Report from Dagstuhl Seminar 23342

Computational Geometry of Earth System Analysis

Susanne Crewell*[1], Anne Driemel*[2], Jeff M. Phillips*[3], and Dwaipayan Chatterjee†[4]

1 **Universität Köln, DE.** susanne.crewell@uni-koeln.de
2 **Universität Bonn, DE.** driemel@cs.uni-bonn.de
3 **University of Utah – Salt Lake City, US.** jeffp@cs.utah.edu
4 **Universität Köln, DE.** dchatter@uni-koeln.de

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 23342 "Computational Geometry of Earth System Analysis". This seminar brought together experts of algorithms and the Earth sciences to foster collaborations that can tackle algorithmic problems in the Earth system by the crossover of expertise in these different areas. The Earth sciences include a manifold of disciplines that deal with atmospheric, oceanic and terrestrial observations to further our understanding of climate processes. New generations of observation systems that are being developed right now provide novel data about the atmospheric and surface conditions at increasing spatial and temporal resolution. This provides unique information to improve weather and climate prediction but cannot always be handled by traditional numerical models. Computational Geometry is rooted in a strong tradition of algorithm and complexity analysis applied to practical geometric problems. Efficient algorithmic methods developed in this field are often tailored to the low-dimensional geometric settings that arise in a multitude of application areas, but have until recently not been applied to problems arising in the Earth system sciences – and in particular not in meteorology.

Seminar August 20–25, 2023 – https://www.dagstuhl.de/23342
2012 ACM Subject Classification Theory of computation → Computational geometry; Theory of computation → Data structures and algorithms for data management
Keywords and phrases Data reduction, Event detection, Feature tracking, Geometric algorithms, Interpolation methods, Sensor placement
Digital Object Identifier 10.4230/DagRep.13.8.91

1 Executive Summary

Anne Driemel (Universität Bonn, DE)
Susanne Crewell (Universität Köln, DE)
Jeff M. Phillips (University of Utah – Salt Lake City, US)

Various disciplines within the Earth sciences deal with measuring and representing the geometry of the Earth's land and sea surface, as well as atmospheric and oceanic conditions, to further our understanding of dynamic processes occurring on the Earth. All around the world society and economy is becoming more and more vulnerable to changing weather. The recent extreme precipitation and flooding in western Germany and Libya, the strong tornado in the Czech Republic or dry spells leading to severe fires in Canada are just few of

* Editor / Organizer
† Editorial Assistant / Collector

many examples which illustrate how anthropogenic climate change is going to influence our weather. Climate describes the statistics of all-weather events and is typically predicted using physical models representing the relevant processes in the Earth system. Understanding these processes is paramount to anticipating and addressing the challenges posed by climate change today. A key aspect of meteorological research and Earth sciences in general is to develop methods to turn atmospheric, oceanic and terrestrial observations into regional information for weather and climate prediction. The observations are being collected with different types of sensors at increasing spatial and temporal resolution which poses computational challenges to meteorologists that cannot always be addressed with traditional methods. More and more observation systems (e.g. from commercial aircraft or new satellite series including radio occultation) but also opportunistic crowd-sourced measurements are currently exploited. In addition, there are ground-based remote-sensing networks for operational atmospheric profiling, which can shed light on small-scale processes such as turbulence and cloud physics that cannot be resolved in detail from satellites, but can be used to improve model parameterizations. Thus, on the horizon there is a wealth of new, voluminous observation systems providing unprecedented possibilities for improving weather and climate models, but require innovative and explorative approaches in the areas of data handling, data assessment, information extraction, and data assimilation. The field of computational geometry is concerned with the design, analysis, and implementation of efficient algorithms for geometric and topological problems, which arise naturally in a wide range of application areas. Computational geometry is a vibrant and mature field of research, with several dedicated international conferences and journals and strong intellectual connections with other computing and mathematics disciplines. Within computer science and mathematics, computational geometry lies in the intersection of the theory of algorithms and combinatorial geometry. Despite its theoretical nature, the research in this field is strongly oriented towards and motivated by concrete practical problems that arise in various application areas that deal with geometric data. The related emerging field of geometric data analysis deals with the efficient statistical analysis of geometric data by providing sketches and data summaries with provable guarantees.

Outcome

This Dagstuhl Seminar brought together 26 researchers from the fields computational geometry and the Earth sciences to provide a forum to discuss the unique computational challenges that need to be dealt with and how the geometry underlying the input data can be exploited to obtain efficient algorithms. The 5-day seminar was initially focused on three problem areas (1) data assimilation of weather-related measurements for numerical simulation, (2) tracking and clustering of moving atmospheric features, and (3) the planning and optimization of sensor placements. In addition to these topics, seminar participants contributed research questions from the current state of the art in their fields. There were 13 research talks throughout the seminar. Several longer talks gave an overview over the proposed seminar topics, either from the point of view of Earth scientists or from the point of view of algorithmic techniques. Organized group discussions resulted in several breakout groups on emerging topics. Each group was composed of a mix of participants from the different fields. In addition, the program left ample room for discussions and for new research directions to emerge.

The seminar was seen as a success by organizers and participants. This is nicely summarized by a quote from the survey:

"This seminar brought together a highly motivated group of excellent researchers. It felt as if all of them wanted to make the best use of this seminar to make a difference, in an interdisciplinary application domain. Since this was an interdisciplinary seminar, there was certainly a natural spread in terms of how willing people were to learn about new topics and/or to (re-)explain their work/ideas to researchers from other fields. However, overall this worked well and the atmosphere stimulated new ideas and interdisciplinary exchange."

2 Table of Contents

3 Overview of Talks

3.1 Partitioning a Polygon Into Small Pieces

Mikkel Abrahamsen (University of Copenhagen, DK)

Joint work of Mikkel Abrahamsen, Nichlas Langhoff Rasmussen
Main reference Mikkel Abrahamsen, Nichlas Langhoff Rasmussen: "Partitioning a Polygon Into Small Pieces", CoRR, Vol. abs/2211.01359, 2022.
URL https://doi.org//10.48550/ARXIV.2211.01359

We study the problem of partitioning a given simple polygon P into a minimum number of connected polygonal pieces, each of bounded size. We describe a general technique for constructing such partitions which works for several notions of 'bounded size,' namely that each piece must be contained in a unit square or unit disk, or that each piece has bounded perimeter, straight-line diameter or geodesic diameter. The problems are motivated by practical settings in manufacturing, finite element analysis, collision detection, vehicle routing, shipping, laser capture microdissection and perhaps earth system analysis.

It seems out of reach to compute optimal partitions; even in extremely restricted cases such as when P is a square. Our main result is to develop constant-factor approximation algorithms, which means that the number of pieces in the produced partition is at most a constant factor larger than the cardinality of an optimal partition. Existing algorithms [Damian and Pemmaraju, 2004] do not allow Steiner points, which means that all corners of the produced pieces must also be corners of P. This has the disappointing consequence that a partition does often not exist, whereas our algorithms always produce meaningful partitions. Furthermore, an optimal partition without Steiner points may require $\Omega(n)$ pieces for polygons with n corners where a partition consisting of just 2 pieces exists when Steiner points are allowed. Other existing algorithms [Arkin, Das, Gao, Goswami, Mitchell, Polishchuk and Tóth, 2020] only allows P to be split along chords, whereas we make no constraints on the boundaries of the pieces.

3.2 Flood Risk Analysis on Terrains

Pankaj Kumar Agarwal (Duke University – Durham, US)

An important problem in terrain analysis is modeling how water flows across a terrain and creates floods by filling up depressions. This talk discusses a number of flood-risk related problems: Given a terrain T, represented as a triangulated xy-monotone surface, a rain distribution R, and a volume of rain V, determine which portions of T are flooded and how water flows across T. Efficient algorithms are presented for flood-risk analysis under both single-flow-direction (SFD) as well as multi-flow-directions (MFD) models – in the former, water at a point can flow along one downward slope edge while in the latter, it can flow along multiple downward slope edges; the latter more accurately represent flooding events but it is computational more challenging.

3.3 Algorithms for Movement Analysis

Maike Buchin (Ruhr-Universität Bochum, DE)

Nowadays a lot of movement data of animals, people and weather phenomena are being collected. To analyse this data requires efficient geometric algorithms. Some fundamental tasks such as clustering, segmentation, and simplification occur often in this. I will discuss three of these problems: similarity, clustering, and grouping. Each of these, I will motivate by movement analysis, sketch the algorithmic challenges of solving these, and present some experiments of our algorithms on concrete movement data sets.

3.4 Earth System Modelling

Peter Dueben (ECMWF – Bonn, DE)

I will provide an overview on Earth system modelling discussing the ten most relevant questions from Earth system modelling for Computational geometry. This includes what the Earth system is, how it is modelled, what good Earth system models are, how good they are, how uncertainty is represented, what geometry the models use, how they use high performance computing, and what machine learning will change in Earth system modelling in the future.

3.5 Between Drought and Heavy Rain: Digital Methods for Extreme Water Events

Sándor Fekete (TU Braunschweig, DE)

URL https://www.exdimum.org

In the course of global climate change, not only the averages of precipitation are shifting, but also the extremes: Drought and local heavy rainfall events have become more pronounced. This development leads to challenges that occur simultaneously on different scales. Extreme drought is a long-term process that extends over large regions; conversely, floods are often short-term and localized. It is important that the extremes do not balance out, but often reinforce each other.

In this talk, we give an overview of a number of different aspects related to extreme water events. In particular, we describe challenges and methods within the interdisciplinary research cluster EXDIMUM, which is part of a federal funding program WaX that targets extreme water events.

3.6 A network approach to cloud organization

Franziska Glassmeier (TU Delft, NL)

Joint work of Franziska Glassmeier, Graham Feingold
Main reference Franziska Glassmeier, Graham Feingold: "Network approach to patterns in stratocumulus clouds", Proceedings of the National Academy of Sciences, Vol. 114(40), pp. 10578–10583, 2017.
URL https://doi.org//10.1073/pnas.1706495114

We discuss a network analysis of the organization of cloud constellations based on data from detailed cloud simulations (LES) of a stratocumulus cloud deck. We find the network structure to be neither random nor characteristic to natural convection. It is independent of the typical length scale (atmospheric boundary layer height). The latter is a consequence of entropy maximization (Lewis's Law with parameter 0.16). Non-6 sided cells occur according to a neighbor-number distribution variance of about 2. Reflecting the continuously renewing dynamics of Sc fields, large (many-sided) cells tend to neighbor small (few- sided) cells (Aboav-Weaire Law with parameter 0.9). By developing a heuristic model, we show that stratocumulus cell dynamics can be mimicked by versions of cell division and cell disappearance and are biased towards the expansion of smaller cells.

3.7 Integrated Modeling of Terrestrial Systems

Stefan Kollet (Forschungszentrum Jülich, DE)

Groundwater-to-atmosphere simulations are useful for closing the terrestrial water and cycle at the continental scale and interrogating anthropogenic impacts at different space and time scales. We want to understand how human water use, in this case groundwater pumping and irrigation, changes the natural terrestrial cycles over the European continent including local effects, such as changes in water table depths, evapotranspiration, and air temperature, and non-local effects, such as base flow, continental discharge and precipitation. In particular, we study whether these changes are systematic in space and time, and ultimately impact and potentially redistribute water resources across the continent. We present technical aspects of our work related to model coupling and high-performance computing technologies, and results illustrating the significant impact of human water use beyond individual watersheds.

3.8 SCALGO Live: Computational geometry algorithms and data structures in practice

Thomas Moelhave (SCALGO – Aarhus, DK)

By harnessing the power of global geographical digitisation and using the latest advances in big data processing technology, SCALGO builds innovative digital tools that enable our users to create better, more liveable and sustainable environments where there is space for water.

Our technology is based on decades of basic and applied research in algorithms for processing big geographic data at leading universities in both Europe and the USA. In this talk I will discuss some of the technical challenges we have faced in transitioning from basic computer science research to a powerful user-friendly product for users with varied backgrounds.

3.9 Machine learning for climate science and meteorology: new opportunities for computational geometry to shine?

Peer Nowack (KIT – Karlsruher Institut für Technologie, DE)

Machine learning techniques have gained widespread use in meteorology and climate science. In this talk, I will put forward a few ideas on how computational geometry could help address key methodological and scientific challenges in these application domains. In particular, computational geometry could lead to better input feature design for machine learning applications, or could contribute to the development of new methods to detect and track important weather phenomena.

These ideas could open up new routes for close, solution-oriented, collaborations between computer scientists and domain scientists. For example, I will suggest ways in which such approaches could improve results of machine learning methods to observationally constrain climate change projection uncertainties. I will further discuss important atmospheric dynamical phenomena which are currently still difficult to identify automatically with both high precision and recall. Here, I will emphasize the potential of methods from computational geometry to overcome longstanding event definition and detection barriers. A related advance would be the creation of a wider collection of new event-focused datasets that dynamically follow and characterize important weather systems, to be made available for future use in event-focused studies by domain scientists. Motivating examples include Ceppi and Nowack (2021) and Thomas et al. (2021).

References

1 Thomas, C., Voulgarakis, A., Lim, G., Haigh, J., and Nowack, P. An unsupervised learning approach to identifying blocking events: the case of European summer, Weather and Climate Dynamics 2, 581 – 608 (2021). `https://doi.org/10.5194/wcd-2-581-2021`

2 Ceppi P. and Nowack P. Observational evidence that cloud feedback amplifies global warming, PNAS 118, e2026290118 (2021). `https://doi.org/10.1073/pnas.2026290118`

3 Nowack, P., Ceppi, P., Davis, S.M. et al. Response of stratospheric water vapour to warming constrained by satellite observations. Nature Geoscience 16, 577 – 583 (2023). `https://doi.org/10.1038/s41561-023-01183-6`

3.10 Challenges in analyzing and evaluating next-generation (km-scale) global climate simulations

Vera Schemann (Universität Köln, DE)

In global climate modelling a new generation of models is about to become feasible. By approaching km-scale resolutions, more information about regional and temporal variability of a changing climate system becomes possible. Furthermore, an improved representation of clouds and precipitation is expected and with that more insights into extreme events and their potential changing frequency and properties could be provided. But this new generation is also posing new challenges: The pure amount of data as well as the chaotic structure of the climate system will challenge our traditional way of analyzing climate model output or make it even impossible. New methods for statistical comparisons or identification of specific cluster and pattern have to be developed or adapted. This talk gave a short overview about this topic by starting with our traditional models and methods, introducing ongoing projects on developing new models and workflows and formulating upcoming questions and challenges.

3.11 Coresets for Data Reduction

Christian Sohler (Universität Köln, DE)

After a gentle introduction into analysis of algorithms, I survey some results from the area of coresets in the context of clustering. A coreset is a small summary of a data set with respect to an optimization problem. In the context of clustering, the input is a set of points and a coreset is a much smaller point set such that every solution to the clustering problem has roughly the same objective value for the coreset and the original point set. After introducing the basic definitions, I explain some applications of coresets and techniques to compute them.

3.12 Data-driven stochastic modelling of unresolved scales for weather and climate models

Nikki Vercauteren (University of Oslo, NO)

Limited computer resources lead to the need for a reduced representation of the weather and climate system when studying its dynamical behaviour. We discussed different approaches used to reduce the complexity of atmospheric models. One standard approach is to divide the dynamics into slow, resolved processes for which the fluid dynamical and thermodynamics equations are solved on a numerical grid, and fast or unresolved processes that are described through a simplified model or parameterisation. Another strategy is to simplify the model itself by projecting the equations of motion onto a smaller dimensional state space. We discussed alternative, data-driven strategies to identify clusters, or coarse-grain patterns, in atmospheric datasets, and how to use such coarse-grain patterns, or atmospheric regimes, to inform model reduction approaches. Finally we discussed how to combine such a clustering framework with a stochastic modelling strategy. In this approach, a stochastic parameterisation of unresolved processes is learned from observations. In the learning process, the underlying, unobserved atmospheric regimes that lead to different dynamics of the unresolved variables to be modelled are clustered. The resulting parameterisation of the unresolved or fast degrees of freedom is hence adapted to the occurring atmospheric regimes. The example of application analysed turbulence modelling in context where turbulent mixing can be very intermittent. The stochastic formulation accommodates both the short-term intermittent behavior of turbulent mixing and the long-term average mixing. It could present a way forward for dealing with the complexities of unsteady flows in numerical weather prediction or climate models.

3.13 Topological Characterization and Uncertainty Visualization of Atmospheric Rivers

Bei Wang Phillips (University of Utah – Salt Lake City, US)

Joint work of Fangfei Lan, Brandi Gamelin, Lin Yan, Jiali Wang, Hanqi Guo, Bei Wang

Atmospheric rivers (ARs) are long, narrow regions in the atmosphere that transport water vapor from the Earth's tropics. ARs have been of great interest to climate scientists because they cause a large percentage of precipitation worldwide. In North America, ARs contribute significantly to water supply and flooding risk, especially in the western regions. However, ARs are difficult to characterize due to the lack of a universal definition and their varying shapes and sizes. Many AR detection tools (ARDTs) have been developed for different purposes, producing distinct AR boundaries. In this work-in-progress, we study the ARs detected by an ensemble of AR detection algorithms, quantify and visualize the uncertainty that arises from the output of these algorithms. We propose an uncertainty visualization framework that captures both the exterior and interior variability of an ensemble of ARs. This is build upon a collaboration with the Argonne National Laboratory (ANL). This research is partially supported by grants DOE DE-SC0021015 and NSF IIS-2145499.

4 Working groups

4.1 Sensor placement and path planning problems

Mikkel Abrahamsen (University of Copenhagen, DK), Sándor Fekete (TU Braunschweig, DE), Jürgen Kusche (Universität Bonn, DE), Petra Mutzel (Universität Bonn, DE), and Vera Schemann (Universität Köln, DE)

Problems of and methods for geometric optimization play an important role in the interface between application areas and fundamental theory. In this work group, we discussed a number of such questions. In particular, we discussed the problem of planning flight paths with a Lagrangian focus (sampling the same air mass several times) and the issues rising during the comparison of measured and simulated parameter. Furthermore, we learned about the problems of placing stationary and floating sensors to measure rising sea levels.

4.2 Tracking of cloud fields

Franziska Glassmeier (TU Delft, NL), Kevin Buchin (TU Dortmund, DE), Dwaipayan Chatterjee (Universität Köln, DE), Sándor Fekete (TU Braunschweig, DE), Siddharth Gupta (University of Warwick – Coventry, GB), André Nusser (University of Copenhagen, DK), Frank Staals (Utrecht University, NL), and Bei Wang Phillips (University of Utah – Salt Lake City, US)

Atmospheric phenomena evolve internally, while at the same time being transported by the prevailing winds. The study of internal evolution is simplified by isolating it from transport, i.e., by studying the Lagrangian evolution of atmospheric phenomena along trajectories. We focused on cloud fields, i.e., constellations of many clouds in spatial proximity. The state-of-the-art approach for obtaining trajectories of cloud fields makes use of observationally constrained, simulated wind fields (reanalysis data). Different, purely-observational approaches for identifying the Lagrangian evolution of atmospheric phenomena through tracking are used within the Atmospheric Sciences. We discussed methods for identifying the Lagrangian evolution of cloud fields by tracking constellations of clouds in time-series of satellite images. Based on a range of example case studies from the tropical Atlantic, we identified two different strategies. Both approaches rely on tracking cloud fields as represented by the scalar field of cloud-optical depth and do not require a segmentation into cloud objects and surrounding cloud-free space.

The first approach is mostly topological with some geometric hybridization. It relies on keeping track of specific features of the cloud field (like local maxima of the intensity distribution). This results in a relatively fast and scalable approach for mapping the shape development, with readily available code by some of the participants. The cost is that this feature-based approach relies on some simplifications that may potentially cause discontinuities, so some post-processing of outcomes is indicated.

The second is purely geometric and relies on an Earth mover's distance. Specifically, it uses a weighted matching between consecutively frames to the relocation of the geometric distribution itself. This promises to more closely map the continuous process of shape development, in particular when for the relocation of the cloud field an underlying translation is assumed (either one for the whole field or local translations guided by optical flow). At this point, the challenge is to make the algorithmic realization scalable enough to make it useful for larger data sets.

Thanks to an existing implementation, the topological approach could already be tested during the Dagstuhl Seminar and showed promising results. The implementation of the geometric approach was also started during the seminar. Going forward, we have made arrangements to compare our two approaches with each other and to state-of-the-art approaches in cloud research. This comparison will introduce new methodology from computational geometry to cloud research. In addition, it will establish a benchmarking dataset and metrics and thus provide the foundation for future improvements.

4.3 Detection of blocking events

Peer Nowack (KIT – Karlsruher Institut für Technologie, DE), Maike Buchin (Ruhr-Universität Bochum, DE), Susanne Crewell (Universität Köln, DE), Anne Driemel (Universität Bonn, DE), Jan-Henrik Haunert (Universität Bonn, DE), Benjamin Raichel (University of Texas at Dallas – Richardson, US), Melanie Schmidt (Heinrich-Heine-Universität Düsseldorf, DE), and Bei Wang Phillips (University of Utah – Salt Lake City, US)

Atmospheric blocking events are key drivers of extreme weather in the mid-latitudes, including Europe and the US. The high-pressure systems "block" the normal zonal flow of air masses, and are often associated with major heatwaves during summer and cold spells during winter. Blocking patterns are also coupled to low-pressure systems and extreme rainfall in other regions. For example, this was the case in September 2023 when extreme flooding in Greece was coupled to a heatwave over Central to Northern Europe. An important science question is how the frequency and character of blocking events might change in the future in the presence of anthropogenic climate change. For this purpose, climate scientists need to analyse blocking events in century-scale climate model simulations, which requires automatic blocking detection methods.

In this break-out group, we discussed how computational geometry approaches could use a recently published ground-truth dataset for European summer blocking (Thomas et al., 2021) to develop more reliable blocking detection methods, compared to current blocking indices (which commonly disagree in their detection).

References

1 Thomas, C., Voulgarakis, A., Lim, G., Haigh, J., and Nowack, P. An unsupervised learning approach to identifying blocking events: the case of European summer, Weather and Climate Dynamics 2, 581 – 608 (2021). https://doi.org/10.5194/wcd-2-581-2021

4.4 Data reduction

Nikki Vercauteren (University of Oslo, NO), Dwaipayan Chatterjee (Universität Köln, DE), Peter Dueben (ECMWF – Bonn, DE), Siddharth Gupta (University of Warwick – Coventry, GB), Jeff M. Phillips (University of Utah – Salt Lake City, US), and Christian Sohler (Universität Köln, DE)

Despite the complexity of the climate system, the existence of repeating and persisten patterns or weather regimes is known from observations. Hence one should expect that the very high dimensionality of data provided by climate simulations can be reduced to study such quasi-persistent weather regimes. In this working group we discussed possible data-driven approaches to identify persistent weather regimes and derive reduced models for the dynamics, based on climate reanalysis datasets. Different approaches to reduce the dimension of input data where discussed. Once identified from data, the reduced models can be used to study the dynamics of transitions between weather regimes and identify factors favouring regime transitions. That can be pursued based on statistical causality approaches that help to investigate the impact of weather regimes on other atmospheric features such as cloud organization, or the occurrence of atmospheric rivers.

Participants

- Mikkel Abrahamsen
University of Copenhagen, DK
- Pankaj Kumar Agarwal
Duke University – Durham, US
- Kevin Buchin
TU Dortmund, DE
- Maike Buchin
Ruhr-Universität Bochum, DE
- Dwaipayan Chatterjee
Universität Köln, DE
- Susanne Crewell
Universität Köln, DE
- Anne Driemel
Universität Bonn, DE
- Peter Dueben
ECMWF – Bonn, DE
- Sándor Fekete
TU Braunschweig, DE
- Franziska Glassmeier
TU Delft, NL
- Siddharth Gupta
University of Warwick – Coventry, GB
- Jan-Henrik Haunert
Universität Bonn, DE
- Stefan Kollet
Forschungszentrum Jülich, DE
- Jürgen Kusche
Universität Bonn, DE
- Thomas Moelhave
SCALGO – Aarhus, DK
- Petra Mutzel
Universität Bonn, DE
- Peer Nowack
KIT – Karlsruher Institut für Technologie, DE
- André Nusser
University of Copenhagen, DK
- Jeff M. Phillips
University of Utah – Salt Lake City, US
- Benjamin Raichel
University of Texas at Dallas – Richardson, US
- Vera Schemann
Universität Köln, DE
- Melanie Schmidt
Heinrich-Heine-Universität Düsseldorf, DE
- Christian Sohler
Universität Köln, DE
- Frank Staals
Utrecht University, NL
- Nikki Vercauteren
University of Oslo, NO
- Bei Wang Phillips
University of Utah – Salt Lake City, US