Predicting Extremes by Data-driven Analytics (PrExDA)
Report of Workshop (Sep 30 – Oct 2, 2020, online)

Summary

This report summarizes the results of a workshop entitled "Predicting Extremes by Data-driven Analytics (PrExDA)" and sponsored by NSF Convergence Accelerator initiative. The participants (30) from seven countries (Brazil, Canada, Chile, Germany, India, UK and USA) are engaged in space weather, terrestrial climate and weather, insurance, finance and economics, and social work. They represented academia, industry and commerce, small business, federal agencies and laboratories, and non-profit organizations. They identified challenges in modeling and prediction of extreme events in natural and anthropogenic systems, and those that can be addressed by transdisciplinary convergence research focused on data-driven approaches.

The Convergence Agenda centered on the assessment of modeling and prediction of complex systems as a data-driven analytics platform, and its suitability and readiness for transition to use-inspired products. The presentations covered multidisciplinary modeling and prediction of complex systems, machine learning, complex networks, and information theory. The focus was on the complex systems framework as a self-consistent approach for a convergence of research, use-inspired product and partnership for societal impact. The advances enabled by this approach in many areas, including space weather and terrestrial climate and weather, provide the foundational research for the development of a platform for harnessing massive data of multiple phenomena.

The Users of the products will be the communities engaged in meeting the challenges of natural hazards and other extreme events. Hazards such as wildfires, weather extremes and space storms, are recurring challenges for government agencies, policy and decision makers, and communities engaged in response and resilience efforts. The prediction of extreme events in natural, commercial and social sectors are directly relevant to commerce and industry, including small business.

The Deliverables are centered on a comprehensive platform for prediction of extremes by data-driven analytics (PrExDA) that will transition the foundational research in data-driven modeling and prediction. Such a platform can be a licensed or open-access product on the internet/cloud or in a stand-alone mode. The planning phase (9 months) will focus on the proof-of-concept in stand-alone mode and will use data of selected phenomena (wildfires, heat waves and space storms). The implementation phase (2 years) for the comprehensive system will develop the infrastructure for data preparation, modeling and prediction, and user-inspired products.

The Partnerships needed for the transition from research to use-inspired product will be achieved by engaging the research (academia), development of products (industry, small-business) and users (policy and planning, hazard response, commercial, non-profit). The workshop highlighted the essential role of partnerships, identified potential partners (regional and urban governments, hazard response entities, small business) and pathways to building these partnerships, such as with UR (understandingrisk.org) and Thriving Earth Exchange (thrivingearthexchange.org).

The Cohort of teams for harnessing massive data for developing products of societal impact have a strong common interest in a thriving ecosystem. First, the data availability and readiness for seamless ingestion into different products is a common requirement that can leverage collaborative efforts. Second, the data-driven analytics based on different frameworks, such as machine learning and statistical analysis, has deep synergy with PrExDA. Third, the stakeholders in transitioning research to use-inspired products by harnessing data are largely common to the cohort, thus providing a strong motivation for working across the teams.

Workshop Structure

The workshop theme of prediction of extremes by data-driven analytics requires evaluation of research outcomes for transition to products, development of required tools for ready access and use, and building partnerships for an ecosystem engaged in solutions with high societal impact.

The workshop agenda (https://prexda.astro.umd.edu) was planned with the objectives of:
1) bringing together participants from different disciplines and communities to form teams with well-defined common goals and objectives,
2) identifying data availability and data needs for different systems,
3) defining proof-of-concept of integrative data-driven modeling of extremes in multiple systems,
4) evaluating requirements for a proto-type prediction system for forecasting and risk assessment.

The workshop activities centered on the modeling and prediction in the complex systems framework by harnessing the data revolution and its readiness to address problems of high societal impact. This approach is based on the recognition that the dynamics of a nonlinear dissipative system can be reconstructed from time series data of a limited number of variables, independent of modeling assumptions. The combination of this capability with the growing big data enables the modeling and prediction of extremes of multiple phenomena in the complex systems framework [Fig. 1].

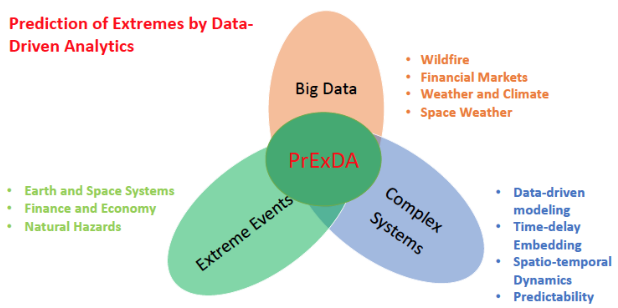

Fig. 1. Complex systems framework for Prediction of Extremes by Data-driven Analytics

The presentations in the workshop on the data-driven modeling and prediction provided critical overviews of the current methods and techniques, in particular the reconstruction of spatio-temporal dynamics from data. This complex systems approach led to the first predictions of space weather, in terms of the geomagnetic indices, and has developed techniques for quantifying the associated predictability with conditional probabilities and fluctuation analysis. A combination of these techniques with other machine learning approaches such as neural nets was addressed to explore a comprehensive modeling and prediction system. In the modeling of intraseasonal climate (Indian monsoon) that used more than 100 years of gridded rainfall data, this approach yielded better predictions than numerical models and provided a clear example of spatio-temporal modeling and prediction, and quantification of predictability. The challenges of modeling extreme events in complex nonequilibrium systems were discussed in general and in the context of terrestrial climate and space weather. The current research is largely based on the distribution of events to quantify the likelihoods of extremes from their heavy-tail features. In another technique, the waiting times yield the likelihoods of events of specified strengths in different time windows. The complementary aspects of these techniques and advantages and ways to an integrated methodology were discussed to identify and assess their strengths and weaknesses. The modeling and prediction using statistical methods such as NARMAX for space storms and cryosphere systems (Greenland ice sheet) presented at the workshop provided a way to compare them with the complex systems approach and for exploring ways to enhance the capabilities.

Evaluative Activities

The transdisciplinary nature of the workshop theme brought together participants from diverse areas, including space weather, earth system modeling, weather and climate, insurance, finance and economics, complex networks, machine learning, and satellite imagery and monitoring. This enabled a critical evaluation of the maturity of the research and potential for transition to use-inspired product. Keeping in perspective the developments in data-driven modeling and prediction in many areas, those in space weather have the common thread of complex systems approach using ground-based and satellite-borne observations. These advances, including the *first predictions of space weather*, provide a comprehensive set of research and discovery that form a coherent framework for convergence acceleration. Integrated with research in intraseasonal climate, this approach is suitable for transition to use-inspired product with potential for societal impact.

The modeling and prediction of space weather formed the major part of the presentations and addressed the key aspects, viz., data availability, modeling and prediction, and quantification of predictability. The data from ground-based and spacecraft-borne instruments provide a comprehensive database, covering many decades, for modeling and prediction of space storms in terms of the widely used indices. A key challenge of data-driven space weather modeling is the spatial coverage because of the very large volume of the physical system. While extensive data is available from networks of ground magnetometer stations and ionospheric monitoring stations, more data will be needed to provide improved regional and local forecasts.

The spatio-temporal modeling and prediction of intraseasonal climate provide an example of data-driven modeling and prediction, in particular the use of gridded data. The modeling using multi variable singular spectrum analysis (MSSA) is suitable for development of predictions using spatially distributed data, including satellite images. Another important feature of this modeling is the comparison with first-principle based numerical models, providing the complementary aspects.

The multi-spectral imaging data from satellites, e. g., NOAA Suomi NPP, provide extensive data of spatio-temporal features, including wildfires. The presentation on NOAA imaging data of wildfires in California from Suomi NPP showed daily development and such data are available continuously from October 2011. The data of wildfires at higher time resolution are provided by NOAA geostationary GOES-R series satellites (GOES-17/West,and GOES-16/East). These imaging data are well suited for the modeling and prediction, following the approach using gridded data for intraseasonal climate.

The conceptual framework for prediction of extreme events in natural and anthropogenic systems as an integrated platform is shown in Fig 2. Extreme events and hazards are outliers in the distribution of events in different phenomena, e. g., wildfires, heat waves, urban heat, floods, space storms, power grid and telecommunication disruptions. From the systems point of view, these are complex systems which exhibit coherent as well fluctuating features and their historical data can be used for reliable modeling and prediction. The space weather prediction discussed earlier provide such a case. The satellite imaging data of wildfires, complemented by ground-based observations, is another case that is considered suitable for the prediction system. In the proof-of-concept phase these two cases, viz., wildfires and space weather, can provide sufficient range and depth for developing a comprehensive product that can be used in a range of extreme events. The implementation phase will be the integration of horizontal and vertical components of PrExDA, as shown schematically in Fig. 2. Although the verticals in this framework are shown as independent systems there are couplings among them and these need to be included. For example, wildfires, heat waves and urban heat can influence each other and an integrated modeling is essential for improved and reliable predictions. In a similar manner space storms, power grids and telecommunications are coupled, thus requiring an integrated approach. With a range of spatial and temporal scales of the physical processes in these phenomena, an integrated modeling based on the first-principles is a daunting task. Even so modeling in the complex systems framework can provide a well-defined platform for integrated prediction system. While the underlying processes of the horizontals in Fig. 2 can be independent, the extreme events can have confluence. For example, space storms during any other extreme event will exacerbate the damaging effects. An integrated system is thus expected to provide an ecosystem for bringing together stakeholders in natural hazards, social, commercial, bio-medical and public health systems.

The partnership for building a thriving ecosystem will bring together academia, industry, small-business and non-profit entities. Examples including AGU Thriving Earth Exchange (thrivingearthexchange.org) and MIT Climate CoLab (climatecolab.org) provide effective ways of engagement. As an open global community of over 9000 experts and practitioners in different aspects of disaster risk, Understanding Risk (understandingrisk.org) facilitates active collaboration. The biennial UR Forums provide a dynamic ecosystem for collaboration and PrExDA will explore partnerships through such events.

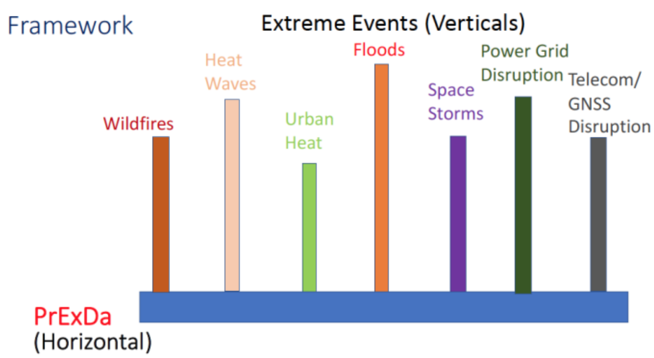

Fig. 2. Concept for Prediction of Extremes by Data-driven Analytics Framework.

Conclusions and recommendations

The adoption of the complex systems framework for modeling and prediction of extremes is well founded and timely based on the following recent developments.

- Advances in the theory of nonlinear dissipative complex systems
- Reconstruction of dynamics from time series data (based on the embedding theorem)
- Massive data of natural and anthropogenic systems

This approach enables modeling and prediction of a variety of systems from data, encumbered by modeling assumptions in the respective first-principles models. Reconstruction of phase space from time series data with its mathematical foundation of the embedding theorem, derives this capability from the nonlinear and dissipative nature of complex systems, enabling low-dimensional representations based on the leading features inherent in the data. Modeling of the dynamics naturally yields predictions of the future states of the system from the characteristics of similar states in the past. Predictions are based on the leading features inherent in the data and is an averaged behavior, and the extremes form the heavy-tail of the deviations. The trans-disciplinary nature of the complex systems framework makes it eminently suitable for modeling and prediction of extreme events in many systems by harnessing the massive data from the growing observing platforms. Such a prediction system will address a range of needs of disaster risk assessment and response in natural and anthropogenic extreme events, and is expected to have a broad spectrum of users, from government, industry and non-profit entities. This will provide a forum to build an ecosystem through partnerships to deliver solutions with high societal impact.