"The CAMELS data set: catchment attributes and meteorology for large-sample studies" *by* Nans Addor et al.

Response to Anonymous Referee #2 by Nans Addor et al.

We thank the reviewer for her/his helpful and positive comments. Before we address them, we would like to stress that since our manuscript was submitted, we added a new set of attributes to CAMELS. We extracted geological characteristics from the GLiM and GLHYMPS data sets and produced catchment-scale averages for the 671 catchments. We think that this addition enables a more complete description of the landscape of the CAMELS catchments, and that it will provide useful insights into hydrological processes. These new attributes are now introduced and discussed in the new Section 8, Figure 7 and Table 6 – see the end of this document.

We are also pleased to report that the catchments attributes are now freely available online: https://dx.doi.org/10.5065/D6G73C3Q

This manuscript presents a nice extension of Newman et al., 2015b on a catchment dataset across the U.S. My group have used the Newman dataset in our research new process understanding, so I am happy to see the extension of it. I feel this manuscript can be published at HESS after addressing the following comments.

1. MOPEX dataset is a very good example of such a catchment dataset. It has been extensively used by the hydrology and land surface modeling communities leading to at least over 100 journal articles. It'd be interesting to see a more in-depth or more detailed comparison between MOPEX and the new dataset here, i.e., a table would be nice.

We rephrased and extended the end of Section 9. Comparison with the MOPEX data set (previously Section 8):

"Overall, the data used for CAMELS are more recent than those used for MOPEX. The period covered by hydro-meteorological times series is 1948-2003 for MOPEX and 1980-2015 for CAMELS, so given the fast rate of human development and the impacts caused by climate change, CAMELS provides a more current picture of hydrological processes in the United States. Further, CAMELS leverages new data sets, which were not available when the MOPEX data were released, for instance to characterize soils (Pelletier et al., 2016) and geology: GLiM (Hartmann and Moosdorf, 2012) and GLHYMPS (Glesson and et al., 2014). And importantly, data used for CAMELS are not only more recent, but also tend to be better documented. A clear example is that CAMELS meteorological time series come from three widely-used gridded data sets (Daymet, Maurer and NLDAS), while for MOPEX, station measurements were aggregated to provide catchment-scale estimates (Schaake et al, 2006)".

We also produced a new table outlining the main differences and similarities between MOPEX and CAMELS (now Table 8):

Table 8: Main differences and similarities between MOPEX and CAMELS

	MOPEX	**CAMELS**
Number of catchments and spatial repartition	438 catchments, principally from the Eastern half of the CONUS and with an under-representation of the Rocky Mountains	671 catchments, with a relatively even repartition over the CONUS
Catchments in common to MOPEX and CAMELS	52	
Period covered by the hydro-meteorological time series	1948 - 2003	1980 - 2015
Minimal anthropogenic influence and long streamflow records assessed by:	Wallis et al. (1991) and the hydro-climatic data network (HCDN, Slack and Landwehr, 1992)	Updated version of the HCDN (HCDN-2009, Lins, 2012)
Raingauge density used as catchment selection criterion	Yes	No
Meteorological data	Point observations from National Climate Data Center daily Cooperative Observer Network (COOP) and SNOTEL stations. Missing observations were filled. Long-term precipitation averages computed using 1961-1990 PRISM data	Gridded data from Daymet (Thornton et al., 2012), Maurer (Maurer et al., 2002) and NLDAS (Xia et al., 2012)
Streamflow data	USGS streamflow measurements	
Soil data	SATASGO (Miller and White, 1998)	STATSGO (Miller and White, 1998) and Pelletier et al. (2016)
Vegetation data	North America Land Data Assimilation (NLDAS)	MODIS imagery
Geology data	Not available	GLiM (Hartmann and Moosdorf, 2012) and GLHYMPS (Glesson and et al., 2014)
Reference papers	Duan et al. (2006) and Schaake et al. (2006)	Newman et al. (2015) and this paper

2. I'd like to see more (perhaps quantitative) discussion on whether and how the catchments included in this dataset are free of human impacts. One good example is Wang and Hejazi, 2011.

We now mention in the first sentence of the abstract that the catchments are minimally impacted by human activities:

"We present a new data set of attributes for 671 catchments in the contiguous USA (CONUS) minimally impacted by human activities."

We now also stress this on P2, L27-30, and provide a reference to the section of the N15 paper explaining how the catchments were selected, and in particular, how the impacts of human activities were assessed:

"All those catchments have 20 years of continuous discharge record from 1990 to 2009 and are minimally impacted by human activities (see Section 2.1 in Newman et al., 2015)."

Geological characteristics – new section and associated table and figure

8 Geology

8.1 Data and methods

We used two complementary global sets to characterize the geology of each catchment. The first data set is the Global Lithological Map (GLiM) by Hartmann and Moosdorf (2012). GLiM synthesizes lithological data from 92 regional maps spread across the globe. The spatial resolution is remarkable, as GLiM relies on ~1.2 million polygons to discretize the Earth surface. Three levels of details are available. In this study, we focus on the first level, while the two other levels provide further details that could be processed at a later stage. The first level differentiates between 16 lithological classes (see the list of classes in the legend of Figure 7). We determined the contribution of each lithological classes to the area of each catchment, and recorded the first and second most frequent class within the catchment, as well as the fraction of the catchment they cover. The class "carbonate sedimentary rocks" is particularly relevant from a hydrological perspective (it designates areas likely to host karst systems), we hence also recorded the fraction of each catchment associated with this class. Finally, note that although a 0.5°-gridded version of GLiM is available, we used the more detailed polygon-based version for this study.

The second data set we used to characterize catchment geology is the GLobal HYdrogeology MaPS (GLHYMPS) of the subsurface permeability and porosity by Gleeson et al. (2014). GLHYMPS is based on GLiM spatial polygons, so its level of spatial details is equally high. Gleeson et al. (2014) principally relied on GLiM lithologic classes to derive quantitative estimates of two key characteristics of the geologic units below soil horizons: porosity and permeability (i.e., the ease of fluid flow through porous rocks and soils). For CAMELS, we produced catchment-averages of these two variables, the contribution of each spatial polygon being weighted by the fraction of catchment it covers. The arithmetic mean was used for porosity, but for permeability, we followed Gleeson et al. (2011) and used the geometric mean instead. The geological attributes are summarized in Table 6.

A clear advantage of these high-resolution global lithological maps is that they can be used to extract catchment-scale attributes for diverse parts of the globe. Yet, data quality is spatially variable and caveats of the GLiM and GLYHMPS (outlined in the Section 3 of Gleeson et al., 2014) should be kept in mind. In particular, there are unrealistic spatial discontinuities coinciding with jurisdictional boundaries in GLiM maps, which by construction also affect GLYHMPS maps (for instance in the region of North and South Dakota).

8.2 Results and discussion

The four most frequent dominant geological classes in CAMELS catchments are siliciclastic sedimentary rocks (34% of the catchments), unconsolidated sediments (19%), metamorphics (16%) and carbonate sedimentary rocks (12%). Unconsolidated sediments dominate in catchments along the Gulf Coast and along the southern to middle Atlantic Coast (Figure 7a). In those catchments, both the subsurface porosity (Figure 7f) and permeability (Figure 7g) are high. The Pacific Coast and the region north of the Appalachian Mountains features catchments rich in siliciclastic sedimentary rocks, leading to a comparatively low subsurface permeability. To the south of the Appalachian Mountains, metamorphic rocks are dominant, resulting in a particularly low subsurface porosity. Finally, the catchments with the highest proportion of carbonate sedimentary rocks are principally located in Central-Western Texas, in the region stretching from Lake Michigan to and including Missouri (Figures 7a and 7f) and to some extent in the Appalachian Mountains (Figure 7b). In addition to these three main regions, there are also isolated catchments with a high proportion of carbonate rocks, for instance in Florida, Nevada and Vermont. The subsurface permeability of those catchments is high. Overall, in 18% of the CAMELS catchments, there is only one GLiM lithological type (Figure 7d), while in 11% of the catchments, the dominant geological class accounts for less than 50% of the catchment area (Figure 7e).

New references

Gleeson, T., Moosdorf, N., Hartmann, J., & van Beek, L. P. H. (2014). A glimpse beneath earth's surface: GLobal HYdrogeology MaPS (GLHYMPS) of permeability and porosity. *Geophysical Research Letters*, *41*, 3891–3898. https://doi.org/10.1002/2014GL059856

Hartmann, J., & Moosdorf, N. (2012). The new global lithological map database GLiM: A representation of rock properties at the Earth surface. *Geochemistry, Geophysics, Geosystems*, *13*(12), 1–37. https://doi.org/10.1029/2012GC004370

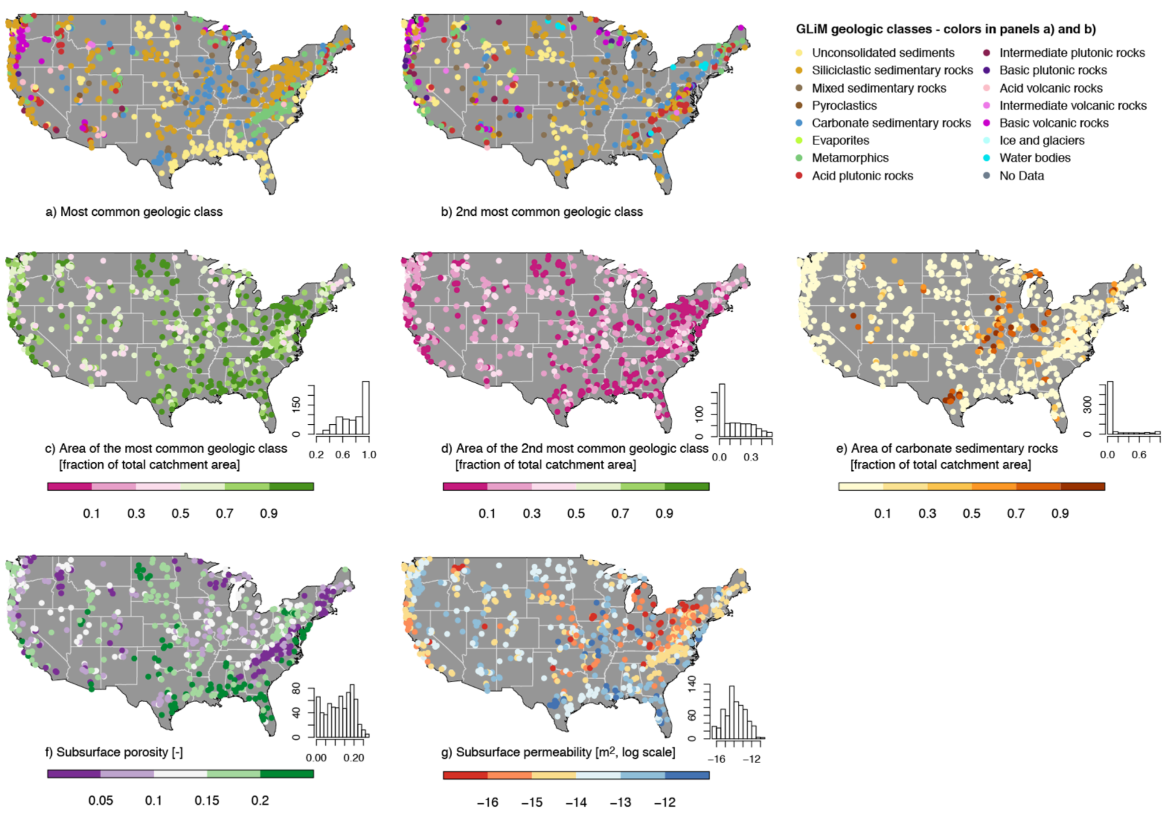

Figure 7: Maps of the geological characteristics over the CONUS. The histograms indicate the number of catchments (out of 671) in each bin.

Table 6: Geological characteristics.

Attribute	*Description*	*Unit*	*Data source*	*References*
glim_1st_class	most common geologic class in the catchment	qualitative	GLiM	Hartmann and Moosdorf (2012)
glim_1st_frac	fraction of the catchment area associated with its most common geologic class	-	GLiM	Hartmann and Moosdorf (2012)
glim_2nd_class	2nd most common geologic class in the catchment	qualitative	GLiM	Hartmann and Moosdorf (2012)
glim_2nd_frac	fraction of the catchment area associated with its 2nd most common geologic class	-	GLiM	Hartmann and Moosdorf (2012)
glim_carb_rocks_frac	fraction of the catchment area characterized as "Carbonate sedimentary rocks"	-	GLiM	Hartmann and Moosdorf (2012)
glhymps_porosity	subsurface porosity	-	GLHYMPS	Gleeson et al. (2014)
glhymps_permeability	subsurface permeability (log10)	m2	GLHYMPS	Gleeson et al. (2014)