

ICTRS 2016

In Memory of Blagovest Shishkov

Fifth International Conference on Telecommunications and Remote Sensing



Proceedings

Milan, Italy • 10-11 October 2016

Organized by:



Co-Organized by:

ARSIT



Technical
Co-Sponsorship by:

BULATSA

Under the
Auspices of:



Cooperating Organizations:

AMAKOTIA



BULGARIAN ACADEMY
OF SCIENCES

ICTRS 2016

**Proceedings of the
Fifth International Conference on
Telecommunications and Remote Sensing**

Milan, Italy
10-11 October 2016

Organized by
**IICREST - Interdisciplinary Institute for Collaboration and
Research on Enterprise Systems and Technology**

Co-Organized by
ARSIT - Association for Radio Systems and Intelligent Telecommunications
AUTH - Aristotle University of Thessaloniki

In Collaboration with
AMAKOTA Ltd.
**IMI - Institute of Mathematics and Informatics of
Bulgarian Academy of Sciences**

Under the Auspices of
URSI - International Union of Radio Science

Technical Co-sponsorship by
BULATSA - Bulgarian Air Traffic Services Authority

Copyright © 2016 SCITEPRESS - Science and Technology Publications
All rights reserved

Edited by:
Andon Lazarov, Dimitris Mitrakos, Boris Shishkov

Graphics Production by Canka Petrova

Printed in Bulgaria

ISBN: 978-989-758-200-4

Depósito Legal: 414815/16

<http://www.ictrs.org>

secretariat@iicrest.org

BRIEF CONTENTS

Chairs and Program Committee	V
Best Papers Selection	VIII
Foreword	IX
Contents	XI

CHAIRS AND PROGRAM COMMITTEE

PROGRAM CO-CHAIRS

Andon Lazarov, Burgas Free University, Bulgaria / K.N. Toosi Univ. of Technology, Iran

Dimitris Mitrakos, Aristotle University of Thessaloniki, Greece

Naoki Shinohara, Kyoto University, Japan

GENERAL CHAIR

Boris Shishkov, Bulgarian Academy of Sciences / IICREST, Bulgaria

PROGRAM COMMITTEE

Oleksiy Agapitov, Taras Shevchenko Kyiv
National University, Ukraine

Catherine Algani, CNAM, France

Makoto Ando, Tokyo Institute of
Technology, Japan

Mauro Assis, Brazilian Committee of
URSI, Brazil

Akalam Baki, Independent University,
Bangladesh

Vera Behar, Bulgarian Academy of
Sciences, Bulgaria

Maurice Bellanger, CNAM, France

Jun Cheng, Doshisha University, Japan

Lyubka Doukovska, Bulgarian Academy
of Sciences, Bulgaria

Yoshiharu Fuse, USEF, Japan

Hristo Kabakchiev, Sofia University “St.
Kl. Ohridski”, Bulgaria

Mesut Kartal, Istanbul Technical
University, Turkey

Kazuya Kobayashi, Chuo University,
Japan

CHAIRS AND PROGRAM COMETEE (CONT.)

Karol Kozak, Dresden University of Technology, Germany

Mohamed Latrach, ESEO, France

Frank Little, Texas A & M University, USA

Marco Luise, University of Pisa, Italy

Francois Lefeuvre, LPC2E / CNRS, France

Olga Maltseva, Southern Federal University, Russia

Andrea Massa, University of Trento, Italy

Galia Marinova, Technical University of Sofia, Bulgaria

Wolfgang Mathis, Leibniz Universität Hannover, Germany

Lyudmila Mihaylova, Lancaster University, UK

Tomohiko Mitani, Kyoto University, Japan

Tadao Nagatsuma, Osaka University, Japan

Shoichi Narahashi, NTT DOCOMO, Inc., Japan

Marin Nenchev, Technical University of Sofia, Bulgaria

Martin O'Droma, University of Limerick, Ireland

Takashi Ohira, Toyohashi University of Technology, Japan

Jacques Palicot, SUPELEC / IETR, France

Hermann Rohling, Hamburg University of Technology, Germany

Sana Salous, Durham University, UK

Hamit Serbest, Cukurova University, Turkey

Boris Shishkov, Bulgarian Academy of Sciences / IICREST, Bulgaria

Alexander Shmelev, Russian Academy of Sciences, Russia

CHAIRS AND PROGRAM COMETEE (CONT.)

Marten van Sinderen, University of
Twente, The Netherlands

Angela Slavova, Bulgarian Academy of
Sciences, Bulgaria

Jun-Ichi Takada, Tokyo Institute of
Technology, Japan

Hiroyuki Tsuji, NIICT, Japan

Ing Widya, University of Twente, The
Netherlands

Satoshi Yagitani, Kanazawa University,
Japan

Tsuneki Yamasaki, Nihon University,
Japan

BEST PAPERS SELECTION

The authors of several selected ICTRS'16 best papers will be invited to submit revised and extended versions of their papers to a special issue of the international journal

ANNALS OF TELECOMMUNICATIONS

FOREWORD

Bringing together telecommunications (through areas, such as radio communications and signal processing) and remote sensing (through areas, such as pattern recognition and context awareness), and enriching this with Data-Science-related topics, is characterizing ICTRS 2016 – the *Fifth International Conference on Telecommunications and Remote Sensing*.

This book contains the proceedings of ICTRS 2016, held in Milan, Italy, on 10-11 October 2016. The proceedings consists of 12 high-quality research and experience papers that have not been published previously. Those papers have undergone a detailed peer-review process and were selected based on rigorous quality standards.

Since 2012, we have enjoyed four successful ICTRS editions, namely: *Sofia 2012*, *Noordwijkerhout 2013*, *Luxembourg 2014*, and *Rhodes 2015*, inspired by the heritage of *Prof. D.Sc. Blagovest Shishkov* who established and led this conference until his decease in 2015.

ICTRS 2016 was organized and sponsored by the *Interdisciplinary Institute for Collaboration and Research on Enterprise Systems and Technology* (IICREST), being co-organized by the *Association for Radio Systems and Intelligent Telecommunications* (ARSIT) and *Aristotle University of Thessaloniki* (AUTH), and technically co-sponsored by the *Bulgarian Air Traffic Services Authority* (BULATSA). Cooperating organizations were *AMAKOTA Ltd.* and the *Institute of Mathematics and Informatics* of the Bulgarian Academy of Sciences. Further, the conference is performed under the auspices of the *International Union of Radio Science* (URSI), covering topics relevant to the work in a number of URSI Commissions, namely commissions A, B, C, D, E, F, G, H, and K.

Publisher of the current proceedings is *SCITEPRESS*. Besides printed proceedings we also deliver an electronic version – all presented papers will be made available on the *SCITEPRESS Digital Library* by December 2016. Furthermore, the authors of several selected papers presented at ICTRS 2016 will be invited to submit revised and extended versions of their papers to a special issue of the international journal *Annals of Telecommunications*.

As it can be seen from the papers, ICTRS 2016 is addressing a large number of topics relevant to the areas of Telecommunications and Remote Sensing: such as specific ionospheric propagation effects, novel electromagnetic wave technologies, satellite image fusion and topographic mapping, remote sensing based on unmanned aerial vehicles and signal processing for next generation networks. This is complemented by an eHealth related paper touching upon medical diagnosis protocols based on specific sensing-at-a-distance approaches. The special session on Data Science and Applications, addressing in turn topics, such as: NoSQL databases, data mining, and data matching is further enriching ICTRS'16. This all allows for an inspiring interdisciplinary technical program that we hope will be interesting and stimulating for all participants.

FOREWORD (CONT.)

ICTRS 2016 received 22 paper submissions from which 12 papers were selected (including several invited papers) for publication in the current proceedings. Further, 10 of those papers were selected for ICTRS oral presentation and 2 papers were selected for the Special Session on Data Science and Applications, leading to an ICTRS'16 acceptance ratio of 45%. Finally, the ICTRS'16 authors are from: Bulgaria, China, France, Greece, Germany, Iran, Italy, Mexico, Morocco, Poland, Russia, and USA (listed alphabetically); that makes a total of 12 countries.

The high quality of the ICTRS 2016 program is enhanced by a panel discussion on *Area and Border Security*. That discussion and also other (informal) discussions stimulating community building and facilitating possible R&D project acquisition initiatives, will definitely contribute to maintaining the event's high quality.

Organizing this interesting and successful conference required the dedicated efforts of many people. Firstly, we must thank the authors, whose research and development achievements are recorded here. Next, the program committee members each deserve credit for the diligent and rigorous peer-reviewing. Further, we would like to mention the excellent organization provided by the IICREST team, and especially Canka Petrova and Aglika Dikova plus the Co-Organizers, ARSIT and AUTH (supported by the logistics partner, AMAKOTA Ltd.) – the team did all the necessary work for delivering a stimulating and productive event, and we have to acknowledge also the inspiring support of our colleagues from Thessaloniki and our local partners in Milan, especially Angelo Bellanova. Last but not least, we are grateful to SCITEPRESS for their willingness to publish the current proceedings and we would like to especially mention Vitor Pedrosa with whom we collaborated excellently on the proceedings preparation.

We wish you all an inspiring conference and an enjoyable stay in the beautiful city of Milan. We look forward to seeing you next year in Delft, The Netherlands, for the Sixth International Conference on Telecommunications and Remote Sensing (ICTRS 2017), details of which will be made available on <http://www.ictrs.org>.

Boris Shishkov

Bulgarian Academy of Sciences / IICREST, Bulgaria

Andon Lazarov

Burgas Free University, Bulgaria / K.N. Toosi Univ. of Technology, Iran

Dimitris Mitrakos

Aristotle University of Thessaloniki, Greece

CONTENTS

IONOSPHERIC PROPAGATION EFFECTS

- Efficiency of the Equivalent Slab Thickness of the Ionosphere to Set Radio Wave Propagation Conditions
Olga Maltseva and Natalia Mozhaeva 5
- Assessing the Impact of TEC Fluctuations on ALOS-PALSAR Images
Luca Spogli, Elvira Musicò, Claudio Cesaroni, John Peter Merryman Boncori, Giorgiana De Franceschi and Roberto Seu 15

SATELLITE AND AERIAL SENSING

- New Wavelet Based Spatiotemporal Fusion Method
Amal Ibnelhobyb, Ayoub Mouak, Amina Radgui, Ahmed Tamtaoui, Ahmed Er-Raji, Driss El Hadani, Mohamed Merdas and Faouzi Mohamed Smiej 25
- Multi-Satellite Interferometric SAR System
Andon Lazarov, Chavdar Minchev and Dimitar Minchev 33
- Bathymetric Mapping of Shallow Rivers with UAV Hyperspectral Data
Valeria Gentile, Marek Mróz, Marie Spitoni, Jérôme Lejot, Hervé Piégay and Luca Demarchi 43

ELECTROMAGNETIC WAVE TECHNOLOGIES

- Physically Inspired Digital Predistortion for Dual Input Doherty Power Amplifiers and Automatic in-Situ Identification
Giovanni Donati, Christian Musolff, Giuseppe Gottardo, Daniel Popp and Georg Fischer 53
- Is it Possible to Detect the Stealth Flying Objects by the Millimetre Wave Radiometer?
Jinghui Qiu, Shengchang Lan, Hao Liu, Xinyu Yin, Alexander Denisov, Xiao Qing, Zhao Man and Francesco Soldovieri 61

SIGNAL PROCESSING FOR NEXT GENERATION NETWORKS

- Capacity Analysis of Radio Frequency Interconnect for Manycore Processor Chips
Eren Unlu, Christophe Moy, Yves Louet and Jacques Palicot 71

CONTENTS (CONT.)

Hidden Markov Model Traffic Characterisation in Wireless Networks <i>Evangelos Spyrou and Dimitris Mitrakos</i>	78
--	----

MEDICAL SENSORS APPLICATIONS

Brain Rehabilitation in Clinical Trials Setup by Eye-Tracking <i>Bartosz Kunka, Robert Kosikowski, Jessica Barlinn and Karol Kozak</i>	89
---	----

SPECIAL SESSION ON DATA SCIENCE AND APPLICATIONS

Adopting NoSQL Databases using a Quality Attribute Framework and Risks Analysis <i>Hilda Mackin, Gonzalo Perez and Charles C. Tappert</i>	97
--	----

A Comparative of Spanish Encoding Functions Effectiveness on Record Linkage <i>María del Pilar Angeles and Noemi Bailón-Miguel</i>	105
---	-----

AUTHOR INDEX	117
--------------	-----

PAPERS

**IONOSPHERIC
PROPAGATION
EFFECTS**

Efficiency of the Equivalent Slab Thickness of the Ionosphere to Set Radio Wave Propagation Conditions

Olga Maltseva and Natalia Mozhaeva

Institute for Physics, Southern Federal University, Stachki, 194, Rostov-on-Don, Russia
mal@ip.rsu.ru

Keywords: GPS, Total electron content TEC, Ionospheric models, Equivalent slab thickness.

Abstract: Now the total electron content TEC is a key parameter characterizing conditions of the ionosphere. TEC is widely used for an estimation of positioning accuracy, definition of index of ionospheric storm activity. Data of TEC is very important for systems of satellite communication and navigation. The advantages of the TEC measurement are the systems of a large number of receivers, the possibility of continuous global monitoring of the ionosphere, the availability of data on the Internet. For many systems (HF-communication, HFDF, HFGINT) it is necessary to know the maximum density of the ionosphere NmF2 or, that is equivalent, a critical frequency foF2. To obtain NmF2, it is necessary to know the proportionality coefficient $\tau = \text{TEC}/\text{NmF2}$, which is the equivalent slab thickness of the ionosphere. Before occurrence of navigational satellites, no special attention was given to this parameter and there were many inaccuracies in the papers devoted to τ . The possibility of the global monitoring of NmF2 with use of TEC, measured by navigational satellites, makes to give the more close attention to its study. In the present paper, data of more than 50 ionospheric stations and several global maps of TEC are used to investigate behavior of a median $\tau(\text{med})$ of the observational equivalent slab thickness $\tau(\text{obs})$. Comparison of $\tau(\text{med})$ with the equivalent slab thickness $\tau(\text{IRI})$ of the IRI model, $\tau(\text{NGM})$ of the Neustrelitz global model and others has shown essential differences between these values. Approaches for developing a global model of $\tau(\text{med})$ are offered. The most amazing are following results: (1) for a large amount of stations, the use of observational TEC and $\tau(\text{IRI})$ worsens values of foF2 compared to the initial IRI model, (2) there are no fundamental quantitative differences in the use of $\tau(\text{med})$ for all regions of the world, (3) the IRI model and maps of TEC (in the absence of GPS receivers) for the most northern Nord station (Greenland) showed surprisingly good agreement with the experimental values of foF2.

1 INTRODUCTION

The critical frequency of the ionosphere foF2 was the main parameter determining the state of the ionosphere and radio wave propagation conditions in the last century. It is connected with the maximum density NmF2 of the ionosphere by the relationship $\text{NmF2} = 1.24_{10} \times \text{foF2}^2$ and with the maximum usable frequency MUF through the propagation factor M(D): $\text{MUF} = \text{M(D)} \times \text{foF2}$. This frequency is measured by ground ionosondes. In the 21st century, the total electron content TEC becomes the main parameter. TEC is measured by means of navigation satellites in units of $\text{TECU} = 1_{10}16 \text{ e/m}^2$. The advantages of the TEC measurement are regional systems of a large number of receivers, the possibility of continuous global monitoring of the ionosphere, the availability of data on the Internet. Naturally,

there was a proposal to use TEC to determine the maximum density of the ionosphere NmF2 and foF2. To do this, we need to know the proportionality coefficient $\tau = \text{TEC}/\text{NmF2}$, which is the equivalent slab thickness of the ionosphere. Despite the fact that the measurement of NmF2 was begun with the invention of ionosondes (Breit and Tuve, 1926) and TEC measurements were begun with the first artificial satellite launch (Aitchison and Weekes, 1959), and the opportunity to use τ to obtain knowledge of the ionosphere parameters and the atmosphere was immediately appreciated, interest in this parameter was increased only with the advent of satellite navigation systems GPS, GLONASS, allowing measurement of TEC. Despite the huge amount of publications, unified picture of the behavior of the experimental equivalent slab thickness $\tau(\text{obs})$ does not exist because different data

is used to determine TEC. Distinctions concern both technical characteristics, and heights of satellites. Traditionally the International Reference Ionosphere model IRI (Bilitza, 2001; Bilitza et al., 2014) and its equivalent slab thickness τ (IRI) are used to calculate foF2. However, on one hand, τ (IRI) is not an empirical model in a statistical sense, but, on the other hand, almost nobody compared values of τ (obs) and τ (IRI) and suggested the use of experimental median τ (med) to calculate foF2. In (Maltseva et al. 2011), it was proposed to use τ (med) together with the experimental values TEC(obs) and it was shown that this use allows to obtain foF2 closer to the experimental value foF2(obs) than by means of τ (IRI), and to fill gaps of the experimental data of foF2.

The aim of this article is to estimate: (1) the features of the behavior of the median equivalent slab thickness, (2) the effectiveness of its use together with the total electron content to obtain the critical frequencies in comparison with the equivalent slab thickness of current models of the ionosphere, (3) the opportunity of developing a global model of τ (med).

2 EXPERIMENTAL DATA AND CALCULATED VALUES

The feature of the current stage of research is the availability of online databases of experimental data that allows us to obtain the results on a global scale (Maltseva, 2015). Data of foF2 of 56 ionosondes of vertical sounding were used together with 5 global maps JPL, CODE, UPC, ESA, IGS (Hernandez-Pajares et al., 2009). The disadvantage of ionospheric data is their sketchy character however there are stations for which long-term measurements are available. Data of foF2 was taken from the databases SPIDR (<http://spidr.ngdc.noaa.gov/spidr/index.jsp>) and DIDBase (<http://ulcar.uml.edu/DIDB/>). TEC values were calculated from IONEX files (<ftp://cddis.gsfc.nasa.gov/pub/gps/products/ionex/>). As models, we used IRI2001, IRI2012 (Bilitza, 2001; Bilitza et al., 2014), which have the upper limit of 2,000 km, and IRI-Plas (Gulyaeva, 2003; Gulyaeva and Bilitza, 2012) located on the site <http://ftp.izmiran.ru/pub/izmiran/SPIM/> and allowed determining N(h)-profile up to heights of navigation satellites h(GPS) by taking into account a plasmaspheric part of the profile. The magnitude of the equivalent slab thickness τ is calculated in accordance with the relationship $\tau = \text{TEC}/\text{NmF2}$ for the model and experimental parameters TEC and

NmF2. In this paper, values of τ (IRI) of the IRI model and the median τ (med) of observational τ (obs) are calculated and compared. To evaluate the effectiveness of their use jointly with observational TEC(obs) we have introduced corresponding efficiency coefficients. These coefficients are determined by using the deviations of calculated foF2 from the observational foF2(obs). The value of $|\Delta\text{foF2}(\text{IRI})| = |\text{foF2}(\text{obs}) - \text{foF2}(\text{IRI})|$ is the difference between the instantaneous values of the IRI model and observational foF2(obs). Monthly averages were calculated. This difference is in the denominators of the efficiency coefficients. $|\Delta\text{foF2}(\tau(\text{IRI}))| = |\text{foF2}(\text{obs}) - \text{foF2}(\tau(\text{IRI}))|$ is the difference between the values calculated using τ (IRI) and TEC(obs) and the observational foF2(obs). Deviation $|\Delta\text{foF2}(\tau(\text{med}))| = |\text{foF2}(\text{obs}) - \text{foF2}(\tau(\text{med}))|$ is the difference between values calculated using τ (med) and TEC(obs) and foF2(obs). Coefficient $K_{\tau\text{IRI}} = |\text{foF2}(\Delta\text{IRI})| / |\Delta\text{foF2}(\tau(\text{IRI}))|$ is the coefficient of efficiency of joint use of τ (IRI) and TEC(obs). Coefficient $K_{\text{eff}} = |\Delta\text{foF2}(\text{IRI})| / |\Delta\text{foF2}(\tau(\text{med}))|$ is the coefficient of efficiency of joint use of τ (med) and TEC(obs). These coefficients are given together with the line $K = 1$ to visually evaluate the effectiveness of using TEC(obs): if the coefficient is 1, this means that the use of TEC(obs) leads to the same results of foF2, as the initial IRI model without TEC(obs). In this case $|\Delta\text{foF2}(\text{IRI})| = |\Delta\text{foF2}(\tau(\text{IRI}))|$. If the coefficient > 1 , then the use of TEC(obs) leads to results better than the initial model. If the coefficient is higher than 1, then the use of TEC(obs) worsens the results of the model. These values determine how strongly the deviation of the initial model differs from the deviation of values calculated using τ (IRI) and τ (med) together with TEC(obs).

3 FEATURES OF τ BEHAVIOR

In the literature, there are certain disagreements in the describing of such features of τ as: 1) diurnal variation, 2) seasonal variation, 3) latitudinal dependence 4) dependence on solar activity, 5) behavior during disturbed conditions. The most important disagreement is the absence or weakness of the latitudinal dependence of τ , noted in many publications (Kouris et al., 2008; Sardar et al., 2012; Vryonides et al., 2012). Real situation is given in Figure 1.

If the latitudinal dependence of τ (med) was absent, the value of τ (med) for one station could be used in obtaining foF2 with the observational TEC in

the whole region. Of particular importance is the study of the behavior of $\tau(\text{obs})$ during disturbances because it is different from the behavior of $\tau(\text{med})$. In paper (Maltseva et al. 2011), a hyperbolic approximation of τ was introduced as $\tau(\text{hyp}) = b_0 + B1/Nm$ to build a regression relation in which $Nm = foF2 * foF2$ (foF2 in MHz). Such function is calculated for each map. An example is shown in Figure 2. Approximation of $\tau(\text{hyp})$ was built for a more accurate determination of τ during the disturbances.

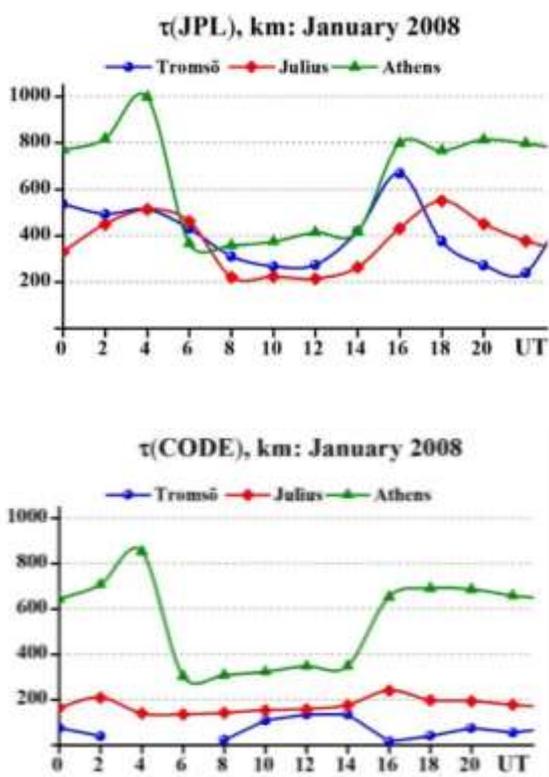


Figure 1. Example of the diurnal variation of $\tau(\text{obs})$ for the different latitude and two maps JPL and CODE.

The most important is the difference between $\tau(\text{IRI})$ and $\tau(\text{med})$. Numerous examples are given in papers (Maltseva and Mozhaeva, 2014, 2015) for stations in all regions of the world with long-term data. Example for the etalon station Juliusruh is shown in Figure 3.

Large difference is seen not only in magnitude but also in the diurnal variation. Namely this difference determines the difference of the critical frequency foF2(rec) reconstructed from the observational values TEC(obs) using $\tau(\text{med})$ and $\tau(\text{IRI})$.

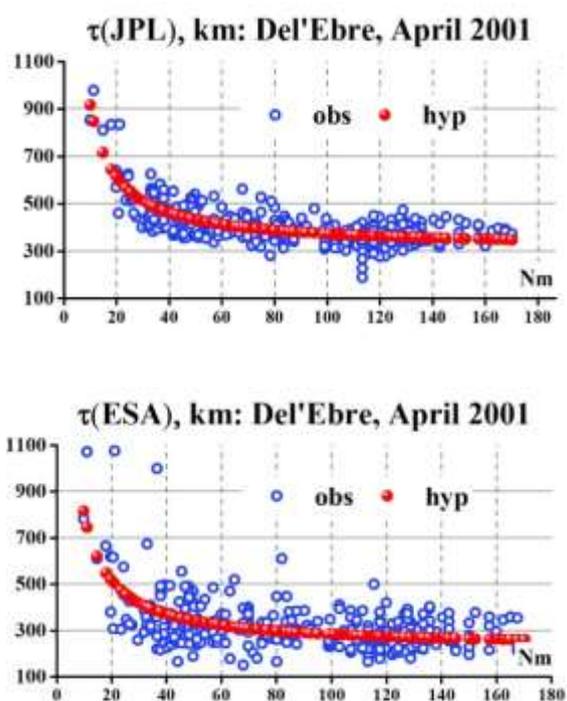


Figure 2: An example of a hyperbolic dependence of τ for the disturbed month.

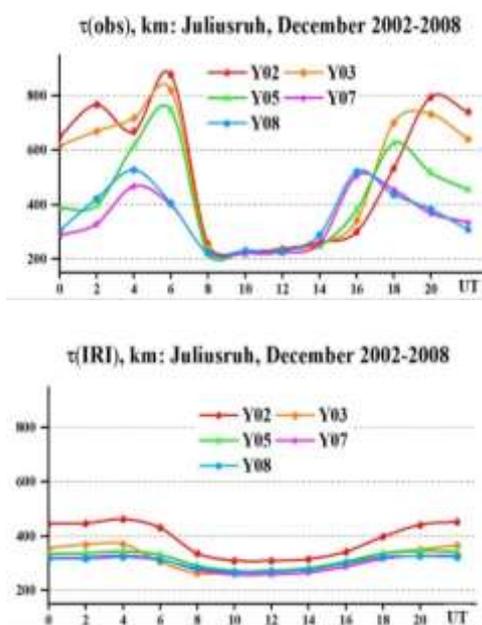


Figure 3: Differences of the model and the experimental equivalent slab thicknesses in the example for the mid-latitude station Juliusruh of European region and JPL map.

4 EFFICIENCY OF JOINT USING τ (IRI) AND τ (MED) AND TEC(OBS)

Figure 4 shows the deviation of the calculated values of foF2 from foF2(obs) and efficiency coefficients for stations with long-term observations: high- latitude

Thule, mid-latitude Juliusruh and equatorial Kwajalein stations of the northern hemisphere and the high-latitude Mawson station of the southern hemisphere. Black dots show the results for the IRI model, triangles present the results of joint using τ (IRI) and TEC(obs), circles give the results of joint using τ (med) and TEC(obs).

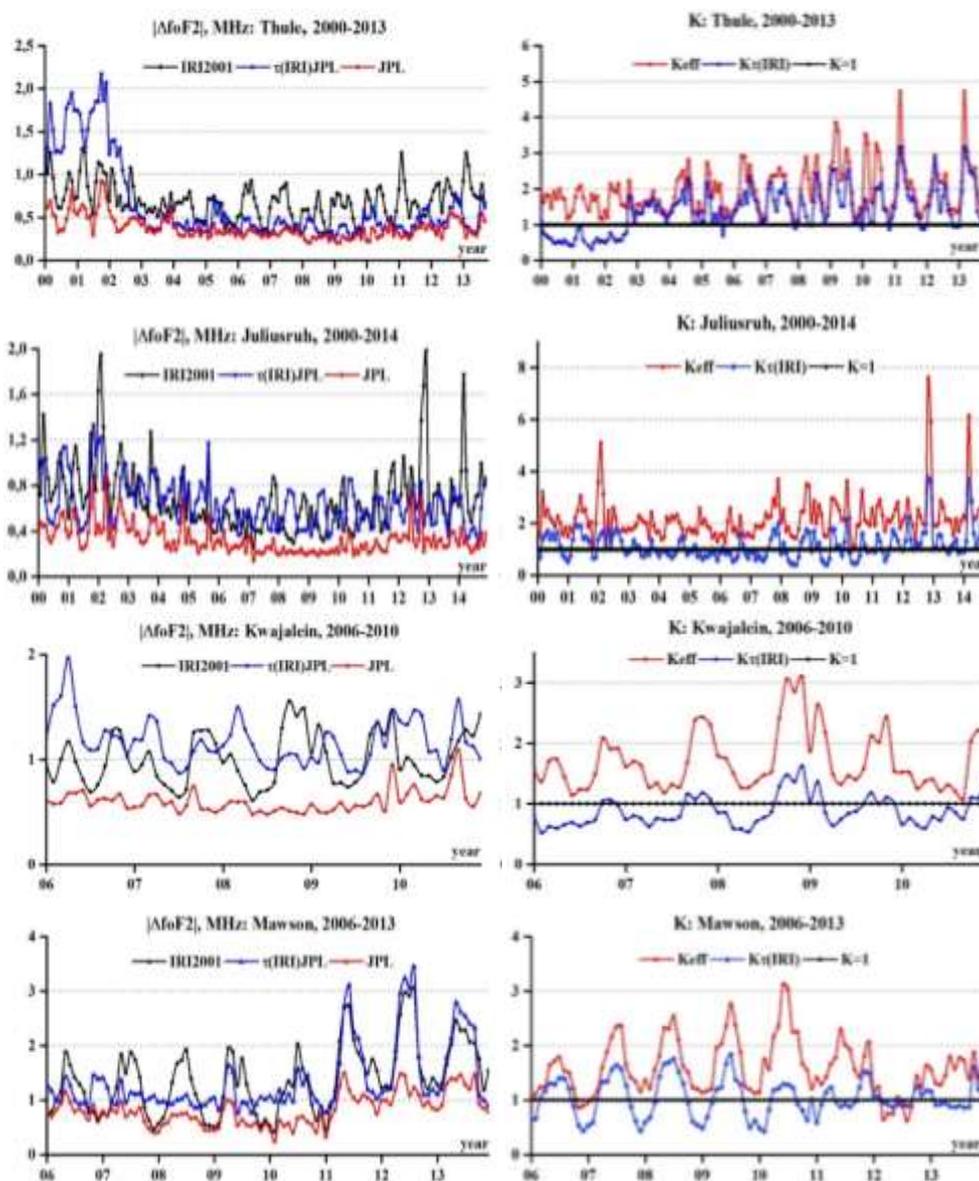


Figure 4: Deviations $|\Delta foF2|$ and efficiency coefficients for the IRI model and two cases of using τ (IRI) and τ (med) together with the TEC(obs).

It can be seen that the joint use of τ (IRI) and TEC(obs) ($K < 1$) can significantly worsen the

calculation of foF2 compared with the model ($K < 1$). By using τ (med) and TEC(obs) deviations $|\Delta foF2|$ do

not exceed 1.0 MHz in most cases even in problematic areas, such as high and equatorial latitudes, and the coefficients are always exceed 1. In the previous 2-3 years, a large amount of ionospheric data has become available. This allows us to check and compare the results simultaneously on many stations on a global scale, in particular, to obtain these results in the points in which they have not been obtained and the IRI model was not tested itself. We selected April 2014 and March 2015, because they included geomagnetic disturbances.

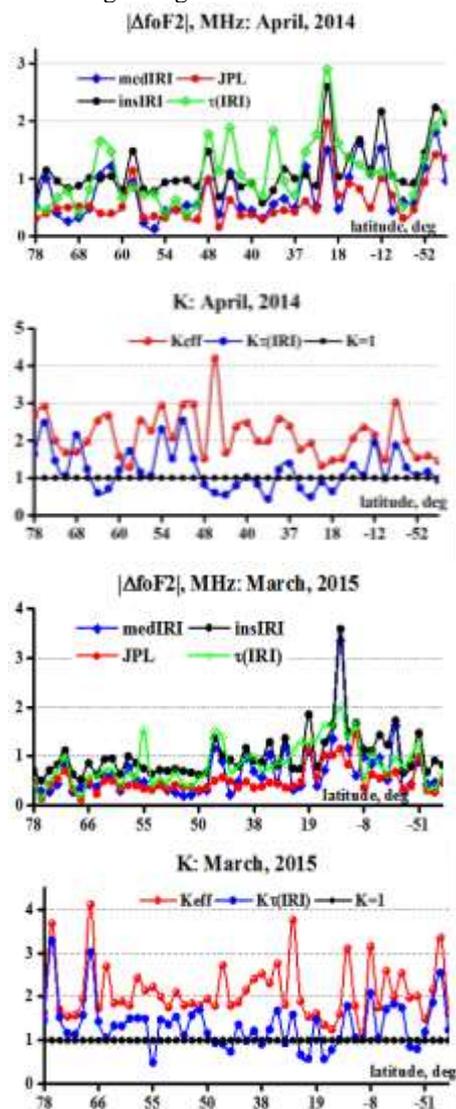


Figure 5: Correspondence between the model and the observational values of foF2 and efficiency coefficients on a global scale in April 2014 and March 2015.

Figure 5 illustrates the effectiveness of using $\tau(\text{med})$ globally by example of data for April 2014 (minimum Dst = -81 nT) and March 2015 (minimum

Dst = -223 nT). The samples shown on the x axis are with a variable step, as stations are located not uniformly.

It is evident that in the northern hemisphere the amount of stations is larger than in the southern one. In all cases, the use of $\tau(\text{med})$ and TEC(obs) improves matching calculated foF2(rec) with foF2(obs) compared with two options: using the initial IRI model and joint using $\tau(\text{IRI})$ and TEC(obs). Joint using $\tau(\text{IRI})$ and TEC(obs) may provide poor results compared to the initial model. The best results were obtained for mid-latitudes. For high latitudes they were not worse, but in the equatorial latitudes problems for the model are seen, although the joint use of $\tau(\text{med})$ and TEC(obs) mitigates these problems. Figure 6 gives the results for the Nord station (Greenland) which are of particular interest because it is the most northern station. Unfortunately data of this station in the DIDbase were downloaded recently and in a very limited extent (a few months and not each year).

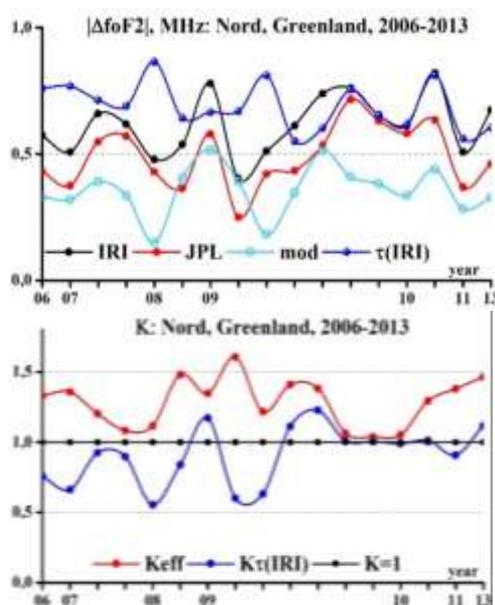


Figure 6: Long-term statistics for the Nord station (Greenland) for available months of years indicated on the x axis.

The results are almost similar to the results for Thule, Tromso and other lower high-latitudinal stations. Since the critical frequency of the Nord station has never been compared with the model the upper-hand plot of Figure 6 presents the curve showing the deviation of the model from the experimental medians of foF2. Results indicate a surprisingly good performance of the IRI model in this region.

5 ABOUT THE CONSTRUCTION OF A GLOBAL MODEL OF $\tau(\text{MED})$

The mention of the opportunity to construct a model of τ is practically absent in the papers, but in recent years some articles were published on the use of TEC and the ionospheric equivalent slab thickness τ to determine NmF2, what confirms the importance of this problem. In (Gerzen et al., 2013), the authors have proposed the use of two Neustrelitz models of TEC and NmF2 (Hoque and Jakowski, 2011; Jakowski et al., 2011) for the calculation of foF2, but without sufficient testing. We use these models under one name NGM (from Neustrelitz Global Model). It is obvious that authors of (Gerzen et al., 2013) have also used the value of $\tau(\text{NGM}) = \text{TEC}(\text{NGM})/\text{NmF2}$, which can serve as an empirical model of τ . However, testing this model in (Maltseva et al., 2013, 2014) showed that in many regions the value of $\tau(\text{NGM})$ is very different from the experimental $\tau(\text{med})$. In (Muslim et al., 2015), a model of the average values of τ was proposed as expansion in Fourier series according to the TEC of the global map CODE and foF2 for 21 stations, however, the assumptions made in constructing the model: (1) linear dependence of all parameters of the TEC, foF2 and τ on solar activity, (2) lack of longitudinal dependence of these parameters at the same local time LT, (3) regularity of τ in quiet and disturbed conditions need confirmation. This allows drawing a conclusion about the need to develop the model of τ . To build a model of $\tau(\text{med})$ on a global scale two approaches are proposed: two-parameter model based on a hyperbolic approximation $\tau(\text{hyp}) = b_0 + b_1/\text{NmF2}$ and the use of the coefficient $K(\tau) = \tau(\text{med})/\tau(\text{IRI})$, since the construction of the model using the values themselves is not possible because of the large variability of values (in particular, the pre-sunrise peak on some latitudes). A hyperbolic dependence and approximation coefficient $K(\tau)$ were calculated for March 2015. The results are shown for two regions 2 ($15^\circ\text{E} < \lambda < 40^\circ\text{E}$) with 8 stations and 4 ($110^\circ\text{E} < \lambda < 170^\circ\text{E}$) with 9 stations and two wider zones (Lat1 and Lat2). Area Lat1 includes stations, located mostly in American continent of northern and southern hemispheres. Area Lat2 includes stations from European, Siberian and South-Eastern regions. Behavior of coefficients b_0 and b_1 is shown in Figure 7 for these regions.

The test results of this model are shown in Table 1

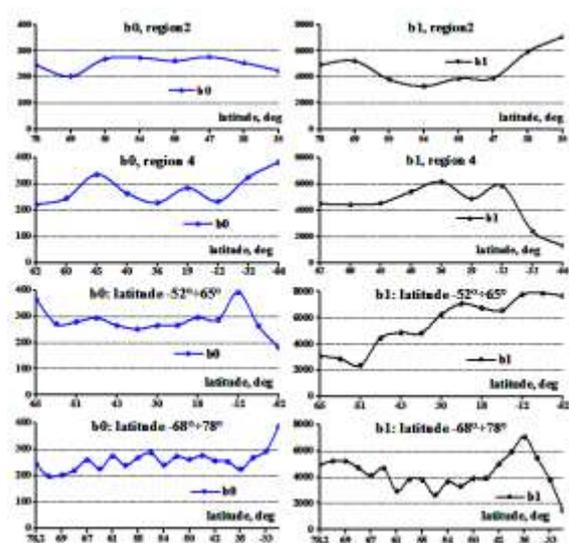


Figure 7: Behavior of hyperbolic approximation coefficients in various regions

Table 1 includes different deviations. Column 1 indicates the station name and the region to which it belongs. The second column shows the coefficients of the hyperbolic approximation of $\tau(\text{obs})$ for the corresponding station. The third column specifies conditions to which the two rows of values belong. The top line (full) indicates average for all days of the month, the bottom line (dist) gives average for disturbed days (from 16 to 21 March). The fourth column shows the results for the initial IRI model, the fifth column presents the absolute difference between the foF2(obs) and the values calculated using $\tau(\text{med})$ and TEC(obs). Column 6 contains the frequency deviation, calculated using the coefficients b_0 and b_1 of the hyperbolic approximation for a given station. The remaining columns give results of using coefficients of areas referred to in the line title. All of these values should be compared with the values for the IRI model in bold. This is a test of the effectiveness of the model. It can be seen that all values are higher in disturbed days and the largest differences concern to the initial IRI model. It is seen that frequencies of one region could be used to calculate the coefficients for other region. This demonstrates a global character of the model of $\tau(\text{med})$.

Another method of constructing a global model of $\tau(\text{med})$ would be to use the coefficients $K(\tau) = \tau(\text{med})/\tau(\text{IRI})$. Definite advantage of this model might be in the fact that its denominator is the value of $\tau(\text{IRI})$, which has a global character, and a small change in $K(\tau)$ in areas with close latitudes.

The distinction is development of a model for each hour. The degree of proximity of τ is better

illustrated in the circular diagram. An example of some diagrams is shown in Figure 8 for region 4 for UT = 0, 6, 12, and 18 on March 2015. The red line shows the value of coefficient $K(\tau)$, green triangles

are the average values, the blue quadrates concern circles with radius $R = 1$.

Table 1: Deviation of frequencies, calculated by the hyperbolic dependence, from the experimental values in March 2015.

1	2	3	4	5	6	7	8	9	10
station	b1, b0		IRI	rec	stat	reg2	reg4	Lat1	Lat2
Juliusruh	3295.5	full	0.73	0.41	0.43	0.68	0.67	1.03	0.57
reg2	273.2	dist	1.44	0.52	0.49	0.67	0.68	1.07	0.64
Athens	5929.3	full	0.91	0.36	0.46	0.56	0.48	0.52	0.58
reg2	253.2	dist	1.31	0.44	0.74	0.52	0.59	0.87	0.51
Grahamstown	3788.7	full	0.80	0.40	0.54	0.77	0.59	0.73	0.73
Lat2	293.2	dist	1.54	0.46	0.62	0.84	0.75	0.82	0.77
Longyearbyen	4947.1	full	0.70	0.43	0.62	0.60	0.58	0.82	0.58
Lat2	244.2	dist	0.69	0.49	0.73	0.69	0.69	1.01	0.63
Thule	692.7	full	0.51	0.14	0.15	0.56	0.42	0.47	0.59
	437.6	dist	0.55	0.10	0.13	0.51	0.46	0.64	0.54
Millstonehill	4864.4	full	0.90	0.50	0.47	0.48	0.46	0.67	0.49
Lat1	265.4	dist	1.38	0.67	0.67	0.65	0.81	0.81	0.80
Bejing	5402.8	full	1.17	0.49	0.61	0.61	0.58	0.70	0.62
reg4	263.9	dist	1.99	0.42	0.64	0.45	0.51	0.84	0.45
Kokubunji	6176.7	full	1.29	0.47	0.65	0.61	0.69	0.85	0.62
reg4	228.4	dist	2.11	0.55	0.66	0.56	0.70	0.96	0.56
Niue Island	4874.7	full	1.85	1.15	1.36	1.35	1.28	1.43	1.29
reg4	285.0	dist	1.67	0.71	1.00	0.73	0.85	1.11	0.67
Cocos Island	5467.3	full	1.43	0.55	0.68	0.86	0.62	0.65	0.82
Lat2	267.8	dist	1.66	0.52	0.77	0.88	0.67	0.80	0.83
Mawson	1466.2	full	0.91	0.27	0.37	1.00	0.85	1.02	0.92
Lat2	386.8	dist	1.12	0.12	0.21	0.80	0.98	0.98	0.81

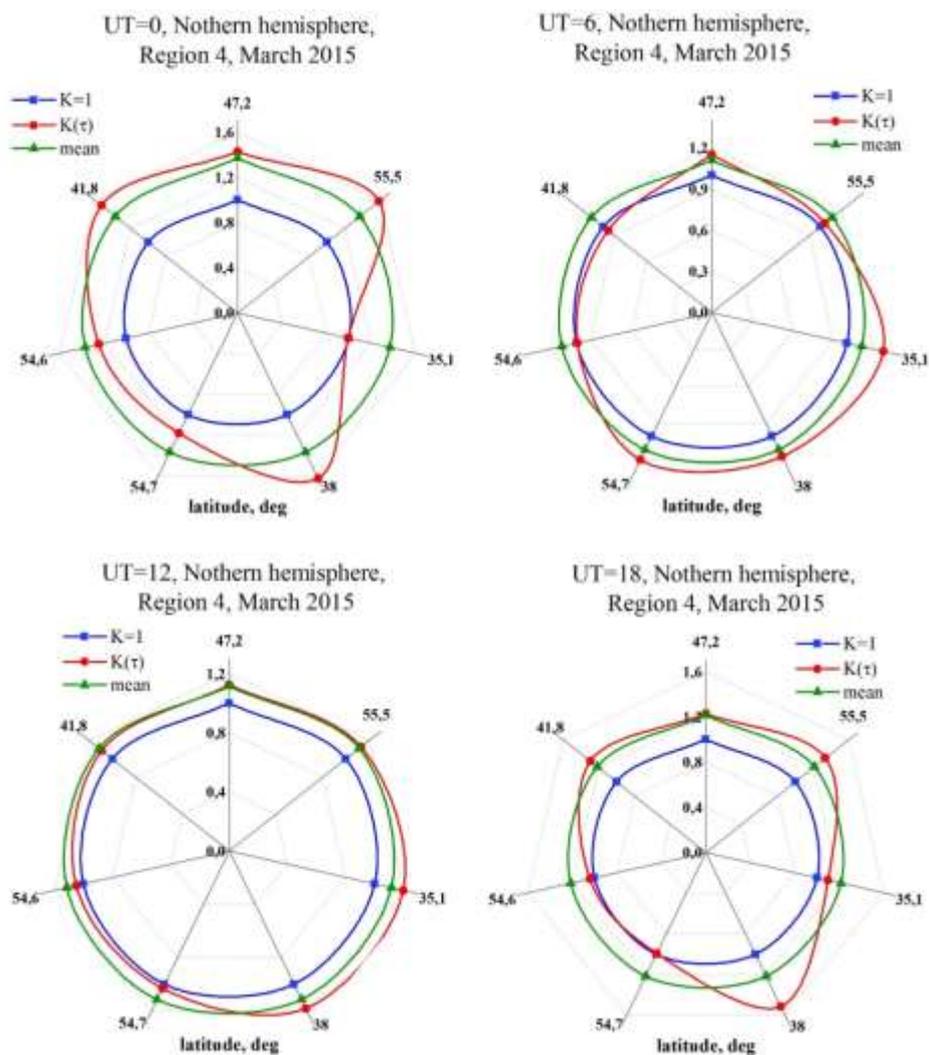


Figure 8: Illustration of charts for the coefficient $K(\tau)$.

The model is the average value $K(\text{mean})$. The algorithm of its use is calculation of the new value $\tau(K\tau) = K(\text{mean}) \times \tau(\text{IRI})$ and the use of this new value together with $\text{TEC}(\text{obs})$ to calculate foF2. To test the efficiency of the algorithm, averages of $K(\text{mean})$ were calculated for 7 stations for region consisted of 8 stations and were used to calculate foF2 for the 8th station. Results are presented in Figure 9 as deviations of the calculated values from observational foF2(obs) for the four stations: Pruhonice, Gorkovskaya, Tunguska and Ramey.

Table 2. Average deviations of calculated foF2 from foF2(obs) using various options of models and τ .

station	IRI-off	IRI-on	$\tau(\text{IRI})$	new	$\tau(\text{med})$
Pruhonice	0.63	0.65	0.54	0.34	0.32
Gorkovsk	0.64	0.60	0.43	0.37	0.36
Tunguska	0.95	0.86	0.72	0.57	0.50
Ramey	0.78	0.85	1.36	0.78	0.58

All plots contain curves of deviations for the initial model without the use of TEC. Since both options (off and on) were used, the curves were indicated by IRI-off and IRI-on. Curves with the icon τ (IRI) show results of joint using of τ (IRI) and TEC(obs).

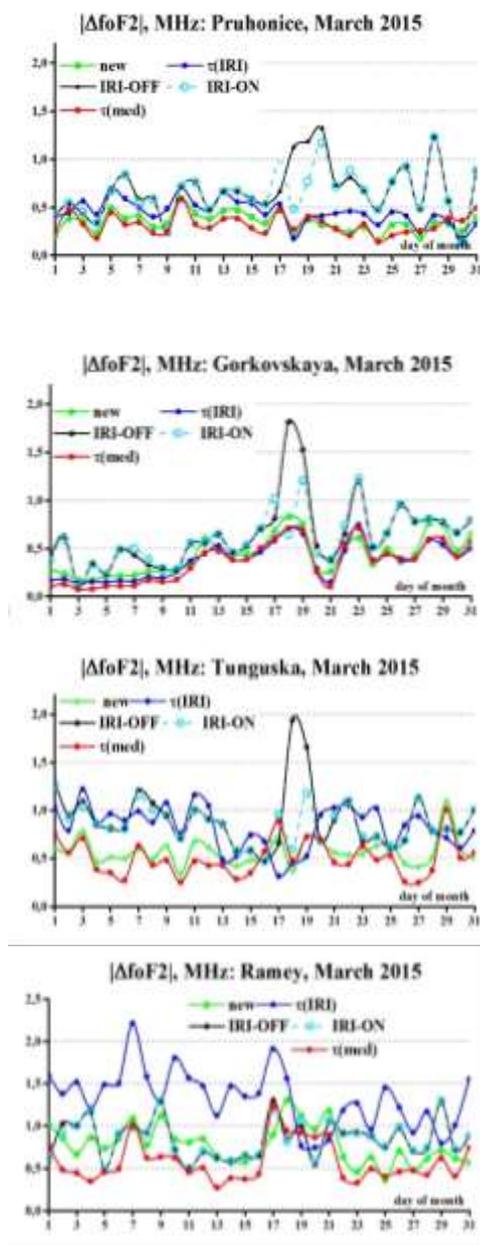


Figure 9: Deviation of calculated foF2 values from foF2(obs) using various options of τ .

The curves marked by “new” show test values for the $K\tau$ model. Asterisks present results for τ (med) and TEC(obs). Most values of $|\Delta\text{foF2}|$ do not exceed 1

MHz. Significant deviations are only visible to the IRI model in disturbed days on March 18-21 for the case τ (IRI) and for the Ramey station. Quantitative characteristics are given in Table. 2.

6 CONCLUSIONS

Data of more than 50 ionospheric stations and several global maps of TEC were used to study behavior of a median τ (med) of the observational equivalent slab thickness τ (obs) and comparisons with existing models τ (IRI), τ (NGM) and others. Essential differences between them, leading to the large deviations of the calculated values of foF2 from the experimental ones are shown. As a quantitative estimation, the effectiveness coefficients of joint using TEC(obs) and τ (med) in comparison with joint using TEC(obs) and τ (IRI) are used. It is shown that the effectiveness coefficients practically always exceed 1 for joint using of TEC(obs) and τ (med). There are several striking results: (1) for a large amount of stations, the use of observational TEC and τ (IRI) worsens values of foF2 compared to the initial IRI model, (2) there are no fundamental quantitative differences in the use of τ (med) for all regions of the world, (3) the IRI model and maps of TEC (in the absence of GPS receivers) for the most northern Nord station (Greenland) showed surprisingly good agreement with the experimental values of foF2. In this sense, results of HF propagation modeling on high-latitude paths based on the IRI model (Blagoveshchensky et al., 2016) seem no longer surprising. Two approaches for developing a global model of τ (med) are offered.

ACKNOWLEDGEMENTS

The authors thank the scientists who provided data of SPIDR and DIDBase, global maps of TEC, operation and modification of the IRI model, Southern Federal University for financial support (grant N 213.01-11/2014-22).

REFERENCES

Aitchison, G.J., Weekes, K., 1959. Some deductions of ionospheric information from the observations of emissions from satellite 1957a2—I: The theory of the analysis, *J. Atm. Terr. Phys.*, 14(3–4), 236–243. doi:10.1016/0021-9169(59)90035-2.

- Bilitza, D., 2001. International Reference Ionosphere, *Radio Sci.*, 36(2), 261-275.
- Bilitza, D., Altadill, D., Zhang, Y., Mertens, C., Truhlik, V., Richards, P., McKinnell, L.-A., Reinisch, B., 2014. The International Reference Ionosphere 2012 – a model of international collaboration, *J. Space Weather Space Clim.*, 4, A07, 12p. DOI: 10.1051/swsc/2014004.
- Blagoveshchensky, D.V., Maltseva, O.A., Anishin, M.M., Rogov, D.D., Sergeeva, M.A., 2016. Modeling of HF propagation at high latitudes on the basis of IRI, *Adv. Space Res.*, 57, 821–834.
- Breit, G., Tuve, M.A., 1926. A test for the existence of the conducting layer, *Phys. Rev.*, 28, 554-575.
- Gerzen, T., Jakowski, N., Wilken, V., Hoque, M.M., 2013. Reconstruction of F2 layer peak electron density based on operational vertical total electron content maps, *Ann. Geophys.*, 31, 1241-1249. doi:10.5194/angeo-31-1241-2013.
- Gulyaeva, T.L., 2003. International standard model of the Earth's ionosphere and plasmasphere, *Astron. and Astrophys. Transaction*, 22, 639-643.
- Gulyaeva, T., Bilitza, D., 2012. Towards ISO Standard Earth Ionosphere and Plasmasphere Model, In: *New Developments in the Standard Model*, (R.J. Larsen ed.). NOVA, Hauppauge, New York, 1–48.
- Hernandez-Pajares, M., Juan, J. M., Orus, R., Garcia-Rigo, A., Feltens J., Komjathy, A., Schaer, S.C., Krankowski, A. 2009. The IGS VTEC maps: a reliable source of ionospheric information since 1998, *J. Geod.*, 83, 263–275.
- Hoque, M.M., Jakowski, N., 2011. A new global empirical NmF2 model for operational use in radio systems. *Radio Sci.*, 46, RS6015, 1-13.
- Jakowski, N., Hoque M.M., Mayer C., 2011. A new global TEC model for estimating transionospheric radio wave propagation errors, *J. Geod.*, 85(12), 965-974.
- Kouris, S.S., Polimeris, K.V., Cander, L.R., Ciraolo L., 2008. Solar and latitude dependence of TEC and SLAB thickness, *J. Atmos. Solar-Terr. Phys.*, 70, 1351-1365.
- Maltseva, O.A., 2015. Usage of the Internet resources for research of the ionosphere and the determination of radio-wave propagation conditions. *Proceedings of the Fourth International Conference on Telecommunications and Remote Sensing, Rhodes, Greece, 17-18 September, 7-17.*
- Maltseva, O.A., Mozhaeva, N.S., 2014. Features of behavior and usage of a total electron content in the Indian region, *International Journal of Engineering and Innovative Technology*, 4(4), October 2014, 1-9. www.ijet.com, ISSN: 2277-3754.
- Maltseva, O.A., Mozhaeva, N.S., 2015. Obtaining Ionospheric Conditions according to Data of Navigation Satellites, *International Journal of Antennas and Propagation*. 1-16. <http://dx.doi.org/10.1155/2015/804791>.
- Maltseva, O.A., Mozhaeva, N.S., Glebova, G.M., 2011. Global maps of TEC and conditions of radio wave propagation in the Mediterranean area, *PIERS Proceedings, Marrakesh, MOROCCO, March 20-23, 1P9_0422, 1-5.*
- Maltseva, O.A., Mozhaeva, N.S., Nikitenko, T.V., 2014. Validation of the Neustrelitz Global Model according to the low latitude *Adv. Space Res.*, 54 () 463–472.
- Maltseva, O.A., Mozhaeva, N.S., Nikitenko T.V., 2015. Comparative analysis of two new empirical models IRI-Plas and NGM (the Neustrelitz Global Model), *Adv. Space Res.*, 55, 2086–2098.
- Maltseva, O., Mozhaeva, N., Vinnik, E., 2013. Validation of two new empirical ionospheric models IRI-Plas and NGM describing conditions of radio wave propagation in space. *Proceedings of Second International Conference on Telecommunications and Remote Sensing, Noordwijkerhout, The Netherlands, 11-12 July, 109-118.*
- Muslim, B., Haralambous, H., Oikonomou, Ch., Anggarani, S., 2015. Evaluation of a global model of ionospheric slab thickness for foF2 estimation during geomagnetic storm. *Ann. Geophys.*, 58(5), A0551. doi:10.4401/ag-6721.
- Sardar, N., Singh, A.K., Nagar, A., Mishra, S.D., Vijay, S.K., 2012. Study of Latitudinal variation of Ionospheric parameters - A Detailed report, *J. Ind. Geophys. Union*, 16(3), 113-133.

Assessing the Impact of TEC Fluctuations on ALOS-PALSAR Images

Luca Spogli^{1,2}, Elvira Musico¹, Claudio Cesaroni¹, John Peter Merryman Boncori¹,
Giorgiana De Franceschi¹ and Roberto Seu³

¹*Istituto Nazionale di Geofisica e Vulcanologia, Via di Vigna Murata 605, Rome, Italy*

²*SpacEarth Technology, Rome, Italy*

³*Electronics and Telecommunications, Sapienza University of Rome, Via Eudossiana 18, 00184 Rome, Italy*

luca.spogli@ingv.it, elvira.musico@ingv.it, claudio.cesaroni@ingv.it, john.merryman@ingv.it,

giorgiana.defranceschi@ingv.it, roberto.seu@uniroma1.it

Keywords: Ionosphere, SAR, InSAR, Total Electron Content, GNSS, Mid-latitude Ionosphere.

Abstract: Trans-ionospheric waves experience delay proportional to the Total Electron Content (TEC), being the number of free electrons present along a satellite-receiver ray path. TEC is a highly variable quantity, influenced by different helio-geophysical parameters, such as solar activity, season, time of the day, etc. Such large variability may lead to TEC spatial and temporal fluctuations over different scales, affecting the quality of the Synthetic Aperture Radar (SAR) signals and, in turn, limiting further developments of interferometric techniques, such as InSAR (Interferometric SAR). In the specific, the need of catching qualitative and quantitative correspondences between TEC fluctuations and InSAR image streaks is a key point to drive the development of future mitigation techniques to improve the quality of the SAR imaging. In this paper calibrated TEC values, derived from the RINEX data provided by the RING (Rete Integrata Nazionale GPS) network of GPS receivers, are analysed to assess the ionosphere conditions during the ALOS (Advanced Land Observing Satellite) – PALSAR (Phased Array type L-band Synthetic Aperture Radar) passages over central Italy.

1 INTRODUCTION

The InSAR is among the most used techniques for geophysical applications, because it is able to provide high-resolution imaging, independently on all tropospheric weather conditions (Bürgmann et al., 2000). SAR signals emitted by ALOS-PALSAR in L-band frequency range cross the bulk of the ionosphere, because they are emitted at about 700km altitude, i.e. in the topside ionosphere. By consequence, such signals experience changes in their phase and group velocity induced by the ionosphere. Changes are proportional to the distribution of free electrons encountered during their travel from SAR satellite to ground and back. The InSAR technique extracts the position and/or the displacement of a surface by calculating the phase difference between two SAR images acquired

in different days, but at the same local time and on the same geometrical area (Ferretti et al., 2007).

Thus, to ensure the performance of the InSAR imaging and to assess the ionospheric impact on it, the status of the ionosphere in both passages must be taken into account, as it is known to be responsible of the appearance of streaks on SAR co-registration images. In the specific, the shift of the co-registration is proportional to the ionospheric TEC gradients. (Chen & Zebker, 2014).

TEC can be obtained by means of GNSS dual frequency receivers located at ground. GNSS slant TEC (STEC) is defined as the total number of free electrons within a cylinder with a cross section of 1 m² and height equal to the signal ray path, from satellite to ground (in the GNSS case). STEC is then a measure of the integrated electron density Ne according to the formula:

$$STEC = \int Ne(s) ds, \quad (1).$$

where s is the ray path of the satellite-receiver link. Being GNSS satellites located at an altitude of about 20200 km, TEC measured with suitable ground receivers provides information also above the ALOS altitudes. However, the contribution to TEC of the ionosphere above 700 km is of few %'s (Kelley, 2009).

TEC is a highly variable quantity influenced by different helio-geophysical parameters such as the solar activity, the season and the time of the day. This must be taken into account when considering the ionospheric conditions of the days to which the images obtained by InSAR techniques refer. Many researches about ionospheric effects on SAR have been carried out, most of which are based on numerical simulations, use of global ionospheric models like WBMOD (Meyer & Agram, 2015) or large-scale interpolated ionospheric maps that do not have the spatial resolution needed for InSAR (Hanssen, 2001). Currently, only few studies investigated the relationship between ionospheric variability and SAR/InSAR imaging by means of independent measurements (like GNSS) and, in general, a final word has not been told about the ionospheric influence on real SAR images (Zhu et al 2016). For this reason, high-resolution determination of TEC and of its spatial-temporal fluctuations is needed. In this work, we show the preliminary results obtained by comparing ALOS-PALSAR images of central Italy and TEC measurements obtained by means of the Rete Integrata Nazionale GPS (RING, <http://ring.gm.ingv.it/>) network of GNSS receivers, managed by the Istituto Nazionale di Geofisica e Vulcanologia (Italy).

The paper is organized as follows: section 2 presents the ALOS-PALSAR and GNSS data and how they have been treated to obtain the results, discussed in section 3. Then, conclusions are provided in section 4.

2 DATA AND METHODS

2.1 ALOS-PALSAR

InSAR is an imaging technique that evaluates the pixel-to-pixel phase difference between two SAR images, acquired over the same area, to produce an interferogram. In this work, two SAR images (a “master” and a “slave”) are acquired with a single receiver in two different epochs with almost the

same incidence angle and on the same area, i.e. in the so called “repeat-pass interferometry mode”. Since the orbital positions of the two SAR passages are slightly different, it is necessary to coregistrate the master and slave images, with the accuracy of sub-pixel, to calculate the phase difference of two corresponding pixels and, then, obtaining high quality InSAR images. In the present work, we concentrate on the results of the co-registration, without looking at the final InSAR product (interferometric phase), with the aim to minimize the tropospheric error affecting phase measurements (Hassen, 2001) and to focus on the ionospheric contribution only.

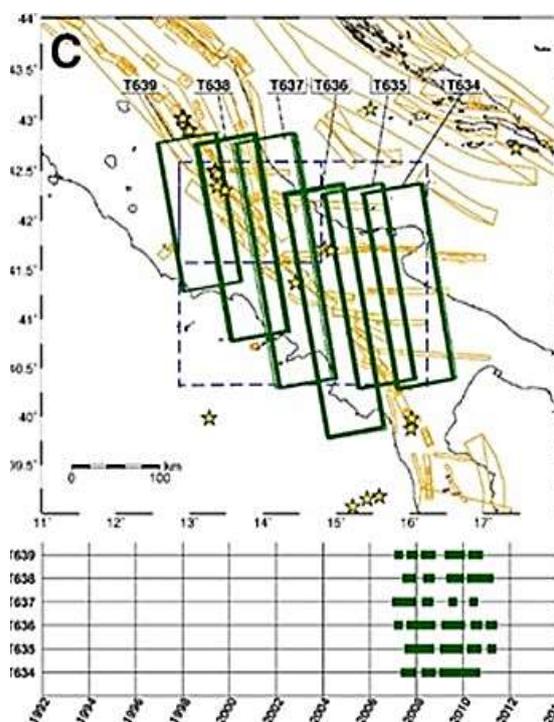


Figure 1: ALOS ascending tracks over central Italy

We consider the azimuth shifts obtained by the coregistration of three images acquired by ALOS-PALSAR over Italy. The observed area is that covered by the ground track 638 of ALOS, which is located over central Italy. Shifts are estimated by using the Multiple Aperture InSAR (MAI) method. The MAI method, based on split-beam InSAR processing, is able to extract along-track displacements from InSAR data more efficiently than the pixel amplitude correlation (Bechor & Zebker 2006).

2.2 GNSS

As mentioned before, GNSS signals passing through the ionosphere experience transmission delay proportional to TEC. Starting from RINEX data, dual frequencies GNSS receivers are able to calculate TEC along the slant path, according to the following formula:

$$STEC = \frac{1}{40.3} \left(\frac{L1^2 L2^2}{L2^2 - L1^2} \right) (P_2 - P_1) - \varepsilon, \quad (2)$$

where $L1$ (1575.42 MHz) and $L2$ (1227.60 MHz) are the two frequencies of the signal transmitted, $P1$ and $P2$ are the corresponding pseudoranges and ε represents the biases induced by the receiver, satellite, multipath etc. In order to minimize the biases, it is necessary to calibrate the STEC (Cesaroni et al., 2015a).

To such scope, the GOPI calibration software (<http://seemala.blogspot.it/2011/04/rinex-gps-ec-program-version-22.html>) has been used. The GOPI software provides TEC values projected to the vertical, under the assumption that the ionosphere can be represented by a single, thin, ionized layer located at 350 km (Mannucci et al., 1998) that is suitable for a quiet mid-latitude ionosphere. Hereafter, we refer to calibrated and verticalized TEC as TECcal. The projection to the vertical allows having TEC values not dependent on the position of the GNSS receivers at ground. Since ALOS ascending passages of T638 over the area of interest occurred during night, it is reasonable to assume that a mid-latitude ionosphere in the nighttime is “frozen”, i.e. not meaningfully changing, during a time interval of 5 minutes around the two passages of ALOS.

To find the portion of the ionosphere crossed by the SAR signal, some geometrical consideration are here reported and sketched in figure 2.

The distance D between the point P at ground and the projection of the point P_i (which is the intersection of the SAR signal and the ionosphere) at ground is approximately given by $H_i \cdot \tan(\alpha)$ (Reuveni et al 2015), where the look angle, α , is related to the incidence angle, θ , through the following formula:

$$\alpha = \sin^{-1} \left(\frac{Re}{Re+H_s} \sin \theta \right); \quad (3)$$

In which Re is the Earth's radius and H_i is the altitude of the thin layer approximated ionosphere.

Then, to highlight the difference (if any) between the ionospheric features in correspondence with the

two ALOS passages, maps of TECcal values have been considered. To obtain a fine representation of the ionospheric features as derived from regional TEC maps, it is necessary to have a very dense GNSS network. To the scope, the RING geodetic network has been selected, being composed by more than 100 receiver spread all over the Italian territory (Figure 3). TECcal maps boundaries have been defined to correspond to the SAR images boundaries, following the geometry described in Figure 2.

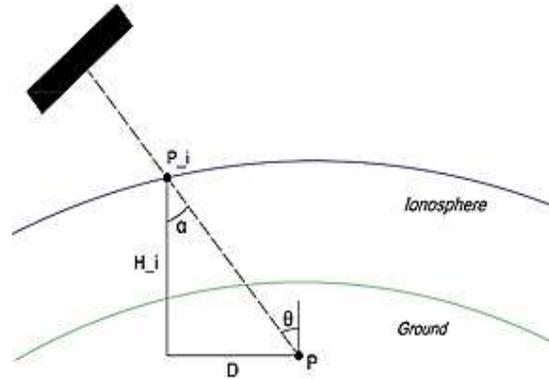


Figure 2: Ionosphere crossed by SAR signal. P_i is the intersection between the SAR signal and the ionosphere, D is the distance between the point P at ground and the projection of the point P_i on the ground, α and θ are respectively the look angle and the incidence angle, H_i is the altitude of the thin layer ionosphere.



Figure 3: Location of the GPS stations of the RING network

In such maps, TECcal are calculated at each IPP (Ionospheric Pierce Point) for every satellite in view by each receiver over a 5-minute interval (frozen ionosphere) centred in the time of the track. TECcal are then interpolated with the natural neighbour method (Sibson, 1981). Such interpolation technique has been selected according to Cesaroni (2015b) and to Foster and Evans (2008), in which the authors demonstrate that the natural neighbour interpolation is the best choice when regional TEC maps are considered. To avoid border effect due to the interpolation, maps have been calculated all over Italy and, then, restricted to fit the SAR image boundaries, as mentioned before. To evaluate the TECcal differences between the two passages, the values of the maps that correspond to the master and slave days are subtracted and a map of difference (Δ TECcal) is produced.

3 RESULTS AND DISCUSSION

The variations of the ionosphere between the different passes of the SAR cause an azimuth shift between the positions of the pixels of the master and slave image. This effect, also known as “azimuth streaks”, influences the optimal coregistration of the interferometric pair (Gray et al., 2000).

In this work, we concentrate on 3 ALOS-PALSAR images of central Italy, whose master/slave dates and parameters are summarized in Table 1. For each coregistered image, azimuth shifts have been determined and reported in figures 4 to 6.

Table 1: Images and relative parameters used in this work, where θ_m and θ_s are respectively the incidence angle of the master and the slave image, Bt stands for the temporal baseline and Bp stands for perpendicular baseline.

Image #	1	2	3
Master	01/07/2007	01/07/2007	16/08/2007
θ_m (°)	38.7191	38.7191	38.7395
Slave	16/08/2007	01/10/2007	01/10/2007
θ_s (°)	38.7395	38.7223	38.7223
Bt(days)	46	92	46
Bp (m)	279.9401	539.2134	259.2523

In all three cases, the shift varies from -3 m to 3 m. By comparing such figures, a strong similarity between image #1 (fig. 4) and #2 (fig. 6) can be noticed. In particular, a similar pattern of the shift are present, but with opposite sign. The sign shift is

because the day 16 August 2007 is used as slave in figure 4, while in figure 6 it is used as master.

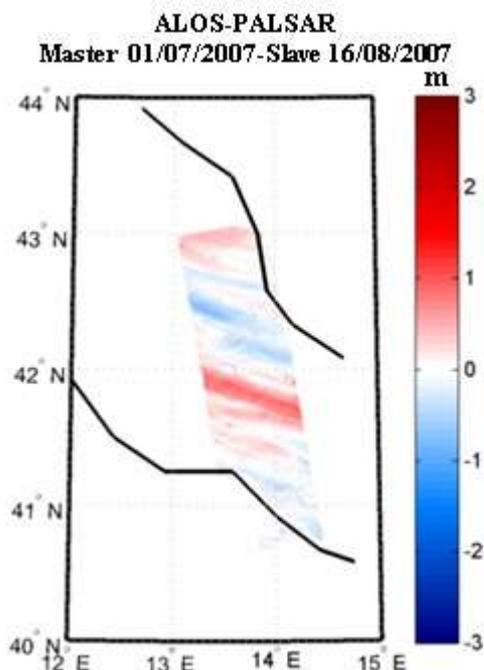


Figure 4: Shift obtained by coregistration of the master (01/07/2007) and slave (16/08/2007) images

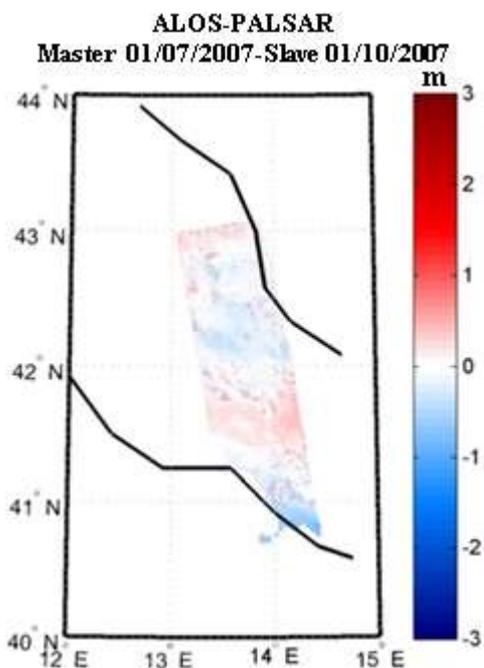


Figure 5: Shift obtained by coregistration of the master (01/07/2007) and slave (01/10/2007) images

By looking closely to the red strike located at around 41.8°N in figure 4 (or the corresponding blue strike in figure 6), it is possible to note that it divides into 3 smaller sub-strikes.

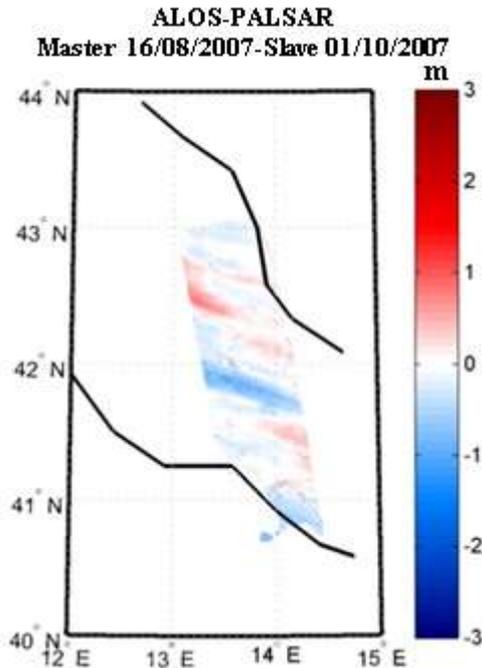


Figure 6: Shift obtained by coregistration of the master (16/08/2007) and slave (01/10/2007) images

In Figure 5, the streaks are broader and less structured than those in Figures 4 and 6. In the case of Figure 5, 4 large streaks can be identified:

- a positive shift around 43°N,
- a negative shift between 41.7°N and 42.8°N,
- a positive shift between 41.5°N and 41.7°N,
- a negative shift between 40.7°N and 41.5°N.

Since the SAR image of the 16 August is used both in Figures 4 and 6 but not in Figure 5, the ionosphere in 16 August 2007 seems to be the principal responsible for the streaks appearance. According to Dst (Figure 7) and Kp (not shown) indices, 16 August 2007 can be considered quiet from the geomagnetic point of view (red box of middle panel). Also 1 July 2007 (red box, top panel of Figure 7) and 1 October 2007 (red box, bottom panel of Figure 7) can be considered quiet.

In Figures 8 to 10, the $\Delta\text{TEC}_{\text{cal}}$ maps corresponding to the images #1 to #3, respectively, are reported. All maps presents structures of

$\Delta\text{TEC}_{\text{cal}}$, that are likely responsible for the presence of streaks in the ALOS-PALSAR images.

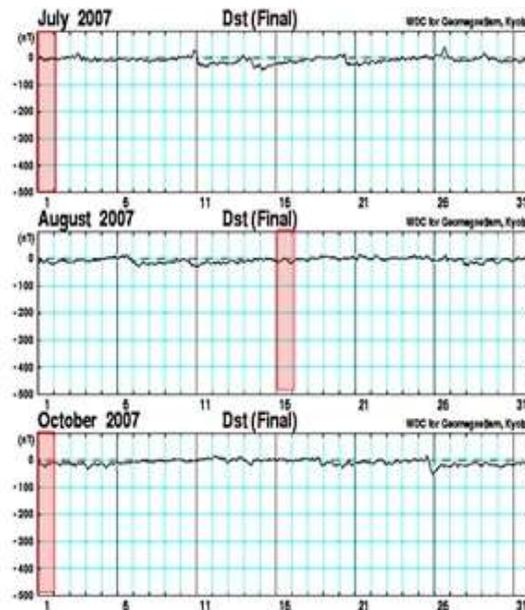


Figure 7: Dst index of July, August and October 2007. (http://wdc.kugi.kyoto-u.ac.jp/dst_final/). Days of the images are highlighted in the red boxes

We remind that, to highlight the correspondence between azimuth streaks and gradients of $\Delta\text{TEC}_{\text{cal}}$, the ionosphere maps are fitted on SAR area, according to the method described in Section 2.2. By comparing $\Delta\text{TEC}_{\text{cal}}$ behaviour in the figures, we can see that the $\Delta\text{TEC}_{\text{cal}}$ values show largely different patterns among the three days, reinforcing the idea on how variable can be the ionospheric conditions in different days, even if at mid latitudes. Figures 8-10 show patterns of $\Delta\text{TEC}_{\text{cal}}$ characterized by large $\Delta\text{TEC}_{\text{cal}}$ gradients mainly in the N-NW direction of all over the maps, and, in particular, gradients of 1-2 TECu in the latitudinal range between 42.2°N and 41.5°N, where the strikes, highlighted in figures 4-6, are present. This confirm how variation of few TECu's and below have a great impact on the SAR imaging. It is important to note that $\Delta\text{TEC}_{\text{cal}}$ map in figure 8 is obtained by differencing TEC data acquired in the same season, while the maps in Figure 9 and 10 have been obtained by differencing summer and fall data.

Considering that the geospace conditions can be considered quiet for all the 3 days, the seasonal variation seems to be a key point in understanding the impact of the ionosphere on the streaks appearance in the analysed case.

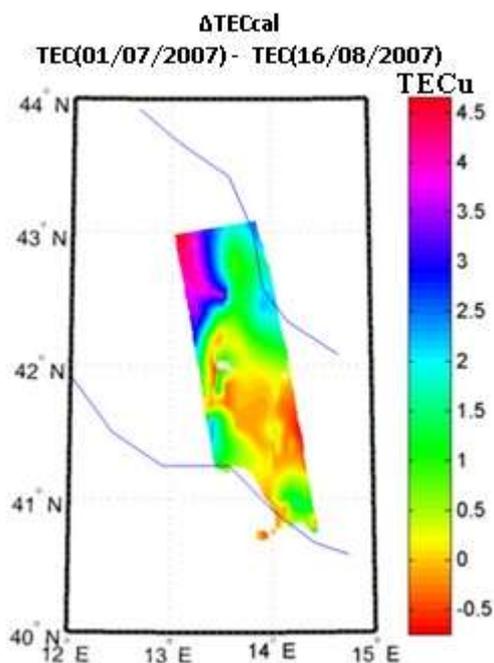


Figure 8: ΔTECcal maps (01/07/2007-16/08/2007) The color bar shows the differences of the ΔTECcal values in TECu

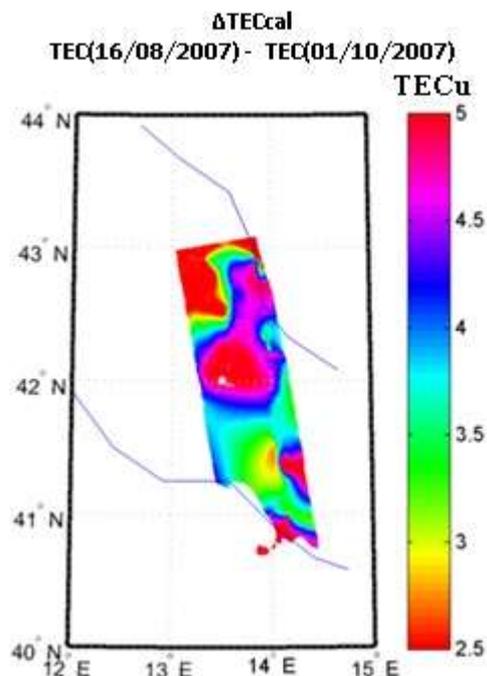


Figure 10: ΔTECcal maps (16/08/2007-01/10/2007) The color bar shows the differences of the ΔTECcal values in TECu

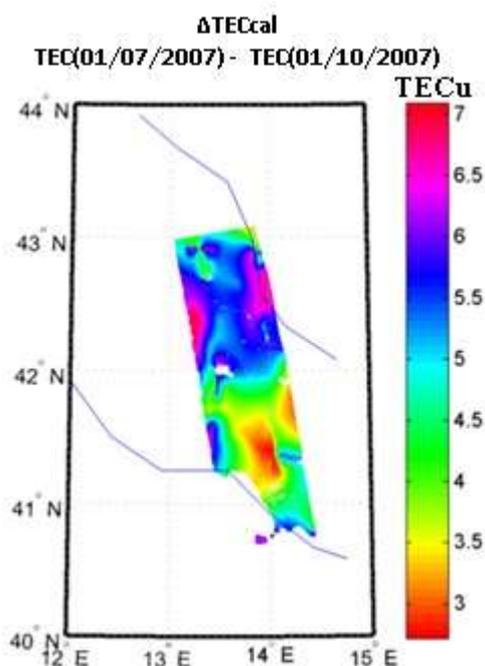


Figure 9: ΔTECcal maps (01/07/2007-01/10/2007) The color bar shows the differences of the ΔTECcal values in TECu .

The spatial scale of the identified ΔTECcal gradients has been reached thanks to the fine spatial density of the GNSS receivers at ground. Such resolution is not reachable with the standard, global, ionospheric models commonly used.

4 CONCLUSIONS

TEC fluctuation can significantly affect InSAR applications. To investigate these ionospheric effects on ALOS-PALSAR images, we analysed 3 maps of azimuth shift and 3 maps of TEC differences (ΔTECcal) over central Italy.

The InSAR images present meaningful streaks, with shift ranging from -3 m to 3 m.

The ΔTECcal maps were obtained by:

- using dual frequency GPS data of the RING network,
- assuming that the ionosphere is a thin layer located at an altitude of 350 km above the Earth's surface,
- processing the GPS data with a calibration technique to reduce the errors introduced by satellite receiver biases

- interpolating the data with natural neighbour interpolation

In order to show the condition of the ionosphere in terms of TEC fluctuation in correspondence with the coregistration of two L-band SAR images (Fig. 4-6), the maps of $\Delta\text{TEC}_{\text{cal}}$ were projected on the same area. Results indicate how variations below 1-2 TECu observed with $\Delta\text{TEC}_{\text{cal}}$ provides the appearance of streaks. The map (Fig. 9) in which

- the maximum value of $\Delta\text{TEC}_{\text{cal}}$ is larger than those of the other 2 maps
- the gradient are less and smaller than those in Figures 8 and 10

corresponds to the case with less strikes.

Furthermore, the results show that the 16 August 2007 seems to be the principal responsible for the streaks appearance.

These case study suggests how variations of few TECUs can affect considerably ALOS-PALSAR images. Further refinement of the work will aim at compensating for the seasonal variation, aiming at catching finer ionospheric effects and different geospace conditions and geographical sectors will be investigated. A comparison between the results with the TEC mapping here shown and common ionospheric models (like GIM, WBMOD) is currently ongoing. Future refinement of the work will aim at quantitatively assess the impact of TEC fluctuation on L-band SAR, by evaluating the TEC gradient along the azimuth direction. Effect of the zonal and meridional $\Delta\text{TEC}_{\text{cal}}$ gradients will be also investigated.

ACKNOWLEDGEMENTS

ALOS PALSAR data were provided by the European Space Agency through Category-1 Proposal 26350. Authors are grateful to Dr. Antonio Avallone for his support with RING data.

REFERENCES

- Bechor, N. B., H. A. Zebker, *Measuring two-dimensional movements using a single InSAR pair*, Geophys. Res. Lett., 33, L16311, 2006, DOI:10.1029/2006GL026883.
- Bürgmann, R., Rosen, P. A., Fielding, E. J., *Synthetic aperture radar interferometry to measure Earth's surface topography and its deformation*. Annual review of earth and planetary sciences, 28(1), 169-209, 2000, DOI:10.1146/annurev.earth.28.1.169.
- Cesaroni, C., Spogli, L., Alfonsi, L., De Franceschi, G., Ciraolo, L., Monico, J. F. G., Scotto, C., Romano, V., Aquino, M., Bougard, B., *L-band scintillations and calibrated total electron content gradients over Brazil during the last solar maximum*. Journal of Space Weather and Space Climate, 5, A36, 2015a, DOI:10.1051/swsc/2015038.
- Cesaroni C., *A multi instrumental approach to the study of equatorial ionosphere over South-America*, PhD dissertation, 2015b, DOI: 10.6092/unibo/amsdottorato/6889.
- Chen, A. C., Zebker, H. A., *Reducing ionospheric effects in InSAR data using accurate coregistration*. IEEE Transactions on Geoscience and Remote Sensing, 52(1), 60-70, 2014, DOI:10.1109/TGRS.2012.2236098.
- Ferretti, A., Monti Guarnieri, A., Prati, C., Rocca, F., Massonnet, D., *InSAR Principles: Guidelines for SAR Interferometry Processing and Interpretation*. The Netherlands, Noordwijk: ESA Publications. TM-19, 2007, ISBN: 92-9092-233-8.
- Foster, M.P., A.N. Evans. *An evaluation of interpolation techniques for reconstructing ionospheric TEC maps*. IEEE Trans. Geosci. Remote Sens., 46 (7), 2153–2164, 2008, DOI: 10.1109/TGRS.2008.916642.
- Gray, A. L., Mattar, K. E., Sofko, G., *Influence of ionospheric electron density fluctuations on satellite radar interferometry*, Geophys. Res. Lett., 27(10), 1451–1454, 2000, DOI: 10.1029/2000GL000016.
- Hanssen, R.F., (2001) *Radar Interferometry: Data Interpretation and Error Analysis*. Kluwer Academic Publishers, Dordrecht, 328 pp. ISBN: 0-7923-6945-9.
- Kelley, M. C., *The Earth's Ionosphere: Plasma Physics and Electrodynamics* (2nd ed.). Academic Press, 2009, ISBN: 9780120884254.
- Mannucci, A.J., B.D. Wilson, D.N. Yuan, C.H. Ho, U.J. Lindqwister, T.F. Runge. *A global mapping technique for GPS-derived ionospheric total electron content measurements*. Radio Sci., 33 (3), 565, 1998, DOI: 10.1029/97RS02707.
- Meyer, F. J., Agram, P. S. *Developing an Error Model for Ionospheric Phase Distortions in L-Band SAR and InSAR Data*. In FRINGE'15: Advances in the Science and Applications of SAR Interferometry and Sentinel-1 InSAR Workshop, Frascati, Italy, 23-27 March 2015.
- Reuveni, Y., Bock, Y., Tong, X., Moore, A. W., *Calibrating interferometric synthetic aperture radar (InSAR) images with regional GPS network atmosphere models*, Geophys. J. Int. 202, 2106–2119, 2015, DOI: 10.1093/gji/ggv253
- Sibson, R., *A brief description of natural neighbor interpolation (Chapter 2)*. In V. Barnett. *Interpreting Multivariate Data*. Chichester: John Wiley. pp. 21–36, 1981.
- Zhu, W., Ding, X. L., Jung, H. S., Zhang, Q., Zhang, B. C., & Qu, W. (2016). *Investigation of ionospheric effects on SAR Interferometry (InSAR): A case study of Hong Kong*. Advances in Space Research, 58, 564–576 2016, DOI: 10.1016/j.asr.2016.05.004.

**SATELLITE AND
AERIAL
SENSING**

New Wavelet Based Spatiotemporal Fusion Method

Amal Ibnelhobyb¹, Ayoub Mouak¹, Amina Radgui¹, Ahmed Tamtaoui¹, Ahmed Er-Raji²,
Driss El Hadani², Mohamed Merdas² and Faouzi Mohamed Smiej²

¹*STRS Lab, Institut National des Postes et Télécommunication-INPT, Rabat, Morocco*

²*Centre Royal de Télédetection Spatiale-CRTS, Rabat, Morocco*

{ibnelhobyb, mouak, radgui, tamtaoui}@inpt.ac.ma, {er-raji, elhadani, merdas, smiej}@crts.gov.ma

Keywords: Spatiotemporal fusion, Landsat, MODIS, NDVI, STARFM, ESTARFM, WSAD-FM, WESTARFM.

Abstract: Satellite image sensors are able to give images at high temporal resolution as the MODIS sensor that gives an image every day but with low spatial resolution, or at high spatial resolution as the Landsat sensor that gives images at 30m but with a revisit cycle of 16 days. Thus, these sensors are not able to give images with both high spatial and high temporal resolution. This need has become more and more absolute for many applications. Therefore spatiotemporal fusion methods were proposed. By applying these methods on images from different sensors with different spatial and temporal resolution, we can take the advantage of the high spatial and high temporal resolution of these sensors. As a result we get an image with both high spatial and high temporal resolution. We introduce in this paper a new method, the Wavelet base Enhanced Spatial and Temporal Adaptive Reflectance Fusion Model (WESTARFM), which is an improvement of the ESTARFM method. It uses the principle of wavelet transform with the original ESTARFM method. We have applied our method to predict daily NDVI in a study site in an irrigated zone in the region of TADLA in MOROCCO. Results have been compared with other methods.

1 INTRODUCTION

Satellite images are more and more used in many applications such as vegetation monitoring, ecosystem disturbance and land cover mapping. However, a tradeoff exists between spatial and temporal resolution in available satellite data. Satellite data obtained by moderate resolution sensors like the Moderate resolution Imaging Spectroradiometer (MODIS) gives daily observations of the entire earth but with a low spatial resolution attending 1 km (Gao et al., 2014). Whereas, data obtained by Landsat sensors gives more spatial details with a spatial resolution of 30 m but they have a long revisit cycle of 16 day and their use is limited by the presence of clouds. In order to get full use of advantageous characteristics of these sensors, fusion methods were proposed to combine satellite data from different sensors. By using spatiotemporal fusion we can obtain satellite images

with both high spatial and high temporal resolution. Many fusion methods have been proposed. They can be classified into four categories(Chen, Huang, & Xu, 2015): i. Transformation based methods (Ghannam, Awadallah, Abbott, & Wynne, 2014), ii. Learning based methods (Huang & Song, 2012)-(Song & Huang, 2013), iii. Reconstruction based methods(Gao, Masek, Schwaller, & Hall, 2006)-(Zhu, Chen, Gao, Chen, & Masek, 2010)-(Zhu et al., 2016)-(Hilker, Wulder, Coops, Linke, et al., 2009)-(Fu, Chen, Wang, Zhu, & Hilker, 2013), iv. Data assimilation based methods (Chemin & Honda, 2006). Spatial and Temporal Adaptive Reflectance Fusion Model (STARFM) (Gao et al., 2006) is one of the most common methods widely used for spatiotemporal fusion. It is a reconstruction based method that was proposed by Feng Gao on 2006. This method introduced the use of neighbouring pixels and windowing to predict Landsat-like images. However it was convenient only for homogenous regions. Enhanced STARFM (ESTARFM) (Zhu et al., 2010)

was proposed after that to overcome the limitation of STARFM and introduced a conversion coefficient that makes it applicable for heterogeneous regions using two pairs of Landsat and MODIS data. A wavelet base method has been used with the STARFM method (WSAD-FM) (Ghannam et al., 2014) it decompose the Landsat image at high and low frequencies and predict each part separately using only Landsat image for high frequencies and Landsat and MODIS images for low frequencies.

The paper presents a new fusion model based on the same concept of the WSAD-FM(Ghannam et al., 2014) but uses the wavelet transform with the ESTARFM(Zhu et al., 2010) method. This fusion method is applied on NDVI data in the region of Tadla in Morocco. We have used actual Landat 8 NDVI and MODIS NDVI data for evaluation and calculate commonly used statistic parameters RMSE, AAD and R2 to compare the accuracy of our method with the STARFM, ESTARFM and WSAD-FM methods. First the theoretical basis and the proposed method will be introduced, after, evaluation of our method will be explained. At the end, the results of this evaluation will be discussed.

2 THEORETICAL BASIS

2.1 ESTARFM

As an improvement of the STARFM method (Gao et al., 2006), the Enhanced Spatial and Temporal Adaptive Reflectance Fusion Model was proposed to overcome its limitation in prediction on heterogeneous and changing regions, this by using a conversion coefficient that presents the heterogeneity of coarse pixels (Chen et al., 2015). The STARFM supposes the presence of one land cover type on coarse pixels (pure pixels) which the case of homogenous regions. But for heterogeneous regions different land convers types are present on a coarse pixel. Therefore the ESTARFM considers the presence of mixed pixels and apply the linear mixture model to calculate the reflectance change of each present class. The sum of this changes is the change of the coarse mixed pixel between two days. It requires two pairs of Landsat and MODIS images and a MODIS image from the predicted day. It is described in the following steps:

- For a fine central pixel of a moving window we use thresholding or a classification map to find spectrally similar pixels.
- A weighting function W_i is calculated for these n similar pixels after being filtered. This

weighting function is based on spectral similarity, temporal difference and spatial distance.

- Calculate the conversion coefficient v_k presenting the ratio of reflectance change for a class k represented by the fine pixel L_k to the reflectance change of the coarse pixel M between two days t_m and t_n :

$$v_k = \frac{L_{kn} - L_{km}}{M_n - M_m} \quad (1)$$

- Predict the Landsat image L at time t_p for the central pixel $(x_{w/2}, y_{w/2})$ within the moving window of size w , based on pair input L and M from time t_m to have L_m and from time t_n to have L_n :

$$\begin{aligned} L_m(x_{w/2}, y_{w/2}, t_p) = & \\ L(x_{w/2}, y_{w/2}, t_m) + & \\ \sum_{i=1}^N W_i \cdot v_i \cdot \left(\begin{aligned} & M(x_i, y_i, t_p) - \\ & M(x_i, y_i, t_m) \end{aligned} \right) & \end{aligned} \quad (2)$$

$$\begin{aligned} L_n(x_{w/2}, y_{w/2}, t_p) = & \\ L(x_{w/2}, y_{w/2}, t_n) + & \\ \sum_{i=1}^N W_i \cdot v_i \cdot \left(\begin{aligned} & M(x_i, y_i, t_p) - \\ & M(x_i, y_i, t_n) \end{aligned} \right) & \end{aligned} \quad (3)$$

- The predictions at times t_m and t_n are summed and temporally weighted to calculate the final prediction. This temporal weight T presents the contribution of each pair:

$$\begin{aligned} L(x_{w/2}, y_{w/2}, t_p) = & T_m \times L_m(x_{w/2}, y_{w/2}, t_p) \\ & + T_n \times L_n(x_{w/2}, y_{w/2}, t_p) \end{aligned} \quad (4)$$

2.2 Wavelet-based ESTARFM

In this section we present our proposed method, the Wavelet based Enhanced Spatial and Temporal Adaptive Reflectance Fusion Model used to combine data with different spatial resolution and different temporal resolution in order to predict an image with both high spatial and high temporal resolution.

Our method is based on the ESTARFM method but it uses the wavelet transform to predict more details on the image (Ghannam et al., 2014). In the original ESTARFM the whole image is used for prediction, but in our proposed method the Landsat image is decomposed into high and low frequencies using the wavelet transform. After that each component is used for prediction separately. As the ESTARFM our method requires two pairs of Landsat and MODIS data and a MODIS image from the prediction day. The WESTARFM is implemented following these steps:

- The two Landsat images from time t_m and t_n are decomposed into high and low frequencies using wavelet transform.
- Predict approximation coefficients L_a of Landsat image at time t_p using low frequency components of Landsat images at time t_m and t_n and MODIS images at time t_m , t_n and t_p :

$$\begin{aligned} L_{am} & \left(x_{w/2}, y_{w/2}, t_p \right) \\ & = L_a \left(x_{w/2}, y_{w/2}, t_m \right) \\ & + \sum_{i=1}^N W_i \cdot v_i \cdot \left(M(x_i, y_i, t_p) - M(x_i, y_i, t_m) \right) \end{aligned} \quad (5)$$

$$\begin{aligned} L_{an} & \left(x_{w/2}, y_{w/2}, t_p \right) \\ & = L_a \left(x_{w/2}, y_{w/2}, t_n \right) \\ & + \sum_{i=1}^N W_i \cdot v_i \cdot \left(M(x_i, y_i, t_p) \right. \\ & \left. - M(x_i, y_i, t_n) \right) \end{aligned} \quad (6)$$

- Predict detail coefficients L_d of Landsat images at time t_p using only high frequency components of Landsat images at time t_m and t_n :

$$\begin{aligned} L_{dm} & \left(x_{w/2}, y_{w/2}, t_p \right) \\ & = \sum_{i=1}^N W_i \cdot L_d \left(x_{w/2}, y_{w/2}, t_m \right) \end{aligned} \quad (7)$$

$$\begin{aligned} L_{dn} & \left(x_{w/2}, y_{w/2}, t_p \right) \\ & = \sum_{i=1}^N W_i \cdot L_d \left(x_{w/2}, y_{w/2}, t_n \right) \end{aligned} \quad (8)$$

- Calculate the final images at time t_m and t_n by applying the inverse wavelet transform on the predicted low frequency and high-frequency components.
- Estimate the final predicted image at time t_p using temporal weight T (Zhu et al., 2010).

3 RESULTS AND DISCUSSION

3.1 Data and Pre-processing

High Spatial Resolution Landsat 8 images and High Temporal Resolution MODIS images are required for evaluating the accuracy of the proposed method. In order to generate a daily prediction of NDVI, the MODIS Surface Reflectance Daily 250m (MOD09GQ bands 1,2) was selected. As shown in Figure 1. Eight Landsat 8 and MODIS images representing days: 05 June 2015(156), 21 June 2015(172), 07 July 2015(188), 23 July 2015(204), 08 August 2015(220), 24 August 2015(236), 09 September 2015(252) and 25 September 2015(268) were selected for evaluation. The Landsat 8 images were downloaded from the USGS GLOVIS website, and the MODIS Surface Reflectance data were downloaded from the Reverb ECHO website. This Data cover an irrigated area of 800m x 800m in the region of Tadla (32° 28 N, 5° 38 W) situated in central Morocco.

Erdas imaging were used as a preprocessing tool to calculate the Landsat 8 Surface Reflectance from Landsat 8 data, and to generate the Landsat NDVI images from Landsat 8 reflectance and MODIS NDVI from MODIS Surface Reflectance. Landsat and MODIS input data should have the same projection and same pixel size, thus Erdas imaging was used for UTM projection and ARCMAP was used for resampling of MOD09 in order to have the same resolution as Landsat images (30m).

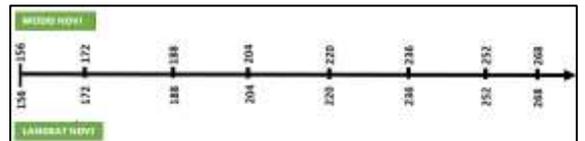


Figure 1: DOYs of Landsat 8 and MODIS Data used for evaluation.

3.2 Evaluation and Results

The proposed method was used for predicting Landsat NDVI images. This method is applied on NDVI images since it gives more accurate results and less complexity than applying the fusion methods on RED and NIR bands used for calculating NDVI (Jarihani et al., 2014)

Two pairs of Landsat NDVI and MODIS NDVI images are needed with the MODIS NDVI from the prediction day, in the original method of ESTAFM the author used the pairs of Landsat and MODIS from the start and the end of the period as input of the method, for our evaluation we have used the two pairs from previous days. For WSADAFM method only one pair of Landsat and MODIS NDVI is needed.

The prediction was performed using eight Landsat and MODIS NDVI images in DOYs 156, 172, 188, 204, 220, 236, 252, 268 when Landsat NDVI images are available in order to evaluate the accuracy of the

proposed method and compare it with STARFM, ESTARFM and WSAD-FM methods.

Figure 2 shows the prediction results of the methods for the selected days compared with real images. We have the results for all the days for STARFM, ESTARFM, WSAD-FM and WESTARFM except DOYs 156 and 172 since they are used as first inputs to the fusion. Visually, we can see that the Landsat-like images contain all the details of the region. However, for some days, as an example DOY 220, details of the image are lost in some zones and the image is noisy as a result of clouds. The quality of prediction depends on the quality of input images(Hilker, Wulder, Coops, Seitz, et al., 2009), good results can be obtained if Landsat and MODIS NDVI input images are clean from clouds also results are affected by the inconsistency existing between the Landsat and MODIS sensors(Gevaert & García-Haro, 2015) .

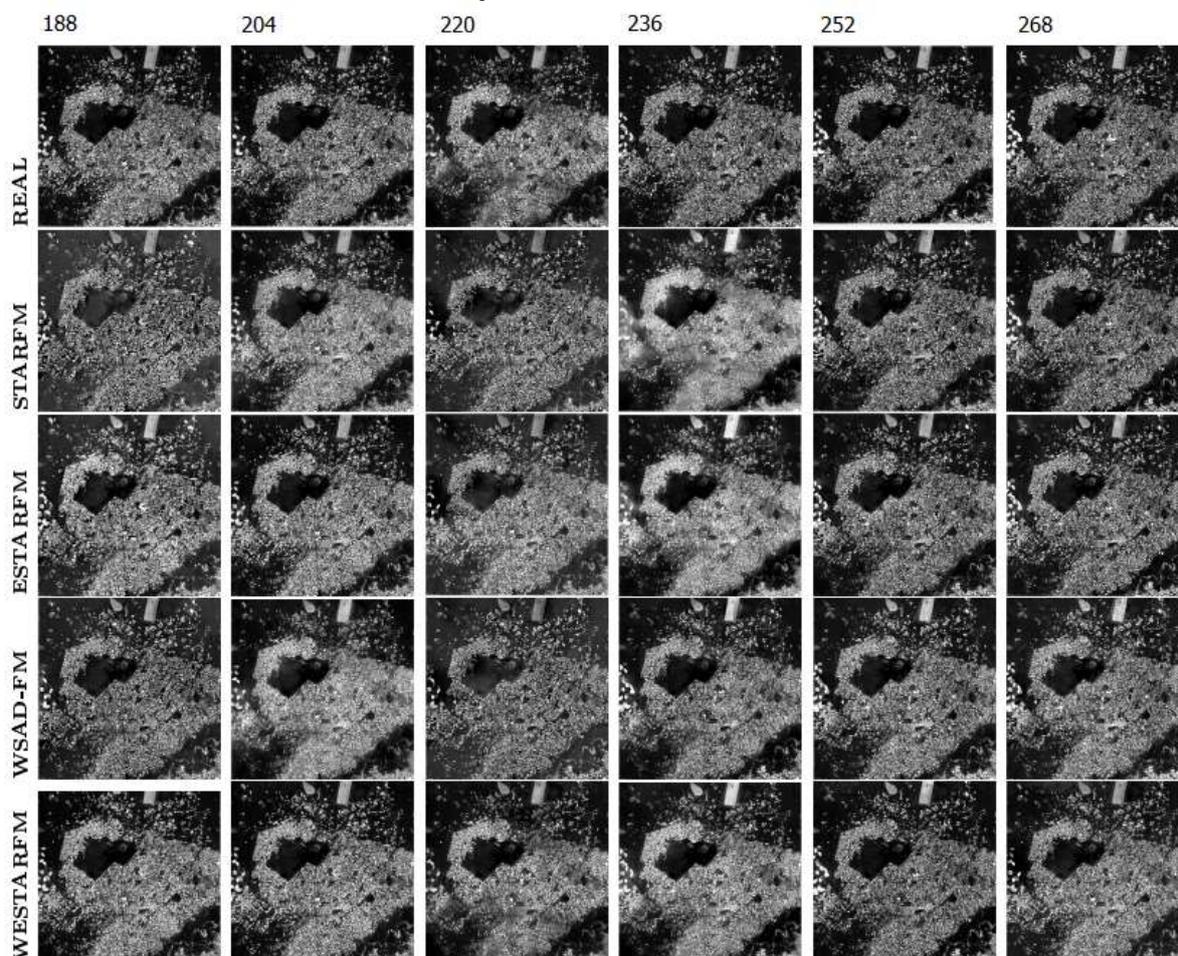


Figure 2: NDVI prediction results using ESTAFM, WSADFM and WESTARM.

Figures 3, 4, 5, 6 show real Landsat NDVI and predicted Landsat NDVI using our proposed method WESTARFM for DOY 252. We can see that the predicted image contains most of the details and visually it is almost similar to the real image. If we zoom in a particular zones like a citrus zone, a rainfed zone and an irrigated zone, as it is illustrated on figures (5,6,8,9,11,12), we notice, visually, that the WESTARFM method was able to predict almost the same NDVI information as the real image.

Figures 7, 10 and 13 show the correlation between the real and predicted NDVI values using the WESTARFM method for the 3 selected zones at time 252. Results show that the prediction is better for the citrus zone. We can say that one of the parameters that affect the performance of the prediction is the value of input NDVI, more the value of NDVI is high more the prediction is better. We assume also that the prediction is better for citrus zone since it is more stable between the input and the prediction days than culture areas.

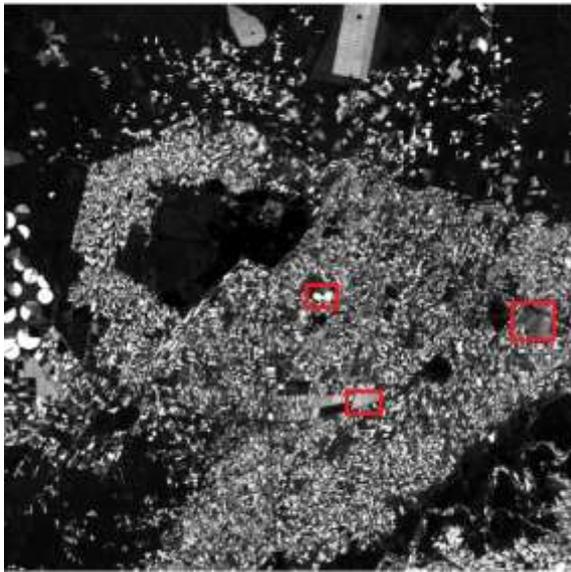


Figure 3: Real Landsat NDVI DOY 252.

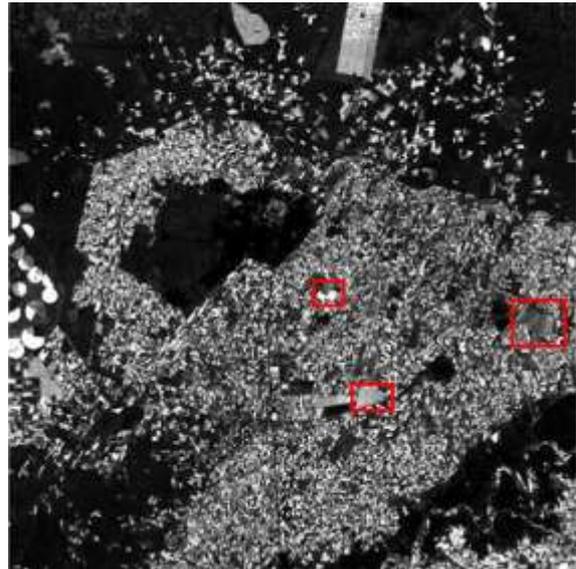


Figure 4: Predicted Landsat NDVI DOY 252 using WESTARFM.



Figure 5: Zoom in a citrus zone in real Landsat NDVI image DOY 252.



Figure 6: Zoom in a citrus zone in predicted Landsat NDVI image using WESTARFM DOY 252.

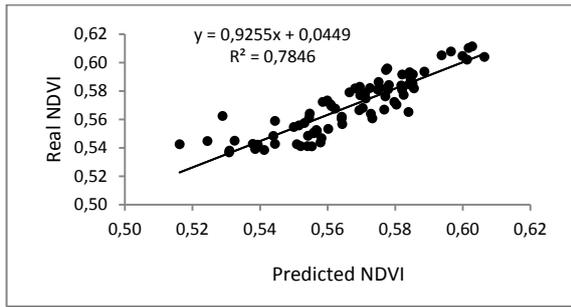


Figure 7: Correlation between real and predicted Landsat NDVI in a citrus zone DOY 252



Figure 11: Zoom in a rainfed zone in real Landsat NDVI image DOY 252.



Figure 8: Zoom in an irrigated zone in real Landsat NDVI image DOY.



Figure 12: Zoom in a rainfed zone in predicted Landsat NDVI image using WESTARFM DOY 252.



Figure 9: Zoom in an irrigated zone in predicted Landsat NDVI image using WESTARFM DOY 252.

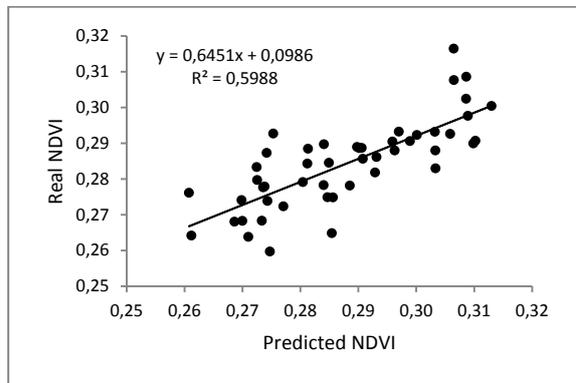


Figure 13: Correlation between real and predicted Landsat NDVI in a rainfed zone DOY 252

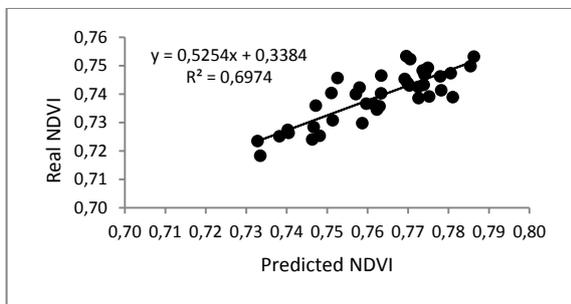


Figure 10: Correlation between real and predicted Landsat NDVI in an irrigated zone DOY 252.

We have calculated the Root Mean Square Error (RMSE), the Absolute Average Difference (AAD) and the Coefficient of Determination (R2) to validate the prediction results of the three methods.

Results presented in table 1 have shown that the WESTARFM method gives more accurate results than the ESTARFM and WSAD-FM with a RMSE attending 0.05, AAD of 0.02 and a R2 of 0.64

4 CONCLUSION

A new fusion model found on Wavelet transform and ESTARFM method was presented (WESTARFM). The model utilizes the Wavelet transform to decompose the Landsat data into approximation and detail coefficients. Each of these components is, after that, predicted separately with the ESTARFM method. Two pairs of Landsat and MODIS NDVI from previous days and a MODIS NDVI from

prediction date were needed as inputs of the WESTARFM to predict an unavailable Landsat NDVI image. The WESTARFM was tested on NDVI and compared with other methods. Results have shown that the proposed method gives more accurate results for most of evaluated dates. Therefor working on frequency domain improves the prediction and predicts more image details. This method was tested on NDVI but it can be applicable also on Landsat and MODIS bands.

Table 1: Statistic validation of prediction results using the four fusion methods.

DOY		STARFM	ESTARFM	WSAD-FM	WESTARFM
188	RMSE	0,11	0,09	0,08	0,06
	AAD	0,08	0,07	0,06	0,05
	R2	0,39	0,42	0,52	0,64
204	RMSE	0,08	0,06	0,09	0,06
	AAD	0,06	0,03	0,06	0,03
	R2	0,25	0,20	0,34	0,37
220	RMSE	0,07	0,06	0,09	0,08
	AAD	0,05	0,03	0,06	0,06
	R2	0,38	0,26	0,32	0,29
236	RMSE	0,12	0,08	0,11	0,06
	AAD	0,02	0,06	0,08	0,05
	R2	0,34	0,24	0,24	0,41
252	RMSE	0,08	0,07	0,09	0,05
	AAD	0,04	0,04	0,07	0,02
	R2	0,28	0,25	0,31	0,41
268	RMSE	0,07	0,06	0,1	0,09
	AAD	0,03	0,03	0,06	0,05
	R2	0,22	0,19	0,55	0,6

REFERENCES

- Chemin, Y., & Honda, K. 2006. Spatiotemporal fusion of rice actual evapotranspiration with genetic algorithms and an agrohydrological model. *IEEE Transactions on Geoscience and Remote Sensing*, 44(11), 3462–3469
- Chen, B., Huang, B., & Xu, B. 2015. Comparison of Spatiotemporal Fusion Models: A Review. *Remote Sensing*, 7(2), 1798–1835.
- Fu, D., Chen, B., Wang, J., Zhu, X., & Hilker, T. 2013. An improved image fusion approach based on enhanced spatial and temporal the adaptive reflectance fusion model. *Remote Sensing*, 5(12), 6346–6360.
- Gao, F., Hilker, T., Zhu, X., Anderson, M., Masek, J., Wang, P., & Yang, Y. 2014. Monitoring, *Fusing Landsat and MODIS data for vegetation*, (september), 47–60.
- Gao, F., Masek, J., Schwaller, M., & Hall, F. 2006. On the Blending of the MODIS and Landsat ETM + Surface Reflectance, 20771(2), 20794.
- Gevaert, C. M., & Garcia-Haro, F. J. 2015. A comparison of STARFM and an unmixing-based algorithm for Landsat and MODIS data fusion. *Remote Sensing of Environment*, 156, 34–44.
- Ghannam, S., Awadallah, M., Abbott, a. L., & Wynne, R. H. 2014. Multisensor Multitemporal Data Fusion Using Wavelet Transform. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XL-1(November)*, 121–128.
- Hilker, T., Wulder, M. a., Coops, N. C., Linke, J., McDermid, G., Masek, J. G., White, J. C., Gao, F. 2009. A new data fusion model for high spatial- and temporal-resolution mapping of forest disturbance based on Landsat and MODIS. *Remote Sensing of Environment*, 113(8), 1613–1627.

- Hilker, T., Wulder, M. A., Coops, N. C., Seitz, N., White, J. C., Gao, F., Masek, J. G., Stenhouse, G. 2009. Generation of dense time series synthetic Landsat data through data blending with MODIS using a spatial and temporal adaptive reflectance fusion model. *Remote Sensing of Environment*, 113(9), 1988–1999.
- Huang, B., & Song, H. 2012. Spatiotemporal reflectance fusion via sparse representation. *IEEE Transactions on Geoscience and Remote Sensing*, 50(10 PART1), 3707–3716.
- Jarihani, A. A., McVicar, T. R., van Niel, T. G., Emelyanova, I. V., Callow, J. N., & Johansen, K. 2014. Blending landsat and MODIS data to generate multispectral indices: A comparison of “index-then-blend” and “Blend-Then-Index” approaches. *Remote Sensing*, 6(10), 9213–9238.
- Song, H., & Huang, B. 2013. Spatiotemporal satellite image fusion through one-pair image learning. *IEEE Transactions on Geoscience and Remote Sensing*, 51(4), 1883–1896.
- Zhu, X., Chen, J., Gao, F., Chen, X., & Masek, J. G. 2010. An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions. *Remote Sensing of Environment*, 114(11), 2610–2623.
- Zhu, X., Helmer, E. H., Gao, F., Liu, D., Chen, J., & Lefsky, M. A. 2016. A flexible spatiotemporal method for fusing satellite images with different resolutions. *Remote Sensing of Environment*, 172, 165–177.

Multi-Satellite Interferometric SAR System

Andon Lazarov^{1,2}, Chavdar Minchev³ and Dimitar Minchev¹

¹ *Burgas Free University, 62 SanStefano Str., Burgas, Bulgaria*

² *K.N. Toosi University of Technology, Tehran, Iran*

³ *Shumen University, 115 Universitetska Str. Shumen, Bulgaria*

lazarovr@bfu.bg, chavdar_minchev@yahoo.com, mitko@bdu.bg

Keywords: InSAR, Multi-satellite system, InSAR geometry, InSAR signal model.

Abstract: In the present work a multi-satellite SAR system is considered. Between every pair of SAR satellites an interferometric concept is implemented. It allows the height of each pixel on the surface to be evaluated with high precision and a three dimensional map to be created. InSAR geometry is analytically described. Mathematical expressions for determination of current distances between SAR's and detached pixels on the ground, and principal InSAR parameters are derived. A model of linear frequency modulated (LFM) SAR signal with InSAR applications, reflected from the surface is developed. Correlation and spectral SAR image reconstruction algorithms and co-registration procedure are described. To verify the correctness of the signal model and image reconstruction and co-registration algorithm numerical experiment is carried out.

1 INTRODUCTION

Synthetic Aperture Radar (SAR) is a coherent active microwave imaging instrument. Back scattered information of a target is recorded as a complex signal with amplitude and phase information. Interferometric SAR (InSAR) technique makes use of phase difference information extracted from two complex valued SAR images acquired from different orbit positions. This information is useful in measuring several geophysical quantities such as topography, slope, deformation (volcanoes, earthquakes, ice fields), glacier studies, vegetation growth etc (Rott, H., et al., 2003, Rott, H., et al. 1999, Massonet, D., et al. 1998, Henry, E. et al. 2004, Rott, H. et al. 2000, Berardino, P., et al. 2002, Berardino, P., et al. 2003).

SAR interferometry was first used for topographic mapping by Graham (Graham, L.C, 1974). The first practical results were obtained by Zebker and Goldstein (Zebker, H.A., Goldstein, R.M., 1986) using side looking airborne radar. Studies on interferometric SAR were extended after the launch of ERS-1 and ERS-2, and Envisat

satellites (Feigl, K.L., et al. 2002). The number of scientific research on InSAR technology has also exploded since the launch of new satellites such as of class Sentinel.

In order to obtain SAR interferometric data, two spatially separated antennas, the physical separation of which is called the interferometric baseline, are mounted on a single platform or one antenna is mounted on a satellite and data sets are acquired by passing the same area twice. In the latter case, the interferometric baseline is formed by relating radar signals on repeat passes over the target area. This approach is called repeat-pass interferometry. SAR interferometry is a promising tool for mapping topography, detecting small surface displacement caused by earthquake, volcanic activity, and landslides and ice movement (Weston, J., et al. 2012, Feng, G. et al, 2010). There are various publications on estimating earthquake parameters using InSAR measurements (Gens, R., Van Genderen, J.L., 1996, Reilinger, R.E., 2000, Wright, T.J., et al. 2003, Liu-Zeng, J., et al. 2009, Sudhaus, H., et al. 2009).

The current emphasis in the satellite industry is on replacing large satellite platforms with one or more smaller satellites, i.e. there is increasing interest in the potential capabilities and applications of so-called mini satellites. A multi-satellite system is a group of mini satellites that fly within very close range of each other (e.g., 250 m - 5 km). These satellites coordinate their activities, so that they can use sparse array interferometry and synthetic aperture techniques to simulate a single, very large satellite. The multi-satellite system operates as a “virtual” satellite with a very large effective aperture, without the need for the heavy infrastructure that would be required to have a monolithic satellite with the equivalent aperture. The multi-satellite system approach has many advantages over a single large satellite: each spacecraft is smaller, lighter, simpler, and simpler to manufacture; economies of scale enable a multi-satellite system of many satellites to be less expensive to manufacture than a single satellite; multi-satellite system can adapt to the failure of any individual satellites, and failed satellites can be incrementally replaced; multi-satellite system can reconfigure the orbits of the satellites in the multi-satellite system to optimize for different missions.

In the present work a multi-satellite SAR system is considered. The multi-satellite system consists of mini satellites performing different roles. On one of them a SAR transmitter and receiver are activated, on the other only SAR receivers. Between every pair of SAR satellites an interferometric concept is implemented. It allows the height of each pixel on the surface to be evaluated with high precision and a three dimensional map to be created. The main goal of the work is to describe InSAR geometry with multiple satellites and derive mathematical expressions for definition of current distances between SAR’s and detached pixels, and principal InSAR parameters and, as a result, a model of linear frequency modulated SAR signal with InSAR applications, reflected from the surface to be developed. In order to extract complex SAR images and intergerograms, correlation and spectral SAR image reconstruction algorithms is suggested and implemented.

The rest of the paper is organized as follow. In section II a 3-D InSAR geometry is analytically described. In section III SAR LFM transmitted waveform and model of deterministic SAR signal return are defined. In section IV SAR image reconstruction, co-registration, SAR interferogram generation, and iterative procedure for pixel height determination are analytically described. In section

V results of numerical experiment are presented. In section VI conclusions are made.

2 InSAR GEOMETRY

2.1 Geometry and Kinematics’ Equations

Assume a multi-satellite SAR system viewing three dimensional (3-D) surface presented by discrete resolution elements, pixels. Each of these pixels is defined by the third coordinate $z_{ij}(x_{ij}, y_{ij})$ in 3-D coordinate system $Oxyz$. The surface is illuminated by linear frequency modulated waveforms emitted by a transmitter mounted on one satellite. $A_1, A_2 \dots A_n$ represent SAR receive antennas mounted on satellites viewing the same surface simultaneously. Between every pair A_m, A_n of satellites, where $m \neq n = \overline{1, N}$, N is the number of satellites,

$C_N^2 = \binom{N}{2}$ InSAR baselines can be drawn.

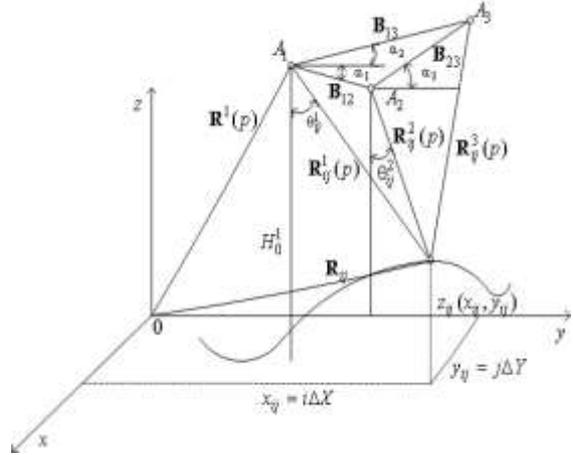


Figure 1: Geometry and kinematics of the interferometry SAR scenario.

The main geometrical characteristic of the SAR signal is the distance vector from the SAR system to each point scatterer of the scene of interest, defined as the vector difference

$$\mathbf{R}_{ij}^n(p) = \mathbf{R}^n(p) - \mathbf{R}_{ij} = [x_{ij}^n(p), y_{ij}^n(p), z_{ij}^n(p)]^T \quad (1)$$

where $n = 1 \div N$ is the number of SAR receiver \mathbf{R}_{ij} is the distance vector of ij th point scatterer from the scene, and $\mathbf{R}^n(p)$ is the distance vector to the n -th SAR receiver, the point of observation in coordinate system $Oxyz$ (Fig. 1); $x_{ij}^n(p)$, $y_{ij}^n(p)$, $z_{ij}^n(p)$ are the current coordinates of the pixel from the scene in respect of the n -th SAR receiver, measured at p th moment of observation and defined as

$$\begin{aligned} x_{ij}^n(p) &= x^n(p) - x_{ij}, \\ y_{ij}^n(p) &= y^n(p) - y_{ij}, \\ z_{ij}^n(p) &= z^n(p) - z_{ij}, \end{aligned} \quad (2)$$

where $x_{ij} = i\Delta X$, $y_{ij} = j\Delta Y$, $z_{ij} = z_{ij}(x_{ij}, y_{ij})$ is the discrete coordinate defining a surface of the scene of interest, $x^n(p)$, $y^n(p)$, $z^n(p)$, are the instant coordinates of the n -th SAR receiver, defined be following equations

$$\begin{aligned} x^n(p) &= x_0^n - V_x p T_p, \\ y^n(p) &= y_0^n - V_y p T_p, \\ z^n(p) &= z_0^n - V_z p T_p, \end{aligned} \quad (3)$$

where x_0^n , y_0^n , z_0^n are the initial coordinates of the n -th SAR, measured at the initial moment, T_p is the time repetition period; p is the current number of the emitted pulse, $\mathbf{V} = [V_x, V_y, V_z]^T$ is the vector velocity of the InSAR multi satellite system, $V_x = V \cos \alpha$, $V_y = V \cos \beta$, $V_z = V \cos \delta$ are the components of vector velocity, $\cos \alpha$, $\cos \beta$, $\cos \delta = \sqrt{1 - \cos^2 \alpha - \cos^2 \beta}$ are the guiding cosines, and V is the module of the vector velocity, \mathbf{V} . Modulus of the current distance vector of the pixel in respect of the n th SAR's receiver antenna is defined by the equation

$$R_{ij}^n(p) = \left\{ [x_{ij}^n(p)]^2 + [y_{ij}^n(p)]^2 + [z_{ij}^n(p)]^2 \right\}^{\frac{1}{2}} \quad (4)$$

The expression (4) can be used to model a SAR signal return from the ij -th point scatterer to the n -th receiver by calculation of the respective time delay and phase of the signal.

2.2 InSAR Geometrical Relations

The relations between two the distances to ij th pixel from m th and n th SAR receivers can be defined from Fig. 1 by application of the cosines theorem, i.e.

$$R_{ij}^n = \left\{ \left(R_{ij}^m \right)^2 + B_{mn}^2 - 2B_{mn}R_{ij}^m \cos \left[\frac{\pi}{2} - [\theta_{ij}^m - \alpha_{mn}] \right] \right\}^{\frac{1}{2}}, \quad (5)$$

where B_{mn} is the modulus of baseline vector, θ_{ij}^m is the look angle, α_{mn} is a priori known tilt angle, the angle between baseline vector and plane xOy . From (5) can be extracted expressions for calculation of the look angle θ_{ij}^m and height h_{ij}^m of a ij -th pint scatter on the surface with respect to m -th receiver

$$\theta_{ij}^m = \alpha_{mn} + \arcsin \frac{[R_{ij}^m]^2 + B_{mn}^2 - [R_{ij}^n]^2}{2B_{mn}R_{ij}^m}, \quad (6)$$

$$z_{ij} = H^m_0 - R_{ij}^m \cos \theta_{ij}^m. \quad (7)$$

The path length from the transmitter A_1 to ij -th pixel on the scene and back to the receiver A_n where ($n = 1 \div N$), is $R_{ij}^{1n} = R_{ij}^1 + R_{ij}^n$. The difference between two the distances R_{ij}^{1n} and R_{ij}^{1m} is equal to $R_{ij}^{1n} - R_{ij}^{1m} = R_{ij}^n - R_{ij}^m$.

Denote the path difference between R_{ij}^{mn} and R_{ij}^n as ΔR_{ij}^{mn} , i.e.

$$\Delta R_{ij}^{mn} = R_{ij}^n - R_{ij}^m, \text{ then}$$

$$R_{ij}^n = R_{ij}^m + \Delta R_{ij}^{mn}.$$

Substitute $R_{ij}^n = R_{ij}^m + \Delta R_{ij}^{mn}$ in (6) and (7), then

$$\theta_{ij}^m = \alpha + \arcsin \left[\frac{B_{mn}}{2R_{ij}^m} - \frac{\Delta R_{ij}^{mn}}{B_{mn}} - \frac{[\Delta R_{ij}^{mn}]^2}{2B_{mn}R_{ij}^m} \right] \quad (8)$$

$$\begin{aligned} z_{ij} &= H^m_0 - R_{ij}^m \times \\ &\times \cos \left\{ \alpha_{mn} + \arcsin \left[\frac{B_{mn}}{2R_{ij}^1} - \frac{\Delta R_{ij}^{mn}}{B_{mn}} - \frac{[\Delta R_{ij}^{mn}]^2}{2B_{mn}R_{ij}^m} \right] \right\} \quad (9) \end{aligned}$$

The path difference ΔR_{ij}^{mn} in equations (8) and (9) can be expressed by the corresponding phase difference $\Delta\phi_{ij}^{mn} = \frac{2\pi}{\lambda}\Delta R_{ij}^{mn}$ as $\Delta R_{ij}^{mn} = \frac{\lambda}{2\pi}\Delta\phi_{ij}^{mn}$. If the current distance R_{ij}^m can be measured expressions (8) and (9) can be written as follows

$$\theta_{ij}^m = \alpha_{mn} + \arcsin \left[\frac{\frac{B_{mn}}{2R_{ij}^m}}{-\frac{\lambda}{2\pi B_{mn}}\Delta\phi_{ij}^{mn} \left(1 + \frac{\lambda}{4\pi R_{ij}^m}\Delta\phi_{ij}^{mn} \right)} \right] \quad (10)$$

$$z_{ij} = H^m_0 - R_{ij}^m \times \cos \left\{ \alpha_{mn} + \arcsin \left[\frac{\frac{B_{mn}}{2R_{ij}^m} - \frac{\lambda}{2\pi B_{mn}}\Delta\phi_{ij}^{mn} \times \left(1 + \frac{\lambda}{4\pi R_{ij}^m}\Delta\phi_{ij}^{mn} \right)}{\left(1 + \frac{\lambda}{4\pi R_{ij}^m}\Delta\phi_{ij}^{mn} \right)} \right] \right\} \quad (11)$$

Using expressions (10) and (11) the elevation look angle θ_{ij}^m and the height z_{ij} of the particular pixel can be calculated.

3 SAR TRANSMITTED LFM WAVEFORM AND DETERMINISTIC SIGNAL MODEL

The SAR transmits a series of electromagnetic waveforms to the surface, which are described analytically by the sequence of linear frequency modulation (LFM) (chirp) pulses as follows

$$S(t) = \sum_{p=1}^M A \exp \left\{ -j \left[\omega(t - pT_p) + b(t - pT_p)^2 \right] \right\}, \quad (12)$$

where A is the amplitude of the transmitted pulses, T_p is the pulse repetition period, $\omega = 2\pi\frac{c}{\lambda}$, is the signal angular frequency, $p = \overline{1, M}$ is the current

number of emitted LFM pulse, M is a full number of emitted pulses during aperture synthesis, $c = 3.10^8$ m/s is the speed of the light, ΔF is the bandwidth of the transmitted pulse that provides the dimension of the range resolution cell, i.e. $\Delta R = c/2\Delta F$, $b = \frac{\pi\Delta F}{T}$ is the LFM rate, T is the time duration of a LFM pulse.

The deterministic component of SAR signal, reflected by ij -th pixel can be determined by expression

$$S_{ij}^n(t) = a_{ij}(z_{ij}) \text{rect} \frac{t - t_{ij}^n}{T} \times \exp \left\{ -j \left[\begin{array}{l} \omega(t - t_{ij}^n) \\ + b(t - t_{ij}^n)^2 \end{array} \right] \right\} \quad (13)$$

where

$$\text{rect} \frac{t - t_{ij}^n(p)}{T} = \begin{cases} 0, & \frac{t - t_{ij}^n(p)}{T} \leq 0 \\ 1, & \frac{t - t_{ij}^n(p)}{T} \leq 1 \\ 0, & \frac{t - t_{ij}^n(p)}{T} > 1 \end{cases}, \quad (14)$$

where $a_{ij}(z_{ij})$ is the reflection coefficient of the pixel from the surface. The parameter $a_{ij}(z_{ij})$ is a function of surface geometry; $t_{ij}^n(p) = \frac{R_{ij}^1(p) + R_{ij}^n(p)}{c}$ is the signal time delay from the ij th point scatterer measured on the n th receiver.

The deterministic components of the SAR signal return are derived by applying the physical optic's principle of Huggens-Fresnel, according to which the SAR signal return can be calculated as a sum of elementary signals reflected by point scatterers from the surface, i.e. the time record of data $S^n(p, k)$ can be written as

$$S^n(t) = \sum_i \sum_j a_{ij}(z_{ij}) \text{rect} \frac{t - t_{ij}^n}{T} \exp \left\{ -j \left[\begin{array}{l} \omega(t - t_{ij}^n) \\ + b(t - t_{ij}^n)^2 \end{array} \right] \right\} \quad (15)$$

The time dwell t of the SAR signal return for each transmitted pulse p can be expressed as

$t = (k - k_{ij \min})\Delta T$, where $k = k_{ij \min}(p), k_{\max}^n(p)$ is the sample number of the SAR return, $k_{ij \min} = \text{int}[t_{ij \min}(m)/\Delta T]$, $\Delta T = 1/2\Delta F$ is the sample time duration of the return, ΔF is frequency bandwidth, $k_{\max}^n(p)$ is the number of the last range bin where SAR return is registered in n -th receiver for each emitted pulse. Hence, SAR return registered in n -th receiver in discrete form can be written as

$$\hat{S}^n(p, k) = \sum_i \sum_j a_{ij}(z_{ij}) \text{rect} \frac{t - t_{ij}^n}{T} \times \exp \left\{ -j \left[\omega \left((k-1)\Delta T - t_{ij}^n(p) \right) + b \left((k-1)\Delta T - t_{ij}^n(p) \right)^2 \right] \right\} \quad (16)$$

The expressions (1) - (16) can be used for modeling the SAR signal return in case the satellites are moving on rectilinear trajectory in 3-D coordinate system.

4 SAR IMAGE RECONSTRUCTION, COREGISTRATION AND SAR INTERFEROGRAM GENERATION

4.1 SAR Image Reconstruction Algorithm

Image reconstruction is constituted by following operations: frequency demodulation; range compression; coarse range alignment and precise phase correction, and azimuth compression. The result is a complex image of the scene.

In each SAR receiver a frequency demodulation is carried out. It is performed by multiplication of the right term of equation (10) with a complex conjugate exponential function

$$\exp \left\{ j \left(\omega(k-1)\Delta T + b[(k-1)\Delta T]^2 \right) \right\}.$$

Thus, the range distributed frequency demodulated SAR return in n -th receiver for p -th pulse can be written as

$$\hat{S}^n(p, k) = \sum_i \sum_j a_{ij}(z_{ij}) \text{rect} \frac{(k-1)\Delta T - t_{ij}^n}{T} \times \exp \left\{ -j \left[\omega t_{ij}^n(p) + b \left((k-1)\Delta T - t_{ij}^n(p) \right)^2 \right] \right\} \quad (17)$$

SAR data refer to a set of data that has a real (Cosine, In-phase) and imaginary (Sine, Quadrature) component. Both this components of backscattered signals are measured by each SAR receiver. Therefore, the SAR data stream from each receiver can be writes as

$$\hat{S}^n(p, k) = I^n(p, k) + jQ^n(p, k) \quad (18)$$

where

$I^n(p, k) = \text{Re}[\hat{S}^n(p, k)]$ is the In-phase part of the SAR signal in the n th receiver,

$Q^n(p, k) = \text{Im}[\hat{S}^n(p, k)]$ is the Quadrature part of the SAR signal in the n th receiver,

Range compression is implemented by cross correlation of the complex SAR signal and reference function, $\exp \left\{ j b [(k-1)\Delta T]^2 \right\}$ or by Fourier transform of LFM demodulated SAR signal, using the expressions as follows.

- cross correlation

$$\hat{S}_R^n(p, \hat{k}) = \sum_{k=1}^K \hat{S}^n(p, k) \exp \left\{ j b [(k - \hat{k} - 1)\Delta T]^2 \right\} \quad (19)$$

where K is the full number of LFM samples, range bins where SAR signal is registered;

- Fourier transform

$$\hat{S}_R^n(p, \hat{k}) = \sum_{k=1}^K \hat{S}^n(p, k) \cdot \exp \left(j \frac{2\pi k \hat{k}}{K_{\max}^n} \right), \quad (20)$$

for each $p = \overline{1, M}$ and $\hat{k} = \overline{1, K}$.

Range alignment and phase correction is implemented by well known range alignment and focusing procedures [Lazarov, A. 2002].

Azimuth compression is implemented by Fourier transform of the range compressed signal, $\hat{S}_R^n(p, \hat{k})$. It yields a complex image based on the n -th receiver data and can be expressed as

$$\hat{I}^n(\hat{k}, \hat{p}) = \sum_{p=1}^M \hat{S}_R^n(p, \hat{k}) \exp \left(j \frac{2\pi p \hat{p}}{M} \right), \quad (21)$$

for each $\hat{p} = \overline{1, M}$, $\hat{k} = \overline{1, K}$.

It is worth noting that the complex pixel of the SAR image in n -th receiver preserves the phase information defined by path length from the transmitter to the ij -th pixel and back to the n -th SAR receiver. Based on the pixels phase information and precise co-registration of two complex images a complex interferogram can be generated.

4.2 Co-registration of Two SAR Complex Images

To generate an interferogram between two complex images first a precision under pixel co-registration for any pair of two complex images has to carry out (Guizar-Sicairos, M., 2008).

Let $I^n(\hat{k}, \hat{p})$ and $I^m(\hat{k}, \hat{p})$ be two complex images obtained by the n th and m th receiver, respectively, and let $i^n(q, r)$ and $i^m(q, r)$ be two complex spectrums of the images defined by the expressions

$$i^m(q, r) = \sum_{\hat{p}=1}^M \sum_{\hat{k}=1}^K I^m(\hat{k}, \hat{p}) \exp \left[-j \left(\frac{2\pi\hat{k}q}{K} + \frac{2\pi\hat{p}r}{M} \right) \right],$$

$$i^n(q, r) = \sum_{\hat{p}=1}^M \sum_{\hat{k}=1}^K I^n(\hat{k}, \hat{p}) \exp \left[-j \left(\frac{2\pi\hat{k}q}{K} + \frac{2\pi\hat{p}r}{M} \right) \right].$$

In case the image $I^m(\hat{k}, \hat{p})$ is displaced in respect of the image $I^n(\hat{k}, \hat{p})$ on intervals $\Delta\hat{k}$ and $\Delta\hat{p}$ it can be written

- in the space domain

$$I^m(\hat{k}, \hat{p}) = I^n(\hat{k} + \Delta\hat{k}, \hat{p} + \Delta\hat{p})$$

- in the frequency domain

$$i^m(q, r) = i^n(q, r) \exp \left[j \left(\frac{2\pi\Delta\hat{k}.q}{K} + \frac{2\pi\Delta\hat{p}.r}{M} \right) \right].$$

The level of coincidence of the complex images can be calculated in space and/or frequency (spectral) domain by cross-correlation and inverse Fourier transform of the multiplication of the two complex spectrums as follows

- in the space domain

$$C(\Delta\hat{k}, \Delta\hat{p}) = \sum_{\hat{p}=1}^M \sum_{\hat{k}=1}^K I^n(\hat{k}, \hat{p}) \cdot I^m(\hat{k} + \Delta\hat{k}, \hat{p} + \Delta\hat{p}),$$

where $\Delta\hat{k}$ varies from 0 to the pixel's dimension with step $1/10$ (pixel's dimension on \hat{k} axis), $\Delta\hat{p}$ varies from 0 to the pixel's dimension with step $1/10$ (pixel's dimension on \hat{p} axis);

- in the frequency domain

$$C(\Delta\hat{k}, \Delta\hat{p}) = \sum_{r=1}^M \sum_{q=1}^K i^m(q, r) i^n(q, r) \times \exp \left[j \left(\frac{2\pi(\hat{k} + \Delta\hat{k})q}{K} + \frac{2\pi(\hat{p} + \Delta\hat{p})r}{M} \right) \right].$$

Maximum of the correlation function corresponds to a maximum coincidence, i.e. the complex images are best co-registered.

After SAR image reconstructions complex interferograms can be generated by pixel-wise multiplication of the two complex SAR images. For example, the complex interferogram between m -th and n -th SAR complex image can be calculated by the expression

$$i^{mn}(\hat{k}, \hat{p}) = i^n(\hat{k}, \hat{p}) \otimes [i^m(\hat{k}, \hat{p})]^*, \quad (22)$$

where \otimes denotes elementwise product; the sign "*" denotes complex conjugate.

4.3 Iterative Procedure for Pixel Height Determination

Step $k = 0$

1. Consider ij -th pixel with coordinates $x_{ij}, y_{ij}, z_{ij} = 0$ (the pixel is on the base plane xOy).
2. Compute the distance R_{ij}^m

$$R_{ij}^m(0) = \left[\begin{aligned} & (x_0^m - x_{ij})^2 + (y_0^m - y_{ij})^2 \\ & + (z_0^m - z_{ij})^2 \end{aligned} \right]^{\frac{1}{2}},$$

where x_0^m, y_0^m and z_0^m are coordinates of m -th SAR at the moment $p=N/2$ (moment of image extraction).

Step $k = k$

3. Compute the distance R_{ij}^n

$$R_{ij}^n(k) = \left[\begin{aligned} & (x_0^n - x_{ij})^2 + (y_0^n - y_{ij})^2 \\ & + (z_0^n - z_{ij})^2 \end{aligned} \right]^{\frac{1}{2}},$$

where x_0^n, y_0^n and z_0^n are coordinates of n -th SAR at the moment $p=N/2$ (moment of image extraction).

4. On k -th step compute the interferometric phase:

$$\Delta\varphi_{ij}^{mn}(k) = \frac{2\pi}{\lambda} \Delta R_{ij}^{mn},$$

where $\Delta R_{ij}^{mn} = R_{ij}^m - R_{ij}^n$.

5. If $\Delta\varphi_{ij}^{mn}(k) < \Delta\varphi_{ij}^{mn}$, then

Step $k = k + 1$.

6. On $(k + 1)$ -th step consider a pixel with coordinates $x_{ij}, y_{ij}, z_{ij}(k+1) = z_{ij}(k) + \Delta z$.

7. Go to point 2 and 3.

8. If $\Delta\varphi_{ij}^{mn}(k) > \Delta\varphi_{ij}^{mn}$, then consider the pixel with coordinates

$$x_{ij}, y_{ij}, z_{ij}(k+1) = z_{ij}(k) + (\Delta z / r),$$

where $r = 2, 3, 4, \dots$

9. Go to point 2 and 3.

10. If $\Delta\varphi_{ij}^{mn}(k) = \Delta\varphi_{ij}^{mn}$, then $z_{ij}(k) = h_{ij}$.

Stop.

The height's estimation procedure can be applied either for all pixels or only for one pixel from the surface of interest. In the latter case a scale coefficient can be defined in order to transform unwrapped phase surface to topographic map.

5 NUMERICAL EXPERIMENT

To verify the correctness of the signal model and image reconstruction and co-registration algorithms numerical experiment is carried out.

Assume multi satellite InSAR system comprising three satellites, one with transmitter and receiver, two – only with receivers, with initial space coordinates as follows:

$$\begin{aligned} x_0^1 &= 0 \text{ m}; y_0^1 = 10 \cdot 10^3 \text{ m}, z_0^1 = 100 \cdot 10^3 \text{ m}, \\ x_0^2 &= 0 \text{ m}, y_0^2 = 10,1 \cdot 10^3 \text{ m}, z_0^2 = 100 \cdot 10^3 \text{ m}, \\ x_0^3 &= 0 \text{ m}, y_0^3 = 10,2 \cdot 10^3 \text{ m}, z_0^3 = 100 \cdot 10^3 \text{ m}, \end{aligned}$$

and coordinates of vector-velocity of the three satellites: $v_x = 0 \text{ m/s}; v_y = -600 \text{ m/s}; v_z = 0 \text{ m/s}$.

Multi satellite SAR system observes a surface (Fig. 2) depicted by equation

$$\begin{aligned} z_{ij} &= 3(1 - x_{ij})^2 \exp[-x_{ij}^2 - (y_{ij} + 1)^2] - \\ &- 10 \left(\frac{x_{ij}}{5} - x_{ij}^3 - y_{ij}^5 \right) \exp(-x_{ij}^2 - y_{ij}^2) - \\ &- \frac{1}{3} \exp[-(x_{ij} + 1)^2 - y_{ij}^2] \end{aligned}$$

where $x_{ij} = i\Delta X, y_{ij} = j\Delta Y, i = \overline{1, I}, j = \overline{1, J}, I = 128 \text{ pixels}; J = 128 \text{ pixels}; \Delta X; \Delta Y$ - the spatial dimensions of the pixels.

Normalized amplitude of reflected signals from every pixel $a_{ij} = 0.001$. Dimensions of the pixel are $\Delta X = \Delta Y = 2 \text{ m}$. Wavelength is 0.03 m . Carrier frequency is 3.10^9 Hz . Frequency bandwidth is $\Delta F = 250 \text{ MHz}$. Pulse repetition period is $T_p = 25 \cdot 10^{-3} \text{ s}$. LFM pulse duration is $T = 5 \cdot 10^{-6} \text{ s}$.

Sample time duration is $\Delta T = 1,95 \cdot 10^{-8} \text{ s}$. LFM samples are $K = 512$. Emitted pulses are $M = 512$. Digital geometry description and SAR signal modeling are performed based on the theory in sections III and IV. Complex images through correlation range compression and FFT azimuth compression are retrieved. Based on a priory known kinematical parameters of satellites and coordinates of reference point from the surface autofocusing phase correction of the SAR signals registered in the both receivers can be implemented.

The real and imaginary part of the complex signal measured in the first receiver is depicted in Fig. 2.

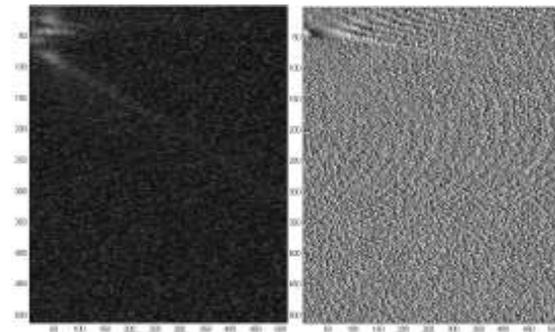


Figure 2: Real and imaginary part of the complex signal measured in the first receiver.

The amplitude and phase of the reconstructed complex image obtained in the first receiver is depicted in Fig. 3. The position of the surface's image in the frame is defined by the position of the receiver satellite in the point of imaging of the

surface, as a rule, this point is in the middle of the synthetic aperture length.

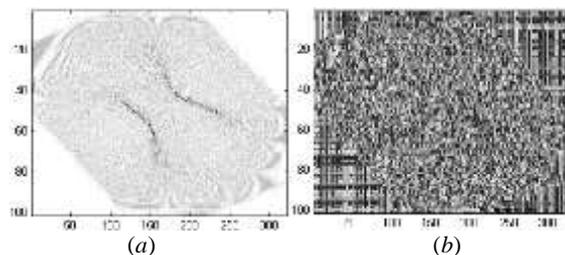


Figure 3: The amplitude and phase of the reconstructed complex image obtained in the first receiver.

The real and imaginary part of the complex signal measured in the second receiver is depicted in Fig. 4.

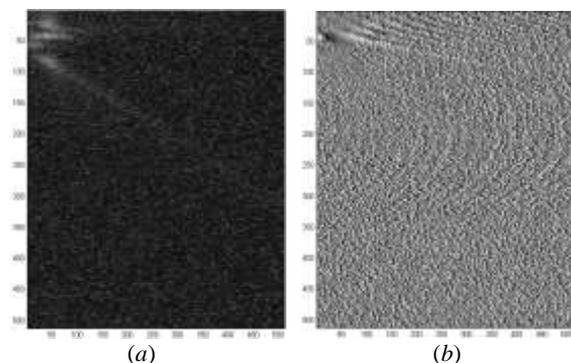


Figure 4: Real (a) and imaginary (b) part of the complex signal measured in the second receiver.

The amplitude and phase of the reconstructed complex image obtained in the second receiver is depicted in Fig. 5. It can be noticed the shape of the surface is similar to the shape of the image obtained by the first receiver. In contrast, the phase pictures are different based on the different positions of both satellites in respect of the surface.

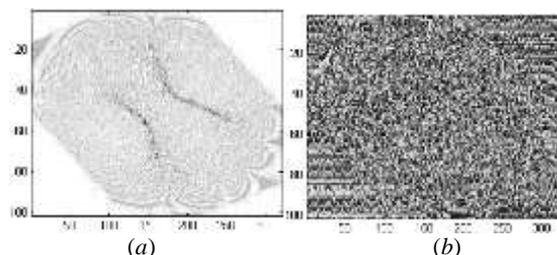


Figure 5: The amplitude and phase of the reconstructed complex image obtained in the second receiver.

After co-registration of complex images obtained in the first and second receiver and complex interferogram generation the result as a coherent map (a) and interferometric phase (b) is depicted in Fig. 6.

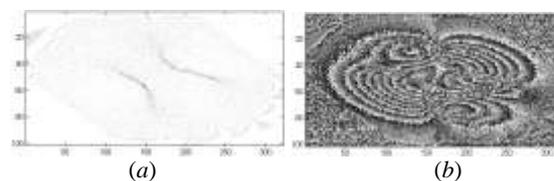


Figure 6: Complex interferogram: coherent map and interferometric phase generated by the first and second complex images.

The real and imaginary part of the complex signal measured in the third receiver is depicted in Fig. 7.

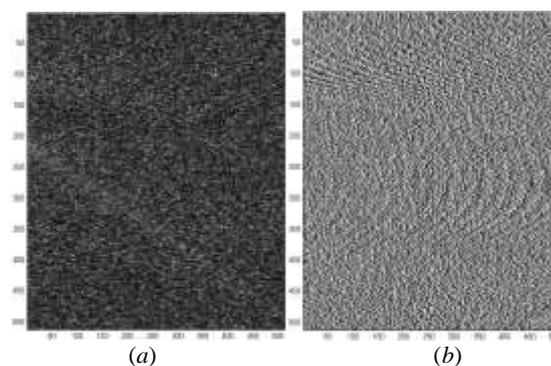


Figure 7: Real and imaginary part of the complex signal measured in the third receiver.

The amplitude and phase of the reconstructed SAR complex image obtained in the third receiver is depicted in Fig. 8. It can be noticed the shape of the surface is similar to the shape of the image obtained by the first receiver. In contrast, the phase pictures are different based on the different positions of both satellites in respect of the surface.

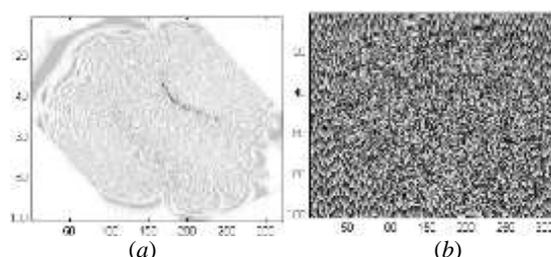


Figure 8: The amplitude and phase of the reconstructed complex image obtained in the third receiver.

After co-registration of complex images calculated by the data obtained in the first and third receiver, and complex interferogram generation, the result as a coherent map (a) and interferometric phase (b) is depicted in Fig. 9.

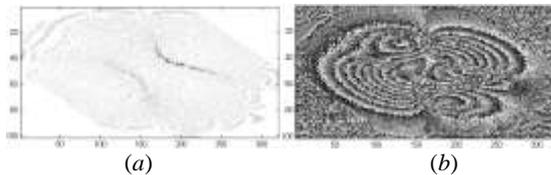


Figure 9: Complex interferogram: coherent map and interferometric phase generated by the first and third complex images.

Comparing the interferometric phases presented in Fig. 6 and Fig. 9 it can be concluded that they are similar, i.e. very close to each other. It is a result of precise under pixel co-registrations of the first and second, and the first and third complex images.

6 CONCLUSIONS

In the paper a multi-satellite InSAR system has been analysed and numerically experimented. The geometry of InSAR scenario and kinematics of multiple SAR satellites has been analytically described. Mathematical expressions for current distances between SAR satellites and surface point scatterers are derived and principal InSAR parameters are defined. A model of linear frequency modulated SAR signal with InSAR applications, reflected from the topographic surface has been developed. Correlation and spectral SAR image reconstruction algorithms, co-registration, and iterative pixel height determination procedures have been described. Based on geometrical and kinematical models numerical interferograms of a topographic surface have been created.

REFERENCES

- Rott, H., Nagler, T., Rocca, F., et al., 2003. InSAR techniques and applications for monitoring landslides and subsidence, in: Benes (Ed.), *Geoinformation for European-wide integration*, In Proceedings of *EARS&L Assembly, Prague*, June 2002. Millpress, Rotterdam, 25–31.
- Massonet, D., Feigl, K.L., 1998. Radar interferometry and its application to changes in the Earth's surface. *Rev. Geophys.*, 36, 441–500, 1998.
- Henry, E., Mayer, C., Rott, H., 2004. Mapping mining-induced subsidence from space in a hard rock mine: example of SAR interferometry application at Kiruna mine. *CIM Bull.* 97 (1083), 1–5.
- Rott, H., Scheuchl, B., Siegel, A., et al., 1999. Monitoring very slow slope movements by means of SAR interferometry: a case study from a mass waste above a reservoir in the Ötztal Alps, Austria. *Geophys. Res. Lett.* 26, 1629–1632.
- Rott, H., Mayer, C., Siegel, A., 2000. On the operational potential of SAR interferometry for monitoring mass movements in Alpine areas, In Proceedings of *3rd European Conference on Synthetic Aperture Radar (EUSAR2000)*, Munich, May 2000, 43–46.
- Berardino, P., M. Costantini, G. Franceschetti, A. Iodice, Pietrnera L. and Rizzo, V., 2002. Differential SAR interferometry for the study of slope instability at Maratea, Italy, in Proceedings of *International Geoscience and Remote Sensing Symposium*, Toronto, Canada, 2693–2695.
- Berardino, P., M. Costantini, G. Franceschetti, A. Iodice, Pietrnera L. and Rizzo, V., 2003. Use of differential SAR interferometry in monitoring and modelling large slope instability at Maratea (Basilicata, Italy), *Eng. Geol.*, vol. 68, 2003, 31–51.
- Graham, L.C., 1974. Synthetic Interferometer Radar for Topographic Mapping, *Proceedings of the IEEE*, vol.62, no.6, 1974. 763.
- Zebker, H.A., Goldstein, R.M., 1986. Topographic Mapping from Interferometric Synthetic Aperture Radar Observations, *Journal of Geophys.Res.*, vol.91, no. B5, 1986, 4993–4999.
- Feigl, K.L., Sarti, F., Vadon, H. Durand, P., McClusky S. et al., 2002. Estimating slip distribution for the Izmit mainshock from coseismic GPS, ERS-1, RADARSAT and SPOT measurements, *Bull. Seism. Soc. Amer.*
- Gens, R., Van Genderen, J.L., 1996. Review article: SAR interferometry-issues, techniques, applications. *International Journal for Remote Sensing*, 17(10), 1803–1835.
- Reilinger, R.E., Ergintav, S., Bürgmann, R., McClusky, S., Lenk, O., et al., 2000. Coseismic and postseismic fault slip for the 17 August 1999, M=7.4, Izmit, Turkey earthquake, *Science*, 289, 1519–1524.
- Wright, T.J., Fielding, E.J., Parsons, B.E., and England, P.C., 1999. Triggered slip: observations of the 17 August 1999 Izmit (Turkey) earthquake using radar interferometry, *Geophys. Res. Lett.*, 28, , 2001, 1079–1082.
- Weston, J., Ferreira, A.M.G., Funning, G. J., 2012. Systematic comparisons of earthquake source models determined using InSAR and seismic data, *Tectonophysics* 532–535, 61–81.
- Feng, G., Hetland, E., Ding, X.L., Li, Z., Zhang, L., 2010. Coseismic fault slip of the 2008 Mw 7.9 Wenchuan earthquake estimated from InSAR and GPS measurements, *Geophysical Research Letters*, 37, L01302.

- Liu-Zeng, J., Zhang, Z., Wen, L., Tapponnier, P., Sun, J., Xing, X., Hu, G., Xu, Q., Zeng, L., Ding, L., Ji, C., Hudnut, K.W., van der Woerd, J., 2009. Co-seismic rupture of the 12 May 2008, Ms 8.0 Wenchuan earthquake, Sichuan: east-west crustal shortening on oblique, parallel thrusts along the eastern edge of Tibet. *Earth and Planetary Science Letters* 286, 355–370.
- Sudhaus, H., Jonsson, S., 2009. Improved source modelling through combined use of InSAR and GPS under consideration of correlated data errors: application to the June 2000 Kleifarvatn earthquake Iceland. *Geophysical Journal International* 176, 389–404.
- Lazarov, A.D., Minchev, Ch., 2002. Correlation-autofocusing-spectral 2-D ISAR Image Reconstruction from Linear Frequency Modulated Signals, *21st Digital Avionics Systems Conference (DASC)*, Irvine, California, October 2002.
- Guizar-Sicairos, M., Thurman, S.T., Fienup, J. R., 2008. Efficient subpixel image registration algorithms, *Opt. Lett.*, vol. 33, 156-158.

Bathymetric Mapping of Shallow Rivers with UAV Hyperspectral Data

Valeria Gentile^{1,2}, Marek Mróz³, Marie Spitoni², Jérôme Lejot⁴, Hervé Piégay² and Luca Demarchi⁵

¹Dept. of Information, Electronics & Telecom. Eng. (DIET), Univ. of Rome "Sapienza", via Eudossiana 18, Rome, Italy

²Université de Lyon, UMR 5600 EVS, Ecole Normale Supérieure, Lyon Cedex 07, France

³Dept. of Photogrammetry & Remote Sensing, Univ. of Warmia and Mazury in Olsztyn, ul. Oczapowskiego, Olsztyn, Poland

⁴Université de Lyon, UMR 5600 EVS, Université Lumière Lyon 2, Campus Porte des Alpes, Bron Cedex, France

⁵Warsaw University of Life Sciences Nowoursynowska 166, 02-787 Warsaw, Poland
valeria.gentile@uniroma1.it, {valeria.gentile, marie.spitoni, herve.piegay}@ens-lyon.fr,
marek.mroz@uwm.edu.pl, jerome.lejot@univ-lyon2.fr, demarchi.luca.ld@gmail.com

Keywords: Bathymetry, High spatial resolution, Hyperspectral images, Fluvial morphology, Remote sensing.

Abstract: Airborne images have long been used to support environmental monitoring due to their synoptic capability to cover wide areas with high spatial and temporal resolution. The potential for bathymetric mapping by airborne remote sensing has been addressed and demonstrated in several studies by means of imaging and non-imaging techniques. In this paper we evaluate the potential to retrieve water depth of shallow river from high resolution hyperspectral images using an empirical model, applicable under a range of specific field conditions and in a definite interval of wavelengths.

1 INTRODUCTION

The necessity to preserve water resources and ecosystems has led to an increasing interest in monitoring the morphological status of water bodies. Through a constant data collection on the long term, it is possible to determine trends in monitored parameters and to decide suitable strategies in order to prevent river channel degradation or to restore its original status.

Remote Sensing (RS) has long been used to support environmental monitoring of fluvial environments due to its synoptic capability to cover wide areas with high spatial and temporal resolution and to detect features that are not rapidly and easily evaluable with *in situ* measurements. Remote sensing techniques have been also widely applied to assess bathymetry of water body (Carbonneau, Lane and Bergeron, 2006, Fonstad and Marcus, 2005, Lane, Westaway and Murray Hicks, 2003), being the only effective alternative to measurements collected by echo sounder mounted on boat, in very shallow and braided rivers, impossible to be entirely

navigated. Furthermore ground surveys are extremely time-consuming, require a consistent deployment of manpower and provide a low spatial sampling of acquired data despite to their accuracy.

As reviewed by Feurer, Bailly, Puech, Le Coarer and Viau (2008), besides echo sounder and GPR (Ground Penetrating Radar), both requiring ground surveys, three remote sensing approaches exist for mapping water depth through imaging and non imaging techniques (Gao, 2009). These are spectral methods, photogrammetry and bathymetric LIDAR (Light Detection and Ranging). Spectral methods exploit the attenuation of electromagnetic wave through the water interface in order to derive water depth. Their capability for mapping bathymetry has been addressed in several studies, using data acquired in the visible spectrum from UAV platforms (Lejot *et al.*, 2007, Feurer *et al.*, 2008) or Airborne Thematic Mapper data simulated from ground based measurements collected through spectroradiometer (Gilvear, Hunter and Higgings 2007) or AISA (Airborne Imaging Spectrometer for Applications) data (Legleiter, Roberts and Lawrence, 2009).

In our study we evaluate the potential to retrieve water depth of shallow river from very high resolution hyperspectral images, using a simple empirical model, applicable under a range of specific field conditions. In doing so, we take advantages first from the capability of UAV platforms to acquire very high resolution images, combined with the possibility given by hyperspectral sensors to investigate the model behaviour over a wide range of wavelengths in addition to the visible spectrum.

In our paper we provide a detailed description of the study site in section 2. In section 3 is explained the overall process to obtain the final results: the acquisition campaign with used sensors and platform is described in subsection 3.1, in subsection 3.2 the entire processing chain to derive the final orthomosaics from raw hyperspectral data cubes is illustrated, in subsection 3.3 the physical model adopted to derive the relation between the radiometric pixel value and the water depth is explained. In subsection 3.4 processing applied to the final orthomosaics to establish the goodness of the relation previously described is explained. Results and conclusion follow in sections 4 and 5 respectively.

2 STUDY SITE

The acquisition campaign was carried out along a channel reach of the Ain River in the south-east part of France. The Ain River drains a watershed surface of about 3700 km² along 200 km. It rises in the Jura Mountains, then it flows through a steep mountainous relief, before reaching its Lower Valley (Liébault and Piégay, 2002). The river in its Lower Valley flows through 50 km, in alluvial deposits (Bravard, 1986) where it is free to laterally and vertically adjust. Its depth ranges between 0 to 5 m (Lejot *et al.*, 2007). Its hydrology is dominated by snowmelt mixed with rainfall. The mean annual discharge is 120 m³s⁻¹, ranging between 17 m³s⁻¹ to 1600 m³s⁻¹ (1-in-50 year flood) at Pont d'Ain and Chazey-sur-Ain gauging stations according to the *banque HYDRO* (<http://www.hydro.eaufrance.fr/>). A chain of 5 main hydroelectric dams were built until the 70's in its middle V-shape valley section. These dams have undergone important changes in the Lower Valley, *e.g.* reduction of peak flows and channel narrowing or degradation (Liébault and Piégay, 2002).

The study site (Figure 1), approximately 700 m long, between Pont d'Ain and Priay, is located in the

Ain Lower Valley, northeast of the city of Lyon. It was chosen because of its fairly morphological and channel stability (paved riverbed and low lateral mobility). Due to the lack of *in situ* water depth data synchronous with imagery data, we used the simulated hydrological parameters from the numerical model developed by Paquier, Camenen, Le Coz and Béraud (2014). This model runs over ADCP and GPS cross-sectional surveys performed in 2013 and 2014 (Naudet, Le Coz, Camenen and Paquier, 2015). Riverbed changes were assumed to be negligible on the study reach since the last 3 years (the mean absolute error for the modelled 2013 water level elevation is 15 cm).

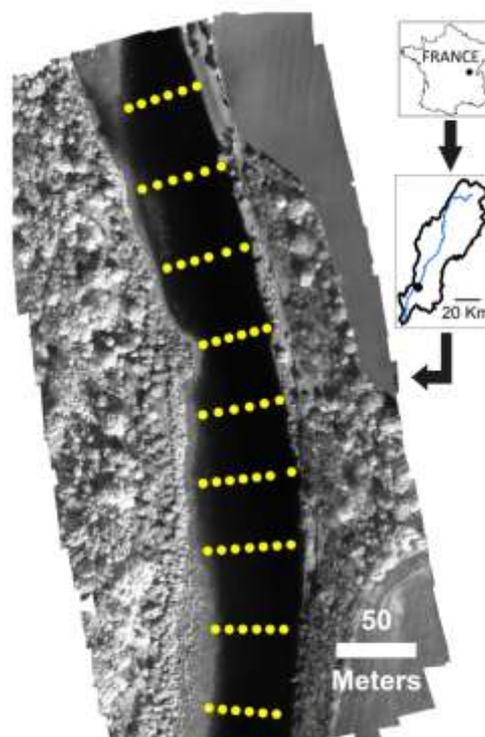


Figure 1: Orthomosaic at the central wavelength $\lambda=776\text{nm}$ and geographical location of the study site; in yellow sampling points of 2D hydraulic model.

3 METHODS

3.1 Data Collection

The study area was imaged twice on 28th September 2015 in the interval 12h00-12h54 (CEST) using two coupled cameras mounted on the UAV md4-1000 quadcopter (table 1):

- digital RGB OLYMPUS EP-2 camera

- Rikola 2D spectral sensor (Makelainen, Saari, Hippo, Sarkeala and Soukkamaki, 2013).

The Rikola 2D imaging system is a VNIR sensor based on the Piezo-Actuated Fabry-Perot Interferometer (FPI), working in the spectral range 500 nm - 900 nm. This allows the user to select the central wavelengths of the bands to be recorded by setting up the appropriate “air gaps” in the interferometer. The CCD/CMOS matrix consists of 1024x1024 detectors. Each sensing element has the size of 5.5x5.5 μm . The camera is characterized by FOV=37°, focal length f=9mm and F-number=2.8. The ADC (analog-to-digital converter) is operating in 12 bit mode. The system is equipped with GPS receiver and hemispherical irradiance sensor. The described camera model and software version permit to acquire 16 bands in full-frame mode or 24 bands in the half-frame mode (1024x648 pixels) for one “hypercube” (single frame). The user can also choose one of the two FWHMs (full width at half maximum): narrow or wide. The precise values of the FWHM for each band are determined by the interferometer itself.

Table 1: Set of spectral bands recorded in the experiment.

First flight – Full frame mode			Second flight – half frame mode		
Band no.	Central wavelength [nm]	FWHM [nm]	Band no.	Central wavelength [nm]	FWHM [nm]
1	500	15	1	500	13
2	523	20	2	516	13
3	546	19	3	532	11
4	569	18	4	548	10
5	591	18	5	564	11
6	614	19	6	580	13
7	643	13	7	596	15
8	661	19	8	612	14
9	684	17	9	628	18
10	707	18	10	644	13
11	730	18	12	676	13
12	753	17	13	692	11
13	776	18	14	708	11
14	799	17	15	724	12
15	822	19	16	740	12
16	845	18	-	-	-

The pictures were taken at the altitude MHOG=100 m (mean height over ground) forming regular blocks of strips with the end lap of 70% and side lap of 30% for hyperspectral images. The mean ground resolution of Rikola images was about 6 cm.

The Rikola camera takes the hypercubes with a constant time interval Δt which was set in our experiment at 5 s.

For each block of RGB images the overlapping was bigger (80% and 40% respectively) and the ground sampling distance (GSD) was about 2 cm. The number of acquired hypercubes was bigger than 100 for each of two flights, and about 66 of RGB pictures.

3.2 Data Pre-processing

Acquired RGB images underwent orthorectification process with Agisoft Photoscan Professional software. The process consisted of digital aerotriangulation, image matching, 3D cloud point and Digital Surface Model generation and the final ortho-correction. The final RGB orthophotomap had pixel size 5x5 cm and it was considered as a background supporting the geometric processing of acquired hyperspectral data.

The exposition time for Rikola camera is usually set between 10 and 25 ms depending on sunlight intensity. In our experiment the exposition was set at 15 ms. Such a value is suitable for taking non-blurred pictures from moving platform but the technology of image formation and recording on the memory card leads to the situation where every band of the given hypercube has a slightly different position and external orientation. In these circumstances there are two alternative ways for further geometric processing:

- to adjust all bands of the hypercube to a common frame first and to produce in the next step all spectral orthomosaics in one photogrammetric run;
- to split all bands of each hypercube and to process all frames taken at the same wavelength in separated photogrammetric runs forming a set of independent spectral mosaics.

We adopted the second way because the automatic geometric adjustment/matching of the bands taken in visible and infrared spectrum is very problematic for the scenes without structural points. Therefore hyperspectral frames were processed similarly like RGB photos giving as a result a set of monochromatic orthophotomaps at the resolution of 10 cm with, unfortunately, slightly different georeferencing. The last step in geometric processing was the adjustment of all spectral orthophotomaps to the common frame by affine transformation based on RGB orthophotomap.

Prior to the orthorectification at each spectral band, the radiometric processing needed to be performed. The first step was the radiometric calibration of each hypercube to remove the influence of the black current from measured signals. The second step was the radiometric normalization *i.e.* comparison of the recorded spectral luminance for each band with the luminance of the white standardized target. In our case the Zenith Lite™ panel 50x50 cm covered by BaSO₄-based white paint was used. No other atmospheric or radiometric corrections were applied. Some spectral bands from second Rikola dataset taken in half-frame mode were eliminated due to the encountered errors in pictures recording.

3.3 Bathymetric Model

The spectral radiance observed at the remote sensor detector $L_T(\lambda)$ for any wavelength λ is expressed as the sum of four components (Legleiter *et al.*, 2009):

$$L_T(\lambda) = L_B(\lambda) + L_C(\lambda) + L_S(\lambda) + L_P(\lambda) \quad (1)$$

where $L_B(\lambda)$ is the radiance reflected from bottom, $L_C(\lambda)$ is the radiance from water column, $L_S(\lambda)$ is the radiance reflected from water surface and $L_P(\lambda)$ is the path radiance from the atmosphere. Under the conditions of homogeneous water properties, shallow river, opportune viewing geometry, low acquisition altitude, favourable atmospheric conditions, highly reflective and homogeneous streambed and relatively clear water, we can consider negligible the radiance components $L_C(\lambda)$, $L_S(\lambda)$, $L_P(\lambda)$ (Legleiter *et al.*, 2009):

$$L_T(\lambda) \approx L_B(\lambda) \quad (2)$$

where $L_B(\lambda)$ is (Philpot, 1989, Legleiter *et al.*, 2009):

$$L_B(\lambda) = E_d(\lambda)C(\lambda)T(\lambda)(R_b(\lambda) - R_c(\lambda))\exp(-k(\lambda)d) \quad (3)$$

$E_d(\lambda)$ is the downwelling solar irradiance, $C(\lambda)$ is a constant for transmission across air water interface, $T(\lambda)$ is the transmittance of atmosphere, $R_b(\lambda)$ is the reflectance of river bottom, $R_c(\lambda)$ is the volume reflectance of water column, $k(\lambda)$ is an attenuation coefficient that accounts for absorption and scattering of light within the water column (Maritorena, Morel, Gentili, 1994, Legleiter *et al.*, 2009), d is the water depth. Solving with respect to water depth, we obtain:

$$\ln(L_B) = \ln(E_d C T (R_b - R_c)) - kd \quad (4)$$

where we have not considered the dependence on λ to simplify the notation. The relation (4) suggests that under the above-mentioned acquisition conditions and for certain wavelengths, a relation between the remotely sensed variable L_B and the water depth can be derived and used for mapping river bathymetry. Replacing L_B with the corresponding value of digital number registered by the sensor and opportunely calibrated, after several adjustments we can rewrite (4) as a linear relation between the natural logarithm of pixel values in the image and the corresponding values of water depth:

$$d_{i,j} = a_{0,k} + a_{1,k} \ln P_{i,j,k} \quad (5)$$

where $d_{i,j}$ is the water depth in correspondence of pixel i,j in the image, $a_{0,k}$ and $a_{1,k}$ are the coefficients of linear relation related to k -th spectral band and $P_{i,j,k}$ is the value of pixel i,j at k -th spectral band.

3.4 Data Processing

Before deriving coefficients $a_{0,k}$ and $a_{1,k}$ of linear relation (5) for each orthomosaic, through a linear regression, a median filter with a window of 5x5 was applied to remove residual noise after images pre-processing.

For each spectral band, the pixel values were extrapolated from orthomosaics, in correspondence of the geographical coordinates of bathymetric values given by the numerical model of Paquier *et al.* (2014) applied to the Ain River (Naudet, Camenen, Le Coz, Paquier and Piégay, 2014). This 2D hydraulic model provides the riverbed elevation, the water level elevation and the water depth, based on topographic cross-sections surveyed every 50 m, increased to every 25 m where the riverbed geometry rapidly changes.

The coefficients of the linear regression of the empirical model were calculated with 70% of the samples randomly extracted from the set of samples derived in the previous step. The remaining 30% of samples were used to assess the validity of the model.

This method was repeated for each orthomosaic at each spectral band. The goodness of fitting was assessed by means of the coefficient of determination calculated on the 70% of samples and mean absolute error computed on the remaining 30% of samples.

4 RESULTS

In Figures 2, 3, 4 and 5 a subset of results from the first survey is shown. The sign of linear regression slope changes from positive to negative values, as wavelength increases. This behaviour is due to the weak correlation between water depth values and pixel values in the relation (5) at shorter wavelengths, that increases at longer wavelengths (Red and Near Infrared), when the absorption due to water column becomes stronger compared to reflectance. The increasing trend of correlation versus wavelength is more evident in Figure 6 where the coefficient of determination R^2 and the mean absolute error with respect to spectral band are shown.

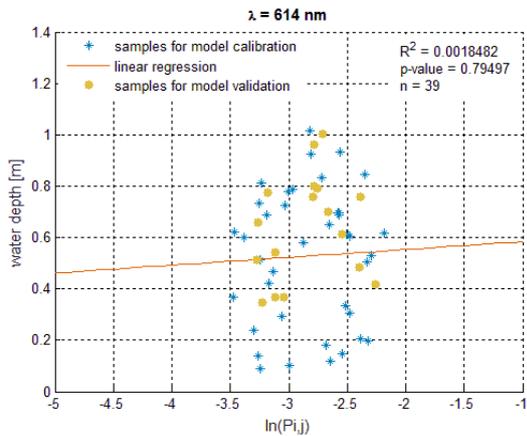


Figure 2: Linear regression at $\lambda = 614$ nm.

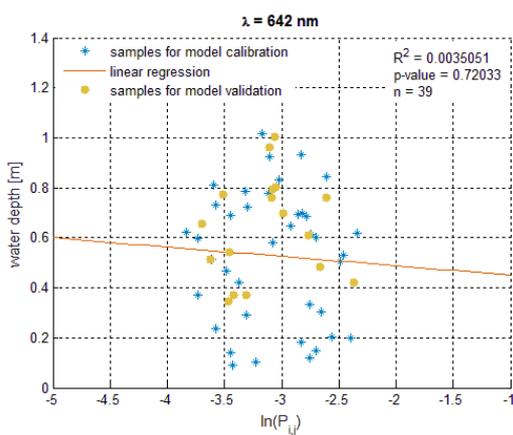


Figure 3: Linear regression at $\lambda = 642$ nm.

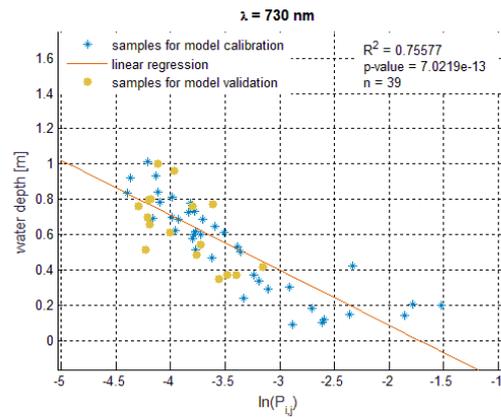


Figure 4: Linear regression at $\lambda = 730$ nm.

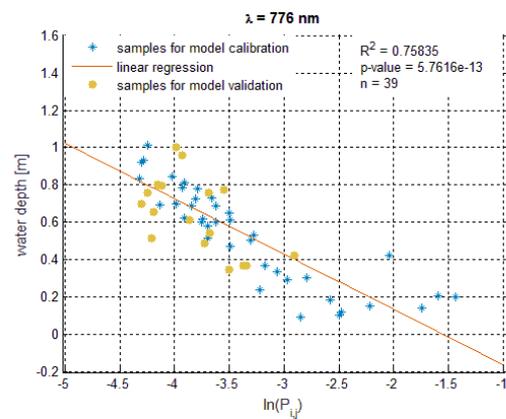


Figure 5: Linear regression at $\lambda = 776$ nm.

The best correlations are obtained in the spectral range from 700 nm to 800 nm.

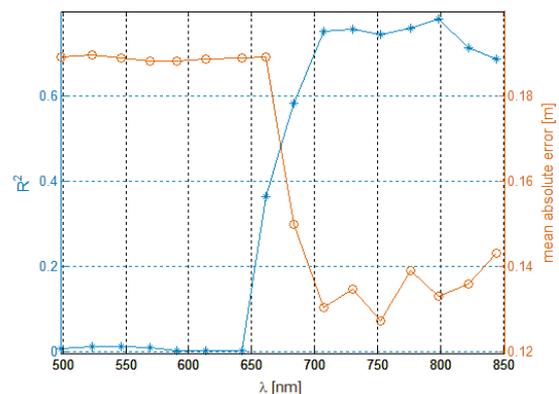


Figure 6: Trend of coefficient of determination and mean absolute error versus wavelength.

In Figures 7, 8, 9 and 10, a subset of results from the second survey is shown. In Figure 11 trends of coefficient of determination and mean absolute error with respect to spectral band are shown, confirming

both the best values of correlation in the spectral range from 700 nm to 750 nm and the increase of correlation versus wavelength, obtained in the first survey.

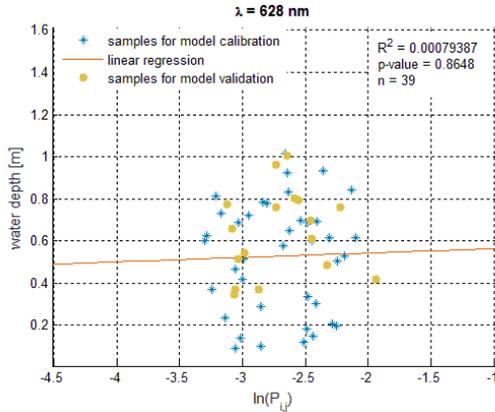


Figure 7: Linear regression at $\lambda = 628$ nm.

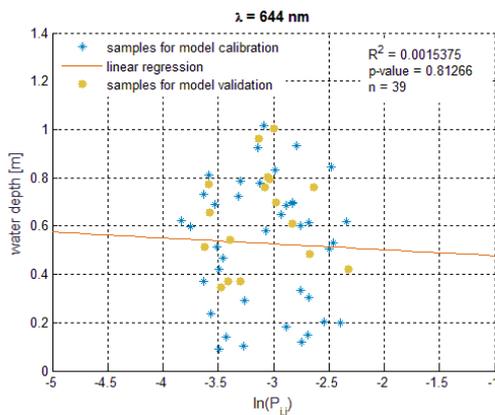


Figure 8: Linear regression at $\lambda = 644$ nm.

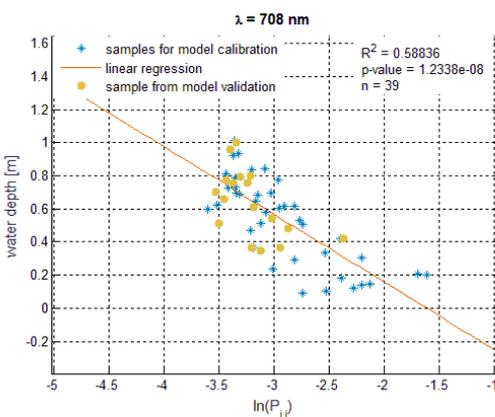


Figure 9: Linear regression at $\lambda = 708$ nm.

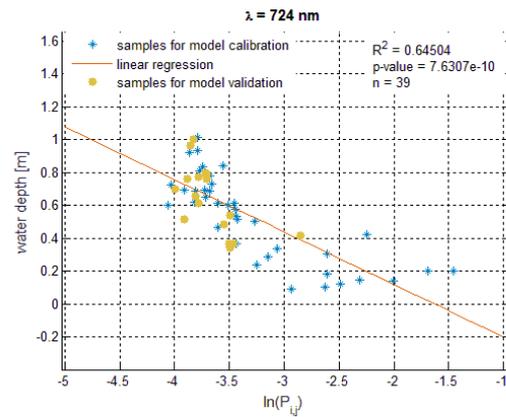


Figure 10: Linear regression at $\lambda = 724$ nm.

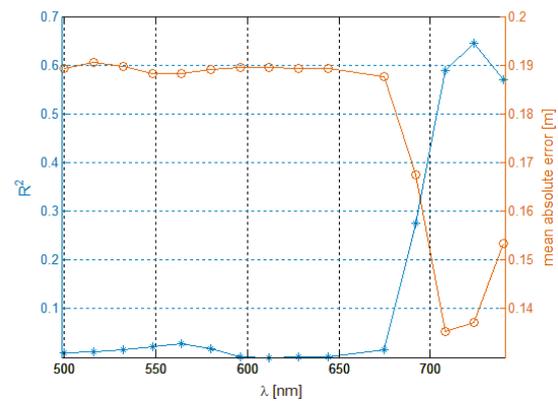


Figure 11: Trend of coefficient of determination and mean absolute error versus wavelength.

5 CONCLUSIONS

The results show the potential of UAV hyperspectral data for bathymetric mapping at centimetre resolution. The empirical model fits well the water depth values derived from the hydraulic model in the spectral range from 700 nm to 800 nm with an average error less than 0.13 m in the best case when the water depth ranges from 0.09 m to 1.01 m.

In further studies we intend to apply the proposed methodology over imagery acquired on other longer reaches of the Ain River with a wider range of water depth in order to confirm the model behaviour with respect to wavelength, to investigate its applicability over a range of wider environmental conditions, such as changes in river bottom morphology and composition, concentration in suspended sediment, water deepness and finally to examine obtained results on the basis of sensor

configuration and acquisition mode as pixel ground resolution and bandwidth.

REFERENCES

- Bravard, J. P. (1986). La basse vallée de l'Ain: dynamique fluviale appliquée à l'écologie. *Recherches Interdisciplinaires sur les Ecosystèmes de la Basse-Plaine de l'Ain: Potentialités Evolutives et Gestion. Doc. Cartogr. Ecol*, 29, 17-43.
- Carbonneau, P. E., Lane, S. N., & Bergeron, N. (2006). Feature based image processing methods applied to bathymetric measurements from airborne remote sensing in fluvial environments. *Earth Surface Processes and Landforms*, 31(11), 1413-1423.
- Feurer, D., Bailly, J. S., Puech, C., Le Coarer, Y., & Viau, A. (2008). Very-high-resolution mapping of river-immersed topography by remote sensing. *Progress in Physical Geography*, 32(4), 403-419.
- Fonstad, M. A., & Marcus, W. A. (2005). Remote sensing of stream depths with hydraulically assisted bathymetry (HAB) models. *Geomorphology*, 72(1), 320-339.
- Gao, J. (2009). Bathymetric mapping by means of remote sensing: methods, accuracy and limitations. *Progress in Physical Geography*, 33(1), 103-116.
- Gilvear, D., Hunter, P., & Higgins, T. (2007). An experimental approach to the measurement of the effects of water depth and substrate on optical and near infra-red reflectance: A field-based assessment of the feasibility of mapping submerged instream habitat. *International Journal of Remote Sensing*, 28(10), 2241-2256.
- Lane, S. N., Westaway, R. M., & Murray Hicks, D. (2003). Estimation of erosion and deposition volumes in a large, gravel-bed, braided river using synoptic remote sensing. *Earth Surface Processes and Landforms*, 28(3), 249-271.
- Lejot, J., Delacourt, C., Piégay, H., Fournier, T., Trémélo, M. L., & Allemand, P. (2007). Very high spatial resolution imagery for channel bathymetry and topography from an unmanned mapping controlled platform. *Earth Surface Processes and Landforms*, 32(11), 1705-1725.
- Legleiter, C. J., Roberts, D. A., & Lawrence, R. L. (2009). Spectrally based remote sensing of river bathymetry. *Earth Surface Processes and Landforms*, 34(8), 1039-1059.
- Liébault, F., & Piégay, H. (2001). Assessment of channel changes due to long-term bedload supply decrease, Roubion River, France. *Geomorphology*, 36(3), 167-186.
- Makelainen, A., Saari, H., Hippi, I., Sarkeala, J., & Soukkamaki, J. (2013). 2D Hyperspectral frame imager camera data in photogrammetric mosaicking. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 263-267.
- Maritorena, S., Morel, A., & Gentili, B. (1994). Diffuse reflectance of oceanic shallow waters: Influence of water depth and bottom albedo. *Limnology and oceanography*, 39(7), 1689-1703.
- Naudet, G., Camenen, B., Le Coz, J., Paquier, A., & Piégay, H. (2014, June). Numerical modelling contribution to sedimentary redynamisation projects in the Ain River. In *Conférence Internationale IsRivers 2015*.
- Naudet, G., Le Coz, J., Camenen, B., Paquier, A. (2015, Septembre). Modélisation numérique hydrosédimentaire 1D et 2D de la basse vallée de l'Ain. Technical report.
- Paquier, A., Camenen, B., Le Coz, J., & Béraud, C. (2014). Comparison of two models for bed load sediment transport in rivers.
- Philpot, W. D. (1989). Bathymetric mapping with passive multispectral imagery. *Applied optics*, 28(8), 1569-1578.

**ELECTROMAGNETIC
WAVE
TECHNOLOGIES**

Physically Inspired Digital Predistortion for Dual Input Doherty Power Amplifiers and Automatic in-Situ Identification

Giovanni Donati, Christian Musolff, Giuseppe Gottardo, Daniel Popp and Georg Fischer

Friedrich Alexander University of Erlangen-Nuremberg, Cauerstrasse 9, 91058 Erlangen, Germany

{giovanni.donati, christian.musolff, giuseppe.gottardo, daniel.popp, georg.fischer}@fau.de

Keywords: Digital Signal Processing, Digital Predistortion, Dual Input Doherty, Power Amplifiers

Abstract: A novel approach for the linearization of dual input Doherty power amplifier (DPA) is derived by taking inspiration from the operation of the device. A new point of view is evaluated and an automatic identification procedure is developed, reducing the complexity of the predistorter, improving linearity and efficiency at the same time and avoiding the usage of crest factor reduction (CFR). It is also emphasized the importance of a power dependent phase relationship between the inputs of the dual input DPA. A validation of the theory is performed by means of a MATLAB/ADS envelope co-simulation.

1 INTRODUCTION

Modern telecommunication systems are aiming for very high performance demanding high linearity, efficiency and wide bandwidth. A key component of the Base-Station (BS) operation is the power amplifier, which is the major contributor to the whole power consumption of the system. The signals used in order to improve the spectral efficiency have high peak to average power ratio (PAPR) and are decisive in terms of the average efficiency of the whole system. For this reason Doherty Power Amplifiers (DPA) are representing a primary choice for BS applications. DPAs deliver an extended efficiency range and are perfect candidates for amplification of signals with a very high dynamic range. Furthermore, the advances in the design techniques led to new power amplifier (PA) architectures able to deliver high efficiency over a large bandwidth of operation. Class J and JF realizations of the DPA are making use of harmonic terminations to improve the bandwidth. It has already been demonstrated that separating the input branches and driving them separately is of huge benefit in terms of bandwidth and reconfiguration of the power added efficiency (PAE) (Gustafsson, Anderson and Fager, 2012; 2013). Also the concept behind the algorithms for digital pre-distortion (DPD) of dual input DPA were documented (Cahuana et al., 2014), reporting very good

performance and showing the advantages of using this architecture in BSs. The previous research has demonstrated the advantage of driving MAIN and PEAK amplifiers by means of a static splitter, implemented in the digital domain, delivering the optimal combinations of the inputs signals to the PA in order to obtain high efficiency (up to 42%). Despite the improvements, the effect of the phase difference between the two input branches is still not fully characterized yet. In this article we are showing the importance of the power dependent phase correction to be applied at the input of the dual input DPA architecture, and we are presenting a way to automatically identify the optimum combinations of the inputs in-situ. We have reached an average power efficiency of 51%, and we developed a simple physics-inspired algorithm for the pre-distortion that we tested with a 5 MHz LTE signal. The results are demonstrated and validated using envelope simulations in Advanced Design System (ADS).

2 THEORY

In the work of Gustafsson, Anderson and Fager, (2013) it was demonstrated how, the efficiency in power back-off (PBO) of a DPA could be improved and maintained over a wide fractional bandwidth, by introducing the back-off point ϵ_b as a design parameter. All the control parameters of the device

can be expressed as a function of the normalized drive level ε , in particular the output power is expressed as:

$$P_{out}(\varepsilon, f, \theta) = \frac{I_{max_1} RL}{8} \left[\begin{array}{c} \varepsilon^2 + 2\varepsilon \cos\left(\frac{\pi f}{2}\right) \cos(\theta) + \\ k^2 \cos^2\left(\frac{\pi f}{2}\right) \end{array} \right] \quad (1)$$

where I_{max_1} is the maximum current deliverable by the MAIN PA, $\varepsilon_b < \varepsilon < 1$, θ is the phase difference between MAIN and PEAK amplifier input, f is the fractional bandwidth and k is the drive level of the PEAK that sets the efficiency bandwidth of the DPA. In particular, as also demonstrated by the work of Gustafsson, Anderson and Fager (2013), the relationship determining the acceptable values for \bar{f} is:

$$\frac{1 - \varepsilon_b}{1 + \varepsilon_b} = \sin\left(\frac{\pi \bar{f}}{2}\right) \quad (2)$$

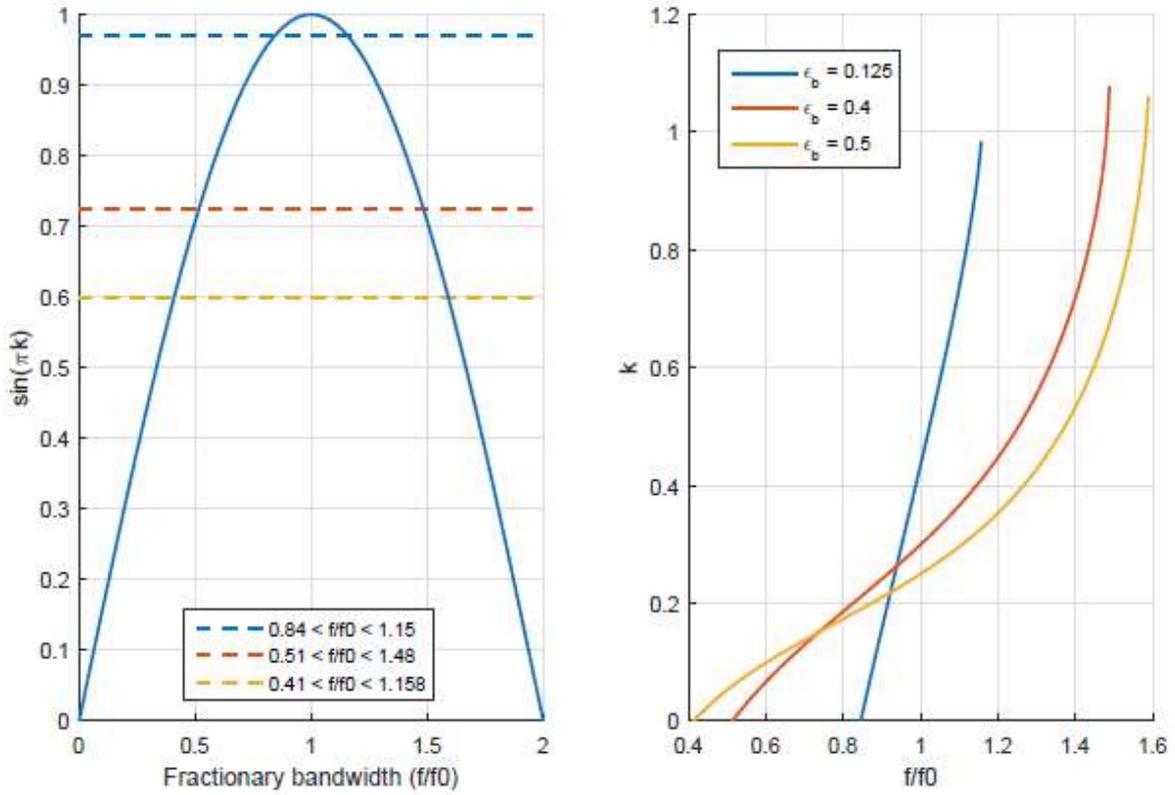


Figure 1: Dependency of the DPA bandwidth on back off poin choice

Figure 1 is graphically showing how the choice of ε_b is decisive for determining the efficiency bandwidth. In addition it was presented a closed formula (Gustafsson, Anderson and Fager, 2013), for the phase relationship between main and peak power amplifier (PA), where the dependency on the fractional bandwidth and the drive level of the PEAK is exploited (3):

$$\theta = \arccos\left(\frac{-k \cos\left(\frac{\pi \bar{f}}{2}\right)}{2\varepsilon}\right) \quad (3)$$

It is clear that the ability to control the input signals of the DPA architecture makes it possible to reach high efficiency both at maximum power and

PBO, causing a active load modulation that assures high performance on an extended frequency range. The theory previously developed by Gustafsson, Anderson and Fager (2013) states that improving the efficiency bandwidth of the DPA architecture leads to a degree of freedom in the reconfiguration of the PA efficiency, depending on the particular frequency where a signal is residing. This is very important for modern BS where we are continuously dealing with multistandard signals and scenarios involving multiband transmission. In Cahuana et al. (2014), this theory was used to implement a digital power splitter to get the maximum power efficiency out of the designed DPA

3 REVISION OF THE IDENTIFICATION PROBLEM

The previous scientific work has opened a lot of possibilities in terms of reconfiguration and bandwidth enlargement. Also if the articles in literature are giving a very good insight of the problem and deliver a solution, it is still unclear how to identify the input combinations in a flexible and automatic way. If we imagine the problem as the identification of a drive function for the DPA, we could depict it as the black box (Figure 2):

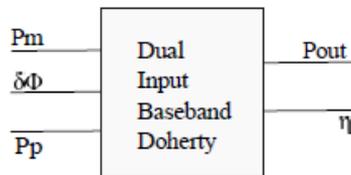


Figure 2: Block scheme for the identification of a drive function.

where P_m , P_p and $\delta\phi$ are respectively the values of the input powers for MAIN and PEAK amplifier, and the phase relationship between them. At the output of the model we have the power and the power added efficiency. Seen in this way, the identification of a single drive function can be exploited as a multidimensional optimization problem. Specifically if we could try all the possible triplets (P_m , P_p , $\delta\phi$), we would end up discovering that there is a theoretically infinite number of combinations leading to the same output power. The challenge relies in identifying the triplets maximizing the efficiency for a specific power level and frequency. In order to do so performed an analysis on the model of a dual input DPA with 300

MHz of bandwidth (700 to 1000 MHz) and a maximum output power of 100 W. A set of measurements in ADS was obtained by means of harmonic balance (HB) simulations, performing a large power sweep over the possible combinations of the input parameters at four frequency points. The results of the simulations at 900 MHz are shown in Figure 3:

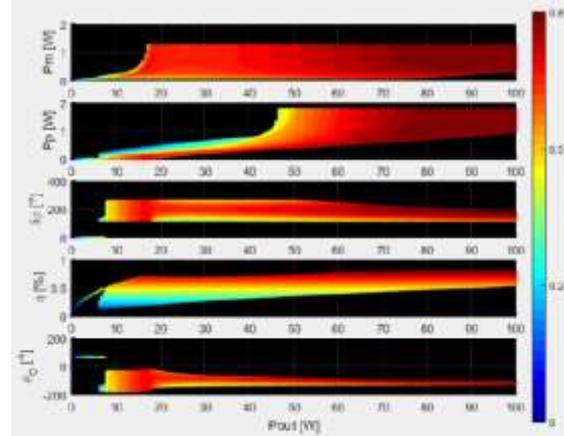


Figure 3: Multidimensional identification dataset for drive function.

where η is the power added efficiency and ϕ_O is the output absolute phase of the PA. The color is coding the information about the efficiency value assumed by a specific point, with a resolution of about 2%. Analyzing the results in figure 3 it becomes obvious how several combinations of the inputs lead to the same output power but not to the maximum efficiency. This appears much more evident in the middle power region. The image was generated by separating the data into bins in order to reduce dimensionality of the dataset, which is otherwise composed of more than 90.000.000.000 points. Despite the separation into bins the data are quite dense, so in order to better appreciate the magnitude of the problem we should zoom in. Figure 4 presents a closer view of the data. Using different triplets we could generate 38 W at the output of the PA, but in a very small range around it (4 mW) we could drop the efficiency to 30% or less by choosing a suboptimal triplet.



Figure 4: Detail of the characterization space for P_m .

In (Cao et al, 2012) an algorithm for the DPD of dual input power amplifiers was presented. From the

developed theory, it turns out that the knowledge of an optimized signal for the peak power amplifier is necessary, together with the desired RF output signal, in order to determine the shape of the predistorter signal of the main PA. This technique has a higher complexity than the classic approach, but because the nonlinear order used was low it was still considered acceptable. This algorithm shows a good linearization performance and also tries to optimize the efficiency, but it seems that the compromise between the two goals avoids obtaining a very good efficiency in PBO.

4 NOVEL LINEARIZATION APPROACH

We can improve the identification of the digital static splitter by optimizing it for both efficiency and linearity. Below the compression point the response of the system is depending only on the behavior of the MAIN PA. Defining the drive function in the power domain, we can build a model for the predistorter of the MAIN branch with a piecewise lookup table (LUT).

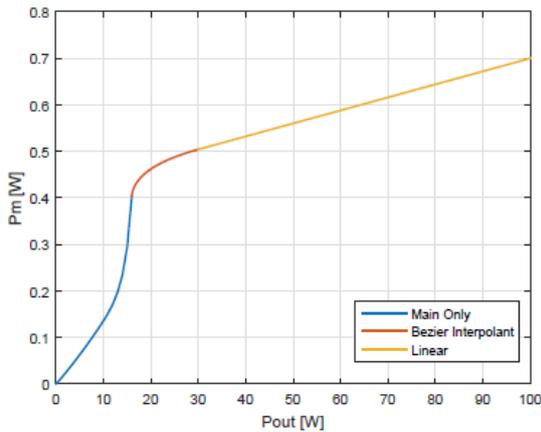


Figure 5: Example of MAIN drive function.

Referring to Figure 5 we have three different sections of the curve. The first one is obtained by turning the PEAK PA off and sending a power ramp to the MAIN PA. We can determine its saturation power and choose the back-off point. Interestingly we are free to reconfigure the power efficiency of the dual input DPA by moving the back-off point. There is a trade off between linearity and efficiency when choosing the back-off point. By deciding the compression point of the MAIN PA we can allow a

certain amount of nonlinear effects included in the system. The drawback observed is a decrease of the back-off efficiency. The third part of the curve is linear because we don't want to generate nonlinear effects by predistorting.

The two curves could be directly blended but this would generate problems in terms of bandwidth expansion of the input signal. The knee between the the two curves is a discontinuity in the first order derivative of the function and is responsible for the generation of very high frequency components in the spectrum. The bandwidth expansion generates issues in the signal reconstruction path because the DAC has a bandwidth limited by its sampling frequency which is not infinite. By using a Bezier interpolant (Ping and Guozhao, 2011) to connect the two curves, we are reducing the nonlinear effects of the curve and improving the spectral efficiency by introducing G2 continuity at the blending points. By defining the behavior of the MAIN PA predistorter we reduce the dimensionality of the identification problem and we detect the right value of output power for the DPA in that point. Since we are trying to linearize the device, we would like to have an output characteristic which is linear in the amplitude and possibly constant in phase. Using the information about the output phase at the back-off point, we can set a target for the algorithm performing the identification of the drive functions. As in Cao et al. (2012), we can define a target output power (equally spaced power points) with the condition of constant output phase joined with a LUT of powers to drive the MAIN PA. The first main difference with the work of Ca, is that we are not introducing any DPD mathematical model, we are instead using a certain number of supports (for instance 200) to extract the LUTs used later as references to interpolate between the values. LUTs can be considered less performing, especially in terms of bandwidth expansion, when compared to polynomials, but it is was already demonstrated (Barradas et al., 2014) that they can be reformulated in order to be as efficient as the polynomial model. In order to identify the missing predistorter functions, an intelligent algorithm can be applied. We need to define a cost function and minimize it in order to find the value of the PEAK and phase predistorted signals. A good candidate for such function is the pure error vector module EVM (not the classic EVM used for constellations), which can be easily calculated as:

$$EVM = \frac{|S_r - S_i|}{|S_r|} \quad (4)$$

which is expressing the ratio between the error vector and the original vector. The EVM also accounts for the output phase, so by minimizing it we are able to find the correctly aligned output vector with the wanted power and efficiency. It should be emphasized that the minimization problem uses a mono dimensional cost function. The efficiency optimization is performed by choosing the target output phase that we want to maintain at the PA output. Gradient based algorithms such as the ones used to train the classic DPD models could be used to minimize a specific cost function, but they have experimentally exhibited slow convergence on this kind of problem. In addition they could converge to global minima, preventing the discovery of an absolute minimum. Particle Swarm Optimization (Kennedy et al., 1995) is a very attractive algorithm for this sort of application because it is simple and allows control over the power range swept at the input of the DPA. PSO is really useful in a real application in order to avoid damaging the devices due to wrong drive levels. In the literature there are already documented uses of PSO in the field of DPD (Abdelhafiz et al., 2013) for the computation of the coefficients of DPD models. Here we would use it for identifying the optimum triplets directly. The idea behind PSO is very simple, a single particle is described by a set of parameters:

- Position (X): described by the decision variables
- Velocity (V): velocity of the particle during its motion, defines also the direction of search
- Local Best (L): local best met by the particle, updated each time the local best is improved during the search
- Global Best (G): global best found by the whole swarm, this information is shared between all the particles

Each particle is moving in the 2-dimensional space updating its speed according to the social interactions with the rest of the swarm. The formulas to update the velocity and position are expressed as in 5 and 6.

$$\begin{aligned}
 V^{n,d} &= \chi V^{n,d} + c_1 \text{rand}(\cdot)(L^{(n,d)} - \\
 &X^{(n,d)}) + c_2 \text{rand}(\cdot)(G^{(n,d)} - \\
 &X^{(n,d)})
 \end{aligned} \quad (5)$$

$$X^{n,d} = X^{n,d} + V^{n,d} \quad (6)$$

Table 1: PSO algorithm

Algorithm 1 PSO

- 1: Initialize a population of particles with random values positions and velocities from D dimensions in the search space
 - 2: Return the global minimum G and the values of the input variables generating it
 - 3: **while** Termination condition not reached **do**
 - 4: **for** Each particle i **do**
 - 5: Adapt velocity of the particle using Equation 5
 - 6: Update the position of the particle using Equation 6
 - 7: Evaluate the fitness $f(\vec{X}_i)$
 - 8: **if** $f(\vec{X}_i) < f(\vec{P}_i)$ **then**
 - 9: $\vec{P}_i \leftarrow \vec{X}_i$
 - 10: **if** $f(\vec{X}_i) < f(\vec{P}_g)$ **then**
 - 11: $\vec{P}_g \leftarrow \vec{X}_i$
-

The coefficients appearing in the equation of the velocity are controlling the "memory" of the particles to lead them towards the best solution that has been found so far. Coefficient c_1 is controlling the tendency of the particle to search in the direction of its own best found solution, while c_2 manages the

social interaction of the particle with the rest of the swarm members to let its position drift towards the global best found by the whole swarm. The flow of the optimization algorithm 1 is very simple to implement and can be adapted to a large range of problems by performing a proper sensitivity study of

the parameters. To make sure that the particle swarm optimizer does not stop if no convergence is met, a maximum number of iterations can be established. This also sets the speed of the algorithm in the worst case, when no optimum is found or when the cost function does not reach the expected precision

5 APPLICATION OF PSO TO STATIC DRIVE FUNCTION IDENTIFICATION

In our approach, the PSO algorithm described above was adapted to the problem of the identification of a single static drive function: Position X is a bi-dimensional vector made of P_p (expressed in W) and $\delta\Phi$ (expressed in degrees); a vector for the target output power is generated by defining 50 equally spaced values; the target output phase is established by using the procedure explained in the previous section.

6 SIMULATION SETUP

In order to prove the concept, a co-simulation MATLAB/ADS was performed. Signal generation, baseband signal processing and analysis of the results were performed in MATLAB while the whole physical part of the simulation was performed by ADS which was configured to perform circuit envelop simulation. The workflow represented in Figure 6 starts with a MATLAB script generating the drive function for the MAIN PA with only 50 points, defining the linear power target points and identifying the back-off point in the same way described in the previous sections.

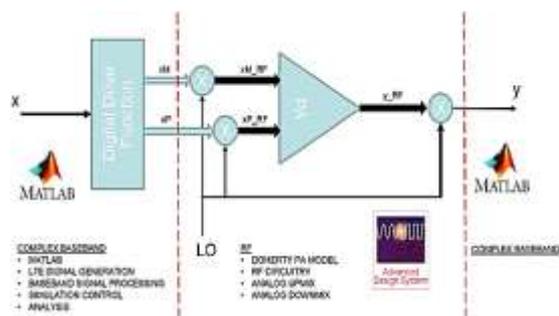


Figure 6: Simulation setup interfacing MATLAB and ADS.

When the algorithm reaches the back-off point it reads the output phase of the PA and starts the particle swarm optimizer to sequentially compute the right pre-distorted phase and power for the peak PA. Since we want to characterize the static correction for the DPA, in MATLAB we are generating constant power points of a duration which avoids memory effects to be sensed at the output. It was experimentally seen that the memory effects of the PA model are visible just up to 300 ns. For this reason we have set the duration of the power pulses to 1 μs . The convergence of PSO is improved by using more particles, but this also means a bigger number of evaluations and slower results. We used $N = 20$ particles as a good compromise between the simulation time and the precision of the results. To limit the maximum duration of the simulation we have set the number of maximum iterations to 50. The whole simulation takes about:

$$T_{sim_{max}} = N_{particles} * N_{targets} * N_{max-iterations} * T_{cycle} \quad (7)$$

Unfortunately such simulation is time consuming because of the calls to ADS made from MATLAB. The simulation for one point takes about two seconds, so the whole duration, in the worst case, is about 27 hours. The solution is found after less than 10 iterations, so the simulation can be completed in 2 to 5 hours. We expect that by performing the measurement in-situ using an FPGA system, the time to measure each point can be reduced to the limits of memory effects, resulting in a complete identification of the drive function in 15 ms. This process can be applied around several carrier frequencies, creating a raster of drive functions to obtain a wideband model. This workflow generates a LUT, which is used to evaluate the results by running a simulation with an LTE signal. Results are shown in the next section.

7 RESULTS

We have simulated the identification of the drive function by setting the carrier frequency to 900 MHz. The identified drive function is shown in Figure 7.

Looking at the phase relation between MAIN and PEAK driving signals, we noticed that phase is ranging from 200 degrees, at low output power, to 166 degrees at the maximum power. There are 340

of total phase variation, which is not a negligible quantity to account for. This shows the importance of the relationship between P_p and $\delta\phi$, which is why we should choose dual input DPAs for BS operation. By correctly tweaking P_p we select a specific phase relationship, at the input of the structure, maximizing linearity and efficiency for a specific carrier frequency.

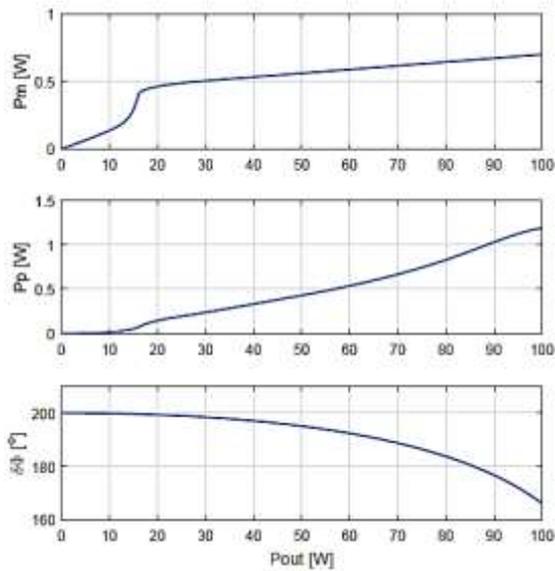


Figure 7: Drive function for the DPA at 900 MHz.

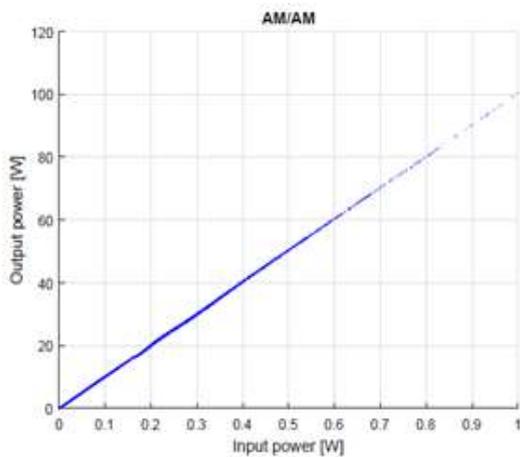


Figure 8: Amplitude Input Output characteristic of DPA + DF.

Looking at the phase relation between MAIN and PEAK driving signals, we noticed that phase is ranging from 200 degrees, at low output power, to 166 degrees at the maximum power. There are 34° of total phase variation, which is not a negligible

quantity to account for. This shows the importance of the relationship between P_p and $\delta\phi$, which is why we should choose dual input DPAs for BS operation. By correctly tweaking P_p we select a specific phase relationship, at the input of the structure, maximizing linearity and efficiency for a specific carrier frequency.

Using the drive function to pre-distort the DPA driven by a 5 MHz LTE signal with a 10 dB PAPR (no CFR was applied), we obtained a linear output characteristic (Figure 8), where the dispersion around the curve is due to the memory of the device.

As we can see in Figure 9, the whole system still shows a non-linear phase characteristic

This is due to the fact that the identification process is done on the two separated inputs of the DPA. When applying a dynamic signal, the splitter does not account for the phase difference between the input of the DF and the output of the DPA.

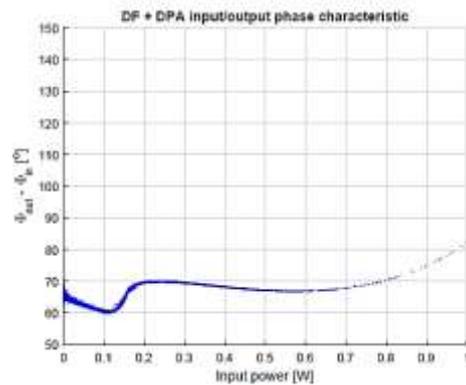


Figure 9: Phase characteristic of DPA + DF

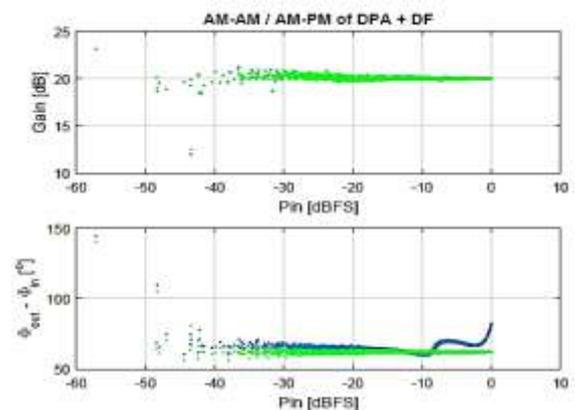


Figure 10: AM-AM and AM-PM characteristic of DF + DPA after static phase correction

Such phase difference can be statically corrected by a common identification technique. We can see the effect of the phase linearization in Figure 10.

After applying the static phase correction (SPC) we obtained a drastic improvement of the ACPR, from 30 to 54 dBc (Figure 11) and an average power added efficiency of 51:1%.

This represents a very good result, considering that no CFR was applied to the input signal.

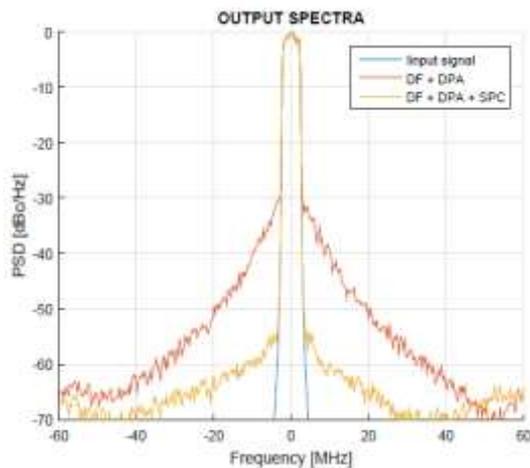


Figure 11: Resulting output spectra

8 CONCLUSIONS

In this article a novel approach for the linearization of dual input DPAs was presented.

We showed how, taking inspiration from the device operation, flexible reconfiguration of the efficiency can be obtained by means of a digital drive function. In addition to that, we developed an automatic identification procedure eligible for FPGA implementation, due to its simplicity.

The algorithm performs an optimization of the input signals for the device and avoids a multi-object approach delivering maximum efficiency by keeping the linearity conditions at its output.

The procedure was validated by making use of a MATLAB/ADS cosimulation environment and testing the workflow on a 5 MHz LTE signal with a 10 dB PAPR. Results have demonstrated that very good performance can be reached without introducing CFR in the transmit chain.

Therefore this opens a huge number of possibilities to adapt the performance to the signal's probability density function and to future multiband approaches for the linearization of this architecture.

Combined together, the identification of several drive functions at different frequencies and, novel techniques for the detection of the instantaneous frequency of non-stationary signals, can lead to a new wideband approach to the DPD.

REFERENCES

- Gustafsson, D., Andersonn C.M. and Fager, C., 2013, "A Modified Doherty Power Amplifier With Extended Bandwidth and Reconfigurable Efficiency", IEEE Transactions on Microwave Theory and Techniques, 533-542.
- Gustafsson, D., Andersonn C.M. and Fager, C., 2012, "A Novel Wideband and Reconfigurable High Average Efficiency Power Amplifier", Microwave Symposium Distest (MTT), IEEE MMT-S International, 1-3.
- Cahuana, J.C., Landin, P.N., Gustaffson, D. and Fager C., 2014, "Linearization of Dual Input Doherty Power Amplifiers", IEEE Integrated Nonlinear Microwave and Millimetre-wave Circuits (INMMic), 1-3.
- Chao, H., Qureshi, J., Eriksson, Fager, C., and de Vreede L., 2012, "Digital predistortion for dual-input Doherty amplifiers", Power Amplifiers for Wireless and Radio Applications (PAWR), IEEE Tropical Conference on, 45-58.
- Ping Z., Guozhao, W., 2011, "Non-uniform hyperbolic blending B-spline curve", Computer Science and Automation Engineering (CSAE), IEEE Internation Conference on, 88-92.
- Barradas, F.M., Cuhna, R.T., Lavrador, P.M., Pedro, J.C., 2014, "Polynomials and LUTs in PA Behavioral Modeling: A Fair Theoretical Comparison", IEEE Transactions on Microwave Theory and Techniques, 3247-2185.
- Kennedy, J., Eberhart, R., 1995, "Particle Swarm Optimization", Neural Networks, IEEE Internationl Conference on, 1942-1948.
- Abdelhafiz, A.H., Hammi, O., Zerguine, A., Al-Awami, A.T., Ghannouchi, F.M., 2013, "A PSO Based Memory Polynomial Predistorter With Embedded Dimension Estimation", IEEE Transactions on Broadcasting, 665-673.

Is it Possible to Detect the Stealth Flying Objects by the Millimetre Wave Radiometer?

Jinghui Qiu, Shengchang Lan, Hao Liu, Xinyu Yin and Alexander Denisov
Department of Microwave Engineering, Harbin Institute of Technology, P.R.China

Xiao Qing and Zhao Man
Southwest Institute of Electronic Equipment, Chengdu, P.R.China

Francesco Soldovieri
Istituto per il Rilevamento Elettromagnetico dell'Ambiente Consiglio Nazionale delle Ricerche, Napoli, Italy

Keywords: Microwave radiometer, Passive millimetre wave imaging system, Remote sensing control, Antiradar coating

Abstract: This work evaluates the possibility of using the passive millimeter waves (PMMW) radiometric discriminator for the remote control and the detection of stealth aircrafts.

1 INTRODUCTION

Microwave radiometry is concerned with measurements of natural electromagnetic radiation of an object of physical temperature above 0° K.

A large literature describes the main working principles of microwave radiometry (Reinwater, 1978; Moffa et al, 2001; Goldsmith et al,1993; Appleby and Lettington, 1991; Piechl, 2004; Poradish and Habbe, 1982; Esepkina et al, 1973; Skou, 1989; Shuchardt et al, 1981). With respect to special application fields, it is interesting also to read old publications and patents regarding military designs and applications based on the deployment of millimeter wave bands during the Cold War (Shuchardt,1978; Moore et al., 1976; Parnell, 1988; Seashore, Miley and Kearns, 1979; Corrado, 1988; www.giws.de). After, due to the large potential market, there has been a large diffusion of modern microwave firms for the equipment of the airport security systems and the detection of the concealed objects.

Based on the growing problem of the terrorism, especially after 11 September 2001, a large amount of money has been devoted to special programs for the design of special passive millimeter wave imaging systems (PMMW) (Proc..of SPIE a lot of, Huguenin,2006; Appleby,2007; Internet,Dill et al. 2009).

The focus of the present work is on the possibility

offered by PMMW radiometric systems in order to detect remote aircraft with antiradar surfaces coverage (Figure 1).

A relevant statement about the use of an anti-radar surfaces adapted to the stealth ship is provided just below. (wikipedia.org/wiki/Stealth ship). “In designing a ship with reduced radar signature, the main concerns are radar beams originating near or slightly above the horizon (as seen from the ship) coming from distant patrol aircraft, other ships or sea-skimming anti-ship missiles with active radar seekers. Therefore, the shape of the ship avoids vertical surfaces, which would perfectly reflect any such beams directly back to the emitter. Retro-reflective right angles are eliminated to avoid causing the cat’s eye effect. A stealthy ship shape can be achieved by constructing the hull and superstructure with a series of slightly protruding and detruding surfaces”.

Anyway, it is sufficient to change the word from “ship” to “aircraft”.

The stealth coating it is very suitable to avoid the specular reflection for active systems, but at the same time is practically useless for radiometric systems, because nature produces radiations, which re-reflect from this surfaces to antenna of radiometer from the every directions (Figure 2 and 3)



Figure 1: Photos of various modern famous special aircrafts with antiradar surfaces.

Objects reflect and emit radiation in the millimeter wave range as they do it in the infrared and visible ranges. The degree to which the object reflects or emits is characterized by emissivity ϵ . A perfect radiator (absorber) has $\epsilon = 1$ and is known as a blackbody (Esepkina, 1973). A perfect reflector (non-absorber) has $\epsilon = 0$. The earth and the sky can be approximated as blackbody, whereas the metal object is a reflector. Intermediate values of the emissivity ϵ depends on several parameters as, dielectric properties of the objects, angle of observation (for example, for the water surface), the polarization parameters, the surface roughness or coatings, the wavelength and other factors.

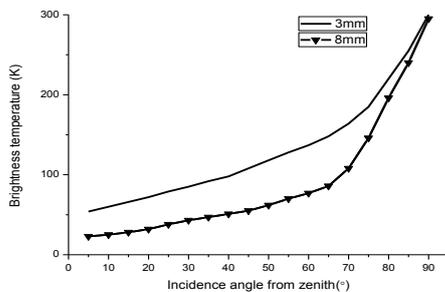


Figure 2: Brightness temperature of sky at the various angles of relatively zenith.

The measurement of such a radiation is more correct and understandable in terms of radio brightness (simply brightness) *temperature*, which is expressed in temperature T .

According to the usual definition (Esepkina, 1973), a radiometer (Figure 4) is a receiving device designed for the measurements of the level of noise radiation in an assigned band of the frequencies Δf .

The main functionality of the receiver in a radiometer is to provide a measure of the input noise power, expressed as an antenna temperature, in equivalent black body temperature units. The sensitivity of the radiometer ΔT_{sens} , is defined as the minimum detectable signal and is determined by amplitude of the fluctuations presented at the output indicator in the absence of the signal.

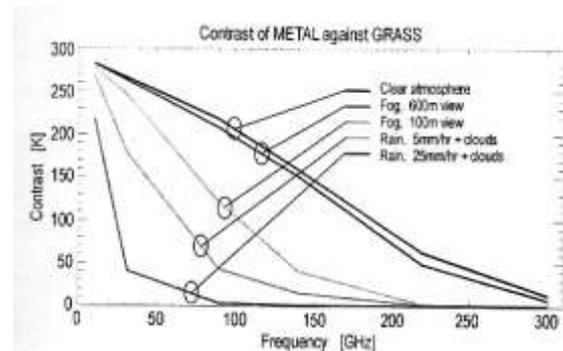


Figure 3: Possible radio-brightness radiometric contrast between metal and grass at the zenith angle for the case of the observing an objects on the earth surface



Figure 4: Picture of the 8 mm radiometer (Gorishnyak et al, 2004).

Usually, the temperature sensitivity of radiometer is evaluated for post-detection time 1 sec. More details about various situations concerning ΔT_{sens} (calculated and measured) can be found in (Esepkina, 1973; Skou 1989).

The bottom of the aircrafts will reflect the radiation of the hot Earth and accordingly it will be seen by the radiometer as an hot object with respect to the background of the cold sky (space).

Even in case of the coating of all surfaces by special absorption (small probability) material (“painting”), such a “blackbody” will have surface

brightness temperature as from outside of the aircraft. Remember standard information at the board during the air flight: “Temperature of air overboard makes minus 56 degrees.

In this paper, we evaluate the possibility of using PMMW radiometric system for the remote control and the finding of stealth aircrafts. The choice of the real working frequency depends on the real size of antenna and the microwave losses in atmosphere on path radiometer – aircraft (Figure.5 and Table 1).

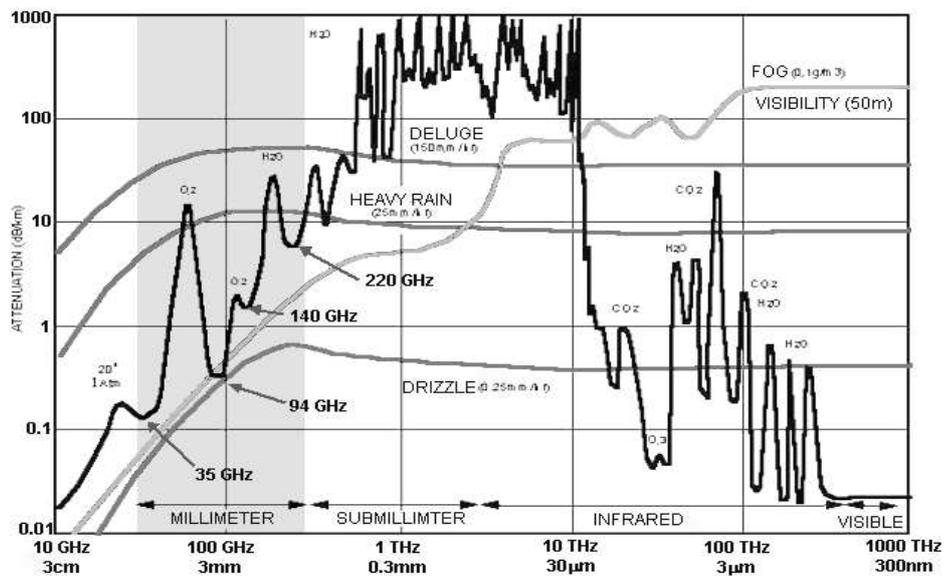


Figure 5: Absorption of the electromagnetic waves in the atmosphere

Table 1: Absorption in atmosphere for Three Window Frequencies (Seashore et al, 1979)

CHARACTERISTIC	FREQUENCY		
	35GHz	94GHz	140GHz
Wavelength	8.6mm	3.2mm	2.2mm
Clear Air Attenuation	0.12dB/km	0.4 dB/km	1.6 dB/km
Rain Attenuation			
0.25mm/hr	0.07 dB/km	0.17 dB/km	0.2 dB/km
1.0 mm/hr	0.25	0.6	0.7
4.0 mm/hr	1.0	3.0	3.2
16.0 mm/hr	4.0	8.0	9.0
Fog Attenuation			
Light 0.01g/m ³	0.006 dB/km	0.035 dB/km	0.07 dB/km
Thick 0.1 g/m ³	0.06	0.35	0.7
Dense 1.0 g/m ³	0.6	3.5	7.0
Apparent Sky Temperature			
Clear	23°K	50°K	81°K
Moderate Overcast	65	120	200
Rain	110	220	250

2 TECHNICAL DETAILS

2.1 Passive Millimeter Wave Imaging

A very increasing interest towards the design and production various PMMW imaging system is due to the real possibility to have fourth type of the remote control system in addition to the existing optical, radar and infrared (IR) systems. Attractive feature of PMMW imaging systems is the capability to operate under adverse weather conditions and to be sensitive to non-metallic targets. In addition, since PMMW sensor is passive, it can be operated in all locations including friendly and hostile ports where RF emissions may be disruptive to local systems. (Moffa et al, 2001).

Figure 6 presents images produced by the 32 sensors 8 mm PMMW system with antenna diameter 90 cm (Gorishnyak et al, 2004; Denisov et al, 2009). For this system, sensor sensitivity is 0,01 K for post detecting time 1 sec and the sensor works without any modulation-calibration at the entrance. This system according to the theory works with full power radiometers. In figure 6, it is worth noting that the black spots on the image of city, at the top of building, represent the mobile phone transmission stations, producing harmonics in 8 mm band. Radio-images in 90 GHz (3mm wavelength) frequency band presented on Figure 7.

PMMW system are now gaining of deep cooling based on superconductors for the space tasks (Proc.of SPIE,2014) and the improvement of the passive images, with the help of advanced data processing, for the super resolution (Luxin et al. 2006). In the near future, it is expected the modern application of the processed radio-images in the various spectral bands with the help of a correlation analysis.

2.2 Technical Peculiarities of the Radiometric Discriminator

Good performance of radio-image systems is based on the necessity to have identical sensors; this is not challenging in case of a direct amplifier based on the Monolithic Microwave Integrated Circuits (MMIC) with enough high dynamic range. Instead, various issues arises in the data processing, when we have to turn from the measurements with a large number of sensors in an image in a digital form or in optical up-converting “looks”. Issuers regard also special cooling (Moffa, 2001) and temperature stabilization, which really increases the cost of PMMW system. For example, prime cost of 8 mm sensor in Figure 4 has value around 700 \$, and for modern European direct amplifiers on 3 and 2,2 mm the cost is 5-20 times more expensive.

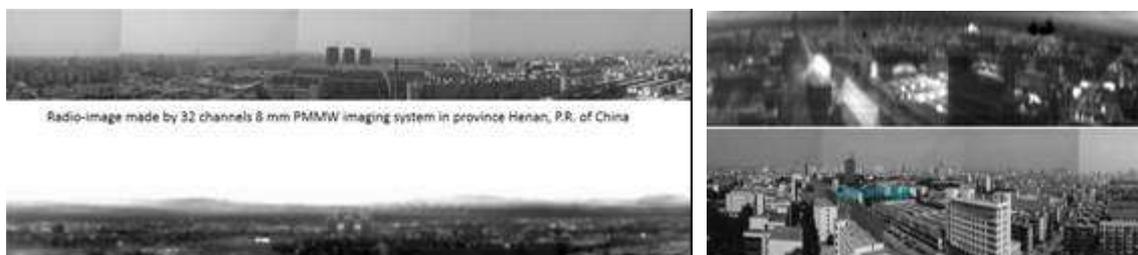


Figure 6: Radio-images in 8 mm wave band in comparing with the optical images of the same scenes.



Figure 7: Radio-images in 3 mm wave band

sensors, because the relevant angle of view is enough small in this case.

Here, we focus on the differential modulation radiometer, which is also named by discriminator.

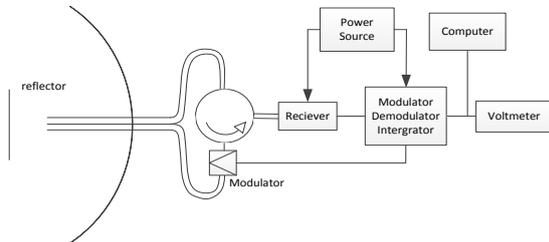


Figure 8: Block-scheme of the simplest 8 mm microwawe discriminator.

The discriminator (see Figure 8) exploits two feeders combined with the one antennas surface. The special antenna system forms two space beams directed on the two neighboring directions of the space in the horizontal plane (the same device can be done for the vertical plane). The angle between the two beams depend on the technical specification. The incoming signals from these two directions are at the input of the switch and later to the radiometer

(Figure 8). Usually, it is convenient to deploy a circulator in replacement of the switch where radiometer is connected to the third gate of the circulator. In this configuration, at the input of the radiometer there are two microwave signals at the modulation frequency (for example, 10 kHz). After amplification (around 56 dB), the signal are detected and are going as the meander with modulation frequency. Afterward, the resulting signal can be demodulated by the synchronous detector with time constant around $\tau = 0.01\text{sec}$. Finally, at the output of the integrator (discriminator), there is a signal proportional to the difference of the power in the measured beams. This signal difference arises only in case of observation of different observed scenes. Therefore, as the result of the scanning, there is a picture that resembles the two neighboring angles of the space and accounts for the contour of the observed object.

2.3 The Job of the Discriminator

For a simple explanation of the working principle of the discriminator, we can refer to Figure 9.

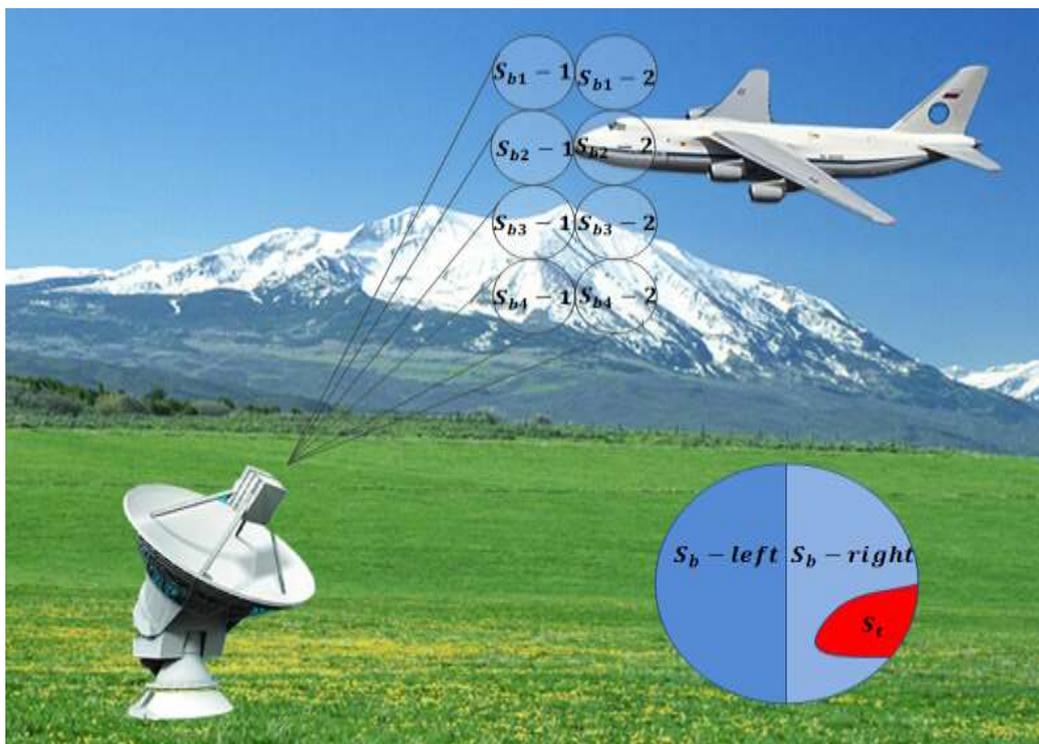


Figure 9: Pictorial description of the working principle of a discriminator.

There the four channels PMMW discriminator system produces the scanning of scene. The first

spot of the antenna beam S_{b-left} , does not intercept any part of the observing object whereas the object

There the four channels PMMW discriminator system produces the scanning of scene. The first spot of the antenna beam S_{b-left} , does not intercept any part of the observing object whereas the object is imaged in the second right spot $S_{b-right}$.

For an evaluation of these two following observing spots, the brightness temperatures T of the object is used. The brightness temperature can be evaluated according to the expression concerning the microwave power $P = kT\Delta f$, where k – Boltzmann’s constant, T – brightness temperature, Δf – the band of the receiving frequencies, which income to the radiometer from the two antenna’s spots. Of course, this task requires a critical technology (Huguenin,2006, Dill et al,2009). In Eq.1 there is $S_{back} = \pi A^2$, where A is the area of the aperture antenna beam spot at the distance L till the aircraft. Other main parameters are: the brightness temperature of background (cold sky) is T_{back} ,

(concerned with S_{b2-1} , S_{b1-1} , S_{b1-2} on fig.9), this value is dependent on the atmosphere conditions and in particular on the absorption in atmosphere, which increases the brightness temperature of sky compared to the case of clear air. The diameter of the antenna’s beam spot is $A = 1,22 L \sin \lambda/D$. S_t – area of the aircraft surface inside of the beam spot S_{b2-2} with the brightness temperature of the aircraft depending on the reflection from the earth or water T_t .

In the inset of Figure.9, we schematically present two parts of the antenna beam spots S_{b-left} and $S_{b-right}$. Both S_{b-left} and $S_{b-right}$ are two half of the full surface of S_b . Discriminator compares the received signals from these two beam spots; the difference between the received microwave power in terms of the brightness temperatures of these two spots can be evaluated in the simplest approximation as :

$$\left| T_{back} - \frac{T_{back} (S_{b-right} - S_t) + T_t \cdot S_t}{S_{b-right}} \right| = \left| T_{back} - T_t \right| \cdot \frac{S_t}{S_{b-right}} \quad (1)$$

So, if the receiver of discriminator has a sensitivity better that the result of eq. (1), it is possible to detect the distinction inside of this direction and the objects can be found by PMMW discriminator system.

Every antenna has the efficiency η (for example, in percents). On the track object – radiometer, attenuation in atmosphere and losses from the antenna to the input of the radiometer can be accounted by the parameter α (expressed in dB or times). Every measuring system must have a reserve in the probability of detection of κ (signal/noise).

If we consider $\Delta T_{Contrast} = T_{back} - T_{sky}$, it is possible to recognize an aircraft, under a simple approximation in case of :

$$\Delta T_{Contrast} \geq \kappa \alpha \Delta T_{sens} S_{b-right} / S_t \eta \sigma \quad (2)$$

Obviously, it is desirable to have a large contrast $\Delta T_{Contrast}$, which depends on the environmental conditions, the brightness temperature of sky (Figure 2), the polarization effects, the working frequency, the material and geometry of the reflecting surfaces of the observed flying objects and position of observing object of relatively horizon, zenith and sun. Right part of Eq.2 depends on the sensitivity

ΔT_{sens} , microwave losses in atmosphere α and $(S_{b-right}/S_t)$. If we consider the sensitivity of radiometer for post-detection time 1 sec, real value is 0,01 K for 8 mm and in 3 mm practically too. There are no big problems to do the radiometric system at the base of modern MMIC with good thermal stabilization, which is the key factor to filter out possible amplification drift between the multitude of sensors. In principle by using MMIC with more small noise factor or special cooling, it is possible to reach best sensitivity, as for example, for the space investigation (Proc. of 22th ISSTT), but this is not of interest in the application considered in this paper. By turning to the value $S_t/S_{b-right}$ which defines the percent of the filling an aperture antenna’s beam spot by the surface of aircraft, which according to aerodynamics must have enough big figure in comparing with an armor objects, for example, on the earth (www.giws.de). Table 2 presents the diameter of the antenna’s beam spot at the various distances from an antenna as the function of its diameter. According to the picture in fig.5 and Table 2, it is possible in good approximation to evaluate the real microwave losses on path between radiometric discriminator and the aircraft to be observed.

Table 2: Diameter of antenna beam spot versus the distance from the antenna for the 8 mm wavelength

Distance (km) \ Diameter of antenna (m)	2	4	5	10	15	20
0.9	22	43	57	108	162	216
1.0	19	39	49	97	146	194
1.2	16	32	40	81	121	162
1.5	13	26	32	65	97	130
2.0	10	19	24	49	73	97

2.4 Simplest Calculation of the Possible Distance

Let's try to use Eq.2 in two real cases, concerning possible contrast levels $\Delta T_{Contrast}$ as 250 K and 100 K (fig.9).

Range of radio brightness contrast $\Delta T_{Contrast}$ for the situation presented in Figures 2,3,9 and Table 1 can be from 23 till 270 K.

Real parameters of the discriminator:

- Radiometer sensitivity is 0.01 K for post detection time 1 sec, at the real time for the analysis (doing pixel) as $\tau = 10$ msec, ΔT_{sens} for this scanning rate will be 0.1 K.

- Wavelength $\lambda = 8$ mm

- The diameter antenna discriminator D , for example - 200 cm (not so big problem to do it for 8 mm, if the accuracy of the surface must be worse that $\lambda/10$). Diameter of an antenna spot at the distance L is

$$\frac{0,1 K <}{50 \times 10^4 \text{ cm}^2 \times 0,8 \times 0,8 \times (100 \dots 250 \text{ K}) \times (200 \text{ cm})^2 \times 4 / 5 \times 6 \times 0,5 \times 3,14 \times L^2 \times (1,22)^2 \times 0,64 \text{ cm}^2} \quad (3.1)$$

$$L^2 < (100 \dots 250) \times 50 \times 0,64 \times 16 \times 10^8 / 15 \times 0,64 \times 4,68 \times 0,1 \quad (3.2)$$

..

According to (3-1) the value L for the best case (contrast equal to 250 K) will be **16,87 km**. For $\Delta T_{Contrast}$ of 100 K, the distance decrease at about 10,67 km. These evaluations have been made under the assumptions of $S_t = 50 \text{ m}^2$, but according to Internet the real wing surface of the left aircraft on

$A = 1.22L (\lambda/D)$ and accordingly S_b is $\pi A^2 / 4 = \pi [1.22L (\lambda/D)]^2 / 4$. $S_{b-left} = S_{b-right} = 0,5 \pi A^2 / 4$, according to inset on fig.4.

- The antenna efficiency, for example, $\eta = 0,8$.
- The factor of the object position σ ($\sigma = 0,8$)
- The size of the appearing object inside of the antenna beam spot S_t is $5 \times 10 \text{ m}^2$ (middle size ship).
- The probability of detection of κ (S/N) = 3-10 times. (Accept $\kappa = 5$)
- The attenuation or microwave losses between discriminator and observed ship α (3....10 dB). (Accept $\alpha = 6$ times).

In this case it will be Eq.3:

$$\Delta T_{sens} < S_t \eta \sigma \Delta T_{Contrast} / \kappa \alpha 0,5 \pi (L 1,22 \lambda/D)^2 / 4 \quad (3)$$

Arithmetical calculations will provide

the Figure 1 is 73 m^2 , and the right one has surface 478 m^2 !

It is worth noting that for the evaluations in the case of a UAV, if we use the reflecting surface of about 1 m^2 (really it is more smaller for the observing UAV), the value L will be around 2,4 Km for the same discriminator antenna size.

3 CONCLUSIONS

For the cases where the PMMW kvazi image is not so principle it can be used simple microwave discriminator which is variety of the differential modulation radiometer for the detection of an objects. In this case, the receipted results can repeat the contour of the observing objects.

REFERENCES

- Reinwater J. 1978, Radiometers: Electronic Eyes that "See Noise", *Microwaves*, September, p.58-62.
- Moffa P., Yujiri L., Agravante H., deAmici D., Fornaca S., Jackson C., Jaeger T., Jordan K., Quonh., Rasmussen K., Samec T., Shoucri M., 2001, A large aperture passive millimeter wave pushbroom camera, *Proc. of SPIE*, vol.4373, p. 1-6
- Goldsmith P., Hsieh C., Huguenin G., Kapitzky J., Moore E. 1993, Focal plane imaging systems for millimeter wavelength", *IEEE Transaction on Microwave Theory and Techniques*, Vol. 41, 10
- Appleby R., Lettington A. 1991, Passive millimeter wave imaging, *Electronics & Communication Engineering Journal*, February, p.13- 16
- Piechl M., H.Suss H., Greiner S., Jirousek M. 2004, Imaging Technologies and Applications in Microwave Radiometry, *EuRAD - 2004, Amsterdam*, p.269-272
- Paradish J., Habbe J., 1982 Millimeter wave radiometric imaging, *Proc. of SPIE*, vol.337, p.170-181
- Proc.of SPIE , Passive Millimeter Wave Imaging Technology, Numbers- 3064, 3378,3703, 4032, 4373, 5077, 5619, 5789, 6211,6548,7837, 7854, 8544, 8716, 9253, 9252.
- Esepkina N., Korol'kov D., Pariiskiy Y., 1973 [*Radioteleskopy i radiometry (Radiotelescopes and Radiometeres)*], Nauka Press, Moscow
- Skou N., 1989, Microwave Radiometer Systems, *Design and Analysis*, Artech House.
- Schuchardt J., Newton J., Morton T., Galliano J., 1981, The coming of mm-wave Forward Looking Imaging Radiometers, *Microwave Journal*, June, p.45-62
- Schuchardt J., Stratigos J., 1978, Detected noise levels guide radiometer design, *Microwaves*, September, p.64-7
- Moore R., Hawthorn C., Hoover M., Gravlin A., 1976, Position Updating with Microwave radiometric sensors, *NAECO-76, Record, USA*, p.13-19
- Parnell W., 1988, Multiband Design boosts resolution of imaging radar, *Microwaves & RF*, September, p.85-94.
- Seashore C., Miley J., Kearns B., 1979, MM-wave radar and radiometer sensors for guidance systems. *Microwave Journal*, August, p. 47-58.
- T. Corrado T., 1988, Smart munitions-are they affordable. *MSN*, December, p.32-36, Site: www.giws.de
- Huguenin G. 2006, The detection of hazards and screening fir concealed weapons with passive millimeter wave imaging concealed threat detectors, *Press release*, Site: www.millivision.com
- Appleby R., Wallace H.B., 2007, Standoff detection of weapons and contraband in the 100 GHz to 1 THz region, *IEEE Trans. on Ant. and Propagation*, vol.55, no.11, p.2944-2956 Sites: www.brijot.com, www.millivision.com, www.xytrans.com, www.thruvision.com
- Dill S., Piechl M., Jirousek M., Suss H., 2009, Further analysis and evaluation of the results of Common Shield –DAT № 7 experiment- Defence Againstbn Terrorism, *Proc.of SPIE*, vol.7485, 74850B-2. Site: en.wikipedia.org/wiki/Stealth_ship
- Gorishnyak V., Denisov A., Kuzmin S., Radzikhovskiy V., Shevchuk B., 2004 , 8 mm Passive Imaging System with 32 channels , *EurRAD , Amsterdam, Netherland*, 2-14 October.
- Denisov A., Gorishnyak V., Kuzmin S., Miklashevich V., Obolonsky V., Radzikhovskiy V., Shevchuk B., Yshenko B., Uliz'ko V., Son J.Y., , 2009, Some experiments concerning resolution of 32 sensors passive 8mm wave imaging system, *Proceedings of the International Symposium on Space Terahertz Technology, Charlottesville, USA*.
- Proc. of 22th International Symposium on Space Terahertz Technology, 2011, Tucson, USA, and 2014, *Proc. of SPIE*, vol. 9153
- Luxin Y., Tianxu Z., Sheng Z., Jian H., Jianmao Z., 2006, Study on multichannel passive millimeter-wave radiometric imaging and superresolution, *Int. J. Infrared Milli. Waves*, vol.27, p. 1403-1414
- Salmon N., Masan I., Wilkinson P., Taylor C., Scinluna P., 2010, First imagery generated by near-field real-time aperture synthesis passive millimeter wave imagers at 94 GHz and 183 GHz, *Proc. of SPIE*, vol.7837, 78370I
- Salmon N., Wilkinson P., Taylor C., 2012 Interferometric aperture synthesis for next generation passive millimeter wave imagers, *Proc. of SPIE*, vol.8544, 854405

**SIGNAL
PROCESSING FOR NEXT
GENERATION
NETWORKS**

Capacity Analysis of Radio Frequency Interconnect for Manycore Processor Chips

Eren Unlu, Christophe Moy, Yves Louet and Jacques Palicot
CentraleSupélec-IETR, F-35576 Cesson-Sevigne Cedex, France
eren.unlu@supelec.fr

Keywords: Capacity analysis, Manycore processor chip, OFDMA.

Abstract: We study here the case of a 2048 cores chip, where cores are spread into 32 tilesets of 16 tiles containing 4 cores each. Each of the 32 tilesets has a RF access point to the serpentine transmission line across the chip. Inside tilesets, a 2-D mesh is used between the 16 tiles where a crossbar switch joints 4 cores, one RAM block and one DMA unit. Total 1 Terabytes memory is physically distributed into all tiles but logically shared between all cores and managed by a distributed hybrid cache coherency protocol (DHCCP). From the RF-NoC point of view, a 20 GHz bandwidth between 20 and 40 GHz is shared into 1024 carriers between all 32 RF access nodes. The novelty of our work is that we have derived, in previous publications, algorithms able to dynamically share the RF resources between the 32 nodes. It has been stated by simulations that the channel transfer function is flat in the 20-40 GHz frequency band and just depends on the distance between nodes. The scope of this paper is to make a capacity analysis on the different links between nodes and to derive mean capacity evaluation of the RF NoC. We state that only -42 dBm of transmission power on the RF line is necessary to reach a 6 bits/s/Hz spectral efficiency.

1 INTRODUCTION

In order to drive the ever increasing computational demands of the 21st century, multiprocessors are being preferred more and more over single processors which have reached their limits due to thermal and physical issues. With developing lithographic techniques and semiconductor technologies, the number of cores in a single chip is expected to reach *thousands*, before the end of next decade (Borkar, 2007). These architectures constituted of *sea of processors*, where each of them are simpler, lower frequency cores providing higher computational power by exploiting parallelism, power efficiency and robustness. They are referred as *Chip Multiprocessors (CMPs)* or as *manycore processors* in the community (Olukotun et al., 2007).

Traditional, relatively simple communication mediums for connecting on-chip processing elements were buses or crossbars. However, with increasing number of cores it has become impractical to implant dedicated point-to-point wires and congestion problem has arisen with the buses. Researchers had to introduce a new framework known as Network-on-

Chip (NoC), where communication layer is detached from the data generated by on-chip nodes, and packetized transmission is performed via buffered routers as it can be done in large scale telecommunication networks (Cota et al., 2011). This modular approach does not only increase bandwidth, but also enables a wider spectrum of choice for designers, as various topologies, routing and arbitration algorithms can be applied. NoC has changed the approach of on-chip community to the problem, where brand new research interests have emerged such as optimization and dimensioning these router based architectures. However, as hundreds, thousands of cores are on the horizon, conventional electrical networks are not able to sustain the communication demands of these massive chips in terms of latency, bandwidth and power efficiency.

In order to provide the necessary breakthrough, designers have focused on developing optical and RF interconnects, recently (Pasricha & Dutt, 2010). These interconnects serve as high bandwidth, low latency *communication highways* between each core or group of several cores. Photonic interconnects are considered as an effective technology to reduce the latency, however they require constantly operated on-

chip or off-chip laser sources and they are incompatible with the CMOS technology. On the other hand proposed RF interconnects are based on fully CMOS compatible components and a mature technology (Deb, 2012).

In this paper, firstly we present our hierarchical 2048-core CMP with its *Orthogonal Frequency Division Multiple Access* (OFDMA) based RF interconnect. State-of-the-art optical and RF on-chip interconnects rely on numerous amount of electronic circuits such as microring resonators, local oscillators etc. to generate orthogonal communication channels, which limit their scalability. However, proposed OFDMA interconnect in this paper has the potential to overcome the scalability issue by encoding data on frequency domain digitally by not requiring high amount of circuitry, providing broadcasting capability and high bandwidth reconfigurability, thanks to the cutting edge components being designed in the vicinity of the project. After we present the multiprocessor architecture, we introduce certain details of the RF interconnect connecting 32 tilesets (group of several cores) and the attached RF frontends on each of these tilesets. The main aim of this paper is to study the information theoretic capacity and determine the associated minimum transmission powers for this on-chip RF interconnect.

2 CMP ARCHITECTURE AND OFDMA INTERCONNECT

2.1 WiNoCoD Project

Considering the CMOS incompatibility of optical interconnects and lack of efficient, reconfigurable bandwidth allocation of existing designs, *Wired RF Network-on-Chip / Reconfigurable on Demand* (WiNoCoD) project has been initiated in 2012 by the French national research agency (ANR) (Briere et al, 2015). It uses an innovative Orthogonal Frequency Division Multiple Access (OFDMA) based communication via RF signals on a shared transmission line.

OFDMA offers the ability to redistribute the bandwidth among sharers in nanosecond scale rapidness, with fine granularity on digital domain. In contrast with other state-of-the-art FDMA interconnects, it does not require to employ massive number of redundant static CMOS elements such as filters, mixers etc. and does not rely on these circuits to share bandwidth dynamically.

2.2 3-level Hierarchical CMP Architecture

Our project provisions a 2048 core generic massive CMP. A shared memory principle is adopted, so that the address space is accessible by all processing elements, where a 1 TByte of total RAM is distributed evenly to 512 tiles. A global Distributed Hybrid Cache Coherency Policy is preferred (Lija, 2016; Kurian et al., 2010). A hierarchy of 3 levels is chosen for scalability, where each level has a different type of interconnect concerning the needs. At the lowest level 4 generic cores, a 2 GByte slice of the total RAM and a memory controller are grouped and interconnected by an electric crossbar, where this entity is called as a tile. At the next level, each 16 of these tiles are grouped as a 2-D mesh topology and connected via electrical routers, where we refer it as a tileset. Finally, 32 tilesets are interconnected via a guided RF transmission line. Each of these tilesets have a necessary RF front-end to transmit and receive information on this interconnect. The hierarchical architecture of our CMP is illustrated in Fig. 1.

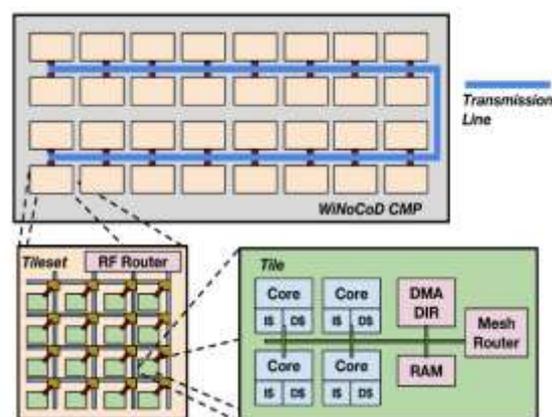


Figure 1: 3 level hierarchy of our CMP architecture incorporating 2048 cores. Each level has a different dedicated NoC infrastructure.

2.3 OFDMA Based RF Interconnect

In the WiNoCoD chip proposed in (Briere et al., 2015), 32 tilesets are interconnected via an serpentine, U-shaped, state-of-the-art RF microstrip transmission line for the inter-tileset communication. Close ended, circular loop transmission line shape is avoided in order not to cause self-interference. The packets that are generated inside a tile in tileset, which are destined to a tile in another tileset, traverses the electrical mesh network and reaches to the RF

access point. In reception, the received packets are checked, treated and send to interior tiles by the mesh routers. Provided converter technology by our partner NXP envisions a 20 GHz bandwidth for the system. Based on the design constraints and circuit simulations, most suitable spectrum is chosen between 20-40 GHz (Hamieh et al., 2014).

It is decided to have 1024 subcarriers, thus 1024-point FFT and IFFT blocks are required. Hence, as we have a 20 GHz bandwidth with 1024 subcarriers, we have subcarrier frequency spacing of 19.53 MHz, where a symbol duration is $T = 1/19.53 \text{ MHz} = 51.2 \text{ ns}$. Fig. 2, illustrates the architecture of the RF front-ends. The up-conversion mixers combine a baseband signal with a local oscillator signal. Mixing occurs in a MOSFET, whose gate and drain are respectively fed by the local oscillator and the baseband signal. As the local oscillator frequency is 30 GHz, which is the middle of our 20 GHz bandwidth, it needs to be suppressed. Thanks to the differential outputs of the DAC, two IQ-Modulators can work together to do so. Besides avoiding interference caused by image frequencies they can reduce the LO level in the output. As we use the same local oscillator for both of IQ-modulators and opposite I-Q signals, the IQ-Modulators outputs are subtracted in a differential amplifier to perform this suppression. Then this signal is amplified by a Low-Noise Amplifier (LNA) and transmitted on waveguide.

The reception is done synchronously every $T = 51.2 \text{ ns}$ as transmission, too. The received signal from the transmission line is amplified and fed to a separator circuit, mixers and 30 GHz local oscillator to obtain in-phase and quadrature components. Low Pass Filters (LPF) are used for down-conversion. Then I and Q components are converted to digital domain by our Analog-to-Digital (ADC) components. After Serial to Parallel conversion this vector of I and Q values are converted to frequency domain by an FFT block. Utilized FFT/IFFT processors are estimated to be manufactured with 120 nm CMOS technology. We estimate the area of each of these modules as 0.31 mm^2 and power consumption of 67.5 mW. Each of ADCs and DACs are designed with 120 nm technology and have an estimated surface area of 0.12 mm^2 and power consumption of 81 mW (Briere et al., 2015). It was shown that WiNoCoD interconnect has a 0.2-0.3 dB/mm attenuation with distance over 20-40 GHz band. These results are derived in the scope of WiNoCoD project (Briere et al., 2015; Hamieh et al., 2014).

3 INFORMATION THEORETIC ANALYSIS OF THE PROPOSED RF INTERCONNECT

3.1 RF Interconnect Capacity Derivation

In this section, we introduce a brief analysis of achievable communication capacities between tilesets and the associated minimum transmission powers based on the *information theory*. Information theory, which is founded by C. Shannon's seminal paper (Shannon, 2001) provides the bound for maximum achievable transmission rate on a communication channel with the given signal power, where the probability of error approaches the zero. This theoretical bound is independent of the utilized signal protection or correction mechanisms and may provide a good insight for designers for dimensioning a reliable communication on a channel. The information theoretic capacity of a channel can be written in bits/sec as:

$$C_0 = B \log_2(1 + SNR) \quad (1)$$

where B is bandwidth in Hz and SNR is Signal-to-Noise Power Ratio in linear. The power of the noise P_N , depends on the temperature and bandwidth. SNR can be written as the ratio of the received signal power to the noise power as $P_R = P_N$. P_N is the Additive White Gaussian Noise (AWGN) power in the bandwidth. The AWGN power spectral density in standard room temperature is -174 dBm/Hz, which we also accept this value in our calculations (Shankar, 2002).

As we have a immobile and minuscule environment in contrast with general wireless communications, we can assume that the only loss on transmitted signal power is due to distance between tilesets. As the frequency response is relatively nonfluctuating, we can assume a single value for attenuation per distance over all bandwidth. For our calculations we assume a 0.25 dB/mm attenuation on the transmission line, which is the average of minimal and maximal values of 0.2 and 0.3 dB/mm. Hence, the received signal power can be written as the ratio of the transmitted signal power to the attenuation due to the distances between tileset- i and tileset- j :

$$P_R = P_T / A(d_{ij}).$$

d_{ij} being distance in mm between tileset- i and tileset- j , the resulting attenuation in dB becomes $0.25d_{ij}$. Converting this expression in scalar, we can rewrite the received signal power as a function of distance:

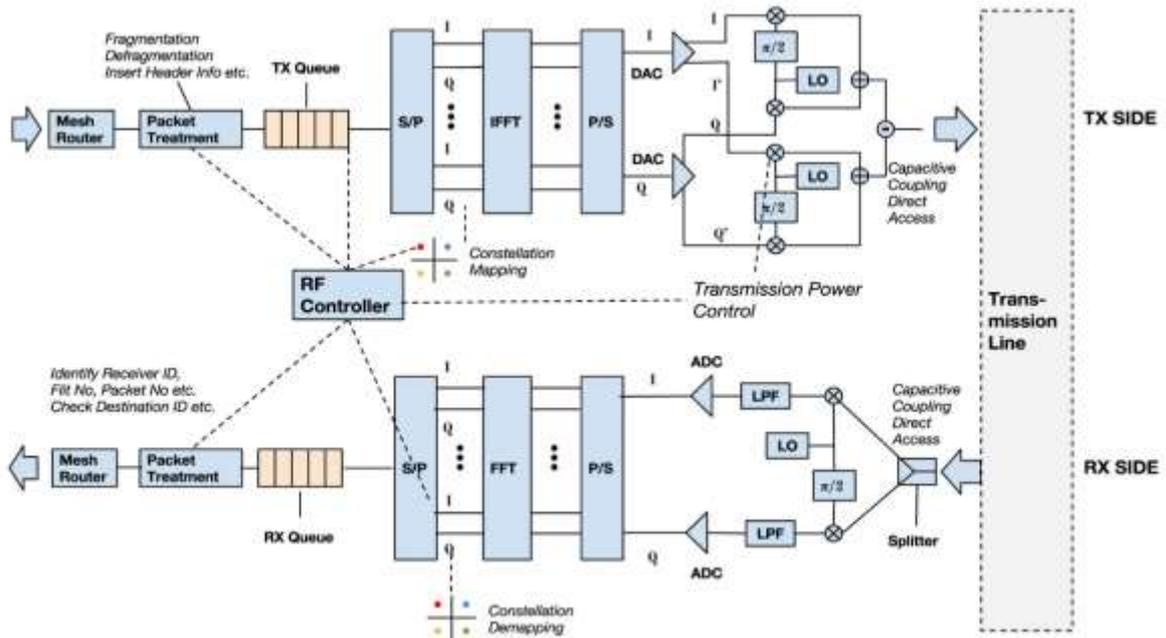


Figure 2: The detailed illustration of transmission and reception RF interface of a tileset.

$$P_R = \frac{P_T}{10^{(0.25d_{ij}/10)}} \quad (2)$$

Spectral efficiency in bits/sec/Hz, $SE = C_0 / B$, is a metric that we use more frequently, which defines the achievable transmission rate per bandwidth. By inverting the capacity formula, one can also derive the required minimum transmission power (linear), P_T as a function of desired spectral efficiency, SE :

$$P_T = P_N 10^{(0.25d_{ij}/10)} (2^{SE} - 1) \quad (3)$$

This calculated minimum transmission power shall provide a good rationale for users on the requirements of error-free reliable communication.

In this paper, we will analyze the information theoretic channel capacities and associated minimum transmission power values for our RF based on-chip interconnect. In addition to this, required transmission powers for different bit error rates for different modulation orders and uncoded communications are evaluated. These two type of indicators can be compared to dimension the required communication energy for the proposed on-chip RF interconnect. Adjacent tileset access points have a displacement of 8 mm and we assume that vis-a-vis

tilesets' access points have a displacement of 1 mm. We also assume that each tileset is allocated evenly 32 subcarriers for transmission, which corresponds to 640 MHz of bandwidth. Hence, the noise power for the bandwidth can be calculated by multiplying it with the assumed AWGN power spectral density and can be found as approximately -86 dBm. Note that one can derive desired figures for different bandwidths simply by scaling linearly. Information spectral capacity densities can be regarded as good indicators for the utilizable modulation orders with error free communication, such as 1 bits/s/Hz corresponding to BPSK, 2 bits/s/Hz corresponding to QPSK etc.

However, we also investigate the required transmission powers for various bit error rates for different modulation orders under uncoded transmission. Note that, information theoretic capacity defines the maximum achievable rate with a bit error rate approaching to 0, with a hypothesized perfect channel coding mechanism. Therefore, even though required transmission power for a desired channel capacity density is defined for an error rate of 0, it may require less power than the power required for various bit error rates. This is due to robustness of channel coding. Next, we develop the expressions for the required transmission powers for different modulation orders. For BPSK and QPSK, the BER or

probability of bit error (p_b) can be written as (Goldsmith, 2005):

$$p_b = Q\left(\sqrt{2SNR_b}\right) \quad (4)$$

where $Q(\cdot)$ is the Gaussian tail function and SNR_b is the received SNR per bit. For square M-QAM constellation symbols (such as 16-QAM, 64-QAM ..), p_b can be written approximately as (Goldsmith, 2005):

$$p_b = Q\left(\sqrt{\frac{3SNR_b b}{2^b - 1}}\right) \quad (5)$$

where b is the number of bits per constellation such as 4 bits for 16-QAM, 8 bits for 256-QAM etc. Let us calculate the SNR per bit at first. As we did in capacity formula in (2), we can write the received SNR as the ratio of transmission power to ambient noise power and attenuation by distance. For BPSK and QPSK. By using the assumption made in (Hamieh et al., 2014), a noise factor, F of 3 dB is used, which is approximately 2 in linear scale. Therefore, the linear noise power density can be calculated from the -174 dBm and the additional 3 dB noise factor as:

$$810^{-21} \sim 10^{(-174+3)/10} \quad (6)$$

By using this, we can calculate the BER values for BPSK:

$$p_b = Q\left(\sqrt{\frac{2P_T}{10^{0.025d} B 810^{-21}}}\right) \quad (7)$$

and for M-QAM:

$$p_b = \frac{4}{b} Q\left(\sqrt{\frac{3bP_T}{10^{0.025d} B 810^{-21} (2^b - 1)}}\right) \quad (8)$$

By inverting these equations we can calculate the required minimum transmission powers with a given BER. The minimum required transmission power for BPSK and QPSK:

$$P_T = 0.510^{0.025d} B 810^{-21} \left(Q^{-1}(p_b)\right)^2 \quad (9)$$

and minimum required transmission power for M-QAM constellations:

$$P_T = \frac{1}{3} 10^{0.025d} B 810^{-21} \cdot \left(Q^{-1}(0.25p_b)\right)^2 (2^b - 1) \quad (10)$$



Figure 3: Placement of tilesets with their ID number on the U-shaped transmission line.

The placement of tilesets with their ID number is shown on Fig. 3. For instance, one of the maximum distances among tileset connection is between tileset-1 and tileset-32, which is 120 mm (spacings of 15 adjacent tilesets). We have analyzed the information theoretic limits for each of the 32x31 unicast communication combination between tilesets. Fig. 4 shows the distances between tilesets for each of these combinations according to tileset ID numbers.

The capacity of the each transmission combination can be written in matrix format assuming transmission power P_T is allocated for each them, where d_{ij} is the distance between tileset- i and tileset- j in mm:

$$C_0^{ij} = B \log_2 \left(1 + \frac{P_T}{10^{(-174-3+30+10\log(B)+0.25d_{ij})/10}} \right) \quad (11)$$

3.2 Results

Combining (2) and (3), we derive the required minimum transmission powers in dBm for different spectral capacity densities from 1 bits/s/Hz to 6 bits/s/Hz corresponding to modulation orders between BPSK and 64-QAM. One can see that from Fig.5, for achieving a spectral efficiency of 1 bits/s/Hz between the most distant tilesets, we need approximately -40 dBm transmission power and between the closest tilesets, we need approximately -70 dBm transmission power. For achieving a spectral efficiency of 6 bits/s/Hz between the most distant tilesets, we need approximately -20 dBm transmission power and between the closest tilesets, we need approximately -50 dBm transmission power. Note that, the transmission power in dBm varies linearly with spectral efficiency and distances between tilesets, consistent with the equations above.

Fig. 6 shows the total required transmission power, considering each of the unicast communication combination has capacity densities 1-6 bits/s/Hz. We can see that for achieving a spectral efficiency of 1 bits/s/Hz, we need approximately -57 dBm total transmission power and for a spectral efficiency of 6 bits/s/Hz, we need approximately -42 dBm average transmission power.

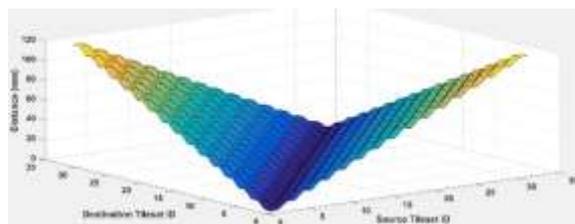


Figure 4: Distance between each 32x31 unicast communication pair in our U-shaped transmission line.

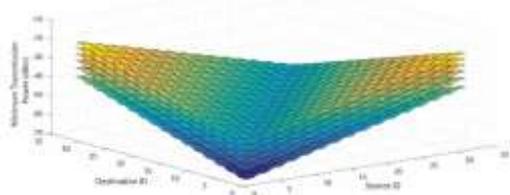


Figure 5: Required Transmission Power for each 32x31 unicast communication for the linear (U-shaped) shaped transmission line for capacity densities 1-6 bits/s/Hz.

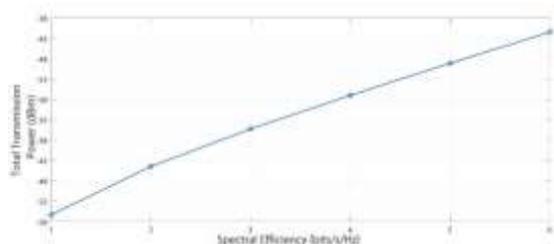


Figure 6: Total Required Transmission Power of 32x31 unicast communication for the U-shaped shaped transmission line for capacity densities 1-6 bits/s/Hz.

And finally, Fig.7 shows the minimum required overall transmission power from each tileset to their 31 destinations, for bit error rates of 10^{-1} ; 10^{-3} ; 10^{-5} and 10^{-7} for the lowest and highest modulation orders; BPSK and 64-QAM. As a reference, overall required transmission powers for information theoretic capacity densities of 1 bits/s/Hz and 6 bits/s/Hz (which are associated to BPSK and 64-QAM

respectively) are shown. Note that, minimum required overall power for bit error rates of 10^{-3} ; 10^{-5} ; 10^{-7} for BPSK and 64-QAM are higher than the power required for information theoretic capacity densities of 1 bits/s/Hz and 6 bits/s/Hz, respectively. This is due to capacity's definition for perfect channel coding as mentioned previously.

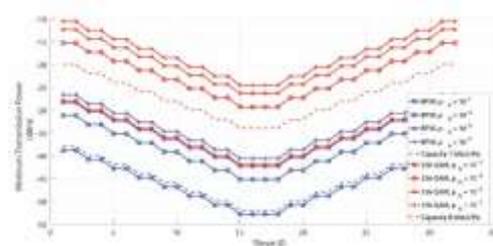


Figure 7: Average required transmission power from each tileset to their 31 destinations, for bit error rates of: 10^{-1} ; 10^{-3} ; 10^{-5} ; 10^{-7} under BPSK and 64-QAM and information theoretic capacity densities of 1 bits/s/Hz and 6 bits/s/Hz.

4 CONCLUSIONS

We have briefly presented the architecture of a 2048-core generic multiprocessor, which is being developed in the scope of WiNoCoD project, which attempts to employ an OFDMA based on-chip RF interconnect for the first time, to the best of our knowledge. With their matured manufacturing techniques, full CMOS compatibility and ever increasing transistor frequencies, RF interconnects are considered viable candidates of future massive on-chip platforms. We have given details of this 20 GHz OFDMA infrastructure to be used by 32 tilesets, where each of them incorporates 64 cores. Information theory defines the limits for the achievable transmission rate with given power budget and noise characteristics. Information theoretic capacities and the associated minimum transmission powers for this interconnect have been evaluated in this paper.

ACKNOWLEDGEMENTS

This work is supported by French National Research Agency (ANR) under No. ANR-GUI-AAP-05. The authors would also like to thank to LIP6, ETIS, ENSEA laboratories and NXP Semiconductors.

REFERENCES

- Borkar, S., "Thousand core chips: a technology perspective," in Proceedings of the 44th annual Design Automation Conference. ACM, 2007, pp. 746–749.
- Olukotun, L., K. Hammond, and J. Laudon, "Chip multiprocessor architecture: techniques to improve throughput and latency," Synthesis Lectures on Computer Architecture, vol. 2, no. 1, pp. 1–145, 2007.
- Cota, E., A. de Moraes Amory, and M. S. Lubaszewski, Reliability, Availability and Serviceability of Networks-on-chip. Springer, 2011.
- Pasricha S., and N. Dutt, On-chip communication architectures: system on chip interconnect. Morgan Kaufmann, 2010.
- Deb, S., "Millimeter-wave wireless network-on-chip: A cmos compatible interconnection infrastructure for future many-core processors," Ph.D. dissertation, Washington State University, 2012.
- Brière, A., J. Denoulet, A. Pinna, B. Granado, F. Pêcheux, E. Unlu, Y. Louët, and C. Moy, "A dynamically reconfigurable rf noc for manycore," in Proceedings of the 25th edition on Great Lakes Symposium on VLSI. ACM, 2015, pp. 139–144.
- Lilja, D. J. "Cache coherence in large-scale shared-memory multiprocessors: issues and comparisons," ACM Computing Surveys (CSUR).
- Kurian, G., J. E. Miller, J. Psota, J. Eastep, J. Liu, J. Michel, L. C. Kimerling, and A. Agarwal, "Atac: a 1000-core cache-coherent processor with on-chip optical network," in Proceedings of the 19th international conference on Parallel architectures and compilation techniques. ACM, 2010, pp. 477–488.
- Hamieh, M., M. Ariaudo, S. Quintanel, and Y. Louët, "Sizing of the physical layer of a rf intra-chip communications," in Electronics, Circuits and Systems (ICECS), 2014 21st IEEE International Conference on. IEEE, 2014, pp. 163–166.
- Shannon, C. E. "A mathematical theory of communication," ACM SIGMOBILE Mobile Computing and Communications Review, vol. 5, no. 1, pp. 3–55, 2001.
- Shankar, P. M. Introduction to wireless systems. Wiley New York, 2002.
- Goldsmith, A. Wireless communications. Cambridge university press, 2005.

Hidden Markov Model Traffic Characterisation in Wireless Networks

Evangelos Spyrou and Dimitris Mitrakos

School of Electrical and Computer Engineering, Aristotle University of Thessaloniki,
Egnatia Odos, Panepistimioupoli, 54124 Thessaloniki, Greece
{*evang.spyrou, mitrakos*}@eng.auth.gr

Keywords: Information Based Similarity, Euclidean distance, Wireless Traffic, Shannon Entropy, Signal-to-Interference-and-Noise-Ratio.

Abstract: Quality of service wireless traffic that often exhibits burstiness, occasionally occurring due to mobility, provides a critical networking issue. Traffic patterns in wireless networks are not of a traditional nature. Nodes transmit their information in batches over a short period of time, before they lose connection. Prediction of wireless incoming load plays an important role in the design of wireless local area networks. The issues of load balancing and Quality of Service constraints are a major problem, which is responsible for the increase of throughput of the network; thus, predicting traffic can be of a great assistance in the aforementioned research directions, leading to a significant optimisation of the wireless network operation. This paper addresses the problem of traffic prediction using Hidden Markov Models. The data is clustered using the Information Based Similarity index that classifies different types of traffic. We show the limitation of this approach and we finally select Euclidean distance for data clustering. Together, they provide an efficient solution towards the solution of wireless traffic characterisation and prediction. We show the efficiency of our scheme in a series of simulations

1 INTRODUCTION

In the field of wireless networks, requirements such as quality of service over normal or bursty links, often originating from mobility, are of great significance (Jiang and Dovrolis, 2005; Alizai et al., 2009). Wireless devices do not always follow specific traffic patterns; on the contrary, they attempt to transmit their packets in a bursty fashion, meaning all of them in a short period of time, before they lose or make the connection not reliable (Papadopoulos et al., 2015). Such dynamic and bursty traffic cause certain deficiencies in the network and provide an interesting problem to investigate.

Wireless traffic prediction is a key factor in the analysis and design of wireless local area networks (WLAN)s (Papadopouli et al., 2005). For example, the well-known problem of load balancing can be addressed using Access Point (AP) traffic prediction. In this way, a new connection with this AP may be decided based on the prediction of the AP load. This

is an improvement of the throughput of the network; furthermore, Quality of Service (QoS) constraints can be satisfied.

Recently, wireless network traffic has proven to exhibit self-similarity or long-range dependence (Park and Willinger, 2000; Jiang et al., 2001). Self-similarity in traffic is crucial and cannot be analysed by traditional models. Traffic prediction from past measurements constitutes an efficient way to acquire traffic control when self-similar loads occur. Optimality in forecasting still remains a major issue (Beran, 1994). This constitutes the location of efficient self-similar models for prediction of future traffic fluctuations a fundamental problem.

In this paper we address the problem of classifying wireless traffic by attempting to cluster the data utilising the information based similarity (IBS) index (Yang et al., 2003). We show the efficiency of this approach as well as the problems that may arise when integrating it to a Hidden Markov Model (HMM). Based on the limitations of this

approach we select the Euclidean distance to cluster the data in our machine learning approach.

We attempt to characterise the pattern of the traffic by feeding the packets generated by a wireless network to a Hidden Markov Model. The distances we find using the Euclidean distance assist us to cluster the data to normal, bad and bursty traffic. Combining the two, we are able to provide substantial information and a good characterisation and prediction on the traffic we will be experiencing. More specifically, we show the following contributions:

- We show the difference in similarity of different types of wireless traffic using the IBS index;
- We show that the IBS index can be a good methodology to cluster the incoming data and characterise traffic;
- We show the limitations of this approach when integrating it to an HMM
- We employ Euclidean distance to cluster the traffic data.
- We utilise the HMM to characterise and predict the upcoming traffic and show indexes of the efficiency of our approach;
- We show in a series of simulations the efficiency of our approach;

This paper is structured as follows: Section 2 provides the related work, Section 3 gives the background of the IBS index and some results, Section 4 provides a brief background on the Euclidean distance, Section 5 gives a background on our HMM proposal, Section 6 provides the simulation results of our approach, and Section 7 provides the conclusions.

2 RELATED WORK

In (Ni et al., 2015), the authors address the resource management for dynamic control of channel resources and energy efficiency in modern cellular systems. They identify that early monitoring and forecasting of basestation traffic volumes play a key role in the management. Spatiotemporal network traffic analysis is the means to a successful prediction of traffic. To this end, they examine the spatiotemporal features of cellular traffic generated in a practical scenario in China. The authors analyse basestations that exhibit similar features and they cluster them. They elaborate on the sliding windows

sizes and encapsulate the Elman Neural Network with wavelet transform to accomplish traffic precision.

In (Maheshwari et al., 2013), the authors aim to tackle the issue of Quality of Service of end-user calls for realistic traffic models. They mainly address modelling wireless internet traffic using realistic traffic traces. This traffic are collected from networks and they perform forecasting on the end-to-end Quality of Service parameters for the networks. The traffic model is designed based on Hidden Markov Model by taking into account the joint distribution of end-to-end delay, packet variation and packet size. The states that are identified are mapped to four traffic classes. These are conversational, streaming, interactive, and background.

In (Yadav and Balakrishnan, 2014), the issue of traffic modeling is investigated. Typical networking issues, such as resource allocation, quality of service, bandwidth and congestion control are addressed. A comparison of modeling techniques is carried out of adaptive neuro fuzzy inference system (ANFIS) and autoregressive integrated moving average (ARIMA) for modeling of wireless network traffic in terms of typical statistical indicator and computational complexity. Furthermore, a comparative performance evaluation is undertaken in traffic modeling showing that ANFIS constitutes a good methodology for prediction with respect to statistical indicators. Moreover, it provides a reasonable description of the conditions of a wireless network in the time domain. The main drawback is the complexity ANFIS introduces, even though it performs better than the ARIMA model in the scenarios they investigate.

In (Li et al., 2014), the authors aim to examine traffic prediction on cellular radio networks. The need for a traffic-aware energy efficient architecture is highlighted. To this end, traffic prediction is modeled a traffic-aware networking is addresses. For the former, entropy theory is utilised towards the analysis of traffic predictability. In terms of the latter practical prediction performance is demonstrated using the best methodologies in the literature. Finally, the authors suggest a blueprint regarding a traffic-oriented software-defined cellular radio network architecture and they show the potential applications of traffic prediction in this architecture.

In (Rutka and Lauks, 2015), the utilisation of neural networks for internet traffic prediction is investigated. Specifically, the investigate traffic prediction in the presence of self-similarity, which is an important feature of traffic in high-speed networks that may not be obtained by traditional traffic models. The major aim of this work is the performance and

prediction error investigation using feed forward neural networks.

In (Loumiotis et al., 2014), the efficient management of the backhaul resources in 4G networks is examined. The authors raise this issue in the case that the backhaul network has been leased by the mobile operator. Hence, the backhaul resource allocation issue at the basestation is investigated and aggregate traffic demand scheme is proposed using artificial neural networks. Finally, the authors provide evidence of the efficiency of their scheme in terms of absolute percentage error of downlink and uplink traffic.

In (Pan et al., 2013), the stochastic cell transmission model is extended, in order to include spatiotemporal characteristics of traffic and predict short-term traffic. Initially, the authors utilise a multivariate normal distribution-based best linear predictor as an auxiliary dynamical system predict boundary variables and/or supply functions. Thereafter these variables and functions are input to the stochastic cell transmission model for short-term traffic state prediction. Stochastic cell transmission model is relaxed by utilising the covariance structure calibrated from the spatial correlation analysis for probabilistic traffic state evaluation. Prediction is carried-out in a rolling horizon manner, which is handy for setting the predicted traffic state using real-time measurements.

3 IBS FOR WIRELESS TRAFFIC

We consider a wireless network where nodes communicate with respective APs to transmit their load. We assume that the network is mobile; hence, there is a possibility of loss of connection and transmission in a bursty fashion. This is reflected in the Signal-to-Interference-plus-Noise Ratio (SINR) between the transmitter and the receiver. We employ the SINR model appearing in (Spyrou and Mitrakos, 2015) to construct our case.

We denote as $\gamma_{k,j}$ as the SINR of the transmission for node k to node j and it is given by

$$\gamma_{k,j} = \frac{H_{k,j}p_k}{\sum_{t \neq k, j \neq t} p_t H_{t,j} + N_0} \quad (1)$$

where $H_{k,j}$ is the channel gain between nodes k and j, p_k is the transmission power of node k transmitting, p_t is the interfering node t's transmission power, $H_{t,j}$ is the channel gain between the interferer and the

receiver and N_0 is the noise. For a packet to be successfully received the following condition must be satisfied

$$\gamma_{k,j} \geq \gamma_{thr} \quad (2)$$

Where γ_{thr} is the SINR threshold for successful reception of the packet. In practical scenarios, a packet is successfully received when an acknowledgement is received by the sender. We consider a packet reception series $\{x_1, x_2, \dots, x_N\}$ where x_i is the packet i. We classify each packet into two states that represents the successful reception or not of the packet, as it can be identified by its acknowledgement value, as below

$$I_n = \begin{cases} 1, & \gamma \geq \gamma_{thr} \\ 0, & otherwise \end{cases} \quad (3)$$

We map $m + 1$ successive intervals to a binary sequence of length m, called an m-bit "word." Each m-bit word, w_k , therefore, represents a unique pattern of fluctuations in a given time series. By shifting one data point at a time, the algorithm produces a collection of m-bit words over the whole time series. Therefore, it is plausible that the occurrence of these m-bit words reflects the underlying dynamics of the original time series. Different types of dynamics thus produce different distributions of these m-bit words.

The resulting rank-frequency distribution, therefore, represents the statistical hierarchy of symbolic words of the original time series. For example, the first rank word corresponds to one type of fluctuation, which is the most frequent pattern in the time series. In contrast, the last rank word defines the most unlikely pattern in the time series. To define a measurement of similarity between two signals, we plot the rank number of each m-bit word in the first time series against that of the second time series.

If two time series are similar in their rank order of the words, the scattered points will be located near the diagonal line. Therefore, the average deviation of these scattered points away from the diagonal line is a measure of the "distance" between these two time series. Greater distance indicates less similarity and vice versa. In addition, we incorporate the likelihood of each word in the following definition of a weighted distance, D_m , between two symbolic sequences, S_1 and S_2 .

$$D_m(S_1, S_2) = \frac{1}{2^{m-1}} \sum_{k=1}^{2^m} |R_1(w_k) - R_2(w_k)| F(w_k) \quad (4)$$

where

$$F(w_k) = \frac{1}{Z} [-c_1(w_k) \log c_1(w_k) - c_2(w_k) \log c_2(w_k)] \quad (5)$$

Here $c_1(w_k)$ and $R_1(w_k)$ represent probability and rank of a specific word, w_k , in time series S_1 . Similarly, $c_2(w_k)$ and $R_2(w_k)$ stand for probability and rank of the same m-bit word in time series S_2 . The absolute difference of ranks is multiplied by the normalized probabilities as a weighted sum by using Shannon entropy as the weighting factor. Finally, the sum is divided by the value $2^m - 1$ to keep the value in the same range of $[0, 1]$. The normalization factor Z in Equation 5 is given by

$$Z = \sum_k [-c_1(w_k) \log c_1(w_k) - c_2(w_k) \log c_2(w_k)] \quad (6)$$

Next, we will provide an example using realistic traffic models that will show some initial results of this approach.

We have taken measurements from a wireless network, where we obtained the SINR values of 1000 packet transmissions. We made sure that the configuration of the nodes were as such that there was significant fluctuation on the signal; simulating thus, mobility and bursty traffic. Furthermore, we have obtained packet information from nodes' communication, where the network is fully connected and not connected.

Initially, we have undertaken experiments to show the information similarity index between the three configurations, namely no network, full network and bursty traffic network. The values that we collect from the comparison of the three types of traffic leads to thresholds for the characterisation of traffic; in short, we cluster the data based on these values, as we can see in table 1 below:

Table 1: IBS index for different traffic classes

No – Full	No - Bursty	Full - Bursty
0.140923	0.075158	0.158906

Note that No corresponds to No Network with low SINR, Full Network is a fully connected network with high SINR, and Bursty is a bursty traffic network.

We selected an 8 value word, in order to mimic the bursty traffic that wireless network usually

exhibit. We see that the IBS index for the Bursty Traffic - Full Network is similar to the Full Network -No Network configuration. This is the case, due to the existence of the bursts that lead the IBS index method to move the index towards the existence of a fully connected network.

Subsequently, we may include these values as thresholds in our machine learning implementation, in order to identify different states of traffic. However, using IBS to distinguish between these two traffic classes may result in an ambiguity in the state identification. This shows that the IBS may not locate the states that correspond to these different traffic patterns. Moreover, another limitation is that the IBS index method requires to collect at least 9 samples to be able to calculate a distance between the current state and the distance of the new value, in order to compare it to the thresholds. This does not allow us to investigate our problem at the single value granularity that will allow us to see the events that will occur.

4 EUCLIDEAN DISTANCE FOR WIRELESS TRAFFIC

The potential limitations of the IBS index lead us to employ the Euclidean distance for the data clustering. This method allow us to investigate the data at the single value, since it calculates the distance between the current state and the incoming value to compare their distance with our defined threshold. This will give us the opportunity to examine events at the value level, without requiring a set of values to be compared against another set.

The Euclidean distance in terms of machine learning is the distance measure between a pair of samples p and q in an n -dimensional feature space and it is given by

$$d = \sqrt{\sum_{i=1}^n (q_i - p_i)^2} \quad (7)$$

The Euclidean is often the default distance used in approaches such as the K-means clustering (MacQueen,1967) to locate the k closest point of a sample point.

5 HMM FOR WIRELESS TRAFFIC

We selected the Hidden Markov Models (Eddy, 1996) to classify and predict traffic in our network due to the fact the predictions may be made using the last recorded value, as opposed to other machine learning techniques, such as neural networks (Kosko, 1992), which requires a plethora of historic data, in order to be trained..

Hidden Markov models (HMMs) are the most popular means of temporal classification. Informally speaking, a hidden Markov model is a variant of a finite state machine. However, unlike finite state machines, they are not deterministic. A normal finite state machine emits a deterministic symbol in a given state. Further, it then deterministically transitions to another state. Hidden Markov models do neither deterministically, rather they both transition and emit under a probabilistic model (Rabiner and Juang, 1986).

Formally, an HMM is essentially a Markov model where a series of observed outputs

$$x = \{x_1, x_2, \dots, x_T\}$$

is available drawn from an alphabet

$$V = \{u_1, u_2, \dots, u_T\}.$$

Furthermore, he have the existence of states

$$z = \{z_1, z_2, \dots, z_T\}$$

provided by a state alphabet

$$S = \{s_1, s_2, \dots, s_{|S|}\}, z_t \in S, t = 1 \dots T.$$

The transition between states i and j is represented by the respected value in the state transition matrix A_{ij} . Moreover, the probability of generating the output observation is modelled as a hidden state. To this end, we define

$$P(x_t = u_k | z_t = s_j) = p(x_t = u_k | x_1, \dots, x_T, z_1, \dots, z_T) = B_{jk}$$

where B_{jk} is the matrix which encodes the probability of the hidden state producing the output u_k provided that the state at the corresponding time was s_j .

We used a similar approach as the one in (Dunham et al., 2004). Essentially, the HMM we developed is a time-varying Markov Chain, which

consists of entities that perform tasks, in order to reach a predicted value.

Initially, HMM performs the clustering action. The data that arrives from the wireless devices join a specific cluster labeled by the centroid is calculated using the following equation:

$$\bar{x} = \sum_{i=1}^n x_i \quad (8)$$

Where x_i is a set of n points of a dimension

$$dx_i, i = 1, 2, \dots, n.$$

In order to store an incoming value into a cluster, it is essential to calculate the distance between already existent states and the incoming value. The distance is found using equation (7) described in the previous section.

The present value is declared as a new state if the value of its distance with the already existent states is bigger than the value of a defined threshold. On the other hand, the incoming value is similar to an existent state, whose values are closer to the incoming value. The completion of the clustering initiates the building of the HMM. Given the Markov Chain at time t and the clustering result at $t + 1$ the Markov Chain is updated at time $t + 1$. First, the state transition probability between two successive points is calculated. Thereafter, the time sequence is updated with the state transition probability. Furthermore, the HMM includes a procedure of self-evaluation by calculating certain metrics of its performance such as the Normalised Absolute Ratio Error (NARE) and the Root Means Square (RMS) error.

$$NARE = \frac{\sum_{t=1}^N |(O_{(t)} - P_{(t)})|}{\sum_{t=1}^N O_{(t)}} \quad (9)$$

$$RMS = \sqrt{\frac{\sum_{t=1}^N (O_{(t)} - P_{(t)})^2}{N}} \quad (10)$$

where $O_{(t)}$ is the observed profile, $P_{(t)}$ is the predicted profile, N is the length of the dataset and t is the time variable of the t^{th} tuple in the input dataset.

Then the HMM reaches the prediction phase. Initially, the transition probability of the current state is calculated. The product of the transition probability with the states vector for each sensor recording provides the predicted value of the wireless network traffic. If a node has no connections with another

node, then the HMM assumes that the current node is connected to itself.

6 RESULTS

We obtained data from a wireless network that consisted of nodes that exhibited bursty traffic and no connectivity due to mobility. Thereafter, we input the data to our HMM to check the identification of different states. We set the threshold that the HMM recognises a new state to be the SINR threshold for successful transmission of a packet.

As we can see in Figure 1, the HMM finds two states, which correspond to the no connection between the

nodes and the burst of successful packet transmissions. There is an identification of events that shows the frequency of the identification of the states. Finally, we see a similar result in the increment of the states in the third figure, moving from the state of no connection to the state of connected network. Similarly, in Figure 2 we see the identification of a single state, since the data we put in our HMM does not extend from the value of the SINR threshold. In the events identification subfigure of figure 1 we see that different events are not emerging when a network is not connected; on the contrary, a single state exists dictating that the HMM can identify the presence of a disconnected network. In the same way, we see the next two subfigures of figure 1 that dictate the existence of a single state of a not connected network

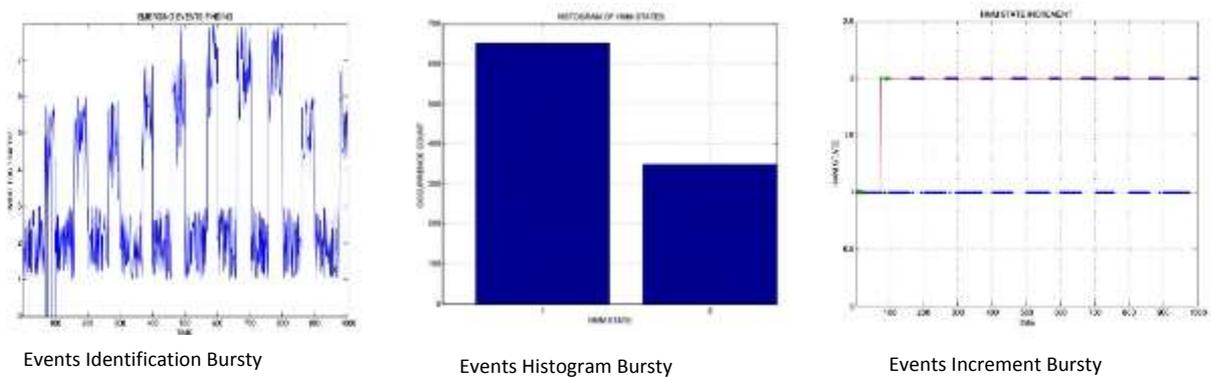


Figure 1: Results for HMM with Bursty Traffic

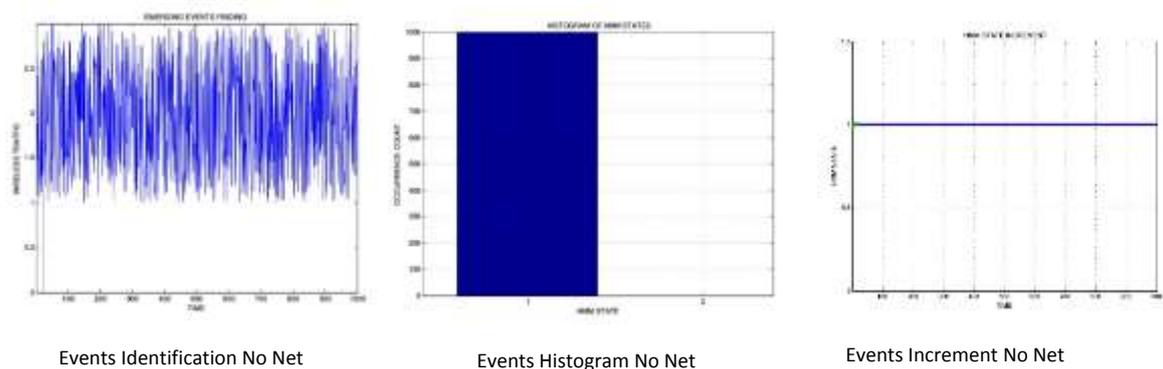


Figure 2: Results for HMM with No Net Traffic

In terms of performance, we calculated the RMS and NARE for each of the two scenarios. As we can see in table 1, the performance of the HMM is reasonably good with NARE values in the range of 0.20 – 0.26 and RMS from 0.58 – 0.88.

Table 2: RMS and NARE of the two scenarios.

	NARE	RMS
No Net	0.2576	0.5838
Bursty	0.2091	0.8730

Subsequently, we see that the events identification does not show any peaks to indicate that there is a state that has not been found. The above show us that we may have a reasonable mechanism that will be able to classify and predict traffic in a wireless network.

7 CONCLUSIONS

In this paper, we addressed the classification and prediction of wireless traffic using HMMs. We employed two clustering techniques, in order to clarify the states of the data to be input in the HMM.

The first one was the IBS index, which is usually used in physiological signals. We performed three experiments, obtaining the distances between three types of network traffic, namely No Network, Full Network and Bursty Traffic. We have seen that we get a difference in two of the three experiments; thus, resulting in a clear threshold identification for the identification of different traffic states. However, two of the three traffic classes exhibit a very similar distance; hence, the recognition of a new state by the HMM will be ambiguous. Furthermore, the nature of the IBS require bunches of signal values to be examined in order to locate the distance of the data to be evaluated.

Hence, we decided to use the Euclidean distance, which allows us to get a distance between the current state and the incoming value at a single value level; thus, identifying at a greater granularity the traffic patterns.

We put our approach to the test using two traffic files, one of the showing bursty traffic and a second with no connection. The HMM was able to locate the traffic patterns and identify the right number of states and events. We believe that this is an efficient approach for traffic pattern classification and prediction.

For future work, we aim to put our approach to a real network implementation, in order to obtain useful information regarding the operation of the HMM to low-power energy constraint devices.

REFERENCES

- Alizai, M. H., Landsiedel, O., Link, J. A. B., Goetz, S., and Wehrle, K. (2009). *Bursty traffic over bursty links*. In Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems, pages 71–84. ACM.
- Beran, J. (1994). *Statistics for long-memory processes*, Volume 61. CRC press.
- Dunham, M. H., Meng, Y., and Huang, J. (2004). *Extensible markov model*. In Data Mining, 2004. ICDM'04. Fourth IEEE International Conference on, pages 371– 374. IEEE.
- Eddy, S. R. (1996). *Hidden markov models*. Current opinion in structural biology, 6(3):361–365.
- Jiang, H. and Dovrolis, C. (2005). *Why is the internet traffic bursty in short time scales?* In ACM SIGMETRICS Performance Evaluation Review, volume 33, pages 241–252. ACM.
- Jiang, M., Nikolic, M., Hardy, S., and Trajkovic, L. (2001). *Impact of self-similarity on wireless data network performance*. In Communications, 2001. ICC 2001. IEEE International Conference on, volume 2, pages 477–481. IEEE.
- Kosko, B. (1992). *Neural networks and fuzzy systems: a dynamical systems approach to machine intelligence/book and disk*. Prentice Hall, Upper Saddle River.
- Li, R., Zhao, Z., Zhou, X., Palicot, J., and Zhang, H. (2014). *The prediction analysis of cellular radio access network traffic: From entropy theory to networking practice*. IEEE Communications Magazine, 52(6):234–240.
- Loumiotis, I., Adamopoulou, E., Demestichas, K., Kosmides, P., and Theologou, M. (2014). *Artificial neural networks for traffic prediction in 4g networks*. In International Wireless Internet Conference, pages 141–146. Springer.
- Maheshwari, S., Mahapatra, S., Kumar, C. S., and Vasu, K. (2013). *A joint parametric prediction model for wireless internet traffic using hidden markov model*. Wireless networks, 19(6):1171–1185.
- MacQueen, J. (1967, June). Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* (Vol. 1, No. 14, pp. 281–297).
- Ni, F., Zang, Y., and Feng, Z. (2015). *A study on cellular wireless traffic modeling and prediction using elman neural networks*. In 2015 4th International Conference on Computer Science and Network Technology (ICCSNT), volume 1, pages 490–494. IEEE.
- Pan, T., Sumalee, A., Zhong, R.-X., and Indra-Payoong, N. (2013). *Short-term traffic state prediction based on temporal-spatial correlation*. IEEE Transactions on Intelligent Transportation Systems, 14(3):1242–1254.
- Papadopouli, M., Shen, H., Raftopoulos, E., Ploumidis, M., and Hernandez-Campos, F. (2005). *Short-term traffic forecasting in a campus-wide wireless network*. In 2005 IEEE 16th International Symposium on Personal, Indoor and Mobile Radio Communications, volume 3, pages 1446–1452. IEEE.

- Papadopoulos, G. Z., Kotsiou, V., Gallais, A., Chatzimisios, P., and Noël, T. (2015). *Wireless medium access control under mobility and bursty traffic assumptions in wsn*s. *Mobile Networks and Applications*, 20(5):649–660.
- Park, K. and Willinger, W. (2000). *Self-similar network traffic and performance evaluation*. Wiley Online Library.
- Rabiner, L. and Juang, B. (1986). *An introduction to hidden markov models*. *IEEE ASSP Magazine*, 3(1):4–16.
- Rutka, G. and Lauks, G. (2015). *Study on internet traffic prediction models*. *Elektronika ir Elektrotechnika*, 78(6):47–50.
- Spyrou, E. D. and Mitrakos, D. K. (2015). *On the Homogeneous transmission power under the sinr model*. In 2015 ICTRS. SCITEPRESS.
- Yadav, R. K. and Balakrishnan, M. (2014). *Comparative evaluation of arima and anfis for modeling of wireless network traffic time series*. *EURASIP Journal on Wireless Communications and Networking*, 2014(1): 1-8.
- Yang, A. C.-C., Hseu, S.-S., Yien, H.-W., Goldberger, A.L., and Peng, C.-K. (2003). Linguistic analysis of the human heartbeat using frequency and rank order statistics. *Physical review letters*, 90(10):108103.

**MEDICAL
SENSORS
APPLICATIONS**

Brain Rehabilitation in Clinical Trials Setup by Eye-Tracking

Bartosz Kunka¹, Robert Kosikowski¹, Jessica Barlinn² and Karol Kozak²

¹*AssisTech Sp. z o.o., R&D Department, ul. Trzy Lipy 3, 80-172 Gdańsk, Poland*

²*Clinic for Nuerology, Medical Faculty, Dresden Univ. of Technology, Fetscherstrasse 74, 01307 Dresden, Germany
{bartosz.kunka, robert.kosikowski}@assistech.eu, {jessica.barlinn, karol.kozak}@uniklinikum-dresden.de*

Keywords: Brain rehabilitation, Neurology, Stroke, Human-computer interaction

Abstract: The number of patients with traumatic brain injury in Germany is about 280,000 per year. Eighty percent of the patients hospitalized in these cases exhibit minor traumatic brain injury, while approximately 10 percent are suffering from moderate and another 10 percent from severe traumatic brain injury. The goals of rehabilitation are to help survivors become as independent as possible and to attain the best possible quality of life. For the last few years, eye tracking has been used as an assistive tool, especially as a tool for alternative communication. Within the paper we explore new patent pending approach in brain injury rehabilitation. However, eye tracking questionnaire need a full implementation into clinical studies and medical documentation systems. In this paper we present integration of cognitive test into eye tracking technology based on electronic case report form.

1 INTRODUCTION

Increasing number of people who undergo brain damage is one of the most characteristic features of our contemporary society. “Brain injury” is a term used in terms of traumatic brain injury (TBI), a cerebral stroke, or changes occurring in the brain that are consequent from cerebral hypoxia (e.g. due to perinatal incidents, sudden cardiac arrest (SCA) or suicidal strangulation). These examples of TBI lead to serious neurological disorders directly related to cognitive disturbances and need to be assessed in an objective way.

Neurological rehabilitation pursues different goals for recovery that can be divided into several steps: in the acute phase of TBI restitution of neurobiological processes should be facilitated by different therapeutic strategies. If improvement of functional deficits is not achieved or is not expected to occur (e.g. because of a large brain lesion), it should be aimed at compensation strategies, for instance by the use of assistive tools. Individuals may need to learn how to communicate and express their own feelings. If sensomotor, language or cognitive deficits cannot be compensated, the patient’s environment should be adapted to his needs

(Turner-Stokes, 2007). As recently recommended for stroke patients, the patient’s functional and cognitive status must be accurately assessed within the first few days after stroke to identify special needs for further therapeutic and rehabilitative strategies (Hebert, 2016).

Eye tracking technology might be useful in several steps of brain rehabilitation process, from diagnosis to therapeutic implications, especially when eyeballs movements is the only channel of communication. Nowadays eye tracking technology is well-known and it is reasonable to use it in medical purposes, especially in supporting neurological examination of patients with serious communication barriers (Doležal, 2015) (Kunka, 2014). Therefore, within the paper we present a new approach for objectivized neurological assessment procedure based on tests included in a case report form and adopted for eye tracking.

Patients with cognitive disturbances originating from TBI, should be immediately assessed in regards to the extensiveness of the damage and the level of their consciousness. Properly conducted diagnostics is absolutely crucial, since the whole rehabilitation management will be later based on it. Furthermore, the proposed approach based on eye

tracking technology supporting diagnosis of neurological patients could be also employed at their intensive rehabilitation by stimulating particular structures of the central nervous system (CNS).

Neurorehabilitation concerned with cognitive functions stimulation is very important. It is a time-consuming process requiring cyclic, systematic work and motivation of all involved people. When the CNS is constantly stimulated by different stimuli and intellectual efforts, the repair mechanism can be activated in a damaged brain. The main mechanism is neuroplasticity which could compensate damaged brain centers' functions due to creation of new neural connections. Repair mechanism of brain consists of reorganization, adaptation, changeability, self-repair, learning and remembering of neurons (Doležal, 2015).

It is worth mentioning that cyclic monitoring and observation of rehabilitation progress would be also supported by proposed in this paper neurological assessment utilizing eye tracking system.

Eye tracking is a technology allowing for determining user's gaze direction (it measures coordinates of eye fixation point). Eye tracking follows the path of an observer visual attention thus it enables controlling of mouse cursor in the system, as well as detection and analysis of user's regions of interests.

Currently eye tracking interfaces are based on video processing. They utilize near-infrared technology along with a high-resolution camera to track gaze direction.

In the paper we focus on the remote eye trackers, especially on the C-Eye system described in C-Eye System section.

Eye tracking technology has many applications in different areas. The most known are related to human-computer interaction (HCI), entertainment (Drewes, 2010), (Jacob, 2003), and research, including psychology and neuroscience (Kooiker, 2015), (Hahn, 2015), as well as medicine (Kunka, 2014).

Neuroscience and psychology employ eye tracking technology to analyze the scan paths (gaze patterns) and heat maps to gain deeper insights into cognitive processes underlying attention, learning, and memory. Another research shows that eye tracking gives us insights into i.a. word processing, particularly how eye movements during reading are affected by the emotional content of the texts (Urry, 2010).

Eye tracking in combination with standard research methods or other biometric sensors can also support diagnosis of neurological diseases such as

Autism Spectrum Disorder (ASD), Attention Deficit Hyperactivity Disorder (ADHD), Schizophrenia, Parkinson's (PD), and Alzheimer's disease (AD). It is worth mentioning that iMotions exists (iMotions, 2016). It is the integrated web biometric research platform integrating best-in-class biosensors and synchronizes eye tracking, facial expression analysis, EEG, GSR, EMG, ECG and surveys. iMotions platform helps its users conduct state-of-the-art human behavior research in the areas of psychology, neuroscience, human factors engineering, education, health, business and HCI. The platform is used worldwide by leading universities such as Harvard, Yale and Stanford.

This section presents possibilities of eye tracking technology offered to different branches, especially for specialized medical examinations and objective assessment of the measured data. It has been proven that eye tracking can be successively used both in diagnostic and therapeutic applications.

A case report form (or CRF) (Bellary, 2014) is a paper or electronic questionnaire specifically used in clinical trial research. The Case Report Form is the tool used by the sponsor of the clinical trial to collect data from each participating site. All data on each patient participating in a clinical trial are held and/or documented in the CRF, including adverse events. Originally all case report forms were made on paper. But recently there is a changing trend to perform clinical studies using an electronic case report form (eCRF). Commonly encountered challenges in CRF designing are consistency in the design, collection of precise data and user-friendliness also from assisting devices. These challenges can be overcome by proper planning by a using clinical supporting system. This way of working has many advantages: faster and efficient, high security, environmentally friendly. In this paper we present integration of cognitive test into eye tracking technology based on electronic case report form.

In this paper we present a concept of integration of eCRF with eye tracking technology for clinical trials in TBI rehabilitation.

2 C-EYE SYSTEM

Clinical trials involves many steps, one of the most time-consuming elements of conducting clinical studies is the entry of clinical data onto case report forms (CRFs). Traditionally, this is done by clinical research coordinators (CRCs) at various research centers who use a pen to write the data on paper-

based CRFs, which are then faxed to clinical monitors (CRAs) where they are examined for potential errors that may skew the accuracy of statistical data required to evaluate a drug's performance. The most important for facilitating study management are electronic data capture by users (doctors and patient) and clinical trial management software. This paper describes the advantages of integration C-Eye and study management.

As eCRFs are created within the C-Eye user interface, all the field definitions, data types, control positions, and validation rules are stored in a single table that saves CRF definitions for all trials, no matter how different they are. Data on patient-specific CRFs is stored in a similar manner – all field values are checked for data type compliance at the application level and written to a table as records that can easily be extracted using a single reporting tool for all trials. Elements of the system for study management in TBI domain:

2.1 Data Capture Section

A data element in an eCRF represents the smallest unit of observation captured for a subject in a clinical investigation. Examples of data elements include IQ test, color recognition, object recognition, or other clinical observations made and documented during a study Data capture interface allow:

1. Electronic Source Data Origination
2. Test Data Capture
3. Data Element Identifiers
4. Modifications and Corrections
5. Use of Electronic Prompts, Flags, and Data Quality Checks in the eCRF

Many data elements in a clinical investigation can be obtained at a study visit and can be entered directly into the C-Eye eCRF form by an authorized data originator. This direct entry of data can eliminate errors by not using a paper transcription step before entry into the eCRF.

The forms with eye tracking system are providing a possibility to collect specific records:

- Heat maps: aggregations of gaze points and fixations revealing the distribution of visual attention.
- Scan paths (or Fixation sequences): sequence representing the order of subjects' looking and how much time they spend

- Time of interesting: parameter quantifies the amount of time that subjects have spent on Areas of Interest (AOIs) being predefined subregions of displayed content (e.g. subregions representing the right answer)
- Time To First Fixation (TFFF): the time to first fixation indicates the amount of time it takes a respondent to look at a specific AOI.

Typically, images (eye motion, eye symptoms, face images) are not included as data elements in an eCRF, but rather the clinical interpretation of the image is included as a predefined data field.

2.2 Data Review

To comply with the requirement to maintain accurate clinical test, clinical investigator should review and electronically sign the completed eCRF for each subject before the data are archived or submitted to clinical research organization (CRO). To comply with the requirement to maintain accurate test histories, data elements might call for modification or correction during data review. Either the clinical investigator can enter the revised data element. Modified and/or corrected data elements must have data element identifiers that reflect the date, time, originator, and reason for the change, and must not obscure previous entries.

If changes are made to the eCRF after the clinical investigator(s) has already signed, the changes should be reviewed and electronically signed by the clinical investigator(s).

2.3 Retention of Records by Clinical Investigator

The clinical investigator(s) should retain control of the records (i.e., completed and signed eCRF or certified copy of the eCRF).

2.4 C-Eye System as Study Management Interface

The C-Eye system is a fully integrated certified medical device supporting the evaluation of the state of consciousness of patients suffering from any central nervous system disorder, enabling neurorehabilitation for people with neurological dysfunctions and impaired development. The C-Eye system also supports alternative communication thanks to eye tracking technology implemented as a remote interface.

The evaluation and neurorehabilitation of a patient suffering from neurological disorders and impaired development consists in performing special tasks based on multimedia contents. The subject establishes interaction with contents displayed on the screen using their sight, i.e.: graphics, photographs and captions. This way, specific centers of the central nervous system are both evaluated and stimulated, especially centers responsible for sight, hearing, speaking and cognitive functions.

As presented in the Figure 1, the system consists of the processing unit (PU), which is the integrated into the touch display (TD) and the speakers (S), which are integrated into the touch display (TD).

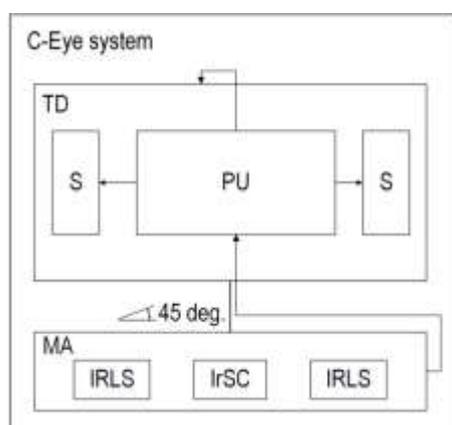


Figure 1: C-Eye – structure of the system

The system is equipped with two infrared light sources (IRLS), that enable to indicate the visual fixation point position through generating infrared light reflections, that are reflected from the surface of subject's cornea and acquired by the infrared sensitive camera (IrSC). The infrared light sources (IRLS) are integrated with the infrared sensitive camera (IrSC) in the way that the infrared light sources (IrLS) are located symmetrically and uniaxially on the both sides of the camera (IrSC) and put together into the longitudinal cover to be formed into the movable attachment (MA). The movable attachment (MA), which is connected with the processing unit (PU) and located in the lower part of the display (TD), is up and down tiltable in a range of 45 degrees in relation to the perpendicular location of the attachment towards the display (TD).

The patient with potential cognitive disorder is located before the C-Eye system. The C-Eye is attached to the movable extension arm and adjusted to the subject through the adjusting movements of the movable extension arm in this way, that the

subject is located 60 cm before the system. The C-Eye is parallel to the patient's interpupillary line, so that the patient's eyes are situated in the angle of view of the camera (IrSC), as it was presented in Figure 2.



Figure 2: Examination session with the C-Eye system

2.5 C-Eye and Integrated Medical System for Study Management

There are a set of many tasks included in eCRFs that can be fully, objectively performed only with support of eye tracking technology. The C-Eye and eCRF approach assist the physician to use as system in clinical practice. C-Eye could be a platform for a comprehensive patient management and the integration of study documentation into clinical practice – orientation guide for ideal progression control of the therapy. The current version of the C-Eye contains various tasks that correspond to the tasks located in specific subsections of the eCRF being used in everyday clinical practice. C-Eye combines clinical documentation, medical records register, specific therapy documentation, and research projects in one platform. It is necessary to adapt some of them to the structure and template of content presentation to the C-Eye system. Adaptation of eCRF tasks being dedicated for eye tracking-based interaction is associated with content development and its proper implementation. There is a scope for interdisciplinary cooperation between AssisTech engineers and representatives of the medical world.

Conducting full adequate and objective neurological assessment of patient after TBI requires efficiency evaluation of the communication senses (vision and hearing). Sometimes patients who have suffered craniocerebral injury experience visual impairment, and hemispatial neglect, unilateral

(partial) visual inattention (agnosia), and unilateral "neglect" of space occurs. Due to the preliminary assessment we may take into account the patient's difficulties with perception in the half of space opposite to the damaged brain hemisphere.

In the next step, we may conduct simplified hearing examination. It is very important, as patients following craniocerebral trauma are auditory oversensitive (Landon, 2012). At times, such sounds cause physical pain, and significantly reduce the patient's comfort. Therefore, it is very important to adjust loudness of all sounds produced by the C-Eye system to suit the patients' needs.

The C-Eye can effectively expand the use of existing assessment forms for cognitive functions of patients that have been deprived contact with the world and increase the efficiency of their evaluation. Especially, the following sections of Assessment Forms for Cognitive Rehabilitation should be mentioned here: attention problems, visual processing problems, memory problems, information processing problems, executive functions problems. Cognitive assessment procedure adopted to the C-Eye system with considering objectivized evaluation of the communication senses was presented in Figure 3.

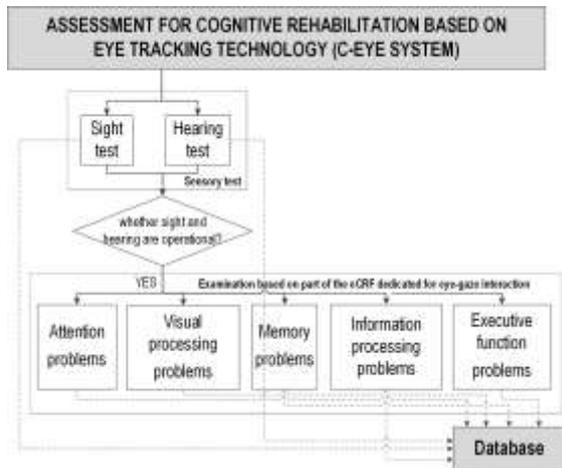


Figure 3: Adaptation of eCRF test to the C-Eye system

Furthermore, the C-Eye can be an effective tool for rehabilitation that allows to perform the following exercises dedicated for patients communicating by eyeballs movements, as well as engaged in Cognitive Stimulation Program (Ribeiro, 2011) at home: spatial/writing skills, search and find, recall of pictures and places, recall of story material, visual scanning, language exercises, categorization exercises, reading comprehension,

time sense. It is worth mentioning that the C-Eye being a medical device is simple to use - feature especially important in everyday practice. It doesn't require any software installation or configuration. The C-Eye is fully operable couple seconds after turning it on, and does not require calibration procedure which, in fact, would disqualify the use of the system in case of patients in severe state after TBI. Our approach presented in the paper is comprehensive. We propose employing specialized eye tracking system for holistic assessment, as well as neurorehabilitation (Figure 4).

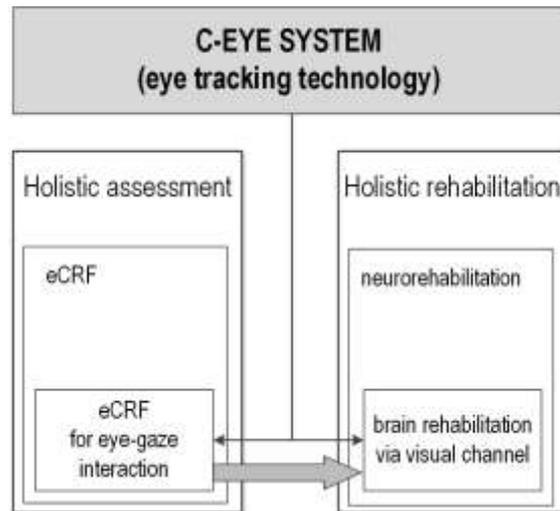


Figure 4: C-Eye system in neurorehabilitation and cyclic, constant assessment based on profiled eCRF

3 RESULTS

Cognitive Test data includes all information in original records and certified copies of original records of rehabilitation procedure, observations in a rehabilitation and diagnostic after TBI.

Access to cognitive test data is critical to the review and inspections of clinical investigations. The review of cognitive test data by both the clinic and sponsor is important to ensure adequate protection of the rights, welfare, and safety of human subjects and the quality and integrity of the clinical investigation data. Cognitive clinical test should be attributable, legible, contemporaneous, original, and accurate and must meet the regulatory requirements for record keeping.

Capturing eye tracking cognitive test data electronically and transmitting it to the eCRF will help:

- Eliminate unnecessary duplication of data
- Reduce the possibility for transcription errors
- Encourage entering source data during a subject's visit, where appropriate
- Eliminate transcription of source data prior to entry into an eCRF
- Facilitate remote monitoring of data
- Promote real-time access for data review
- Facilitate the collection of accurate and complete data

4 CONCLUSION

Cognitive Stimulation and rehabilitation allows for great flexibility so that patients can tailor their program of rehabilitation and follow individual schedules. TBI survivors may participate in an intensive level of therapy several hours per week or follow a less demanding regimen. Eye tracking rehabilitation efforts to address the continuum-of-care needs of TBI patients are being developed. Eye tracking service providers and researchers will need to put in place service delivery plans backed by strong research components, which include control populations, prospective evaluations, and rigorous methodology for the assessment of functional vision. In conclusion, our study yielded relevant information related to a structured TBI rehabilitation service and represents an alternative for patients and families afflicted by TBI, enabling the generation of clinical protocols in eCRF Format.

ACKNOWLEDGMENTS

We are very grateful to GoGlobal Team from Fraunhofer Center for International Management and Knowledge Economy for their insights and consultancy into the German medical system and the role cognitive devices in this country.

REFERENCES

Turner-Stokes L, Disler PB, Nair A et al. *Multidisciplinary rehabilitation for acquired brain injury in adults of working age*. Cochrane Database Syst Rev 2007; 2: CD004170

Hebert D, Lindsay MP, McIntyre A, Kirton A, Rumney PG, Bagg S, et al, Canadian stroke best practice recommendations: Stroke rehabilitation practice

guidelines, update 2015. Int J Stroke. 2016 Apr 14. pii: 1747493016643553.

J. Doležal, V. Fabian, *Application of eye tracking in neuroscience*, Clinical Neurophysiology, Volume 126, Issue 3, pp. e44, 2015.

B. Kunka, T. Sanner, A. Czyżewski, A. Kwiatkowska, *Consciousness study of subjects with unresponsive wakefulness syndrome employing multimodal interfaces*. In: Ślęzak, D., Tan, A.-H., Peters, J.F., Schwabe, L. (eds.) BIH 2014. LNCS, vol. 8609, pp. 57–67. Springer, Heidelberg (2014).

H. Drewes, *Eye Gaze Tracking for Human Computer Interaction*, LFE Medien-Informatik der Ludwig-Maximilians-Universität, Monachium 2010.

R. J. K. Jacob, K. S. Karn, *Eye tracking in Human-Computer Interaction and usability research: ready to deliver the promises (section commentary)* in “The Mind's Eyes: Cognitive and Applied Aspects of Eye Movements”, (J. Hyona, R. Radach, H. Deubel), Elsevier Science, Oxford 2003.

M. J. G. Kooiker, J. J. M. Pel, J. van der Steen, *The Relationship Between Visual Orienting Responses and Clinical Characteristics in Children Attending Special Education for the Visually Impaired*. Journal of Child Neurology, 30(6), 690–697, 2015.

N. Hahn, J. Snedeker, H. Rabagliati, *Rapid Linguistic Ambiguity Resolution in Young Children with Autism Spectrum Disorder: Eye Tracking Evidence for the Limits of Weak Central Coherence: Rapid lexical ambiguity resolution in ASD*. Autism Research, 8(6), 717–726, 2015.

H. L. Urry, *Seeing, thinking, and feeling: emotion-regulating effects of gaze-directed cognitive reappraisal*, Emotion, 10(1), pp. 125-135, 2010.

iMotions platform, available at: <https://imotions.com/>, last access: April 14th, 2016.

S. Bellary, B. Krishnankutty and M. S. Latha: *Basics of case report form designing in clinical research*. Perspect Clin Res. 2014 Oct-Dec; 5(4): 159–166.

J. Landon, D. Shepherd, S. Stuart, A. Theadom, S. Freundlich, *Hearing every footstep: Noise sensitivity in individuals following traumatic brain injury*, Neuropsychological Rehabilitation, 22(3), pp. 391-407, 2012

F. R. Freire, F. Coelho, J. R. Lacerda, M. F. da Silva, V. T. Gonçalves, S. Machado, B. Velasques, P. Ribeiro, L. F. Hindi Basile, A. Maynard, P. Oliveira, Wellingson S. P., P. A. Medeiros Kanda, R. Anghinah: *Cognitive rehabilitation following traumatic brain injury*. Dement Neuropsychol 2011 March;5(1):17-25

SPECIAL SESSION ON

DATA

SCIENCE AND

APPLICATIONS

Adopting NoSQL Databases Using a Quality Attribute Framework and Risks Analysis

Hilda Mackin, Gonzalo Perez and Charles C. Tappert

*School of Computer Science and Information Systems, Pace University, 861 Bedford Road, Pleasantville, NY, USA
hd79424w@pace.edu, hdiaz19@hotmail.com, 76gonzalo@gmail.com, ctappert@pace.edu*

Keywords: Distributed Databases, NoSQL Databases, Relational Databases, Quality Attributes.

Abstract: NoSQL has emerged in recent years to provide increased scalability and performance, and organizations have a problem in choosing between traditional SQL and NoSQL databases. This research gives software engineers and architects a way to select a NoSQL database for a particular big data environment and domain. It proposes a Quality Attribute Framework and Risk Analysis of NoSQL databases that can measure quality metrics associated with availability and security, which are critical to choosing the right NoSQL database for a given domain and to making better software development and design decisions. The framework will help IT departments align perceived risks of NoSQL database adoption with actual risks, helping IT managers in their database adoption and in the identification of risk factors that affect the new database technologies. The framework developed here will be finalized through a qualitative analysis of risk vectors via surveys of top IT leaders and IT companies.

1 INTRODUCTION

In the past, relational databases were used for all tasks to be processed by a database because of their query capabilities, and transaction management features.

Traditional Relational databases were designed for different hardware and software times and are facing challenges in meeting the performance and scale requirements of Big Data (Grolinger et al., 2013).

The vital factor for a change in data storage was the need to support large volumes of data in distributed environments. Two companies in particular Google and Amazon have been influential with their Big Table from Google and Dynamo from Amazon papers.

One of the first solutions to increase the amount of data that could be stored by a DBMS system was to spread the data among several database servers or clusters instead of just one server.

A second important factor for a change was that distributed systems were not very efficient performing transactions and join operations using a Relational Database System (RDBMS).

Since Relational databases need to maintain strict consistency using the transactional ACID model and

must be highly available a new family of NoSQL clustered databases emerged.

As more companies' considered adopting Big Data Solutions, discussions about the most appropriate NoSQL database for their use case, application or environment originated.

NoSQL databases are highly scalable, have good performance, are designed to store and process a significant amount of unstructured data faster than Relational Databases. So why to be cautious about adopting a NoSQL database? Availability and Security are major concerns for IT Enterprise Infrastructures. There are perceived Security Risks associated with the new database technologies (Obijaju, 2015).

2 RELATED WORK AND RESEARCH CONTRIBUTIONS

NoSQL data stores are seen as data processing alternatives that can handle considerable volume of data and provide better scalability, but attributes and risks associated with these new technologies are not well understood. Because of the large number and diversity of NoSQL solutions, it is challenging to

choose an appropriate NoSQL solution for a specific task or Use Case. Some studies have identified NoSQL challenges including the immense diversity and inconsistency of terminologies, limited documentation, sparse comparison and benchmarking criteria, and lack of standardized query languages (Grolinger et al., 2013).

Many researchers have focused on Performance Evaluation but have not included other Software Quality Attribute requirements (Lourenço et al., 2015).

A previous study created a comparison of NoSQL databases, identifying the software attributes that would aid a software engineer's decision process (Lourenço et al., 2015). It identified several desirable quality attributes to evaluate NoSQL databases: Availability, Consistency, Durability, Maintainability, Read Performance, Recovery Time, Reliability, Robustness, Scalability, Stabilization Time and Write Performance. It also selected some popular NoSQL databases: Aerospike, Cassandra, Couchbase, CouchDB, HBase, Mongo DB, and Voldemort. Their study resulted in a summary table (Table 1) to aid software engineers and architects in their decision process when selecting a given NoSQL database according to certain quality attributes.

Our research investigates, in a qualitative manner, the attributes of an effective NoSQL Quality Attribute Framework. This work will make two new contributions to the state-of-the-art.

1) Provide a straightforward and coherent way for IT Managers to understand NoSQL database quality attributes and gain some insight on mitigation strategies for current NoSQL databases risks.

2) Provide guidance to rationalize risks associated with Security and Business Availability.

This paper improves and complements the study (Lourenço et al., 2015).

1) Quality attributes are clearly defined (Section 4.2) and classified and evaluated by four types of NoSQL database types: Key-Value, Document, Columnar and Graph Database (Section 4.4).

2) This research adds Security attributes, enhances Availability attributes and adds Additional Quality Attributes to provide a complete set of quality attributes. (Table 4).

The framework will be completed shortly through a quality-attribute-focused survey based on NoSQL database type (Key Value, Document, Columnar, and Graph), where databases are compared with regards to their suitability for quality attributes.

2.1 Research Approach

The goal of this research is to gain a thorough understanding of NoSQL databases to clarify the risks associated with these new technologies and to provide a framework and recommendations to mitigate risks. Two main research problems and a sub problem are identified:

First Research Problem

What NoSQL databases Quality Attributes have the most positive and negative impacts on Security, and Business Availability risks?

What Software Quality Attributes are required to evaluate and adopt NoSQL databases?

What is the perceived Business Availability and Security Risks associated with these new database technologies in the context of Big Data?

Second Research Problem

Are there any misalignments between actual and perceived NoSQL database risks uncovered through surveys and the risks perceived by experts and IT professionals?

What are the technical and non-technical impacts on the organization?

First Research Sub-problem

Are Relational and NoSQL databases totally mutually exclusive?

How do applications intersect between Relational and NoSQL databases?

If the technologies do not overlap, but intersect choose between the databases?

3 NOSQL DATABASES

The NoSQL movement is a contemporary approach to data persistence using novel storage methods (Franks, 2012). NoSQL databases were built to deal with the increasing amount of complex data, required in some real-time applications, to address availability over consistency, and to allow horizontal scalability, using a distributed architecture and open source (Bazar & Cosmin, 2014).

The common characteristics of NoSQL databases are (Sadalage & Fowler, 2013):

- Not using the relational model
- Running well on clusters
- Open source projects
- Built for the 21st-century web estates
- Schema-less.

3.1 Data Model

There are four accepted types of NoSQL databases:

Key Value Stores, Consist of keys and their corresponding values, data is stored in a schema-less way. This allows search millions of values in a fraction of the time needed by conventional storage (Franks, 2012).

Document Stores, Consist of a set of documents possibly nested. Data can be structured in a schema-less way or in the form of collections. Popular examples are Couch Base, Mongo DB (Franks, 2012).

Column Stores or Extensible Record Stores, Consist of tables which may have different schema for each row having one huge extensible column containing the data. Column stores do not declare null fields as relational databases do (Franks, 2012).

Graph Stores, Consist of a set of graph nodes linked together by edges (providing index-free adjacency of nodes) (Barmpis & Kolovos, 2012).

Each type of NoSQL database is tailored for storing a different and specific type of data (Barmpis & Kolovos, 2012).

3.2 Advantages of NoSQL Databases

Most NoSQL data stores run on clusters. Relational databases use ACID transactions to handle consistency across the entire database. In contrast in a cluster environment, NoSQL databases offer a range of options for uniformity and distribution.

NoSQL databases operate without a schema, allowing one to freely add fields to database records without having to define any changes in structure first. This is particularly useful when dealing with non-uniform data and custom fields.

As a result, Relational databases are seen as just one option for data storage. The result is that many organizations will have a mix of data storage technologies or Polyglot Persistence.

Besides handling data access with sizes and performance that demand a cluster, the other important reason for NoSQL technology is the impedance mismatch problem. Big Data concerns have created an opportunity for people to think in a different way about their data storage needs. Some development teams see that using a NoSQL database can help their productivity by simplifying their database access. This can be achieved even if they have no reason to scale beyond a single machine, improving the productivity of application development by using a more suitable data interaction style (Sadalage & Fowler, 2013).

4 WORK PLAN

The work plan will be composed of the following sections:

1. Presenting a previous study's comparison of NoSQL engines and their quality attributes.
2. Defining a framework by enlarging the quality attributes to include Security attributes, enhancing Availability attributes and adding Additional Quality Attributes.
3. Two additional NoSQL engines are added to the study, illustrating two examples each of the four types of NoSQL databases.
4. NoSQL databases proposed are categorized into the four NoSQL database types.
5. The quality attributes of the two examples of each type are consolidated to arrive at a direct comparison of the four NoSQL database types.
6. Finally, the set of attributes are enlarged to include Security, Availability and Additional Attributes, obtaining a framework to form the basis for a subsequent qualitative study.

4.1 Previous NoSQL Database Study

A previous study created a comparison of NoSQL databases, identifying the software attributes that would aid a software engineer's decision process (Lourenço et al., 2015).

Their study resulted in Table 1 to aid software engineers and architects in their decision process when selecting a given NoSQL database according to certain quality attributes.

The table uses a 5-point scale ranging from "Great" to "Bad" to allow a direct comparison among databases. For example, Cassandra is written based on a performance-oriented approach, more than Couch base. Worse grades were assigned when a database was not an ideal pick, according to their authors' literature revision. This does not mean that the database lacked the attribute entirely, but it was not the best compared to the others databases (Lourenço et al., 2015).

The quality attributes have the following meanings:

Availability: Downtime was used as a primary measure.

Consistency: Graded according to ACID semantics consistency and how much consistency can be fine-tuned.

Durability: Measured according to the use of single or multi-version concurrency control schemes,

the way the data are persisted to disk.

Maintainability: Measured for ease of setup and use, accessibility of tools to interact with the database.

Read Performance: Considered studies about the fine tuning of each database.

Write Performance: Considered studies about the fine tuning of each database.

Recovery Time: Related to availability, took results from a previous study.

Reliability: Looked at synchronous propagation modes

Robustness: Considered the tendency of databases to have problems dealing with crashes or attacks.

Scalability: Looked at each database elasticity, horizontal scaling, and ease of online scalability.

Stabilization Time: Related to availability.

Table 1: Quality attributes for popular NoSQL engines (redrawn from (Lourenço et al., 2015)). Legend: “G” = Great, “+” = Good, “A”= Average, “_” = Mediocre, “B” = Bad, and “?” = Unknow/NA.

	Aerospike	Cassandra	Couchbase	CouchDB	HBase	MongoDB	Voidmort
Availability	G	G	G	G	-	-	G
Consistency	G	G	+	+	A	G	+
Durability	-	+	+	-	+	+	+
Maintainability	+	A	+	+	-	A	-
Read Performance	+	-	G	A	-	G	+
Write Performance	+	G	+	-	+	-	G
Recovery Time	G	B	+	?	?	G	?
Reliability	-	+	-	+	+	G	?
Robustness	+	+	A	A	B	A	?
Scalability	G	G	G	-	G	-	+
Stabilization Time	B	+	+	?	?	B	?

The study concluded that even though there was a variety of other research and evaluations of NoSQL technology, there was still not enough information to verify how fit a NoSQL database is in a specific scenario or system, making the following recommendation for future work: “The development of a framework for assessing most of these quality attributes would greatly benefit the database adoption of software engineers and architects” (Lourenço et al., 2015).

4.2 Proposing a Framework

The framework proposed covers Availability and Security attributes highly valued by NoSQL databases users. Definitions of the quality attributes included in the study are given below:

Availability: The percentage of time a system is operating correctly. Is the data always accessible? Is the data permanently available? In the context of the CAP Theorem Availability was evaluated versus

Consistency and Availability versus Partition Tolerance.

Consistency: The valid and reliable data that is saved in every cluster node. Is the data the same in every replication on every cluster node? In the context of CAP Theorem Consistency is assessed when all nodes see the same data at the same time.

Partitioning: Defined as the data divided in smaller segments to be allocated in different data store tables. Is Horizontal Partitioning or Vertical Partitioning allowed?

Replication: Keeping a copy of the data in different databases and servers. Is Replication transparency allowed? Is Replication considered in different layers? Does Master slave and master replication have one instance or have different instances?

Scalability: Related to horizontal scaling. Is Scaling achieved by replicating the data synchronously or asynchronously?

Shared Nothing: Are all replica nodes allowed to continue working even if they are disconnected?

Recovery Time: Related to the time it takes for several NoSQL systems to recover from a node failure.

Stabilization Time: Related to the time it takes for the system to stabilize when that node re-joins the cluster.

Reliability: System’s probability of operating without failures for a given period.

Robustness: Defined as the ability of the database to cope with error during execution

Durability: Property that guarantees that a transaction that has been saved/commit in the database will be committed even in the event that the system crashes

Maintainability: Does the NoSQL database provide features for easy maintainability, administration, management and operation?

Read Write Performance: Is the NoSQL database more robust on Reading operations than writing operations?

Security: Defined as the software and the set of management tools that protect the database against attacks, hackers, and viruses. Some properties that need to be analysed in the study are:

Authentication: Related to password and user’s login. What types of authentication does the NoSQL database provides?

Authorization: Defined as a set of read and write permissions request on tables, creation of users. What administrative functions the NoSQL database provides?

Encryption: Does the NoSQL database provide mechanisms that allow encryption techniques that preserve data confidentiality? If not what mechanisms are used to enforce data confidentiality?

Three levels will be explored in the study: Data at rest, Client to server communication, Server to Server connection.

Auditing: Does the NoSQL database provide mechanisms that allow writing to the database or Audit Logs?

Data integrity: Does the NoSQL database provide mechanisms that allow data integrity such as ACID or eventually consistent BASE? Could the NoSQL database achieve different levels of Data Integrity?

Confidentiality: Data Confidentiality. Does the NoSQL database provide different mechanisms to preserve data confidentiality?

Documentation: Does the NoSQL database provide End User Documentation? What levels of documentation does it provide?

Additional Quality Attributes –Additional attributes will be addressed in the study as part of the contribution envisioned. These include attributes related to the ease of developing when using NoSQL databases:

Popularity Depending on the Type on NoSQL database (Key Value, Document, Columnar or Graph) some databases are more popular than others. Different aspects need to be analysed to determine what database is applicable to the business case and from that infer the popularity.

Maturity: Considering Maturity of the API, time in the market, and Enterprise adoption.

Query Possibilities: Does the NoSQL database provide SQL like query possibilities?

Concurrency Control: Does the NoSQL database provide features to manage Concurrency Control? Does it provide Optimistic or Multi Version concurrency control?

Conflict Resolution: Does the NoSQL database provide mechanisms to manage Conflict Resolution?

4.3 Selecting NoSQL Databases

Considering four NoSQL database types (Key Value, Document, Columnar, and Graph) the study selects eight popular NoSQL databases used by enterprises, two databases to represent each type. These database types are now compared with regards to their suitability for quality attributes.

The database selection was based on literature research and data collected from preliminary interviews: Voldemort, Redis, Mongo DB, Couch DB, Cassandra, HBase, Neo4J, OrientDB.

4.4 NoSQL Database Types

We now categorize these eight popular NoSQL engines into the four NoSQL database types, taking five of the engines from Table 1 (Voldemort, MongoDB, CouchDB, Cassandra, HBase) and adding three new engines (Redis, Neo4J, and OrientDB).

Table 2: Quality attributes for eight popular databases, two of each type. Legend: “G” = Great, “+” = Good, “A” = Average, “-” = Mediocre, “B” = Bad, and “?” = Unknown/NA.

	Key Value		Document		Columnar		Graph	
	Voldemort	Redis	MongoDB	CouchDB	Cassandra	HBase	Neo4J	OrientDB
Availability	G	+	-	G	G	-	G	+
Consistency	+	+	G	+	G	A	+	G
Durability	+	+	+	-	+	+	+	+
Maintainability	-	+	A	+	A	-	+	+
Read Performance	+	+	G	A	-	-	+	G
Write Performance	G	G	-	-	G	+	A	G
Recovery Time	?	?	G	?	B	?	?	?
Reliability	?	+	G	+	+	+	+	G
Robustness	?	+	A	A	+	B	+	+
Scalability	+	+	-	-	G	G	G	A
Stabilization Time	?	?	B	?	+	?	?	?

Two examples illustrate each of the four types, as shown in Table 2.

The attributes for the engines from Table 1 come from (Lourenço et al., 2015), the attributes for Redis and Neo4J from (Redmond, 2012), and the attributes for Orient Database from (Orient Database 2016)

The quality attributes of the two examples of each type are consolidated to arrive at a direct comparison of the four NoSQL database types (Table 3). Averaging the attribute values for the two example engines of each type was done conservatively by leaning toward the lower rating on the 5-point scale.

4.5 Expert Opinions and Positions

Due to the diversity of NoSQL solutions and four different types of NoSQL databases, making the choice of the most appropriate data store for a given use case scenario will be easier with the framework proposed. Table 3 is now extended by adding the security and additional quality attributes discussed above to obtain Table 4.

With the framework illustrated in Table 4, we will reach out for expert opinions and positions through interviews and surveys. A quality-attribute-focused survey will be created, based on NoSQL database type (Key Value, Document, Columnar, and Graph), where databases are compared with regards to their suitability for quality attributes. The anticipated survey results should allow completion of blank spaces on Table 4.

This research selected a methodology based on finding qualitative measures and understanding quality attributes of NoSQL databases by leveraging the knowledge of NoSQL database specialists and other early adopters users. It will involve:

- Survey of enterprises and experts
- Anonymity of Participants
- Iterations
- Controlled Feedback
- Statistical Aggregation of Group Responses
- Research Tasks Inputs and Outputs.

Table 3: Averaged quality attributes for the four NoSQL database types. Legend: “G” = Great, “+” = Good, “A” = Average, “-” = Mediocre, “B” = Bad, and “?” = Unknown/NA.

	Key Value	Document	Columnar	Graph
Availability	+	A	A	+
Consistency	+	+	+	+
Durability	+	A	+	+
Maintainability	A	A	-	+
Read Performance	+	+	-	+
Write Performance	G	-	+	+
Recovery Time	?	+	B	?
Reliability	A	+	+	+
Robustness	A	A	-	+
Scalability	+	-	G	+
Stabilization Time	?	B	A	?

4.6 Surveys

Example of the Survey with three brief sections is listed below.

Please select the database used:

- Voldemort(Key Value)
- Redis (Key Value)
- Mongo DB (Document)
- Couch DB (Document)
- Cassandra(Columnar)
- HBase (Columnar)
- Neo4J(Graph)
- OrientDB (Graph)

Table 4: Averaged quality attributes for the four NoSQL database types with the framework proposed. Legend: “G” = Great, “+” = Good, “A” = Average, “-” = Mediocre, “B” = Bad, and “?” = Unknown/NA.

	Key Value	Document	Columnar	Graph
Availability	+	A	A	+
Consistency	+	+	+	+
Durability	+	A	+	+
Maintainability	A	A	-	+
Read Performance	+	+	-	+
Write Performance	G	-	+	+
Recovery Time	?	+	B	?
Reliability	A	+	+	+
Robustness	A	A	-	+
Scalability	+	-	G	+
Stabilization Time	?	B	A	?
Authentication				
Authorization				
Encryption				
Auditing				
Data integrity				
Confidentiality				
Documentation				
Popularity				
Maturity				
Query Possibilities				
Concurrency Control				
Latency				
Conflict Resolution				

4.6.1 Availability Survey

1. Is Horizontal Partitioning allowed? Yes, No
2. Is Vertical Partitioning allowed? Yes, No
3. Is Scaling achieved by replicating the data? (Please Select)
Synchronously
Asynchronously
Both
4. Are all replica nodes allowed to continue working even if they are disconnected? Yes, No
5. Is the NoSQL database more robust on Reading operations? Yes, No
6. Is the NoSQL database more robust on writing operations? Yes, No
7. Is the data permanently available? Yes, No
8. Is the data the same in every replication on every cluster node? Yes, No
9. Does the NoSQL database provide mechanisms that allow ACID data integrity? Yes, No
10. Does the NoSQL database provide mechanisms that allow eventually consistent BASE? Yes, No

4.6.2 Security Survey

1. Does the NoSQL database provide mechanisms that allow encryption techniques? Yes, No
If “Yes” Please name the encryption technique
If “No” what mechanisms are used to enforce data confidentiality?
2. What encryption level does the NoSQL database provide? (Please Select)
 - Data at rest
 - Client to server communication
 - Server to Server connection
3. Does the NoSQL database provide mechanisms that allow Auditing or Audit Logs? Yes, No
4. Does the NoSQL database provide mechanisms that allow Authentication? Yes, No
If “Yes” please mention the Authentication method
5. Does the NoSQL database provide mechanisms that allow Authorization? Yes, No
If “Yes” please mention the Authorization method
6. Is Replication on the database allowed? Yes, No
7. How often are backups tested? Daily, Weekly
8. How often is Disaster Recovery Infrastructure tested? Daily, Weekly

4.6.3 Additional Quality Attributes Survey

1. What is the maturity of the Database - API in the market according to your experience)? (Please Select)
 - Mature (5- 10 years)
 - Growing (2 - 5 years)
 - Start Up (0 to 2 years)
2. In what Business Area the NoSQL Database is used in your environment? (Please Select)
 - Medical
 - Financial
 - Retail
 - Social Media
 - Other (Please Describe)
3. How long have the database been used in the Business Area selected previously? (Please Select)
 - Popular (5- 10 years)
 - Growing (2 - 5 years)
 - Start Up (0 to 2 years)
 - Other (Please Describe)
4. Does the NoSQL database provide SQL like query possibilities? Yes, No
If “Yes” please mention the SQL Query name
5. Does the NoSQL database provide features to manage Concurrency Control? (Please Select)

Optimistic concurrency

Multi Version concurrency

Other

If “Other” please mention the Concurrency Control

6. Is the NoSQL database coder friendly? Yes, No
7. Is the NoSQL database used as a caching Layer? Yes, No
8. Is the NoSQL database used on Real Time Analysis? Yes, No
9. Is the NoSQL database used on Analytics? Yes, No
10. Does the NoSQL database provide End User Documentation? Yes, No

The result of this methodology will be a quality attribute framework and risks analysis of adopting NoSQL databases, which will aid software engineers and architects in their decision process when selecting a NoSQL database according to their software quality, attributes requirements.

5 CONCLUSION

There are a number of NoSQL data stores that can be classified into four different types. However, there is no Quality Software Framework that can help managers decide which NoSQL databases are the most appropriate for their Business Use Case.

The diversity of NoSQL data stores present challenges to differentiate and get a perspective of which databases is the most suited, establishing paths and opportunities for future research.

Sophisticated Security and Privacy provisions need to be explored. At the corporate level, companies and institutions need to develop software technology that offers Security features at the minimum similar if not better than the ones used by Relational Databases.

Considering a previous study’s comparison of NoSQL databases and their quality attributes, the contribution of this research includes Security attributes, enhances Availability attributes and adds Additional Quality Attributes to define a Quality Attribute Evaluation and Risk Analysis of NoSQL databases framework that will benefit the NoSQL adoption in the long term.

The framework proposed will help IT departments align perceived risks of NoSQL database adoption with actual risks measuring quality metrics associated with Availability and Security, which are critical to choosing the right NoSQL database for a given domain and to making better software development and design decisions, giving software

engineers and architects a better way to select a NoSQL database for a particular big data environment and domain.

REFERENCES

- Barmpis Konstantinos, Kolovos Dimitrios S. 2012, 'Comparative Analysis of Data Persistence Technologies', Proceedings of the 2012 Extreme Modeling Workshop, New York.
- Bazar Christina, Cosmin Sebastian 2014, 'The Transition from RDBMS to No SQL. A comparative Analysis of three popular No Relational solutions: Cassandra, Mongo DB and CouchBase', Database Systems Journal, University of Economic Studies, Bucharest, Romania, Romania.
- Catell Rick 2010, 'Scalable SQL and No SQL Data Stores', ACM SIGMOD Record, ACM Digital Library, New York.
- Doshi Kshitij A., Zhong Tao, Lu Zhongyan, Tang Xi 2013, 'Blending SQL and NewSQL Approaches, Reference Architectures for Enterprise Big Data Challenges', International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC).
- Fowler Adam 2014, 'Why use a NoSQL Database, and why not?', Adam's Blog.
- Franks Bill 2012, Taming the Big Data Tidal Wave, finding opportunities in Huge Data Streams with Advanced Analytics, 1st edn, Wiley.
- Grolinger Katarina, Higashino Wilson A, Tiwari Abhinav, Capretz Miriam 2013, 'Data management in cloud environments: No SQL and New SQL data stores', Journal of Cloud Computing, Springer Open, London.
- Gualtieri Mike 2010, 'No SQL and Elastic Caching Platforms Are Kissing Cousins', Forrester, http://blogs.forrester.com/application_development/2010/02/nosql.html.
- Himmel Azua Maria 2012, Qualitative Analysis of Cloud Computing Risks and Framework for the Rationalization and Mitigation of Cloud Risks, Pace, New York.
- Intersimone David 2010, 'The end of SQL and relational databases?', Computer World, February 2010.
- Kroenke David M., ADJ, Database Processing, 13th edn.
- Lakshman Avinash 2008, 'Cassandra – A structured storage system on a P2P Network', <https://www.facebook.com/notes/facebook-engineering/cassandra-a-structured-storage-system-on-a-p2p-network/24413138919/>.
- Lourenço João Ricardo, Cabral Bruno, Carreiro Paulo, Vieira Marco, Bernardino Jorge 2015, 'Choosing the right NoSQL database for the job: a quality attribute evaluation', Springer Open Journal, no. Journal of Big Data.
- Nepa Suryal, Ranjan Rajiv, Kwang Kim, Choo Raymond 2015, 'Trustworthy Processing of Healthcare Big Data in Hybrid Clouds', IEEE Cloud Computing.
- Obijaju Mike 2015, 'No SQL NoSecurity – Security issues with No SQL Database', June 2015, <https://blogs.perficient.com/enterpriseinformation/2015/06/22/nosql-nosecurity-security-issues-with-nosql-database/>.
- Orend Kai 2012, Analysis and Classification of No SQL Databases and Evaluation of their ability to replace an Object Relational Persistence Layer, University of Munchen.
- Orient Database 2016, 'Orient DB', <http://orientdb.com/orientdb/>.
- Penchikala Srini 2011, 'Distributed Cache as a No SQL Data Store?', InfoQ, <http://www.infoq.com/news/2011/11/distributed-cache-nosql-data-sto>.
- Redmond Eric, WJR 2012, Seven Databases in Seven Weeks, The Pragmatic Programmers.
- Robin Hecht, SJ 2011, 'No SQL Evaluation', International Conference on Cloud and Service Computing.
- Robin Hecht 2011, 'No SQL Evaluation', International Conference on Cloud and Service Computing, IEEE Computer Society, Hong Kong.
- Sadalage Pradmod J., Fowler Martin 2013, NO SQL Distilled. A brief guide to the emerging world of Polyglot Persistence, 1st edn, Addison Wesley.
- Stonebracker Michael, 'New SQL: An Alternative to NoSQL and Old SQL for New OLTP Apps'.
- Tudorica Bogdan George., Bucur Christian 2011, 'A comparison between several NoSQL databases with comments and notes', 2011 10th International Conference RoEduNet, IEEE.
- Wikipedia, 'Cassandra', Wikipedia, <https://en.wikipedia.org/wiki/Cassandra>.

A Comparative of Spanish Encoding Functions

Efectiveness on Record Linkage

María del Pilar Angeles and Noemi Bailón-Miguel

Facultad de Ingeniería, Universidad Nacional Autónoma de México, Ciudad de México, México
pilarang@unam.mx, mimibailon@hotmail.com

Keywords: Data mining; Data matching; Record linkage; Data cleansing.

Abstract: Many business within big data projects suffer from duplicate data. This situation seriously impedes to managers to make well informed decisions. In the case of low data quality written in Spanish language, the identification and correction of problems such as spelling errors with English language based coding techniques is not suitable. In the case of Spanish language, written information is pronounced equal. There are phonetic techniques for duplicate detection that are not oriented to the Spanish language. Thus, the identification and correction of problems such as spelling errors in Spanish texts with such techniques is not suitable. In this paper we have implemented, modified and utilized in SEUCAD (Angeles, 2014) three Spanish phonetic algorithms to detect duplicate text strings in the presence of spelling errors in Spanish. The results were satisfactory, the Phonetic Spanish algorithm performed the best most of the time, demonstrating opportunities for an improved performance of Spanish encoding during the record linkage process.

1 INTRODUCTION

Data matching allows the following enterprise big data characteristics: a) optimizing the use of storage resources by eliminating redundant and possibly inconsistent data, hence reducing storage costs; b) enhancing the enterprise data quality through tighter governance on the consolidated hub; c) in order to execute more applications on this big data resource, you should be able to develop more powerful analytics, via MapReduce, YARN, R, and other programming frameworks.

Data de-duplication, results in better cache utilisation and less disk I/O. De-duplication is useful at any scale. In fact, most modern data warehousing products use column-based compression to achieve high de-duplication ratios and to improve performance. In the case of "text" big data, data de-duplication is highly recommendable. After all, the fastest and effective an I/O is, the least I/O required.

The way the companies handle their data makes its information more compressible too. For instance, the record linkage algorithms allow a better use of physical storage, reduce RAM, the information retrieval and its analysis are enhanced as there is no need to store the name of a person twice, besides the risk of being inconsistent.

Compression and deduplication play a key role in big data; In terms of economics, if a business system demands more storage resources than the competing systems, and the analysis takes longer, it will struggle to compete. The problem of detection and classification of duplicate records during the integration of disparate data sources affects business competitiveness. A number of encoding, comparison and classification methods have been utilized until now, but there still some work to do in terms of effectiveness and performance.

The present research was focused on the implementation and enhancement of Spanish encoding functions in order to improve the performance of the encoding phase during entity resolution when data has been written in Spanish language.

We have developed a prototype called Universal Evaluation System of Data Quality (SEUCAD) (Angeles, 2014) on the basis of the Freely Available Record Linkage System (FEBRL) (Christen, P. 2008). Within SEUCAD, there has been previously compared the Phonex, Soundex, and Modified Spanish phonetic functions in (Angeles, 2015). The Spanish phonetic coding was proposed in (Amon, 2015), which is an extended Soundex coding, where Spanish characters have been added. Besides, we have modified the Spanish Phonetic Algorithm so

the encryption code is resizable, and all white spaces are removed during encoding. The previous comparison showed that the modified version of the Spanish Phonetic Algorithm had a better performance in terms of precision. However, during the present research we have implemented two more Spanish encoding functions: the Spanish Metaphone algorithm (Philips, 2000), (Mosquera, 2012), and a second version of such algorithm, which applies same code to similar sounds derived from very common misspellings.

The present paper is organized as follows: The next section briefly explains the data matching process. Section 3 explains the phonetic encoding functions proposed from previous research, the enhancements we have implemented on some of them, along with their role within the process of data matching. Section 4 presents the experiments carried out, and analyses the results. Finally, the last section concludes the main topics achieved regarding the performance of the encoding functions and the future work to be done.

2 RELATED WORK

The data matching process is mainly concerned to the record comparison among databases in order to determine if a pair of records corresponds to the same entity or not (Christen, 2012). It is also called record linkage or de-duplication. This process in general terms consists on the following tasks:

A standardization process (Christen, 2012), which refers to the conversion of input data from multiple databases into a format that allows correct and efficient record correspondence between two data sources.

Phonetic encoding is a type of algorithm that converts a string into a code that represents the pronunciation of that string. Encoding the phonetic sound of names avoids most problems of misspellings or alternate spellings, a very common problem on low quality of data sources.

The indexing process aims to reduce those pairs of records that are unlikely to correspond to the same real world entity and retaining those records that probably would correspond in the same block for comparison; consequently, reducing the number of record comparisons. The record similarity depends on their data types because they can be phonetically, numerically or textually similar. Some of the methods implemented within our prototype SEUCAD are for instance, Soundex (Odell, 1918), Phonex, Phonix (Christen, 2012), NYSIIS

(Borgman, 1992), Double Metaphone (Philips, 2000).

Field and record comparison methods provide degrees of similarity and define thresholds depending on their semantics or data types. In the prototype, the algorithms Qgram, Jaro - Winkler Distance (Jaro, 1989), (Winkler, 1990), longest common substring comparison are already implemented.

The classification of pairs of records grouped and compared during previous steps is mainly based on the similarity values that were already obtained, since it is assumed that the more similar two records are, there is more probability that these records belong to the same entity of the real world. The records are classified into matches, not matches or possible matches.

The aim of the following section is to briefly explain the phonetic encoding functions that we have implemented and enhance in order to quantify and compare their performance during the record linkage process.

3 PHONETIC ENCODING PROPOSALS TO COMPARE

3.1 Phonetic Coding Functions

Phonetic encoding is a type of algorithm that converts a string (generally assumed to correspond to a name) into a code that represents the pronunciation of that string. Encoding the phonetic sound of names avoids most problems of misspellings or alternate spellings, a very common problem on low quality of data sources.

3.2 Spanish Phonetic

The Spanish phonetic coding function compared in the present document is a variation of the Soundex algorithm. Soundex is a phonetic encoding algorithm developed by Robert Russell and Margaret Odell in (Odell, 1918), and patented in 1918 and 1922. It converts a word in a code (Willis, 2002). The Soundex code is to replace the consonants of a word by a number; if necessary zeros are added to the end of the code to form a 4-digit code. Soundex choose the classification of characters based on the place of articulation of the English language.

The limitations of the Soundex algorithm have been extensively documented and have resulted in several improvements, but none oriented to the

Spanish language. Furthermore, the dependence of the initial letter, the grouping articulation point of the English language, and the four characters coding limit are not efficient to detect common misspellings in the Spanish language. The Spanish phonetic coding was proposed in (Amon, 2012), it is an extended Soundex coding, where Spanish characters have been added. In general terms the algorithm is as follows:

The string is converted to uppercase with no consideration of punctuation signs. The symbols "A, E, I, O, U, H, W" are eliminated from the original word. Assign numbers to the remaining letters according to Table 1.

Table 1: Spanish Coding

Characters	Digit
P	0
B, V	1
F, H	2
T, D	3
S, Z, C, X	4
Y, LL, L	5
N, Ñ, M	6
Q, K	7
G, J	8
R, RR	9

We have modified the Spanish Phonetic Algorithm (Angeles, 2014) so the encryption code is resizable, and all white spaces are removed during encoding. This model allows us to analyse a larger number of cases where we can have misspellings. The modified Spanish phonetic algorithm is called as soundex_sp in our SEUCAD prototype.

3.3 The Spanish Metaphone Algorithm

The Metaphone is a phonetic algorithm for indexing words by their English sounds when pronounced, it was proposed by Lawrence Philips in 1990 (Philips, 2000). The English Double-Metaphone algorithm was implemented by Andrew Collins in 2007 who claims no rights to this work. The Metaphone port adapted to the Spanish Language is authored by Alejandro Mosquera in (Mosquera, 2012); we have implemented this function and called as Esp_metaphone in our SEUCAD prototype. Some

of the changes applied in order to adjust to the Spanish language are shown in Table 2, which considers typical cases of the Spanish language with letters such as á, é, í, ó, ú, ll, ñ, h.

Table 2: Spanish Metaphone

Char	Replacement
á	A
ch	X
C	S
é	E
í	I
ó	O
ú	U
ñ	NY
ü	U
b	V
Z	S
ll	Y

3.4 Modified Spanish Metaphone Coding Function

In Spanish language there are words such as "oscuro", "obscuro" or "combate", "convate" that should share the same code because even they are written different, their sound is similar and the misspelling is common. The second version of Esp_metaphone contains the following enhancements:

The Royal Academy of the Spanish Language reviewed words that originally were written with "ps" as "psicología", and introduced some changes, because "the truth is that in Castilian the initial sound ps is quite violent, so the ordinary, both in Spain and in America, it is simply pronounced as "sicología". Moreover, our language, differing French or English, is not greatly concerned to preserve the etymological spelling; He prefers the phonetic spelling and therefore tends to write as it is pronounced." (Toscano-Mateus, 1965). Words that begin with "ps" can be written and pronounced as "s", and are called silent letters; for example, words "psicólogo" and "sicólogo". We have added some cases to the Spanish Metaphone algorithm in order to consider these possible variations in Spanish written words and to assign the same code in both

cases. Therefore, in case there is a word that starts with “ps”, it will be replaced by “s”. A special case with silent letter is presented with words like “oscuro” and “obscuro”, where both words have the same meaning so that the use of both is correct. In this case both its meaning and pronunciation is usually the same. Then, in case there is a word that starts with “bs”, it shall be replaced by “s”. One case of a common misspelling in Spanish language is given with words like “tambien” and “tanbien” were the latter is orthographically wrong, but phonetically is very similar to the former, and in case of typos, the letter “n” is close to letter “m” in a keyboard. Thus, we have decided to replace "mb" by "nb" and assign the same code. We have decided to replace "mp" by "np" and assign the same code in case of words such as “tampoco” and “tanpoco”. The words that begin with “s” followed by a consonant are replaced by 'es' such as “scalera” and “escalera”. Table 3 shows the additions contained in the Spanish Metaphone version 2.

Table 3: Modified Spanish Metaphone

Char	Replacement
mb	nb
mp	np
bs	S
ps	s

Table 4 shows coding from Metaphone and Metaphone_v2, the former is not able to apply the same code to words “psiquiatra“, “siquiatra“, “oscuro“, “obscuro“, “combate“, “convate“, “conbate“. All these words have the same meaning and in order to identify duplicates they should have the same code. In the case of code generated by Metaphone_v2 the code is the same, although there are not identical texts because of spelling mistakes but same meaning.

Table 4: Spanish Metaphone and Spanish Metaphone V2 coding

Word	Metaphone	Metaphone_v2
Caricia	KRZ	KRZ
Llaves	YVS	YVZ

Word	Metaphone	Metaphone_v2
Paella	PY	PY
Cerilla	ZRY	ZRY
Empeorar	EMPRR	ENPRR
Embotellar	EMVTYR	ENVTYR
Hoy	OY	OY
Xochimilco	XXMLK	XXMLK
Psiquiatra	PSKTR	ZKTR
siquiatra	SKTR	ZKTR
Obscuro	OVSKR	OZKR
Oscuro	OSKR	OZKR
Combate	KMBT	KNVT
Convate	KNVT	KNVT
Conbate	KNBT	KNVT
Comportar	KMPRTR	KNPRTR
Conportar	KNPRTR	KNPRTR
Zapato	ZPT	ZPT
Sapato	SPT	ZPT
Escalera	ESKLR	ESKLR
scalera	ESKLR	ESKLR

4 EXPERIMENTS

We have been developed and executed a set of experiments within the record linkage process through four scenarios; each scenario contains a different data-source. These experiments are aimed to identify for each data-set which encoding function has the best performance. The performance of the record linkage process is measured in terms of how many of the classified matches correspond to true real-world entities, while matching completeness is concerned with how many of the real-world entities that appear in both databases were correctly matched (Christen, 2012), (Churches, 2002). Each of the record pair corresponds to one of the following categories: True positives (TP): These are the record pairs that have been classified as matches and are true matches. These are the pairs where both records refer to the same entity. False positives (FP): These are the record pairs that have been classified as matches, but they are not true matches. The two records in these pairs refer to two different entities. The classifier has made a wrong decision with these record pairs. These pairs are also known as false matches. True negative (TN): These are the record pairs that have been classified as non-matches, and they are true non-matches. The two records in pairs in this category do refer to two different real-world entities. False negatives FN): These are the record pairs that have been classified as non-matches, but they are actually true matches. The two records in these pairs refer to the same entity. The classifier has made a wrong decision with these record pairs. These pairs are also known as false non-matches. Precision: calculates the proportion of how many of the classified matches (TP + FP) have been correctly classified as true matches (TP). It thus measures how precise a classifier is in classifying true matches (Odell, 1918). It is calculated as: $precision = TP / (TP + FP)$. F-measure graph: An alternative is to plot the values of one or several measures with regard to the setting of a certain parameter, such as a single threshold used to classify candidate records according to their summed comparison vectors, as the threshold is increased, the number of record pairs classified as non-matches increases (and thus the number of TN and FN increases), while the number of TP and FP decreases.

An ideal outcome of a data matching project is to correctly classify as many of the true matches as true positives, while keeping both the number of false positives and false negatives small. Based on the number of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN),

different quality measures can be calculated. However, most classification techniques require one or several parameters that can be modified and depending upon the values of such parameters, a classifier will have a different performance leading a different numbers of false positives and negatives. Figure 1 shows the structure and sample source data utilized for experimentation.

nombre	apellido_pat	apellido_mat	calle
santiago		gonzalez	calle de san gumerindo
david	hernandez	cruz	calle de amedillo
jessica	perez	martinez	calle de barbara de braganz
martha	sanchez	lopez	calle de jordi sole tura
patricia	garcia	aviles	calle de santa maria reina
alfonso	garcia	hernandez	calle del iridio
adriana	vazquez	gonzalez	calle de jose espellius
tania	mendez	lopez	plaza de arguelles
vicente		reyes	calle de infiesto
angelica	hernandez	brito	calle de los hermanos carpi
maria elena	perez	ramirez	calle de los bascones
isaac	martinez	gutierrez	calle de julia garcia bostan
berenice	ramirez	reyes	calle de elvira bardios
alejandro	alonso	flores	calle de la anunciacion
enrique	cordero	ramirez	calle del gladiolo

Figure 1: Sample of data source

The configuration of indexing, comparison and classification for all scenarios has been the same and repeated for each encoding function (Esp-Metaphone, Esp_metaphone_v2 and Soundex_sp). Such configuration is presented as follows:

1. Indexing:

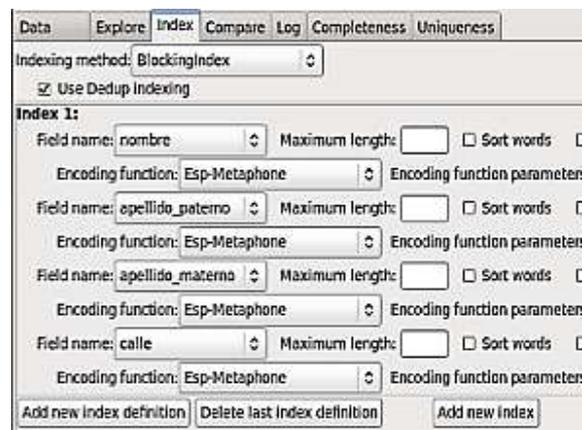


Figure 2: Indexing and encoding configuration

Fields that form the record require to be encoded and indexed in order to avoid a large number of comparisons between records whose fields are not even similar. Then, during the coding phase, we have executed for each experiment one of the coding functions: esp-metaphone, esp_metaphone_v2 or

soundex_sp. We have chosen "Blocking index" as indexing method based on fields: "nombre", "apellido paterno", "apellido materno", "calle".

Figure 2 shows the configuration utilized for indexing and encoding methods.

2. Comparison: Once records have been ordered and grouped in terms of the previous fields specified. Each encoded field will be compared.

In order to obtain quality measures during the comparison step, we have chosen an exact function "Str-Exact", with "nombre", "apellido paterno", "apellido materno", "calle" fields.

Figure 3 shows the comparison specification for the experiments.

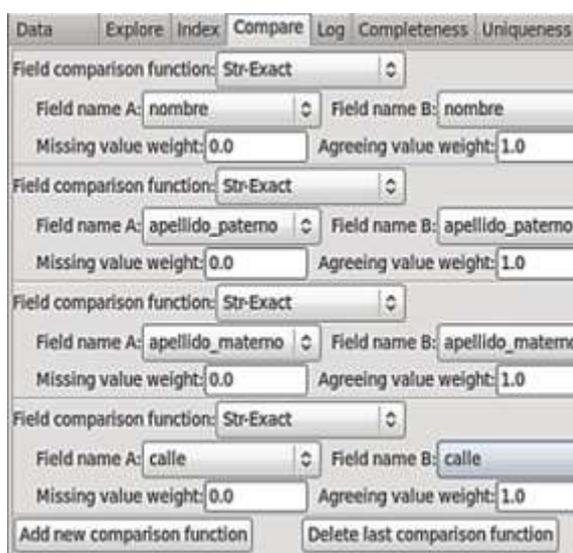


Figure 3: Comparison by String Exact method

3. Classification: In the case of pairs of record classification, we have selected the Optimal Threshold method, with a minimized false method of Positives and negatives, and a bin width of 40 for the range of values to be considered for the output graphic.

Figure 4 shows the classification configuration for the experiments.

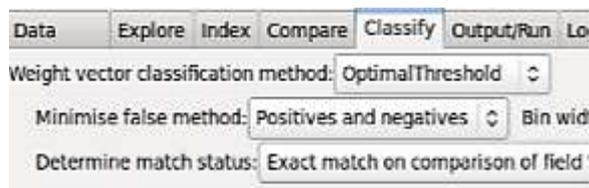


Figure 4: Classification by Optimal Threshold

4.1 Scenario I

The first file was generated with a total length of 1000 records, 100 duplicated records, one duplicated record for an original record as maximum, one change field per item as maximum, one maximum record modification, with a uniform probability distribution for duplicates.

The quality metrics obtained for each encoding method are presented in Table 5.

Table 5: Quality Metrics for Scenario I

Encode Method	Total Classif.	TP	FP	Precision	F-measure
Metaphone_sp	68	65	3	0.95588	0.977443
Metaphone_v2	69	66	3	0.95652	0.977777
Soundex_sp	76	73	3	0.96052	0.979865

According to the outcomes obtained from the first scenario, we can observe that in the case of the Modified Spanish coding function (soundex_sp), there were 76 record pairs classified, with 73 duplicated record pairs as true positives and 3 record pairs as false positives. Therefore, this method was 96% precise, slightly higher than the rest.

4.2 Scenario II

The second data source contained a total length of 5000 records, 500 duplicated records, one duplicated record for an original record as maximum, one change field per item as maximum, one maximum registry modification, with a uniform probability distribution for duplicates.

The quality metrics obtained for each encoding method are presented in Table 6.

Table 6: Quality Metrics for Scenario II

Encode Method	Total Classif.	TP	FP	Precision	F-measure
Metaphone_sp	320	319	1	0.9968	0.9984
Metaphone_v2	341	340	1	0.99706	0.99853
soundex_sp	353	352	1	0.99716	0.99581

From Table 6 we can observe that the Modified Spanish function classified 353 record pairs, with 352 duplicated record pairs as true positives and 1 record pair mistakenly classified as true match, corresponding then as one false positive. Therefore, this method was 99.7% precise, with more records

classified than the Metaphone_sp and Methaphone_v2 with 320 and 341 records classified respectively.

4.3 Scenario III

The third data source contained a total length of 10000 records, 5000 duplicated records, one duplicated record for an original record as maximum, one change field per item as maximum, one maximum registry modifications, with a uniform probability distribution for duplicates. The process of record linkage under this scenario showed that the Modified Spanish coding function classified 3622 record pairs out of a total of 5000 potentially to detect, with 3620 duplicated record pairs as true positives and 2 record pairs mistakenly classified as true match. Therefore, this method was 99.94% precise. The Metaphone_sp and Methaphone_v2 phonetic functions obtained less records classified and more false positives than Spanish soundex function. The quality metrics obtained for each encoding method are presented in Table 7.

Table 7: Quality Metrics for Scenario III

Encode Method	Total Classif.	TP	FP	Precision	F-measure
Metaphone_sp	3333	3324	9	0.997299	0.9986
Metaphone_v2	3489	3480	9	0.99742	0.9987
Soundex_sp	3622	3620	2	0.99944	0.9997

4.4 Scenario IV

The fourth file has a total length of 1000 records, 100 duplicated records, one duplicated record for an original record as maximum, two changed fields per item as maximum, three maximum registry modifications, with a uniform probability distribution for duplicates.

The Modified Spanish coding function allowed that 964 record pairs could be classified; the total number of duplicates was actually 2500 records. However, this method did not present any false positive. The rest of the phonetic algorithms were 99% precise with two false positives, but the number of classified records was lower than those with Soundex_sp. The outcomes obtained for each encoding method under scenario IV are presented in Table 8.

Table 8: Quality Metrics for Scenario III

Encide Mrtod	Total Clas s	TP	FP	Precision	measure
Metapho- ne_sp	812	810	2	0.9987536	0.998766
Metapho- ne_v2	884	882	2	0.99773	0.99886
Soundex_ sp	964	964	0	1	1

4.5 Analysis of Outcomes

According to the outcomes shown in previous section, we can observe that the Modified Spanish Phonetic algorithm was always more precise than the rest of the algorithms. Therefore, the Modified Spanish-Phonetic algorithm allows a higher proportion of how many of the classified matches (TP+FP) have been correctly classified as true matches.

The Spanish phonetic algorithm allows a total similarity greater than the remaining algorithms in all cases, because is more effective codifying Spanish words.

The Spanish phonetic algorithm achieved a slightly higher f-measure than the two versions of the Spanish Metaphone algorithm.

The graphics presented in this section, have been generated according to the variation of the coding function in order to observe the behaviour of the algorithms.

The precision obtained from each encode method for all the scenarios have been compared, graphed and shown in Figure 5, which shows the trend of the contribution of each encoding method to the precision of the classification.

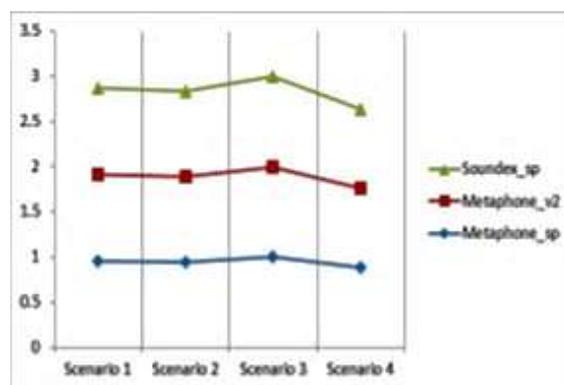


Figure 5: Precision of encode function

Figure 6 shows the trend of the contribution of each encoding method to the completeness of the classification.

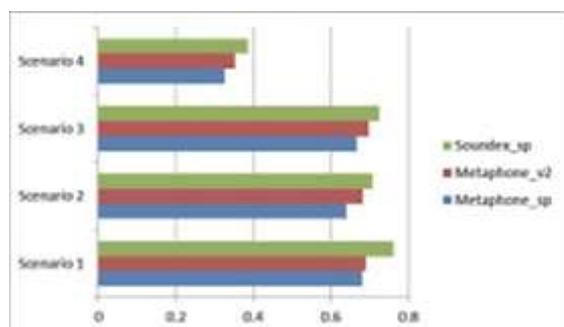


Figure 6: Completeness of each encoding method per scenario.

In other words, the proportion of record pairs classified against the total number of duplicates per scenario.

According to the outcomes shown in previous section, we can observe that the Modified Spanish Phonetic algorithm was always more precise than the two versions of Metaphone. Therefore, the Modified Spanish-Phonetic algorithm allows a higher proportion of true matches. The Spanish phonetic algorithm allows a total similarity greater than the remaining algorithms in all cases, because is more effective codifying Spanish words. The Spanish phonetic algorithm achieved a slightly higher f-measure than the rest. As we can observe from Figure 6, the Spanish phonetic algorithm obtained a larger number of pairs of records classified than the rest of the phonetic algorithms.

5 CONCLUSION

There are very real costs derived from duplicated customer data within big data.

Depending on the functional area (marketing, sales, finance, customer service, healthcare, etc.) and the business activities undertaken, high levels of duplicate customer data can cause hundreds of hours of manual reconciliation of data, sending information to wrong addresses, and decrease confidence in the company, increase mailing costs, increase resistance to implementation of new systems result in multiple sales people, sales teams or collectors calling on the same customer.

The present work has evaluated the record linkage outcomes under a number of different

scenarios, where the true match status of record pairs was known. We have obtained precision, recall, and f-measure because they are suitable measures to assess data matching quality.

The Modified Spanish Soundex function presented a better performance than the rest of the phonetic functions during most of the experiments. However, it takes the longest execution time with a difference of some milliseconds.

It is important to be aware that the performance of a de-duplication system or technique is dependent on the type and the characteristics of the involved data sets, having good domain knowledge is relevant in order to achieve good matching or deduplication results.

We have previously concluded in (Angeles, 2015) that the Modified Spanish Phonetic algorithm was always more precise and complete than Soundex y Phonex.

Under a new set of experiments we have carried out against a Spanish version of the Metaphone algorithm and an enhanced version of the Spanish Metaphone, the Modified Spanish Phonetic algorithm still having the best performance in terms of precision in the majority of the cases we have experimented during the present research.

ACKNOWLEDGEMENTS

This work is being supported by a grant from Research Projects and Technology Innovation Support Program (Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica, PAPIIT, UNAM Project 1N114413 named Universal Evaluation System Data Quality (Sistema Evaluador Universal de Calidad de Datos).

REFERENCES

- Angeles, P., et al., 2014. Universal evaluation system data quality, In *DBKDA 2014 : The Sixth International Conference on Advances in Databases, Knowledge, and Data Applications*, vol. 32, pp. 13–19.
- Angeles, P., J. García-Ugalde, A. Espino-Gamez, and J. Gil-Moncada, Comparison of a Modified Spanish Soundex, and Phonex Coding function during datamatching process, In *International Conference on Informatics, Electronic and Vision, ICIEV, Kytakyushu, Fukuoka Japan, ISBN:978-1-4673 6901-5*, DOI:10.1109/ICIEV.2015.7334028, IEEE, pp.1-6,2015.

- Borgman, C.L. & S. L. Siegfried, 1992. Getty's synonym TM and its cousins: A survey of applications of personal name-matching algorithms. In *Journal of the American Society for Information Science* 43(7), 459-476.
- Christen, P., 2008. Febrl A Freely Available Record Linkage System with a Graphical User Interface. *Second Australasian Workshop on Health Data and Knowledge Management (HDKM 2008)*, 80, 17-25. Figure 6: Completeness of each encoding method per scenario.
- Christen, P., 2012. Data Matching: Concepts and Techniques for Record Linkage, Entity Resolution and Duplicate Detection. *Springer Data-Centric Systems and Applications*.
- Churches, T., P. Christen, K. Lim, & J. X. Zhu, 2002. Preparation of name and address data for record linkage using hidden Markov models. In *BMC Medical Informatics and Decision Making* 2 (1), 9.
- Cohen, W. W., P. Ravikumar, & S. E. Fienberg, 2003. The book, *A comparison of string distance metrics for name matching tasks*, 73-78.
- Rahm E. & H. Do, 2000. Data cleaning: Problems and current approaches. In *IEEE Data Engineering* 23 (4), 3-13.
- Amon F. M. I. & J. Echeverria, 2012. Algoritmo fonético para detección de cadenas de texto duplicadas en el idioma español. In *Ingenierías Universidad de Medellin* 11 (20), 120-138.
- Jaro, M. A., 1989. Advances in record-linkage methodology applied to matching the 1985 Census of Tampa, Florida. In *Journal of the American Statistical Association*, 84, 414-420.
- Mosquera, A., E. Lloret, & P. Moreda, 2012. Towards Facilitating the Accessibility of Web 2.0 Texts through Text Normalisation. In *Proceedings of the LREC workshop: Natural Language Processing for Improving Textual Accessibility*, 9-14.
- Odell, M. & R. Russell, 1918. The book, *The soundex coding system*. (1261167).
- Philips, L., 2000. The double metaphone search algorithm. In *C/C++ Users J* 18 (6), 38-43.
- Toscano-Mateus, H., J. B. Powcrs, (ed.), 1965. *Hablemos del lenguaje*.
- Winkler, W., 1990. String comparator metrics and enhanced decision rules in the Fellegi-Sunter model of record linkage. In *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 354-359.

AUTHOR INDEX

AUTHOR INDEX

Angeles, M.	105	Minchev, D.	33
Barlinn, J.	89	Mitrakos, D.	78
Boncori, J.	15	Mouak, A.	25
Cesaroni, C.	15	Moy, C.	71
Demarchi, L.	43	Mozhaeva, N.	5
Denisov, A.	61	Mróz, M.	43
Donati, G.	53	Musicò, E.	15
El Hadani, D.	25	Musolff, C.	53
Er-Raji, A.	25	Palicot, J.	61
Franceschi, G.	15	Perez, G.	97
Friedrich, G.	53	Piégay, H.	43
Gentile, V.	43	Popp, D.	53
Gottardo, G.	53	Qing, X.	61
Ibnelhobyb, A.	25	Qiu, J.	61
Kosikowski, R.	89	Radgui, A.	25
Kozak, K.	89	Seu, R.	15
Kunka, B.	89	Smiej, F.	25
Lan, S.	61	Soldovieri, F.	61
Lazarov, A.	33	Spitoni, M.	43
Lejot, J.	43	Spogli, L.	15
Liu, H.	61	Spyrou, E.	78
Louet, Y.	71	Tamtaoui, A.	25
Mackin, H.	97	Tappert, C.	97
Maltseva, O.	5	Unlu, E.	71
Man, Z.	61	Yin, X.	61
Merdas, M.	25		
Miguel, N.	105		
Minchev, C.	33		

ICTRS 2016



Proceedings of ICTRS 2016
Fifth International Conference on Telecommunications and Remote Sensing
ISBN: 978-989-758-200-4
<http://www.ictrs.org>