This study, “GRUN: An observations-based global gridded runoff dataset from 1902 to 2014,” presents a new global dataset of monthly runoff fields from 1902 to 2014, as the title says. The runoff fields are produced by upscaling the observed discharge in small river basins (smaller than a grid cell) using the time series of monthly precipitation and temperature, and their time lag operators.

I find the data product timely (actually, well overdue) and very relevant for the evaluation of global scale land surface and hydrological models. The paper is well-written with clear explanations of the methods and data. Additionally, it includes some suitable examples/case studies, and shows that the data is able to capture the main features discussed therein.

I, however, have several concerns and suggestions (see below) which should be addressed before the manuscript is published. The additional analysis and discussion will strengthen the data product.

Data and Methods:

- The area of basins selected for fitting the random forest model is well below the size of a grid cell. That means that the discharge in these basins might be driven by sub-grid precipitation (and climate) that is not reflected at the grid-scale. As a sanity check, one should compare the observed discharge against the gridded precipitation. Perhaps, even filter out the basins where the annual discharge is higher than annual precipitation? If not, the results should show the distribution of discharge over precipitations in basins with observed discharge.
- The study would benefit by including a figure showing/comparing the distribution of data values in whole and training dataset. It is known that random forest cannot predict outside the range of data values used in training.
- There is no explanation on why the predictors included precipitation and temperature only, especially considering that the co-authors have produced a long-time series of terrestrial water storage variations at the global scale. In addition, several other factors such as vegetation properties, topographic indices, soil properties, and so on contain useful information on runoff generation mechanisms. Previously, these variables have been shown to include information on baseflow characteristics globally (see H. Beck et al, 2013 (WRR), and 2015 (JHM)).

Data Evaluation:

- The validation of the predicted runoff against GRDC should include the comparison of mean flow in addition to the time series (and metrics based on it). The mean discharge and runoff over the basin should be equal if there is no long-term change in river storage. If the mean runoff from GRUN is similar to GRDC observed discharge, one can infer that the poorer performance in the time series is due to exclusion of river routing.
- Is it possible that the poorer performance in the humid basin is related to lack of data values similar to those in Amazon and Congo in the training dataset? Additionally, in the comparison of global volumes, the GRUN also falls in the lower range. I am curious if this underprediction is coming from the lower values in the tropics which dominate the contribution to global runoff. This can be checked by having a grid-to-grid comparison (something like hexbins in Figure 3) against previous products or model simulations. If the underprediction is clear, it makes me wonder if predicting the runoff ratio (runoff to precipitation) would be beneficial. A discussion on this would help the future studies on a similar topic.
- The GRDC discharge observations are likely to be affected by anthropogenic uses, as the manuscript rightly points out several times. Therefore, wouldn't it make sense to evaluate the GRUN product against:
    - previous mean total runoff from Beck et al., 2015 which uses fewer observation stations but more predictors (also based on neural network),
    - mean ET from, say, LandFlux-EVAL, part of which is coming from satellite observations?

References:

Beck, Hylke E., et al. "Global patterns in base flow index and recession based on streamflow observations from 3394 catchments." *Water Resources Research* 49.12 (2013): 7843-7863.

Beck, Hylke E., Ad De Roo, and Albert IJM van Dijk. "Global maps of streamflow characteristics based on observations from several thousand catchments." *Journal of Hydrometeorology* 16.4 (2015): 1478-1501.