Published in "Permafrost and Periglacial Processes 28(1): 275–285, 2017"
which should be cited to refer to this work.

Gap-Filling Algorithm for Ground Surface Temperature Data Measured in Permafrost and Periglacial Environments

Benno Staub,[1]* *Andreas Hasler,*[1,2] *Jeannette Noetzli*[3,4] *and Reynald Delaloye*[1]

[1] University of Fribourg, Department of Geosciences, Geography, Fribourg, Switzerland
[2] SensAlpin GmbH, Davos Dorf, Switzerland
[3] University of Zurich, Department of Geography, Zurich, Switzerland
[4] WSL Institute for Snow and Avalanche Research SLF, Unit Snow and Permafrost, Davos Dorf, Switzerland

ABSTRACT

Ground surface temperatures (GST) are widely measured in mountain permafrost areas, but their time series data can be interrupted by gaps. Gaps complicate the calculation of aggregates and indices required for analysing temporal and spatial variability between loggers and sites. We present an algorithm to estimate daily mean GST and the resulting uncertainty. The algorithm is designed to automatically fill data gaps in a database of several tens to hundreds of time series, for example, the Swiss Permafrost Monitoring Network (PERMOS). Using numerous randomly generated artificial gaps, we validated the performance of the gap-filling routine in terms of (1) the bias resulting on annual means, (2) thawing and freezing degree-days, and (3) the accuracy of the uncertainty estimation. Although quantile mapping provided the most reliable gap-filling approach overall, linear interpolation between neighbouring values performed equally well for gap durations of up to 3–5 days. Finding the most similar regressors is crucial and also the main source of errors, particularly because of the large spatial and temporal variability of ground and snow properties in high-mountain terrains. Applying the gap-filling technique to the PERMOS GST data increased the total number of complete hydrological years available for analysis by 70 per cent (>450-filled gaps), likely without exceeding a maximal uncertainty of ± 0.25 °C in calculated annual mean values.

KEY WORDS: mountain permafrost; ground surface temperatures (GST); long-term monitoring; gap-filling; uncertainty estimation; Swiss Alps

INTRODUCTION

Ground surface temperatures (GST) are widely measured to investigate the spatial variability of thermal conditions in potential permafrost areas (Bonnaventure and Lewkowicz, 2008; Etzelmüller *et al*., 2007; Guglielmin, 2006; Hoelzle *et al*., 1999; Wu *et al*., 2013). Indices such as mean annual ground surface temperatures (MAGST) and thawing and freezing degree-days (TDD, FDD, respectively) are frequently used to compare sites and to reveal interannual changes (Gubler *et al*., 2011; Isaksen *et al*., 2011; Luetschg *et al*., 2008), especially in mountainous terrains, where topo-climatic and snow characteristics vary greatly. To calculate such aggregates or indices reasonably requires uninterrupted time series. Although many GST time series within the Swiss Permafrost Monitoring Network (PERMOS) cover 10–15 years (PERMOS, 2013), the majority of the data series contain gaps of hours to years. To increase the number of complete time series, the data need to be resampled to a common time interval (e.g. daily means) and gaps filled. Ideally, estimated values of the missing data should be combined with a statistical estimate of their uncertainties to consider the propagation of these uncertainties for calculating aggregates and indices.

This paper presents a fully automatic algorithm to treat gaps in large sets of GST data. It considers typical terrain and snow characteristics of high-mountain areas as well as differences related to thermistors and their placement in

* Correspondence to: B. Staub, University of Fribourg, Department of Geosciences, Geography, Ch. du Musee 4, CH-1700, Fribourg, Switzerland.
E-mail: benno.staub@unifr.ch

the field. The algorithms use daily means as for input and output to ensure a consistent calculation. Gap-filled, synthesised output data should, however, only be used for aggregates and not for analysing features at the daily scale (e.g. to derive additional information such as snow characteristics) (Schmid *et al*., 2012; Staub and Delaloye, 2016). We do not intend to carry out reanalysis of past GST variations nor use physically based models, and in principle, we prefer simple techniques. We validate the applicability of our approach using artificial gaps in the PERMOS GST data.

DATA DESCRIPTION

Characteristics of GST Time Series in Permafrost

The seasonal pattern and interannual variations of GST are dominated by snow characteristics (Delaloye, 2004; Gisnås *et al*., 2014; Park *et al*., 2014; Staub and Delaloye, 2016; Zhang, 2005). When the ground is snow covered, GST usually remain close to 0 °C at locations without permafrost and/or where fine-grained and moist substrates occur. By contrast, when intense ground cooling occurs (e.g. on ridges less influenced by snow or in coarse-blocky and porous ground material favouring convection), GST may be subject to large interannual variations in FDD or MAGST (Gisnås *et al*., 2014; Gubler *et al*., 2011; Hasler *et al*., 2015). During the snow-free period, air temperature is the dominating factor and the differences among GST time series are much smaller (Guglielmin *et al*., 2012). The remaining spatial variability mainly results from radiation and shading effects, the placement depth of the sensor, as well as terrain and subsurface properties (e.g. roughness, humidity, porosity and thermal conductivity). The melting period in spring is also characteristic of specific measurement locations, when phase changes keep GST at 0 °C (zero curtain, cf. Outcalt *et al*., 1990). Zero curtain effects can occasionally be observed in autumn, if snow falls on still warm ground and basal melting occurs. Steep bedrock or windblown locations show little or no zero curtains and generally follow the variations of air temperature. All these characteristics imply that the similarity of different GST time series depends more on topography and snow characteristics than spatial proximity. Therefore, gap-filling approaches are particularly promising when applied to large data-sets, where similar characteristics for several time series become more likely.

Accuracy and Pre-Processing

GST are usually recorded with low-cost miniature temperature loggers (Hoelzle *et al*., 1999; Lewkowicz, 2008). Wherever zero curtains occur, offsets to 0 °C are minimised to calibrate the sensors. Potential inhomogeneities, both within or between time series, can relate to the sampling resolution (ranging between 30 min and 6 h); the precision of the sensors (0.23–0.27 °C for UTL-1 (75% of PERMOS GST data); $<$ 0.1 °C for UTL-3 (8%); 0.5 °C for iButton® DS1922L (6%); 0.1 °C for Geoprecision M-Logs (11%)); the placement of the loggers in the field; and changes in terrain morphology (e.g. on fast-moving rock glaciers). While the last points cannot be improved during post-processing, different measurement intervals can at least partly be homogenised. Here, we linearly interpolated the data on regular time intervals (e.g. 2 h) in a first step and aggregated them to daily means in a second step. Days with missing raw data or a sampling resolution of less than 4 h were excluded. However, any other systematic data preparation procedure could be used to achieve a consistent and comparable data basis. Time series of at least 5 years of measurements were used for validation (see the Validation with Artificial Gaps section). Working with daily means allows accurate calculation of MAGST or degree-day indices.

Gap Characteristics and Completeness of the PERMOS GST Data

GST time series obtained within PERMOS start between 1994 and 2011 and have gaps ranging from some days to more than a year. For the 278 loggers and the hydrological years 2000/2001–2013/2014 shown in Figure 1a, the 150 longest GST time series are interrupted by about 300 gaps with a mean duration of 145 days. But about 50 per cent of the gaps are shorter than 1 month and 30 per cent are no longer than 5 days (Figure 1c). Most gaps occurred in summer, probably because maintenance was sometimes not done in the field but later in the laboratory (Figure 1b). Because in the snow-free period GST are influenced by common, atmospheric signals, covariance between different time series is high and empirical-statistical algorithms (see the Gap-Filling Algorithm section) are expected to perform most reliably for filling data gaps.

GAP-FILLING ALGORITHM

Approaches

Dealing with missing data is a major problem in many research disciplines and the need for standardisation has been identified (Falge *et al*., 2001; Gudmundsson *et al*., 2012). The potential approaches range from process-based models to empirical-statistical transformations, notably with very different requirements about input data, time windows and intervals, the applicability to reproduce the mean conditions and/or to retain the short-term variability, their ease of implementation and performance (e.g. Moffat *et al*., 2007). However, no studies compare gap-filling methods explicitly for ground temperature data measured in permafrost and periglacial environments. Gisnås *et al*. (2014) used air temperature to fill short gaps of only some days during the snow-free period. Hasler *et al*. (2015) replaced missing data with mean GST values and mean uncertainties for

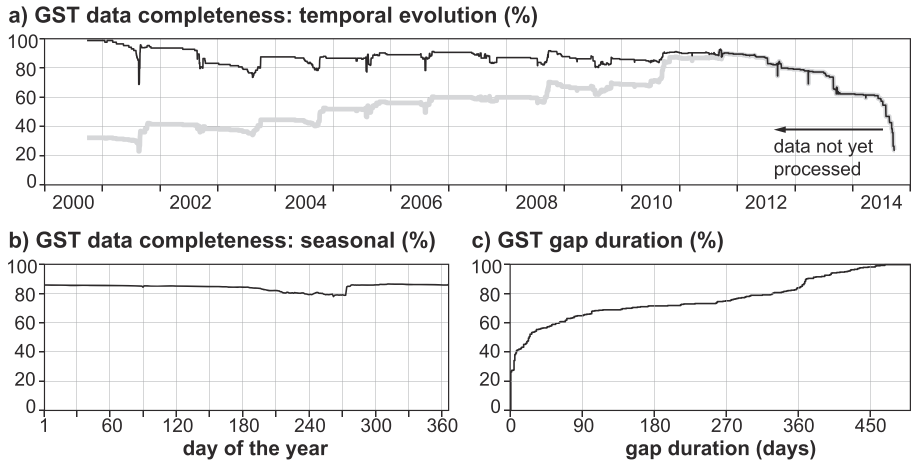

Figure 1 Data completeness and gap characteristics of the Swiss Permafrost Monitoring Network (PERMOS) ground surface temperature (GST) data for the hydrological years 2000/2001–2013/2014 (gaps after 2012 are mainly because the data were not yet processed): (a) data completeness relative to the total number of GST series already started (black line) and relative to the maximal number of GST loggers (in %, 278 in 2011/2012, grey line); (b) seasonal pattern of data completeness with most data gaps in summer; (c) empirical cumulative density function of the gap duration for gaps < 500 days.

comparing annual means. Magnin *et al*. (2015) filled gaps < 5 days by linear interpolation (LI) between the nearest available data points. These authors filled gaps up to 1.5 months with the average value of 30 days on each side of the gap, as done by Hasler *et al*. (2011). Within PERMOS, gaps in reference GST time series used for reporting were filled by linear regression on the basis of monthly means or monthly degree-day sums to calculate the indices MAGST, TDD and FDD (PERMOS, 2013). A more flexible use of the measured and synthesised data requires daily means with daily uncertainty estimates.

Process-based models (Ekici *et al*., 2015; Lehning *et al*., 2006; Westermann *et al*., 2013) are not applicable for large quantities of GST time series because of the stringent requirements on input variables (meteorological data and calibration parameters) and the high computational effort. Empirical-statistical methods (Gudmundsson *et al*., 2012; Themessl *et al*., 2011) are more appropriate. Statistical transformations are the most commonly used for bias correction and the spatial transfer of meteorological parameters (Gudmundsson *et al*., 2012; Rajczak *et al*., 2015; Tardivo and Berti, 2014).

The techniques that we used for this study are as follows. **Linear Interpolation** (LI) from GST values prior to and after the gap of the same logger is independent from other data and is mainly applied to fill short gaps of a few days. For gaps longer than synoptic weather patterns, we required a more sophisticated method based on similar regressor time series and using techniques for spatial transfer and bias correction, which adjust for the shape of the distribution (Gudmundsson *et al*., 2012; Rajczak *et al*., 2015; Themessl *et al*., 2011). The **quantile mapping** (QM, nonparametric statistical transformation) method was used to fill long GST gaps of up to several months. The QM technique is widely used to adjust simulation results (e.g. coarse-gridded RCM output) to local station observations, and gaps in meteorological data have been successfully filled for several PERMOS sites (Rajczak *et al*., 2015). QM uses empirical distribution of the data and therefore does not depend on a pre-defined distribution. QM assumes that the differences between the regressor and target logger are stationary, which is a simplification regarding interannual variations and long-term trends. However, we consider that QM could account for the high temporal and spatial variability of high-mountain GST realistically as long as the regressor and target logger are influenced by the same processes. Therefore, we selected the QM approach for the filling of long GST gaps.

First, the general work flow checks the gap characteristics, and selects the appropriate gap-filling approach and (only for QM) the most suitable regressor logger (Figure 2). Second, GST values as well as the resulting uncertainties are estimated for the target logger. The operational programming routine (written in R; R Core Team, 2015) and a small demo data-set are available as Supplementary Information.

LI of Values in Short Gaps

Calculation of Expected GST Values.

Based on GST average values before (GST_{prior}) and after the gap (GST_{after}), missing data (GST_{synt}) at date i in the gap of length n can be filled by LI (Equation 1 and Figure 3).

$$GST_{synt,i} = GST_{prior} + \left(\frac{GST_{after}\text{-}GST_{prior}}{n+1}\right)i, \quad \text{with } i = \{1, 2, \ldots, n\} \qquad (1)$$

While gaps of only 1 day are ideally interpolated just between the two nearest neighbours (Figure 3c), longer gaps require longer aggregation windows before and after the gap to calculate representative values for GST_{prior} and GST_{after} (Figure 3d). Therefore, the length m of the aggregation

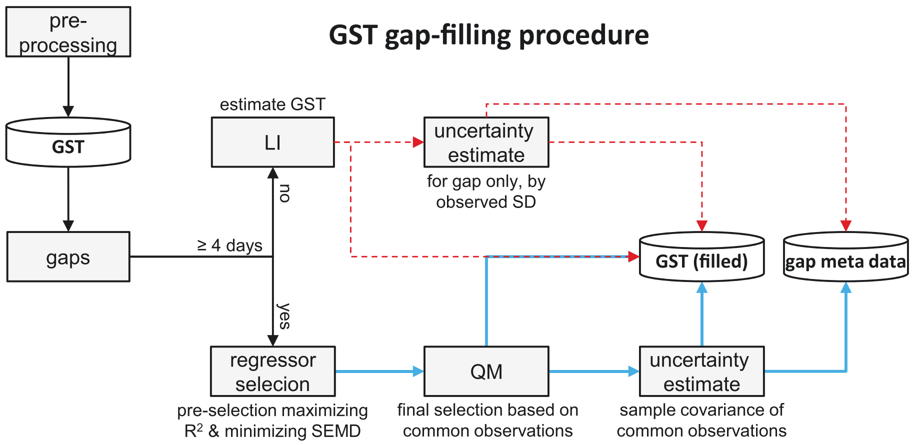

Figure 2 Flow chart describing the general approach for the ground surface temperature (GST) gap-filling. Depending on the gap duration, either linear interpolation (LI, for gaps of < 4 days) or quantile mapping (QM) were used to estimate missing GST data. Besides of the GST estimates (daily means), the resulting uncertainties can also be stored as meta-information (regressor logger, residuals, etc.). The figure illustrates how the gap-filling routine could be implemented into a monitoring database structure. SEMD = Standard error of the mean differences; SD = standard deviation.

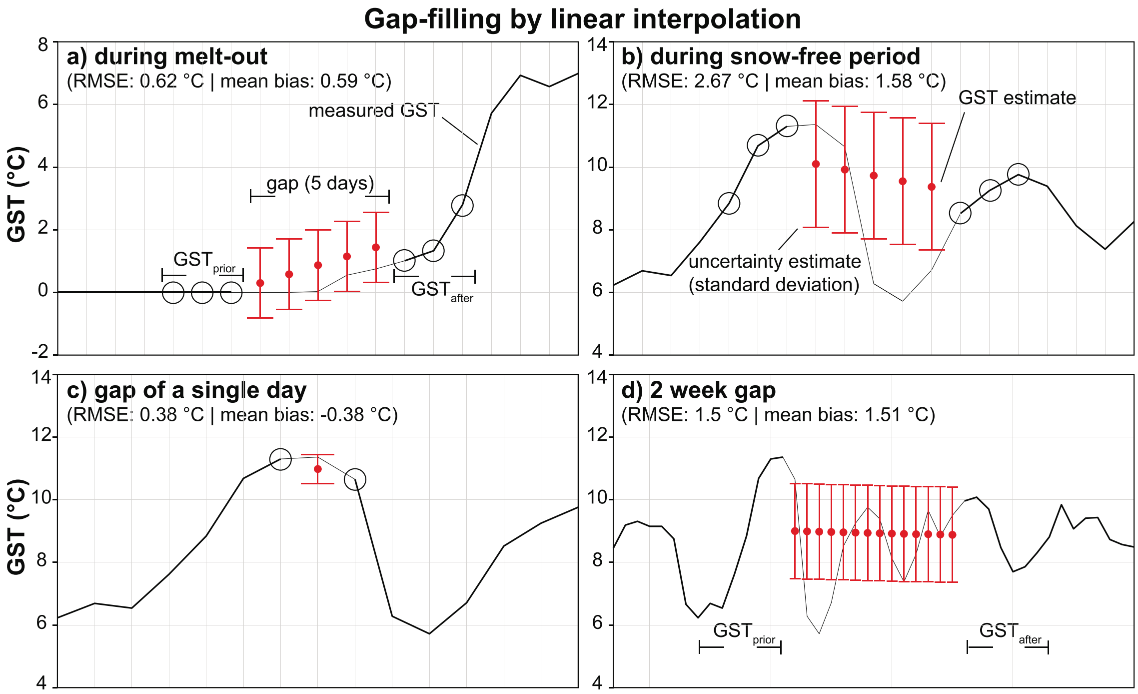

Figure 3 Examples for artificial gaps filled using linear interpolation: (a) end of the snowmelt period in early summer; (b) when the ground is snow-free; (c) a gap of a single day; (d) a gap of 2 weeks. The red dots represent the interpolated ground surface temperature (GST) values, and the mean uncertainty (error bars) was approximated based on the standard deviation of GST_{prior}/GST_{after} (see text for details). The root mean square error (RMSE) values and mean bias are calculated between the interpolated, synthetic GST in comparison to the real values (black lines).

windows is defined as:

$$m = \left\lceil \frac{n}{2} \right\rceil \in \mathbb{N} \tag{2}$$

Because GST data are highly autocorrelated during the snow-covered periods, when the ground is thermally insulated from air temperature variations, LI would technically be suitable also to fill gaps of up to several weeks during this time of the year. Such gaps were not observed, however, in the PERMOS data because they usually persist until the next maintenance takes place in summer (no remote control and no access to replace the logger during winter). The analysis of artificial gaps (see the Overview and Comparison of the LI and the QM Approach for Short Gaps section) showed that on average, the QM approach is more reliable for gaps longer than 3 days regarding the estimated GST values.

Calculation of Uncertainty.

For short gaps filled by LI (Equation 1), the mean uncertainty σ (°C, cf. Equation 3) can be estimated with the standard deviation (SD) calculated over all days in GST_{prior} and GST_{after}. This method accounts for differences in the short-term GST variability, resulting in greater uncertainties if this variability is larger (Figure 3a, b). These uncertainties are supposed to be stochastic and partly self-compensating, which is why the mean uncertainty of a GST aggregate can be divided by the square root of observations (see the Uncertainty Propagation for Indices section).

QM using the Best Regressor for Long Gaps

Selection of the Best Regressor Logger.

The selection of the best regressor logger (cf. Figure 2) is the crucial task for filling gaps using the QM approach, particularly in mountainous terrain with its extreme spatial heterogeneity. The robustness of the transformation increases further with the number of observations. The regressor logger is selected based on the following criteria:

1) Completeness of measurement data during the entire gap is required.
2) The season of the gap needs to be covered by the target and regressor logger in at least 5 years. Gaps shorter than 30 days are extended to a minimum length of 30 days to assess similarities with other loggers.
3) Calculation of the maximum of the Pearson product-moment and Spearman's rank correlation coefficients as well as the minimum of the standard error of the mean differences (SEMD, see below for further details) for the remaining observations (Figure 4b). The best 5 per cent of potential regressors for these three criteria are selected for criterion 4.
4) Application of the QM method to the selected regressor loggers and calculation of residuals between fitted values and observations. The logger with the minimal standard error of these residuals is selected as the final regressor logger (Figure 4a, b).

By maximising the Pearson correlation in criterion 3, the regressors of best covariance with the target logger can be selected. This is particularly important for QM because the statistical transformation corrects for errors in the mean and the percentiles, but not the daily correspondence between the regressor and the target logger (Rajczak *et al*., 2015). The Spearman's rank correlation is added to identify potential regressors, for which a strong but non-linear relationship to the data of the target logger is observed. Since correlation is only defined for finite SD > 0 (which is not necessarily given, for example, during periods of snowmelt), the SEMD is also used, as it is computationally efficient and accounting for the number of observations. For this study, the top 5 per cent of all statistical measures were pre-selected in criterion 3. This pre-selection of regressor loggers mainly reduces computation time for large data-sets because the statistical transformation (Figure 4c, d) is only evaluated for a subset of the potentially most suitable regressor loggers. To account for seasonally differing snow and meteorological conditions, only the days of the year where the gap occurs were used to fit the QM model. Within criterion 2, the minimal gap duration of short gaps is therefore set to 30 days to gain a representative sample of values for fitting the QM model.

Calculation of Expected GST Values.

The relatively homogeneous evolution of GST at different sites (PERMOS, 2013) allows the use of QM even for filling gaps of several months (Figure 4). Most challenging are the highly variable snow conditions influenced by precipitation, wind, radiation and avalanches, which may substantially differ over distances of a few metres. Therefore, each season of the gap period should be equally represented in the calibration data (Figure 4a). The robustness to fit means and extremes is the great advantage of the QM approach (Themessl *et al*., 2011). Gudmundsson *et al*. (2012) provide a more detailed description of the QM technique, and Gudmundsson (2014) the implementation and use in R.

Calculation of Uncertainty.

Hasler *et al*. (2015) estimated uncertainties for entire gaps in the form of a standard error of the mean. To gain accuracy regarding the seasonal pattern and to be more flexible for quantifying mean uncertainties of aggregates and indices (see the Uncertainty Propagation for Indices section), we suggest daily uncertainty estimates. Based on the residuals ε between the observations and the fitted values during the calibration period, uncertainty estimates (σ) can be approximated for each day of the year. To account for the length of the calibration period, the corrected sample SD is used with n_{year} being the number of years with common observations (Equation 3).

$$\sigma = \pm\sqrt{\frac{1}{n_{year}\text{-}1}\sum \varepsilon^2} \tag{3}$$

Assuming that the calibration period (common observations in the respective season, cf. Figure 4a) covers different meteorological and snow conditions, the interannual variability is to some extent included in this estimate. The uncertainty estimates are larger because of larger interannual variability (Figure 4c, e), especially at the beginning and the end of the melting period, but also during winter, when the ground is snow covered. In general, the residuals are more likely to be systematic (and not self-compensating) when snow insulates the ground. Therefore, we suggest using a simplified snow detection filter similar to those suggested by Schmid *et al*. (2012) and Staub and Delaloye (2016). We considered days as snow covered when the weekly SD of daily mean GST were ≤ 0.25 °C either for the target or the regressor logger. Based on this snow information, stochastic and systematic errors can be separated (see the Uncertainty Propagation for Indices section).

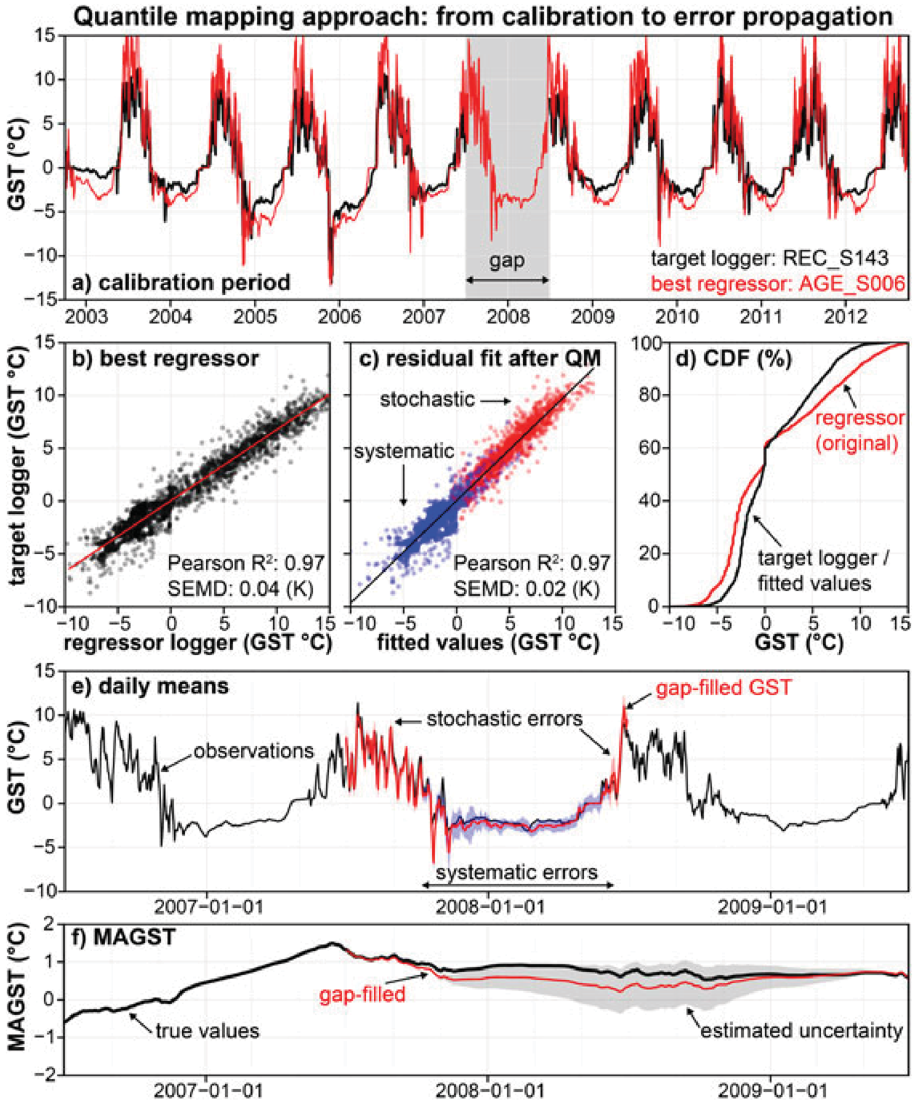

Figure 4 Illustration of the quantile mapping (QM) approach to fill data gaps in an artificial 365 day gap: (a, b) comparison of the target logger and the best regressor for the entire calibration period; (c) comparison of the fitted values after the QM transformation, based on the remaining residuals, stochastic and systematic uncertainties are quantified; (d) illustration of the cumulative density function (CDF) of the regressor and the target logger; the results of the gap-filling are illustrated as (e) daily means and (f) running mean annual ground surface temperatures (MAGST) in comparison with observations (black). GST = Ground surface temperatures; SEMD = standard error of the mean differences.

Uncertainty Propagation for Indices

According to the error propagation laws, the uncertainties resulting from gap-filling can be calculated also for aggregates and indices. Theoretically, stochastic errors (σ_{sto}) should be close to a normal distribution and self-compensating with increasing sample size n_{sto} (Figure 4e, f). Therefore, the mean error for a given period $i = \{1, 2, \ldots, n_{sto}\}$ can be divided by the square root of the total number of observations or rather gap length (Equation 4). Systematic errors (σ_{sys}) are additive and need to be averaged over all samples j in n_{sys} to calculate the mean uncertainty of the gap (σ_{gap}). The stochastic errors are restricted to the snow-free period and identified as described in the QM using Best Regressor for Long Gaps section.

$$\sigma_{gap} = \pm\sqrt{\left(\frac{\sigma_{sto}}{\sqrt{n_{sto}}}\right)^2 + \sigma_{sys}^2} \tag{4}$$

$$= \pm\sqrt{\frac{\left(\frac{1}{n_{sto}}\sum_{i=1}^{n_{sto}} \sigma_{sto,i}\right)^2}{n_{sto}} + \left(\frac{1}{n_{sys}}\sum_{j=1}^{n_{sys}} \sigma_{sys,j}\right)^2}$$

An example of uncertainty estimates for running annual means (MAGST) is illustrated in Figure 4 f. Measurement errors due to the limited accuracy of the sensors and the precision of the loggers (cf. the Accuracy and Pre-Processing section) could be included analogously.

VALIDATION WITH ARTIFICIAL GAPS

Overview and Comparison of the LI and QM Approach for Short Gaps

We validated the gap-filling routine with artificial gaps using complete GST observations from 18 different field sites and 173 different loggers over 10 hydrological years (2002/2003–2011/2012). Most of these time series were measured on gently inclined terrain and only some originated from steep bedrock. Hence, the difficulties related to the temporally and spatially variable snow conditions (see the Characteristics of GST Time Series in Permafrost section) made the application and validation of the gap filling challenging. The starting dates (day of the year), loggers and gap durations were randomly selected but forced to match the characteristics of the existing gaps in the PERMOS data (Figure 1). In a first simulation, the performance of the LI (see the LI of Values in Short Gaps section) and the QM approach (see the QM using Best Regressor for Long Gaps section) was compared for short gaps of up to 30 days between June and October (n ~ 24 000 gaps). This first simulation aimed to define the threshold for the maximal gap duration, until which LI could be chosen as an alternative for QM (cf. Figure 2). The validation showed that the QM method performed significantly better regarding the GST estimates than LI for gaps of 4 days or more. Based on this threshold, the second validation run applied the final gap-filling routine for gap durations between 1 and 500 days (n ~18 000 gaps, ~ 2.5 million days with missing values). For temperature data with less short-term variability (e.g. ground temperatures at depth or GST during winter), LI would be a reasonable choice for longer gaps as well.

Validation of the Gap-Filling Algorithm

Calculation of Expected GST Values.
Overall, the synthetic (simulated/gap-filled) GST data agreed well with the observations. The time series of the simulated daily mean GST was close to the observations, with an R^2 of 0.93 (root mean square error (RMSE) ~ 1.6 °C), although reconstructing the correct daily values was not the primary aim. More relevant was the performance for aggregates and indices. Because errors partly compensate, the RMSE of simulated and observed GST mean values for the entire gap was even smaller (Figure 5a). The degree-day sums FDD and TDD calculated on the basis of gap-filled and observed data showed a high covariance and reasonable RMSE values (Figure 5b). The temporal variation (SD) of TDD and FDD values was about ± 140 and ± 180, respectively. The spatial variations were even larger, with SD around ± 230 for TDD and ± 255 for FDD. The largest outliers occurred during the transition period between snowmelt and the snow-free period (cf. scatters along the 0 °C lines in Figure 5a). With larger gap sizes, the resulting bias on annual means tended to increase (Figure 5c). In 95 per cent of the situations, the maximal bias on MAGST (if running MAGST would be computed based on the gap-filled data) remained below 0.5 °C, even for gaps of 6–8 months. In the best 50 per cent of instances, the maximal MAGST bias did not exceed 0.25 °C. Interannual MAGST variations observed in the PERMOS data ranged between an amplitude of 2 and 3 °C (SD ~ 0.5 °C) (cf. PERMOS, 2013). More than 450 gaps could be filled in the PERMOS GST data (Figure 1), very likely without exceeding a maximal MAGST bias of ± 0.25 °C. This increases the number of complete hydrological years by *ca*. 70 per cent, with uncertainties in the order of the precision of the most popular miniature logger UTL-1. Since GST measured on steep bedrock are less influenced by snow and closely follow the variations of air temperature, these data-sets can be filled, theoretically, with smaller uncertainties as long as a suitable regressor time series is available. However, this was not observed due to the limited number of time series used that represent steep bedrock.

Calculation of Uncertainty.
At the daily scale, the differences between estimated uncertainties and the observed absolute bias (synthetic-observed GST values) were smallest during winter and largest at the end of the spring zero curtain because of the spatially and temporally variable snow disappearance (Figure 6a). The median difference remained close to 0 °C all year, which means that uncertainty was not systematically over- or underestimated. The remaining scatter visible in Figure 6a was also due to the limited significance of the uncertainty estimation at a daily scale (see the Uncertainty Propagation for Indices section). Regarding the aggregation level of entire gaps, the estimated uncertainties were slightly larger than the observed absolute bias, 0.045 °C on average. For ~ 60 per cent of all gaps (~70% of all reconstructed daily mean values), the observed absolute errors were smaller than the estimated uncertainty and the differences were generally small (RMSE = 0.21 °C) and normally distributed (Figure 6b, c). However, the correlation between the estimated uncertainties and the observed absolute bias was limited (R^2 = 0.37).

The uncertainty estimation for the shortest gaps of up to 3 days, for which the mean uncertainty was estimated based on the short-term variability before and after the gap according to the LI approach (see the LI of Values in Short Gaps section), is very accurate (R^2 = 0.86; RMSE = 0.27 °C). The high autocorrelation of the GST time series at short timescales probably accounts for this high covariance between estimated uncertainty and observed bias. But the first validation run, which compared the LI and the QM approach for

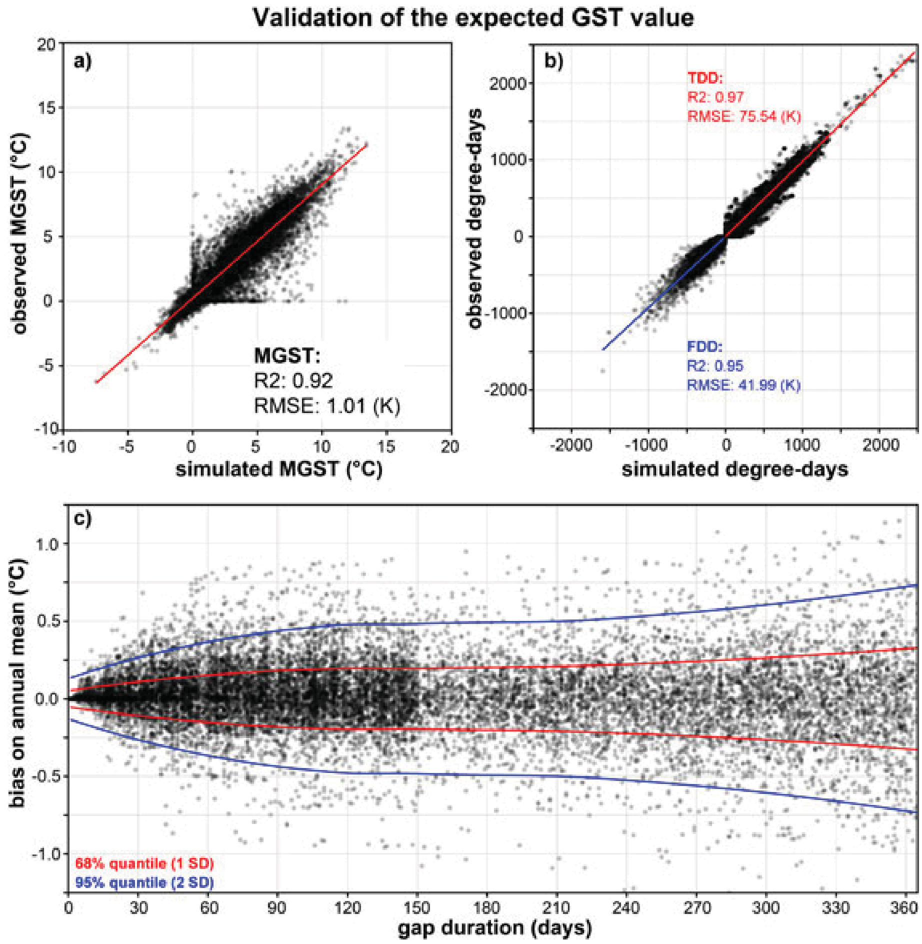

Figure 5 Comparison of simulated (synthetic/gap filled) and observed ground surface temperatures (GST) for ~ 18 000 gaps (~2.5 million daily mean values): (a) scatterplot based on mean ground surface temperatures (MGST) for entire gaps; (b) scatterplot comparing simulated and observed thawing (TDD) and freezing degree-days (FDD); (c) maximal bias resulting on mean annual ground surface temperature as a function of the gap duration: each point corresponds to one artificial gap, the red lines show the limits of one standard deviation (SD) (68% quantile) and the blue lines for two SD (95% quantile) of the absolute annual bias. RMSE = Root mean square error.

gap durations between 1 and 30 days, revealed that in ~ 60 per cent of the situations the uncertainty estimation of the LI method performed better than that of QM. However, to fill gaps in GST time series, we recommend the QM method for gap durations in the order of weeks because it estimates the GST values more accurately. But for data with low daily and seasonal variation (e.g. ground temperatures measured at depth or GST during the snow-covered period), the LI method also holds promise for uncertainty estimation.

To improve the uncertainty estimation for the QM approach, the interannual variability could be assessed in more detail using separated validation data (e.g. by means of cross-validation techniques). Each year of the common observation period could be validated separately by excluding it from calibration. However, this would only make sense for time series of at least 10 years. Alternatively, expected GST values could be calculated for numerous regressor loggers and the daily uncertainty values estimated from the resulting differences. The latter approach would probably be more adequate for the usually short GST time series and take into account the spatial variability.

The validation using artificial data gaps also gave insights about the similarities of the different GST time series. For almost 50 per cent of the gaps where QM was applied, the selected regressor logger was from the same study site (usually located within 50–500 m distance). For gaps during winter, this proportion was slightly higher. For short gaps during summer, only ~ 35 per cent of the regressor loggers originated from the same site, which means that in the majority of situations, the most similar GST time series were potentially several tens or even hundreds of kilometres away. Moreover, each target logger used on average seven different regressor loggers (4% of all available loggers). Among these preferred regressor loggers, the most frequently used filled 57 per cent of all gaps of the corresponding target logger. These observations demonstrate how a large set of distributed time series can be used to fill data gaps, especially in heterogeneous mountain terrain. They also illustrate the

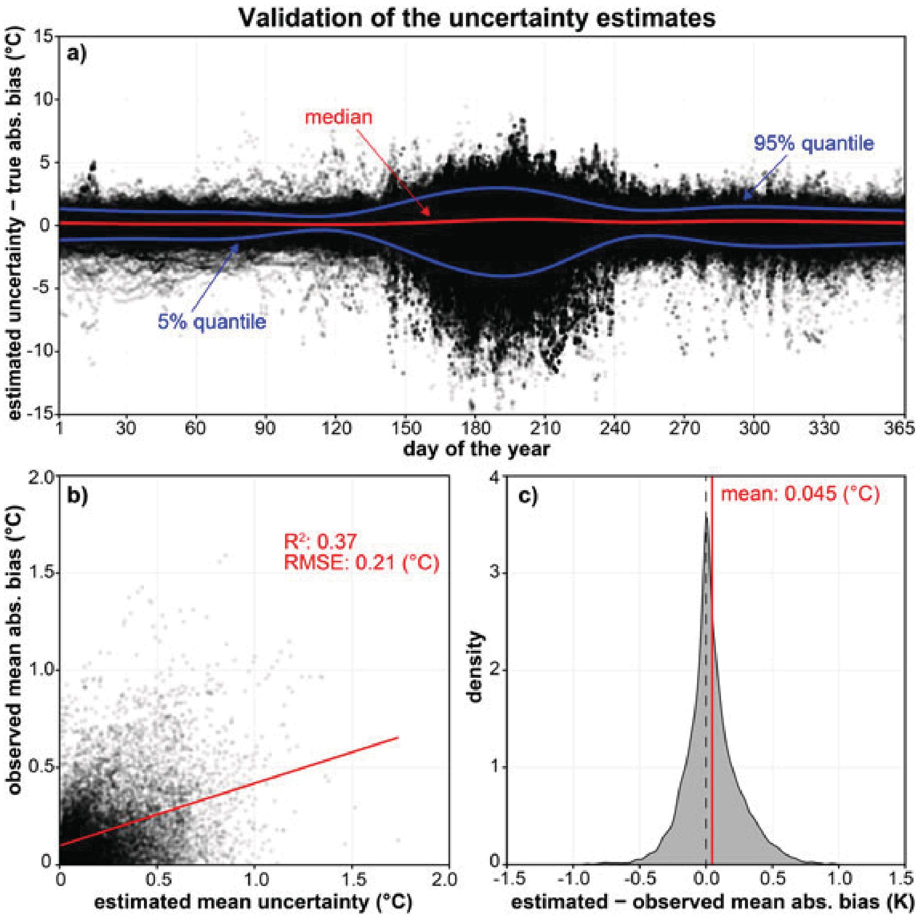

Figure 6 Comparison of estimated uncertainties and the true absolute (abs.) bias: (a) seasonal pattern based on daily mean values (the red line indicates the median, the blue lines illustrate the 5% and 95% quantiles); (b) scatterplot comparing the estimated mean uncertainties and the mean absolute bias for entire gaps; (c) density distribution of the same data as shown in (b). RMSE = Root mean square error.

relevance of snow and the point-specific topo-climatic conditions in high-mountain terrains for the evolution of GST at various timescales. In this regard, the diversity of the potential regressor time series is more important than their quantity, especially for filling gaps during winter.

It could be interesting to assess in more detail the minimal number of time series required to apply the gap-filling approach (e.g. for the entire Swiss Alps). Possibly, 30–50 continuous time series measured in very diverse ground and snow properties and characterising the precipitation effects of three to five main regions could be sufficient to fill the majority of data gaps in the PERMOS GST data-set. The more time series are available, the higher the chances are to find appropriate regressors.

CONCLUSIONS

We introduce an automatic procedure to fill gaps in GST data sets on the basis of daily mean values. The approach is optimised to consider spatial and temporal variations which result from the heterogeneous terrain and snow characteristics of high-mountain regions. A validation was done using artificial gaps and GST data from PERMOS. The main conclusions on the use and limitations of the presented approach are:

- Linear interpolation (LI) works well to fill gaps shorter than 1 week. Analysis with artificial gaps showed that LI performed as good as or better than the more complex quantile mapping (QM) approach for gaps up to 3 days′ duration. The uncertainty estimation of the LI method was slightly more robust than that of the QM approach for gap durations of up to 1 month.
- Since QM corrects well for mean biases, even gaps of several months′ duration may be filled with reasonable bias on MAGST in the order of the precision of the common UTL-1 data logger. Maximal MAGST bias resulting from 1 year gaps usually did not exceed ± 0.5 °C, which corresponds to the SD of interannual MAGST variations. A high number of potential regressors as well as long periods with common observations probably make the approximation of GST values and the quantification of uncertainties more reliable.

- Validation with artificial gaps showed that the resulting mean error was usually smaller than the estimated uncertainty. Although the uncertainty estimation and error propagation approach is conservative regarding the effects of snow, the results are imperfect, particularly during snow disappearance. The uncertainties of the GST estimates might be quantified more reliably based on an ensemble of several simulations using numerous regressor loggers.
- Selection of the regressor logger is crucial for accurately representing the specific ground thermal conditions with the QM approach. Interannual variations in the snow and meteorological conditions need to be captured. The QM technique also performed well for GST measured on steep bedrock and the uppermost thermistors of boreholes. To increase the quantity of potential regressors, air temperatures and ground temperatures measured reasonably close to the surface can be included, even from distant locations. In the Swiss Alps, several GST time series showed high covariance even between different regions of up to 200 km distance, especially during the snow-free period.
- The application of the gap-filling approach on the PERMOS GST data illustrated the high relevance of point-specific terrain, snow and micro-meteorological characteristics for the evolution of GST at timescales of days to several months.

The processing routine (R code) as well as a small sample GST data-set are available as Supplementary Information.

ACKNOWLEDGEMENTS

The Swiss National Science Foundation funded the Sinergia TEMPS project (no. CRSII2 136279). The data time series were provided by PERMOS and the TEMPS project. We thank all the individual researchers from the partner institutions, for their work in the field and the pre-processing, as well as the two projects, for the data preparation and access. Moreover, we would like to thank the Editor, Julian Murton, as well as two anonymous referees for their detailed and helpful comments and suggestions.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher's web site.

REFERENCES

Bonnaventure PP, Lewkowicz AG. 2008. Mountain permafrost probability mapping using the BTS method in two climatically dissimilar locations, northwest Canada. *Canadian Journal of Earth Sciences* **45**: 443–455. DOI:10.1139/E08-013

Delaloye R. 2004. Contribution à l'étude du pergélisol de montagne en zone marginale. PhD thesis, Department of Geosciences, University of Fribourg, GeoFocus No. 10.

Ekici A, Chadburn S, Chaudhary N, Hajdu LH, Marmy A, Peng S, Boike J, Burke E, Friend AD, Hauck C, Krinner G, Langer M, Miller PA, Beer C. 2015. Site-level model intercomparison of high latitude and high altitude soil thermal dynamics in tundra and barren landscapes. *The Cryosphere* **9**: 1343–1361. DOI:10.5194/tc-9-1343-2015

Etzelmüller B, Farbrot H, Guðmundsson Á, Humlum O, Tveito OE, Björnsson H. 2007. The regional distribution of mountain permafrost in Iceland. *Permafrost and Periglacial Processes* **18**: 185–199. DOI: 10.1002/ppp.583

Falge E, Baldocchi D, Olson R, Anthoni P, Aubinet M, Bernhofer C, Burba G, Ceulemans R, Clement R, Dolman H, Granier A, Gross P, Grünwald T, Hollinger D, Jensen NO, Katul G, Keronen P, Kowalski A, Lai CT, Law BE, Meyers T, Moncrieff J, Moors E, Munger JW, Pilegaard K, Rannik Ü, Rebmann C, Suyker A, Tenhunen J, Tu K, Verma S, Vesala T, Wilson K, Wofsy S. 2001. Gap filling strategies for defensible annual sums of net ecosystem exchange. *Agricultural and Forest Meteorology* **107**: 43–69. DOI:10.1016/S0168-1923(00)00225-2

Gisnås K, Westermann S, Schuler TV, Litherland T, Isaksen K, Boike J, Etzelmüller B. 2014. A statistical approach to represent small-scale variability of permafrost temperatures due to snow cover. *The Cryosphere* **8**: 2063–2074. DOI:10.5194/tc-8-2063-2014

Gubler S, Fiddes J, Keller M, Gruber S. 2011. Scale-dependent measurement and analysis of ground surface temperature variability in alpine terrain. *The Cryosphere* **5**: 431–443. DOI:10.5194/tc-5-431-2011

Gudmundsson L. 2014. qmap. Statistical transformations for post-processing climate model output. *R package qmap version* **1**: 0–3.

Gudmundsson L, Bremnes JB, Haugen JE, Engen-Skaugen T. 2012. Technical Note: Downscaling RCM precipitation to the station scale using statistical transformations - a comparison of methods. *Hydrology and Earth System Sciences* **16**: 3383–3390. DOI:10.5194/hess-16-3383-2012

Guglielmin M. 2006. Ground surface temperature (GST), active layer and permafrost monitoring in continental Antarctica. *Permafrost and Periglacial Processes* **17**: 133–143. DOI:10.1002/ppp.553

Guglielmin M, Worland MR, Cannone N. 2012. Spatial and temporal variability of ground surface temperature and active layer thickness at the margin of maritime Antarctica, Signy Island. *Geomorphology* **155-156**: 20–33. DOI:10.1016/j.geomorph.2011.12.016

Hasler A, Gruber S, Haeberli W. 2011. Temperature variability and offset in steep alpine rock and ice faces. *The Cryosphere* **5**: 977–988.

Hasler A, Geertsema M, Foord V, Gruber S, Noetzli J. 2015. The influence of surface characteristics, topography and continentality on mountain permafrost in British Columbia. *The Cryosphere* **9**: 1025–1038. DOI:10.5194/tc-9-1025-2015

Hoelzle M, Wegmann M, Krummenacher B. 1999. Miniature temperature dataloggers for mapping and monitoring of permafrost in high mountain areas: first experience from the Swiss Alps. *Permafrost and Periglacial Processes* **10**: 113–124. DOI: 10.1002/(SICI)1099-1530(199904/06)10:2<113::AID-PPP317>3.0.CO;2-A

Isaksen K, Ødegård RS, Etzelmüller B, Hilbich C, Hauck C, Farbrot H, Eiken T, Hygen HO, Hipp TF. 2011. Degrading Mountain Permafrost in Southern Norway: Spatial and Temporal Variability of Mean Ground Temperatures, 1999-2009. *Permafrost and Periglacial Processes* **22**: 361–377. DOI:10.1002/ppp.728

Lehning M, Völksch I, Gustafsson D, Nguyen TA, Stähli M, Zappa M. 2006. ALPINE3D:

a detailed model of mountain surface processes and its application to snow hydrology. *Hydrological Processes* **20**: 2111–2128. DOI:10.1002/hyp.6204

Lewkowicz AG. 2008. Evaluation of miniature temperature-loggers to monitor snowpack evolution at mountain permafrost sites, northwestern Canada. *Permafrost and Periglacial Processes* **19**: 323–331. DOI:10.1002/ppp.625

Luetschg M, Lehning M, Haeberli W. 2008. A sensitivity study of factors influencing warm/thin permafrost in the Swiss Alps. *Journal of Glaciology* **54**: 696–704. DOI: 10.3189/002214308786570881

Magnin F, Deline P, Ravanel L, Noetzli J, Pogliotti P. 2015. Thermal characteristics of permafrost in the steep alpine rock walls of the Aiguille du Midi (Mont Blanc Massif, 3842 m a.s.l). *The Cryosphere* **9**: 109–121. DOI:10.5194/tc-9-109-2015

Moffat AM, Papale D, Reichstein M, Hollinger DY, Richardson AD, Barr AG, Beckstein C, Braswell BH, Churkina G, Desai AR, Falge E, Gove JH, Heimann M, Hui D, Jarvis AJ, Kattge J, Noormets A, Stauch VJ. 2007. Comprehensive comparison of gap-filling techniques for eddy covariance net carbon fluxes. *Agricultural and Forest Meteorology* **147**: 209–232. DOI:10.1016/j.agrformet.2007.08.011

Outcalt SI, Nelson FE, Hinkel KM. 1990. The zero-curtain effect: Heat and mass transfer across an isothermal region in freezing soil. *Water Resources Research* **26**: 1509–1516. DOI:10.1029/WR026i007p01509

Park H, Sherstiukov AB, Fedorov AN, Polyakov IV, Walsh JE. 2014. An observation-based assessment of the influences of air temperature and snow depth on soil temperature in Russia. *Environmental Research Letters* 97 pp. DOI:10.1088/1748-9326/9/6/064026

PERMOS. 2013. Permafrost in Switzerland 2008/2009 and 2009/2010, Noetzli J (ed), Glaciological Report Permafrost No. 10/11 of the Cryospheric Commission of the Swiss Academy of Sciences, Zurich, Switzerland.

Core Team R. 2015. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing: Vienna, Austria.

Rajczak J, Kotlarski S, Salzmann N, Schär C. 2015. Robust climate scenarios for sites with sparse observations: a two-step bias correction approach. *International Journal of Climatology* **36**: 1226–1243. DOI:10.1002/joc.4417

Schmid MO, Gubler S, Fiddes J, Gruber S. 2012. Inferring snowpack ripening and melt-out from distributed measurements of near-surface ground temperatures. *The Cryosphere* **6**: 1127–1139. DOI:10.5194/tc-6-1127-2012

Staub B, Delaloye R. 2016. Snow-indices for mountain permafrost applications derived from near-surface ground temperature data. *Permafrost and Periglacial Processes* DOI:10.1002/ppp.1890

Tardivo G, Berti A. 2014. The selection of predictors in a regression-based method for gap filling in daily temperature datasets. *International Journal of Climatology* **34**: 1311–1317. DOI:10.1002/joc.3766

Themessl MJ, Gobiet A, Leuprecht A. 2011. Empirical-statistical downscaling and error correction of daily precipitation from regional climate models. *International Journal of Climatology* **31**: 1530–1544. DOI: 10.1002/joc.2168

Westermann S, Schuler TV, Gisnås K, Etzelmüller B. 2013. Transient thermal modeling of permafrost conditions in Southern Norway. *The Cryosphere* **7**: 719–739. DOI:10.5194/tc-7-719-2013

Wu T, Zhao L, Li R, Wang Q, Xie C, Pang Q. 2013. Recent ground surface warming and its effects on permafrost on the central Qinghai-Tibet Plateau. *International Journal of Climatology* **33**: 920–930. DOI:10.1002/joc.3479

Zhang T. 2005. Influence of the seasonal snow cover on the ground thermal regime: An overview. *Reviews of Geophysics* **43**RG4002: DOI:10.1029/2004RG000157