We have addressed all of the points raised by the reviewer (copied here and shown in red text), and include our responses to each point below (in black text).

Reviewer 3

General comments

The manuscript prepared by Tuononen et al. evaluates the surface downwelling solar irradiance from the IFS model by comparing 4 years of observations from one location in Helsinki, Finland, with model output at the nearest grid point. Overall, the model bias in the surface solar irradiance is positive. This positive bias results from a combination of negative biases in less-frequent clear-sky conditions and positive biases in more-frequent overcast conditions. As part of the analysis, a new algorithm is also presented for improved detection of cloud base, precipitation and fog from ceilometer observations, which can be applied at other sites. The paper is very nicely written, with clear motivation and aims, well thought out methods, and concise results. The paper is almost ready and I recommend publication after addressing the minor comments outlined below.

We thank the reviewer for their comments on our submitted manuscript "Evaluating solar radiation forecast uncertainty". Based on the comments and suggestions by the reviewer, we have revised our manuscript.

Specific comments

P6, L21: “obtained from the closest land grid point to the measurement site”. How close is this exactly? And how much does this distance change when the resolution of the model increased from 16 to 9 km? I think numbers should be mentioned here.

We use the Meteorological Archival and Retrieval System (MARS) at ECMWF to obtain the data and we were advised by ECMWF to use a grid resolution of 0.125° in the retrieval API used for downloading data. This means that the data we receive is transformed from the internal model representation (spherical harmonic) and resolution, to a regular lat/lon grid. Therefore, our retrieved model grid point has the same lat/lon before and after the model resolution change. We have modified the text to describe this, here: ".. obtained from the closest land grid point to the measurement site, 2.1 km away."; and in Table 1: "Obtained via the Meteorological Archival and Retrieval System (MARS) at ECMWF using a grid resolution of 0.125°."

P7, L25–26: “one hour averaging corresponds to advection speeds of 4.5 m s -1 or 2.5 m s-1”. I generally like the idea to use temporal averaging of the observations to better match the spatial scale of the model, but I think this could have been handled

better. Specifically, I think the analysis would have been more consistent if observed (or even modelled) wind speeds were used to define the appropriate averaging time of the observations on a case-by-case basis. I do not suggest the authors change their analysis, but they should provide a sentence or two to support their decision. For example, are the corresponding advection speeds of 4.5 m s^{-1} and 2.5 m s^{-1} at least close to climatological wind speeds at this site?
Yes, we agree that spatial averaging based on the wind speed at the cloud level would be ideal. However, this would require the download of wind and cloud fields on model levels; we take the single-level cloud forecasts as we were developing a simple and robust methodology which can be applied rapidly to numerous sites globally. Hence, we used one-hour averaging, which has been used by many other researchers. We suspect that one-hour averaging is substantially longer than required to meet the advective-averaging time scale, which likely corresponds to advection speeds above 10 m s^{-1} at the cloud altitudes at this location. We do mention in the text that care should be taken.

P8, L2–3: "Additionally, a cloud base may not be detected in strong precipitation due to the attenuation of the lidar signal". I found this statement a bit contradictory to the earlier results presented in Fig. 2. Perhaps it could be rephrased or, if this is now an infrequent issue, it could be left out to avoid confusion.
Heavy precipitation may be a frequent occurrence in some locations so the possibility of this situation occurring should always be kept in mind. Note that Fig. 2 does not represent heavy precipitation. We have rephrased this sentence: "In strong precipitation, the lidar signal may be sufficiently attenuated so that the liquid cloud base can no longer be detected."

P8, L16–17: "which may result in a slight overestimate". Seems a bit vague. Could a reference be provided here?
We have added a reference:
Hogan, R. J. and Illingworth, A. J. (2000), Deriving cloud overlap statistics from radar. Q.J.R. Meteorol. Soc., 126: 2903-2909. doi:10.1002/qj.49712656914

P8, L18: "5.2". For the comparison of surface shortwave irradiance between model and observation, I think one important difference has been overlooked. The observations see the entire hemisphere above the given location and are therefore inherently 3D. In contrast, the model output is likely a result of 1D radiative transfer, using only the atmospheric properties of the vertical column at the given location. Under homogeneous conditions (eg. clear-sky or overcast), this may not be important. But

Fig. 4 shows the prevalence of broken cloud in summer for which 3D effects can be large. I think the authors need to acknowledge that they are aware of this difference (3D vs. 1D), even if they are not able to account for it.
Thanks for this comment. Yes, we are aware that the model may not include all 3D radiative transfer effects. We have added a sentence in section 5.2 to acknowledge this issue:"It should be noted that the model radiative transfer scheme is unlikely to completely account for the 3-dimensional nature of radiative transfer as experienced by the observations."

P9, L19: “Thus, the amount of solar radiation at the top of the atmosphere is much higher during summer”. Not just because the length of the day is longer in summer, but also because the sun reaches higher in the sky (will scale as the cosine of the solar zenith angle).
We have revised the text to say: "Due to the change in the solar zenith angle, the length of the shortest day of the year (winter solstice on 21st or 22nd December) is less than 6 hours and the length of the longest day (summer solstice between 20th and 22nd June) is almost 19 hours. The amount of solar radiation at the top of the atmosphere is much higher during summer when the solar zenith angle is also much higher (Fig. 4b, solid line)."

P9, L22: “clouds and the atmosphere”. Better to mention aerosols explicitly here. Perhaps “clouds, aerosols and atmospheric gases”.
We made this change.

Technical corrections
P3, L16: “cloud contain” -> “cloud contains”
Corrected.

P3, L25: “(Kotthaus et al., 2016)” -> “Kotthaus et al. (2016)”
Corrected.

P6, L20: “corresponding” -> “correspond”
Corrected.

P6, L26: “(LCC; Table 1) ... (MCC)”. Seems inconsistent, should probably cite Table 1 in both brackets or not at all.
We have removed the reference to Table 1, as we already stated "A list of the model variables we use is given in Table 1" in line 22.

P8, L27–28: “therefore penalizing larger errors more than small but more common differences”. This doesn’t make sense to me, consider re-phrasing.
We have revised this sentence: "MAESS uses the absolute difference between forecast and observed value, and MSESS uses the squared difference, which for two forecasts with the same absolute error, will penalize the forecast with one or two large errors more than the forecast with many small errors."

P10, L6: “forecasts” -> “forecast”
Corrected.

P11, L24: “cloud radiative properties” -> “cloud radiative effect”
Changed.

P13, L29: “to remove the observed bias” -> “to remove the bias”
Changed.