**Responses to Reviewer #2**

**RC :** The authors took two surface layer schemes (Noah-MYJ and Noah-MP) out of WRF, and simplified them as stand-alone modules to evaluate their performance for temperature inversions over forests in the Arctic winter. Additionally, a conceptual model was also developed to investigate the impact of individual variables. To correct the limits of the WRF schemes on turbulent collapse, some modifications were inserted. The research provides some ideas on improving the surface layer models, especially under stable conditions. However, I found the structure of the manuscript is not well organized. The interpretation of the results is limited too. So, I would like to recommend a major revision before it can be published in this journal.
**AC :** The authors would like to thank the Reviewer#2 for his/her careful review of our manuscript. We addressed each comment individually and have revised the manuscript accordingly.

**1)** **RC :** The title is misleading, since you didn't evaluate the surface layer schemes inside WRF. Especially, what's the difference of the surface layer models in the version 4.5.1 compared to the previous versions ?
**AC :** We agree that the title probably wasn't the most appropriate. We have modified it to "Evaluation and development of surface layer scheme representation of temperature inversions over boreal forests in Arctic wintertime conditions" to avoid confusion. Indeed, the goal of the paper wasn't to compare the surface layer models in this version of WRF compared to previous versions, but to compare their performance, specifically in wintertime Arctic conditions, and suggest improvements.

**2)** **RC :** In section 2 the authors first introduce the conceptual model, then the two schemes from WRF, and then in section 3 the modified schemes are described after the measurements. I feel this organization is not straightforward and confusing. The connection between the conceptual model and WRF schemes is not clear.
**AC :** The aim of the conceptual model is to gain insight into the behaviour of a 2-layer surface layer model. By calculating the strongly and weakly stable limits, the differences with a 1-layer model are put forward and the comparison with the measurement data is clearer.
We agree with the reviewer that the current organization was confusing. We have re-organised the article in the following manner :

2. Conceptual model (previous part 2.1)
3. Measurements at the Ameriflux Poker Flats Research Range (previously part 3.1 et 4.1)
4. Description of and suggested modifications to the WRF surface layer models (previously parts 2.2 et 3.2)
5. Results (previously part 4.2)

**3)** **RC :** The validation of the models (Figure 7) is based on a lot of input parameters regarded as constant values ; however, these parameters should change with time. It's difficult to draw the conclusion that the modifications improve the model performance. This part should be expanded to gain more confidence.
**AC :** It is true that in Figure 7, most input parameters are set as constant values, for illustrative purposes. To gain insight about the reliability of the model, we have selected all available observations and binned them according to their wind speed $U_a$ values in intervals of width 0.5 m s$^{-1}$. This eliminates assumptions regarding input parameters such as the net radiation at the surface. Results are shown in Fig. 9.
It clearly shows that MYJ, whether in its original (oMYJ) or its modified (mMYJ) versions, reproduces a too sharp transition due to the fact that it only considers a single layer, and strongly differs

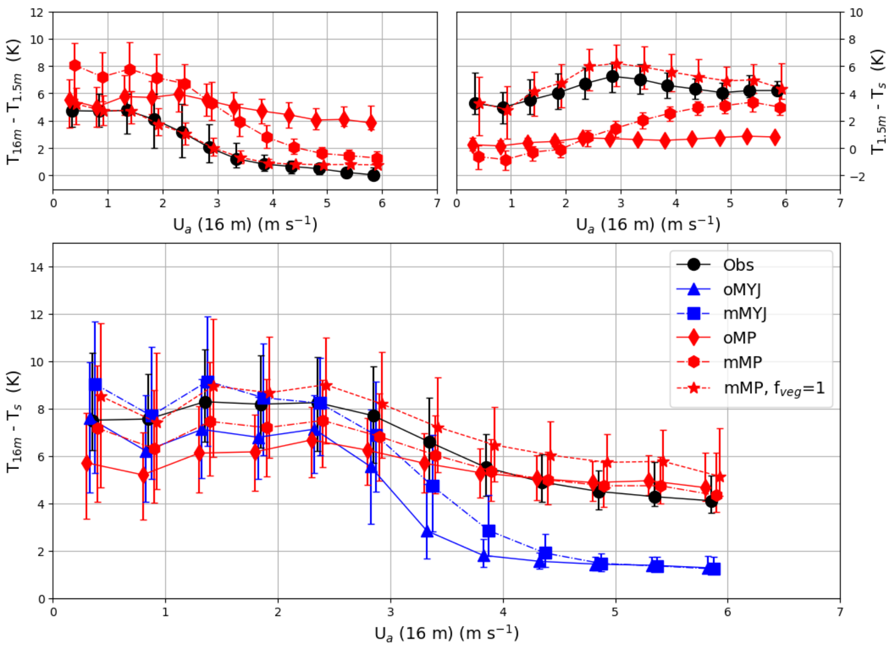


**Figure 1** (a) Median temperature difference between $z_a = 16$ m and 1.5 m ($\Delta T_{ac}$) as a function of wind speed at 16 m. Black line indicates measurements binned them according to their wind speed $U_a$ values in intervals of width 0.5 m s$^{-1}$. (b,c) Same as (a), but for $\Delta T_{cs}$ and $\Delta T_{as}$, respectively. The blue continuous and dotted lines correspond to the output of the oMYJ and mMYJ models respectively. The red continuous and dotted lines correspond to the output of the oMP and mMP models respectively. The red dashed line corresponds to the same simulation as the dotted red line, except that $f_{veg} = 1$. The error bars on the measured or modelled values represent the interquartile range (25$^{\text{th}}$ and 75$^{\text{th}}$ percentiles).

from the observations when $U_a$ values become larger than 2.5 m s$^{-1}$. mMYJ is in better agreement with the observations when the wind speed is weaker because the modelled $\Delta T_{as}$ values are obtained in a constant regime and enhanced by 2 K. This is the consequence of removing the limitation of $\zeta$ values to 1.

Regarding the 2-layer models, oMP slightly underestimates the strength of the inversion for small values of the wind speed $U_a$, even though the results are not too far from the error bars : the interquartile intervals barely overlap with those of the observed values. On the other hand, it appears obvious that it is actually due to compensation errors on the two layers taken individually : $\Delta T_{ac}$ is overestimated while $\Delta T_{cs}$ is underpredicted.

The two versions of mMP provide by far the best results compared to the observations, especially when $f_{veg} = 1$. It captures the dependency of the two individual layers (atmosphere-canopy and canopy-surface) on the wind speed well.

Furthermore, Fig. 8 shows the performance of the models over all PRR site data (curated as explained in Sect.3), using the actual measurements as input parameters. The modified versions both perform better than the original in reproducing the temperature gradient, as evidenced by a more even distribution around the 1 :1 and the lower RMSE.

This discussion has been added into the revised manuscript.

**4) RC :** The investigations on the model results are limited, especially the last part with all measurements input. The discussion should be extended.

**AC :** We are of course open to suggestions on how to improve our discussion on the results, however it is unclear to us at the moment what else we could include to extend it. The model output was analysed and the RMSE calculated over several years of measurements (Sect. 6.2), showing that the modified versions better reproduce the temperature gradient as the original. Furthermore, the behaviour of the temperature gradient in the individual layers was shown (Sect. 6.1) and the reasons for the differences in behaviour between the models are discussed.The behaviour of the modified models in other conditions (for example, other forest covers, or with shortwave radiation) is still an open question, as mentioned in the conclusion, however we consider it to be outside the scope of the present paper.

**5)** Minor comments :

— **RC :** The full name of WRF should be mentioned somewhere.
  **AC :** True. It has been added as the first occurrence of WRF.

— **RC :** Line 56 : LMDZ model should be explained.
  **AC :** The acronym of the LMDZ model stands for "Laboratoire de Météorologie Dynamique - Zoom" model. It has been detailed in the text.

— **RC :** The language needs to be improved. There are some spelling and grammar mistakes. For example, line 227 & 404 : a comma should be inserted before which. Line 401 is confusing.
  **AC :** The language has been checked carefully.
  L227 : "which" has been replaced by "that", which does not require a comma.
  L404 : A comma has been added.
  L401 has been rephrased : « Although the PRR site is classified as "Evergreen Needleleaf Forest" by the MODIS land-use categories, its characteristics are actually rather similar to a Wooded or Mixed Tundra : its trees are indeed very short and spaced out and its emissivity and roughness length are quite low for a forest site. »