Prediction of the Impact of Climate Change on Drought: an evaluation of six UK catchments using two stochastic approaches

Abstract

Under future climate scenarios, possible changes of drought patterns pose new challenges for water resources management. For quantifying and qualifying drought characteristics in the United Kingdom, the drought severity indices (DSI) of six catchments are investigated and modelled by two stochastic methods: autoregressive integrated moving average (ARIMA) models and the generalised linear model (GLM) approach. From the ARIMA models, autocorrelation structures are first identified for the drought index series, and the unexplained variance of the series is used to establish empirical relationships between drought and climate variables. Based on the ARIMA results, mean sea level pressure and possibly the North Atlantic Oscillation (NAO) index are found to be significant climate variables for seasonal drought forecasting. Using the GLM approach, occurrences and amounts of rainfall are simulated with conditioning on climate variables. From the GLM simulated rainfall for the 1980s and 2080s, the probabilistic characteristics of the drought severity are derived and assessed. Results indicate that the drought pattern in the 2080s is less certain than for the 1961-1990 period, based on the Shannon entropy, but that droughts are expected to be more clustered and intermittent. The 10th and 50th quantiles of drought are likely higher in the 2080s scenarios but there is no evidence showing the changes in the 90th quantile extreme droughts.

Keywords: drought, generalised linear model (GLM), autoregressive integrated moving average (ARIMA), extremes, persistence, forecast

Introduction

Although the intensity of precipitation events is projected to increase under future climate scenarios (Solomon et al., 2007), an increased risk of widespread drought is also expected (Bates et al., 2008), and higher drought frequency has been observed in several recent regional (e.g. Timbal, 2004) and global studies (e.g. Dai et al., 2004; Neelin et al., 2006). Drought impacts are multi-faceted, including for example effects on water resources (e.g. McIntyre et al., 2003; Shiau, 2006), agriculture (e.g. Wang, 2005), ecosystems (e.g. Harrison, 2000; Reichstein et al., 2002) and fire hazard (Keetch and Byram, 1968; Hessl et al., 2004), as well as socioeconomic aspects such as insurance (Bardsley et al., 1984). Hence drought monitoring and management are important issues (Wilhite, 1996), and it is necessary to develop decision-support drought modelling based on the best currently available information for climate change adaptation and mitigation. The main tools for hydrological investigation of climatic impacts related to anthropogenic emissions are Global Circulation Models (GCMs) (Wheater, 2002). Several recent studies (e.g. Blenkinsop and Fowler, 2007; Sheffield and Wood, 2008) use GCM or Regional Circulation Model (RCM) outputs to assess drought characteristics. This paper addresses the use of climate model outputs for drought assessment under scenarios of climate change, and in particular the development of appropriate methods for statistical downscaling of GCM outputs for drought impact assessments.
Defining drought is difficult because of complicated interactions between meteorological conditions, the natural environment and human factors. A universal definition of drought for various applications is not available, even though the concept of drought has been widely reviewed (e.g. Mawdsley et al., 1994; Rossi, 1992; Hisdal and Tallaksen, 2001; Heim, 2002). In general, drought at different temporal and spatial scales can be grouped into four categories: meteorological, agricultural, hydrological and socioeconomic (American Meteorological Society (AMS), 1997). Most definitions are related to the water deficit of precipitation, streamflow, soil moisture (Dracup et al., 1980) or groundwater (Hisdal and Tallaksen, 2001), considered individually or in combination.

Drought indices are commonly used as a quantitative measure to summarise and characterise drought properties including duration, intensity and severity (cumulative deficit) (Dracup et al., 1980). For example, McIntyre et al. (2003) considered various drought indices and used the lowest 100-day average river flow each year to measure drought severity in an analysis of the capacity of UK reservoirs. Thornthwaite (1948) proposed an early drought index based on the difference between precipitation and evapotranspiration. The basic criteria used to define drought indices are related to the scale and the applicability of relevant data across space and time (e.g. Heim, 2002). Numerous indices with varying degrees of complexity (e.g. Palmer, 1965) have been proposed by including various climate variables which are either difficult to measure or are not available. However, a drought index derived from accumulated rainfall deficit has broad applicability. It can be applied when other data do not exist (Mawdsley et al., 1994), and, moreover, most types of drought originate from a precipitation deficit (Wilhite and Glantz, 1985; Heim, 2002).

Drought modelling studies are mainly related to either (1) analysis and modelling of drought statistical characteristics, such as the frequency of drought severity and duration (e.g. Phillips and McGregor, 1998), or (2) drought forecasts (e.g. Panu and Sharma, 2002). With respect to the former, we note that an observed weather sequence is only one realisation of the natural process (Richardson, 1981), hence it is very likely that possible extreme droughts of practical importance (e.g. for water resource planning) will not be included in an observed weather series. Therefore, many studies (e.g. Richardson, 1981, Racsko et al., 1991, Cancelliere and Salas, 2004) have proposed statistical models to overcome the shortcomings of finite historical records. In an early study, Gumbel (1954) studied the extreme distribution of droughts based on an analysis of 13 North American rivers. Subsequently, various statistical theories have been proposed to represent the characteristics of different types of drought (e.g. Dracup et al., 1980; McIntyre et al., 2003). In addition, the relationships between drought statistical characteristics and meteorological processes have been investigated, normally with the assumption that the type and parameters of the distributions of the relevant weather factors are independent of the length of the series (Racsko et al., 1991), This type of analysis can readily be extended to evaluate potential effects of changing climate (Jones et al., 2009).

Regarding drought forecasts, the literature is mainly focussed on possible predictors for drought (e.g. Cordery and McCall, 2000). Similarly to the forecasting of floods (e.g. Young 2002) and rainfall (e.g. Onof et al., 1998), drought forecasting is important for the development of early warning systems and response procedures (Panu and Sharma, 2002). Many studies (e.g. Piechota and Dracup, 1996; Cordery and McCall, 2000 and Panu and Sharma, 2002) have found that climate variables and circulation indices (e.g. the Southern Oscillation Index) are useful indicators of the probable timing of drought. Drought characteristics and mitigation measures can also be
investigated under climate change scenarios, using GCM outputs and downscaling relationships between climate variables and drought. The validity of these methods depends of course on the robustness and sensitivity of the empirical relationship between drought and climate variables, but some such assumption is a feature of all statistical methods (e.g. Wilby et al., 1998).

Generally, stochastic studies for either projecting or forecasting drought characteristics adopt one of two approaches: (i) data-driven modelling of drought properties or (ii) rainfall and/or hydrological process modelling based on climate variables as inputs. In the first approach, the drought properties (e.g. drought index, rainfall below a threshold or minimum discharges) are investigated or modelled directly. Many drought probabilistic studies use this kind of black box approach (e.g. Mathier et al., 1992; Sen, 1998), commonly based on the concept of runs, derived from an early definition of statistical point drought in Yevjevich (1967). In the concepts of runs, a run is referred to as a chain of successive similar events. For example, in Sen (1976), a run of droughts is made up of successive negative water deviations from averages. One of main advantages of using runs is that analytical solutions of water deficit length, intensity and cumulative deficit can be determined by assuming certain statistical distributions for variables such as precipitation or flow (e.g. Sen, 1976 and 1977; Rao and Rao, 1986). Using Monte Carlo simulation along with state modelling (e.g. Markov processes), a probabilistic formulation of spatial-temporal drought pattern can be deduced (e.g. Tase, 1978; Sen, 1998). Moreover, droughts can be modelled using point or renewal process (e.g. Abi-Zeid et al., 2004). These models can include external variables which take into account dependent or non-stationary characteristics.

In the approach based on rainfall or hydrological process modelling, synthetic time series (e.g. precipitation or streamflows) are generated from rainfall, runoff or other models, and drought characteristics are derived from these time series realisations. This is a common approach for recent climate studies concerning change in drought characteristics (e.g. Blenkinsop and Fowler, 2007; Lopez-Moreno and Beniston, 2009).

In the UK, historical drought events have long been of interest in climate studies (e.g. Bryant et al., 1992, 1994; Jones et al., 1997; Burt and Shahgedanova, 1998; Phillips and McGregor, 1998). A long-term trend in drought has not been observed in England and Wales based on annual rainfall totals (Marsh et al., 2007), although Blenkinsop and Fowler (2007) noted that considerable uncertainty surrounds future drought characteristics in the British Isles. However, there are important concerns for the intensification of droughts in the UK under future climate (Environment Agency, 2009), and hence a need to develop appropriate tools to evaluate UK drought risk under climate change.

Since droughts generally have close relationships with large scale climate variables (Hoerling and Kumar, 2003), it is likely that the empirical relationships between historical drought sequences and observed large scale climate variables can be estimated using current climate characteristics, which is a technique common to many statistical downscaling approaches (e.g. Wilby et al., 2004). Using such empirical relationships established from the current climate in conjunction with climate projections based on GCMs or RCMs, future drought conditions can be simulated, under the assumption that the relationships retain validity for future climate conditions. Therefore, to support such analyses, there is a need for the relationships between drought characteristics and global or regional climate circulation variables to be better identified for the UK for drought projections. Moreover, the drought associated with rainfall is better defined as a risk (i.e. probabilistically),
rather than simply using a temporally-aggregated mean, to provide a more complete drought representation. Instead of classifying a drought event based on a fixed threshold of severity (e.g. Phillips and McGregor, 1998), using the probability distribution of severity under particular climate conditions appears to characterise better the uncertain nature of droughts.

Although there have been attempts to provide fine scale climate projections for assessment of extremes in the UK (e.g. Kilsby et al., 2007; Jones et al., 2009), the latter authors pointed out that these projections have limited capability in terms of reproducing long dry spells. It is therefore necessary to develop a new framework or tool for probabilistic drought risk to address possible water resource challenges under scenarios of future climate in the UK.

In the present paper, drought severity index (DSI) series for six UK catchments have been investigated, using both data-driven and hydrological modelling approaches. In the first method, the DSI series derived from the observed rainfall are modelled by autoregressive integrated moving average (ARIMA) models. Although ARIMA models are mainly used for forecasting, they have been extensively used to study various drought indices (e.g. Mishra and Desai, 2006, Durdu, 2010). No studies using ARIMA models to represent the DSI have been found in the literature. Moreover, using ARIMA models to study DSI can provide insights of the important climate variables for drought prediction when the autocorrelation of a drought index is considered. For the second approach, the DSI series are computed from rainfall simulated using the Generalised Linear Model (GLM) approach (Chandler and Wheater, 2002), which has been demonstrated to be able to generate adequate rainfall series in a framework that can be applied to future climate states (Chandler et al., 2006; Yang et al., 2005). In the subsequent sections, the data and methodologies are explained, the validation and prediction performance of the two approaches are presented, and finally, results and other considerations are summarised in a concluding discussion.

**Data and Methodology**

Six catchments (Figure 1) with coincident rainfall and streamflow data series were selected for study from a database developed by Young (2000) and previously used by Lee (2006). The selected catchments are distributed over the North West, Midlands and South East of England, i.e. all with an oceanic climate. They are located near to a set of areas for which GLM-based models of daily rainfall were developed by Chandler et al. (2006), and hence the GLM structure developed in that project is considered to be adequate here (Chun et al., 2009a&b). The size of the catchments is in the range of 60 to 250 km² which may be deemed to be small for water resources but is still significant in terms of low flow impacts on local communities and the water environment. Further details of these catchments and the possible changes in rainfall and streamflow series under climate change scenarios can be found in Chun et al. (2009a&b).

As droughts are considered to be closely tied with multiyear persistence of atmospheric circulation patterns (Hoerling and Kumar, 2003) and teleconnections (Bates et al., 2008), atmospheric circulation variables appear to be useful predictors for providing nonstationary climate signals in drought models. Therefore, climate data from the US National Centres for Environmental Prediction (NCEP) reanalysis dataset (Kalnay et al., 1996) and North Atlantic Oscillation (NAO) index data from
the Climatic Research Unit (CRU, 2009) are used as potential external predictors of drought in the
GLM and ARIMA models. NCEP data are selected to calibrate and validate the proposed drought
model because they are contemporaneous gridded data with similar resolution to GCMs. Regarding
future climate variables, the Hadley Centre Hadcm3 GCM data (Gordon et al., 2000) are used for
2080s drought projections under the A2 emissions scenario. Details of the emission scenarios are
explained in Nakicenovic et al. (2000).

DSI

A Drought Severity index (DSI) developed by Phillips and McGregor (1998) is used as a proxy to
understand possible changes in drought characteristics. This DSI is based on the precipitation deficit
concept of Bryant et al. (1992), and has been used to investigate possible changes in UK drought in
several previous studies (e.g. Fowler and Kilsby, 2002 and 2004, Blenkinsop and Fowler, 2007).
Phillips and McGregor (1998) quantify the precipitation deficit by rainfall anomaly ($X_t$) which is the
difference between observation and the average rainfall in month $t$.

Three and six-monthly drought indices are selected, defined as DSI3 and DSI6 respectively. These are
of variable duration, but use three and six month periods to define the drought initiation and
termination criteria, as specified in Fowler and Kilsby (2004) and Blenkinsop and Fowler (2007). The
adopted initiation rule for the n-monthly index (DSIn) is when the consecutive rainfall in an n month
period is lower than the n-monthly mean and $X_t$ is negative. The corresponding termination rule is
when the n-monthly mean has been exceeded. Both DSI3 and DSI6 are standardised by dividing the
absolute value of deficit by the site mean-annual precipitation, i.e. they express the accumulated
deficit for a particular period as a percentage of mean annual precipitation (Phillips and McGregor,
index is more suitable for evaluation of surface water drought in the British Isles whereas a 6+
month index is more appropriate to represent groundwater drought. In this paper, DSI3 and DSI6 are
used to demonstrate the feasibility of modelling the drought index at seasonal or semi-annual scale.

As the total lengths of available rainfall series for the six catchments are different, the monthly
precipitation anomaly is defined in terms of the monthly averages calculated from slightly different
periods. Three catchments have data period from 1961 to 1990 but the Maifold at Ilam (28031), the
Loddon at Sheepbridge (39022) and the Dean at Stanneylands (69008) have shorter records, which
are 1968-1990, 1965-1990 and 1976-1990 respectively. Admittedly, the investigation period is not
very long and it can only realistically be used to study the drought characteristics at seasonal and
annual but not at inter-decadal scale. Nevertheless, the data period includes three UK drought
periods of potential significance: Phillips and McGregor (1998) considered the catchments they
examined in south west England to have a drought of borderline significance in 1983/4; and Byrant
et al. (1994) supposed that south and east England were subjected to drought conditions in
drought was a severe dry episode over their several investigated stations, despite wide spatial
variations.
ARIMA models

Autoregressive integrated moving average (ARIMA) models are commonly employed to analyse drought time series (e.g. Yurekli and Kurunc, 2006; Modarres, 2007, Fernandez et al., 2008). Many previous studies (e.g. Mishra and Desai, 2006) have used ARIMA models to study and forecast drought indices (e.g. Standardised precipitation index (McKee et al., 1993)) directly. Although alternating renewable processes or run theory (e.g. Abi-Zeid et al., 2004) may be good models for the DSI series, the autocorrelation and the specific ‘termination’ rules of DSI appear to be better modelled by ARIMA models. In this study, DSI is modelled using external regressors, after Hyndman (2009). The general multiplicative seasonal ARIMA model (Box et al., 1994) with order \((p,d,q)(P,D,Q)s\)
can be expressed as:

\[
\phi_p(B)\phi_p(B^s)(1-B)^d(1-B^s)^P Y_t = \theta_q(B)\theta_q(B^s)a_t  
\]

(1)

where \(Y_t\) is the observed series and \(a_t\) is the random error at time \(t\). \(p\), \(d\) and \(q\) are the order of non-seasonal autoregression, differencing and moving average respectively, and \(P\), \(D\) and \(Q\) are the corresponding seasonal orders. \(s\) gives the length of the season. \(\phi_p(B)\), \(\phi_p(B)\), \(\theta_q(B)\) and \(\theta_q(B)\) are the polynomial functions of the backshift operator (B).

From Equation (1),

\[
P(B)Y_t = a_t 
\]

(2)

where \(P(B)\) is the polynomial in the backshift operator (B) from ARIMA (i.e.

\[
P(B) = \phi_p(B)\phi_p(B^s)(1-B)^d(1-B^s)^P \phi_q(B^s)^{-1}(B)\phi_q(B^s)^{-1}(B^s)
\]

As possible effects of external climate variables are the interest, dynamic regression to incorporate series of independent variables into ARIMA is performed (e.g. Fernandez et al., 2008), summarised as:

\[
P(B)Y_t = \beta X_t + a_t 
\]

(3)

where \(X_t\) is a vector of external regressors and \(\beta\) is a vector of regression coefficients.

The ARIMA approach is used to model the DSI series in a three-stage Box-Jenkins approach (Pfaff, 2008): identification, estimation and diagnosis. An automated stepwise approach (Hyndman and Khandakar, 2008) based on Akaike’s information criterion (AIC) (1974) is adopted. In any case, identifying ARIMA model structure (i.e. identifying \(p\), \(P\), \(q\), \(Q\), \(d\) and \(D\)) is generally deemed to be subjective (Hyndman and Khandakar, 2008).

The identification and estimation of an ARIMA model with external factors can be done by a one-stage or a two-stage approach. In the former, the order of the ARIMA model and parameters for external regressors can be identified and estimated together (e.g. the ARIMA selection algorithm in Hyndman and Khandakar (2008)). For the two-stage approach, the orders of ARIMA models are identified without external regressors, and the parameters of the ARIMA with external regressors are estimated afterwards. With respect to diagnosis, the suitability of the final model specifications and the ARIMA assumption of independence of the error terms \(a_t\) is validated by examining the autocorrelations and using the Ljung-Box test (Ljung and Box, 1978). In the forecast stage, Kalman
filtering is used to compute point estimates and corresponding standard errors (Durbin and Koopman, 2001). By using recursive Kalman filtering, the drought index series can be specified by a pair of equations for each time $t \geq 1$ with a Gaussian prior distribution $(B_0 \sim N(m_0, C_0))$ (i.e. an initial condition):

\[
Y_t = F_t \theta_t + \nu_t \quad \nu_t \sim N(0, V_{\nu_t})
\]

\[
\theta_t = G_t \theta_{t-1} + \omega_t \quad \omega_t \sim N_p(0, W_{\omega_t})
\]

In the state space model terminology, the first equation is called the observation or measurement equation, and the second equation is called the state equation. $F_t$ is a design matrix and $G_t$ is a transition matrix. $\nu_t$ and $\omega_t$ are two independent Gaussian random series with zero mean and known variance matrices $V_{\nu_t}$ and $W_{\omega_t}$ respectively.

**GLMs**

Generalised linear models (GLMs) as described by Nelder and Wedderburn (1972) provide a feasible general framework for data analysis and simulation. Many studies (e.g. Chandler and Wheater, 2002; Yang et al., 2005 and Furner and Katz, 2007) have shown that the GLM approach can be used to project and downscale meteorological series. In daily rainfall modelling, the GLM approach (Stern and Coe, 1984; Chandler and Wheater, 2002) has been shown to provide realistic rainfall simulations based on various validation approaches, such as seasonal plots (e.g. Stern and Coe, 1984) and residual analysis (e.g. Chandler and Wheater, 2002), and it has been applied to various climatic regions such as in Africa (e.g. Stern and Coe, 1984; Kenabatho et al., 2008), Asia (e.g. Mirshahi et al., 2008), Australia (e.g. Frost, 2007), Europe (e.g. Chandler and Wheater, 2002; Yang et al., 2005) and South America (e.g. Furrer and Katz, 2007). Chandler and Wheater (2002) showed that geophysical variables such as the North Atlantic Oscillation index can be readily incorporated into the GLM rainfall framework. Yang et al. (2005) demonstrated the possibility of simultaneous multisite rainfall simulation using the GLM approach.

Based on the GLM approach in Chandler and Wheater (2002), the daily rainfall model consists of an occurrence and an amounts model. The conditioned occurrence probability is in the form of a logistic regression

\[
\ln \left( \frac{p_i}{1-p_i} \right) = x_i^T \beta
\]

where $p_i$ is the probability of rain for the $i$th day; $x_i$ is a transposed predictor vector and $\beta$ is a coefficient vector.

Given a rainfall event, the expected daily rainfall amount is expressed as a gamma distribution with mean on the $i$th wet day given by

\[
\ln(\mu_i) = \frac{x_i^T \beta}{2}
\]
where $\xi_i^T$ is a transposed predictor vector and $\gamma$ is a coefficient vector.

The daily rainfall model structure adopted here is developed by Leith (2005). It includes seasonal covariates, an autoregressive structure and three external climatic parameters (i.e. sea level pressure, temperature and relative humidity). The general performance of the model is presented in Chun et al. (2009a&b). However, the relationship between water deficit (rainfall anomaly) and duration (persistence), which are the main characteristics of drought (e.g. Mawdsley et al., 1994), were not studied in the previous work (e.g. Chun et al., 2009a&b). Hence, the drought characteristics of the simulated rainfall series from the GLM approach are examined here.

The daily rainfall model structure is first simulated using the GLM approach. The distribution of the DSI3 and DSI6 series are deduced from monthly rainfall series accumulated from 100 GLM-simulated daily series. Robust estimates of the quantiles are possible with 10 simulations of six catchments as the simulated distribution reach stationary for 10 GLM simulations.

Results

The DSI3 and DSI6 values obtained from the six observed records are broadly consistent with the values for Devon and Cornwall in Phillips and McGregor (1998) and for Yorkshire in Fowler and Kilsby (2002). The DSI values are high from 1988 to 1990, especially for the Cole at Coleshill (28066) and the Medway at Chafford Weir (40007), which is consistent with the 1988/1992 drought observed in Bryant et al. (1994). Moreover, the 1975/6 drought is well captured by the DSI for the six catchments.

ARIMA

In the ARIMA analysis, both log transformed (Eq. 5) and non-transformed DSI series are modelled.

$$\text{TDSI} = \log(DSI + \min(DSI) + 1)$$

(5)

where TDSI is transformed DSI and $\min(DSI)$ is the minimum in the DSI series.

In the one-stage ARIMA approach, the parameters for external climate predictors are optimised together with the order selection of the ARIMA model. However, the ARIMA selection algorithm implemented here (Hyndman and Khandakar, 2008) may not give optimum ARIMA models because of early termination of the optimisation process. The estimated sample deviances, which are defined as twice the difference of log likelihood (Nelder and Wedderburn, 1972), can be problematic as a result. Because of limitations of the optimisation process, theoretically impossible negative deviances are obtained when the saturated model (which is a general model with the maximum number of parameters that can be estimated) is compared to the nested simpler ARIMA model. Moreover, there were convergence problems in the one-stage approach for the Medway at Chafford Weir (40007) and the Dean at Stanneylands (68005). It is relatively frequent for failure to complete a run to occur because of numerical difficulties of convergence.

In contrast, the two-stage optimisation appears to be more robust with respect to the model order selection and external predictor identification. The best ARIMA models without external explanatory climate variables are first identified (Hyndman and Khandakar, 2008) based on AIC, and the external
parameters (exogenous variables) are evaluated based on the statistical significance of the likelihood-ratio and the Wald test (c.f. Chatfield, 1995).

As the uncertainty of external parameters from reanalysis data and climate models affects the precipitation forecast, the climate predictors used in the modelling are required to be reliable and provide a signal of climate change (Wheater, 2006). Three climate variables (i.e. temperature, sea level pressure and relative humidity) from the NCEP reanalysis dataset are used as potential predictors because they are considered to be reasonably well represented by the climate models (Leith, 2005). Although other climate variables (e.g. wind speed) may be considered, their reliability is less certain as the spatial dependence and heterogeneities become concerns for these variables and the regional features need to be modelled to reveal these (See Yan et al. (2002) for similar discussion). In addition to the three climate variables, Chandler and Wheater (2002) noted that the NAO index can influence seasonal rainfall structure. Hence the NAO index is also included as a potential exogenous predictor for drought forecast even though it is a direct measure of sea level pressure difference between two locations.

The orders of the derived ARIMA models are summarised in Table I. As subjectivity is an intrinsic issue in order selection (Hyndman and Khandakar, 2008) and ARIMA model specification, the differences between the final ARIMA models obtained for the transformed and non-transformed DSI series are not unexpected. Using likelihood ratio tests, relative humidity is not found to be a significant external climate variable for droughts (for both DSI3 and DSI6). Conversely, concurrent sea level pressure is a significant variable according to the likelihood ratio test for all catchments. With respect to temperature and the NAO index, the selection is more complicated. The likelihood ratio tests are not sufficient to decide on the inclusion of these as external variables. Hence, the Wald statistic (Dobson, 1990) is used to select the external climate variables in the ARIMA models (Table II). The inclusion of temperature and the NAO index is catchment specific, and is difficult to generalise. For the DSI3 models, one catchment has the NAO index as an explanatory variable, one catchment has temperature as an explanatory variable, and there are four catchments with only sea level pressure as the external climate variable. Regarding the DSI6 models, sea level pressure is the significant climate variable for all catchments. Generally, fewer climate variables are required to explain the variance in DSI6, because the use of longer rainfall records to calculate the series smooths the variability and the autocorrelation structure of the drought index explain more variance in the series. Nevertheless, the final selected external climate variables are largely the same for both transformed and non-transformed DSI3 series, and they are fairly similar to the DSI6 series.

Although the physical reasons for a particular catchment having specific climate variables need further study, the signs of the explanatory variables in the ARIMA provide an interesting indication of the relationship between the DSI series (drought) and climate variables. All of the estimated coefficients for sea level pressure are negative but the coefficients for temperature and the NAO index are positive. This implies that drought severity increases when the temperature and the NAO index decrease and the sea level pressure increases. The pattern is consistent with the relationship between the NAO index, rainfall and temperature over northern Europe presented in Hurrell (1995) and Shorthouse and Arnell (1997).

Although the NAO index is derived from sea level pressure, it is a teleconnection variable indicating simultaneous variation in climate over a large area of the earth’s surface (Potter et al., 2004)
whereas sea level pressure provides information about regional variation. This may explain why both the NAO index and sea level pressure are significant parameters for the DSI3 model for Dean at Stanneylands (69008) despite the inter-connection between the NAO index and sea level pressure.

The final ARIMA models were checked with diagnosis plots. Figure 2 presents the histograms of the residuals of the Manifold at Ilam (28031) and provides grounds for using the log transform because of the improvement in the normality of the residuals. Figure 3 shows the diagnostic plots of the transformed DSI3 for the Cole at Coleshill (28066), which indicate that the ARIMA assumptions (i.e. normality and uncorrelatedness of residuals) are generally valid. Generally, the diagnostic plots for other catchments are similar except for the Weaver at Audlem (68005) which requires exclusion of sea level pressure to achieve an adequate Ljung-Box statistic.

Regarding independence assumptions based on the Ljung-box statistics, the identified ARIMA model may be valid only when the lag is less than 6 months. The Ljung-Box statistic shows that the independence assumption for the DSI3 series cannot be rejected for the maximum number of lags (i.e. more than 20 months) but holds for only up to around 6 months for the DSI6 series of the Loddon at Sheepbirdge (39022), Medway at Chafford Weir (40007) and Dean at Stanneylands (69008), i.e. the independent assumption of residuals for DSI6 breaks down when the lags between DSI6 residuals are longer than half a year. The reason may be that the DSI6 series are derived from longer duration rainfall totals, with associated higher smoothing. As a result, the forecast lead time for the DSI6 ARIMA is limited to 6 months, the range of validity of the assumption of residual independence.

The actual, fitted and forecast DSI3 and DSI6 values using the transformed ARIMA models are shown in Figure 4. The 1990s forecast band in Figure 4 is the result of using a Kalman filter in a state space model framework (Commandeur and Koopman, 2007). The expected values and variance for 12 months in the 1990s are estimated recursively by Kalman filter using the DSI values and Kalman gain of December 1989. Generally, the fitted curves represent the observed DSI well (The Pearson correlation is around 0.8). The ARIMA model based on the non-transformed DSIs give better fitted curves. However, the ARIMA models based on the transformed DSIs provide better confidence intervals, bounding the DSI series during the prediction period.

Despite the high goodness of fit in the training period, the quality of forecast deteriorates rapidly with the increase of the forecast lead time and depends on the initial DSI in particular weather conditions for Kalman filter. Figure 5 shows the variance of the forecast level from the Kalman filter. The variance of forecast level increases quickly for lead times larger than one month. The ARIMA models of the DSI series can only be used for drought warning forecasting if an adequate initial drought condition is available. In the next section, we shall examine whether the GLM is able to provide a comprehensive probabilistic representation of drought.

GLM

As noted above, the GLM structure adopted for daily rainfall modelling was developed by Chandler et al. (2006). Based on the results for a range of stations over the UK from Leith (2005), sea level pressure, temperature and relative humidity were used as the driving regional climate variables of the adopted GLM structure. Daily temporal memory effects of the six catchments are modelled by
the preceding 3-day rainfall pattern and seasonal effects are modelled by a 2 parameter Fourier series representation. Details of the model are summarised in Table III. The model structure was tested and found to be suitable for the six catchments by Chun et al. (2009a&b). The significant atmospheric variables for the ARIMA model and the GLM are different because the monthly drought index modelled by the ARIMA model and daily rainfall modelled by the GLM are related to different temporal scales. At a monthly scale, temperature and relative humidity become insignificant for drought projection although they are important for daily rainfall simulation.

Figure 6 provides the distribution of DSI3 and DSI6 derived from the GLM simulated daily rainfall. Generally, the simulated distribution of DSI3 provides good representation of the observed DSI series. The simulated distribution of DSI6 also gives a fairly good representation of the observed drought characteristic, apart from some poor fits for the Medway at Chafford Weir (40007) in 1979 and Weaver at Audlem (68005) in 1965. The deficiencies in the GLM-simulated DSI6 distributions may be a result of imperfect modelling of consecutive drought years spanning two wet seasons, which is not likely to happen to DSI3. Modelling inter-annual variability appears to be challenging for the current approach as only 3 climate variables used as exogenous parameters provide an inter-annual signal to the model. The main problem in studying inter-annual characteristics for the proposed approach is the limitation of the length of available data because long-term records are required (Barring and Fortuniak, 2009).

In Figure 7, the quality of models is assessed using quantile-quantile (QQ) plots between the observed DSI and the DSI calculated from the simulated rainfall driven by the NCEP data. Generally, the QQ plots of DSI6 are better than DSI3. The performance of the model starts to degrade for the DSI3 above 25 even though this represents only a small proportion (less than 6.5%) of the observed DSI3 values. Overall, the approximate straight lines in the QQ plots show that the models provide acceptable results, albeit with some overestimation and underestimation of the high drought index values which varies between catchments. As the DSI3 results of the GLM simulated distribution are generally acceptable, the drought index distributions in 2080s are simulated using GLM simulated rainfall driven by the GCM data (Hadcm3).

Although using a threshold for DSI can help interpret the drought characteristics of the DSI series (Blenkinsop and Fowler 2007), valuable information from DSI values below the threshold may be lost in this approach. A probabilistic approach is appealing as an alternative to the hard threshold approach (e.g. over 10% of mean annual rainfall). Instead of assessing drought duration and severity separately, the drought index distribution derived from the GLM approach provides a more comprehensive representation of the probabilistic changes in characteristics and magnitude across the simulation period. Figure 8 shows the simulated distribution of DSI in the 2080s (from 100 realisations) and indicates that the drought events are more clustered and intermittent than the events in the 1980s which are shown in Figure 6. Using information theory (Shannon, 1948), the degree of clustering can be quantified by entropy ($H(x)$). The entropy is expressed as

$$H(x) = \sum_{i=1}^{n} p(x_i) \log(p(x_i))$$

where $p(x_i)$ is probability (proportion) of $x_i$. Generally, higher entropy means less certainty when an event with happen. For the 12,000 simulated DSI values, the maximum entropy is 9.4 when the DSI values are constant. In the six catchments, the entropy of the simulated 1980s DSI3 values is around 8.5 whereas the entropy of the simulated 2080s DSI3 values is around 8.6. This signifies that the
inter-annual drought will happen in a more uncertain manner in the 2080s. Although the entropy gives a quantitative measure of uncertainty, it is not an absolute measure. It depends on scale and the size of the kernel width. Therefore, investigating how the DSI series varies with different scales and periods using entropy to explore the stationarity of DSI series is worthy of further investigation.

Using the expected quantiles of drought index as indicators, the changes to the different frequencies of drought events can be studied. The expected quantiles of drought index and their standard error can be estimated by integrating over the simulated quantile curves and dividing by the integrated period. In Table IV, the expected 10th, 50th and 90th quantiles of DSI3 calculated from the NCEP data (1980 -1989) and the Hadley Centre model data (2080-2089) are summarised. The expected NCEP 50th quantiles are consistent with the observed DSI3. The expected Hadley model 10th and 50th quantiles are higher than the expected NCEP 10th and 50th quantiles respectively, but there is no statistically significant difference between them. Concerning the 90th quantiles, the difference between the NCEP and Hadley distribution is even smaller. It should be noted that the results here related to high percentiles (e.g. 90th quantiles) have generally less confidence as the corresponding standard errors are usually higher. Overall, the severity of the low and median drought events of DSI3 is generally higher but the changes in the 90th extreme severity of DSI3 are relatively minor.

Conclusions

The non-stationary drought characteristics of six UK catchments have been investigated by two types of stochastic model. In the first approach, ARIMA models are employed to analyse drought using the DSI series directly. Although these models are site specific, sea level pressure has been identified as the most important external parameter whereas, for all the studied catchments, relative humidity is not a crucial exogenous parameter for drought forecasting. It is interesting to note that when simple regression between the DSI series and climate variables has been performed to investigate the relationship between drought and climate variables, relative humidity is the principal significant factor for the DSI series. It is not surprising that relative humidity is low during a drought. However, relative humidity is also highly correlated to the previous month’s DSI values, and it becomes an insignificant variable because the autoregression in the ARIMA model explains the variance of relative humidity in the DSI series.

On the whole, the ARIMA approach may underestimate the magnitude of drought severity (such as the 1975/6 drought in Figure 4) but it provides fair one month ahead fit (i.e. short term forecasts) if a suitable current state is provided. The quality of DSI forecast using ARIMA depends on the initial states and number of steps ahead. The uncertainty of the predictions increases with number of steps ahead as shown by the increase in the variance of the forecasts. Modelling the transformed DSI index is recommended for the ARIMA approach because the transformation increases the normality of the residuals. Based on the Ljung-Box statistics, a forecast of longer than 6 months ahead by the ARIMA method is not recommended for the six catchments because the independent assumption is not valid.

Using the rainfall process model, the GLM approach can simulate rainfall for drought index series estimation. Although there is room for improvement in the persistence of rainfall simulation for more than 6 months, the good DSI3 results show that the GLM approach is suitable for local drought assessments at a seasonal scale. The limitation of the GLM DSI6 result is very likely related to the
study period here because of inter-decadal variation of drought characteristic, the used records (although they are 30 years long) may not be enough to estimate annual average adequately. The DSI values for the six catchments are mildly sensitive to annual average calculated from different control periods. Based on the GLM approach, the expected median and lower extreme of severity will probably be higher in the 2080s, although the 90th extreme severity of DSI3 may not be more severe (It should be noted that the result for the high percentiles (e.g. 90th) may not be reliable because of high uncertainty of these results). GLMs thus provide a useful tool for regional drought study across the British Isles, contingent on the performance of the underlying climate models. Moreover, the GLM approach can also be used for forecasting, by carrying out sequential simulations, i.e. simulations based only on the past and by using forecasts of the explanatory variables.

The main limitation of the current work is that only a three decade drought series have been examined. The inter-decadal characteristics of persistent drought have not been investigated. Although it is not likely that the ARIMA approach can have the skill for long term drought persistence prediction, the GLM approach appear to be applicable for studying long drought series if long observations series are available for model calibration. Moreover, spatial drought has also not been studied here. Further work, similar to Lloyd-Hughes (2002), for exploring inter-decadal and fine spatial drought is needed.

Apart from the ARIMA and GLM approaches, other stochastic models may be suitable for DSIs. For example, point and renewal processes (e.g. Abi-Zeid et al., 2004) and recursive neural networks (Mishra and Desai, 2006) may be interesting models providing a better representation for the physical mechanisms of DSIs. Nevertheless, the ARIMA approach is useful for a future drought forecast system and the probabilistic predictions provided by the GLM approach give distinct pictures of possible changes in drought under future climate scenarios and facilitate risk-based water management (e.g. McIntyre et al., 2003).

Acknowledgements
The first author is supported by a research studentship from the Grantham Institute for Climate Change Research at Imperial College. Access to the catchment data was kindly provided by A. Young of CEH Wallingford.


Lee, H.S., 2006. *Regionalisation of rainfall-runoff models in the UK*, Imperial College, the University of London.


Young, A., 2000. *Regionalisation a daily rainfall-runoff model with the United Kingdom*, University of Southampton, UK.


Figure 1 Locations of six catchments
Figure 2 Histograms of the residuals of the Manifold at Ilam (28031) show the normality of residuals. The distribution of residual of transformed DSI3 is more symmetric and normal.
Figure 3 Diagnostic plots of the transformed DSI3 for the Cole at Coleshill (28066) show that the independent assumptions of ARIMA are satisfactory. The blue lines of empirical autocorrelation function (ACF) plot for residual shows the asymptotic 95% interval bounds. As a rough test, the autocorrelation is deemed to be violated if the empirical autocorrelation of residuals exceed the blue lines.

The p values for Ljung-box statistic are used to test the independence between residuals. The null hypothesis of the Ljung-box statistic is that the residuals are independent, with an approximation that the degrees of freedoms are simply lag. Therefore, for the estimated p value below the blue line (95% significant level) on the Ljung-Box statistic diagnostic plot, the independence assumption is considered to be invalid.
Figure 4 The fitted and forecasted values of the transformed DSI3 and DSI6 using the ARIMA model. The thick black lines show the observed values. The red lines are the fitted curves in the calibration period calculated from one step ahead point estimates. The confident bands in 1990 (forecast period) are computed from point estimates and standard errors with the Gaussian assumption. The bands correspond to the 5th, 10th, 25th, 50th, 75th, 90th and 95th percentiles of 12 month forecast starting from January 1990.

a) DSI3
b) DSI6

28031

28066

39022
Figure 5: The forecast level variance of the Manifold at Ilam (28031) from the Kalman filter
Figure 6: The drought index and the modelled distribution driven by the NCEP data illustrate the probabilistic characteristics of drought in the 1980s based on the proposed GLM approach. The bands correspond to the 5th, 10th, 25th, 50th, 75th, 90th and 95th percentiles and the thick lines show the observed values.

a) DSI3

![DSI3 (28031)](image1)

![DSI3 (28066)](image2)

![DSI3 (39022)](image3)

![DSI3 (40007)](image4)
b) DSI6

DSI6 (28031)

Year


DSI

0 20 40 60

DSI6 (28066)

Year


DSI

0 20 40 60

DSI6 (39022)

Year


DSI

0 20 40 60

DSI6 (40007)

Year


DSI

0 20 40 60
Figure 7: The QQ plots show the GLM simulated drought index driven by the NCEP data against the observed index.
Figure 8: The drought index and the modelled distribution driven by the Hadley data show the more clustered and intermittent drought characteristics in the 2080s. The bands correspond to the 5th, 10th, 25th, 50th, 75th, 90th and 95th percentiles and the thick lines show the observed values.
Table I

The ARIMA models are summarised by the notation of \((p,d,q)(P,D,Q)\) in Eq. 1 where \(p, d\) and \(q\) are the order of non-seasonal autoregression, differencing and moving average respectively, and \(P, D\) and \(Q\) are the corresponding 12-month seasonal orders.

### The ARIMA models for DSI3

<table>
<thead>
<tr>
<th>Location</th>
<th>ARIMA Model</th>
<th>ARIMA Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manifold at Ilam (28031) (1968-1989)</td>
<td>Non-transformed ((2,0,0)(0,0,0))</td>
<td>Transformed ((1,0,2)(0,0,0))</td>
</tr>
<tr>
<td>Cole at Coleshill (28066) (1961-1989)</td>
<td>((1,0,1)(0,0,0))</td>
<td>((1,0,0)(0,0,0))</td>
</tr>
<tr>
<td>Loddon at Sheepbridge (39022) (1965-1989)</td>
<td>((1,0,2)(0,0,0))</td>
<td>((2,0,1)(1,0,0))</td>
</tr>
<tr>
<td>Medway at Chafford Weir (40007) (1961-1989)</td>
<td>((1,0,0)(0,0,0))</td>
<td>((2,0,0)(0,0,0))</td>
</tr>
<tr>
<td>Weaver at Audlem (68005) (1961-1989)</td>
<td>((2,0,0)(0,0,0))</td>
<td>((2,0,0)(0,0,0))</td>
</tr>
<tr>
<td>Dean at Stanneylands (69008) (1976-1989)</td>
<td>((1,0,1)(0,0,0))</td>
<td>((2,0,1)(0,0,0))</td>
</tr>
</tbody>
</table>

### The ARIMA models for DSI6

<table>
<thead>
<tr>
<th>Location</th>
<th>ARIMA Model</th>
<th>ARIMA Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manifold at Ilam (28031) (1968-1989)</td>
<td>Non-transformed ((1,1,1)(0,0,0))</td>
<td>Transformed ((1,0,0)(0,0,0))</td>
</tr>
<tr>
<td>Cole at Coleshill (28066) (1961-1989)</td>
<td>((1,0,1)(0,0,0))</td>
<td>((2,0,0)(0,0,0))</td>
</tr>
<tr>
<td>Loddon at Sheepbridge (39022) (1965-1989)</td>
<td>((2,0,0)(0,0,0))</td>
<td>((1,0,0)(0,0,0))</td>
</tr>
<tr>
<td>Medway at Chafford Weir (40007) (1961-1989)</td>
<td>((2,0,0)(0,0,0))</td>
<td>((1,0,0)(0,0,0))</td>
</tr>
<tr>
<td>Weaver at Audlem (68005) (1961-1989)</td>
<td>((1,0,0)(0,0,0))</td>
<td>((1,0,0)(2,0,1))</td>
</tr>
<tr>
<td>Dean at Stanneylands (69008) (1976-1989)</td>
<td>((1,0,0)(0,0,0))</td>
<td>((1,0,0)(0,0,0))</td>
</tr>
</tbody>
</table>

---

833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852
Table II

The significant external climate variables based on the likelihood-ratio and the Wald test at a 95% level.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Sea level pressure</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Temperature</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NAO Index</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sea level pressure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Temperature</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NAO Index</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Sea level pressure is excluded from the Weaver at Audlem (68005) for the final DSI3 model because of Ljung-Box tests.
Table III

a) Covariates used in the GLM occurrences model

The listed components are the predictors of the logistic regression in Eq.3. \( I(.) \) is an indicator function that takes the value of 1 if the equality inside the function is true. \( \text{Ln}(.) \) is an logarithm transformation. \( Y[t-d] \) is the lag d rainfall amount.

<table>
<thead>
<tr>
<th>Components</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>1</td>
</tr>
<tr>
<td>Sea level pressure</td>
<td>2</td>
</tr>
<tr>
<td>Temperature</td>
<td>3</td>
</tr>
<tr>
<td>Relative humidity</td>
<td>4</td>
</tr>
<tr>
<td>Index</td>
<td>5</td>
</tr>
<tr>
<td>( I(Y[t-1]&gt;0) )</td>
<td>6</td>
</tr>
<tr>
<td>( I(Y[t-2]&gt;0) )</td>
<td>7</td>
</tr>
<tr>
<td>( I(Y[t-3]&gt;0) )</td>
<td>8</td>
</tr>
<tr>
<td>( I(Y[t-k]&gt;0;\ k=1\ to\ 2) )</td>
<td>9</td>
</tr>
<tr>
<td>Daily seasonal effect, cosine component</td>
<td>10</td>
</tr>
<tr>
<td>Daily seasonal effect, sine component</td>
<td>11</td>
</tr>
<tr>
<td>Smooth February effect</td>
<td>12</td>
</tr>
<tr>
<td>( \text{Ln}(1+Y[t-1]) )</td>
<td></td>
</tr>
<tr>
<td>2-way interaction: covariates 2 and 9</td>
<td></td>
</tr>
<tr>
<td>2-way interaction: covariates 2 and 10</td>
<td></td>
</tr>
<tr>
<td>2-way interaction: covariates 2 and 11</td>
<td></td>
</tr>
<tr>
<td>2-way interaction: covariates 3 and 9</td>
<td></td>
</tr>
<tr>
<td>2-way interaction: covariates 3 and 10</td>
<td></td>
</tr>
<tr>
<td>2-way interaction: covariates 3 and 11</td>
<td></td>
</tr>
<tr>
<td>2-way interaction: covariates 1 and 5</td>
<td></td>
</tr>
<tr>
<td>2-way interaction: covariates 4 and 5</td>
<td></td>
</tr>
<tr>
<td>2-way interaction: covariates 4 and 6</td>
<td></td>
</tr>
<tr>
<td>2-way interaction: covariates 12 and 9</td>
<td></td>
</tr>
<tr>
<td>2-way interaction: covariates 12 and 10</td>
<td></td>
</tr>
<tr>
<td>2-way interaction: covariates 12 and 11</td>
<td></td>
</tr>
</tbody>
</table>
b) Covariates used in the GLM amount model

The listed components are the predictors of the amounts model in Eq.4. $I(.)$ is an indicator function that takes the value of 1 if the equality inside the function is true. $\ln(.)$ is an logarithm transformation. $Y[t-d]$ is the lag $d$ rainfall amount.

<table>
<thead>
<tr>
<th>Components</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>1</td>
</tr>
<tr>
<td>Sea level pressure</td>
<td>2</td>
</tr>
<tr>
<td>Temperature</td>
<td>3</td>
</tr>
<tr>
<td>$5\ln(1+Y[t-1])$</td>
<td>4</td>
</tr>
<tr>
<td>$\ln(1+Y[t-2])$</td>
<td>5</td>
</tr>
<tr>
<td>Daily seasonal effect, cosine component</td>
<td>6</td>
</tr>
<tr>
<td>Daily seasonal effect, sine component</td>
<td>7</td>
</tr>
<tr>
<td>$I(Y[t-k]&gt;0: k=1$ to $2)$</td>
<td>8</td>
</tr>
<tr>
<td>2-way interaction: covariates 1 and 5</td>
<td></td>
</tr>
<tr>
<td>2-way interaction: covariates 1 and 6</td>
<td></td>
</tr>
<tr>
<td>2-way interaction: covariates 3 and 5</td>
<td></td>
</tr>
<tr>
<td>2-way interaction: covariates 3 and 6</td>
<td></td>
</tr>
<tr>
<td>2-way interaction: covariates 1 and 3</td>
<td></td>
</tr>
<tr>
<td>Dispersion Parameter</td>
<td></td>
</tr>
</tbody>
</table>

Table IV

The expectation and variance of the quantiles of DSI3

<table>
<thead>
<tr>
<th>Obs</th>
<th>NCEP</th>
<th></th>
<th></th>
<th></th>
<th>Hadley</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10</td>
<td>50</td>
<td>90</td>
<td>10</td>
<td>50</td>
<td>90</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\mu$</td>
<td>se</td>
<td>$\mu$</td>
<td>se</td>
<td>$\mu$</td>
<td>se</td>
<td>$\mu$</td>
<td>se</td>
</tr>
<tr>
<td>28031</td>
<td>3.8</td>
<td>5.0</td>
<td>0.3</td>
<td>1.1</td>
<td>3.6</td>
<td>4.5</td>
<td>11.6</td>
<td>7.1</td>
</tr>
<tr>
<td>28066</td>
<td>3.8</td>
<td>4.9</td>
<td>0.4</td>
<td>1.3</td>
<td>4.2</td>
<td>4.7</td>
<td>12.8</td>
<td>7.6</td>
</tr>
<tr>
<td>39022</td>
<td>3.8</td>
<td>4.9</td>
<td>0.7</td>
<td>2.1</td>
<td>4.7</td>
<td>5.5</td>
<td>14.5</td>
<td>9.2</td>
</tr>
<tr>
<td>40007</td>
<td>5.7</td>
<td>6.4</td>
<td>0.6</td>
<td>1.9</td>
<td>5.0</td>
<td>6.2</td>
<td>15.3</td>
<td>11.1</td>
</tr>
<tr>
<td>68005</td>
<td>4.7</td>
<td>5.3</td>
<td>0.4</td>
<td>1.4</td>
<td>3.9</td>
<td>4.7</td>
<td>11.9</td>
<td>6.6</td>
</tr>
<tr>
<td>69008</td>
<td>4.0</td>
<td>5.1</td>
<td>0.5</td>
<td>1.4</td>
<td>3.4</td>
<td>4.4</td>
<td>11.7</td>
<td>7.2</td>
</tr>
</tbody>
</table>