The state of the earth: estimating physical parameters from noisy and incomplete earth observation data

Arp, L.R.

Summary

Estimating physical parameters from observations is a core theme in AI for science, with broad relevance across diverse scientific fields and real-world applications. These parameters, which define the state of physically governed systems, are often not directly observable: we can only infer them from their observed outcomes. In this thesis we focus on developing AI solutions for reliably estimating Earth system parameters from satellite data, which are critical for, e.g., environmental protection, disaster response, and agriculture. Two core challenges complicate this estimation: 1) spatio-temporal gaps in the available data hamper model training and consistency, and 2) multiple physical states of the parameters can produce the same observations, leading to multiple possible solutions for an observation. In this dissertation, we address four research questions, two of which are related to the first challenge and two others to the second.

The first research question addresses spatio-temporal gaps in ground truth datasets and how to fill them in effectively. In Chapter 3, we propose a novel spatial interpolation method to mitigate this issue. Ground truth data, representing the 'true' parameter values that we need to estimate, often covers only sparse measurement points, while we require a complete map of predictions at every point in an area. Inspired by Markov reward processes, our method iteratively propagates the available information through a system of unknown values, producing an interpolated grid of ground truth values. We demonstrate the strength of our proposed method on simulated data and real-world GDP and COVID-19 datasets.

The second research question addresses spatio-temporal gaps in satellite data, such as those caused by cloud cover. In Chapter 4, we propose a novel method that adapts our spatial interpolation approach from Chapter 3 for image reconstruction. Our method uses a fully observable, but outdated, reference image to guide the reconstruction process. We demonstrate its effectiveness on a popular cloud removal dataset and a diverse dataset that we collected, which we make publicly available for further scholarly work.

The third research question explores multiple solutions in parameter estimation. While using a common radiative transfer model (RTM) inversion approach, we unexpectedly found only a single solution, contradicting prior expectations. In Chapter 5, we present an empirical study on RTM-simulated data to investigate how problems with a single solution might still exhibit multiple solutions in practice. Our analysis shows that the problem meets all the criteria of well-posedness (i.e., a unique solution exists), and experiments on real-world satellite data confirm that this property is retained in non-simulated settings. Further experiments revealed that observation noise is likely the primary cause of multiple apparent solutions. Even if a unique solution exists for a given observation, errors in the observation itself can lead to incorrect solutions.

The final research question addresses computing the set of solutions for parameter estimation problems. In Chapter 6, we formalise and theoretically support a framework to make this computation tractable, and we propose a method based on constrained black-box optimisation to automatically approximate the set of possible solutions. This solution set enables a wide array of novel analyses and applications. We validate both the theoretical framework and our practical approximation method through empirical experiments across physical, statistical, and machine learning models.

In conclusion, in this dissertation we address two core challenges in Earth system parameter estimation: spatio-temporal data gaps and multiple solutions for the same observations. Through four distinct contributions contained in Chapters 3—6, we advance the understanding of these challenges and improve the reliability of estimation methods. While the impact of these challenges likely cannot be eliminated altogether within the scope of a single dissertation, our work provides tangible advances and actionable pointers for future research. We encourage other scholarly work to build on these findings and contribute toward interdisciplinary efforts aimed at estimating the state of the Earth.