New Concepts in Wind Power Forecasting Models

Vladimiro Miranda, Ricardo Bessa, João Gama, Guenter Conzelmann, Audun Botterud

Abstract – This paper reports new results in adopting entropy concepts to the training of mappers such as neural networks to perform wind power prediction as a function of wind characteristics (mainly speed and direction) in wind parks connected to a power grid. Renyi's Entropy is combined with a Parzen Windows estimation of the error pdf to form the basis of three criteria (MEE, MCC and MEEF) under which neural networks are trained. The results are favourably compared with the traditional minimum square error (MSE) criterion. Real case examples for two distinct wind parks are presented.

***Index Terms*—Wind power forecasting, neural networks, correntropy, entropy, Parzen windows.**

I. Introduction

THE demand for more accurate short term wind power forecasting models has led to solid and impressive development in recent years. Two basic requirements are behind this drive: security of operation and cost.

It is at present consensual in the scientific community that 3-day-ahead wind power prediction requires some form of a two-step approach: a meteorological meso-scale model to predict general characteristics of wind speed and direction at a given geographical location and a local model that may transform wind speed predictions into wind power generation of a wind farm.

However, it is surprising that little attention has been given so far to the requirements behind the wind power forecasting exercise. Instead, most research activity has been conducted under what we may call the "signal processing paradigm", where the sole concern is to try to generate a prediction that matches the actual signal. This paradigm is mostly concerned with minimizing the prediction error. However, not much discussion has been witnessed on the nature and consequences of the errors.

The short-term forecasting techniques use models that try to map a function linking wind to power whose analytic form is unknown. These models depend on internal parameters that have to be tuned in a supervised training mode and there has been a generalized adoption of a criterion for measuring the performance quality in terms of the MSE (Minimum Square Error) of the error distribution.

This criterion is only an optimal choice if the error distribution is Gaussian. The reason is that MSE is equivalent to minimizing the variance of the error distribution. Gaussian distributions are the only ones that contain all information in their first two moments (linked with mean and variance); all other cumulants are zero. On the contrary, other distributions have meaningful skewness, kurtosis and other indices related with higher moments; therefore, mean and variance do not represent all information contained in the distribution.

It is a fact recognized by researchers that wind predictions, namely coming from the meso-scale meteorological models, exhibit errors that are not Gaussian in nature. If one considers that a wind farm also introduces a non-linear input-output relation in terms of wind speed/power generated, errors from the wind power production model are surely not Gaussian in nature – therefore, the adoption of the MSE criterion is certainly not the best tool available to build an accurate wind forecasting system.

However, the published material on wind power prediction does not seem to take this fact in account and the MSE criterion is widely used when it comes to tuning parameters, namely when one builds a model around a basic concept of neural networks.

The signal processing community, however, having recognized this issue in the generalized problem of training mappers, has produced an approach, sometimes called ITL – Information Theoretic Learning, which resorts to the concept of information content of the error distribution, instead of variance. And as Entropy is a concept that can be translated as a measure of the information content of a probability density function, criteria based on minimizing Entropy (like MEE – Minimum Error Entropy) of functions related with entropy were devised.

The rationale behind this approach lies in the fact that an entropy evaluation may allow one to extract more information from data that just using a variance-based method, because entropy takes in account all the moments of a probability distribution. Therefore, one may hope in wind power prediction to obtain error distributions with higher frequencies close to zero, thus approximating a Dirac function, which has minimum entropy.

Furthermore, this approach will be less sensitive to outliers and noise, which is a valuable property when building a model for the real world, where data are usually contaminated.

The first works in this direction with real data on wind power have given extremely promising results. Paper [1]

V. Miranda is with INESC Porto, Instituto de Engenharia de Sistemas e Computadores do Porto, Portugal, and also with FEUP, Faculty of Engineering of the University of Porto, Portugal (email: vmiranda@inescporto.pt).

R. Bessa is with INESC Porto, Portugal (email: rbessa@inescporto.pt).

J. Gama is with INESC Porto LA/LIAAD, Laboratory of Artificial Intelligence and Decision Aid, and with FEP, Faculty of Economics of the Universitry of Porto, Portugal (email: jgama@liaad.up.pt).

G. Conzelmann and A. Botterud are with Argonne National Laboratory, USA, Decision and Information Sciences Division (emails: guenter@anl.gov, abotterud@anl.gov).

reports that a Takagi-Sugeno Fuzzy Inference System trained under an Entropy criterion originated a better prediction than the same system trained under MSE. In [2], one confirms this for neural networks and extends the research to other criteria related with entropy but that display better computing efficiency.

In this paper we summarize the concepts behind the Entropy-based approaches and illustrate with some results how one can produce better models, in the signal processing sense, by using Entropy instead of the Minimum Square Error criterion. But the paper is not limited to this description. Once having at hand two distinct methodologies to generate wind power predictions, one has paid attention to the consequences of the prediction exercise.

First of all, it is important to recognize that there are different users of wind power predictions, the most important groups being wind farm owners that sell in the market and ISOs that must ensure the security of the system. These groups may have conflicting interests. Then, we must realize that prediction systems have errors, will always have errors. Therefore, at some point one must ask if the nature of the errors (translated into their probability density function) is neutral to the users or if, on the contrary, users may take advantage (and illicit advantage) from the specific nature of the error distribution.

This paper gives some contribution to the discussion of this topic, by comparing, in the Spanish (Iberian) market, the economic results of adopting either the MSE or an Entropy criterion in a prediction system. The real cases analyzed demonstrate that, in fact, the choice of a training criterion is not neutral and that a choice that may be convenient to wind power generators may not be acceptable to ISOs.

The work reported in this paper is one of the motivations behind the project led by ANL – Argonne National Laboratory, together with INESC Porto from Portugal, to develop a wind power prediction platform that may serve as test bed and as benchmarking for other models.

II. Remembering That Wind Power Prediction Errors Are Not Gaussian

To build our case, first, we illustrate that errors in wind speed predictions coming from meso-scale models are not Gaussian. Wind speed predictions 3 days ahead are generated from a NWP (numerical weather prediction) model, namely the well known MM5 mesoscale model for mean wind speed and wind direction, for a reference point in the wind farm (see http://www.mmm.ucar.edu/mm5/).

Figure 1 displays one year of the NWP/MM5 prediction errors against real wind speed values measured by a metering station at a wind farm in Portugal. It might seem that the NWP error could be approximated by a Gaussian distribution, but if we apply the *Kolmogorov-Smirnov* test to the error data then the null hypothesis (error with normal distribution) is rejected.

This distribution will be passed through a non-linear model relating wind speed to power output in the wind farm. It is now clear that the wind power prediction errors cannot be Gaussian.

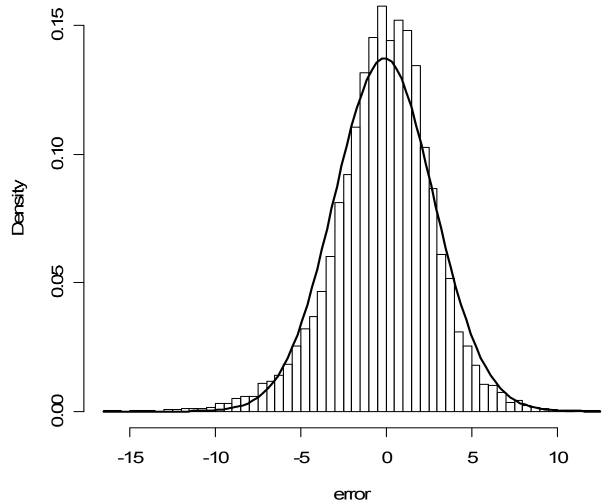

Figure 1 - MM5 wind speed prediction error for a wind farm in Portugal and the normal distribution (curve) with mean of -0.10 and standard deviation of 2.90 – they do not match.

III. Training Mappers

A mapper is a term used to designate a Neural Network, a Fuzzy Inference System or, in general, any system that emulates an input-output transfer function and whose performance depends on the tuning of internal weights or parameters.

We can identify in a mapper three basic modules: its internal structure, the performance criterion and the mechanism of training. Three types of actions can be applied in order to condition the training of a mapper (Figure 2) :

a) in the internal structure, by modifying the weights
b) in the performance criterion, by selecting an adequate measure of performance
c) in the training mechanism, by choosing an algorithm or procedure to close a feedback loop that updates the weights as a function of the performance criterion.

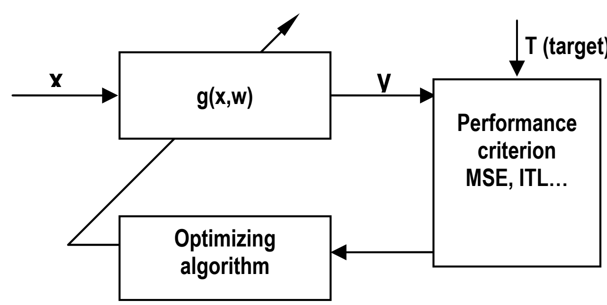

Figure 2 – Basic arrangement of a mapper identifying its three main modules

The most important aspect dealt with in this paper is the performance criterion. The basic idea is the following: if the error distribution of the output would become a Dirac function (meaning that all errors would be equal), we would have reached a predictor whose output would reproduce exactly the actual data series – by just adding to the results a bias corresponding to the mean of the pdf of the errors, i.e., the deviation from zero (see Figure 3).

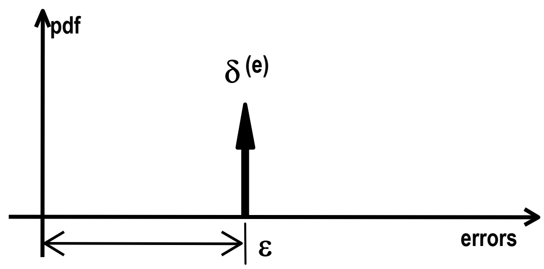

Figure 3 – A mapper producing a systematic error ε for all inputs will display a error density function like a Dirac function

But it so happens that the Dirac function has minimum Entropy. If the objective of model development is to discover weights W that lead to a pdf of errors as much approximated as possible to a Dirac function, then we would be minimizing the Entropy of the error distribution. The success of ITL is in having discovered a cost function representing this objective and having set up a manageable procedure to compute the solution.

IV. Entropy And Parzen Window Pdf Estimation

The most well known Entropy definition, in Infornation Theory, is Shannon's Entropy. However, its formulation is not convenient to build a computing algorithm for training a neural network. Instead, the ITL approach used Renyi's definition of entropy.

Renyi's entropy [3] of a discrete probability distribution P = (p_1, p_2,…, p_n) is defined as

$$H_{R\alpha} = \frac{1}{1-\alpha} \log \sum_{k=1}^{N} p_k^{\alpha} \quad \text{with} \quad \alpha > 0, \alpha \neq 1 \tag{1}$$

Renyi's entropy is a family of functions $H_{R\alpha}$ depending on a real parameter α. When α = 2, we have what is called quadratic entropy

$$H_{R2} = -\log \sum_{k=1}^{N} p_k^2 \tag{2}$$

This definition can be generalized for a continuous random variable Y with pdf $f_Y(z)$:

$$H_{R2} = -\log \int_{-\infty}^{+\infty} f_Y(z)^2 dz \tag{3}$$

The operational advantage over Shannon's definition becomes apparent: instead of an integral of the logarithm of the pdf, we have the logarithm of the integral, a formulation much more amenable to be massaged into an algorithm.

The estimation of the pdf of data from a sample constituted by discrete points $y_i \in R^M$, i=1,…,N in a M-dimensional space, may be done by the Parzen Window method [4]. This technique uses a kernel function centered on each point; it looks at a point as being locally described by a probability density Dirac function, which is replaced or approximated by a continuous set whose density is represented by the kernel. If a Gaussian kernel is used, the expression of the estimation $\hat{f}_Y$ for the real pdf f_Y of a set of N points is a summation of individual contributions:

$$\hat{f}_Y(z) = \frac{1}{N} \sum_{i=1}^{N} G\left(z - y_i, \sigma^2 I\right) \tag{4}$$

where G(.,.) is the Gaussian kernel and $\sigma^2 I$ is the covariance matrix (here assumed with independent and equal variances in all dimensions). In each dimension we have

$$G\left(z_k - y_{ik}, \sigma^2\right) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(z_k - y_{ik})^2} \tag{5}$$

It is easy to understand that the "size" of the window, here defined by the value of σ, is important in obtaining a smoother or more "spiky" estimate for f_Y. This property is used in controlling the behaviour and convergence of algorithms.

V. MEE – Minimum Entropy Criterion

Combining Renyi's entropy with a pdf estimate with Parzen windows [5][6] generates a practical formulation – this approach was called Information Theoretic Learning. An entropy estimator for a discrete set of data points {y} is

$$H_{R2}(y) = -\log \int_{-\infty}^{+\infty} \hat{f}_Y(z)^2 dz = -\log V(y) \tag{6}$$

where, using (4)

$$V(y) = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \int_{-\infty}^{+\infty} G\left(z - y_i, \sigma^2 I\right) G\left(z - y_j, \sigma^2 I\right) dz \tag{7}$$

In this expression we recognize the convolution of Gaussian functions, and the integral of two Gaussians with equal standard deviations is a Gaussian with twice the standard deviation. Then we have the following result:

$$V(y) = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} G\left(yi - yj, 2\sigma^2 I\right) \tag{8}$$

which allows the practical evaluation of entropy by simply calculating the Gaussian function values of the vector distances between pairs of samples. V(y) is called the information potential (IP) of the data set.

Optimizing a mapper with minimum output entropy [7] becomes the MEE criterion, for Minimum Entropy Error.

The discovery of weights in a mapper may be done by applying a suitable optimization method that will discover the weights *w* that minimize the objective function

$$\min_{w} H_{R2}(e) \tag{9}$$

This can be achieved by the classical back-propagation algorithm [7][8] but in [1] we have applied instead an evolutionary algorithm to minimize entropy: EPSO, Evolutionary Particle Swarm Optimization. The entropy concept is independent of the training method and several avenues can be explored in order to build efficient algorithms to optimize mappers because different objective functions may lead to distinct performance in different algorithms. In this paper, however, we will use a classic back-propagation scheme, adapted to deal with entropy cost functions.

The problem with MEE is that it is a time-consuming method, because of the need to calculate all differences of errors in (8), while the MSE only requires the calculation of simple errors.

VI. OTHER ENTROPY CRITERIA - CORRENTROPY

To address the computation time burden associated with the MEE criterion, other information content measures were devised. We now summarize some of the most important criteria that we have used to train neural networks in the wind power prediction problem:

1. MSE – Minimum Square Error. This is the classical criterion that minimizes the variance of the error distribution and has the form

$$MSE(\varepsilon) \Leftrightarrow min\ \frac{1}{N}\sum_{i=1}^{N}\frac{1}{2}\varepsilon_i^2 \tag{1}$$

where $\varepsilon = (T_i - y_i)$ is the error of sample i relative to the target value T_i.

2. MEE – Minimum Error Entropy. This is the fundamental ITL criterion where the minimization of the entropy of the error distribution is equivalent to maximize the information potential

$$MEE(\varepsilon) \Leftrightarrow max\ V = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} G\left(\varepsilon_i - \varepsilon_j, 2\sigma^2 I\right) \tag{2}$$

where G is a Gaussian function, in this case with a variance given by a value represented by $2\sigma^2$.

3. MCC – Maximum Correntropy. This criterion is based on a generalized similarity measure called correntropy [9][10] and may be translated by

$$MCC(\varepsilon) \Leftrightarrow max \frac{1}{N}\sum_{i=1}^{N} G(\varepsilon_i, \sigma^2 I) \tag{3}$$

4. MEEF – Minimum Error Entropy with Fiducial Points. This criterion [9] intends to anchor the error distribution to a zero mean by defining a compromise between minimizing entropy and maximizing correntropy through a cost function

$$\begin{aligned} MEEF(\varepsilon) \Leftrightarrow max\ \gamma \frac{1}{N}\sum_{i=1}^{N} G(\varepsilon_i, \sigma^2 \boldsymbol{I}) + \\ +(1-\gamma)\frac{1}{N^2}\sum_{j=1}^{N}\sum_{i=1}^{N} G\left(\varepsilon_j - \varepsilon_i, 2\sigma^2 \boldsymbol{I}\right) \end{aligned} \tag{4}$$

where γ is a weighting constant between 0 and 1.

Among these criteria, Correntropy plays an important role. Correntropy is a generalized similarity measure between two arbitrary scalar random variables X and Y defined by:

$$V_\sigma(X, Y) = E\left[k_\sigma(X - Y)\right] \tag{10}$$

where k_σ is the kernel function (usually Gaussian).

Correntropy is directly related to the probability of how similar two random variables are in a neighbourhood of the joint space defined by the kernel bandwidth, and provides the probability density of the event p(X=Y). Using Parzen windows, the bandwidth controls the observation window in which the similarity is assessed but makes one unable to assess similarity in the whole joint space.

In [10] one may find a discussion on the properties of correntropy. It is proved that a measure CIM(X,Y) = k(0)-V(X,Y) related with correntropy satisfies all the properties of a metric. CIM may be divided in three different regions: when the error is close to zero CIM is equivalent to L2 norm; when the error grows CIM becomes a L1 norm; when the error is very large CIM becomes a L0 norm, the metric saturates and becomes very insensitive to large errors. This property shows the robustness of CIM and the importance of kernel bandwidth. A small kernel size leads to a small Euclidean zone while a large kernel size will increase the Euclidean region where the metric behaves like the MSE criterion.

One can see that the MCC criterion implies that one tries to maximize the value of the pdf of the errors at $\varepsilon = 0$. If the data allows, this tends to transform the error pdf into a Dirac function at zero – therefore MCC tends to minimize entropy. However, the MCC criterion does not require the calculation of the differences in all pairs of errors and therefore the computational burden becomes similar to the MSE criterion.

VII. ON-LINE SELF-ADAPTIVE MODELS

The Entropy criteria have been modified to suit an on-line self-adaptive strategy for the neural networks to be used in the prediction system. The aspiration to have on-line training is not new in wind power prediction systems – however, models to adequately dealing with data streaming are still a matter for research. The need to adapt to new information has been readily recognized – but the need to forget learning acquired with outdated information is fairly recent.

The methodology for the online training under the MEE criterion requires that the information potential of the error is computed with a recursive formula [11]. When a new measure arrives, the prediction error of the neural network is computed and added to a time window with M errors of previous predictions. The information potential of the error is then recursive estimated using the following equation

$$V_{k+1} = (1-\lambda)V_k + \frac{\lambda}{M}\sum_{i=k-L+\backslash}^{k} k_\sigma\left(x_i - x_{k+1}\right)$$

where λ is a forgetting factor with values between 0 and 1. A window with the M most recent errors was also used. The recursion of the information potential uses the gradient from the previous time step.

VIII. COMPARING ITL CRITERIA WITH MSE

In this section we present results for training of neural networks in real wind power forecasting problems, comparing the performance of MSE and three ITL inspired criteria (MCC, MEE and MEEF). We have trained a feed forward MLP neural network with only one hidden layer comprising 7 neurons, using a hyperbolic tangent activation function. We will not describe details of the training, because the purpose is to compare results. The inputs of the MLP are NWP meteorological forecasted values: mean wind speed values, mean wind direction values. Duo to the cyclic character of the wind direction, this variable comprised two components, i.e. the sine and cosine components.

To validate the results we have run 25 simulations in each study, randomly generating weights in each case. The final conclusions derive from the average of all simulations. When comparing criteria, the same weights were used in all cases.

The real test cases refer to two wind parks located in Portugal in mountain regions, far from each other. Wind park A has a rated power of above 20 MW and comprises equal wind turbines of under 2 MW each. Wind park B has a rated power of above 15 MW and comprises a set of equal older turbines of above 0.5 MW and another set of equal turbines of about 2 MW. The data collected from the wind parks included SCADA registers with mean electric power delivered by the wind park into the substation connecting it to the electric power network.

Other model's input variables include forecasts generated from a NWP (numerical weather prediction) model for mean wind speed and wind direction, for a reference point in the wind park.

A. 24 hour ahead forecasting

The first case (Figure 4) compares the pdf of errors obtained with MEEF, MCC and MSE in batch-sequential off-line training for wind farm A for 24 hour ahead prediction.

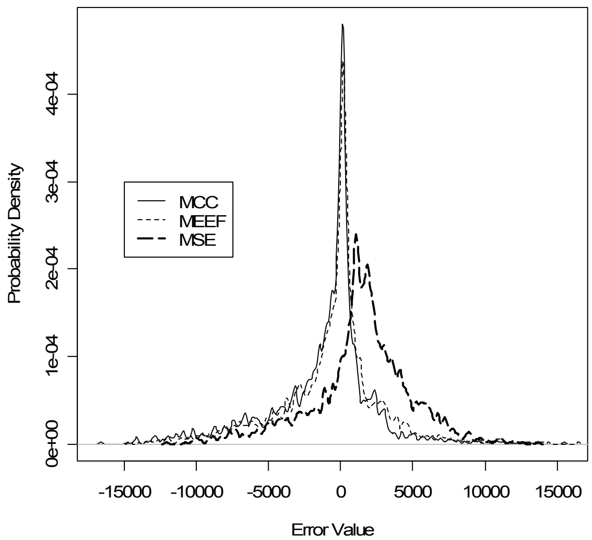

Figure 4 - Comparison of the error pdf generated by NN trained with MEEF (dotted), MCC (solid) and MSE (dotted line width) criteria for wind park A

Just by examining the figure, we confirm that the prediction errors are not Gaussian. The neural network trained under the entropy-related criteria made it possible to obtain a narrower pdf (extremely similar in results) than with the variance-based criterion MSE. In agreement with the theory, it was possible to design a mapper that produces a predictor with a higher frequency of errors close to zero.

Figure 5 shows the forecasted values for the mean 30-minute electric power for the same wind park obtained with three criteria (2 ITL criteria and MSE), when compared to the real value registered in the park SCADA. The wind speed registered is also shown. It's clear that the entropy-related criteria made a better fit and followed better the SCADA measured. In some points the MSE produced small errors, but as we see in this figure the models trained with entropy criteria produce a larger number of errors close to zero.

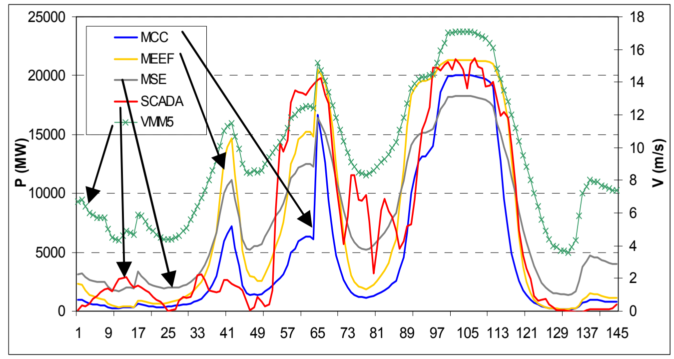

Figure 5 - Comparison among predictions by three criteria on 2005 May 26, 27 and 28 for wind park A. The graph also displays the wind speed prediction provided by the MM5 model (wind speed scale on the right)

B. 3 day ahead forecasting

The second case is about 3 day ahead forecasting. The error measure adopted to evaluate performance was the NMAE (normalized mean absolute error). This error average was calculated over all 3-day horizon windows available in the test set months.

The rationale for this choice is two-fold: it is a criterion often used by researchers working in wind power forecasting [12] and it would not introduce a bias in comparisons, because it's a criterion not used in the calculations. The errors are also compared with a classical reference model: persistence, which corresponds to admit as forecast the last known value of the generation time series.

In Figure 6 we can see the result for the 3 day ahead prediction at a specific date for wind park A, obtained with a neural network trained offline.

It is evident that the entropy-based criteria produced smaller average error than the MSE criterion. We can show with this example that the results from applying MCC and MEEF are virtually identical. Therefore, in practice the MCC criterion is the good choice, because it is much less demanding in computing effort (comparable to MSE in this respect).

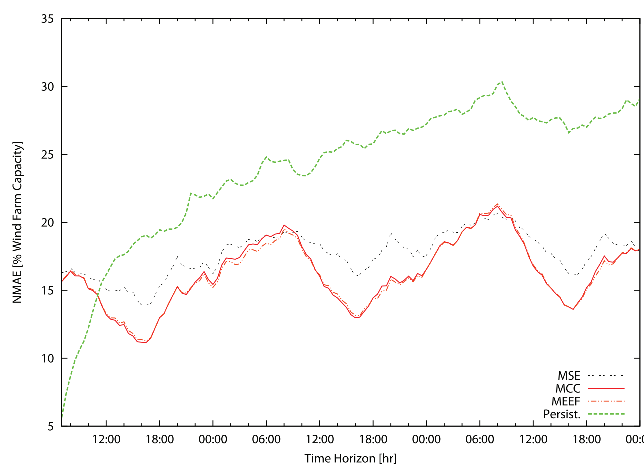

Figure 6 – NMAE error for 3 day ahead forecasts in specific days for wind park A (off-line training).

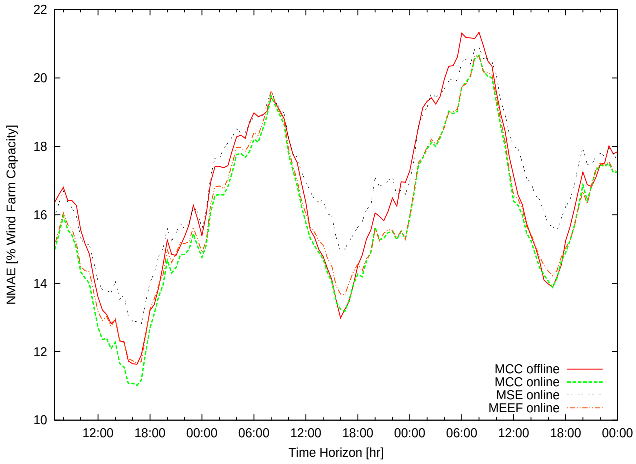

Figure 7 - Wind Park A: NMAE with distinct training criteria

In Figure 7 we can compare, for a specific day, the 3 day ahead prediction performance for the MSE criterion with on-line adaptive training, the MCC criterion for off-line training and for on-line adaptive training and the MEEF criterion for on-line adaptive training. The figure illustrates for a specific case what we have witnessed as a rule: that models that undergo self-adaptive on-line training perform better than models trained off-line. This is a result of the fact that the data streaming of wind speed values exhibits properties of concept drift – the underlying probabilistic model that could explain wind behavior is changing and therefore an off-line trained model eventually diverges from the reality it tries to emulate.

This case is very interesting because all cases with training driven by an entropy-based criterion performed better that the neural network trained on-line under the MSE criterion – even the off-line MCC criterion was better.

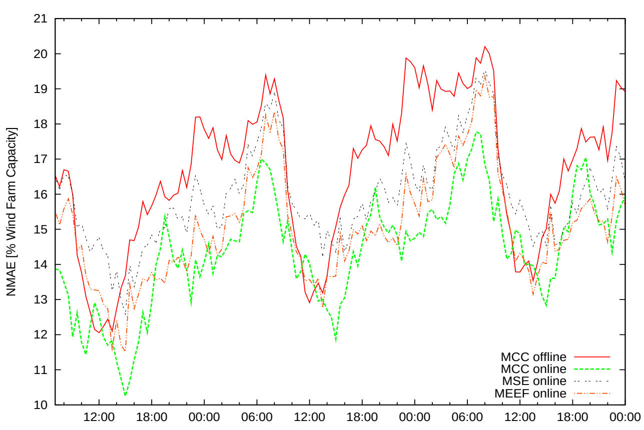

Figure 8 – Wind Park B: NMAE with distinct training criteria

In Figure 8 we see the advantage of a self-adaptive on-line training model. This case is selected from data form wind park B. A new wind generator had been added and the off-line trained model did not adapt – so, the MCC criterion with off-line training displays now clearly a worse result, much worse than even the MSE criterion, which has been applied in an adaptive on-line training fashion.

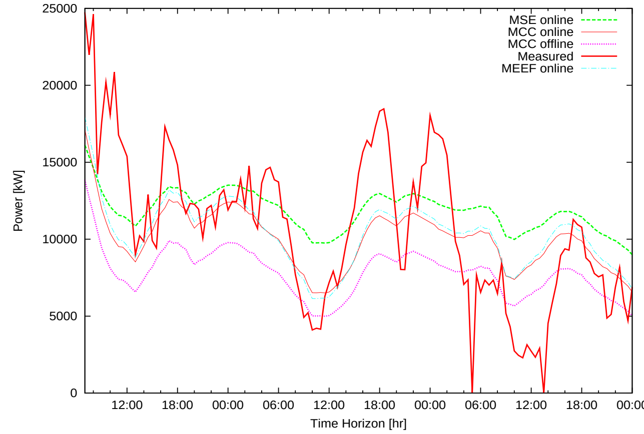

Figure 9 - Wind Park B: forecasted and measured values of the electric power obtained with online and offline training (12-14 November 2005).

In Figure 9 we can evaluate the prediction performance in a case in November 2005. The wind power 3 day ahead prediction that seems to be a better fit to the actual values later measured is the one resulting from the entropy-based criterion. One must bear in mind that the prediction curves are strongly influenced by the NWP/MM5 wind speed predictions and this explains why they seem to be somewhat parallel.

IX. Cost of Errors

Both Figure 4 and Figure 9 give evidence to a characteristic that we have confirmed in all systems studied: that predictor systems trained with Minimum Entropy and the MSE criteria display different biases in the forecasting values they generate. As a general trend, a system trained with MSE tends to produce forecasts that slightly overestimate the actual values while the Entropy criteria lead on average to inferior values. In Figure 4 this bias is quite evident for the pdf associated with the MSE criterion – there is a shift to the right of the pdf peak, when compared with the pdf associated with the Entropy criteria.

The fact that predictions are not perfect and that any system will produce some distribution of errors is trivial. However, in a market system where wind power producers may play, prediction errors have a cost or penalty associated with them. These costs are very much dependent on the rules of the market and no general conclusion can be offered. However, one can clearly state that if the penalties are asymmetric then an opportunity arises for a player to benefit from using a biased model instead of a true fair and neutral model.

Take the Spanish market (OMEL), for instance. In this market, wind power producers offer power in the day ahead market and are paid at the market clearing price. Then, if the wind generation results above or below their offer, they will be subject to a penalty:

- If the actual generation is above the offered power, the excess in paid at a discounted value.
- If the actual generation remains below the offered power, a penalty is applied according to the cost of the generation the ISO has to purchase to compensate for the lack of power.

The following Table shows the annual mean energy prices in 2007 and 2008:

	2007	**2008**
Marginal Price [€/MWh]	40.29	65.49
Down-regulation price [€/MWh]	8.97	9.10
Up-regulation price [€/MWh]	-1.98	-2.99

These are very asymmetric cases and this is the general rule: the down-regulation price has an absolute value several times greater than the up-regulation price.

We have taken one of the wind farms and simulated its participation in the Spanish market, offering the power prediction coming from prediction systems trained with MSE and with MCC. The following table summarizes the results obtained:

simulation	MSE	MCC
Contracted Energy [GWh]	50.40	35.38
Surplus [GWh]	16.32	22.45
Shortage [GWh]	17.64	8.74
Down-regulation costs [10^3 €]	137.59	195.47
Up-regulation costs [10^3 €]	52.91	26.68
Total Revenue	1752.46	1720.80

The conclusion is unmistakable: a strategy of offering power in excess to the real value gives a higher reward that a strategy that would have a more neutral attitude and higher frequency of errors close to zero. In other words: although the Entropy-trained system produces a prediction that is closer to the actual power, because of the asymmetry in prices for up and down errors, a system trained with MSE could give a revenue advantage to the wind power producer.

However, this gain is made at the loss of the ISO (admitting that it trusts the wind power prediction of the wind farm owner). Furthermore, because the rewarding strategy or the producer leads to a bias or systematic over-estimation of the wind power, this is contrary to the interests of the ISO in terms of security of the system.

Therefore, a wind power prediction system that provides the maximization of revenue rewards (or minimization of penalties) to wind power producers cannot be trusted or accepted by ISOs.

One can even realize that the excess reward paid to wind power producers will ultimately be passed to the consumers – so, regulators that have the mission to protect the interests of the consumers and the fairness of the market will have to consider the properties of the error distribution of commercial wind power prediction systems.

X. Conclusions

Wind power prediction for wind parks displays obvious non-Gaussian characteristics in the error probability distributions and therefore training input-output systems (mappers) to perform wind power prediction should move away from criteria based on Variance (such as MSE) and should instead adopt criteria based on Entropy concepts.

The application of these concepts to real wind parks in Portugal allows one to extract clear conclusions:

- Entropy training criteria lead to predictions with a higher frequency of errors closer to zero than the Minimum Square Error criterion
- Self-adaptive models produce better predictions than off-line trained models
- Off-line Entropy-based criteria are competitive with on-line MSE trained systems
- On-line self-adaptive training wit Entropy criteria lead to the best predictions, as evaluated by the error distributions
- Entropy training criteria and the MSE exhibit different biases on the average value of the predictions they produce
- In a market with asymmetric prices for up and down errors, a neutral prediction system does not provide the best revenue to a wind power producer: a bidding strategy that over-estimates wind power will be preferred
- The MSE criterion is not neutral
- In such a market, a forecasting system serving the purposes of wind power producers is against the interest of Independent System Operators

This paper contributes with evidence from real cases supporting the above conclusions. These will be taken in account during the development of the project, led by ANL – Argonne National Laboratory, in cooperation with INESC Porto, that aims at the implementation of new concepts in a wind power forecasting platform that may be taken as reference by all agents in the power business.

VI. References

[1] V. Miranda, C. Cerqueira, C. Monteiro, "Training a FIS with EPSO under an Entropy Criterion for Wind Power prediction", Proceedings of the *International Conference on Probabilistic Methods Applied to Power Systems*, PMAPS 2006, Stockholm, 11-15 June 2006.

[2] R. Bessa, V. Miranda and J. Gama, "Wind Power Forecasting With Entropy-Based Criteria Algorithms", Proceedings of the *International Conference on Probabilistic Methods Applied to Power Systems*, PMAPS 2008, Rincón, Puerto Rico, 25-29 May 2008.

[3] A. Renyi, "Some Fundamental Questions of Information Theory", *Selected Papers of Alfred Renyi*, vol 2, pp. 526-552, Akademia Kiado, Budapest, 1976.

[4] E. Parzen, "On the estimation of a probability density function an the mode", Annals Math Statistics, vol 33, 1962

[5] J. C. Principe and Dongxin Xu, "Introduction to information theoretic learning", *Proc. International Joint Conference on Neural Networks (IJCNN'99)*, Washington DC, USA, 10-16 July 1999, pp. 1783-1787.

[6] J. C. Principe and D. Xu "Information-theoretic learning using Renyi's quadratic entropy", *in J.-F. Cardoso, C. Jutten, and P. Loubaton, editors, Proceedings of the First International Workshop on Independent Component Analysis and Signal Separation*, Aussois, France, pages 407-412, 1999.

[7] D. Erdogmus and J. C. Principe, "Generalized Information Potential Criterion for Adaptive System Training", *IEEE Transactions on Neural Networks*, vol. 13, no. 5, September 2002, pp. 1035-1044.

[8] R.A. Morejon, J.C. Principe, “Advanced search algorithms for information-theoretic learning with kernel-based estimators”, *IEEE Transactions on Neural Networks*, vol. 15(4), pp. 874 - 884, July 2004

[9] Liu Weifeng, P.P. Pokharel, J.C. Principe, “Error Entropy, Correntropy and M-Estimation”, *16th IEEE Signal Processing Society Workshop on Machine learning for Signal Processing*, pp. 179–184, 2006

[10] W. Liu, P. Pokharel, J. Principe, “Correntropy: properties and applications in non-Gaussian signal processing”, *in IEEE Transactions on Signal Processing*, 2007.

[11] D. Erdogmus “Information Theoretic Learning: Renyi’s Entropy and Its Applications to Adaptive System Training”, *Ph.D. thesis*, University of Florida, 2002.

[12] H. Madsen, P. Pinson, G. Kariniotakis, Henrik Aa. Nielsen, Torben Skov Nielsen, “Standardizing the performance evaluation of short-term wind prediction models”, *Wind Engineering*, vol. 29, no.6, pp. 475-489, Dec. 2005.