**Data description:**

As part of this dataset, 44064 simulations have been performed, one for each of the following discretized operating conditions: fault location (line on which the fault happens-34 lines), system loading (range 0.7-1.2 in 0.1 steps), active power output of the 3 wind generators (range 0-1 in 0.2 steps). This gives in total 34\*6\*6\*6\*6=44064 different operating conditions.

Description of the filename:

no#: Number of simulation (range 1-44064)

line: line on which the initial fault occurs

load: system loading (range 0.7-1.2 in 0.1 steps)

wind1, wind2, wind3: Active power output of wind generators: NSG\_1, NSG\_2, NSG\_3 (range 0-1 in 0.2 steps)

File name format: no#\_line\_load\_wind1\_wind2\_wind3.csv

Input: Time series data

File name example: “1\_Line 04 - 05\_load=0.7\_wind=0\_0\_0.csv”

Output: Cascading events sequence

File name example: “1\_Line 04 - 05\_load=0.7\_wind=0\_0\_0 results\_cascs.csv”

Summary of the simulations:

flag: A variable set to describe the output of the simulation that can take one of the following values:

* 0 if no cascading event appeared in the simulation
* 1 if a cascading events sequence appeared
* 2 if the algorithm did not converge
* 3 if an error appeared during the simulation

(Simulations flagged as ‘2’ or ‘3’ should not be used for the dataset)

In ‘results\_summary.csv’ there is a summary of all the simulations and the events that appeared, with the columns representing the variables in the following format:

no#, line, load, wind1, wind2, wind3, flag, cascading events sequence (empty if no sequence appeared)

Input-Output connection:

For the multi-class classification problem this file and the time series files are necessary:

In ‘results\_sequences.csv’ there is a list with the cascading event sequences that have appeared sorted by times of appearance. An index (1, 2,…) is assigned to each sequence for the multi-class classification problem. The no# of the simulations that this sequence has appeared in are also listed in the file, to link the input files to the cascading events sequences. In this file also there is a list with the simulations that no cascading events happened, in order for the algorithm to be able to identify ‘normal conditions’ as well.

If the prediction of the whole cascading events sequence is not feasible, the file ‘results\_sequences\_group\_5.csv’ contains the same information as stated above, taking into account only the first 5 events of the sequence. The file ‘results\_sequences\_group\_load.csv’ contains the same information, grouping also load under-frequency events.