Earth Surf. Dynam. Discuss.,
https://doi.org/10.5194/esurf-2017-75-AC2, 2018

Earth Surface Dynamics Discussions
Open Access
EGU

Interactive comment on "The R package "eseis" – a comprehensive software toolbox for environmental seismology" *by* Michael Dietze

M. Dietze

mdietze@gfz-potsdam.de

Received and published: 1 May 2018

[11pt, a4paper]article pdfpages hyperref listings

Response to referee comments

[The R package 'eseis' – a software toolbox for environmental seismology]
May 1, 2018

I would like to thank the referee for the encouraging and helpful comments, all of them

Printer-friendly version

Discussion paper

obviously devoted to improve the quality and impact of the manuscript.

Referee 2.1*: My main concern is mainly about the introduction part of the paper and the motivations for the "eseis" package: - The author has to better explain his motivation for developing this new R package even through there exists several other seismic data (pre)processing and analysis open source solutions, among them (Obspy, Seisan, SeismicHandler,...) several are intensively used by a broad community. Basically I did not find in the introduction clear answers to some simple questions like: 1) If I a am new to seismic data processing (and/or if I analyze seismic data related to surface processes), why should I use "eseis" rather than another solution? 2) what is specific to "eseis" compared to other packages (pros/cons)? 3) if I am a "R-lover", why should I use "eseis" rather than the RSEIS package (https://cran.rproject.org/web/packages/RSEIS/index.html)?*

Reply: I agree and see some of the arguments fully justified. I clarified the introduction, especially the second paragraph, in several sections to address where: i) R differs from other software commonly used in seismology, ii) that the approach is not "from seismology to environmental disciplines" but rather the other way around: "opening the door to utilising seismic data for a diverse range of disciplines that are very much used to working with R" and iii) one would benefit from working consistently in one software environment rather than processing and routing data from one isolated software with its specific syntax or GUI to another. Hence, the pros – or justifications – are elaborated on in the introduction but also section 2 and 3.1. The drawbacks are now pointed out in section 3.2, during the descriptions of each of the data processing steps, as demanded in the referee's points 2.7, 2.9, 2.10, 2.11 and 2.12.

With respect to the R package 'RSEIS', my package differs fundamentally in the way

data is handled as eseis objects, how the processing chain is organised and implicitly documented, and how the package is managed. I did not find a conflict-free way in the manuscript to point at the comparably ”less well organised“ collection of functions (several appear to be doubled but with different names and partly different output structure), not to speak of the documentation of and examples for each function. The best recommendation would be, please try to work with alternatives to 'eseis' in R and report your experiences, but this is, again, not a polite and helpful phrase in a manuscript I think. Actually I benefited a lot from working myself through the source code of many 'RSEIS' functions but it is an experience I would like not everyone to share. I added a more elaborated sentence just before opening section 3.1 (p.4, l. 17-19) to clearly point out in which dimensions 'eseis' differs from other R packages devoted to seismology.

Referee 2.2*: The author argues in the introduction that it is ”essential to find a common language“ in environmental seismology. Why such a statement? Seismic data used for environmental seismology are not different than other (passive) seismic data and are (pre)processed in similar ways (this is actually what mostly does ”eseis”). “Classic seismologists” (not focusing on surface and subsurface processes) also work on non impulsive sources (“eg. tremors”), ambient noise... For example, the ambient noise interferometry approach mentioned by the author (p2, l.16-18) is applied in a wide range of seismological studies, including environmental seismology but not only! And the MSnoise package (http://www.msnoise.org/) already does the job quite well!*

Reply: I completely agree and think it is a misunderstanding. I did not mean to imply it is essential to find a common language in environmental seismology. Rather, the common language needs to be arranged among the scientific disciplines (geomorphology, hydrology, meteorology, glaciology, and so on) with respect to the seismic approach. And the term language refers to both, jargon and programming language. I explicitly mention the scope of the statement now in the text (p. 2, l. 4).

***Referee 2.3**: For me, the introduction is too much an apologia for the R language. Other languages, and especially Python, are almost not cited although they are used by a wide and growing range of seismologists (and scientists) having the same motivations as the author.*

Reply: I now explicitly mention the role of Python in seismology (p. 2, l. 1) and would as well refer to point 2.1., where I clarify that the idea of R and the 'eseis' package is about "opening the door to utilising seismic data for a diverse range of disciplines that are very much used to working with R" instead of "yet another software solution for seismologists interested in study Earth surface dynamics". In order to contribute to this topic in addition to my replies to point 2.1, I gave more emphasis on the two references that provide examples of available tools for seismology (p. 1, l. 24).

***Referee 2.4**: My advice is to be less ambitious in the introduction avoiding too general (or oriented) statements and to present the "eseis" package as a promising R solution for easy data processing with some specific modules (that users will not be able to find in other solutions; like the model_turbulence module) dedicated to the analyses of environmental surface sources that produce seismic signals.*

Reply: I assume my replies to points 2.1-2.3 cover this topic.

***Referee 2.5**: In the following sections, it might also be good to better separate the aspects related to "standard" processing of seismic data (including preprocessing, temporal/spectral plotting, sta/lta,...) from modules purely dedicated to the analysis of surface processes.*

Reply: I inserted a paragraph to address this topic (p. 11, l. 18-20).

Referee 2.6*: Part 2 (p2 l21 to p3 l21) could be condensed. Although I agree with most of the author's statements related to data/code sharing policy and general principles in coding, these problematics are for most beyond the scope of this paper.*

Reply: I am not sure about the statement that the problematics are beyond the scope of my paper. I would argue that they in fact are the motivation and justification of the R package and working with free and open software in general. Section 2 is arranged in a way that each paragraph addresses a different dimension of open and reproducible science and they feed into details explained in section 3.1. I initially had the material from section 2 organised in the introduction but realised that it would have caused a mismatch in focus (and amount of words for each of the sub topics of the introduction) as it makes up a section on its own. I would prefer to keep the section in its current form.

Referee 2.7*: Header of the "eseis" objects : How "eseis" is handling the fact that SAC files and miniseed files do not have the same information in their headers (SAC being more event oriented whereas miniseed is more dedicated to continuous streams) ? Is it possible to add information in the headers (such as events information)?*

Reply: Yes, it is always possible to modify, add or remove elements of R objects, including the header and meta elements. Thereby, the header element is the original material imported from the seismic data file, and the meta element is another list that is generated for R-internal purpose, to consistently work with the 'eseis' object. However, since the package currently allows only to write SAC files, and mseed and SAC

are not fully comparable there might be some trade off. I mention now explicitly this point in section 3.2.1 (p. 6, l. 15-17).

***Referee 2.8**: Low level programming languages (p5, l11-15) : Note that there are a lot of other analysis techniques, not yet developed in "eseis", which would benefit from the use of low level languages (for example for continuous scanning of waveform parameters)*

Reply: The misleading sentence has been removed. I was not about to say that low level code should not be used, for example in future package functionalities. Of course, whenever there is good reason for implementing code of another language this shall and can be done in R.

***Referee 2.9**: Data structure (p6 l20) : The Year/Julian day file structure is not so common. Lot of seismologists use a Seiscomp "Standard Data Structure (SDS)" (Year/Net/Sta/Chan) with day long files.*

Reply: Very good point. I added this to the list of feature requests for upcoming versions of the package.

***Referee 2.10**: Metadata : It seems that "eseis" does not have the ability to read/write standard FDSN metadata formats (Seed dataless, StationXML). They are used by a wide variety of seismologists and they include information that may be crucial for some processing. If "eseis" does not accept such type of metadata formats, the author should mention the implications and potential limitations of their dedicated way of handling*

metadata.

Reply: Indeed, these meta data formats are not yet part of the package. I now explicitly mention this at the end of section 3.2.1 (p. 7, l. 22 – p. 8, l. 2) along with the consequences. Again, for a future version of the package it is envisioned to provide such support as it is obviously a vital goal to increase the acceptance of the package in different communities, as discussed in the dedicated section 5.1.

***Referee 2.11**: Deconvolution (p7 l4-10) : following the previous comment, it seems that the digitizers have only a "gain" parameter. Not taking into account stages such as anti-aliasing filter coefficients may lead to some misinterpretation of the signal in time or frequency domain.*

Reply: The logger list contains information as noted below. So with respect to the deconvolution the digitizers are characterised by their AD value, while the gain parameter is set as argument of the deconvolution function. But the referee is right, no filter coefficients etc. are accounted for, at the moment. Thus, I mention this point in the text (p. 8, l. 10-11) and would refer to point 2.10 that with upcoming package versions and dataless seed and StationXML support this shall be resolved.

```
eseis::list_logger()$Cube3ext
$ID
[1] "Cube3ext"
$name
[1] "Cube 3ext"
$manufacturer
[1] "Omnirecs"
```

```
$type
[1] "n.n."
$n_components
[1] 3
$comment
[1] ""
$AD
[1] 2.4414e-07
```

***Referee 2.12**: Metadata / channel naming: Legend 1 indicate all the "relevant metadata" but I don't see information like the "location code", "channel name", the orientation of the sensor, etc. For example how works the signal_rotate module (or others) if the orientation of the components are not provided in the input metadata file?*

Reply: The figure shows the actual output of the function `write_report()` when applied to an imported seismic file. What is shown is the meta data (not the header data). And in this case there was no network code provided to the Omnirecs Cube logger config file. Likewise, the imported SAC file did not contain any information about location or sensor and logger type in its header part. Thus, the import assigned NA values for these parameters. The channel name is represented by the element "component" (here "p0"). I added a clarifying sentence pointing at the missing information and how it can be documented (and any changes traced) by these reports (p. 9, l. 11-13).

With respect to the second example question, the function `signal_rotate()` has predefined argument values for the channel order that can be changed manually if needed. All these information are given in the function documentation manual or online help.