

Testing the use of standardised indices and GRACE satellite data to estimate the European 2015 groundwater drought in near real-time

Anne F. Van Loon^{1,*}, Rohini Kumar^{2,*}, and Vimal Mishra³

*These authors contributed equally to this work

¹School of Geography, Earth and Environmental Sciences, University of Birmingham, UK

²UFZ-Helmholtz Centre for Environmental Research, Leipzig, Germany

³Civil Engineering, Indian Institute of Technology, Gandhinagar, Gujarat, India

Correspondence to: Anne Van Loon (a.f.vanloon@bham.ac.uk)

Abstract. In 2015, central and eastern Europe were affected by a severe drought. This event has recently been studied from meteorological and streamflow perspective, but no analysis of the groundwater situation has been performed. One of the reasons is that real-time groundwater level observations often are not available. In this study, we evaluate two alternative approaches to quantify the 2015 groundwater drought over two regions in southern Germany and eastern Netherlands. The first approach is based on spatially-explicit relationships between meteorological conditions and historic groundwater level observations. The second approach uses the Gravity Recovery Climate Experiment (GRACE) Terrestrial Water Storage (TWS) and groundwater anomalies derived from GRACE-TWS and (near-)surface storage simulations by the Global Land Data Assimilation System (GLDAS) models. We combined the monthly groundwater observations from 2040 wells to establish the spatially-varying optimal accumulation period between the Standardised Groundwater Index (SGI) and the Standardised Precipitation Evapotranspiration Index (SPEI) at a 0.25 gridded scale. The resulting optimal accumulation periods range between 1 and more than 24 months, indicating strong spatial differences in groundwater response time to meteorological input over the region. Based on the estimated optimal accumulation periods and available meteorological time series, we reconstructed the groundwater anomalies up to 2015 and found that in Germany a uniform severe groundwater drought persisted for several months in this year, whereas the Netherlands appeared to have relatively high groundwater levels. The differences between this event and the 2003 European benchmark drought are striking. The 2003 groundwater drought was less uniformly pronounced, both in the Netherlands and Germany. This is because slowly responding wells (the ones with optimal accumulation periods of more than 12 months) still were above average from the wet year of 2002, which experienced severe flooding in central Europe. GRACE-TWS and GRACE-based groundwater anomalies did not capture the spatial variability of the 2003 and 2015 drought events satisfactorily. GRACE-TWS did show that both 2003 and 2015 were relatively dry, but the difference between Germany and the Netherlands in 2015 and the spatially-variable groundwater drought pattern in 2003 were not captured. This could be associated to the coarse spatial scale of GRACE. The simulated groundwater anomalies based on GRACE-TWS deviated considerably from the GRACE-TWS signal and from observed groundwater anomalies. The uncertainty in the GRACE-based groundwater anomalies mainly results from uncertainties in the simulation of soil moisture by the different GLDAS models. The GRACE-based groundwater anomalies are therefore not suitable for use in real-time groundwater drought monitoring in our case study regions. The alternative approach based on the spatially-variable relationship between meteorological conditions

and groundwater levels is more suitable to quantify groundwater drought in near real-time. Compared to the meteorological drought and streamflow drought (described in previous studies), the groundwater drought of 2015 had a more pronounced spatial variability in its response to meteorological conditions, with some areas primarily influenced by short-term meteorological deficits and others influenced by meteorological deficits accumulated over the preceding two years or more. In drought management, this information is very useful and our approach to quantify groundwater drought can be used until real-time groundwater observations become readily available.

1 Introduction

In the summer of 2015 large parts of Europe were affected by a severe drought (Van Lanen et al., 2016; Orth et al., 2016). This drought event has been analysed from climatological (Ionita et al., 2016; Orth et al., 2016) and hydrological (Laaha et al., 2016) perspective, which gives a useful overview of the causes, development, extent and severity of the drought event. Ionita et al. (2016), for example, found that the below-normal precipitation amounts in the summer of 2015 were related to positive geopotential height anomalies over Europe, large negative SST anomalies in the North-Atlantic Ocean and positive SST anomalies in the Mediterranean Sea. The effects of the below-normal precipitation and high temperatures during summer 2015 resulted in streamflow droughts with return periods of 100 years and more in central Europe (Laaha et al., 2016). Impacts of the drought were felt across many sectors, including agriculture, drinking water supply, electricity production, navigation, fisheries, and recreation (Van Lanen et al., 2016).

Van Lanen et al. (2016) and Laaha et al. (2016) note that hydrological information is key to understanding and managing the impacts of drought events and that hydrological drought needs to be monitored. Hydrological drought encompasses both below-normal river flow and below-normal groundwater levels (Van Loon, 2015). For the 2015 event, however, Laaha et al. (2016) could only analyse streamflow drought because of the absence of near real-time groundwater level data for Europe. Monitoring groundwater drought, however, is highly relevant for some drought-sensitive sectors such as water supply and irrigation. Additionally, groundwater is linked to other hydrological variables such as soil moisture and streamflow; monitoring groundwater during drought gives important information about drought persistence in these related variables.

Timely groundwater drought analysis is often hampered by a lack of real-time data. Groundwater drought studies are therefore often done on historic data sets (e.g., Chang and Teoh, 1995; Peters et al., 2006; Tallaksen et al., 2009; Shahid and Hazarika, 2010; Bloomfield and Marchant, 2013; Kumar et al., 2016). For adequate groundwater drought monitoring and early-warning, however, more actual information is essential (Bachmair et al., 2016). Up to date, different approaches have been explored that aim to quantify groundwater drought in the absence of near real-time groundwater level data, using either long-term precipitation, streamflow, lumped conceptual models, land surface models or GRACE satellites data (e.g., Thomas et al., 2015; Laaha et al., 2016; Jackson et al., 2016; Li and Rodell, 2015; Thomas et al., 2014).

An approach that is often used is accumulating precipitation anomalies over a long period that is assumed to be representative for groundwater fluctuation (Zeng et al., 2008; Raziei et al., 2008; Thomas et al., 2015; Asoka et al., 2017). The assumption is that a precipitation-based index (like SPI) with a long accumulation period can adequately represent anomalies in groundwater

level. Good correlations are found when groundwater time series of a high number of wells are averaged over a large region or when only annual averages are considered (Joetzjer et al., 2013). Although groundwater on average responds to precipitation on long timescales, several studies have found that groundwater drought often has a heterogeneous pattern and can hardly be represented by a spatially uniform accumulation period (e.g., 6 or 12 months) for precipitation (Bloomfield et al., 2015; Kumar et al., 2016). The approach of accumulating precipitation over a fixed period completely disregards spatial variability in the response time of aquifers to precipitation (Stoelzle et al., 2014; Staudinger et al., 2015).

Laaha et al. (2016) suggest that “in absence of groundwater data, [streamflow] deficit volumes, representing the reduced outflow of stored sources in the catchment, may also be indicative for groundwater resources” (p.15). We agree that this approach might be more suitable as an indication of groundwater drought than a spatially-uniform SPI, but streamflow still integrates hydrological information over the whole catchment, disregarding spatial variability of groundwater levels within the catchment. This is especially important in large catchments and regions with high spatial variability of landscapes and geology (e.g., Troch et al., 2009; Fan et al., 2013). Additionally, streamflow data are often not available in near real-time either, as pointed out by Laaha et al. (2016).

Others have explored the use of land surface models to quantify groundwater drought (Li and Rodell, 2015), which worked well on regional scales, but had much lower reliability on local scales. Like the approaches mentioned above, land surface models can not capture the true spatial variability of groundwater droughts (Van Loon et al., 2012). Li and Rodell (2015) assimilated GRACE derived TWS anomalies in their land surface model and improved the results in terms of temporal correlation between modelled and observed groundwater drought, but because regionally-averaged GRACE data were used the spatial representation of modelled groundwater droughts was not improved. This raises the questions whether direct use of GRACE fields would provide enough resolution to accurately characterise the spatial variability of groundwater droughts (AghaKouchak et al., 2015). GRACE data has been used to quantify groundwater drought in a number of recent studies (Thomas et al., 2014; Breña Naranjo and Pedrozo Acuña, 2016), but their ability to provide better representation of the spatial variability of groundwater droughts remains to be tested.

At the local scale more sophisticated approaches exist. For example, spring yields are probably one of the best alternatives for groundwater level data to evaluate the groundwater drought situation (Fiorillo, 2009). But these can only be used in regions where springs occur (i.e. not in lowland areas) and the availability and sharing of spring data is even more problematic than that of groundwater level data (Fendeková and Fendek, 2012). At local scales data are available to develop groundwater models, ranging from lumped conceptual models to physically-based distributed models. Physically-based distributed models are most suitable for a correct representation of complex groundwater drought situation (e.g., Peters et al., 2006; Tallaksen et al., 2009), but these require a good understanding of the aquifer system and a lot of data. Lumped conceptual models are much easier and quicker to develop. These models have shown skill in extrapolating beyond the observed range, both in hindcasting (Jackson et al., 2016) and in forecasting (Mackay et al., 2015) groundwater drought. Lumped conceptual models can be more easily set up for multiple sites covering a larger region than physically-based distributed models, but they are still site-specific, i.e. based on site-specific aquifer characteristics and calibrated on point-based groundwater level observations. Scaling up models to regional or continental scale is challenging, because parameter uncertainty is large.

Summarising, the major drawback of all large-scale approaches that are used as alternative for near real-time groundwater level data is lack of spatial variability, and the challenge with site-specific observation (e.g. spring yields) or modelling is their low representativity on larger scales. There is an urgent need for a new large-scale approach that can account for spatial variability in groundwater drought. Until groundwater level observations become widely available in near real-time, we have to find pragmatic solutions for quantifying groundwater drought during or shortly after a drought event. In this paper, we evaluate two potential ways forward. We propose an approach that makes use of known spatial variation in groundwater based on historic groundwater level datasets, assuming that heterogeneity of landscape and geology does not change over short timescales. If we know the spatial variation in the response time of the groundwater to precipitation from analysis of historical drought events, we can use that relationship to extrapolate to monitor a current drought. These relationships have been determined for various parts of the world, e.g. Khan et al. (2008) in Australia, Mendicino et al. (2008) and Fiorillo and Guadagno (2012) in Italy, Bloomfield et al. (2015) in England, Li and Rodell (2015) in the USA, and Kumar et al. (2016) in the Netherlands and Germany. In this way, we combine readily available data (precipitation) with knowledge about the spatial variability of groundwater response derived from historic observations. Additionally, we are interested to explore the use of GRACE-TWS to monitor groundwater drought. Previous studies (e.g., Thomas et al., 2014) have used anomalies in the GRACE-TWS signal to quantify hydrological drought severity. Others use land-surface models to decompose the GRACE total water storage into components of the water balance to study groundwater anomalies in isolation (e.g., Breña Naranjo and Pedrozo Acuña, 2016). The use of models adds time and uncertainty to the groundwater drought monitoring, but may add accuracy by disregarding anomalies in surface storage and soil moisture.

In this paper, we aim to evaluate two methods that make use of available data to monitor groundwater drought in near real-time on a large scale and test their ability to capture required spatial variability ensuring their local usefulness. The two methods are, 1) a novel approach that exploits the pre-determined historical relationships between precipitation and groundwater levels, and 2) patterns of GRACE-TWS and derived groundwater anomalies. With these methods we analyse the spatio-temporal characteristics of the 2015 groundwater drought in Europe. The objectives of this study are: i) to test large-scale approaches to obtain near real-time groundwater drought information for the 2015 European drought event, ii) to contrast the pattern of the 2015 groundwater drought with that of the 2003 groundwater drought, and iii) to discuss the differences between the groundwater drought and the meteorological drought (Ionita et al., 2016) and streamflow drought (Laaha et al., 2016). We focus on two regions in Europe, the eastern part of the Netherlands and southern Germany, for which groundwater level data were available and historical relationships between precipitation and groundwater levels can be determined following the method of Kumar et al. (2016). The GRACE data for the 2003 and 2015 drought events were studied for the same domain.

In the following section (Sect. 2), we describe the study areas and data. In Section 3, we explain the methodology of deriving historical relationships from standardised meteorological conditions and groundwater levels, and the GRACE data analysis. In Section 4 we show the 2015 groundwater drought for both regions and compare it with the 2003 groundwater drought. In Section 5 we discuss our methodology and findings, and compare them with the meteorological and streamflow drought analysed by Ionita et al. (2016) and Laaha et al. (2016). Concluding remarks are given in Sect. 6.

2 Study areas and data

This study focusses on two hydrogeologically different regions within Europe, i.e. the eastern Netherlands and southern Germany (Fig. 1 - upper panels). Both regions have an oceanic climate (Köppen-Geiger: Cfb), with the southern Germany region experiencing less moderating influence from the ocean. The study area in the Netherlands is composed of sedimentary deposits with varying thickness and composition, ranging from low-lying river valleys with clay and loam sediments and shallow groundwater, to ice-pushed ridges consisting of coarse sand and gravel with an unsaturated zone thickness of up to 30 m. The study area in Germany has a large variability in aquifers, due to a hilly to mountainous terrain and a wide range of unconsolidated and consolidated geological formations (Kumar et al., 2016).

These regions were affected differently by the 2015 drought and are therefore suitable to study spatial differences in groundwater drought using different methods. In our study, we used data of groundwater levels, meteorological variables (precipitation and temperature), and GRACE-TWS. For the meteorological datasets we used the free E-OBS data set (v13.0) from the European Climate Assessment & Dataset project (Haylock et al., 2008). These datasets are available at a $0.25^\circ \times 0.25^\circ$ spatial resolution and were created using the external drift kriging interpolation technique based on ground-based observation networks. We extracted daily precipitation totals and daily temperatures (average, maximum and minimum) for the grid cells indicated in the green areas in Fig. 1. The daily precipitation totals were summed up to monthly precipitation totals. The temperature data were used to estimate daily potential evapotranspiration (PET) with the Hargreaves and Samani method, which only requires average, maximum and minimum daily temperature (Hargreaves and Samani, 1985). Other PET estimation methods could not be used, because of the limited availability of high-resolution meteorological variables over a long time period (1950-onwards). The daily PET values were then aggregated to monthly values.

We analysed monthly groundwater levels from a total number of 1991 and 49 observations wells located in southern Germany and the eastern Netherlands, respectively (Fig. 2 - upper left panels). The data for the German wells were acquired from the Bavarian Environment Agency (LfU Bayern) and the State Institute for Environment, Measurements and Nature Conservation Baden-Württemberg (LUBW), whereas the information for the Dutch wells was taken from the Dutch institute TNO (www.dinoloket.nl/). Most of these wells (90%) are located in shallow aquifers with an average depth to the water table lower than 20 m. The length of record varied from well to well with a minimum of 10 years, starting from the year 1951 for the German wells and 1988 for the Dutch wells and ending in the year 2013. Kumar et al. (2016) performed a sensitivity analysis on the effect of record length and did not find differences in results between wells with short or long record lengths. Since we used the same groundwater level dataset, we here assume no effect of record length. The selected wells are screened to have minimal anthropogenic influences, allowing the understanding of the natural response of the groundwater to the precipitation signal. Readers may refer to Kumar et al. (2016) for further details on the location, selection and/or pre-processing of groundwater well records.

The seasonal and inter-annual variation of terrestrial water storage (TWS) was derived from the remotely sensed anomalies of Earth's gravity field retrieved by GRACE (version 5.0; Landerer and Swenson, 2012). The monthly anomalies are computed by removing the long-term mean estimated over the baseline period of Jan-2004 to Dec-2009 (NASA, 2016). The monthly

GRACE derived TWS anomalies are available at $1^\circ \times 1^\circ$ spatial resolution from three different processing centres: the GeoForschungsZentrum (GFZ; Potsdam, Germany), the Center for Space Research at the University of Texas (CSR; Austin, United States), and the Jet Propulsion Laboratory (JPL; United States). We used the ensemble mean of these three GRACE-TWS products to reduce the noise (and scatter) among different TWS products. We used all monthly GRACE-TWS records that are available within the period 2002-2015.

3 Methodology

3.1 Analysis of standardised meteorological conditions and groundwater levels

In our novel approach, we analysed the spatio-temporal variability of the 2015 groundwater drought from the long-term relations between groundwater and driving meteorological conditions. For this purpose, the observations of monthly precipitation and groundwater levels were converted to the corresponding standardised anomalies (or probabilistic drought indices) denoted as Standardized Precipitation Evapotranspiration Index (SPEI; Vicente-Serrano et al., 2010) and standardised groundwater index (SGI; Bloomfield and Marchant, 2013). These indices represent standardised anomalies with respect to their own climatological estimates. SPEI accounts for both atmospheric water supply (precipitation) and demand (potential evapotranspiration; PET), which are both crucial variables for groundwater recharge, and therefore SPEI is assumed to correlate best to groundwater levels. We also considered the commonly used meteorological drought indicator Standardized Precipitation Index (SPI; McKee et al., 1993) that only takes into account precipitation.

Both meteorological drought indices (SPI and SPEI) can be estimated for different time scales by accumulating the corresponding monthly values (precipitation or precipitation minus PET) over different periods (e.g., 3, 6, 12, or 24 months) to reflect short and long-term meteorological droughts (e.g., Raziei et al., 2013; Potop et al., 2014; Stagge et al., 2015; Barker et al., 2016). The SGI, on the other hand, is estimated at a monthly time scale (Bloomfield et al., 2015; Kumar et al., 2016) since groundwater often exhibits a high persistence (large memory), and therefore accounts for effects of anomalous conditions during the preceding months.

Following Kumar et al. (2016), we used a non-parametric method to estimate the SGI, SPEI and SPI. Unlike fitting a pre-defined distribution function (e.g., Gamma function), the method uses a kernel density estimator to compute the cumulative probability distribution of the accumulated precipitation and groundwater data to estimate drought indices. The Gaussian kernel is selected here because of its unlimited support and the corresponding bandwidth is estimated via cross-validation approach (see Samaniego et al., 2013, for details). Furthermore, the bandwidth is computed for every calendar month and location separately to ensure comparability of resulting drought indices over time and space and to account for the high seasonality exhibited by both variables. The computed indices are bounded between [0-1] with values below (above) 0.5 denote dry (wet) conditions. A drought is defined when the index value fall below a threshold of 0.2 or the 20th percentile, indicating an event with a re-occurrence period of five years. We note that the quantile-based indices are now increasingly used in drought studies (Sheffield et al., 2004; Andreadis et al., 2005; Vidal et al., 2010; Samaniego et al., 2013; Kumar et al., 2016), and they can be easily transformed to the unbounded range of the standard normal distribution (Vidal et al., 2010). For example, the usual

drought categories (see <http://droughtmonitor.unl.edu>) of moderate (e.g., $SGI \leq 0.2$), severe ($SGI \leq 0.1$), extreme ($SGI \leq 0.05$) and exceptional ($SGI \leq 0.02$) drought can easily be estimated.

We performed a regional scale groundwater drought analysis using indices computed at 0.25° grid resolution, at which the meteorological data are available. The SPI and SPEI were directly computed at this grid resolution for multiple accumulation periods ranging between 1 and 48 months. Following Kumar et al. (2016), the gridded fields of SGI were estimated by first averaging the well-specific SGIs from all those wells that lie within the selected grid cell. And then, the averaged SGI values at each grid were converted into a percentile-based drought index using the non-parametric kernel density estimator approach. The number of wells within each 0.25° grid cell varied between 1 and 115 for Germany and between 1 and 6 for the Netherlands, with a median value of around 6 and 3 wells, respectively (Fig. 2 - upper left panels). The south German study region is covered by 166 grid cells, whereas the Dutch part is covered by 17 cells.

We use a cross-correlation analysis to establish the linkage between meteorological and groundwater drought indices (i.e., SPI vs SGI and SPEI vs SGI) corresponding to different accumulation periods of SPI and SPEI (1-48 months). The goal is to identify the optimal accumulation period of meteorological drought indices (SPI and SPEI) that aligns with the groundwater drought index (SGI). The Spearman rank correlation coefficient (r) is used to quantify the strength of a monotonic relationship between drought indices. This relationship is established at every grid cell separately, recognising that the optimal accumulation period varies from grid cell to grid cell, and a spatially fixed a-priori selection of the accumulation period (for computing the SPI and SPEI) would result in inadequate characterization of groundwater droughts (Kumar et al., 2016). The identified optimal accumulation periods of SPI and SPEI in the historical time period (1951-2013 for Germany and 1988-2013 for the Netherlands) are then used to reconstruct the 2015 groundwater drought over the study domain.

3.2 Analysis of GRACE-TWS and groundwater anomalies

GRACE-TWS shows a pronounced seasonal variation in western Europe that should be removed if we want to study drought in a similar way as with standardised indices such as SPEI and SGI. Because GRACE data is available for a relatively short period, we only applied a simple deseasonalisation or standardization, by estimating the z-score based on the monthly mean and standard deviation of the TWS-anomalies for each calendar month and cell, separately.

GRACE-TWS represents variability in both groundwater and (near) surface water. We obtained (near) surface water storage (sum of soil moisture, canopy storage, and surface water) from four land surface models (VIC, Noah, MOSAIC, and CLM) available from the Global Land Data Assimilation System (GLDAS; Rodell et al., 2004). Monthly groundwater anomalies were then constructed after subtracting the GLDAS model derived (near) surface water storages from the TWS anomaly. To reduce the noise from individual GLDAS model outputs, we use the ensemble mean of the groundwater anomalies in our analysis.

For the two study regions, we estimated area-averaged ensemble mean and range (max-min) for TWS (based on the three GRACE products: JPL, CSR, and GFZ) and soil moisture (from the four GLDAS models: VIC, Noah, CLM, and MOSAIC). Then ensemble mean groundwater anomalies and associated uncertainty were estimated using the 12 ensemble members, i.e. the combination of three TWS and four GLDAS products.

4 Results

4.1 Characterisation of the 2003 and 2015 droughts in case study regions from precipitation, GRACE-TWS and groundwater observations

In this section we will characterise the 2003 and 2015 drought in our case study regions in the Netherlands and Germany from readily available data, namely precipitation and GRACE-TWS. For 2003, we will also show the groundwater drought derived from borehole observations.

Average annual precipitation is relatively similar in the two case study regions (Fig. 1 - upper left panel), but drought can have a clearly distinct pattern. In 2003, both regions were relatively dry, whereas in 2015 southern Germany was the hotspot of the precipitation anomaly while the Netherlands had a relatively wet summer (Ionita et al., 2016; Laaha et al., 2016). This difference is visible both in the spatial patterns and in the annual anomalies (Fig. 1 - upper and lower panels, respectively). Moreover, the southern Germany region has endured two consecutive precipitation deficits of more than 10% during 2014 and 2015, which may translate to a (larger) deficit in sub-surface water components (e.g., groundwater). In contrast, the region in the Netherlands endured two consecutive positive precipitation anomalies during the recent years. The two regions also exhibited a contrasting behaviour in terms of precipitation anomalies over the last decade: southern Germany experienced negative anomalies (drier conditions), while the eastern Netherlands show positive anomalies (wetter conditions) in the majority of years during the last decade (Fig. 1 - lower panels).

Unfortunately, GRACE-TWS was not available for October and November 2015 due to technical problems with the satellites. Therefore, we mapped the months of August, September and December in Fig. 3, both for 2003 and 2015 and for both study regions. As evident from the figure, both years had a negative terrestrial water storage anomaly in August and September. The TWS anomaly for the Netherlands was lower than that in Germany in both summers, ranging between -1 and -7 cm for the Netherlands and -6 and -14 cm for Germany. There was no apparent difference in the TWS spatial pattern between 2003 and 2015, with similar cells in the domain showing relatively higher or lower values. The only differences are the high TWS values in Scandinavia in 2015 and the drought recovery in December, which shows a southwest - northeast gradient in 2003 and a north - south pattern in 2015.

Figure 4 shows that after deseasonalisation, spatial patterns of TWS in the summers of 2003 and 2015 are still similar over the study regions. Again, only the recovery in December is clearly different between the two years. The timeseries for the grid boxes of Germany and the Netherlands (Fig. 4 - lower panels), do indicate different antecedent conditions in 2003 compared to 2015. Standardised TWS anomaly values were between 2 and 3 at the start of 2003 and between 0.5 and 1.5 in 2015, but this does not seem to have any significant effect on the drought severity, with standardised TWS anomalies for Germany around -1.5 in both summers. The timeseries of TWS in Fig. 4 (grey lines in lower panels) also show that the spatial variability in standardised TWS anomaly values is limited, due to a relatively coarser resolution of GRACE dataset.

For 2003, the gridded SGI values derived from borehole observations show a pattern of high and low groundwater levels (Fig. 5). In Germany, the groundwater drought is clustered in the southeast, southwest and centre of the study area. In the Netherlands, the east and west are drier than normal and the centre of the study area is wetter than normal.

4.2 Relationship between standardised meteorological conditions and groundwater levels

The first step in quantifying the 2015 groundwater drought from meteorological data is determining the spatially explicit relationship between SGI and SPEI. Figure 2 (upper middle panels) shows that both in the Netherlands and Germany the optimal accumulation period of SPEI to match SGI varies substantially, between 1 and 48 months (4 years). The spatial pattern is similar to that of Kumar et al. (2016), although the precipitation dataset and resolution were different (see Sect. 5 for more details). In the Netherlands, we see relatively high SPEI accumulation periods in the centre and low accumulation periods in the west and east. In Germany, accumulation periods are relatively low in the south and high in the north, intermediate accumulation periods occur in the centre and high values in the east and west.

As expected, the maximum correlation (the correlation obtained for the optimal SPEI accumulation period to match grid-average SGI) did not show a specific spatial pattern (Fig. 2 - upper right panels). In the Netherlands, all correlations are above 0.6, while in Germany values below 0.6 occasionally occur, but never below 0.4. There is no relationship between the number of wells in the grid cell and the maximum correlation.

Since we calculated both SPI and SPEI for all grid cells in the study areas, we could compare the accumulation periods of SPI and maximum correlations between SPI and SGI with those of SPEI (Fig. 2 - lower panels). The clustering of points around the 1:1-line illustrates that there is hardly any difference between the optimal accumulation period of SPI and SPEI to match SGI. However, the Dutch grid cells tend to be all located above the 1:1-line indicating on average slightly higher accumulation periods and corresponding maximum correlations when using SPEI instead of SPI.

4.3 The 2015 groundwater drought derived from the relationship with meteorological conditions

From the relationship between SPEI and SGI for the historic period, we calculated SGI values for the period for which we did not have groundwater level observations (i.e. after 2013). This allowed us to map the SGI, derived based on the spatially varying optimally accumulated SPEI, during the drought event of 2015 (Fig. 6 and 7). The 2015 groundwater drought in Germany was extremely severe throughout the whole domain (i.e. SGI below the drought threshold of the 20th percentile for almost all grid cells in August, September and October 2015, and only a few grid cells in the Alps higher with values around 0.3 - 0.4; Fig. 6). The Netherlands, however, showed relatively high SGI values in 2015 (above the 50th percentile, so wetter than normal, for all grid cells in the domain; Fig. 7).

When we compare the 2015 groundwater drought with the 2003 event, we notice that in both study regions there was a much larger spatial variability in SGI levels in 2003. In the Netherlands, 2003 actually showed up as a severe drought in some grid cells (Fig. 7 - upper panels), although on average the region was not classified as drought (Fig. 7 - lower panel). The reason for this is that some wells in the centre of the domain, those with optimal accumulation periods of more than 24 months (Fig. 2), had higher than normal groundwater levels, even up to the 90th percentile. The same happened in Germany in 2003. Many grid cells in the northwest and east of the domain had above normal SGI (50th - 80th percentile; Fig. 6 - upper panels), while the grid cells in the centre, southwest and southeast had extremely low values. Although the values are slightly different, the SGI pattern derived from the SPEI and historic relationships between SGI and SPEI (Figs. 6 and 7 - upper panels) is very similar

to the one derived from borehole observations directly (Fig. 5). Despite the fact that the spatial average for the German region showed a drought in the summer of 2003 (below the 20th percentile, Fig. 6 - lower panel; in contrast to that of the Dutch region, Fig. 7 - lower panel), the SGI values ranged from 0 to around 0.95 (the 95th percentile), which is a very large spread. In the summer of 2015, the spatially averaged SGI was around 0.05 (5th percentile) for a few months and the range in SGI values among grid cells was exceptionally low, i.e. in August 2015 only up to 0.3.

The crucial difference between the 2003 and 2015 summer droughts was the antecedent conditions. Since the majority of the groundwater wells responds to meteorological conditions accumulated over 6 to 24 months, with some exceptions going up to 48 months (Fig. 2), the relative wetness or dryness over the previous 6-48 months strongly influences the groundwater drought condition. Before 2003, the preceding three years had above average groundwater levels, with spatially averaged SGI fluctuating around 0.7 (Fig. 7 - lower panel), culminating in a period of uniformly extremely high SGI values at the end of 2002 and start of 2003 (above 0.8). The antecedent conditions for 2015 show a number of years with below average SGI values and quite a severe drought in spring 2014. The timeseries of SGI in Germany also show that by the end of 2015 the groundwater had not recovered from the severe summer drought with spatially average values still around 0.05 (Fig. 7 - lower panel).

4.4 The 2015 groundwater drought derived from GRACE

The GRACE-TWS data have been converted into groundwater anomalies by subtracting GLDAS model results of soil moisture and surface water stores. The resulting GRACE-GLDAS groundwater anomalies during the 2003 and 2015 drought events (Fig. 8, top panels) show a scattered pattern over Europe with our study area in the Netherlands showing wetter than normal groundwater storage during both droughts and our study area in Germany showing both wet and dry anomalies. The pattern of GRACE-GLDAS groundwater anomalies does not correspond to patterns found in the gridded groundwater anomaly estimates from boreholes (Fig. 5) and does not show any similarity to other patterns of the 2003 and 2015 meteorological or hydrological drought (Ionita et al., 2016; Laaha et al., 2016) or to known differences in the aquifer characteristics or groundwater depth over Europe (Wendland et al., 2008; Fan et al., 2013).

The bottom two panels in Fig. 8 depict the monthly GRACE-GLDAS groundwater anomalies for every grid cell located within the study domains as well as their spatial averages. Except for a peak in 2014, the temporal variation in GRACE-GLDAS groundwater anomalies is limited. When compared to time series of groundwater anomalies from borehole observations (Fig. 5 in Kumar et al., 2016), no agreement is found. The removal of the seasonal cycle from the GRACE-GLDAS groundwater anomaly estimates (Fig. 9) only slightly exaggerates the spatial pattern exhibited in Fig. 8 and also the time series do not change significantly. The individual GLDAS models show similar results (see Appendix A).

5 Discussion

The objectives of this study were: i) to test large-scale approaches to obtain near real-time groundwater drought information for the 2015 European drought event, ii) to contrast the pattern of the 2015 groundwater drought with that of the 2003 groundwater

drought, and iii) to discuss the differences between the groundwater drought and the meteorological drought (Ionita et al., 2016) and streamflow drought (Laaha et al., 2016).

5.1 Testing a novel approach based on standardised indices to obtain near real-time groundwater drought information

5 We assessed two approaches to monitor groundwater drought in the absence of near real-time groundwater level observations. The first approach is based on pre-determined relationships between precipitation and observed historical groundwater levels. The method assumes that heterogeneity of landscape and geology does not change over the short timescales of the extrapolation (in this case 2013 to 2015) and that the relationship between standardised meteorological conditions (SPEI) and standardised groundwater levels (SGI) is relatively robust. The grid scale of $0.25^\circ \times 0.25^\circ$ used for the analysis averages out sub-grid spatial variability, for example Kumar et al. (2016) showed a lower range in correlations between precipitation and groundwater from gridded data than from station data. However, the grid cells we used in this study are finer than those used in that previous study (0.25 instead of 0.5 degree spatial resolution), which means that fewer groundwater boreholes are used in the determination of optimal accumulation period of SPEI to match SGI and has the advantage that less spatial variability in groundwater is averaged out. The other big advantage of using the gridded E-OBS dataset in this study in contrast to local observation networks is that these meteorological data for Europe are freely available and are updated regularly.

The highly variable spatial pattern of the optimal accumulation period of SPEI again shows the importance of using a spatially-variable accumulation period to represent groundwater compared to using a spatially-uniform accumulation period. This confirms the results of Bloomfield et al. (2015) and Kumar et al. (2016) on the significance of the underlying land surface properties (geology) in transmitting the precipitation signal to groundwater levels.

20 The analysis using SPI instead of SPEI to represent meteorological conditions gave very similar results. This means that adding the effect of potential evaporation does not change the relationship between meteorological forcing and groundwater levels; the delay in groundwater drought occurrence compared to meteorological drought occurrence is similar. This may be different in more arid regions where PET is a more important component in the water balance. For regions similar to the ones we analysed here, we expect that in absence of PET data SPI can be used instead of SPEI. The PET estimations used in this study are based on approximations of meteorological variables from average, maximum and minimum daily temperature. Temperature-based PET estimates in drought indices have been found to result in spurious trends in global drought (Sheffield et al., 2012). However, the temperature-based PET estimate used for estimation of SPEI is not expected to have influenced our results, because SPI showed similar accumulation periods as SPEI.

5.2 Testing the use of GRACE satellite data to obtain near real-time groundwater drought information

30 The second approach used GRACE-TWS and groundwater anomalies derived from GRACE satellite data and GLDAS models. The GRACE-TWS anomalies indicated drier than normal conditions in our case study regions, whereas GRACE-GLDAS groundwater anomalies showed both wetter and drier grid cells. Evaluation against gridded SGI values derived from well

observations for the 2003 drought indicated that the spatial pattern of both the GRACE-TWS anomalies and GRACE-GLDAS groundwater anomalies did not show any agreement with observations.

The range (max-min) of TWS, modelled soil moisture and groundwater anomalies estimated from three GRACE-TWS products and four GLDAS models (Figs. 10 and 11) show that the uncertainty in groundwater anomalies is largely due to uncertainty in modelled soil moisture from the GLDAS models. Since soil depth and number of soil layers in the GLDAS models vary, their total column soil moisture has large inter-model differences (Long et al., 2013). Moreover, due to the differences in physical process equations in the GLDAS models, there are uncertainties in the simulated soil moisture (for more information see Syed et al., 2008). None of the GLDAS models currently simulate groundwater storage and flow explicitly, which makes groundwater the rest term of the models and results in the checkerboard pattern groundwater anomalies in Appendix A. This is in line with the findings of Long et al. (2013) for the 2011 drought in Texas and with Syed et al. (2008) for river basins around the world. Long et al. (2013) found that both the direction and the order of magnitude of groundwater anomalies were different for different model combinations, leading to the conclusion that the groundwater anomaly estimates were "highly unreliable" (Long et al., 2013, p.3399). Our findings assert their conclusion that "basic TWS may provide a more reliable indicator of drought, rather than disaggregated SMS [soil moisture storage] and GWS [groundwater storage]" (Long et al., 2013, p.3399).

Other reasons for the unsuitability for GRACE-GLDAS groundwater anomaly results for real-time groundwater drought monitoring in these case study regions are: i) GRACE resolution may not be appropriate to understand the localized groundwater variability, and ii) model results are not always available in near real-time. From this analysis, we conclude that the GRACE-GLDAS groundwater anomalies currently have large uncertainties that limit their application in real-time groundwater drought monitoring in this part of Europe. We stress that further research is needed to evaluate the reasons for the discrepancies between the GRACE-based model outcomes and groundwater observations and that incorporating a fully-coupled groundwater component in the GLDAS models might be the way forward (e.g., Niu et al., 2007; De Graaf et al., 2017).

5.3 Contrasting the pattern of the 2015 groundwater drought with that of the 2003 groundwater drought

In comparison to the benchmark 2003 drought event, the 2015 groundwater drought covered the entire German study domain, whereas the Netherlands were relatively wet. In 2003, both regions showed a patchy pattern with both extremely dry and (extremely) wet groundwater conditions. These differences in spatial pattern of the groundwater drought are related to antecedent conditions that affect different locations differently depending on the response time of the aquifer to meteorological conditions. In 2015, antecedent conditions were relatively dry over a long period (several years), so that both the quickly and slowly responding aquifers had low groundwater levels during the 2015 summer drought. However, in 2003, antecedent conditions were relatively dry over short timescales (a few months), causing quickly responding aquifers to show severe drought in summer 2003, but extremely wet over longer timescales (a year and longer), causing slowly responding aquifers to show wetter conditions in summer 2003. The reason for the extremely wet antecedent conditions in 2003 are the 2002 summer floods in Central Europe (e.g., Ulbrich et al., 2003a, b). These floods mainly affected our study region in Germany (Fig. 1 and 6), although also the Netherlands showed a relatively wet year in 2002 (Fig. 1 and 7). Without the 2002 floods, the groundwater drought of 2003

would have been more severe in southern Germany than that of 2015. This study also indicates that the recovery from drought in groundwater can be very patchy, with relatively fast recovery in locations with quickly responding aquifers (e.g. the eastern and western parts of the study area in the Netherlands in October 2003, Fig. 7, and the Alpine region in October 2015, Fig. 6) and slow recovery for locations with an aquifer response time of a few years (e.g. no recovery for many grid cells in Germany by the end of 2015, Fig. 6).

5.4 Discussing the differences between with meteorological and streamflow drought

In three previous papers, the 2015 European drought has been investigated from a meteorological and hydrological (i.e., streamflow) perspective and focussing on impacts (Ionita et al., 2016; Laaha et al., 2016; Van Lanen et al., 2016). The 2015 groundwater drought in Europe, as analysed in this paper, showed some striking similarities and differences with the meteorological and streamflow drought. Ionita et al. (2016) analysed 3-month accumulation periods for SPI and SPEI in June, July and August and concluded that an important difference between the 2003 and 2015 drought was that the winter and spring of 2015 were rather wet compared to dry preconditions in 2003. Laaha et al. (2016) indicated that also for streamflow drought no exceptionally dry conditions were found for the winter prior to the 2015 summer drought. Our analysis shows that for groundwater drought we need to look further back than the previous spring or winter, because accumulation periods of above 2 years are not uncommon and spatial variability in the memory of the groundwater to meteorological conditions is large.

Laaha et al. (2016) also found that the 2015 drought was less spatially extensive and more "patchy" in streamflow than that noticed in the 2003 drought. This was partly confirmed for groundwater with 2015 drought conditions apparent in Germany, but not in the Netherlands. However, the 2015 groundwater drought in Germany was much more spatially uniform than the 2003 groundwater drought because of much drier preconditions. Laaha et al. (2016) also mention the effect of the 2002 flood event in Central Europe on the 2003 streamflow drought, but they only detect this effect in a shift of the onset of the drought. The reason for this could be that they used a fixed threshold to calculate streamflow drought, whereas standardised indices such as SGI are more comparable to a monthly-variable threshold (Van Loon, 2015).

Like streamflow drought, the severity of groundwater drought is determined by the combined effect of initial conditions (on different timescales), catchment functioning and aquifer characteristics, and atmospheric conditions (Laaha et al., 2016). In 2015, the combination of these factors resulted in a uniformly severe groundwater drought in Germany, which lead to socio-economic impacts in the region, e.g. water shortage for cattle due to dried boreholes (Van Lanen et al., 2016). These impacts are expected to be less severe in more patchy groundwater drought events, like the 2003 event, because in these cases other boreholes in the region would still exhibit above normal groundwater levels.

6 Conclusions

In this research we evaluated two approaches to monitor groundwater drought as alternative to real-time groundwater observations. The first method, based on extrapolation of the pre-determined relationship between meteorological conditions and observed groundwater levels, seemed suitable, as it reproduced the 2003 groundwater drought in both case study regions

(derived from borehole observations) and showed expected patterns of spatial variability for 2015, related to aquifer characteristics, antecedent conditions and meteorological drought. This novel approach performed much better than the satellite-based GRACE-TWS and GRACE-GLDAS groundwater anomalies. GRACE-TWS has a too coarse resolution to represent the observed spatial variability in groundwater drought and GRACE-GLDAS models were not able to simulate groundwater anomalies realistically for our study areas. We found a high spatial variability in the optimal accumulation period of SPEI to represent anomalies in groundwater levels (SGI), which had significant influence on the spatial pattern of the 2003 and 2015 groundwater drought. These results signify that a spatially-uniform accumulation period should not be used to represent hydrological drought over larger areas. The analysis of both SPI (representing accumulated precipitation anomalies) and SPEI (representing accumulated anomalies in precipitation minus potential evaporation) showed similar results, indicating that the propagation from meteorological drought to groundwater drought in our study areas is unchanged by including potential evaporation.

Our analysis shows that the 2015 groundwater drought was uniformly severe over southern Germany due to a combination of a relatively dry summer and dry preceding seasons and years. This is in contrast with the 2003 groundwater drought, which showed a severe drought in areas with quickly responding aquifer systems, both in the Netherlands and in Germany, in response to a dry summer, but wetter than normal conditions in areas with slowly responding aquifer systems, in response to a wet preceding two years. Especially in Germany, the 2002 summer floods still influenced some groundwater levels in the summer of 2003.

Groundwater drought is notably different from meteorological drought and streamflow drought. For example, the meteorological drought of 2015 was actually found to have originated much later than that of 2003, with persistent anticyclonic circulation starting at the end of spring and the end of winter for 2015 and 2003, respectively (Ionita et al., 2016). The streamflow droughts, which respond relatively quickly to precipitation deficits, matched up quite well with these meteorological droughts, with the 2015 streamflow drought having in general shorter durations than the 2003 streamflow drought (Laaha et al., 2016). However, (spatial) differences in the severity of streamflow drought within the larger European study area were apparent and were found to be related to differences in catchment functioning that determine how far back antecedent conditions influence streamflow. This effect is even stronger in groundwater: in many areas extending back to the previous winter is not enough to adequately capture the groundwater response to precipitation signals.

We recognise that our approach of using the pre-determined relationship between meteorological conditions and observed groundwater levels is crude and has uncertainties. It does, however, provide a first-order look into the spatio-temporal patterns of current and recent groundwater droughts based on meteorological indices. This is crucial for pro-active drought management in absence of real-time groundwater observations. With this work, however, we also want to promote more long-term groundwater measurement and international sharing of groundwater level data.

7 Code availability

The code for calculating the standardised indices and gridded estimates of groundwater drought are available from the authors. Please contact Rohini Kumar at rohini.kumar@ufz.de.

8 Data availability

The data used in this study are freely available from different sources. The Dutch groundwater level data are available online from the Dutch institute TNO (www.dinoloket.nl/). The E-OBS dataset is available from the EU-FP6 project ENSEMBLES and the European Climate Assessment & Dataset project (<http://ensembles-eu.metoffice.com> and <http://www.ecad.eu>). GRACE-TWS and GRACE-GLDAS data are free to download from <http://grace.jpl.nasa.gov/>. The German groundwater level data not publicly available, because they have been collected by the Bavarian Environment Agency (LfU Bayern) and the State Institute for Environment, Measurements and Nature Conservation Baden-Württemberg (LUBW) for use in GLOWA-Danube project (www.glowa-danube.de), funded by the German Ministry of Education and Research (BMBF). These institutes have restricted further distribution of the data.

10 Appendix A: GRACE-GLDAS groundwater anomalies simulated by four different models

To test the effect of using GLDAS model simulations, we used the CLM, NOAH, VIC and MOSIAC models to remove the total (near) surface water storage anomaly ($TOW = \text{Soil moisture} + \text{Canopy Storage} + \text{Snow water equivalent}$) from the GRACE-TWS anomaly. We calculated the ensemble mean of the resulting GRACE-GLDAS groundwater anomalies of the four models, but also plotted at the four models separately (Figs. A1 - A8). None of the models was able to reproduce the observed groundwater drought pattern for 2003 (Fig. 5) or the expected pattern for 2015 (Figs. 6 and 7).

Author contributions. A.F. Van Loon and R. Kumar designed the study, V. Mishra provided and processed GRACE data, R. Kumar performed the analyses with feedback from A.F. Van Loon and V. Mishra. A.F. Van Loon prepared the manuscript with contributions from all co-authors.

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

Acknowledgements. We would like to thank the Bavarian Environment Agency (LfU Bayern) and the State Institute for Environment, Measurements and Nature Conservation Baden-Württemberg (LUBW) who provided the southern Germany groundwater data sets for use in GLOWA-Danube (www.glowa-danube.de), a project funded by the German Ministry of Education and Research (BMBF). We also thank the Dutch institute TNO for making the Dutch groundwater level data available (www.dinoloket.nl/). We acknowledge the E-OBS dataset from the EU-FP6 project ENSEMBLES (<http://ensembles-eu.metoffice.com>) and the data providers in the European Climate Assessment & Dataset project (<http://www.ecad.eu>). We thank NASA and the GLDAS model groups for providing the GRACE-TWS and GRACE-GLDAS data (<http://grace.jpl.nasa.gov/>). We also want to acknowledge A. Akarsh for his help in GRACE and GLDAS data processing. This research is partly funded by NWO Rubicon project no. 2004/08338/ALW. This paper was developed within the framework of the UNESCO-IHP VIII FRIEND programme (EURO-FRIEND - Low flow and Drought group) and supports the Helmholtz Climate Initiative REKLIM project.

References

- AghaKouchak, A., Farahmand, A., Melton, F. S., Teixeira, J., Anderson, M. C., Wardlow, B. D., and Hain, C. R.: Remote sensing of drought: Progress, challenges and opportunities, *Reviews of Geophysics*, 53, 452–480, doi:10.1002/2014RG000456, 2014RG000456, 2015.
- Andreadis, K. M., Clark, E. A., Wood, A. W., Hamlet, A. F., and Lettenmaier, D. P.: Twentieth-Century Drought in the Conterminous United States, *Journal of Hydrometeorology*, 6, 985–1001, doi:10.1175/JHM450.1, 2005.
- Asoka, A., Gleeson, T., Wada, Y., and Mishra, V.: Relative contribution of monsoon precipitation and pumping to changes in groundwater storage in India, *Nature Geosci*, 10, 109–117, doi:10.1038/ngeo2869, 2017.
- Bachmair, S., Stahl, K., Collins, K., Hannaford, J., Acreman, M., Svoboda, M., Knutson, C., Smith, K. H., Wall, N., Fuchs, B., Crossman, N. D., and Overton, I. C.: Drought indicators revisited: the need for a wider consideration of environment and society, *Wiley Interdisciplinary Reviews: Water*, 3, 516–536, doi:10.1002/wat2.1154, 2016.
- Barker, L. J., Hannaford, J., Chiverton, A., and Svensson, C.: From meteorological to hydrological drought using standardised indicators, *Hydrology and Earth System Sciences*, 20, 2483–2505, doi:10.5194/hess-20-2483-2016, 2016.
- 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, doi: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, doi:10.5194/hess-19-4327-2015, 2015.
- Breña Naranjo, J. A. and Pedrozo Acuña, A.: Can the Gravity Recovery and Climate Experiment (GRACE) mission detect hydrological droughts?, in: *EGU General Assembly Conference Abstracts*, vol. 18, p. 11479, 2016.
- Chang, T. J. and Teoh, C. B.: USE OF THE KRIGING METHOD FOR STUDYING CHARACTERISTICS OF GROUND WATER DROUGHTS1, *JAWRA Journal of the American Water Resources Association*, 31, 1001–1007, doi:10.1111/j.1752-1688.1995.tb03416.x, 1995.
- De Graaf, I. E., van Beek, R. L., Gleeson, T., Moosdorf, N., Schmitz, O., Sutanudjaja, E. H., and Bierkens, M. F.: A global-scale two-layer transient groundwater model: Development and application to groundwater depletion, *Advances in Water Resources*, 102, 53 – 67, doi:10.1016/j.advwatres.2017.01.011, 2017.
- Fan, Y., Li, H., and Miguez-Macho, G.: Global Patterns of Groundwater Table Depth, *Science*, 339, 940–943, doi:10.1126/science.1229881, 2013.
- Fendeková, M. and Fendek, M.: Groundwater drought in the Nitra river basin-identification and classification, *Journal of Hydrology and Hydromechanics*, 60, 185–193, 2012.
- Fiorillo, F.: Spring hydrographs as indicators of droughts in a karst environment, *Journal of Hydrology*, 373, 290 – 301, doi:10.1016/j.jhydrol.2009.04.034, 2009.
- Fiorillo, F. and Guadagno, F. M.: Long karst spring discharge time series and droughts occurrence in Southern Italy, *Environmental Earth Sciences*, 65, 2273–2283, doi:10.1007/s12665-011-1495-9, 2012.
- Hargreaves, G. H. and Samani, Z. A.: Reference crop evapotranspiration from temperature, *Applied engineering in agriculture*, 1, 96–99, doi:10.13031/2013.26773, 1985.
- Haylock, M. R., Hofstra, N., Klein Tank, A. M. G., Klok, E. J., Jones, P. D., and New, M.: A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006, *J. Geophys. Res.*, 113, doi:10.1029/2008JD010201, 2008.

- Ionita, M., Tallaksen, L. M., Kingston, D. G., Stagge, J. H., Laaha, G., Van Lanen, H. A. J., Chelcea, S. M., and Haslinger, K.: The European 2015 drought from a climatological perspective, *Hydrology and Earth System Sciences Discussions*, 2016, 1–32, doi:10.5194/hess-2016-218, 2016.
- Jackson, C. R., Wang, L., Pachocka, M., Mackay, J. D., and Bloomfield, J. P.: Reconstruction of multi-decadal groundwater level time-series using a lumped conceptual model, *Hydrological Processes*, 30, 3107–3125, doi:10.1002/hyp.10850, hYP-15-0340.R1, 2016.
- Joetzier, E., Douville, H., Delire, C., Ciais, P., Decharme, B., and Tyteca, S.: Hydrologic benchmarking of meteorological drought indices at interannual to climate change timescales: a case study over the Amazon and Mississippi river basins, *Hydrology and Earth System Sciences*, 17, 4885–4895, doi:10.5194/hess-17-4885-2013, 2013.
- Khan, S., Gabriel, H. F., and Rana, T.: Standard precipitation index to track drought and assess impact of rainfall on watertables in irrigation areas, *Irrigation and Drainage Systems*, 22, 159–177, doi:10.1007/s10795-008-9049-3, 2008.
- Kumar, R., Musuza, 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, doi:10.5194/hess-20-1117-2016, 2016.
- Laaha, G., Gauster, T., Tallaksen, L. M., Vidal, J.-P., Stahl, K., Prudhomme, C., Heudorfer, B., Vlnas, R., Ionita, M., Van Lanen, H. A. J., Adler, M.-J., Caillouet, L., Delus, C., Fendekova, M., Gailliez, S., Hannaford, J., Kingston, D., Van Loon, A. F., Mediero, L., Osuch, M., Romanowicz, R., Sauquet, E., Stagge, J. H., and Wong, W. K.: The European 2015 drought from a hydrological perspective, *Hydrology and Earth System Sciences Discussions*, 2016, 1–30, doi:10.5194/hess-2016-366, 2016.
- Landerer, F. W. and Swenson, S. C.: Accuracy of scaled GRACE terrestrial water storage estimates, *Water Resour. Res.*, 48, doi:10.1029/2011WR011453, 2012.
- Li, B. and Rodell, M.: Evaluation of a model-based groundwater drought indicator in the conterminous U.S., *Journal of Hydrology*, 526, 78 – 88, doi:10.1016/j.jhydrol.2014.09.027, drought processes, modeling, and mitigation, 2015.
- Long, D., Scanlon, B. R., Longuevergne, L., Sun, A. Y., Fernando, D. N., and Save, H.: GRACE satellite monitoring of large depletion in water storage in response to the 2011 drought in Texas, *Geophysical Research Letters*, 40, 3395–3401, doi:10.1002/grl.50655, 2013.
- Mackay, J., Jackson, C., Brookshaw, A., Scaife, A., Cook, J., and Ward, R.: Seasonal forecasting of groundwater levels in principal aquifers of the United Kingdom, *Journal of Hydrology*, 530, 815 – 828, doi:10.1016/j.jhydrol.2015.10.018, 2015.
- McKee, T. B., Doesken, N. J., and Kleist, J.: The relationship of drought frequency and duration to time scales, in: *Proceedings of the 8th Conference on Applied Climatology*, vol. 17, pp. 179–183, American Meteorological Society Boston, MA, 1993.
- Mendicino, G., Senatore, A., and Versace, P.: A Groundwater Resource Index (GRI) for drought monitoring and forecasting in a mediterranean climate, *Journal of Hydrology*, 357, 282 – 302, doi:10.1016/j.jhydrol.2008.05.005, 2008.
- NASA: GGRACE monthly mass grids — Land. Jet Propulsion Laboratory, accessed May 2016. [Available online at <http://grace.jpl.nasa.gov/>], 2016.
- Niu, G.-Y., Yang, Z.-L., Dickinson, R. E., Gulden, L. E., and Su, H.: Development of a simple groundwater model for use in climate models and evaluation with Gravity Recovery and Climate Experiment data, *Journal of Geophysical Research: Atmospheres*, 112, n/a–n/a, doi:10.1029/2006JD007522, d07103, 2007.
- Orth, R., Zscheischler, J., and Seneviratne, S. I.: Record dry summer in 2015 challenges precipitation projections in Central Europe, *Scientific Reports*, 6, 28334–, 2016.
- Peters, E., Bier, G., van Lanen, H., and Torfs, P.: Propagation and spatial distribution of drought in a groundwater catchment, *Journal of Hydrology*, 321, 257 – 275, doi:10.1016/j.jhydrol.2005.08.004, 2006.

- Potop, V., Boroneanț, C., Možný, M., Štěpánek, P., and Skalák, P.: Observed spatiotemporal characteristics of drought on various time scales over the Czech Republic, *Theoretical and applied climatology*, 115, 563–581, 2014.
- Raziei, T., Saghafian, B., Paulo, A. A., Pereira, L. S., and Bordi, I.: Spatial Patterns and Temporal Variability of Drought in Western Iran, *Water Resources Management*, 23, 439–455, doi:10.1007/s11269-008-9282-4, 2008.
- 5 Raziei, T., Bordi, I., and Pereira, L. S.: Regional Drought Modes in Iran Using the SPI: The Effect of Time Scale and Spatial Resolution, *Water Resources Management*, 27, 1661–1674, doi:10.1007/s11269-012-0120-3, 2013.
- Rodell, M., Houser, P. R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C.-J., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M., Entin, J. K., Walker, J. P., Lohmann, D., and Toll, D.: The Global Land Data Assimilation System, *Bulletin of the American Meteorological Society*, 85, 381–394, doi:10.1175/BAMS-85-3-381, 2004.
- 10 Samaniego, L., Kumar, R., and Zink, M.: Implications of Parameter Uncertainty on Soil Moisture Drought Analysis in Germany, *Journal of Hydrometeorology*, 14, 47–68, doi:10.1175/JHM-D-12-075.1, 2013.
- Shahid, S. and Hazarika, M. K.: Groundwater Drought in the Northwestern Districts of Bangladesh, *Water Resources Management*, 24, 1989–2006, doi:10.1007/s11269-009-9534-y, 2010.
- Sheffield, J., Goteti, G., Wen, F., and Wood, E. F.: A simulated soil moisture based drought analysis for the United States, *Journal of*
 15 *Geophysical Research: Atmospheres*, 109, n/a–n/a, doi:10.1029/2004JD005182, d24108, 2004.
- Sheffield, J., Wood, E. F., and Roderick, M. L.: Little change in global drought over the past 60 years, *Nature*, 491, 435–438, doi:10.1038/nature11575, 2012.
- 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, doi:10.1002/joc.4267, 2015.
- 20 Staudinger, M., Weiler, M., and Seibert, J.: Quantifying sensitivity to droughts – an experimental modeling approach, *Hydrology and Earth System Sciences*, 19, 1371–1384, doi:10.5194/hess-19-1371-2015, 2015.
- Stoelzle, M., Stahl, K., Morhard, A., and Weiler, M.: Streamflow sensitivity to drought scenarios in catchments with different geology, *Geophysical Research Letters*, 41, 6174–6183, doi:10.1002/2014GL061344, 2014GL061344, 2014.
- Syed, T. H., Famiglietti, J. S., Rodell, M., Chen, J., and Wilson, C. R.: Analysis of terrestrial water storage changes from GRACE and
 25 GLDAS, *Water Resources Research*, 44, n/a–n/a, doi:10.1029/2006WR005779, w02433, 2008.
- Tallaksen, L. M., Hisdal, H., and Van Lanen, H. A.: Space-time modelling of catchment scale drought characteristics, *Journal of Hydrology*, 375, 363 – 372, doi:10.1016/j.jhydrol.2009.06.032, 2009.
- Thomas, A. C., Reager, J. T., Famiglietti, J. S., and Rodell, M.: A GRACE-based water storage deficit approach for hydrological drought characterization, *Geophysical Research Letters*, 41, 1537–1545, doi:10.1002/2014GL059323, 2014GL059323, 2014.
- 30 Thomas, T., Jaiswal, R. K., Nayak, P. C., and Ghosh, N. C.: Comprehensive evaluation of the changing drought characteristics in Bundelkhand region of Central India, *Meteorology and Atmospheric Physics*, 127, 163–182, doi:10.1007/s00703-014-0361-1, 2015.
- Troch, P. A., Carrillo, G. A., Heidbüchel, I., Rajagopal, S., Switanek, M., Volkmann, T. H. M., and Yaeger, M.: Dealing with Landscape Heterogeneity in Watershed Hydrology: A Review of Recent Progress toward New Hydrological Theory, *Geography Compass*, 3, 375–392, doi:10.1111/j.1749-8198.2008.00186.x, 2009.
- 35 Ulbrich, U., Brücher, T., Fink, A. H., Leckebusch, G. C., Krüger, A., and Pinto, J. G.: The central European floods of August 2002: Part 1 - Rainfall periods and flood development, *Weather*, 58, 371–377, doi:10.1256/wea.61.03A, 2003a.
- Ulbrich, U., Brücher, T., Fink, A. H., Leckebusch, G. C., Krüger, A., and Pinto, J. G.: The central European floods of August 2002: Part 2 - Synoptic causes and considerations with respect to climatic change, *Weather*, 58, 434–442, doi:10.1256/wea.61.03B, 2003b.

- Van Lanen, H. A., Laaha, G., Kingston, D. G., Gauster, T., Ionita, M., Vidal, J.-P., Vlnas, R., Tallaksen, L. M., Stahl, K., Hannaford, J., Delus, C., Fendekova, M., Mediero, L., Prudhomme, C., Rets, E., Romanowicz, R. J., Gailliez, S., Wong, W. K., Adler, M.-J., Blauhut, V., Caillouet, L., Chelcea, S., Frolova, N., Gudmundsson, L., Hanel, M., Haslinger, K., Kireeva, M., Osuch, M., Sauquet, E., Stagge, J. H., and Van Loon, A. F.: Hydrology needed to manage droughts: the 2015 European case, *Hydrological Processes*, pp. n/a–n/a, doi:10.1002/hyp.10838, hYP-16-0162, 2016.
- 5
- Van Loon, A.: Hydrological drought explained, *Wiley Interdisciplinary Reviews: Water*, 2, 359–392, doi:10.1002/wat2.1085, 2015.
- Van Loon, A. F., Van Huijgevoort, M. H. J., and Van Lanen, H. A. J.: Evaluation of drought propagation in an ensemble mean of large-scale hydrological models, *Hydrology and Earth System Sciences*, 16, 4057–4078, doi:10.5194/hess-16-4057-2012, 2012.
- Vicente-Serrano, S. M., Beguería, S., and López-Moreno, J. I.: A Multiscalar Drought Index Sensitive to Global Warming: The Standardized Precipitation Evapotranspiration Index, *Journal of Climate*, 23, 1696–1718, doi:10.1175/2009JCLI2909.1, 2010.
- 10
- Vidal, J.-P., Martin, E., Franchistéguy, L., Habets, F., Soubeyroux, J.-M., Blanchard, M., and Baillon, M.: Multilevel and multiscale drought reanalysis over France with the Safran-Isba-Modcou hydrometeorological suite, *Hydrology and Earth System Sciences*, 14, 459–478, doi:10.5194/hess-14-459-2010, 2010.
- Wendland, F., Blum, A., Coetsiers, M., Gorova, R., Griffioen, J., Grima, J., Hinsby, K., Kunkel, R., Marandi, A., Melo, T., Panagopoulos, A., Pauwels, H., Ruisi, M., Traversa, P., Vermooten, J. S. A., and Walraevens, K.: European aquifer typology: a practical framework for an overview of major groundwater composition at European scale, *Environmental Geology*, 55, 77–85, doi:10.1007/s00254-007-0966-5, 2008.
- 15
- Zeng, N., Yoon, J.-H., Marengo, J. A., Subramaniam, A., Nobre, C. A., Mariotti, A., and Neelin, J. D.: Causes and impacts of the 2005 Amazon drought, *Environmental Research Letters*, 3, 014002, 2008.

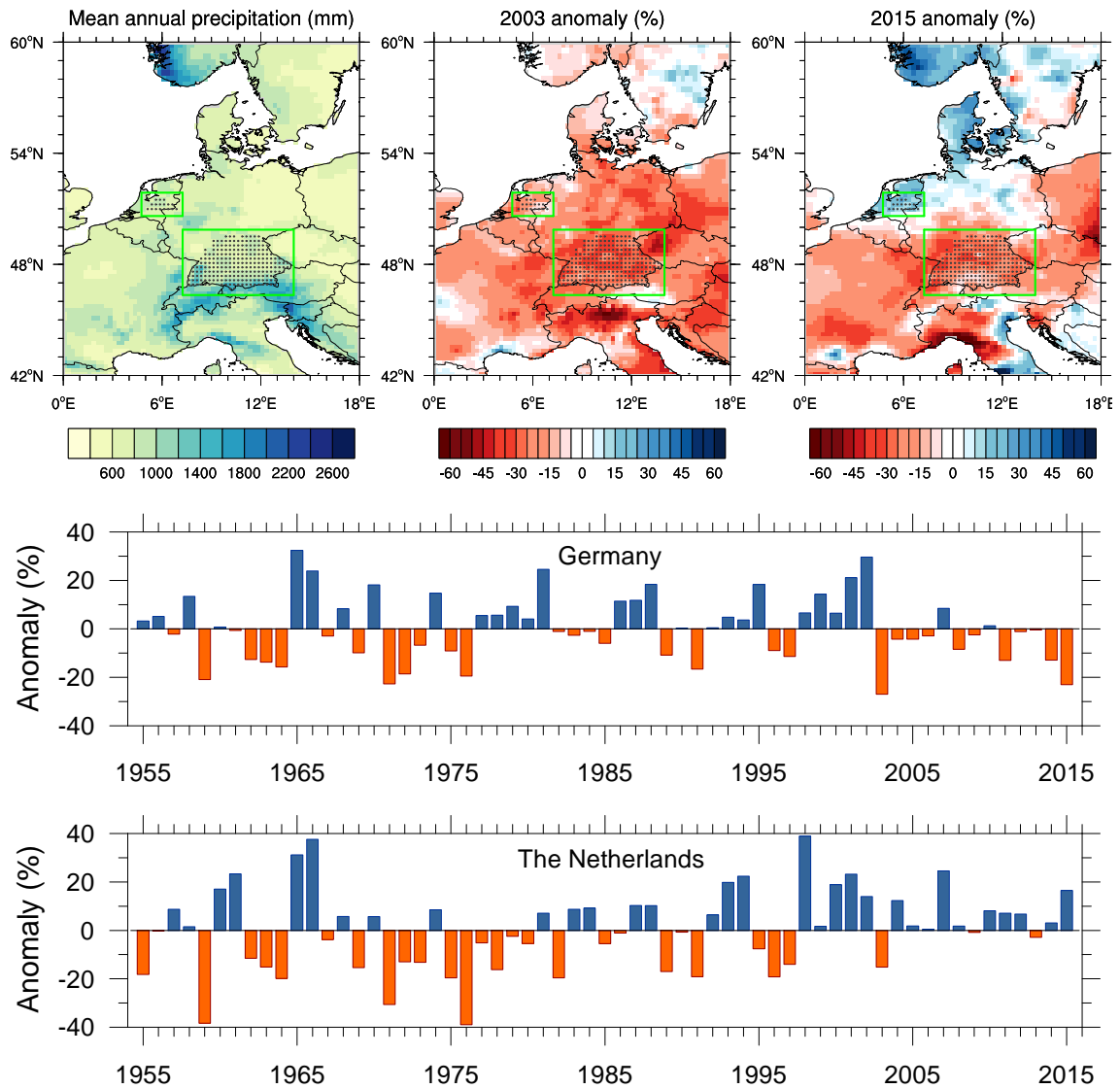


Figure 1. Long-term mean annual precipitation (1950-2015) and precipitation anomalies in the years 2003 and 2015 (% difference from long-term mean) for our study domain in Europe. The green boxes indicate the study regions and the black dots represent the grid cells used for the analysis. The bottom two panels show the annual precipitation anomalies (% deficit from the long term mean) for the two study regions located in southern Germany and the eastern Netherlands.

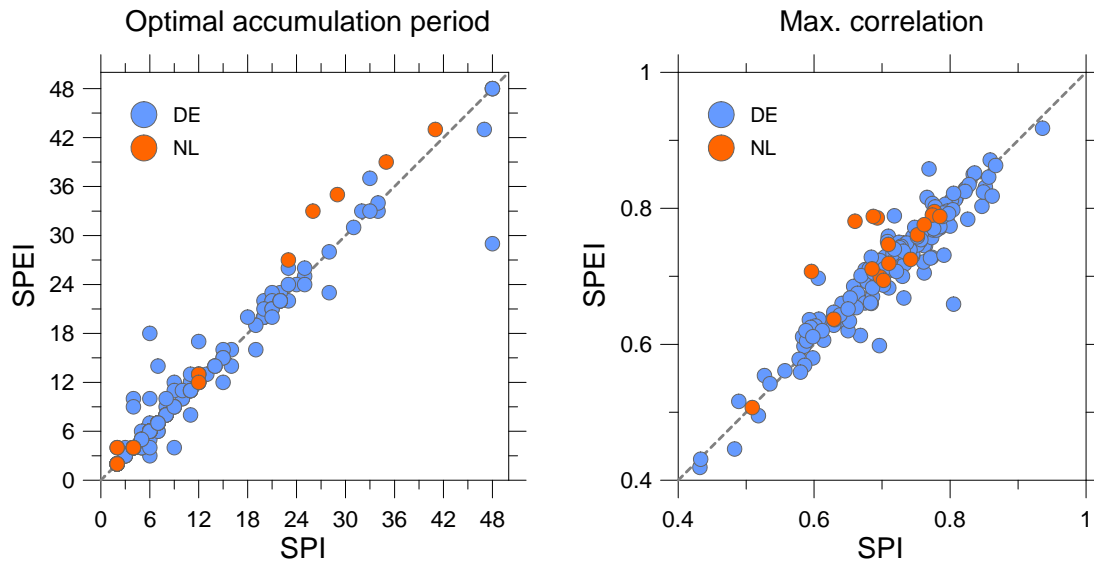
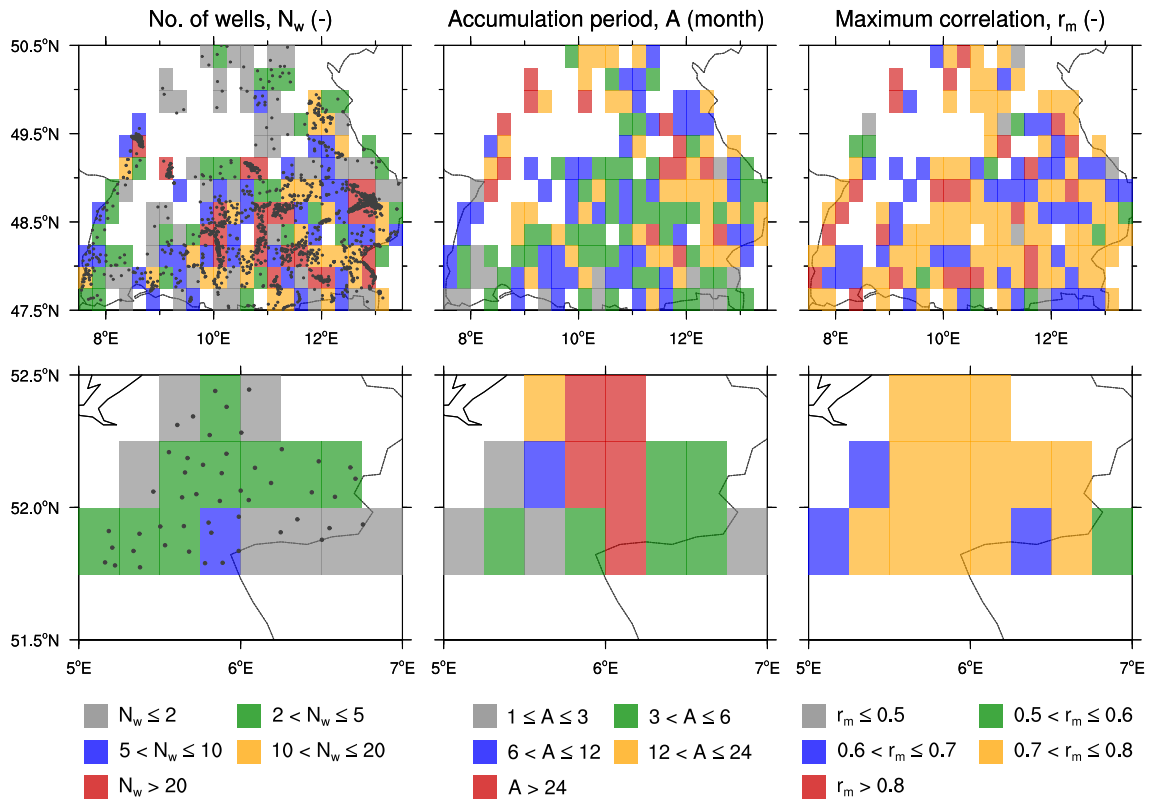


Figure 2. The location of groundwater wells and number of groundwater wells in every 0.25° grid cell, the optimal accumulation A (month) and the maximum correlation $r_m(-)$ between the gridded SGI and SPEI for German (top panel) and Dutch (middle panel) data sets. The bottom panels show the correspondence of A and r_m estimated from (SPEI vs. SGI) and (SPEI vs. SGI).

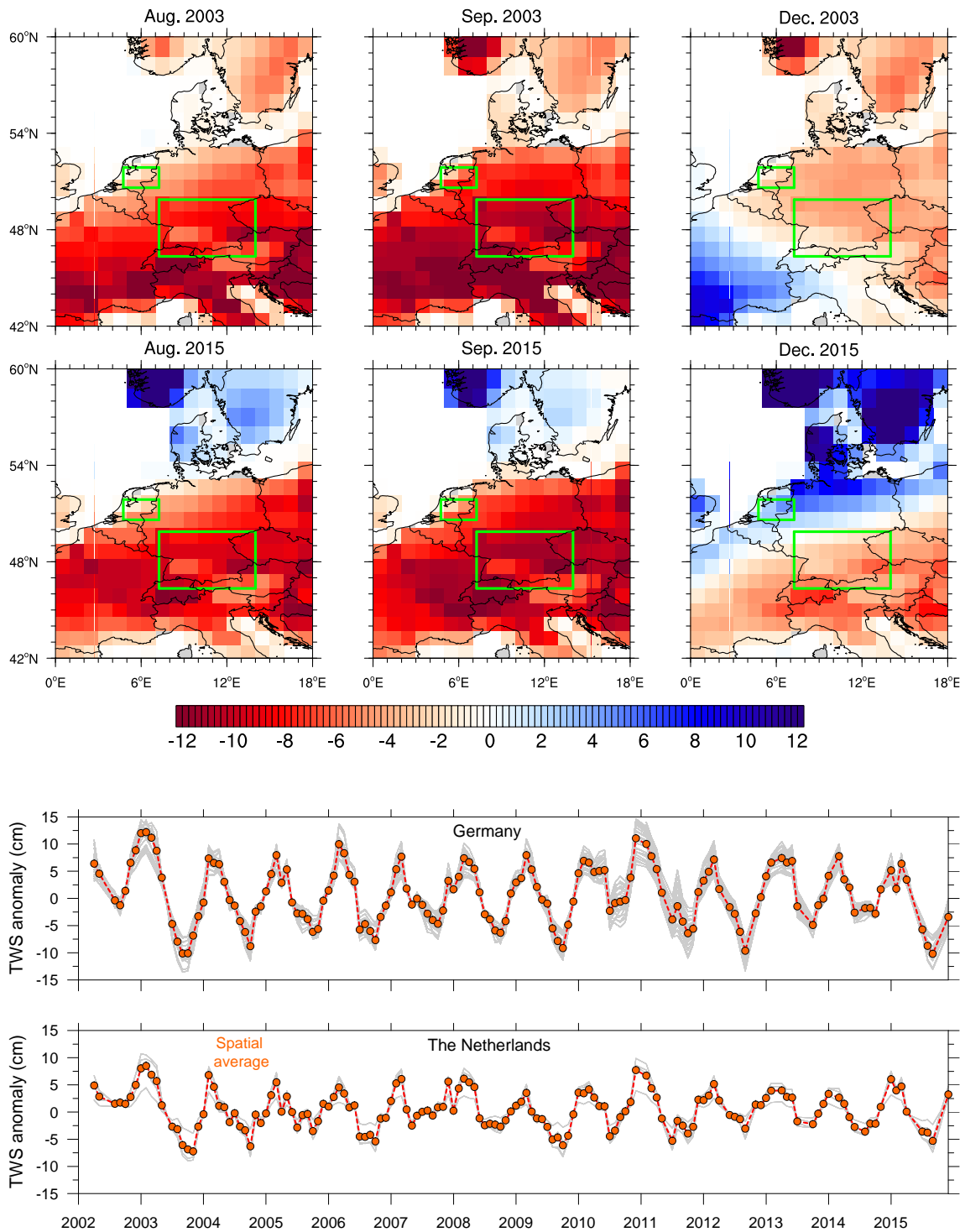


Figure 3. The GRACE derived Terrestrial Water Storage (TWS) anomalies (cm) for the available three common months (August, September, and December) during the 2003 and 2015 drought events (top panel). The bottom two panels depict the monthly TWS anomalies for every grid cell located within the study domains (shown in the green boxes in the top panels; south Germany and central Netherlands) as well as their spatial averages.

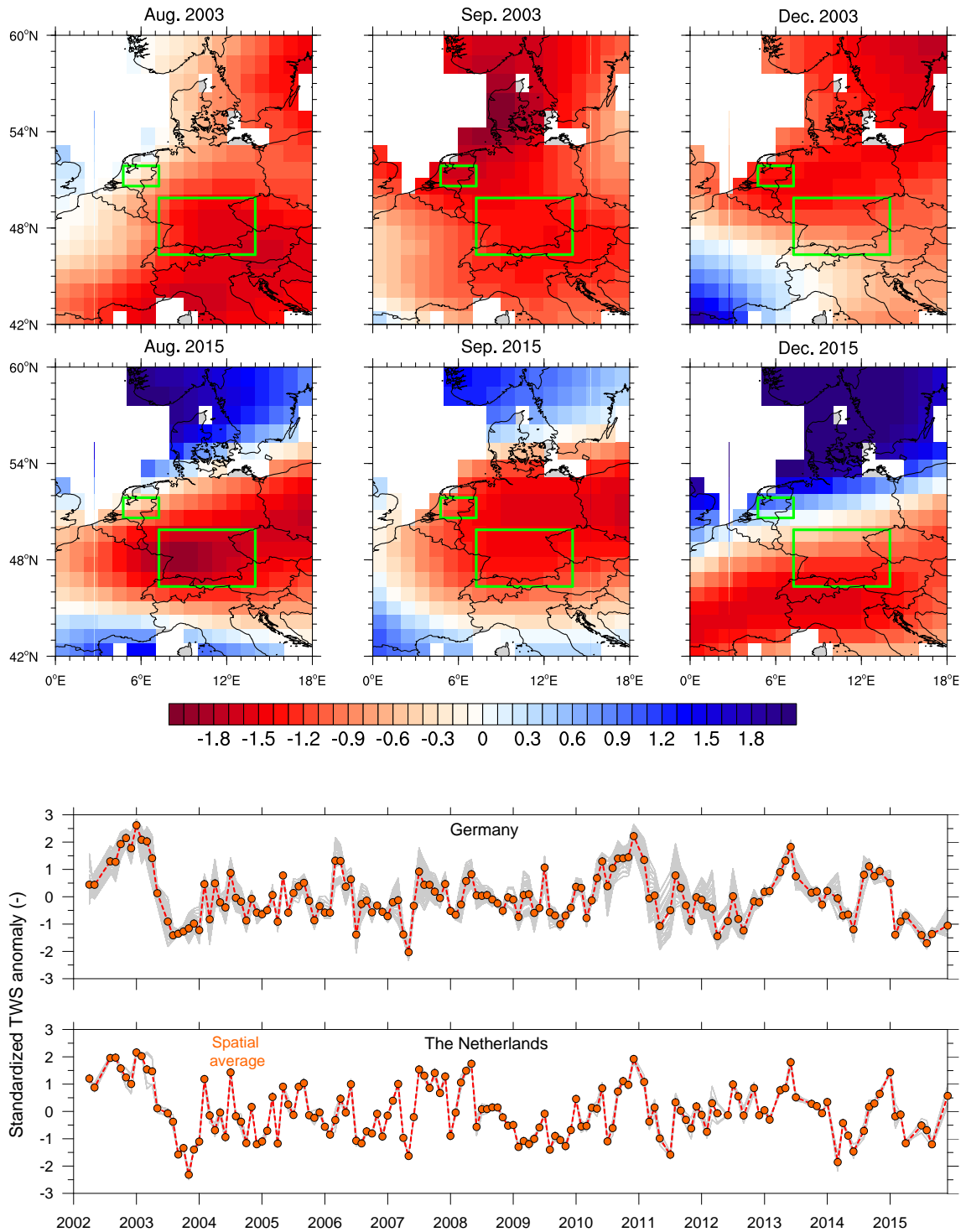


Figure 4. Same as Fig. 3, but after standardisation of the TWS anomalies to de-emphasise the seasonality component of TWS anomalies.

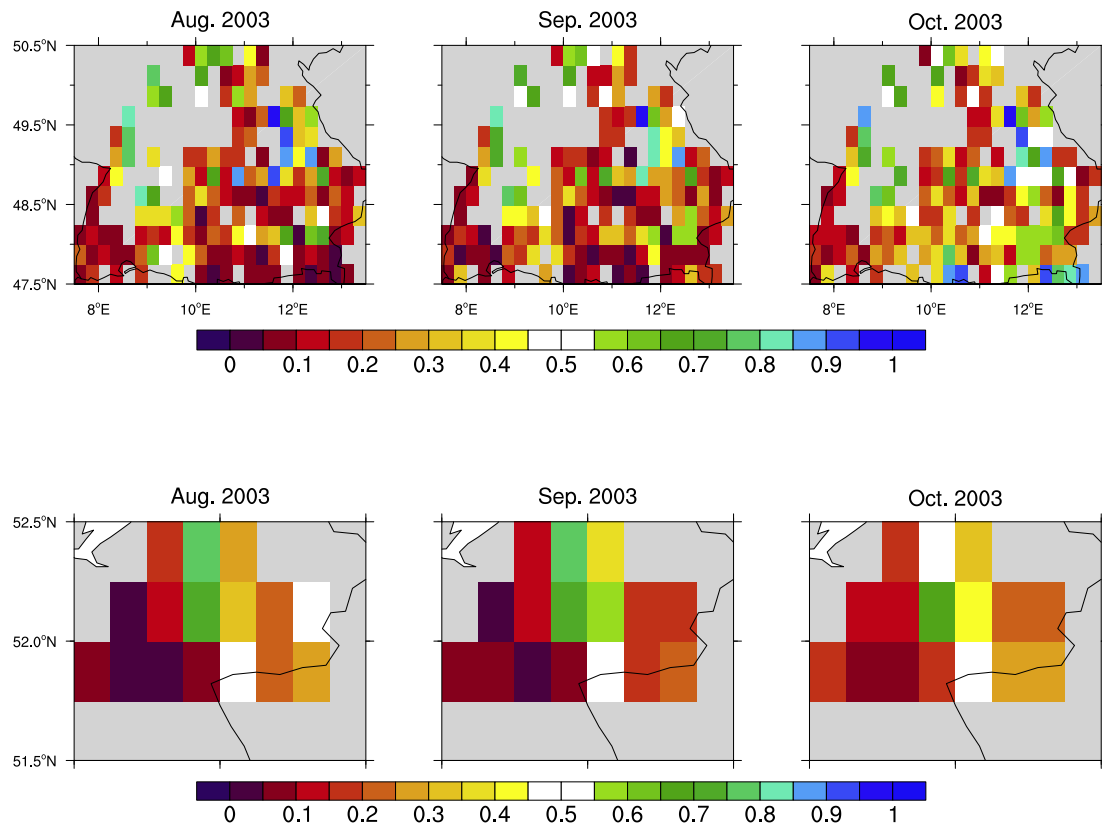


Figure 5. The groundwater drought (SGI) during the months of August, September, and October for the 2003 drought event over the south Germany and east Netherlands study domains, derived from borehole observations and recalculated to 0.25° grid cells for comparability.

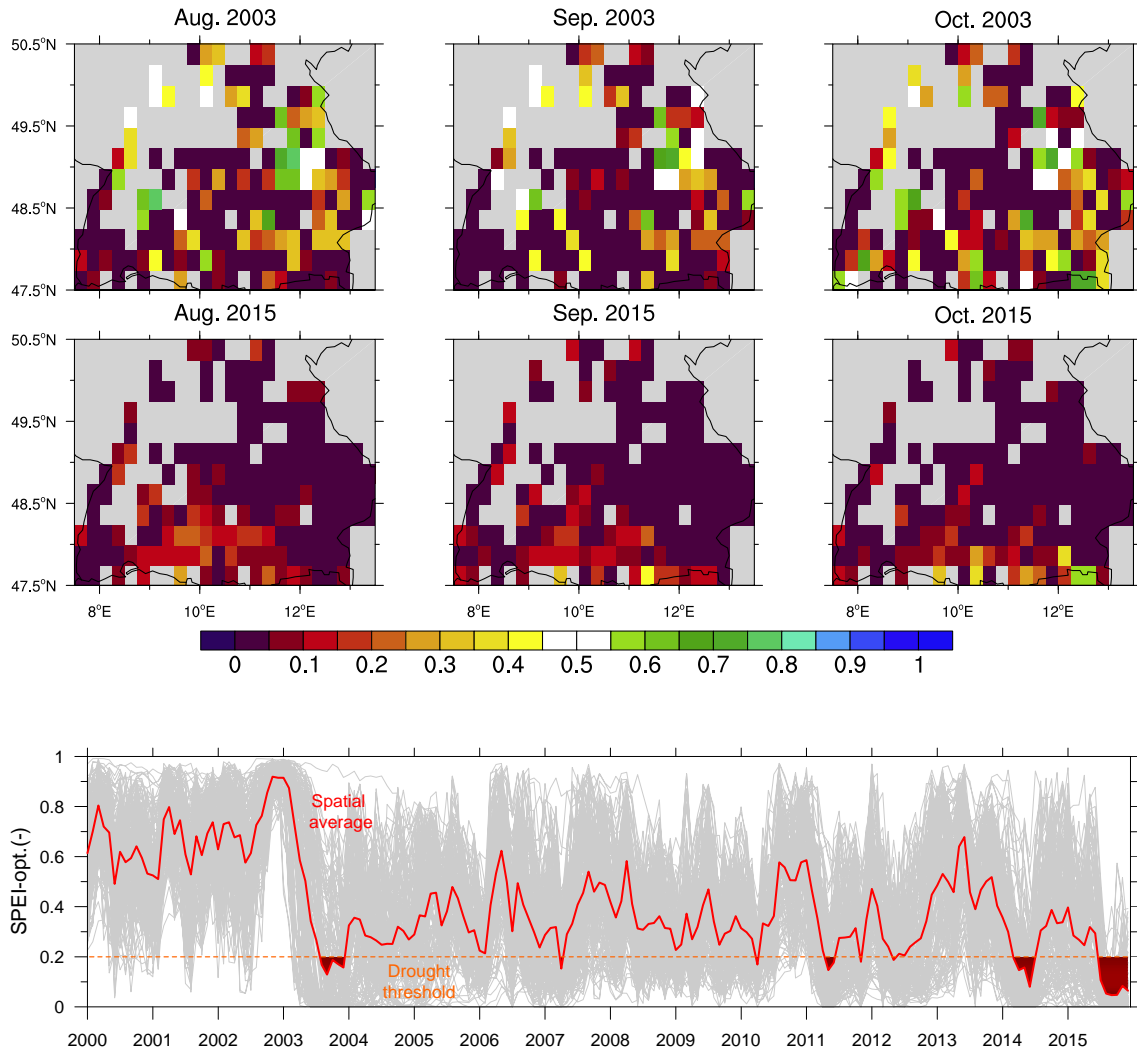


Figure 6. The reconstructed groundwater drought (SGI) based on the SPEI with spatially varying (optimal) accumulation periods during the months of August, September, and October for the 2003 and 2015 drought event over the south Germany study domain. The monthly time series of the optimally accumulated SPEI for every 0.25° grid cell and the respective spatial averages (dark red) are depicted in the bottom panel. The orange dashed line depicts the drought threshold (τ) of 0.2.

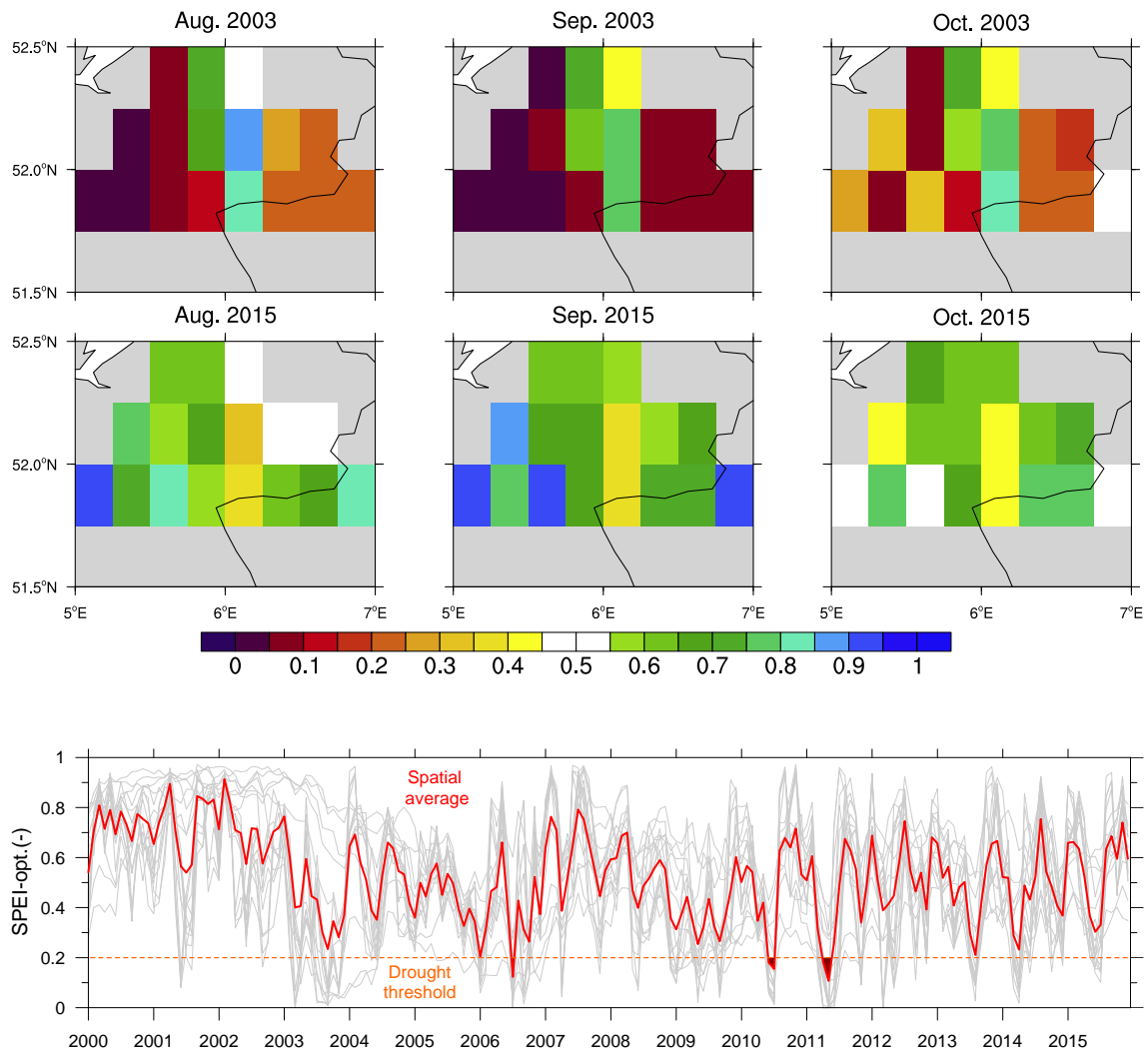


Figure 7. Same as Fig. 6, but for the central Netherlands study domain.

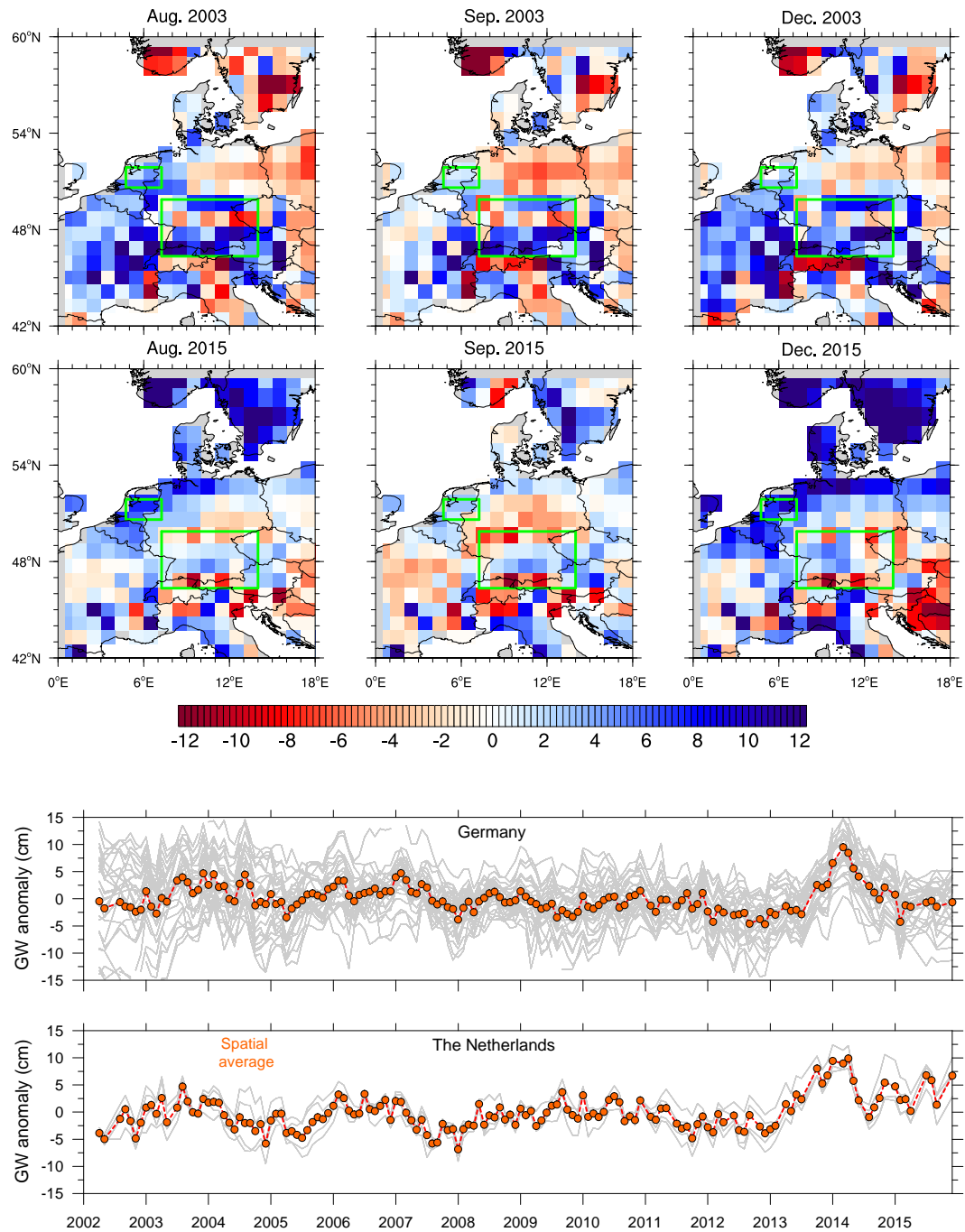


Figure 8. Groundwater anomalies (cm) based on GRACE and GLDAS model outputs for the available three common months (August, September, and December) during the 2003 and 2015 drought events (top panel), estimated by removing the total surface water storage anomaly (TOW = Soil moisture + Canopy Storage + Snow water equivalent) from the GRACE-TWS anomaly (Fig. 3). The TOW estimates are based on mean simulations from four GLDAS models (CLM, NOAH, VIC and MOSIAC). The plots are based on the ensemble mean of these four model outputs. The bottom two panels depict the monthly GRACE-GLDAS groundwater anomalies for every grid cell located within the study domains (shown in the green boxes in the top panel; south Germany and central Netherlands) as well as their spatial averages.

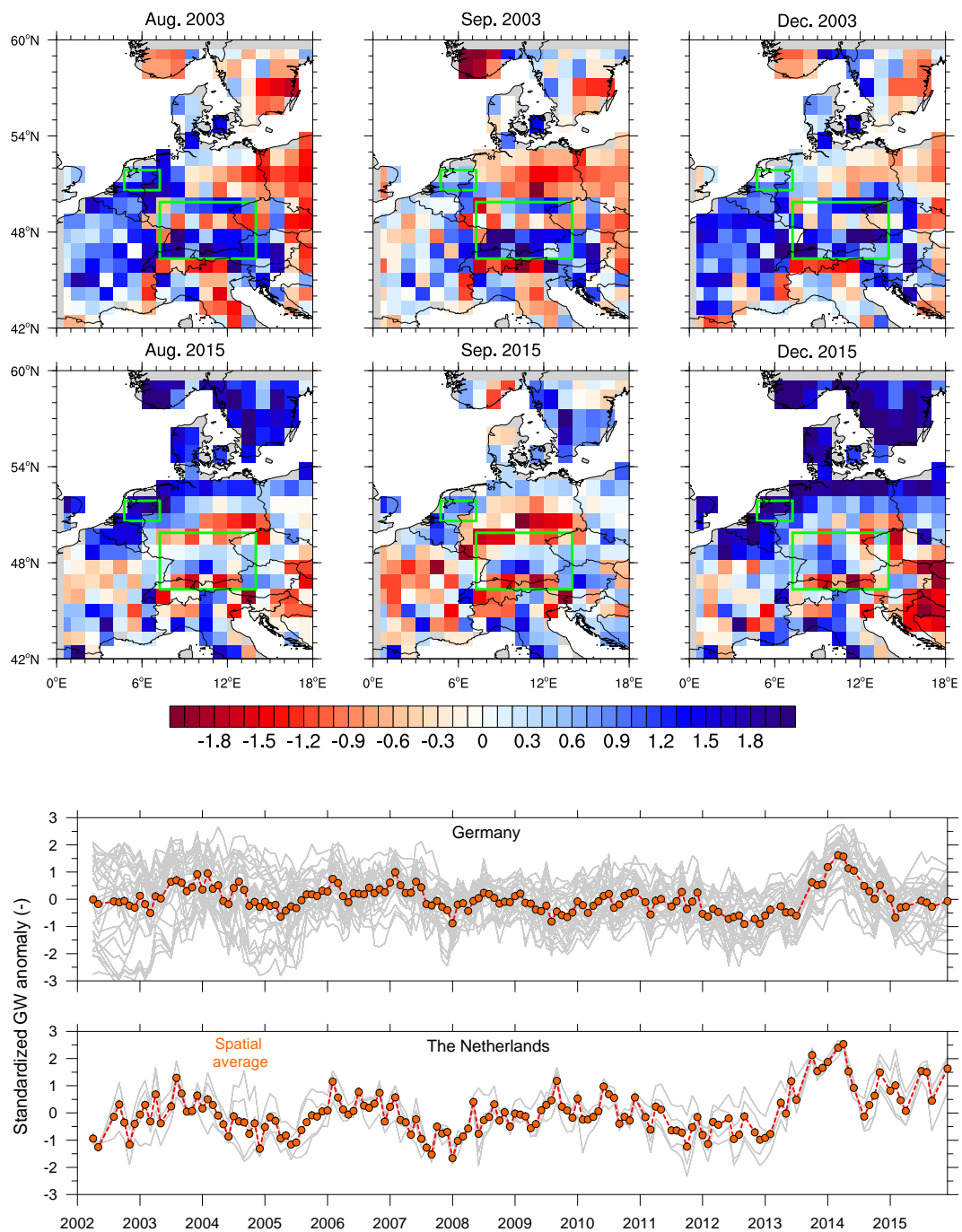


Figure 9. Same as Fig. 8, but after standardisation of the GRACE-GLDAS groundwater anomalies to de-emphasise the seasonality component of the anomalies.

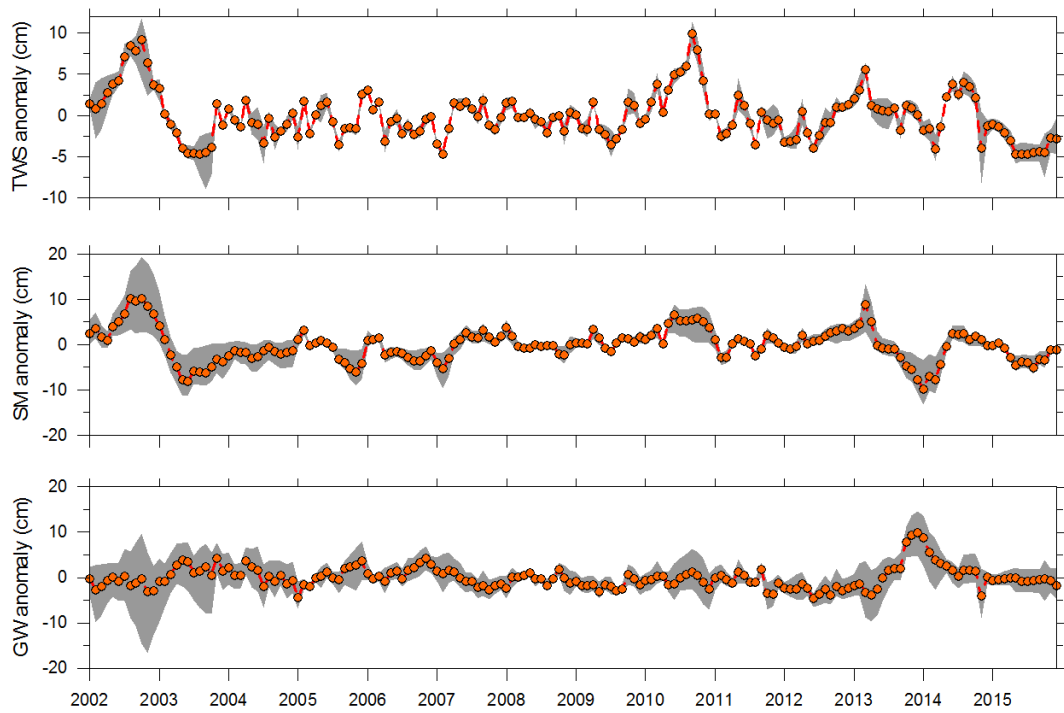


Figure 10. Ensemble mean and range (min-max; shaded) of GRACE-TWS (from JPL, CSR, GFZ), simulated soil moisture (SM; from GLDAS models VIC, CLM, Noah, and MOSAIC), and resulting groundwater (GW; from 12 combinations, i.e. three TWS products and four models) area-averaged anomalies for Germany for the period 2002–2015.

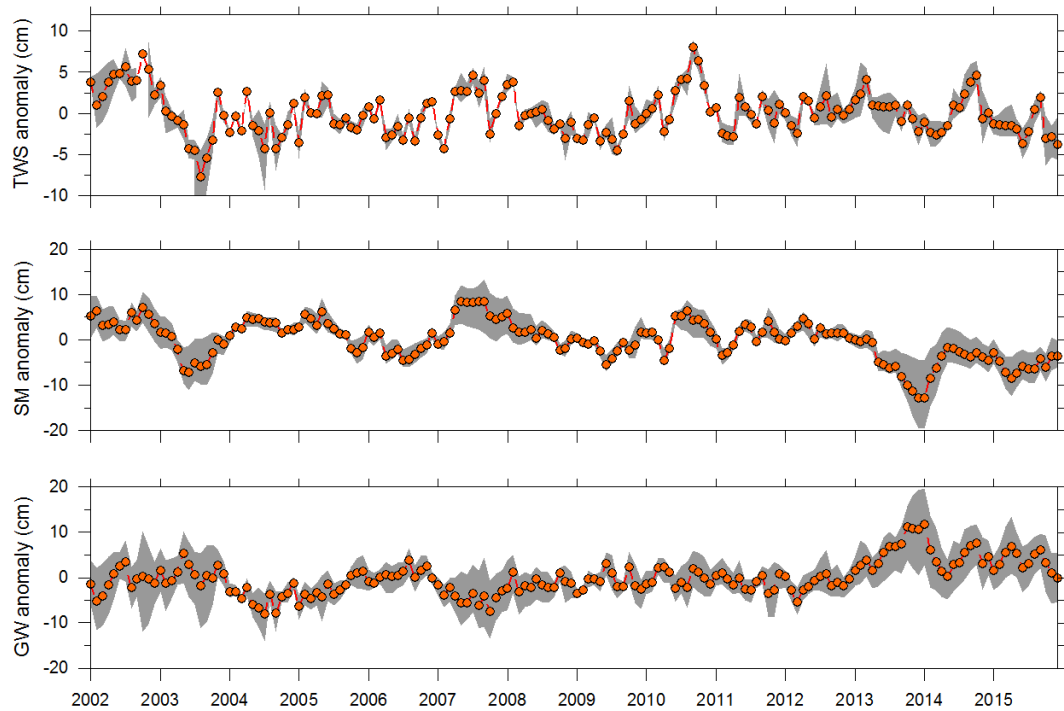


Figure 11. Same as Fig. 10 but for the study region in the Netherlands.

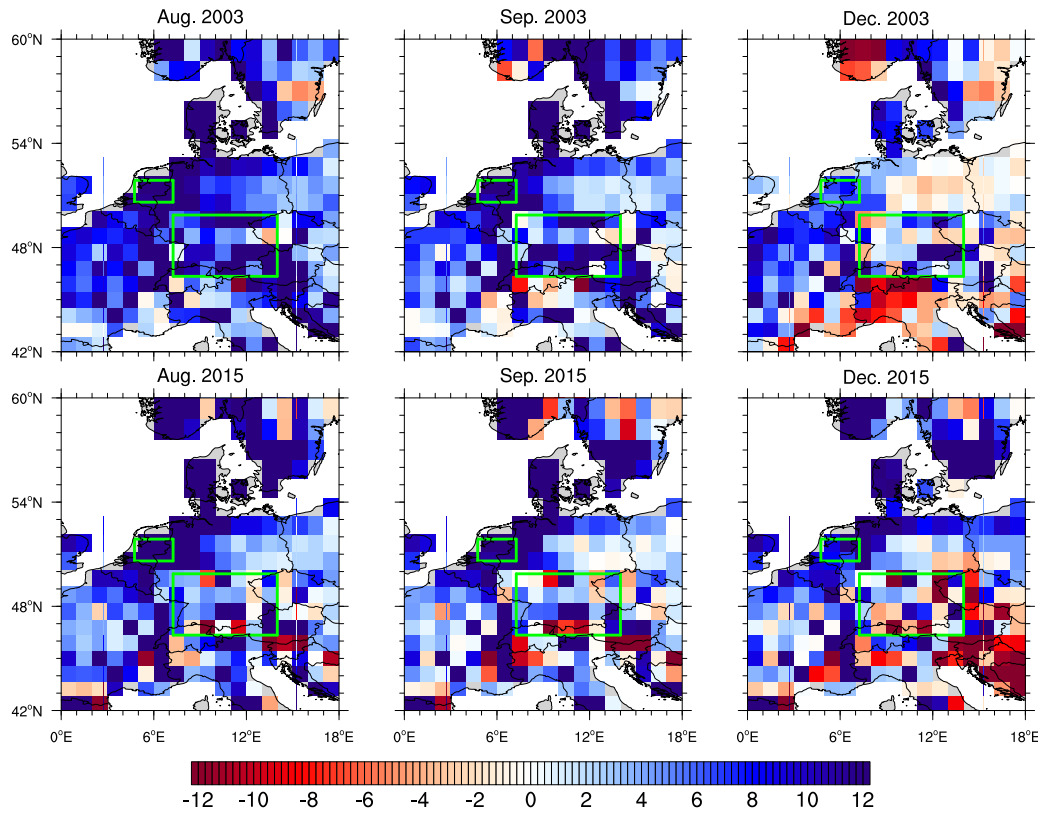


Figure A1. Groundwater anomalies (cm) based on GRACE and VIC model outputs for the available three common months (August, September, and December) during the 2003 and 2015 drought events (top panel), estimated by removing the total surface water storage anomaly (TOW = Soil moisture + Canopy Storage + Snow water equivalent) from the GRACE-TWS anomaly (Fig. 3).

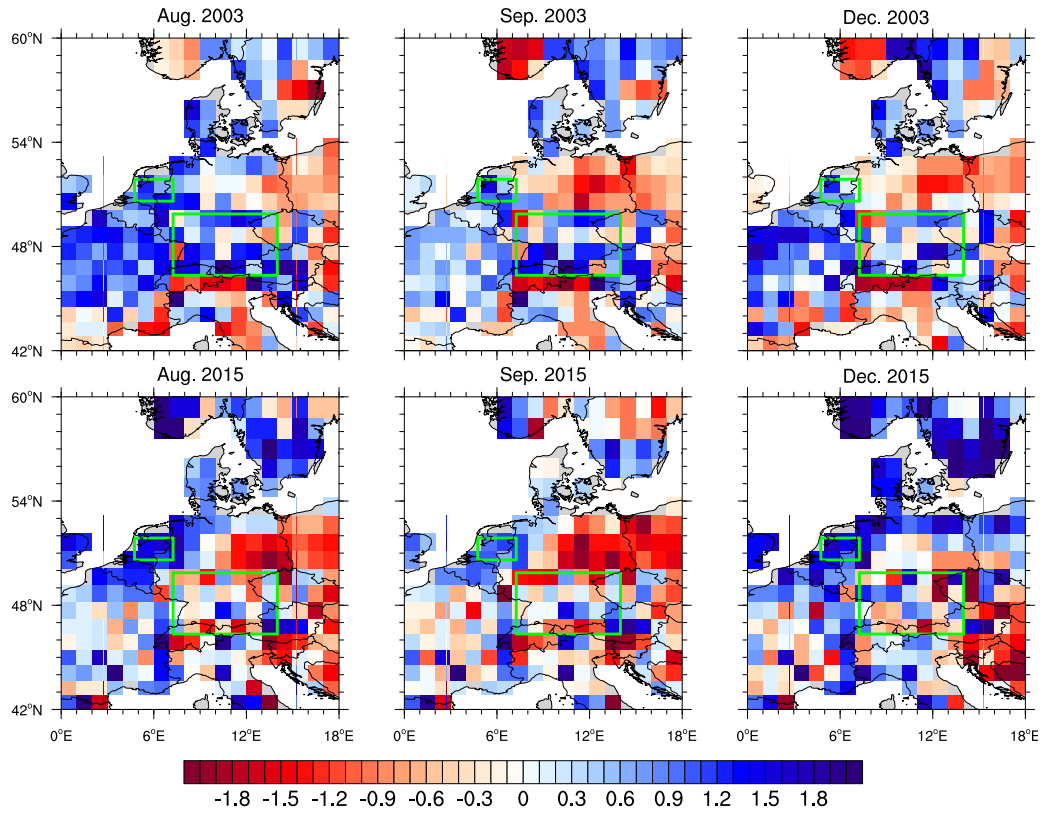


Figure A2. Same as Fig. A1, but after standardisation of the GRACE-VIC groundwater anomalies to de-emphasise the seasonality component of the anomalies.

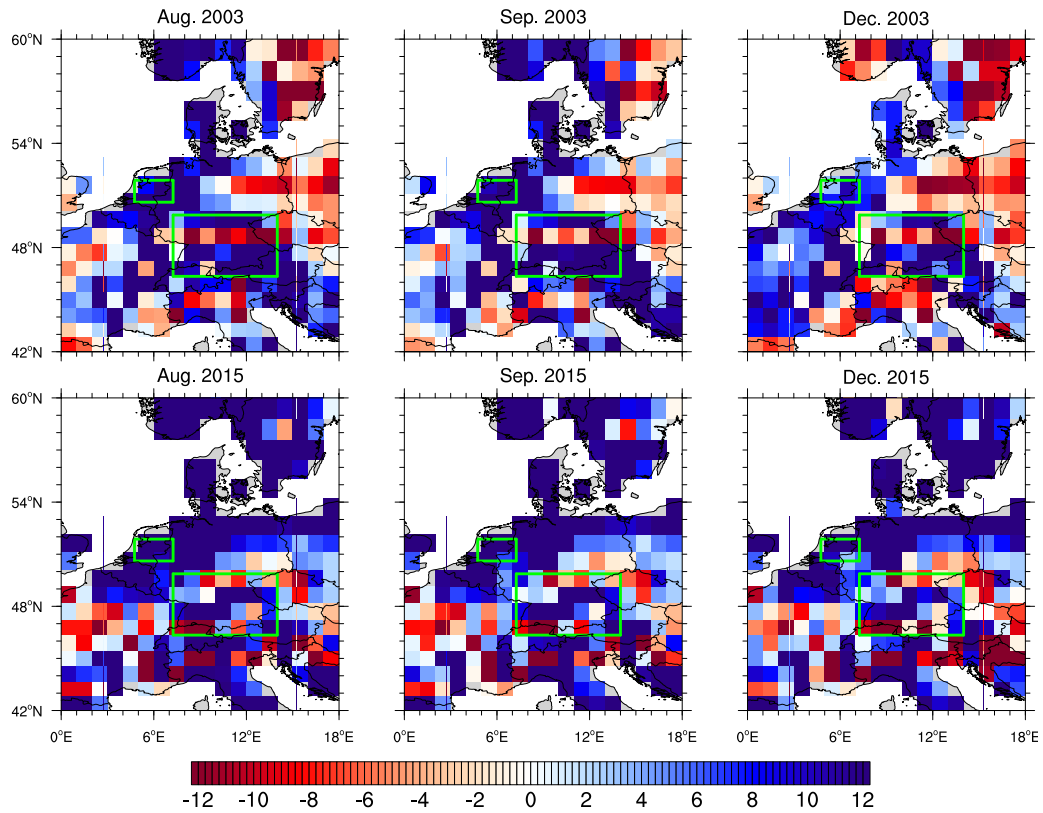


Figure A3. Groundwater anomalies (cm) based on GRACE and MOSAIC model outputs for the available three common months (August, September, and December) during the 2003 and 2015 drought events (top panel), estimated by removing the total surface water storage anomaly (TOW = Soil moisture + Canopy Storage + Snow water equivalent) from the GRACE-TWS anomaly (Fig. 3).

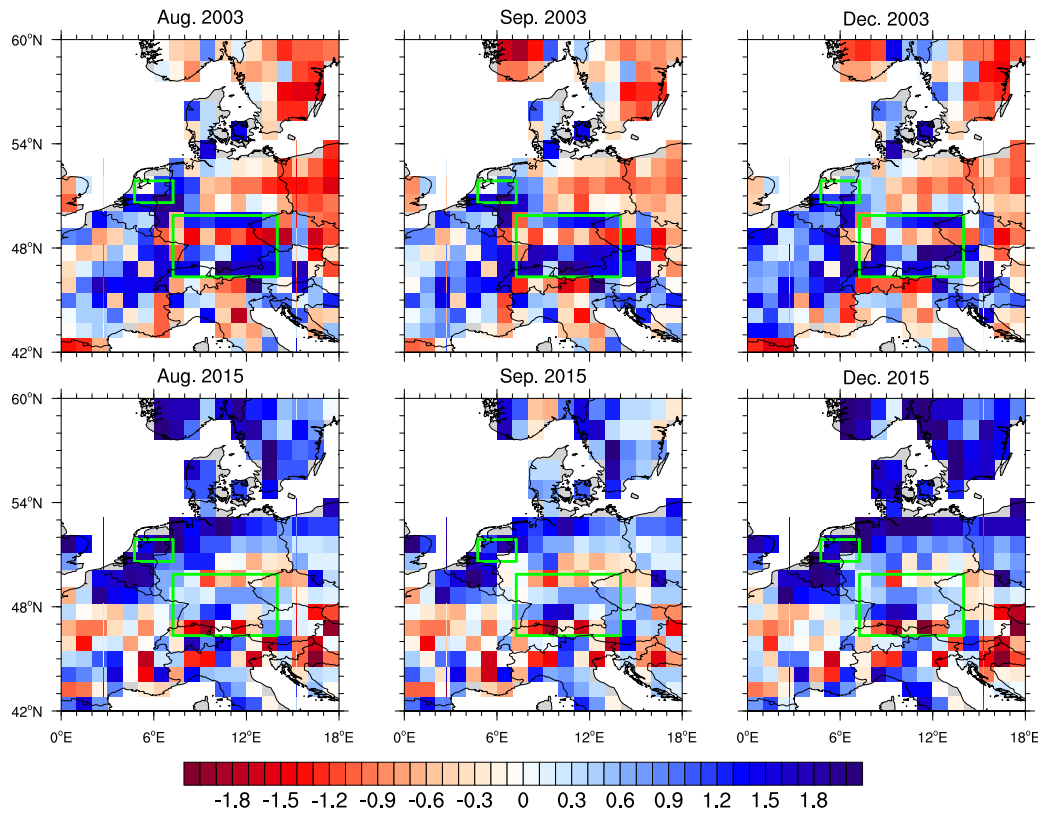


Figure A4. Same as Fig. A1, but after standardisation of the GRACE-MOSAIC groundwater anomalies to de-emphasise the seasonality component of the anomalies.

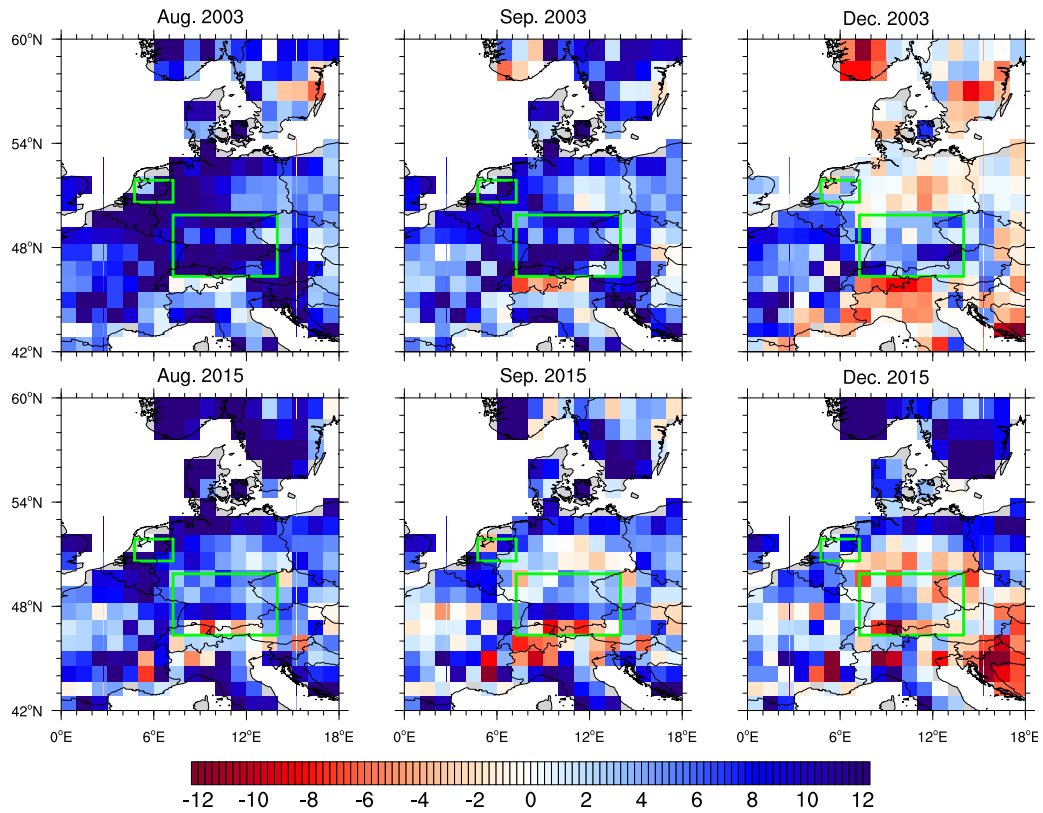


Figure A5. Groundwater anomalies (cm) based on GRACE and NOAH model outputs for the available three common months (August, September, and December) during the 2003 and 2015 drought events (top panel), estimated by removing the total surface water storage anomaly (TOW = Soil moisture + Canopy Storage + Snow water equivalent) from the GRACE-TWS anomaly (Fig. 3).

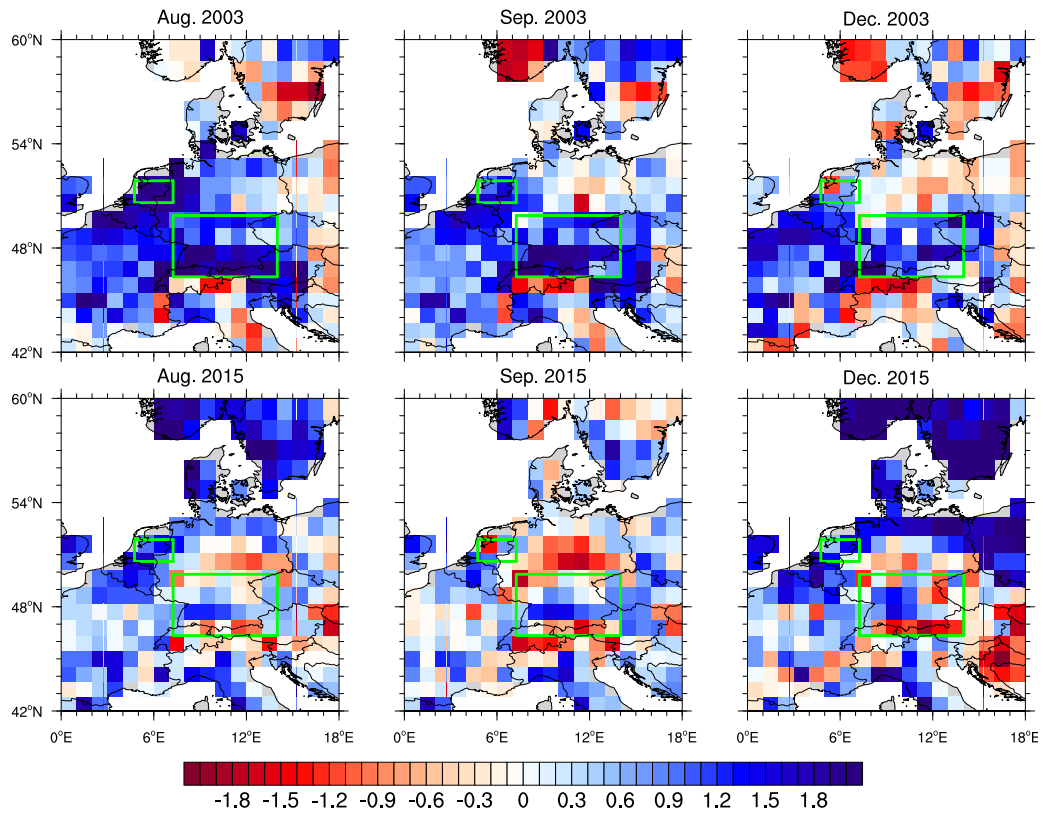


Figure A6. Same as Fig. A1, but after standardisation of the GRACE-NOAH groundwater anomalies to de-emphasise the seasonality component of the anomalies.

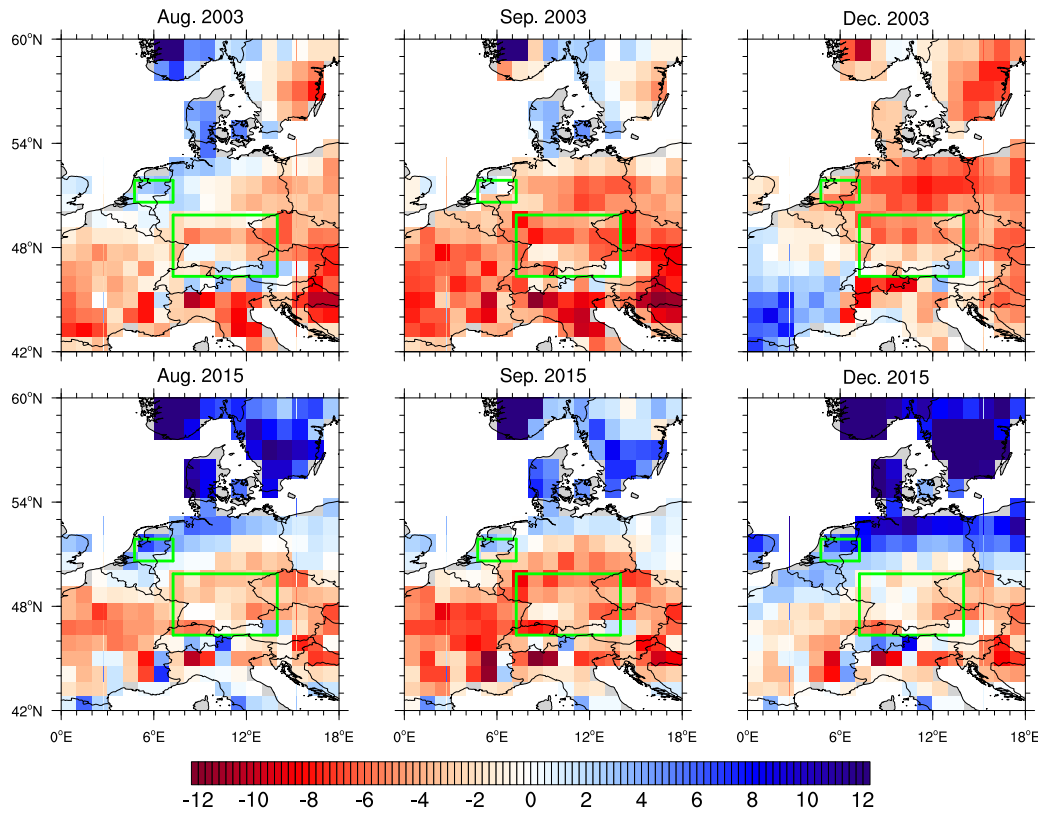


Figure A7. Groundwater anomalies (cm) based on GRACE and CLM model outputs for the available three common months (August, September, and December) during the 2003 and 2015 drought events (top panel), estimated by removing the total surface water storage anomaly (TOW = Soil moisture + Canopy Storage + Snow water equivalent) from the GRACE-TWS anomaly (Fig. 3).

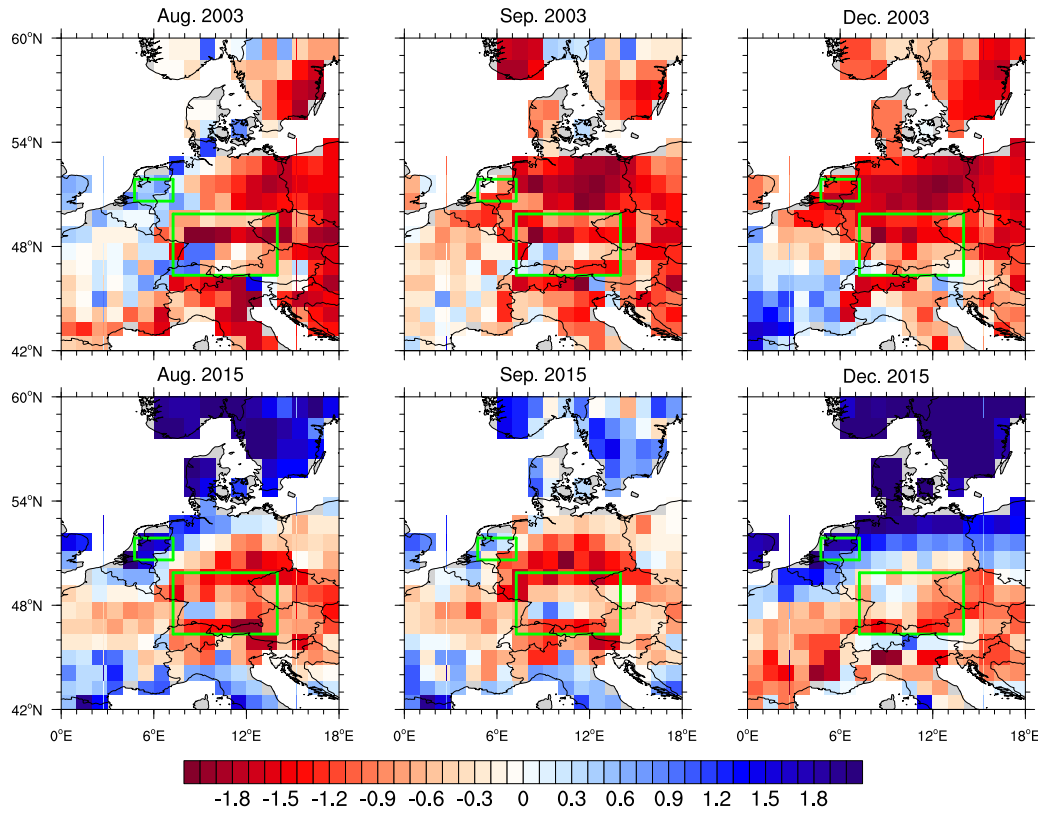


Figure A8. Same as Fig. A1, but after standardisation of the GRACE-CLM groundwater anomalies to de-emphasise the seasonality component of the anomalies.