

***Interactive comment on “Content analysis for  
Agricultural Records of the United Kingdom,  
AD 200 to 1977: a study of frequency in human  
records concerning climate phenomenon” by  
D. H. Holt***

**A. Kiss**

kissandi@earth.geo.u-szeged.hu

Received and published: 30 September 2011

I do apologize for being so late with the review.

First of all, I think it is important to express that I do appreciate the significant efforts and the great amount of work spent on the statistical evaluation, coding of a huge database. I do agree with the author that once a scientist carries out such an investigation (if using proper database based on a representative set of reliable contemporary evidence with clearly no preference on any type of weather-related information), content analysis can

C1530

be an appropriate method.

Nevertheless, due to other, major problems, I cannot recommend this paper to be published in this scientific journal.

Main reasons are as follows: I. the database does not reach basic source quality requirements, and therefore cannot form the basis of any natural scientific analysis II. low quantity of data, fundamental differences in data coverage in time (and space) + unknown level of representativeness

Reasons and comments in more detail

I. the database does not reach basic source quality requirements

The type of analysis the author carried out is very case sensitive: this means that – in my opinion – we should be very careful to ask such a question even in case of a dataset, clearly based on contemporary original sources, and even in case of a ‘complete’ temporal and spatial coverage. In reality, such a question could be asked with reason only in the case when the collected, contemporary source material applied really means a full dataset of all documentation written by human beings (and all preserved) in a given period of time. A basic question has to be raised: Does this ‘compilation’ in any of its parts fulfill these requirements?

The author applied a documentary dataset published as a book entitled *Agricultural Records A.D. 220-1968*. The origin of this book goes back to another one, collected and put together by a farmer, Thomas H. Baker, on the events of his locality. This was the book entitled *Records of the Seasons, Prices of Agricultural Products and Phenomena Observed in the British Isles*, published in 1883. This later book was updated with many new data concerning a much broader area up to 1977. As the titles suggest, their main viewpoint was not necessarily climate- or weather-based, although majority of information still referred to some weather(-related) phenomena. As we know, compilations have several disadvantages such as misdating, doubling, tripling of events etc.

C1531

– we can find several references in the scientific literature discussing this problem in detail. In our case, however, the problem is more complex.

Concerning the real nature of the book, applied by the author as a basis of analysis, may I just quote from a review of the above-mentioned book (I personally would not even call it as a compilation, because source compilations are usually of higher quality with giving massive hints where the collectors collected their data), from the following article, written by Hilary Ely on *Vulpes Libris* (London), a bibliophil website:

"The sad defect of this book is apparent right from the start – there are no notes, and no references. There is absolutely no information about where these facts were found, until close to the end, when for a few years information published in *The Times* is reproduced with permission. . . . I am convinced that the records that date from the lifetime of Thomas Baker and John Stratton are partly at least based on their own records or diaries – but nowhere does it say so. When it comes to information from the Dark Ages, it is really frustrating not to know how this information has been transmitted."

Being a natural scientist and a historian, I personally had the opportunity to get acquainted (among others) with the same type of 'private' collections (without any reference on sources, collecting impressive amount of all sorts of weather- and agriculture-related evidence – whatever it was, wherever it came from), and unfortunately – except for approximately the last one century they referred to – they were clearly unreliable and of low (or no) scientific value, and by no means comparable to the 'real' data coming from real contemporary (primary) source evidence. In such collections, data is collected in a rather subjective way, especially concerning early periods. It is really up to the collector's personal choice and also greatly of his possibilities what to include and what to neglect (therefore, in the past, social status of the collector did matter a lot). Here we have to make a differentiation, because it is especially true for historical times well before the collection was made. If the collector collects contemporary evidence, then there is wide choice of information available. However, going back in time, data is much more selected, and even those selection criteria are completely unknown,

C1532

and might be independent from the collector in person (i.e. the collector himself cannot usually decide over the reliability of his collected data). This independence, of course, does not mean objectivity in itself.

For comparison: as if a dendrochronologist or a dendroclimatologist would be forced to work with tree samples (and should count the rings and measure ring width) provided only on photos, in which case one cannot even decide whether there are real tree samples on the photo (then the angle of taking the picture or shadows etc. can also matter) or it is just a photo-montage or even a complete fake tree made of plastic, and rings painted on. It is rather impossible to call the upcoming results as real scientific results, and draw serious scientific conclusions.

All in all, this type of data collection is well under the value even of a collection called 'weather compilation' (e.g. *Briton for Britain*; whose direct application would be already problematic in such an analysis). Thus, in terms of quality, the database is clearly inappropriate for such an evaluation.

II. the results achieved cannot be interpreted due to low quantity of data and fundamental differences in data coverage in time (and space)

1. There is a very low density of data; the coverage is inappropriate to answer any sort of mass-analysis questions. Looking at Fig. 1., data before 1000, but partly up to the mid-17th century have anyway practically no importance, even compared to the known documentary source evidence preserved in reality. This also means that for the "Little Climate Optimum" only a very marginal amount of data is available in this database at all – not appropriate for any comparison. 2. In this respect, I am wondering why only the LCO and the LIA are compared, and why not the recent warm phase (with visibly highest density of data) with the LIA and/or the LCO? The 20th century, for example, I cannot detect in the comparative analysis, although majority of all data applied comes from this late period. Of course, representativeness in a statistical sense does not necessarily mean huge amounts. But the sample is representative only if pro-

C1533

vides information which is not (primarily) disturbed by any selection criteria (in subject, time and space etc.). In our case the level of 'disturbance' is, for example, completely unknown. Therefore, the level of representativeness is unknown.

I have a number of other questions and comments, such as the meaning of the category of 'Climate Recorded' in Fig. 3, or the relevance of the category 'Earthquake' in the present investigation etc. Nevertheless, I think their importance is at the moment somewhat marginal compared to the main points listed above.

---

Interactive comment on Clim. Past Discuss., 7, 2555, 2011.

C1534