Linking drought indices to impacts to support drought risk assessment in Liaoning province, China

Yaxu Wang¹,²,³, Juan Lv¹,², Jamie Hannaford³,⁴, Yicheng Wang¹,², Hongquan Sun¹,², Lucy J. Barker³, Miaomiao Ma¹,², Zhicheng Su¹,², and Michael Eastman¹

¹China Institute of Water Resources and Hydropower Research, Beijing 100038, China
²Research Center on Flood and Drought Disaster Reduction, Ministry of Water Resources, Beijing 100038, China
³The UK Centre for Ecology & Hydrology, Oxfordshire, OX10 8BB, UK
⁴Irish Climate Analysis and Research UnitS (ICARUS), Maynooth University, Dublin, W23 F2K8, Ireland

Correspondence: Juan Lv (lujuan@iwhr.com)

Received: 19 September 2019 – Discussion started: 30 September 2019
Revised: 13 February 2020 – Accepted: 27 February 2020 – Published: 30 March 2020

Abstract. Drought is a ubiquitous and recurring hazard that has wide-ranging impacts on society, agriculture and the environment. Drought indices are vital for characterising the nature and severity of drought hazards, and there have been extensive efforts to identify the most suitable drought indices for drought monitoring and risk assessment. However, to date, little effort has been made to explore which index (or indices) best represents drought impacts for various sectors in China. This is a critical knowledge gap, as impacts provide important ground truth information for indices used in monitoring activities. The aim of this study is to explore the link between drought indices and drought impacts, using Liaoning province (northeast China) as a case study due to its history of drought occurrence. To achieve this we use independent, but complementary, methods (correlation and random forest analysis) to identify which indices link best to drought impacts for prefectural-level cities in Liaoning province, using a comprehensive database of reported drought impacts in which impacts are classified into a range of categories. The results show that the standardised precipitation evapotranspiration index with a 6-month accumulation (SPEI6) had a strong correlation with all categories of drought impacts, while the standardised precipitation index with a 12-month accumulation (SPI12) had a weak correlation with drought impacts. Of the impact datasets, “drought-suffering area” and “drought impact area” had a strong relationship with all drought indices in Liaoning province, while “population and number of livestock with difficulty in accessing drinking water” had weak correlations with the indices.

The results of this study can support drought planning efforts in the region and provide context for the indices used in drought-monitoring applications, so enabling improved preparedness for drought impacts. The study also demonstrates the potential benefits of routine collection of drought impact information on a local scale.

1 Introduction

Drought is one of the most pervasive natural hazards and can cause numerous and severe societal impacts. Drought impacts are mainly non-structural, widespread over large areas and often have a delayed onset in relation to the start of the drought event; therefore, it is challenging to properly define, quantify and manage drought (Mishra and Singh, 2010). There are a number of types of drought (Wilhite and Glantz, 1985), such as meteorological, agricultural, hydrological, social and ecological. Meteorological drought is defined as a deficit of rainfall for a period in respect to the long-term mean (Houérou, 1996). As these rainfall deficits propagate through the hydrological cycle, the other drought types occur as deficits occur in river flows, soil moisture and groundwater. Eventually impacts become manifest in the environment and society. China has experienced numerous droughts, which have caused great impact in many sectors since the 1950s, especially in Liaoning province in the dry northeast of the country (Zhang, 2004). Liaoning province experienced a severe drought from spring 2000 to
autumn 2001 which captured a large amount of attention from stakeholders and caused serious impacts as a result of the consecutive years of drought (Chen et al., 2016).

The costly nature of droughts means it is essential to plan and prepare for droughts proactively. Drought risk assessment is an essential prerequisite of this proactive approach (Wilhite, 2000; Wilhite and Buchanan, 2005) and provides the methods to predict the potential drought risk to society and the environment. Some risk assessment efforts focus primarily on the meteorological indices of drought, e.g. assessing the risk of a given severity of meteorological drought using historical precipitation data (Potopová et al., 2015). However, to adequately assess drought risk it is also necessary to characterise the consequences of drought occurrence, i.e. the impacts of drought on society, the economy and the environment (UNISDR, 2009).

There are a wealth of drought indices in the literature (Lloyd-Hughes, 2014); however they have predominantly been used for drought monitoring and early warning (e.g. Bachmair et al., 2016b) rather than drought risk assessment applications. The range of drought indices reflects the different types of drought which can be monitored, e.g. meteorological, hydrological and agricultural (Erhardt and Czado, 2017). Many indices, such as the standardised precipitation index (SPI), can be calculated over different timescales. This enables deficits to be assessed over different periods and can help in the monitoring of different types of drought. For example, shorter timescales, such as the SPI for 3 or 6 months, are used for agricultural drought monitoring, while SPI accumulations for 12 or 24 months are often applied to monitor hydrological droughts (Hong et al., 2001; Seiler et al., 2002). In China, many indices are used for drought monitoring, such as the Palmer drought severity index (PDSI), standardised precipitation evapotranspiration index (SPEI), SPI, China-Z index, and relative soil moisture and remote sensing indices (Hong et al., 2001; Wang and Chen, 2014; Wu et al., 2012; Yanping et al., 2018). Li et al. (2015) found that serious drought events occurred in 1999, 2000, 2001, 2007 and 2009 in China using the SPEI. Zhao et al. (2015) compared drought-monitoring results between the self-calibrating PDSI and SPEI in China with emphasis on the difference of timescales. Wu et al. (2013) developed an integrated surface drought index for agricultural drought monitoring in mid-eastern China. Drought-monitoring efforts in China tend to focus on meteorological and agricultural drought monitoring. Based on this and previous drought studies, the SPI, SPEI, soil moisture index and normalised difference vegetation index (NDVI) were selected in this research to characterise meteorological and agricultural drought. The relationship between drought indices and drought impacts, established by statistical methods (e.g. Bachmair et al., 2016a), can be used for drought risk assessment and appraisal of vulnerability. Vulnerability is by its nature difficult to define and measure, but in effect, drought impacts are symptoms of drought vulnerability and provide a proxy for vulnerability appraisal by demonstrating the adverse consequences of a given drought severity (Blauhut et al., 2015a).

There are many different types of drought impacts that affect many aspects of society and the environment, but drought impacts are rarely systematically recorded (Bachmair et al., 2016b). Some countries and regions have established drought impact recording systems to analyse historical drought impacts. A leading example of this is the US Drought Impacts Reporter (Svoboda and Hayes, 2011) which was launched as a web-based system in July 2005. More recently, the European Drought Impact Report Inventory (EDII) has been established (Stahl et al., 2016). Such databases are an important step forward, but the information in them is necessarily partial and biased, as a result of being effectively crowdsourced text-based information based on reported impacts from a range of sources (the media, grey literature, etc.). In contrast to many other countries, China has a relatively complete and systematically assembled quantitative drought impact information collection system. Data are collected and checked at the county level by the Drought Resistance Department via a formalised network of reporters, who collect information on drought impacts on agriculture, the industrial economy and water supply in every village. These data then are fed up to the national government and held by the State Flood Control and Drought Relief Headquarters (SFDH). This consistent collection of impact reporting provides a rich resource for drought risk assessment. However, impacts by themselves are not fully instructive, and to help inform risk assessment there is a need to understand their relationship with quantitative drought indices.

Understanding the relationship between drought indices and drought impacts and drought vulnerability is a vital step to improving drought risk management (Hong and Wilhite, 2004). However, whilst there have been many studies that develop, apply and validate drought indices, relatively few studies have assessed the link between indices and observed impacts. Bachmair et al. (2016b) noted that this literature tended to be dominated by studies focused on agricultural drought, generally linking indices like the SPI or SPEI and crop yield. Examples appraising multisectoral impacts are much sparser – there are several recent studies in Europe, utilising impact reports from the EDII. Stagge et al. (2014) and Bachmair et al. (2016b) used drought impacts from the EDII and various timescales of SPI, SPEI and streamflow percentiles. They found that the relationships between indices and impacts varied significantly by region, season, impact types, etc. whilst Blauhut et al. (2015a, b) developed a quantitative relationship between drought impact occurrence and SPEI using logistic regression in four European regions. However, all four studies assumed drought impacts were only measured by the drought impact occurrence (i.e. whether there was or was not an impact in a given month), the number of impacts or a combination of both. This means that all drought impacts had an equal weight without considering the...
duration, intensity or spatial extent of the individual impacts. In contrast, Karavitís et al. (2014) analysed drought impacts transformed into monetary losses to measure drought impacts in Greece; however, it is challenging to transform all drought impacts into monetary units – especially the indirect impacts of droughts.

In China, previous studies have also focused on agricultural drought risk assessment. Hao et al. (2011) applied the information diffusion theory to develop a drought risk analysis model which used affected crop area to measure the drought disaster. Zhao et al. (2011) established the relationship between drought frequency and simulated crop yield data in Henan Plain, and Jia et al. (2011) used the water stress coefficient and duration to establish a drought index. Li et al. (2009) analysed the links between historical crop yield and meteorological drought and established a meteorological drought risk index by combining the drought frequency, intensity, yield loss and extent of irrigation. The drought index was found to explain 60%–75% of the major crop yield reduction. In drought impact studies, Xiao-jun et al. (2012) collected information on the annual drought-affected area, damaged area, and annual losses in food yield at a national level from China Water Resources bulletins to explore the water management strategies during droughts. In Hao et al. (2011), drought impacts were only measured by the affected crop area at the 10 d time step at the county level. In our research, eight types of drought impacts are collected to measure drought impacts at the city unit (i.e. prefectural) level in Liaoning province, including not only consideration of the drought-affected area, damaged area and yield loss but also drought impacts on humans, livestock and the agricultural economy.

In summary, previous studies have focused on linking impacts to only one characteristic of drought (such as intensity or duration of occurrence) with most focusing on meteorological drought and agricultural impacts with little application of the results to drought vulnerability assessments, with the exception of Blauhut et al. (2015a, 2016) and Anderson et al. (2019), for example. Here we link drought indices to drought impacts in 14 cities in Liaoning province, northeast China, showcasing the use of the Chinese drought impact data from the SFDH. Using the drought impact–index linkage, we evaluate drought vulnerability in Liaoning province and assess what factors affect drought vulnerability. A drought vulnerability evaluation method that can be extended to other areas is then developed. The objectives of this paper are

1. to identify when and where the most severe droughts occurred between 1990 and 2013 in Liaoning province;
2. to identify which drought indices best link to drought impacts in Liaoning province;
3. to determine which city or area has higher drought vulnerability in Liaoning province; and
4. to ascertain which vulnerability factor or set of vulnerability factors have a higher contribution to drought vulnerability, as quantified in objective 3.

2 Materials

2.1 Study area

Located in northeastern China, Liaoning province, comprised of 14 prefectural cities, has a temperate continental monsoon climate with an annual average precipitation of 686 mm, which is unevenly distributed both temporally and spatially (Cai et al., 2015). Figure 1 shows the annual average rainfall across Liaoning, the southeast receiving on average more than 1000 mm yr$^{-1}$, whilst the northwest receives less than 500 mm yr$^{-1}$.

The annual average volume of freshwater resources in Liaoning province is 34.179 billion m$^3$, and the annual average per capita volume of water resources is 769 m$^3$ – about one-third of the per capita water resources for the whole of China (Sun et al., 2012). Freshwater resources are unevenly distributed within Liaoning province, with more freshwater resources in the southeast than the northwest (Liu and Guo, 2009; Cao et al., 2012). Thus, Liaoning province is one of the provinces with severe water shortages in northern China. Liaoning province is also a highly productive area for agriculture; spring maize is the dominant crop in agriculture production which makes it an important high-quality maize production area (Liu et al., 2013; Ren and Zhou, 2009). Due to these characteristics, when drought occurs, as has frequently been the case in Liaoning province, it causes a significant reduction in agricultural production (Yan et al., 2012). According to the SFDH, between 2000 and 2016 the average annual yield loss due to drought was 1.89 million t in Liaoning province, with an average annual direct agricultural economic loss of CNY 1.87 billion.

2.2 Data

2.2.1 Meteorological data

Daily precipitation and temperature data for each city in Liaoning province for the period 1990–2013 were obtained from the China Meteorological Administration (http://data.cma.cn/, last access: 3 January 2018). Although there are 52 meteorological stations in Liaoning province, due to the quality and length of the records and location of the stations, one meteorological site in each city (shown in Fig. 1) was selected to represent the meteorological conditions for the whole city in order to derive drought indices.

2.2.2 Soil moisture data

Daily soil moisture (SoilM) data for 96 soil moisture stations in Liaoning province (shown in Fig. 1) from 1990 to 2006.
were obtained from the Liaoning Provincial Department of Water Resources. Daily soil moisture was measured at three different depths, 10, 20 and 30 cm using frequency domain reflectometry soil moisture sensors. Soil moisture data were not available at most stations between November and February due to freezing conditions.

2.2.3 Normalised difference vegetation index (NDVI) data

Monthly MODIS normalised difference vegetation index (NDVI) data from 2000 to 2013 were collected in Liaoning province from the Geospatial Data Cloud (http://www.gscloud.cn/, last access: 10 February 2019); the daily maximum data were used to derive the monthly average NDVI.

2.2.4 Impact data

In contrast to many other countries, China has a systematic, centralised drought impact information collection system. Drought statistics include drought impacts; drought mitigation actions; and benefits of action to agriculture, hydrology and civil affairs. During a drought event, impact statistics are collected from every day to every 3 weeks, according to the drought warning level (Wang, 2014). When a drought warning is not triggered, drought impact data are collected after an event has ended which could be several months afterwards; and no data are collected when there is no drought event. Statistics for eight drought impact types were collected from the SFDH between 1990 and 2016 and aggregated into annual totals. The impact types used are listed in Table 1.

2.2.5 Vulnerability factors

The drought impacts described in Sect. 2.2.4 are mainly focused on the agriculture sector. As a result of this, the availability of data, and the findings of Junling et al. (2015) and Kang et al. (2014), vulnerability factors relevant to these impacts were selected. Vulnerability factors were collected from the 2017 Liaoning Statistical Yearbook to assess their contribution to the drought vulnerability (Liaoning Province Bureau of Statistics, 2017) and are shown in Table 2 for each city unit.
Table 1. The eight drought impact categories for Liaoning province used in this study collected by the SFDH.

<table>
<thead>
<tr>
<th>Impact</th>
<th>Abbreviation</th>
<th>Description</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drought-suffering area</td>
<td>DSA</td>
<td>The area that was officially declared in drought</td>
<td>kha</td>
</tr>
<tr>
<td>Drought-impacted area</td>
<td>DIA</td>
<td>The area that suffered crop yield loss by 10 % or more</td>
<td>kha</td>
</tr>
<tr>
<td>Disaster area</td>
<td>DA</td>
<td>The area that suffered crop yield loss by 30 % or more</td>
<td>kha</td>
</tr>
<tr>
<td>Recessed area</td>
<td>RA</td>
<td>The area that suffered crop yield loss by 80 % or more</td>
<td>kha</td>
</tr>
<tr>
<td>Population with difficulty in accessing drinking water</td>
<td>PHD</td>
<td>Rural populations that cannot have normal access to drinking water</td>
<td>10 k</td>
</tr>
<tr>
<td>Number of livestock with difficulty in accessing drinking water</td>
<td>NLH</td>
<td>Number of livestock that cannot have normal access to drinking water</td>
<td>10 k</td>
</tr>
<tr>
<td>Yield loss due to drought</td>
<td>YLD</td>
<td>The amount of yield loss due to drought</td>
<td>10 k t</td>
</tr>
<tr>
<td>Direct economic loss in agriculture</td>
<td>DELA</td>
<td>Direct losses to agricultural economy caused by drought</td>
<td>CNY 0.1 billion</td>
</tr>
</tbody>
</table>

Table 2. Vulnerability factors for each city in Liaoning province collected from the 2017 Liaoning Statistical Yearbook (Liaoning Province Bureau of Statistics, 2017).

<table>
<thead>
<tr>
<th>City</th>
<th>Per capita gross domestic product (kCNY)</th>
<th>Population per capita (10 k)</th>
<th>Crop cultivated area (ha)</th>
<th>Annual per capita water supply (m³)</th>
<th>Per unit area of fertiliser application (kg ha⁻¹)</th>
<th>Effective irrigation rate (%)</th>
<th>Number of electromechanical wells (k)</th>
<th>Reservoir total storage capacity (m³)</th>
<th>Per unit area of major agricultural products (kg ha⁻¹)</th>
<th>Livestock production (10 kt)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shenyang</td>
<td>755.8</td>
<td>733.9</td>
<td>656.0</td>
<td>91.5</td>
<td>1000.4</td>
<td>40.0</td>
<td>27.6</td>
<td>686.6</td>
<td>7090.5</td>
<td>64.5</td>
</tr>
<tr>
<td>Dalian</td>
<td>1143.4</td>
<td>595.6</td>
<td>327.0</td>
<td>73.4</td>
<td>1437.2</td>
<td>22.8</td>
<td>19.0</td>
<td>2523.0</td>
<td>4914.3</td>
<td>70.8</td>
</tr>
<tr>
<td>Anshan</td>
<td>422.9</td>
<td>345.7</td>
<td>247.9</td>
<td>42.3</td>
<td>1031.8</td>
<td>30.1</td>
<td>4.1</td>
<td>91.9</td>
<td>6641.6</td>
<td>36.7</td>
</tr>
<tr>
<td>Fushun</td>
<td>402.7</td>
<td>214.8</td>
<td>116.1</td>
<td>94.7</td>
<td>776.9</td>
<td>37.4</td>
<td>1.8</td>
<td>2575.5</td>
<td>6342.9</td>
<td>10.4</td>
</tr>
<tr>
<td>Benxi</td>
<td>511.1</td>
<td>150.0</td>
<td>58.0</td>
<td>167.9</td>
<td>756.3</td>
<td>29.9</td>
<td>0.4</td>
<td>6078.8</td>
<td>6606.3</td>
<td>9.3</td>
</tr>
<tr>
<td>Dandong</td>
<td>315.8</td>
<td>237.9</td>
<td>190.4</td>
<td>28.0</td>
<td>1049.7</td>
<td>41.7</td>
<td>1.4</td>
<td>16202.8</td>
<td>6056.9</td>
<td>20.2</td>
</tr>
<tr>
<td>Jinzhou</td>
<td>341.8</td>
<td>302.2</td>
<td>457.2</td>
<td>46.6</td>
<td>915.4</td>
<td>41.3</td>
<td>18.7</td>
<td>977.9</td>
<td>6825.7</td>
<td>64.0</td>
</tr>
<tr>
<td>Yingkou</td>
<td>496.7</td>
<td>232.8</td>
<td>109.4</td>
<td>42.4</td>
<td>1564.6</td>
<td>67.7</td>
<td>12.3</td>
<td>269.6</td>
<td>7325.0</td>
<td>13.5</td>
</tr>
<tr>
<td>Fuxin</td>
<td>215.9</td>
<td>188.9</td>
<td>479.4</td>
<td>39.7</td>
<td>881.9</td>
<td>30.1</td>
<td>26.6</td>
<td>545.0</td>
<td>5243.6</td>
<td>49.6</td>
</tr>
<tr>
<td>Liaoyang</td>
<td>373.4</td>
<td>178.6</td>
<td>162.8</td>
<td>42.4</td>
<td>1002.6</td>
<td>44.8</td>
<td>4.0</td>
<td>1418.8</td>
<td>7202.2</td>
<td>11.0</td>
</tr>
<tr>
<td>Panjin</td>
<td>778.3</td>
<td>130.1</td>
<td>143.0</td>
<td>70.2</td>
<td>937.0</td>
<td>68.7</td>
<td>1.0</td>
<td>141.5</td>
<td>8918.3</td>
<td>23.8</td>
</tr>
<tr>
<td>Tieling</td>
<td>196.5</td>
<td>299.9</td>
<td>548.5</td>
<td>12.2</td>
<td>960.2</td>
<td>32.0</td>
<td>18.1</td>
<td>2174.5</td>
<td>8397.1</td>
<td>46.0</td>
</tr>
<tr>
<td>Chaoyang</td>
<td>210.1</td>
<td>341.1</td>
<td>464.5</td>
<td>15.8</td>
<td>874.7</td>
<td>42.0</td>
<td>17.4</td>
<td>2085.6</td>
<td>6292.0</td>
<td>63.6</td>
</tr>
<tr>
<td>Huludao</td>
<td>230.8</td>
<td>280.5</td>
<td>249.7</td>
<td>18.7</td>
<td>976.8</td>
<td>28.9</td>
<td>14.0</td>
<td>892.7</td>
<td>4852.3</td>
<td>35.4</td>
</tr>
</tbody>
</table>

2.3 Methods

2.3.1 Drought indices

Two meteorological indices were selected, the standardised precipitation index (SPI; McKee et al., 1993) and the standardised precipitation evapotranspiration index (SPEI; Vicente-Serrano et al., 2010). These standardised indices are widely used in drought-monitoring applications around the world, and the SPI is recommended by the World Meteorological Organization for monitoring meteorological drought (Hayes et al., 2011). This is due to the flexibility of being able to derive the SPI over different timescales and because it can be compared across time and space.

The SPI, in its default formulation, assumes that precipitation obeys the gamma (Γ) skewed distribution, which is used to transform the precipitation time series into a normal distribution. After normalisation, classes of drought can be defined with the cumulative precipitation frequency distribution (Botterill and Hayes, 2012; Hayes et al., 1999). The SPEI uses the same standardisation concept using the climatic water balance (that is, precipitation minus potential evapotranspiration; PET) instead of precipitation. In this study PET is calculated by the Thornthwaite method (Thornthwaite, 1984), using observed temperature and sunlight hours (estimated from latitude) as inputs. The SPEI is calculated by normalising the climatic water balance using a log-logistic probability distribution (Vicente-Serrano et al., 2010).
The SPI and SPEI are easily calculated and can fit a wide range of accumulation periods of interest (e.g. 1, 3, 12, 24, 72 months; Edwards, 1997). The SPEI has the added advantages of characterising the effects of temperature and evapotranspiration on drought. In this study, the SPI and SPEI were calculated for five accumulation periods, 6, 12, 15, 18 and 24 months, from 1990 to 2013 for 14 meteorological stations (i.e. one in each city – as shown in Fig. 1). Generally, precipitation in Liaoning province is concentrated between April and September which corresponds to the growing stage of spring maize. Considering the climatology and crop growth period, the SPI6 and SPEI6 ending in September were selected for this study, i.e. calculated using precipitation during April to September. The 12, 15, 18 and 24 months SPI and SPEI ending in December were also analysed with the annual drought impacts during 1990 to 2013.

Using the daily soil moisture of 10, 20 and 30 cm depths, the daily average soil moisture for each station was calculated using Eqs. (1) and (2) (Lin et al., 2016).

\[ \theta_1 = \theta_{10}, \quad \theta_2 = \frac{\theta_{10} + \theta_{20}}{2}, \quad \theta_3 = \frac{\theta_{20} + \theta_{30}}{2}, \]

where \(\theta_1\) is the soil moisture of the first layer \((i = 1, 2, 3)\), \(\theta_{10}, \theta_{20} \) and \(\theta_{30}\) are the measured values at different depths \((10, 20 \) and \(30 \) cm); \(\bar{\theta}\) is the average soil moisture; \(h_1\) is the thickness of the second layer of soil, and \(H\) is the total thickness of the measured soil.

Some of the daily soil moisture data were missing; however this was limited to 17\% of the total soil moisture data. In some cases there were missing data for one depth of soil moisture measurement. In these cases, the average soil moisture of the other two layers was calculated, and where there was only one layer of soil moisture available it was used to represent the average soil moisture. The annual average soil moisture was calculated based on the available daily soil moisture (March to October) and was analysed with the annual drought impact data during 1990 to 2006. As each city has more than one soil moisture station, the annual soil moisture of each station was calculated and then averaged to one value for each city.

The area-averaged NDVI at the city unit was calculated using the monthly NDVI. The critical stage of the spring maize growth in Liaoning is in July, so the area-averaged NDVI in July was selected for the analysis with the annual drought impacts from 2000 to 2013.

2.3.2 Correlation analysis

The Pearson correlation method was used to characterise the correlation between indices and the selected drought impacts (Özger et al., 2009). Due to the limited availability of soil moisture data, correlation analysis of soil moisture and drought impact data was only carried out in nine cities. The linkage between drought indices and impacts was used to assess the drought vulnerability in Liaoning province. It can be inferred that the greater the impact caused by droughts at a specific severity (measured according to the SPI and SPEI), the higher the drought vulnerability of the city.

2.3.3 Random forest modelling

Decision trees are regularly implemented for machine learning tasks. They resemble flowcharts, consisting of a series of branches, internal nodes and leaf nodes. Internal nodes typically represent binary conditions of the explanatory variables. These nodes are connected to other internal nodes by branches, which represent the outcome of the previous internal node. Leaf nodes represent the outcome classes. Internal nodes are eventually connected to leaf nodes, which represent the outcome classes of the classification task. Whilst quick to train and interpretable, decision trees are limited by overfitting to the training set. Random forests (RFs) reduce overfitting by fitting an ensemble of uncorrelated decision trees. This is achieved using bootstrap aggregation with replacement (bagging) and by only considering a random subset of features for splitting at each internal node (Breiman, 1996). As well as the reduction in overfitting compared to decision trees, the advantages of RFs include its fast training speed, good accuracy and relative efficiency (Mutanga et al., 2012). Additionally, once RF models are established, the values of the predictor that correspond to the first split in the decision tree can be extracted as thresholds corresponding to impact occurrence (Bachmair et al., 2016a).

In this analysis random forests were built for regression. This is achieved by assigning categorical outcomes at each leaf node and using the mean prediction as the outcome. The R package randomForest was employed to identify the relationship of drought indices to drought impacts (Kursa, 2017; Liaw and Wiener, 2002) using 5000 decision trees for each RF model. The variance explained was used to determine the goodness of fit of the random forest model (Fukuda et al., 2016a). The mean squared error (MSE), Eq. (3), was used to evaluate the importance of each index:

\[ \text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2, \]

where \(y_i\) and \(\hat{y}_i\) are the observed drought impacts and the estimated drought impacts of each city, \(i\), respectively; \(n\) is the length of the time series.

The percent change in MSE (MSE \%) is the difference in accuracy when the effect of the variable is excluded (i.e. if the SPEI6 is excluded from the model, the MSE \% of the model may increase). Higher MSE \% represents a higher index importance. The first splitting values of each decision tree were also extracted.
Soil moisture and the NDVI were not analysed using the random forest approach due to their short time series and prevalence of missing data.

2.3.4 Standardisation of drought impacts and vulnerability factors

To ensure comparability and to facilitate their visualisation, the drought impacts and vulnerability factors were standardised to a value from 0 to 1 using Eqs. (4) and (5) (Below et al., 2007).

\[
\text{SDI}_i = \frac{\text{DI}_i - \text{min DI}}{\text{max DI} - \text{min DI}},
\]

(4)

\[
\text{SVF}_j = \frac{\text{VF}_j - \text{min VF}}{\text{ma VF} - \text{min VF}},
\]

(5)

where \(\text{SDI}_i\) and \(\text{DI}_i\) are the standardised drought impacts and the drought impacts of year \(i\) in Liaoning province, respectively; \(\text{max DI}\) and \(\text{min DI}\) are the maximum and minimum values of drought impacts in all years for the given impact type; \(\text{SVF}_j\) and \(\text{VF}_j\) are the standard vulnerability factors and vulnerability factors of city \(j\) in Liaoning province; and \(\text{ma VF}\) and \(\text{min VF}\) are the maximum and minimum values of each category of vulnerability factors in all cities.

3 Results

3.1 Drought monitoring and drought impacts

Figure 2 shows the SPEI and the drought impact data for Liaoning province from 1990 to 2013; the sum of SDI is the sum of all types of standardised drought impacts in the 14 cities for each year.

Figure 2 shows that the most severe droughts occurred in 2000, 2001 and 2009, whilst in 1994, 1995, 2012 and 2013 there was above-normal precipitation. From a visual inspection, the largest impacts are generally associated with the lowest index values. This suggests that there is a relationship between the drought indices and drought impacts. This relationship is explored quantitatively in the next sections.

Figure 3 shows the spatial distribution of the average of each drought impact type collected between 1990 and 2016. It shows that the most severe drought impacts were recorded in the drier northwestern part of Liaoning province than in eastern parts of the province; the NLH was highest in Dalian, whilst Shenyang had the biggest PHD.

3.2 Correlation of indices with impacts

The Pearson correlation coefficient (\(r\)) for each city and drought impacts is shown in Fig. 4. In most cases the drought index is negatively correlated with the drought impacts, suggesting that the lower (and more severe) the drought index, the greater the drought impact. However, correlation strength and direction varied between the cities and impact types, ranging between \(-0.890\) and \(0.621\). In most cities of Liaoning province, the NDVI and SoiM have a weak correlation with most types of drought impacts. In Dalian, Chaoyang and Fuxin, all drought indices had a strong correlation with DA, whilst there was a significant correlation for drought-impacted area in Jinzhou, Fuxin and Dalian, where most of the correlations were significant (\(p < 0.01\)). The strongest correlation was found between indices and PHD in Dalian, while it was the weakest in Dandong. There is a positive correlation between PHD and the NDVI in Fushun, while NLH has a positive correlation with the NDVI in Anshan. Generally, the SPEI6 had the strongest correlation with all types of drought impacts, whilst the SPI12 had the weakest correlation. The SPEI typically exhibited stronger correlations with drought impacts than the SPI with the same accumulation period.

DSA and DIA had a strong correlation with all drought indices in Liaoning province, while PHD and NLH had a weak correlation. The average correlation coefficient across all drought indices and DSA in Liaoning was \(-0.43\), while the average correlation coefficient with PHD and NLH was \(-0.22\) and \(-0.27\), respectively. Drought indices showed a moderate correlation with RA and YLD with average correlation coefficients of \(-0.32\) and \(-0.37\), respectively.

The performance of soil moisture varied significantly between cities and impact types (Fig. 4); it had a strong correlation with the impacts in Chaoyang and a weak correlation in Huludao. In Chaoyang, the correlation between soil moisture and drought impacts was significant (\(\alpha = 0.01\)), whilst other cities were not significantly correlated.

3.3 Drought index importance in random forest models

Each drought impact type was selected as the response variable in the random forest. On average the random forests explained 41% of the variance observed within the drought impacts. The MSE % for each city and impact type is shown in Fig. 5. The MSE % can be seen to vary between different impact types. DIA and YLD have higher MSE % than other impact types, with an average MSE % of 3.02 and 3.01, respectively. The PHD and NLH had lower MSE %, with an average MSE % of 1.58 and 1.39, respectively. DSA and RA had a moderate relationship with drought indices. The SPEI performed better than the SPI with the same durations; the SPEI6 had the highest importance with drought impacts. The SPI12 was the least important index in terms of drought impacts. Indices had a higher importance with impacts in Anshan and Dalian and lower importance in Yingkou and Dandong.

The variables’ MSE % identified from the random forest analysis generally match those with strong negative correlations. This supports the statement that indices are negatively related to impacts. The threshold of impact occurrence based on the indices was also identified in the RF analysis using the first splitting value. Figure 6 shows the distribution of
first splitting values of each decision tree within the RF. The average first splitting values for the SPI18 and SPI24 were higher than those of the SPI6, SPI12 and SPI15 (i.e. a more negative index value and more severe meteorological drought state) for all categories of drought impacts. For the SPEI, the results were similar (i.e. long-term deficits must be more severe to result in equivalent impacts compared to short-term deficits) but there was more variability between accumulations. When viewed in terms of impact types, DSA had a low threshold, indicating that DSA impacts occur more readily than DA or RA, as may be expected. The impact occurrence of index values increase for DSA, DIA, DA and RA; and YLD and DELA tended to occur for more severe water deficits, with the highest severity threshold being for NLH, indicating that only very severe drought conditions triggered impacts on livestock.

3.4 Drought vulnerability evaluation

The results of correlation analysis and random forest analysis suggest that, in most parts of Liaoning province, the SPEI with the 6-month accumulation period has the strongest relationship with drought impacts. The SPEI6 was therefore selected to assess the drought vulnerability of the 14 cities. Regression analysis was performed on the SPEI6 for each category of drought impact, and an example is given in Fig. 7 which shows the linear regression of DSA with the SPEI6 in the 14 cities. It can be surmised that the more serious the drought impacts for a specific drought severity (as defined by the SPEI6), the higher the drought vulnerability. Fuxin, Tieling, Chaoyang, Jinzhou and Shenyang have a higher vulnerability to DSA compared to the other cities.

Similar analyses were performed for all impact types, and Fig. 8 displays the drought impacts each city in Liaoning province is most vulnerable to. It can be seen from Fig. 8 that there is little difference between cities in terms of sensitivity to various categories of drought impacts. Considering the various impacts, Chaoyang, Jinzhou, Tieling, Fuxin and Shenyang had the highest drought vulnerability – these cities are all located in the northwest part of Liaoning province. Dalian was most vulnerable to NLH.

3.5 Vulnerability analysis

A stepwise regression model was built to explain the variation in each type of standardised drought impact using vulnerability factors (listed in Table 2) as predictors. Drought impacts are symptoms of vulnerability and so can be used to estimate vulnerability to drought (Blauhut et al., 2015a). The vulnerability to drought can be assessed by maintaining a constant severity of drought (i.e. particular drought index values) and comparing the resultant impacts. More serious impacts correspond to higher vulnerability. Thus, the standardised drought impacts corresponding to a severe meteorological drought (SPEI6 equal to $-1.5$) were regressed on the standardised drought vulnerability factors for 2017 to assess drought vulnerability for each drought impact type. Table 3 shows the results of the stepwise regression model, demonstrating the contribution of vulnerability factors to each category of drought impact; the results varied for each impact type.

The relatively high $R^2$ values demonstrate the ability of the vulnerability factors to explain the variability exhibited by each drought impact. This is particularly the case for DSA, drought-suffering area, and PHD, population with difficulty in accessing drinking water, which had associated $R^2$ values of 0.894 and 0.805, respectively. Population, crop cultivated area and livestock production explained 89.4 % of

**Figure 2.** Standardised precipitation evapotranspiration index (SPEI) for 6-, 12-, 15-, 18- and 24-month accumulation periods, and the sum of the standardised drought impacts (SDIs) for each impact type listed in Table 1 for Liaoning province from 1990 to 2013.
the variation in DSA. Population, in combination with per unit area of fertiliser application and reservoir total storage capacity, also contributed to the DA model, explaining 80.5% of DA variation.

Population, crop cultivated area and livestock production were identified as significant predictors in four, five and three models, respectively, more than other vulnerability factors. Crop cultivated area, the most frequently significant predictor of drought impacts, also exhibited relatively high regression coefficients, demonstrating the strong relationship between the areas cultivated for crops and the vulnerability to drought impacts. These results are paralleled in a composite drought vulnerability tool, which assigns relatively high weighting to area of land irrigated (Quinn et al., 2014).

Population exhibited negative regression coefficients for three of four drought impacts, suggesting that, as the population increased, the vulnerability to drought impacts decreased. However, population exhibits correlation with crop cultivated area and livestock production. This, paired with potential unaccounted for interactions between population...
and other predictors, may have resulted in inaccurate population coefficient estimation. This is supported by a positive population coefficient for predicting NLH. Population was used exclusively to predict NLH; thus correlation with other predictors and interaction effects were unable to influence the coefficient. Furthermore, Fig. 9 demonstrates that, as population increases, DSA, DIA and DELA increase. The composite drought vulnerability tool of Quinn et al. (2014) does not explicitly account for population, making a direct comparison not possible. However, it does assign a positive relationship between the ratio of rural population and drought vulnerability, which may explain the unexpected negative coefficients presented here. The number of electromechanical wells also exhibited a positive coefficient, suggesting that, as the number of wells increases, drought impacts increase. However, it is possible that electromechanical wells are more prevalent in more drought-prone areas; thus, the positive coefficient may simply demonstrate an association between electrometrical wells and RA.

Whilst drought vulnerability factors were able to explain 47.4%–89.4% of the variability in drought impacts, annual per capita water supply, effective irrigation rate and per unit area of major agricultural products were not identified as significant predictors of any drought impact type. It is important to consider, however, that correlation between these other vulnerability factors could result in them not being identified as significant, as the information is already contained within other vulnerability factors. This is supported by drought impact correlations with per capita water supply and effective irrigation rate. However, minimal correlations between drought impacts and unit area of major agricultural products were observed, suggesting that the absence of a detected relationship may be a true reflection.

4 Discussion

The methodology in this research has a number of distinctive characteristics in relation to previous drought im-
Figure 5. The MSE % of drought indices (SPI and SPEI) with drought impacts (identified by the codes in Table 1) in Liaoning province using random forest.

Figure 6. Box plots showing the splitting value (i.e. the thresholds of impacts) in the random forest construction across all impact types for each index (a), and across all indices for each impact type (b) in Liaoning province.
Figure 7. Linear regression results of drought-suffering area (DSA) with the SPEI6 in each of the 14 cities in Liaoning province.
considered not only the occurrence of drought impacts but also direct economic losses in agriculture. In addition, we collect data enabling regional comparisons. The drought impact data used here included impact variables that are rarely available in other settings, e.g. population with difficulty in accessing drinking water, yield loss due to drought and number of livestock with difficulty available in other settings. The extensive drought impact data were systematically collected at county level, which is a consistent and reliable data source enabling regional comparisons. The drought impact data is most vulnerable to, based on the results of the linear regression.

**Figure 8.** Map showing the drought impacts each city in Liaoning province is most vulnerable to, based on the results of the linear regression.

**Table 3.** The vulnerability factors selected for the stepwise regression model and the $R^2$ of the resulting model for each impact type (identified by the codes in Table 1).

<table>
<thead>
<tr>
<th>Drought impacts</th>
<th>Predictors vulnerability factors</th>
<th>Standardised coefficients</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>DSA</td>
<td>Crop cultivated area</td>
<td>0.814</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Population</td>
<td>−0.476</td>
<td>0.894</td>
</tr>
<tr>
<td></td>
<td>Livestock production</td>
<td>0.451</td>
<td></td>
</tr>
<tr>
<td>DIA</td>
<td>Crop cultivated area</td>
<td>1.098</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Population</td>
<td>−0.451</td>
<td>0.743</td>
</tr>
<tr>
<td>DA</td>
<td>Livestock production</td>
<td>0.691</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Per capita gross domestic product</td>
<td>−0.436</td>
<td>0.731</td>
</tr>
<tr>
<td>RA</td>
<td>Number of electromechanical wells</td>
<td>0.629</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Per capita gross domestic product</td>
<td>−0.452</td>
<td>0.541</td>
</tr>
<tr>
<td>PHD</td>
<td>Crop cultivated area</td>
<td>0.949</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Reservoir total storage</td>
<td>0.472</td>
<td>0.805</td>
</tr>
<tr>
<td></td>
<td>Per unit area of fertiliser appli-</td>
<td>0.352</td>
<td></td>
</tr>
<tr>
<td></td>
<td>cation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NLH</td>
<td>Population</td>
<td>0.720</td>
<td>0.474</td>
</tr>
<tr>
<td>YLD</td>
<td>Crop cultivated area</td>
<td>0.798</td>
<td>0.606</td>
</tr>
<tr>
<td>DELA</td>
<td>Crop cultivated area</td>
<td>0.556</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Population</td>
<td>−0.879</td>
<td>0.786</td>
</tr>
<tr>
<td></td>
<td>Livestock production</td>
<td>0.793</td>
<td></td>
</tr>
</tbody>
</table>

The vulnerability factors selected for the stepwise regression model and the $R^2$ of the resulting model for each impact type (identified by the codes in Table 1).

The results from the correlation analysis were consistent with the results from the RF analysis. Drought-suffering area (DSA) and drought impact area (DIA) had strong correlations with all drought indices in Liaoning province, while PHD and NLH had a weak correlation with indices. This was because DSA and DIA are direct impacts of agricultural drought, whilst PHD and NLH are related to many additional factors, such as drinking water source location and the quality of water resources; for example, livestock can drink wa-
ter from the river directly, but the quality of the river water means it is not suitable for humans – for this reason, NLH showed the least sensitivity to water deficits.

The random forest algorithms presented in this paper explained an average of 41% of the variance observed within the drought impact data. This is relatively modest and may be partially due to limitations associated with the impacts data. The collinearity of the drought indices (e.g. SPI6 is correlated with SPEI6) is also a potential cause of the low MSE %. The correlation coefficients calculated for drought indices and NLH in Yingkou and PHD in Fushun were positive. This was unexpected given the interpretation of these indices as estimations of drought severity and the majority of reported correlation coefficients being negative. Therefore, it seems likely this result is not representative of the true relationships between these indices and impacts and instead is an artefact of imperfect impact data. To explore this, years with the highest numbers of impacts were removed before the correlation coefficients were estimated. This resulted in a negative correlation coefficient, providing further evidence for the positive correlation coefficients not being representative of the true relationships in these cities. The availability of more data would enable a better approximation of the true relationships between indices and impacts.

For all the drought impacts, Dalian and Fuxin showed the highest correlation coefficients among drought impacts and drought indices in all cases. The most vulnerable cities were Fuxin, Tieling, Chaoyang, Jinzhou and Shenyang, which are all located in the northwestern part of Liaoning province indicating there is a high drought vulnerability and drought risk in northwestern Liaoning. This is consistent with existing research by Yan et al. (2012) and Zhang et al. (2012), which established a drought risk assessment index system to assess drought risk in northwestern Liaoning. This is consistent with existing research by Yan et al. (2012) and Zhang et al. (2012), which established a drought risk assessment index system to assess drought risk in northwestern Liaoning. Zhang et al. (2012) used indicators such as precipitation, water resources, crop area, irrigation capacity and drought resistance cost to measure drought risk and found that Fuxin, Chaoyang and Shenyang had a high drought risk.

The above results are also in general agreement with Hao et al. (2011), who used 10 d affected crop area data as the drought impacts to assess drought risk in China at the county level. Their results showed that the west Liaohe Plain had a high risk. The results presented in our paper identify Chaoyang and Fuxin as having the highest drought vulnerability – the majority of these two cities’ areas are located on the west Liaohe Plain.

As the accumulation period increased, the first splitting value extracted from the random forest model tended to decrease, suggesting that, over longer accumulation periods, larger water deficits are required for equivalent impacts to occur. There are severe water deficits associated with RA occurrence since they caused more yield loss compared to DIA and DA. Drinking water for livestock requires lower water quality compared to that for humans; for example, livestock can drink water from the river directly, but the water quality of the river cannot meet the human drinking needs. For this reason, NLH showed the least sensitivity to water deficits.

The relationships analysed in this research support the use of drought indices as a predictor of drought impacts, and the impact thresholds identified can also support improved drought warning and planning. The drought vulnerability map (Fig. 8) can be used to support drought risk planning, helping decision makers to implement appropriate drought mitigation activities through an improved understanding of the drivers of drought vulnerability – for example, by sinking more wells to enhance resilience to drought (noting of course, that this measure has potential longer-term implications on, for example, groundwater exploitation; see e.g. Changming et al., 2001).

The methods used here can be applied in other areas to better understand drought impacts and drought vulnerability where similar data (e.g. drought impacts, meteorological data) are collected. While systematic statistical archives of drought impact are comparatively rare, there are numerous other potential sources of impact data globally that could be used (e.g. see Bachmair et al., 2016b).
5 Conclusion

This study used correlation analysis and random forest methods to explore the relationship between drought indices and drought impacts. It assessed drought risk in Liaoning province and proposes a drought vulnerability assessment method which can be applied to study the contribution of various socioeconomic factors to drought vulnerability. Here, we return to the original objectives of the study to summarise the key findings.

1. When and where did the most severe droughts occur between 1990 and 2013 in Liaoning province? Based on the drought-monitoring results of the SPI, severe droughts occurred in 2000, 2001, and 2009. In 2000–2001, drought resulted in many impacts in Liaoning province, particularly in the northwestern part of Liaoning province. The drought-monitoring data shown corresponded well with the recorded drought impacts.

2. Which drought indices best link to drought impacts in Liaoning province? The results showed that the indices varied in their capacity to identify the different type of drought and impacts. The strongest correlation was found for the SPEI at 6 months, whilst the SPI12 had a weak correlation with drought impacts. The SPEI was found to better link to drought impacts than the SPI for the same accumulation period. The NDVI and soil moisture index showed some links with impacts in some cities, but the results were generally weaker and less consistent than for either the SPI or the SPEI – primarily reflecting the limitations in the soil moisture index and NDVI datasets.

3. Which city or areas has a higher drought vulnerability in Liaoning province? Chaoyang, Jinzhou, Fuxin, Shenyang, and Tieling had higher drought vulnerability, all of which are located in the northwestern part of Liaoning province, indicating that drought vulnerability is higher in these regions than in other parts, which is consistent with previous research. However, in contrast with past work, the present research provides a much more comprehensive assessment based on the occurrence of observed impact data.

4. Which vulnerability factor or set of vulnerability factors has a higher contribution to drought vulnerability? Population and crop cultivated area were strongly associated with drought vulnerability, suggesting these factors are good indicators of drought vulnerability. However, the complexities of these relationships with drought vulnerability require further investigation.

The results shown here give a clearer understanding about drought conditions in Liaoning province. The linkage developed can be used to assess drought risk and to map vulnerability. It can also be used to help develop early-warning systems and predict drought impacts, which are vital tools for drought management. The results of the vulnerability analysis can guide management measures to mitigate drought impacts – an important step to shift from post-disaster recovery to proactive pre-disaster prevention.

Data availability. Some data used during the study are proprietary or confidential in nature and may only be provided with restrictions (e.g. drought impacts data, soil moisture). Daily meteorological data can be explored (available by contacting Hongquan Sun, sunhq@iwhr.com). NDVI data can be obtained from the Geospatial Data Cloud (available by contacting Yaxu Wang, wangyxiwhr@sina.com). Vulnerability data were from the Liaoning Statistical Yearbook which can be obtained from the Liaoning Province Bureau of Statistics (http://www.ln.stats.gov.cn/sjsj/sjcx/nds/otherpages/2017/indexch.htm, last access: March 2020).

Author contributions. YW, JL, JH, YiW and LB discussed and developed the aims of the paper. YaW was responsible for the data analysis and visualisation and prepared the original manuscript, with contributions from HS, LB, JH, MM, ZS and ME.

Competing interests. The authors declare that they have no conflict of interest.

Special issue statement. This article is part of the special issue “Recent advances in drought and water scarcity monitoring, modelling, and forecasting (EGU2019, session HS4.1.1/NH1.31)”.

Acknowledgements. The authors gratefully acknowledge funding support for this research provided by the Chinese National Key Research and Development Programme (no. 2017YFC1502402), the China Institute of Water Resources and Hydropower Research (JZ0145B592016, JZ0145B582017), and the China Scholarship Council. Jamie Hannaford, Lucy Barker and Michael Eastman were supported by the NERC National Capability Official Development Assistance project SUNRISE (Sustainable Use of Natural Resources to Improve Human Health and Support Economic Development; grant no. NE/R000131/1).

Financial support. This research has been supported by the National Key Research and Development Project (grant no. 2017YFC1502402), the China Institute of Water Resources and Hydropower Research (grant nos. JZ0145B592016 and JZ0145B582017), the China Scholarship Council (grant no. 201808110260), and the NERC National Capability Official Development Assistance project SUNRISE (grant no. NE/R000131/1).
References


