We are grateful to the referee for devoting their time to our manuscript. The valuable comments and suggestions will help us to improve the paper.

We will here respond to the main comments made:

In the abstract, the authors wrote there is an intrinsic limit of predictability after 22 days, this conclusion is made by fitting the modified power law function to the ECWMF data. If the function form in Zhang et al 2019 is used, then the predictability limit becomes 15 days as mentioned in the text. The difference between these two estimated limits is sort of large. In the context of the ECMWF operational forecast system, the reviewer did not find any advantage of using the modified power law function rather than the function form in Zhang et al. Is there any reason for the reader to believe that this 22-day limit is more accurate than the 15-day limit?

This is indeed a valid and relevant question. As we write in our manuscript (lines 416-424), from the data of error growth experiments performed with the ECMWF forecast system, it cannot be decided which law for the error growth is more appropriate. We provided some evidence that the extended power law has some theoretical justification and matches well the observations from our toy-model. The two competing error growth laws fitted to the ECMWF data yield different results for the prediction horizon, namely 15 versus 22 days. Due to the theoretical justifications behind the extended power law, we are inclined to follow its results and to be optimistic. Please also consider that the current forecasts do lose their skill after about 10-15 days, and that the quadratic law is a fit to such forecasts without extrapolation to smaller initial condition errors. Our 22 days are an upper bound for the prediction horizon for hypothetical forecasts with perfect initial conditions. Also, these 22 days are in nice agreement with Krishnamurthy (2019).

Scale-dependence is the key for understanding atmospheric predictability. The authors proposed this three-scale toy model. Could it be possible to verify this three-scale model with the ECMWF data? E.g., to connect X1 with synoptical errors, X2 with meso-scale error and X3 with turbulent motions. If such filters are applied to the ECMWF data and verify the errors of different scales with the toy model, then the results would be much more convincing.

We agree that it would be nice to build such a three-scale model from the ECMWF model.
However, we do not think that reality has only 3 scales. The atmosphere has a large variety of spatial and temporal scales, and even a crossover from essentially 2-dimensional transport on very large spatial scales to 3-d on small scales. Also, from the practical point of view, we do not know how to do that: We could adapt and then apply our filters to the ECMWF model in order to construct 3 data sets representing different scales. But that would require a suitable fine-tuning of parameters in our filters, and therefore would be a project of its own which will be worth to be started. Our toy model only represents something like a latitude circle and not the globe. Hence, our toy-model is much simpler than any atmospheric model and serves as a kind of didactical example of for multi-scale models. However, we have a set of similarities with real atmospheric models which we want to stress here again:

1. Our variable \( X_1 \) has 5 to 7 main highs and lows that correspond to planetary waves (Rossby waves) and several smaller waves corresponding to synoptic-scale waves.
2. Parameters of our three-scale toy model are chosen in order for the medium scale (\( X_2 \)) amplitude to be approximately ten times smaller than the large scale amplitude and the small scale amplitude (\( X_3 \)) again approximately ten times smaller than the medium scale amplitude. Our variables also have different oscillation periods (lines 168-170), which can be roughly compared to synoptic, meso, and turbulent scales (lines 39-47).
3. Bednar (2020) showed a similarity in the error growth of one-scale toy model (Eq. (2)) and ECMWF data.

Once the parameter of the three-scale toy model is set, then its error growth behavior is also determined. Is the results shown here sensitive to the value of the parameters in the equation (e.g. \( F, b, c, I \))?

We have chosen parameters such that all levels behave chaotically (the largest Lyapunov exponent of each level is positive) and that all levels have a significant difference in amplitudes and fluctuation rates (lines 151-153). For such conditions, the general results are not sensitive to parameter values in quality, but the values of fit-coefficients of the different error growth laws will differ.

Line 14-15: a theoretical justification of function form in Zhang et al. is recently provided by Sun and Zhang 2020 (https://doi.org/10.1175/JAS-D-19-0271.1).

We are grateful for pointing out this reference, of which we were not aware. We changed lines 14 - 15 to: Although the quadratic hypothesis cannot be completely rejected and could serve as a first guess, the hypothesis parameters are not theoretically justifiable in the model. We added the mentioned citation: Lines 348; 351; 405 and 474-475.

Line 180: perfect model assumption is used, right?

Yes, we added this to our text (Line 182).

Line 214: what does this initial transient behavior look like? Is the error decreasing with time? What would happen if the initial error is further reduced towards 0?

Yes, before the perturbed trajectory has relaxed back to the attractor, the error is decreasing with time. When the magnitude of the initial perturbation is reduced, the relaxation time of the perturbed trajectory also decreases. We were using initial error magnitudes which were optimized for having minimal but non-zero errors after the end of the transient.

A bracket is missing in Equation. (20)

Many thanks, we corrected this.
Line 365: the extended quadratic model Eq. (21) is more accurate than the extended exponential growth? How to reconcile this with the three-scale toy model results?

Originally, the extended exponential error growth was designed to describe the error growth and error growth rate in 1-dimensional models, where “extended” means that it also captures the saturation of error growth on large scales. So the extended exponential model was made for a scale-independent error growth rate, while for our 3-scale model, we had implemented the “extension” to saturation into the power law error growth. Since the ECMWF systems exhibit multi-scale dynamics with scale dependent error growth, the extended exponential error growth model is less accurate than the extended quadratic model, but it was our intention to compare the observed error growth to the extended power law. Bednar et al. (2020) showed that the extended quadratic model Eq. (21) is more accurate than the extended exponential growth for ECMWF data, but in that work the extended power law was not tested.

Line 415-425: The authors seem to hint that the extended power-law form is better compared to the extended quadratic form here. This is true for the toy model. But it is not supported by the ECMWF, right?

Yes, this is right, for the ECMWF model, both fits are of similar quality. In lines 415-425 we argue what we can conclude if we assume that nonetheless the extended power law is the correct description of the data.

References: