Review of "On the suitability of current atmospheric reanalyses for regional warming studies over China" by C. Zhou, Y. He and K. Wang

This paper compares surface air temperatures over China from observations and many atmospheric analyses, and seeks to improve understanding of biases in terms of deficiencies in the representation of forcing factors in the assimilating models used by the analyses. It shows that the effects of homogenising the observations are small compared with the differences between the analyses and the observations. It merits publication, but requires improvement to the presentation and discussion of results.

The comments that follow are not in order of priority.

(1) The language is generally clear, but needs a little sub-editorial refinement.

(2) Page 3, lines 44 to 50. ERA-20CM uses a newer version of the ECMWF model and sea-surface temperature analyses that are more homogeneous over time than ERA-Interim. The comparability of its pattern of trend biases with that of ERA-Interim cannot solely or necessarily be ascribed to its use of an ensemble technique. Note also that ERA-20CM used perturbed sea-surface temperature analyses, and did not include perturbations of the prescribed CMIP5 forcing. As such, classifying its approach as a "perturbed physical ensemble technique" does not seem appropriate.

(3) Page 4, line 58. Satellites should be included in the list.

(4) Page 4, line 66. The models used to produce reanalyses do more than fill gaps in observations. They are important for the quality control and bias adjustment of observations, which is especially important when merging the information provided by many different types of observation.

(5) Page 5, lines 83 to 92. MERRA-2 should be included in this list.

(6) Page 8, line 150, and in later places, including the labelling of figures. It is wrong to label the century-scale analyses that assimilate only surface pressure (and perhaps surface wind) observations as "climate quality" compared with the shorter "NWP" reanalyses that assimilate comprehensive sets of observations. Climate quality is something that has to be demonstrated, and is not just a matter of the type of reanalysis that is carried out. The ERA-Interim and JRA-55 "NWP" reanalyses give the best climate trends for surface air temperature, as the paper shows over China. The century-scale reanalyses still suffer from changes in observations over time, as the number of surface pressure and wind observations has increased enormously over the past one hundred or more years. Also, these reanalyses use sea surface temperature and sea ice analyses that depend on observations that have changed over time. Moreover, their avoidance of upper-air observations means that their climatological states are more subject to model biases than in "NWP" analyses in which observations help determine these states. The term "climate quality" should not be used to categorize some of the reanalyses for which results are presented.

(7) Page 10, lines 184 to 192. It would be helpful for the reader to be informed how many of the 2200 or so stations provide data that are exchanged globally under the auspices of WMO. ERA-Interim and JRA-55 analyse surface air temperature data from those stations for which data are transmitted internationally, and perhaps some additional data to which they have access for early

years, but data from a significant fraction of the 2200 or so stations were probably not used by these reanalyses. It would also be helpful to know whether the observational data used in this study are publicly available to anyone who might wish to carry out such a study, or for use in future reanalyses.

(8) Page 11, line 213 and pages 47 and 48. Table 1 needs some correction and tidying up. ERA-20CM did not use a 4D-VAR assimilation system as it was an ensemble of model runs. It used prescribed sea surface temperature and sea-ice analyses, but they were not produced by 4D-VAR. One column is headed "Related assimilated surface observations", but the entry for ERA-Interim includes reference to upper-air observations, and that for MERRA-2 includes reference to aerosol observations that are not surface ones. It is stated that ERA-Interim assimilated "land surface temperature" data. It did not. It did analyse "surface air temperature data over land", which is not the same variable. As discussed below in comment (10) it is probably better to refer to these data as analysed not assimilated.

(9) Page 11, line 222. Surface pressure observations are not distributed homogeneously in space or (especially) time. Also, sea surface temperature and sea ice analyses are not of homogeneous quality, due to observational changes. See also comment (6).

(10) Pages 12 and 13, lines 246 to 251. The explanation of what ERA-Interim and JRA-55 do could be clearer. A background surface air temperature, at a height of two metres, is produced using a processing of the model-level background forecast with the help of Monin-Obukhov similarity profiles. The observations of surface air temperature are then analysed using a relatively simple analysis scheme. It is best not to use the word assimilated as the two-metre temperatures do not affect the starting atmospheric state for the next background forecast. But they are not simply post-processed products either – in contrast to the products from other reanalyses. Some information is retained (assimilated) in that where appropriate the increments in surface temperature and corresponding ones in relative humidity are used to update soil temperature and humidity, and these do carry over into the next background forecast. It is nevertheless probably better to refer to the observations as analysed rather than assimilated.

(11) Page 18, line 375. The better performance of ERA-Interim and JRA-55 is described as "mainly due to the post-processing of assimilated surface air temperature". If this statement is retained is should read "mainly due to their analysis of surface air temperature data", as discussed in comment (10). The statement is probably correct, but do the authors have evidence that this is the case? Perhaps ERA-Interim and JRA-55 simply have a better background forecast of surface air temperature due to other aspects of their data assimilation system. If the statement is to be retained, it needs to be backed up by showing that the background forecast surface air temperatures from ERA-Interim and JRA-55 are not significantly better than the surface air temperatures from the other reanalyses. In that case, analysing the surface air temperature observations must be the main reason they provide a better product.

(12) Page 19, lines 383-385. It should be noted that CERA-20C used a newer model cycle than ERA-20C, and some problems that were found to affect ERA-20C were fixed in CERA-20C. So CERA-20C's better performance than ERA-20C cannot be ascribed entirely to the use of a coupled forecast model and data assimilation.

(13) Page 19, line 389. The type of analysis presented in section 3.3 needs to be interpreted carefully when it comes to ERA-Interim and JRA-55. This is because their surface air temperature products involve analyses of surface air temperature observations, and values depend on the analysis increments to the background as well as to contributions via the background forecasts from key physical factors that influence surface air temperature. For example, the sentence in lines 475 to 477 on page 23 reads as if the trend biases in surface air temperature have contributions from biases in various physical forcings. But in ERA-Interim and JRA-55 such biases in physical forcing will tend to be counterbalanced by the changes the observations bring to the background forecasts. The balancing will not be perfect, so ERA-Interim and JRA-55 may inherit some of the deficiencies in forcing, but these deficiencies are likely to be much weaker than would be the case if surface air temperature observations had not been analysed.

(14) Pages 21 and 22, lines 443 to 445. Again (see comment (11)) it is asserted that the better performance of ERA-Interim and JRA-55 is due to the assimilation [analysis] of surface air temperature [observations]. This is almost certainly part of the story, but unlikely to be the only reason these two reanalyses perform better than the others. A phrase such as “in part, at least,” is needed after the word “due”.

(15) Page 28, line 28. It is stated that ”only vegetation is included as climatology”. This is wrongly worded. Perhaps the authors mean “vegetation is only included as [a] climatology”. A number of fields other than vegetation are specified climatologically.

(16) page 29, lines 607 to 614. I simply do not understand this paragraph.

(17) Pages 29 to 30, lines 615 to 626. It is misleading to label the models used for the century-scale reanalyses “climate models” and the models used for shorter reanalyses “NWP models”. The same ECMWF models are used for the two types of reanalysis, apart from a tendency for more recent reanalyses to use newer model versions.

(18) Page 30, lines 629. The reference to ERA-20CM is incorrect, as its circulation is not controlled by pressure data. No meteorological observations are assimilated in ERA-20CM.

(19) Page 32, line 682. High temporal resolution in situ and satellite observations of precipitation are available only for recent years, so their use in reanalysis to refine trend estimates will be limited until longer time series of observations have been accumulated.

(19) Page 33, lines 690 to 692. See comment (2) regarding the nature of the perturbations applied in ERA-20CM.

(20) Page 33, line 704. It is not clear why the Argo system is mentioned here, as it is not primarily an observing system for SST and sea ice, and data have been available in substantial numbers for little over a decade, posing a problem for homogeneity.