

## S1 POD-FAR analysis method

The skill of a model can be assessed using hits and false alarms (Table S1), and, further, by calculating the probability of detection (POD) and the false alarm ratios (FAR). The POD and FAR values can be calculated as follows (Roebber, 2009):

$$POD = \frac{a}{a+c} \quad (1)$$

$$FAR = \frac{b}{a+b} \quad (2)$$

**Table S1.** The contingency table.

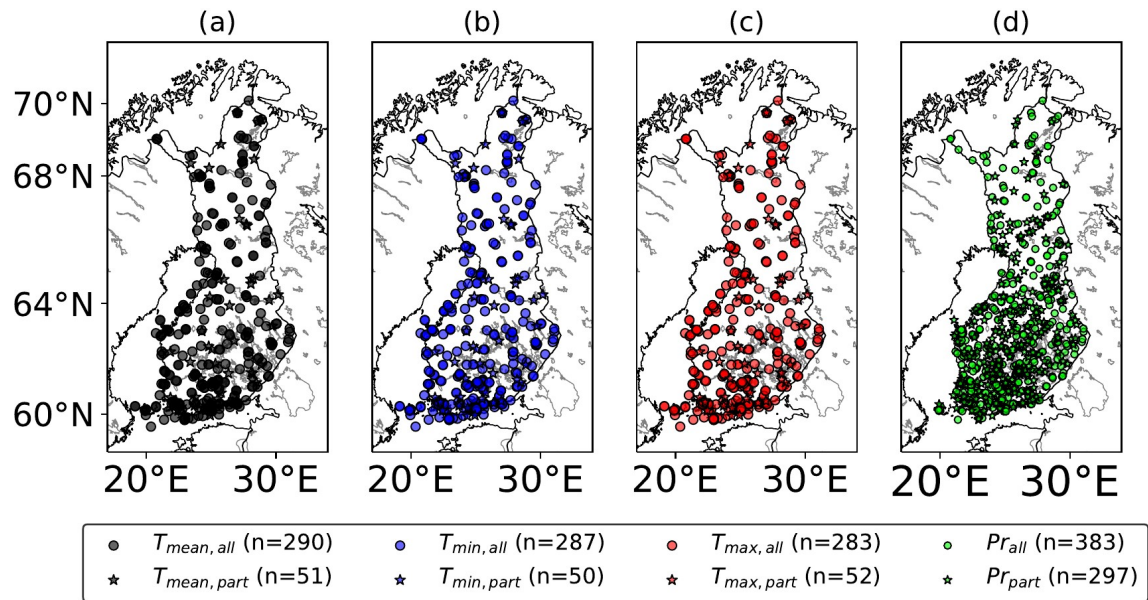
		Event observed	
		Yes	No
Event modeled	Yes	a	b
	No	c	d

The results of the POD-FAR analysis can be illustrated using a categorical performance diagram (Roebber, 2009). The POD describes the proportion of the times when the event occurred and was also captured by the model. In contrast, the FAR defines the number of false alarms divided by the number of all cases when the event is modeled. This implies that the performance of the model is the better the closer the POD is to 1 and the FAR to 0. Therefore, the best values can be found in the upper-right corner of the diagram when the y-axis shows the POD values and the x-axis the success ratio which means the FAR values in the reversed order (1–FAR). The frequency bias (Eq. (3)) indicates overestimation (underestimation) if the values are higher (lower) than 1. The critical success index (CSI; Eq. (4)), in turn, represents the hits in relation to the number of cases when the event was either observed or modeled. Ideally, the CSI values should be close to 1.

$$Frequency\ bias = \frac{a+b}{a+c} \quad (3)$$

$$CSI = \frac{a}{a+b+c} \quad (4)$$

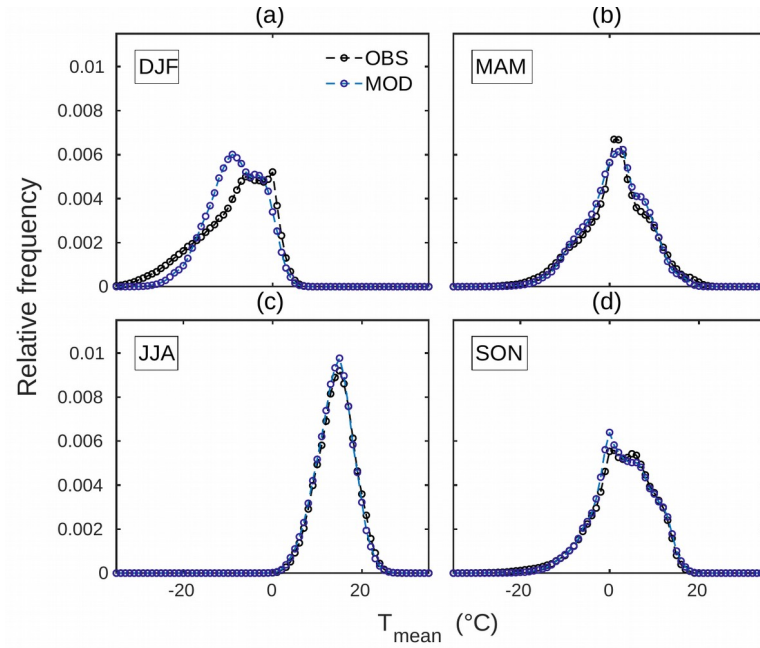
## S2 E-OBS stations



**Figure S1:** The locations of E-OBS stations for (a) the daily mean near-surface air temperature ( $T_{mean}$ ), (b) daily minimum near-surface air temperature ( $T_{min}$ ), (c) daily maximum near-surface air temperature ( $T_{max}$ ), and (d) daily precipitation ( $Pr$ ). The stations that were operative through the study period 2002–2014 were counted as “all” and stations that were operative only a part of the study period as “part” and are marked as stars.

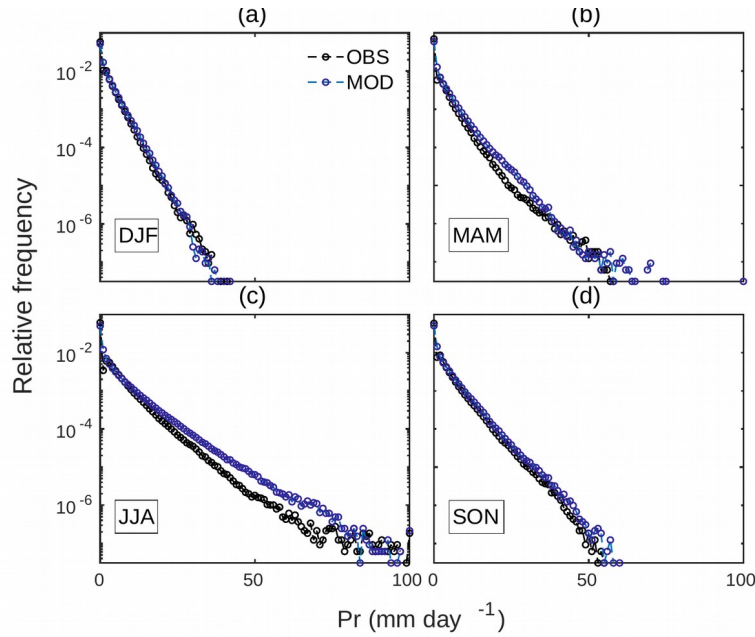
### S3 Probability density functions

#### S3.1 Daily mean near-surface air temperature



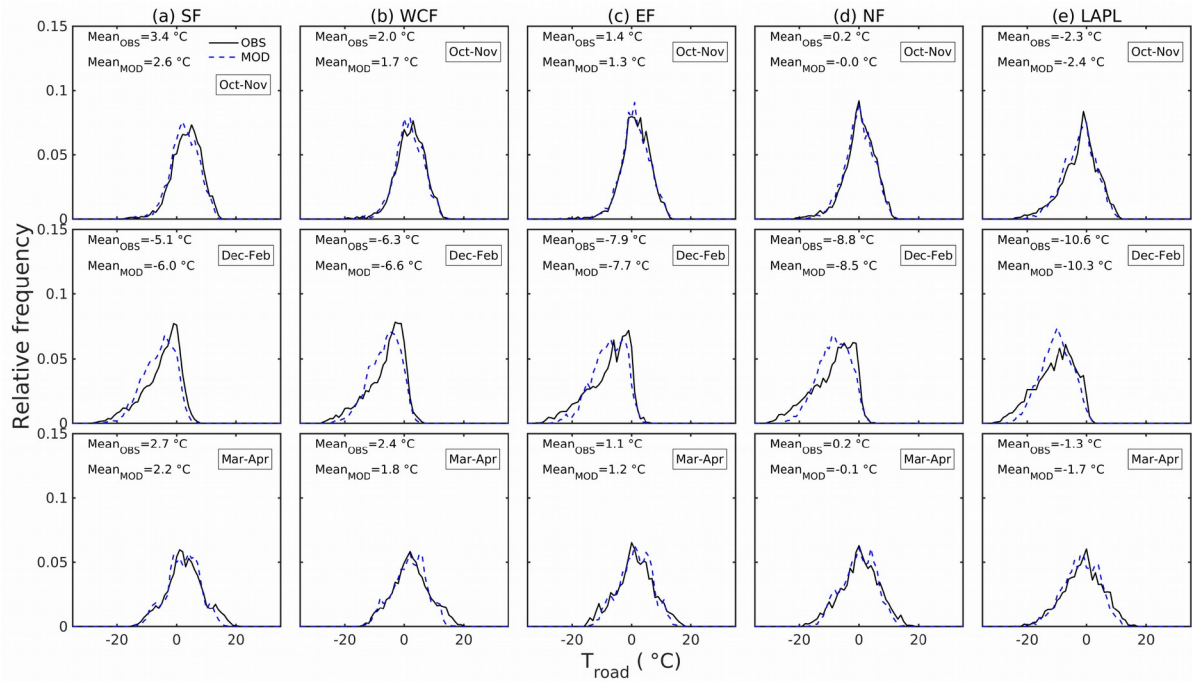
**Figure S2:** The probability density functions of daily mean near-surface air temperature ( $T_{mean}$ ) for the model and observations during (a) the winter (DJF), (b) spring (MAM), (c) summer (JJA), and (d) autumn (SON) in 2002–2014 over Finland.

#### S3.2 Daily precipitation



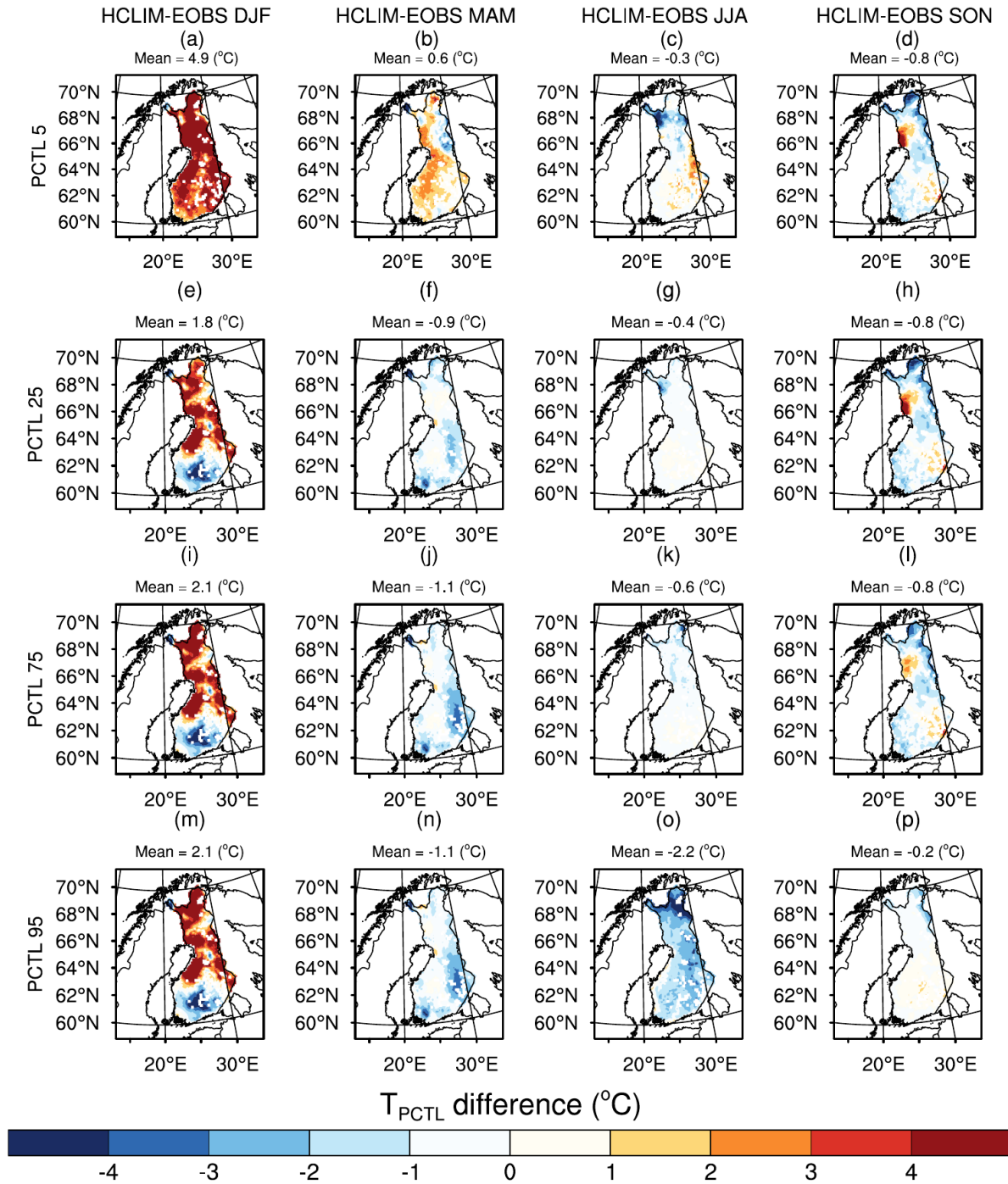
**Figure S3:** The probability density functions of daily precipitation sum for the model and observations during (a) the winter (DJF), (b) spring (MAM), (c) summer (JJA), and (d) autumn (SON) in 2002–2014 over Finland.

### S3.3 Daily road surface temperature



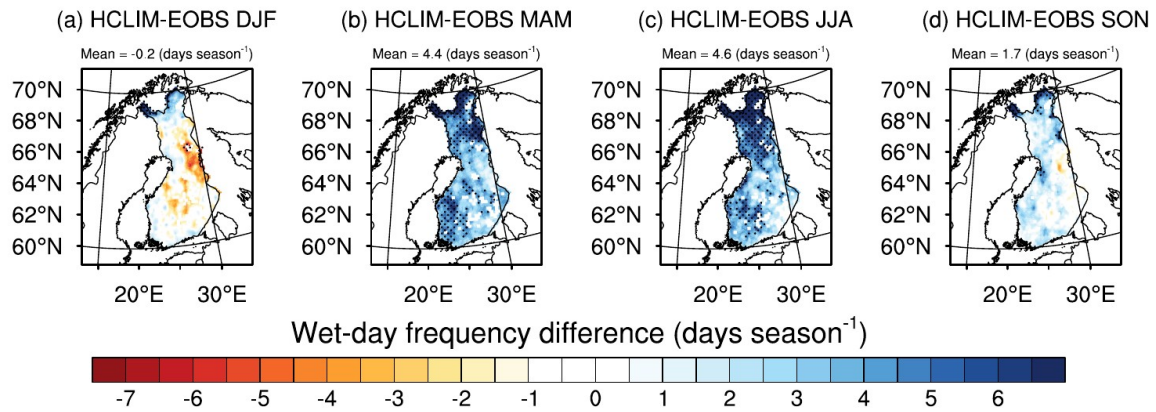
**Figure S4:** The probability density functions of daily mean road surface temperature ( $T_{road}$ ) for the model and observations during the autumn (October–November), winter (December–February), and spring (March–April) in 2002–2014 at the stations located in (a) Southern Finland (SF), (b) Western and Central Finland (WCF), (c) Eastern Finland (EF), (d) Northern Finland (NF), and (e) Lapland (LAPL). The means of the observed and modeled distribution are marked as Mean<sub>OBS</sub> and Mean<sub>MOD</sub>, respectively.

## S4 Percentiles of mean near-surface air temperature



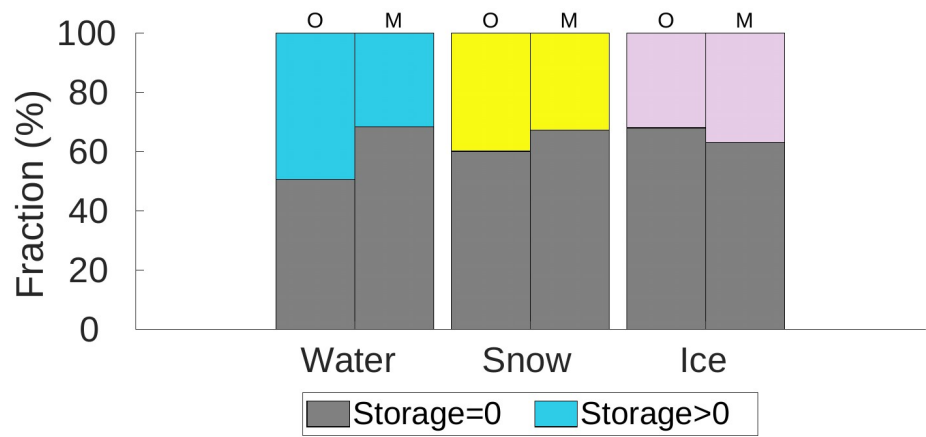
**Figure S5:** The biases in the simulated seasonal means of (a–d) the 5th, (e–h) 25th, (i–l) 75th, and (m–p) 95th percentiles of the daily mean near-surface air temperature with a reference to E-OBS. The seasonal mean biases were calculated over Finland for the time period of January 2002–December 2014.

## S5 Wet-day frequency



**Figure S6:** (a–d) The biases in the simulated seasonal means of the wet-day frequency (with a threshold of 1 mm day<sup>-1</sup>) with a reference to E-OBS. The seasonal mean biases were calculated over Finland for the time period of January 2002–December 2014. Stippled areas represent statistically significant differences with  $p$  values  $< 0.05$ .

S6 The frequency of storage terms



**Figure S7:** The observed (O) and modeled (M) fractions of days with water, snow, or ice on the road at 11 road weather stations which have an optical sensor (see Table 1). Absolute values of the modeled and observed maximum daily storages were not used directly, but instead, the daily value was set to one if the maximum value was more than zero and to zero if the maximum value was zero. The months between October and April were included in the analysis.

## References

Roebber, P.: Visualizing Multiple Measures of Forecast Quality, *Weather Forecast.*, 24, 601–608, <https://doi.org/10.1175/2008WAF2222159.1>, 2009.