Unfortunately, the authors have failed to respond to the questions raised. Below I will emphasise the two main points which along the other points mentioned earlier in the previous comments makes the method proposed in this paper to be wrong.

Main points:

1) The authors deviate from answering the main concern. A coarse representation has been used in which $A_{coarse} = I$; however, to interpret this equation correctly the state vector MUST be a full state vector on a fine grid. So, the TRUE averaging kernel is defined on the fine grid: $A = WG_{coarse}K$ (please refer to 10.2 of Rodgers).
   We also assume that, the fine grid is fine enough that the effect of discretization is unimportant for the practical purposes (thus can show the variabilities in the atmosphere). Moreover, the smoothing error in the fine grid can be calculated correctly and can represent the deviation of the retrieved profile from the true state (in a fine grid which is fine enough). In the proposed method, the first OEM run is on a fine grid in which the smoothing error and the averaging kernel can be calculated. But, how much does the retrieved profile in the second run of the OEM deviate from the true state in the fine grid? The question should be answered thoroughly. Unlike, the proposition of the authors these two retrievals are NOT independent. The coarse grid retrieval is produced using the averaging kernel of the first retrieval. How come calculating that how much the coarse grid retrieval is far from the true state (in the fine grid which was the base of the original retrieval) becomes an irrelevant comparison? Thus, to communicate the smoothing characterizations of the final retrieval in the coarse grid, the calculated averaging kernel in the coarse grid MUST be interpolated to the fine grid. Furthermore, the whole discussion on the definition of the "absolute true state" is purely philosophical and has no practicality. This is not a scientific response.

2) The authors argue that they want to answer the question that "What can we learn from the measurement without involvement of any a priori information". Even looking at this question any data user can realize that the answer cannot

be retrieving a profile which exceeds the OEM retrieval height by few kilometers. So, at the end the authors claim that by taking information out of the system (the a priori) they retrieve higher altitudes. If the degree of freedom of the new method is the same as the OEM (which is difficult to believe), how come the acceptable height has increased? The claim that the cut-off height is an add-hoc quantity is not correct, as it is really clear that the number of degrees of freedom of retrieval can tell us how much information can be extracted (and until what altitude). In this paper, an ad-hoc cut off of 0.9 is used as the cut-off height for the OEM retrieval, and another ad-hoc cut off of 60% uncertainty is used for the ML retrieved acceptable height. Then it is concluded that the acceptable height in the ML method is higher and this method should be used for UTLS studies. This is not a fair and acceptable comparison between the two methods.

Sincerely,

John Kgaran