Response to referees

We thank both referees for their detailed reading of the manuscript and for their valuable comments. Please find below our responses to both general and specific comments.

In addition to the modifications suggested by the referees, during the review process we found a minor bug in our code related to leaf area index that has now been fixed. This changed the results slightly (model performance in Sections 3.1 and 3.2) whereas the overall model performance has stayed the same.

Referee #1

General Comments

This manuscript describes the development and testing of a snow scheme for the SUEWS model. The snow scheme itself is of appropriate sophistication for a model intended for use in a mesoscale model or used independently for urban planning, incorporating the primary effects of snow on albedo, heat storage, melting and re-freezing, and latent heat/evaporation, and fractional snow cover. A couple of the parameterizations are perhaps less process-based than snow submodels that can be found in land surface models, e.g., melting and freezing is calculated using the degree day method as opposed to an energy balance method, however, the authors acknowledge the need to assess site specific degree day parameters. On the other hand, the model accounts for the effects of heat release by rain on snow, an important process not always represented in land surface models. The model also accounts for snow-clearing on paved and built surfaces, although does this commonly occur on roofs? The model is evaluated for several cold climate sites with multi-year data. After being tuned, the model does quite a good job of reproducing observations.

Snow clearance from roof depends on the city and area. In Helsinki, the removal of snow is commonly observed as otherwise it may fall on the walkways and people walking there are in danger. Also, there are weight limits on roofs that they can safely hold.

Specific Comments

1. The discussion about modeling leaf area index seems out of place here. I assume LAI is influencing evapotranspiration but there is no description here as to how this is accomplished in the model, plus it is not clear how LAI (for non-deciduous vegetation) might interact with snow, e.g., is snow intercepted by vegetation, does it modify optical properties of vegetation? Furthermore, I don’t see any reference to observations for LAI, yet it is concluded that the leaf growth algorithm is improved and appropriate in both Helsinki and Montreal. The authors should elaborate on this aspect of the model.

As the focus of the manuscript is the snow, we decided to move the text related to LAI to an Appendix to make it less out of place. Unfortunately we did not have measured LAI available, but the leaf-on and leaf-off timings are based on visual inspection. We have indicated the approximate times in a table to provide some understanding of the behavior of LAI. Text related to this was added to the Appendix *"Unfortunately no measurements of LAI were available. The values are from visual surveys of leaf-on and leaf-off timings."* (P47, L17-18).

In addition, we added line *"In addition to snow, the parameterization of leaf area index has been improved to be more applicable for cold climate cities (Appendix A)."* (P3, L23-25).

2. Does the model account for how interactions between vegetation and snow affect albedo. e.g., due to the height of vegetation, albedo can be a mixture of vegetation and snow albedo until the vegetation is fully buried?
Snow does not interact physically with vegetation (e.g. shrub/tree bending), but rather via the depletion curves and therefore partly snow covered vegetation is taken into account. E.g. the reflected shortwave radiation is calculated as a weighted averaged from the snow covered and snow-free surfaces.

3. There are numerous model parameters being adjusted for each of the sites. Perhaps the authors could comment on the implications for the transferability of the model to other sites, particularly those that don't have extensive observational data.
We have now added a sensitivity analysis of the model for some of the snow parameters in the new Section 3.4. This gives some information to which parameters and the degree to which the model is sensitive. This helps to provide guidance to what are the critical parameters and their impact with respect to transferability of the model to other sites.

Concluding Comment
In general, the study is quite comprehensive, the model performs well despite or perhaps due to its complexity, and the manuscript is very well written and organized. I recommend that it be published after addressing the comments above.

Referee #2
The manuscript describes modifications to an existing surface energy and water balance urban model to include snow processes. Model parameters are optimized and evaluated against data from Helsinki and Montreal and the model is found to perform well against observations.

The manuscript is generally well written and easy to follow. The authors provide a through description of the model parameterizations and details of parameter values, which I believe would facilitate the reproducibility of the model. The study addresses a research area that still receives little attention in urban modeling and snow hydrology and therefore provides an important contribution to both fields. As such, I believe that the study should be published providing that the authors address the following important points:

1) The SWE_{LIM} parameter is currently a grey area in the model:
- Section 2.1.7 states that "SWE is assumed to be reduced to SWE_{LIM} " but in Section 2.1.8 it is stated that "human activities redistribute snow. For example large roadside

snow piles are created ...". Therefore, Section 2.1.7 suggests that the model does not conserve mass and “removes” SWE, but Section 2.1.8. suggests it is merely "moved" for fs purposes. Which is it? This needs to be consistent in both the model and the manuscript.

Both processes, snow removal and snow redistribution, take place in the model. The depletion curves describe the redistribution of snow, whereas $S_{WE,LIM}$ gives the limit when snow is fully removed from the study area. Thus, the redistribution does not affect the mass balance, but the snow removal does (see also eq. 2).

Text in Section 2.1.7 was modified to (P8, L24-2):

“Snow from paved and built surfaces ($T_{R,i}$) can be transported out of the study area. The amount removed is calculated as amount of excess snow above a user defined threshold ($S_{WE,Lim}$). This behaviour is neighbourhood specific (e.g. city or neighbourhood ordinances, snow clearance priorities). The S_{WE} is assumed to be reduced to the $S_{WE,Lim}$ at the next site specific snow clearing time period. People are also assumed to redistribute snow (e.g. paths are cleared and the snow is piled elsewhere) within the study area and this is considered via depletion curves (eq. 15a-c).”

- SWE_{LIM} values are the same in both Montreal and Helsinki. How did you obtain these values and can you justify why they are the same given that, in the text, you note that clearing is "neighbourhood specifc"?

The limits 40 mm in roofs and 100 mm in paved are based on rough estimates obtained from Helsinki at the Pasila catchment. Unfortunately no actual measurements were available. We use the same values at all sites as no better information was available. In addition, the sensitivity analysis (New section 3.4) shows how the turbulent fluxes and runoff are fairly insensitive to $S_{WE,LIM}$.

We have added text “*Table 3 lists the parameters, both for the snow covered and snow-free surface, used in the model runs. The snow parameters are optimized (above), whereas the limit values for the snow transport ($S_{WE,Lim}$) were estimated based on maximum mass allowances at the Pa catchment. Same values were used at all sites as no better information was available. According to the sensitivity analysis in Section 3.4, the model is fairly insensitive for these limits.*” (P16, L10-14) to clarify this.

- If mass is not conserved and simply “removed”, could SWE_{LIM} (too low for example?) account for the difference in snow depth between measurements and observations (assuming both depth and mass are lost) discussed in Section 3.4 and Figures 7 and 8?

Snow removal is part of the mass balance and therefore no mass is disappearing (equation 2). T_r cannot explain the underestimation of snow depth in Figure 7 and 8a (now 8 and 9a) as the modelled values are plotted against values for lawns where no removal is taking place.

2) Many of the parameters are site specific and, despite having read the manuscript 3 times, I have failed to track how many of those parameters were optimized with

data specific to this study and how many parameters were optimized with data from one town. I appreciate seeing a sensitivity study but I actually find this section hard to read. I think it would be helpful to include scatterplots of range of value tested vs. RMSE; this would be a very efficient way for the reader to instantly see how sensitive the model is to specific parameters. However, as there are many parameters, the authors may prefer providing a table with all the optimized parameters, the range of value tested, the range of RMSE, and the final value decided upon.
As suggested, we have added a new Table 4 to clarify what parameters were optimized using datasets from this study. Hopefully this clarifies the section.

Also, it seems that parameter optimization was performed one parameter at a time. In Section 2.3 the authors constrain four parameters against runoff. Given the large number of optimized parameters, it would be interesting to run a multi-parameter sensitivity analysis to assess possibilities of equifinality.
We do not understand what the reviewer means by four parameters as only the melt water coefficients a_r and a_t were optimized using runoff. To clarify the sensitivity of the model we have now added a new Section 3.4, where sensitivity analysis for some snow parameters are made. These include the form of the depletion curve in the vegetation surface, limits for snow removal, snow heat storage coefficients and melt coefficients.

3) Advection. The authors need to strengthen their argument for not including advection in their model:
(i) p1068, l14-16: Lemonsu et al. (2010) is quoted but the actual reference is Lemonsu et al. (2008) (Lemonsu et al., 2010 only references the 2008 paper) where Lemonsu et al. (2008) don't assume advection to be negligible but actually demonstrate that the footprint is homogeneous enough to omit QA from the energy balance. Looking at the surface cover fractions in Table 1, the modelled areas in this study look more heterogeneous than in Lemonsu et al. (2008). Furthermore, p 1087 the authors of the present study attribute advection as a possible source of error in the turbulent and radiative fluxes. Please show that neglecting QA as in Lemonsu et al. (2008) is appropriate for this study or discuss the potential errors associated with neglecting these processes more thoroughly.
The issue with advection was not clearly written. In the model, the local scale advection is not implicitly resolved as a separate component, but net micro-scale advective processes are embedded in the modelled turbulent fluxes. Text related to this was modified in P5, L1-5.

(ii) p1073 acknowledges that "One of the most important factors controlling the energy balance and snowmelt is the patchiness of snow". However, many studies have shown that this patchiness leads to high advective fluxes (e.g Shook and Gray, 1997, HP; Granger et al., 2006,). Presumably, lacking advection the model only uses fs to weight the energy and water balance in terms of snow-covered and snow-free fractions? The authors may like to rephrase and avoid the word "advection" (fs represents much more than that anyway) and simply state that it represents snow heterogeneity at sub-grid scale. (I would also move section 2.1.8 to 2.1.2 and clarify that fs is used for the fractional weighting of the energy balance described in Section

2.1.2. 2.1.8 is a bit far and, without context, the reader forgets why it is there at all).
Please, see response to the previous comment.

Minor comments:
- Last sentence in the abstract and p1090, l5-7 There is absolutely no proof that the model can be nested in large scale atmospheric model nor that it can be used in urban planning. The model is highly calibrated and the manuscript does not demonstrate that its parameters are transferable to other cities. The sentence should be removed altogether from the abstract and re-written in the Conclusion section.
The sentence from the abstract was removed as suggested and text *"…at sites with varying surface cover fractions"* was added (P2, L5). However, as the model can be nested and has already been used in urban planning we left the sentence to conclusions but added the word *"potential"* (P22, L22).

Abstract, l18: What do the authors mean by "accumulation"? Do they mean the rate of accumulation? Given that snow depth is underestimated and that there are no SWE observations to assess mass accumulation, this is confusing.
By accumulation and melt events we mean the timing of these. The sentence was modified to be *"At all three sites the model simulates the timing of the snow accumulation and melt events well, but underestimates the total snow depth by …"* (P1, L28).

Equation 2: How is T_R calculated? What is its value?
TR is the amount of snow removed from the study area above a certain threshold ($S_{WE,LIM}$). To clarify this we modified the text in P8, L23-25 to *"Snow from paved and built surfaces ($T_{R,i}$) can be transported out from the study area. The amount removed is calculated as amount of excess snow above a user defined threshold ($S_{WE,Lim}$)".*

Eqs. 3 and 4: Why are Lemonsu et al. (2010) referenced? They have used these equations, like many others, but not developed them. Same for Koivusalo and Kokkonen l12; Verseghy (1991) also reset albedo above a certain snowfall threshold and would be a more appropriate reference if one is needed.
Lemonsu et al. (2010) was removed as suggested by the referee. Koivusalo and Kokkonen refers more to the limit 2 mm that was successfully used in Finnish boreal forest. The location of the reference was thus moved to better correspond to the location of the content (P5, L26).

p1072, 18: Word missing between "surface" and "ra"?
We meant r_a of the snow surface and the text was now changed to *"the r_a of the snow surface"* to be more clear (P8, L8).

p1074 "the Ek et al. (2003) form of the function is used with coefficients derived from Swenson and Lawrence's (2012)". Which coefficients? 1.3? The study Swenson and Lawrence (2010) investigated fs for a climate model and they acknowledge that the shape of fs may not be appropriate for smaller scale studies. I can see no specific mention of the value 1.3 in their paper either. Please clarify.

There was a mistake in the text and the references were the wrong way around. We used the values from Ek et al.'s Figure 1 to estimate the Swenson and Lawrence's (2012) coefficients N_{Melt}. This is now corrected. There was also a mistake with the coefficient 1.3 as this should be 1.7. We are aware that both the Ek et al. and S&L give the depletion curves for larger horizontal scales, but with no better information available we decided to use these. We have added text related to the horizontal scales to P9, L16-17: *"As this function was developed for climate models its application to smaller scales does require caution."*

Section 2.1.8: A figure showing the different fs shapes for given SWE would be welcome.
A plot of the depletion curves was added to Appendix B.

Eq. 18: Whilst not incorrect, normalizing RMSE by the range is unusual in model evaluation studies. The denominator is more usually the standard deviation of the observed variable (see Taylor, 2001; Moriasi et al., 2007; Glecker et al., 2008). Given that the range is larger than the standard deviation, the nRMSE, looks much smaller when using the range. I admit that I would much rather see nRMSE with standard deviation as denominator or no nRMSE at all as I think that the range just provides extremely small numbers which don't explain much.
We chose to use the nRMSE as we want to be able to compare the different seasons and sites easily. This parameter is also the same that has been used in Järvi et al. (2011). However, for the convenience of comparing our results with climate modeling studies we have also added the RMSE normalized with the standard deviation to Table 4.

p1077 l20: Dates for runoff are not consistent with dates p1076 l18.
The first date gives the measurement period and the latter the modeling period. Text *"…,as observations are available for a shorter period"* was added (P11, L27-28) to clarify this.

Section 3.1: This section does not describe results, I would move to the previous section as 2.5.
The section was moved to Section 2.5 as suggested.

p1080, l19: reference to Eq. 4 in Järvi et al. (2011). Is it the same as Eq. 16 in this manuscript using the b1, b2, c1 and c2 given in l19? If so, maybe reference Eq. 16 instead so the reader does not need to find Järvi et al. (2011).
This section was moved to Appendixes and the extra references to eq. 4 in Järvi et al. (2011) have been removed.

p1080, l25: Replace Eq. 15 by Eq. 16.
This section was moved to Appendixes and the comment is no longer valid.

p1083: I think that it would be clearer if 3.3 was something like "Model Evaluation" and 3.3.1. was "Surface Runoff" and so on...
We have changed the title of 3.2 to *"Model evaluation"* and 3.2.1 is now *"Surface*

runoff". Similarly, *"snow depth"* is now 3.2.2 and *"Turbulent and radiative energy fluxes"* is 3.2.3.

p1085, l5: I can see other possible reasons why snow depth is underestimated; how reliable are the precipitation data? What about SWE_{LIM} (discussed above)?
We agree the reviewer that precipitation measurements, especially when it comes as snow, are typically underestimated. This could indeed explain some of the underestimation in the snow depth. We added text *"The underestimation of s_d is also impacted by the precipitation measurements, as the difference between observed and modelled values begins during the snow accumulation period. Precipitation measurements are known to underestimate snowfall especially due to wind effects (Goodison et al., 1998; Savina et al., 2012)"* to P18, L21-24.
As explained above, $S_{WE,LIM}$, can only affect underestimation of snow depth on the paved and roof surfaces, not lawn where most of the comparisons are made.

p1088, l2: this is the first time QF is mentioned in the text. How does it fit in Equation 1? How is it calculated? Are there town-specific parameters involved? Given that it seems to be rather important in Section 3.6, far more context is needed.
The missing anthropogenic heat flux (Q_F) from eq. 1 has been added. Detailed description of Q_F is out of the scope of the manuscript and its description can be found from Järvi et al. (2011). Short explanation *"Q_F is the anthropogenic heat flux,…"* was added to P4, L27-28 and *"Q_F is calculated based on cooling and heating degree days (Järvi et al., 2011)."* in P4-5, L31-1.

Table 3 is not referenced in the text.
Table 3 (new Table 4) is now referred to P16, L10.

Table 4: I think there is too much in this table. I would remove S and I (they are not mentioned in the results section anyway) and maybe nRMSE (see previous comment).
We removed *S* and *I* from Table 4 (new Table 5), but left nRMSE. However, in addition to this, we added RMSE normalized with standard deviation of the observations to the table.

Figs 7 and 8: Why are the Figures starting on 1 January when the end of the spin-up time was before the beginning of the snow season? It would be helpful to know how close to observations the model accumulation is. It may also give a few clues as to why the model underestimates snow depth (it currently doesn't help that there is 200 mm difference between model and obs at the start of Figure 7a).
The figure was re-plotted as suggested, with the spin-up times marked. The underestimation of snow depth starts to take place during the precipitation events so the underestimation of precipitation is the likely cause at least for some of the underestimation. Text is added to P18, L21-22 to explain this.

Fig 10: I find that the symbols are too large so the red symbols mask the black symbols rather a lot.

The symbols were removed from the observed values and hopefully both figures 10 and 11 are now more clear.