Reply to the RC of the anonymous referee
Oliver Wetter and Christian Pfister

Dear anonymous referee

Thank you for your many thought-provoking impulses that allowed a significant improvement of our paper!

GENERAL COMMENT:

1. RC: This paper relies on a new Swiss series of reconstructed summer temperature. As such, it is interesting.

Reply: Thank you for stating our paper to be interesting.

2. RC: However, I think it does not meet the standard of Climate of the Past in its present form. The description of the data is really insufficient. The methodology is very poorly explained. The authors indeed do not explain neither how they treat the data (the homogeneisation procedure which is a very important stage is not describe at all, and the description of the dataset is very mean), nor how they do transform them into temperature.

Reply: We agree that data description is not sufficient and we therefore will provide the evidence needed to replicate the results (all data will be given in a few days). The procedure was mentioned on page 2705, line 20-22, albeit too shortly: The methodology of our paper agrees with the standard palaeo-climatological procedure of calibration and verification (Cook et al. 1994, Dobrovolny et al. 2010), The alternative methodological survey by Garcia de Cortazar et al. (2010) favoured by the research group of Dr. Pascal Yiou focuses upon the process-based approach developed by Chuine et al. (2004). This model was constructed using 20th century instrumental and phenological data which are not contained in the supplementary data on internet and therefore not available for replication. The standard paleo-climatological approach of calibration and verification is not addressed in this paper. The authors just mention "calibration" omitting verification though this approach allows testing the suitability of the model on the basis of historical proxy data. Garcia de Cortazar et al. (2010) admit that "estimating an average error connected to socio-political factors is *a priori* impossible to achieve in a regular statistical framework, because those phenomena never follow well-defined biological and probability laws." However, record breaking extreme events such as 1540 and 2003 just lead to situations which plants and societies are just outside well defined biological and probability laws. In particular, human agency matters in such situations.

RC 3: The individual series should be provided in an appendix.

Reply: All series will be provided in the supplementary material.

RC 4: While the authors in the introduction remind why GHD may not be reliable data regarding climate reconstruction, they use GHD and other related data to reconstruct temperature. After having raised so many reservations it is surprising that they apply

the same method they have denigrated without discussing much the uncertainties on their reconstruction, nor the possible errors or limitations.

Reply: Nowhere in our introduction is stated that GHD "may not be reliable data regarding climate reconstruction" and we reject the claim that we denigrated the GHD approach. In order to avoid this impression we will insert a paragraph providing a short review acknowledging the pioneering role of Emmanuel Le Roy Ladurie and the recent improvement of GHD analysis with reference to the paper by Daux et al. (2011) and Labbé and Gaveau (2011).
On the other hand, we maintain our view that time series analysis of proxy data **per se** is not sufficient to adequately assess temperatures in record breaking warm and dry years such as 1540 and 2003. Our point of view is also confirmed by the short comment posted by Ivo Maran. As is well known, record breaking extreme values are outliers that cannot be dealt with just by using standard statistical methodologies. In the tree-ring based study by Battipaglia et al. (2010) it was demonstrated that the assessment of pre-instrumental extreme values from natural archives needs to be supported, if available, by narrative documentary evidence. A combination of both, time series analysis and study of documentary evidence leads to significant improvements. In the case of GHD, which are man-made proxies the context also needs to be analysed under which the evidence was generated. This issue was stressed by Garnier and co-authors (2010) being mostly members of the team led by the mathematician Pascal Yiou. Critical analysis of the data prior to model building is essential (see also Labbé and Gaveau 2011). The same care is required in the case of data from natural archives, such as for example tree-rings where insect attacks and other pathogens need to be taken into account (e.g. Schweingruber 1996). We could demonstrate that the decision to begin the vintages in 1540 hinged on precipitation, not temperatures. An extended discussion of the Chuine et al. (2004) paper is contained in the reply to Dr. Pascal Yiou.

Uncertainties were discussed in different parts of the current open discussion paper, e.g. on page 2702, line 24-26, page 2705, line 23-25 and on page 2709, line 17-21. Standard Error of Estimate was mentioned on page 2669, line 12-15 and on page 2709, line 17-18. They will be discussed in extenso in the revised paper.

RC5: They discuss extensively the fact that 1540 could have been the hottest summer of the last 500 years. Is it an issue? I think a more interesting approach would be to discuss the variation through time of extreme events frequency and amplitude.

Reply: The referee doubts, whether dealing with record breaking extreme events is an issue. This is somewhat odd considering the enormous impact of such events on economies and societies both past and present and the fact that the Chuine et al. (2004) article itself put an emphasis on the 2003 extreme event. In an additional paper to be submitted we will assess precipitation in 1540 to demonstrate that the extreme temperatures in 1540 were related to a drought that was by far longer lasting and severe than in 2003. We doubt whether a statistical analysis of the extreme events frequency and amplitude in the new Swiss series would yield results substantially different from what is known from the analysis of known GHD series.

SPECIFIC COMMENT:

RC 6: Line 23: "He demonstrated that their model overestimated the measured temperature for 2003 by 2.4 _C, whilst at the same time severely underestimating temperatures in other warm summers such as 1947, 1952 and 1945": The formulation leads one to believe that many examples of underestimations are available whereas Keenan (2007) did not produce any other one but 1947, 1952 and 1945.

Reply: We will omit this statement because it detracts from the main argument.

RC 7: Line 24: "Beniston (2004) based on documentary evidence by Pfister (1984) and Glaser et al. (199) argued that the year 1540 was warmer than 2003" . I think the authors should read more carefully Benston (200')'s work. The only reference to 1540 in Beniston's paper is in the introduction : "Research by Pfister et al. [1999], indeed suggest that 2003 is likely to have been the warmest summer since 1540". It does not mean that 1540 is warmer than 2003. It is a statement in the introduction and not a demonstration. Glaser is not even referred to!

Reply: Beniston reads as follows: "Research by Pfister et al. [1999], [he actually referred to Glaser et al. 1999] ***indeed suggest that 2003 is likely to have been the warmest summer since 1540 [emphasis added by C.P.]***. Climate reconstructions, based on written historical archives, for that particular summer suggest that Europe was under the influence of a strong high pressure system centered on the English Channel that resulted in prolonged drought and heat" (p. L02202, 1). The quote clearly indicates that the arguments brought forward by Glaser et al. (1999) seemed plausible to Beniston. Otherwise he would not have quoted them in extenso. Likewise, Schär et al. (2004) "neither excludes the possibility that such warm summers might have occurred in the more distant historical past, for instance in the Medieval Warm Period, in 1540 or in 1757". On the other hand, the paper by Chuine et al. (2004) does not refer to this possibility though the book by Pfister (1999) was quoted in which the heat and drought was described in extenso. This point of view was confirmed in papers by Glaser et al. (1999) and Jacobeit et al. (1999) which are both also not quoted by Chuine et al. (2004).

Glaser et al. (1999) is referred to on page 2697, line 24-26.

RC 8: Line 28: "The assessment of these authors is supported with the reconstruction of monthly temperature in Central Europe since 1500 (Dobrovolny et al., 2010)". The authors should calculate the T AMJJA with Dobrovolny's data. In the Dubrovolny's paper, the year 1540 is not specifically addressed, and according to their figure 11, it is not obvious than a combination involving AMJJA will be higher in 1540 than in 2003.

Reply: The data were provided on request by lead author Petr Dobrovolny. We found that average April-July temperature anomalies as well as April-August anomalies might have been higher in 1540 than in 2003. August temperatures are not relevant for the date of grape ripening, in particular under the extreme conditions prevailing in 2003 and 1540.(see reply to RC 28).

April-August 1540	April-August 2003 (Temp; Dijon)
3.39 °C (± 0.96 SE)	3.25 °C

April-July 1540	April-July 2003 (Temp; Dijon)
3.54 °C (± 1.02, SE)	2.80 °C

We are sure that Petr Dobrovolny would also provide these data to the referee on request.

RC 9: Line 10 page 2698: "Coherent narrative: : :" What are the references?

Reply: This sentence is part of the outline of the paper where references are not provided. We refer to table 1.

RC 10: Line1 page 2701: "no less than 132 divergences with their corrected version". Labbé and Gaveau (2011) add the: "83 differences are inferior to 5 days, in which case there is no effect on meteorological conclusions". A fair citation would mention this aspect.

Reply: We will address this issue in the revised paper.

RC 11: Line 10 page 2701: Your series is not a GHD only series as it contains 40% of other type data (wages, phenological data). In particular, year 1540 which you concentrate on is not a GHD.

Reply: The term GHD is used in a more general way encompassing all evidence dealing with the beginning of grape harvest. We will address this issue in the revised paper. Our main series is composed from Wage Payment Dates- (WPD), Institutional Grape Harvest Dates- (classic GHD related to the lifting of harvest ban), Historic Phenological Data (HPD) and Phenological Network Observations (PNO). WPD in fact represent the actual beginning of grape harvest. Because day labourers of the hospital of Basel were paid at the end of each working day, the first appropriate entry in in the accounting books represent the beginning of harvest. HPD are phenological observations by individuals about the beginning grape harvest. PNO are standardised observations carried out within the Swiss phenological network. All four kinds of GHD are thus principally equal to each other and may be seen as "GHD" even though the generation of data was different. By the way, the harvest ban was abolished in Dijon in 1836 and re-inroduced in 1898. The intermediate data refer to the date of allowance given to vine-growers to bring their grapes into town. (Labbé and Gaveau 2011).

The date of lifting of the vintage ban is a proxy for grape maturity. In exceptional years this criterion did not apply as the study by Garnier et al.(2010) demonstrates. For the focus of our paper it is essential that grapes in 1540 were (over)ripe much in advance of the lifting of the harvest ban. This act was related to to precipitation rather than to grape maturity. Therefore, the date of full grape maturity in 1540 needed to be reconstructed based on three different approaches: a phenological based- (relation between occurrence of sweet berries and GHD), a vine growing traditional- (relation between first vine most and GHD) and a HPD based (observation of begin of grape harvest by individuals) approach. Each of these approaches tend to grape harvest around mid August. HPD are equal to GHD and confirm the result of the two other approaches. Additional considerations are provided in the paper.

RC 12: Line 17 page 2701: How do you explain that the two curves are not consistent prior to 1730-1740 (precise date unreadable on this figure).

Reply: First it has to be mentioned that these are 11-yr moving average curves. The not corrected GHD by Meier et al. (2007) thus play a role for the next 5 years after 1700.

Meier et al. (2007), on the other hand, used a somewhat different set of data. We will shortly address this issue in the final version of the paper.

RC 13: Line 28 page 2701: How scattered are the 80 villages? Are they possibly distributed at different altitudes? Do the harvest begin in all the villages at the same time? Is the name of the village where the labourer was employed mentioned in the books of expenditures? Is the final wage series an amalgam of the 80 villages?

Reply: The hospital of Basel only hired day labourer**s** to cultivate and harvest their own vineyards. Three different vineyards cultivated by the hospital are mentioned in the books of expenditures in terms of Wage Payment Dates (WPD). The location of these vineyards could be reconstructed. One was in Biel-Benken (352 m asl.), one called "Spittelschüre" in Basel (275 m asl.) and one in Fischingen (313 m asl.). Each vineyard had a southern exposure and was located in a perimeter of 10 km around Basel. Weather and climate in this area are thus more or less the same. As it is not always clear from the accounting books which vineyard was harvested first, we decided to average the three altitudes (= 313 m asl.) and then corrected the Basel hospital GHD series to the average altitude of the whole data set (WPD, GHD, HPD, PNO) i.e. 436 m asl. by linear regression.

RC 14: Line 19 page 2702: Give the relevant information about the series: How long are they? What are the oldest/youngest dates? Where are they precisely from?

Reply: In Fig. 2, page 2724 all relevant information about the 17 GHD series are given in the box.

RC 15: Line 27 page 2702: What are these phenological data? If you have no metadata accompanying them, how do you take into account altitude effects? Describe the series please: number of data, length of the records, starting-ending date, location, variety and phenological stage..

Reply: see reply to RC 14 above. Except variety and phenological stage all information are given in Fig. 2 (see box), page 2724. With regard to metadata, our formulation is indeed misleading. As we know the name of the related community, the mean altitudes were derived from present-day vineyards still being situated in the same favourable locations as centuries ago.
Varieties of Pinot Noir were grown both in the German speaking and the French speaking part of Switzerland known under different local denominations (Schlegel 1973) that cannot be resolved any more, today. Early Red Burgundy grapes grown in the Basel region and in Alsace being ripe about 17 days earlier than other Pinot Varieties are an exception (see page….). Other information about grape variety is only available in the 20th Century, as in most known GHD series. This is, of course a source of uncertainty we will address.

RC 16: Line 9: For a discussion on grape maturity determination you should also refer to Garcia de Cortazar et al., 2010.

Reply: As this recent paper provides a review of the literature including a discussion of methodologies we will include it in the final version of the paper.

RC 17: Line 14 page 2703: "Instead, the time of full grape-maturity was estimated from HPD as a substitute, which is also more adequate for comparisons with the 2003 harvest date in Dijon, on which the individual vine-growers decided in place of the village community". Your reasoning is not understandable at this stage of the paper. You must reach this conclusion after having explained why you think GHD would not correspond to maturity in this case.

Reply: As this paragraph seems to be misleading, we will delete it. We will try to explain more convincingly in the revised paper why GHD was not a reliable indicator for grape maturity in the record breaking hot and dry year 1540. It needs to be stressed that the begin of harvest depends on human decision-making (see Legrand 1978). It is not necessarily related to physical parameters. Garnier et al.(2011) provide many examples of how human decision-making in the context of extraordinary situations – wars and epidemics. In the case of 1540 it was possibly a once in five centuries drought during from February to November. In early August, grapes were overripe looking like raisins and grape harvest was postponed.In many communities vine growers waited for some relieving rain to raise the berries content of juice.

RC 18: Line 12 page 2704: Besançon is not in western France

Reply: We will correct this mistake in the revised paper.

RC 19: Line 4 page 2705: You could also compare these dates with those reported in Daux et al. (2012) for Jura and Northern Rhône valley.

Reply: This would be possible. But as indicated in RC 15 decisions about the begin of harvest did mainly depend on rainfall due to the desiccation of the berries. The date given by Chuine et al. (2004) for 1540 is DOY 278 (4th October). How is this compatible with date(s) of 3rd September provided by Daux et al.?)

RC 20: Line 12 page 2705: "around mid-August" this is rather vague. Schweinfurt is about 400 km far from Switzerland. Do you think it is an appropriate comparison?

Reply: Please consider the context (p. 2705): "The three approaches [described above] are qualitatively consistent with a maturity related grape harvest in 1540 that might have started between 12th and 24th August." Nobody is able to assess full grape maturity in this exceptional year more precisely. The reference to Schweinfurt is, by the way, just one on several coherent proofs.

RC 21: Line 20 page 2705: Your series is not a GHD only series: 60% percent of the data are GHD, the other 40% are not. Please correct throughout the manuscript. You cannot just refer to the homogenization procedure. It must be summarized here.

Reply: see reply to RC 11. This is a terminological issue. In our understanding the term of GHD covers all appropriate evidence, whether related to the harvest ban or produced by eye observations (HPD, PNO) or related to bookkeeping procedures (WPD). We will explain in the final version how to understand the term. Instead of GHD we might speak of "ban-related GHD".

RC 22: Line 26 page 2705: How did you correct for the altitude? Did you use Basel altitude (which is not the place where the harvest took place) or each village altitude?

Reply: see reply to RC 13.

RC 23: Line 16 page 2706: "merging together different, partly short, GHD series". You must explain in details how you merge the series.

Reply: In our opinion it is explained in detail on the pages of 2706, line 16 ff and 2706 up to line 19. Homogenisation of data type should be deleted because of the reasons explained in reply to RC 21.

RC 24: Line 2 page 2707: Can you estimate an uncertainty on the 17 days difference?

Reply: We refer to table 2. (esp. Std)

RC 25: Line 5 fwd page 2707: Page 2700, you develop why GHD are not good temperature proxies. You cite Guerreau (1995) who concluded that the relation between grape harvest and temperature is not stationary. Why should Swiss series be an exception to this rule?

Reply: This is a claim that you read into the text. Of course, Swiss GHD suffer from the same problems.

RC 26: Line 5 page 2707: What is the estimated uncertainty on your averaged GHD? At least you should discuss the value of the standard deviation. Garcia de Cortazar et al. (2010) and Daux et al. (2012) have discussed the question of uncertainty on GHD extensively.

Reply: We will look into their arguments. The standard deviations of the series are given in Fig. 2 (see box). They amount between 7.59 and 11.89. Standard deviation of the averaged GHD series (average DOY of all available GHD series per year) amounts to 10.1 days.

RC 27: Line 6 page 2707: there are many ways to calibrate : : : you must explain how you proceed precisely

Reply: The temperature reconstruction was achieved using a standard linear calibration- and verification approach as described e.g. by Brázdil et al. (2010). A transfer function was established between the GHD series (predictor) and the mean April-July temperature HISTALP NW series (predictand) over the period. In the absence of any trend the approach of linear regression is adequate. Equations were calculated for the best correlating 50 yr sub periods (1774-1824; and 1955-2005). Both equations were then used to reconstruct mean April-July temperature anomalies during the instrumental period. The linear regression equation derived from the 1774-1824 sub period was verified in the 1955-2005 and vice versa. Better verification results were found with the 1955-2005 linear regression equation ($r = 0.74$, $p = 0.001$). This equation thus was used to reconstruct mean April-July temperature anomalies from the homogenised Swiss GHD series from 1444-2011. The equations will be provided within the supplementary material on internet.

RC 28: Line 10 page 2707: your target is AMJJ while Chuine et al 's was AMJJA. Can it explain some differences between you and them knowing that August was very hot in 2003 (as mentioned in the introduction)?

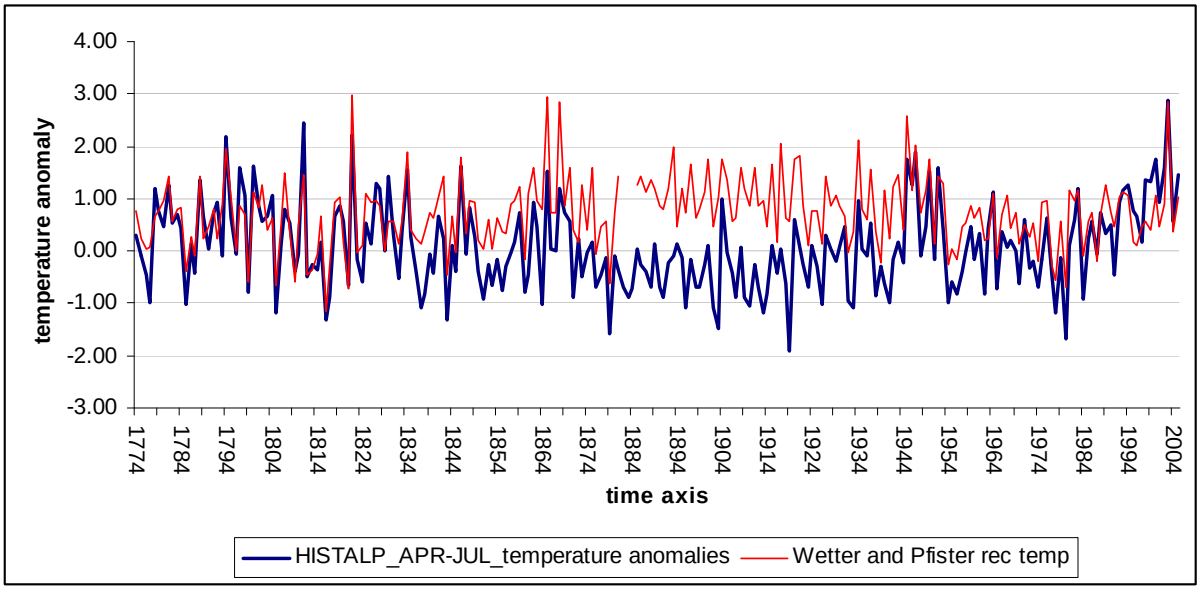
Reply: According to the article by Jones (2003) in the standard textbook on phenology by Schwartz (which is quoted in Chuine et al. 2004) véraison is "near the end of July or the first week of August," Garcia de Cortazar et al. (2010) mention that the delay between véraison and harvest date was found to be almost constant. These facts agree with the point of view by Dr. John Gladstones (2011) who writes about a "widely observed phenomenon that temperatures of the first two or three growing season months, or alternatively the date of flowering, can usually predict quite closely the dates of véraison and maturity to follow. …] The later phenological intervals [i.e. véraison] show little response to temperature, and tend to be constant from year to year." (p. 17). These considerations are highly relevant in the case of 2003 and 1540, when harvesting took place in mid August. How can you unconditionally support the point of view of Dr. Pascal Yiou that temperatures during the first decade of August 2003 still mattered for grape maturity under the extreme conditions during that year? Statistical analyses support the point made by Jones (2003), Garcia de Cortazar et al. (2010) and Gladstones (2011). Likewise, Guerreau (1995) confirmed conclusions by Pfister (1984) that the grapevine harvest date depends on temperature from April to July. We could replicate this result from a stepwise regression of our new Swiss series, and also from the data given in Chuine et al. (2004).

RC 29: Line 19 page 2707: 0.52_C: this is a minimum value which takes into account only the model error.

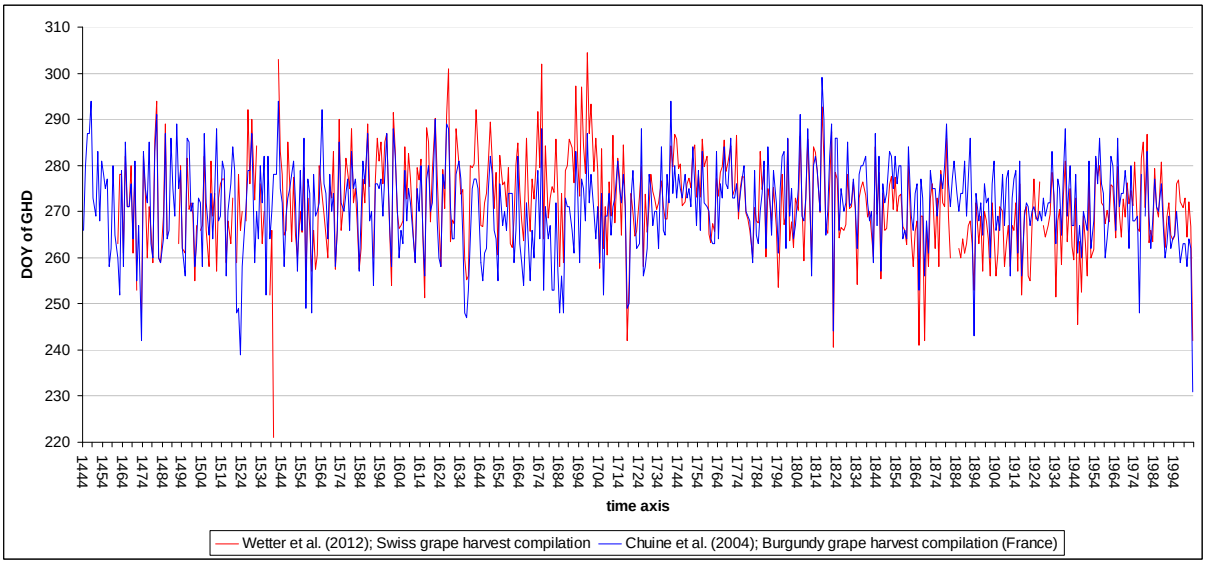
Reply: You are right. In our AMJJ-temperature anomaly reconstruction we thus chose the double SEE as confidence interval which is the standard procedure in temperature reconstruction papers (see Fig. 5; 2 x σ 1774-1824 verification).

RC 30: Line 25 page 2707: Except 2003, how do your extreme years compare with those in Chuine et al.? and with HISTALP data?

Reply: HISTALP – Wetter & Pfister (2012) see Fig. below:



You can see that there is quite a good overlap between our reconstructed and HISTALP temperature anomalies from 1774 to 1850 and from 1950 to 2005 (also in the warm and cold extremes). The lower compliance during the intermediate period is mainly caused by low GHD availability during that period. For the late 19th and early 20th century only very few and scattered HPD are available which furthermore only can be obtained in their printed version, because the original handwritings are lost. Critical verification with the originals, as it was done with the other GHD series, thus was not possible in that period. Major errata might be the reason for the bad compliance. We will address this weakness in the final version of the paper. The situation rapidly improves after MeteoSwiss decided to do phenological network observations (PNO) in 1951. Compliance between Chuine et al. (2004) and Wetter and Pfister (2012) see in figure below



Comparisons between 'ordinary' extreme years are not conclusive with regard to the comparison between the two record breaking summers of 1540 and 2003. In 1540 conditions were radically different from those in all other extreme years since that time as we demonstrated in our article.

RC 31: Line 13 page 2708: "critically verified": this aspect cannot be appreciated in this manuscript. Line 17 fwd page 2708: provide the number of data for the correlation, please. All the correlations are decreased in the 18th century. How do you interpret this point?

Reply: All GHD series were critically verified, expect PNO series from MeteoSwiss (we believe them to be reliable) and some GHD series between 1850 and 1950 were original series are not available anymore (see reply to RC 30). But if this term really should be a bone of contention we will agree to delete it in the revised paper version.

Correlations between Wetter and Pfister (2012) and other GHD series (see Table 3):
Wetter and Pfister (2012) – Garnier et al. (2011); 0.81, p = 0.01, n = 253
Wetter and Pfister (2012) – Dobrovolny et al. (2010); -0.75, p = 0.01, n = 343
Wetter and Pfister (2012) – Labbé and Gaveau (2011); 0.73, p = 0.01, n = 403
Wetter and Pfister (2012) – Wetter and Pfister (2011); 0.65, p = 0.01, n = 480
Wetter and Pfister (2012) – Chuine et al. (2004); 0.65, p = 0.01, n = 528
Wetter and Pfister (2012) – Maurer et al. (2009); 0.54, p = 0.01, n = 314
Wetter and Pfister (2012) – Kiss et al. (2011); 0.54, p = 0.01, n = 194
Wetter and Pfister (2012) – Battipaglia et al. (2010); -0.53, p = 0.01, n = 530
Wetter and Pfister (2012) – Büntgen et al. (2006); -0.48, p = 0.01, n = 530

The decreased 31-yr moving correlations in the 18th century between our GHD- and GHD series from other authors clearly indicate that in this period our series must have some problems. So far we have no explanation for the weakness of our GHD in that century. Nevertheless most of the 31 yr moving correlations stay above the critical 0.6 correlation coefficient value. Battipaglia same as Büntgen et al. MXD series (not shown) have more or less overall weak correlation coefficients. Maurer et al. (2009) state that they have problems with their data in the 18th century and Kiss et al. (2011) base almost on the same data set as Maurer et al.(2009).

RC 32: Line 25 page 2709: "the relationship between temperature and grape maturity is non linear": isn't your reconstruction based on a linear calculation?

Reply: This quote is taken out of context: We read: "Considering the good agreement between the estimated and the measured temperature anomaly of 2.9 ºC in 2003 (see section 3) we may, however, conclude that spring-summer (AMJJ) temperatures were very likely higher in 1540. In interpreting this result it may also be worth noting, that the relationship between temperatures and grape maturity is non-linear. Grapes rate of net photosynthesis e.g. decreases significantly above 35°C. The absolute limit of CO_2 absorption is reached if 40°C are achieved (Currle et al., 1983). The plant in such cases temporarily stops its vegetative activity." After all, this statement refers just to the extraordinary situation in 1540.

RC 33: Line 24 page 2711: "The temperature is 2.4_C higher than measured temperatures in Paris". The Chuine et al.(2004) paper does not pretend reconstructing the temperature in Paris but in Burgundy. The difference of temperature between the two stations can be sizable. For instance, the August Tmax in Dijon was 32.2_C, while it was 29.9 in Paris (Meteofrance web site). The difference between Tmean (not given in the website) may not be so big. Anyway, it is clear that the comparison between reconstructed and measured temperatures should definitely be made with Dijon station and not with Paris.

Reply: We have already dealt with the issue of August temperature in the reply to RC28

The reference to the Dijon temperatures is easily overlooked, because it is only included in the supplementary material on internet, whereas Paris temperatures are mentioned in the article itself. We will correct this mistake in the corrected version of our paper.

RC 34: *Figures : x-axis are unreadable or even non-existent (fig. 7). Chose a 2, 5 or 10-year interval and begin by a 10 or 5 multiple. The figures are too small. Units must be provided.*

Reply: We will provide new figures according to your instructions in the revised paper.

RC 35: *Figure 2: Situate year 1540. In the figure caption define what mean and SD refer to.*

Reply: All information is contained in the box (Fig 2). As Fig 2 is not a time series, the year 1540 cannot be located.

RC 36: *Figure 3 and related text: the regression is strongly influenced by the two outsiders. You should choose another way of calculating it (rank correlation for instance).Remove the horizontal lines.*

Reply: One of the two lowest points in figure 3 is Basel from which most of the data prior to the 1550s originate. We don't think that the approach of rank correlation would alter the main results of our paper. If we did e.g. a simple homogenisation of long term mean GHD of the 17 series (as Chuine et al. did), the correction would be more or less the same as with the used linear regression approach. Given the fact of such a strong correlation between altitude and GHD (after dating type- and grape variety corrections have been done) we prefer to do the altitude correction with the linear regression approach. Horizontal lines are provided to improve readability but can be removed if requested (?)

RC 37: *Figure 4: there is some kind of divergence from 1990 onwards between calibrated grape harvest and measured temperature. Do you have any idea why? When you calibrate on 1774-1824, the verification in the 1955-2005 shows a systematic 0.5 _C bias from 1955 to ca 1990. Can you comment on it?*

Reply: The reasons for this feature would have to be analysed in a particular analysis outside of the focus of this paper. For the same reason we cannot deal with biases in the 1999-2005. However, we think that this problem is not relevant for the focus of our paper.

RC 38: *Tables: Table 2: Provide the dates of start and end of the series please.*

Reply: Basel hospital WPD series; 1444-1705
Mulhouse GHD series; 1694 – 1728
Swiss GHD series; 1444-2011

RC 39: *Technical comments Line 20 page 2696: Â´n As such: : :, long records Â˙z Two sentences*

Reply: Thank you. We will correct this in the revised paper.

RC 40: Line 24 page 2702: HPD's? or HPDs?

Reply: we will delete the "s" in the revised paper, because HPD is already a plural form.

RC 41: Line 7 page 2703: define BBCH

Reply: The BBCH scale is generally applied in phenology since more than 20 years (see Meier et al. 2009). As explained in his paper, the abbreviation BBCH stands for **B**iologische **B**undesanstalt, Bundessortenamt and **CH**emical industry (see BBCH Monograph). This scale is also addressed by Jones (2003) who is referenced in the Chuine et al (2004) article.

RC 42: Line 18 page 2706: "location, etc:" already said in lines 12-

Reply: We cannot find the double saying about location on page 2706. Could you please be more precise about this query!

References:

Battipaglia, G., Frank, D., B¨untgen, U., Dobrovoln´y, P., Br´ azdil, R., Pfister, C., and Esper, J.: Five centuries of Central European temperature extremes reconstructed from tree-ring density and documentary evidence, Global Planet. Change, 72, 182–191, 2010.

Büntgen, U., Frank, D. C., Nievergelt, D., and Esper, J.: Summer temperature variation in the European Alps, AD 755–2004, J. Climate, 19, 5606–5623, 2006.

Chuine, I., Yiou, P., Viovy, N., Seguin, B., Daux, V., and Le Roy Ladurie, E.: Grape ripening as a past climate indicator, Nature, 432, 89–290, 2004.

Cook ER, Briffa KR, Jones PD (1994) Spatial regression methods in dendroclimatology: a review and comparison of two techniques. Int J Climatol 14(4):379–402. doi:10.1002/joc.3370140404

Currle, O., Bauer, O., Hofa¨cker, W., Schumann, F., and Frisch, W.: Biologie der Rebe. Aufbau, Entwicklung, Wachstum, Meininger Verlag, Neustadt an der Weinstrasse, 301 pp., 1983.

Dobrovolny, P., Moberg, A., Brazdil, R., Pfister, C., Glaser, R., Wilson, R., van Engelen, A., Limanowka, D., Kiss, A., Halıckova, M., Mackova, J., Riemann, D., Luterbacher, J., and Böhm, R.: Monthly, seasonal and annual temperature reconstructions for Central Europe derived from documentary evidence and instrumental records since AD 1500, Climatic Change,101, 69–107, 2010.

Federal Biological Research Centre for Agriculture and Forestry: Growth stages of mono- and dicotyledonus plants. BBCH Monograph (ed. by. Meier, U.), 1–158, 2001.

Garnier, E., Daux, V., Yiou, P., and Garcıa de Cortazar-Atauri, I.: Grapevine harvest dates in Besancon (France) between 1525 and 1847: social outcomes or climatic evidence?, Climatic 25 Change, 104, 703–727, 2011.

Gladstones, J. Wine, terroir and climate change. Kent Town, Wakefield Press, 2011

Jones GV. Winegrape phenology. In: Schwartz MD (ed.) Phenology: An integrative environmental science. Dordrecht: Kluwer, 2003, 523–39.

Labbé , T. and Gaveau, F.: Les dates de bans de vendange a` Dijon: etablissement critique et révision archivistique d'une série ancienne, Revue historique, 657, 19–51, 2011.

Keenan, D.: Grape harvest dates are poor indicators of summer warmth, Theor. Appl. Climatol., 87, 255–256, 2007.

Kiss, A., Wilson, R., and Bariska, I.: An experimental 392-yr documentary-based multi-proxy (vine and grain) reconstruction of May–July temperatures for Ko¨szeg, West-Hungary, Int. J. Biometeorol., 5504, 595–611, 2011.

Maurer, C., Koch, E., Hammerl, C., Hammerl, T., and Pokorny, E.: BACCHUS temperature reconstruction for the period 16th to 18th centuries from Viennese and Klosterneuburg grape harvest dates, J. Geophys. Res., 114, D22106, doi:10.1029/2009JD011730, 2009.

Meier, N., Rutishauser, T., Pfister, C., Wanner, H., and Luterbacher, J.: Grape harvest dates as a proxy for Swiss April to August temperature reconstructions back to AD 1480, Geophys. Res. Lett., 34, L20705, doi:10.1029/2007GL031381, 2007.

Mullins, M. G., Bouquet, A., and Williams, L. E. (2007): Biology of the Grapevine, Cambridge University Press, Cambridge.

Schlegel, W. (1973). Der Weinbau in der Schweiz. Wiesbaden.

Schweingruber, F. H. (1996). Tree Rings and Environment. Dendroecology. Birmenstorf.

Wetter, O. and Pfister, C.: Spring–summer temperatures reconstructed for northern Switzerland and southwestern Germany from winter rye harvest dates, 1454–1970, Clim. Past, 7, 1307–1326, doi:10.5194/cp-7-1307-2011, 2011.

Wetter, O., Pfister, C., Weingartner, R., Luterbacher, J., Reist, T., Troesch, J.: The largest floods in the High Rhine basin since 1268 assessed from documentary and instrumental evidence. Hydrological Sciences Journal 56(5), 733-758, 2011.