ITISE 2019

International Conference on Time Series and Forecasting

PROCEEDINGS OF ABSTRACTS

ITISE 2019
International Conference on Time Series and Forecasting

Proceedings of Abstract
25-27 September 2019
Granada (Spain)

Editors and Chairs
Olga Valenzuela
Fernando Rojas
Héctor Pomares
Ignacio Rojas

I.S.B.N: 978-84-17970-79-6
Legal Deposit: Gr 1210-2019
Edit and Print: Godel Impresiones Digitales S.L.

Preface

We are proud to present the set of final accepted papers for the 6th International conference on Time Series and Forecasting (ITISE 2019) held in Granada (Spain) during September, 25th-27th, 2019.

The ITISE 2019 seeks to provide a discussion forum for scientists, engineers, educators and students about the latest ideas and realizations in the foundations, theory, models and applications for interdisciplinary and multidisciplinary research encompassing disciplines of computer science, mathematics, statistics, forecaster, econometric, etc, in the field of time series analysis and forecasting.

The aims of ITISE 2019 is to create a friendly environment that could lead to the establishment or strengthening of scientific collaborations and exchanges among attendees, and therefore, ITISE 2019 solicits high-quality original research papers (including significant work-in-progress) on any aspect time series analysis and forecasting, in order to motivating the generation, and use of knowledge and new computational techniques and methods on forecasting in a wide range of fields.

The list of topics in the successive Call for Papers has also evolved, resulting in the following list for the present edition:

1. **Time Series Analysis and Forecasting**.
   - Nonparametric and functional methods
   - Vector processes
   - Probabilistic Approach to Modeling Macroeconomic Uncertainties
   - Uncertainties in forecasting processes
   - Nonstationarity
   - Forecasting with Many Models. Model integration
   - Forecasting theory and adjustment
   - Ensemble forecasting
   - Forecasting performance evaluation
   - Interval forecasting
   - Econometric models
   - Econometric Forecasting
   - Data preprocessing methods: Data decomposition, Seasonal adjustment, Singular spectrum analysis, Detrending methods, etc.
2. **Advanced method and on-Line Learning in time series**.
   - Adaptivity for stochastic models
   - On-line machine learning for forecasting
   - Aggregation of predictors
   - Hierarchical forecasting
   - Forecasting with Computational Intelligence
   - Time series analysis with computational intelligence

- Integration of system dynamics and forecasting models

3. **High Dimension and Complex/Big Data**.

   - Local Vs Global forecast
   - Techniques for dimension reduction
   - Multiscaling
   - Forecasting Complex/Big data

4. **Forecasting in real problem**.

   - Health forecasting
   - Telecommunication forecasting
   - Modelling and forecasting in power markets
   - Energy forecasting
   - Financial forecasting and risk analysis
   - Forecasting electricity load and prices
   - Forecasting and planning systems
   - Real time macroeconomic monitoring and forecasting
   - Applications in: energy, finance, transportation, networks, meteorology, health, research and environment, etc.

After a careful peer review and evaluation process (each submission was reviewed by at least 2, and on the average 2.9, program committee members or additional reviewer). In this proceedings we are presetting the abstract of the contribution to be presented during ITISE-2019 (accepted for oral, poster or virtual presentation,according to the recommendations of reviewers and the authors' preferences).

In this edition of ITISE, we are honored to have the following invited speaker:

1. Prof. Per Bjarte Solibakke, Professor and Associate Dean for Education, Faculty of Economics, Norwegian University of Science and Technology — NTNU Department of International Business.Vice Dean for Education, Faculty of Ecdonomics and Management, Department of International Business.

2. Prof. Thorsten Lehnert, Full professor of Finance. Luxembourg School of Finance (LSF)

3. Prof. Dieter Nautz, Professor Freie Universitt Berlin. Fachbereich Wirtschaftswissenschaft. Chair of Econometrics.

4. Prof. Dr. Stephan Schlter, Professor. Faculty of Mathematics, Natural and Economic Sciences University of Applied Sciences Ulm .

5. Prof. J. Hinaunye Eita, Professor and Head of Academics School of Economics College of Business and Economics . University of Johannesburg .

This new edition of ITISE was organized at the Universidad de Granada, with the help of the Spanish Network Time Series (RESET). We wish to thank to our main sponsor the institutions Faculty of Science, Dept. Computer Architecture & Computer Technology and CITIC-UGR from the University of Granada for their support. We wish also to thank to the Dr. Veronika Rosteck and Dr. Eva Hiripi, Springer, Associate Editor, for their interest in the future editing a book series of Springer from the best papers of ITISE 2019.

We would also like to express our gratitude to the members of the different committees and to the reviewer for their support, collaboration and good work.

September, 2019
Granada

ITISE Editors and Chairs
Olga Valenzuela
Fernando Rojas
Hector Pomares
Ignacio Rojas

.

ITISE 2019 Short Program

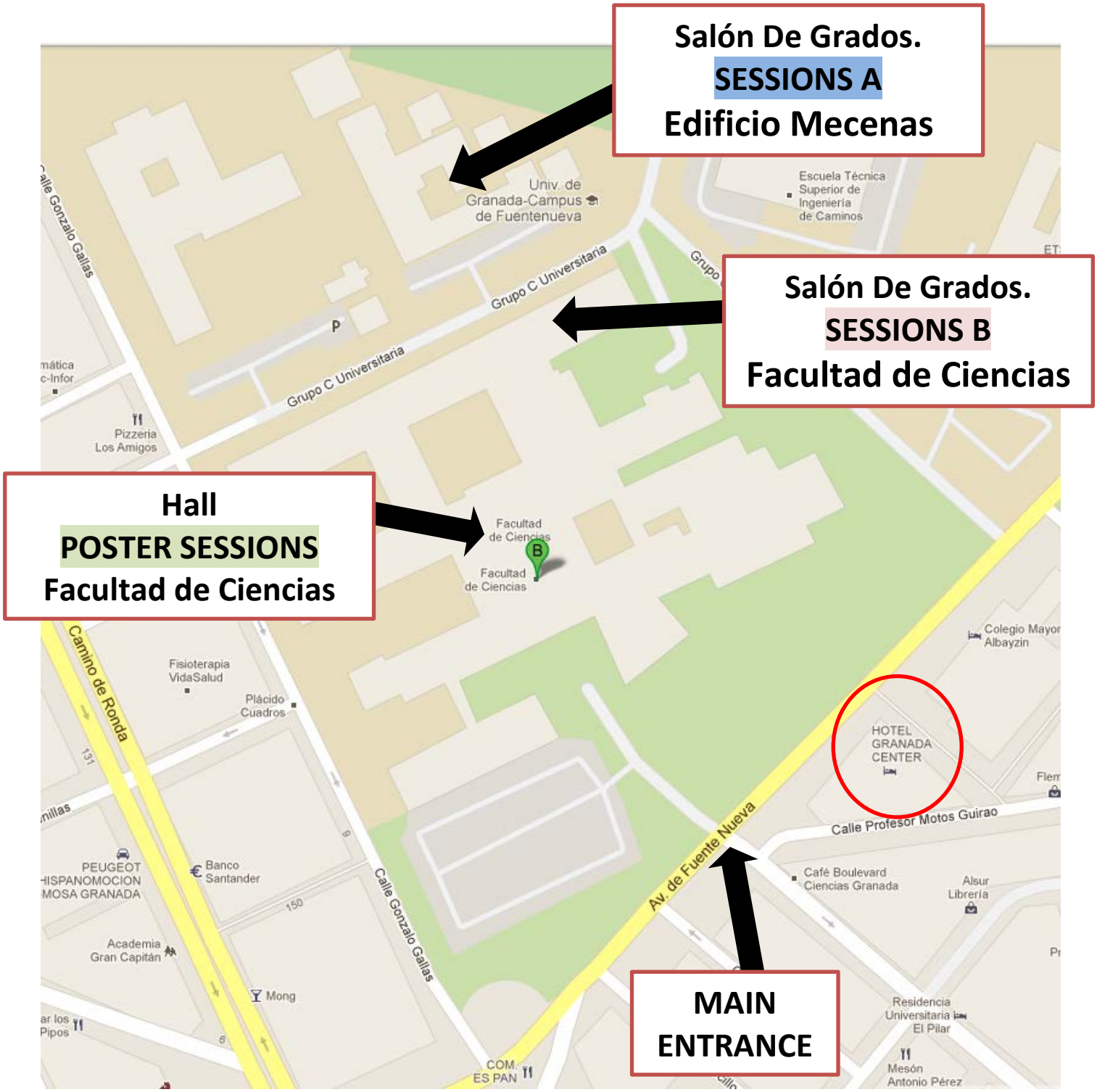
Wednesday, September 25th, 2019		
8:00-8:30	**REGISTRATION DESK** ***(start at 8h but it is opened during all the conference)***	
8:30-10:00	**Session A.1: Econometric models (Part I)**	**Session B.1: Time series analysis with computational intelligence**
10:00-10:30	**COFFEE BREAK**	
10:30-11:30	**OPENING PLENARY LECTURE.** **Prof. Per Bjarte Solibakke**	
11:30-12:45	**Session A.2: Nonstationarity time series (Part I)**	**Session B.2: Nonparametric and functional methods**
12:45-13:30	**Session A.3: Energy forecasting (Part I)**	**Session B.3: Forecasting theory and adjustment**
13:30-15:00	**LUNCH & COFFEE**	
15:00-16:30	**Session A.4: Dimension reduction techniques in Time Series**	**Session B.4: Real macroeconomic monitoring and forecasting (Part I)**
16:30-17:00	**COFFEE BREAK**	
17:00-18:00	**PLENARY LECTURE.** **Prof. Thorsten Lehnert**	
18:00-19:30	**Session A.5: Forecasting performance evaluation**	**Session B.5: Applications in Time Series (Part. I)**

NOTES:

- All **Sessions A** will be held in Salón de Grados, Edificio Mecenas (just 50 meters from the Facultad de Ciencias).
- All **Sessions B** will be held in Salón de Grados, Facultad de Ciencias.
- The **Poster Sessions** will be held in the Hall of Facultad de Ciencias.
- Social event (departure): Buses will be at the main entrance of Hotel Granada Center (26th September at 20:30 for the Gala Dinner at Hotel Alhambra Palace and 27th September at 17:00 for the visit to Alhambra).

Thursday, September 26th, 2019		
8:00-8:30	**REGISTRATION DESK** ***(start at 8h but it is opened during all the conference)***	
8:30-10:00	**Session A.6: Econometric Forecasting**	**Session B.6: Applications in Time Series (Part.II)**
10:00-10:30	**COFFEE BREAK**	
10:30-11:30	**PLENARY LECTURE.** **Prof. Dieter Nautz**	
11:30-12:45	**Session A.7: Atmospheric science forecasting**	**Session B.7: Forecasting with Many Models**
12:45-13:30	**Session A.8: Econometric models (Part II)**	**Session B.8: Financial forecasting and risk analysis**
13:30-15:00	**LUNCH & COFFEE**	
15:00-16:30	**Session A.9: Forecasting Complex/Big data (Part I)**	**Session B.9: Vector processes**
16:30-17:00	**COFFEE BREAK**	
17:00-18:00	**PLENARY LECTURE.** **Prof. Stephan Schlüter**	
18:00-19:00	**Session A.10: Nonstationarity time series (Part II)**	**Session B.10: Real macroeconomic monitoring and forecasting (Part II)**
19:00-20:25	**Session A.11/B.11: Poster Session.**	
20:30	**Gala Dinner at Hotel Alhambra Palace**	

Friday, September 27th, 2019		
8:00-8:30	**REGISTRATION DESK** ***(start at 8h but it is opened during all the conference)***	
8:30-10:30	**Session A.12: Data preprocessing methods in Time Series**	**Session B.12: Applications in Time Series (Part. III)**
10:30-11:00	**COFFEE BREAK**	
11:00-12:15	**Session A.13: Energy forecasting (Part II)**	**Session B.13: Applications in Time Series (Part. IV)**
12:15-12:50	**Session A.14: Forecasting Complex/Big data (Part II)**	
13:00-14:00	**CLOSING PLENARY LECTURE.** **Prof. J. Hinaunye Eita**	
14:00-17:00	**FREE TIME**	
17:00	**Visit to the Alhambra**	


Salón De Grados.
SESSIONS A
Edificio Mecenas
Salón De Grados.
SESSIONS B
Facultad de Ciencias
Hall
POSTER SESSIONS
Facultad de Ciencias
MAIN
ENTRANCE
Univ. de Granada-Campus de Fuentenueva
Escuela Técnica Superior de Ingeniería de Caminos
Grupo C Universitaria
Calle Gonzalo Gallas
Pizzería Los Amigos
Facultad de Ciencias
Camino de Ronda
Fisioterapia VidaSalud
Plácido Cuadros
Colegio Mayor Albayzín
HOTEL GRANADA CENTER
Calle Profesor Motos Guirao
Av. de Fuente Nueva
Café Boulevard Ciencias Granada
Alsur Librería
PEUGEOT HISPANOMOCION MOSA GRANADA
Banco Santander
Academia Gran Capitán
Mong
Residencia Universitaria El Pilar
Mesón Antonio Pérez
COM. ES PAN

Proceedings of Abstracts. ITISE-2019.Table of Contents

Session: Plenary Lecture

Session A.1: Econometric models (Part I)

Session B.1: Time series analysis with computational intelligence

Session A.2: Nonstationarity time series (Part I)

Session B.2: Nonparametric and functional methods

Session A.3: Energy forecasting (Part I)

Session B.3: Forecasting theory and adjustment

Session A.4: Dimension reduction techniques in Time Series

Session B.4:Real macroeconomic monitoring and forecasting (Part I)

Session B.6: Applications in Time Series (Part.II)

Session A.7: Atmospheric science forecasting

Session B.10: Real macroeconomic monitoring and forecasting (Part II)

Session A.11/B.11: Poster Session

Session A.12: Data preprocessing methods in Time Series

Session B.12: Applications in Time Series (Part. III)

Session A.13: Energy forecasting (Part II)

Session B.13: Applications in Time Series (Part. IV)

Session A.14: Forecasting Complex/Big data (Part II)

Session Virtual

Why is the market skewness-return relationship negative?

Thorsten Lehnert

Abstract: The observed negative relationship between market skewness and excess return in the time-series or the negative price of market skewness risk in the cross-section of stock returns is somewhat counterintuitive when we consider the usual interpretation of e.g. option-implied skewness as an indicator of jump risk or downside risk. One possible explanation for this inconsistency is that there are factors affecting option-implied market skewness other than jump risk in the stock market. In this paper, I find that price pressure associated with "crowded trades" of mutual funds is an important endogenous factor. Given that retail investors are prone to herding, the directional trading of mutual funds is correlated, and their collective actions can generate short-term price pressure on aggregate stock prices. Short sellers systematically exploit these patterns not only in the equity lending market, but also in the options market. In line with this economic channel, I find that firstly, the significant negative relationship between market skewness and returns becomes insignificant, once I control for price pressure. Secondly, the negative relationship is only present for the "bad" downside component of risk-neutral skewness, associated with out-of-the-money put options. For the "good" upside component of risk-neutral skewness, associated with out-of-the-money call options, the relationship is always positive. Thirdly, price pressure affects the skewness-return relationship, which can be clearly distinguished from the impact of flows on the volatility-return relationship in terms of the leverage effect.

Two Algorithms to Identify Outliers in Large Climate Time Series

Stephan Schlueter and Milena Kresoja

Abstract: Nowadays, most technical devices generate and store data in the form of time series. They are widely used in climatology, where e.g. water levels are collected for dike construction, in agriculture or the energy sector, for water management, and many other industries. Such time series help farmers, for example, to calculate the risk of frost in April, or an energy company to estimate the photovoltaic production, which - among other things - depends on the temperature. Most of the surface climate data such as air temperature, wind speed, or solar radiation are recorded in high frequency by automated instruments at many sites. These series are likely to contain outliers or irregular patterns caused e.g. by measurement errors, transmission errors, or systems changes, which, as a result, leads to false model specification, skewed parameter estimates, and forecasting errors. Data quality is crucial, which is exacerbated by the fact that climate is a complex system showing elements like high dimensionality, multiple seasonality, and serial dependence. From a practitioner's point of view, setting up a model is the easy part, it's the data preprocessing step which requires most attention and is the most time-consuming part of data analysis. Hence, having a reliable preprocessing method is essential. Literature offers a huge variety of methods for detecting anomalies in time series. For an overview, please refer to Gupta et al. (2014). Here we propose two further ones especially designed to treat clime time series. To overcome the problem of complexity of most algorithms, we first introduce a rather simple approach for anomaly detection based on autoregressive cost updates (ACU). The algorithm is easy to implement and to communicate, and it detects anomalies without the necessity of calibrating a model. Besides, we test a more sophisticated alternative, which is based on a method from signal processing called wavelet transform (see e.g. Mallat, 2003). Here we split up a time series into a linear combination of different frequencies and identify irregularities among the highest frequencies. We apply the model to temperature data. However, the model can be easily adapted to other climate time series such as humidity, precipitation, wind speed, but also to time series generated in other fields and applications. The performance of our proposed algorithms is tested on synthetic time series and benchmarked with existing preprocessing algorithms. Thereby we see that there is no method that performs best in all scenarios and for all tested quality criteria. If a time series is supposed to contain mostly solitary outliers, we recommend to apply the ACU algorithm. Otherwise, for multiple consecutive outliers, the wavelet-based algorithm is the best choice among the tested alternatives. Additionally, the algorithms are applied to a multivariate data set of temperatures at various sites in the City of Novi Sad, Serbia. We thereby

especially focus on identifying irregular patterns that are not clear outliers, i.e. hard to identify. We see that both algorithms effectively handle large data sets, which is the basis for applying further algorithms e.g. for clustering (i.e. dimension reduction) or modeling the data. Literature: Gupta M, Gao J, Aggarwal C, Han J (2014). Outlier detection for temporal data. Synthesis Lectures on Data Mining and Knowledge Discovery, 5(1); pp. 1-129. Mallat S. A wavelet tour of signal processing, 2nd ed. . Academic Press: Manchester, 2003.

Divisia Monetary Aggregates for a Heterogeneous Euro Area

Maximilian Brill, Dieter Nautz and Lea Sieckmann

Abstract: We introduce a Divisia monetary aggregate for the euro area that accounts for the heterogeneity across member countries both, in terms of interest rates and the decomposition of monetary assets. In most of the euro area countries, the difference between the growth rates of the country-specific Divisia aggregate and its simple sum counterpart is particularly pronounced before recessions. The results obtained from a panel probit model confirm that the divergence between the Divisia and the simple sum aggregate has a significant predictive content for recessions in euro area countries.

Stochastic volatility model's predictive relevance for Equity Markets

Per B Solibakke

Abstract: This paper builds and implements multifactor stochastic volatility models. The main objective is volatility prediction and its relevance for equity markets. The paper outlines stylised facts from volatility literature showing density tails, persistence, mean reversion, asymmetry and long memory, all contributing to systematic dependencies. Applying long simulations from stochastic volatility (SV) models and filter volatility using a form of nonlinear Kalman filtering, the unobservables of the nonlinear latent variables can be forecasted with associated fit characteristics. The paper uses European equity data from United Kingdom (Ftse100) and Norway (Equinor) for relevance arguments and illustrational prediction purposes. Multifactor SV models seem to enrich volatility predictions empowering equity market relevance.

Productivity and Real Exchange Rate: Investigating the Validity of the Balassa-Samuelson Effect in Five African Countries

Joel Hinaunye Eita, Zitsile Zamantungwa Khumalo and Ireen Choga

Abstract: Productivity in any economy is important because it increases economic performance and promotes economic growth. Given the importance of productivity, the study investigated the validity of the Balassa-Samuelson Effect in five African countries. The focal point of the Balassa-Samuelson (BS) effect is the relationship between productivity growth and real exchange rate appreciation. The study estimated the equilibrium real exchange with total factor productivity as an explanatory variable. Further, real exchange rate misalignment was derived and its effects on economic performance tested. The study employed more than one measure of economic performance in assessing the effects of real exchange rate misalignment on economic performance. The results revealed a valid Balassa-Samuelson effect in the selected countries and negative coefficients for real exchange rate misalignment. Recommendations emanating from this study include countries pursuing policies that contain misalignment of the real exchange rate because sustained misalignment hampers economic growth, performance and competitiveness.

Estimating the Equilibrium Real Exchange Rate, Misalignment and Economic Performance in Selected African Countries

Joel Hinaunye Eita, Zitsile Zamantungwa Khumalo and Ireen Choga

Abstract: This study estimates the equilibrium real exchange rate and the resulting misalignment for selected African countries. It then tests the impact of real exchange rate misalignment on economic performance. The fundamental approach model was used to estimate, the equilibrium real exchange rate and resulting misalignment. The results revealed that the real exchange rate was misaligned. Additionally, both negative and positive coefficients for real exchange rate misalignment for the different models and samples are revealed, indicating periods of undervaluation and overvaluation of the real exchange rate.

Low frequency estimation of Lévy-driven moving averages

Mikkel Slot Nielsen

Abstract: In this paper we consider least squares estimation of the driving kernel of a moving average and argue that, under mild regularity conditions and a decay condition on the kernel, the suggested estimator is consistent and asymptotically normal. On one hand this result unifies scattered results of the literature on low frequency estimation of moving averages, and on the other hand it emphasizes the validity of inference also in cases where the moving average is not strongly mixing. We assess the performance of the estimator through a simulation study.

Backtesting Basel III: Evaluating the Market Risk of Past Crises through the Current Regulation

Marcelo Zeuli and André Carvalhal

Abstract: Are the recommendations from the Bank for International Settlements (BIS) effective to a broad set of financial crises? We submitted two of the main Basel III recommendations for market risk to a back test: the capital requirements and the Value at Risk (VaR) methodology that includes the BIS's Stressed VaR. We tested the main Brazilian currency exchange (U.S. Dollar to Brazilian Reais) and currency exchange swaps contracts through volatility-based VaR methodologies in the period that comprises the so-called Brazilian confidence crisis, which occurred in the second half of 2002. While the Stressed VaR revealed inapplicable, due to historical data shortage, the capital requirements level appeared innocuous, due to the high levels of daily volatility - daily oscillation limits may have a significant role on crisis mitigation. To circumvent the lack of either historical information or optimal window for stress patterns, we suggest to calibrate the Stressed VaR or the recently announced Expected Shortfall with a historical VIX (Volatility Index, Chicago Board Options Exchange), working as a volatility scale. We suggest modeling with other densities, apart from the BIS recommended standard normal.

Testing normality for unconditionally heteroscedastic macroeconomic variables

Hamdi Raissi

Abstract: In this paper the testing of normality for unconditionally heteroscedastic macroeconomic time series is studied. It is underlined that the classical Jarque-Bera test (JB hereafter) for normality is inadequate in our framework. On the other hand it is found that the approach which consists in correcting the heteroscedasticity by kernel smoothing for testing normality is justified asymptotically. Nevertheless it appears from Monte Carlo

experiments that such a methodology can noticeably suffer from size distortion for samples that are typical for macroeconomic variables. As a consequence a parametric bootstrap methodology for correcting the problem is proposed. The innovations distribution of a set of inflation measures for the U.S., Korea and Australia are analyzed.

Regional Development and Inequalities in Latin American Countries: Econometric Analysis

Evgeniya Muzychenko

Abstract: For centuries, problems of socio-economic development of Latin American countries, in particular, the issues of poverty and income inequality have been a central topic in the studies. At the same time, the aspect of regional (sub-national) development of Latin America appears to be quite new and understudied as there is a lack of literature on this topic. Therefore, the aim of this research is to examine the dynamics of regional development from 2000 to 2016 and determine the main factors of regional growth of Latin America. In order to analyze regional development and assess the dynamics, the classification of selected countries' regions will be conducted. The results of the study have shown that despite the improvements in multiple socio-economic indicators, the problem of regional inequality is still relevant for the countries in the region, including the most developed ones. Moreover, some factors of territorial development were detected and the hypothesis of industrial regional growth was confirmed.

Structural stability of infinite-order regression

Abhimanyu Gupta and Myunghwan Seo

Abstract: We develop a class of tests for the structural stability of infinite order regression models, when the time of a structural change is unknown. Examples include the infinite order autoregressive model, the nonparametric sieve regression and many others whose dimensions grow to infinity. When the number of parameters diverges, the traditional tests such as the supremum of Wald, LM or LR statistic or their exponentially weighted averages diverge as well. However, we show that a suitable transformation of these tests converges to a proper weak limit as the sample size n and the dimension p grow to infinity simultaneously. In general, this limit distribution is different from the sequential limit, which can be obtained by increasing the order p of the standardized tied-down Bessel process in Andrews (1993). More interestingly, our joint asymptotic analysis discovers that the joint asymptotic distribution depends on a higher order serial correlation. We also establish a weighted power optimality property of our tests under certain regularity conditions. A new result on partial sums of random matrices is established. We examine finite-sample performance in a Monte Carlo study and illustrate the test with a number of empirical examples.

Customers of Future: How do They Spend their Bitcoins

Huber Nieto-Chaupis

Abstract: Once the Bitcoins are established in the worldwide economy, from the point of view of markets would appear the question: what are the products that a potential customer would acquire with Bitcoins. Particularly developing economies might have not clear what market or product follow to. In this paper we present a Monte Carlo simulation that allows us to have a rough idea about the preferences of future customers. Thus, our results would indicate that electronics might be a favorable market for Peruvians with a confidence of order of 97%.

Mimicking the Mechanisms of Language for the Unsupervised Detection of Hierarchical Structure in Time Series

Christopher Josef Rothschedl, Paul O'Leary and Roland Ritt

Abstract: The purpose of this paper is to investigate, how mechanisms of natural language can support the analysis of time series data emanating from human-operated physical systems. There may be no physical models to describe human behavior reliably; nevertheless, there is commonly structure in the data acquired from observing this behavior. This paper focuses on investigating a new approach to the unsupervised detection of structure in such data. Symbolic analysis and linear differential operators are adopted, to derive results bearable for resemblance with the metaphorical concept of language and its mechanisms. To support this, phenomenology is introduced and discussed. The general significance of metaphors - and why they may be of specific relevance to analyzing time series data - is discussed and references to exemplary applications are presented to ascertain the validity of the approach.

Prediction of Transformer Temperature for Energy Distribution Smart Grids Using Recursive Neural Networks

Francisco Jesús Martínez-Murcia, Javier Ramirez, Fermin Segovia, Andres Ortiz, Susana Carrillo, Javier Leiva, Jacob Rodriguez-Rivero and Juan Manuel Gorriz

Abstract: The near future of energy is conformed by a plethora of heterogeneous sources as well as an increasing demand. This poses new challenges for energy production and distribution, in which it will be essential that Medium Voltage/Low Voltage (MV/LV) distribution networks are planned, operated and monitored in an analogous way to what transport networks have been doing for decades. Within this context, an accurate tool for anticipating transformer overload and potential network problems is of paramount importance. Here, a system that can predict transformer temperature -a critical indicator of potential problems in a transformer. is key to the development of versatile and autonomous network control strategies that allow for a smarter distribution of energy. In this work we propose a transformer temperature prediction system based on long short-term memory (LSTM) networks that uses data from the previous 100 minutes to predict the transformer temperature in the next 100 minutes. The system is able to predict with a low error the temperature value using only the active power of the three transformer lines together with the ambient temperature. This makes it possible to discover trends towards anomalous temperature values in different transformers and act accordingly by planning a redistribution of the workload, avoiding possible incidents or service interruptions.

Knowledge Extraction (KnoX) in Deep Learning: Application to the Gardon de Mialet Flash Floods Modelling

Bob E. Saint Fleur, Guillaume Artigue, Anne Johannet and Severin Pistre

Abstract: Flash floods frequently hit Southern France and cause heavy damages and fatalities. To better protect persons and goods, official flood forecasting ser-vices in France need accurate information and efficient models to optimize their decision and policy. Since heavy rainfalls that cause such floods are very heterogeneous, it becomes a serious challenge for forecasters. Such phenomena are typically nonlinear and more complex than classical floods events. That problem leads to consider complementary alternatives to en-hance the management of such situations. For decades, artificial neural net-works have been very efficient to model nonlinear phenomena, particularly rainfall-discharge relations in various types of basins. They are applied in this study with two main goals: first modelling flash floods on the Gardon de Mialet basin; second, extract internal information from the model by using the Knowledge eXtraction method to provide new ways to improve models. The first analysis shows that the kind of nonlinear predictor influences strongly the representation of

information: e.g. the main influent variable (rainfall) is more important in the recurrent and static models than in the feed-forward one. For understanding flash floods genesis, recurrent and stat-ic models appear thus as better candidates, even if their results are better.

The Study of Recurrent Neuron Networks based on GRU and LSTM in Time Series Forecasting

Tatiana Afanasieva and Pavel Platov

Abstract: When forecasting time series, artificial neural networks are often being used. Recently, interest in recurrent neural networks with GRU or LSTM units, which are considered as deep learning tools, has increased. Compared to LSTM, the GRU is characterized by some simplifications, leading to the study of a smaller number of weights. Despite the interest in recurrent neural networks with GRU or LSTM units, their ability to predict one-dimensional time series has not been adequately explored. The article focuses on a comparative study of the prediction accuracy of the above mention configurations of recurrent neural networks in dependence of the time series behavior and regardless of their behavior. The study was conducted on a dataset of 30 artificial time series. A methodology for such an investigation and corresponding software for time series modelling and forecasting have been proposed. The obtained results allowed us to expand our understanding of the applicability of recurrent neural networks with GRU or LSTM units in time series forecasting.

Optimal and Efficient Model Selection Criteria for Parametric Spectral Estimation

Abass Taiwo

Abstract: Traditional Parametric spectral estimation methods have been widely used to obtain spectral estimate, resolution and variance distribution in signals or time series data across several fields. But a problem of how to choose an optimal and efficient model order selection was encountered. In this research work, modified forms of Final Prediction Error, Akaike Information Criteria, Bayesian Information Criteria and Minimum Description Length which involved the replacement of variance error with sample autocorrelation function that has capabilities of detecting non-randomness in time series data were proposed, to choose an optimal and efficient model order selection. The results of the modified and traditional information criteria were used to choose AR(11) and AR(7) as the optimal model. The spectral estimate and resolution of AR(11) indicated the modified methods outperformed the traditional methods for analyzing Heart-beat readings. Conclusively, the proposed method outperformed the traditional Information criteria for selecting optimal and efficient model order for Heart-beat readings with lower values of parameters.

Forecasting Stock Market Data using a Hybrid EMD-HW Method

Ahmad Awajan and Sadam Al Wadi

Abstract: In this study, a hybrid method based on Empirical Mode Decomposition and Holt-Winter (EMD-HW) is used to forecast stock market data. First, the data are decomposed by EMD method into Intrinsic Mode Functions (IMFs) and residual components. Then, all components are forecasted by HW technique. Finally, forecasting values are sum together to get the forecasting value of stock market data. Empirical results showed that the EMD-HW outperform individual forecasting models. The daily Spain stock market time series data are applied to show the forecasting performance of the EMD-HW. The strength of this EMD-HW lies in its ability to forecast non-stationary and non-linear time series without a need to use any transformation method.

Moreover, EMD-HW has a relatively high accuracy comparing with eight existing forecasting methods based on the five forecast error measures.

The Non-Stationary INARMA(1,1) Model with Generalized Innovation

Yuvraj Sunecher

Abstract: This paper considers modelling of a non-stationary integer- valued autoregressive moving average of order 1 (INARMA(1,1)) model by assuming that the innovation follows a Poisson and negative binomial distribution. Two simulation experiments are also conducted to assess the performance of conditional maximum likelihood (CML) and generalized quasi-likelihood (GQL) estimation methods.

Numerical Study of the Conditional Time Series of the Average Daily Heat Index

Nina Kargapolova

Abstract: In this paper results of the numerical study of the statistical properties of the con-ditional time series of the average daily heat index (ADHI) are presented. Both point and interval conditions on the values of the ADHI were imposed. The study was carried out on a basis of the real data as well as the stochastic models of the non-stationary non-Gaussian conditional time series of the ADHI. To simulate the time series with point conditions, the conditional distributions method and the inverse distribution method were applied. In case of interval conditions, the straightforward enumeration of the non-conditional trajectories was used. It is shown that the trajectories of the models proposed are close in their statistical properties to the real time series of the ADHI. Simulated trajectories of the ADHI are used to study statistical properties of some unfavorable (in sense of the ADHI) weather events.

Real time prediction of irregular periodic time series data

Chi Tim Ng

Abstract: By means of a novel time-dependent cumulated variation penalty function, a new class of real-time prediction methods is developed to improve the prediction accuracy of time series exhibiting irregular periodic patterns, in particular, the breathing motion data of the patients during the robotic radiation therapy. It is illustrated that for both simulated and empirical data involving changes in mean, trend, and amplitude, the proposed methods outperform existing forecasting methods based on support vector machines and articial neural network in terms of prediction accuracy. Moreover, the proposed methods are designed so that real-time updates can be done efficiently with O(1) computational complexity upon the arrival of a new signal without scanning the old data repeatedly.

New test for a random walk detection based on the arcsine law

Konrad Furmańczyk, Marcin Dudziński and Arkadiusz Orłowski

Abstract: In our work, we construct a new statistical test for a ran- dom walk detection, which is based on the arcsine laws. Additionally, we consider a version of the unit root test for autoregressive process of order 1, which is also related to the arcsine laws. We also conduct some simulation study in order to check the quality of the proposed test.

Analysis of non-stationary time series based on modelling stochastic dynamics considering self-organization, memory and oscillations

Dmitry Zhukov, Tatiana Khvatova and Leonid Istratov

Abstract: Behaviors of stock and raw material market indexes are most often described using non-stationary time series. The processes observed therein show the potential for self-organization and the presence of memory of previous events having occurred within the system, both conditioned by the human factor. On the one hand, the presence of numerous players engenders uncertainty about the effects on processes and creates stochasticity. However, on the other hand, it also creates opportunities for self-organization, and moreover predefines the presence of memory about previous activities. Therefore, the present authors argue that stock exchange processes cannot be fully explained using the theory of chaos, and this should be considered when developing forecasting models. Consequently, this paper proposes a newly developed model of stochastic dynamics for forecasting stock indexes considering self-organization pro- cesses, the presence of memory and possible oscillations.

From Long Memory to Oscillatory Modes - The Potentials of Detrended Fluctuation Analysis

Philipp G. Meyer and Holger Kantz

Abstract: Detrended fluctuation analysis is a popular method for the detection of long range correlations in time series. Recently further insights were gained which allow a theoretical discussion of its results. Since then the method is no longer restricted to long memory, but can explain the whole correlation pattern of the signal and enables us to infer characteristic timescales. We show its traditional usage in a toy model for long range correlations and thereby discuss the relation to other anomalous statistical properties. Then we summarize the uncovered potentials of detrended fluctuation analysis for short range correlated systems showing examples from atmospheric science.

The correspondence between stochastic linear difference and differential equations

D. Stephen G.Pollock

Abstract: The relationship between autoregressive moving-average (ARMA) models in discrete time and the corresponding models in continuous time is examined in this paper. The linear stochastic models that are commonly regarded as the counter- parts of the ARMA models are driven by a forcing function that consists of the increments of a Wiener Process. This function is unbounded in frequency. In cases where the periodogram of the data indicates that there is clear up- per bound to its frequency content, we propose an alternative frequency-limited white-noise forcing function. Then, there is a straightforward translation from the ARMA model to a differential equation, which is based of the principle of impulse invariance. Whenever there is no perceptible limit to the frequency content, the trans lation must be based on a principle of autocovariance equivalence. On the website of the author, there is a computer program that effects both of these discrete-to-continuous translations.

Multifractal Detrended Fluctuation Analysis combined with Singular Spectrum Analysis

Anton Karmatskii

Abstract: This paper addresses the problem of finding appropriate method for analysis of non-stationary time series with complex trends and possibly with distinct periodic components in power spectrum. A well established Multifractal Detrended Fluctuation Analysis (MDFA) combined with Singular Spectrum Analysis (SSA) used on several artificial examples to demonstrate the capabilities of the method. In this combined method SSA is used for nonparametric periodic components extraction and for adaptive automatic nonparametric trend extraction. As usual, the local fractal characteristics of the signal are studied by utilizing the MFDFA algorithm. It is shown that combined method is capable of accurately extracting fractal features when signal is contaminated with broadband and narrowband noise. The main task that the author focused on when creating the method was the analysis of currents through ion channels in cell membrane.

Metamodeling Based Approach for District Heat Network Aggregation

Nihad Aghbalou

Abstract: To deal with the problem of (DHN) modeling, this manuscript introduce a refinement approach to develop a novel method for district heat network aggregation based on soft computing techniques (Neural Networks and metaheuristics). Reducing size and complexity of DHN by preseving the dynamic properties of district heating whitout physical description (network topology, pipe diameter and insulation, pump characteristics etc.) is a main challenge for simulating various operational strategies. Data from reel DHN system was used to investigate the ability of the proposed approach. This approach embed the time of flow transportation in the inputs of the model. The time of flow transportation between production plant and consumers substations is dermined using the cross wavelet transform. The results show that the proposed model leads to acceptable values but this can be more enhanced (i.g. by considering the quality of the measured dada) and then finds application for DHN modeling.

Theoretical foundation of detrending methods for fluctuation analysis such as detrended fluctuation analysis and detrending moving average

Marc Höll, Ken Kiyono and Holger Kantz

Abstract: We present a bottom-up derivation of methods for fluctuation analysis with detrending and claim their basic principles. Such methods detect long-range correlations in time series even in the presence of additive trends or intrinsic nonstationarities. Long-range correlations which are often characterized by a power law decaying autocorrelation function are omnipresent in a tremendous amount of data sets. But especially in nonstationary time series, their existence has a non-negligible impact on the analysis, modeling and prediction of the time series. Therefore advanced methods are necessary to extract information about the correlation structure. The well-known detrended fluctuation analysis (DFA) and detrending moving average (DMA) are such methods but were introduced ad hoc. Both methods sample over mean-squared errors between the summed time series and its fitting polynomials in different time windows. We show that DFA and DMA fulfill our claimed basic principles of the general framework of detrending methods for fluctuation analysis. The bottom-up derivation of these principles starts with the observation that the standard sample estimator of the autocorrelation function has at least two fundamental statistical problems. First, the estimator fluctuates strongly around zero for large time lags which makes it difficult to observe a possible power law decay. Second, the estimator is only meaningful for stationary time series. Therefore we consider instead the mean-squared displacement of the summed time series. It contains the same information about long-range correlations as the

autocorrelation function but overcomes the first problem as it is an increasing function depending on the time lag. However, the estimation of its scaling behavior on a single time series is affected not only by additive trends on the data but also by intrinsic nonstationarities. Exemplary, fractional Brownian motion without additive trends is an intrinsic nonstationary process. We show in detail, how the estimation of the mean-squared displacement of the summed time series is disturbed by these two types of nonstationarity. We relate this estimator with the autocorrelation function of the random signal of the time series so that we are able to identify analytically the reason of failure of correct estimation. If we write the mean-squared displacement of the summed time series with respect to the autocorrelation function we call this the increment representation as the time series is the increment process of the summed time series. A nonstationary process either having additive trends or being only fractional Brownian motion has a nonstationary autocorrelation function which therefore introduces a bias in the estimation of the mean-squared displacement of the summed time series. This bias is exactly the reason why stationary detrending methods such as fluctuation analysis (FA) or DFA with zero order detrending always observe the same scaling behavior in the presence of nonstationarities no matter what random process is underlying. To erase the influence of the bias we introduce a weighting kernel for the mean-squared displacement of the summed time series in the increment representation which is the only possible presentation to formulate a general form of the weighting kernel. We call this function the fluctuation function of the time series and will later show that DFA and DMA are specific examples with each having their own weighting kernel. Now we can claim two basic principles of detrending methods for fluctuation analysis. First, we claim that the fluctuation function scales like the mean-squared displacement of the summed time series. We preserve the information of the latter. Second, we claim that the estimator of the fluctuation function is unbiased even in the presence of nonstationarities. Hence the weighting kernel removes the influence of the bias during the estimation procedure. This is called detrending in methods for fluctuation analysis which is usually introduced by removing fitting polynomials from the summed time series. Our novel and more general understanding of detrending provides a unified picture for different types of nonstationarity and especially explains detrending for the intrinsic nonstationary fractional Brownian motion. Both principles define the general framework for detrending methods and serve as basis to derive important characteristics of them such as the following three. First, we can derive a superposition principle assuming that the time series consists of independent additive processes so that then the fluctuation function is the sum of individual fluctuation functions. Second, we can relate our theory to the field of wavelet transforms as a factorizable weighting kernel of the fluctuation function is identical to the Haar wavelet transform. And third, our framework allows to formulate the practical implementation of detrending methods in a general way where only the specific form of weighting kernel is optional. Here DFA and DMA are suitable candidates because both are able to detect the correct scaling behavior in the presence of nonstationarities. However, their original description is different than above general framework meaning their original forms of the fluctuation function are not written in the increment representation. For both DFA and DMA we analytically derive their weighting kernels and show that above two principles are indeed fulfilled. Therefore DFA and DMA are examples of detrending methods. Although both methods are different their basic techniques of detrending works very similar. The main difference between them is the method of sampling. But no matter what sampling procedure one chooses the influence of the bias introduced by nonstationarities is oppressed by their weighting kernels. The knowledge of the weighting kernel allows us not only to show the fulfillment of the basic principles but furthermore also to calculate analytically the fluctuation function for different stochastic models. Notably, for autoregressive models the fluctuation function exhibits a crossover behavior in scaling. This crossover behavior demands long enough data sets which is often not the case in real applications. Summarized, we provide a theoretical framework of detrending methods for fluctuation analysis which serves as basis to investigate many application problems. This framework also allows to find new methods with more suitable weighting kernels for certain problems in fluctuation analysis of nonstationary time series.

Seasonal Models for Forecasting Day-Ahead Electricity Prices

Catherine McHugh, Sonya Coleman, Dermot Kerr and Daniel McGlynn

Abstract: Prediction algorithms are increasingly popular techniques within the energy industry to assist in forecasting future prices and to help reduce costs. When predicting day-ahead energy prices several significant

external factors that influence electricity prices need to be considered. Initially we use the transparent Nonlinear AutoRegressive Moving Average model with eXogenous inputs (NARMAX) to model the relationship of factors that contribute to energy costs and day-ahead electricity prices. Energy data from 2017 were analyzed separately for Spring, Summer, and Autumn periods to observe the effect of factors on seasonality. The seasonal NARMAX models identify important input factors and their correlated lagged values such as price, energy demand, generation type, and environmental temperature to be significant factors in accurately predicting day-ahead prices for all three seasons. The identified factors are used to refine a Seasonal AutoRegressive Integrated Moving Average model with eXogenous variables (SARIMAX). To conclude the seasonal NARMAX models proved useful for removing insignificant external factors and the seasonal refined SARIMAX models permitted accurate prediction of day-ahead prices.

A Lotka-Volterra model for diffusion of electric vehicles in the US: competition and forecasting

Mariangela Guidolin

Abstract: The transport sector requires a reduction in CO2 emissions, so that industry, policy makers and researchers are forced to think about the diffusion of plug-in electric vehicles. Market forecasting is a well-developed field of study, but in the electric vehicle domain this is a complex task due to the relative newness of the market. In this paper, we propose and apply a new Lotka-Volterra model to the diffusion of electric vehicles in the US market. Specifically, we study the competitive interaction between the two main market players, namely Tesla and Nissan Leaf, by considering the time series of monthly sales. After a suitable model reduction, the results of the selected model statistically confirm a significant interaction between the two and indicate that Tesla has been able to cannibalize the market of Nissan Leaf. At the same time, the model suggests that Nissan Leaf exerted a collaborative effect towards Tesla, rather a competitive one. The analysis is completed with short term forecasts and a SARMAX refinement which accounts for seasonal and autodependent components.

Extreme Value Analysis of Power System Data

Per Westerlund and Wadih Naim

Abstract: In the electric system, the consumption varies throughout the year, during the week and during the day. The consumption should be balanced by the production, which is not easy with solar power and wind power, as they lack storage like the dams for hydropower. Then these sources should be modelled as time series. The time series analyzed is the solar and wind power production in Sweden during 5 months. The method is called Peaks over Threshold or POT and it calculates the frequency of peaks above a certain threshold. It also determines the distribution of the size of the peaks, which is the generalized Pareto distribution in this case. Two different clustering methods are tried; one by Lindgren and the other one a modification of Leadbetter's method. The latter one gives the best fit. However, some ten years of data should be analyzed in order to include the seasonal effects.

Reconstruction of the transition probability density function from persistent time series

Zbigniew Czechowski

Abstract: In some models (e.g., diffusion Markov models) the probabilistic forecast is given by the transition probability density function. For the case of the standard Langevin equation (which describes a time evolution of Markov processes) Siegert et al. (1998) introduced the numerical procedure of reconstructing from time series

the drift and diffusion functions that are creating the transition probability density for small times. Then the method was developed in other papers (see Friedrich et al. 2011). However, the standard Langevin equation describes Markov processes only. In this work we introduce the generalized discrete Langevin equation for some class of non-Markov processes, namely for persistent time series of order p. Therefore, non-local effects must be considered. We assume that the next state of the process is dependent not only on the present state but also on signs of p previous jumps. To this aim, the standard discrete Langevin equation is modified by introducing a new random function which determines the sign of the diffusion term. The function depends on the vector random variable (i.e., the chain of p previous signs), the random scalar variable with the uniform distribution in [0, 1] and on the vector persistence parameter (with 2^p components). The term is keeping the tendency of increase/decrease of the process in the next step according to given persistence parameters. When all the parameters are equal to 0.5 the modified equation reduces to the standard Langevin equation. The proposed model is a significant extension of our previous approach (Czechowski 2016) in which persistent processes of order p = 1 were taken into account. The generalization opens a wide possibilities of nonlinear modeling of data in which persistence and antipersistence of different orders can be mixed in a time series under investigation. In order to construct the transition probability density function the forms of drift and diffusion functions are needed. The standard procedure (Siegert et al. 1998, Friedrich et al. 2011) of reconstruction of the Langevin equation from time series leads to the proper estimation of the diffusion function but to the wrong reconstruction of the drift function in the case of the modified equation for persistent processes. To estimate the deviation in the drift we propose a new reconstruction procedure in which the standard procedure is used three times. The algorithm can be summarized in five steps. At the beginning the vector persistence parameter is estimated by using a histogram method. In the second step the standard procedure is applied to the input persistent time series. As the result first reconstructions of the diffusion function and the drift function are obtained, however, the reconstructed drift is deviated from the input drift. In the next step a new time series generated by the modified Langevin equation with the estimated persistence parameter and first reconstructions of the drift and diffusion functions is treated as the input to the second use of standard procedure. In this way the second reconstruction of the drift function is obtained. The fourth step of the procedure is similar to the third step, however, here the second reconstruction of the drift (instead of the first reconstruction) is applied. At the result the third reconstruction of the drift function is found. The last step of the procedure bases on the assumption (which has been verified analytically for the case p = 1) that ratios of deviations are dependent on the persistence parameter only. Therefore, the ratio calculated from deviations found in the third and fourth step is used for estimation of the deviation between the unknown input drift function and the first reconstruction. This leads to the final reconstruction of the drift function from the input p-persistent time series. In order to test an efficiency of the procedure many time series were generated by using the modified Langevin equation with different drift and diffusion functions and different persistences. This enables to compare the input functions and parameters to the reconstructed ones. A good efficiency of the modified reconstruction procedure has been shown. Having the proper forms of reconstructed drift and diffusion functions enables derivation of the short-time transition probability density function. For the case of Markov processes the function has a Gaussian form, however non-Markovian features of persistent processes make the problem more complex. Therefore, a correction term appears in the formula for short-time transition probability density function. For the persistence of order 1 (i.e., for p = 1) the correction term can be derived analytically but for higher orders numerical estimations are necessary. It should be underlined that the presented method applied to forecasting time series generates a probability distribution for the next point, rather than a single point estimate as in autoregression. The parameters in the modified Langevin model are dynamically estimated from past data. An important advantage of the proposed approach is that it offers simultaneously the reconstruction of the stochastic model of the phenomena under investigation and the method of probabilistic forecast. Acknowledgements This work has been financed by the project of the National Science Centre (contract No. 2016/21/B/ST10/02998). 1. Siegert S., R. Friedrich, and J. Peinke, 1998, Analysis of data sets of stochastic systems, Phys. Lett. A 243, 275-280. 2. Friedrich R., Peinke J., Sahimi M., Reza Rahimi Tabar M., 2011. Approaching complexity by stochastic methods: From biological systems to turbulence, Physics Reports 506, 87-162. 3. Czechowski Z., 2016, Reconstruction of the modified discrete Langevin equation from persistent time series, CHAOS 26, 053109.

A covariance function for time dependent Laplacian fields in 3D

Gyorgy Terdik

Abstract: A homogeneous and isotropic Laplacian field is considered. We compare the decays of the covariance functions of an AR model in 1D, 2D and 3D. It is well known that the decay is exponential for a stationary AR time series and it has slower decay in 2D by a celebrated result of Whittle. We show that the decay is again exponential in 3D. A covariance function for spatio-temporal 3D-Lapalcian fields will also be considered in time-frequency domain, see Subba Rao -Terdik: A New Covariance Function and Spatio-Temporal Prediction, JTSA 12017.

Do Google Trends Forecast Bitcoins? Stylized Facts and Statistical Evidence

Argimiro Arratia and Albert López Barrantes

Abstract: In early 2018 Bitcoin prices peaked at US\$ 20,000 and, almost two years later, we still continue debating if cryptocurrencies can actually become a currency for the everyday life or not. From the economic point of view, and playing in the field of behavioral finance and sentiment analysis, this paper analyses the relation between Bitcoin prices and the search interest on “Bitcoin” since 2014. We questioned the forecasting ability of Google Bitcoin Trends for the behavior of Bitcoin price by performing linear and nonlinear dependency tests, and exploring performance of ARIMA and Neural Network models enhanced with this social sentiment indicator. Our analyses and models are founded upon a set of statistical properties common to financial returns that we establish for Bitcoin, Ethereum, Ripple and Litecoin.

Random Forest-controlled Sparsity of High-Dimensional Vector Autoregressive Models

Dmitry Pavlyuk

Abstract: This paper introduces a new random forest-based approach to control the sparsity of vector autoregressive (VAR) models. VAR models are the popular forecasting tool in many applied areas, but high dimensionality of data is widely acknowledged as a significant hurdle for their efficient application. Sparse specifications of VAR models are designed to overcome this problem and keep the model specification parsimonious and interpretable. We consider the control of VAR models' sparsity as a special case of the feature filtering problem and propose application of the random forest learning for its solution. The proposed approach includes preliminary ranking of features (model autoregressive and cross-sectional lags) by the random forest algorithm and further application of obtained feature importance values for a sparse specification of the VAR model. We tested the proposed approach against other specifications of VAR models' sparsity (refined VAR, penalized VAR) using the real-world urban traffic data set and demonstrated its statistical advantages: higher and more stable forecasting accuracy, manageable level of sparsity, and good computational performance for extremely high-dimensional time series.

Unsupervised Anomaly Detection in Time Series with Convolutional-VAE

Emanuele La Malfa and Gabriele La Malfa

Abstract: We propose an unsupervised machine learning algorithm for anomaly detection that exploits self-learnt features of monodimensional time series. A Variational Autoencoder, where convolution takes place of dot product, is trained to compress each input to a low-dimensional point from a normal distribution, detecting

an anomaly as low probability and high density sequence. We validate our work on different public datasets, obtaining results that shed new light on Variational Autoencoders applied to anomaly detection.

Feature Selection based Multivariate Time Series Forecasting: An Application to Antibiotic Resistance Prediction

Jose Palma, Fernando Jimenez, Gracia Sánchez, David Marín García, Francisco Palacios and Lucía López-Rodríguez

Abstract: In this paper we propose a methodology to build a model for predicting future outbreaks of \textit{Methicilin-resistant Staphylococcus aereus} (MRSA). Infection incidence forecasting is approached as a \textit{feature selection based time series forecasting} problem using multivariate time series composed of incidence of \textit{Staphylococcus Aereus} and MRSA infections, influenza incidence and total days of therapy of both of \textit{Levoflaxin} and \textit{Oseltamivir} antimicrobials. Data were collected from the University Hospital of Getafe (Spain) from January 2009 to January 2018, using months as time granularity. The proposed methodology includes the application of wrapper multivariate feature selection methods on transformed datasets with a different number of lagged variables, where the search strategy is based on \textit{multi-objective evolutionary algorithms} and the evaluators are based on the most powerful state-of-the-art regression algorithms. The performance of the feature selection methods has been measured using both \textit{root mean square error} ($\mathcal{RSME}$) and \textit{mean absolute error} ($\mathcal{MAE}$) metrics. Finally, in order to select the most satisfactory regression model, including \textit{ARIMA} and \textit{VAR} autoregressive models, a novel multi-criteria decision-making process is proposed. Results show that the best model according to the proposed multi-criteria decision making process provides a $\mathcal{RSME}=(0.1349,0.1304,$ $0.1325)$ and a $\mathcal{MAE}=(0.1003,0.096,0.0987)$ for 1, 2, and 3 steps-ahead predictions.

Multi-Objective Evolutionary Optimization for Time Series Lag Regression

Fernando Jimenez, Joanna Kaminska, Estrella Lucena-Sánchez, Josè Tomàs Palma and Guido Sciavicco

Abstract: It is well-known that in some regression problems the effect of an independent variables on the dependent one(s) may be delayed; this phenomenon is known as lag. Lag regression is one of the standard techniques for time series explanation and prediction. However, using lagged variables to transform a multivariate time series so that a propositional algorithm such as a linear regression learner can be used requires to decide, at preprocessing time, which independent variables must be lagged and by how much. In this paper, we propose a novel optimization schema to solve this problem. We test our solution, implemented with a multi-objective evolutionary algorithm, on real data taken from a larger project that aims to construct an explanation model for the study of atmospheric pollution in the city of Wroclaw (Poland).

Stochastic dimension reduction techniques for time-point forecasting data

Shrikant Pawar and Aditya Stanam

Abstract: Dimension reduction techniques are essential for forecasting datasets. The reduction can be achieved through vector processes, probabilistic approaches to modeling macroeconomic uncertainties, nonstationarity, model integration, forecasting theory and adjustment, rnsemble forecasting, forecasting performance evaluation, interval forecasting, data decomposition etc. Using stochastic modeling techniques, we can retrieve accurate probabilities about a given system or event. This article compares five ("direct", "multitaper", "mvspec", "pgram"; and "wosa") different reduction techniques for stochastic distributed forecast data, while

there are other parametric dimension reduction techniques, "direct" and "mvspec" were found to be the most effective reduction techniques (lowest entropies) for M4 Forecasting Competition dataset.

Towards a Better Nowcasting and Forecasting of Tunisian GDP Growth: The Relevance of Sovereign Ratings Data

Adel Karaa and Azza Bejaoui

Abstract: In this paper, we propose a formal and unified statistical framework in order to help performing reliable nowcasts, short and long-term forecasts of the real GDP growth in the Tunisian context. To do so, we use a set of available monthly macro-financial data and takes into account the sovereign ratings assigned to Tunisia by the four biggest Credit Rating Agencies (i.e. R&I, Fitch, Moody's and Standard & Poor's) since 1994. These data are used to estimate an appropriate multivariate unobserved componants times series model. The empirical results clearly show that combining of macro-financial indicators and sovereign ratings data seems to produce valid and consistent latent factors which track well GDP growth realizations throughout the estimation period. The short and long-term factor-based forecasts of the real GDP growth obtained by applying the AutoRegessive Distributed lag (ARDL) bound testing approach to cointegration by Pesaran et al. (2001) are also more accurate than those produced by other classical benchmark forecasting models. This framework, which we consider as a first step in setting up a real-time monitoring system for the macro-financial situation in Tunisia, already allows us to highlight the main flash indicators capturing the dynamics of the Real GDP growth.

How Well Does Economic Uncertainty Forecast Economic Activity?

John Rogers and Jiawen Xu

Abstract: It is difficult to overstate the reach and influence of the literature on economic and economic policy uncertainty over the last decade (www.policyuncertainty.com). On-going research relates uncertainty to macroeconomic phenomena such as inflation and GDP growth, microeconomic issues concerning firm-level investment and export market entry and exit, and finance considerations like corporate strategy and equity returns. In this paper, we address one surprisingly under-researched topic: what is the forecasting performance of economic uncertainty measures? We examine both in-sample and out-of-sample forecasting, both real and financial outcome variables, sub-sample stability, and real-time considerations. Our measures of uncertainty include U.S. Economic Policy Uncertainty (Baker, Bloom, and Davis (2016)), U.S. macroeconomic uncertainty (Jurado, Ludvigson, and Ng (2015)), and U.S. financial uncertainty (Ludvigson, Ma, and Ng (in press)). We begin by showing that there is substantial explanatory power both in-sample and out-of-sample over the 128 outcome variables in the McCracken and Ng (2016) data set. Next, we document that uncertainty sometimes has additional predictive content over the widely-used excess bond premium (Gilchrist and Zakrajsek (2012)) and the Chicago Fed's National Financial Conditions Index (NFCI). We then use quantile regressions to examine whether the explanatory power of the above measures varies over different parts of the GDP growth distribution. There is good reason to expect that uncertainty would forecast recessionary conditions better than expansions (Adrian et. al. (2019)). We find that both in sample and out of sample, the EPU and MU measures of uncertainty show strong predictive power, especially at lower quartiles, consistent with the prior. Finally, we construct a real-time measure of macroeconomic uncertainty, using vintages of data back to 1999. The real-time uncertainty measure fares much worse than the original measure that is ex-post in its construction, in terms of forecasting performances both in sample and out of sample.

The impact of oil prices on products groups inflation: is the effect asymmetric?

Ligia Elena Topan, Miguel Jerez Mendez and Sonia Sotoca Lopez

Abstract: In this paper we assess the oil price pass-through into both, the global inflation in Spain and the inflation derived from the non-deterministic prices of the standard european classification of product groups, during the period 2002-2018. To this end we fit a transfer function to the inflation in each group, extended to allow for an asymmetry in the transmission of positive/negative oil cost shocks, that is, a ``rockets and feathers effect''. Our results show that most often there is a significant asymmetry, which can be explained by the degree of competition in each market. Even more, we show that allowing for asymmetric effects yields a remarkable improvement in the precision of inflation forecasts.

Forecasting macroeconomic processes with missing or hidden data

John Mashford

Abstract: An approach to utilizing Bayesian methodology to model and forecast macroeconomic processes over a region of interest and multiple planning horizons in the case where some of the data is unavailable is proposed. The generalized state space associated with such a model is defined. The stochastic model considered is autoregressive in time and has a simple standard covariance structure for the spatial component. It is proposed simply to fill in missing values by utilising standard forecasting techniques on complete data blocks to extend the block up to just before the next complete data block for any set of data corresponding to any given spatial index.

Imputing monthly values for quarterly time series. An application performed with Swiss business cycle data

Klaus Abberger, Oliver Müller, Michael Graff and Boriss Siliverstovs

Abstract: This paper documents an applied investigation into strategies and algorithms to deal with the problem of missing higher frequency data. We refer to Swiss business tendency survey (BTS) data, in particular the KOF manufacturing surveys, which are conducted in both monthly and quarterly frequency. As a result, some information is available at quarterly frequency only. There is a wide range of ways to address this problem comprising univariate and multivariate approaches. We resort to different multivariate imputation algorithms and apply them to generate monthly series out of quarterly series from the KOF BTS in the Swiss manufacturing sector and compare the results. Our strategy to compare the suitability of the different approaches is to make sure that we do possess adequate reference series for the model selection stage. To this end, we apply our procedures to series that are monthly, from which we create artificial quarterly data by deleting the same two out of three data points from each quarter. The candidate series for the imputation of the missing (i.e. deleted) observations are given by the entire set of time series that are resulting from the monthly KOF manufacturing BTS survey. In this way, we resort to a set of indicators that share the common theme, which is a reflection of the Swiss business cycle. With this set of potential indicators, we conduct the different imputations. On this basis, we then run standard tests of forecasting accuracy by comparing the imputed monthly series to the original monthly series. Descriptive statistics like the correlation and the absolute mean or root square forecast error allow ranking the algorithms; statistical tests like the encompassing test reveal whether the different methods are significantly superior/inferior.

Hybrid Method Forecasting Stock Market Data

Sadam Alwadi and Ahmed Awajan

Abstract: Since the industrial data theatres important element in any financial development. Therefore, in this article the events of productivity of the industry stock market data from Amman Stock Exchange (ASE) will be explored and forecasted using some of traditional model which is Exponential model compound with Wavelet transform functions (WT) using 4 functions which are (Haar, Daubechies, Coiflet and La8) for the available data from Jan. 2012 to Dec. 2018 in order to improve the forecasting accuracy. First, the series of dataset will be decomposed by WT functions in order to capture the significant affect based on detailed coefficients, then the smooth's series will be predicted using Exponential functions and exponential functions with WT .

Measuring the Effect of Unconventional Monetary Policies on Market Volatility

Demetrio Lacava and Edoardo Otranto

Abstract: As a response to the great recession, ECB resorts to unconventional monetary policies, i.e. central bank's balance sheet expansions. Our research aims to analyse the impact of unconventional monetary policy by ECB on stock market volatility in four Eurozone countries (France, Germany, Italy and Spain) within the Multiplicative Error Model framework. Basing on Otranto (2015), we further modify this model to allow volatility to depend on unconventional monetary policy: in particular, we quantify the part of market volatility depending directly on unconventional policies by distinguishing between the announcement effect and the implementation effect. While we observe an increase in volatility on announcement days, we find a negative implementation effect, which causes a remarkable reduction in volatility in the long term. Moreover, we extend the analysis implementing a Markov Switching model to test the ECB ability to keep volatility in low and high regimes. In this case, it emerges an average duration of the QE effects on volatility of about 15 days for France, Italy and Spain. The effect lasts more in Germany, probably because of more favourable economic conditions characterizing this country, during the sample period.

Comparative Investigation of Tests in Modeling Process in Univariate Time Series

Reşat Kasap and Sibel Sancak

Abstract: Forecasting is one of the most important concepts in time series. To forecast truly the model should be determined that identifies the data set best. However one or more outliers in the model affect the parameters of the model and forecasting. In the scope of this study, firstly, time series and the outliers in time series concepts are identified. The effect of outliers is investigated on the ARMA model parameters and forecasting. For this reason, data of TUIK are used to research outliers and forecasting.

Modelling the Nigerian Market Capitalization Using Vector Error Correction Model

Nura Isah, Dr. Sani Ibrahim Doguwa and Basiru Yusuf

Abstract: This research work intends to empirically develop a model for Nigerian stock Market Capitalization using Vector Error Correction Model. Forecast performance of the estimated Model is analyzed using Quarterly data on Real Gross Domestic Product (RGDP), Inflation rate (INF), Exchange Rate (EXR), Money Supply (MS) and Nigerian Stock Exchange market Capitalization (NSEC) from 1985Q1 to 2016Q3. The result for Augmented Dickey Fuller test indicates that all the variables are stationary after taking the first difference I(1).

The Johansen co-integration test revealed that the variables are co-integrated, that is VEC Model is more appropriate to represent the time series data. The long run equation indicates that all the variables have a significant long run relationship with NSEC. However, for the short run equation (VEC Model) only RGDP and MS are significant in the Model while INF and EXR are insignificant to NSEC. The result for forecast performance for estimated VEC Model indicates that, the model has a Root Mean Square Forecast Error (RMSFE) of 22.05, Mean Absolute Forecast Error (MAFE) 17.65 and Mean Absolute Forecast Percentage Error (MAFPE) of 55.72%. The result for Likelihood Method (LM) test indicates that the null hypothesis of no serial correlation at lag 1 to 8 at 1% significant level, has been accepted.

Modelling and Predicting Air Quality in Visakhapatnam using Amplified Recurrent Neural Networks

Lavanya Devi Golagani and Srinivasa Rao Kurapati

Abstract: Air quality refers to the condition of the air within our surroundings. Good air quality relates to the degree which the air is clean, clear and free from pollutants such as smoke, dust and fog among other vaporous impurities in the air. Emission of pollutants by nature and human made developments have drastically increased all over the globe in the present circumstances. Study of the pollutant concentrations, developing models to predict the future air quality has become areas of interest for the research and industry community. The air pollutants data is a temporal sequence type data and hence proper initiatives are to be taken to handle it. The system which is able to predict the concentration of air pollutants with sufficient anticipation can provide public authorities the time required to manage the emergency. Great progress has been made in the prediction of the concentration of air pollutants over the past decades by using conventional techniques. However, it is still challenging to accurately predict the concentration of air pollutants due to the complex influential factors such as meteorological parameters. On the other hand, the air quality is unique for distinct geographical locations. Hence, it is important to study, analyze and predict the air quality parameters for specific geographical area of interest. This paper aims to study and predict the quality of air at city, Visakhapatnam, India through persistent deep learning technique, amplified recurrent neural networks (ARRN) model to predict air quality. The results obtained were evaluated with the state-of-the-art models. It has been observed that the that proposed frame work significantly improves the prediction compared to widely used benchmarks models.

Environmental policies analysis for CO2 emission reduction: evidence across countries 1980-2014

Yi Zheng and Dess Pearson

Abstract: We employ a difference-in-difference regression model to analyse the efficiency of carbon taxation policy and its interaction with an Emission Trading System (ETS). We provide evidence for 26 of the most developed countries in the world in the period 1980-2014. Following the work of Dynarski [2004], Cameron et al. [2008, 2012] and Hoechle [2007], we compare the results under three methods of regression analysis. We confirm that carbon taxation has preformed well in the 35 years under observation in that they efficiently control CO2 emissions. We find that the longer duration a country uses carbon taxation, the greater the reduction in CO2 emissions. For the countries that use both an ETS and carbon taxation, we find an even more efficient CO2 emission reduction. The results are robust to heteroskedasticity, autocorrelation and cross-sectional dependence.

View of the hydrological determination of turbomachinery potential in current

Levent Yilmaz

Abstract: At the stationary flow along current lines the hydropower is determined huge but the application about measurements Show some uncertainties. To improve these difficulties in scope of hydrology some hydrodynamical relations will be given. At the end we will find a new correct way for determining the hydropotential of sea currents. There is a definite correlation between pressure P and flow velocity v. If the flow is assumed as stationary, the current lines have identical similarity to the trajectory of the fluid particles distribution.

Hybrid Orbit Propagator based on Time Series Forecasting: Predictive Interval

Montserrat San-Martín, Iván Pérez, Rosario López and Juan Félix San Juan

Abstract: The orbital motion of an artificial satellite or space debris object is perturbed by a variety, and sometimes not well-modeled, exter- nal forces [1,2]. The hybrid methodology can be used to predict these unmodeled effects or the uncertainty associated with this process. In this work, a Hybrid Orbit Propagator based on SGP4 [3–7] and a state space formulation of the exponential smoothing method as the forecast- ing technique is developed. The error terms of the forecasting technique are considered Gaussian noise what allows us to use the maximum like- lihood method to estimate the parameters of the exponential smoothing model, as well as computing the point forecast and the reliable predic- tive intervals. Finally, this Hybrid Orbit Propagator is applied to data from a satellite of the Galileo constellation. This Propagator improves the accuracy of the classical SGP4 and it is particularly good for short forecast horizons.

Hybrid Orbit Propagators based on Neural Network

Iván Pérez, Rosario López, Montserrat San-Martín and Juan Félix San Juan

Abstract: Space Situational Awareness current needs demand innova- tive solutions to the orbit propagation problem, so as to find new algo- rithms which are simultaneously accurate and fast. The hybrid method- ology for orbit propagation constitutes a recent approach based on mod- eling the error of any orbit propagator with the aim of complementing its calculations and hence enhancing its precision. Diverse sources of inaccuracy can exist in propagators, such as incomplete perturbation models, forces not considered, low-order of the series expansions, etc. The creation of a time series with the differences be- tween ephemerides computed with low-accuracy propagators and their corresponding real observations (or precisely computed ephemerides) allows applying time- series forecasting techniques so as to create a model that includes any dynamics not contained in the original propagator. Then, the adjusted model can be used in order to correct other future predictions. We present an application of the hybrid methodology, in which the time-series fore- casting process is performed by means of machine-learning techniques, to the well-known SGP4 propagator [1, 2]. We have adjusted the result- ing Hybrid SGP4 propagator [3–5], HSGP4, to the case of Galileo-type orbits. We will show how the use of HSGP4 can reduce the position error of SGP4, hence extending the validity of Two-Line Elements (TLE) from Galileo satellites.

A Stochastic Drift Model for Electrical Parameters of Semiconductor Devices

Horst Lewitschnig and Lukas Sommeregger

Abstract: During their development, semiconductor devices are put to accelerated stress tests in order to simulate their life in a short time. Devices are initially tested, stressed for a certain time, then tested again, stressed again, and so on. Such tests before, during and after the stress are called readouts. In the course of such readouts, electrical parameters are tested and their behavior over time, i.e. their drift, can be observed. Potentially they could also drift outside of their specified limits. In order to avoid this, tighter test limits are introduced at the final production test of semiconductors. These tighter limits are called guardbands.

Guardbands should be placed to limit the likelihood that electrical parameters drift outside the specification limits during their lifetime while maximizing the production yield at the same time. We show a way to describe such drift behavior, which serves as a basic framework from which guardbands can be derived. The method is easy to implement and reflects the knowledge that is gained from stress tests and parametric drift investigations.

Chaos and Slow Earthquakes Predictability

Adriano Gualandi, Jean-Philippe Avouac, Sylvain Michel and Davide Faranda

Abstract: Slow Slip Events (SSEs) are episodic slip events that play a significant role in the moment budget along a subduction megathrust. They share many similarities with regular earthquakes, and have been observed in major subduction regions like, for example, Cascadia, Japan, Mexico, New Zealand. They show striking regularity, suggesting that it might be possible to forecast their size and timing, but the prediction of their extension and exact timing is still yet to come. They certainly are a great natural system to study how friction works at scale of the order of hundreds or thousands of km, and their recurrence time being much shorter than that of regular earthquakes, they give us the possibility to study multiple cycles and test their predictability. Here we focus on the Cascadia region, where SSEs recur every about 1 or 2 years, depending on the latitude. We use a catalog containing dozens of SSEs derived from the study of GPS position time series during the time span ranging from 2007 to 2017. The data show a clear segmentation with a few major patches interacting with one another, a behavior that recalls that of a discrete body system. We use both classical embedding theory and extreme value theory applied to the study of dynamical systems to show that, where the signal to noise ratio is sufficiently high, a low-dimensional (< 5) non-linear chaotic system is more appropriate to describe the dynamics than a stochastic system. We calculate major properties of the strange attractor like its correlation and instantaneous dimension, its instantaneous persistence and a possible range for the metric entropy of the system. For the better resolved segments, the onset of large SSEs can be correctly forecasted by high values of the instantaneous dimension. Longer-term deterministic prediction seems intrinsically impossible. In conclusion, SSEs in Cascadia can be described as a deterministic, albeit chaotic, system rather than as a random process. As SSEs might be regarded as earthquakes in slow motion, regular earthquakes might be similarly chaotic and predictable but with a predictable horizon of the order of their duration.

Load Forecast by Multi Task Learning Models: designed for a new collaborative world

Leontina Pinto, Jacques Szczupak and Robinson Semolini

Abstract: This paper proposed a forecasting model designed for lack-of-data problems, based on Multi tasking learning techniques. It is especially useful for evolutionary markets and systems, where new paradigms (like renewable penetration or prosumers) significantly impact behavior and dynamics, creating unforeseen responses, unpredictable from past (possibly obsolete) historical data. A case study targeting the recent Brazilian load changes illustrate the approach performance: it was possible to combine data from four different distribution companies, creating a learning network, yielding reliable results where all other models failed.

Forecasting inflation in the euro area: countries matter!

Claudia Pacella and Angela Capolongo

Abstract: We construct a Bayesian vector autoregressive model including the key drivers of inflation, cross-country dynamic interactions, and country-specific variables. The model provides good forecasting accuracy with respect to the popular benchmarks used in the literature. We perform a step-by-step analysis to shed light

on which information is more crucial for forecasting euro area inflation. The complete model performs better in forecasting inflation excluding energy and unprocessed food. Our empirical analysis reveals the importance of including the key drivers of inflation and taking into account the multi-country dimension of the euro area.

On the automatic identification of Unobserved Components Models

Diego J.Pedregal and Juan R. Trapero

Abstract: This paper presents an approach to automatically identify Unobserved Components models. This is a necessity once the big data era has come and is saying among us. This has become obvious for many companies and public entities that has passed from a crafted analysis of each individual problem to handle a tsunami of information that have to be processed efficiently, online and in record time. Automatic identification tools are the usual way to go in the Machine Learning area and also in some other statistical approaches, but it never has been tried out in Unobserved Components models (UC). There are many reasons for this. Firstly, UC have been developed mainly in academic environments with the purpose of research, with little dissemination among practitioners for their everyday use in business and industry. Secondly, there is a widely-held feeling that UC models do not really have anything relevant to add to exponential smoothing methods has deterred its use in practice. Third, UC models are usually identified by hand, with automatic identification being very rare. Finally, software is rather scarce compared with other methods. This paper introduces a new piece of software developed for the automatic identification and forecasting of UCs with some enhancing features. The package searches among many combinations of different specifications for each of the components according to a step wise algorithm to reduce the universe of possible models and selects the one with the best metrics. The forecasting results suggest that UC models are powerful potential forecasting competitors to other well-known methods. Though there are several pieces of software available for UC modeling, this is the first implementation of an automatic algorithm for this class of models, to the authors knowledge.

Theory and Simulaion of Procrastination: The Before and After the Releasing of a Cash Credit

Huber Nieto-Chaupis

Abstract: Commonly the apparition of procrastination in some social layers are not well understood in all. In this paper we present some stochastic simulations by which customers fail to pay the monthly fee when have accessed to a cash credit from a bank. In according to these computational studies, one of the main causes is the expended time either the bank or customer to study and analyze the possible scenarios by which one of them is in risk. Thus, customers might be in a kind of pseudo-procrastination even the cash is not yet released because the rapid decision to access the cash becomes a type of loss.

Theory of Blockchain Based on Quantum Mechanics

Huber Nieto-Chaupis

Abstract: Because the cryptocurrency dynamics inside the framework of Bitcoins would require of advanced schemes to encrypt information, the usage of powerful technologies aimed to guarantee anonymous information becomes a must. In this paper we present a theory of blockchain entirely based on Quantum Mechanics formalism the which might be used as part of advanced algorithms to generate random ensembles that is in fully concordance with the purpose of the markets based on cryp- tocurrency. Simulations have demonstrated the reliability of a simple model of blockchain of up to a 75% fact that supports the idea for using the Quantum Mechanics formalism in these advanced and highly secured modern markets.

Different frequencies in term structure forecasting

Alexander Matthies

Abstract: Government bond interest rates determine the ability of a country to finance itself and are important factors in the fixture of other interest rates. Modelling the term structure helps in understanding this economic base variable and determine financial risk. Unseen dynamic factors of the term structure are estimated in a data driven approach using principal components analysis in rolling time windows to produce yield curve forecasts. Statistical and economic evaluation is provided for different time windows, factor numbers, and data frequencies. The data consists of daily, weekly, and monthly observations of government bond interest rates for Germany, the UK, and the US for the time period from 1980 to 2016. Implicitly this approach tests the basic assumptions of Nelson-Siegel type economic factor models of the term structure. Term structure forecasts are evaluated in terms of three complementary criteria or loss functions, namely the statistical mean squared forecast error criterion, and the two more economic criteria of directional accuracy and big hit ability. Factor analysis supports the idea that a level, slope, and curvature factor underly the yield curve. In a data set with all term structures we find evidence of a global level factor. The selection of the ideal time windows with respect to mean squared forecast errors depends on the forecast horizon. Data sets with lower frequencies appear to produce better forecasts. Statistical and economic criteria suggest that more than one factor should be employed for forecasting. The results support the concepts of of the Nelson-Siegel Model.

Methods of Detection of Non-Technical Energy Losses with the Application of Data Mining Techniques and Artificial Intelligence in the Utilities

Marco Toledo and Carlos Álvarez

Abstract: This research use the technical, commercial and social information of consumers of electric power distribution utilities through the application of data mining techniques and artificial intelligence proposes to determine groups of potential consumers who present infractions or damages in the measurement equipment, in such a way that the theft of energy is minimized and the income in the utilities is increased.

End of charge detection of batteries with high production tolerances

Andre Loechte, Ole Gebert and Peter Gloesekoetter

Abstract: When developing new battery technologies, fundamental research means assembling new batteries by hand since a production line is not worthwhile for building and testing individual cells. This causes high production tolerances to occur because manual manufacturing is not as precise as machine-made. When putting these prototypes into operation, problems can arise due to the varying parameters. One of the most important exercise is finding a criterion of a full battery. This can be challenging when parameters like the capacity or the end of charge voltage are not precisely known due to the tolerances. Furthermore, new battery types do not necessarily rely on the same stopping criteria. For example zinc-air secondary batteries do not offer an end of charging voltage. Its charging current is not going to decrease when the battery is full and the charging voltage is held at a fixed value. But instead of de-oxidising zinc oxide, hydrogen is produced. In the majority of cases overcharging should be avoided as it harms the battery. Another even more dangerous consequence is the possibility of an explosion. Especially lithium based batteries are known for their need of compatible ambient and charging parameters. This paper proposes a new criterion for detecting the end of charge that is based on the rate of change of electrochemical impedance spectra of the examined batteries. Device parameter

fluctuations influence every measurement. Therefore, using the rate of change offers the possibility to not depend on these fluctuations.

Climate change: climate missing data processing, modeling rainfall variability of Soummam watershed (Algeria)

Amir Aieb, Khalef Lefsih, Marco Scara, Brunella Bonacorso and Khodir Madani

Abstract: The time series of monthly rainfall during 51 years of observations (from 1967 until 2018) for 24 stations in Soummam watershed of Algeria were analyzed for trends and aridity state of the area. The choice of the Weibull distribution law was justified by comparing the fitting of different probability distributions law used in literature reviews. This paper proposed a new imputation algorithm to fill missing climate data, based on the optimization of some regression methods, which are hot deck, k-nearest-neighbors imputation, weighted k-nearest-neighbors imputation, multiple imputation, linear regression and simple average method. The choice of these methods was justified by qualitative and quantitative statistical tests analysis. However, the reliability of obtained results depends mainly on percentage of missing data, choice of neighboring stations and data missingness mechanism, which should be missing at random.

Conversion of geological model (fine-mesh) to dynamic (coarse-mesh) hydrocarbon model with the nature approach in simulation of thermal recovery in a fractured reservoir

Mehdi Foroozanfar

Abstract: Operation and proper management of reservoirs requires the prediction of reservoir performance. This prediction is generally done by computer simulation. Since simulation software is capable of generating static models with the number of millions and even billions, using simulation methods and dynamic simulation on these models is difficult and sometimes impossible, multi-scale generation is necessary. In this study, we introduce a methodology which inspired by Earth’s grid to make multi-scale grid generation on a multi-phase, heterogeneous reservoir, the enhanced oil recovery process is steam injection. The multi-scale grid generation that has been introduced in this study is base- on Earth’s model. From a reservoir engineer’s point of view Earth is equivalent to a multi-scale grid model that could be a pattern for multi-scale grid generation in a hydrocarbon reservoir to minimize CPU-Time for dynamic simulation. The principle of multi-scale grid generation in hydrocarbon reservoir is; regions with high Darcy velocities should remain fine-scale and other segments with low intensity of heterogeneity regions could resize into up-scale. Intersection of latitude and longitude lines causes Earth’s discretization surface could be a practical pattern for dynamic simulation of hydrocarbon reservoirs. The interpretation of Earth’s discretization is; intersection of latitude and longitude lines create segments with fine-size like South and North Poles that equal to injection and production wells in a reservoir model and other segments have verity intensity of coarse-size. Earth's magnetic field which enters from South and exits from North Pole leads to we can consider earth as a hydrocarbon model. The results of the multi-scale grid generation method which inspired by Earth’s pattern were compared with the fine-mesh (Geological Model) model; these results show that the proposed method predicts close accuracy of the fine-mesh network model with less run-time.

Analysis of periodicities of cosmic ray time series located at different geomagnetic locations

Jose F. Valdes and Marni Pazos

Abstract: We studied the neutron monitor data bases of Mexico City, Oulu, Finlad and Moscow, Russia from 1990 to 2017 to find periodicities in the intensity variations of the cosmic ray flux. We used the wavelet transform to identifiy mid-term variations present in the records. The corresponding confidence levels are given to the periodicities, as well as the contribution to the total power spectrum of such variations. Results are consistent with previous analysis done for other cosmic ray detectors, showing the relevance of mid-term variations, probably related to phenomena occuring below the solar atmosphere. As a reference, we compare these results with those of classical Fourier analysis based on the discrete Fourier transform, and a fractal analysis, giving consistent results. To the best of our knowledge, this is the first time that a comparative analysis of this kind is done for these three neutron monitor series representing low, medium and high geomagnetic latitudes.

Wind-power intra-day multi-step predictions using polynomial networks solutions of general PDEs based on Operational Calculus

Ladislav Zjavka, Stanislav Mišák and Lukáš Prokop

Abstract: Precise intra-day predictions of wind-power are challenging due to its intermittent nature and high correlation with large-scale atmospheric chaotic circulation processes. NWP systems solve sets of differential equations to predict a time-change of each 3D-grid cell in several atmospheric layers. Their surface forecasts of wind speed are not entirely adapted to specific local characteristics and anomalies, which largely influence its temporal-flow. AI methods using historical observations can convert and refine the daily forecasts in consideration of wind farm siting, terrain asperity and ground level (hub height). Their independent wind-power predictions in horizon of several hours are also more precise then NWP model forecasts as these are usually produced every 6 hours. The designed method uses Polynomial neural networks to decompose and substitute for the general linear Partial Differential Equation being able to describe n-variable functions of unknown complex dynamic systems. It solves specific 2-variable 2nd order PDEs, formed in PNN nodes, using a polynomial conversion based on Operational Calculus. The inverse Laplace transformation is applied to the resulting rational terms to obtain the originals of node functions whose sum gives the complete PDE model. The composite PDE models are developed with data samples from the estimated optimal numbers of training days to represent spatial data relations in current weather, necessary for applicable predictions. They can predict wind power up to 12 hours ahead according to a trained data inputs->output time-shift. Intra-day multi-step predictions using the PDE models are more precise than those based on NWP model forecasts or statistical techniques allowing using local time-series of several variables only.

Stochastic Weather Generators in Czechia: 25 Years of Development and Applications

Martin Dubrovsky, Radan Huth, Ondrej Lhotka, Jiri Miksovsky, Petr Stepanek, Jan Meitner and Miroslav Trnka

Abstract: Stochastic weather generators (WGs) are software tools, which can produce synthetic weather time series statistically similar to real-world weather data. This means that the relevant statistics representing probabilistic distributions of individual variables, correlations between variables, and temporal & spatial structure derived from the synthetic series are as close as possible to those derived from observations. WGs can do different tasks. Most typically, they are used to produce synthetic weather series serving as an input to various models (often agricultural and hydrological models; these will be called “impact models” in next) employed in simulating weather-dependent processes. In this role, WGs are commonly used in cases when observational data are not available or they are not sufficiently long or when we want to assess variability or changes in characteristics simulated by the impact models in response to weather variability or projected climate change. In the later case (climate change impact studies), parameters of the generator are derived from the baseline (present climate) weather data in the first step, and then they are changed according to climate change scenarios derived from the dynamical climate models (either Global Climate Models or Regional Climate Models). WGs

may be applied also in other tasks. Specifically, our M&Rfi generator may be linked with the weather forecast and used for a probabilistic crop yield forecasting. Our other generator SPAGETTA (being a multi-site WG) has been used for studying the collective significance of local trends at multiple sites and developing a new test for examining this significance by analysis of data from multiple mutually correlated sites; all WG parameters representing the statistical structure of the series are prescribed by the user in this mode of operation. Although no WG can produce perfectly realistic weather data, they have important advantages, which make them a favourite tool used in applied climatology: (i) WGs are very fast and may produce a large number of realisations of arbitrarily long weather series in a reasonable time (in contrast with complex but slow dynamical climate models. (ii) WGs may be interpolated so that they can produce data even for sites where there are no weather observations available for WG calibration. (iii) WGs are very flexible and may handle various combinations pf weather variables needed to feed the impact models. (iv) WGs may produce weather series representing both present and future climates (even for emission scenarios, for which outputs from dynamical climate models are not available), as well as the weather series which conform to the probabilistic weather forecast. Various WGs are available, they may differ in several aspects: (a) The modeling approach may be parametric (most common approach) or semi-parametric. In the parametric WGs, the temporal structure is often based on Markov chains (for precipitation occurrence) and autoregressive models (non-precipitation variables) and the variables are assumed to have a specific distribution function (for example, Gamma distribution or mixed-exponential distribution are used for daily precipitation sums). The non-parametric generators are based on resampling. Apart from that, some generators may be called semi-parametric as they use both parametric and non-parametric approaches (e.g. LARS generator developed by M. Semenov). To make the temporal structure of the synthetic series more realistic, the surface weather may be generated conditionally on large-scale circulation characteristics or on weather series simulated by WG running at longer (e.g. monthly) time step. (b) Number of variables: the generators may simulate only a single variable, but they are mostly multivariate. Specifically, in agrometeorological applications, up to 6 variables are commonly needed: daily temperature minimum and maximum, daily sums of precipitation and incoming solar radiation, and daily means of air humidity and wind speed. (c) Time step ranges from continuous time-scale or very short time intervals (minutes or shorter) models (using different statistical models instead of Markov chains and AR models) to hourly, daily and monthly or even annual generator; most commonly, the daily weather generators are used in both agrometeorological and hydrological applications. (d) Spatial aspect: in agrometeorology, single-site generators are usually used as the simulated processes mostly depend only on the site-specific weather. On the contrary, in hydrological modelling, where the proper simulation of rainfall-runoff processes requires realistically spatially coherent data, multi-site (the sites may be regularly or irregularly distributed in space) WGs are needed. In our contribution, we will give a brief overview of the weather generators developed since 1994 by the main author in a close co-operation with agrometeorologists, hydrologists, programmers and other statistical climatologists. The WGs mostly (but not only) run at daily time step, they are based on both parametric and non-parametric approaches, and those developed before 2016 are only single-site - only the most recent SPAGETTA generator is multi-site; all of them are multi-variate. The focus will be put on two generators most commonly used by our (but not only) team: (A) M&Rfi is a parametric single-site multi-variate daily generator, which has been involved in many agrometeorological experiments since 1994. In this generator, precipitation occurrence is modelled with the Markov chain, precipitation amount is sampled from the Gamma distribution, and the non-precipitation variables are simulated using the first-order AR model, whose parameters are conditioned on precipitation occurrence. (B) SPAGETTA is a spatial generator, whose development (started in 2016) was motivated by its involvement in hydrological modelling and in studying the collective significance of local trends at multiple mutually correlated sites. SPAGETTA is based on a single-site M&Rfi generator, which was spatialised using the Wilks' approach. As a part of our presentation, we will show selected results obtained while validating the two generators in terms of various climatological indices (for example in terms of occurrence of hot/cold/wet/dry spells as well as the compound hot-dry/hot-wet/cold-dry/cold-wet spells). Some results obtained with these generators while being employed in climate change impact experiments will be also shown. In addition, we will demonstrate the use of the generators in other applications: application of M&Rfi in seasonal forecasting of crop yields, and application of SPAGETTA in developing the test for examining a collective significance of local trends at multiple sites. Acknowledgement: The development and applications of our weather generators have been funded by several projects since the end of 20th century. The most recent experiments are made within the frame of project SustES ("Adaptation strategies for sustainable ecosystem services and food security under adverse environmental conditions"; CZ.02.1.01/0.0/0.0/16_019/0000797) and

project GRIMASA ("Development of high-resolution spatial weather generator for use in present and future climate conditions" project no.18-15958S) funded by Czech Science Foundation.

Wind and Solar Forecasting for Renewable Energy System using SARIMA-based Model

Marwa Haddad, Jean Marc Nicod, Yacouba Boubacar Maïnassara, Landy Rabehasaina and Zeina Al Masry

Abstract: In order to limit global warming and decrease CO_2 consumption, renewable energy sources (RES) such as photovoltaic panel (PV), wind turbine (WT) and hydrogene system composed of fuel cell (FC) and electrolyzer (EL) are more and more commonly used to generate electricity for different purposes (houses, street lights, companies, data centers..). However, reasonable sizing for these system could improve the power supply reliability and reduce significantly the annual cost. In this paper, we propose different system components of hybrid energy system and develop a general model to find an optimal sizing based on mathematical models and the power efficiency of the components to completely power a data center.

Deterministic weather forecasting with a newly developed non-hydrostatic global atmospheric model

Song-You Hong

Abstract: Numerical weather prediction based on computer programs of weather phenomena is a grand challenge and has been continuously progressed for past half century. Korea Institute of Atmospheric Prediction Systems (KIAPS) has embarked a national project in developing a new global forecast system in 2011. The ultimate goal of this 9-year project is to replace the current operational model at Korea Meteorological Administration (KMA), which was adopted from the United Kingdom's Meteorological Office's operational model. As of January 2019, the 12-km Korean Integrated Model (KIM) system that consists of a spectral-element non-hydrostatic dynamical core on a cubed sphere and the state-of-the-art physics has been launched in a real-time forecast framework, with the initial conditions obtained from the advanced 4-DEnvar over its native grid. A background on the KIAPS mission and the development strategy of KIM toward a world-class global forecast system are described, along with a future plan for operational deployment.

The Generalized STAR Model with Spatial and Time Correlated Errors to Analyze the Monthly Crime Frequency Data

Utriweni Mukhaiyar, Udjianna Sekteria Pasaribu, Kurnia Novita Sari and Debby Masteriana

Abstract: Most of space-time series models assume that errors are independent or at least uncorrelated. Here the errors are considered have both spatial and time correlated errors (STCE). This assumption defines that errors in a certain location is affected by its neighbors errors of previous times. Such errors are attached to Generalized Space-Time Autoregressive (GSTAR) model. Due to the errors dependence, generalized least squares is applied to estimate the parameters. For case study, the monthly crime frequency data is used. It is obtained that the GSTAR($1;1$) with STCE give the lower Akaike Information Criteria (AIC) than higher GSTAR model with time correlated errors.

The Generalized STAR Model with Adjacency-Spatial Weight Matrix Approach to Investigate the Vehicle Density in Nearby Toll Gates

Utriweni Mukhaiyar, Kurnia Novita Sari and Nur Tashya Noviana

Abstract: A spatial weight matrix represents dependency among observed locations. In Generalized Space Time Autoregressive (STAR) modelling, the non-uniform type of this matrix is mostly built based on Euclidean distance, such that it become fixed all the time. This study use adjacency matrix approach which is constructed based on correlation among time series data of each location. From the adjacency matrix, a minimum spanning tree of observed locations is acquired along with the degree of neighbours. This degree determines how many and large the neighbour locations influence the observed locations in various spatial lag. Then the more representative spatial weight matrix which represents the behaviour of real observations, is obtained. This approach is applied to analyze the pattern of number vehicles which enter the Purbaleunyi of toll gates in West Java, Indonesia. It is obtained that the simplest of GSTAR model, GSTRA($1;1$), is the most appropriate model to be applied for forecasting. Although adjacency matrix approach does not give the best result, but its performance is alike with the best obtained model.

Landslide Debris-Flow Prediction using Ensemble and Non-Ensemble Machine-Learning Methods

Praveen Kumar, Priyanka Sihag, Ankush Pathania, Shubham Agrawal, Naresh M, Pratik Chaturvedi, Ravinder Singh, Uday K V and Varun Dutt

Abstract: Landslides and associated soil movements (debris-flow) are the common natural calamities in the hilly regions. In particular, Tangni in Uttrakhand state between Pipalkoti and Joshimath has experienced a number of landslides in the recent past. Prior research has used certain machine-learning (ML) algorithms to predict landslides. However, a comparison of ensemble and non-ensemble ML algorithms for debris-flow predictions has not been undertaken. In this paper, we use ensemble and non-ensemble machine-learning (ML) algorithms to predict debris-flow at the Tangni landslide. Non-ensemble algorithms (Sequential Minimal Optimization (SMO), and Autoregression) and ensemble algorithms (Random Forest, Bagging, Stacking, and Voting) involving the non-ensemble algorithms were used to predict weekly debris-flow at Tangni between 2013 and 2014. Result revealed that the ensemble algorithms (Bagging, Stacking, and Random Forest) performed better compared to non-ensemble algorithms. We highlight the implications of predicting debris-flow ahead of time in landslide-prone areas in the world.

Unemployment and Poverty as Disordered Social Observables in the Shannon Entropy Theory

Huber Nieto-Chaupis

Abstract: In this paper we use the well-known Shannon entropy to simulate the identication of areas in Lima city where unemployment and poverty are largely correlated each other. Under the assumption that these social problems are seen as a disordered complex system, the logarithm modeling applies. Our approach is well adjusted to the ocial data with a discrepancy of order of 7% that supports the fact that the social anomalies are actually entropic systems with a high number of freedom degrees.

Spatial integration of agricultural markets in the EU: Complex Network analysis of non-linear price relationships in hog markets

Christos Emmanouilides and Alexej Proskynitopoulos

Abstract: We present work in progress on the spatial price causality structure between the hog markets of 24 European countries using weekly time series data from 2007 to 2018 and non-linear Granger causality. The EU hog market is studied as a dynamic complex network of linkages between prices in member states. We investigate the temporal development of the spatial network of price relationships, and through the dynamics of its major structural characteristics we draw insights about the horizontal agricultural market integration process in the EU. Of particular interest is the evolution of the degree of market inter-connectedness, the strength and reciprocity of price relationships, the development of influential markets (hubs) and of market clusters with strongly interacting components.

Comparative Study of Models for Forecasting Nigerian Stock Exchange Market Capitalization

Basiru Yusuf and Nura Isah

Abstract: This paper proposes two forecasting models for the Nigerian Stock Exchange Market Capitalization using the Autoregressive Integrated Moving Average (ARIMA) process and an Autoregressive Distributed Lag (ARDL) process. A better model was selected by comparing the forecast evaluation for the estimated models using pseudo-out of sample forecasting procedure over 2013q4 to 2016q3. The statistical loss functions 〖MAE〗_(t) 〖,RMSE〗_t and 〖MAPE〗_(t) for the t forecast horizon (t=1, 2, . . . , 12) are used to compare the forecast performance of the two estimated models. The results show that ARIMA model outperform ARDL model in three to four quarters forecast horizon. On the other hand, ARDL model outperform ARIMA in one to two quarters, five to seven quarters as well as nine to twelve quarters forecast horizon. Therefore, in forecasting Nigerian Stock Exchange Market Capitalization in both short and long horizons, it can be concluded that ARDL is better model to be used.

Models predicting corporate financial distress and industry specifics

Dagmar Camska

Abstract: This paper is focused on tools predicting corporate financial situation. There have been constructed plenty of models whose aim is to predict possible corporate default or distress. These models will be examined. Traditionally analyses would be focused on the explanatory power or models' accuracy. The aim of this paper is different. Although the models can be mainly used generally there are many specifics which affect results and gained conclusions. The specific highlighted in this paper is an industry branch. Companies operate in different industry areas which influence their performance and overall financial results and ratios and therefore it has an impact on the models' result. The paper works with three industry branches: Manufacture of fabricated metal products, except machinery and equipment (CZ-NACE 25), Manufacture of machinery and equipment (CZ-NACE 28) and Construction (CZ-NACE F). The results will be based on three data sample, specifically financial healthy companies 2012, insolvent companies 2012 and companies 2017. The results of different models predicting financial distress will be computed and compared. The main tools of descriptive statistics will be applied. It should prove or disapprove if industry specifics influence the models significantly.

Analyzing Extreme Financial Risks: A Score-driven Approach

Rodrigo Herrera

Abstract: This paper develops a new class of dynamic extreme value models, driven by the score of the conditional distribution with respect to both the duration between extreme events and the magnitude of these events. This data-driven framework is a feasible method for capturing their time-varying arrival intensity, and magnitude. It is also shown how exogenous variables such as realized measures of volatility can easily be incorporated. An empirical analysis based on a set of major equity indices shows that both the arrival intensity and the size of extreme events vary greatly during times of market turmoil. The proposed framework performs well relative to a competing approach in terms of forecasting extreme tail risk measures.

Freedman's Paradox: an Info-Metrics Perspective

Pedro Macedo

Abstract: In linear regression models where there are no relationships between the dependent variable and each of the potential explanatory variables – a usual scenario in real-world problems – some of them can be identified as relevant by standard statistical procedures. This incorrect identification is usually known as Freedman's paradox. To avoid this disturbing effect in regression analysis, an info-metrics approach based on normalized entropy is discussed and illustrated in this work. The results suggest that normalized entropy is a powerful alternative to traditional statistical methodologies currently used by practitioners.

Powers of Texts

Diana Gabrielyan, Lenno Uuskula and Jaan Masso

Abstract: In this paper we take advantage of technological advances and use high frequency multidimensional textual news data available in the internet and propose a new index of inflation expectations. We utilize the power of text mining and its ability to convert large collections of text from unstruc-tured to structured form for in-depth quantitative and qualitative analysis of Guardian news data. Main contribution of his paper is to explore online news as novel data source to capture the inflation expectations in real time. We do so by building an index of inflation expectations and capture the intensity and uncertainty of expectations as well as the quantitative value. The preliminary results show that the new inflation index is correlat-ed with the actual inflation dynamics. Moreover, the inflation news pre-cedes actual inflation by a few months. To validate our results, we build a linear regression using our newly built indices and market-based inflation expectations and confirm that our methodology results in a model with good forecasting power.

Big Data: Does it really improve Forecasting techniques for Tourism Demand in Spain?

Miguel Ángel Ruiz Reina

Abstract: In this study, innovative forecasting techniques and data source from Big Data are used for the study of Hotel Overnight Stays for Spain, from January 2012 to December 2018. The unstoppable development of the tourism sector, together with the application of Big Data technologies, allow to make efficient decisions by economic agents. In this paper, univariate forecasting methodologies such as SARIMA and SSA are used. The use of the data obtained from the Google Data Mining tools allows to obtain knowledge. The ARDL models with seasonality explain easily when economic agents will make their decisions. ECM allows make forecasting

for short-term and long-term. This fact means that tourist offers and demands can be perfectly adjusted at every moment of the year. As a criterion for the selection of models, the innovative Matrix U1 Theil is proposed, this allows to quantify how much a model is better than another in terms of forecasting.

Estimation of parameters and reconstruction of hidden variables for a semiconductor laser from intensity time series

Mikhail Prokhorov, Ilya Sysoev, Vladimir Khorev and Vladimir Ponomarenko

Abstract: We propose a method for the reconstruction of dynamical systems with time-delayed feedback, having several hidden variables, including a hidden variable with a time delay. The method is based on the original approach, which allows one to significantly reduce the number of starting guesses for a hidden variable with a delay. The method is applied to the reconstruction of the model system of Lang-Kobayashi equations, which describes the dynamics of a single-mode semiconductor laser with time-delayed feedback, from periodic and chaotic intensity time series. The dependence of the quality of the system reconstruction on the accuracy of the assignment of starting guesses for unknown parameters and hidden variables is investigated. It is shown that for periodic regimes, the region of starting guesses, which provides high quality of reconstruction, is greater than for chaotic regimes.

Will the spanish converge in the near future?

Sofía Tirado Sarti, Rafael Flores de Frutos and Manuel León Navarro

Abstract: This paper analyses interregional effects on capital stock in Spain in order to get results that allow us to evaluate the efficiency of public investment policies and provide new insights into the analysis of regional disparities. To this aim, the Vector Autoregressive (VAR) methodology is used to estimate the dynamic effects of capital on output and employment and evaluate the relevance and magnitude of regional spillover effects. The empirical results for Spanish economy suggest that there are two groups of regions. GDP and Employment of small Spanish regions seem to be I(1) variables, while their corresponding Capital Stocks seem to be I(2) variables. This contrasts with the statistical properties of the same variables for the Spanish big regions, all being I(2).This result implies that, for the smaller regions, only permanent changes in the rate of growth of Capital Stock can produce permanent effects on the levels of GDP and/or Labor. The same occurs when the investment takes place outside the smaller regions. A small region cannot beneficiate, in the long run, from investments, neither inside nor outside the region itself. These statistical properties suggest that a kind of circular cumulative effects, á la Myrdal, take place which could difficult the convergence (in terms of GDP) of small regions. We conclude that to achieve real convergence in Spanish regions, a long sequence of big pushes in investment is required for the smaller regions to change their growth path and to become a development engine.

Estimation of Vector Long Memory Processes

Hao Wu and Peiris Shelton

Abstract: Time series modelling has been shown to be an effective tool for analyzing data from macroeconomics and finance. There seems to be an increased interest to extend univariate models to multivariate case in various application domains. This paper provides an overview of the most important development in parametric multivariate long memory time series modelling, with estimation mainly based on a Gaussian likelihood. It discusses the model specification and estimation methodology for vector autoregressive fractionally integrated moving average (VARFIMA) model and vector Gegenbauer ARMA (VGARMA) model, in both methodology and empirical applications. It standardizes the state space representation of multivariate time series models and

presents the simulation results of quasi-maximum likelihood estimator via Kalman filter. The performance of different estimation methods in both time domain and frequency domain are also compared. Finally, the likely directions of future research are identified.

A robust method for estimating the number of factors in an approximate factor model

Higor Henrique Aranda Cotta, Valdério Reisen and Pascal Bondon

Abstract: This paper considers the approximate factor model for high-dimensional time series with additive outliers. We propose a robustification procedure of the information criteria proposed by Bai and Ng (2002). The robust estimator of the number of factors is obtained by replacing the standard covariance matrix with M-covariance matrix. Simulations are carried out under the scenarios of multivariate time series with and without additive outliers to assess the impact of additive outliers on the standard information criteria and to analyze the finite sample size performance of the proposed robust estimator of the number of factors.

Monotonicity Assumptions for Recession Forecasting

David Kelley

Abstract: The probability of recession given an economic indicator is often modeled using a probit model. These approaches can produce poorly calibrated estimates in applied settings. The issue is particularly acute when estimating the probability of recession one year ahead using the slope of the yield curve. Replacing the single index assumption of the probit model with a Gaussian processes model provides estimates that are well calibrated. Results from Gaussian process models restricted to be monotonically increasing in the indicator imply that breaking the monotonicity assumption and not just the linearity assumption of a single index model is required for well-calibrated probability estimates.

The Tsallis Statistics Faces Social Problems in Developing Countries

Huber Nieto-Chaupis

Abstract: We apply the well-known Tsallis statistics to model crimi- nality and vehicle chaos in Lima city. Despite of the fact tha Peru has shown interesting indicators of a reduced generalized poverty, the appari- tion of high rates of criminality and high vehicle trac is considered as a logic consequence due to the economical progress of the country. In this paper we argument that the geographical apparition of these social issues has as origin the phase transitions as seen in the Tsallis entropy, from a low to middle social-economical layer, fact that is not contemplated in the regional and urbanistic evolution of large cities such as Lima with a population of around 10M of habitants.

Common trends in producers' expectations: implications for GDP forecasting in Uruguay

Bibiana Lanzilotta, Lucía Rosich and Juan Gabriel Brida

Abstract: This paper examines the interdependence between expectations and economic growth in Uruguay, for the last two decades (1998-2018). To this aim, this research considers the expectation surveys collected by the “Cámara de Industrias del Uruguay” and macroeconomic series of National Account System. The study

achieved two main findings. Firstly, that there is a long-run relationship between producers' expectations and Uruguayan GDP growth controlling by the usual production factors. Secondly, by estimating multivariate structural models we found that there is a common level between the expectation indicators of the different industrial sectors grouped according to production specialization and international trade insertion. The expectation indicator of which the others depends is the one of the more tradable industries. The level component of this indicator of expectations drives the level component of the other groups. The research shows that expectation indicators could be an accurate leader indicator of sectoral and aggregate growth. Moreover, the driver of the producers' expectations is the aggregate indicator of expectation of the more tradable industries.

Latent precursors of delayed river ice-jam shattering: An anthropogenic factor

Alexandre Chmel and Lyubov Banshchikova

Abstract: Dynamics of the ice-jam flooding was analyzed statistically as the process in the non-equilibrium system with the varying conservation. The distributions of the daily average water heights during spring breakups in the basin of Lena river were constructed. It has been found that the water heights during water rising in jam-free years were distributed in a random manner that is in accordance with an exponential (Poissonian-like) function. In contrast, the water height distributions in the periods of ice-jam flooding followed a power law in common with various multiscale hydrological phenomena. The analysis of the height distributions showed some latent perturbations in the water-ice system caused by natural causes or blasting intended to destroy jams. The efficiency aspect of the blasting actions during ice-jam flooding was considered.

Time Series Causality Based on Complex Net-works for the Study of Air-Sea and Climate-Epidemics Coupled Systems

Teddy Craciunescu, Andrea Murari, Michela Gelfusa and Emmanuele Peluso

Abstract: A recently developed measure for the characterization of interconnected dynamical systems is used for the study of several coupled phenomena related to the influence of the El Niño Southern Oscillation on other atmospheric systems and to the influence of climatic factors on malaria epidemics. The method is based on the representation of time series as weighted cross-visibility networks. The weights are introduced as the metric distance between connected nodes. This allows the representation of the adjacency matrix as an image. The structure of the networks, depending on the coupling strength, is quantified via the image entropy

Short-term Temperature Forecasts using Deep Learning – an Application to Data from Ulm, Germany

David Kreuzer, Michael Munz, Samuel Peifer and Stephan Schlüter

Abstract: With the increasing importance of solar energy we need more precise forecasts of its amount in order to guarantee network stability. Forecasting models are normally based on solar radiation, the major impact factor, but also on temperature which impacts the level of efficiency of the solar modules. Since conventional weather models (like the Lorenz model) are extremely chaotic, data driven models are getting more popular. Among those models the field of deep learning, i.e. neural networks with multiple hidden layers, is becoming more relevant. Especially convolutional neural networks which do not have the need of a feature extraction stage are a solid alternative to classic approaches. One of the main reasons is that computation power for massive parallel computing (i.e. GPGPU computing) is increasing. Authors like Dong et al. (2018) or Xiaoyun et al. (2016) use a recurrent neural network to predict temperature, wind speed or radiation, for example. Especially long short-term memory (LSTM) networks are often used, due to the fact that recurrent networks generally

show good results when dealing with time series data (Lopez et al., 2016) and LSTM cells, in particular, have the “ability to bridge very long time lags” (Hochreiter & Schmidhuber, 1997), which is crucial when dealing with seasonal data. To the best of the authors knowledge, nobody applied the combination of convolutional and LSTM layers in this context so far, which is what we do in this work in order to forecast temperature data. Therefore we use weather data from Ulm between 2015 and 2018. The motivation for applying a convolution LSTM network is the assumption that individual measurements are highly correlated. Hence, the use of 2D-convolutional operations is likely to lead to more stable results, since interdependencies between the channels are taken into account. The recurrent connections of the LSTM, again, incorporate the time dependencies. As input data we use wind speed, wind direction, temperature, humidity, dew point temperature, air pressure, global radiation, and diffuse radiation, whereby we have a temporal granularity of 10 minutes. To ensure a better generalization of the network, the data is split day by day and shuffled, meaning that the order of the days, being fed to the network, varies in every epoch. To compensate the risk of losing the annual seasonality and to simplify training, domain knowledge is used. The Integration of prior knowledge is done by incorporating additional information to the networks input representation, specifically month, daytime, and monthly mean temperature according to every data point. We are using L2-regularisation in all layers and dropout only in the dense layers. For training the sliding window approach is used. The model is created with Python 3.6 and Tensorflow. To benchmark our deep neural network, the results are compared with classical methods of time series forecasting such as naive forecasts or the seasonal autoregressive integrated moving average (SARIMA) model. The SARIMA is chosen as we expect the temperature to show distinct daily seasonality. The model is an autoregressive model which accounts for dynamics in the error term but also for trends and seasonal effects, which are captured by a lag term. The degree of integration is determined using the partial autocorrelation function; all further parameters are identified by (conditional) maximum likelihood estimation (see e.g. McNeil et al., 2016). For the case study we choose three different time horizons, namely 6, 12 and 24 hours. As performance measures we use the root mean squared error (RMSE) and the standard deviation of the mean absolute error (MAE) for the first up to the 144th time step (i.e. 24 hours). Both models show good results for consecutive days with similar patterns. Since the neural network also uses information from other channels, it shows better results than the SARIMA model when the weather changes drastically in comparison to the day before. Nevertheless feeding more information is not always beneficial – we see that the SARIMA model often outperforms the neural network in the first few hours. However, for larger forecasting horizons, the neural network delivers more accurate results. Regarding the prediction time, neural network prediction is independent from the newest data and has to be trained only once. After training, which takes much longer than the training of the SARIMA, it only requires data from the last six hours for prediction, while SARIMA always has to be fit to the data of the last week. Literature Hochreiter S, & Schmidhuber J (1997). Long Short-term Memory. Neural Computation; 9, pp. 1735-1780. Dong D, Sheng Z, Yang T (2018). Wind Power Prediction Based on Recurrent Neural Network with Long Short-Term Memory Units. 2018 International Conference on Renewable Energy and Power Engineering (REPE), Toronto, ON, Canada, pp. 34-38. Lopez L, Valle C, Allende H (2016). Recurrent Networks for Wind Speed Forecasting. International Conference on Pattern Recognition Systems (ICPRS-16), Talca, pp. 1-6. McNeil A J, Frey R, Embrechts P. Quantitative Risk Management: Concepts, Techniques, and Tools. Princeton University Press: 2006. Xiaoyun Q, Xiaoning K, Chao Z, Shuai J, Xiuda M (2016). Short-term prediction of wind power based on deep Long Short-Term Memory. 2016 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC), Xi'an, pp. 1148-1152.

Multiple change-point estimation of multi-path panel data via EM algorithm

Jaehwi Kim and Jaehee Kim

Abstract: This paper addresses the problem of estimating multiple common change-points in multi-path panel data with the unknown num- ber of change-points. The EM algorithm is used for maximization of the likelihood for mixture distributions. With the unknown number of change-points in parameters, the variable dimension of parameters causes computational diculty in searching and resulting appropriate parti- tions. We suggest the tail-cutting algorithm which is recursive and repet- itive as an alternative for binary search algorithm. We also extend our method to the multiple common change-points estimation to multi-path time series panel data. Simulations to evaluate the performance of the estimators are provided.

Time Series Generation using a 1D Wasserstein GAN

Kaleb Smith and Anthony Smith

Abstract: Time series data is an extremely versatile data type that can relate many real world events; however the acquisition of it requires special sensors, devices, and man power to be able to record the events and translate to one dimensional (1D) data. This is costly in effort on the labor side and in many cases the events do not occur plentifully which results in a lack of time series data describing these events. This paper looks to address that issue of a shortage of time series data by implementing a one dimensional improved training Wasserstein Generative Adversarial Network utilizing a residual network architecture in the generator. With this framework, we learn specific time series data distribution and are able to autonomously generate synthetic data resembling these actual real world events covering a multitude of different time series data types (sensors, medical, imaging, etc). We justify our method visually by comparing the mean envelope of the real data versus the actual generated synthetic data, and we also look to justify it by a machine learning approach, having our synthetic data classified by a state-of-the-art time series classifier into its appropriate class. Our method shows to perform well on generating time series data classifying in the same percentage of accuracy as the original data sets had.

Forecasting using Big Data: The case of Spanish Tourism Demand

Miguel Ángel Ruiz Reina

Abstract: The development of new technologies, especially in the digital field applied to Tourism Industry, has meant a revolutionary change in Data Analysis. Specifically, due to data obtained from social networks, environment, images, videos, sound or any tool from the Internet of Things (IoT). In this article, we will work in an applied way on the hotel demand in Spain and through the dynamic correlations generated worldwide. The Big Data set will come from Google applications. Granger-Causality test extended to seasonality allows determine causality relationships and through VAR (VECM) analysis when consumers decide to confirm their hotel booking. The endogenous or exogenous dynamics effects will be analysed by Impulse-Response Analysis. All this analysis will be compared with the results of ADRL multivariate models with seasonality, and univariate techniques such as SARIMA and Hierarchical Neural Networks. The best model of forecasting will be selected by Matrix U2 Theil. Tourism market is provided of knowledge through Data Analysis.

Estimation of the crustal velocity field in the Balanegra fault from GPS position time series in 2006 - 2018

Antonio J.Gil

Abstract: In 2006 a non-continuos GPS network was installed to monitor the Balanegra fault, one of the most active faults recognized in the Campo de Dalias area. This is a zone with relevant seismicity associated with the active tectonic deformations of the southern boundary of the Betic Cordillera. The goal of our research is to constrain the activity of this fault from high quality GPS measurements to obtain precise deformation rates. Here we show the first results computed from five GPS campaigns in the timespan 2006-2018. These results agree with the general movement of the area. Moreover, we have identified a residual velocity field with values ranging from 1.7 to 3.0 mm/yr. There is a clear geodetic evidence of recent and active deformation by folding and faulting in the eastern Betic Cordillera, which currently undergoes the NNW-SSE convergence of the Eurasian and Afri-can plates, while also affected by orthogonal extension. The integration of further geologic and seismic information with the GPS site rates will allow a better un-derstanding of the kinematics of this interesting geologic structure.

Electricity Load Forecasting - An Evaluation of Simple 1D-CNN Network Structures

Christian Lang, Florian Steinborn, Oliver Steffens and Elmar W. Lang

Abstract: This paper presents a convolutional neural network (CNN) which can be used for forecasting electricity load profiles 36 hours into the future. In contrast to well established CNN architectures, the input data is one-dimensional. A parameter scanning of network parameters is conducted in order to gain information about the influence of the kernel size, number of filters, and dense size. The results show that a good forecast quality can already be achieved with basic CNN architectures. The method works not only for smooth sum loads of many hundred consumers, but also for the load of apartment buildings.

A study of variable importance in multiclass classification problems based on the Volume Under the Surface measure

Ismael Ahrazem Dfuf, José Manuel Mira McWilliams and Mª Camino González Fernández

Abstract: When dealing with more than two classes, classification problems become more complex, and hence the variable importance analysis. In this article, we present a comparative study of variable importance analysis in the multiclass case. The applied methodology uses the Conditional Inference Tree algorithm (CIT) (as it perfoms well in pronounced imbalanced data sets) along with the permutation importance technique. The comparison consists of evaluating the methodology using different performance measures: The Volume Under the Surface (VUS) and its computation alternatives, the G-Mean measure and the generalization of the Are Under the Curve (AUC). Several simulated scenarios characterized by their linearity and multiclass imbalance levels are built and tested.

A machine learning-based approach to forecasting alcoholic relapses

Nikola Katardjiev, Steve McKeever and Andreas Hamfelt

Abstract: This research aims to explore alcoholic relapses by modelling four types of machine learning algorithms on clinical trial data of patients in an alcohol addiction treatment plan provided by an Uppsala-based company called Kontigo Care, with the goal of predicting downwards trends in the data that could indicate a relapse. The learning algorithms were support vector machine, random forest, decision trees, and a K-nearest neighbour regressor. Results indicated that using a random forest model, it is possible to forecast future data entries in the particular data set by beating the benchmark of a baseline predictor, suggesting that alcoholism could indeed be predicted, and thus future work leveraging larger data sets and deep learning methods could improve relapse predictions even further.

Improved Extreme Rainfall Events Forecasting Using Neural Networks and Water Vapor Measures

Matteo Sangiorgio, Stefano Barindelli, Riccardo Biondi, Enrico Solazzo, Eugenio Realini, Giovanna Venuti and Giorgio Guariso

Abstract: In the last few years, many studies claimed that machine learning tools would soon overperform the classical conceptual models in extreme rainfall events forecasting. In order to better investigate this statement,

we implement advanced deep learning predictors, such as the deep neural nets, for the forecasting of the occurrence of extreme rainfalls. These predictors are proved to overperform more simple models such as the logistic regression, which are traditionally used as a benchmark for these tasks. Also, we evaluate the value of the information provided by the Zenith Tropospheric Delay. We show that adding this variable to the traditional meteorological data leads to an improvement of the model accuracy in the order of 3-4 %. We consider an area composed by the catchments of four rivers (Lambro, Seveso, Groane, and Olona) in the Lombardy region, northern Italy, just upstream from the metropolitan area of Milan, as a case study. Data of convective extreme rainfall events from 2010 up to 2017 (more than 600 extreme events) have been used to identify and test the predictors.

Statistical Approach to Predict Meteorological Material for Real-time GOCI Data Processing

Hyun Yang

Abstract: The Geostationary Ocean Color Imager (GOCI) can be utilized to analyze subtle changes on oceanic environments because it observes ocean colors around the Northeast Asia hourly, for 8 times a day. To realize this, the Korea Ocean Satellite Center (KOSC) which is the main operating agency of GOCI has a role to receive, process, and distribute its data within an hour. In this situation, we need several meteorological materials (e.g., ozone, wind, relative humidity, pressure, etc.) to successfully process the GOCI atmospheric corrections. Meteorological materials from National Aeronautics and Space Administration (NASA) Ocean Biology Processing Group (OBPG) are used when the GOCI atmospheric corrections are processed. Unfortunately, however, these materials cannot be used for the real-time GOCI data processing because they cannot be provided in real-time. In this paper, therefore, we propose a statistical approach for predicting the meteorological material and analyzed its accuracy.

New Technique for Risk Measurement: Beyond Conventional Methods

Maryam Zamani, Ali Namaki, Gholamreza Jafari and Holger Kantz

Abstract: Risk management is a crucial factor to consider before any investment decision. Risk is measured based on the deviation of the prices from our expectation, most common way to measure risk is using standard-deviation. In this abstract, we introduce a new method for risk measurement using Level-crossing analysis. In this method, we calculate the waiting time and the frequency of different events in the markets. The results are compared with the corresponding fBm data with the same Hurst exponent as the time series of stocks. For instance US stock markets Dow Jones Industrial Average ($DJIA$) and $S\&P$ have Hurst-exponent around $H=0.5$, Shanghai Stock Exchange (SSE) and Tehran Stocks ($TEPIX$) have Hurst Exponent around $H=0.6$. Therefore, according to the Hurst exponents of the markets, the frequency of the different values in fBm data with the same Hurst exponent is an appropriate standard reference point for our expectation of that market . Results show, although US stock markets are more active and have less average waiting time in compare with other two markets, they have a high compatibility with fBm data, specially in extreme events demonstrate the low level of risk in these markets. At the other hand Shanghai and Tehran markets are less active, having higher average waiting time and deviate from its corresponding fBm data, show the high level of risk in these markets.

Power transformer monitoring based on a non-linear autoregressive neural network model with exogenous inputs

Javier Ramirez, Francisco J. Martinez Murcia, Fermin Segovia, Susana Carrillo, Javier Leiva, Jacob Rodriguez-Rivero and Juan M. Gorriz

Abstract: Next generation of low- and medium-voltage distribution networks will demand to be better planned, operated and supervised as transportation networks have been managed for decades. The adaptation of these networks will require to incorporate much more intelligence, sensorization, broadband communications, optimal control and intelligent reporting among other emerging technologies. This paper shows a non-linear autoregressive neural network with exogenous inputs (NARX) for time series forecasting in these scenarios. The network model provides a description of the system by means of a non-linear function of lagged inputs, outputs and prediction errors that can be interpreted as a recurrent dynamic network, with feedback connections enclosing several layers of the network. It consists of a multilayer perceptron (MLP) in the hidden layer that takes as input a window of past independent (exogenous) inputs and past outputs followed by an output layer that finally forecast the target time series. The proposed NARX network was trained and evaluated in open-loop and closed-loop modes for prediction of the temperature of the transformer based on other correlated exogenous inputs that were recorded at the power transformation centers. The results show that the NARX networks can accurately predict and monitor the operation of power transformers.

Partial Least Squares for the Characterization of Meditation and Attention States

Jorge García-Torres, Juan Manuel Górriz, Javier Ramírez and Francisco Jesús Martínez-Murcia

Abstract: Electroencephalographic activity (EEG) has allowed the exploration and development of a new non-muscular communication channel under the name of brain-computer interfaces (BCI), which eases the study of neurological disorders and brain functions. This paper deploys machine learning algorithms based on partial least squares feature extraction to characterize two easily recognizable cognitive states: attention and meditation. The classification problem is addressed by using the power spectrum density of the raw EEG values provided by a low-cost BCI system as the input data of our experimental approach. Accuracies around 80% over a dataset of 30 subjects are obtained by combining the two first PLS components with a RBF kernel, significantly improving the results provided by other approaches.

A time-varying Markov-switching regimes in a financial stress transmission. Evidence from Non-Eurozone Visegrad Group Countries

Magdalena Ulrichs

Abstract: The main purpose of the article is to identify structural changes in the reactions of macroeconomic variables due to occurring financial stress episodes. In our empirical analysis, we use Marcov-switching vector autoregression models with time-varying transition probabilities. The main attention of our analysis is on the monetary policy and banking sector and its reactions on the financial stress shocks during two regimes: normal times and financial stress episodes. The obtained empirical results indicate that, during periods corresponding to the high levels of financial stress index, reactions of variables included in the system are stronger and more persistent than during periods of lower stress level.

Preparation of training data by filling in missing vessel type data using deep multi-stacked lstm neural network for abnormal marine transport evaluation

Julius Venskus and Povilas Treigys

Abstract: Highly-loaded seaports have extremely complex and intensive marine vessel traffic, which generates large volumes of traffic data. Meteorological conditions and maritime vessel type influence maritime traffic and they must also be taken into account in order to train the model capable of recognizing the abnormal movement of the sea transport. Real data often misses some data values, such as type of vessel or its status. Data that influences maritime traffic must also be taken into account in order to train the model capable of recognizing the abnormal movement of the sea transport. Typically, in every region missing data differs. This paper reviews methods of obtaining vessel traffic and meteorological data and filling missing vessel type data in regions such as Rotterdam, Gdansk, Malmö, Copenhagen, Kiel and Lübeck. Author describes possibilities to obtain technical parameters of ships and existing limitations while taking into account the available vessel data. By assessing the characteristics of the received traffic data, the regionalization of the meteorological data and the assignment to a specific maritime traffic data vector is performed by using the method of the closest neighbor, depending on the time of traffic in a given region. A deep multi-stacked LSTM neural network model is trained to fill the missing vessel type data. This model is trained with available vessel type data and used to predict missing values. This paper describes creation and evaluation of this model.

Calendar based forecast of emergency department visits

Cosimo Lovecchio, Mauro Tucci, Sami Barmada, Andrea Serafini, Luigi Bechi, Mauro Breggia, Simona Dei and Daniela Matarrese

Abstract: In this paper we use a well established method for short-term forecasting to predict the amount of hourly Emergency Department (ED) visits in thirteen different hospitals in the south-east area of Tuscany. Our algorithm belongs to the class of similar shape algorithms and perform the forecast in an unsupervised manner. It exploit an historical dataset containing the patient arrival data, in which similar pattern, filtered on the base of a calendar condition, are selected to predict the incoming visit volume for a tunable number of day ahead.

Short-term solar power forecasting using clustered VAR model over South Korea

Jin-Young Kim, Chang Ki Kim, Hyun-Goo Kim, Yung-Seop Lee and Yong-Heack Kang

Abstract: Abstract. Solar power forecasting is a key role for large-scale photovoltaic penetration to distributed generation and management system. This paper proposes an empirical statistical model for forecasting hourly power genera-tion up to three hours as a function of remotely sensed irradiance, UASIBS/KIER (University of Arizona Solar Irradiance Based on Satellite-Korea Institute of Energy Research) over the South Korea. The model uses the vector autoregression for multivariate irradiance time series over a period of about three years (2014 to 2016). Forecasting hourly power generations were additionally conducted using k-mean clustering to classify the national scale forecasting system. The results showed that the mean absolute percent-age error with maximum 87.8%.

Forecasting Energy Consumption in Residential Buildings using ARIMA Models

Muhammad Fahim and Alberto Sillitti

Abstract: Forecasting energy consumption in residential buildings is an important information for energy providers to meet the demands of consumers and plan properly grid resources. Moreover, it can be useful to

support residents to reduce their utility bills and save energy. In this paper, we presented preliminary short-term forecasting results by utilizing autoregressive integrated moving average (ARIMA) models. Our analysis also connects the forecasting to the number of residents, employment status, number of electrical appliances, and size of the house. We evaluated our model on a publicly available dataset and optimal parameters were obtained through grid search. The mean absolute percentage error (MAPE) is calculated to quantify the forecasting error at 17.25%. The obtained results confirmed the applicability of the model to real-life applications.

Predicting hospital admissions with integer-valued time series

Radia Spiga, Mireille Batton-Hubert and Marianne Sarazin

Abstract: Introduction: Prediction of seasonal epidemics have been widely treated in the medical literature, with various methods to forecast the future cases of a given disease, when it's about infectious diseases: compartmental methods that forecast the number of persons at each state of the epidemic (susceptible, infected, resistant) are used, as well as methods based on time series (ARMA ,ARCH,…), this last method can be successful with large sample of data, assuming their normal distribution. Here, we propose to test time series for count data when the continuous time series are not adapted to predict health activity. The aim of this work is to predict the number of hospital admissions of future weeks at local scale, using methods for integer-valued time series: INAR(p) and INGARCH(p,q) methods, we have also test the autoregressive methods for continuous data. Data: Weekly admissions for pulmonary diseases at the Saint Etienne University Hospital (France), and weekly number influenza like illness (ILI) provided by The French Sentinelles Network (the network of general practitioners(6)), from 2011 to mars 2018. The first step is to describe and characterize these two data, then to apply the appropriate count time series, next multivariate method for count time series will be tested. Data characterization: Overdispersion was observed for both ILI and hospital data, suggesting a negative binomial distribution, and there were 133 count of 0 on ILI data. Form this results we assume a negative binomial distribution for hospital data, and zero-inflated negative binomial distribution for ILI data. Data fitting and forecasting: Hospital admission time series have been predicted with a better accuracy with the NB INGARCH (1,1). For the ILI time series ZINB INAR have failed to predict the outbreak of epidemics, higher order of ZINB INAR will be tested. Perspective: The present work is based on an empirical approach; we are planning to develop a similar approach based on Bayesian methods, combining specificity of the integer value data. The final aim is to predict hospital admissions, including ILI series as an exogeneous variable in the count tie series model.

Neural Network approaches for Air Pollution Prediction

Marijana Cosovic and Emina Junuz

Abstract: According to World Health Organization exposure to elevated levels of air pollutants is directly linked to health issues. Air pollution in the countries of Western Balkans is an evident problem. Sarajevo, the capital of Bosnia and Herzegovina, is at the end of 2018 presented as the world's most polluted city. Previous years it was a runner up being amongst five most polluted cities together with cities from Macedonia and Montenegro. Air pollution forecasting could be performed using machine learning techniques. Deep learning techniques are proving useful since weather behavior can be analyzed as time series data. We observed hourly, daily, weekly and seasonal periodicities of the air pollutant and meteorological data. Considering the data collection process was not the most reliable in the past, although improving recently, we need to evaluate and address the missing data issue to deliver recommendations on air quality using machine learning techniques. In addition, this paper evaluates performance of several neural network architectures applied to weather data in greater Sarajevo area for the 2014-2018 period. Urban air quality is of outmost importance to the public hence we compare prediction accuracies of the proposed forecasting models.

Long and Short Term Prediction of PowerConsumption using LSTM Networks

Juan Carlos Morales, Salvador Moreno, Carlos Bailón, Héctor Pomares, Ignacio Rojas and Luis Javier Herrera

Abstract: This work presents an adapted deep learning approach for short and long term prediction of a national power consumption time series. A modified LSTM network based on direct prediction of four hours horizon is presented, and the improvements in efficacy by using external inputs for the prediction are shown. Additionally, an alternative based on Convolutional Neural Networks is applied to the same objective, and results are compared. Finally, the performance of those models for long term time series forecasting is assessed, and a modification of the training process on the LSTM network adapted to long term prediction is shown, leading to relevant improvements in accuracy for this task.

Time Series Classification of Automotive Test Drives Using an Interval Based Elastic Ensemble

Felix Pistorius, Daniel Grimm, Marcel Auer and Eric Sax

Abstract: The development of autonomous vehicles leads to an increasing number of necessary test drives with different scenarios and driving maneuvers. These are needed to validate the growing number of functions and to ensure safe driving behaviour. During a test drive, data from a variety of sensors is recorded for vehicle development. A particularly interesting use case for this data is the analysis of individual driving situations of the entire test drive, which are relevant for the refinement of the vehicles under development. For the chassis calibration, for example, it is of special importance to consider track sections with certain surfaces and curve characteristics, since only in these time periods of the recording relevant information for the development can be found. Therefore, this paper presents a method to automatically and robustly retrieve specific sequences in test drives. The method proposed by us is able to identify any desired time segments and thus to select only the relevant segments from the dataset of all test drives. It should be particularly emphasized that our method finds the desired time segments even for datasets that originate from different vehicles, driving styles and routes. This paper therefore makes a valuable contribution to reducing the development effort for vehicles. Our concept is based on an efficient ensemble of the elastic time series classification algorithms Dynamic Time Warping (DTW), Derivative DTW, Weighted DTW and Move-Split-Merge (MSM). Our evaluation results on both synthetic and real-world data are promising.

Modeling recession curves in a karstic aquifer

Roger Gonzalez-Herrera, Carlos Zetina-Moguel and Ismael Sánchez Y Pinto

Abstract: The analysis of recession curves (RC) is used in research, administration and management of water resources. In karst systems it has been used to study the dynamics of groundwater, particularly in the process of groundwater discharge during the dry season. There is a diversity of models, published in the literature, based on different physical principles; examples of them are the Boussinesq (BM) and Maillet (MM) models, which were derived from the theoretical equations of groundwater flow. A study was conducted whose objective was to determine the model that best represents the hydrodynamics of a karst aquifer located in a plane area. To achieve the above, statistical criteria were defined to select the model; a statistical analysis strategy was implemented to compare the estimates of the parameters of greatest hydrological interest of the chosen decay models. The data used come from three wells located in the northern plain of Yucatan, Mexico, where hydraulic heads were daily measured for different periods of time. The scenarios considered for the analysis were: I.- Two wells with the same time intervals with standardized data of: different initial hydraulic head (ho); ho similar; ho modified. II.- A single well, starting from an established hydraulic head and with the same period of discharge time. The adjustment of the models to the observed RCs was carried out by means of numerical

methods and parameters estimation by means of a maximum likelihood function. To choose the appropriate model, the following statistical criteria were established: the sum of squared errors (SSE); the adjusted coefficient of determination (R2adj); a function of maximum likelihood and the Akaike criterion. For the Boussinesq model, estimates of the α parameter were explored, using fixed values of the parameter n, in the range of 0.5 to 5. The comparison between the estimates of α was made based on hypothesis tests (t to compare between two values of α and F for the comparison between more than two values of α); quasiprobability distributions obtained by a maximum likelihood estimator were also used. The Boussinesq model proved to be the most appropriate. The best estimates of α (BM) for hydrological interpretations are obtained with values of the parameter n between 0.5 and 1.5.

The HJ-Biplot Visualization of the Singular Spectrum Analysis Method

Alberto Silva and Adelaide Freitas

Abstract: Time series data usually emerge in many scientific domains. The extraction of essential characteristics of this type of data is crucial to characterize the time series and produce, for example, forecasts. In this work, we take advantage of the trajectory matrix constructed in the Singular Spectrum Analysis, as well as of its decomposition through the Principal Component Analysis via Partial Least Squares, to implement a graphical display employing the Biplot method. In these graphs, one can visualize and identify patterns in time series from the simultaneous representation of both rows and columns of such decomposed matrices. The interpretation of various features of the proposed biplot is discussed from a real-world data set.

Linear regression model for prediction of multi-dimensional time-point forecasting data

Shrikant Pawar and Aditya Stanam

Abstract: Machine Learning (ML) has been promising in predicting financial series, the direction of the medicine, stock market, macroeconomic variables, accounting balance sheet information and many other applications. Regression models a target prediction value based on independent variables. Different regression models differ based on the kind of relationship between dependent and independent variables they are considering and the number of independent variables being used. The time series signature is a collection of useful features that describe the time series index of a time-based data set. It contains a wealth of features that can be used to forecast time series that contain patterns. The user can learn methods to implement machine learning to predict future outcomes in a time-based data set. Testing the data with linear regression provided higher R-Squared, Adj R-Squared and F-Statistic values. The t-statistic was greater than 1.96 with a p-value less than 0.05. MAPE (Mean absolute percentage error) and MSE (Mean squared error) were reasonably lower. The paper concludes that linear regression can be an effective ML for forecast data provided dimension reductions have been taken into considerations.

Occupancy Forecasting using two ARIMA Strategies

Tiên Dung Cao, Laurent Delahoche, Bruno Marhic and Jean-Baptiste Masson

Abstract: We present an occupancy forecast method in a smart home context based on the exploitation of environmental measures such as CO2, sound or relative humidity. This article presents our machine learning algorithm and prediction strategy. It is based on two levels of data exploitation. The first level is “supervised learning” to obtain past occupancy from sensor measurements. It is achieved with a multiple logistic regression algorithm. The second level consists in two main steps. During the first step ARIMA learns and trains the model,

using the past occupancy data from level 1. During the second step ARIMA predicts the future occupancy. The innovative part of our paper is that we compare two different ARIMA's (de-seasonalised). The first is the "day-sequence-time-series" (a serial ARIMA). The second is the "daily-slice-time-series" (a parallel ARIMA). We conclude by analysing the performance of our occupancy prediction paradigm.

Engineering Data for Business Forecasting

Klaus Spicher

Abstract: Section 2 describes two innovative ideas/conceptions, enabling the development of new forecasting methods. One related method (algorithm) based on the described ideas is presented in Section 3. Section 4 deals with the development of a new heuristic approach enabling (beyond Croston-/ and WSS-Methods) forecasting the number of sporadic events along the forecast period and a related metric measuring the result.

Evaluating the effectiveness of transportation information provision in the sharing economy context

Joshua Paundra, Jan van Dalen, Laurens Rook and Wolfgang Ketter

Abstract: Information communication and technology (ICT) based transportation applications have facilitated the provision of transportation information related to environmental impact, trip price, and traffic situations to commuters. ICT advancement is also central in the rise of sharing economy transportation services, such as ridesharing. These developments raise the possibility of encouraging people to consider alternative transportation options other than private vehicles. This study considers the use of information provision as a soft policy intervention in the presence of sharing economy transportation options. Based on an experiment, we showed that providing information on environment, price, and traffic enabled people to respectively consider the more environmentally friendly, cheaper, and lower travel time options. These effects, in part, depend on people's individual psychological ownership. Providing relevant transportation information seemed to be a promising soft policy approach.

The influence of local terrain variations on spectral analysis of insolation time-series in Sierra Nevada (Granada province, southern Spain)

José Sánchez-Morales, Eulogio Pardo-Igúzquiza and Francisco J. Rodríguez-Tovar

Abstract: The widespread availability of marine and terrestrial climate proxies record-ing global events allows the identification of past and present climatic cycles. Basically, these proxies can be understood as different ecological and geological processes ruled in first instance by orbital phenomena, transforming energy inputs (usually from the Sun) into noisy signals or time-series. Characterizing past insolation by means of orbital solutions helps to improve the understanding of the global climatic context inherent to these systems. However, we want to raise the attention on a type of noise, which is the effect produced by the terrain on blocking the Sun light in mountainous areas. This article focuses on the effects of different orography settings on the local insolation calculations and ultimately, it shows how these alterations can produce different spectral signals.

Linking high-resolution marine data sets and the field of time series analysis – The long-term observational records from Helgoland and Sylt (North Sea)

Mirco Scharfe

Abstract: Long-term observations are key to the detection and understanding of changes in the marine environment. The observations at the coastal stations Helgoland and Sylt, starting in year 1962 and year 1973, represent a unique combination of time series length, temporal resolution, and level of detail (range of parameters). Some of these time series, as water temperature at Helgoland Roads, cover around 13500 daily measurements in the past 57 years. Our continuous monitoring documents climatic change, such as the stronger than global warming trend at Helgoland Roads (0.034°C/yr, 1962-2018), and clearly shows the strong trends in nutrient loadings in the past decades. Moreover, time series of hydrographic parameters, nutrients and food-web components serve to develop new hypotheses, as a basis for experimental and model approaches to investigate anthropogenic and naturally driven long-term changes in the marine ecosystem. Our time series provide excellent properties to gain new knowledge by interdisciplinary approaches, e.g. by linking to approaches of non-linear time series analysis as well as approaches to derive causality from observations alone. I will show an overview of the different data series, including their main statistical features and will provide examples on how these long-term observations have contributed to the analysis of change in the North Sea. I will discuss approaches to further develop the explanatory power of the data records, e.g. by using time series techniques to derive new information content inherent to the time series. One important goal related to this is to gain more knowledge about the past and future response of the North Sea system to climate variability. Such findings will contribute to the improvement of information base for scientific use and the decision-making by society, politics, and authorities.

The Prediction Analysis of Zero Inflated Poisson Autoregression Model for the Number of Claims in General Insurance

Utriweni Mukhaiyar, Adilan Widyawan Mahdiyasa, Sapto Wahyu Indratno and Maudy Gabrielle Meischke

Abstract: The number of claims happen in a fixed interval time, is highly possible to be a rare event. It make the series data has many zeros frequency, and the variance data is much higher than its mean, which called as over dispersion. Commonly, to model the probability distribution of frequency data, the poisson distribution is favorable. However for over-dispersed model, it no longer appropriate. The Zero Inflated Poisson (ZIP) Autoregression be the strong candidate to solve it. This model offer prediction of upcoming count data through its probability distribution. Here, this prediction method is equipped with the analysis of cumulative distribution function behaviours which assumed to follow beta distribution. Through this approach, the at most upcoming count data can be predicted as a single number instead of its probabilty distribution. For case study, the number of general insurance happen in Jakarta City is used.

Very Short Term Time-Series Forecasting of Solar Irradiance Without Exogenous Inputs

Christian Hans and Elin Klages

Abstract: This paper compares different forecasting methods and models to predict average values of solar irradiance with a sampling time of 5 min over a prediction horizon of up to 3 h. The methods considered only require historic solar irradiance values, the current time and geographical location, i.e., no exogenous inputs are used. Nearest neighbor regression (NNR) and autoregressive integrated moving average (ARIMA) models are tested using different hyperparameters, e.g., the number of lags, or the size of the training data set, and data from different locations and seasons. The hyperparameters and their effect on the forecast quality are analyzed

to identify properties which are likely to lead to good forecasts. Using these properties, a reduced search space is derived to identify good forecasting models much faster.

The effect of Daylight Saving Time on Spanish Electrical Consumption

Eduardo Caro Huertas, Jesús Juan Ruiz, Jesús Rupérez Aguilera, Carlos Rodríguez Huidobro, Ana Rodríguez Aparicio and Juan José Abellán Pérez

Abstract: In this work, a detailed simulation-based analysis is conducted to assess the impact of adopting Daylight Saving Time (DST) on the consumption in the Spanish electric power system. To carry out this study, it has been used the short-term electric load forecasting software currently used by the Spanish transmission system operator, simulating the load in case of removing the DST in Spain. Obtained results denote that the DST may have a positive impact on the reduction of the electric energy demand.

Wind Speed Forecasting Using Kernel Ridge Regression

Mohammad Alalami, Maher Maalouf and Tarek El Fouly

Abstract: Wind speed forecasting is considered a challenging task as wind energy data is highly variable. Therefore, powerful methods are required in the prediction of wind speed. In this paper, a kernel ridge regression model is proposed. The model performance accuracy, for wind speed forecasting, is compared with two reference prediction models, namely, the persistent model and the least squares model. For the least squares and kernel regression model, moving window cross-validation is used. Historical wind speed data from the Canadian weather station is used to validate the model performance for three different forecasting horizons (1-hour, 12-hours and 24-hours ahead). Results show that forecasts made with the kernel ridge regression produced the highest accuracy compared to the least squares model and the persistent model.

Evaluating the impact of solar and wind production uncertainty on prices using quantile regression

Mauro Bernardi and Francesco Lisi

Abstract: A long abstract is provided in a separated pdf file.

Interpretation of Kuwait Power System through ARIMA Model

Sarah Alosaimi and K.J. Sreekanth

Abstract: Kuwait is considered one of the most important oil producing countries in the world whose economy depends heavily on the oil sector, by nearly 90%. The electricity production, transmission and distribution, and consumption for the state of Kuwait needs a thorough understanding in order to identify the various factors that influences the power sector scenario. By analyzing power data and recognizing the effect of factors such as per capita income, number of buildings, and annual production of electricity on the total primary energy supply of the state, the need for a comprehensive and sustainable power system in Kuwait can be forecasted. In this research, regression analyzes was used to determine the factors influencing between the various elements, so as to present the different future scenarios, using ARIMA, to predict energy consumption during the period from 2016 to 2030. The result shows that GDP and electricity generated have higher influence than number of building in 2030 based on the study.

Estimating the Unknown Parameters of a Chaos-Based S-Box from Time Series

Salih Ergun

Abstract: A novel estimate system is proposed to discover the security weaknesses of a chaos based S-Box and a chaos based random number generator (RNG) it contains. Convergence of the estimate system is proved using auto-synchronization. Secret parameter of the target chaos based RNG are recovered where the available information are the structure of the chaos based S-Box and a scalar time series observed from the target chaotic system. Simulation and numerical results verifying the feasibility of the estimate system are given such that, next bit can be predicted while the same output bit sequence of the chaos based RNG used for S-Box generation can be regenerated.

GNSS based Automatic Anchor Positioning in Real Time Localization Systems

Andreas Heller, Ludwig Horsthemke, Marcel Gebing, Goetz Kappen and Peter Gloesekoetter

Abstract: This work describes the setup of an ultrawideband (UWB) realtime localization system (RTLS) in which the determination of an- chor coordinates is automated by assistance of global navigation satellite systems (GNSS).

Comparison of machine-learning methods for multi-step-ahead prediction of wave and wind conditions

Mengning Wu, Zhen Gao, Christos Stefanakos and Sverre Haver

Abstract: Short-term prediction of wave and wind conditions is important for decision-making during the execution of marine operations. Typically, short-term wave and wind conditions can be forecasted using physics-based models or data-driven models. This paper addresses machine-learning methods for weather forecast. So far, one-step-ahead forecasting can be handled well based on the various methods, while multi-step-ahead forecasting of weather conditions is still challenging and rarely studied. This paper aims to present a comparison of the application of different machine-learning methods in multi-step-ahead wave and wind predictions. Before prediction, two pre-processing techniques, namely decomposition technique and empirical mode decomposition (EMD) are considered on the time series to reduce the prediction complexity. Several widely known data-driven models including the autoregressive integrated moving average (ARIMA), artificial neural network (ANN), recurrent neural network (RNN) and the adaptive-network-based fuzzy inference system (ANFIS) are employed. To perform multi-step forecasting, three multi-step-ahead models (M-1, M-N and M-mN model) are considered. The data set used to assess the accuracy of the methods comprises hourly time series of the mean wind speed Uw, the significant wave height Hs and the spectral peak period Tp from 2001 to 2010 at the North Sea center. After prediction, an uncertainty quantification analysis is conducted to compare the forecasting performance of the aforementioned methods. The results indicate that the considered machine-learning methods can provide acceptable predictions for weather conditions for the first several steps ahead. However, all methods encountered an essential problem that the forecast uncertainty would increase significantly with the forecast horizon. In this case, more research is needed to develop a method that incorporates physical processes into the data-driven model to improve the weather forecast accuracy.

Forecasting Anomalous Events And Performance Correlation Analysis In Event Data

Sonya Leech and Bojan Bozic

Abstract: Classical and Deep Learning methods are quite common approaches for anomaly detection. Extensive research has been conducted on single point anomalies. Collective anomalies that occur over a set of two or more durations are less likely to happen by chance than that of a single point anomaly. Being able to observe and predict these anomalous events may reduce the risk of a server's performance. This paper presents a comparative analysis into time-series forecasting of collective anomalous events using two procedures. One is a classical SARIMA model and the other is a deep learning Long-Short Term Memory (LSTM) model. It then looks to identify if an influx of message events have an impact on CPU and memory performance. The findings of the study conclude that SARIMA was suitable for time series modeling due to the elimination of heteroskedasticity once transformations were implemented, however it was not suitable for anomaly detection based on an existing level shift in the data. The deep learning LSTM model resulted in more accurate time-series predictions with a better ability to be able to handle this level shift. The findings also concluded that an influx of event messages did not have an impact on CPU and memory performance.

GPU forecasting for big data problems

Juan Ramon Trapero, Enrique Holgado, Francisco Ramos and Diego J. Pedregal

Abstract: High Performance Computing via General Purpose Graphical Processing Unit (GPU) is a potential instrument to speed up computational times. In a world where big data is becoming a revolution, GPU could play an important role. This work intends to analyze the performance of GPU by implementing the calculation of probabilistic forecasts based on single exponential smoothing in conjunction with simulated predictive distributions. Essentially, supply chain companies must deal with a high number of forecasts at SKU level. In this context, reducing the computational times can be a source of a competitive advantage. Since the forecasts are usually made independently between SKUs, this problem can be easily parallelized and GPU computing can exploit such parallelization. To the best of authors knowledge, this is the first time GPU is applied to a supply chain demand forecasting context. Firstly, we will show how to adapt the programming of probabilistic forecasts in a parallel fashion. Then, real data coming from a manufacturer company will be used to illustrate the differences between GPU and traditional CPU computing. The results show that GPU can significantly increase the computational speedup ratio more than 30 times with respect to traditional CPU computing.

Could the supply of a chain big data analytics market register a better forecast performance for the Stock Markets? – A comparative software analysis

Diana Mendes, Nuno Ferreira and Vivaldo Mendes

Abstract: The dimension of the finance industry that has been most influenced by technological advances in the last decade it is the speed and frequency with which financial transactions are decided and executed. Data analytics, business intelligence, machine learning, algorithmic trading are the leading tech trends of this decade and the most active and innovative segments of the information technology market. In consequence, we can assist in a quick development process to explore new methodologies and to extract and analyse the rich information that the big data market sets contain. For example, in 2019, the biggest U.S. enterprises tend to migrate their data to hybrid cloud solutions helping agents to centralise management of big-data assets distributed between private and public clouds. In this paper, we use the G7 indexes stock markets prices for periods between 10 and 50 years of daily historical data (from Yahoo Finance). The purpose of the analysis it is twofold: first - several forecasting methods are applied, and we search for the minimum forecasting error,

second – we are interested in the time duration and the software velocity attained in the forecasting process. For the first purpose, we use supervised deep learning methods, namely Recurrent Neural Networks and Long Short-Term Memory (LSTM) architectures. We have found that the LSTM configuration works the best out of all the combinations we have tried (for our dataset). LSTMs are very powerful in sequence prediction problems because they can store past information, which is crucial in predicting its future price. The forecast error is measured by the Mean Absolute Percent Error and by the Mean Square Error. We use Python software, where Keras, TensorFlow and Pandas are the packages with the main role. The second important point that we analyse in this article it is the execution time. We use different software, namely R, Julia and Python, and we aim to measure the trade-off between the algorithm complexity and the speed of execution. A comparison of the below results with ARIMA models forecast accuracy and execution time it is also considered.

Photovoltaic Power Forecasting Using Back-Propagation Artificial Neural Network

Hamza Couscous, Abderrahman Benchekroun, Khaled Almaksour, Arnaud Davigny and Dhaker Abbes

Abstract: Known for being a reliable alternative, microgrids have been widely deployed recently in power distribution field in order to guarantee a constant power supply especially in isolated zones. Moreover, microgrids have increasingly known the penetration of renewable energy as environmentally friendly energy sources. However, the intermittency of these sources oblige specialists to think about tools allowing determining their potential, over a predetermined time interval, in order to ensure energy security. For this reason, renewable energy forecasting is crucial. Thus, in this paper, a feed forward back-propagation neural network is used to forecast next 24 hours photovoltaic (PV) power of one of the catholic university buildings "îlot RIZOMM''. The accuracy of the model built is evaluated with some performance metrics. Thereafter, prediction results of PV power are compared to those provided by SteadySat, an industrial solution developed by the company SteadySun. It is shown that the prediction Mean Absolute Errors (MAEs) of the model are of 3.05% in a clear sky day, 4.95% in a cloudy day and 5.98% in a partly cloudy day.

Likelihood Estimation for Hunter Syndrome using ZIP Model and Simulated Data

Behrouz Ehsani-Moghaddam

Abstract: Hunter syndrome is a rare disease, caused by a deficiency in the activity of the lysosomal enzyme, alpha-L-iduronidase. Expression of the disease is inconstant with symptoms emerging at different ages. Diagnosis and management of the disease is challenging and often is delayed until individuals are from 2 to 4 years old. The primary objective of this study was creating a large size dataset using a simulation-based approach. Such dataset can be used for further studies of the disease and for free exchange of information without any need for ethical approval. Additionally, the ability to produce and share the dataset can shorten the idea-to-insight time from years to hours and it can also reduce research expenses. Determining the practicality of the zero-inflated Poisson (ZIP) model in evaluating patients for a rare disease by defining the precision metrics of the predictive model was the second objective. We created a dataset using Monte Carlo simulation technique containing 18 variables including 16 symptoms, 1 variable as patient age and 1 variable as disease status. A ZIP regression model was created to estimate the likelihood of having Hunter syndrome in individual patients using the simulated data. Out of 17 predictors, 14 variables were selected and remained in the final model. The result of ZIP model validation by data splitting revealed that the overall accuracy, sensitivity, specificity, positive predictive value and negative predictive value were 1.0, 0.983, 1.0, 0.892 and 1.0, respectively. The factor analysis detected 8 factors and 11 symptoms that accounted for nearly 60% of the total variance in the dataset. These findings suggest that these 11 symptoms should be considered as the most important features when evaluating new cases for diagnosis and the simulated dataset can be used successfully for future studies of the disease such as for predictive modeling.

Double Seasonal Holt-Winters to forecast electricity consumption in a hot-dip galvanizing process

J. Carlos García-Díaz and Oscar Trull

Abstract: The hot-dip galvanizing process is a very common process in the automotive industry and enables steel bands to be protected against corrosion by coating with zinc. The key process in the galvanizing is the zinc bath, and especially the heating inductors, which have a high electricity consumption. The recent price rises in Spain have led industries to take more care about the electricity they consume, as well as future demand in order to manage their energy costs efficiently. Time series forecasting is a powerful tool for forecasting future demand. This paper focuses on the modelling and exploitation of a multiple seasonality Holt–Winters model to perform these actions. An application of this method to a real example utilising data from an industrial factory in Spain is explained and discussed.

Numerical estimation of GARCH models through a constrained Kalman filter

Abdeljalil Settar, Nadia Idrissi and Mohammed Badaoui

Abstract: In this work, a research framework oriented towards the parameters estimation method of the GARCH model was proposed, based on the constrained Kalman filter and on the simultaneous perturbation stochastic approximation (SPSA) as an optimization process. Kalman fillter with constraints on the conditional variance is used in order to take into account the information eventually known a priori on volatility which nonnegativity represents a particular case. Consequently, the constraints imposed on the parameters ensuring the nonnegativity of the volatility are relaxed and restricted only to these of stationarity and the existence of moments. Simulations of this algorithm applied to the GARCH(p,q) with $p \leq 2$ and $q \leq 2$, are presented and the eciency of the proposed method revealed.

Using Time-Series and Forecasting to Manage Type 2 Diabetes Conditions (GH-Method: Math-Physical Medicine)

Gerald Hsu

Abstract: Introduction: This paper describes the author's application of Time-Series Analysis and Forecasting to manage Type 2 Diabetes (T2D) conditions. Method: The author utilizes the GH-Method: math-physical medicine to manage metabolic disorder diseases especially diabetes. Initially, he observed various disease phenomena. Therefore, he recorded big volume of related data, derived necessary and applicable mathematical equations, utilized suitable computational tools, including time-series analysis, spatial analysis, frequency domain analysis, and artificial intelligence. As a result, he combined them with medical domain knowledge in order to forecast the forthcoming outcomes to interpret the new findings or discoveries regarding human health. In this paper, he disregards the theoretical discussion of time-series analysis in order to focus on application and certain results from his diabetes research by using time-series analysis and forecasting method. Results: Here are some of the results from time-series analysis and forecasting: (1) Weight: He developed a weight prediction model based on food portion, exercise, and certain metabolism respects and achieved 99.8% linear accuracy with a correlation coefficient (R) of 90% to compare with actual weight. Weight contributes ~85% of FPG formation. (2) Fasting plasma glucose (FPG) in early morning: Using time-series analysis, he obtained R=70% between weight and FPG. (3) Postprandial plasma glucose (PPG) at two hours after fist-bite of meal: Using time-series analysis, he obtained R= +45% between carbs/sugar intake and PPG and R= -59% between post-meal walking and PPG. Combined carbs/sugar and walking contributes ~80% of

PPG formation. He achieved 100% linear accuracy and R=85% between predicted and actual PPG. (4) Hemoglobin A1C or HbA1C (A1C): The medical community uses A1C as the measuring yardstick to determine the severity of the patients' diabetes conditions. There are no consistent conversion ratios available between glucose and A1C values. Therefore, the author applied statistics tools (including time-series analysis, spatial analysis, and frequency-domain analysis) and engineering approximation modeling to build up an effective A1C forecasting model. In comparison of this mathematically forecasted A1C results and lab tested A1C results (quarterly data due to insurance constraints), he achieved a linear accuracy of 96% and R=54%. Conclusion: By using time-series analysis method, this clinical case study of more than four years, encompassing 1,488 days and ~500,000 big data, has demonstrated its powerful forecasting capability on weight, glucose, and diabetes control.

Inflation Rate Forecasting: Extreme Learning Machine as a Model Combination Method

Jeronymo Marcondes Pinto and Emerson Fernandes Marçal

Abstract: Inflation rate forecasting is one most discussed topics on timeseries analysis due to its importance on macroeconomic policy. The majority of these papers findings point out that forecasting combination methods usually outperform individual models. In this sense, we evaluate a novel method to combine forecasts based on Extreme Learning Machine Method (Huang et al., 2004), which is becoming very popular but,to the best of our knowledge, has not been used to this purpose. We test Inflation Rate forecasting for four Latin American countries, for one, two,three, ten, eleven and twelve steps ahead. The models to be combined are automatically estimated by R forecast package, as SARIMA, Exponential Smoothing, ARFIMA, Spline Regression, and Artificial Neural Networks. Another goal of our paper is to test our model against classical combination methods such Granger Bates, Linear Regression and Average Mean of models as benchmarks, but also test it against basic forms of new models in the literature, like Diebold and Shin (2018),Garcia et al. (2017) and Wang et al. (2018). Therefore, our paper also contributes to the discussion of forecast combination by comparing versions of some methods that have not been tested against each other. Our results indicate that none of these methods have an indisputable superiority against the others, however the Extreme Machine Learning Method proved to be the most efficient of all, with the smaller Mean Absolute Error and Mean Squared Error for its predictions.

Dynamic behavior in the fractional scope of agricultural commodities price series vis-a-vis ethanol prices

Claudio Inacio and Sergio A. David

Abstract: Having in mind that the share of ethanol as biofuel has increased in recent decades, concern about the impacts of its prices on agricultural commodity prices has gained relevance. These concerns are due to the fact that the energy and environmental benefits derived from the use of biofuels can occur at the expense of the impact of agricultural commodity prices. This is because most of the raw materials currently used to produce biofuels, such as corn in the US, sugarcane in Brazil and oilseeds in Europe, are also important commodities globally. When it comes to ethanol, it could not only influence the price level of these agricultural commodities but may also affect the volatility of these prices. Price volatility reflects the volatility of current and expected future values of production, consumption, and inventory demand. With the recognition that there are other measures of volatility associated with consumption, production or stocks, in this study, the focus is on price series and the main objective is to use numerical modelling and numerical simulation tools in the fractional scope in order to analyze the relationship between the behavior of ethanol price dynamics and the price dynamics of some agricultural commodities, such as corn, sugar and soybean. It is hoped that the results may contribute to the clarification of price relationships between ethanol and such commodities, thus allowing agents in these markets to be clearer in making decisions that involve hedging, risk exposure and investment incentives.

Patent Analysis as a Tool for Revealing Promising Trends of Technological Development

Vladimir Avdzeyko, Vladimir Karnyshev and Evgenia Pascal

Abstract: The paper proposed to use the method of patent analysis for revealing promising trends of technological development. In the authors' view such approach may be useful for short-term forecasting of new products. This method is considered to be perspective and accurate due to the fact that patent information goes ahead of industrial implementation of technological solutions. The given approach is demonstrated by using the International Patent Classification (IPC) and analysis of relevant US patents based on the example of the IPC main group H02M3/00 "Conversion of DC power input into DC power output". The results are based on time series of US patents issued in from 1976 to 2017 and covers the develop-ment trends of DC to DC power converters corresponding to IPC subgroups H02M3/02 – H02M3/42.

Time series analysis of rainfall from climate models under the future warming scenarios over the western Himalayan region

Sudip Kundu and Charu Singh

Abstract: Due to widely reported global warming, events like melting glaciers, rising sea level and changing weather patterns are being observed across the globe. The monsoon season rainfall in the western Himalayan region become more unpredictable as there is a considerable variation between the duration of monsoon and the amount of rainfall in the different places of that region. In this context, the present study deals with the time series analysis of rainfall under various future warming scenarios namely RCP 4.5 and RCP 8.5 (2076-2100) with respect to the reference historical period (1976-2000) during the principal monsoon season (i.e.1st June to the 30th September) over the western Himalayan region. To serve this purpose, we have considered the daily monsoon season rainfall data based on the archives of five coupled climate models which participated in Climate Model Inter-comparison Project Phase 5 (CMIP5). Considering huge inter-model variations amongst the CMIP5 models, here results are also reported based on the multi-model mean (MME). Characteristics of the rainfall pattern in term of time series during the historical time period are compared with that of the two future warming scenarios viz; RCP 4.5 and RCP 8.5 by means of statistical techniques. There is a possibility of an increase in the mean monsoon season rainfall at daily scale under RCP 4.5 and RCP 8.5 over that region compared to the historical MME (HMME). Both of the western Himalayan Indian states Uttarakhand (UK) and Himachal Pradesh (HP) would experience more enhancement in the daily intensity of rainfall under the warmest future warming scenario (i.e. RCP 8.5). Based on the present analysis, intense rainfall events are anticipated over this mountainous region which could act as an initiator for various meteorological hazards over the western Himalayan region in future.

On the evaluation of similarity for time series

Silvia María Ojeda, Juan Carlos Bellassai Gauto and Marcos A. Landi

Abstract: ON THE EVALUATION OF SIMILARITY FOR TIME SERIES Silvia María Ojeda Juan Carlos Bellassai Gauto Marcos Alejandro Landi Abstract. The search and detection of similarities is a central problem in the analysis and processing of time series databases. The issue is relevant, for ex- ample, in problems of classification of time series and in situations in which a predictive process must be evaluated, or when it is necessary to compare two or more prediction methods. Many of the works oriented to the evaluation of simi-larity in time series have focused on the notion of dynamic distortion, with good results in the quantification of similarity, but with a high computational cost. As a result, the interest in the development of new similarity

indexes and the im- provement of existing similarity measures remains in force; even more consid- ering the remarkable increase and availability of time series databases and the urgency that applications demand daily. The expectation about the new pro- posals is that they are able to quantify quickly and not only effectively the simi- larity between time series, in response to different application problems. There- fore, an interesting alternative is to investigate about simple mathematical for- mulation measures, which have proven useful for measuring the similarity in two-dimensional scenarios and assess their adaptation to measure similarity be- tween time series. One of the proposals to measure similarity between two- dimensional scenarios is the SSIM similarity index, defined to quantify simi- larity between digital images. The development was presented by Wang et al. in 2004 and has shown excellent results to evaluate the similarity between two digital images. SSIM has the advantage over other proposals, its simple mathe- matical formulation. In effect, this index is calculated from the product of three factors: the luminance, the contrast and the correlation between the images to be compared. These factors represent, respectively, simple relations between the means, the contrast and the correlation between the images. In this work, we adapted the SSIM index for images to the problem of evaluating the similarity in time series, obtaining a temporal similarity index called SSIMT. Keywords: Time Series, Classification, Clustering.

On-The-Fly Dynamic Ensembles for Time Series Forecasting

Ahmed Elshami, Aliaa Youssef and Mohamed Fakhr

Abstract: This paper proposes two approaches for dynamic creation of prediction ensem-bles. Instead of using fixed regions in the space, we propose the “on-the-fly” idea which is employed in 2 approaches. In the first approach, global models are trained, and a dynamic validation set is created on-the-fly for each test vector based on random forest (RF) defined neighborhood. The dynamic ensemble weights are obtained from the models performance on this validation set. In the second approach, the test vector RF-neighborhood is divided randomly in half, where the first part is used to train the models and the other is used to validate and calculate the dynamic weights. 5 prediction models are used and the proposed approaches are tested on the monthly time series data from M3 competition with 1045 different time series, and with multi-step horizons ranging from 1 to 18. The proposed approaches show very competitive results with other machine learning and statistical forecasting techniques while significantly better results than the state of the art dynamic ensemble methods.

Assessing Wavelet Analysis for Precipitation Forecasts Using Artificial Neural Networks in Mediterranean Coast

Javier Estévez, Xiaodong Liu, Juan A. Bellido-Jiménez and Amanda P. García-Marín

Abstract: Precipitation is one of the most important variables needed in different hydrological models: infiltration, soil loss, droughts, overland flow production, floods, etc. To predict its behavior is complex due to it is highly intermittent over time. Because of the adequate time-frequency representation of wavelet techniques, they are being widely applied to different hydrological resources applications. In this work, wavelet analysis has been applied in order to forecast monthly precipitation data in Mediterranean Coast (Málaga, Southern Spain) using Artificial Neural Networks (ANN) models. Several mother wavelets have been evaluated at different decomposition levels for rainfall predictions using a standard multilayer perceptron architecture. The results obtained indicate that the Daubechies wavelet transforms of order 5 (db5) used at level 3 are the most appropriate for this case study, deriving the more effective performance of all the models assessed.

A robust Hodrick-Prescott filter for smoothing high-frequency time series

Ilaria Lucrezia Amerise and Agostino Tarsitano

Abstract: This paper introduces a new robust method for detecting and replacing extreme spikes in high-frequency time series data and provides guidelines for researchers who wish to use this procedure in empirical applications. The method consists primarily of a nonparametric procedure that pursues a balance between fidelity to the observed data and smoothness. The main tool to obtain the new smoother is a re-formulation of the Hodrick-Prescott smoother procedure in terms of absolute values, which can be very effective in constructing reliable reference time series affected by outliers from a variety of sources. Additionally, the proposed filter makes only a single pass through the time series. Outliers are detected by examining the difference between original and smoothed values. A value is declared to be an outlier if the corresponding discrepancy exceeds a fixed threshold. Properties of the new method are investigated through the accuracy of simultaneous prediction limits. Experiments with SARMAX models show that the new method can be validly used as a data preparation tool to ensure that time series modeling is supported by clean data, particularly in a complex context such as high-frequency data.

Big-Learn 2.5: Using Lucidworks and SolrJ to Improve Online Search in Big Data Environment

Karim Aoulad Abdelouarit, Boubker Sbihi and Noura Aknin

Abstract: In this article, we present an implementation of the Big-Learn system in its version 2.5. It is a Big Data solution for online search, used by a learner in a context of distance learning or information searching. The main objective of this solution is to process massive data and evaluate the reliability and quality of the information returned by the online search system, in order to improve the learning process and thus allow a reliable consumption of data. The version 2.5 of the Big-Learn tool is based on the combination of two systems: Lucidworks and SolrJ. Thus, the massive data generated by the Big Data layer will be processed at the Spark technique of the Lucidworks tool, in addition to their indexing by the Lucene engine and finally their exploitation by the SolrJ tool from the Solr interface.

Traffic demand and longer term forecasting from real-time observations

Alexandros Sopasakis

Abstract: We propose an end to end system comprised of real-time image processing in combination with neural networks to describe upcoming traffic demand in order to forecast short as well as longer term traffic evolution and congestion from traffic camera images. The neural networks use both current and historic traffic information in the form of traffic camera images from a number of roads in an actual traffic network. We process and analyze these images in order to estimate vehicle density at different time points into the future at each of the camera locations on this network. We collect and record minute to minute traffic images over a period of 2 months from cameras overlooking city traffic. We apply a number of filters on each such image collected in order to estimate and record vehicle densities. We use this dataset to train three different neural networks in order to test which of them has a better capability to learn patterns in traffic densities from the camera data. Specifically we design and train a long short-term memory, a gated recurrent unit and a stacked autoencoder network. We train these networks on data from a single camera location and use each of the three networks to predict traffic density by processing images arriving in real time at all the other camera locations in this traffic network. The results reveal that such a system could be helpful to provide information about traffic demand and formation of congestion for hours into the future. The traffic data used is collected from the traffic network of Goteborg in Sweden.

The Impact of Signed Jump Variation in Forecasting Realized Variance

Ioannis Papantonis, Elias Tzavalis and Leonidas Rompolis

Abstract: It is a well-known stylized property that statistical distributions of financial time-series exhibit significant variation over time, as well as prevalent deviations from normality. The vast majority of earlier financial econometric studies has employed GARCH and SV model-classes in order to capture the empirically-observed time-variation (clustering) in volatility. This has also offered a flexible parametric framework to address other empirically-observed properties of the data, such as leverage and/or feedback effects -among others- which are partly responsible for inducing asymmetries in the distributions of returns. Another major source of asymmetry and tail-heaviness has been attributed to the existence of latent jump dynamics in both returns and variance. Recent advances to the econometrics literature of high-frequency data have popularized approaches that permit identification of the underlying jump dynamics. These realized jump-driven estimators carry significant information not only for the returns, but also for volatility. This paper extends Hansen's Realized-GARCH framework to allow for the impact of jumps (and other realized estimators) in the conditional variance dynamics of returns. First, we build intraday evenly-spaced observations from tick-by-tick data that we use as building blocks for our realized estimators. We decompose realized variation into its continuous (quadratic) and discontinuous (jump) components and we utilize the realized semi-variances to construct realized signed jump variation measures. Next, we employ a collection of (not necessarily nested) asymmetric GARCH-type processes that we augment to incorporate our realized estimators. At the same time, we explore the parametric specification of the model that can better explain the joint dynamics of returns and realized variance. This methodology allows us to test several interesting hypotheses; among these we examine if conditional variance responds differently to the upside and downside semi-variances, if there's a material jump contribution to variance, and also if there is an asymmetric behavior behind the impact of signed jump variation. Last, our main objective is to assess the performance of this joint approach in forecasting realized variances at multiple horizons/frequencies. Our paper also implicitly tries to bridge the gap between the GARCH models of conditional variance and the more-recent HAR/HEAVY models of realized variance, and it has really important implications for risk-management, option-pricing and macro-economic uncertainty forecasting.

Copper price variation forecasts using genetic algorithms

Raul Carrasco

Abstract: This work suggests that it is possible to apply genetic algorithms to predict the variation in copper prices, in order to improve the degree of certainty by incorporating the inverse of the percentage of sign prediction (PSP).

On the stress of testing credit default

Viani Djeundje Biatat and Jonathan Crook

Abstract: Analyzing and predicting credit default continues to attract a lot of interest, both in academia and among practitioners. In this work, we look at uncertainty and stress using account-levels auto-regressive models within a survival analysis framework. Within this framework, we (i) identify a number of components contributing to the distribution of credit loss, and then (ii) quantify the impact of each component on standard stress metrics.

An Automated Lane Change Strategy for Autonomous Vehicles Based on QoS Forecasting

Jamal Raiyn

Abstract: A large number of accidents are caused by human lane change behaviors. Delays in drivers' reaction times and errors in judgement are the main causes of the vast majority of accidents. Autonomous vehicles (AVs) can decrease the occurrence of the accidents through platooning, as well as increase road capacity. Platooning is defined as the gathering together of AVs, which interact with each other to coordinate the safe joining, exiting, and changing lanes. It manages the traffic on urban roads by reducing distance between AVs as much as possible. This paper discusses the impact of the lane changing of AVs in platooning congestion. A strategy is introduced for successful lane changing using V2I communications and intelligent speed assistance (ISA). ISA forecasts the travel speed in real-time and detects when traffic congestion is due to an accident or repeated failure to change lanes. The performance of the V2I communication is measured by QoS parameters, such as delay and interference. Latency and inter-carrier interference can have an impact on cooperative communication based V2I; therefore, an automated lane change strategy is proposed to overcome the limitations of cooperative Communication.

Stochastic Analysis and Modeling of Local Temperature Fluctuations

Faeze Minakhani and Mohammad Dehghan Niri

Abstract: We examine the Markov property of temperature fluctuation in four meteorological stations. Using stochastic analysis, we characterize the complexity of the temperature fluctuation by means of the single and multi-variable Fokker-Planck equation and find the Kramers-Moyal coefficients. The Langevin equation enables us to reconstruct temperature time series with similar statistical properties compared with the observed fluctuations in real temperature sequence.

Selective Attention in Exchange Rate Forecasting

Svatopluk Kapounek and Zuzana Kučerová

Abstract: The paper investigates exchange rate forecasting performance of main-stream macroeconomic fundamentals, uncertainty, attention and news. Although currency markets react to many different information, we hy-pothesize that smaller sizes models offer better fitting and forecasting re-sults because economic agents accept only a limited amount of infor-mation. Our results confirm selective attention hypothesis of currency market participants and explain behavioural biases in investment decision-making process. We employ Dynamic Model Averaging approach to reduce model-selection uncertainty and to identify time-varying probability to include regressors into our models. We analyse 100.5 thousand news articles about the 6 most traded Forex currency pairs in the period from 1979 to 2016 and confirm growing impact of the news about foreign trade and monetary policy issues on the Euro/U.S. Dollar exchange rate after the financial crisis. The other analysed currencies react mostly to output differentials and stock market returns induced by portfolio rebalancing.

Can the Machine Learn Capital Structure?

Jack Strauss

Abstract: We use machine-learning tools to analyze and predict capital structure using data from approximately 180,000 firms. LASSO, Random Forecast, Gradient Boosting, Neural net methods and a general linear model are used to select variables over a 1970-2000 in-sample and training period. We then construct out-of-sample mean squared forecast errors from 2001-2016. We show that Random Forest methods significantly, substantially and consistently outperform the benchmark linear model as well as a benchmark Lasso model. We then examine variable importance, and show that more than a dozen variables reliably predict corporate leverage over the past fifteen years. Our work highlights the importance of incorporating nonlinearities for predicting capital structure. And yes the machine can learn and predict capital structure.

Evaluating Auto-encoder and Principal Component Analysis for Feature Engineering in Electronic Health Records

Shruti Kaushik, Abhinav Choudhury, Nataraj Dasgupta, Sayee Natarajan, Larry Pickett and Varun Dutt

Abstract: Feature engineering is an important mechanism where we transform and represent the high-dimensional data into a lower-dimensional space. These representations can then be used to efficiently train machine-learning models. Auto-encoders are widely used in research for unsupervised feature learning. However, the application of auto-encoders for electronic health records (EHRs) containing features with binary values (binary-valued features) has been less studied. The primary objective of this research was to compare an auto-encoder with principal component analysis (PCA), a popular feature engineering technique, for feature selection in different (US and Indian) EHR datasets containing binary-valued features. The US dataset contained thousands of binary-valued features, and the Indian dataset contained nineteen binary-valued features. Results revealed that feature selection by the auto-encoder followed by different classification algorithms gave the highest accuracy on both the datasets compared to feature selection by PCA. We highlight the implications of using auto-encoders for learning features in EHR datasets.

Improving the management of public transport through modeling and forecasting passenger occupancy rate

Tulio Vieira, Paulo Almeida, Magali Meireles and Renato Ribeiro

Abstract: The improvement of public transport in large cities is a fundamental factor for the quality of life. Poor transportation leads to an increased of greenhouse gases generation, hinders access to essential services and emphasizes the difference between social classes. One possible way to improve traffic in large cities is to encourage people to use public transport. By improving the quality of public transport systems and reducing tariffs, more people can use it as a means of getting around in urban centers. This article performs an analysis between different strategies (Neural Recurrent Network using LSTM and GRU, Convolutional Neural Network and ARIMA models) to model the variation of the occupancy rate (PTO) of the metropolitan buses in order to improve the planning and management of public transport. Results show that ARIMA models present better results to PTO forecasting and to describe the behavior of time series. This kind of approach can be used,in practice, to adjust the number of buses and population demand, for a given period.

Applications of Statistical and Machine Learning Methods for Predicting Time-Series Performance of Network Devices

Naveksha Sood, Usha Rani, Srikanth Swaminathan, George Abraham, Dileep A.D. and Varun Dutt

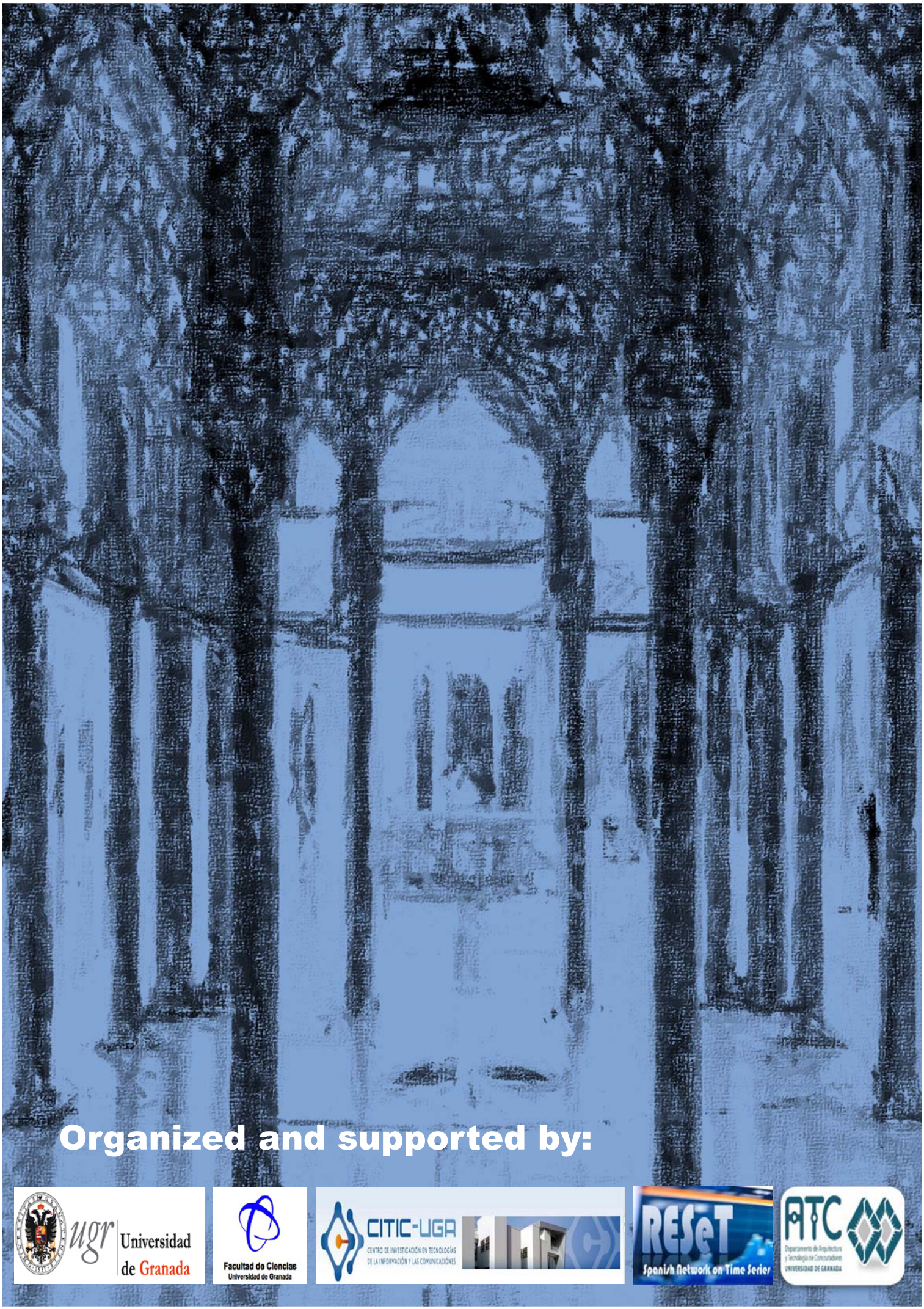
Abstract: Prediction of performance of network devices, which control the flow of data, is of utmost importance to be able to manage the network efficiently. In this paper, we used a statistical auto-regressive integrated

moving average (ARIMA) model and a machine learning (ML) multi-layer perceptron (MLP) model to forecast the performance of a network device 15-minutes ahead in time. Forecasting was done with one-feature (univariate) and multiple features (multivariate). Also, we evaluated the effects of highly correlated features on our predictions in MLP and ARIMA models. Results revealed that the ARIMA model performed better compared to the MLP model for univariate data; however, the MLP model performed better compared to the ARIMA model with exogenous variables for multivariate data. In addition, keeping highly correlated features in the models improved model predictions on multivariate data. We highlight the real-life implications of using statistical and ML models for time-series forecasting of network data.

Applying Diebold-Mariano test for performance evaluation between individual and hybrid time series models for modeling bivariate time series data and forecasting the unemployment rate in the USA

Moamen Abbas Mousa Al-Sharifi and Firas Ahmmed Mohammed Al-Mohana

Abstract: Unemployment rate forecasting has become a particularly promising domain of comparative studies in recent years because it's a major issue facing the economic forecasting process. Since the time series data are rarely pure linear or nonlinear obviously, sometimes contain both components jointly. Therefore, this study introduces a hybrid model combines between two commonly used models, namely the Linear Autoregressive Moving Average with exogenous variable (ARMAX) model and nonlinear Generalized Autoregressive Conditional Heteroskedasticity with exogenous variable (GARCHX) model whose conditional variance follows a General error distribution (GED). i.e. build a hybrid (ARMAX-GARCHX-GED) model employed in modeling bivariate time-series data of the unemployment rate and exchange rate. Usually, the forecasting performance evaluation based on the common classical forecast accuracy criteria such as Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percent Error (MAPE), have some specific limitations in application to choosing the optimal forecasting model. Therefore, in this paper, we employed a modern evaluation criterion based on the methodology advocated by Diebold-Mariano (DM) known as (DM-test) as a new criterion for evaluation based on statistical hypothesis tests. This (DM-test) has been applied in this study to distinguish the significant differences in forecasting accuracy between hybrid (ARMAX-GARCHX-GED) and individual ARMAX models. From the case study results, and according to DM-test it is observed that the differences between the forecasting performances of models are significant and the hybrid model (ARMAX-GARCHX-GED) more efficient from the individual competitive ARMAX model for the unemployment rate forecasting.



Organized and supported by:

Universidad de Granada

Facultad de Ciencias
Universidad de Granada

CITIC-UGR
CENTRO DE INVESTIGACIÓN EN TECNOLOGÍAS DE LA INFORMACIÓN Y LAS COMUNICACIONES

RESeT
Spanish Network on Time Series

ATC
Departamento de Arquitectura y Tecnología de Computadores
UNIVERSIDAD DE GRANADA