

## ***Interactive comment on “Statistical issues about solar-climate relations” by P. Yiou et al.***

**Anonymous Referee #2**

Received and published: 10 May 2010

The manuscript is basically a rebuttal of a previous papers on the solar influence on European temperature by Le Mouel et al. and Courtillot et al. The authors of the present manuscript try to estimate the significance of the correlation between temperature and solar activity under a more realistic null-hypothesis, encompassing the autocorrelation structure of the temperature series and the geomagnetic series, used as a proxy for solar activity. The main conclusion is that the statistical significance of these relationships under this null hypothesis is very low.

Although I think that the contents of the manuscript are correct, I found it difficult to read. The structure is confusing and the language very imprecise in critical passages. The English formulation does not contribute to bring the message across clearly.

In my view, one critical problem in the manuscript is that it is aimed to rebute a former analysis without saying it explicitly. The introduction thus appears unfocused, criticizing

C167

in general statistical studies that analyze the solar-temperature relationships. Only after having read half of the manuscript would a reader realize that the criticism is directed to a particular methodology exposed in a particular paper. I think the manuscript would be much more clear if it would state this from the very beginning, then summarize the methods employed by Le Mouel and Courtillot, state their conclusions and the set off to unveil their flaws or deficiencies, essentially the lack of a proper analysis of statistical significance.

The abstract is too general and non- informative: 'The goal of the paper is to provide a framework to control the spurious results that statistical tools can generate'. Do the authors refer to \*all\* statistical methods ? do they refer to \*all\* statistical studies about the solar-temperature relationship? I do not think so.

The introduction is devoted initially to a short review of the physically based studies of the solar-climate relationship. This review is quite biased, citing only papers that tend to be critical to this relationship. As this manuscript is mainly concerned with statistical issues, I do not see the need for this part of the introduction, unless the authors would like to go into deeper detail, which they explicitly do not wish to. After briefly mentioning the papers by Le Mouel et al., the introduction turns too general again, and states that the goal of the manuscript is 'to explore the statistical consequences of transformation of data'. But the manuscript only deals with one particular data filtering, that apparently applied by Le Mouel et al. The expression 'transformation of data is in my view also too unspecific. The Le Mouel et al. study was previously unknown to me, and I think any reader in my situation would find this introduction quite confusing.

The method section starts with a fairly basic description of autoregressive processes, to settle down later on the most simplest of those as a null-hypothesis. From this, several data filters are derived (the mean interannual squared variation and the mean squared daily variation). It turns out that some functional forms of the expectation of these filters provide an estimate of the lag-1 autocorrelation of a series. The reader would try to scour the manuscript for the reason of this detour. Couldnt one just estimate the lag-

C168

1 autocorrelation from the standard Yule equations or least-square estimators (in this case, just the correlation at lag 1? ) Is this done because Le Mouel et al. also applied these filters ? Is this done because these estimators are better by some unspecified measure ?

Data: why are daily local data of the magnetic field used as a proxy for solar activity? There are several published and assessed reconstructions of solar activity around. What is the reason to use local data ? solar activity certainly does not depend on the location of the Earth where the magnetic field is measured? is it necessary to use daily data of the magnetic field? Few people would believe that that daily variations in solar activity can influence local daily temperature in a noticeable manner. Is this again because the UK series is deemed more reliable or because Le Mouel et al used it ?

The choice of data confirms that the purpose of the manuscript is not to explore the sun-temperature relationships per se -this choice of data is certainly not optimal-, but to rebut the Le Mouel et al. study. I have therefore some concerns that Climate of the Past is the right framework for this.

The manuscript goes on to apply the quite entangled estimators for the lag-1 autocorrelation of a time series to the long temperature and magnetic field observation series, obtain values for the lag-1 autocorrelation, and generate surrogate series are generated with the same lag-1 autocorrelation. Then the same filters are applied to the surrogate series and a distribution of the correlations between the filtered series under the AR-1 null hypothesis is derived. The reader would think if this long detour is really necessary. For which reasons in particular are these filters applied ? What information do these filters provide ? why are they interesting ? The justification is that they seem to be the used by Le Mouel et al. If this is the case, it should be stated up front, otherwise the reader will be confused.

I found the last section 4.4 particularly unclear. This section deals with the general of problem of causality between two time series. According to the manuscript, a possible

C169

proof of casualty of series x on series y is if the lag-1 autocorrelation of y, which may slowly evolve in time, depends on the values of x. Why is this really convoluted model of causality being addressed ? The reader will remain baffled until some paragraphs later he is informed that this is what Le Mouel et al. did. Again, I found this very confusing.

In the following I list the passages that remained unclear to me:

Page 464 line 7: 'we computed the daily mean' - > spatial average for each day

line 17: this problem could be circumvented by using centered data' ?

line 25: 'most of the stations are tagged as suspect in the EC&D data base, but the authors would deal with this problem in a forthcoming manuscript'. I found this very striking. Shouldn't first the quality of the data be checked, previously to any analysis, specially if they are tagged as suspicious ?

Page 467 line 4 : 'mathematical expectation' → expected value

line 5: ' for a sufficiently large Zeta'. I think the equation after this line is valid for any value of the window width.

Equation 5. Again I think the equation is valid for any window width, provided that the correlation at lag 365 is neglected, which I think is what the authors are doing

Page 468 line 22, 'the probability of failure when one rejects the null hypothesis' → the risk of wrongly rejecting the null hypothesis.

Page 469

line 18 : 'N=300000 increments' → time steps

Page 470

line 21: 'hence the large variations of Q(t) are meaningful for temperature and could be interpreted as such'? Imprecise language. what does 'meaningful' mean ?

line 24: ' their significance with respect to a AR1 process has the same feature as Q'.

C170

Imprecise language. what does feature mean here ?

Page 475 line17: 'the analysis and interpretation of L(t) are potentially irrelevant' ? I cannot understand this sentence.

In general, I found the language imprecise. I would recommend to avoid buzz words such as 'feature', 'characteristic' and words of the same kind. Last but not least, as a non-native speaker I think the manuscript would benefit from copy-editing by a native speaker.

---

Interactive comment on Clim. Past Discuss., 6, 461, 2010.

C171