

Clim. Past Discuss., author comment AC1
<https://doi.org/10.5194/cp-2021-52-AC1>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.



Reply on RC1

Martin Skoglund

Author comment on "Climate variability and grain production in Scania, c. 1702–1911" by
Martin Karl Skoglund, Clim. Past Discuss., <https://doi.org/10.5194/cp-2021-52-AC1>, 2021

In the following, I have divided the reviewer's comments into different parts, answering each part on a running basis with references shown after each answer when relevant.

Reviewers comment part 1:

"This study analyses the relationship between climate variability and grain production in southern Sweden for the 18th and 19th centuries. In the long introductory part, the author provides detailed information about the farming system of the study area, historical background, natural conditions, grain crops, and their varieties. He stressed the importance of crop diversity for stable crop production. It is interesting that reversed relationship of the crop production to temperature compared to other parts of Scandinavia, as well as other parts of Europe, was found in this study.

Historical database of the tithe records is used for the period before 1865 while official statistics on county level were utilized after that year. Data on grain production from the historical database were pre-processed in several steps attempting to solve some biases and sources of uncertainty (normalizing, de-trending). However, not all pre-processing steps are sufficiently explained. For instance, what is the role of the "threshing coefficient". While the reasons for aggregating data are well explained, the application of the cluster analysis (Section 2.2) is rather strange. How it was decided that exactly three clusters are optimal? The final number of clusters was decided subjectively or any objective measure was used? While individual villages are clustered, the map in Figure 3 presents administrative units belonging to different clusters. It would be useful to explain more clearly, whether the BISOS data from the 1865–1911 period were standardized in a similar way as the HDSA data."

Answer:

Historically, the amount of grain obtained after threshing varied in different localities across Scania (in fact this is a common phenomenon in historical agriculture in general). The tithe in Scania was collected *before* threshing. Olsson & Svensson (2017a, referenced in the article), have thus constructed threshing coefficients for different parishes and type of farming districts in Scania, allowing a more accurate conversion from *harvested volumes* to *threshed volumes*, the latter of course being a more accurate measure of the final output of grain production. These threshing coefficients are based on actual threshing accounts found in various sources, e.g. manorial archives, church archives and probate

inventories. The threshing coefficients can be found at <https://www.ekh.lu.se/en/research/economic-history-data/HDSA-1702-1881> in the "STATA do-file.txt" (last accessed 2021-06-16). In the article, the threshing coefficients thus influences the relative production levels in the analyzed villages for the period 1702-1865, which for instance can be seen in the average grain production for each cluster shown in Fig. 5.

Regarding the clusters, it is correct that the hierarchical cluster analysis (HCA) sorts individual villages. The administrative units shown in Fig. 3 are parishes, which in general are constituted by a number of villages. In some parishes, different villages were sorted into different clusters in the HCA, leading to some parishes being represented by multiple clusters. This is not an issue, since in this example, the parishes are merely employed as a descriptive tool (like the type of farming district and the institutional status of villages) to describe the more abstract (from an historical point of view) clusters. This is discussed in the article (lines 326-329). Finally, regarding the number of clusters in the HCA. The HCA is often referred to as an explorative approach, given that the number of "true" clusters is not known. Even though there have been multiple methods proposed for optimizing the number of clusters, it remains difficult to know *a priori* what the best metric to use is. I used one of the more common metrics, e.g. the *gap statistic* (Tibshirani et al. 2001). Specifically, I used the `clusGap` function in the `factoextra` package in *r* setting a maximum number of potential clusters at 10, yielding the optimal number of three clusters after 100 bootstrap samples (Kassambra & Mundt, 2020). Following the logic of trying to reduce the interpretative and descriptive issues related to HCA (discussed in lines 326-329), this number can be compared to the most common categorization employed in the historical literature when discussing relative specialization in production, i.e. the type of farming district. In Scania three different types of farming districts have been conceptualized, namely the forest districts (sw. *skogsbygd*), the mixed districts (sw. *risbygd*) and the plain districts (sw. *slättbygd*) (see references in the article, e.g. Campbell, 1928, Dahl, 1989 or Bohman, 2010). As can be seen in Fig. 4 there is some relationship between the clustering results, although it is quite limited.

References:

- Bohman, M.: *Bonden, bygden och bördigheten: Produktionsmönster och utvecklingsvägar under jordbruksomvandlingen i Skåne ca 1700-1870*. Diss. Lund University: Lund University, oai: DiVA.org:umu-99296, 2010.
- Campbell, Å.: *Skånska bygder under förra hälften av 1700-talet : etnografisk studie över den skånska allmogens äldre odlingar, hägnader och byggnader*. Uppsala: Lundequistiska bokhandeln, 1928.
- Dahl, S.: *Studier i äldre skånska odlingsystem*. Stockholm: Stockholms universitet, 1989.
- Kassambra, A. & Mundt, F.: *Package 'factoextra'*. <https://rpkgs.datanovia.com/factoextra/index.html> last accessed 23 June 2021. R package version 3.1.2, 2020.
- Tibshirani, R., Walther, G. & Hastie, T.: *Estimating the number of clusters in a data set via the gap statistic*. J. R. Statist. Soc. B, 2001.

Reviewers comment part 2:

The manuscript is not well structured. While the introductory part is rather long, with numerous details, methods are mixed with results in Sections 2 and 3. Section 2.4 is

followed by Section 3.1. Some section titles do not correspond to the following paragraphs (e.g. Section 1.4.1). Results of correlation analysis (Section 3.1) are mixed with discussion (e.g. lines 496–504). Results are presented in the form of several correlation matrices (Figs. 9–13). Correlation coefficients between crop production series and temperature/precipitation/drought characteristics were calculated for two different periods, for different clusters and different crops. And these correlations are repeated for “dry” and “wet” years. The description and interpretation of such results are rather long and not very synoptic. It is very hard to orient in the text and to find any signal in presented correlations. I miss any information which correlations are statistically significant (those in colored boxes?) and on which level (p-values)?

Answer:

The motivation or idea for the initial structure of the paper was that the hierarchical cluster analysis and other data homogenization procedures was part of how the data was managed, while the results were more delimited to the statistical correlations between grain production indicators and climate indicators. However, given that parts of the method-section contains actual results this initial structure could be improved in line with the commentators comments by more clearly separating methods and results. I agree with the commentators note that the lines 496-504 more appropriately belongs to the discussion section.

The commentator points out some other errors in the structure of the pre-print. For example Section 2.4 should be followed by a heading titled Section 3 and not Section 3.1. An important error of omission in Figs. 9-13 pointed out by the commentator is that it is not specified what correlations are significant. In the notes to each figure it should be clearly stipulated that only significant correlations are colored ($p \leq 0.05$). Adding this will also clarify the signals discussed in the text. The ‘signals’ corresponding to the patterning of colorations in Figs. 9-13.

Reviewers comment part 3:

In spite of a relatively long and detailed introductory part, I miss any direct information on harvest dates (or threshing dates?). This date may indicate the time when the grain production of the given year was determined. Harvest dates are mentioned only indirectly in August and September (lines 199 – 201). In this sense, a correlation of a grain harvest with Oct, Nov, and Dec climate of a given year seems to have no meaning. Contrary, it would be more meaningful to correlate the grain harvest of a given year with Dec, Nov, and Oct climate of a previous year. It would be useful especially for crops sown in autumn.

Answer:

Regarding the months of October, November and December, it is described in lines 453-454 that I use lagged (i.e. previous year) values. In the previous lines, 452-453, I briefly mention that harvesting was usually *completed* in late August or during September. However, I would agree with the commentator that the time of harvesting (and sowing) are quite important and some elaboration is probably in order.

There is generally a lack of sources on harvesting (and threshing dates) for Scania before the 20th century. Even in the official statistics, BiSOS 1865-1911, there is a lack of harvest dates (even though reporting on harvest and sowing dates were part of the forms that were sent out for the collection of agricultural statistics). Most likely, there was a relatively large variation across parishes and villages when actual harvesting was begun. However, in farmers’ diaries and some parish descriptions from the 19th century one can find some specific as well as general notations on harvest, sowing and threshing dates. For example, in the parish description of Husie socken in southwestern Malmöhus from 1826, the parish

priest Carl Carlsson noted that the climate of Husie *village* (the village and parish had the same names) was mild and relatively humid and undoubtedly beneficial for arable farming. Sowing was usually started between the middle of April until the middle of May (Andersson, 1986). In a local farmers diary from the same village, the farmer Anders Andersson on the farm Skrävlinge nr 1 wrote down sowing and harvesting dates during a some years in the early 1820s (Andersson, 1986). In 1821, sowing of barley and mixed-grains (barley and oats) started on the 18th and 15th of May, respectively, and lasted roughly a week (ended on the 22th of May). In 1823, peas, oats and vetches were first sown on the 14th of April and barley and mixed-grains on the 15th of May (finished 27th of May). In 1826 barley was sown starting on the 12th of May. In most of these years, rye sowing started in the beginning of October. Regarding harvests, they were begun on the 21st of August in 1821. In 1826, due to excessive heat and drought, the harvest (of at least rye) was begun on the 25th of July.

In a collection of parish descriptions from Malmöhus Län in 1828 there are short summaries of general sowing and harvesting dates for a little more than a dozen parishes. According to these descriptions, sowing of oats started between the middle of April until late May, depending on the parish and current weather conditions. Barley sowing was initiated from late April until early June, albeit mostly in the 2nd half of May. Sowing of autumn-rye began somewhere between the middle of August until early September, and in some instances in late October, however mostly in the latter part of September. Harvesting generally started during August or sometimes in early September, or in late July for the autumn-rye crop (Bringeus, 2013).

There appears to be some differences between the forest and plains district as well as parishes in different latitudes. However the number of available sources presented here are too few to say anything with certainty in this regard. In relation to the discussion in the article of the potential similarity of barley landraces in southern Sweden and more northerly parts of Scandinavia, it is interesting to note that sowing and harvesting dates of barley in the northerly province of Jämtland are quite similar to those available from Scania, i.e. sowing in the middle of May and harvesting in August. This could possibly be inserted into the discussion. Furthermore, in line with the commentators points, I could add general sowing and harvesting dates for Scania (to the extent they are available), as well as shorten the introduction.

References:

Andersson, H.: Huse- Lantsoken som blev stadsdel. Några anteckningar kring en sockenbeskrivning från 1828. Elbogen 16 (1-4), 1-24, 1986.

Bringeus, N. A.: Sockenbeskrivningar från Malmöhus Län 1828. Arcus Förlag, 2013.

Reviewers comment part 4:

While the analysis of the relationship between grain production and climate is based on simple correlation analysis, a relatively extensive discussion mentions a number of aspects that were not analyzed in this study. For instance, lines 584-585: "... the absence of a climate signal in the spring and autumn months, as well as the last summer month of August to some extent. This could be interpreted as spring or autumn frosts not being a systematic threat ..." However, correlation does not mean causality and it would be correct to add some info about the frequency of spring/autumn frosts.

Similarly, lines 630-631: "If conditions were relatively wet or dry in the early summer, the effects from subsequent precipitation and temperatures later in June and especially July would theoretically have been amplified." Such claims should be supported by their own results and / or citations from other similarly focused studies.

Answer:

The formulation referred to on lines 584-585 could be amended to further clarify that it is merely correlations and not causal relationships being discussed. However, I would emphasize that what I write on the specified lines is that it *could be interpreted*, i.e. it is merely one possible interpretation of the statistical associations. Furthermore, I agree that it would be correct to add information in the article on the frequency of spring/autumn frosts. This can be done using data from the second half of the 19th century when daily temperature data is available.

Regarding the comment on lines 630-631, it is a theoretical stipulation but also partly based on the results in the article (see Fig. 12 and Fig. 13) and the literature (see for example Brunt, 2004, on the benefit of rain spread out across the growing season). It is theoretical, or hypothetical in the sense that it is based on my speculations on the possibility that there is a skew towards early summer conditions in the SPEI-reconstruction used in the article. Perhaps an amendment is in order to clarify the hypothetical nature of what is being discussed on the specified lines?

Reviewers specific comments:

- Map of the study area with outlined geography mentioned in the text would be very useful

Answer: This could possibly be added, however it is a relatively comprehensive task given that there exists to good geographical data relevant to 18th and 19th century land use. For example, there was large changes (reductions) in forested land and (increases) in arable land.

- Section 2.1 refers to clusters. However, clustering is explained later in Section 2.2.

Answer: Should be amended.

- Homogenization of the Lund temperature measurements from 1753 is an important by-product of this study. However, this series was also extended further back to 1701 and there is no information about the validation of this earliest part of the "calculated" series. Looking at Figs. 6 and 7, the calculated series (before 1753) seems to have lower variability compared to part of measured temperatures (after 1748) for all seasons. Was the variability of the calculated series adjusted in any way?

Answer: The larger variability seen in Figs. 6 and 7 from 1753 can only be seen in the graphs representing the raw series; looking at only the graphs for the homogenized series no such discernible shift can be seen.

- It would be useful to add some information to Table 2 about the length (N) of the period that was used to calculate correlations between the Lund series and the other temperature series. Are there all correlations statistically significant? At that level?

Answer: Common periods were used to calculate correlations, which are all significant at the ≤ 0.05 level. The spatial correlation coefficient in this context is mainly important for the purposes of the ACMANT procedure, see answer below.

- The problem of homogenization of the early instrumental temperature measurements closely relates to the so called “the early instrumental warm-bias” (see e.g. Böhm al., 2010). It would be useful at least to comment on it. While in the Greater Alpine Region warm bias was found especially in summer, this study found “cold” bias for Lund.

Answer: Such a comment could be added.

- Table 2 – It is not clear, how the spatial correlation was calculated and why it is listed on the last row of the table. One would expect that it is a correlation between Lund and neighboring stations (or some spatial temperature field?). In this sense, it would be listed as a separate column, not a row.

Answer: The notes for Table 2 shortly describes the method for how the spatial correlation is calculated in ACMANT. Only one spatial correlation coefficient is obtained per station, which is why it is added as a row and not a column in Table 2. The spatial correlation is important because it shows whether a station series is a valid input into the ACMANT homogenization procedure. It is described in more detail in the AMANT manual and ACMANT scientific description referred to in Domonkos & Coll (2017).

- From correlation analysis it follows that explained common variance (r-squared) is mostly negligible. For instance, when the correlation coefficient $r=0.3$, grain harvest and climate share less than 10% ($r^2=0.09$) of the common variance. It would be useful to explain more (or even quantify) the role of other factors. Some of them are mentioned in the discussion.

Answer:

It is not unproblematic to jump from a correlation coefficient obtained from a correlation analysis to an r^2 without a properly specified regression model (which carries with it some different assumptions). This article is mainly concerned with estimating the extent of relationship between climate variability and grain production. It is not, as the commentator correctly observed in a previous comment, to establish specific causalities between climate and grain production. Whether a correlation coefficient of 0.3 in this context is mostly negligible or not is difficult to determine, however one can subjectively argue one way or the other. If a correlation coefficient of similar magnitudes is obtained across the same months, in different samples, data-sets, time periods as well as different crops, I would argue that the relationship was probably not *historically* negligible. I would further emphasize that in the context studied here, the margins in production were important. There is also the issue discussed by Beillouin et al (2020), and Edvinsson et al (2009) noted on lines 236-237 in the article, that a lack of detailed climate data usually leads to an underestimation of the impacts of climate variability on grain production. For example, in this article, I do not have access to daily temperature and precipitation data in the pre-1860 period, nor do I have minimum and maximum values at a sub-annual level.

Regarding the direct inclusion of other factors, this is a task that is beyond the scope of this article and would require a study of its own, probably at a much lower level of analysis (e.g. looking at only a smaller subset of villages). One of the more important factors in this context is probably soil. To obtain more detailed soil data would require extensive efforts to rectify soil maps with historical land use maps, and even then it would remain unclear which crop was grown at a particular plot of land in a given year. Nor is historical land use maps available for all villages in the sample, and for those villages that there are available maps, they cover only a few select years. Other possible and largely time-invariant factors include institutional set-up, farming system and other geographical factors like type of farming district. However, besides soil types, there is a theoretical gap

in the literature as to why and how such other factors would affect the relationship between climate variability and grain production. This being said, most of these factors are actually indirectly included in the analysis, see the description of the different clusters in Fig. 3 and 4, Table 1 and lines 333-383.

References:

Beillouin, D., Schauburger, B., Bastos, A., Ciais, P. and Makowski, D.: Impact of extreme weather conditions on European crop production in 2018. *Phil. Trans. R. Soc. B* 375:20190510, doi: 10.1098/rstb.2019.0510, 2020.

Edvinsson, R., Leijonhufvud, L. and Söderberg, J.: Väder, skördar och priser i Sverige. In: Liljewall, B., Flygare, I., Lange, U., Ljunggren, L. and Söderberg, J. (eds) *Agrarhistoria på många sätt. 28 studier om människan och jorden*. Kungl. skogs- och lantbruksakademien, 2009.

Reviewers' technical corrections:

- Lines 25-25: reference to "Huhtamaa & Helama, 2017b" is mentioned twice in the list

Answer: Should be fixed.

- Line 34-35: correct to (Osvald, 1959; Persson, 2015).

Answer: Should be fixed.

- Line 78: correct to: ... century.

Answer: Should be fixed.

- Line 97: "...the suly of winter fodder". Please check. Is it correct?

Answer: Should be corrected to 'supply'.

- Line 109: „starting in 1749/1757" – this is not clear

Answer. Should be clarified to 1757.

- Footnote 4: „The share of oats WAS quite low"

Answer: Should be fixed.

- Line 289: please unify: BiSOS or BISOS?

Answer: Should be unified to BiSOS.

- Line 302. Normalized production anomalies is abbreviated asNPAa, but differently in formula (1) and in the text (line 306)

Answer: Should change normalized production anomalies on line 306 to NPA.

- Line 347: The data ... has subsequently incorporated ... Please check

Answer: Appears to be in order to the author.

- Line 358: "four clusters" – Should not be "three"?

Answer: Correct, should be three.

- Line 382: "...increase in an ascending order ..."

Answer: Appears to be correct to the author.

- Line 447-8: What is the meaning of „simple“ climate variables?

Answer: "Simple" climatic variables in this context refers to the mean values at seasonal, monthly or annual levels of temperature, or the sum of precipitation at the same time-scales. An example of adding complexity would be to convert these values into deviations from a long-term or moving average. Please see the discussion and references on lines 443-448.

- Line 459: What is the meaning of "...most consistent coefficients"?

Answer: That the direction, magnitude and significance of coefficients are consistent across samples, time-periods and different crops. Could be changed to for example: "producing the clearest signal".

- Please check figure and table captions, correct and complete. For instance, there are no „Descriptive statistics“ in Table 2, Figure 5 – "...estimated loess" is not clear. The loess function is used here as a low-pass filter. Figure 7 – correct the caption – figure relates to DJF and SON seasons. Figures 9 – 13 – please add information about the statistical significance of the correlation coefficients.

Answer: Descriptive statistics could be added to Table 2. The extent of the series are described on lines 392-397. Fig. 5 shows the estimated loess (smoothed lines) and the NPAs (point-lines) for each cluster and crop (see legend). Legend could be adjusted to e.g. "barley NPA" and "barley loess"? Fig. 7 should be fixed. Fig. 9-13 should be fixed (see above).