Atmospheric Lagrangian transport structures and their applications to aerobiology

Amir Ebrahim Bozorg Magham

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Engineering Mechanics

Shane D. Ross
David G. Schmale III
Mark R. Paul
Mark A. Stremler
Traian Iliescu

February 5, 2014
Blacksburg, Virginia

Keywords: Finite-time Lyapunov exponent (FTLE), Lagrangian coherent structure (LCS), chaotic atmospheric transport, unresolved turbulence, stochastic FTLE field, ensemble forecasting, uncertainty analysis, local FTLE time-series, maximal diversity monitoring

Atmospheric Lagrangian transport structures and their applications to aerobiology

Amir Ebrahim Bozorg Magham

ABSTRACT

Exploring the concepts of long range aerial transport of microorganisms is the main motivation of this study. For this purpose we use theories and concepts of dynamical systems in the context of geophysical fluid systems. We apply powerful notions such as finite-time Lyapunov exponent (FTLE) and the associated Lagrangian coherent structures (LCS) and we attempt to provide mathematical explanations and frameworks for some applied questions which are based on realistic concerns of atmospheric transport phenomena. Accordingly, we quantify the accuracy of prediction of FTLE-LCS features and we determine the sensitivity of such predictions to forecasting parameters. In addition, we consider the spatiotemporal resolution of the operational data sets and we propose the concept of probabilistic source and destination regions which leads to the definition of stochastic FTLE fields. Moreover, we put forward the idea of using ensemble forecasting to quantify the uncertainty of the forecast results. Finally, we investigate the statistical properties of localized measurements of atmospheric microbial structure and their connections to the concept of local FTLE time-series.

Results of this study would pave the way for more efficient models and management strategies for the spread of infectious diseases affecting plants, domestic animals, and humans.

Dedication

To my Mom and Dad,

and to my lovely wife Parisa who is my best friend

Acknowledgments

There are some people who made this journey more attractive and more intellectually satisfying with words of encouragement. I offer my regards and blessings to all of those who supported me in any respect during the completion of this work.

I would like to take this opportunity to express my sincerest gratitude to my exceptional advisor and committee chair, Dr. Shane Ross, for all his special supports during these years at Virginia Tech. His constant encouragements to explore new ideas and commitment to research while allowing me the room to work in my own way have been an inspiration for me. His patience with my special case of life has been crucial to my intellectual development.

I would also like to thank my committee co-chair, Dr. David Schmale III, for giving me the opportunity to work on the interesting problem of atmospheric transport of microorganisms. This tangible problem which could help improving the standard of life of the human beings is a motivation for me to explore the concepts of dynamical systems from a very practical point of view.

I would also like to thank my committee members. Dr. Mark Paul for his enthusiasm and encouragement both in and out of the class which helped me to explore many of the theoretical concepts used in this work. Dr. Mark Stremler for his excellent style of teaching, especially in advance fluid mechanics course, which reconnect me to the world of fluids after

years of working in control and design fields. And Dr. Traian Iliescu for his suggestions, advises and questions that helped me to polish the thesis.

My special appreciation goes to Lisa Smith, who has helped me with all the administrative procedures in ESM.

I would like to show my gratitude to my family, who have always given me their all and stood by my side through every chapter of my life. Their constant encouragements of intellectual pursuits have inspired me towards a scientific point of view which is the goal of any educational system.

Finally, I especially thank my wife Parisa, the joy of my life, who has been a source of love, inspiration and support. Without her patience I have not been able to take this long journey to this point.

Attribution

This dissertation is composed of an introduction, four main chapters and a conclusion. The four chapters are written in journal format and either have or will be submitted to archival journals for separate publication. The following individuals are co-authors on one or more chapters.

Dr. Shane D. Ross, Ph.D., Associate Professor in department of Engineering Science and Mechanics (ESM), Virginia Tech. Served as the committee chair and provided advice on the works presented in each chapter. Dr. Ross is a coauthor on each of the papers.

Dr. David G. Schmale III, Ph.D., Associate Professor in department of Plant Pathology, Physiology, and Weed Science, Virginia Tech. Served as the committee co-chair and provided advice on the works presented in each chapter. Dr. Schmale is a coauthor on chapters 2, 4 and 5.

Dr. Binbin Lin, Ph.D., department of Plant Pathology, Physiology, and Weed Science, Virginia Tech. Dr. Lin is a coauthor on chapter 4.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This study is inspired by aerial transport phenomena of microbial populations of a significant group of fungi known as *Fusarium*. This group of fungi contains a number of important plant and animal pathogens and they cause huge annual losses in agricultural and domestic animal industries. Thus, for having a sustainable and healthy food cycle it is essential to have a deep understanding about them. For this aim, there are many biological and aerobiological researches associated to fusaria to describe their genetic diversities, life cycles, related infections and diseases and other connected subjects. An important part of the life cycle of fusaria is their aerial transport from their source points to their target habitats. Research results show that this transport could be in different ranges, from small scales (meters, farm size) to long range mesoscale distances. Focus of this study is on long range transport phenomena of passive tracers such as fusaria.

There are some traditional methods for this purpose, e.g., backward and forward trajectory calculations which shows spaghetti of the particles' pathlines (spores). This approach does not provide any big picture about the underlying dynamics and for each case it is necessary to repeat all the calculations. To avoid this drawback we use some fundamental concepts

of dynamical systems which give us a big picture of the system and the motion of particles along important pathways which govern mixing and transport of passive particles. For this purpose we mainly use the concept of finite time Lyapunov exponent (FTLE) and the associated hyperbolic Lagrangian coherent structures (LCSs). Tersely, FTLE shows the separation rate between nearby particles in a moving fluid flow and LCS are the material surfaces with maximum attraction or repulsion rates in non-autonomous dynamical systems. These features are analog to invariant stable and unstable manifolds in ODE systems.

Thus, the main objective of this work is to explore some applied questions regarding the aerial transport of passive microorganisms in geophysical (atmospheric) fluid domain. This work is outlined as follows.

In Chapter 2 we investigate the forecast results of FTLE fields and LCS features with respect to various forecast parameters. We specifically explore questions such as: how accurate and precise are forecast FTLE-LCSs? And what are the quantitative methods for comparing the forecast-based with the archive-based features? Also, what are the effective parameters on the quality of FTLE-LCSs forecasting? Answering these questions would be important when we want to apply this approach to real situations, for example, to predict the incursion of a high threat plant pathogen into susceptible regions from a distant source.

In Chapter 3 we use a Lagrangian particle dispersion model to determine the probabilistic source (destination) regions and we introduce the concept of stochastic FTLE fields. In addition, we show that the spatiotemporal dependence of the stochastic velocity component to the time–varying deterministic background velocity field plays an important role in the determination of the probabilistic distribution of the end–position of particles and consequently on the stochastic FTLE fields. Also, we propose a practical approach to measure uncertainty of forecast FTLE–LCSs by using ensemble forecasting concepts.

In Chapter 4 we investigate the results of localized samplings of fusaria to determine how statistically their collections vary between consecutive samplings in short period of time. This work is essential for understanding whether changes in the recovery of fusaria in the lower atmosphere may be attributed to large-scale transport phenomena.

In Chapter 5 we study the concept of local FTLE and successive (aerial) particle samplings at a fixed location. The proposed theorem in this chapter helps us to have a better understanding about the distribution of the source points of sampled particles and also to investigate the long range transport phenomena as a possible cause of abrupt characteristic changes among the successive collected samples. We also consider the spatiotemporal limitations of the velocity field data and the uncertainties of the location of the source and destination points. Finally, we show how the notion of local FTLE could be used in cases that unresolved turbulence has considerable importance.

Our contributions to the literature are:

- We quantify the sensitivity of FTLE-LCS predictions with respect to the involved forecasting parameters of operational data sets,
- We show the effects of unresolved turbulence on probabilistic source/destination regions, also we distinguish two numerical procedures for calculations of those probabilistic regions,
- We propose a definition of stochastic FTLE field,
- We introduce the ensemble FTLE fields as a practical method for uncertainty analysis of forecast FTLE-LCSs,
- We investigate the statistics of the consecutive samples of *Fusarium* and compute the autocorrelation coefficient of the colony counts,

- We present a new interpretation about the local FTLE and we show its applications for describing the differential distances between the source/destination points and also for maximal-diversity monitoring purposes.

The following chapters are presented in journal format with the second chapter accepted into archival publication in Physica D, nonlinear phenomena and the forth chapter accepted into archival publication in Aerobiologia. The remaining chapters have been or are in the process of being submitted for publication.

We place all the references corresponding to different chapters at the end of this document.

Chapter 2

Real-time prediction of atmospheric Lagrangian coherent structures based on forecast data: An application and error analysis

Note: The following chapter was formatted to facilitate publication in Physica D Nonlinear Phenomena. This work was originally published by Bozorg Magham, Ross, and Schmale in Volume 258, September 2013, pp.47-60. (DOI doi:10.1016/j.physd.2013.05.003) of Physica D Nonlinear Phenomena.

Abstract

The language of Lagrangian coherent structures (LCSs) provides a new means for studying

transport and mixing of passive particles advected by an atmospheric flow field. Recent observations suggest that LCSs govern the large-scale atmospheric motion of airborne microorganisms, paving the way for more efficient models and management strategies for the spread of infectious diseases affecting plants, domestic animals, and humans. In addition, having reliable predictions of the timing of hyperbolic LCSs may contribute to improved aerobiological sampling of microorganisms with unmanned aerial vehicles and LCS-based early warning systems. Chaotic atmospheric dynamics lead to unavoidable forecasting errors in the wind velocity field, which compounds errors in LCS forecasting. In this study, we reveal the cumulative effects of errors of (short-term) wind field forecasts on the finite-time Lyapunov exponent (FTLE) fields and the associated LCSs when realistic forecast plans impose certain limits on the forecasting parameters. Objectives of this paper are to (a) quantify the accuracy of prediction of FTLE-LCS features and (b) determine the sensitivity of such predictions to forecasting parameters. Results indicate that forecasts of attracting LCSs exhibit less divergence from the archive-based LCSs than repelling features. This result is important since attracting LCSs are the backbone of long-lived features in moving fluids. We also show under what circumstances one can trust the forecast results if one merely wants to know if an LCS passed over a region and does not need to precisely know the passage time.

2.1 Introduction

The emergence of Lagrangian coherent structures (LCSs) along with related concepts from dynamical systems theory during the past decade has aided in providing a better understanding of the geometric mechanisms of transport and mixing of particles in moving fluids (Lekien et al., 2005; Haller and Poje, 1998; Haller and Yuan, 2000; Lekien et al., 2007; Shadden et al., 2005; Olascoaga and Haller, 2012; Vogel et al., 2011). Hyperbolic LCSs are

material surfaces with maximum attraction or repulsion rates in non-autonomous dynamical systems as viewed in the context of extended phase space (Haller, 2011) and can be considered analog to invariant stable and unstable codimension one manifolds.

One of the novel applications of this new concept is in the study of long range transport of airborne microorganisms passively advected by atmospheric flow (Tallapragada et al., 2011; Schmale et al., 2012; Lin et al., 2013). The diseases borne by microorganisms can have severe economical and ecological effects. For example, Hurricane Ivan brought soybean rust from South America to the Gulf coast of the U.S. (Pan et al., 2006; Schneider et al., 2005), and long distance transport is believed to play a role in the dispersal of tobacco blue mold in the U.S. (Aylor et al., 1982). Experimental verification of long distance transport of airborne microorganisms from known sources is underway (Prussin et al., 2013). Given the biological and economical importance of invasive plant diseases, there is an increasing interest in predicting the distribution patterns of pathogens. This will be beneficial for better informing national and local managements. For example, the USDA Pest Information Platform for Extension and Education is an extensive program which provides the latest information on soybean rust and soybean aphid including observations, management recommendations, and scouting information (http://sbr.ipmpipe.org).

It has been hypothesized that LCSs may play an important role in the long distance and non-uniform spread of microbes (Tallapragada et al., 2011). This assumption is built on the essential properties of the repelling and attracting LCSs which act as the backbone of mixing, providing moving partitions of the fluid domain into regions which move coherently (Haller and Sapsis, 2011; Haller and Yuan, 2000). This hypothesis has been tested using autonomous unmanned aerial vehicles (UAVs) equipped with microbe-sampling devices to collect viable spores of a specific group of fungi known as *Fusarium* (Schmale et al., 2012). This group of fungi contains a number of important plant and animal pathogens (Leslie and

Summerell, 2006). Collections were performed during daylight hours at random times over a fixed geographic location. Counts of *Fusarium* were compared to hyperbolic LCS features obtained from archived wind velocity data from the NAM-218 model. Statistical results from several sampling flights show that when a punctuated change in spore concentration occurs, there is a high probability that an LCS passes over the sampling location between the two sampling times (Tallapragada et al., 2011).

Sampling flights are costly experiments in terms of equipment and personnel and have been performed without any forecast of LCSs. While this approach avoids certain biases in the data collection for initial hypothesis testing, a more efficient strategy would be to choose the sampling times with respect to the expected passage times of LCSs over the sampling location. Thus, instead of selecting arbitrary sampling times for UAVs, one can use wind velocity forecast data to predict LCS features. This would enable investigators (such as ourselves) to optimize a sampling strategy to collect samples on either side of a predicted LCS.

Another motivation for using forecast results is for risk assessment. If the association of LCSs and long-range transport of agricultural pathogens holds up under further scrutiny, then predictions of LCSs can be incorporated in management strategies, i.e., by short-term prediction based on wind data. If the differences between the predicted and true wind velocity field are small in the sense of Haller (2002), then hyperbolic LCSs will be topologically stable, and the strategy of using wind forecast data is sound.

In this context, we seek to study the accuracy of forecast LCS features. In this paper, our results reveal the cumulative effects of errors of wind field forecasts on the finite-time Lyapunov exponent (FTLE) fields and related LCS patterns (Olcay et al., 2010).

We specifically investigate questions such as: How accurate and precise are forecast FTLE-

LCSs? And what are the quantitative methods for comparing the forecast-based with the archive-based features? Also, what are the effective parameters on the quality of FTLE-LCSs forecasting? Answering these questions would be vital when we want to apply this approach to real situations, for example, to predict the incursion of a high threat plant pathogen into susceptible regions from a distant source (Madden and Wheelis, 2003).

The paper is outlined as follows. In § 2.2, we discuss the conceptual and mathematical definitions of the FTLE field, ridges and hyperbolic repelling and attracting LCSs. In § 2.3, we discuss the time scheduling of real-time extraction of hyperbolic LCSs from NAM-218 data set. In § 2.4, we study the errors of wind field forecasts as the main cause of imprecise FTLE-LCS predictions. In § 2.5, we propose five methods for quantitative comparison of forecast-based with archive-based LCSs with respect to effective parameters (pointwise comparison of the FTLE field, 2D cross correlation, proper orthogonal decomposition (POD) method (Antoulas, 2005; Sirovich, 1987; Rowley, 2005), composite correlation filter (Kumar et al., 2005) and modified Hausdorff distance method (Dubuisson and Jain, 1994)). In § 2.6, we make conclusion of the comparison methods and the results from each of them.

2.2 LCS computation and conceptual motivation

In the description below, we follow the notation and terminology of Shadden et al. (2005). Consider a velocity field of the form

$$\dot{\mathbf{x}} = \mathbf{v}(\mathbf{x}, t),\ \mathbf{x} \in U \subset \mathbb{R}^n \tag{2.1}$$

At each instant of time t, a trajectory of the system (2.1) is defined by $\mathbf{x}(t, t_0, \mathbf{x}_0)$, starting from the initial position $\mathbf{x}_0$ at time t_0. The flow map $\phi_{t_0}^{t}(\mathbf{x}_0)$ maps the initial position $\mathbf{x}_0$ at

time $\mathbf{t}_0$ into the position at time t advected under the flow,

$$\phi_{t_0}^t : \mathbf{x}_0 \rightarrow \mathbf{x}(t, t_0, \mathbf{x}_0). \tag{2.2}$$

We use the deformation gradient (Jacobian)

$$D\phi_{t_0}^t(\mathbf{x}_0) = \frac{\mathrm{d}\phi_{t_0}^t(\mathbf{x})}{\mathrm{d}\mathbf{x}}|_{\mathbf{x}=\mathbf{x}_0}, \tag{2.3}$$

to define the finite-time right *Cauchy-Green* strain tensor $C_{t_0}^t(\mathbf{x}_0)$ as

$$C_{t_0}^t(\mathbf{x}_0) = D\phi_{t_0}^t(\mathbf{x})^* \, D\phi_{t_0}^t(\mathbf{x})|_{\mathbf{x}=\mathbf{x}_0} \tag{2.4}$$

where the superscript * refers to matrix transpose. From the strain tensor $C_{t_0}^t(\mathbf{x}_0)$, the largest *finite-time Lyapunov exponent (FTLE)* corresponding to the trajectory $\mathbf{x}(t, t_0, \mathbf{x}_0)$ over the time interval $[t_0, t]$ is defined as

$$\sigma_{t_0}^t(\mathbf{x}_0) = \frac{1}{|t - t_0|} \log \left\| D\phi_{t_0}^t(\mathbf{x}_0) \right\| = \frac{1}{|t - t_0|} \log \sqrt{\lambda_{max} \left[C_{t_0}^t(\mathbf{x}_0) \right]} \tag{2.5}$$

where $\|.\|$ is the norm operator and $\lambda_1 < \lambda_2 < ... < \lambda_{n-1} < \lambda_n = \lambda_{max}$ are the eigenvalues of $C_{t_0}^t(\mathbf{x}_0)$. Since the strain tensor is positive definite by definition, all the eigenvalues are real and positive. When $t > t_0$, we refer to $\sigma_{t_0}^t(\mathbf{x}_0)$ as the *forward* FTLE and for $t < t_0$, we refer to it as the *backward* FTLE.

Shadden et al. (2005) and Lekien et al. (2007) defined Lagrangian coherent structures as the ridges of the FTLE field. Later, Haller (2011) showed that the ridges of the FTLE field are hyperbolic Lagrangian coherent structures only if they satisfy additional criteria. Based on Haller (2011) and Karrasch (2012), for a fixed time interval $[t_0, t]$ a compact hypersurface

$\mathrm{R}(t_0) \subset U$ is defined as an FTLE ridge if for all $\mathbf{x}_0 \in \mathrm{R}(t_0)$ we have

$$\begin{aligned} &\langle D\lambda_{\max}(\mathbf{x_0}, t_0, t), \xi_{\max}(\mathbf{x}_0, t_0, t)\rangle = 0 \\ &\left\langle \xi_{\max}(\mathbf{x}_0, t_0, t), D^2\lambda_{\max}(\mathbf{x}_0, t_0, t)\xi_{\max}(\mathbf{x}_0, t_0, t)\right\rangle < 0 \end{aligned} \tag{2.6}$$

where $\xi_{\max}$ is the eigenvector corresponding to the largest eigenvalue of $C^t_{t_0}(\mathbf{x}_0)$, $\langle\cdot,\cdot\rangle$ is the inner product on $\mathbb{R}^n$ and D^2 represents the Hessian of a scalar field.

Provided that $\mathrm{R}(t_0)$ is an FTLE ridge (2.6) at the initial time, Haller (2011) showed the sufficient and necessary conditions for $\mathrm{R}(t) = \phi^t_{t_0}(\mathrm{R}(t_0))$ to be a hyperbolic repelling LCS during the interval $[t_0, t]$ are

$$\begin{aligned} &\lambda_{n-1}(\mathbf{x}_0, t_0, t) \neq \lambda_{\max}(\mathbf{x}_0, t_0, t) > 1 \\ &\xi_{\max}(\mathbf{x}_0, t_0, t) \perp T_{\mathbf{x}_0}\mathrm{R}(t_0) \\ &\mu^*\mathbf{L}(\mathbf{x}_0, t_0, t)\mu > 0. \end{aligned} \tag{2.7}$$

where $T_{\mathbf{x}_0}\mathrm{R}(t_0)$ is the tangent space of $\mathrm{R}(t_0)$, $\mathbf{L}$ is a matrix defined in Haller (2011), and μ is any non-zero column vector of real numbers (the positive definite condition for $\mathbf{L}$).

Later, Karrasch (2012) showed that if eigenvectors of $C^t_{t_0}(\mathbf{x}_0)$ are differentiable (which is naturally observed in smooth enough velocity fields) then (2.7) simplifies to

$$\begin{aligned} &\lambda_{n-1}(\mathbf{x}_0, t_0, t) \neq \lambda_{\max}(\mathbf{x}_0, t_0, t) > 1 \\ &\xi_{\max}(\mathbf{x}_0, t_0, t) \perp T_{\mathbf{x}_0}\mathrm{R}(t_0). \end{aligned} \tag{2.8}$$

We use the sets of equations (2.6) and (2.8) to extract the actual hyperbolic LCS features from the list of candidates.

Hyperbolic repelling and attracting LCSs have the very important characteristic of being locally the strongest repelling or attracting material surface over time interval $[t_0, t]$ (Haller,

2011). Since they are material surfaces, they divide the domain of motion of particles into different regions of qualitatively different motion (or origin or fate), so we consider the LCSs as *atmospheric transport barriers* (Lekien and Ross, 2010; Senatore and Ross, 2011; Tallapragada et al., 2011; Schmale et al., 2012). Fig. 2.1 shows the conceptual features of the repelling and attracting LCSs. In the case of the attracting LCS feature, the particles inside an initially straddling fluid blob move in such a way that they are attracted to the LCS in forward time. A repelling LCS feature is the same as an attracting one for an inverse time direction.

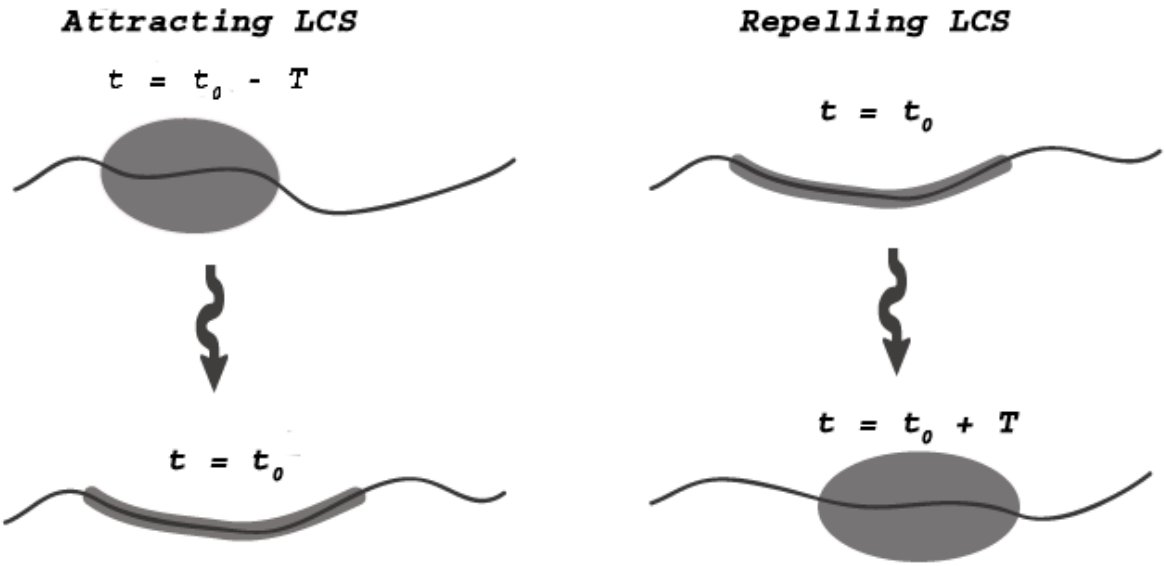


Figure 2.1: Schematic drawing of hyperbolic attracting and repelling LCSs.

2.3 Real-time LCSs extraction

In this section we discuss the data and the relevant data structure we use for numerical calculations of the FTLE-LCSs in the atmospheric flow of interest. In the first part, we introduce NAM-218 integrated system as the main data source for our calculations and

the forecast and archive data sets. In the second part, we define the time schedule and the constraints for real-time FTLE-LCS forecasting. Finally, the last part describes the structure of velocity data.

2.3.1 NAM-218 data set

Computing the FTLE field requires the flow map, (2.2), $\phi_{t_0}^{t} : \mathbf{x}_0 \rightarrow \mathbf{x}(t, t_0, \mathbf{x}_0)$. For generating this map we use numerical data provided by National Oceanic and Atmospheric Administration (NOAA) and National Centers for Environmental Prediction's (NCEP) Operational Model Archive and Distribution System (NOMADS) project. NOMADS product utilizes observational data from radar stations, weather balloons and data from satellites as inputs for its meteorological models. We use one of the outputs of this nonhydrostatic mesoscale model, which is the North America Mesoscale, NAM-218, with data given on a grid of 614×428 points spaced at about 12.1 km covering North America. This model contains 70 variables such as temperature, humidity and components of velocity on 70 levels. Among the 70 levels, 44 correspond to pressure levels (up to 10 mb), while other levels mostly refer to various heights above the ground level. NAM-218 data are given on a plane given by Lambert conformal projection, which projects the points given by their latitude and longitude (ϕ, λ) on the sphere to a Cartesian coordinate (Snyder, 1987). We consider a 6230×4670 km rectangular area as our wind velocity field and we compute the LCSs over an interrogation zone of 1000×1000 km size with 256×256 grid points. We consider the sampling area (Virginia Tech's Kentland Farm near Blacksburg, VA) as the center of that square ($37°11'$ N latitude and $80°35'$ W longitude), which is approximately 16 km southwest of the Virginia Tech campus in Blacksburg, Virginia.

Essentially two types of data are used in our FTLE-LCSs calculations: archive data and

forecast data. The recorded state of the atmospheric system is called *archive data*. To generate this data set, numerous measurements from, e.g., weather stations, weather balloons, satellites and any available atmospheric observations, are assimilated and used to drive large-scale oceanic-atmospheric geophysical fluid models. Outputs of this process over the domain of interest, e.g., North America, might be different from measured values at specific measurement points but they satisfy objective functions of the data assimilationsimulation process. By *forecast data*, we mean the output of geophysical models which use the most current outputs of the data assimilation process (archive data) as an input to estimate the state of the system for (near-)future times (Kalnay, 2003). In NAM-218 data set, the temporal resolution of archived data is 6 hours corresponding to 00:00, 06:00, 12:00 and 18:00 UTC of each day. Accordingly, the forecast chain can start based on each of these archive slices. Temporal resolution for forecast data is 3 hours and at each moment, the maximum available forecast data is for 84 hours lead time.

2.3.2 Time scheduling for real-time FTLE-LCSs calculations

As mentioned in the Introduction, we want to use forecast data to predict the LCS features. Based on the availability of wind velocity forecast data and processing time, we have a time schedule for each day. Fig. 2.2 illustrates the details of such a schedule for a typical day. For our sampling purposes, LCSs are desired in the time interval 08:00 to 17:00 local time (Eastern Daylight Time (EDT)), equivalent to 12:00 to 21:00 UTC, the daylight hours during which UAV flights are permitted. We call this time interval the *interrogation window* $\mathcal{I} = [t_1, t_2] \subset \mathbb{R}$. All the results presented in this paper are related to this time interval and the selected day is 29 Sep 2010, i.e., $t_1 =$ 12:00 UTC 29 Sep 2010 and $t_2 =$ 21:00 UTC 29 Sep 2010. We imagine we want to plan sampling flights during the interrogation window, but we are making the forecast before t_1 by some number of hours. To perform real-time extraction

of LCSs over the interrogation window, we must use a combination of forecast and archived wind data. Based on NAM-218 timetable, archived data are available up to 18:00 UTC of the day before t_1. For times after 18:00 UTC of the day before t_1, we must use forecast data (Fig. 2.2).

We call the time interval between the last frame of available archive data and the start of the interrogation window the *gap time*, T^*, e.g., the time interval between 18:00 UTC 28 Sep 2010 and 12:00 UTC 29 Sep 2010 (note that $T^* > 0$ for all real-time calculations, but in some cases we *artificially* use negative values; perfect continuous-time LCSs forecasting; to show its effects on the quality of LCSs forecasting, e.g., Fig. 2.9).

As mentioned previously, this procedure is general and the time schedule would be similar for other days.

Based on the time schedule Fig. 2.2, the parameters that control the forecasting procedure are gap time, T^*, and integration time, $T = t - t_0$, following the notation of §2.2 for t_0 and t, where $t_0 \in \mathcal{I}$. We are interested in studying the effects of the parameters (T^*, T) on LCS forecasting.

We set the maximum integration time T to be 48 h, since our results in §2.4 suggest that for larger integration time, patterns of forecast FTLE-LCS are significantly different from the true answers obtained from archived data. Intuitively and from the chaotic dynamics of atmospheric systems (Lorenz, 1963; Palmer, 2000) we know that by increasing the portion of trajectory integration which depends on the forecast wind velocity field, the forecast particle trajectories, and hence Lagrangian structures, will diverge from true ones. The important result of this paper is the observation of divergence of Lagrangian features based on short forecast lead time (i.e., T^* small) when errors of the Eulerian forecast field are expected to be small.

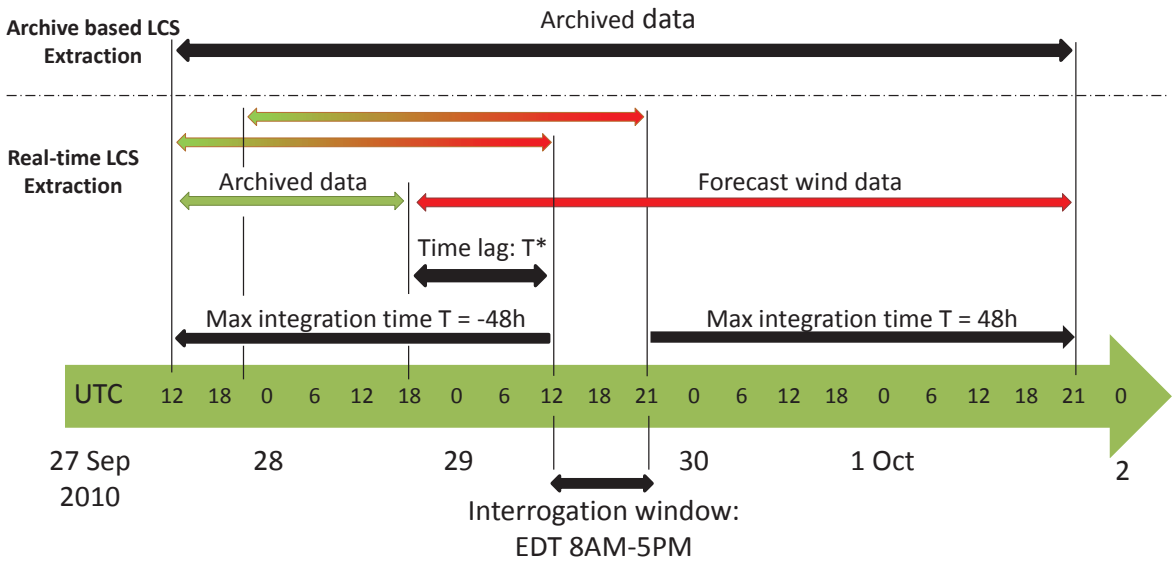


Figure 2.2: Schematic time schedule for calculation of real-time and archive-based FTLE-LCSs with respect to NAM-218 timetable. Archive-based features are calculated from pure archive data. Real-time FTLE-LCSs are calculated based on a combination of forecast and archive data. Interrogation window (12:00 to 21:00 UTC) is the time interval that UAV flights are permitted. For optimal sampling we have to know the LCS features during this interval. *Gap time* (T^*) is the time between the last frame of available archive data and the start of the interrogation window. FTLE-LCSs are calculated every 15 minutes in interrogation window in forward and backward directions.

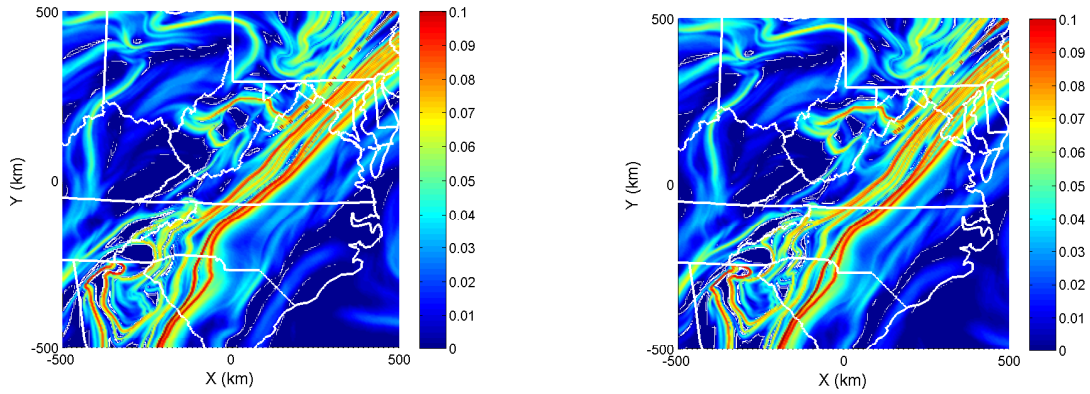
The main contribution of this paper is to quantify (a) the accuracy of prediction of LCSs and (b) the sensitivity of such predictions with respect to the parameters (T, T^*).

2.3.3 Velocity (wind) field data structure

Based on the previous work (Haagenson et al., 1990; Kahl et al., 1989; Stohl, 1998; Kahl and Samson, 1986; Tallapragada and Ross, 2013; Tallapragada et al., 2011), we chose flow data on a 900 mb pressure surface which also lies within the range of our prescribed UAV sampling height of 100-400 m above the ground level at the elevation of Kentland Farm near Blacksburg, VA (farm ground level is approximately 540 m above the sea level (Buergler et al., 2005)). We make an isobaric approximation, considering only the components of

velocity parallel to the pressure surface, since the vertical velocity is about three orders of magnitude less than the horizontal components for the range of elevation of interest, and averaged over the mesoscale. Considering pressure as the vertical coordinate, the mesoscale averaged rate of change of pressure levels is about 0.03 Pa/s. Thus, during $T_{max} = 48$ h, pressure may change by as much as $\Delta p \sim \pm 52$mb. Under this stratification assumption, we would expect FTLE fields separated by Δp to be similar over integration times $T = \pm T_{max}$ (Tallapragada, 2010). To illustrate the similarity of neighboring pressure levels, in Fig. 2.3 the backward FTLE field for 900 mb and 850 mb levels are given at the time $t =$ 12:00 UTC 29 Sep 2010 and $(T, T^*) = (-T_{max}, 18$ h).

In addition, we can consider this study as a diagnostic approach for approximately 2D flows to show the effects of cumulative errors of the Eulerian velocity field on the resultant FTLE-LCS features.



(a) (b)

Figure 2.3: Eastern U.S. with the state of Virginia and Kentland farm (sampling location) in the center. Backward FTLE field for (a) 850 mb and (b) 900 mb pressure levels at time 12:00 UTC 29 Sep 2010 with $(T, T^*) = (-48$ h, 18 h), vertical color bar indicates FTLE magnitude (h^{-1}). Note the similarity.

2.4 Forecast and archive based FTLE-LCSs

The main interest in prediction of FTLE-LCSs is based on the results, e.g., (Tallapragada et al., 2011) suggesting that there exists a relationship between passing atmospheric LCSs on the mesoscale and locally detected changes in airborne microbial concentrations. If the population structure of airborne microorganisms is determined by large-scale atmospheric features which the LCSs represent, then the ability to accurately predict atmospheric LCSs is of interest for prediction in microorganism dispersal.

To represent the time-dependent FTLE-LCSs (both forecast and archive based) over the 9 h interrogation window $\mathcal{I}$, we use snapshots of these features every 15 minutes, so during 9 h, the total number of FTLE-LCS snapshots is 37.

The upper panels of Fig. 2.4 show two snapshots of the attracting LCS features from forecast (left) and archived (right) data over the specified zone at the beginning of the interrogation window (12:00 UTC). In this case $(T, T^*) = (-24 \text{ h}, 18 \text{ h})$ and conditions (2.6) and (2.8) are considered for extracting the hyperbolic LCSs. In this figure we see very similar patterns; however, the features are not exactly the same as we expected due to the errors in the wind forecast data.

The lower panels of Fig. 2.4 show snapshots of the forecast- (left) and archive-based (right) LCSs at 21:00 UTC with the same calculation parameters as the upper panels.

One observes that the quality of forecasting decreases for further lead time, i.e., the LCS patterns from archive and forecast data become further apart by some metric from the beginning of the interrogation window, t_1, to the end, t_2, since 9 additional hours of forecast contributed to t_2 compared to t_1. Due to cumulative effects of wind forecast errors on the *trajectories* of particles, the resultant LCS features are less accurate, as we will attempt to quantify in § 2.5 by applying different hybrid Eulerian-Lagrangian approaches.

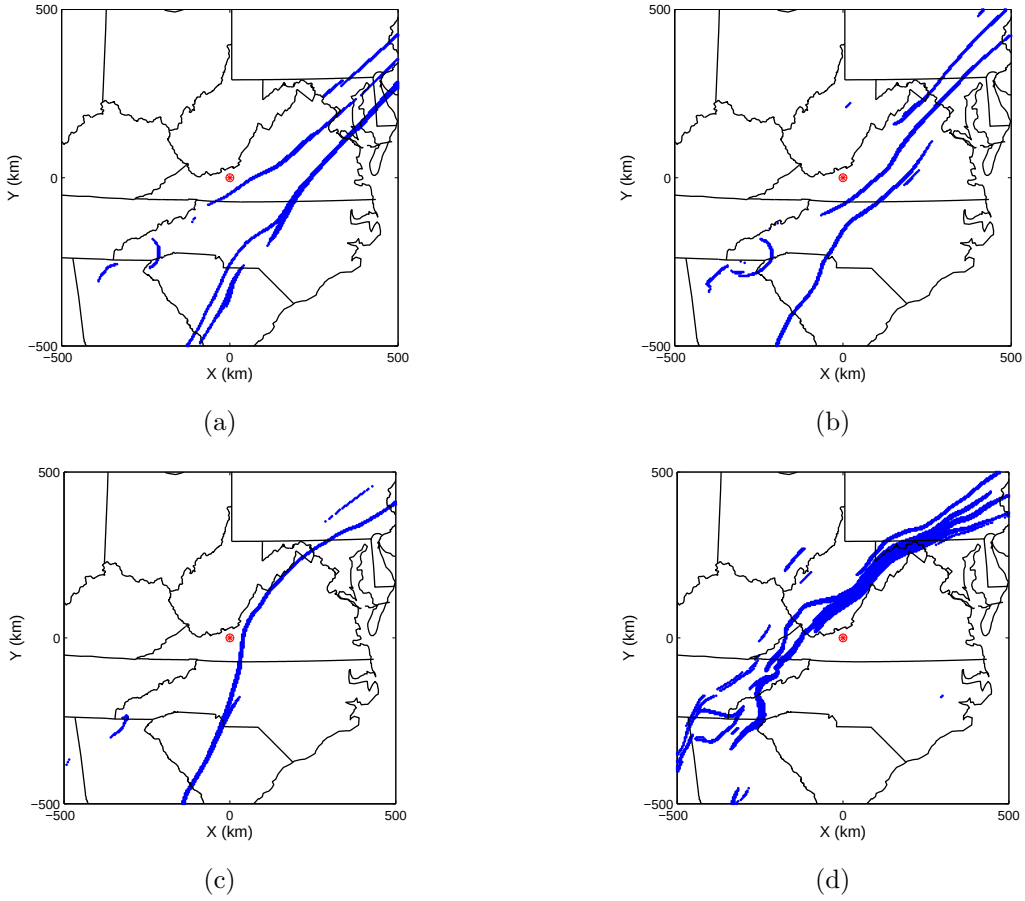


Figure 2.4: Eastern U.S. with the state of Virginia and Kentland farm (sampling location) in the center (red point). (a, c) forecast- and (b, d) archive-based hyperbolic attracting LCSs. Upper panels: 12:00 UTC; lower panels: 21:00 UTC 29 Sep 2010. $(T, T^*) = (-24\text{ h}, 18\text{ h})$. In this case, LCSs displacements are SE to NW with different velocities. Some branches diminish during interrogation window such as the upper branch of the forecast results (panel (a)) and some remains strong, e.g., the lower branch (panels (a) and (c)).

2.4.1 Sources of errors

In this section we discuss some reasons for the disagreement between forecast- and archive-based LCS features at the level of Eulerian velocity fields.

Spatiotemporal finite resolution of input data is an important reason. As mentioned previously, the temporal resolution provided by the NAM model is 6 hours (considering short-term

forecast, temporal resolution would be 3 hours) and the best spatial resolution is about 12.1 km over North America (we say best as these spacings are non-uniform). For numerical integration of particle trajectories, spatiotemporal interpolation of the velocity field is needed. Interpolation leads to some differences between the true and the calculated particle paths (Griffa et al., 2004; Ozgokmen et al., 2000) which lead to different FTLE fields. Since we use the same method of interpolation and numerical integration for both the forecast and archived data (third order splines / fourth order Runge-Kutta), we do not consider the differences between forecast- and archive-based particle paths to be caused by the interpolation and integration methods.

We are led to conclude that the errors of the forecast fields are due to the inherent chaotic behavior of the atmospheric system. Recalling Fig. 2.4, when we desire FTLE-LCS features for further lead time, we get less accurate results. This fact is the outcome of limits of predictability of chaotic motion of particles in atmospheric flow (Lorenz, 1963; Palmer, 2000). In this paper, the goal is not to study the limits of forecasting in presence of deterministic or stochastic chaos, but merely to quantify the effects of existing forecast errors on the calculated FTLE-LCSs.

2.4.2 Quantitative comparison of forecast and archived velocity (wind) fields

To understand the error distribution of the forecast velocity field, the essential source of errors in forecasting LCSs, we compare the wind forecast data with the corresponding archived data sets.

We consider the common scenario of real-time extraction of LCSs for our purposes, for which $T^* = 18$ h and the chain of forecast data starts from 18:00 UTC of the day before

the interrogation window (referring to Fig. 2.2). At each time slice and for each spatial grid point, we find the error as the difference between the archive and the forecast velocity components in the XY plane (u and v are the components of velocity in X and Y directions, respectively). Fig. 2.5 shows the results for three frames related to the interrogation window, highlighting the persistence and large size of the regions with large and growing amplitude error.

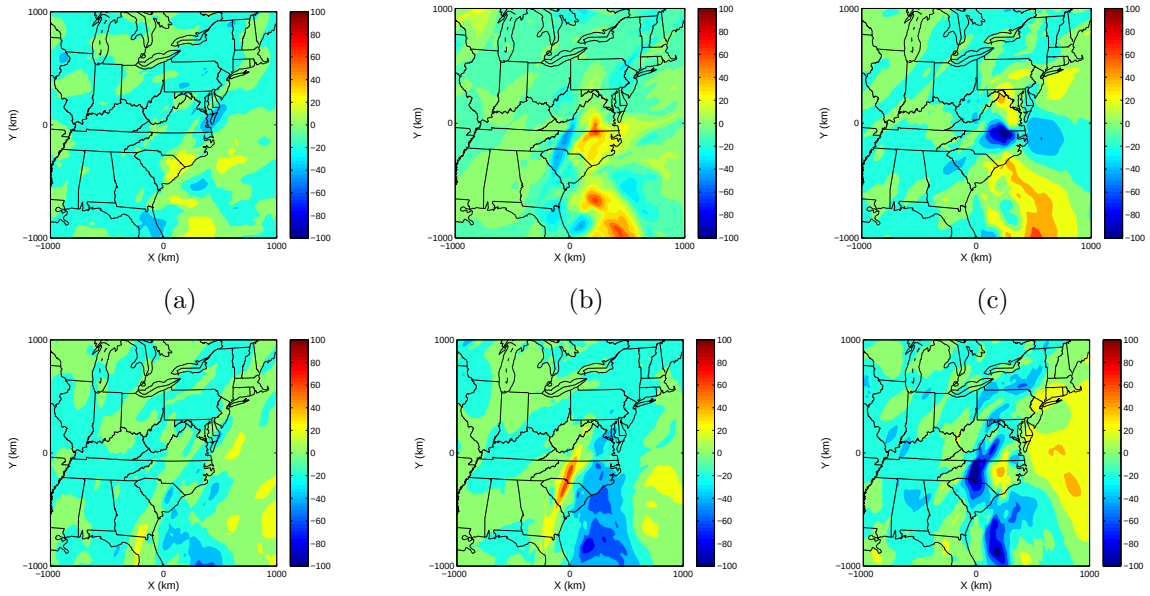


Figure 2.5: Error of forecasting (km/hr) over North America, (a, d) 12:00 UTC 29 Sep 2010, (b, e) 21:00 UTC 29 Sep 2010, and (c, f) 12:00 UTC 30 Sep 2010. Upper panels: u component, lower panels: v component. Regions with persistence large error are responsible for deviations of forecast LCSs from true archive-based results.

Calculated LCS features from forecast data should resemble the true features by maximum distance of Δ (eq. (19) of (Haller, 2002)) *if forecast errors are localized in time*. The existence of large and prolonged high error regions, as shown in Fig. 2.5, reveal that the errors are not localized in time and thus could lead to significant differences between forecast- and archive-based FTLE-LCS results.

To investigate the statistics of the error we find the mean absolute error (MAE) and mean squared error (MSE) over the domain of interest (Fig. 2.6, panels (a) and (b), respectively) (Jolliffe and Stephenson, 2003). Along with these two standard measures, we apply a method of nonlinear weighted averaging which yields a normalized scalar number as a measure of quality of forecasting in a Eulerian sense. This metric decreases with time. Also, the standard deviation of errors grows as a function of the forecast lead time, showing the increasing spread and divergence of forecasting results from the true state of the system.

For the nonlinear weighted averaging, we normalize the relative error of velocity forecast with respect to a Gaussian filter,

$$\frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-\left(\chi-\mu\right)^2}{2\sigma^2}\right)$$

where $\sigma = (2\pi)^{-1/2}$, $\mu = 0$ and χ represents the corresponding component of relative error at each point. The output of this process is a normalized value for each grid point belonging to $(0, 1]$. This value is one for a perfect forecast and approaching zero for an infinite error. We set the quality of forecasting (η) over the entire domain as the mean value of this normalized relative error field,

$$\eta = \frac{1}{N}\sum_{i,j} \exp(-\pi\chi_{i,j}^2) \tag{2.9}$$

where N is the total number of grid points.

Fig. 2.6(c) shows the quality of forecasting (η) decreasing with time (corresponding to the normalized field). In addition, Fig. 2.6(d) shows the standard deviation values related to the spatial distribution of error (original values) for all forecast time slices (note that in this example the value of MAE, MSE and standard deviation is zero before 18:00 UTC 28 Sep 2010 since we use the archived data for the time interval before it).

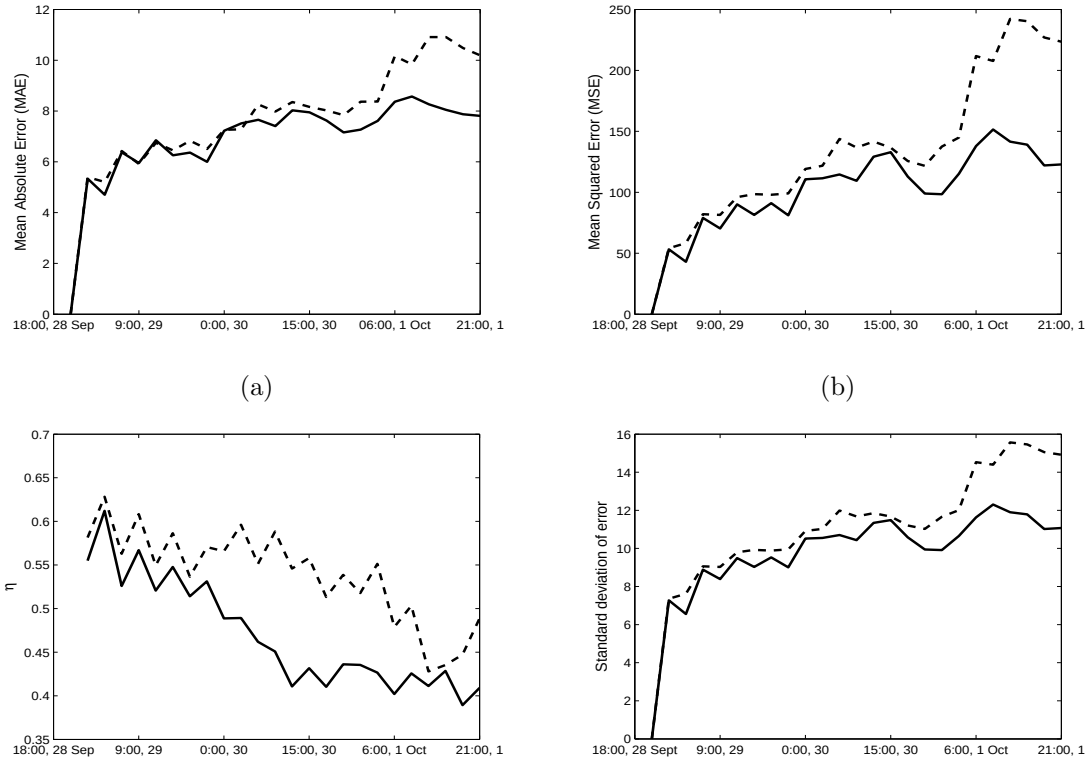


Figure 2.6: Some statistical measures of forecast error over the domain of velocity field (a) mean absolute error (MAE), (b) mean squared error (MSE), (c) decay of quality of forecasting and (d) growth of standard deviation with respect to time: u component (solid line), v component (dashed line).

The important point of this figure is the general trend of decreasing forecast quality and increasing MAE, MSE and standard deviation of error with time. In the next section, we discuss the effect of these errors on FTLE-LCS features.

2.5 Results

Referring to the differences between forecast- and archive-based LCS features e.g., Fig 2.4, a natural question that arises is: How can one quantify the differences between the forecast-

and archive-based LCSs? Other related questions that emerge are about the value of the forecast results for scheduling flights to collect samples before and after the passage time of the forecast LCSs. In other words, how well do the forecast LCSs describe the real case of LCSs passages, particularly the passage times? Also, regarding limitations on availability of necessary data for predicting the LCSs, what is the best choice for effective parameters such as the integration and gap times?

We use five methods to compare the forecast and archive LCSs. We focus on the resultant FTLE-LCSs as the Lagrangian objects which record the history of the system (since they are calculated from flow maps, i.e., trajectory of particles) We compare the 37 frames of the FTLE-LCSs corresponding to every 15 minutes of the interrogation window (12:00-21:00 UTC). These comparisons could be regarded as hybrid Lagrangian-Eulerian methods since they compare snapshots of Lagrangian features. We notice that for a comprehensive verdict, we need to consider the result of all these methods. None of them alone quantify the similarities and/or differences.

2.5.1 Pointwise comparison of FTLE fields

The first and the simplest approach is pointwise comparison of the values of FTLE fields at a fixed geographical location, e.g., our sampling site Kentland farm, which is shown in Fig. 2.7 for the case of attracting LCSs (backward time integration). If the forecasts were perfect (i.e., they match the archive-based FTLE-LCSs exactly), we expect a complete match between two time-series curves. Although this approach seems to be over-simple, we can employ a useful rule of thumb for detecting the validity of our forecasts before doing any additional calculations on archived data to extract the archive-based LCSs.

Results from several simulations for different days suggest a rule of thumb which may be

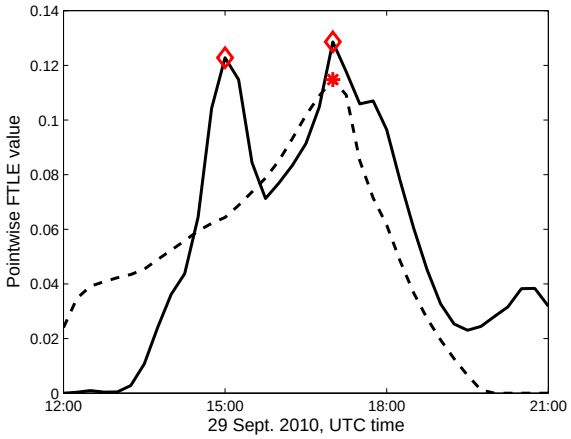


Figure 2.7: Forecast- (dashed) and archive-based (solid) FTLE value at the sampling location. $T = -24$ h and $T^* = 18$ h. Corresponding hyperbolic LCSs are shown by asterisk and diamonds.

beneficial for early warning systems, e.g., for vast crop fields where the exact passage times of LCSs are not important, but it is important for one to know if hyperbolic LCSs pass over a region, since these have been associated with fluctuations in microbial populations (Tallapragada et al., 2011). The rule of thumb is: if the maximum forecast FTLE value during the interrogation window is above a certain threshold then we expect to observe at least one archive-based LCS in the same time interval. The threshold we have determined by observation is 0.07 h^{-1}, i.e., if the peak forecast FTLE is > 0.07 h^{-1}, then the peak will likely be preserved (and likely an LCS) in the archive-based calculation. Further statistical tests would be needed to bear out this rule of thumb.

Note that Fig. 2.7 shows that the forecast FTLE-LCS captured one of the archive peaks (close to 17:00 UTC) but not the other one which is close to 15:00 UTC. We understand the context of this mismatch by considering the temporal evolution of the entire FTLE field. In the forecast, one observes two major ridges at 12:00 UTC (Fig. 2.4, panel (a)), the upper ridge has a motion from SE to NW, but the other ridge does not pass the sampling location

and has a slower motion (Fig. 2.4, panel (c)), so one peak of the FTLE field is observed (Note that after the upper ridge passes over the sampling location, it becomes less strong and at 21:00 UTC (panel (c)) it is vanished). In contrast, for the archive case all ridges move in the same direction and two major features pass over the sampling location (Fig. 2.4, panel (b)), so we record two peaks.

2.5.2 Cross correlation between forecast and archive based LCSs

Cross correlation is a standard method of estimating the degree to which two scalar fields are correlated (Gonzalez and Woods, 2007). Considering 2D shifts and discrete data points, we use

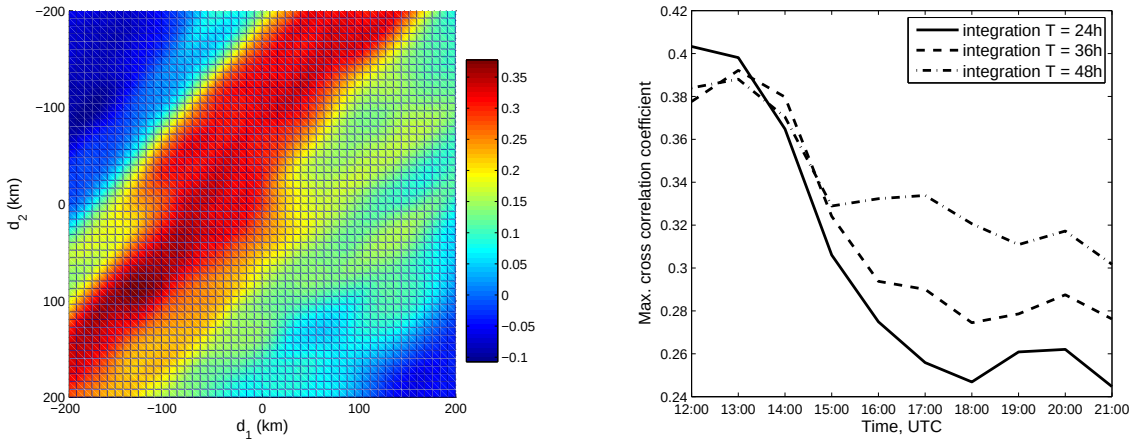
$$r(d_1, d_2) = \frac{\sum_m \sum_n \left[\left(\sigma_{m-d_1,n-d_2} - \bar{\sigma}\right) \left(\psi_{m,n} - \bar{\psi}\right) \right]}{\sqrt{\sum_m \sum_n \left[\left(\sigma_{m-d_1,n-d_2} - \bar{\sigma}\right)^2 \left(\psi_{m,n} - \bar{\psi}\right)^2 \right]}} \tag{2.10}$$

to get a 2D surface of cross-correlation coefficients, where σ and ψ represent the value of the FTLE field at each point of forecast and archive fields and $\bar{\sigma}$ and $\bar{\psi}$ are the spatial average of σ and ψ, and d_1 and d_2 represent the shift in the X and Y directions, respectively.

Fig. 2.8(a) shows the cross correlation coefficient between forecast and archived attracting LCS features at 12:00 UTC 29 Sep 2010, for (T, T^*)=(-36h,18h). In this figure, we consider shifts up to ±200km in both directions. Using this approach and by looking for spatial shifts, we focus on the quality of forecast over the interrogation window. As an example, Fig. 2.8 (b) shows how the quality of forecast tends to decrease with time (T^* =18 h and $T = -24$ h, -36 h and -48 h). This result is typical (recall Fig. 2.4).

Generally, we can apply this method to compare the forecast- and archive-based LCSs with respect to different parameters such as gap time T^* and integration time T.

Fig. 2.9(a) shows the typical effect of T^* on the quality of forecasting for attracting and



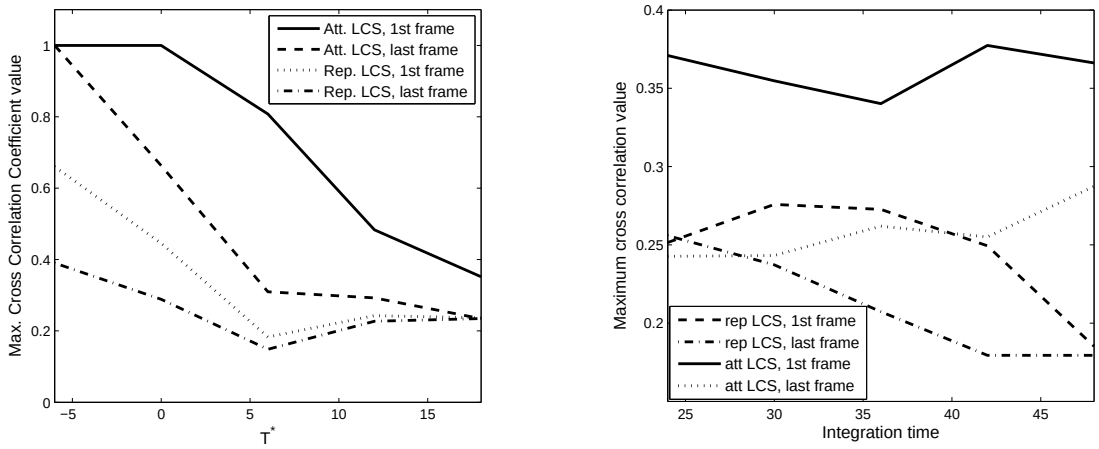
(a) (b)

Figure 2.8: (a) Maximum cross correlation coefficient between archive and forecast attracting LCSs at 12:00 UTC 29 Sep 2010, considering 2D shift ($\pm 200km$ in both X and Y), (T, T^*)=(-36h,18h). (b) Ensemble maximum cross correlation coefficient during 12:00 to 21:00 UTC 29 Sep 2010, $T^* = 18$h, integration time $T = -24$ h, -36 h and -48 h.

repelling LCSs. To generate this figure we fix the integration time to $|T| = 24$ h and then find the maximum value of cross correlation coefficient between the archive-based FTLE field and forecast-based FTLE field, for various values of T^* ('1st frame' refers to 12:00 UTC 29 Sep 2010 and 'last frame' refers to 21:00 UTC 29 Sep 2010). This figure also illustrates a general trend; when we decrease the gap time, we see better agreement between forecast- and archive-based LCSs. Note that negative values of T^* show perfect continuous time LCSs forecasting.

Considering repelling LCSs ($T > 0$), one might think that T^* has no effects on the quality of forecasts, since the direction of integration is positive and we do not need the past (archive) data for calculating the FTLE field. However, this is not the case as Fig. 2.9(a) demonstrates. Note that for smaller T^*, the forecast is for a smaller duration, so the quality of wind forecast and the corresponding LCS features show better agreement with purely archive-based features.

Fig. 2.9(b) shows the effect of integration time T on the agreement of the archive- and forecast-based FTLE fields. Generally by increasing the integration time (T), one sees sharper (and more) ridges, but based on the quality of the wind forecast, the quality of the forecast LCSs could get better or worse, i.e., if we have a reliable wind forecast, we expect cross-correlation coefficients to increase when we increase T but if the wind forecast is not reliable, we anticipate less accurate features. From panel (b) we observe that by increasing the integration time the quality of attracting LCSs improves slightly. In contrast, for repelling LCSs the best choice is to minimize the integration time.



(a) (b)

Figure 2.9: Comparing the forecast- and archive-based FTLE-LCSs, studying the effect of (a) T^* when T =24h and (b) integration time (T) on the maximum value of cross correlation coefficient when T^* =18h.

From Figs. 2.8 and 2.9 we notice small correlation coefficients even for the cases where LCS patterns seem to be similar by eye. This technique is not adequate for practical pattern recognition since the coefficient degrades rapidly when the patterns (forecast results) deviate from the references (archive-based results). In cases where we want to investigate the quantitative correlation of slightly offset patterns this approach is beneficial. If the overall qualitative similarity is important, then we have to consider other approaches such as robust

correlation filters (discussed in a later section) (Kumar et al., 2005).

Moreover, the nature of cross-correlation method forces us to compare single snapshots of extracted LCSs (i.e., spatial correlation considered pointwise in time) and it cannot be applied to the whole series of 37 LCS snapshots (i.e., full spatiotemporal variability). More generally, we expect the forecast-based FTLE field to be both shifted and deformed compared to the archive-based FTLE field. For example, for the backward ($T < 0$) FTLE field, we expect

$$\bar{\sigma}_{t_0}^{t_0+T}(\mathbf{x}_0) = \zeta_{T^*}(\sigma_{t_0}^{t_0+T}(\mathbf{x}_0)) \tag{2.11}$$

where $\bar{\sigma}_{t_0}^{t_0+T}(\mathbf{x}_0) \in \mathcal{F}(U, \mathbb{R})$ is the forecast-based FTLE field with gap time T^*, $\sigma_{t_0}^{t_0+T}(\mathbf{x}_0) \in \mathcal{F}(U, \mathbb{R})$ is the archive-based FTLE field, and $\zeta_{T^*} : \mathcal{F}(U, \mathbb{R}) \to \mathcal{F}(U, \mathbb{R})$ is a one-parameter family of diffeomorphisms on the function space $\mathcal{F}(U, \mathbb{R})$, with parameter T^*, where ζ_0 is the identity.

Spatial shifts are only one limited possibility for ζ_{T^*}. In the next section, we use the POD method to enable us to consider more general ζ_{T^*} by comparing the archive- and forecast-based FTLE fields as a series of successive time-slices rather than, at each fixed time, considering spatial shifts of two-dimensional scalar fields.

2.5.3 POD analysis

Proper orthogonal decomposition (POD) is a technique to analyze, e.g., a time-varying scalar field, and can reduce the order of complex systems (Antoulas, 2005; Sirovich, 1987; Rowley, 2005). We apply this technique to the time-varying forecast- and archive-based FTLE fields to get the principal mode shapes as the building blocks of the original high-dimensional system. We then compare these mode shapes to study the effects of different parameters on the quality of forecasting. This approach yields a wider and deeper view, since the mode

shapes encapsulate the overall data of all the time slices.

Considering the scalar function $z(\mathbf{x}, t)$ over some finite domain, we want to approximate this function as a superposition of spatial modes with time-varying coefficients, as

$$z(\mathbf{x}, t) \approx \sum_{i=1}^{K} a_i(t)\Phi_i(\mathbf{x}), \tag{2.12}$$

where $\Phi_i(\mathbf{x})$ represents the spatial mode shapes and we expect as $K \to \infty$, the summation yields the exact value of $z(\mathbf{x}, t)$. Conventionally, the Φ's are chosen to be orthonormal, so

$$\int_{\mathbf{x}} \Phi_i \Phi_j d\mathbf{x} = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases} \tag{2.13}$$

Using eq. (2.13), we find the time-varying amplitudes of the mode shapes as

$$a_i(t) = \int_{\mathbf{x}} z(\mathbf{x}, t)\Phi_i(\mathbf{x}) d\mathbf{x} \tag{2.14}$$

Generally, finding the Φ's depends on our choice of basis functions (e.g, Fourier series with orthogonal trigonometric functions). In the case of the POD method, the Φ's are chosen such that the approximation of (2.12) for *each individual and arbitrary* K is the best approximation in a least squares sense.

When we have $z(\mathbf{x}, t)$ as a set of discrete numerical data, we can use the singular value decomposition (SVD) technique to find the mode shapes as well as the corresponding coefficients.

Considering M to be a $m \times n$ real matrix (note that FTLE fields are always real, so we only

consider real matrices), we can decompose M as

$$M = U\Sigma V^*, \quad M \in \mathbb{R}^{m\times n} \tag{2.15}$$

where U is $m \times m$, Σ is $m \times n$ diagonal and nonnegative and V^* is the transpose of V, and an $n \times n$ matrix. Since M is real, all U, Σ and V would be real, also U and V are unitary matrices. The diagonal elements of Σ are called the *singular values* of the matrix M and the number of them is equal to the $\min(m, n)$, also they are conventionally placed in descending order. For obtaining the mode shapes, we can rewrite eq. (2.15) as

$$M = QV^* = \sum_{i=1}^{n} q_i v_i^* \tag{2.16}$$

where $Q = U\Sigma$ and q_i is the i^{th} column of Q. We arrange the numerical data from $z(\mathbf{x}, t)$ into a matrix like M; then the a_i's are equal to the q_i's and the mode shapes are equal to the v_i^*'s.

Determination of mode shapes

For the 37 snapshots representing our interrogation window, we get 37 mode shapes from the POD method. Note that we can *regenerate* all the snapshots by those 37 mode shapes via (2.12) where the required coefficients can be generated by the inner product of input data M and the related mode shape as described by (2.14). Recalling that the importance of the mode shapes depends on the magnitude of their corresponding singular value, Fig. 2.10 shows the first and the second mode shapes for both the forecast- and archive-based FTLE field, where (T, T^*)=(−24 h,18 h). One advantage of POD is that one can do the comparison for the most important mode shapes only, rather than for all of them, since the contribution of a mode to the original time-varying field depends on the value of the related singular

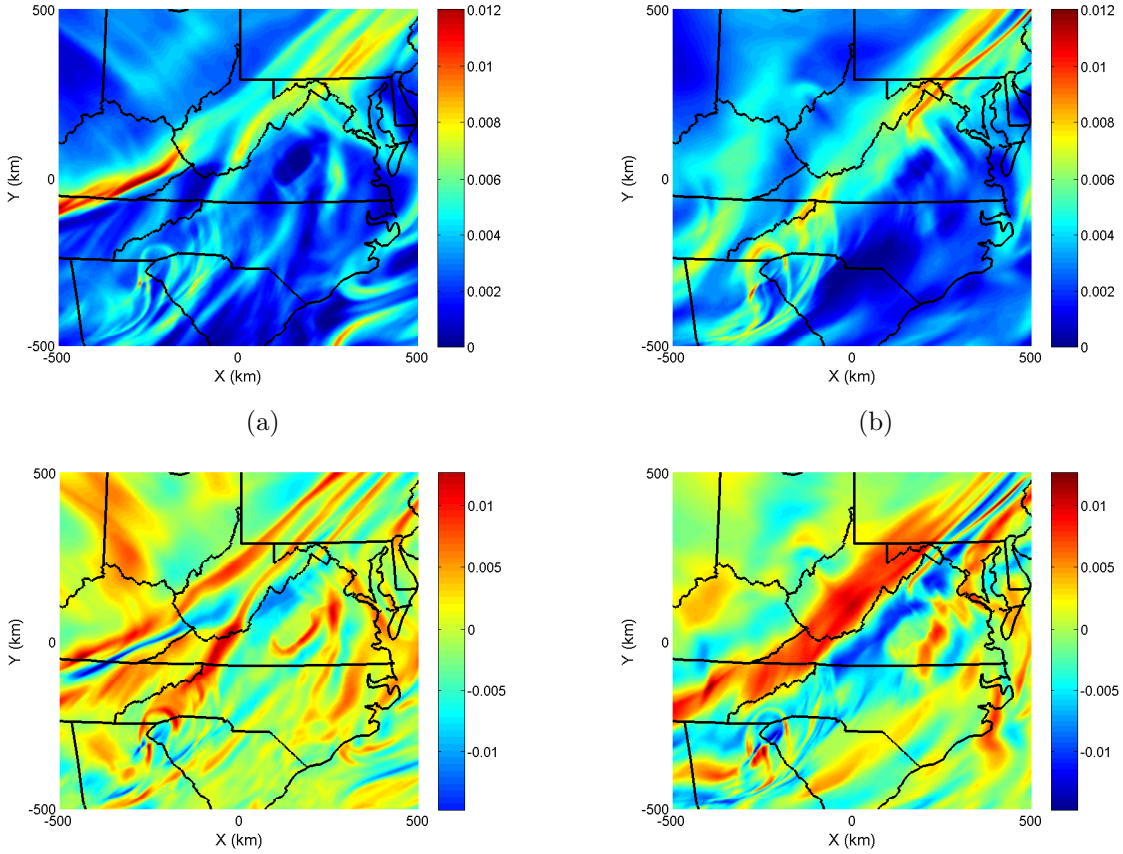


Figure 2.10: First and second mode shapes for the (a, c) forecast- and (b, d) archive-based backward FTLE fields, (T, T^*)=(−24 h,18 h).

value.

Fig. 2.11 shows the cumulative contribution of mode shapes for typical forecast-based backward and forward FTLE fields and the associated attracting and repelling LCS features, corresponding to the interrogation window with $(|T|, T^*)$=(24 h,18 h). From this figure it is clear that if we want to have up to 90% of the energy of the system, we have to consider only five mode shapes.

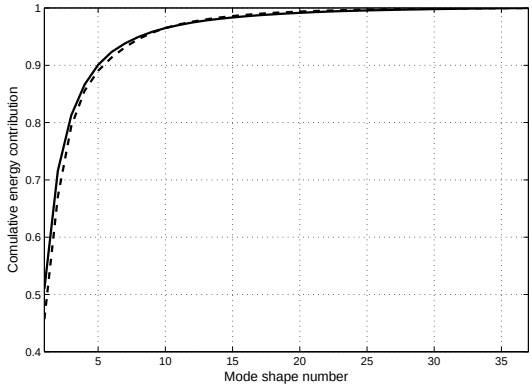


Figure 2.11: Cumulative contribution of the mode shapes, backward FTLE (dashed line), forward FTLE (solid line), $(T, T^*) = (48, 18)$.

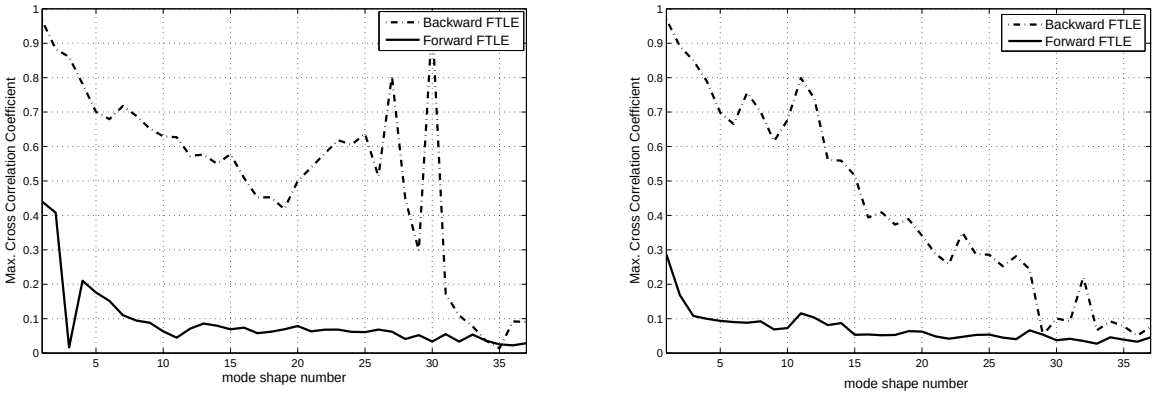
Effect of parameters on the quality of forecasting using POD modes

In this section, we investigate the contribution of parameters T and T^* on the quality of forecast LCSs. We compare the principal mode shapes of forecast- and archive-based FTLE fields, using the cross correlation coefficient, to find the degree of similarity between them.

Fig. 2.12 shows the effect of integration time T on the measure of similarity of mode shapes of backward and forward FTLE fields when we fix the T^* at zero (perfect real-time forecast case) and let the integration time $|T|$ be 24 or 48 hours.

We observe that for backward FTLE, increasing T from 24 h to 48 h does not change the quality of mode shapes much. We see some differences for higher mode shapes, but these contribute little to the FTLE field.

We also observe that the cross-correlation coefficients for the forward FTLE are always less than the coefficients of the backward FTLE. Also, for a given forward FTLE mode shape, the similarity coefficient decreases with increasing T, which is understandable as the reliance on forecast wind data increases. Note that for the $T^* = 0$ case, the cross-correlation



(a) (b)

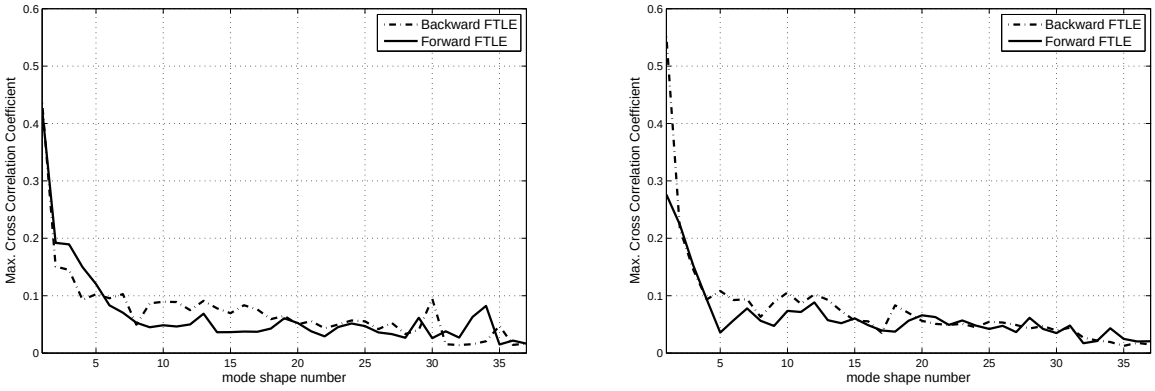
Figure 2.12: Maximum cross correlation coefficient between different mode shapes, perfect real-time forecasting case; T^*=0, integration time (a) $|T|$ =24 h and (b) $|T|$ =48 h.

coefficient for the first backward FTLE mode shape is close but not exactly one. Recall that mode shapes encapsulate information from *all* time snap shots, so as later snap shots of the forecast wind field diverge from the archive ones, we expect an increasingly imperfect match between FTLE mode shapes.

For the next comparison, we choose T^* to be 18 hours which is the *worst case* of real-time forecasting we have considered. Fig. 2.13 shows the value of cross correlation coefficients for different mode shapes.

Considering the forward FTLE (repelling LCS features), Fig. 2.13 and 2.12 show that when we decrease the integration time, we generally get better results regarding the correlation coefficients, i.e., for shorter integration times T, cumulative errors of predicted trajectories will be smaller. While the results presented are for the time interval 12:00-21:00 UTC 29 Sep 2010, comparison with other days and times (not reported here) suggests that these results are typical.

Results of this section are based on correlating the mode shapes, thus, similar to § 2.5.2, we



(a) (b)

Figure 2.13: Maximum cross correlation coefficient between different mode shapes, real-time forecasting; $T^* = 18$h, integration time (a) $|T|$ =24 h and (b) $|T|$ =48 h.

encounter the sensitivity of the correlation method to deformation of LCSs, leading to small correlation coefficients for what may seem similar FTLE field patterns. In § 2.5.4 we use a type of composite correlation filter to overcome this weakness.

2.5.4 Composite correlation filter

We see some weakness of cross correlation techniques in previous sections, e.g., when two sets of patterns are similar to our eye but the correlation coefficient is small. In this regard, robust pattern recognition methods can overcome some of these limitations and provide measures of similarity in better agreement with our visual perception.

In this section we apply synthetic discriminant function (SDF) filter for correlating whole patterns (Kumar et al., 2005; Hester and Casasent, 1980). To design this kind of filter, we need to have multiple views (training images) of a single object. If we consider each frame of FTLE-LCS as *one* object, then we do not have enough training images to construct the filter, but if we consider the whole 37 frames as different views of one time-varying object

then we would have enough data to design the composite correlation filter. By convoluting this filter with all the forecast FTLE-LCS frames, one can find the overall similarity of each frame to the whole set of reference frames. In this approach, the filter is designed such that it generates a pre-specified value in response to each training image, e.g., 1 for the reference frames.

The governing equation of this filter is

$$\mathbf{h} = \mathbf{X} \left(\mathbf{X}^* \mathbf{X} \right)^{-1} \mathbf{u} \tag{2.17}$$

where X represents the all training images collected together,

$$\mathbf{X} = \left[\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n \right], \tag{2.18}$$

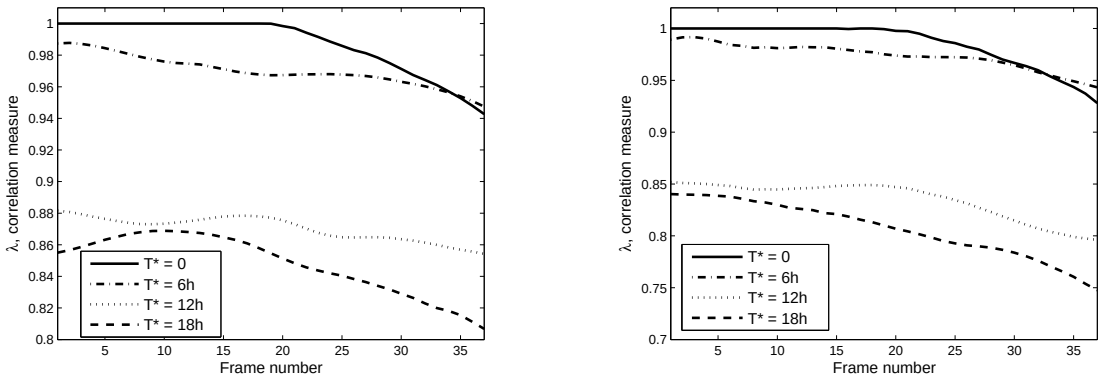
note that each $\mathbf{x}_i$ represents a training image ($1 \leq i \leq n$) as a $d \times 1$ vector, where d is the number of data points in each frame and $\mathbf{u} = [u_1, u_2, ..., u_n]$ is an $n \times 1$ vector containing the chosen peak values for the training frames.

The similarity measure of each forecast FTLE-LCS frame, λ_i, to the whole time-varying reference is obtained by

$$\lambda_i = \mathbf{Y}_i^* \mathbf{h}, \tag{2.19}$$

Fig. 2.14 shows an example of applying this method for attracting LCSs associated with different T^* and Ts. This figure displays the measure of similarity of each frame (horizontal axis) to the whole series of archive-based LCSs. One observes that how the overall similarity decreases as T^* increases. Also it is concluded that similarity measure in this sense is not sensitive to integration time between -24 to -48 h. Similar results are expected for different interrogation windows.

In the case of repelling LCSs (not reported), the range of similarity measure is smaller ($\sim 0.6 - 0.8$) for integration time $24 \leq T \leq 48$ h and similar to attracting features that measure is not sensitive to integration time.



(a) (b)

Figure 2.14: Measure of similarity of each forecast frame (attracting LCS) to the whole archive-based FTLE-LCS snapshots. Integration time (a) $T = -24$ h (b) $T = -48$h and $T^* = 0$ h, 6 h, 12 h and 18 h.

One should note that this method is appropriate when LCS features do not change much during the interrogation window, in other words, the rate of change of the patterns should be small in that time interval. Applying this method in cases where the LCS change is large yields unpredictable correlation measures which are not useful for measuring correlation between forecast- and archive-based patterns. In addition, this approach is suitable when overall similarity is important, but if one requires a measure of distance between LCS features, other methods should be considered. In next section, we quantify the distance between LCS patterns by applying the modified Hausdorff method.

2.5.5 Modified Hausdorff distance

Hausdorff distance is an extension to the Euclidean metric (Rockafellar and Wets, 2005). This measure describes how far two subsets of a metric space are from each other. The original definition of Hausdorff distance requires the objects to be closed and bounded (satisfying the axioms of metric space); however, LCSs are not closed features, so we cannot apply the Hausdorff method for them.

The modified Hausdorff method is designed to overcome this weakness (Dubuisson and Jain, 1994). By using this method one can calculate the *distance* between LCS features.

We denote the Euclidean distance between two points α and β as $d(\alpha, \beta) = \|\alpha - \beta\|$. The distance between a single point (α) and a set of points $\mathfrak{B} = \{\beta_1, \beta_2, ..., \beta_n\}$ is defined as $d(\alpha, \mathfrak{B}) = min_{\beta \in \mathfrak{B}} \|\alpha - \beta\|$.

In the modified Hausdorff method, distance between two sets $\mathfrak{A} = \{\alpha_1, \alpha_2, ..., \alpha_m\}$ and $\mathfrak{B} = \{\beta_1, \beta_2, ..., \beta_n\}$ is defiend as

$$\mathfrak{D} = max\left(d(\mathfrak{A}, \mathfrak{B}), d(\mathfrak{B}, \mathfrak{A})\right), \tag{2.20}$$

where $d(\mathfrak{A}, \mathfrak{B})$ is

$$d(\mathfrak{A}, \mathfrak{B}) = \frac{1}{m} \sum_{\alpha \in \mathfrak{A}} d(\alpha, \mathfrak{B}). \tag{2.21}$$

This choice for defining the distance is based on two requirements: sufficient discriminatory power and increasing distance measure when two sets are more distant.

We have the hyperbolic LCSs as a set of discrete points from criteria (2.6) and (2.8), so we can find the distance between features of two frames and consider it as a measure of similarity; smaller distance means better match between patterns.

Figure 2.15 shows the distance between archive- and forecast-based LCSs for integration times $T = 24, 48$h and different T^*s. This figure shows how the patterns of forecast results diverge from the true answers. Panels (a) and (b) of this figure show that by increasing the integration time for attracting LCSs (backward FTLE), the forecast results would be closer to the true features.

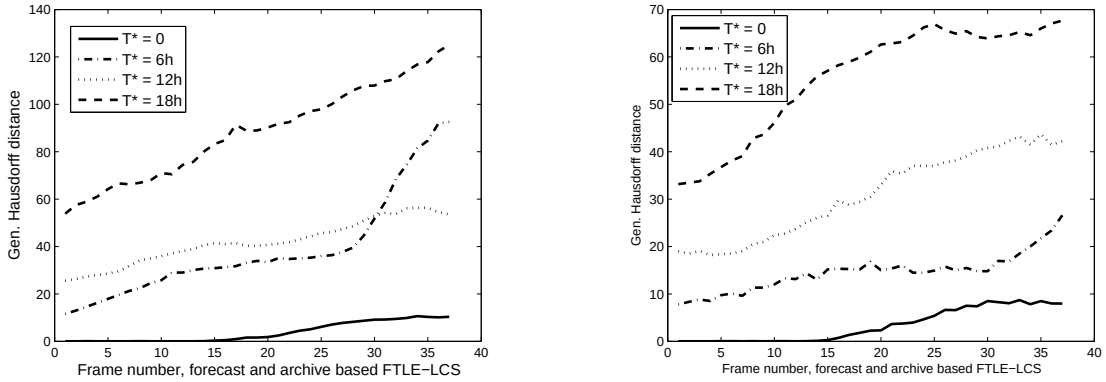


Figure 2.15: Modified Hausdorff distance between each frame of forecast- and archive-based LCSs. (a) integration time $T = -24$ h, (b) integration time $T = -48$ h and $T^* = 0$ h, 6 h, 12 h and 18 h.

An interesting point of this figure is the abrupt change in the curve corresponding to $T^*=$ 6 h. A probable reason could be an emerging error in the forecast velocity fields due to the input data associated with that T^*. This noticeable change is a result of nonlinear and chaotic dynamics of atmospheric models in which small changes of the input data could yield large differences at future times. By increasing the integration time to -48 h, the portion of archive data increases, so errors in the flow map are suppressed, so as one observes that the LCS patterns associated with $(T^*, T) = (6$ h, -48 h) have less distance to the archive-based results. Results (not reported) show that the repelling features are more vulnerable to the forecast input data. In contrast to the attracting features, by increasing the integration time, they become more distant from the archive-based LCSs.

2.6 Discussion

This study was motivated by recent observations suggesting that LCSs govern the large-scale atmospheric motion of airborne microorganisms (Tallapragada et al., 2011; Schmale et al., 2012). Such observations have the potential to aid in development of early warning systems for high risk plant pathogens in the future. As a part of this comprehensive system, UAVs are implemented to investigate the association of atmospheric LCS and microbial populations. To optimize the sampling of microbes at a fixed geographic location it is necessary to predict the attracting and repelling LCSs, which requires the use of wind forecast data. We use mesoscale forecast data over North America provided by NOAA-NCEP via NAM-218 to predict the passage of hyperbolic LCSs over the sampling location and plan for collecting samples with UAVs. This is directly linked to the NAM-218 timetable for online posting the forecast and archive data.

A part of this study sought to compare the forecast-based FTLE-LCSs with archive-based features to investigate the effectiveness of this approach in choosing correct flight times. In addition, quantifying the sensitivity of FTLE-LCSs predictions regarding the involved parameters such as (T, T^*) is the other goal.

Based on hyperbolicity of LCSs features, if the errors between the forecast and archive wind fields satisfy stringent criteria, the forecast LCS features mimic the archive-based ones with some minor differences (Haller, 2002). However, we showed that these criteria are not met in practice. As a result, forecast- and archive-based LCS features could show significant differences in time and space.

In this paper, we considered five methods for comparing the forecast- and archive-based FTLE fields: (1) fixed spatial point with temporal comparison, (2) fixed time with spatial cross-correlation, (3) spatiotemporal POD mode cross-correlations, (4) composite correlation

filter and (5) modified Hausdorff distance.

Results from the first method allowed us estimate the validity of the forecast in the absence of the archive results. This may be valuable for early warning systems, e.g., for vast crop fields, where the exact passing times of LCSs are not important, but it is important to know the overall shape and features of the LCSs and the probability of passing colonies of microorganisms formed by LCSs.

The second and third methods showed that attracting LCS features are typically more robust to wind field forecast errors compared to repelling structures. This point is significant since we know that the attracting LCSs are the backbone of mixing in moving fluids, governing important future events (Haller and Sapsis, 2011). In the presence of unavoidable velocity field errors, we expect the attracting LCSs to be more reliable than repelling ones.

Although the cross correlation technique is a well-established method, our results show that this approach of measuring FTLE field similarity is very sensitive to nonlinear deformations of the patterns, resulting in small coefficients for patterns that are similar by visual perception. The fourth applied method addresses this issue. Composite correlation filters capture the overall similarity of compared patterns better than simple correlation techniques but they are not perfect since they need to be trained with enough number of slow-varying input images, i.e., if the rate of change of the FTLE-LCS features is small enough this approach works well. Finally, the modified Hausdorff distance is considered as an additional method for comparing LCS features. A combination of these five methods could help us to reach a comprehensive verdict about the effects of forecasting parameters on the quality of forecast FTLE-LCSs.

The most challenging result of this study, which motivates future research, is as follows: if we want to forecast LCSs, we must take into account the *uncertainty* of the wind field forecasts

and the resulting uncertainty of the flow maps. By considering the concept of uncertainty, we would expect probabilistic predictions of the FTLE-LCS features which would be more consistent for real applications (e.g., crop pest monitoring).

Having reliable predictions of LCSs along with a network of sampling centers which provide data about infected crop areas (potential sources of inoculum) would contribute to an early warning system. This study has shed some light on the effect of key parameters on the quality of FTLE-LCS forecasts and hopefully it will lead to robust applications of LCS-based management strategies.

Acknowledgments

This material is based upon the work supported by the National Science Foundation under Grant Numbers DEB-0919088 (Atmospheric transport barriers and the biological invasion of toxigenic fungi in the genus *Fusarium*) and CMMI-1100263 (Dynamical Mechanisms Influencing the Population Structure of Airborne Pathogens: Theory and Observations). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. We thank Phanindra Tallapragada for providing codes for pre-processing the numerical data and Ali Aryanpour for his constructive comments.

Chapter 3

Atmospheric Lagrangian coherent structures considering unresolved turbulence and forecast uncertainty

Abstract

To obtain more realistic approximations of atmospheric Lagrangian coherent structures, the material surfaces which form a template for the Lagrangian transport, two concepts are considered. First, the effect of unresolved turbulent motion due to finite spatiotemporal resolution of velocity field data is studied and the resulting qualitative changes on the FTLE field and LCSs are observed. Stochastic simulations show that these changes depend on the probabilistic distribution of position of released virtual particles after backward or forward time integration. We find that even with diffusion included, the LCSs play a role in structuring and bifurcating the probability distribution. Second, the uncertainty of the forecast FTLE fields is analyzed using ensemble forecasting. Unavoidable errors of the forecast velocity data due to the chaotic dynamics of the atmosphere is the salient reason for errors of the

flow maps from which the FTLE fields are determined. The common practice for uncertainty analysis is to apply ensemble forecasting and here this approach is extended to FTLE field calculations. Previous work has shown an association between LCS passage and fluctuations in microbial populations and we find that ensemble FTLE forecasts are sufficient to predict such passages one day ahead of time with an accuracy of about 2 hours.

3.1 Introduction

The notion of hyperbolic Lagrangian coherent structures (LCSs) provide a framework for understanding transport and mixing phenomena especially in the case of passive particles in fluid systems (Haller and Poje, 1998; Haller and Yuan, 2000; Shadden et al., 2005). These structures are codimension 1 manifolds (or material surfaces) which effectively separate two regions of fluid with different qualitatively different past histories or fates. Several such critical material surfaces are present in any given geophysical flow and govern transport patterns. In the atmosphere, they persist from a few hours to a few days.

The present study is motivated by the role of atmospheric LCSs in aerial transport of microorganisms and the statistical correlation between sudden changes in aerobiota density and passage of LCS features over a fixed location (Tallapragada and Ross, 2013; Tallapragada et al., 2011; Tallapragada, 2010; Schmale et al., 2012; Schneider et al., 2005; Lin et al., 2013). Considering that role, if one can predict the LCSs to a good degree of accuracy and reliability, then important knowledge about the front propagation of microorganisms would be available.

Considering this fact, in a previous study, forecast LCSs were compared with reanalysis (pastcast) results to ascertain the accuracy and reliability of the forecasts (BozorgMagham et al., 2013). Based on that study we infer the need for including more realistic considerations

of atmospheric flows such as (i) finite spatiotemporal resolution of the available fluid velocity data, (ii) the accuracy of the fluid velocity data and (iii) the uncertainty of the forecast velocity fields. There have been studies about each of these concerns. For example, Griffa et al. (2004) and Ozgokmen et al. (2000) studied the predictability of Lagrangian trajectories and Kahl and Samson (1986) considered the uncertainty in trajectory calculations due to low resolution data. In addition Wilson and Sawford (1996) and Stohl (1998) studied different models for generating Lagrangian stochastic trajectories and Palmer (2000) and Ehrendorfer (1997) reviewed the concepts of predictability and uncertainty of atmospheric forecasts. Also Kalnay (2003) represented various approaches for data assimilation, ensemble forecasting and uncertainty analysis. Regarding LCSs, Haller (2002) discussed the errors of approximate velocity field and their effect on hyperbolic LCS features, Lermusiaux et al. (2006) described the uncertainty of oceanic LCSs and their numerical studies indicated that the more intense FTLE ridges are usually more certain. Moreover, Olcay et al. (2010) studied the role of flow field resolution and random errors on LCS identification. Finally, Turbulent dispersion velocity at length scales smaller than wind data grid was considered by Peng and Peterson (2012) and it was shown that attracting LCS structures coincide with the regions of high particle concentration. In that study, finite time Lyapunov exponent (FTLE) fields and the associated LCSs were calculated by using deterministic atmospheric flow map and the volcanic ash particles' dispersion was calculated by adding random walk and deposition velocities to the background velocity field.

In this study we connect notions such as unresolved turbulence and uncertainty of flow field to the atmospheric LCSs. For this aim we consider two concepts. First we show the effects of unresolved turbulence on the FTLE scalar field and the resultant LCSs (we will often refer to FTLE–derived LCSs as FTLE–LCSs) by adding the stochastic component of displacement to the deterministic flow map. As a result, the trajectories will be stochastic

and non–differentiable instead of deterministic and smooth. We observe significant changes in the (probabilistic) position of particles and the associated FTLE fields. We show that the spatiotemporal dependence of the stochastic velocity component to the time–varying deterministic background velocity field (Draxler and Hess, 1998; Fay et al., 1995; Legg and Raupach, 1982) plays an important role in the determination of the probabilistic distribution of the end–position of released virtual particles and consequently on the stochastic FTLE fields. We should note that Peng and Peterson (2012) just considered the dispersion velocity for the ash particles and the FTLE–LCSs were calculated by deterministic velocity field.

Second, we study the quantification of uncertainty and reliability of the FTLE–LCS when forecast velocity data are used to generate flow maps. The effects of forecast velocity data on the accuracy of forecast FTLE–LCSs have previously been observed (BozorgMagham et al., 2013). In that paper we compared FTLE–LCSs from forecast velocity data with archive based FTLE–LCSs. Results of that study show the sensitivity of FTLE–LCS forecasting to different parameters such forecast lead time, but because we used a deterministic forecast velocity field (NAM–218 data set) we were not able to measure the uncertainty of the forecast FTLE–LCS results. In the present study we propose a practical approach to measure uncertainty of forecast FTLE–LCSs by using ensemble forecasting concepts and considering the distribution of local FTLE values. In addition, we study the distribution of error of LCS forecast passage time over a fixed geographical location to have a measure of reliability of the forecast results.

3.2 Effects of unresolved turbulence on the FTLE field; stochastic FTLE field

In this section we show how including the stochastic component of unresolved turbulent velocity changes the FTLE field.

In atmospheric applications, the spatial resolution of operational data could vary from the order of 10 km to more than 250 km and the temporal resolution is usually of the order of 3 to 12 hours. For example, in the operational model NAM–218, which we use, the spatial resolution is about 12 km and the temporal resolution is 6 hours (a short term forecast at each intermediate 3 hours is also available). As a result of this coarse resolution, important *unresolved motions* with significant effects on the particle flow map and the resultant FTLE field could exist. To investigate this, we use a particle dispersion model to calculate the stochastic turbulent velocity component of the flow field. These components are functions of turbulent diffusivity (Draxler and Hess, 1998; Fay et al., 1995). Considering the Lagrangian frame, a fluid particle trajectory $\mathbf{x}(t)$ is governed by the stochastic differential equation,

$$\frac{d\mathbf{x}}{dt} = \mathbf{v}[\mathbf{x}(t)] \tag{3.1}$$

where the velocity of the particle is composed of the deterministic grid scale velocity, $\bar{\mathbf{v}}(\mathbf{x}, t)$, and the stochastic turbulent fluctuation component, $\mathbf{v}_t(\mathbf{x}, \bar{\mathbf{v}}, t)$, respectively,

$$\mathbf{v}(\mathbf{x}, t) = \bar{\mathbf{v}}(\mathbf{x}, t) + \mathbf{v}_t(\mathbf{x}, \bar{\mathbf{v}}, t). \tag{3.2}$$

Using the Langevin equation and assuming a Markovian process (Øksendal, 2003; Lemons

and Langevin, 2002), one can parameterize the turbulent component of velocity as,

$$dv_{t_i} = a_i(\mathbf{x}, \mathbf{v_t}, t)\ dt\ +\ \Sigma_{j=1}^{3} b_{i,j}(\mathbf{x}, \mathbf{v_t}, t)\ dW_j \tag{3.3}$$

where drift (a_i) and diffusion ($b_{i,j}$) are functions of time, position and turbulent velocity and v_{t_i} represents the i^{th} component of $\mathbf{v}_t\,(\mathbf{x}, \bar{\mathbf{v}}, t)$ (Thomson, 1987; Wilson and Sawford, 1996; Stohl, 1998; Stohl et al., 2005).

There are different methods for estimating the drift and diffusion terms with respect to the available data and turbulent regime of the flow field. In this study we follow the work of Thomson (1987); Fay et al. (1995); Legg and Raupach (1982); Stohl et al. (2005); Draxler and Hess (1998).

3.2.1 Grid scales velocity data

For generating the deterministic flow map we use numerical data provided by the National Oceanic and Atmospheric Administration (NOAA) and National Centers for Environmental Prediction's (NCEP) Operational Model Archive and Distribution System (NOMADS) project. We use one of the outputs of this nonhydrostatic mesoscale model, which is the North America Mesoscale, NAM–218, with data given on a grid of 614×428 points spaced at about 12.1 km covering North America. We consider a 6230×4670 km rectangular area as our wind velocity field and we use the smooth extension of the velocity field outside the boundaries whenever a particle leaves this domain (Haller, 2002).

Two types of data are used in our FTLE–LCSs calculations, archive data and forecast data. *Archive data* is the state of the atmospheric system determined by a combination of numerous measurements from, e.g., weather stations, weather balloons, satellites and any available atmospheric observations through data assimilation methods which use large–scale oceanic–

atmospheric geophysical fluid models. By *forecast data*, we mean the output of geophysical models which use the most current outputs of the data assimilation process (archive data) as input to estimate the state of the system for (near–)future times (Kalnay, 2003; Houtekamer and Mitchell, 1998; Evensen, 1994). In the NAM–218 data set, the temporal resolution of archived data is 6 hours corresponding to 00:00, 06:00, 12:00 and 18:00 UTC of each day. Accordingly, the forecast chain can start based on each of these archive time snapshots. Temporal resolution for forecast data is 3 hours and at each moment, the maximum available forecast data is for 84 hours lead time, i.e., ahead of the most recent archive data.

In this paper we will show the results for 2D flows over constant pressure surfaces which could be a simplified representation of the full 3D cases. Isobaric calculation is common in Lagrangian studies such as in the state–of–the–art implementation HYSPLIT[1] and it generates a first approximation to trajectories while avoiding the complexity of three–dimensional flow field integrations (Stohl, 1998; Stohl et al., 1995; Danielsen, 1961). In addition, these results can be regarded as examples of what is expected if more realistic conditions are considered for FTLE field computations. Finally, from (Branicki et al., 2011; Sulman et al., 2013) we know that 3D LCS features are nearly vertical within a layer whose thickness is sufficiently smaller than the ratio of the average horizontal velocities to their average vertical gradients. In that layer we may approximate the 3D structures with 2D features with a satisfactory accuracy. In our case of study that ratio can be estimated by the horizontal approximation of the wind velocity profile (Kaltschmitt et al., 2007) showing the upper and lower limits of that ratio $(\bar{u}/(\overline{\frac{\partial u}{\partial z}}))$ is between 250 to 1500m. So, we may conclude that within a layer centered at 850mb pressure surface, whose thickness is in the range of 25 to 150m (10% of that ratio), the 2D LCS features are acceptable approximation of the true 3D structures. Meanwhile more careful investigation about this approximation is left for future work.

[1] `http://ready.arl.noaa.gov/HYSPLIT.php`

3.2.2 Unresolved turbulent velocity component

By considering essential hypotheses such as the Kolmogorov similarity hypothesis (Rodean, 1987) and the criteria described by Thomson (1987), various solutions have been suggested for the unresolved turbulent velocity term. We use one which is implementable with respect to the available data (Fay et al., 1995; Draxler and Hess, 1998). In this method, components of $\mathbf{v}_t(\mathbf{x}, \bar{\mathbf{v}}, t)$, which are shown as v_{t_i}, are estimated by a Markov–chain formulation as,

$$v_{(t+\delta t)_i} = R_{\delta t} v_{t_i} + \left(1 - R_{\delta t}^2\right)^{0.5} \xi, \tag{3.4}$$

where $R_{\delta t}$ is the turbulent velocity autocorrelation coefficient. This coefficient is approximated from the Lagrangian time scale of the flow system as,

$$R_{\delta t} = \exp\left(-\delta t / T_L\right), \tag{3.5}$$

where δt is the integration time step and the Lagrangian time scale (T_L) is considered to be constant over the pressure surface, assuming δt to be sufficiently smaller than T_L. In our calculations δt is between 5 to 15 minutes and T_L is 180 minutes (Draxler and Hess, 1998). For calculation of the second term of (3.4), only lateral diffusion within constant pressure surfaces is considered and density variations over those surfaces are neglected (Smagorinsky, 1963). The ξ term is estimated with respect to the standard deviation of the turbulent velocity σ, as,

$$\xi = \mathcal{N}(0, 1)\, \sigma, \tag{3.6}$$

where $\mathcal{N}(0, 1)$ is the standard normal distribution and σ at each point is estimated as,

$$\sigma = \sqrt{\kappa / T_L}. \tag{3.7}$$

where the sub–grid scale horizontal mixing coefficient, κ, is computed from the velocity deformation tensor as,

$$\kappa = 2^{-0.5} (c\chi)^2 \left[\left(\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y} \right)^2 + \left(\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y} \right)^2 \right]^{0.5}, \tag{3.8}$$

where χ is the meteorological data grid size and c is an empirical constant equal to 0.14 (Draxler and Hess, 1998).

The set of equations (3.1)–(3.8) provides all one needs to calculate the stochastic velocity and stochastic trajectory of particles with respect to the background 2D velocity field. This process can be repeated for 3D flows by considering the vertical component of the velocity (Stohl et al., 2005) as well.

3.2.3 Stochastic FTLE field (SFTLE)

Similar to the deterministic case, the essential step for defining the stochastic FTLE (SFTLE) field is to define the stochastic flow map as,

$$\Phi_{t_0}^{t} : \mathbf{x}_0 \rightarrow \mathbf{X}(t, t_0, \mathbf{x}_0) \tag{3.9}$$

where conventionally deterministic (sure) values are written in lower case and random (stochastic) variables are represented by capital letters. Regarding the stochastic trajectory definition, the stochastic FTLE field is *defined* for the random variable $\Phi_{t_0}^{t}(\mathbf{x}_0)$ as,

$$\Sigma_{t_0}^{t}(\mathbf{x}_0) = \frac{1}{|t - t_0|} \log \left\| D\Phi_{t_0}^{t}(\mathbf{x}_0) \right\| = \frac{1}{|t - t_0|} \log \sqrt{\Lambda_{max} \left[\mathfrak{C}_{t_0}^{t}(\mathbf{x}_0) \right]} \tag{3.10}$$

where the right Cauchy–Green strain tensor, $\mathfrak{C}_{t_0}^t(\mathbf{x}_0)$, and the deformation gradient (Jaccobian), $D\Phi_{t_0}^t$, are defined with respect to the stochastic map $\Phi_{t_0}^t$ and consequently they are random variables.

Assuming that the initial position of a particle, $\mathbf{x}_0$, is known with infinite precision, the main difference between deterministic and stochastic FTLE fields is that when the deterministic velocity field is considered, the source or destination (referring to backward and forward integration respectively) of a released particle would be a point, but when unresolved stochastic turbulence is considered that single point becomes a probabilistic source or destination distribution.

For calculation of the probabilistic distribution of the end–position of trajectories, we divide the domain into small square boxes $\{B_1, B_2, \cdots\}$ of side length d. Then by using the Monte Carlo method in forward or backward integration, the number of final landings in each box is counted. We re–index all the boxes by number of landings per box $\left\{B^{(1)}, B^{(2)}, \cdots\right\}$ such that $n_{B^{(1)}} \geqslant n_{B^{(2)}} \geqslant \cdots$, where $n_{B^{(i)}}$ is the number of landings in box i. For $\alpha \in (0,1)$, we define the probabilistic α–source (or destination) area as,

$$A_\alpha\left(\mathbf{x_0}, t, t_0\right) = \bigcup_{i=1}^{k} B^{(i)}, \quad \text{where} \quad k = \inf\left\{ j \in \mathbb{N} \mid \frac{\sum_{i=1}^{j} n_{B^{(j)}}}{N} \geq \alpha \right\} \tag{3.11}$$

and N is the total number of released virtual particles from each reference point $\mathbf{x_0}$ (number of realizations of the Monte Carlo method). The interpretation is that A_α contains a fraction α of the ensemble of particles, starting with the most dense boxes. Our simulations show that area(A_α) converges to a fix number for each case by choosing sufficient number of realizations and considering small size for boxes (i.e., we have convergence as $d \to 0$).

For convenience, let us define the integration time $T = t - t_0$. For $T < 0$ (backward integration) we call A_α the "probabilistic α–source area" and for $T > 0$ (forward integration),

A_α is called "probabilistic α–destination area". Note that for both cases, N particles are released from each grid point of the domain ($\mathbf{x_0}$) and then by time integration the end positions are calculated.

Fig. 3.1 shows the backward solutions starting from the reference point (shown by the black circle) for two cases. The dark collection of points shows the probable source points for integration time -12 h and the cyan region shows the probabilistic region for integration time -24 h. As observed, the distance between the deterministic solutions (red circles) and the centroid of the stochastic realizations (red diamonds) increase with time, also it is noticeable that since the probable source areas are stretched, the deterministic solution or the centroid of the stochastic solutions are not good representatives for those regions. In §3.2.4, the qualitative effects of probable source regions on stochastic FTLE fields are discussed for different cases.

In general, the final probabilistic distribution of arbitrary individual particles after integration time T is not Gaussian since drift and diffusion terms are functions of position, velocity and time. As a result, the deterministic final position of a particle after the integration time is not at the same position of the centroid of the stochastic realizations. In general, to find the exact probability distribution one needs to solve the Fokker–Planck equation, however as a practical approach one can use the Monte Carlo method to make an estimate of that distribution (Barkai et al., 2000; Kinzelbach, 1988).

3.2.4 Results; stochastic FTLE fields

As a first case, we consider the backward FTLE field for 12:00 UTC 29 Sep 2010 over the eastern part of the United States. We consider quasi–2D flow on pressure surfaces and all the presented results correspond to the 850 mb surface.

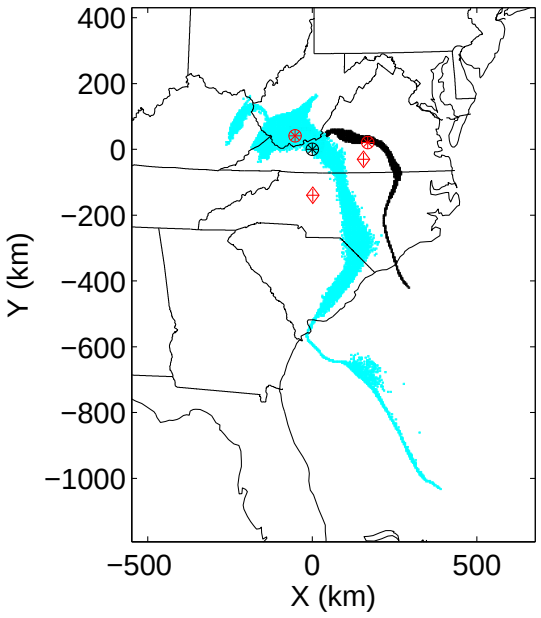
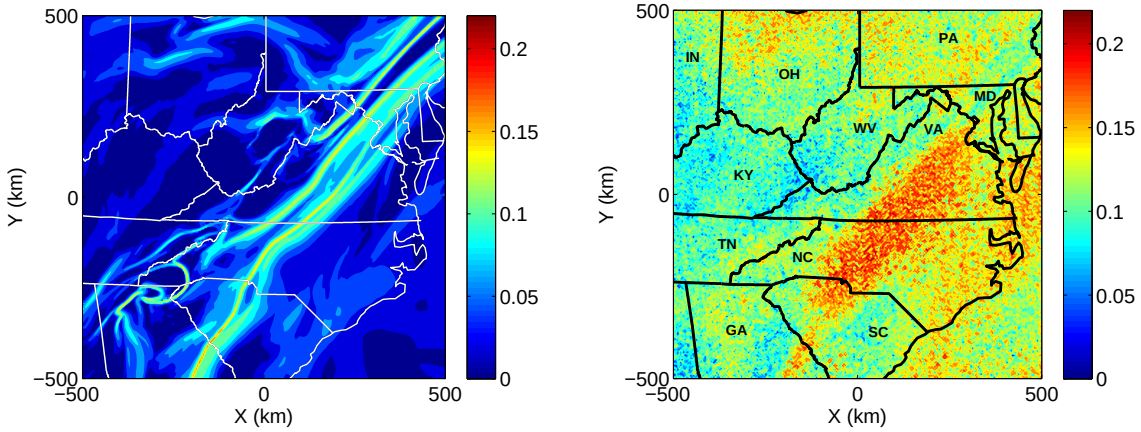


Figure 3.1: Collections of the probabilistic source points ($A_{90\%}$), backward integration times $T = -12$ h (black) and -24 h (cyan), black circle: reference point at (37°11′ N latitude and 80°35′ W longitude), red circles: deterministic solutions, red diamonds: centroid of the stochastic solutions. For this case 10^6 virtual particles are released from the reference point on the 850 mb pressure surface at time t_0 = 14:37 UTC 29 Sep 2010.

In Fig. 3.2(a) the deterministic FTLE is shown and we see sharp ridges which are candidates for hyperbolic attracting LCS features. This FTLE field was generated with archived data and integration time equals to -24 h. We can extract the true hyperbolic LCS features from this figure but our focus in this paper is on the FTLE fields so we just compare the FTLE fields (for more information about the hyperbolic LCSs of this figure refer to (Karrasch, 2012; BozorgMagham et al., 2013)).

Fig. 3.2(b) shows *one particular* realization of the stochastic FTLE field when the stochastic component of the unresolved turbulent velocity field is taken into account. We notice different features comparing the two panels of Fig. 3.2. For example, the sharp ridges of the deterministic case (panel (a)) over VA, NC and some part of SC become a broad high



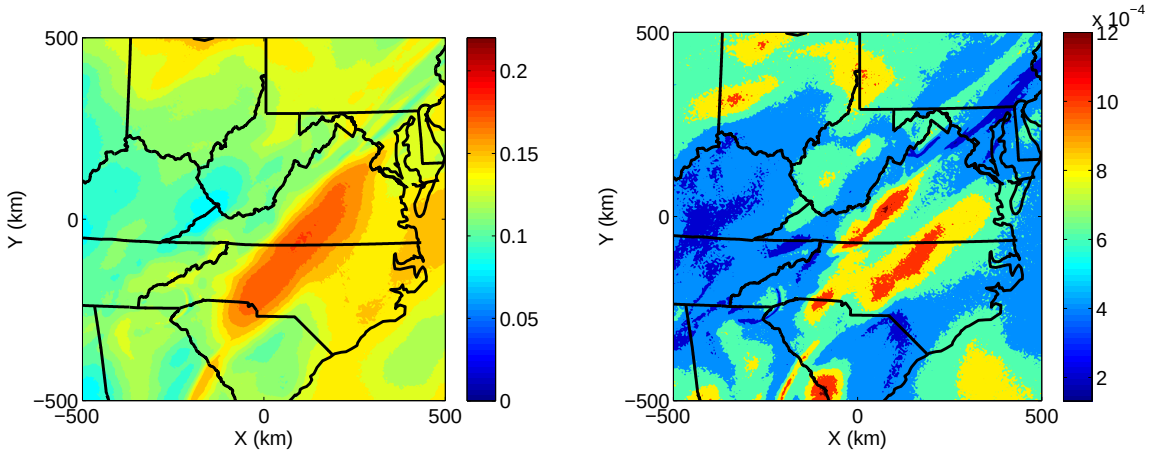
(a) (b)

Figure 3.2: A typical comparison between deterministic and stochastic FTLE field, (a) Deterministic backward FTLE field (b) One realization of the stochastic backward FTLE field; Integration time $T = -24$ h, $t_0 = 12{:}00$ UTC 29 Sep 2010, and the velocity field corresponds to the 850 mb pressure surface. Notice different regions comparing the two panels and also the higher level of FTLE coefficient in panel (b). The states are labeled by their postal code for brevity.

value region in panel (b). Meanwhile, one can observe that sharp and strong deterministic ridges over PA and MD are washed out in panel (b) which suggests it is not necessary to have a broad and high value cloud of stochastic FTLE wherever deterministic high value ridges exist. This observation is important since it shows different behaviors when two similar cases—parallel and strong ridges of FTLE field—are considered. We can also see that the curly ridge over south west NC does not exist in panel (b). Note that the graininess of Fig. 3.2(b) is due to the inherent and discontinuous stochastic behavior.

To investigate the common patterns of the stochastic FTLE fields we use the expected value of those fields. Fig. 3.3(a) is the expected value of the stochastic FTLE fields for 1000 particular realizations obtained by the Monte Carlo method. As is common, the expected value smooths out the graininess of individual realizations. Fig. 3.3(a) shows an important point which is difficult to observe in Fig. 3.2(b). In this panel we observe a narrow cloud of high value SFTLE over GA at the same place where we have a strong deterministic ridge.

This observation shows that it is possible to have sharp deterministic FTLE ridges which remain as narrow high value regions of SFTLE; but not always.



(a) (b)

Figure 3.3: 1000 stochastic realizations are used for calculation of (a) expected value of the stochastic backward FTLE fields (b) standard deviation of the stochastic backward FTLE fields; Integration time $T = -24$ h, $t_0 = 12{:}00$ UTC 29 Sep 2010.

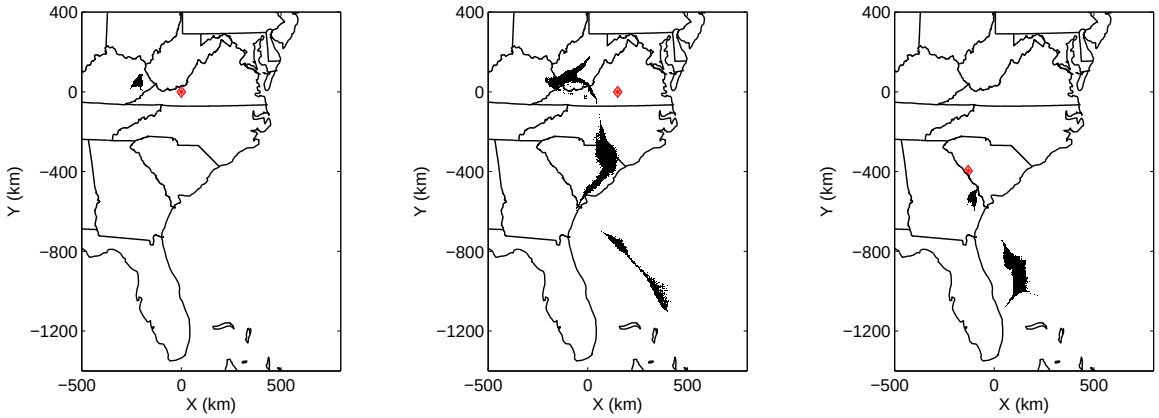
Fig. 3.3(b) is the standard deviation of the stochastic FTLE fields. We observe that the value of standard deviation is much smaller than the expected values with a maximum order of 6%. This observation means that the stochastic realizations are not too "separated" from each other. This is an important remark since the numeric behavior of typical stochastic differential equations (SDEs) depends on their coefficients and by changing the coefficient one can theoretically get significantly different distributions, e.g., in Ornstein–Uhlenbeck processes (Risken and Eberly, 1985). Based on results corresponding to different days, we observe that the standard deviations in atmospheric SFTLEs are typically small comparing to their expected value.

To have a better understanding about different regional changes, we investigate the sensitivity of the probability distribution of forward and backward trajectories with respect to the reference point, i.e., the sensitivity of $A_\alpha(\mathbf{x_0}, t, t_0)$ with respect to $\mathbf{x_0}$. We observe

that source (or destination) areas corresponding to different target (or starting) locations are qualitatively different. In those cases that the source (or destination) areas are small and close to the deterministic solutions, the *local* FTLE field's changes are small, but in cases that the source (or destination) areas are large and/or the centroid of the source (or destination) areas are far from the deterministic solutions, local changes of FTLE fields are noticeable.

Fig. 3.4 shows three cases of the probabilistic 50%–source area (each sub–figure represents 10^6 stochastic trajectories obtained from the Monte Carlo method) corresponding to different qualitative changes observed in Fig. 3.2(a) and Fig. 3.3(a). These cases correspond to different initial locations with the same initialization time. In the first case, Fig. 3.4(a), the source area is shown where $\mathbf{x_0}$ is located at Virginia Tech Kentland Farm $(37°11'N, -80°35'W)$ which is considered as the reference point (0,0). Fig. 3.4(b) shows the probabilistic source area when $\mathbf{x_0}$ is placed at $(37°10'30''N, -78°55'48''W)$ equal to 150 km eastward relative to the reference point, and the last case, Fig. 3.4 (c), refers to the source area when $\mathbf{x_0}$ is at $(33°40'48''N, -82°W)$ which corresponds to 131 km west and 396 km south respectively. For all three cases the backward integration time is -24 h. These three points and their corresponding probabilistic source areas illustrate the qualitative changes compared to the deterministic FTLE field when unresolved turbulent velocity is added. In the first case the source area is small and compact. We observe the same qualitative behavior for the other nearby local points (points inside a circle of radius $\mathcal{O}(d)$). So, we expect very small change in the local FTLE field around the point $(0, 0)$. The second case, Fig. 3.4(b), which is the corresponding probable source area for a point between two sharp ridges, where a band of high value SFTLE is formed, is large and stretched. And finally, the third source area refers to a region where the width of the sharp FTLE ridge is increased but not as much as the second case (referring to the ridges over GA in Fig. 3.2(a) and Fig. 3.3(a)). In Fig. 3.4(c)

the source area is not as large as the second one but larger then the first one (red diamonds show $\mathbf{x_0}$ for each scenario and the collections of black points show the probabilistic source areas). Before considering other features of SFTLE fields, we consider Fig. 3.4(b) as



(a) (b) (c)

Figure 3.4: Investigating the qualitative changes in a stochastic FTLE field due to the distribution of the probabilistic source areas, (a) probable source area for the reference point $(37°11'N, -80°35'W)$ corresponding to (0,0) (b) probable source area for $(37°10'30''N, -78°55'48''W)$ or 150 km east to of the reference point (c) probable source area for $(33°40'48''N, -82°W)$ equal to 131 km west and 396 km south respectively; for all the cases $\mathbf{x_0}$ is shown by a red diamond, $A_{50\%}$, $T = -24$ h, $t_0 =$ 12:00 UTC 29 Sep 2010 and box size d =4 km.

an example with separated source areas to see the effect of LCSs on shaping those areas. Fig. 3.5(a) shows the expected stochastic forward FTLE field for 12:00 UTC 28 Sep 2010, obtained from 500 sample members. When we *back* track the probable source points for $\mathbf{x_0}$ (red diamond) on Fig. 3.4(b), we expect that they settle down on *parts* of a stochastic repelling LCS feature since the integration direction is backward (repelling in forward time is equal to attracting in backward time). This attraction is noticeable when Fig. 3.4(b) is compared with Fig. 3.5(a). We notice that almost all the probable source areas lie down on the *repelling* feature. The separation of islands are governed by the stochastic attracting LCS features acting as repeller when applying inverse integration time direction. Fig. 3.5(b)

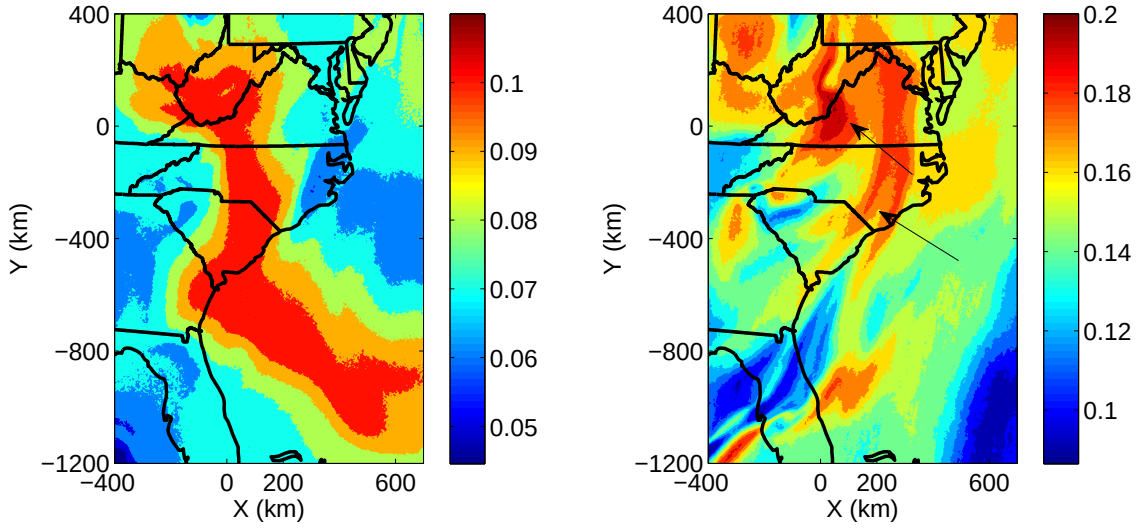


Figure 3.5: Role of stochastic FTLE fields in shaping the probable source area of the virtual particles, (a) expected forward SFTLE field. High SFTLE regions act as the source location of particles (b) expected backward SFTLE field. High SFTLE regions acts as cutting edges, thus density of the virtual particles is low near them. Both cases are obtained by averaging 500 members of SFTLEs, integration time is 24 h, time shown is 12:00 UTC 28 Sep 2010.

shows that isolated and bifurcated islands are indeed the results of backward "cutting" by attracting LCS features which are pointed by two arrows.

Panels of Fig. 3.6 illustrates how the probabilistic bifurcated source areas are influenced by the corresponding contemporaneous attracting LCSs during 24 hours of integration (three panels correspond to 24, 12 and 6 h of backward integration). In this figure we show the position of all 10^6 virtual particles. To keep the figure clear we only show the deterministic LCSs, as we know that these features usually act as the backbone of the stochastic LCS features. The cutting effects of LCS features manifests when smaller percentage of particles is shown, for example if we show 50% of particles, then the result would be Fig. 3.4(b) on which the bifurcated and isolated islands of probable source areas due to the LCSs are distinct. In the language of Bayesian statical inference, we can say the LCSs play a role in structuring the prior probability distribution of the source location.

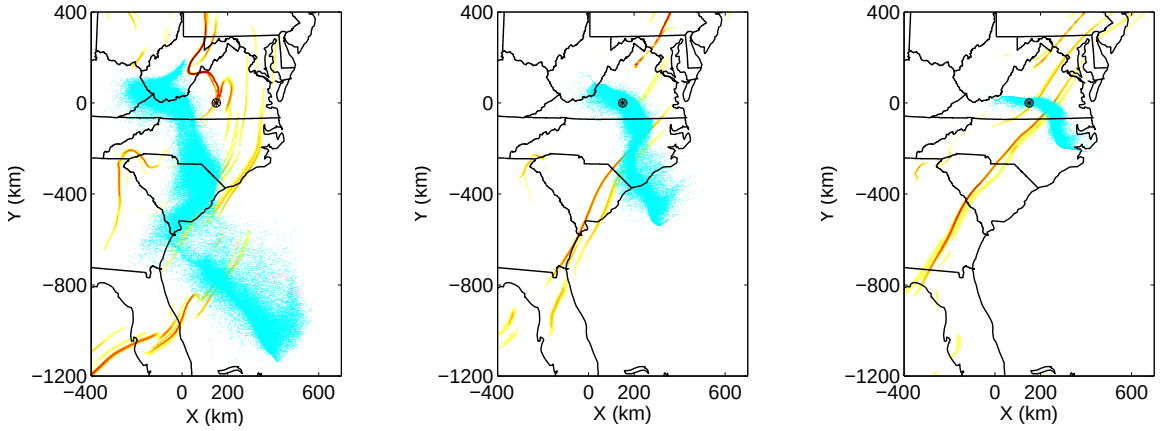
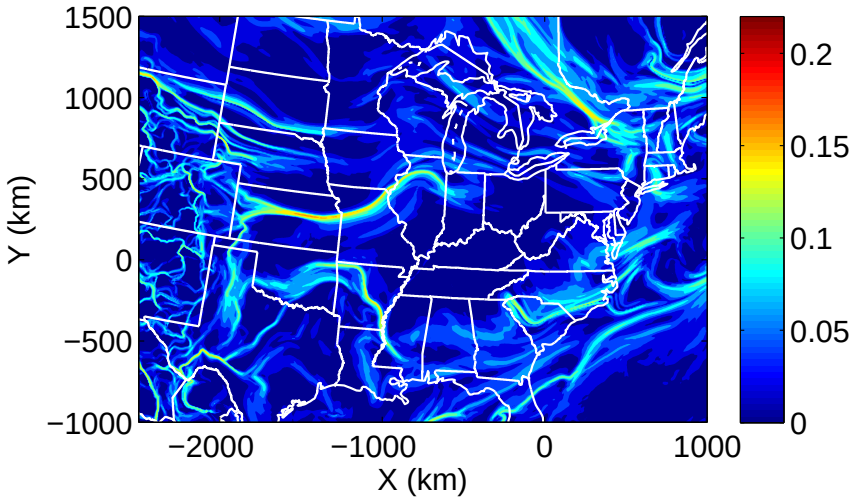


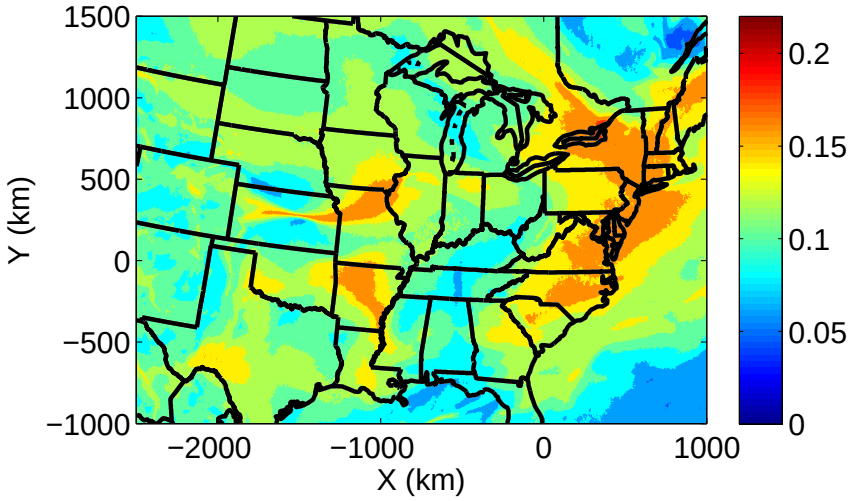
Figure 3.6: Probabilistic source area ($A_{99\%}$) at (a) T=−24 h, (b) T=−12 h, (c) T=−6 h and the role of attracting LCS features in forming the bifurcated probabilistic areas; initialization time t_0 = 12:00 UTC 29 Sep 2010, reference point $\mathbf{x_0} = (37°10'30''N, -78°55'48''W)$.

To study other probable features in stochastic FTLE fields we consider Fig. 3.7(a) which is the backward FTLE field for 12:00 UTC 8 Apr 2012. We can see a very strong, sharp and lengthy ridge over KS, MO and IL. The analog of this feature exists in Fig. 3.7(b), which is the expected value of SFTLE fields for 1000 realizations.

We can see that the tail of the long ridge over KS remains fairly sharp and the head becomes a broad region over MO. The most interesting feature of these two panels is over the eastern coast of the United States. In Fig. 3.7(a) there is no high value ridge over the eastern part but in Fig. 3.7(b) we notice a vast high value area covering all the eastern part of the figure; starting from NC continuing to the northern part of NY. We conclude that if we consider realistic conditions by adding unresolved turbulent motion, it is possible to have regions with high SFTLE values where the deterministic FTLE values are low. In fact these regions are regions of strong attraction or repulsion which can be important for describing the distribution of passive tracers. If we neglect the unresolved turbulence, then we may underestimate the importance of these regions.



(a)



(b)

Figure 3.7: (a) Deterministic backward FTLE field, (b) expected value, backward SFTLE field. Integration time $T = -24$ h, $t_0 = 12{:}00$ UTC 8 Apr 2012. Note the existence of the high value could of SFTLE field over the east coast of US where the deterministic FTLE field does not capture it. This region could be attracting area for passive tracers. Also note the varying width of the high vale SFTLE cloud over KS, MO and IL.

We summarize the observations into four categories:

1. Sharp ridges of the deterministic FTLE field deform into bands of high value expected SFTLE,

2. Sharp ridges of the deterministic FTLE field remain fairly sharp in stochastic cases,

3. Ridges of the FTLE field are washed out and disappear when unresolved stochastic motion is considered,

4. Low FTLE value regions in deterministic field become high expected SFTLE value regions.

An application of stochastic FTLE in geophysical flows is the prediction of spread of volcanic ash and dust. If we consider the deterministic solutions, we only obtain co–dimension one LCS features that are strongest in attraction or repulsion of nearby particles. But if we consider the stochasticity of the flow field, we may find regions of high attraction or repulsion strength that may not be observed by deterministic solutions. By using the archive data we can detect the backward SFTLE and their associated attracting LCSs where the particles settle down. Also by considering the stochasticity of the flow field we declare the boundaries between coherent sets as *fuzzy* objects instead of bright lines. Since unresolved turbulence is similar to diffusion process, in cases that we observe considerable diffusion such as atmospheric transports of aerosols, we could expect to see qualitative changes to the FTLE field analogues to our results for SFTLE fields.

3.3 Uncertainty analysis

We have considered the effect of unresolved turbulence but another important issue when atmospheric forecast velocity data is used to generate the forecast FTLE fields is the level of errors of the results with respect to the archive–based (observation-based) FTLE fields. In a previous paper (BozorgMagham et al., 2013) we observed that even if one uses the best resolution operational forecast data for nearly real–time forecasting, one would face inevitable

errors in resultant FTLE fields and their associated LCSs. This fact is due to the cumulative errors in the integration process for calculating the tracer particle flow map. One might compare Eulerian forecast and archive velocity fields and find that the level of error is small and the forecasts are skillful (Murphy, 1988), however, in the process of time integration, small errors accumulate and the forecast flow map diverges dramatically from the flow map obtained from archive/reanalysis data which is taken as the best representation of the true state of the system. So the quality of the velocity forecast field is essential. However it is well known that improving the quality of forecasts is limited by the atmosphere's inherent chaotic behavior (Lorenz, 1963). Regarding forecast FTLE fields and the associated LCSs, one has the option to either depend on the best resolution deterministic velocity fields or attempt to quantify the uncertainty of the forecast results.

The current practice for the study of uncertainty of forecast results is to use an ensemble forecasting approach. In this approach the outputs of the data assimilation process are perturbed and fed to the geophysical models which are assumed to represent the dynamics of the system. Different perturbation techniques are used in this regard, e.g., NCEP uses the breeding method while ECMWF implements the singular vector approach (Kalnay, 2003; Toth and Kalnay, 1997). The output members of the ensemble forecasting approach are suitable for different purposes, e.g., improving the quality of forecasting by ensemble averaging, providing indications of the reliability of the forecast or providing a quantitative basis for probabilistic forecasting (Kalnay, 2003; Toth and Kalnay, 1997).

We apply a similar approach for calculations of FTLE fields and the associated LCSs. In this approach, the ensemble members of the forecast velocity field are used to generate ensemble members of the FTLE field. By comparing these fields we have a measure of the uncertainty of the forecast fields over different geographical regions.

3.3.1 Velocity ensemble data set

To perform this case study we use the GFS ensemble forecast data in which the ensemble forecasts and the breeding cycles are initialized at 00:00 UTC each day and the control run is initialized from NCEP/NCAR reanalysis data.[2]

When the breeding approach is used in ensemble forecasting, a random initial perturbation (the random seed) with a small fixed initial norm is added and subtracted to the control input data, generating a forecast pair; see Fig. 3.8(a). The random seed is added at the beginning of the procedure, then the model is integrated for the both control and the perturbed data sets during a pre–determined length of time Δt. Results from the control set integration are subtracted from the perturbed one. For the next step the difference is scaled down to the initial perturbation norm and this new perturbation vector is added to the new state of the system calculated from the control set to produce the new perturbed initial condition. These steps are repeated for all the ensemble members until the pre–specified maximum integration time (Kalnay, 2003). This approach is conceptually similar to the process of finding the leading order Lyapunov vector/exponent in systems of ordinary differential equations. Moreover the perturbations depend on the dynamical responses of the underlying system and it is known that for atmospheric ensemble forecasting the bred vectors do not converge to a *single* leading bred vector (Kalnay et al., 2002).

At NCEP a self–breeding approach is used; see Fig. 3.8(b). In this approach the difference between positive and negative perturbed forecasts is divided by 2 and then scaled down to the initial perturbation norm. The scaled difference is added and subtracted from the control set, generating initial condition for the new pair of forecasts. For GFS ensemble data, 7 pairs are calculated, so we have 14 ensemble members plus one control case (Toth and Kalnay, 1997; Caplan et al., 1997; Uppala et al., 2005).

[2] `http://www.esrl.noaa.gov/psd/forecasts/reforecast/data.html`

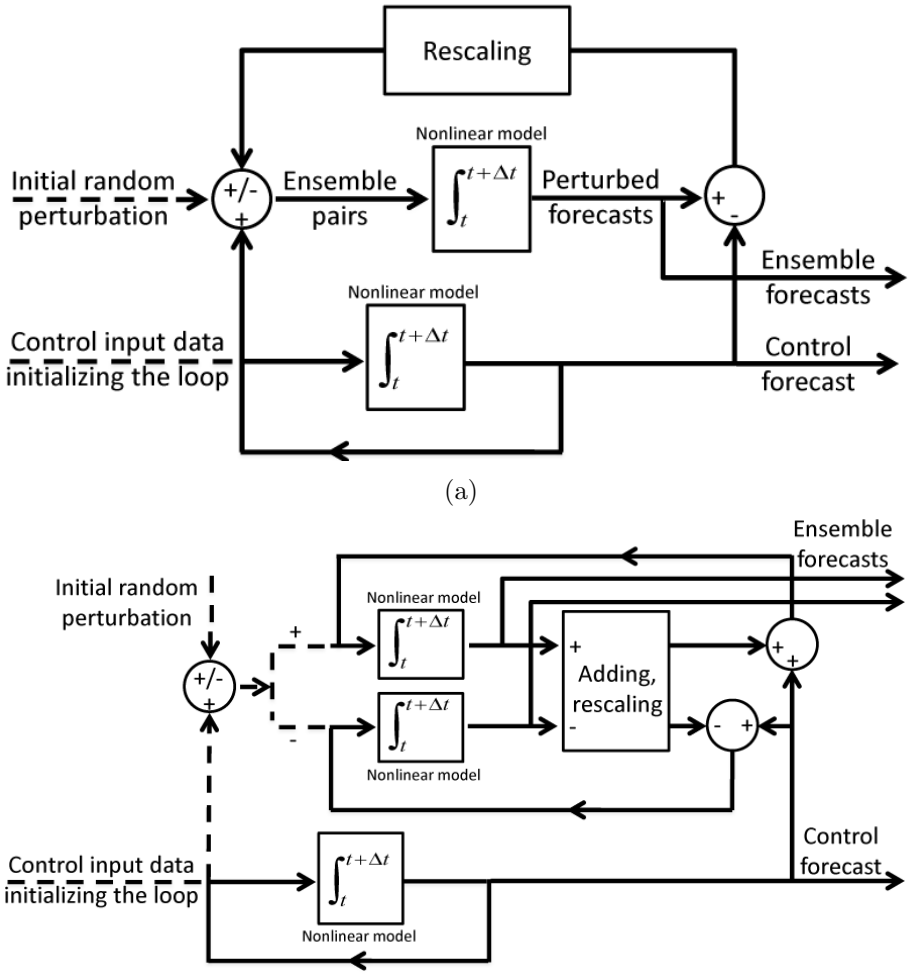


Figure 3.8: Schematic of the breeding cycle, (a) control run is integrated without any perturbation. Bred modes are initialized by adding random perturbation to the control case. Each ensemble member is integrated (solving a set of PDE equation with this initial condition) for a specified length of time, and then each field is compared with the control field. The difference between control case and each ensemble member is rescaled. This rescaled field is added/subtracted to the control case to generate the new ensemble members, and the integration cycle is repeated until the end of forecast time, (b) self–breeding pair of ensemble forecasts used at NCEP. In this approach corresponding pairs are used to generate the new initial condition for the new pair of forecasts. For both cases the initiation signals are shown by dash lines.

3.3.2 Ensemble FTLE field results

To demonstrate the applications of ensemble FTLE fields, we use the 850 mb pressure surface data for the velocity field. The time resolution of GFS data is 12 hours and the spatial resolution is 2.5°. The target time for the FTLE field calculation is 12:00 UTC 29 Sep 2010 the same as the first case shown in Fig. 3.2(a) and the backward integration time is −24 h. According to this target time, the forecast chain was initialized at 00:00 UTC 28 Sep 2010 and extended to 00:00 UTC 30 Sep 2010. Fig. 3.9 shows all 15 members of the ensemble FTLE fields calculated from ensemble members of the GFS velocity field. Fig. 3.9(a) is the FTLE field from the control velocity field and the other 14 members are the results from 7 pairs of bred modes. From this collection of FTLE fields we recognize the main patterns of the forecast result and the associated candidate LCSs. Also we notice the differences in ensemble members, e.g., the FTLE–LCS results from Fig. 3.9(j) and (k) are very different, but they correspond to a single pair of bred modes.

Fig. 3.10(a) and (b) represents the average and standard deviation (as a measure of spread of FTLE value over each point) of the represented ensemble FTLE field members in Fig.3.9. This figure shows that over some regions of the Florida panhandle one can strongly expect a high value FTLE ridge but over some other regions such as north–west South Carolina and especially south–east Canada the resultant FTLE–LCS are uncertain ([the last one is not due to the edge effect since the data domain is larger than what is considered for the FTLE field and also we apply the linear extension technique]). Panel (c) of Fig. 3.10 is the backward FTLE field result (integration time −24 h) from GFS reanalysis velocity data with spatiotemporal resolution 2.5° and 6 h. By using panels (a) and (c) of this figure, we can compare the results from forecast ensemble mean with the pure reanalysis based FTLE field. For example we observe a bifurcation in the LCS structures over the east coast which the mean of the forecast ensemble does not show. In addition there is a weaker but

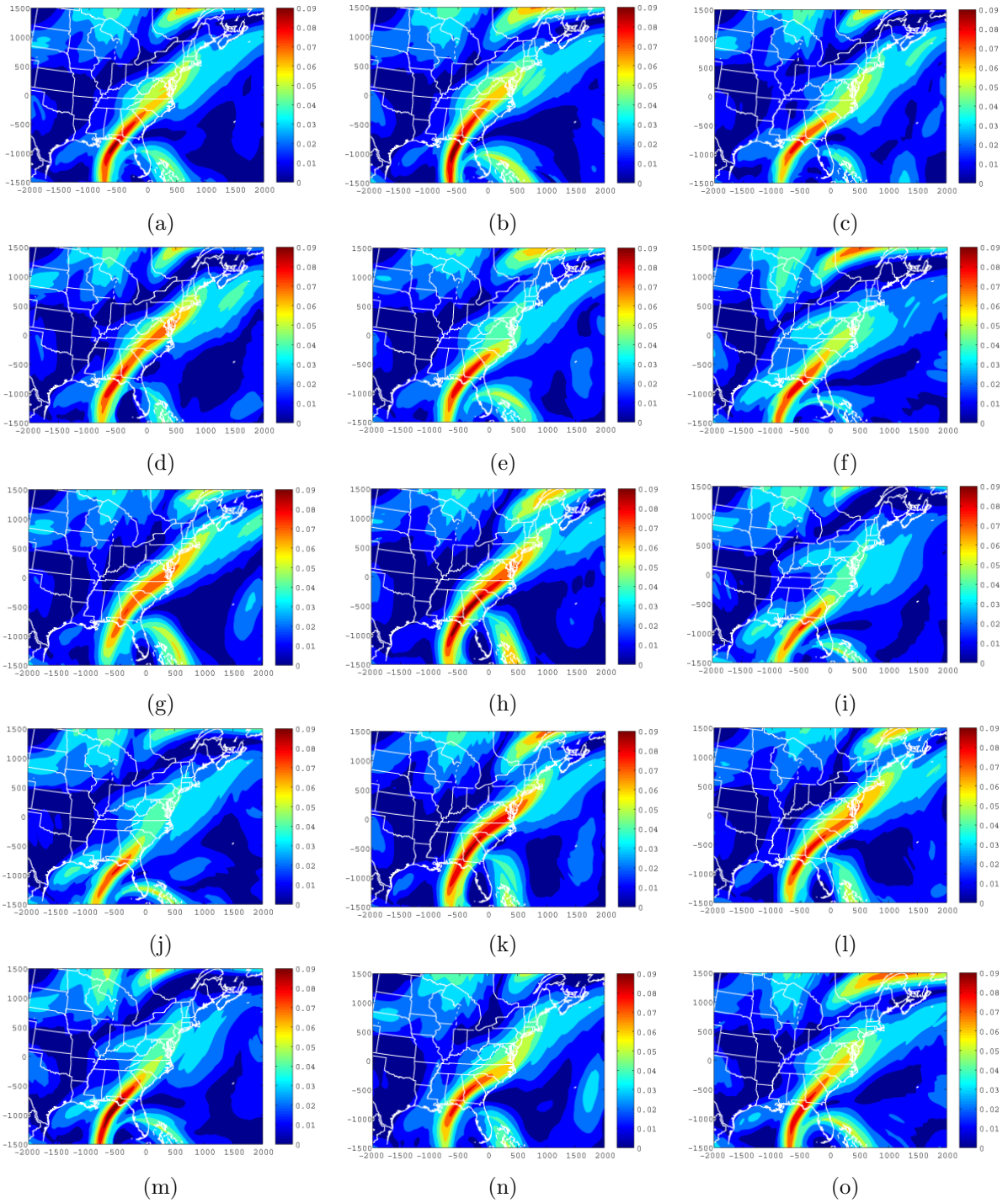


Figure 3.9: 15 ensemble members of forecast FTLE field for 12:00 UTC 29 Sep 2010, (a) FTLE from the control velocity field (b ... o) 14 other ensemble members of forecast FTLE filed obtained from perturbed velocity fields, integration time is -24 h for all the cases.

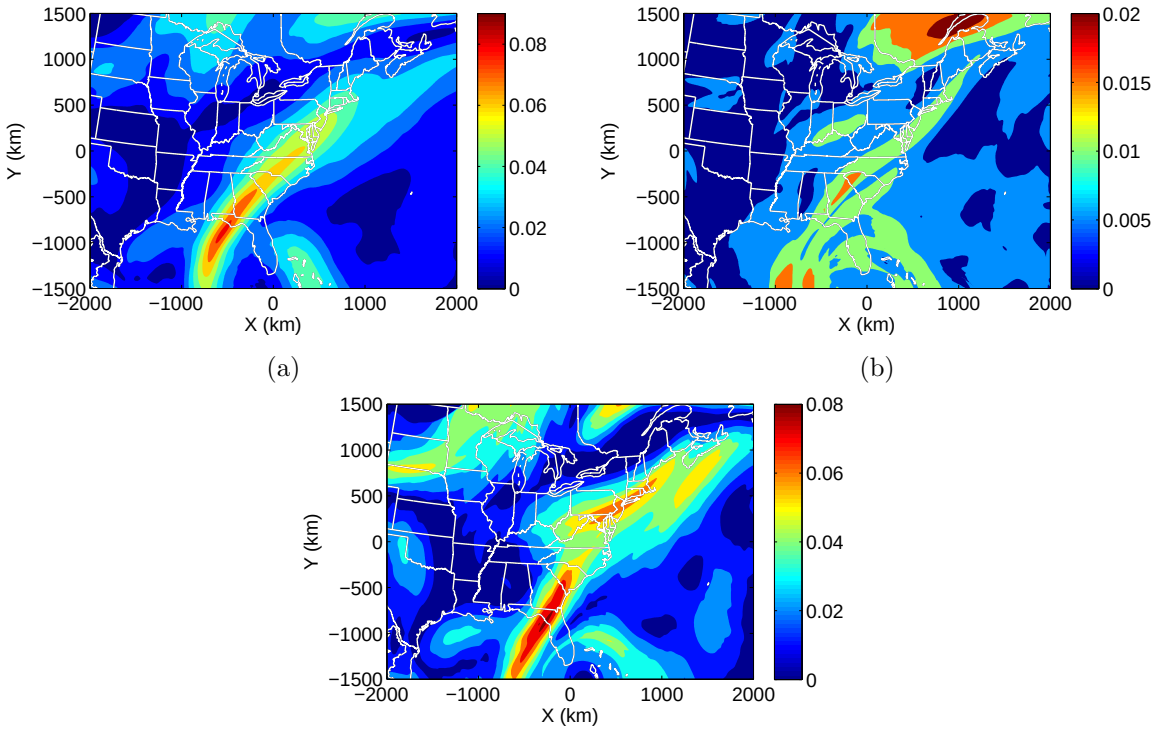


Figure 3.10: (a) Mean of the forecast ensemble members (b) standard deviation of the forecast ensemble members (c) FTLE from GFS reanalysis data; corresponding time and integration time are 12:00 UTC 29 Sep 2010 and −24 h respectively.

long and wide feature over Minnesota, South Dakota and Nebraska in the reanalysis–based FTLE field which is absent in the mean ensemble forecast. One should note that the overall similarity between Fig. 3.2(a) and Fig. 3.10(a) in common windows is due to the fact that both FTLE fields correspond to the same time, meanwhile the differences are mainly due to the spatiotemporal resolution of the velocity fields and the fact that the former is calculated from archive data and latter is generated from *pure* forecast data initialized by reanalysis data set at 00:00 UTC 28 Sep 2010.

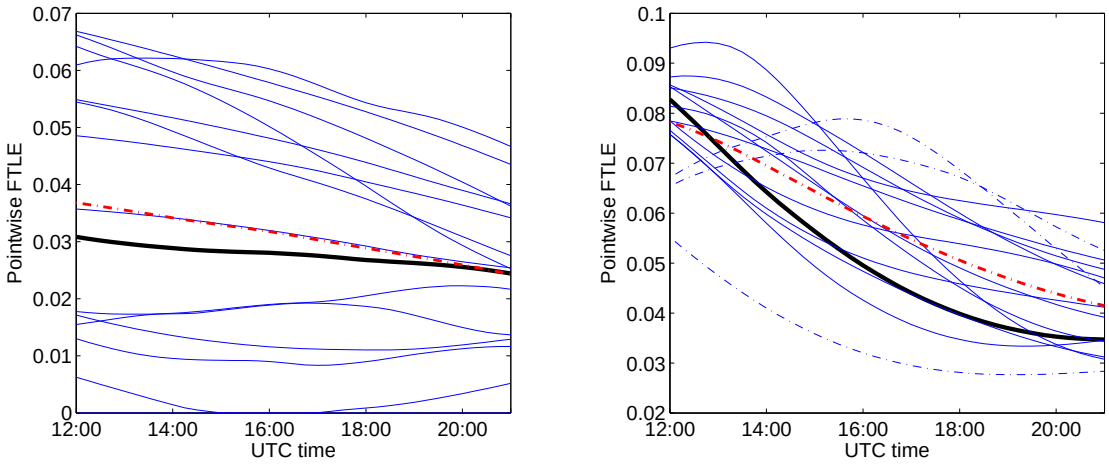
To study the uncertainty of the FTLE fields we use the pointwise FTLE value vs. time at a fixed location. The local maxima of this time series could be regarded as representative of candidate hyperbolic LCS features that might pass over the specified location

(BozorgMagham et al., 2013). The uncertainty of the pointwise results represents the uncertainty of presence and timing of strong LCS features passing over a location of interest. As an example, Fig. 3.11(a) shows the pointwise FTLE value at $(48°22'N, -68°10'48''W)$ equal to (1000, 1400) km with respect to the reference point (Virginia Tech Kentland Farm) for the previous forecast interrogation window of 12:00 to 21:00 UTC 29 Sep 2010 where the first frames of the FTLE field for the 15 members are shown in Fig. 3.9. We observe that the results from ensemble members are significantly different during the interrogation time window. This figure shows high uncertainty of the forecast FTLE value over that location when different forecast velocity fields are used to calculate the FTLE–LCS features. Fig. 3.11(b) shows another case of pointwise FTLE vs. time at $(30°27'N, -85°3'W)$ equal to (-430, -750) km with respect to the same reference point and during the same time interval. In this figure we observe that, except for three members that are shown by dashed line, the other ones follow a similar pattern and their differences are small, meaning that forecasts at that location are fairly consistent during the interrogation time interval.

In addition to uncertainty quantification, this observation could lead to the concept of probabilistic forecasting of the LCS features. To obtain reliable forecast of LCS features, we will assume that one needs to have enough ensemble members that show maxima in a common time interval. Referring to Fig. 3.12 as a schematic of forecast ensembles, we define the probability of the passage of an LCS feature over a region during a defined time interval as,

$$P_{\left(t-\frac{\Delta t}{2}, t+\frac{\Delta t}{2}\right)} = n_p / N_{ens} \tag{3.12}$$

where n_p is the number of maxima beyond a pre-determined threshold (for more information about that threshold refer to BozorgMagham et al. (2013)) in the time interval $\left(t-\frac{\Delta t}{2}, t+\frac{\Delta t}{2}\right)$ and N_{end} is the total number of ensemble members providing a unique peak for each ensemble member in that time interval.



(a) (b)

Figure 3.11: Ensemble of pointwise FTLEs vs. time. Black line represents the control case, dashed red line shows the average of ensembles and the other 14 members are from each FTLE calculation, (a) highly uncertain forecast, FTLE measurement at $(48°22'N, -68°10'48''W)$ (b) fairly consistent forecast, FTLE measurement at $(30°27'N, -85°3'W)$. These two points correspond to (1000, 1400) and (-450,-750) km with respect to the reference point.

Figure 3.13(a) shows the ensemble forecast members for 29 Sep 2010 at $(30°27'N, -85°3'W)$. Since the spatiotemporal resolution of ensemble GFS data is coarse (2.5°, 12 h) the ensemble members of the pointwise FTLEs do not show sharp peaks, however this figure shows the main concept of probabilistic FTLE–LCS forecasting. Figure 3.13(b) shows the corresponding probability distribution of having a candidate LCS during that time interval when $\Delta t =$ 2 h. This figure shows that the probability of an LCS passage is maximum between 10:00 and 12:00 UTC. We should note that to generate Fig. 3.13(a) we use the most updated reanalysis data as the initial condition for each forecast data frame. Thus, we use reanalysis data corresponding to 00:00 UTC 28 and 29 Sep 2010 as the initial conditions for the forecast velocity fields. Note that Fig. 3.11(b) and Fig. 3.13(a) are slightly different in their common window (after 12:00 UTC) because in the former one we just use the 00:00 UTC 28 as the initial condition while in the latter one we use the most updated initial conditions.

To test this approach for more days and to know more about statistical distribution of

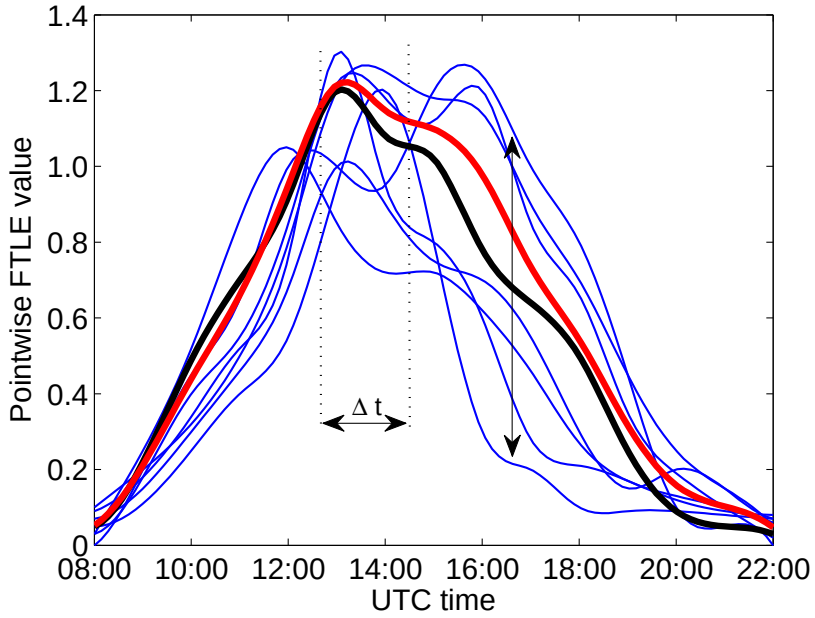
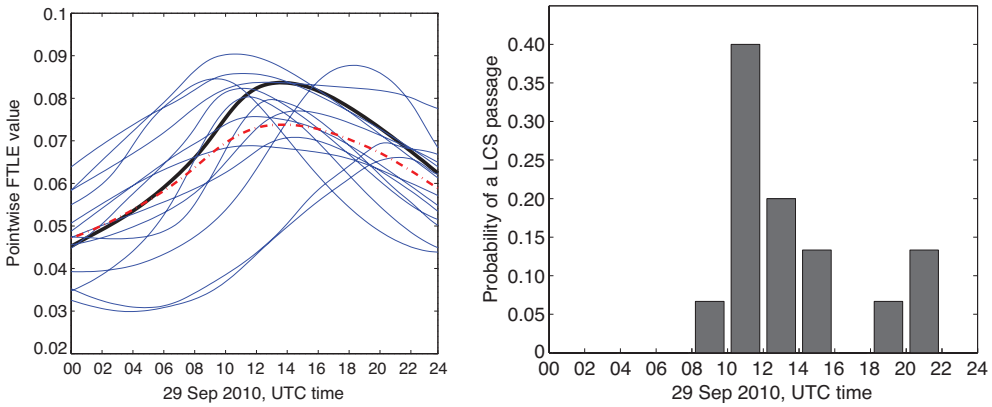


Figure 3.12: Schematic of ensembles pointwise forecasts FTLE value at a specified location. Black: control case, red: mean of the ensembles, blue: other ensemble members. Number of maxima in time interval Δt is proportional to the probability of passage of a LCS over the specified location.



(a) (b)

Figure 3.13: (a) Ensembles of pointwise forecasts FTLE field value at a $(30°27'N, -85°3'W)$, integration time: -24 hrs, red line: mean curve, and black line: control case (b) probability of a LCS passage during 2 h windows.

errors we use GFS ensemble forecast data to calculate the forecast FTLE–LCS features (the same procedure as described for Fig. 3.13), and then we compare these results with reanalysis based FTLE–LCSs which are obtained from the NCEP North American Regional Reanalysis (NARR). Note that spatiotemporal resolution of NARR data is 0.3 degrees and 3 hours. For both cases of forecasts and reanalysis results we consider a time interval of 24 h and an integration time for FTLE calculations of -24 h (backward FTLE). Results in Table 3.1 correspond to four weeks, one in each season of 2012. The selected weeks are randomly chosen. We report all days of each week that major LCS features have been observed over the Virginia Tech Kentland Farm.

We use a time resolution of 15 minutes for the FTLE–LCS calculations, so for each day we have 97 time slices. For each day all 15 members of the GFS based forecast FTLE–LCS are calculated (the total number of frames per day is $1455 = 97$ slices times the 15 members) and then the probability distribution of having a LCS over the reference point is obtained. The result of this process is similar to Figure 3.13(b) for each time slice. Table 3.1 contains data for 15 days, so it is the abstract of comparing 21825 ensemble forecast frames (1455 frames per day times 15 days) with 1455 reanalysis based frames of FTLE–LCSs. We divide each day into 2 hours intervals. The probability of LCS passage over the reference point is calculated as introduced in (3.12). Table 3.1 shows the corresponding probabilities as factors of $(\frac{1}{15})$ since the number of ensemble members is 15, so the numerator shows the number of ensemble peaks in each time interval. Note that in some cases, such as Sep 20 and Nov 8, the summation of peaks is less than 15 which means that some member of the ensemble does not have any maximum during the 24 hour interrogation window or the peak is very flat. The passage time of the reanalysis (NARR) based LCSs over the reference point for each day is shown by "$*$" in each column.

Results of this comparison show that in 9 out of 15 cases, the probabilistic forecast passage

	April				June			September				November			
UTC	22	23	25	27	4	6	7	16	17	20	22	5	7	8	9
00:00-02:00	0	$\frac{10}{15}^*$	0	$\frac{1}{15}$	0	$\frac{1}{15}$	0	0	0	0	0	0	$\frac{1}{15}$	0	0
02:00-04:00	0	$\frac{2}{15}$	0	$\frac{7}{15}^*$	0	$\frac{1}{15}$	0	$\frac{2}{15}$	0	0	0	0	$\frac{6}{15}^*$	0	0
04:00-06:00	0	$\frac{3}{15}$	$\frac{7}{15}$	$\frac{4}{15}^*$	0	$\frac{9}{15}^*$	0	$\frac{11}{15}^*$	0	0	0	0	$\frac{8}{15}$	0	0
06:00-08:00	$\frac{1}{15}$	0	$\frac{6}{15}^*$	$\frac{3}{15}^*$	0	$\frac{2}{15}$	0	0^*	$\frac{1}{15}$	0	0	$\frac{1}{15}$	0	0	0
08:00-10:00	$\frac{2}{15}$	0	$\frac{2}{15}$	0	0	$\frac{2}{15}$	0	$\frac{2}{15}$	$\frac{14}{15}^*$	0	0	$\frac{1}{15}$	0	0	0
10:00-12:00	$\frac{8}{15}$	0	0	0	0	0	0	0	0	0	0	$\frac{1}{15}$	0	0	0
12:00-14:00	$\frac{3}{15}^*$	0	0	0	0	0	0	0	0	0	0	$\frac{10}{15}^*$	0	0	0
14:00-16:00	0^*	0	0	0	$\frac{1}{15}$	0	$\frac{2}{15}$	0	0	0	0	$\frac{2}{15}$	0	0	$\frac{2}{15}$
16:00-18:00	$\frac{1}{15}$	0	0	0	$\frac{12}{15}^*$	0	0^*	0	0	0	$\frac{1}{15}$	0	0	$\frac{1}{15}^*$	$\frac{6}{15}$
18:00-20:00	0	0	0	0	$\frac{2}{15}$	0	$\frac{6}{15}$	0	0	$\frac{1}{15}$	$\frac{2}{15}$	0	0	0	$\frac{3}{15}^*$
20:00-22:00	0	0	0	0	0	0	$\frac{7}{15}$	0	0	$\frac{6}{15}^*$	$\frac{11}{15}^*$	0	0	$\frac{11}{15}^*$	$\frac{2}{15}$
22:00-24:00	0	0	0	0	0	0	0	0	0	$\frac{7}{15}$	$\frac{1}{15}$	0	0	$\frac{1}{15}$	$\frac{2}{15}$

Table 3.1: Four randomly selected weeks, one in each season of 2012. Days that major LCS features observed are reported. Each entity represents the probability of passing LCS features over the reference point (Virginia Tech Kentland farm) during the specified time interval. These values are calculated based on 15 ensemble members of GFS forecast data set. "$*$" represent the passage time of LCS features over the reference point which is calculated by using the reanalysis NARR data set. In some days we observe more than one LCS over the reference point. Note that in some cases the summation of numerators in each column is less than 15 which means that some of the ensemble members do not have any peak during 24 hours of the interrogation window.

times are in the same time 2–hour interval as the reanalysis based results, in 5 cases they just differ one time interval (maximum error 4 hours) and in 1 case the difference is equal or more than 2 time intervals. Also in 4 days we have more than one LCS passage based on reanalysis velocity field where for 3 days, the ensemble forecasting captures at least one of them. A summary of this comparison is represented in Fig. 3.14. For days with more than one LCS occurrences, we consider the closest one to the forecast interval result with highest probability. Table 3.1 and Fig. 3.14 help us to have quantitative measure of the distribution and reliability of the forecast results which is not available with single deterministic forecast outputs. As an example we observe that for about 60% of the cases the error of ensemble forecasting with respect to the reanalysis result is within 2 hours.

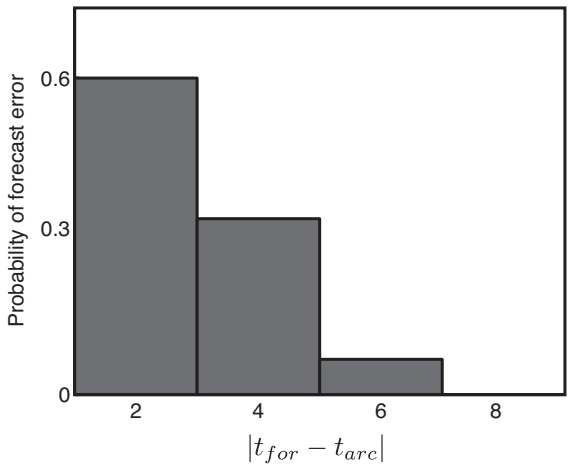


Figure 3.14: Summary of Table 3.1. A histogram of the absolute value of the difference between forecast (t_{for}) and archive (t_{arc}) LCS passage times. For days with more than one LCS occurrence, we consider the one closest to the forecast interval result with highest probability.

3.4 Discussion

The aim of this study was to consider two issues related to prediction of atmospheric FTLE fields. The first concern was the effect of unresolved turbulence on the resultant FTLE fields. By using the Monte Carlo method and adding a stochastic turbulent velocity term, we introduced the notion of stochastic FTLE (SFTLE) field. A significant difference between the deterministic solution and the centroid of the (in general, non–Gaussian) probability distributions was observed. The role of FTLE-LCSs in shaping the probability distributions was discussed.

The second concern originated from previous observations about comparisons between archive– and forecast–based FTLE fields (BozorgMagham et al., 2013). These comparisons show that the major source of errors in forecast FTLE fields is errors in the forecast velocity field which is unavoidable due to the chaotic dynamics of atmosphere. Thus, a major concern is the level of confidence regarding the forecast results. We introduced an ensemble method for FTLE–

LCSs for quantitative study of uncertainty and probabilistic forecasting, using relatively low spatiotemporal resolution GFS data. We also introduced a method for probabilistic forecasting of LCS features, particularly the passage time of an LCS over a particular geographic point, which was previously found to be associated with aeroecological events (Tallapragada et al., 2011). We found that one can predict the passage time with an accuracy of 2 hours 60% of the time which lends confidence to the practical application of LCS passage prediction for forecasting movement of biological and chemical fronts.

Subgrid scale stochastic effects and uncertainty will be important in real applications of transport phenomena such as the monitoring of volcanic ash and large scale dust storms or aeroecology of microorganisms related to human health and agriculture (topics which can be interrelated; Kellogg and Griffin (2006)). For the latter, given the association of FTLE fields and associated LCS features to moving fronts of microorganisms (Tallapragada et al., 2011), there is a possibility of predicting significant changes in the population structure of airborne pathogens, or finding constraints on their dispersal range, with implications for improved management decision making by stakeholders.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant Number CMMI–1100263 (Dynamical Mechanisms Influencing the Population Structure of Airborne Pathogens: Theory and Observations).

Chapter 4

Small fluctuations in the recovery of fusaria across consecutive sampling intervals with unmanned aircraft 100 m above ground level

Note: The following chapter was formatted to facilitate publication in Aerobiologia. This work was originally published by Lin, Bozorg Magham, Ross, and Schmale in Volume 29, Issue 1 (March 2013) pp.45-54. (DOI 10.1007/s10453-012-9261-3) of Aerobiologia.

Abstract

Many *Fusarium* spp. are well-suited for atmospheric dispersal, yet the aerobiology of these fungi are poorly understood. Recent work has highlighted the role of Lagrangian coherent

structures (LCSs) in the movement of fusaria in the lower atmosphere. Here, we extend this work by examining the relationship between the length of atmospheric sampling intervals with autonomous unmanned aerial vehicles (UAVs) and the recovery of fusaria (Lin, 2013). UAVs were equipped with an array of eight microbe-sampling devices with four inner sampling arms and four outer sampling arms. Each set of arms was used to collect consecutive aerobiological samples for periods of 10 minutes at 100 m above ground level at Kentland Farm in Blacksburg, Virginia. Fifty-one flights (102 consecutive sampling intervals) were conducted in 2010 and 2011. A correlation analysis showed that counts of fusaria did not vary between the inner and outer sampling arms from consecutive sampling periods of 10 min ($r = 0.93, P < 0.001$), and the frequency of colony counts had similar distributions for samples from the inner and outer sampling arms. An analysis of the temporal variation in collections of *Fusarium* showed that the similarity between collections decreased over time. This work supports the idea that atmospheric populations of fusaria are well-mixed, and large changes in the recovery of fusaria in the lower atmosphere may be attributed to large-scale phenomena (e.g., LCSs) operating across varying temporal and spatial scales. This work may contribute to effective control measures for diseases causes by fusaria in the future.

4.1 Introduction

Fusarium is one of the most important genera of fungi on Earth (Leslie and Summerell, 2006). Members of this genus cause a number of devastating plant diseases, and can threaten the health of both domestic animals and humans through the production of mycotoxins (Berek et al., 2001; McMullen et al., 1997). Many fusaria are transported through the atmosphere from one habitat to another (Schmale et al., 2012; Tallapragada et al., 2011).

Previous work has shown that large scale atmospheric features known as Lagrangian coherent structures (LCSs) or atmospheric transport barriers (ATBs) are associated with the long distance transport of *Fusarium* in the lower atmosphere (Schmale et al., 2012; Tallapragada et al., 2011). ATBs are moving boundaries that effectively separate air masses of qualitatively different dynamics and may play a significant role in the movement of microbes among habitats (Senatore and Ross, 2011). Tallapragada et al. (2011) showed that LCSs (ATBs) were associated with changes in atmospheric counts of *Fusarium*. Though the work by Tallapragada et al. (2011) was the first to demonstrate that large fluctuations in atmospheric counts of *Fusarium* could be attributed to the passage of ATBs, it was unable to account for small-timescale fluctuations that might explain natural fluctuations among collections of *Fusarium*.

Recently, members of our research team have developed technologies with autonomous unmanned aerial vehicles (UAVs) to track the movement and structure of populations of microbes such as *Fusarium* in the lower atmosphere (Schmale Iii et al., 2008). The UAVs were equipped with microbe-sampling devices that contained a total of four Petri plates that were opened and closed by remote control from the ground once the UAV was aloft (Schmale Iii et al., 2008). In the present study, we used a new array of sampling devices that contained a total of eight Petri plates, with four *inner* sampling arms and four *outer* sampling arms that were used to collect consecutive aerobiological samples for periods of 10 minutes at 100 m above ground level. This method was used to test the null hypothesis that the recovery of fusaria would not vary across consecutive (a 10 min sample on the inner arms, immediately followed by a separate 10 min sample on the outer arms) aerobiological sampling intervals with UAVs 100 m above ground level. Thus, large fluctuations in the recovery of fusaria could be attributed to a suite of factors including the passage of LCSs (ATBs) and/or the contribution of local sources, and not random fluctuations in counts of *Fusarium* that would

be representative of a *natural* condition. The specific objective of this study was to determine if collections of fusaria vary between the inner and outer sampling arms of a UAV from consecutive sampling periods of 10 min. This work is prerequisite for understanding if changes in the recovery of fusaria in the lower atmosphere may be attributed to large-scale phenomena (e.g., LCSs) operating across varying temporal and spatial scales and may contribute to effective control measures for diseases causes by fusaria in the future.

4.2 Materials and methods

4.2.1 Autonomous unmanned aerial vehicles (UAVs) for sampling

Autonomous (self-controlling) UAVs were used to collect *Fusarium* from the atmosphere above Virginia Tech's Kentland Farm in Blacksburg, VA, USA. The UAVs consisted of a Sig Rascal © airframe equipped with an autopilot computer and a suite of onboard telemetry devices (Schmale Iii et al., 2008) and were programmed to fly a circular sampling pattern at a target altitude of 100 m above ground level and a nearly constant speed of 90 km/h. Each UAV carried eight collection plates containing a *Fusarium* selective medium on the wings. The eight sampling plates were separated into inner and outer sampling arms (4.1).

For consecutive sampling flights, a 10 min sample was collected using the inner arms (4 plates were exposed during this sampling interval), immediately followed by a separate 10 min sample using the outer arms (4 plates exposed were exposed during this sampling interval) (4.1). Sampling flights were also conducted with the inner and outer sampling devices open at the same time (8 plates exposed during the sampling interval).

4.2.2 Culturing and identification of *Fusarium*

A *Fusarium* selective medium (FSM) (Schmale III et al., 2006) was used to bias our atmospheric collections for fungi in the genus *Fusarium*. Immediately following a sampling flight, the exposed plates were removed from the UAV and placed in small plastic containers for transport to the laboratory. The plates were incubated for 5 to 7 days at room temperature to allow white, fuzzy colonies of *Fusarium* to develop. Colonies of *Fusarium* were counted and subcultured to plates of 1/4 strength potato dextrose agar (PDA) medium for further identification.

4.2.3 Statistical analyses

We hypothesized that the recovery of fusaria would typically not vary significantly across consecutive aerobiological sampling intervals of short duration (a 10 min sample on the inner arms, immediately followed by a separate 10 min sample on the outer arms). If we fail to reject this hypothesis, then large fluctuations in the recovery of fusaria over short to intermediate timescales could be attributed to a suite of factors, such as the passage of LCSs (ATBs) and/or the contribution of a strong local source. By *short* timescale, we mean short compared with the Lagrangian timescale, discussed below. It is also important to note that since the atmosphere is moving, short timescales are also related to short spatial scales. To test our hypothesis, colony counts of *Fusarium* obtained from different flights were assembled to perform statistical analyses (Tables 4.1 and 4.2). For flights with simultaneous inner and outer arm sampling (Tables 4.1), we estimated the variability in sampled colony counts of Fusarium, yielding an estimate of error for colony counts. For flights with consecutive inner and outer arm sampling (plates exposed during consecutive sampling periods of 10 minutes) (Table 4.2), a simple linear regression was used to determine the relationship between colony

counts of *Fusarium* collected for the inner and outer sampling arms. A scatter plot and a frequency plot were also used to show this relationship. Statistical analyses were performed using JMP 4.0. The correlation between colony counts from the inner and outer sampling arms was also explored as a function of time-lag between sampling intervals (i.e., comparisons of colony counts between consecutive flights separated by 10 minutes, and between other flights separated by longer periods throughout a sampling day).

Flights	Date	Time open (in & out)	Time closed (in & out)	Counts in	Counts out	Time sampling (min)
F137	10-Mar-10	0929	0944	3	1	15
F138	10-Mar-10	1035	1042	1	8	7
F139	10-Mar-10	1130	1145	7	4	15
F140	10-Mar-10	1300	1315	5	2	15
F141	10-Mar-10	1400	1415	4	9	15
F142	11-Mar-10	1005	1020	3	3	15
F143	15-Jul-10	0950	1005	23	16	15
F144	15-Jul-10	1155	1210	35	37	15
F145	16-Jul-10	0925	0940	20	26	15
F146	16-Jul-10	1045	1100	27	23	15
F147	28-Sep-10	0959	1014	2	3	15
F148	28-Sep-10	1118	1133	6	2	15
F149	28-Sep-10	1412	1427	17	8	15
F150	28-Sep-10	1532	1540	9	6	8
F151	29-Sep-10	0915	0926	3	7	11
F152	29-Sep-10	1029	1044	4	1	15
F153	29-Sep-10	1323	1338	7	6	15
F154	01-Oct-10	0908	0923	8	3	15
F155	01-Oct-10	1203	1218	9	8	15
F156	01-Oct-10	1428	1443	31	21	15
F157	01-Oct-10	1700	1708	10	6	8

Table 4.1: Colony counts of *Fusarium* from simultaneous sampling (inner arms and outer arms were opened at the same time) with UAVs 100 m above ground level at Virginia Tech's Kentland Farm

4.3 Results

4.3.1 Simultaneous sampling with eight plates

In order to compare samples collected from inner and outer arms during different time periods, it is essential to show that samples do not vary significantly between inner and

outer arms during the same time period. In other words, we must examine the potential role (if any) that plate position on the UAV has on the recovery of fusaria. To do this, we conducted 21 simultaneous sampling flights in which all eight sampling devices (inner and outer sampling arms opened at the same time) were exposed during the same sampling interval (Tables 4.1). For these 21 flights, 433 colonies were recovered across all 21 sampling intervals; 234 colonies were collected across the inner arms, and 199 colonies were collected across the outer arms. Results of our correlation analysis for this sampling method showed that counts from the plates on the inner arms were positively correlated with counts from plates on the outer arms ($r = 0.89, P < 0.001, n = 21$). Thus, the location of the plates (inner versus outer sampling arms) did not impact the collection of fusaria.

We also used samples from this method to estimate the variation in colony counts for the inner and outer sampling arms. Considering only the 15 minute samples of Tables 4.1, we calculated the total colony count, c, and the magnitude of the difference between the inner and outer plates, which is the variation in colony counts, δc. In Fig. 4.2, we plot the fractional variation, $\delta c/c$, versus c, and notice a trend. The curve corresponds to $1/\sqrt{c}$, the fractional standard deviation for a Poisson distribution. Thus, the probability of a viable airborne *Fusarium* spore impacting the samplers is well approximated as an inhomogenous Poisson process with an arrival rate varying on a timescale long compared to the sampling duration. Thus, when a colony count, c, is obtained, the margin of error can be approximated as $\pm\sqrt{c}$.

4.3.2 Consecutive sampling with four inner plates and four outer plates

To determine if collections of fusaria varied between consecutive sampling periods of 10 min, we conducted 102 consecutive sampling intervals (51 flights) 100 m above ground level in

2010 and 2011 (Table 4.2). Five hundred and ninety two colonies were recovered across all 102 sampling intervals; 275 colonies were collected across the inner arms, and 312 colonies were collected across the outer arms (excluding flights F207 and F208, which were clear outliers in the dataset). A significant positive correlation was observed for colony counts of *Fusarium* between the inner plates and the outer plates ($r = 0.93, P < 0.001, n = 98$). Scatter plots and a simple linear regression of consecutive sampling intervals are shown in Fig. 4.3.

Two flights however, F207 and F208 (Table 4.2), were excluded from the scatter plot, since these flights were considered outliers and are the subject of additional discussion below. It should still be noted, however, that when these flights were included in the correlation analysis, a significant positive correlation was still observed ($r = 0.38, P < 0.01, n = 102$). A frequency plot showed that the distribution of colony counts was similar for the inner and outer sampling arms over the range of colony counts (Fig. 4.4).

The correlation between colony counts from the inner and outer sampling arms was also explored as a function of time-lag, τ, between sampling intervals. This approach allowed us to examine the temporal variation of colony counts. From Table 4.2 (excluding F207 and F208), pairwise comparisons of colony counts were determined for five time intervals: 0.17 h (n = 49 pairs), 0.5 h to 1.5 h (n = 83 pairs), 1.5 h to 3 h (n = 94 pairs), 3 h to 6 h (n = 84 pairs), and 6 h to 9 h (n = 16 pairs). An autocorrelation coefficient was determined as follows (autocorrelation coefficient for zero time-lag is defined as 1)

$$R(\tau) = \frac{E\left[\left(c_t - \mu\right)\left(c_{t+\tau} - \mu\right)\right]}{\sigma^2} \tag{4.1}$$

where c_t is the colony count at time t, $c_{t+\tau}$ is the colony count at time $t+\tau$, μ represents the mean value, σ represents the standard deviation of the colony counts, and $E\left[\bullet\right]$ represents

expectation value. The results of this analysis are shown in Fig 4.5; the similarity between collections of *Fusarium* decreases over time. Error bars were calculated using results from section § 4.3.1 with the assumption that colony counts are Poisson distributed. A Gaussian distribution weighted method was used to obtain the error bars with sufficient number of simulations converging to a constant limit.

4.3.3 Anomalous punctuated changes in colony counts for two consecutive flights

In flights F207 and F208, we observed a significant departure from the usual 10-minute time-lag correlation. Flight 207 started sampling at 12:00 PM on 25 October, 2011. Eighty colonies were recovered from the inner arms, but only 31 colonies were recovered from the outer arms. Flight 208 started sampling at 1:16 PM on the same day. Twelve colonies were recovered from the inner arms, but 120 colonies were recovered from the outer arms. The inner and outer samples from these flights were not correlated, and cannot be explained by the statistics of a slowly varying inhomogeneous Poisson process. We view these two flights as anomalies that are in need of further explanation. We hypothesized that an ATB could have contributed to the observed changes in colony counts (e.g., Tallapragada et al. (2011)), but archived weather-based computations did not reveal the presence of any strong LCSs (data not shown). Furthermore, HYSPLIT back trajectories for these samples suggested that all of these samples originated from a similar location in West Virginia (within the scale of accuracy of the computations, on the order of 10-100km) (Fig. 4.6).

4.4 Discussion

Recent work has highlighted the role of Lagrangian coherent structures (LCSs) in the movement of fusaria in the lower atmosphere (Schmale et al., 2012; Tallapragada et al., 2011). Here, we extend this work by examining the relationship between the length of atmospheric sampling intervals with autonomous unmanned aerial vehicles (UAVs) and the recovery of fusaria. UAVs were equipped with an array of eight microbe-sampling devices with four inner sampling arms and four outer sampling arms. Each set of arms was used to collect consecutive aerobiological samples for periods of 10 minutes at 100 m above ground level at Kentland Farm in Blacksburg, Virginia. A total of 102 consecutive sampling intervals (51 flights) was conducted in 2010 and 2011. Results showed that counts of fusaria did not vary across consecutive aerobiological sampling intervals. This work supports the idea that atmospheric populations of fusaria are well-mixed, and large changes in the recovery of fusaria in the lower atmosphere may be attributed to large-scale phenomena (e.g., LCSs) operating across varying temporal and spatial scales.

Counts of *Fusarium* were not significantly different between plates located on inner and outer sampling arms in which all eight sampling devices were exposed during the same sampling interval. Thus, collections of *Fusarium* with UAVs were not influenced by the position of the plates. Random collections of *Fusarium* across all of the sampling surfaces is consistent with the idea that atmospheric populations of *Fusarium* are well-mixed (Schmale III et al., 2006). The fractional variation in colony counts revealed that the statistical distribution of colony counts across the inner and outer sampling arms is well approximated by a slowly varying inhomogeneous Poisson process. Colony counts from consecutive sampling intervals separated by 10 minutes did not vary significantly, but the correlation drops to nearly zero

for flights separated by 9 hours. The Lagrangian (autocorrelation) time-scale

$$T_L = \int_0^\infty R(\tau)\, d\tau \tag{4.2}$$

is approximately 3 hours, which is on the order ($\sim 10000s$) estimated for velocity autocorrelations in atmospheric turbulence (Gifford, 1987), and is also the Lagrangian timescale for layer (stratus) clouds. The time T_L provides the timescale for the variation of the arrival rate for the slowly varying inhomogeneous Poisson process assumption, and we note that this is indeed long compared with the sampling duration (10 minutes), further justifying the Poisson assumption. For purely stochastic motion, the autocorrelation is an exponential, $R(\tau) = exp(-\tau/T_L)$ (Csanady, 1973; Dosio et al., 2005). With horizontal winds on the scale of 2-10 m/s, this timescale suggests that there are coherent *clouds* of *Fusarium* with horizontal dimensions on the scale of 20-100 km. This idea is consistent with the observations of Tallapragada et al. (2011) based on mesoscale atmospheric simulations, who found that the typical size for a coherent air mass was on a similar scale (50-150 km), based on the average passage of an LCS over the sampling location every 5 to 7 hours.

Tallapragada et al. (2011) showed that LCSs (atmospheric transport barriers or ATBs) were correlated with changes in atmospheric counts of *Fusarium*. Schmale et al. (2012) suggested that LCSs were likely to influence the population structure of F. graminearum. With the exception of flights 207 and 208, we did not observe any significant variation in colony counts among consecutive sampling flights. Thus, we are now able to exclude fluctuations over short periods of time as potential contributors to changes in the atmospheric counts of Fusarium. Thus, large changes in populations of *Fusarium* in the lower atmosphere may be attributed in part to large-scale phenomena (such as LCSs) or strong local sources operating across varying temporal and spatial scales.

The inner and outer samples from flights F207 and F208 were not correlated, and represented a significant departure from the usual 10-minute time-lag correlation. Archived weather-based computations did not reveal the presence of any strong LCSs, and HYSPLIT back trajectories for these samples suggested that they both originated from a similar same location in West Virginia (Fig. 4.6). Less than 90 minutes separated the start of F207 and the end of F208, but there was significant variation over a timescale of 10 minutes during both flights. This suggests a patchy (heterogeneous) distribution of *Fusarium* in the atmosphere (Okubo and Levin, 2001). With the horizontal wind speeds at that time being approximately 2 m/s at ground level, the patchiness spatial scale was at most 1 km over a cloud on the order of 10-20 km (bracketed by the more typical flights F206 and F209). It is possible that these high values were triggered by a local, and possibly temporally non-uniform, source. Future work aimed at identifying the species recovered in these flight populations may provide clues about the approximate origin and mixing of these populations.

Future work by our research team aims to examine meteorological events that might provide signatures for the life history of populations of *Fusarium* in the lower atmosphere. Such work may contribute to an increased understanding of the spread of plant diseases in the future (Aylor, 2003).

Acknowledgments

We thank John Cianchetti for his excellent technical assistance with the construction, maintenance, and operation of the UAVs described in this work. We also thank Zolton Bair for his help with fungal cultures and Pavlos Vlachos for helpful discussions. This material is based upon work supported by the National Science Foundation under Grant Numbers DEB-0919088 (Atmospheric transport barriers and the biological invasion of toxigenic fungi

in the genus Fusarium) and CMMI-1100263 (Dynamical mechanisms influencing the population structure of airborne pathogens: Theory and observations). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

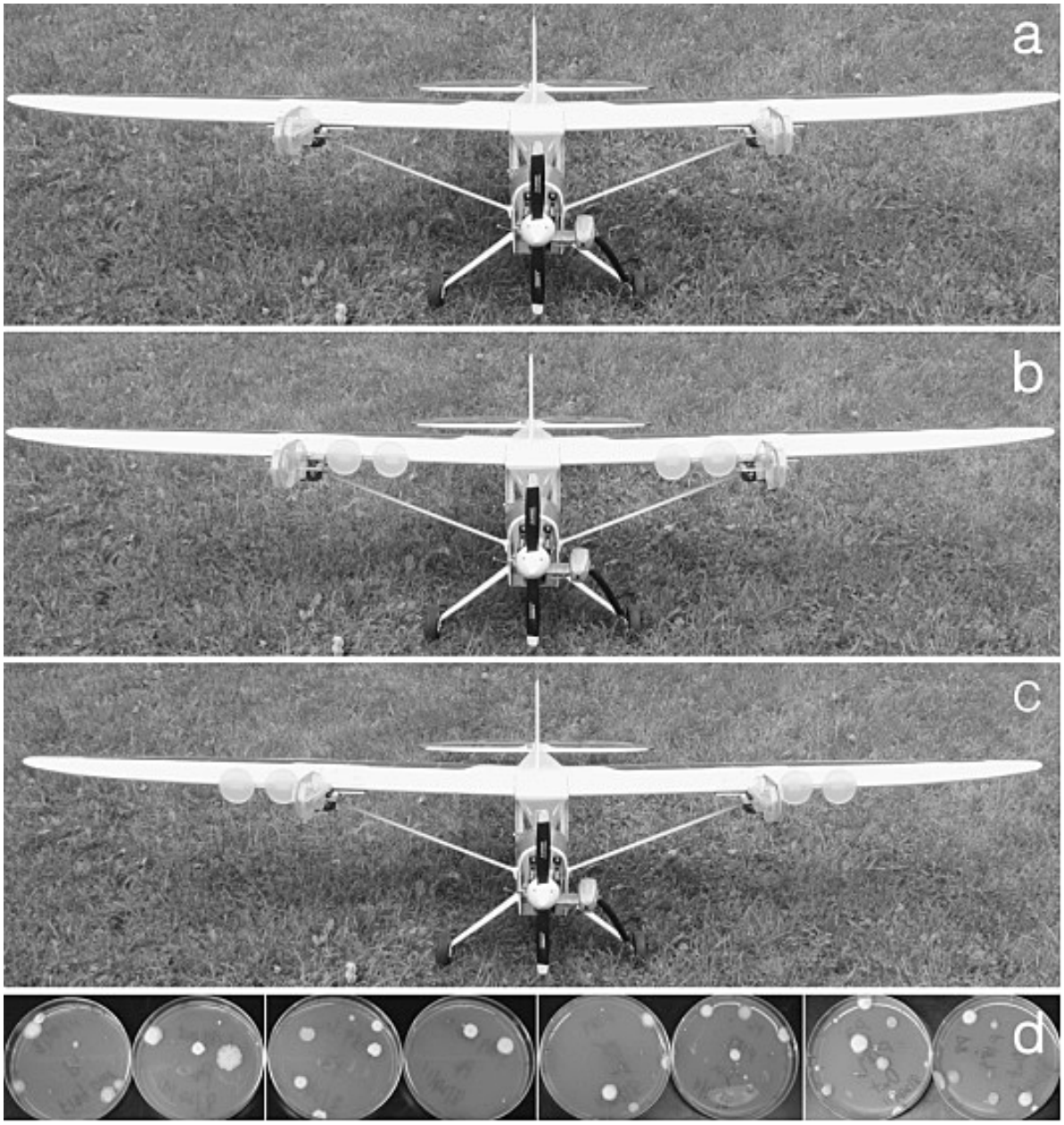


Figure 4.1: An autonomous unmanned aerial vehicle (UAV) equipped with an array of eight microbe-sampling devices with four inner sampling arms and four outer sampling arms. Each arm carries two Petri plates containing a *Fusarium* selective medium. During takeoff and landing, the sampling devices are closed (a). After reaching the target altitude of 100 m, the inner sampling arms are opened for 10 min (b). These inner arms are closed, and the outer arms are opened for 10 min immediately following the first collection (c). Colonies of *Fusarium* are recovered in the laboratory and recorded for each of the plates (shown here from flight F189) (d). By author, 2013, Aerobiologia.

Flights	Date	Open (in)	Closed (in)	Open (out)	Closed (out)	Counts (in)	Counts (out)	Time sampling out (min)	Time sampling in (min)
F158	06-Apr-11	0913	0923	0923	0933	6	10	10	10
F159	06-Apr-11	1027	1037	1037	1046	14	12	10	9
F160	06-Apr-11	1557	1607	1607	1616	18	26	10	9
F161	07-Apr-11	0952	1002	1002	1012	27	33	10	10
F162	07-Apr-11	1344	1354	1354	1404	16	13	10	10
F163	07-Apr-11	1514	1524	1524	1534	15	11	10	10
F164	07-Apr-11	1611	1621	1621	1631	9	13	10	10
F167	08-Apr-11	1407	1422	1422	1435	5	9	15	13
F168	08-Apr-11	1510	1520	1520	1530	4	5	10	10
F169	11-Apr-11	0952	1002	1002	1012	22	21	10	10
F171	16-May-11	1449	1459	1459	1509	7	7	10	10
F173	18-May-11	1426	1436	1436	1446	2	1	10	10
F174	18-May-11	1600	1610	1610	1620	4	2	10	10
F175	19-May-11	0943	0953	0953	1003	4	5	10	10
F176	19-May-11	1045	1055	1055	1105	2	3	10	10
F177	19-May-11	1152	1202	1202	1212	1	2	10	10
F178	19-May-11	1312	1322	1322	1332	3	2	10	10
F179	19-May-11	1435	1445	1445	1455	6	3	10	10
F180	22-Aug-11	1000	1010	1010	1020	1	3	10	10
F181	22-Aug-11	1224	1234	1234	1244	3	2	10	10
F182	23-Aug-11	0942	0952	0952	1002	2	2	10	10
F183	23-Aug-11	1040	1050	1050	1100	2	2	10	10
F184	23-Aug-11	1149	1159	1159	1209	2	3	10	10
F185	23-Aug-11	1309	1319	1319	1329	4	5	10	10
F186	23-Aug-11	1424	1434	1434	1444	2	2	10	10
F187	23-Aug-11	1601	1611	1611	1621	4	5	10	10
F188	24-Aug-11	0915	0925	0925	0935	2	2	10	10
F189	24-Aug-11	1031	1041	1041	1051	4	5	10	10
F190	24-Aug-11	1159	1209	1209	1219	4	6	10	10
F191	24-Aug-11	1314	1324	1324	1334	4	5	10	10
F192	24-Aug-11	1426	1436	1436	1446	5	7	10	10
F193	25-Aug-11	0914	0924	0924	0934	1	2	10	10
F195	25-Aug-11	1200	1210	1210	1220	2	5	10	10
F197	26-Aug-11	1115	1125	1125	1135	3	2	10	10
F198	26-Aug-11	1313	1323	1323	1333	3	5	10	10
F199	26-Aug-11	1419	1429	1429	1439	7	12	10	10
F200	24-Oct-11	1018	1028	1028	1018	3	3	10	10
F201	24-Oct-11	1133	1143	1143	1153	4	5	10	10
F202	24-Oct-11	1303	1313	1313	1323	6	5	10	10
F203	24-Oct-11	1418	1428	1428	1438	5	4	10	10
F204	24-Oct-11	1530	1540	1540	1550	6	7	10	10
F205	25-Oct-11	0915	0925	0925	0935	1	2	10	10
F206	25-Oct-11	1031	1041	1041	1051	3	6	10	10
F207	25-Oct-11	1200	1210	1210	1220	80	31	10	10
F208	25-Oct-11	1316	1326	1326	1336	12	120	10	10
F210	25-Oct-11	1543	1553	1553	1603	8	6	10	10
F211	26-Oct-11	0956	1006	1006	1016	3	3	10	10
F212	26-Oct-11	1205	1215	1215	1225	1	2	10	10
F213	26-Oct-11	1316	1326	1326	1336	8	9	10	10
F214	26-Oct-11	1432	1442	1442	1452	4	4	10	10
F215	27-Oct-11	0858	0908	0908	0918	3	3	10	10

Table 4.2: Colony counts of *Fusarium* from consecutive (a 10 min sample on the inner arms, immediately followed by a separate 10 min sample on the outer arms) aerobiological sampling intervals with UAVs 100 m above ground level at Virginia Tech's Kentland Farm in 2010 and 2011.

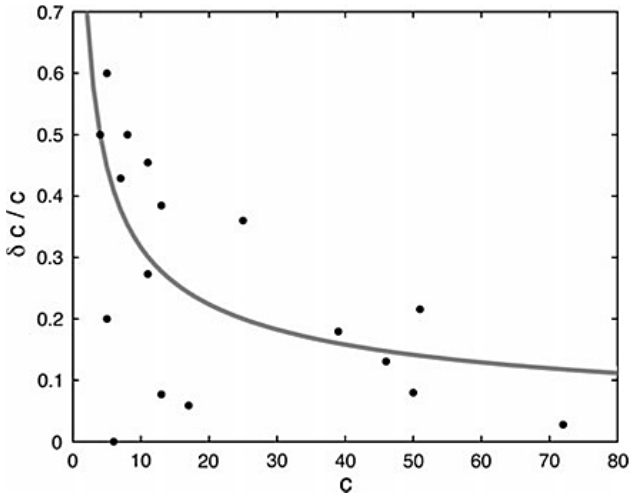


Figure 4.2: Fractional variation in sampled colony counts of *Fusarium* (c is the total colony counts from inner and outer plates, and δc is the variation in colony counts) based on simultaneous sampling using inner (4 plates) and outer (4 plates) arms of a UAV during 15 min sampling periods. Flights were conducted 100 m above ground level during 2010. The probability of a viable airborne *Fusarium* spore impacting the samplers is approximated as a slowly varying inhomogenous Poisson process

Figure 4.3: Scatter plot and simple linear regression of consecutive sampling of *Fusarium* with inner (4 plates) and outer (4 plates) arms of a UAV. A significant correlation ($r = 0.93, P < 0.001, n = 98$) was observed between colony counts of *Fusarium* from the inner and outer arms. Flights were conducted 100 m above ground level during 2010 and 2011. Flights F207 and F208 were outliers and were removed from the analysis

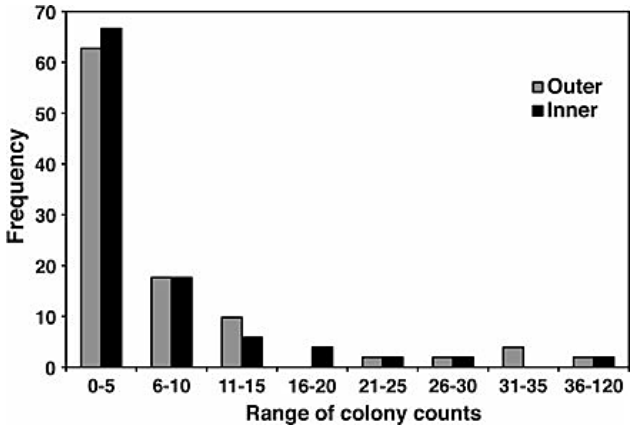


Figure 4.4: Frequency plot of colony counts of *Fusarium* from consecutive sampling periods of 10 mins. Flights were conducted 100 m above ground level during 2010 and 2011. The plot shows that the distribution of colony counts was similar for the inner and outer sampling arms over the range of colony counts

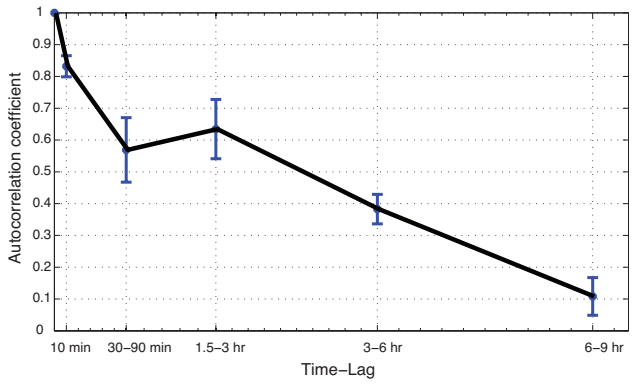


Figure 4.5: The colony count autocorrelation coefficient $R(\tau)$ versus the time-lag τ between sampling intervals. A high correlation is observed for a small time-lag. The similarity between collections decreases over time; typical behavior for Lagrangian trajectories of particles (spores) in atmospheric turbulence

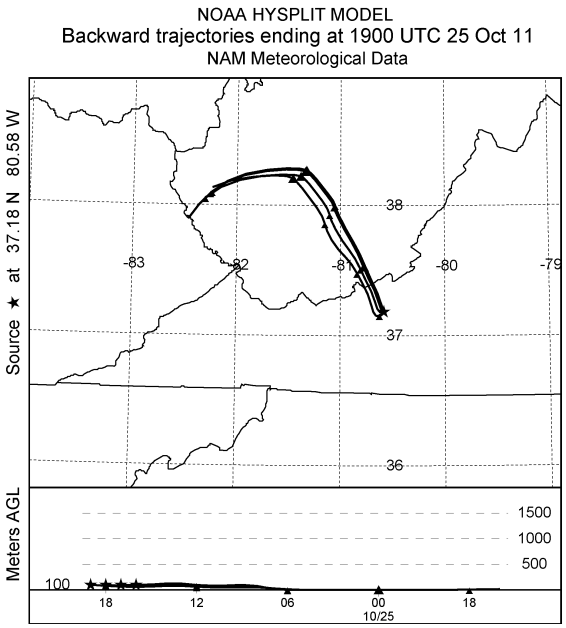


Figure 4.6: HYSPLIT backward trajectories for flights 207 and 208 on based on 25 Oct 2011. Trajectories were calculated hourly for 1600-1900 UTC and suggest that trajectories suggest that all of the samples originated from a similar location in West Virginia (within the scale of accuracy of the computations, on the order of 10-100km)

Chapter 5

Local finite time Lyapunov exponent, sampling strategy and probabilistic source regions

Abstract

We present a new interpretation about the local finite time Lyapunov exponent (FTLE). The suggested notion enables us to estimate the true local FTLE or the differential distances between source (destination) points of the sampled (released) particles when sampling (or release) in fluid field takes place at a fixed location and in a consecutive sequence. Results of this study help us to plan the localized geophysical samplings for maximal diversity monitoring purposes such that the collected particles come from the most possible separated source locations. Also, we may apply the proposed theorem to explain characteristic variations among the successive collected samples if they are caused by long range transport phenomena. In case of low resolution data sets, we consider the unresolved turbulence for identifying the probabilistic source (destination) regions. We emphasize on the

differences between numerical procedures for correct calculations of the probabilistic source and destination regions. Finally we show that separation of the probabilistic source regions corresponding to the sampled particles on either side of an attracting LCS happens even in presence of the unresolved turbulence.

5.1 Introduction

The classical interpretation of finite time Lyapunov exponent (FTLE) fields and associated hyperbolic Lagrangian coherent structures (LCSs) provides valuable information about transport and mixing of passive tracers in a flow domain (Haller and Poje, 1998; Haller and Yuan, 2000; Shadden et al., 2005; Haller, 2011). There are increasing number of studies which apply various concepts of FTLE–LCSs to describe and predict the time evolution of Lagrangian systems. For example in some of these studies large scale geophysical information such as wind or oceanic velocity fields are used as input data then the Lagrangian results such as oil spill or volcanic ash distribution are compared with true response of the system (Peng and Peterson, 2012; Olascoaga and Haller, 2012). In some applications there is little information about the system meanwhile there is a huge interest for knowing about the past or future of mesoscale Lagrangian configurations. For example, this study is motivated by a series of aerobiological researches of that nature where long range transport of microbial populations is studied. In those studies there are usually few localized (sometimes consecutive) measurements about microbial structure of atmosphere while knowing about the earlier or upcoming distribution of pathogens is anticipated. These outcomes are important for many practical means such as early warning and integrated risk management systems (Tallapragada et al., 2011; BozorgMagham et al., 2013; Lin et al., 2013).

In this study we connect the concept of FTLE and successive (aerial) particle samplings at

a fixed location by presenting a new interpretation about the local FTLE. This description is useful in cases such as sampling the microbial structure of atmospheric flow by equipped unmanned aerial vehicles (UAVs). This point of view would help us to have a better understanding about the distribution of the source points of sampled particles or the destination points of the released tracers. In addition, the suggested interpretation of the local FTLE is useful for planning the schedule of samplings at a fixed location for the aim of maximal diversity monitoring such that the collected particles come from the most possible separated source locations. Moreover, this analysis would help us to investigate the long range transport phenomena as a possible cause of abrupt characteristic changes among the successive collected (airborne microbial) samples.

Since this study is motivated by aerial measurements in realistic condition, we consider the spatiotemporal limitations of the available velocity field data. These limitations are manifested in unresolved turbulence and impose uncertainties on the location of the source and destination points. We use a Lagrangian particle dispersion model to determine the probabilistic source (destination) regions and we show how the notion of local FTLE could be used in cases that unresolved turbulence has considerable importance.

The paper is outlined as follows. In § 5.2 we study the relationship between local (point-wise) FTLE values and the dispersion of source (or destination) points in deterministic flow fields. Some applications of this notion is considered in the same section. In § 5.3 we consider the effects of unresolved turbulence and the stochastic velocity as an additive term and we investigate the uncertainty of the deterministic solutions for backward (source regions) and forward (destination regions) integration cases (Fay et al., 1995; Draxler and Hess, 1998; BozorgMagham and Ross, 2014).

5.2 Local finite time Lyapunov exponent

In this section we consider a new interpretation about the local FTLE for time–varying vector fields which is conceptually an extended notion of the local Lyapunov exponent in ODE systems (Oseledec, 1968; Abarbanel et al., 1992). By local FTLE we mean the probed time–series of the FTLE value at an arbitrary point in the field.

Classically, the time–varying FTLE field shows the maximum separation rate between nearby particles when they are released in the flow field at the *same* time. Figure 5.1 refers to this classical description.

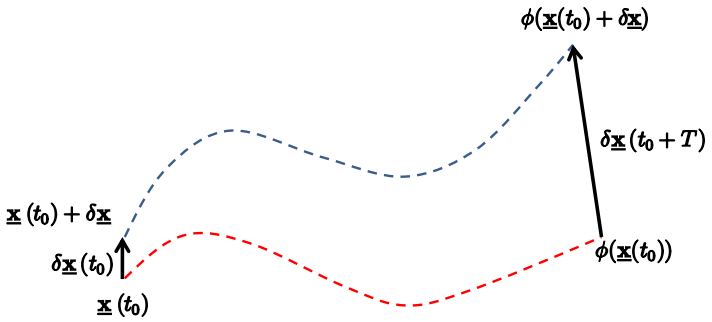


Figure 5.1: Separation of the nearby particles during time interval T due to the flow map ϕ when the two particles are released in the flow field at the same time t_0.

This figure shows two particles which are close to each other at initial time t_0 and then under the effect of the flow field their distance increases. The new vector between these two particles is described by

$$\delta\underline{\mathbf{x}}\left(t_0+T\right)=\phi_{t_0}^{t_0+T}\left(\underline{\mathbf{x}}+\delta\underline{\mathbf{x}}\right)-\phi_{t_0}^{t_0+T}\left(\underline{\mathbf{x}}\right)=D\phi_{t_0}^{t_0+T}\left(\underline{\mathbf{x}}\right)+\mathrm{O}\left(\left\|\delta\underline{\mathbf{x}}\left(t_0\right)\right\|^2\right) \tag{5.1}$$

where ϕ is the flow map and $D\phi_{t_0}^{t_0+T}=d\phi_{t_0}^{t_0+T}\left(\underline{\mathbf{x}}\right)/d\underline{\mathbf{x}}$ is the Jacobian of the flow map

corresponding to the initial configuration.

The maximum possible separation between the new positions of released particles after time interval T, assuming small–enough initial distance between them and with respect to an appropriate norm $\|\cdot\|$, is proportional to the square root of the maximum singular value of the right Cauchy–Green strain tensor which is calculated with respect to the Jacobian of the flow maps as

$$max \left\|\delta \underline{\mathbf{x}} \left(t_0 + T\right)\right\| = \sqrt{\lambda_{max} \left(\Delta\right)} \left\|\delta \underline{\mathbf{x}} \left(t_0\right)\right\| \tag{5.2a}$$

$$\Delta = D\phi_{t_0}^{t_0+T} \left(\underline{\mathbf{x}}\right)^{tr} D\phi_{t_0}^{t_0+T} \left(\underline{\mathbf{x}}\right) \tag{5.2b}$$

The finite time Lyapunov exponent (FTLE), a spatiotemporal extended notion of the Lyapunov exponent (LE), is defined as

$$\sigma_{t_0}^{T} \left(\underline{\mathbf{x}}, t_0\right) = \frac{1}{|T|} ln \sqrt{\lambda_{max} \left(\Delta\right)} \tag{5.3}$$

Similar to the calculation of maximum separation between two initially neighborhood points in an ODE system and the corresponding maximum Lyapunov exponents (LE) we use $\sigma_{t_0}^{T}$ to describe the $max \left\|\delta \underline{\mathbf{x}} \left(t_0 + T\right)\right\|$ as

$$max \left\|\delta \underline{\mathbf{x}} \left(t_0 + T\right)\right\| = exp \left(\sigma_{t_0}^{T} \left(\underline{\mathbf{x}}, t_0\right) |T|\right) \left\|\delta \underline{\mathbf{x}} \left(t_0\right)\right\| \tag{5.4}$$

We put forward a different point of view about the local FTLE. In this notion we are interested in particles that are sampled (or released) consecutively at a fixed location. Thus, the standard notion of a FTLE field, separation rate of nearby points, is not applicable since the considered particles are not corresponding to a same time. We show that we can recover an approximation of the true local FTLE by using the differential distances of the successive

source (or destination) points and also, we can estimate the differential distance of the source (or destination) points by having the true local FTLE and velocity time-series.

In the following sections we focus on the backward FTLE fields and the location of source points because this situation has more importance for us in the case of sampling and studying the risks of arriving aerial microorganisms and pathogens by atmospheric flow, however all the results are applicable for forward FTLE fields and spread of passive tracers.

Theorem: The true local FTLE value could be approximated by

$$\sigma_{t_0}^{T}\left(\underline{\mathbf{x}}, t_0\right) \approx \lim_{|T|\gg\delta t} \lim_{\delta t\to 0} \frac{1}{|T|} ln \frac{\delta\left(t_0, \underline{\mathbf{x}}_0, T, \delta t\right)}{\left\|\underline{\mathbf{v}}\left(\underline{\mathbf{x}}, t_0\right) \delta t\right\|} \tag{5.5}$$

where $|T|$ is the integration time, $\delta\left(t_0, \underline{\mathbf{x}}_0, T, \delta t\right)$ is the differential distance between successive source (or destination) points after integration time $|T|$, $\underline{\mathbf{v}}\left(\underline{\mathbf{x}}, t_0\right)$ is the average velocity vector at sampling (or release) intervals and at the sampling (or release) location and δt is the time interval between successive samplings (or releases).

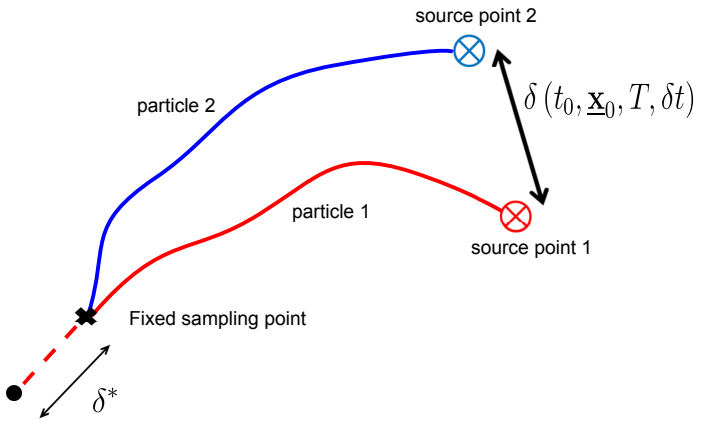


Figure 5.2: Two successive sampled particles at a fixed location shown by $\times$. Time interval between two successive sampling is δt and the integration time between the source and sampling points is T for both particles. The excessive travel length of the first particle during δt, is shown by δ^*.

Proof: We convert this statement to the classical theorem by introducing δ^* (shown by the dashed line in Fig. 5.2) which is the distance between the position of the first sampled particle (shown by a solid circle) and the sampling location at the moment the second particle is observed at that place. If sampling time interval is short enough ($\delta t \to 0$ or $\delta t \ll$ Lagrangian time scale of the velocity field) then (i) δ^* would be sufficiently small and (ii) it converges to $\|\underline{\mathbf{v}}\left(\underline{\mathbf{x}}, t_0\right)\delta t\|$. Meanwhile, The total integration time for the first particle $T + \delta t$ is approximately the same as the integration time of the second particle (since $T \gg \delta t$), thus if we look backward in time, we can consider these two particles as two isochronic and initially nearby particles which are released from two nearby points (cross and solid circle) with approximately same integration time $T \approx T + \delta t$. Finally, we need to show that $\delta\left(t_0, \underline{\mathbf{x}}_0, T, \delta t\right)$ is the maximum possible distance between the two particles (recall equation 5.4). From dynamical system concepts we know that a blob of initial conditions will stretched into the direction of the maximum Lyapunov vector. In addition, we know that if the integration time is long enough, then particles are attracted to the attracting LCSs where the local stretching rate is maximum. Thus, by choosing a sufficient integration time we conclude that the two separated particles lie along (or very close to) the major axis of an ellipse containing the blob of the initial conditions. So, $\delta\left(t_0, \underline{\mathbf{x}}_0, T, \delta t\right)$ would be close to the maximum possible stretching distance between two particles. Thus, as $\delta t \to 0$ and for sufficiently large $|T|$, we can recover the true value of FTLE at the sampling (release) point by equation (5.5).

This theorem enables us to either recover the true local FTLE if we have the distance between the source (destination) points and the local time–varying velocity or more importantly to estimate the differential distance between the source (or destination) position of sampled (or released) particles if we have the true local FLTE at the sampling location as $exp(|T|\,\sigma_{t_0}^{T}\left(\underline{\mathbf{x}}, t_0\right))\|\underline{\mathbf{v}}\left(\underline{\mathbf{x}}, t_0\right)\delta t\|$. Finally, we should note that the results of this theorem are

independent of the flow field dimension.

5.2.1 Sampling at a fix location and local FTLE

We apply this theorem to compare true and recovered local FTLE and also true and estimated differential distances of the source points corresponding to the particles that are sampled at Virginia Tech's Kentland Farm (located at 37°11′ N and 80°35′ W) where we have collected a large variety of microbial samples by using UAVs during 2007 to 2013 (Schmale et al., 2012). We refer to this point as $(0, 0)$ in our plots.

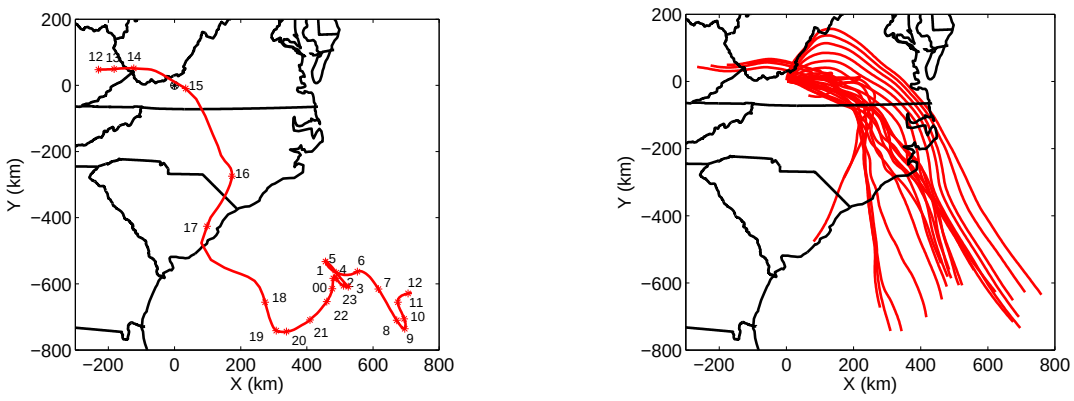
For calculations of the required flow maps we use numerical data corresponding to the North America Mesoscale, NAM–218 provided by the National Oceanic and Atmospheric Administration (NOAA) and National Centers for Environmental Prediction's (NCEP) Operational Model Archive and Distribution System (NOMADS) project[1]. Spatial resolution of this data set is about 12.1 km and its temporal resolution is 3 hours. In addition, all the trajectories are calculated by a fourth order Runge-Kutta integrator with a constant integration time step equals to 5 min. We use third order splines for all the necessary spatiotemporal interpolations.

Figure 5.3 (a) shows the initial positions of the corresponding sampled particles which are collected at the sampling location in the time interval 12:00 UTC 29 Sep to 12:00 UTC 30 Sep 2010. We call this time interval the interrogation window. The frequency of samplings is 1 hour and backward time integration is 24 hours for all the particles. In addition, for simplicity we do the integration on quasi–2D 850mb pressure surface (BozorgMagham and Ross, 2014). Indexing on this figure indicates the sequence of the initial positions with respect to the sampling time, so for example the index #12 refers to the initial position

[1] `http://nomads.ncdc.noaa.gov/data.php`

of the particle that started at 12:00 UTC 28 Sep and is sampled 24 later at the sampling location.

In terms of streaklines, we may say that this line is composed of contemporaneous points (24 hours) from the assembly of streaklines which pass through the sampling point. We call this line the *isochron source–line* since the integration time from all points on it to the sampling location is the same. Figure 5.3 (b) shows the back–trajectories of the previously indexed particles from their initial position toward the sampling point. We note strong changes in source locations and the pathline shapes during the 24 hours of samplings.



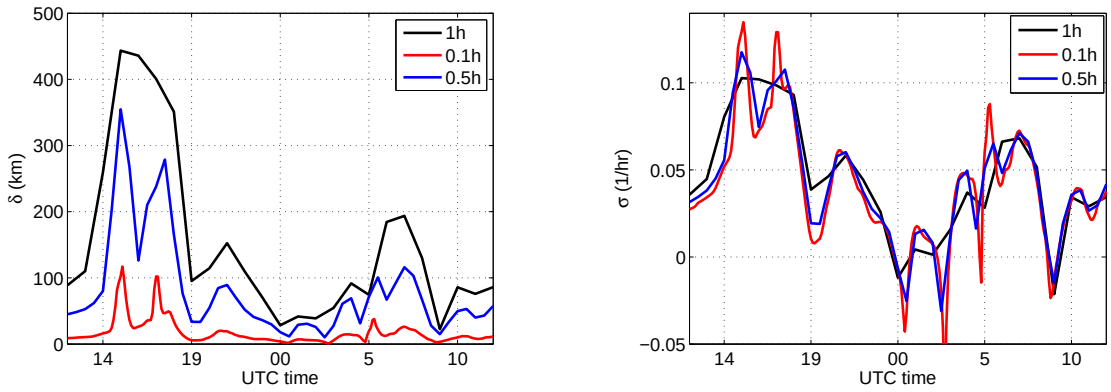
(a) (b)

Figure 5.3: (a) Sequential source points and the isochron source–line, (b) back trajectories of the sampled particles during 24 hours of integration. Sampling frequency is one hour between 12:00 UTC 29 Sep to 12:00 UTC 30 Sep 2010 and the sampling point is located at (0,0) (Virginia Tech Kentland Farm 37°11′ N and 80°35′ W).

To follow the assumptions of local recovery theorem ($\delta t \ll$ Lagrangian time scale of the velocity field which is $\approx$ 10000s for horizontal turbulence cases (Draxler and Hess, 1998)) we have to choose small sampling periods. For this aim the frequency of samplings is selected from 0.1 to 1 hour and all the integrations are done in the same interrogation window.

Figure 5.4 (a) shows the differential distance between successive source points, $\delta\left(t_0, \underline{\mathbf{x}}_0, T, \delta t\right)$, during the interrogation window. Note that in this figure we calculate the true differential

distances from the available velocity field data. We use the average velocity at the sampling point to calculate δ^* as $\|\underline{\mathbf{v}}\,(\underline{\mathbf{x}}, t_0)\,\delta t\|$. Depending on the ratio of the sampling period to the time scale of the flow field, the average velocity term, $\underline{\mathbf{v}}\,(\underline{\mathbf{x}}, t_0)$, could be considered as the average of the velocities at two successive sampling (release) times or the velocity at the mid time of the two samplings (releases). Figure 5.4 (b) show the time series of the recovered point–wise FTLE for each sampling period time.



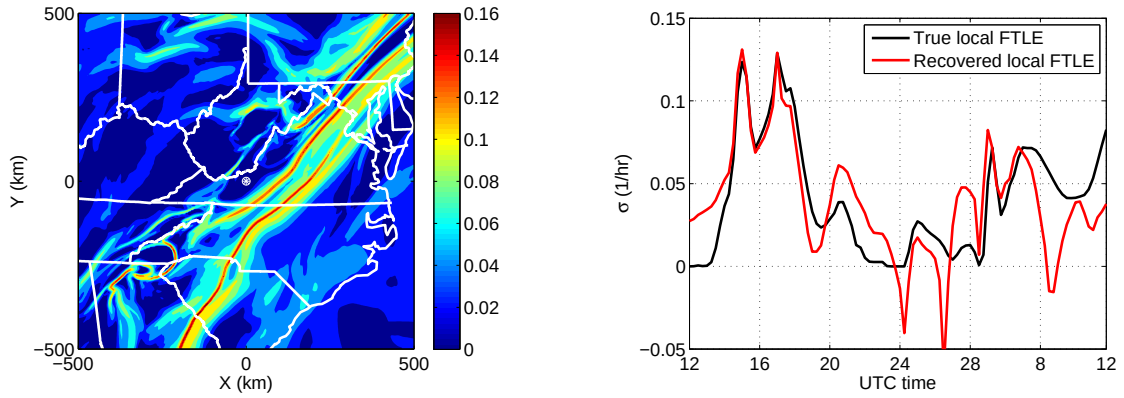
(a) (b)

Figure 5.4: (a) δ as the true differential distance between successive source points. Horizontal axis represents the averaged time corresponding to each successive pairs, (b) recovered local FTLE for different δt's form 6 min to 1 hour. Interrogation window is 12:00 UTC 29 Sep to 12:00 UTC 30 Sep 2010.

Figures 5.3 (a) and 5.4 elucidate that we are interpreting the local FTLE time-series as differential stretching of line elements along an isochron source-line.

To check the validity of this result and to study the effect of different δt's on the recovered local FTLE time-series we calculate the true backward FTLE field for the interrogation window with integration time equal to 24 hours. Figure 5.5 (a) show the first snap shot of the time–varying FTLE field corresponding to 12:00 UTC 29 Sep 2010. To have a sense about the changes of this FTLE field, we may describe the motion of the strong ridges of the field in figure 5.5 (a) toward North–West direction (upper-left corner of the figure).

Figure 5.5 (b) shows the true local FTLE value (black line) at the Kentland Farm during the interrogation window. To generate this plot we calculate the backward FTLE field every 15 minutes, then the time varying value of FTLE at $(0, 0)$ is extracted. Also for comparing the results, the recovered FTLE time–series corresponding to $\delta t = 0.1$ hr is displayed in the same panel by the red line. Figures 5.4 (b) and 5.5 (b) indicate that as the sampling period time (δt) becomes smaller the recovered time series become more similar to the true answer. For $\delta t = 0.1$ hr we observe that the two time series are highly correlated and also the maxima of the two time series (corresponding to the local maxima of the FTLE field) are at the same times by a good precision. Thus, for small-enough δt's the recovered local FTLE time–series can accurately capture the passage times of a moving ridges of a FTLE field. Detecting those ridges is important since they are the first candidates for the hyperbolic LCSs in many geophysical applications (Haller, 2011; BozorgMagham et al., 2013).



(a) (b)

Figure 5.5: (a) The first snap shot (12:00 UCT 29 Sep 2010) of the backward FTLE field during the interrogation window. Integration time is 24 hours for FLTE calculations, (b) the corresponding true (black) and recovered (red) local FTLE time series at the reference point (0,0). For the recovered time series (red), δt is equal to 0.1 hr.

Finally, we investigate whether we can estimate the differential distances by the local FTLE theorem providing some information about the velocity and FTLE data (e.g., from forecasts). Fig. 5.6 is a numerical example which shows that the true differential distance between the

successive source points (black line) is well approximated by the local FTLE theorem (red line) as $\delta \approx exp(|T|\,\sigma_{t_0}^{T}\,(\underline{\mathbf{x}}, t_0))\|\underline{\mathbf{v}}\,(\underline{\mathbf{x}}, t_0)\,\delta t\|$. Note that in this case we have the data of the true local FTLE and the local velocity. In this figure we see that for $\delta t = 15$ min, the estimated differential distance time-series is very close to the true answer and it captures the correct times of the local maxima.

This is an empirically important result. For example in case of sampling the aerial tracers, one can plan the schedule of the UAV's flights based on the available forecast FTLE fields and local wind velocity such that the collected particles of successive flights originate from the most possible diverse places (see § 5.2.2).

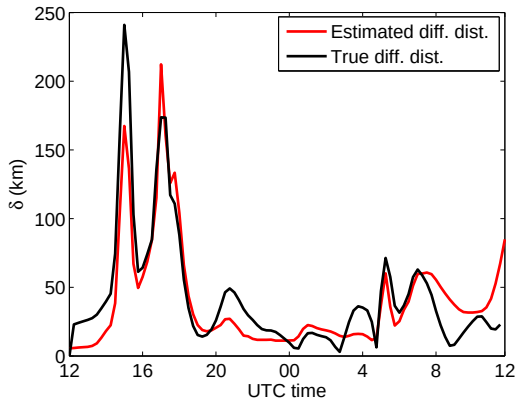


Figure 5.6: Differential distance between the successive source points on the isochron source–line corresponding to $\delta t = 0.25$ hr. The black line shows the true answer and the red line shows the approximated time series which is calculated by using the local FTLE theorem. The backward integration time for calculations of the flow maps is 24 hr and the interrogation window is 12:00 UTC 29 Sep to 12:00 UTC 30 Sep 2010.

5.2.2 Applications of the Local FTLE

Planning for maximal diversity monitoring such that the collected particles come from the most possible separated source locations is a direct result from the local FTLE recovery

theorem. Assume that we want to minimize the number of successive samplings meanwhile we have the intention to maximize the source–diversity of collected samples. Results of the local FTLE theorem tell us that the best time for collecting a limited number of samples such that they originate from the most diverse locations is when we have a high value of local FTLE at the sampling location (note the high correlation between the differential distance and the local FTLE time–series in Fig. 5.6 and Fig. 5.5 (b)). Moreover, to ensure that the particles are coming from significantly separated locations we may use the topology of the FTLE field and collect the samples on either side of a strong attracting LCS feature, providing short enough sampling period time. In this condition, the high value of $\sigma^T_{t_0}$ between the sampling moments which represents as the exponent in equation (5.4) is the key reason for having a large value of δ. Figure 5.7 schematically shows this strategy for a fixed location and a moving LCS feature.

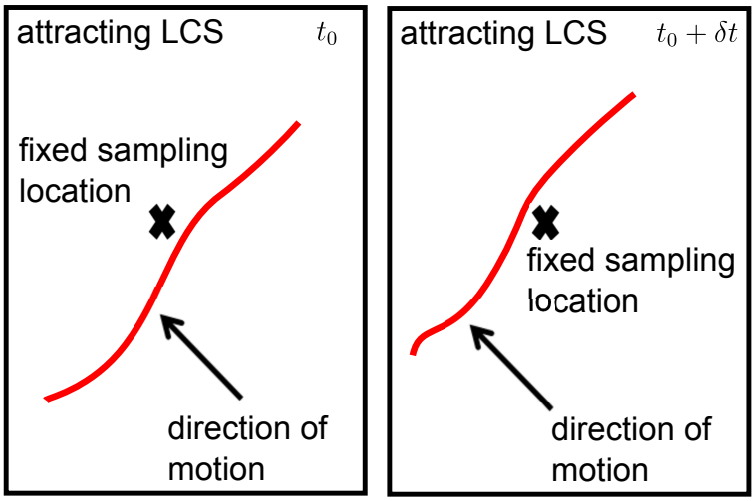


Figure 5.7: Cross sign shows the fixed location of samplings where an attracting LCS feature passes over that point. If we collect samples on either side of this feature, we would have particles from two sufficiently separated locations.

As an example in realistic geophysical flow, figure 5.8 show three backward trajectories of three particles. Integration time for all particles is 40 hours. All three particles are sampled

at $(0, -100)$ km with respect to the reference point. The sampling times are 13:40 UTC for the red particle, 14:00 UTC for the blue particle and finally 14:10 UTC time for the green particle. The green and the blue particles are sampled on one side of a LCS but the red particle is sampled on the other side of the same LCS. As we observe the source points corresponding to blue and green particles are very close meanwhile the source point of the red particle is significantly far from the other two particles. Another interesting feature of this figure is that the separation of the trajectories does not start from the sampling point, but as it is shown, the three trajectories remains close to each other for about 200 km and then the separation starts. This observation is directly related to the basic concept of the FTLE field which tells us about the final separation between nearby particles but it does not mention anything about the moment of separation.

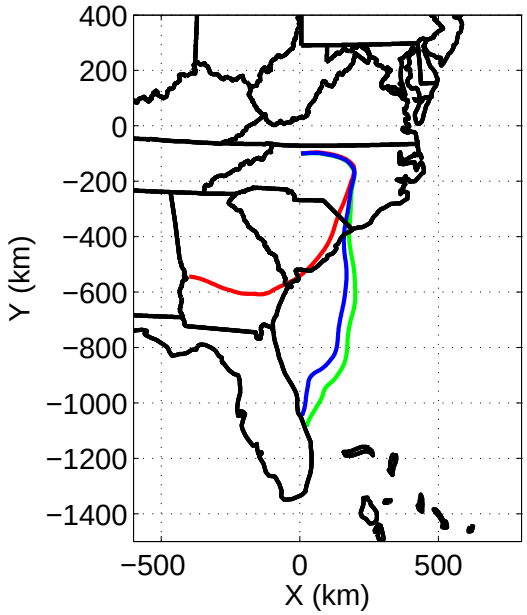


Figure 5.8: Three trajectories of collected samples, red and blue pathline correspond to samples on either side of a LCS, blue and green correspond to samples on one side of the same LCS. Sampling times are 13:40, 14:00 and 14:10 UTC time for the (virtually) red, blue and the green particles respectively.

Another important application of the local FTLE is that it may help us to explain the observation of significant changes of characteristics (genetic types or aerial density) of the collected microbial samples at the sampling location during short intervals (Lin et al., 2013; Lin, 2013). Referring to this notion, observing diverse samples may be associated with diverse source points as a result of sampling in presence of a high value local FTLE and/or the passage of a strong LCS over the sampling area.

In addition, a direct result of this theorem is that when the local backward FTLE value remains small during the sampling process, then one can conclude that the (aerial) sampled particles originate from some close areas, assuming that all the sampled particles have approximately the same flight time. This might be the reason that the characteristics of the microbial samples remain quasi-constant in many consecutive UAV's collections. This situation is similar to sampling from a *coherent set* where the FTLE values are generally small (Froyland et al., 2010; Tallapragada and Ross, 2013) and the particles have similar characteristics. Moreover, in cases that we observe significant changes in collected samples while the local FTLE value is small, we may conclude that those changes are due to some local causes and not because of long range transport phenomena.

5.3 Unresolved turbulence

In this section we study the uncertainties of the source (destination) point calculation results due to low resolution data of the velocity field. Finally we investigate validity of some results of the previous section in presence of unresolved turbulence.

Precise determination of the source location of any sampled particle and calculations of the flow maps require a high resolution data of the velocity field. But in reality, available geophysical data sets are spares due to many technical limitations and we only have data on

grid points and at discrete times. For example, spatial and temporal resolution of operational data set varies from the order of 10 to hundreds of kilometers and 3 hours to longer intervals respectively. Meanwhile, spatiotemporal scales of atmospheric flows are usually small and we may lose important Lagrangian phenomena such as turbulent diffusion and small size eddies if we just consider the available grid data (Csanady, 1973; Rodean, 1987). Thus, in calculation of the source (destination) points it is necessary to consider the uncertainty of the trajectories and pathlines. For this aim we consider a Lagrangian particle dispersion model (LPDM) which let us to calculate the stochastic component of the velocity with respect to the available deterministic grid data set. In this model the stochastic term is added to the deterministic background flow field to compensate the effects of the un-captured probable fluctuating velocities. Equation (5.6) shows that the velocity vector is composed of the deterministic component of the velocity, $\bar{\mathbf{v}}(\mathbf{x}, t)$, and a random variable, $\mathbf{V}(\mathbf{x}, \bar{\mathbf{v}}, t)$ which depends on the instantaneous position of the particle and its deterministic velocity at that location. As we see latter this dependency is very important for calculation of the probability distribution of the source or destination regions (BozorgMagham and Ross, 2014).

$$\mathbf{v}(\mathbf{x}, t) = \bar{\mathbf{v}}(\mathbf{x}, t) + \mathbf{V}(\mathbf{x}, \bar{\mathbf{v}}, t). \tag{5.6}$$

In this study we mainly focus on the probabilistic *source regions* (corresponding to backward trajectories) since there is a delicate point in stochastic velocity calculations which could easily be neglected and thus mislead the calculations. In contrast to the case of probabilistic source region calculations, finding the probabilistic destination region (corresponding to forward trajectories) is straightforward and much easier and for this reason we just show the principal of this procedure. In addition and regarding the local FTLE theorem, we revisit the problem of successive samplings when we consider the effects of unresolved turbulence. Our numerical results show that even in presence of unresolved turbulent, if successive samplings

are performed on either side of a LCS, then the probabilistic source regions are significantly separated by the LCS features as in the case of deterministic flow fields.

5.3.1 Probabilistic source and destination regions

To focus on the main conceptual concerns and to avoid complexity we prefer to proceed with 2D field, like the previous sections, however this approach could be extended to 3D fields by adding the vertical component of the field and an appropriate stochastic term in that direction (Rodean, 1987). In addition, there are various atmospheric dispersion models which are useful for different condition of the atmospheric flow. The one that we apply, considering practical simplicity and availability of the required data was introduced and used by Legg and Raupach (1982); Fay et al. (1995); Draxler and Hess (1998); Stohl et al. (2005). This Lagrangian particle dispersion model describes the stochastic velocity term as a random variable which is a function of the tensor of velocity deformation and the Lagrangian time scale of the flow field. In a compact form we may write down the stochastic term of the equation (5.6) as a Markov–chain process as

$$V_{(t+\Delta t)} = R_{\Delta t} V_t + \left(1 - R_{\Delta t}^2\right)^{0.5} \mathcal{N}(0,1) \sqrt{\kappa / T_L} \tag{5.7}$$

where V shows each component of the stochastic velocity term ($\mathbf{V}$) and $R_{\Delta t}$ is a measure of the assossiation between stochastic velocity values of successive steps. Equation (5.8) shows the correlation between two steps as a function of integration time step length, Δt, and the Lagrangian time scale of the flow field, T_L

$$R_{\Delta t} = \exp\left(-\Delta t / T_L\right). \tag{5.8}$$

Referring to equation (5.7) we see that κ is a key element in calculation of V. This term depends on the gradient of the instantaneous deterministic velocity, $\bar{\mathbf{v}}$, meteorological data grid size, χ, and an empirical constant, c, as

$$\kappa = 2^{-0.5}\,(c\chi)^2 \left[\left(\frac{\partial \bar{v}}{\partial x} + \frac{\partial \bar{u}}{\partial y} \right)^2 + \left(\frac{\partial \bar{u}}{\partial x} - \frac{\partial \bar{v}}{\partial y} \right)^2 \right]^{0.5}. \tag{5.9}$$

Because the final velocity term is a sum of the deterministic grid scale (background) velocity and the stochastic term as a random variable the trajectories would be continuous but not differentiable. Also each realization of the set of equations shows one possible solution for the particle's trajectory. Referring to our main problem and based on the dependency of κ to the gradient of the background velocity we have two distinct cases, (i) finding the destination regions of the released particles, (ii) finding the source regions of the sampled particles.

(i) Probabilistic destination region is the probability distribution of the final positions of the (virtually) released particles after integration time T when the initial position is known precisely as a Dirac delta function. Thus, the case of forward integration and related calculations of probabilistic distribution is equivalent to solving the Fokker–Planck or Kolmogorov forward equations (Rodean, 1987; Risken and Eberly, 1985) which describe the future of a probability distribution function of a known initial condition that evolves under the dynamics of a system e.g., diffusion process. Since the time–varying vector field is complicated and analytical solutions are not practical we have to perform numerical solution to find the distribution of end points. For this aim, first we discretize the domain of motion into sufficiently small boxes and then we use the Monte Carlo method by releasing sufficient number of independent (virtual) particles from the box which includes the release point. Figure 5.9 shows this procedure. By choosing an appropriate integration time step we may produce segments of the trajectory of each virtual particle. By completion of the integration process

we would have a distribution of particles ending in some boxes. By considering a mesh grid over the domain, we may count the number of virtual particles in each box. If the total number of released particles is large and the boxes' dimensions are small enough, then the ratio of the virtual particles in each box to the total number of released particles show the probability distribution of the destination region. By increasing the number of virtual particles and decreasing the size of the boxes the resultant distribution become invariant.

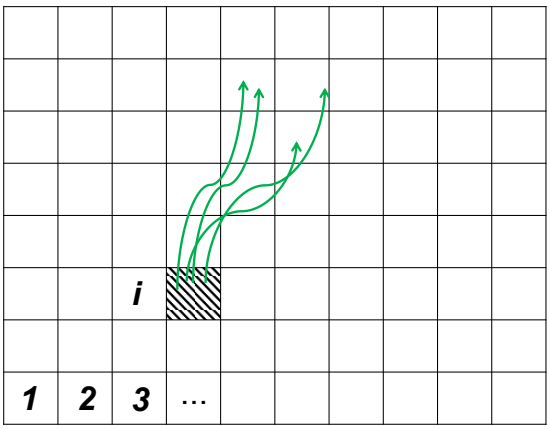


Figure 5.9: Solution for probability distribution of the forward case, starting time is t_0 and integration time is T. Virtual particles are released from one specific box (containing the release location) and then the distribution of the landings would specify the probabilistic destination region. Calculation of the probabilistic destination region is equivalent to the solution of the Fokker-Planck equation for finding the future probability distribution of an initially known distribution. Trajectories from the release box are shown in green.

(ii) Probabilistic source region cannot be determined by direct backward integration, similar to previous procedure, since *any* assumption about the particle's position (i.e., sampling location at the beginning of backward integration process) will determine κ term which consequently yields a new position of the particle. As this procedure continues a false trajectory is generated for each realization of the stochastic differential equation. In other word, since the vector field is time–varying, if we start from the sampling time and location, and go backward in time all of our calculations are based on false "future" data.

In this case finding the probabilistic source region is conceptually the same as solving the Kolmogorov backward problem (Risken and Eberly, 1985). In mathematical term, at time $t_0 - T$ (note that T is the integration time) we investigate for a specific source configuration such that in future time t_0 the system will be in a given target set which we call it the sampling box.

Similar to the forward case we have no analytical solution for general time–varying vector field, thus we have to use a numerical algorithm to find the probabilistic source distribution. For this aim, we discretize the domain of the flow field into small boxes. Then we shift the starting time to $t_0 - T$. By this means, we convert this problem into a forward integration problem from $t_0 - T$ to t_0. At $t_0 - T$ we release large number of independent virtual particles from *all* boxes of the domain. By forward integration to time t_0 we find the landing location of each released (virtual) particle. The important particles in this procedure are those which land inside the sampling (target) box. Figure 5.10 show this process. In this figure those boxes that have partial contribution to the landed particles in target box are hatched. As we observe, there may be particles from some contributing boxes that do not land in the target box. This fact is due to the randomness of stochastic velocity term and is common in this kind of calculations.

In figure 5.10 the sampling box is shown by the index j and other boxes are shown by $i = 1, 2, \cdots$. We denote The number of particles which start from box i and land in box j by $n_{i \to j}$. We calculate the relative contribution of each source box as,

$$\gamma_i = \frac{n_{i \to j}}{\sum_i n_{i \to j}}, \tag{5.10}$$

where $\sum_i n_{i \to j}$ shows the total number of particles which land in sampling (target) box j. Thus, γ_i shows the chance of a sampled particle to come from a specific box i. Consequently,

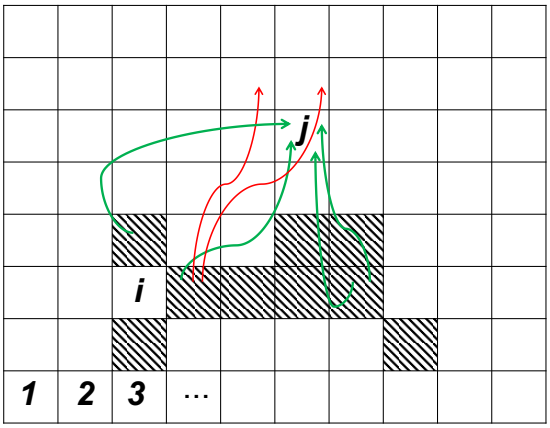


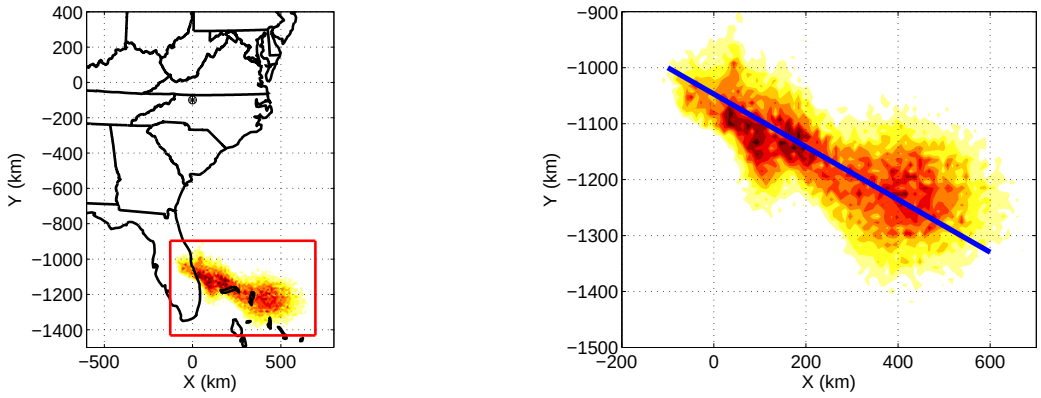
Figure 5.10: Solution for probability distribution of the backward case, starting time is shifted to $t_0 - T$ and the integration time is T. Virtual particles are released from *all* the boxes in the domain. Important particles are those who land in target box which include the sampling location. Solution of the probabilistic source region is conceptually the same as the solution of backward Kolmogorov equation where an initial probability distribution is the desired solution such that in future time the system will have a specified probability distribution. Trajectories of particles which land in the target box are shown by green, other trajectories are shown by red.

numerical distribution of γ over the domain shows the probability distribution of the source region.

Although this procedure works and yields the probabilistic distribution of source points, its numerical efficiency is low since we release many independent particles (e.g., 10^6) from each box of the domain but only those particles which land in a specific box (target or sampling box) are of our interest. Thus, there would be huge number of particles that we calculate their trajectory but they do not land in our sampling box, thus they are not important for our purpose of finding the source distribution and we have to leave them out. We may increase the efficiency of this procedure by sequential release of particles from all the boxes and identifying the regions with a minimum contribution. After that, we may limit our search procedure to those boxes and then we can increase the number of releases to distinguish the

spatial patterns of the probabilistic source points.

Figure 5.11 shows one example of the probabilistic source distribution where the color intensity shows the relative contribution of each source box. In this case the sampling point is located at $(0, -100)$ km with respect to our reference point. Sampling time is 14:15 UTC 29 Sep 2010 and the integration time for probabilistic source region is 40 hours. This figure is the stochastic equivalent of the source point of the particle which its pathline was shown in green (Fig. 5.8). For this calculation 10^5 particles are released from each box which is a square of 10×10 km, so we have 10^3 particles per square kilometer. The search area for this specific probabilistic source region is a $900 \times 600\,\mathrm{km}^2$ rectangular. Considering the size of the boxes we would have 5400 boxes, thus the total number of released particles and calculated trajectories would be 5.4×10^8 in each integration time step.



(a) (b)

Figure 5.11: (a) Probabilistic equivalent of the source point of the (virtually) green particle in Fig. 5.8, the sampling point is located at $(0, -100)$ km with respect to our reference point and sampling time is 14:15 UTC 29 Sep 2010, (b) details of the probabilistic source region which is composed of 5400 boxes, each $10 \times 10\,\mathrm{km}^2$. Color intensity shows the relative contribution of each source box. To investigate the distribution, relative contribution of the boxes along the specified diagonal line is shown in Fig. 5.12.

An important point for both source and destination region cases is that although at each time step the stochastic velocity term has a normal distribution (equation (5.7)), but the final

distribution is not necessarily Gaussian. The reason for this fact is the cumulative effects of the variability of the variance of normal distribution ($\sqrt{\kappa/T_L}$) which is a function of the gradient of instantaneous velocity. In general, for small integration time the probability distribution of the source (destination) region is close to a Gaussian distribution but as the integration time increases, the corresponding distribution diverges from a normal one. As an example, visual inspection of Fig. 5.11 (b) and also Fig. 5.12 which shows the relative contribution of the source boxes along the specific diagonal line indicate that the final distribution of the probable source points is not Gaussian.

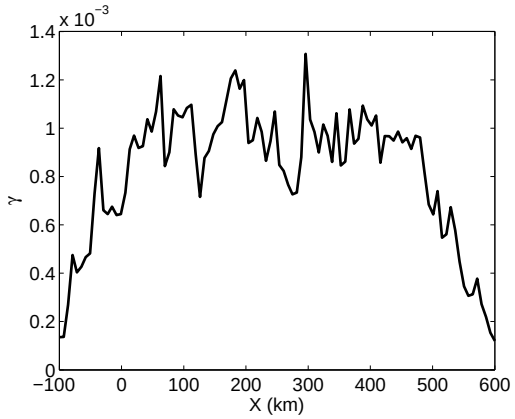


Figure 5.12: γ, relative contribution of source boxes along the diagonal line in Fig. 5.11 (b) projected on the horizontal direction.

Now, we revisit the problem of successive samplings on either side of a LCS. In figure 5.8 we show that the sources location of the two sampled particles on either side of a LCS is much further apart than the source points of the other two successive sampled particles on one side of the same LCS. We want to investigate whether in presence of unresolved turbulence this result is still valid. If the previous results hold, we conclude that two sampled particles on either side of a LCS originate from two significantly separated regions regardless of deterministic or stochastic solutions. This fact is important in practical applications such as the case of sampling the microbial structure of the atmosphere since it enable us to

have some judgments about the source of the sampled particles just based on deterministic analysis and without doing the cumbersome calculations regarding the probabilistic source regions.

We study the separation of probabilistic source regions corresponding to successive samplings on either side of a LCS through an example and we leave the general conclusion for future work. Results of this example shows satisfactory results and assures us that by sampling on either side of a deterministic LCS the probabilistic source regions are significantly separated and the sampled particles originate from far apart locations.

Figure 5.13 show the evolution of the probabilistic source regions "A" and "B" (shown in panel (a)) corresponding to (virtually) red and blue particles of Fig. 5.8 respectively. The total integration time for this example is 40 hours. In each panel of this figure we show the contemporaneous attracting hyperbolic LCSs which are calculated by the method of Haller (2011) and Karrasch (2012). For calculation of each probabilistic region of this figure, 10^5 particles are released from each small 10×10km box and the procedure that introduced earlier is followed. As we observe in panel (a) of this figure, the two source regions corresponding to the sampled particles on either side of a LCS are significantly separated from each other. Also we see how the two probabilistic clouds contract and become closer to the attracting LCS as they get closer to the sampling point. One noticeable feature in this figure is the difference between the shapes of the two source regions, while they correspond to a very close sampling times (less than 30 min).

5.4 Discussion

In this paper we propose a theorem which describes a new interpretation about the local FTLE. Mathematical concept of the local FTLE enable us to either recover the true local

FTLE time-series if we have the local velocity and differential distance data or more importantly, to estimate the differential distances between the source (or destination) locations of successive sampled (released) particles at a fixed location if we have the corresponding local velocity and FTLE time-series data. The suggested notion is useful in cases that we have a collection of particles which are sampled at a fixed location and we want to find an association between the characteristics of the collected particles, their source locations and the local FTLE time–series. We show that if samplings are performed in short interval of time then the differential distance between the source locations is governed by the local FTLE and velocity. This result may help us to explain the observation of characteristic variations in collected (microbial) samples. Moreover, we show that the concept of local FTLE could be useful for scheduling successive samplings to obtain maximal–diversity collections.

In the last part of this paper we investigate the unresolved turbulence and the stochastic description of the source (destination) regions. We use the idea of box method and discuss the important differences between calculation methods of the probabilistic source and destination regions. Finally, we study the probabilistic source regions corresponding to successive sampled particles on either side of a strong hyperbolic LCS and we show that similar to the deterministic flow field, the source regions are significantly separated.

Result of this study could help us to have a better understanding about the outcomes of localized samplings of the geophysical fluids. Also it may provide beneficial information for optimal monitoring of passive tracers.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant Number CMMI–1100263 (Dynamical Mechanisms Influencing the Population Structure of

Airborne Pathogens: Theory and Observations). Part of this research was done during a visit by SDR to Instituto de Ciencias Matemáticas, Madrid Spain. He thanks ICMAT for its hospitality and support from MINECO: ICMAT Severo Ochoa project SEV-2011-0087.

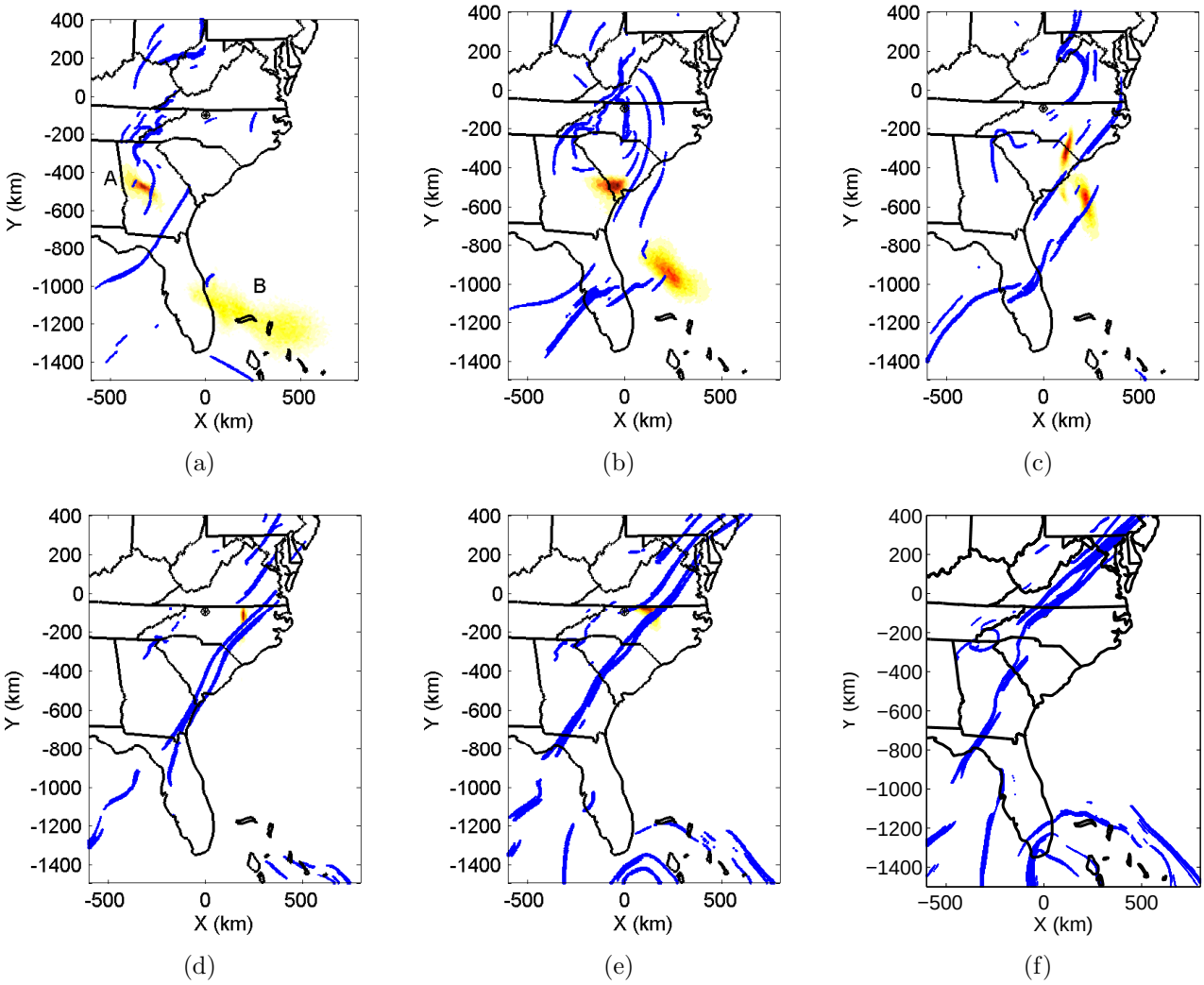


Figure 5.13: Sequence of hyperbolic LCSs (blue) and two clouds of probabilistic source regions corresponding to two successive samples. Probabilistic regions "A" and "B" (panel (a)) correspond to the virtually red and blue particles in Fig. 5.8. These six panels correspond to 40, 30, 20, 10, 5 and 0 hours of probabilistic source regions, respectively.

Chapter 6

Conclusion

This study is motivated by the atmospheric transport of microorganisms. To describe and analysis this time-varying phenomena, we use the concepts of dynamical system such as finite-time Lyapunov exponent and Lagrangian coherent structures which are powerful tools for describing and exploring non-autonomous fields and provide valuable information about the global structures of particles' motion in the moving fluid domain. Considering the main objective of this research and also realistic condition of our data, we focus on five major subjects: forecasting, uncertainty analysis, ensemble forecasting, statistical analysis of the local measurements and finally the connection between the local measurements and the local FTLE time-series.

Since the LCSs govern the mixing and transport phenomena and because we want to use them to forecast the particles' motion, we need to have reliable prediction about them. Thus, we investigate questions such as: how accurate and precise are forecast FTLE-LCSs? This question is important since we encounter the cumulative effects of the Eulerian errors in our Lagrangian calculations of FTLE fields and the associated LCSs. So, in the first part we evaluate the prediction results of FTLE-LCS features. Our results show how the errors of

wind field forecasts degrade the forecast FTLE-LCSs.

Unavoidable errors of the wind field forecasting which induce considerable errors in calculated FTLE-LCS features show that deterministic forecast results may not be applicable for reliable predictions and we need to know about the uncertainties of the results. This fact inspires us to propose and use ensemble forecasting methods for FTLE-LCS calculations. We show that by using this approach we would be able to measure the uncertainty of the results and also to improve the quality of forecasting compering to single deterministic calculations. In addition, we investigate the uncertainty of particles' trajectories due to low resolved operational velocity data. The observed uncertainty of the source (destination) points is the motivation for defining the stochastic FTLE fields. We show that uncertainty of the source (destination) points highly depends on the bulk (deterministic) velocity field. We show that a stochastic version of a FTLE field is not necessarily a uniform fade of the deterministic field, but it depends on the time-varying velocity field and we may see different changes over the field. The last two topics of this work are related to what we locally observe, collected samples from microbial structure of atmosphere and also the local FTLE time series. Regarding to the local collection of microbial samples, we study some statistical characteristics of the collected samples such as the autocorrelation coefficients, type of the stochastic process and also the variations of aerial density during short time intervals. Then we investigate a mathematical framework by proposing a theorem to connect the variations of characteristics of the collected samples to the concept of local FTLE time-series.

We may suggest some future researches that can be followed from this thesis,

- Development of an automatic system of data pre-processing, FTLE-LCS calculations and post-processing based on the operational and online data sets, capable of handling different data formats and different data bases,

- Development of a framework for obtaining FTLE-LCSs from 3D velocity fields, especially for geophysical fluids, along with the necessary pre-processing steps (e.g., WRF) for generating 3D velocity fields from operational data sets,

- Development of a numerical algorithm for a fast and reliable computation of the probabilistic source regions (optimal numerical solver for the backward Kolmogorov problem in presence of an arbitrary flow field),

- Collecting microbial samples based on the results of the local FTLE theorem (maximal-diversity observation) and investigating their characteristic changes with respect to the local FTLE values,

- Studying the association between the LCS features and some meteorological phenomena such as temperature fronts and/or precipitation regions.

Bibliography

Abarbanel, H. D., Brown, R., Kennel, M. B., 1992. Local lyapunov exponents computed from observed data. Journal of Nonlinear Science 2 (3), 343–365.

Antoulas, A., 2005. An overview of approximation methods for large-scale dynamical systems. Annual Reviews in Control 29, 181–190.

Aylor, D., Taylor, G., Raynor, G., 1982. Long range transport of tobacco blue mold spores. Agricultural Meteorology 27, 217–232.

Aylor, D. E., 2003. Spread of plant disease on a continental scale: role of aerial dispersal of pathogens. Ecology 84 (8), 1989–1997.

Barkai, E., Metzler, R., Klafter, J., 2000. From continuous time random walks to the fractional fokker-planck equation. Physical Review E 61 (1), 132.

Berek, L., Petri, I., Mesterhazy, A., Téren, J., Molnár, J., 2001. Effects of mycotoxins on human immune functions in vitro. Toxicology in vitro 15 (1), 25–30.

BozorgMagham, A. E., Ross, S. D., 2014. Atmospheric lagrangian coherent structures considering unresolved turbulence and forecast uncertainty. ommunications in Nonlinear Science and Numerical Simulation -, –.

BozorgMagham, A. E., Ross, S. D., Schmale, D. G., 2013. Real-time prediction of atmospheric Lagrangian coherent structures based on uncertain forecast data: An application and error analysis. Physica D 258, 47–60.

Branicki, M., Mancho, A. M., Wiggins, S., 2011. A lagrangian description of transport associated with a front–eddy interaction: Application to data from the north-western mediterranean sea. Physica D: Nonlinear Phenomena 240 (3), 282–304.

Buergler, A. L., Fike, J. H., Burger, J. A., Feldhake, C. R., McKenna, J. A., Teutsch, C. D., 2005. Botanical composition and forage production in an emulated silvopasture. Agronomy Journal 97 (4), 1141–1147.

Caplan, P., Derber, J., Gemmill, W., Hong, S.-Y., Pan, H.-L., Parrish, D., 1997. Changes to the 1995 NCEP operational medium-range forecast model analysis-forecast system. Weather and Forecasting 12 (3), 581–594.

Csanady, G. T., 1973. Turbulent Diffusion in the Environmnet. Vol. 3. Springer.

Danielsen, E. F., 1961. Trajectories: Isobaric, isentropic and actual. Journal of Atmospheric Sciences 18, 479–486.

Dosio, A., Guerau de Arellano, J. V., Holtslag, A. A., Builtjes, P. J., 2005. Relating eulerian and lagrangian statistics for the turbulent dispersion in the atmospheric convective boundary layer. Journal of the atmospheric sciences 62 (4), 1175–1191.

Draxler, R., Hess, G., 1998. An overview of the hysplit_4 modelling system for trajectories, dispersion and deposition. Australian meteorological magazine 47, 295–308.

Dubuisson, M.-P., Jain, A. K., 1994. A modified hausdorff distance for object matching. Proceedings of the 12th IAPR International Conference, IEEE 1, 566–568.

Ehrendorfer, M., 1997. Predicting the uncertainty of numerical weather forecasts: a review. Meteorologische zeitschrift 6, 147–183.

Evensen, G., 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. Journal of Geophysical Research: Oceans (1978–2012) 99 (C5), 10143–10162.

Fay, B., Glaab, H., Jacobsen, I., Schrodin, R., 1995. Evaluation of Eulerian and Lagrangian atmospheric transport models at the deutscher-wetterdienst using ANATEX surface tracer data. Atmospheric environment 29, 2485–2497.

Froyland, G., Lloyd, S., Santitissadeekorn, N., 2010. Coherent sets for nonautonomous dynamical systems. Physica D: Nonlinear Phenomena 239 (16), 1527–1541.

Gifford, F., 1987. The time-scale of atmospheric diffusion considered in relation to the universal diffusion function, f1. Atmospheric environment 21 (6), 1315–1320.

Gonzalez, R. C., Woods, R. E., 2007. Digital Image Processing. Prentice Hall.

Griffa, A., Piterbarg, L., Ozgokmen, T., 2004. Predictability of Lagrangian particle trajectories: Effects of smoothing of the underlying Eulerian flow. Journal of Marine Research 62, 1–35.

Haagenson, P. L., Gao, K., Kuo, Y., 1990. Evaluation of meteorological analyses, simulations, and long-range transport using ANATEX surface tracer data. Journal Applied Meteorology 29, 1268–1283.

Haller, G., 2002. Lagrangian coherent structures from approximate velocity data. Physics of Fluids 14, 1851–1861.

Haller, G., 2011. A variational theory of hyperbolic Lagrangian coherent structures. Physica D 240, 574–598.

Haller, G., Poje, A., 1998. Finite time transport in aperiodic flows. Physica D 119, 352–380.

Haller, G., Sapsis, T., 2011. Lagrangian coherent structures and the smallest finite-time Lyapunov exponent. Chaos 21, 023115.

Haller, G., Yuan, G., 2000. Lagrangian coherent structures and mixing in two-dimensional turbulence. Physica D 147, 352–370.

Hester, C. F., Casasent, D., 1980. Multivariant technique for multiclass pattern recognition. Applied Optics 19, 1758–1761.

Houtekamer, P. L., Mitchell, H. L., 1998. Data assimilation using an ensemble kalman filter technique. Monthly Weather Review 126 (3), 796–811.

Jolliffe, I. T., Stephenson, D. B., 2003. Forecast verification. Wiley.

Kahl, J. D., Harris, J. M., Herbert, G. A., Olson, M. P., 1989. Intercomparison of three long-range trajectory models applied to arctic haze. Tellus Series B-Chemical and Physical Meteorology 41, 524–536.

Kahl, J. D., Samson, P. J., 1986. Uncertainty in trajectory calculations due to low resolution meteorological data. Journal of Climate and Applied Meteorology 25, 1816–1831.

Kalnay, E., 2003. Atmospheric Modelling, Data Assimilation and Predictability. Cambridge.

Kalnay, E., Corazza, M., Cai, M., 2002. Are bred vectors the same as lyapunov vectors? In: EGS General Assembly Conference Abstracts. Vol. 27. p. 6820.

Kaltschmitt, M., Streicher, W., Wiese, A., 2007. Renewable energy: technology, economics and environment; 66 tables. Springer.

Karrasch, D., 2012. Comment on “a variational theory of hyperbolic Lagrangian coherent structures, Physica D 240 (2011) 574-598”. Physica D 241, 14701473.

Kellogg, C. A., Griffin, D. W., 2006. Aerobiology and the global transport of desert dust. Trends in Ecology & Evolution 21 (11), 638 – 644.

Kinzelbach, W., 1988. The random walk method in pollutant transport simulation. In: Groundwater flow and quality modelling. Springer, pp. 227–245.

Kumar, V. B., Mahalanobis, A., Juday, R. D., 2005. Correlation Pattern Recognition. Cambridge.

Legg, B., Raupach, M., 1982. Markov-chain simulation of particle dispersion in inhomogeneous flows - the mean-drift velocity induced by a gradient in Eulerian velocity variance. Boundary-layer meteorology 24, 3–13.

Lekien, F., Coulliette, C., Mariano, A., Ryan, E., Shay, L., Haller, G., Marsden, J., 2005. Pollution release tied to invariant manifolds: A case study for the coast of florida. Physica D 210, 1–20.

Lekien, F., Ross, S. D., 2010. The computation of finite-time Lyapunov exponents on unstructured meshes and for non-Euclidean manifolds. Chaos 20, 017505.

Lekien, F., Shadden, S. C., Marsden, J. E., 2007. Lagrangian coherent structures in n-dimensional systems. Journal of Mathematical Physics 48, 065404.

Lemons, D. S., Langevin, P., 2002. An introduction to stochastic processes in physics. JHU Press.

Lermusiaux, P. F., Chiu, C.-S., Gawarkiewicz, G. G., Abbot, P., Robinson, A. R., Miller, R. N., Haley, P. J., Leslie, W. G., Majumdar, S. J., Pang, A., et al., 2006. Quantifying uncertainties in ocean predictions. Tech. rep., DTIC Document.

Leslie, J. F., Summerell, B. A., 2006. The Fusarium laboratory manual. Blackwell Publishing.

Lin, B., 2013. Movement and structure of atmospheric populations of Fusarium. Ph.D. thesis, Virginia Tech.

Lin, B., BozorgMagham, A. E., Ross, S. D., Schmale, D. G., 2013. Small fluctuations in the recovery of fusaria across consecutive sampling intervals with unmanned aircraft 100 m above ground level. Aerobiologia 29 (1), 45–54.

Lorenz, E. N., 1963. Deterministic nonperiodic flow. Journal of the Atmospheric Sciences 20, 475–480.

Madden, L. V., Wheelis, M., 2003. The threat of plant pathogens as weapons against U.S. crops. Annu. Rev. Phytopathol. 41, 155–176.

McMullen, M., Jones, R., Gallenberg, D., 1997. Scab of wheat and barley: a re-emerging disease of devastating impact. Plant disease 81 (12), 1340–1348.

Murphy, A. H., 1988. Skill scores based on the mean square error and their relationships to the correlation coefficient. Monthly weather review 116 (12), 2417–2424.

Øksendal, B., 2003. Stochastic differential equations. Springer.

Okubo, A., Levin, S. A., 2001. Diffusion and ecological problems: modern perspectives. Vol. 14. Springer.

Olascoaga, M. J., Haller, G., 2012. Forecasting sudden changes in environmental pollution patterns. PNAS 109, 4738–4743.

Olcay, A. B., Pottebaum, T. S., Krueger, P. S., 2010. Sensitivity of Lagrangian coherent structure identification to flow field resolution and random errors. Chaos 20, 017506.

Oseledec, V. I., 1968. A multiplicative ergodic theorem. lyapunov characteristic numbers for dynamical systems. Trans. Moscow Math. Soc 19 (2), 197–231.

Ozgokmen, T., Griffa, A., Mariano, A., Piterbarg, L., 2000. On the predictability of Lagrangian trajectories in the ocean. Journal of Atmospheric and Oceanic Technology 17, 366–383.

Palmer, T., 2000. Predicting uncertainty in forecasts of weather and climate. Reports on Progress in Physics 63, 71–116.

Pan, Z., Yang, X. B., Pivonia, S., Xue, L., Pasken, R., Roads, J., 2006. Long-term prediction of soybean rust entry into the continental united states. Plant Disease 90, 840–846.

Peng, J., Peterson, R., 2012. Attracting structures in volcanic ash transport. Atmospheric Environment 48, 230–239.

Prussin, A. J., Li, Q., Malla, R., Ross, S. D., Schmale III, D. G., 2013. Monitoring the long distance transport of *fusarium graminearum* from field-scale sources of inoculum. ( ), submitted.

Risken, H., Eberly, J., 1985. The fokker-planck equation, methods of solution and applications. Journal of the Optical Society of America B Optical Physics 2, 508.

Rockafellar, T. R., Wets, R. J.-B., 2005. Variational Analysis. Springer-Verlag,.

Rodean, H. C., 1987. Stochastic Lagrangian models of turbulent diffusion. American Meteorological Society.

Rowley, C. W., 2005. Model reduction for fluids, using balanced proper orthogonal decomposition. International Journal of Bifurcation and Chaos 15, 997–1013.

Schmale, D. G., Ross, S. D., Fetters, T., Tallapragada, P., Wood-Jones, A., Dingus, B., 2012. Isolates of fusarium graminearum collected 40 to 320 meters above ground level cause fusarium head blight in wheat and produce trichothecene mycotoxins. Aerobiologia 28, 1–11.

Schmale Iii, D. G., Dingus, B. R., Reinholtz, C., 2008. Development and application of an autonomous unmanned aerial vehicle for precise aerobiological sampling above agricultural fields. Journal of Field Robotics 25 (3), 133–147.

Schmale III, D. G., Leslie, J. F., Zeller, K. A., Saleh, A. A., Shields, E. J., Bergstrom, G. C., 2006. Genetic structure of atmospheric populations of gibberella zeae. Phytopathology 96 (9), 1021–1026.

Schneider, R. W., Hollier, C. A., Whitam, H. K., Palm, M. E., McKemy, J. M., Hernandez, J. R., Levy, L., DeVries-Paterson, R., 2005. First report of soybean rust caused by phakopsora pachyrhizi in the continental united states. Plant Disease 89, 774–774.

Senatore, C., Ross, S. D., 2011. Detection and characterization of transport barriers in complex flows via ridge extraction of the finite time Lyapunov exponent field. International Journal for Numerical Methods in Engineering 86, 1163–1174.

Shadden, S. C., Lekien, F., Marsden, J. E., 2005. Definition and properties of Lagrangian coherent structures from finite-time Lyapunov exponents in two-dimensional aperiodic flows. Physica D 212, 271–304.

Sirovich, L., 1987. Turbulence and the dynamics of coherent structures. Quarterly of Applied Mathematics 45, 561–571.

Smagorinsky, J., 1963. General circulation experiments with the primitive equations. Monthly Weather Review 91, 99–164.

Snyder, J. P., 1987. Map Projections-A Working Manual. U. S. Government Printing Office.

Stohl, A., 1998. Computation, accuracy and applications of trajectories - a review and bibliography. Atmospheric Environment 32, 947–966.

Stohl, A., Forster, C., Frank, A., Seibert, P., Wotawa, G., 2005. Technical note: The Lagrangian particle dispersion model FLEXPART version 6.2. Atmospheric chemistry and physics 5, 2461–2474.

Stohl, A., Wotawa, G., Seibert, P., Kromp-Kolb, H., 1995. Interpolation errors in wind fields as a function of spatial and temporal resolution and their impact on different types of kinematic trajectories. Journal of Applied Meteorology 34 (10), 2149–2165.

Sulman, M. H., Huntley, H. S., Lipphardt Jr, B., Kirwan Jr, A., 2013. Leaving flatland: Diagnostics for lagrangian coherent structures in three-dimensional flows. Physica D: Nonlinear Phenomena.

Tallapragada, P., 2010. Identifying dynamical boundaries and phase space transport using Lagrangian coherent structures. Ph.D. thesis, Virginia Tech.

Tallapragada, P., Ross, S. D., 2013. A set oriented definition of finite-time Lyapunov exponents and coherent sets. Communications in Nonlinear Science and Numerical Simulation 18, 1106–1126.

Tallapragada, P., Ross, S. D., Schmale, D. G., 2011. Lagrangian coherent structures are associated with fluctuations in airborne microbial populations. Chaos 21, 033122.

Thomson, D., 1987. Criteria for the selection of stochastic-models of particle trajectories in turbulent flows. Journal of Fluid Mechanics 180, 529–556.

Toth, Z., Kalnay, E., 1997. Ensemble forecasting at NCEP and the breeding method. Monthly Weather Review 125 (12), 3297–3319.

Uppala, S. M., Kållberg, P., Simmons, A., Andrae, U., Bechtold, V., Fiorino, M., Gibson, J., Haseler, J., Hernandez, A., Kelly, G., et al., 2005. The ERA-40 re-analysis. Quarterly Journal of the Royal Meteorological Society 131 (612), 2961–3012.

Vogel, B., Pan, L. L., Konopka, P., Guenther, G., Mueller, R., Hall, W., Campos, T., Pollack, I., Weinheimer, A., Wei, J., Atlas, E. L., Bowman, K. P., 2011. Transport pathways and signatures of mixing in the extratropical tropopause region derived from lagrangian model simulations. Journal of Geophysical Research-Atmospheres 116, .

Wilson, J., Sawford, B., 1996. Review of Lagrangian stochastic models for trajectories in the turbulent atmosphere. Boundary-layer meteorology 78, 191–210.