Workshop Programme

Information Theory & The Earth Sciences

Schneefernerhaus | Zugspitze

April 25th-27th, 2016

Workshop Organizing Committee:

Uwe Ehret (KIT Karlsruhe Institute Of Technology)

Hoshin Gupta (University of Arizona)

Grey Nearing (NASA Goddard)

Florian Wellmann[1] (RWTH Aachen University)

Rohini Kumar (Helmholtz Center for Environmental Research - UFZ)

Ben Ruddell (Arizona State University)

Steven Weijs (University of British Columbia)

Beth Jackson (Victoria University Wellington)

Gab Abramowitz (University of New South Wales)

Support by DFG is greatly appreciated!

[1]Editor of Workshop Programme

Contents

1 Workshop overview

1.1 About the workshop

There is a growing understanding in both basic physics and the applied physical sciences that information is a fundamental aspect of the nature and workings of reality, and is also fundamental to our understanding of reality (Knuth, 2004, 2010). This realization has resulted in a resurgence of interest in the topic of Information Theory (hereafter "IT") and its implementation in the Earth and Environmental Sciences.
Information theoretic analyses are essentially general in nature and can be applied to all parts of the scientific endeavor: complex systems (Tononi, 2011), models of those systems (Nearing and Gupta, 2015), observational data (e.g., Datcu et al., 1998), and the synthesis of these things (MacKay, 2003). Rooted firmly in mathematics and statistical theory, IT provides a compelling basis for expanding upon tools and methods that typically make simplifying assumptions of linearity and Gaussianity to address problems of inference. Because of this, IT has the potential to help us understand emergent behavior of complex systems in ways that more traditional analyses cannot (e.g., synergistic information, extreme non-linearity, networks etc.). Additionally, IT allows us to study any and all parts of a system (real or modeled) under a common dynamical framework, so that, in principle, no a priori assumptions must be made to understand relationships between a large number of diverse dynamical processes (Ruddell and Kumar, 2009).
This workshop aims to foster this renewed and growing interest by following and building upon a series of successful AGU and EGU sessions titled "On the Interface between Models and Data", and "Data & Models, Induction & Prediction, Information & Uncertainty: Towards a Common Framework for Model Building and Predictions in the Geosciences", respectively. These sessions have sought to bring greater insight and understanding to the topics of:

- How to properly include in models the things we already know (e.g., all physical laws, and not just conservation of mass and energy, etc.).
- How to evaluate the usefulness and robustness of data and models for a given task (their information content) in a generalized way.
- How to evaluate the appropriateness (generality, parsimony) of models given the data and the purpose?
- How to evaluate the interplay of data-, model structure- and predictive uncertainty, i.e., the flow of information from data through models to decision-makers?
- How to learn from the encounter of models and data; i.e., how to detect, diagnose and correct model structural errors?

1.2 Goals & Outcomes

The goals of this workshop are to promote the innovative use of Information Theoretic concepts in the service of discovery, modeling and decision-making in the Earth and Environmental Sciences and, through high level presentations and open discussion, to inspire revolutionary advances in the theories of modeling/learning, inference, and diagnostic evaluation.
Specific outcomes: will include one or more multi-authored white papers on the topic(s). The format will include a number of targeted presentations, but will mostly consist of moderated discussion groups focused around specific pre-prepared questions/issues. Each participant is requested to bring a poster about his or her research, and to be prepared to give a brief "Speed Presentation" on either the first or second day.

1.3 Additional background reading

We recommend a small amount of advance reading in preparation for this workshop. The following two articles provide a concise overview of the foundations and motivation for this workshop (they are also highlighted in boldface in the reference list below):

1. Theoretical Foundations: Knuth (2008) describes the theoretical basis for information theory, as well as its relationship with probability theory.

2. Overview of Relevance to Earth Sciences: Ruddell et al. (2013) provide a very short and recent overview of applications of information theory in the earth sciences.

For those interested in more in-depth reading, the broad themes that we see emerging in Earth Science are as follows:

- Model Evaluation and Uncertainty: Nearing & Gupta (2015) describe the general theory behind using information theory for model evaluation, and for reconciling models with observational data. A notable example of what this looks like is given by Majda & Gershgorin (2010).

- Analysis of Data: Weijs et al (2013) describe how information theory can be used for time-series analysis, and they do this by starting with one of the most fundamental theorems in science (Solomonoff, 1964). Another notable application to the analysis of spatiotemporal data fields is given by Brunsell (2010). (It is important to note that the Weijs et al. article also takes a very fundamental approach to the model uncertainty problem.)

- Analysis of Networks: It has been argued extensively in several fields of science that information theory provides a fundamental way to analyze complex system networks. This concept has been applied to understand emergent behavior in ecological networks (e.g., Ruddell et al., 2015).

- Information as a Physical Law: There are certain information theoretic principles that can be justified as basic laws of physics (indeed from which the more common laws of physics may be derived: Caticha & Cafaro, 2007, also see Knuth, 2015). Wang & Bras (2011) developed a model of the surface energy balance using this principle, and Nearing et al (2012) showed how addition of information-theoretic principles could be used to reduce the number of free parameters in a physics-based model.

1.4 References

Brunsell, N. (2010) 'A multiscale information theory approach to assess spatial-temporal variability of daily precipitation.' Journal of Hydrology 385(1), 165–172.

Caticha, A. and Cafaro, C. (2007) 'From information geometry to Newtonian dynamics.' arXiv preprint arXiv:0710.1071.

Datcu, M., Seidel, K. and Walessa, M. (1998) 'Spatial information retrieval from remote-sensing images. I. Information theoretical perspective', IEEE Transactions on Geoscience and Remote Sensing, 36(5), 1431–1445.

Jaynes, E.T. (2003) 'Probability Theory — The logic of Science', Cambridge University Press, ISBN 978-0521592710.

Knuth, K. H. (2004) 'What is a question?', In: C. Williams (ed.) Bayesian Inference and Maximum Entropy Methods in Science and Engineering, Moscow ID 2002, AIP Conference Proceedings vol. 659, American Institute of Physics, Melville NY, 227–242.

Knuth, K. H. (2008) 'On the origin of probability and entropy.' Presented at the Bayesian Inference and Maximum Entropy Methods in Science and Engineering at Sao Paulo, Brazil: `http://knuthlab.rit.albany.edu/papers/knuth-me08-final.pdf`.

Knuth, K. H. (2010) 'Information physics: The new frontier', arXiv preprint arXiv:1009.5161.

MacKay, D. J. (2003) Information theory, inference and learning algorithms. Cambridge University press, ISBN-13: 9780521642989.

Majda, A. and Gershgorin, B. (2010) 'Quantifying uncertainty in climate change science through empirical information theory.' Proceedings of the National Academy of Sciences 107(34), 14958–14963.

Nearing, G. S. and Gupta, H. V. (2015) 'The quantity and quality of information in hydrologic models', Water Resources Research, 51(1), 524–538.

Ruddell, B. L. and Kumar, P. (2009) 'Ecohydrologic process networks: 1. Identification', Water Resources Research, 45(3), 10.1029/2008wr007279.

Ruddell, B, L., Brundell N. and P. Stoy, P. (2013) 'Applying information theory in the geosciences to quantify process uncertainty, feedback, scale.' Eos, Transactions American Geophysical Union 94(5), 56-56.

Ruddell, B. L., Yu, R., Kang, M., amd Childers, D. L. (2016) 'Seasonally varied controls of climate and phenophase on terrestrial carbon dynamics: modeling eco-climate system state using Dynamical Process Networks.' Landscape Ecology, 31(1), 165-180.

Solomonoff, R. (1964) 'A formal theory of inductive inference. Part I.'Information and control 7(1), 1–22.

Tononi, G. (2011) 'Integrated information theory of consciousness: an updated account', Archives italiennes de biologie, 150(2–3), 56–90.

Wang J. and Bras, R. (2011) 'A model of evapotranspiration based on the theory of maximum entropy production.' Water Resources Research 47(3).

Weijs, S., N. Van De Giesen, N., and Parlange, M. (2013) 'Data compression to define information content of hydrological time series.' Hydrology and Earth System Sciences 17. EPFL-ARTICLE-188013: 3171-3187.

The conference organizers have collected several other papers that describe both the theoretical origins and implications of information theory for the physical sciences, as well as several example applications in the Earth Sciences. These are hosted on a shared drive here: `https://arizona.box.com/s/elh3l6s59votnduw2ymuu19xz9pf479j`

2 Schedule

2.1 Sun April 24th: Arrival

2.2 Mon April 25th: What is information theory and why should we care?

► Morning session (chaired by Grey Nearing)	
08:00 – 08:15	Coffee
08:15 – 08:30	Introduction (Hoshin Gupta, Grey Nearing, Uwe Ehret)
08:30 – 09:00	Group Introduction
09:00 – 10:00	Invited Talk (Kevin Knuth): On the Relationship between Information Theory and Physics
10:00 – 10:15	Coffee
10:15 – 10:45	Invited Talk (Michal Branicki): Interplay between information theory, uncertainty quantification, and improving reduced-order predictions
10:45 – 11:30	Plenum Discussion
11:30 – 13:00	Lunch
► Poster session I (chaired by Florian Wellmann and Rohini Kumar)	
13:00 – 13:45	Speed Presentations by Poster Presenters
13:45 – 15:30	Poster Session I
► Afternoon session (chaired by Ben Ruddell)	
15:30 – 16:00	Talk (Hoshin Gupta): On the Relationship between Information Theory and the Hydrological Sciences
16:00 – 16:30	Plenum Discussion

16:30 – 18:00	Breakout Sessions: What are the core questions in the earth sciences and how can we inform these questions? 3 groups sessions (led by Kevin Knuth, Praveen Kumar, Jingfeng Wang)
18:00 –	Dinner & Socializing

2.3 Tue April 26th: Information in Data, Models, and Systems

► Morning session (chaired by Steven Weijs)	
08:00 – 08:15	Coffee
08:15 – 09:00	Reports from the 3 breakout sessions
09:00 – 09:30	Talk (Grey Nearing): On the Information Content in Data
09:30 – 10:00	Invited Talk (Wei Gong): On the Information in Models
10:00 – 10:15	Coffee
10:15 – 10:45	Talk (Ben Ruddell): On the Information in Networks
10:45 – 11:30	Plenum Discussion
11:30 – 13:00	Lunch
► Poster session II (chaired by Florian Wellmann and Rohini Kumar)	
13:00 – 13:45	Speed Presentations by Poster Presenters
13:45 – 15:30	Poster Session II
► Afternoon session (chaired by Uwe Ehret)	
15:30 – 16:00	Talk (Steven Weijs): On Information and Complexity
16:00 – 16:30	Plenum Discussion
16:30 – 18:00	Breakout Sessions: How can information theory help us understand the interface between models and data? 3 groups sessions (led by Michal Branicki, Wei Gong, Joon Kim)
18:00 –	Dinner & Socializing

2.4 Wed April 27th: Physical Models from an information perspective

► Morning session (chaired by Bethanna Jackson)	
08:00 – 08:15	Coffee
08:15 – 09:00	Reports from the 3 breakout sessions
09:00 – 09:30	Invited Talk (Jingfeng Wang): Maximum Entropy Production
09:30 – 10:00	Invited Talk (Praveen Kumar and Allison Goodwell): Information Sharing in Eco-hydrologic Systems: Synergy, Uniqueness, and Redundancy
10:00 – 10:15	Coffee
10:15 – 10:45	Talk (Bethanna Jackson): Entropy-based metrics to evaluate physical models
10:45 – 11:30	Plenum Discussion
11:30 – 13:00	Lunch
13:00 – 16:30	Visit the research facilities of the Schneefernerhaus and the summit of the Zugspitze
► Afternoon session (chaired by Hoshin Gupta and Ben Ruddell)	
16:30 – 18:00	Group discussion, Workshop Conclusion, Future Planning, Paper Preparation
18:00 –	Dinner & Socializing

2.5 Thurs April 28th: Morning Departure

3 Talks

3.1 Abstracts

Information Physics and the Application of Information Theoretic Concepts to Earth Science

Kevin Knuth[1]
[1]*University at Albany*

Traditionally the Laws of Physics are conceived as laws decreed by Mother Nature that must be discovered by scientists. However two major revolutions in physics in last century, quantum mechanics and relativity, have made clear the importance of the role of the observer in physical law. This is difficult to reconcile with the concept of physical laws as being dictated by Nature. At this point in time, there is a great deal of evidence that the laws of physics represent rules for the optimal processing of information about the world around us and its relationship to us as participatory observers. This way of thinking about physics has come to be called “Information Physics” or “Information-Based Physics”.
In this talk I will present a brief overview of Information Physics, and describe a modern way to conceive of probability theory as a logically consistent means of quantifying the degree to which one logical statement implies another. Similarly, the quantity known as entropy can be shown to consistently quantify the degree to which one question answers another.
I will conclude by describing past work where we have used information theoretic quantities, such as mutual information, to attempt to identify relevant climate variables, as well as quantities such as transfer entropy to identify causal relations in complex dynamical systems.

Interplay between information theory, uncertainty quantification, and improving reduced-order predictions

Michal Branicki[1], **A. J. Majda**[2]
[1]*School of Mathematics, University of Edinburgh, UK*
[2]*Courant Institute of Mathematical Sciences, NYU, New York, USA*

The issue of mitigating model error in reduced-order prediction of high-dimensional, PDE driven dynamics is particularly important when dealing with turbulent geophysical systems with rough energy spectra and intermittency near the resolution cut-off of the corresponding numerical models. I will discuss a new framework which allows for information-theoretic quantification of uncertainty and mitigation of model error in imperfect stochastic/probabilistic predictions of non-Gaussian, multi-scale dynamics. In particular, I will outline the utility of this framework in derivation of a sufficient condition for improving imperfect predictions via a popular Multi Model Ensemble approach; this formulation also gives rise to systematic guidelines for optimising data assimilation techniques which are based on multi model ensembles. Time permitting, the role and validity of "fluctuation-dissipation" arguments for improving imperfect predictions of externally perturbed non-autonomous turbulent systems will also be addressed.

Information Theory and the Hydrological Sciences: Models, Data, Uncertainty and Learning

Hoshin V Gupta[1]
[1]*Department of Hydrology & Atmospheric Sciences The University of Arizona*

Applications of Information Theoretic concepts to hydrology go back to at least the early 1970's. However, the past decade has seen a dramatic resurgence of interest, particularly in the contexts of prediction (decision making), and of data assimilation and model development (the basis for scientific investigation). This talk will briefly review (by no means comprehensively) the recent hydrological literature to reveal some emergent themes and then discuss how Information Theory can help us understand the utility of "Dynamical Environmental Systems Models (DESM)" as vehicles for learning. We conclude that (a) models encode information as an hierarchical sequence of hypotheses, assumptions or decisions, (b) all DESM models should be probabilistic, and (c) it is necessary to take a Maximum Entropy (MaxEnt) approach to the assimilation and representation of information. One approach to this is via the use of MaxEnt process parameterization equations that properly (honestly) represent what we actually know (or can reasonably assume) about the underlying physical processes operating at the modeling scale of interest. While conceptual and computational issues still need to be addressed, the approach enables an inferential process that is more consistent with the philosophy of science than the current strategy based primarily in unjustifiably strong hypotheses about what we actually know.

Information in Data

Grey Nearing[1]
[1]*NASA/NCAR/UMBC*

Observations are part of the physical universe and therefore contain information in an ontological sense. To translate this into epistemological information, we must define an encoder and a decoder based on our own understanding of the natural world — this process is called science. Further, information in data may be used to (a) test hypotheses, or (b) make predictions for decision-making. In this talk I will cover the basic principles behind how both science and science arbitration efforts use information available from observations.
First, I will cover how we might understand the concept of falsification from an information theoretic perspective, and suggest that this has the potential to reconcile our most famous demarcation criteria with a fundamentally Bayesian epistemology. That is, if we understand the amount of information available to some science experiment — some hypothesis test — then we can ask the

coherent question "*does the model use all of the available information?*" If the answer is "no", then our model has failed the test, and there is potential to use the information in observations to improve the model. We will discuss briefly the basic theory about how we might go about extracting such information. If the answer is "yes", then our model has passed all tests that we have put to it and remains unfalsified.
Second, I will discuss the foundational concept behind using information from observations to make predictions of dynamical systems. I will outline a basic theory about how we may determine whether our prediction systems are using all of the information available to them. Testing such prediction systems using a scientific method is a tri-variate problem (model, observation, experiment), and an information-theoretic treatment of this problem completely generalizes triple collocation. Further, because information is linear (a basic ontological fact) we may, in principle, estimate the efficiency of our ability to extract information from data.

On the Information in Models: quantification of model structure adequacy with information based metrics

Wei Gong[1]
[1]*College of Global Change and Earth System Science (GCESS), Beijing Normal University, Beijing, 100875, and Joint Center for Global Change Studies, Beijing, 100875, China*

The uncertainty of dynamic models can be categorized as (1) aleatory uncertainty that caused by random error of observation (2) epistemic uncertainty that caused by lack of knowledge about physical processes and can possibly be reduced by improving the model physics and parameterization schemes. To discriminate these two kinds of uncertainties, a model-structure independent approach based in information theory has been proposed to estimate the best achievable predictive performance of a model (for a system given the available data), and quantify the room for model structural improvements. In developing the information theory based metrics we have solved the following challenging problems (1) the approaches of manifold learning are attractive but not suitable for characterizing aleatory and epidemic uncertainty; (2) By reducing the dimensionality to 1D marginal distribution, Independent Component Analysis (ICA) is able to do this but with a linear approximation; (3) Discuss some technical issues of probability density estimation and 1D entropy estimation; (4) On extending the ICA based approach to nonlinear problem, what's the problem ahead? (5) On extending the information theory based approaches to large complex dynamic models, such as land surface / weather / climate models, where is the way ahead? (6) On the way to challenge the big BOSS: Earth System Model, how to guarantee the inherent consistency of archiving good performance with the correct way. Consequently, there has been a growing demand of developing appropriate performance metrics to summarize the physical characteristics of hydrology / land surface / weather and climate models, such as the daily/annual cycle, climate sensitivity, water/energy patterns, etc. Information theory based metrics may have a great opportunity to do so!
Furthermore, I would like to give a brief introduction about my recent research about surrogate-based optimization and MCMC for large complex dynamic models. We have achieved great breakthrough of simultaneously improve every aspect (objective, or physical processes) of a coupled dynamic model with only a few hundreds of model evaluations. Now we have almost been

ready to challenge ESM if there are enough funding, and inherent consistent model performance metrics.

Climate sensitivity of global terrestrial ecosystems' subdaily carbon, water, and energy dynamics

Benjamin L. Ruddell (With Rong Yu, Dan Childers, Minseok Kang, and Joon Kim)[1]
[1]*Fulton Schools of Engineering, Arizona State University*

Under the context of global climate change, it is important to understand the direction and magnitude of different ecosystems respond to climate at the global level. In this study, we applied dynamical process network (DPN) approach combined with eco-climate system sensitivity model and used the global FLUXNET eddy covariance measurements (subdaily net ecosystem exchange of CO_2, air temperature, and precipitation) to access eco-climate system sensitivity to climate and biophysical factors at the flux site level. Eco-climate system sensitivity at flux timescales was estimated at the global flux sites and extrapolated to all possible land covers by employing artificial neural network approach. The extrapolation utilizes MODIS phenology and land cover products, the long-term climate GLDAS-2 product, and the GMTED2010 Global Grid elevation dataset. We found that the eco-climate system dynamical process structures are more sensitive to seasonal temperature, than to radiation, phenology, or (lowest sensitivity) precipitation. Interestingly, if global temperature continues rising, the temperature-to-NEE process coupling may increase in tropical rain forest areas while decreasing in tropical desert or Savanna areas. At the same time, phenology showed a positive effect on the temperature-to-NEE process coupling at all pixels, so increased greenness increases the importance of temperature to carbon dynamics and consequently carbon sequestration globally. This work is unique in that it provides a theoretically independent and complex system based means of assessing the sensitivity of global ecosystem processes to climate change, and it can therefore be used to critique or corroborate the findings of process based ecosystem models.

Information, complexity, description length, data compression and their application in hydrology

Steven V. Weijs[1]
[1]*Department of Civil Engineering, UBC, Vancouver, Canada*

Science is data compression. In this talk I highlight some of the links between learning from data, information content and description length from the perspectives of information theory and algorithmic information theory. These concepts are useful in quantifying how much information is contained in data, how much is added by models, and how much by prior knowledge. In minimizing description length of data, we automatically find a balance between model complexity and the accuracy of that model describing the data: A compact model with little unexplained uncertainty about the known data is more likely to be a good predictor. This is expressed by Solomonoff's Algorithmic probability, which can describe a discrete prior distribution over all computable models/functions. This prior assigns the highest weight to the simplest models. I describe some applications of data compression algorithms to find approximations of minimum description length, a quantity that is fundamentally incomputable.

Maximum Entropy Production Model of Heat Fluxes — An Application of Information Theory in Earth System Modeling

Jingfeng Wang[1]
[1] *School of Civil and Environmental Engineering Georgia Institute of Technology*

The development of the maximum entropy production (MEP) model of evapotranspiration (ET) and heat fluxes is made possible by the information theory with two crucial concepts: generalized thermodynamic entropy (production) and the maximum entropy (production) principle as a physical law as well as an inference algorithm. Generalization of the thermodynamic entropy production is needed to formulate the MEP model of ET using the maximum entropy production principle as a derivative of the maximum entropy principle applied to non-equilibrium thermodynamic processes such as the exchange of water and heat between the Earth and atmosphere. It also allows boundary layer turbulence model, a key component for modeling surface fluxes, to be incorporated into the MEP formalism. The application of information theory has led to the first-principles based model of ET and heat fluxes that (1) closes surface energy budgets at all space and time scales; (2) is independent of temperature and moisture gradients, wind speed and surface roughness parameters, (3) holds for the entire range of soil moisture, and (4) is parsimonious in model input and parameter free of empirical tuning parameters. The MEP model may be used as a retrieval algorithm for remote sensing of surface fluxes and a physical scheme of surface energy budgets in coupled regional and global land/ocean-atmosphere models.

Information Sharing in Eco-hydrologic Systems: Synergy, Uniqueness, and Redundancy

Praveen Kumar[1], **Allison Goodwell**[1]
[1] *University of Illinois*

Ecosystems are complex systems in which interactions between atmospheric, soil, and vegetation components evolve due to perturbations in addition to diurnal and seasonal patterns. Information theory measures, such as conditional mutual information, can identify synchronized and time-lagged relationships between these components and how they shift over time. In a process network, a detected time dependency between a "source" and "target" time-series node constitutes a link. The shared information identified as the link could be partially or completely (a) unique, meaning that no other source provides the same information to the target, (b) redundant with other links, or (c) synergistic with other links, meaning that two sources provide more information to the target than the union of their individual contributions. We extend a recently introduced information decomposition approach to develop measures of unique, redundant, and synergistic information within a process network. We apply these to analyze a 1-minute resolution dataset from a weather station in central Illinois. We see that patterns of synergistic, redundant, and unique information quantities are different from those of combined information measures, indicating their potential as predictors of eco-hydrologic characteristics.

Entropy-based metrics to evaluate physical models

Bethanna Jackson[1]
[1]*School of Geography, Environment and Earth Sciences, Victoria University of Wellington, PO Box 600, Wellington 6140, New Zealand*

This talk considers how entropy and other information theory concepts can be placed within a physically relevant context, and used to help quantify model structure, performance and data worth. Challenges and recent progress in extracting hydrologically meaningful information theory measures from modelled and measured flow signals are discussed. Such measures can provide insights into structure and extent of memory within catchment systems, inform where model structural components may be inappropriate, the point to where data is most limiting, and sometimes inform on where previously unsampled conditions and/or events have recently been collected; prompting re-analysis. Fundamentally, it is hoped that these physically relevant information theory measures will provide new insights into 1) inherent limitations in characterisation and predictability of catchment system response and 2) into our ability to predict given specific resource and data limitations.

4 Poster presentations

4.1 Poster session I (Mon)

Information Theory and the Hydrological Sciences: Models, Data, Uncertainty and Learning Case study and illustrative example for conceptual rainfall-runoff models

Shervan Gharari[1], **Hoshin Gupta**[2], **Grey Nearing**[3], **Uwe Ehret**[4], **Fabrizio Fenicia,**[5], **Markus Hrachowitz**[5], **Patrick Matgen**[5], **Saman Razavi**[5], **Hubert Savenije**[5], **Howard Wheater**[5]
[1]*TU Delft*
[2]*University of Arizona*
[3]*NASA*
[4]*KIT Karlsruhe*
[5]*(not specified)*

From a systems point of view, a model building process can be explained based on the hierarchical assemblage of model-information. Based on this view, the model building practice can be summarized in three main steps:

1. A decision should be made on the boundaries of the system and on the ingoing and outgoing mass and energy fluxes.
2. A decision should be made on the sub-system architecture or the geometry of the model; whereby the number of control volumes, e.g. reservoirs, and their interconnecting fluxes should be designed.
3. A decision on model parameterization, i.e. state-discharge relations or constitutive functions, should be made in order to be able to track fluxes from states and states from fluxes.

Building on previous studies we answer the following questions in this illustrative example applied to conceptual rainfall-runoff models. The research questions we are keen to answer are:

1. How much information is provided in each step of model building, especially in the choice of model structure (the second level)?
2. How much information is provided by assuming the shape of the model parameterization as reflected in the model (the third level)?
3. What is the best complexity for a given set of data (model adequacy)?
4. And how a decision on model structure (the second level) can affect model parameterization (the third level) if no assumptions are made concerning the shape of these parameterizations (inverse problem)?

Monitoring and interpreting the flows of energy, matter and information in ecosystems

Joon Kim[1], **Ko Flux**[2]
[1]*Department of Landscape Architecture and Rural Systems Engineering / Interdisciplinary Program in Agricultural and Forest Meteorology, Seoul National University, Seoul, South Korea*
[2]*National Center for AgroMeteorology, Seoul, South Korea*

The United Nations Food and Agriculture Organization pursues the climate smart agriculture (CSA) with a hope of triple wins: (1) sustainably increasing agricultural productivity and income of farm household; (2) reducing and/or removing greenhouse gases emissions, where possible; and (3) adapting and building resilience to climate change. Using the time series of the long-term flux data at cropland and forest sites in AsiaFlux (i.e., a regional network of eddy covariance tower-based networks of carbon/water/energy flux monitoring), a suite of biotic, network, and thermodynamic indicators are evaluated. These ecological indicators are then integrated to scrutinize the ecosystems in terms of the CSA's triad strategies (i.e., efficiency, resilience, and carbon footprint) sustainable intensification while ensuring ecosystem integrity. Sustainability is all about maintaining ecosystem integrity which is preserved when the ecosystems' self-organizing processes are preserved. One of the challenges here is how to quantify information flows and how to interpret their roles in self-organizing processes in ecosystems.

A spatially consistent seamless predictions of continental-scale hydrologic fluxes and states

Rohini Kumar[1], **Juliane Mai**[1], **Oldrich Rakovec**[1], **Matthias Zink**[1], **Matthias Cuntz**[1], **Stephan Thober**[1], **Sabine Attinger**[1], **Martin Schroen**[1], **David Schaefer**[1], **Luis Samaniego**[1]
[1]*Helmholtz Centre for Environmental Research - UFZ, Leipzig, Germany*

One of the major challenges in the contemporary hydrology is to establish a continental-scale hydrologic model that can provide spatially consistent, seamless prediction of hydrologic fluxes and states to better characterise extreme events like floods and droughts. In this study, we demostrate the applicability of a mesoscale hydrologic model parameterized using a multiscale regionalization technique to derive daily gridded fields of hydrologic fluxes/ states over the Pan-EU domain since 1950. A multi-basin parameter estimation (MBE) strategy that utilizes observed streamflows from a set of hydrologically diverse basins is introduced to infer a representative set of regional calibration parameters applicable over the entire EU domain. We tested three sampling schemes to select a set of calibration basins to maximize information content during the parameter calibration process. These are based on the 1) random selection procedure, 2) gradient along the hydro-climatic regimes, and 3) diversity in hydro-climatic and basin physiographical properties (e.g., terrain, soil, land cover properties).
Results of the MBE approach are contrasted against the benchmark at-site calibration strategy across 400 EU basins varying from approximately 100 to 500,000 km^2. Single-site calibrated parameters performed best for site-specific streamflow predictions, but their transferability to other sites resulted in poor performance. Moreover, the at-site calibration strategy generated a patchy,

spatially inconsistent distribution of parameter fields that further induced large discontinuities in simulated hydrologic fields of soil moisture among other sates/fluxes. These limitations were overcome by the MBE strategy that provided a compromise solution with improved model performance compared to at-site cross-validated estimates. The gridded fields of hydrologic parameters, states and fluxes from MBE were spatial continous and much more meaningful compared to those of the at-site calibration strategy. The selection of calibration basins that include diversity in both hydro-climatic and basin physical properties provided consistently better results compared to other two strategies. This selection strategy extracted most of available information from hydrologically diverse basins to infer a consistent set of model parameters for performing a spatially consistent seamless predictions of continental hydrological fluxes and states.

Hydrological model complexity depends on the magnitude of its parameters

Saket Pande[1]
[1]*Department of Water management, Delft University of Technology, Netherlands*

Three measures of complexities are computed for SIXPAR model on a MOPEX basin data set. One of the measures is based on an approximation of marginal likelihood function, while the other two are based on measuring stability of system representation by SIXPAR. Results suggest that the 3 measures are equivalent, indicating that if the Bayesian measure is a valid measure of model complexity then the other two are valid measures of complexity as well. Results further suggest that SIXPAR complexity varies with the magnitude of its parameters.

Revisit of the Global Surface Energy Balance Using the MEP Model of Surface Heat Fluxes and Remote Sensing Observations

Shih-Yu Huang[1], **Yi Deng**[2], **Jingfeng Wang**[1]
[1]*School of Civil and Environmental Engineering, Georgia Institute of Technology*
[2]*School of Earth and Atmospheric Sciences, Georgia Institute of Technology*

The climatology of global surface heat fluxes is re-evaluated using the maximum entropy production (MEP) model and surface radiation fluxes and temperature data of NASA Clouds and the Earth's Radiant Energy System (CERES) supplemented by surface humidity data from Modern-Era Retrospective analysis for Research and Applications (MERRA). The MEP model of surface fluxes have unique properties: (1) closure of surface energy budgets at all space and time scales; (2) independent of temperature and moisture gradients, wind speed and surface roughness parameters, (3) covering entire range of soil moisture over land surfaces, and (4) parsimonious in model input and parameter free of empirical tuning parameters. The new MEP-based global heat fluxes over lands agree with the previous estimates. The new estimate of ocean evaporation is lower than previous estimates, while the new estimate of ocean sensible heat flux is higher than previously reported. The MEP model produces the first global ocean surface heat flux product.

Predict renewal times in an aquifer: what is the informative content carried by atmospheric tracer data measurements? A synthetic case study of the crystalline aquifer of Ploemeur, Brittany

J. Marcais[1,2], **J.-R. de Dreuzy**[2], **T.R. Ginn**[3], **P. Rousseau-Gueutin**[4], **S. Leray**[5]
[1]*Agroparistech, 16 rue Claude Bernard, 75005 Paris, France*
[2]*Géosciences Rennes (UMR 6118 CNRS), Université de Rennes 1, Campus de Beaulieu, 35042 Rennes Cedex, France*
[3]*University of California, Department of Civil and Environmental Engineering, 1 Shields Avenue, Davis, CA 95616, USA*
[4]*EHESP Rennes, Sorbonne Paris Cité, Paris, France*
[5]*IFP Energies nouvelles, 1-4 avenue de Bois-Préau, 92852 Rueil-Malmaison Cedex, France*

Transit time distributions (TTDs) play a key role in the transport processes, the interpretation of atmospheric tracer data and the predictions made on groundwater resources. However TTDs are not accessible from field measurements. Some information on these TTDs can be obtained from anthropogenic tracer concentrations, also commonly called "groundwater ages". We evaluate the predictive capabilities of the information contained in anthropogenic tracer concentrations on groundwater renewal times through the use of Lumped Parameter Models (LPMs) instead of fully developed aquifer models. Towards this end, we develop an assessment methodology in three steps. First, a synthetic crystalline aquifer model representing the site of Ploemeur (Brittany, France) is used to give, at any point, references for observables quantities (anthropogenic tracer concentrations of CFC-11, ^{85}Kr and SF_6), for non-observables quantities (the TTDs), and for the prediction objective (groundwater renewal times). Second, several LPMs are considered with one, two or three parameters, and are parameterized by fitting the reference anthropogenic concentrations. Third, the reference renewal times obtained from the synthetic model are compared to the renewal times obtained independently from the LPMs. Statistical analyses over the aquifer show that a good fit of the anthropogenic tracer concentrations is a necessary but not sufficient condition for acceptable predictions. The use of only one anthropogenic tracer gives poor predictions differing by 7 to 12 years to the references. The use of two sufficiently different anthropogenic tracers not only reduces the errors but surprisingly yield to very accurate predictions with errors smaller than 3 years. The additional use of a third anthropogenic tracer does not improve the predictive capabilities. Careful a posteriori analyses reveal that reference TTDs have widely varying shapes from well peaked in recharge zones where flows are diverging to broadly distributed in more converging zones. Nonetheless, the same LPM (whether it is an inverse Gaussian or a shifted exponential) gives excellent predictions everywhere. In such circulation patterns where dispersive and mixing processes prevail at different scales, broad distributions seem to be more suited than multi-modal or shape-free models.

A network approach to determine ecosystem vulnerability to extremes

Allison Goodwell[1], **Praveen Kumar**[1]
[1]*University of Illinois*

Ecosystems evolve due to complex interactions over various space and time scales. Process networks, in which nodes are time-series variables and links are information theoretic measures, allow us to analyze a system in terms of unique and redundant time dependencies. It has been shown that process networks composed of measured and modeled ecohydrologic time-series variables exhibit breakdowns in feedbacks during periods of climate extremes such as drought. In this study, we use an information decomposition approach to partition shared information between ecohydrologic time-series nodes into redundant and unique components. Redundant information is information shared between multiple source nodes with a target node, while unique shared information is only attributable to a single source node. In an ecohydrologic network, unique shared information between two nodes is likely to represent a critical ecosystem link, and redundant shared information indicates synchronization or internally induced feedbacks between variables. We form networks using flux tower and ecohydrologic model output variables over a range of natural and intensively managed ecosystem types. Methods to compute information transfer components with short datasets are used in order to observe shifts in network behavior that vary with weather conditions and extreme events on hourly to weekly timescales. This analysis shows that a network approach can detect critical linkages that dictate ecosystem vulnerability to extreme events.

4.2 Poster session II (Tue)

Surrogate-based Multi-Objective Optimization and Uncertainty Quantification Methods for Large, Complex Geophysical Models

Wei Gong[1], **Qingyun Duan**[1]
[1]*College of Global Change and Earth System Science (GCESS), Beijing Normal University, Beijing, 100875, and Joint Center for Global Change Studies, Beijing, 100875, China*

Parameterization scheme has significant influence to the simulation ability of large, complex dynamic geophysical models, such as distributed hydrological models, land surface models, weather and climate models, etc. with the growing knowledge of physical processes, the dynamic geophysical models include more and more processes and producing more output variables. Consequently the parameter optimization / uncertainty quantification algorithms should also be multi-objective compatible. Although such algorithms have long been available, they usually require a large number of model runs and are therefore computationally expensive for large, complex dynamic models. In this research, we have developed surrogate-based multi-objective optimization method (MO-ASMO) and Markov Chain Monte Carlo method (MC-ASMO) for uncertainty quantification for these expensive dynamic models. The aim of MO-ASMO and MC-ASMO is to reduce the total number of model runs with appropriate adaptive sampling strategy assisted by surrogate modeling. Moreover, we also developed a method that can steer the search process with the help of prior parameterization scheme derived from the physical processes involved, so that all of the objectives can be improved simultaneously. The proposed algorithms have been evaluated with test problems and a land surface model - the Common Land Model (CoLM). The results demonstrated their effectiveness and efficiency.

Can causality be detected in noisy environmental systems?

Christopher J Tennant[1], **Laurel G Larsen**[1]
[1]*Department of Geography, University of California, Berkeley, CA, USA*

Identifying causal relationships is a fundamental scientific goal. Complexity, emergence, nonlinearity, and delayed forcing complicate attribution of causality. For example, even simple sets of ordinary differential equations (ODEs), where the rules are explicitly formulated, produce complexity that belie their seemingly simplistic nature. These parametric equations are often used to model complex environmental systems where causality detection using data streams is further complicated by the addition of observational noise (instrumental sensor error) and process noise (dynamics not captured by models). The Lorenz system is a set of ordinary differential equations,

$$\begin{aligned} dX/dt &= \sigma(y-x) \\ dY/dt &= x(\varrho - z) - y \\ dZ/dt &= xy - \beta z \ , \end{aligned}$$

where σ, ρ, and β determine whether the system exhibits chaotic, non-chaotic, or transient chaotic dynamics and provides an example of a set of ODEs that has been used to describe state-space dynamics for atmospheric convection. Here, we use the Lorenz to evaluate the ability of transfer entropy to detect causal relationships with increasing levels of observational and process noise added to the time series of one or all of the state variables. Specifically, observational noise was added to Xt by first solving the Lorenz system and then adding white noise to Xt's time series. This was done iteratively to produce a suite of Xt time series with progressively increasing noise levels; noise was not added to Yt and Zt in the observational noise experiment. Process noise was added by discretizing the Lorenz system and iteratively solving the three equations; at each time step noise was added to Xt and was propagated to Yt and Zt through their dependence on Xt. The same noise levels used in the observational noise experiment were used in the creation of process noise.

At noise levels equal to zero (i.e. no noise added to the system) Xt and Yt transfer information to each other at historical lags (≈1 to 45 time steps) but the transfer is less than their shared information at lag zero. Interestingly, Yt and Zt both transfer significantly more information to each other at time lags from 1 to 55 and 1 to 45, respectively, that is greater than their mutual information at lag zero. Xt and Zt transfer significant information to each other but their information transfer never exceeds their mutual information. The results with no noise added to the system, suggest that Yt and Zt are the strongest drivers with respect to information transfer. While Xt does transfer information to Yt and Zt, the information transferred never exceeds their mutual information, suggesting that Xt evolves more synchronously with Yt and Zt.

As observational noise is added to Xt, its transfer of information to Yt and Zt quickly declines and no significant information transfer occurs at any lags. That is, as Xt becomes less deterministic and more stochastic, it no longer reduces the uncertainty of other system variables. However, the loss of propagation of information by Xt is accompanied by an increase in the ability Yt and Zt to reduce X's uncertainty. This reflects the strong embedding of Xt in the dynamics of Yt and Zt, which results in a "memory" of Xt that is retained in the system. This memory effect strengthens as the noise level of Xt increases. These observations imply that, while a data series with a low signal to noise ratio could be an important driver of system dynamics, this quality (low S:N), makes causal influences of a noisy variable difficult to detect or infer. These results also suggest that causally linked variables with much greater signal to noise can greatly reduce the uncertainty of noisy variables.

The effects of process noise on system dynamics are inherently different from those of observational noise because the noise affects all variables in the system. Whereas observational noise leads to Yt and Zt transferring more information to Xt, the increase in process noise causes the opposite, with Xt transferring more information to Yt and Zt. Information transfer between Yt and Zt also increases as the noise level increases. In all cases, the timing of peak information transfer shifts to shorter time lags. These results imply that a variable with strong stochastic dynamics (Xt), can appear to be the most important system driver, even if in the underlying deterministic system the variable is not as influential as other state variables. The results also imply that stochastic spikes to a deterministic system can push timescales of significant information transfer to shorter lags, even if the underlying noise-free system has a longer memory. Together these results provide an important framework for how we assimilate observations and attribute causality. A firm understanding of signal to noise and characterization of the stochastic dynamics of system variables is warranted when estimating causal linkages.

Space in the informationscape: Resolution of critical spatial scales and functional connectivity in heterogeneous landscapes

Laurel Larsen[1], **Mollie van Gordon**[1], **Christopher Tennant**[1], **Saalem Adera**[1], **Dino Bellugi**[1], **Hong-xu Ma**[1], **Theresa Oehmke**[2]
[1]*Department of Geography, UC Berkeley, Berkeley, CA*
[2]*Department of Civil and Environmental Engineering, UC Berkeley, Berkeley, CA*

Resolution of spatial scales over which dominant processes drive landscape functionality is critical for model development and assessing resilience. Extensions of transfer entropy to spatially explicit datasets hold promise for delineating dominant spatial scales. However, their application to heterogeneous environments with limited spatial or temporal data coverage is far from straightforward. Approaches span a gradient in the extent to which space is represented explicitly. Here, we share insights gained from application of entropy-based analyses to a range of questions about the scales, directionality, and impact of feedback and forcing in spatially explicit systems. Data sources include numerical models of sandpiles and turbulence, lidar imagery of the Wax Lake Delta, sparsely distributed rain gauge and streamflow networks, coupled with climate and land cover data in the Amazon and Sahel, and water quality monitoring stations in the Everglades.
Where data are abundant in space and time, pairwise transfer entropies across all points can be formulated in both sets of dimensions, and the challenge becomes one of dimensionality reduction of the resulting process network. Preservation of information about the geographic anchoring of pairwise entropy calculations is important in attempting to identify geographic sources or sinks of information (e.g., of landslides in the sandpile model) or resolve region-specific functional connectivity pathways (e.g., hydrologic connections in the Amazon or Sahel). However, process-domain scale summing or averaging of transfer entropies over ranges of temporal or spatial lags is appropriate for determining salient system-scale properties (e.g., power-law scaling of information transfer in turbulence or the sandpile model). In a landscape experiencing a strong directional forcing (e.g., flow of water), spatial and temporal dimensions are not fully independent, and snapshots of the data field (e.g., vorticity) may be sufficient to resolve critical spatial scales of information transfer. However, the temporal dimension must be introduced to infer the critical directionality of information transfer.
Where data are abundant in space but not time (e.g., lidar sensing imagery), process connectivity may still be revealed with transfer entropy using a space-for-time substitution, but doing so requires assumption of process homogeneity or proper stratification of process domains. Performing analyses on subsets of data corresponding to patches (e.g., of vegetation) is one potential approach for satisfying this assumption. In the most limited datasets — such as small-scale networks of manually sampled sites — data spatiotemporal resolution is often insufficient to perform causal inference using transfer entropy. In such cases, functional connectivity can be inferred through non entropy-based techniques (e.g., event synchronization, generalized linear modeling of residuals) that operate under a similar strategy of tracking the propagation of perturbations, even though they are less adept at resolving nonlinear relationships.
Regardless of how they are resolved from data, spatially explicit functional and/or process connectivity networks may be analyzed through other entropy-based statistics to predict their resilience to small perturbations. Here we show how a network mutual information statistic and

Shannon entropy-based metrics can quantify the capacity, redundancy, and scale distribution of spatial functional connectivity networks, thereby suggesting their potential for resilience.

Structogram: A method to quantify the structuredness and complexity of data sets

Uwe Ehret[1]
[1]*Institute of Water Resources and River Basin Management, Karlsruhe Institute of Technology - KIT, Karlsruhe, Germany*

This poster presents a method to quantify the structuredness and complexity of spatial and/or temporal data sets based on a Structogram. 'Structuredness' relates to how information on the data set is distributed among its elements, 'complexity' relates to how this information is distributed over the extent of the data set. The core procedure is to approximate the data set by a subset of its elements and, via interpolation, to estimate the values of all elements. The related estimation error quantifies the information on the data set contained in the subset. Increasing the subset size from zero by always adding the most informative element yields an ordered list of subset size and related error, the plot of which is the Structogram. Highly structured data sets can be represented by small subsets; little structured data sets require large subsets. For each data set, a pareto optimal subset can be identified which jointly minimizes subset size and the related error. This indicates the optimal level of (lossy) compression for the data set. The complexity of the data set is estimated by evaluating the distribution of the elements in the pareto optimal subset across the extent of the data set: Irregular distributions indicate complex data sets, uniform distributions point to little complex data sets. This is quantified by applying the Structogram procedure to the pareto optimal subset, but replacing the element values by the inter-element distances.
The Structogram approach is applied to both artificial and real-world data sets. It is shown that the Structogram is invariant to additive and multiplicative transformation and the length of the data set, and that the resulting ranking of the data sets with respect to structuredness and complexity is in accordance with expectations. Useful applications of the Structogram include the characterization of data sets, design and evaluation of monitoring networks, and, comparable to a variogram-based Kriging, for geostatistical interpolation. Differently from the latter, the Structogram does not assume second-order stationarity or the intrinsic hypothesis. The Structogram approach can be applied to time series and spatial data sets of arbitrary dimension

What Constitutes a "Good" Sensitivity Analysis? Elements and Tools for a Robust Sensitivity Analysis with Reduced Computational Cost

Saman Razavi[1], **Hoshin Gupta**[1]
[1]*University of Arizona*

Global sensitivity analysis (GSA) is a systems theoretic approach to characterizing the overall (average) sensitivity of one or more model responses across the factor space, by attributing the variability of those responses to different controlling (but uncertain) factors (e.g., model parameters, forcings, and boundary and initial conditions). GSA can be very helpful to improve the credibility and utility of Earth and Environmental System Models (EESMs), as these models are continually growing in complexity and dimensionality with continuous advances in understanding and computing power. However, conventional approaches to GSA suffer from (1) an ambiguous characterization of sensitivity, and (2) poor computational efficiency, particularly as the problem dimension grows. Here, we identify several important sensitivity-related characteristics of response surfaces that must be considered when investigating and interpreting the "global sensitivity" of a model response (e.g., a metric of model performance) to its parameters/factors. Accordingly, we present a new and general sensitivity and uncertainty analysis framework, Variogram Analysis of Response Surfaces (VARS), based on an analogy to "variogram analysis", that characterizes a comprehensive spectrum of information on sensitivity. We prove, theoretically, that Morris (derivative-based) and Sobol (variance-based) methods and their extensions are special cases of VARS, and that their SA indices are contained within the VARS framework. We also present a practical strategy for the application of VARS to real-world problems, called STAR-VARS, including a new sampling strategy, called "star-based sampling". Our results across several case studies show the STAR-VARS approach to provide reliable and stable assessments of "global" sensitivity, while being at least 1-2 orders of magnitude more efficient than the benchmark Morris and Sobol approaches.

Characterizing Delta-Scale Process Connectivity Using Information Theory

Alicia Sendrowski[1], **Paola Passalacqua**[1]
[1] *University of Texas at Austin, Austin, TX*

In deltaic settings, hydrologic connectivity refers to the physical transport of water fluxes within the distributary channel network and onto deltaic wetlands. These fluxes also bring sediment and nutrients into the system, contributing to distinct vegetation structures and an evolving delta morphology. Another important facet of delta evolution is process connectivity- the interaction between deltaic variables and external forcings across temporal and spatial scales that is enhanced by the physical connectivity of the system. Process connectivity is essentially the flow of information between water, sediment, nutrients, and external forcings such as river discharge, wind, and tides within the delta. These relationships affect important deltaic functions, such as nitrate dynamics and sediment deposition patterns. In this work, we apply information theoretic principles to understand process connectivity across five different deltaic islands in Wax Lake Delta, a naturally prograding river delta in coastal Louisiana. Water level loggers collected water level and temperature measurements at six locations in the delta, over the course of three months in the summer of 2014. We use the mutual information and transfer entropy metrics to quantify relationships between water level, water temperature, and tides, wind, and river discharge. Mutual information is a measure of the synchronization between variables, while transfer entropy quantifies the directional flow of information between couplings. By using a moving window of time over the dataset, this method is able to capture the changes in the network, from event scale interactions, to seasonal relationships with river discharge. While discharge is a predominant force in the delta, wind and tides show significant influence at a range of temporal scales.

An entropy based quantification of delta channel network complexity

Alejandro Tejedor[1], **Anthony Longjas**[1], **Ilya Zaliapin**[2], **Efi Foufoula-Georgiou**[1]
[1]*National Center for Earth-surface Dynamics and St. Anthony Falls Laboratory, University of Minnesota, Minneapolis, MN, USA 55414*
[2]*Department of Mathematics and Statistics, University of Nevada, Reno, NV, USA 55414*

River deltas across the world exhibit an astonishing variety of morphologies in response to different forcings (e.g., river, tides and waves), sediment composition, incoming flow variability, sea level rise, etc. Understanding and quantifying the patterns imprinted on the landform would enable us to infer processes from observed imagery. Galloway [1975] introduced a qualitative diagram to classify deltas, showing how the balance of upstream (fluvial) and downstream (waves and tides) forcings dictates the delta form, depicted most distinctively in the coastline morphology. Recently, we presented a rigorous framework [Tejedor et al., 2015a,b] based on spectral graph theory to study delta channel networks, enabling us to extract important structural and dynamics-related information of river deltas. Using that information, we are able to introduce a suite of metrics to quantify channel network complexity, including entropic-based metrics measuring the complexity in terms of the uncertainty in the splitting and rejoining paths and fluxes, enhancing the comparison of deltas and process from form inference. Finally using the above mentioned framework, we are able to construct vulnerability maps that depict the relative change of sediment and water delivery to the shoreline outlets in response to possible perturbations in hundreds of upstream links. We show that an inverse relationship exists between entropy and vulnerability, reinforcing the idea that entropy is a surrogate of the capacity of the system to undergo changes.

Analysis of information correlation in geological models

Miguel De la Varga[1], **J. Florian Wellmann**[1]
[1]*Graduate School AICES, RWTH Aachen University*

The quantification and analysis of uncertainties is important in all cases where maps and models of uncertain properties are the basis for further decisions. Once these uncertainties are identified, the logical next step is to determine how they can be reduced. Information theory provides a framework for the analysis of spatial uncertainties when different subregions are considered as random variables. In the work presented here, joint entropy, conditional entropy, and mutual information are applied for a detailed analysis of spatial uncertainty correlations. The aim is to determine (i) which areas in a spatial analysis share information, and (ii) where, and by how much, additional information would reduce uncertainties. As an illustration, a typical geological example is evaluated: the case of a subsurface layer with uncertain depth, shape and thickness. Mutual information and multivariate conditional entropies are determined based on multiple simulated model realisations. Even for this simple case, the measures not only provide a clear picture of uncertainties and their correlations but also give detailed insights into the potential reduction of uncertainties at each position, given additional information at a different location. The methods are directly applicable to other types of spatial uncertainty evaluations, especially where multiple realisations of a model simulation are analysed. In summary, the application of information theoretic measures opens up the path to a better understanding of spatial uncertainties, and their relationship to information and prior knowledge, for cases where uncertain property distributions are spatially analysed and visualised in maps and models.

5 Workshop participants

Name	Affiliation	Email
Ahrens, Bodo	University of Frankfurt	bodo.ahrens@iau.uni-frankfurt.de
Branicki, Michal	Edinburgh University	m.branicki@ed.ac.uk
Brunsell, Nathan	University of Kansas	brunsell@ku.edu
de la Varga, Miguel	Aachen University	miguel.de.la.varga@rwth-aachen.de
Ehret, Uwe	KIT Karlsruhe	uwe.ehret@kit.edu
Foufoula-Georgiou, Efi	University of Minnesota	efi@umn.edu
Gharari, Shervan	TU Delft	S.Gharari@tudelft.nl
Gong, Wei	Bejing Normal University	gongw03@gmail.com
Goodwell, Allison	University of Illinois	allison@goodwell.net
Gupta, Hoshin	University of Arizona	hoshin@email.arizona.edu
Jackson, Bethanna	Victoria University, NZ	bethanna.jackson@vuw.ac.nz
Kim, Joon	Seoul National University	Joon@snu.ac.kr
Knuth, Kevin	University at Albany	kknuth@albany.edu
Kumar, Praveen	University of Illinois	kumar1@illinois.edu
Kumar, Rohini	UFZ Leipzig	rohini.kumar@ufz.de
Larsen, Laurel	U of California, Berkeley	laurel@berkeley.edu
McMillan, Hillary	NIWA, NZ"	hilary.mcmillan@gmail.com
Moore, Catherine	GNS Science, NZ	c.moore@gns.cri.nz
Nearing, Grey	NASA	grey.s.nearing@nasa.gov
Nijssen, Bart	University of Washington	nijssen@uw.edu
Pande, Saket	TU Delft	s.pande@tudelft.nl
Ruddell, Ben	Arizona State University	bruddell@asu.edu

Name	Affiliation	Email
Shafiei, Mojtaba	EWERI, Iran	moj.shafiei@gmail.com
Tejedor, Alejandro	University of Minnesota	alej.tejedor@gmail.com
Tennant, Chris	U of California, Berkeley	chris.tennant@berkeley.edu
Tian, Yudong (pending)	NASA	yudong.tian-1@nasa.gov
Vessilinov, Monty (pending)	Los Alamos National Lab	vvv@lanl.gov
Wang, Jingfeng	Georgia Tech	jingfeng.wang@ce.gatech.edu
Weijs, Steven	U of British Columbia	steven.weijs@ubc.ca
Wellmann, Florian	Aachen University	wellmann@aices.rwth-aachen.de