Preface to the 2010 Workshop on Knowledge Discovery from Climate Data: Prediction, Extremes, and Impacts

The 2010 Workshop on Knowledge Discovery from Climate Data: Prediction, Extremes, and Impacts is the second workshop on this topic convened in conjunction with the IEEE International Conference on Data Mining. Through this workshop, we bring together leading researchers from academia as well as practitioners from industry and government organizations; areas of expertise include data mining and computational data sciences for spatial/spatio-temporal data, computational statistics, climate modeling and visualization, as well as earth and atmospheric sciences.

The analysis of climate data, both observed and model-generated, poses a number of unique challenges: (i) massive quantities of data are available for mining, much of which has never been analyzed due to lack of a suitable methodology, (ii) the data is spatially and temporally correlated so that the IID assumption underlying many learning algorithms does not apply, (iii) the data-generating processes are known to be non-linear, (iv) the data is potentially noisy, uncertain and highly imbalanced, and (v) extreme events are known to exist within the data. These characteristics present fundamental research challenges to the data mining community. We posit that the interface with climate data sciences will not only present a real-world data challenge bereft of such issues, but also provide an application that has the potential to impact physical and natural systems in the world.

Climate data mining is based on spatio-temporal data and inherits the attributes of space-time data mining. In addition, climate relationships are nonlinear, spatial correlations can be over long range (teleconnections) and have long memory in time. The processes are nonlinear, even chaotic or sensitive to initial conditions, with seasonal effects, as well as with low-frequency, even 1/f, noise. Thus, in addition to new or state of the art tools from temporal, spatial and spatio-temporal data mining, new methods from nonlinear modeling and analysis are motivated along with analysis of massive data for teleconnections and long-memory dependence.

The workshop has six accepter papers, and an invited talk by Dr. Claire Monteleoni, Associate Research Scientist at the Center for Computational Learning Systems and Assistant Professor of Computer Science at Columbia University. Dr. Monteleoni is a machine learning researcher with a background in Earth & Planetary sciences and her research is at the intersection of machine learning and climate science.

We are indebted to our program committee members who helped with the peer review process. We would also like to thank all the authors who submitted papers to the workshop; they provided us with an excellent program.

Workshop Organizers

Nitesh V Chawla
Auroop R Ganguly
Vipin Kumar
Michael Steinbach
Karsten Steinhaeuser