# Reviewers comment #1

Helena Gerdener, Olga Engels and Jürgen Kusche

The author's responses are indicated in red color, as well as old text passages. New text passages are indicated in green color.

The authors developed a framework for deriving synthetic terrestrial water storage change (TWSC) from the GRACE observations for computing drought indicators. The synthetic TWSC enables comparisons of existing drought indicator methods and analyses of the influence of GRACE trend and noise on the drought detection. I think that the topic is important for further understanding of hydrological drought and the findings are valuable, shedding lights to characteristics of different drought identification methods. The manuscript is fairly well written; however, I have some concerns and need some clarifications. Thus, I recommend major revision and the specific comments are listed below.

#### Response:

Thank you very much for the time spent in reviewing the manuscript and for the really useful reviewer comments, which certainly helped us to improve the manuscript.

### Comment 1.

The authors chose three existing indicators of Zhao et al (2017), Houborg et al (2012), and Thomas et al (2014) methods because they are based on the monthly GRACE data. The comparison of the methods is interesting, but I don't think this framework is a fair ground for evaluating their skills, especially for the Houborg-method. As I understand it, the CDF (which is the basis for the percentile computation) is based on the historical simulation of 1948-2010. This analysis focuses on the GRACE period of 2002 to 2016 and it mentions about disregarding the bias correction. Also for the Zhaomethod, a bias correction is not applied. I understand that direct comparison of these indicators are not possible as Houborg is regional, but the indicators in their final product form (as opposed to the method concepts) may be able to detect the drought that were missed in this study.

#### Response:

Thanks, we understand the reviewer's point. However, we believe the primary aim of this paper is to provide a fair comparison of the indicators based on a synthetic environment, which is derived by computing simulated TWSC during the GRACE period. Within this controlled synthetic environment the fairest comparison for the Houborg-method is a comparison without using a bias correction. We agree with the reviewer that a bias correction would be appropriate when considering real observations of TWSC.

We believe that this issue would require further discussion which we cannot provide here. Our simulations confirm that without applying any correction, indeed the limited duration of the synthetic time series may render the computation of the biased indicator. Nonetheless, the bias correction as suggested in Houborg's paper would have to come from a cumulative distribution function (CDF) derived from long runs of hydrological models, and these are far from representing reality as new studies show (e.g. Scanlon et al., 2018).

On balance, as our main focus is on the synthetic environment, we prefer to keep our current indicator computation, but we modified a sentence in the description of the Houborg-method, to precisely state our focus.

#### Old text:

Here, we focus on a TWSC from GRACE only and, as explained in Sec. 2.1, we therefore disregard the bias correction.

#### New text:

Here, we focus on a simulated TWSC environment for the GRACE period only and, as explained in Sec. 2.1, we therefore disregard the bias correction.

#### Comment 2.

Is the GRACE-specific noise dependent on the instrument or the solution? As it is an important term and needs to be characterized well, I am wondering if it would be different when using different GRACE-TWSC solutions such as mascon solutions from JPL or CSR. Is the same approach (equation 21) applicable to other GRACE data? What is the grid size of the TU GRAZ data (0.5 degree)?

#### Response:

Even after 17 years of GRACE data, a full understanding of the GRACE noise characteristics, let alone of the individual sources, has not been reached. The noise in the GRACE solutions depends on the GRACE orbit configuration, on the instrument performance (which changed significantly over time due to technical issues such as the switch-off of the thermal stabilization of the accelerometers in 2010), of the realism of the background models, and on the data editing and estimation strategy itself which differs between institutions. One could either use a diagonal or a non-diagonal solution variance-covariance matrix to describe the noise model. By accounting for a non-diagonal solution variance-covariance matrix, the noise model accounts for latitudinal variation of noise levels dependency due to orbit convergence. However certain errors like the noise introduced by background model errors are difficult to know and, currently, there is no way of accounting for them. So the short answer is one would probably be able to work with the same error characterization for other GRACE solutions.

However, the use of the mascon solutions creates another difficult problem; the mascon solution exhibit a better S/N ratio as compared to the conventional solutions but this is to a large extent due to the fact that these solutions use constraints derived from geophysical models, and it would be difficult to characterize the biases introduced by these constraints.

Here, we use the TU GRAZ solutions that are provided in monthly geopotential coefficients (spherical harmonics, SH), this means the data is not given in the spatial domain. We then transform these coefficients by using spherical harmonic synthesis to monthly TWSC grids (here we use 0.5 degree grid). These grids inherit the native GRACE resolution which is somewhere about 300 km. The geopotential coefficients can also be derived by other processing centers, for example CSR, GFZ and JPL. Therefore, we could also apply our approach on these data.

The reviewer raises a very important point about the significance of a proper noise description. The SHs used to compute TWSC are provided along with corresponding standard deviations (ftp://ftp.tugraz.at/outgoing/ITSG/GRACE/ITSG-Grace2016/monthly/monthly\_n60/). In the submitted version, we propagated this information to a grid, which led to a full variance-covariance matrix (used in Eq. 21) for the TWSC. Following the reviewer comment, we now use a full variance-covariance matrix (normal equations provided by TU GRAZ: ftp://ftp.tugraz.at/outgoing/ITSG/GRACE/ITSG-Grace2016/monthly/monthly\_n90\_normals/) of the SHs to estimate the full variance-covariance matrix of the TWSC. This procedure better represents the GRACE-specific noise, because the correlations between the SHs are taken into account. Thus, at the moment the full variance-covariance matrix of the SHs is in our opinion the best solution available to describe the GRACE-specific noise.

After this adjustment, some passages, values and figures have been slightly changed, but the interpretation of all derived results remained exactly the same. These changes are mentioned below under the section "Changes in the noise term".

### Comment 3.

[Page4;Ln9-18] The equations 1-3 are not referred later in the text. I agree that TWSC corresponds to precipitation anomaly accumulation in many cases, but it does not seem to tie in with the rest of the discussions.

#### Response:

Following the reviewers comment, we reference the Eq. 1 that is used to better describe the relation of Eq. 2 (Page4;Ln16). The Eq. 2 and 3 are referred to accumulated (Page5;Ln21-24) and differenced (Page 5-6;Ln26-3) TWSC, correspondingly. For the sake of completeness and to avoid any misunderstandings regarding the connection between fluxes and storages, we would like to keep these equations.

#### Comment 4.

[Page9;Ln5-7] Identifying regional clusters seems very important and I wonder where else clusters are located.

#### Response:

We agree this needs an additional figure, which we added to the Appendix (B1) and adjusted the text of the manuscript correspondingly.

#### Old text1:

As a result of this procedure, we chose three clusters located in East Brazil (EB), South Africa (SA), and West India (WI), which were also affected by droughts in the past (e.g. Parthasarathy et al., 1987; Rouault and Richard, 2003; Coelho et al., 2016).

New text1:

As a result of this procedure, we identified three clusters located in East Brazil (EB), South Africa (SA), and West India (WI), which were indeed affected by droughts in the past (e.g. Parthasarathy et al., 1987; Rouault and Richard, 2003; Coelho et al., 2016). Location and shape of the three chosen clusters are shown in Fig. 3, and a global map of all clusters is provided in Fig. B1.



Figure B1. Clusters based on Expectation Maximization (EM) clustering applied to the global autoregressive model (AR) coefficients.

#### Comment 5.

[Page9;Ln16-17] It would be helpful to present the list of droughts included in step 3, in a table or supplement.

#### Response:

Thanks, for the suggestion. We added a table to show the considered TWSC period for the corresponding drought periods.

#### Old text:

Searching for drought duration and magnitude (step 3) led to four droughts seen in GRACE-TWSC: The 2005 and 2010 droughts in the Amazon (e.g. Chen et al., 2009; Espinoza et al., 2011), the 2011 drought in Texas (e.g. Long et al., 2013), and the 2003 drought in Europe (e.g. Seitz et al., 2008).

#### New text:

Performing literature research for drought duration and magnitude (step 3) led to four droughts seen in GRACE-TWSC (Tab. 4): The 2005 and 2010 droughts in the Amazon (e.g. Chen et al., 2009; Espinoza et al., 2011), the 2011 drought in Texas (e.g. Long et al., 2013), and the 2003 drought in Europe (e.g. Seitz et al., 2008).

Table 4. Drought events in Europe, Amazon river basin and Texas with corresponding duration taken from literature.

Region	Year of drought	Considered TWSC months	Examples of literature		
Europe	2003	June to August	Andersen et al. (2005)		
			Rebetez et al. (2006)		
			Seitz et al. (2008)		
Amazon river basin	2005	May to September	Chen et al. (2009)		
			Frappart et al. (2012)		
	2010	June to September	Espinoza et al. (2011)		
			Frappart et al. (2013)		
			Humphrey et al. (2016)		
Texas	2011	February to October	Humphrey et al. (2016)		
			Long et al. (2013)		

### This table contains two new references, which is added to the reference list as follows:

Frappart, F., Papa, F., Santos da Silva, J., Ramillien, G., Prigent, C., Seyler, F. and Calmant, S.: Surface freshwater storage and dynamics in the Amazon basin during the 2005 exceptional drought, Environmental Research Letters, 7(4), 044010, doi:10.1088/1748-9326/7/4/044010, 2012.

Rebetez, M., Mayer, H., Dupont, O., Schindler, D., Gartner, K., Kropp, J. P. and Menzel, A.: Heat and drought 2003 in Europe: a climate synthesis, Annals of Forest Science, 63(6), 569–577, doi:10.1051/forest:2006043, 2006.

### Comment 6.

[Page10;Ln20] It was not clear to me if the study addressed the last point of this framework's benefit "... identify the most suitable indicator for a specific application".

Response:

We thank the reviewer for this comment. There is a large number of different hydrological regimes for which the TWSC-based indicators would show very different results as soon as e.g., trends are contained in the TWSC time series. Unfortunately, we are not able to elaborate all these applications within this manuscript. With this last point we would like to explain that our aim is to identify strengths and weaknesses of different indicators using our synthetic framework. The identified strengths and weaknesses allow us to decide which indicator might be the most suitable ones (or is not recommended) for a particular application. For example, if the drought period is much shorter than the analyzed TWSC time span and the observations contain a pronounced? trend, we would encourage using indicators like e.g. DSID6 based on the results of our synthetic framework (Sec. 4.1). We modified the corresponding sentence.

### Old text:

...; iii) comparing different indicator outputs allows us to identify the most suitable indicator for a specific application.

### New text:

...; iii) the synthetic framework enables us to identify strengths and weaknesses of each analysed indicator, and thereby enables us to choose the most suitable indicator for a specific application.

### Comment 7.

[Figure 5] I understand that the purpose of this figure is to show the importance of spatial GRACE noise, especially in SA. However, the TWSC time series for EB and WI have low TWSC amplitudes that are equally as low as that of during the simulated drought period in later 2016 (EB) and 2003/2004 (WI). Can you add an explanation to what distinguishes 2005 from these low TWSC?

### Response:

Yes, the reviewer is right. The synthetic TWSC for the cluster located in East Brazil (EB) are in later 2016 as low as the TWSC within the simulated drought period in 2005. The same concerns the synthetic TWSC for the cluster located in West India (WI). The synthetic TWSC for WI are in 2003/2004 as low as the TWSC within the simulated drought period. These low TWSC for EB in later 2016 can be explained by the negative acceleration, which was used to generate the synthetic time series. In contrast, the low TWSC for WI in 2003/2004 are based on a positive trend, which has a strong influence here. We do not discuss this influence on the TWSC in detail because we analysed in Sec. 4.1 how trends and accelerations affect drought detection by different indicators.

We hope we addressed this comment by referring the low TWSC in later 2016 in EB and in 2003/2004 in WI to linear trends and constant accelerations.

#### Old text:

Furthermore, a strong trend and acceleration is contained in the synthesized time series for East Brazil and West India (Tab. 3).

#### New text:

Furthermore, a strong negative acceleration is contained in the synthesized time series for East Brazil (Tab. 3) leading to strong negative TWSC towards the end of the time series. For West India a strong positive trend leads to low TWSC at the begin of the time series.

#### Comment 8.

[Section 4.4] I am a bit confused how Figure 11 using real GRACE data is very different from the corresponding figure 10 center, right using the synthetic TWSC. Related to point 5 above, the synthetic data can detect only 2005 drought by design?

Response:

The drought indicators derived by synthetic data should only detect the drought as we designed it, here it was the drought in 2005 but we are able to design different drought duration and magnitudes as we did in one of the experiments described in Sec. 4.3. This drought is by design not equal to the detected real drought that we found in the real GRACE data (Figure 11).

At this point we also need to distinguish the synthetic TWSC data from the real GRACE TWSC data. The synthetic data were computed within a cluster. These clusters are based on regions with similar residual TWSC correlation within the interannual and subseasonal signal. We hypothesized these to be more likely affected by the same hydrological processes, for example by droughts.

We denoted the clusters according to the region where they are located, but the polygons are not exactly equal to, for example, the political boundaries of South Africa, which was used to estimate the results for the real GRACE TWSC. In turn, the polygons of the clusters are not used for the real GRACE TWSC application in Sec. 4.4 because the spatial interpretation of indicators based on political boundaries is better comparable to other research results than to our clusters. Furthermore, we do not intend to compare synthetic data to real data.

To better clarify, that the clusters have specific polygons that are different from the political boundaries, we added an explanation to the methodology part of the framework section.

Old text:

As a result of this procedure, we chose three clusters located in East Brazil (EB), South Africa (SA), and West India (WI), which were also affected by droughts in the past (e.g. Parthasarathy et al., 1987; Rouault and Richard, 2003; Coelho et al., 2016).

#### New text:

As a result of this procedure, we identified three cluster located in East Brazil (EB, South Africa (SA), and West India (WI) (Fig. 3), which were indeed affected by droughts in the past (e.g. Parthasarathy et al., 1987; Rouault and Richard, 2003; Coelho et al., 2016). Location and shape of the three chosen clusters are shown in Fig.3, and a global map pf all clusters is provided in Fig. B1. Cluster delineations from the above procedure should not be confused with political boundaries or watersheds.

### Comment 9.

[Page21;Ln20] It is not clear what "simplified noise models" mean. Please elaborate.

### Response:

By "simplified noise models" we mean error estimates that do not account for the peculiar way how the GRACE data are collected. For example, by simply assuming globally uniform error does not account for latitude-dependency, density of satellite orbits and data, time dependency of noise levels due to instrument problems or missing data, or the strong error correlation between neighboring grid cells. Here, we exemplarily add one simple example to name one possibility.

### Old text:

GRACE studies have been often based on simplified noise models (e.g. Zaitchik et al., 2008; Girotto et al., 2016),; however it is important to account for realistic error and signal correlation, in particular for drought studies where one will push the limits of GRACE spatial resolution.

### New text:

GRACE studies have often been based on simplified noise models (e.g. Zaitchik et al., 2008; Girotto et al., 2016), where the GRACE noise model is not derived from the used GRACE data but, for example, from literature and assumed to be spatially uniform and uncorrelated. However, it is important to account for realistic error and signal correlation (e.g. Eicker et al., 2014), in particular for drought studies where one will push the limits of GRACE spatial resolution. This signal correlation includes information about, for example, the geographic latitude, the density of the satellite orbits, the time-dependencies of mission periods or North-South-dependencies.

### Comment 10.

[Page22;Ln10-11] I do not follow "when we did not simulate a trend". When did you?

### Response:

Thanks, we modified the corresponding sentence.

### Old text:

When we did not simulate a trend, all indicators were able to detect drought, but they identified different timing, duration, and strength.

#### New text:

When we simulated smaller trends or accelerations, all indicators were able to detect drought, but they identified different timing, duration, and strength; for example for the Southafrican cluster (trend of 4.98 mm/year, accelerations of -0.38 mm<sup>2</sup>/year).

# Comment 11.

[Page23;Ln10] It will be helpful to name the four new indicators (or refer to equations).

### Response:

We assume that the reviewer suggested to add the four new indicators to [Page23;Ln20]. Please correct us if we are wrong.

#### New text:

Four new GRACE-based indicators (DSIA, DSID, DIA, and DID) were derived and tested; these are modifications of the above mentioned approaches based on time-accumulated and -differenced GRACE data.

### Minor edits:

[Equation 19] dot typo?

#### Response:

The equation is part of the previous sentence. We use the dot in the equation to finish the sentence.

[Table 3] What are the two values for Annual and Semi-annual?

### Response:

The coefficients in the table represent the same coefficients as used in Equation 18. In this equation, we have two coefficients for the annual and semi annual signal because these signals are computed using a sine and a cosine wave. So, the values in the table represent b1 and b2 coefficient for annual and c1 and c2 for the semi-annual signal. We included the coefficients in the table and the reference to the equation in the caption of the table.

#### New caption:

Coefficients (a\_0 to c\_1 from Eq. 18 and phi\_1 from Eq. A1) for signals contained in GRACE-TWSC that were extracted within the clusters of East Brazil, South Africa, and West India. These coefficients are used to simulate synthetic TWSC.

[Page14;Ln16] DSID appears twice. Thanks, done. [Page15;Ln1] This sentence is incomplete.

#### Response:

Please correct us if we are wrong, but we believe the sentence seems incomplete due to the word "results" as a verb instead of a subject and might led to confusion. We replaced it by "derive" to avoid confusion.

#### Old text:

Applying the Thomas-method to simulated GRACE TWSC results in magnitude, duration and severity of drought, which we show in Fig. 8 for the EB region.

#### New text:

The Thomas-method is applied to simulated TWSC data to derive magnitude, duration and severity of drought, which we show in Fig. 8 for the EB region.

[Figure6] DSI appears as black line in the plots while legend for DSI is blue. Done, legend is black now. [Page21;Ln7] GRACE and the DSIA6  $\rightarrow$  GRACE DSIA6? Thank you.

### Changes in the noise term

The noise for the synthetic TWSC is derived by using a full variance-covariance matrix. Since this matrix is now derived using the full variance-covariance matrix of the spherical harmonics (computed from normal equations provided by TU GRAZ) instead of using a variance matrix (main diagonal only), the noise levels in Fig. 5, 6, 7, 8, 9 and 10 were slightly updated. However, we would like to emphasize

that these changes do not yield to any changes in our conclusions. Following lines and values have been updated.

O= Old text, N= New text

## [Page3;Ln6]

O: ...(2) correlated spatial noise that is related to GRACE, ...

N: ...(2) correlated spatial noise that is related to the peculiar GRACE orbital pattern, ...

# [Page15;Ln5]

O: ... up to 28 months (Fig. 8, center) and a severity of about -2500 mm months (Fig. 8, bottom). N: ... up to 38 months (Fig. 8, center) and a severity of about -4000 mm months (Fig. 8, bottom).

# [Page17;Ln2]

O: However, for other cases differences can be more significant, which might lead to misinterpretation (e.g. February and April 2005 for the DI East Brazil, Fig. 9).

N: However, for other cases these differences can be more significant. These may lead to misinterpretation (e.g. May and July 2005 for the DI East Brazil, Fig. 9).

## [Page17;Ln14]

O: The DSI shows exceptional drought within the drought period with a maximum of 38 % of the grid cells, i.e. it does not detect exceptional drought in all grid cells.

N: Within the simulated drought period, the DSI indicator identified no more than 14 % of all grid cells as being affected by exceptional drought where it should be 100 %.

### [Page20;Ln5]

O: As a reference, the synthetic time series for West India, without any trend or acceleration signal, ranges from about -335 to 76 mm.

N: As a reference, the synthetic time series for West India, without any trend or acceleration signal, ranges from about -323 to 87 mm.

# [Page20;Ln14]

O: The severity class with the strongest drought type (i.e. exceptional drought) is only classified by the Zhao- and Houborg-method for East Brazil when using a drought magnitude of -120 mm; this is related to the trend and acceleration signal contained in the simulated TWSC and was already found in Sec. 4.1.

N: Exceptional drought is only classified by the Zhao-method for East Brazil for a simulated drought magnitude of 120 mm; this is related to the trend and acceleration signal contained in the simulated TWSC and was already found in Sec. 4.1.

### [Page20;Ln19]

O: Thus, a magnitude of -80 mm in severe drought all applied drought periods (3 to 24 months), while a magnitude of -60 mm leads to moderate dry events and a magnitude of -40 mm to abnormal dry events.

N: Thus, simulating a magnitude of -100 or -120 mm is identified as severe drought for all simulated drought periods (3 to 24 months), while simulating a lower magnitude (-80 mm and -60 mm) causes moderate or abnormal dry events to be identified.

# Reviewers comment #2

### **General comments**

In this paper the authors developed a framework that potentially contributes to the understanding of how drought signals propagate through various GRACE drought indicators. By applying three methods (GRACE-based indicators), the authors assessed the skills of newly derived GRACE drought indicators under rather more controlled conditions. This work is significant, as the study is a considerable addition to the existing literature about drought identification methods. Also, the topic is within the scope of Hydrology and Earth System Sciences. Overall, the experimental design is clear, and for the most part, the authors' conclusion are supported by their findings. However, I outline several general concerns, followed by a range of specific comments, which prevent me from recommending this manuscript for publication in its current form. I do hope through that the authors will be able to adequately address my comment and when that is done, this paper should be acceptable for publication.

### Response:

Thank you very much for your positive assessment and for your helpful feedback. We hope that we found good solutions to adequately address your comments and to improve the manuscript.

### Comment 1.

The paper is relatively poorly written. There is a significant number of grammatical/syntactic errors that are present throughout the entire body of the manuscript. I specify several of these in the "Specific Comments" section below, but the authors need to thoroughly check the entire text, as similar or other mistakes may exist elsewhere.

#### Response:

We thank the reviewer for this comment. The comments in the "Specific Comment" section will be addressed (see below), and we will thoroughly double check the entire text for revision.

#### Comment 2.

Page 3 Line 14 "As can be expected, TWSC and 6 months SPI appear moderately similar (correlation 0.43), characterised by positive peaks at the beginning of 2013. This motivates us to modify common GRACE indicators..." I find the evidence not supportive enough to safely conclude that this link/association between TWSC and SPI is always (or everywhere) the case. The authors should test this on several different regions characterized by varying hydro-climatic conditions. Making such conclusive statements using only one example is scientifically inaccurate.

#### Response:

We agree with the reviewer that one example is not sufficient to warrant such a conclusive link association between TWSC and the SPI. In fact, we tested this link for other regions, and indeed we found considerable correlations between TWSC and SPI (e.g. Missouri river basin, South Africa,

Maharashtra in West India). This was not illustrated (with figures) in the previous version due to space limitations, but we realize we should at least mention these results. Thus, a short sentence about some other regions including correlations is added.

### Old text:

As can be expected, TWSC and 6 months SPI appear moderately similar (correlation 0.43), characterised by positive peaks e.g. at the beginning of 2004 and at the end of 2009, and negative peaks at the beginning of 2013. This motivates us to modify common GRACE indicators ... .

#### New text

As can be expected, TWSC and 6 months SPI appear moderately similar (correlation 0.43), characterised by positive peaks, for example at the beginning of 2004 and at the end of 2009, and negative peaks at the beginning of 2013. We also found correlations between TWSC and 6 months SPI in regions with different hydro-climatic conditions for the Missouri river basin (0.31), Maharashtra in West India (0.46), and South Africa (0.45) among other regions. This motivates us to modify common GRACE indicators ....

### Comment 3.

More information is required for the cluster identification. How exactly were the three clusters determined? The authors also need to clearly specify their exact geographic location.

#### Response:

We believe that a detailed description of the EM-clustering is given in the literature, so we would like to avoid explaining the EM-algorithm in the main part of the paper. However, we would like to follow the reviewer's suggestion to provide some information to interested readers so we add the main idea and equations of the EM-clustering to the appendix.

Thanks for pointing it out, the information about the polygons can indeed easily be missed out. We adjusted the text and changed the color of the polygons to make them better detectable. We also added the global distribution of all clusters to Fig. B1 in the appendix.

#### Old text1:

As a result of this procedure, we chose three clusters located in East Brazil (EB), South Africa (SA), and West India (WI), which were also affected by droughts in the past (e.g. Parthasarathy et al., 1987; Rouault and Richard, 2003; Coelho et al., 2016).

#### New text1:

As a result of this procedure, we identified three clusters located in East Brazil (EB), South Africa (SA), and West India (WI), which were indeed affected by droughts in the past (e.g. Parthasarathy et al., 1987; Rouault and Richard, 2003; Coelho et al., 2016). Location and shape of the three chosen clusters are shown in Fig. 3, and a global map of all clusters is provided in Fig. B1.

#### Old text2:

The EM algorithm by Chen (2018) is modified to identify regional clusters by maximizing the likelihood of the data (Alpaydin, 2009).

#### New text2:

The EM algorithm by Chen (2018) is modified to identify regional clusters. The EM-algorithm alternates expectation and a maximization steps to maximize the likelihood of the data (e.g. Dempster, 1977; Redner, 1984; Alpaydin, 2009). More details about EM-clustering are provided in App. B.

#### Appendix B: EM-Clustering

Expectation maximization (EM) represents a popular iterative algorithm that is widely used for clustering data. EM partitions data into cluster of different sizes and aims at finding the maximum likelihood of parameters of a predefined probability distribution (Dempster, 1997). In case of a Gaussian distribution the EM-algorithm maximizes the Gaussian mixture parameters, which are the Gaussian mean  $\mu_k$ , covariance  $\Sigma_k$  and mixing coefficients  $\pi_k$  (Szeliski 2010). The algorithm then iteratively applies two consecutive steps to maximize the parameters: the expectation step (E-step) and the maximization step (M-step). Within the E-step we estimate the likelihood that a data point  $x_t$  is generated from the k-th Gaussian mixture by

E-step:

$$z_{ik} = \frac{1}{Z_i} \pi_k N(x|\mu_k, \Sigma_k) \quad ,$$

The M-step then re-estimates the parameters for each Gaussian mixture: M-step:

$$\mu_{k} = \frac{1}{N_{k}} \sum_{i} z_{ik} x_{i}$$
  

$$\Sigma_{k} = \frac{1}{N_{k}} \sum_{i} z_{ik} (x_{i} - \mu_{k}) (x_{i} - \mu_{k})^{T}$$
  

$$\pi_{k} = \frac{N_{k}}{N}$$

by using the number of points assigned to each cluster via

$$N_k = \sum_i z_{ik}$$
.

Using the maximized parameters EM assigns each data point to a cluster. The final global distributed clusters of the AR-parameters (Fig. 3) are shown in Fig. B1. These clusters were derived by modifying and applying an EM-algorithm provided by Chen (2018).

This appendix section contains a new reference, which is added to the reference list as follows: Szeliski, R.: Computer Vision: Algorithms and Applications, Springer Science and Business Media, 2010





#### Comment 4.

The authors should provide more detailed information (characteristics) about specific droughts mentioned in their methodology section.

#### Response:

To elucidate the chosen drought events, we added a table containing the specific regions and the corresponding considered drought year and TWSC months.

#### Old text:

Searching for drought duration and magnitude (step 3) led to four droughts seen in GRACE-TWSC: The 2005 and 2010 droughts in the Amazon (e.g. Chen et al., 2009; Espinoza et al., 2011), the 2011 drought in Texas (e.g. Long et al., 2013), and the 2003 drought in Europe (e.g. Seitz et al., 2008).

#### New text:

Performing literature research for duration and magnitude (step 3) led to four droughts seen in GRACE-TWSC (Tab. 4): The 2005 and 2010 droughts in the Amazon (e.g. Chen et al., 2009; Espinoza et al., 2011), the 2011 drought in Texas (e.g. Long et al., 2013), and the 2003 drought in Europe (e.g. Seitz et al., 2008).

Table 4. Drought events in Europe, Amazon river basin and Texas with corresponding duration taken from literature.

Region	Year of drought	Considered TWSC months	Examples of literature		
Europe	2003	June to August	Andersen et al. (2005)		
			Rebetez et al. (2006)		
			Seitz et al. (2008)		
Amazon river basin	2005	May to September	Chen et al. (2009)		
			Frappart et al. (2012)		
	2010	June to September	Espinoza et al. (2011)		
			Frappart et al. (2013)		
			Humphrey et al. (2016)		
Texas	2011	February to October	Humphrey et al. (2016)		
			Long et al. (2013)		

#### This table contains two new references, which is added to the reference list as follows:

Frappart, F., Papa, F., Santos da Silva, J., Ramillien, G., Prigent, C., Seyler, F. and Calmant, S.: Surface freshwater storage and dynamics in the Amazon basin during the 2005 exceptional drought, Environmental Research Letters, 7(4), 044010, doi:10.1088/1748-9326/7/4/044010, 2012.

Rebetez, M., Mayer, H., Dupont, O., Schindler, D., Gartner, K., Kropp, J. P. and Menzel, A.: Heat and drought 2003 in Europe: a climate synthesis, Annals of Forest Science, 63(6), 569–577, doi:10.1051/forest:2006043, 2006.

#### **Specific Comments**

#### Abstract

"Thus, this study aims at a better understanding of how drought signals, in the presence of trends and GRACE-specific spatial noise, propagate through GRACE drought indicators": This phrase is perhaps the essence of the abstract; therefore it should be able to provide the necessary information on its own. The authors need to specify which trends they are referring to.

#### **Response:**

Thanks, we are referring to linear trends and constant accelerations in the paper, which are described with a linear term  $a_1 (t - t_0)$  and a quadratic term  $a_2 \frac{1}{2} (t - t_0)^2$  in Eq. 18. Linear trends and possible constant accelerations in GRACE TWSC can result from many different hydrological processes, for example, accelerations can results from linear trends in the fluxes precipitation, evapotranspiration and runoff. To specify the terms, we added linear trend and constant accelerations to the abstract.

#### New text:

Thus, this study aims to better understand of how drought signals propagate through GRACE drought indicators in the presence of linear trends, constant accelerations, and GRACE-specific spatial noise.

According to this comment, we specified the meaning of trends and accelerations for the subsequent usage of the terms.

Page 7 Line 16

O: The signal is computed by ... at time t with a constant  $a_0$ , linear trend  $a_1$  and acceleration  $a_2$  terms, an annual signal  $b_1$  and  $b_2$ , and similar for a semi-annual signal  $c_1$  and  $c_2$ .

N: The signal is computed by ... [equation] ... at time t with a constant  $a_0$ , linear trend term  $a_1$ , constant acceleration  $a_2$  terms, annual signal terms  $b_1$  and  $b_2$ , and similarly semi-annual signal terms  $c_1$  and  $c_2$ . Trends and possible accelerations in GRACE TWSC can result from many different hydrological processes. For example, accelerations can results from trends in the fluxes precipitation, evapotranspiration, and runoff (e.g. Eicker et al. 2016). In the following, the linear trends are denoted as trends and constant accelerations are denoted as accelerations.

Line 10 application-dependent Yes, corrected, thanks.

Line 10 large differences Corrected.

Line 11 particularly Addressed.

Line 12 We show that trend and accelerations – what do the authors mean by "accelerations"?

Response:

We mean possible constant accelerations contained in the analysed time series that is described by the quadratic term  $a_2 \frac{1}{2} (t - t_0)^2$  in Eq. 18. We hope this is more clear now by specifying the trends, as the reviewer recommended in the first comment of the "Specific Comments" section (above).

Page 1 Line 17 affect the Done, thanks. Line 18 replace "reach" with "range" Done. Line 24 led Yes, thanks, corrected.

Page 2

Line 4 depends on the accumulation period considered - unclear

Response:

Yes, we see that the term accumulation period leads to confusion here, because it is introduced at a later point. We remove this part of the sentence.

Old text:

For South Africa, due to a complex rainfall regime, areas and percentage of land surface affected by drought can vary strongly (Rouault and Richard, 2005) and their identification depends on the accumulation period considered.

### New text:

For South Africa, due to a complex rainfall regime, areas and percentage of land surface affected by drought can vary strongly (Rouault and Richard, 2005).

Line 16 Much fewer Done.

### Line 23 and the first data are expected

#### Response:

We updated this sentence, because the first data is now available and not "expected to become available in May 2019".

### Old text:

Meanwhile, GRACE has been continued with the GRACE-FO mission and the first data are expected to become available in May 2019.

### New text:

Meanwhile, GRACE has been continued with the GRACE-FO mission from which the first data are now available.

Line 27 they found good agreement to net precipitation minus evaporation. - unclear

### Response:

We agree this needs clarification. The agreement between TWSC and the combination of the net precipitation and evaporation is meant.

### Old text:

For example, Seitz et al. (2008) investigated the 2003 heat wave over seven Central European basins using GRACE timeseries; they found good agreement to net precipitation minus evaporation.

#### New text:

For example, Seitz et al. (2008) investigated the 2003 heat wave over seven Central European basins using GRACE timeseries; they found a good agreement between TWSC and the combination of net precipitation and evaporation.

### Line 34 without utilizing external information - please specify

#### Response:

Separating a specific compartment from GRACE TWSC data requires knowledge from other observation techniques or model outputs, because GRACE can only measure the sum of all compartments.

### Old text:

However, neither GRACE nor GRACE-FO enable one to separate different compartments such as groundwater storage without utilizing external information, and their spatial (about 300 km for GRACE) and temporal (nominally one month) resolution are limited.

#### New text:

However, neither GRACE nor GRACE-FO enable one to separate different storage compartments, such as groundwater storage, without utilizing additional (e.g. compartment-specific) observations or model outputs, and their spatial and temporal resolution (about 300 km and nominally one month respectively for GRACE) are limited.

Page 3 Line 4 delete "e.g." Done, thanks. Line 7 "smoothing" Done. Line 17 What are "differencing periods"

Response:

We agree that the term here is confusing, because it was not introduced before. We change the sentence.

Old text:

This motivates us to modify common GRACE indicators to account for accumulation and differencing periods.

New text:

This motivates us to modify common GRACE indicators to account for accumulation periods of input data, e.g. used with 6 months SPI, but also periods that are based on differences of input data.

Line 21 spatially averaged Done, thanks. Line 26 will complete the paper Done.

Page 4 Line 2 explore Thanks, corrected. Line 10 more regularly Corrected.

Page 8 Line 10 we construct Done, thanks. Line 13 including the introduced (in Sec. 2.3) signal ... Done. Line 26 ... following A et al. (2013) ... is there something missing here?

Response: Indeed it might lead to confusion but A is the full last name.

Page 11

Line 8 drought onset and end Corrected.

Lines 10-14 these thresholds are rather arbitrarily made. It seems to me that a single value for the drought duration and magnitude should not be used for different hydrologic regimes.

Response:

We do not agree with the reviewer that these values for the threshold are arbitrary because we identified these values by analysing different historical droughts that were detected in literature using GRACE TWSC. Of course, one can not assume that one value for drought duration and magnitude can be detected in different hydrological regimes, but this is not what we intended with this analysis. We aim at simulating a signal that is similar to existing drought signals contained in GRACE, which is able to show up as exceptional drought in at least one indicator.

Page 12

Line 5 inappropriate use of English for a scientific paper Corrected.

Old text

However, seen these difficulties, we decided to stick to the most simple TWSC drought model, i.e. a constant water storage deficit within a given time span.

New text

However, due to these difficulties, we decided to use the most simple TWSC drought model, i.e. a constant water storage deficit within a given time span.

Page 13 Line 10 delete "would" Corrected, thanks.

Page 14 Line 17 for the 3, and 6 months differenced DSID Sorry we do not see a difference.

Page 20

Line 24 climatic phenomenon Yes, thanks, corrected. Line 24 delete "related to climatic conditions" as it is redundant Corrected.

Page 21

Line 9 in the northeastern Thanks, we changed it to "Northeastern".

Due to this comment we also changed following text:

Old text:

Fig. 3 shows the estimated AR-model coefficients, which represent the temporal correlations, ranging from very low up to 0.3, e.g. over the Sahara or in South West Australia, to about 0.8, for example in Brazil or in South Eastern U.S. EM-clustering is then based on these coefficients.

New text:

Fig. 3 shows the estimated AR-model coefficients, which represent the temporal correlations, ranging from very low up to 0.3, e.g. over the Sahara or in South West Australia, up to about 0.8, e.g. in Brazil or in the Southeastern U.S. EM-clustering is then based on these coefficients.

Page 23 Line 22 particularly Done. Line 25 the onset and end Done.

# List of changes – hess-2019-268

We undertook following <u>major</u> changes in the manuscript according to the reviewer's comments:

Page/Line(s)	Action	Reviewer No./Comment No.		
P 3/L 15-19	Rephrased sentence and added sentences	R 2/C 2		
P 7/L 16-17	Rephrased sentence	R 1/C 1		
P 10/L 4-7	Rephrased sentence and added sentences	R 1/C 4, R 1/C 8, R 2/C 3		
P 11/L 7-9	Rephrased sentence	R 1/C 6		
P 12/L 5-7	Rephrased sentence and added Tab. 4	R 1/C 5, R 2/C 4		
P 12/L25 - P 13/L 2	Rephrased sentence	R 1/C 7		
P 22/L17 - P 23/L 2	Added sentences	R 1/C 9		
P 23/L 13-14	Rephrased sentence	R 1/C 10		
P 24/L 22-24	Rephrased sentence	R 1/C 11		
P 26/L 1-19	Added paragraph and Fig. B1	R 1/C 4, R 2/C 3		

# A framework for deriving drought indicators from GRACE

Helena Gerdener<sup>1</sup>, Olga Engels<sup>1</sup>, and Jürgen Kusche<sup>1</sup>

<sup>1</sup>Institute of Geodesy and Geoinformation, University of Bonn, Bonn, Germany.

Correspondence: Helena Gerdener (gerdener@geod.uni-bonn.de)

Abstract. Identifying and quantifying drought in retrospective is a necessity for better understanding drought conditions and the propagation of drought through the hydrological cycle, and eventually for developing forecast systems. Hydrological droughts refer to water deficits in surface and subsurface storage, and since these are difficult to monitor at larger scales, several studies have suggested to exploit total water storage data from the GRACE (Gravity Recovery and Climate Experiment) satellite

- 5 gravity mission to analyse analyze them. This has led to the development of GRACE-based drought indicators. However, it is unclear how the ubiquitous presence of climate-related or anthropogenic water storage trends , which has been found from GRACE analyses , found within GRACE analyses masks drought signals. Thus, this study aims at a better understanding of to better understand how drought signals , propagate through GRACE drought indicators in the presence of trends-linear trends, constant accelerations, and GRACE-specific spatial noise, propagate through GRACE drought indicators. Synthetic data are
- 10 constructed and existing indicators are modified to possibly improve drought detection. Our results indicate that while the choice of the indicator should be application dependent, larger application-dependent, large differences in robustness can be observed. We found a modified, temporally accumulated version of the Zhao et al. (2017) indicator in particular particularly robust under realistic simulations. We show that trends and linear trends and constant accelerations seen in GRACE data tend to mask drought signals in indicators, and that different spatial averaging methods required to suppress the spatially correlated
- 15 GRACE noise affect the outcome. Finally, we identify and <u>analyse analyze</u> two droughts in South Africa using real GRACE data and the modified indicators.

Copyright statement. TEXT

#### 1 Introduction

Droughts are recurrent natural hazards that affect the environment and economy with potentially catastrophic consequences.

20 Drought impacts reach range from reduced streamflow, water scarcity, and reduced water quality to increased wildfires, soil erosion, and increased quantities of dust, crop failure, and large-scale famine. With climate change and population growth, the frequency and impact of droughts are projected to increase for many regions of the world (IPCC, 2013). Drought types can be distinguished depending on their effect on the hydrological cycle (e.g. Changnon, 1987; Mishra and Singh, 2010). In this study we focus on hydrological drought, a multiscale problem which may last weeks or many years, and which may affect local or

continental regions. For example, the severe drought between mid-2011 and -2012 affected millions of people in the entire East Africa region (Somalia, Djibouti, Ethiopia and Kenya)and let, and led to famine with an estimate of estimated 258,000 deaths (Checchhi and Robinson, 2013). From 2012 to 2016, the US state of California experienced a historical drought that adversely affected groundwater levels, forests, crops, fish populations, and led to widespread land subsidence (Mann and Gleick, 2015;

5 Moore et al., 2016). In contrast, European droughts<del>as e.g., for example</del> in 2018, typically last a few months in exceptionally dry summers. For South Africa, due to a complex rainfall regime, areas and percentage of land surface affected by drought can vary strongly (Rouault and Richard, 2005)<del>and their identification depends on the accumulation period considered</del>.

Hydrological drought refers to a deficit of accessible water, i.e. water in natural and man-made surface reservoirs and subsurface storages, with respect to normal conditions. The propagation of drought through the hydrological cycle typically

- 10 begins with a lack of precipitation, developing leading to runoff and soil moisture deficit, followed by decreasing streamflow and groundwater levels (Changnon, 1987). However, no unique standard procedures exist for measuring the deficit of each of these factors and for defining the normal conditions. In order to arrive at operational definitions, e.g. which are required for triggering a response according to drought class for example, a large variety of drought indicators has been defined which typically seek to extract certain sub-signals from observable fields (Bachmair et al., 2016; Wilhite, 2016; Mishra and Singh,
- 15 2010; Van Loon, 2015). Reviews of hydrological drought indicators are contained in Keyantash and Dracup (2002); Wilhite (2016); Mishra and Singh (2010); Tsakiris (2017). Streamflow is the most frequently used observable in these studies.

Drought detection is mostly restricted to single fluxes (precipitation or streamflow) or storages (surface soil moisture, reservoir levels) that are easy to measure. Much less-fewer measurements are available to assess water content in deeper soil layers and groundwater storage deficit, or the total of all storages. The NASA/DLR Gravity Recovery and Climate Experiment

- 20 (GRACE) satellite mission, launched in 2002, has changed this situation since GRACE-derived monthly gravity field models can be converted to total water storage changes (TWSC, Wahr et al., 1998). GRACE consisted of two spacecraft following each otherand linked with, which were linked together by an ultra-precise microwave ranging instrument; these ranges are routinely processed to provide monthly gravity models and further to mass changemaps thus maps of mass change. Since other mass transports in the atmosphere and ocean are removed during the processing, GRACE indeed provides quantitative measure
- of surface and subsurface water storages (Chen et al., 2009; Frappart et al., 2013). Meanwhile, GRACE has been continued with the GRACE-FO mission and first data is expected to become available in May 2019. from which the first data are now available.

Studies of drought detection with GRACE TWSC can be summarized in three groups: (i) using monthly maps of TWSC directly, (ii) partitioning TWSC timeseries into sub-signals that include drought signatures, or (iii) using indicators. For ex-

- 30 ample, Seitz et al. (2008) investigated the 2003 heat wave over seven Central European basins using GRACE timeseries; they found good agreement to net precipitation minus a good agreement between TWSC and the combination of net precipitation and evaporation. Other studies focused on drought detection using TWSC sub-signals, e.g. trends were used to identify drought in Central Europe (Andersen et al., 2005) and for the Tigris-Euphrates-Western Iran (Voss et al., 2013). After decomposing GRACE TWSC into a seasonal and non-seasonal signal, Chen et al. (2009) were able to detect the 2005 drought in the Central
- 35 Amazon river basin while Zhang et al. (2015) identified two droughts in 2006 and 2011 in the Yangtze river basin. In the

latter study, the El Niño/Southern Oscillation (ENSO) was identified as a possible driver for drought events in the Yangtze river basin. However, neither GRACE nor GRACE-FO enable one to separate different compartments storage compartments, such as groundwater storage without utilizing external information, without utilizing additional (e.g. compartment-specific) observations or model outputs, and their spatial and temporal resolution (about 300 km for GRACE) and temporal (and nomi-

5 nally one month ) resolution respectively for GRACE) are limited. Several efforts are therefore focusing focused on assimilating GRACE TWSC maps into hydrological or land surface models (e.g., Zaitchik et al., 2008; Eicker et al., 2014; Girotto et al., 2016; Springer, 2019).

Thus, perhaps not surprisingly, a number of GRACE-based drought indicators have been suggested (e.g. Houborg et al., 2012; Thomas et al., 2014; Zhao et al., 2017), typically either based on e.g. normalization or percentile rank methods. However,

- 10 a comprehensive comparison and assessment of these indicators is still missing, in particular in particularly in the presence of (1) trend signals as picked up by GRACE in many regions that may reflect non-stationary 'normal' conditions, (2) correlated spatial noise that is related to GRACE the peculiar GRACE orbital pattern, and (3) the inevitable spatial averaging applied to GRACEresults to smooth, which results in smoothing out noise (Wahr et al., 1998). From a water balance perspective, GRACE TWSC variability mainly represents monthly total precipitation anomalies (e.g., Chen et al., 2010; Frappart et al., 2013). It is
- 15 thus obvious that GRACE drought indicators will contain signatures that are visible in meteorological drought indicators, yet the difference should tell about the magnitude of other contributions (e.g. increased evapotranspiration due to radiation) to hydrological drought.

Fig. 1 shows a time series of region-averaged, de-trended and de-seasoned GRACE water storage changes over Eastern Brazil (Ceará state) compared to the region-averaged 6 months Standard Precipitation Indicator SPI (McKee et al., 1993) to illustrate the potential of GRACE TWSC for drought monitoring. As can be expected, TWSC and 6 months SPI appear

- moderately similar (correlation 0.43), characterised characterized by positive peakse.g., for example at the beginning of 2004 and at the end of 2009, and negative peaks at the beginning of 2013. We also found correlations between TWSC and 6 months SPI in regions with different hydro-climatic conditions for the Missouri river basin (0.31), Maharashtra in West India (0.46) and South Africa (0.45) among other regions. This motivates us to modify common GRACE indicators to account for accumulation
- 25 and differencing periods . periods of input data, e.g. used with 6 months SPI, but also periods that are based on differences of input data. To our knowledge, this is the first study where (modified) indicators are tested in a synthetic framework based on a realistic signal that includes a hypothetical drought. We hypothesize that in this way we can (i) assess indicator robustness, with respect to identifying a 'true' drought of given duration and magnitude, and (ii) understand how trend signals and spatial noise propagate into indicators and mask drought detection. In addition, we investigate to what extent the spatial averaging that
- 30 is required for <u>analysing analyzing</u> GRACE data affects indicators. For this, we compare spatially <u>average averaged</u> gridded indicators to indicators derived from <u>spatial</u> spatially averaged TWSC.

This contribution is organized as follows: in section 2 we will review three GRACE-based drought indicators and modify them to accommodate either multi-month accumulation or differencing, while in section 3 our framework for testing GRACE indicators in a realistic simulation environment will be explained. Then, section 4 will provide simulation results and finally

35 the results from real GRACE data. A discussion and conclusion will elose complete the paper.

20



Figure 1. De-trended and de-seasoned GRACE TWSC [mm](orange) and the SPI[-] of 6-month accumulated precipitation (blue), spatially averaged for Ceará, Brazil.

#### 2 Indicators for hydrological drought

5

20

Hydrological drought indicators are mostly based on observations of single water storages or fluxes, e.g. for precipitation, snowpack, streamflow, or groundwater. In general, indicator definitions can be arranged in four categories: 1) data normalization, 2) threshold-based, 3) quantile scores, and 4) probability-based (e.g., Zargar et al., 2011; Keyantash and Dracup, 2002; Tsakiris, 2017).

Since total water storage deficit may be viewed as a more comprehensive information for drought, with source on drought, the advent of GRACE total water storage changes (TWSC) data new indicators have been has led to new indicators being developed. For example, Frappart et al. (2013) developed a drought indicator based on yearly minima of water storage and a standardization method for standardization, and Kusche et al. (2016) computed recurrence times of yearly minima

- 10 through generalized extreme value theory. Other indicators explored explore the monthly resolution of GRACE, e.g. the Total Storage Deficit Index (TSDI, Agboma et al., 2009), the GRACE-based Hydrological Drought index (GHDI, Yi and Wen, 2016), the Drought Severity Index (DSI, Zhao et al., 2017), and the Drought Index (DI, Houborg et al., 2012). Further, Thomas et al. (2014) presented a water storage deficit approach to detect drought magnitude, duration, and severity based on GRACE-derived TWSC. To our knowledge, only the Zhao et al. (2017), Houborg et al. (2012), and Thomas et al. (2014) methods are able to
- 15 detect drought events from monthly GRACE data without any additional information. Therefore, these three indicators will be discussed further.

In order to stress the link between GRACE-based and meteorological indicators, we first describe the relation of TWSC and precipitation. Assuming evapotranspiration (*E*) and runoff (*Q*) vary more regular regularly as compared to precipitation (i.e.  $\Delta E = 0$ ,  $\Delta Q = 0$ ), the monthly GRACE TWSC ( $\overline{\Delta s}$ ) corresponds to precipitation anomalies ( $\Delta P$ ) accumulated since the GRACE storage monitoring began

$$\overline{\Delta s}(t) = \Delta t \sum_{t_0}^t \Delta P , \qquad (1)$$

where  $\Delta t$  is the time from  $t_0$  to  $t_1$ . In contrast to Eq. 1, the difference between GRACE months

$$\overline{\Delta s}(t_2) - \overline{\Delta s}(t_1) = \Delta t \sum_{t_1}^{t_2} \Delta P \tag{2}$$

corresponds to the precipitation anomaly accumulated between these months. Accumulating monthly TWSC corresponds thus Accumulated monthly TWSC thus corresponds to an iterative summation over the precipitation anomalies described by

5 
$$\sum_{t_0}^t \overline{\Delta s}(t) = \Delta t \sum_{\tau=t_0}^t \sum_{t_0}^\tau \Delta P.$$
 (3)

In the following, we will discuss and extend the definition of Zhao et al. (2017), Houborg et al. (2012), and Thomas et al. (2014) GRACE-based indicators, which are then hence referred to as the Zhao-method, Houborg-method, and Thomas-method, respectively.

#### 2.1 Zhao-method

In the approach of Zhao et al. (2017), one considers GRACE-derived monthly gridded TWSC for n years, 10

$$x_{i,j} = \overline{\Delta s}(t_{i,j}) \tag{4}$$

with

$$t_{i,j} = i + \left(j - \frac{1}{2}\right) \frac{1}{12}$$
  $i = 1, \dots, n$   $j = 1, \dots, 12$ . (5)

Let us define the monthly climatology, i.e. mean monthly TWSC,  $\tilde{x}_j$  with j = 1, ..., 12 and the standard deviation  $\tilde{\sigma}_j$  of the anomalies in month j with respect to the climatological value as 15

$$\tilde{x}_j = \frac{1}{n} \sum_{i=1}^n x_{i,j} \tag{6}$$

$$\tilde{\sigma}_j = \left(\frac{1}{n} \sum_{i=1}^n (x_{i,j} - \tilde{x}_j)^2\right)^{1/2}.$$
(7)

Zhao et al. (2017) define their drought severity index 'GRACE-DSI' as the standardized anomaly

20 TWSC-DSI<sub>*i*,*j*</sub> = 
$$\frac{x_{i,j} - \tilde{x}_j}{\tilde{\sigma}_j}$$
 (8)

of a given month  $t_{i,j}$  and provide a scale from -2.0 (exceptional drought) to +2.0 (exceptional exceptionally wet), as shown in Tab. 1. There is no particular probability distribution function (PDF) underlying the method, however if we assume the anomalies for a given month follow a Gaussian PDF it is straightforward to compute the likelihood of a given month falling in one of the Zhao et al. (2017) severity classes: For example, 2.1 % of months would be expected to turn out as exceptional

25 drought and 2.1 % as exceptionally wet. This can be applied to any other PDF. Drought severity, however, should be related to the duration of a drought. For example McKee et al. (1993) showed how typical time scales of 3, 6, 12, 24, and 48 months of precipitation deficits are related to their impact on usable water sources. To account for the relation between severity and duration in the Zhao et al. (2017) approach, we consider q-months accumulated TWSC, which is approximately related to precipitation in Eq. (3) as

5 
$$x_{i,j,q}^+ = \sum_{k=1}^q \overline{\Delta s}(t_{i,j+1-q})$$
 (9)

with  $t_{i,j+1-q} = t_{i-1,j+13-q}$  for j+1-q < 1, or equivalently written for q-months averaged TWSC as

$$x_{i,j,q}^{+} = \frac{1}{q} \sum_{k=1}^{q} \overline{\Delta s}(t_{i,j+1-q}) .$$
(10)

For example for q = 3, we would look for the 3 months running mean Dec-Jan-Feb, Jan-Feb-Mar, and so on. In the next step, one computes e.g., for example, the climatology and anomalies as with the original method. On the other hand, we can relate hydrological to meteorological indicators using Eq. (2). To develop a TWSC indicator that can be compared to indicators based on accumulated precipitation, one should rather consider the q months differenced TWSC

$$x_{i,j,q}^{-} = \overline{\Delta s}(t_{i,j}) - \overline{\Delta s}(t_{i,j+1-q}).$$
<sup>(11)</sup>

Thus, equivalent to as with TWSC-DSI<sub>*i*,*j*</sub> in Eq. (8), through standardization we can define two new multi-month indicators (TWSC-DSIA and TWSC-DSID) through standardization by using accumulated (A) and differenced (D) TWSC (Eq. 9 and 11) as

TWSC-DSIA<sub>*i*,*j*,*q*</sub> = 
$$\frac{x_{i,j,q}^+ - \tilde{x}_{j,q}^+}{\tilde{\sigma}_{j,q}^+}$$
 (12)

and

10

15

$$\text{TWSA-DSID}_{i,j,q} = \frac{x_{i,j,q}^- - \tilde{x}_{j,q}^-}{\tilde{\sigma}_{j,q}^-}.$$
(13)

Finally, it is obvious that sampling the full climatological range of dry and wet months is not yet possible with the limited
GRACE data period. Therefore, Zhao et al. (2017) suggest applying a bias correction to avoid the under- or overestimation of drought events. This implies using TWSC from multi-decadal model runs, which is feasible but not in the focus of this study.

#### 2.2 Houborg-method

Houborg et al. (2012) define the drought indicator 'GRACE-DI' via the percentile of a given month,  $t_{i,j}$ , with respect to the cumulative distribution function (CDF). The GRACE-DI is applied to TWSC by

25 TWSC-DI<sub>*i*,*j*</sub> = 
$$\frac{\sum_{i} (x_j \le x_{i,j})}{\sum_{i} x_j} \cdot 100,$$
 (14)

	TWSC-DSI [-]		
Drought Severity Level	Min.	Max.	
Abnormal	-0.8	-0.5	
Moderate	-1.3	-0.8	
Severe	-1.6	-1.3	
Extreme	-2.0	-1.6	
Exceptional		-2.0	

Table 1. Drought severity level of the TWSC-DSI (Zhao et al., 2017). The values of TWSC-DSI are unitless.

i.e. all years containing month j are counted for which TWSC is equal or lower than TWSC in month j and year i, and normalized by the number of the years that contain month j. The indicator value is assigned to five severity classes as shown in Tab. 2. For example, exceptional droughts occur up to 2 % of the entire time period at any location.

Again, to relate drought severity to duration, we proceed to via multi-month accumulation (Eq.9) and differences (Eq.11) 5 resulting in the definition of two new indicators based on TWSC-DI<sub>*i*,*j*</sub> in Eq. (14):

TWSC-DIA<sub>*i*,*j*</sub> = 
$$\frac{\sum_{i} (x_{j,q}^{+} \le x_{i,j,q}^{+})}{\sum_{i} x_{j,q}^{+}} \cdot 100$$
 (15)

$$\text{TWSC-DID}_{i,j} = \frac{\sum_{i} (x_{\overline{j},q} \le x_{\overline{i},j,q})}{\sum_{i} x_{\overline{j},q}} \cdot 100.$$
(16)

Assuming again that the CDF equals to the cumulative Gaussian, for example 0.6 % of months would be detected as exceptionally dry or and 9.5 % of months as abnormally dry. Houborg et al. (2012) applied the percentile approach also separately to surface soil moisture, root zone soil moisture and groundwater storage, which were derived by assimilating GRACE-derived TWSC into a hydrological model, and the CDFs were adjusted to a long-term model run. Here, we focus on TWSC from GRACE a simulated TWSC environment for the GRACE period only and, as explained in Sec. 2.1, we therefore disregard the bias correction.

#### 15 2.3 Thomas-method

Thomas et al. (2014) define a drought by considering the number of consecutive months below a threshold of TWSC. Given TWSC observations  $x_{i,j}$  and a threshold c, we can compute anomalies by

$$\Delta x_{i,j} = \begin{cases} 0 & \text{for } x_{i,j} \ge c \\ x_{i,j} - x_j & \text{for } x_{i,j} > c. \end{cases},$$
(17)

	TWSC-DI [%]		
Drought Severity Level	Min.	Max.	
Abnormal	20	30	
Moderate	10	20	
Severe	5	10	
Extreme	2	5	
Exceptional	0	2	

Table 2. Drought severity level of the TWSC-DI (Houborg et al., 2012). The values of TWSC-DI are given in %.

The While the threshold can be derived following from different concepts, however, Thomas et al. (2014) use the monthly climatology  $x_i$  (Eq. 6). Here, we also consider using a fitted signal for defining the threshold. The signal is computed by

$$x(t) = a_0 + a_1(t - t_0) + a_2 \frac{1}{2}(t - t_0)^2 + b_1 \cos(\omega t) + b_2 \sin(\omega t) + c_1 \cos(2\omega t) + c_2 \sin(2\omega t)$$
(18)

at time t with a constant  $a_0$ , linear trend a linear trend term  $a_1$  and acceleration, a constant acceleration term  $a_2$  terms, and

- 5 annual signal, annual signal terms  $b_1$  and  $b_2$ , and similar for a similarly semi-annual signal terms  $c_1$  and  $c_2$ . Trends and possible accelerations in GRACE TWSC can result from many different hydrological processes. For example, accelerations can result from trends in the fluxes precipitation, evapotranspiration, and runoff (e.g. Eicker et al., 2016). In the following, the linear trends are denoted as trends and constant accelerations are denoted as accelerations. The Thomas-method then identifies drought events through the computation of their magnitude, duration, and severity: the magnitude or water storage deficit
- 10 equals is equal to  $\Delta x_{i,j}$  (Eq. 17) and the duration  $d_{i,j}$  is given by the number of consecutive months where TWSC is below the threshold. Thomas et al. (2014) propose a minimum number of 3 consecutive months that are-required for the computation of drought duration. By using the deficit  $\Delta x_{i,j}$  and the duration  $d_{i,j}$ , the severity  $s_{i,j}$  of the drought event can finally be computed by

$$s_{i,j} = \Delta x_{i,j} d_{i,j} . \tag{19}$$

15 Severity is therefore a measure of the combined impact of the water storage deficits and duration duration and magnitude of water storage deficit, see Thomas et al. (2014) and Humphrey et al. (2016).

#### 3 Framework to derive synthetic TWSC for computing drought indicators

#### 3.1 Methods

In order to <u>analyse analyze</u> the performance of drought indicators, we <u>suggest to first</u> construct a synthetic timeseries of 'true' 20 total water storage changes (TWSC) , on a grid, first. We base our drought simulations on the GRACE data model

$$\Delta s(t) = x(t) + \eta(t) + \epsilon(t) \tag{20}$$



Figure 2. Concept of synthetic framework to generate synthetic TWSC

including the introduced (in Sec. 2.3introduced) signal x (Eq. 18) (which contains seasonality, a constant, linear, and time varying trend, and seasonalityEq. 18), an interannual signal  $\eta$ , which (which has been de-trended and de-seasoned and which will carry the simulated 'true' drought signature and which has been de-trended and de-seasoned), and a GRACE-specific noise term  $\epsilon$ . To simulate the 'true' signal as realistically as possible using Eq. (20), we first analyse analyze real GRACE-

5 TWSC following the steps summarized in Fig. 2. We derive 1) the signal components constant, trend, acceleration, annual, and semi-annual sine wave, 2) temporal correlations, 3) a representative drought signal quantified by strength and duration, and 4) spatially correlated noise , the latter from GRACE error covariance matrices. While the first three steps are generic and can be used for simulating other observables, step 4 is directly related to the measurement noise , (in this case the GRACE noise).

As an input to the simulation, GRACE-TWSC are derived by mapping monthly ITSG-GRACE2016 gravity field solutions of

- 10 degree and order 60, provided by TU GRAZ (Mayer-Gürr et al., 2016), to TWSC grids. As per standard practice, we add degreeone spherical harmonic coefficients from (Swenson et al., 2008) Swenson et al. (2008) and degree 2, order 0 coefficients from laser ranging solutions, (Cheng et al., 2011). Then, we remove the temporal mean field, apply a DDK3-filtering (Kusche et al., 2009) to suppress excessive noise, and map coefficients to TWSC via spherical harmonic synthesis. We also remove the effect of ongoing glacial isostatic adjustment (GIA) following A et al. (2013).
- 15 Droughts are a multiscale phenomenon, and for a realistic simulation we must first define the largest spatial scale to which we will apply the model of Eq. (20). In other words, we first need to identify coherent regions in the input data for which our approach is then applied at grid-scale prior to step 1. For this, we apply two consecutive steps: we first compute temporal signal correlations by fitting an autoregressive (AR) model (Appendix A; Akaike, 1969) to detrended and deseasoned GRACE data. These TWSC residuals contain interannual and subseasonal signals including real drought in-

formation. Temporal Next, temporal correlation coefficients are then used as input for an Expectation Maximization (EM) clustering (Dempster et al. (1977), Redner and Walker (1984)), because regions with similar residual TWSC correlation within the interannual and subseasonal signal are hypothesized here to be more likely affected by the same hydrological processes. The EM algorithm EM algorithm by Chen (2018) is modified to identify regional clustersby maximizing. The

5 EM-algorithm alternates an expectation and a maximization step to maximize the likelihood of the data (Alpaydin, 2009).
 (e.g. Dempster et al., 1977; Redner and Walker, 1984; Alpaydin, 2009). More details about EM-clustering are provided in App. B.

As a result of this procedure, we <u>chose\_identified</u> three clusters located in East Brazil (EB), South Africa (SA), and West India (WI), which were <u>also indeed</u> affected by droughts in the past (e.g. Parthasarathy et al., 1987; Rouault and Richard, 2003;

10 Coelho et al., 2016). Location and shape of the three chosen clusters are shown in Fig. 3, and a global map of all clusters is provided in Fig. B1. Cluster delineations from the above procedure should not be confused with political boundaries or watersheds. The following simulation steps are then applied to each of these three clusters.

In step 1 we estimate the signal coefficients according to Eq. (18) through least squares fit for each grid cell within the cluster. The coefficients are then spatially averaged to create a signal representative for of the mean conditions within the

- 15 region, and they then are used to create the constant, trends, and the seasonal part of parts of the synthetic time series. To simulate realistic temporal correlations at the region regional scale (step 2), we use the AR-model identified beforehand (Fig.2) and again average AR-model coefficients within the cluster. Then, we apply an AR model with the estimated optimal order and the averaged correlation coefficient (Eq. A1) to the synthetic time series to add temporal correlations.
- Simulating realistic drought events in step 3 is challenging because, to our knowledge, no unique procedure to simulate realistic drought periods for TWSC exists. For this reason, we first perform a literature review to identify representative drought periods and magnitudes for selected regions. Among others, this includes the 2003 European drought and the drought in the Amazon basin in 2011 (e.g., Seitz et al., 2008; Espinoza et al., 2011)(e.g., Seitz et al., 2008; Espinoza et al., 2011, respectively). TWSC within the identified drought period are then eliminated from the time series. In the next step, the parameters describing the constant, trend, acceleration and seasonal signal components before and after the drought are used to 'extrapolate' these sig-
- 25 nals during the drought period. By computing the difference of the original GRACE-TWSC time series and the continued signal in the drought period, we can separate non-seasonal variations from the data, which represent the drought magnitude. Our hypothesis is that the non-seasonal variations that we derive from the procedure possibly show a systematic behaviour behavior that can be parameterized. To extract this systematic behaviourbehavior, all extracted droughts are transformed to a standard duration. To compare the different drought signals, a standard duration and a standard magnitude are arbitrarily set
- 30 to 10 months and -100 mm, respectively. Finally, a synthetic drought signal  $\eta$  is generated by using the extracted knowledge of drought duration, drought magnitude, and systematic behavior and systematic behavior and it is added to the synthetically generated signal (Eq. 20).

In step 4 we add GRACE-specific spatially correlated and temporally varying noise  $\epsilon$  (Eq. 20). First, for each month t we extract a full variance-covariance matrix  $\Sigma$  for the region grid cells from GRACE-TWSC. Next Then, whenever  $\Sigma$  is positive

35 definite, we apply Cholesky decomposition  $\Sigma = R^T R$ , while if  $\Sigma$  is only positive semi-definite we apply eigenvalue decom-



Figure 3. AR(1)-model coefficients [-] for global GRACE-TWSC. The polygons of the clusters of East Brazil, South Africa and West India are added in dark greenmagenta.

position (Appendix C). Second, we generate a Gaussian noise series v of the length n, where n represents the number of grid cells within the cluster. Finally, spatial noise in month t is simulated through

$$\epsilon = R^T v. \tag{21}$$

The final synthetic signals for each grid cell within a cluster will thus exhibit the same constant, trend, acceleration, seasonal

5 signal, temporal correlations, and drought signal, but spatially different and correlated noise. In the following, we will test the hypothesis that GRACE indicators depend on the presence of trend and random input signals using the generated synthetic time series.

We believe that our synthetic framework based on real GRACE data has multiple benefits: i) we are able to identify the skill ability of an indicator by comparing the 'true' drought duration and magnitude (step 3) to the indicator results; ii) we are able to

10 detect the influence of other typical GRACE signals on the drought detection; iii) comparing different indicator outputs allows the synthetic framework enables us to identify strengths and weaknesses of each analyzed indicator, and thereby enables us to choose the most suitable indicator for a specific application.

#### 3.2 Synthetic TWSC

Here, we will briefly discuss the TWSC simulation following methods described in the previous section.

- 15 When estimating AR models for detrended and deseasoned global GRACE data, we find that for more than 70 % of the global land TWSC grids are best represented by an AR(1) process (App. Fig. A1). Therefore, we apply the AR(1) model for each grid. Fig. 3 shows the estimated AR-model coefficients, which represent the temporal correlations, ranging from very low up to 0.3, e.g. over the Sahara or in South West Australia, <u>up</u> to about 0.8, <u>for example e.g.</u> in Brazil or in <u>South Eastern the Southeastern U.S. EM-clustering is then based on these coefficients.</u>
- The selected three clusters (Fig. 3) show differences between the signal coefficients of the functional model (step 1, Eq. 18), which are exemplarily shown hence discussed for the linear trend. We find a mean linear trend for the East Brazil cluster of 1.0

mm TWSC per year; South Africa shows, a higher trend of 5.0 mm per year in South Africa, and for West India the trend is a trend of 56.3 mm per year (Tab. 3). The trends for East Brazil and South Africa in GRACE TWCS have been identified before (e.g. Humphrey et al., 2016; Rodell et al., 2018). We did not find confirmations for the strong linear trend in West India , e.g. Humphrey et al. (2016) found, for example, by Humphrey et al. (2016) who identified about 7 mm per year within this region.

5 We assume that in this study the linear trend for West India is estimated as strong positive because we additionally identify a strong negative acceleration of -8.03 mm per year<sup>2</sup> in West India. However, our simulation will cover weak and strong trends. In fact, all coefficients show such strong differences, which suggests that we cover different hydrological conditions when simulating TWSC for the three regions. In step 2 we identify correlations of 0.74 in East Brazil, 0.79 in West India, and 0.42

**Table 3.** Coefficients ( $a_0$  to  $c_2$  from Eq. 18 and  $\phi_1$  from Eq. A1) for signals contained in GRACE-TWSC that were extracted within the clusters of East Brazil, South Africa, and West India. These coefficients are used to simulate synthetic TWSC.

Cluster	Constant	Linear Trend	Acceleration	An	nual	Semi-a	nnual	AR-correlation
	$a_{0}$	$\widetilde{a_1}$	$\underline{a_2}$	$\underbrace{b_1}{\underbrace{b_1}}$	$\underbrace{b_2}{\sim}$	$\stackrel{c_2}{\sim}$	$\sim 2$	$\phi_{1}$
East Brazil	34.85	1.02	-1.77	6.83	106.12	4.69	9.47	0.74
South Africa	-24.00	4.98	-0.38	-4.31	-2.34	-1.23	1.07	0.42
West India	-139.37	56.30	-8.03	30.23	-122.69	-24.22	25.24	0.79

in South Africa (Tab. 3).

- Searching Performing literature research for drought duration and magnitude (step 3) led to four droughts seen in GRACE-TWSC (Tab. 4): The 2005 and 2010 droughts in the Amazon (e.g. Chen et al., 2009; Espinoza et al., 2011), the 2011 drought in Texas (e.g. Long et al. (2013))(e.g. Long et al., 2013), and the 2003 drought in Europe (e. g. Seitz et al. (2008)). (e.g. Seitz et al., 2008). To extract the drought duration, we compared drought begin and end onset and end identified in these and other papers. We found that different studies do not exactly match, with inconsistencies likely due to different method-
- 15 ologies used. Furthermore, some authors only specified the year of drought. Droughts finally extracted from the literature had a duration of 3 to 10 months (Fig. 4a-d). Unless otherwise specified, we decided to base our simulations on a duration of 9 months to represent a clear identifiable drought duration. Extracted drought magnitudes range from about -20 to -350 mm TWSC (Fig. 4a-d). Therefore, in order to simulate a drought magnitude that has a clear influence on the synthetic time series, we set the magnitude to -100 mm.
- As described in Sec. 3.1, we transform these water storage droughts to a standard duration and magnitude to understand whether a typical signature can be seen. However, Fig. 4e remains inconclusive as in particular there are, in particular, four standardized droughts, which show a very different temporal behaviourbehavior: Toulouse in 2003, Obidos in 2010, and Houston and Dallas in 2011. When we remove those four timeseries (Fig. 4f), a systematic behaviour behavior can be identified and parameterized using a linear or quadratic temporal model. However, seen due to these difficulties, we decided to stick to use
- 25 the most simple TWSC drought model, i.e. a constant water storage deficit within a given time span.

Region	Year of drought	Considered TWSC months	Examples of literature
Europe	2003	June to August	Andersen et al. (2005)
			Rebetez et al. (2006)
			<u>Seitz et al. (2008)</u>
Amazon river basin	2005	May to September	<u>Chen et al. (2009)</u>
			Frappart et al. (2012)
	2010	June to September	Espinoza et al. (2011)
			Frappart et al. (2013)
			Humphrey et al. (2016)
Texas	2011	February to October	Humphrey et al. (2016)
			Long et al. (2013)



**Figure 4.** Extracted drought periods from GRACE-TWSC for the droughts in (a) Europe 2003, (b) Amazon river basin 2005, (c) Amazon river basin 2010, (d) Texas 2011. (e) All droughts from (a-d) were transformed to standard severity and duration. (f) as (e) but after removing four timeseries with a significant different temporal behaviour behavior.

In step 4, we project the simulation on a 0.5° grid and add spatially correlated GRACE noise. A few representative time series of the gridded synthetic total water storage change are shown in Fig. 5 for East Brazil (EB), South Africa (SA), and West India (WI) for the GRACE time period from January 2003 to December 2016. The effect of realistic GRACE noise (dark blue vs. light blue) is clearly visible, in particular particularly for the SA case with low annual amplitude. The synthetic drought period is placed from January to September 2005 (light brown) in all three regions. Synthetic TWSC variability includes considerable (semi-) annual variations for EB based on Tab. 3. Furthermore, a strong trend and negative acceleration is contained in the

13

5



**Figure 5.** Synthetic TWSC [mm] without (light blue) and with spatial GRACE noise (dark blue) using average parameters for the clusters in East Brazil (EB), South Africa (SA), and West India (WI). Light brown shows the simulated drought period.

synthesized time series for East Brazil and West India (Tab. 3) -leading to strong negative TWSC towards the end of the time series. For West India a strong positive trend leads to low TWSC at the begin of the time series.

#### 4 Indicator-based drought identification with synthetic and real GRACE data

#### 4.1 Synthetic TWSC: masking effect of trend and seasonality

- 5 Here, we analyse analyze how non-drought signals, such as a linear or accelerated water storage trend and the ubiquitous seasonal signal, propagate through the Zhao-, Houborg-, and Thomas- GRACE-indicators (Sec. 2) and potentially mask a drought. To this end, we select representative time series from each of the three synthetic grids of total water storage changes (TWSC) for East Brazil (EB), South Africa (SA), and West India (WI), and apply the three methods. Since all results are based on TWSC, we refer to TWSC-DSIA, TWSC-DSID, TWSC-DIA, and TWSC-DID as DSIA, DSID, DIA and DID, and DID.
- 10 respectively.

We first assess the temporal characteristics of the Zhao-method (Sec. 2.1). Figure 6 (left) shows time series for the DSI and DSIA (with 3, 6, 12 or 24 months accumulated TWSC). It is obvious that trend and acceleration propagate into both DSI and DSIA (see East Brazil and West India). Resulting indicator values (e.g. for the years 2015 and 2016are lower as) are lower than compared to a small trend (South Africa) and this may lead to misinterpretations because a severe to mild drought is identified

(-2 to -0.5) while none is actually simulated. In contrast, the actual simulated drought in 2005 is only identified as a moderate drought (values up to -1.0) for EB.

In the presence of a small trend (5.0 mm/year) and acceleration (-0.38 mm/year<sup>2</sup>, Tab. 3, SA), we do identify <u>an</u> exceptional drought (Fig. 6 DSIA for South Africa). This shows that the drought strength that we chose does indeed <u>would</u>-lead to a correct

- 5 identification of exceptional drought in case-if no masking occurs (but in the presence of GRACE noise), so at this point we can determine that exceptional drought represents the 'true' drought severity class. As expected, a trend and/or an acceleration signal that are frequently observed in GRACE analyses analyzes can lead to misinterpretations in the indicators. However, the influence of the trend or acceleration also depends on the timing of the drought period within the analysis window. For example, assuming we simulate the time series with the same trend or acceleration but the drought would were to occur in
- 10 2014, the drought detection would not have been as much influenced influenced as much. Therefore, we decided to set up an additional experiment and discuss the influence of different trend strengths for the drought detection (Sec. 4.3).

The analysis reveals that DSI and DSIA indicators are sensitive with respect to trends, while they are less sensitive to the annual and semi-annual signal. The seasonal signal is clearly dampened (compare e.g. compare Fig. 5 and to the DSIA in Fig. 6). This is caused by removing the climatology within the Zhao-method (Eq. 8). Comparing DSIA3, DSIA6, DSIA12,

- 15 and DSIA24, e.g. for East BazilBrazil, suggests that with a longer accumulation period, indicator time series are increasingly smoothed and less severe droughts are identified (Fig. 6, left). Furthermore, the drought period appears shifted in time and its duration is prolonged. This can lead to missing a drought identification if a trend or an acceleration is contained in the analyzed timeseries, for example for the 24 months DSIA for East Brazil. We find that all DSIA are able to unambiguously detect a drought close to 2005 assuming that neither trend nor acceleration is apparent (Fig. 6 DSIA for South Africa). In
- 20 particularParticularly, the 3 and 6 months DSIA identify the drought close to 2005 for South Africa, and its computation appears to dampen the temporal noise that is present in the DSI.

In contrast we find that the 3, 6, 12, or and 24 months TWSC-differencing DSID exhibit stronger temporal noise as compared to the DSIA and the DSI. This can be seen in the light of Eq. (2) - these indicators are closer to meteorological indicators and thus do not inherit the integrating property of TWSC. The DSID does neither not propagate a trend nor and acceleration, annual

- signal or semi-annual signal. All DSID and DSID time series, for example for East Brazil (Fig 6, right), show a strong negative peak within the drought period, but this peak does not cover the entire drought period for the 3, and 6 months differenced DSID. The negative peak within the drought period is always followed by a strong positive peak, when we consider Eq. 2 this lends to the interpretation that a pronounced drought period is normally followed by a very wet event to return to 'normal' water storage condition. Despite higher noise and the positive peak and contrary to the DSIA, all DSID (DSID3, DSID6, DSID12, DSID6, D
- 30

All different DSID time series for WI identify at least a moderate drought.

Analysis of the Houborg-method shows a broadly similar behaviour behaviour as compared to the Zhao-method: The sensitivity of drought detection to an included trend or acceleration depends on the indicators type. Using the DIA we can confirm the large influence of the trend or acceleration on the indicator value, which is not the case for DID (e.g. Fig. 7 DIA and DID

and DSID24) correctly identify the drought within 2005 to be exceptional exceptionally dry for East Brazil and South Africa.

35 for East Brazil). Annual and semi-annual water storage signals are all considerably weakened in the Houborg-method because



Figure 6. A representative example of the synthetic DSI, DSIA, DSID [-] for the East Brazil (EB), South Africa (SA), and West India (WI) cluster over the periods of 3, 6, 12, and 24 months. Light brown shows the synthetic constructed drought period.

they are effectively removed when computing the empirical distribution for each month of the year. Differences to the Zhaomethod appear when comparing more general properties, e.g. we find that DI is more noisy and the range of output values is restricted to about 7 % to 100 % (Fig. 7). This restriction is caused by the length of the time series, e.g. assuming we strive to identify an event with exceptional dry values ( $\leq 2\%$ ), we would need at least 50 years of monthly observations. Yet, with CD 1 CE

- 5 GRACE we only have about 14 years of good monthly observations, so the simulation was also restricted to this period. If we then take the driest value that might occur only once, we can compute the minimum value of DI to be 7.14 %. Hence the detection of exceptional or extreme drought is not possible when referring to the duration of the GRACE TWSC time series. As mentioned in Sec. 2.2, Houborg et al. (2012) applied a bias correction to the empirical CDF to mitigate this restriction. We do not follow Houborg's approach here in order to focus on realistic observation availability the synthetic environment instead
  10 of the availability of model outputs.
- to of the availability of model outputs.

Applying the The Thomas-method to simulated GRACE TWSC results in is applied to simulated TWSC data to derive magnitude, duration and severity of drought, which we show in Fig. 8 for the EB region. We find that the linear trend and acceleration propagate into the magnitude (Fig. 8, top) when using TWSC deficits with climatology removed (blue, Eq. 6) instead of compared to using TWSC deficits with removed trends(linear and time-varying) and, accelerations and seasonality

15 (red, Eq. 18). When using non-climatological TWSC (blue), we identify a strong deficit in 2015 and 2016 (Fig. 8, top) which



**Figure 7.** A representative example of the synthetic DI, DIA, DID [%] for the East Brazil (EB), South Africa (SA), and West India (WI) cluster over the periods of 3, 6, 12, and 24 months. Light brown shows the synthetic constructed drought period.

suggests a duration of up to  $\frac{28}{38}$  months (Fig. 8, center) and a severity of about  $\frac{-2500}{4000}$  mm months (Fig. 8, bottom). Using the detrended and deseasoned TWSC (red), drought is mainly detected in the 'true' drought period (2005) and not at the end of the time series. Thus we conclude that a trend or acceleration indeed modifies the drought detection.

5

10

Results so far were derived by imposing a minimum duration of 3 months (blue and red). When moving to a minimum duration of 6 consecutive months (green, Fig. 8, middle and bottom) we find this would lead to a decrease in identified severity by half, and the beginning of the drought period shifts 3 months in time. This is in line with Thomas et al. (2014). The same findings are made for South Africa and West India.

#### 4.2 Synthetic TWSC: effect of spatially correlated GRACE errors

Here, we investigate how robust the Zhao-, Houborg- and Thomas-indicators are with respect to the spatially correlated and time-variable GRACE errors. However, any analysis must take into account that GRACE results cannot be evaluated directly at grid resolution.

In our first analysis, <u>indicator indicators</u> based on (synthetic) TWSC grids are thus spatially averaged through two different methods (Sec. 3.1). We find that regional-scale DSI , <u>DI indicators and DI indicators</u>, as well as the outputs derived by the Thomas-method for South Africa computed from <del>1)</del>-averaging TWSC first (darkblue Fig. 9) is indeed different to the <del>2)</del>



**Figure 8.** Drought magnitude [mm], duration [mo] and severity [mm·months] for the cluster of East Brazil (EB) using TWSC with removed climatology (dark blue) and TWSC with removed trend and seasonal signal (red). The minimum duration (MD) is set to 3 months (blue and red) or 6 months (green). Light brown shows the synthetic constructed drought period.

averaging indicators computed at grid scale from TWSC (lightblue, Fig. 9). These differences can be explained by the inherent non-linearity of the indicators. Since the synthetic data have been constructed from the same constants, trends, seasonal signal, temporal correlations, and drought signal, we isolate the effect of GRACE noise on regional-scale indicators here. Outside of the drought period we conclude that the sequence how-in which we spatially average causes larger differences for DI as

- 5 compared to DSI: for . For South Africa, the range of averaged DI is about 7 100 % while the range of the DI of averaged TWSC is about 7 80 %. Within the drought period the DI exhibits little differences difference between both averaging methods. The DSI from averaged TWSC does suggest a weaker severity in the drought period compared to averaged DSI. In this case, both indicator averages identify the same (exceptional) drought severity class. Yet we find that for both DSI and DI the identification of drought severity is not sensitive to the choice of the averaging method for this cluster. However, for
- 10 other cases these differences can be more significant, which might. These may lead to misinterpretation (e.g. February and April-May and July 2005 for the DI East Brazil, Fig. 9). For the Thomas-method, we cannot distinguish which result is more significant, since we have no comparable 'true' severity amount for that indicator.

To determine the influence of the GRACE-specific spatial noise on the detected drought severity, a second analysis is applied. This analysis computes the share of area, for each time step , for which a given drought severity class is identified (Fig. 10).

15 Since different grid cells for one time step only differ in their spatial noise, it is important to understand that identifying more than one severity class is directly related to the noise. Only one class of drought would be detected for one epoch, assuming the grid cells have no or exactly the same noise. For example, we identify all classes of droughts (abnormal to exceptional) in



**Figure 9.** DSI and DI average in South Africa (SA, top and top center), severity average for the Thomas-method (SA, bottom center) and DI average in East Brazil (EB, bottom) by applying two different methods: the average of the indicators for all grids (light blue) and the indicators of averaged TWSC (dark blue). The grey shaded area represents the bandwidth for all grids. Light brown shows the synthetic constructed drought period.

December 2015 by using DSI for the East Brazil cluster (Fig. 10, top left). Thus, the spatial noise has a large influence on the drought detection. To establish which indicator is mostly most affected, the indicators are compared with each other.

We note that large differences are found between the DSI, the 6 months accumulated DSIA, and the 6 months differenced DSID within the given drought period for the East Brazil region (Fig. 10, left). All three indicators manage to identify the

5 drought, but with different duration and percentage of affected area. The DSI shows exceptional drought within the drought period with a maximum of 38 % of the grid cells, i.e. it does not detect exceptional drought in all grid cells. Within the simulated drought period, the DSI indicator identified no more than 14 % of all grid cells as being affected by exceptional drought where it should be 100 %. On the contrary other hand, the DSIA does not detect exceptional drought in any grid cell. Apparently, It is apparent that this indicator misses the exceptional dry event because of the included trend and acceleration.



Figure 10. Drought affected area of the DSI, DSIA, and DSID [%] considering the different drought severity classes within the clusters of East Brazil (EB) and South Africa (SA).

When comparing DSIA of East Brazil to the DSIA of South Africa (Fig. 10, center), we find that DSIA is able to detect the drought strength correctly when there is a small trend or acceleration present. However, DSIA appears more robust against spatial noise, since it identifies (at least) severe drought severe drought or drier in more than 90 % of grid cells, while the DSI indicator identifies only about 60 %. As described in Sec. 4.1, longer accumulation periods lead to smoother and thus more robust indicators. We find that the DSID is more successful in detecting exceptional drought: more than 60-80 % of the DSID grid cells show exceptional drought, but the indicator appears more noisy than the DSIA. Finallyas what regards, with regard to the drought duration, we find that only DSI detects the 'true' period correctly. When identified via DSIA, the duration appears longer and when identified in DSID, the period was found shorter as compared to the 'true' drought period.

Overall, we find that the different indicators DSI, DSIAor, and DSID all come with advantages and disadvantages regarding
the presence of spatial and temporal noise. The same findings were made for the indicators of the Houborg-method (results not shown). This analysis is not applied to the Thomas-method, because the method does not refer to severity classes (Sec. 2.3).

#### 4.3 Synthetic TWSC: experiments with variable trend, drought duration and severity

5

Two experiments were additionally constructed to examine the influence of trends and drought parameters on the indicator skillscapability. First, we consider how strong a linear trend in total water storage must be to mask drought in the indicators.

15 For this, we test different trends from -10 mm/year to 10 mm per year for DSI, DSIA, DI, DIA, and the Thomas-method in the West India region (since these indicators were identified as being affected by trends, Sec. 4.1). No acceleration is included for these tests. We find that trends between -1 and 1 mm per year cause no influence on all indicators, while differences start to

appear when simulating a trend higher than 2 mm per year. This propagates into the DSI, DSIA, DI, and DIA indicators but did not affect the drought period.

What A question we must ask is what would be the largest trend magnitude that does not affect the correct detection of drought duration and drought severity, and how can we verify this? An obvious influence within the drought period in 2005 is

- 5 found when simulating a trend of -6-7 mm or lower per year. It is important at this point to understand that there is a relation between the timing of the drought and the sign of the trend, i.e. a positive or a negativetrend whether the trend is positive or negative. Assuming that a positive trend exists and the drought occurs closer to the end of the time series, the trend may lead to a drought that is identified as more dry than the actual 'true' drought. But if the trend is negative, the drought is identified more easily.
- 10 Other factors, e.g. the length of the time series, have an influence on the masking by the trend and, as a result, affect drought detection. The longer the input time series, the more sensitive is the drought detection is to the trend. At the same time, the magnitude of the trend needs to be considered relative to the variability or range of the TWSC. E.g. For example, a -6 mm per year trend has a larger influence on the drought detection assuming if the range of TWSC being is -50 to 50 mm as compared to -200 to 200 mm. As a reference, the synthetic time series for West India, without any trend or acceleration signal, ranges
- 15 from about  $\frac{-335 \text{ to } 76 323 \text{ to } 87}{-323 \text{ to } 87}$  mm. So, deriving a general quantity for these dependencies is difficult.

In a second experiment, we assess which input drought duration and magnitude would at least be visually recognized in the indicators. We choose 3, 6, 9, 12, and 24 months for the simulated duration and -40 mm, -60 mm, -80 mm, -100 mm, and -120 mm for the drought magnitude, and apply both the Zhao- and the Houborg-method. We compare the changes for one indicator time series for the East Brazil region. The drought always begins in January 2005 for the first tests. In general, we found that the identification of the severity class is less sensitive to changes in the drought duration, since a drought duration of 3, 6, 9, 12, and 24 months mostly results in equal drought severity classes for example for , for example, a drought magnitude of 120

mm. Thus, we concentrate our analysis on changes in drought magnitude.

The severity class with the strongest drought type (i.e. exceptional drought) Exceptional drought is only classified by the Zhao- and Houborg-method Zhao-method for East Brazil when using a for a simulated drought magnitude of -120-120 mm;

- 25 this is related to the trend and acceleration signal contained in the simulated TWSC and was already found in Sec. 4.1. For the Zhao-method, extreme drought is identified when simulating a drought magnitude of at least -100 mm, while only severe and moderate drought is identified when simulating a magnitude of -80 mm and -60mm. The Houborg-method fails to identify extreme and exceptional drought, as described in Sec. 4.1. Thus, simulating a magnitude of -80 mm in severe drought all applied -100 and -120 mm is identified as severe drought for all simulated drought periods (3 to 24 months), while a magnitude
- 30 of simulating a lower magnitude (-80 mm and -60 mmleads to moderate dry events and a magnitude of -40 mmto-) causes moderate or abnormal dry events to be identified. We find that the both methods are not able to clearly detect a drought that has a magnitude of -40 mm or higher weaker, if the duration is between 3 and 24 months. This experiment supports our findings in Sec. 3.2.

#### 4.4 Application to real GRACE data: South Africa droughts

For South Africa, droughts are a recurrent <del>climate phenomenon related to climatic conditions</del> <u>climatic phenomenon</u>. The complex rainfall regime <u>has</u> led to multiple <u>extents occurrences</u> of drought events in the past, for example to a strong drought in 1983 (e.g. Rouault and Richard, 2003; Vogel et al., 2010; Malherbe et al., 2016). <u>The These past</u> droughts appeared in varying

5 climate regions<del>at different timing, at different times</del> of the year, and with a different severity. Since 1960, many of them were linked to El Niño (e.g. Rouault and Richard, 2003; Malherbe et al., 2016).

Based on the simulation results, we chose the 6 months accumulated DSIA to identify droughts for (the administrative area of) South Africa (*GADM*, 2018) in retrospective in the GRACE total water storage data. DSIA has proven to be more robust with respect to the peculiar, GRACE-typical spatial and temporal noise as compared to the other tested indicators (Sec. 4.2 and 4.1).

GRACE-DSIA6 suggests two drought periods, from mid of 2003 to mid of 2006 mid-2003 to mid-2006, and from 2015 to 2016 (Fig. 11). The first drought event is identified to affect at least 70 % of the area of South Africa. While 2003 was indeed a year of abnormal to severe dry conditions, in 2004 until mid of 2006 also extreme drought occurred during the period of 2004 to mid-2006. Figure 11 reveals that a small area (about 7976 km<sup>2</sup>, close to Lesotho) experienced even exceptional drought in

- 15 even experienced exceptional drought during 2004. This period is confirmed by The Emergency Events Database (EM-DAT , 2018) recording a drought event in 2004 , see e. g. Masih et al. (2014). (e.g. Masih et al., 2014). Extreme drought in 2004 mainly occurred in the Central and South East of South Africa; this is exemplarily shown for April 2004 in exemplified Fig. 12a for April 2004. Another confirmation is found in Malherbe et al. (2016), who identified a drought period from 2003 to 2007 by using the SPI.
- The second drought in 2015 and 2016 is identified to have affected Despite affecting less area (about 50 to 70 %, Fig. 11), but it the second drought in 2015 and 2016 is perceived as more intense than the 2003 to 2006 drought. Based on GRACE and the the GRACE DSIA6, we conclude that in 2016 at least 30 % of South Africa were was affected by extreme drought and about 20 % experienced an exceptional droughtin 2016... The 2016 drought occurred in the North Eastern Northeastern part of South Africa (Fig 12b). For comparison, the EM-DAT database also listed similarly identified 2015 as drought event but not 2016. a
- 25 drought event, but did not classify 2016 as such. We speculate that the differences are due to the drought criteria of the EM-DAT database (disasters are included when, for example, 10 or more people died or 100 or more people were affected). However, the EM-DAT database lists 2016 as a year of extreme temperature, which might be related to our detected drought. Furthermore, we can confirm the 2015/2016 drought by a lower maximum precipitation in these years than in other years (about 65 mm) and by meteorological indicators indicating severe to extreme drought (SPI, Standardized Precipitation Evapotranspiration Index
- 30 (Vincente-Serrano et al., 2010), and Weighted Anomaly Standardized Index (Lyon and Barnston, 2015)).

#### 5 Discussion

10

The framework developed in this study enables us to simulate GRACE-TSWC data with realistic signal and noise properties, and thus to assess the skills-ability of GRACE drought indicators to detect drought events in a controlled environment with



Figure 11. Percentage of drought affected area of the 6 months DSIA[-] considering the different drought severity classes. Application on real GRACE-TWSC over South Africa from 2003 to 2016.



Figure 12. DSIA6 [-] for real GRACE-TWSC within South Africa (black line, GADM (2018)) for (a) April 2004 and (b) March 2016.

known 'truth'. This will be extended to GRACE-FO in the near future. GRACE studies have been often often been based on simplified noise models (e.g. Zaitchik et al., 2008; Girotto et al., 2016) ; however where the GRACE noise model is not derived from the used GRACE data but, for example, from literature and assumed to be spatially uniform and uncorrelated. However, it is important to account for realistic error and signal correlation (e.g. Eicker et al., 2014), in particular for drought studies where

5 one will push the limits of GRACE spatial resolution. This signal correlation includes information about, for example, the geographic latitude, the density of the satellite orbits, the time-dependencies of mission periods or North-South-dependencies.

However, identifying a drought signal from real GRACE-TWSC is indeed challenging since we do not know in advance how-what the signature of a drought looks like; a parametric drought model does not yet exist and our experiment (Sec. 3.2) to extract such a model from TWSC data and known droughts did not lead to conclusive results. Still we believe that this first

10 – to our knowledge – approachidentified a similar systematic behaviour of different drought periods, although, despite being based on a small number of drought periods, identified a similar systematic behavior of different drought periods and should

be pursued further. Based on literature and our own experiments (Sec. 4.3) we chose to define our 'box'-like GRACE drought model as an immediate and constant water storage deficit.

When <u>analyzing</u> the Zhao-, Houborg- and Thomas-methods, we find that trends and accelerations in GRACE water storage maps tend to bias <u>not only</u> the DSI, DI and the Thomas-indicator that uses (which use non-climatological TWSC), but

- 5 also for-the DSIA and DIA (which use accumulated TWSC. Indicators). The indicators DSID and DID, which utilize timedifferenced TWSC, were not found biased by trends and accelerations; the same goes for the Thomas-method when based on detrended and deseasoned TWSC. When we did not simulate a trendsimulated smaller trends or accelerations, all indicators were able to detect drought, but they identified different timing, duration, and strength; for example for the SA cluster (trend of 4.98 mm/year, acceleration of -0.38 mm/year<sup>2</sup>). This suggests removing the trend in GRACE data first, but this must be done
- 10 with care, since it can also influence the detection of, for example, long-term droughts. The same is true for removing the trend and seasonal signal prior of applying the Thomas-method, although in this study we found that the removal of these signals simplified the correct drought detection (Sec. 4.1).

An experiment was then set up to understand the influence of the trend on the detected drought duration and severity. Several factors play a role here, e.g. the length of the time series, the TWSC range in relation to the trend magnitude, and the sign of the trend. We found that providing a general rule appears nearly impossible.

15

As expected, we find time-series for the modified time-differencing GRACE indicators DSID and DID as much noisier when compared to the time-accumulating indicators DSIA and DIA; this can be linked to precipitation (Sec. 2) driving total water storage. The drought period was identified to be shorter than the 'true' simulated drought periodfor, e.g. for DSID3 and DSID6. After these drought periods, strongly wet periods were detected. In the Regarding future applications, we suggest a direct comparison of the DSID and metagralazing indicators in martinular for confirming on microtary drought duration and

20 direct comparison of the DSID and meteorological indicators, in particular for confirming or rejecting drought duration and the following wet periods.

On the contraryother hand, computing accumulated indicators implies a temporal smoothing and causing the drought period will to appear lagged in time, albeit; however for accumulation periods of 3 and 6 months the lag was found insignificant. DSIA and DIA are thus more robust against temporal and spatial GRACE noise as compared to DSID and DID, and again we

25 would suggest <u>utilizing</u> 3 or 6 months accumulation periods. In general, we found the Zhao- and Thomas-indicators <u>performing</u> <u>performed</u> better in detecting the correct drought strength than the Houborg-method, at least <u>seen for</u> the limited duration of the GRACE time series that we have at the time of writing.

By simulating the effect of spatial noise on drought detection, we found that some indicators appear less robust. Analysis of the percentage of drought affected area showed that the GRACE spatial noise limits the correct drought detection. Again,

- 30 the DSIA was identified to be more robust as compared to DSI and DSID it was the only indicator that identified exceptional drought in nearly all grid cells. A second experiment was applied to examine , conducted to examine if the influence of the spatial noise can be reduced by using spatial averages. We found that spatially averaging DSI and DI appears less robust against the spatial noise compared to computing the indicator of averaged TWSC. At this point we therefore suggest to compute the indicator from spatially averaged TWSC. Since the DI showed stronger difference between both averaging methods than the
- 35 DSI, we conclude that the DI is generally less robust against spatial noise than the DSI. In our real-data case study, due to these

findings, the DSIA6 was then thus applied to GRACE-TWSC, and it identified two drought periods: mid 2003 to mid 2006 in Central and South East and 2015 to 2016 in North East of South Africa.

#### 6 Conclusions and outlook

A framework has been developed that enables understanding the masking of drought signals when applying the Zhao et al. (2017), Houborg et al. (2012) and Thomas et al. (2014) methods. Four new GRACE-based indicators (DSIA, DSID, DIA and DID) were derived and tested; these are modifications of the above mentioned approaches and work with based on time-accumulated and -differenced GRACE data. We found that indeed most indicators were mainly sensitive to water storage trends and to the GRACE-typical spatial noise.

Among these various indicators, we identified the DSIA6 as in particular particularly well-performing, i.e. it is less sensitive
 to GRACE noise and with good skills in capability towards identifying the correct severity of drought, at least in absence of trends. However, the choice of the indicator should always be made in the light context of the application.

We see ample possibilities to extend our framework. Future work should focus on better defining the begin onset and end of a drought and developing a signature for TWSC drought. One will should also consider other observables in the simulation such as e.g. groundwater, such as groundwater for example, which can be derived from GRACE and by removing other storage

15 contributions from direct modelling or through data assimilation.

In the GRACE community, efforts are currently being made to 'bridge' the GRACE timeseries to the <u>begin\_beginning</u> of the GRACE-FO data period (e.g. Jäggi et al., 2016; Lück et al., 2018). These gap-filling data will inevitably have much higher noise and spatial correlations that may be very different from GRACE data, and drought detection <u>skills\_capability</u> should be investigated through simulation first. On the <u>contraryother hand</u>, GRACE-FO is supposed to provide more precise measure-

20 ments, and thus less influence of spatial noise on the drought detection may be expected. The combination of GRACE-FO data and a thorough understanding and 'tuning' of GRACE drought identification methods, possibly through this framework, might then enable us to identify water storage droughts more precisely.

#### Appendix A: AR model coefficients computations

To extract temporal correlations from the GRACE total water storage changes (TWSC) we apply an autoregressive(AR) model, which is described by

$$X(t) = \phi_1 X(t-1) + \dots + \phi_p X(t-p) + \epsilon_t,$$
(A1)

where X represents the observed process at time t, p is the model order,  $\phi$  are the correlation parameters, and  $\epsilon$  is a white noise process (Akaike, 1969). Here, detrended and deseasoned TWSC are used as the observed process X(t), because the remaining residuals contain interannual and subseasonal signal as the drought information, which we want to extract with this

30 approach. The approach is then applied for different model orders. The optimal order of the AR-model is adjusted by means of the information criteria, for example the Akaike information criterion (AIC), and the Bayes information criterion (BIC). Then, by using the optimal order, the AR-model coefficients  $\phi$ , which represent the temporal correlations, can be computed using a least squares adjustment.

The results for the optimal order of interannual and subseasonal TWSC is shown in Fig. A1. The most of the global land grids of detrended and deseasoned TWSC shows an optimal order of 1 (about 70%).



Figure A1. Histogram of the optimal order of an AR model for global detrended and deseasoned GRACE-TWSC on land grids.

#### 5 Appendix B: EM-Clustering

Expectation maximization (EM) represents a popular iterative algorithm that is widely used for clustering data. EM partitions data into cluster of different sizes and aims at finding the maximum likelihood of parameters of a predefined probability distribution (Dempster et al., 1977). In case of a Gaussian distribution the EM-algorithm maximizes the Gaussian mixture parameters, which are the Gaussian mean  $\mu_k$ , covariance  $\Sigma_k$ , and mixing coefficients  $\pi_k$  (Szeliski, 2010). The algorithm then

10 iteratively applies two consecutive steps to maximize the parameters: the expectation step (E-step) and the maximization step (M-step). Within the E-step we estimate the likelihood that a data point  $x_t$  is generated from the k-th Gaussian mixture by E-step:

$$z_{ik} = \frac{1}{Z_i} \pi_k \mathcal{N}(x|\mu_k, \Sigma_k), \tag{B1}$$

The M-step then re-estimates the parameters for each Gaussian mixture:

15 M-step:

$$u_k = \frac{1}{N_k} \sum_i z_{ik} x_i, \tag{B2}$$

$$\Sigma_k = \frac{1}{N_k} \sum_{i} z_{ik} (x_i - \mu_k) (x_i - \mu_k)^T,$$
(B3)

$$\pi_k = \frac{N_k}{N} \tag{B4}$$



Figure B1. Clusters based on EM-clustering applied to the global AR-model coefficients.

by using the number of points assigned to each cluster via

$$N_k = \sum_{i} z_{ik}.$$
(B5)

Using the maximized parameters EM assigns each data point to a cluster. The final global distributed clusters of the AR-parameters (Fig. 3) are shown in Fig. B1. These clusters were derived by modifying and applying an EM-algorithm provided by Chen (2018).

#### Appendix C: Eigen value decomposition

5

The decomposition of the variance-covariance matrix  $\Sigma$  by using Cholesky decomposition fails, when  $\Sigma$  is positive semi definite. To still be able to decompose the matrix, we can use eigen value decomposition, but this is accompanied by a loss of information due to the rank deficiency. The decomposition is then examined by  $\Sigma = UDU^T$ , where U is a matrix with the eigenvectors of  $\Sigma$  in each column and D is a diagonalmatrix of the eigenvalues. In this case, a decomposed matrix can be related to  $R^T$  introduced in Sec 3.1.  $R^T$  can be computed by  $U\sqrt{D}$ . In Sec. 3.1, we multiply  $R^T$  with a normal distributed noise time series of the same length as the rows of  $\Sigma$ . In this case, the number of normal distributed noise time series n is then replaced by the rank of  $\Sigma$ .

*Author contributions.* HG, OE, and JK designed all computations and HG carried them out. HG prepared the manuscript with contributions from OE and JK.

*Acknowledgements.* We acknowledge the funding from the Federal Ministry of Education and Research in Germany (BMBF) for the 'GlobeDrought' project through through its funding measure Global Resource Water (GRoW) (FKZ 02WGR1457A).

#### References

5

A, G., Wahr, J., Zhong, S.: Computations of the viscoelastic response of a 3-D compressible Earth to surface loading: an application to Glacial Isostatic Adjustment in Antarctica and Canada, Geophysical Journal International, 192, pp. 557—572, 2013.

Agboma, C. O., Yirdaw, S. Z. and Snelgrove, K. R.: Intercomparison of the total storage deficit index (TSDI) over two Canadian Prairie catchments, Journal of Hydrology, 374(3–4), 351–359, doi:10.1016/j.jhydrol.2009.06.034, 2009.

Akaike, H.: Fitting autoregressive models for prediction, Annals of the Institute of Statistical Mathematics, 21(1), 243–247, doi:10.1007/BF02532251, 1969.

Alpaydin, E.: Introduction to machine learning, MIT press., 2009.

Andersen, O. B., Seneviratne, S. I., Hinderer, J. and Viterbo, P.: GRACE-derived terrestrial water storage depletion associated with the 2003

European heat wave, Geophysical Research Letters, 32(18), doi:10.1029/2005GL023574, 2005.
 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: Drought indicators revisited, Wiley Interdisciplinary Reviews: Water, 3(4), 516–536, doi:10.1002/wat2.1154, 2016.

Changnon, S. A.: Detecting Drought Conditions in Illinois, Circular, 169, 41, 1987.

- 15 Checchi, F. and Robinson, W. C.: Mortality among populations of southern and central Somalia affected by severe food insecurity and famine during 2010-2012, Food and Agriculture Organization of the United Nations., 2013.
  - Chen, J. L., Wilson, C. R., Tapley, B. D., Yang, Z. L. and Niu, G. Y.: 2005 drought event in the Amazon River basin as measured by GRACE and estimated by climate models, Journal of Geophysical Research, 114(B5), doi:10.1029/2008JB006056, 2009.
- Chen, J. L., Wilson, C. R., Tapley, B. D., Longuevergne, L., Yang, Z. L. and Scanlon, B. R.: Recent La Plata basin drought conditions
  observed by satellite gravimetry, Journal of Geophysical Research, 115(D22), doi:10.1029/2010JD014689, 2010.
- Mo Chen: EM Algorithm for Gaussian Mixture Model (EM GMM) (https://www.mathworks.com/matlabcentral/fileexchange/26184-emalgorithm-for-gaussian-mixture-model-em-gmm), MATLAB Central File Exchange. Retrieved September, 2018.

Cheng, M., Ries, J. C. and Tapley, B. D.: Variations of the Earth's figure axis from satellite laser ranging and GRACE, Journal of Geophysical Research, 116(B1), doi:10.1029/2010JB000850, 2011.

- 25 Coelho, C. A. S., de Oliveira, C. P., Ambrizzi, T., Reboita, M. S., Carpenedo, C. B., Campos, J. L. P. S., Tomaziello, A. C. N., Pampuch, L. A., Custódio, M. de S., Dutra, L. M. M., Da Rocha, R. P. and Rehbein, A.: The 2014 southeast Brazil austral summer drought: regional scale mechanisms and teleconnections, Climate Dynamics, 46(11–12), 3737–3752, doi:10.1007/s00382-015-2800-1, 2016.
  - Dempster, A. P., Laird, N. M. and Rubin, D. B.: Maximum Likelihood from Incomplete Data via the EM Algorithm, Journal of the Royal Statistical Society, 39(1), 1–38, 1977.
- 30 Eicker, A., Schumacher, M., Kusche, J., Döll, P. and Schmied, H. M.: Calibration/Data Assimilation Approach for Integrating GRACE Data into the WaterGAP Global Hydrology Model (WGHM) Using an Ensemble Kalman Filter: First Results, Surveys in Geophysics, 35(6), 1285–1309, doi:10.1007/s10712-014-9309-8, 2014.
- Eicker, A., Forootan, E., Springer, A., Longuevergne, L. and Kusche, J.: Does GRACE see the terrestrial water cycle "intensifying"?:
   WATER CYCLE INTENSIFICATION WITH GRACE, Journal of Geophysical Research: Atmospheres, 121(2), 733–745,
   doi:10.1002/2015JD023808, 2016.
  - EM-DAT: The Emergency Events Database Université catholique de Louvain (UCL) CRED, D. Guha-Sapir www.emdat.be, Brussels, Belgium, accessed 2018.

- Espinoza, J. C., Ronchail, J., Guyot, J. L., Junquas, C., Vauchel, P., Lavado, W., Drapeau, G. and Pombosa, R.: Climate variability and extreme drought in the upper Solimões River (western Amazon Basin): Understanding the exceptional 2010 drought, Geophysical Research Letters, 38(13), n/a-n/a, doi:10.1029/2011GL047862, 2011.
- Frappart, F., Papa, F., Santos da Silva, J., Ramillien, G., Prigent, C., Seyler, F. and Calmant, S.: Surface freshwater storage and dynamics in the
- Amazon basin during the 2005 exceptional drought, Environmental Research Letters, 7(4), 044010, doi:10.1088/1748-9326/7/4/044010, 2012.
- Frappart, F., Ramillien, G. and Ronchail, J.: Changes in terrestrial water storage versus rainfall and discharges in the Amazon basin, International Journal of Climatology, 33(14), 3029–3046, doi:10.1002/joc.3647, 2013.

GADM database, URL: www.gadm.org, version 3.4, 2018.

5

- 10 Girotto, M., De Lannoy, G. J. M., Reichle, R. H. and Rodell, M.: Assimilation of gridded terrestrial water storage observations from GRACE into a land surface model, Water Resources Research, 52(5), 4164–4183, doi:10.1002/2015WR018417, 2016.
  - Houborg, R., Rodell, M., Li, B., Reichle, R. and Zaitchik, B. F.: Drought indicators based on model-assimilated Gravity Recovery and Climate Experiment (GRACE) terrestrial water storage observations, Water Resources Research, 48(7), doi:10.1029/2011WR011291, 2012.
- 15 Humphrey, V., Gudmundsson, L. and Seneviratne, S. I.: Assessing Global Water Storage Variability from GRACE: Trends, Seasonal Cycle, Subseasonal Anomalies and Extremes, Surveys in Geophysics, 37(2), 357–395, doi:10.1007/s10712-016-9367-1, 2016.
  - IPCC: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535 pp, 2013.
- 20 Jäggi, A., Dahle, C., Arnold, D., Bock, H., Meyer, U., Beutler, G., and van den IJssel, J.: Swarm kinematic orbits and gravity fields from 18 months of GPS data, Adv. Space Res., 57, 218–233, https://doi.org/10.1016/j.asr.2015.10.035, 2016.
  - Keyantash, J. and Dracup, J. A.: The Quantification of Drought: An Evaluation of Drought Indices, Bulletin of the American Meteorological Society, 83(8), 1167–1180, 2002.

Kusche, J., Schmidt, R., Petrovic, S. and Rietbroek, R.: Decorrelated GRACE time-variable gravity solutions by GFZ, and their validation

- 25 using a hydrological model, Journal of Geodesy, 83(10), 903–913, doi:10.1007/s00190-009-0308-3, 2009.
  - Kusche, J., Eicker, A., Forootan, E., Springer, A. and Longuevergne, L.: Mapping probabilities of extreme continental water storage changes from space gravimetry, Geophysical Research Letters, 43(15), 8026–8034, doi:10.1002/2016GL069538, 2016.
    - 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(13), 3395–3401, doi:10.1002/grl.50655, 2013.
- 30 Lück, C., Kusche, J., Rietbroek, R., and Löcher, A.: Time-variable gravity fields and ocean mass change from 37 months of kinematic Swarm orbits, Solid Earth, 9, 323-339, https://doi.org/10.5194/se-9-323-2018, 2018.

Lyon, B. and Barnston, A. G.: ENSO and the Spatial Extent of Interannual Precipitation Extremes in Tropical Land Areas, Journal of Climate, 18(23), 5095–5109, doi:10.1175/JCLI3598.1, 2005.

Malherbe, J., Dieppois, B., Maluleke, P., Van Staden, M. and Pillay, D. L.: South African droughts and decadal variability, Natural Hazards,

35 80(1), 657–681, doi:10.1007/s11069-015-1989-y, 2016.

Mann, M. E. and Gleick, P. H.: Climate change and California drought in the 21st century, Proceedings of the National Academy of Sciences, 112(13), 3858–3859, doi:10.1073/pnas.1503667112, 2015.

- Masih, I., Maskey, S., Mussá, F. E. F., and Trambauer., P.: A review of droughts on the African continent: a geospatial and long-term perspective, Hydrology and Earth System Sciences, 18(9), 3635-3649, 2014.
- Mayer-Gürr, T., Behzadpour, S., Ellmer, M., Kvas, A., Klinger, B., Zehentner, N.: ITSG-Grace2016 Monthly and Daily Gravity Field Solutions from GRACE. GFZ Data Services. http://doi.org/10.5880/icgem.2016.007, 2016.
- 5 McKee, T. B., Doesken, N. J. and Kleist, J.: The relationship of drought frequency and duration to time scales, pp. 179–183, American Meteorolocial Society, Anaheim, CA., 1993.
  - Mishra, A. K. and Singh, V. P.: A review of drought concepts, Journal of Hydrology, 391(1–2), 202–216, doi:10.1016/j.jhydrol.2010.07.012, 2010.

Moore, J., Woods, M., Ellis, A. and B. Moran, B.: Aerial survey results: California, 2016.

- 10 Parthasarathy, B., Sontakke, N. A., Monot, A. A., and Kothawale, D. R.: Droughts/floods in the summer monsoon season over different meteorological subdivisions of India for the period 1871–1984, Journal of Climatology, 7(1), 57-70, 1987.
  - Rebetez, M., Mayer, H., Dupont, O., Schindler, D., Gartner, K., Kropp, J. P. and Menzel, A.: Heat and drought 2003 in Europe: a climate synthesis, Annals of Forest Science, 63(6), 569–577, doi:10.1051/forest:2006043, 2006.

Redner, R. A. and Walker, H. F.: Mixture Densities, Maximum Likelihood and the EM Algorithm, SIAM Review, 26(2), 195-239,

15 doi:10.1137/1026034, 1984.

25

Rodell, M., Famiglietti, J. S., Wiese, D. N., Reager, J. T., Beaudoing, H. K., Landerer, F. W. and Lo, M.-H.: Emerging trends in global freshwater availability, Nature, 557(7707), 651–659, doi:10.1038/s41586-018-0123-1, 2018.

Rouault, M., and Richard, Y.: Intensity and spatial extension of drought in South Africa at different time scales, water SA, 29(4), 489-500, 2003.

- 20 Rouault, M. and Richard, Y.: Intensity and spatial extent of droughts in southern Africa, Geophysical Research Letters, 32(15), 2005.
  - Seitz, F., Schmidt, M. and Shum, C. K.: Signals of extreme weather conditions in Central Europe in GRACE 4-D hydrological mass variations, Earth and Planetary Science Letters, 268(1–2), 165–170, doi:10.1016/j.epsl.2008.01.001, 2008.
    - Springer, A.: A water storage reanalysis over the European continent: assimilation of GRACE data into a high-resolution hydrological model and validation, Ph.D. thesis, Rheinische Friedrich-Wilhelms Universität Bonn, Bonn. [online] Available from: urn:nbn:de:hbz:5n-53930, 2019.
    - Kronberg, E. A.: Dynamics of the Jovian Magnetotail, Ph.D. thesis, International Max Planck Research School, Universities of Braunschweig and Göttingen, Germany, 133 pp., 2006.

Swenson, S. C., and Wahr, J.: Post-processing removal of correlated errors in GRACE data, Geophysical Research Letters, 33(8), 2006.

Swenson S. C., Chambers, D. P., and Wahr, J.: Estimating geocenter variations from a combination of GRACE and ocean model output,

Journal of Geophysical Research-Solid Earth, 113(B8), 2008.
 Szeliski, R.: Computer Vision: Algorithms and Applications, Springer Science and Business Media, 2010.
 Svoboda, M.and Fuchs, B.: Handbook of Drought Indicators and Indices, in Drought and Water Crises, pp. 155–208, CRC Press., 2016.
 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(5), 1537–1545, doi:10.1002/2014GL059323, 2014.

- Tsakiris, G.: Drought Risk Assessment and Management, Water Resources Management, 31(10), 3083–3095, doi:10.1007/s11269-017-1698-2, 2017.
  - Van Loon, A. F.: Hydrological drought explained: Hydrological drought explained, Wiley Interdisciplinary Reviews: Water, 2(4), 359–392, doi:10.1002/wat2.1085, 2015.

- 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(7), 1696–1718, doi:10.1175/2009JCLI2909.1, 2010.
- Vogel, C., Koch, I. and Van Zyl, K.: "A Persistent Truth"—Reflections on Drought Risk Management in Southern Africa, Weather, Climate, and Society, 2(1), 9–22, doi:10.1175/2009WCAS1017.1, 2010.
- 5 Voss, K. A., Famiglietti, J. S., Lo, M., de Linage, C., Rodell, M. and Swenson, S. C.: Groundwater depletion in the Middle East from GRACE with implications for transboundary water management in the Tigris-Euphrates-Western Iran region, Water Resources Research, 49(2), 904–914, doi:10.1002/wrcr.20078, 2013.

Wahr, J., Molenaar, M. and Bryan, F.: Time variability of the Earth's gravity field: Hydrological and oceanic effects and their possible detection using GRACE, Journal of Geophysical Research: Solid Earth, 103(B12), 30205–30229, doi:10.1029/98JB02844, 1998.

10 Wilhite, D. A.: Droughts: A Global Assessment, Routledge., 2016.

Yi, H. and Wen, L.: Satellite gravity measurement monitoring terrestrial water storage change and drought in the continental United States, Scientific Reports, 6(1), doi:10.1038/srep19909, 2016.

- Zaitchik, B. F., Rodell, M. and Reichle, R. H.: Assimilation of GRACE Terrestrial Water Storage Data into a Land Surface Model: Results for the Mississippi River Basin, Journal of Hydrometeorology, 9(3), 535–548, doi:10.1175/2007JHM951.1, 2008.
- 15 Zargar, A., Sadiq, R., Naser, B. and Khan, F. I.: A review of drought indices, Environmental Reviews, 19(NA), 333–349, 2011. Zhang, Z., Chao, B. F., Chen, J. and Wilson, C. R.: Terrestrial water storage anomalies of Yangtze River Basin droughts observed by GRACE and connections with ENSO, Global and Planetary Change, 126, 35–45, doi:10.1016/j.gloplacha.2015.01.002, 2015.

Zhao, M., Velicogna, I. and Kimball, J. S.: A global gridded dataset of GRACE drought severity index for 2002-14: Comparison with PDSI and SPEI and a case of the Australia millenium drought, Journal of Hydrometeorology, 18(8), 2117–2129, 2017.