A High-Performance Computing System for Probabilistic Weather Forecasts

Weiming Hu[1], Guido Cervone[1], Vivek Balasubramanian[2], Matteo Turilli[2], and Shantenu Jha[3]

[1]Pennsylvania State University Main Campus
[2]Rutgers University
[3]Rutgers University Newark

November 23, 2022

Abstract

Numeric weather prediction is undergoing a revolution resulting from the continuous advances in scientific knowledge and technologies. With dozens of weather models emerging that all generate different predictions from each other, forecasts have been gradually shifting from a deterministic form to a probabilistic form which shows the increasing concerns of, not just the absolute prediction values, but the confidence of predictions and the uncertainty of models. As a computational problem, generating uncertainty information can be an expensive task. Conventionally, prediction models are initiated with slightly perturbed parameters and then the diversion of model results can be a measure of model uncertainty. However, the multi-simulation approach drastically increases the computational requirement so that it can potentially exceed the ability of the state-of-art high-performance computing platforms. Meanwhile, if spatial and temporal resolutions are of concern, this approach is far from being efficient and viable. The Parallel Ensemble Forecast system is designed to generate probabilistic weather forecasts by using the revolutionary numerical weather prediction technique, Analog Ensemble. It is a data-driven method that derives probability information of a deterministic prediction model using past forecasts and observations without multiple simulation runs. Integrated with high-performance platforms, the system distributes computational tasks among nodes and therefore further boosts the data simulation process.

A High-Performance Computing System for Probabilistic Weather Forecasts

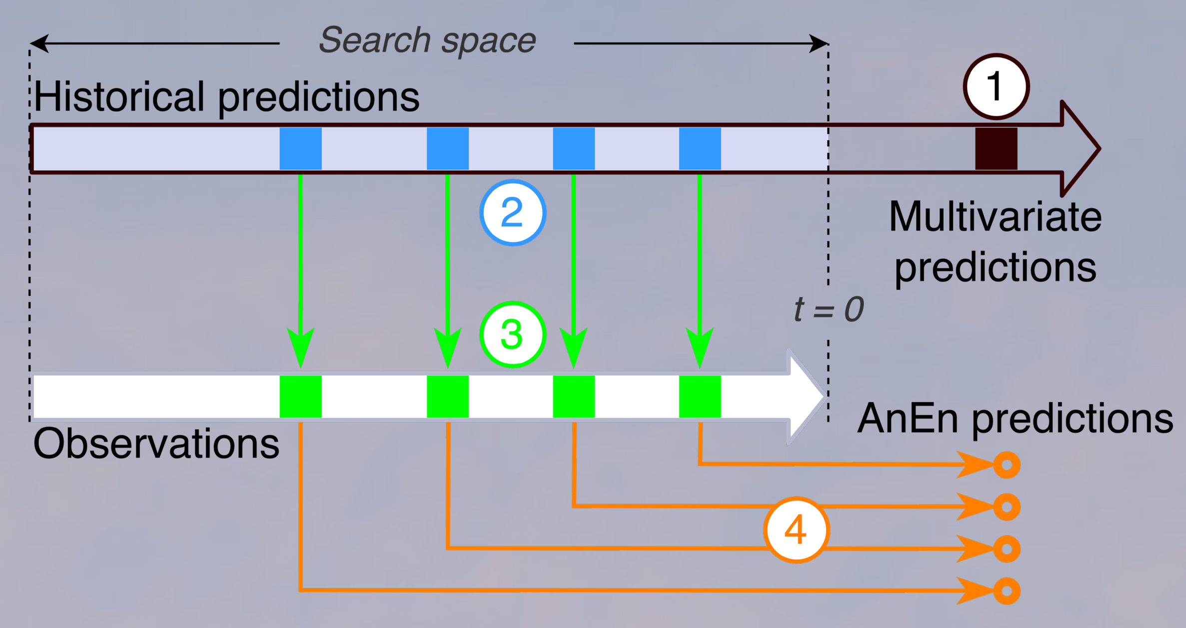
Weiming Hu*, Guido Cervone*, Vivek Balasubramanian†, Matteo Turilli†, Shantenu Jha†
*Geoinformatics and Earth Observation Laboratory, weiming@psu.edu
Dept. of Geography and Institute for CyberScience, The Pennsylvania State University
†Research in Advanced Distributed Cyberinfrastructure and Applications, The State University of New Jersey
PENNSTATE Institute for CyberScience
RUTGERS

Introduction

Numeric weather prediction is undergoing a revolution as a result of continuous advances in scientific knowledge and technologies. With the emergence of numerous weather models, the model uncertainty and the forecast probability are receiving growing attention. Conventionally, weather models are initiated with slightly perturbed parameters and then the diversion of model results can be an estimation of model uncertainty. However, this multi-simulation approach drastically increases the computation requirement so that it can potentially exceed the capacity of the state-of-art high-performance computing systems.

To approach to the above problems, this poster presents the coupled system of the Analog Ensemble (AnEn), a probabilistic forecasting technique, and the Ensemble Toolkit (EnTK), an application that provides scalable, inter-operable and sustainable solutions on high-performance computing platforms.

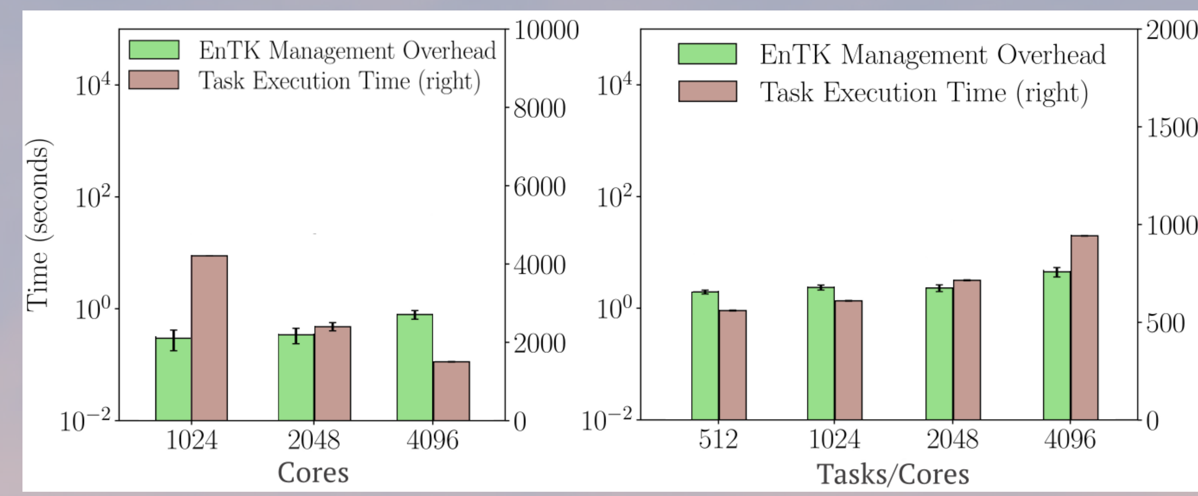
Analog Ensemble



1. Starts with a current deterministic prediction, a set of historical predictions, and the corresponding observations;
2. Finds the most similar historical predictions to the current prediction based on a multivariate metric;
3. Takes the observations associated with the historical predictions;
4. Generates the prediction ensemble for the current time.

Ensemble Toolkit

- Abstracts execution and resource management
- Supports multi-node and GPU tasks
- Supports heterogeneous computing infrastructures
- Scales up to 10^5 tasks on diverse HPC systems
- Fault-tolerant at scale



Acknowledgments

Ensemble Toolkit is available on Github at http://radicalentk.readthedocs.io/.

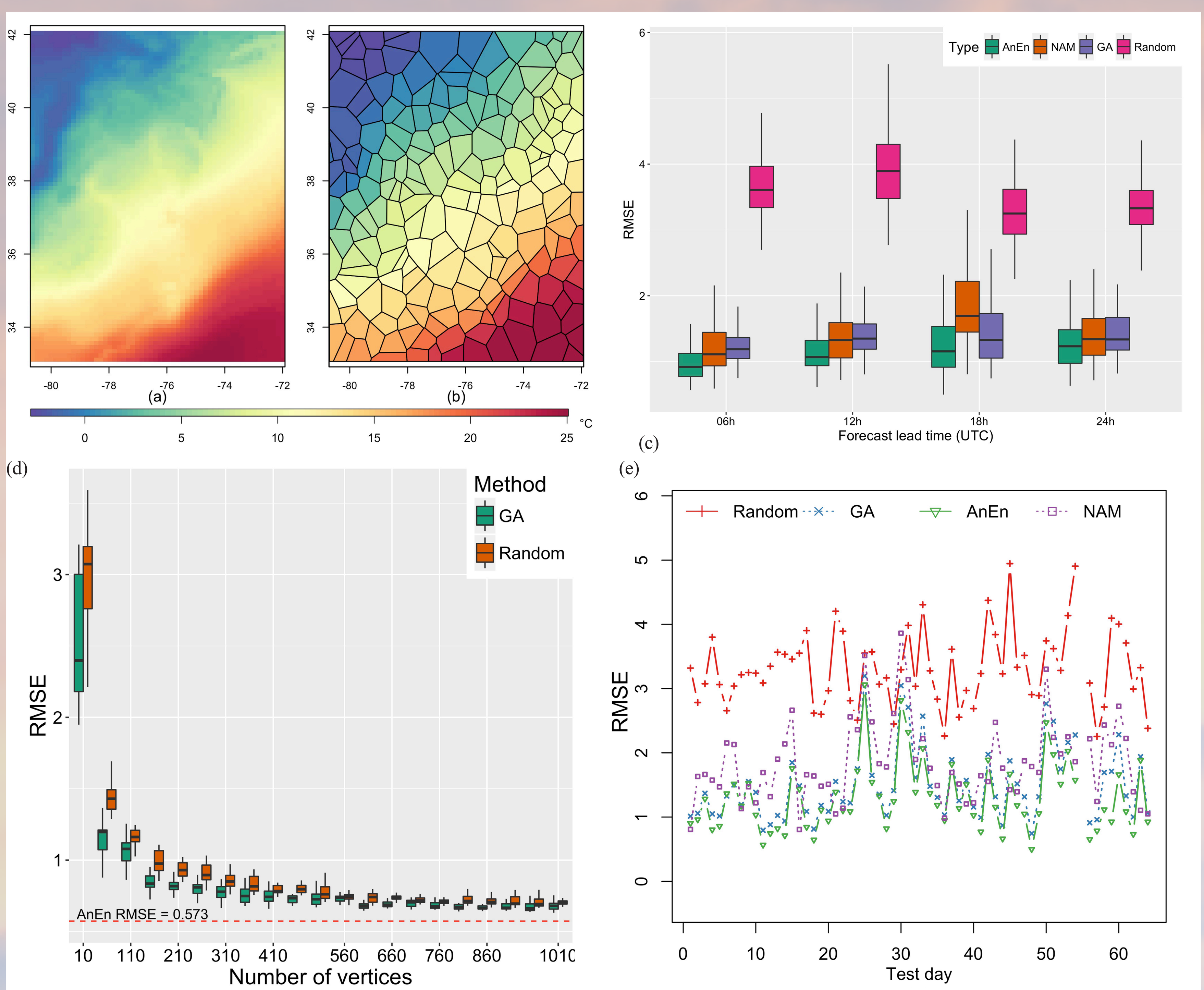
Analog Ensemble package is available for download on Github at https://weiming-hu.github.io/AnalogsEnsemble/.

This work is funded by NSF program and EarthCube.

Use Case 1 - Dynamically Optimized Unstructured Mesh

Since different regions of interest might have different weather variability, the computation requirement varies with locations. Together with a Genetic Algorithm (GA), an optimized unstructured mesh can be trained to minimize the overall forecast accuracy. It saves computation by generating AnEn on fewer vertices than a regular mesh would have.

North American Mesoscale model forecasts and analysis between 2016 to 2018 are collected. The following figures show the results for short-term temperature predictions for 24-hour ahead with a 6-hour interval.



Lessons Learned

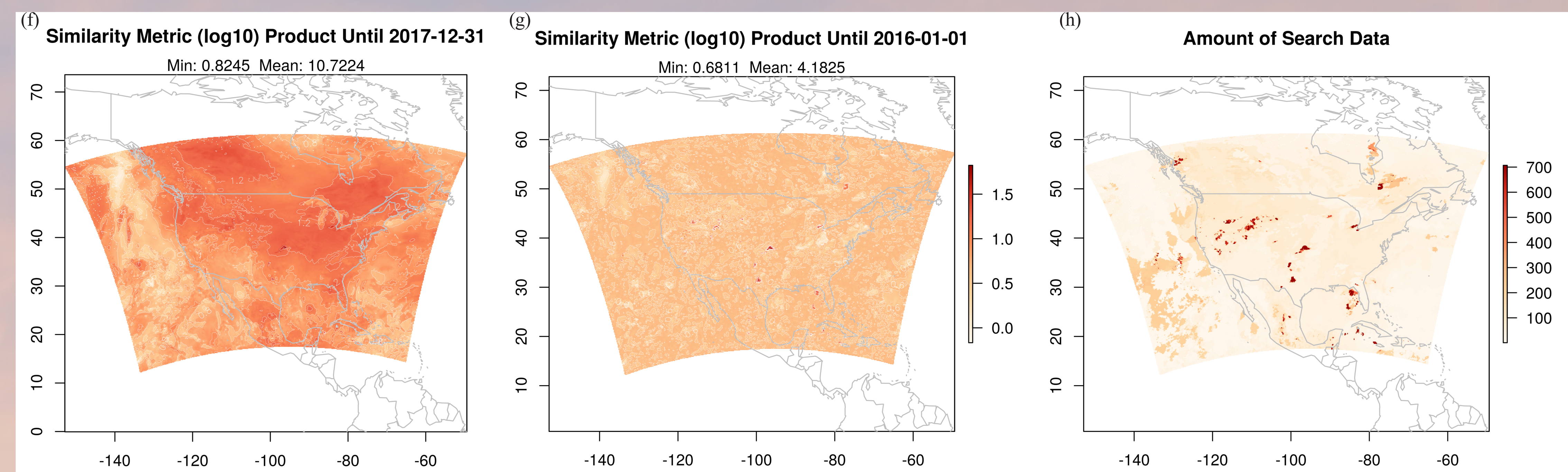
- The overall prediction accuracy can be improved by dedicating computing resources to particular regions.
- Using an unstructured mesh constrains the error propagation with time.
- An unstructured mesh helps balance between the amount of resources required and the desired prediction accuracy.

Legend

- *NAM: NAM on a regular grid;*
- *AnEn: AnEn on a regular grid;*
- *GA: AnEn on a GA-generated mesh;*
- *Random: AnEn on a random mesh.*

Use Case 2 - Adaptive Search with a Spatial Metric

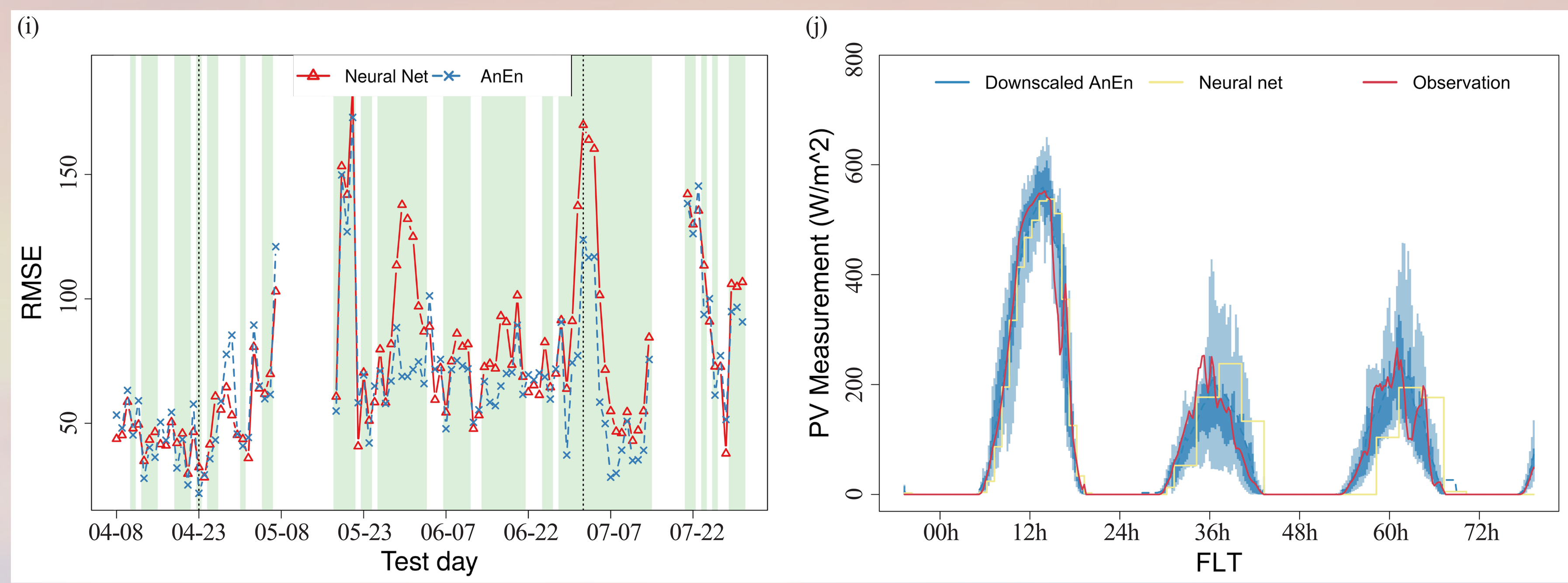
In theory, a perfect analog can always be found with an infinite amount of historical forecasts. In practice, different regions might require a different amount of historical forecasts to find good analogs. This study investigates an adaptive algorithm that allows a subregion to stop searching when a good analog has already been found.



Lessons Learned

- Most grid points can find good analogs within relatively small historical data.
- Computation can be dedicated to the rest of the grid points where good analogs are harder to find.

Use Case 3 - High-Resolution Photovoltaic Energy Production Forecast



AnEn is applied to generating 3-day forecasts with a 3-hour interval for photovoltaic (PV) energy production and then used to downscale the forecasts to the temporal scale of the power measurement. Results are compared with an Artificial Neural Network (ANN).

Lessons Learned

- AnEn outperforms ANN on 67% of days in the test dataset of year 2018 on a small scale PV panel that is mounted on a personal weather station.
- The downscaling technique using AnEn shows promising results in correcting model output and capturing shor-term PV energy production variability.

Balasubramanian, Vivekanandan, et al. "Ensemble toolkit: Scalable and flexible execution of ensembles of tasks." Parallel Processing (ICPP), 2016 45th International Conference on. IEEE, 2016; Cervone, Guido, et al. "Short-term photovoltaic power forecasting using Artificial Neural Networks and an Analog Ensemble." Renewable Energy 108 (2017): 274-286.