Discovery Informatics for Climate Science

Prepared for the NSF Discovery Informatics Workshop
held February 2-3, 2012 in Washington, DC

Karsten Steinhaeuser
Computer Science & Engineering
University of Minnesota
ksteinha@umn.edu | www.umn.edu/~ksteinha

Climate change (or "*global change"*) has become one of the defining issues of our era, as greenhouse gas emissions from fossil fuel combustion and deforestation will likely lead to an increase in the average temperature of the world over the course of the 21st century (IPCC 2007). Potentially dramatic consequences include increased occurrence of extreme weather events, shocks in food and water supplies, rising sea levels, spread of diseases, and so on, which have created an urgent need to better understand the Earth's climate and its evolution to the present state as well as predict future changes in response to natural and anthropogenic causes (DOE 2008). Thus, while developments in first principles-based physical models of the Earth system and computational tools for implementing them remain a high priority, now is the time to explore alternate strategies for characterizing uncertainties and risks and reducing them in order to inform actionable mitigation and/or adaptation strategies.

Climate and earth sciences have recently experienced a rapid transformation from a historically data-poor to a data-rich environment, thus bringing them into the realm of the *Fourth Paradigm* of scientific discovery – a term coined by the late Jim Gray (Hey et al. 2009), the other three being theory, experimentation and computer simulation. In particular, climate-related observations from remote sensors on satellites and weather radars, *in situ* sensors and sensor networks, as well as outputs of climate or Earth system models from large-scale simulations, provide terabytes of spatio-temporal data. These massive and information-rich datasets offer a huge opportunity for advancing the science of climate change and its impacts, yet current analysis techniques are not able to fully realize their potential benefits. I am part of an NSF Expeditions in Computing project titled *Understanding Climate Change: A Data Driven* Approach, which aims to fill this void by applying the knowledge discovery process to climate and ecosystem data. Specifically, the goal of this project is to dramatically advance the state of the art in spatio-temporal data mining while addressing the unique challenges of climate and ecosystem data including:

- Spatio-temporal dimensions – traditional i.i.d. assumption does not hold
- Long-range spatial dependencies (or teleconnections) – violates assumption of spatial statistics/ data mining that near things are more closely related than distant
- Long-memory temporal dependencies – violates assumption of time series analysis that more recent events are more relevant
- Nonlinear and non-stationary processes
- Low-frequency variability
- Multi-scale processes and dependencies
- …

For more information about the project, visit our website at http://climatechange.cs.umn.edu

There has also been a concerted effort by a group of scientists representing both computer science and climate science (myself included) to build a community around this emerging research area of discovery informatics for climate science, or *climate informatics*, including the organization of several workshops and conference tracks including:

- IEEE ICDM Workshop on Knowledge Discovery from Climate Data
  http://www.nd.edu/~dial/climkd11/
- International Workshop on Climate Informatics
  http://www.nyas.org/climateinformatics
- NASA Conference on Intelligent Data Understanding – Earth & Environmental Systems Track
  https://c3.nasa.gov/dashlink/events/1/
- Climate Knowledge Discovery Workshops
  https://redmine.dkrz.de/collaboration/projects/ckd-workshop
- AAAI Special Track on Computational Sustainability and Artificial Intelligence
  http://www.aaai.org/Conferences/AAAI/2012/aaai12sustainabilitycall.php
- ICCS Workshop on Data Mining in Earth System Science
  http://www.climatemodeling.org/workshops/dmess2012/
- Workshop on Understanding Climate Change from Data
  http://climatechange.cs.umn.edu/annual.php

While these events are all targeted at one particular subject area, they also represent an instance of discovery informatics in action and an example of the type of multi-disciplinary collaborations necessary for success in these types of activities.

References

DOE (2008). *Identifying Outstanding Grand Challenges in Climate Change Research: Guiding DOE's Strategic Planning*. Department of Energy Biological and Environmental Advisory Committee Workshop Report.

Hey, T., S. Tansley, K. Tolle, *eds*. (2009). *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Microsoft Research.

IPCC (2007). *Climate Change 2007: The Fourth Assessment Report from the Intergovernmental Panel on Climate Change.*