

**QUESTION1: I mentioned to update your reconstruction to 2015. You have presented it to 2012. It would be nice to have the last couple of years as well and I can imagine, that data would be available. This would mean to update the figures also. I think it would be worthwhile to do so.**

RESPONSE1: GHD and temperature series have been completed to 2015 and in the new version of the article all texts and figures have been also updated where relevant.

**QUESTION2: Another point I mentioned was to add a couple of references and comparison with for instance with Luterbacher et al. 2004 and 2016. I saw you did include this in the figure but the main text does not include a judgment/assessment how well your reconstruction performs compared to all the other data you have used. It would be important to add a bit of text about this comparison.**

RESPONSE2: The text concerning comparison of our new reconstruction with other existing proxy reconstructions was extended significantly in Section 5 (Discussion) as follows:

As follows from Section 4.3, overall correlations between the new GHD-based temperature series and 15 different European proxy reconstructions are relatively high and statistically significant. However, those relations are not stable over time and periods with evident drops in running correlations occur quite frequently. The most evident periods of lower shared variance refer to the same time intervals in most of the series compared. This indicates that there must be some common causes, examined below for the four most obvious periods (Fig. 9).

Periods with statistically significant correlations were addressed by means of the Monte Carlo simulation to take into account the degree of freedom from autocorrelations. This approach confines to an acceptable level the probability that significant correlations might be found only “by chance” (i.e. type I errors). The significant running correlations between our new GHD series and 15 existing reconstructions may be further affected by other factors, such as significant local trends in compared series. This may be the case for the second half of the 20th century, which shows high and largely significant running correlations for all the series compared. More obvious delimitation of the time intervals for which the new GHD reconstruction shows the best agreement with other proxy reconstructions (Fig. 9) may be expressed by the mean of running correlations ( $\bar{r}$ ). In general, the best performance was found for the 1530s–1630s, the 1780s–1830s and from the 1950s onwards. Higher correlations with documentary-based reconstructions compared with those based on tree-rings underlines the importance of a multi-proxy approach. In this sense, direct comparison of the new GHD Czech reconstruction with the European multi-proxy JJA temperature reconstruction of Luterbacher et al. (2016) provides promising results. Running correlations are mostly significant, with the exception of only two relatively short periods (the 1650s and mid-18th century), which are discussed above (Fig. 9, B and C). This result also demonstrates the distinctly spatially representative nature of the new Czech GHD reconstruction, which is not confined to the territory of the Czech Lands. Moreover, the GHD series also shows high potential for the study of the long-term variability of droughts (Cook and Wolkovich, 2016; Možný et al., 2016).

**QUESTION3: In Fig 5 and subsequent, it is not clear to me why over the past decades the one recon is systematically higher than the other one. It would be nice if you could explain this in the main text more pronounced.**

RESPONSE3: This feature was explained in detail in Discussion. The following text was added:

At first sight, the LR and VS methods used in this analysis appear to have provided a very high level of reconstruction skill, as summarised in Table 1. However, closer examination of Figs. 5 and 6, where reconstructed and target temperatures for early/late calibration periods are compared, reveals certain inconsistencies. When the earlier period (1811–1910) is used for calibration, reconstructed April–August temperatures clearly underestimate those measured in the most recent decades. This is more pronounced for the LR method (Fig. 5) than for the VS method (Fig. 6). This probably relates to the fact that there has been a highly significant positive temperature trend since the 1970s, whereas the calibration period (1811–1910) shows lower temperatures, lacking any long-term positive trend. This anomaly is a common feature of regression-based methods, since they are able to reproduce well only that which they have “experienced”. Moreover, differences between the measured and reconstructed temperatures in recent decades are also partly related to the fact that regression-based methods suppress the variability of reconstructed values – a feature discussed in Section 3. The effect described here is partly eliminated by the final reconstruction being based on calibration employing the full overlap period (1811–2010). In spite of this, lower reconstructed temperatures compared to those measured are evident in Fig. 8, especially in smoothed series for recent decades. This feature demonstrates further the exceptionality of recent global warming at a central European scale over the last five hundred years.

**QUESTION4: Please also correct in Fig 5 verificaton to verification**

**RESPONSE4:** Corrected in Figs. 5 and 6

**QUESTION5: In Fig 9 you use 31 year running correlations but you dont mention the significance for it taking also in consideration the degree of freedom from autocorrelations. It would be important to modify the text related to any figure with running correlations and just mention those periods for which the running correlation exceeds the significance levels.**

**RESPONSE5:** In the new version we completely changed the way how the significant running correlations were estimated. We used Monte Carlo approach to avoid multiple comparisons problem. Explanation of the method is provided in the Section 3 (Methods) as follows:

The significance level was estimated by means of the Monte Carlo simulation to oversee the effect of multiple comparisons (Livezey, 2010). Running correlations were calculated for 1000 permutations of the time series compared and the highest correlation for each trial established. Finally, the 95th quantile from the 1000 values of the highest correlation coefficients determined the critical value at  $p = 0.05$  for the one-tailed test.

Moreover, Fig. 9 was re-drawn and the text concerning interpretation of running correlations was also completed. See the text in RESPONSE2 above.

**QUESTION6: you might also add the following reference of a paper that was published recently. B. Cook et al. Nature Climate Change and references therein**

**RESPONSE6:** This paper was mentioned in the new version of the article (see your QUESTION2).