Finding Hidden Patterns in High Resolution Wind Flow Model Simulations

Students: Yuqiao Zhao, Ming Chi Wong (City University of Hong Kong), Yulin Wei (Chinese University of Hong Kong)
Anna Fortenberry, Spencer Smith (University of North Texas)
Mentor: Dr. Kwai Wong

Introduction

Background

Accurate microscale wind flow dynamic data is essential to the design of wind farms. Wind flow dynamics can be highly sensitive to terrain irregularities, and wind conditions may drastically change from one location to another over small distances.

Datasets

ERA5: a global weather model with a resolution around thirty kilometers. Derived from boundary conditions of ERA5 data, we have **Large Eddy Simulation (LES)** datasets, which simulate turbulence at a reasonable cost.

Questions

Exploratory data analysis, dimensionality reduction of the grid and visualization of latent space, and upscaling from a low resolution to high resolution grid.

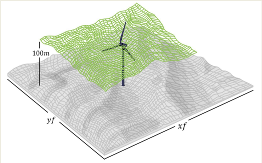

Fig 1: Visualization of data: terrain-following slice

Methodology

Add labels

To interpret the features of the output latent spaces, we first added labels to the original high resolution LES dataset. The selected scale for this process was wind power density because it is important data to the wind industry. Additionally, it combines the *temp* and *vel* variables for the calculation. There are eight wind power density classifications. The criteria are based on the classification method established by Onea et al. We added an eighth label for when the wind power density was over 1148.75, labeled C8.

Overall Mthods

We used **Pearson Correlation Coefficient** to check linear correlation between ERA5 and LES datasets. **Principal Component Analysis (PCA)** and **t-distributed stochastic neighbor embedding (t-SNE)** were implemented for compression. Additionally, an **integrated 2D and 3D CNN**, an **interpolation model**, and a **ResUnet model** were compared for the upscaling task.

- **t-SNE** The two parameters we adjusted were perplexity and learning_rate, which related to the shape of latent space. Larger datasets usually have a lager perplexity, and larger learning rate usually preserves original shape better.
- **3DCNN** After constructing the model, a few optimizations were performed. Gradient computation was turned off during validation, some parts of the data were successfully run in parallel on GPU, automatic mixed precision was used, and different learning rates were tested. The code is available at https://github.com/CheukHinHoJerry/3DCNN-SUPER-2021-pytorch.
- **Interpolation** We had more success in our approach with the interpolation method. The resize() function in OpenCV allowed us to test each of the three flags by simply setting interpolation to each one. We tried nearest-neighbor, bilinear, and bicubic interpolation.
- **ResUnet** The ResUnet proved to be problematic, as the CNN did. When we attempted to train the model with the full dataset, the model failed.

Results

PCA preserves most information of original datasets and t-SNE extract the features of the original data.

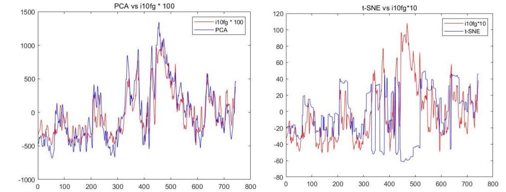

Fig 2: 1D latent space by PCA (left) and t-SNE (right) (blue line) compare with enlarged ERA5 (red line)

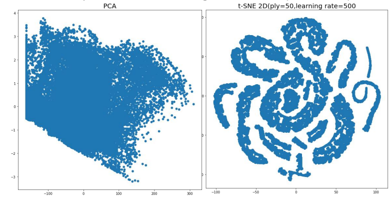

Fig 3: 2D latent space by PCA (left) and t-SNE (right)

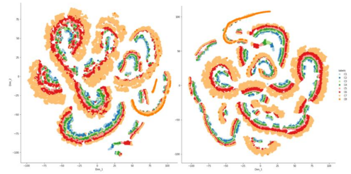

Fig 4: 1-hour data with 6 variables (left) and 4 variables (right)

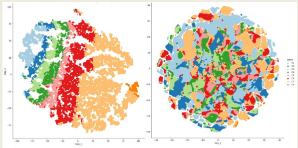

Fig 5: with 5 variables one-hour data (left) and 24-hour data (right)

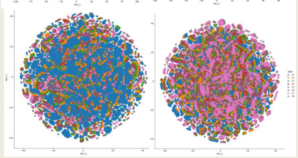

Fig 6: with 4 variables 24-hour data on 01-14(left) and 01-22(right)

The six variables are *temp, vel, u, v, std*, and *absolute_height*. The five variables are without *absolute_height*, and four variables are without *u, v*. The shape of latent space with different clusters, as Fig 4 shows, is due to the *absolute_height*. Therefore, without it, the magnitude of wind power density for one-hour data can be observed easily (decreases from left to right in the left part of Fig 5). Although the latent space of 24-hour data is like a sphere, Fig 6 also illustrates the strength of wind power for one day (blue represents poor wind and pink is for superb wind).

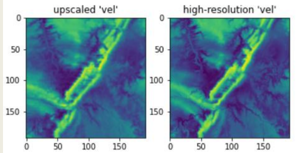

Fig 7: : left: upscaled by cubic interpolation right: original high-resolution

The PSNR of nearest-neighbor, bilinear and cubic interpolations are 23.48, 24.12, and 24.13, respectively. Their SSIM is 0.64, 0.67, and 0.68. Figure 7 compares the upscaled low-resolution LES with the *vel* variable by cubic interpolation and the original high-resolution LES.

Future Work

Use other methods such as wavelet-based CNN for super resolution.
Explore more compression techniques and interpret latent space better.

Acknowledgements

This project was sponsored by the National Science Foundation through Research Experience for Undergraduates (REU) award, with additional support from the Joint Institute of Computational Sciences at University of Tennessee Knoxville. This project used allocations from the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by the National Science Foundation. In addition, the computing work was also performed on technical workstations donated by the BP High Performance Computing Team.

References

Maaten, L. V., Hinton, G (2008).: Visualizing Data using t-SNE. Jornal of Machine Learning Research, vol 9.
Kasper, D.: Wind Energy and Power Calculations.
Ho, C., Li, J (2021).: Video Super Resolution with Integrated 2D and 3D Convolution Neural Network.
OpenCV resize image using cv2.resize(). TutorialKart. (2020).
Hall, N (2021).: Earth Atmosphere Model. NASA. .
Onea, F., Ruiz, A., Rusu, E (2020).: An Evaluation of the Wind Energy Resources along the Spanish Continental Nearshore. Energies.