Agricultural and Forest Meteorology 237–238 (2017) 105–122

Contents lists available at ScienceDirect

Agricultural and Forest Meteorology

journal homepage: www.elsevier.com/locate/agrformet

Meteorological drought forecasting for ungauged areas based on machine learning: Using long-range climate forecast and remote sensing data

Jinyoung Rhee[a], Jungho Im[b,*]

[a] Climate Research Department, APEC Climate Center, Busan, Republic of Korea
[b] School of Urban and Environmental Engineering, Ulsan National Institute of Science and Technology (UNIST), Ulsan, Republic of Korea

ARTICLE INFO

Article history:
Received 28 September 2016
Received in revised form 31 January 2017
Accepted 6 February 2017
Available online 10 February 2017

Keywords:
Drought forecasting
Machine learning
Climate forecast data
Remote sensing
Spatial interpolation

ABSTRACT

A high-resolution drought forecast model for ungauged areas was developed in this study. The Standardized Precipitation Index (SPI) and Standardized Precipitation Evapotranspiration Index (SPEI) with 3-, 6-, 9-, and 12-month time scales were forecasted with 1–6-month lead times at $0.05 \times 0.05^{\circ}$ resolution. The use of long-range climate forecast data was compared to the use of climatological data for periods with no observation data. Machine learning models utilizing drought-related variables based on remote sensing data were compared to the spatial interpolation of Kriging. Two performance measures were used; one is producer's drought accuracy, defined as the number of correctly classified samples in extreme, severe, and moderate drought classes over the total number of samples in those classes, and the other is user's drought accuracy, defined as the number of correctly classified samples in drought classes over the total number of samples classified to those classes. One of the machine learning models, extremely randomized trees, performed the best in most cases in terms of producer's accuracy reaching up to 64%, while spatial interpolation performed better in terms of user's accuracy up to 44%. The contribution of long-range climate forecast data was not significant under the conditions used in this study, but further improvement is expected if forecast skill is improved or a more sophisticated downscaling method is used. Simulated decreases of forecast error in precipitation and mean temperature were tested: the simulated decrease of forecast error in precipitation improves drought forecast while the decrease of forecast error in mean temperature does not contribute much. Although there is still some room for improvement, the developed model can be used for drought-related decision making in ungauged areas.

1. Introduction

Droughts have caused significant losses and damages. Defining the types of drought helps to monitor droughts and to build strategies for the preparation and response to droughts (Ward et al., 2006; Harou et al., 2010). Wilhite and Glantz (1985) addressed the conceptual definition of drought as well as the operational definition of drought. The operational definition includes the onset, severity, termination, and frequency of drought. They also defined meteorological, agricultural, hydrological, and socio-economic droughts based on their review of more than 150 studies. More recently, Mishra and Singh (2010) reviewed the definitions of previous studies. In this study, we adopted the definition of the Intergovernmental Panel on Climate Change Special Report (IPCC SREX, 2012): "*a period of abnormally dry weather long enough to cause a serious hydrological imbalance*."

Real-time or near real-time drought monitoring is invaluable for timely response to drought events. Drought early warning systems help to perform drought assessment and monitoring, and to make appropriate decisions in a timely manner by providing drought information, and to ultimately reduce drought damages. Drought forecasts, if feasible, can provide drought information in advance, enabling the reduction of drought impacts by securing appropriate resources and planning effective allocation of them.

There are not many existing drought forecasting systems, but a variety of methods have been studied to provide drought forecast data to be used for drought-related decision making. Mishra and Singh (2011) listed the components of drought

* Corresponding author.
E-mail address: ersgis@unist.ac.kr (J. Im).

http://dx.doi.org/10.1016/j.agrformet.2017.02.011
0168-1923/

forecasting as hydro-meteorological variables, drought indices, large-scale climate indices, methodologies, and output. The hydro-meteorological variables used for drought forecasting can be determined according to the type of drought of interest; precipitation is the most important variable for meteorological drought, while soil moisture and reservoir level are essential for agricultural and hydrological droughts, respectively. Drought indices such as the Standardized Precipitation Index (SPI; McKee et al., 1993) and the Palmer Drought Severity Index (PDSI; Palmer, 1965), or the combination of any related variables, are used as indicators for drought conditions. Large-scale climate variabilities such as the El Nino-Southern Oscillation (ENSO) and Arctic Oscillation (AO) may be used for long-range drought forecast. Methodologies of drought forecasting include regression models (e.g., Leilah and Al-Khateeb, 2005), time series models (e.g., Han et al., 2010), probability models (e.g., Cancelliere et al., 2007), artificial neural network models (e.g., Morid et al., 2007), and hybrid models (e.g., Mishra et al., 2007), and each method produces outputs to quantify the drought condition determining the initiation and termination of drought, the nature of severity, and the probability of occurrence, among other values (Mishra and Singh, 2011).

Long-range climate forecast data for variables such as precipitation, air temperature, and relative humidity with lead times up to 6 months can be used for drought forecasting. The seasonal drought outlook of the National Oceanic and Atmospheric Administration (NOAA) of the US has already used 3-month long-range climate forecast data as well as the 1-month precipitation and temperature forecast of the Climate Prediction Center (CPC), short-term climate forecast data of the Weather Prediction Center, soil moisture model results, the probability of the termination and reduction of PDSI, climatology, and initial conditions to produce seasonal drought outlook results (CPC, 2016). Drought forecast data based on long-range climate forecasts are reprocessed sometimes to be easily understood by end users; Steinemann (2006) developed the Forecast Precipitation Index (FPI) so that decision makers can easily use drought forecast results after conducting a questionnaire survey on water resources managers to bridge the gap between climate science and social decisions.

Each method introduced above has its own advantages to predict drought-related variables. In order to forecast drought indices that can be calculated based on hydro-meteorological variables from climate model outputs, the use of long-range climate forecast is a good choice since it is not necessary to use any proxy variable or to depend on any indirect method. Mo and Lyon (2015) recently performed meteorological drought prediction based on SPI using the North American Multi-Model Ensemble (NMME) and the Global Precipitation Climatology Center (GPCC) precipitation data at $1 \times 1^{\circ}$ (about 100×100 km near the equator). This resolution is too coarse though to provide detailed drought information for decision makers. Weather station data can replace the GPCC precipitation. In this case, there should be a model to provide information for ungauged areas between the locations of weather stations.

In this study, a high resolution meteorological drought forecast model was developed to provide drought forecast information based on SPI and Standardized Precipitation-Evapotranspiration Index (SPEI) for ungauged areas. The high resolution ($0.05 \times 0.05^{\circ}$) of the developed model enables the provision of detailed data for regional and local uses and ultimately the reduction of drought impact. The objectives of this study are to (1) develop a drought forecast model based on the combination of remote sensing and long-range forecast data using machine learning for ungauged areas, and (2) provide improved ranges of drought forecast in case of the improvement of forecasting skill of the long-range forecast data.

2. Study area and data

2.1. Study area

The study area is South Korea, located in northeast Asia (Fig. 1). The area of South Korea is about 100,284 km^2. About 19% of the area is composed of rice paddies and other crop lands, and about 64% consists of forests (KOSIS, 2016a). The annual average temperature of the five main cities, Seoul, Incheon, Gangneung, Mokpo, and Busan ranges from 11.4–14.1 °C, and average annual precipitation ranges from 1108–1445 mm (KOSIS, 2016b). The temporal scope of this study is from January 2003 to August 2015 and one to six months of lead time were used.

The methodology of this study may be applied to any region with appropriate input variables related to drought. South Korea is located in the mid-latitude region with partial influence of large-scale atmosphere-ocean interactions such as ENSO, and the forecast skill of the long-range climate forecast data is also known to be lower compared to tropical regions (e.g., Wang et al., 2001; Wang et al., 2004). The study area was thus selected to examine the performance of machine learning models under the limited skill of the long-range forecast data. The study area will be expanded to areas with distinct influence of large-scale atmosphere-ocean interactions such as Southeast Asia in the following studies.

There have been serious historical droughts in South Korea. The events in 1939, 1968, 1978, and 1982 were considered as the most extreme drought events before the 1990s (Sim, 2009). After the 1990s, the area experienced a nation-wide drought during 1994–1995 (Park and Schubert, 1997). The 2001 and 2008–2009 drought events were also recorded as devastating ones especially in Kangwon Province located in the northeast of South Korea where more than 50,000 residents experienced drinking water shortage (Sim, 2009). More recently, a short-term spring drought occurred in 2012 resulting in less than 30% of normal precipitation, and a long-term drought initiated in 2013 continued to 2015, especially in the northwest part of South Korea.

2.2. Data

2.2.1. Automatic synoptic observation system (ASOS) data

Drought conditions in this study were measured using drought indices, and the reference drought index values were calculated from the ASOS data. Monthly precipitation and temperature data from 61 ASOS stations in the study area (Fig. 1) were obtained from the Korea Meteorological Administration for the period of January 1973–August 2015. Potential evapotranspiration data were obtained from mean air temperature using the Thornthwaite method (Thornthwaite, 1948). Despite that this method could be misleading during winter seasons in the study area because it assumes zero potential evapotranspiration for below zero degree Celsius temperatures, the physically-based methods such as Penman-Monteith or Hargreaves could not be used because the minimum data requirements are the minimum and maximum temperature while long-range climate forecast only provides mean temperature.

2.2.2. Drought indicators

Many drought indices have been developed since the mid-20th century. The most widely used drought indices include the PDSI developed in the 1960s and the SPI developed in the 1990s. The Standardized Precipitation Evapotranspiration Index (SPEI), which was developed by Vicente-Serrano et al. (2010) considering not only the precipitation used for SPI but also atmospheric moisture demand, has also been used in many studies. Usually drought index values are calculated at weather stations locations from time series