Review jss4780

I read the manuscript "cubble: An R Package for Organizing and Wrangling Multivariate Spatio-temporal Data" with great interest. In the following, I will provide the outcome of my review of the manuscript and the R-package.

The Manuscript

The manuscript is in general clearly structured. However, an important reference to the R-package sftime is missing that also deals with the representation of spatiotemporal data. To me, the motivation and mutual benefits of the R-package cubble are not entirely clear. It might be beneficial to provide illustrative examples in the manuscript that clearly compare between a cubble representation and representations of existing packages. Which data sets/structures cannot (or with a greater effort) be represented with existing R-packages such as e.g. stars and sftime? For spatial data, the coordinate reference system is essential metadata information. In one example of the manuscript, the input data does have information on the CRS but it is not discussed in the manuscript how that is handled within cubble. How would coordinate transformations be handled with cubble?

The use-case of temporal matching based on features of the time series appears an interesting, but also a very specific one. How could this be generalized to a more generic approach?

Additionally, several typos and language issues arise and limit the readability of the manuscript. Some examples are:

- doubled/missing words/none correct sentences:
    - "*components* spatio-temporal *components*"
    - "... fits works ..."
    - "... be activate rows ..."
    - "... highlighted *in* the ..."
    - "... *using* 2020 measurements *using* match_sites() function."
    - "An example of this using is included in the Appendix"
    - "... the data in -a multiple of ways on-the-fly"
- Surprising references: in Section 4.4 it says "... Glyph maps (Section 3.4)", but Section 3.4 says "... glyph maps will be explained in Section 4.4".
- Should "... it's temperatures are more **consistent**" rather be "... it's temperatures are more **constant**"?
- Typos:
    - "... spatial and **tmeporal** information are available."
    - The polar vortex, signalled by the high **speicfic** humidity, splits into two on 2002-09-26 and **further-s- split__s__** into four on 2002-10-04.
    - the data in **-a-** multiple **-of-** ways on-the-fly.

The package

The package generally works as expected, I have been a bit puzzled by the "print" of a nested cubble in its temporal face. The row `# temporal: date [date], prcp [dbl], tmax [dbl], tmin [dbl]` appears at first sight (also in comparison with the print of the spatial face), as if the temporal domain is given by all those variables. To me, a notion such as

```
# temporal: date [date]
# variables: prcp [dbl], tmax [dbl], tmin [dbl]
```

would have been more intuitive. Possibly also including the temporal range (as for the spatial domain its bbox). Following the cube principle, space and time are only the ordering dimensions, but the observed phenomenon, the variables, are the content, no matter which projection (face) is used.

In the manuscript and in the `reproducible-script.R`, a data set `historical_tmax` is introduced. In the package, the corresponding data set seems to be `tmax_hist`.

The manuscript cannot be reproduced, as the `reproducible-script.R` (part of the submission to JSS) only uses a subset, but does not provide a separate result file that would allow to compare the reproduced outputs with the desired output for the subset.

Conclusion

In view of my above comments, I suggest to rate the manuscript as "resubmit for review".