FourCastNet Summary 2

Wen-wen Tung et al.

Abstract

▶ FourCastNet: Fourier ForeCasting Neural Network, is a global data-driven weather forecasting model

▶ accurate short (0.5-2.5-day) to medium-range (3-7-day) global predictions at 0.25° resolution.

    ▶ high-resolution, fast-timescale variables: surface wind speed, precipitation, and atmospheric water vapor.

▶ It has important implications for planning wind energy resources, predicting extreme weather events such as tropical cyclones, extra-tropical cyclones, and atmospheric rivers.

▶ Performance matches the forecasting accuracy of the ECMWF Integrated Forecasting System (IFS) at short lead times for large-scale variables, while outperforms IFS for small-scale variables, including precipitation.

▶ FourCastNet generates a week-long forecast in less than 2 seconds, orders of magnitude faster than IFS.

    ▶ enables rapid and inexpensive large-ensemble forecasts with thousands of ensemble-members for improving probabilistic forecasting.

Traditional Models

▶ The beginnings of modern numerical weather prediction (NWP) can be traced to the 1920s

▶ The quality of weather forecasts has been steadily improving over the past decades (c.f. Bauer et al. [2015], Alley et al. [2019]).

▶ In addition to better computing capabilities, improvements in weather forecasting have been achieved through better parameterization of small-scale processes through deeper understanding of their physics and higher-quality atmospheric observations.

  ▶ The latter has resulted in improved model initializations via data assimilation.

Data-Driven Weather Models

▶ There is now increasing interest around developing data-driven Deep Learning (DL) models for weather forecasting owing to their orders of magnitude lower computational cost as compared to state-of-the-art NWP models [Schultz et al., 2021, Balaji, 2021, Irrgang et al., 2021, Reichstein et al., 2019].

▶ Data-driven models have great potential to improve weather predictions by overcoming model biases present in NWP models and by enabling the generation of large ensembles at low computational cost for probabilistic forecasting and data assimilation.

▶ Most data-driven weather models, however, use low-resolution data for training, usually at the $5.625°$ resolution as in Rasp and Thuerey [2021b] or $2°$ as in Weyn et al. [2020]

▶ For data-driven models to be truly impactful, it is essential that they generate forecasts at the same or greater resolution than current state-of-the-art numerical weather models, which are run at $\approx 0.1°$ resolution.

Approach of FourCastNet

- ▶ global data-driven forecasts of key atmospheric variables at a resolution of 0.25°
  - ▶ which corresponds to a spatial resolution of roughly 30 km × 30 km near the equator
  - ▶ a global grid size of 720 × 1440 pixels.
- ▶ allows making a direct comparison with the high-resolution Integrated Forecasting System (IFS) model of the European Center for Medium-Range Weather Forecasting (ECMWF).
- ▶ FourCastNet is about 45,000 times faster than traditional NWP models on a node-hour basis.
- ▶ The energy required to train FourCastNet is approximately equal to the energy required to generate a 10-day forecast with 50 ensemble members using the IFS model.
- ▶ Once trained, however, FourCastNet uses about 12,000 times less energy to generate a forecast than the IFS model.

Example
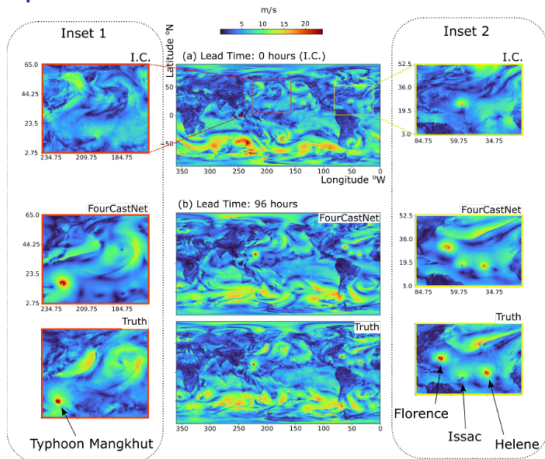
Figure 1. global near-surface wind forecast generated by FourCastNet initialized with an initial condition from the out-of-sample test dataset on September 8, 2018 at 00:00 UTC. The model was run for 16 time steps (6 hours × 16) from this initial condition in inference mode.

FourCastNet Blurb

- ▶ a Fourier transform-based token-mixing scheme [Guibas et al., 2022] with a vision transformer (ViT) backbone [Dosovitskiy et al., 2021].
- ▶ This approach is based on the recent Fourier neural operator that learns in a resolution-invariant manner and has shown success in modeling challenging partial differential equations (PDE) such as fluid dynamics [Li et al., 2021a].
- ▶ ViT backbone is capable of modeling long-range dependencies well.
- ▶ Combining ViT with Fourier-based token mixing yields a state-of-the-art high-resolution model that resolves fine-grained features and scales well with resolution and size of dataset.

Training FourCastNet

- ECMWF Reanalysis ERA5
  - hourly from 1979 to the present day
  - at horizontal resolution of $0.25° \times 0.25°$
  - 37 vertical levels from the surface of the earth to about 100 km in height
- FourCastNet focuses training on
  - 10-m wind velocities
  - 6-hourly total precipitation
  - Model also forecasts geopotential height, temperature, wind velocity, and relative humidity at a few different vertical levels, surface pressure and mean-sea-level pressure, and total column water vapor.
- Rationale
  - surface influences from small-scale and fast physical processes might be where FourCastNet can shine
  - previous DL-based weather prediction have not been able to produce global forecasts for these variables at full ERA5 resolution

FourCastNet Model Description

- ▶ Adaptive Fourier Neural Operator (AFNO) model [Guibas et al., 2022]
  - ▶ specifically designed for high-resolution inputs
  - ▶ synthesizes several key recent advances in DL into one model: Fourier Neural Operator (FNO) learning approach of Li et al. [2021a] shown to perform well in modeling challenging PDE systems, with a powerful vision transformer (ViT) backbone.
- ▶ ViT in AFNO
  - ▶ ViT's performance is attributed to *multi-head self-attention mechanism* in these networks, allowing the network to model interactions between features (tokens) globally at each layer in the network.
  - ▶ Spatial mixing via self-attention is quadratic in the number of tokens, and thus quickly becomes computationally infeasible.
  - ▶ AFNO model frames the mixing operation as continuous global convolution, implemented efficiently with FFTs
  - ▶ AFNO model thus allows modeling dependencies across spatial and channel dimensions flexibly and scalably.
  - ▶ Thus, spatial mixing complexity is reduced to $O(NlogN)$, where N is the number of image patches or tokens.

Procedure

▶ First, the input variables on the 720×1440 lat-lon grid are projected to a 2D grid ($h \ times w$) of patches (with a small patch size $p \times p$, where e.g., $p = 8$), with each patch represented as a d-dimensional token.

▶ Then, the sequence of patches are fed, along with a positional encoding, to a series of AFNO layers.

▶ Each layer, given an input tensor of patches $X \in \mathbb{R}^{h \times w \times d}$, performs spatial mixing followed by channel mixing.

FourCastNet Model Architecture
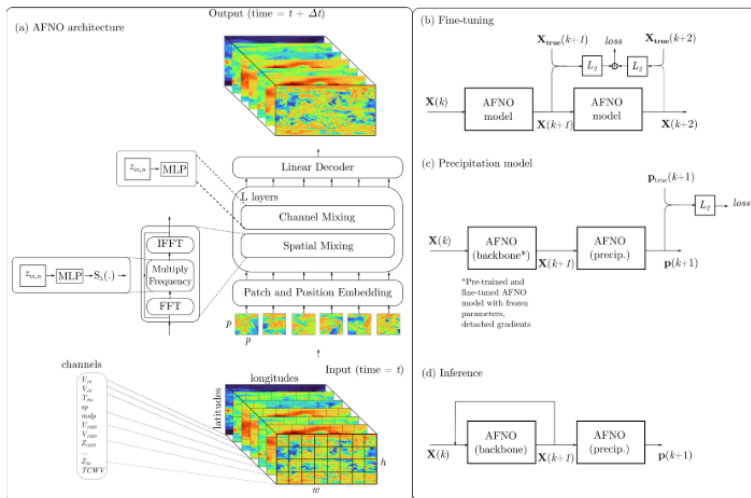
Figure 2. The Multi-layer transformer architecture

Results Section 3

Inset 1

Inset 2
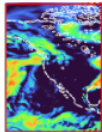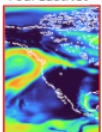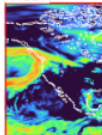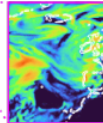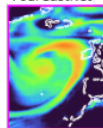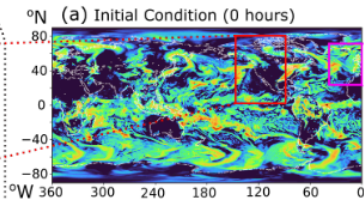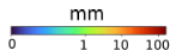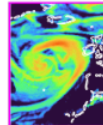
mm

0   1   10   100

°N (a) Initial Condition (0 hours)

I.C.

I.C.

FourCastNet

(b) Lead Time: 36 hours
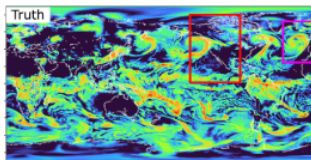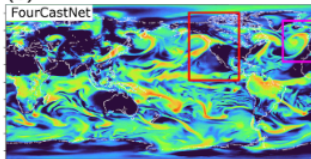
FourCastNet

FourCastNet

Truth

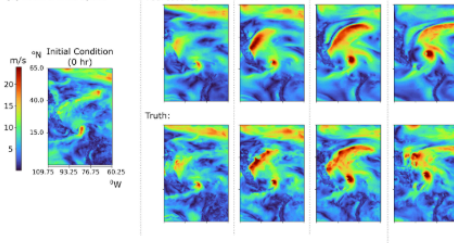Truth

Truth

°W 360   300   240   180   120   60   0
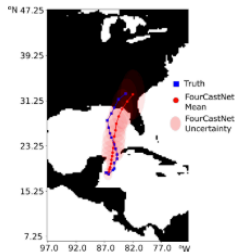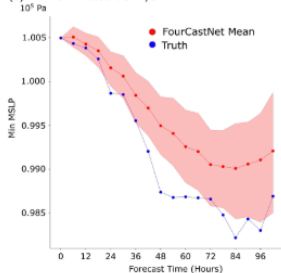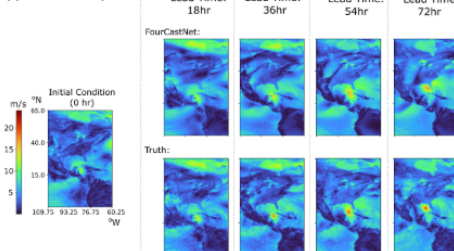
Results Section 3.1

Figure 4. Hurricanes

Results Section 3.2
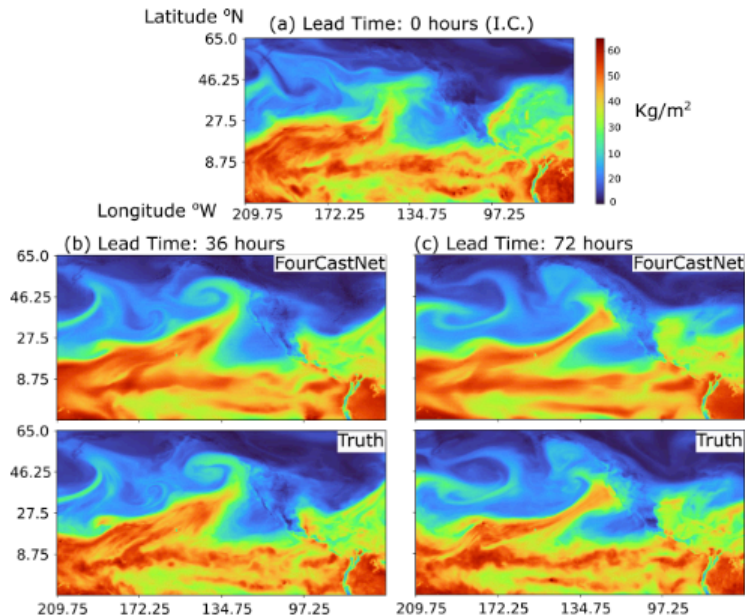
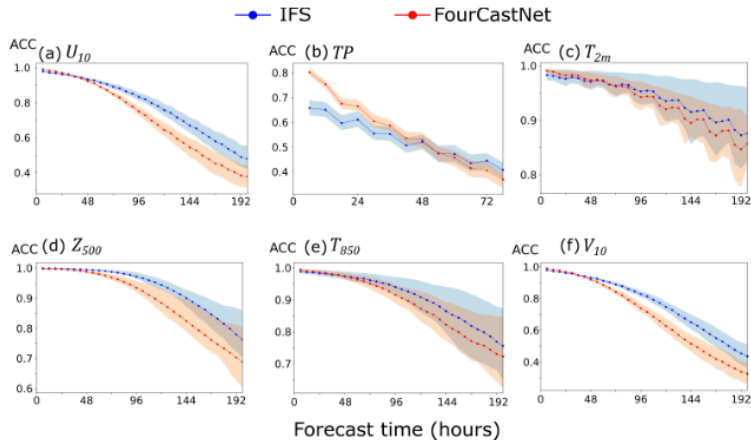Figure 5. Atmospheric Rivers

Results Section 3.3 Quantitative Skill

Figure 6. Latitude weighted ACC
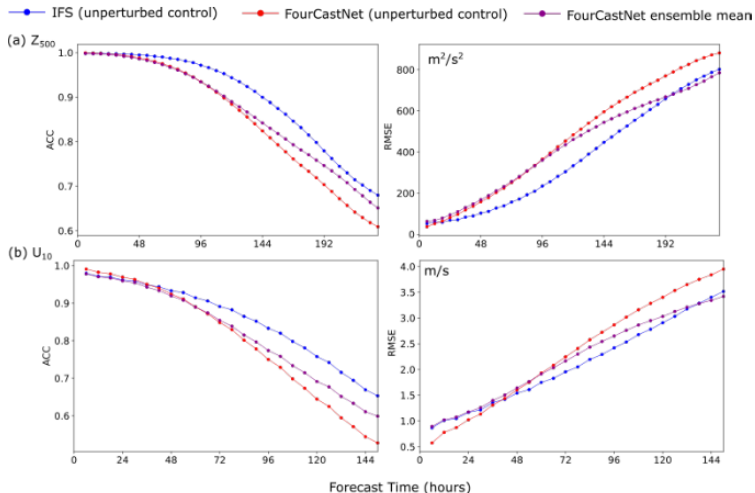
Results Section 3.4 Ensemble Forecasts

Figure 7. Large ensembles

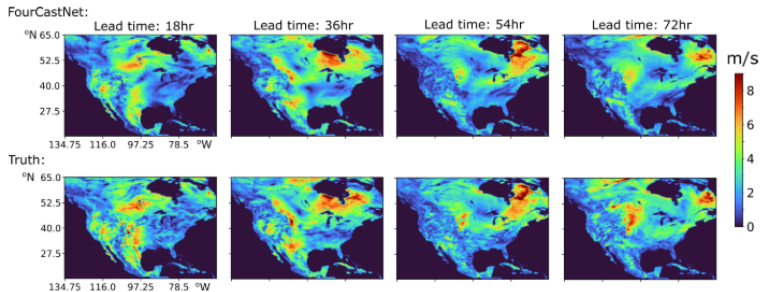# Results Section 3.5 Forecast Skill Over Land For Near-surface Wind Speed

Figure 8. Skill Over Land

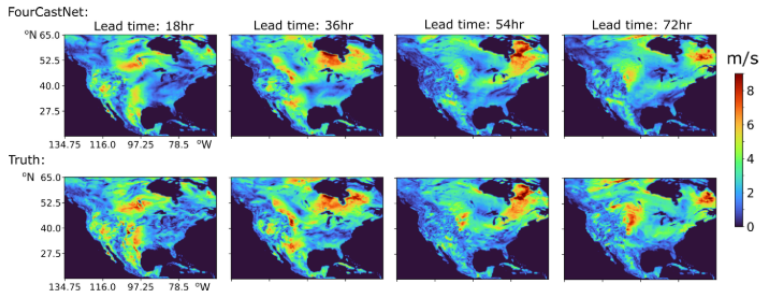# Results Section 3.5 Forecast Skill Over Land For Near-surface Wind Speed
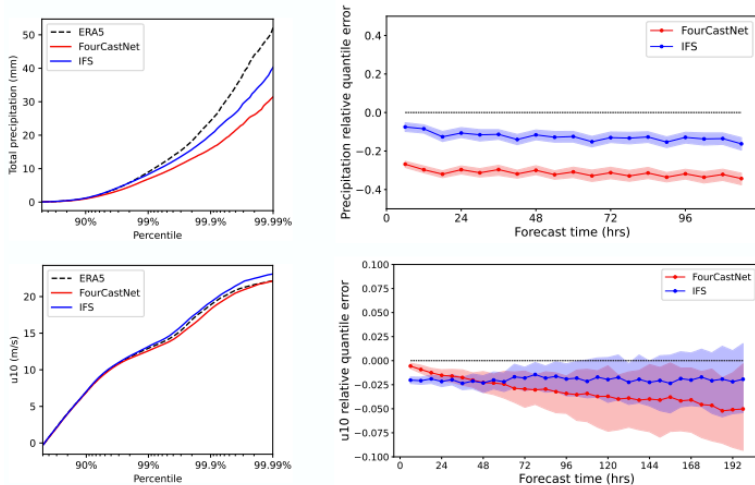

Figure 8. Skill Over Land

Results Section 3.6 Extremes

Figure 9. Extreme percentiles