Forecasting models for short and long term gas price

A Data Science point of view

2nd AIEE Energy Symposium, Rome, Italy

November 3, 2017

Introduction

Macroeconomic, Energy Analysis and Forecasting (MEAF)

Administration Finance Control of ENEL S.p.A.

Pregagnoli Claudio (main author), Troise Raffaele (co-author and presenter), Dicembrino Claudio (co-author) and Frate Michele (co-author)

Outline

Introduction:

- Brief description of the **gas markets** modeled
- **References** about the forecasting models of the natural gas prices: **new methods to be considered ?**

Methodology:

- **Method**: essential features of the model
- **Strengths and weakness** of the model
- **Results**: model output and forecasting KPIs

Models:

- Vector Auto Regressive **(VAR)**
- Neural Network **(NN)**
- Generalized Additive Model **(GAM)**
- Gradient Boosting Model **(GBM)**

Deployment and further developments:

- **Model deployment** into a Big Data architecture, namely a **Big Data point of view**
- **Further models** to be considered for the gas price forecasting

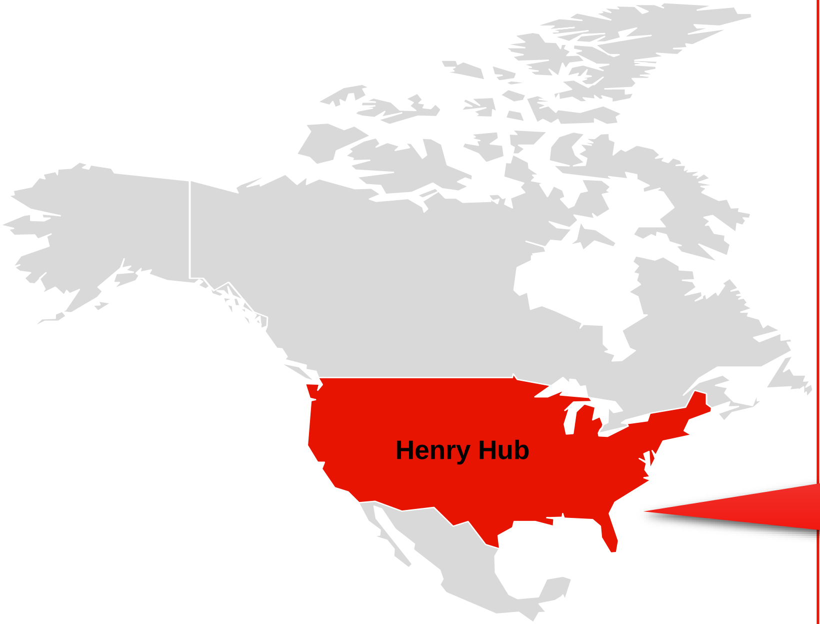
Introduction

- A brief description of the gas markets
- A review of the forecasting models for the gas prices

Introduction

A gas markets brief description (a global point of view)

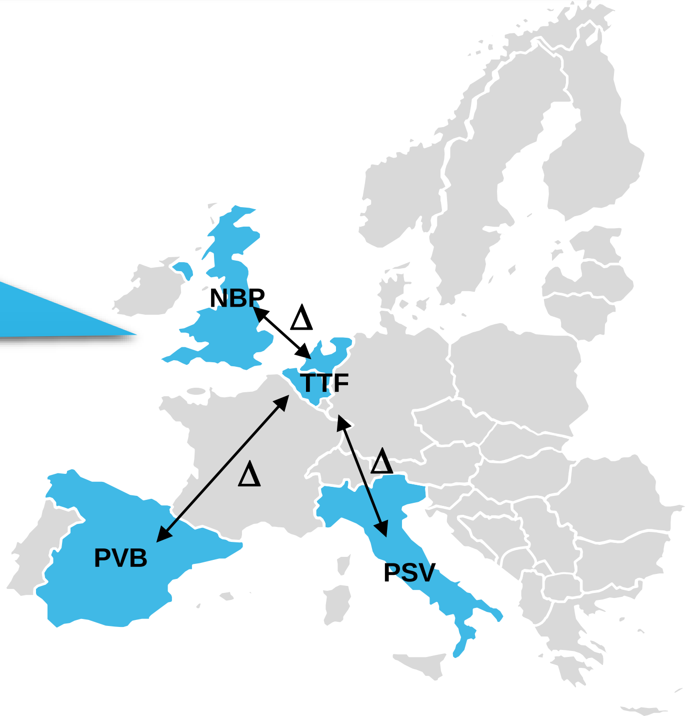
The Henry Hub US market



Henry Hub
- ✓ Average volume 30 day: 1,200 Mmbtu
- ✓ Open interest: 277.600

The NBP EU market

NBP
- ✓ Average volume 30 day: 457 Mil. therms
- ✓ Open interest: 63.800
- ✓ Historical Average Spread:
  - NBP-TTF: 0,2 €/MWh

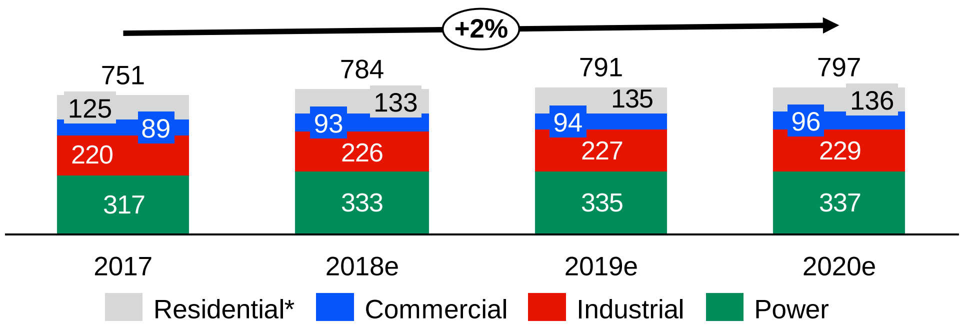


Introduction

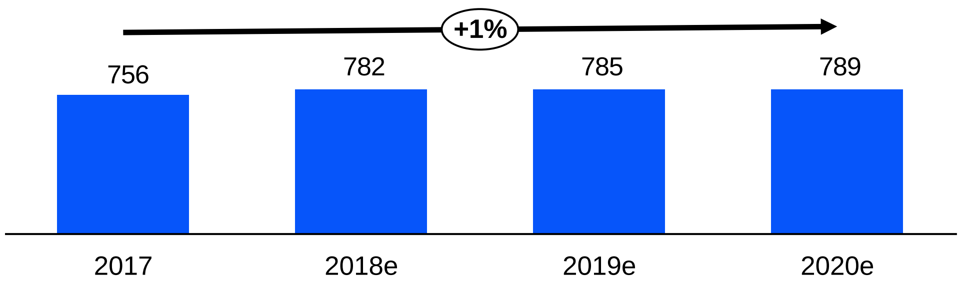
A gas markets brief description (fundamentals analysis)

The US gas fundamentals

Demand (bcm)

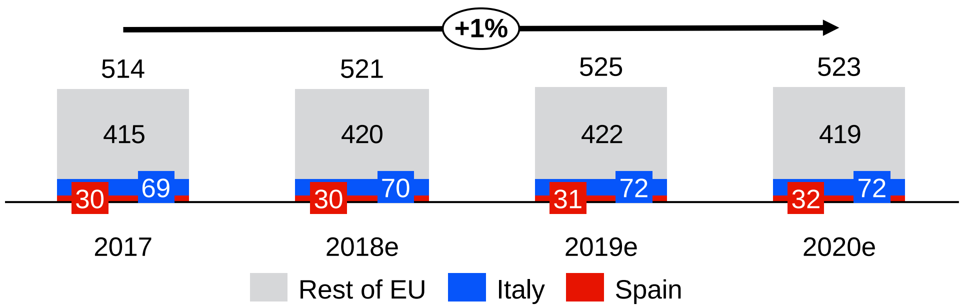


Supply (bcm)

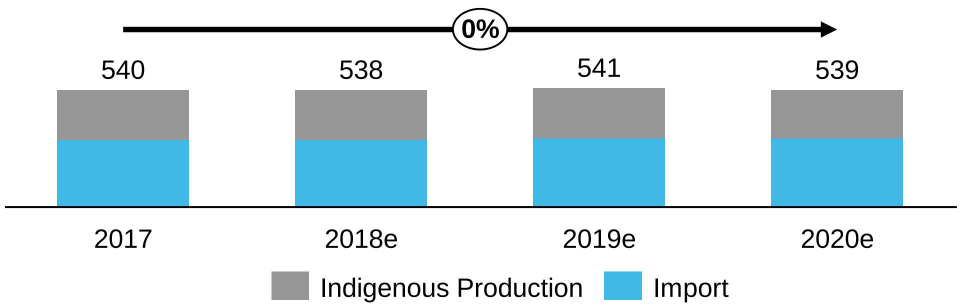


The EU gas fundamentals

Demand (bcm)



Supply (bcm)



Introduction

References about the forecasting models of the natural gas prices

Some references

linear models have poor performance in prediction due to the complex behavior of the system which, could be recognized as nonlinearity (Agbon & Araque, 2003).

Two forecasting models have been developed by Nogales, et al. (2002) to predict the daily price of natural gas. They have used the time series analysis approach to establish dynamic regression

Doris (1999) [1] provided a methodology for prediction of short-term natural gas prices using econometrics and neural network. According to the author, developing a

implemented time series models to predict natural gas prices. Reiter and Economides (1999) applied both multivariate econometrics and neural network models to forecast the natural gas price in the short term.

Complex systems

Econometric models

Time series analysis

Artificial Intelligence algorithms

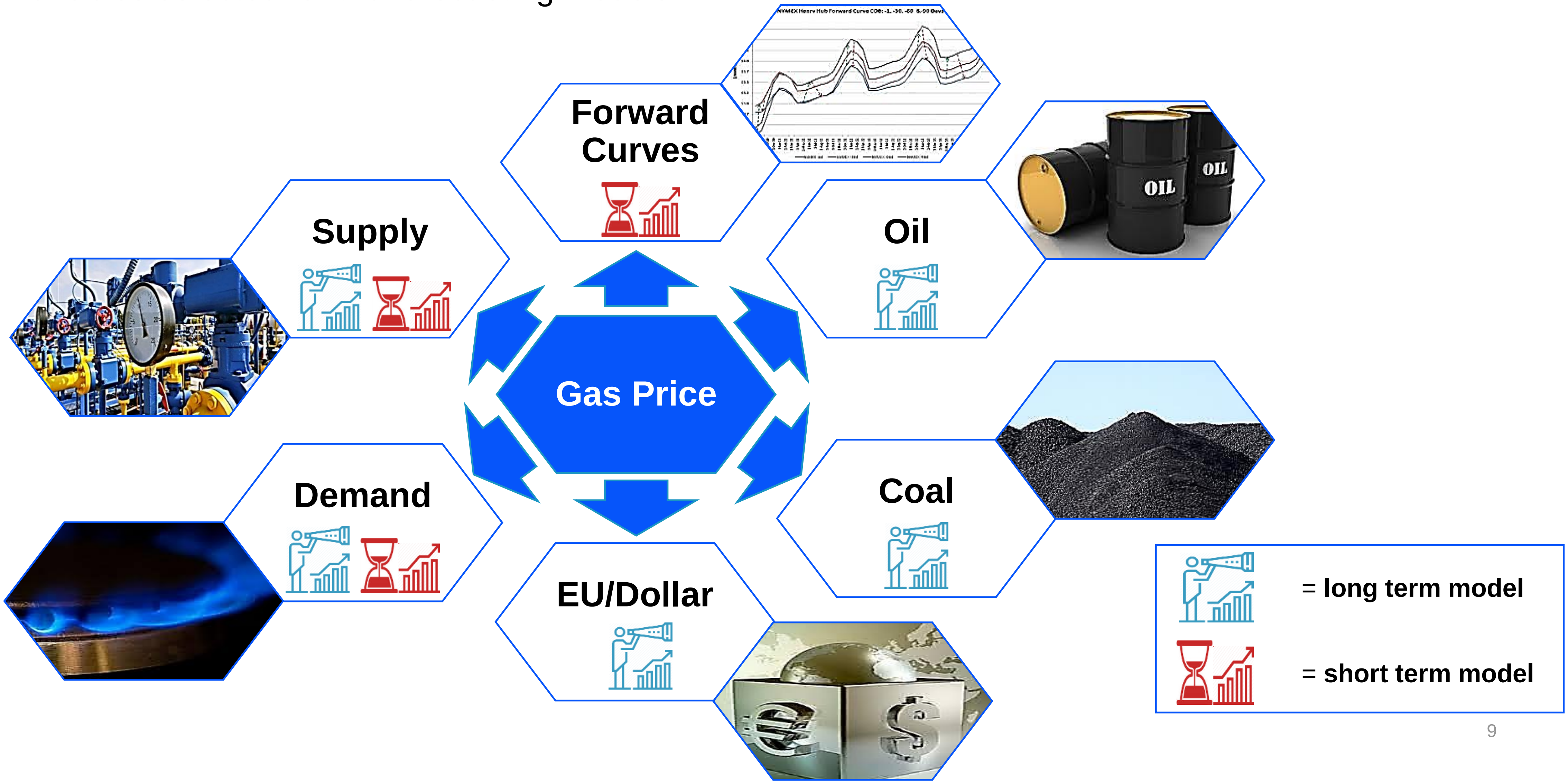
Introduction

The general features of the models proposed

Fundamental models	• Supply and demand determining the price dynamics • The impact of important physical and economic factors on the market equilibrium price
External variables	• The exogenous inputs of the models have to be predicted together with the gas price
Forecast horizon	• Short term of about 15 month ahead • Long term of about 20 years ahead

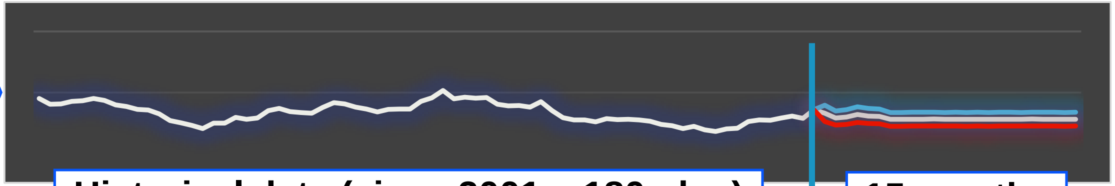
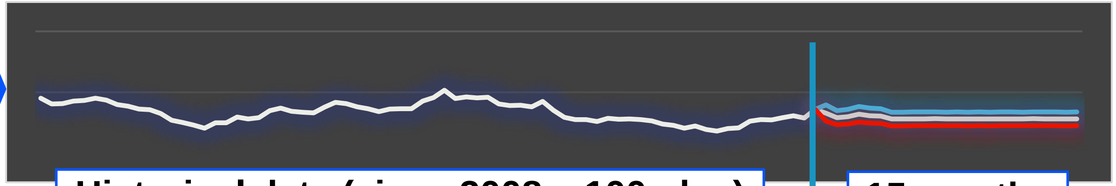
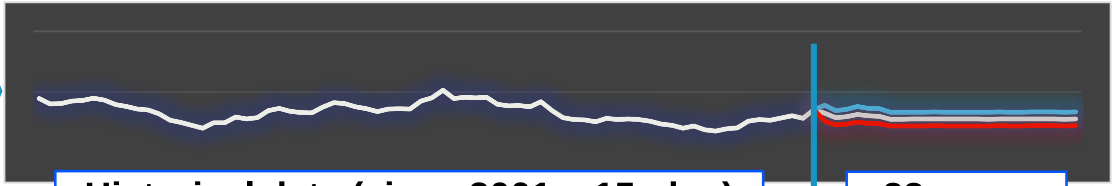
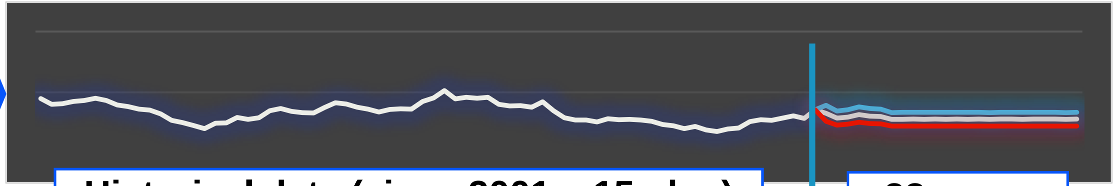
Introduction

The variables selected for the forecasting models



Introduction

Historical data and the forecast horizons

Gas price	Model	Time Series Granularity	Historical data and forecast horizon
Henry Hub	Short Term	Monthly	 Historical data (since 2001 ~ 180 obs) / 15 months
NBP	Short Term	Monthly	 Historical data (since 2008 ~ 100 obs) / 15 months
Henry Hub	Long Term	Yearly	 Historical data (since 2001 ~ 15 obs) / 23 years
NBP	Long Term	Yearly	 Historical data (since 2001 ~ 15 obs) / 23 years

Methodology

- Essential features of the models and their strengths/weakness
- Results of the forecasting models

Methodology

Essential features and strengths/weakness: **VAR** models

Features

- Vector Auto Regressive model (**VAR**) selected to take into account the **co-integration** between the variables: gas price, demand and supply

- **Auto Regressive** part of the price dynamics is taken into account → **short term** models

- **Linear equations** of the model:

$$Y_t = \Pi_1 Y_{t-1} + \Pi_2 Y_{t-2} + \cdots + \Pi_p Y_{t-p} + \Phi D_t + G X_t + \varepsilon_t$$

with the **D**'s variables as **dummies**, the **X** variables as **exogenous variables** (i.e. the forward curves)

Strengths/weakness

- Relationships between the gas price and the input variables in terms of **simple coefficients**

- **Statistical tests** to be evaluated in order to validate the model

- **Non-linear effects** can not be effectively realized

Variables

To take into account the **monthly demand/supply equilibrium** related to the gas price trend:

- Gas demand
- Gas supply

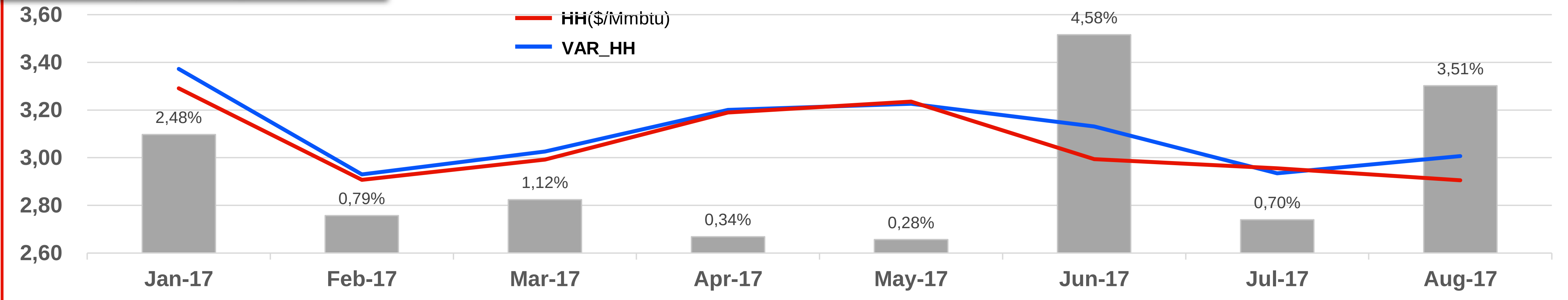
The selected exogenous variables are the monthly **forward curves of the gas price**:

- maturity up to **36** months for **HH**
- maturity up to **10** months for **NBP**

Methodology

Results: the **short term** model for the **HH** gas prices

The Henry Hub gas price



Co-integration factor	Test statistic	10pct	5pct	1pct
r<=2	19,72	10,49	12,25	16,26
r<=1	53,26	16,85	18,96	23,65
r=0	316,20	23,11	25,54	30,34

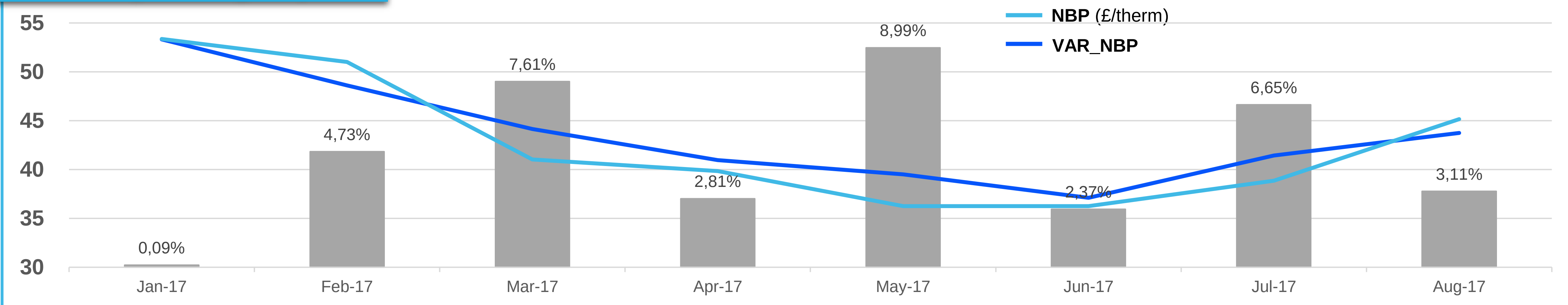
Test	AR(1)	AR(2)	AR(3)
AIC	29,11	29,10	29,13

Curve	Model	MAPE
VAR_HH	Vector Auto Regressive Model	1,73%

Methodology

Results: the **short term** model for the **NBP** gas prices

The Henry Hub gas price



Co-integration factor	Test statistic	10pct	5pct	1pct
r<=2	11,60	10,49	12,25	16,26
r<=1	35,26	16,85	18,96	23,65
r=0	121,52	23,11	25,54	30,34

Test	AR(1)	AR(2)	AR(3)
AIC	32,4	31,5	32,5

Curve	Model	MAPE
VAR_NBP	Vector Auto Regressive Model	4,54%

Methodology

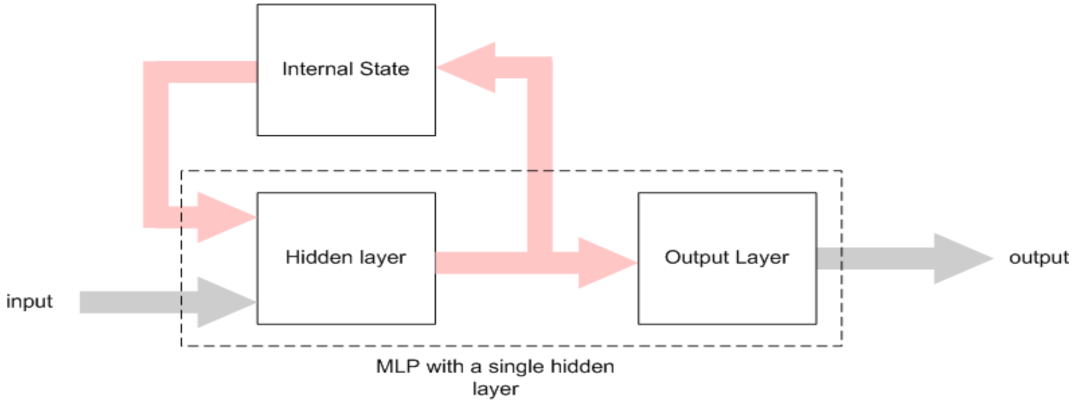
Essential features and strengths/weakness: **Neural Network** models

Features

- **Non linear effects** given by the evaluation of a **complex** function of the input variables:

$$Y=F(X_1, X_2,\ldots, X_3)$$

- The F- function given by **the minimization** of the **RMSE**
- The Neural Network used for the gas price forecast has an **Elmann recurrent** topology



Strengths/weakness

- **Non linear effects** that can be realized in terms of complex functions
- **High number of input variables** can be managed in the model
- **Black-box** modelling

Variables

To take into account the **yearly demand/supply equilibrium**:

- Gas demand
- Gas supply

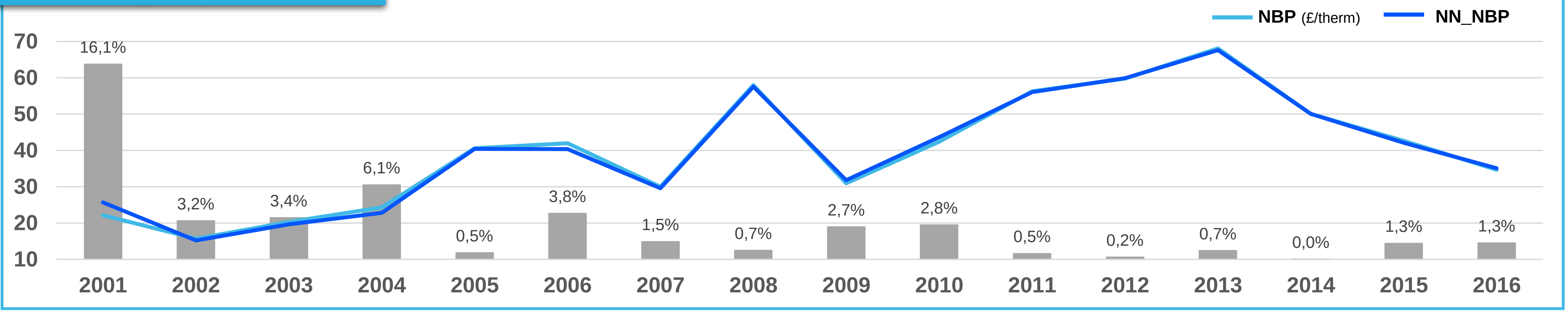
Exogenous variables related to the gas prices:

- Brent oil price for **NBP** and WTI oil price for **HH**
- Coal price
- Euro/Dollar FX

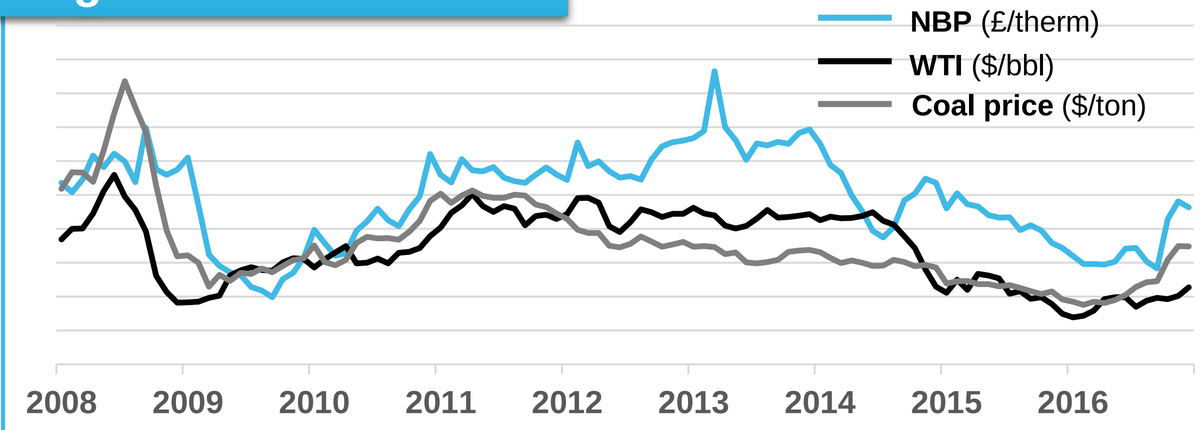
Methodology

Results: the **long term** model for the **NBP** gas prices

The NBP gas price



Exogenous variables



Curve	Model	MAPE
NN_NBP	Neural Network	2,8%

Methodology

Essential features and strengths/weakness: **GBM** models

Features

- **Non linear effects** given by the evaluation of a **complex** function of the input variables:

$$Y=F(X_1, X_2,\dots, X_3)$$

- The starting F- function is expressed in terms of a set of **complex basis-functions**
- The F-function is calculated by the minimization of a loss function by **boosting** the **gradient descendent algorithm**:

Init $F_0(x) = a_0 = argmin_\gamma L(\gamma, \gamma, \dots, \gamma)$

For $m = 1$ to M:

– Compute negative gradient:

$$[r_i] = -\left[\frac{\partial L(F_{m-1}(x_i))}{\partial F_{m-1}(x_i)}\right] \; i = 1..N$$

– Feed a base learner the training data $\{(x_i, r_i)\}_{i=1}^{N}$ to get a base function $h_m(x)$

– Compute the multiplier with line search strategy:

$$a_m = argmin_\gamma L(F_{m-1}(x) + \gamma h_m(x))$$

– Update: $F_m(x) = F_{m-1}(x) + a_m h_m(x)$

Result is $F_M(x)$

Strengths/weakness

- **Non linear effects** that can be realized in terms of complex functions
- **A high number of input variables** can be managed in the model
- **A high number of parameters** to be set in the definition of the model

Variables

To take into account the **yearly demand/supply equilibrium**:

- Gas demand
- Gas supply

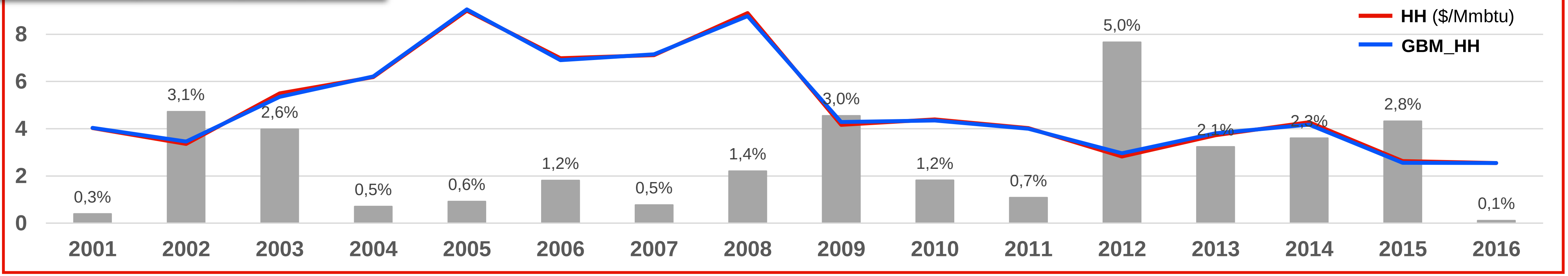
Exogenous variables related to the gas prices:

- Brent oil price for **NBP** and WTI oil price for **HH**
- Coal price
- Euro/Dollar FX

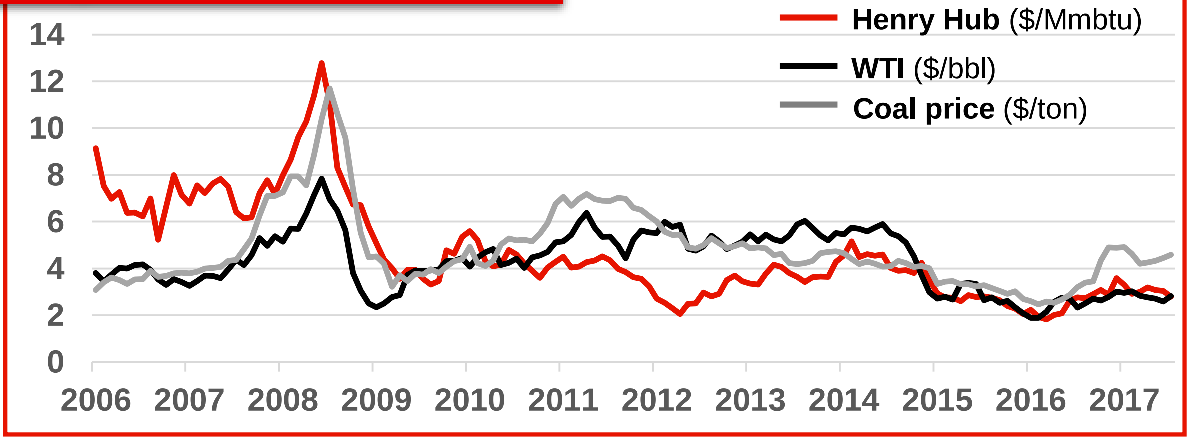
Methodology

Results: the **long term** model for the **HH** gas prices

The Henry Hub gas price



Exogenous variables



Curve	Model	MAPE
GBM_HH	Gradient Boosting Models	1,7%

Methodology

Essential features and strengths/weakness: **GAM** models

Features

- **Non linear effects** given by the transformation of the input variables
- The starting F- function is expressed in terms of a set of **complex basis-functions** of the input:

$$\mathbf{f(x_t)=\sum_i b_i(xt)\beta_i}$$

where the x's are the input variables and the functions f's are "smooth functions" of the input variables

- The b's functions in the above equation are *s-plines* of the input variables
- The dependent variable Y to be modelled is expressed in the following form:

$$\mathbf{E(y_t)=\beta_0+f_1(x_{1t})+\cdots+f_p(x_{pt})}$$

Strengths/weakness

- **Non linear effects** that can be realized in terms of complex functions
- **A high number of input variables** can be managed in the model
- **A high number of parameters** to be set in the definition of the model

Variables

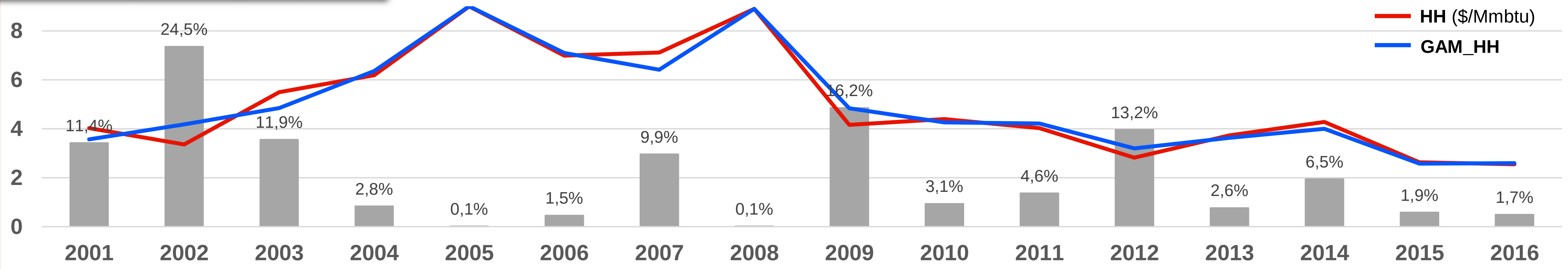
To take into account the **yearly demand/supply equilibrium**:
- Gas demand
- Gas supply

Exogenous variables related to the gas prices:
- Brent oil price for **NBP** and WTI oil price for **HH**
- Coal price
- Euro/Dollar FX

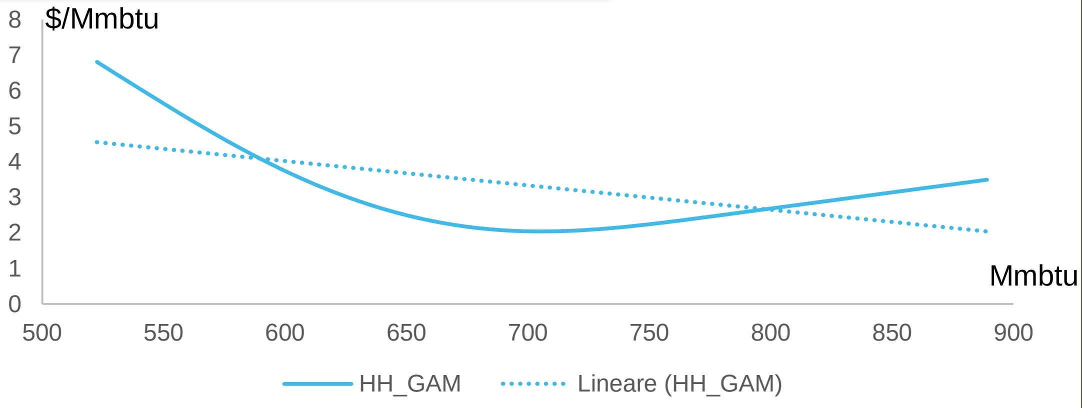
Methodology

Results: the **long term** model for the **HH** gas prices

The Henry Hub gas price



Non linearity effect (Supply)



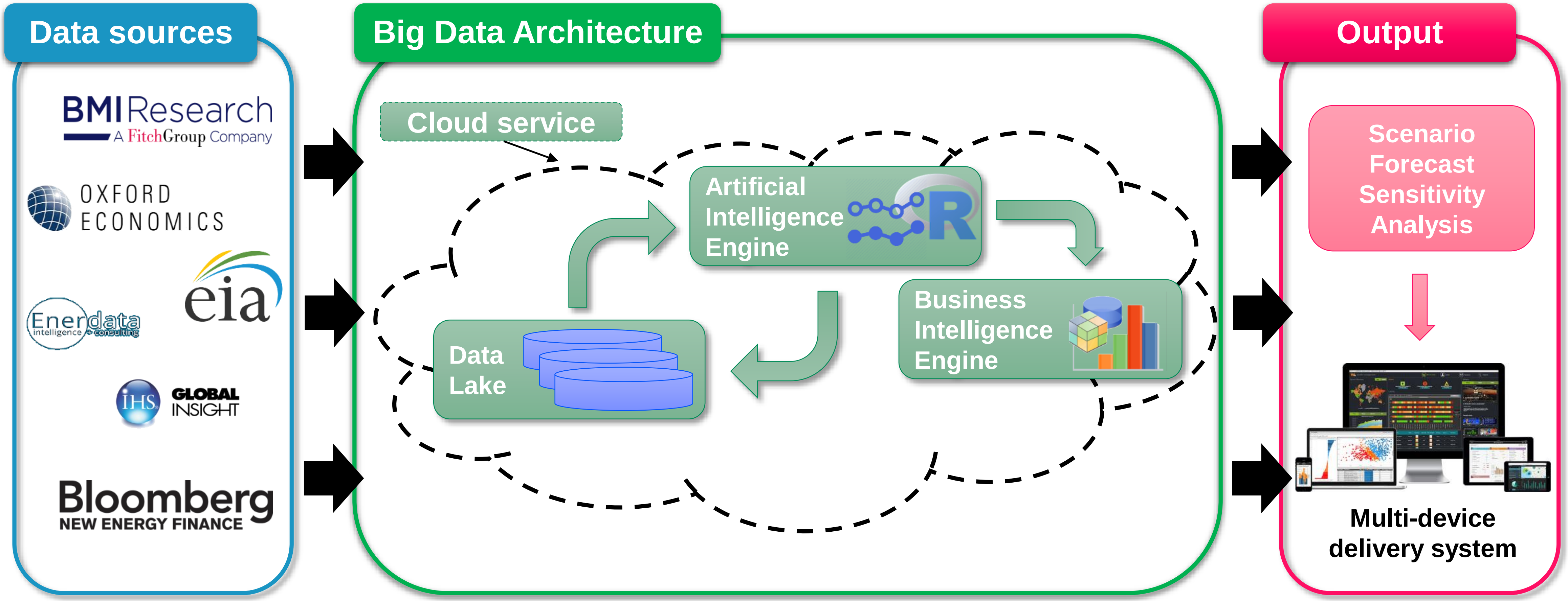
Curve	Model	MAPE
GAM_HH	General Additive Models	7,0%

Deployment and further developments

- Deployment in a Big Data architecture
- Further models

Results & Deployment

Deployment of the model outputs: the **digitalization project**



Further models

A regression tree model for the gas price forecasting

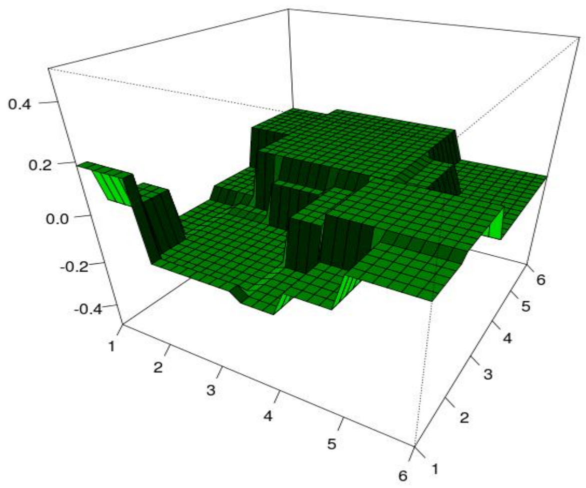
Features

- The predicted value as the result of the application of rules such as:

IF X > (or <) to one specific critical value

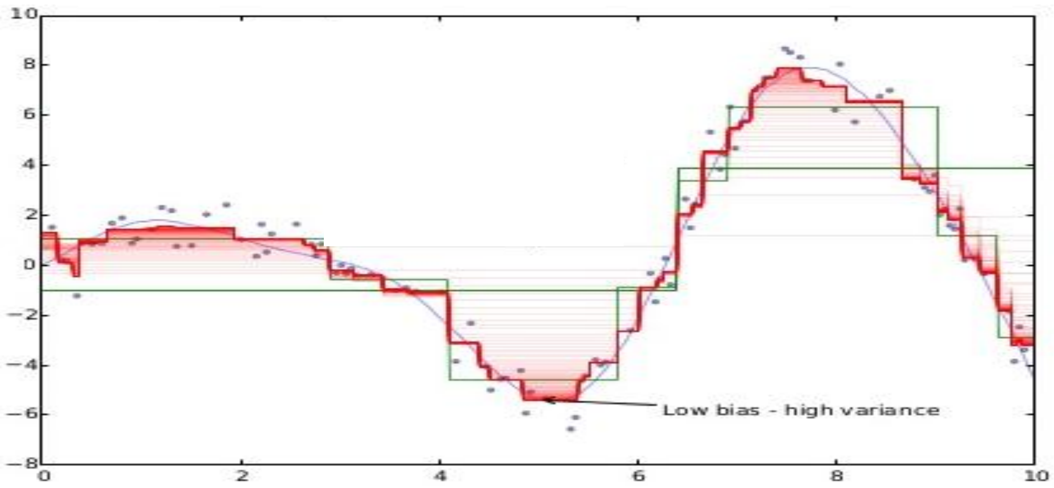
given by the calculation of the model

- **Discretization of the input variables space:**



Strengths/weakness

- **Non linear effects** are taken into account
- **A huge number of input variables** can be managed in the model
- With **more input variables** the accuracy of the regression tree model can be increased
- The form of the **predicted price curve** is as follow:



Conclusions

- Necessity to consider a wider class of models in addition to those of the classical statistical econometrics
- Complex dynamics realized by machine learning algorithms and non-linear models
- A data science approach allows us to manage a huge number of variables in order to consider all the effects that affects the prices
- Developing an effective tool able to reduce the forecasting error is key to build up a resilient and reliable energy market

Thank you