“Optimizing the precision of infrared measurements using the Eppley Laboratory, Inc. model PIR pyrgeometer” by Michalsky et al. evaluates existing approaches for transferring calibrations from PIR standards to sensors used for field operations. The authors conduct a careful and transparent analysis that provides useful results for long term networks with a historical investment in the PIR sensor. The study is a good fit for AMT. I have some comments that should be addressed in a revision before publication.

General comments:

(1) What equation is used to produce calibrated fluxes in the WISG sensors? If the WISG also uses eq (3), then wouldn’t the results shown here be interpreted firstly as an indication of consistency, but not necessarily an indication of accuracy?

Yes, the WISG sensors use Eq. (3) coefficients, which is, in fact, the original Albrecht and Cox (1977) formulation, i.e., Eq. (1). This paper is about precision transfer of calibration to other pyrgeometers, and which of four forms of Eq. (1) best does this. We have removed the occurrences of the word ‘accuracy’ in regards to irradiance from our text to clarify that this is about precision.

(2) Figure 7: Some additional analysis here is warranted. The most obvious candidate for explaining the difference between the first and second half of the cal period is the cloud fraction (the repeatability of pyrgeometer measurements is much worse under clear skies than stratiform clouds; e.g., https://doi.org/10.5194/amt-14-1205-2021), but perhaps mean temperature or precipitable water vapor could also explain it. Figures 6 and 7 could mean that the small differences reported earlier (from any equation) are overly optimistic compensation between opposing errors given fortuitous proportions of conditions within the cal period.

We agree with the referee that the small differences depicted in Figs. 6 and 7 are likely caused by outliers associated with changing meteorological conditions that were not excluded in either of the two halves of the data set. Our objective approach to exclude points was to use a robust fitting technique that reduces the influence of outliers without having to undertake a detailed, subjective approach for which data to exclude. In the end the differences are fairly small. The referee makes the point that in his study in the Arctic that the most stable conditions were for uniform stratus clouds. In fact, the WMO broadband infrared calibration transfer is performed under clear and stable conditions (Groebner--and--Wacker--2015--WMO120).

(3) It would be good to apply t-tests to determine which means are different from one another, or from zero, where appropriate. The analyzed differences are small enough that they may not be significant.

If one assumes that there are no significant differences in the calculation of infrared irradiances using the Philipona et al. (1995) formula versus each of the other three methods discussed in this paper, this assumption is rejected with 95% confidence in 15 of the 18 cases studied (six calibrated PIRs and three formulae). The three cases where the null hypothesis cannot be rejected with 95% confidence are for three of the six PIRs using the Reda et al. (2002) formula. Changes have been made to the text.

(4) L320-322: Regarding conclusions, what about the fact that the differences amongst transfer equations is so much smaller than either the WISG uncertainty or (speculatively, see 2 above) the uncertainty caused by the sampling of conditions during outdoor calibrations?

The WISG uncertainty is 2.6 $Wm^{-2}$ (Groebner--and--Wacker--2015--WMO120). The median differences in all formulae used to transfer calibration from the WISG are considerably smaller than this. Further, Fig. 7, which shows differences caused by the failure to sufficiently de-weight all outliers are also much smaller than the WISG uncertainty. Undoubtedly, in non-stable conditions the uncertainties for field measurements will be larger. However, our goal in this paper is to minimize the uncertainty where we can, specifically, in calibration transfers.

Specific comments:

L14: For clarity, "broadband thermal IR…"

Change made to clearer terminology

L40: Maybe clarify that the dome is designed to partly transmit only in the range of 3.5-50 um.

A sentence was added to clarify the dome transmission issues.

Figure 1: A few suggestions to improve the communication in this figure: (a) Label "dome" in the picture as you have done with the thermopile so that it is not interpreted as schematic of example paths in the sky (as I did at first). (b) In the caption after the word "rays" clarify that these are the numbered vectors in the picture. (c) Tb is not actually at the base of the thermopile, but is potted in the bronze casing nearby, so it would be helpful to depict the upper part of the case to show that Tb and Tr are not measuring the same thing. (d) Label Td and Tb as being thermistor measurements to distinguish from Tr, which is estimated (see also my comment at L92, which could also refer back to this figure).

Fig. 1 has been changed according to referee's suggestions.

L88: Since it is not clear from this text what Reda et al.'s justification was for including k0, it is also not clear what the present study's justification is for dropping it.

Likely, the Reda et al. (2002) equation is used at the National Renewable Energy Laboratory, however, the ARM program, for which NREL calibrates PIRs does not use the constant term in their calibration transfer to field PIRs. That is why it is not used here.

L90: I think this paragraph would benefit from a leading statement expressing the problem this paper is solving. That statement might be supported by another that explains the reason prior studies modified the original Albrecht and Cox approach. As is, the text presumes too much insider knowledge on the historical context and current gap in understanding.

Simple introductory sentence added.

L92: The fact that YSI44031s are used to measure the temperature, and which temperatures are measured this way, should be included in Figure 1.

Added to the Fig. 1 caption.

L112: Eq. 7 is odd. Can you write "c" instead of 0 in the equation to be more consistent with the Section 2 analysis/figure and then clarify in the text that in the classical form, S-H set c = 0?

Done.

L119: I'm confused about the use of the quadratic term. It looks like c = 0 for all lines in Figure 2. Where in the figure is the full cubic found? If it is the dashed blue line, it seems to be defined differently, as there is a minus sign in both the legend and the y-axis (Is the dashed blue line actually comparable to the other lines?) Also, what is c when it isn't 0, and when it is not 0, are a, b, and d the same or do they also change?

Fig. 2 has been changed to clarify these points with added text in the caption.

L120: "Interestingly…" I don't understand this statement. It seems like it would be much more surprising that changing the units yields a different result. The paper is not very long. Perhaps the appendix can be returned to the main text.

Text was changed to clarify this point.

Figure 2: An error of 0.01 C in the thermistor will produce an error < 0.05 Wm2 at 0 C, which is negligible compared to other uncertainties (similar, in fact, to the error produced by the conventional, though incorrect, assumption that sigma is 5.6700e-8). Isn't it true that the most relevant problem attributable to the YSI44031 is not the calibration method, but instead either the representativeness of its placement in the sensor in the case or the variance amongst individual thermistors in conforming to the calibration coefficients? So, I'm left not being entirely sure what the purpose of this exercise is. Is the take-away message that the YSI calibration isn't the problem with the flux calibration? If so, make that clear. [Returning to this point after reading the conclusion, I appreciate the point you made at L299-304, though it might be worth commenting on the other issues with the thermistor in the conclusion. At very least, I suggest making the purpose of the thermistor section clearer in Section 2.]

The point about the uncertainty in the temperature measurements using a 0.2 K thermistor is added to the text. I know nothing about the uncertainty caused by thermistor replacement within the brass body.

L164: When you say "using these standards", do you mean that the average of the standards was used for the calibration?

Fixed.

L247, 266: I think Figures 6 and 7, which show larger differences than Figures 3 and 4, suggest that the conditions under which outdoor calibrations (clarify somewhere that these are indeed outdoor?) are carried out are responsible for larger calibration uncertainty than the choice of equation. Yet, I think the community has historically been more focused on methodology. Maybe a recommendation to be made there?

Paragraph added to summary section suggesting larger errors should be expected for general conditions.

L313: "…are small." Specifically, the differences are an order of magnitude smaller in the transfer of relative calibrations than the reported uncertainty of the WISG.

Okay, point added.

L360, 364: Are these equation references supposed to be to A#?

Fixed.

L417: Is this Grobner (2025) from the main text?

Dropped. This information only available on calibration sheets not available to public.