

Response to Reviewer #2

We would like to thank the reviewer for their helpful comments on our manuscript. In light of the reviewer comments, we propose to include several improvements to the package to provide a more flexible and generalised framework. These include providing:

1) Additional statistical gap-filling methods. The package currently gap-fills forcing variables using ERA-Interim estimates. The new statistical methods (such as linear interpolation or synthesis methods; Abramowitz et al., 2012) will allow the package to be applied to datasets for which ERA-Interim (or similar) estimates are not available and can be used to gap-fill both forcing and evaluation variables.

2) A model-specific look-up table for outputting site vegetation type as a plant functional type in addition to the IGBP vegetation type currently provided by the package.

3) An option to aggregate the data to a different output time step (e.g. daily).

4) The ability to use the package to also process the La Thuile Synthesis Dataset (<http://fluxnet.fluxdata.org/data/la-thuille-dataset/>). This dataset precedes the FLUXNET2015 data release that the R package is designed for but includes additional sites and may thus be the preferred dataset for some users.

We hope these will further enhance the utility of the package and provide for a wider range of applications. Below we address each of the reviewer comments in more detail.

General comments:

Ukkola et al. document an R package to convert FLUXNET data into forcing data for land surface models. This tool might be useful for all land surface models and may lead to more frequent use of FLUXNET data for model evaluation. As the general steps described here are necessary for using FLUXNET data for any land surface model, this can develop into a frequently cited reference. Reading data files, converting the units and writing them into netcdf is however not a big issue for most scientists. I therefore have some suggestions that could generalize the package more and hopefully lead to a more frequent use of the package.

1) the authors convert the driving data with respect to the units. It might also be useful to provide aggregation to different time steps, not all land surface models use the same time step in their forcing.

We thank the reviewer for this suggestion and agree that it would be useful to provide the option to aggregate the forcing data to a different time step. The package currently outputs the data at the same time step as the flux tower data is provided in but we will add an option to customise the time step in the revised package.

We note that the aggregation would require the driving variables (such as precipitation) to be gap-filled in order to perform aggregation. The package currently provides the option to gap-fill these using downscaled ERA-Interim estimates. We propose to also add additional methods for statistical gap-filling (see point 1) above) for applications where ERA-Interim estimates are not available or statistical gap-filling is preferred.

2) The authors only mention that they include the IGBP vegetation classification. Many models however use plant functional types. For the package to be applicable in this respect for most land surface models a conversion to plant functional types would be necessary.

We acknowledge the reviewer's point that a conversion to plant functional types (PFT) may be necessary for other modelling groups. We will therefore add a new functionality to the package to output the vegetation type as a PFT.

As the PFTs used vary between individual models, we do not wish to provide an automated translation between the IGBP vegetation types and PFTs. The IGBP classification also does not distinguish between common model PFT types, such as C₃ and C₄ grass. Moreover, the choice of vegetation type for each site is model specific and at times a subjective choice. For example, a savanna site, such as Howard Springs, could be modelled as a C₄ grass, an evergreen broadleaf tree, or a combination of both PFTs depending on the model configuration and application (De Kauwe et al., 2015; Whitley et al., 2016).

To overcome these challenges, we will provide a model-specific look-up table for PFTs for each site, with the PFT type nominated by the user. This will be integrated with the site metadata file currently provided with the package (P5 L13). This will provide the user with flexibility to set each site's PFT to suit their model and application.

3) It would be interesting to check whether the unit conversion that is applied here is the one required for other models. The authors could gather a list with units for the most widely used land surface models and check whether additional unit conversions are necessary, and if so extend the package accordingly.

Our package uses the ALMA (Assistance for Land surface Modelling Activities; <http://www.lmd.jussieu.fr/~polcher/ALMA/>) convention. This has been the standard for land surface models since the mid-1990s (Polcher and Shao, 1996) and has been used for a number of model intercomparison projects, such as PILPS (Project for Intercomparison of Land surface Parameterisation Schemes; Lettenmaier, 2003; Polcher and Shao, 1996), and more recently PLUMBER (The Protocol for the Analysis of Land Surface Models (PALS) Land Surface Model Benchmarking Evaluation Project; Best

et al., 2015) and GSWP3 (Global Surface Wetness Project Phase 3; <http://hydro.iis.u-tokyo.ac.jp/GSWP3/>). The outputs are also CF (Climate and Forecast) compliant. This is the prevailing metadata convention used across the climate science and forecasting community (<http://cfconventions.org/>). We recognise that some models (e.g. JULES) use a different format, but developing a package that accounts for every eventuality is not feasible. Instead what we have done is develop a generic community tool that could be easily adapted to a specific scenario (e.g. a model not compatible with ALMA or NetCDF). We are of course willing to work with individual groups, helping where possible.

We would also like to note that the output variable names are fully customisable and need not comply with those used in ALMA (as detailed on P6 L5). While only ALMA unit conversions are currently provided, addition of new conversions to the package is trivial. We will provide instructions for this in the package README file (<https://github.com/aukkola/FluxnetLSM>).

Specific comments:

p.2, l. 16-26: I think it would be good to distinguish between the forcing data and the flux measurements used to evaluate the model. The flux measurements do not necessarily need to be gapfilled if the model is compared with these data in high temporal resolution. Then you can simply only use the datapoints that were measured. Of course if you want to evaluate the annual sum the fluxes also need gapfilling.

The flux measurements are not currently gap-filled by the package, as stated on P6 L24. Only meteorological variables are gap-filled using ERA-Interim estimates (if this option is chosen by the user). We propose to add additional statistical gap-filling functions to the package that do not rely on ERA-Interim data to give the user the option to also gap-fill flux variables (see point 1) above). A number of gap-filling options (such as linear interpolation, copy-fill

and synthesis based on other variables) are already provided in the PALS R package and will be integrated with the new package.

p.2 l. 35: what are Tier 1 sites?

The Fluxnet2015 data release has two data tiers with different data usage policies. The Tier 1 sites are those with an open data policy and are thus likely to be those used by the majority of users. We will clarify this in the revised manuscript.

p.3 l.21: is there any reference for this R package?

We will provide the appropriate reference (Abramowitz, 2012) for this package in the revised manuscript, in addition to the github repository already provided.

p.3 l. 22: "encourages screening of flux tower sites for model applications", what do you mean ? can you be more specific what this screening does?

The flux tower data have been gap-filled to various degrees and may have missing data periods. In many circumstances, these are not desirable for modelling applications. Our package provides an automated method for screening gap-filled and missing data. However, this may not detect all data periods and/or sites that are not desired in a particular application. The diagnostic plots generated by the package provide a final quality control step to complement the automated screening to verify that the data are realistic and as expected. For example, this will allow the user to check the magnitude and nature of variability of particular variables. We will clarify this in the revised manuscript.

p.3 l. 26-30: please be more precise: "encourages better documentation", basically this paper is the documentation of the methods, right?

What we intended to say is that the use of the package will allow the data processing methods to be fully reproducible (by including as much metadata as possible in the data files, as well as metadata about the processing used to generate the files) and easily documented in a manuscript. We will clarify this in the revised manuscript.

p.7, l. 30: please include all variables that are not gap filled.

We will name all variables that are not gap-filled in the revised manuscript.

p.10, l. 30: did you verify that the format is really directly usable by (many) LSMs? Formats might differ considerably between different models.

Our package uses the ALMA convention. This has been the standard format for the land surface modelling since the mid-1990s (Polcher and Shao, 1996) and has been used in several previous model intercomparison studies. See our response to reviewer comment #3 above for full details.

p.11, l.1: what are these specific applications?

The applications can range from model evaluation studies to addressing scientific questions using models at the site scales. For example, the user may wish to process the data differently if interested in evaluating models during short-term phenomena (such as heat waves) as opposed to longer seasonal to annual scales. We will clarify this in the revised manuscript.

Reference: Poulter, B., Ciais, P., Hodson, E., Lischke, H., Maignan, F., Plummer, S., and Zimmermann, N. E.: Plant functional type mapping for earth system models, *Geosci. Model Dev.*, 4, 993-1010, doi:10.5194/gmd-4-993-2011, 2011.

References:

Abramowitz, G.: Towards a public, standardized, diagnostic benchmarking system for land surface models, *Geosci. Model Dev.*, 5, 819–827,

doi:10.5194/gmd-5-819-2012, 2012.

Abramowitz, G., Pouyanné, L. and Ajami, H.: On the information content of surface meteorology for downward atmospheric long-wave radiation synthesis, *Geophys. Res. Lett.*, 39, 1–5, doi:10.1029/2011GL050726, 2012.

Best, M. J., Abramowitz, G., Johnson, H. R., Pitman, A. J., Balsamo, G., Boone, A., Cuntz, M., Decharme, B., Dirmeyer, P. A., Dong, J., Ek, M., Guo, Z., Haverd, V., van den Hurk, B. J. ., Nearing, G. S., Pak, B., Peters-Lidard, C., Santanello, J. A., Stevens, L. and Vuichard, N.: The plumbing of land surface models: benchmarking model performance, *J. Hydrometeorol.*, 16, 1425–1442, doi:10.1175/JHM-D-14-0158.1, 2015.

De Kauwe, M. G., Kala, J., Lin, Y.-S., Pitman, A. J., Medlyn, B. E., Duursma, R. A., Abramowitz, G., Wang, Y.-P. and Miralles, D. G.: A test of an optimal stomatal conductance scheme within the CABLE land surface model, *Geosci. Model Dev.*, 8, 431–452, doi:10.5194/gmd-8-431-2015, 2015.

Lettenmaier, D. P.: Preface, *Glob. Planet. Change*, 38, vii–ix, doi:10.1016/S0921-8181(03)00002-X, 2003.

Polcher, J. and Shao, Y.: A standard format for reporting PILPS experiments, *Glob. Planet. Change*, 13, 217–223, doi:10.1016/0921-8181(95)00047-X, 1996.

Whitley, R., Beringer, J., Hutley, L., Abramowitz, G., De Kauwe, M. G., Duursma, R., Evans, B., Haverd, V., Li, L., Ryu, Y., Smith, B., Wang, Y.-P., Williams, M. and Yu, Q.: A model inter-comparison study to examine limiting factors in modelling Australian tropical savannas, *Biogeosciences*, 13, 3245–3265, doi:10.5194/bg-12-18999-2015, 2016.