Chapter 13
Knowledge Discovery from Sensor Data For Scientific Applications

Auroop R. Ganguly, Olufemi A. Omitaomu, Yi Fang, Shiraj Khan, and Budhendra L. Bhaduri

Abstract The current advances in sensors and sensor infrastructures offer new opportunities for monitoring the operations and conditions of man-made and natural environments. The ability to generate insights or new knowledge from sensor data is critical for many high-priority scientific applications especially weather, climate, and associated natural hazards. One example is sensor-based early warning systems for geophysical extremes such as tsunamis or extreme rainfall, which can help preempt disaster damage. Indeed, the loss of life during the 2004 Indian Ocean tsunami may have been significantly reduced, if not totally prevented, had sensor-based early warning systems been in place. One other example is high-resolution risk-mapping of insights obtained through a combination of historical and real-time sensor data, with physics-based computer simulations. Weather, climate and associated natural hazards have established history of using sensor data, such as data from DOPPLER radars. Recent advances in sensor technology and computational strengths have created a need for new approaches to analyzing data associated with weather, climate, and associated natural hazards. Knowledge discovery offers tools for extracting new, useful and hidden insights from data repositories. However, knowledge discovery techniques need to be geared towards scalable and efficient implementations of predictive insights, online or fast real-time analysis of incremental information, and solution processes for strategic and tactical decisions. Predictive insights regarding weather, climate and associated natural hazards may require models of rare, anomalous and extreme events, nonlinear phenomena, and change analysis, in particular from massive volumes of dynamic data streams. On the other hand, historical data may also be noisy and incomplete, thus robust tools

A.R. Ganguly · O.A. Omitaomu · Y. Fang · S. Khan · B.L. Bhaduri
Computational Sciences and Engineering Division, Oak Ridge National Laboratory, 1 Bethel Valley Road, Oak Ridge, TN 37831 USA

A.R. Ganguly
e-mail: gangulyar@ornl.gov

need to be developed for these situations. This chapter describes some of the research challenges of knowledge discovery from sensor data for weather, climate and associated natural hazard applications and summarizes our approach towards addressing these challenges.

Key words: Sensors, Knowledge discovery, Scientific applications, Weather extremes, Natural hazards

13.1 Introduction

Predictive insights generated from sensor data, in conjunction with data obtained from other sources like computer-based simulations, can facilitate short-term decisions and longer-term policies. Remote sensors [35], such as Earth-observing satellites, weather radars, large-scale sensor infrastructures [38] and environmental wireless sensor networks [5], yield massive volumes of dynamic and geographically distributed sensor data at multiple space-time resolutions. We define sensors broadly to include wireless sensor networks, in-situ sensor infrastructures and remote sensors. The raw data need to be converted to summary information and subsequently used to generate new knowledge or insights, ultimately leading to faster and more accurate tactical and strategic decisions. Therefore, we define knowledge discovery as the overall process where raw data from sensors or simulations are ultimately converted to actionable predictive insights for decision and policy makers. In addition to observations, scientific applications demand that information about the known physics, or data-dictated process dynamics, be taken into account. The scientific domains are diverse and requirements for sensor-based data processing and analysis can be fairly broad on one hand and domain specific on the other. This chapter focuses on applications of knowledge discovery from sensor data for weather, climate and geophysical hazards; these applications may be useful for hazards mitigation [17]. However, we present a broader view of knowledge discovery as compared to the traditional definitions by the data mining community, but we include the data mining and other data sciences as key aspects of the overall process.

Hazards can be natural [50], such as weather extremes including rainfall, hurricanes and heat waves; they can be technological, such as leakage and spread of toxic plumes from industrial facilities [23]; or they can be adversarial, as in security [1] and war. This chapter focuses primarily on hazards due to weather or climate extremes [33]. The idea of using intelligent data sciences and sensor data for hazards mitigation has been demonstrated in a proof-of-concept way. For example, in October 2006, a small satellite, Earth Observing 1 (EO-1) [41], collected data on its own after noticing a plume of smoke on the island of Sumatra, Indonesia [43]. Such automatic sensor-based data collection efforts could provide insights into what happened hours before a natural hazard; in this case, before the eruption of a volcano. The overall goal is to develop objective-based rather than subjective-based models, high-resolution rather than low-resolution models, and large-scale rather than low-scale models that can form bases for extracting useful and insightful knowledge for immediate and future hazard-mitigation purposes.

This chapter is organized as follows. Section 13.2 proposes a broader knowledge discovery framework. Section 13.3 presents a brief introduction to natural hazards and sensors used for natural hazards. Section 13.4 discusses the significance and challenges of knowledge discovery from sensor data for natural hazards. Section 13.5 focuses on some applications of knowledge discovery approaches in natural hazards. Section 13.6 presents some preliminary discussions of the applications of knowledge discovery insights for hazard mitigation. Section 13.7 summarizes the chapter.

13.2 A Broader Knowledge Discovery Framework

Knowledge discovery offers tools for extracting new, useful and hidden insights from massive sensor and historic data. However, knowledge discovery techniques need to be geared towards scalable and efficient implementations of offline predictive insights, fast real-time analysis of incremental information, and solution processes for tactical and strategic decisions. Therefore, we propose a somewhat broader knowledge discovery framework (see Fig. 13.1), which describes an end-to-end process for knowledge discovery for natural disasters.

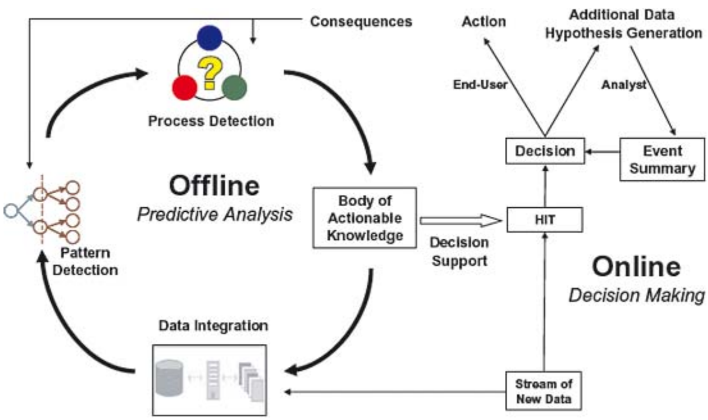


Fig. 13.1 A holistic approach to knowledge discovery

The components of the proposed framework, in the context of natural hazards, are stated in this section; a description of the two broad areas of the framework are discussed briefly in the following subsections. The proposed framework consists of two sub-frameworks:

1. Offline Predictive Analysis
   - Data Integration
     – Remote sensors, wired and wireless in-situ sensor networks

  - Numerical physics-based computer model outputs
  - Ancillary information and encoded domain knowledge

- Pattern Detection
  - Offline data mining from sensor observations and models
  - Computational efficiency and scalability to massive data
  - Anomalies, extremes, nonlinear processes, in space and time
  - Probabilities, intensities, duration, frequency, risks of observations

- Process Detection
  - Numerical models with sensor data assimilation schemes
  - Extraction of dynamics from massive sensor observations
  - Extraction of dynamics from incomplete, noisy information

2. Online Decision Making

- Decision Support
  - Online (real-time) analysis from models and observations
  - Algorithmic efficiency for dynamic, distributed processing of sensor observations
  - Resiliency, vulnerability, and impacts of observations
  - Visualization and decision or policy aids models

13.2.1 Requirements for Offline Predictive Analysis

We propose the broad requirements of the offline predictive analysis, other than the need for both capacity and capability computing, as including the following:

1. Multidisciplinary: Multiple aspects of problems solved using a set of individual tools, each motivated from one or more disciplinary area.
2. Interdisciplinary: Comprehensive solutions developed based on blend of methodologies spanning traditional disciplinary areas.
3. Process-based: Larger overall problem partitioned into component processes and solved using physics and a suite of data science tools.
4. Holistic: Approaches for an application from raw sensor and model data to decision and policy aids.

The distinguishing features compared to conventional knowledge discovery areas are the following (sub-bullets list the primary differences from the conventional):

- Data Mining
  - Enhanced focus on scientific rather than business data
  - Algorithms for anomalies, extremes, rare and unusual events rather than predicting regular events
  - Geographic, time series, spatial, space-time *relational* specific data

- Statistics and Econometrics
  - Focus on computational efficiency and scalability for distributed sensors
  - Methods for nonlinear processes and representations
  - Statistics of rare events and extremes/anomalies
- Nonlinear Dynamics and Information Theory
  - Robust to limited or incomplete and noisy information
  - Scalability to massive data for centralized sensors
  - Spatial, space-time and geographic specific data
- Signal Processing
  - Nonlinear dynamical, even chaotic, system behavior
  - Colored, even 1/f, noise
  - Noisy and incomplete information

13.2.2 Requirements for Online Decision Making

The *decision support* component is composed of online (real-time) knowledge discovery and the decision sciences. The online knowledge discovery processes need to be efficient in terms of memory usage and analysis times (especially for distributed sensors); must be able to handle incremental information in real-time; and must generate time-phased or event-based decision metrics, at multiple geographically based locations and possibly times, such that metrics can be used for automated alert mechanisms or to facilitate the task of the *human in the loop* in the space-time context.

One new example of this application is the concept of ubiquitous sensing. Ubiquitous sensing describes a situation where one has either an array of many sensors that generate high flows of data, much of which may be null (background, uninteresting or contradictory), or where one has a few mobile sensing platforms that need to be deployed in a cost-effective way. Examples of the first category include arrays that detect contraband crossing borders or unauthorized persons entering restricted areas. Examples of the latter category include satellite or air-breathing remote-sensing assets. Some offline modeling and decision-support tools have been semi-coupled in real-time to direct the next sequence of data acquisition. In one example of these decision-support tools, Bayesian approaches formulate hypotheses (such as a missile launch being detected) and marshal the data from other elements of the array (in the first example) or to move the mobile platform to the next location (in the second example) in order to gain the next most valuable data point that would reduce uncertainty once an event is detected. Some of the applications and challenges of ubiquitous sensing has been discussed in the literature [53].

Efficient real-time algorithms are required to react to the real-time, dynamic and distributed nature of knowledge discovery as well as direct this data acquisition within the time cycle of an event. Overall, the online approaches need to be algorithmically efficient, that is, the mathematical algorithms must be amenable to robust online implementations, which implies that they be fast, storage-efficient, memory efficient, adaptive and possess real-time or near-real-time analytic capacity.

However, there is a trade-off between computational efficiency, algorithm performance and domain requirements. A good example of such a trade-off is the SPIRIT algorithm [48], which is essentially an incremental version of the Principal Component Analysis (PCA) technique, but the weight estimates are slightly different from the principal directions in conventional (offline) PCA. The difference in computational requirements does not affect the algorithm performance. Therefore, a good understanding of application domains is the key to achieving such compromise. The decision-science component encompasses the development of decision metrics in space and time for visualization and visual analytics, utilization of predictive insights from offline analysis and real-time distributed discovery processes. Additionally, the decision-science component processes dynamic and event-based streams of data in conjunction with offline discovery and real-time analysis, provides feedback loops from prior decisions or policies, and provides a framework for decision metrics, uncertainty and impacts of risks, including the determination of resiliency and consequences in the context of natural disasters.

13.3 Weather, Climate, and Associated Natural Hazards

The exposure of human life and economy to natural hazards—from hurricanes, volcano, tornadoes, tsunamis and earthquakes to heat waves, cold spells, droughts and floods or flash floods—appears to have increased even as world economies have developed and prospered [50,65]. However, one natural hazard impacts the other, which generates multidimensional scenarios. In this section, we discuss impacts of some natural hazards and highlight some sensors that can collect relevant data, giving us a better understanding of the causes of, and interactions among, natural hazards.

13.3.1 Natural Hazards Impacts and Weather Extremes

Even though the occurrence of natural hazards cannot be prevented, understanding the interactions between natural hazards is significant for extracting new insights. In this section, we discuss climate impacts on weather extremes and the human impacts of weather extremes.

1. *Climate Impacts on Weather Extremes*: The tremendous uncertainty surrounding some of the issues regarding climate-weather linkage—for example, the links of global warming to the increase in the number and intensity of hurricanes—suggests that a closer inspection is necessary. Specifically, historical weather-

sensor observations and climate data gathered from various sources need to be analyzed, in an offline mode, in significant detail and with much greater care. When climate models and indicators are used to understand and quantify the impacts of climate on weather extremes, there is a need to delineate the impacts of natural climate variability before or during the quantification of the impacts of climate change. These issues are described in detail below:

- *Natural Climate Variability and Weather Extremes*: In the longer-term, natural climate variability—for example, the inter-annual El Niño phenomena—can have significant impact on weather and hydrologic extremes [19,27,57]; In fact, the 2006 hurricane season turned out to be much quieter than anticipated and the most plausible hypothesis is the occurrence of the El Niño [8], although the influence of African dust storms has also been suggested as an added factor [40]. Incidentally, as of this writing, a *very active* 2007 hurricane season is being predicted by forecasters [60].
  The need to understand and quantify the impacts of natural climate variability on weather extremes is underscored through the previous examples. The ability to quantify climate variability, including climate anomalies like El Niño, requires processing massive amounts of geographic data obtained from remote sensors like satellites and aircraft. Sensors for ocean temperature or salinity and sensor networks like ocean monitoring instrumentation also play a role. The ability to relate such large-scale geophysical phenomena to weather or hydrologic extremes and regional change, both for offline discovery and online analysis, requires holistic knowledge discovery approaches and massive computational capabilities.
  The need to quantify the impact of natural climate variability also stems from the requirement to delineate and isolate the effects of global or regional climate change [24], and in particular possible anthropogenic effects [20], on weather extremes and natural hazards.
  The impacts and current wisdom in the insurance sector (e.g., see [70] and [59], for two interesting viewpoints) may provide some indications to how the financial world may be adapting—or may be anticipating the need to adapt—to climate change. The importance of human factors rather than climate change has been emphasized as the primary driving cause for recent natural disasters [10]. Although consensus is lacking on the relative impacts of change in climate versus human factors (e.g., [59,70]), the problem points to two different lines of research. First, there is a need to understand the relative and complementary roles of climate- and human-induced changes and their combined impact on natural disaster losses. Second, there is a need to understand how future changes in climate may influence the variability of the weather extremes as well as related disaster losses. Finally, there is a need to combine the risks, consequences, vulnerabilities and anticipatory damage assessments on impacts within one policy tool which can provide metrics and visual guidance to policy makers.

- *Climate Change and Weather Extremes*: Future projections—especially at sufficiently long terms when projections based on past trends or current observations may no longer be valid—need to rely directly or indirectly on climate model simulations. State-of the-art climate models like the Community Climate System Model, Version 3, (CCSM3) can generate precise climate reconstructions and predictions. For example, CCSM3 can give estimates of climate variables at three dimensions and one-degree spatial grids, from the year 1870 to 2100 [6]. These estimates can be given at daily or even six-hour intervals However, precision does not necessarily imply accuracy, and precise predictions may be only as good as the temporal and spatial scales of the coupled atmospheric and earth system processes that such systems can model.
  The first step is to compare the model outputs with observations for time periods when both are available. This comparison is needed to understand the problems in the model outputs and to quantify the inherent uncertainties in space and time. Since any simulation model is an imperfect realization of reality and tends to smooth out the outliers and extremes, this can be a hard test for climate models. However, simplified tests may help prove a point. Thus, while the parameters of extreme value theory obtained from observations and model simulations of temperature may or may not be statistically similar, the number, frequency and duration of heat waves based on user-defined criteria and thresholds may align well and this alignment may provide sufficient information in some cases. There have been attempts to compare modeled and observed extremes [31], even based on detailed statistical analysis of extreme values [32].
  The next step is to investigate trends and patterns within climate model projection and quantify the uncertainties based on the results of model-observation comparisons. The *Science* paper by Meehl and Tebaldi [40] demonstrated how insights about future weather extremes—in their case heat waves—can be obtained in this fashion. However, this is a good starting point in terms of actionable predictive insights from a combination of observations and models. If temporally, spatially and geographically aware knowledge discovery tools [28,30,32], specifically tailored for earth science applications, are *let loose* on the massive volumes of sensor-observed and model-simulated data, we can hope to validate, and perhaps discover, insights about weather extremes. This is an urgent and high-priority research area whose time has clearly come.

2. *Human Impacts and Weather Extremes*:

   - *Globalization and Change in Human Factors*: While our discussions have focused on weather extremes alone, there have been claims that the current increase in disaster losses is more due to human factors, such as the impact of human actions on the global environment. However, there is also an understanding that anticipated climate change may begin to change the relative impacts. In any case, a link needs to be firmly established (or rejected) between the anticipated change in weather extremes, whether caused by inherent climate variability or human-induced change, to the corresponding impacts on human population [14,3] and adaptability.

- *Resiliency, Vulnerability and Policy Tools*: Weather disaster impacts relate to the design and safety of infrastructures and the resiliency of vulnerable societies [50,64,67]. An integrated policy tool is needed for various levels of strategic decisions. Thus, insights from hurricane or rainfall extremes based on archive sensor data may be used to design more resilient hydraulic structures in the short-term near the coasts, or stronger foundations for offshore structures. In addition, these policy aids can be used to assess the extreme variability, risks or consequences, resiliency and overall damage caused by anticipated extremes (for a proof-of-concept example, see [14,47,61]).
- *Policy Impacts*: A quantitative assessment of climate-weather links has direct influence on the design of highway and infrastructure sensing or monitoring systems, building redundancies for contingency planning, enhancing readiness of societies through early warning systems and public education, and planning human habitations such that vulnerabilities may be reduced. Longer-term planning based on climate projections influences human habitation and demographics through policy regulations; for example, planned movement of populations from vulnerable regions. This may be a feedback loop as the quantitative assessment may help guide climate policy.

13.3.2 Utilization of Sensors for Weather and Climate

The occurrence of natural hazards cannot be prevented; but their occurrences and interactions can be studied for useful insights into what drives them. Sensors have long been used to collect data about weather (e.g., DOPPLER), climate, and natural hazards. Recent advances in computational techniques and the current advances in satellite, telecommunication and sensor technologies are providing access to, and analysis of, massive data that can provide better knowledge of what drives these hazards. The types of natural hazards are many but the most common (in alphabetical order) include asteroid, avalanche, drought, earthquake, flood, heat wave, hurricane, landslide, salinity, tornado, tsunami, volcanism, and wildfire. Each of these natural hazards has been an integral part of the human experience. However, their occurrences, as well as their effects on lives, properties and infrastructures, are becoming more dramatic. For example, drought may not be the most dramatic occurrence, but it is one of the most damaging disasters. Since 1967, drought alone has been responsible for millions of deaths and has cost hundreds of billions of dollars in damage worldwide [41]. Some of the sensors that can be used to track these hazards and collect related data are:

1. DOPPLER Radars: These are weather-related sensors that send out radio waves and are one of the oldest weather-related sensors.
2. Earth Observing Sensors: These are satellite-based sensors used by NASA to monitor the Earth. An example of these sensors is Earth Observing 1 (EO-1).

These sensors are useful for weather/climate-related hazards such as heat waves, volcano, and hurricanes [41].
3. The National Ecological Observatory Network (NEON) sensors: This array of sensors are used to understand how land-use change and climate variation affect ecological systems [22].
4. Remote Sensors: Remote sensors are used to measure global ice cover changes and carbon deposits, which can help track hurricanes, forest fires, and many other climate-related hazards. Two types of remote sensing are used by NASA: passive remote sensing, such as radiometers, and active remote sensing, such as RADAR and Lidar [42].
5. River Sensor Network: This network of sensors developed at the University of Lancaster monitors water depth and flow, which can be used to predict impending floods. Some of the sensors in the network measure pressure from below the water line in order to determine depth; others monitor the speed of river flow to track objects and ripples moving along the surface from the riverbank [44].
6. Satellite Imagery: Satellite imagery is a multi-sensor system useful for capturing forces of nature such as hurricanes [11].

Other examples are in-situ sensors, such as the Prompt Assessment of Global Earthquakes for Response (PAGER) system, developed by the US Geological Survey, which automatically estimates human impact following significant earthquakes [66]. This system also provides important information to help emergency relief organizations, government agencies and the media plan their response to earthquake disasters. There are also efforts to distribute data from environmental and ecological sensors to interested communities for further analysis. For example, the sea-viewing Wide Field-of-view Sensor (SeaWiFS) project provides quantitative data on global ocean bio-optical properties to the earth science community [62]. These data are useful for extracting insightful knowledge that can help us understand the causes driving these hazards. These sensors and others provide data that can be used for insightful predictions of when an hazard will happen and the potentially affected areas.

13.4 Challenges of Knowledge Discovery from Sensor Data for Natural Hazards

The earth science community has developed and used traditional statistics [71], non-traditional statistical models like extreme value theory [25,26], spatial and space-time statistics [7,45], and nonlinear dynamics and signal processing [46,54,71], for conventional data-analysis applications. However, the availability of advanced sensor technology with high-performance computational facilities provides opportunities for developing solutions beyond statistical analysis of isolated data sets. Some applications of knowledge discovery for the earth sciences have been reported in the last few years [21,52,63]. However, in view of the amount of sensor data available, contributions from data mining have been limited. For example, the process of using remotely sensed and normalized difference vegetation index (NDVI) data for land cover change detection is well known. Potere et al. [51] demonstrated a novel knowledge discovery approach where time series data for NDVI (derived from the MODIS sensor [37] between 2000 and 2005) was used to detect changing landscape and land use from construction of WalMart stores. However, such an approach, originating from traditional earth science perspective, was rather visually driven and not easily transferable from a knowledge discovery *approach* to a knowledge discovery *process* through online implementation of statistical reasoning. As a result, an online knowledge discovery approach was developed by Fang et al. [12] as a first step towards automating landscape and land-use change detection process. The online approach was motivated by statistical process control methods for change detection. The automated approach, which is an adaptation of simulated annealing for change-point detection, was validated with WalMart store openings data (Fig. 13.2) and has encouraging results.

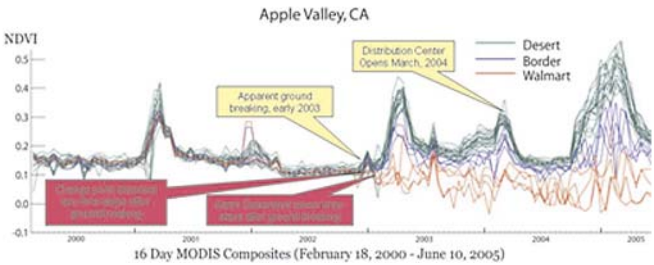


Fig. 13.2 Online change detection and alarms [12]

The online approach is a real-time approach in the sense that incremental data can be analyzed efficiently as soon as they become available. An extension of this approach can be used in the context of real-time natural disaster management by identifying regions in space and time with significant and rapid change in land cover. The analysis methods can be used to investigate geographic and remote sensing data and then zero in on areas where change is occurring or has occurred in the recent past. In addition, the methods can be used to identify changes due to deforestation. The efficiency of the approach and its incremental approach differentiates it from many traditional approaches used in earth science and remote sensing change detection applications. The interdisciplinary online approach [12] is efficient and lends itself to full automation unlike visualization-based manual validation [51] traditionally used in remote sensing. In addition, the ability to make immediate decisions based on incremental information is a distinguishing feature. The approach can be further developed to automatically detect land cover changes from large-scale and high-resolution geospatial-temporal data, as well as to automatically zoom in on the specific locations where such change may have occurred. This capability is important to assess natural disaster damage by investigating remotely sensed images before and after an event.

Newer developments in data mining include the ability to deal with dependence of learning samples or "relational" data [9,39,55], mining rare events from dynamic sequences [68,69], as well as anomaly detection (primarily in the context of cyber-security, e.g., [34]) and change-point detection [18].

Computationally efficient methods for anomalies, extreme values, rare events, change and nonlinear processes need to be developed for massive geographically based sensor data. An interdisciplinary focus within the data sciences is necessary, so that new tools can be motivated from a combination of traditional and non-traditional statistics, nonlinear dynamics, information theory, signal processing, econometrics and decision theory, and lead toward application solutions that span these areas. One key challenge of natural hazards is the multi-scale nature of the problem, which may mean that the physical processes, parameterization and parameter values, as well as the fitted data-generation models or distributions, do not remain invariant across space-time scales. In a section describing the *interdisciplinary nature of knowledge discovery in databases (KDD)*, Fayyad et al. [13] mentioned that KDD evolves from the *intersection of research fields* like *machine learning, pattern recognition, databases, statistics, AI, knowledge acquisition for expert systems, data visualization, and high performance computing*.

The time may have come to further broaden the interdisciplinary roots. Scientific applications are built on domain knowledge and are typically embedded with numerical models. The sensor technology and knowledge discovery communities also have to work together with modelers to develop efficient strategies for real-time data assimilation within physically based computer models. In addition, analysis from observations needs to be effectively synthesized with model simulations. While the challenges are tremendous, the scientific opportunities [4] and benefits can be immense.

Some of the reasons there have been limited research efforts in weather and climate extremes using knowledge discovery approaches are:

1. *Limited multidisciplinary span*: There is a need to extend knowledge discovery approaches to include tools from other disciplinary areas such as spatial statistics, extreme value theory and nonlinear dynamics, information theory and decision sciences.
2. *Limited Interdisciplinary span*: Solutions should be based on a blend of methodologies including traditional areas of statistics and machine learning.
3. *Lack of holistic-based solutions*: The knowledge discovery approaches should be formulated with solutions in mind rather isolated predictive analyses.
4. *Approaches for strategic and tactical decisions*: The knowledge discovery approaches must be focused on both short-term decisions and long-term planning rather than immediate implications.

As a result, we propose a broader and holistic knowledge discovery framework in Sect. 13.2.

13.5 Knowledge Discovery Approaches for Weather, Climate and Associated Natural Hazards

This section provides an overview of the knowledge discovery approaches developed or being developed by the authors and their collaborators in the area of weather and climate extremes and related natural hazards. While the sample approaches are neither exhaustive nor intended to be directional, they illustrate the areas where knowledge discovery, broadly construed, can help in the context of natural hazards.

The examples presented in this section attempt to emphasize the potential of knowledge discovery approaches. In addition, these are intended to provide examples of both a closed-loop knowledge discovery process depicted in Fig. 13.1 as well as for knowledge discovery approaches that may be useful for decision and policy making in natural hazards mitigation.

13.5.1 Knowledge Discovery Approaches for Weather, Climate and Associated Natural Hazards

This section provides an overview of the knowledge discovery approaches developed or being developed by the authors and their collaborators in the area of weather and climate extremes and related natural hazards. While the sample approaches are neither exhaustive nor intended to be directional, they can help illustrate the possibilities that knowledge discovery, broadly construed, can help in the area of natural hazards.

The examples presented in this section attempt to emphasize the potential of knowledge discovery approaches. In addition, these are intended to provide examples of both a closed-loop knowledge discovery process depicted in Figure 1.1 as well as for knowledge discovery approaches that may be useful for decision and policy making in natural hazards mitigation.

Some of the building blocks which can ultimately lead to a closed-loop process for tactical and strategic decision-making in the context of weather or climate related hazards are:

1. Short-term prediction from remotely sensed observations [16].
2. Trends in weather extremes from dynamic data streams [30].
3. Prediction from short and noisy sensor data [29].
4. Natural variability and impacts on local geophysical phenomena [27].
5. Comparison of model simulations and historical observations [32].
6. Real-time change detection from remotely sensed data [12].
7. Quantification and visualization of human impacts [14].

Here, we highlight a few of these building blocks through the following results:

1. *Short-term prediction from remotely sensed observations*: Figure 13.3 depicts the short-term rainfall prediction methodology developed by Ganguly and Bras [16].

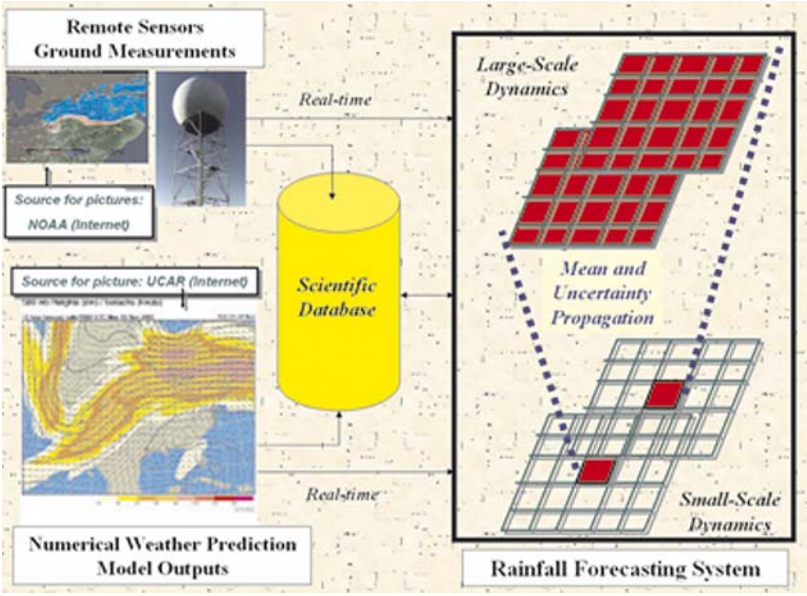


Fig. 13.3 Short-term rainfall prediction [16]

Numerical weather prediction model outputs and remote sensor observations from weather radar were blended for high-resolution forecasts at radar resolutions and for one- to six-hour lead times.

The approach relied on a process-based strategy, where the overall problem was partitioned into component processes based on domain knowledge and exploratory data analysis and then the results were re-combined. The forecasting strategy used a combination of weather physics like advection, as well as a suite of traditional and new adaptations of data-dictated tools like exponential smoothing, space-time disaggregation and Bayesian neural networks. Case studies with real data indicated [15,16] that the methodology was able to outperform the state-of-the-art approach at that time. The ability to generate short-term (0–6 hour) and high-resolution (order of a km or less in space and hourly or less in time) quantitative precipitation forecasts, especially for convective storms, is important for heavy rainfall events and hurricane activity, primarily to quantify the potential risks and damage from flash flood and flood-related hazards. Advance information can be used to control hydraulic flows, take preparatory measures and issue flood advisories.

2. *Trends in weather extremes from dynamic data streams*: Figure 13.4 exhibits a preliminary result from the approach developed by [30] for computing the spatio-temporal trends in the volatility of precipitation extremes. The methods were used for large-scale, geographically dimensioned data at high spatial resolutions. The geospatial-temporal indices were computed at each grid point in South America for which rainfall time series was available. The change in the indices can be

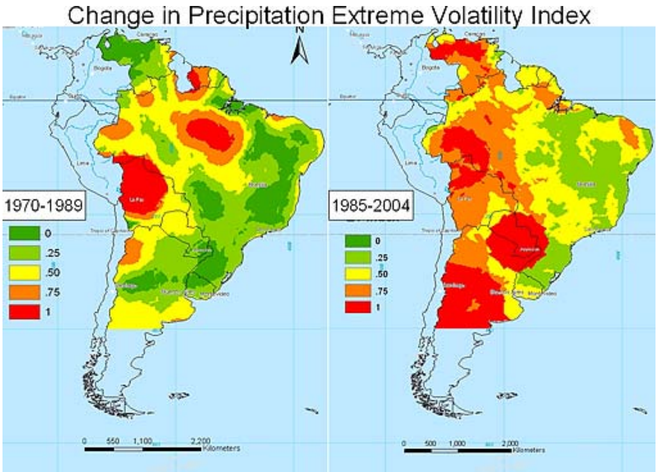


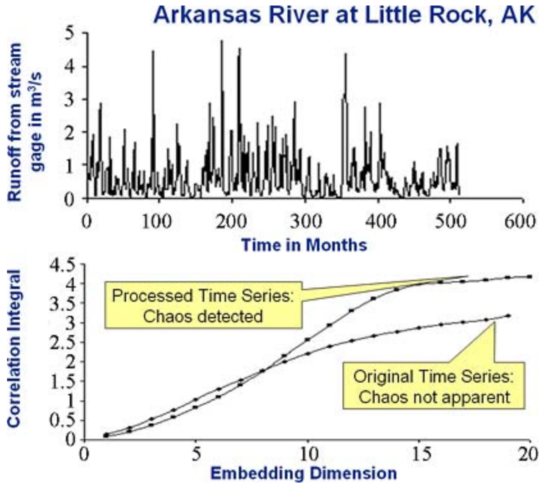
Fig. 13.4 Space-time trends in extremes volatility [30]

quantified and visualized with multiple time windows of data. The color scheme in Fig. 13.4 and the GIS-based visualization was done as part of the [14] study described later.

The red-amber-green color combination is used to denote high (red) to low (green) volatility for the extremes. The extremes volatility index was a new measure based on the ratios of return levels, computed at each grid. The index presented here was normalized to scale between zero and unity by [14]. The ability to quantify and visualize weather and hydrologic extreme values and their properties (e.g., 100-year levels) in space and time is an important first step to studying the impacts on these extremes on infrastructures and human societies. One implication of this study for natural hazards is that it can help evaluate the threat posed by failure by critical infrastructures, such as dams. The extreme volatility index is a measure of the anticipated degree of surprise, or "threat", due to natural extremes. The measure relates, in an aggregate sense, to the expected impacts of extremes. Thus, if critical infrastructures such as dams or levees have been designed to withstand rare 100-year rainfall events (or a rainfall intensity of 0.01 probability of exceedance) then a rarer and more intense event (e.g., a 500-year rainfall) may cause significant damage only if the 500-year intensity is significantly different from the 100-year intensity. This second-order information about relative intensity of extremes is important both for natural variability of the climate system and in situations where global change may cause the extremes to grow more intense. The information can be used for risk-benefit analysis during the design of hydraulic structures and response systems.

3. *Prediction from short and noisy sensor data*: The ability to deal with massive volumes of geographic data from remote and in-situ sensors needs to be complemented by the ability to derive predictive insights from short and noisy geophysical and weather- or climate-related observations. Some of the most relevant his-

Fig. 13.5 Predictability in short and noisy data [29]



torical geophysical data and indices may be limited or incomplete, however the presence of nonlinear dynamics and chaos on the one hand, and colored or even 1/f noise with seasonal fluctuations on the other, cannot be ruled out a priori. In fact, the ability to detect the underlying nonlinear dynamical signals from such data may be of significant value for studies in short- and long-term predictability. Khan et al. [29] developed a methodology to extract the seasonal, random and dynamical components from short and noisy time series, and applied the methods to simulated data and real river flows. Figure 13.5 shows the application to the Arkansas River. The methodology was based on a combination of tools from signal processing, traditional statistics, nonlinear prediction and chaos detection. This methodology can be used to determine how much information is actually contained in the data; such quantification can help determine the appropriate preprocessing approaches and which knowledge discovery techniques to use. The ability to quantify how much predictive insights can be generated from data has direct implications for anticipatory risk-mitigation strategies. Thus, when the available data are completely random, a risk-benefit analysis based on standard deviations may be appropriate, while for completely deterministic signals, an investment in the development of better predictive models followed by recommendations of specific mitigation strategies may be the better strategy. However, for nonlinear dynamics and chaos, the trade-offs between short-term predictability and longer-term error growth need to be carefully balanced, depending on how much information can be extracted from data, especially when the data are noisy and/or limited. Thus, information on the type, quality and quantity of predictive insights that can be generated from data may lead to a determination of preventive actions that can be taken in anticipation of climate, weather and hydrologic extremes.

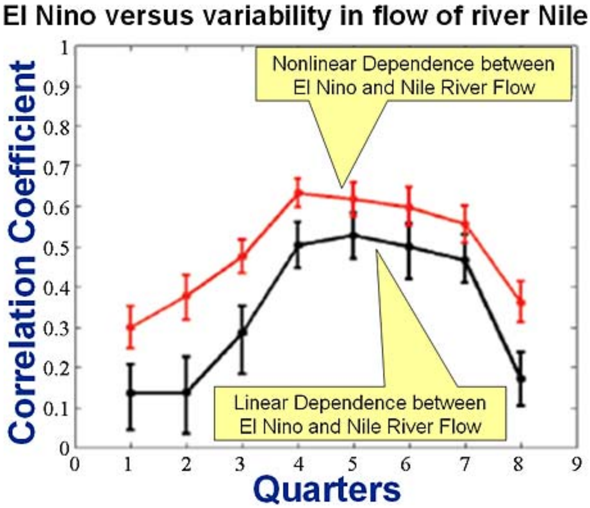


Fig. 13.6 Nonlinear impact of El Niño on rivers [27]

4. *Natural variability and impacts on local geophysical phenomena*: The ability to quantify nonlinear dependence from historical data, even in situations where such data are short and noisy, are critical first steps in studies on predictability, predictive modeling, and physical understanding of weather, climate and geophysical systems.
   Khan et al. [27] developed an approach based on nonlinear dynamics and information theory, along with traditional statistics, to develop and validate new adaptations of emerging techniques. The approach was tested on simulated data, and then applied to an index of the El Niño climate phenomena and the variability in the flow of tropical rivers. The approach revealed more dependence between the variables than previously thought. The ability to quantify the impacts of natural climate variability on weather and hydrologic variables, as shown in Fig. 13.6, can help refine our understanding of the impacts of climate change. The methodology, which can have significant broader impact beyond the case study considered here, was refined and expanded in a another work [28]. Climate systems' response to global changes often leads to natural hazards. We note that the individual and combined impacts of El Niño and global warming have often been advanced as causes for relatively hot or cold summers, as well as the activity of the hurricanes season, in the continental United States. An extraction of causality may be an open research area; however, previous researchers have suggested that natural variability in climate systems, as well as global environmental change, may cause hydrologic or weather extremes at local or regional scales. The ability to quantify the dependency among natural or changing climate phenomena and natural extremes or hazards can help point to appropriate information sources that may guide predictive analyses. This is especially true when larger-scale climate effects can be predicted in advance from data or from

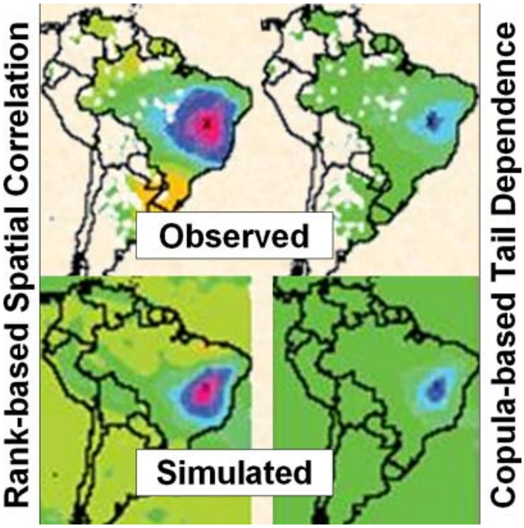


Fig. 13.7 Geospatial-temporal extreme dependence [32]

simulations, which in turn can be used to provide predictive insights on natural hazards.

5. *Comparison of model simulations and historical observations*: Kuhn et al. [32] developed a new approach to quantify the geospatial-temporal dependence among extreme values from massive geographic data. The methodology was motivated by recent developments in multivariate extremes, and hence can be applied to quantify the dependence among extremes of multiple variables—for example, heat waves and precipitation extremes—in space and time. In addition, this copula-based measure can be useful in analyzing simultaneous occurrence of extremes, which may be indicators of possible change.
Thus, if two 100-year events which have zero extremes dependence were to occur simultaneously, this would be a 10,000-year event (see [32], for details), whereas if the extremes have complete dependence, then the simultaneous occurrence still represents a 100-year event. The new measure was utilized on geospatial-temporal rainfall observations and climate model simulations for intercomparisons and model evaluation as shown in Fig. 13.7. In addition, the extremes dependence in space and time was compared with the corresponding spatial correlation values, obtained here through a rank-based correlation measure. Co-occurrence of extremes like heat waves and prolonged droughts can have a combined impact on human lives and economies that is greater than the sum of the individual impacts. The co-occurrence of extremes over space and time may imply a larger regional impact, for example, co-occurrence of extreme rainfall over larger areas may increase the chances of widespread flooding. The relation among extremes in time can be useful for predictive insights regarding the extremes of one variable based on observations of extremes in related variables. The simultaneous and/or frequent occurrence of multiple extremes in space and

time may suggest local, regional or global change in the underlying dynamics of the weather or climate system.

13.6 The Significance of Utilizing Knowledge Discovery Insights for Hazards Mitigation

While debates may persist on the exact causes of natural hazards that led to increased losses of human life and property in recent years, the fact that enhanced predictive insights can help mitigate the impacts of such hazards through improved decisions and policies is becoming relatively well accepted. Pearce [49] described the need for a *shift [in] focus from response and recovery to sustainable hazard mitigation* and laments that in current practices *hazard awareness is absent from local decision-making processes*. However, such sustainable hazard mitigation depends on useful predictive insights from historical data. An interesting early article by Sarewitz and Pielke, Jr. [58] discusses the art of making scientific predictions relevant to policy makers, and uses weather extremes and natural hazards as examples of possible applications. The idea is to move beyond post-disaster consequence management and humanitarian aid disbursal toward preemptive policies based on predictive insights. Decision makers need to realize that while natural disasters may or may not be *acts of God*, their consequences affect humans and can be mitigated by policies; this was highlighted by two recent *Science* magazine articles [3,36].

Hazards mitigation, using predictive insights in some rudimentary form, has been attempted since time immemorial with varying success. However, what has changed dramatically in recent years is the availability of massive volumes of historical and real-time data from sensors. These data, combined with advanced techniques and high-performance computational tools that can be used to extract actionable predictive insights, enhance our understanding of physical processes. This leads to improved short- and longer-term computer simulation models, which in turn are initialized and updated in real time with sensor observations for improved accuracy. The result is a yield of massive volumes of simulation outputs. In this sense, both the new opportunities and the key challenges in generating predictive insights for natural hazard mitigation rely on extracting knowledge from massive volumes of dynamic and distributed sensory data as well as large volumes of computer-based simulations.

An initial investigation in this area was performed by Fuller et al. [14]. Their work investigated how a combination of variables—specifically, the precipitation extremes volatility as defined by Khan et al. [30], the high-resolution population maps described by Bhaduri et al. [2] and used by Sabesan et al. [56], as well as measures representing development or financial indices like the GDP—could be used in conjunction with each other to quantify and visualize the human impacts on natural disasters, specifically those caused by rainfall extremes in South America. Figure 13.8 is a map showing the impacts of weather-related disaster on the human population based on their investigation.

Fig. 13.8 Human impacts of weather related disasters [14]

The computed geospatial indices included the probabilities of truly unusual rainfall extremes, the risks to human population associated with such extremes and the resiliency, or the ability of a region to respond to the disaster. Anticipatory information on disaster damage based on refinements of this study can aid policy makers. Risk metrics can be designed and quantified in space and time based on *threat* or degree of surprise caused by natural disasters, as well as consequences to human population, economies and critical infrastructures. Resiliency metrics for infrastructures and societies can be used in conjunction with risks to develop geospatial and temporal metrics for anticipated impacts. The metrics can provide an overall and objective assessment of potential disaster damage to emergency planners and policy makers at high-resolutions in space and time over large space-time scales. In addition, the various metrics can help planners perform root-cause analysis to determine the critical responsible factors. The overall assessment can help policy makers optimize the level of resource allocations in space and time while the root-cause analysis can help design appropriate mitigation strategies based on the allocated resource at any specific location in any given time.

13.7 Closing Remarks

Predictive insights generated from sensor data, in conjunction with data obtained from other sources like computer-based simulations, can facilitate short-term decisions and longer-term policies. The overall process where raw data from sensors or simulations are ultimately converted to actionable predictive insights for decision and policy makers is defined as knowledge discovery in this chapter. This chapter presents a broader view of knowledge discovery compared to the traditional definitions by the data mining community, but with the data mining and other data sciences as key aspects of the overall process. In addition, we have defined sensors broadly to include wireless sensor networks, in-situ sensor infrastructures and remote sensors. The challenges and opportunities for knowledge discovery based on data from sensors and simulations were described. In particular, we have presented a vision of knowledge discovery in the context of scientific applications. This chapter describes how knowledge discovery from historical and real-time sensor data and computer model simulations can lead to improved predictive insights about weather, climate and associated natural hazards, which can in turn be combined with metrics for disaster risks, consequence, and vulnerability. Scientific applications and scientific knowledge discovery may make sense primarily in the context of a specific domain. Our focus is weather, climate and geophysical hazards. While prediction of natural hazards and mitigating their consequences have been attempted since the dawn of human civilization with varying degrees of success, the possibility of enhanced knowledge discovery from ever-increasing and improving sensor and simulation data make us optimistic that significant breakthroughs may be possible in the near future.

Acknowledgements This research was funded by the Laboratory Directed Research and Development (LDRD) Program of the Oak Ridge National Laboratory (ORNL), managed by UT-Battelle, LLC, for the U.S. Department of Energy under Contract DE-AC05-00OR22725. We are grateful to Dr. Vladimir Protopopescu, Dr. David J. Erickson III and Mark Tuttle, all of ORNL, for their reviews of the manuscript within ORNL's internal publication tracking system. In addition, we are thankful to Drs. Ranga Raju Vatsavai and Steven J. Fernandez, both of ORNL, for their informal reviews and comments. We would also like to thank our co-authors and others who have helped and are continuing to help in our recent and ongoing research highlighted in Sect. 13.6 of this chapter. In particular, we thank Dr. Gabriel Kuhn, Aarthy Sabesan, Kathleen Abercrombie, and Christopher Fuller, who worked on some of the cited projects while Gabriel and Aarthy were at ORNL as post-masters associates, and Kate and Chris as undergraduate interns, respectively. We are grateful to Drs. Vladimir Protopopescu and Brian Worley of ORNL for their contributions to Fig. 13.1, to Professor George Cybenko of Dartmouth for the process detection emblem, and David Gerdes, formerly of ORNL, for his help with the figure. We also gratefully acknowledge the suggestions by the original *PRIDE* team comprising collaborators at ORNL and multiple university partners for their helpful suggestions. Table 1 is an adapted version of the requirements proposed by the first author at a working group meeting of the Weather Extremes Impacts Workshop organized by Los Alamos National Laboratory and the National Center for Atmospheric Research at Sante Fe on February 27–28, 2007. Shiraj Khan and Auroop Ganguly would like to acknowledge the help and support of Professor Sunil Saigal, currently at the University of South Florida (USF), for his support, and for facilitating the completion of Shiraj Ph.D. program at USF with Auroop as his major supervisor. This chapter has been authored by UT-Battelle, LLC, under contract DE-

AC05-00OR22725 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

References

[1] A. Agovic, A. Banerjee, A.R. Ganguly, V.A. Protopopescu, Anomaly detection in transportation corridors using manifold embedding. In: Proceedings of the First International Workshop on Knowledge Discovery from Sensor Data, ACM KDD Conference, San Jose, CA.
[2] B. Bhaduri, E. Bright, P. Coleman, J. Dobson, LandScan: locating people is what matters. Geoinformatics, 5(2):34–37, 2002.
[3] J. Bohannon, Disasters: searching for lessons from a bad year. Science, 310(5756):1883, 2005.
[4] C.E. Brodley, L. Terran, T.M. Stough, Knowledge discovery and data mining, computers taught to discern patterns, detect anomalies and apply decision algorithms can help secure computer systems and find volcanoes on Venus. American Scientist, 87(1):54, 1999.
[5] D. Butler, 2020 computing: everything, everywhere. Nature, 440:402–405, 2006. doi: 10.1038/440402a.
[6] W.D. Collins, C.M. Bitz, M.L. Blackmon, G.B. Bonan, C.S. Bretherton et al., The community climate system model version 3 (CCSM3). Journal of Climate, 19(11):2122–2143, 2006.
[7] N. Cressie, Statistics for Spatial Data. Wiley–Interscience, New York, 1993.
[8] Drye, Mild U.S. hurricane season defied predictions. National Geographic, 30 November 2006.
[9] S. Dzeroski, N. Lavrac, Relational Data Mining. Springer, Berlin, 2001.
[10] D.R. Easterling, G.A. Meehl, C. Parmesan, S.A. Changnon, T.R. Karl, L.O. Mearns, Climate extremes: observations, modeling and impacts. Science, 289(5487):2068–2074, 2000.
[11] ESA http://www.esa.int/esaCP/SEMPMB0XDYD_index_0.html. Downloaded: 30 May 2007.
[12] Y. Fang, A.R. Ganguly, N. Singh, V. Vijayaraj, N. Feierabend, D.T. Potere, Online change detection: monitoring land cover from remotely sensed data, ICDMW, In: Sixth IEEE International Conference on Data Mining—Workshops (ICDMW'06), pp. 626–631, 2006.
[13] U. Fayyad, G. Patietsky-Shapiro, P. Smyth, From data mining to knowledge discovery in databases. AI Magazine, 17(3):37–54, 1996.
[14] C. Fuller, A. Sabesan, S. Khan, A. Kuhn, G. Ganguly, D. Erickson, G. Ostrouchov, Quantification and visualization of the human impacts of anticipated precipitation extremes in South America, Eos Transactions, American Geophysical Union, 87(52), Fall Meeting Supplement, Abstract GC44A-03, 2006.
[15] A.R. Ganguly, A hybrid approach to improving rainfall forecasts. Computing in Science and Engineering, 4(4):14–21, 2002.
[16] A.R. Ganguly, R.L. Bras, Distributed quantitative precipitation forecasting using information from radar and numerical weather prediction models. Journal of Hydrometeorology, 4(6):1168–1180, 2003.
[17] R.L. Grossman, C. Kamath, P. Kegelmeyer, V. Kumar, R.R. Namburu, Data Mining for Scientific and Engineering Applications. Kluwer Academic, Dordrecht, 2001.
[18] V. Guralnik, J. Srivastava, Event detection from time series data. In: Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 33–42, 1999.
[19] A. Hannachi, Quantifying changes and their uncertainties in probability distribution of climate variables using robust statistics. Climate Dynamics, 27(2–3):301–317, 2006.

[20] G. Hegerl, C. Zwiers, F.W. Stott, V.V. Kharin, Detectability of anthropogenic changes in annual temperature and precipitation extremes. Journal of Climate, 17(19):3683–3700, 2004.
[21] T.H. Hinke, J. Rushing, H. Ranganath, S.J. Graves, Techniques and experience in mining remotely sensed satellite data. Artificial Intelligence Review, 14(6):503–531, 2000.
[22] M. Hopkin, Ecology: Spying on nature. Nature, 444:420–421, 2006.
[23] C. Huang, T. Hsing, N. Cressie, A.R. Ganguly, V.A. Protopopescu, N.S. Rao, Statistical analysis of plume model identification based on sensor network measurements. Accepted by ACM Transactions on Sensor Networks, to be published, 2007.
[24] IPCC, Climate Change 2007: The Physical Science Basis. Summary for Policymakers, Intergovernmental Panel on Climate Change, 18 pages, 2007.
[25] R.W. Katz, Stochastic modeling of hurricane damage. Journal of Applied Meteorology, 41:754–762, 2002.
[26] R.W. Katz, M.B. Parlange, P. Naveau, Statistics of extremes in hydrology. Advances in Water Resources, 25:1287–1304, 2002.
[27] S. Khan, A.R. Ganguly, S. Bandyopadhyay, S. Saigal, D.J. Erickson, III, V. Protopopescu, G. Ostrouchov, Nonlinear statistics reveals stronger ties between ENSO and the tropical hydrological cycle. Geophysical Research Letters, 33:L24402, 2006.
[28] S. Khan, S. Bandyopadhyay, A.R. Ganguly, S. Saigal, D.J. Erickson, III, V. Protopopescu, G. Ostrouchov, Relative performance of mutual information estimation methods for quantifying the dependence among short and noisy data. Physical Review E, to appear, 2007.
[29] S. Khan, A.R. Ganguly, S. Saigal, Detection and predictive modeling of chaos in finite hydrological time series. Nonlinear Processes in Geophysics, 12:41–53, 2005.
[30] S. Khan, G. Kuhn, A.R. Ganguly, D.J. Erickson, G. Ostrouchov, Spatio-temporal variability of daily and weekly precipitation extremes in South America. Water Resources Research, to appear, 2007.
[31] D. Kiktev, D.M.H. Sexton, L. Alexander, C.K. Folland, Comparison of modeled and observed trends in indices of daily climate extremes. Journal of Climate, 16(22):3560–3571, 2003.
[32] G. Kuhn, S. Khan, A.R. Ganguly, M. Branstetter, Geospatial-temporal dependence among weekly precipitation extremes with applications to observations and climate model simulations in South America. Water Resources Research, to appear, 2007.
[33] K.E. Kunkel, R.A. Pielke Jr., S.A. Changnon, Temporal fluctuations in weather and climate extremes that cause economic and human health impacts: a review. Bulletin of the American Meteorological Society, 80:1077–1098, 1999.
[34] W. Lee, S.J. Stolfo, K.W. Mok, A data mining framework for building intrusion detection models. IEEE Symposium on Security and Privacy, sp, p.0120, 1999.
[35] T.M. Lillesand, R.W. Kiefer, J.W. Chipman, Remote sensing and image interpretation, 5th edn. Wiley, New York, 2004.
[36] J. Linnerooth-Bayer, R. Mechler, G. Pflug, Refocusing disaster aid. Science, 309(5737):1044–1046, 2005.
[37] MODIS, Moderate Resolution Imaging Spectroradiometer. http://modis.gsfc.nasa.gov/about/, 2007.
[38] J. McEnery, J. Ingram, Q. Duan, T. Adams., L. Anderson, NOAA's advanced hydrologic prediction service: building pathways for better science in water forecasting. Bulletin of the American Meteorological Society, 86:375–385, 2005.
[39] H.J. Miller, J. Han, Geographic Data Mining and Knowledge Discovery. Taylor and Francis, London, 2001.
[40] NASA, Did dust bite the 2006 hurricane season forecast? NASA News, 2007.
[41] NASA, Earth Observatory Natural Hazards. http://earthobservatory.nasa.gov/NaturalHazards. Downloaded: 30 May 2007.
[42] NASA, Earth Science Remote Sensing. http://science.hq.nasa.gov/earth-sun/technology/remote_sensing.html. Downloaded: 30 May 2007.
[43] NASA, A Growing Intelligence Around Earth, Science@NASA, Headline News Feature, 2006.

[44] NewScientistTech, http://www.newscientisttech.com/article/dn10360. Downloaded: 30 May 2007.
[45] D. Nychka, C.K. Wikle, J.A. Royle, Multiresolution models for nonstationary spatial covariance functions. Statistical Modelling: An International Journal, 2:315–331, 2002.
[46] G.R. North, K.-Y. Kim, S.S.P. Shen, J.W. Hardin, Detection of forced climate signals. Part I: Filter theory. Journal of Climate, 8(3):401–408, 1995.
[47] Oak Ridger, ORNL project aims to lessen impact of natural disasters, Oak Ridger, 2007.
[48] S. Papadimitriou, J. Sun, C. Faloutsos, Streaming pattern discovery in multiple time-series, VLDB, pp. 697–708, 2005.
[49] L. Pearce, Disaster management and community planning, and public participation: how to achieve sustainable hazard mitigation. Natural Hazards, 28(2–3):211–228, 2003.
[50] M. Pelling, Natural disasters and development in a globalizing world. Rutledge, 2003.
[51] D. Potere, N. Feierabend, E. Bright, A. Strahler, Walmart from space: a new source for land cover change validation. Photogrametric Engineering and Remote Sensing, accepted for publication, 2007.
[52] C. Potter, P.-N. Tan, M. Steinbach, S. Klooster, V. Kumar, R. Myneni, V. Genovese, Major disturbance events in terrestrial ecosystems detected using global satellite data sets. Global Change Biology, 9(7):1005–1021, 2003.
[53] D. Puccinelli, M. Haenggi, Wireless sensor networks: applications and challenges of ubiquitous sensing. IEEE Circuits and Systems Magazine, 5(3):19–31, 2005.
[54] R.H. Reichle, D.B. McLaughlin, D. Entekhabi, Hydrologic data assimilation with the Ensemble Kalman filter. Monthly Weather Review, 130(1):103–114, 2002.
[55] J.F. Roddick, K. Hornsby, M. Spiliopoulou, An updated bibliography of temporal, spatial and spatio-temporal data mining research. Lecture Notes in Computer Science, 147 pages, 2001.
[56] A. Sabesan, K. Abercrombie, A.R. Ganguly, B.L. Bhaduri, E.A. Bright, P. Coleman, Metrics for the comparative analysis of geospatial datasets with applications to high resolution grid-based population data. GeoJournal, to appear, 2007.
[57] A. Sankarasubramanian, U. Lall, Flood quantiles in a changing climate: seasonal forecasts and causal relations. Water Resources Research, 39(5):1134, 2003.
[58] D. Sarewitz, R. Pielke, Jr., Prediction in science and policy. Technology in Society, 21:121–133, 1999.
[59] Q. Schiermeier, Insurers' disaster files suggest climate is culprit: rising costs hint at weather effect. Nature, 441:674–675, 2006.
[60] ScienceDaily, Very active 2007 hurricane season predicted. Science Daily, 2007.
[61] ScienceDaily, Study aims to ease natural disaster impact. Science Daily, 2006.
[62] SeaWiFS, Background of the SeaWiFS Project. http://oceancolor.gsfc.nasa.gov/SeaWiFS. Downloaded, 30 May 2007.
[63] M. Steinbach, P.-N. Tan, V. Kumar, C. Potter, S. Klooster, Discovery of climate indices using clustering. In: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 446–455, 2003.
[64] K.J. Tierney, M.K. Lindell, R.W. Perry, Facing the Unexpected: Disaster Preparedness and Response in the United States. The National Academies Press, 318 pages, 2001.
[65] UNDP, Reducing Disaster Risk: A Challenge for Development—A Global Report. United Nations Development Programme, 164 pages, 2004.
[66] USGS, http://earthquake.usgs.gov/eqcenter/pager/. Downloaded, 30 May 2007.
[67] L.J. Vale, T.J. Campanella, M.W. Fishwick, The resilient city: how modern cities recover from disaster. The Journal of American Culture, 28(4):456, 2005.
[68] G.M. Weiss, Mining with rarity: a unifying framework. SIGKDD Explorations, 6(1):7–19, 2004.
[69] G.M. Weiss, H. Hirsh, Learning to predict rare events in event sequences. In: Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining, pp. 359–363, 1998.

[70] R. White, D. Etkin, Climate change, extreme events and Canadian insurance. Natural Hazards, 16(2–3):135–163, 2004.
[71] F.W. Zwiers, H. von Storch, On the role of statistics in climate research. International Journal of Climatology, 24(6):665–680, 2004.