Hydrol. Earth Syst. Sci. Discuss., 9, C5040–C5043, 2012
www.hydrol-earth-syst-sci-discuss.net/9/C5040/2012/

Hydrology and Earth System Sciences Discussions

Interactive comment on "Data expansion: the potential of grey literature for understanding floods" *by* S. Uhlemann et al.

Anonymous Referee #1

Received and published: 1 November 2012

Overview

The authors stated purpose of this paper is to address the lack of consideration of event- specific flood research which is often published in non-peer reviewed publications ('grey literature') which effectively eliminates approximately 80% (their estimate) of the available information which would otherwise be available. The authors highlight their experience in obtaining this lost information from German-dominated governmental publications, and provide a characterisation of the grey literature which was obtained as a result of the extensive data mining performed. On a general note, I must commend the authors in their effort to access and assimilate the large body of information available within the grey literature. There is no doubt that this information resource is hugely

undervalued and underutilised, and it is therefore encouraging that effort is being made to incorporate the data into the scientific community. However, I feel that the authors did not utilise the available information to its fullest potential, predominantly due to limited search scope and analysis of the information obtained from the search, as outlined below.

Specific Comments

1. Being an academic within the Government sector myself, I am familiar with the size and scope of grey literature produced by Responsible Authorities. As such, I realise that despite our greatest efforts, the majority of information and relevant reports are often sat on desks and bookshelves rather than listed and available within search engines, predominantly due to issues with confidentiality, lack of resources and awareness of the use of this data to those outside of government departments. This is particularly true for those reports produced prior to the 'digital age'. As such, I feel that there is a fundamental flaw in the data mining process the authors have undertaken within this study. This, perhaps, is not insurmountable, provided the authors can provide an indication of how much (and particularly, what types of) literature may have been missed. If this MS is to be a 'seminal' work, highlighting the use and application of grey literature in flood research, I feel such validation is necessary. 2. Related to the point above, due to a lack of validation of the completeness and representative sampling of the grey literature, I feel that the results presented in Figure 4a & 6, and conclusions drawn from it (e.g. sentences 23-29 on pg 11071, and further paragraph on page 11072) cannot be made. It is possible that the trends observed in these figures are the result of the dawn of the digital age and data resulting accessibility, rather than a real result of increased reporting. This possibility is alluded to in sentence 23/24 on page 11072, but without validation, I don't feel this conclusion can be drawn, and would disagree with the reasoning dismissing this possibility. 3. Similarly, I would question the limitation of the search to those produced by high level German authorities only. In my experience, those reports produced by local, regional and district level equiva-

lents in the UK tend to include a larger scale consideration of local events to those in other (often intra-basin) areas. Reports from adjoining countries may also be of use, but given the search terms used in the mining, I feel that such publications would have been excluded. Exclusion of these reports at the very beginning of the data mining may have discounted a great deal of grey information, and may warrant further investigation to ensure this is not the case. 4. Considering the range of grey literature available; from brochures and websites to multi million Euro investigations, I feel that there should be some qualification or scoring system for the varying data sources, or at the very least, consideration of the perceived accuracy of each data source (in a similar way to the characterisation of the reports in Table 3). This would be particularly relevant for any openly accessible data base which the authors mention within the Introduction, but would also provide additional significance and novelty to the MS. Validation of the grey literature in this way would allow fellow researchers to gauge the usefulness and relevance of inclusion of grey literature into their own research. The authors may want to consider the approach of Norris et al which may assist in this manner.

Technical Comments

1. The MS requires a great deal of re-arranging and tidying-up, as a lot of the information within the methods should go in to either the intro or results (e.g. background information on systematic reviews, pgs 11057-11059); the conclusion section is far too long. Separation of the results and discussion may provide a better framework for laying out the MS, as otherwise, I find it long winded. There are also a few odd 'turns of phrase' throughout the MS, which hopefully, would be picked up by the editor. 2. Tables & Figures. Generally- why are some figures in colour and others only in black and white? This is especially confusing in Fig 3 a&b. Figure 2; I'm not quite sure what this figure is showing. Where there is no black bar present, does that imply there was no publication in that decade? If an event from that decade is published later, is it included in a black bar in another decade, or not represented at all? Figure 5; T 20a/ 50a should be defined in the caption. It would be interesting to know what the regressions

for 5b look like for comparison. Figure 6; difficult to grasp, especially so in black & white. Could a change in point shape be used to clarify? To conclude, I feel that the premise of the MS is a valid one, but I would question the validity of the search terms and scope of the resulting grey literature obtained. Due to the lack of validation of the choices in search terms/ methods used, I would view the resulting data and analysis to be skewed, and hence, may not support the conclusions and application of the study. Without such validation, I don't feel the MS could be accepted in its current form, nor would it provide a good example for fellow researchers seeking to include the grey literature into their own research. âĂČ References

Norris, R.H., Webb, J.A., Nichols, S.J., Stewardson, M.J. & Harrison, E.T. (2012). Analyzing cause and effect in environmental assessments: using weighted evidence from the literature. Freshwater Science 31(1): 5-21.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 9, 11049, 2012.