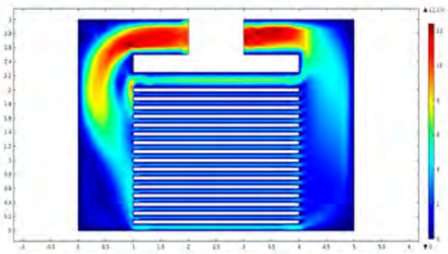
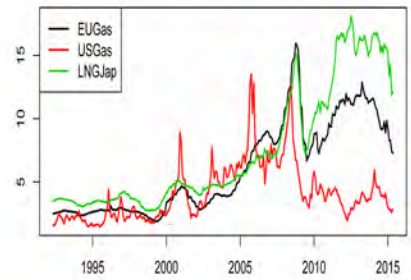
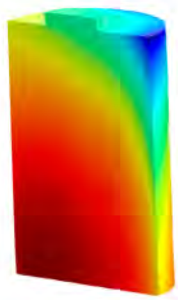
110A European Study Group with Industry







Laureano Escudero Bueno
Manuel Febrero Bande
José Antonio García Rodríguez
Javier Roca Pardiñas
Julio Rubio García
Editors

Facultad de Matemáticas | Santiago de Compostela
June 29 - July 2, 2015

[math-in].net
Red Española Matemática-Industria

ITMATI

PROCEEDINGS OF THE

110A EUROPEAN STUDY GROUP WITH INDUSTRY (110A ESGI)

Santiago de Compostela, 29th June-2nd July, 2015

Editors

Laureano Escudero Bueno
University Rey Juan Carlos and math-in
laureano.escudero@urjc.es

Manuel Febrero Bande
University of Santiago de Compostela
manuel.febrero@usc.es

José Antonio García Rodríguez
University of A Coruña
jagrodriguez@usc.es

Javier Roca Pardiñas
University of Vigo
roca@uvigo.es

Julio Rubio García
University of La Rioja
julio.rubio@unirioja.es

PROCEEDINGS OF THE

110A EUROPEAN STUDY GROUP WITH INDUSTRY (110A ESGI)

This volume contains the proceedings of the 110A European Study Group with Industry (110A ESGI) held in Santiago de Compostela, Spain, from June 29th to 2nd July, 2015.

The 110A ESGI was jointly organized by the Technological Institute for Industrial Mathematics (ITMATI) and the Spanish Network for Mathematics and Industry (math-in).

It was co-sponsored by the Thematic Network RTmath-in, granted by the Ministry of Economy and Finance within the dynamic actions "Networks of Excellence" at 2014 call and by the Industrial Mathematics Technology Network (Red TMATI), granted by the Department of Culture, Education and University of the Xunta de Galicia. This 110A ESGI was also framed as activity in the Joint Research Unit ITMATI-Repsol which has funding of the Galician Agency for Innovation and the Ministry of Economy and Finance in the framework of the Spanish Strategy Innovation in Galicia.

Index

Introduction

Initiated in Oxford in 1968, Study Groups with Industry provide a forum for industrial scientists to work alongside academic mathematicians on problems of direct industrial relevance. They are an internationally recognized method of technology and knowledge transfer between academic mathematicians and industry, usually lasting one week.

The success of the ESGI lies in its unique format which has been copied around the world, and which allows Mathematicians to work on reduced groups to study problems presented by industry. These problems arise from any economic sector thanks to the versatility of Mathematics.

The objective is to present the capabilities of Mathematics and its applicability in a large part of the challenges and needs that industry presents. It aims to bring small, medium and large companies a technology with great potential, used by highly qualified researchers and which does not require large investments to use.

Therefore, collaboration between industry experts and researchers is key to address technological innovation issues by using successful mathematical techniques. ESGI contributes to the promotion of mathematics and helps companies to use Mathematics to improve their processes.

The goals which want to be reached at the ESGI are:

- found solutions and insights into existing industrial problems;
- established lasting and productive working links between research applied mathematicians and industry;
- propose new lines of research based on business challenges;
- reinforce the importance of mathematics in industry and mathematical profiles companies;
- stimulated greater awareness in the wider community of the power of mathematics in providing solution paths to real-world problems.

Finally, it should be pointed out that 56 researchers, students, professors and company technicians contributed to a successful 110A ESGI.

Santiago de Compostela on 2nd July, 2015

Members of the Scientific Committee:

- Laureano Escudero Bueno. Professor of Statistics and Operations Research at the University Rey Juan Carlos.Member of the Management Board of the Spanish Network for Mathematics and Industry (math-in)

- Manuel Febrero Bande. Professor of the Department of Statistics and Operations Research at the University of Santiago de Compostela. Affiliatede researcher of ITMATI.

- José Antonio García Rodríguez. Associate Professor of Applied Mathematics at the University of A Coruña. Affiliated researcher of ITMATI.

- Javier Roca Pardiñas. Associate Professor of Statistics and Operations Research at the University of Vigo. Affiliated researcher of ITMATI.

- Julio Rubio García. Professor of Computer Science and Artificial Intelligence at the University of La Rioja. Secretary of the math-in Executive Board.

Model order reduction for Li-ion batteries simulation at cell scale

Academic Coordinator Jerónimo Rodríguez
University University of Santiago de Compostela

Business Coordinators José Francisco Rodríguez Calo and David Aller Giráldez
Company Repsol Technology Center

Specialist Fernando Varas Mérida
Company Technical University of Madrid and Technological Institute for Industrial Mathematics

Team Alfredo Bermúdez de Castro (University of Santiago de Compostela and Technological Institute for Industrial Mathematics), Teresa Cao (University of A Coruña and Technological Institute for Industrial Mathematics), Noemi Esteban (University of Santiago de Compostela), Pedro Fontán Muiños (Technological Institute for Industrial Mathematics), Francisco Pena (University of Santiago de Compostela and Technological Institute for Industrial Mathematics), María Luisa Seoane (University of Santiago de Compostela)

Model order reduction for Li-ion batteries simulation at cell scale

David Aller, José Francisco Rodríguez, Jerónimo Rodríguez, Fernando Varas

Abstract

The aim of this study group was to propose order reduction methods for simulating the performance of batteries that are of interest for use in optimization methods associated with these simulations.

1 Introduction

The fundamental idea of this working group was to develop methods to accelerate the process of design of energy storage systems optimization, specifically based on lithium-ion technology batteries. The optimization of these designs required simulation of the processes of charging and discharging of the battery, seeking to maximize its life maintaining operational requirements as directed. The simulation of the processes of loading and unloading required solving a coupled system of partial differential equations with the presence of nonlinear terms result of chemical reactions that occur during operation of the battery. Optimizing designs or charging of the batteries requires making decision on the optimum design variables in a range from well below the actual simulated period time, which in turn imposes very stringent limits on time resolution of the simulations.

Analysis of the influence of the air speed and the temperature in the quality and in the energetic efficiency of the wood drying process for Galician pine

---

Academic Coordinators
José Antonio García Rodríguez and Javier Roca Pardiñas,
University Associate Professor of Applied Mathematics in University of A Coruña and Affiliated researcher of ITMATI and Associate Professor of Statistics and Operations Research in University of Vigo and Affiliated researcher of ITMATI.

Business Coordinator
Gonzalo Piñeiro Veiras,
Company Technician in CIS Madera, Axencia Galega de Innovación.

Team Aderito Araújo (University of Coimbra), Alejandro Lopez (University of A Coruña), Andrés Gómez Tato (CESGA), David González Peñas (Technological Institute for Industrial Mathematics), Manuel Cruz (PT-MATHS-IN), Peregrina Quintela (University of Santiago de Compostela and Technological Institute for Industrial Mathematics)

Analysis of the influence of the air speed and the temperature in the quality and in the energetic efficiency of the wood drying process for Galician pine

José Antonio García Rodríguez, Javier Roca Pardiñas, Gonzalo Piñeiro Veiras

Abstract

The first goal of this study group is to know the influence of the speed and the temperature during the drying process in order to provide an uniform distribution into the kiln. The second goal is to use this to optimize the design of the shape of a kiln.

1 Introduction

Currently, there are bibliographic references about the influence of the air speed in the drying process, especially in conifers from northern of Europe, but few exist in the case of the drying of Galician autochthones conifers as the Galician Pine (*Pinus pinaster*). Know the influence of the air speed and of the temperature in the wood drying will permit its adjustment during the process. In fact, this information will permit to obtain optimal parameters of fan speed, air temperature, spaced between the lumbers and distances among wooden stacks, that help to decrease the drying time, increase the moisture content uniformity, and to obtain an improvement in the final quality of the dried wood and, as consequence, a reduction of the electric consumption[1][2].

2 Mathematical Modelling

To model this problem at least four phenomena must be considered :

- Fluid dynamic,
- Heat transfer processes,
- Hydro dynamic processes with phase changes,
- Mechanical process corresponding to non isotropic materials.

Figure 1: In the images a kiln can be observed, in the first an inside view, and an outer view at the second.

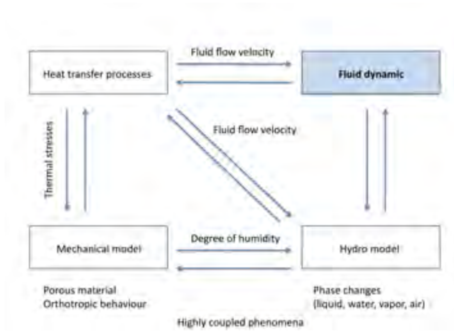


Figure 2: Phenomena involved in the drying process

As we can see in the 2, in the involved processes a high level of coupling of these phenomena occurs. In the study group only hydro fluid dynamic phenomena have been considered.

With the aim to obtain the mathematical model corresponding to the drying process the following conservation laws had been considered (see [1], [2] for a more detailed description):

- **Mass conservation**:

$$\frac{\partial \rho_G}{\partial t} + \mathrm{div}(\rho_{\mathrm{G}} \overrightarrow{u_G}) = \dot{\mathrm{m}}, \qquad \text{in } \Omega_{\mathrm{G}}, \qquad (*1)$$

where ρ_G is the density of the gas mixture, u_G is the velocity of the gas, $\dot{m}$ is the mass proceeding from other phases per unit time and volume and Ω_G is the gas domain inside the kiln.

We consider

$$\rho_G = X_A \rho_A + X_V \rho_V = (1 - X_V)\rho_A + X_V \rho_V,$$

being ρ_A, X_A, ρ_V and X_V the densities and the volume fractions of the air and the vapeur of water constituting the gas. Assuming that

$$\dot{m} = 0,$$

and the process is stationary,

$$\frac{\partial \rho_G}{\partial t} = 0,$$

equation(*1) can be written as

$$\mathrm{div}(\rho_{\mathrm{G}}\overrightarrow{u_G}) = 0.$$

- **Equilibrium law for the fluid model**:

$$\rho_G \frac{D\overrightarrow{u_G}}{Dt} = \mathrm{div}(\sigma_{\mathrm{G}}) + \overrightarrow{F_G},$$

where the total derivative of $\overrightarrow{u_G}$ with respect to time is given by:

$$\frac{D\overrightarrow{u_G}}{Dt} = \frac{\partial \overrightarrow{u_G}}{\partial t} + (\overrightarrow{u_G} \cdot \nabla)\overrightarrow{u_G},$$

and stress tensor is obtained through the relation

$$\sigma_G = -\pi_G I + 2\eta_G \epsilon(\overrightarrow{u_G}) + \xi_G \mathrm{div}\overrightarrow{u_G}\mathrm{I},$$

where π_G is the pressure, η_G and ξ_G are the viscosity coefficients, with $\xi_G = -\frac{2}{3}\eta_G$, $\epsilon(\overrightarrow{u_G})$ is the linearised strain rate given by

$$\epsilon(\overrightarrow{u_G}) = \frac{1}{2}(\nabla\overrightarrow{u_G} + \nabla\overrightarrow{u_G}^T),$$

and $\overrightarrow{F_G} = 0$.

Then, we obtain the compressible Navier-Stokes equation for the perfect gas

$$\rho_G \frac{D\overrightarrow{u_G}}{Dt} + \nabla\pi_G - \mathrm{div}\Big(\eta_{\mathrm{G}}(\nabla\overrightarrow{u_G} + \nabla\overrightarrow{u_G}^{\mathrm{T}}) - \frac{2}{3}\eta_{\mathrm{G}}\nabla\cdot\overrightarrow{u_G}\mathrm{I}\Big) = 0$$

with $\pi_G = \rho_G RT$, where $R = 8314\ J/kmolK$ is the universal gas constant.

- **Liquid Mass Balance**

$$\frac{\partial\left(\varepsilon_L \rho_L\right)}{\partial t} + div\left(\rho_L \overrightarrow{u_L}\right) + \dot{m} = 0 \tag{1}$$

where ε_L is the liquid to total volume ratio (m^3 liq/m^3 sol), ρ_L is the density of the liquid (kg/m^3), t is time (s), $\overrightarrow{u}_L$ is the velocity vector of the liquid (m/s) and $\dot{m}$ is the rate of mass evaporation per unit volume (kg/m^3s). The remaining terms of equation (1) are: storage mass of liquid in the control volume, flow of liquid mass through the system and rate of volumetric evaporation of the liquid/vapor phase change.

- **Water Vapor Mass Balance**

$$\frac{\partial\left[\left(\varepsilon - \varepsilon_L\right) X_V \rho_G\right]}{\partial t} + div\left(X_V \rho_G \overrightarrow{u_G} + \overrightarrow{J_V}\right) - \dot{m} = 0, \tag{2}$$

being J_V the diffusive term of the vapor concentration (kg/m^2s) defined as

$$\overrightarrow{J_V} = -\rho_G\left(\varepsilon - \varepsilon_L\right) D_{EFF} \nabla X_V. \tag{3}$$

- **Air Mass Balance**

$$\frac{\partial\left[(\varepsilon-\varepsilon_L)X_A\rho_G\right]}{\partial t}+\nabla\left(X_A\rho_G\overrightarrow{u_G}-\overrightarrow{J}_V\right)=0, \tag{4}$$

where X_A is the concentration of air in the air-vapor mixture (kg air/kg air vapor mixture). The remaining terms of equation (4) are: storage of air mass in the control volume, and flow of air mass through the system by convection and diffusion.

- **Liquid Momentum Equation (Darcy's Law)**

$$\overrightarrow{u_L}=-\left(\frac{\alpha_L}{\mu_L}\right)\nabla\left(\pi_G-\pi_C\right), \tag{5}$$

where α_L is the directional permeability of the liquid (m^2), μ_L is the dynamic viscosity of the liquid ($N.s/m^2$), π_G is the pressure of the gas phase (Pa) and π_C is the capillary pressure of the liquid (Pa). The pressure gradient by capillary effects was neglected, since no liquid movement was regarded in the present model, therefore the water vapor air mixture momentum equation can be simplified to

Water Vapor Air Mixture Momentum Equation (Darcys Law)

$$\overrightarrow{u_G}=-\left(\frac{\alpha_G}{\mu_G}\right)\nabla\left(\pi_G\right) \tag{6}$$

where α_G is the directional permeability of the gas phase (m^2) and μ_G is the dynamic viscosity of the gas phase ($N.s/m^2$).

- Boundary conditions

  To complete model the following boundary conditions have been considered (see Figure 2)

  1. On the kiln walls, Γ_0, a no slip condition: $\vec{u}=0$.
  2. On the air inlet boundary, Γ_1: $\vec{u}_G=-u_0\vec{n}$, with a inlet speed $u_0=10\,m/s$.
  3. On the air outlet boundary, Γ_2, a null pressure $\pi_G=0$.

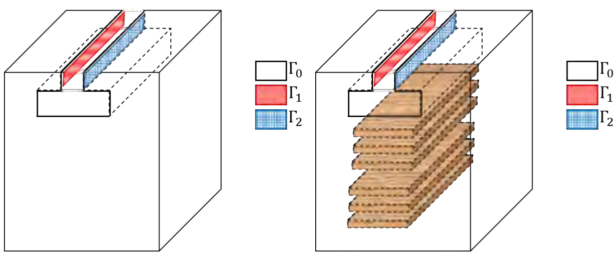


Figure 3: Computational domain of the dryers and its boundaries

A two-dimensional simulation has been performed. In Figure 3 a mesh of a 2D vertical section of the kiln is shown. For the mesh, a medium mesh was used. The narrower areas between wood planks have been refined in the mesh to better simulate the wall effects of the circulating air.

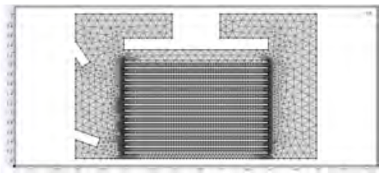


Figure 4: Mesh of a 2d vertical section

3 Numerical simulation results

Two different test were performed using COMSOL software package, one without deflectors and the other with deflectors on the sides. In the image of Figure 4, the velocity magnitude can be observed if no deflectors exist at the lateral wall of the kiln, while in the right image when they exist.

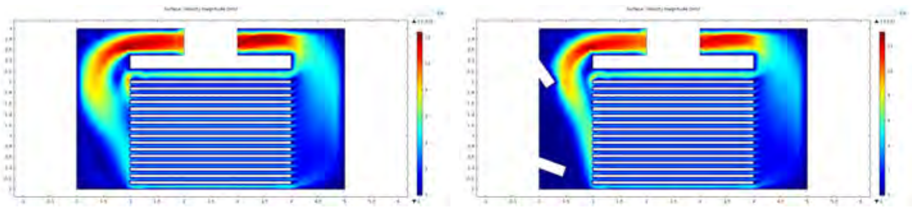


Figure 5: Velocity magnitud: the first image without deflectors and the second one with deflectors

There are significant differences in the distribution of air inside the kiln according to whether or not there are deflectors as can be seen in Figure 4. If there are deflectors, the incoming air (which will be hot) will reach the wood tables more quickly, increasing the efficiency of the process.

4 Conclusion

The 2d simulation seems to be useful as first approach to study the real problem, and different kiln designs can be numerically tested in a quick way. The geometry and the use of deflectors has a key importance for the distribution of the air flow.

The 3d geometry can be fully parameterized so that it will be easy to test different geometries, deflectors type and their orientations and charge distribution of the kiln, although the calculation times will be considerably longer. Therefore, it is recommended to make a first approximation of the optimal design considering 2d simulations.

In the future, it could be also interesting introducing the other phenomena of the drying process in the simulation: heat transfer, mass transfer and the mechanical model.

References

[1] P. C. C. Pinheiro, T.J. Raad and M.I. Yoshida, *Model of a process for drying Eucalyptus spp at high temperatures*, Braz. J. Chem. Eng. vol. 15 n. 4 Sao Paulo Dec. 1998.

[2] F. Thibault, D. Marceau, R. Younsi and D. Kocaefe , *Numerical and experimental validation of thermo-hidro-mechanical behaviour of wood during drying process*, International Communications in Heat and Mass Transfer Volume 37, Issue 7, Pages 756-760, August 2010.

LNG import forescast in Spain

Academic Coordinator Manuel Febrero Bande
University Professor of Statistics and Operations Research in University of Santiago de Compostela and affiliated researcher of ITMATI.

Business Coordinator Laurent Moriceau
Company Responsible for European Regulation. REGANOSA

LNG import forecasts in Spain

Manuel Febrero–Bande, Laurent Moriceau

Abstract

The goal of this study group is obtain forecasts for the demand/consumption and prices of Liquefied Natural Gas (LNG) for the short and medium term. The problem was proposed by Reganosa which is the principal company in the northwestern of Spain devoted to all tasks related with LNG: Vessel loading/unloading, regasification, storage and the distribution to customers through pipelines or using trucks. The price and the consumption of LNG are affected not only by local rules but by a lot of restrictions derived from global markets: oil prices, long term agreements among companies and even natural disasters.

Laurent Moriceau was the business coordinator and Manuel Febrero–Bande the academic one. The complete list of collaborators (in alphabetical order) was: Bartomeu Coll Vicens, Eliana Costa e Silva, Pablo Díez Albert, Laureano Escudero Bueno, María José Ginzo Villamayor, Isabel Cristina Lopes, Adela Martínez Calvo, Andrea Meilán Vila, Manuel Oviedo de la Fuente and Carlos Vázquez Cendón.

1 Introduction

The consumption and/or the price of LNG depends, in general, on many factors derived from a competitive global market. Some of them could be related with other sources of energy and its relative cost compared with LNG. Other factors are completely uncontrolled like private agreements between producers and redistributors/customers or even natural events that can alter the composition of the energetic pool of a country. Our goal is to produce forecasts using the information/data publicly or easily available in the web. There are many other sources of information but typically the data obtained are fragmented, incomplete and cannot be mixed with other variates due to its resolution temporal or spatial. Also, the naive point of view of considering only data publicly available is applied to the statistical models used for the analysis. We begin with the most basic models/tools for treating the information and change to more sophisticated ones when needed.

2 Time Series. Application for GNL Demand

We are trying to establish optimal forecasts for consumption and prices. The basic idea is that the pattern followed by the data in the past will maintain

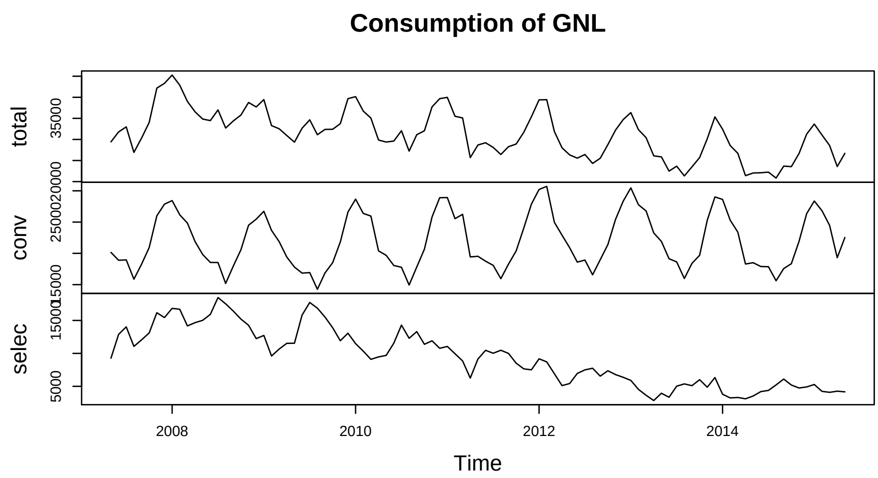


Figure 1: Consumption of GNL in Spain by years

unchanged in the future and so, the time series framework is a good choice to begin with.
A time series $\{X_t\}_{t\in T}$ is an ordered sequence of random variables that are supposed to be dependent or correlated with respect to the index t (typically time). We begin considering Regular Time Series, i.e. the observations are taken only at specific time intervals equally spaced. In our case, we have monthly observations of consumption of GNL (in GWh) since April, 2007 divided into two parts: the conventional one ('conv') for households and industries and the one ('selec') specifically consumed by the power generation plants in the electric network. The total demand is the sum of this two components. The conventional one shows a clearly annual pattern with peaks in winter and valleys in summer whereas the part whereas the selec component have a long decay from 2007.
Typically, a time series is decomposed in three parts: $X_t = T_t + S_t + Z_t$ where T_t is the long-term *trend* (deterministic), S_t is the *seasonal* component (deterministic) and Z_t is the *residual, irregular or random effect* (stochastic). The first two parts must be identified and removed before any stochastic model can be applied to the third one which is supposed to be stationary.
A time series is said *stationary (2nd order)* if

$$\begin{aligned} \mathbb{E}\left[Z_t\right] &= \mu \quad \forall t \\ \operatorname{Cov}\left(Z_t, Z_{t-j}\right) &= \mathbb{E}\left[(Z_t-\mu)(Z_{t-j}-\mu)\right] = \gamma_j \quad \forall t, j \\ \rho_j &= \frac{\operatorname{Cov}\left(Z_t, Z_{t-j}\right)}{\sqrt{\operatorname{Var}\left(Z_t\right)\operatorname{Var}\left(Z_{t-j}\right)}} = \frac{\gamma_j}{\gamma_0} \end{aligned}$$

These properties establish that the stochastic part has a constant mean and variance and the dependence does not depend on t but on the lag between times. To check the stationarity in mean and variance, an usual tool is to compute the mean and variance for long periods of time. In this case, and given that our data has a monthly structure, it is usual to compute the annual mean and variance to detect if there is a long term trend or the variance is not constant.

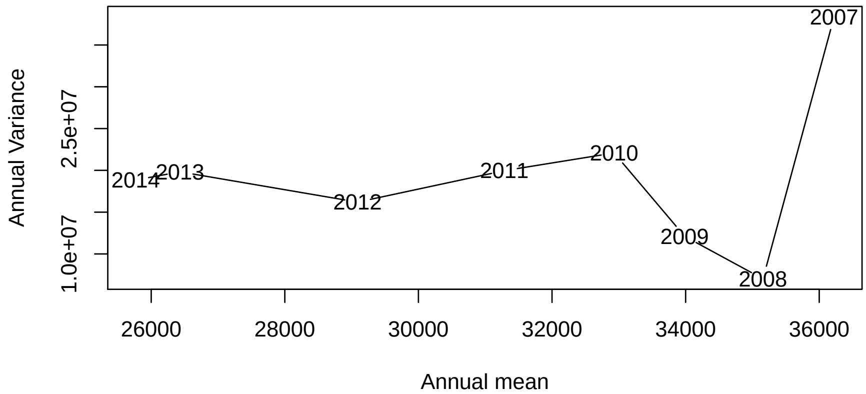
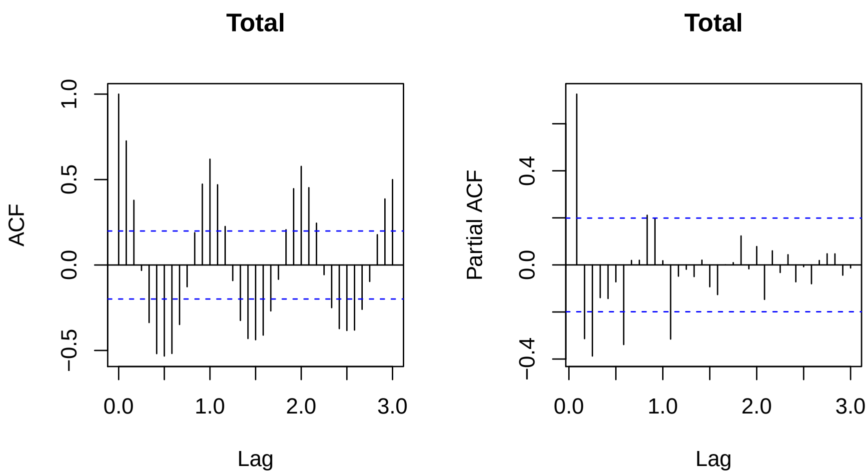


Figure 2: Annual Mean and Variance for Total Consumption of GNL

¿From Figure 2 it can deduced that there is no relation between the annual mean and the annual variance which seems constant along the years. There is a slow decay pattern in the mean but, probably, not enough to be considered as a deterministic pattern. This can be checked estimating a regressión of the data as a function of the time, i.e. $X(t) = \beta_0 + \beta_1 t + \epsilon$ where the parameter β_1 provides the inference about the trend.

```
>                beta0       beta1
> Coef.      2911652.0 -1432.0633
> Std.Error  499275.8    248.2311
```

As can be deduced from the previous table, there is a slightly significant decay along the years (about 1432.063 per year) that must be removed before estimating an stochastic model.



The usual tools for the identification of a time series model are the ACF (the

sample version of the autocorrelation function) and the PACF (partial autocorrelation function) which is the correlation between (z_t, z_{t+j}) given the information of $(z_{t+1}, \ldots, z_{t+j-1})$.
The classical time series models for stationary process are derived from the classic book [1], ([2, for an updated revision]) and are called ARMA models.
Let ϵ_t a white noise process (i.e. $\mathbb{E}[\epsilon_t] = 0$, $\text{Var}(\epsilon_t) = \sigma^2$, and $\rho_j = 0, \forall j \neq 0$) and B the backward operator $B^j(Z_t) = Z_{t-j}$

- An autoregressive model of order p (AR(p)) is defined as: $Z_t = c + \phi_1 Z_{t-1} + \ldots + \phi_p Z_{t-p} + \epsilon_t$ where $\Phi_p(B)z_t = c + \epsilon_t$ with $\Phi_p(B) = 1 - \phi_1 B - \ldots - \phi_p B^p$
- A moving average model of order q (MA(q)) is defined as: $Z_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \ldots + \theta_p \epsilon_{t-q}$ or equivalently, $Z_t = \mu + \Theta_q(B)\epsilon_t$ with $\Theta_q(B) = 1 + \theta_1 B + \ldots + \theta_q B^q$
- An ARMA(p,q) model is defined as: $Z_t = c + \phi_1 Z_{t-1} + \ldots + \phi_p Z_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \ldots + \theta_p \epsilon_{t-q}$ or in simplified form $\Phi_p(B)z_t = c + \Theta_q(B)\epsilon_t$

The identification of a model for a particular data set is then performed computing the sample ACF and PACF functions and comparing its shape with the theoretical pattern under the different models:

- AR(p) models
  - ACF: $\rho_k = \sum_{i=1}^{m} G_i^k \sum_{j=1}^{d_i-1} A_{ij} k^j$ with G_i^{-1} the roots of $\Phi_p(B) = 0$: Mixture of exponential and/or sinusoidal decays.
  - PACF: $P_k \neq 0, k \leq p$, and $P_k = 0, k > p$
- MA(q) models
  - ACF: $\rho_k \neq 0, k \leq q$, and $\rho_k = 0, k > q$
  - PACF: Mixture of exponential and/or sinusoidal decays.
- ARMA(p,q) models
  - ACF: First q non zero autocorrelations and then mixture of exponential and/or sinusoidal decays.
  - PACF: First p non zero autocorrelations and then mixture of exponential and/or sinusoidal decays.

When the time series is not stationary in mean, it can be converted to a stationary one through d differentiations or through estimating a long term regression. In the first case, the model is called ARIMA(p, d, q) including the number of differences needed. Indeed, more tests about stationarity can be done using the roots of the characteristic polynomials $(\Phi_p(B), \Theta_q(B))$.
In this case, the ACF function shows a clear periodic annual pattern (the bars rise every 12 lags) and it seems to be good explained by an AR(3). The trend is modeled as a linear component which can be computed with the following command in `R`.

```
res.arma = arima(total[1], order = c(3, 0, 0), seasonal = list(order = c(1,
    0, 0), period = 12), xreg = t[1:92])
```

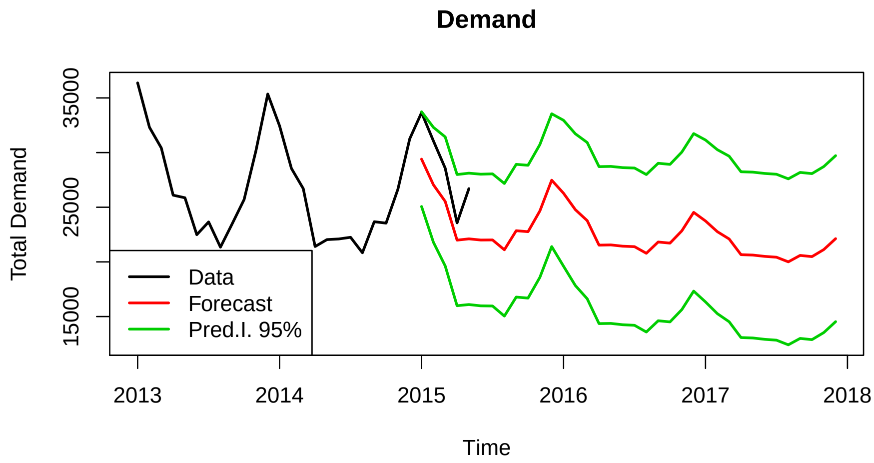


Figure 3: Predictions of the Total GNL Demand

```
>
> Call:
> arima(x = total[1], order = c(3, 0, 0), seasonal = list(order = c(1, 0, 0),
>     period = 12), xreg = t[1:92])
>
> Coefficients:
>          ar1     ar2      ar3    sar1  intercept     t[1:92]
>       0.6857  0.1510  -0.2816  0.6185  3125227.6  -1538.4735
> s.e.  0.1093  0.1247   0.1068  0.0979   734799.4    365.3635
>
> sigma^2 estimated as 4696359:  log likelihood = -840.55,  aic = 1695.09
```

After checking all the structural hypothesis with satisfactory results, the time series model can be employed for constructing the forecasts shown in Figure 3. The same analysis could be done for the two components of the GNL demand. There are slightly differences between the models although the most interesting one is that there is no significative trend for the conventional demand. The predictions are shown in Figure 4.

```
res.arma = arima(conv[1], order = c(3, 0, 0), seasonal = list(order = c(1, 0,
    0), period = 12), xreg = t[1])

>
> Call:
> arima(x = conv[1], order = c(3, 0, 0), seasonal = list(order = c(1, 0, 0), period = 12),
>     xreg = t[1])
>
> Coefficients:
>          ar1     ar2      ar3    sar1  intercept        t[1]
>       0.7156  0.1612  -0.2833  0.7395   301595.0   -139.1678
> s.e.  0.1292  0.1238   0.1383  0.1086   592464.5    294.6035
>
> sigma^2 estimated as 1971374:  log likelihood = -802.52,  aic = 1619.04
```

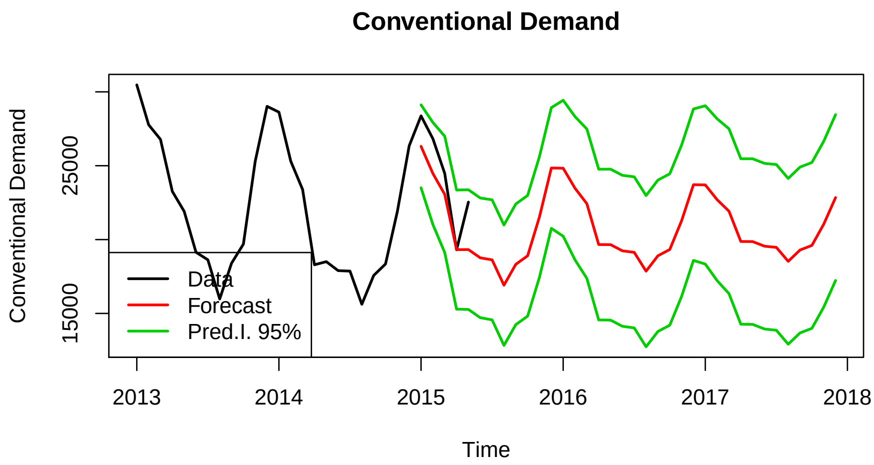


Figure 4: Predictions of the Conventional GNL Demand

The trend for the Electric Network component has a strong decay and a small periodic contribution. The predictions are shown in Figure 5 that reaches zero at the end of 2017. This is an effect of the interval considered and probably of the evolution of the domestic electric market in the last three years. Recall that these models only use the past of the time series as information without any reference to other variables (for instance, the price of other sources of energy) that can be useful.

```
res.arma = arima(selec[1], order = c(3, 0, 0), seasonal = list(order = c(1,
    0, 0), period = 12), xreg = t[1])

>
> Call:
> arima(x = selec[1], order = c(3, 0, 0), seasonal = list(order = c(1, 0, 0),
>     period = 12), xreg = t[1])
>
> Coefficients:
>          ar1     ar2      ar3    sar1  intercept        t[1]
>       0.7552  0.1310  -0.1473  0.3421  2867116.6  -1420.9524
> s.e.  0.1155  0.1436   0.1097  0.1223   597743.7    297.1931
>
> sigma^2 estimated as 1660929:  log likelihood = -790.62,  aic = 1595.24
```

In order to reduce the variability of the predictions, some other variables could be considered. In particular, in Figure 6 the temperature in Madrid–Barajas is plotted jointly with a couple of Industrial Production Indexes. Temperature can help in the conventional component as the consumption of GNL on many households depends on the outer temperature as the GNL is used as the principal heating system. On the other hand, the Industrial Production Indexes can help in the prediction of the Electric Network component because the GNL power stations mainly work when the industries needs energy and these indexes are reflecting in some way, the vitality of the industries.
The following is the result of modeling the conventional demand with a temper-

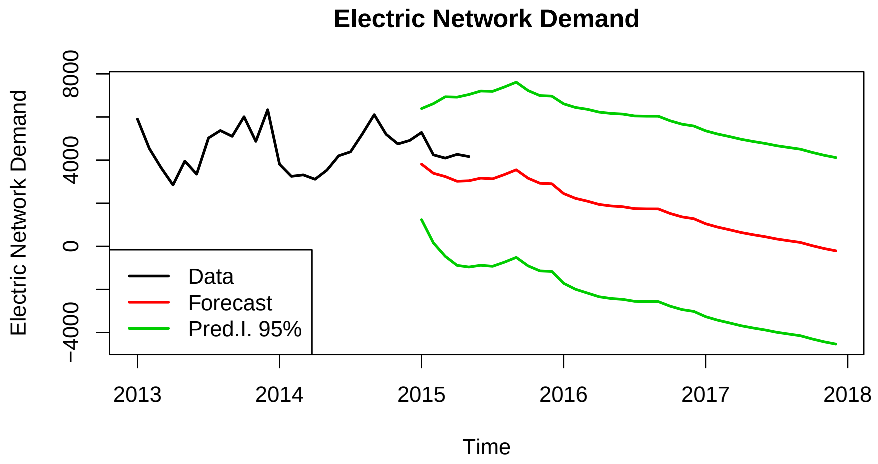


Figure 5: Predictions of the GNL Demand for power generation

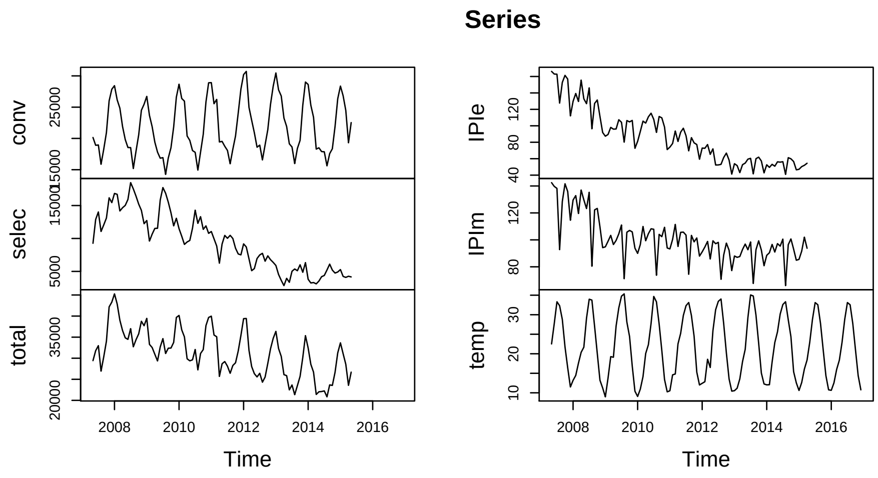


Figure 6: Components of GNL jointly with temperature and Industrial Production Indexes

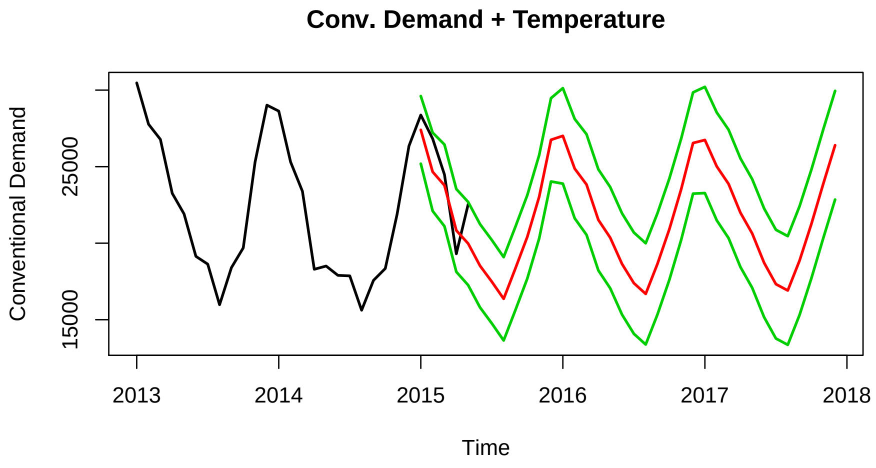


Figure 7: Predictions for Conventional Demand including Temperature

ature indicator (Madrid–Barajas temperature). This variate has a significative contribution to the model and dramatically reduces the variability of predictions as can be seen in Figure 7. In this example, we have considered 2014 monthly mean temperature as the predicted values for the next three years in order to construct forecasts.

```
>
> Call:
> arima(x = conv[1], order = c(1, 0, 0), seasonal = list(order = c(1, 0, 0), period = 12),
>     xreg = data.frame(temp = temp[1]))
>
> Coefficients:
>          ar1     sar1  intercept       temp
>       0.5845   0.6879  30371.501  -398.4995
> s.e.  0.0897   0.0791   1264.983    47.7016
>
> sigma^2 estimated as 1220509:  log likelihood = -779.28,  aic = 1568.56
```

For the prediction of the Electric Network Demand (END), we must have in mind the significative decaying trend along the last three or four years and the two Industrial Production Indexes (for energy (e) and for industries (m)). A model with both IPI's have the result that both IPI's are not significative (recall that the value of the parameter is less than two times the standard error). But, when we use only the IPIm jointly with the trend, this index is close to be relevant.

```
>
> Call:
> arima(x = selec[1], order = c(1, 0, 0), seasonal = list(order = c(1, 0, 0),
>     period = 12), xreg = data.frame(ipie = ipie[1], ipim = ipim[1], t = t[1]))
>
> Coefficients:
>          ar1     sar1  intercept      ipie     ipim           t
>       0.7924   0.3718  2644551.9  -11.3445  31.8461  -1311.4631
> s.e.  0.0724   0.1123   824432.8   20.7958  22.3427    409.7594
```

```
>
> sigma^2 estimated as 1626768:  log likelihood = -789.85,  aic = 1593.7
```

```
>
> Call:
> arima(x = selec[l], order = c(1, 0, 0), seasonal = list(order = c(1, 0, 0),
>     period = 12), xreg = data.frame(ipie = ipie[l], t = t[l]))
>
> Coefficients:
>          ar1    sar1  intercept     ipie          t
>       0.7867  0.3255  2344323.0  13.6721  -1161.6520
> s.e.  0.0739  0.1073   768448.3  11.9604    381.8213
>
> sigma^2 estimated as 1673219:  log likelihood = -790.91,  aic = 1593.81
>
> Call:
> arima(x = selec[l], order = c(1, 0, 0), seasonal = list(order = c(1, 0, 0),
>     period = 12), xreg = data.frame(ipim = ipim[l], t = t[l]))
>
> Coefficients:
>          ar1    sar1  intercept     ipim          t
>       0.7942  0.3572  2437371.9  21.9568  -1208.4484
> s.e.  0.0726  0.1090   736332.8  12.6922    365.9383
>
> sigma^2 estimated as 1634521:  log likelihood = -790,  aic = 1592
```

Both models share similar results respect to the value of the parameters being the one with IPIm slightly better than the one with IPIe. The predictions are shown in Figure 8 and are almost indistinguishable. In both cases, we use the most recent values of both indexes to fill the future. This is a pretty pessimistic scenario and probably, a better future scenario can be considered because these two indexes in the last two years show a long decay. In any case, there is a slow decay of the END that almost reaches zero at the end of 2017.

The last attempt to forecast END is by including the Electric Consumption in Spain as a covariate. The effect of this covariate is significative and also improves the forecasting with the Industrial Production Indexes (about a 11% less variability). Again the model is a $AR(1) \times AR_{12}(1)$ with a significative long–term decay (about 1180 units per year)

```
>
> Call:
> arima(x = selec[l], order = c(1, 0, 0), seasonal = list(order = c(1, 0, 0),
>     period = 12), xreg = data.frame(celec = celec[l], t = t[l]))
>
> Coefficients:
>          ar1    sar1  intercept   celec          t
>       0.7750  0.3158    2368519  6.6296  -1178.8381
> s.e.  0.0741  0.1164     601870  1.6845    299.0129
>
> sigma^2 estimated as 1444005:  log likelihood = -784.06,  aic = 1580.12
```

For the future values of electric consumption, a rise of 1.7% is considered for the next three years.

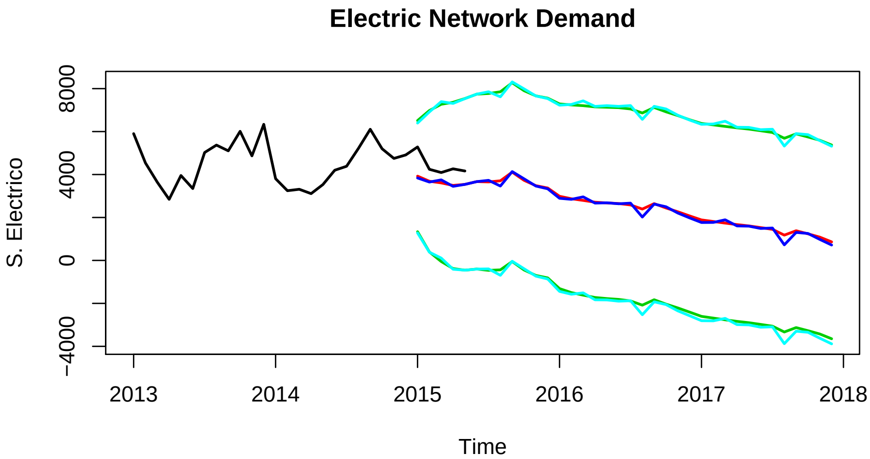


Figure 8: Prediction for Electric Network Demand using Industrial Production Indexes

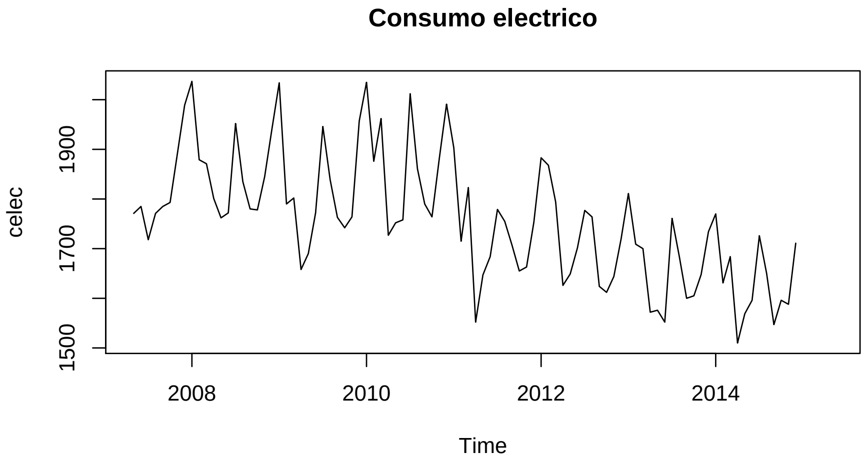


Figure 9: Electric consumption in Spain

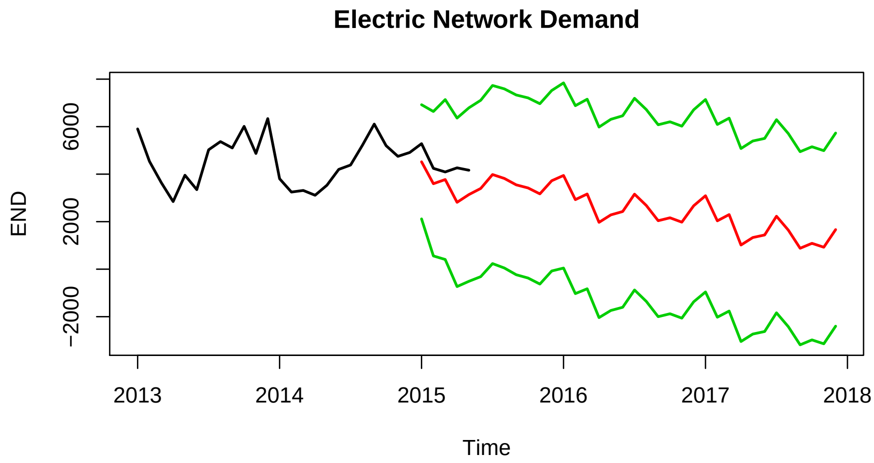


Figure 10: Prediction of Electric Network Demand using Electric Consumption

3 Multivariate Time Series. Application for GNL Prices

As a second application, we want to forecast the price of Gas in the European Union using other covariates. The price depends on many factors but mainly, the cost of alternative sources plays an important role, in particular, the price of the oil barrel. The evolution of the price in different regions of the world is provided in Figure 11. The prices up to 2008 are more or less the same with only isolated peaks for the USGas. But after 2008, the influence of certain political decisions (authorization of fracking in the US or the natural disaster of Fukushima in 2010 in Japan followed by the stop of nuclear power stations) may change the relations between prices as a global market producing a divergence between these series. Also in the last months, a cheaper oil is also making to reduce the Gas prices.
The effect of a cheaper oil can be seen in Figure 12 jointly with Gold. The original units are dollars per GWh, dollars per ounce and dollars per barrel, respectively, although in the figure some transformations were made in order to adjust the scale. Here, the price for Gold is multiplied by 0.005 and the price for a barrel of Brent by 0.15. It seems that the Brent series anticipates the evolution of EUGas whereas the Gold seems to be a long term trend.

Now, for modeling jointly these three variables, we must employ the theory of Stationary Multivariate Time Series. A good and extensive reference for this part is [3]. The main ideas are summarize in the following:
A time series, $\mathbf{Z}_t = (Z_{1,t}, \ldots, Z_{k,t})$, is a stationary k-dimensional time series (stationary vector time series) iif every component is a stationary univariate time series. So, for all t, k, l, j

$$\begin{aligned} \mathbb{E}\left[\mathbf{Z}_t\right] &= \vec{\mu} \\ \mathrm{Cov}\left(\mathbf{Z}_t, \mathbf{Z}_{t-j}\right) &= \mathbb{E}\left[(\mathbf{Z}_t - \vec{\mu})(\mathbf{Z}_{t-j} - \vec{\mu})'\right] = \Gamma(j) = \left(\gamma_{k,l}(j)\right)_{k,l} \\ \gamma_{k,l}(j) &= \mathbb{E}\left[(Z_{k,t} - \mu_k)(Z_{l,t-j} - \mu_l)\right] \\ P(j) &= \left(\rho_{k,l}(j)\right)_{k,l} = \left(\frac{\gamma_{k,l}(j)}{\sqrt{\gamma_{k,k}(j)\gamma_{l,l}(j)}}\right)_{k,l} \end{aligned}$$

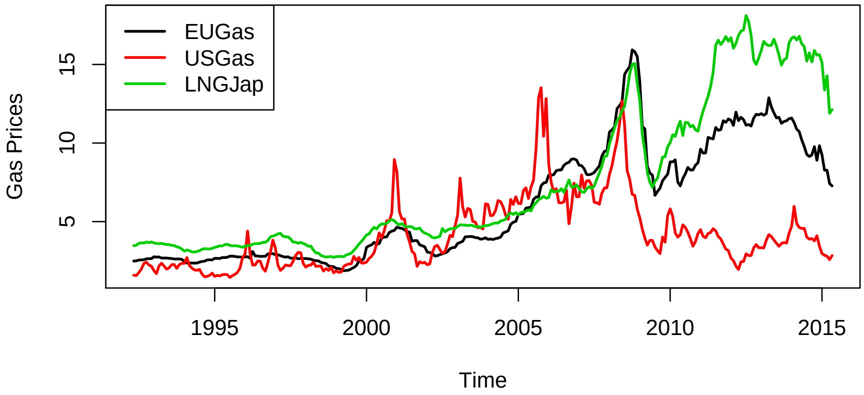


Figure 11: Evolution of Gas prices in US, EU and Japan

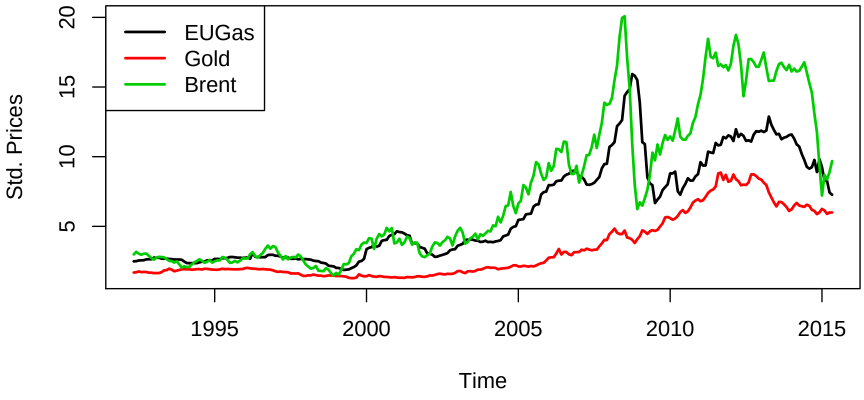


Figure 12: Evolution of Gas prices EU compared with Brent

The decomposition of a vector time series is similar to the univariate case and the ideas can be easily translated.

3.1 Vector ARMA Models

A series $\{\mathbf{Z}_t\}$ is a VARMA(p,q), if

$$\begin{aligned} \Phi(B)\left(\mathbf{Z}_t - \vec{\mu}\right) &= \Theta(B)\vec{\epsilon}_t \\ \Phi(B) &= \mathbf{I} - \Phi_1 B - \Phi_2 B^2 - \ldots - \Phi_p B^p \\ \Theta(B) &= \mathbf{I} - \Theta_1 B - \Theta_2 B^2 - \ldots - \Theta_q B^q \end{aligned}$$

where Φ_i, Θ_i are $k \times k$ matrices, $\vec{\epsilon}_t$ is a vector white noise ($\mathbb{E}\left[\vec{\epsilon}_t\right] = \vec{0}$, $\mathbb{E}\left[\vec{\epsilon}_t \vec{\epsilon}_t{}'\right] = \Sigma$) and B is the backward operator.
A VARMA model can be represented in the following alternate ways:

- Let $\Sigma = \mathbb{E}\left[\epsilon_t \epsilon_t'\right]$ and $\Phi_0^{\#}$ a lower triangular matrix with ones in the diagonal such that $\Phi_0^{\#} \Sigma \Phi_0^{\#'} = \Sigma^{\#}$ is a diagonal matrix with positive diagonal elements. Then,
$$\Phi^{\#}(B)\left(\mathbf{Z}_t - \vec{\mu}\right) = \Theta^{\#}(B)\vec{a}_t$$
with $\Phi^{\#}(B) = \Phi_0^{\#}\Phi(B)$, $\Theta^{\#}(B) = \Phi_0^{\#}\Theta(B)$ and $\vec{a}_t = \Phi_0^{\#}\epsilon_t$ having a diagonal covariance matrix, i.e. $\Sigma^{\#} = \Phi_0^{\#}\Sigma\Phi_0^{\#'}$
- Let $\Theta(B)^{-1} = \{1/\det\left(\Theta(B)\right)\}\,\Theta^*(B)$ where $\Theta^*(B)$ is the adjoint of the matrix $\Theta(B)$. Then,
$$\Theta^*(B)\Phi(B)\left(\mathbf{Z}_t - \vec{\mu}\right) = \Pi(B)\left(\mathbf{Z}_t - \vec{\mu}\right) = \det\left(\Theta(B)\right)\epsilon_t$$
- ARMAX models or transfers functions could also be represented in this way

3.2 VAR, VECM and Cointegration

Being $\mathbf{D}_t$ deterministic terms (trend, seasonal, constant, ...) we can define the following multivariate models:

- *VAR(p)*: $\mathbf{Z}_t = \mathbf{C}\mathbf{D}_t + \Phi_1\mathbf{Z}_{t-1} + \cdots + \Phi_p\mathbf{Z}_{t-p} + \vec{\epsilon}_t$
- *Structural VAR(p)*: $\mathbf{A}\mathbf{Z}_t = \mathbf{C}^*\mathbf{D}_t + \Phi_1^*\mathbf{Z}_{t-1} + \cdots + \Phi_p^*\mathbf{Z}_{t-p} + \mathbf{B}\vec{\epsilon}_t$
- *VECM*: $\Delta\mathbf{Z}_t = \mathbf{C}\mathbf{D}_t + \boldsymbol{\Phi}\mathbf{Z}_{t-1} + \Gamma_1\Delta\mathbf{Z}_{t-1} + \cdots + \Gamma_{p-1}\Delta\mathbf{Z}_{t-p+1} + \vec{\epsilon}_t$
where $\boldsymbol{\Phi} = \sum_{i=1}^{p}\Phi_i - \mathbf{I}$ and $\Gamma_i = -\sum_{j=i+1}^{p}\Phi_j, i = 1, \ldots, p-1$
- *Structural VECM*: Change $\vec{\epsilon}_t$ by $\mathbf{B}\vec{\epsilon}_t$

The multivariate nature of the said now leads to the more interesting definition of *cointegration* which falls between the concepts of stationarity and non-stationarity.
A series $\mathbf{Z}_t$ is said *cointegrated* if $\mathbf{Z}_t \sim I(1)$ (non–stationary) but there exist a vector $\boldsymbol{\beta}$ such that $\boldsymbol{\beta}'\mathbf{Z}_t \sim I(0)$ (stationary)
Given the preceding ideas, we can establish the relations Relations between VAR, VECM and Cointegration in the following way:

- $\mathbf{Z}_t \sim I(0) \Rightarrow rank(\mathbf{\Phi}) = k$
- $\mathbf{Z}_t \sim I(1)$ and $rank(\mathbf{\Phi}) = 0 \Rightarrow \mathbf{Z}_t$ is a VAR$(p-1)$ in differences and not cointegrated
- $\mathbf{Z}_t \sim I(1)$ and $0 < rank(\mathbf{\Phi}) = r < k \Rightarrow \mathbf{Z}_t$ has r linearly independent cointegrating relations. Then $\mathbf{\Phi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$. The columns of $\boldsymbol{\beta}$ form the basis of r cointegrations.

In this case, the three series show a non-stationary evolution but, an important issue now is to determine the cointegration relations between covariates (if any). For that purpose, the tool is the Johansen's test ([4])
The Johansen's test is based on likelihood ratio test for r, the rank of $\mathbf{\Phi}$(= number of non-zero eigenvalues of $\mathbf{\Phi}$) and it is formulated as:
$H_0(r) : r = r_0$ vs. $H_1(r_0) = r > r_0$

$$LR(r_0) = -n \sum_{i=r_0+1}^{k} \log\left(1 - \hat{\lambda}_i\right)$$

with $\hat{\lambda}_i$ the estimated eigenvalues of $\mathbf{\Phi}$
Johansen proposes also a sequential testing for r.

1. Set $r_0 = 0$
2. Test $H_0(r_0)$ against $H_1(r > r_0)$
3. If null is rejected $r_0 = r_0 + 1$ & go to 2
4. The number of cointegrated vectors is r_0

The application of this methodology to the data is done using a quite large $K = 5$ to ensure to catch the effect of the price of the oil in the future of EUGas.

```
>
> ######################
> # Johansen-Procedure #
> ######################
>
> Test type: trace statistic , without linear trend and constant in cointegration
>
>           test 10pct  5pct  1pct
> r <= 2 |  1.71  7.52  9.24 12.97
> r <= 1 |  8.31 17.85 19.96 24.60
> r = 0  | 46.33 32.00 34.91 41.07
>
> Eigenvectors, normalised to first column:
> (These are the cointegration relations)
>
>              EUGas.l1     Gold.l1     Brent.l1    constant
> EUGas.l1  1.000000000  1.00000000   1.00000000  1.00000000
> Gold.l1   0.004335406 -0.01023207   0.01663152 -0.03717146
> Brent.l1 -0.154225814  0.06564913   0.20967740  0.13545791
> constant -1.256916940 -1.78263092 -58.13868347  6.87773871
>
....
```

The Johansen-Procedure confirms the presence of one cointegration relation (the tests for $r \leq 2$ and $r \leq 1$ are accepted and the test for $r = 0$ is rejected and so, we must conclude that $r = 1$) which is $EUGas = 1, Gold = 0.004, Brent = -0.154, Intercept = -1.257$. This means that this combination is stationary and we can use it for prediction.
The parameters associated with the model are reflected in the next matrices. Every row represents the response variate and the matrices are the corresponding $\mathbf{\Phi}_i$ shown above. With this information, the model for the prediction of EUGas can be reconstructed as a function of the other assets.

```
>
> Coefficient matrix of lagged endogenous variables:
>
> A1:
>         EUGas.l1      Gold.l1    Brent.l1
> EUGas  0.7694248 0.0005050132 0.002521418
> Gold  -4.9255653 1.1528051959 1.058761925
> Brent -0.1127406 0.0066571702 1.318815152
>
>
> A2:
>        EUGas.l2       Gold.l2    Brent.l2
> EUGas 0.2157127 -0.0006556704  0.01498292
> Gold  3.9683535 -0.1927169037 -0.97914696
> Brent 2.1275974 -0.0015168916 -0.30377273
>
>
> A3:
>         EUGas.l3      Gold.l3     Brent.l3
> EUGas  0.2498603 0.0008091053  0.002710852
> Gold   1.6089356 0.1484702768 -1.354443128
> Brent -1.5127809 0.0079013118 -0.086997224
>
>
> A4:
>         EUGas.l4       Gold.l4    Brent.l4
> EUGas -0.4682386 -0.0006198872  0.01041780
> Gold   6.4144558 -0.1612003328  1.96392071
> Brent -0.6935846  0.0052354179 -0.03564787
>
>
> A5:
>         EUGas.l5       Gold.l5    Brent.l5
> EUGas  0.1194300 -0.0005319766 -0.01308044
> Gold  -7.8924496  0.0490595475 -0.56166037
> Brent  0.9768782 -0.0148721125 -0.01352158
>
>
> Coefficient matrix of deterministic regressor(s).
>
>         constant
> EUGas  0.1430506
> Gold   1.0385528
> Brent -0.9871442
```

The forecasting for 2015, 2016 and 2017 is shown in Figure 13 in the format of a Fan Chart. This graph shows in dark gray the mean and the core of the prediction interval and in lighter gray, the area of the prediction interval that corresponds with high and low quantiles.
The forecasts for the three variates show an almost constant prediction for the next three years with values for EUGas around 7.2, around 1184 for Gold and

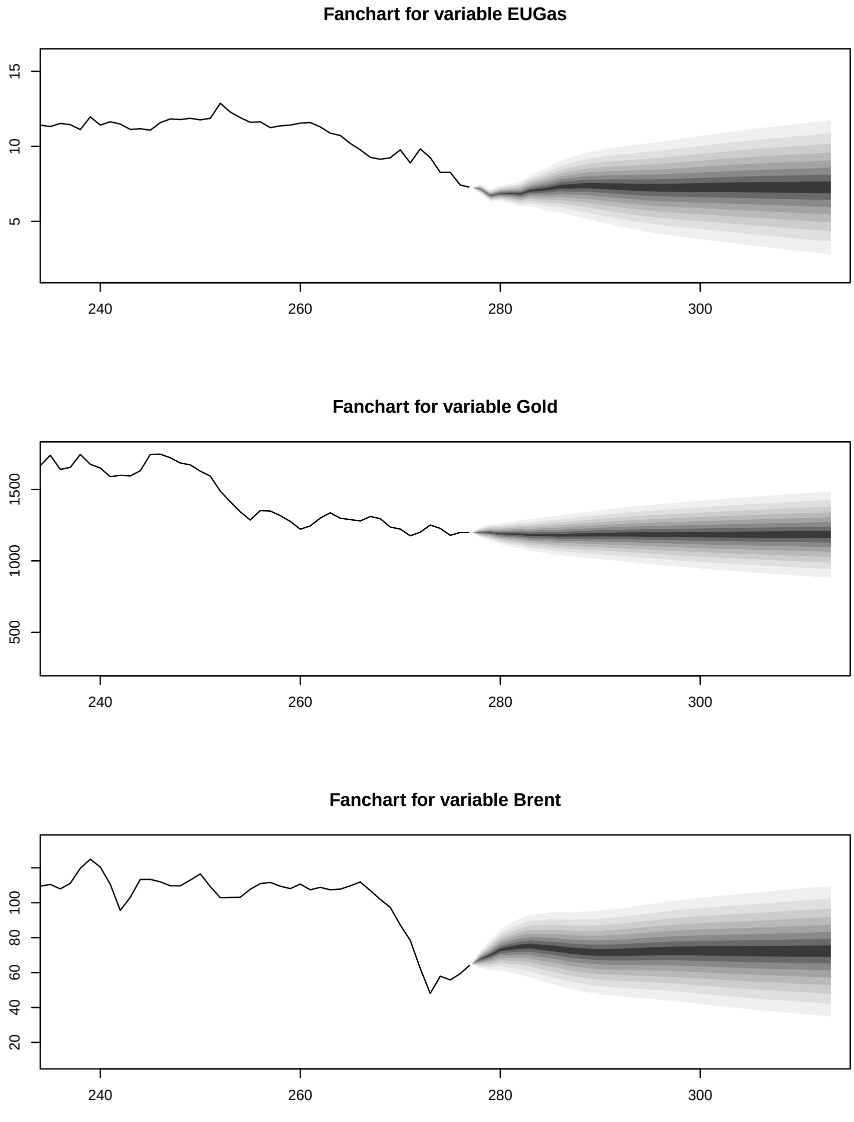


Figure 13: Forecasting for EUGas, Gold and Brent up to the end of 2017

around 72 dollars for Brent.

4 Conclusion

This study group finishes only showing a couple of possibilities for predicting the variables associated with GNL: Demand and price. The analysis is pretty classic using the usual tools from Time Series Methodology with the following characteristics:

- The availability of information is far for being homogeneous and although there are several institutions that provides future scenarios and plots, rarely they provide the original data or the source. So, the information can be found broken into pieces: the available time series are excessively short or don't have the same temporal aggregation or intervals (for instance annual covariates for a monthly one) and in many cases is not available for public knowledge.
- The forecasts can be improved including new informative variables but, as the direct variables are not available, it is possible to find out new ones indirectly related with them. For instance, the temperature in Madrid–Barajas is an indirect information about winter and the use of heating systems.
- Some other models can be incorporated depending on the goals (GLS, Dynamic Regression, Kalman Filter, Non Linear Time Series, ...). The use of one or another depends on the amount of available information. Many times a complex methodology performs poorly than a simple one just only because the quality of the available information cannot be exploited successfully by the complex one.
- Finally, the type of application would be also interesting. This report is a retrospective study but it can be improved, for instance, updating the models and the forecasts every month. Also this procedure can be easily automatized through a web page using, for instance, the Shiny apps. During this study group, a shiny apps was created by Manuel Oviedo de la Fuente to show the potentialities of the analysis that can be done. The app is available following the next link `https://manueloviedo.shinyapps.io/markdown/aa.Rmd` or directly by contacting the author.

References

[1] George EP Box and Gwilym M Jenkins. *Time series analysis: forecasting and control, revised ed.* Holden-Day, 1976.

[2] George EP Box, Gwilym M Jenkins, and Gregory C Reinsel. *Time series analysis: forecasting and control*, volume 734. John Wiley & Sons, 2011.

[3] James Douglas Hamilton. *Time series analysis*, volume 2. Princeton university press Princeton, 1994.

[4] Soren Johansen. Likelihood-based inference in cointegrated vector autoregressive models. *OUP Catalogue*, 1995.

Acknowledgements

The Scientific Committee wishes to thank to the company speakers, the academic coordinators and the researchers of each working team for their invaluable contributions to the scientific success of the 110A European Study Group with Industry.

This 110A ESGI was framed as activity in the Joint Research Unit ITMATI-Repsol which has funding of the Galician Agency for Innovation and the Ministry of Economy and Finance in the framework of the Spanish Strategy Innovation in Galicia. We would also like to thank to the Faculty of Mathematics of the University of Santiago de Compostela which held the 100A ESGI.

Finally, we would also like to express our gratitude to Adriana Castro Novo, Technology and innovation transfer at ITMATI, Rubén Gayoso Taboada, manager of ITMATI, Peregrina Quintela Estévez, director of ITMATI and president of Math-in and Fe Sampayo Fernández, technology translator of Math-in, whose meticulous work and dedication have contributed to the success of this 110A European Study Group with Industry.

INDEX

Universidade Vigo

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

UNIVERSIDADE DA CORUÑA