An algorithm for the mid-term forecast and scenario generation of natural gas and fuel oil prices

Emilio Gómez-Villalva, *Member, IEEE,* and Andrés Ramos

***Abstract*—The liberalization of energy markets and, as a consequence, the openness of retail energy markets, has allowed consumers to sign new types of energy contracts which are subject to uncertainty. To evaluate these contracts, consumers need easily implementable and reliable tools for the forecast of energy prices. With this purpose in mind, we propose an original method for the forecast and scenario generation of natural gas and fuel oil monthly average prices with an annual scope in Spain. The algorithm is based on the strong linear correlations existing between crude oil Brent spot prices and the energy prices to be estimated. Thus, the algorithm first generates future Brent spot price scenarios by sampling from probability distributions constructed with historical data of Brent futures and spot prices. An example of the capabilities of the algorithm is presented for the natural gas and fuel oil price forecast of 2003.**

***Index Terms*—Energy price forecast, scenario generation, Monte Carlo simulation.**

I. Introduction

In the framework of the current liberalized energy markets, an industrial consumer who is responsible for energy contracting decisions is obviously greatly affected by energy price uncertainty. Thus, he or she must consider this factor in the decision-making process [1].

In this paper, we present industrial consumers with a model for representing the natural gas and fuel oil price uncertainty. This uncertainty is modeled by a scenario tree to be used as input data in a stochastic optimization model for supporting decisions related to optimal energy management.

Stochastic optimization is a very powerful technique for making decisions since it allows for the representation of uncertainty associated with the parameters of the problem. Despite the broad scope of stochastic optimization models, there has so far been no general methodology established for the generation of scenario trees. The method to be used depends greatly on factors such as the optimization model into which the tree is introduced and the availability of input data. A summary of different techniques of scenario generation depending on the available information can be found in [2].

In our model, the tree generated by the algorithm we propose is used to feed a two-stage stochastic mixed-integer linear optimization problem. Since the first-stage variables (decision variables) are unique for the whole time frame, the model generates independent time series of fuel prices.

Ideally, the technique to be used for scenario generation should obtain a unique solution of the decision variables of the stochastic optimization problem with any possible scenario tree generated. If several solutions of the stochastic optimization problem are obtained, it would be convenient to increase the number of scenarios to be generated despite increasing the execution time. Nevertheless, the intrinsic stability of the solution of the stochastic optimization problem is independent from the hypotheses posed for the generation of the scenario tree.

In general, scenario generation is performed by either advanced stochastic models (e.g. [3]) or time series models (e.g. [4], [5]) which are used for sampling scenarios.

More sophisticated is the method proposed by Høylan *et al.* [6] for scenario generation. In order to obtain a tree with certain statistical properties, they use non-linear programming to minimize the square error between the moments of the marginal distributions of the tree to generate and those of the distributions entered as data. The resulting non-linear optimization problem is hard to solve since it may contain several local minima.

In this model, all of the scenarios are generated simultaneously, which makes the method slow if the number of random variables is high. To overcome this obstacle, the same authors propose a model [7] in which the tree is generated decomposing the problem and dealing with each marginal probability distribution separately.

Another method for scenario generation consists of performing sampling from historical data distributions [8]. This is the easiest method and can especially be applied when the random variables of interest have the same behavior in the past as well as in the future. Following this technique, Takriti *et al.* [9] obtain an electricity demand tree using observations from the past of power plant failures and deviation in demand forecast under circumstances similar to those expected in the future.

The mid-term forecast of natural gas and fuel oil prices for industrial consumers is not an easy task. These prices depend on hard-to-predict macroeconomic variables which determine crude oil prices as well as on transportation and distribution costs. In addition, these prices are influenced by political decisions, wars, development of alternative energy sources and other unpredictable factors.

In this paper, we present an original model which takes advantage of the high correlations between the prices to be estimated and crude oil Brent prices.

II. Previous considerations & hypotheses

In Spain, natural gas prices for consumers are indexed to Brent spot prices while fuel oil prices are indexed to different fuel oil market prices. Nevertheless, since fuel oil

E. Gómez-Villalva is with Gómez-Villalva Ingenieros, Granada, Spain (e-mail: emlgvg@eresmas.net).

A. Ramos is with the Instituto de Investigación Tecnológica, Escuela Técnica Superior de Ingeniería (ICAI), Universidad Pontificia Comillas, Madrid, Spain (e-mail: andres.ramos@iit.upco.es).

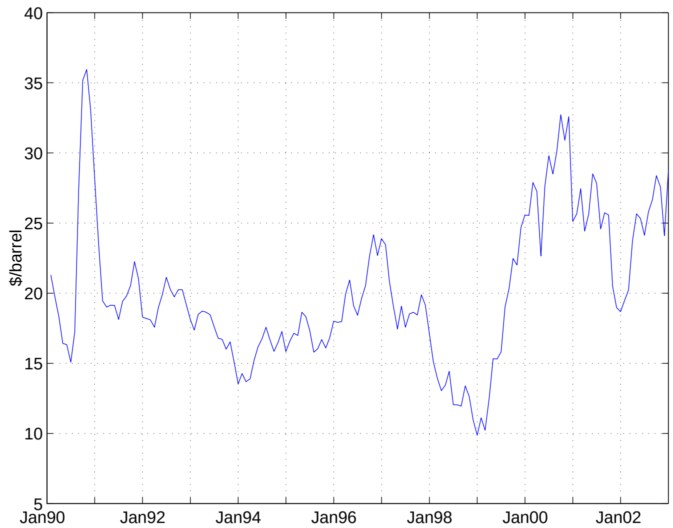

Fig. 1. Monthly average Brent spot prices 1990-2002.

market prices and Brent prices are also highly correlated, both fuels are estimated by the Brent spot forecast.

In this section, we explain the basics of the Brent spot price forecast and the characteristics of the technique used for scenario generation.

A. *Brent spot price forecast*

The algorithm estimates for each scenario g monthly (periods $k = 1,..,12$) natural gas $p_g^g(k)$ and fuel oil $p_f^g(k)$ prices with annual scope.

Historical prices of natural gas $p_g(i)$ and fuel oil $p_f(i)$ (with i the period index of historical data) are highly linearly correlated with Brent spot prices $s(i)$, and therefore the forecast of the prices $p_g^g(k)$ and $p_f^g(k)$ is done via the estimation of monthly average Brent spot prices $s^g(k)$. Specifically, the correlation coefficients between $p_f(i)$ and $s(i-1)$, on the one hand, and between $p_g(i)$ and $\sum_{r=i-7}^{i-1} s(r)$, on the other hand, are high (above 0.95). In this way, estimating only Brent spot prices $s^g(k)$, the prices $p_g^g(k)$ and $p_f^g(k)$ and the correlation between them are calculated jointly.

To forecast Brent spot prices, in a first approach we represented Brent historical data $s(i)$ (figure 1) by means of an ARIMA [1] univariate time series model according to the Box-Jenkins methodology [10]. The results were not satisfactory. Alternatively, and using solely the Brent spot time series, a GARCH [2] model could be a possible methodology to estimate Brent spot prices [10], [11].

Given the difficulty in forecasting Brent spot prices, it is advisable to take into account additional information to the Brent spot time series. In this sense, historical Brent futures prices contain information that can be used for the Brent spot price forecast $s^g(k)$. Pilipović [12] links both prices defining the equation:

$$f_r(i) = e^{-\alpha(i-r)} E_r[s(i)] \tag{1}$$

[1] AutoRegressive Integrated Moving Average.
[2] General Autoregressive Conditional Heteroscedasticity.

where the parameter α includes terms such as the risk-free rate and volatility, the value $f_r(i)$ represents the Brent futures price observed in period r with the time of expiration in period i, and the value $E_r[s(i)]$ represents the expected value of the Brent spot price $s(i)$ in period i observed from period r.

We tried to determine a function θ (not necessarily exponential) to relate historical prices $f_{i-k}(i)$ and $s(i)$ for each period k between 1 and 12 in order to calculate Brent spot prices with the equation:

$$s^g(k) = \theta[f_0(k)] + \varepsilon^g(k) \tag{2}$$

In this equation, the values $f_0(k)$ are Brent futures prices observed in the month (period 0) previous to the forecast time frame with expiration in each period of the forecast scope and therefore are known prices, and the values $\varepsilon(k)$ are the error distributions derived from the determination of θ. It was not possible to apply this method since the prices $f_{i-k}(i)$ and $s(i)$ are not correlated.

Due to the lack of success forecasting mid-term Brent spot prices, we propose an original methodology based on the following hypotheses:

- Brent futures prices provide information about Brent spot prices $s^g(k)$ that is useful for the forecast.
- The relation between Brent futures and spot prices in the past is considered to be the same as in the future.

These hypotheses are consistent with the use of the algorithm by industrial consumers since the information needed as input data is easily available in public sources.

B. *Scenario tree*

The model presented in this paper has been conceived to represent price uncertainty as input for a stochastic optimization model with a quite large deterministic version [13] and thus it is advisable to have a reduced number of scenarios. Alternatively, it is possible to construct a tree of any size and apply scenario reduction techniques [14], [15]. Specifically, each node of the tree that the algorithm generates belongs to only one scenario, with all the scenarios having the same probability of occurrence.

To generate the price probability distributions to introduce in the scenario tree, it is necessary to identify the sources of uncertainty in the method stated. These are the following:

- The error derived from the linear regressions between fuel oil prices $p_f(i)$ and Brent spot prices $s(i-1)$ on the one hand, and between natural gas prices $p_g(i)$ and Brent spot prices $\sum_{r=i-6}^{i-1} s(r)$ on the other hand.
- The error derived from the Brent spot price forecast.

The latter error is significantly greater than the former and therefore the scenario generation is focused on representing the uncertainty derived from the Brent spot price forecast. The error in the linear correlations between Brent prices and both fuels is neglected.

III. ALGORITHM DESCRIPTION

The Brent spot price forecast is based on the hypothesis that the relative error distributions of Brent spot and

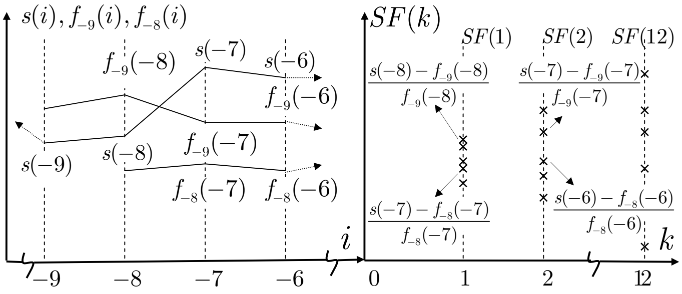

Fig. 2. Generation of the random variables $SF(k)$ from historical data.

futures prices is the same in the last few years as in the year of the forecast. Since Brent futures prices are available for the forecast time frame, spot prices are calculated from futures prices.

For each period, the moments of the distribution of the generated samples are compared to those of *the error distributions between historical Brent spot* $s(i)$ *and futures* $f_{i-k}(i)$ *prices* so as to validate the values obtained.

The variation of Brent spot prices in consecutive periods should be reflected in the algorithm since, for instance, high prices in one period will lead to a higher probability of high prices in the following period than of low ones. In order to consider *the price relation in consecutive periods*, the model uses linear regressions obtained from relative errors between historical spot and futures prices in consecutive periods.

The *determination of the error distribution of the first period* is done in a different way from those of the other periods k. To generate the error distribution of the first period, the Brent spot price in the month preceding the forecast time frame is taken into account.

Finally, once the Brent spot price scenarios $s^g(k)$ are obtained, the *determination of natural gas* $p_g^g(k)$ *and fuel oil* $p_f^g(k)$ *prices* is done via the linear regressions that link Brent spot prices with natural gas and fuel oil prices.

Each of the items in italics mentioned above is explained in detail below.

A. Price distribution in each period

The relation between Brent spot and futures prices is defined for each period k by the random variable $SF(k)$, whose sample space S is composed of the following historical data:

$$S[SF(k)] = \left\{ \frac{s(i) - f_{i-k}(i)}{f_{i-k}(i)} \right\} \tag{3}$$

Figure 2 shows an example of the generation of two values of each of the distributions $SF(1)$ and $SF(2)$ from historical prices. The graphic on the left represents historical values in periods $i = \{-9, -8, -7, -6\}$. These periods are numbered in decreasing order starting with the period farther away in time. The last period preceding the forecast time frame is period 0. Historical Brent prices represented are spot prices $s(i)$, futures prices $f_{-9}(i)$ in period -9 for the following three periods ($f_{-9}(-8)$, $f_{-9}(-7)$, $f_{-9}(-6)$) and futures prices $f_{-8}(i)$ in period -8 for the following two periods ($f_{-8}(-7)$, $f_{-8}(-6)$).

In any period i, for example in period -7, in addition to the prices $s(-7)$, $f_{-9}(-7)$ and $f_{-8}(-7)$ depicted, there are historical data of futures prices in previous periods to period -9 with the time of expiration in period -7. In this way, the value $f_{-10}(-7)$ is the futures price in period -10 for period -7, the value $f_{-11}(-7)$ is the futures price in period -11 for period -7, and so on until we reach the futures price twelve periods preceding period -7, which is the value $f_{-19}(-7)$. All these values are not depicted so as to ease the comprehension of the figure.

With these historical data represented in figure 2, it is possible to determine the values $\frac{s(-8)-f_{-9}(-8)}{f_{-9}(-8)}$ and $\frac{s(-7)-f_{-8}(-7)}{f_{-8}(-7)}$ (right side of the same figure) of the random variable $SF(1)$. This variable is a measure of the difference between futures prices in period k expiring in period $k+1$ and spot prices in period $k+1$. Similarly, $\frac{s(-6)-f_{-8}(-6)}{f_{-8}(-6)}$ and $\frac{s(-7)-f_{-9}(-7)}{f_{-9}(-7)}$ (also in figure 2), belong to the random variable $SF(2)$.

Using the random variables $SF(k)$ for the Brent spot price forecast has the following advantages:

- The model takes advantage of the information provided by Brent futures prices in relation to Brent spot prices.
- Since the samples from the random variables $SF(k)$ are normalized, there is no need to apply any type of discount rate to the current value of money.
- In order to estimate the parameters of the model, no restrictions are imposed concerning the stabilization of mean and variance (like in the case of ARIMA models), which is something desirable given the characteristics of the Brent spot time series.

As explained in the following sections, the model generates the samples $sf^g(k)$. Given that

$$sf^g(k) = \frac{s^g(k) - f_0(k)}{f_0(k)} \tag{4}$$

and the futures prices $f_0(k)$ in period 0 for each of the forecast time frame periods are known, Brent spot prices $s^g(k)$ can easily be calculated.

To guarantee that the samples generated $sf^g(k)$ have the same statistical properties as the distributions $SF(k)$ obtained from historical data, a *deviation measure* function is formulated to compare the moments of both distributions. This function is

$$\sum_{z \in Z} \phi_z \left| \frac{m_z(k) - m'_z(k)}{m_z(k)} \right| \tag{5}$$

where $m_z(k)$ is the moment of order z of the random variable $SF(k)$ in period k, $m'_z(k)$ is the moment of order z of the samples generated $sf^g(k)$ in period k and ϕ_z is the weight assigned to each moment. The resulting function is adimensional, which allows the comparison of moments of different order.

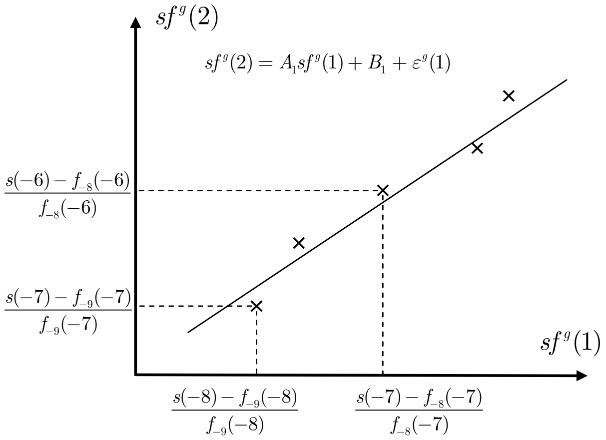

Fig. 3. Price relation between periods 1 and 2.

The number of moments z of the distributions $SF(k)$ which have to be reflected in the scenario tree depends on how the risk is measured in the stochastic optimization model. For instance, with Markowitz mean-variance models, only the first two moments affect the result of the problem. The algorithm stated in this work calculates 4 moments: mean, variance, skewness and kurtosis. Nevertheless, by means of the weights ϕ_z, it is possible to assign the moments different importance in the resulting price distributions. In order to validate the samples generated $sf^g(k)$, the deviation measure function should be below a certain threshold ζ.

B. Price relation in consecutive periods

The observations of the random variables $SF(k)$ represent Brent futures and spot relative differences in each of the 12 months of the forecast time frame. These distributions do not take into account the price relation in consecutive periods, a factor necessary to determine the samples $sf^g(k)$.

In order to consider the relative error dependency in consecutive periods (relation between the random variables $SF(k)$ and $SF(k+1)$), the following 11 straight lines ($k = 2,..,12$) are constructed:

$$sf^g(k) = A_{k-1} sf^g(k-1) + B_{k-1} + \varepsilon^g(k-1) \qquad (6)$$

Each of these lines is obtained through the regression analysis with the following pair of historical differences: $\left(\frac{s(i)-f_{i-k}(i)}{f_{i-k}(i)}, \frac{s(i+1)-f_{i-k}(i+1)}{f_{i-k}(i+1)}\right)$.

Brent futures prices used in each pair correspond to prices in the same period with expiration time in consecutive periods whereas Brent spot prices correspond to those in one month and the following one. Going back to the example in the previous section with the historical prices represented in figure 2, the pairs of values $\left(\frac{s(-8)-f_{-9}(-8)}{f_{-9}(-8)}, \frac{s(-7)-f_{-9}(-7)}{f_{-9}(-7)}\right)$ and $\left(\frac{s(-7)-f_{-8}(-7)}{f_{-8}(-7)}, \frac{s(-6)-f_{-8}(-6)}{f_{-8}(-6)}\right)$ are used to determine the linear relation between samples of the first period $sf^g(1)$ and those of the second period $sf^g(2)$ (see figure 3 as well). The linear regression between the first two periods has the expression $sf^g(2) = A_1 sf^g(1) + B_1 + \varepsilon^g(1)$.

By means of the mentioned linear regressions, the model reflects the relation in consecutive months of Brent spot and futures prices in period 0 with the time of expiration in each of the periods of the forecast time frame. For instance, if in period $k-1$ the sample $sf^g(k-1)$ has a high value, the spot price $s^g(k-1)$ will be clearly higher than the futures price $f_0(k-1)$ (see equation 4) and thus the most likely thing to happen in the following period is that the value $sf^g(k)$ will not change its sign and therefore the spot price $s^g(k)$ will also be higher than the futures price $f_0(k)$. On the contrary, if the sample $sf^g(k-1)$ is close to zero, meaning similar values of the spot $s^g(k-1)$ and futures $f_0(k-1)$ prices, the probability of the sample $sf^g(k)$ having a different sign from that of $sf^g(k-1)$ is high, and the spot price $s^g(k)$ will be higher or lower than the futures price $f_0(k)$ but probably of similar value.

The adjustment error of the linear regressions is represented by the empirical distributions of residues $\varepsilon(k-1)$. Back again to figure 3, $\frac{s(-6)-f_{-8}(-6)}{f_{-8}(-6)} - (A_1 \frac{s(-7)-f_{-8}(-7)}{f_{-8}(-7)} + B_1)$ and $\frac{s(-7)-f_{-9}(-7)}{f_{-9}(-7)} - (A_1 \frac{s(-8)-f_{-9}(-8)}{f_{-9}(-8)} + B_1)$ are values of the random variable $\varepsilon(1)$.

The generation of error samples is performed by Monte Carlo simulations using the inverse transformation technique, in the same way as the sampling from the distribution $SF(1)$ for the determination of the values of the first period (next section). This method consists of generating a value u from the uniform distribution $[0, 1)$ and finding the corresponding value from the error distribution function $F[\varepsilon(k-1)]$ so as to obtain an error value $\varepsilon^g(k-1)$ equal to $F^{-1}[u]$. For each value $sf^g(k-1)$, only one sample of the random variable $\varepsilon(k-1)$ is obtained since, as said in section II. B, each node belongs to only one scenario.

C. Determination of the relative error distribution of the first period, $sf^g(1)$

The method described so far is used to obtain the distribution of spot and futures prices in each period $SF(k)$ (section A) as well as to obtain the relation that these prices have to fulfill in consecutive periods, relation between the distributions $SF(k)$ and $SF(k+1)$ (section B).

The price scenarios determined under these hypotheses do not consider in a specific manner spot prices in the last few periods before period 1, the first of the forecast scope. The historical price distributions $SF(k)$ are formed independently of the period in which those prices took place.

In order to consider recent past prices, the Brent spot price $s(0)$ in the period preceding the first one of the forecast time frame is employed. This price is used for the generation of the samples of the first period $sf^g(1)$, as described in this section, whereas the samples $sf^g(k)$ of the other periods are generated according to the criteria explained in sections A and B.

To determine the relative error $sf^g(1)$ between spot and futures prices in period 1, the algorithm first samples the values

$$sf^g(0) = \frac{s^g(0) - f_{-1}(0)}{f_{-1}(0)} \quad [3] \tag{7}$$

from the distribution function $F[SF(1)]$. In this distribution, the value $\frac{s(0)-f_{-1}(0)}{f_{-1}(0)}$ is introduced replacing the closest sample of $sf^g(0)$. This introduced value contains the known Brent spot price $s(0)$ and is used as the root for the scenario generation.

The samples $sf^g(0)$ (including the value $\frac{s(0)-f_{-1}(0)}{f_{-1}(0)}$) belong to period 0. To refer them to period 1 (i.e. $sf^g(1) = \frac{s^g(1)-f_0(1)}{f_0(1)}$), two transformations are needed:

1. The first one obtains the samples

$$sf'^g(1) = \frac{s^g(1) - f_{-1}(1)}{f_{-1}(1)} \tag{8}$$

   from $sf^g(0)$ via the linear regression that links the random variables $SF(1)$ and $SF(2)$ (equation (6) with $k = 2$). These values belonging to the distribution $SF(2)$ correspond to spot prices in period 1 and futures prices in period -1 for period 1.
2. To relate these samples generated $sf'^g(1)$ to the desired $sf^g(1)$, a linear regression analysis is done with the pairs $\left(\frac{s(i)-f_{i-2}(i)}{f_{i-2}(i)}, \frac{s(i)-f_{i-1}(i)}{f_{i-1}(i)}\right)$. Again with the example of the historical data depicted in figure 2, one of these pairs of values is $\left(\frac{s(-7)-f_{-9}(-7)}{f_{-9}(-7)}, \frac{s(-7)-f_{-8}(-7)}{f_{-8}(-7)}\right)$. These observations relate relative differences between spot and futures prices in period -7, with the futures prices being from 1 and 2 preceding periods (periods -8 and -9 respectively).

The adjustment error in the former regression is reflected in the distribution of residues ε'. For each sample $sf'^g(1)$, G values are generated from the distribution function $F[\varepsilon']$, with G the number of scenarios to be obtained. Therefore, the size of the samples $sf^j(1)$ is $G \cdot G$, and those starting in period 0 from the value $\frac{s(0)-f_{-1}(0)}{f_{-1}(0)}$ are the ones corresponding to the samples $sf^g(1)$. The algorithm generates $G \cdot G$ scenarios since in each period the samples must behave as the random variables $SF(k)$ do. Of all these $G \cdot G$ scenarios, only those G which consider the known spot price $s(0)$ constitute the output of Brent spot prices.

This process as well as the generation of samples in period 2 is depicted in figure 4. Circles from period 1 on, represent the samples $sf^g(k)$ obtained from the known value $\frac{s(0)-f_{-1}(0)}{f_{-1}(0)}$. Samples generated in each period $sf^j(k)$ are those depicted as both circles and crosses. The linear regressions which relate values of consecutive periods are also represented in the figure. The dummy period $1'$ contains the samples $sf'^g(1)$ which link values of periods 0 and 1 as explained in the previous paragraphs.

D. Determination of natural gas and fuel oil scenarios

Once the samples $sf^g(k)$ are generated, Brent spot price estimation is immediate through the equation:

[3] Index -1 corresponds to 2 periods prior to the first period of the forecast time frame.

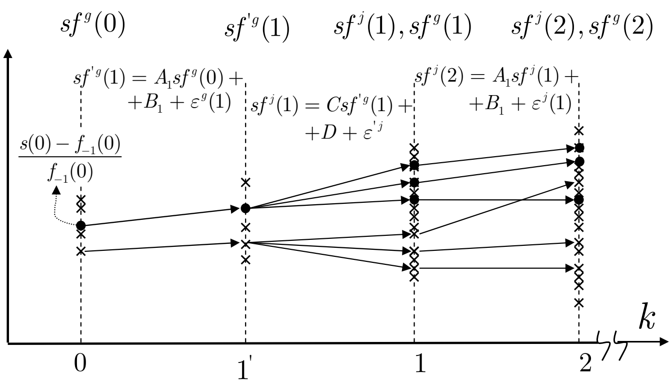

Fig. 4. Determination of samples of periods 1 $sf^g(1)$ and 2 $sf^g(2)$.

$$s^g(k) = f_0(k)(1 + sf^g(k)) \tag{4'}$$

With these prices, natural gas $p^g_g(k)$ and fuel oil $p^g_f(k)$ prices are calculated as follows:

$$p^g_g(k) = E \sum_{r=k-6}^{k-1} \lambda(r) s^g(r) + F \tag{9}$$

and

$$p^g_f(k) = H\lambda(k-1)s^g(k-1) + L \tag{10}$$

with $\lambda(\cdot)$ the €/\$ exchange rate and E, F, H and L the parameters of the linear regressions.

IV. Sequential formulation

To facilitate the comprehension of the algorithm, its sequential formulation is presented below.

1. Generation of the samples $sf^j(1)$ of the first period of the forecast scope.
   (a) Generation of the samples $sf^g(0)$ from the distribution function $F[SF(1)]$ with

$$sf^g(0) = \frac{s^g(0) - f_{-1}(0)}{f_{-1}(0)} \tag{7}$$

   These values from the random variable $SF(1)$ are Brent spot prices in period 0 (the one preceding the first period of the forecast scope) and Brent futures prices for period 0 observed from the preceding period -1.
   (b) The observation $\frac{s(0)-f_{-1}(0)}{f_{-1}(0)}$ is replaced by the sample of the distribution $sf^g(0)$ with the closest value. In this way, the historical observation from which the future scenarios are generated is introduced in the distribution obtained in (a).
   (c) Checking that the generated samples $sf^g(0)$ have the same statistical properties as the distribution $SF(1)$:

$$\sum_{z \in Z} \phi_z \left| \frac{m_z(1) - m'_z(0)}{m_z(1)} \right| \le \zeta \tag{11}$$

   If this inequality is not satisfied, the algorithm goes back to (a).

(d) From each sample $sf^g(0)$ one sample in period $1'$ is obtained:

$$sf'^g(1) = A_1 sf^g(0) + B_1 + \varepsilon^g(1) \quad ((6) \text{ with } k{=}2)$$

These samples $sf'^g(1)$ belong to the random variable $SF(2)$ with

$$sf'^g(1) = \frac{s^g(1) - f_{-1}(1)}{f_{-1}(1)} \tag{8}$$

(e) Futures prices in $sf'^g(1)$ are from period -1. To refer them to period 0, the most recent futures prices before the forecast time frame, the linear regression

$$sf^j(1) = Csf'^g(1) + D + \varepsilon'^j \tag{12}$$

is used.

(f) Checking that the generated samples $sf^j(1)$ have the same statistical properties as the distribution $SF(1)$:

$$\sum_{z \in Z} \phi_z \left| \frac{m_z(1) - m'_z(1)}{m_z(1)} \right| \leq \zeta \tag{13}$$

If this inequality is not satisfied, the algorithm goes back to (d).

2. Generation of the samples $sf^j(k)$ of the remaining periods of the forecast scope.

(a) From $k = 2$ to 12:

(b) Determination of the samples $sf^j(k)$ for each period via the linear regressions linking consecutive periods:

$$sf^j(k) = A_{k-1} sf^j(k-1) + B_{k-1} + \varepsilon^j(k-1) \tag{6}$$

(c) Checking that the generated samples $sf^j(k)$ have the same statistical properties as the distribution $SF(k)$:

$$\sum_{z \in Z} \phi_z \left| \frac{m_z(k) - m'_z(k)}{m_z(k)} \right| \leq \zeta \tag{14}$$

If this inequality is satisfied, the algorithm goes back to (a) while $k < 12$, otherwise it goes to (b).

3. Calculation of Brent spot prices $s^g(k)$.

(a) Among all the samples $sf^j(k)$ generated, $sf^g(k)$ contain the Brent spot prices $s^g(k)$ coming from $\frac{s(0)-f_{-1}(0)}{f_{-1}(0)}$ (see item 1.(b)):

$$s^g(k) = f_0(k)(1 + sf^g(k)) \tag{4'}$$

4. Calculation of natural gas $p_g^g(k)$ and fuel oil $p_f^g(k)$ prices.

(a) Natural gas and fuel oil prices are obtained from Brent spot prices with the equations:

$$p_g^g(k) = E \sum_{r=k-6}^{k-1} \lambda(r) s^g(r) + F \tag{9}$$

and

$$p_f^g(k) = H\lambda(k-1)s^g(k-1) + L \tag{10}$$

respectively.

V. Case example

The algorithm stated in this paper has been implemented in the programming environment MATLAB 6.5. To demonstrate its performance, we present an example in which the scenario tree of natural gas and fuel oil monthly average prices are obtained for 2003. Period 0, the one preceding the first month of the forecast time frame, is December 2002.

The liberalization of the natural gas market is now taking place in Spain. At this point, consumers can choose between negotiating their contracts with retailers and signing contracts based on tariffs established by the Government. Due to this recent market openness, prices of the contracts in the new framework are indexed to tariffs. These latter prices represent a reference for the prices of the contracts under the open market and thus are the natural gas prices we use as historical data.

Fuel oil prices for large consumers have been liberalized now for a few years and thus we use historical fuel oil prices of industrial consumers in the open market as input data for the model.

The data we have used as input for the forecast comes from the years 1999-2003. This data contains enough information to calculate the linear regression between Brent spot and natural gas prices as well as the linear regression between Brent spot and fuel oil prices. The correlation coefficients obtained in these regressions are high (see Table II). In the data from these years, 1999 to 2003, the increase in mean and variance of Brent spot prices which has occurred in the last few years can be observed (figure 1). In addition, this period consisting of four years does not constitute a high number of years, since it would be difficult to reflect future situations in years farther away in the past.

In the proposed example, ten ($G = 10$) scenarios are generated. The other configurable parameters corresponding to the *deviation measure* between the distributions of historical data $SF(k)$ and the values generated are: $\phi_1 = \phi_2 = 1$, $\phi_3 = \phi_4 = 0.2$ and $\zeta = 0.15$. The mean and variance have been weighted more heavily than the other moments since they are the most important of the distributions. The value for ζ has been chosen as a tradeoff between time execution and magnitude of the *deviation measure*. Tables I and II depict the correlation coefficients ρ of the regressions used in the algorithm.

TABLE I
Correlation coefficients of samples of relative errors between Brent spot and futures prices in consecutive periods, $sf^g(k-1)$ and $sf^g(k)$

k	2	3	4	5	6	7
ρ_{k-1}	0.8315	0.8579	0.8940	0.9113	0.9359	0.9505

k	8	9	10	11	12
ρ_{k-1}	0.9516	0.9553	0.9574	0.9479	0.9436

The price scenarios of Brent spot, natural gas and fuel

TABLE II
OTHER CORRELATION COEFFICIENTS

Determination of period 1	Brent spot Natural gas	Brent spot Fuel oil
$(sf'^{g}(1), sf^{j}(1))$	$(s^{g}(k), p_{g}^{g}(k))$	$(s^{g}(k), p_{f}^{g}(k))$
0.7706	0.9897	0.9692

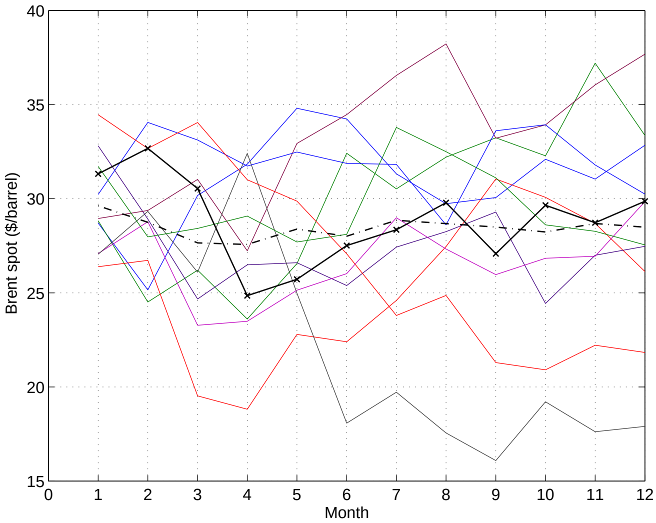

Fig. 5. Brent spot price scenarios for 2003.

oil obtained for the 2003 forecast are depicted in figures 5, 6 and 7, respectively. The dash-dot lines in these figures represent average values of the scenarios generated whereas the solid lines with crosses are the real prices which occurred in 2003. Changes in natural gas prices between consecutive periods are smoother than those of fuel oil prices since the former are estimated through the correlation with average Brent spot prices over 6 months.

The terms $\phi_z\left|\frac{m_z(k)-m_z'(k)}{m_z(k)}\right|$ resulting from the *deviation measure* for each forecast period are stated in table III. This table also shows the sum of these terms, whose values should be below ζ to verify that the samples obtained $sf^j(k)$ are distributed according to the random variables $SF(k)$.

VI. Summary

The model presented in this paper aims at obtaining a scenario tree of monthly average natural gas and fuel oil prices as a support for industrial consumers in making their contracting decisions in liberalized energy markets. Although the model can be easily adapted to fulfill the requirement of other countries, it has been developed according to the Spanish peculiarities.

The algorithm is conceived to generate a reduced number of scenarios since this is used as input data for a large two-stage stochastic optimization model.

First of all, the algorithm performs the Brent price forecast and scenario generation. For this purpose, we propose an original method based on the use of the relation between

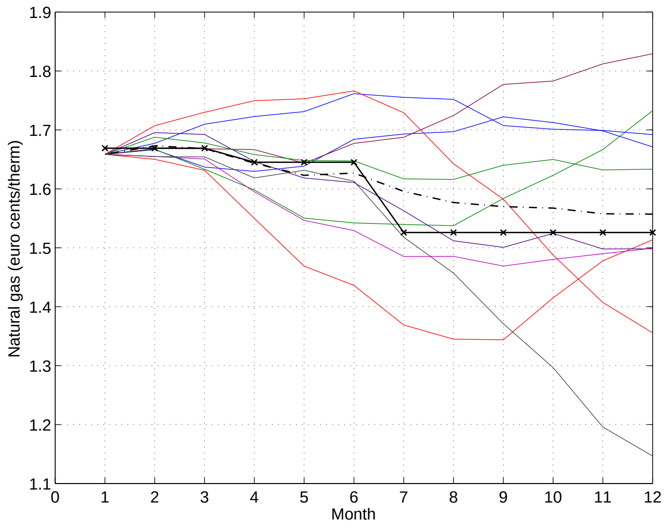

Fig. 6. Natural gas price scenarios for 2003.

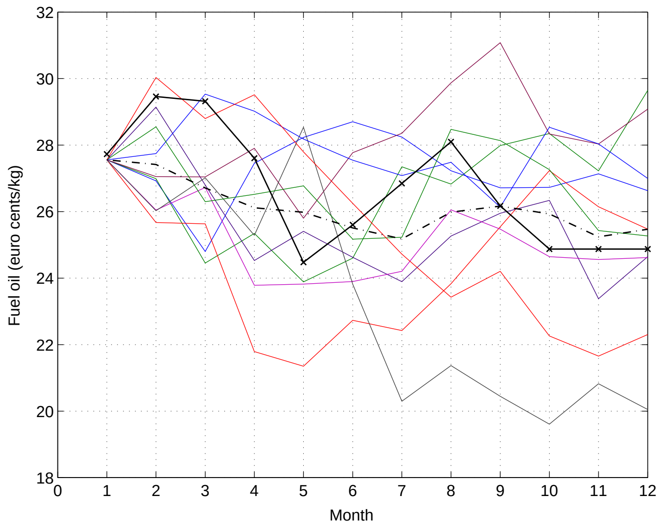

Fig. 7. Fuel oil price scenarios for 2003.

Brent futures and spot historical prices for the forecast of future Brent spot prices.

To check that the tree generated has the same statistical properties as the distributions constructed with historical Brent futures and spot prices, the moments of the distributions used as input data and those of the tree generated are compared. The iteration process ends when the difference of the moments of the distributions compared is below a configurable threshold.

Once the Brent spot price tree is generated, linear regressions between Brent spot prices and the desired fuel prices are employed for the scenario generation of natural gas and fuel oil prices.

Finally, an example of the forecast and scenario generation of energy prices for 2003 is presented to demonstrate how the model works. The results obtained are satisfactory and reliable for supporting industrial consumer decisions.

TABLE III
Terms of the deviation measure $\phi_z\left|\frac{m_z(k)-m'_z(k)}{m_z(k)}\right|$

k	$z=1$	$z=2$	$z=3$	$z=4$	Total
1	0.0098	0.0786	0.0177	0.0255	0.1316
2	0.0045	0.0346	0.0392	0.0328	0.1111
3	0.0481	0.0514	0.0163	0.0240	0.1398
4	0.0231	0.0581	0.0410	0.0019	0.1241
5	0.0051	0.0929	0.0099	0.0092	0.1171
6	0.0121	0.1213	0.0045	0.0081	0.1460
7	0.0182	0.0722	0.0049	0.0178	0.1131
8	0.0191	0.0057	0.0319	0.0709	0.1276
9	0.0443	0.0135	0.0194	0.0643	0.1415
10	0.0503	0.0139	0.0203	0.0544	0.1389
11	0.0407	0.0318	0.0696	0.0035	0.1456
12	0.0376	0.0018	0.0758	0.0061	0.1213

References

[1] H. Brand, E. Thorin, C. Weber, J. Hlouskova, S. Kossmeier, M. Obersteiner, and A. Scnabl, "Methodology to identify the relevant uncertainties," Deliverable d3.1, OSCOGEN (Optimization of cogeneration systems in a competitive market environment), October 2001, http://www.oscogen.ethz.ch.

[2] J. Dupacŏvá, G. Consigli, and S. W. Wallace, "Scenarios for multistage stochastic programs," *Baltzer Journals*, June 2000, http://www.iot.ntnu.no/iok_html/users/sww/vanc-ny.pdf.

[3] B. Johnson and G. Barz, *Energy Modelling and the Management of Uncertainty*, Risk Books, 1999.

[4] A. Berkelaar, H. Hoek, and A. Lucas, "Arbitrage and sampling uncertainty in financial stochastic programming models," http://www.eur.nl/WebDOC/doc/econometrie.

[5] O. B. Fosso, A. Gjelsvik, A. Haugstad, B. Mo, and I. Wangensteen, "Generation scheduling in a deregulated system. The norwegian case," *IEEE Transactions on Power Systems*, vol. 14, no. 1, pp. 75–80, February 1999.

[6] K. Høyland and S. W. Wallace, "Generating scenario trees for multi-stage decision problems," *Management Science*, pp. 295–307, 2001.

[7] K. Høyland, M. Kaut, and S. W. Wallace, "A heuristic for generating scenario trees for multistage decision problems," February 2002, http://home.himolde.no/∼wallace/reports.htm.

[8] R. Kouwenberg and S. A. Zenios, "Stochastic programming models for asset liability management," *Working Paper 01-01, School of Economics and Management, University of Cyprus*, vol. http://mscf.gsia.cmu.edu/BPM/zenios.pdf, May 2001.

[9] S. Takriti, J. R. Birge, and E. Long, "A stochastic model for the unit commitment problem," *IEEE Transactions on Power Systems*, vol. 11, no. 3, pp. 1497–1508, August 1996.

[10] G.E. Box, G.M. Jenkins, and G.C. Reinsel, *Time series analysis: forecasting and control*, Prentice Hall, 1994, third edition.

[11] C. Batlle, *A model for electricity generation risk analysis*, Ph.D. thesis, Universidad Pontificia Comillas de Madrid. Escuela Técnica Superior de Ingeniería (ICAI), 2002.

[12] D. Pilipović, *Energy Risk. Valuing and Managing Energy Derivatives*, McGraw-Hill, 1997.

[13] E. Gómez-Villalva and A. Ramos, "Optimal energy management of an industrial consumer in liberalized markets," *IEEE Transactions on Power Systems*, vol. 18, no. 2, pp. 716–723, May 2003.

[14] H. Heitsch and W. Römisch, "Scenario reduction algorithms in stochastic programming," *Computational optimization and applications*, vol. 24, pp. 187–206, 2003.

[15] N. Gröwe-Kuska J. Dupacŏvá and W. Römisch, "Scenario reduction in stochastic programming. An approach using probability metrics," *Mathematical programming*, vol. Ser. A 95, pp. 493–511, 2003.

Acknowledgments

The authors would like to thank Carmen Gómez-Villalva from Repsol-YPF for her valuable explanations on the current Spanish energy market framework.

Emilio Gómez-Villalva (S'98-A'00-M'02) was born in Granada, Spain, in 1973. He received the degree in Electrical Engineering from the Universidad Pontificia Comillas (ICAI), Madrid, Spain, in 1997. He has developed his professional career in engineering and technical consulting within the electricity industry. He worked for NorSistemas (Unión Fenosa) and Iberinco, and today he is working for Gómez-Villalva Ingenieros. He is also completing his Ph.D. degree at the Universidad Pontificia Comillas (ICAI).

Andrés Ramos was born in Guadramiro, Spain, in 1959. He received the degree in Electrical Engineering from the Universidad Pontificia Comillas (ICAI), Madrid, Spain, in 1982 and the Ph.D. degree in Electrical Engineering from the Universidad Politécnica de Madrid, Spain, in 1990. From 1982 to 1984 he was a Junior Research Staff at the Instituto Tecnológico para Postgraduados. Since 1984 he has been a Senior Research Staff at the Instituto de Investigación Tecnológica, ICAI. He was a visiting scholar at Stanford University in 1991-1992 academic year. Today, he is a Full Professor at the Departamento de Organización Industrial (DOI), ICAI.

His areas of interest include operations, planning and economy of electric power systems, operations research applied to power systems and software development.