

Dear reviewer,

We are grateful for the time that you put into evaluating our manuscript and for your critical comments, which are very valuable to improve the manuscript. Our general author comment (a separate document) provides an overview of the modifications that will be implemented in the revised version of our manuscript. In this author comment we will address more specifically your major and conceptual comments. A more detailed point-by-point response will be given when submitting the revised version of the manuscript. Your comments are repeated in black while our comments are colored in blue.

This study examines the hypothesis posed by Lee and Held (1993) that “Because the packet can remain coherent despite chaotic internal dynamics, the packet envelope should be more predictable than the individual weather systems”. As the authors suggested, as far as I can tell, this hypothesis has not been quantitatively assessed up to now. The authors compared normalized error growth of the envelope and the underlying wave fields, and their results suggested that the envelope errors might reach saturation slightly faster than errors of the wave field, suggesting that the hypothesis posed above should be rejected. They also examined the mechanisms responsible for error growth and found that the same mechanisms (divergent flow and nonlinearity) dominate the error growth of both the envelope and the waves. They also suggested that higher amplitude waves may be more predictable even outside of the RWP envelopes. Overall, I think the topic is of interest to the forecasting and dynamics community, and this paper should eventually be published, but I do have some major issues that need to be resolved before this paper can be accepted for publication.

Major comments:

1) The authors made use of an approximation to the RWP envelope, E (defined in equation 1), which was computed based on along-latitude circle Hilbert transform of PV . Comparing this with the more commonly used RWP envelope function computed based on along-stream velocity (Fig. 3 vs. Fig. 1), E is clearly noisier, since it is computed from a noisier variable (PV), and without the along stream calculations to smooth out kinks due to the undulating properties of RWPs. It is not clear to me that the error characteristics of E are the same as those of the RWP envelope. In my opinion, the authors need to show that the error growth in the smoother RWP envelope is similar to that of E to make the results of this paper more convincing. This can probably be done by creating normalized error growth curves similar to Fig. 6b for the envelope function in place of E .

- The authors argued (line 130) the refinements would impact on both the envelope and the underlying wave pattern in a consistent manner”, which I do not agree with, since the along stream refinement by Zimin et al (2006) does not really affect the waves, but does serve to produce much smooth and more coherent envelope signals.

Thank you very much for being critical with our assumption that the method to extract the envelope would impact on the envelope and on the underlying wave in a consistent manner. Your critique was instrumental to improve the manuscript, yielding a more nuanced and arguably more interesting result. We still think that the anomalies of the underlying wave

become more coherent also when the wave signal is diagnosed with respect to streamlines of a background flow. We now agree with you, however, that the predictability of the envelope benefits more than that of the underlying wave. Our revised results indicate that enhanced predictability of the envelope is associated with the envelope values being high-amplitude anomalies with respect to climatology. It now seems plausible to us that using a more sophisticated method to extract a more coherent envelope signal also enhances the amplitude of this signal with respect to climatology, yielding higher predictability.

Unfortunately, we cannot provide an analysis dedicated to the impact of the method defining the RWP envelope. The wind data from which previous work has computed the envelope and the phase now used in the revised version is no longer readily available to us to compute an “approximation” of the RWP envelope (along latitude circles). And for PV, we are not aware of a method similar to that of Zimin et al. 2006. In addition, evaluating phase and envelope errors within RWP objects defined by the Zimin et al. 2006 method introduces further difficulties for a fair comparison. As shown in the original version, the agreement between these RWP objects and approximate (PV) envelopes is good only in a qualitative sense. RWP objects thus sample a wider range of (PV) envelope values, whereas the (wind) envelopes defined by the Zimin et al. 2006 method by design sample values in the high-percentile range only. Finally, as noted in your comment, PV has more power than wind at the higher wavenumbers in the considered wavenumber range, reducing predictability of the PV envelope in comparison to the wind envelope due to the general scale dependence of predictability. We thus refrain from a more detailed comparison of predictability of the “approximated” PV RWP envelopes and of the RWP wind envelopes.

In the revised version, the “approximated” PV envelopes will be considered to examine differences in the error-growth dynamics of the envelope and the phase. The revised version will be clear that the PV envelopes are an approximation to the RWP (wind) envelopes only and that their predictability characteristics are different, in particular in the sense that in the PV framework the envelope does not exhibit higher predictability than the phase during the lead times under consideration (not shown). To gain insight into the error-growth dynamics during early, “high-skill” lead times (see general author comment), it may seem to be justified to accept the approximation.

Also, even if we apply the Wolfe and Wirth (2015) transform to compute the envelope, it is not clear that to quantify error growth of the underlying wave field, we should consider the error growth of the transformed wave field rather than the error growth of the original wave field.

We agree that there is no clear answer to the question how to define an “individual weather system” in the sense of Lee and Held (1993). In the revised version, we compare predictability of the (RWP) envelope and the phase of the (underlying) carrier wave. Because both envelope and phase are direct functions of the analytical signal Z (see general author comment), which is computed from the filtered wave signal, it seems most natural to compare the predictability of phase and envelope subject to the same filter procedure. Our revised version will briefly reflect this discussion when introducing the PV perspective and in the final discussion of the results.

2) The authors justified using MSE-type error metrics instead of using object-based methods for verifying coherent RWP features stating that “It is not clear to us, however, how a comparable object-based method could be designed for the underlying wave pattern” (line 50). I’m not sure I understand this argument, since many previous studies have examined troughs and ridges using feature-based methods. In fact, the hypothesis posed by Lee and

Held (1993) relates to “individual weather systems”, which, at least to me, suggested that the comparison should be between RWPs and troughs and ridges as objects, rather than the “wave” field. Hence, to me, this study does not really directly assess the hypothesis posed by Lee and Held (1993). My interpretation of their hypothesis would be something like “Can RWPs be predicted further ahead than individual troughs and ridges”?

We agree that troughs and ridges can, of course, be identified in an object-based framework. Upon reflection, we thus agree that the underlying wave pattern could be identified in an object-based framework, too. We still refrain from object-based verification in this study, though. First, as a major drawback of object-based verification, there is ambiguity in matching objects at lead times when skill has substantially decreased. Second, from the computational perspective for our climatological study, we prefer the efficiency of MSE-based metrics. Third, we perform conditional verification, an aspect that was not stressed in the original version but will be in the revised version. For our main result, we restrict verification to the region of RWP objects. Within this region, the envelope is a smoothly varying field. (The phase is a smoothly varying field anyways.) The envelope is thus not subject to double penalty that arises when a distinct feature is erroneously displaced within the verification domain. We will revise our argument in favor of the MSE perspective in the revised version accordingly.

We further agree with your interpretation of the Lee and Held hypothesis. Our general author comment argues that our revised version does address this very question (see Fig. R2 and related discussion in that document).

3) In section 4.3, the authors claimed that higher amplitude PV anomalies (in terms of E) exhibit higher predictability even outside of RWPs. The following may be due to my misunderstanding. In comparing Fig. 1 and Fig. 3, it seems to me that areas outside of RWPs (defined in Fig. 1) with very high values of E (shown in Fig. 3) may frequently be located just outside of the RWP objects and are likely related to PV anomalies that are associated with the RWP, so it is not clear to me that these should be considered as being “outside” of the RWPs. Did the authors make sure that what they considered “outside” are really not associated with RWPs, and how?

We investigated maps of the occurrence frequency of high-amplitude anomalies outside of RWPs and compared them with maps of RWP occurrence frequency. We found spatially distinct patterns which led us to the conclusion that the high-amplitude anomalies outside of RWPs are not associated with RWPs. The revised version will no longer compare the predictability of PV anomalies inside and outside of RWP objects. The main idea of our associated result, however, will be retained: High-amplitude anomalies exhibit higher predictability than their smaller-amplitude counterparts, and RWPs constitute such high-amplitude anomalies.

4) To me, the example shown in Fig. 9 is potentially misleading. Let us consider the simplest wave packet, which is formed by the superposition of 2 waves: $\cos(\phi_1) + \cos(\phi_2)$. Mathematically, this is identical to $2 \cdot \cos(\phi_1 - \phi_2) \cdot \cos(\phi_1 + \phi_2)$, where $\phi_1 = k_1 x - \omega_1 t + C_1$, and $\phi_2 = k_2 x - \omega_2 t + C_2$, where the k 's are the wavenumbers, and the ω 's are the angular frequency. This is in the form of a wave packet, with the first factor being the envelope (lower frequency and smaller wavenumber), and the second factor the carrier wave (higher frequency and higher wavenumber). The authors' example is similar to the case in which the errors of ϕ_1 and ϕ_2 are of opposite signs. This makes the error in $\phi_1 + \phi_2$ small, while the error in $\phi_1 - \phi_2$ will be large. The result is that the envelope will shift, but the phase of the carrier wave is not shifted, which is what is shown in Fig. 9. In this case, the error in the

underlying wave pattern will largely be amplitude errors instead of phase errors, while clearly the envelope error is a phase shift. This example is in fact inconsistent with the authors' contention that for the underlying wave, phase errors dominate over amplitude errors. Now in the case in which the errors in ϕ_1 and ϕ_2 are of the same sign, then the phase error in the carrier wave will be amplified, while the phase error of the envelope will be small. So, while this example illustrates that phase errors in the wave field could lead to envelope errors, to me, this is not necessarily a typical case, and does not really "explain" the results found in this paper, unless the authors could demonstrate that the phase errors in the different wavenumbers are correlated with each other in a manner that is consistent with their contrived example.

Our intention with this schematic was not to imply that errors in the phase of the underlying wavenumbers project *predominantly* on the envelope. Our intention was rather to demonstrate that phase errors of the individual (monochromatic) underlying waves *may* project onto envelope errors. In consideration of the reviews, in particular comments by Maarten Ambaum, we have realized that it is confusing to consider the phase of the individual (monochromatic) underlying waves in the context of RWPs, because the wave packet's carrier wave is also characterized by 'its own' phase. The revised version follows the more common approach to consider the phase of the carrier wave. The schematic Fig. 9 will be omitted from the revised version of the manuscript.

5) Regarding mechanisms responsible for error growth (Section 4.2), before I read the paper, I expected that since RWP propagation is linked to wave dispersion, the largest errors in RWP envelopes would likely be due to mechanisms that act to disrupt wave dispersion, as well as mechanisms that are generally not well handled by models, which, to me, would suggest that the error growth would be dominated by nonlinearity as well as the impact of diabatic heating (or DIV). I expect linear dynamics to be better handled by model forecasts, so is it really surprising that the dominant mechanisms for error growth in both the envelope and wave fields are nonlinearity and divergent flow?

Thank you very much for your insightful thoughts regarding error-growth mechanisms. We will adopt your interpretation regarding the disruption of wave dispersion in our revised version.

The revised version investigates the differences between the error-growth dynamics of the envelope and of the phase (see our general author comment). We find again that nonlinearity dominates error growth in general for both envelope and phase. We further find that the divergent flow makes an important contribution to the error growth of both in the first few days of lead time. Our results thereby fully support your expectation about dominant growth mechanisms. We now, however, find also substantial differences. The phase (within RWPs) is less affected by nonlinearity than the envelope for early lead times, but more so for later lead times. The phase (within RWPs) is furthermore less affected by the divergent flow than the envelope for early lead times. We observe these differences only within RWPs/ for high percentiles of the envelope, but not for small-amplitude waves. Arguably, these differences are blurred in the original manuscript for the underlying wave pattern, which can be considered a combination of both phase and amplitude information.

6) In many of the comparisons (e.g. Fig. 6) there were no estimations of uncertainties, so it is not clear to me whether the differences between the different curves are significant or not. The authors should conduct an uncertainty analysis to examine how significant the results are.

The revised version will use two-sided, paired t-tests to assess the statistical significance of the main results (see Fig. R3 and R6 in our general author comment). The statistical uncertainty of our empirical estimate of the saturation level will be given in terms of the relative standard error of the mean, which is below 2.5% and 3.5% for the envelope and the phase, respectively, for all lead times.

Other comments: [We will respond to these comments when we submit the revised version of the manuscript.](#)

a) Lines 15-16: After reading this paper, it seems to me that this is merely a hypothesis posed near the end of the paper (lines 373-375) without much support from the results presented. As such, this should probably not belong to the abstract.

b) Line 155: “where the amplitude of E exceeds” – is that the instantaneous value of E? In some discussions I thought it might be the climatological value of E. Can the authors please clarify?

c) Line 170-171: “whereas the climatological mean of E is comparable to that of the variance”. Not sure what this means. E and its variance would have different dimensions. How can they be comparable?

d) Equation 12: Divergence instead of gradient?

e) Line 230: “tropospheric” – should be tropopause?