Models to predict client's phone calls to Iberdrola Call Center

VI Modelling Week UCM

PARTICIPANTS (UCM):
Cazallas Piqueras, Rosa
Gil Franco, Dolores M
Gouveia de Miranda, Vinicius
Herrera de la Cruz, Jorge
Inoñan Valdera, Danny Javier

Advisor: Ivorra, Benjamin

IBERDROLA

Outlook

1. Problem description
2. Data analysis
3. Modelling
4. Validation and prediction
5. Conclusions and perspectives

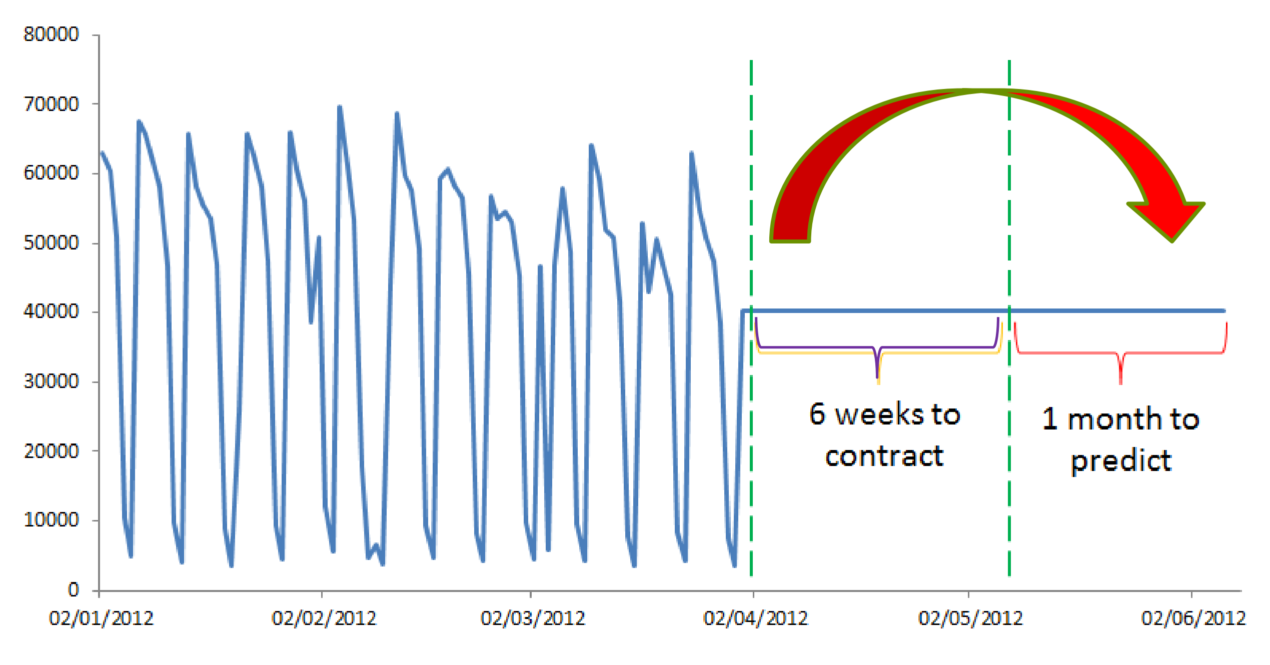
1. Problem description

Problem description

- Iberdrola is a company that provides electricity and/or gas to the customers.
- Customers can contact Iberdrola, for instance, by letters, internet and **Phone calls** **(Call Center)**.
- In a Call Center: to avoid **over-staff** and **under-staff** Iberdrola needs a good prediction about the number of operators.
- **Objective**: Find a model to predict the volume of customer calls.

Problem description

- Forecasting methodology: for operational reasons, **we need to predict the volume of calls for next month, six weeks in advance.**



2. Data analysis

Data analysis

Weekly data

- **Portfolio 1:** regulated market
- **Portfolio 2:** free market
- **Portfolio 3**: gas market (since 01/2007)
- **Bill type 1**: estimated consumption
- **Bill type 2**: real consumption
- **Bill type 3**: fixed quota without consumption
- **Bill type 4**: other (corrections, material, etc.)

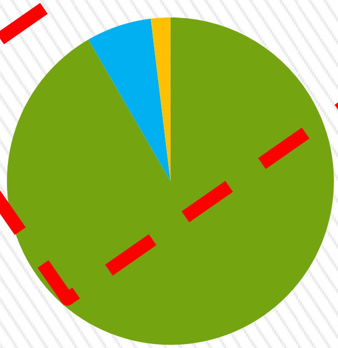
Transformed to daily data

Daily data

- **Calls:** number of calls → *target variable*

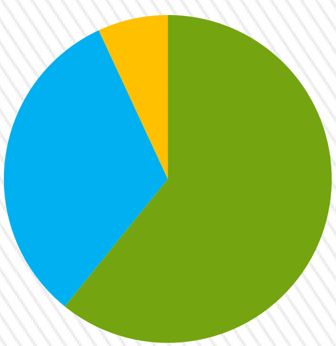
Data analysis

2009



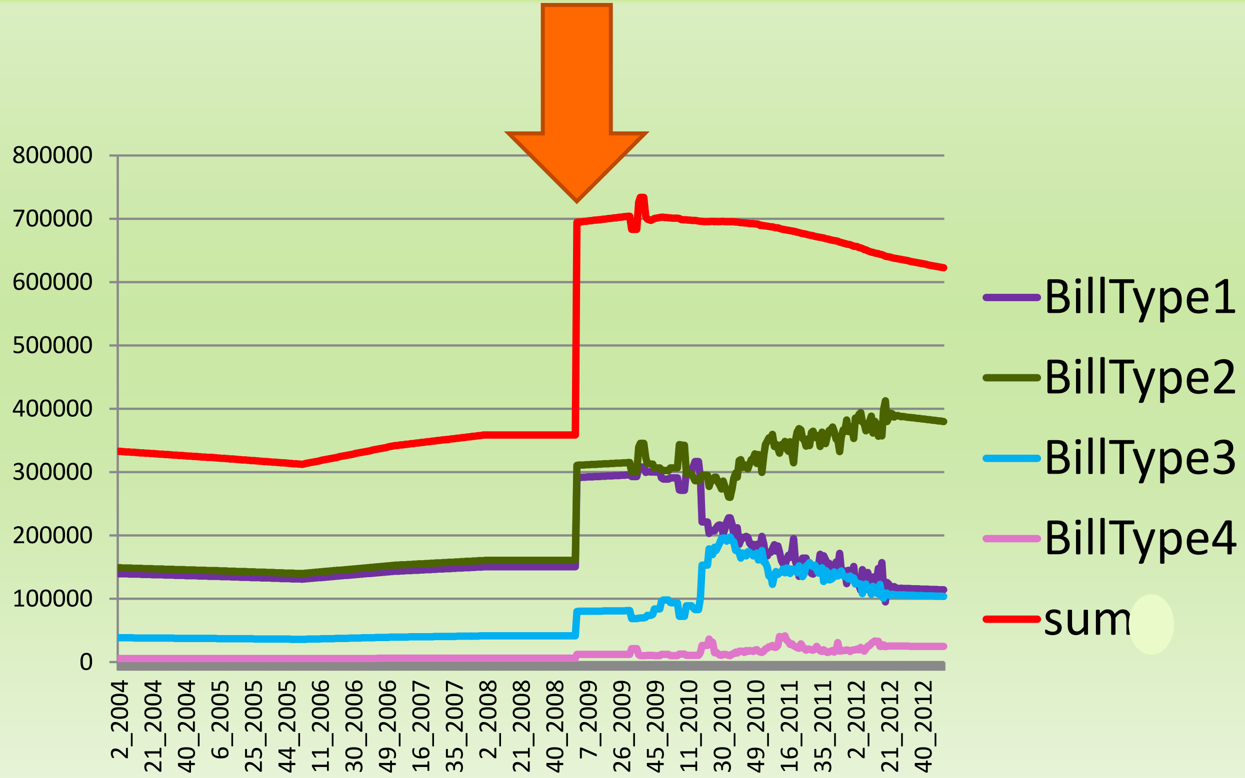
EVOLUTION

2012



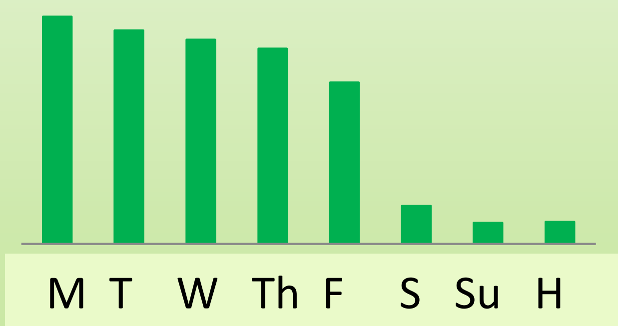
Data Analysis

2009: Change due to an European regulation (monthly bills instead of twice-monthly bills).

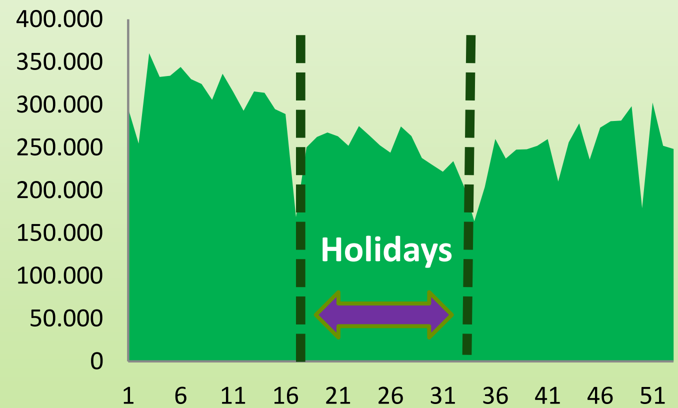


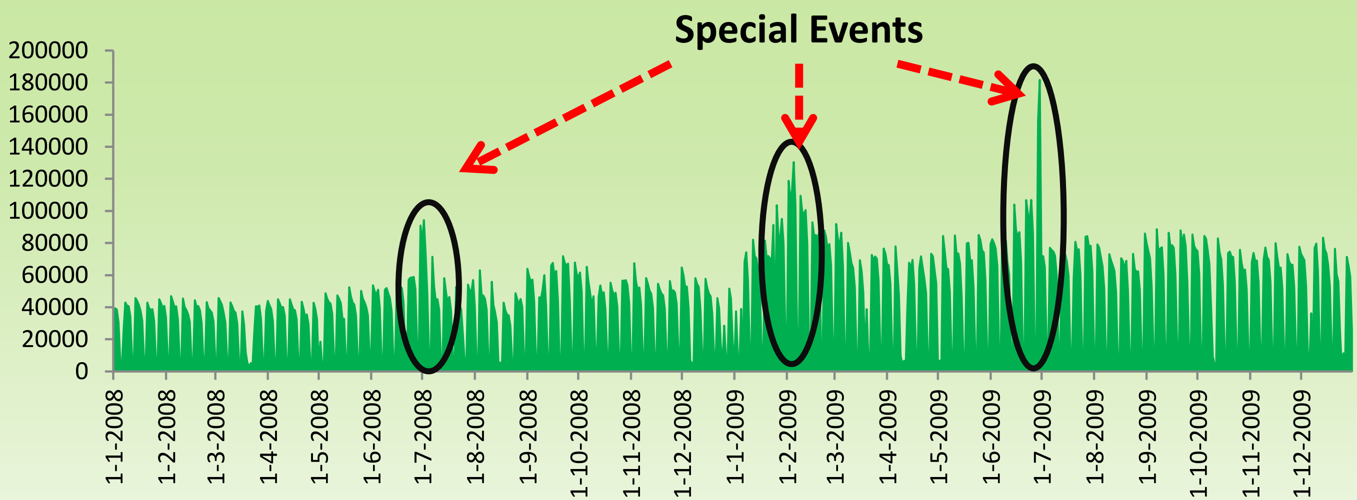
Data Analysis

Calls distribution during weeks and holidays



Calls seasonality pattern



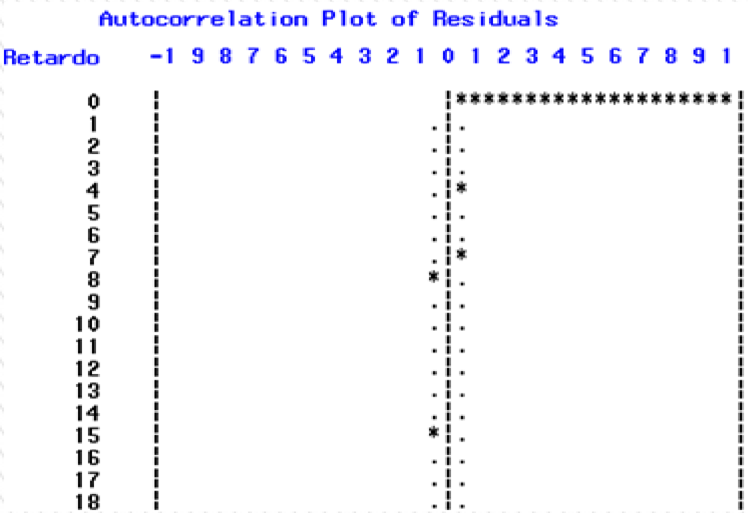


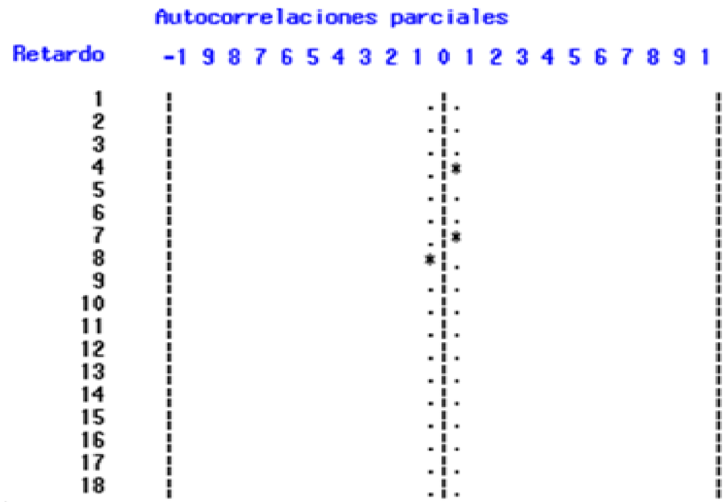
3. Modelling

SARIMA MODEL $(1,1,1)(0,1,1)_7$

$$\log(calls_{\,t}) = seasonal\ Inputs + \eta_{\,t}$$

$$\eta_t = \frac{(1-0{,}73L)(1-0{,}64L^7)\hat{\varepsilon}_t}{(1-0{,}38L)(1-L)(1-L^7)}$$



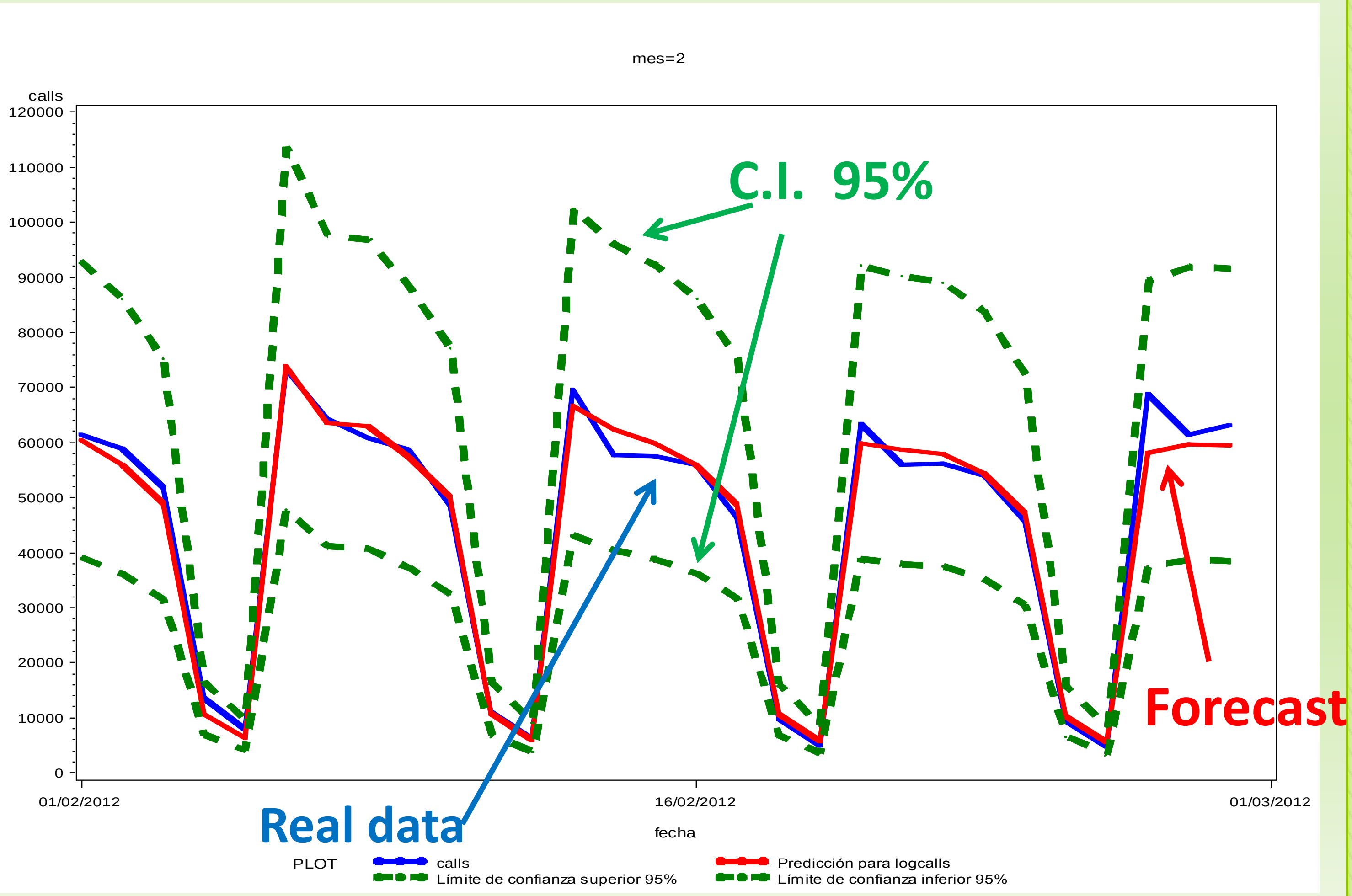


White noise

Para retardo	Chi-cuadrado	DF	Pr > ChiSq	--------------------Autocorrelaciones-------------------					
6	1.87	3	0.5989	-0.010	0.012	0.023	0.031	-0.021	0.003
12	5.48	9	0.7904	0.054	-0.029	0.000	0.004	-0.020	-0.005
18	6.52	15	0.9696	-0.002	-0.007	-0.027	-0.010	0.012	0.013
24	31.74	21	0.0622	-0.136	-0.060	-0.077	0.010	0.009	0.032
30	36.60	27	0.1027	-0.011	-0.023	0.031	-0.016	0.056	0.024
36	48.78	33	0.0377	0.025	0.010	-0.003	-0.060	0.096	-0.015
42	53.50	39	0.0609	0.008	-0.002	0.007	0.047	0.031	-0.046
48	57.04	45	0.1075	0.047	0.023	0.029	0.012	0.013	0.009
54	66.35	51	0.0730	0.031	0.003	0.047	0.043	0.049	-0.054
60	68.40	57	0.1434	0.032	-0.017	0.028	0.011	0.001	-0.005

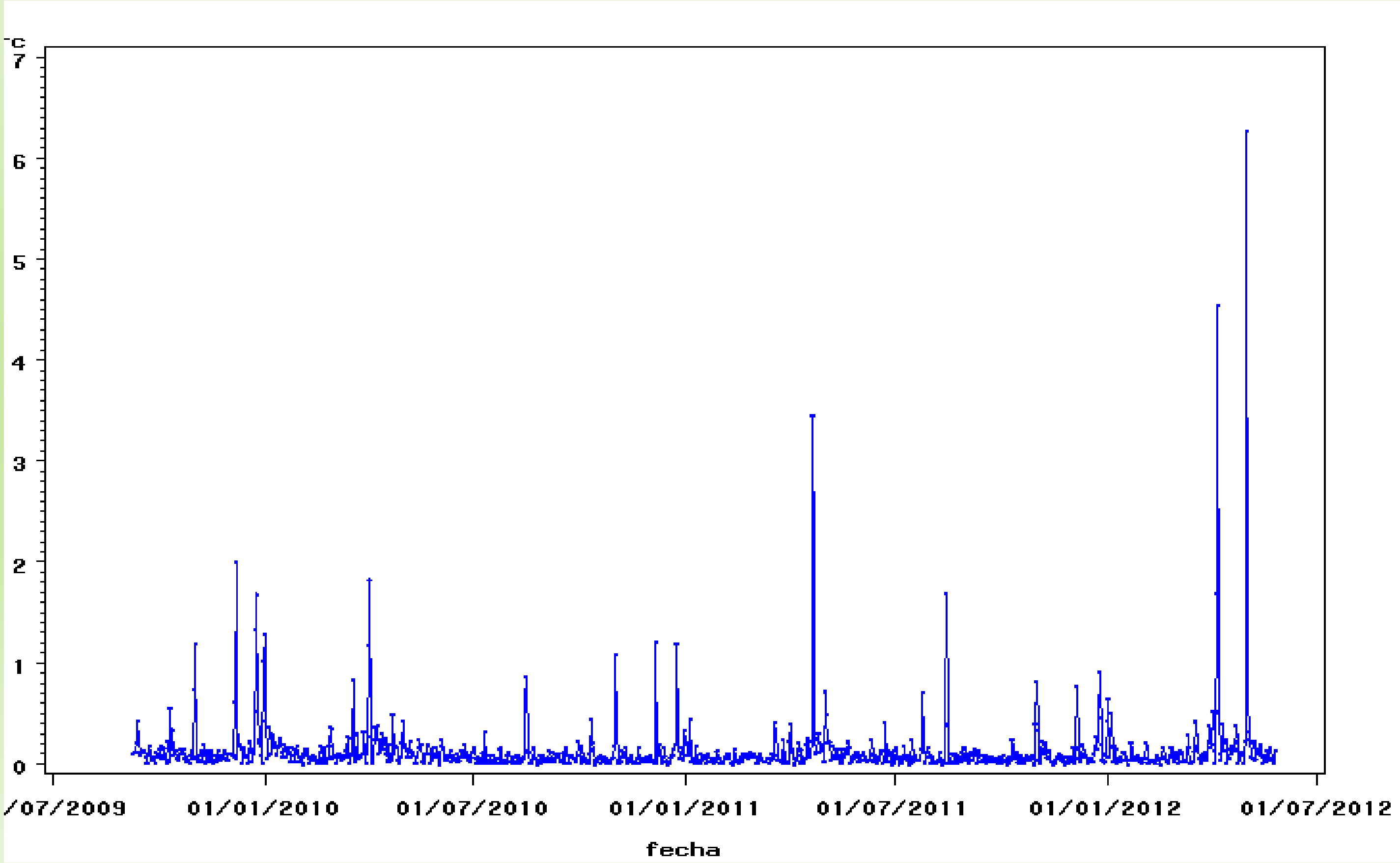
SARIMA Model

Example of a forecast:

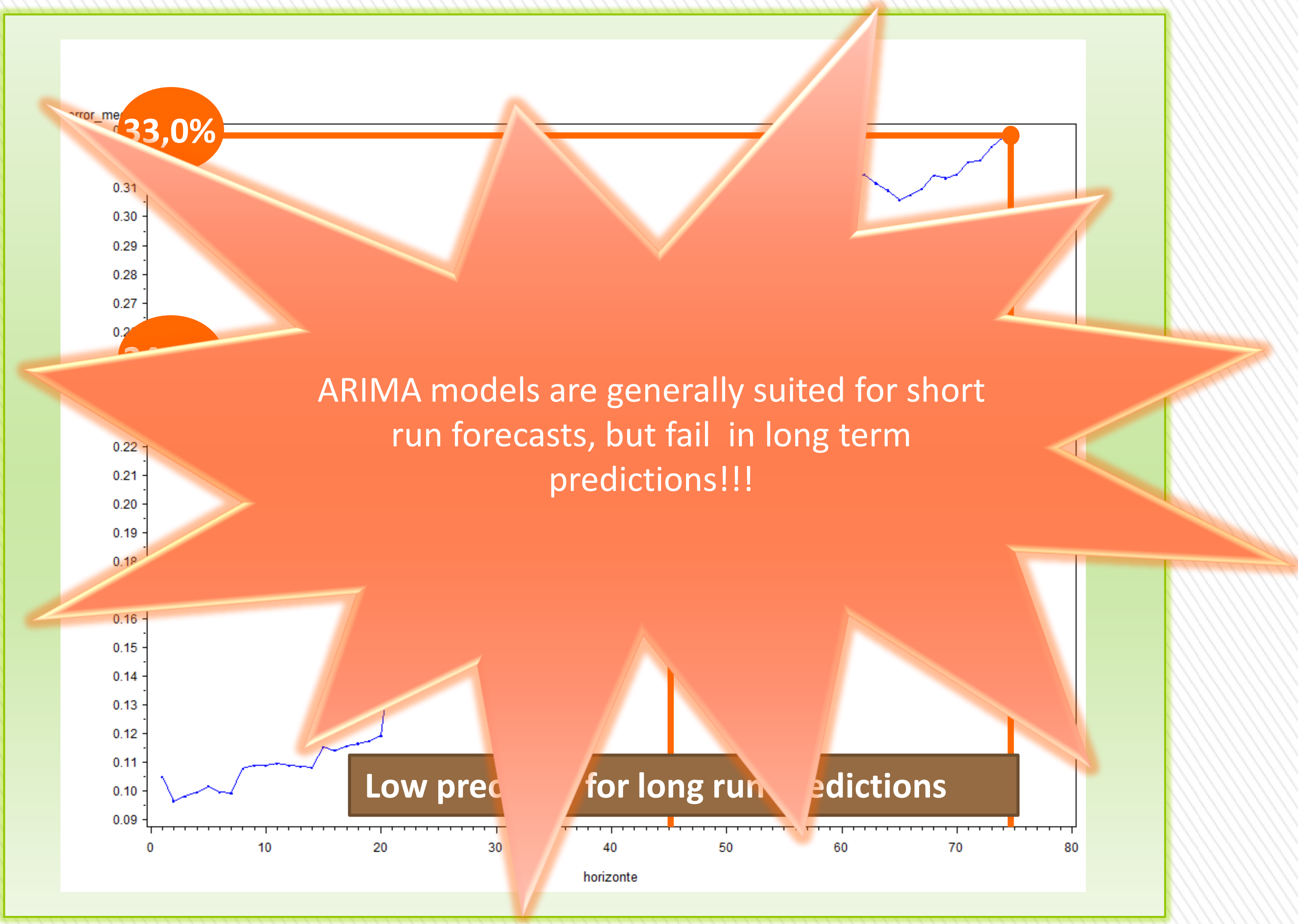


SARIMA Model

Example of forecast error:



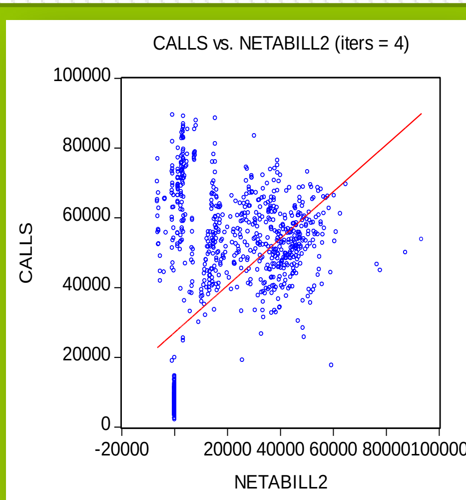
SARIMA Model

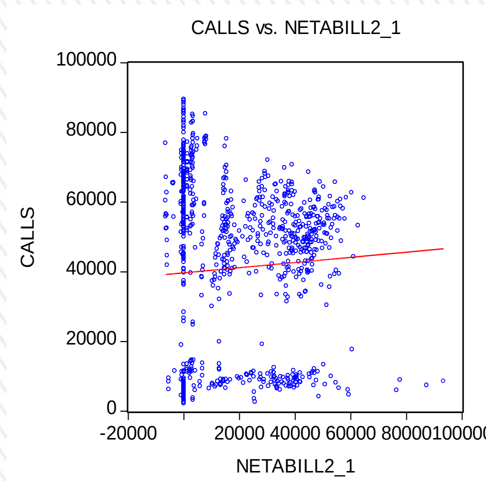


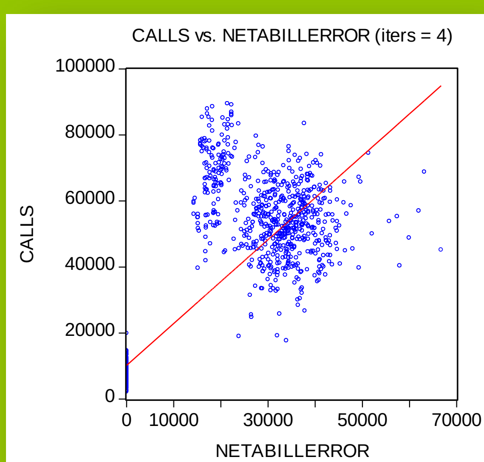
A Structural model

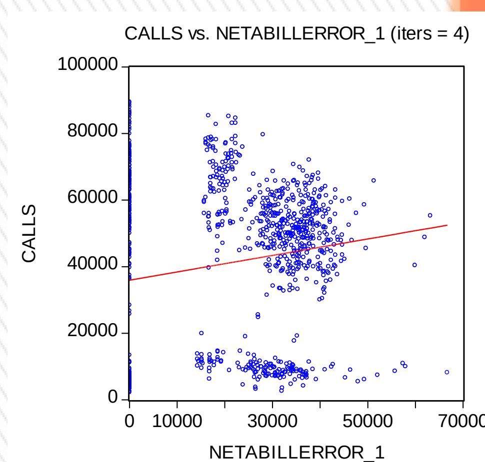
- The purpose is to estimate the **relationship** among some **relevant inputs** and the **output:**
  - Using a dynamical model for inputs.
  - Adding some seasonal events.
  - Estimating an additional model for the error term.
- In order to simplify the inputs, we have created two relevant variables:
  - NetBill2: the difference between Billtype2 and Billtype1
  - Neterror: the sum of Billtype3 and Billtype4

- We have found strong contemporary relation among target and inputs









Could be due to measure errors in our daily generation of weekly inputs.

Structural Model

Variable	Coefficient (lag)	Explanation
$\Delta\log(Calls_{\ t})$		Target variable
$\Delta\log(NetError_{\ t})$	0.20 (0)	An increase of 1% in billtype 3 or 4 **increases** 0.20% the calls
$\Delta(NetBill2_{\ t})$	-0.285(0) 0.141 (1)	An increase of 1000 billtype 2 over billtype1, **decreases** 1.4% the calls
Long term seasonality	$-0.0021\sin\left(\frac{2\pi t}{365}\right) + 0.0010\cos\left(\frac{2\pi t}{365}\right)$	Is a parsimonious way to take long run cyclical movements
Daily pattern	$0.47\Delta Mon + 0.39\Delta Tue + 0.35\Delta wed + 0.12\Delta fri + 0.0\Delta sat + 0.0\Delta sun$	Dummy variables to catch deterministic daily fluctuations
Error term	$\frac{(1-0.36L-0.54L^2)(1-0.20L^7)}{(1+0.06L-0.23L^2)}$	Capturates autocorrelation pattern

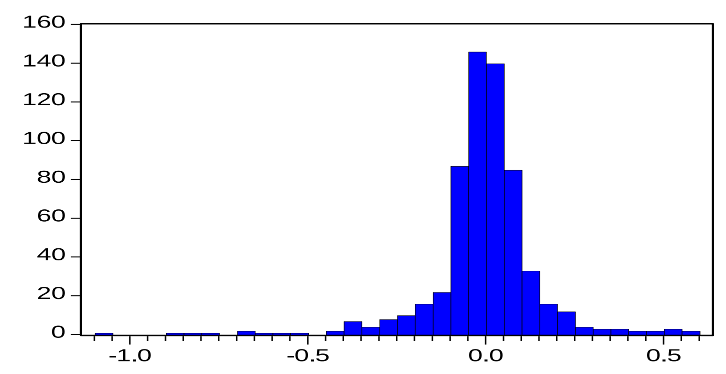
Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.010	-0.010	0.0661	
		2	-0.014	-0.014	0.1801	
		3	0.051	0.050	1.7663	
		4	0.024	0.025	2.1354	
		5	-0.043	-0.041	3.2928	
		6	0.077	0.075	7.0023	0.008
		7	0.002	-0.000	7.0053	0.030
		8	0.034	0.040	7.7393	0.052
		9	-0.063	-0.068	10.218	0.037
		10	-0.026	-0.032	10.644	0.059
		11	-0.049	-0.050	12.182	0.058
		12	0.053	0.052	13.950	0.052
		13	-0.040	-0.032	14.944	0.060
		14	0.019	0.016	15.175	0.086
		15	0.008	0.011	15.215	0.124

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.288	0.288	51.268	
		2	0.145	0.068	64.312	
		3	0.032	-0.029	64.937	
		4	0.037	0.029	65.812	
		5	0.053	0.041	67.537	
		6	0.031	0.001	68.153	0.000

We obtain a white noise residual correlogram…

…however…

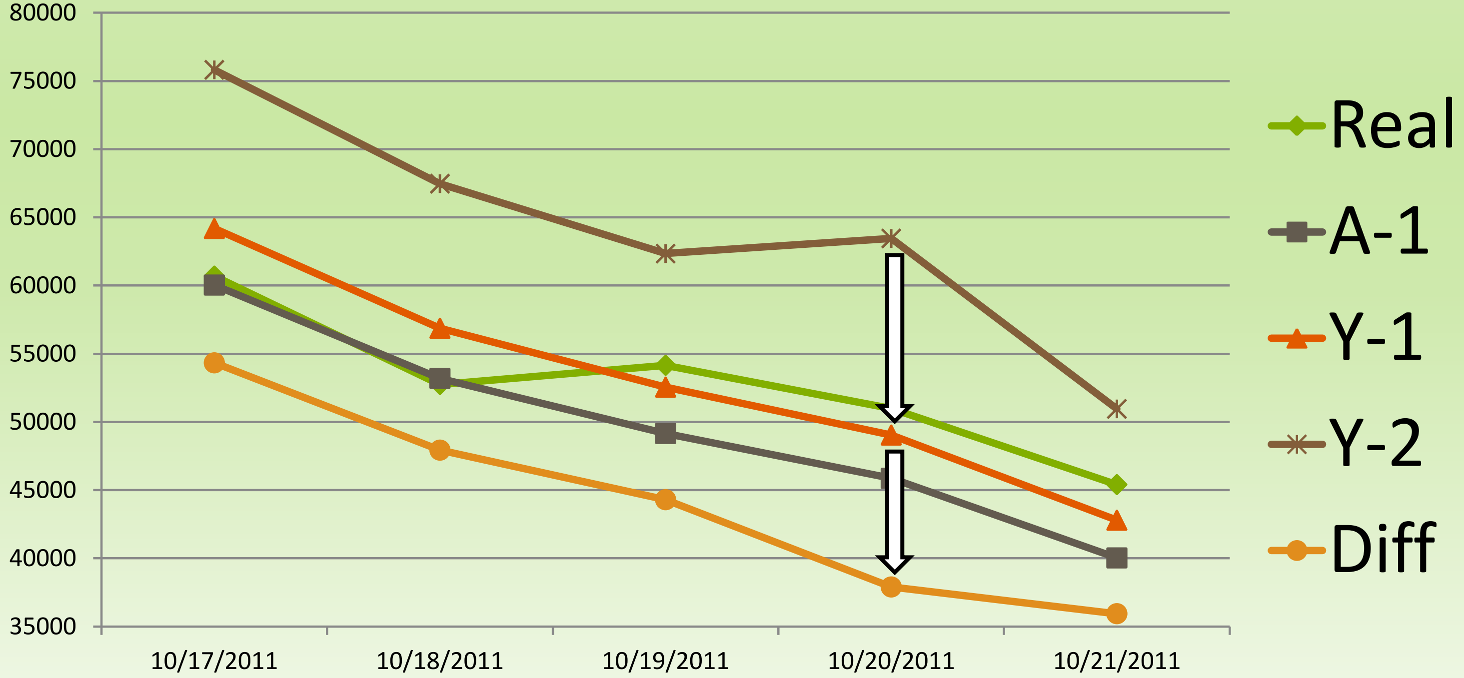
…residuals are not normal. Squared residual correlogram are not white noise



In a next step, we should study the posibility of any GARCH-ARCH structure for error term to model the error volatility.

Some Naïves Models

- Consider the number of calls of previous Year (**Y-1**)
- Adjust the **Y-1** model with a coefficient proportional to the difference with the real data (**A-1**)
- Apply the same variation (%) of the number of calls from Y-1 to Y-0 than from Y-1 to Y-2 (**Diff**)

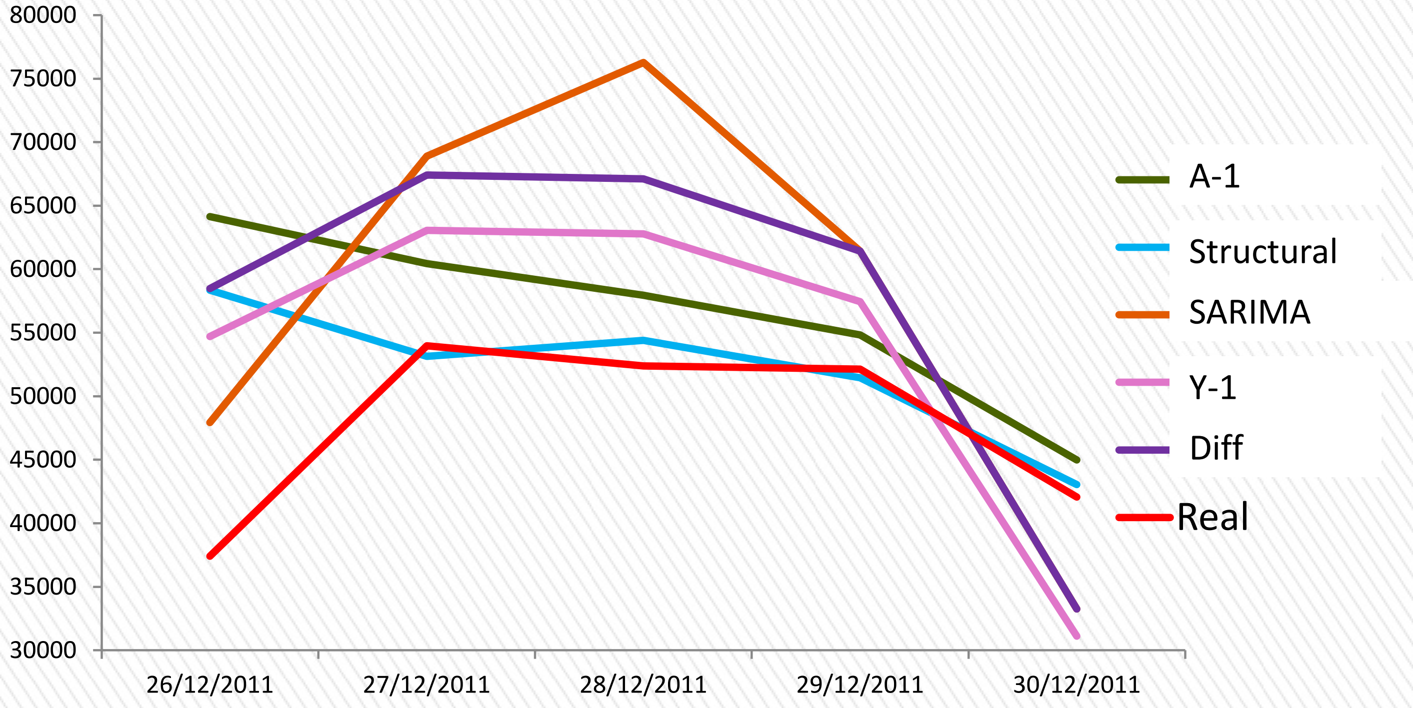


4. Validation and prediction

Validation and prediction

Mean relative error

Structural	SARIMA	Diff	A-1	Y-1
9,31	14,97	12,40	10,32	11,68



Weightened Models

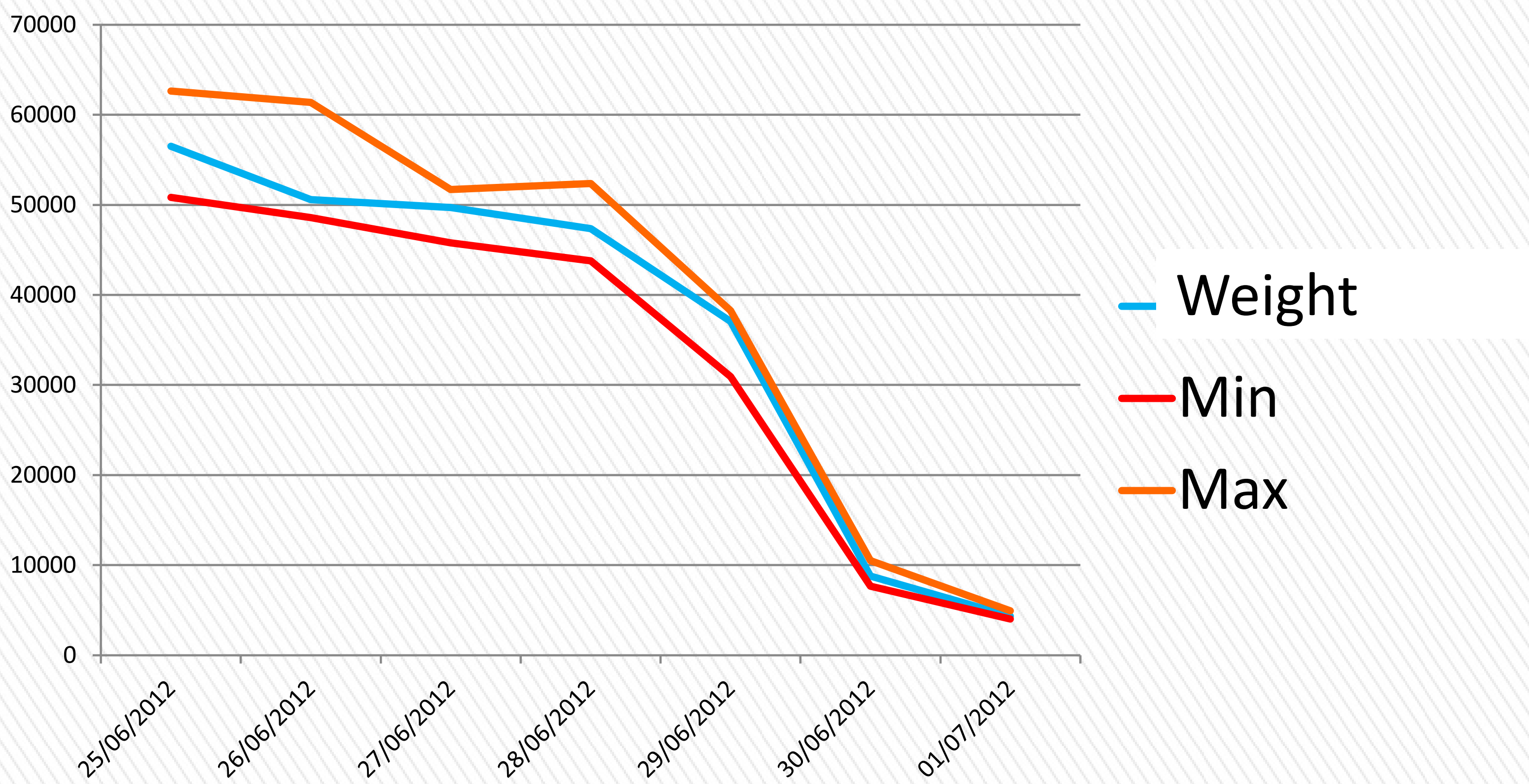
- Following *Clements and Hendry* we have considered a model based on the weigths of the solutions generated by all the models (Structural, SARIMA, Differential, A-1).
- We define $J_{err}(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$ the relative error function which is computed considering the weights $(\alpha_1, \alpha_2, \alpha_3, \alpha_4) \in \Theta=[0,1]^4$.
- We minimize J_{err} in Θ by considering a Newton method. Then, we obtain:

Structural	SARIMA	Diff	A-1
$\alpha_1 = 0.6$	$\alpha_2 = 0$	$\alpha_3 = 0$	$\alpha_4 = 0.4$

- Mean Error : **8,45** (vs. 9,31 for Structural)

Validation and prediction

We have predicted from 28/05/2012 to 30/12/2012.
For instance, a week obtained from inputs generated by Iberdrola.



5. Conclusions and perspectives

Conclusions and perspectives

Conclusions

- We have generated models with error values between 8%-15%.
- In addition, we have generated a prediction for future number of calls from an unknown time interval.

Perspectives

- Compare our solution with the real data when available.
- We would like to have additional variables such as campaigns, discretional events, etc.

Thank You!