

Reviewer 1

Sterle et al. present a compiled novel dataset of water quality solutes and atmospheric deposition inputs for the CAMELS catchments. Their work augmenting existing and widely used CAMELS datasets is needed for further research in analyzing spatial and temporal water quality trends in minimally impacted watersheds. Existing papers have done large-scale water quality analyses, but few have provided open-access datasets and the breadth of solutes.

At its core, this is a data paper. As such, I think the methods need to be expanded. My comments primarily pertain to the data and methods, as I see that this is the paper's novelty.

The CAMELS dataset is widely used, and the addition of water chemistry provides the opportunity for analysis. However, I think the dataset could be improved substantially. This paper by Sterle et al. provides an excellent resource for the community.

Authors response: We thank the reviewer for this assessment and agree that this database presents a great resource for the community and has already found application in research resulting in publications. We are looking forward to seeing how the revised version of this manuscript will be received.

Reviewer comment: In the paper, the dataset is stated to end in 2014 (see Data Comments below because I'm not sure if this is accurate). However, in many cases, solute and discharge data are available until the at least end of the NADP reporting period. Soon 2014 will be 10-years ago, and I worry the data will not be quickly obsolete and not used to its fullest capacity. The value of this dataset would be exponential if the authors were able to harmonize data from various agencies.

Authors response: We appreciate this comment and agree. We have included data until (and including) the year 2018. We have updated the description where the data range is mentioned.

For example in the introduction, where we state: "Stream data goes back to the early 1900 in some cases and we integrated these data into the dataset. However, since CAMELS covers the timeframe from 1980 to 2014, we showcase stream data starting 1980 and have included the earliest data on atmospheric deposition data by the National Atmospheric Deposition Program (NADP), available since 1985."

Reviewer comment: The USGS has developed methods to generate the longest time series possible for watersheds. However, I do not see these methodologies applied to these CAMELS watersheds.

1. The following USGS report looked at the statistical differences between similar solutes. They were then able to merge solutes that were registered under different parameter codes but were statistically similar, allowing for a richer dataset.
 2. Secondly, the USGS report also has a methodology for pairing gauges, and sampling locations are not co-located.
 3. With the dataRetrieval package in R you can pull USGS, state, tribal, and NGO water quality measurements. This can potentially expand the dataset. However a lot of manual data cleaning is required because of the poor metadata for non-USGS agencies (ex. nitrate can be reported as N or NO₃, and not explicitly stated in the metadata) (Sprague et al. 2017). With the data harmonization from multiple sources, I would expect a section on the methodology of data cleaning.
- Oelsner, G. P., Sprague, L. A., Murphy, J. C., Zuellig, R. E., Johnson, H. M., Ryberg, K. R., Falcone, J. A., Stets, E. G., Vecchia, A. V., Riskin, M. L., Cicco, L. A. D., Mills, T. J., & Farmer,

W. H. (2017). Water-Quality trends in the nation's rivers and streams , 1972 – 2012 — Data Preparation, Statistical Methods, and Trend Results Scientific Investigations Report 2017 – 5006. U.S. Geological Survey Scientific Investigations Report 2017–5006. <https://doi.org/10.3133/sir20175006>

Authors response: We agree with the reviewer that excellent resources exist for expanding data sets, including methodologies that significantly improve common pitfalls such as those mentioned above. We also agree that we did not offer this context for the reader in this manuscript and made changes accordingly. We also were remiss in not citing this important paper and have rectified this issue.

Our objective for this relational database is to supplement water quality and atmospheric data in conjunction with the catchment attributes of the original CAMELS dataset; therefore, for this publication, we did refrain from adding sites and additional harmonization. However, this will offer opportunities for further expansion and analysis of this database going forward.

However, we agree on merging solutes when possible and to clarify what parameter codes are included (Table 1). For example for chloride, sulfate and a nitrate, data from unfiltered samples can be used as these ions do not have a particle form and are dissolved independent of filtering. We have included this in the table.

Reviewer comment: Figures capture the various hydrological/biogeochemical metrics for select solutes in section 4. However, as a user, I find it challenging to evaluate whether the data would be sufficient for my use case. The authors have included some analysis of data coverage in section 3.1, however, since the strength of this paper lies in the data there could be more information to help users evaluate whether this dataset suits their needs.

Authors response: We agree with the reviewer that we chose a somewhat unconventional way of presenting data and possible applications and that both can be improved. We therefore have added new sections (3.2, 3.3, and 3.4) and comprehensive maps for all solutes for the reader to be able to evaluate the dataset. This includes information on the consistency of sampling across time, seasons, and discharge and the co-occurrence of samples. Additionally, we show maps on data availability over the time frame that CAMELS and stream solute data are available (since 1985). These include the number of samples and length of records, mean concentrations and coefficient of variation, atmospheric deposition for all variables, % range of flow duration curves, and maps of the 66 percentile of flow duration curve. Further, in supplementary materials we provide all data (since beginning of the record).

The example applications are intentionally included (with the removal of the previous example 2) to inspire use of this dataset by a broad range of disciplines that otherwise might not consider using these data. Especially interdisciplinary fields, such as Ecology and Critical Zone science that are in need of data that capture the system from “tops of the tree canopy to the bedrock and groundwater”, to investigate how atmospheric drivers interact with the critical zone to impact dissipative products e.g. as stream waters. Our examples showcase an array of example applications to facilitate these types of data-base approaches. As such, our manuscript is a hybrid of a paper that accompanies the dataset, but also an outreach tool for communities to use. We added this framing to L. 24-25” To motivate use by the larger scientific community across a variety of disciplines, we exemplify the potential use of this dataset for three analyses”, and L. 200-201” To motivate adoption of this dataset by the larger scientific community, especially in fields that span multiple disciplines, we intentionally decided to offer examples in three themes:”.

Reviewer comment: There should be summary figures for all solutes, so users can adequately assess whether the dataset is appropriate for their use. I think the paper would benefit tremendously if the dataset had more metadata and signatures/summary statistics. Specifically, in Addor et al. (2017), the authors

summarized many indices and described the indices in great detail (see Table 2 and Table 3 in Addor et al). I suggest summarizing information about (1) missing data/data gaps, years of continuous data, (2) low/high flow distribution, (3) FDC the WQ spans (figure 7), (4) seasonality of hydrology and solutes, and other metrics that the authors deem useful.

Authors response: We appreciate the reviewers comments that are geared towards making this database usable. In addition to the comprehensive maps described above, we now provide requested summary statistics in new tables (Table 3) that we integrated into the main manuscript and supplementary materials.

These comments pertain to the files on the google drive.

Reviewer comment: Data provided has inconsistencies in the way the date is reported. Example from camel_chem_v3.

- Gauge ID 14309500: sample_timestamp reads 8/15/67 15:00 whereas other dates are in 2010-12-11 format.
- Gauge ID 6447000: sample_start_dt appear to indicate that the data from 1950s. If so, this falls outside of the time period listed in the paper.

Authors response: We thank the reviewer for this comment and have made the changes as requested.

Reviewer comment: It appears that the data available in camel_chem_q_v1 ends in 2018. Please update the manuscript with the correct dates if this data is available.

Authors response: We clarified in the MS that the data of the database is updated regularly to include more recent data. However, for the Figures and summary statistics offered in this publication we use data until and including the year 2018.

Reviewer comment: The original CAMELS dataset provides Shapefiles, however, to allow for seamless merging of data, the header used to identify which column the watershed IDs are the same as the name used by the original CAMELS dataset.

Authors response: we have ensured seamless merging of our data with the original CAMELS dataset. We have used the same column name for the watershed IDs in our dataset as the name used by the original CAMELS dataset, which is "gauge_id". This should allow for easy integration and analysis of our data with the original dataset.

Reviewer comment: The dataset provided should be able to stand on its own without needing the other CAMELS dataset. The watershed metadata should thus be included (area, outlet latitude and longitude, USGS gauge number with leading 0s).

Authors response: Thank you for your feedback on our dataset. We agree that this is an important consideration and have made efforts to ensure that our dataset contains all necessary metadata.

In response to your specific comment about the inclusion of watershed metadata such as area, outlet latitude and longitude, we would like to confirm that these variables are indeed included in our dataset. The columns

"area_gages2", "gauge_lat", and "gauge_lon" contain the area, outlet latitude and longitude information respectively.

Individual scientific questions/issues ("specific comments"),

Reviewer comment: Line 84-87: Authors state that they have WQ data from 1980-2014. Later the authors state that the data is "for the same time period" as NADP data (1985-2019). Dates should be consistent.

Authors response: we agree and have revised this throughout the MS. The data we showcase ranges from 1980 to 2018. However, the NADP data is only available since 1985, thus we clarified "(...) since CAMELS covers the timeframe from 1980 to 2014, we showcase stream data starting 1980 and included the earliest atmospheric deposition data from a product developed by the National Atmospheric Deposition Program (NADP, available since 1985)".

Reviewer comment: Line 103: Section 2.2 should be written in a high-level abstract way. I find it unclear how these frameworks are applied to this data in the way it is currently written. I would be better to see more specificity and names of the database (ex. NADP, NWIS).

Authors response: Thank you for this suggestion. We have rewritten the section on our Data Harvesting and Integration approach (Section 2.2) to more directly speak to why and how the Extraction, Transformation and Loading framework was applied to generate the CAMELS-Chem relational database. We also moved the order of this section to follow a presentation of the sourcing and characteristics of the specific data sets integrated within CAMELS-Chem (i.e., NWIS, NADP)

Reviewer comment: Line 125: It is unclear whether "daily average discharge" means a continuous dataset or just discharge measurements for the data that there are solutes. The dataset provided suggests the latter, but I think there would be value in providing the daily discharge timeseries for the same timespan of the solute data.

Authors response: The daily discharge we reference are the USGS average discharge values. The instantaneous data reference hourly or finer time scales. The continuous daily discharge data are part of the original CAMELS datasets, and are therefore easily available.

Reviewer comment: Line 135: In Table 2, deposition units are reported as mg/L. However, NADP reports their deposition in both concentration and kg per hectare. Are the units in Table 2 a mistake? If not, can you add some detail on the methods used to convert concentration to an area normalized load?

Authors response: we agree with the reviewer and have reported both the concentration data and the deposition units as kg/ka in Table 4 .

Reviewer comment: Table 1: Consider added the NWIS parameter code. For example, is "Nitrate, water filtered" the nitrate plus nitrite (00631) or just nitrate (00618)? There are many parameters for slightly similar solutes and it would help with reproducibility if the parameters codes were included. Also, consider listing the difference between pH in the field and pH in the lab for users.

Authors response: We agree with this suggestion and have included the parameter codes to the table. Further, we were able to increase the number of observations for Cl, nitrate and sulfate by adding unfiltered samples as these are species where filtration does not impact concentration. To help the user we added all of this information, including column abbreviations, to the table.

Reviewer comment: Table 1: Also consider adding more detail to units. For example, is nitrate mg-NO₃/L or mg-N/L.

Authors response: We agree and added this information for nitrate (mg/L N) to the Table. For all other solutes the units are not ambiguous but we provide the parameter codes as requested.

Reviewer comment: Line 204: EPA link is broken. I have had many issues with direct links where they are archived and become a dead end. I highly encourage the authors to find a paper with a DOI to support this sentence. As a starting point, you can consider:

- Baumgardner, R. E., Lavery, T. F., Rogers, C. M., & Isil, S. S. (2002). Estimates of the Atmospheric Deposition of Sulfur and Nitrogen Species: Clean Air Status and Trends Network, 1990–2000. In *Environmental Science & Technology* (Vol. 36, Issue 12, pp. 2614–2629). <https://doi.org/10.1021/es011146g>
- Lloret, J., & Valiela, I. (2016). Unprecedented decrease in deposition of nitrogen oxides over North America: the relative effects of emission controls and prevailing air-mass trajectories. *Biogeochemistry*, 129(1-2), 165–180.

Authors response: We agree with the reviewer and have replaced the weblink with references as suggested.

Technical corrections ("technical corrections": typing errors, etc.).

Reviewer comment: Subscripts for solutes should be consistent throughout the manuscript.

Authors response: We agree and have fixed the subscripts for NH₄, NO₃, SO₄ etc throughout the MS and supplementary materials.

Reviewer comment: Table 3 formatting caused solutes to be cut off.

Authors response: We have reformatted the table.

Reviewer comment: Line 18 and 149-151: 18 solutes listed in the abstract, 17 listed in Table 1 and in Line 149, and 16 listed in the text. Make them all consistent.

Authors response: We agree that this is confusing and stemmed from the omission of temperature in the original table. We have rectified the issue and revised with the correct # of solutes (18) throughout.

Reviewer comment: Line 57: Remove (?).

Authors response: Revised as suggested.

Reviewer comment: Line 60: Remove "CITE"

Authors response: Revised as suggested.

Reviewer comment: Line 168, 174 and others: When referencing figure (ex Figure 2), please add the panel letters (a,b,c, etc.).

Authors response: We agree and added the panel letters as requested for all figures throughout the ms.

Reviewer comment: Figure 2: Panel a, why does daily average discharge only have 393 watersheds while the original CAMELS dataset uses USGS discharge in the original CAMELS dataset?

Authors response: we agree that this is confusing and we clarified in the ms. The discrepancy is coming from the fact that not all gauges have paired Q and water quality data available for the time frame we report in this paper. For the timeframe of the dataset (1985-2018) and of the original 671 CAMELS watersheds, 506 watersheds returned instantaneous or daily mean discharge data and 488 returned both discharge and water quality data.

Reviewer comment: Figure 6: Regarding NO₃, if arid and humid sites are a subset of all sites, I am unsure how the slope for all sites can be larger than both arid and humid.

Authors response: We agree with the reviewers and found that this example does not add substantially to the Manuscript. We decided to remove this figure from the MS.

Reviewer 2:

This paper describes efforts to amend the known and quite successful US-wide CAMELS dataset with hydrochemical and deposition data. This effort is overdue and would greatly enhance the usage of the original CAMELS data as well.

Authors response: We thank the reviewer for this assessment and agree that this database presents a great resource for the community and appreciate the comments that are geared to making this a better and more user-inspired resource. We are looking forward to seeing how the revised version of this manuscript will be received.

Reviewer comment: Having stated this I have to admit that the manuscript is not convincing to me. It fails to state what data (e.g., constituent codes in the USGS data) was used in what exact way (missing description of data evaluation, conversion, filtering). It visualizes data coverage but does not state number of observations and number of stations in a consistent way. So overall the reader is left rather unclear about the whole data handling process and the outcome.

Authors response: We agree that the user will need this information and have added the parameter codes to the table and also offered more context for data evaluation and processing. We further clarified how many gauges are included as “n” in all figures and offered additional maps for all solutes (Figure 2, 3, 4, 7 and 9).

Reviewer comment: While I like the idea of three examples this data could be used for, this leaves me rather puzzled. This is submitted as a research paper but does not really come up with research. For me this manuscript would rather fit the purpose of ESSD (Environmental Science System Data) as a dataset description than HESS as a research paper. For the latter it would have an appropriate structure and depth but still would need to acknowledge the detailed comments below.

Authors response: We agree with the reviewer that we chose a somewhat unconventional way of presenting data and possible applications that bring us to a gray area between a data paper and a research paper. Working in an interdisciplinary group with domain and data scientists has repeatedly confronted us with the disconnect between domain knowledge across multiple disciplines and database science. We therefore made the intentional decision to go beyond a traditional data paper by offering examples that demystifies the application of such a relational database especially for interdisciplinary fields such as ecology and critical zone community. These examples help illuminate how these data can be used and therefore facilitate the integration of ideas and data from different disciplines. This is essential for generating new ideas as it is often challenging to see connections across disciplinary boundaries.

Especially interdisciplinary fields, such as Ecology and Critical Zone science are in need of data that capture the system from “tops of the tree canopy to the bedrock and ground water”, to investigate how atmospheric drivers interact with the critical zone to impact dissipative products e.g. as stream waters. Our examples showcase an array of example applications to facilitate these types of data-based approaches.

However, we agree that we failed to offer enough information that a typical data publication has to offer and revised our presentation accordingly for user evaluation. We have added new sections of text (3.2, 3.3, and 3.4) comprehensive maps for all solutes for the reader to be able to evaluate the dataset. This includes maps on data availability (number of samples and length of records, Fig. 2), mean concentrations and coefficient of variation (Fig. 3), atmospheric deposition for all variables (Fig 4), % range of flow duration curves (Fig. 7), and maps of the 66 percentile of flow duration curves (Fig 9), as well as additional summary statistics.

Reviewer comment: The abstract needs to transport the content and motivation in a better way. It misses quite some information e.g. resolution of provided data. It is not clear what is really provided in this new data set and what is already available from CAMELS.

Line 18: The linkage to the original US CAMELS dataset remains unclear in the abstract. Is this an addition to the original one as induced by title and text here or is this something completely new.

Authors response: We agree with the reviewer that we did not provide enough background and framing for the reader in the abstract and appreciate this comment. Indeed, CAMELS-Chem is both an amendment of existing CAMELS dataset and also something new, as the relational nature of this dataset allows a completely new set of applications. The relevant section in the abstract now reads:”Large sample datasets are transforming hypothesis testing and model fidelity in the catchment sciences, but few large stream water chemistry datasets exist with complementary streamflow, meteorology, and catchment physiographic attributes. The existing CAMELS (Catchment Attributes and Meteorology for Large-sample Studies) dataset includes data on topography, climate, streamflow, land cover, soil, and geology across the continental U.S. With CAMELS-chem, we now pair these existing catchment attribute data with atmospheric deposition data from the National Atmospheric Deposition Program and water chemistry data and instantaneous discharge from U.S. Geological Survey over the period from 1980 through 2018 (...)”

Reviewer comment: Line 24: This is odd. Either give the exact number or >. Rounding to something easier to read would probably best.

Authors response: We agree and have removed “>” which was an editing error.

Reviewer comment: From my point of view the introduction should state some examples of existing water quality databases. There are recent advances here such as: GRQA: Global River Water Quality Archive (Virro et al., ESSD), GLORICH (Hartmann, J., Lauerwald, R., and Moosdorf, N.: A Brief Overview of the GLOBAL RIVER Chemistry Database, GLORICH, *Proced. Earth Plan. Sc.*, 10, 23–27, <https://doi.org/10.1016/J.PROEPS.2014.08.005>, 2014. a, b) or QUADICA (<https://doi.org/10.5194/essd-2022-6>)

Authors response: We were indeed remiss in referencing these datasets and have rectified this oversight in the current version. We have added reference of these datasets to L. 43 ff. Thai section now reads ”The Global River Water Quality archive (GRQA) and GLObal RIVER Chemistry Database (GLORICH) have offered opportunities for water quality analyses across time and land scale (Hartmann et al. 2014, Virro et al. 2021) and the Catchment Attributes for Large-Sample Studies (CAMELS, Newman et al., 2014) compile high quality streamflow measurement in 671 unimpaired catchments, as well as climate forcing datasets

(e.g. daily precipitation and temperature) and physiographic properties (e.g. land cover, topography, etc., Addor et al., 2017). CAMELS has seen widespread adoption by the hydrological community as a benchmarking tool for hydrological models (Melsen et al., 2018; Mizukami et al., 2019; Pool et al., 2019), in the development of hydrological signatures and new information theory-based approaches, and the application of novel machine learning tools (Kratzert et al., 2018).” The combination of catchment attributes and generation of matching datasets on stream water chemistry has recently been developed for Germany (Ebeling et al. 2022), however, for the CONUS this approach has not not seen as much development. Furthermore, atmospheric deposition data is typically not included in such data sets, despite the significant impact of atmospheric contribution to stream chemistry.”

Reviewer comment: Line 31: There are quite a number of examples of nation-wide to continental scale water quality studies using more than a single catchment. I disagree here that availability of datasets does not go hand in hand with usage of this.

Authors response: We actually meant to emphasize the challenge of analyzing processes across bio-geo-hydrosphere where single sites often have data, but might not have the breadth for regional, continental or global studies. Essentially it is the problem of regional and continental patterns vs. site specific process identification and “uniqueness of place”, however, we realized we did not frame this well. We revised accordingly, the first sentence now reads: Earth surface processes include coupled and complex processes that involve atmosphere, biosphere, lithosphere and hydrosphere, however, tracking these important processes across time, space and disciplines remains a challenge that is, amongst others, related to data availability and data connectedness.”

Reviewer comment: Line 35: Check citation formatting here.

Authors response: revised as requested

Reviewer comment: Line 35: Can you be more specific on the "issues" mentioned here?

Authors response: We clarified that the uniqueness of place issue is the disconnect between catchment specific processes vs larger scale patterns. The section now reads: ”One of the key advantages of aggregating and harmonizing data into larger sample size datasets is to test how model hypotheses reproduce observed behavior across variable conditions and sites to reduce the uniqueness of place-based issues that plague catchment science (Gupta et al., 2014, Hubbard et al., 2020), that is, individual (unique) catchments might not be generalizable to explain a larger scale pattern and vice versa”.

Reviewer comment: Line 42: Year in citation missing.

Authors response: Fixed as requested.

Reviewer comment: Line 43: Use the complete name of CAMELS here.

Authors response: We agree and introduced the abbreviation at first use here.

Reviewer comment: Line 56: Check the question mark here.

Authors response: This was an editing error that we have removed.

Reviewer comment: Line 60: Check citation here.

Authors response: Fixed as requested.

Reviewer comment: Line 83: I think you should better introduce the idea to provide atmospheric deposition data.

Authors response: We agree and have provided framing in the introduction: L. 30ff “Earth surface processes include coupled and complex processes that involve atmosphere, biosphere, lithosphere and hydrosphere, however, tracking these important processes across time, space and disciplines remains a challenge that is, amongst others, related to data availability and connectedness” and L. 50 ff: “The combination of catchment attributes of matching datasets on stream water chemistry has recently been developed for Germany (Ebeling et al. 2022), however, for the CONUS this approach has not not seen as much development. Furthermore, atmospheric deposition data is typically not included in such data sets, despite the significant impact of atmospheric contribution to stream chemistry”.

Reviewer comment: Line 86: Can you specify why to stop in 2014? Because the original CAMELS data also have that time frame?

Authors response: We agree that this requires clarification. The CAMELS catchment attributes timeframe spans 1980 to 2014, but our stream data goes back to the early 1900 in some cases. The NADP data starts from 1985. We include all data in the dataset and feature it in supplementary materials, however in the main body we showcase data for the time where CAMELS and stream data is available (since 1980). We clarified that in the introduction and also methods. Further, the database can be readily updated with more recent data, but we chose a cutoff for the data we showcase in the publication. We now show all data until and including the year 2018.

Reviewer comment: Line 94: I would expect a citation to the NWIS data sources mentioned here.

Authors response: We agree and have added the citation

Reviewer comment: Line 97: What is meant with the "two datasets"? “For each observation” seems unnecessary here.

Authors response: We have clarified that we mean two data sources and have removed “for each observation”. The section now reads “NWIS data also have unique gauge identifiers for each observation reducing the complexity of merging the two data sources”.

Reviewer comment: Line 97: The match with the CAMELS station does not explain the promised "geographical coverage". Please adjust this sentence. An option could be a reference to a figure showing the distribution of stations, maybe including the ones from the CAMELS data that are not amended with chemical data.

Authors response: We agree and have included maps to showcase the capacity of CAMELS-Chema across a number of attributes (e.g. Figure 1, 2, 3, 4, 5 and 6).

Reviewer comment: Line 102: From my point of view the 2.2 chapter on the methods is not informative and needs some revision. The reader does learn some fancy names but not what exactly was done with the data. The linkage of methods to the provided text-files remains unclear. This does not fit the previously stated wish to make everything reproducible by providing scripts. See also details below.

Authors response: Based on this comment and similar ones from Reviewer #1, we have rewritten a section on our Data Harvesting and Integration approach (Section 2.2) to more directly speak to why and how the Extraction, Transformation and Loading framework was applied to generate the CAMELS-Chem relational database. We also moved the order of this section to follow a presentation of the sourcing and characteristics of the specific data sets integrated within CAMELS-Chem (e.g., NWIS, NADP).

Reviewer comment: Line 103: I cannot follow this argumentation.

Authors response: These lines have now been replaced by the rewritten section.

Reviewer comment: Line 109: Is ETL an established method? If so, please reference this. If not: Is there really a need for a fancy name for a standard data transformation method needed that uses a scripting language in a reproducible way?

Authors response: We added more information: in section 2.2.: Extraction Transformation and Loading is a framework that has been established in the business information world for some decades (Kimbal R, Caserta J. 2004. The Data Warehouse ETL Toolkit. USA: Willey Publishing, Inc), but only more recently has it been applied in environmental data management. In our rewritten section, we focus less on ETL as a novel method, and instead provide specifics of why and how the ETL framework was applied to generate the CAMELS-Chem database.

Reviewer comment: Line 112: I do not see the link from this method to the resulting data provided, which are easy-to-use but simple text-files. This paragraph seems to be overly complicated.

Authors response: These lines have now been replaced by the rewritten section.

Reviewer comment: Line 116: Similar as before. Either reference this as established methods or leave it. These abbreviations are not used further in the manuscript - so why introducing them here?

Authors response: We agree. These lines have now been replaced by the rewritten section.

Reviewer comment: L131: "once per day" at the same time? Is this what is meant here?

Authors response: We agree that this sentence needed clarification and have reworded "We address this by making the assumption that field technicians generally collect samples for multiple solutes at the same time."

Reviewer comment: Line 135: Again a reference to the dataset would makes sense for me. It would be, moreover, helpful for the readers to know about the absolute number of stations the interpolation is based on.

Authors response: We agree and have added this information to the MS (e.g. in abstract and methods).

Reviewer comment: L 137: I do not understand what "align" means in this context. Please be more specific.

Authors response: We reworded this section for clarification: Data rasters were positioned to correspond to CAMELS catchment shape files to determine total watershed deposition for 10 species for a given year.

Reviewer comment: Line 139: Somehow strange to mention Table 2 first in the text while Table 2 comes second.

Authors response: We agree that this formatting is not helpful and made sure that table 1 is referenced before table 2. For the final version the Journal will clarify the formatting and placement of tables.

Reviewer comment: Table 1: It would be helpful, if not done elsewhere, to state the USGS parameter codes as this often causes confusion. The caption is not fully matching the column names, e.g. is abbreviation=attribute? Also it would be good if this is also consistent with or clearly linked to the column names in the provided data files, which seems to be not the case.

Authors response: We agree and have added all USGS parameter codes and abbreviations to the table.

Reviewer comment: Line 145: The "+" seems unnecessary.

Authors response: Fixed as requested

Reviewer comment: Line 151: This description says all what is needed. No further need for a table here.

Authors response: We can see what the reviewer means, but prefer to keep the Table so that the reader has a quick comparison at hand on what parameters are included for atmospheric inputs vs water quality.

Reviewer comment: Line 151: Please use subscript in the constituent's abbreviations similar to the table where the constituents are introduced. This applies to the entire manuscript.

Authors response: We agree and have revised throughout.

Reviewer comment: Table 3: I do not get the meaning and idea of Table 3.

Authors response: We agree, it was redundant and we have removed the table.

Reviewer comment: Line 164: This chapter would profit from a table (maybe combined with table 1) that lists number of stations and number of observations and maybe median number of observations per station for each of the constituents.

Authors response: We agree that the number of observations is a very important component for user evaluation of the data. We chose to supply this information with our maps, which now include all solutes.

Reviewer comment: Line 171: I was not aware of this varying foci. Is there a reference for that?

Authors response: The shift in foci is localized in some cases and we have provided a reference "CAMELS-Chem offers long-term records for trend analysis and broad geographic coverage in catchments (Figure 3 as well as S4 in supplementary materials). Because USGS sampling foci varied between decades, temporal biases in the sampling record exist (Shanley et al. 2015). "

Reviewer comment: Line 175: This sentence needs to be checked - the logic is not clear.

Authors response: We agree and clarified: "The NO₃ data are more abundant in the Midwest and east coast where agriculture is generally more intense and focussed sampling for nutrients is common.

Reviewer comment: Line 186: While the previous chapter states reasons for different data density and distribution. Here, however, suddenly spatial pattern in the deposition is described. This belongs elsewhere. The chapter should describe the meta-information only.

Authors response: We agree and have revised the entire section. Atmospheric deposition is now under a new header as section 3.5. “Atmospheric deposition needs to be considered when evaluating water chemistry patterns and especially for weathering studies, the contribution of atmospheric deposition needs to be corrected for (Berner and Berner 2012). For example, Cl and Na deposition values are higher in coastal areas (Figure 7e and j), while NH₄ and NO₃ deposition values are higher in places where anthropogenic inputs of fertilizer are high (Figure 7b and c). Ca typically has higher values away from coastal areas and is strongly impacted by local bedrock and soil composition (Berner and Berner 2012). In many cases these patterns are consistent with patterns in stream chemistry (e.g. Figure 1j for stream NO₃)”.

Reviewer comment: Line 226: "not" missing?

Authors response: Agreed, we fixed this section: “As expected, plotting trends in SO₄ stream chemistry and wet deposition for an earlier timeframe (1985-1992, Figure 8a) decreasing trends in SO₄ deposition and corresponding decreasing trends in SO₄ stream chemistry are apparent. Wet deposition trends remain decreasing in the following two decades (1992-2002 and 2002-2010), but without much response in SO₄ stream chemistry”.

Reviewer comment: Line 234: This reads quite strange. You compute cq relationships to check if there are cq relationships? Consider revising.

Authors response: We agree and this has been fixed through rewriting the entire section on C-Q relationships (now section 3.3).

Reviewer comment: Line 250: I do not get the meaning of "overlapping solutes" here.

Authors response: We agree and this has been fixed through rewriting the entire section.

Reviewer comment: Line 251: Isn't this part of the A (attributes) in the CAMELS dataset?

Authors response: Yes, we agree that we have not set this up well earlier in the description. The way we designed CAMELS-Chem is that all catchment attributes are now connected to these additional datastreams and can be queried as such. This is actually the main contribution of the database.

Reviewer comment: Line 127: I very much like the analysis of FDC coverage. For me this is, however, not just an addition to the cq analysis but rather part of the dataset description in the former chapters.

Authors response: We agree and have added the FDC analyses as part of the dataset description and include maps in Fig 3 and Fig. 6.

Reviewer comment: Line 269: "response of the relationship"? Consider revising.

Authors response: We agree and have revised “the slope” of the concentration-discharge relationship.

Reviewer comment: Line 270: What are uneven collection dates?

Authors response: We agree and this has been fixed through rewriting the entire section

Reviewer comment: Line 299: This statement seems to be unrelated to stream chemistry and thus not directly a result of CAMELS-Chem.

Authors response: We agree and this has been fixed through rewriting the entire section

Reviewer comment: Line 302: This sentence is very long. What exactly is meant by large-scale response? The observed response is always local, isn't it?

Authors response: We agree and this has been fixed through rewriting the entire section

Reviewer comment: Line 404: This sentence is redundant, please revise.

Authors response: We have removed this sentence.

Reviewer comment: I think the authors may already give an idea which repository they are aiming at.

Authors response: We have decided to make the dataset (1980-2018) available on Hydroshare and a doi is available and included in the MS. The database is housed at the University of Vermont as it will be continuously updated.