This project has received funding from the European Union's Horizon 2020 research and innovation program under

Deliverable No. 5.2

Project acronym: **PrimeFish**

Project title:

"Developing Innovative Market Orientated Prediction Toolbox to Strengthen the Economic Sustainability and Competitiveness of European Seafood on Local and Global markets"

Grant agreement No: **635761**

This project has received funding from European Union´s Horizon 2020 research and innovation program.

Start date of project: **1st March 2015**

Duration: **48 months**

Due date of deliverable:	28/02/2018
Submission date:	27/02/2018
File Name:	PrimeFish_D5_2_Boom and bust model
Revision number:	01
Document status:	Final[1]
Dissemination Level:	CO[2]

Revision Control

Role	Name	Organisation	Date	File suffix[3]
Authors	Cristina Mora	UNIPARMA	21/02/2018	CM
Authors	Marco Riani	UNIPARMA	21/02/2018	MR
Authors	Fabrizio Laurini	UNIPARMA	21/02/2018	FL
Authors	Gianluca Morelli	UNIPARMA	21/02/2018	GM
WP leader	Sveinn Agnarsson	UIce	22/02/2018	SA
Coordinator	Guðmundur Stefánsson	Matis	26/02/2018	GS

[1] Document will be a draft until it was approved by the coordinator

[2] PU: Public, PP: Restricted to other programme participants (including the Commission Services), RE: Restricted to a group specified by the consortium (including the Commission Services), CO: Confidential, only for members of the consortium (including the Commission Services)

[3] The initials of the revising individual in capital letters

This project has received funding from the European Union's Horizon 2020 research and innovation program under

Deliverable 5.2

"Boom and Bust Model"

Demonstrator/Protoype

February 2018

This project has received funding from the European Union's Horizon 2020 research and innovation program under

Executive Summary

The main aim of the PrimeFish project is to develop an innovative decision support framework, PrimeDSF, which contains economic and statistical models and a decision support system, PrimeDSS, that can be used by the industry and policymakers to better predict consequences based on existing knowledge and simulation/forecasting models. Starting from the analysis carried out in Task 2.3 "Boom and Bust Analysis" and D2.4, this deliverable describes how the FSDA toolbox will be used to predict price behaviour in the growth risk analyser (GRA) of PrimeDSS. To simplify the calculations to be done in the GRA, this deliverable is based on different statistical models to the one used in WP2. The Kalman filter used there generated 4-5 graphs for each analysis that had to be read together. As only statisticians could probably fully grasp the information revealed in the graphs, it was deemed necessary to develop more easily accessible presentations. For this purpose, new models were compiled for use in WP5 and WP6 that summarize the price forecast in a single chart joined by a table that shows the price expected for each month and the extreme the price can take.

This document describes the methodology behind the price prediction tool developed in WP5, the GRA. The purpose of the GRA is to predict price over time (up to 12 months) for a species based on historical data (at least 36 months series) and on research of previous "boom and bust" cycles in the fisheries market. Finally, this deliverable also provides instructions, already circulated to the Syntesa partners, on how to implement this tool into the PrimeDSS in WP6.

This project has received funding from the European Union's Horizon 2020 research and innovation program under

Contents

This project has received funding from the European Union's Horizon 2020 research and innovation program under

1. Introduction

This deliverable discusses the methodology to develop simulation and prediction models to be used to predict price behaviour and to be integrated in the GRA of the PrimeDSS and PrimeDSF.

2. Robust monitoring of time-series

The statistical model used for the price prediction is based on Robust Monitoring of Time-series approach. Time-series often contain outliers and level shifts or structural changes, and these unexpected events are of the utmost importance in the forecasting of prices. The presence of such unusual events can easily mislead conventional time-series analysis and yield erroneous conclusions. The model provides a unified framework for detecting outliers and level shifts in short time-series that may have a seasonal pattern. The methodology was developed to detect potential fraud cases in time-series of imports into the European Union, and we have borrowed the methodology as it is particularly well suited to the type of data and phenomena this deliverable deals with (Barabesi et al, 2016; Fried et al., 2012; Galeano and Pena, 2013; Riani et al., 2012), Rousseeuw and van Driessen, 2006; Salini et al., 2015; FSDA).

The formal approach of the model is described in Perrotta et al. (2018), which contains a methodology for robustly analysing a time-series which contains a trend, a seasonal component (possibly time varying) and a level shift in an unknown position, as well as isolated or consecutive outliers.

The model is particularly suitable for the task at hand because it introduces a new robust approach to model and monitor non-linear time-series with a possible level shift. A fast algorithm was developed and applied to several real and artificial datasets. The automatic detection of level shift avoids the alternative and most common way of splitting the parts before and after the break, after which each part can be analysed separately.

The model is especially relevant for this study because it can be applied to data sets which are not very long (less than 36 months), such as frequently found in the fisheries sectors. The models also make it possible to manage any significant price changes over the period observed and therefore to make better forecasts. It is also quite easy to interpret the results for persons with non-statistical background. The model also offers the possibility to calculate robust confidence bands in the forecasts, therefore to define the risk linked to the forecasts. Finally, the model is innovative, as the method applied represents an improvement on previous models.

This project has received funding from the European Union's Horizon 2020 research and innovation program under

The methodology applied is coherent with the current development of the international research in this field. However, other methodologies could have been applied, such as ARCH and GARCH models. ARCH (autoregressive conditionally heteroscedastic) model is a model for the variance of a time-series. ARCH models are used to describe a changing, possibly volatile variance. Although an ARCH model could possibly be used to describe a gradually increasing variance over time, most often it is used in situations in which there may be short periods of increased variation. GARCH (generalized autoregressive conditionally heteroscedastic) model uses values of the past squared observations and past variances to model the variance at time t.

The main strength of this method is that it works work well with not very long times series (less than 36 months), as are often found in fisheries, the ability to manage any significant price changes over the period observed and therefore to make better forecasts, the ease of interpretation of results for persons with non-statistical background and the possibility to calculate the confidence bands in the forecasts, therefore to define the risk linked to the forecasts and its robustness to the presence of isolated or consecutive outliers.

The weaknesses of the model, in relation to the aims of the project, are its inability to not foresee the boom and bust cycles in future periods compared to the observed data. This incapacity is not related to the model chosen in particular, but it is common to each statistical model.

The methodology applied was chosen to suit the task at hand and the nature of the data collected. The model applied provides, in our opinion, the best possible estimate of the cycle in prices time-series for the species analysed.

Future work should be based on data collected using a well-defined methodology and set up in a well organised database.

This study was based on the following variables taken from EUMOFA; period (month, year), monthly prices, country, flow type (import or export), partner country (imported from or exported to), fish species, market (first sale/landing, wholesale, retail).

3. Prototype of the GRA from the web site of Prime Fish

The end-users will be able to enter in the private area of the tool, now through the link :

This project has received funding from the European Union's Horizon 2020 research and innovation program under

http://www.dss.primefish.eu/index.php/gra?view=results

and choose the link to the GRA (Fig. 1).

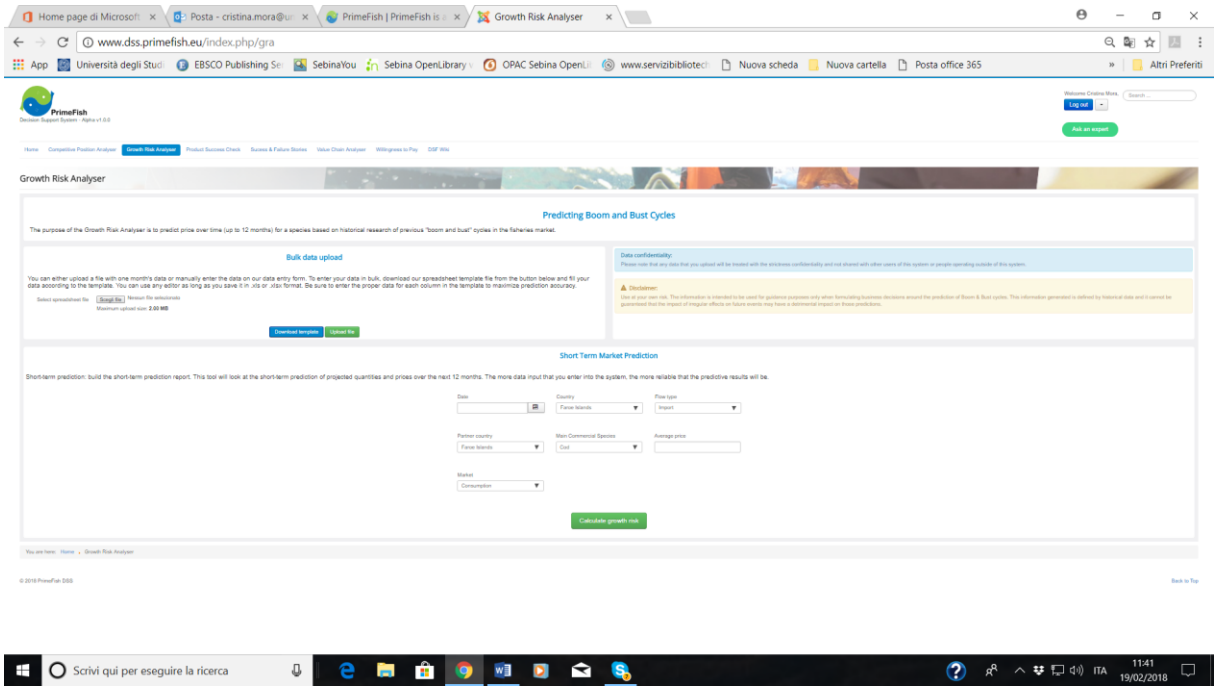


Figure 1: The GRA - Growth risk analyser

The purpose of the GRA is to predict price over time (up to 12 months) for a species based on historical data (at least 36 months series) and on research of previous "boom and bust" cycles in the fisheries market.

The data needed will be uploaded by end-users or by the company. The data must be "good data", but can come from any source, and must be suitable with the structure of the data set. The data is connected to the DSS tools which are linked with the FSDA-toolbox (forward search for data analysis) that extends MATLAB and the Statistics Toolbox for the robust analysis of data sets affected by different sources of heterogeneity.

This project has received funding from the European Union's Horizon 2020 research and innovation program under

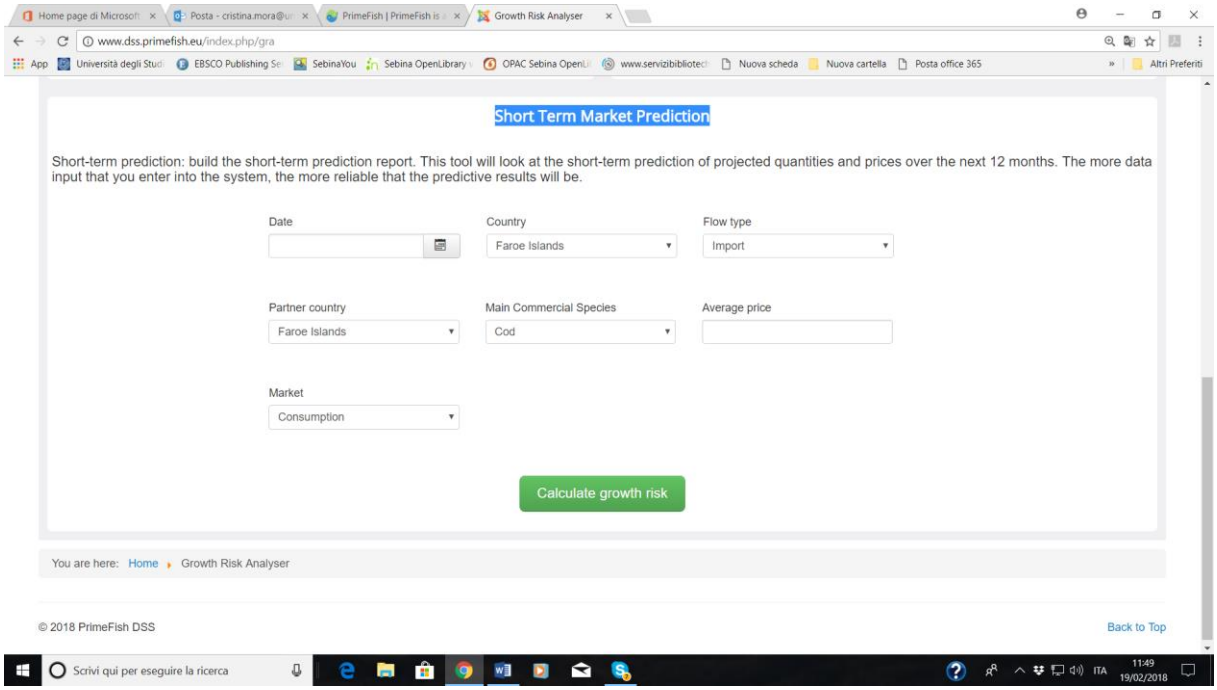
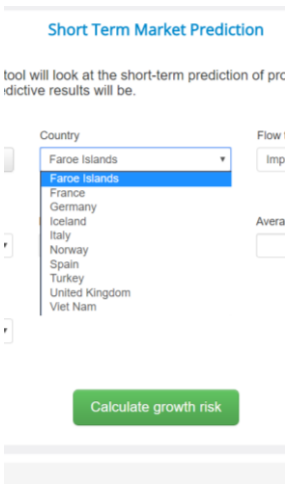


Figure 2: Short term market prediction frame

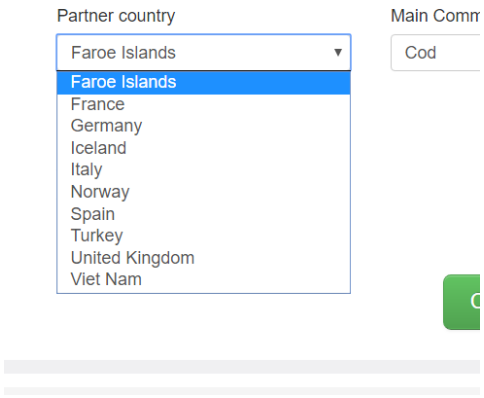
Short-term market prediction

The growth risk analyser yields a prediction of the development of prices over the next 12 months. The more data input that is entered into the system, the more reliable the predictive results will be.

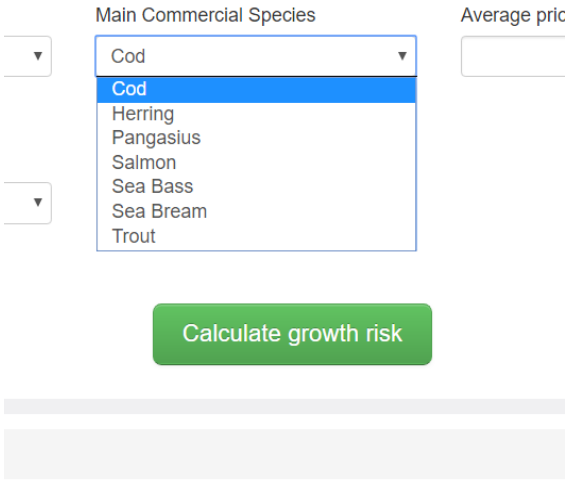
The end-user will use pull-down menus to indicate the nature of the data set. This will proceed in several steps. 1) choose the country (see below)



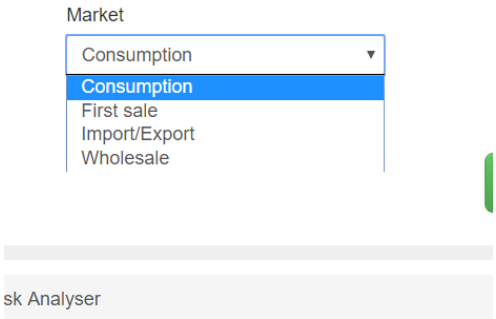
2) Choose the flow type (import and export) and the partner country (see below):



3) Choose the main commercial species:



4) Choose the type of market.



The page displaying the GRA predictive results is currently under revision. This version of the help/explanation page refers to a previous approach which was based on the Kalman filter (the methodology used in WP2).

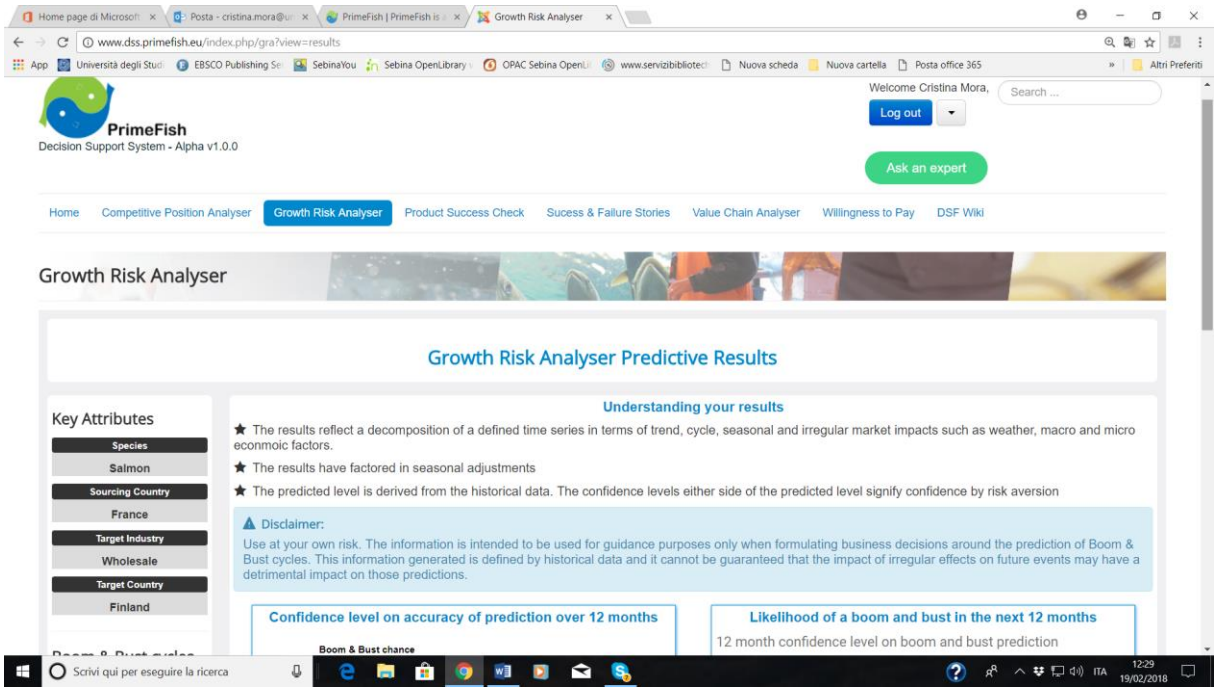


Figure 3: Help menu for GRA

This project has received funding from the European Union's Horizon 2020 research and innovation program under

4. Robust monitoring of time-series

The methodology discussed above is illustrated, using as an example the problem of forecasting prices of seabass in the Italian consumption market. The end-user chooses the attributes of the data at hand as shown above. The tool should produce a filtered dataset which will then be saved as csv-file in a folder specified by the user.

A Matlab code is used to capture the data file and analyse the data provided. For each time-series, future values are provided for an out of sample period consisting of 12 months (value 1 in Fig. 4). Confidence bands are also provided (value 2 and value 3in Fig. 4).

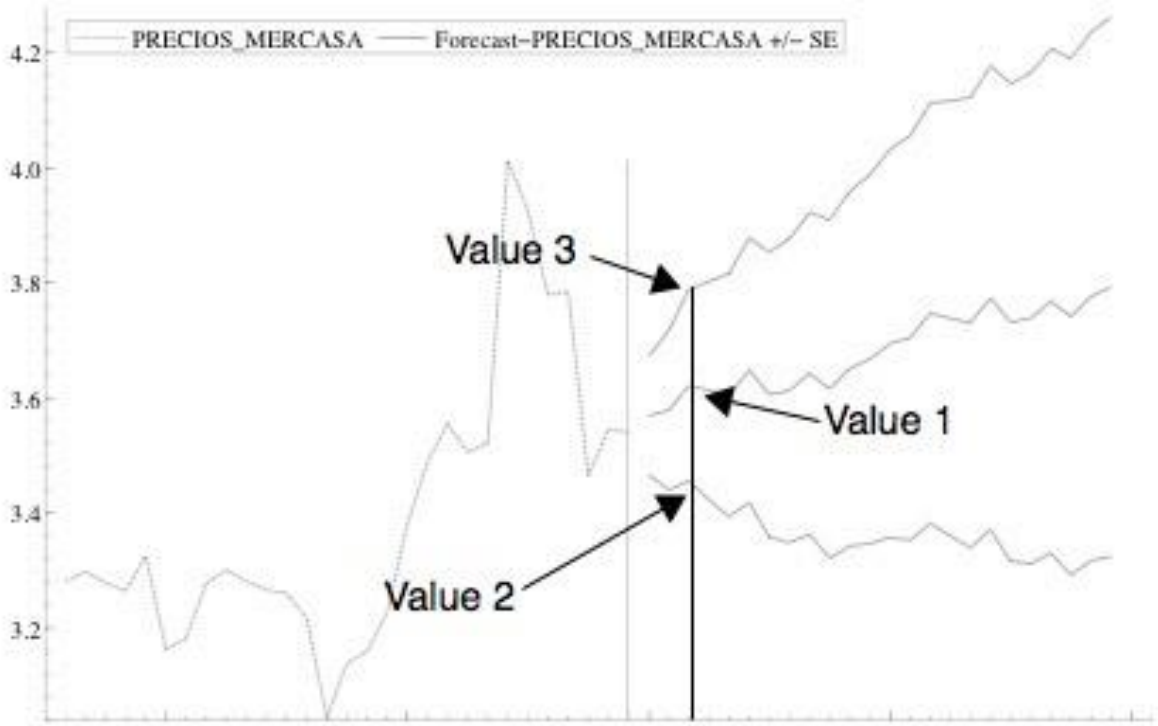


Figure 4: Price prediction

5. From robust monitoring of time-series in FSDA to GRA in PrimeDSS

The FSDA toolbox (Fig. 5) will be used with a Matlab algorithm (the least trimmed squares (LTS) functions[4]) to perform the calculations. Integrating the Matlab code and PHP[5] is a task that will be completed in WP6. The algorithm of the analysis developed for the study of price forecasts is defined as LTS. The codes will be connected with the GRA to allow the PrimeDSS to engage with the algorithm.

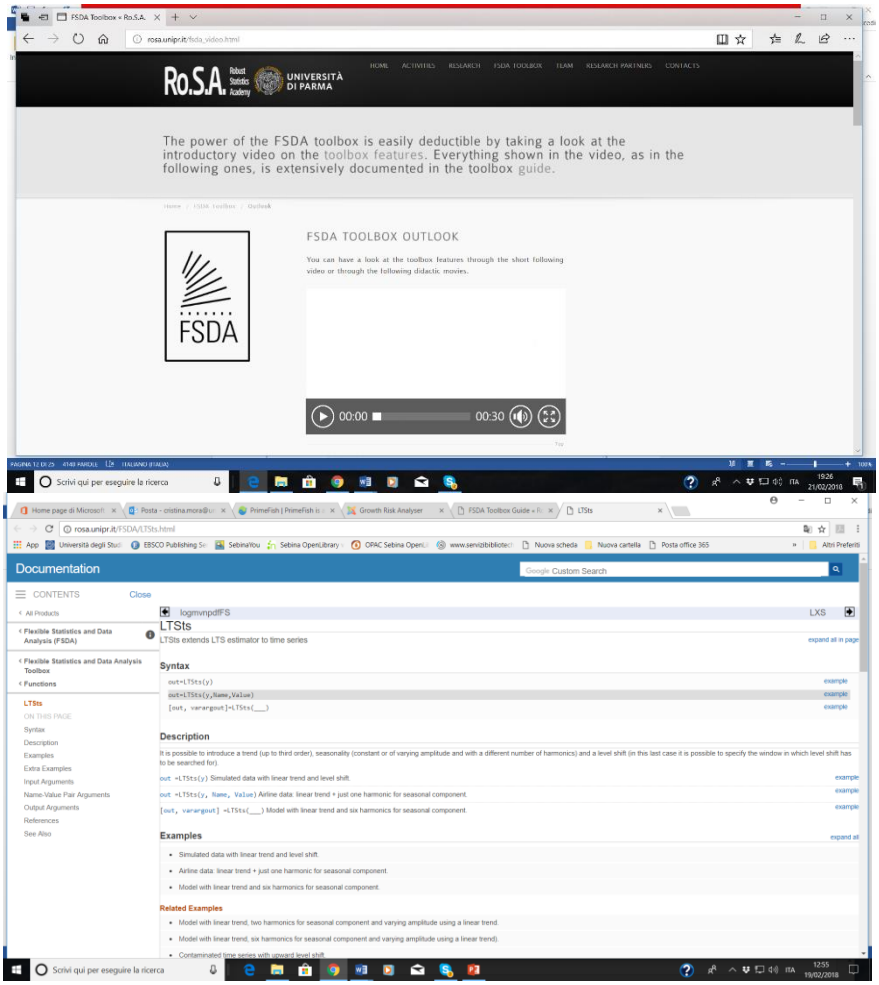


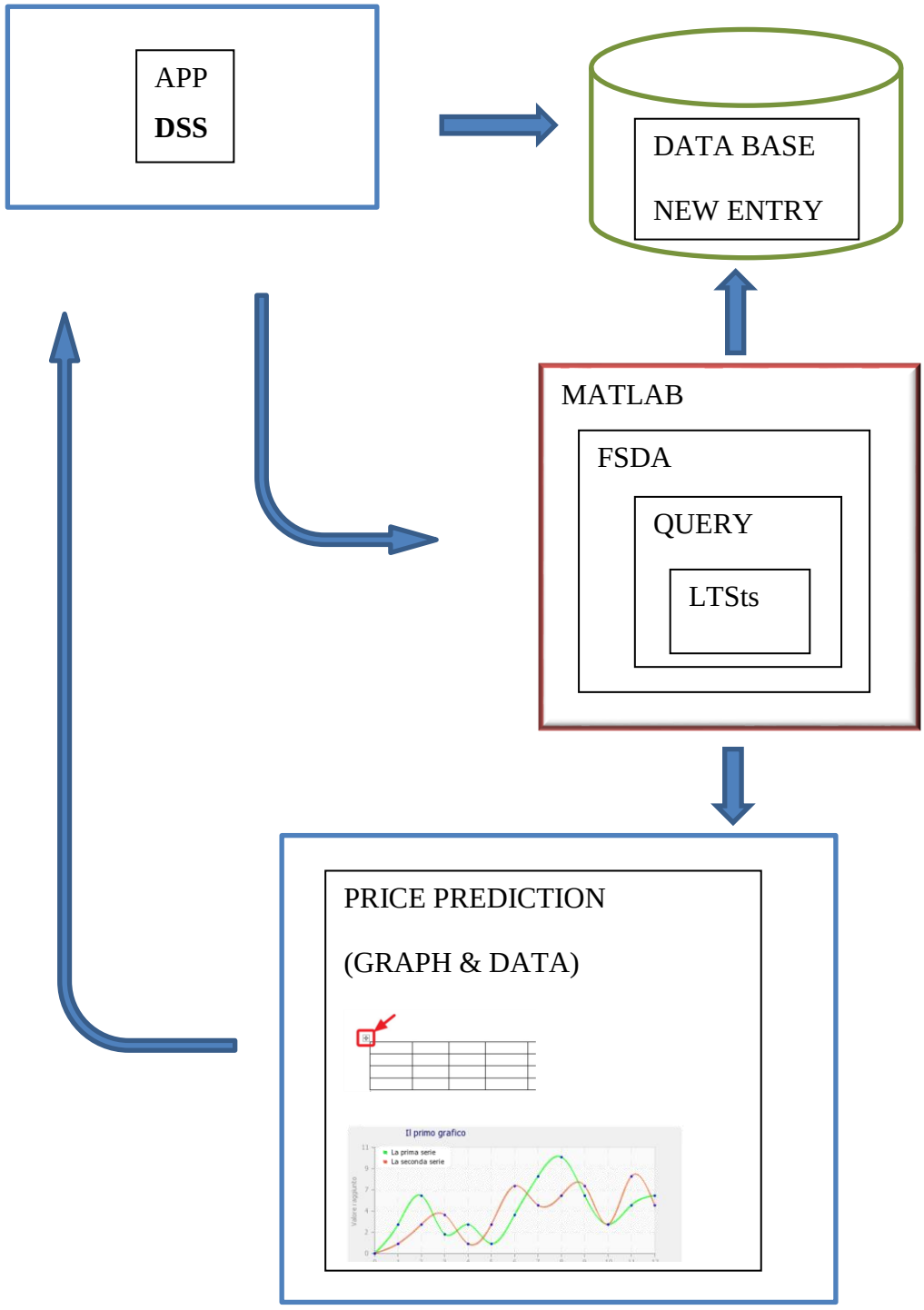
Figure 5: FSD toolbox web page (at http://rosa.unipr.it/team.html)

[4] Extends LTS estimator to time series. It is possible to introduce a trend (up to third order), seasonality (constant or of varying amplitude and with a different number of harmonics) and a level shift (in this last case it is possible to specify the window in which level shift has to be searched for). http://rosa.unipr.it/FSDA/guide.html: LTSts

[5] PHP is a server-side scripting language designed for web development.

This project has received funding from the European Union's Horizon 2020 research and innovation program under

Figure 6: GRA/ FSDA Matlab Model



This project has received funding from the European Union's Horizon 2020 research and innovation program under

6. Conclusion

This deliverable discusses the methodology developed to compile the simulation and prediction models to be used to predict price development. The models will be integrated in the GRA of the PrimeDSS and PrimeDSF, an approach also for the understanding the results.

7. References

Barabesi, L., A. Cerasa, D. Perrotta, and A. Cerioli (2016). Modelling international trade data with the Tweedie distribution for anti-fraud and policy support. European Journal of Operational Research 258, 1031–1043.

Fried, R., K. Schettlinger, and M. Borowski (2012). Package 'robfilter'. CRAN. URL:

http://www.statistik.tu-dortmund.de/fried.html.

FSDA Toolbox: http://rosa.unipr.it/fsda.html.

Galeano, P. and D. Peña (2013). Finding outliers and unexpected events in linear and nonlinear possible multivariate time-series data. In C. Becker, R. Fried, and S. Kuhnt (Eds.), Robustness and Complex Data Structures. Festschrift in Honour of Ursula Gather,pp. 255–273. Springer.

Perrotta, D., Riani. M., Rousseeuw, P. and Hubert, M. (2018). "Robust Monitoring of Time-series with Application to Fraud Detection". Manuscript submitted.

Riani, M., Perrotta, D. and Torti, F. (2012). FSDA: A MATLAB toolbox for robust analysis and interactive data exploration, CHEMOMETRICS AND INTELLIGENT LABORATORY SYSTEMS, 116, p. 17-32, doi:10.1016/j.chemolab.2012.03.017.

Rousseeuw, P. and K. Van Driessen (2006). Computing LTS regression for large data sets. Data Mining and Knowledge Discovery 12, 29–45.

Salini, S., A. Cerioli, F. Laurini, and M. Riani (2015). Reliable robust regression diagnostics. International Statistical Review 84, 99–127.