PrimeFish

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 635761

Deliverable No. 5.6

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 programme.

Start date of project: **1st March 2015**
Duration: **48 months**

Due date of deliverable:	31/12/2018
Submission date:	06/02/2019
File Name:	PrimeFish_Deliverable 5.6
Revision number:	02
Document status:	Draft[1]
Dissemination Level:	PU[2]

Revision Control

Role	Name	Organisation	Date	File suffix[3]
Authors	Cristina Mora	UNIPARMA	05/02/2019	CM
Authors	Marco Riani	UNIPARMA	05/02/2019	MR
Authors	Fabrizio Laurini	UNIPARMA	05/02/2019	FL
Authors	Gianluca Morelli	UNIPARMA	05/02/2019	GM
WP leader	Sveinn Agnarsson	UIce	05/02/2019	SA
Coordinator	Guðmundur Stefánsson	Matis	06/02/2019	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 programme under grant agreement No 635761

Deliverable D5.6

Scientific papers on "Early warning signs for boom and bust cycles"

February 6th, 2019

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 635761

Executive Summary

This report consists of a brief summary of the "boom and bust" tool developed by Parma University in PrimeFish and two peer reviewed papers about the MATLAB implementation of the method in the FSDA toolbox developed by the University of Parma. A third paper is under revision. The intended audience of the deliverable and the journal article is primarily the scientific community, but the executive summary is written in plain language and is therefore understandable to other stakeholder groups.

Based on work carried out in task 2.3 of PrimeFish "Identifying and characterising "boom and bust cycles" and D2.4, the Flexible Statistics and Data Analysis (FSDA) toolbox was used to predict price behaviour in the growth risk analyser (GRA) of PrimeDSS. To simplify the calculations carried out in GRA, a simple statistical model based on Robust Monitoring of Time Series approach was used. 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, new models were compiled for use in WP5 and WP6 that summarise the price forecast in a single chart joined by a table that shows the price expected for each month and the extreme prices that can occur.

The first scientific article included in the deliverable was authored by Marco Riani and has been available on-line from the 19th of November 2018:

RIANI M., ATKINSON A.C., CERIOLI A., CORBELLINI A. (2019). *Efficient robust methods via monitoring for clustering and multivariate data analysis*, **Pattern Recognition**, 88, p. 246–260. https://doi.org/10.1016/j.patcog.2018.11.016

Abstract: Monitoring the properties of single Sample robust Analyses of Multivariate data as a function of break- down Point or efficiency Leads to the adaptive choice of the best values of these parameters, eliminating arbitrary decisions about their values and so increasing the quality of estimators.

Journal Metrics : Impact Factor: 3.962

The second scientific article included in the deliverable was co-authored by Marco Riani and has been available on-line from the 30th of July 2018:

TORTI F. PERROTTA D., RIANI M., CERIOLI A (2018). Assessing Trimming Methodologies for Clustering Linear Regression Data, **Advances in Data Analysis and Classification**, https://doi.org/10.1007/s11634-018-0331-4

Journal Metrics : Impact Factor 1.653

Abstract: We assess the performance of state-of-the-art robust clustering tools for regression structures under a variety of different data configurations. We focus on two methodologies that use trimming and restrictions on group scatters as their main ingredients. We also give particular care to the data generation process through the development of a flexible simulation tool for mixtures of regressions, where the user can control the degree of overlap between the groups. Level of trimming

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 635761

and restriction factors are input parameters for which appropriate tuning is required. Since we find that incorrect specification of the second-level trimming in the Trimmed CLUSTering REGression model (TCLUST-REG) can deteriorate the performance of the method, we propose an improvement where the second-level trimming is not fixed in advance but is data dependent. We then compare our adaptive version of TCLUST-REG with the Trimmed Cluster Weighted Restricted Model (TCWRM) which provides a powerful extension of the robust cluster wise regression methodology. Our overall conclusion is that the two methods perform comparably, but with notable differences due to the inherent degree of modelling implied by them.

A draft abstract for the third paper "Detection of "boom and bust" cycles and making predictions for price using Efficient Robust Methods" is also included in the deliverable.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 635761

Contents

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 635761

Introduction

The main aim of the PrimeFish project is to develop an innovative decision support framework, PrimeDSF, which contains economic and statistical model 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.

Based on work carried out in task 2.3 of PrimeFish "Identifying and characterising "boom and bust cycles" and D2.4, the Flexible Statistics and Data Analysis (FSDA) toolbox was used to predict price behaviour in the growth risk analyser (GRA) of PrimeDSS. To simplify the calculations carried out in GRA, a simple statistical model based on Robust Monitoring of Time Series approach was used. 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 summarizes the price forecast in a single chart joined by a table that shows the price expected for each month and the extreme prices that 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 with the Syntesa partners, on how to implement this tool into the PrimeDSS in WP6.

As part of WP5 "Development of robust simulation and prediction models" in PrimeFish, this report present:

- the methodology to develop simulation and prediction models used to predict price behaviour and integrated in the Growth Risk Analyzer (GRA) of the PrimeDSS and PrimeDSF;

- the activity for the Dissemination of the tool;

- two peer reviewed papers already published about the MATLAB implementation of the method used for the Growth Risk Analyzer (Annex);

- the Draft Abstract of a third paper.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 635761

Methods

1. Robust Monitoring of Times 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 methodology for robustly analysing trend containing time series, 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 nonlinear 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 in 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 are 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.

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.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 635761

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, 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. The tools were based on the following variables taken from EUMOFA databank; 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).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 635761

2. 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 throught the link :

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

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

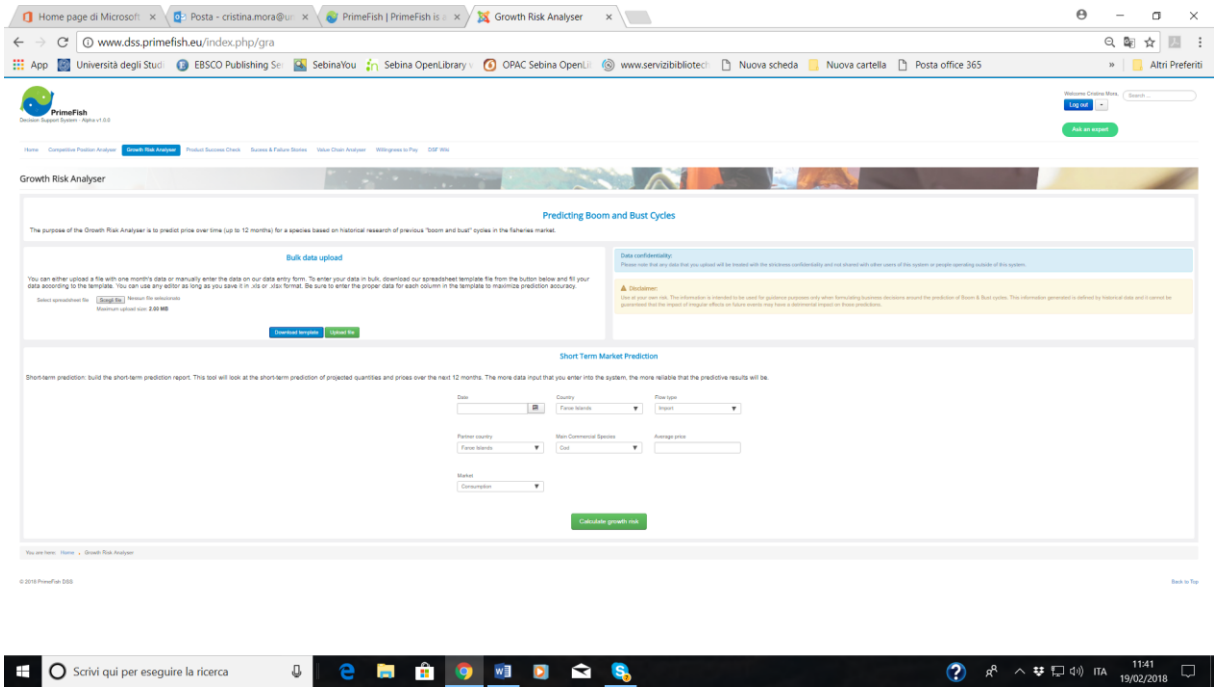


Fig. 1: The GRA - Growth Risk Analyser

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.

Bulk data upload

The data needed will be uploaded by end-users or by the company and the source is not predermined, but it must be "good data" and suitable with the structure of the data set, connected to the DSS tools linked with the FSDA.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 635761

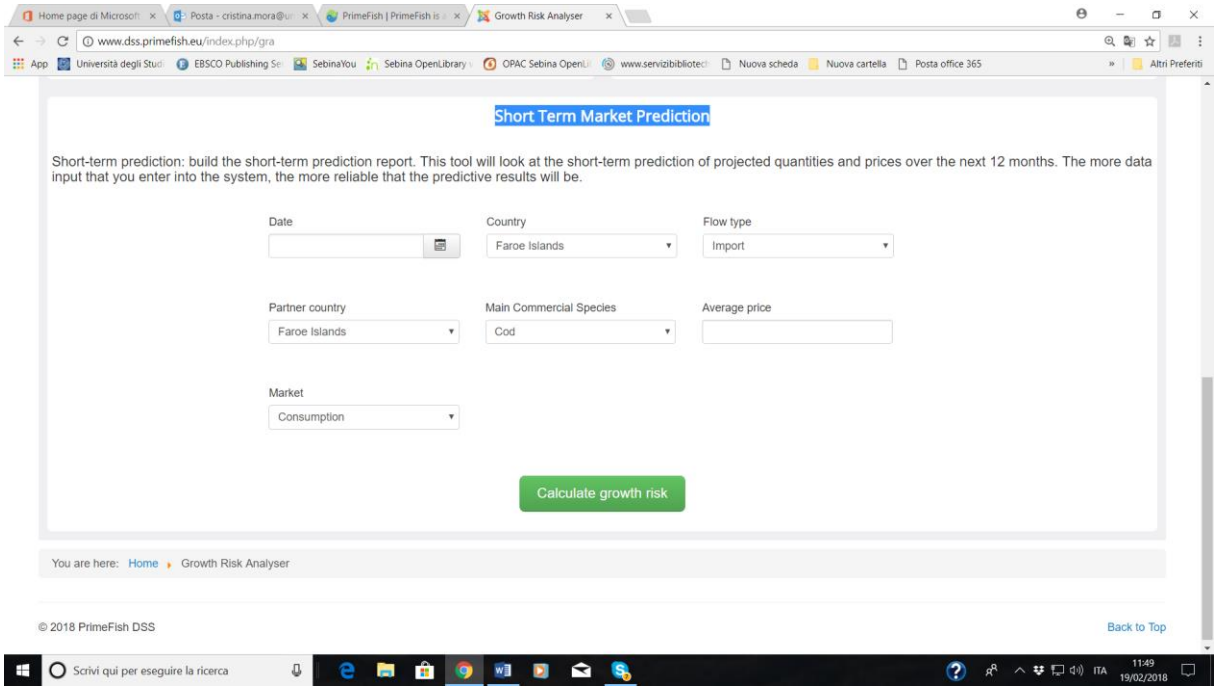


Fig. 2: Short term market prediction frame

Short-term market prediction

The Tools will return a short-term market prediction. This tool will look at the short-term forecast of prices over the next 12 months. The more data input that is entered into the system, the more reliable that the predictive results will be.

The tool analyses a time series of data of any value type such as profit per month, average sales quantity per day/week/fortnight/etc, average catch/landings per month/semester/year/etc, containing a minimum of 24 values (to ensure accuracy).

http://www.dsf.primefish.eu/wiki/GRA

3. Robust monitoring of time-series

The methodology is here discussed 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. 3). Confidence bands are also provided (value 2 and value 3 in Fig. 3).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 635761

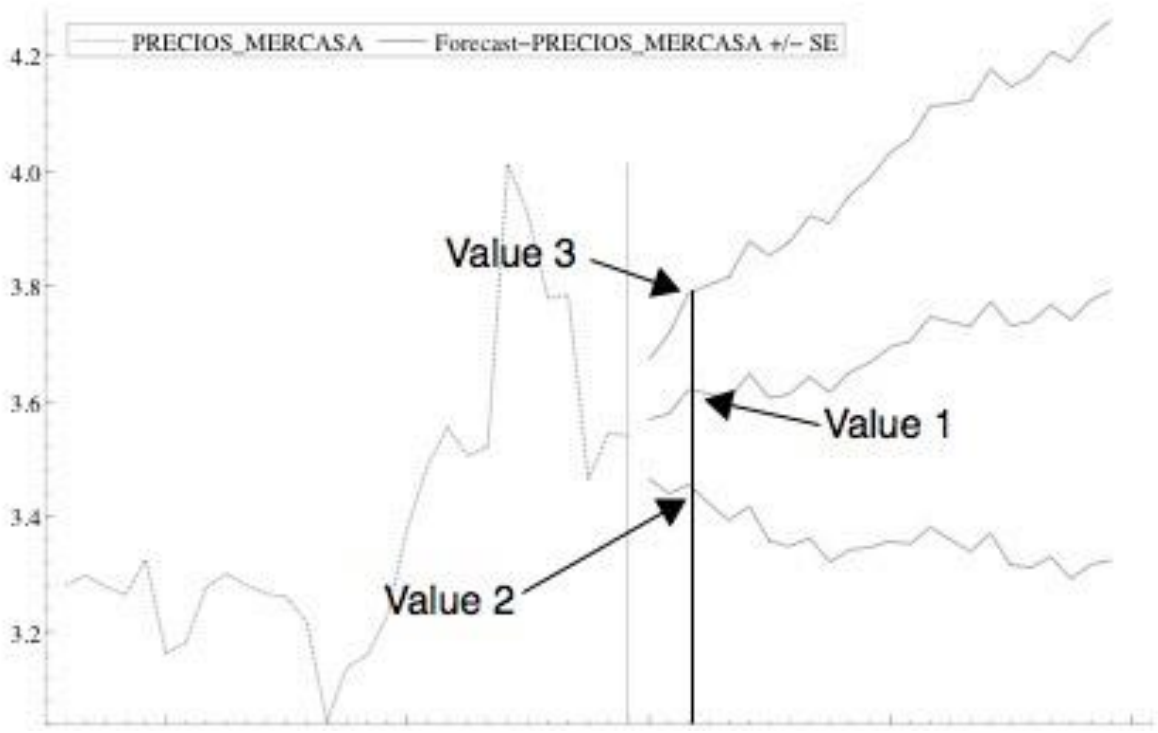


Fig. 3: Price prediction

One of the outcomes of the research is the Growth risk analyser tool, implemented through simulation/forecasting models for analysing and forecasting pricing trends in the short term: up to 12 months with a confidence band of 95% in the case of this tool. Forecasts given outside of 12 months become less reliable as the forecast horizon increases (up to 24 months are generated by the tool).

The Growth Risk Analyser statistical model is designed to predict pricing trends in the short term: up to 12 months with a 95% confidence band in the case of this tool. The tool forecasts 24 values, but forecasts given outside of the first 12 months become less reliable as the forecast horizon increases. A user must enter a time series of data of any value type such as profit per month, average sales quantity per month or average catch/landings per month. You must input a minimum of 24 values, but for better accuracy we recommend a minimum of 36 values.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 635761

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

The FSDA toolbox (Fig. 4) will be used with a Matlab algorithm (the LTSts functions[4]) to perform the calculations. Integrating the Matlab code and PHP[5] is a task that remains to 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 to the GRA to allow the PrimeDSS to engage with the algorithm.

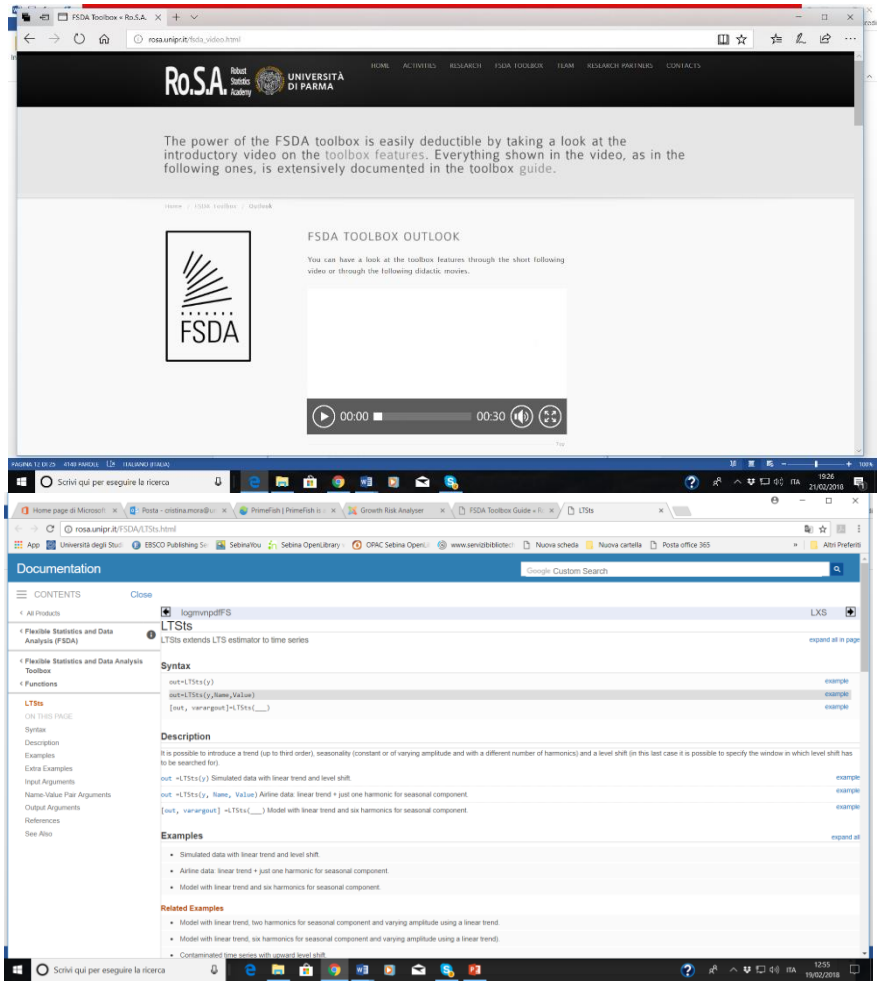


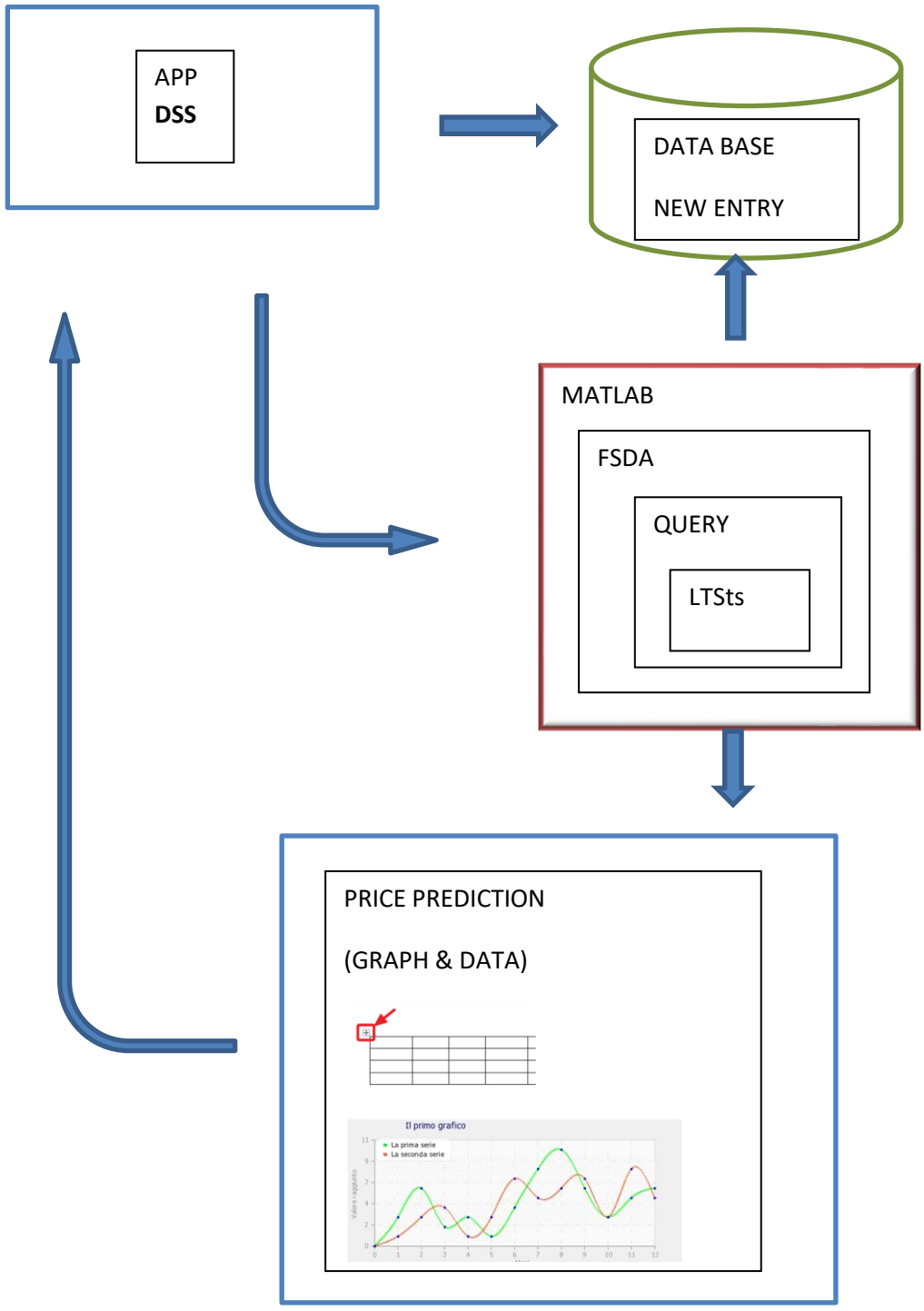
Fig. 4: FSD toolbox web page (at http://rosa.unipr.it/team.html)

[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 programme under grant agreement No 635761

Fig. 5 : GRA/ FSDA Matlab Model



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 635761

5. Dissemination of the tool

A webinar was prepared and recorded for future use the 11 October 2018. Here the programme of the webinar.

What is the Price Trend? PrimeFish and Price Development– 11 Oct		
Time	**Content**	**Responsible**
3'	Welcome and presentation of PrimeFish. Intro to tool + calendar of webinars & feedback	CETMAR
10'	Explanation of method, data and highlights	Uni. Parma
5'	Showcase with PrimeDSS - overview of results	Syntesa
10'	Q&A (including remote questions)	All

The Primefish project is properly acknowledged in the laboratory for robust statistics leaded by Prof. Riani (Ro.Sa, http://rosa.unipr.it/project.html).

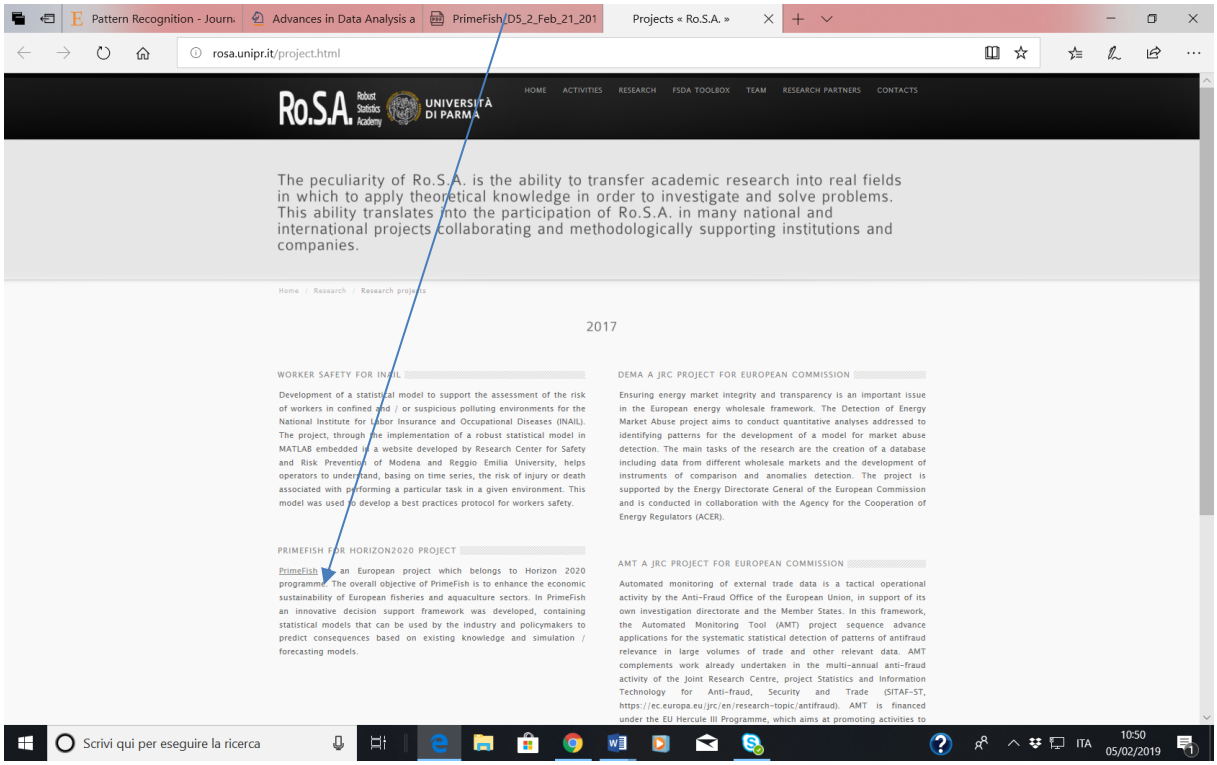


Fig. 6: Primefish project is properly acknowledged

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 635761

Results

Scientific Articles

The scientific articles included in the deliverable was co-authored by Marco Riani, University of Parma. The two papers are Open Access.

The first scientific article included in the deliverable was authored by Marco Riani (Corresponding Author) and has been available on-line from the 19th of November 2018:

RIANI M., ATKINSON A.C., CERIOLI A., CORBELLINI A. (2019). Efficient robust methods via monitoring for clustering and multivariate data analysis, Pattern Recognition, 88, p. 246–260. https://doi.org/10.1016/j.patcog.2018.11.016

Abstract: Monitoring the properties of single Sample robust Analyses of Multivariate data as a function of break- down point or efficiency Leads to The adaptive choice of the best values of these parameters, eliminating arbitrary decisions about their values and so increasing the quality of estimators.

The second scientific article was co-authored by Marco Riani and has been available on-line from the 30th of July 2018:

TORTI F. PERROTTA D., RIANI M., CERIOLI A (2018). Assessing Trimming Methodologies for Clustering Linear Regression Data, Advances in Data Analysis and Classification, https://doi.org/10.1007/s11634-018-0331-4

Abstract: We assess the performance of state-of-the-art robust clustering tools for regression structures under a variety of different data configurations. We focus on two methodologies that use trimming and restrictions on group scatters as their main ingredients. We also give particular care to the data generation process through the development of a flexible simulation tool for mixtures of regressions, where the user can control the degree of overlap between the groups. Level of trimming and restriction factors are input parameters for which appropriate tuning is required. Since we find that incorrect specification of the second-level trimming in the Trimmed CLUSTering REGression model (TCLUST-REG) can deteriorate the performance of the method, we propose an improvement where the second-level trimming is not fixed in advance but is data dependent. We then compare our adaptive version of TCLUST-REG with the Trimmed Cluster Weighted Restricted Model (TCWRM) which provides a powerful extension of the robust clusterwise regression methodology. Our overall conclusion is that the two methods perform comparably, but with notable differences due to the inherent degree of modeling implied by them.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 635761

A third paper applied to Fish Prices Data sets is under revision.

Title:

Detection of "boom and bust" cycles and making predictions for price using Efficient Robust Methods

Authors:

Marco Riani, Fabrizio Laurini, Gianluca Morelli, Cristina Mora – University of Parma

Abstract:

A well-known problem in the fishery industry is the ``boom and bust'' cycles, where high prices on market encourage individual firms to increase production significantly (boom); however, when all companies react in this manner the market is over-flooded, leading to overproduction, low prices and declining profits and bankruptcies (bust). Given the nature of the data available in fisheries market, the time series have some peculiarities, for example the presence of outliers, of level shift and often their depth is poor. Faced with these data features, we propose a study that aims to compare the Kalman filter, a widespread model for predictions, with an innovative model able to predict the fluctuations of prices when the time series are short and present outliers and level shift. The development of an effective forecasting tool allows manufacturers to improve their production plan thanks to the possibility of forecasting and protecting themselves against strong market price fluctuations.

To be submitted in March 2019 to Fisheries

Impact Factor: 3.000

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 635761

Acknowledgement

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 635761

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.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 635761

Annex (Open Access)

RIANI M., ATKINSON A.C., CERIOLI A., CORBELLINI A. (2019). *Efficient robust methods via monitoring for clustering and multivariate data analysis*, **Pattern Recognition**, 88, p. 246–260. https://doi.org/10.1016/j.patcog.2018.11.016

TORTI F. PERROTTA D., RIANI M., CERIOLI A (2018). Assessing Trimming Methodologies for Clustering Linear Regression Data, Advances in Data Analysis and Classification, https://doi.org/10.1007/s11634-018-0331-4