# SPECTRAL ANALYSIS REVEALS A PHYSICAL MECHANISM FOR INSTABILITY IN CONVOLUTION-BASED DATA-DRIVEN WEATHER MODELS

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

Data-driven weather prediction (DDWP) with deep learning (DL) models has seen a surge of interest in recent times. While these models are still not as accurate as operational weather forecasting models, recent successes have seen DDWP models being comparable to them in terms of short-term accuracy at orders-of-magnitude less computational cost. These successes of DDWP models promises us of improved performance in terms of subseasonal-to-seasonal and extreme weather predictions. However, one major drawback of DDWP models that currently hinders this, is the absence of long-term stability (when integrated autoregressively) and a physical interpretation of the mechanism and cause of this instability. In this paper, we show that a spectral analysis in Fourier space reveals a physical mechanism by which convolution-based DDWP models become unstable. We systematically explore the physics by which instability is introduced in these models by exploring the connection between the representation skills of these networks in Fourier space and a consistent growth of meridional heat flux that leads to an increase in transfer of eddy momentum flux to the mean flow. We further speculate how convolution-based models may have inductive bias that would always lead to such instability and further explorations in novel loss functions or architectures should be explored for autoregressive predictions of weather, climate, and generally fully turbulent flow.

## 1 INTRODUCTION

The increase in interest around building data-driven weather prediction (DDWP) models with deep learning (DL) is primarily due to the low computational cost of DDWP models as compared to operational numerical weather prediction (NWP) models (Schultz et al., 2021; Balaji, 2021; Reichstein et al., 2019). Such DDWP models can facilitate generating a large number of ensembles (owing to their low computational cost) that can enable both probabilistic forecasting to improve subseasonal-to-seasonal prediction (Weyn et al., 2021) or data assimilation (Chattopadhyay et al., 2021). These data-driven models are usually either trained on reanalysis data which are similar to observations and are generally limited in quantity (Weyn et al., 2020; Chattopadhyay et al., 2021; Pathak et al., 2022) or pre-trained on a large amount of climate model outputs and fine-tuned on observations/reanalysis (Rasp & Thuerey, 2021). The main motivation behind building DDWP models is that: (1) if trained (or fine-tuned) on observational data, the DDWP models would not suffer from the same biases that often exist in physics-based weather/climate models due to inaccurate parameterization of small-scale and/or poorly-understood processes in the atmosphere and (2) DDWP models may seamlessly be integrated to obtain seasonal and climate forecasts, which are generally not possible with NWP models. Current DDWP models while coming close to (and sometimes outperforming in case of certain small-scale variables such as precipitation (Pathak et al., 2022)) NWP models in terms of short-term forecast accuracy, cannot remain stable for long-term autoregressive integration to yield physically-consistent climatology. While various solutions have been proposed in the past to alleviate instability in these DDWP models (Chattopadhyay et al., 2020; Weyn et al., 2021), a physical interpretation of the mechanism that induces instability in these models has not been explored.

In this paper, we consider convolution-based DDWP models that are developed on data simulated from a simple two-layer quasi-geostrophic (QG) flow which represents the dynamics of the mid-

latitude of Earth's atmosphere. The QG system, while being simpler than a real climate model, allows us to carefully interpret the DDWP models in terms of the physics that the system represents. We consider two set-ups, (1) a base system, that mimics a climate model that has incorrect representation of certain physics, but from which we have a large amount of data and (2) a target system, resembling the true atmosphere from which we have a limited amount of observations.

In this paper, we consider a few DDWP models (differing in architecture and other details in the set-up) that are pre-trained on $M_{tr}$ samples from the base system and then fine-tuned on $N_{tr}$ samples from the target system ($N_{tr} \ll M_{tr}$). We interpret the cause and provide a mechanism for the instability that arises in these DDWP models by representing their outputs in Fourier space during autoregressive predictions. This reveals a consistent mis-representation of the small-scale part of the Fourier spectrum which results in an inverse cascade of energy which ultimately results in the instability. We further show that the best (in terms of short-term accuracy) and the worst DDWP models are governed by how well they represent the large-scale spectrum while remaining equally incompetent in representing the small-scale spectrum which is responsible for instability.

## 2 Systems and Models

### 2.1 QG systems

The non-dimensional dynamical equations of the two-layer QG flow have been developed following Lutsko *et al*. Lutsko et al. (2015), and Nabizadeh *et al*. Nabizadeh et al. (2019). The model consists of two constant density layers with a $\beta$-plane approximation in which the meridional temperature gradient is relaxed towards an equilibrium profile. The model equation is described as

$$\begin{aligned}\frac{\partial q_k}{\partial t} + J(\psi_k, q_k) = &-\frac{1}{\tau_d}(-1)^k\left(\psi_1 - \psi_2 - \psi_R\right) \\ &-\frac{1}{\tau_f}\delta_{k2}\nabla^2\psi_k - \nu\nabla^8 q_k.\end{aligned} \tag{1}$$

Here, $q$ is potential vorticity and is expressed as

$$q_k = \nabla^2\psi_k + (-1)^k\left(\psi_1 - \psi_2\right) + \beta y, \tag{2}$$

where $\psi_k$ is the stream function of the system. In Eqs. (1) and (2), $k$ denotes the upper ($k = 1$) and lower ($k = 2$) layers. $\tau_d$ is the Newtonian relaxation time scale while $\tau_f$ is the Rayleigh friction time scale, which only acts on the lower layers represented by the Kronecker $\delta$ function ($\delta_{k2}$). $J$ denotes the jacobian, $\beta$ is the $y-$gradient of the coriolis parameter, and $\nu$ denotes the hyperdiffusion coefficient. Our base system has a baroclinically unstable jet at the center of a horizontally (zonally) periodic channel by setting $\psi_1 - \psi_2$ to be equal to a hyperbolic secant centered at $y = 0$. When eddy fluxes are absent, $\psi_2$ is identically zero, making zonal velocity in the upper layer $u(y) = -\frac{\partial\psi_1}{\partial y} = -\frac{\partial\psi_R}{\partial y}$ where we set

$$-\frac{\partial\psi_R}{\partial y} = \operatorname{sech}^2\left(\frac{y}{\sigma}\right), \tag{3}$$

$\sigma$ being the width of the baroclinic region. Further details of the parameters can be found in Nabizadeh *et al.* (Nabizadeh et al., 2019).

$\beta = 0.192$ for the base system. On the other hand, the target system has a double jet configuration by changing the equilibrium jet profile of $u(y) = \operatorname{sech}^2(\frac{y-10}{\sigma}) + \operatorname{sech}^2(\frac{y+10}{\sigma})$ with an increased value of $\beta^* = 3\beta$. We assume that we have long-term climate simulations of the base system while we have only a few observations of the target system. The difference between the base and target systems is shown in Fig. 1.

### 2.2 DDWP models

In this paper, we train several DDWP models ($\mathbf{P}$) on the base system, to predict the state of the QG flow (in this case, $\psi_k$) from time step $t$ to $t + 40\Delta t$. Here, we investigate a fully convolutional neural

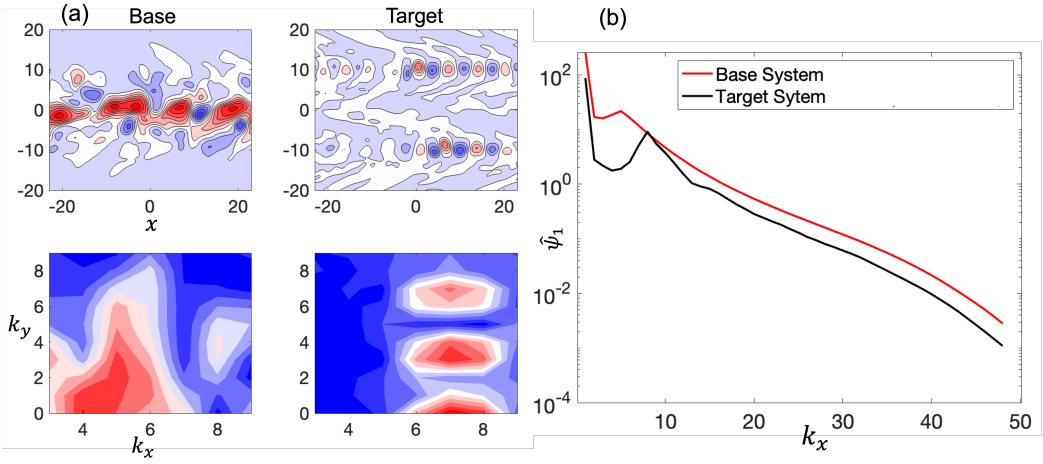


Figure 1: Base and Target systems for the QG flow. (a) Top rows shows individual snapshots of $\psi_1$, while the bottom panels show the 2D Fourier spectrum of $\psi_1$. (b) Zonally averaged Fourier spectrum ($\hat{\psi}_1$) of $\psi_1$.

| **Model Name** | **Architecture** | **Type of Loss** |
|---|---|---|
| M1 | Fully convolutional | single time-step |
| M2 | U-NET | single time-step |
| M3 | U-NET | multi-time-step |

Table 1: Various model configurations.

network and a U-NET with two different configurations, (1) a regular single time-step loss and (2) a multi-time-step loss. For (1), we compute the model output $\psi_k(t+40\Delta t) = \mathbf{P}\left(\psi_k(t)\right)$ and optimize the loss $\mathcal{L} = ||\psi_k^{true}(t+40\Delta t) - \psi_k(t+40\Delta t)||_2$ (true labels obtained from the base system). For (2), we obtain the model outputs at two successive time steps, $\psi_k(t+40\Delta t) = \mathbf{P}\left(\psi_k(t)\right)$ and $\psi_k(t+80\Delta t) = \mathbf{P}\left(\mathbf{P}(\psi_k(t))\right)$. Following this, we optimize the combined losses at these two time steps, $\mathcal{L} = ||\psi_k^{true}(t+40\Delta t) - \psi_k(t+40\Delta t)||_2 + ||\psi_k^{true}(t+80\Delta t) - \psi_k(t+80\Delta t)||_2$ (Weyn et al., 2020). We have implemented the multi-time-step loss for only the U-NET architecture, although it can be implemented for the fully convolutional network as well and the conclusions will not change. For fine-tuning on the small number of observations on the target system, we perform transfer learning by retraining the first 3 layers of each of the architectures. These set-ups for each of the architectures have been shown in Table 1. Throughout the paper, we would refer to BNN (base neural network) as the DDWP model trained on the base system and applied to the target system *without* fine tuning on observations from the target system and TLNN (transfer learnt neural network) as the DDWP model that is fine-tuned on a small number of observations on the target system. For fine-tuning, we have considered, $N_{tr} = M_{tr}/10$.

## 3 Results

### 3.1 Mis-representation of small scales in Fourier space

In this section, we show the short-term performance of all the models M1, M2, and M3 in terms of RMSE error as well the zonally averaged Fourier spectrum of the predicted $\psi_1$ as compared to the truth. Figure 2 shows the relative short-term performances of the different models. TLNN-M3 and TLNN-M2 shows the best performance while BNN-M1 shows the worst performance. However, the best and the worst model have equally bad representation skills in the small-scale part of the

spectrum even at day 1 where the RMSE error for the predictions are small for all the models (Fig. 2 (b)). This mis-representation of small-scale eddies pushes energy to the large scales through an inverse cascade that increases the overall energy of the predicted outputs.

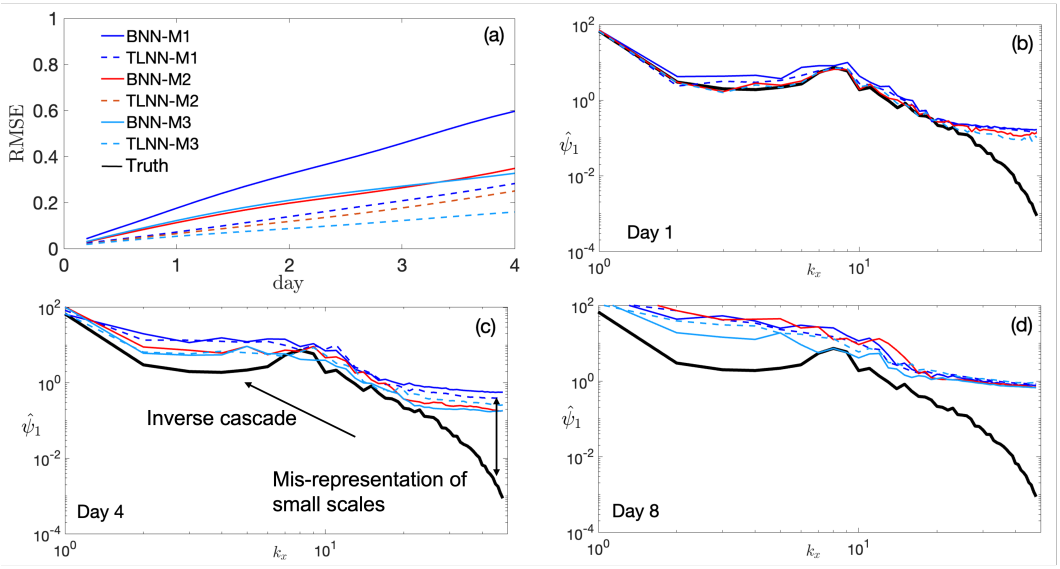


Figure 2: Short-term predictions of M1, M2, and M3 for both BNN and TLNN. (a) RMSE of different models. (b) Zonally averaged Fourier spectrum ($\hat{\psi}_1$) of $\psi_1$ for different models at day 1. (c) same as (b) for day 4. (d) same as (b) for day 8.

### 3.2 Inaccurate Baroclinic Waves' Life Cycle

The increase in value of small-scale $\hat{\psi}_1$ indicates an increase in the energy of the small-scale eddies. This affects the generation of eddies and consequently the eddy life-cycle. The mis-representation of the small-scale eddies results in the inaccurate transport of meridional heat flux, $J = (\psi_2 - \psi_1)\, v$, where $v$ is the meridional velocity of the flow. This transport of meridional heat flux would transfer energy back to the mean flow (Thompson & Woodworth, 2014) through the eddy momentum flux, $u^{'} v^{'}$. Figure 3 shows that all the models show an increase in $J$, while the true $J$ remains roughly constant. It is however to be noted, that even though some of the models, e.g., TLNN-M2 shows good short-term performance would still become unstable due to the increase in $J$. It also shows, that relative to TLNN-M2, TLNN-M1 which is a poorer model in terms of short-term performance, might remain stable for longer, despite losing accuracy much more quickly. Figures 2 – 3 also show that short-term accuracy is determined by how well the small wavenumbers are captured while long-term instability is caused due to mis-representation of large wavenumbers that ultimately results in larger growth of $J$.

In Fig. 4, we show how the increase in $J$ for various models lead to the increase in zonally averaged eddy momentum flux, $\langle u'v' \rangle$, where $\langle . \rangle$ refers to zonal averaging. This eddy momentum flux is transferred to the mean flow that leads to unstable predictions. It should be noted that both BNN-M3 and TLNN-M3 show an abrupt growth in eddy momentum consistent with the abrupt increase in $J$ although they have good short-term predictions.

## 4 Conclusion

In this paper, we have interpreted the cause and proposed a physical mechanism by which instability is induced into DDWP models. We have shown that, consistent with established physics, a mis-representation of small-scale eddies results in the growth of meridional heat flux which increases the convergence of eddy momentum flux near the jets that leads to instability. We have further shown that the best DDWP model in terms of short-term accuracy may not always be the most stable one,

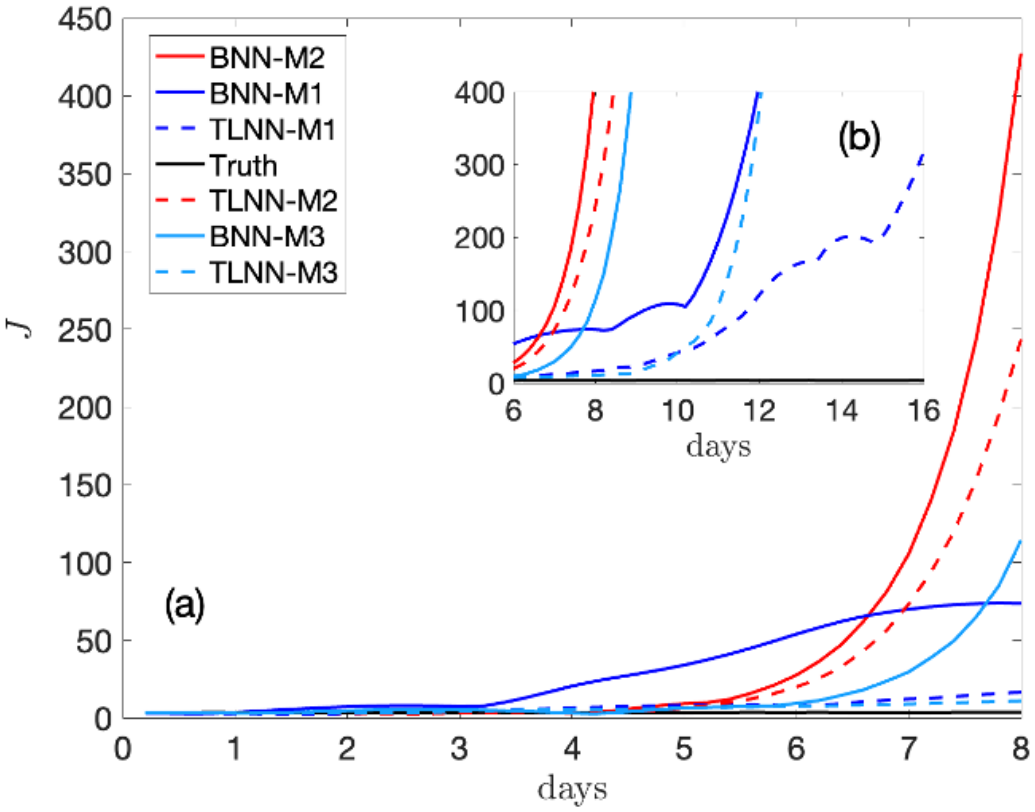


Figure 3: Meridional heat flux, $J$, as a function of days for various models. (a) Growth of $J$ until 8 days. (b) Growth of $J$ between 6 and 16 days. Model TLNN-M3 although shows good short-term performance results quickly in growth of $J$ resulting in instability. All models show significant deviation of predicted $J$ from true $J$.

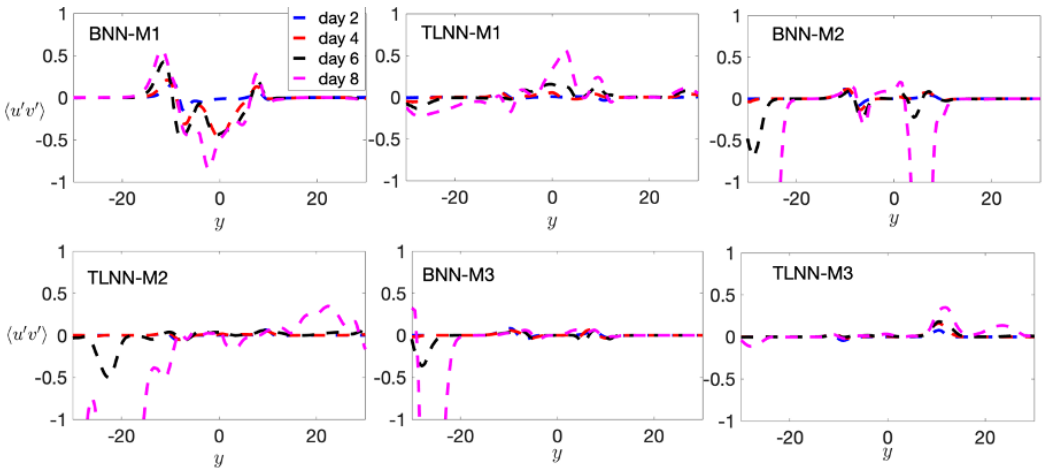


Figure 4: Zonally averaged eddy momentum flux ($\langle u'v' \rangle$) for different models at various days up to 8 days. Increase in momentum flux leads to unstable blow-up.

and that all the analyzed models are infact, equally incompetent in representing small-scale eddies which are responsible for long-term instability.

## REFERENCES

V Balaji. Climbing down charney's ladder: machine learning and the post-dennard era of computational climate science. *Philosophical Transactions of the Royal Society A*, 379(2194):20200085, 2021.

Ashesh Chattopadhyay, Mustafa Mustafa, Pedram Hassanzadeh, and Karthik Kashinath. Deep spatial transformers for autoregressive data-driven forecasting of geophysical turbulence. In *Proceedings of the 10th International Conference on Climate Informatics*, pp. 106–112, 2020.

Ashesh Chattopadhyay, Mustafa Mustafa, Pedram Hassanzadeh, Eviatar Bach, and Karthik Kashinath. Towards physically consistent data-driven weather forecasting: Integrating data assimilation with equivariance-preserving spatial transformers in a case study with era5. *Geoscientific Model Development Discussions*, pp. 1–23, 2021.

Nicholas J Lutsko, Isaac M Held, and Pablo Zurita-Gotor. Applying the fluctuation–dissipation theorem to a two-layer model of quasigeostrophic turbulence. *Journal of the Atmospheric Sciences*, 72(8):3161–3177, 2015.

Ebrahim Nabizadeh, Pedram Hassanzadeh, Da Yang, and Elizabeth A Barnes. Size of the atmospheric blocking events: Scaling law and response to climate change. *Geophysical Research Letters*, 46(22):13488–13499, 2019.

Jaideep Pathak, Shashank Subramanian, Peter Harrington, Sanjeev Raja, Ashesh Chattopadhyay, Morteza Mardani, David Hall Thorsten Kurth, Zongyi Li, Kamyar Azizzadenesheli, Pedram Hassanzadeh, Karthik Kashinath, and Animashree Anandkumar. Fourcastnet: A global data-driven high-resolution weather model using adaptive fourier neural operators. *arXiv preprint arXiv:2202.11214*, 2022.

Stephan Rasp and Nils Thuerey. Data-driven medium-range weather prediction with a resnet pretrained on climate simulations: A new model for weatherbench. *Journal of Advances in Modeling Earth Systems*, pp. e2020MS002405, 2021.

Markus Reichstein, Gustau Camps-Valls, Bjorn Stevens, Martin Jung, Joachim Denzler, Nuno Carvalhais, et al. Deep learning and process understanding for data-driven earth system science. *Nature*, 566(7743):195–204, 2019.

MG Schultz, C Betancourt, B Gong, F Kleinert, M Langguth, LH Leufen, Amirpasha Mozaffari, and S Stadtler. Can deep learning beat numerical weather prediction? *Philosophical Transactions of the Royal Society A*, 379(2194):20200097, 2021.

David WJ Thompson and Jonathan D Woodworth. Barotropic and baroclinic annular variability in the southern hemisphere. *Journal of the Atmospheric Sciences*, 71(4):1480–1493, 2014.

Jonathan A Weyn, Dale R Durran, and Rich Caruana. Improving data-driven global weather prediction using deep convolutional neural networks on a cubed sphere. *Journal of Advances in Modeling Earth Systems*, 12(9):e2020MS002109, 2020.

Jonathan A Weyn, Dale R Durran, Rich Caruana, and Nathaniel Cresswell-Clay. Sub-seasonal forecasting with a large ensemble of deep-learning weather prediction models. *arXiv preprint arXiv:2102.05107*, 2021.