Calibrating stochastic models for understanding solar activity

Dr. Simone Ulzega
Senior Research Scientist,
ulzg@zhaw.ch

Research project
Bayesian Inference with Stochastic Models (BISTOM)

Lead:
Dr. Carlo Albert (EAWAG, ETHZ)

Role ZHAW:
Development of inference algorithms, parallelization for high-performance computing

Duration:
2 years
(01.04.2018–31.03.2020)

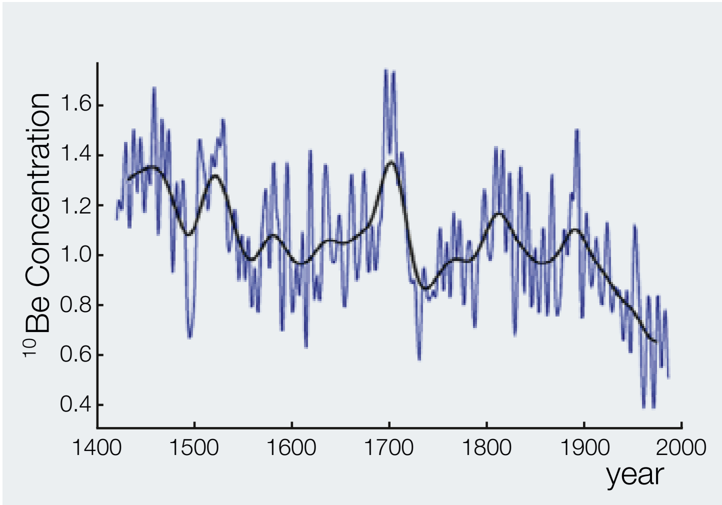
Partners:
SDSC, Eawag, ZHAW, Università della Svizzera Italiana (USI), Paul Scherrer Institute (PSI)

Funding:
Swiss Data Science Center (SDSC, ETHZ)

Research Group Biomedical Simulation

One of the most fundamental questions in essentially all applied sciences is how to predict the dynamics of complex systems. Complex systems are everywhere, in chemistry, biology, physics, engineering, economy, environmental, life and social sciences.

To tackle complexity, scientists often have to resort to simplified conceptual models that incorporate only a small selection of system variables and parameters. In such framework, only a few dominant dynamic processes occurring on our observation scale are described by deterministic differential equations, while all other processes are included in the model as *noise.* This naturally leads to *stochastic differential equation models.* Stochastic models take uncertainties intrinsic to dynamic processes into account thus providing more realistic descriptions of real systems. However, for making reliable probabilistic predictions, model parameters need to be soundly calibrated to measured data and their uncertainty needs to be quantified. Parameter inference, as this process is called, is a fundamental problem in data-driven modeling. Bayesian statistics is a consistent framework for parameter inference where knowledge about model parameters is conveniently expressed through probability distributions and updated using measured data. However, Bayesian inference with non-trivial stochastic models can become mathematically and computationally extremely challenging, and it is therefore hardly ever applied. In recent years, sophisticated and scalable algorithms have emerged, which have the potential of making Bayesian inference for complex stochastic models feasible, even for very large data sets. In the framework of a 2-year project funded by the Swiss Data Science Center (SDSC), in collaboration with Eawag, USI and PSI, we will explore the power and versatility of two classes of Bayesian inference algorithms, that is, Approximate Bayesian Computation (ABC) and Hamiltonian Monte Carlo (HMC) methods. The former is well-known and technically easy to apply but yields only approximate results, while the latter requires much more tailoring to a particular problem but has the potential of yielding exact results. The HMC algorithm, as recently proposed by Carlo Albert, Simone Ulzega and Ruedi Stoop (Albert et al. Phys. Rev. E 93, 2016), is raising great attention in various scientific communities due to its exceptional efficiency and high parallelizability. An efficient parallelization of the HMC algorithm, in particular, will be of paramount importance for making Bayesian inference amenable to a «Big Data» context.



Berillium-10 concentration (10^4 atoms/gram of ice) in polar ice cores, Greenland. Solar activity varies inversely with the concentration of the radioisotopes.

We will focus on a real case study in solar physics. Time-series of cosmogenic radionuclides, that is, radioactive Carbon-14 and Berillium-10 nuclei produced in the Earth's atmosphere by galactic cosmic rays and stored in wood and polar ice cores, are an exceptional proxy for solar activity on multi-millennial time-scales. Cosmic rays are in fact modulated by solar magnetic fields and the production rates of these isotopes is thus modulated by the solar magnetic activity. These time-series exhibit a number of interesting and mostly not-yet-understood features such as stable cycles and intermittency.

Solar physicists have put a lot of effort into the development of stochastic solar dynamo models, which need to be calibrated to the observations. Parameter inference for stochastic dynamo models on long time-series of radionuclides is currently an open and highly topical question in solar physics. Achieving more reliable predictions of the solar activity may have important implications also for our understanding of the Earth's climate. ■

Dreaming or ready to buy? Predictive analytics for online hotel booking platform

Dr. Krzysztof Kryszczuk, Head of Predictive Analytics Group, krys@zhaw.ch, **Dr. Stefan Glüge and Adrian Busin,** Research Associates

In the course of the last several years, the importance of data-driven facility management in the hospitality industry has been on a steady rise. Several leading hotel chains have made substantial investments into advanced analytics in order to better understand the needs and behavior of their customers. At the same time, the growing prevalence and popularity of chain-independent hotel booking platforms, such as booking.com, is exerting growing pressure on the hotel chains to find new ways of attracting directly-paying customers, and avoid sharing profit margins – which, in the hospitality industry, are already low (~5 %). According to a study by Google, the customers of the hospitality industry follow the cycle of «5 stages of travel», consisting of the Dreaming, Planning, Booking, Experiencing and Sharing phases (Fig.). Travelers manifest their activities corresponding to each of the 5 Stages through their online behavior. While the last two stages of Experiencing and Sharing are largely visible through social media, stages of Dreaming, Planning and Booking are directly linked to interactions with hotel booking platforms. In collaboration with our Implementation Partner, the company UCOB VENTURES AG from Steinhausen, Switzerland, researchers from the Predictive Analytics Group of the IAS deploy advanced analytics tools to model the online behavior of customers of several large international hotel chains, seeking to find patterns in individual paths from dreaming to booking and, eventually, purchasing. For instance, we found that the longer a customer is interacting with the booking site, the more likely he or she is to purchase. The potential to accurately identify the intention of the customer and their position in the 5 Stages of Travel cycle will drive a targeted personalization of the content of the online booking platform. The novel development of dynamic recommendation algorithms for the hospitality business are developed at IAS with the support from the CTI grant 19319.2 PFES-ES. ■

Fig.: Google's 5 stages of travel. Source: researchgate.net/figure/Googles-five-stages-of-travel_fig1_290172211

Komplexität in Industrie 4.0 beherrschen mit Simulation

Dr. Lukas Hollenstein, Dozent, Forschungsgruppe Simulation & Optimierung, hols@zhaw.ch

Die Umsetzung der Zukunftsvision Industrie 4.0 rückt für viele Unternehmen in greifbare Nähe. Gleichzeitig erhöht sich die Komplexität von Wertschöpfungsketten durch die zunehmende Vernetzung der Produktionsprozesse und durch die horizontale Integration der Organisationsstrukturen. Eine Erhöhung der Komplexität ist einerseits notwendig, um genügend Produktvarianten anbieten zu können, und ist somit ein Treiber für den Umsatz des Unternehmens. Andererseits treten dabei auch Komplexitätskosten auf. Gefragt ist daher ein Assistenzsystem für komplexe Entscheidungen im Kontext der Industrie 4.0, das Handlungsempfehlungen abgibt, z. B. auf die Frage, ob ein neues Produkt ins Portfolio aufgenommen werden soll oder ob es sich lohnt, die Prozessstrukturen auf Kosten der Produktvielfalt zu vereinfachen. Erste Schritte in Richtung einer Optimierung der Komplexität hat das Institut für Angewandte Simulation (IAS) im Rahmen des KTI-Projekts *Complexity 4.0* in Zusammenarbeit mit Partnern aus dem ZHAW-Datalab und dem Institut für Technologiemanagement (ITEM) der Universität St. Gallen unternommen. Dabei wurde eine Metrik für die Komplexität der Produktionsprozesse einer Supply Chain entwickelt. Anhand eines neuronalen Netzes wird die Komplexität direkt aus den Logfiles einer nach Industrie 4.0 vernetzten Fabrik berechnet. Traditionelle Komplexitätsmasse hingegen benötigen aggregierte Indikatoren, z. B. die Anzahl Produktvarianten und Anzahl Prozessschritte pro Produkt. Weil zum heutigen Zeitpunkt noch keine solchen Logfiles vorhanden sind, hat das IAS ein Simulationsmodell entwickelt, um die benötigten Daten zu erzeugen. Dabei werden komplexe Produktionssysteme (Fabrik oder Supply-Chain) nach realistischen Vorgaben zusammengewürfelt und die Auftragsdaten dazu generiert. Im Simulator werden die virtuellen Produktionsprozesse durchgerechnet und die Logfiles erzeugt. Mit den simulierten Daten konnte schliesslich ein erfolgreiches Proof of Concept der Komplexitätsmessung durchgeführt werden. ■

Neue Projekte

PhD Network in Data Science
Leitung: sven.hirsch@zhaw.ch
Dauer: 01.06.17–31.12.20
Beteiligte Departemente: N, T, W
Projektpartner: Swissuniversities, Bern; mitfinanziert durch SBFI, Bern

Umsetzungshilfe zum Thema Waldrandaufwertungen 2017–2019
Leitung: petra.lustenberger@zhaw.ch und manuel.babbi@zhaw.ch
Dauer: 01.10.17–31.12.19
Beteiligte Institute: IAS, IUNR
Projektpartner: Bundesamt für Umwelt BAFU, Bern

simCAT: A mobile aid to recombinant enzyme production and whole-cell biocatalysis using microorganisms
Leitung: roland.gassmann@zhaw.ch
Dauer: 01.12.17–31.01.19
Beteiligte Institute: IAS, ICBT; mitfinanziert durch SBFI, Bern

Qualität 4.0 – IoT System für Qualitätsmanagement und Lebensmittelsicherheit
Leitung: sven.hirsch@zhaw.ch
Dauer: 01.12.17–31.12.19
Beteiligte Institute: IAS, ILGI
Projektpartner: Axino Solutions AG, Solothurn; Genossenschaft Migros Zürich, Zürich; mitfinanziert durch Innosuisse (KTI), Bern

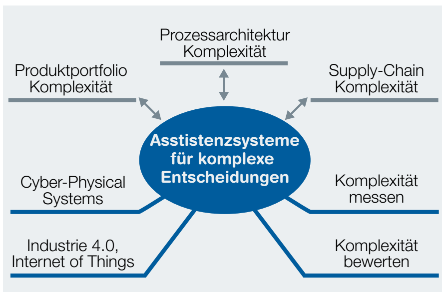
Frequentist estimation of the evolutionary history of sequences with substitutions and indels
Leitung: maria.anisimova@zhaw.ch
Dauer: 01.02.18–01.03.23
Projektpartner: Schweizer Nationalfonds SNF, Bern

Weitere Projekte
zhaw.ch/ias/projekte

Weiterbildung

Herbstsemester
CAS Digital Basics for Life Sciences

Infos und Anmeldung
zhaw.ch/ias/weiterbildung



Diverse Komplexitätstypen (oben), Datenquellen (links) und Schritte (rechts) im Kontext des Komplexitätsmanagements.