



Probabilistic Electric Load Forecasting Model for the Uruguayan Interconnected Electrical System

E. Cornalino¹ and R. Chaer¹⁻²

¹ Administración del Mercado Eléctrico (ADME), Uruguay Phone/Fax number: +0598 2901 1630, e-mail: <u>Ecornalino@adme.com.uy</u>, <u>Rchaer@adme.com.uy</u>

² Instituto de Ingeniería Eléctrica, Facultad de Ingeniería, Universidad de la República, Uruguay Phone/Fax number: +0598 2711 0974, e-mail: <u>Rchaer@fing.edu.uy</u>

Abstract. The aim of this research is to improve the capacity to represent and forecast the electric demand for next week's scheduling. Currently the demand forecast used for this purpose is deterministic, which is not representative of reality, even if an ideal temperature forecast was available.

The current context of the Uruguayan electrical system has high probability of exportable surplus energy. For this reason, improvements to the procedure used to calculate systems supply costs and the quantity of exportable energy are welcome, in order to maximize the benefit we can get from resources.

The methodology applied is based on previous developments for simulation of stochastic variables within the SimSEE platform [2]. It combines daily step CEGH model [3] with a k-means clustering method [4].

Obtained results were satisfactory both from the point of view of the representation of the temporal behavior of the power demand, and from the point of view of the error obtained in the predictions. What is more, this improvements helps to reduce risks involved when making energy commitments with neighbouring countries.

Key words

Probabilistic load forecasting, Simulation of stochastic variables, Dispersion, Decision at risk.

1. Introduction

The aim of this research is to improve the capacity to represent and forecast the electric demand for next week's scheduling. The current context of the Uruguayan electrical system has high renewable energy penetration and high probability of exportable surplus energy. See Fig.1.

There are a lot of new aspects to consider for the calculation of the optimal operation system due the strong incorporation of renewable sources that, accompanied by the big existing share of hydroelectric, can bring scenarios like shown in Fig. 1 with relatively high frequency. As a result, taking into consideration the uncertainty on the calculation and simulation of the input variables becomes increasingly important [1].



Fig. 1 Composition by source of Uruguayan electricity supply for three days of last spring.

The more precise estimate of expected demand, together with the forecast of wind and solar energy generation, as well as the hydraulic contributions, the more certainty on the amount of energy exports availability determination. This certainty reduces the risk associated with the early negotiation of energy exports through interconnected countries, maximizing the benefits.

The main advantage of the model developed is the nondeterministic representation of the demand, allowing multiple simulations of the system supply to make decisions based on probabilistic results.

The expected results of the application of this model, represent a substantial enhancement for internal energy supply forecasting, to the extent that it allows to valorise the power demand and its supply costs on the basis of probabilities.

2. Methodology

The methodology to apply in order to build the model is based on previous developments for simulation of stochastic variables within the SimSEE platform [2].

The SimSEE platform is used to perform simulations of the electricity power system of a region or country. In the simulation, the power plants, the interconnections with other electrical systems and all system constraints and specifics are represented. That implies the representation of a reality full of details and uncertainties. Multiple simulations are performed to represent statistically the infinite possible realizations of the stochastic processes involved. The system's function cost is evaluated using Monte Carlo simulations and, consequently, the more number of draws used, the more confidence on the estimated value [3].

For the simulation of stochastic variables such as demand, generation based on renewable resources, the price of fossil fuels, etc., the platform includes a specific tool. This application serves to analyse time series data and identify a correlation model based on Gaussian Space with Histogram - CEGH [4] that can be used in the representation of a dynamic system. The identified model has the property of having a structure that enables its inclusion in the modelling of the system dynamics and generating synthetic series. These synthetic series preserve self-correlations and cross-correlations in a transformed space, and the amplitude histograms in real space.

A. Methodology for the daily behaviour identification

Daily behaviour identification is performed by a daily step CEGH model extracted from the maximum and minimum temperatures forecasted for the day itself and the previous day registered demand.

The performed model should take into account the seasonality and not only the temperature in isolation since the relationship between demand and temperature is not linear along the year. In any case it could be assumed linear in short intervals of temperatures. Based on this, a *variable filter* CEGH model with a 365 step was calculated. That is, a filter is identified for each day of the year, with reasonable buffer days around the corresponding day to smooth the filter, so the coefficients of the correlation matrix to be applied are changing continuously along the year.

The data series correspond to the period July 1, 2012 to April 30, 2017. Initially, on the series of hourly data, the increasing trend was estimated and removed in order to obtain an approximately stationary series, with energy values equivalent to July 2012.

To be more precise in the determination of covariations of the daily energy and temperature, the daily energy is separated into three time intervals.

TIME INTERVALS						
Valley	0 to 5					
Level	6 to 17					
Peak	18 to 23					

Table 1- Time intervals for daily energy demand

These three time intervals represent differentiated sections in terms of the shape of the hourly curve (see Fig. 2) and, in addition represent sections of the day where the human activities that have greater influence over the use of electric power are very different, therefore, its dependence on temperature is also different.



The weekly cycle of the data series was estimated and removed before identifying the model. It was observed that the average coefficients for each day of the week (and for every season) in relation to the average demand for work days (Monday to Friday) keep approximately constant along the year. This allows generating a model based on working days which is then affected by the corresponding week day coefficients.

B. Methodology for the hourly behaviour identification

To represent the hourly behaviour, the energy is distributed according to the expected form of the power demand curve, which is very stable, i.e., it does not have high frequency oscillations. Based on a clustering method [5] of historical power demand curves, the procedure involves the selection and interpolation from an index of historical curves using as input information the average temperature, type of day (work day, weekend, public holiday) and the day of the year (from 1 to 365). Four clusters were made for each curve section.

3. Results

The representation achieved by the proposed method has the advantage of allowing to obtain multiple simulations of load evolution, appropriate to the historical average temperature (and dispersion) for this day of the year, and according to the type of day (holiday or working day). What is more, in case of availability of temperature forecasts, it allows to reduce the variance of the result.

Regarding the hourly behaviour, the weekly cycle and the annual seasonality the results are also satisfactory. The mean daily power load behaviour during a complete

year, from 1000 simulations performed with the developed model, can be seen in Fig. 3.



4. Error Assessment

In order to quantify the errors, a period of data - not included in the original data set used to perform the model- was used. Concretely, the period used for this assessment was from December 12, 2016 until August 31, 2017.

For each day included in that period, 1000 simulations with a horizon of 14 days were performed within the SimSEE platform.

The data generating source of power demand, was configured - from the developed CEGH model - introducing forecasts of daily maximum and minimum temperatures.

The forecasts were introduced forcing the expected value of the simulations of temperatures to be equal to the forecast along the 13 steps of simulation.

The scattering around the mentioned expected value, was configured with a linear growth in 13 steps. It means, at the beginning, the forecasted value is deterministic, then the scattering width, spreads, linearly, step by step, reaching the historical scattering at the last step of simulation.

As a result the significance level regarding the forecast lowers with the increment of the time horizon.

Errors were averaged, obtaining a medium error and exceedance probability averaged¹ for each day of the forecast, from 1 to 14. The results in P.U. of the daily demand power value are portrayed in Fig. 4.



The accumulated error of the demand is not sufficiently damped (see Fig. 5) due to the fact that errors in the temperature forecast have some permanence. If at the beginning of a forecast a slight error is produced, for example, overestimating temperature, it is probable that this error maintains or amplifies with the time horizon of the forecast, as seen in Fig. 6.





Accumulated Error of Max. Temp. Forecast (°C)



5. Relative impact of model improvements

As mentioned before, the currently situation of the Electric Uruguayan System requires the handling of system operation based on probabilities, since the major part of the variables involved are aleatory. This implies that for improving the dispatch of the system and optimizing costs and benefits, it is required the consideration of all the uncertainties involved as accurate as possible.

Of all the variables involved, the one that stands the most variability for its amplitude and frequency is wind power. Also solar power presents an important short term dispersion. However, until nowadays there is no very important incidence due to the low installed capacity.

Currently, solar power capacity reaches 225 MW while wind power capacity is about 1400 MW. FiguresFig. 7 and Fig. 8 show the daily mean power for both energy sources simulated during a year, in expected value and with 10% and 90 % exceedance probability.

of the corresponding exceedance probabilities of each day included in the period.

¹ That is depicted in figures 4 to 6 are not the exceedance probabilities concerning the whole period but the averages



Expected value and 10% and 90 % exceedance probability



Fig. 8 Solar Power. Daily mean production during a year. Expected value and 10% and 90 % exceedance probability

In order to have a reference of the involved dispersion level, note as it is showed from Fig. 5, that the average daily wind generation, with a confidence level of 80%, has a dispersion in the order of 500 MW.

On the other hand, the hydroelectric power is managed by the system operator. According to this, you can appreciate in Fig. 9 that the hydro power production could be slightly accompanying the evolution of the power demand, in its weekly and also its annual cycle. See again Fig. 3 to compare.



Fig. 9 Hydro Power. Daily mean production during a year. Expected value and 10% and 90 % exceedance probability

In other aspects, the operation of the hydroelectric power plants is conditioned by the availability of hydric affluent to their respective lakes. This could result in a significant difference of cumulative generation for a year, like is shown in Fig. 10. There is depicted the dispersion of the accumulated generation based on 1000 simulations.

In summary, it can be observed that the main sources of uncertainty, due to the magnitude of its powers and their dispersions, are wind and hydraulic generation.

The power demand dispersion locates in an order near to 10% of its magnitude. This is equivalent to around 150 MW.



Fig. 10 Hydro Power. Accumulated production along a year. Expected value and 10% and 90 % exceedance probability

In fact, the demand uncertainty, is not the most important parameter at the time of cuantificating uncertainty and risk in the system dispatch decisions. However its input is not negligible and contributes to a global improvement in the modelling of the system, as seen below.

6. Risk reduction assessment

A comparison made between the two models based on the evaluation of relevant variables used to define the export offers, is described below.

For maximizing the benefits of energy exchanges with neighboring countries, the negotiation is starting to adopt the form of the offering of energy blocks for next week.

This energy blocks are calculated and offered at a certain price, being calculated in function of own costs and the neighboring pricings of their system in order to obtain the maximum profit. It is, without taking risks of increasing overmuch the cost of the own system and achieving an attractive offer for the neighboring system.

As main indicator for the calculation of the offer, the Marginal Exportation Cost (CME) is defined. It represents the cost of the last MW delivered to supply the exported energy, analogous to the Marginal Cost defined for the system. [6].

For the evaluation that is carried out in this research, a value of CME = 35 USD/MWh was taken as a reference to decide the amount of energy to be exported and 9 scenarios with different levels of energy availability were tested.

The decision is made based on the expected value of the CME. Considering that receipts from sales can be around 90 USD/MWh, there is an acceptable margin of coverage against the risk of higher costs.

In any case, once the decision of the amount of energy to be offered has been made based on these assumptions, it is important to observe the dispersion presented by the CME around its expected value.

The deterministic demand forecast model used to date does not correctly represent the dispersion that the demand can reach. Therefore reduces the expected dispersion of system costs.

The probabilistic model, by incorporating uncertainty around the expected value of demand, gives a more realistic view of high cost risks, allowing better quantification of risks.

Fig. 11 shows a comparison of the distribution width of the expected CME, according to results obtained with the deterministic model (DET) and the probabilistic model (CEGH).



Fig. 11 Compared dispersion width of CME distribution for DET model vs. CEGH model

As can be seen, for the case shown in Fig. 11, the CME with probability of exceedance of 5% varies by approximately 8 USD / MWh between the results of one model and the other. Although it seems an amount of lesser order, it is not negligible when making the decision of the price to bid. It is beacuse the greater certainty regarding to the expected profit margin and to avoid unforeseen risks.

The 9 scenarios were constructed by selecting for each variable -hydric affluent and wind and solar generationthree chronicles whose cumulative total was at 25%, 50% and 75% of the probability distribution to define the low, med and high levels respectively.

The DDH-CME variable that represents the difference in the dispersion of the CME histograms between one model and the other, was calculated. That is, the difference between the values with 5% and 95% exceedance probability was calculated first as an indicator of the dispersion. Then the difference was made between the dispersion obtained with the DET model and that obtained with the CEGH model.

The results obtained for each scenario are shown in Table 2.

	Hydraulic									
	Low Wind + Solar			Med			High			
				Wind + Solar			Wind + Solar			
	Low	Med	High	Low	Med	High	Low	Med	High	
Energy Exports (GWh)	0	0	0	9	29	48	28	48	48	
DDH-CME (USD/MWh)	8	6	7	9	9	11	27	14	6	
CME MEDIO (USD/MWh)	126	97	73	35	35	35	35	16	6	
DDH-CME (%)	7%	6%	10%	24%	25%	31%	77%	90%	97%	

Table 2 Results of the difference among both models of the dispersion indicator (interval between exceedance probability 5% and 95%).

In the last row, the values of the dispersion difference are presented as a percentage of the average CME value. In the most relevant cases, difference of between 25% and 77% of the average CME can be observed.

These values represent an appreciable improvement in the quantification of risk when defining sales prices with the desired level of confidence.

To appreciate it more clearly, the comparison of CME values with 5% probability of exceedance is presented in Fig. 12 for each of the 9 scenarios evaluated.

The cases in the graph of Fig. 12 are arranged as in Table 2. The notation LH_HW indicates low hydro with high wind and so on.

There it can be seen that, from the point of view of the risk of high export costs, the most critical case is HH_LW. In this case the CME with 5% exceedance probability is 70% higher than the expected value if evaluated with the model DET and 120% higher if evaluated with CEGH model.



Fig. 12 Compared CME costs with 95% of confidence for DET model vs. CEGH model

4. Conclusion

The application of the developed model will produce a substantial improvement to the procedure used to forecast the internal energy supply and the exportable energy to the extent that it allows decision making on the basis of probabilities.

Obtained results were satisfactory both from the point of view of the representation of the daily and annual behavior of the power demand, and from the point of view of the error obtained in the predictions.

The error with 5% of exceedance probability is reasonably increasing as the forecast horizon progresses. It goes from approximately 4% and reaches 11% on day 14 of forecast. The accumulated error remains around 4% for up to two weeks of accumulation, for the same level of confidence.

Due to the dispersion of the other variables involved, is greater than the dispersion of demand, the impact on the dispersion of global results is not very noticeable. In any case, it is not negligible and, in the face of the negotiation of exports, especially the definition of the prices to be offered, the risk reduction is significant since the margins in which the dispersion of the expected costs increases for a level of 90% confidence - they hover above values of 25% of the average value of the CME.

As a general conclusion, beyond the model's evaluation results presented in this paper, it is noted the need to go forward with evaluation, calibration and improvement of the developed model, in all its aspects. In particular, it is necessary to work on the definition of parameters and control mechanisms that allow keeping the model's correlation functions, calibrated to maintain or increase its accuracy in the forecast.

References

[1] T. Hong, S. Fan. "Probabilistic electric load forecasting A tutorial review", International Journal of Forecasting 32 (2016) 914-938

[2] R. Chaer et al. Memoria Final Proyecto ANII-FSE2009-18 pág. 45 / 176

[3] W.R. Gilks, S. Richardson, D. Spiegelhalter. "Markov Chain Monte Carlo in Practice". Chapman & Hall. 1996.

[4] R Chaer, "Fundamentos de modelo CEGH de procesos estocásticos multivariable". Technical Report, IIE-Fing, Udelar -2011, Montevideo.

[5] A.K. Jain, "Data clustering 50 years beyond K-means", in Pattern Recognition Letters 31 (2010) 651–666.

[6] Steven Stoft. "Power System Economics: Designing Markets for Electricity" IEEE Press & WILEY-INTERSCIENCE A JOHN WILEY & SONS, INC., PUBLICATION. 2002.