The paper describes a dataset of eddy covariance data collected at 20 urban sites spread across Australia, Eastern Asia, Europe, and North America. Quality control and gapfilling methods applied to the data were described, as well as extended versions of driving climatic drivers based on reanalysis data harmonized with site data. This dataset is a unique and important contribution to eddy covariance, micrometeorology, and urban climate research.

The argument of using the dataset for climate models is somewhat limited by the fact that climate models mostly cannot take full advantage of datasets like this yet. However, the data presented here being available helps reduce the barriers for models to better take advantage of these data. The bias correction specific to urban environments is an innovative contribution, helping understand the mismatch between in-situ and larger scale representations of these sites; in particular, the hourly corrections are important for analyses of short duration and very localized phenomena, which are not always well represented in these types of datasets.

One of the main limitations related to this dataset is that it was not made available within any of the many research networks that could have offered it a home, but rather it was made available via a general platform—Zenodo. Most notably, FLUXNET offers a clear path to community created datasets, with one such dataset having been described in this very journal: see Delwiche et al. 2021, ESSD. The dataset published in this manner not only makes it harder to find, but also makes it inherently less compatible with/comparable to similar datasets, increasing the difficulty of combined analyses. The datasets being useful for models is a desirable characteristic, as is being easy to use in synthesis-type analyses.

Although the methods presented in the paper for gapfilling the data using different bias-corrections is very interesting, the comparison to alternative methods is not exactly a fair one. The goals for the ERA5 reanalysis dataset and, to some extent, also the WFDE5, are not to be used as single-point data streams, so the usefulness of the comparison of the
methods proposed to these directly should not be interpreted as the proposed method being an improvement over these. The comparison to the FLUXNET2015 method, on the other hand, is very appropriate and led to interesting results. There are clear improvements from the linear method used for FLUXNET2015, which are clearly shown in the description. However, there are also variables/metrics that seem to not do quite as well, e.g., for short-wave radiation and, to some degree, wind variables. Exploring why these variables haven’t seen the same level of improvement would be very interesting indeed, perhaps leading to some form of a hybrid approach.

Finally, some form of summary of the data might have been of interest in a paper like this. For example, a plot showing the carbon or energy fluxes through time for all 20 sites, maybe in a summary form? Or maybe fluxes grouped by climate or latitude or land cover-type? Maybe simple scatter plots of all data (or groups) of fluxes against driving variables (temperature or precipitation). Any of these would help highlight the coverage and usefulness of the dataset, which is quite clearly significant for someone who is familiar with the data, but maybe not as much to someone with more limited exposure to the domain.

A few more specific details come next.

In the QC method presented, step 2 (setting nocturnal shortwave radiation set to zero) might seem like a reasonable approach, clearly no such thing as negative radiation. However, this might also inadvertently introduce a bias to the corrections: the sensor has its accuracy, which might lead to negative radiation values at night, and artificially forcing them to zero will reduce the impact of the sensor’s accuracy limitations to the bias corrections that come later. Perhaps it is indeed better to have this step, but this limitation should at least be clearly stated.

Still in the QC method, now for step 4, would it be necessary to adjust the number of standard deviations and/or window size for each variable? Wind varies in a very different way than atmospheric pressure, for example.

A little more detail could have been given as to how the values for Table 4 were obtained, and also how they are used. The information currently listed in the paper is somewhat limited.

In the hourly and daily corrections, section 2.6.1, step 4 involves extending 1-year worth of data by duplicating that year before and after; wouldn’t this approach artificially create “jumps” towards the beginning and end of the dataset? Wouldn’t it be better to use the actual years before and after the current year whenever available?

The evaluation of the gapfilling presented in section 3 focuses on the differences to ERA5.
However, as mentioned above, the differences to the other methods, in particular the one for FLUXNET2015, would be of wider interest.

The statement regarding daylight savings in section 4.2 (“does not account for day light savings”) is somewhat confusing: does this mean it is always UTC or that the conversion from local time to UTC didn't correct for it so time might be shifted when DST? I believe it is supposed to be the former, however rephrasing that statement might help clarify this point.

Finally, it would have been useful to have a netCDF file with data for all sites for easy access. This would be just a shortcut, but a very handy one.