Review: A quantitative analysis to objectively appraise drought indicators and model drought impacts, by S**. Bachmair, C. Svensson, J. Hannaford, L. J. Barker, and K. Stahl,** *Hydrology and Earth System Science Discussions*, **12**, 9437-9488.

And

Exploring the link between drought indicators and impacts, by S. Bachmair, I. Kohn, and K. Stahl, *Natural Hazards and Earth System Science*, **15**, 1381

W. Pozzi, Group on Earth Observations (GEO)

The authors are to be congratulated on developing and deploying the European Drought Impact Report Inventory over the course of the Drought R & SPI project; the English co-authors of this paper, as well, have been carrying out long standing, important work reconstructing the chronology and impacts of drought episodes in the United Kingdom. To summarize the content of this paper, surface-water storage responds rapidly to changes in high frequency precipitation events daily or across the season, but soil moisture responds more slowly and is thereby only influenced by longer term precipitation deficits (perhaps weekly or within a season), causing agricultural droughts. On the other hand, groundwater storage responds very much more slowly in comparison with soil moisture and is usually changed by changes in precipitation extending from the seasonal to annual and interannual time scale. The groundwater response lags behind those in surface water and soil moisture, and groundwater takes significantly longer to be replenished and recover (Bloomfield and Marchant 2013). Any type of activity in which groundwater is involved such as recharge of base flow of streams will thereby be impacted by a protracted drought that diminishes groundwater supply. The subsequent declining water levels in rivers can impede river transportation, water supplies drawn from rivers (or the groundwater source), or impact thermal power plants drawing water from rivers. Other activities, such as seasonal or annual agricultural would also be impacted by soil moisture over a shorter, seasonal time scale. The authors also attempt to construct multiple indicators for these stores of water, i.e., groundwater level percentiles and streamflow percentiles. By assembling impact data, the authors are able to corroborate and test these relationships.

Unfortunately, subsequent to deployment of the RF methodology (see paper), the authors begin expressing drought in terms of a single drought variable which conflicts with their earlier approach taken within the paper of using multiple indicators to monitor more of drought's different impacts. On top of that, single drought indicators (Standardized Precipitation Index or Standardized Precipitation Evapotranspiration Index) are then identified as possible "triggers," a so-called threshold in a drought management plan at which some form of remediation is evoked. Why is identification of a single threshold with a so-called “drought trigger” a misstep? Realistic proactive drought management plans are prepared in extensive workshops or surveys in which the subgroups within a population are identified through interviews or surveys, in order to discern the vulnerable population group subject to the highest hazard. Furthermore, during different drought episodes, different subgroups may be affected (the very multiple processes identified in some of the impact reports by the authors). However, the problem is that preparation of the European Drought Impact report Inventory is a very labor-intensive process. In addition, the project is not continually funded as is the Drought Reporter in the USA. As a result, the authors do not have a sufficiently high sample size; correspondingly, they are required to “scale up” to the NUTS1 level in order to increase their sample size. Even at the coarse spatial level of a federal state in Germany, they are only able to discern the differences between hydrological impacts and “other impacts,” given the somewhat small sample size (sample size is larger for specific drought episodes, such as the 2003 European heat wave and the 2011 drought episode). Drought management plans—especially designation of triggers—has to be from the ground up, identifying all the groups affected. Having data based at the state level only, and not having access to data at finer spatial scales (for example, NUTS2 and NUTS3) (due to lack of adequate sample size), the discussion of “thresholds” and “triggers,” particularly for local, high resolution, site specific activities such as impedance of waterborne navigation and power plant outages (i.e., “hydrological” impacts), is not germane and relevant.

The authors are to be commended for attempting to examine the cross correlation among different stores of water versus their hydrological and agricultural impacts entries. I’m proposing to the editor that the sections of their paper dealing with single indicator drought thresholds and triggers be dropped (see below), but allowing the remainder of the paper to continue, the correlation and machine learning technique sections testing the relationships among drought indicators and, among the three categories of drought impacts: that of total drought impacts, “hydrological” drought impacts, and drought impacts other than hydrological. The authors are invited to offer *local* evidence, at the NUTS3 scale, that triggers can be developed out of their numerical frequency of impacts data (or alternatively through detailed case study in individual droughts) at that higher spatial resolution.

The real main conclusion of their study is: “***Agricultural and hydrological drought impacts were generally best linked to shorter and longer SPI (and SPEI) time scales, respectively. Here, shorter and longer refer to 1-4 (Germany) and 7-8 months (England).*** Having multiple time scales present means that a drought management plan needs to be formulated with accommodation made for the different time scales over which impacts manifest. A drought management plan would have in place different remedial actions operative over different time scales.

The authors could, as well, conclude with an assessment of how many reports would have to actually be prepared (how larger a sample size) would be required in order to resolve some of these impacts at the NUTS2 and NUTS3 level.

Where triggers and single thresholds are cited in the Paper

"The correlation analysis for Germany yielded more powerful results. *However, the method does not provide further information, such as thresholds of impact occurrence, such as provided by the RF algorithm.* ***We regard splitting values during recursive partitioning as estimates of thresholds of impact occurrence, because they provide guidance on critical predictor values triggering a consequence*** (P.9463 line 3). *The reason why we consider free splitting values as meaningful thresholds is Bachmair et al 2015 found a similar threshold pattern (*differences between southern and central Germany and northern Germany) *using the same impact data but a different approach* (9464).

"Triggers" and any effort to establish a drought management plan which involves either users or people impacted by a drought has to be a bottom up process, i.e., grounded at the local level. The local level is not the federal state level. The process of establishing so-called "triggers" requires extensive workshop participation among stakeholders and the different parties affected during droughts. Studies in vulnerability are legion within the development community (and, in point of fact, are as extensive a literature as the drought indicator literature). Studies in vulnerability and impact assessment are commonly based upon methodologies of interviewing and questioning different subgroups within a country in order to identify vulnerable groups which will be targeted during a drought (and as the authors point out, different droughts may target different groups, adding to the complexity in formulating drought management plans). Vulnerability studies have been carried out by the European Framework Drought Early Warning System for Africa (DEWFORA) and numerous other studies, supported by multiple countries. The drought impact report inventory is not a "short cut" bypassing the need for workshops with affected water users. To provide a concrete example, during some of the droughts in Germany, water surface elevation within a river declined to the point that power plant outages occurred, as, for example, might occur due to lack of availability of cooling water to supply a once-through cooling system. Alternatively, a declining river level would also impede waterborne transportation. Obviously, river channel cross sections and hydrography will vary throughout the state or district, so that impacts (and suggestions for guidelines at which water level mandated actions might occur) will have to be assessed at the local level. One cannot assess these at such an aggregate level as NUTS1, at the federal state level.

As a point of information, the authors had previously acknowledged in their earlier paper (Bachmair et al 2015) that "*An individual drought indicator is not capable of* *representing the diversity and complexity of drought conditions across space and time."* They also state *"Our analysis reveals a single one size fits all indictor threshold does not exist."* This position is correct, but, for some reason, they seem to be somewhat reversing themselves within this paper, by proposing a single indicator as a trigger. If there are multiple impacts occurring over multiple time scales, then obviously multiple indicators are required. I understand that they are trying to find an index with which to identify differences among impact types. That is a research question. To utilize language involved in drought management plans, on the other hand, is an application. The data stores present within the two papers is identical, excepting the addition of the UK data whose analysis is also carried out at the NUTS1 level.

While more impact reports are available during the 2003 European heat wave and the 2011 drought episodes, apparently the number of cases per month during these two drought episodes is still too low to support data analysis at a higher resolution, i.e., at the NUTS2 level. However, while such impacts are local, the authors might be able to make a case for situations in which an underlying aquifer has a larger spatial domain, as well as a spatial extent approaching that of the federal state level. Under these conditions, impacts arising from the aquifer, in common across the spatial domain, might result in commonality among the impacts in a larger area, approaching that of NUTS1. This has already been shown to be true over large parts of the United Kingdom, in which many drought episodes impacted groundwater (and streamflow) across much of England (Marsh, Cole, and Wilby 2007). This conclusion would be true, so long as the variability present at each borehole would not prohibit a common drought signature being detected over a larger area (see below).

Why the Study was Aggregated to the NUTS1 Level

An illustration of the number frequency of impact entries per month at the NUTS1, NUTS2, and NUTS3 level is provided in Stahl et al. (Figure 1):

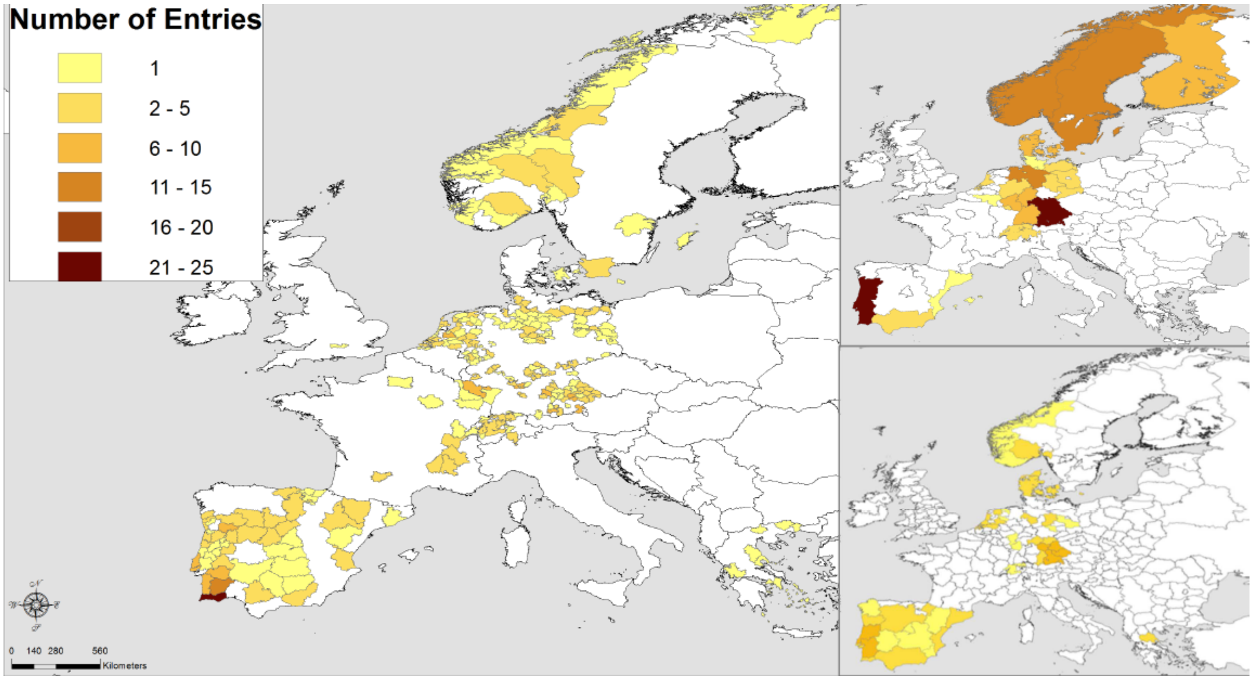

Figure 1 A comparison of the distribution of number frequency of EDII impacts at the NUTS-3, NUTS-2 and NUTS-1 level (Stahl, et al 2012:"A European Drought Impact Report Inventory (EDII): Design and Test for Selected Recent Drought Events in Europe," Drought R&SPI Technical Report Number 3.

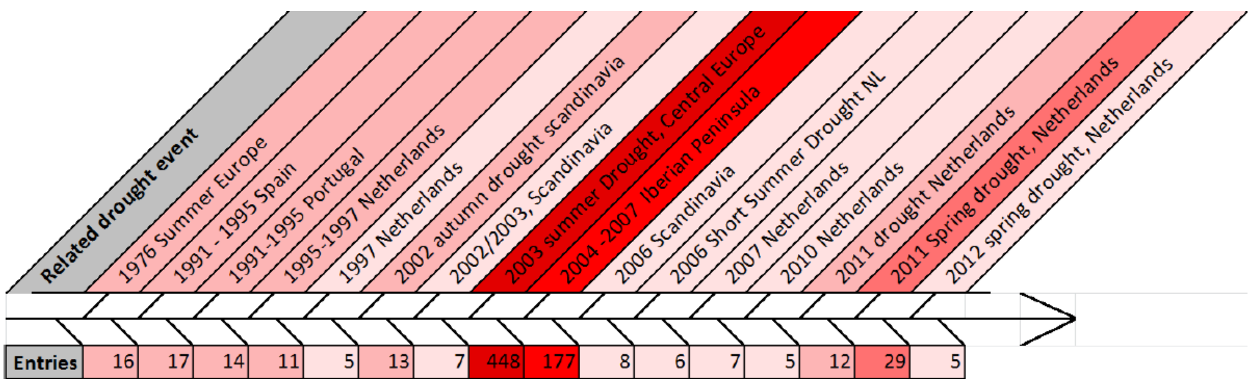

Figure 2 Distribution of number frequency of EDII impacts falling within recorded drought episodes over Europe (Stahl, et.al. 2012)

To be fair, the authors compiled their data at a later date than these two figures (Figures 1 and 2), so that more reports and entries were available. How many more? This is shown in Table 1 of the paper under review (see Table 1).

Country	NUTS1 region name	NUTS1 region abbr.	N_I	Length of censored timeseries (months)	Percentage of months with $N_I > 0$	No. streamflow stations	No. groundwater stations
UK	North East	NEE	28	48	22.9	9	–
UK	North West	NWE	400	120	35.8	16	–
UK	Yorkshire and the Humber	YHU	213	108	32.4	11	–
UK	East Midlands	CEE	345	120	37.5	13	–
UK	Wales	WAL	884	156	35.9	20	–
UK	West Midlands	CWE	310	96	42.7	12	–
UK	East of England	EE	545	156	50.0	12	–
UK	South West	SWE	456	156	57.1	23	–
UK	South East	SEE	1079	168	57.1	23	–
UK	London	LND	291	144	45.1	1	–
DE	Schleswig-Holstein	SH	34	60	25.0	9	9
DE	Mecklenburg-Western Pomerania	MP	54	96	28.1	7	4
DE	Lower Saxony	LS	107	132	28.0	38	42
DE	Saxony-Anhalt	ST	46	96	22.9	16	14
DE	Brandenburg	BB	114	96	30.2	21	18
DE	Berlin	BE	57	72	23.6	–	–
DE	North Rhine-Westphalia	NW	143	84	34.5	23	18
DE	Hesse	HE	95	60	43.3	19	18
DE	Saxony	SX	50	96	31.3	23	10
DE	Rhineland-Palatinate	RP	182	84	35.7	20	18
DE	Saarland	SL	42	36	30.6	3	–
DE	Baden-Wuerttemberg	BW	228	84	39.3	28	15
DE	Bavaria	BV	382	72	33.3	69	26

Table 1(from the paper under review)

A table from the second paper divides up the Germany impact data by drought episode (Table 2)(Bachmair et al 2015).

Table 2. Information on selected drought events: duration and number of drought impact occurrences and onsets.

Drought event	Jun–Dec 1971	Feb–Aug 1976	Mar–Aug 1992	Feb 2003–Feb 2004	Jun–Aug 2006	Jan–Dec 2011
Duration (months)*	7	7	6	13	3	12
$n\,I$	48	146	72	931	33	155
$n\,I_h$	18	82	18	736	12	111
$n\,I$ onset	15	55	30	422	19	88
$n\,I_h$ onset	5	33	6	340	12	63

* Event delineation based on impact occurrence (see Sect. 2.3). $n\,I$ = number of drought impact occurrences; $n\,I_h$ = number of hydrological drought impact occurrences; $n\,I$ onset = number of drought impact onsets; $n\,I_h$ onset = number of hydrological drought impact onsets.

As can be seen in Figure 1, by aggregating or scaling up from NUTS3 to NUTS2 to NUTS1, the sample size is being increased. The NUTS1 level of resolution is the resolution of the federal states in Germany; at this high level of aggregation, whether one can actually identify local impacts is questionable. Given the total number of federal states within Germany and the divisions of UK, the total number of cells (units within which impact reports are collected) is somewhat small.

The Groundwater Issue: Does the larger spatial extent of aquifers allow impacts to be resolvable at a NUTS1 level?

Bloomfield and Marchant, 2013, HESS, 17, 4769 presented actual raw time series plots for monthly groundwater levels from 17 wells across different aquifers in central and eastern England. The really important result from their study is that standardization is required, along with considerable non-parametric processing in order for the commonalities found among all the wells to emerge. And they did, indeed, find that drought signatures were present within the Standardized Groundwater Index. However, at the same time, they also discovered that differences existed among wells, and one possible conclusion was that the Standardized Precipitation Index time scale (accumulation period of SPI-1 for one month, SPI-3 for three months, etc) increased as the thickness of the unsaturated zone (vadose zone) increased. At the same time, unresolved factors are also still at work.

Kumar et al, 2015, "multiscale evaluation of the Standardized Precipitation Index as a groundwater drought indicator," HESS Discussions, 12, 7405 carried out a somewhat similar, but smaller study over the overlapping southern Germany spatial domain, as that of the current paper under review. Their Standardized Groundwater Indices, calculated somewhat differently, for monitoring wells in southern Germany, utilizing a different non-parametric technique, found similar conclusions as Bloomfield and Marchant 2013: they found common drought signatures, with German and Netherlands wells being impacted under different drought episodes across a common area; and they also discovered differences among the wells.

While drought signatures may exist in common across a region, there is a difference between impacts arising out of common water supplies being derived from a regional aquifer in common (although differences can exist there as well) and impacts arising from power plants sited along rivers. River channel cross sections will change throughout the regions, so that water surface elevation will be different at low stage along different bends and channel reaches. The impacts of power plants and navigation will be local, and a drought management plan would be prepared in stakeholder workshops with users at the local level, with each power plant operator providing the unique local conditions at his or her site. This is a higher resolution than the NUTS1 level..

With these studies in context, we can now turn to the paper under review. A very weak correlation was found between SPI and groundwater level percentiles. The first question might be whether this poor correlation arises because standardization methodology was not undertaken as had been the case with Bloomfield and Marchant (2013) and Kumar et al 2015. The second question might be whether the poor correlation arises due to aggregation of the data to the federal state level (including the hydrological index).