geoENV2024 Book of Abstracts

Editors:

J. Jaime Gómez-Hernández

Emmanouil Varouchakis

Dionissios T. Hristopulos

George Karatzas

Philippe Renard

Maria João Pereira

Editors:
J. Jaime Gómez-Hernández
Emmanouil Varouchakis
Dionissios T. Hristopulos
George Karatzas
Philippe Renard
Maria João Pereira

Editorial:
Internationa Association of Hydrogeology – Spanish Chapter

ISBN: 978-84-920529-9-8

geoENV, the Conference on Geostatistics for Environmental Applications, shows the state of the art of geostatistics in environmental applications with new cases, results and relevant discussions from leading researchers and practitioners around the world.

geoENV conferences have been held biannually at venues across Europe. From the first conference in Lisbon in 1996, the event has been staged in Valencia (1998), Avignon (2000), Barcelona (2002), Neuchâtel (2004), Rhodes (2006), Southampton (2008), Gent (2010), Valencia (2012), Paris (2014), Lisboa (2016), Belfast (2018), Parma (2021 virtual, 2022) and has become established as a leading forum for scientists across a broad range of disciplines to share their experiences on the application of geostatistics to environmental problems.

This 24 marks the 15th edition and this book of abstracts is just a small window on the presentations made in Chania from June 19th to 24th.

Jaime Gómez-Hernández
Emmanouil Varouchakis
Dionissios T. Hristopulos
George Karatzas
Philippe Renard
Maria João Pereira
Editors

Organizers

Sponsors (with thanks)

Organizing Committee

Emmanouil A. Varouchakis
Assistant Professor
Technical University of Crete
School of Mineral Resources Engineering

George P. Karatzas
Professor
Technical University of Crete
School of Chemical & Environmental Engineering

Dionissios T. Hristopulos
Professor
Technical University of Crete
School of Electrical & Computer Engineering

J. Jaime Gómez-Hernández
Professor
Universitat Politècnica de València
School of Civil Engineering

Maria João Pereira
President of geoENVia
Universidade de Lisboa
Instituto Superior Técnico

Philippe Renard
Secretary of geoENVia
University of Neuchâtel

Scientific Committee

Rachid Ababou
Institut de Mecanique des Fluides de Toulouse – France

Teresa Albuquerque
IPCB ICT CERNAS – Portugal

Denis Allard
BioSP, INRAE – France

Peter Atkinson
Lancaster University – United Kingdom

Leonardo Azevedo
CERENA, Instituto Superior Técnico, Universidade de Lisboa – Portugal

Peter Atkinson
Lancaster University – United Kingdom

Sofia Barbosa
NOVA School of Science and Technology – Portugal

Andreas Bardossy
University of Stuttgart – Germany

Anastasia Baxevani
University of Cyprus – Cyprus

Patrick Bogaert
UCLouvain – Belgium

Pauline Collon
Université de Lorraine – France

Eduardo Cassiraga
Universitat Politècnica de València – Spain

Pauline Collon
Université de Lorraine – France

Nadim Copty
Bogazici University – Turkey

Gerald Corzo
IHE – Delft – Netherlands

Early Career Scientists Scientific Committee

Vanessa A. Godoy
Universitat Politècnica de València – Spain

Svenja Fischer
Universität Bochum – Germany

Sara Kasmaeeyazdi
University of Bologna – Italy

Stephanie Thiesen
Universidade Federal de Santa Catarina – Brazil

Valeria Todaro
University of Parma – Italy

Table of Contents (ordered alphabetically by title)

A Comparative Analysis of Conventional and Contemporary Stochastic Techniques in Multivariate Geostatistics for a Copper-Molybdenum Deposit

Ayana Karakozhayeva[1], Nasser Madani[1]

[1]*Nazarbayev University*

Keywords: Conditional cosimulation, PPMT, Accuracy plots, Prediction cross-plots

Abstract. The modeling of Cu and Mo in the mining industry is a comprehensive approach that enables the implementation of responsible mining techniques throughout the extraction of valuable metals. This paper intends to employ a range of geostatistical methods, particularly conventional conditional cosimulation using a linear model of coregionalization and a factor-based approach known as projection pursuit multivariate transform (PPMT), to jointly simulate two valuable metals in a copper porphyry deposit. The objective of this work is to examine the effectiveness of traditional conditional cosimulation and the PPMT approach in creating a more accurate block model for resource evaluation. A homotopic dataset of 29,450 samples containing both Cu and Mo from the ore deposit has been exploited in this study.

The exploratory data analysis over this dataset identified a significantly strong correlation between the two continuous variables. This correlation was replicated effectively by both conventional conditional cosimulation and the PPMT approach. To determine the best cosimulation results, the jackknife procedure was also utilized, wherein the dataset was partitioned into separate test and train subsets. The scatter plots comparing the true grades and predicted grades for Cu and Mo demonstrated that the simulation results, which were obtained using traditional conventional conditional cosimulation and the PPMT approach, are both reliable. This is evidenced by the fact that they accurately recreated the desired correlation with a reasonably small margin of error. Nevertheless, the level of accuracy was increased slightly using conventional conditional cosimulation for Mo. Furthermore, while analyzing the accuracy plots of each element, it was illustrated that the conventional

technique of conditional cosimulation is somewhat more accurate than the PPMT approach in this particular instance.

Acknowledgment: The authors appreciate Nazarbayev University for funding this work via the Collaborative Research Program 2023-2025 under contract No. OPCRP2023013.

A comparison between Empirical Bayes combined with Akaike's Bayesian Information Criterion and Ensemble Smoother with Multiple data assimilation to evaluate hydrogeophisical data

Camilla Fagandini[1], Marco D'Oria[1], Valeria Todaro[1], Andrea Zanini[1]

[1]*Department of Engineering and Architecture, University of Parma*

Keywords: Groundwater, ebaPEST, ES_MDA, ERT

Abstract. Knowledge of subsurface properties is critical for appropriate groundwater resource management and predicting solute transport in aquifers. Due to the necessary field investigations such as drilling logs, pumping tests, and so on, subsurface characterization necessitates a significant amount of economic resources. As a result, this subject has received a lot of attention in the literature. Electrical resistivity tomography (ERT) is a noninvasive approach for identifying the subsurface structure based on the material's electrical resistivity.

The ERT surveys yield pseudo-resistivity values and require the use of an inverse technique to estimate the resistivity field of the studied area. Two inverse procedures are applied and compared: the Empirical Bayes approach combined with Akaike's Bayesian Information Criterion (ebaPEST) and Ensemble Smoother with Multiple Data Assimilation applied via the software package genES-MDA.

ebaPEST is model-independent, written in Fortran using the PEST framework, and capable of identifying both the unknown field and the prior distribution's hyper-parameters.

The genES-MDA software package is a Python tool with a flexible workflow that can be easily modified for the solution of general inverse problems using a Kalman-based method.

These inverse processes were tested on a synthetic example of a saturated aquifer with a resistivity field built with an exponential covariance function with a variance of 0.5 and a correlation length of 30 m. A 2D model of 96 m length and 20 m depth with a 2m x 2m computational grid represents the unknown field. In both packages, the forward problem is tackled using a 2.5-D electrical

resistivity model developed with Matlab® software. It consists of computing pseudo-resistivity values while considering fully known boundary conditions and electrode configuration beginning from a resistivity field.

Both approaches can quantify uncertainty, ebaPEST via the posterior probability distribution of the target quantities and genES-MDA via the ensemble of parameter realizations.

The results demonstrate that both methods performed well. In this work, the advantages and drawbacks of the approaches are explored.

A comparison of ANN and CNN generalizability for flood susceptibility mapping

Rahma Khalid[1], Usman Khan[1]

[1]York University

Keywords: GIS, ensemble, floodplain

Abstract. Floods are the most common and severe natural disasters globally, adversely impacting environmental, social and economic welfare, through the destruction of natural ecosystems, loss of human life and infrastructure damage. Identifying flood vulnerable regions, through Flood Susceptibility Mapping (FSM), assists in proactive implementation of flood mitigation and protection strategies, facilitating risk management decision making. Machine Learning (ML) models are commonly used in FSM studies; however, their generalizability is seldom studied. To develop accurate ML models that can be used for large scale applications, this gap in knowledge should be addressed.

Artificial Neural Networks (ANNs) are frequently used in FSM studies due to their accuracy, commonly composed of one input, hidden and output layer. Recently, deep learning neural networks, such as Convolutional Neural Networks (CNNs), are becoming increasingly popular within the field of FSM, due to their efficiency and ability to preserve spatial relationships. Therefore, given their widespread use, and to facilitate comparison across FSM studies, this research investigates the generalizability of CNNs and a simple three layer ANN. An ensemble of each model was trained and tested on the Don River watershed in Southern Ontario, Canada, and their generalizability is then tested on four independent watersheds in the region.

The proposed CNN and ANN models use a variety of geospatial data in the training watershed which are obtained from remote sensing data such as digital elevation models and landcover datasets, and meteorological datasets. Each dataset has a 30m-by-30m resolution, resulting in approximately 407,000 datapoints used in the training, validation and testing of the models on the Don River watershed.

The commonly used metrics of Area Under the Curve-Receiver Operating Characteristic (AUC-ROC) and Overall Accuracy (OA) were used to assess model performance and generalization potential. Initial results demonstrate both

approaches perform well across the training and independent watersheds. Though some physical dissimilarity is present between the five watersheds, the models are able to generalize and retain the geo-physical relationships between the given inputs and flood susceptibility, resulting in accurate FSMs.

Model generalizability is critical as it facilitates FSM across multiple regions, which is particularly important in data sparse areas. Additionally, it reinforces ML models as an alternative to existing time-consuming, and computationally intensive physically-based modelling approach to FSM. Future research opportunities include applying interpretability techniques to further the understanding of the region's physical characteristics their propensity in flood susceptibility.

Acknowledgment: Natural Sciences and Research Council CGS scholarship

A copula-based approach for modelling the SST/Chl-a distributions in the Strait of Gibraltar

Amina Bensalah[1], Bouchra Zellou[2], Mofdi El Amrani[1], Mohammed Seaid[3], Nabil El Mocayd[4]

[1]Faculty of Science and Technology University Abdelmalek Essaadi, Tangier, Morocco
[2]College of Computing, University Mohammed VI Polytechnic, Benguerir, Morocco
[3]Department of Engineering, University of Durham, South Road, DH1 3LE, UK
[4]College of Agriculture and Environmental Sciences, University Mohammed VI Polytechnic, Benguerir, Morocco

Keywords: Strait of Gibraltar, Copula method, Marginal distributions, Goodness of fit

Abstract. In the present study, we propose a novel copula method as a powerful statistical tool for modelling the dependence between the Sea-Surface Temperature (SST) and Chlorophyll-a (Chl-a) in the Strait of Gibraltar region over the last 20 years. Satellite data generated by MODISA are used in our analysis for which the first step was to interpolate the data using spatial averaging followed by a spatial revealing the discordance between these variables. Subsequently, univariate marginal distributions of the SST and Chl-a are identified by fitting Normal, Log-normal, Gamma, Weibull, and Gumbel probability distributions. Initially, the current work focuses on the analysis of annual disseminations of the SST and Chl-a concentration. However, we observed bimodal distributions and p-values below 0.05 when testing for virtue of fit, indicating a severe complexity in the structure of the annual data. Based on these results, we next analyze the data by season, and we identify the MAM (March-April-May) and SON (September-October-November) seasons as the most favorable for the Strait of Gibraltar. The Normal distribution is also identified as the best fitting distribution for the SST in the SON season while the Gumbel distribution is determined as the most appropriate distribution for the MAM season for the SST. In both seasons for Chlorophyll-a, we also apply the Kolmogorov-Smirnov, Anderson-Darling and Cramer-von Mises tests.

Next, the most suitable copula function that best suits the bivariate relation among the two seasons is studied in detail. In the current work, the minimum

Akaike information criterion and the Bayesian information criterion are used to decide the most suitable copula function. For the considered data in the Strait of Gibraltar, the rotated Tawn type 2 270° copula is the best suited copula function for the SON season, and the rotated Tawn type 1 270° for the MAM season. We have also found a significant relationship between the variables, but it is not extremely strong, which could be attributed to the strong sea circulation in the Strait of Gibraltar. In the light of these results, an im-portant step for our future research will be to investigate the potential impact of climate change on the observed dynamics of the sea-surface temperature and Chlorophyll-a concentration in the region of the Strait of Gibraltar by studying how these variables might change over time.

Acknowledgment: This work was supported by the MICODIR project (Modeling the Impact of Ocean Circulation on Fish Resource Availability)

A Hierarchical Sequential Gaussian Cosimulation Algorithm with Acceptance-Rejection Sampling Technique for Mine Tailings Evaluation

Alikhan Ibraimov[1], Nasser Madani[1], Mohammad Maleki[2], Annika Parviainen[3]

[1]*School of Mining and Geosciences, Nazarbayev University*
[2]*Department of Metallurgical and Mining Engineering, Universidad Católica del Norte*
[3]*Departamento de Edafología y Química Agricola, Universidad de Granada, Avda. Fuente Nueva s/n*

Keywords: Acceptance-rejection sampling, Hierarchical Gaussian cosimulation, Inequality constraint, Mine Tailings

Abstract. Mine tailings, while posing a significant environmental threat, can potentially host substantial amounts of valuable and critical minerals that present an opportunity for re-valorization and re-mining. However, evaluation of tailings becomes a complex task, particularly when dealing with multi-element compositions. Conventional approaches often fall short of accurately reproducing the original bivariate relationship between elements of interest. In response to this challenge, our research focuses on the application of a hierarchical sequential Gaussian cosimulation algorithm coupled with acceptance-rejection sampling techniques. The proposed algorithm focuses on accurately replicating inequality constraints between two critical elements, copper and gold in a tailing deposit. In our study, we compared the outcomes of the proposed hierarchical approach against conventional methods of cosimulation, highlighting the nuanced differences in their results. Notably, our findings indicated that the proposed algorithm outperforms traditional approaches by reproducing the linearity constraints in a more precise way. By incorporating an acceptance-rejection sampling technique, the proposed technique ensures the reproduction of values within specified linearity constraints. For implementing this algorithm, first regression analysis was used to assess the linearity constraint between copper and gold in the dataset and to derive coefficients for a formula that represents the desired inequality constraint. Subsequently, the obtained formula introducing the inequality constraint was employed to cosimulate the values of the copper and gold conditions to their bivariate inequality relations by accepting or rejecting the

simulated values following whether they satisfy the given constraint. As a result, the proposed algorithm provides a more accurate and reliable representation of the bivariate relationship between copper and gold compared to traditional method of cosimulation, which ignores the inequality restriction. However, application of hierarchical Gaussian cosimulation is limited to the chemical elements with a poor correlation.

Acknowledgment: The first and second authors appreciate Nazarbayev University for funding this work via the Collaborative Research Program 2023-2025 under contract No. OPCRP2023013.

A new optimality criterion for designs for simultaneous kriging predictions

Helmut Waldl[1]

[1]Johannes Kepler University Linz, Department of Applied Statistics

Keywords: optimal experimental design

Abstract. In this talk, we further investigate the problem of selecting a set of design points for universal kriging, which is a widely used technique for spatial data analysis. We have chosen universal kriging as it is more general than ordinary kriging and there are numerous examples where this approach should be used in practical applications. Our goal is to select the design points in order to make simultaneous predictions of the random variable of interest at a finite number of unsampled locations with maximum precision. Specifically, we consider as response a correlated random field given by a linear model with an unknown parameter vector and a spatial error correlation structure. We propose a new design criterion that aims at simultaneously minimizing the variation of the prediction errors at various points. We also present an efficient technique for incrementally building designs for that criterion scaling well for high dimensions. Thus the method is particularly suitable for big data applications in areas of spatial data analysis such as mining, hydrogeology, natural resource monitoring, and environmental sciences or equivalently for any computer simulation experiment.

We compare the efficiencies of the optimal designs with respect to the newly proposed optimality criterion with those of G-optimal and V-optimal designs and we demonstrate the effectiveness of the proposed designs through an illustrative example on the optimal positioning of meteorological stations in the province of Upper Austria.

Acknowledgment: This work was partly supported by project INDEX(INcremental Design of EXperiments) I3903-N32 of the Austrian Science Fund(FWF).

A new strategy for enhancing speed of Direct Sampling

Mathieu Gravey[1]

[1]*Institute for Interdisciplinary Mountain Research (IGF), Austrian Academy of Sciences (ÖAW), Innsbruck*

Keywords: Direct Sampling, Optimisation

Abstract. In the evolving field of geostatistics, especially in multiple point statistics (MPS), the balance has historically teetered between enhancing computational speed and simulation quality. This paper introduces FastDS, a novel approach in Direct Sampling simulations, targeting this longstanding issue. Traditional algorithms struggled with features like high pattern variability, extreme events, and continuous data, requiring extensive computation time. FastDS revolutionizes this process by employing a strategy rooted in Direct Sampling, effectively reducing the simulation time for complex spatial features.

Our approach is premised on the observation that a significant portion of pattern comparisons in Direct Sampling are redundant, primarily serving to identify unsuitable patterns. By subsampling potential patterns based on their unique, complex features, FastDS drastically reduces the number of computations needed.

We delve into the history of MPS, noting the gradual evolution from the early ENESIM to SNESIM, IMPALA, DS and recent advancements like QuickSampling. Each iteration brought improvements but also introduced new challenges. FastDS builds upon this history, specifically targeting the inefficiencies in pattern matching. The methodology involves a selective process of eliminating obviously unsuitable patterns, streamlining the search for optimal matches. This is particularly effective in handling continuous data, where good patterns are exceedingly rare.

Our results indicate that FastDS significantly reduces the number of computations without compromising too much the quality of simulations. While the approach shows promise for monovariate continuous variables, extensions to multivariate cases and categorical variables are also discussed.

In conclusion, FastDS offers an optimization of Direct Sampling, focusing on improving computational efficiency. This method potentially contributes to more rapid and effective geostatistical simulations, addressing the growing demands of complex and data-intensive applications.

A Risk Assessment Methodology for Supporting Decision Making on the Climate Proofing of Electricity Distribution Networks

Aikaterini Gkika[1], Ourania-Anna Lymperi[2], Emmanouil Varouchakis[2], Andreas Pavlides[2], Efstratios Zacharis[3], Efthymios Lekkas[1]

[1]*National and Kapodistrian University of Athens*
[2]*Technical University of Crete*
[3]*Hellenic Electricity Distribution Network Operator S.A.*

Keywords: Climate Change, Electricity Distribution Networks, Extreme Weather Events, Resilience

Abstract. Modern electric power systems are critical and highly complex infrastructures that are designed to ensure a reliable electricity supply to consumers under normal operating conditions and common failures or expected disturbances. Although these infrastructures have achieved high degrees of reliability, they are however not capable of effectively dealing with unpredictable extreme situations with low-frequency rates but severe impacts, usually caused as a result of extreme weather events. In parallel, electricity grids are aging and were designed for climate bands that are now obsolete because of the changing climate and the subsequent weather extremes; this has left infrastructure operating outside of its tolerance levels. Consequently, the adverse effects of climate change on networks and their associated financial costs make it essential to establish long-term action plans and adaptation strategies to enhance network resilience.

In this paper, a methodological framework is developed to support decision-making for enhancing the climate resilience of the Hellenic Electricity Distribution Network Operator (HEDNO), allowing for the promotion of risk knowledge, and the evaluation and prioritization of the required technical interventions. The selected hazards are divided into acute climate risks (i.e., wildfire, flood, rainfall, high wind, lightning strike, snowfall, heatwave) and long-term climate risks (i.e., drought, humidity, ambient temperature rise, sea level rise, ground instability/ landslide, coastal erosion). The approach includes the performance of a comprehensive and systematic assessment of vulnerability and risk for each of HEDNO's administrative areas, following international best

practices and in particular the European Commission's technical guidance on the climate proofing of infrastructure.

The process followed consists of 2 separate phases:

• Climate vulnerability assessment: This phase includes an initial screening for identifying potential climate risks through sensitivity, exposure, and vulnerability analysis based on scientific evidence and geographical considerations.

• Climate risk assessment: A more detailed analysis is performed to assess the impact of climate risks and determine the adaptation solutions required. The purpose is to examine the longer cause-effect chains linking climate risks to how a network infrastructure project performs across several dimensions (technical, health and safety, environmental, social, financial, etc.) and to investigate interactions between different factors. During this analysis, issues that might not have been picked up by the vulnerability assessment can be identified.

This methodology has led to the development of vulnerability and risk maps per hazard for each HEDNO's region, highlighting climate change hotspots. Moreover, a more advanced approach has been adopted by conducting Monte Carlo simulations to estimate the spread of the risk for various areas in Greece as a result of three hazard groups: acute climate risks, long-term climate risks, and major risks for the electricity distribution network. The investigated areas include the geographical regions corresponding to the 5 HEDNO's Regional Departments, i.e., Attica, Central Greece, Peloponnese-Epirus, Macedonia-Thrace, and Islands. More than 10,000 simulations have been conducted on each of the investigated regions for the above hazard groups and the results are provided as histograms.

The proposed framework aims to provide a methodology resource for power utilities to implement targeted climate change adaptations and improve network resilience.

Adsorption and Optimization of a Basic Dye into NaX Zeolite: Addressing Ecotoxicological and Health Concerns of Textile Dyes with Potential Remediation Approaches for Environmental Safety

Hadda Aya Hammoudi[1], Djamel Nibou[1], Samira Amokrane[1], Ali Khan Moonis[2], Feyda Srarfi[3]

[1]*University of Science and Technology Houari Boumediene*
[2]*King Saud University*
[3]*University of Tunis El Manar*

Keywords: Adsorption, ANOVA, Water treatment, Health implications

Abstract. This study extensively characterizes NaX zeolite using various analytical methods such as X-ray diffraction (XRD), scanning electron microscopy (SEM), laser granulometry, infrared spectroscopy (IR), and X-ray fluorescence spectrometry (XRF). The focus is on understanding the physicochemical structure and properties of NaX zeolite to facilitate the removal of methylene blue, a dye commonly present in wastewater, while addressing ecotoxicological and health concerns associated with textile dyes.

A full 4-factor, 2-level factorial design and a three-dimensional response surface methodology (RSM) were used in the experiment to look at the effects of each parameter, including how they interacted with each other. The results of this analysis were presented in a Pareto chart, underscoring the critical importance of optimization within this context.

An analysis of variance (ANOVA) was employed to quantitatively measure the significant impact of each factor on the efficiency of methylene blue adsorption into NaX zeolite. These findings hold substantial relevance for the treatment of water contaminated with undesirable dyes, considering ecotoxicological and health aspects.

The conclusions of this study emphasize the pivotal role of optimization in enhancing adsorption efficiency, which could have substantial implications for water treatment technologies and the reduction of dye-induced pollution.

Acknowledgment: I warmly thank you for the good news and the opportunity to participate in the conference. I am grateful for your financial support. Thank you once again for this recognition.

Advancing Sustainable Mining: The Role of Legislation and Certification in Ecological Reclamation

Alexandra Gkoutzou[1], Anthoula Vasileiou[1], Michail, Galetakis[1], Emmanouil Varouchakis[1], Christos Roumpos[2], Georgios Louloudis[2]

[1]*Technical University of Crete*
[2]*Public Power Corporation of Greece*

Abstract. Ecological reclamation of mined areas is a crucial step in restoring ecosystem in an appropriate level that environmental balance could be achieved by 2050. Restoration process should be along sided through legislations and appropriate certification measures, which are critical tools for authorising sustainable management of mining areas before, during, and after mining operations. The European Parliament has formed regulations as instructions in order to establish a management framework. The European Directive Mining Waste (2006) outlines measures, procedures, and guidelines for preventing or reducing any adverse environmental impacts related to mining waste management. Legislation such as the Soil Monitoring Law (2023) is also crucial, emphasizing soil quality preservation and sustainable management on environmental and social levels. Certification standards play a decisive role in implementing directives, providing a framework for assessing and verifying the effectiveness of restoration projects. Some prominent certifications include the Society for Ecological Restoration (SER) certification and the International Organization for Standardization (ISO) environmental management system certification. The SER certification typically assesses projects on factors like ecological integrity, historical and cultural evaluations, and sustainable practices, aiming to enhance biodiversity, soil health, and the resilience of ecosystems. In contrast, older ISO certifications, especially ISO 14001, prioritize environmental management systems, while ISO 26000 emphasizes social responsibility. While these do not directly address ecological restoration, they offer guidelines for organizations to adopt and enforce environmentally sound practices. Recent standards like ISO 21795:2021 outline the framework and processes necessary for planning the closure and reclamation of mines, both new and operational, providing essential requirements and recommendations. These certifications are

crucial in legitimizing ecological reclamation efforts and ensuring they adhere to recognized standards, furthering the broader objective of sustainable mining practices. This study investigates the development of standards in the certification of mine rehabilitation, highlighting an increasing recognition of environmental impacts and a movement towards sustainable methods.

An integrated framework for estimating wind-wave energy potential by means of numerical and geostatistical methods.

Katerina Spanoudaki[1], Georgios V. Kozyrakis[1], Emmanouil A. Varouchakis[2]

[1]*Institute of Applied and Computational Mathematics, Foundation for Research and Technology – Hel-las, Iraklion, Crete, Greece*
[2]*School of Mineral Resources Engineering, Technical University of Crete, Greece*

Keywords: geostatistical modelling, kriging, wave energy potential, wind wave modelling

Abstract. The current work focuses on the spatiotemporal wave data analysis with the use of coupled dynamic downscaling of climatic data, numerical wave modelling and geostatistical methodologies, for estimating the Wave Energy Potential for the Aegean and Ionian Seas. The main areas of interest are the Aegean and Ionian Sea islands, with unsustainable energy production. WRF model is used to downscale the raw reanalysis ERA5 input data grid to a finer 3 x 3 Km results grid for the period 1980-2019. The downscaled climatic data are used to force NOAA/NCEP's WAVEWATCH III to obtain the significant wave height and mean wave period for the areas of interest over a fine grid of 3 x 3 Km resolution. Geostatistical modelling by means of ordinary kriging is employed to estimate significant wave height variability and wave energy potential at finer coastal scales. Results of the geostatistical analysis are cross validated with existing measurements as well as with the results obtained from the computational methods. Spatiotemporal geostatistical or other stochastic spatiotemporal approaches have not been used with marine data and detailed studies of temporal, seasonal and spatial distribution analysis of significant wave height and wave energy potential is being conducted for the first time using these methods for the Hellenic region. Updated spatial maps of the significant wave height and period and of the wave energy potential distribution for long-term seasonal changes are provided, based on results spanning a 40-year period, which reaches up to current years (2019). The proposed integrated framework is relocatable to other areas of the Mediterranean Sea and provides insight into the application of the less computationally intensive geostatistical modelling in ocean wave data.

Analysing Crop Yield Patterns on the Arable Land Using XAI

Marija Kopanja[1], Dragana Blagojević[1], Branislav Pejak[1], Sanja Brdar[1], Bernardo Maestrini[2]

[1]*BioSense Institute*
[2]*Wageningen University Research*

Keywords: Explainable Artificial Intelligence, SHAP, Yield Prediction, Remote Sensing

Abstract. Many different machine learning models have been successfully applied for crop yield forecasting, an increasingly important and popular problem in recent years due to the growing need for ensuring food security globally. The way the predictions are derived by the machine learning models is not directly evident. Providing explanations for the decision-making process, apart from increasing transparency of the model, can be beneficial in many contexts. One of them is a comprehensive understanding of spatial field variability, as some parts of a field produce more relative to the rest of the field. This study aims to provide explanations for yield fluctuations within the field by employing a technique from the domain of eXplainable Artificial Intelligence (XAI), known as SHAP (Shapley Additive Attribution Method). The experimental part involves three phases, training the machine-learning model, generating a post-hoc explanation model on the sample (pixel) level, and aggregating obtained explanations by taking into consideration spatial aspects of the pixels. Experiments are conducted on commercial wheat fields utilizing pixel-level yield monitoring data. Several machine learning regression models are trained by using vegetation indices (NDMI, NDVI, GNDVI, DVI, EVI, GCI, SAVI etc.) across crop-growth stages jointly with soil and physical-geographical properties. The models are evaluated and the best one is chosen as relevant for the other two phases. For analysis, the yield values are divided into four categories: low, medium-low, medium-high, and high yield. In the first aggregation, the obtained explanations are used to understand the main features of the decision-making process of the machine learning model per each field. Results show that GNDVI in the critical phase of the wheat growing season (grain filling in May) is the most important feature for most fields, excluding two fields, where the most influential features are DVI and GCI in the same growing phase. Accordingly, it

can be concluded that the values of several vegetation indices for the critical growing phase of wheat have a major effect on the decision-making process of the machine learning yield prediction model. To dive more into the decision-making process and spatial yield variability, the explanations are firstly aggregated for correctly and wrongly predicted yields per each yield category. The analysis for one field discovered that the wrong decisions can be attributed to the high values of the top 3 rated features, where high values of GNDVI and GCI lead to overestimation of low and medium-low yield. For correct predictions of these two yield categories, explanation plots indicate that low values of GNDVI, GCI, and DVI are main contributors to low yield correct classification. Similarly, for some other fields, it is observed that GNDVI and GCI contribute to the correct decision for all yield categories. The underestimation of the model for fields with predominantly high yield and medium-high yield categories can be explained by high silt consistency. These examples demonstrate the potential of using XAI methods to conclude the main factors for spatial crop yield variability that can be used during the season for actionable decisions.

Acknowledgment: This work is supported by the ANTARES project that has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement SGA-CSA No. 739570 under FPA No. 664387, https://doi.org/10.3030/739570 and Ministry of Education, Science and Technological Development of the Republic of Serbia, grant agreement 451-03-47/2023-01/200358.

Annual NDVI curves reveal different growth dynamics of mountain rangelands in response to weather variability

Fabio Oriani[1], Manuel K. Schneider[1], Helge Aasen[2]

[1]*Grazing Systems, Agroscope*
[2]*Water Protection and Substance Flows, Agroscope*

Keywords: mountain pastures, remote sensing, vegetation

Abstract. Mountain pastures host rich biodiversity organized in various distinct habitats. Monitoring their seasonal growth is of primary importance for nature conservation and forage production planning in the remote alpine zones. We establish an analytical framework based on Sentinel-2 images to track the growth pattern of mountain rangeland habitats over seven years (2016-2023), in relation to snow persistence and elevation changes.

We consider a 1000-km² study zone the mid-to-high elevation mountain rangelands surrounding the Swiss National Park in the Grisons Canton, Switzerland. We analyse annual curves of the spectral index NDVI, a proxy for living vegetation, for different habitats and elevation.

In order to capture the spatial and temporal variability of sparse vegetation over a wide area, the NDVI values are collected in a data dictionary with pixel attributes. This pixel-analysis approach allows easily relating local growth to the habitat type and elevation. Moreover, we compute statistical descriptors (growth slope based on the Gompertz function, maximum NDVI, season start/end, and area under the curve) and their correlation to link the seasonal growth pattern to climate variability.

The habitats show a marked difference in their growth pattern in function of their wetness until 2400 m a.s.l, while they seem to homogenize at higher elevations. In dry pastures and dwarf shrubs, growth appears to be strongly correlated to elevation changes. Growth is mainly controlled by snow persistence and partially compensated by a quick growth after snowmelt.

This workflow presents as an effective strategy to monitor the seasonal and long-term evolution of mountain rangelands in the complex alpine domain.

Assessing the impact of hydraulic conductivity variability on freshwater lenses using numerical simulations, machine learning and dimensionality reduction methods

George Kopsiaftis[1], Vasileios Christelis[3], Anastasios Doulamis[2], Nikolaos Doulamis[2], Dimitrios Kalogeras[1], Ioannis Rallis[1], Margarita Skamantzari[1]

[1]*Institute of Communication & Computer Systems (ICCS)*
[2]*School of Rural and Surveying Engineering, National Technical University of Athens*
[3]*British Geological Survey, Keyworth, United Kingdom*

Keywords: freshwater lens, SEAWAT, Machine Learning, dimensionality reduction

Abstract. On relatively small islands, freshwater often takes the form of lens-shaped water bodies surrounded by seawater. These freshwater lenses are susceptible to extended deterioration of water quality as a result of intense water pumping that exceeds sustainability limits. To ensure freshwater availability under the current anthropogenic and climate variability stresses, it is crucial to determine the individual components of the hydrogeological conditions and quantify their impact on the delicate balance of such a complex environment.

In this work, we investigate the effect of the aquifer heterogeneity with respect to hydraulic conductivity on the size and shape of the freshwater lens in a hypothetical aquifer resembling a typical Mediterranean island. A series of numerical simulations under yearly average recharge and pumping regimes are conducted using the SEAWAT numerical code to represent the variable-density flow processes in coastal aquifers. Often limited field data are available to properly characterise the hydrogeological features of subsurface flow and in particular hydraulic conductivity, which is an important variable that affects salinity plume extent in coastal aquifers. To that end, a python-based random field generation tool is employed to create realisations of hydraulic conductivity fields and compare simulations results with those obtained from a homogeneous aquifer representation. In this way, we explore the impact of

preferential flow paths to pumping wells due to spatial variability of high and low conductivity zones.

A major challenge in simulations using variable density and solute transport numerical codes is the resulting computational burden, which might hinder repetitive simulation tasks, such as Monte Carlo analysis. To mitigate this, we have employed machine learning (ML) techniques, in particular the Gaussian process and the support vector machine regressors, to efficiently predict salinity distribution in lieu of the physics-based and computationally expensive numerical model. However, the nonlinear nature of seawater intrusion simulations and the curse of dimensionality affect the predictions skills of machine learning methods as emulators of coastal aquifer models. In particular, special treatment is required to train machine learning models using multi-dimensional input/output obtained from the application of spatially distributed groundwater numerical models. To overcome the computational complexity associated with such a large number of input variables, we employed several dimensionality reduction techniques, such Principal Component Analysis, the Independent Component Analysis and the Factor Analysis, or other manifold learning related techniques.

An interesting interplay is observed between pumping well locations and pumping rates and hydraulic conductivity patterns. We explore the vulnerability of the freshwater volumes due to uncertain hydraulic conductivity and recharge reduction. Regarding the combination of machine learning and dimensionality reduction methodologies, preliminary numerical experiments provide multiple insights in terms of applicability, scalability and robustness, for the proposed pipeline.

Acknowledgment: This work is part of the DaWetRest-Danube Wetlands and flood plains Restoration through systemic, community engaged and sustainable innovative actions project. This project has been funded by the European Union through the Horizon 2020 research and innovation programmes under Grant Agreement No 101113015.

Assessment of established interpolation methods for the 3D spatial reconstruction of buried peat using legacy borehole data

Pablo De Weerdt[1], Jeroen Verhegge[1], Philippe De Smedt[1], Ellen Van De Vijver[1]

[1]*Department of Environment, Faculty of Bioscience Engineering, Ghent University*

Keywords: peat, subsurface modelling, uncertainty propagation

Abstract. The landscape-scale reconstruction of subsurface peat(y) layers in geological or pedological models attracts increasing attention in land development projects and related policy-making processes. Considering the importance of accurate volume and depth estimations, and their associated uncertainty, there is a clear incentive to further develop probabilistic modelling tools. In this study, the performance of three different methods to create a 3D subsurface model of buried peat is evaluated. A highly heterogeneous 1000-ha area in the embanked floodplain of the river Scheldt in Belgium constitutes the study area and includes 130 legacy boreholes, collected over more than 100 years, of which descriptions have been encoded into lithology fractions. We applied inverse distance weighting (IDW), as a deterministic interpolation method, and ordinary and indicator kriging (OK and IK), as probabilistic geostatistical methods, because of their established reputation in various applications. The optimization of interpolation parameters (including the directional power variation in IDW and variogram model parameters in OK and IK) was performed through cross validation. We assessed the resulting 3D models of peat by independent validation with 26 boreholes, which were selected based on conditional subsampling. In conclusion, all methods succeeded in the reconstruction of a buried peat layer using only limited legacy borehole data. Geostatistical methods predicted peat occurrence and derived volume estimations slightly more accurately. The obtained kriging variance aids in the localization of additional borehole samples to reduce uncertainty in the subsurface model. Besides, it facilitates uncertainty propagation in the derived volume estimation model, and it could also facilitate this in e.g. derived greenhouse gas emission models. Nevertheless, there remains potential for further methodological improvement. Future research will be directed towards comparing these established interpolation techniques with more advanced modelling methods such as sequential simulations and investigating the added

value of implementing other (secondary) legacy data sources such as cone penetration testing or surface geophysics.

Chemostratigraphy and Geostatistical Analysis of Geochemical and Lithological Data from the Trypali Unit, Omalos Region, Western Crete, Greece

Emmanouil Manoutsoglou[1], Ilias Lazos[1], Nikolaos Kallithrakas – Kontos[1], Christos Kanellopoulos[2], Emmanouil Steiakakis[1], Emmanouil Varouchakis[1]

[1]*School of Mineral Resources Engineering, Technical University of Crete, Greece*
[2]*Faculty of Geology and Geoenvironment, National and Kapodistrian University of Athens, Greece*

Keywords: Chemostratigraphy, Karstification, Trypali Unit, Western Crete

Abstract. The Alpine basement of Crete is composed of a number of fault bounded tectonic units with different lithologies, paleogeographical origin and metamorphic grade. Among others, a group of metamorphic rocks, composed mainly of phyllite and quartzite, has been described through time with variable names. This Group named Phyllite Nappe lies between Tripolitza nappe and Crete-Mani Unit/Plattenkalk Group or/and the Trypali Unit. The Trypali Unit, characterized by carbonate formations and significant karstification, forms a unique hydrogeological entity within the area. The term "Trypali Unit" primarily refers to carbonaceous rocks such as dolomites, dolomitic limestones, occasional pure limestones, carbonaceous breccia, rauhwacke, and white, glace-textured marbles, predominantly found in Western Crete. These rocks have experienced varying degrees of recrystallization over time. The stratigraphic evolution of this unit, coupled with the presence of spar fossils like algae, corals, and gastropods, suggests a shallow water depositional environment for Trypali unit. Furthermore, its specific lithology and tectonic positioning relative to the impermeable formations of the underlying Plattenkalk Group establish it as a distinct hydrogeological entity in the region, characterized by intensive karstification and hosting a productive aquifer. The water chemistry in this area is influenced by the composition of the rocks that form the aquifer system. The major aquifers in western Crete are hosted within the karstified rocks of the Trypali Unit for which there is no clear stratigraphic arrangement due to its intense tectonism.

Considering the lithological variety described above, a significant number of samples were collected to perform the geochemical and mineralogical analyses. The sampling was focused on the surrounding Omalos plateau (north of Samaria Gorge). In particular, 123 different samples were collected from the northern (FO, SM and OM samples), southern (NOM samples), eastern (AOM samples), western (DOM samples) and southwestern (PG samples) parts of Omalos plateau. All these samples were geochemically analyzed by implementing the XRF (X-Ray Fluorescence) method, while the mineralogical study was conducted by implementing the XRD (X-Ray Diffraction) method. These methods resulted to the measurement of the major, trace element content and the identification of major mineral phases, respectively.

The aim of this work is twofold: to classify the chemistry of metamorphic carbonates and to use the results of the statistical processing of rock chemistry in chemostratigraphic approaches of this Unit in the region of Omalos. To enhance the analytical approach, geostatistics and K-means clustering will be integrated into the study. Geostatistics, which involves the spatial analysis of geochemical and lithological data, and K-means clustering using a machine learning algorithm for partitioning datasets into distinct groups, will contribute valuable insights into the distribution and classification of rock compositions within the Trypali Unit in the Omalos region.

Acknowledgment: This research is financially supported by the Green Fund “Forest Protection and Upgrading 2019” under “Other Nationals. Green Fund”.

Co-simulation of Geological Facies and Ore Grades in the Molaoi Polymetallic Orebody, Greece: A Geostatistical Application Using Advanced Modeling Techniques

George Valakas[1], Konstantinos Modis[1]

[1]National Technical University of Athens

Keywords: Geostatistical Simulation, Ore Reserve Estimation, Co-simulation, Sustainable Exploitation

Abstract. This study focuses on the Molaoi area in Lakonia, Greece, aiming to estimate ore reserves using geostatistical simulation, taking into account the spatial distribution of lithofacies and metal grades. The project's relevance is heightened by the fact that over 15 strategic and critical raw materials listed in the European Union's catalog are found within Greece's Public Mining Areas, including zinc, silver, gallium, and germanium, elements crucial for the EU's energy transition industry. The Molaoi area, part of South-Eastern Peloponnese, contains significant polymetallic deposits, mainly sphalerite, with previously estimated proven reserves of 2.655 million tonnes and high concentrations of Zn and Pb. This study utilizes core data from over 60 boreholes, encompassing geological descriptions and metal grades. The geological formations of the area play a pivotal role in the spatial estimation of ore reserves, influencing the distribution and concentration of metals and guiding the efficient and accurate identification of ore reserves. For the purpose of this study, we co-simulate geological facies and metal concentrations for accurate grade estimation. Through Plurigaussian modeling, the geological facies are first transformed into Gaussian Random Fields. Then, a Linear Coregionalization Model is established to account for the dependencies between facies and the Normal scores of grades. This approach is particularly well-suited to addressing the complexities inherent in the Molaoi area's geology, where the spatial distribution and correlation of metal grades significantly vary across different geological domains. By simulating both categorical (lithological domain) and continuous (ore grades) variables, our goal is to enhance confidence in resource evaluation, ensuring accurate consideration of geological uncertainty. This enables a more nuanced and realistic joint distribution modeling. The proposed geostatistical

methodology aims to refine ore reserve estimations, which is essential for the sustainable exploitation of deposits, the reduction in licensing times, and attracting investments. The outcome of this research provides critical insights into resource estimation techniques, aligning with contemporary environmental, economic, and strategic objectives within the European Union and Greece, thus contributing to a more self-reliant and sustainable future in critical raw materials.

Combined Scenarios for Reducing Water Scarcity through a Water-Energy-Food-Ecosystems Nexus Approach

Ioannis Trichakis[1], Berny Bisselink[1]

[1]European Commission, Joint Research Centre

Keywords: water management, water scarcity, LISFLOOD, hydrological model

Abstract. Lack of water is a growing concern in many European Union (EU) Member States, with approximately 52 million people living in water-scarce regions. The majority of these people live in southern European countries, where water demand often exceeds supply during episodes of low water availability. Water saving measures can help to reduce water scarcity and ensure water resilience.

The LISFLOOD-EPIC coupled hydrological crop model has successfully simulated Europe's water resources for the period 1990-2018 using a meteorological reference scenario and scenarios for water-saving measures. This analysis includes five policy-relevant water-saving measures, namely improving irrigation efficiency, reducing urban water leakage, re-using treated wastewater, reducing energy water withdrawals, and desalination. Conclusions are drawn from two water scarcity indicators, the Water Exploitation Index Plus (WEI+), calculated as consumed water over total water availability, and the Water Exploitation Index (WEI) calculated as abstracted water over total water availability. The study's focus is on the relative differences in the indicator values between the reference scenario and scenarios with the proposed measures.

To combat water scarcity and secure water availability for all sectors that require water, EU Member States propose various measures under the Water Framework Directive. The study evaluated these measures in the context of the Business As Usual (BAU) and High Ambition Scenarios (HAS), and checked their expected effect in current climate. The BAU considered the planned investments on irrigation efficiency improvement and urban water savings, a small-scale reuse of wastewater, and the projection for water needs for energy production with slow energy transition. The HAS considered additional planned investments for irrigation efficiency, a goal of 5% losses for all urban water networks, higher

reuse of wastewater, a faster energy transition (with reduced water needs for cooling) and some use of desalination.

The results showed that irrigation efficiency and desalination were important measures in reducing water quantity pressures presented with the WEI+ and WEI water scarcity indicators. Other measures, such as water reuse and urban water efficiency improvements, had a positive effect on the WEI, but not for the WEI+. Scenarios with combined measures were the most effective in addressing water scarcity issues. However, the combined water-saving measures in the BAU scenario were not sufficient to compensate for the projected increase in water scarcity due to climate change, particularly in already existing water-scarce areas in southern Europe. For reducing water scarcity in these areas more ambitious measures, like in the HAS, are needed. The study's findings highlight the importance of combined water-saving measures in addressing water scarcity issues. While individual measures such as improving irrigation efficiency or reducing urban water leakage are important, their combined impact is more significant.

This work has shown that there is plenty of room to further enhance the modelling approach by using geostatistical methods for the spatial distribution of many parameters.

Combining Global Climate Projections: A Stochastic Graph Cut Approach with Multivariate Distributions

Lucas Schmutz[1], Soulivanh Thao[2], Mathieu Vrac[2], Denis Allard[3], Gregoire Mariethoz[1]

[1]*Institute of Earth Surface Dynamics (IDYST), University of Lausanne*
[2]*Laboratoire des Sciences du Climat et de l'Environnement (LSCE-IPSL), Université Paris-Saclay*
[3]*Biostatistics and Spatial Processes (BioSP), INRAE*

Keywords: Multi-Model Ensembles, CMIP6, Climate Models

Abstract. Numerous methods have been proposed to combine individual climate models and extract a robust signal from an ensemble, including the Multi-Model Mean (MMM) and the weighted MMM. These methods typically apply a global weighting to the models, overlooking the fact that certain models may excel in specific regions. This observation suggests that a method acknowledging the superior regional performance of individual models could significantly enhance climate projections derived from model ensembles.

To date, the Graph Cut optimization method (Thao et al., 2022) stands out as the sole technique effectively leveraging the local capabilities of different models across multi-decadal periods to produce global projections. This method involves selecting the most suitable model for each grid point while also ensuring the spatial consistency of the resulting fields. Despite its promising results, surpassing other ensemble combination techniques, it is restricted to optimizing for a single variable, thereby causing inconsistent model selection across variables in multivariate scenarios. This leads to a loss of the multivariate relationships encoded in models. Furthermore, this technique was limited to multi-decadal averages, and is thus unable to capture the distributional characteristics of climate variables, including extreme events.

Here, we introduce significant enhancements to the Graph Cut optimization method, allowing for the combination of distributions of daily means while preserving multivariate relationships, thus better capturing the complete span of climate dynamics. By employing the Hellinger distance to assess model performance, we can identify, at each grid point, the model that most accurately

represents the joint distribution of target variables (e.g., temperature, pressure, and precipitation), minimizing the emergence of unrealistic discontinuities in the combined fields. Although every Graph Cut optimization produces a deterministic map where each grid point is assigned to one model, there exists a large number of solutions that have equivalent performance. By initializing the optimizations stochastically, we generate a range of analogous model combinations. For each grid point, we then proportionally combine the distributions of models based on their frequency of appearance, thereby weighting the models' contributions in accordance with their relative performance at both local and global scales. Compared to the original method, the resulting projections offer a more realistic representation of extremes and compound events.

Comparison of two causal inference methods and relation between NAO and rainfall

Dionissios T. Hristopulos[1], Petros Symeonidis[1], Sofia Tsakiridou[1], Emmanouil Varouchakis[1]

[1]*Technical University of Crete*

Keywords: causal inference, rainfall, Wiener-Granger causality, Liang information flow rate

Abstract. Methods of causal inference have applications in various scientific fields, including the earth and climate sciences, neuroscience, and finance. Data-driven causal inference methods aim to identify relations of cause and effect between potential drivers and responses based on the available data, which are typically in the form of time series. In contrast with correlation analysis, the results of causal inference can be asymmetric, i.e., it is possible that series A influences series B but not vice versa. This should be juxtaposed with the outcome of correlation analysis: if A is correlated with B, then B is also correlated with A. Hence, causal inference allows determining drivers and responses based on statistical analysis of the data. We compare the method of Liang information flow rate (LIFR) with the classical method of Wiener-Granger causality (WGC). LIFR is a novel method which is formulated on the basis of entropy exchange between the components of an interacting system. WGC is a standard tool of causal analysis which models a system of time series by means of vector autoregressive models. These methods have quite different origins but have not yet been compared in the scientific literature. Our analysis investigates causal links in synthetic data derived from a simulated first-order stochastic differential equation system using both methods (LIFR and WGC). We also seek for potential causal links between the North Atlantic Oscillation (NAO) index and the amount of monthly rainfall in Greece. Towards this goal, we use reanalysis data of rainfall amount from two different geographic areas of Greece as well as rainfall measurements from the island of Crete (Greece). Finally, we also report on the statistical significance of the causal measures based on the non-parametric method of permutation testing.

Compositional signature of trace elements in stream sediments along the Babahoyo/Daule riverbeds (Ecuador): A geostatistical and Bayesian approach

Marlon Cabrera[1], María Pazo[2], Saki Gerassis[2], Teresa Albuquerque[3], Carlos Sierra[4], Carlos Boente[1]

[1]E.T.S.I. Minas y Energía de Madrid, Universidad Politécnica de Madrid, C/ Ríos Rosas 21, Madrid, 28003, Spain
[2]CINTECX, Universidade de Vigo, Grupo de Xestión Segura e Sostible de Recursos Minerais, Dpto. De Enxeñaría dos Recursos Naturais e Medio Ambiente, 36310, Vigo, Spain
[3]Research Centre for Natural Resources, Environment and Society (CERNAS) - Instituto Politécnico de Castelo Branco, Polytechnic University Castelo Branco, Portugal.
[4]European University on Responsible Consumption and Production, Campus de Vegazana, 24006, León, Spain

Keywords: Surface water, geochemistry, Compositional Data, Bayesian network

Abstract. The high concentrations of Potentially Toxic Elements (PTEs) in stream sediments may pose a significant threat since the pollutants may travel long distances to the sea, affecting aquatic environments and public health. When rivers cross large cities, mining, or cultivated areas, it is usually observed an increasing concentration of PTEs downstream, reaching hazardous levels due to the ability and persistence of heavy metals by bioaccumulation in the environment.

This research comprises a comprehensive sampling campaign composed of 75 samples of stream sediments from the Babahoyo and Daule riverbeds (Ecuador, South America), downstream and upstream of Guayaquil. 63 chemical major and trace elements were analyzed in each sample. The research aims to define a preliminary expert-based geochemical signature for the area to ascertain the degree of anthropization in two riverbeds exposed to intense industry and agriculture.

In the first step, the selected attributes underwent a compositional transform (CoDa) by means of a clr-transformation of geochemical data, followed by an unsupervised Bayesian approach. From the exploratory analysis of the unsupervised network, the potential relationships between geochemical data can be explored in reality, transferring them to the model. This great potential allows us to carry out a global analysis of the problem, detect which PTEs have the greatest influence and identify data redundancy. The definition of spatial hot spots of pollution downstream of the Babahoyo and Daule rivers was computed through Sequential Gaussian Simulation and further interpreted on the PTEs' geochemical fate framework, facing the lithological and land use patterns. The preliminary results point to a division between those trace elements that are linked to geogenic enrichment (Co, Cr, V), the ones related to anthropogenic nature, such as the related to industrial or agricultural origin (Ba, Cd, Mo), and the ones linked to both processes (As, Cu, Mn).

Contaminant plume reconstruction through joint geophysical and concentration data assimilation

Camilla Fagandini[1], Valeria Todaro[1], Leonardo Azevedo[2], J. Jaime Gómez-Hernández[3], Andrea Zanini[1]

[1]*University of Parma*
[2]*Instituto Superior Técnico, Universidade de Lisboa*
[3]*Universitat Politècnica de València*

Abstract. Pollutants represent a significant environmental threat, impacting soil and groundwater quality and posing risks to human health. The inherent complexity of subsurface processes raises challenges to the precise location and definition of the extent of contaminant plumes. Hydrogeophysics emerges as a powerful tool by enabling the simultaneous combination of non-invasive geophysical techniques (i.e., electrical resistivity, seismic survey…) and hydrological variables. This field contributes to a comprehensive approach to understanding and managing contaminant plumes, ensuring effective responses to environmental problems.

While previous research has demonstrated the ability of hydrogeophysics to estimate aquifer characteristics, only a few studies have focused on its use to predict groundwater contamination. This work aims to accurately estimate contaminant spread and distribution within the subsurface system. To this aim, a novel approach that combines Ensemble Smoother with Multiple Data Assimilation (ES-MDA) with Convolutional Neural Network (CNN) is introduced. ES-MDA is applied to solve the inverse problem by assimilating Electrical Resistivity Tomography (ERT) data and sparse concentrations observed at monitoring wells, while a CNN replaces the forward model to reduce the computational effort. ES-MDA integrates observational data into numerical models to increase model accuracy and reduce uncertainties, whereas CNNs use spatial data to detect and characterize contamination plumes, particularly in imagery datasets.

A 2D synthetic case study simulating a tracer test in a fully saturated unconfined aquifer is conducted to test the methodology. The study examined five different datasets to assess the performance of the proposed approach, allowing for a

thorough examination of the advantages of combining data from multiple sources, as well as the effects of various observation datasets on the accuracy of plume distribution assessments. The first scenario interpolated 15 concentration values using a kriging-based approach, while subsequent scenarios tested the suggested inverse hydrogeophysical approach's capabilities. The second scenario used only apparent resistivity data as observations into the ES-MDA; while the third to fifth scenarios combined apparent resistivity data with different subsets of concentration values: 15, 9, and 3, respectively. The third scenario, which combines apparent resistivity with 15 concentration values, proved to be the most accurate and precise setup. The least accurate estimations were obtained using kriging interpolation (Scenario 1) and ES-MDA with only apparent resistivity data (Scenario 2). These findings suggest that a limited dataset may not provide sufficient information to capture the spatial variability of subsurface concentration maps accurately. Furthermore, they emphasize the importance of a more comprehensive approach and the need to combine multiple data sources when precise mapping of tracer plumes is desired. The results underscore the effectiveness of the proposed integrated approach in accurately estimating the spatial distribution of a concentration plume, offering a valuable and cost-effective tool for supporting optimal strategies in contaminated site remediation and contributing to the sustainable management of groundwater resources.

Copula-based simulation of random fields with directional asymmetry

Sebastian Hoerning[1], Andras Bardossy[2]

[1]*Centre for Natural Gas, The University of Queensland*
[2]*Institute for Modelling Hydraulic and Environmental Systems, University of Stuttgart*

Keywords: Copula, Asymmetry

Abstract. Certain physical processes, such as advection-diffusion, can lead to directional spatial asymmetry. The general concept of directional asymmetry is well-known in time series analysis where a time series such as a stream flow record is assumed to be irreversible (or also called directional) if its properties depend on the direction of time. The spatial extension of reversibility is directional spatial asymmetry. Measures, such as spatial copulas or the directional asymmetry function can be used to detect and analyse directional asymmetry, however, the simulation of random fields that exhibit directional asymmetry is not trivial, especially if these random fields require conditioning. In this contribution we will introduce a novel, copula-based approach to simulate conditional random fields with directional asymmetry. The approach uses a combination of FFT-MA, spatial copulas, Random Mixing, and geometric transformations for the simulation, and it offers a great flexibility in terms of its parameterization. The methodology allows a very fast conditional simulation of asymmetrical non-Gaussian random fields both in 2 and 3 dimensions. A groundwater contaminant transport example will be used to demonstrate the approach and to highlight its benefits.

Deep Learning for Downscaling Climate Model Data - A Comparative Analysis with Traditional Approaches

Valeria Todaro[1], Daniele Secci[2], Marco D'Oria[1], Maria Giovanna Tanda[1]

[1]*University of Parma*
[2]*University of Parma, Universitat Politècnica de València*

Keywords: Downscaling, CNN, Climate Models

Abstract. To address the challenges posed by climate change, the use of climate models has become common practice. Although numerous climate models from various institutes are readily accessible, challenges persist in achieving the optimal spatial resolution required for impact assessment studies at subregional or local scales. To enhance the spatial resolution of climate models, conventional methods typically employ computationally expensive dynamical downscaling procedures, while alternative approaches utilize statistical downscaling techniques. This study presents a deep learning based approach to perform spatial downscaling of climate models, specifically employing a Convolutional Neural Network (CNN). The model is trained to learn complex relationships between large-scale climate features and local characteristics, focusing on climate variables like precipitation and temperature. The proposed methodology can be employed whenever there is a need to enhance the spatial resolution of climate variables. The aim of this work is to improve the accuracy of climate outputs from CMIP6 initiative. For these models, there is a heightened urgency for downscaling, as only General Circulation Models (GCMs) are available, and they lack the necessary resolution to perform regional to local analyses. The study focuses on the Emilia-Romagna region in northern Italy, with the proposed methodology designed to be easily adaptable to other regions. A comparative analysis is conducted against alternative methods, such as a second order conservative interpolation and nearest-neighbour interpolation, to assess their effectiveness and determine the most suitable approach for the given context. Choosing between these approaches requires careful consideration of specific project requirements and available resources.

Acknowledgment: This work was carried out within the PNRR MUR project ECS_00000033_ECOSISTER and supported by OurMED PRIMA Program project funded by the European Union's Horizon 2020 research and innovation under grant agreement No. 2222.

Deep neural networks in surrogate hydrogeological modeling: An application for transient groundwater flow combined with a geostatistical spectral algorithm for inverse problem-solving

Dany Lauzon[1]

[1]*Polytechnique Montréal*

Keywords: Stochastic Hydrogeology, Inverse Problem, Surrogate Modeling, U-Net Architecture

Abstract. Characterizing groundwater flow parameters is crucial for understanding complex aquifer systems. Inverse techniques are key for modeling hydrogeological parameters and assessing uncertainties. However, using a flow simulator can be time-consuming, especially with many model parameters. To address this, surrogate models are proposed, increasingly leveraging deep learning. However, their training relies on a large database of models, often lacking diversity and requiring significant time.

A recent proposal suggests replacing the transient groundwater flow model with a U-Net encoder-decoder architecture. This reduces execution time and enables uncertainty quantification with geostatistical methods. The substitute is trained using limited forward model evaluations to understand the physical relationship between hydraulic conductivity fields and transient hydraulic heads measured on-site. Physical principles, like boundary conditions and source terms, are mapped as inputs to enhance the model's understanding of transient groundwater flow equations.

We explore the possibility of generating drawdowns at any given time by training a U-Net architecture on a subset of the spatiotemporal drawdown series. We propose a methodology to reduce training times while maintaining good emulation quality. Mapping boundary conditions and source terms introduce the physical knowledge of the problem. The novelty pertains to the introduction of an estimation map to mimic the pumping area. Once the model is trained, we use a spectral geostatistical method to solve the inverse problem

using the surrogate model to estimate uncertainties associated with hydraulic conductivity and boundary conditions.

Our study demonstrates that the U-Net accurately reproduces the drawdown inside the training range, and in terms of computational demand, using U-Net as a substitution model reduces the required calculation time by about an order of magnitude for the defined field. The proposed approach provides an efficient and accurate method for characterizing groundwater flow parameters. The quantification of uncertainties in complex aquifer systems is thus determined more rapidly.

Acknowledgment: We thank the Natural Sciences and Engineering Research Council of Canada for their contribution to the funding of this research project, as well as Polytechnique Montréal for their involvement and financial support.

Deep neural operators for surrogate modeling of multi-physics environmental flow

Sourav Dutta[1], Peter Rivera-Casillas[2], Shukai Cai[1], Mark Loveland[2], Jonghyun Lee[3], Matthew Farthing[2], Clint Dawson[1]

[1]*The University of Texas at Austin*
[2]*UISACE Engineer Research Development Center*
[3]*University of Hawaii at Manoa*

Keywords: model order reduction, physics-informed machine learning, shallow water equations

Abstract. In coastal hydrology, appropriate characterization of spatially varying fields such as Manning's roughness coefficient and bathymetry is crucial for building an accurate overland flow and storm surge model, which can, in turn, help devise effective strategies for mitigating flood damage. However, estimating spatially distributed and temporally varying properties from sparse and potentially noisy measurements is a notoriously challenging, ill-posed inverse problem. A key challenge in this process is the computational expense of high-fidelity numerical models commonly employed to simulate coupled physical processes in coastal and near-shore regions. The need to accurately capture complex, nonlinear interactions between wave hydrodynamics, meteorological, and hydrological processes using high-resolution, unstructured meshes renders most of the existing high-fidelity numerical models intractable for parameter estimation, ensemble-based forecasting, or hazard assessment.

In this talk, we will present a deep neural operator framework, inspired by the deep operator network (DeepONet) philosophy, that can efficiently emulate the underlying solution operators of complex systems of partial differential equations and generate fast and accurate forecasts for coupled, multi-physics flows. We will explore different neural operator model formulations and first demonstrate their merits and drawbacks using benchmark analytical problems. In addition, a reduced-order emulator of a real-world example of 2D tide-driven flow in Shinnecock Inlet Bay, New York will also be presented. Such flows are typically simulated by a coupled system of the depth-averaged shallow water equations and the wave action balance equations, parameterized by various combinations of meteorological and morphological conditions. Moreover, a typical design and development pipeline for a model problem will be presented

to highlight how such a framework might fit as a general approach for reduced-order modeling in multi-physics engineering applications.

Development of an Operational Approach for the Exploitation of In Situ Radiological Measurements through Non-stationary Geostatistics: Example in the Fukushima area.

Meryem Meziane[1], Léa Pannecoucke[2], Xavier Freulon[3], Chantal De Fouquet[3]

[1]*Institut de Radioprotection et de Sûreté Nucléaire (IRSN), PSE-ENV/SPDR/LT2S ; MINES Paris PSL, Centre de géosciences*
[2]*Institut de Radioprotection et de Sûreté Nucléaire (IRSN), PSE-ENV/SPDR/USDR*
[3]*MINES Paris PSL, Centre de géosciences*

Keywords: in situ gamma spectrometry, spatial non-stationarity, geostatistics, Fukushima

Abstract. Radiological characterisation is a key step in managing the consequences of nuclear activities, be it in post-accidental situations or in the perspective of rehabilitating contaminated sites. These situations may require responses such as evacuating specific areas based on contamination thresholds or implementing decontamination actions. To obtain contamination maps that serve as decision supports, in situ gamma spectrometry is a non-destructive survey method allowing the acquisition of data at high spatial frequency. Geostatistics then enable the exploitation of the acquired data for “optimal” estimation. The methods usually implemented assume some stationarity properties of the quantity to estimate: for instance, the spatial structure of the phenomenon is considered shift invariant. Stationarity hypotheses are often questioned in practice: the contamination plumes present very high concentration contrasts, and the variations of geographical characteristics and acquisition conditions may cause the spatial structure to vary locally. Taking this non-stationarity into account may yield more realistic estimations and a reduction in uncertainties. From an operational standpoint, it may allow for the adaptation of the sampling scheme according to location, opening the perspective of time saving in post-accidental situations. It may also help orient the acquisition of data through more expensive and time-consuming methods such as soil sampling. Therefore, the aim of this work is to explore the potential of non-stationary geostatistics in improving radiological characterisation.

The study is mainly based on airborne data acquired following the Fukushima Daiichi Nuclear Power Plant accident in March 2011. The accident led to the dispersion in the environment of a massive amount of radionuclides, including Cesium-137, a particularly preoccupying isotope due to its human and environmental toxicity and to its relatively long half-life (30 years). The data consists of an airborne survey of Cesium-137 deposit acquired in October/November 2011. The measurements cover a half-disc of 80 km radius around the power plant and are organized in flight lines of two perpendicular directions, with a typical spacing of 1.8 km between flight lines and 40 m between two consecutive observations along a flight line. This data is completed by land use and topography information as covariates, in addition to a set of Cesium-137 deposit simulations, which do not rely on the measurements mentioned above.

Correlations and links between the deposit and the covariates are explored and variography is undertaken on the whole and some resampled data sets (due to the heterogeneity of the sampling scheme). Then, non-stationary estimation methods are implemented and compared with ordinary kriging through cross-validation procedures: external drift kriging using different covariates, (co-)kriging with numerical variograms and an SPDE approach. Some of these methods have been developed recently or are not commonly used. Implementing them on a dataset of such density is ideal to put them to test and demonstrate their interest. The geographical characteristics of the territory under study are taken into consideration at local scale through the covariates, as well as data of different nature such as soil samples or carborne measurements. Operational recommendations regarding the sampling scheme are derived from the analysis of the results.

Acknowledgment: This project was funded by the French government as part of the « France Relance » plan.

DischargeKeeper (DK) an Emerging Tool for Non-Contact Volumetric Flow Monitoring: A Case Study in a Semi-Arid Wadi

Anam Amin[1], Mohammed Magdy[1], Ibrahim Elsaify[1], Mohammed Gad[1]

[1]*TDS (Technical Development Solutions Co.), Riyadh, Saudi Arabia*

Keywords: DischargeKeeper Camera, Semi-Arid Environments, Wadis, Volumetric Flow Monitoring

Abstract. The inherent challenge associated with monitoring volumetric flow in semi-arid and arid regions is due to their inhospitable and ever-changing environmental features, characterized by fluctuating temperature reaching extremes and continuous occurrence of intense sandstorms. Within these geographical areas, hydrological phenomena sometimes occur as heavy rainfall and sudden flash floods with a time span of years between occurrences of significant events.

Ephemeral drainage courses are widespread in these regions and mainly act as the primary conduits for water drainage in arid landscapes. These seasonal water channels, known as Wadis, play a fundamental role in managing water flow through these zones. Wadis typically remain dry except during rainfall periods, triggering quick and intense flash floods, where the peak values occur within the first few minutes of the rainfall.

It is crucial to capture flood peaks for flood control and early warning systems therefore to utilize robust devices capable of continuous recording at short time intervals (i.e., minutes or even less) to be able to capture the steep rise of the flood peak.

Between 31st December 2021 and 4th January 2022, heavy rainfall triggered a significant event in Wadi Naqab located in the northern region of United Arab Emirates, resulting in surface runoff for 3.5 hours captured by an advanced image-based DischargeKeeper (DK) system. The Wadi is approx. 50 meters wide with an elevation range 165-328 m, geology characterized by dolomitic limestones of the Ghail formation and has been dry for most of its existence.

DK system's validation is carried out by correlating the camera-based flow readings with those obtained from an Acoustic Doppler Current Profiler (ADCP)

as a standard. The deviation between DK and ADCP readings was under 5% in 40% of the conducted measurements (total: 11 comparisons). Deviations of less than 10% and 15% were achieved in 60% and 80% of the measurements, respectively. Discrepancies mainly ranged from 0.01 m^3/s to 0.1 m^3/s.

Calibration of DK requires specific input parameters: river/channel cross-section, four detectable reference points far-off the shore within the camera's view, and initial water level equivalent to the current waterline position. These parameters are necessary only during the establishment of a new monitoring site initially.

Methods encompass monitoring and logging of flow velocity, water level, water flow and data analysis. This paper presents the results of a significant event that occurred on 3rd January 2022 at 15.05, recorded a discharge peak (QPEAK) equal to 78 m3/s within 15 minutes after the initiation of water flow. The hydrograph showed the recession of the event started at 15.20 (Q = 51 m3/s).

This is the first pilot project for measuring the discharge by optical camera in the middle east. Our results support the use of DK technology to overcome the limitations of traditional monitoring methods, indicating a new era of precision and efficiency in water monitoring and management for the region.

Droughts and Floods: Statistical Characterization of Extreme Rainfalls

Rachid Ababou[1], Nabil Elmocayd[2]

[1]IMFT - Institut de Mecanique des Fluides de Toulouse, France
[2]UM6P - University Mohammed 6 Polytechnic, Ben Guerir, Morocco

Keywords: Hydro-meteorological processes, Statistical analyzis and synthezis, Random processes, Extreme rainfall events

Abstract. Droughts and floods remain the most recurrent climate-related natural hazards, inflicting severe consequences on individuals worldwide and impeding their ability to sustain daily activities. In many vulnerable regions, the accessibility of data for analyzing, understanding, predicting, and projecting the impact of climate change on such phenomena is frequently limited or entirely absent. Statistical approaches can be employed for regionalizing rainfalls obtained from rain gauge stations. The widespread availability of open gridded data globally has facilitated addressing the challenges and predictive capabilities concerning rainfall signals, even in regions where traditional measurement methods based on rain gauges may be inexistent, unfeasible or impractical. The value of such regionalized or gridded rainfall datasets to evaluate extreme precipitation has been investigated for instance in Tuel, ElMocayd et al. 2023. On the other hand, in some regions, long records of daily rainfalls are available locally, at some rain gauges. The rainfall signals at these rain gauges can be analyzed with appropriate statistical methods to account for high or low rainfall events of interest over decades. Overall, with the expanded access to data, new opportunities for developing a variety of higher-order analyses have emerged, and the demand for sophisticated, data-driven and statistical algorithms is now increasing in order to analyze and mitigate the negative impact of flood and drought events in major vulnerable areas.

This paper focuses on statistical and probabilistic analyses of hydro-meteorological events that can be qualified as "droughts" and "floods", that is, low rainfall periods ("droughts") and high rainfall periods ("floods"), using two types of rainfall data signals. These low/high periods, or events, are analyzed either directly from daily precipitation time series P(t), or else indirectly from normalized quantities like the monthly Standardized Precipitation Index SPI(t) obtained from a transformation of gridded rainfall data. In this study, all such

hydro-meteorological signals are treated as stationary random processes, that is, random functions of time that are statistically invariant under time translations. The methods used here include estimation of probability distributions, analyses of autocorrelated random processes, their level set crossings, excursions and extrema (based in part on Rice theory), and algorithms for synthetic generation of rainfall processes or rainfall events of interest.

Effective probability distribution approximation for non stationary non Gaussian random fields

Anastassia Baxevani[1], Dionissios Hristopulos[2], Christos Andreou[1]

[1]University of Cyprus
[2]Technical University of Crete

Keywords: Random field models, Effective probability distribution, non stationary, precipitation

Abstract. Most environmental data, like precipitation wind etc, may be modelled in terms of stochastic fields. These fields need to posses not only complex spatial and temporal dependence structure, but also to be non-stationary. Some of the non-stationarities may be due to the dynamic nature of the phenomena, and some other due to the fact that phenomena may posses different properties depending on the location and the time of the year. Moreover, while environmental data often exhibit significant deviations from Gaussian behavior, only a few non-Gaussian joint probability density functions admit explicit expressions. In addition, random field models are computationally costly for big datasets. We propose an effective distribution approach, which is based on the product of univariate conditional probability density functions modified by local interactions. The effective densities involve

local parameters that are estimated either by means of kernel regression or using origins equations. The prediction of missing data is based on the median value from an ensemble of simulated states generated from the effective distribution model. The latter can capture non-Gaussian dependence and is applicable to large spatial datasets, since it does not require the storage and inversion of large covariance matrices. We finish with some application to precipitation data.

Efficient Inference for Non-Deterministic Fractures

Tim Bruennette[1], Wolfgang Nowak[1]

[1]*Department of Stochastic Simulation and Safety Research for Hydrosystems (LS3), Institute for Modelling Hydraulic and Environmental Systems (IWS), University of Stuttgart*

Abstract. Real-world fracture patterns often show irregular geometries. The underlying causes for this are multi-scalic and not explicitly tractable. Nonetheless, these effectively random geometric aspects are essential for the characterization of mechanical and hydromechanical properties of fractures, relevant in fields such as nuclear waste disposal, enhanced geothermal systems and geologic carbon storage.

Parametric inference allows insights into the fracture creating conditions and correspondingly a sensible resampling of additional fractures. We give a brief overview on modeling approaches to incorporate this non-deterministic nature and discuss how to efficiently perform inference tasks on them. First, we argue why the problem setting is inherently hard for traditional methods, then we consider two distinct remedial approaches:

(a) we explore the option of data-driven surrogates. While these surrogates are trained on sets of single fractures, they are employed to predict relevant statistical properties of (hydro-)mechanical quantities of interest. Cutting out the need to predict specific geometries makes the surrogates fast to evaluate and therefore recovers the possibility of employing computationally intensive inference methods. Notably, in contrast to existing methods, the surrogates are not employed for fracture networks nor to predict full geometries of single fractures.

However, as these fast surrogates are purely data-driven, this comes at the cost of sacrificing first principles in the model formulation. Therefore, as an alternative (b), we propose an implementation of a randomized fracture generator based on a phase-field model for brittle fractures with a directional stress split. The variational formulation is leveraged to randomly sample split directions according to the energy state at every time step. This sampling during

runtime enables a forced tracing of specific geometries, thus avoiding costly Monte Carlo computations for the likelihood.

Both approaches (a) and (b) are effectively means to the same end, bringing inverse problems for hard-to-predict fracture geometries into computational reach.

Acknowledgment: German Research Foundation (DFG), Collaborative Research Centre SFB 1313, Project Number 327154368

Enhancing Geostatistical Analysis of Natural Resources Data with Complex Spatial Formations through Non-Euclidean Distances

Maria Koltsidopoulou[1], Andreas Pavlides[1], Emmanouil A Varouchakis[1]

[1]*Technical University of Crete*

Keywords: geostatistics, covariance functions, non-Euclidean distances

Abstract. Natural resources are critical in global and regional economies and are required for most of our society's consumer goods. Accurately representing and assessing natural resources is crucial for sustainable management and economic growth. The traditional kriging family geostatistical methods often give adequate reserves and uncertainty estimations. However, in some cases, the investigated deposit is spatially distributed in complex formations, posing significant challenges for subsequent analysis using the traditional approach. Traditional Euclidean distance measures often fail to capture the true spatial relationships in complex formations.

The utilization of Ordinary Kriging with advanced covariance functions able to accommodate non-Euclidean distances emerges as a promising approach to address these issues. This work attempts to couple Ordinary Kriging with non-Euclidean distance-based covariance functions to improve the modelling and analysis of such spatially complex data of natural resources and test different distances along with the Euclidean distance. By extending the application of Ordinary Kriging to non-Euclidean distances, our approach attempts better to assess the intricate spatial characteristics of the resources.

Our study involves implementing an advanced covariance function capable of forming permissible matrices under non-Euclidean distance metrics. This function is then tested with various non-euclidean distances to model the spatial distribution of natural resources data. Ordinary Kriging and the suggested covariance function are used to estimate the content of natural resource deposits. Leave-one-out cross-validation results for the various distances are compared and evaluated. The paper discusses the mixed results obtained from the case study. These findings highlight the potential benefits and challenges of integrating non-Euclidean distances in geostatistical analysis. Our insights into the effectiveness and challenges of this approach contribute to the future

development of more accurate geostatistical estimations to facilitate more efficient resource management strategies.

The research project is implemented in the framework of H.F.R.I call “Basic research Financing (Horizontal support of all Sciences)” under the National Recovery and Resilience Plan “Greece 2.0” funded by the European Union – NextGenerationEU (H.F.R.I. Project Number: 16537).

Acknowledgment: The research project is implemented in the framework of H.F.R.I call “Basic research Financing (Horizontal support of all Sciences)” under the National Recovery and Resilience Plan “Greece 2.0” funded by the European Union – NextGenerationEU (H.F.R.I. Project Number: 16537).

Ensemble Surrogate Modeling of Advective-Dispersive Transport with Intraparticle Pore Diffusion Model for Column Leaching Test

Amirhossein Ershadi[1], Michael Finkel[1], Binlong Liu[1], Olaf Cirpka[1], Peter Grathwohl[1]

[1]*Department of Geosciences, University of Tübingen*

Keywords: Simulation-Based Inference, Surrogate Model, Sorption Kinetics, Model Calibration

Abstract. Column-leaching tests are a common approach for evaluating the leaching behavior of contaminated soils and waste materials, which are often reused for various construction purposes. The observed breakthrough curves of the contaminants are influenced by the complex dynamics of solute transport and kinetic inter-phase mass transfer. Disentangling these interactions necessitates numerical models. However, inverse modeling and sensitivity analysis can be time-consuming, especially when sorption/desorption kinetics are explicitly described by intraparticle diffusion, which requires discretizing the domain both along the column axis and inside the grains. To circumvent the need for such computationally intensive models, we have developed a machine-learning-based ensemble surrogate model that employs two distinct ensemble methods: random-forest stacking and inverse-distance weighted interpolation of base surrogate models covering different parts of the parameter space. The base surrogate models use the Extremely randomized Trees (ExtraTrees) method. The defined parameter range is based on the German standard for column-leaching tests. To optimize the base surrogate models, we utilized adaptive sampling methods based on three distinct infill criteria: maximizing the expected improvement, staying within a certain Mahalanobis distance to the best estimate (both for exploitation), and maximizing the standard deviation (for exploration). The ensemble surrogate model demonstrates excellent performance in emulating the behavior of the original numerical model, with a relative root mean squared error of 0.08. We applied our proposed surrogate model to estimate the complete posterior parameter distribution using Simulation-Based Inference, specifically Neural Posterior Estimation (NPE), to determine the full parameter distribution conditioned on copper leaching data

from two different soils. Samples drawn from the posterior distribution align perfectly with the observed data for both the surrogate and original models.

Estimating Contaminated Soil Volumes Using a Generative Neural Network: A Hydrocarbon Case in France

Herbert Rakotonirina[1], Paul Honeine[2], Olivier Atteia[3], Antonin VAN EXEM[4]

[1]*Univ Rouen Normandie, INSA Rouen Normandie, Université Le Havre Normandie, Normandie Univ, LITIS UR 4108 . Tellux company*
[2]*Univ Rouen Normandie, INSA Rouen Normandie, Université Le Havre Normandie, Normandie Univ, LITIS UR 4108*
[3]*Univ Bordeaux, CNRS, Bordeaux INP, EPOC, UMR 5805*
[4]*Tellux Company*

Keywords: Polluted Soil Volume Estimation, Deep Learning, Hydrocarbon Soil Contamination, Simulation

Abstract. The estimation of the volumes of contaminated soil to be treated is a crucial step in soil remediation. Numerous techniques exist for estimating the distribution of pollutants in soils, such as inverse distance weighting, kriging, Gaussian sequential simulation, and sequential indicator simulation. Unfortunately, these methods require significant computational resources to achieve precise estimations. Moreover, both kriging and Gaussian simulation require the transformation of non-normal distributions, often seen in hydrocarbon contamination, to produce accurate results. In response, we propose a generative neural network to generate 3D maps of contaminant distributions without prior training, This differentiates it from other Deep Learning approaches that necessitate training data. The approach relies on a convolutional neural network for image reconstruction method. Rather than solely depending on the concentration of chemicals determined in the laboratory, we utilize hyperspectral imaging data from soil cores to achieve a more precise depiction of soil contaminants. We have assessed this approach using a real case of hydrocarbon pollution on a polluted site in France. The method has shown competitive performance with controlled computation times thanks to the utilization of a GPU accelerator. Our study offers a new, practical way to improve soil pollution management using fast, and data-based techniques.

Acknowledgment: The authors would like to thank the agency of ecological transition ADEME in France and TELLUX company for the funding of this research work.

Evaluating nitrate pollution dynamics and identifying potential causes in karst and alluvial aquifers of the Apulian Region (Southern Italy)

Ayesha Nawaz[1], Maria Rosaria Alfio[1], Gaetano Daniele Fiorese[2], Andreas Panagopoulos[3], Gabriella Balacco[1]

[1]*Department of Civil, Environmental, Land, Construction and Chemistry, Polytechnic University of Bari*
[2]*Department of Agricultural and Environmental Sciences, University of Bari Aldo Moro*
[3]*Soil and Water Resources Institute, Hellenic Agricultural Organization-Demeter*

Keywords: Nitrate pollution, Groundwater, Karst and alluvial aquifers, Kriging

Abstract. Groundwater constitutes the most extensive distributed freshwater reservoir globally, serving a crucial role in sustaining ecosystems and facilitating human adaptation to climate variability and change. Nitrate pollution stands as the predominant form of groundwater contamination worldwide, presenting a particularly sensitive issue, especially in heterogeneous and anisotropic aquifers.

The aim of this research is to examine the spatial and temporal evolution of nitrate pollution in the groundwater circulating in the main aquifers of the Puglia region, using geostatistical methods.

The Apulian region (Southern Italy), with a coastline stretching approximately 785 km, exhibits marked variability in geological, morphostructural, and environmental characteristics, resulting in different hydrogeological conditions. From a hydrogeological perspective, it is possible to distinguish five main areas: Gargano, Murge, Salento, Tavoliere, and the Daunian Apennine sector. For the purposes of this analysis, the Daunian Apennines and the Gargano aquifers have been omitted due to unavailability of data. Murgia and Salento aquifers are karst systems while the Tavoliere consists of a succession of marine and alluvial deposits lying on clays grey-blue. The geological, structural, and morphological features of the Apulia region avoid the formation of superficial water resources, except for the Tavoliere. Therefore, water demand for drinking and irrigation purposes relies on groundwater bodies, primarily contained within carbonate successions. Nitrates represent the most important form of pollution human-induced widespread in the region. The significant degree of anisotropy resulting

from fracturing and karstification, together with regional flow patterns, favours the formation of preferential pathways for pollutants, contributing to widespread degradation of groundwater quality in the Murgia and Salento aquifers. On the other hand, the Tavoliere is the region's most significant agricultural production area, and the past spasmodic use of fertilisers has yet to be removed from the underlying aquifer.

Considering the significant role of groundwater in the economic development of the Apulian region, 31,205 chemical samples from 1986 to 2021 were collected and examined with the aim of investigating the spatial and temporal evolution of nitrate concentrations and identifying their potential causes.

Two domains were investigated (Murgia and Salento karst and the Tavoliere alluvial systems) over three periods (1995-1997, 2007-2011, 2015-2021). The areas more vulnerable to nitrate pollution were identified through Ordinary Kriging. Furthermore, relationships between nitrates and chemical (main ions), natural (lithology), and human (land use) factors were investigated based on Co-Kriging, incorporating the spatial correlation between nitrates and the aforementioned factors sampled at the same locations.

The data collected indicate a widespread upward trend in nitrate levels in recent years, despite the implementation of policies at both national and regional levels and efforts to reduce fertilizer distribution. This approach holds significant relevance in pinpointing areas that are more prone to groundwater pollution and identifying potential causes. Results underscore the significance of implementing regular monitoring programs to acquire reliable assessments of the hydrochemical space-time variability of groundwater, particularly in response to climate change and human practices.

Examining the Interplay of Crucial Geochemical Factors on the Sorption of Bisulfide onto Bentonite: A Statistical Perspective

Magdalena Krol[1], Sifat Papry[1], Pulin Mondal[1], Tarek Rashwan[2], Mehran Behazin[3], Peter Keech[3]

[1]*York University*
[2]*Open University*
[3]*Nuclear Waste Management Organization*

Abstract. Given the extensive use of nuclear energy, used nuclear fuel management is a global challenge. Deep geological repositories (DGRs) have been internationally recognized as the best approach for long-term storage of used nuclear fuel. Similar to many other nations, Canada's DGR is planned to incorporate copper-coated used fuel containers (UFCs) within a highly compacted bentonite (HCB) buffer and bentonite backfill as part of its engineered barrier system (EBS). However, under long term anaerobic conditions expected in a DGR, there is a potential for microbiologically influenced corrosion (MIC) of the UFC, where sulfate is reduced to bisulfide (HS-) and is then transported to the UFC surface, leading to copper corrosion. The extent of MIC is expected to be dependent on the flux of HS- to the UFC surface. Thus, it is important to examine the HS- transport as well as sorption mechanisms onto bentonite that can minimize the risk of corrosion and ensure the long-term performance of EBS. However, a DGR may undergo geochemical changes over its lifespan, potentially impacting the sorption characteristics of bentonite. Therefore, the aim of this research is to assess what factors significantly affect HS- sorption onto bentonite. To this end, batch sorption experiments were performed to investigate HS- sorption onto bentonite slurries as a function of temperature (10-40°C), pH (9-11), and ionic strength (0.01 M-1 M NaCl). These conditions were selected to mimic the potential geochemical conditions of a DGR. The experiment results indicate that HS- sorption increases with rising temperature but decreases with increasing pH and ionic strength. To assess whether these changes were significant, three-way ANOVA (analysis of variance) test was conducted on the experimental data to evaluate the individual and combined effects of the factors on HS- sorption onto bentonite.

The results indicate that the individual and two-way interaction effects are statistically significant and, hence, need to be considered in examining sorption mechanisms of HS- onto bentonite. Altogether, the study's findings are valuable for evaluating the impact of various DGR conditions on bentonite performance as a barrier system for used nuclear fuel storage. The results will also be used to develop a reactive transport model of HS- through bentonite under the evolving geochemical conditions of DGRs worldwide.

Acknowledgment: Natural Science and Engineering Research Council of Canada, Ontario Research Fund - Research Excellence, Nuclear Waste Management Organization

Explainable machine learning in satellite precipitation product correction

Hristos Tyralis[1], Georgia Papacharalampous[1], Nikolaos Doulamis[1], Anastasios Doulamis[1]

[1]*School of Rural, Surveying and Geoinformatics Engineering, National Technical University of Athens*

Keywords: Feature engineering, Machine learning, Spatial interpolation, Variable inportance

Abstract. We present recent experiments that utilized explainable machine learning methods in the context of satellite precipitation product correction. These experiments compared the importance of multiple predictor variables using random forest and boosting algorithms. Large datasets providing information across the contiguous United States supported the original comparisons, which were conducted at the monthly and daily timescales. The comparisons addressed both the common task of mean-value prediction (point prediction) and the less frequent task of quantifying predictive uncertainty (probabilistic prediction). The data included gauge measurements and gridded data from PERSIANN (Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks) and IMERG (Integrated Multi-satellitE Retrievals for GPM), and spanned a 15-year period for the monthly timescale and a 2-year period for the daily timescale.

We formulated the problem as a regression setting. Gauge measurements were the dependent variable, and satellite and topography data were the predictor variables. Two types of experiments were conducted. In the first type, the predictors included precipitation values from the four closest PERSIANN and the four closest IMERG grid points to the gauge, the respective distances from the gauge and station elevation. The second type of experiments used distance-weighted precipitation values from the four closest PERSIANN and the four closest IMERG grid points to the gauge, along with station elevation, as predictors, thereby introducing a new feature engineering method that effectively reduces the number of predictor variables.

In most of the examined cases, IMERG precipitation data was proven the most important for predicting ground-measured precipitation, followed by PERSIANN

precipitation data and station elevation. Specifically, for these experiments of the first type that refer to the monthly timescale and the point prediction problem, XGBoost gain scores revealed that the four IMERG precipitation values were the most important predictors (ranked according to their distance from the gauge, with closer points being more important than farther ones), followed by station elevation, the four PERSIANN precipitation values, and finally the IMERG and PERSIANN distances in a random order. For the experiments of the second type, which refer to the case of probabilistic prediction of monthly precipitation, LightGBM gain scores showed that the distance-weighted IMERG precipitation data were the most important for predicting various quantiles of the predictive probability distribution, followed by the distance-weighted PERSIANN data and station elevation. Similar results were observed in other experiments (referring, e.g., to the daily timescale).

Overall, this presentation highlights the importance of computing predictor variable importances in spatial interpolation settings, thereby potentially aiding the selection of appropriate gridded datasets for specific applications.

Acknowledgment: The research project was supported by the Hellenic Foundation for Research and Innovation (H.F.R.I.) under the “3rd Call for H.F.R.I. Research Projects to support Post-Doctoral Researchers” (Project Number: 7368).

Exploring the Influence of Prior Parameter Uncertainty on Groundwater Model Calibration: A Case Study from the Rhodope Aquifer in Northern Greece

Elias Gravanis[1], Phaedon Kyriakidis[1], Charalampos Doulgeris[2], Evangelos Tziritis[2]

[1]*Department of Civil Engineering and Geomatics, Cyprus University of Technology, Cyprus*
[2]*Soil and Water Resources Research Institute, Hellenic Agricultural Organization, Greece*

Keywords: geohydrology, FePEST, geostatistics, inverse problems

Abstract. Uncertainty characterization and quantification is a complex issue of major importance in assessing the reliability of hydrologic model outputs in the context of water resources management and science-based decision-making. Parameter estimation error is the most studied type of uncertainty arising in groundwater problems, understood as a solution error in (regularized) inverse problems in geohydrology.

The present work approaches the problem of quantifying parameter (hydraulic conductivity) estimation uncertainty through (1) conditioning on both hydraulic head observations and conductivity measurements and (2) exploring the influence of input parameter uncertainty by calibrating an ensemble of prior parameter fields conditioned on the conductivity measurements. In the context of a groundwater case study, the above are implemented in the environmental modelling software (Fe)PEST (Parameter ESTimation), which relies on non-intrusive/model-independent approaches, using FEFLOW as forward problem solver. The pilot points method is employed, as an integral part of PEST, while Tikhonov regularization in the preferred value 'fall-back position' is adopted, to constrain the solution process through the initial (input) values of conductivity and smooth out the solution. Pilot points were distributed roughly one correlation length apart, to avoid excessive overfitting, with

somewhat higher density in the vicinity of the measurement locations. To enforce prior knowledge on the calibration process, it is also necessary to introduce bounds on the parameter field. The results, given in the form of mean

value and variance maps, are compared with the post-calibration analysis of FePEST, which produces uncertainty estimates through perturbing the calibrated field in the null directions for a given initial parameter field. The results show that post-calibration significantly

underestimates parameter uncertainty.

The influence of the prior (e.g., measurements and variogram) on parameter estimation, both at the level of honoring measurements as well as uncertainty, has been long emphasized. But it is only recently that powerful and user-friendly software, such as FePEST, allow the implementation of these concepts with relatively ease by groundwater modelling practitioners.

Acknowledgment: This research was conducted in the context of MEDSAL Project (www.medsal.net), which is part of the PRIMA Programme supported by the European Union’s Horizon 2020 Research and Innovation Programme and funded by the national funding agencies of RIF (grant number 0318-0031), GSRT (grant number 2018-7), BMBF, MIUR (grant number 94I18000230007), MHESR (grant number 2018-12) and, TÜBITAK (grant number 118Y366).

Exploring the Uncertainty of Landslide Susceptibility via a Probabilistic Neural Network

Helga Jordão[1], Amílcar Soares[1]

[1]CERENA. DER. Instituto Superior Técnico. Universidade de Lisboa

Keywords: Deep neural network, Landslide susceptibility, Uncertainty

Abstract. It is irrefutable that Earth´s climate is changing rapidly. Consequently, extreme precipitation events are not only becoming more frequent but also intensifying in magnitude, spatial extent, and duration. These changes pose an increased risk of landslides, particularly in mountainous regions. Therefore, the creation of landslide susceptibility maps is crucial for assessing the probability of landslide occurrence under specific geo-environmental conditions. Identifying areas susceptible to landslides is essential, as it enables stakeholders to implement targeted mitigation strategies aimed at minimizing impacts on the economy, public safety, and ecosystems.

Numerous studies have focused on landslide susceptibility assessment, employing various approaches such as physical models, heuristic models, statistical models, and machine learning (ML) techniques. Within ML methods, deep learning (DL) has emerged as particularly promising due to its capability to effectively learn complex nonlinear relationships inherent in landslide causes and spatial heterogeneity.

Despite recent advancements in DL, many studies prioritize optimizing loss functions to enhance accuracy, often neglecting to account for prediction confidence and uncertainty. Additionally, they remain susceptible to overfitting, limiting their generalizability. Networks are still predominantly used as black-box function approximators, mapping a given input to a single classification output, thereby making their decisions less understandable and challenging to justify or explain. In traditional deep neural networks (DNNs), the learnable parameters, known as weights, are single values or point estimates. They are adjusted by optimizing a loss function, which measures the error between the predicted output and the target output of the model. This optimization process leads to deterministic network output for a given input, thereby lacking information about uncertainty estimates and limiting the ability to assess risk. These drawbacks pose significant challenges for the application of neural networks in

real-world scenarios, particularly in decision-making and risk assessment applications.

To address these limitations, we propose a Bayesian neural network (a neural network with probabilistic layers and employing the principles of Bayes' theorem) to assess landslide susceptibility and quantify the associate uncertainty. Bayesian learning offers a robust framework for training and evaluating uncertainty-aware neural networks. Unlike traditional DNNs, Bayesian neural networks assign probability distributions to all parameters, including weights and biases, rather than single point estimates. By learning the posterior distribution over weights based on training data and prior distribution, Bayesian neural networks provide a way to quantify uncertainty.

We applied this approach to develop a landslide susceptibility model and quantify uncertainty for Funchal and Ribeira Brava counties in Madeira Island, Portugal, using historical landslide data as the foundation. We constructed a model that not only predicts susceptibility but also offers insights into the level of uncertainty associated with each prediction, thereby enhancing decision-making processes and risk assessment capabilities in real-world applications.

Finding the preferred spatial resolution

Nicholas Hamm[1]

[1]University of Nottingham

Keywords: ELSA, LISA, spatial resolution

Abstract. Fine resolution datasets are often preferred for environmental applications; however, increasing the spatial resolution presents challenges. There is a higher data requirement and the larger size of the dataset may lead to it being more expensive and more difficult to work with. The high spatial resolution is unnecessary if the phenomenon of interest does not vary at this scale or if the scientific or practical application does not require it. Therefore, we are interested in developing methods to identify the most suitable spatial resolution for a given scenario.

We expect that the suitable spatial resolution may be affected by the scientific application, the phenomenon that is being investigated and data quality and modelling. For remote sensing and other raster datasets spatial resolution is usually equated with the pixel size.

This research begins by addressing the first problem. This builds on the entropy-based local indicator of spatial association (ELSA), which we have published previously. ELSA evaluates the local spatial association based on the composition (Shannon entropy) and dissimilarity in a window surrounding the point or pixel of interest. The change in ELSA with decreased spatial resolution depends on the variable, the window size as well as data quality. This is expected from theory and was also demonstrated by experimental results. We conducted structured experiments based on simple datasets (elevation, air quality, land cover) for different landscapes, pixel sizes and window sizes. We calculated ELSA for different window sizes, thus allowing us to investigate the effect of aggregation or 'upscaling'. Hence, we could identify areas with high spatial association at different scales.

Increasing the window size by different factors (2, 4, 6, 8) allowed us to identify areas that maintained a high degree of spatial association (e.g, ELSA < 0.05) at different scales. Current research aims to provide a clearer conceptual basis and to evaluate this with a wider range of datasets.

Acknowledgment: This research was supported by the Ningbo Natural Science Foundation (Project ID 2022J173).

Flood risk in Attika from building-block spatial resolution survey to pluvial rainfall-runoff simulation

Panayiotis Dimitriadis[1], Alexia Tsouni[2], Theano Iliopoulou[1], Stavroula Sigourou[2], G.-Fivos Sargentis[1], Vassiliki Pagana[2], Romanos Ioannidis[1], Dimitra Dimitrakopoulou[1], Marcos-Julien Alexopoulos[1], Demetris Koutsoyiannis[1], Charalampos Kontoes[1]

[1]*National Technical University of Athens*
[2]*National Observatory of Athens*

Keywords: flood risk, rainfall-runoff, pluvial flood simulation, building-block spatial resolution

Abstract. In this work, we present a new methodology for flood-risk evaluation, which is comprised of five stages that cover the most important sources of uncertainty in flood-risk assessments. Starting from a step-by-step topographic survey at a building-block spatial scale, we focus on the existing hydraulic conditions of the drainage and other hydraulic works along the main hydrological network. The high impacts of this survey are reflected through various techniques to the Digital-Elevation-Map and by combining the practical knowledge from previous studies and historical flood events at the area of interest, as acquired from the collaboration of the associated municipalities and the experience of the residents through questionnaires. Finally, the flood risk and hazard maps are created by sensitivity analyses of the most high-variability flood parameters, through a new 2D pluvial rainfall-runoff simulation scheme and by innovative spatial and country-based ombrian (i.e., intensity-duration-frequency) curves. The above five-stages methodology is applied at the Attika region in Greece, and the results and implications are discussed by showing how to tackle existing issues in the literature and by introducing new questions to be answered towards a more sustainable and robust flood risk assessment.

Acknowledgment: This work is performed within the frame of the research project Risk Assessment for Earthquakes, Fire and Flood in the Attica Region conducted by/in collaboration with the BEYOND research center of the National Observatory of Athens.

Flood risk mapping using an ExGan approach under climate change impact in Morocco.

Nabil El Mocayd[1], Rafia Belhajjam[2], Abdelaziz Chaqdid[3], Mohammed Seaid[4]

[1]*College of Agriculture and Environmental Sciences, University Mohammed VI Polytechnic*
[2]*Faculté des sciences, University Abdelmalek Essaadi*
[3]*Institute of Applied Physics, University Mohammed VI Polytechnic*
[4]*Department of Engineering, University of Durham*

Keywords: Climate Change, Flood risk, Generative Machine Learning, Northern Morocco

Abstract. Floods are prevalent natural disasters worldwide, impacting large populations worldwide and designing effective flood defenses and identifying at-risk areas remain among ongoing challenges that require careful consideration. Extreme precipitation serves as a key trigger for floods in many areas, and despite increased attention on this topic, understanding the influence of climate change on flood occurrences remains a complex task. The persistence of the stationarity assumption in many flood-related studies poses a significant hurdle to the realism of results.

The present research emphasizes on the effectiveness of physical-based methodologies in flood risk assessment. The central idea of this investigation involves the application of generative machine learning techniques, particularly the ExGans, to address challenges of the synthesizing extreme precipitation scenarios that accurately reflect local climatology nuances. Using refined temporal disaggregation, the ExGans demonstrate exceptional proficiency in reproducing a diverse spectrum of extreme precipitation patterns specific to the vulnerable region under scrutiny. Using these synthesized scenarios as inputs in a thoroughly calibrated hydrodynamical model would facilitate a comprehensive flood risk mapping exercise. Rigorous testing and validation within the highly susceptible Martil River basin, located in the northern Mediterranean region of Morocco, underscores the robustness of the considered model and the suitability for this region of increased vulnerability. The ability to generate multiple synthetic extreme precipitation scenarios would also allow to tackle uncertainties of the model. In our current study, the space-time uncertainty on the generated risk maps is also discussed.

Once the model is developed and well established, future scenarios of extreme precipitation, as indicated by dynamical models, are employed to force the same model, and explore the impact of climate change on flood extent in the region. This approach challenges the stationary assumption commonly associated with statistical-based studies, as the proposed ExGans approach account for the forecasting of dynamic scenarios and it can replicate the corresponding future scenarios for extreme events. The uncertainty associated with the generated risk maps are also produced which therefore allow to attribute a spatially distributed level of confidence.

Acknowledgment: The present work was supported financially by the MICODIR project, funded by the APRD Program. Thus, we extend our gratitude to the Moroccan Ministry of Higher Education, Scientific Research, and Innovation, as well as the OCP Foundation. The PhD thesis work of Abdelaziz Chaqdid was funded by the UMRP project, supported financially by OCP.

Fracture network modeling using GPR data and multiple-point geostatistics

Ain Lee[1], Seokhoon Oh[1], Hyoung-Seok Kwon[1]

[1]*Kangwon National University*

Keywords: Ground penetrating radar, Machine learning, Fracture, Underground research laboratory

Abstract. Various types of field experiments in the underground research laboratory (URL) are essential to ensure the safety of the deep geological repository for high-level radioactive waste (HLW). Fractures and faults in the rock are the primary fluid flow paths and influence rock strength and deformation characteristics. To predict radionuclide transport, it is necessary to identify the spatial characteristics of the fracture to create a fracture model. In underground spaces such as URLs, visual observation or imaging can be used to characterize the spatial characteristics of fractures, which has limitations in determining the fracture structure within the exposed rock face. Ground penetrating radar (GPR) surveys are a highly effective geophysical survey method for non-destructive imaging of fracture structures within exposed rock faces. The objective of this study is to investigate the 2D mapping of fracture planes and profiles within URLs using GPR surveys and to compare simulations of fractures that are suitable for URLs.

In geological settings, fractures vary in orientation and size, requiring analysis on a 2D horizontal plane or on a vertical profile. Traditionally, the extraction of 2D fracture data has relied on the visual observation of outcrops, imaging, and borehole survey data. Recent research has focused on fast extraction of fracture planes using machine learning techniques. Machine learning algorithms detect fractures in rocks from images with varying levels of noise, with high performance on irregular patterns. This study used a machine learning algorithm to automatically detect fractures from photographic data and extract 2D fracture planes. At the Hard Rock Laboratory (HRL) in Äspö, Sweden, GPR extracted 2D fracture profiles from the rock surface survey, clearly showing individual fractures on a meter scale. GPR provides high-resolution information about the underground subsurface by reflecting waves from discontinuities such as fractures, voids and groundwater. Electromagnetic waves reflected from fractures have larger amplitudes, allowing calculation of fracture depth and

scale, applicable to fracture network modeling. This study explores the fracture scanning method, considering the geometric structure and spatial continuity of GPR images for site characterization for URLs. We aim to propose the fastest and most effective GPR-based 2D fracture profile generation method.

To assess groundwater flow and verify URL safety, it is essential to generate 2D or 3D fracture simulations using extracted data. The discrete fracture network (DFN) model is a widely used simulation tool for fracture networks that accurately reflects their characteristics without homogenization. However, DFN models often overlook the spatial dispersion and correlation of fractures, which limits their ability to describe roughness and impacts deformation behavior. The multiple point statistic (MPS) based model is an alternative that uses probabilistic simulation to represent complex fracture patterns and demonstrate connectivity between fractures. MPS uses a training image (TI) representing geological fracture characteristics and statistical properties, obtained through various surveys. This study conducts land surface-based fracture mapping using GPR data from Äspö HRL, Sweden, and proposes an improved modeling technique through comparative analysis. This paper proposes an optimal fracture simulation method for characterizing hydraulic behavior of fractured zones within URLs, providing insights for future studies on fracture network modeling in HLW geological disposal scenarios.

Acknowledgment: This work was supported by the Institute for Korea Spent Nuclear Fuel (iKSNF) and Korea Foundation of Nuclear Safety(KOFONS) grant funded by the Korea government(Nuclear Safety and Security Commission, NSSC)(RS-2021-KN066110);and the Basic Science Research Programme through the National Research Foundation of Korea (NRF), funded by the Ministry of Education (Grant No. 2019R1A6A1A03033167)

Geospatial Data Science for Public Health Surveillance

Paula Moraga[1]

[1]*KAUST*

Abstract. Geospatial health data are essential to inform public health and policy. These data can be used to understand geographic and temporal patterns, identify risk factors, measure inequalities, and quickly detect outbreaks. In this talk, I will give an overview of geostatistical methods and computational tools for geospatial data analysis and health surveillance. I will discuss data biases and availability issues, and present modeling advancements to integrate complex data from different sources and resolutions to predict disease risk and detect outbreaks. Finally, I will describe how digital health systems can inform better surveillance using an example of dengue nowcasting in Brazil.

Geostatistical and Machine-learning Regionalization Methods for the Designation of Areas with Groundwater Nitrate Pollution in Germany

Alexander Brenning[1], Jonathan Frank[1], Thomas Suesse[2]

[1]*Friedrich Schiller University Jena, Department of Geography, Jena, Germany*
[2]*National Institute for Applied Statistics Research Australia, School of Mathematics and Applied Statistics, University of Wollongong, NSW 2522 Australia*

Keywords: geostatistical regionalization, spatial random forest, exceedance area, groundwater quality

Abstract. Implementing the EU Nitrates Directive (91/676/EEC), national regulations in Germany mandate the use of geostatistical or deterministic regionalization methods for the designation of areas with nitrate-polluted groundwater. In this context, we assess methods for a possible national-scale area designation based on an in-depth problem analysis and empirical as well as theoretical model assessments, identifying shortcomings in methods that are currently widely used. In addition, we explore the utility of various spatial random forest regionalization techniques in this application.

Suitable methods must be able to identify exceedance regions – as opposed to simply regionalizing nitrate concentration. Nonstationarity in the spatial trend and spatial dependence as well as generally varying distributional characteristics across a variety of hydrogeological settings pose additional challenges in this study with more than 13,500 measurements country-wide. The specific regulatory framework results in constraints on the admissible ancillary variables and additional user requirements.

Within the existing regulatory framework, kriging with external drift, geostatistical regression-kriging and conditional geostatistical simulations offer an established methodological toolbox that fulfils these requirements and enables transparent decision-making. Nonstationarity can be accounted for through the combination of linear or nonlinear trend models, accounting for 10-59% of the variance, with semivariograms modelled within hydrogeological

regions. To address the often-overlooked problem of quantity disagreement between designated and actual polluted areas, an unbiased estimate of the total exceedance area with pollutant levels above the legal threshold can be obtained from these models and accounted for in the decision-making process. An empirical comparison highlights possible biases in the size of exceedance areas obtained with traditional approaches that ignore local prediction uncertainty and focus on spatial prediction of nitrate concentration instead of exceedance modeling. Potentials and challenges of combining geostatistical techniques with nonlinear machine-learning models in a regulatory context are discussed, considering empirical results obtained with selected spatial random forest implementations, some of which offer a small potential for improvements in the root-mean-square prediction error although at the cost of a poorer traceability of predictions made by these black-box models.

Acknowledgment: This research was supported by the German federal environmental agency (Umweltbundesamt).

Geostatistical investigations for the suitable mapping of the Piezometric level: A case study (Skhira aquifer (Southeastern Tunisia)

Besma Tlili[1], Nabila Halouani[2], Moncef Gueddari[2]

[1]*University of Manouba*
[2]*University Tunis El Manar*

Keywords: Geostatistis, Groundwater level, Spartran variogram, Skhira aquifer (Tunisia)

Abstract. In many countries of the world, groundwater is the main source of water in arid and semi-arid regions. Water scarcity stands out as a significant issue in Tunisia, especially considering the pronounced decrease in rainfall, particularly in the southern and central parts of the country, as projected by regional climate change models.

The Skhira coastal aquifer, situated in the central-eastern region of Tunisia, faces excessive exploitation due to the increase in the number of wells. This exacerbates the alarming state of groundwater, which is under serious threat from nonpoint source pollution from saltwater intrusion resulting from overpumping. This study aims to introduce novel modeling tools for improved monitoring and prediction of groundwater levels in sparsely gauged basins. Spatial variability of water table levels is analyzed based on hydraulic heads recorded during the dry period of the hydrological year 2002–2003.

For the first time, the Spartan variogram family is applied to data from this aquifer, demonstrating its optimality for stochastic interpolation of the dataset. Finally, maps of groundwater levels and kriging variance are presented based on the optimal spatial model.

The optimal spatial interpolation approach for groundwater-level spatial variability in the Skhira aquifer is based on ordinary kriging, with the Spartan variogram model. The present findings are supported by the results of cross validation analysis. In addition, the error map identifies the locations of the Skhira aquifer with the highest kriging standard deviation (SD). According to OK, the south and weast borders of the aquifer can benefit from further sampling. Overall, the Spartan variogram model displays an excellent fit with the

experimental variogram, offering a reliable method for assessing spatial dependence in groundwater-level data interpolation studies.

Thus, accurate estimation of water table depth is essential for integrated groundwater resource management plans, particularly in aquifers with limited monitoring, such as the Skhira aquifer.

Geostatistical modelling of stratigraphical layers and geotechnical parameters: Cone Penetration Test, Modulus and Limit Pressure from Menard pressure-meter test

Pedram Masoudi[1], Cyril Simon[2], Claire Faucheux[1], Margaux Blanckaert[2], Hélène Binet[1]

[1]*Geovariances*
[2]*EDF-DI-TEGG*

Keywords: Geotechnics, Stratigraphy

Abstract. On a site under investigation, stratigraphical information is usually more abundant than geotechnical information. Geotechnical parameters depend on stratigraphical layers. So, the constructed stratigraphical model could be used in creating a robust geotechnical model. The objective in this study is to construct 3D block models of Cone Penetration Test (CPT), Modulus (Em) and Limit Pressure (Pl) from Menard Pressure-meter tests, on a nuclear power plant site under development. From a geological viewpoint, the site is on recent sediments (less than 2000 year) with three sandy strata N1, N2 and N3 with total thickness of about 30 m. N2 is sandwiched between N1 from top and N3 from bottom, and N3 is laid on a shale deposit. There is a thin stratum of silt (thickness between 0.5 m and 2.5 m) that is deposited between N2 and N3. The geotechnical behaviour of the silty thin stratum is very different from the adjacent sandy layers, e.g., the CPT is much lower in the silt. Therefore, this thin stratum is the main concern of the geotechnical investigations.

The theory of geostatistics provides a wide range of modelling tools to create 2D maps and 3D block models of categorical variables, like stratigraphical units as well as continuous variables, like geotechnical parameters. Following an exploratory data analysis, completed by analysing the spatial stationarity, three approaches were designed for stratigraphical modelling: (I) interpolating top strata by cokriging, (II) interpolating strata thickness by kriging, and (III) generating 100 possible realizations by Truncated Gaussian Simulation (TGS). The approaches (I) and (II) provide the most probable stratigraphical model, whereas the approach (III) provides a probabilistic evaluation of stratigraphical units.

In order to create geotechnical models for the CPT, Em and Pl, three approaches were designed: (A) interpolating the parameters independent from the stratigraphical model, (B) interpolating the parameters in each stratigraphic unit, individually, and (C) estimating the parameters by considering probability of each stratigraphical unit, previously calculated during approach (III). In all the approaches, the interpolation method was ordinary kriging for CPT, and ordinary cokriging for Em and Pl. The variographical and interpolation steps were performed in a flattened space following applying an unfolding method.

After comparing the results, it is concluded that the approach (C) provides the best result because geotechnical parameters depend on the stratigraphical model. In this approach, both stratigraphical model and its uncertainty are considered in geotechnical modelling. In addition, stratigraphical information is more abundant and more smooth than geotechnical parameters. Hence, the stratigraphical model serves as splitting the geotechnical information into more homogenous space.

Geostatistical modelling of total precipitation for hydrological forecast over south shore river basins of Quebec to improve hydroelectric power management

Dominique Tapsoba[1], Pedram Masoudi[2]

[1]Hydro-Québec Research Institute
[2]Geovariances

Keywords: hydrological modelling, hourly total precipitation estimation, multivariate geostatistics, ERA5-Land

Abstract. The ERA5 is a reanalysis global dataset, monitoring the earth climate and weather, hourly, since 1940. It has good geographical coverage and high spatial resolution; however, no uncertainty is associated to the available data. The ERA5 dataset is widely used in different studies. The goal of this project is to use multivariate geostatistical methods in modelling the ERA5 data and site observations, i.e., hourly measurements of total precipitation at meteorological stations. The study area is south of Quebec covering several drainage basins. The estimation of total precipitation in each drainage basin is important in hydrological forecasts and in improving hydroelectric power management. The results will be used for calibration of hydrological model of flow simulation, as well as meteorological post-processing.

From meteorological viewpoint, the advantage of site observations is being direct measurements, while ERA5 dataset is an indirect information. From the viewpoint of data configuration, the site measurements are irregularly distributed and the distance between the adjacent stations could vary from a few kilometres to more than one hundred kilometres; meanwhile the ERA5 data are regularly spaced with distance of about 10 km. In the language of statistics, the coefficient of variation of site observations is in the order of magnitude of 0.1 while the coefficient of variation of ERA5 is in the order of magnitude of 0.001. It means that ERA5 data are smoother than site observations, and the smoothness is a general disadvantage of indirect information, here ERA5.

The site observations and ERA5 dataset are heterotopic, i.e., there is a distance between these two datasets. Hence, application of statistical multivariate

analysis is limited, and a prerequisite could be interpolating one dataset to the other, so cross-variogram and co-simulation could be used. The implemented workflow contains two main steps: (i) Sequential Indicator Simulation (SIS) for calculating the probability of precipitation, and (ii) Turning Bands Co-Simulation (TBS) for predicting total precipitation in areas with probability of precipitation above zero. So, quantile maps of total precipitation are provided in an hourly frequency. Close to the principal data, i.e., site observations, the geostatistical interpolation tends to the direct measurement of total precipitation, and far from the principal data, spatial behaviour of total precipitation would be inferred from the auxiliary data, i.e., ERA5, which has a better spatial coverage. The geostatistical analysis is applied for 24 hours over a period of several weeks to confirm the robustness of results over the weeks.

Acknowledgment: Hydro-Québec Research Institute is acknowledged for the publication permission and related financial supports. The results will be exploited in two internal projects Climatologie and SIMPTOM (Système d'Information Météorologique Prévisionnel pour Terrain, Opérations et Main-d'œuvre). Mr. Rousseau-Rizzi, Raphaël (PhD) meteorologist at Hydro-Quebec is acknowledged for his feedback on the obtained results.

Geostatistical Models of Ground-Level PM2.5 and AOD from Satellite Sensors for Exposure Assessment of Myocardial Infarctions Events in Lisbon Region

Anna Kamińska-Chuchmała[1], Manuel C. Ribeiro[2]

[1]*Computational Intelligence Group, Dept. CCIA, Faculty of Informatics, University of the Basque Country UPV/EHU*
[2]*Centro de Recursos Naturais e Ambiente, Instituto Superior Técnico, Universidade de Lisboa*

Keywords: Spatial Prediction, Geostatistical Methods, Sensors Data, Myocardial Infarctions

Abstract. Detecting air pollution problems in metropolitan areas remains a major challenge. Particulate matter (PM), especially PM2.5 (PM with diameter < 2.5um), is one of the main drivers adversely affecting human health. Taking this into account, the following study examined the effects of PM2.5 concentrations on cardiovascular diseases. Investigating air pollution exposure over a large area using ground sensors is not trivial, as measurements are usually sparsely distributed, with low spatial resolution, leading to relevant gaps in population's exposure assessment. To improve spatial coverage, the capabilities of satellite instruments are currently being combined with ground sensors. For PM2.5 exposure assessment, good correlations have been found between PM2.5 measured in ground stations and AOD (Aerosol Optical Depth) values retrieved from MODIS sensor onboard NASA satellites (Terra, Aqua). Research presented in this investigation mainly focus on analysing PM2.5 measured in air quality ground stations and AOD taken from MODIS. The data covers the Lisbon region in Portugal from October 2018 to November 2019. Preliminary and structural analysis are prepared with using geostatistical approach like variogram and cross-validation. Next, authors prepare geostatistical simulations models combining ground-level PM2.5 and AOD from satellite sensors data for exposure assessment of Myocardial Infarctions events in Lisbon Region. The results obtained will be discussed before presenting a constructive summary describing the validity of the use of geostatistical models and the analysis of the direct impact of PM2.5 on health.

Geostatistical Predictive Model of Drought Severity. A Case Study of Southern Portugal

Miguel Gomes[1], Amílcar Soares[1], Ana Meira[2], Rita Durão[3]

[1]*CERENA-IST*
[2]*CERENA-ISEP*
[3]*IPMA*

Keywords: Drought, Drought indicator, Southern Portugal

Abstract. Some indicators of the drought phenomenon over the last 40 years have been analysed based on precipitation records from a set of monitoring stations in southern Portugal. For this study, the consecutive number of dry days (CDD) was adopted as the indicator that best reveals the drought dynamics over that period. Three levels of drought severity were classified in relation to CDD: A – corresponding to periods of CDD exceeding 70 days; B – corresponding to episodes of CDD lasting between 70 and 20 days; C – corresponding to shorter episodes of less than 20 days.

The objective of this study is to characterize a short-term drought severity prediction model based on short-term and long-term occurrences for the analysed period. We consider the month as the minimum unit for the predictive model. We aim to determine, for the future month t0, the probability of it being in severity class A, B, or C, based on two levels of knowledge: i) Short-term: knowing the proportions of severity levels A, B, and C for the past 3 months, t0-1, t0-2, and t0-3; ii) Long-term: knowing the proportions of A, B, and C for the target month t0 in previous years.

To achieve this, we constructed a Machine Learning (ML) predictor, Random Forest (RF) of the probabilities of A, B, and C in t0 for each monitoring station, based on short-term and long-term independent variables. After predicting (RF) the probability of drought severity level in each monitoring station in the short term, we will have, for each monitoring station, for the test set of the ML algorithm, a bivariate distribution of predicted and equivalent observed real values.

In a second phase, the predicted probabilities of short-term drought severity levels are simulated (direct sequential simulation with bivariate distributions) for

the entire study area. Predicted average maps of drought severity as well as corresponding uncertainty and risk are presented.

Acknowledgment: The author acknowledges the financial support provided by the Portuguese Science and Technology Foundation (FCT) through grant UI/BD/153765/2022 under CERENA Strategic Project FCT UIDP/04028/2020

Geostatistical-based multiscale analysis of surface and image texture: experiments in the Taklimakan desert

Sebastiano Trevisani[1], Peter L. Guth[2]

[1]*University IUAV of Venice (Italy)*
[2]*retired from US Naval Academy Department of Oceanography (USA)*

Keywords: geomorphometry, image texture, surface roughness

Abstract. A simplified geostatistical-based approach for multiscale analysis of surface roughness and image texture (Haralick et al., 1973; Herzfeld and Higginson, 1996; Atkinson and Lewis, 2000; Pike, 2000; Trevisani et al., 2009; Lindsay et al., 2019) is applied in the peculiar landscape of the Taklimakan desert. The algorithm has been devised to be robust to data contamination (e.g., related to nonstationarity, noise, hotspots and other spatial discontinuities) and to reduce at a minimum the users-based decisions (Trevisani and Rocca, 2015; Trevisani et al., 2023). Differently from the variogram, the modified spatial variability estimator uses the median of absolute differences instead to the mean of the squared differences. Moreover, increments of order 2 (i.e., differences of differences) are considered to remove the impact of local trends when computing short-range spatial viability measures. The surface roughness/texture analysis is conducted on the global digital elevation model (DEM) Copernicus-ESA GLO30. The image texture analysis is conducted on the band 12 of a Esa Sentinel 2 imagery. The approach permits to describe at different scales two key features of the spatial variability structure of the spatial fields considered: omnidirectional variability and anisotropy. These two basic indices calculated at different scales are highly informative and can then be used for landscape interpretation. In addition, these indices can be used as input features in unsupervised and supervised learning approaches. The results of the analysis are presented and discussed outlining the main findings and future developments. One of the key messages is the high informative content of anisotropy, which is less correlated across spatial scales with respect to omnidirectional roughness. The geostatistical-based tool is available with MIT license in Zenodo: https://zenodo.org/doi/10.5281/zenodo.7132160.

Acknowledgment: The participation to the conference has been possible with the funds of the project "MERGE" financed by the University IUAV of Venice (call 2023).

Geostatistics for smartfarming – Implementation of a virtual weather stations network for agricultural decision support systems

Damien Jean Rosillon[1], Sébastien Dandrifosse[1], Alban Jago[1], Michel Journée[2], Jean Pierre Huart[1], Patrick Bogaert[3], Viviane Planchon[1]

[1]*Walloon Agricultural Research Centre*
[2]*Royal Meteorological Institute of Belgium*
[3]*UCLouvain – Earth and Life Institute*

Keywords: Agriculture, DSS, Kriging, Agrometeorology

Abstract. Weather data have always been important for agriculture, but they become crucial with the emergence of precision farming and a data-driven approach to agricultural production. All applications for pest management, irrigation scheduling or even crop yield prediction depend on weather data. Through spatial interpolation, it is possible to evaluate weather conditions at any location within a region and therefore to provide key data for the development and implementation of control systems in agriculture.

This study aimed to develop an operational application for near real-time spatial interpolation of air temperature and relative humidity, both at hourly and daily scales, to support agricultural decision support systems (DSS) in Belgium. Four operational constraints were considered to select the best interpolation methods: accuracy (they must produce accurate weather data to properly estimate the risk of pest development), robustness (they must demonstrate resilience facing with unusual weather conditions and perform well in locations with peculiar topoclimatic context), reliability (they cannot fail in real-time conditions) and latency (they must run quickly, typically within a few minutes, to produce every hour weather maps with a resolution of 1 km² representing 16.000 points of prediction)

Two main datasets including six years (2016 – 2021) of hourly and daily temperature and relative humidity were used. The first dataset contained weather observations recorded by a regional network of 30 agrometeorological weather stations. The second dataset contained climate model outputs so-called ‘meteorological reanalyses’. Five spatial interpolation algorithms were investigated: nearest neighbour (NN), inverse distance weighing (IDW), multiple

linear regression (mulLR), ordinary kriging (OK) and kriging with external drift (KED). Models were calibrated for each hour (or day) and evaluated by leave-one-out cross validation.

A first objective was to select a spatial interpolation algorithm with longitude, latitude and elevation. Models based on kriging with elevation as external drift were the optimal choice with respect to the four operational constraints. Special attention was given to the selection of the semivariogram model. Three approaches were tested for its estimation, and the chosen one was to estimate a single semivariogram that is relevant for all hourly (or daily) data and that minimised the mean absolute error of predictions on the whole studied period.

A second objective was to investigate the potential benefits of using meteorological reanalyses as dynamic predictor variables in addition to position and elevation, two static predictors commonly used for spatial interpolation. It appeared that reanalyses did not improve the interpolation.

These models were integrated into an online application (http://agromet.be) where 'virtual weather stations' (i.e. weather data obtained by interpolation by opposition to physical weather stations) are available in near real-time. Such data can improve current agricultural DSS which are still largely based on the nearest weather station and help farmers to take better decisions.

In the future, the integration of farmers' weather stations is a promising option for improving interpolation, but several issues need to be addressed like data quality of and accessibility.

Acknowledgment: This project was funded by the Walloon Agricultural Research Centre (CRA-W) by a tax exemption operation of research in Belgium under the Moerman law.

Graph Recurrent Neural Networks for Stochastic Simulation of Karst Network Topology and Properties

Julien Straubhaar[1], Dany Lauzon[2], Philippe Renard[1]

[1]*University of Neuchâtel*
[2]*Polytechnique Montréal*

Keywords: Machine Learning, Graph, Karst network

Abstract. Climate change, leading to extreme events such as severe droughts and torrential rains, significantly impacts karst environments. Due to their specific structure, these environments can react abruptly to meteorological events, resulting in the rapid transport of pollutants. Understanding their behavior during such extreme events is crucial, and a probabilistic approach would enable the quantification of these risks.

Karstic systems form intricate interconnected networks through carbonate rock dissolutions. A critical aspect of modeling these networks involves generating and simulating their geometry and topology. Representing them as graphs, with property information stored in nodes and edges representing intersections and conduits, facilitates the exploration of their distributions. Consequently, modeling and graph generation play pivotal roles in studying these networks. However, modeling the complex distributions of karstic systems as graphs and efficiently sampling from them presents a challenge due to their restrained observation and comprehension, the high-dimensional nature of karst networks, and the intricate, non-local dependencies among edges within a given karst network.

This presentation outlines a methodology for generating karst network topology and simulating their properties using a sequences-of-sequences scheme and recurrent neural network. In the first step, we employ a graph recurrent neural network (GraphRNN) to learn the topology of a karst system, covering nodes and edge connections. GraphRNN deconstructs the graph generation process into a sequence of node and edge formations, conditioned on the generated graph structure up to that point. The karst topology is then generated sequentially one node and one edge at a time. After training, this enables the generation of multiple graphs based on the learned karst topology. Statistical

metrics are then used to analyze the topology of the karst systems generated by GraphRNN in comparison to the natural network to validate and test the approach.

The second step is to condition the graph on the input and output data, as well as the karst network properties associated with the nodes and edges. Geostatistical simulation algorithms are used to populate nodes and edges based on observations. The result is a set of karst networks where each member is unique, but whose statistical properties and distribution of nodes and edges adequately reflect the known natural system. This ensemble can then be used to calculate flow simulations and evaluate the uncertainties linked to the rapid transport of pollutants. The presentation highlights the key points of our methodology and illustrates our approach, primarily focusing on the graph generation part with an emphasis on potential geostatistical methods for populating karst network nodes and edges.

Acknowledgment: Funded by the European Union (ERC, KARST, 101071836).

Groundwater resources, climate change and artificial intelligence models

Daniele Secci[1], Maria Giovanna Tanda[2], Marco D'Oria[2], Valeria Todaro[2]

[1]*Università degli Studi di Parma. Universitat Politècnica de València*
[2]*Università degli Studi di Parma*

Keywords: Non-linear Autoregressive Neural Network, Long-Short Term Memory Neural Network, Convolutional Neural Network, Groundwater Resources

Abstract. This study utilizes three different artificial intelligence (AI) models, the Non-linear Autoregressive Neural Network (NARX), Long-Short Term Memory Neural Network (LSTM), and Convolutional Neural Network (CNN), to investigate the impact of climate change on groundwater resources. The neural networks are trained to replicate observed groundwater levels using monthly-scale precipitation and temperature records as input; the NARX additionally incorporates antecedent monthly groundwater levels. The trained models are then used to forecast groundwater levels until the end of this century, based on precipitation and temperature projections from an ensemble of 13 Regional Climate Models (RCMs) of the EURO-CORDEX initiative. Two emission pathways, the RCP4.5 and RCP8.5, are considered. The methodology is applied to an aquifer in northern Italy but can be easily extended to other regions of interest. The evaluation of the performance of the AI models indicates that all of them exhibit satisfactory metrics during the training phase. However, the NARX demonstrates less effectiveness in both the validation and testing phases when compared to the other models. In the future period, both the NARX and LSTM predict a decline in groundwater levels, especially under the RCP8.5 scenario, while the CNN suggests minimal changes. Given the lack of deep learning capabilities of the NARX and the challenges that the CNN may encounter in extrapolating beyond the training range, the LSTM emerges as the optimal choice for assessing the impact of climate change on groundwater resources through the analysis of time-series data.

Acknowledgment: This work was supported by OurMED PRIMA Program project funded by the European Union’s Horizon 2020 research and innovation under grant agreement No. 2222.

Hydrogeological Insights Through Complex Kriging of Groundwater Head Gradients

Graciela del Socorro Herrera[1], Hugo Enrique Júnez-Ferreira[2], Mario Alberto Hernández-Hernández[3], Claudia Capello[4], Sabrina Maggio[4], Sandra De Iaco[6]

[1]*Instituto de Geofísica, UNAM*
[2]*Licenciatura en Ciencia y Tecnología del Agua y Doctorado en Ciencias de la Ingeniería, Universidad Autónoma de Zacatecas*
[3]*Investigador por México CONAHCYT, Instituto de Geofísica, Universidad Nacional Autónoma de México*
[4]*Department of Economic Sciences, University of Salento*
[5]*Department of Economic Sciences, University of Salento, National Future Center of Biodiversity, National Centre for HPC, Big Data and Quantum Computing*

Keywords: Groundwater, Hydraulic head gradients, Complex kriging, Basin of Mexico

Abstract. Understanding groundwater head gradients is crucial for hydrogeology and groundwater management. Together with hydraulic conductivity, it provides valuable insights into the speed and direction of groundwater flow within an aquifer, which is important for various applications such as water resource management, contamination assessment, and aquifer characterization. Complex kriging is designed to estimate complex-valued random fields and has found novel applications in various domains, including the estimation of vectorial datasets in two-dimensional space, such as wind velocities and ocean currents. In a pioneering effort, the capability of complex kriging is employed in the present work to estimate groundwater gradients for the years 2002 and 2007, focusing on the upper aquifer of the southern part of the Basin of Mexico aquifer system, which is characterized by high water stress.

This study presents a comparative analysis of two kriging methodologies for estimating hydraulic head gradients. The first method, referred to as the "traditional kriging approach," uses ordinary kriging to estimate hydraulic head values. Subsequently, hydraulic gradients are derived from these values by computing slopes in both the x and y directions. In contrast, the second method, termed the "complex kriging approach," directly uses hydraulic head data to compute gradients. This is achieved through a triangulation process based on the spatial positions of the data, utilizing the Delaunay triangulation method. In

this approach, each triangle's hydraulic head gradient is approximated by the gradient of the plane that passes through the hydraulic head values at each vertex of the triangle. The resulting hydraulic gradients are then assigned to the triangle centroids and represented as vectors. Complex ordinary kriging is applied to analyze these vectors and generate vectorial estimates across the same grid utilized in the traditional approach.

Overall, the empirical evidence demonstrates that the use of the complex approach improves the estimate's reliability compared to the traditional one. Indeed, the goodness-of-fit of the complex covariance model for assessing the spatial correlation of the investigated vectorial components is confirmed by the statistics results, expressed in terms of Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) between the actual values and the estimated ones for the two components

Hypothetical Optical Image Generation Using Multi-sensor Image Fusion for Environmental Monitoring

No-Wook Park[1], Soyeon Park[1], Seokhoon Oh[2]

[1]*Inha University*
[2]*Kangwon National University*

Abstract. The rapid growth of sensor technology has made a lot of Earth observation satellite images more available for environmental monitoring. Despite the increasing availability of optical satellite images, a trade-off between spatial and temporal resolutions often makes it difficult to collect satellite images at the optimal resolutions for a particular application. For example, monitoring of small-scale croplands requires image time series at high spatial resolution to fully account for crop growth cycles at a local scale. However, satellite images with high spatial resolutions typically have low temporal resolutions, making it often challenging to construct complete image time series at high spatial resolution. In contrast, some satellite sensors, such as MODIS and Sentinel series, provide dense image time series, but their spatial resolutions are too coarse to be applied for local analyses. From a data availability perspective, optical satellite images are often contaminated by clouds and shadows. This limited availability of cloud-free optical images at desired times is another obstacle to environmental monitoring. Synthetic aperture radar (SAR) images can provide surface information regardless of weather conditions, but there is a difficulty in the visual interpretation of the SAR imagery compared to optical images. To resolve the above two issues in environmental monitoring using optical satellite images, hypothetical optical image generation can be one alternative to acquiring cloud-free optical imagery. To this end, multi-sensor image fusion, which utilizes the individual advantages of each sensor imagery, can be a promising approach.

This presentation introduces advanced multi-sensor image fusion approaches to generate hypothetical optical imagery, focusing on spatiotemporal image fusion and SAR-to-optical image translation. First, we present high spatial resolution image prediction by spatiotemporal image fusion. After introducing the

principles of spatiotemporal image fusion, an advanced method for spatiotemporal fusion of relatively high spatial resolution images is demonstrated by an experiment in cropland. Second, we present optical image prediction by deep learning-based SAR-to-optical image translation for optical image reconstruction. Generative adversarial networks (GANs) are employed to predict reflectance values from SAR-based features. The potential of the GAN-based hypothetical optical image generation is demonstrated via an early crop mapping experiment using Sentinel-1 and -2 images. Finally, we discuss future research directions and potential application fields of multi-sensor image fusion for hypothetical image generation.

Acknowledgment: This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. NRF-2022R1F1A1069221).

Identification of hydrological droughts in small catchments: an innovative approach to supplementing missing hydrological data using extra trees and spatial clustering

Bożena Pius[1], Krzysztof Leki[1]

[1]*Nicolaus Copernicus University in Toruń*

Keywords: hydrological droughts, random trees, time series, types of droughts

Abstract. In the structure of river flow, extreme events play a significant role. Hydrological droughts, unlike floods, develop very slowly. Research on hydrological droughts includes monitoring climate change and analysing hydrological data. Based on these studies, adaptive strategies, remedial measures, and actions for sustainable water resource management are developed. Hydrological droughts are most often described as a spatial and temporal process, resulting from meteorological, soil, and hydrogeological droughts. The noticeable deficit of water resources in the environment leads to damage to the economy and environment. Research on hydrological droughts, limited to small catchments, is rare because it requires access to appropriate historical data on river flows. Small rivers are extremely sensitive to climate and anthropogenic changes in catchments. Therefore, it is important for research to cover as many catchments as possible. The main goal of the study was to identify and isolate types of hydrological droughts in Poland.

Data from the period 1961-2020 used in the study were provided by the Institute of Meteorology and Water Management in Poland (IMGW). They included 124 time series from small catchments with an area of up to 500 km². The main problem was filling gaps in the measurement data. Out of the 124 time series, it was necessary to fill measurement gaps in 51 cases. Data were only supplemented when the period of missing data did not exceed 24 months. To solve this problem, the Extra Trees algorithm was used. It operates similarly to the Random Forest algorithm, but the main difference with the Extra Trees algorithm is that the splits in the nodes are made randomly. This aims to reduce variance without increasing the error level. For selected stations, treated as predictors, a model of Extra Trees was created based on the Bayesian optimisation method, aimed at the best possible prediction of missing values.

The similarity of the physiographic conditions of the catchments was also taken into account, limiting the possibility of supplementing data based on stations located in different hydroclimatic conditions.

Then, in the time series, separately for each month, the threshold flow characteristic of hydrological droughts and other parameters, such as duration, start, and end of the phenomenon, were identified. The spatial variability of hydrological droughts in small catchments is very high and depends on many factors. Therefore, searching for spatial regularities is extremely difficult. For this purpose, catchment clustering was conducted using the method of agglomerative feature clustering. In the first step, a distance matrix between features was created, based on information about the occurrence of hydrological drought on a given day. Then, the distance between stations was calculated, measured as the proportion of the number of days on which different phenomena occurred between two stations relative to the total number of days in the study period. Using this distance matrix, hierarchical feature clustering was conducted using Ward's method, isolating 8 groups with the highest similarity between stations. As a result of the clustering, the spatial distribution of hydrological drought groups turned out to be related to the physiogeographic conditions of the catchments and the impact of human pressure, especially areas associated with coal mining.

Improving the forecasts of a numerical wave model based on Kalman Filters and Artificial Neural Networks.

Athanasios Donas[1], Ioannis Famelis[3], George Galanis[2], Dionysia Kolokotsa[4]

[1]*microSENSES Laboratory,Department of Electrical and Electronics Engineering,Ancient Olive Grove Campus,University of West Attica Mathematical Modeling and Applications Laboratory, Hellenic Naval Academy, Hatzikiriakion*
[2]*Mathematical Modeling and Applications Laboratory, Hellenic Naval Academy, Hatzikiriakion*
[3]*microSENSES Laboratory,Department of Electrical and Electronics Engineering,Ancient Olive Grove Campus,University of West Attica School of Science & Technology, Hellenic Open University*
[4]*Energy Management in the Built Environment Research Lab, School of Chemical and Environmental Engineering, Technical University of Crete*

Keywords: Kalman Filters, Artificial Neural Network, Significant Wave Height, WAM-cycle 4

Abstract. Improving numerical wave prediction models' forecasting capabilities is essential since they have an impact on numerous activities, such as: maritime transportations, search and rescue, renewable energy sources and tourism. To this end the proposed work introduce a new optimization technique for numerical wave predictions, utilizing both Kalman Filters (KFs) and Artificial Neural Networks (ANNs). Kalman Filters are a useful tool to eliminate systematic biases as they are concerned the statistically optimum sequential estimate process that combines observations with recent forecasts recursively. However, in many cases such filters fail to detect the remaining non-systematic error and thus the providing predictions are unstable. To avoid this drawback this study implements a Feed-Forward Neural Network as a self-adaptive dynamic system to adjust crucial parameters of the filter. More specifically, in every step of the filtering the covariance matricies are estimated as a non-linear combination of their prior state and the direct output of a Feed-Forward Neural Network. Through that procedure the modified Kalman Filter seeks to enhance the model's predictions by interpreting both the systematic and the non-systematic error of the simulation. The proposed methodology was applied in various

numerical weather predictions models, but this study focuses on results concerning Significant Wave Height forecasts from the WAM-cycle 4 numerical wave model.

Inferring Features of Random Fields Using Dimensionality Reduction and the Correlated Pseudo-Marginal Method

Lea Friedli[1], David Ginsbourger[1], Niklas Linde[2]

[1]Institute of Mathematical Statistics and Actuarial Science, University of Bern, Switzerland
[2]Institute of Earth Sciences, University of Lausanne, Switzerland

Keywords: Bayesian Inversion, Dimensionality Reduction, Latent Variables

Abstract. We consider non-linear Bayesian inversion problems involving high-dimensional random fields describing hydrogeological or geophysical properties and indirect data. Fully inverting target features of the field (such as hyperparameters or primary dimensionality reduction components) alongside the local properties, often entailing thousands of unknowns, poses significant computational challenges. Reducing the dimensionality of the parameter space by neglecting local properties might introduce biases. To mitigate these errors and to avoid the computational challenges associated with high-dimensional full inversions, we employ the correlated pseudo-marginal (CPM) method. In this approach, we infer the features of interest, treating the remaining random variables as latent. The intractable likelihood of observing the data given the target features of the field is then approximated through Monte Carlo averaging across realizations of the latent variables. One drawback of the CPM method is the need of well-chosen importance sampling distributions to make the method practical in data-rich settings with low noise levels. This presented a challenge in prior research employing the CPM method, specifically if inferring only the hyperparameters of a high-dimensional field while treating all the local properties as latent variables. By jointly inferring the primary dimensionality reduction components of the field along with the hyperparameters, the residual randomness of the field has a diminished impact on the observed data, thereby reducing the criticality of a precisely tuned importance density. We assess the effectiveness of our approach in estimating hyperparameters associated with hydrogeological or geophysical property fields. However, integrating the CPM method with a dimensionality reduction approach not only facilitates efficient and unbiased hyperparameter estimation but also introduces a framework

capable of mitigating bias in general high-dimensional inversion problems that employ dimensionality reduction methods.

Innovating for Urban Wellbeing: The Integration of Blue-Green Regenerative Technologies

Charalampos Zafeiropoulos[1], Anastasios Temenos[1], Anastasios Doulamis[1], Nikolaos Doulamis[1], Aikaterini Angeli[2]

[1]*National and Technical University of Athens*
[2]*National and Kapodistrian University of Athens*

Keywords: Blue-Green Solutions, Public Health, Wellbeing, Machine Learning

Abstract. European and international cities face major challenges resulting from global geopolitical, economical, climate and other changes. This intensifies inequalities in health, due to several factors such as living conditions, health-related behaviour, education, occupation and income. Urban areas are under huge pressure to enhance policies in order to become more sustainable and liveable, as well as to provide environments and social fabric that support Public Health (PH) and Wellbeing (WB). The conventional approach to urban and revitalisation planning is based mostly on profit criteria with routine methodologies, often lacking advanced integrated methods and concepts with emphasis on health, societal and environmental aspects (including Climate Change-CC). Moreover, local communities' needs are neglected and hence cities often end up with solutions that are not embraced by local communities and citizens. To address these challenges, the proposed work aims to comprehensively address factors influencing PH&WB, provide evidence base improving urban health and reducing health disparities through changes in individual behaviour of citizens, stimulated by different policies. The proposed framework includes (a) emotional sensing devices for capturing the emotion and phycological states of the citizens, (b) wearable physical activities sensors for measuring different types of citizens' exercise, (c) a Chatbot software platform for recommendation purposes and acquisition of emotional and Quality of Life parameters which challenging to get measured by the wearable sensors, and (d) the Data Management System platform for providing advanced decision making and support by exploiting deep Machine Learning tools. Analysis and processing software toolkits are employed to integrate multi-dimensional and heterogeneous data, extracting meaningful information from the above technologcal inovative tools harmonized for each patient's daily routine. In this regard, ensemble learning regression models like Random Forest, LightGBM, and

XGBOOST can effectively analyze these datasets, unveiling correlation patterns and interdependencies among clinical, personal, environmental, and air quality data.

Acknowledgment: The research work was supported by the European Union funded project: ‘HEAlthier Cities through Blue-Green Regenerative Technologies: the HEART Approach’—‘HEART’ under the European Union and the grant agreement No. 945105.

Integrated cluster classification of small catchments considering catchment characteristics and extreme value analysis,

Anika Hotzel[1], Christoph Mudersbach[1]

[1]*Bochum University of Applied Sciences, Department of Hydraulic Engineering and Hydromechanics*

Keywords: cluster analyses, catchment hydrology, flood risk

Abstract. An in-depth understanding of how a catchment reacts to extreme hydrological events is crucial for assessing hazards and risks to humans and the environment more precisely. Historically, floods have caused considerable damage, especially in small, unmonitored catchments. To identify these risks early and initiate countermeasures, this approach presents an integrated classification method. It examines the reactions of small catchment areas to hydrological extremes using specific catchment characteristics (geographical and morphometric factors) and runoff extreme values. The aim is to find an optimal combination of characteristics that best represents the catchment's response and can be utilized for further risk assessments in ungauged areas.

The analysis was conducted using datasets from the federal state of North Rhine-Westphalia. With the aid of a digital terrain model and GIS analyses, catchment characteristics such as morphological factors (elongation ratio, shape factor, etc.) were determined. Additionally, extreme value statistics were compiled for the catchment areas using runoff data from the State Office for Nature, Environment, and Consumer Protection of North Rhine-Westphalia. Through numerous k-means cluster analyses, various combinations of these characteristics were explored to identify the most significant groupings indicating specific runoff behaviors in the catchment areas. These results are then compared with the runoff data, and the clusters are hierarchized based on the runoff statistics and catchment characteristics within each cluster.

This geostatistical approach leverages the spatial variability of catchment characteristics and the distribution of extreme runoff values to classify catchment areas effectively. By integrating geospatial data and hydrological extremes into the k-means clustering process, it enhances the understanding of catchment response to extreme events. Furthermore, this methodological

framework serves as a basis for future research in hydrological modeling and risk assessment, particularly in ungauged areas where data scarcity poses significant challenges. The incorporation of geostatistical techniques enriches the analysis by considering spatial relationships and variability, providing valuable insights for improved decision-making and management of hydrological risks.

Integrating massive and heterogeneous spatio-temporal data in environmental science.

Baptiste Alglave[1], Marie-Pierre Etienne[2], Youen Vermard[3], Mathieu Woillez[3], Etienne Rivot[3]

[1]*Lab-STICC*
[2]*IRMAR*
[3]*UMR DECOD*

Keywords: hierarchical modelling, change of support, preferential sampling, marine ecology

Abstract. In all the fields of environmental sciences, massive and heterogeneous datasets are progressively becoming available. These datasets open possibilities to infer spatio-temporal processes at a fine spatio-temporal resolution, and urge to address methodological challenges needed for this integration.

First, sampling is not always standardized ; sampled locations can be concentrated in areas of higher process intensity. This is often referred to as preferential sampling and it can lead to biased predictions if not accounted for in inference. Second, the data can be aggregated at various spatial scales. This is usually called change of support (COS). At this step, most COS approaches handle additive random variables such as Poisson, Gaussian or Gamma variables. However, no approach exists to handle complex data such as highly tailed or zero-inflated data which are common in environmental science. Finally, the combination of these heterogeneous data sources requires a careful insight of their relative contribution to inference so as to avoid model misspecification.

A generic framework has been developed to tackle these methodological issues. The approach is illustrated through applications in marine ecology, geography and climate science.

The spatial preference of sampling agents can depend on many factors. For instance, sampling agents that intend to record birds for citizen science programs will most probably target the species of interest and will also be affected by factors such as habits, closeness to a road or hiking trails. Our model accounts for the preference to the variable under study as well as additional covariates that can be included in the model. The model we introduce allows to unbias model outputs while being generic and parsimonious.

The COS component of the model allows to account for complex data such as zero-inflated or highly tailed data. Such component is crucial to estimate properly the spatial predictions. Our model allows to unbias the predictions and to provide unbiased estimates of the parameters relating the response variable and the predictors.

Finally, the combination of the distinct data sources is realized within a hierarchical framework. Such framework allows to combine several datasets and to assess the consistency between these through standard statistical tests. We demonstrate that if COS is not accounted properly in the model, the statistical test will most probably reject consistency which emphasizes that one model component is misspecified.

At this step, no generic framework allows to tackle jointly these issues while they are very common in environment. In this presentation, we propose a framework that serves as a basis for integrating massive environmental science while accounting for potential bias arising from these data.

Integration of Exploration and Production Data Towards an Improved Resource Model of a Marble Quarry

Ioannis Kapageridis[1], Charalampos Albanopoulos[2], Evangelos Gialamas[1]

[1]*Department of Mineral Resources Engineering, University of Western Macedonia*
[2]*Iktinos Hellas SA*

Keywords: marble, normal scores, declustering, resource estimation

Abstract. Estimation of marble quarry resources is commonly based on exploration data (drillholes and quarry faces), characterised by experienced personnel (or more recently by some machine-learning based process) as to the aesthetic parameters relevant to the commercially exploitable marble products of the specific quarry, and to the density of natural joints and fractures that affect the percentage of recoverable blocks from the quarried units (slabs). Aesthetic parameters include marble features such as background colour, vein texture, density and homogeneity, and the presence and density of visual defects such as stains, marks, and glass. These parameters are categorical and their representation in exploration databases is based on discrete, non-scaled values. The density of natural joints and fractures is measured as to the prominent joint sets, and the resultant densities are represented in the database as separate parameters of scaled or non-scaled values.

As the drill core surface area is very small compared to the surface area of the slabs, it is difficult, even for experienced personnel, to characterise core samples objectively and consistently as to the considered aesthetic parameters. This issue resembles the change of support problem found in geostatistical resource estimation studies of other mineral deposits, with the additional challenge that the aesthetic variables are not continuous or even numerical in the case of marble. Similar issues are present in the natural joints and fractures density variables. These problems lead to very different distributions for all considered variables between exploration samples and produced slabs, and difficulty in formulating a reliable resource estimation strategy. Spatial clustering of drillholes due to non-regular drilling campaigns further increases the difference between production and exploration data.

The procedure described in this paper takes advantage of georeferenced production data from the extracted marble slabs. This data consists of separate percentage variables for each of the marble product qualities considered. Each slab is represented as a “drillhole sample” in the database and marble product quality percentages are recorded for the entire length (slab height) of the “sample”. Exploration drillhole data is also processed to derive the product quality percentages for each interval using the aesthetic and fracture density variables, and the original intervals are then composited to lengths similar to the heights of the slabs. Both production and exploration data are then transformed to normal scores, as their distributions are highly skewed. This transformation also helps exploration data become compatible with production data so that they can be combined for resource estimation. Variogram models of the normal scores are derived from the combined data set, a block model is estimated using ordinary kriging, and the estimates are transformed back to the original distribution based on the production data transformation. Extra de-clustering weights are calculated and used during estimation.

This procedure ensures that the estimates follow closely the percentages reported from quarry production and eliminates the bias introduced by the non-regular drilling pattern and characterization of drillhole samples based on human interpretation. A case study on a dolomitic marble deposit demonstrates the benefits of this approach.

Interactive web-based dashboard for analyzing geographical and temporal spread of COVID-19 in South Korea

Hanbyul Song[1], Kyulhee Han[1], Ashok Krishnamurthy[2], Taesung Park[1]

[1]*Seoul National University*
[2]*Mount Royal University*

Keywords: COVID-19, R Shiny, Interactive Dashboard, Mathematical model

Abstract. The COVID-19 pandemic has necessitated the development of robust tools for tracking and modeling the spread of the virus. We present 'K-Track-Covid,' an interactive web-based dashboard developed using the R Shiny framework, designed to analyze the geographical and temporal spread of COVID-19 in South Korea through dynamic UI elements. Our dashboard integrates validated epidemiological models (SEIRD and SVEIRD) alongside population count data at 1km x 1km resolution. The Simulation tab applies a compartment model within each grid cell, with transitions among disease states driven by integro-differential equations that account for spatial interactions through an exponential decay function of distance. Additionally, the dashboard's Map and Regional Trend tab use daily regional data covering all 17 of South Korea's major administrative regions. It allows users to customize views by selecting specific time frames, geographic regions, and demographic groups, enabling detailed analysis of daily fluctuations, cumulative case counts, and mortality statistics through interactive visualizations. The dashboard is designed to assist researchers, policymakers, and the public in understanding the spread and impact of COVID-19, thereby facilitating informed decision-making. All data and resources related to this study are publicly available to ensure transparency and facilitate further research.

Acknowledgment: This research was supported by research grants from the Ministry of Science and ICT, South Korea (No.2021M3E5E3081425).

Karst Spring Discharge Prediction using Multiple Point Statistics and Deep Learning Methods

Xiao Xia Liang[1], Erwan Gloaguen[1], Philippe Renard[2], Julien Straubhaar[2]

[1]*Institut national de la recherche scientifique*
[2]*Université de Neuchâtel*

Keywords: Karst systems, Graph neural network, Long short-term memory, Multiple point statistics

Abstract. Karst aquifer systems are an important water resource for human consumption, and industrial purposes such as manufacturing, and resource extraction. However, these systems are highly heterogenous and anisotropic with fast water flow through conduits, and slow water seepage through rock matrix and fractures. Therefore, it is difficult to make accurate spring discharge forecasting for future sustainable use. Here, we compare the multiple point statistics (MPS) and deep learning (DL) methods for forecasting discharges at 3 spring locations given observations at 1 upstream station and meteorological data from a local station. With the MPS method we use a direct sampling, non-parametric model called Geesse. This model consists in populating the time series (in a random order) to be simulated by borrowing patterns from the training datasets. While in DL, we utilize a recurrent neural network (RNN) model called Long Short-Term Memory (LSTM) and graph neural network (GNN) model called Graph WaveNet (GWN). The LSTM model is a data-driven, recursive network that was first designed to address long time series gradient vanishing problems. Since its development, the LSTM model has produced many successes working with time series forecasting. The GWN model is a spatio-temporal, data-driven model that employs a graph structure. The graph structure contains nodes and edges that are used here to map the relative spatial connectivity of all the measuring stations (total of 4 stations with 3 springs and 1 upstream station). The nodes represent the stations and edges represent the connectivity between the stations. In addition to the spatial connectivity of the nodes, temporal data is embedded into the nodes for forecasting. Data measured from the Milandre karst system in Switzerland is used in this study. We trained the 3 models with temporal data such as temperature, precipitation, seasonal trends, and flow rates to make the springs' discharge predictions. We showed that alternate advanced methods that are

less labour intensive than physical modeling can be used to obtain accurate spring discharge forecasting for complex karst systems.

Acknowledgment: Natural Resources Canada

Leveraging Machine Learning for Enhanced Analog Weather Forecasting

Stylianos Hadjipetrou[1], Phaedon Kyriakidis[1]

[1]*Cyprus University of Technology*

Keywords: Near-surface variables, Numerical Weather Prediction (NWP), Machine Learning, Analog Forecasting

Abstract. Accurate forecasting of weather variables, such as temperature and precipitation, is crucial for various environmental applications, ranging from agriculture and energy management to disaster preparedness and air quality monitoring. In this study, we propose a novel approach that combines machine learning techniques and analog forecasting to enhance the prediction accuracy of near-surface variables. The proposed model aims to provide accurate predictions for near-surface weather variables over a time horizon of a few days. The methodology entails discerning weather analogs—past instances exhibiting similar atmospheric conditions and patterns—from historical Numerical Weather Prediction (NWP) model outputs intertwined with auxiliary variables to augment predictive efficacy.

The model uses ensemble learning algorithms and Neural Networks (NN) to discern complex interrelationships in space and time between input variables and target near-surface parameters. NNs can extract intricate features and patterns from multivariate input predictors time-series through nonlinear transformations. The framework employs NNs to learn empirical transformations, facilitating the clustering of similar forecasts in a transformed space. Analog forecasting techniques are integrated into the model framework to identify past weather situations closely resembling current atmospheric conditions. By comparing current NWP data with historical records, the model identifies analog situations and adapts them to the present context for forecasting. This approach provides valuable insights into the short-term evolution of near-surface variables based on weather analogs.

The case study over the island of Cyprus demonstrates the effectiveness of the proposed approach. The performance of the proposed model is evaluated using cross-validation techniques. Preliminary results demonstrate promising accuracy and reliability in short-term near-surface variable predictions. The model's

ability to capture subtle variations in weather patterns and incorporate auxiliary information enhances its utility in environmental applications, particularly in scenarios requiring, spatially explicit, near-surface variable forecasts for decision-making. The findings contribute to the broader field of geostatistics by demonstrating the efficacy of machine learning and weather analog identification in enhancing spatial forecasting of near-surface variables using NWP model outputs. The approach holds promise for improving decision-making processes in sectors reliant on accurate and localized weather information.

LISTEN-EO research project: The contribution towards a better understanding of the land surface interactions

George P. Petropoulos[1], Spyridon E. Detsikas[1], Christina Lekka[1], Vasileios Anagnostopoulos[2]

[1]*Department of Geography, Harokopio University of Athens, Greece*
[2]*Agile Actors Hellas S.A., Halandri, Greece*

Keywords: Earth Observation, Land Biosphere Models, Land Surface Interactions, SimSphere

Abstract. Today, particularly in the face of climate change, the study of land-atmosphere interactions is an increasingly important but challenging field. Earth observation (EO), simulation process models and ground instrumentation, as well as advances in other fields of geoinformation such as cloud computing and software tool development, have shown a great promise towards paving the way for addressing the challenges existent today associated to the study of the land-atmosphere interactions. The use of land-biosphere models has played a key role in expanding our ability to study Earth system processes and improve our understanding of how different components of the system interact. LISTEN-EO research project, funded by the Hellenic Foundation for Research and Innovation programme, aims to unravel the potential of the above technologies to advance our knowledge and also to develop a series of innovative methodologies, software platforms and modelling tools to radically support the forefront of research and practical applications related to land surface interactions. In this context, the current presentation provides an overview of the project, presenting its objectives and expect contribution in the view of the current state-of-the-art and scientific challenges. It also introduces the main software and modelling tools that will form the backbone of the project to achieve the proposed scientific objectives. Last but not least, it informs on the latest publications and software tools produced from the research project so far.
The research presented herein has been conducted in the framework of the project LISTEN-EO (DeveLoping new awareness and Innovative toolS to support efficient waTer rEsources management Exploiting geoinformatiOn technologies), funded by the Hellenic Foundation for Research and Innovation programme (ID 15898).

Acknowledgment: The research presented herein has been conducted in the framework of the project LISTEN-EO (DeveLoping new awareness and Innovative toolS to support efficient waTer rEsources man- agement Exploiting geoinformatiOn technologies), funded by the Hellenic Foundation for Research and Innovation programme (ID 15898).

Machine learning for dealing with extreme precipitation events in satellite product correction

Georgia Papacharalampous[1], Hristos Tyralis[1], Nikolaos Doulamis[1], Anastasios Doulamis[1]

[1]*School of Rural, Surveying and Geoinformatics Engineering, National Technical University of Athens*

Keywords: Distributional regression, Extremes, Machine learning, Spatial interpolation

Abstract. We present a study that introduced multiple distributional regression machine learning algorithms from the GAMLSS (Generalized Additive Models for Location Scale and Shape) family, as well as combinations of these algorithms, for correcting satellite precipitation measurements by simultaneously quantifying predictive uncertainty with an emphasis on extreme events. This study builds upon previous ones that employed quantile regression in the field. It leverages the favourable properties of distributional regression, which enables extrapolation but at the cost of reduced flexibility due to imposing a conditional probability distribution on the dependent variable. Predictive probability distributions include both widely used ones, such as the Gaussian, and those with a positive support and mass at zero, with the latter effectively modelling spatial rainfall intermittency

For extensively comparing the new approaches, we used big precipitation data that span over a 15-year period and across the contiguous United States. These data include ground-measured data and information from the PERSIANN (Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks) and IMERG (Integrated Multi-satellitE Retrievals for GPM) gridded datasets.

We formulate the problem as a regression one. Gauge observations are the dependent variable, while distance-weighted precipitation from the four closest PERSIANN and the four closest IMERG grid points, along with gauge elevation, are the predictor variables.

The comparison conducted additionally relied on metrics particularly suited for the task of evaluating probabilistic predictions, specifically on consistent scoring functions and proper scoring rules, and allows the identification of the best algorithms amongst those introduced. Both scoring functions and scoring rules are types of loss functions. Scoring functions take point predictions and observations as input, while scoring rules handle probability distribution predictions and observations. Consistency and propriety are desirable properties for scoring functions and scoring rules, respectively, that ensure the honest evaluation of predictions.

The identification of the best algorithms paves the way for new precipitation products that offer improved uncertainty estimation and model extreme precipitation events more effectively. These products would be produced by integrating satellite precipitation data with elevation information at every location.

Acknowledgment: The research project was supported by the Hellenic Foundation for Research and Innovation (H.F.R.I.) under the “3rd Call for H.F.R.I. Research Projects to support Post-Doctoral Researchers” (Project Number: 7368).

Mapping the seagrass beds - UAV Reveal the Secrets of Zostera noltii in the River Mira Estuary

Pedro Nogueira[1], Helena Adão[2], Miguel Quelhas[3]

[1]*ICT- Institute of Earth Sciences, Évora Pole; Department of Geosciences, Evora University*
[2]*Department of Biology, Evora University; MARE – Marine and Environmental Sciences Centre*
[3]*Department of Biology, Evora University*
[4]*MARE – Marine and Environmental Sciences Centre*

Keywords: Remote Sensing, Multispectral image, Drone, Seagrass beds

Abstract. The application of advanced remote sensing technologies has significantly enhanced our understanding and management of coastal and marine ecosystems. In this context, the use of drones equipped with multispectral cameras offers a promising advanced tool for detailed mapping and ecological studies. This study focuses on the mapping of Zostera noltii, an essential seagrass species, in Mira estuary (SW coast, Portugal), employing drones with multispectral imaging capabilities.

The Zostera noltii plays a crucial role in coastal ecosystems, providing high diversity and abundance of the assemblages and contributing to sediment stabilization, and supporting high nutrient cycling rate. However, anthropogenic pressures and environmental changes pose threats to its distribution and health. Traditional mapping methods are often limited by their spatial resolution and the extent to which they can capture the complexity of seagrass habitats. The use of drones overcomes these limitations, offering high-resolution, accurate, and comprehensive data over extensive areas.

This methodology involved systematic drone flights over the estuary, capturing multispectral imagery that was later processed to delineate and characterize Zostera noltii beds . The analysis enabled not only the identification of the spatial distribution of the seagrass beds but also provided insights into its health and biomass. Furthermore, the multispectral data facilitated the examination of the interactions between Zostera noltii and adjacent ecosystems, including their response to estuarine gradient and anthropogenic effects.

Preliminary results demonstrate the effectiveness of drones with multispectral cameras in mapping Zostera noltii with high precision. The detailed maps produced in this study offer valuable information for conservation planning, habitat management, and the formulation of management and conservation strategies to mitigate environmental change effects. Additionally, our findings underscore the potential of drone-based multispectral imaging in understanding the complex dynamics between seagrass beds and other coastal ecosystems.

This research contributes to the growing body of literature on the use of unmanned aerial vehicles (UAVs) in marine and coastal ecology, highlighting their potential to enhance our understanding of the seagrass bed dynamic. Moreover, it sets a precedent for the application of this technology in the monitoring and management of other critical estuarine and marine habitats around the world.

Acknowledgment: FCT - Fundação para a Ciência e a Tecnologia, I.P., under the projects UIDB/04683/2020 (https://doi.org/10.54499/UIDB/04683/2020) e UIDP/04683/2020 (https://doi.org/10.54499/UIDP/04683/2020).

Modeling N02 concentrations over the Brussels Capital Region, Belgium: a citizen science approach

Patrick Bogaert[1]

[1]*Université catholique de Louvain*

Abstract. Air pollution is a major human health issue for most major cities worldwide. Thanks to the use of a stricter EU legislation and the adoption of limited traffic zones, the situation largely improved in Brussels Capital Region (BCR) and the city of Brussels, Belgium, over the last two decades for the most common pollutants. However, the N02 concentrations still occasionally exceeds the EU recommendations. Moreover, the planned EU regulations will align with the OMS recommendations by reducing the N02 mean annual threshold down to 10 µg/m3, and all measuring stations in BCR are currently exceeding this limit. It is also known that N02 concentrations largely vary over the region, and it is thus important to identify the most problematic areas for proposing targeted mitigation measures. Unfortunately, the limited number of reference measuring stations prevents a comprehensive monitoring of these variations over space.

In order to bridge this gap, a large citizen science project was conducted in October 2021 for collecting 3000 air quality samples spread over the whole region over a one-month time span using passive tube samplers (see https://curieuzenair.brussels/en/home/). It constitutes a unique data collection at the international level, with an average density of 18 samples per square kilometer. Preliminary analysis shows a ten-fold increase between the lowest (6.2 µg/m3) and highest concentrations (60.5 µg µg/m3) depending on the area, with peak concentrations close to the center of the city. Previous studies suggest that local traffic emissions, poor ventilation in street canyons and socio-economic factors appear to be the main drivers of the observed differences in air quality. While disentangling their effect was unrealistic due to the lack of dense spatial information so far, this is now possible thanks to this new data set.

This oral communication will present the result of a spatial analysis and modeling of this rich dataset by combining various spatial techniques (geostatistics and machine learning among others), with the aims of (i)

identifying the respective effect of each contributing factor to the total N02 concentration and (ii) predicting at best local N02 concentrations based on the useful factors or proxies at the street level. Our results will also be matched with those obtained from the mechanistic SIRANE model (see https://doi.org/10.1016/j.atmosenv.2011.07.008), a popular choice for modeling the N02 concentration in cities down to the street scale at hourly time intervals. The limitations of the citizen science project for monitoring air quality will also be identified.

Acknowledgment: The author gratefully thanks Pr. Filip Meysman from the Universiteit Antwerpen, Belgium for allowing him to use the CurieuzenAir project data set.

Multi-source Modelling of the Ocean

Ana F. Duarte[1], Roberto Miele[1], Leonardo Azevedo[1], Renato Mendes[2]

[1]CERENA/DER, Instituto Superior Técnico (IST), Universidade de Lisboa, Av. Rovisco Pais 1, 1049-001, Lisbon, Portugal
[2]+ATLANTIC CoLAB, Museu das Comunicações, Rua do Instituto Industrial 16, 1200-225, Lisbon, Portugal; Underwater Systems and Technology Laboratory (LSTS), Associate Laboratory for Energy, Transports and Aerospace (LAETA), Faculdade de Engenharia, Universidade do Porto (FEUP), Porto, Portugal

Keywords: Seismic Oceanography, Modelling the Ocean

Abstract. Oceans have a vital role in the regulation of Earth's temperature, serving as a significant thermal reservoir by absorbing and storing substantial amounts of heat. However, the impact of climate change on the oceans is becoming increasingly severe. Sampling the ocean directly and extensively between scales of 1 to 10 kilometers is challenging resulting in a lack of understanding about the ocean dynamics at these spatial scales. Nevertheless, these spatial scales are crucial for ocean productivity and the facilitation of vertical carbon export down to the deep ocean, which is essential for mitigating the impacts of climate change.

Seismic oceanography is a remote sensing technique, which utilizes multichannel reflection seismic method to produce high-resolution images that enable the examination of fine-scale ocean processes over large distances. The seismic acoustic response is dependent on variations in ocean temperature and salinity, and the resulting seismic images provide a means to track theinterfaces between these thermohaline layers both horizontally and vertically. By interpreting the observed seismic reflections, valuable insights into oceanographic phenomena and mixing processes occurring at different depths within the water column can be gained. However, in order to gain a deeper understanding of these processes, it is necessary to be able to predict the temperature and salinity of the ocean. In this study, we demonstrate the application of a spatial data science framework to assimilate alternative data (e.g., observations, earth observation data and geophysics) to spatially predict ocean temperature and salinity. This methodology is applied to a real case study in the SW margin of Portugal.

Acknowledgment: AFD and LA gratefully acknowledge the support of the CERENA through its project FCT-UIDB/04028/2020. AFD is funded by the Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) through the project PRT/BD/154661/2023. RM was supported by JUNO—Robotic Exploration of Atlantic Waters project—Refª 2021/0008—from the Luso-American Development Foundation.

Multiple-point geostatistics-based spatial downscaling of heavy rainfall fields

Wenyue Zou[1], Pau Wiersma[1], Grégoire Mariethoz[1], Nadav Peleg[1]

[1]*Institute of Earth Surface Dynamics, University of Lausanne, Lausanne, Switzerland*

Keywords: Downscaling rainfall, MPS model, hourly gridded rainfall, Rainfall spatial structure

Abstract. High-resolution gridded rainfall products at sub-daily and 1 km scales are required for hydrological applications in mountainous and urban catchments. As most catchments are ungauged, gridded rainfall data are often obtained through remote sensing. However, their spatial resolution is often too coarse (at 10 km) and requires to be downscaled to a finer resolution. The challenge is not only to downscale the rainfall intensity to a finer scale by considering areal reduction factors, but also the spatial structure of the storm, as both elements are equally important to the assessment of the surface hydrological response. As a result of the lack of training data, the latter is difficult to obtain. Further development of the stochastic multiple-point geostatistics (MPS) framework is presented to downscale long-term satellite-derived gridded rainfall series using only a few years of high-resolution rainfall observations. We demonstrate how the MPS framework can be used to downscale the satellite-derived CMORPH rainfall from 8 to 1 km resolution for 1998-2019, taking the city of Beijing as a case study, with a specific focus on extreme rainfall events. The high-resolution multisource-merged CMPAS dataset (1 km, hourly), available for 2015-2020, is used as the source of the training images. We show that the downscaling framework preserves the observed mean areal rainfall (with a bias of 2%), reproduces the spatial coefficient of variance (with a similar bias), and also retains extreme rainfall at the 99th percentile (with a bias of 6%). Furthermore, it adequately reproduces the rainfall spatial structure, preserving the variograms of the rainfall fields. Similarities were also observed comparing the 2- to 30-year return period maps of the downscaled rainfall extreme with ground observations, with half of the stations (10 out of 19) agreeing on the location and intensity of the extreme rainfall for all return periods. The results indicate that our framework downscales rainfall intensities and preserves the spatial structure well, especially for heavy rainfall,

even if limited data is available. The proposed approach can be applied to other rainfall datasets and regions.

Acknowledgment: China Scholarship Council (CSC. Grant number: 202106040028) SNSF. Grant number: 194649, Rainfall and floods in future cities

Multiplicative cascades as generators of log-normal fields; Case study: Environmental radon

Peter Bossew[1]

[1]*retired*

Keywords: log-normal fields, cascades, multifractals, environmental radon

Abstract. Traditional methods of spatial estimation and mapping, such as kriging and regression against predictors, notoriously suffer from ill-estimation of extremes and anomalies. In some cases this may lead to undesirably non-conservative estimates. Based on some remarkable properties of environmental fields, a methodology is envisaged which may better capture such phenomena.

Approximate spatial log-normality is ubiquitous in fields of environmental variables which represent concentrations of resources and pollutants. The property is scale persistent across a domain, that is, it reproduces across spatial scales and within sub- and supersets. A number of explanations for this property have been proposed; a particularly convincing one appears to be conceptualization of state variables (the concentrations) as outcomes of cascades in the underlying physical realm, that is, repeated stochastic partition of the support according to a partition law. Certain cascades lead intrinsically to persistent asymptotic (with generations) log-normal fields and properties such as multifractality and particular variogram shapes, which are indeed observed in real fields, for example of environmental radon.

In this contribution, the concepts of multiplicative cascades and multifractals will be briefly recalled and their properties demonstrated on real-world radon data, which have analytic properties as can be expected of cascade-generated multifractals.

The objectives of this research are two aspects:

(a) An interpretation and a tentative physical explanation of the cascade nature of the generating process is proposed, in this case the one that results in the diversity of geological units. The emphasis of this contribution is on this aspect.

(b) A possible practical use is that the multifractal framework may allow better capturing of apparent anomalies (as long as they do not result from the observation process), which tend to be underestimated by traditional

geostatistics (the kriging family) as well as by regression approaches against external predictors, including machine learning (which is the best performing family of methods these days, if sufficient predictors are available). Multifractal interpolation and estimation has been proposed by modified kriging-type procedures and by conditional simulation. However, for real-world data these methods do not seem to be at hand easily.

Multivariate Analysis of Multispectral Satellite Images for Potentially Toxic Elements Identification in Abandoned Mine Areas

Barbara Fonseca[1], Joaquim Góis[1], António Guerner Dias[2], Henrique Garcia Pereira[3]

[1]*CERENA - FEUP: Centro de Recursos Naturais e Ambiente - Faculdade de Engenharia da Universidade do Porto*
[2]*Faculty of Sciences of the University of Porto*
[3]*CERENA - IST: Centro de Recursos Naturais e Ambiente - Instituto Superior Técnico*

Keywords: Multivariate Analysis, Abandoned mining areas, tungsten

Abstract. Abandoned mining areas is a worrying environmental problem to the local and scientific community, which has made necessary to implement methodologies capable to monitor the degradation associated to these areas. Sentinel-2 multispectral images use selected wavelengths of the electromagnetic spectrum according to the needs to be investigated. The aim of this work is to investigate the contribution of using multispectral satellite images to assess Potentially Toxic Elements (PTE) in abandoned mining areas.

The methodology used in this work consisted of the sequential overlay, in the QGIS software, of three WMS (Web Map Service) layers, the orthophoto layer, and the layers of multivariate analysis and mineralogical indices existing in a database. The orthophoto is an image that allows to obtain precise and detailed structures on the Earth's surface, being used on this work as a background image. Regarding the multivariate analysis layer, is an image from National Laboratory of Energy and Geology (LNEG) database that provides a multivariate analysis on a mosaic of Sentinel-2 multispectral images, using the Minimum Noise Fraction (MNF) algorithm, based on the trichromatic theory of Young-Helmholtz and Maxwell. As a result, is obtained an image with a spectrum of colors, in which color represents a characteristic given by the user in order to differentiate and to characterize the different areas within the image. The mineralogical indexes image highlights the predominance of different minerals with contrasts, supporting lithological discrimination. The evidence, in a generalized way, of the spectral signatures of different minerals is carried out

through band quotients in a single composite raster, with visualization in the Red, Green and Blue channels (RGB, English acronym) R – clay minerals, G – ferrous minerals, B – iron oxides.

The first procedure regards the overlay of the multivariate analysis image with the orthophoto image, in order to characterize the distinct colors existing on the image and define the potential mining areas, corresponding to the green color. The following procedure consisted on the overlaying of the potential tailing zones in the abandoned mining areas under study and the mineralogical indexes image. As a result, it was shown an almost exact overlap between the mineralogical groups ferrous minerals and the iron oxides with the previous defined potential tailing areas, which corroborates the definition of the tailing areas, once this mineralogical groups are the most common in these type of areas. The last procedure refers to the Principal Component Analysis (PCA) of soil samples collected in the abandoned mining areas under study, in order to verify the intercorrelation between the Potentially Toxic Elements as well as their distribution on different cluster. The PCA analysis revealed that the PTE, usually associated with sulfides, are connected with iron, that is, related to the areas associated with the mineralogical groups ferrous minerals and iron oxides. The Principal Component Analysis corroborates the procedures previously carried out, as well as the definition of mining heap areas in the abandoned mining areas under study. The implementation of this methodology is effective for delimiting mine tailing areas and to assess the PTE associated with.

Acknowledgment: This work was financially supported by: Base Funding - UIDB/04028/2020 and Programmatic Funding - UIDP/04028/2020 of the Research Center for Natural Resources and Environment – CERENA - funded by national funds through the FCT/MCTES (PIDDAC).

Neural Process Regression

Stefania Scheurer[1], Wolfgang Nowak[1]

[1]*Department of Stochastic Simulation and Safety Research for Hydrosystems (LS3), Institute for Modelling Hydraulic and Environmental Systems (IWS), University of Stuttgart*

Abstract. The estimation of parameter values in unsampled locations plays a significant role in various environmental applications. Environmental scientists predict air quality, groundwater contamination, soil pollution, or (hydro-)geological material properties over a wide area solely based on individual measurements at specific locations. These are just a few examples that share a common solution approach: Gaussian Process Regression (GPR), also known as Kriging in the field of geostatistics. GPR is particularly valuable when dealing with spatial data that exhibit autocorrelation. It not only enables the creation of informed predictions at unobserved locations but also quantifies the associated uncertainty.

GPR has the advantage that it operates well in a data-poor environment and at the same time provides an interpretable structure. However, this GPR structure is quite rigid in its learning procedure. Alternatively, regression tasks can also be solved with artificial neural networks (ANNs). They offer a high degree of flexibility in the learning process, but sacrifice interpretability and require a significant amount of training data in return.

Our goal is to combine the advantages of GPR and ANNs in Neural Process Regression (Neural PR), i.e., to create a flexible, interpretable structure that operates in a data-poor environment. We develop neural PR starting from a basic GPR and gradually add flexibilities: (1) The choice of the variogram has a substantial effect on the result of the GPR. Nonetheless, this step is often performed manually and usually requires an understanding of the process. A first step is to learn the variogram without prior assumptions, which provides primary flexibility in modeling. (2) GPR comes with high computational cost when operating in data-rich regimes due to the computationally intensive step of the covariance matrix inversion. We try to avoid this step by learning the inverse covariance matrix in parallel, thus reducing the computational cost. (3) GPR is restricted to two-point distance variograms (since the process is

considered to be Gaussian). We break away from this limitation by introducing not only a distance-based but also a value-based correlation, thus making the step from GPR to neural (general) PR.

Acknowledgment: Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through EXC2075 – 390740016 under Germany's Excellence Strategy.

Oceanic surface currents reconstruction with the Direct Sampling method: a test case on the high frequency radar data set of the Gulf of Naples

Alessandro Comunian[1], Roberta Di Lemma[2], Pierpaolo Falco[3], Mauro Giudici[1], Giuseppe Messina[1], Enrico Zambianchi[2]

[1]*Università degli Studi di Milano*
[2]*Università degli Studi di Napoli Parthenope*
[3]*Università Politecnica delle Marche*

Keywords: coastal high frequency radar, reconstruction, direct sampling, multiple-point geostatistics

Abstract. Measurements of the oceanic surface collected by coastal high frequency (HF) radars can have a crucial role in the study of coastal areas. For example, data collected by these instrument can be very useful for the study of wind induced currents, wave properties and tidal currents, or Lagrangian transport models. Unfortunately, measurement collected by these instruments are sometimes affected by large errors due to a number of reasons, including for example unfavorable weather conditions. Due to these large errors, for some time steps and at some locations, the grid of these measurements contains gaps. In some situation, these gaps can cover and important portion of the area measured by the HF radar.

To fill these gaps, diverse techniques can be used. Previous studies explored the possibility to reconstruct the missing velocity values with Self Organizing Maps (SOM), Open Boundary Model Analysis (OMA) and Data Interpolating Empirical Orthogonal Functions (DINEOF). In particular, some of these studies focused on a data set collected by the University Parthenope with a HF radar system located in the Gulf of Naples, operating since 2004.

In this new study, a multiple-point statistics (MPS) based algorithm is applied for the reconstruction of the aforementioned data set and its results are compared against the results obtained with the previously mentioned methods. In particular, the Direct Sampling (DS) algorithm is used for the reconstruction of missing measurements of velocity due to its flexibility in handling multi-variate data. In fact, current data, with their two components, represent a multi-variate

reconstruction problem that can be easily handled with the DS algorithm by considering a multi-variate training image.

To test the reconstruction performance of the DS, artificial gaps are created in some (complete) maps for some reference time steps, to compare the reconstructed current values against the measured ones. Diverse gap configurations are tested, both with different spatial distribution (randomly scattered or with structured gaps) and different data coverage (from the original data set a fraction of measurements ranging from 10% to 40% is removed).

The results show that the DS method allows to better reconstruct the incomplete velocity maps when the number of missing data remains below the 20%. For higher portions of missing data, other methods, like for example DINEOF, provide better reconstruction, although they require a more complex preliminary setup.

On spatially indexed response models and its applications

Alvaro I. Riquelme[1]

[1]CSIRO, Canberra, ACT, Australia

Keywords: Spatial modelling, Machine learning, Mineral potential mapping, Mineral resources

Abstract. In various branches of geosciences, including remote sensing, mineral exploration, geometallurgical characterization and mineral resource modeling on established mining operations, regionalized input variables are reasonably highly sampled across the spatial domain of interest. This leaves little room for further spatial uncertainty quantification, risk assessment and reconciliation. Despite their critical nature, response functions such as environmental pollution, mineral potential in a specific province, mining throughput and metallurgical recovery are often modeled from a much-restricted subset of testing samples, obtained at certain specific geographic locations, due to the greater acquisition costs. Due to both the multi-dimensional complexity and our limited understanding of these response functions, which can hide further (possibly inaccessible) dependencies to additional variables, we are likely to observe changes in such response models according to the spatial location under analysis. This presentation introduces an ad-hoc mathematical framework on which different response models established at different locations in the area under study can be compared, and the subsequent spatial statistical inference can be worked out. The framework assumes that the response model itself is a regionalized variable, which can be regarded therefore as one among many possible realizations of a random function and, consequently, enabling both the spatial inference of full response models and their uncertainty assessment through simulation methods at unsampled/unexplored locations, in a similar fashion to Gaussian simulation procedures. The parameters required for the adjustment of this response random field are discussed, such as the modeling of the prior probability density in the space of response models and the variography. By calibrating the different local response models using machine learning techniques and subsequently conducting spatial inference, the framework provides a pathway to rigorously integrate machine learning approaches with geostatistics. A realistic case study demonstrating how the

approach is implemented into mineral prospectivity mapping is presented, together with examples on how to extend it to other various fields.

Acknowledgment: The author acknowledges the funding provided by the CSIRO Early Research Career (CERC) Fellowship program.

Physics-informed neural networks for a numerical meshless solution of partial differential equations with moving boundary conditions: an application to groundwater flow in phreatic aquifers.

Daniele Secci[1], Vanessa A. Godoy[2], J. Jaime Gómez-Hernández[1]

[1]*Universitat Politècnica de València*
[2]*Università degli studi di Parma*

Keywords: Aquifer behavior, Unconfined aquifer, Machine learning, Deep learning

Abstract. Consider unconfined aquifers as underground water systems lacking barriers to contain water, where the phreatic surface depends on the spatiotemporal distribution of piezometric heads. In other words, the phreatic surface depends on the solution of the Partial Differential Equation (PDE) that explains how groundwater flows, making it a tough challenge to handle. Think of it as if we were trying to solve a puzzle, but every time we move the special phreatic surface piece, it changes the shapes of the other pieces we have already fit. This back-and-forth makes the puzzle tricky to solve, mirroring the complexity of the nonlinear PDE governing groundwater flow. Traditional Artificial Neural Networks (ANNs) demonstrate proficiency in mapping intricate relationships between input and desired output. Nevertheless, their black-box structure lacks inherent physical interpretability. Dependent solely on data, ANNs may face limitations in accurately modeling and predicting the behavior of complex physical systems, particularly in environments with restricted data, potentially leading to outcomes that are less precise or reliable. In contrast, Physics-Informed Neural Networks (PINNs) address this limitation by incorporating fundamental physical laws into their structure. This integration enhances their ability to provide interpretable insights into complex systems, making PINNs a valuable alternative for applications where understanding the underlying physics is crucial, such as in unconfined aquifers. This study delves into recent PINN developments, covering their formulation and training algorithms, and explores their application in solving the intricate problem of

estimating the phreatic surface and spatiotemporal distribution of piezometric heads in mildly heterogeneous and anisotropic unconfined aquifers.

Acknowledgment: This work was supported by OurMED PRIMA Program project funded by the European Union's Horizon 2020 research and innovation under grant agreement No. 2222.

Pixel-wise Synthetic Hydrological Data Consistent with Climate Reanalysis to Enable Long-term Hydrological Modelling

Loïc Gerber[1], Grégoire Mariéthoz[1]

[1]University of Lausanne

Keywords: Satellite time series, Climate reanalysis, Hydrological modelling

Abstract. Many earth science models rely on satellite imagery as inputs to inform key hydrological variables such as evapotranspiration, soil moisture, and terrestrial water storage. However, there are often periods with significant data gaps, particularly before the year 2000. In this work, we generate synthetic satellite images to bridge such gaps, thereby extending the temporal scope of available data. This effort is crucial for modelling water resources and understanding the impacts of climate change on hydrological processes, especially in regions with sparse data coverage.

Existing synthetic image generation methods proceed by resampling past satellite images and using them to fill observation gaps. However, this can lead to a lack of variability in the generated spatial patterns. We propose an approach that treats individual pixels independently rather than analysing images as cohesive wholes as is currently done in existing methods. It considers per-pixel time series, using a k-Nearest Neighbour algorithm coupled with a process-specific similarity metric to match current climate predictors with the most analogous historical states. This results in the selection of distinct analogues for each pixel, collectively forming "compound synthetic images.”

The proposed method also incorporates random sampling with replacement within the pixel analogue ensembles, introducing a degree of stochasticity. The derived stochastic ensembles are used to estimate the uncertainty associated with each synthetic pixel on any given day, providing valuable insights into the variability and reliability of the generated images.

The method is assessed in the Volta River Basin in West Africa where water resources are critically impacted by climate change. The synthetic images are fed to a spatially distributed hydrological model for calibration and validation, and their quality is assessed by their ability to reproduce historical streamflow

time series. The goal of this testing phase is to improve the generation technique and produce synthetic images that closely approximate unobserved processes.

Preliminary Insights from the Assessing Infectious Disease Outbreaks with Pandemic and Epidemic Risk (APES) Project

Angela Fanelli[1], Emanuele Massaro[3], Alessandro Cescatti[3], Juan Carlos Ciscar[2], Gregoire Dubois[4], Dolores Ibarretta[2], Nicola Riccetti[1], Marine Robuchon[4], Ilaria Capua[4], Wojtek Szewczyk[4]

[1]*European Commission, Joint Research Centre (JRC) Scientific Development Programmes Unit, Ispra, Italy and Economics of Climate Change, Energy and Transport Unit, Seville, Spain*
[2]*European Commission, Joint Research Centre (JRC) Economics of Climate Change, Energy and Transport Unit, Seville, Spain*
[3]*European Commission, Joint Research Centre (JRC) Forests and Bio-Economy Unit, Ispra, Italy*
[4]*European Commission, Joint Research Centre (JRC) Nature Conservation and Observations Unit, Ispra, Italy*
[5]*Johns Hopkins University SAIS Europe, Bologna, Italy*

Keywords: Spatial Epidemiology, Pandemics, Epidemics, Anthropogenic drivers

Abstract. The project "Assessing infectious disease outbreaks with pandemic and epidemic risk" (APES) aims to bridge the research gap in emerging infectious disease events by integrating established and newly available data sources and, further, to analyse spatial and temporal dynamics using advanced statistical methods. The objective is to elucidate the global interplay between environmental and anthropogenic factors and emerging infectious diseases (EIDs) focusing on historical changes from 1975 to 2020. The approach acknowledges several key factors contributing to EIDs, namely land use change, human population, human-forest distance, biodiversity loss, environmental degradation, climate change, species richness, livestock distribution

The initial phase of the APES project, as reported in this paper, utilized a spatial model to investigate the impact of human-induced environmental and climatic changes on the outbreak risk of nine infectious diseases with pandemic and epidemic potential (CCHF, Ebola, Lassa fever, MERS, SARS, MVD, NiV, RVF, and Zika). Spanning from 1975 to 2020, our study examined ten indicators reflecting human impacts on the environment. Utilizing an ensemble of 100 Bayesian

Regression trees (BART) models, each comprising 200 trees and 1,000 iterations, we produced spatial predictions of risk, assessed uncertainties, and determined the influence of the drivers on the likelihood of an outbreak occurrence. We validated model performance using 3-fold cross-validation and computing several performance metrics. Our assessment of risk predictions involved analyzing human outbreak data from 2021 to January 2024 and mapping the risk distribution geographically. We found that environmental degradation, livestock density, and a rapid change in population density exacerbate the likelihood of an outbreak occurring while temperature trend, used as proxies for climate change, showed a complex risk pattern. We also introduced a novel variable that quantifies changes in human-forest proximity, serving as an indicator for wildlife-human interactions. Our findings reveal that outbreaks are more likely to occur as densely populated areas encroach upon forested regions. We identified that roughly 11 % of the globe, predominantly, in Southern Asia, South-eastern Asia, and Sub-Saharan Africa, is at very high and high risk. Furthermore, we introduced a vulnerability index that refines risk estimates taking into account the country capacity to manage zoonotic events, suggesting that well-prepared areas might be less susceptible despite facing a high risk. This research emphasizes the relationship between anthropogenic changes and the risk of infectious disease outbreaks with pandemic and epidemic potential, providing valuable insights for policymaking, despite the inherent uncertainties in the findings.

Acknowledgment: This work was supported by the European Commission's Joint Research Center (JRC) under the Exploratory Research Project "Assessing infectious disease outbreaks with pandemic and epidemic potential" (APES).

Process-based, Surrogate and System Dynamics Modeling for Enhanced Management of Groundwater Resources

Daniele Secci[1], Ali Kerem Saysel[3], Izel Uygur[2], Onur Cem Yoloğlu[2], Andrea Zanini[1], Nadim K Copty[2]

[1]*Department of Engineering and Architecture, University of Parma*
[2]*Institute of Environmental Sciences, Bogazici University*
[3]*Department of Geography, System Dynamics Group, University of Bergen*

Keywords: groundwater flow, numerical modeling, groundwater management

Abstract. Numerical modeling is widely used to aid in the management of groundwater systems. This paper describes three modeling approaches- process-based, surrogate data-driven and system-dynamics modeling- that have been applied to analyze groundwater use in the Konya Closed Basin (KCB), a major agricultural region located in central Türkiye. Water resources in the basin have been under significant stress in recent years due to the expansion of irrigated lands and the switch to more-water demanding crops. The process-based model relies on the latest mathematical representation of the water flow processes to simulate water flow through the saturated/unsaturated zones over the entire basin. The model was calibrated using observed water level data and is shown to be a useful tool for simulating hydrological processes and for scenario analysis, yet it faces two main challenges: the difficulty of defining all input parameters needed to characterize the aquifer system and the relatively large computational effort. The data-driven model aims to find the relation between the dependent variable, the hydraulic head, and climatic and water use patterns without the need to incorporate formal representation of the hydrological processes. It relies on extensive data for training; however, it is highly efficient making it particularly suitable for various purposes such as real-time decision support systems. The system dynamics model examines the water management problem from a broader perspective involving various stakeholders with potentially conflicting interests. The model accounts for the hydrology of the basin as well as a series of endogenous decision making rules and is suitable for scenario analysis and for policy development. While the models differ in model formulation and data requirements, it is shown that these three models developed for the KCB are interdependent. The advantages

and limitations of each modeling approach is discussed. When coupled, these three modeling approaches can potentially lead to enhanced groundwater resource management.

Acknowledgment: This work was developed under the scope of the InTheMED project. InTheMED is part of the PRIMA program supported by the European Union's Horizon 2020 research and innovation program under grant agreement No 1923.

Process-driven asymmetries in spatial data

András Bárdossy[1], Sebastian Hoerning[2]

[1]*Institute for Modelling Hydraulic and Environmental Systems, University of Stuttgart*
[2]*Centre for Natural Gas, The University of Queensland*

Keywords: Asymmetry, non-Gaussian

Abstract. Spatial structures of natural variables are often very complex due to the different physical, chemical, or biological processes which contributed to the emergence of these structures. The spatial dependence of these structures is often non-Gaussian, i.e., asymmetric. This spatial asymmetry can manifest itself in two distinctly different forms; the first being a value-dependent, or order-asymmetry and the second being a directional asymmetry. In this contribution, we are going to introduce both types of asymmetries. Measures, which go beyond classical variograms, to detect and to quantify the magnitude of the different types of asymmetries will be presented. Several different datasets of process simulations and real-life observations will be used to demonstrate the concept. The importance of considering asymmetry in a geostatistical context for simulation, interpolation and uncertainty assessment will be demonstrated through examples.

Quantifying industrial impact on coastal environments: A multivariate geostatistical method to investigate pollution in seafloor and beaches

Marlon Cabrera[1], Carlos Boente[1], Carlos Sierra[2], Germán Flor-Blanco[3], José Luis R. Gallego[4]

[1]E.T.S.I. Minas y Energía de Madrid, Universidad Politécnica de Madrid, C/ Ríos Rosas 21, Madrid, 28003, Spain
[2]European University on Responsible Consumption and Production, Campus de Vegazana, 24006, León, Spain
[3]Geocantabrica Fecyt Group, Department of Geology, University of Oviedo, Arias de Velasco, UNESCO IGCP 725 Forecasting Coastal Change, s/n, 33005 Oviedo, Spain
[4]INDUROT and Environmental Biogeochemistry & Raw Materials Group, Campus de Mieres, Universidad de Oviedo, Mieres, Asturias, Spain

Keywords: Coast, Potentially Toxic Elements, Empirical Bayesian Kriging, Sparse Principal Component Analysis

Abstract. Coastal sedimentary ecosystems are particularly vulnerable to the impacts of human activities stemming from nearby industrial centres. In the case of Gijón Bay (Asturias, Spain), it faces pollution from various sources, including a seaport and extensive metallurgical operations, leading to the accumulation of Potentially Toxic Elements (PTEs) in the sediments of beaches and seafloor.

To address this complex issue, we conducted a comprehensive study involving the collection of 50 sediment samples from various zones within the bay, spanning from the beach system to the nearshore areas. Utilizing advanced multivariate statistical techniques, we examined the geochemical interactions among different elements to pinpoint the sources of pollution and assess the extent of environmental degradation in the bay.

Our methodology involved the integration of classical and booming statistical and geostatistical tools, including Principal Components Analysis (PCA), Sparse Principal Components Analysis (SPCA), CUR Decomposition, Clustering Disjoint Biplot (CDBiplot), and geostatistical Empirical Bayesian Kriging (EBK). But the

compositional nature of geochemical concentrations are scale-invariant objects requiring careful consideration when examining relationships between the elements. For this reason, the analysis are referred to proportions attending to a centered-log-ratio (clr) transformation of data, revealing hot-points of enrichment for main pollutants. All mathematical analysis were cross-validated.

By amalgamating these techniques, we could identify significant variables, discern underlying patterns, and draw valuable insights, conclusively demonstrating the undeniable correlation between industrial activities and sediment contamination in the region.

Based on our findings, we advocate implementing measures to mitigate the emission of PTEs and establishing regular monitoring protocols, particularly in seaport areas. These proactive steps are essential for safeguarding the ecological integrity of Gijón Bay and ensuring the sustainability of its coastal ecosystem in the face of anthropogenic pressures.

Soil Quality Variations in Reclaimed Lignite Mine Areas - Insights from the Western Macedonia Lignite Centre, Greece

Ioannis Kapageridis[1], Tryfon Barmpas[2], Christos Roumpos[3], Francis Pavloudakis[1], Themistoklis Katsalis[2]

[1]*Department of Mineral Resources Engineering, University of Western Macedonia*
[2]*Western Macedonia Lignite Centre, Public Power Corporation of Greece*
[3]*Department of Mining Engineering and Closure Planning, Public Power Corporation of Greece*

Keywords: soil quality, geostatistical analysis, land reclamation, lignite mining

Abstract. Depending on the planned post-mining land uses, the reclamation of surface lignite mining areas requires a thorough understanding of the spatial variations in soil quality parameters. In this study, the spatial dynamics of soil characteristics within the reclaimed surfaces of a lignite mining complex located in the region of Western Macedonia, Greece, were investigated. Using the capabilities of commercial mining planning software, the study covered over 15 years of soil sample collection and analysis, which were conducted in accordance with the environmental permits governing the mining operations.

Two key aspects were investigated in particular: soil fertility and pollution related to elevated concentrations of heavy metals. By examining the spatial distribution of numerous parameters, meaningful connections with the origin of the materials comprising the final soil cover on waste heaps were established. Additionally, the influence of vegetative cover and the age of the waste heap on soil quality variations were determined.

The results of this geostatistical analysis revealed a compelling interplay between soil quality and the geological composition of overburden rocks and lignite-bearing strata. Significant spatial variations emerged among waste heaps of different mines, further confirming the relation between soil characteristics and geological composition. Moreover, the concentrations of heavy metals remain below the screening values referred in various soil quality standards in all the investigated reclaimed mine sites, allowing the selection of new land uses without restrictions.

This study contributes to the broader discourse on geostatistics for environmental applications, offering an innovative perspective on the spatial variations of soil quality parameters in post-mining landscapes. It also underscores the need for site-specific land reclamation trategies. Furthermore, the analysis presented herein highlights the crucial role of knowledge of the relationship between geological origins, vegetative cover, waste heap age, and soil quality towards sustainable land reclamation and environmental stewardship in lignite mining regions.

Spatiotemporal prediction for daily river levels in Ireland using nearest neighbor Gaussian processes

Victor Nagahama[1], Niamh Cahill[1], James Sweeney[2]

[1]*Maynooth University*
[2]*University of Limerick*

Keywords: Space-time, Gaussian process, Water level, Bayesian modelling

Abstract. Obtaining accurate river level predictions are essential for water resource management and implementing flood mitigation strategies. Although several data-driven models can be found in the literature, statistical modelling frameworks to address the spatiotemporal nature of large collected datasets, as well as coherently account for sources of predictive uncertainty, are largely absent. From a nonparametric modelling context, Gaussian Processes (GP) can capture complex space-time interactions and account for predictive uncertainty. However, GPs are computationally expensive and suffer from poor scaling as the number of locations increases due to covariance matrix inversions. To overcome the computational bottleneck, the Nearest Neighbor Gaussian Process (NNGP) introduces a sparse precision matrix providing scalability without inferential compromises. This reduces the computational complexity of matrix inversion from O(N^3) to O(NM^2), where N is the number of locations and M the neighbour set size. We investigate the application of a Bayesian spatiotemporal NNGP model in a rich dataset of daily river levels from stations located in Ireland, consisting of 291 stations over 90 days. The proposed approach allows forecasting for both monitored stations and unobserved locations by using spatial interpolation and estimating the exceedance probabilities of a certain river level at each location.

Stochastic Inverse Modeling of Diclofenac Dynamics in Groundwater

Laura Ceresa[1], Monica Riva[1], Alberto Guadagnini[1]

[1]*Politecnico di Milano*

Keywords: Groundwater contamination, Diclofenac reactive transport in groundwater, Stochastic model calibration and identification, Uncertainty Quantification

Abstract. Groundwater contamination by pharmaceuticals is a potential threat to the integrity of natural ecosystems and human health. Here, we focus on the medical drug diclofenac (Dcf), which poses concerns due to its recalcitrance and documented accumulation in aqueous bodies. Experimental evidences yield controversial results on the real extent of Dcf natural attenuation in groundwater. The role of mechanisms/processes deemed as potential responsible for such attenuation is also debated. For example, biotransformation of Dcf seems to be key in driving reversible attenuation patterns observed in batch experiments under biotic denitrifying redox conditions. In this context, we recently proposed a conceptualization (and ensuing mathematical formulation) rigorously accounting for the full molecular dynamics of Dcf biotransformation in the above-mentioned setting. Our geochemical model is fully consistent with the observed behavior of Dcf and has been calibrated in a stochastic context against a limited amount of available data.

Recently performed column experiments conducted under flow-through conditions suggest that Dcf adsorption is a dominant process driving its fate in anoxic systems hosting a wide range of reducing conditions (i.e., in the presence of redox zonation). Here, we propose a novel and comprehensive hydrogeochemical model that is capable to interpret reactive transport dynamics of Dcf (and major ions involved in speciation in water) in the considered scenario. We then embed the resulting model formulation in a stochastic context. Inverse modeling rests on a Maximum Likelihood approach assisted by modern sensitivity analyses. The latter aim at identifying parameters that are most influential to the calibrated model performance. Similar to what we observed in the context of batch scenarios, preliminary calibration results show severe difficulties in the simultaneous estimation of all uncertain model

parameters. This is related to (i) the model complexity (i.e., high level of model parameterization), (ii) the likely cross-dependence among uncertain parameters, and (iii) the insufficient amount (and/or quality) of available information. Overcoming these issues is generally challenging, especially when dealing with ill-posed problems, a situation that is commonly faced under the above circumstances. As such, (a) development of comprehensive modeling frameworks and (b) establishment of rigorous workflows enabling one to estimate uncertain model parameters while rigorously quantifying estimation uncertainty, and (c) demonstration of the potential and applicability of such workflows to realistic scenarios are critical research tasks. Our study is framed in this context and successfully demonstrates the reliability of our modeling framework when transferred from batch to flow-through scenarios effectively incorporating a wide range of physical and/or biochemical processes. Our results can then be employed in the context of stochastic simulation techniques conditional on available observations.

Acknowledgment: Funding from the European Union Next-Generation EU (National Recovery and Resilience Plan – NRRP, Mission 4, Component 2, Investment 1.3 – D.D. 1243 2/8/2022, PE0000005) in the context of the RETURN (Multi risk science for resilient communities under a changing climate) Extended Partnership is acknowledged.

Stochastic Simulation of Karst Network Skeleton Patterns with KarstNSim

Augustin Gouy[1], Pauline Collon[1], Vincent Bailly-Comte[2], Philippe Landrein[3]

[1]*University of Lorraine, CNRS, Georessources*
[2]*BRGM, University of Montpellier. G-eau, INRAE, CIRAD, IRD, AgroParisTech, Supagro, BRGM.*
[3]*Andra*

Keywords: Karst, Simulation, Skeleton, Network

Abstract. In karst aquifers, groundwater flow heavily depends on underground conduits called karst networks. To model this flow spatially accurately, explicit representation of these conduits is necesary through distributed approaches. However, the thickness of conduits varies greatly, making their direct exploration and mapping with methods like seismic reflection or electrical resistivity quite challenging. Stochastic simulation of discrete karst networks is a helpful approach to overcome the large uncertainties on conduit position and geometry. Yet, among existing methods in the literature, many are limited in capturing the full range of possible cave patterns determined by the specific conditions of karstification, including branchwork, anastomotic and angular patterns.

To overcome this issue, we propose to use the newly developed karst simulation method KarstNSim, which solves a shortest path problem between sinks and springs – respectively the inlets and outlets of the network – with the use of an anisotropic cost function defined on an unstructured mesh conformal to geological and structural heterogeneities. This cost function represents the physico-chemical processes that govern speleogenesis – such as erosion and chemical weathering – providing simplified control over the morphometry of the generated networks. The method encompasses geological parameters such as inception surfaces, fractures, permeability, ghost-rocks and solubility of layers, along with considering the hydrological context of recharge by assigning relative weights to the inlets.

We will explore the variability of morphologies that the method can reproduce through different tests based on simple (hydro)geological models with varying

geology and boundary conditions, represented by different parameters for the cost function.

Acknowledgment: This work was performed in the frame of a research agreement between University of Lorraine, BRGM and ANDRA , and as a part of the "Calcaires du Barrois" program led by ANDRA.

Stochastic simulation of precipitation combining heavy-tails and long-range dependence

Panayiotis Dimitriadis[1], Theano Iliopoulou[1], Demetris Koutsoyiannis[1]

[1]*National Technical University of Athens*

Keywords: precipitation, stochastic simulation, heavy tails, long-range dependence

Abstract. The precipitation process is well-known to exhibit both heavy-tail and long-range dependence, especially in climatic scales. When combining both in stochastic simulation, several modelling issues emerge, such as the explicit and genuine preservation of high-order moments along with the intermittent and dependence structure of the process. Such difficulties cannot be fully supported by the established models in the literature, where most can preserve explicitly the marginal structure and only approximately the dependence one, leading to some-degree underestimation of the uncertainty of precipitation. In this work, we present a new stochastic scheme based on the Hurst-Kolmogorov dynamics through the preservation of cumulants, rather than of the classical moments, which tackle some of the above issues by also permitting lower computational burden, and we apply it to a long-length precipitation timeseries in Greece to test its advantages and limitations as well as its simplicity vs. its strength to capture a wide range of trends and clusters often observed in rainfall timeseries.

Acknowledgment: This research is funded in the context of the project “Development of Stochastic Methods for Extremes (ASMA): identification and simulation of dependence structures of extreme hydrological events” (MIS 5049175).

The effects of viewing angle and point spread function on downscaling monthly nighttime light imagery at the city-level

Nikolaos Tziokas[1], Ce Zhang[2], Peter M. Atkinson[1]

[1]*Lancaster University*
[2]*University of Bristol*

Keywords: downscaling, satellite nighttime light imagery, random forest area-to-point Kriging, geostatistics

Abstract. Since the 1990s, nighttime light (NTL) remote sensing data have been widely used to research socioeconomic issues, primarily at a coarse spatial scale (e.g., a country- or county-scale). Given that over half of the world's population lives in cities, these places are known as "hotspots of human economic activities". This means that scientists and decision-makers have a rare chance to use NTL imaging as a proxy for socioeconomic factors at city scale. The main benefits of satellite NTL imaging are its wide geographic coverage, freely available, and cross-national comparability regardless of statistical capacity. Despite the advantages of NTL, the accuracy of applying its brightness to the assessment of human activities is restricted by its coarse spatial resolution and severe blooming effect (also known as the point spread function, or PSF). A unique and intrinsic phenomenon in NTL imagery is called "blooming" which occurs when several pixels in urban regions are illuminated by lights over the neighboring urban pixels. Such blooming effects bring major challenges in utilizing nighttime lights to evaluate socioeconomic applications, and socioeconomic elements tend to be underestimated in urban areas. Furthermore, discrepancies occur in observations made over the same region due to the angular composites being spatially linked to phenomena such as the overglow. To address the coarse spatial resolution, the blooming effect and the impact of the viewing angle on the overglow effect, this paper aimed to downscale the monthly nocturnal satellite imagery by accounting for the PSF and the viewing angle using a geostatistical solution. Random Forest Area-to-Point Kriging (RFATPK) is adopted with PSF explicitly incorporated in NTL downscaling process, and the viewing angle is considered to establish any effect on the PSF. The studies carried out in downscaling two megacities (Delhi, India and Los Angeles, U.S.), with results showing that downscaling can be

significantly improved by taking the PSF impact into account using a geostatistical technique. The PSF could vary in size based on the viewing angle, suggesting that the latter has an impact on the PSF. Besides, the downscaled images following blooming effect mitigation exhibit larger spatial variability than the original NTL data. More specifically, the width of the PSF (measured by pixels) for Delhi, was 0.8 and 1 for the near-nadir and off-nadir, respectively. Whereas Los Angeles, was 0.9 and 1 for the near-and-off nadir angles, respectively.

The role of fruit trees and orchards plantations in carbon sequestration in different climatic conditions in South Africa

Phumudzo Tharaga[1], Lindsay Banda[1], Nompumelelo Mobe[2]

[1]University of the Free State
[2]Council for Scientific and Industrial Research

Keywords: carbon sequestration, Eddy covariance, Eulerian, Fluxes

Abstract. The removal of natural forests for agricultural purposes has led to a reduction in the rate of carbon sequestration or net ecosystem exchange. However, the establishment of orchards in the tropical, Mediterranean, and temperate climate regions of South Africa has led to improvement in the reduction of net ecosystem exchange, which was caused by the clearing of vegetation. The study used data from irrigated fruit tree orchards and rainfed orchards for several years. The deciduous fruit trees, evergreen trees, citrus, and tropical fruits such as avocados, Pinus Pinaster, and Pinus radiata were also used to compare the sequestration by natural deciduous alien invasive species introduced in the country in South Africa for commercial reasons and food Security. The net ecosystem exchange was measured using the open path and closed path eddy covariance systems deployed in orchards and plantations in the following towns in South Africa (George, Malelane, Ceres, Wolsey Ficksburg). Eddy flux data were treated using the eddy-covariance data acquisition and processing software package Eddy Pro Software. The footprint probability distribution function of the measured fluxes was estimated with footprint models.

However, some models' complexity and large computational demand restrict their practical applicability. The multi-site EC datasets, using a somewhat simplified footprint model that retains the ability to discern spatial footprints, are advantageous. In this study, we used the Simple Analytical Footprint model on Eulerian coordinates for scalar Flux (SAFE-F), to compute the footprint probability distribution function. This analytical footprint model considers atmospheric stability and uses the wind velocity power law above the canopy, allowing it to apply to various atmospheric conditions. The SAFE-F model input includes the EC sensor height (hm), canopy height (hc), roughness length (z0),

friction velocity (u∗), u∗ threshold for EC flux calculation (u*th), wind direction (WD), wind speed (u), standard deviation of lateral wind speed (σv), and sensible and latent heat fluxes measured at the EC sensor height. The input meteorological variables were measured in 2006, generally considered a normal weather year for most sites. Missing flux and meteorological data were filled using the gap-filling method. Most deciduous species maintained the same ecosystem water use efficiency, the rate trees take up carbon and release water. Soil water rehydration in autumn and a significant drop in vapour pressure deficit led to increased water use efficiency in all species except for evergreen citrus, which only moderately increased. Eddy-covariance data showed that the orchard ecosystem was a substantial carbon sink during the early summer and less towards the beginning of autumn when the trees start losing leaves. This carbon sequestration decreased gradually with the onset of the dry season, although a positive carbon balance was maintained until the first rains in mid-autumn. his CO2 uptake contribution follows an expected seasonal trend in accordance with soil water availability and assimilation rate limitation. The temperature sensitivity of the early autumn net ecosystem exchange was more than 12 gC m−2 ◦C−1, whereas it was close to 5 gC m−2 ◦C−1 in the late autumn. The cumulative growth primary production was practically independent of the temperature in early autumn for all fruit trees, especially in temperate and arid regions.

Acknowledgment: Thuthuka- NRF; Water Research Commission

Towards an equitable and multisectoral water management: The Bode catchment in central Germany as a showcase

Amir Rouhani[1], Yao Li[1], Michael Rode[1], Seifeddine Jomaa[1]

[1]*Department of Aquatic Ecosystem Analysis and Management, Helmholtz Centre for Environmental Research - UFZ, Magdeburg 39114, Germany*

Keywords: Water managment, Hydrology, Water quality, Groundwater

Abstract. Ensuring access to freshwater in sufficient quantity and quality for diverse sectors while maintaining ecological functions is increasingly challenging, especially under global warming conditions. This situation calls for the urgent need for water management to shift from a mono-sectoral water management approach based on trade-offs to a more balanced multisectoral management that considers the requirements of all stakeholders. To this end, integrating real-time monitoring, information gathered from different sources, such as camera and remote sensing and modelling results into a digital twin system is proposed as an innovative approach to support decision-making. Developing enhanced and multisectoral water management strategies at a river basin scale requires detailed knowledge of hydrological and biogeochemical processes and their ecological targets towards sustainable water use. The Bode River basin, with its long-term monitoring data and infrastructure, can serve as a unique showcase for multisectoral and fair water distribution, supporting domestic use, food production, and ecosystem needs. The Bode catchment has been affected in recent years by a prolonged drought, leading to severe deforestation and adversely affecting both water quantity and quality. The accumulated scientific knowledge during drought can be considered as a unique and natural scenario of extreme conditions. This gained knowledge is crucial to define the ecological boundary conditions of different ecosystem services in the Bode catchment.

In this study, we combined the mHM-Nitrate model, which assesses hydrological processes and water quality, with the MODFLOW groundwater model. This integration allowed us to simulate groundwater levels and discharge, across the entire basin. To achieve this, we replaced the baseflow component of the mHM-Nitrate model with the River Package of MODFLOW groundwater model. Thanks to this modelling framework, we can comprehensively evaluate the

dynamics of surface and subsurface water interactions, especially under low-flow conditions. The coupled modelling approach is used to investigate the relationship between groundwater and rivers and to capture the complexity of the hydrological and biogeochemical interactions. A digital twin platform for the Bode catchment is being developed to enable the representation and integration of data from various sources, including model simulation results, in-situ measurements, and remote sensing data, ensuring timely and continouse information for better decision making. The platform will be updated daily and further developed for forecasting capabilities considering different weather-forcing data.

Acknowledgment: This study is supported by the OurMED project (https://www.ourmed.eu/), which is part of the Partnership for Research and Innovation in the Mediterranean Area (PRIMA) Programme supported by the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No 2222.

Tracing the Spatial Origins and Spread of Omicron Lineages in South Africa

Graeme Dor[1], Houriiyah Tegally[1], Eduan Wilkinson[1], Richard J. Lessells[2], Darren P. Martin[3], Tulio de Oliveira[1]

[1]*Centre for Epidemic Response and Innovation (CERI), School for Data Science and Computational Thinking, Stellenbosch University*
[2]*KwaZulu-Natal Research Innovation and Sequencing Platform (KRISP), Nelson R. Mandela School of Medicine, University of KwaZulu-Natal*
[3]*Division of Computational Biology, Department of Integrative Biomedical Sciences, Institute of Infectious Diseases and Molecular Medicine, University of Cape Town*

Keywords: SARS-CoV-2, Omicron, BA lineages, phylogeography

Abstract. Since November 2021, five genetically distinct Omicron lineages (named BA.1 through BA.5) have emerged in southern Africa, with four of the lineages (BA.1, BA.2, BA.4 and BA.5) spreading throughout the world and collectively dominating global SARS-CoV-2 populations. During 2023, BA.2.86, another highly divergent BA.2 lineage was detected and rose to prominence globally. Although BA.2.86 was first discovered in Israel and Denmark in July 2023, the high diversity of sampled southern African BA.26 sequences in July and August of 2023 indicate that, as with the progenitors of all the BA lineages, it too likely emerged in southern Africa. Here we integrate spatial mapping and advanced Bayesian phylogeographic techniques to infer the geographical origins and dispersion patterns of the respective BA lineages. Such an approach is critical in determining the presence, or lack thereof, of shared dispersal patterns and/or a common geographical origin for these highly divergent and globally significant lineages. We find that South Africa's Gauteng province, situated in the northeastern part of the country, is the most likely origin of BA.1, BA.3, BA.4, BA.5, and BA.2.86 and is the local dispersal hub for all six of the lineages. It remains unclear whether Gauteng province is the true origin of these lineages or whether, given its high population density and prominent transport hub, it is an amplification zone of highly transmissible new variants. It remains highly plausible therefore that one or more less well sampled regions of southern Africa are the actual source(s) of the lineages. Although the degree of SARS-CoV-2 sampling across southern Africa in the months bounding the emergence of the

analysed lineages is insufficient to pinpoint the precise geographical origins of the lineages at sub-provincial or district resolution, our results strongly support the hypothesis that an epidemiologically unique reservoir of rapidly evolving SARS-CoV-2 variants exists somewhere in southern Africa: likely demarcated by an area encompassing the north/north east of South Africa, southern Zimbabwe, southern Mozambique and the Kingdom of Eswatini. The implication of these results underscores the urgent need for broader genomic surveillance across the region, which could yield critical data for preempting the emergence of highly divergent, and potentially globally consequential, SARS-CoV-2 variants.

Training surrogate models using input dimension reduction and Bayesian active learning techniques for inverse modelling in heterogeneous media applications

Maria Fernanda Morales Oreamuno[1], Sergey Oladyshkin[1], Wolfgang Nowak[1]

[1]*Department of Stochastic Simulation and Safety Research for Hydrosystems, Institute for Modelling Hydraulic and Environmental Systems, University of Stuttgart*

Abstract. Groundwater models often require the calibration of numerous parameters to accurately represent various subsurface processes. Bayesian inverse modeling, using observation data to update prior knowledge on parameter distributions, is an insightful approach for calibration. However, this method becomes computationally prohibitive when dealing with high-dimensional models due to factors like heterogeneous media, multiple processes, and uncertainties in contaminant sources. In such cases, running computationally expensive models numerous times becomes impractical. To address these challenges, surrogate models offer an efficient concept. Surrogate models facilitate the exploration of parameter spaces and the identification of optimal parameter sets reducing the need for repetitive full model outcomes. These models are trained using input-output data to emulate the response of the full complexity model (FCM). When dealing with Bayesian inverse problems, selecting appropriate training points for surrogates becomes important. The challenge lies in efficiently training surrogates for high-dimensional problems, which involves addressing increased computational costs during training and determining an optimal, and computationally feasible, number of training points.

In particular, for heterogeneous media applications, Input-Dimension Reduction (IDR) methods, such as the Karhunen-Loéve expansion (KLE), are applied to reduce the dimensionality of the input parameter field. The reduced input is associated to a smaller percentage of the total input variance and thereby introduces a truncation error when training the surrogate model. Alternative IDR

methods such as Autoencoders also introduce errors into the training process, however they are not as easily quantifiable.

To address these challenges associated to high-input dimensional problems, we explore Bayesian Active Learning (BAL) techniques for the optimal training of surrogate models. Our goal is to analyze the effectiveness of different BAL criteria in conjunction with IDR techniques for surrogate models in Bayesian inverse modeling. To this end, we use Gaussian Process Emulators (GPEs) as surrogates, to provide a distribution over the model outputs and to enable the use the Bayesian3 Active Learning (B3AL) approach, as proposed in Oladyshkin et al (2020). B3AL proposes different Bayesian and information-theoretic criteria to select training points, considering the predictive uncertainty of the surrogate. We employ an advective-diffusive groundwater transport model under heterogeneous media as a case study, comparing different B3AL criteria and evaluating the resulting surrogates in terms of posterior predictive accuracy. The ultimate goal is to optimize computational resources, measured in terms of FCM runs, while maintaining strong performance in Bayesian inverse applications for heterogeneous media problems.

Acknowledgment: The authors would like to acknowledge the Federal Company of Radioactive Waste Disposal (BGE) within the URS research cluster, including all participating projects (research contract: STAFuE-21-4-Klei) and the German Research Foundation (DFG) for financial support of the project within the Cluster of Excellence "Data-Integrated Simulation Science" (EXC 2075 – 390740016)

Uncertainty Assessment by using Multi-Solution Geostatistical Seismic Inversion

Amilcar Soares[1], João Lucas Alves[2], João Felipe Costa[2]

[1]*CERENA, Instituto Superior Técnico, Universidade de Lisboa, Portugal*
[2]*Universidade Federal do Rio Grande do Sul, Brazil*

Keywords: Multi-Solution Geostatistical Seismic Inversion, Niching Genetic Algorithm

Abstract. Seismic inversion methods, like most methods for solving an inverse problem, aim to derive a solution – a model characterizing petrophysical properties. This inverted model targets the reproduction of the observed data – real seismic. Inherent to this class of inversion problems is their ill-posed nature, lacking a unique solution. The major limitation of most existing seismic inversion methods - Deterministic, Bayesian, and Geostatistical Seismic inversion methods - is that they offer a unique solution, with a limited uncertainty around the obtained result.

In this study, a new approach is presented based on the Niching Genetic Algorithms method, to allow a diverse set of optimal solutions in a multi-modal space, thus overcoming the major limitation of seismic inversion methods, which is access to the uncertainty of the final models. In the new seismic inversion approach proposed in this study, in each iteration of the optimization process, niches of petrophysical properties models are calculated with a Machine Learning Clustering method based on a distance measure of the similarity between synthetic seismic models in a Multidimensional scaling (MDS) space. For the set of niches, the evolutionary process (Niching Genetic Algorithm) will produce different solutions, but close to the observable data, the real seismic, giving rise to a multi-solution seismic inversion methodology. The new method is illustrated in a case study with the objective of risk assessment at early stages of exploration of a complex geological target.

Uncertainty evaluation of simulated extreme floods in Switzerland

Eleni Kritidou[1], Martina Kauzlaric[2], Marc Vis[1], Maria Staudinger[1], Jan Seibert[1], Daniel Viviroli[1]

[1]*University of Zurich, Zurich, Switzerland*
[2]*Institute of Geography and Oeschger Centre for Climate Change Research, University of Bern, Switzerland*

Abstract. Floods are high-impact natural hazards, causing loss of human lives, environmental deterioration and damage to infrastructure. Reliable flood estimates are necessary to reduce their impact on society and develop effective flood risk management.

Although traditional flood estimation methods rely on historical records of streamflow and, therefore, on observed hydrographs, the relatively short length of the observations limits the robustness of common extrapolation techniques. In addition, missing data and assumptions about the antecedent catchment conditions lead to insufficient representations of the underlying flood processes. Therefore, estimations about rare floods become highly unce

To overcome these limitations, a combination of stochastic precipitation models with hydrological models is often adopted. Here, we use a framework based on continuous simulations with a hydrometeorological modeling chain to estimate rare floods in several large Swiss catchments (> 450 km²) and further explore the associated uncertainties. The first element of the modeling chain is the multi-site stochastic weather generator GWEX, focusing on generating extremely high precipitation events. Then, the bucket-type hydrological model HBV is adopted to simulate the discharge time series. Lastly, the RS Minerve (RSM) hydrologic routing model is used to implement simplified representations of river channel hydraulics and floodplain inundations. After examining the sensitivity of the simulations to altered weather generator inputs, we further explore the impact of the hydrological model on the robustness of the simulated flood estimates. A remarkable part of uncertainties in modeled discharge is expected to arise from the hydrological model. Therefore, an experimental set-up that focuses on HBV's potentially most sensitive elements is adopted, while GWEX and RSM remain

unchanged. First, a model configuration with a different response function is used. Second, the modeled precipitation lapse rates, adjusting mean catchment precipitation input from the weather generator and then distributing it over the different elevation zones are varied within 0–10%.

This model chain advances our understanding of uncertainties in flood estimates associated with changes in hydrological modeling and aids in identifying geographic regions of higher sensitivity. The results will serve as a basis for follow-up studies related to hazard assessment, safety control planning, and hydraulic engineering projects.

Uncertainty Quantification of Averages and Totals of Soil Spatial Predictions

Gerard Heuvelink[1], Alexandre Wadoux[2]

[1]*Wageningen University and ISRIC - World Soil Information*
[2]*INRAE*

Keywords: Digital Soil Mapping, Change of Support, Block Kriging, Monte Carlo Integration

Abstract. Digital soil mappers take pride in routinely quantifying the uncertainty of maps produced, by computing quantiles of the predictive distributions and prediction intervals. Quantification of the prediction uncertainty, derived from the kriging variance in geostatistical mapping, or through methods like quantile regression forest in machine learning, is well-established. However, this uncertainty pertains to point support predictions, i.e. predictions that have the same support as the observations used for model training and interpolation. Yet, many users seek information about spatial averages or totals of soil properties, such as the mean clay content in a field or the total soil organic carbon stock in a region. While deriving predictions of spatial averages and totals from point predictions is straightforward, determining the associated uncertainty is challenging, due to spatial autocorrelation of prediction errors. Block kriging addresses this in geostatistical modelling, but for soil property maps created using machine learning algorithms, the solution is less obvious.

In this presentation we propose a new model-based approach that sidesteps the numerical complexity of block kriging, making it feasible for large-scale studies employing machine learning for soil mapping. Our approach uses Monte Carlo integration to derive uncertainty of spatial averages or totals from point support prediction errors. In a first case study, we employed block kriging and show that uncertainty in predicted topsoil organic carbon in France decreases as the spatial support increases. We illustrate the broad applicability of the Monte Carlo integration method with a non-soil example in a second case study. We estimated the uncertainty of spatial aggregates from a machine learning map of above-ground biomass in Western Africa, finding it to be small due to weak spatial autocorrelation of standardized map errors.

This work introduces a scalable method that is of key importance to studies that aim to evaluate the statistical significance of predicted differences in aggregated soil properties and other environmental variables between regions or over time.

Uncertainty Quantification of Rock and Fluid Properties in Near Surface Geophysics Studies

Dario Grana[1], Andrew Parsekian[1], Ang Li[1], Emmanuel Oladeji[1]

[1]University of Wyoming

Keywords: Inverse theory, Near surface geophysics, Rock physics, Bayesian methods

Abstract. Near surface geophysics methods are often used to assess the variability of rock and fluid properties in the critical zone. The goal of studying the near surface environment is to understand the structure and the physical processes of the subsurface layers from the top of the soil to the bottom of the groundwater. We focus on mountain watersheds and investigate the near surface on hillslopes where snow melt flows through the soil and porous rocks to ultimately recharge the underground aquifers. The structure of the critical zone can be described using models of continuous rock and fluid properties, such as porosity and water saturation, as well as categorical water-relevant classifications of geologic substrates often referred to as hydrofacies.

Geophysical methods such as seismic refraction and electrical resistivity tomography measure geophysical properties, such as seismic velocity and electrical resistivity. Rock physics models are then used to convert these variables into rock and fluid properties such as porosity and water saturation or their hydrofacies classification. Rock and fluid properties are generally obtained as a result of an inverse problem where the rock physics model is calibrated using core measurements and inverse theory methods are applied to estimate the spatial distribution of the properties of interest. Hydrofacies are then obtained by applying semi-supervised clustering methods either in the geophysical or in the petrophysical property domain. The mathematical inverse and classification problems are often formulated in a Bayesian framework, where the solution is the posterior probability distribution of the rock and fluid properties conditioned to the measured geophysical data. The results allow predicting the most likely model as well as assessing the uncertainty of model predictions. Several sources of uncertainty affect the accuracy and precision of the posterior estimation, including noise in the data, prior knowledge of the model, assumptions of unknown hyperparameters, approximations of the

physical models, and natural variability and heterogeneity of the subsurface properties.

In the proposed study, the uncertainty is estimated for each modeling step and propagated from the input data to the posterior distribution. Seismic velocity and electric resistivity inversion processes are mostly affected by two types of uncertainty: a) the lack of accurate knowledge of the initial model and the inversion regularization parameters adopted to constrain the inversion and reduce the non-uniqueness of the solution, and b) the limited precision of the measured data due to measurement and data processing errors. Consequently, the accuracy and precision of the results of the rock and fluid property inversion and hydrofacies classification depend on the uncertainty in the predictions of seismic velocity and electric resistivity as well as the accuracy of the rock physics model approximation. We demonstrate the proposed approach to inversion, classification, and uncertainty quantification on a real geophysical dataset measured along a section of mountain hillslope near Laramie, Wyoming, USA. The model predictions and the associated uncertainty models are crucial for stakeholders to make more informed decisions on water management.

Utilization of multispectral satellite image to define potential tailing areas in abandoned mining areas

Barbara Fonseca[1], Joaquim Góis[1], António Guerner Dias[2], Henrique Garcia Pereira[3]

[1]CERENA - FEUP: Centro de Recursos Naturais e Ambiente - FEUP
[2]Faculty of Sciences of the University of Porto
[3]CERENA - IST: Centro de Recursos Naturais e Ambiente - IST

Keywords: abandoned mining areas, multispectral images, Sentinel-2

Abstract. Multispectral images, using different wavelengths of the electromagnetic spectrum, make it possible to obtain additional information about a given object. Depending on the objectives of the study to be carried out, wavelengths can be selected using filters or isolated through the instruments sensitive to specific wavelengths. The use of multispectral images constitutes a powerful analysis tool in Earth Sciences, allowing its use in studies involving different soil occupations, lithological surface characterization, mineralogical mapping, the identification of contaminated areas, the recognition of geological structures, the identification of geothermal systems, etc, once is .

This work is part of the first stage of an ongoing doctoral project, entitled “Geomathematical assessment of environmental degradation caused by potentially toxic elements resulting from abandoned mining operations. Study cases inserted in the Iberian Metallogenetic Tin-Tungsten Province”, and aim to investigate the contribution of using multispectral satellite images, once in old abandoned mining farms, it is sometimes difficult, due to the growth of the surrounding flora, to identify the geographical locations of the heaps.

This work consisted on the analysis, in the QGIS software, of a multivariate analysis satellite image, which is a data set referring to exploratory data analysis applied to a mosaic of multispectral images, from Sentinel-2 satellite, using the Minimum Noise Fraction (MNF) algorithm. The use of this layer allows to define and delimit the existence of different zones in the target areas of study, and can subsequently classify them. The multivariate analysis satellite image is based on the trichromatic theory of Young-Helmholtz and Maxwell, which consists on viewing a given image in three colors, that, applying to the study, this satellite

image allows the visualization of the mining areas under study through a color system, RGB (Red, Green and Blue). The RGB system does not limit the color spectrum to just the three previously mentioned, but rather to a spectroscopic range sensitive to the human eye, thus providing a color gradation in the areas under study. Each color is assigned a certain attribute, in order to distinguish and characterize the target areas of study. In this way, for each study area a color map is obtained which, depending on its characteristics, is categorized according to the objectives of interest for this study. In this study it was possible to identify urbanization areas, forest areas and potential mining waste areas, which are the areas of greatest interest and main focus of this research. The potential tailing zones were identified with the green color in the Multivariate Analysis satellite image, with subsequent investigation on the abandoned mining areas under study.

The implementation of a multivariate analysis satellite image is effective as a preparatory step regarding the delimitation of the potential tailing areas, once the analysis of this information allows a primary identification and characterization of the abandoned mining areas under study.

Acknowledgment: This work was financially supported by: Base Funding - UIDB/04028/2020 and Programmatic Funding - UIDP/04028/2020 of the Research Center for Natural Resources and Environment – CERENA - funded by national funds through the FCT/MCTES (PIDDAC).

X-Ray Flourescence Spectrometry (pXRF) as a tool for high resolution mapping of terrestrial pollution at legacy mine sites in East Greenland.

Christian Frigaard Rasmussen[1], Jens Søndergaard[1], Anders Mosbech[1], Christian Juncher Jørgensen[1], Arnold Downey[2]

[1]Aarhus University, Department of Ecoscience, Arctic Environment
[2]Université de Montréal, Montréal (Québec) Canada

Keywords: pXRF, Mining, Environmental investigations

Abstract. In Greenland, legacy mines provide testimony to how pollution still affects the sensitive Arctic environment. These legacy mines serve as valuable study sites that can improve future predictions on environmental consequences of mining operations in Greenland and other areas in the Arctic. Environmental studies at legacy mine sites in Greenland have previously focused on the leaching of pollutants into the marine environment and little is known about the spatial distribution of pollution in the terrestrial environment at these sites.

In the current study, we present preliminary data from an environmental survey at the Blyklippen legacy mine. Blyklippen was a lead and zinc mine in East Greenland that operated between 1956-1963. Mining operations caused substantial pollution of lead (Pb), zinc (Zn), and other heavy minerals such as cadmium (Cd), barium (Ba) and copper (Cu), still measurable today. The primary sources of pollution today are the tailings storage facility and remains of ore concentrate spills along the haul road and at the quay areas at the harbor. Pollution is dispersed from these sites into the surrounding environment by wind and water. A total number of 995 discrete sediment samples were collected over 10 days at the Blyklippen mine site and surrounding area of Mestersvig from 178 sampling locations at depth intervals of 5 cm.

The aim of the current study is to investigate and map the spatial distribution of heavy metals in the environment surrounding the Blyklippen legacy mine using both field measurements by portable X-Ray Fluorescence spectrometry (pXRF) on depth specific sediment samples at in situ conditions and laboratory measurements on freeze-dried samples to investigate the effect of soil moisture on the accuracy of the field screening. Field measurements were conducted

using short measurement times of 5-10 seconds, whereas laboratory measurement times were 180 seconds.

Our results demonstrate the effectiveness of pXRF as a field measurement tool for identifying the spatial delineation of soil pollution by comparing in situ measurements against pre-defined natural background values for heavy metals such as Pb and Zn. Comparison between Pb and Zn concentrations measured in the field versus in the laboratory on freeze-dried samples showed a good agreement for mineral soils. On organic and/or wet samples, field concentrations were underestimated for some elements.

Overall, the approach shows that a fast and cost-effective large-scale field survey at legacy mines is obtainable using pXRF, enabling an effective identification of pollution ‘hotspots’ directly in the field. In combination with geostatistical mapping, the approach can improve the overall accuracy of environmental monitoring and mapping of pollution with enhanced environmental protection at both legacy, recent and future mines.

Author Index

A

B

C

D

J

K

L

M

N

O

P

Q

R

S

T

U

V

W

Y

Z