Data Challenge 3: Finding Hidden Patterns in High Resolution Wind Flow Model Simulations

Smoky Mountains Computational Sciences Data Challenge (SMCDC22)
Proposed by OAK RIDGE National Laboratory
Team: Yuqiao Zhao, Anna Fortenberry, Yulin Wei, Wilson Wong, Spencer Smith
Mentor : Dr. Kwai Wong

Contents

Introduction

Background

Wind flow dynamics at **micro-scale** are of paramount importance in the wind energy industry.

- Historically, wind farm designs rely on precise measurements from few meteorological masts over an entire site.
- However, at microscale, wind flow dynamics can be very sensitive to the terrain irregularities and wind conditions can drastically change from one location to another even over small distances.
- **Computational Fluid Dynamics (CFD)** is a promising approach for assessing atmospheric flow properties over a domain of interest.
- **Large Eddy Simulation (LES)** is one of the most advanced mathematical models used in CFD for resolving turbulences at a reasonable cost.

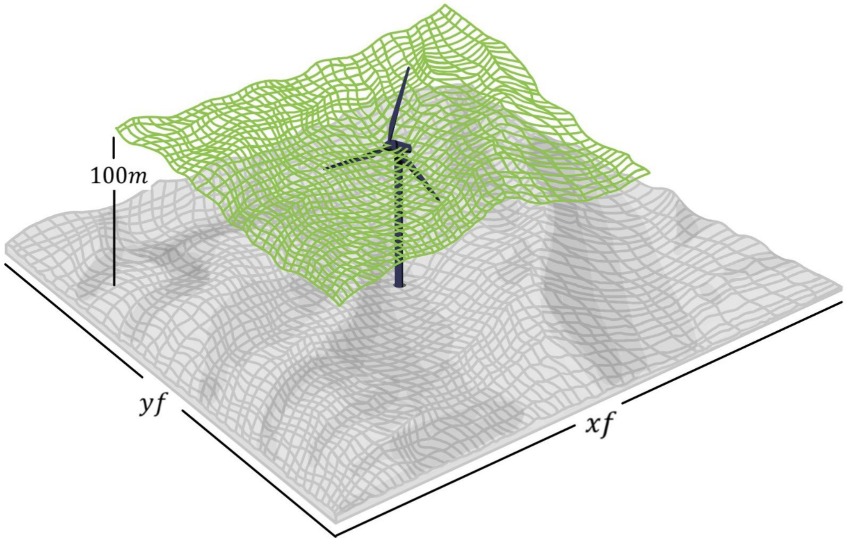


Fig 1: Visualization of data: terrain-follow

Introduction

Data sources

1. **ERA5:** a global weather model at a resolution of ~30km with hourly estimates of atmospheric variables. For each site provided, the corresponding timeseries is provided.
2. **Large Eddy Simulation (LES):**
   a simulation model driven by boundary conditions derived from ERA5 data and then resolve the local wind farm site wind field at much higher resolution in space and time.

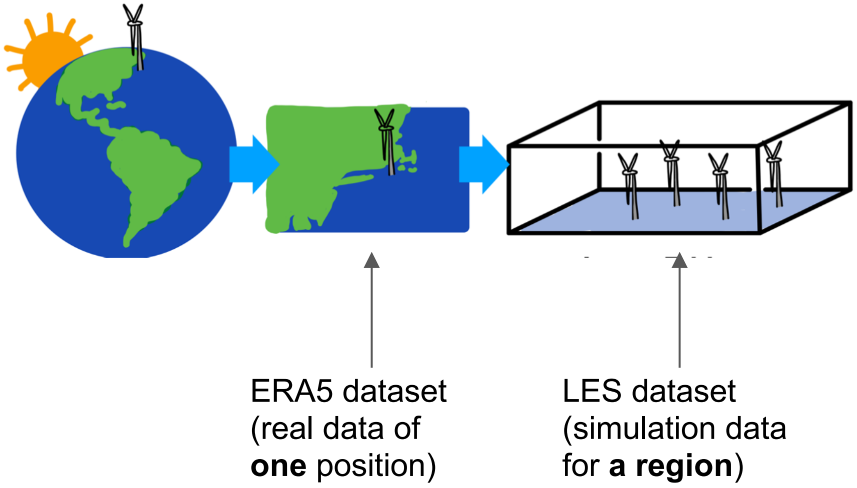


Fig 2: Demonstration of ERA5 & LES datasets

Introduction

Datasets

- **ERA5: hourly data** for the year 2020. The data has been extracted **at single point** (-7.737°E, 39.7°N) since its **spatial resolution** is about 30km.

- **LES:** 100m height above ground level, with two different spatial resolutions:
  - **high-resolution**: 80m x 80m at 1H frequency
  - **low-resolution**: 160m x 160m at 1H frequency

Introduction

Datasets

Table1: dimensions of datasets

datasets	time	xf	yf	variables
ERA5	8784	1	1	4
LES (high-resolution)	8784	192	192	6
LES (low-resolution)	8784	96	96	6

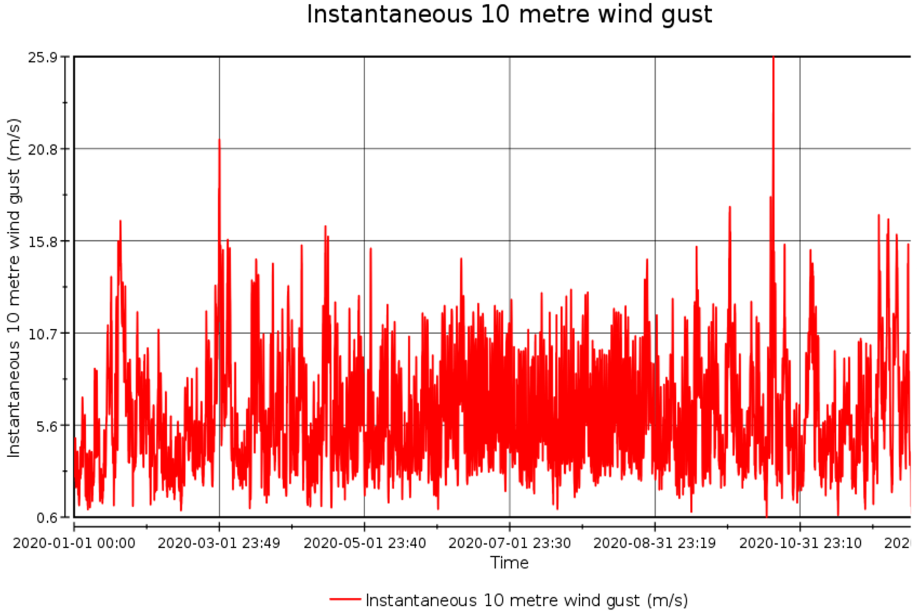
- **'time'**: Timestamps at 1H frequency
- **'xf'**: Horizontal cartesian coordinate in meter of the simulated domain (**West to East**)
- **'yf'**: Vertical cartesian coordinate in meter of the simulated domain (**South to North**).

Introduction

Table 2: Variables of ERA5 & LES

Variables (ERA5)	Variable interpretations	Variables(LES)	variable interpretations
‘t2m’	2 meter above ground level temperature in K	**‘temp’**	1H average of temperature in Kelvin
‘u100’	100 meter above ground level U wind component in m/s	**‘u’**	1H average of U component of wind speed (along ‘xf’) in m/s
‘v100’	100 meter above ground level V wind component in m/s	**‘v’**	1H average of V component of wind speed (along ‘yf’) in m/s
‘i10fg’	10 meter above ground level instantaneous wind gust	**‘vel’**	1H average of horizontal wind speed in m/s
		‘std’	1H average of standard deviation of horizontal wind speed in m/s
		‘absolute_height’	Height above sea level in meter, note that this variable only depends on (xf, yf) not on time

Introduction



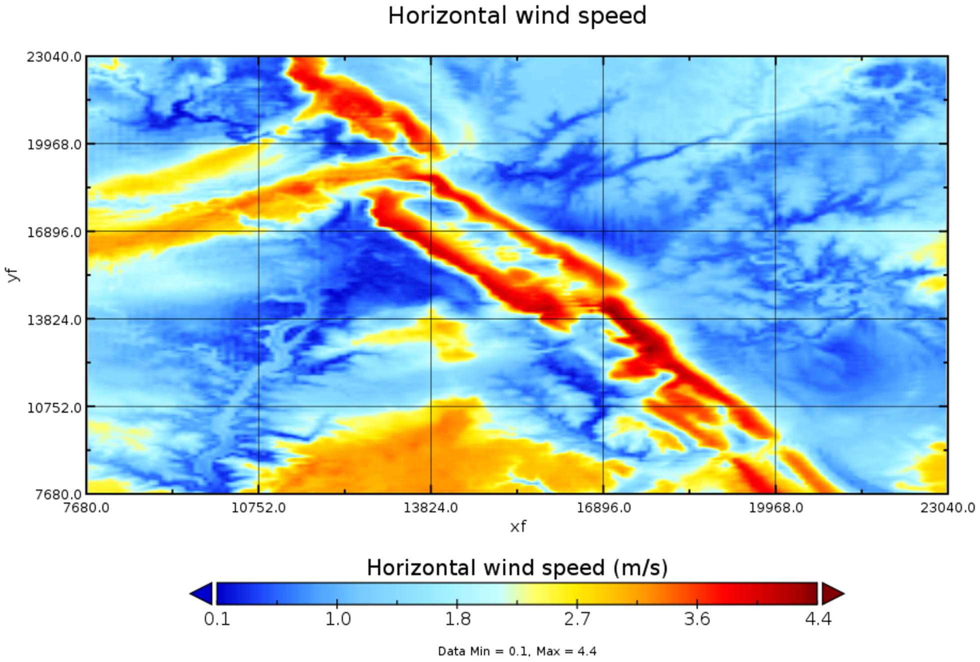


Fig 4: ‘i10fg’ of ERA5 (left), ‘vel’ of LES at a fixed time (right)

Question 1 (Compare ERA5 & LES simulation)

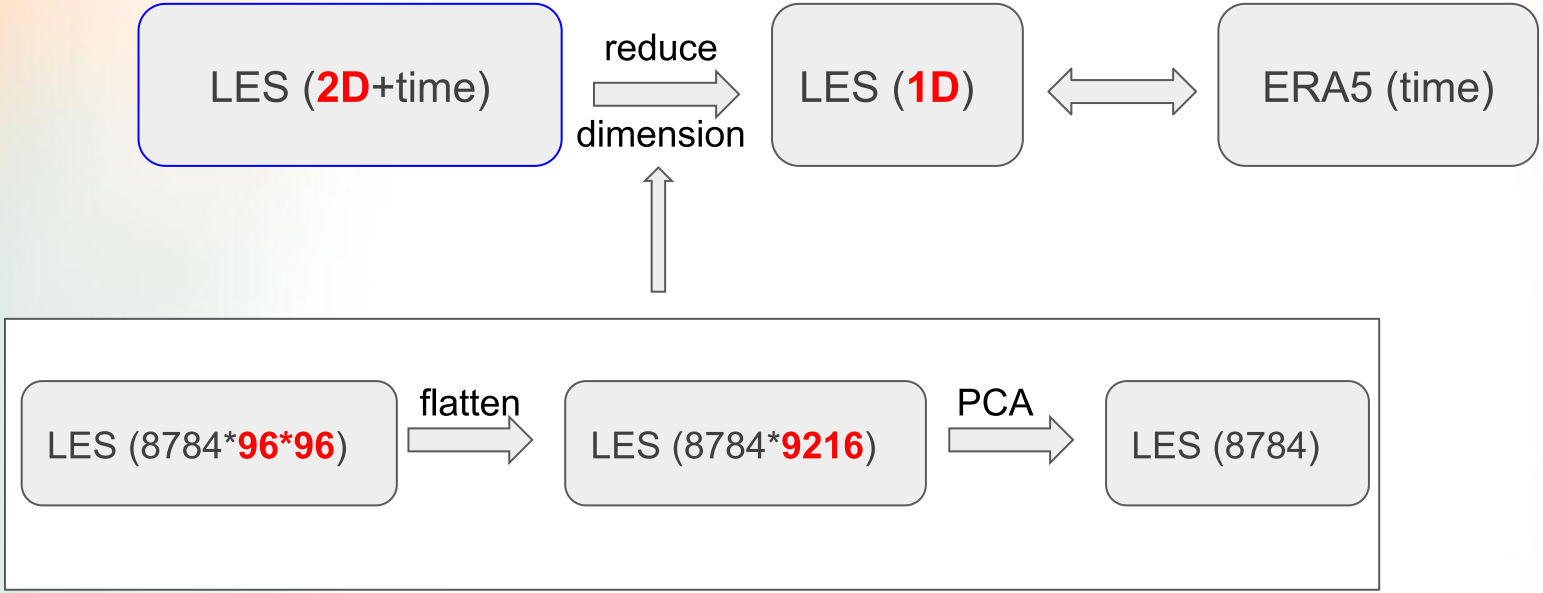
- Is there any **systematic bias**?
- What's the **correlation**?

Methods:

$$\text{bias} = \frac{\sum (x - y)}{number\ of\ x}$$

Pearson correlation coefficient:

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}}$$


LES (2D+time)
reduce
dimension
LES (1D)
ERA5 (time)
LES (8784*96*96)
flatten
LES (8784*9216)
PCA
LES (8784)

Results:

Table 3: Bias and correlation coefficient of ERA5 & LES

ERA5 v.s. High_res	**temp (t2m & temp)**	**Wind speed (i10fg & vel)**	**U component of wind speed (u100 & u)**	**V component of wind speed (v100 & v)**
bias	-2.13×10^{-5}	-1.66×10^{-7}	3.48×10^{-8}	5.29×10^{-8}
correlation	0.8833	0.8065	-0.9677	0.9561

Conclusion:

- LES is **unbiased** since the bias of all variables are small
- LES has **strong linear correlation** with ERA5, since the magnititutes are all close to one

Question 2 (compress to a lower demension)

- Compare **standard** method and **deep learning** method.
- What's the interpretability & visual insights of the **latent space**?

Methods:

Principal Component Analysis (PCA)

v.s.

t-distributed stochastic neighbor embedding (t-SNE)

PCA(Principal Component Analysis)

Linear dimensionality reduction using **Singular Value Decomposition (SVD)** of the data to **project** it to a lower dimensional space.

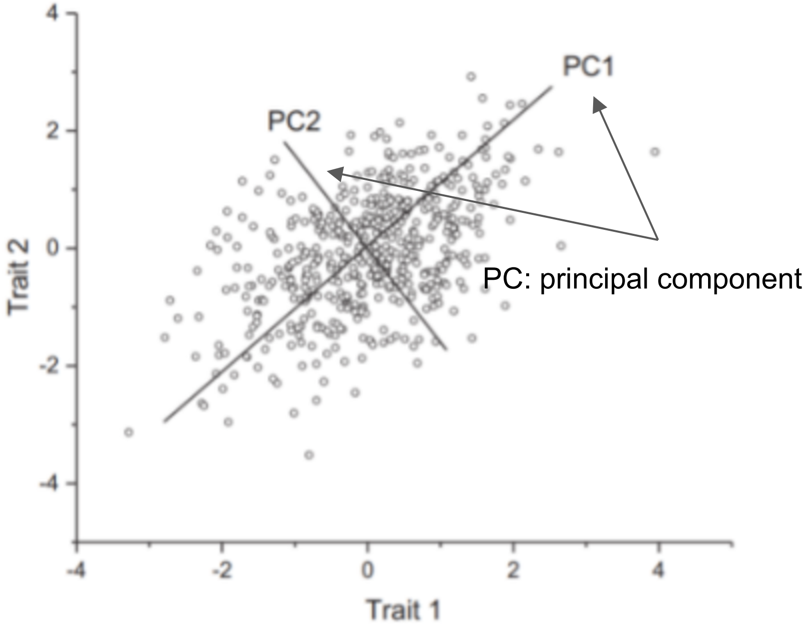


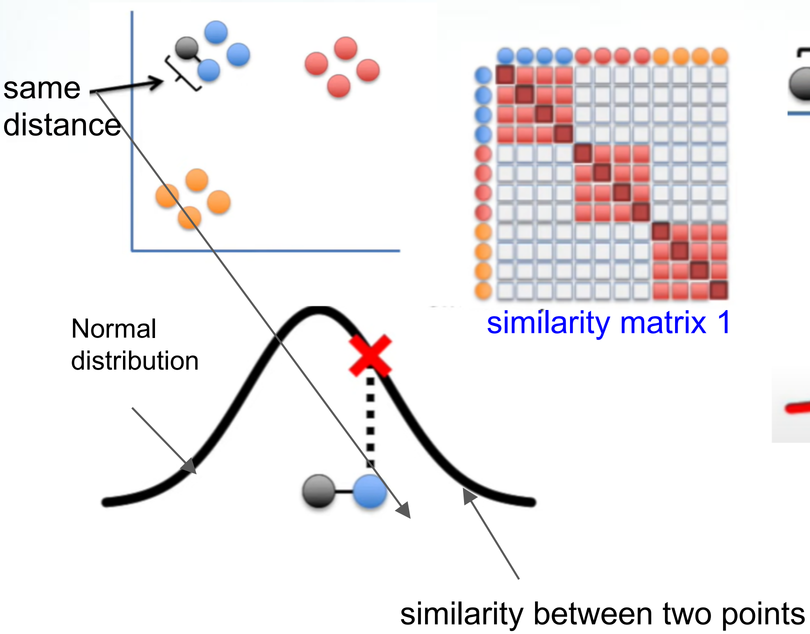
Fig 5: Schematic diagram of PCA

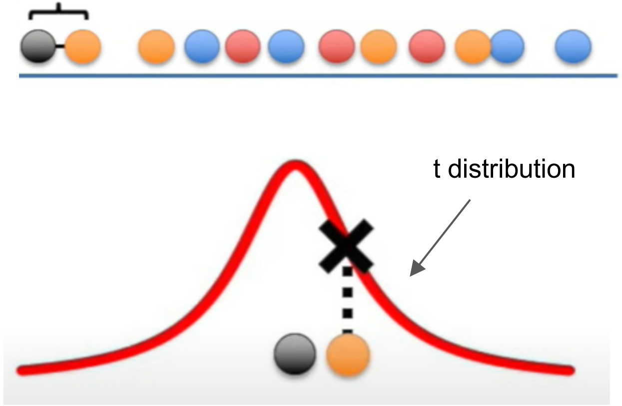
standardize
↓
compute covariance matrix
↓
compute eigenvectors & eigenvalues
↓
decide principle components used
↓
recast data

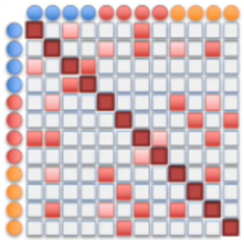
Resource:https://builtin.com/data-science/step-step-explanation-principal-component-analysis

t-SNE(t-distributed stochastic neighbor embedding)

It is a **non-linear** dimensionality reduction technique that is particularly suited for the visulaization of high-dimensional datasets.







similarity matrix 2

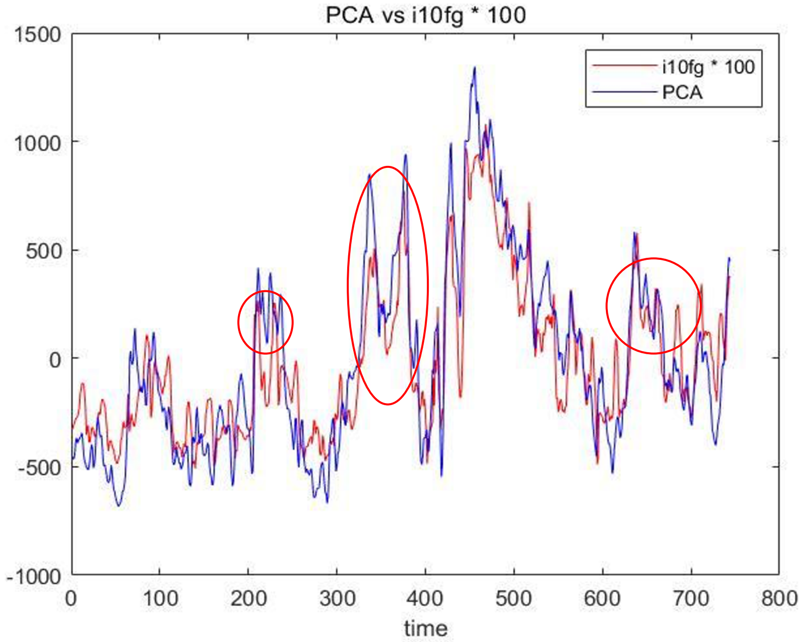
scaled score = score / sum of all scores

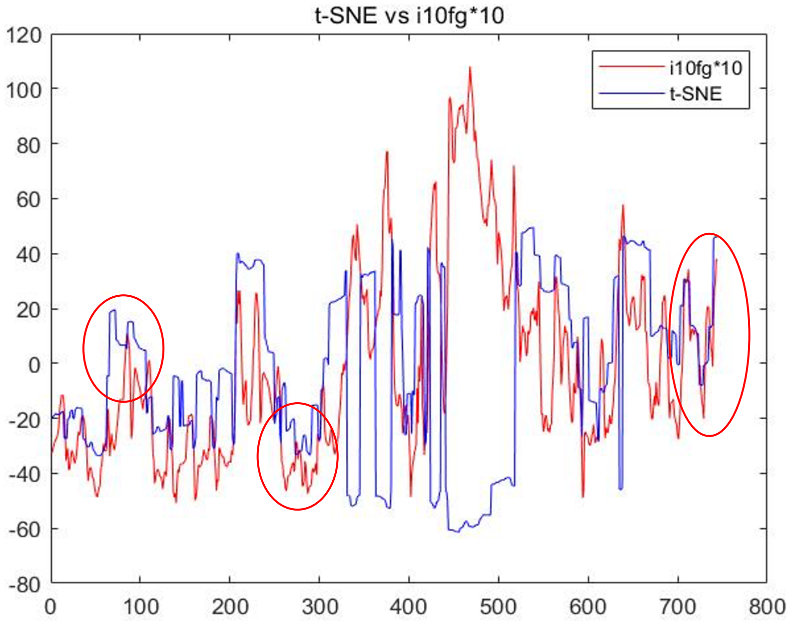
- calculate “unscaled similarity” score of the points in original dimension
- scale them to get “similarity matrix 1”
- project points onto a line randomly
- calculate “unscaled similarity” score of the points in the line
- scale them to get “similarity matrix 2”
- move points to make the matrix 2 like matrix 1

Resource: https://www.youtube.com/watch?v=NEaUSP4YerM

Compare PCA and t-SNE (1D)

Reduce one-month high-resolution data with “**vel**” variable to 1D (744*192*192 → 744) by **PCA** and **t-SNE** independently (blue line), and compare with enlarged ERA5 (“**i10fg**”, red lines). Multiple “i10fg” by 100 times and 10 times respectively, to better compare the shape of compressed data and true data.

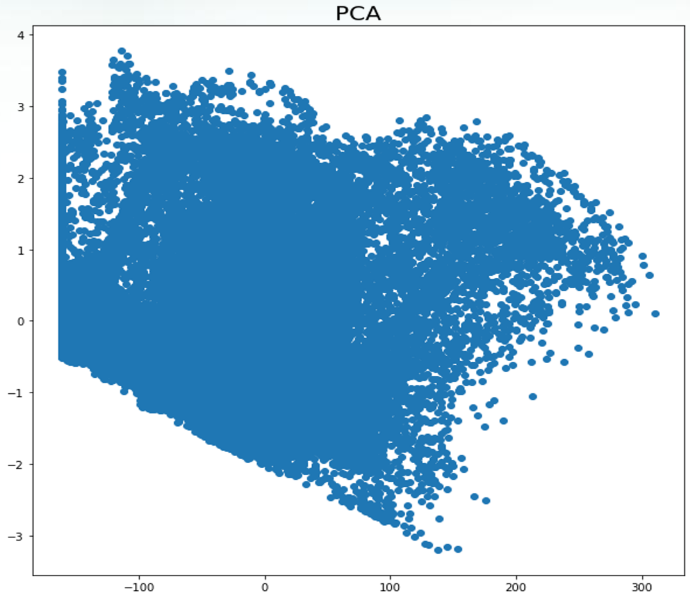


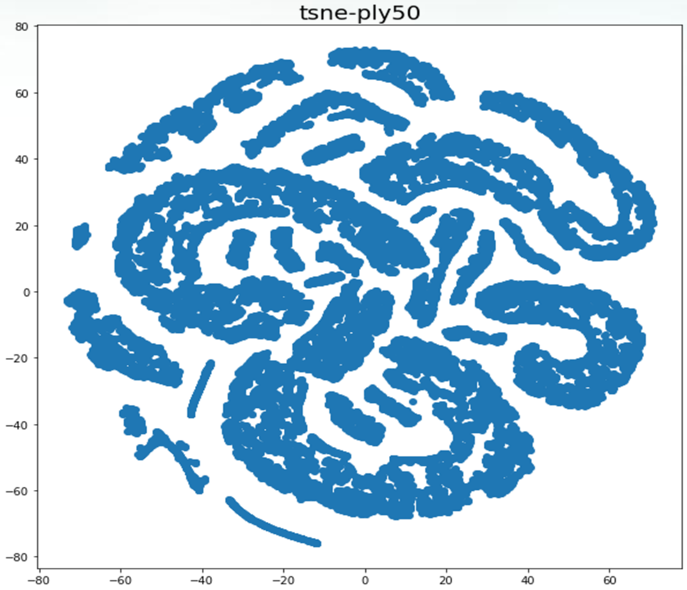


Conclusion: With similar shapes around peaks, PCA **preserves most information** of the original data, but t-SNE keeps **some information**.

Compare PCA and t-SNE (2D)

Reduce one-month high-resolution data to 2D (36864*$\mathbf{6}$ → 36864*$\mathbf{2}$)





The image after PCA compressing forms **one cluster**.

The t-SNE compressed image is divided into **different clusters**.

Conclusion: t-SNE **extract the features** of the original data, but PCA does not.

Add **labels** to LES

Choose wind power density as the indicator because it combines variables “temp” and “vel” together, and it is important to the wind energy industry. Wind turbines convert the kinetic energy in the wind into mechanical power.

Wind power density(w/m^2) wp/s =½ * ρ *v^3
wp - wind power, s - unit area, **ρ - air density, v - wind speed**

calculate air density:

altitude of our data are all under 11000m

For h < 11000 (Troposphere)

$$T = 15.04 - .00649\,h$$

$$p = 101.29 * \left[\frac{T + 273.1}{288.08}\right]^{5.256}$$

ρ = density (kg/cu m)
p = pressure (K-Pa)

$$\rho = p\,/\,(.2869 * (T + 273.1))$$

T = temperature (^{0}C)
h = altitude (m)

Table 4: Wind energy scale at elevation of 100m

Wind Power Class	Wind Power Density (W/m²)
C1 (Poor)	0-114.87
C2 (Marginal)	114.87-172.31
C3 (Fair)	172.31-229.75
C4 (Good)	229.75-287.19
C5 (Excellent)	287.19-344.62
C6 (Outstanding)	344.62-459.50
C7 (Superb)	459.50-1148.75
C8 (Out of Superb)	over 1148.75

Resources: https://www.grc.nasa.gov/www/k-12/airplane/atmosmet.html; http://educypedia.karadimov.info/library/Lesson1_windenergycalc.pdf; https://www.researchgate.net/publication/343382220_An_Evaluation_of_the_Wind_Energy_Resources_along_the_Spanish_Continental_Nearshor

latent space

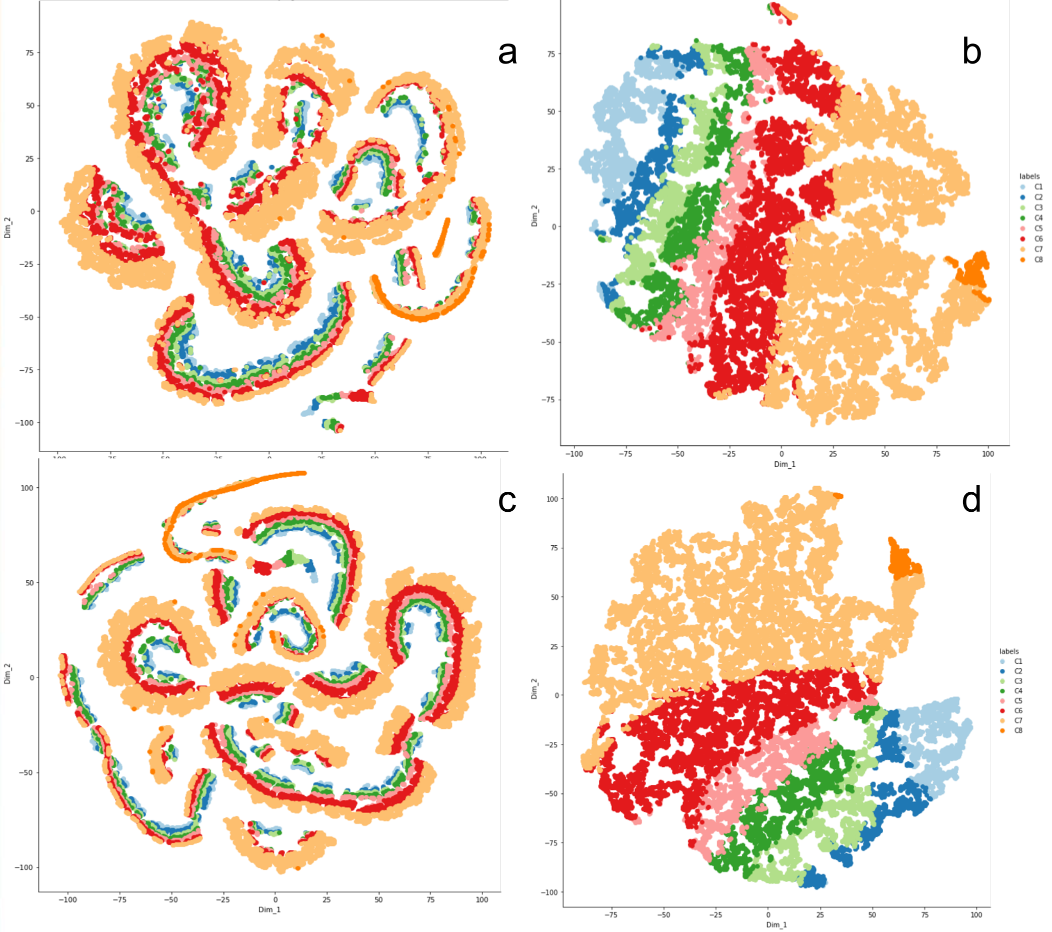


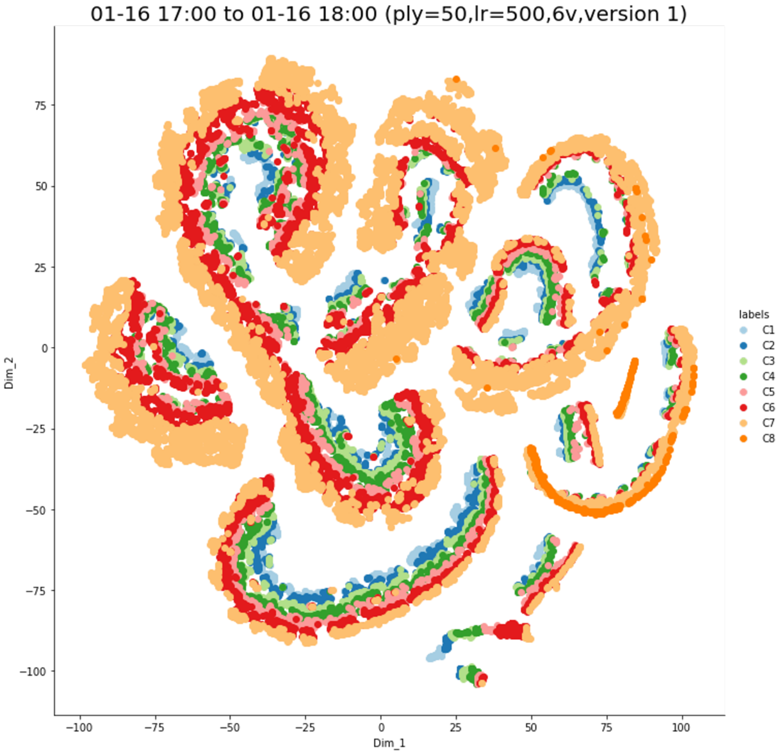
Table 5: variables included in each figure

	a	b	c	d
“vel”	√	√	√	√
“temp”	√	√	√	√
“std”	√	√	√	√
“u”	√	√		
“v”	√	√		
“absolute _height”	√		√	

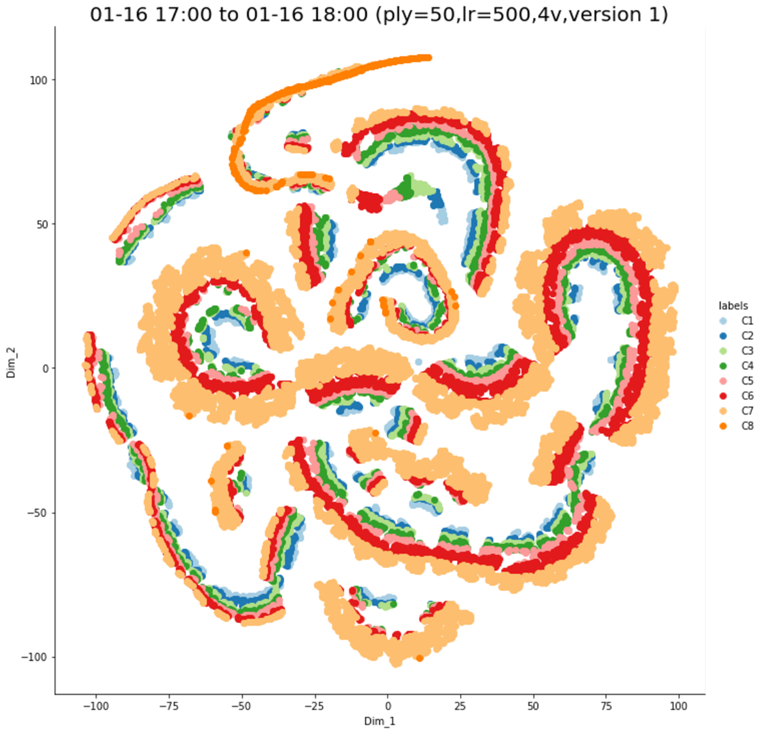
Conclusion:

- The multiple clusters in (a) and (c) are due to “absolute_height”.
- Each cluster in all figures separates different classes clearly.

latent space



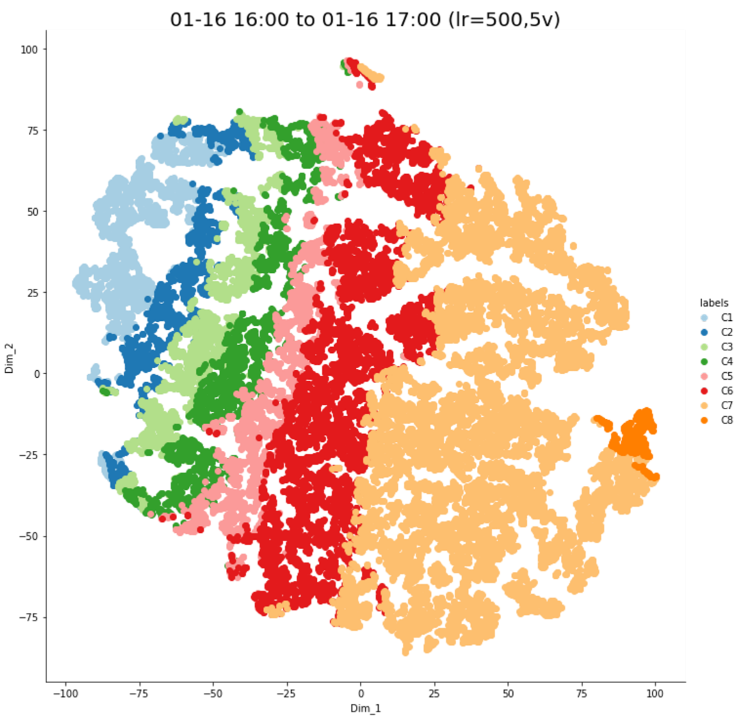
6 variables



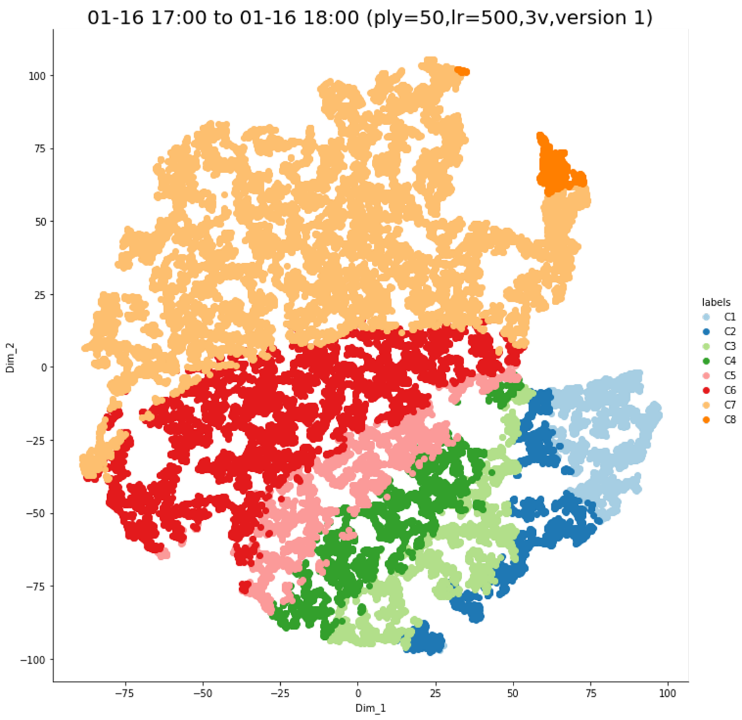
4 variables (drop “u”, “v”)

Observation: The **boundaries become smother** without “u”, “v” variables

latent space



5 variables (drop “absolute_height”)



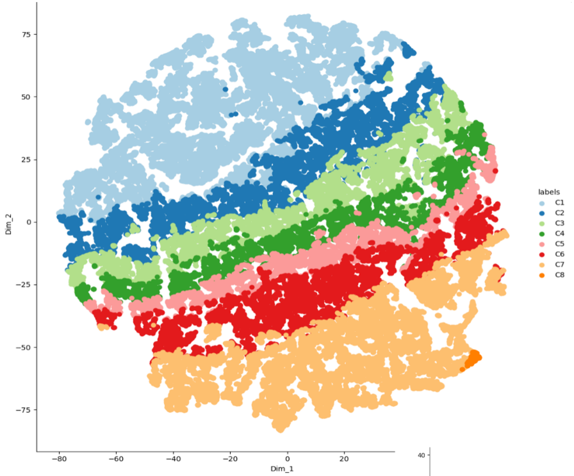
3 variables (drop “absolute_height”, “u”, “v”)

Observations:

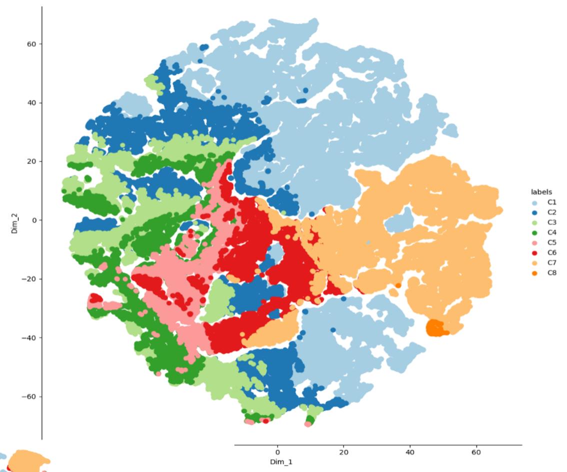
- The boundaries become smoother without “u”, “v” variables.
- Different classes are separated clearly without “absolute_height” variable.
- The proportion of data in each class can be observed intuitively.

latent space

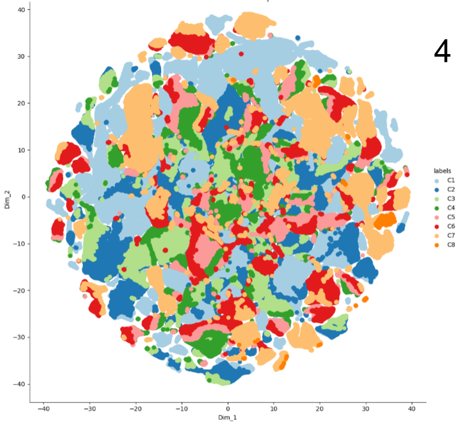
- 5 variables (drop “absolute_height”)



2 hours data



4 hours data

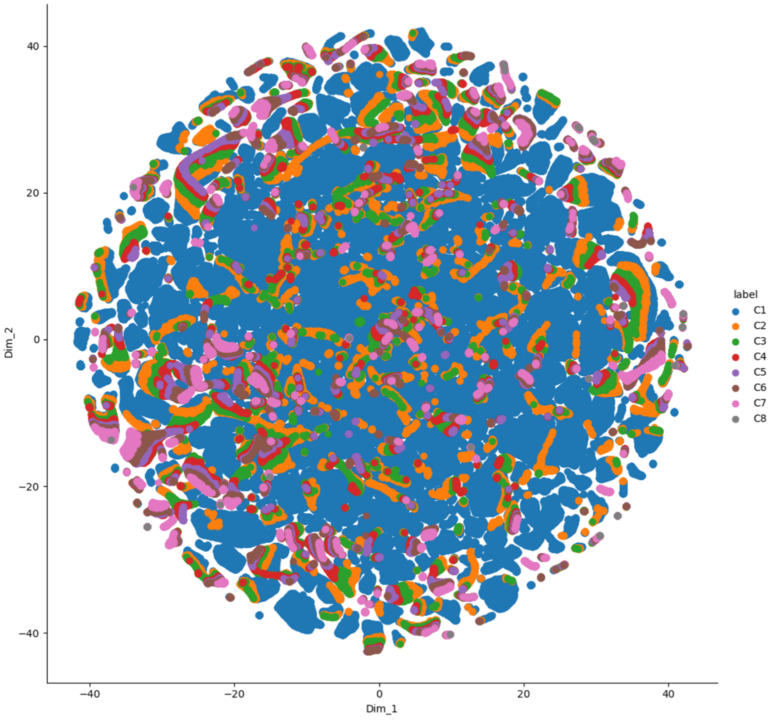


24 hours data

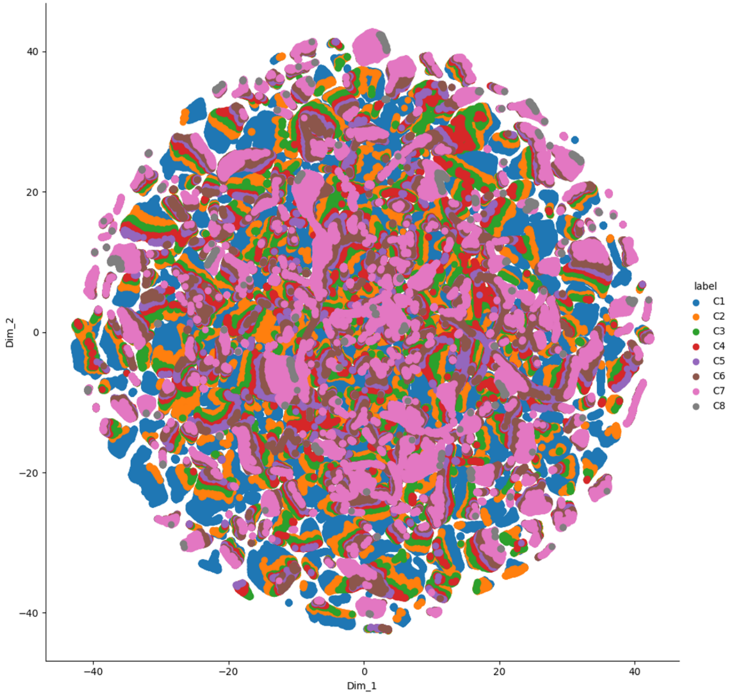
Conclusion: Different classes are no longer separated by ribbons with over 4 hours data as input, but begin to form spherical.

latent space

- 24 hours data
- 4 variables
"vel", "temp", "std", "absolute_height"



01-14



01-22

Conclusion: The strength of wind power density on different days can be told by latent space of 24 hours data.

Question 3 (Upscaling from a low-resolution to high-resolution grid)

- Can we make a **prediction** of a high-resolution dataset based on a low-resolution input dataset?
- What's the **accuracy** of upscaling?

Methods: Interpolation (nearest-neighbor, bilinear, and bicubic interpolations) & Unet

Interpolation is a statistical method using related **known** values to **estimate unknown** values.

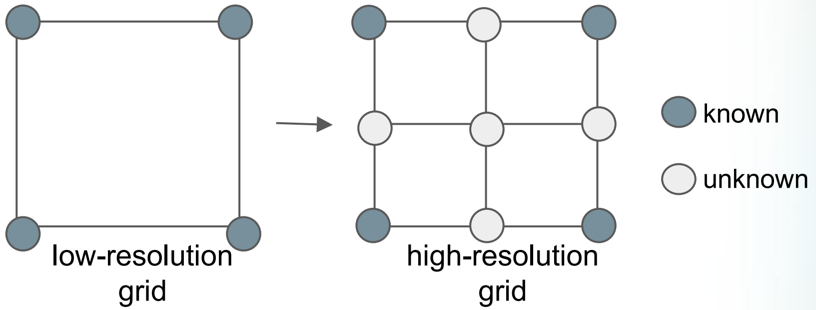


Fig 6: Simple demonstration of upscaling

Table 6 Comparison of interpolation methods

Interpolation method	**Nearest neighbor interpolation**	**Bilinear interpolation**	**Bicubic interpolation**
Pixel value	Use value of **nearest pixels**	Use weighted **average of two pixels**	Use weighted **average of four pixels**
Subjective Feelings	Mosaic phenomenon	Blurring, not sharp	Sharper and fuzzy
Image visibility	Not clear	Jaggy, not clear	Better than bilinear
Performance	Worst	Poor	**Better**
Computation time	Less	Less than bicubic	more
Speed	Simple and **fast**	Slightly slower	fast

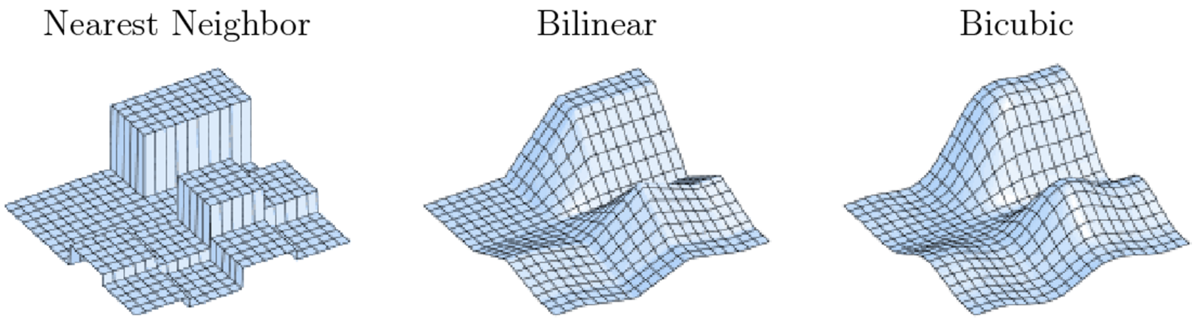


Fig 7: Patterns formed by different interpolation methods

source: https://www.semanticscholar.org/paper/Survey-on-Image-Interpolation-Kaur-Kaur/bad7a7dde3c13d6bfd7bbddfc3455022854b4934

Table 7: Performance of different interpolation methods

	Nearest-neighbor interpolation	Bilinear interpolation	Bicubic interpolation
PSNR	23.48	24.12	24.13
SSIM	0.64	0.67	0.68

- **Peak signal-to-noise ratio (RSNR)**

$$PSNR = 10 \cdot \log_{10}\left(\frac{MAX_I^2}{MSE}\right)$$

- **Structural similarity (SSIM)**

$$\mathrm{SSIM}(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}$$

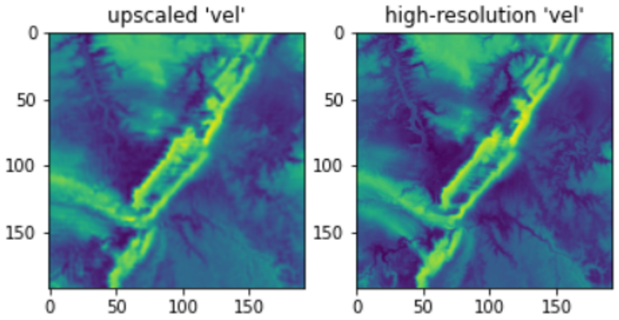
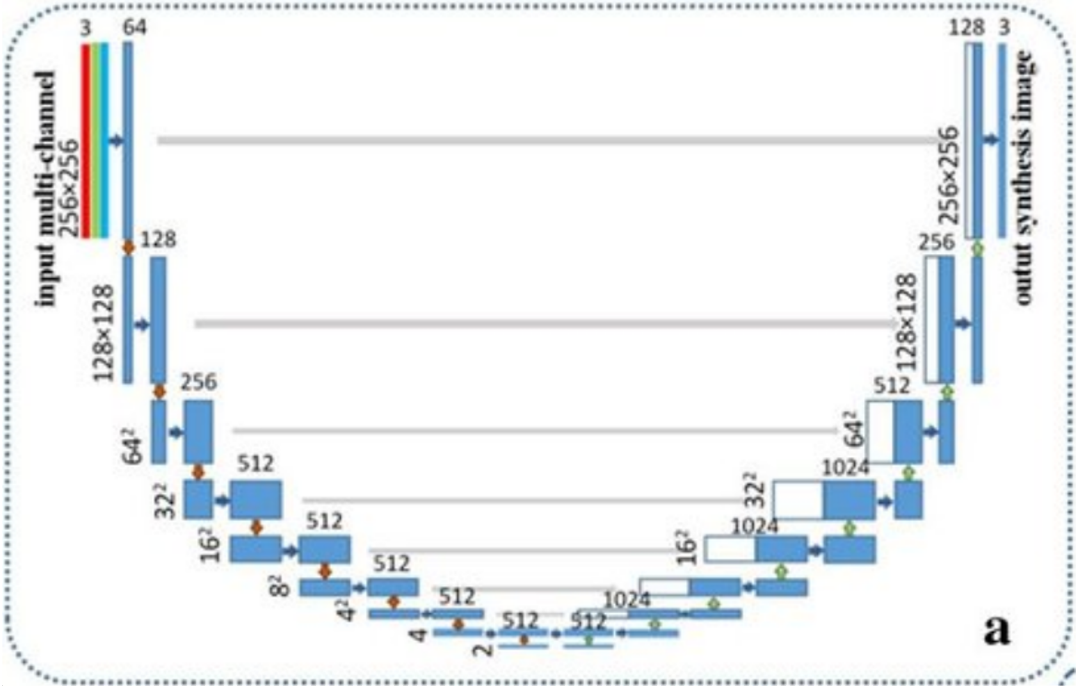


Fig 8: upscaling of 'vel' variable using bicubic interpolation

Unet

Method

- Normally use for segmantation
- Extract features of the image

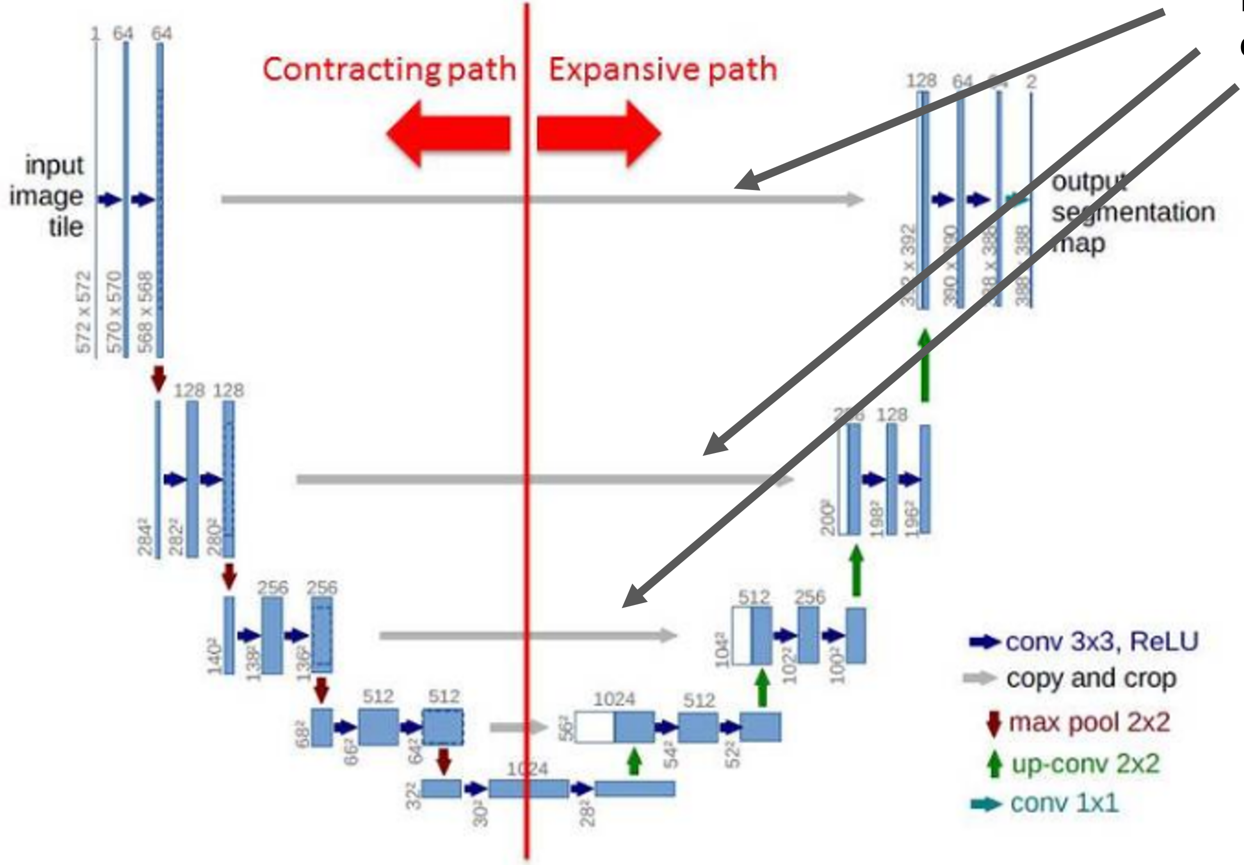


Unet

Method

Network Architecture

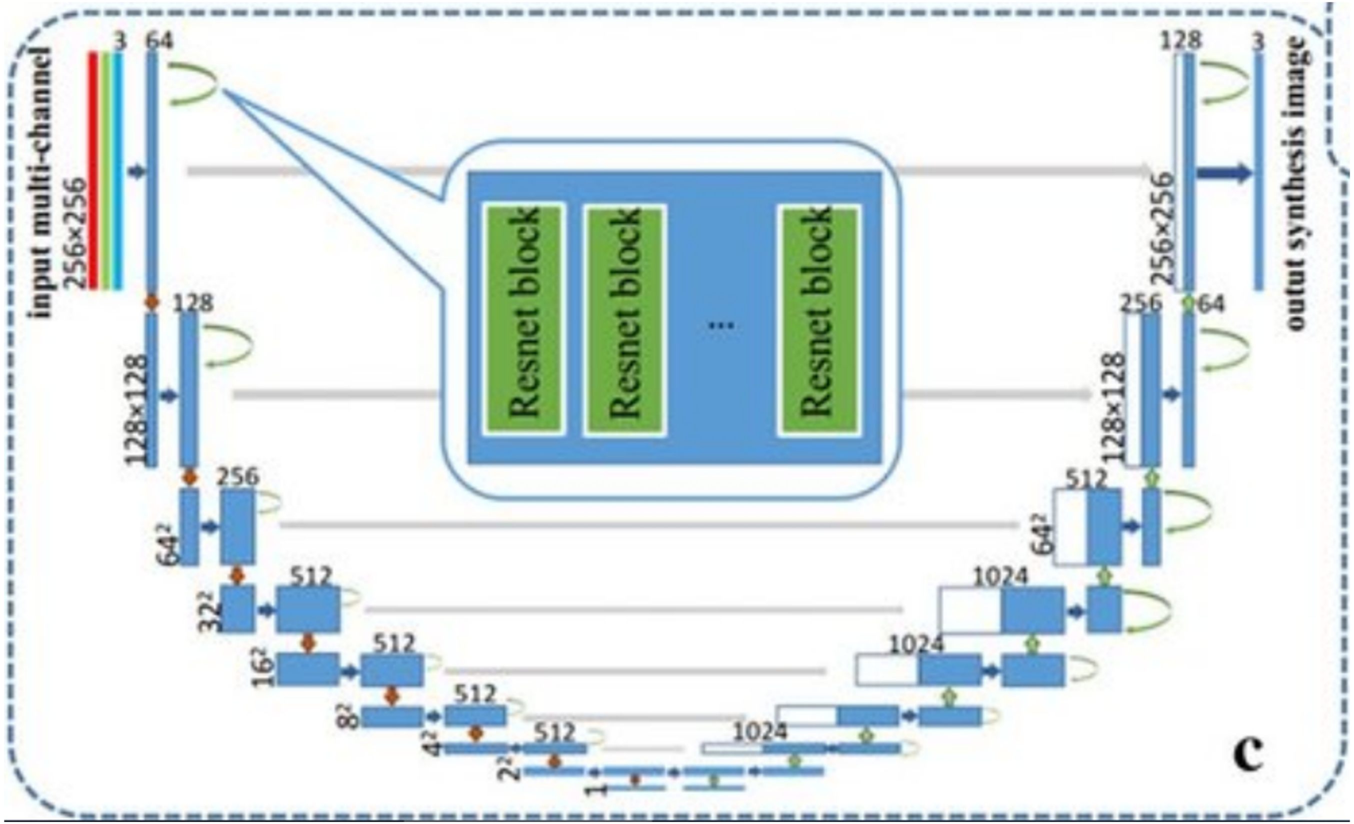
Concatenation for feature extraction



Unet

Method

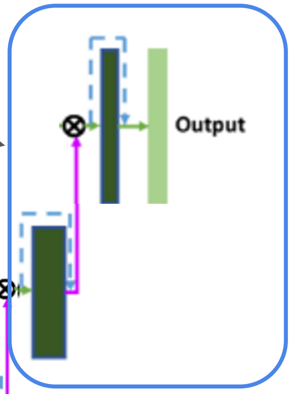
ResNet Unet

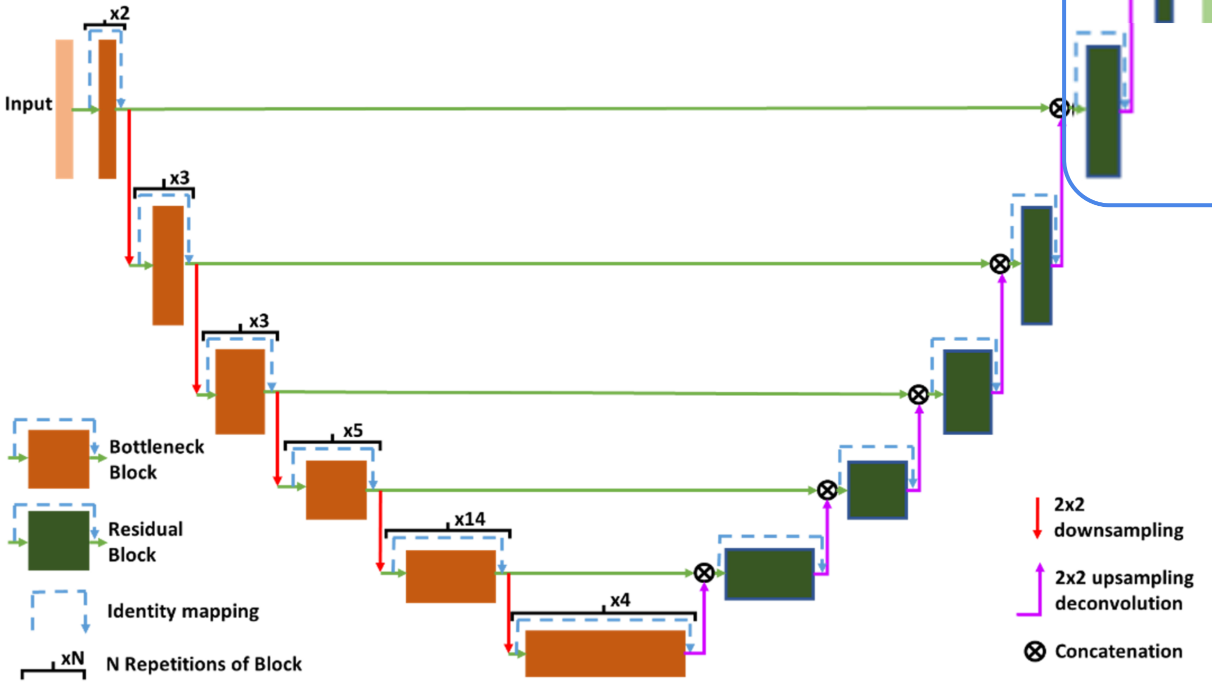


Unet

Method

- Add one more upsampling layer





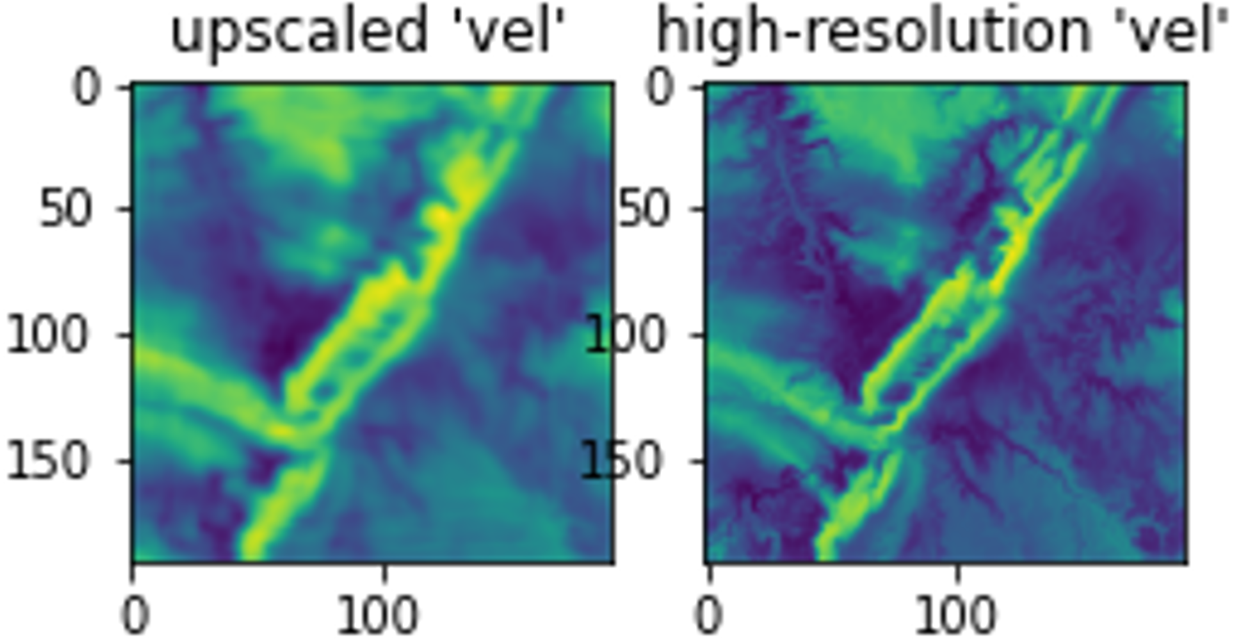
Unet

Method

Result compare to bicubic interpolation

	Unet	Bicubic interpolation
PSNR	16.88	24.13
SSIM	0.49	0.68

Unet Method


upscaled 'vel'
high-resolution 'vel'
0
50
100
150
0
100
0
50
100
150
0
100

Unet

Method

Possible reasons

- **Not enough training (140 epoch for 8hrs)**
- **ResNet Unet is not good for image upsampling**
- **Parameter settings:**
  - **Learning rate**
  - **Optimization algorithm**

Super-resolution

Method

USRNet

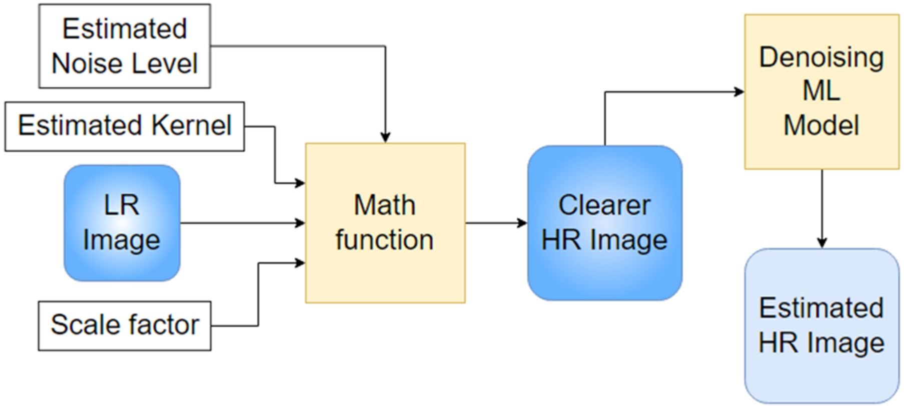
K. Zhang, L. V. Gool, and R. Timofte, “Deep unfolding network for image super-resolution,” in CVPR, 2020, pp. 3214–3223.

Input:

- LR image
- Estimated Kernel:

  kernel width = 0.01
- Estimated Noise Level:

  sigma = 4.5
- Scale factor = 2



Super-resolution

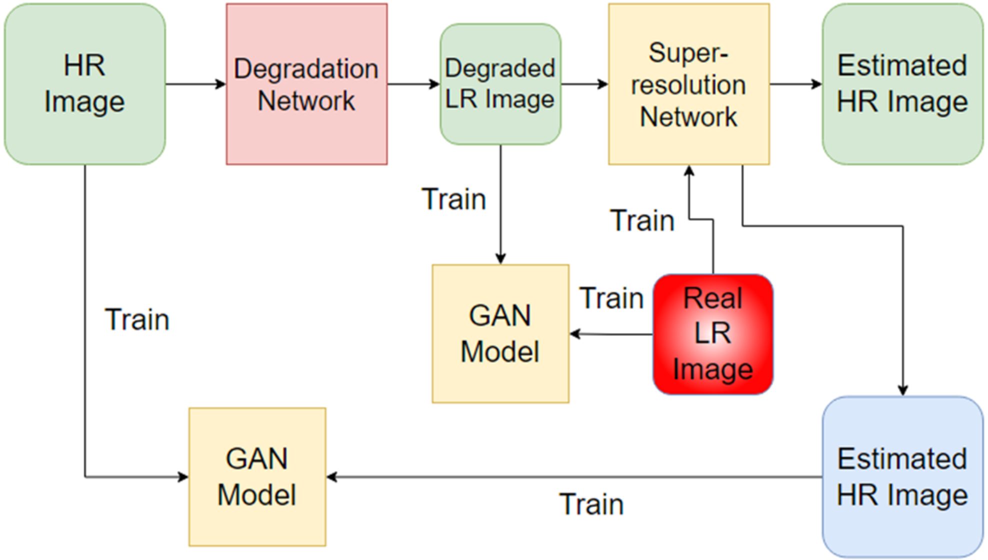
Method

DASR

S. Y. Kim, H. Sim, and M. Kim, "Koalanet: Blind super-resolution using kernel-oriented adaptive local adjustment," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2021, pp. 10611–10620.

Input:

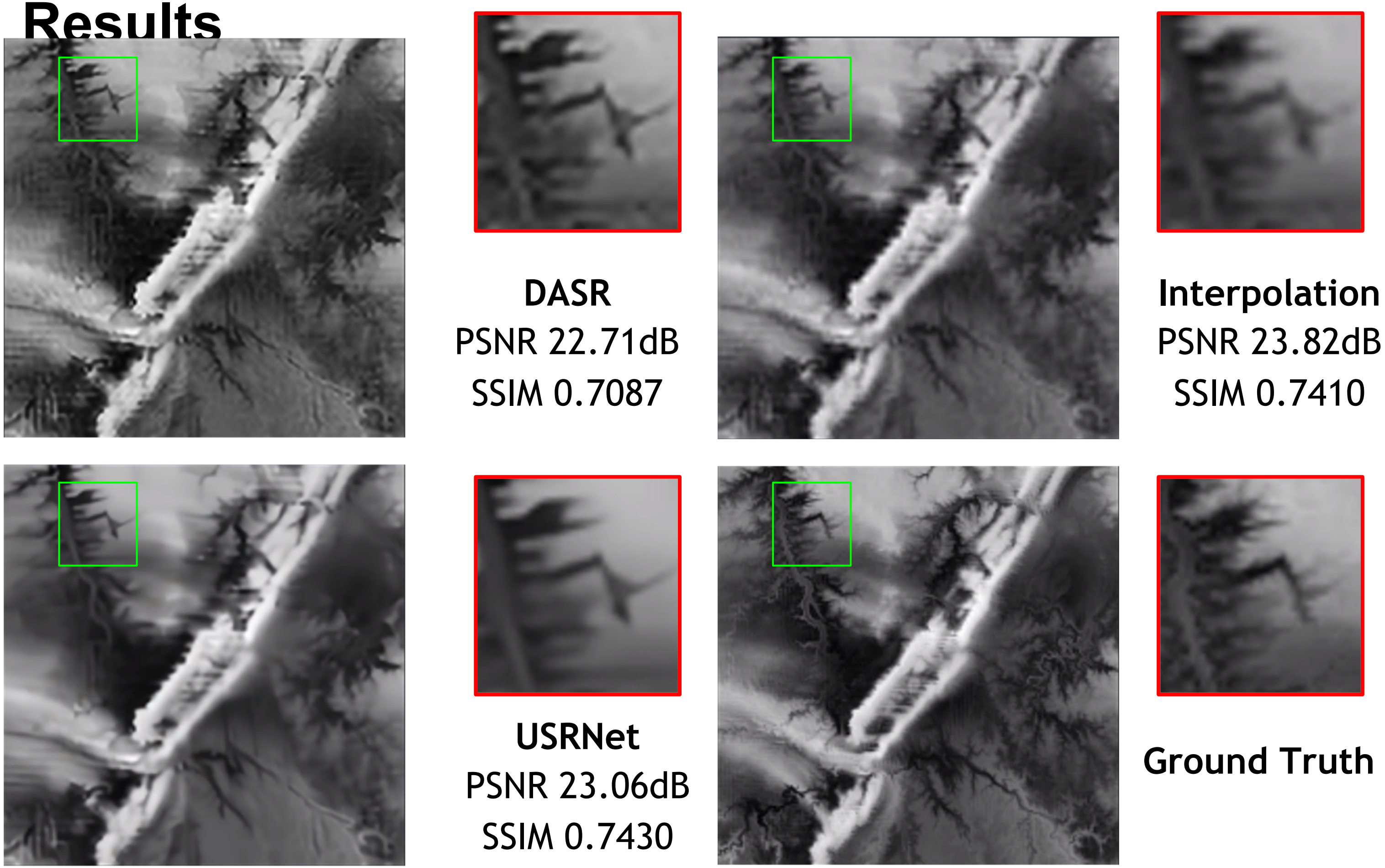
- LR image only



Super-resolution

Method

Results



Super-
resolution

Method

Possible reasons

- Model normally use for general photos
- Looking clear != good performance
- Not enough training

Reference

- https://builtin.com/data-science/step-step-explanation-principal-component-analysis
- https://www.youtube.com/watch?v=NEaUSP4YerM
- https://towardsdatascience.com/t-sne-clearly-explained-d84c537f53a
- https://www.nature.com/articles/s41467-019-13056-x
- http://educypedia.karadimov.info/library/Lesson1_windenergycalc.pdf
- https://www.grc.nasa.gov/www/k-12/airplane/atmosmet.html
- https://www.researchgate.net/publication/343382220_An_Evaluation_of_the_Wind_Energy_Resources_along_the_Spanish_Continental_Nearshore
- https://www.semanticscholar.org/paper/Survey-on-Image-Interpolation-Kaur-Kaur/bad7a7dde3c13d6bfd7bbddfc3455022854b4934

Thank you

Q&A