# Influence of raw data analysis for the use of neural networks for wind farms productivity prediction

M. Beccali*, S. Culotta*, J. M. Galletto* and A. Macaione**
*Dipartimento dell'Energia, Università degli Studi di Palermo Viale delle Scienze, Edificio 9, 90128 Palermo (*Italy*)
**Asja Ambiente Italia S.p.A. (*Italy*)

***Abstract*—In the last decade wind energy had a strong growth because of cost effectiveness of the technology and the high remunerative of investments.**

**The increase of wind power penetration in power grids, however, makes necessary the development of instruments for prediction of productivity of a wind farm.**

**This paper presents a study dealing with the capability of neural network to forecast short term production of a wind farm by the correlation of wind and energy production data. Available measures of wind parameters were related to productivity data of a real wind farm. Also wind data not strictly related to the site have been used in order to assess their possible influence on the production. After a first step of data pre-processing a statistical analysis has been done.**

**The model of input-output correlation is based on the use of artificial neural networks.**

***Index Terms*—Artificial neural networks, multi layer perceptron, wind data, wind energy production.**

## I. Introduction

Among renewable energy technologies, wind farms became more and more attractive in the last decade in many countries. Wind farms installations dramatically increased in areas where public incentives together with climatic conditions, topography and environment have allowed their development. High wind energy penetration has in many cases caused problem of grid stability and balance due to the fluctuating and unpredictable nature of its generation.

Obviously, a good knowledge of the wind characteristics is the prerequisite for good planning and implementation of any project of wind energy [1,2].

In future, the presence of a high wind power in the electricity grid will cause problems in many grids, at least in terms of network management

The strategies faced in this work aimed to being complementary to the tools of management of the power systems.

This work deals with the demonstration of the potential capability of predictive algorithms based on artificial neural networks (NN) [3] for the forecast of energy production of a given wind farm.

The strength of these methods is based on their ability to help the system operator during the on line scheduling of the network in particular when wind energy contribution to the energy mix is high (period where consumption is low while wind speed is high). In such situations the network's balance is difficult to maintain[4].

Neural networks have the capability to process complex input-output data sets in order to predict events from other observed phenomenon even if not clearly or physically correlated. This can be done through the creation of "intelligent" systems able to find, through a dynamic process based on multiple iterations, the relationship between environmental variables (wind speed and direction, temperature, pressure, humidity) and output parameters (electricity production)[5].

The case study presented here aims to estimate the energy production of the farm under consideration through the correlation of the production and wind data from public and private networks (SIAS-Sicilian Agrometeorological Information Service and CNMCA-National Meteorology and Climatology Aeronautics Center )

The use of wind data not closely related (for position and height of measurement) to the investigated location is a brand new approach. In fact, the use of wind distribution physical models is generally limited to micrositing studies mainly voted to optimize the layout of the plant. Models implementing complex terrain characteristics for large areas surrounding the wind farm are very difficult to manage.

It is known that, especially during micrositing phase, wind data needed to estimate the productivity, are collected in the areas closest to the plant and at heights comparable with those of wind turbines (typically 50 m).

## II. Case Study

### A. Wind farm characteristics

This work was carried out using wind and power data of a wind farms located in Sicily composed by 11 turbines of 850 kW. Wind parameters have been measured by the station CNMCA of Trapani and from those of SIAS in Mazara del Vallo, Trapani Fulgatore and Castelvetrano located in the area of the wind farm whose position is shown in picture 2.

978-1-4244-8927-5X/11/$26.00 ©2011 IEEE

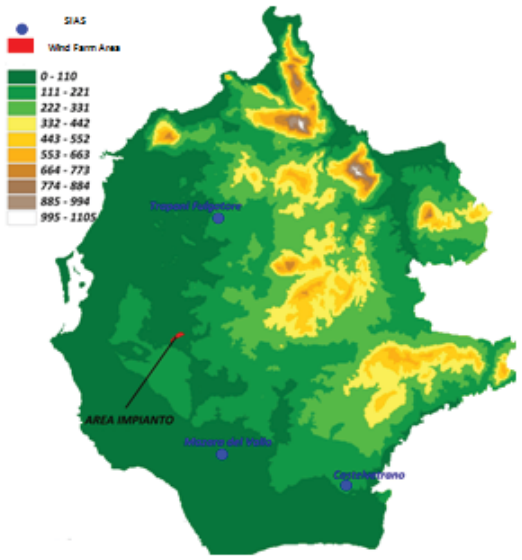


Fig. 1. Topography of the province of Trapani, the plant area and location of stations SIAS

Figure 1 shows that the provi nce of Trapani presents a large area with lev el o rography (altitu des rang ing between 0 an d 250 m ) as th ey fa ll with in th e anemometric stations and the farm object of this study.

Wind farm act ually g rows on a h illside ri dge with a peak hei ght equal to 136 m , while t he surrounding areas are all below that level.

The e xpected wind fa rm load f actors is ab out 200 0 hours per year.

### *B. Data Characteristics*

In order to implement a fo recast model for wind farm the authors have used artificial neural networks. The limit of t he neural network i s often re presented by t he dat a quality.

The dat a (averaged i n e very 10 m inutes) were collected by t he si te manager. Tw o wi nd speed sens ors are installed at h eight 50 m. In case of one sensor failing, the o ther still op erates, thus en suring con tinuous data collection. Two wind direction sensors are in stalled at height 50 m . The power data from each turbi ne a re available from 28 April 2005 to March 2008.

The other data used are wind speeds data measured by the a nemometer station CNMCA and the network of SIAS. Table I s hows a sum mary of t he a vailable anemometric data.

TABLE I
SUMMARY OF WIND DATA

| Station Code | | Network | Sensor height (m) | Data Collection Period | Acquisition Time |
|---|---|---|---|---|---|
| Mazzara del Vallo | M SI | AS | 10 | 02-07 | 1h |
| Castelvetrano | C SI | AS | 10 | 02-07 | 1h |
| Trapani Fulgatore | TF SI | ASs | 10 | 02-07 | 1h |
| Trapani Birgi | T | CNMCA | 10 | 02-06 | 1h |
| Mast 50 Sud | V1 | Private | 50 | May 05-Dec 07 | 10’ |
| Mast 50 Nord | V2 | Private | 50 | May 05-Dec 07 | 10’ |

First it was necessary to process the raw data in order to d etermine th e days on which su rveys were no t available, t o id entify i ncorrect data , to remove days/months/years i n w hich t he n umber of missing dat a was hi gh. A fterwards, t he time of rel iable data has been reduced to 2 years (2005 and 2006).

Figure 2 shows that the trend of power data is similar from 2005 to 2006 in all turbines and very different from these in 200 7 in w ind tur bines T007 , T008, T0 09, T0 10 and T011.

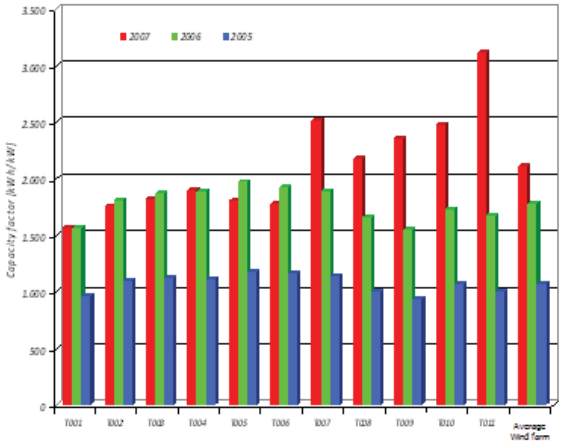


Fig. 2. Capacity factor

## III. DESIGN OF ARTIFICIAL NEURAL NETWORKS MODEL

Neural Netwo rks (NN) h ave th e ability to learn from past e xperiences an d t hen ap ply t heir knowledge t o new circumstances. Th is is p ossible th anks to the ab ility to create a represe ntative syste m of the m ultiple relationships between ra ndom vari ables of a c omplex system an d a h igh ap titude to exp ress assessm ents o n a regular b asis ab out ap propriate situ ations app arently chaotic.

The process of creating a nd eval uation of neural systems [6] is shows in figure 3.

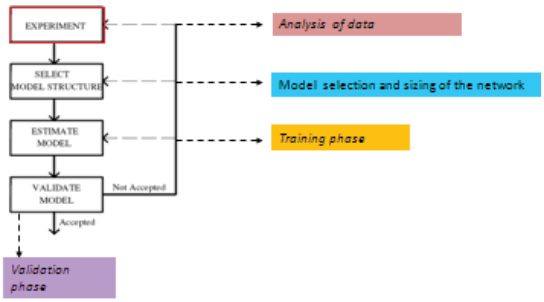


Fig. 3. Procedure for the creation and evaluation of neural systems

Because the available data for every year are 8760 and the nu mber of inp uts parameters is v ariable (fro m a minimum of 3 t o a m aximum of 6 ) t he aut hors has decided to use in a first step a networks with 100 neurons and a number of iterations, in the training phase, equal to 100 [7].

By li miting the field of so-called MLP (Multi Layer Perceptrons) i n th is st udy were tested two typ es of models :

- Models NN ARX (Neural Netw ork Autoregressive exogenous signal);
- Models NNA RMAX ( Neural N etwork Au to Regressive M oving Ave rage eXosogenous signal)

The software u tilized for the NNs m odeling was Matlab [8].

In order to create the m odel it is also ne cessary t o define t he set of t he t raining data t o be used f or t he "learning process". F or t he val idation of t he m odel another set of d ata is u sed. The v alidation set tests th e ability of generalization of the network [9].

### A. Training Phase

Authors used for the learni ng pro cess th e dataset o f 2006. T his y ear was t he o nly one having the com plete series of wind and power data.

The forecacast processing was carried out for different time steps from 2 to 5 hours.

The performance of the NN model evaluated using the NSSE ( Normalized Sum Sq uare Er ror) de fined by t he equation (1):

$$NSSE = \frac{\sum_{k=1}^{n}\left(Y_{REAL} - Y_{PREDICTION}\right)^2}{\sum_{k=1}^{n}\left(Y_{REAL} - \overline{Y}_{REAL}\right)^2} \quad (1)$$

Where:

- $Y_{REAL}$ is energy power of the wind farm;
- $Y_{PREDICTION}$ is the output value from the neural network after training;
- $\overline{Y}_{REAL}$ is the value of average power of all training;
- $n$ is the number of values of power output (8760 minus the time step).

Low values of this index imply best performances.

Table II shows a s ummary state ment for NNARX networks f or scan ning t ime wi th 1 00 neurons, i nput variables and time step 2-3 hours.

TABLE II
SUMMARY STATEMENT NNARX MODEL

| Station | Input e | NSSE time step 2 | NSSE time step3 |
|---|---|---|---|
| Mazara M | -V1 | 0,072 | 0,059 |
| Castelvetrano | C-V1 0, | 079 | 0,066 |
| Trapani Fulgatore T | F-V1 | 0,079 | 0,061 |
| Dir 50m | D-V1 | 0,093 | 0,082 |
| SIAS M | -C-TF | 0,054 | 0,042 |
| SIASL M | -C-TF-V1 | 0,051 | 0,037 |
| SIAS2L M | -C-TF-V1-V2 | 0,049 | 0,034 |
| SIAST2L M | -C-TF-T-V1-V2 | 0,045 | 0,030 |

It is worth noting that the higher the number of inputs the lower the error while the use of wind direction data has a small influence on the NN performance. Moreover, it can be o bserved th at NNAR X m odels using b oth spee d an d direction data at 5 0 m ab ove gr ound lev el ( D-V1) has higher errors that the other models running with the same time step.

With the m odel NNARMAX have been carried out o nly simulations f or t he networks t hat h ave given better performance (Table III)

TABLE III
SUMMARY STATEMENT NNARMAX MODEL

| Station I | nput | NSSE time step 2 | NSSE time step3 |
|---|---|---|---|
| SIASL M | -C-TF-V1 | 0,050 | 0,039 |
| SIAS2L M | -C-TF-V1-V2 | 0,046 | 0,038 |
| SIAST2L M | -C-TF-T-V1-V2 | 0,043 | 0,033 |

Results o f the training ph ase d idn't id entify a th e b est model between NARX and NARMAX because the NSSE are very similar.

The performance evaluation of neural networks, in the training ph ase, is co mpleted by t he c omparison between the predicted and the real output (Figure 4).

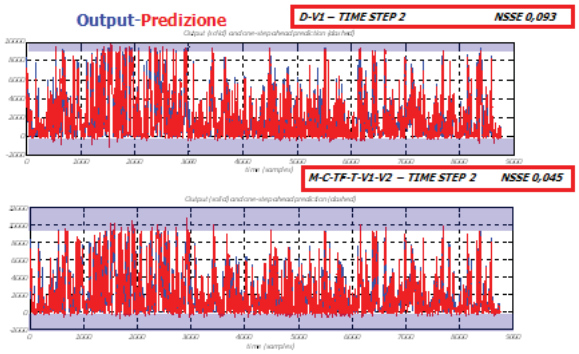


Fig. 4. Comparison of Output-Prediction in network D -V1 and SIAST2L time step 2 (NNARX)

Another big issue was detected in the management of data rel ated t o nul l or v ery l ow production by t he wi nd farm. This is due to t he presence i n the t raining set of events with wind spee d o r po wer pr oduction n ull. This data are ge nerally rel ated t o fai lures or not ordinary service. T hese events are misrepresented duri ng the training phase of the networks [3].

### B. Validation Phase

For the validation phase it was decided to use a subset of data (speed and production) detected into 2005.

This set it was ch osen with th e go al to avoid outputs with nu ll v alues of pro duction and with a good correlation between power and velocity acco rding to t he rated power curve of the turbines

Figure 5 s hows t he t rend o f wi nd p ower production and wind spe eds i n t he pe riod bet ween 8: 00 am of 16/12/2005 and 1 5:00 of 1 9/12/2005 wi th 80 events per hour.

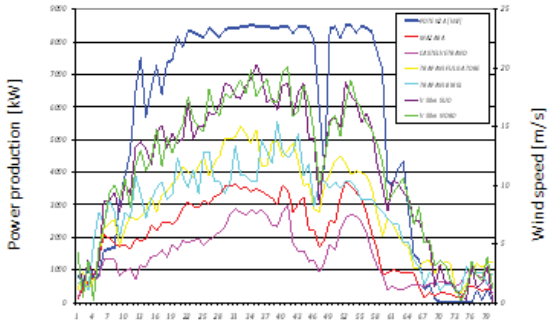


Fig. 5. Trend of wind power production and wind speeds

The results (two examples are shown in Figures 6 and 7) show some deficiency of the network in productivity forecasting.

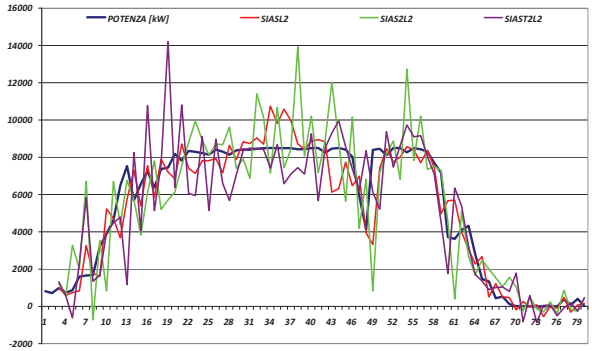


Fig. 6. Comparison output of NNARX neural network with time step 2 and actual data

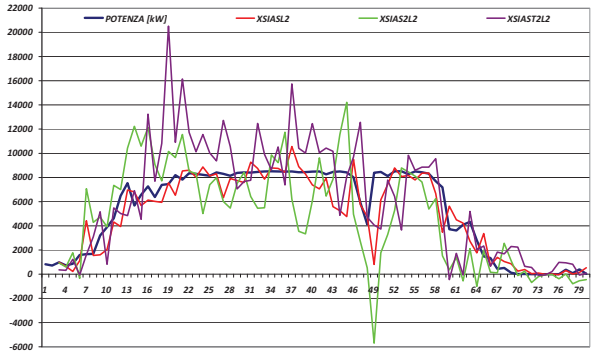


Fig. 7 Comparison output of NNARMAX neural network with time step 2 and actual data

It is worth noting that NNs with complex architectures while are capable to well describe the phenomena during the training phase, give unsatisfactory results when run with a new set of data [7].

Starting from this assumption, authors tried to optimize the model reducing its complexity with the aim to find the best architecture.

### *C. Artificial Neural Networks with 10 Neurons*

The authors have therefore decided to repeat the test with a less complex neural network with only 10 neurons, numbers of input more than 3 and time step less than or equal to 3 hours. It is expected that the reduction of the complexity of the network leads to higher errors in training phase but to a better performance in the validation phase.

Results of the NARX training phase confirmed that the increase of the number of input had a good influence on NSSE (Table IV).

Otherwise, this fact is not clearly represented for the NNARMAX1 model. (Table V).

During the training phase NSSE of both models (with 10 neurons) are very similar and higher of the equivalent networks with 100 neurons

TABLE IV
SUMMARY STATEMENT NNARX MODEL WITH 10 NEURONES

| Station I | nput | NSSE time step 2 | NSSE time step3 |
|---|---|---|---|
| SIASL2_10 M | -C-TF-V | 0,0962 | 0,0919 |
| SIAS2L2_10 | M-C-TF-V-V 0, | 0960 | 0,0914 |
| SIAST2L2_10 | M-C-TF-T-V-V 0, | 0939 | 0,0908 |

TABLE V
SUMMARY STATEMENT NNARMAX MODEL WITH 100 NEURONES

| Station I | nput | NSSE time step 2 | NSSE time step3 |
|---|---|---|---|
| XSIASL2_10 M | -C-TF-V | 0,0951 | 0,0923 |
| XSIAS2L2_10 M | -C-TF-V-V | 0,0949 | 0,0915 |
| XSIAST2L2_10 | M-C-TF-T-V-V | 0, 0928 | 0,1152 |

For the validation of the NN authors have used the same set of data used for the networks with 100 neurons.

The figures 8 and 9 show the results of NNARX and NARMAX networks with time step 2 hours.

It can be observed that there is an improvement in the forecast capability. In fact the fluctuations around the real value decrease. This effect is true for all the networks investigated.

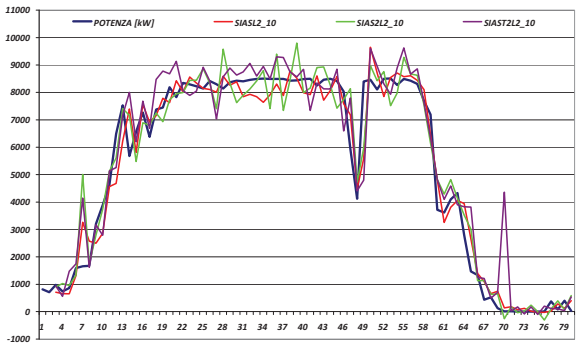


Fig. 8. Comparison of output of NNARX neural network (10 neurons and time step 2) and actual data

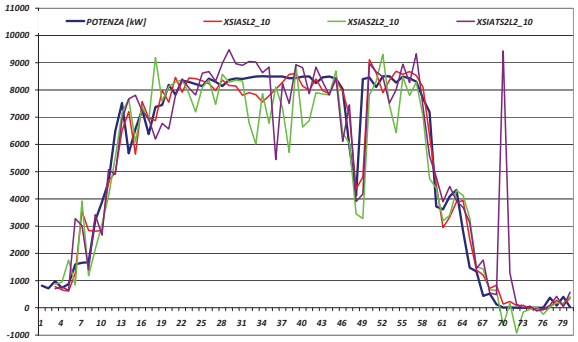


Fig. 9 Comparison of outputs of NNARMAX neural network (10 neurons and time step 2) and actual data

In order to make a more detailed comparison among the simulation outputs obtained by the several NN models the normalized error was calculated using the following equation:

$$E_{iNORMALIZED} = \frac{\left|Y_{iREAL} - Y_{iPREDICTION}\right|}{Y_{iREAL}} \quad (2)$$

where Y represents the energy production in the time step.

Maximum, minimum, sum, average, variance and standard deviation values of the errors have been calculated in order to assess the performances of each models. Results are given in the following tables.

Following tables show an analysis of the performances for the different models under investigation.

TABLE VI
ANALYSIS OF THE NORMALIZED ERRORS OF NNARX MODEL 100 NEURONS WITH TIME STEP 2

| SIASL2 | | SIAS2L2 | SIAST2L2 |
|---|---|---|---|
| Max 193, | 92 | 749,36 | 996,22 |
| Min 0, | 000 | 0,001 | 0,003 |
| Sum 479, | 433 | 1398,035 | 1815,404 |
| Average 6, | 147 | 17,924 | 23,274 |
| Variance 912, | 71 | 7963,35 | 14399,55 |
| St deviation | 30,211 | 89,238 | 119,998 |

TABLE VII
ANALYSIS OF THE NORMALIZED ERRORS OF NNARX MODEL 10 NEURONS WITH TIME STEP 2

| SI | ASL2_10 | SIAS2L2_10 | SIAST2L2_10 |
|---|---|---|---|
| Max 47, | 27 | 604,16 | 427,40 |
| Min 0, | 001 | 0,003 | 0,000 |
| Sum 182, | 252 | 779,731 | 1080,337 |
| Average 2, | 337 | 9,997 | 13,850 |
| Variance 58, | 58 | 4734,69 | 4459,46 |
| St deviation | 7,654 | 68,809 | 66,779 |

It is worth noting that by reducing the number of neurons, and thus the complexity of the neural network, the NNARX model is certainly the one that best describe the events presenting for all cases a lower average normalizes error (ranging from 2.34 to 13.85).

TABLE VIII
ANALYSIS OF THE NORMALIZED ERRORS OF NNARMAX MODEL 100 NEURONS WITH TIME STEP 2

| XSIAS | L2 | XSIAS2L2 | XSIAST2L2 |
|---|---|---|---|
| Max 60, | 138 | 761,100 | 1971,960 |
| Min 0, | 000 | 0,000 | 0,011 |
| Sum 197, | 363 | 1193,698 | 2668,408 |
| Average 2, | 530 | 15,304 | 34,210 |
| Variance 73, | 205 | 8042,871 | 50675,225 |
| S deviation | 8,556 | 89,682 | 225,112 |

TABLE IX
ANALYSIS OF THE NORMALIZED ERRORS OF NNARMAX MODEL 10 NEURONS WITH TIME STEP 2

| XSI | ASL2_10 | XSIAS2L2_10 | XSIATS2L2_10 |
|---|---|---|---|
| Max 132, | 456 | 469,260 | 926,954 |
| Min 0, | 005 | 0,003 | 0,002 |
| Sum 285, | 894 | 978,263 | 1317,276 |
| Average 3, | 665 | 12,542 | 16,888 |
| Variance 252, | 584 | 4046,147 | 11267,524 |
| S.deviation | 15,893 63, | 609 | 106,149 |

Different result has been experienced for NNARMAX model. In this case the reduction of the complexity of the network is not automatically related to an improvement of the performances. This fact is due to the influence of the data set content.

### *D. The Effect of Preliminary Data Management in NN Performance*

Previous results showed that the worst performances of the NN models are mainly related to presence of negative values in the output of networks. These values are generally related to measured data of the power production near to zero. For this reason authors decided to improve a step of analysis of raw data with the aim to reduce the noise of information processed by the neural network.

The data with a low energy production or having a relevant (error > 20%) difference form the rated power curve have been deleted from the test set (Fig.10).

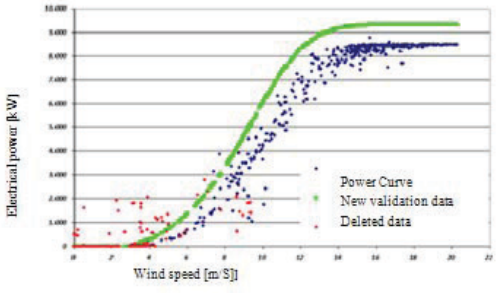


Fig. 10 Actual and theoretical working points of the wind farm in the set of validation

With the new data sets derived from a more precise filtering of raw data, network performance is significantly improved. The average normalized error ranges from 0.123 to 0.200 for networks NNARX (10 neurons) (figure 11) and from 0.127 to 0.196 (figure 12) for networks NNARMAX (10 neurons).

The 12 bars reported in figures 11 and 12 represents for each tested model the two errors obtained by the

simulations either with the original data set either with the reduced data set (the bars with the arrow pointers).

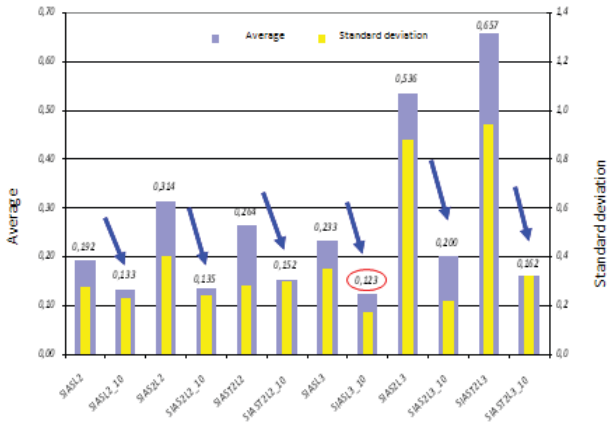


Fig. 11 Average normalized errors for NNARX networks with original and reduced data sets (with arrow pointers)

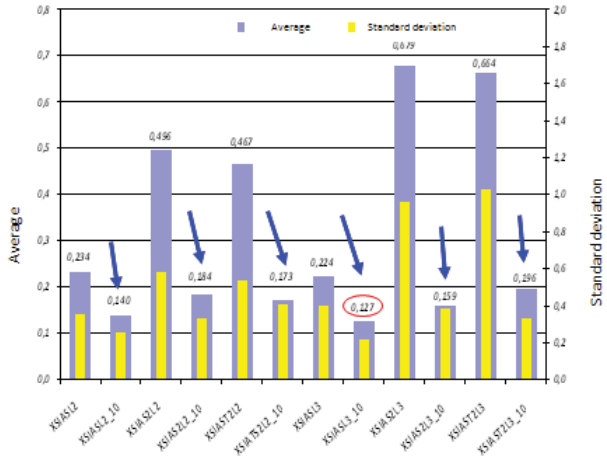


Fig. 12 Average normalized errors for NNARMAX networks with original and reduced data sets (with arrow pointers).

## IV. Conclusions

The aim of this work is show how artificial neural network can be utilized as a tool to predict the wind energy output using wind data not strictly related to the plant site as input. On the other hand the paper showed how input data quality can influence the NN performance.

Different configurations of NN were generated and test under several conditions. Results obtained using raw data appeared to be unfair in order to suggest NN as a real prediction tool. Therefore the authors have decided to improve the analysis of raw data with a new and more accurate data filtering. The aim was to reduce the noise of information processed by the neural network.

It should be stressed that the presence in the input and output data of subsets of inconsistent events (not null with wind speed and null energy production) has a big influence on the neural network performance.

For this reason it is very important during the step of validation of data to consider the power curve of the turbines in order to discard data not fitting with it.

Also singularities deriving from plant maintenance, grid interruption, etc. must be carefully considered.

Obviously, an ordinary maintenance plan repeated in the several years in the same period or days, would be "recognized" by the network with an egligible effect on final errors. On the other hand the presence of negative output values was observed during both training and test.

These events occurred in the presence of sudden decreases in wind speed or in presence of no wind data or no production data.

With the new data set derived from a more precise filtering of raw data, network performance was significantly improved.

## References


[1] Sathyajith M., Pandey K.P., Kumar.V. A. *Analysis of wind regimes for energy estimation*. Renewable Energy 25 (2002), pp.381–399.

[2] Cancino-Solorzano Y., Xiberta-Bernat J. *Statistical analysis of wind power in the region of Veracruz (Mexico)*. Renewable Energy 34 (2009), pp.1628–1634.

[3] Carolin Mabel M., Fernandez E. *Analysis of wind power generation and prediction using NN: A case study* , Renewable Energy 33 (2008),pp.986–992.

[4] Blombou R.. *Very short-term wind power forecasting with neural networks and adaptive Bayesian learning* . Renewable Energy 36 (2011), pp.1118–1124.

[5] Cadenas E., Rivera W. *Short term wind speed forecasting in La Venta, Oaxaca, Mexico, using artificial neural networks*. Renewable Energy 34 (2009), pp.274–278.

[6] Nørgaard M., Neural Network Based, System Identification, TOOLBOX, version 2, Department of Automation, Department of Mathematical Modeling, Technical University of Denmark, 2000

[7] Marvuglia A. *Utilizzo di reti neurali artificiali a supporto della pianificazione energetica* , Tesi di Dottorato, Dottorato in Fisica Tecnica Ambientale, 2007;

[8] Vesanto J, Himberg J, Alhoniemi E, Parhankangas J., (2000). SOM toolbox for Matlab 5, Tech Rep A57, Helsinki University of Technology;

[9] Grassi G., Vecchio P.. *Wind Energy prediction using a two- hidden layer neural network*. Commun Nonlinear Sci Numer Simulat 15 (2010), pp. 2262-2266