Data Analytics Application Series I: Wind Energy Part I

Zijun Zhang

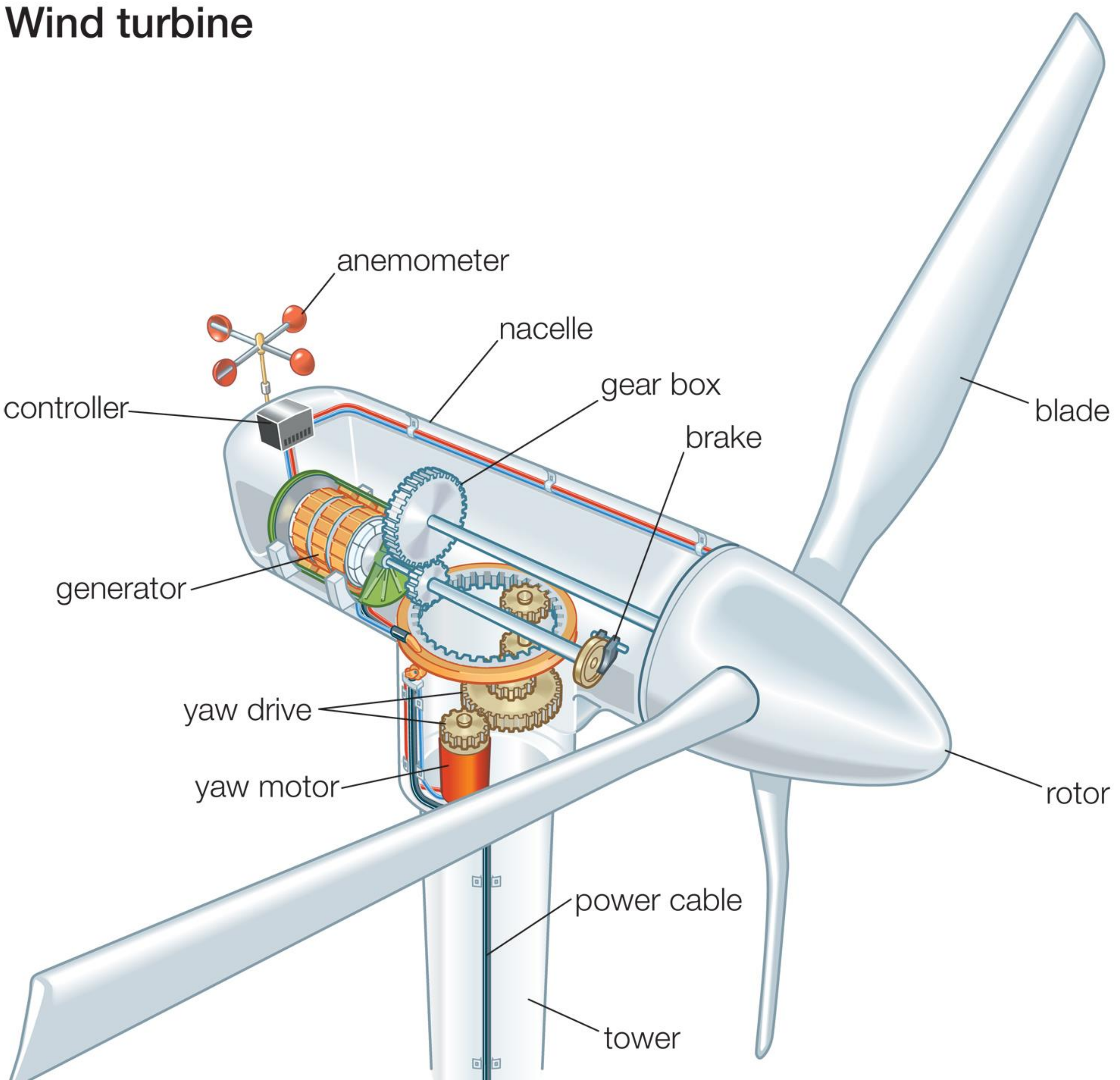
Why wind energy?

- Renewable energy plays an important role in the city sustainability
- Wind energy is one of most successfully commercialized renewable energy in the world
- Wind farm is composed of a population of distributed units, difficult to manage
- Wind turbine is working under harsh environment
- Wind is uncertain
- Wind energy asset is expensive
- Wind turbine is designed and manufactured to be fully instrumented

A Commercial Wind Farm

A Wind Turbine System

- Composed of a number of mechanical and electrical components
- Energy conversion process is highly nonlinear
- A high dimension of system parameters
- Uncertain wind speed as free fuel

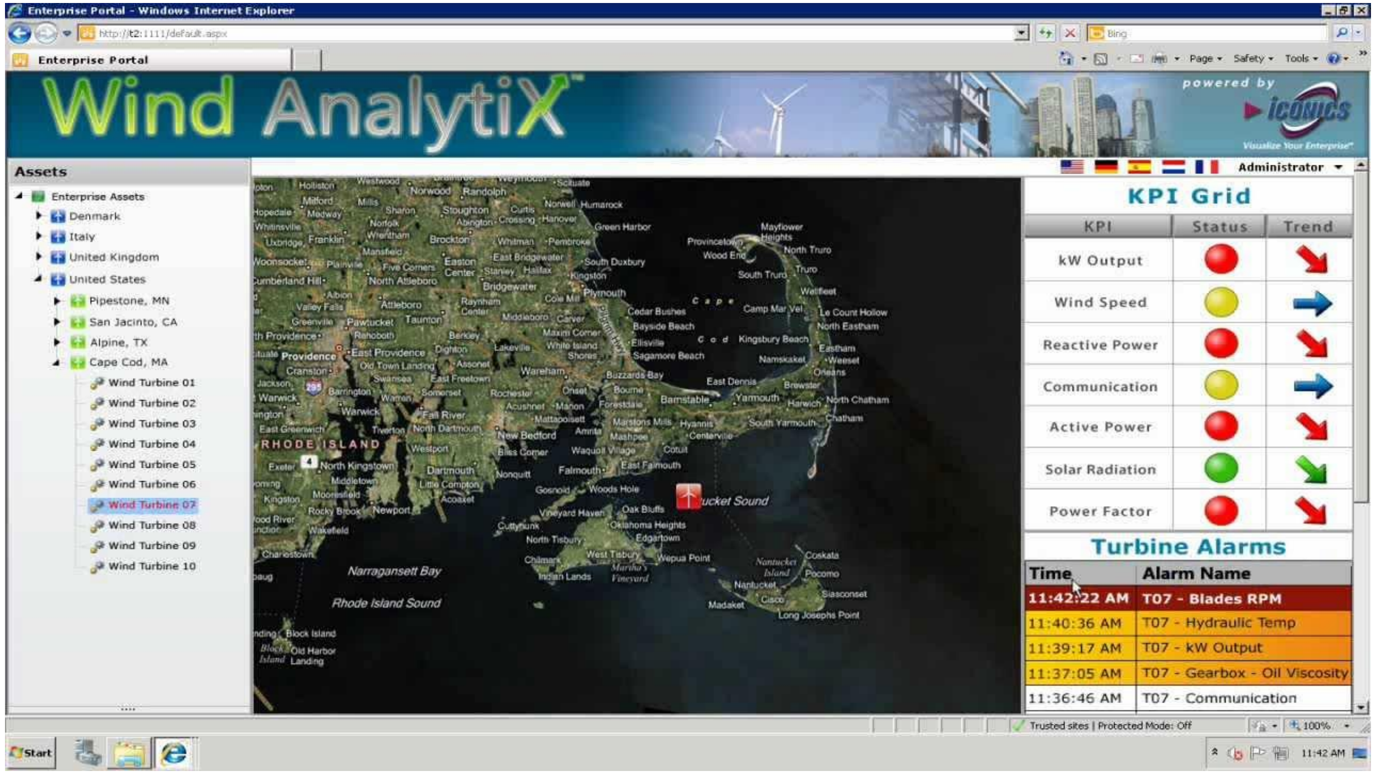


Supervisory Control and Data Acquisition

- SCADA system, designed for monitoring WT conditions but it collects a massive volume of WT data
- Over hundred system and environmental parameters involved in wind power generation processes are measured
- Sampling frequency can be quite high but data typically stored in 10-min intervals due to the storage constraint
- SCADA data enables the era of wind energy big data analytics

SCADA System

Can be considered as an IoT technology



Wind Energy Problems Receiving Data Driven Solutions

Problems Only Involve Machine Learning:

- Uncertain factor forecasting, such as wind speed, wind power.
- Data-driven wind turbine modeling

Problem Involve Machine Learning + Another Field:

- Data-driven wind turbine intelligent control
- Data-driven wind farm operations
- Data-driven wind turbine condition monitoring

Wind Energy ML Problems

- Wind speed/power prediction v.s. wind system modeling
- Problem formulation:
  - A data-driven modeling process considering physics based and mechanical principles
  - Based on domain knowledge, you need to know system inputs and system outputs
  - Input parameters – controllable and non-controllable
  - Output parameters – numerical or categorical based on the definition

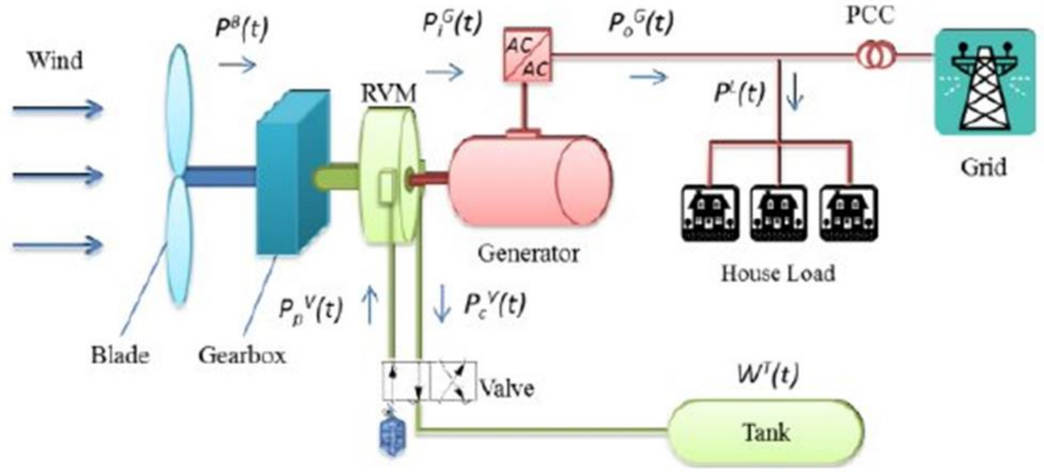
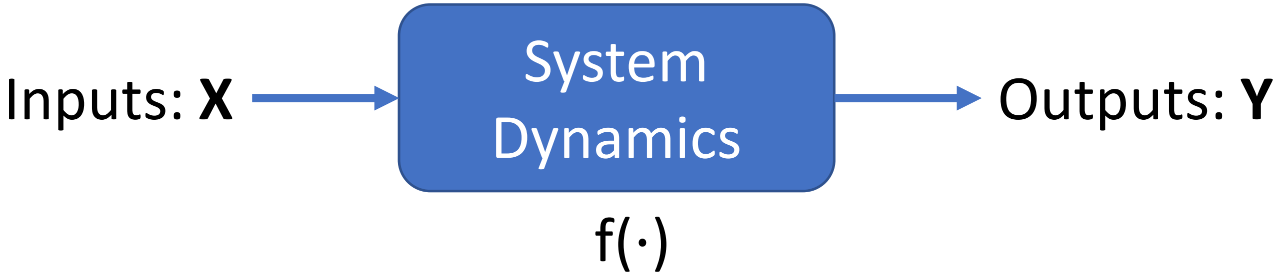


Figure 2. Proposed wind conversion system.

Wind Energy ML Problem

A System



Problem: **X**, **Y** are known as data, while f(·) needs to be obtained

Given a loss function L, a data-driven system modeling task typically aims at identifying a non-parametric model $g_A(\cdot)$ via algorithm A by minimizing the $L(g_A(\mathbf{X}), \mathbf{Y})$.

Wind Energy ML Problem

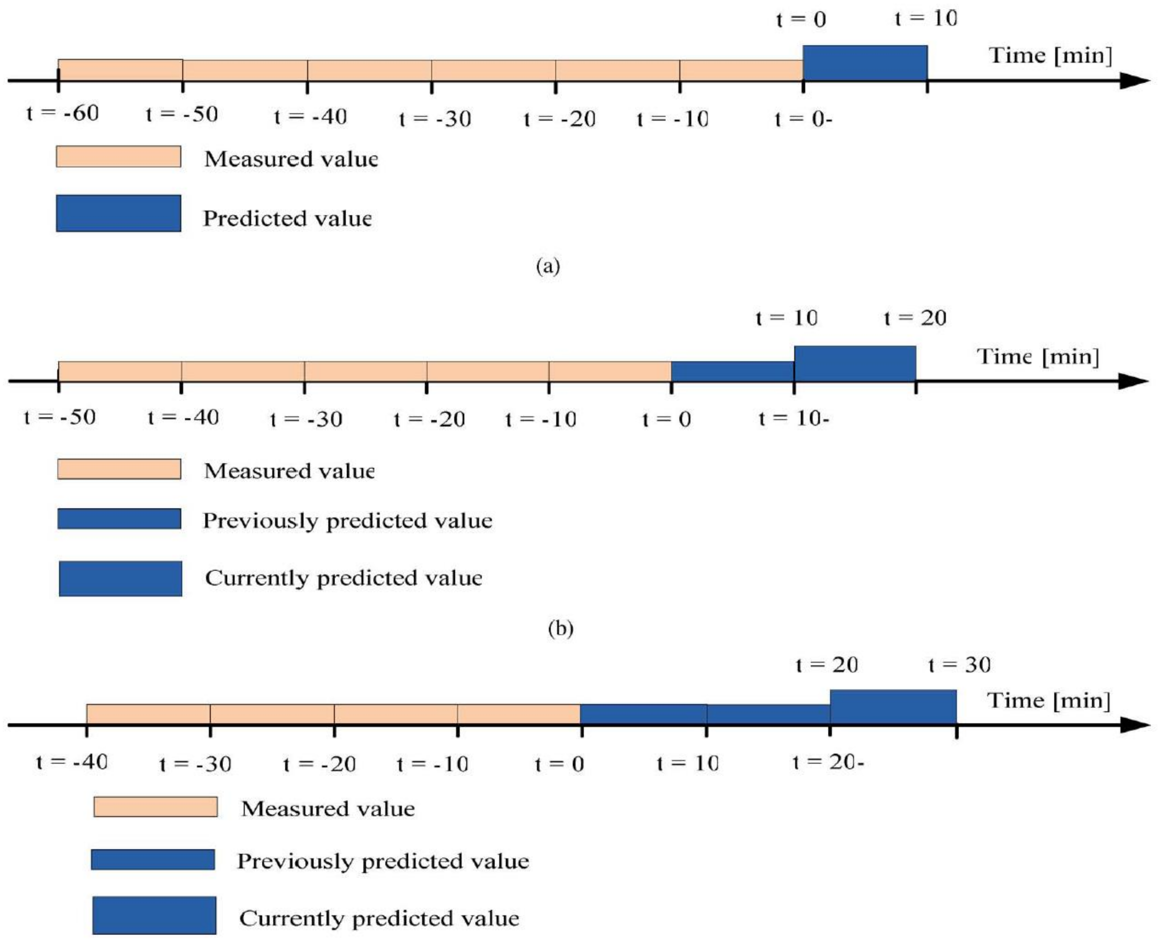
- Two types of problems can be generalized under the same computational framework
- Differences typically include the considered input parameters and the output parameters
- In modeling, we aim to use X_t to estimate Y_t
- In prediction, we aim to use X_t to estimate Y_{t+n}, $n \geq 1$
- The learning process is the same, which aims to find a $f(\cdot)$ correctly describing the relationship between X and Y based on data

Wind Speed/Power Prediction

Source: A. Kusiak, H. Zheng, and Z. Song, “Short-Term Prediction of Wind Farm Power: A Data Mining Approach,” IEEE Trans on Energy Conversion, Vol. 24, No. 1, pp. 125-136, 2009

Time-series model formulation

$$\hat{y}(t+T) = f(y(t), y(t-T), \ldots, y(t-mT))$$



Wind Speed/Power Prediction

Problems:

How to select appropriate input parameters (features) for doing data-driven modeling?

How to select the appropriate algorithm for developing/learning the data-driven models?

Wind Speed/Power Prediction

Feature selection

- Based on domain knowledge/principles
- Trial-an-error methods
- Linear models with feature selection capability, LASSO
- Nonlinear case
  - boosting tree – importance rank
  - NN – global sensitivity analysis
- No theoretical guarantee of the global optimal choice due to two reasons: 1) features measured might be insufficient; 2) feature selection and modeling algorithms can be different

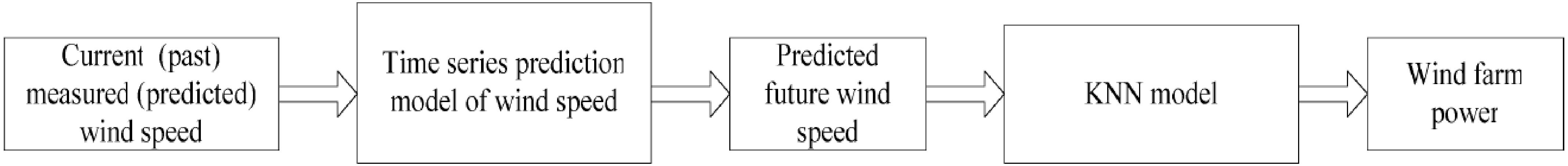
Wind Speed/Power Prediction

Algorithm selection

- A pool of candidate algorithms need to be determined
- Computational experiments for comparative analytics
- Assessment metrics, MAE, MSE, MAPE, and standard deviations
- All computational should be conducted on training dataset
- Training dataset can be split into training and validation

Wind Speed/Power Prediction

- Ideas can be more inspiring

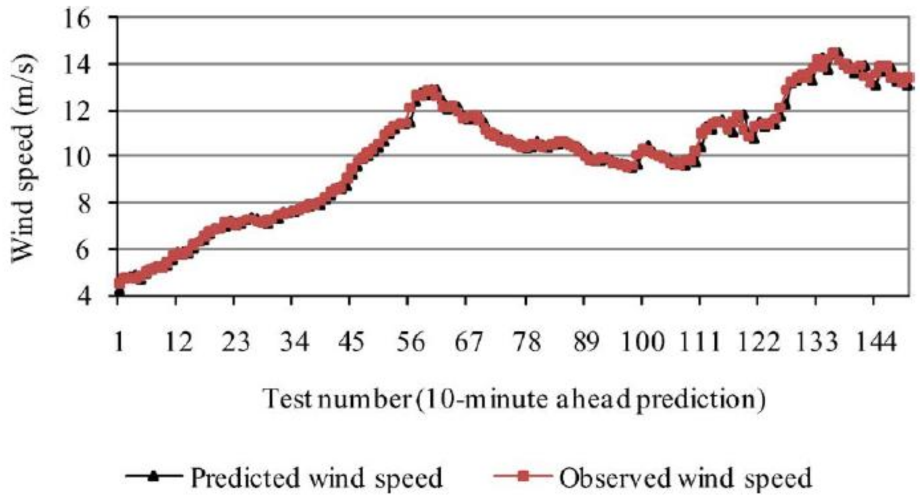


Wind Speed/Power Prediction

- Illustration of results – Tables and Figures

TABLE III
ERROR STATISTICS OF DIFFERENT MODELS BASED ON DATASET 3 OF TABLE I

Algorithm	Mean Absolute Error (m/s)	Absolute Error Std (m/s)	Mean Relative Error (%)	Relative Error Std (%)
SVMreg	0.198	0.181	3.514	5.730
MLP	0.266	0.388	4.661	6.739
M5P tree	0.206	0.267	3.953	6.035
REP tree	0.249	0.256	4.576	6.564
Bagging tree	0.202	0.198	3.665	5.723



Wind Speed/Power Outputs

- kNN based integrated prediction model
  - For new data coming in, divide it into X and Y
  - Compute the distance of the new data to past records in the whole dataset based on X
  - Find K nearest historical records close to X
  - Average their predictions of ys to make it as a prediction of y of the new data, or stack a new regression model on top of it

Wind Speed/Power Prediction

- Time series models only consider past records of the same parameter

- E.g., X = {yt-1, yt-2, …, yt-m} in the paper

- It is also to incorporate exogenous inputs, but must be records of them in the past as well