The Effects of Unresolved Scales on Analogue Forecasting Ensembles

Paul Platzer and Bertrand Chapron

1 Introduction

Analogues are nearest neighbors of a system's state according to a given similarity measure. The term "analogue" originally refers to the idea that certain geophysical systems are likely to undergo similar states at distant times, a notion close to the recurrence theorem of Poincaré [1] which states that certain dynamical systems will almost surely return infinitesimally close to any initial condition in the attractor. The term "analogue" was coined by Lorenz [2] who built a proxy of atmospheric predictability from the time-growth of the Euclidean distance between any observed geopotential height field and its closest neighbour (analogue) in other years of record. Since then, analogues have been used in other contexts and can be more broadly understood as similar observations of a system's state occurring at distant times either within a given long trajectory of the system, or within independent simulations of the same system.

More specifically, analogues are found in a database of events often called a "catalogue", such as a long series of observations [3], a reanalysis [4], past deterministic forecasts [5], one long numerical simulation [6], or ensemble simulations [7, 8]. One key advantage of analogue methods is that they are fast and therefore allow a cheap generation of ensembles. As far as forecasting is concerned, analogues have been used for post-processing of deterministic forecasts with "the Analogue Ensemble (AnEn)" [9], in weather generators with the "Analogue Weather Generator (AnaWEGE)" [10], in data assimilation with "the Analog Data Assimilation (AnDA)" [11], and finally for the forecast of dynamical systems with "Kernel Analog Forecasting (KAF)" [12].

P. Platzer (✉) · B. Chapron
Laboratoire d'Océanographie Physique et Spatiale (LOPS), Ifremer, Plouzané, France
e-mail: paul.platzer@ifremer.fr

B. Chapron et al. (eds.), *Stochastic Transport in Upper Ocean Dynamics III*,
Mathematics of Planet Earth 13, https://doi.org/10.1007/978-3-031-70660-8_10

There are several mechanisms driving uncertainties of analogue forecasts. First, analogue forecasting generally assumes that the distance between the analogues and the target[1] image is small, however these distances are never zero [13]. These non-trivial distances introduce forecasting errors. Depending on the nature of the system under study and on the target image, this initial dissimilarity may increase or decrease with time. In a chaotic system, the mismatch in initial conditions grows according to scaling laws given by Lyapunov exponents [14]. This type of deterministic analogue forecasting error was studied thoroughly [15–17].

Second, for real systems one does not have access to the full state of the system, and the similarity measure used to find analogues is bound to be under-representative of the system's complexity. That is, analogues and corresponding distances are computed on observables of the system, not in the system's true state-space. In particular, fine-scale structures of atmospheric or ocean flows are generally unknown, and these may impact the time-evolution of the system. Therefore, even if the analogues were at distance zero to the target initial large-scale image according to the chosen similarity measure, their time-evolution would likely differ from the one of the target. In the present study we focus on this second source of uncertainty, referred to here as the effect of unresolved spatial scales on analogue forecasting ensembles.

In real-life case studies, these two sources of uncertainty are non-linearly combined and hard to disentangle. Here, we use numerical simulations to separate the effect of unresolved spatial scales from the effect of initial analogue-to-target distance. We base our analysis on numerical simulations of a stochastic version [18] of the well-known three-variable chaotic Lorenz system [19]. In this set of stochastic ordinary differential equation, the effects of unresolved spatial scales on the low-order model of Lorenz are rigorously derived from a reformulated material derivative, and expressed as stochastic multiplicative noise terms. This gives a physically consistent yet numerically affordable multi-scale system which can easily be used to understand the effects of unresolved spatial scales on analog forecasting ensembles.

Section 2 details the numerical simulations of the stochastic Lorenz system under location uncertainty as well as the different ensemble forecasting procedures and statistical tests considered in this work. Section 3 outlines numerical experiments and their results. Section 4 draws conclusion and gives perspectives for future work.

[1] The "target", also called "query" in machine-learning terminology, is the (vector of) value(s) of the observed variable(s) of which analogues are sought for. In the example of this article, the target is the three-dimensional vector of values of the coordinates of the stochastic Lorenz at a given time t.

2 Methods

2.1 Lorenz System Under Location Uncertainty

Flow modelling under location uncertainty introduces errors through the assumption that Lagrangian fluid particle displacement results from a smooth deterministic component and another component which is uncorrelated in time but correlated in space [20, 21]. This framework allows to derive the Lorenz system under location uncertainty [18] (hereafter "stochastic Lorenz") as a stochastic representation of the classical Lorenz-63 model [19] (hereafter "deterministic Lorenz") describing incompressible flow undergoing Rayleigh–Bénard convection in a simplified 2D atmospheric convection scenario.

Traditionally, large-scale geophysical flow representations rely on Reynolds decomposition and eddy viscosity models, which dissipate energy without considering local backscattered energy or inhomogeneous turbulence. In contrast, the stochastic approach from which we borrow decomposes the flow into a large-scale, smooth component and a small-scale, fast oscillating velocity component modeled as a random field. It is derived rigorously from physical conservation principles and offers efficient exploration of the attractor at low computational cost [18]. The derivation of the stochastic Lorenz follows closely the original derivation of the deterministic Lorenz, except that the location uncertainty framework allows to express the effects of truncated higher-order spatial modes (i.e., small spatial scales) on the dominant large-scale modes through multiplicative stochastic terms. Therefore, the noise terms in the model represent the effects of small spatial scales on the three large-scale variables, and will be loosely referred to as "small-scale variables" in the following.

Stochastic differential equations for this system are as follows:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = Pra\,(y - x) - \frac{2}{\Upsilon}x\,, \tag{1}$$

$$\mathrm{d}y = \left((\rho - z)x - \left(1 + \frac{2}{\Upsilon}\right)y\right)\mathrm{d}t + \frac{\rho - z}{\Upsilon^{1/2}}\mathrm{d}B_t\,, \tag{2}$$

$$\mathrm{d}z = \left(xy - \left(b + \frac{4}{\Upsilon}\right)z\right)\mathrm{d}t + \frac{y}{\Upsilon^{1/2}}\mathrm{d}B_t\,. \tag{3}$$

It comprises a deterministic differential equation for the velocity variable x and two coupled stochastic differential equations associated with the small-scale temperature fluctuations, written in Itô formulation. Scalar parameters include the ones from the deterministic Lorenz model, which are here set to the usual chaotic values $Pra = 10$, $\rho = 28$, $b = 8/3$, and one parameter that controls noise amplitude, set to $\Upsilon = 10$ (this value is labelled "strong noise" by [18]). The noise terms $\mathrm{d}B_t$ are Wiener processes, which can be viewed heuristically as independent,

identically distributed, centered normal random variables with variance $\mathrm{d}t$, where $\mathrm{d}t$ is the time-increment of the numerical scheme.

Here, this system is numerically simulated with Euler–Maruyama integration for the stochastic differential equations with time step 10^{-5} as in [18].

In the following, we will use a parallel between the non-dimensional time of the stochastic Lorenz system and the typical time of evolution of atmospheric synoptic circulation. We will consider, as in [11], that 0.08 non-dimensional time intervals of the stochastic Lorenz system correspond to 6 hours of atmospheric time. This choice is close to the one of [22] who set non-dimensional time intervals of length-1 of the Lorenz system to correspond to 5 days. This parallel is a rough estimation for comparison purposes and should be interpreted with care.

2.2 Ensemble Forecasts

We place ourselves in the context of the forecast of a "true" trajectory, which corresponds to the numerical integration of the stochastic Lorenz system ((1)–(3)) with time step 10^{-5}, from a given initial condition $\mathbf{x}_t = (x_t, y_t, z_t)$ on the large-scale variables, and with a given sequence of noises $\{\mathrm{d}B_t, \ldots, \mathrm{d}B_{t+h}\}$ from time t to time $t + h$, where h is the forecast horizon.

- True trajectory :

Large-scale variables (initial condition) = $\mathbf{x}_t$,

Small-scale variables = $\{\mathrm{d}B_t, \ldots, \mathrm{d}B_{t+h}\}$.

To forecast this true trajectory using analogues, we assume that one is given a "catalogue" which is here a sub-sampling of a long numerical resolution of the stochastic Lorenz system ((1)–(3)), generated with one particular realization of the noises. Each catalogue has a set of elements $\left\{\mathbf{x}_{t'}, t' \in \mathcal{C}\right\}$ in which analogues are sought for, and a corresponding set of elements $\left\{\mathbf{x}_{t'+h}, t' \in \mathcal{C}\right\}$ in which the successors of the analogues can be found. To generate the different catalogues used in this study, fixed-size random subsamples were drawn from the same long numerical solution of ((1)–(3)) from time $t' = 0$ to $t' = 7008$ in non-dimensional time-unit. The notation t' is here to emphasis that the time of the catalogue is different from the time of the true trajectory (typically, analogues are sought for in a catalogue of past observations). As the catalogues are random subsamples, the time-difference between different analogues are irregular, but each analogue $\mathbf{x}_{t'}$ is associated to its successor $\mathbf{x}_{t'+h}$ with the same horizon h.

We employ the simplest type of analogue ensemble forecasting where the forecast is given by an ensemble of K successors $\mathbf{x}_{t'_i+h}$ of analogues $\mathbf{x}_{t'_i}$ of the true initial point $\mathbf{x}_t$, and each successor is given the same weight $\frac{1}{K}$. In all numerical experiments, the number of analogues is set to 50. Analogues are computed as K-

nearest neighbours, in terms of Euclidean distance, of the large-scale variables' initial value $\mathbf{x}_t$, which are assumed to be observed, while the noises $\mathrm{d}B_t$ are unknown. Other similarity measures could be envisaged (see e.g., [23–25]), but this is beyond the scope of this study. The analogue ensemble forecast is then made of the set successors $\left\{\mathbf{x}_{t'_1+h} \ \ldots \ , \mathbf{x}_{t'_K+h}\right\}$. Therefore, although $\mathbf{x}_{t'_i}$ should be close to $\mathbf{x}_t$, the sequence of noises $\left\{\mathrm{d}B_{t'_i}, \ldots, \mathrm{d}B_{t'_i+h}\right\}$ is completely different (independent) from $\{\mathrm{d}B_t, \ldots, \mathrm{d}B_{t+h}\}$.

- Analogue ensemble forecast :

Large-scale variables (initial conditions) = $\left[\mathbf{x}_{t'_1}, \ldots, \mathbf{x}_{t'_K}\right]$

Small-scale variables = $\left[\left\{\mathrm{d}B_{t'_1}, \ldots, \mathrm{d}B_{t'_1+h}\right\}, \ldots, \left\{\mathrm{d}B_{t'_K}, \ldots, \mathrm{d}B_{t'_K+h}\right\}\right]$

All catalogues are sub-samples of a long trajectory from time $t' = 0$ to $t' = 7008$ in non-dimensional time-unit, which corresponds to 60 years in atmospheric time. Note that our stochastic Lorenz system does not include seasonality, while in practice one must search for analogues in historical data within a seasonality condition, typically ± 2 months, and therefore our 60-years "season-less" catalogue will allow us to emulate the amount of data that one can access with $\sim$180 years of data. Sub-catalogues of different sizes will be drawn from this large initial catalogue to asses the effect of catalogue size on analogue ensemble. For each catalogue size, different catalogues can be drawn from the original large catalogue, allowing to assess the uncertainties in analogue ensemble forecasts even when the catalogue size is known.

To understand the contribution of different types of uncertainty to the analogue ensemble forecast, we also compute two other ensemble forecasts.

First, we build what we refer to as the "small-scale ensemble forecast". This ensemble is given the same size as the analogue ensemble : $K = 50$ members. Every member is initiated from the same true initial condition on the large-scale variables $\mathbf{x}_t$ (i.e., for horizon $h = 0$ the ensemble is a Dirac-delta function centered on $\mathbf{x}_t$). Each member is propagated with numerical simulations of the stochastic Lorenz and randomly generated noises $\left\{\mathrm{d}B'_t, \ldots, \mathrm{d}B'_{t+h}\right\}$ that are independent, different realizations of the same process as the true noises $\{\mathrm{d}B_t, \ldots, \mathrm{d}B_{t+h}\}$. Each member i corresponds to a particular realization $\left\{\mathrm{d}B_t^{\prime(i)}, \ldots, \mathrm{d}B_{t+h}^{\prime(i)}\right\}$ of the noises. Members are given equal weights $\frac{1}{K}$.

- Small-scale ensemble forecast :

Large-scale variables (initial conditions) = $[\mathbf{x}_t, \ldots, \mathbf{x}_t]$

Small-scale variables = $\left[\left\{\mathrm{d}B_t^{\prime(1)}, \ldots, \mathrm{d}B_{t+h}^{\prime(1)}\right\}, \ldots, \left\{\mathrm{d}B_t^{\prime(K)}, \ldots, \mathrm{d}B_{t+h}^{\prime(K)}\right\}\right]$

This ensemble represents the effect of uncertainty associated with the unresolved small spatial scales modelled by the noises. Since analogue forecasting is based only on the knowledge of resolved large-spatial-scale variables, one desired property of analogue forecasting ensemble is that it resembles the small-scale ensemble.

Second, we build what we call the "large-scale ensemble forecast". We compute ensemble forecasts from numerical simulations of the stochastic Lorenz starting at the analogues' positions in large-scale variables space $\left[\mathbf{x}_{t'_1}, \ldots, \mathbf{x}_{t'_K}\right]$, but using the same noise $\{\mathrm{d}B_t, \ldots, \mathrm{d}B_{t+h}\}$ as the "ground truth" trajectory. Therefore, while each member of the small-scale ensemble differs by the realization of the noises, on the contrary each member of the large-scale ensemble differs by the initial condition which corresponds to the initial value of the analogue ensemble. Members are again given equal weights $\frac{1}{K}$.

- Large-scale ensemble forecast :

Large-scale variables (initial conditions) = $\left[\mathbf{x}_{t'_1}, \ldots, \mathbf{x}_{t'_K}\right]$

Small-scale variables = $[\{\mathrm{d}B_t, \ldots, \mathrm{d}B_{t+h}\}, \ldots, \{\mathrm{d}B_t, \ldots, \mathrm{d}B_{t+h}\}]$

This large-scale ensemble allows to assess what the analogues would have evolved to if they were forced by the same small-scale as the ground truth trajectory. This ensemble allows to isolate the effect of mismatch in large-scale initial condition, also called "analogue-to-target distance" [13].

The properties of the different ensembles with respect to the ground truth and the analogues are summarized in Table 1 and Fig. 1. Figure 2 shows a schematic example of ensemble forecast trajectories in large-scale and small-scale space compared to the ground-truth. This figure is only illustrative and was not generated using the real stochastic Lorenz system. It highlights that the large-scale ensemble shares large-scale initial conditions with the analogue ensemble, while the small-scale ensemble shares large-scale initial condition with the ground truth. Conversely, the large-scale ensemble shares small-scale trajectory with the ground truth, while the analogue ensemble and the small-scale ensemble have different, independent small-scale trajectories.

Note that this procedure would be hardly feasible with a real physical model where spatial and temporal scales are not separated but intertwined. The stochastic Lorenz model not only allows to perform numerical simulations at a low cost, but

Table 1 Properties of the different forecast ensembles considered

	Initial condition on resolved large-scale variables (x, y, z)	Trajectory of unresolved small-scale variables $\mathrm{d}B_t$
Analogue ensemble	Past large-scale analogues of ground-truth	Past realizations associated with large-scale analogues of ground truth
Small-scale ensemble	Same as ground-truth ("present")	Random ("that could have been")
Large-scale ensemble	Same as analogues ("past")	Same as ground truth ("present")

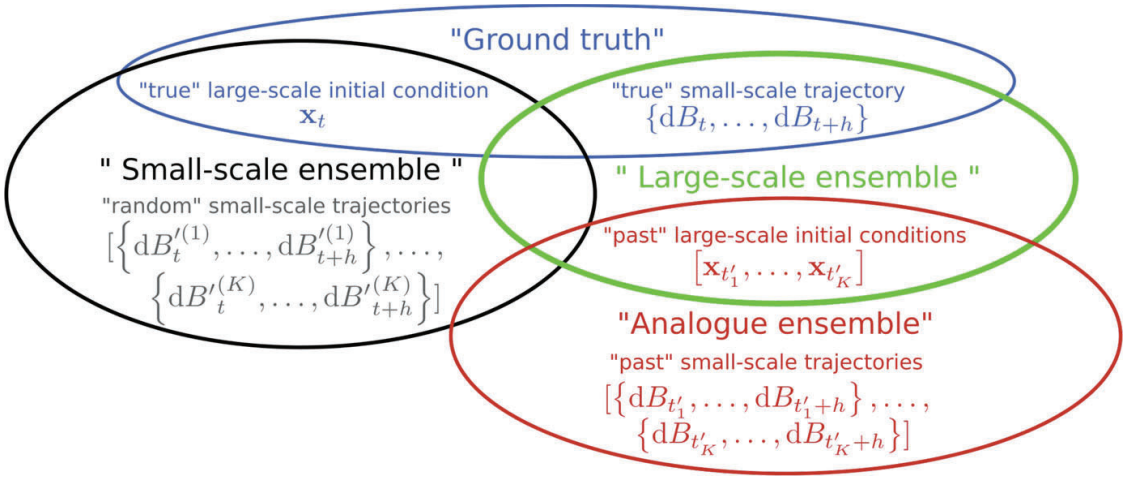

Fig. 1 Schematic of the shared properties of the different forecast ensembles considered and the ground truth trajectory

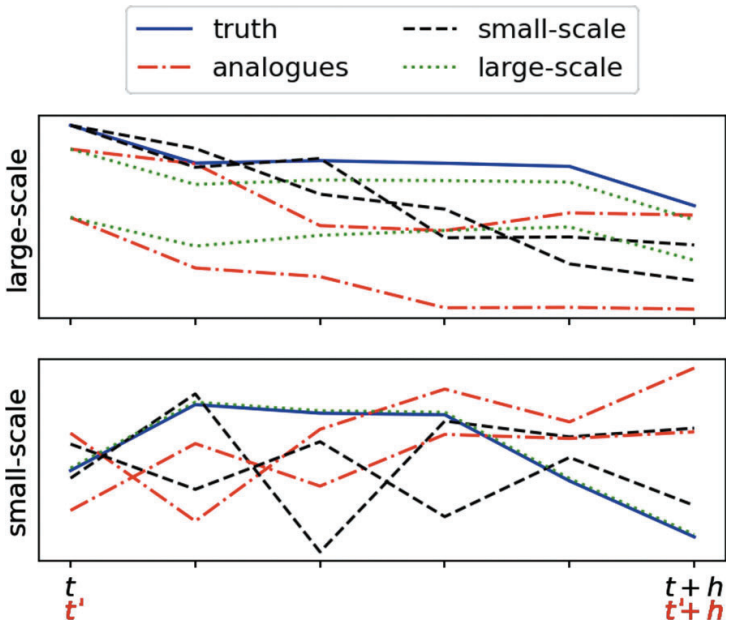

Fig. 2 Schematic of the trajectories of the different forecast ensembles considered and the ground truth trajectory, with only two ensemble members ($K = 2$) for readability purposes

also facilitates the separation between the effects of small-scale and large-scale components of the flow.

The ground truth to which these ensemble forecasts is compared is made of one long trajectory of 1168 non-dimensional time steps which corresponds to 10 years in atmospheric time-scale. The noises $\mathrm{d}B_t$ associated with this ground-truth trajectory are saved in memory to allow for the generation of the large-scale ensemble forecasts.

2.3 Statistical Tests to Compare Ensembles

In Sect. 3.2, we will make use of three statistical tests to assess the potential differences between two given forecast ensembles. The tests will be performed on each large-scale variable separately, either x, y or z.

The first test is called Welch's t-test [26], and is used to test the null hypothesis that the two ensembles' background distributions have the same mean value without necessarily having the same variance. It is based on the assumption that the two samples are normally distributed.

The second test is called Levene's test [27] and is used to test the null hypothesis that the two ensembles' background distributions have the same variance. This test does not make assumption of normality.

The third test is called the Kolmogorov-Smirnov test [28] and is used to test the null hypothesis that the two ensembles' background distributions are equal. It does not make assumption of normality, and can also be used for complex distributions such as multi-modal distributions, for which the notions of mean and variance are inoperative.

For all three tests, we used built-in functions from Python's SciPy package, namely "ttest_ind", "levene" and "kstest". We use the p-value as an indicator, compared to the reference values 0.01, 0.05 and 0.1. p-values below 0.05 indicate that the distributions are different (high confidence if below 0.01), while p-values above 0.05 indicate that the distributions are equal (high confidence if above 0.1).

3 Numerical Experiments

3.1 Small-scale and Large-scale Variability

Figure 3 shows an example of the three different forecast ensembles outlined in the previous section, along with the ground truth trajectory, for a given initial condition inside the attractor, with forecast ensembles of size $K = 20$ (we use a smaller value for visualization purposes only). The forecast horizon of 0.1 corresponds to between 7 and 8 hours in atmospheric time scale. The large-scale ensemble is less dispersed as the two other ensembles, and is close to the ground truth. The fact that this ensemble appears to shrink with time for small forecast horizon is a consequence of the dissipative nature of the Lorenz system. The analogue and small-scale ensembles are close to each other at horizon h =0.1, which is a desired property of analogue ensemble forecasting because it is based only on the knowledge of large-scale variables x, y, z, and it is not informed by the values of the small-scale variables (here, the noises). However, for smaller time horizon one can witness that the small-scale ensemble tends to the initial position, which is not the case of the analogue and large-scale ensembles. This undesired property of analogue forecasting errors for very short forecast-horizon was described previously in [17].

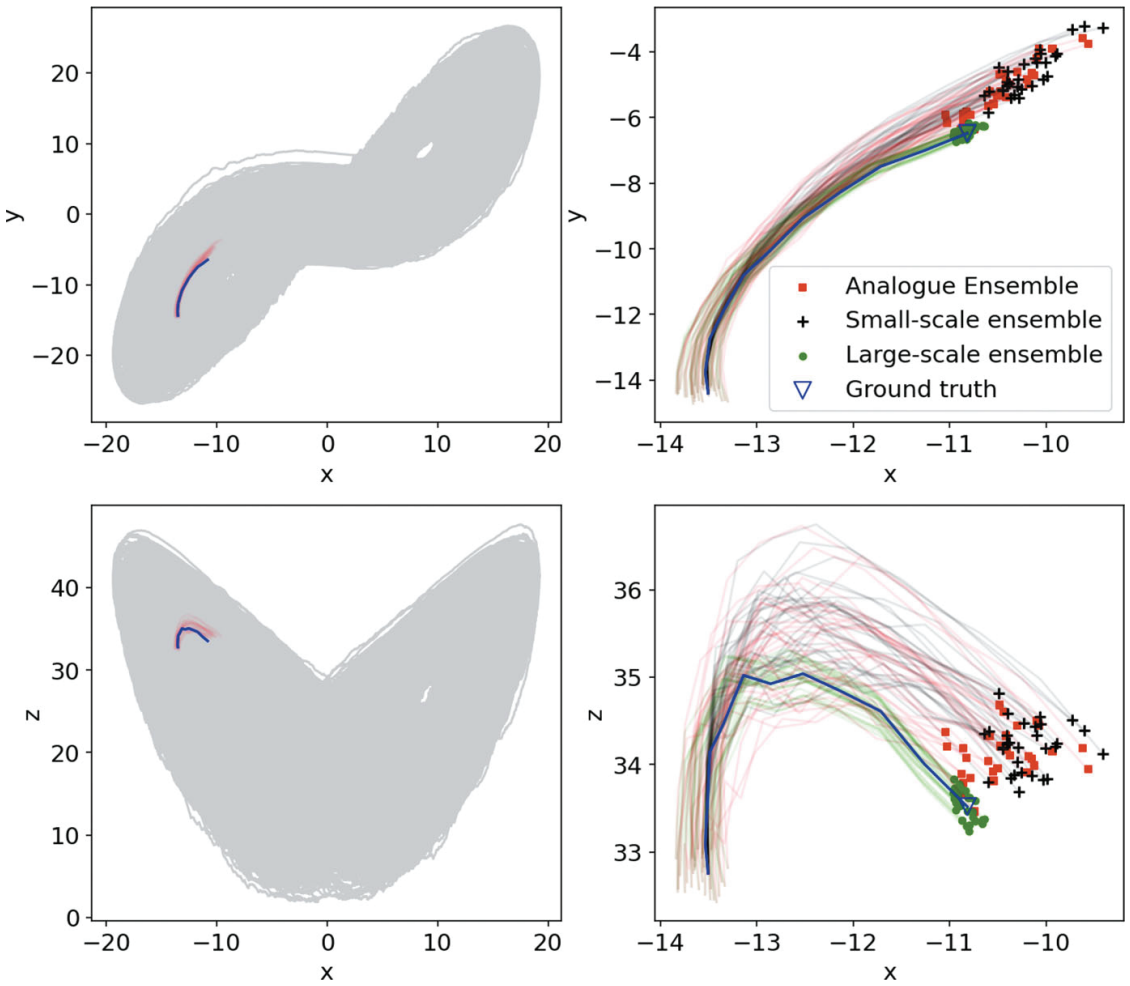

Fig. 3 Showing an example of three types of ensemble forecasts: small-scale (black lines and crosses), large-scale (green lines and circles) and analogue forecasting (red lines and squares) ensembles, versus ground truth trajectory (blue line and empty triangle). Final positions are given by the symbols. The forecast horizon is 0.1. We show only 20 members out of 50 for readability purposes

Focusing on ensemble spread, we now define the generalized variance (GV) as the determinant of the ensemble covariances, noted $|\Sigma_s|$ (small-scale), $|\Sigma_l|$ (large-scale), and $|\Sigma_a|$ (analogues).

Depending on the initial point, the situation may differ from Fig. 3. To test this assertion, Fig. 4 shows the ratio $(|\Sigma_s|/|\Sigma_l|)^{1/2}$ in log-scale, depending on the position of the initial point in the attractor, using 1000 points from the 10-years ground truth trajectory. Figure 4 shows that the relative spread of the large-scale and small-scale ensembles strongly depends on the values of the large-scale variables at the initial point. In particular and at this forecast horizon, small-scale variability dominates for values of $z < 20$, and large-scale variability dominates at the wing's borders for $z > 25$ and $-5 < y < 5$. Investigating the variability of $|\Sigma_s|$ and $|\Sigma_l|$ with initial point position shows that what drives high values of the ratio $|\Sigma_s|/|\Sigma_l|$ for low values of z is $|\Sigma_s|$, with a strong influence of small-scale noises in this area (not shown), while the smaller values of the ratio $|\Sigma_s|/|\Sigma_l|$ at the wings' borders are

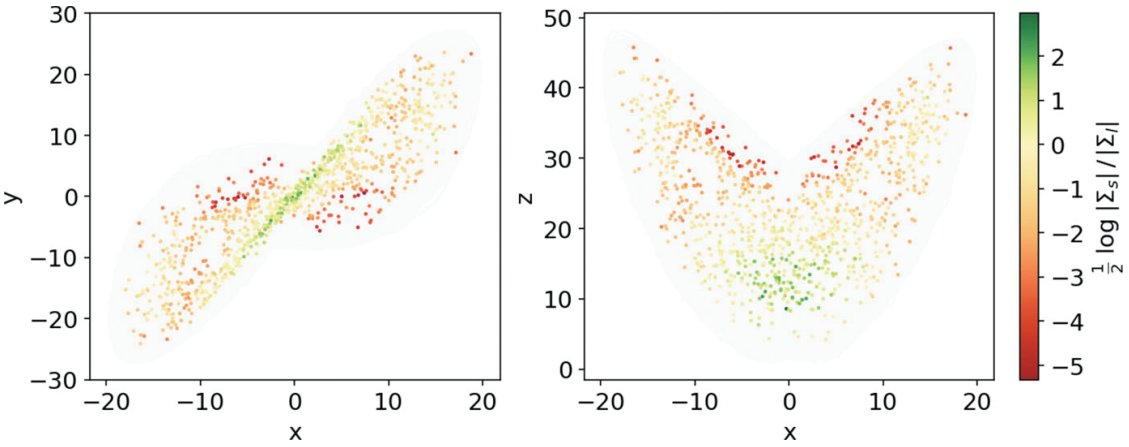

Fig. 4 Showing the ratio of small-scale to large-scale generalized standard-deviation ensemble variability (square-root of the covariance matrix's determinant), for forecast horizon 0.06 (equivalent to 4.5 hours in atmospheric time-scale), as a function of the initial point's position, in logarithmic scale

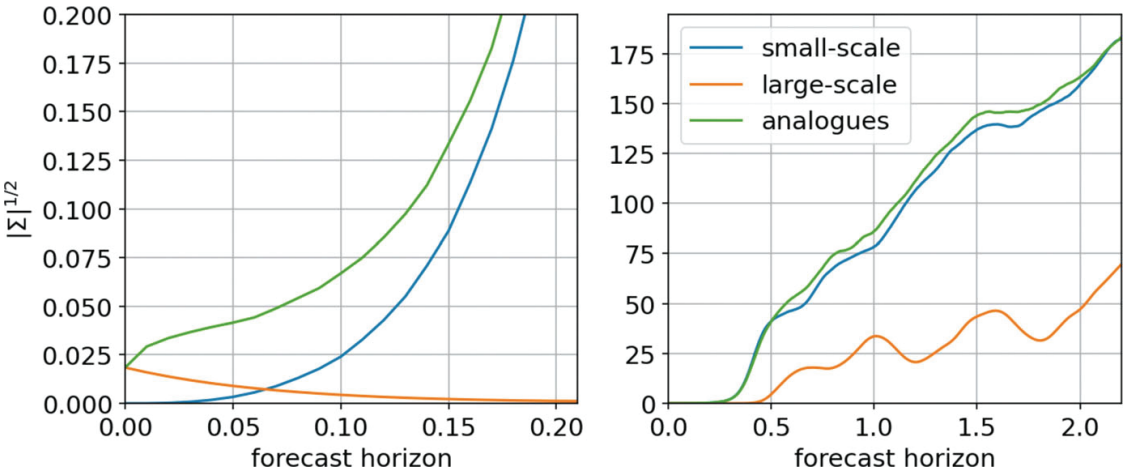

Fig. 5 Showing the attractor-averaged ensemble variability (given by the square root of the covariance matrix's determinant), against forecast horizon, for small-scale, large-scale and analogue ensemble forecasts. Left: zoom on small forecast horizon (<0.2). Right: forecast horizon up to 2.24

due to a larger values of $|\Sigma_l|$ which are caused by a smaller density of points (not shown). This experiment shows that analogue forecasting ensemble spread may be influenced both by the varying intensity of perturbation by unresolved spatial scales and by the varying density of points in the attractor (and therefore of analogue-to-target distances), both of which highly depend on the position in the attractor.

As shown in the particular case of Fig. 3, the ration between these ensemble GV depends not only the initial large-scale position but also on forecast horizon. The square root of GVs (which can be viewed as a generalized standard-deviation) $|\Sigma_s|^{1/2}$, $|\Sigma_l|^{1/2}$ and $|\Sigma_a|^{1/2}$, are averaged over 1000 points in the attractor (randomly drawn from the 10-year ground truth trajectory), and plotted against forecast horizon in Fig. 5. By construction, the small-scale ensemble GV tends to zero for small forecast horizon. Large-scale GV decreases as a function of time for small time-horizons because the Lorenz system is dissipative, however it remains

non-zero. The analogue and large-scale ensemble GV are equal at horizon-zero by construction, and then the analogue ensemble GV grows similarly to small-scale ensemble GV. This shows that the analogue ensemble variability is highly linked to the small-scale ensemble variability, although it naturally overestimates the latter for short forecast-horizon because of the non-zero distance between the analogues and the initial point. Therefore, the only way that analogues can adequately represent the small-scale driven variability is to decrease analogue-to-target distance by increasing the catalog size.

3.2 Convergence of Analogues Towards the Small-scale Ensemble

In Figs. 6 and 7, two examples of convergence of the distribution of analogues to the small-scale ensemble distribution are shown, at fixed forecast horizon 0.1 (7–8 hours). The analogue forecast ensembles converge the small-scale ensembles in the limit of large catalogue size. In addition to showing quantiles of distribution as a function of catalogue size, we also show p-values for three statistical tests where the null hypothesis is that the analogue and small-scale ensembles are drawn from the same background continuous probability distribution. The catalog size needed for convergence depends on the point of reference and on the chosen variable. For instance it is fastest for variable z and Fig. 6, and slowest for variable z and Fig. 7. This is because in the second example the initial point is in an area of smaller attractor density. In Fig. 6, the average of the distribution converges faster than the standard deviation, while in the second example all three statistical tests indicate a convergence for approximately the same catalogue size. When the initial point is in an area of large attractor density, the analogue ensemble is likely to have a precise mean even for small catalogue size, although it will necessarily overestimate the variance of the small-scale ensemble. On the contrary, for small catalogue sizes and in areas of small attractor density such as the tail of the butterfly's wings, the analogue average is biases towards the attractor's average.

The features highlighted in the particular cases of Figs. 6 and 7 allow to better understand the general law given in Fig. 8. In the latter, we test the convergence of the analogue ensemble to the small-scale ensemble (for horizon 0.13, ∼10 hours, and horizon 0.43, ∼32 hours) according to the three statistical tests. The p-values are stored for 10 catalogues for each catalogue size and for 1000 points in the attractor taken from the ground truth trajectory, and quantiles are extracted for each catalogue size. The observed behaviour for the top plots (a-c, horizon 0.13, ∼10 hours) resembles the one of Fig. 6, which is the case where forecast horizon is small, and the attractor density is high at the initial point. This shows that for short forecast horizon the analogue ensemble demands a high catalogue size to adequately represent the standard-deviation while the average can be well represented even with small catalogue size. However, this must be tempered by the fact that the

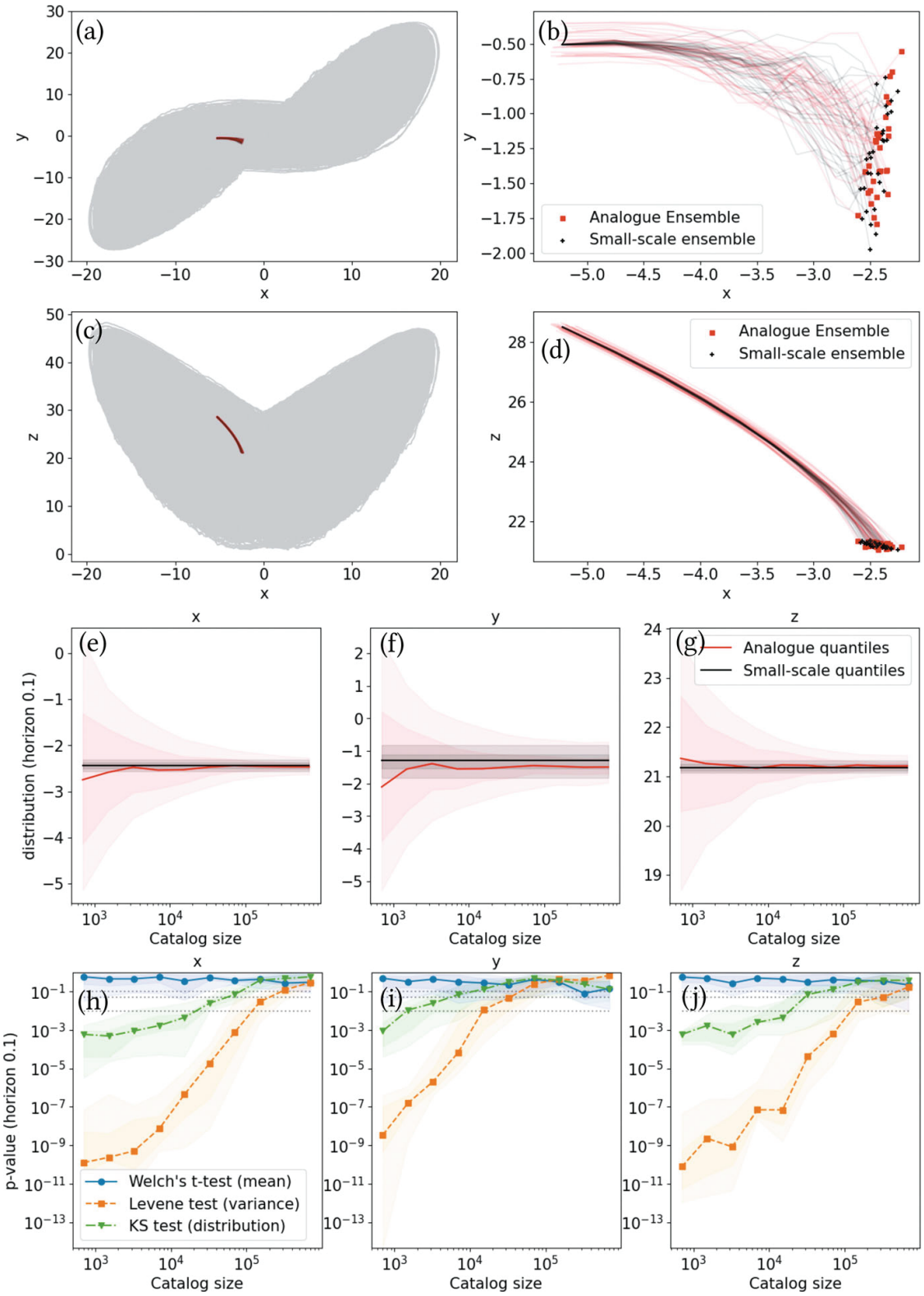

Fig. 6 One example of analogue ensemble forecast versus small-scale perturbed forecast, at forecast horizon 0.1. Top (**a**)–(**d**): forecast trajectories, unzoomed (left) and zoomed (right). Middle (**e**)–(**g**): analogue ensemble distribution (red) as a function of catalogue size, and small-scale ensemble distribution (black) shown for comparison, for each variable (x: (a), y: (b), z: (c)), showing quantiles 0.05-0.25-0.5-0.75-0.95. Bottom (**h**)–(**j**): p-value for Welch's t-test (full blue line and circles), Levene's test (dashed orange line and squares), and the Kolmogorov-Smirnov test (dash-dotted green line with triangles), versus catalogue size, where the null hypothesis is that the analogue ensemble and the small-scale ensemble are drawn from the same distribution, for each variable (x: (a), y: (b), z: (c)), showing quantiles 0.05-0.25-0.5-0.75-0.95 for different draws of catalogues for each catalogue size. Reference p-values of 0.01-0.05-0.1 are shown (dotted gray lines)

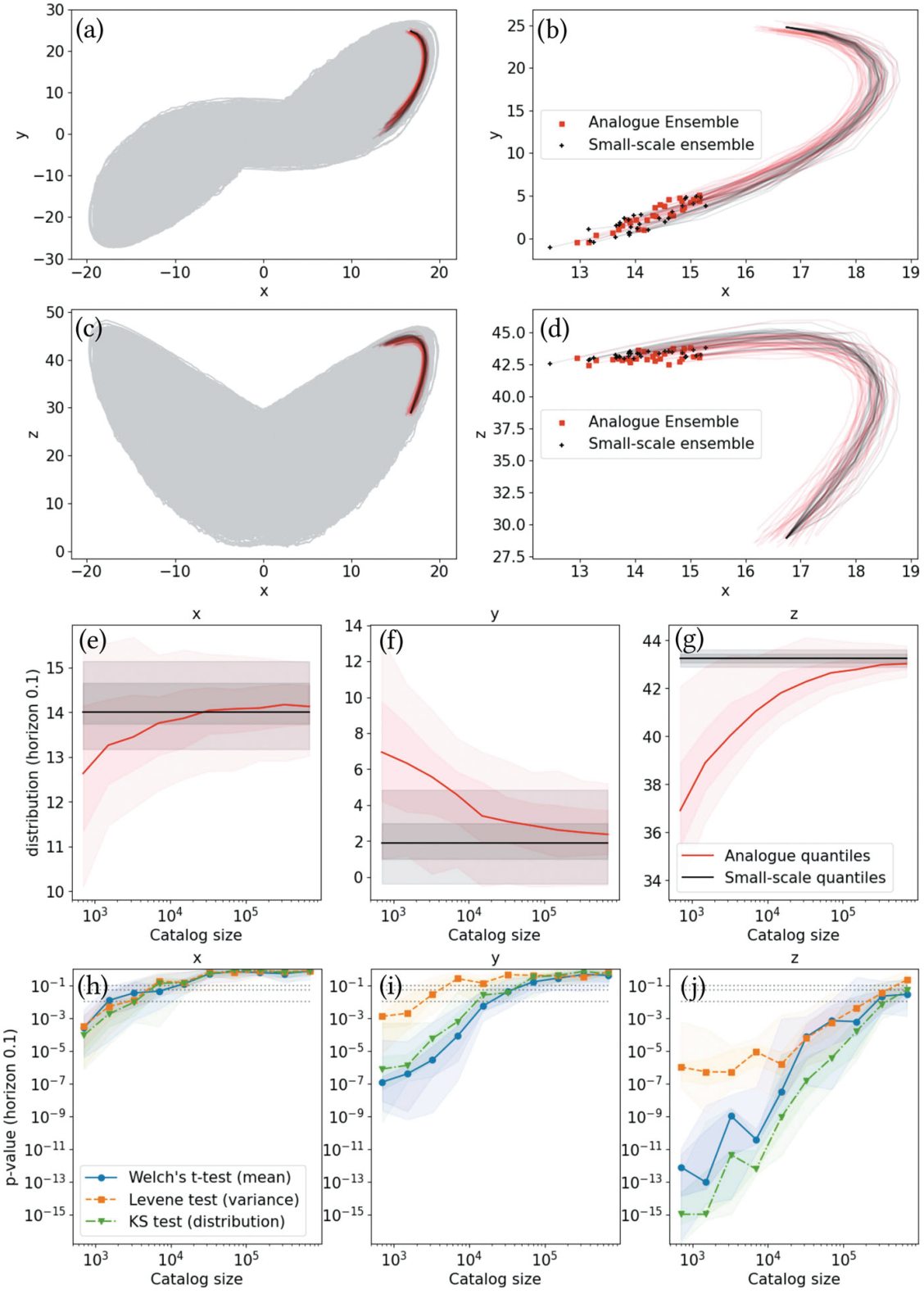

Fig. 7 Same as Fig. 6 but for a different initial condition

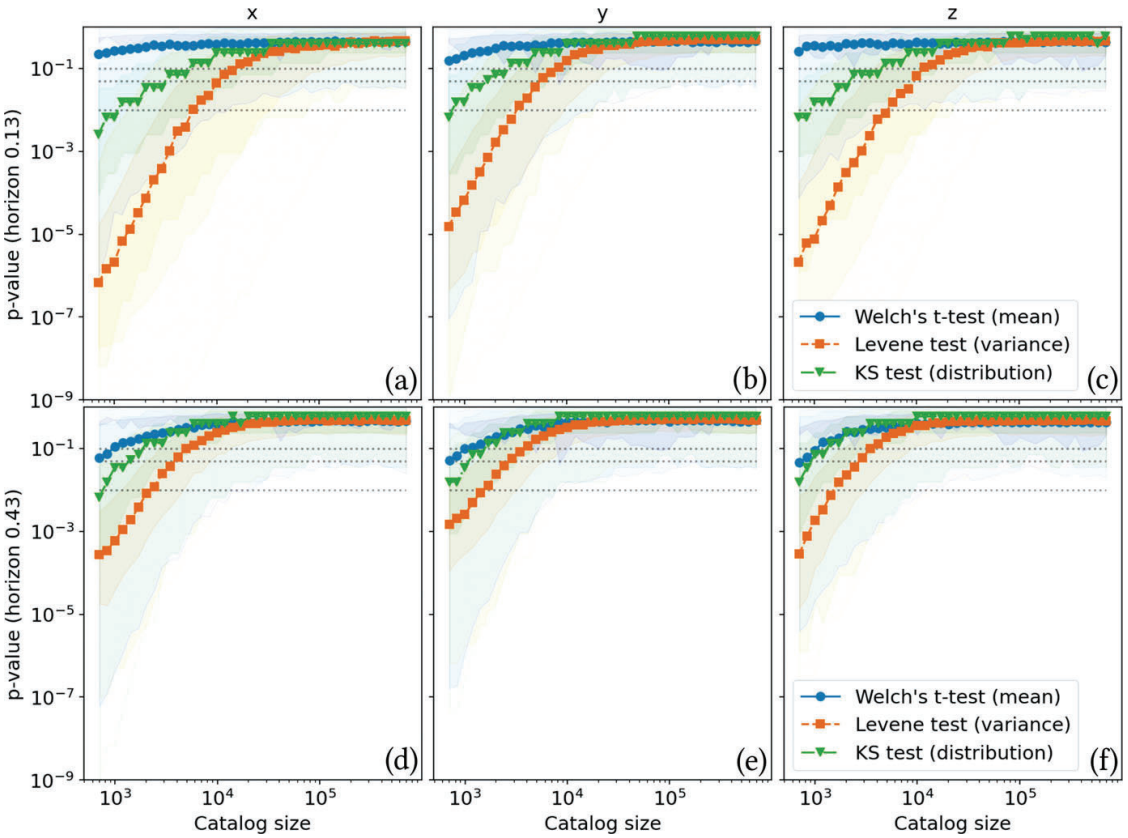

Fig. 8 Same as bottom plots of Fig. 6h–j and Fig. 7h–j, but quantiles of p-values are derived from statistics over the whole attractor (1000 points from a long trajectory) rather than for one point. Top (**a**)–(**c**): forecast horizon=0.13. Bottom (**d**)–(**f**): forecast horizon=0.43. Left (a,d): x. Middle (b,e): y. Right (c,f): z

quantiles around the median p-value are highly scattered, indicating the need to adapt analogue forecasting strategies beyond the simple brute-force ensemble when dealing with extreme or rare events.

For larger forecast horizon (bottom plots of Fig. 8), Welch's test converges more slowly while the Levene test converges faster with catalogue size. It indicates that the forecasting of the average is more challenging for such horizons, which can be attributed to the chaotic growth of small initial errors at these time scales (as can be witnessed in the beginning of the growth of the large-scale GSTD around horizon 0.5, see Fig. 5). On the contrary, the analogue ensemble's spread is closer to the small-scale ensemble spread, as the latter grows with time, and the effect of analogue-to-target distance starts to fade at these horizons (see again in Fig. 5 how the small-scale and analogue ensemble GSTD get closer to each other at large forecast horizon).

Building from the results of Fig. 8, we define a minimum catalogue size for each statistical test and large-scale variable, as the catalogue size for which a given quantile of p-value exceeds the threshold value of 0.1, where quantiles of p-value are calculated over 1000 points on the attractor and for 10 catalogues of equal size. To make this computation, we consider 40 different catalogue sizes

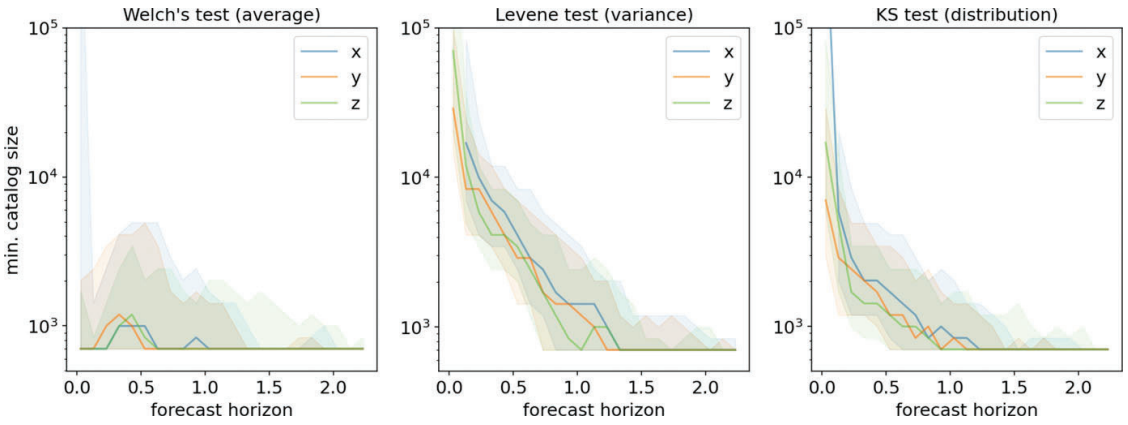

Fig. 9 Minimum catalogue size so that the quantile (0.3,0.5,0.7) of p-value over the attractor (1000 points from a long trajectory) exceeds the threshold value of 0.1, versus forecast horizon, for three tests (Welch: left, Levene: middle, Kolmogorov-Smirnov: right) and variables (x: blue, y: orange, z: green)

ranging from 700 to 700,000 independent elements, equally-spaced on a log-scale. Results for each variable (x, y, z) and for three quantiles of p-value (0.3,0.5,0.7) are shown in Fig. 9. To interpret these minimum catalog sizes, we make the parallel with atmospheric circulation data. We assume that one can find 60 independent circulation patterns as "analogue candidates" per year of data, if one restricts the search to a "season" of 4 months and if two circulation patterns are uncorrelated after 2 days on average.

The minimum catalogue size for Welch's t-test is highest around forecast horizon 0.4, for which a catalogue size of 10^3 is needed for the median p-value to exceed 0.1. Using our interpretation in terms of atmospheric data, a catalog size of 10^3 elements amounts to 20 years of data, which is reasonable with respect to nowadays's typical length of observational records. The minimum catalogue size for Welch's t-test then decreases with time horizon, which would not be the case for a deterministic system. Here, since the unresolved small-scale induce inevitable uncertainty in the forecast, the mismatch in initial condition due to the finite catalogue size has a lesser influence on the analogue ensemble average for larger forecast horizon. We consider horizons up to 2.24 which is equivalent to 1 week in atmospheric time-scale. For such large forecast horizons it appears that one does not require perfect knowledge of the large-scale initial condition in order to evaluate the average forecast under small-scale perturbation, which is due to the fact that the uncertainty due to the unknown small-scale features has more influence on the forecast than the initial large-scale perturbation due to the analogue-to-target distances. This particular experiment shows that, for chaotic multi-scale systems, analogue forecasting may actually be more successful in predicting the average of the stochastic large-scale system, than to predict the deterministic multi-scale system altogether.

The curves for the Levene and Kolmogorov-Smirnov tests are very similar to each other, but very different from the curves of Welch's t-test, indicating that what the analogue ensemble struggles to estimate is not the average but the spread

of the small-scale distribution. As we have seen, for small horizons and small catalogue size the analogue ensemble overestimates the ensemble spread attributed to the unresolved small-scales. According to Fig. 9, this effect cannot be neglected until horizon 1.2, which corresponds to ∼4 days in atmospheric time-scale. Taking this analogy further, for a 1-day (horizon 0.4) analogue ensemble forecast one would need a catalogue of 6000 independent elements, which would correspond to 120 years of data, in order for the analogue ensemble spread to correspond to the small-scale perturbed ensemble spread. Indeed, if one uses a smaller catalogue size, there is a risk that although the average would be well estimated, the spread would not, which is an unwanted property.

This result urges for the development of new ways of estimating the forecast spread based on the analogue ensemble, beyond using directly the spread of the analogue ensemble. Although optimizing the value of K might help in reducing the analogue ensemble spread, it will not be able to reduce the non-zero analogue-to-target distance due to finite catalogue size. Indeed, analogue-to-target distances depend on the analogue rank K as $(K/L)^{1/d}$ where d is the local dimension of the attractor and L is the catalogue size [13]. Therefore the distance between two analogues is lower-bounded by a constant proportional to $L^{-1/d}$, which gives a lower-bound for the analogue ensemble spread as well. Moreover, for moderately high-dimensional systems such as large-scale atmospheric circulation, the dimension ranges from 10 to 20 [29], and therefore analogue-to-target distances increase very slowly with the number of analogue K, therefore decreasing the number of analogues is likely not to reduce efficiently the analogue ensemble spread.

4 Conclusion and Perspectives

We have studied the effects of unresolved spatial scales on analogue forecast ensembles, using numerical simulations of a set of stochastic differential equations that emulate the effects of small-scales on the three large-scale variables of the chaotic Lorenz system in a physically consistent way. We have interpreted the behaviour of analogue ensemble forecasts based on two other ensembles, the large-scale ensemble which has the same noises (i.e. the same small-scale) as the ground truth but the same large-scale initial conditions as the analogues, and the small-scale ensemble which has the same large-scale initial condition as the ground truth but different noises (i.e. different small-scales). We have shown that the ratio of large-scale to small-scale ensemble spread strongly depends on the initial position in the attractor and on forecast horizon. In particular, the large-scale (respectively, small-scale) spread dominates for short-term (respectively, long-term) forecasts. The analogue ensemble spread is dominated by large-scale effects for short term forecasts and gradually tends to the small-scale forecast spread for larger time-horizons. The analogue ensemble converges in distribution to the small-scale ensemble in the limit of large catalogue size, which is a desired property of analogue

ensemble forecasts. The rate of convergence depends on the position in the attractor and on the forecast horizon. The convergence of the analogue ensemble average to the small-scale ensemble average is rather fast, while the convergence of ensemble spread is slow. In particular, for short-term forecast of 1–2 days in atmospheric time-scale, our results in this idealized setup suggest that the catalogue size needed for the analogue ensemble to adequately represent the small-scale driven spread might exceed hundreds of years of data. In order for analogue ensembles to reliably estimate uncertainties associated to unresolved small-scales, one might benefit from the development of methodologies that go beyond the simple spread of the analogue ensemble.

This work should be extended to higher-dimensional physical systems with scale separation, and systematic tests of the adequacy between the analogue ensemble spread and the uncertainty of the analogue ensemble forecast with respect to the ground truth trajectory should be performed using atmospheric and ocean circulation data such as reanalysis products. We advocate for an adaption of uncertainty quantification from analogue ensembles using leave-one-out procedures, famous in the machine-learning community [30], which could also be used to produce unbiased analogue ensemble average forecasts in data-scarce areas of the attractor. Finally, proper tuning of the distance using metric-learning algorithms could help calibrate the analogue ensemble towards the desired distribution.

Acknowledgments This work was financially supported by the ERC project 856408-STUOD. We thank Pierre Dérian for providing code snippets on numerical simulations of the Lorenz System under location uncertainty. We thank two anonymous reviewers for helpful comments and suggestions.

References

1. H. Poincaré, "Sur le problème des trois corps et les équations de la dynamique," *Acta mathematica*, vol. 13, no. 1, pp. A3–A270, 1890.
2. E. N. Lorenz, "Atmospheric predictability as revealed by naturally occurring analogues," *Journal of Atmospheric Sciences*, vol. 26, no. 4, pp. 636–646, 1969.
3. R. Alexander, Z. Zhao, E. Székely, and D. Giannakis, "Kernel analog forecasting of tropical intraseasonal oscillations," *Journal of the Atmospheric Sciences*, vol. 74, no. 4, pp. 1321–1342, 2017.
4. P. Yiou and C. Déandréis, "Stochastic ensemble climate forecast with an analogue model," *Geoscientific Model Development*, vol. 12, no. 2, pp. 723–734, 2019.
5. S. Alessandrini, L. Delle Monache, S. Sperati, and G. Cervone, "An analog ensemble for short-term probabilistic solar power forecast," *Applied energy*, vol. 157, pp. 95–110, 2015.
6. D. Burov, D. Giannakis, K. Manohar, and A. Stuart, "Kernel analog forecasting: Multiscale test problems," *Multiscale Modeling & Simulation*, vol. 19, no. 2, pp. 1011–1040, 2021.
7. H. Ding, M. Newman, M. A. Alexander, and A. T. Wittenberg, "Diagnosing secular variations in retrospective enso seasonal forecast skill using cmip5 model-analogs," *Geophysical Research Letters*, vol. 46, no. 3, pp. 1721–1730, 2019.
8. Y. Zhen, P. Tandeo, S. Leroux, S. Metref, T. Penduff, and J. Le Sommer, "An adaptive optimal interpolation based on analog forecasting: application to ssh in the gulf of mexico," *Journal of Atmospheric and Oceanic Technology*, vol. 37, no. 9, pp. 1697–1711, 2020.

9. L. Delle Monache, F. A. Eckel, D. L. Rife, B. Nagarajan, and K. Searight, "Probabilistic weather prediction with an analog ensemble," *Monthly Weather Review*, vol. 141, no. 10, pp. 3498–3516, 2013.
10. P. Yiou, "Anawege: a weather generator based on analogues of atmospheric circulation," *Geoscientific Model Development*, vol. 7, no. 2, pp. 531–543, 2014.
11. R. Lguensat, P. Tandeo, P. Ailliot, M. Pulido, and R. Fablet, "The analog data assimilation," *Monthly Weather Review*, vol. 145, no. 10, pp. 4093–4107, 2017.
12. R. Alexander and D. Giannakis, "Operator-theoretic framework for forecasting nonlinear time series with kernel analog techniques," *Physica D: Nonlinear Phenomena*, vol. 409, p. 132520, 2020.
13. P. Platzer, P. Yiou, P. Naveau, J.-F. Filipot, M. Thiebaut, and P. Tandeo, "Probability distributions for analog-to-target distances," *Journal of the Atmospheric Sciences*, vol. 78, no. 10, pp. 3317–3335, 2021.
14. A. Wolf, J. B. Swift, H. L. Swinney, and J. A. Vastano, "Determining lyapunov exponents from a time series," *Physica D: nonlinear phenomena*, vol. 16, no. 3, pp. 285–317, 1985.
15. J. D. Farmer and J. J. Sidorowich, "Predicting chaotic time series," *Physical review letters*, vol. 59, no. 8, p. 845, 1987.
16. J. D. Farmer and J. J. Sidorowichl, "Exploiting chaos to predict the future and reduce noise," in *Evolution, learning and cognition*, pp. 277–330, World Scientific, 1988.
17. P. Platzer, P. Yiou, P. Naveau, P. Tandeo, J.-F. Filipot, P. Ailliot, and Y. Zhen, "Using local dynamics to explain analog forecasting of chaotic systems," *Journal of the Atmospheric Sciences*, vol. 78, no. 7, pp. 2117–2133, 2021.
18. B. Chapron, P. Dérian, E. Mémin, and V. Resseguier, "Large-scale flows under location uncertainty: a consistent stochastic framework," *Quarterly Journal of the Royal Meteorological Society*, vol. 144, no. 710, pp. 251–260, 2018.
19. E. N. Lorenz, "Deterministic nonperiodic flow," *Journal of the atmospheric sciences*, vol. 20, no. 2, pp. 130–141, 1963.
20. V. Resseguier, E. Mémin, and B. Chapron, "Geophysical flows under location uncertainty, part i random transport and general models," *Geophysical & Astrophysical Fluid Dynamics*, vol. 111, no. 3, pp. 149–176, 2017.
21. V. Resseguier, E. Mémin, and B. Chapron, "Geophysical flows under location uncertainty, part ii quasi-geostrophy and efficient ensemble spreading," *Geophysical & Astrophysical Fluid Dynamics*, vol. 111, no. 3, pp. 177–208, 2017.
22. T. N. Palmer, "Extended-range atmospheric prediction and the lorenz model," *Bulletin of the American Meteorological Society*, vol. 74, no. 1, pp. 49–66, 1993.
23. Z. Toth, "Intercomparison of circulation similarity measures," *Monthly weather review*, vol. 119, no. 1, pp. 55–64, 1991.
24. K. Fraedrich and B. Rückert, "Metric adaption for analog forecasting," *Physica A: Statistical Mechanics and its Applications*, vol. 253, no. 1–4, pp. 379–393, 1998.
25. C. Matulla, X. Zhang, X. Wang, J. Wang, E. Zorita, S. Wagner, and H. Von Storch, "Influence of similarity measures on the performance of the analog method for downscaling daily precipitation," *Climate Dynamics*, vol. 30, pp. 133–144, 2008.
26. B. L. Welch, "The generalization of 'student's'problem when several different population varlances are involved," *Biometrika*, vol. 34, no. 1–2, pp. 28–35, 1947.
27. H. Levene, "Robust tests for equality of variances," *Contributions to probability and statistics*, pp. 278–292, 1960.
28. J. W. Pratt, J. D. Gibbons, J. W. Pratt, and J. D. Gibbons, "Kolmogorov-smirnov two-sample tests," *Concepts of nonparametric theory*, pp. 318–344, 1981.
29. D. Faranda, G. Messori, and P. Yiou, "Dynamical proxies of north atlantic predictability and extremes," *Scientific reports*, vol. 7, no. 1, pp. 1–10, 2017.
30. D. W. Aha, "Lazy learning," in *Lazy learning*, pp. 7–10, Springer, 1997.