



Meteorological and evaluation datasets for snow modelling at ten reference sites: description of in situ and bias-corrected reanalysis data

- 5 Cecile B. Menard¹, Richard Essery¹, Alan Barr², Paul Bartlett³, Jeff Derry⁴, Marie Dumont⁵, Charles Fierz⁶, Hyungjun Kim⁷, Anna Kontu⁸, Yves Lejeune⁵, Danny Marks⁹, Masashi Niwano¹⁰, Mark Raleigh¹¹, Libo Wang³, Nander Wever^{6,12}

¹School of Geosciences, University of Edinburgh, Edinburgh, United Kingdom

- 10 ² Climate Research Division, Environment and Climate Change Canada, Saskatoon, Canada;
Global Institute for Water Security, University of Saskatchewan, Saskatoon, Canada

³ Climate Research Division, Environment and Climate Change Canada, Toronto, Canada

⁴ Center for Snow and Avalanche Studies, Silverton, Colorado, USA

- ⁵ Univ. Grenoble Alpes, Université de Toulouse, Météo-France, CNRS, CNRM, Centre
15 d'Etudes de la Neige, Grenoble, France

⁶ WSL Institute for Snow and Avalanche Research SLF, Davos, Switzerland

⁷ Institute of Industrial Science, University of Tokyo, Tokyo, Japan

⁸ Finnish Meteorological Institute, Space and Earth Observation Centre, Sodankylä, Finland

⁹ Northwest Watershed Research Center, Agricultural Research Service, Boise, Idaho, USA

- 20 ¹⁰ Climate Research Department, Meteorological Research Institute, Tsukuba, Japan

¹¹ National Snow and Ice Data Center (NSIDC), University of Colorado Boulder, Boulder,
Colorado, USA

¹² Department of Atmospheric and Oceanic Sciences, University of Colorado Boulder,
Boulder, CO, USA



Abstract

30

This paper describes in situ meteorological forcing and evaluation data, and bias-corrected reanalysis forcing data, for cold regions modelling at ten sites. The long-term datasets (one maritime, one arctic, three boreal and five mid-latitude alpine) are the reference sites chosen for evaluating models participating in the Earth System Model-Snow Model Intercomparison Project. Periods covered by the in situ data vary between seven and twenty years of hourly meteorological data, with evaluation data (snow depth, snow water equivalent, albedo, soil temperature and surface temperature) available at varying temporal intervals. 30-year (1980-2010) time-series have been extracted from a global gridded surface meteorology dataset (Global Soil Wetness Project Phase 3) for the grid cells containing the reference sites, interpolated to one-hour timesteps and bias corrected. Although applied to all sites, the bias corrections are particularly important for mountain sites that are hundreds of meters higher than the grid elevations; as a result, uncorrected air temperatures are too high and snowfall amounts are too low in comparison with in situ measurements. The discussion considers the importance of data sharing to the identification of errors and how the publication of these datasets contributes to good practice, consistency and reproducibility in Geosciences. Supplementary material provides information on instrumentation, an estimate of the percentages of missing values, and gap-filling methods at each site. It is hoped that these datasets will be used as benchmarks for future model development and that their ease of use and availability will help model developers quantify model uncertainties and reduce model errors. The data are published in the repository PANGAEA and available at: <https://doi.pangaea.de/10.1594/PANGAEA.897575>.

50

55



1. Introduction

60 In the past decade, several long-term datasets aimed at providing high quality continuous meteorological and evaluation data for cold regions modelling have been published (Table 1). The importance of such datasets is twofold. Their primary value is scientific: they help us to understand key surface processes by enabling the development and evaluation of existing and new geophysical models for climate research and forecasting. The second, perhaps less obvious, value of having
65 multiple long-term datasets is for meta-research; the smaller the studies or sample size, the less likely research findings are to be true (Ioannidis, 2005). In a snow modelling context, this is corroborated by Rutter et al. (2009) who found low correlations in performance statistics for the same snow models but in different years.

Here, we describe ten long-term datasets (Table 1) from reference sites chosen to force and to
70 evaluate models participating in the Earth System Model-Snow Model Intercomparison Project (ESM-SnowMIP) (Krinner et al., 2018), an international coordinated modelling effort that investigates snow schemes. ESM-SnowMIP is closely aligned with the Land Surface, Snow and Soil Moisture Model Intercomparison Project (LS3MIP; van den Hurk et al. 2016), which is a contribution to the Coupled Model Intercomparison Project Phase 6 (CMIP6) including global offline land model experiments with
75 meteorological forcing data provided by phase 3 of the Global Soil Wetness Project (GSWP3; Kim, 2017). Two meteorological datasets are described for each site: one compiled from on-site measurements, the other derived from GSWP3. Previous iterations of SnowMIP have provided 19 site-years of data from four sites in SnowMIP1 (Essery and Etchevers 2004) and 9 site-years of data from five sites in SnowMIP2 (Rutter et al., 2009); ESM-SnowMIP totals 136-site years of in situ data from
80 ten sites and 300 site-years derived from GSWP3.

Measurement details at five of the sites have been described in dedicated publications within the last eight years. The other five sites are partially described in a number of publications which, combined, give a broad but not comprehensive overview of the data. All of the in situ measurements and the GSWP3 data are freely available either on the web or on request but, previously, post-processing
85 would have been required to homogenize the in situ datasets compiled by different teams or to downscale the reanalyses. This situation causes two major issues. Firstly, different modelling teams are likely to apply different post-processing methods, leading to numerous versions of the same dataset being used for scientific studies. Secondly, although time spent identifying and processing data has never been quantified in scientific literature to our knowledge, it is a well-known but under-
90 acknowledged time consuming task for modelers.



The aim of this collaborative work is to provide easy-to-use, quality-controlled data in a format adopted by the climate modelling community to facilitate consistency, continuity and reproducibility in snow research (Menard and Essery, 2019). As such, it complies with efforts in geosciences to foster best practices on data accessibility and documentation (Gil et al., 2016). The seven teams who collated
95 the in situ datasets have provided updates since previous publications and details about instrumentation, gaps in the original data and methods for gap filling. Such additions are first steps towards being able to quantify uncertainty in observed data, without which “meaningful evaluation of a model is impossible” (Clark et al., 2011). Similarities and differences between sites are discussed in individual measurement sections.

100 **Table 1: Data ownership and reference papers for the sites. Asterisks denote dedicated data description papers; the others are modelling papers in which a short description of a site is included.**

2. Data

105 Of the ten sites, two are in the European Alps (Col de Porte and Weissfluhjoch), three are in the mountains of the Western USA (Reynolds Mountain East, Senator Beck and Swamp Angel), three are in the Canadian boreal forest (the Boreal Ecosystem Research and Monitoring Sites, BERMS, the acronym hereafter collectively describing the Old Aspen, Old Black Spruce and Old Jack Pine sites), one is sub-arctic (Sodankylä) and one is urban (Sapporo). Broad geographic characteristics of the sites
110 are described in Table 2. The climate of each site is described by a snow cover classification and the Köppen climate classification, based on seasonal precipitation and air temperature. Information about data ownership and references are given in Table 1. Most sites are in artificial forest gaps or in sheltered environments. While this facilitates measurements of precipitation prone to large errors in windy environments, implications for the other meteorological variables are discussed in their
115 dedicated sections.

Three of the sites (SAP, SOD and WFJ) are located near staffed research stations, which allows frequent (daily to sub-weekly) and regular maintenance of the instruments. Col de Porte, Reynolds Mountain East and the sites in the Senator Beck basin (SWA and SNB) are accessible from nearby research facilities allowing regular (weekly to fortnightly) maintenance visits. Intensive monitoring associated
120 with the BERMS project took place in the first years after the instruments were installed, but visits to the sites during winter have become sporadic.



Both meteorological and evaluation data contain uncertainties and errors, partly due to instrument accuracy and calibration, gap-filling of missing data or subjective choices; precipitation is a notable example (See Section 2.1.3 for details). Fully quantifying these uncertainties and errors is beyond the scope of this paper, but the comprehensive list of instruments and information about missing data and gap-filling available in the supplementary material provides data users with an indication of weaknesses in the data at each site.

Table 2: Geographic characteristics of the ten sites.

2.1 Meteorological forcing data

All of the models participating in ESM-SnowMIP (Krinner et al. 2018) operate on energy balance principles, requiring incoming shortwave and longwave radiation fluxes, solid and liquid precipitation rates, air temperature, humidity, wind speed and air pressure as forcing data. Pressure is used by models to calculate air density and vapour pressure, but the temporal coefficient of variation in pressure is always very small; averages for the site elevations can be used where continuous measurements are not available (CDP, RME, SNB and SWA). At Col de Porte and Sapporo, where data outside of the snow season have not been published, all meteorological data are filled with downscaled (CDP) and bias-corrected (SAP) meteorological reanalysis data (publication of summer data for Col de Porte started in 2015; Lejeune et al., 2018). Figure 1 shows monthly averages of all meteorological forcing variables except air pressure at all sites. Details of each variable are given in the following sections.

Figure 1: Climatological monthly averaged meteorological forcing data. Wind speeds at all sites are normalised at 10 m height.

2.1.1 Air temperature

The range of air temperature at all sites is shown in

Figure 2. Sapporo has the highest annual mean (9.3°C) and minimum (-15.8°C) temperatures, although Col de Porte is generally warmer from December to February. The lowest and only annual mean temperature below freezing (-1°C) is at Senator Beck. The coldest winters are at BERMS (the lowest



temperature recorded, -41°C , is at Old Black Spruce) and Sodankylä, where most years see temperatures down to -35°C .

In terms of instrumentation, Col de Porte is the only site at which the temperature sensor is moved
155 (approximately weekly) to keep it at a constant height above the snow; otherwise, it is recommended
that measurement heights use in models should be adjusted according to observed or simulated snow
depths because this can have a significant impact on turbulent flux computations. Depending on wind
speed and solar radiation, unventilated instruments can overestimate air temperature by up to a
daytime average of 2.5°C (Georges and Kaser, 2002) or up to 10°C for individual measurements
160 (Huwald et al., 2007). Such errors are not corrected for in Reynolds Mountain East, the Senator Beck
basin sites or Sodankylä (temperature sensors at the other sites are artificially ventilated).

Figure 2: Boxplots of air temperature including means (red dashed line) at all sites. Outliers beyond 1.5 times the interquartile range (25th to 75th percentiles) are marked with circles.

2.1.2 Incoming shortwave and longwave radiation

165

Sodankylä is the only site situated above the Arctic Circle and therefore has continuous periods
without incoming solar radiation in winter (14 days) and uninterrupted daylight in Spring/Summer (44
days). Longwave radiation depends, amongst others, on air temperature, water vapour, cloud cover
and altitude. In general, the order of incoming longwave radiation (LW) between sites is very close to
170 that of air temperature (T_a). However, other influences can be seen in the Senator Beck Basin sites,
which have lower LW but higher T_a than BERMS because the former are drier. The same reversal
occurs between Sodankylä and Reynolds Mountain East.

With the exception of precipitation, radiation measurements are the most prone to errors and/or
missing data because snow can settle on upward-looking sensors. In the absence of wind to displace
175 the snow or if the instruments do not have a heating and ventilation system to prevent snow
accumulation, data are only reliable after the instruments have been wiped clean. Maintenance
frequencies are described in the introduction to Section 2. Methods for gap-filling during snowfall
events or while instruments are obstructed by snow vary; details for all sites are in the supplementary
material.

180 Radiation in Sodankylä is measured above the canopy, but evaluation data are measured in a nearby
clearing. For consistency, shortwave radiation was modified to account for the effects of shortwave
shading and longwave emission from nearby trees (Essery et al. 2016). At Reynolds Mountain East,
longwave radiation measurements started in 2002. For consistency across the RME dataset, which



185 starts in 1988, all longwave radiation is modelled but measured data are used to provide information
on seasonal and diurnal variations (e.g. cloud cover, turbidity, canopy and terrain exposure
conditions). Details of the methods used to model *LW* are in Reba et al. (2011).

2.1.3 Precipitation

190 Snowfall measurements are often underestimates and prone to large errors because much is lost to
sublimation or displaced by wind. Such difficulties are acknowledged by the World Meteorological
Organization (WMO) which, rather than imposing a standardized method, advises that adjustment
methods be chosen depending on environmental conditions and gauge types (Goodison et al., 1998;
Nitu et al., 2018). As detailed in the supplementary material, precipitation at all sites is measured
195 either with tipping buckets or weighing gauges and six different methods are applied by the seven
collecting teams to correct for undercatch: yearly or constant scaling factors, model simulations,
matching against SWE or replicate gauges. Furthermore, as weighing gauges do not provide
information on the type of precipitation, choices also have to be made about how to partition snowfall
and rainfall. Figure 3 shows how the different methods used at each site affect the solid fraction of
200 precipitation as a function of air temperature; total precipitation at Swamp Angel and Senator Beck
are assumed to be same because of their proximity so only the latter only is shown. Partitioning
methods include using dew point (RME, SAP, SWA, SNB) or air temperature (BERMS, SOD, WFJ)
functions or thresholds, and ancillary data such as snow depth and albedo measurements (CDP); more
information is provided in the supplementary material.

205 **Figure 3: Fraction of precipitation falling as snow at different temperatures, as imposed on the in
situ data and fitted to the GSWP3 data.**

Frequent summer snowfall at Weissfluhjoch and early autumn snowfall at Col de Porte can form snow
cover that melts before the winter snow pack accumulates, thus causing discrepancies between
annual snowfall and peak SWE. At BERMS, such discrepancies are mostly accounted for by snowfall
210 intercepted by the canopy that is lost through sublimation.

2.1.4 Wind speed

Wind speeds provided in the datasets are measured at variable heights but were normalised to 10 m
height assuming a logarithmic wind profile for Fig. 1-g such that



$$215 \quad u(z_2) = u(z_1) \frac{\ln((z_2 - d)/z_0)}{\ln((z_1 - d)/z_0)}$$

where u is wind speed at measured (z_1) and normalised (z_2) heights, d is a displacement height (2/3 of vegetation height at BERMS, 0 at other sites) and z_0 is a roughness length (1/10 of vegetation height at BERMS and 0.1 m at the other sites). The three sites with the lowest wind speed in Fig. 1-g are situated in forest gaps. There is little difference between mean wind speeds at Sodankylä and Col de
220 Porte. Wind speed at Sodankylä is measured above the canopy but was scaled down to 2 m height against an anemometer installed for one week in the ~2400 m² clearing. The approximately 60 x 50 m dedicated experimental area at Col de Porte is situated in the southeast corner of a larger clearing (270 x 360 m) within a spruce forest. As mentioned in Morin et al. (2012), all trees sheltering the north side of the experimental area were cut in summer 1999; mean wind speed at 10 m height was 1 m s⁻¹
225 prior to the event but 1.26 m s⁻¹ afterwards. Nevertheless, a Mann-Kendall (MK) test shows a significant increasing trend in wind speed from 1999 onwards despite the tree growth mentioned in Lejeune et al. (2018). Average wind speed at the exposed Senator Beck is the highest of all the sites and is almost four times more than at the nearby sheltered Swamp Angel. MK shows a significant decreasing trend in wind speed at Reynolds Mountain East and increasing trends at the Senator Beck
230 basin sites. At Reynolds Mountain East and Weissfluhjoch, but more prominently at Sapporo and Senator Beck, wind speed is higher in winter; it is highest in spring at the BERMS Old Aspen and Old Black Spruce. Col de Porte, Swamp Angel and Sodankylä do not show strong seasonal variability.

2.1.5 Relative Humidity

235 Humidity is measured at all of the sites except Weissfluhjoch using capacitive sensors. These sensors respond to changes in relative humidity (Anderson 1995), but vapour fluxes in models are driven by specific humidity gradients. Conversion is therefore required from relative to specific humidity. At temperatures below 0°C, there are two possible definitions of relative humidity because of the
240 different saturation vapour pressures over water and ice. Sensors calibrated following the WMO convention of reporting relative humidity with respect to water at all temperatures are used at most of the sites. The consequences of this choice (shown in Fig. 4) are not very significant at warmer sites such as Col de Porte or Sapporo but are clear in data from colder sites such as Old Jack Pine and Swamp Angel; relative humidity is never observed much above the ice saturation point for a particular
245 temperature. A chilled mirror dew point hygrometer is used at Weissfluhjoch, and reported relative humidity can reach 100% or higher even at sub-zero temperatures. In homogenizing the datasets,



relative humidity has been limited to a maximum of 100% and converted to specific humidity using the site calibrations.

Figure 4: Example scatter plots of relative humidity against temperature for four of the sites. The solid lines show ice saturation at temperatures below 0°C and water saturation above. Lines of constant specific humidity near the upper end of the data ranges are dashed.

2.2 Evaluation data

2.2.1 Snow depth and water equivalent

255

Figure 5: Monthly climatological averages of manual snow water equivalent measurements.

Figure 6: Daily climatological averages of snow depth measurements at all sites.

Although automatic sensors are increasingly being used to measure SWE, the most reliable methods to obtain snow mass are still manual (Pirazzini et al., 2018). They work by weighing snow mass in samplers of known volume or area, such as small cutters in snow pits or tubes to extract vertical snow cores. Nevertheless, such measurements are prone to errors: wet snow can stick to instruments, manual measurements can never be replicated in the same place because they are destructive, and subjectivity and skill do play a part so consistency can be hard to achieve if multiple people collect the data.

265 Replicate measurements of SWE and snow depth can be used to estimate uncertainty, which can be caused by measurement errors, spatial variability or a combination of both. At Col de Porte, three replicate weekly snow pits are available, two of which are used to calibrate automatic SWE measurements. Their mean standard deviation is 17 kg m⁻²; although standard deviation increases with increasing snow amount, it is generally less than ten percent of mean SWE. At Reynolds Mountain
270 East, there is a snow pillow next to a snow course that is visited approximately 10 to 15 times during the snow season. Root mean square difference (RMSD) between the two methods is 40 kg m⁻², for annual maximum SWE ranging from 186 kg m⁻² (1992) to 838 kg m⁻² (1989). RMSD in snow depth can be calculated at all sites as all have both automatic and manual measurements. Results are shown in Table , along with maximum and minimum peak snow depth to normalise the difference. At Senator



275 Beck, the snow pits cannot be collocated with the automated snow depth so the spatial variability of snow is intrinsic to any comparisons between the manual and automated measurements.

Table 3: Root mean square difference between manual and automatic snow depth measurements, maximum yearly snow depth and minimum yearly snow depth for all sites.

Climatological averages of measured SWE and snow depth are shown in Figs. 5 and 6 respectively.
280 Although all sites are situated in the Northern Hemisphere and only one is above the Arctic Circle, the snow season characteristics provide a diverse range of scenarios for the evaluation and development of snow models e.g. cold sites (e.g. SNB, SWA, SOD) with a well-defined snow season (snowpack building in autumn and winter, melting in spring/summer), warmer sites with occasional early- to mid-season snowmelt (CDP and SAP), and forest sites with interception of snowfall by the canopy (BERMS).

285

2.2.2 Albedo

Figure 7: Daily averaged albedo over time (a) and as a function of snow depth (b) at all sites except RME and SOD.

290 Reflected shortwave radiation is measured at all sites except Sodankylä and Reynolds Mountain East, thus allowing calculations of albedo. Daily effective albedos (Fig. 7-a) have been calculated at all sites with reflected shortwave radiation measurements by the method used by Morin et al. (2012) for Col de Porte. Hourly data are rejected during snowfall, if incoming shortwave radiation is less than 20 W m^{-2} or if reflected shortwave radiation is less than 2 W m^{-2} . For days with more than five hours of data
295 remaining after rejection, an albedo is calculated by dividing the sum of reflected shortwave radiation measurements by the sum of incoming shortwave radiation measurements.

Peak snow depth and highest albedo do not coincide at BERMS in Fig. 7-b because the highest albedo occurs when the most snow is intercepted by the canopy during the coldest month of the year (January), not at the end of the snow accumulation period. Curves of albedo against snow depth show
300 hysteresis at all of the sites, with snow cover of the same depth having lower albedo when melting than when accumulating. Albedos of melting snow can be much lower at Senator Beck and Swamp Angel than at other non-forested sites with comparable snow depths because of frequent dust storms dirtying the snow surface (Painter et al., 2012). Although it is not obvious from Fig. 7 in the absence of nearby sites without snow impurities, high concentrations of black carbon are found in the Sapporo
305 snowpack (Aoki et al., 2011). Model simulations suggest that impurities reduce albedo at Sapporo by



0.05 in winter and 0.18 during melt (Niwano et al., 2012). Information about errors and uncertainties in albedo due to incoming radiation measurements is in Section 2.1.2.

2.2.3 Surface and Soil temperature

310

Figure 8: Daily climatological averages of surface temperature (a), soil temperature (b), and differences between air and soil temperatures (c). Soil temperatures are shown at 30 cm depth at RME and at 10 cm depth at all other sites.

315 Successive IPCC reports have noted that Earth System Models often underestimate soil temperatures at high latitudes (Randall et al., 2007; Flato et al., 2013; Koven et al., 2013). This has implications for assessing the permafrost carbon feedback, i.e. the amplification of surface warming from carbon emissions released by thawing permafrost. Long term datasets are therefore essential to evaluate model performance and to improve model representations of soil / atmosphere interactions.

320 Surface temperature (Fig. 8-a) and soil temperatures (Fig. 8-b) are available at eight of the sites. Surface temperature was calculated from measured outgoing longwave radiation assuming blackbody radiation except at the Senator Beck basin sites where infrared temperature sensors measurements are used. Unlike at Col de Porte, Sapporo and Weissfluhjoch where the pyranometers measuring outgoing longwave radiation are above snow cover, the instruments at BERMS are above the canopy.

325 The strong insulating effect of snow is apparent in Fig. 8 b-c for all sites with average winter air temperatures below 0°C. Even at a shallow depth (10 cm), daily averaged winter soil temperatures remain above freezing at all sites except Old Jack Pine and Sodankylä, although the soil does freeze in some individual years. They are also the only two sites not to have soil temperatures at 10 cm depth (or 20 cm, not shown) plateau during the snow season. They do show the highest annual ranges of temperatures, with climatologically averaged winter temperatures down to -5°C and summer
330 temperatures above 12°C.

3. Large-scale meteorological forcing data for reference site simulations

335 The Land Surface, Snow and Soil Moisture Model Intercomparison Project (LS3MIP; van den Hurk et al. 2016) contribution to CMIP6 includes global offline land model experiments with meteorological forcing data provided by phase 3 of the Global Soil Wetness Project (GSWP3; Kim, 2017). GSWP3



forcing data were generated by a run of the Global Spectral Model at T248 (approximately 50 km) resolution nudged at each pressure level with meridional- and zonal-wind and air temperature from the 20th Century Reanalysis (Compo et al. 2011), followed by bias corrections using observations. All
340 of the variables required for forcing land surface models are provided on a 0.5° global grid and three-hour timesteps.

For ESM-SnowMIP, 1980-2010 forcing data have been extracted for GSWP3 grid cells containing reference sites and interpolated to one-hour timesteps. The longer time period provides more variability for investigating the sensitivity of models to trends in forcing data. These data would also
345 allow rerunning LS3MIP experiments at reference sites with models that do not have capabilities for global runs, but a complication is immediately apparent from the comparisons of site and grid data in Fig. 9. The maritime, boreal and Arctic sites (SAP, OAS, OBS, OJP, SOD) are in areas with low relief and lie close to the mean elevations of their GSWP3 grid cells, but snow study sites in mid-latitude mountains (CDP, RME, SNB, SWA, WFJ) are typically established at higher elevations with longer snow
350 seasons; most of the ESM-SnowMIP mountain sites are hundreds of metres higher than grid elevations (Fig. 9-a). Consequently, GSWP3 temperatures at the mountain sites are too high (Fig. 9-b), total precipitation is too low (Fig. 9-c), snowfall is much too low (Fig. 9-d), and bias corrections are required for model forcing.

Figure 9: Comparisons between elevations (a), temperatures (b), total precipitation (c) and snowfall (d) at ESM-SnowMIP reference sites and corresponding GSWP3 grid cells. Triangles identify mountain sites.
355

Bias corrections have been applied to all GSWP3 meteorological variables at all sites. Quantile mapping was used to correct relative humidity within the 0-100% range, but only mean biases for
360 overlapping data periods were removed from the other variables to retain the interannual and shorter variability in the large-scale forcing; the aim is to stay as close as possible to the global GSWP3 simulations without introducing gross elevation-dependent errors in site simulations. Offsets were applied to air temperature, pressure and longwave radiation data, and multipliers were applied to precipitation, wind speed and shortwave radiation data to avoid negative or spurious non-zero values.
365 Site wind speeds were first normalized to the GSWP3 10 m reference height using a logarithmic profile and an assumed 0.1 cm roughness length.

Total precipitation rate P_r in each timestep was repartitioned into snowfall rate $S_f = f_s P_r$ and rainfall rate $R_f = (1 - f_s) P_r$ depending on corrected air temperature T using a logistic curve



$$f_s = \frac{1}{1 + \exp[(T - T_0)/T_1]}$$

370 with site-dependent parameters T_0 and T_1 fitted to unadjusted GSWP3 data (Table 4). Figure 3 shows that the logistic curve fits the GSWP3 data well at all sites with the exception of Sapporo, which has the unusual feature of some precipitation at low temperatures falling as rain. Inspection of the gridded GSWP3 data shows occurrences of freezing rain in coastal areas.

Table 4: Precipitation phase factors fitted to GSWP3 data at site locations (Senator Beck and Swamp Angel are located within the same 0.5° grid cell).
375

Annual mean temperature and snowfall variations are shown in Figs. 10 and 11 for the in situ and bias-corrected GSWP3 data at all sites. Although only mean errors for the periods of overlap have been removed, there is generally good correlation between annual means of GSWP3 data and site
380 observations for overlapping years. Table gives linear trends fitted to the in situ and bias-corrected GSWP3 annual mean temperatures and snowfall. 1998-2009 observations at BERMS show decreasing temperatures and increasing snowfall after the Saskatchewan drought of the early 2000s, but there are negligible trends in the longer GSWP3 series. Sapporo also has increasing snowfall in recent years but little trend in GSWP3. Some sites show stronger warming trends in the GSWP3 data, which will be
385 useful for investigating modelled snow responses to warming.

Table 5: Trends in annual mean temperatures and snowfall from in situ and GSWP3 data. Bold trends are statistically significant (Mann-Kendal $p < 0.05$).

**Figure 10: Annual mean temperatures and fitted trends for years starting on 1 October at reference sites from GSWP3 and in situ data. Numbers show correlation (r) between GSWP3 and in situ air
390 temperature for the n complete years of overlap.**

Figure 11: Annual snowfall and fitted trends for years starting on 1 October at reference sites from GSWP3 and in situ data.

4. Discussion

395 A number of errors were identified in the datasets in the course of the study. Firstly, we noted that snowfall at the Old Aspen was much lower than at the Old Black Spruce and Old Jack Pine during the 2007 / 2008 winter. It was subsequently found that a gauge malfunction in November and December 2007 was not identified at the quality control (QC) stage. Secondly, two other errors were identified



by decomposing time series: trend analyses showed an increase in wind speed at the Senator Beck
400 basin sites from October 2012 to the end of the dataset in October 2015. Both sites measure wind
speed at two heights; the lower wind speed measurements were used for the first seventeen years of
the dataset, but the upper wind speed was accidentally used for the last three years. At the same sites,
instrument re-calibration led to a small but statistically significant increasing trend in longwave
radiation. These errors were included in the preliminary ESM-SnowMIP results shown in Krinner et al.
405 (2018); erroneous years will either be neglected in future publications or models will be forced with
the corrected datasets which are published alongside this paper (see Section 6).

While unfortunate, such errors are symptomatic of long-term data sets for which consistent
maintenance and data collection is problematic. Firstly, by definition, long-term monitoring stations
might have been installed before metadata were kept electronically (and before the word “metadata”
410 was invented in 1983; Merriam-Webster, 2018) and when information about changes of instruments
or re-calibrations were in notebooks which might never have been digitised, have now been lost or
never even existed. Equally, improvements in data storage capacities mean that temporal sampling
intervals are shorter than they were. For example, measurements at Reynolds Mountain East were
initially made every 15 minutes and averaged to hourly values; currently, 10 second samples and 5
415 minutes averages are aggregated to hourly values for most variables. Such factors are known to affect
the values of meteorological variables (Hupet and Vanclooster, 2001) but it is beyond the scope of this
study to attempt to quantify their contributions to errors or variations in the datasets. Secondly,
immediate use of the data allows instrument malfunctions to be identified quickly. For example, a
power supply failure was not identified at Sodankylä for 52 days in September and October 2011
420 because data were being collected but not used; more frequent QC checks are now in place. Thirdly,
long-term monitoring stations are susceptible to funding cycles and to changes in climate change
policies by successive governments. For example, the BERMS sites, which were established in 1994,
had the most frequent site visits from 2001 to 2008, but changing priorities led to less frequent snow
surveys after 2008 with only one in the 2009/2010 snow season. Finally, while automated QC
425 protocols are in place, some checks require a subjective interpretation of the data and can therefore
depend on just one person to identify errors due to malfunction, snow deposition on instruments etc.
Reliance on subjectivity or local knowledge – which in some cases is advocated as mentioned in
Section 2.1.3 to choose the best method to correct undercatch in precipitation – diminishes the
likelihood of the dataset being reproducible. In a discipline like geoscience where uncertainties and
430 errors are required to be quantifiable, it is important to acknowledge that subjectivity is not. The
closest estimate comes from a survey in which more than 40% of scientists in the field of Earth and
Environment admitted to failing to have reproduced their own experiments (Baker, 2016); the figure



increased to more than 60% when trying to reproduce other researchers' experiments. Nevertheless,
human errors, or more appropriately "mistakes", are not exclusive to data processing: Menard et al.
435 (2015) identified mistakes in the description files of the land surface model JULES that caused it to
underperform considerably.

A recent and growing push towards standardising methods for data sharing and publishing may lead
to errors being identified more systematically as more people have access to data. One of the
advantages of open source software is that bugs are reported by users and their correction is, at times,
440 a community effort which allows software to be improved quickly (Wu et al., 2016). Sharing of
geoscientific models' source code, although still a fairly recent development compared to the field of
engineering software, has equally led to model improvements through the identification and fixing of
bugs beyond the model development teams (David et al., 2016; Samuel Morin, personal
communication about the Crocus snow model). One might expect a similar trend for data sharing
445 where identifying errors becomes an asset to the community because, as mentioned by Gil et al.
(2016) in their proposal for a framework for best practices in the publication of data papers, "data
sharing makes authors double-check their work, improving science at the first stage as well as future
reuse". The more data are used, the more likely it is that mistakes, errors and uncertainties are
identified, and the less likely it will be that model results can, according to Clark et al. (2011) "at best
450 be merely attributed to a nebulous mix of data and structural errors"; to this we can also add human
errors.

5. Conclusion

It is hoped that one of the legacies of ESM-SnowMIP will be for the datasets presented in this paper
455 to be used as benchmarks for model development and to facilitate improvements in snow modelling.
Cold region processes have been a major source of uncertainties in previous IPCC reports. The sparsity
of long-term high quality datasets in cold regions in the past may have contributed to this if one
considers that ESMs are run globally but their snow schemes are generally evaluated at a small
number of sites; the first iteration of SnowMIP (Etchevers et al., 2002) sixteen years ago included only
460 one long-term (15-year) dataset and three short-term (less than two snow seasons) ones. Meta-
research argues that it is misleading to emphasize statistically significant findings of any single team;
what matters instead is the totality of the evidence (Ioannidis, 2005). It is equally misleading to draw
conclusions on model performance when models are evaluated only at one or two sites for one or two
years. The ease-of-use and availability of the datasets presented here, as well as further ESM-



465 SnowMIP reference sites which will be located in more challenging conditions, should help model
developers quantify – and reduce – model uncertainties and errors.

6. Data availability and archiving

The data presented and described in this paper are available in the data repository PANGAEA:
470 <https://doi.pangaea.de/10.1594/PANGAEA.897575>

Author Contribution

CM led the writing of the paper and archived the data. RE prepared the data in the standardized format
described in Section 6, bias-corrected GSWP3 data and wrote Section 3. The supplementary material
475 (metadata) and data provision are attributable to: MD and YL for CDP, AB and PB for BERMS, DM for
RME, MN for SAP, JD and MR for SNB and SWA, AK and RE for SOD and CF for WFJ. HK provided GSWP3
data, which were extracted for sites and interpolated by LW. All co-authors provided comments which
contributed to the paper.

480 Acknowledgments

The authors would like to thank all the staff and students who have collected the data presented in
this paper over the years. Work by CM and RE was supported by NERC grant NE/P011926/1.
CNRM/CEN is part of Labex OSUG@2020 (ANR10 LABX56). CDP is part of Observatoire des Sciences
de l'Univers de Grenoble (OSUG), Observation pour l'Experimentation et la Recherche en
485 Environnement CryObsClim and Systemes d'Observation et d'Experimentation au long terme pour la
Recherche en Environnement des glaciers, GlacioClim. CDP contributes to OZCAR (Observatoires de la
Zone Critique Applications et Recherches), one of the French components of the eLTER European
Research Infrastructure (International Long-term Ecological Research Networks). It is also labeled as a
member of the World Meteorological Observation Global Cryospheric Watch Cryonet network and of
490 the INARCH network. H. Kim acknowledges support by Grant-in-Aid for Specially promoted Research
16H06291 from JSPS. M. Niwano was supported in part by (1) the Japan Society for the Promotion of
Science through Grants-in-Aid for Scientific Research numbers JP16H01772 (SIGMA project),
JP15H01733 (SACURA project), JP17K12817, JP17KK0017, JP18H03363, and JP18H05054; (2) the
Ministry of the Environment of Japan through the Experimental Research Fund for Global
495 Environmental Research Coordination System; (3) the Institute of Low Temperature Science, Hokkaido



University, through the Grant for Joint Research Program (18S007 and 18G035). The sites at the Senator Beck Basin are maintained by the Center for Snow and Avalanche Studies with development funding from the U.S. National Science Foundation (ATM-0431955) and the USDA-Forest Service. A. Barr and P. Bartlett acknowledge financial support from the Climate Research Division of Environment and Climate Change Canada, and field and data management support from Joe Eley, Charmaine Hrynkiw, Dell Bayne, Natasha Neumann, Erin Thompson and Steve Enns. Sodankylä is a member of WMO Global Cryosphere Watch Cryonet network.

References

- 505 Anderson, P.: Mechanism for the behaviour of hydroactive materials used in humidity sensors, *Journal of Atmospheric and Oceanic Technology*, 12, 662-667, 2005.
- Aoki, T., Kuchiki, K., Niwano, M., Kodama, Y., Hosaka, M. and Tanaka, T.: Physically based snow albedo model for calculating broadband albedos and the solar heating profile in snowpack for general circulation models, *J. Geophys. Res.*, 116, D11114, doi:10.1029/2010JD015507, 2011.
- 510 Baker M.: 1,500 scientists lift the lid on reproducibility. *Nature*, 533:452–4, doi:10.1038/533452a, 2016.
- Clarke, M.P.A, Kavetski, D. and Fenicia, F.: Pursuing the method of multiple working hypotheses for hydrological modeling. *Water Resources Research*, 47, W09301, doi:10.1029/2010WR009827, 2011.
- 515 Compo, G.P., Whitaker, J.S., Sardeshmukh, P.D., Matsui, N., Allan, R.J., Yin, X., Gleason, B.E., Vose, R.S., Rutledge, G., Bessemoulin, P., Brönnimann, S., Brunet, M., Crouthamel, R.I., Grant, A.N., Groisman, P.Y., Jones, P.D., Kruk, M., Kruger, A.C., Marshall, G.J., Maugeri, M., Mok, H.Y., Nordli, Ø., Ross, T.F., Trigo, R.M., Wang, X.L., Woodruff, S.D. and Worley, S.J.: The Twentieth Century Reanalysis Project. *Quarterly J. Roy. Meteorol. Soc.*, 137, 1-28. DOI: 10.1002/qj.776, 2011.
- 520 David, C., Famiglietti, J., Yang, Z-L., Habets, F., Maidment, D.: A decade of RAPID—Reflections on the development of an open source geoscience code, *Earth and Space Science*, 3:226-244, doi:10.1002/2015EA000142, 2016.
- Essery, R., Pomeroy, J., Parviainen, J., and Storck, P.: Sublimation of Snow from Coniferous Forests in a Climate Model, *Journal of Climate*, 16:1855–1864, doi:10.1175/1520-0442(2003)016<1855:SOSFCF>2.0.CO;2, 2003.



- 525 Essery, R., Rutter, N., Pomeroy, J., Baxter, R., Stähli, M., Gustafsson, D., Barr, A., Bartlett, P., and Elder, K.: SNOWMIP2: An Evaluation of Forest Snow Process Simulations, *B. Am. Meteorol. Soc.*, 90, 1120–1135, doi:10.1175/2009BAMS2629.1, 2009.
- Essery, R., Kontu, A., Lemmetyinen, J., Dumont, M., and Menard, C. B.: A 7-year dataset for driving and evaluating snow models at an Arctic site (Sodankylä, Finland), *Geosci. Instrum. Method. Data Syst.*, 5, 530 219–227, doi:10.5194/gi-5-219-2016, 2016.
- Etchevers, P., Martin, E., Brown, R., Fierz, C., Lejeune, Y., Bazile, E., Boon, A., Dai, Y.-J., Essery, R., Fernandez, A., Gusev, Y., Jordan, R., Koren, V., Kowalczyk, E., Nasonova, R., Pyles, D., Schlosser, A., Shmakin, A., Smirnova, T.G., Strasser, U., Verseghy, D., Yamazaki, T., and Yang, Z.-L.: SnowMIP, an intercomparison of snow models: first results. In: *Proceedings of the International snow science workshop*, Penticton, Canada, 29 Sep.-4 Oct., 2002. 535
- Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S., Collins, W., Cox, P., Driouech, F., Emori, S., Eyring, V., Forest, C., Gleckler, P., Guilyardi, E., Jakob, C., Kattsov, V., Reason, C., and Rummukainen, M.: Evaluation of climate models. In Stocker, T., Qin, D., Plattner, G., K., Tignor, M., Allen, S., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P., editors, *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, 2013. 540
- Georges C., and Kaser, G.: Ventilated and unventilated air temperature measurements for glacier-climate studies on a tropical high mountain site, *Journal of Geophysical Research Atmospheres*, 107, ACL15-1:ACL15-10, doi:10.1029/2002JD002503, 2002.
- 545 Gil, Y., David, C., Demir, I., Essawy, B., Fulweiler, R., Goodall, J., Karlstrom, L., Lee, H., Mills, H., Oh, J-H, Pierce, S., Pope, A., Tzeng, M., Villazimar, S and Yu, X.: Toward the Geoscience Paper of the Future: Best practices for documenting and sharing research from data to software to provenance, *Earth and Space Science*, 3, 388–415, doi:10.1002/2015EA000136, 2016.
- Goodison, B.E., Louie, P.Y.T., and Yang, D.: WMO solid precipitation measurement intercomparison. 550 WMO Instruments and Observing Methods Rep. 67, WMO/TD-872, 212 pp, 1998.
- Hupet, F. and Vanclooster, M.: Effect of the sampling frequency of meteorological variables on the estimation of the reference evapotranspiration, *Journal of Hydrology*, 243:192–204, doi:10.1016/S0022-1694(00)00413-3, 2001.



Huwald, H., Higgins, C. W., Boldi, M.-O., Bou-Zeid, E., Lehning, M., and Parlange, M. B.: Albedo effects
555 on radiative errors in air temperature measurements, *Water Resources Research*, 45, W08431,
doi:10.1029/2008WR007600, 2009.

Ioannidis J.P.A.: Why Most Published Research Findings Are False. *PLoS Med* 2(8): e124.,
doi:10.1371/journal.pmed.0020124, 2005.

Kim, H.: Global Soil Wetness Project Phase 3 Atmospheric Boundary Conditions (Experiment 1) [Data
560 set]. Data Integration and Analysis System (DIAS), doi:10.20783/DIAS.501, 2017.

Köppen, W.: Das geographische System der Klimate. In Köppen, W., Geiger, R. (Eds) *Handbuch der
Klimatologie*, Berlin, Borntraeger, 1936.

Koven, C.D., Riley, W.J., and Stern, A.: Analysis of Permafrost Thermal Dynamics and Response to
Climate Change in the CMIP5 Earth System Models. *Journal of Climate*, 26, 1877–1900,
565 doi:10.1175/JCLI-D-12-00228.1, 2013.

Krinner, G., Derksen, C., Essery, R., Flanner, M., Hagemann, S., Clark, M., Hall, A., Rott, H., Brutel-
Vuilmet, C., Kim, H., Menard, C. B., Mudryk, L., Thackeray, C., Wang, L., Arduini, G., Balsamo, G.,
Bartlett, P., Boike, J., Boone, A., Chéruy, F., Colin, J., Cuntz, M., Dai, Y., Decharme, B., Derry, J.,
Ducharne, A., Dutra, E., Fang, X., Fierz, C., Ghattas, J., Gusev, Y., Haverd, V., Kontu, A., Lafaysse, M.,
570 Law, R., Lawrence, D., Li, W., Marke, T., Marks, D., Nasonova, O., Nitta, T., Niwano, M., Pomeroy, J.,
Raleigh, M. S., Schaedler, G., Semenov, V., Smirnova, T., Stacke, T., Strasser, U., Svenson, S., Turkov,
D., Wang, T., Wever, N., Yuan, H., and Zhou, W.: ESM-SnowMIP: Assessing models and quantifying
snow-related climate feedbacks, *Geosci. Model Dev. Discuss.*, doi:10.5194/gmd-2018-153, 2018.

Landry, C. C., Buck, K.A., Raleigh, M.S., and Clark, M.P.: Mountain system monitoring at Senator Beck
575 Basin, San Juan Mountains, Colorado: A new integrative data source to develop and evaluate models
of snow and hydrologic processes, *Water Resources Research*, 50, 1773–1788,
doi:10.1002/2013WR013711, 2014.

Lejeune, Y., Dumont, M., Panel, J.-M., Lafaysse, M., Lapalus, P., Le Gac, E., Lesaffre, B., and Morin, S.:
580 57 years (1960–2017) of snow and meteorological observations from a mid-altitude mountain site (Col
de Porte, France, 1325 m alt.), *Earth Syst. Sci. Data Discuss.*, doi:10.5194/essd-2018-84, Accepted,
2018.

Menard, C.B. and Essery, R.: ESM-SnowMIP meteorological and evaluation datasets at ten reference
sites (in situ and bias corrected reanalysis data), PANGAEA,
<https://doi.pangaea.de/10.1594/PANGAEA.897575>, 2019.



585 Menard, C.B., Ikonen, J., Rautiainen, K., Aurela, M., Arslan, A.N., and Pulliainen, J.: Effects of Meteorological and Ancillary Data, Temporal Averaging, and Evaluation Methods on Model Performance and Uncertainty in a Land Surface Model. *J. Hydrometeor.*, 16, 2559–2576, doi:10.1175/JHM-D-15-0013.1, 2015.

Merriam-Webster: Metadata | Definition of Metadata by Merriam-Webster, <https://www.merriam-webster.com/dictionary/metadata#h1>, Publication date unavailable, Accessed on 18/12/2018.

590 Morin, S., Lejeune, Y., Lesaffre, B., Panel, J.-M., Poncet, D., David, P., and Sudul, M.: An 18-yr long (1993–2011) snow and meteorological dataset from a mid-altitude mountain site (Col de Porte, France, 1325 m alt.) for driving and evaluating snowpack models, *Earth Syst. Sci. Data*, 4, 13–21, doi:10.5194/essd-4-13-2012, 2012.

595 Nitu, R., Roulet, Y.-A., Wolff, M., Earle, M., Reverdin, A., Smith, C., Kochendorfer, J., Morin, S., Rasmussen, R., Wong, K., Alastrué, J., Arnold, L., Baker, B., Buisán, S., Collado, J.L., Colli, M., Collins, B., Gaydos, A., Hannula, H.-R., Hoover, J., Joe, P., Kontu, A., Laine, T., Lanza, L., Lanzinger, E., Lee, G.W., Lejeune, Y., Leppänen, L., Mekis, E., Panel, J.-M., Poikonen, A., Ryu, S., Sabatini, F., Theriault, J., Yang, D., Genthon, C., van den Heuvel, F., Hirasawa, N., Konishi, H., Nishimura, K., and Senese, A.: WMO
600 Solid Precipitation Intercomparison Experiment (SPICE) (2012 - 2015), World Meteorological Organization Instruments and Observing Methods Report No. 131, 2018.

Niwano, M., Aoki, T., Kuchiki, K., Hosaka, M., and Kodama Y.: Snow Metamorphism and Albedo Process (SMAP) model for climate studies: Model validation using meteorological and snow impurity data measured at Sapporo, Japan, *J. Geophys. Res.*, 117, F03008, doi:10.1029/2011JF002239, 2012.

605 Painter, T. H., Skiles, S. M., Deems, J.S., Bryant, A.C., and Landry C.C.: Dust radiative forcing in snow of the Upper Colorado River Basin: 1. A 6 year record of energy balance, radiation, and dust concentrations, *Water Resour. Res.*, 48, W07521, doi:10.1029/2012WR011985, 2012.

Pirazzini, R., Leppänen, L., Picard, G., Lopez-Moreno, J.I., Marty, C., Macelloni, G., Kontu, A., von Lerber, A., Tanis, C.M., Schneebeli, M., de Rosnay, P., Arslan, A.N.: European In-Situ Snow
610 Measurements: Practices and Purposes, *Sensors*, 18, 2018.

Randall, D.A., Wood, R.A., Bony, S., Colman, R., Fichet, T., Fyfe, J., Kattsov, V., Pitman, A., Shukla, J., Srinivasan, J., Stouffer, R.J., Sumi, A. and Taylor, K.E.: Climate Models and Their Evaluation. In: *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* [Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K.B., Tignor, M., and Miller H.L. (eds.)]. Cambridge University Press,
615 Cambridge, United Kingdom and New York, NY, USA, 2007.



Reba, M. L., Marks, D., Seyfried, M., Winstral, A., Kumar, M., and Flerchinger, G.: A long-term data set for hydrologic modeling in a snow-dominated mountain catchment, *Water Resources Research*, 47, W07702, doi:10.1029/2010WR010030, 2011.

620 Rutter, N., Essery R.L.H., Pomeroy, J.W. and 48 others: Evaluation of forest snow processes models (SnowMIP2), *J. Geophys. Res.*, 114, D06111, doi:10.1029/2008JD011063, 2009.

Tenopir C., Allard S., Douglass K., Aydinoglu A.U., Wu L., Read, E., Manoff, M. and Frame, M.: Data Sharing by Scientists: Practices and Perceptions. *PLoS ONE* 6(6): e21101. doi:10.1371/journal.pone.0021101, 2011.

625 van den Hurk, B., Kim, H., Krinner, G., Seneviratne, S. I., Derksen, C., Oki, T., Douville, H., Colin, J., Ducharne, A., Cheruy, F., Viovy, N., Puma, M. J., Wada, Y., Li, W., Jia, B., Alessandri, A., Lawrence, D. M., Weedon, G. P., Ellis, R., Hagemann, S., Mao, J., Flanner, M. G., Zampieri, M., Matera, S., Law, R. M., and Sheffield, J.: LS3MIP (v1.0) contribution to CMIP6: the Land Surface, Snow and Soil moisture Model Intercomparison Project – aims, setup and expected outcome, *Geosci. Model Dev.*, 9, 2809-630 2832, doi:10.5194/gmd-9-2809-2016, 2016.

Wever, N., Schmid, L., Heilig, A., Eisen, O., Fierz, C., and Lehning, M.: Verification of the multi-layer SNOWPACK model with different water transport schemes, *The Cryosphere*, 9, 2271-2293, doi:10.5194/tc-9-2271-2015, 2015.

635 Wu, H., Shi, L., Chen, C., Wang, Q., Boehm, B.: Maintenance Effort Estimation for Open Source Software: A Systematic Literature Review, 2016 IEEE International Conference Software Maintenance and Evolution (ICSME), pp. 32-43, 2016.

640

645



Tables

Site	Short name	Data provider	Reference paper
Col de Porte France	CDP	Météo-France, France	Morin et al. (2012)* Lejeune et al. (2018)*
Old Aspen, Old Black Spruce Old Jack Pine Saskatchewan, Canada	OAS OBS OJP (BERMS collectively)	Environment and Climate Change Canada, Canada	Bartlett et al. (2006)
Reynolds Mountain East Idaho, USA	RME	USDA Agricultural Research Service, USA	Reba et al. (2011)*
Sapporo Japan	SAP	Meteorological Research Institute, Japan Meteorological Agency, Japan	Niwano et al. (2012)
Senator Beck Swamp Angel Colorado, USA	SNB SWA (Senator Beck basin collectively)	Center for Snow and Avalanche Studies, USA	Landry et al. (2014)*
Sodankylä Finland	SOD	Finnish Meteorological Institute, Finland	Essery et al. (2015)*
Weissfluhjoch Switzerland	WFJ	WSL Institute for Snow and Avalanche Research, Switzerland	Wever et al. (2015)

Table 1: Data ownership and reference papers for the sites. Asterisks denote dedicated data description papers; the others are modelling papers in which a short description of a site is included.

650

655



Short name	Latitude (°N)	Elevation (m)	Vegetation type	Soil type	Snow cover classification	Köppen climate classification
CDP	45.30	1325	Grassy meadow surrounded by coniferous forest	Sandy clay loam	Alpine	Warm-summer humid continental climate
OAS	53.63	600	21 m high aspen forest. Thick understory of 2 m high hazelnut.	10 cm organic litter and peat over sandy clay loam	Taiga	Warm-summer humid continental climate
OBS	53.99	629	12 m high black spruce forest. Sparse understory.	Peat over sand and sandy loam	Taiga	Warm-summer humid continental climate
OJP	53.92	579	14 m high forest. Sparse understory.	Sand	Taiga	Warm-summer humid continental climate
RME	43.19	2060	Clearing (short grass) in an alpen/fir grove	Silty clay	Alpine	Warm-summer humid continental climate
SAP	43.08	15	Short grass	Clay	Maritime	Hot summer continental climates
SNB	37.91	3714	Alpine tundra	Thin soil and exposed bedrock	Alpine	Polar and alpine (montane) climates
SOD	67.37	179	Clearing (short heather and lichen) in coniferous forest	Sand	Taiga	Subarctic climate
SWA	37.91	3371	Clearing (short grass) in subalpine forest	Colluvium	Alpine	Subarctic climate
WFJ	46.83	2536	Barren	Moraine	Alpine	Polar and alpine (montane) climates

660 *Table 2: Geographic characteristics of the ten sites.*



Sites	RMSD (m)	Max peak yearly Snow depth (m)		Min peak yearly snow depth (m)	
		Manual	Automatic	Manual	Automatic
CDP	0.11	2.09	2.03	0.53	0.60
OAS	0.06	0.60	0.68	0.32	0.34
OBS	0.05	0.61	0.55	0.29	0.30 670
OJP	0.06	0.54	0.61	0.25	0.31
RME	0.08	2.02	2.14	1.06	1.02
SAP	0.08	1.22	1.20	0.62	0.52
SNB	0.27	2.37	2.30	1.52	1.37
SOD	0.04	1.03	1.02	0.65	0.61
SWA	0.11	2.66	2.90	1.66	1.78
WFJ	0.05	3.56	2.95	1.82	1.82

675 Table 3: Root mean square difference between manual and automatic snow depth measurements, maximum yearly snow depth and minimum yearly snow depth for all sites.

Site	T_0 (°C)	T_1 (°C)
CDP	3.08	1.13
OAS	-1.73	1.63
OBS	-1.14	1.81
OJP	-1.32	1.76
RME	-2.00	1.48
SAP	3.72	1.48
SNB / SWA	-3.01	2.05
SOD	2.52	1.16
WFJ	0.39	1.47

680 Table 4: Precipitation phase factors fitted to GSWP3 data at site locations (Senator Beck and Swamp Angel are located within the same 0.5° grid cell).

Site	Temperature trend (°C/year)		Snowfall trend (%/year)	
	In situ	GSWP3	In situ	GSWP3
CDP	0.01	0.04	-0.56	-1.25
OAS	-0.11	0.01	1.54	-0.02
OBS	-0.16	0.01	3.98	-0.07
OJP	-0.15	0.01	2.02	-0.07
RME	0.02	0.06	0.31	-1.42
SAP	0.02	0.04	5.01	-0.06
SNB	0.05	0.01	-2.10	-0.55
SOD	0.08	0.07	0.05	-0.29
SWA	0.05	0.01	-1.56	-0.53
WFJ	0.03	0.03	-1.47	-0.88

Table 5: Trends in annual mean temperatures and snowfall from in situ and GSWP3 data. Bold trends are statistically significant (Mann-Kendal $p < 0.05$).



685 Figures

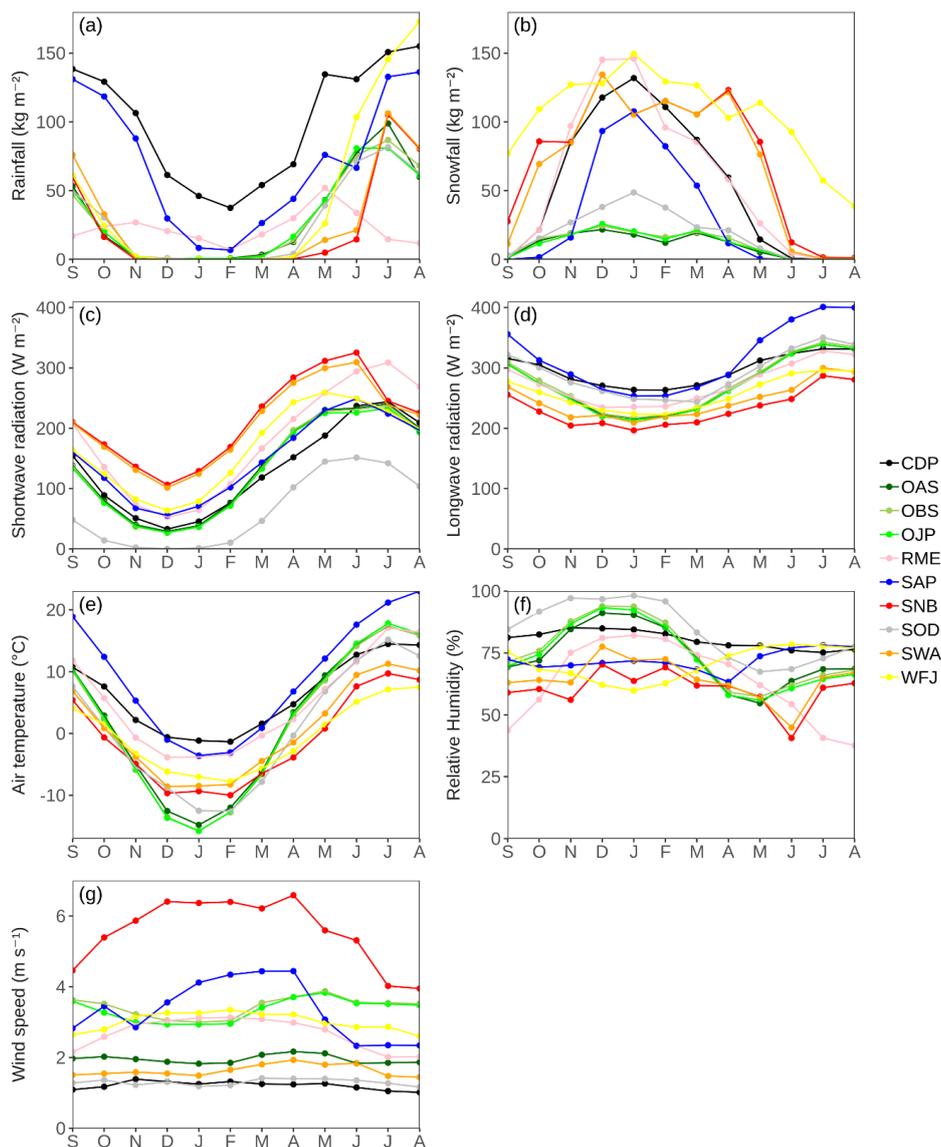
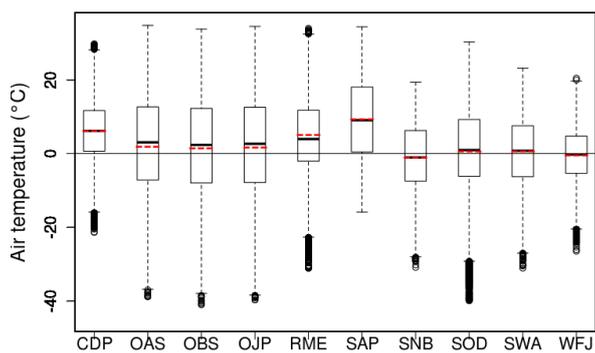


Figure 1: Climatological monthly averaged meteorological forcing data. Wind speeds at all sites are normalised at 10 m height.



690 *Figure 2: Boxplots of air temperature including means (red dashed line) at all sites. Outliers beyond 1.5 times the interquartile range (25th to 75th percentiles) are marked with circles.*

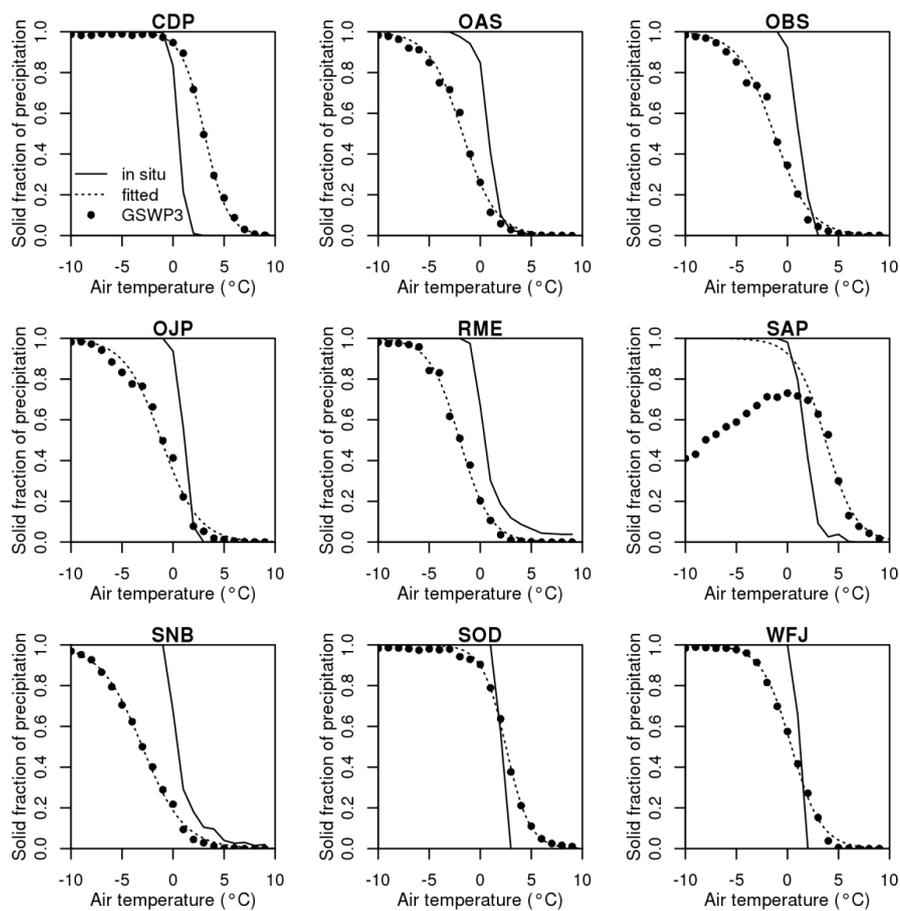
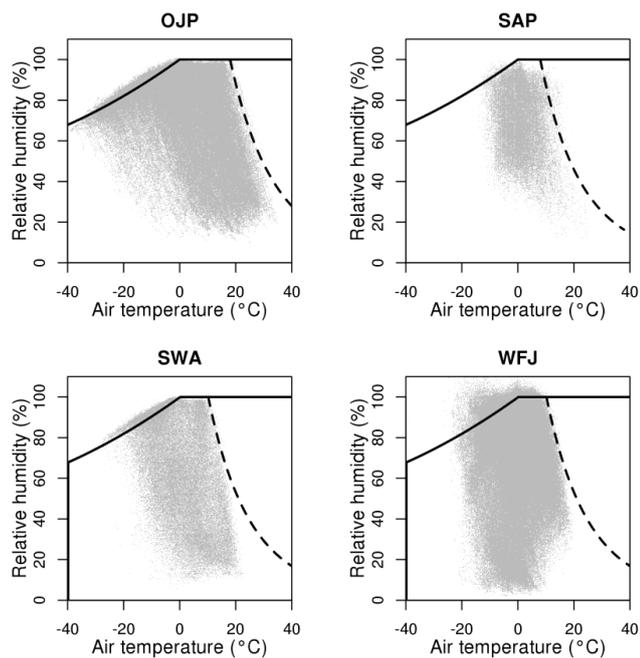
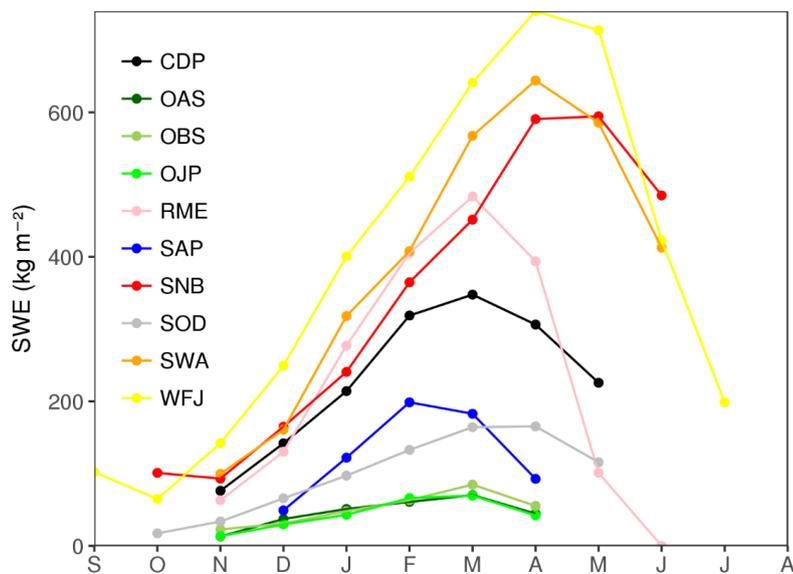


Figure 3: Fraction of precipitation falling as snow at different temperatures, as imposed on the *in situ* data and fitted to the GSWP3 data.



695

Figure 4: Example scatter plots of relative humidity against temperature for four of the sites. The solid lines show ice saturation at temperatures below 0°C and water saturation above. Lines of constant specific humidity near the upper end of the data ranges are dashed.



700

Figure 5: Monthly climatological averages of manual snow water equivalent measurements.

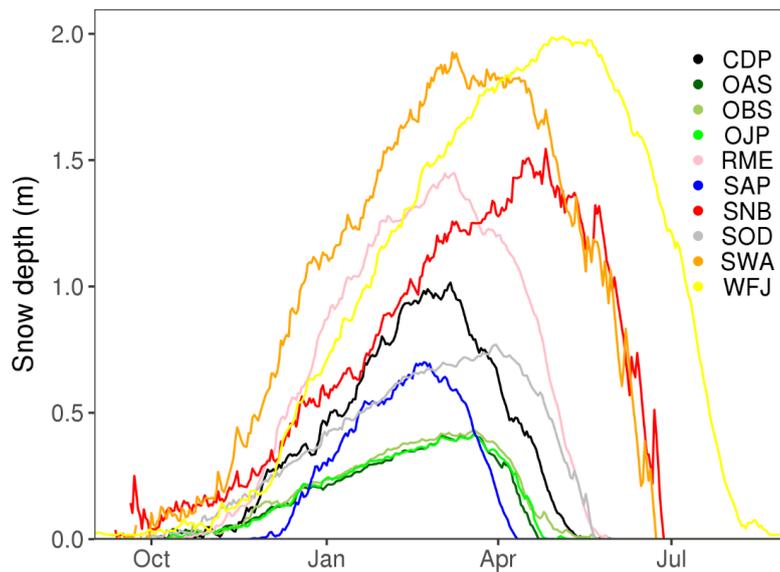
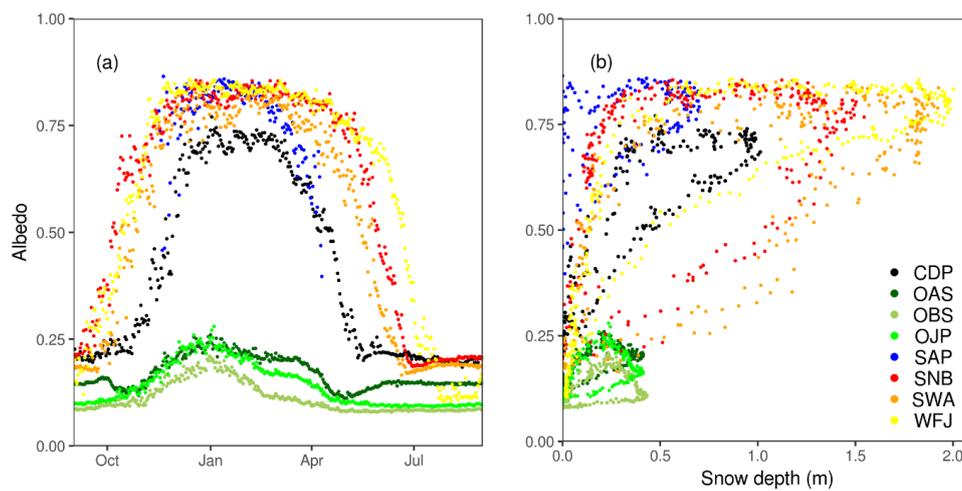
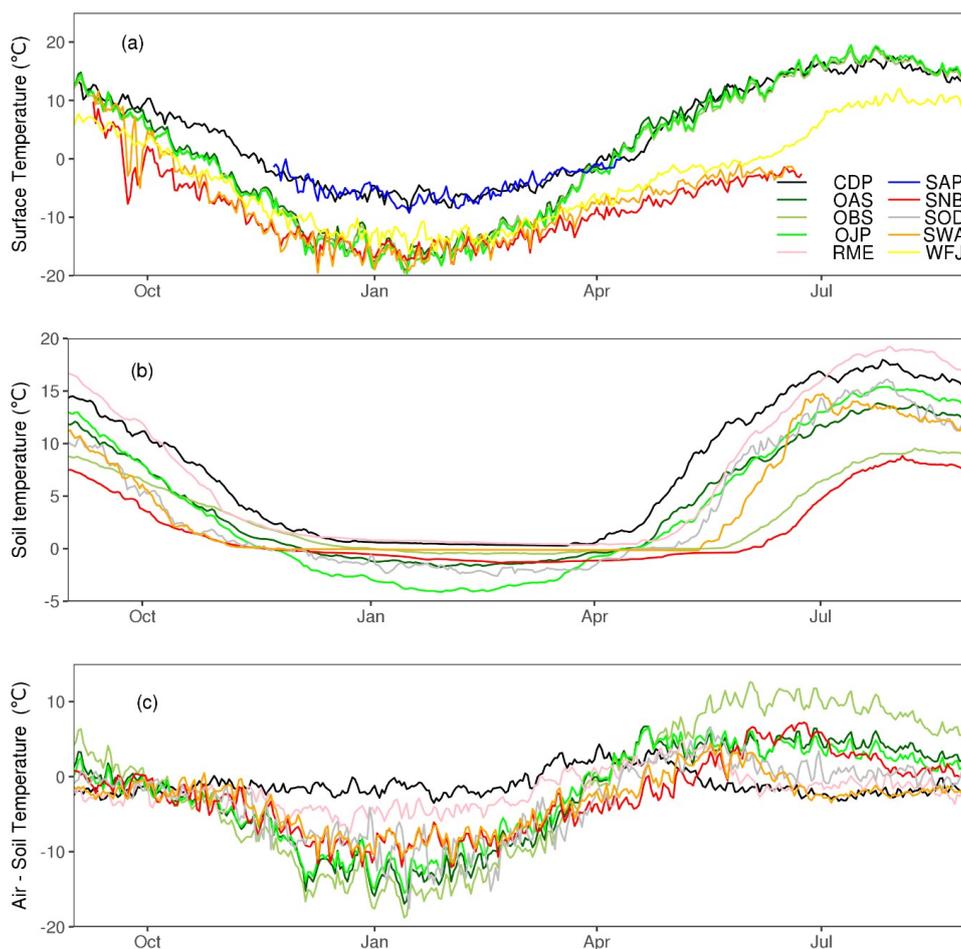


Figure 6: Daily climatological averages of snow depth measurements at all sites.

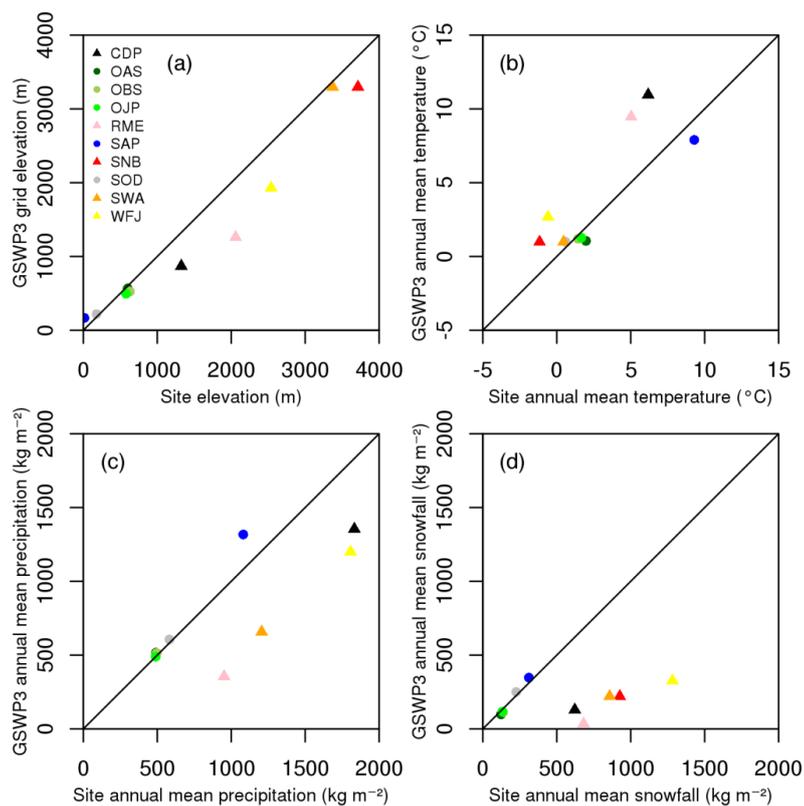


705

Figure 7: Daily averaged albedo over time (a) and as a function of snow depth (b) at all sites except RME and SOD.



710 *Figure 8: Daily climatological averages of surface temperature (a), soil temperature (b), and differences between air and soil temperatures (c). Soil temperatures are shown at 30 cm depth at RME and at 10 cm depth at all other sites.*



715 *Figure 9: Comparisons between elevations (a), temperatures (b), total precipitation (c) and snowfall (d) at ESM-SnowMIP reference sites and corresponding GSWP3 grid cells. Triangles identify mountain sites.*

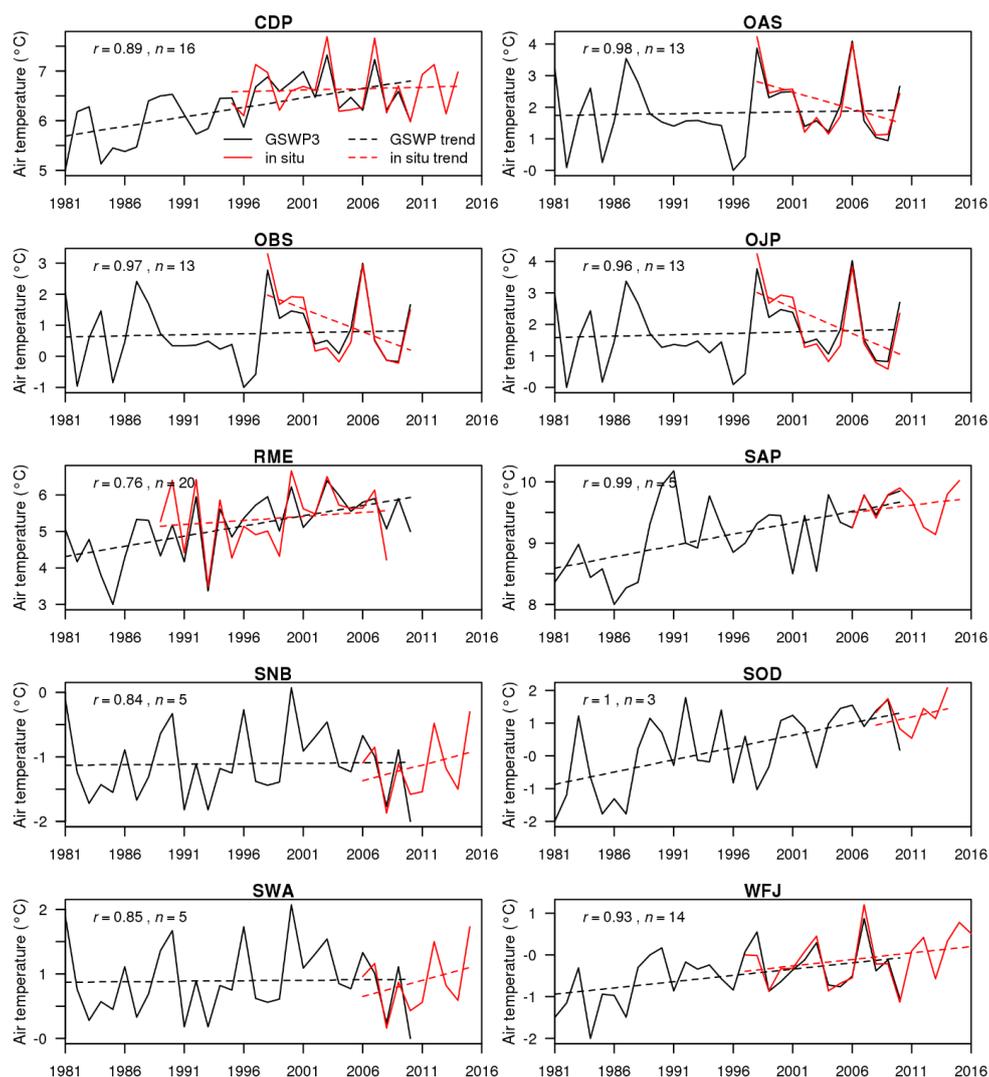


Figure 10: Annual mean temperatures and fitted trends for years starting on 1 October at reference sites from GSWP3 and in situ data. Numbers show correlation (r) between GSWP3 and in situ air temperature for the n complete years of overlap.

720

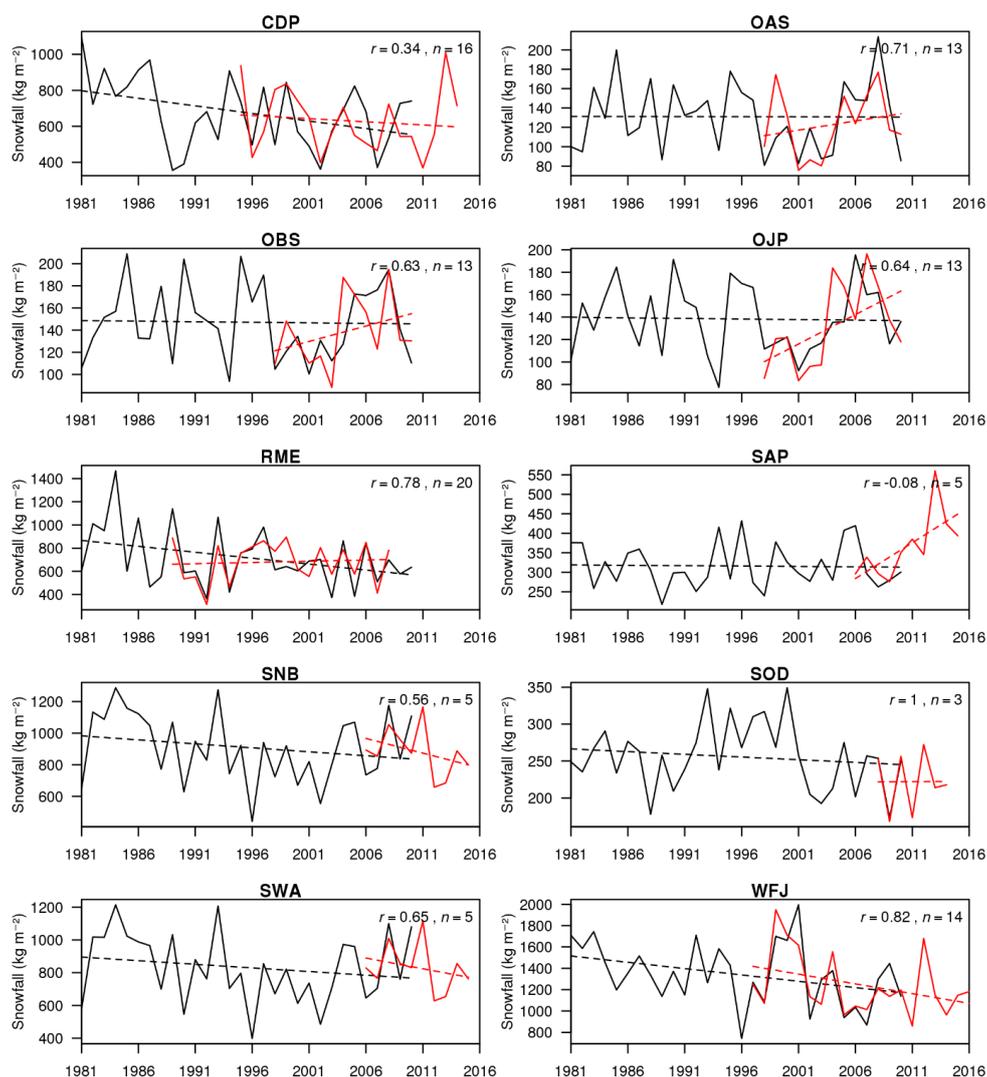


Figure 11: Annual snowfall and fitted trends for years starting on 1 October at reference sites from GSWP3 and in situ data. Numbers show correlation (r) between GSWP3 and in situ snowfall for the n complete years of overlap. The legend is as in Figure 10.

725