General response to Reviewer 2

We would like to thank Reviewer 2 for their careful reading and constructive comments. These comments will certainly help to increase the clarity of the manuscript. Most of the reviewer's comments are related to requests for additional explanation regarding the methods and uncertainty in the results. In this general response to the comments, we provide more explanatory examples and improve the description of the methods and results addressing all major comments of Reviewer 2. The suggested modifications of this general response will be included in the new version of the manuscript. Minor reviewers comments are not considered herein as these can be accommodated straightforwardly (by rephrasing or by adding points of minor clarification) in a revised manuscript.

Comment 1

> In the methods/discussion please add some comment on the choice of gamma distribution used to calculate SPI. Other studies have shown that this is often not the most appropriate distribution for precipitation data and it would be good to discuss the impacts of this (see Svensson et al. 2017 for example).

We thank Reviewer 2 for their comment. We now tested the alternative distributions for the SPI calculation and will include a comment on the choice of gamma distribution used to calculate the SPI. We understand that uncertainties are introduced when fitting a gamma distribution to precipitation totals (Stagge et al. 2015; Svensson et al. 2017; McKee et al. 1993). This uncertainty will be addressed more explicitly in the next version of the manuscript by improving the phrasing in L115-117. However, we would like to emphasise that the SPI is primarily used in combination with the SGI to find the optimal correlation between the SPI and SGI. For this correlation, primarily long accumulation periods ($>$ 12months) of the SPI were used (see the mean of optimum accumulation periods in L214:218 for near-natural wells and the range of accumulation periods in S3 in the supplementary material of the current manuscript). Considering the use of long accumulation periods, the 'best' fitting distribution varies (Svensson et al. 2017). High rejection rates are found for multiple distributions (Stagge et al. 2015), which suggests we need to test which distribution performs best.

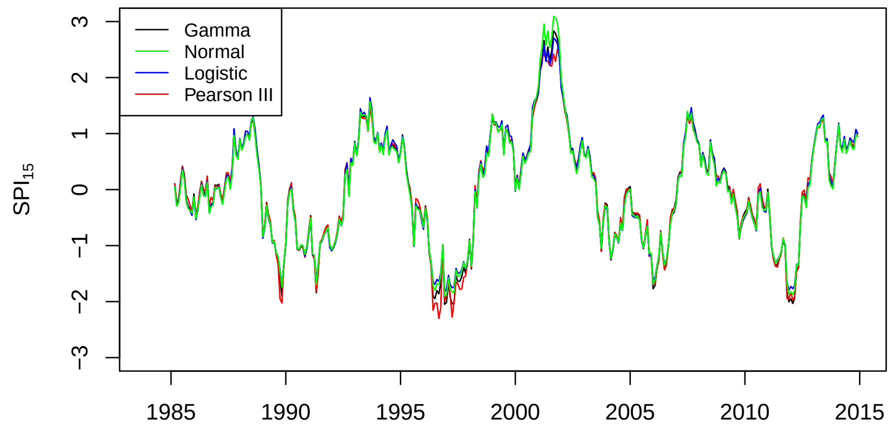
We have tested different distributions on a subset of the total dataset (45 precipitation grids matching to groundwater monitoring sites in the Chilterns). Three alternative distributions were tested: Normal, Pearson III, and Logistic distribution and results are presented in a similar way to Figure 5d in Svensson et al. (2017) (Figure 1). Figure 1 shows an example of a SPI_{15} in which a slight variation is seen in the calculated SPI values during droughts using different distributions. This variation did, however, not result in better or worse SPI_Q-SGI correlation. For the subset of the data (45 monitoring sites), the range of correlations using the Gamma distribution was 0.41-0.89 with a mean of 0.794. The mean of the calculated correlation remained the same when using different distributions (Normal, Pearson III, and Logistic distribution). The range of the 45 correlations showed minimal changes compared to Gamma distribution (0.41-0.89): 0.40-0.89 (Normal & Logistic), 0.40-0.90 (Pearson III). Reviewing the minimal change in SPI_Q-SGI correlation, we think that the use of alternative distributions instead of the current distribution (Gamma) would not change the results of this study given the use of the SPI only in the correlation analysis.

Comment 2

> From the methods section, it seems that you compare the SPI from a single grid cell with the corresponding groundwater well location (this should be clarified in the text). It would be good to add to the discussion the impacts of comparing a 1km2 grid cell of SPI with SGI that is a product of a regional groundwater aquifer system and regional rainfall patterns.

We agree with Reviewer 2 that the phrasing could be improved regarding the SPI and SGI comparison. In the comparison, we have indeed used the location of a groundwater monitoring sites with a single cell of the GEAR dataset ($1km^2$). We will rephrase L114-115 to clarify this.

We understand the concern raised by Reviewer 2 questioning our comparison of a point measurement (gridded precipitation, $1km^2$) to the result of regional groundwater recharge (groundwater level observation). However, we would like to emphasize



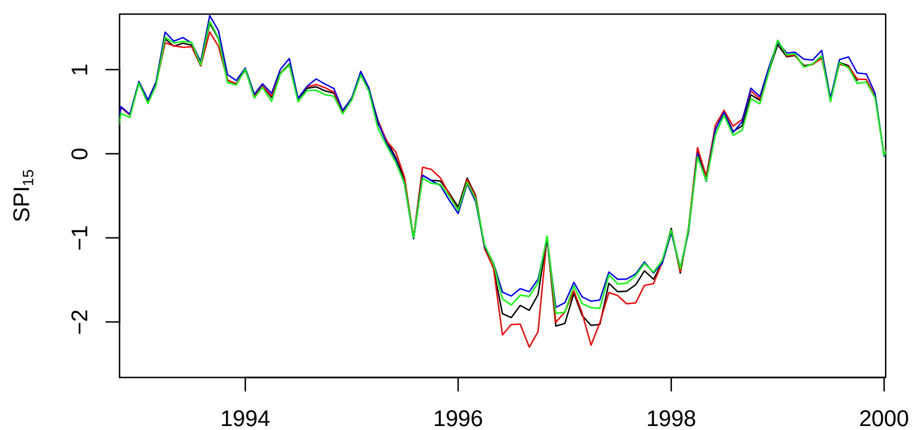


Figure 1. SPI_{15} computed using different distributions for a precipitation estimate located in the Chilterns (the location corresponding to groundwater site (SP90.27).

that the regional extend of groundwater recharge varies and that the precise extent of this recharge area associated with a given observation borehole is usually unknown. In contrast to surface water boundaries, there is no consistent source of information regarding the recharge area for groundwater monitoring sites in either the Hydrometric Register (Marsh and Hannaford 2008) or the water management units. The unknown recharge area is a common uncertainty for groundwater studies and other studies have either used a regional aggregate to overcome this unknown recharge area (Haas et al. 2018) or used a point-scale analysis under the assumption that the influence of precipitation is largest surrounding the groundwater monitoring site (Bloomfield and Marchant 2013; Bloomfield et al. 2015; Li and Rodell 2015; Kumar et al. 2016). Even though a regional precipitation product would potentially results in a more accurate correlation, it could also introduce larger uncertainties given the unknown recharge area. In addition to this, high correlations between SPI_Q-SGI have been previously obtained using the point-based precipitation estimates in different climate regions and by different authors (Bloomfield and Marchant 2013; Bloomfield et al. 2015; Li and Rodell 2015; Kumar et al. 2016). Also in our study, high correlations are found for near-natural wells and the majority of the groundwater monitoring wells [L212:213 and L226:229]. Therefore, considering the unknown recharge area and the reasonable results, consistent with previous studies using point-based precipitation estimates, we don't propose to modify our methodology. However, we will include an extended justification for our approach in the next version of the manuscript. We agree with Reviewer 2 that this is an interesting issue, which might need further investigation in future studies. This would, however, require more detailed information and potentially fewer sites to quantify the associated uncertainty with the current method.

Comment 3

> The methods (in places) were not clear – in particular, the SPI_Q-SGI correlations and the use of the near-natural wells, uninfluenced and influenced monitoring sites. It would be useful to have a worked example of how the SPI-SGI correlations work in practice (showing an example for two sites – one influenced and one non-influenced and how they compare to the near natural reference cluster). It would also be useful to have a map of the influenced and non-influenced wells (this is maybe already included in Figure 1 but this figure is quite busy so it is hard to tell!)

In response to Comment 3, we selected four wells to show the SPI_Q-SGI correlations in practise (Figure 2). The first site is a reference groundwater site (Aylesbury). The second site is an uninfluenced site in the paired water management unit (Lincolnshire) and the last two sites are influenced monitoring sites in the same water management unit (Lincolnshire). The reference well is one of the Index wells in the Hydrologic Register and is representing near-natural conditions. This Index well is included in the reference cluster for Lincolnshire water management unit (see Figure 1 in the current manuscript). For this reference cluster, the lowest correlation is 0.75 (L252) and we assume that monitoring sites with a similar or higher correlation are relatively uninfluenced. The monitoring site in the second panel is thus considered uninfluenced, as the SPI_Q-SGI correlation is 0.831 using the optimal accumulation period (17 months). The last two panels show monitoring sites that are considered to be influenced, as the correlations are 0.561 and 0.566 for these wells, which is below the reference cluster threshold of 0.75. The SGI hydrographs of both wells are remarkably different and don't synchronise well with the long-term SPI_Q (dotted in Figure 2).

We agree with Reviewer 2 that it would be very interesting to show the spatial patterns of detected influenced wells. In earlier versions of the manuscript, we considered showing the locations of a few influenced wells, but explaining spatial pattern of influenced wells requires detailed knowledge regarding the hydrogeological setting of each monitoring well and knowledge of the use of abstraction wells close by. This is because the cone of influence of an abstraction well highly depends on the storage and transmissivity of the aquifer in that particular location, dominant groundwater flow, distance from the abstraction well, volume and timing of abstraction, and possible flow barriers between abstraction wells and monitoring wells. It is possible to analytically calculate the cone of influence for each groundwater monitoring site using a detailed groundwater model designed to replicate the local hydrogeology in detail and using (historical) groundwater abstraction records. Given the unknown use of groundwater abstraction wells (abstraction records were unavailable for the study), it is impossible to deduct the cone of influence of abstraction wells. In other words, we agree that it would be interesting to investigate the spatial patterns of influenced wells, but that goes beyond the scope of the research and would not be possible with current data availability.

> [Comment 4 follows after 5a + 5b]

Comment 5a

> Like Reviewer 1, I am somewhat sceptical of attributing the shorter droughts in Lincolnshire, Shropshire and the Chilterns to water use and/or hotter Summers. Firstly the years that were identified in L263-265 did not have particularly hot summers (or this is certainly not consistent for these years) [continued]

We understand the concern raised regarding the attribution of shorter droughts to increased water use by Reviewer 2 and earlier by Reviewer 1 (comment 4). The timing of observed low groundwater levels and droughts differ indeed slightly for the four water management units and the information regarding the reported drought impacts and observations do not directly refer to these water management units. We will rephrase L267-269 to highlight the literature that we used to hypothesise increased water use because of the exceptional weather. For example, during the first drought event (1988-93), it was found that evapotranspiration was exceptionally high with hot summers for some years (Durant et al. 2015). Mandatory water use restrictions

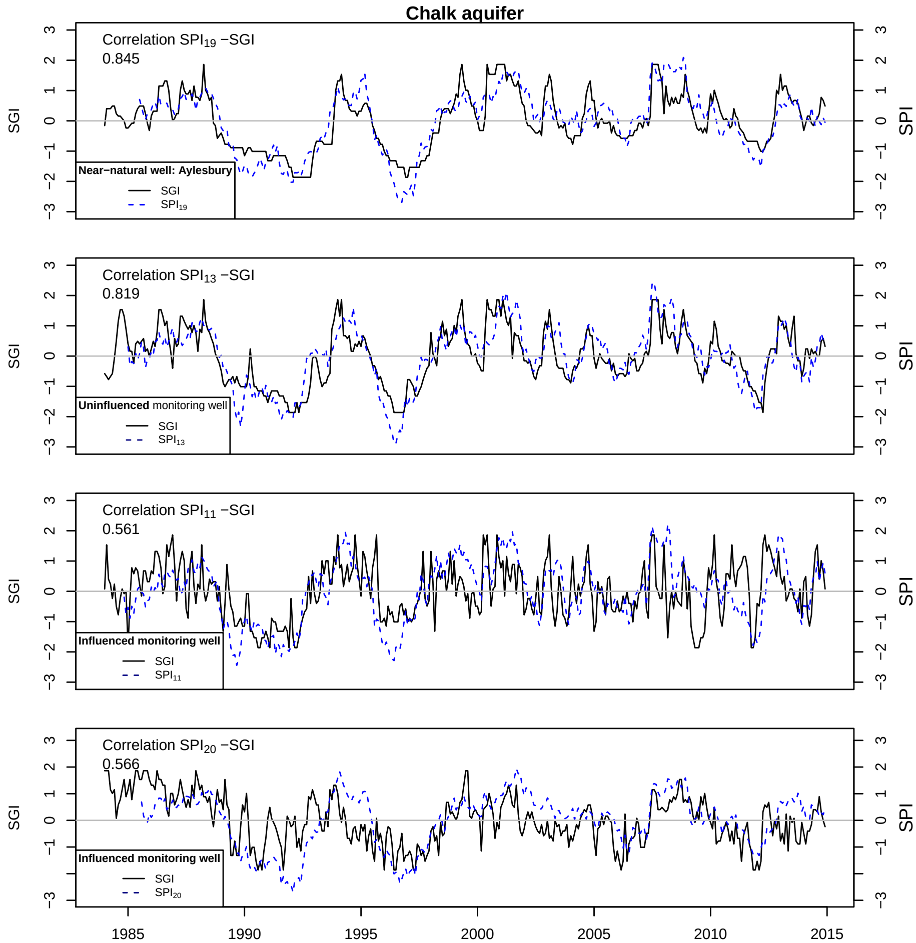


Figure 2. SPI$_Q$-SGI comparison for a near-natural Index site (top panel), an uninfluenced monitoring site (second panel), and two influenced monitoring sites in Lincolnshire (third and fourth panel). The SGI and SPI$_Q$ are shown in black and blue (dashed). The correlation between the SPI$_Q$-SGI is shown in the top left corner of the hydrograph.

were in introduced to reduce the water demand and there is mention of additional groundwater supply and aggravated droughts in regions heavily used for groundwater abstraction (Durant et al. 2015). For the second drought event in 1995-97, an extreme rise in water use was reported by Walker and Smithers (1998) for the summer of 1995 and to a lesser degree in 1996. During the 2003-06 and 2010-12 droughts, a sudden increase in groundwater use was found and attributed to dry springs and hot summers in the work of Marsh et al. (2007; 2013) and Durant et al. (2015).

In the work of Rey et al. (2016), low SPI$_3$ values were found in summer months for 1995, 1996, 2003-2006, and 2010-2011 highlighting exceptional dry weather. These low SPI$_3$ values were related to the increased water use prior to a drought and adaptations in agriculture during droughts (L269-271). By rephrasing L267-261 in the Result section and L319-322 in the Discussion section, we will improve the use of literature to support the current hypothesis, as examples of the literature have not been directly related to the results. We agree with Reviewer 2 that the attribution could be improved and we will thus provide additional support in the literature for this attribution.

Comment 5b

> '... , many of these drought events can also be identified in the uninfluenced wells. These uncertainties need to be reflected in the discussion or the methods for identification need to be more robust.'

Both Reviewer 1 (comment 5) and Reviewer 2 mention the observation of short drought events identified in the uninfluenced wells. We will illustrate why these drought events can be identified in these wells that are in fact *periodically influenced*. In the current analysis, we focus on identifying long-term or continuous human-influence on groundwater level observations, which is why these periodically influenced wells are not separately classified.

When using the SPI_Q-SGI correlation of 30-year time series to classify monitoring sites into uninfluenced or influenced sites, the aim is to identify long-term human-influence of groundwater abstraction can disturb the relation between precipitation and groundwater. Short periods of abstraction (periodic human-influence) resulting in sudden lower SGI or a dry well might not directly affect the long-term relation between groundwater and precipitation. To illustrate this periodic influence, we show three time series of Index and Observation wells in the Chalk, each of which are differently influenced by groundwater abstraction (Figure 3). Influence of groundwater abstraction is reported for these wells, since these Index and Observation wells are part of the Hydrometric Register [R122-125; Marsh and Hannaford, 2008). The first example site represents near-natural groundwater conditions (Therfield) and correlates well with the SPI (SPI_Q-SGI correlation of 0.841). For the second site (the Holt), there is periodic influence of groundwater abstraction reported. There is a sudden drop in SGI in 2000 and in 2003 (within the red box in Figure 3). The long-term correlation remains high (0.812) suggesting that the record is only affected for these very short periods of time. The long-term relation between precipitation and groundwater remains and the well would have been classified as uninfluenced in the analysis. The last site is located close to a groundwater pumping station (Towerhills pumping station) and is assumed to be continuously influenced resulting in a lower correlation (0.505), as the relationship between driving precipitation and groundwater has been disturbed over a much longer period.

This example highlights the robustness of the method to detect long-term human-influence on groundwater time series. However, periodic influence goes (partly) undetected, which is the primary reason for finding short drought events in relatively uninfluenced groundwater monitoring sites. We will rephrase L173-177 in the current manuscript to clarify this, as we recognise that the current phrasing of the strengths and weaknesses is incomplete. Figure 3 will be included in the supplementary material to illustrate the strengths and weaknesses of the used method.

Comment 4

> There is a lot of variation in the groundwater levels between sites and this needs to be better reflected in the results. I suggest that the authors report the min/max or 5th/95th percentile of their results alongside the average in Table 2 and elsewhere in the text.

We agree with Reviewer 2 that there is a lot of variation in the groundwater data. This is inherit in a regional groundwater study using groundwater level observations from a range of different hydrogeological settings. In addition to the regional differences, the differences in time are considerable, as groundwater droughts are episodic and drought events may vary in timing, intensity and duration. On top of the spatial and temporal differences, the human-influence on the groundwater monitoring sites changes in time, as illustrated in Figure 3 showing the difference between periodic and continuous human-influence on groundwater. We intend to improve the phrasing in the result section describing these three different facets in the drought characteristics [L230-246].

In the current manuscript, we have tried to show spatial and temporal variability in the results by highlighting the differences in Figure 2 in the current manuscript. We understand that there might be a further need for showing the variability in the data. To highlight the variability, we have added an example of the distribution of recorded drought frequency across the water management units (Figure 4). In this figure, the distribution of recorded drought frequency per site is shown that (on average) increases for influenced sites in Lincolnshire, Chilterns, and Shropshire. The distribution plots show a larger range for monitoring sites in the water management units compared to reference sites. For example, in the Chilterns, an increase in drought frequency is found for a small proportion of the sites compared to a larger proportion in Lincolnshire and the Shropshire. In the Midlands, the reference drought frequency is exceeded in both directions. Both lower and higher drought frequencies are

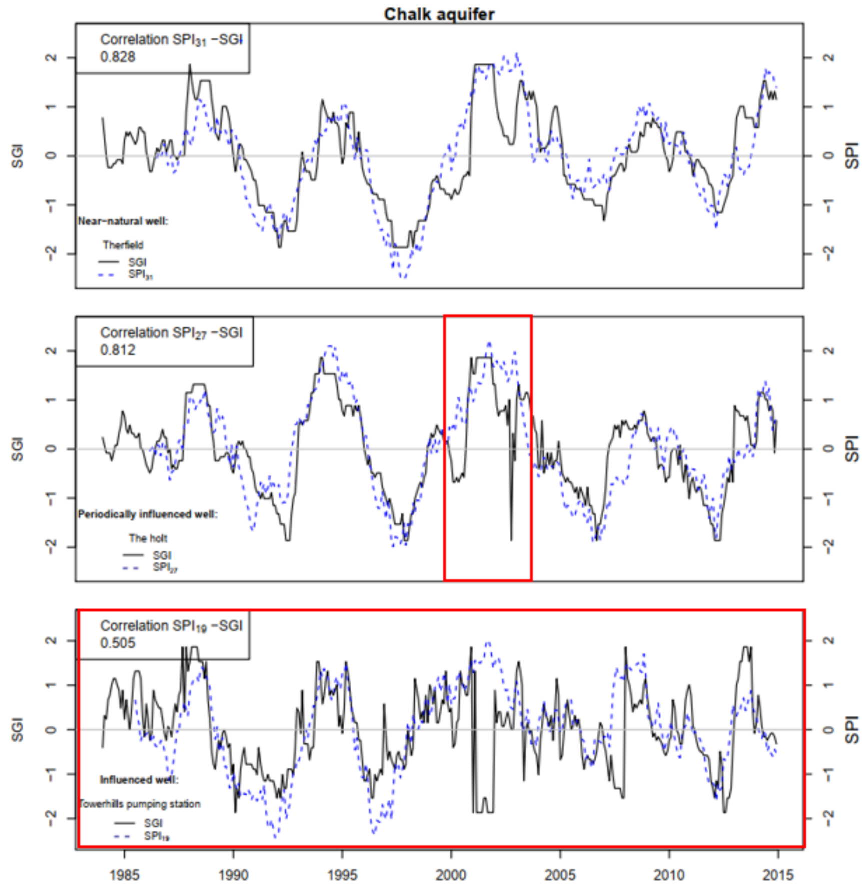


Figure 3. Three groundwater hydrographs and their corresponding SPI are shown to illustrate the impact of periodic influence of abstraction (second panel) and continuous influence of abstraction (third panel) on the SPI_Q-SGI correlation

found in this water management unit compared to the reference wells. On average, drought frequency decreases representing the majority of sites that observe fewer droughts. Figure 4 shows thus the variability in drought frequency, illustrating how not only the min/max has changed as Reviewer 2 highlights, but also the distribution and the range of recorded drought frequency. We will use this explanatory figure to improve the phrasing in L230-246, describing the variability in drought characteristics. The figure will be included in the supplementary material.

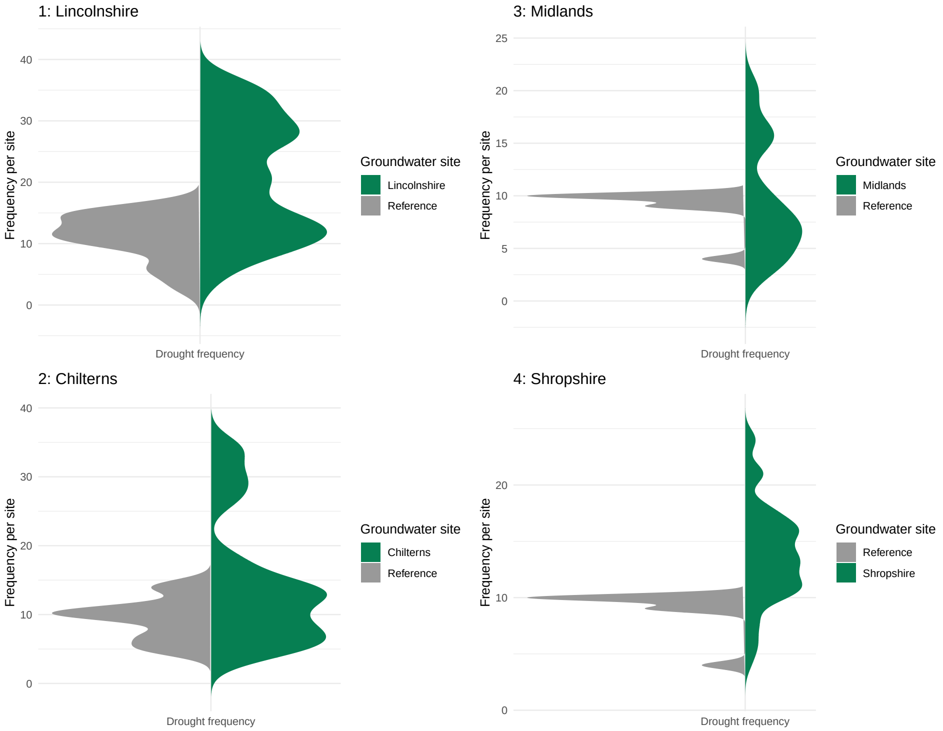


Figure 4. Drought frequency distribution of the four water management units. Drought frequency of the near-natural wells is shown in grey. Drought frequency of monitoring sites in the water management units is shown in green.

Literature

Bloomfield, J. P. and Marchant, B. P.: Analysis of groundwater drought building on the standardised precipitation index approach, Hydrology and Earth System Sciences, 17, 4769–4787, https://doi.org/10.5194/hess-17-4769-2013, 2013.

Bloomfield, J. P., Marchant, B. P., Bricker, S. H., and Morgan, R. B.: Regional analysis of groundwater droughts using hydrograph classification, Hydrology and Earth System Sciences, 19, 4327–4344, https://doi.org/10.5194/hess-19-4327-2015, 2015.

Durant, M.: Description of groundwater droughts in the UK: 1890 to 2015, Nottingham, UK, British Geological Survey, (OR/15/007), 52, 2015

Haas, J. C. and Birk, S.: Characterizing the spatiotemporal variability of groundwater levels of alluvial aquifers in different settings using drought indices, Hydrology and Earth System Sciences, 21, 2421–2448, https://doi.org/10.5194/hess-21-2421-2017, 2017.

Kumar, R., Musuuza, J. L., Van Loon, A. F., Teuling, A. J., Barthel, R., Ten Broek, J., Mai, J., Samaniego, L., and Attinger, S.: Multiscale evaluation of the Standardized Precipitation Index as a groundwater drought indicator, Hydrology and Earth System Sciences, 20, 1117–1131, https://doi.org/10.5194/hess-20-1117-2016, 2016

Li, B. & Rodell, M. Evaluation of a model-based groundwater drought indicator in the conterminous U.S. Journal of Hydrology, 526, 78-88, 2015

Marsh, T., Cole, G., and Wilby, R.: Major droughts in England and Wales, 1800-2006, Royal Meteorological Society, 62, 87–93, https://doi.org/10.1002/wea.67, 2007.

Marsh, T. and Hannaford, J., eds.: UK hydrometric register. A catalogue of river flow gauging stations and observation wells and boreholes in the United Kingdom together with summary hydrometric and spatial statistics, Hydrological Data UK, Centre for Ecology & Hydrology, 2008.

Marsh, T., Parry, S., Kendon, M., and Hannaford, J.: The 2010-12 drought and subsequent extensive flooding: a remarkable hydrological transformation, Tech. rep., Wallingford, 2013

Mckee, T. B.; Doesken, N. J. & Kleist, J. The relationship of drought frequency and duration to time scales AMS 8th Conference on Applied Climatology, 179-184, 1993,

Rey, D., Holman, I. P., and Knox, J. W.: Developing drought resilience in irrigated agriculture in the face of increasing water scarcity, Regional Environmental Change, 17, 1527–1540, 2017.

Stagge, J.H., Tallaksen, L.M., Gudmundsson,L. Van Loon, A.F. and Stahl K. Candidate Distributions for Climatological Drought Indices (SPI and SPEI) INTERNATIONAL JOURNAL OF CLIMATOLOGY 35: 4027–4040, 2015

Svensson, C., Hannaford, J., and Prosdocimi, I. Statistical distributions for monthly aggregations of precipitation and streamflow in drought indicator applications, Water Resour. Res., 53, 999– 1018, doi:10.1002/2016WR019276. 2017

Walker, S. and Smithers, H.: A review of the 1995–96 drought in the North West, Water and Environment Journal, 12, 273–279, 1998.