2004

Drought Monitoring Using Data Mining Techniques: A Case Study for Nebraska, USA

Tsegaye Tadesse
University of Nebraska-Lincoln, ttadesse2@unl.edu

Donald A. Wilhite
University of Nebraska - Lincoln, dwilhite2@unl.edu

Sherri K. Harms
University of Nebraska-Lincoln, harmssk@unk.edu

Michael J. Hayes
University of Nebraska-Lincoln, mhayes2@unl.edu

Steve Goddard
University of Nebraska-Lincoln, goddard@cse.unl.edu

Published in *Natural Hazards* 33 (2004), pp. 137–159.

Submitted April 4, 2003; accepted October 17, 2003.

Drought Monitoring Using Data Mining Techniques: A Case Study for Nebraska, USA

Tsegaye Tadesse,[1] Donald A. Wilhite,[1] Sherri K. Harms,[2]

Michael J. Hayes,[1] and Steve Goddard[3]

1. National Drought Mitigation Center, University of Nebraska, Lincoln, Nebraska, 68583-0728, USA; email ttadesse2@unl.edu, dwilhite1@unl.edu, mhayes2@unl.edu
2. Department of Computer Science and Information Systems, University of Nebraska, Kearney, Nebraska, 68849, USA; email harmssk@unk.edu
3. Computer Science and Engineering, University of Nebraska, Lincoln, Nebraska, 68588, USA; email goddard@cse.unl.edu

Abstract

Drought has an impact on many aspects of society. To help decision makers reduce the impacts of drought, it is important to improve our understanding of the characteristics and relationships of atmospheric and oceanic parameters that cause drought. In this study, the use of data mining techniques is introduced to find associations between drought and several oceanic and climatic indices that could help users in making knowledgeable decisions about drought responses before the drought actually occurs. Data mining techniques enable users to search for hidden patterns and find association rules for target data sets such as drought episodes. These techniques have been used for commercial applications, medical research, and telecommunications but not for drought. In this study, two time-series data mining algorithms are used in Nebraska to illustrate the identification of the relationships between oceanic parameters and drought indices. The algorithms provide flexibility in time-series analyses and identify drought episodes separate from normal and wet conditions, and find relationships between drought and oceanic indices in a manner different from the traditional statistical associations. The drought episodes were determined based on the Standardized Precipitation Index (SPI) and Palmer Drought Severity Index (PDSI). Associations were observed between drought episodes and oceanic and atmospheric indices that include the Southern Oscillation Index (SOI), the Multivariate ENSO Index (MEI), the Pacific/North American (PNA) index, the North Atlantic Oscillation (NAO) Index, and the Pacific Decadal Oscillation (PDO) Index. The experimental results showed that among these indices, the SOI, MEI, and PDO have relatively stronger relationships with drought episodes over selected stations in Nebraska. Moreover, the study suggests that data mining techniques can help us to monitor drought using oceanic indices as a precursor of drought.

Key words: drought indices, oceanic indices, drought, data mining, decision making

1. Introduction

Drought is an adverse environmental phenomenon that influences almost all aspects of society. It is a normal feature of climate and its occurrence is inevitable (Wilhite, 2000a; Rosenberg, 1978). A drought's impact and extent varies from one region to another. However, the strategies and principles underlying the task of managing drought risk are similar, regardless of the specific situations that countries face (O'Meagher et al., 2000). This implies that extracting information about drought characteristics that include its spatial extent, severity, and frequency are important everywhere to decision makers such as government officials and other users with wide regional and national interests.

To identify relationships between different climatic parameters and to distinguish patterns that may be used to predict drought, large historical data sets are essential. In light of this, it is critical to have an efficient way to extract information from large databases and to deliver relevant and effective information for drought risk management. One of the recently developed techniques for such purposes is data mining.

Data mining uses a variety of data analysis tools to discover patterns and relationships of physical variables in different data sets (Two Crows, 1999). This technique is used in multidisciplinary fields. For example, the method is used for commercial applications by many companies for designing strategic benefits to increase profitability (Groth, 1998). One of the applications of data mining for these commercial companies is predicting (or identifying) the customers most likely to buy certain products. This helps decision makers to effectively identify market demand. It is also used to understand trends in the marketplace. This reduces costs and improves timeliness of products reaching the market. Recent studies consider this method to be one of the best tools to identify the demand and supply of any product, as well as the patterns of the customers to be profitable in an "emerging market" (Cabena et al., 1998; Groth, 1998; Kryszkiewicz, 1998). In this study, we employ data mining to identify complex relationships involving atmospheric and oceanic variables that potentially cause droughts over selected stations and statewide areas of Nebraska, USA (fig. 1).

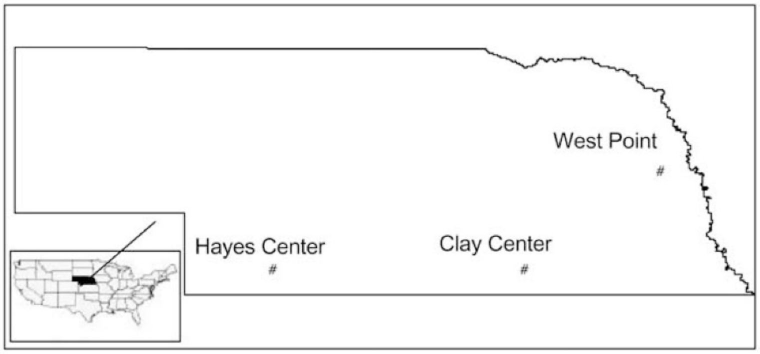

Figure 1. Map of selected stations to illustrate the use of data mining in Nebraska, USA.

2. The Problem

In studying drought characteristics and impacts, Wilhite (1993) emphasized that the occurrence of extended drought over several months, seasons, or years often results in serious economic, environmental, and social consequences. Moreover, multiyear drought events result in significant economic impacts in many sectors such as agriculture, transportation, and energy.

According to Dewey (1996), the world's vulnerability to drought has increased steadily over the centuries primarily because of an ever-increasing population that puts heavy demand on water and natural resources. Over the past century, virtually all portions of the United States experienced several extended severe droughts as well as many short-term droughts, resulting in considerable losses (Svoboda et al., 2002; Wilhite, 2000b; Hayes et al., 1999).

In their study, Ross and Lott (2000) pointed out that the eight major droughts that occurred in the US between 1980 and 1999 accounted for the largest percentage (43%) of weather-related monetary losses. The second largest percentage (30%) was due to hurricanes and tropical storms. This shows that the damage due to drought is far greater than the damage from other weather-related disasters. In 1995, the Federal Emergency Management Agency (FEMA) estimated that droughts cause $6–8 billion in losses in the US each year, more than any other weather-related disaster (FEMA, 1995). In 2002, about 30 states in Great Plains, much of the eastern US, and western states were in drought (NCDC, 2002). The preliminary estimates of damages/costs due to this drought were more than $10 billion. Because of these significant impacts, drought monitoring is an integral part of drought preparedness planning and is essential in making sound and timely decisions that will reduce drought impacts in the future.

An increased understanding of drought events will improve the development and implementation of drought planning and mitigation actions (Wilhite, 2000b). One of the challenges in understanding drought is the large volume of data for numerous climate and hydrologic variables and indices, and the variety of spatial and time scales for which these data exist. Data mining is one recently developed technique that might help solve this problem, improving drought monitoring by identifying spatial and temporal patterns of drought characteristics as well as finding the association of these characteristics with oceanic processes. This information can be used in making knowledge-driven decisions to reduce the impacts of drought through better monitoring.

It is the goal of this study to find the relationships between oceanic/atmospheric indices and drought; and to identify the drought episodes within a certain time lag of the occurrences of oceanic and atmospheric indices so that the product can be used by decision makers in Nebraska, USA. This will increase the knowledge of drought characteristics for more effective drought planning and management. In this study, two time-series data mining algorithms are used in Nebraska to illustrate the methodology used to identify the relationships between oceanic/atmospheric parameters and drought indices. It is hoped that this method can provide additional information that will enhance decision makers' ability to take appropriate and timely mitigation actions.

3. Background

3.1. Brief Explanations of the Indices Used in This Study

In monitoring climatic parameters such as precipitation, some studies (Glantz, 1994; Ogallo, 1994; Philander, 1990; Ropelewski and Halpert, 1987) attempted to find the global relationships of ocean-atmosphere interactions and climatic parameters such as precipitation variability. The results indicate that the upper air and surface synoptic meteorological conditions are affected by this strong relationship. For example, a large, persistent upper-level high-pressure ridge over the central part of the United States contributes to dryness in Nebraska. This "blocking" upper-air pattern can be linked to sea surface temperature anomalies in the Pacific and Atlantic Oceans (Ropelewski and Halpert, 1986). Based on such relationships, it is important to consider the impacts of the variability of the oceanic/atmospheric parameters using the oceanic/atmospheric indices while monitoring drought. Moreover, oceanic parameters such as Sea Surface Temperature (SST) and oceanic processes such as El Niño and Southern Oscillation (ENSO) are changing in a much longer timescale than processes over land (Wallace and Vogel, 1994; Diaz and Markgraf, 1992). Considering this characteristic, it is logical to use the oceanic data to monitor climatic parameters such as precipitation deficit based on its associations with oceanic parameters. Furthermore, the models that are used to predict these oceanic parameters are improving with more observations and technological advancement (McPhaden et al., 1998; Halpert and Ropelewski, 1992; Philander, 1990). The following most common oceanic and atmospheric indices were selected for this study.

3.1.1. The Southern Oscillation Index (SOI)

The SOI is computed using monthly mean sea level pressure anomalies at Tahiti (T), French Polynesia, and Darwin (D), Australia. The standardized monthly mean sea level pressure anomaly SOI [T-D] is an index that combines the Southern Oscillation into one series (Trenberth and Hoar, 1996). A positive SOI means La Niña and negative SOI indicates El Niño conditions.

3.1.2. Pacific/North American (PNA) Index

The PNA index is derived using the formula: PNA = 0.25* [Z(20N,160W) – Z(45N,165W) + Z(55N,115W) – Z(30N,85W)], where Z is a standardized 500 hPa geopotential height value (Wallace and Gutzler, 1981). The values in brackets show the latitude and the longitude position of the geopotential heights. This index shows the upper atmosphere conditions. Monthly average values of the PNA are used in this study.

3.1.3. Multivariate ENSO Index (MEI)

MEI is calculated based on six main observed variables over the tropical Pacific. These six variables are sea-level pressure, zonal and meridional components of the surface wind, sea surface temperature, surface air temperature, and total cloudiness fraction of the sky (Wolter and Timlin, 1993). A positive MEI is associated with El Niño and negative values indicate La Niña conditions.

3.1.4. The Pacific Decadal Oscillation (PDO) Index

The PDO index is defined as the leading principal component of North Pacific monthly sea surface temperature variability poleward of 20N (Francis and Hare, 1994). The monthly values of the PDO index are used to associate the North Pacific conditions with the local drought in Nebraska in this study. The positive values show the warm phase of the North Pacific sea surface temperature, and the negative values show the cold phase.

3.1.5. North Atlantic Oscillation (NAO) Index

The NAO index is defined as the normalized pressure difference between a station on the Azores and one on Iceland (Hurrell, 1995). Since it is computed based on stations to the north (Iceland) and south (Azores) of the middle latitude westerly flow, it could be considered as a measure of the strength of these winds. The positive values show strong mid-latitude westerly flow while the negative values show the weak mid-latitude westerly flow (Hurrell, 1995).

3.2. Data Mining Concepts

Data mining is a recent technology with great potential for identifying the most important information in databases. Data mining is part of a larger process called *knowledge discovery*. Essentially, data mining discovers patterns and relationships hidden within large amounts of data. Data mining may be considered as advances in statistical analysis and modeling techniques to find useful patterns and relationships (Edelstein, 1997). Data mining algorithms are also useful for data automation that is designed to allow users to create "intelligent" data sets by discretizing or converting data from many formats into compatible and user friendly formats. This process of automation makes it easier to execute different functions of the algorithms to the desired output without human interference and within a relatively short time. Thus, data mining tools can answer questions that traditionally were too time-consuming to resolve. These tools search databases for hidden patterns and find predictive information that experts may miss because it lies outside their expectations.

Recent developments in computing have provided the basic infrastructure for fast data access as well as many advanced computational methods for extracting information from large quantities of data. These developments have created a new range of problems and challenges for data analysts as well as new opportunities for intelligent systems in data analysis. According to Thearling (2001), data mining techniques are the result of a long process of research and product development. He indicates that this evolution began when business data were first stored on computers, continued with improvements in data access, and, more recently, generated technologies that allow users to navigate through their data in real time. Data mining takes this evolutionary process beyond retrospective data access and navigation to prospective and proactive information delivery. Data mining algorithms represent techniques that have recently been implemented as mature, reliable, understandable tools that are consistently outperforming older statistical methods in commercial applications (Thearling, 2001). Studies (Bigus, 1998; Cabena et al., 1998) show that data mining tools can also be used in predicting future trends and behaviors, allowing businesses to make proactive, knowledge-driven decisions.

Similarly, data mining algorithms and models such as decision trees, associations, clustering, classification, regression, sequential patterns, and time series forecasting have the potential to identify drought patterns and characteristics. For example, time series data mining can be applied in monitoring patterns of drought events.

4. Time-Series Data Mining

Time-series data mining applications organize data as a sequence of events, with each event having a time of occurrence. In data analysis applications on a sequence of events, one of the main challenges is finding similar situations. This is essential when trying to predict future events and understand the dynamics of the process producing the sequence (Mannila and Seppänen, 2001).

Time-series data mining algorithms are being developed for many applications to identify hidden patterns within time-series data (Berry and Linoff, 2000; Klemettine, 1999; Groth, 1998). These algorithms are designed to characterize and predict nonperiodic complex phenomena (Povinelli, 2000; Huang and Yu, 1999; Keogh and Pazzani, 1998; Edelstein, 1997). Because drought is sensitive to the time sequence of atmospheric and oceanic parameters, time-series data mining techniques can better contribute to identification of drought episodes and quantification of the relationships between the oceanic and atmospheric parameters. Thus, in a real-world application such as drought, it is important to study the relationships of the oceanic and atmospheric parameters that cause drought by considering the time of their occurrences.

In this study, two recently developed time-series data mining algorithms were used to find the relationships of drought and oceanic/atmospheric indices by considering time lags of their occurrences (Harms et al., 2002; Tadesse, 2002). These algorithms are the Representative Episodal Association Rule (REAR) and the Minimal Occurrences With Constraints and Time Lags (MOWCATL). Both algorithms are briefly discussed in the following sections.

4.1. The Representative Episodes Association Rules (REAR) Algorithm

The Representative Episodes Association Rules (REAR) algorithm (Harms et al., 2001b) converts the time-series data into discrete representations and generates association rules. The preprocessing of the time-series data for the REAR algorithm begins by discretizing (segmenting into groups of records) the data. These discretized data form *events* in sequential data. In a data-mining context, a combination of events in an event sequence creates an episode with a time-specified order (Mannila et al., 1997). In other words, an episode occurs in a sequence if events are consistent with the given order, within a given time frame, called *window width*. Thus, an episode *P* is a pair (*V*, *type*), where *V* is a collection of events. In this pair, the type of episode is called *parallel* if no order is specified in a window, and *serial* if the events of the episode have a fixed order. The frequency of an episode is defined as the number of windows in which the episode occurs divided by the number of total windows in the data set. The REAR algorithm finds episodes of events that occur together in a relatively short time interval (window width). To process the data using the REAR algorithm, a *sliding window* is used by sequentially moving the window one step at a time through the data. The user can set the window width and the minimum frequency value. With the complete set of repeated episodes, association rule patterns within the episodes are generated.

The association rule generated is defined as "if *X then Y*," where *X* is the rule antecedent and *Y* is its consequent. The support and confidence of the rule are used to measure the goodness ("interestingness") of the rule. Support of the rule is simply a measure of statistical significance. It is defined as the ratio of the count of the number of windows of the episodes *X* and *Y* occurring together divided by the total number of windows. In association rules, the support of the antecedent is also called the rule coverage. To select the best rule, one of the measures used is the confidence of the rule. For each rule, the confidence is defined as the ratio of support $\{X \cup Y\}$, which is the percentage of events that occur in both *X* and *Y*, divided by the support of *X* (rule coverage). This is a measure of the conditional probability. For example, if *X* occurred 16% of the time and both *X* and *Y* occurred together 10% of the time, then the confidence is 10/16 (63%). In a sequential data mining algorithm, this computation of confidence is different from the traditional statistical computation of confidence.

To generate rules for drought, the REAR algorithm counts occurrences of the drought episodes that occur together with other oceanic parameters within the sliding window. Then the algorithm keeps the repeated episodes that occur at more than the predetermined minimum frequency (Harms et al., 2002). Rules are generated using the antecedent (*oceanic indices*) and consequent (*drought episodes*) constraints to keep track of the target episodes (droughts). This identifies the global oceanic parameters and drought episodes of the association rules in the form of: if *X* (e.g., *global oceanic indices values*) then *Y* (e.g., *moderate, severe, or extreme drought*) occurs with more than a preassigned minimum confidence. For example, the rule may hypothetically show that if the Southern Oscillation Index (SOI) values are greater than 1.5, then severe droughts occur in Nebraska with more than 80% confidence.

Figure 2 shows the flow chart used to generate rules using the REAR algorithm. It should be noted that the number of rules generated depends on the minimum frequency, the window width, and the minimum confidence values. In selecting these parameters one may have to consider the advantages and disadvantages of the parameters on the outputs (i.e., rules that are generated). For example, if a wider window width is selected, more relationships may be found but the analysis and interpretation of the rules may be difficult.

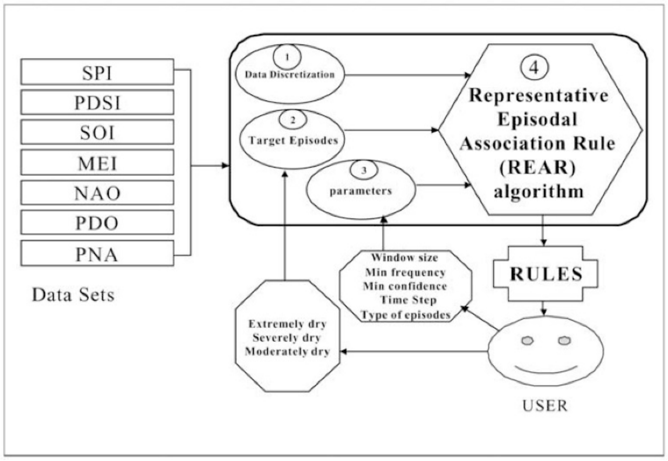

Figure 2. Generation of rules using representative episodal association rule (REAR) algorithm.

4.2. The Minimal Occurrences With Constraints And Time Lags (MOWCATL) Algorithm

The Minimal Occurrences With Constraints And Time Lags (MOWCATL) algorithm is used to find relationships between sequences in the multiple data sets, where a lag in time exists between the antecedent and the consequent. In addition to the traditional frequency and support constraints in sequential data mining, MOWCATL uses separate antecedent and consequent inclusion constraints, and separate antecedent and consequent maximum window widths, to specify the antecedent and consequent patterns that are separated by a time lag. This approach is based on association rules combined with frequent episodes, time lags, and event constraints (Harms et al., 2002). The advantage of this algorithm over the REAR method is that it can handle a lag time between the occurrence of the antecedent and the occurrence of the consequent. The MOWCATL algorithm is attractive for drought applications because it is quite likely that there is a natural time lag between the oceanic parameters (antecedents) and the drought events (consequents).

Technically, the MOWCATL approach identifies minimal occurrences of episodes along with their time intervals. The minimal occurrence of an episode in an event sequence is determined when the episode occurs in the window and does not occur in any other proper subwindow. Then, instead of counting the frequency of the episodes as in REAR, the number of minimal occurrences is counted as the support of the episode. Episodes that do not meet the minimal support threshold are pruned when the rules are generated.

MOWCATL has three window parameters: the maximum window width of the antecedent (*wina*), the maximum window width of the consequent (*winc*), and the time lag.

Using these parameters, the algorithm generates episodal rules where the antecedent episode occurs within a given maximum window width, the consequent episode occurs within a given maximum window width, and the start of the consequent follows the start of the antecedent within a given maximum time lag. A hypothetical example of a generated rule might be that: *if SOI is greater than 1 and MEI is less than –1.5, and these occur within 2 months of each other, then within 3 months they will be followed by severe drought occurring with more than 70% confidence*.

MOWCATL can find serial and parallel episodes. Serial episodes within a time series are episodes that consider the time order of the occurrence of each episode within a window, whereas in parallel episodes the episodes can occur without time order. For serial episodes, the starting time of the oceanic parameters must be greater than or equal to the ending time of the drought events, and must be less than or equal to the starting time of the oceanic parameters plus the time lag. Also, the drought event ending time must be greater than the ending time of the oceanic parameters. For a zero time lag, the REAR algorithm can be used instead of MOWCATL. In contrast, for parallel episodes, the starting time of the drought events must follow the starting time of the oceanic parameters and can differ at most by the time lag. Because the order of the events in the parallel episodes is not important, parallel episodes are used to see if the events in one episode occur "close" to the events in the other episode.

Figure 3 shows the flow chart used to generate rules using the MOWCATL algorithm. In the MOWCATL algorithm, the time lag constraint can be either a fixed time lag constraint or a maximum time lag constraint. With fixed time lag, the occurrence of oceanic parameters (antecedent) and the drought episodes (consequents) are separated with fixed time. This may be used to monitor parameters that occur an exact number of time lags before the consequent. With the maximal time lag constraint, the start of the consequent follows the start of the antecedent after at least one time step, and at most the set lag time steps. This may be used to monitor parameters that occur within a range of time before the consequent.

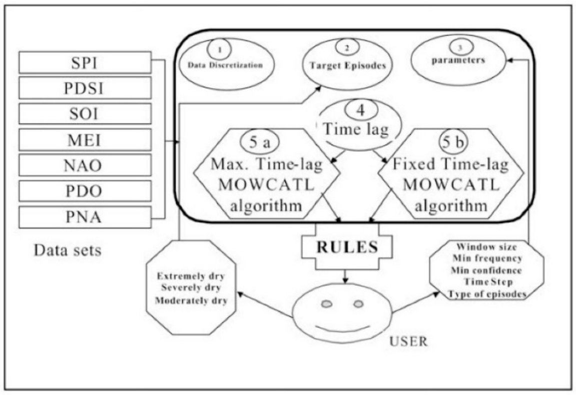

Figure 3. Generation of rules using minimal occurrences with constraints and time lags (MOWCATL) algorithm.

Both REAR and MOWCATL rule discovery methods use the J-measure (Smyth and Goodman, 1992) as an objective measure to select best rules. The formulation of J-measure takes into consideration both frequencies on the left and right sides of a rule. Therefore, it not only favors rules that occur more frequently but also provides a more complex metric for ranking rules in a manner such that the user can trade off rule support and rule confidence.

In general, the MOWCATL approach is well suited for sequential data mining problems that have groupings of events that occur close together, but occur relatively infrequently over the entire data set. It is also well suited for problems that have periodic occurrences when the signature of one or more sequence is present in other sequences, even when the multiple sequences are not globally correlated.

Both REAR and MOWCATL algorithms are designed as exploratory methods. Thus, iterative and interactive application of the approach coupled with human interpretation of the rules is likely to lead to the most useful results, rather than a fully automated approach (Harms et al., 2002). These analysis techniques facilitate the evaluation of the temporal associations between episodes of events and the incorporation of this knowledge into decision support systems. In the following section, this new approach was used to conduct an experiment on selected Nebraska stations.

5. Experimental Results

In monitoring drought, time-series data of meteorological, climatological, and oceanic parameters were used to generate rules that could identify the probable occurrence of drought. The justification for using oceanic parameters to monitor drought is based on the general assumption that ocean-atmosphere relationships have an impact on drought. Since changes in oceanic parameters develop more slowly than surface meteorological parameters, one can understand the trend of the oceanic parameters better than the trend of surface parameters such as surface temperature changes or precipitation deficits (McPhaden et al., 1998). Thus, oceanic parameters can be considered as antecedents and droughts as consequents in finding their relationships.

As an experiment, both algorithms were used to find relationships between drought episodes at several stations in Nebraska and other climatic oceanic indices for the period from 1950 to 1999. The atmospheric and oceanic indices used in this experiment include the PNA, MEI, NAO, PDO, and SOI. The drought indices (i.e., PDSI and SPI) for each station were also used to generate rules. Out of these stations, to demonstrate the uses of these algorithms, three weather stations in Nebraska were selected (fig. 1). These stations were Clay Center (south-central), Hayes Center (southwest), and West Point (northeastern Nebraska). For a comparison and a broader look at the state level, the Nebraska state-averaged climate data were also used.

In this experiment, the monthly values from the oceanic and drought indices were converted into discrete representations and classified into seven categories. The seven drought indices categories are extremely dry, severely dry, moderately dry, normal, moderately wet, severely wet, and extremely wet. The thresholds for classification of the drought indices are shown in table I. The seven SPI and PDSI drought categories are based on the classifications used by the National Drought Mitigation Center (Hayes, 2003). However, in the PDSI classification, we aggregated the values of “incipient” and “mild dry” categories within normal ranges to make it consistent with seven classification categories.

Table I. Threshold values used to classify drought episodes

Drought category	Extremely dry (ed)	Severely dry (sd)	Moderately dry (md)	Normal (n)	Moderately wet (mw)	Severely wet (sw)	Extremely wet (ew)
SPI	≤ –2	–2 < x ≤ –1.5	–1.5 < x ≤ –1	–1 < x < 1	1 ≥ x < 1.5	1.5 ≥ x < 2	≥ 2
PDSI	≤ –4	–4 < x ≤ –3	–3 < x ≤ –2	–2 < x < 2	2 ≥ x < 3	3 ≥ x < 4	≥ 4

SPI = Standardized Precipitation Index; PDSI = Palmer Drought Severity Index

The oceanic parameters were also divided into seven categories using thresholds, as shown in table II, based on their frequency distribution through their historical records. Assuming a normal distribution of the 50 years data, each oceanic and atmospheric parameter values divided into 0.5, 1, and 1.5 standard deviations from the normal frequency distribution. Thus, table II shows the division of these values based on this criterion for classifying the values for each parameter. For example, the SOI data was discretized into 7 clusters as follows: SOI1 is Category 1 in which the values of SOI is greater than 1.5 standardized deviation (STDEV), SOI2 is Category 2 with values within 1 and 1.5 STDEV, SOI3 is Category 3 with values within 0.5 and 1 STDEV, SOI4 is Category 4 with values within 0.5 and –0.5 STDEV, SOI5 is Category 5 with values within –0.5 and –1 STDEV, SOI6 is Category 6 with values within –1 and –1.5 STDEV, and SOI7 is Category 7 with values less than –1.5 STDEV. All other oceanic and atmospheric parameters followed the same rule to determine the thresholds in discretizing the data. However, small adjustments have been made to round off the data values to the nearest 0.5 standard deviation values.

Table II. Threshold values used to classify oceanic and climatic indices

Indices	Category	1	2	3	4	5	6	7
SOI	≥ 1.5	1 ≤ x < 1.5	0.5 ≤ x < 1	–0.5 < x < 0.5	–1 < x ≤ –0.5	–1.5 < x ≤ –1	≤ –1.5	
MEI	≤ –1.5	–1.5 < x ≤ –1	–1 < x ≤ –0.5	–0.5 < x < 0.5	0.5 ≤ x < 1	1 ≤ x < 1.5	≥ 1.5	
NAO	≤ –4	–4 < x ≤ –3	–3 < x ≤ –2	–2 < x < 2	2 ≤ x < 3	3 ≤ x < 4	≥ 4	
PDO & PNA	≤ –2	–2 < x ≤ –1.5	–1.5 < x ≤ –1	–1 < x < 1	1 ≤ x < 1.5	1.5 ≤ x < 2	≥ 2	

SOI = Southern Oscillation Index; MEI = Multivariate ENSO Index; NAO = North Atlantic Oscillation Index; PDO = Pacific Decadal Oscillation index; PNA = Pacific/North American index

After discretization, the drought episodes (extremely dry, severely dry, and moderately dry) were identified as constraints based on the climatic drought indices, i.e., the SPI (Standardized Precipitation Index) and the PDSI (Palmer Drought Severity Index) values. Then, the oceanic and atmospheric parameters were used as the rule antecedents whereas the drought episodes were used as the rule consequents.

After the rules were generated, the “interestingness” (goodness) of the rule was measured by ranking the rules with Smith and Goodman’s J-measure. In addition to confidence values, the J-measure values show the goodness of the rules (Smith and Goodman, 1992) that allows selecting the better rules among the ones that are generated. For drought, which shows relatively lower occurrences in the whole precipitation data sets as compared to the total sum of the occurrences of normal and wet conditions, the J-measure values that were more than 0.04 were generally considered interesting (good) rules (Harms, 2003; Tadesse, 2002).

Using the REAR algorithm with serial (time-ordered) episodes, sample rules generated within a three-month time window are shown in table III. To illustrate the meaning of the rules generated, it may help to look in more detail at some of the rules generated for these selected stations and state-averaged data. For example, for Clay Center, the rules in table I include: *if the MEI value was less than –1.5, and the PDO value was less than –2, then the PDSI value was extremely dry (less than –4) with 64% minimum confidence*. This shows that when the combinations of the two oceanic parameters are both negative in value, then there is a reasonable probability of drought. The same rule applies for Hayes Center and West Point (table I). For the state-averaged data of Nebraska, the rules generated include: *if the PDO value was less than –2, and the MEI value was less than –1.5, then the SPI value was moderately dry with 56% minimum confidence and 0.06 J-measure valu*e. The other significant rule generated for Hayes Center, West Point, and the state-averaged data was: *if the NAO value was between –3 and –2, and the PDO value was less than –2, then the PDSI value was extremely dry with more than 60% confidence and 0.04 J-measure*. These rules show that the global oceanic conditions measured by the MEI, PDO, and NAO indices can be precursors for drought conditions in selected stations and for the state with a given percentage of confidence. The advantage of the REAR algorithm over the MOWCATL is that it can find relationships of events that occur within the specified time window.

Table III. Sample rules generated using the REAR serial algorithm with 3-month window size

Location	Selected serial rules	Confidence	J-measure
Clay Center	MEI1, PDO1 ⇒ PDSIed	0.64	0.08
	NAO3, PDO1 ⇒ SP12sd	0.60	0.04
Hayes Center	MEI1, PDO1 ⇒ PDSIed	0.64	0.08
	NAO3, PDO1 ⇒ PDSIed	0.60	0.04
West Point	NAO3, PDO1 ⇒ PDSIed	0.60	0.04
	MEI1, PDO1 ⇒ PDSIed	0.55	0.07
Nebraska	NAO3, PDO1 ⇒ PDSIed	0.60	0.04
(State average)	PDO1, MEI1 ⇒ SPI6md	0.56	0.06

With the MOWCATL algorithm, to record the frequent episodes and generate rules, a variety of window widths, minimum frequency values, and minimum confidence values were selected for analysis. Table IV and table V show sample rules generated using MOWCATL with maximum time lag and MOWCATL with fixed time lag algorithms, respectively. For example, using MOWCATL, with a two-month antecedent and consequent windows size, and a three-month maximum time lag between the start of the antecedent oceanic parameters and the start of the consequent drought episodes, one of the rules generated for Clay Center, Nebraska, is: *if the MEI was less than –1.5 and the PDO was less than*

–2, then within 3 months the PDSI was extremely dry and the twelve-month SPI was severely dry with more than 88% confidence and 0.08 J-value.

Table IV. Sample rules generated using MOWCATL algorithm for serial episodes, antecedent and consequent windows of 2-months, and the 3 months maximum time lag between the start of the antecedent and the start of the consequent

Location	Selected serial rules	Confidence	J-measure
Clay Center	MEI1, PDO1 ⇒ PDSIed, SP12sd	0.88	0.08
	MEI3, PNA2 ⇒ SPI9sd, PDSIed	0.75	0.04
Hayes Center	MEI1, PDO1 ⇒ PDSIed	0.88	0.09
	MEI1, PDO1 ⇒ SP12md	0.75	0.08
West Point	MEI1, PDO1 ⇒ PDSIed	0.88	0.09
	MEI3, PNA3 ⇒ PDSImd	0.75	0.08
Nebraska	MEI1, PDO1 ⇒ PDSIed	0.88	0.09
(State average)	MEI3, PNA3 ⇒ PDSImd	0.88	0.09

Sample rules generated using the MOWCATL algorithm with a fixed time lag are shown in table V. The advantage of using a fixed time lag with the MOWCATL algorithm is that it provides the rules where the consequent occurs exactly after the specified lag, whereas the maximum lag provides the rules where the consequent occurs within the specified time lag.

Table V. Sample rules generated using MOWCATL algorithm for serial episodes, antecedent and consequent windows of 2-months, and the 3 months fixed time lag between the start of the antecedent and the start of the consequent

Location	Selected serial rules	Confidence	J-measure
Clay Center	MEI1, PDO1 ⇒ PDSIed, SP12sd	0.88	0.08
Hayes Center	MEI1, PDO1 ⇒ PDSIed	0.88	0.09
West Point	MEI1, PDO1 ⇒ SP12sd	0.75	0.08
Nebraska (State average)	MEI1, PDO1 ⇒ SP12sd	0.75	0.08

The confidence and J-values of the rules generated by the MOWCATL were generally better than those of the REAR due to the consideration of time lag. However, since MOWCATL is not designed to identify drought relationships without a time lag, the REAR algorithm handles these cases. This shows that the two algorithms are complementary in generating association rules with or without a time lag.

Generally, the rules for the selected station and state-averaged data of Nebraska showed that most occurrences of drought based on the SPI and PDSI categories are associated with the MEI, PDO, NAO, and PNA with different combinations and confidence factors. In this experiment, the combinations of negative MEI values (La Niña) and negative PDO values implied occurrences of droughts with higher confidence and J-measure values. In other words, the rules indicate that there is a strong relationship between Pacific and Atlantic oceanic conditions and drought in the middle of the United States. These relationships provide important information in drought monitoring.

6. Testing the Results of the Rules on the 2000 Drought

Using the association rules based on 1950–1999 historical data, the recent drought in 2000 was considered to test the results of the relationships between drought and the oceanic parameters. In 2000, the MEI, PDO, and NAO were dominantly negative through out the year. These negative values were the continuation of the La Niña condition in 1999. This situation corresponded with the drier than normal conditions at all of the three selected stations in Nebraska. Although this shows a perfect match with the rules, it should be noted that the technique that is developed does not apply in all cases. In the future, the inclusion of many stations as well as other local and ecological parameters such as the available land cover type and soil moisture may improve the validation of the generated rules to assess and predict drought.

7. Comparison of the Time Series Data Mining with the Statistical Correlation Method

Using data mining techniques, one can identify “local” patterns better than the traditional time-series analysis techniques that largely focus on global models such as statistical correlations (Das, 1998). The infrequent and complex nature of drought requires alternative analysis techniques that emphasize the discovery of local patterns of climate and oceanic data. For example, one may consider the occurrence of drought and its association with climatic and oceanic parameters instead of all precipitation patterns that include wet periods as well. In other words, since drought monitoring is particularly concerned with drought episodes, the data-mining algorithm is needed to discover the associations of drought with oceanic and atmospheric conditions causing drought. This algorithm should identify the drought episodes without the distractions of other “noninteresting” episodes that include normal precipitation and wet episodes within the time series (Harms et al., 2001a).

For a general comparison with the algorithms that we used in this study, one of the traditional statistical correlation methods was used to determine the correlation of the oceanic and climatic parameters. In this traditional statistical method, given a pair of oceanic and climatic parameter values (i.e., X and Y, respectively), the correlation coefficient ($\rho_{X,Y}$) provides an index of the degree to which the paired measures covary in a linear fashion.

Based on this traditional statistical technique, the values of the correlation coefficient are calculated for each climatic and oceanic parameter. Table VI(a) and VI(b) show the correlation of the climatic and oceanic indices for Clay Center, Nebraska, and the statewide average drought indices values of Nebraska. One interesting relationship that can be confirmed with both data mining and traditional methods is the fact that the associations of the climatic drought indices (both SPI and PDSI values) with the SOI, MEI, and PDO are relatively higher than NAO and PNA (table VI(a) and VI(b)).

Table VIa. Linear correlations of drought and atmospheric/oceanic indices for Clay Center, Nebraska

	SPI3	SPI6	SPI9	SPI12	PDSI	SOI	MEI	NAO	PDO	PNA
SPI3	1.00	0.71	0.60	0.54	0.61	–0.15	0.21	–0.04	0.08	0.00
SPI6		1.00	0.82	0.74	0.73	–0.20	0.25	–0.04	0.12	0.05
SPI9			1.00	0.88	0.79	–0.20	0.25	0.01	0.14	0.04
SPI12				1.00	0.79	–0.18	0.23	–0.01	0.15	0.05
PDSI					1.00	–0.27	0.37	0.00	0.31	0.09
SOI						1.00	–0.76	0.05	–0.39	–0.25
MEI							1.00	–0.05	0.58	0.31
NAO								1.00	–0.05	0.01
PDO									1.00	0.38
PNA										1.00

Table VIb. Linear correlations of drought and atmospheric/oceanic indices for Nebraska (state-averaged data)

	SPI3	SPI6	SPI9	SPI12	PDSI	SOI	MEI	NAO	PDO	PNA
SPI3	1.00	0.71	0.60	0.52	0.69	–0.21	0.31	–0.04	0.17	0.03
SPI6		1.00	0.82	0.73	0.78	–0.24	0.37	–0.04	0.23	0.09
SPI9			1.00	0.88	0.81	–0.25	0.37	–0.02	0.27	0.07
SPI12				1.00	0.80	–0.24	0.36	–0.04	0.15	0.09
PDSI					1.00	–0.28	0.40	0.00	0.30	0.08
SOI						1.00	–0.76	0.05	–0.39	–0.26
MEI							1.00	–0.05	0.58	0.31
NAO								1.00	–0.05	0.01
PDO									1.00	0.38
PNA										1.00

However, using the linear correlation method, relationships between the drought indices and oceanic indices are weak (less than 0.40). For example, for Nebraska state average data, the correlation of the PDSI with the SOI was –0.28; with the MEI, 0.40; and with PDO, 0.30. Thus, the population correlation values of the SPI and the PDSI with the atmospheric and oceanic parameters do not show exclusively the relationship with drought.

Thus, there are three main advantages of data mining as compared to the traditional correlation method: (i) instead of a global correlation of the climatic and oceanic data, target episodes such as droughts can be specified separately from normal and wet conditions; (ii) data mining algorithms give flexibility in time series analyses, allowing the discovery of relationships of the parameters using sliding windows, for parameters that occur together within the same time intervals, and with time lags; and (iii) the algorithms allow the analysis of large amounts of data and complicated computations to be executed within a reasonable period of time. This shows that the data mining algorithms that identify the target drought episodes and generate rules are robust tools for monitoring drought.

8. Future Challenges and Prospects

Both REAR and MOWCATL algorithms run on sequential monthly data. They do not address the problems of drought using discrete seasonal data. For example, if we consider the summer seasons of June to September for Clay Center, Nebraska, data that ranges from 1950 to 1999, there are eight-month gaps each year between September and June. This discontinuity in time between the end of the season of one year to the beginning of the season for the following year would not allow the use of sliding windows. Thus, different algorithms should be developed or modified to identify the seasonal relationships between drought episodes and oceanic parameters. Moreover, if algorithms are developed to identify seasonal relationships, they could also be used to identify associations of oceanic parameters with other seasonal time-series data such as crop yields. This will assist in improving the spatial and temporal resolution of drought risk management and crop insurance exposure analysis (Goddard et al., 2003).

The spatial relationships of the rules are also essential in identifying spatial patterns and monitoring drought in certain areas. For example, if one considers drought conditions in Nebraska, the rules that are generated at a few places (stations) are not enough to generalize or make sound decisions. The state-averaged data may provide a feel for general conditions, but they lack details in terms of spatial resolution. Thus, further studies on multiple stations are necessary to understand the results for the state. This suggests that three important improvements in the existing data mining algorithms are needed: (i) automating the rule discovery process to generate rules for all available stations; (ii) developing algorithms for spatial interpolation of the rules to cover areas that do not have observed climate data; and (iii) integrating data mining with other techniques such as GIS and developing visualization methods to make it easy for end users.

In the future, an integrated approach is most likely needed to efficiently identify the influence of all oceanic and atmospheric parameters on drought. Data mining techniques could also help in this respect because well-tailored (i.e., user-oriented) data mining algorithms can be developed to investigate drought problems, and the powerful multiprocessor computers that have high data storage capacity are fast and efficient in running data mining algorithms (IDA Group, 2000). Thus, the complex and nonlinear relationships of the atmosphere and ocean can then be investigated using these two technologies. This can assist in building effective drought decision support systems that include interactive computer-based systems that assist users in management and planning activities.

9. Summary

Data mining algorithms increase efficiency in decision making and allow decision makers to make optimal choices for planning and preparedness. The association rules that are generated at each station or particular place provide information about the associations of these data with a certain degree of confidence. The goal is to improve the quality and accessibility of drought-related data for drought risk management by identifying and monitoring drought characteristics.

In this study, time-series data mining algorithms were used to identify drought episodes and associate these episodes with climatic and oceanic indices. This process integrates temporal knowledge discovery techniques to develop the decision support system using a combination of data mining techniques applied to time series climatic and oceanic data.

In general, there is no specific standard decision making system in drought monitoring. However, in many countries, decisions are taken based on long-range (greater than 10 days) weather outlooks provided by the weather services. In some other cases, the general trend experienced in the sea surface warming of Pacific Ocean (El Niño) conditions is taken into consideration. The rules that are generated using data mining methods complement these existing approaches.

As an experiment, three stations in Nebraska along with state-average data of Nebraska for the period 1950 to 1999 have been used to generate the rules that show the association of drought and oceanic parameters. Rules generated using the REAR and MOWCATL algorithms demonstrate the importance and potential use of data-mining algorithms in monitoring drought using oceanic and atmospheric indices. Because of its flexibility in time, data mining algorithms that include the time lag factor (e.g., MOWCATL) identify a better association of the oceanic and climatic parameters in predicting drought. Since the generated rules indicate the occurrence of drought given certain conditions in the oceanic parameters, the data mining algorithms that identify drought episodes and associate the parameters with a time lag are robust tools in drought monitoring.

The association rules generated using data mining algorithms show that oceanic parameters can be used as a precursor of drought. This allows decisions on appropriate mitigation actions to be made before drought happens, if comprehensive drought preparedness plans are in place. Thus, it can be concluded that data mining could be a useful tool for proactive management of drought and improving the reliability of drought predictions. However, it is important to note that the data mining techniques of identifying drought episodes based on drought's associations with oceanic and climatic parameters are intended to provide additional drought monitoring tools to complement other common techniques already in use.

Acknowledgments – This research was supported in part by the National Science Foundation Digital Government Grant No. EIA-0091530. I would also like to thank Kenneth G. Hubbard and Mark Svoboda for reviewing the manuscript and providing important comments.

References

Berry, J. A. and Linoff, G.: 2000, *Mastering Data Mining: The Art and Science of Customer Relationship Management*, John Wiley & Sons, New York, 494 pp.

Bigus, J. P.: 1998, *Data Mining with Neural Networks: Solving Business Problems from Application Development to Decision Support*, McGraw-Hill, New York, 220 pp.

Cabena, P. H., Stadler, R., Verhees, J., and Zanasi, A.: 1998, *Discovering Data Mining: From Concept to Implementation*, IBM, New Jersey, 195 pp.

Das, G., Lin, K. I., and Mannila, H.: 1998, Rule discovery from time series, in *Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining*, New York, pp. 16–22.

Dewey, K.: 1996, Summer: The Season of the Sun, *NEBRASKAland Magazine's Weather and Climate of Nebraska* 74(1), 57–62.

Diaz, H. F. and Markgraf, V.: 1992, *El Niño: Historical and Paleoclimatic Aspects of the Southern Oscillation*, Cambridge University Press, Cambridge, 476 pp.

Edelstein, H.: 1997, Data mining: Exploiting hidden trends in your data, *DB2magazine*, Spring 2(1). Available online at: <http://www.db2mag.com/db_area/archives/1997/q1/9701edel.shtml> [accessed in December 2002].

FEMA (Federal Emergency Management agency): 1995, National mitigation strategy: Partnerships for building safer communities, Washington, DC.

Francis, R. C., and Hare, S. R.: 1994, Decadal-scale regime shifts in the large marine ecosystems of the Northeast Pacific: A case for historical science, *Fish. Oceanogr.* 3, 279–291.

Glantz, M.: 1994, Usable science: Food security, early warning, and El Niño, in *Proceedings of the Workshop on ENSO/FEWS*, Budapest, Hungary, pp. 3–11.

Goddard, S., Harms, S. K., Reichenbach, S. E., Tadesse, T., and Waltman, W. J.: 2003, Geospatial decision support for drought risk management, *Communication of the ACM* 46(1), 35–37.

Groth, R.: 1998, *Data Mining: A Hands-On Approach for Business Professionals*, Prentice Hall, New Jersey, 264 pp.

Halpert, M. S., and Ropelewski, C. F.: 1992, Surface temperature patterns associated with Southern Oscillation, *J. Climatol.* 5, 577–593.

Harms, S. K.: 2003, Temporal association rule methodologies for geo-spatial decision support, Ph.D. dissertation, University of Missouri, Columbia.

Harms, S. K., Goddard, S., Reichenbach, S. E., Waltman, W. J., and Tadesse, T.: 2001a, Data mining in a geo-spatial decision support system for drought risk management, in *Proceedings of the 2001 National Conference on Digital Government Research*, Los Angeles, CA, pp. 9–16.

Harms, S. K., Deogun, J., Saquer, J., and Tadesse, T.: 2001b, Discovering representative episodal association rules from event sequences using frequent closed episode sets and event constraints, in *Proceedings of the 2001 IEEE International Conference on Data Mining*, San Jose, CA, pp. 603–606.

Harms, S. K., Deogun, J., and Tadesse, T.: 2002, Discovering sequential association rules with constraints and time lags in multiple sequences, in *Proceedings of the 2002 International Symposium on Methodologies for Intelligent Systems*, Lyon, France, pp. 432–441.

Hayes, M.: 2003, Drought indices. Available online at: <http://www.drought.unl.edu/whatis/indices.htm> [accessed on 15 February 2003].

Hayes, M. J., Wilhite, D. A., Svoboda, M., Vanyarkho, O.: 1999, Monitoring the 1996 drought using Standardized Precipitation Index, *Bulletin of the American Meteorological Society* 80(3), 429–438.

Huang, Y., and Yu, P. S.: 1999, Adaptive query processing for time-series data, in *Proceeding of the 5th International Conference on Knowledge Discovery and Data Mining, ACM*, pp. 282–286.

Hurrell, J. W.: 1995, Decadal trends in the North Atlantic Oscillation: regional temperatures and precipitation, *Science* 269, 676–679.

Intelligent Data Analysis (IDA) Research Group: 2000, *Intelligent Data Analysis*. Available online at: <http://web.dcs.bbk.ac.uk/ hui/IDA/home.html> [accessed on 15 February 2003].

Keogh, E. J., and Pazzani, M. J.: 1998, An enhanced representation of time-series which allows fast and accurate classification, clustering and relevance feedback, in *AAAI-98 Workshop on Predicting the Future: AI Approaches to Time-Series Analysis*, pp. 44–51.

Klemettine, M.: 1999, A knowledge discovery methodology for telecommunication network alarm databases, Ph.D. dissertation, University of Helsinki, Finland.

Kryszkiewicz, M.: 1998, Fast discovery of representative association rules, in *Proceedings of the Rough Sets and Current Trends in Computing (RSCTC)*, Warsaw, Poland, pp. 214–221.

Mannila, H., and Seppänen, J.: 2001, Recognizing similar situations from event sequences, *First SIAM Conference on Data Mining*, available online at: <http://www.siam.org/meetings/sdm01/pdf/sdm01_03.pdf> [accessed on 15 February 2003].

Mannila, H., Toivonen, H., and Verkamo, A. I.: 1997, Discovering frequent episodes in sequences, Technical Report, Department of Computer Science, University of Helsinki, Finland, Report C-1997-15.

McPhaden, M. J., Busalacchi, A. J., Cheney, R., Donguy, J. R., Gage, K. S., Halpern, D., Ji, M., Julian, P., Meyers, G., Mitchum, G. T., Niiler, P. P., Picaut, J., Reynolds, R. W., Smith, N., and Takeuchi, K.: 1998, The tropical ocean-global atmosphere observing system: A decade of progress, *J. Geophys. Res.* 103(C7), 14,169–14,240. Available online at: <http://www.pmel.noaa.gov/pubs/outstand/mcph1720/ab stract.shtml> [accessed in December 2002].

NCDC (National Climatic Data Center): 2002, Billion dollar US weather related disasters 1980–2002. Available online at: <http://lwf.ncdc.noaa.gov/oa/pub/data/special/billionz-2002.pdf> [accessed on February 2003].

Ogallo, L. A.: 1994, Validity of the ENSO-related impacts in Eastern and Southern Africa, in *Proceedings of the Workshop on ENSO/FEWS*, Budapest, Hungary, pp. 179–184.

O'Meagher, B., Stafford, M., and White, D. H.: 2000, Approaches to integrated drought risk management: Australia's national drought policy. In D. A. Wilhite (ed.), *Drought: A Global Assessment*, Natural Hazard and Disasters Series, Routledge Publishers, UK, pp. 115–128.

Philander, S. G. H.: 1990, *El Niño, La Niña, and the Southern Oscillation*, Academic Press, San Diego, CA, 289 pp.

Povinelli, R. J.: 2000, Using genetic algorithms to find temporal patterns indicative of time-series events, in *GECCO 2000 Workshop: Data Mining with Evolutionary Algorithms*, pp. 80–84.

Ropelewski, C. F., and Halpert, M. S.: 1986, North American precipitation and temperature patterns associated with the El Niño/Southern Oscillation (ENSO), *Monthly Weather Review* 114, 2352–2362.

Ropelewski, C. F., and Halpert, M. S.: 1987, Global and regional scale precipitation patterns associated with the El Niño–Southern Oscillation, *Monthly Weather Review* 115, 1606–1626.

Rosenberg, N. J.: 1978, *North American Droughts*, Westview Press, Boulder, CO, 177 pp.

Ross, T., and Lott, N.: 2000, A climatology of recent extreme weather and climatic events, US Dept. of Commerce, NOAA/NESDIS, National Climatic Data Center (NCDC), Technical Report 2000-02, 17 pp. Available online at: <http://lwf.ncdc.noaa.gov/oa/reports/billionz.html> [accessed in December 2002].

Smyth, P., and Goodman, R. M.: 1992, An information theoretic approach to rule induction from databases, *IEEE Transactions on Knowledge and Data Engineering* 4(4): 301–316.

Svoboda, M., LeComte, D., Hayes, M., Heim, R., Gleason, K., Angel, J., Rippey, B., Thinker, R., Palecki, M., Stooksbury, D., Miskus, D., and Stephens, S.: 2002, The drought monitor, *Bull. Amer. Meteorol. Soc.* 83(8), 1181–1190.

Tadesse, T.: 2002, Identifying drought and its associations with climatic and oceanic parameters using data mining techniques. Ph.D. dissertation, University of Nebraska, Lincoln.

Thearling, K.: 2001, *An Introduction to Data Mining*. Available online at: <http://www.thearling.com/text/dmwhite/dmwhite.htm> [accessed in December 2002].

Trenberth, K. E., and Hoar, T. J.: 1996, The 1990–1995 El Niño–Southern oscillation event longest on record, *G* 23, 57–60.

Two Crows Corporation: 1999, *Introduction to Data Mining and Knowledge Discovery*, 3rd edn., Two Crows Corporation, Postmac, MD. Available online at: <http://www.twocrows.com/introdm.pdf> [accessed in December 2002].

Wallace, J. M., and Gutzler, D. S.: 1981, Teleconnections in the geopotential height field during the Northern Hemisphere Winter, *Monthly Weather Review* 109, 784–812.

Wallace, J. M., and Vogel S.: 1994, El Niño and Climate Prediction, The University Corporation for Atmospheric Research. Available online at: <http://www.pmel.noaa.gov/toga-tao/el-ninoreport.html#part1> [23 January 2003].

Wilhite, D. A.: 1993, *Drought Assessment, Management, and Planning: Theory and Case Studies*. Kluwer Academic Publishers, Boston, 293 pp.

Wilhite, D. A.: 2000a, Drought as a natural hazard: Concepts and definitions. In D. A. Wilhite (ed.), Drought: A global assessment, *Natural Hazard and Disasters Series*, Routledge Publishers, UK, pp. 3–18.

Wilhite, D. A.: 2000b, Preparing for drought: A methodology. in D. A. Wilhite (ed.), Drought: A global assessment, *Natural Hazard and Disasters Series*, Routledge Publishers, UK, pp. 89–104.

Wolter, K., and Timlin, M. S.: 1993, Monitoring ENSO in COADS with a seasonally adjusted principal component index, in *Proceedings Seventh Annual Climate Diagnostic Workshop*, Norman, Oklahoma, March 1993, pp. 52–57.