Response to Referee #1 comments

The authors have systematically compiled and organized permafrost temperature data from Canada covering the period up to 2025. Given the substantial volume of the dataset and the detailed methodological approach employed, this work not only offers critical baseline data for future simulations of permafrost dynamics in high-latitude regions, but also constitutes a significant reference for the spatial distribution mapping of permafrost across Canada.

The authors compiled 427 GT and 491 GST data sites. It would be good to include a figure in the manuscript illustrating the temperature distribution of GST and GT.

We have added an additional figure (Figure 5 in manuscript) that shows the distribution of observed temperatures and a temperature-depth heatmap of observations. Supporting information has been added to the text:

Temperatures in the dataset range from -49.8 °C to 33.6 °C. The distribution of temperatures is negatively skewed for both ground surface (μ=-2.16 °C, median=-0.84 °C) and deeper ground temperatures (μ=-2.10 °C, median=-0.75 °C). Both distributions are strongly leptokurtic and this is more pronounced for observations below 20 cm than for observations at the surface (Figure 5a). Few boreholes extend past 20 m and there are clear preferential measurement depths (e.g. at 10, 12, 15, and 18 m) corresponding to greater observed ranges (Figure 5b).

Table 1: When referring to Table 1, the measurement depth for GT (Ground Temperature) is not specified. We suggest that the authors provide a range of measurement depths for GT in the table.

A new column with heading "vertical extent" has been added to Table 1. It provides the depth range in metres for the measurements in each contribution.

Line 240: Is this method applicable to all situations? Is there any supporting evidence or reference, and what is the rationale for choosing this particular formula?

The sinusoidal approximation of the diurnal temperature cycle is a simplification based on first principles (e.g., following Carslaw & Jaeger, 1959). We use this because it yields a tractable closed-form expression for bias estimation. We do not claim it is universally applicable: the assumption breaks as the effects of latent heat cause the diurnal cycle to become asymmetric. The intent of this section is to provide a transparent, first-principles bound on the potential bias introduced by sub-daily sampling, rather than an exact characterisa-

tion. We have added the following clarification to the text:

"The assumption of a simple harmonic diurnal cycle is most appropriate for conditions where latent heat effects are negligible. In near-zero temperature conditions where the effects of phase change dampen or distort the diurnal signal, the true bias may be greater than the worst-case estimates provided here. Riseborough and Burgess (1996), working with empirical ground temperature data in a context where latent heat effects are present, report a qualitatively similar pattern of estimation error: uncertainty increases with both temperature amplitude and sampling interval. This corroborates the behaviour predicted by our analytical approach."

Line 260: How were the values of 11%, 23%, and 36% derived? Were the values 70%, 80%, and 90% the only ones derived from the raw data, or were they chosen among others? What is the rationale behind this choice?

The values of 11%, 23% and 36% were derived using Equation 8 and the example completeness ratios $\alpha\epsilon\{0.7, 0.8, 0.9\}$ as follows (here using $\alpha = 0.9$) :

$$\Delta T_{\max} = \left| -\frac{Asin(\pi\alpha)}{\pi\alpha} \right| \qquad \text{(Equation 8)}$$

$$\Delta T_{\max} = |A| \left| \frac{sin(0.9\pi)}{0.9\pi} \right| \qquad (1)$$

$$\Delta T_{\max} = 0.109|A| \qquad (2)$$

Representing the error as a percentage of the amplitude makes the result data-independent for this step. However, empirical median amplitudes are provided later in section 4.2 . The completeness ratios of 70, 80, and 90 percent were chosen as illustrative examples based on what we expect researchers to use for their cut-off values of acceptable levels of missing data.

An explicit reference to equation 8 was added to clear up ambiguity in this section.

5 next steps: I would suggest changing the title. Don't use this one — it reads more like the next step in discussing a matter rather than a proper title.

We have changed the title to: *5. Collaborative steps for permafrost data*