Thank you for your comment and for your suggestions. We detail below how we plan to change the manuscript in order to take them into account.

Referee's comment 1: *The title refers to spread/error and spatial error structures, but mainly uses spectra to examine scale dependence. I would suggest changing the title to more closely match the main results as listed in the introduction.*

We propose to change "spatial error structure of precipitation ensemble nowcasting" to "spatial error representation in precipitation nowcasting models" in the title, so that the title would read "Spread/Error relationship and spatial error representation in precipitation nowcasting: Comparison of STEPS and generative AI". The first part of the title ("Spread/Error relationship") refers to the results of Sec. 3.1 of the manuscript, while the second part ("spatial error representation") refers to the results of Sections 3.2 and 3.3.

Referee's comment 2: *Spread and skill are indeed discussed in section 3.1, but never mentioned again. Could the diagnostics in the later sections be used to explain the features that were pointed out in Fig. 2?*

We view these spread/skill relationship on one hand and the spatial metrics ($\cos\gamma$ and the FSS) on the other hand as two sets of metrics measuring complementary features. The spread/skill relationship measures whether the phase space residual vectors v_i have appropriate sizes with respect to the error, while the spatial metrics quantify whether the directions of the v_i are appropriate (especially in the case of $\cos\gamma$). With this interpretation in mind, no specific feature in Fig. 2 can be linked to the spatial metrics.

We propose to add the following sentence

> In particular, it shows that LDCast and STEPS residual vectors v_i have appropriate sizes with respect to the error since the spread/skill relationships are close to 1. But the directions of v_i in phase space (and thus their localizations in real space) are random, as measured by $\cos\gamma$ and the FSS.

after

> This work investigates the spectral and spatial properties of STEPS ensembles and a pre-trained version of LDCast.

to clarify the interpretation of the results and their relationship.

Referee's comment 3: *I am not familiar with looking at the eigenvalues of the covariance across ensemble members. Could the authors supply reference to previous applications of this method, or if it is novel, give some more motivation for what questions it can answer and why it will be useful here? I am tempted to interpret it by analogy to PCA, in which case it might be interesting to look at the spatial patterns (analogous to EOFs), especially for STEPS where a few leading eigenvalues dominate the response.*

We felt it was quite natural to look at eigenvalues and eigenvectors of the covariance matrix and looking now to the literature, we indeed see that this tool has already been used in this context. Thank you very much for drawing our attention to that.

We propose to add the following at the beginning of Sec. 3.2:

> The ensemble covariance matrix and its eigenvalues and eigenvectors are used in this Section to investigate the morphology of the ensembles. If the ensemble is considered as a cloud of points in phase space, the eigenvectors of the covariance matrix point in the directions in which the ensemble is the most extended (as measured by the variance in these directions, which are given by the associated eigenvalues). The method is therefore equivalent to Principal Component Analysis where all components are kept. This tool was originally introduced by Patil et al. (2001) and then by Oczkowski et al. (2005) to estimate the dimension of ensemble of forecasts ("E dimension"), reducing the information in the covariance matrix to one number. Other works (Satterfield and Szunyogh, 2011; Ono, 2025) followed in that direction. On the other hand, the full eigenvalue spectrum of the covariance of ensemble members was used by Snyder et al. (2003) in an idealized model.
>
> Some of the authors who developed the "E dimension" have also later used the information in the eigenvectors of the ensemble covariance matrix, by using the fraction of error explained by the eigenvectors as a metric to evaluate ensembles of forecasts (Szunyogh et al., 2005; Satterfield and Szunyogh, 2010).

This browsing of the literature brought up additional references, which are relevant for Sec. 3.3, so that we will change the sentence

> The first is $\cos\gamma(e,p)$ is the cosine of the angle between the error e and its projection p on the subspace spanned by the residual vectors v_i.

at lines 190-191 to

> The first is the Perturbation versus Error Correlation Analysis (PECA) score Wei and Toth (2003), which is called $\cos\gamma(e,p)$ here in order to emphasize its geometric interpretation in phase space. Indeed, it is the cosine of the angle between the error e and its projection p on the subspace spanned by the residual vectors v_i.

The sentence

> A similar quantity was considered in Uboldi and Trevisan (2015) to estimate the ability of the breeding vectors of a realistic model to capture the error growth in a convective situation.

at lines 199-200 will also be changed into

> This metric is indeed equal to the PECA score, except that the error and the perturbations are computed with respect to a control run for the PECA score. Apart from that, minimizing the L_2-norm of $e - \sum_i \alpha_i v_i$ amounts to constructing the projection of e on the space spanned by the v_i.
>
> The same quantity was also introduced independently by Szunyogh et al. (2005); Satterfield and Szunyogh (2010) as the Explained Variance (EV) to evaluate the efficiency of data assimilation schemes to provide initial conditions leading, under the perfect model hypothesis,

> to ensembles of forecasts capturing the error. Finally, Uboldi and Trevisan (2015) also considered that metric to estimate the ability of breeding vectors of a realistic model to capture the error growth in a convective situation.

We will also add a visualization of the first eigenvectors of the covariance matrix in the Supplementary file.

Referee's comment 4: *The discussion of 2D versus 3D turbulence around line 159 is not entirely correct. The mesoscale is not like 3D turbulence except in that it appears to share the -5/3 energy spectrum. It is this, not the dimensionality, that matters for the rate of error growth, as shown by Rotunno and Snyder (2008) using the Lorenz (1969) model.*
Thank you for pointing out this. We propose to change

> Indeed, on a theoretical basis, it is expected that the highest scale contaminated by the error evolves as an exponential of the lead time for 2D turbulent fluids (as is seen when working on synoptic scales), while the highest scale reached by the error increases as a power law of the lead time for 3D turbulent fluids (Vallis, 2017). Given the scales of nowcasting, the relevant physical processes are those of 3D turbulence, so that there is a qualitative agreement between these theoretical considerations and the scaling on the error observed in Fig. 3.

to

> Indeed, the fact that the error grows as power law of the time is consistent with the limited predictability at the nowcasting scales. This limited predictability is related to the structure of the energy spectrum of atmospheric fields at these scales, following closely a -5/3 power law Rotunno and Snyder (2008).

Referee's comment 5: *The decrease at small scales with time that is noted at line 185 appears to be only true for LDCast.*
We propose to clarify the sentence

> Spectra at small scales decrease over time. This is because the eigenvectors have unit norm, so that the sum over the scales of each power spectrum is equal to 1.

as

> Note that the eigenvectors have unit norm so that values in a power spectrum are only meaningful relative to other values in that power spectrum. For example, the spectra of LDCast eigenvectors at small scales seem to decrease with lead time, but this is not due to an absolute decrease in the small-scale variability. It is rather due to a decrease in the small-scale variability relative to the large-scale variability in the same vectors.

Referee's comment 6: *The introduction to section 3.3 should include more explanation of the MAAFT ensembles. It is noted that they have the same distribution of rainfall intensities and residual vectors with* *the same power spectra, but the reader has to go to the appendix to find out how they are different, and then work out for themselves if this is an appropriate reference ensemble for the questions considered here. The SPEC ensemble is also an interesting comparison and should be introduced properly.*

Thank you for the comment, the different types of ensembles are indeed not properly introduced. We propose to replace the lines 206-208 by

> The two metrics are computed for STEPS and LDCast ensembles and compared to those of three types of surrogate ensembles. Surrogate ensembles are generated as follows:
>
> – HIST ensembles: for each member x_i in the original ensemble, a HIST surrogate z_i is generated by shuffling all the pixel values of x_i (x_i and z_i have thus the same histogram of values);
>
> – SPEC ensembles: for each x_i (that we decompose as $\bar{x} + v_i$), a SPEC surrogate is generated as $\bar{x} + w_i$, where w_i has the same power spectrum as v_i but with random Fourier phases (SPEC surrogates do not have the same histogram of values as the original members);
>
> – MAAFT ensembles: for each x_i, a MAAFT surrogate is built in such a way that it has the same histogram of pixel values as x_i and the same power spectrum for its residual vector as v_i. The name stands for Modified Amplitude Adjusted Fourier Transform.
>
> MAAFT surrogates are inspired from the Iterated Amplitude-Adjusted Fourier Transform (IAAFT) (Schreiber and Schmitz, 1996). The construction of SPEC and MAAFT ensembles is detailed in Appendix A.

In order to ease the reading of the core part of the manuscript, we propose to keep the detailed steps of the construction of SPEC and MAAFT ensembles in Appendix A, which would read:

> The construction of the surrogates used in this work are detailed in this Appendix.
>
> Surrogate ensembles are generated member by member and lead time by lead time. As explained in the text, HIST ensembles are generated by shuffling the pixel values of each original member x_i. The original member and its surrogate have exactly the same distribution of values.
>
> SPEC surrogates are generated for each original member x_i by
>
> 1. initializing a surrogate z_i by shuffling all the values of x_i (as for a HIST surrogate);
> 2. computing the residual of z_i with respect to the mean as $w_i = z_i - \bar{x}$;
> 3. constructing w'_i by adjusting the power spectrum of w_i to that of $v_i = x_i - \bar{x}$: the Fourier coefficients $W_{i,\boldsymbol{k}}$ of w_i are replaced by $|V_{i,\boldsymbol{k}}|\phi_{\boldsymbol{k}}$, where $V_{i,\boldsymbol{k}}$ are the Fourier coefficients of v_i and $\phi_k = W_{i,\boldsymbol{k}}/|W_{i,\boldsymbol{k}}|$ contains the complex phase of $W_{i,\boldsymbol{k}}$;
> 4. constructing the new version of the surrogate as $z'_i = w'_i + \bar{x}$.

The residual vectors of SPEC surrogates have thus the same power spectra than the v_i, but the distributions of values are not the same.

MAAFT surrogates are built in order to share both the distribution of values of the original members and the power spectrum of the residual vectors. The construction starts as for SPEC surrogates and adds a fifth step:

5. constructing z_i'' by adjusting the distribution of values of z_i' to that of x_i: the highest value in z_i' is replaced by the highest value in x_i, the second highest in z_i' by the second highest in x_i, and so on.

Steps 2 to 5 are then cycled through for 30 iterations, using z_i'' of step 5 as the z_i of the next step 2.

This technique to generate surrogates is inspired from the Iterated Amplitude-Adjusted Fourier Transform (IAAFT) (Schreiber and Schmitz, 1996), and the only difference is that the power spectrum and the distribution adjustments are not done on the same quantity: the distribution is adjusted on the surrogate member z_i itself, while the power spectrum is adjusted on the residual $w_i = z_i - \bar{x}$ of the surrogate with respect to the ensemble mean. This allows to construct surrogate ensembles with the same power spectrum as the v_i and the same distribution of values for each member while keeping the information of the mean in the ensemble.

Referee's comment 7: *Regarding the discussion of phase around line 225, isn't this behaviour produced by design in STEPS?*

Yes, we mentioned this at lines 323-324 of the manuscript. However, a non-parametric method is now used to generate the noise of the STEPS AR(2) processes (Seed et al., 2013) so that the nowcasts better preserve the anisotropic structure of the rain field in addition to matching the power spectrum of observations. The STEPS ensembles in this work were produced with the latter method.

References

Oczkowski, M., Szunyogh, I., and Patil, D. J.: Mechanisms for the Development of Locally Low-Dimensional Atmospheric Dynamics, Journal of the Atmospheric Sciences, 62, 1135 – 1156, https://doi.org/10.1175/JAS3403.1, 2005.

Ono, K.: Assessing the Predictability of Mesoscale Moist Airflow Using the Ensemble Dimension, Weather and Forecasting, 40, 2601 – 2620, https://doi.org/10.1175/WAF-D-25-0042.1, 2025.

Patil, D. J., Hunt, B. R., Kalnay, E., Yorke, J. A., and Ott, E.: Local Low Dimensionality of Atmospheric Dynamics, Phys. Rev. Lett., 86, 5878–5881, https://doi.org/10.1103/PhysRevLett.86.5878, 2001.

Rotunno, R. and Snyder, C.: A Generalization of Lorenz's Model for the Predictability of Flows with Many Scales of Motion, Journal of the Atmospheric Sciences, 65, 1063 – 1076, https://doi.org/10.1175/2007JAS2449.1, 2008.

Satterfield, E. and Szunyogh, I.: Predictability of the Performance of an Ensemble Forecast System: Predictability of the Space of Uncertainties, Monthly Weather Review, 138, 962 – 981, https://doi.org/10.1175/2009MWR3049.1, 2010.

Satterfield, E. and Szunyogh, I.: Assessing the Performance of an Ensemble Forecast System in Predicting the Magnitude and the Spectrum of Analysis and Forecast Uncertainties, Monthly Weather Review, 139, 1207 – 1223, https://doi.org/10.1175/2010MWR3439.1, 2011.

Seed, A. W., Pierce, C. E., and Norman, K.: Formulation and evaluation of a scale decomposition-based stochastic precipitation nowcast scheme, Water Resources Research, 49, 6624–6641, 2013.

Snyder, C., Hamill, T. M., and Trier, S. B.: Linear Evolution of Error Covariances in a Quasi-geostrophic Model, Monthly Weather Review, 131, 189 – 205, https://doi.org/10.1175/1520-0493(2003)131<0189:LEOECI>2.0.CO;2, 2003.

Szunyogh, I., Kostelich, E. J., Gyarmati, G., Patil, D. J., Hunt, B. R., Kalnay, E., Ott, E., and Yorke, J. A.: Assessing a local ensemble Kalman filter: perfect model experiments with the National Centers for Environmental Prediction global model, Tellus A, 57, 528–545, https://doi.org/https://doi.org/10.1111/j.1600-0870.2005.00136.x, 2005.

Wei, M. and Toth, Z.: A New Measure of Ensemble Performance: Perturbation versus Error Correlation Analysis (PECA), Monthly Weather Review, 131, 1549 – 1565, https://doi.org/10.1175//1520-0493(2003)131<1549:ANMOEP>2.0.CO;2, 2003.