Nat Hazards (2013) 65:1275–1284
DOI 10.1007/s11069-012-0406-z

ORIGINAL PAPER

Dynamic risk prediction based on discriminant analysis for maize drought disaster

Qi Zhang · Jiquan Zhang · Denghua Yan · Yulong Bao

Received: 24 August 2012 / Accepted: 12 September 2012 / Published online: 26 September 2012
© Springer Science+Business Media Dordrecht 2012

Abstract This study presents a discriminant analysis-based method for prediction of agriculture drought disaster risk. We selected the Chaoyang city in the Northeast China as the study area. We employed multi-scale standard precipitation index (SPI) to reflect drought hazard. We used the yield losses to indicate the drought disaster risk, which was divided into no, low, or high drought risk. We used the multi-scale SPI and drought disaster risk as the input factors for the discriminant analysis-based risk prediction model. The results showed that the model's prediction accuracy varied between 40 and 82.4 %. The accuracy of high drought disaster risk category was higher than low and no drought disaster risk category. The prediction accuracy of the milky maturity stage was highest. We use leave-one-out cross-validation method to validate the model's accuracy. And the results showed that the model validation accuracy of high drought group could reach 70.6 % in milky maturity stage. This study showed discriminant analysis is an effective and operable method for disaster risk prediction. This model can provide timely information for decision makers to make effective measures for drought disaster management and to reduce the drought effects to yields at the minimum level.

Keywords Dynamic risk prediction · Discriminant analysis · Drought disaster · Multi-scale SPI · Yield

1 Introduction

Drought is one of the most common natural disasters around the world. It receives more attention because of its unique characteristics, that is, the high frequency of occurrence, global impacts, and the great losses. Drought is one of the most significant stress factors in

Q. Zhang · J. Zhang (✉) · Y. Bao
College of Urban and Environmental Sciences, Northeast Normal University, Changchun 130024, China
e-mail: zhangjq022@nenu.edu.cn

D. Yan
China Institute of Water Resources and Hydropower Research, Beijing 100038, China

crop production. It can lead to obvious yield reduction or complete crop failure (Hlavinka et al. 2009). China is a major agriculture country with large population and easily affected by agro-meteorological disasters especially drought and flood. Wheat, maize, and rice are the three main crops in China. Liaoning province is the main maize-growing region in China. Drought is the dominant disaster for maize production especially in the north-western province of Liaoning. Drought disasters have influenced the food security and sustainable development of the region. Consequently, it is important to analyze the relationship between drought and maize yield and to predict drought disaster risk in different maize-growing stage. This is also the basis for early warning of drought disasters and helping make effective measures to reduce the drought losses to the minimum.

Many reported studies used indices to describe droughts, such as standardized precipitation index (SPI) (McKee et al. 1993), palmer drought severity index (PDSI) (Palmer 1965), effective drought index (EDI) (Byun et al. 1999), etc. Those indices performed differently because of the purpose of research and region of study. PDSI is the most commonly used index. Because of its complex, empirical derivation and because the underlying computation is based on the climate of the Midwestern United States, many researchers have reported the low practicability of PDSI in other areas (Keyantash and Dracup 2002; Kim et al. 2009). SPI is an index that precipitation is the only input data. SPI can be calculated in different time scales and is sensitive to drought. It is widely used in China. Huang et al. (2010) used SPI to reflect the spatial and temporal variation of seasonal drought in South China during the past 58a. Logan et al. (2010) used multi-scale SPI to assess spatiotemporal variability of drought over the Kansas River Basin region. The results showed that many areas of increasing wetness throughout the region but only isolated regions were drying.

Droughts can be predominantly distinguished into four types of meteorological, agricultural, hydrological, and socio-economic according to the impacts and time scales (Heim 2002). In this study, we mostly focused on agricultural drought. Only when drought caused agriculture losses, it could be called agriculture drought disaster. There are numerous studies about agriculture drought disaster. The most common method is constructing correlative relationship between drought index and corresponded yield losses to assess agriculture drought disaster. Quiring and Papakryiakou (2003) compared four drought indices and found Palmer's Z-index was the most appropriate index for monitoring agricultural drought and predicting red spring wheat yield in Canada. Trnka et al. (2007) analyzed the relationship between detent yields of spring barley and droughts presented by Palmer Z-index, the results found the growing season water balance significantly influenced the spring barley production. Hlavinka et al. (2009) used sum of Z-index to quantify agricultural drought and concluded drought resulted in significant yield losses of key crops in the Czech Republic. They used the rainfall during the whole growing season. The results could only reflect that whether drought was the main restrictive factor for agriculture production in the certain area. If the fact has been known that drought was the main restrictive factor, how to predict the final yield losses by using timely growth period observation date? Remote sensing data-based methods were used to monitoring drought. Combined with ground data, historical agriculture yield data, it can be used to predict final yield losses timely. Unganai and Kogan (1998) used the advanced very high-resolution radiometer (AVHRR)–based vegetation condition index (VCI) and temperature condition index (TCI) to monitoring vegetation health and productivity. Domenikiotis et al. (2004) used VCI to monitoring cotton agro-meteorological conditions and assess cotton yield before the end of growing season. Vicente-Serrano et al. (2006) used both VCI and SPI to predict crop yield at growing season. The results showed that SPI performed much better than VCI. However, satellite data are largely affected by various sources of error such as satellite changes, satellite orbital drift, sensor degradation, atmospheric perturbation due to aerosols and clouds, and surface nonuniformity change in equator crossing time. The quality of satellite date needs improved when operational application.

The main objective of this study was to find a method to predict potential agriculture losses caused by drought at each growing stage. The potential losses caused by drought were called as drought disaster risk. This was a pre-disaster action. This could guide farmers and decision makers to take emergency and effective actions to relief drought impacts and reduce agriculture losses at the minimum level. Discriminant analysis was introduced to predict potential agriculture losses before the end of growing season. Discriminant analysis is originally developed by Fisher in 1936. It has been mostly used by researchers to predict the group or category to which a subject belongs. It has been successfully used in computer vision and pattern recognition, disease diagnosis, and facial recognition. In recent years, some studies used this method to discriminate different vegetation types in remote sensing image (Guo et al. 2003; Thessler et al. 2008). Thessler et al. (2008) used discriminant analysis and *k* nearest neighbors method to classify tropical lowland forests types from Landsat TM image. The results showed that this method could distinguish floristically and structurally different of rain forest types accurately. Lin et al. (2011) used elevation, slope, aspect, and soil depth data to classify land use type by using discriminant analysis method, and the accuracy can reach 71.0 %. In this paper, we used discriminant analysis to build a predictable relationship between growing stage precipitation and final crop yield losses based on historical data. Then, we could predict potential crop yield losses by using growing stage precipitation.

2 Materials and methods

2.1 Study area

Chaoyang city is situated in northwestern Liaoning province between 118°50′–121°17′E and 40°25′–42°22′N. Annual mean temperature is 5.4–8.7 °C. Annual mean sunshine duration is 2,850–2,950 h. Annual precipitation is 450–580 mm. Maize is the dominant crop in agriculture production. Its growing season begins with planting in May and continues through September. The climate of Chaoyang is favorable for maize growing. However, frequent agro-meteorological disasters, especially droughts, are the limitation factors for maize growing. Since the region is affected by the monsoon climate, the distribution of rainfall is uneven between years and months to this. Maize is susceptible to drought (Fig. 1).

2.2 Data collection

The data consisted of ten-day precipitation data and maize yield recorded during 1970–2009 at Chaoyang city. The precipitation data were collected from China Meteorological Data Sharing Service System, and the yield data collected from Social and Economic Statistical Yearbook of Liaoning province. The precipitation data were used to calculate SPI for indentifying drought hazard. The most commonly used time scale of the SPI is month. However, monthly scale is too long for maize growth. We used 10-day time scale instead. The yield losses data were used to identifying drought disaster.

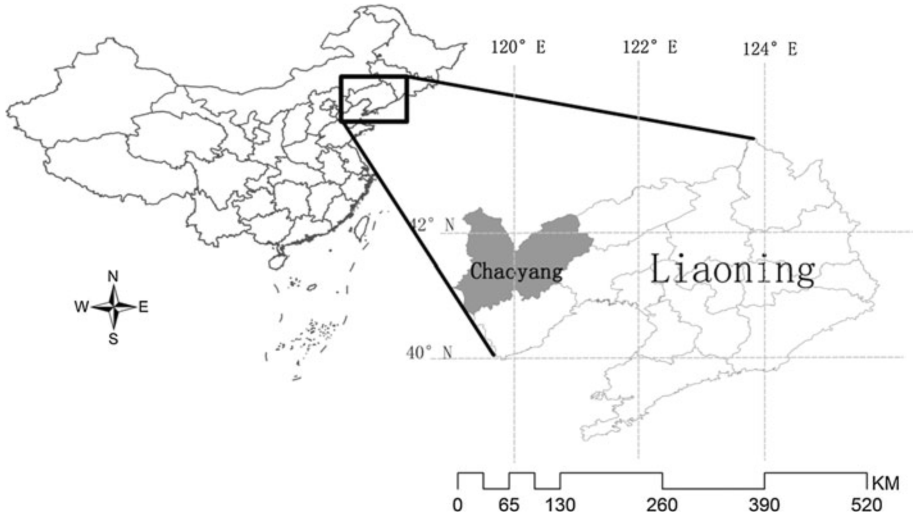

Fig. 1 Study area

2.3 Data analysis

2.3.1 Drought hazard index

Over 40 countries are using SPI for drought monitoring and research. It can be used to monitor drought and flood over a wide spectrum of time scale, allowing users the opportunity to choose the time scale most appropriate for their study (Wu and Wilhite 2004). The SPI at a high-frequency time scales reflects precipitation supply more precisely than that with a lower frequency. In this study, multi-scale SPI was used to describe agricultural drought hazard. The time scales of the SPI include 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, and 110 days, represent as SPI_{10d}, SPI_{20d}, …, SPI_{110d}, respectively. We selected five critical maize-growing stages for drought risk prediction including pre-planting, jointing, tasseling, milky maturity, and maturity. Stage 1 denoted pre-planting as the last 10 days of April. Stage 2 denoted jointing as the last 10 days of June. Stage 3 denoted tasseling as the middle 10 days of July. Stage 4 denoted milky maturity as the last 10 days of August. Stage 5 denoted maturity as the middle 10 days of September. As such, the rainfalls during the last 10 days of April were used to calculate SPI_{10d} of stage 1. The SPI_{20d} of stage 1 calculated by the rainfall from 11 April to 30 April, etc.

2.3.2 Drought disaster index

We define drought disaster as the losses caused by drought. Crop yield losses were the direct effects of agro-meteorological hazards. In this study, we used yield losses as the index of drought disaster. However, yield losses could be resulted from many agro-meteorological hazards which happened during growing season, such as drought, flood, hail disaster, and chilling damage. It is difficult to divide how many yield losses were caused by drought instead of other agro-meteorological hazards. Drought is the dominant disaster for maize production in Chaoyang and other agro-meteorological hazards present little impact on maize production. Figure 2 illustrated that the inter-annual fluctuation has an obvious consistency between maize yield and precipitation in maize-growing season, the correlation coefficient could reach 0.497 at 0.01 significant levels. Therefore, it is reasonable to reflect drought disaster by using yield losses.

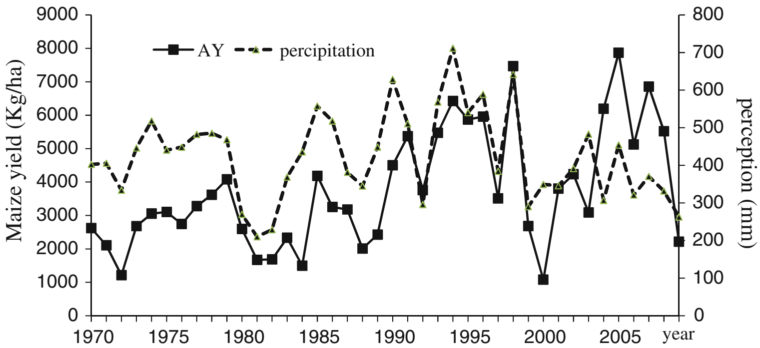

Fig. 2 Precipitation and maize yield in Chaoyang City during 1970–2009. AY is the annual actual maize yield. Precipitation is rainfall from April to September in each year

The actual yield of crop per unit area (AY) is mainly influenced by weather, cultivation technology and management, agricultural policies, and innovations. While weather is the most "random" factor, the others have relatively continuous trend. So, the AY can be divided into two parts according to the factors that influence the crop yield, that is, the trend yield (TY), which is mainly influenced by continuous factors such as cultivation technology and management, agricultural policies, innovations, and climate yield (CY), which is mainly influenced by random factor such as weather (Zhang 2004). The trend yield can be calculated by linear moving average method, it is a commonly used method for fitting crop trend yield. Correspondingly, the climate yield can be expressed by:

$$CY_{\mathrm{i}} = AY_{\mathrm{i}} - TY_{\mathrm{i}}, \tag{1}$$

where, i is the year. A positive value of CY denotes that the climate is favorable for maize production. On the contrary, a negative CY value denotes climate condition is unfavorable. For Chaoyang city, most of the negative values were caused by drought. The rate of yield losses (R) was used as a factor of drought disaster, R can be expressed by:

$$R_{\mathrm{i}} = CY_{\mathrm{i}}/AY_{\mathrm{i}} = (AY_{\mathrm{i}} - TY_{\mathrm{i}})/AY_{\mathrm{i}}, \tag{2}$$

According to actual local conditions and disaster statistical habit, classification of drought disaster risk based on R was as flows:

- No drought disaster risk: $R \geq 0.1$;
- Low drought disaster risk: $0.1 > R \geq -0.1$;
- High drought disaster risk: $R < -0.1$.

2.3.3 Model building

There are two steps for building the discriminant analysis-based risk prediction model. Firstly, we need a large number of training samples to help understand the relationship between a dependent variable and one or more independent variables. SPIs are the independent variables, and drought disaster risks are dependent variables. We used the data from 1970 to 2009, so there were 40 training samples. Then, discriminant functions and territorial map were built to reflect the relationship between independent and dependent variables. The basic form of the function is as follows:

$$D = \beta_1(X_1) + \beta_2(X_2) + \beta_3(X_3) + \cdots + C \tag{3}$$

where D is discriminant function score, β is the function coefficient, C is intercept. X is the value of independent variable. The territorial map is divided into several sections, each section represents a dependent variable category. Usually, there are two discriminant functions. The function scores are used as the coordinates on the territorial map. Second, we input the independent variables and calculate the function scores, then to see the position on territory map to predict this case's category. Those two steps can be calculated in statistical package for the social sciences (SPSS).

3 Results and discussions

3.1 Model in each growing stage

Each growing stage has its risk prediction model. The following two functions were for Pre-planting stage. The corresponding territorial map was illustrated in Fig. 3. The values of SPI_{10d}–SPI_{110d} were used to calculate D_1 and D_2. They were used to find the position on the territorial map to define its category. Take 2001a as an example, we putted the values of SPIs into functions 4 and 5 to calculate D1 and D2, which were 0.48 and −0.09 respectively. We could found (0.48, –0.09) located in high drought risk area (Fig. 3). The function coefficients for other stages were showed in Table 1.

```
Discriminant
function 2 (D2)
     -3.0      -2.0      -1.0        .0       1.0       2.0       3.0
       +---------+---------+---------+---------+---------+--------+
  2.0 +         +        12 +        +         +         +         +
      I                  12                                        I
      I                   12                                       I
      I                    12                                      I
      I                     12                                     I
      I                      12                                    I
  1.0 +         +         +  12      +         +         +         +
      I                       12                                   I
      I                        12      *                           I
      I                         12                                 I
      I                          12                    222222222222I
      I                           1222222222222222222223333333333333I
   .0 +         +         + *       1333333333333333333  +         +
      I                           13                               I
      I                           13     *                         I
      I                           13                               I
      I                          13                                I
      I                          13                                I
 -1.0 +         +         +      13  +         +         +         +
      I                         13                                 I     Symbols  Label
      I                        13                                  I     ----    --------------
      I                        13                                  I
      I                      13                                    I
      I                      13                                    I       1      No drought
 -2.0 +         +         +  13      +         +         +         +       2      Low drought
       +---------+---------+---------+---------+---------+--------+        3      High drought
     -3.0      -2.0      -1.0        .0       1.0       2.0       3.0      *      Indicates a group centroid
                                   Discriminant function 1 (D1)
```

Fig. 3 Territorial map for pre-planting stage

Table 1 Discriminant function coefficients

	Jointing		Tasseling		Milky maturity		Maturity	
	D_1	D_2	D_1	D_2	D_1	D_2	D_1	D_2
SPI_{10d}	−0.194	0.691	−0.052	−0.024	0.173	0.684	0.073	0.071
SPI_{20d}	0.825	−0.491	0.600	−0.290	−0.012	−0.174	0.294	1.192
SPI_{30d}	−3.237	0.630	−1.163	−0.441	−0.574	−0.917	0.235	−1.683
SPI_{40d}	0	0	2.335	0.576	−0.023	0.527	0.245	0.837
SPI_{50d}	0.517	1.488	−4.587	−1.972	0.324	0.446	−0.720	0.579
SPI_{60d}	3.501	−1.194	−0.048	5.285	0.408	−0.334	0.329	−0.475
SPI_{70d}	−0.350	1.452	2.287	0.052	−1.586	−0.314	−0.338	−0.156
SPI_{80d}	−3.021	−3.742	1.138	−3.465	2.625	−0.202	0.678	−0.499
SPI_{90d}	2.807	2.125	0.088	3.707	−5.271	1.178	−1.471	0.649
SPI_{100d}	2.802	1.025	0.164	−2.581	2.333	3.713	2.655	1.611
SPI_{110d}	−3.581	−0.945	−0.164	−0.571	2.164	−4.284	−0.783	−1.853

Table 2 Classification accuracies for pre-planting stage

Actual risk level	Predicted risk by discriminant analysis				Classification accuracy (%)
	No drought	Low drought	High drought	Total	
No drought	7	4	2	13	53.8
Low drought	2	4	4	10	40
High drought	2	7	8	17	47.1

$$\begin{aligned} D_1 = {} & 0.036 \times SPI_{10d} + 1.483 \times SPI_{20d} - 1.262 \times SPI_{30d} + 1.373 \times SPI_{40d} \\ & + 2.995 \times SPI_{50d} - 4.561 \times SPI_{60d} + 0.849 \times SPI_{70d} - 5.096 \times SPI_{80d} \\ & + 4.733 \times SPI_{90d} - 4.7 \times SPI_{10d} + 4.274 \times SPI_{110d} \end{aligned} \tag{4}$$

$$\begin{aligned} D_2 = {} & 1.148 \times SPI_{10d} - 0.25 \times SPI_{20d} - 0.441 \times SPI_{30d} + 0.052 \times SPI_{40d} \\ & - 0.248 \times SPI_{50d} - 4.18 \times SPI_{60d} + 3.779 \times SPI_{70d} - 6.121 \times SPI_{80d} \\ & + 5.065 \times SPI_{90d} - 11.836 \times SPI_{100d} + 13.363 \times SPI_{110d} \end{aligned} \tag{5}$$

3.2 Classification accuracies of drought disaster

From Table 2, we could see that there were 13 years with no drought risk from 1970 to 2009. Among those 13 years, 7 years were divided into no drought risk group by the prediction model. The accuracy for no drought risk was 53.8 %. The classification accuracies of the three drought risk levels varied between 40 and 53.8 % in pre-planting stage. The accuracy rate was low in this stage. Table 3 showed the accuracies in other stages. From pre-planting to milky maturity, the average accuracy of those three classes increased gradually from 47.5 to 77.5 %. The interpretation was that the subsequent weather condition was unknown when predicting in previous stages, it could not reflect the final yield. The accuracy could reach 72.5 % in tasseling stage. That is to say this model could predict the final risk effectively almost 2 month before the end of growing stage.

Table 3 Model's accuracy in each growing stage

Growing stage	Accurate rate (%)			
	No drought	Low drought	High drought	Average
Pre-planting	53.8	40.0	47.1	47.5
Jointing	76.9	40.0	64.7	62.5
Tasseling	69.2	70.0	76.5	72.5
Milky maturity	76.9	70.0	82.4	77.5
Maturity	61.5	70.0	82.4	72.5

Table 4 Model validation assessed by leave-one-out cross-method

Growing stage	Accuracy rate by leave-one-out cross-validation (%)		
	No drought	Low drought	High drought
Pre-planting	23.1	20	29.4
Jointing	38.5	20	29.4
Tasseling	61.5	20.0	41.2
Milky maturity	30.8	40.0	70.6
Maturity	38.5	10.0	64.7

The accuracy for high drought risk group was much higher than the other two classes, which reached 82.4 %. That indicated the model was suitable for regions with frequent droughts. This is because the drought disaster risk used in the model was calculated by yield losses. Although there was no drought disaster occurred but hail, freeze, and other disasters might also influence the final yield. Given that drought was the dominant disaster which influenced the final yield greatly, the accuracy of the model could reach the requirement.

Results in Table 3 showed that from pre-planting to milky maturity the average accuracy increased from 47.5 to 77.5 %. From milky maturity to maturity, the accuracy decreased to 72.5 %. That indicated inputting the precipitation data of stage 5 did not contribute to the accuracy of the model. Comparing to other growing stages, whether drought happened or not in maturity stage did not influence the final yield. This conclusion is consistent with maize physiology.

3.3 Model validation

We used leave-one-out cross-validation method to validate the model's accuracy. Cross-validation is a very important verification technology for discrimination effect. When creating the discriminant function, we removed one case a time and then used the rest cases to establish the discriminant function. Table 4 showed the validation accuracy was low in previous stages. This indicated that the model was unstable in previous stages. This was caused by the correspondence between precipitation, and final yield was not obvious in previous stage. For example, although the precipitation was enough in previous growing stages, the finial yield losses still could be large when drought happened in the following stages. In this condition, the model will misjudge. But as the growing stage advanced and precipitation data added, the model's accuracy of high drought group could reach 70.6 % in milky maturity stage, it performed well.

4 Conclusions

The SPI has been used extensively around the world. Multiple time scale SPI has advantage to reflect how crop influenced by drought that happened in different growing stages and fitted to be used in agriculture drought research. This study introduced discriminant analysis into risk prediction field combined with multi-scale SPI, yield losses analysis to establish the prediction model. The findings reflected that this prediction model has a good performance. The accuracy of the model could reach 82.4 %. This model should be appropriate for areas where drought happens frequently and is the dominant disaster.

This model could predict the final yield losses at each maize-growing stages, it provided timely information for decision makers to make effective management to reduce drought losses at the maximum level. The final results of this model related to maize yield directly, compared to other models it is more intuitive. Precipitation is the only input date in this model, it can be easily obtained. However, additional research could be studied to improve this drought risk prediction model. For example, how to get rid of the impacts of other disasters on the final yield losses could be studied.

Acknowledgments This study is supported by the National Key Technology R&D Program of China under Grant No. 2011BAD32B00-04, the National Grand Fundamental Research 973 Program of China under Grant No. 2010CB951102, the National Natural Science Foundation of China under Grant Nos. 41071326, 40871236, and 41201550, the National Scientific Research Special Project of Public sectors (Agriculture) of China under Grant No. 200903041.

References

Byun HR, Wilhite DA (1999) Objective quantification of drought severity and duration. J Climate 12:2747–2756

Domenikiotis C, Spiliotopoulos M, Tsiros E, Dalezios R (2004) Early cotton yield assessment by the use of the NOAA/AVHRR derived vegetation condition index (VCI) in Greece. Int J Remote Sens 25(14): 2807–2819

Fisher RA (1936) The use of multiple measurements in taxonomic problems. Ann Eugen 7:179–188

Guo X, Price KP, Stiles J (2003) Grasslands discriminant analysis using landsat TM single and multi-temporal data. Photogramm Eng Remote Sensing 69(11):1255–1262

Heim RR Jr (2002) A review of twentieth-century drought indices used in the United States. Bull Am Meteorol Soc 83:1149–1165

Hlavinka P, Trnka M, Semeradova D, Dubrovsky M, Zalud Z, Mozny M (2009) Effect of drought on yield variability of key crops in Czech Republic. Agric Forest Meteorol 149:431–442

Huang W, Yang X, Li M, Zhang X (2010) Evolution characteristics of seasonal drought in the south China during the past 58 years based on standardized precipitation index. Trans CSAE 26(7):50–60 (in Chinese)

Keyantash J, Dracup JA (2002) The quantification of drought: an evaluation of drought indices. Bull Am Meteorol Soc 83:1167–1180

Kim D, Byun H, Choi K (2009) Evaluation, modification, and application of the effective drought index to 200-Yeardroughtclimatology of Seoul, Korea. J Hydrol 378:1–12

Lin C, Chuang C, Chien C (2011) Factors affecting grassland succession retardation in the Juifang area. Nat Hazards 59:987–1002

Logan KE, Brunsell NA, Jones AR, Feddema JJ (2010) Assessing spatiotemporal variability of drought in the U.S. central plains. J Arid Environ 74:247–255

McKee TB, Doesken NJ, Kleist J (1993) The relationship of drought frequency and duration to time scales. Process, Eighth Conference on Applied Climatology, Anaheim, California 179–184

Palmer WC (1965) Meteorological drought. U.S. Weather Bureau Research Paper 45, National Weather Service, 55

Quiring SM, Papakryiakou TN (2003) An evaluation of agricultural drought indices for the Canadian prairies. Agric Forest Meteorol 118:49–62

Thessler S, Sesnie S, Ramos Bendana ZS, Ruokolainen K, Tomppo E, Finegan B (2008) Using k-nn and discriminant analyses to classify rain forest types in a Landsat TM image over northern Costa Rica. Remote Sens Environ 112:2485–2494

Trnka M, Hlavinka P, Semeradova D, Dubrovsky M, Zalud Z, Mozny M (2007) Agricultural drought and spring barley yields in the Czech Republic. Plant Soil Environ 53(7):306–316

Unganai LS, Kogan FN (1998) Drought Monitoring and corn yield estimation in Southern Africa from AVHRR data. Remote Sens Environ 63:219–232

Vicente-Serrano SM, Cuadrat-Prats JM, Romo A (2006) Early prediction of crop production using drought indices at different time-scales and remote sensing data: application in the Ebro Valley (north-east Spain). Int J Remote Sens 27(3):511–518

Wu H, Wilhite DA (2004) An operational agricultural drought risk assessment model for Nebraska, USA. Nat Hazard 33:1–21

Zhang JQ (2004) Risk assessment of drought disaster in the maize-growing region of songliao plain, China. Agric Ecosyst Environ 102:133–153