

Geosci. Model Dev. Discuss., referee comment RC1  
<https://doi.org/10.5194/gmd-2021-368-RC1>, 2022  
© Author(s) 2022. This work is distributed under  
the Creative Commons Attribution 4.0 License.

## Comment on gmd-2021-368

Anonymous Referee #1

---

Referee comment on "Climate Services Toolbox (CSTools) v4.0: from climate forecasts to climate forecast information" by Núria Pérez-Zanón et al., Geosci. Model Dev. Discuss., <https://doi.org/10.5194/gmd-2021-368-RC1>, 2022

---

**Overview.** This paper presents a toolbox facilitating the integration of climate data in sectoral applications. The manuscript is well structured with a clear general introduction, an overview of the toolbox, and some user cases. The R package presented is a valuable contribution. All post-processing steps are integrated in the framework and the processes are compatible with other R libraries. The nested structure of functions, allowing users to adapt specific blocks, and the exhaustive integration of potential operations allowing to process climate data to produce relevant climate forecast information, make this toolbox a valuable contribution in the field of climate services. The resources are open access (Zenodo, Git Hub, Gitlab) and relatively easy to use thanks to the associated R documentation. In addition, the vignettes and use cases made available by the authors are well commented/illustrated and can be launched and (mostly) run. The case studies presented in Section 3 are relevant and well-illustrated by the introduction of two models requiring such post-processing (SNOWPACK and SCHEME). The visualization (Figures 3, 4, 6, 7, 9, A1, B1) of climate data and potential output is of high quality which is also a real contribution for R users.

CSTools functions can be complex, and the handling of the package is not trivial with limited knowledge. In general, R users expect to (i) install and load a package, (ii) run an example notebook, (iii) retrieve an output, in a reasonably limited time. This manuscript should better give the users a clear idea of the available methods and guide them through main difficulties. For example, the authors acknowledge that climate data retrieval/loading and formatting (of the NetCDF files prior the import in R or python environments) is often a blocking point for some users: *"This can be a labour-intensive step when trying to combine multiple datasets such as observations and forecasts from multiple systems"* (l.68). However, the current manuscript simply refers to external notebooks without further comments (l.196-203). In general, there are file paths, in both this paper and vignettes (GitHub), pointing where data is stored (apparently locally) with no reference or link to retrieve the original input files. Following the link Earth Sciences / CDS Seasonal Downloader · GitLab (bsc.es), leads the reader to python functions such as *"download\_seasonal\_cds\_monthly.py"*, which raises the question of the consistency of the tool in terms of programming language (R, CDO, python). From the point of view of a climate risk modeler (in finance or insurance for example), this manuscript is not self-contained enough to allow a quick grasp of the R functions while having a precise idea of the underlying methods.

**Some minor adjustments could help broaden the target audience of the paper.** For example, make sure each specific acronym or term is introduced. This toolbox aims at facilitating the integration of climate data in “*sectoral applications*”, yet the manuscript is hardly accessible to potential interested parties, who would not necessarily have the knowledge of all the techniques recently developed in the field of climate services. Although it is acknowledged that the paper aims at experts (“*applied climate scientists or climate services developers*” l.59), in practice, specialists who already handle climate data frequently have certainly developed their own routines and procedures to perform most of the stated operations, so extending the target audience to non-experts could truly add value to this manuscript. Indeed, especially in the context marked by an increasing concern about climate change, this toolbox would gain from being more understandable by energy system planners (l.530) but also mathematicians, risk modelers, insurers, economists, agricultural engineer, etc. Currently, although the manuscript includes ample references, authors often use specific acronyms or terminology (in particular in 2.2) without proper and non-technical introduction/definitions that would allow the target audience to be broadened (e.g. “Best Estimate Index” l.301, “ignorance score” l.250, “SEAS5” l.430, etc.).

**The overview (section 2) should add value to R package and documentation with further abstraction and description of the underlying processes.** For example, the underlying procedures are not described mathematically in the current version of the manuscript. We could expect this paper to better describe and focus on the “processes” (mathematical specification, parameters, hypothesis), especially as the functions and attributes are already described in the R documentation and vignettes.

**Regarding the presentation, the authors too often use bullet points (listing).** This manuscript sometimes looks more like a complementary user guide made of “lists” than a model description paper. This is particularly true for section 2. This problem of structure affects the substance because even if each of the functions is described in an understandable way, a linear reading of the manuscript makes it difficult for the reader to retain the main mechanisms and methodological choices the package embeds. The structure of the use cases (section 3) should be streamlined to facilitate the reading, e.g. (i) application (why?), (ii) data required/ input and at what resolution / frequency, (iii) process required from source to model, (iv) code guide, (v) output and final data visualization and (vi) interpretation. Some sentences/paragraphs which refer to documentation or with links, could be removed or placed in footnotes. In addition, there are some (not always working) links in the text while we would rather have the information in the document, whereas there are code boxes with path to NetCDF files without the link of website to retrieve the data.

**Contributions.** If one of the contributions is the “gathering” of existing functions in a harmonized toolbox, it is hard to say if some of the processes are original or not in the current version.

**Paper hardly self-contained.** In general, the paper requires to know or check references and nothing can be done from scratch based on the description given in this

manuscript only.

**Conclusion and recommendations:** The paper does not present a model but a toolbox to introduce climate data in several applications. This toolbox fills a clearly identified gap and could help researchers addressing relevant scientific questions within the scope of EGU. This paper proposes no substantial advance I could identify, but from an operational standpoint, the proposed package is within the scope of GMD and the amount and quality of supplementary material is significant. However, in the current manuscript I would recommend the authors to provide more information about (1) the sectoral applications and highlight climate relevant information beyond the three user cases, (2) the modeling structure of underlying functions to help users understand the methods and assumptions, i.e. (2.a) the input, (2.b) the mathematical formulae, (3) the definition of abbreviations, acronyms and technical terms. On the presentation side, (3) avoid excessive use of lists, (4) avoid extensive use of links and (5) streamline the case studies.

Specific comments and typing errors:

- **I-39. "stakeholders":** I would appreciate a series of examples for sectoral applications introduced in the beginning of the paper (agriculture, tourism, consumer discretionary stock planning, climate risk for insurer/ infrastructure, energy (wind but also solar/ thermic etc.).
- **I.41. "tailored climate information":** The transmission channels from climate data to climate relevant information could be slightly more detailed in this section.
- **I.50** To address these needs
- **I.57.** CSTools targets primarily
- **I.100:** R based
- **I.105-110:** first sentence in the end or footnote (from a detailed description ...)
- **I.191. "automatically interpolates all the data onto a common grid":** What is the advantage of `CST_load` (turning `ncdf` into `s2dv_cube`), vs. traditional `ncdf4` package loading `netcdf` object directly? In general, key advantages of the package vs others could be better exposed in the paper rather than in the data description vignettes ("Some benefits of using this function are"). In addition, instead of `CDO`, would it be possible to use internal R functions such as `rasterize` (package `raster`)?
- **I-196: These sentences:** *"Although datasets can be retrieved from OPeNDAP URLs with NetCDF files, in general, the datasets have to be downloaded onto a local repository and formatted to comply with the CST\_Load requirements. Observational reference datasets are stored in a folder in separate monthly NetCDF files (other formats are also possible; see [https://earth.bsc.es/gitlab/es/s2dverification/-/blob/master/vignettes/data\\_retrieval.md](https://earth.bsc.es/gitlab/es/s2dverification/-/blob/master/vignettes/data_retrieval.md) for more information), while seasonal 200 forecasts are stored by start date in distinct folders (see [https://cran.rproject.org/web/packages/CSTools/vignettes/Data\\_Considerations.html](https://cran.rproject.org/web/packages/CSTools/vignettes/Data_Considerations.html)). A python code to download and format the seasonal forecast datasets from the CDS is provided in the repository CDS Seasonal Downloader (<https://earth.bsc.es/gitlab/es/cds-seasonal-downloader>)."*: should be clarified in the paper.
- **I.237. "k-mean":** how k is determined? optimal? parametrized?
- **I.259:** Why five methods? can all downscaling methods be used regardless of the climate variable considered? For instance, if a method is developed for surface (10m) wind (e.g. TORRALBA, 2017), can it be applied to humidity, sea-level pressure? If not,

the authors could list the best suited input for each method.

- **I.266:** Precise applications for each pattern for analog downscaling. What are the main differences, what should we use in which situation? Or is it recommended to use all three and minimize error?
- **I.277 and 285:** maybe recall the minimal mathematical expression of the effect of orography on downscaling (Terzago et al. 2018)?
- **I.291.** *“CST\_AnalogsPredictors function downscales precipitation or maximum/minimum temperature low resolution forecast output data, in a domain centred over Iberian Peninsula”*. The function “Analogs Predictors” works in Spain only?
- **I.292:** in a domain centered over Iberian Peninsula
- **I. 309:** better explain (1) calibration methods (evmos, mse\_min, crps\_min, rpc-based), (2) on what variables / conditions should the choice of the method be based?
- **Visualisation:** Maybe insert some “visualization” (i.e. output of the functions described for each of them so we know what it does, even if it’s in the next section)?
- **I.399:** “Oops, ha ocurrido un error 404 La página a la que intentas acceder al parecer no existe o ha sido eliminada de nuestro sitio web”
- **I.445. Code box.** please add a link where to find the data file referred to in the link to improve reproducibility
- **Figure 3:** isn’t the density shape giving a somewhat misleading idea of the underlying distribution (“smoother” than it is)? Apart from that this figure is very nice.
- **SNOWPACK inputs:** first introduced line 550, while the inputs of the models are introduced line 714 (consider restructuring)
- **I.652-663:** very important issue in climate data manipulation: the size of the data. I think a full subsection could be dedicated to this topic in the section 2, and then simply referred to in the case study section where we want to focus on the application side (and not the technical issue).