Final Response to Neil Macdonald

The use of wells as a proxy for groundwater availability and therefore as a climate indicator is novel and therefore I was excited to review this paper.

Response: We thank you for your constructive comments on our manuscript. Your comments and suggestions give us much help for our future research. Our responses are listed below:

Reviewer's main comments to Authors:

1. The use of well bottom depth is to my knowledge an innovative and interesting approach to detailing past water levels, however they require critical assessment, considering a number of factors, such as do the well bottoms reflect original construction depths, or might they have been excavated several times, as required when water levels dropped further, it also requires an assumption that the recharge in the well was consistent through time and then this was sufficient to meet supply, if demand changes (i.e. population growth) then it may have been necessary to increase the depth.

Response: Thank you for the comment. There are some reasons supporting that the bottom elevations of ancient wells could reflect PUWL in history. First, archaeologists can distinguish the strata according to the soil characteristics and colors and the age of ancient wells according to the shape of wells. In general, they can determine the ages and the bottom of most of the ancient wells. Besides, our data was collected from the archaeological excavation reports that contain specific details of the ancient wells, such as their locations, shapes, head and bottom positions, and so on. Second, archaeologists can distinguish whether those wells were deepened or re-excavated based on the archaeological strata. For instance, the Du-J3 well in Chengdu Sheet can be divided into two parts, namely Du-J3(A) and Du-J3(B). Du-J3(B) was initially built during the Mid-Tang dynasty (A.D. 766-835), while Du-J3(A) was rebuilt during the Wudai dynasty (A.D. 907-960) (Luo and Xu, 1991) (supplementary file). PUWL corresponds to its bottom elevation in the Mid-Tang dynasty since the well was undoubtedly built at that time, and the bottom elevation of Du-J3 represented PUWL in Mid-Tang rather than the Wudai dynasty. Third, there were very few re-excavated wells in our study areas, and the relatively high density of wells reveals that. For instance, the density of ancient wells in the Chenghu Site reached 40.2/ km^2 (Suzhou Museum, 2007), which means that there was a well every 300 meters. Such high well density implies that residents would rather build a new well than deepen the existing wells to meet their water demand. Fourth, when a place was not suitable for living due to drought, people might choose to migrate to other places, abandoning the original wells instead of deepening them. Based on the above reasons, we suggest that the well bottoms can reflect their original construction depths.

References:

Luo, E. H. and Xu, P. Z.: Excavation report of Tang-Song remains in Dufu Thatched Cottage site, Chengdu, Southern Ethnology and Archaeology, 0, 233-298+312-313+319-327, https://doorg/CNKI:SUN:NFMZ.0.1990-00-020, 1991.

Suzhou Museum: New Discoveries in Suzhou Cultural Relic Archaeology, Guwuxuan Press, Suzhou, China, 99-116, 2007.

2. I also expected some discussion of the local hydrogeology at the sites of study to determine whether conditions were similar or comparable, discussion of lag from precipitation to groundwater level and sensitivity to droughts, possibly using contemporary evidence. Unfortunately, none of this was sufficiently explained, as such the principles of using wells as a water table proxy are assumed and not evidenced, as this is a 'novel' approach this is a fundamental requirement. You need to convince the reader, therefore much greater explanation and justification is required for each of the sites studied.

Response: Thank you for the comment. Indeed, previous studies have illustrated the relationship between well bottoms and PUWL, revealing that well bottoms could be a good proxy of PUWL. But, those studies are mostly based on a small area or a short time section, which could not give a long-term and large-spatial picture of the connection between well bottoms and PUWL. Therefore, we take a step forward by using archaeological records to reconstruct the PUWL chronology, showing the environmental change over a long time span and large region. In the revised manuscript, we have strengthed the introduction section (lines 52-58) to make clear the scientific foundation of this study:

"Water wells are built by excavating the ground and digging, driving, boring, or drilling to access groundwater in aquifers, while groundwater was one of the water sources exploited to meet human needs since prehistoric times. Ancient Chinese created sophisticated tools for drilling water wells that are similar to modern machines (Voudouris et al., 2019). It is worth noting that many well-preserved ancient wells were indicators of PUWL (Jorgensen and Walid, 2003), and well bottoms were also recognized been related to sea level (Sivan et al., 2001). Studies of historical wells inferred that well bottoms were well correlated with PUWL: PUWL was 30 cm higher than well bottoms in Caesarea Maritima, Israel (Sivan et al., 2004) and 60 cm above well bottoms in the Northern coast of Israel (Nir and Eldar, 1987; Galili and Nir, 1993), while it is

approximately 80-90 cm higher than well bottoms in the lower Yellow River area (Yu et al., 2018). Although the height differed between PUWL and well bottoms from region to region, the relationship between these two levels has remained similar with time, making it reliable to use the well bottoms to indicate PUWL. Indeed, the variations of well bottoms were employed to indicate sea-level fluctuation during East Han to Ming dynasty in Yancheng city, China (Sheng and Zhu, 2004). Therefore, 482 ancient water wells were collected from published archaeological reports, which were dated by typological and cultural methods, and the mean well bottoms were calculated of each dynasty to build up the variations of PUWL in every dynasty."

Regarding the issue of the local hydrogeology in different regions, we chose our sampled sites that located in the plain regions to control the influence of terrain differences on PUWL. Also, we based on previous studies to assume that well bottoms reflected a constant depth of the PUWL in the plain region. We have also elaborated this point in the revised manuscript:

***Line 75: Changed the sentence "Second, some cities with flat and less undulating terrain were chosen. This is because large undulating terrains influence the groundwater level changes (e.g., Chongqing). In contrast, those cities located in relatively flat areas were chosen (i.e., Chengdu, Changsha, and Nanjing)." to "Second, since a plain limited its PUWL in a steady-state condition of flow (Galili and Nir, 1993), some cities with rugged and more undulating terrain were excluded, because large undulating terrains influenced the groundwater level changes (e.g., Chongqing). In contrast, those cities located in relatively flat areas were chosen (i.e., Chengdu, Changsha, and Nanjing)."**

Regarding the lag from precipitation to groundwater level and sensitivity to droughts, those factors are important in affecting PUWL at the short-term scale (e.g., daily, weekly, and monthly). But, the units of our PUWL reconstruction are dynasties, which are at the centennial scale. As the abovementioned factors are not significant in affecting PUWL at the centennial time scale, they are not covered in this study.

References:

Galili. E. and Nir. Y.: The submerged Pre-Pottery Neolithic water well of Atlit-Yam, Northern Israel, and its palaeoenvironmental implications, Holocene, 3, 265-270, https://doi.org/10.1177/095968369300300309, 1993.

Nir, Y. and Eldar, I.: Ancient Wells and Their Geoarchaeological Significance in Detecting Tectonics of the Israel

Mediterranean Coastline Region, Geology, 15, 3-6, https://doi.org/10.1130/0091-7613(1987)15<3:AWATGS>2.0.CO;2, 1987.

Sivan, D., Lambeck, K., Toueg, R., Raban, A., Porath, Y. and Shirman, B.: Ancient coastal wells of Caesarea Maritima, Israel, an indicator for relative sea level changes during the last 2000 years, Earth Planet. Sci. Lett., 222, 315-330, https://doi.org/10.1016/j.epsl.2004.02.007, 2004.

Sivan, D., Wdowinski, S., Lambeck, K., Galili, E. and Raban, A.: Holocene sea-level changes along the Mediterranean coast of Israel, based on archaeological observations and numerical model, Palaeogeogr. Palaeoclimatol. Palaeoecol., 167, 101-117, https://doi.org/ 10.1016/s0031-0182(00)00234-0, 2001.

Yu, S. Y., Chen X. X., Liu, X. L., Fang, Z., Guo, J. F., Zhan, S. Y., Fang, H. and Chen, F. H.: Ancient water wells reveal a prolonged drought in the lower Yellow River area about 2800 years ago, Sci. Bull., 63, 1324-1327, https://doi.org/10.1016/j.scib.2018.09.017, 2018.

3. I was interested to see how the wells wound be dated, however I am still uncertain, whether these are assigned to dynasties or dated with geo-chronological approaches, I suspect the former, if so this presents challenges in understanding the figures presented. Fig 2 presents the depth and ages of the wells, dynasties across the top, calendar years across the bottom. Therefore I assume that the order that the wells are presented in is meaningful, and that some dating is undertaken? How was this achieved, you must explain this. Similarly the use of dynasties of varying length is challenging when comparing across time, as some long dynasties appear to have trends within these periods e.g. Chengdu – Tang period with increasing depth to the wells towards the end of the period, as such presenting this with a single point is unhelpful. Greater explanation of the methods used to date and organise the wells and potential uncertainties in ages needs to be discussed. You need to discuss the local hydrogeological conditions in more detail, what is the lag time to drawdown/recharge of the well. You discuss historical events, but these are of limited meaning without basic site information.

Response: Thank you for your comment. The wells are dated based on their shape and structure, and the ages of the wells are only assigned to dynasties. Hence, we can only compare the phased changes of PUWL among different dynasties. Although our PUWL reconstruction method has its limitation to cater to those dynasties with a very long duration, there is so far no better approach to reflect the variations of PUWL in historical periods. We have also modified the scales of abscissa to the starting and ending times of historical dynasties rather than equidistant chronological intervals (Figure 2). About the issue of the local hydrogeology in different regions, we chose our sampled sites that located in the plain regions to control the influence of

terrain differences on PUWL (see our response to point #2). Besides, as the ages of the wells are only assigned to dynasties, which are at the centennial time scale, factors such as the lag time to drawdown/recharge of the well that are important in the short-term scale may not be useful in this regard.

4. You include some short or patchy series, remove these from your discussion they offer little e.g. Ezhou. You present little discussion of data/well densities, this has important implications, as new wells impact on older wells if in similar areas, drawing down local water levels. This is needed so that the reader can understand your approaches to data quality in a novel approach such as this.

Response: Thank you for the comment. We have addressed some of those issues (see our responses to point #2). About those sites with short or patchy series (e.g., Ezhou), as they are located in our study area, if possible, we would like to keep them in our discussion.

5. I found your attempt to group the sites together disconcerting, why would Chengdu and Changsha respond or record similar trends, they are >750km away from each other and potentially driven by different climatic processes, in different hydrogeological regions? They certainly have different precipitation patterns (average precip. of ~850mma-1 at Chengdu, ~1800 at Changsha mma-1). It was unclear to me why you attempted to create artificial regions (coastal, inland and transitional). I found that much of the discussion was unfounded and focused on attempting to assign attribution to 'trends' or 'patterns' in the data that simply were not clearly evidenced. These were not statistically or robustly presented, nor justified. I actually felt that this was beyond the scope of the paper, you need to convince the reader first of the robustness of your approach, that should be the focus of the paper.

Response: Thank you for the comment. We calculated the mean values of well bottoms in each dynasty and compared them with the means of all wells in every city. As shown in Figure 2, the periods with high or low PUWL were determined according to the means in a dynasty higher or lower than the mean of all wells in every city. After comparison, it is shown that the variations of PUWL in Chengdu and Changsha are similar, which is likely attributed to their similar climate pattern. As shown in Figure R.1, the two cities are located on the same monsoon front (Zhao and Hao, 2019). Their variations in precipitation are driven by the East Asian Summer Monsoon, and their precipitation also comes from the same vapor source. In addition, the annual precipitation between these two cities is not significantly different. The mean annual precipitation was about 1000 mm in Chengdu (Zhang et al., 2020) and about 1300 mm in Changsha (Li and Zhang, 2015) over the last 50 years.

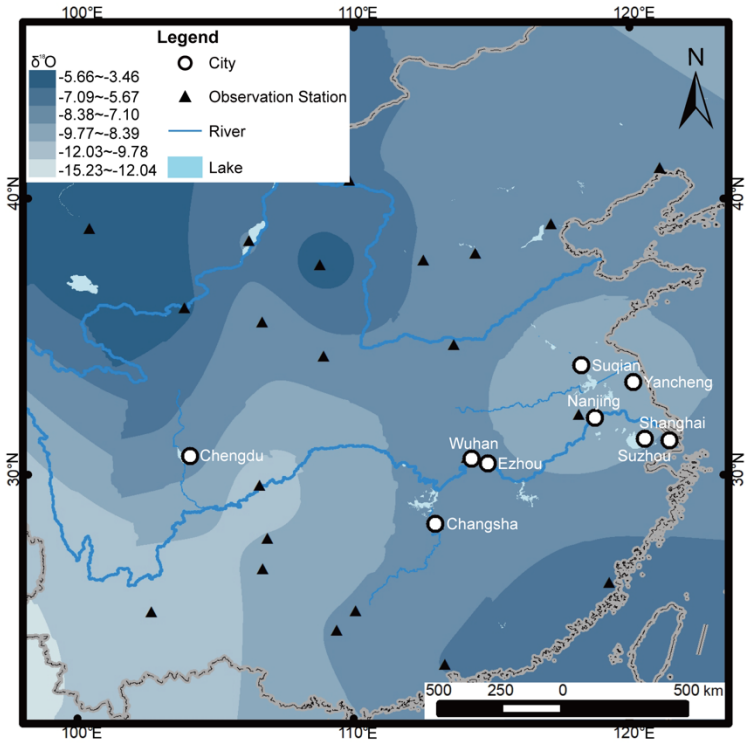

Fig R.1 Distribution of $\delta^{18}O$ in China (Zhao and Hao, 2019)

About the grouping of Suqian and Yancheng, this is based on their coastal locations. Previous studies indicated a significant correlation between the PUWL variations and the sea-level fluctuations in the coastal regions (Shen and Zhu, 2004; Sivan et al., 2004). In addition, the fluctuations of sea levels were driven by the temperature changes (Rohling et al., 2009; Siddall et al., 2010; Kopp et al., 2016), as the rising and falling temperatures would lead to the melting and the growth of glaciers and subsequently result in the rise and the fall of sea levels (Clark et al., 2001). Our results also showed that PUWL in Suqian and Yancheng were similar (both were high at Tang and Song dynasties while low at Ming dynasty) and correlated well with temperature changes in the past 2000 years. The two cities should be grouped together for interpretation based on the above findings.

References:

Clark, P. U., Mix, A. C. and Bard, E.: Ice sheets and sea level of the Last Glacial Maximum, EOS: Earth Space Sci., 82, 241-247. https://doi.org/10.1029/01EO00133, 2001.

Galili. E. and Nir. Y.: The submerged Pre-Pottery Neolithic water well of Atlit-Yam, Northern Israel, and its palaeoenvironmental implications, Holocene, 3, 265-270, https://doi.org/10.1177/095968369300300309, 1993.

Kopp, R. E., Kemp, A. C., Bittermann, K., Horton, B. P., Donnelly, J. P., Gehrels, W. R., Hay, C. C., Mitrovica, J. X., Morrow, E. D., Rahmstorf, S.: Temperature-driven global sea-level variability in the common era, PNAS, 113, E1434-E1441, https://doi.org/10.1073/pnas.1517056113, 2016.

Li, X. L. and Zhang K.: Characteristics and variation trend of precipitation in Changsha in recent 50 years, Rural Economy and Science-Technology, 9, 44-46, https://doi.org/10.3969/j.issn.1007-7103.2015.09.016, 2015.

Nir, Y. and Eldar, I.: Ancient Wells and Their Geoarchaeological Significance in Detecting Tectonics of the Israel Mediterranean Coastline Region, Geology, 15, 3-6, https://doi.org/10.1130/0091-7613(1987)15<3:AWATGS>2.0.CO;2, 1987.

Rohling, E. J., Grant, K., Bolshaw, M., Roberts, A. P., Siddall, M., Hemleben, C. and Kucera, M.: Antarctic temperature and global sea level closely coupled over the past five glacial cycles, Nat. Geosci., 2, 500-504, https://doi.org/10.1038/ngeo557, 2009.

Shen, H. Y. and Zhu, C.: Relationship between the ancient wells and sea-level fluctuation from East Han to Ming Dynasty in Yancheng area, Mar. Geol. Front., 3, 25-29+2, https://doi.org/CNKI:SUN:HYDT.0.2004-03-004, 2004.

Siddall, M., Kaplan, M. R., Schaefer, J. M., Putnam, A., Kelly, M. A. and Goehring, B.: Changing influence of antarctic and greenlandic temperature records on sea-level over the last glacial cycle, Quat. Sci. Rev., 29, 410-423, https://doi.org/10.1016/j.quascirev.20 09.11.007, 2010.

Sivan, D., Lambeck, K., Toueg, R., Raban, A., Porath, Y. and Shirman, B.: Ancient coastal wells of Caesarea Maritima, Israel, an indicator for relative sea level changes during the last 2000 years, Earth Planet. Sci. Lett., 222, 315-330, https://doi.org/10.1016/j.epsl.2004.02.007, 2004.

Sivan, D., Wdowinski, S., Lambeck, K., Galili, E. and Raban, A.: Holocene sea-level changes along the Mediterranean coast of Israel, based on archaeological observations and numerical model, Palaeogeogr. Palaeoclimatol. Palaeoecol., 167, 101-117, https://doi.org/ 10.1016/s0031-0182(00)00234-0, 2001.

Zhang, L. Y., Mao, W. S. and Pang, B.: Characteristics of Climate Change in Chengdu Plain, Journal of Chengdu University of Information Technology, 35, 179-187, https://doi.org/10.16836/j.cnki.jcuit. 2020. 02.009, 2020.

Zhao, W. and Hao, C. Y.: Spatial Characteristics of Stable Isotope in Summer Precipitation in China Mainland, Meteorological and Environmental Sciences, 42, 54-59, https://doi.org/10.16765/j. cnki.1673 -7148.2019.01.008, 2019.

(Another references were listed on the question 2)

6. Overall I was disappointed, I believe you fail to demonstrate and justify the use of wells as a tool for understanding

past climates, however I still beleive there is considerable potential here, therefore the attempt to assign attribution to these changes is unsupported. The idea is novel but is not discussed sufficiently to demonstrate with confidence that it could result in meaningful data, nor could the reader based on this paper replicate your study within another region.

Response: Thank you for the comment. In this study, we systematically explore the relationship between climate change and PUWL based on the archaeological records of 482 wells. We also find that the patterns and variations of PUWL in different regions are attributable to different climate factors. Due to the data limit of ancient wells, we could not generate results that are comparable to high-resolution proxies such as tree-rings and lake sediments. This also implies that this study cannot cover those short-term factors that affect PUWL. But, our pioneering attempt may highlight the potential of using archaeological records in tracing paleo-environmental changes at the centennial time scale. This will be an important step in moving forward with this research topic. Besides, we have also provided the details of our sampled wells in the supplementary information. Readers can use our data to replicate the findings of this study.

Comments in noted manuscript

- *Comments about using "dynasty" as unit*

Line 92: this is a problematic unit to use, explained later. Rather than dynasties, average well depth every 50 or 100 years would be better, this could be presented as a running mean over a temporal window. Or are the wells just dated to dynasty based on their form?

Line 96: I would avoid the use of dynasties, they are not consistent lengths or include consistent climates, indeed some dynasties come to the end because of climate impacts it has been postulated. In providing mean values for the dynasties you end up with different temporal windows, for example at Chengdu, water levels at the end of the Tang are comparable to the Wu dai, but one period is ~400 years and other around 50. The result is that the means are not helping you.

Line 109-114: These are not periods of equal length or densities of information, which makes comparison challenging. For example, consider the Song period in Nanjing, your single point shows little change, however within the dynasty there is considerable noise. Similarly, the Tang period at Chengdu, a 'mean' that is low, but again considerable variability and a trend within the data.

Response: Thank you for the comments. The above issues are related to the limit of the well data. Based on the typological methods and cultural layers in archaeology, the ages of ancient wells can only be dated to dynasties. We are afraid that if we go beyond the data and seek to increase the temporal resolution of PUWL, our results may be biased. Given the above concern, we use dynasties as units.

- *Comments about grouping*

Line 94: these two cities are over 750km apart with a notable difference in elevation and I think different climates, why group them together? Is Suqian not about the same distance from he coast as Suzhou? you need to justify these groupings more clearly.

Line 97-99: by discussing as dynasties you infer that there is a relationship, there isn't, unless you are arguing that there is a trend in technology or practice. However looking at the data in figure 2 there are not consistent depths which would present a clear argument for not grouping together.

Line 108: I do not see this, indeed the trend at Changsha appears similar to Suqian.

Line 134: you have different patterns in a and b, so they alone are not synchronous, I do not see how you then make the leap to the climate proxies, you have not demonstrated sufficiently in my view that the approach produces robust data.

Response: Please kindly refer to our responses to the previous points (especially point #5).

- *Comments about historical documents evidence*

Line 104: this is interesting as widescale drought in Europe and N.America and some indications of severe drought in China too in 1540. Is there any further context.

Response: Thank you for raising this point. We will further explore it in our future studies.

Line 102-103: I think you need to take care in linking potentially short term drought, with longer term water level fluctuations, frequency of well construction may be a better indicator of dry phases. If a severe drought drys the well, why do they not dig it down further? Also each of these cities groundwater responses to inputs will have different response speeds, based on the hydrogeology which you have not discussed.

Response: Thank you for the comment. According to the archaeological excavation report, only a few wells were deepened. On the other hand, although the high frequency of well construction may indicate dry phases, the number of wells is always inflated in later periods because of the population increase. We also realize that the wells records in the Ming and Qing dynasty are far more than those in the earlier dynasty, such as the Warring States and Han. Therefore, we are afraid that the frequency of well construction may not be a good indicator for the dry phases.

Line 115-116: as previously noted, this covers one season not a whole dynasty - this season or year might be wet, but

this does not mean the whole dynasty

Response: We counted and compared the frequency of extreme hydrological events in the whole dynasty, and the sentence has been modified to "There were more severe floods (21) and rainfall events (12) than droughts (27) over the Song Dynasty." (line 114).

Detail Comments

Line 29: finish statement, how?

Response: Thanks for your suggestions. It is summarized from the widely-employed methods of reconstructing paleo hydrological variations mentioned in lines 32-42.

Line 57: so are well bottoms dated by geo-chronological processes, or by the architecture and hence to form? you need to be clear

Line 63-64: this needs to be explained in detail, so that the reader can understand the process - dated by form or geochronological methods?

Response: The ancient wells were dated by using typological methods and cultural layers, which are widely-accepted methods in reconstructing archaeological chronologies.

Line 72-74: does this not represent a challenge as it would mean you have increasing pressure on the wells as the population grows, the result is new wells are drilled which can often result in a reduction of well recharge, therefore the wells have to be redrilled or deeper wells made.

Line 78-80: any further criteria than the above 2 points?

Response: As mentioned in our responses to point #1, only a few wells were deepened according to the archaeological excavation reports.

Line 87-88: this is curious, as it would suggest relative surface have gone up and down at sites (Changsha), rather than just building up, why does this happen?

Response: Due to terrain or other factors (e.g., soil erosion, earthquakes, etc.), cultural layers in one place may change over time (even in the same dynasty).

Line 92: mean of what? current water levels recorded within the well?

Response: It was the overall mean of PUWL. We have clarified it in the revised manuscript.

Line 100: ? about “inrush”

Response: “inrush” is a phenomenon that springs, which used to be drained wells, stream again. It normally implies the rise of PUWL.