Review of article:

Does climate change affect period, available field time and requiredcapacities for grain harvesting in Brandenburg, Germany?

Annette Prochnow, Hilke Risius, Thomas Hoffmann, Frank-M. Chmielewski

METHODOLOGY

Are the tools and methods appropriately applied, and are the data correct?

No

The data seems to be correct, but the methods are not appropriately applied.
I have already discussed the most serious error in the question “Is it allowed to test the significance of linear trends using a t-test if the considered data are moving averages?” (see https://www.researchgate.net/post/Is_it_allowed_to_test_the_significance_of_linear_trends_using_a_t-test_if_the_considered_data_are_moving_averages). I will repeat my argumentation here in a slightly extended form: The authors calculated (simple) moving averages of the considered quantities y(t) („number of hours per year available for harvesting grain with defined moisture content“; coefficient of variation; total sunshine duration; etc.) for time periods of 30 years in steps of one year (i.e., 1961-1990, 1962-1991, …, 1984-2013) first. After that, they derived the trend of these averaged values and used these values to estimate the significance with the help of the „t-test for the regression slope“ (see their section 2.4 and Table 5 and Table 6). This way, most of these trends are proofed to be significant with $p<0.01$.
This method is not allowed because the precondition for the t-test is that the data entering the test are independent (i.e., uncorrelated, because they are also supposed to be normal distributed). The moving averages are highly correlated. This leads to a drastic reduction of the standard deviation of the originally y(t): The standard deviation is reduced by a factor of about 1/15 (see the second figure attached to my question). If one computes the significance of a trend of these averaged values one attains a highly significant result in many cases thanks to the small standard deviation. This can be seen in my first figure which is attached to my question: The upper graph shows a regression with the original data. For this, I generated 53 synthetically values with a normal random generator “gnoise(20)” with a standard deviation of 20: $y(t) = -0.2\ t + gnoise(20)$. Here the regression slope b is not significant different from zero at a 95 %-level. But in the lower graph, where I first averaged the data and then applied the regression, we have a highly significant trend.
How strong the reduction of the standard deviation sdev of the data in the article is, can be seen if one looks at Figure 3 (in the article): As can be seen easily, the coefficient of variation (=sdev/mean) is very small (a few percent) for these moving

averaged data. But in Table 5 the listed values for the coefficient of variation are between 9 % and 63 %. Please note that in this table the trend for the coefficient of variation is computed with the “values of overlapping 30-year time periods” again, and therefore the significance level for the trends is wrong also for these coefficients of variation (because the year to year values are strongly autocorrelated).

To proof that the trends in Table 5 are not significant at all, we can apply the independent two-sample t-test (http://en.wikipedia.org/wiki/Student%27s_t-test#Independent_two-sample_t-test) to the “80%-values” (which are 30-year means according to Equation (1)) for 1961-1990 and 1984-2013) (neglecting the overlap of 7 years) which are listed in Table 5. First we must determine the standard deviation s (resp., sigma). For this, Equation (1) can be used. If we choose the values for winter wheat and grain moisture content ≤ 18 %, one has P(0.2)=140 h/a (from Table 5), mean(x) = 5200 h/(30 a) = 173.3 h/a (from Figure 3) and z(0.2)=0.84 (see paragraph below). With these values one gets sigma=39.7 from Equation (1). Now we can compute t = (mean(y2)-mean(y1))/(sqrt(2)*sigma/sqrt(30)) = (149 – 140)/10.25 = 0.88. The corresponding p-value for this t is 38%. As a result, the difference in the 80%-grain-moisture-content between 1961-1990 and 1984-2013 is not significant!

If one defines the 20%-quantile z(0.2) of a standard normal distribution as a negative value (as one usually does) one has to exchange the minus sign in Eq. (1) by a plus sign.

Another very critical point is the following: The authors used very sophisticated regression functions (see their Table 1) for estimating the „hours within classes for grain moisture contents“. These regressions show coefficients of determination r^2 with values up to 0.99 (the given references are not downloadable; see paragraph “References” below). I assume that these regressions are derived by stepwise multiple regression (which is very dangerous method if one does not perform a subsequent external validation). But the authors did not mention such a procedure. Therefore I assume that these regressions are overfitted and could not withstand an external validation.

Another criticism concerns Table 3 and 4: E.g., in Table 3 we have multiple tests [in total 18 tests (6 seasons and 3 quantities)] for the regression slope (trends). The results are counted as significant if p≤0.05. The probability to get one or more significant results, even if there is no real trend at all, is 60 % (if the tests are considered to be statistically independent; this can be easily derived with the help of the Binomial distribution, n=18, p=0.05), and the probability to get two or more significant results is 23 %. That means that it is not unlikely to achieve a result similar to the one shown in Table 3 (with 2 significant trends). Without any correction for multiplicity (or drastically reduced critical p-values) tables like Table 3 and 4 are worthless.

Please notice, that all trends of the sunshine duration Sd (even in the harvesting months 07-08) are not significant in Table 4 (using original data), but highly significant in Table 6 (in this table moving averaged values were used again!).

The results of Figure 4 (section 3.5) are presented without any probability information. The authors only emphasize that a few bars are twice as long as the average length. If one computes the probability Prob(X≥3) (again with the help of the Binomial distribution; p = 10/53 = 0.1887; n = 11) that a bar in Figure 4 has the length 3 (with unit “number of years”) or more, one gets 34.5 % if one assumes that there is no real trend at all. And the probability that a bar has length 4 or more is

13.7 %. This proofs that the results of Figure 4 are not unlikely and are no indicator for a change in available harvesting hours!

ANALYSES

Have the analyses, including any statistical analyses, been performed appropriately and rigorously?

No

see above (METHODOLOGY)!

REFERENCES

Do the authors refer to the latest research in the field, and is it correctly referenced?

Yes

The references are correct. But the two articles which describe the primary method used for computation of the grain moisture content are very difficult to obtain or not available ("Hoffmann, T., (Doctoral dissertation) 1998" is not available in electronic form and „Hoffmann, T., Fürll, C., 2000" (a proceeding for a conference) is not available any longer).

FINDINGS

Do the authors provide an interesting discussion, and are the findings novel? Partly

The idea to derive available field time and required capacities for grain harvesting as a function of Climate Change is very interesting and innovative. Unfortunately, the authors applied statistical tests to data which do not fulfill the preconditions for the tests. Hence, they achieved results which are only apparently significant. Actually these results are wrong and could be emerged by pure accident. Therefore the whole discussion is senseless because the discussed trends could change to the contrary if one would examine 53 independent years (in the same climate) (what is unfortunately not possible).
The only reliable and significant results of this article are shown in Table 2 and Figure 1 and describe the air temperature at the station Potsdam from 1961-2013 which is examined extensively in a bunch of other papers. The question asked in the title was not answered in this article because all the relevant trends are not significant actually.

CONCLUSIONS

Are the interpretations and conclusions justified by the results?

No

See FINDINGS!