AI for understanding processes in intertwined Earth System Dynamics - break-out session @ phi-week

Organizers and Moderation:
Miguel Mahecha / Leipzig University - RSC4Earth.de / @miguelmahechag1
Gustau Camps-Valls / Valencia Universitat/ IS / @isp_uv_es
Carsten Brockmann / Brockmann Consult GmbH / @BrockmannCon

Motivation

Artificial Intelligence (AI) is advancing many fields of Sciences, and of Earth sciences in particular. The suite of AI methods allows us to revisit classical problems in the Earth system sciences (ESS). These techniques allow us to solve challenging problems like classification, regression or anomaly detection with unprecedented accuracy. Advances are so many and applications so exciting that people have advocated that AI relying on Big Data is replacing Science and Education to some extent. The current mantra is that enough data, fast computers and scalable algorithms suffice to solve whatever data analysis problem and there's no need to understand the first physical principles. But is this as simple as that? Is this really the frontier of science? We believe that much more fundamental epistemic questions need to be formulated and answered, before grasping the full potential of AI for advancing our understanding of the Earth and its intertwined subsystems. This session aimed to explore the conceptual frontiers of AI methods for Earth science applications.

The questions raised here emerged in the final meeting of the https://www.earthsystemdatalab.net/ project funded by the ESA earlier in 2020 and are highly relevant in light of the emerging Digital Twin Earth initiative and the underlying EU strategies around "Destination Earth". We note that not only Earth system scientists are seeking methodological orientation, but also Computer Sciences are increasingly discovering the field of Earth system research as an equally relevant field of application than earlier dedicated to engineering, medical sciences, or natural language processing as it has become evident in the foundation of the European Lab for Learning and Intelligent Systems (ELLIS), and in the the ELLIS Program on Earth and Climate Sciences in particular.

Emerging Topics

The session turned out to be a very lively discussion event, with 10 "lightning talks" by both, very young and very experienced researchers in the field. The session had constantly between 80 and 100 online participants, and we collected multiple questions and had lively discussions From these talks (the full list of talks is provided below), comments and discussions, we noted that a few recurrent topics need to be followed-up in depth:

1. **Equip ESS against naive applications of AI methods:** Many talks (e.g. by Hanna Meyer, Martin Jung, Ribana Roscher, Peter Dueben) called for cautionary application of AI methods. In particular, Meyer showed how critical it is to allow for predictions far

beyond training samples; machine learning models are weak in extrapolations and many recent high-impact papers inexcusably neglected this fundamental principle leading to dubious results that undermine the general reputation of AI and Earth system research. Identifying when machines operate in extrapolation regimes is important, but quantifying what's the extrapolation level (and how to address it) becomes a big challenge nowadays. Another aspect to overcome naive (miss)uses of AI methods is to exactly understand the reasons for their results e.g. a classification result. Ribana Roscher showed clearly that today one can indeed interpret and understand why a certain result is provided from some Deep Learning (DL) algorithm, a field collective known as eXplainable AI (XAI).

2. **Learn from AI what you don't know from first principles:** ESS has made its major breakthroughs via it's predictive capacity, i.e. in climate research based on models that encode our understanding of the underlying dynamics. AI - and this seems to be a consensus - can complement this domain knowledge via different approaches. It was shown by e.g. Fabian Gans et al. and Basil Kraft et al. how such approaches could be complemented when "one part of the equation is unknown". Hybrid modelling is still in its infancy, but certainly the bridge that allows to build on the existing, but constrain it with the unknown such that better prediction will become possible. Hybrid modeling allows us to tackle the problem of extrapolation in data-limited regimes. Where physical knowledge is incomplete, latent variables in e.g. some Deep-Learning model may compensate. Interestingly, however, data-scaristy on ground-truthing emerged as one caveat limiting the potential of such learning approaches. Later it was shown by Jonas Peters that today it is even thinkable to learn Ordinary Differential Equations (ODEs) underlying some complex system from observing its temporal dynamics. "Reverse Engineering" and "Automated Machine Scientists" are the keywords in this context.
3. **Towards causal inference from Earth observations:** The holy grail in the Earth system sciences remains to obtain causal understanding from observations, along with hypotheses and constraints. Peter Watson raised the important question whether we can learn to predict (and to a certain extent to extrapolate) from data sampled from an ever-evolving system? This is also a fundamental motivation for Sebastian Sippel et al., who, for instance, showed the potential of new ML approaches for fortifying detection and attribution questions. In particular they asked whether it is possible to detect externally forced climate changes in observed climate variables. If this is possible, it would be a crucial step towards causal inference from observed data. Along these lines Rune Christiansen then discussed also the issue of spatial confounders in Earth observations and how to control for them to infer general causal mechanisms. All methods in the domain of causal inference have huge potential but are in their infancy and represent a clear frontier of science that needs to be pushed further.

Take-home message

The session ended in a general discussion and advocacy for combining domain sciences to develop customised solutions. However, Peter Deuben finally also made the point that AI can theoretically learn everything from data but(training-)data itself is often the limiting factor. Hence,

one may regard all the above approaches also as an intermediate step towards a much more complete AI-driven understanding of the Earth system that shall be the guiding paradigm as long as we are data-limited - we note that this is at glance a paradoxical statement, but remains to the case until Digital Twin Earth (DTE) is finally fully operational. But even then, the issue of "interpretability" and "causality" will certainly remain the driving force behind AI-empowered ESS. Gustau Camps-Valls finally noted that the three pathways (hybrid modeling, explainable AI, and causality) strictly need (and are based on) active interdisciplinary collaboration where computer scientists and domain experts work together; a common language and emphasis on education of the younger generations will be mandatory for a real, non-incremental advance of the Earth sciences with AI.

The following page is the final programme. Because of the nature of the event being composed of a series of "lighting talks" the discussion partly needed to be extended beyond the actual discussion time and was documented here: https://docs.google.com/document/d/1cPgn7c_-8P3uH2Gv8RmNWHa-6WjKB0pR-e1cuLB8z_0/edit?usp=sharing

PROGRAMME:

9:30 AM Opening, greetings & instructions

- **Miguel Mahecha** - Leipzig University
- **Gustau Camps-Valls** - University of Valencia
- **Carsten Brockmann** - Brockmann Consult GmbH

SESSION 1 - 09:45 - 10:30 EARTH NEEDS OUR HELP - SOME IDEAS TO GET IT DONE

- **Hanna Meyer** - University of Münster - *"Area of applicability of remote sensing-based prediction models"*
- **Fabian Gans** - Max Planck Institute for Biogeochemistry - *"Hybrid modelling gets us closer to understand fundamental parameters of ecosystem functioning - one example"*
- **Basil Kraft** - Max Planck Institute for Biogeochemistry - *"Hybrid modelling the global water cycle"*
- ***Questions+Discussion (15')***

SESSION 2 - 10:30 - 11:15 UNDERSTANDING IS HARDER THAN FITTING: XAI & CAUSALITY

- **Rune Christiansen** - University of Copenhagen - *"Inferring causal links in spatio-temporal data - Conflict and forest loss"*
- **Sebastian Sippel** - ETH Zürich - "*Making detection and attribution robust to multi-decadal internal variability through statistical learning based robustness constraints"*

- **Peter Watson** - Bristol University -*"How important is atmospheric process understanding for projecting climate change impacts?"*
- ***Questions+Discussion (15')***

---- BREAK ---- 11:15 - 11:45

SESSION 3 - OUTLOOK 11:45 - 12:40

- **Ribana Roscher** - University of Bonn - *"Perspectives of explainable AI for Earth system research"*
- **Martin Jung** - Max Planck Institute for Biogeochemistry - *"AI is perfect to prevent and procrastinate the hard logical reasoning needed for enhancing process understanding"*
- **Jonas Peters** - University of Copenhagen - *"Learning stable and predictive structures in kinetic systems"*
- **Peter Dueben** - ECMWF - *"The atmosphere is not a cat video -- on the need to blend machine learning and domain sciences to build customised solutions"*
- *Questions+Discussion (15')*

12:40-13:00 Wrap-up discussion