



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



Amir E. BozorgMagham<sup>a,\*</sup>, Shane D. Ross<sup>a</sup>, David G. Schmale III<sup>b</sup>

<sup>a</sup> Engineering Science and Mechanics, Virginia Tech, Blacksburg, VA, USA

<sup>b</sup> Plant Pathology, Physiology, and Weed Science, Virginia Tech, Blacksburg, VA, USA

## HIGHLIGHTS

- We study the atmospheric Lagrangian coherent structures (LCSs).
- We study the effects of errors of forecast velocity on the finite time Lyapunov exponent (FTLE) field and associated LCSs.
- We propose five methods for comparing forecast- and archive-based FTLE–LCSs.
- We quantify the accuracy of predictions of LCSs.
- We quantify the sensitivity of such predictions with respect to the involved forecasting parameters.

## ARTICLE INFO

### Article history:

Received 25 July 2012

Received in revised form

2 March 2013

Accepted 3 May 2013

Available online 16 May 2013

Communicated by H.A. Dijkstra

### Keywords:

Lagrangian coherent structure

Atmospheric transport

Airborne microorganisms

Uncertain forecast data

FTLE field

Pattern correlation

## 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 the 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.

© 2013 Elsevier B.V. All rights reserved.

## 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 [1–7]. 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 [8] 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 [9–11]. 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 US [12,13], and long distance transport is believed to play a role in the dispersal of tobacco blue mold in the US [14]. Experimental verification of long distance transport of airborne microorganisms from known sources is underway [15]. 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

\* Corresponding author. Tel.: +1 3023672495.

E-mail addresses: [aebm@vt.edu](mailto:aebm@vt.edu), [bozorgmagham@gmail.com](mailto:bozorgmagham@gmail.com) (A.E. BozorgMagham).

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 [9]. 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 [16,3]. 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* [10]. This group of fungi contains a number of important plant and animal pathogens [17]. 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 [9].

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 [18], 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 [19].

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 [20].

The paper is outlined as follows. In Section 2, we discuss the conceptual and mathematical definitions of the FTLE field, ridges and hyperbolic repelling and attracting LCSs. In Section 3, we discuss the time scheduling of real-time extraction of hyperbolic LCSs from NAM-218 data set. In Section 4, we study the errors of wind field forecasts as the main cause of imprecise FTLE–LCS predictions. In Section 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 [21–23], composite correlation filter [24] and modified Hausdorff distance method [25]). In Section 6, we make conclusion of the comparison methods and the results from each of them.

## 2. LCS computation and conceptual motivation

In the description below, we follow the notation and terminology of [5]. Consider a velocity field of the form

$$\dot{\mathbf{x}} = \mathbf{v}(\mathbf{x}, t), \quad \mathbf{x} \in U \subset \mathbb{R}^n. \quad (1)$$

At each instant of time  $t$ , a trajectory of the system (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  $t_0$  into the position at time  $t$  advected under the flow,

$$\phi_{t_0}^t : \mathbf{x}_0 \rightarrow \mathbf{x}(t, t_0, \mathbf{x}_0). \quad (2)$$

We use the deformation gradient (Jacobian)

$$D\phi_{t_0}^t(\mathbf{x}_0) = \left. \frac{d\phi_{t_0}^t(\mathbf{x})}{d\mathbf{x}} \right|_{\mathbf{x}=\mathbf{x}_0}, \quad (3)$$

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

$$C_{t_0}^t(\mathbf{x}_0) = D\phi_{t_0}^t(\mathbf{x})^* D\phi_{t_0}^t(\mathbf{x})|_{\mathbf{x}=\mathbf{x}_0} \quad (4)$$

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

$$\begin{aligned} \sigma_{t_0}^t(\mathbf{x}_0) &= \frac{1}{|t - t_0|} \log \|D\phi_{t_0}^t(\mathbf{x}_0)\| \\ &= \frac{1}{|t - t_0|} \log \sqrt{\lambda_{\max}[C_{t_0}^t(\mathbf{x}_0)]} \end{aligned} \quad (5)$$

where  $\|\cdot\|$  is the norm operator and  $\lambda_1 < \lambda_2 < \dots < \lambda_{n-1} < \lambda_n = \lambda_{\max}$  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. [5] and Lekien et al. [4] defined Lagrangian coherent structures as the ridges of the FTLE field. Later, Haller and [8] showed that the ridges of the FTLE field are hyperbolic Lagrangian coherent structures only if they satisfy additional criteria. Based on Haller [8] and Karrasch [26], for a fixed time interval  $[t_0, t]$  a compact hypersurface  $R(t_0) \subset U$  is defined as an FTLE ridge if for all  $\mathbf{x}_0 \in R(t_0)$  we have

$$\begin{aligned} \langle D\lambda_{\max}(\mathbf{x}_0, t_0, t), \xi_{\max}(\mathbf{x}_0, t_0, t) \rangle &= 0 \\ \langle \xi_{\max}(\mathbf{x}_0, t_0, t), D^2\lambda_{\max}(\mathbf{x}_0, t_0, t)\xi_{\max}(\mathbf{x}_0, t_0, t) \rangle &< 0 \end{aligned} \quad (6)$$

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

Provided that  $R(t_0)$  is an FTLE ridge (6) at the initial time, Haller [8] showed that the sufficient and necessary conditions for  $R(t) = \phi_{t_0}^t(R(t_0))$  to be a hyperbolic repelling LCS during the interval  $[t_0, t]$  are

$$\begin{aligned} \lambda_{n-1}(\mathbf{x}_0, t_0, t) &\neq \lambda_{\max}(\mathbf{x}_0, t_0, t) > 1 \\ \xi_{\max}(\mathbf{x}_0, t_0, t) &\perp T_{\mathbf{x}_0}R(t_0) \\ \mu^* \mathbf{L}(\mathbf{x}_0, t_0, t)\mu &> 0 \end{aligned} \quad (7)$$

where  $T_{\mathbf{x}_0}R(t_0)$  is the tangent space of  $R(t_0)$ ,  $\mathbf{L}$  is a matrix defined in Haller [8], and  $\mu$  is any non-zero column vector of real numbers (the positive definite condition for  $\mathbf{L}$ ). Note that there is no guarantee that all the existing hyperbolic features are identified by this approach.

Later, Karrasch [26] showed that if eigenvectors of  $C_{t_0}^t(\mathbf{x}_0)$  are differentiable (which is naturally observed in smooth enough velocity fields) then (7) simplifies to

$$\begin{aligned} \lambda_{n-1}(\mathbf{x}_0, t_0, t) &\neq \lambda_{\max}(\mathbf{x}_0, t_0, t) > 1 \\ \xi_{\max}(\mathbf{x}_0, t_0, t) &\perp T_{\mathbf{x}_0}R(t_0). \end{aligned} \quad (8)$$

We use the sets of Eqs. (6) and (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  $[t_0, t]$  [8]. 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* [27,28,9,10]. Fig. 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.

### 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.

#### 3.1. NAM-218 data set

Computing the FTLE field requires the flow map, (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  $614 \times 428$  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  $(\phi, \lambda)$  on the sphere to a Cartesian coordinate [29]. We consider a  $6230 \times 4670$  km rectangular area as our wind velocity field and we compute the LCSs over an interrogation zone of  $1000 \times 1000$  km size with  $256 \times 256$  grid points. We consider the sampling area (Virginia Tech's Kentland Farm near Blacksburg, VA) as the center of that square ( $37^\circ 11'$  N latitude and  $80^\circ 35'$  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 assimilation–simulation 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 [30]. In NAM-218 data set, the temporal resolution of archived data is 6 h 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 h and at each moment, the maximum available forecast data is for 84 h lead time.

#### 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 illustrates the details of such a schedule for a typical day. For our sampling purposes, LCSs are desired in the time interval 08:00–17:00 local time (Eastern Daylight Time (EDT)), equivalent to 12:00–21:00 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).

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. 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, the parameters that control the forecasting procedure are gap time,  $T^*$ , and integration time,  $T = t - t_0$ , following the notation of Section 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 Section 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 [31,32] 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.

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  $(T, T^*)$ .

#### 3.3. Velocity (wind) field data structure

Based on the previous work [33–37,9], 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 [38]). We make an isobaric approximation, considering only the components

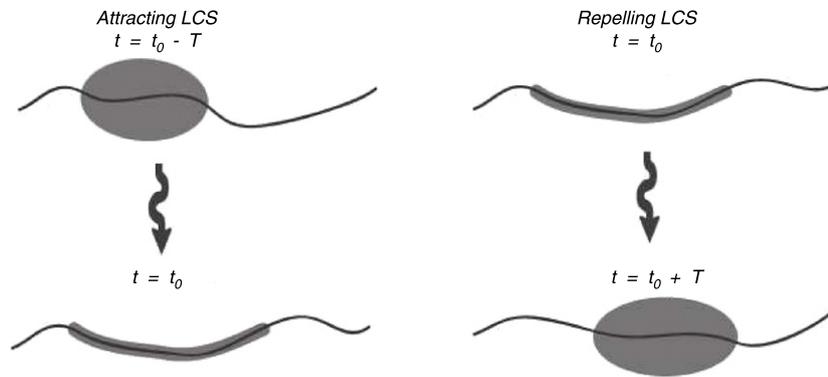


Fig. 1. Schematic drawing of hyperbolic attracting and repelling LCSs.

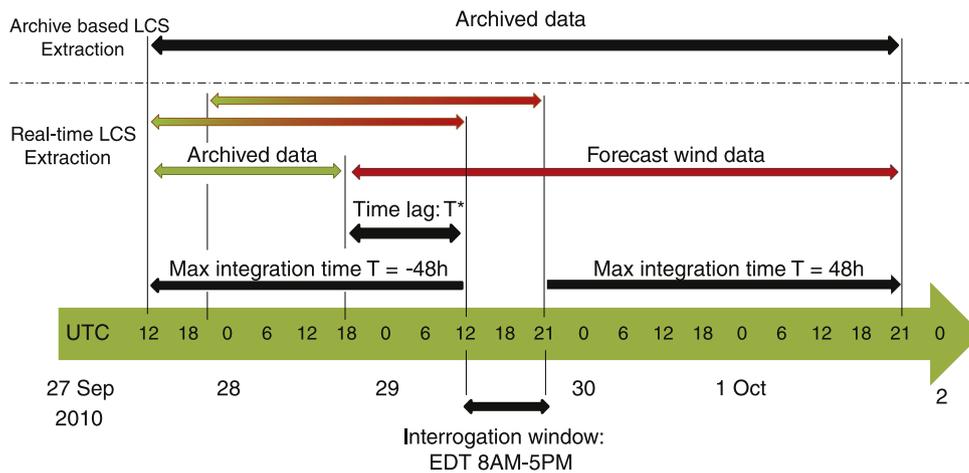


Fig. 2. Schematic time schedule for calculation of real-time and archive-based FTLE-LCSs with respect to NAM-218 timetable. Archive-based features are calculated from pure archive data. Real-time FTLE-LCSs are calculated based on a combination of forecast and archive data. Interrogation window (12:00 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 min in interrogation window in forward and backward directions.

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  $\Delta p$  to be similar over integration times  $T = \pm T_{\max}$  [39]. To illustrate the similarity of neighboring pressure levels, in Fig. 3 the backward FTLE field for 900 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.

#### 4. Forecast- and archive-based FTLE-LCSs

The main interest in prediction of FTLE-LCSs is based on the results, e.g., [9] 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 min, so during 9 h, the total number of FTLE-LCS snapshots is 37.

The upper panels of Fig. 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  $(T, T^*) = (-24$  h, 18 h) and conditions (6) and (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.

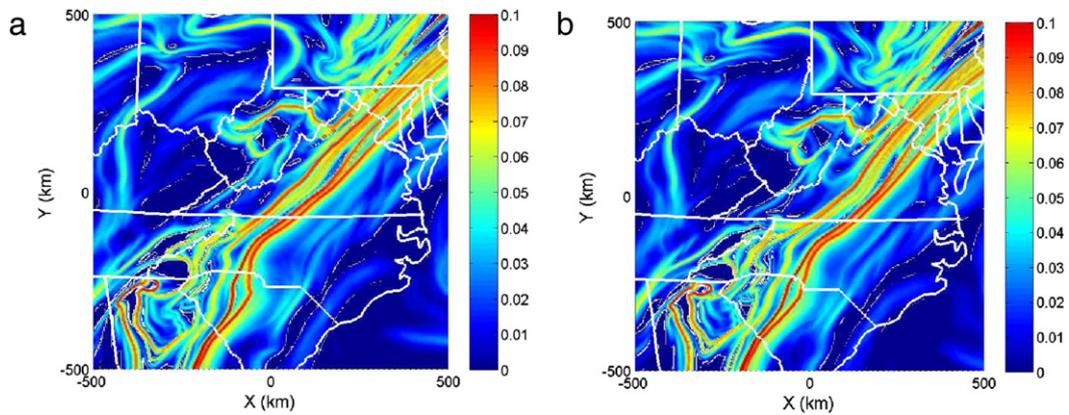
The lower panels of Fig. 4 show snapshots of the forecast- (left) and archive-based (right) LCSs at 21:00 UTC with the same calculation parameters as the upper panels.

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 Section 5 by applying different hybrid Eulerian-Lagrangian approaches.

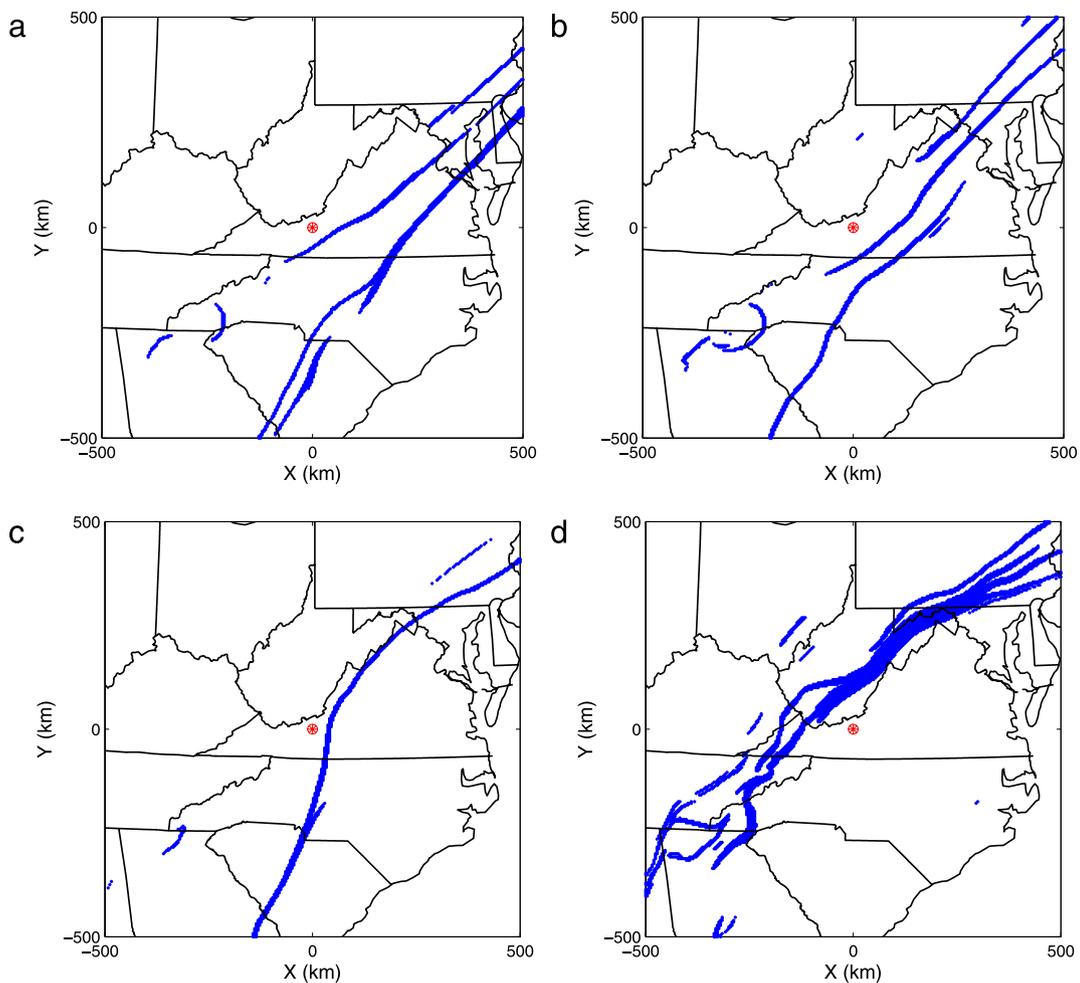
##### 4.1. Sources of errors

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



**Fig. 3.** Eastern US 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  $(T, T^*) = (-48 \text{ h}, 18 \text{ h})$ , vertical color bar indicates FTLE magnitude ( $\text{h}^{-1}$ ). Note the similarity. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

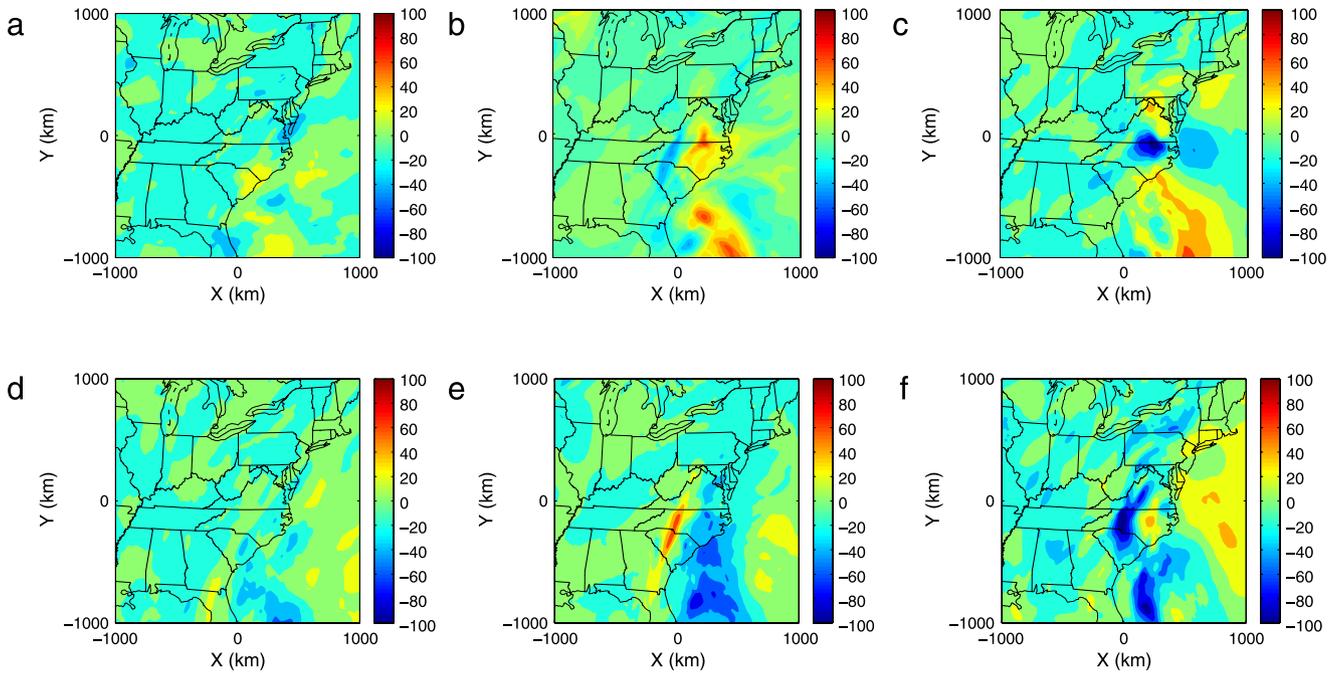


**Fig. 4.** Eastern US with the state of Virginia and Kentland farm (sampling location) in the center (red point). (a, c) forecast- and (b, d) archive-based hyperbolic attracting LCSs. Upper panels: 12:00 UTC; lower panels: 21:00 UTC 29 Sep 2010.  $(T, T^*) = (-24 \text{ h}, 18 \text{ h})$ . In this 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)). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

by the NAM model is 6 h (considering short-term forecast, temporal resolution would be 3 h) 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 [40,41] 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.



**Fig. 5.** Error of forecasting (km/h) 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.

Recalling Fig. 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 [31,32]. In this paper, the goal is not to study the limits of forecasting in the presence of deterministic or stochastic chaos, but merely to quantify the effects of existing forecast errors on the calculated FTLE–LCSs.

#### 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). 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. 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.

Calculated LCS features from forecast data should resemble the true features by a maximum distance of  $\Delta$  (Eq. 19 of [18]) if forecast errors are localized in time. The existence of large and prolonged high error regions, as shown in Fig. 5, reveals 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. 6, panels (a) and (b), respectively) [42]. 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,

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

where  $\sigma = (2\pi)^{-1/2}$ ,  $\mu = 0$  and  $\chi$  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 ( $\eta$ ) over the entire domain as the mean value of this normalized relative error field,

$$\eta = \frac{1}{N} \sum_{i,j} \exp(-\pi \chi_{i,j}^2) \quad (9)$$

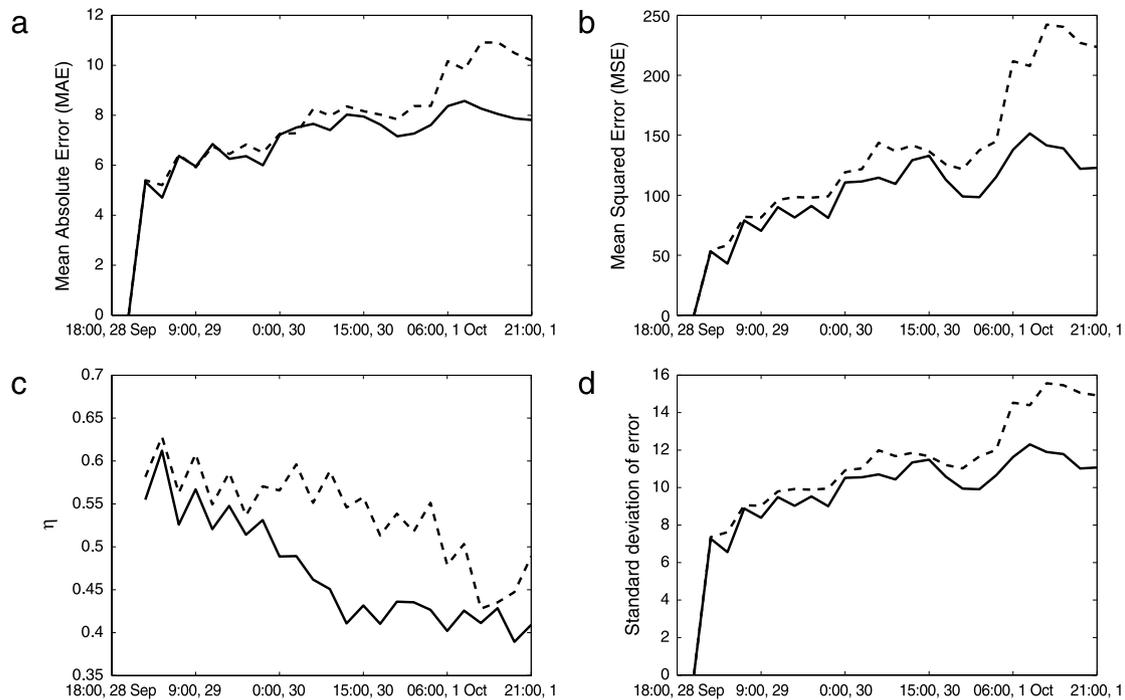
where  $N$  is the total number of grid points.

Fig. 6(c) shows the quality of forecasting ( $\eta$ ) decreasing with time (corresponding to the normalized field). In addition, Fig. 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).

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.

## 5. Results

Referring to the differences between forecast- and archive-based LCS features, e.g., Fig. 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



**Fig. 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).

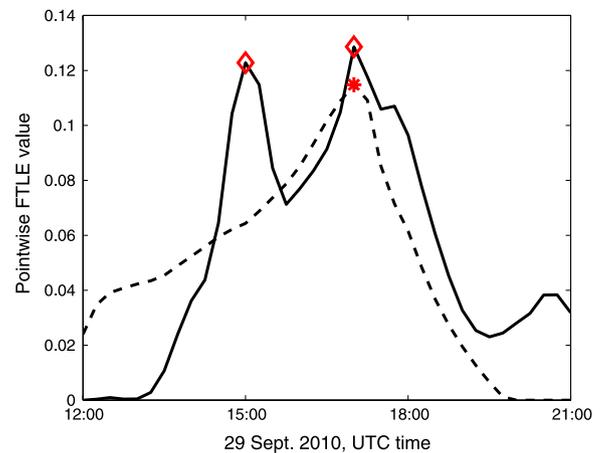
other words, how well do the forecast LCSs describe the real case of LCSs passages, particularly the passage times? Also, regarding limitations on availability of necessary data for predicting the LCSs, what is the best choice for effective parameters such as the integration and gap times?

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 min 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.

### 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. 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.

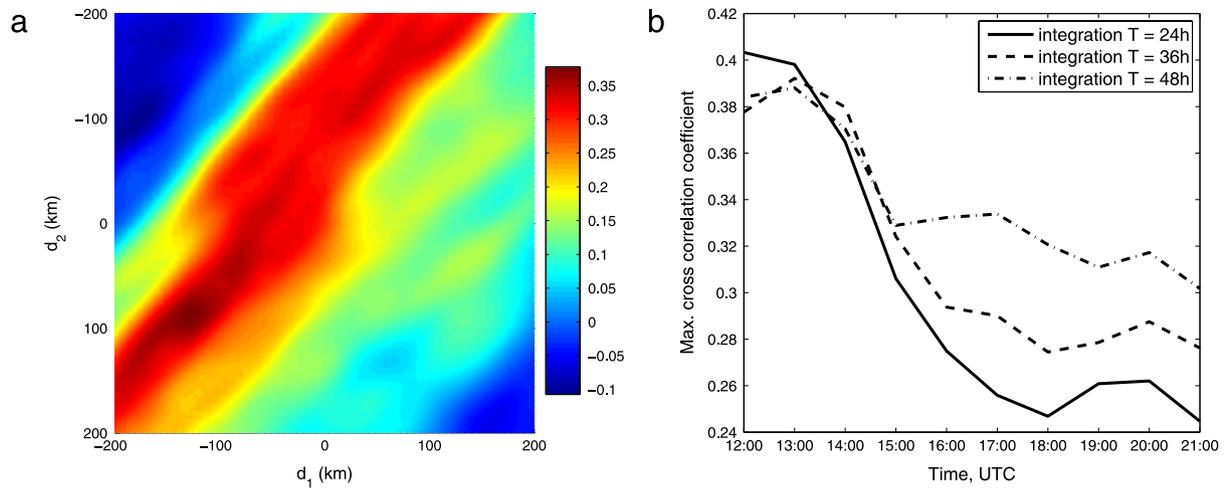
Results from several simulations for different days suggest a rule of thumb which may be beneficial for early warning systems, e.g., for vast crop fields where the exact passage times of LCSs are not important, but it is important for one to know if hyperbolic LCSs pass over a region, since these have been associated with fluctuations in microbial populations [9]. 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



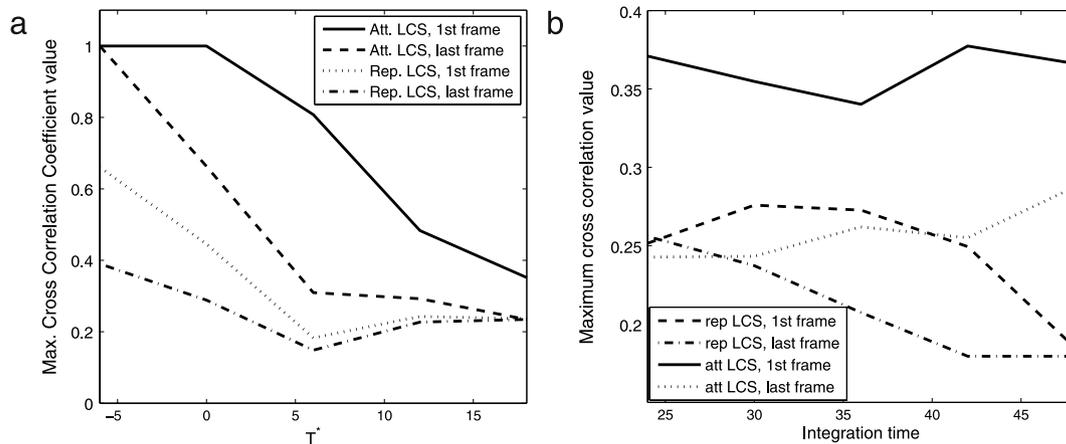
**Fig. 7.** Forecast- (dashed) and archive-based (solid) FTLE value at the sampling location.  $T = -24$  h and  $T^* = 18$  h. Corresponding hyperbolic LCSs are shown by asterisk and diamonds.

threshold we have determined by observation is  $0.07 \text{ h}^{-1}$ , i.e., if the peak forecast FTLE is  $>0.07 \text{ h}^{-1}$ , 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. 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. 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. 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. 4, panel (b)), so we record two peaks.



**Fig. 8.** (a) Maximum cross correlation coefficient between archive and forecast attracting LCSs at 12:00 UTC 29 Sep 2010, considering 2D shift ( $\pm 200$  km in both  $X$  and  $Y$ ),  $(T, T^*) = (-36$  h, 18 h). (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.



**Fig. 9.** Comparing the forecast- and archive-based FTLE–LCSs, studying the effect of (a)  $T^*$  when  $T = 24$  h and (b) integration time ( $T$ ) on the maximum value of cross correlation coefficient when  $T^* = 18$  h.

## 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 [43]. Considering 2D shifts and discrete data points, we use

$$r(d_1, d_2) = \frac{\sum_m \sum_n [(\sigma_{m-d_1, n-d_2} - \bar{\sigma}) (\psi_{m, n} - \bar{\psi})]}{\sqrt{\sum_m \sum_n [(\sigma_{m-d_1, n-d_2} - \bar{\sigma})^2 (\psi_{m, n} - \bar{\psi})^2]}} \quad (10)$$

to get a 2D surface of cross-correlation coefficients, where  $\sigma$  and  $\psi$  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  $\sigma$  and  $\psi$ , and  $d_1$  and  $d_2$  represent the shift in the  $X$  and  $Y$  directions, respectively.

Fig. 8(a) shows the cross correlation coefficient between forecast and archived attracting LCS features at 12:00 UTC 29 Sep 2010, for  $(T, T^*) = (-36$  h, 18 h). In this figure, we consider shifts up to  $\pm 200$  km 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. 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. 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. 9(a) shows the typical effect of  $T^*$  on the quality of forecasting for attracting and 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. 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. 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.

From Figs. 8 and 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) [24].

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

$$\bar{\sigma}_{t_0}^{t_0+T}(\mathbf{x}_0) = \zeta_{T^*}(\sigma_{t_0}^{t_0+T}(\mathbf{x}_0)) \quad (11)$$

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  $\zeta_{T^*} : \mathcal{F}(U, \mathbb{R}) \rightarrow \mathcal{F}(U, \mathbb{R})$  is a one-parameter family of diffeomorphisms on the function space  $\mathcal{F}(U, \mathbb{R})$ , with parameter  $T^*$ , where  $\zeta_0$  is the identity.

Spatial shifts are only one limited possibility for  $\zeta_{T^*}$ . In the next section, we use the POD method to enable us to consider more general  $\zeta_{T^*}$  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.

### 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 [21–23]. 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

$$z(\mathbf{x}, t) \approx \sum_{i=1}^K a_i(t) \Phi_i(\mathbf{x}), \quad (12)$$

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

$$\int_{\mathbf{x}} \Phi_i \Phi_j d\mathbf{x} = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j. \end{cases} \quad (13)$$

Using Eq. (13), we find the time-varying amplitudes of the mode shapes as

$$a_i(t) = \int_{\mathbf{x}} z(\mathbf{x}, t) \Phi_i(\mathbf{x}) d\mathbf{x}. \quad (14)$$

Generally, finding the  $\Phi$ 's depends on our choice of basis functions (e.g. Fourier series with orthogonal trigonometric functions). In the case of the POD method, the  $\Phi$ 's are chosen such that the approximation of (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

$$M = U \Sigma V^*, \quad M \in \mathbb{R}^{m \times n} \quad (15)$$

where  $U$  is  $m \times m$ ,  $\Sigma$  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  $U$ ,  $\Sigma$  and  $V$  would be real, also  $U$  and  $V$  are unitary matrices. The diagonal elements of  $\Sigma$  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. (15) as

$$M = QV^* = \sum_{i=1}^n q_i v_i^* \quad (16)$$

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.

#### 5.3.1. 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 (12) where the required coefficients can be generated by the inner product of input data  $M$  and the related mode shape as described by (14). Recalling that the importance of the mode shapes depends on the magnitude of their corresponding singular value, Fig. 10 shows the first and the second mode shapes for both the forecast- and archive-based FTLE field, where  $(T, T^*) = (-24 \text{ h}, 18 \text{ h})$ . 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 value.

Fig. 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  $(|T|, T^*) = (24 \text{ h}, 18 \text{ h})$ . 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.

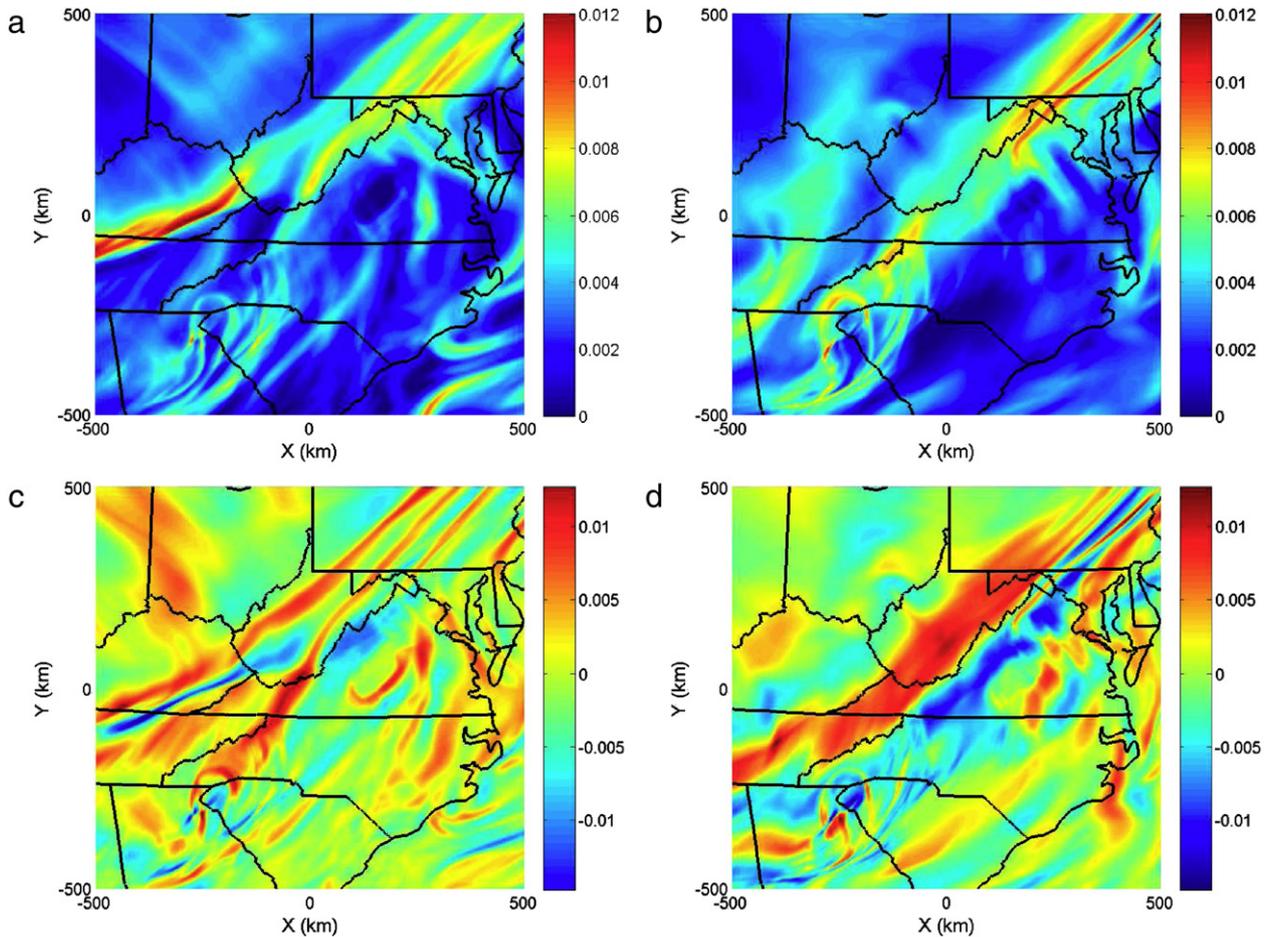
#### 5.3.2. Effect of parameters on the quality of forecasting using POD modes

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.

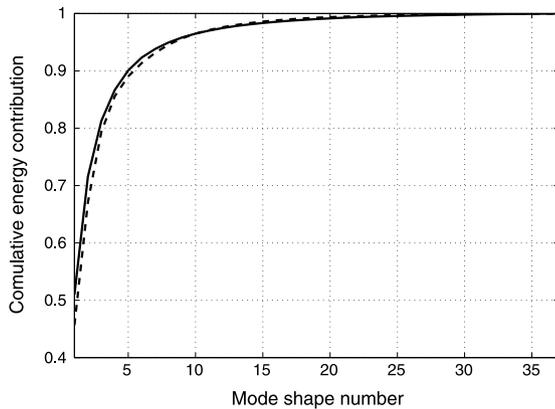
Fig. 12 shows the effect of integration time  $T$  on the measure of similarity of mode shapes of backward and forward FTLE fields when we fix the  $T^*$  at zero (perfect real-time forecast case) and let the integration time  $|T|$  be 24 or 48 h.

We observe that for backward FTLE, increasing  $T$  from 24 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



**Fig. 10.** First and second mode shapes for the (a, c) forecast- and (b, d) archive-based backward FTLE fields,  $(T, T^*) = (-24 \text{ h}, 18 \text{ h})$ .



**Fig. 11.** Cumulative contribution of the mode shapes, backward FTLE (dashed line), forward FTLE (solid line),  $(T, T^*) = (48, 18)$ .

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 coefficient for the first backward FTLE mode shape is close but not exactly one. Recall that mode shapes encapsulate information from *all* time snapshots, so as later snapshots 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 h which is the *worst case* of real-time forecasting we have considered. Fig. 13

shows the value of cross correlation coefficients for different mode shapes.

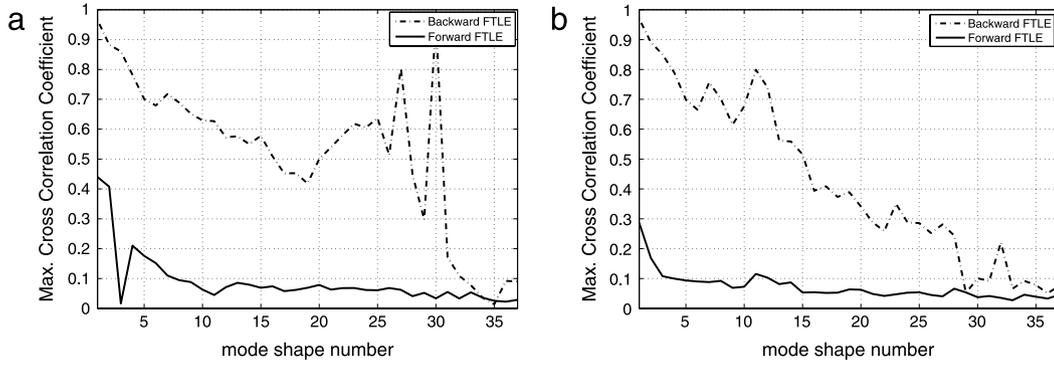
Considering the forward FTLE (repelling LCS features), Figs. 13 and 12 show that when we decrease the integration time, we generally get better results regarding the correlation coefficients, i.e., for shorter integration times  $T$ , cumulative errors of predicted trajectories will be smaller. While 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 Section 5.2, we 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 Section 5.4 we use a type of composite correlation filter to overcome this weakness.

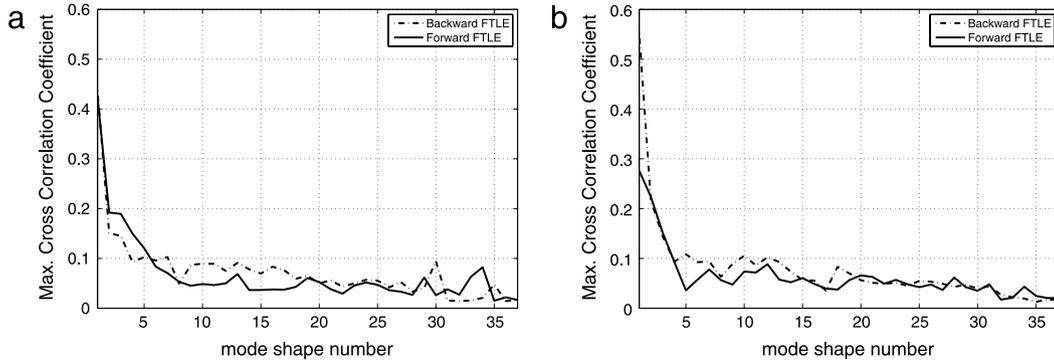
#### 5.4. Composite correlation filter

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 [24,44]. 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,



**Fig. 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.



**Fig. 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.

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

$$\mathbf{h} = \mathbf{X}(\mathbf{X}^* \mathbf{X})^{-1} \mathbf{u} \quad (17)$$

where  $\mathbf{X}$  represents the all training images collected together,

$$\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n], \quad (18)$$

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  $\mathbf{u} = [u_1, u_2, \dots, u_n]$  is an  $n \times 1$  vector containing the chosen peak values for the training frames.

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

$$\lambda_i = \mathbf{Y}_i^* \mathbf{h}. \quad (19)$$

Fig. 14 shows an example of applying this method for attracting LCSs associated with different  $T^*$  and  $T_s$ . 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  $-24$  and  $-48$  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.

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.

### 5.5. Modified Hausdorff distance

Hausdorff distance is an extension to the Euclidean metric [45]. 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 [25]. By using this method one can calculate the distance between LCS features.

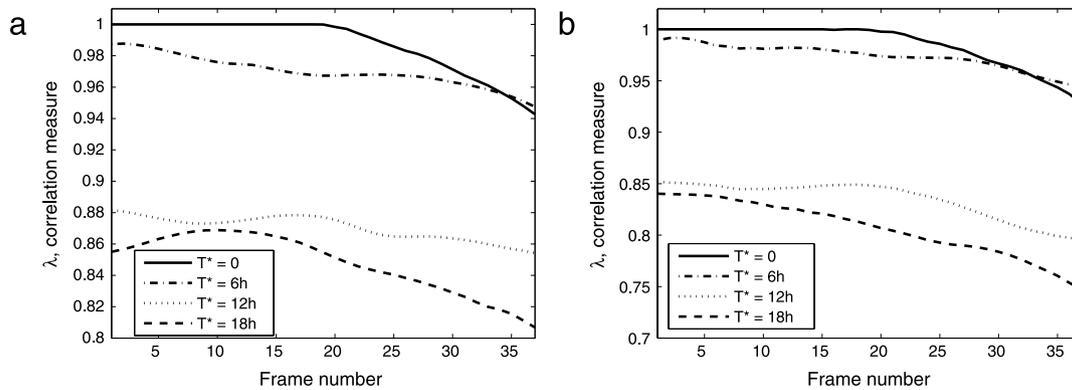
We denote the Euclidean distance between two points  $\alpha$  and  $\beta$  as  $d(\alpha, \beta) = \|\alpha - \beta\|$ . The distance between a single point ( $\alpha$ ) and a set of points  $\mathfrak{B} = \{\beta_1, \beta_2, \dots, \beta_n\}$  is defined as  $d(\alpha, \mathfrak{B}) = \min_{\beta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

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

$$\mathfrak{D} = \max(d(\mathfrak{A}, \mathfrak{B}), d(\mathfrak{B}, \mathfrak{A})), \quad (20)$$

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

$$d(\mathfrak{A}, \mathfrak{B}) = \frac{1}{m} \sum_{\alpha \in \mathfrak{A}} d(\alpha, \mathfrak{B}). \quad (21)$$



**Fig. 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  $T^* = 0, 6, 12$  and  $18$  h.

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 (6) and (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.

Fig. 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.

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  $(T^*, T) = (6 \text{ h}, -48 \text{ h})$  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.

## 6. Discussion

This study was motivated by recent observations suggesting that LCSs govern the large-scale atmospheric motion of airborne microorganisms [9,10]. 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 [18]. 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.

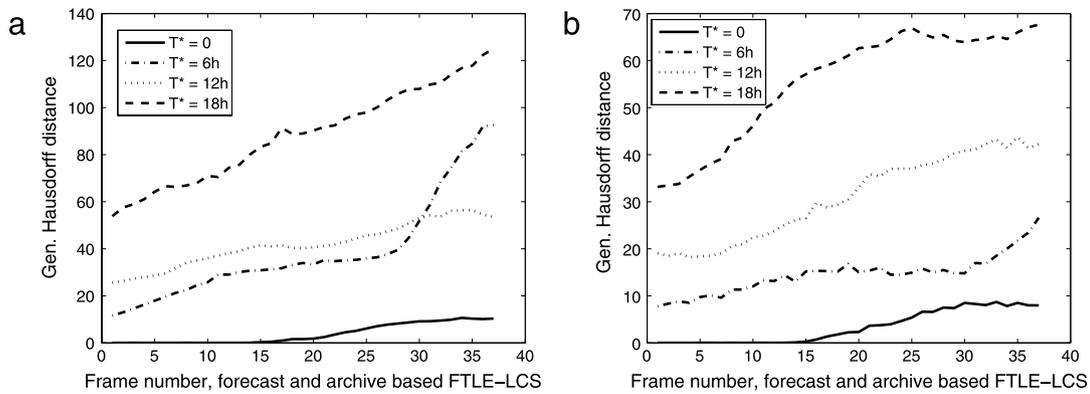
In this paper, we considered five methods for comparing the forecast- and archive-based FTLE fields: (1) fixed spatial point with temporal comparison, (2) fixed time with spatial cross-correlation, (3) spatiotemporal POD mode cross-correlations, (4) composite correlation filter and (5) modified Hausdorff 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 [16]. 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).



**Fig. 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  $T^* = 0, 6, 12$  and  $18$  h.

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.

### References

- [1] F. Lekien, C. Coulliette, A. Mariano, E. Ryan, L. Shay, G. Haller, J. Marsden, Pollution release tied to invariant manifolds: a case study for the coast of florida, *Physica D* 210 (2005) 1–20.
- [2] G. Haller, A. Poje, Finite time transport in aperiodic flows, *Physica D* 119 (1998) 352–380.
- [3] G. Haller, G. Yuan, Lagrangian coherent structures and mixing in two-dimensional turbulence, *Physica D* 147 (2000) 352–370.
- [4] F. Lekien, S.C. Shadden, J.E. Marsden, Lagrangian coherent structures in  $n$ -dimensional systems, *J. Math. Phys.* 48 (2007) 065404.
- [5] S.C. Shadden, F. Lekien, J.E. Marsden, Definition and properties of Lagrangian coherent structures from finite-time Lyapunov exponents in two-dimensional aperiodic flows, *Physica D* 212 (2005) 271–304.
- [6] M.J. Olascoaga, G. Haller, Forecasting sudden changes in environmental pollution patterns, *Proc. Natl. Acad. Sci.* 109 (2012) 4738–4743.
- [7] B. Vogel, L.L. Pan, P. Konopka, G. Guenther, R. Mueller, W. Hall, T. Campos, I. Pollack, A. Weinheimer, J. Wei, E.L. Atlas, K.P. Bowman, Transport pathways and signatures of mixing in the extratropical tropopause region derived from Lagrangian model simulations, *J. Geophys. Res.* -Atmos. 116 (2011).
- [8] G. Haller, A variational theory of hyperbolic Lagrangian coherent structures, *Physica D* 240 (2011) 574–598.
- [9] P. Tallapragada, S.D. Ross, D.G. Schmale III, Lagrangian coherent structures are associated with fluctuations in airborne microbial populations, *Chaos* 21 (2011) 033122.
- [10] D. Schmale III, S. Ross, T. Fetters, P. Tallapragada, A. Wood-Jones, B. Dingus, Isolates of *Fusarium graminearum* collected 40–320 m above ground level cause fusarium head blight in wheat and produce trichothecene mycotoxins, *Aerobiologia* 28 (2012) 1–11.
- [11] B. Lin, A. Bozorgmagham, S.D. Ross, D.G. Schmale III, Small fluctuations in the recovery of fusaria across consecutive sampling intervals with unmanned aircraft 100 m above ground level, *Aerobiologia* 29 (1) (2013) 45–54.
- [12] Z. Pan, X.B. Yang, S. Pivonia, L. Xue, R. Pasken, J. Roads, Long-term prediction of soybean rust entry into the continental united states, *Plant Dis.* 90 (2006) 840–846.

- [13] R.W. Schneider, C.A. Hollier, H.K. Whitam, M.E. Palm, J.M. McKemy, J.R. Hernandez, L. Levy, R. DeVries-Paterson, First report of soybean rust caused by *Phakopsora pachyrhizi* in the continental United States, *Plant Dis.* 89 (2005) 774–774.
- [14] D. Aylor, G. Taylor, G. Raynor, Long range transport of tobacco blue mold spores, *Agric. Meteorol.* 27 (1982) 217–232.
- [15] A.J. Prussin, Q. Li, R. Malla, S.D. Ross, D.G. Schmale III, Monitoring the long distance transport of *Fusarium graminearum* from field-scale sources of inoculum, 2013. submitted for publication.
- [16] G. Haller, T. Sapsis, Lagrangian coherent structures and the smallest finite-time Lyapunov exponent, *Chaos* 21 (2011) 023115.
- [17] J.F. Leslie, B.A. Summerell, *The Fusarium Laboratory Manual*, Blackwell Publishing, 2006.
- [18] G. Haller, Lagrangian coherent structures from approximate velocity data, *Phys. Fluids* 14 (2002) 1851–1861.
- [19] A.B. Olcay, T.S. Pottebaum, P.S. Krueger, Pollution release tied to invariant manifolds: a case study for the coast of florida, *Chaos* 20 (2010) 017506.
- [20] L.V. Madden, M. Wheelis, The threat of plant pathogens as weapons against US crops, *Annual Review of Phytopathology* 41 (2003) 155–176.
- [21] A. Antoulas, An overview of approximation methods for large-scale dynamical systems, *Ann. Rev. Control* 29 (2005) 181–190.
- [22] L. Sirovich, Turbulence and the dynamics of coherent structures, *Quart. Appl. Math.* 45 (1987) 561–571.
- [23] C.W. Rowley, Model reduction for fluids, using balanced proper orthogonal decomposition, *Internat. J. Bifur. Chaos* 15 (2005) 997–1013.
- [24] V.B. Kumar, A. Mahalanobis, R.D. Juday, *Correlation Pattern Recognition*, Cambridge, 2005.
- [25] M.-P. Dubuisson, A.K. Jain, A modified hausdorff distance for object matching, in: *Proceedings of the 12th IAPR International Conference*, Vol. 1, IEEE, 1994, pp. 566–568.
- [26] D. Karrasch, Comment on “a variational theory of hyperbolic Lagrangian coherent structures, *Physica D* 240 (2011) 574–598”, *Physica D* 241 (2012) 1470–1473.
- [27] F. Lekien, S.D. Ross, The computation of finite-time Lyapunov exponents on unstructured meshes and for non-euclidean manifolds, *Chaos* 20 (2010) 017505.
- [28] C. Senatore, S.D. Ross, Detection and characterization of transport barriers in complex flows via ridge extraction of the finite time Lyapunov exponent field, *Internat. J. Numer. Methods Engrg.* 86 (2011) 1163–1174.
- [29] J.P. Snyder, *Map Projections—A Working Manual*, US Government Printing Office, 1987.
- [30] E. Kalnay, *Atmospheric Modelling, Data Assimilation and Predictability*, Cambridge, 2003.
- [31] E.N. Lorenz, Deterministic nonperiodic flow, *J. Atmospheric Sci.* 20 (1963) 475–480.
- [32] T. Palmer, Predicting uncertainty in forecasts of weather and climate, *Rep. Progr. Phys.* 63 (2000) 71–116.
- [33] P.L. Haagenson, K. Gao, Y. Kuo, Evaluation of meteorological analyses, simulations, and long-range transport using an atmospheric tracer data, *J. Appl. Meteorol.* 29 (1990) 1268–1283.
- [34] J.D. Kahl, J.M. Harris, G.A. Herbert, M.P. Olson, Intercomparison of three long-range trajectory models applied to arctic haze, *Tellus Ser. B* 41 (1989) 524–536.
- [35] A. Stohl, Computation, accuracy and applications of trajectories—a review and bibliography, *Atmos. Environ.* 32 (1998) 947–966.
- [36] J.D. Kahl, P.J. Samson, Uncertainty in trajectory calculations due to low resolution meteorological data, *J. Clim. Appl. Meteorol.* 25 (1986) 1816–1831.
- [37] P. Tallapragada, S.D. Ross, A set oriented definition of finite-time Lyapunov exponents and coherent sets, *Commun. Nonlinear Sci. Numer. Simul.* 18 (2013) 1106–1126.
- [38] A.L. Buerger, J.H. Fike, J.A. Burger, C.R. Feldhake, J.A. McKenna, C.D. Teutsch, Botanical composition and forage production in an emulated silvopasture, *Agronomy J.* 97 (4) (2005) 1141–1147.
- [39] P. Tallapragada, Identifying dynamical boundaries and phase space transport using Lagrangian coherent structures. Ph.D. Thesis, Virginia Tech, ESM Department, 2010.
- [40] A. Griffa, L. Piterberg, T. Ozgokmen, Predictability of Lagrangian particle trajectories: effects of smoothing of the underlying Eulerian flow, *J. Mar. Res.* 62 (2004) 1–35.

- [41] T. Ozgokmen, A. Griffa, A. Mariano, L. Piterbarg, On the predictability of Lagrangian trajectories in the ocean, *J. Atmos. Ocean Technol.* 17 (2000) 366–383.
- [42] I.T. Jolliffe, D.B. Stephenson, *Forecast Verification*, Wiley, 2003.
- [43] R.C. Gonzalez, R.E. Woods, *Digital Image Processing*, Prentice Hall, 2007.
- [44] C.F. Hester, D. Casasent, Multivariant technique for multiclass pattern recognition, *Appl. Opt.* 19 (1980) 1758–1761.
- [45] T.R. Rockafellar, R.J.-B. Wets, *Variational Analysis*, Springer-Verlag, 2005.