Review of “Statistical summaries for streamed data from climate simulations: One-pass algorithms (v0.6.2)” by Katherine Grayson, Stephan Thober, Aleksander Lacima-Nadolnik, Ehsan Sharifi, Lorenc Lledo, and Francisco Doblas-Reyes

For consideration in *Geoscientific Model Development*

Recommendation: Major revision

This article describes techniques for on-line statistical analysis of climate model output which can reduce the amount of output data without needing to degrade the spatial resolution of the output. In particular a sophisticated, novel t-digest algorithm is described for on-line generation of percentiles and histograms. The publication is timely given the increasingly large volumes of output being created by increasingly complex and high-resolution numerical models, especially for the emerging class of km-scale models many modeling centers are developing. In particular the quantification of memory savings is a powerful selling point. This paper should be published and GMD is the ideal journal for this purpose. However I was somewhat confused by the discussion of the t-digest method and would like to have some more clarification on a number of points.

I am unfamiliar with some of the terminology. Does the “scale function” k(q) introduced in equation (6) define the “edge values” for each of the bins/clusters? If so, an illustration (showing the domains for each cluster as a function of delta) would be useful. Also, should the same scale function be used for each variable? While having a somewhat equal-spaced set of bins is possibly a good choice for wind speed, this would not be as useful for precipitation since (as is shown in Section 5.3) small values are far more common than large ones, and so the small bin will fill up very quickly.

Also, how the t-digest method goes from its clusters to creating percentiles and histograms is unclear. Could a more concrete demonstration of this be shown? Also, are the numpy percentiles and histograms created by a similar algorithm, or are these the exact percentiles and histograms computed from all of the data without approximation?

I was very glad to see the careful validation of the simpler one-pass algorithms (mean, standard deviation) and the striking demonstration that the errors are at rounding level. I had more trouble understanding Figures 4 and 5. It is not immediately clear that the red and orange dots represent the different locations in panels (b)--(e) without a legend; and the histograms inset in panels (b) and (e) are so small as to be almost impossible to interpret. It would be more useful to overplot these on a larger panel to more directly compare the two. I also don’t quite understand the main result in panels (b) and (e): the shaded area is the typical wind turbine operating tolerance for these locations, and each dot is another percentile larger (first dot is 1st percentile, last is 99th)? Overall I think splitting out the panels and/or insets into different figures, providing better labels and annotations, and enlarging some plots would make it much easier to read and understand.

The discussion of convergence of climatologies in Section 6 is interesting and does get at the vital scientific question of how much data is needed to create a useful climatology. However, given that the one-pass algorithms create nearly-exact statistics (for means and standard deviations) or up to the cluster's sampling uncertainty (t-digest) *for the model run segment being considered*; constructing a climatology for bias correction or other purposes would be a downstream step that would presumably mean aggregating statistics over a large number of segments and/or individual simulations. Would not this analysis then say how long the simulation needs to be to create an accurate climatology, and not have much effect on the algorithms of this paper?

Minor comments:

- There is precedent for on-line calculations of statistics such as means, extrema, standard deviations and so on. See the "data reduction" techniques here, for instance:

https://github.com/NOAA-GFDL/FMS/blob/main/diag_manager/diag_yaml_format.md
https://github.com/NOAA-GFDL/FMS/blob/main/docs/diag_table.md

- Line 171: Should "extremely values" be "extremely small values"?