Review of GMD-2020-148

Surface [Urban] Energy and Water Balance Scheme (v2020a) in non-urban areas: developments, parameters and performance, by Omidvar et al.

This work concerns use of the SUEWS model in non-urban areas. The manuscript includes some recent developments to the SUEWS model, some analysis of observed data from eddy covariance sites, estimation of SUEWS model parameters relevant to latent heat fluxes using the observed datasets and assessment of model performance for different time periods. This study includes many components, yet the overall purpose of the paper is unclear. Is it new model developments, new parameter values, a new method for parameter derivation or an assessment of model performance? It seems likely the authors would like to tackle several, if not all, of these aspects here, but lack of a clear structure makes the manuscript very hard to follow and, in my opinion, none of these aspects are covered in sufficient detail.

One of the main problems is that the manuscript is not well organised. It often reads more like notes than a journal article and it is very hard for the reader to follow what has been done and why. It is not necessary to have six appendices when the main text is only about 450 lines! The figures and tables should also be improved (many are not especially useful and there seems to be a lot of repetition).

This work would be of much greater use if the findings were analysed at a deeper level and set in context against the literature. In general, more evidence of awareness of the literature is necessary. As an example, the authors refer mainly to the body of work on the SUEWS model, in particular the recent model development papers of Järvi et al. (2011) and Ward et al. (2016). One question that arises is why the current study is needed at all, given that much of SUEWS was originally based on non-urban models and parameters. Here the authors take the recent 'urbanised' sub-models and parameters and seem to 'un-urbanise' them again.

More detailed suggestions for improvement are given below, along with questions about various parts of the methodology. Providing these are given consideration, I believe the manuscript can be substantially improved. I therefore recommend publication after major revisions.

General comments

Structure and presentation

Throughout the Methods and Results section, the various aspects of the study are mentioned interchangeably so it is not clear whether observations, model output, calibration or evaluation is being discussed. Currently the model description is spread throughout the Methods section.

- I suggest first adding a section where the model is described along with the equations, including the new developments currently in Appendix A (these seem quite important, especially for a GMD paper, and I'm not sure why these are in the appendix whereas LAI is in the main text). For this to be a standalone publication, more general details about the SUEWS model also need to be given (e.g. what are the required inputs, what are the outputs, what scale does the model operate at). This section should simply describe what the model does, without including any methodological details about the parameterisation or evaluation approaches in this study.
- The readability of the section describing the observations should be improved so that the reader quickly gets an overview of the sites and starts to feel familiar with their different characteristics. For example, mentioning the site names in the text, not just the table, and

giving a very brief description. It would also help if the abbreviations of the sites contained the land cover code instead of the country (which is relatively unimportant). Some details about quality control of observed data should be given. It would also be useful here to mention the representativeness of the site years (e.g. the low rainfall mentioned in L342).

- Then there should be the section on how each of the model parameters were derived. Seeing as this is a key part of the manuscript, more details are needed about how these parameters were fitted. It is currently not at all clear how this was done (using multiple SUEWS runs with various parameter values and minimising the MAE?). What range of parameter values was considered? Was more than one parameter allowed to vary at once to allow for interdependencies? Was any bootstrapping done?
- There also needs to be a section describing the approach used for the SUEWS runs, e.g. spin-up, initial conditions, forcing variables used.

Having so many appendices for a short paper is not helpful. Much of the material in the appendices is not useful anyway and should be edited as strictly as in the main text. Here are some suggestions.

- Appendix A seems to be important and should be in the main text (in the new model description section). If I understood correctly, SUEWS now calculates Q* based on Ts which is based on the sensible heat flux and MOST. This has the potential to cause large errors in Q*, especially in SUEWS since the sensible heat flux is calculated as the residual of the energy balance, and also because of the poor performance of MOST and all the uncertainties of the roughness length for heat, etc. The effect of this change needs to be shown here, and makes it more important that the model's ability to calculate Q* is dealt with.
- Appendix B: this is probably appropriate as an appendix but needs to be rewritten as it is not at all clear what has been done here. Start with one or two sentences describing the purpose of this analysis. It needs to be made clear that this section concerns observed data (not SUEWS). How do the results obtained compare to rule-of-thumb values? How do the roughness length and displacement height values compare to the literature? What do the lines in Fig B1 represent? The discussion of fetch in L523-527 is very unclear and needs rewriting. Make the points smaller and axes ticks consistent in Fig B2.
- Appendix C should be moved to the appropriate place in the main text. Perhaps adding boxplots for different temperature bins would help support your decision. I would suggest rephrasing as there does not seem to be a point where evaporation 'switches off' and three of the sites have very little data below the suggested cut-off. It should be made clear that this is observed data, not model output.
- Appendix D: what do the authors want the reader to take away from these three tables? L564-565 makes no sense. Suggest deleting.
- Appendix E: as explained before, this comparison does not make sense. The Ward et al. (2016) parameters were derived for bulk urban surfaces, and were not intended to be used for non-urban areas. Suggest deleting.
- Appendix F: What does this plot add to what is already shown (and more) in Fig 10? Suggest deleting.

In addition, general readability could be improved by:

- using fewer cross-references: the reader has to work very hard to follow the text when we are constantly directed to Equation/Table/Figure/Appendix X. Use cross-references where necessary and helpful, but try to ensure the reader knows what variable/site you're talking about;

- avoiding vague language; instead specify what you mean (particularly with respect to this study versus previous studies and what is generally true/what is true in the model/what is done here);
- using more words so that the text flows more naturally and is therefore more easily understandable to the reader. This is particularly true for the table and figure captions, many of which don't really make sense;
- don't include key methodological information in captions instead of the text (e.g. L381-382).

A lot of space (Figures and Tables) is given to presenting MAE, MBE and nMAE for the different sites at different times and different states of vegetation. However, there is little insight gained and very little discussion in the text. I therefore suggest removing these figures and, if necessary, compiling the information into a single figure or table

Non-urban SUEWS in context

The background of this work seems to be the SUEWS model – i.e. an urban land surface scheme that has been developed by 'urbanising' sub-models developed over non-urban environments. The latent heat flux calculation is based on the Penman-Monteith equation with the Jarvis formulation of the surface conductance. This work seems to 'start' from SUEWS and then 'un-urbanise' the equations again by setting the anthropogenic heat flux to zero and fitting parameters for non-urban sites. In many places, the manuscript needs adjusting to reflect that these non-urban forms exist – and in fact existed long before SUEWS!

Not only is this acknowledgement missing in the model description, but also in the Introduction, Results and Conclusion. In the Introduction, the motivation for this work needs to be set in the wider context of land surface modelling – i.e. at least a paragraph describing previous work that has been done on albedo, LAI and evaporation in forest, grassland, agricultural environments and for water and bare soil surfaces too. In the Results section, the results obtained here should be compared to previous results obtained in some of these previous non-urban studies. In Table 6, how do the non-SUEWS specific parameters (albedo, LAI, roughness length and displacement height) obtained here compare to the body of literature over the last few decades and why should future users of SUEWS use the values presented here instead of those in the literature? Are the roughness length and displacement height values given in Table 6 really useful? Wouldn't it be more reasonable to use the rule of thumb relating these parameters to vegetation height at the site?

For SUEWS applications in the urban environment, Järvi et al. (2011) and Ward et al. (2016) derived parameters for the surface conductance using datasets collected in urban areas with the aim of better capturing latent heat fluxes in urban environments. The comparison using the Ward et al. (2016) parameter values therefore does not make sense at these urban sites, as those parameter values were never intended to be used for non-urban surfaces! It would make more sense to compare the current results to those derived over non-urban surfaces, such as Ogink-Hendriks (1995) or Stewart (1988) over forest.

A comparison of SUEWS model performance over these non-urban surfaces with previously published results of SUEWS model performance over urban surfaces could also be useful.

Depth of analysis

The analysis/interpretation is generally superficial and needs to be developed substantially. For example, in L353-355 the timing of the decrease in LAI for wheat is much worse than for rice but this is not discussed. In Fig 10 (and even Fig F1), the results for grass seem to show quite poor agreement but this is not discussed in the text at all. For a more complete paper and, crucially, to avoid drawing misinformed conclusions, the observed data must also be analysed. What is the explanation for the observed variation in the albedo of grassland (are the data even reliable)? How is the variation in albedo at US-AR1 related to the variation in LAI – more explanation is required, i.e. what is the mechanism proposed behind the low rainfall in 2011 mentioned in L341-343? Why are the results for water and bare soil not shown? What is the reason for the very different annual cycles for evergreen trees seen in Fig 10. Why are the conductance parameters for the individual sites not shown (L438)? These could potentially provide some important insight into the robustness of the results. Some physical interpretation of the G2-G6 values should be attempted (i.e. why did the fitting procedure result in these values and what does it tell us?).

Specific comments

The Abstract is quite vague and suggests some topics will be analysed more deeply than they are. Some suggestions:

- add 'Here' at the start of the sentence on L34 to make clear this is what you did in this study and is not a general feature of SUEWS
- add 'from multiple sites' to the end of this sentence
- meaning of 'guidance to apply SUEWS are provided' (L38) is unclear
- L39: 'impacts' on what?
- 'The relation between LAI and albedo is explored' (L39-40) – I'm not sure this is really covered in the analysis
- Add 'in the model' after 'captured' in L44
- L44-45: the meaning is unclear and there is no discussion in the manuscript of how latent heat fluxes affects modelled canopy-layer air temperature.

The treatment of snow cover is rather strange. Judging from Fig 10, either the detection of snow cover is not appropriate or the modelled albedo does not capture the true seasonal variability in vegetation characteristics. As snow cover is not considered in this work, perhaps additional sites with much less snow during leaf-off periods would be more valuable.

Why are albedo and LAI parameters derived for each site but surface conductance parameters derived for each land cover type? Analysis of the variation in conductance parameters between sites would be informative and may help to inform about applicability to other sites.

The paper is missing a balanced consideration of shortcomings of this study. For example,

- Would these non-urban parameters be expected to be appropriate for e.g. deciduous trees in residential areas (L66) given the possibility of increased urban temperatures or advective effects?
- Considering the importance placed on seasonal variation the impact of assuming constant OHM coefficients should be addressed.
- By not including snow cover much of the leaf-off periods are not useable, and the significance of accurately capturing seasonal variation in LAI is reduced. If snow cover is not addressed here, these sites seem like a strange choice.
- Previous shortcomings of LSMs have highlighted that parameters derived for particular conditions can lead to bias in model results at other sites. Although this study aims to provide new, generalised parameter values the range of sites used for each land cover type is not very large and does not cover a large geographical area. The limitations of this should at least be mentioned.

- Keep in mind the appropriateness of the statistics used when comparing different sites and different seasons, where the variables may have different magnitudes and different amounts of data available. Depending on the analyses that end up in the revised paper, this may not be an issue, but it is important to consider. Was any consideration given to other statistical measures (e.g. the correlation coefficient would indicate how well the model reproduces the variability, even if the magnitude is wro)?

No uncertainty estimates are provided making it very difficult to judge the robustness and accuracy of these results. This should be rectified in the revised manuscript.

More minor comments

L56-58: Why only mention anthropogenic heat and water here? Urban LSMs have many more additions than this.

L62: make it clear that these land covers can occur in a single grid. Somewhere the intended grid size for SUEWS needs to be mentioned.

L64, 67: what is meant by 'integrated'?

L70: Here would be a good place to bring in some of the non-urban literature.

L71 'bridge this gap' – you have not really talked about a gap so this doesn't really make sense.

L76-79: This overview does not make sense (perhaps reflecting the confused structure)!

L93: 'phenology changes key model parameters' – changes what to what?

L93-102: these paragraphs do not really make sense. Suggest rephrasing and incorporating at the appropriate stage in the model description section

Table 1: I personally find the last 3 columns unhelpful. There are details missing from the Definition column (GDD abbreviation, 'coefficients' alone is not informative, 'shortwave radiation', OHM) but possibly this table could be deleted as everything should be defined and explained in the text anyway.

L163: what is t?

L165-169: This methodology should be separated from the model description and more details should be added, including what timestep and what type of regression. What is the justification for ignoring variation in these OHM coefficients, particularly for this paper on seasonal variation – was any observed? At least a couple of sentences with references should be added here.

L181: measurement height for wind speed appears to be Hu in L271

L187: 'stability scale' → 'stability parameter'

L201: 'Phenological state is critical' – what is the justification for this statement? Presumably all functions are important if they are not correctly parameterised.

L120: It may be easier to follow if this new LAI equation was presented along with Fig 5 so the reader understands why a different parameterisation is needed. Some justification for this equation would be helpful.

L135: In reality or in the model?

L141-142: Meaning unclear. Which 'model parameters'? How does this paragraph fit with L148-154?

L143: And also the longwave radiation components

Table 2: Without discussion in the text this table is not much use. The caption does not make sense. Why were OHM coefficients for some surfaces derived here and others from the literature?

L218: How was it ensured that the surface was dry?

L253-263: Table 3 is vaguely referenced 4 times here. Why not provide the appropriate information in the text if necessary?

Table 3: Why are the DOIs given here? They should be included in the Reference list.

Table 3: How were the study years chosen and then how was it decided which to use for calibration and which for testing? Without information, a cynical reader may suspect the combination was selected which gives the best results…

Figure 2: This figure does not seem very useful. Suggest deleting unless it is better described in the text and provides more insight.

L334-334: Units of LAI missing

L356-359: Meaning unclear.

Figure 7, 12, and 14 take up a lot of space, are not very easy to read and are hardly discussed in the text. Perhaps a more useful way to summarise this information could be found, or only the most relevant results displayed. The same goes for Figure 9d and Fig 11d. Note it is frustrating for the reader to have to adjust to essentially reading the same information as Fig 7 and 12 but presented in a different way.

Panel labels are missing from Fig 9 and 11.

Would it make sense to merge Fig 6 and 9a-c and Fig 10 and 11a-c so that these similar results are presented in a more comparable way? I would suggest perhaps even making the x-axis all one year long so the difference for the crops can be seen more immediately. Please use smaller points.

Table 4: Tplant, GDDv and GDDLAImax would be better as separate columns.

Fig 12 and L403: the MAE and MBE seem small for the grassland sites considering the results shown in Fig 10. Please check.

L427-429: Where is the justification for this statement? Was the performance of other fluxes checked (e.g. net radiation, storage heat flux)? If these fluxes are not modelled adequately, wouldn't they result in inappropriate conductance parameters being derived (e.g. parameters which are tuned to give the right results for the wrong reasons)?

L430: More explanation needed here. Also Q* needs to be reasonably accurate. The assumptions in the resistance approach and the uncertainties in the roughness parameters should be discussed too. The assumption of homogeneous fetch requires more explanation if it is included here.

L433-436: See above for explanation of why this comparison does not make sense.

Fig 13: Difficult to read (make full-page?). Use smaller points. The annual diurnal pattern is of limited use considering the huge seasonal variation which makes the large interquartile ranges. Consider using daily or (if data availability is an issue) monthly evaporation totals instead, which may allow insight into when the latent heat flux is modelled well and when not.

L456-457: This cannot be concluded here as it is not demonstrated in the paper. This hypothesis is suggested as ‘a possible explanation’ in L341-344 but no other possibilities are discussed and there is no further analysis to substantiate or contradict this suggestion.

L459: It should be stated somewhere that (presumably) to obtain fitted parameters for a specific site observations must be available.

L558: ‘Given this’ – given what?

There are also a few typos that would need to be corrected at a later stage.