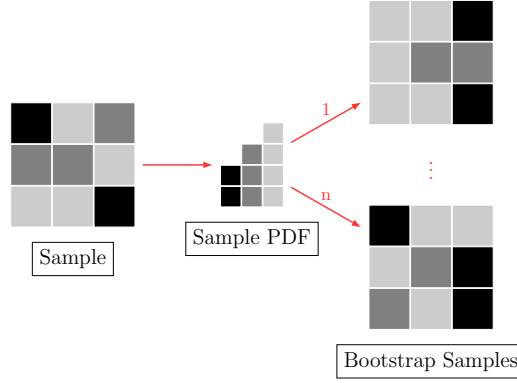
Figure 5.1: The Bootstrap Method.

5.2.2
Proposed Method

Once again, consider a discrete sampling of a 2-dimensional domain on a Cartesian grid structure $\mathbb{S}_{m,n} = \{\mathbf{x}_{i,j} \in \mathbb{R}^2 : 1 \leq i \leq m, 1 \leq j \leq n\}$. Also, suppose that a 2D vector field $\mathbf{V}$ is given, i.e., to each spatial point in $\mathbb{S}_{m,n}$ there is a 2D vector associated.

The main goal of our method is to generate a new vector field that has similar features to the training one. Precisely, the new vector field comprehends a structural perturbation of the training vector field.

5.2.2.0
Method Overview

The first step in our method is to compute the NHHD of the training data $\mathbf{V}$. So, at each point $\mathbf{x}_{i,j} \in \mathbb{S}_{m,n}$ we have the following equality:

$$\mathbf{V}(\mathbf{x}_{i,j}) = \nabla\varphi(\mathbf{x}_{i,j}) + (\nabla \cdot J)\psi(\mathbf{x}_{i,j}) + \mathbf{h}(\mathbf{x}_{i,j}).$$

With the NHHD components of the given training data $\mathbf{V}$ in hand, we stochastically generate other R 2D vector fields based on $\mathbf{V}$. To obtain each realization, we firstly perturb the divergence-free φ^* and rotational-free ψ^*

scalar potentials at b points $\mathbf{x}_{i,j} \in \mathbb{S}_{m,n}$, using a Bootstrap-like technique. From these perturbed scalar potentials, we then compute the corresponding rotational-free and divergent-free terms from their partial derivatives. We add these two terms to the original harmonic term $\mathbf{h}^*$ in order to finally create a vector field realization.

The number b of blocks in which to perform the Bootstrap is defined by a Poisson Distribution (Wasserman, 2004) with rate λ. This rate represents the mean number of blocks that are going to be perturbed. The greater the λ the higher the variability induced in the samples.

When performing the Bootstrap-like approach, as we are calculating derivatives through discrete approximations, one may think of defining two kernels of the same size of the central difference kernels, i.e., 3×1 and 1×3 for the x and y directions, respectively. Then, for each $\mathbf{x}_{i,j}$, two bootstrap samples would be created, one to calculate the x derivative and another to calculate the y derivative. Such an approach, however, is incapable of preserving the orientation of the vector field. As an example, regarding the vector field and its NHHD components presented in Figure 3.1, Figure 5.2 presents a realization generated using this approach, as well as the mean vector field obtained by repeating this technique to generate 100 realizations.

5.2(a): Example of a single sample   5.2(b): Mean of 100 samples
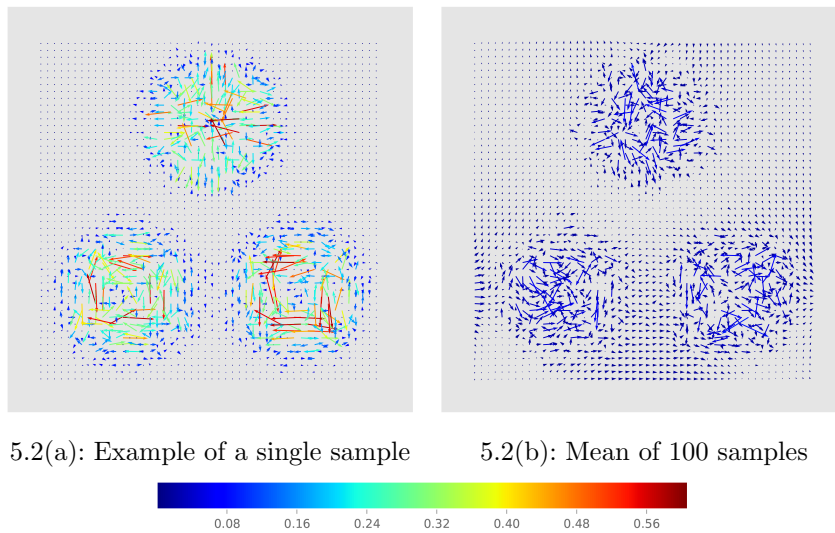
Figure 5.2: Example of vector fields obtained using a kernel for the x derivative and another for the y derivative. The color scale matches the one presented in Figure 3.1 for comparison purposes.

As can be noted, the orientation of the vector field was completely lost.

Given that we are dealing with vector fields, we adopted an strategy to preserve their structure during the resampling step. Such strategy is based on a kernel proposed by Fu et al. (2012) and depicted in Figure 5.3. This kernel

explores the directional coherence of the contours that pass through the central pixel. As can be seen, the kernel divides a $n \times n$ block in 8 subgroups. When performing the Bootstrap-based technique, each of these regions is resampled with replacement separately to obtain a bootstrap sample around the central pixel. The size of the kernel presented in Figure 5.3 is 5×5. The bigger the kernel, the higher the variation of the bootstrap samples in relation to the input data.
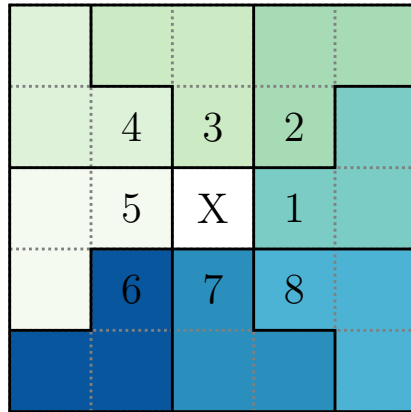
Figure 5.3: Kernel divided into regions to preserve the vector field orientation.

Once again, taking as a realization the vector field depicted in Figure 3.1, one can perceive, through Figure 5.4, that the adopted kernel is capable of preserving the orientation of the vector field used as input for the Bootstrap method. More than that, in regions in which the potentials are practically constant, no noise is added to the vector field samples.
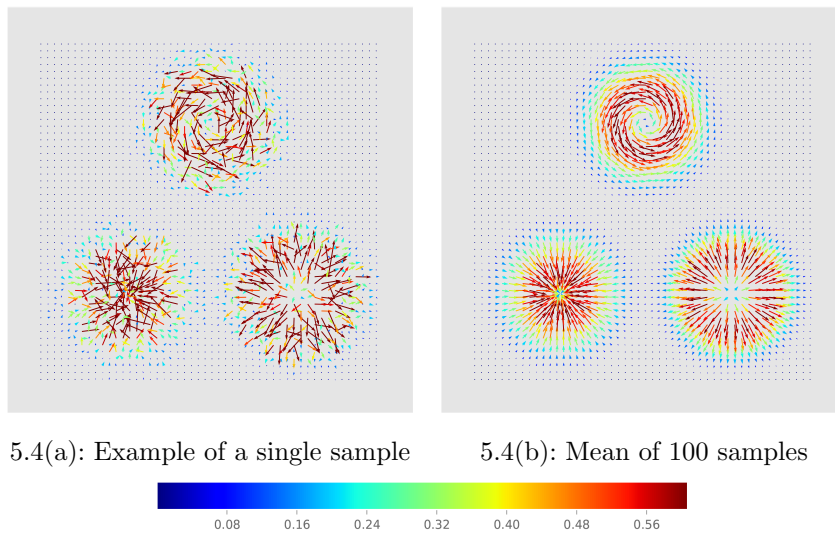
5.4(a): Example of a single sample   5.4(b): Mean of 100 samples

Figure 5.4: Example of vector fields obtained using a kernel divided in regions to preserve the vector field orientation. The color scale matches the one presented in Figure 3.1 for comparison purposes.

With this knowledge, we can now specify that, in this work, λ is defined as a percentage of the training data size divided by the kernel size.

At last, a smoothing step is performed using a *Gaussian Filter* (Gonzalez & Woods, 2006), which standard deviation (σ) can be parameterized, for both x and y dimensions.

5.2.2.0
The Algorithm

We implemented the proposed method according to the pseudocode described in Algorithm 1. This pseudocode generates a stochastic realization $\mathbf{R}^*$ based on the NHHD components of a training data $\mathbf{V}$.

The method has as input the following list of variables:

– the scalar potentials φ^*, ψ^* and the vector field $\mathbf{h}^*$ obtained by the NHHD of the training data $\mathbf{V}$;

– the kernel K of size $l \times l$ used to perform the resampling with replacement on the potentials;

– the number b of blocks in which we will perform the Bootstrap.

---

input : φ^*, ψ^*, $\mathbf{h}^*$, K, b
output: $\mathbf{R}^*$, a vector field realization

1 $\varphi^*_{boot} \leftarrow \varphi^*$;
2 $\psi^*_{boot} \leftarrow \psi^*$;
   /* Derive b indexes in the 2D grid   */ ;
3 $x \leftarrow$ `randInt`$(1, m, b)$;
4 $y \leftarrow$ `randInt`$(1, n, b)$;
   /* For each derived index   */ ;
5 **for** $k \leftarrow 1$ **to** b **do**
6 $\quad$ $i \leftarrow x[i]$;
7 $\quad$ $j \leftarrow y[i]$;
8 $\quad$ $boot\_indices \leftarrow$ `local_bootstrap`(K);
9 $\quad$ $\varphi^*_{boot}(i,j) \leftarrow \hat{\mathrm{F}}\varphi(boot\_indices)$;
10 $\quad$ $\psi^*_{boot}(i,j) \leftarrow \hat{\mathrm{F}}\psi(boot\_indices)$;
11 **end**
12 $\varphi^*_{boot} \leftarrow$ `smooth`(φ^*_{boot});
13 $\psi^*_{boot} \leftarrow$ `smooth`(ψ^*_{boot});
14 $\nabla\varphi^*_{\mathbf{R}} \leftarrow$ `divergent`(φ^*_{boot});
15 $(\nabla \cdot J)\psi^*_{\mathbf{R}} \leftarrow$ `curl`(ψ^*_{boot});
16 $\mathbf{R}^*(\mathbf{x}_{i,j}) \leftarrow \nabla\varphi^*_{\mathbf{R}}(\mathbf{x}_{i,j}) + (\nabla \cdot J)\psi^*_{\mathbf{R}}(\mathbf{x}_{i,j}) + \mathbf{h}^*$;

Algorithm 1: Generating a realization $\mathbf{R}^*$ based on the NHHD components of a training data $\mathbf{V}$ using a Bootstrap-based approach.

The input b represents the number of indexes in the 2D grid and their locations are generated through an Uniform Distribution (Wasserman, 2004) (lines 3 and 4). In other words, the generated indexes respect the dimensions of the training data. These indexes represent positions in the scalar potentials of $\mathbf{V}$ that are going to be perturbed using a Bootstrap-like approach.

Then, for each one of the b indexes pairs, say $\mathbf{x}_{i,j}$, we perform a local Bootstrap (line 8) based on the input kernel K, which results in a new organization of indexes that are going to be used to perturb the neighborhood of $\mathbf{x}_{i,j}$. In other words, those indexes are used to assign new values to $\star_{boot}^{*}(i,j)$, where $\star$ can be either φ or ψ.

In the following, we perform a smoothing step on $\star_{boot}^{*}$, i.e., we obtain a smoothed version of $\hat{\mathrm{F}}_{\star}$. The smoothing step is required because a small change in the potentials can lead to a significant change in the vector field, once this is obtained deriving these potentials. In this work, we used σ equal to 2 pixels in the smoothing step. This value was chosen by experimentation.

After these steps, we can now differentiate the new scalar potentials to obtain new realizations for the divergence-free (line 14) and rotational-free (line 15) components of $\mathbf{V}$. Finally, a new vector field realization is obtained summing these components with the original harmonic component of $\mathbf{V}$ (line 16), following Equation 3-10.

Repeating this procedure R times, we will then have a set of R stochastic realizations of vector fields obtained through the original NHHD components of $\mathbf{V}$.

5.2.3
Results and Discussion

To verify the results that the proposed method can achieve, we make use of a 2D vector field ensemble comprehended by seven multi-method wind forecast realizations $\mathcal{E}$, previously presented in Chapter 4.

As a first step, we apply the NHHD on each realization $\mathbf{R}$ in $\mathcal{E}$ to derive its divergence-free, rotational-free and harmonic components. Those are then perturbed as explained above to obtain new realizations based on the given vector field.

In summary, we repeat the described procedure, i.e., we apply Algorithm 1 on each realization $\mathbf{R}$ in $\mathcal{E}$, 100 times to generate other 100 new realizations.

5.2.3.0
Similarity measure and MDS projection

To provide a way of visually encode the similarity between the vector fields, we make use of the *Multidimensional Scaling* (MDS) (Kruskal, 1964) technique for dimensionality reduction to visualize high-dimensional data in a 2-dimensional space. The MDS method aims to provide insight in the underlying structure and relations between patterns by providing a geometrical representation of their similarities (Honarkhah & Caers, 2010). Mathematically speaking, the MDS translates a dissimilarity matrix into a configuration of points in a n-D Euclidean space.

For two vector fields $\mathbf{A}$ and $\mathbf{B}$, we adopted the following similarity measure, known as the *Cosine Similarity*:

$$similarity_{A,B} = \cos\theta = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \cdot \|\mathbf{B}\|} \tag{5-2}$$

Such a measure states how related two vector fields are given their angles. For similar vectors the similarity coefficient will be close to 1, for opposite vectors, such coefficient will be close to -1. For unrelated vectors, on the other hand, this coefficient will be around 0.

To take into account both the magnitude and orientation of the vector fields $\mathbf{A}$ and $\mathbf{B}$ in the cosine similarity computation, we perform the following transformation.

Firstly, for a vector field $\mathbf{V}$ of dimensions $m \times n$, we unroll it from a 2-dimensional vector field to a 1-dimensional vector. Then, we generate a new vector $\mathbf{V}^* = (\mathrm{V}_1^*, \mathrm{V}_2^*)$ based on $\mathbf{V}$ such as:

$$
\begin{aligned}
\mathrm{V}_1^* &= \mathrm{atan2}(\mathrm{V}_2, \mathrm{V}_1) \\
\mathrm{V}_2^* &= \|\mathrm{V}\| \\
\mathrm{V}^* &= (\mathrm{V}_1, \mathrm{V}_2)
\end{aligned}
$$

For an ensemble $\mathcal{E}$, after this step, we have a new set $\mathcal{E}^*$ in hand. All vectors in $\mathcal{E}^*$ are normalized as follows:

$$
\begin{aligned}
\mathrm{V}_1^* &= \frac{\mathrm{V}_1^*}{\pi/2} \\
\mathrm{V}_2^* &= \frac{\mathrm{V}_2^*}{max(\mathrm{V}_2^* \in \mathcal{E}^*)}
\end{aligned}
$$

After this transformation, we apply the similarity measure for each pair of realizations in $\mathcal{E}^*$.

Figure 5.5 presents the MDS for the wind forecast ensemble and its mean vector μ, after applying the transformation described before. The mean vector field μ, as well as its potentials, are depict in Figures 5.7 and 5.6, respectively.
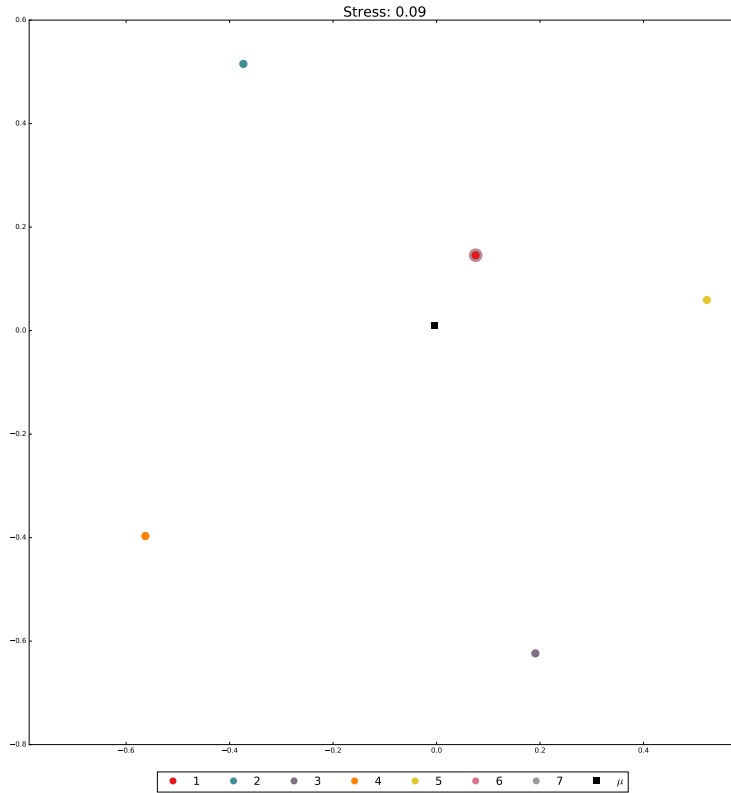
Figure 5.5: MDS visualization for the original ensemble $\mathcal{E}$. Colors represent each realization in $\mathcal{E}$. The black square represent the mean vector of $\mathcal{E}$.

5.2.3.0
Coverage Test

It is relevant to verify whether we can generate a set of realizations that covers the given ensemble set or not. This might state if, from a single realization, it is possible to obtain certain scenarios that could be derived through another simulation process (possible more costly). To do this, we first tried different values for the λ parameter given different Bootstrap kernel sizes to generate 100 new samples from the mean vector field μ. They ranged from 30% to 90% and from 5×5 to 17×17, respectively.

We achieved the best coverage using a λ value of 90% and a kernel of 19×19, as can be seen in Figure 5.8.

For this best scenario, Figures 5.9 and 5.10 show the standard deviation of the potentials and NHHD components obtained through the proposed approach.

From Figure 5.9, we can observe that the potentials obtained trhough this method presented some variation, while, from Figure 5.10, we can note that the higher standard deviation are concentrade in areas where the original vector field presents high magnitude. It is worth noting that, once the harmonic component remains the same as the original vector field, its variation is zero.
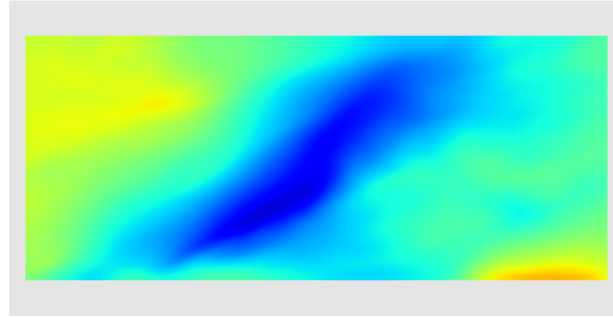
5.6(a): Rotational-free potential μ_ψ
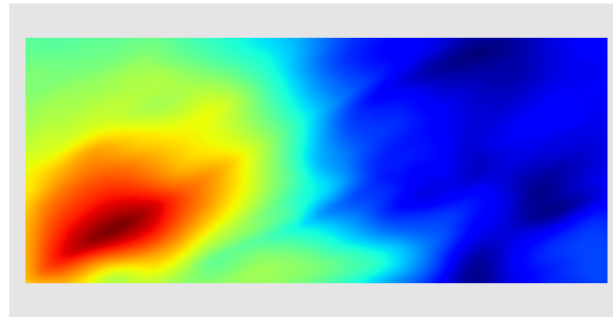
5.6(b): Divergence-free potential μ_ψ
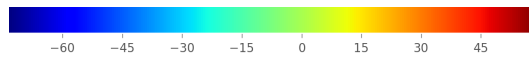
Figure 5.6: Rotational-free and divergence-free potentials of μ.

Given that, Figure 5.11 present the MDS for each vector field in the original set $\mathcal{E}$ and a new set of realizations derived from it using a kernel size and λ as specified before. Markers of same color belongs to the same set, i.e, were generated based on a common realization. Circle markers represent each realization of the set $\mathcal{E}$. Cross markers represent new realizations, and diamond and square markers show both the closest and farthest simulation, in that order, given a base realization – Table 5.1 depict these simulations for each realization in $\mathcal{E}$.

Through this image, we can see that, for each realization $s \in (1, \cdots, 7)$, the resulting set of realizations present some variability in relation to the original vector field used as base for the stochastic simulation method.

Putting all these simulations together, we have the result presented in Figure 5.12. From this image, we can notice that the original set $\mathcal{E}$ is completely surrounded by the new realizations.
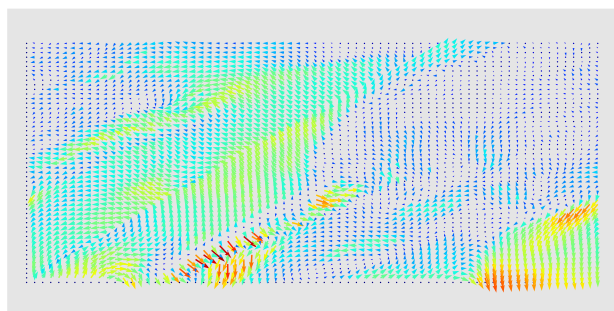
5.7(a): Vector Field
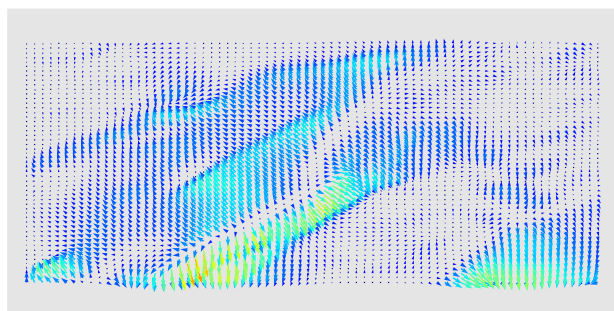
5.7(b): Rotational-free
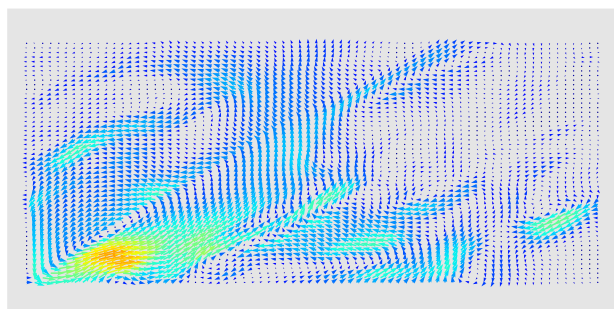
5.7(c): Divergence-free
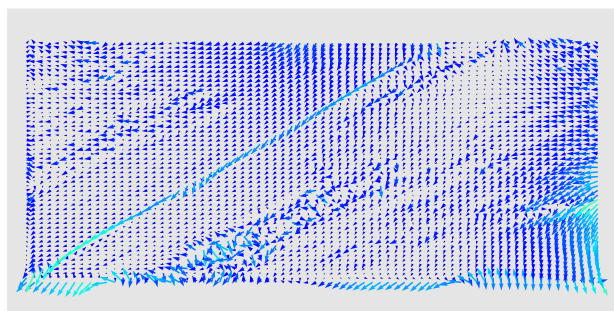
5.7(d): Harmonic

Figure 5.7: Mean vector field μ.
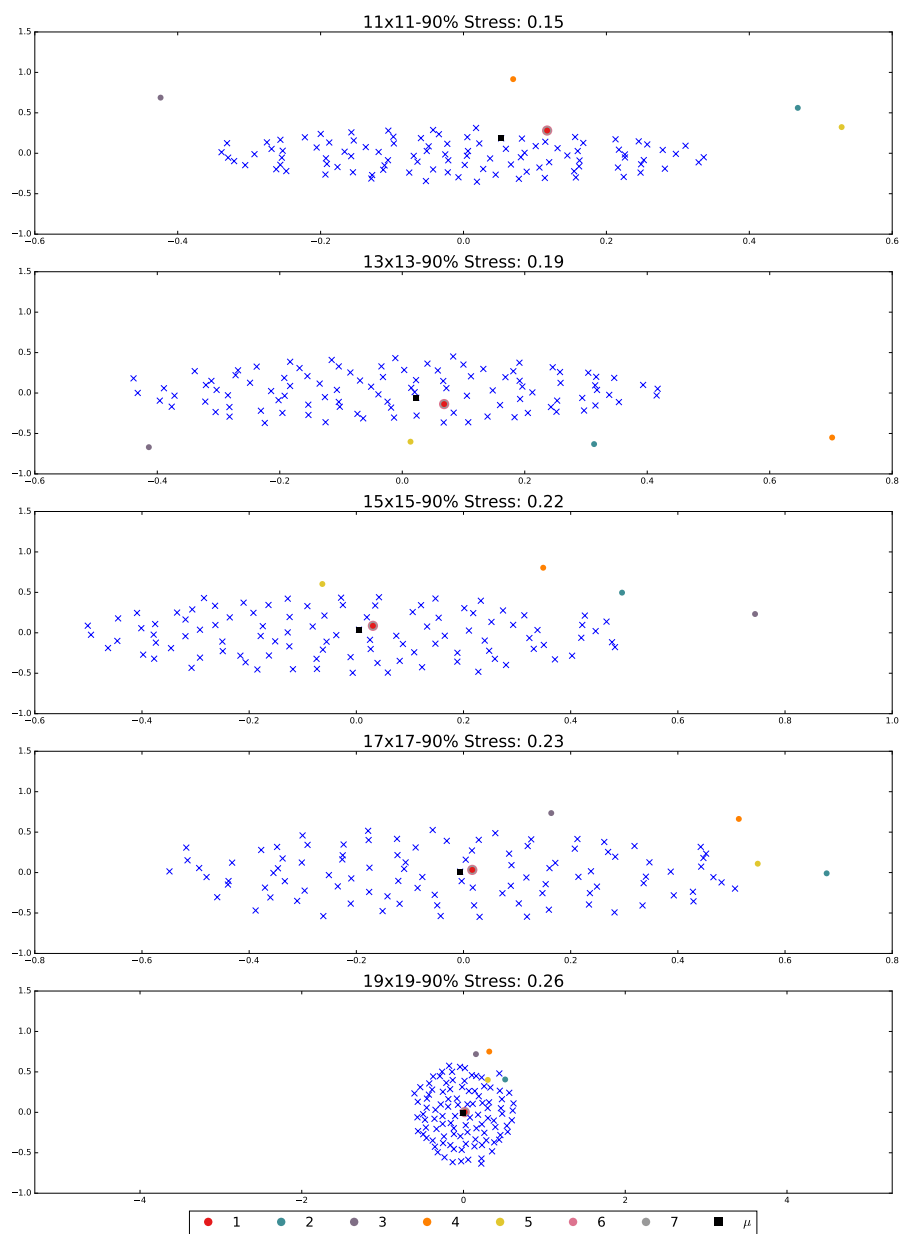
Figure 5.8: MDS visualization between $\mathcal{E}$ and new realizations, obtained using different kernel sizes.

Table 5.1: Original realization and its closest and farthest realizations using the Bootstrap-like technique.

Sample	Closest Simulation	Farthest Simulation
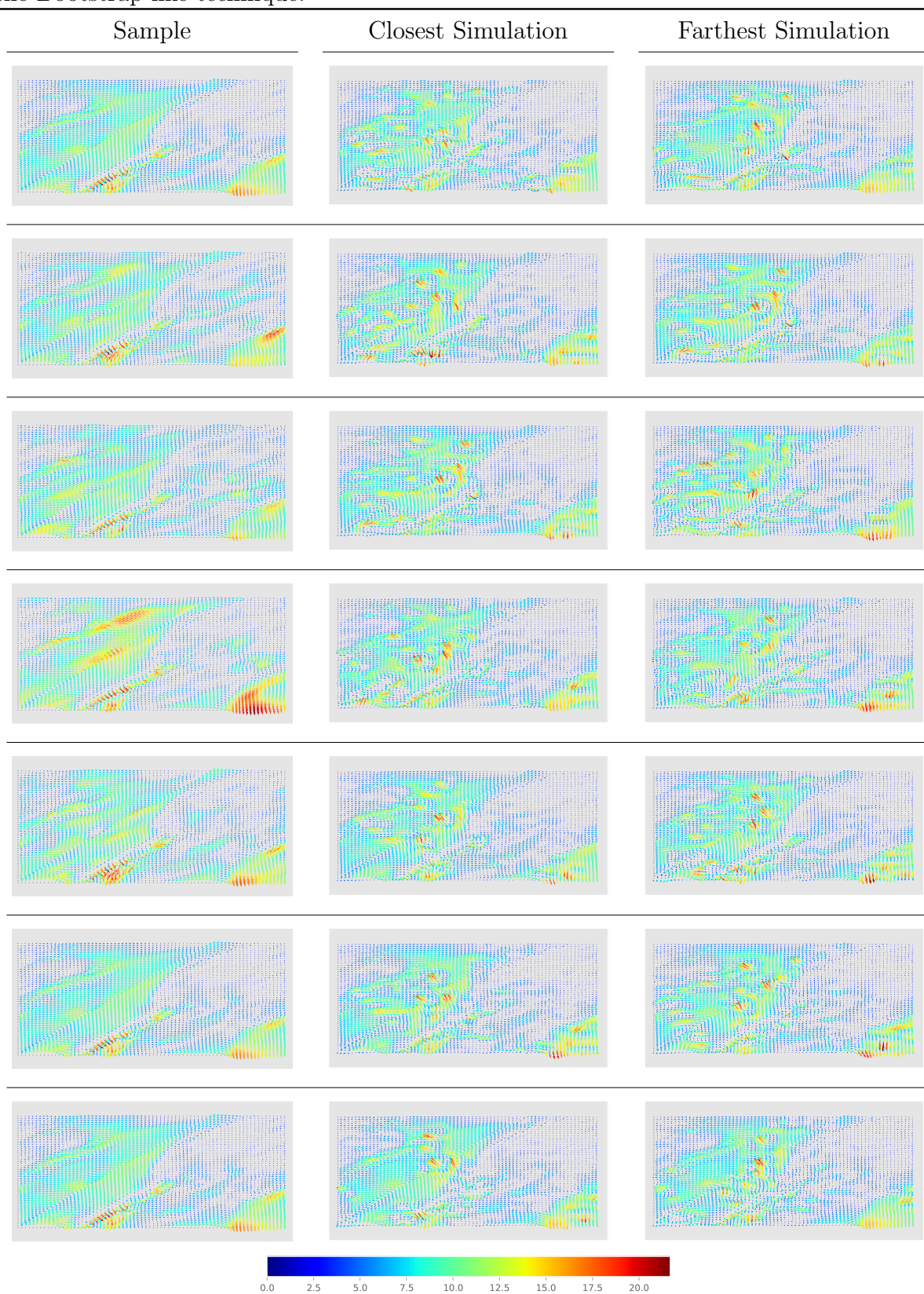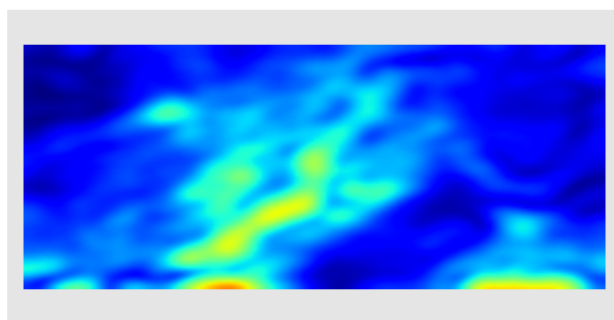
5.9(a): Rotational-free potential standard deviation σ_φ
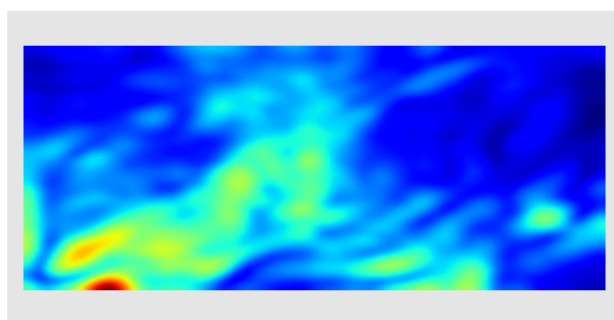
5.9(b): Divergence-free potential standard deviation σ_ψ
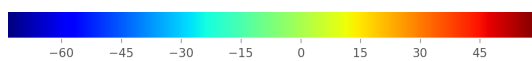
Figure 5.9: Standard deviation of the rotational-free and divergence-free potentials obtained through the Bootstrap-like approach.

5.10(a): Vector-field magnitude standard deviation

5.10(b): Rotational-free magnitude standard deviation

5.10(c): Divergence-free magnitude standard deviation
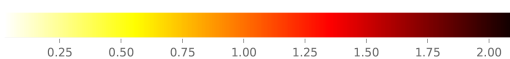
Figure 5.10: Standard deviation of the magnitude of the rotational-free and divergence-free components using the Bootstrap-like approach.
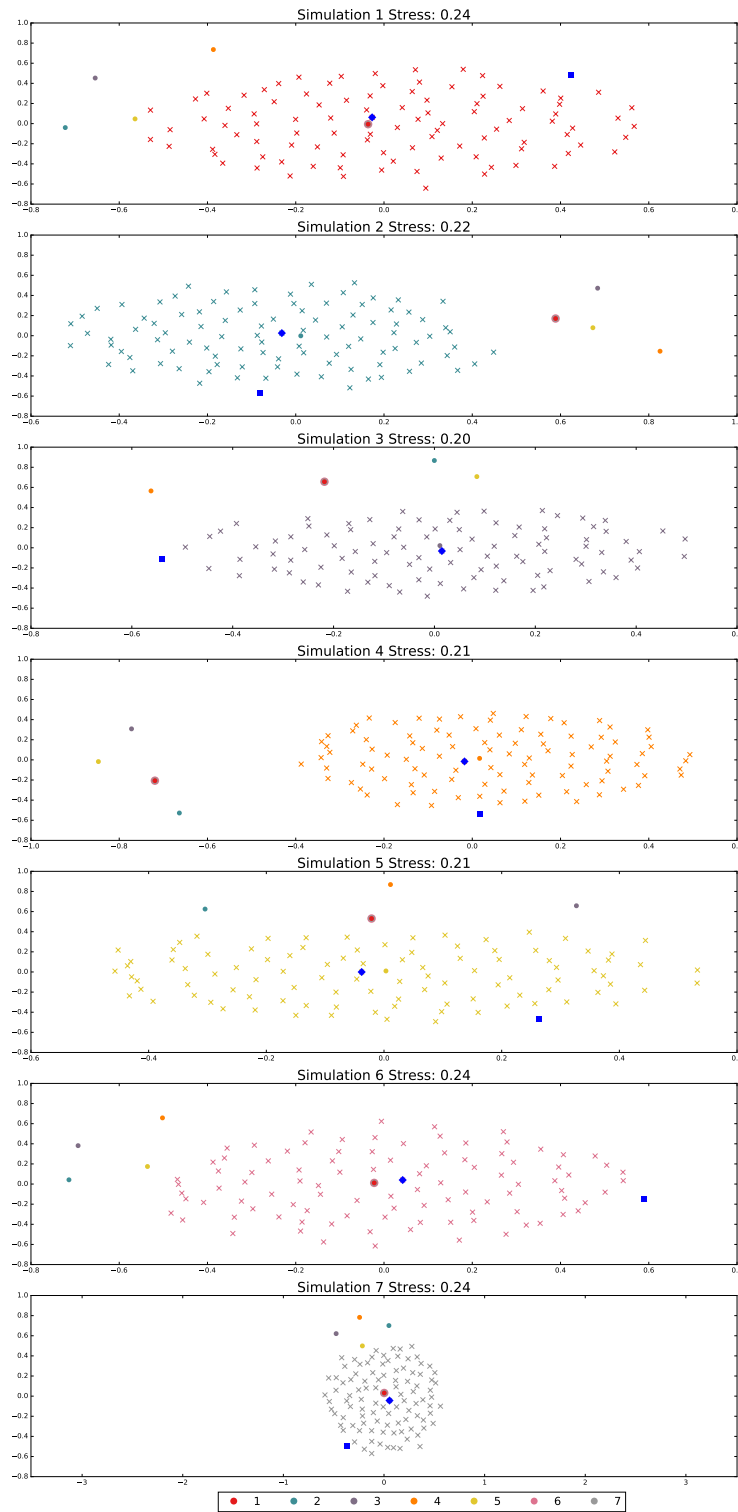
Figure 5.11: MDS visualization between each set of new realizations and the original ensemble $\mathcal{E}$. Colors represent each realization of the set $\mathcal{E}$. Circular markers represent each realization in $\mathcal{E}$. Cross markers represent, for each $\mathbf{V}$ in $\mathcal{E}$, the new realizations derived from $\mathbf{V}$, both represented with the same color.
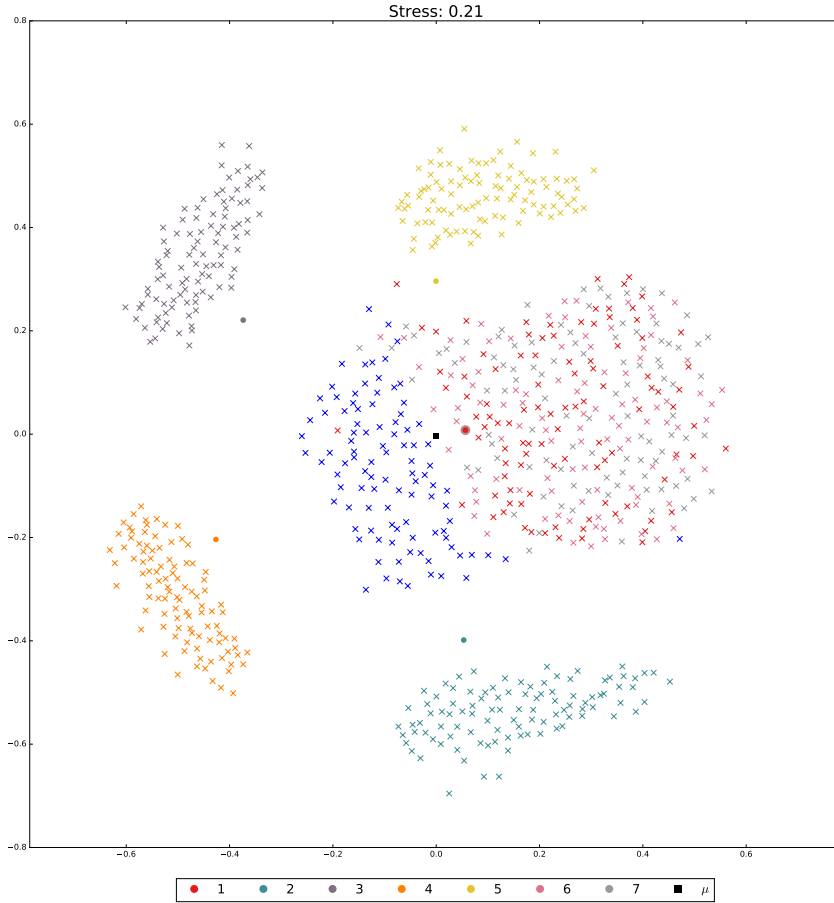
Figure 5.12: MDS visualization between each new set of realizations and the original ensemble $\mathcal{E}$. Colors represent each realization of the set $\mathcal{E}$. Circular markers represent the realizations in $\mathcal{E}$. Cross markers represent new realizations derived from the one presented with a circular marker of the same color.

5.2.4
Applications

In this Section, we present a quantification approach to the algorithm uncertainty related to the application of the curl and divergence discrete differential operators, described in Equation 3-18.

5.2.4.0
Navier-Stokes

Consider the vector field presented in Figure 5.13. This field is defined over a grid of 64×64, with its minimum and maximum as 0.007812 and 0.992188, respectively, in both x and y directions. This field is a result of a *Navier-Stokes* simulation (Chorin, 1968), which aims to describe the motion of viscous fluid flows. Such kind of simulation can be used to model a varied set of physics phenomena, ranging from waves simulation (Abadiea et al., 2010)

to image and video inpainting (Bertalmio et al., 2001). As can be seen, the divergence-free component defines such field (we may consider the rotational-free and harmonic components as noise).

5.13(a): Vector Field   5.13(b): Rotational-free

5.13(c): Divergence-free   5.13(d): Harmonic
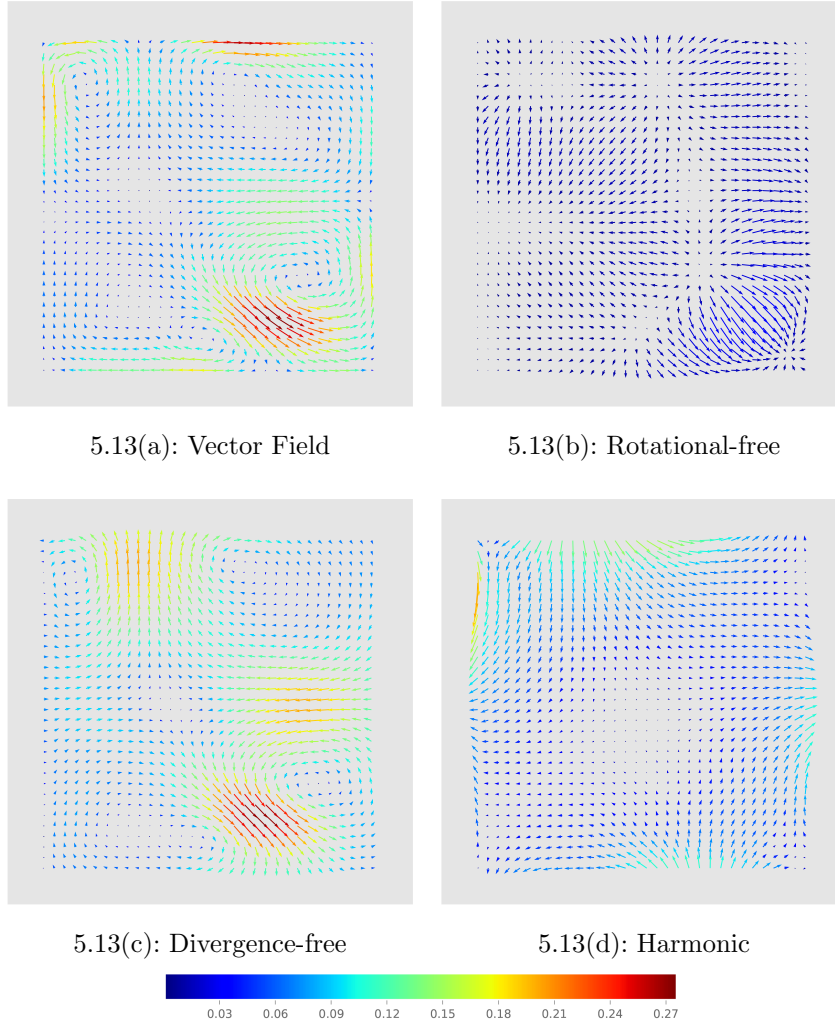
Figure 5.13: Navier-Stokes simulation and its NHHD components.

After generating 100 new realizations through the procedure presented in Algorithm 1, using a kernel of 5×5, we have a set of realizations $\mathcal{E}$. Figure 5.14 shows the MDS for this set. As can be seen, the training data is surrounded by the new ones.

Figure 5.15 shows the closest and farthest simulation derived from the original vector field. They are represented using the same magnitude scale as the original field (Figure 5.13).

With this set in hand, it is now possible to quantify the uncertainty related to the curl operator, which is obtained using partial derivatives. In other words, we can measure the uncertainty related to the kernel used to obtain such attribute. To do so, for each new realization $\mathbf{R} \in \mathcal{E}$, we obtain the
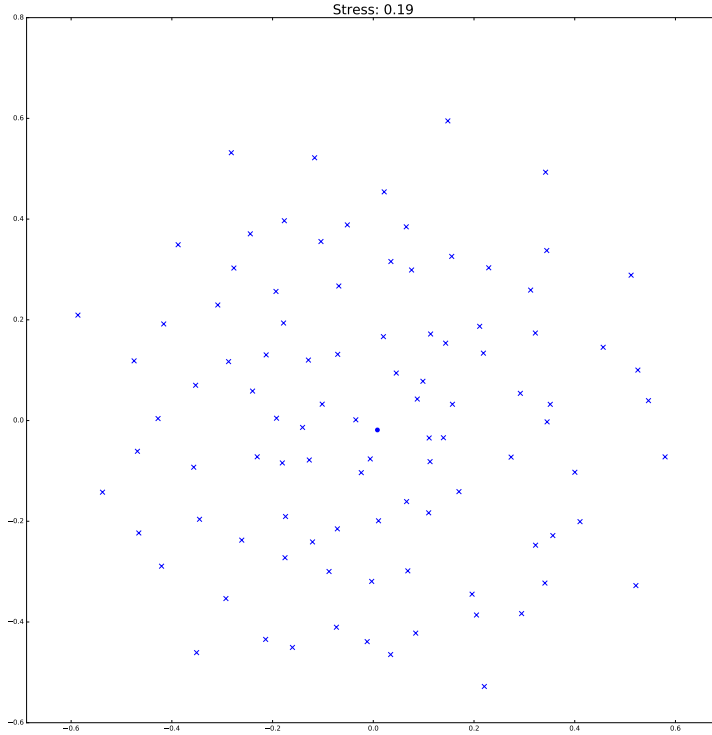
Figure 5.14: MDS visualization between the generated realizations set and the Navier-Stokes vector field.
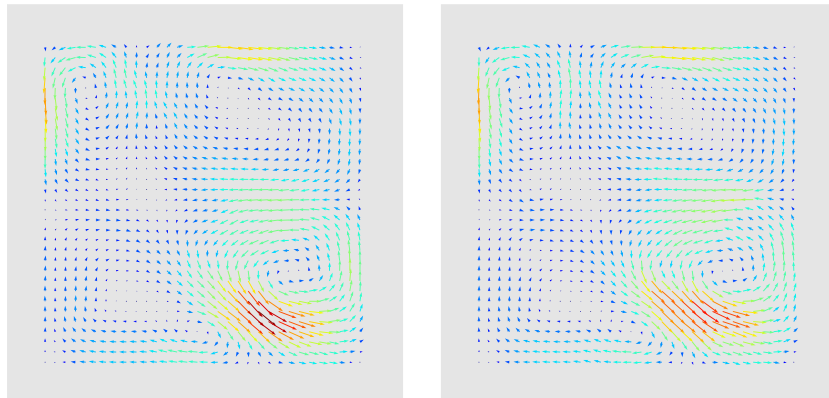
5.15(a): Closest Simulation

5.15(b): Farthest Simulation

Figure 5.15: Closest and farthest realization of the Navier-Stokes vector field.

curl of $\mathbf{R}$. We do the same for the original sample $\mathbf{V}$. To derive the uncertainty of the curl operator, we then compute its root mean squared error (RMSE).

In statistics, the mean squared error, MSE, of an estimator is a way to measure the difference between values implied by an estimator and the true values of its target parameter (Wackerly et al., 2008).

For instance, being $\hat{\mathrm{T}}$ the curl of $\mathbf{V}$ and T_i^*, $i = 1, \ldots, 100$ the curl of each one of the generated samples, the MSE of the predictor $\hat{\mathrm{T}}$ is defined as:

$$\text{MSE}(\hat{\text{T}}) = \frac{1}{100} \sum_{i=1}^{100} (\hat{\text{T}} - \text{T}_i^*)^2 \tag{5-3}$$

The RMSE is given as the square root of the MSE, i.e., $\text{RMSE} = \sqrt{\text{MSE}}$.

Figure 5.16 presents the RMSE of the curl given the generated realizations.

5.16(a): Curl of **V**	5.16(b): RMSE of the curl
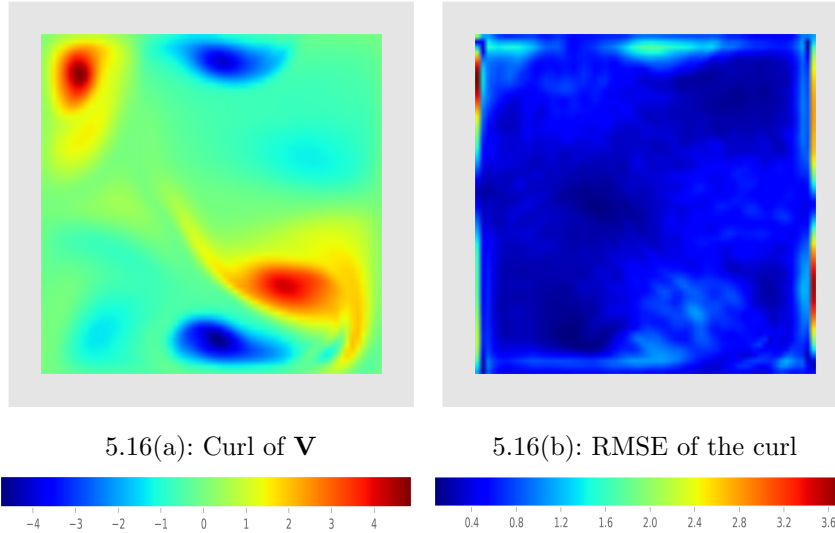
Figure 5.16: Curl of of the Navier-Stokes vector field **V** (a) and RMSE of the curl operator (b) between the set $\mathcal{E}$ and the realization **V**.

Following what was presented in Chapter 4, we can also calculate the correlation between the NHHD components of the derived realizations and the original vector field.

Figures 5.17 (a), (b), and (c) depict the Spearman's Rank correlation between these sets, computed as explained in Section 4.3. Figure 5.17 (d), shows the maximum correlation classification. Color is used to represent which of the three components — rotational-free (blue), divergence-free (white), and harmonic (red) — has the maximum correlation value at each point.

As mentioned before, the original vector field seemed to be dominated by the divergence-free component. Figure 5.17 (b) shows that the divergence-free component of the new realizations is highly correlated with the Navier-Stokes vector field, as expected. We can also note that the harmonic component of the derived realizations has a high degree of correlation with the original vector field, as opposed to the rotational-free component.

5.2.4.0
Particle-Image Velocimetry

Often, PIV applications aims to study the behavior of turbulent flows, analyzing the stability of features such as vortices. Besides providing means to
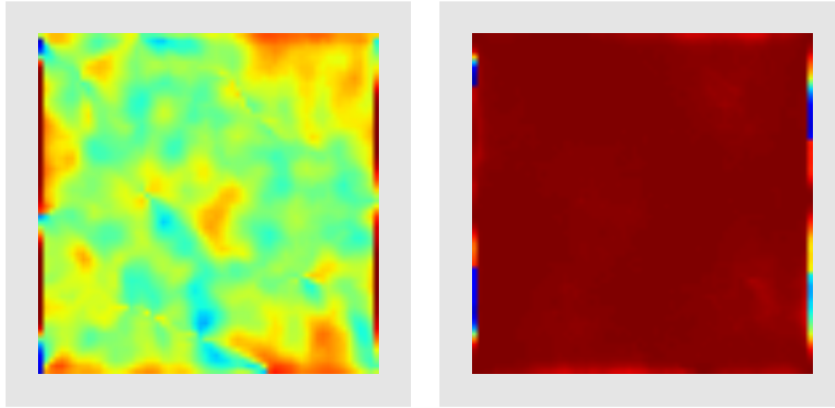
5.17(a): Rotational-free / vector-field correlation

5.17(b): Divergence-free / vector-field correlation
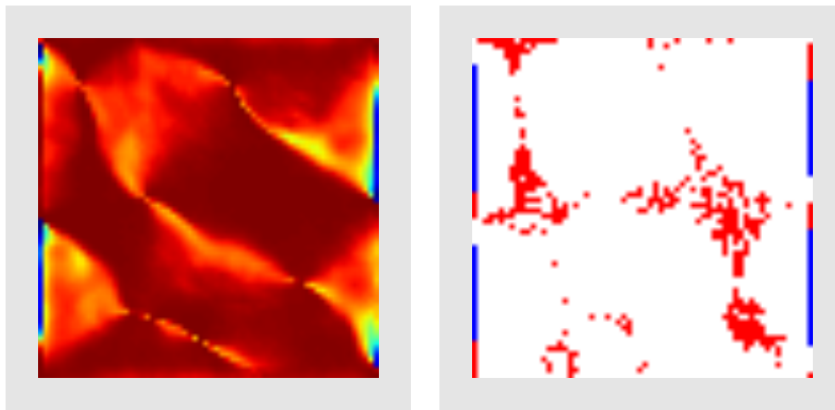
5.17(c): Harmonic / vector-field correlation

5.17(d): Maximum correlation classification
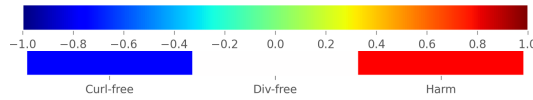
Figure 5.17: Correlation between the Navier-Stokes vector field and each NHHD component of the derived realizations.

perform this kind of study through the generation of different realizations, we can go further with the new samples generated using the proposed technique.

The following PIV simulation is defined over a grid of 124×126. Its horizontal dimension ranges from 0.3824 to 47.4176. On the other hand, its vertical dimension ranges from 0.3824 to 48.1824. Figure 5.18 shows this vector field, as well as its NHHD components. This image corresponds to a velocity field of a gas flow that is continuously injected horizontally on the bottom left corner and that flows on the domain from left to right until it meets a wall (image's right edge). It is possible to observe that the divergence-free component seems to have a high magnitude and basically dominate the flow behavior; we can also notice that the rotational-free component present some

features that characterize it.

5.18(a): Vector Field · · · · · · · · · · · · 5.18(b): Rotational-free

5.18(c): Divergence-free · · · · · · · · · · · 5.18(d): Harmonic
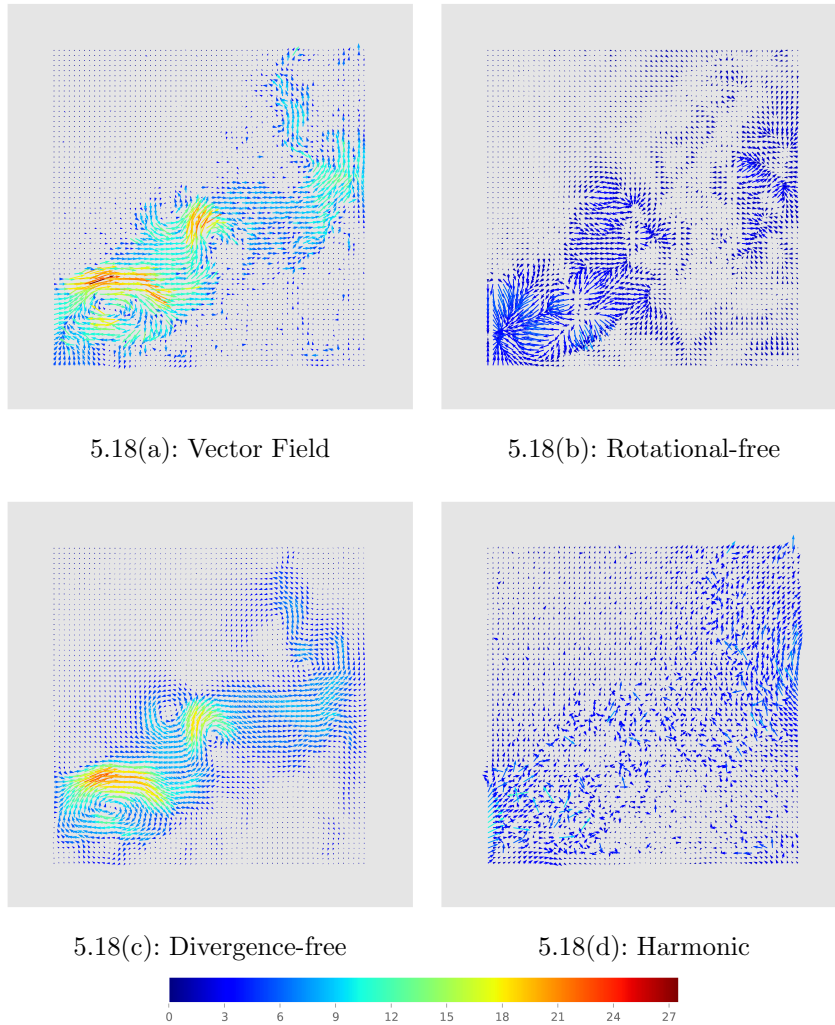
Figure 5.18: PIV simulation and its NHHD components.

Figure 5.19 presents the MDS between the new realizations and the original one.

Figure 5.20 shows the closest and farthest simulation derived from the original vector field. They are also represented using the same magnitude scale as their original field (Figure 5.18).

From Figure 5.21, we can see that, for the curl operator, the RMSE is higher on regions with high magnitude. In such areas, the scalar field also present high values. The same behavior happens with the divergence operator, i.e., we have a higher uncertainty in areas where the magnitude of the vector field is also higher.

Figure 5.23 presents the correlation between the NHHD components of the derived realization and the original field. As with the Navier-Stoke realizations, we can observe that for PIV the divergence-free and harmonic components also present a high correlation with the original vector field
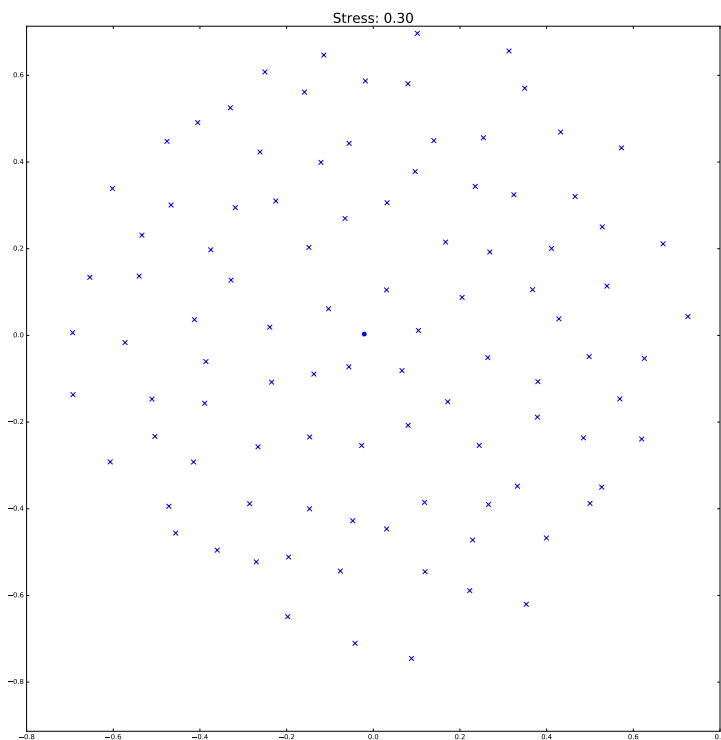
Figure 5.19: MDS visualization between the generated realizations set and the PIV vector field.
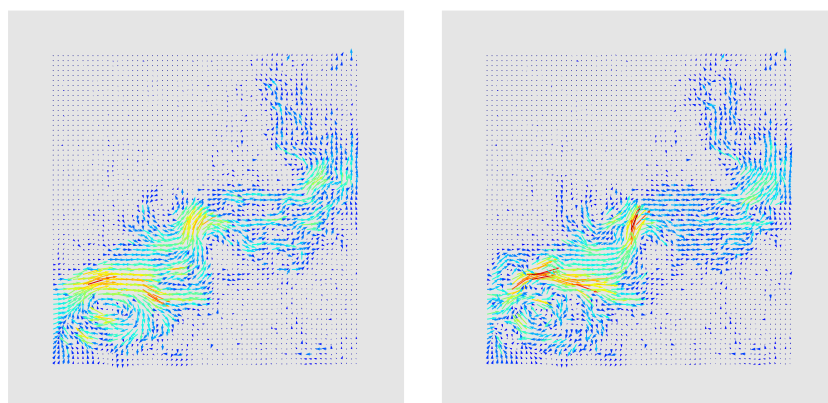
5.20(a): Closest Simulation

5.20(b): Farthest Simulation

Figure 5.20: Closest and farthest simulation of the PIV vector field.

5.21(a): Curl of **V**   5.21(b): RMSE of the curl
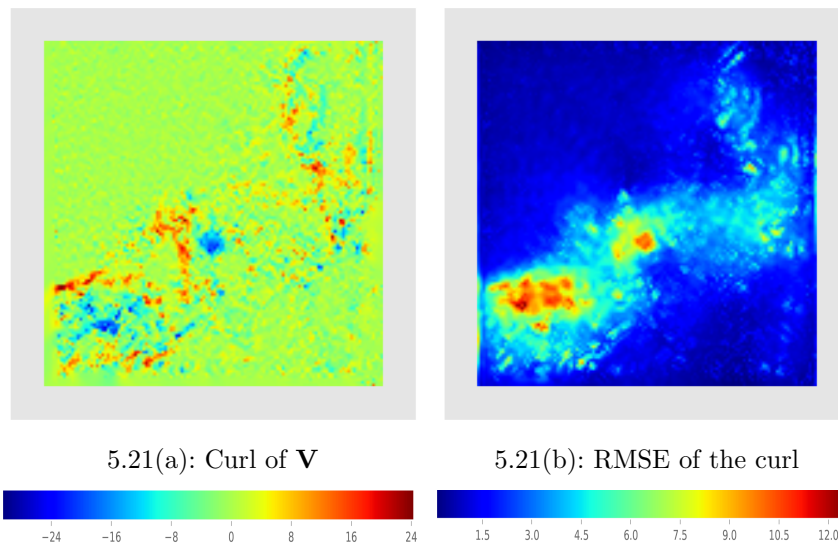
Figure 5.21: Curl of the PIV vector field **V** (a) and RMSE of the curl operator (b) between the set $\mathcal{E}$ and the realization **V**.

5.22(a): Divergence of **V**   5.22(b): RMSE of the divergence
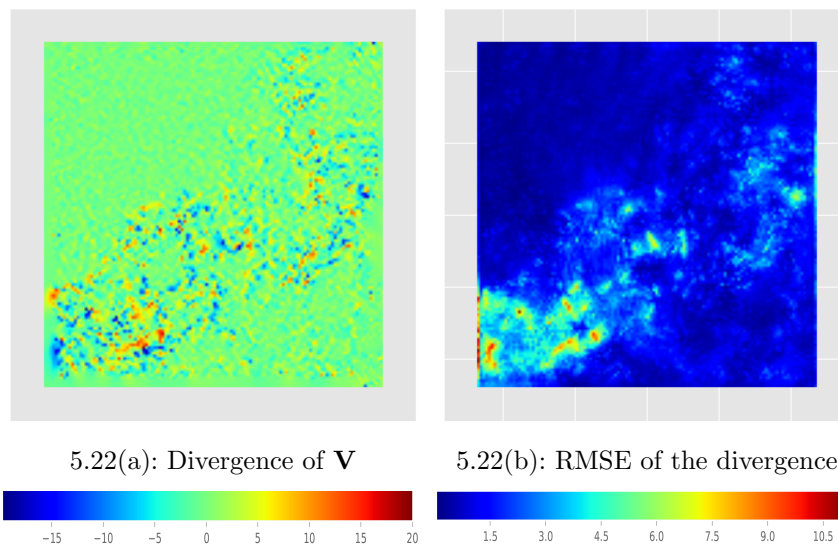
Figure 5.22: Divergence of the PIV vector field **V** (a) and RMSE of the divergence operator (b) between the set $\mathcal{E}$ and the realization **V**.

(Figures 5.23 (b) and (c)). However, in this case, we can also note that the rotational-free component defines important features in those realizations (Figure 5.23 (a)). Nevertheless, when put together, such vector field is still dominated by the divergence-free and harmonic components (Figure 5.23 (d)).
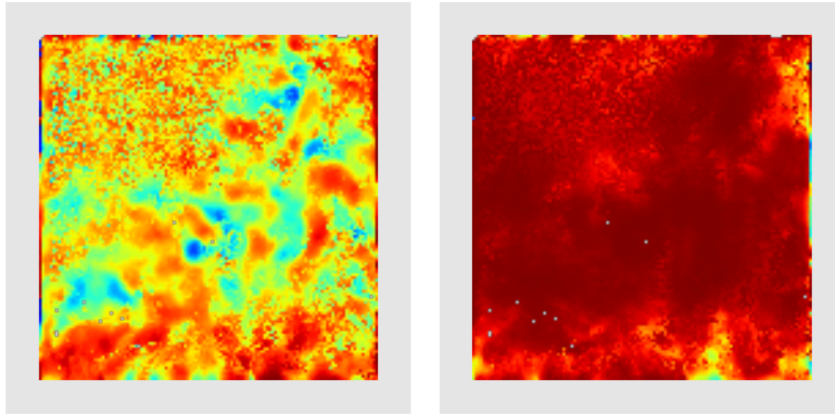
5.23(a): Rotational-free / vector-field correlation

5.23(b): Divergence-free / vector-field correlation
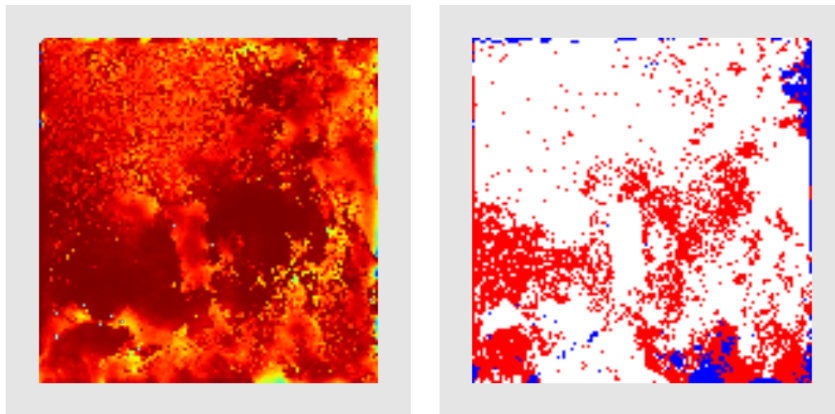
5.23(c): Harmonic / vector-field correlation

5.23(d): Maximum correlation classification
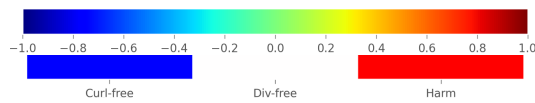
Figure 5.23: Correlation between the PIV vector field and each NHHD component of the derived realizations.

5.3
Method 2: Interpolation Based Stochastic Simulation

In this Section we present a different method to stochastically generate vector fields given a sample. It also uses the Helmholtz-Hodge Decomposition as a base step. The rest of it, on the other hand, is based on using an interpolation method to derive a realization.

5.3.1
Proposed Method

As defined in Section 5.2.2, our method is defined over a discrete sampling of a 2-dimensional domain on a Cartesian grid structure $\mathbb{S}_{m,n} = \{\mathbf{x}_{i,j} \in \mathbb{R}^2 : 1 \leq i \leq m, 1 \leq j \leq n\}$. For a given 2D vector field $\mathbf{V}$, consider it also defined over $\mathbb{S}_{m,n}$, i.e., to each spatial point in $\mathbb{S}_{m,n}$ there is a 2D vector associated. Henceforth, we will call this 2D vector field $\mathbf{V}$ the *training data*.

5.3.1.0
Method Overview

Once again, we firstly compute the NHHD of the training data $\mathbf{V}$, obtaining its components as previously depict in Equation 3-15. Given p random points $\mathbf{x}_{i,j} \in \mathbb{S}_{m,n}$ we create a new realization with pre-known values at those points, according to the values of $\mathbf{V}$ in $\mathbf{x}_{i,j}$. Those points are going to be used to derive the complete realization through an interpolation step. Finally, we perform a smoothing step to filter any noise that may be added.

5.3.1.0
The Algorithm

A pseudocode of the proposed method can be found in Algorithm 2. This pseudocode generates a realization $\mathbf{R}^*$ based on the NHHD components of a training 2D vector field $\mathbf{V}$.

It also expects as input the following list of variables:

- the scalar fields φ^*, ψ^* and the vector field $\mathbf{h}^*$ obtained by the NHHD of the training 2D vector field $\mathbf{V}$;

- number p of points to be randomly chosen.

For a number of points p, p points are randomly generated, being resampled without replacement (lines 4 and 5). These points are going to be fixed and used to interpolate and, thus, derive a new realization. More specifically, those points are going to be set as *known* in the scalar potentials

input : φ^*, ψ^*, $\mathbf{h}^*$, p

output: $\mathbf{R}^*$, a vector field realization

1   $n, m \leftarrow \texttt{size}(\varphi)$;

2   $\varphi_{itp}^* \leftarrow$ empty matrix of $n \times m$;

3   $\psi_{itp}^* \leftarrow$ empty matrix of $n \times m$;

    /* Derive p indexes in the 2D grid   */ ;

4   $x \leftarrow \texttt{randInt}(1, m, p)$;

5   $y \leftarrow \texttt{randInt}(1, n, p)$;

    /* For each derived index   */ ;

6   **for** $k \leftarrow 1$ **to** p **do**

7 | $i \leftarrow x[k]$;

8 | $j \leftarrow y[k]$;

9 | $\varphi^*(i,j)_{itp} \leftarrow \hat{\mathrm{F}}\varphi(i,j)$;

10 | $\psi^*(i,j)_{itp} \leftarrow \hat{\mathrm{F}}\psi(i,j)$;

11   **end**

12   $\varphi_{itp}^* \leftarrow \texttt{interpolation}(\varphi_{itp}^*)$;

13   $\psi_{itp}^* \leftarrow \texttt{interpolation}(\psi_{itp}^*)$;

14   $\varphi_{itp}^* \leftarrow \texttt{smooth}(\varphi_{itp}^*)$;

15   $\psi_{itp}^* \leftarrow \texttt{smooth}(\psi_{itp}^*)$;

16   $\nabla\varphi_{\mathbf{R}}^* \leftarrow \texttt{divergent}(\varphi_{itp}^*)$;

17   $(\nabla \cdot J)\varphi_{\mathbf{R}}^* \leftarrow \texttt{curl}(\psi_{itp}^*)$;

18   $\mathbf{R}^*(\mathbf{x}_{i,j}) \leftarrow \nabla\varphi_{\mathbf{R}}^*(\mathbf{x}_{i,j}) + \nabla \times \varphi_{\mathbf{R}}^*(\mathbf{x}_{i,j}) + \mathbf{h}^*$;

Algorithm 2: Generating a realization $\mathbf{R}^*$ based on the NHHD components of a training data $\mathbf{V}$ using an interpolation approach.

that define the input vector field, i.e., in φ^* and ψ^*. Note that these points can be select independently for these potentials.

Through a cubic interpolation, we obtain the potentials φ^*, ψ^*. After that, we apply a smoothing function on it to reduce any noise that may be inserted. Here, we also use a Gaussian filter with σ equal to 2.

Now we can obtain the divergence-free (line 16) and rotational-free (line 17) components of $\mathbf{V}$. At last, we obtain a new vector field realization (line 18) as stated in Equation 3-15.

5.3.2
Results and Discussion

To verify the achieved results we make use of the same data set previously presented in Section 5.2.3, comprehended by seven multi-method wind forecast realizations $\mathcal{E}$.

Once again, the proposed method firstly obtain the divergence-free, rotational-free and harmonic components of each realization $\mathbf{R}$ in $\mathcal{E}$, using the NHHD. Through Algorithm 2, we derive 100 new realizations for each $\mathbf{R}$

in $\mathcal{E}$. These will be visualized using the MDS technique, following the same procedures defined in Section 5.2.3.

5.3.2.0
Coverage Test

As done in Section 5.2.3, our first evaluation of the proposed method consists in verifying if it is capable of providing some variability while generating valid realizations. Here, we define as *valid* the given realizations present in $\mathcal{E}$. Using the mean vector field of $\mathcal{E}$, μ (previously presented in Figures 5.7 and 5.6), we try to derive a new set of realizations containing the members of $\mathcal{E}$. To do so, we performed Algorithm 2 on $\mathcal{E}$ using different numbers for p. Figure 5.24 shows the obtained results.

From this figure, we might conclude that using p as 0.01% of the vector field dimension results in a set of realizations that covers all members of $\mathcal{E}$. However, using a small p may result in a *bad* interpolation result, bad being defined as the loss of significant features of the vector field, as its singular points. After a visual inspection, which aimed to inspect the similarity between the derived vector field and the training data, we observed that, for μ as the training data, the best results were achieved with p greater or equal to 0.03% of the size of the input vector field (Table 5.2 presents examples of realizations for different values of p.).

For this scenario, Figures 5.25 and 5.26 depict the standard deviation of the potentials and HHD components obtained through this approach.

As can be seen, for both potentials, the regions where the standard deviation is greater than zero are larger then compared with the Bootstrap-like results (Figures 5.9 and 5.10). Also, the maximum value of this deviation, in both cases, is also higher then the one obtained with the best scenario of the Bootstrap-based approach. Once again, there is no variation in the harmonic component. However, as can be seen in Table 5.2, results obtained with the interpolation-based approach are smoother then the ones obtained with the Bootstrap-based technique (Table 5.1).

We proceed generating 100 new realizations for each vector field in the original ensemble $\mathcal{E}$, using p equal to 0.03%. Figure 5.27 shows the correspondent MDS visualization. Markers of the same color belongs to the same set, i.e, were generated based on a common realization. Circle markers represent each realization of the set $\mathcal{E}$. Cross markers represent new realizations, while diamond and square markers show both the closest and farthest simulation, respectively, given a base realization – Table 5.3 depict these simulations for each realization in $\mathcal{E}$.
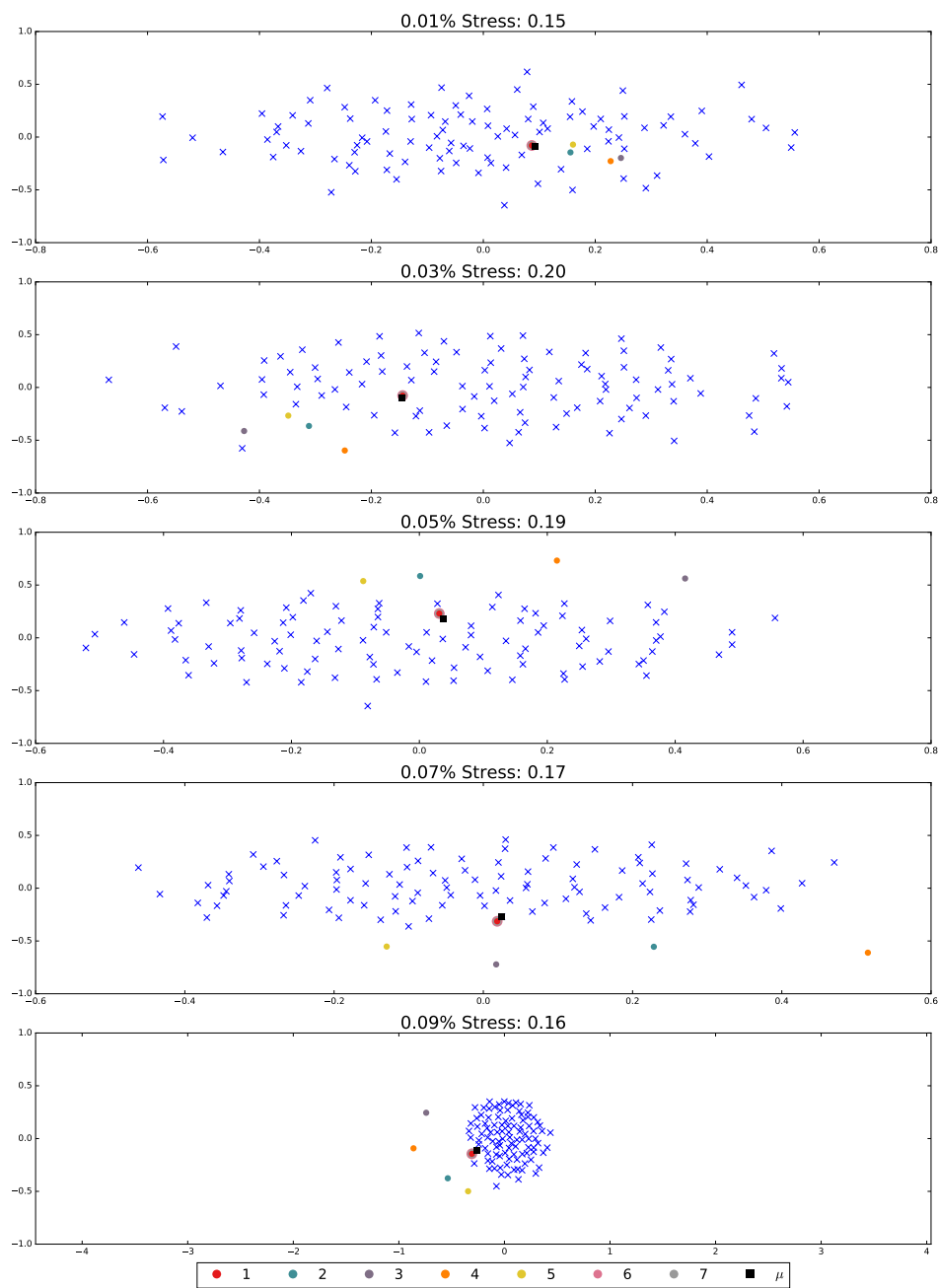
Figure 5.24: MDS visualization between $\mathcal{E}$ and a new realizations, obtained using different values of p.

Table 5.2: Original data and derived realizations for different values of p.

Divergence-free potential	Rotational-free potential	Vector Field

Original Data

$p = 0.01$

$p = 0.03$

$p = 0.05$

$p = 0.07$

$p = 0.09$
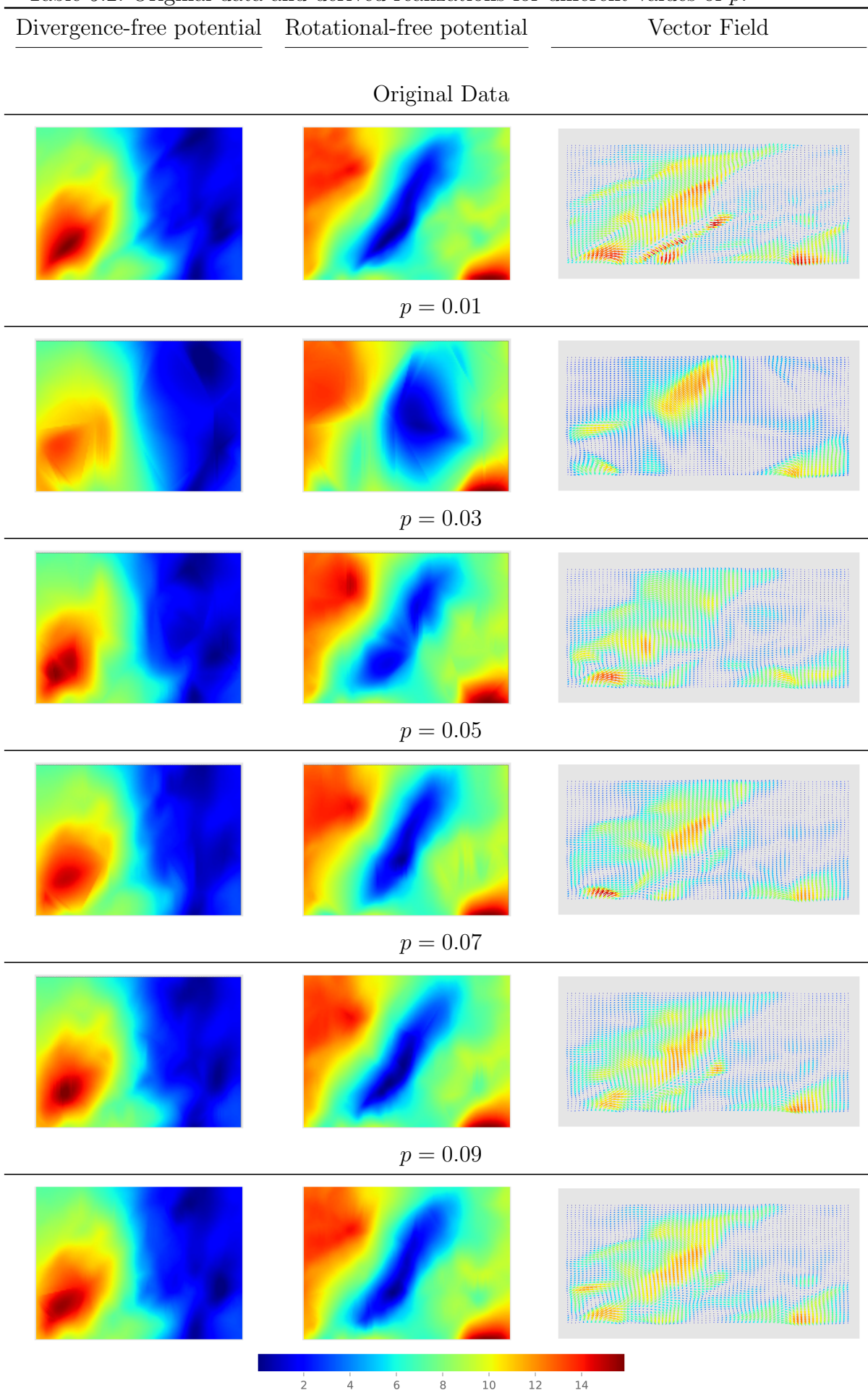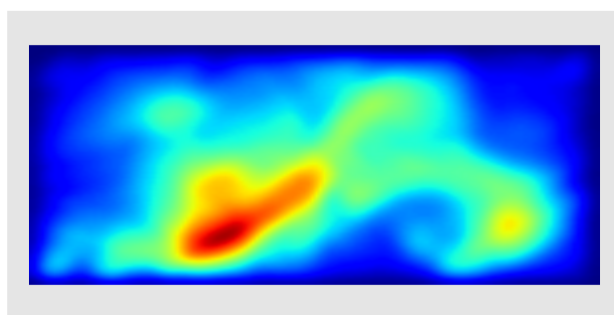
5.25(a): Rotational-free potential standard deviation σ_φ
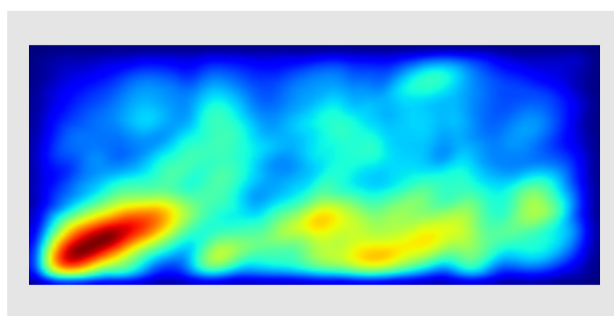
5.25(b): Divergence-free potential standard deviation σ_ψ
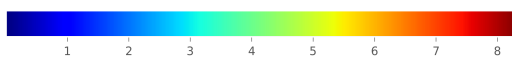
Figure 5.25: Standard deviation of the rotational-free and divergence-free potentials obtained through the Interpolation-based approach.

5.26(a): Vector-field magnitude standard deviation
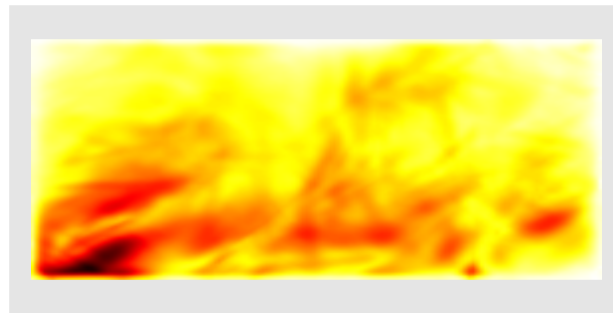
5.26(b): Rotational-free magnitude standard deviation
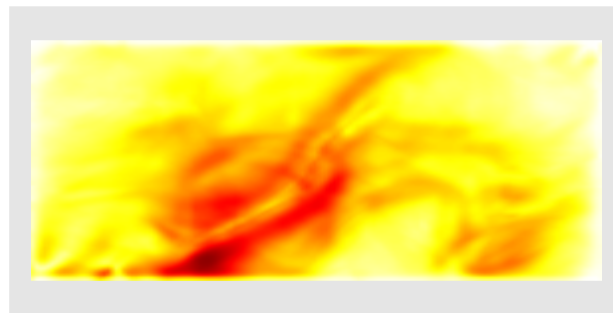
5.26(c): Divergence-free magnitude standard deviation
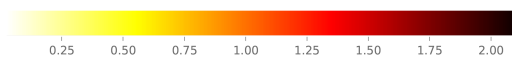
Figure 5.26: Standard deviation of the magnitude of the rotational-free and divergence-free components using the Interpolation-based approach.
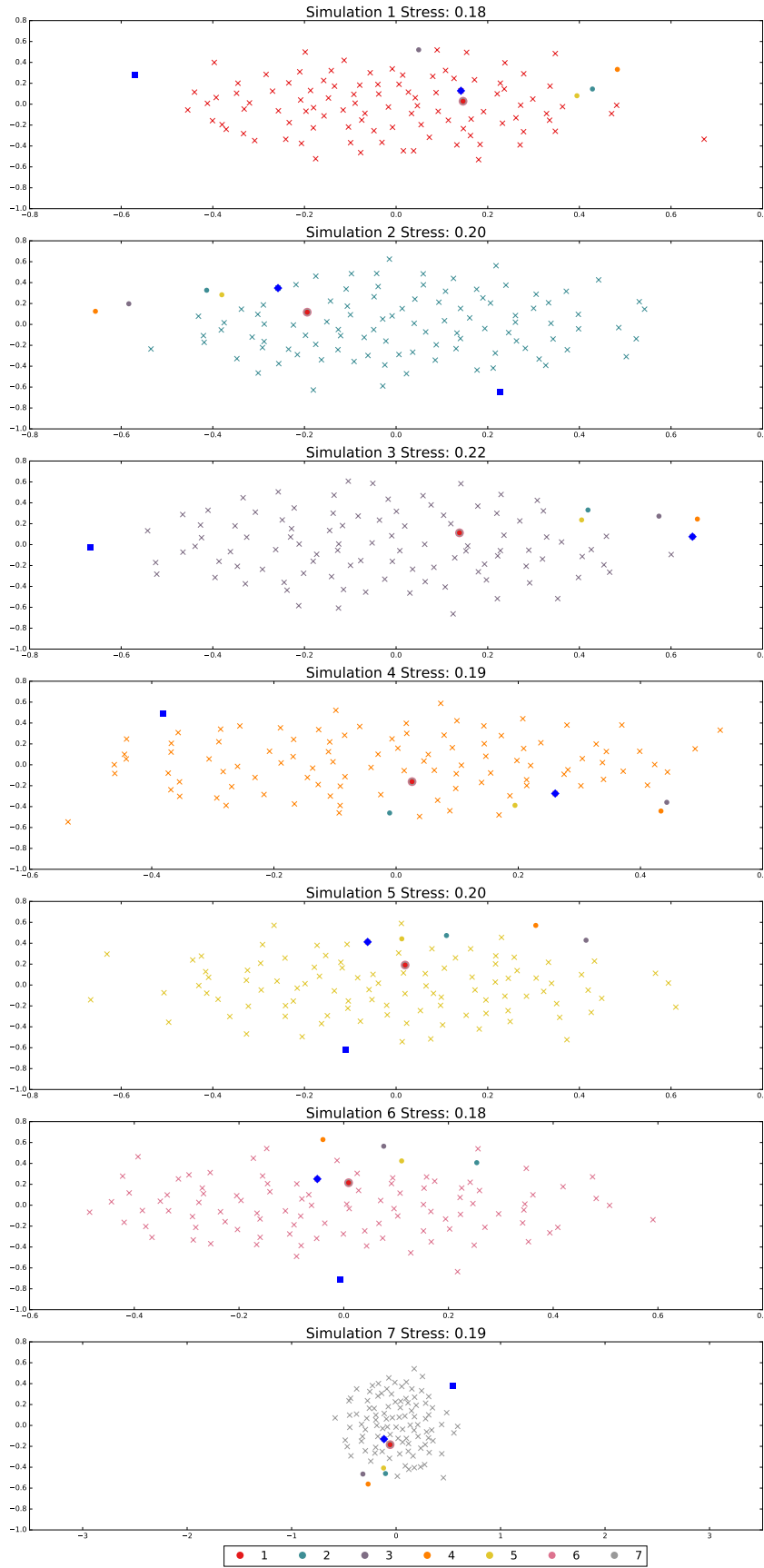
Figure 5.27: MDS visualization between each set of new realizations and the original ensemble $\mathcal{E}$. Colors represent each realization of the set $\mathcal{E}$. Circular markers represent each realization in $\mathcal{E}$. Cross markers represent, for each $\mathbf{V}$ in $\mathcal{E}$, the new realizations derived from $\mathbf{V}$, both represented with the same color.

Table 5.3: Original realization and its closest and farthest realizations using the Interpolation-based approach.

Sample	Closest Simulation	Farthest Simulation
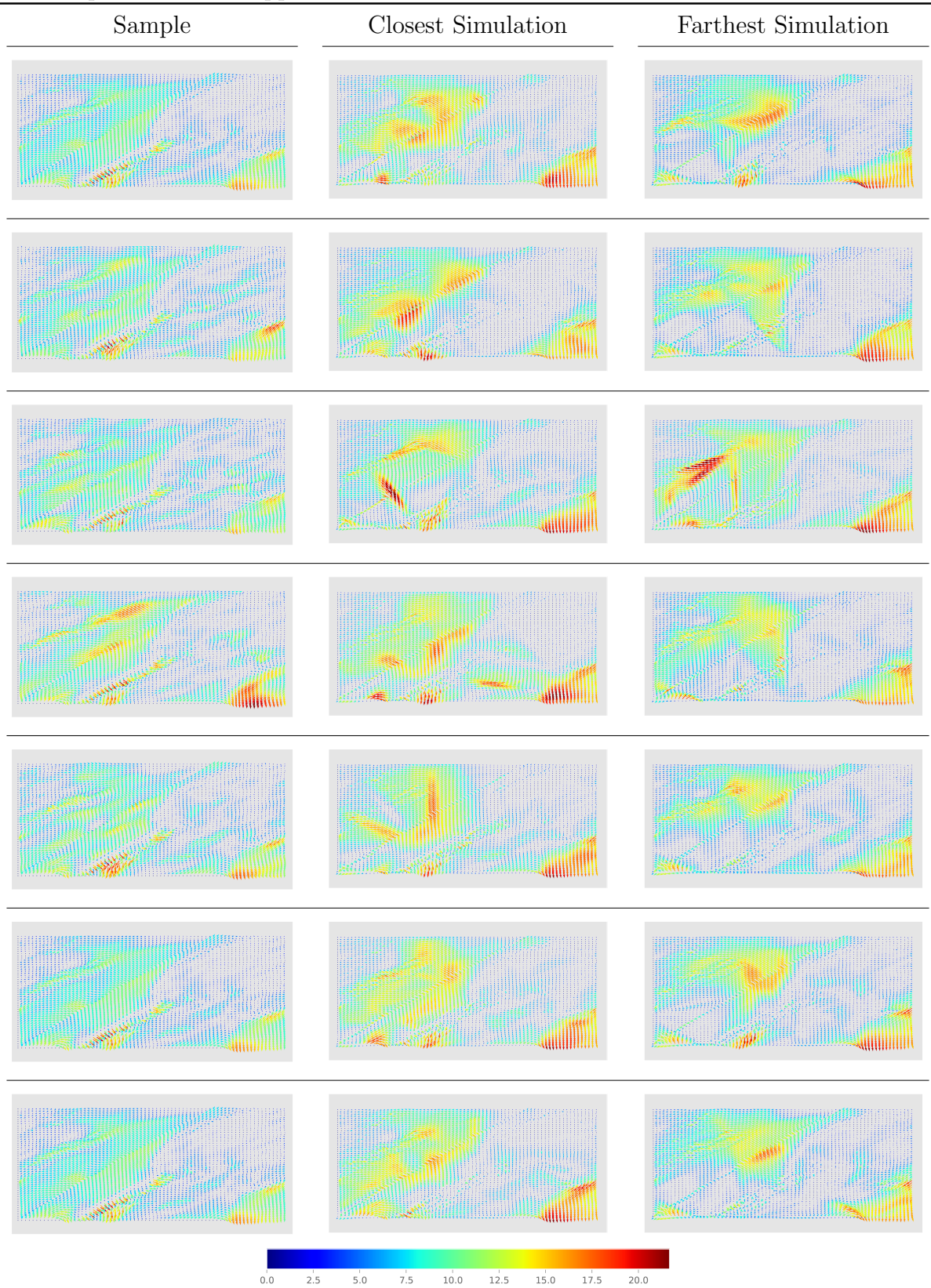
Figure 5.28, on the other hand, present the MDS for all these realizations. As can be seen, they mix together.
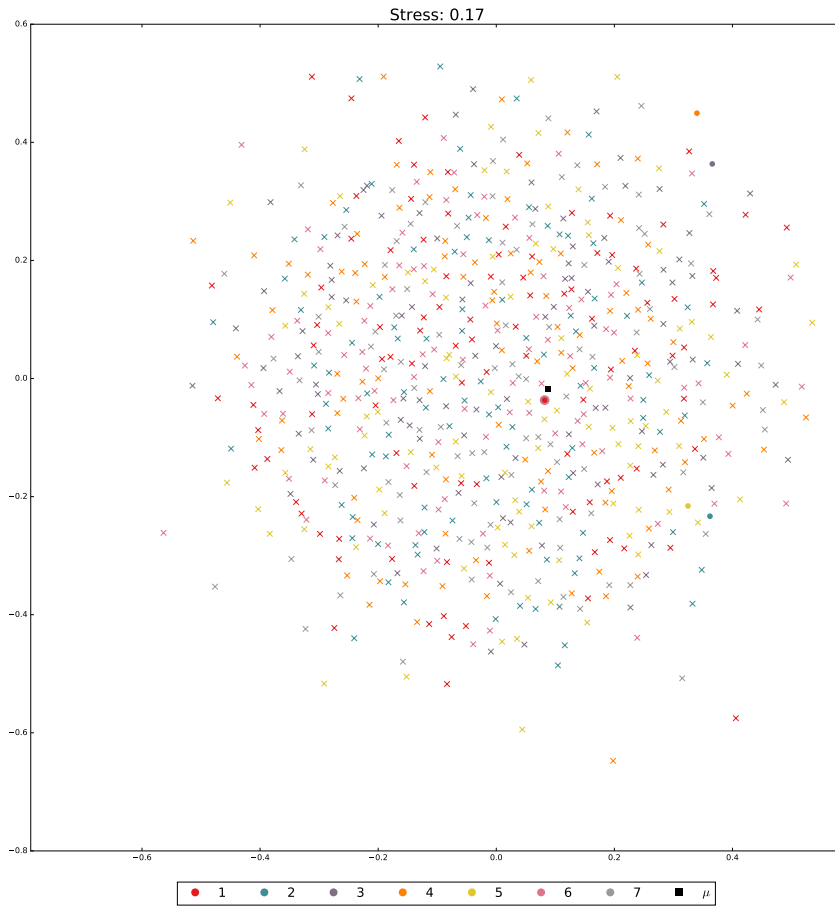
Figure 5.28: MDS visualization between each new set of realizations and the original ensemble $\mathcal{E}$. Colors represent each realization of the set $\mathcal{E}$. Circular markers represent the realizations in $\mathcal{E}$. Cross markers represent new realizations derived from the one presented with a circular marker of the same color.

6
Vector Field Synthesis

In this Chapter we will present an approach to synthesize vector fields realizations from an ensemble. In the field of *Computer Graphics* and *Image Processing*, to synthesize an image consist in creating images from an image description or model. Here, we have a vector field ensemble, consisting of multiple realizations of some specific data, and we intend to use them as the *description* of possible vector fields realizations.

Dimensionality reduction, or *multidimensional projection*, is an approach used to represent a multidimensional data in a low-dimensional space. Its goal consists in providing an overview of similarities between instances of data in a projection space (Amorim et al., 2015), which can then be visually encoded and interpreted. Many algorithms for dimensionality reduction can be found in the literature. Among them, we can cite the well know PCA (Jolliffe, 1986), MDS (Cox & Cox, 2000), t-SNE (van der Maaten & Hinton, 2008) and LAMP (Joia et al., 2011).

Inverse projection, on the other hand, works in the opposite way. I.e., given a multidimensional data set $X \subset \mathbb{R}^m$ and its 2D projection $Y \subset \mathbb{R}^2$, such technique allows the creation of new points in the input space, $\mathbb{R}^m$, given points in the projection space, $\mathbb{R}^2$.

With the knowledge that *vector fields* are used in a varied set of scientific applications, we proposed applying the inverse projection technique to generate new realizations of vector fields from an ensemble. Furthermore, using the Helmholtz-Hodge Decomposition, we also provide an approach to derive vector fields given ensembles of the rotational-free, divergence-free and harmonic components.

The proposed approach can be summarized as follows: the NHHD is calculated to obtain the divergence-free, rotational-free and harmonic components that compose a given set of vector fields. Multidimensional projection is then performed on the original data set as well as on its NHHD components. After that, for any point p in the projection space, it can me mapped into the original space to generate a new vector field realization.

6.1
Motivation

The same motivations described in Section 5.1 apply here. We believe that the main contributions of the technique to be presented consist in:

– A framework that allows the exploration of a multidimensional space to intuitively create new vector fields;

– For several areas of scientific applications, a faster way to create new multidimensional realizations;

– The possibility to generate new realizations from a set of possible ones, instead of a single realization.

6.2
Inverse Projection

A *Radial Basis Function* (RBF) is a function with respect to the origin or a certain point p, i.e., $\psi(x) = f(\|x - p\|)$ with $\| \cdot \|$ being often the Euclidean norm. Any function that satisfies $\psi(x) = \psi(\|x\|)$ is a radial function, or a radial basis kernel.

Recently, Amorim et al. (2015) proposed an inverse projection technique based on RBF interpolation, providing a smooth and global mapping from low to high dimensions. In RBF interpolation, for N data points $X_i \in \mathbb{R}^m$ and their respective function values $Y_i \in \mathbb{R}$, an approximant $s : \mathbb{R}^m \to \mathbb{R}$ is construct in such a way that is possible to derive the value y for any arbitrary point in $x \in \mathbb{R}^m$ in such a way that:

$$s(x) = \sum_{i=1}^{N} \lambda_i \psi(\|X_i - x\|), \tag{6-1}$$

where λ_i are real-valued coefficients.

Here, the choice of the kernel affects the smoothness of the approximant $s(x)$.

Given that, an inverse projection using RBF interpolation will obtain a function $s : \mathbb{R}^2 \to \mathbb{R}^m$ mapping information from the projection space into the original m-dimensional space.

6.3
The Proposed Method

Consider a discrete sampling of a 2-dimensional domain on a Cartesian grid structure $\mathbb{S}_{r,c} = \{\mathbf{x}_{i,j} \in \mathbb{R}^2 : 1 \leq i \leq r, 1 \leq j \leq c\}$. Also, let $\mathcal{E}$ be a 2D vector field ensemble, i.e., a 2-dimensional vector is defined at each spatial point in $\mathbb{S}_{r,c}$ for $\mathbf{V} \in \mathcal{E}$. $\mathcal{E}$ can be viewed as a m-dimensional data set with $m = r \times c$. As described by Amorim et al. (2015), being $Y \in \mathbb{R}^2$ the 2D projection of $\mathcal{E}$, for any point $p \in \mathbb{R}^2$ we want to find its m-dimensional representation, i.e., a point $q \subset \mathbb{R}^m$. Given the concept of RBF interpolation, we seek $s(p)$:

$$s(p)_k = \sum_{i=1}^{N} \lambda_{ki} \psi(\|y_i - p\|), \tag{6-2}$$

where s_k accounts for k-th output dimension. Note that we want a approximant $s : \mathbb{R}^2 \to \mathbb{R}^m$, i.e., $s(y_j) = x_j$. Hence, we can rewrite Equation 6-2 as:

$$s(p)_k = \sum_{i=1}^{N} \lambda_{ki} \psi(\|y_i - y_j\|) = x_{kj} \tag{6-3}$$

To find the scalar coefficients λ_k, s_k reduces to solve the the linear system $\psi \lambda_k = b_k$, where ψ is the interpolation matrix with $\psi_{ij} = \psi_{ji} = \psi(\|y_i - y_j\|)$, $\lambda_k = [\lambda_{k1} \cdots \lambda_{kN}]^\intercal$ and $b_k = [x_{k1} \cdots x_{kN}]^\intercal$. With λ_{ki} in hand the approximant s is complete and can be used to derive $q \in \mathbb{R}^m$ to any given point $p \in \mathbb{R}^2$.

In this work, we are going to a Gaussian kernel $\psi = e^{-\varepsilon r^2}$, where ε is a positive parameter set to 1 and r is the Euclidean distance between a pair of realizations.

6.3.1
Method Overview

We first obtain the NHHD for each vector field in the ensemble $\mathcal{E}$. Having the original vector fields and their NHHD components in hand, we generate four sets of multidimensional projection, one for $\mathcal{E}$ and the other ones for the each NHHD component of $\mathcal{E}$, i.e., for the rotational-free, divergence-free and harmonic components.

To perform a dimensionality reduction on a data set it is necessary to obtain a similarity measure for each pair of instances in the data. Considering an ensemble $\mathcal{E}$ of vector fields, this means that, for each pair of vector fields $\mathbf{V}_i, \mathbf{V}_j \in \mathcal{E}$ it is necessary to define how close they are to each other. Here, we use the same similarity measure as described in Section 5.2.3.

After that, there are two possibilities to generate new vector field realizations: from a point defined in the first projection space, i.e., based on

the original data, or from points defined in the NHHD components projection space, i.e., through the combination of the NHHD components.

In the following, it is necessary to compute the inverse projection as described in Section 6.3. For the first case, the final realization is directly given by this inversion. For the second case, it is given by the sum of the inverse projection of each of the NHHD components, according to Equation 3-15.

6.3.2
The Algorithm

A pseudocode of the proposed approach is presented in 3. This pseudocode generates a new realization $\mathbf{R}^*$ given a point p in a projection space. It consists of two main procedures. The first one, `find_lambda`, is responsible to obtain the λ matrix and takes as input the following list of variables:

– a list of points in the projection space, $p[]$;

– the correspondent vector field, unrolled, for each given point, $v_f[]$.

The second one, `inverse_projection`, obtain a new realization given a point p in the projection space, taking as input the following list of variables:

– a list of points in the projection space, $p_r[]$;

– a point p, also in the projection space, to be inverted;

– the λ matrix.

The algorithm is straightforward. Firstly, it is necessary to obtain the dimensionality reduction for the input data, being it the original vector fields or their NHHD components. After that, it is necessary to obtain the λ matrix, used to obtain the inverse projection for a given point p.

Note that, when generating a vector field from the input data NHHD components, lines 1 and 2 have to be performed for each of the components. This means that we obtain three λ matrices, one for the rotational-free component, one for the divergence-free component and another one for the harmonic component. We will also end up with the three NHHD components, that have to be combined as shown in Equation 3-15 in order to obtain a new vector field realization.

Repeating this procedure R times, for random points in the projection space, one can generate R new vector field realizations.

input : $p_r[]$, $v_f[]$, p

output: R, a vector field realization

```
/* Perform the dimensionality reduction of your choice
   to obtain p_r[]                           */ ;
```

1 $\lambda \leftarrow$ `find_lambda`$(p_r[], v_f[])$;

2 **R** $\leftarrow$ `inverse_projection`$(p_r[], p, \lambda)$;

3 **Procedure** `find_lambda`

   input : $p_r[]$, $v_f[]$

   ```
   /* array of len(p_r) × size(v_f[0])        */ ;
   ```

   output: λ, lambda values that reverse each v_f in $v_f[]$ to the given space

4 | $n \leftarrow$ `len`(p_r) ;

5 | $m \leftarrow$ `size`$(v_f[0])$;

6 | $a \leftarrow$ empty matrix of $n \times n$;

7 | $b \leftarrow$ empty matrix of $n \times m$;

8 | **for** $i \leftarrow 0$ **to** n **do**

9 | | **for** $j \leftarrow 0$ **to** n **do**

10 | | | $a[i,j] \leftarrow a[j,i] \leftarrow$ `kernel`$(\|p_r[i] - p_r[j]\|)$;

11 | | **end**

12 | **end**

13 | **for** $i \leftarrow 0$ **to** n **do**

14 | | **for** $j \leftarrow 0$ **to** m **do**

15 | | | $b[i,j] \leftarrow v[i][j]$;

16 | | **end**

17 | **end**

   ```
   /* Solve the linear system ax = b          */ ;
   ```

18 | **return** x ;

19 **Procedure** `inverse_projection`

   input : $p_r[]$, p, λ

   output: R*, an unrolled vector field realization

20 | $n \leftarrow$ `len`(p_r) ;

21 | $m \leftarrow$ `size`$(\lambda[0])$;

22 | $iv \leftarrow$ empty matrix of $n \times m$;

23 | **for** $i \leftarrow 0$ **to** n **do**

24 | | **for** $j \leftarrow 0$ **to** m **do**

25 | | | $iv[j] \leftarrow iv[j] + \lambda[i,j]*$ `kernel`$(\|p_r[i] - p_r[j]\|)$;

26 | | **end**

27 | **end**

28 | **return** iv ;

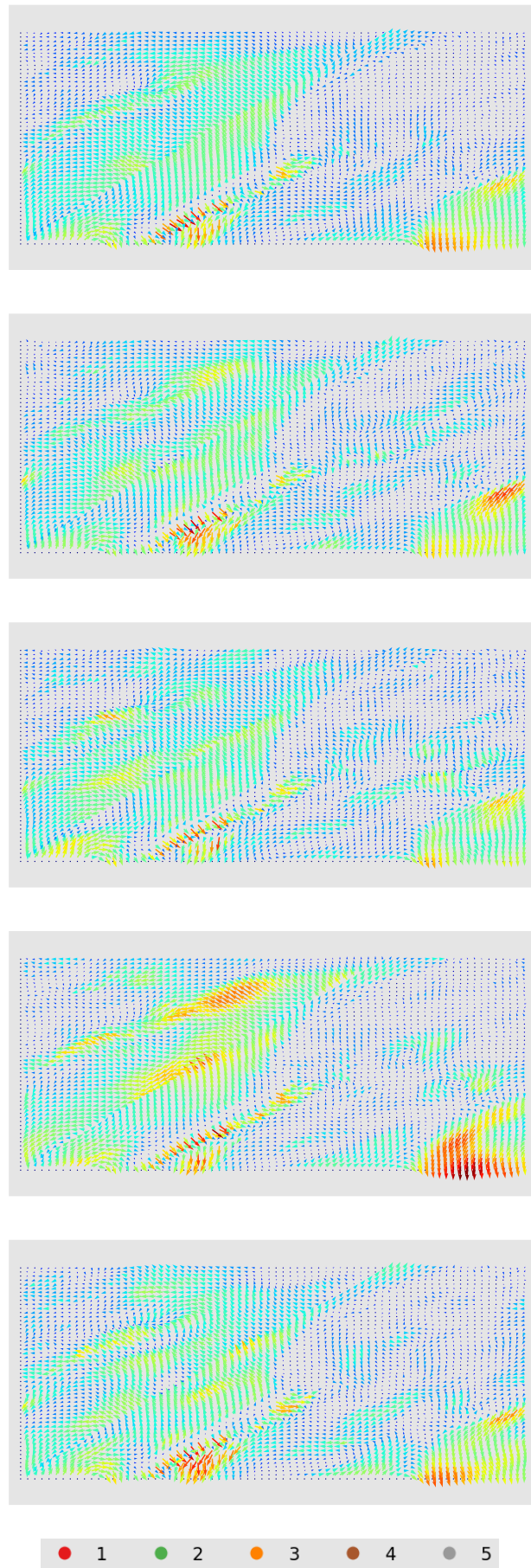**Algorithm 3:** Generating a realization through dimensionality reduction and inverse projection.

Figure 6.1: Multi-method wind forecast realizations. The number/color maps each realization to its representation in the projected space.

6.4
Results and Discussion

To test the proposed approach, we make use of a 2D vector field set comprehended by five multi-method wind forecast realizations $\mathcal{E}$, presented before. These realizations are depict in Figure 6.1 followed by a legend of circular markers showing how they are going to be represented into the projected space visualization.

Firstly, we apply the NHHD on each vector field in $\mathcal{E}$ to obtain its rotational-free, divergence-free and harmonic components. Those are going to be used to derive three different multidimensional projections in $\mathbb{R}^2$ through the MDS technique, plus the projection of the original set. They will provide two different ways of generating new realizations of a wind vector field.

6.4.1
User Interface

The developed interface comprehends two ways of synthesizing a vector field. The first one is accomplished based on the original vector field ensemble. The second one is made through the combination of the NHHD components that compose the original data set. In both cases, we perform a multidimensional reduction, through MDS, transforming the data space from $\mathbb{R}^m$ to $\mathbb{R}^2$. Doing that, we can provide a visualization of those data allowing an user to generate a new realization exploring a low dimensional space, i.e., a new vector field can be generated for any given point in the screen.

For the first case, when a point is selected in the projection space (Figure 6.2), its inverse projection is obtained, generating a new vector field based on the discovered λ values. For the second case, in the other hand, three points – one for each of the NHHD components – are required to generate a new realization (Figure 6.3). Here, a new realization is obtained as the sum of the inverse projection of each one of the given points, accordingly to the Helmholtz-Hodge Decomposition. In both cases, results are obtained in real time.

Using the `Ctrl` key, a selected point is bonded to the closest projected point. This means that, for the NHHD-based vector field generation, an original component will be used. For example, in Figure 6.4, a new realization was created using the rotational-free and harmonic components of the first realization in the ensemble.
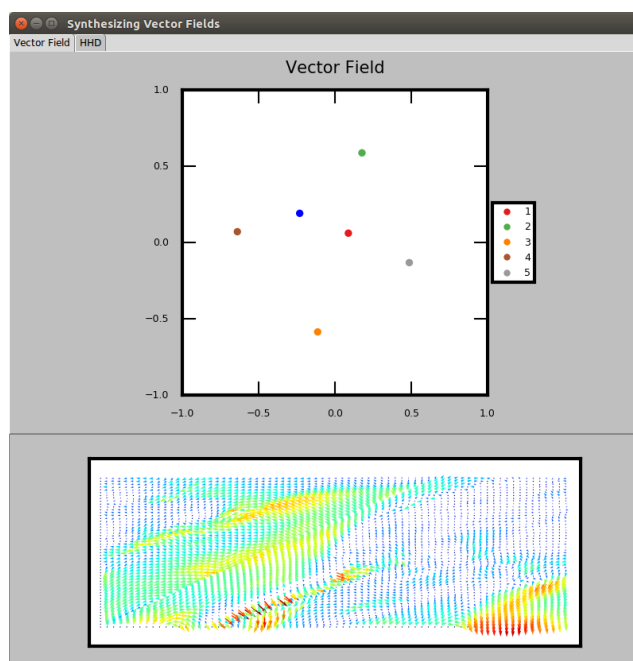
Figure 6.2: A blue point in the top plot represents the chosen point to be used in the inverse projection. The bottom plot shows the resulting vector field.
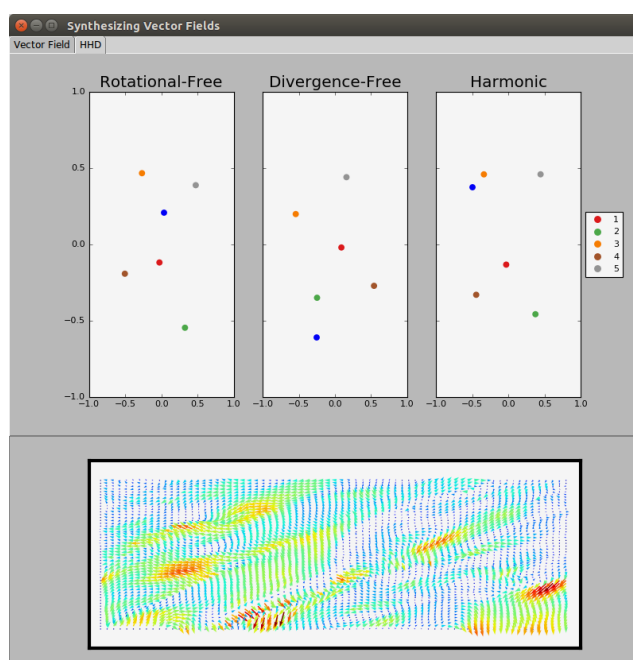

Figure 6.3: For each HHD component, an input point – shown in blue – has to be given. The bottom plot shows the resulting vector field, i.e., the sum of the vector field obtained through the inversion of each of the given points.
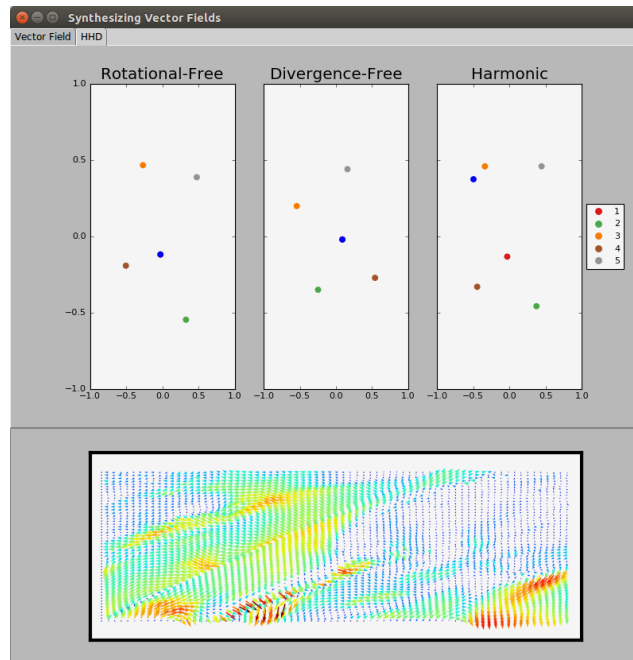
Figure 6.4: Pressing `Ctrl` while inserting a point make it snap to the closest one. This will fix a component of the original ensemble to be used when generating a new realization.

6.4.2
Evaluation

To evaluate the results obtained with the proposed technique, two experiments were performed.

6.4.2.0
Experiment 1

The first experiment consisted in qualitatively analyze the results of the inverse projection given four extreme scenarios. By extreme we mean that these scenarios are well defined and don't have any feature in common. They are depict in Figure 6.5.

Exploring this ensemble projection space, we expect to derive realizations containing features from each one of these scenarios, with their intensity being a factor of the proximity between the projected points and the chosen one. The multidimensional projection for this set, as well as the chosen points to be inverted, can be seen in Figure 6.6.

The results of the inverse projection, on the other hand, can be found in Figure 6.7. As can be seen, the proximity of a given point to an specific vector field (consequently, feature), states its influence in the derived realization. For example, point $(0.3, 0.0)$ is close to the projected point of a vector field containing a vortex. As a result, the derived realization is dominated by such
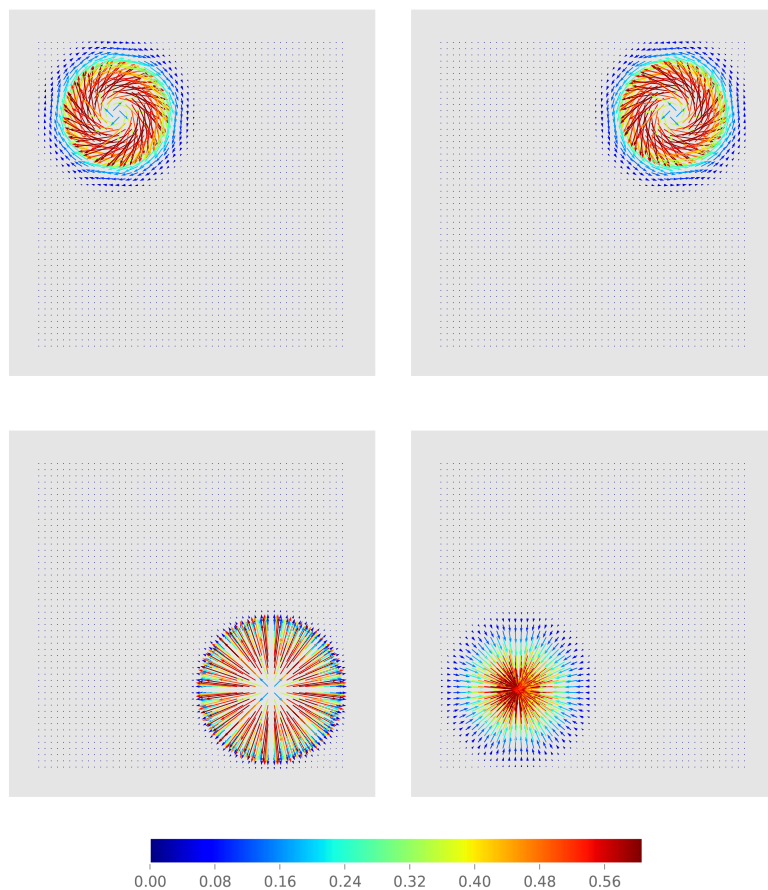
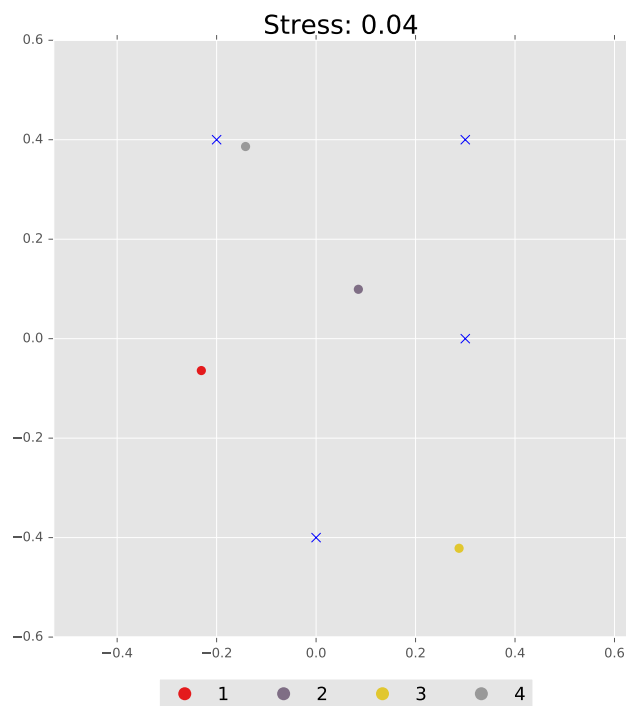Figure 6.5: Synthetic realizations to test the inverse projection technique.

Figure 6.6: MDS of the synthetic ensemble and the chosen points to be inverted. The i-th point is related to the i-th vector field presented in Figure 6.5.

feature. Point $(-0.2, 0.4)$, on the other hand, is closest to a projected point that represents a vector field comprehended by a sink. Thus, the derived realization has this sink as it main feature. Hence, balancing the position of a point defines how similar to each of the vector field inputs the derived realization will be.

6.7(a): $(0.0, -0.4)$ 6.7(b): $(0.3, 0.0)$

6.7(c): $(0.3, 0.4)$ 6.7(d): $(-0.2, 0.4)$
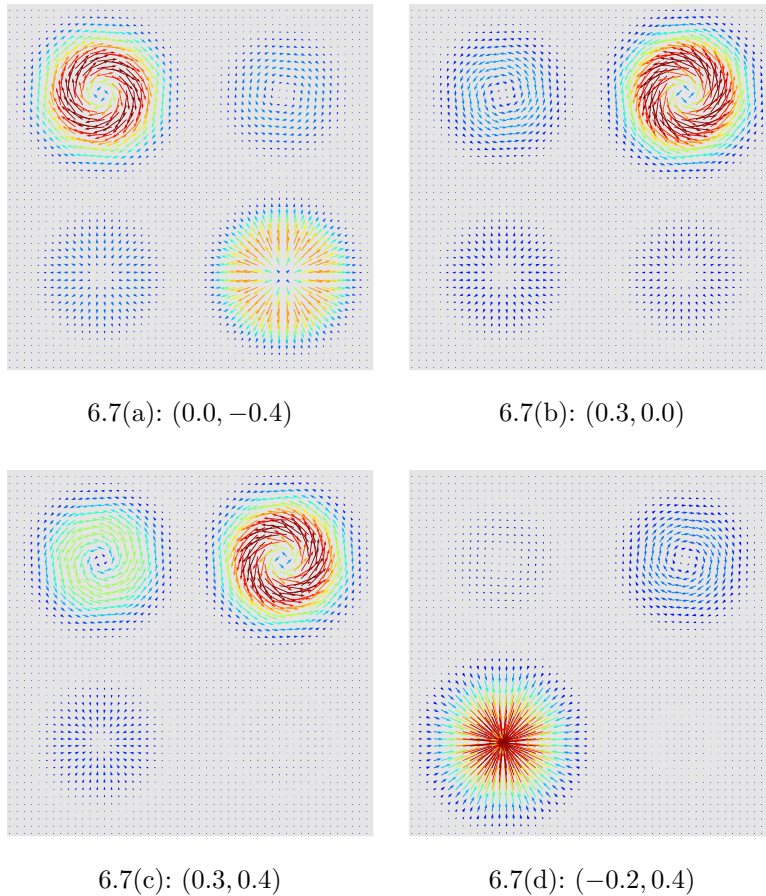
Figure 6.7: Inverse projection of selected points in the projection space.

6.4.2.0
Experiment 2

We generated a set of 100 points in the projection space of the original ensemble $\mathcal{E}$ as well as 100 points in the projection space for each of its components. All points were randomly chosen. After performing the inverse projection (Tables 6.1 and 6.2 shows some of the derived realizations), we obtained the curl and the divergence for each of the resulting vector fields to analyze the variability provided by this technique. As can be seen in Figure 6.8, the curl of the rotational-free realizations as well as the divergence of the divergence-free realizations are both zero, as expected. This means that the inverse projection is not generating any significant artifact in the resulting

vector fields. It can also be noted that the harmonic realizations provided the higher variability, followed by the use of the original ensemble.
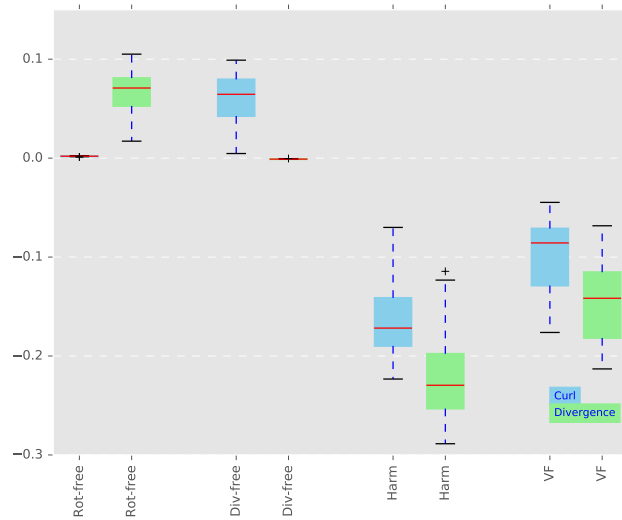
Figure 6.8: Statistics of the curl and divergence of the derived realizations.

We also measured the RMSE (Wackerly et al., 2008) of the curl and divergence between the derived realizations and the mean vector field of $\mathcal{E}$. These are depict in Figure 6.9. As can be seen, the divergence presented higher variability then the curl. This is expected once, for the wind forecast ensemble, the divergence-free component has more influence on the vector fields.
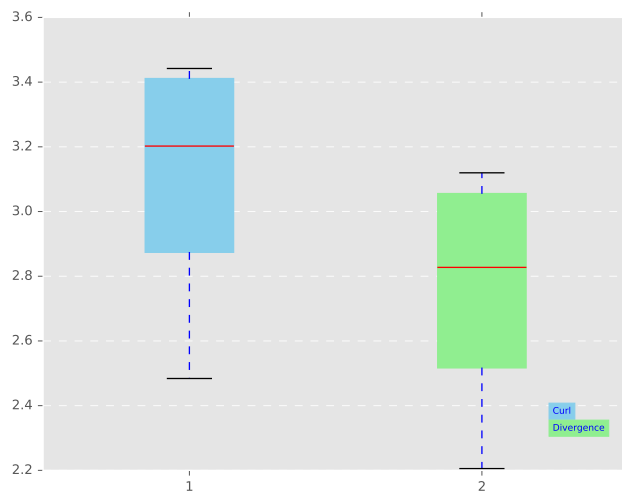
Figure 6.9: Statistics of the curl and divergence operators.

Table 6.1: Derived NHHD components.
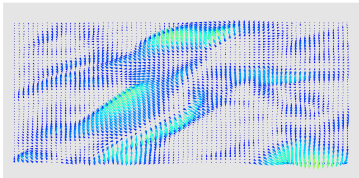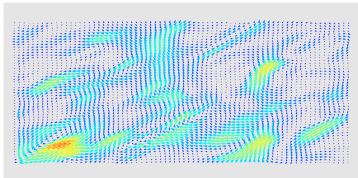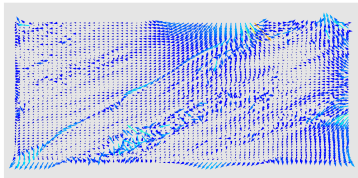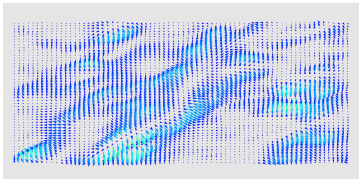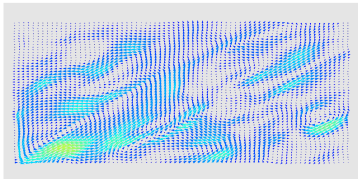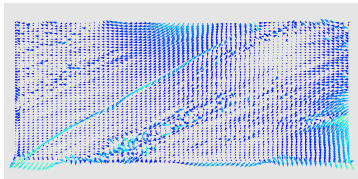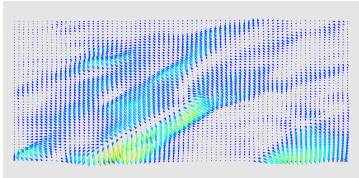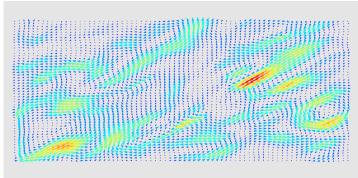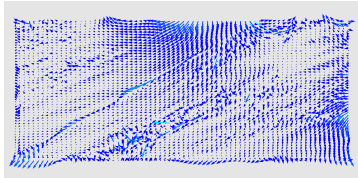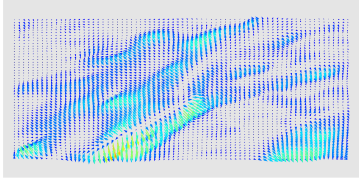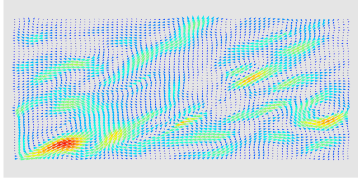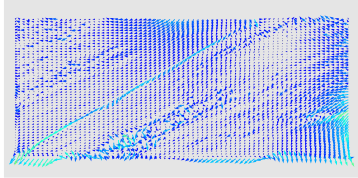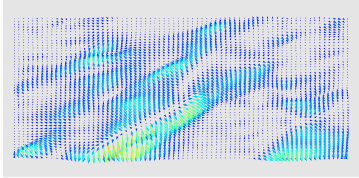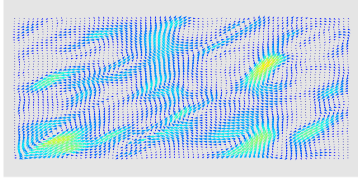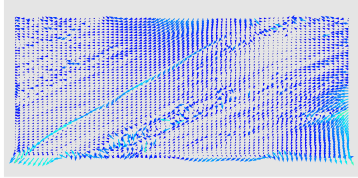
Rotational-free	Divergence-free	Harmonic
$[-0.64, -0.82]$	$[0.32, 0.86]$	$[0.77, -0.70]$
$[-0.29, 0.98]$	$[-0.06, -0.85]$	$[-0.15, 0.43]$
$[0.34, -0.04]$	$[-0.89, -0.37]$	$[-0.89, 0.97]$
$[0.44, -0.53]$	$[-0.74, 0.42]$	$[0.04, 0.24]$
$[0.21, 0.56]$	$[0.76, 0.78]$	$[-0.24, 0.3765]$

Table 6.2: Derived vector fields.

$[0.07, 0.35]$
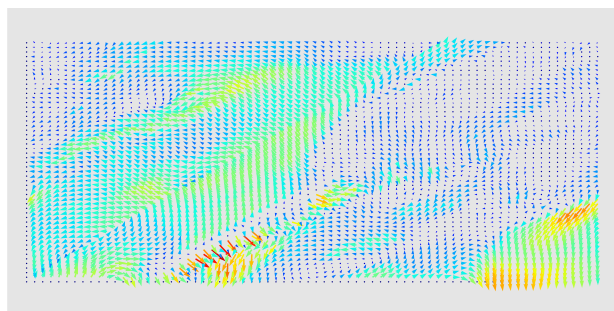
$[0.41, -0.67]$
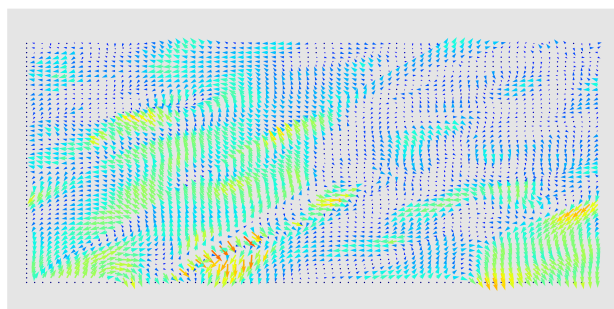
$[-0.1, 7 - 0.71]$
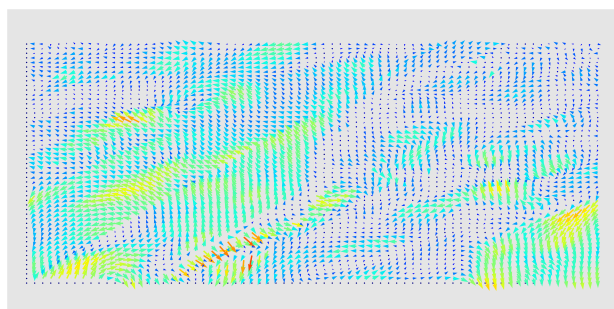
$[-0.66, 0.95]$
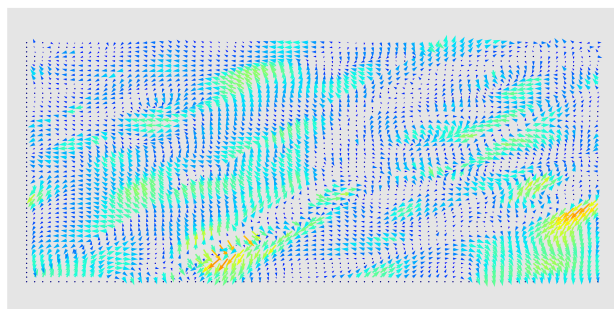
$[0.12, 0.56]$
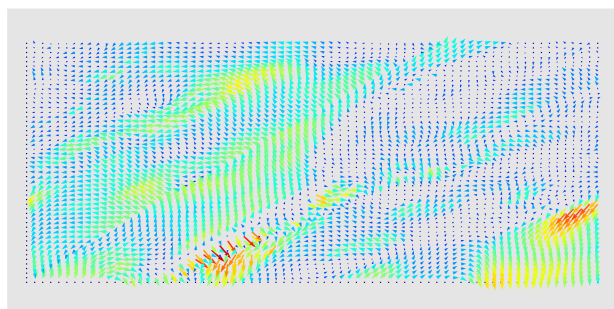
7
Performance

Here we present the performance of the proposed techniques to stochastically generate vector fields from a training data, as well as of the presented approach to synthesize vector fields from an ensemble data. In all cases, tests were performed using a machine running *ubuntu 16.04 LTS* with the configuration presented in Table 7.1.

Table 7.1: Machine configuration.

Memory	62.8 GiB
Processor	Intel® Core™ i7-5820K CPU @ 3.30 GHz ×12
Graphics	GeForce GTX 960/PCle/SSE2
OS Type	64 bit
Disk	55 GB

7.1
Stochastic Generation of Vector Fields

In the following we show the performance achieved with both methods described in Chapter 5.2.

7.1.1
Method 1: Bootstrap Based Stochastic Simulation

For each data set presented in this work, we measured the time necessary to compute the NHHD and to generate new realizations (as the mean of the time spent to generate a set with 100 new samples).

Table 7.2: Performance of the proposed method per sample, in seconds. Tested using λ equal to 90% for all scenarios.

	Forecast[1]	Navier-Stokes[2]	PIV[3]
NHHD	1025.775	86.497	1200.545
Samples Gen.	0.584	0.209	0.992

[1] 19×19 kernel.
[2] 5×5 kernel.
[3] 19×19 kernel.

As can be seen, the NHHD is the most time consuming step. However, this step can be optimized, parallelizing the calculus by each pixel in the data, for example. For more details on the performance of the NHHD, see (Pascucci et al., 2014).

We also tested the effect of different kernel sizes on the samples generation step. This is shown in Table 7.3.

Table 7.3: Performance of the sample generation step for different kernel sizes. Tested with the wind forecast ensemble mean and λ equal to 90%.

11×11	13×13	15×15	17×17
0.613	0.612	0.580	0.577

As we can observe, the size of the kernel didn't cause a significant change in the algorithm performance. It is also interesting to note that, the bigger the size of the kernel the lesser the time consumption. This means that the bootstrap step performance is mostly affected by the number of blocks chosen, instead of the size of the chosen kernel.

7.1.2
Method 2: Interpolation Based Stochastic Simulation

To measure the performance of this technique, we tested Algorithm 2 using the different values of p, as can be seen in Table 7.4.

Table 7.4: Performance, in milliseconds, of the interpolation based method using different p values.

p	1%	3%	5%	7%	9%
	102	120	129	135	154

7.2
Vector Field Synthesis

For the vector field synthesis method, presented in Chapter 6, we achieved the following performance results:

– To obtain the MDS of the wind forecast data took, on average, 19.8 milliseconds.

– To generate a new realization took, on average, 84.2 milliseconds to obtain the λ matrix.

– To compute the inverse projection took and 708 milliseconds.

Naturally, the bigger the vector fields dimension, the more time consuming these procedures will be.

For more details on the performance of this technique, see (Amorim et al., 2015).

8
Conclusion

This work explored the applicability of the Helmholtz-Hodge Decomposition on the uncertainty analyses of 2-dimensional vector fields, as well as on the stochastic generation of vector fields. Also, we proposed a technique to synthesize vector fields from an ensemble. In all cases, results were obtained for synthetic and real data. In the following, we are going to present some final remarks and possible future works.

Uncertainty Analysis of Vector Field Ensembles   Chapter 4 presented an approach to analyze 2D vector field ensembles by decomposing them on the basis of the Helmholtz-Hodge Decomposition. The main advantage of this approach is the ability to separately study the uncertainty of the curl-free, the divergence-free, and the harmonic component ensembles. We also proposed a way to quantify how much each of these component ensembles is correlated to the vector field's original ensemble at each point in the domain. To the best of our knowledge, this is the first work that proposes the use of HHD to help in visually analyzing 2D vector field ensembles.

Our approach was inspired by the works of Pfaffelmoser et al. (2013) and Mihai & Westermann (2014), which deal with uncertainty related to the gradient of scalar fields. In contrast, we consider a general kind of vector field.

We implemented the Natural HHD (Pascucci et al., 2014) in order to better evaluate the HHD on each element of the vector field ensemble. Unlike other techniques that impose boundary conditions to achieve a unique solution, the NHHD does not guarantee the components to be L_2-orthogonal. Although this technique removes the artifacts introduced by the boundary conditions during the decomposition process, providing orthogonality may be crucial to some applications.

The presented approach is extensible to 3D vector fields, since the NHHD can be computed using the appropriate formulae defined by Pascucci et al. (2014), whereas the point-wise correlation analysis simply uses the canonical inner product of two vectors to measure similarity. However, obtaining such decomposition in a large 3D vector field ensemble could be computationally very expensive.

For time-dependent vector field ensembles, we can assume temporal coherence and explore the proposed approach at each time step. As a consequence, the temporal dependence between the component ensembles could be an interesting subject for further investigation.

The correlation analysis shows to be a useful tool to identify regions where a set of vector field realizations is dominated by an extensional flow (rotational equals to zero) or regions where it behaves like an incompressible flow (divergent equals to zero). Thus, it can improve the understanding of the flow kinematics, which is a requirement for the rheologists to establish or to improve constitutive equations. It is in our plan to explore the use of such tool in sampled complex flow analysis with the supervision of a domain expert.

Regarding the wind forecasting analysis, we have an objective to find wind data ensembles on which the uncertainty analysis could reveal a relevant finding with the help of a meteorologist.

In terms of the methodology, we would like to study uncertainty quantification methods that will access the uncertainty of critical points locations in vector fields based on the HHD. We also aim to use the *Multified-Graph* approach (Sauber et al., 2006) to better analyze the correlation between the magnitude and orientation uncertainties among the components. Finally, we plan to quantify the spatial dependencies in each component and measure their relationship with the original ensemble.

Stochastic Generation of Vector Fields In Chapter 5 we proposed a technique to stochastic generate vector fields given a single realization. Such an approach make use of the Helmholtz-Hodge Decomposition and a Bootstrap-based procedure. Results were evaluated using a set of multi-method wind forecast realizations, as well as simulations from Navier-Stokes and PIV. For each data, 100 new scenarios were generated using the presented method. We applied the *MDS* technique to proper visualize the results; we could observe that the simulated scenarios were able to provide a great variability and that they mimic the training data.

To the best of our knowledge, this is the first approach that uses the Helmholtz-Hodge Decomposition to stochastic generate vector fields given a training data.

The applicability of this approach ranges from uncertainty quantification to data assimilation (Kalnay, 2003).

Further studies includes expanding this method for 3-dimensional vector fields. We also would like to explore techniques to measure the spacial correlation, possibly using this correlation to obtain, in an automated way, the size

of the Bootstrap kernel that best fits the input data.

Synthesis of Vector Fields   In Chapter 6 we proposed the application of dimensionality reduction and inverse projection to generate 2D vector fields given an ensemble. Results show that this approach is capable of generating realizations with some variability whilst respecting the given ones. Moreover, through the Helmholtz-Hodge Decomposition, one can generate scenarios choosing different weights for the final vector field divergence-free, rotational-free and harmonic components. This leads to a wide range of final possibilities for vector field realizations.

Of course, this technique could be use to generate random realizations of vector fields based on the ensemble. To do so, we need only to generate a random point in $\mathbb{R}^2$ and perform the inverse projection.

For the future, we intend to test this approach with a multidimensional projection that uses control points, as the LAMP (Joia et al., 2011), for example. We also plan to apply this technique to 2D and 3D temporal vector fields.

Bibliography

Abadiea, S., Morichona, D., Grillib, S. & Glockner, S. (2010), 'Numerical simulation of waves generated by landslides using a multiple-fluid Navier–Stokes model', *Coastal Engineering* **57**(9), 779—-794.

Adler, R. J. (1981), *The geometry of random fields*, Vol. 62, Siam.

Amorim, E., Vital Brazil, E., Mena-Chalco, J., Velho, L., Nonato, L. G., Samavati, F. & Costa Sousa, M. (2015), 'Facing the high-dimensions: Inverse projection with radial basis functions', *Computer & Graphics* **48**(C), 35–47. DOI `10.1016/j.cag.2015.02.009`. ISSN 0097-8493.

Anderson, J. D. & Wendt, J. (1995), *Computational fluid dynamics*, Vol. 206, Springer.

Beccali, M., Cellura, M. & Mistretta, M. (2003), 'Decision-making in energy planning. application of the electre method at regional level for the diffusion of renewable energy technology', *Renewable Energy* **28**(13), 2063–2087.

Bertalmio, M., Bertozzi, A. L. & Sapiro, G. (2001), Navier-stokes, fluid dynamics, and image and video inpainting, *in* 'Proceedings of the 2001 IEEE Computer Society Conference', IEEE, pp. 355–362.

Bhatia, H., Norgard, G., Pascucci, V. & Bremer, P.-T. (2013), 'The Helmholtz-Hodge Decomposition - A Survey', *IEEE Transactions on Visualization and Computer Graphics* **19**(8), 1386–1404.

Blažica, V., Žagar, N., Strajnar, B. & Cedilnik, J. (2013), 'Rotational and divergent kinetic energy in the mesoscale model aladin', *Tellus A*.

Botchen, R. P., Weiskopf, D. & Ertl, T. (2005), Texture-based visualization of uncertainty in flow fields, *in* 'Visualization, 2005. VIS 05. IEEE', IEEE, pp. 647–654.

Casaca, W., Nieto, E. G., Ferreira, C. O. L., Tavares, G., Pagliosa, P. A., Paulovich, F. V., Nonato, L. G. & Paiva, A. (2012), Colorization by multidimensional projection, *in* '25th SIBGRAPI Conference on Graphics, Patterns and Images', pp. 32–38. DOI `10.1109/SIBGRAPI.2012.14`.

Chen, T. C. (1980), 'On the Energy Exchange between the Divergent and Rotational Components of Atmospheric Flow over the Tropics and Subtropics at 200 mb during Two Northern', *Monthly Weather Review* **108**(7), 896–912.

Chen, T. C. & Tribbia, J. J. (1981), 'Kinetic energy spectra of divergent wind in the atmosphere', *Tellus* **33**(1), 102–104.

Cho, K., Meer, P. & Cabrera, J. (1997), 'Performance assessment through bootstrap', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **19**(11), 1185–1198.

Chorin, A. J. (1968), 'Numerical solution of the Navier-Stokes equations', *Mathematics of Computation* **22**, 745–762.

Cox, T. F. & Cox, M. (2000), *Multidimensional Scaling, Second Edition*, 2 ed., Chapman and Hall/CRC.

Croux, C. & Dehon, C. (2010), 'Influence functions of the Spearman and Kendall correlation measures', *Statistical methods & applications* **19**(4), 497–515.

de Silva, C. M., Philip, J. & Marusic, I. (2013), 'Minimization of divergence error in volumetric velocity measurements and implications for turbulence statistics', *Experiments in fluids* **54**(7), 1–17.

Deriaz, E. & Perrier, V. (2009), 'Orthogonal Helmholtz decomposition in arbitrary dimension using divergence-free and curl-free wavelets', *Applied and Computational Harmonic Analysis* **26**(2), 249–269.

do Carmo, F. P. (2008), A Equação de Poisson e a Decomposição de Helmholtz-Hodge com Operadores SPH, Phd thesis, Pontifical Catholic University of Rio de Janeiro.

Efron, B. (1979), 'Bootstrap Methods: Another Look at the Jackknife', *Annals of Statistics* **7**, 1–26.

Efros, A. A. & Freeman, W. T. (2001), Image quilting for texture synthesis and transfer, *in* 'Proceedings of the 28th annual conference on Computer graphics and interactive techniques', ACM, pp. 341–346.

Forsberg, A., Chen, J. & Laidlaw, D. H. (2009), 'Comparing 3d vector field visualization methods: A user study', *Visualization and Computer Graphics, IEEE Transactions on* **15**(6), 1219–1226.

Fu, X., You, H. & Fu, K. (2012), 'A Statistical Approach to Detect Edges in SAR Images Based on Square Successive Difference of Averages', *IEEE Geoscience And Remote Sensing Letters* **9**(6), 1094–1098.

Gao, H., Mandal, M. K., Guo, G. & Wan, J. (2010), Singular point detection using Discrete Hodge Helmholtz Decomposition in fingerprint images, *in* 'ICASSP'10', IEEE, pp. 1094–1097.

Gomez, J. C. & Moens, M.-F. (2012), '{PCA} document reconstruction for email classification', *Computational Statistics & Data Analysis* **56**(3), 741 – 751. DOI `http://dx.doi.org/10.1016/j.csda.2011.09.023`. ISSN 0167-9473.

Gonzalez, R. C. & Woods, R. E. (2006), *Digital Image Processing*, Prentice-Hall, Inc.

Guo, Q., Mandal, M. K., Liu, G. & Kavanagh, K. M. (2006), 'Cardiac video analysis using Hodge-Helmholtz field decomposition', *Computers in Biology and Medicine* **36**(1), 1–20.

Hinkle, J., Fletcher, P. T., Wang, B., Salter, B. & Joshi, S. (2009), 4D MAP image reconstruction incorporating organ motion, *in* 'Information Processing in Medical Imaging', Springer, pp. 676–687.

Holton, J. R. & Hakim, G. J. (2012), *An introduction to dynamic meteorology*, Vol. 88, Academic press.

Honarkhah, M. & Caers, J. (2010), 'Stochastic Simulation of Patterns Using Distance-Based Pattern Modeling', *Mathematical Geosciences* **42**(5), 487–517.

Huang, Q., Hassager, O., Skov, A. L. & Rasmussen, H. K. (2013), *Molecular Rheology of Complex Fluids*, Technical University of DenmarkTechnical University of Denmark.

Jen, D., Parente, P., Robbins, J., Weigle, C., Taylor, R. M., Burette, A., Weinberg, R. et al. (2004), Imagesurfer: A tool for visualizing correlations between two volume scalar fields, *in* 'Visualization, 2004. IEEE', IEEE, pp. 529–536.

Johnson, C. R. & Sanderson, A. R. (2003), 'A Next Step: Visualizing Errors and Uncertainty', *IEEE Computer Graphics and Applications* **23**(5), 6–10.

Joia, P., Coimbra, D. B., Cuminato, J. A., Paulovich, F. V. & Nonato, L. G. (2011), 'Local affine multidimensional projection', *IEEE Transactions on Visualization and Computer Graphics* **17**(12), 2563–2571. DOI `10.1109/TVCG.2011.220`.

Jolliffe, I. (1986), *Principal Component Analysis*, Springer Verlag.

Kalnay, E. (2003), *Atmospheric modeling, data assimilation and predictability*, Cambridge university press.

Kim, K. & Lee, J. (2014), 'Sentiment visualization and classification via semi-supervised nonlinear dimensionality reduction', *Pattern Recogn.* **47**(2), 758–768. DOI `10.1016/j.patcog.2013.07.022`. ISSN 0031-3203.

Kruskal, J. B. (1964), 'Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis', *IEEE Transactions on Visualization and Computer Graphics* **29**(1), 1–27.

Laccarino, G. (2009), Quantification of uncertainty in flow simulations using probabilistic methods, Technical report, DTIC Document.

Laidlaw, D. H., Kirby, R. M., Jackson, C. D., Davidson, J. S., Miller, T. S., da Silva, M., Warren, W. H. & Tarr, M. J. (2005), 'Comparing 2D Vector Field Visualization Methods: A User Study', *IEEE Transactions on Visualization and Computer Graphics* **11**(1), 59–70.

Lall, U., Devineni, N. & Kaheil, Y. (2016), 'An empirical, nonparametric simulator for multivariate random variables with differing marginal densities and nonlinear dependence with hydroclimatic applications', *Risk Analysis* **36**(1), 57–73.

Lantuéjoul, C. (2013), *Geostatistical simulation: models and algorithms*, Springer Science & Business Media.

Liew, K. J., Ramli, A. & Majid, A. A. (2016), 'Searching for the optimum value of the smoothing parameter for a radial basis function surface with feature area by using the bootstrap method', *Computational and Applied Mathematics* pp. 1–16.

Liu, L. & Fieguth, P. W. (2012), 'Texture classification from random features', *Pattern Analysis and Machine Intelligence* **34**(3), 574–586. DOI `10.1109/TPAMI.2011.145`.

Lodha, S. K., Pang, A., Sheehan, R. E. & Wittenbrink, C. M. (1996), UFLOW: Visualizing uncertainty in fluid flow, *in* 'Visualization'96. Proceedings.', IEEE, pp. 249–254.

Luo, C., Safa, I. & Wang, Y. (2012), 'Feature-aware streamline generation of planar vector fields via topological methods', *Computers & Graphics* **36**(6), 754–766.

M., A. & Michel, V. (2010), 'Regularisation of the Helmholtz decomposition and its application to geomagnetic field modelling', *International Journal on Geomathematics* **1**(1), 101–120.

Mariethoz, G. & Caers, J. (2014), *Multiple-point geostatistics: stochastic modeling with training images*, John Wiley & Sons.

Mariethoz, G. & Lefebvre, S. (2014), 'Bridges between multiple-point geostatistics and texture synthesis: Review and guidelines for future research', *Computers & Geosciences* **66**, 66–80.

Martins, R. M., Andery, G. F., Heberle, H., Paulovich, F. V., de Andrade Lopes, A., Pedrini, H. & Minghim, R. (2012), 'Multidimensional projections for visual analysis of social networks', *Journal of Computer Science and Technology* **27**(4), 791–810. DOI 10.1007/s11390-012-1265-5.

Mihai, M. & Westermann, R. (2014), 'Visualizing the stability of critical points in uncertain scalar fields', *Computers & Graphics* **41**(0), 13–25.

Monaghan, J. J. (1992), 'Smoothed particle hydrodynamics', *Annual review of astronomy and astrophysics* **30**, 543–574.

Oliver, M. & Webster, R. (2014), 'A tutorial guide to geostatistics: Computing and modelling variograms and kriging', *Catena* **113**, 56–69.

Otto, M., Germer, T., Hege, H.-C. & Theisel, H. (2010), 'Uncertain 2D Vector Field Topology', *Computer Graphics Forum* **29**(2), 347–356.

Palit, B. (2005), Application of the Hodge Helmholtz Decomposition to Video and Image Processing, Master's thesis, University of Alberta.

Pascucci, V., Bremer, P.-T. & Bhatia, H. (2014), 'The Natural Helmholtz-Hodge Decomposition For Open-Boundary Flow Analysis', *IEEE Transactions on Visualization and Computer Graphics* **99**(PrePrints), 1–11.

Petronetto, F., Paiva, A., Lage, M., Tavares, G., Lopes, H. & Lewiner, T. (2010), 'Meshless Helmholtz-Hodge Decomposition', *IEEE Transactions on Visualization and Computer Graphics* **16**(2), 338–349.

Petz, C., Pöthkow, K. & Hege, H.-C. (2012), 'Probabilistic local features in uncertain vector fields with spatial correlation', *Computer Graphics Forum* **31**(3pt2), 1045–1054.

Pfaffelmoser, T. & Westermann, R. (2012), Visualization of global correlation structures in uncertain 2d scalar fields, *in* 'Computer Graphics Forum', Vol. 31, Wiley Online Library, pp. 1025–1034.

Pfaffelmoser, T. & Westermann, R. (2013), 'Correlation visualization for structural uncertainty analysis', *International Journal for Uncertainty Quantification*.

Pfaffelmoser, T., Mihai, M. & Westermann, R. (2013), 'Visualizing the variability of gradients in uncertain 2d scalar fields', *IEEE transactions on visualization and computer graphics* **19**(11), 1948–1961.

Polthier, K. & Preuß, E. (2003), Identifying Vector Field Singularities Using a Discrete Hodge Decomposition, *in* 'Visualization and Mathematics III', Springer Berlin Heidelberg, pp. 113–134.

Popescu, R., Deodatis, G. & Prevost, J. H. (1998), 'Simulation of homogeneous nongaussian stochastic vector fields', *Probabilistic Engineering Mechanics* **13**(1), 1–13.

Potter, K., Gerber, S. & Anderson, E. W. (2013), 'Visualization of Uncertainty Without a Mean', *IEEE Computer Graphics and Applications* **33**(1), 75–79.

Potter, K., Rosen, P. & Johnson, C. R. (2012), From quantification to visualization: A taxonomy of uncertainty visualization approaches, *in* 'Uncertainty Quantification in Scientific Computing', Springer, pp. 226–249.

Ribeiro, P. C., de Campos Velho, H. F. & Lopes, H. (2016), 'Helmholtz-Hodge decomposition and the analysis of 2D vector field ensembles', *Computers & Graphics* **55**, 80–96.

Sadati, M., Luap, C., Kröger, M. & Öttinger, H. C. (2011), 'Hard vs soft constraints in the full field reconstruction of incompressible flow kinematics

from noisy scattered velocimetry data', *Journal of Rheology (1978-present)* **55**(6), 1187–1203.

Sagis, L. M. & Fischer, P. (2014), 'Nonlinear rheology of complex fluid–fluid interfaces', *Current Opinion in Colloid & Interface Science* **19**(6), 520–529.

Sauber, N., Theisel, H. & Seidel, H.-P. (2006), 'Multifield-graphs: An approach to visualizing correlations in multifield scalar data', *Visualization and Computer Graphics, IEEE Transactions on* **12**(5), 917–924.

Schroff, F., Kalenichenko, D. & Philbin, J. (2015), Facenet: A unified embedding for face recognition and clustering, *in* 'The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)'.

Tong, Y., Lombeyda, S., Hirani, A. N. & Desbrun, M. (2003), Discrete Multiscale Vector Field Decomposition, *in* 'ACM SIGGRAPH 2003 Papers', ACM, pp. 445–452.

van der Maaten, L. & Hinton, G. (2008), 'Visualizing High-Dimensional Data Using t-SNE', *Journal of Machine Learning Research* **9**, 2579–2605.

van Oudheusden, B. W., Scarano, F., Roosenboom, E. W., Casimiri, E. W. & Souverein, L. J. (2007), 'Evaluation of integral forces and pressure fields from planar velocimetry data for incompressible and compressible flows', *Experiments in Fluids* **43**(2-3), 153–162.

Viard, T., Caumon, G. & Lévy, B. (2011), 'Adjacent versus coincident representations of geospatial uncertainty: Which promote better decisions?', *Computers & Geosciences* **37**(4), 511–520.

Wackerly, D. D., Mendenhall, W. & Scheaffer, R. L. (2008), *Mathematical Statistics with Applications*, Vol. 7th, Thomson.

Wang, H. & Deng, J. (2014), Feature extraction of complex ocean flow field using the helmholtz-hodge decomposition, *in* '2013 IEEE International Conference on Multimedia and Expo Workshops', IEEE, pp. 1–6.

Ware, C. (2013), *Information visualization: perception for design*, Elsevier.

Wasserman, L. (2004), *All of Statistics: A Concise Course in Statistical Inference*, Springer.

Wiebel, A. (2004), Feature detection in vector fields using the Helmholtz-Hodge decomposition, Master's thesis, University of Kaiserslautern.

Wittenbrink, C. M., Pang, A. T. & Lodha, S. K. (1996), 'Glyphs for visualizing uncertainty in vector fields', *Visualization and Computer Graphics, IEEE Transactions on* **2**(3), 266–279.

Xiu, D. (2009), 'Fast numerical methods for stochastic computations: a review', *Communications in computational physics* **5**(2-4), 242–272.