



The GRENE-TEA model intercomparison project (GTMIP) Stage 1 forcing data set

T. Sueyoshi^{1,2}, K. Saito², S. Miyazaki^{1,2,a}, J. Mori^{1,2}, T. Ise³, H. Arakida⁴, R. Suzuki², A. Sato⁵, Y. Iijima²,
H. Yabuki^{1,2}, H. Ikawa⁶, T. Ohta⁷, A. Kotani⁷, T. Hajima², H. Sato², T. Yamazaki⁸, and A. Sugimoto⁹

¹National Institute of Polar Research, Tachikawa, Japan

²Japan Agency for Marine-Earth Science and Technology, Yokohama, Japan

³Kyoto University, Field Science Education and Research Center, Kyoto, Japan

⁴RIKEN Advanced Institute for Computational Science, Kobe, Japan

⁵National Research Institute for Earth Science and Disaster Prevention, Snow and Ice Research Center,
Nagaoka, Japan

⁶National Institute for Agro-Environmental Sciences, Tsukuba, Japan

⁷Nagoya University, Graduate School of Bioagricultural Sciences, Nagoya, Japan

⁸Tohoku University, Graduate School of Science, Sendai, Japan

⁹Hokkaido University, Faculty of Environmental Earth Science, Sapporo, Japan

^anow at: Sonic Corporation, Tachikawa, Japan

Correspondence to: K. Saito (ksaito@jamstec.go.jp)

Received: 14 May 2015 – Published in Earth Syst. Sci. Data Discuss.: 24 August 2015

Revised: 10 December 2015 – Accepted: 11 December 2015 – Published: 15 January 2016

Abstract. Here, the authors describe the construction of a forcing data set for land surface models (including both physical and biogeochemical models; LSMs) with eight meteorological variables for the 35-year period from 1979 to 2013. The data set is intended for use in a model intercomparison study, called GTMIP, which is a part of the Japanese-funded Arctic Climate Change Research Project. In order to prepare a set of site-fitted forcing data for LSMs with realistic yet continuous entries (i.e. without missing data), four observational sites across the pan-Arctic region (Fairbanks, Tiksi, Yakutsk, and Kevo) were selected to construct a blended data set using both global reanalysis and observational data. Marked improvements were found in the diurnal cycles of surface air temperature and humidity, wind speed, and precipitation. The data sets and participation in GTMIP are open to the scientific community (doi:10.17592/001.2015093001).

1 Introduction

GTMIP (GRENE-TEA model intercomparison project, <https://ads.nipr.ac.jp/gtmip/gtmip.html>; Miyazaki et al., 2015) is a model intercomparison study in the GRENE-TEA project (Green Network of Excellence – Terrestrial Ecosystem of the Arctic), a subproject of the 2011–2015 Japanese-funded Arctic research project GRENE Arctic. GRENE-TEA aims to understand ongoing changes to the circum-Arctic terrestrial ecosystem and their effects on climate. In order to address this goal, GTMIP attempts to assess inter-model and inter-site (climatic/vegetative) uncertainties,

as well as variations in land surface models (LSMs) including both physical and ecosystem processes. The project consists of two stages: Stage 1 includes one-dimensional (1-D) model simulations for selected sites, while Stage 2 is focused on distributed global/regional model simulation runs for selected climate scenarios. This paper describes the method to prepare the forcing data sets used in GTMIP Stage 1 and their comparison with observational data.

To evaluate the validity and variations of the performance of a model, and to conduct comparisons between the model's outputs and observational data at specific sites, it is essential to drive the model with such data that reflect realis-

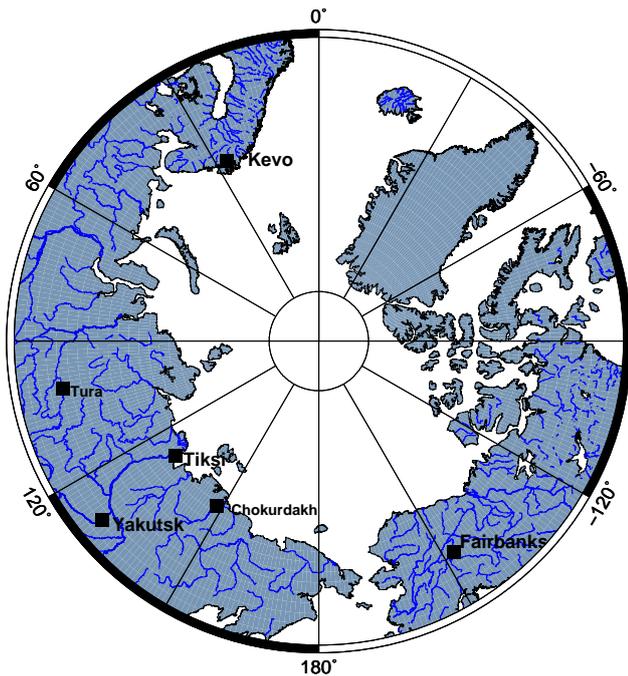


Figure 1. Location of the GTMIP observation sites (after Miyazaki et al., 2015).

tic conditions at those sites. It is very difficult, however, to create a complete data set for driving a model solely from observations, which are prone to data gaps due to instrument problems, human mistakes, or other reasons (Nakai et al., 2013; Kodama et al., 2007; Watanabe et al., 2000; Iwahana et al., 2014; Ohta et al., 2001, 2008, 2014; Kotani et al., 2013; Lopez et al., 2007; Nakai et al., 2008; Sato et al., 2001). Despite efforts to maintain and improve our observational network, data gaps are unavoidable, as the availability of instrumentation is often limited by logistical or funding issues. Many of the Arctic sites are located in remote areas with limited access and power supply (see Fig. 1 for GRENE-TEA observation sites; cf. site distributions of other national/international Arctic research projects, such as PAGE21 (<http://portal.inter-map.com/#mapID=49&groupID=308>; accessed on 20 April 2015) or map of stations used for the Climate Research Unit (CRU) data set (<http://climate.unur.com/cru-station-records/cru-stations-map.html>; accessed on 20 April 2015).

On the other hand, reanalysis data are often used to drive LSMs, as they are always continuous without missing values and physically consistent with regard to all necessary elements. Although there are several good reanalysis data sets available now (e.g. NCEP, ERA, JRA, MERRA), these products, which are essentially model outputs at large horizontal scales (typically on the order of 100 km), are no match for a direct comparison with observational data. Previous efforts of the modelling and empirical research community have produced several data sets available as inputs into LSMs

such as the Global Meteorological Forcing Dataset for land surface modelling (Sheffield et al., 2006; covering for 1948–2000) and WATCH (Water and Global Change; Weedon et al., 2010; covering for 1958–2001), WFDEI (WATCH Forcing Data ERA-Interim; Weedon et al., 2014; covering 1979–2012), or FLUXNET observational data (Vuichard and Papale, 2015; for FLUXNET sites) for debiasing of reanalysis data (ERA-Interim). After consideration of the project's requisites for the study period (i.e. data should cover the period from 1979 to 2013) and locations (i.e. the data set should be ready to cover any GRENE Arctic sites), GTMIP took the approach of preparing its own harmonized data set from reanalysis data. Another motivation for not using the existing data sets was that the processes of creating site-fitted data constituted an essential part of the collaboration between modelling and field scientists, for both of whom it was important to be part of a transparent data development process. Therefore, we decided to use reanalysis data to create temporally continuous data (basis for level 0 data, see Sect. 3) while utilizing observational data in customizing the level 0 data to fit to each site. The final data set (level 1 data) has a high consistency relative to the observational data without gaps.

In this paper, after summarizing information about the study sites and observational period in Sect. 2, we describe the specification of the data in Sect. 3 and the details of the procedure used to produce the two data sets in Sect. 4. The procedure is explained for each variable in the data set, and differences between the sites, when present, are described. Further implications of this work are described in Sect. 5, and access to the data is documented in Sect. 6.

2 Study sites and observational periods

Stage 1 of GTMIP produced a harmonized forcing data set for six sites in Arctic and sub-Arctic regions, which have been, at least partially, studied as part of the GRENE-TEA project (Fig. 1): Fairbanks in Alaska (Nakai et al., 2013), Tiksi (Kodama et al., 2007; Yabuki et al., 2011; Watanabe et al., 2000) and Chokurdakh (Iwahana et al., 2014) in north-east Siberia, Yakutsk in central east Siberia (Ohta et al., 2001, 2008, 2014; Kotani et al., 2013; Lopez et al., 2007; Yabuki et al., 2011), Tura in central Siberia (Nakai et al., 2008), and Kevo in northern Europe (Sato et al., 2001; precipitation and air pressure in Kevo are provided by Finnish Meteorological Institute). The longitudes and latitudes of the sites are summarized in Table 1, as well as the observational periods for data collection.

The sites lie in different geographical zones with different ecoclimatic characteristics. Fairbanks is located in the discontinuous permafrost zone with open black spruce forest; Tiksi and Chokurdakh are in the continuous permafrost zone, the former with tundra vegetation and the latter in the transition between tundra and taiga; Yakutsk and Tura are in the continuous permafrost zone with larch forest; and Kevo

Table 1. Names and coordinates of the observation sites.

Site name	Location	Observation period	Observation height T ; q ; wind
Chokurdakh	148.264° E, 70.563° N	NA*	NA*
Fairbanks	147.488° W, 65.123° N	1 Jan 2011–31 Dec 2012	1.5 m; 1.5 m; 16 m ^a
Kevo	27.010° E, 69.757° N	1 Jan 1995–31 Dec 2001, 1 Jan 1995–31 Dec 2012 (only for precipitation)	2.5 m; 2.5 m; 11.5 m
Tiksi	128.774° E, 71.589° N	1 Jan 1997–31 Dec 2012	2 m; 2 m; 10 m
Tura	100.464° E, 64.209° N	NA*	NA*
Yakutsk	129.618° E, 62.255° N	1 Jan 2005–31 Dec 2011	1.8 m; 1.8 m; 32 m ^b

* Observations at Chokurdakh and Tura were not used since level 1 data were not created for these sites. ^a Average canopy height is 2.91 m (Nakai et al., 2013). ^b Average canopy height is 18 m (Ohta et al., 2001).

is in the non-permafrost (seasonal frozen ground) zone with mixed forest.

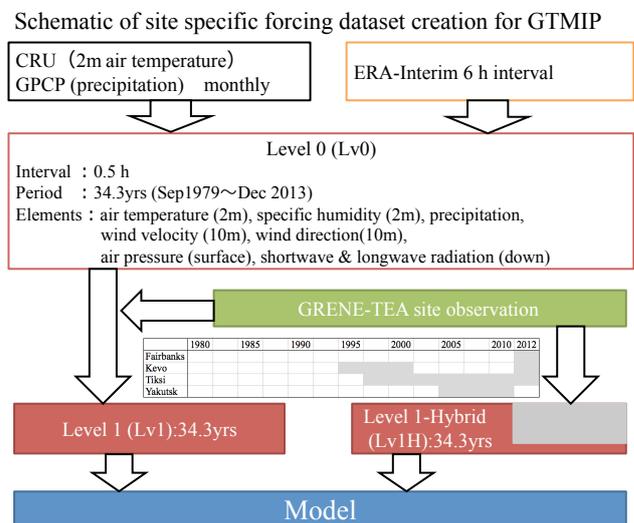
These sites were selected from “super sites” among the GRENE-TEA observation sites, where automatic weather stations were installed that measure a number of meteorological variables and enough site information about soil and vegetation conditions were available. Geographical coverage was also considered, to ensure that the sites were distributed over the entire Arctic and that they covered a combination of tundra and forest.

Although we are planning to expand this network of “harmonized data set sites”, we believe that these six initial sites, already covering typical geographical conditions in the Arctic, are a good starting point for a model intercomparison aiming to improve our understanding of physical and ecosystem land surface processes in the Arctic and sub-Arctic regions.

3 Data set specifications

The data created in this study cover a 35-year period, from 1 September 1979 to 31 December 2013, local time, for all sites. Temporal resolution was set to 30 min, considering the requirements of a majority of the physical land surface models. The start date was set to 1 September, when there should be no snow cover, in order to minimize any potential bias for snow depth at the initial condition.

The GTMIP data set consists of three sets of data, each of which has eight meteorological elements so that a contemporary land surface model can be driven with the data set. The first set is called level 0 (Lv0), consisting of reanalysis data with minimum correction, corresponding to the existing observation-based global data set. The second set is called level 1 (Lv1), with additional corrections applied to Lv0 data, based on the GRENE-TEA site observations. The third set is a detrended data set (DT), which essentially consists of Lv1 data with long-term temporal trends removed. Both Lv0 and Lv1 are forcing data sets. The primary data set for the GTMIP activity is Lv1, while Lv0 was prepared to enable an evaluation of the effect of the observation-based

**Figure 2.** Schematic diagram for data creation during GTMIP Stage 1.

corrections. The DT is intended for use with ecosystem models with long spin-up times. We also provided satellite-based photosynthetically active radiation (fPAR) and leaf area index (LAI) data for models that need these data as prescribed conditions. A schematic diagram for the creation of the Lv0 and Lv1 data is shown in Fig. 2.

The included eight variables and their units are as follows: surface air temperature (K) and specific humidity (kg kg^{-1}) at a reference height (set to 2 m in most models), precipitation ($\text{kg m}^{-2} \text{s}^{-1}$), wind speed (m s^{-1}) and wind direction (at 10 m; clockwise in degrees from north), pressure at the surface (hPa), and downward short-wave and long-wave radiation (W m^{-2}).

4 Methods of forcing data production

This section provides documentation for the (1) selection of reanalysis data, (2) construction procedure for Lv0 data, (3) construction procedure for Lv1 data, and (4) construction

Table 2. Outline of the procedure followed to produce forcing variables for (a) Lv0 and (b) Lv1.

(a) Site	Air pressure	Downward long-wave radiation	Air temperature (T)	Surface specific humidity (q)	Downward short-wave radiation (SW)	Total precipitation (P)	Wind speed
Any site	ERA output (constant for 6h)	ERA output (constant for 6h)	ERA- T + (CRU monthly T - ERA monthly T) linearly interpolated	ERA-relative humidity (i.e. ratio of saturation q at ERA-dew point T to that at ERA- T) multiplied by saturation q at Lv0 T	Theoretical SW (for lat, long, day, and time) multiplied by cloudiness (i.e. ERA daily total SW/theoretical daily total SW)	ERA 6-hourly P assigned at first time slot in every 6h (corrected GPCP monthly P /ERA monthly P)	Calculated from u and v of ERA wind speed (constant for 6h)
(b) Site (altitude of the site, in ERA) (m)	Air pressure	Downward long-wave radiation	Air temperature (T)	Specific humidity (q)	Downward short-wave radiation (SW)	Total precipitation (P)	Wind speed
Fairbanks (210, 390)	Lv0 data corrected for the altitude diff. Lv1 = Lv0 + 24 (2011–2012)	Lv0 data corrected for the emission diff. due to altitude diff. Lv1 = 1.01 · Lv0 + 15 (2011–2012)	Lv0 data corrected for max. and min. based on CRU dtr (daily temperature range)	ERA-relative humidity multiplied by saturation q at Lv1 T	Lv0 data	Lv0 data	Lv0 data
Kevo (100, 279)	Same as above Lv1 = Lv0 + 21 (1995–2001)	Same as above Lv1 = 1.00 · Lv0 + 11 (1995–2001)	Lv0 data corrected for max. and min. based on ERA max. and min. T	Same as above	Lv0 data	Statistical correction on Lv0' (Lv0 data w/o GPCP correction) Lv1 = 0.76 · Lv0 (1995–2012)	Statistical correction Lv1 = 0.6 · Lv0 (for > 1 m s ⁻¹) (1995–2001)
Tiksi (40, 184)	Same as above Lv1 = Lv0 + 19 (1997–2012)	Same as above Lv1 = 0.85 · Lv0 + 50 (1997–2012)	Statistical correction for season on Lv0' (ERA T corrected for daily max. and min. T based on ERA max. and min. T), Lv1 = 0.90 · Lv0' + 0.1, Nov–Mar Lv1 = Lv0', May to Oct	Same as above	Lv0 data	Lv0 data corrected for annual total P (ERA annual P /GPCP annual P)	Statistical correction Lv1 = 1.4 · Lv0 (for > 1 m s ⁻¹) (1997–2012)
Yakusk (220, 176)	Same as above Lv1 = Lv0 - 5 (2005–2011)	Same as above Lv1 = 0.98 · Lv0 + 17 (2005–2011)	Lv0 data corrected for max. and min. based on ERA max. and min. T	Same as above	Lv0 data	Lv0 data w/o GPCP correction	Lv0 data

procedure for additional data (detrended data, and LAI data). The outline of the procedures for Lv0 and Lv1 is given in Table 2. Procedures are explained for each variable in Sects. 4.2 and 4.3. When data were handled differently for different sites, it was also noted.

4.1 Selection of reanalysis data

As the first step of the procedure, we selected the reanalysis data set to be used as the base data. A detailed performance comparison was carried out for the following four data sets: Japanese 55-year Reanalysis (JRA-55), the European Centre for Medium-range Weather Forecasts ERA-Interim (Dee et al., 2011), National Centers for Environmental Prediction (NCEP)/ National Center for Atmospheric Research (NCAR), and NCEP-Department of Energy (DOE). CRU (East Anglia University) data for air temperature (Harris et al., 2014) and the Global Precipitation Climatology Project (GPCP) data set for precipitation (Adler et al., 2003), based on weather station observations and/or satellite data, were used as a reference. Special emphasis was placed on the reproducibility of air temperature and precipitation, as these two variables are known to have a large impact on the performance of LSMs. As a result of the comparison of monthly values, ERA-Interim was chosen as the base data set for GTMIP, reproducing these two variables reasonably with minimum root mean square errors (RMSEs) with respect to the CRU (air temperature) and GPCP (precipitation) data sets (Table 3).

Among the ERA-Interim variables, the following parameters were used in constructing the GTMIP forcing data: 10 m zonal (U) wind component, 10 m meridional (V) wind component, 2 m dew point temperature, 2 m temperature (t_{2m}), surface pressure, surface solar radiation downward, surface thermal radiation downward, and total precipitation. The temporal resolution of the used variables is 6 h, and stamped at 00:00, 00:06, 00:12, and 00:18 UTC (coordinated universal time).

4.2 Construction of level 0 data

4.2.1 General

Construction of Lv0 data (Saito et al., 2014a) from reanalysis data contains the following steps: (1) choose grid and time period (in this case September 1979–December 2013) from the reanalysis data, (2) apply a common correction for all sites for temperature and precipitation, and (3) determine 30 min values using the 6-hourly data. Providing a template data set for modellers to prepare their own read-in modules in advance is the primary purpose of this data set but, as mentioned before, it also serves to evaluate the effects of the site-fitting correction by comparing the performance of LSMs using these two data sets. Minimum corrections of temperature and precipitation using CRU and GPCP are applied so that

Table 3. Root mean square errors for the correlation between monthly reanalysis data and reference data (CRU for air temperature and GPCP for precipitation).

(a) Air temperature at 2 m (°C)				
Site	ERA-Interim	JRA55	NCEP/NCAR	NCEP-DOE
Fairbanks	1.59	3.08	3.44	3.39
Kevo	2.15	2.47	2.86	3.19
Tiksi	6.51	7.87	7.75	8.56
Yakutsk	0.80	2.44	3.20	3.81
(b) Precipitation (mm month ⁻¹)				
Site	ERA-Interim	JRA55	NCEP/NCAR	NCEP-DOE
Fairbanks	21.14	28.63	71.33	57.14
Kevo	15.60	17.15	25.50	24.15
Tiksi	18.11	17.92	29.43	24.47
Yakutsk	15.11	13.45	20.42	22.89

problems with using unrealistic values can be avoided. Differences between sites were not considered for this correction.

For the selection of the data grid, the “nearest neighbour” principle was used without applying any spatial interpolation. From the preliminary analysis, bi-linear interpolation was not always a better solution in the arctic or high-latitude sites. Table 4 shows the coordinates of grid centres of the ERA-Interim, CRU, and GPCP data sets for four sites.

Temporal interpolation is a requisite as the temporal resolution of the target data set is 30 min, while that of the ERA-Interim data is 6 h. A single strategy (e.g. linear interpolation or equal distribution) does not always work, depending on the characteristics of variables and sites. For the variables with intrinsic diurnal cycles (e.g. air temperature and downward solar radiation), however, an effort was made to reconstruct the diurnal cycle as much as possible. For other variables, we took the minimalist approach of adding no additional assumptions (i.e. temporal interpolation) before fixing the data sets, unless this caused unexpected disruptions of our examination simulations by LSMs. Before we finalized the data set to make a call for MIP, a test run was performed using a LSM MIROC-ESM, which did not produce harmful results. The process of temporal adjustment is described for each variable in the next sections.

4.2.2 2 m air temperature

For Lv0 air temperature (T_{Lv0}), monthly mean values derived from 6-hourly ERA-Interim air temperature (K) were adjusted using the CRU TS3.22 data set of the month, so that the monthly mean of Lv0 data coincides with the CRU temperature data, T_{CRU}^{month} (°C):

$$T_{Lv0}^{6h} = T_{ERA}^{6h} + \left(T_{CRU}^{month} + 273.15 \right) - \left(T_{ERA}^{6h} \right)_{\text{month}}. \quad (1)$$

Table 4. Names and coordinates of the grid centres of ERA-Interim, CRU, and GPCP data for the corresponding sites shown in Table 1.

Site name	ERA-Interim grid centre	CRU grid centre	GPCP grid centre
Fairbanks	147.750° W, 65.250° N	147.250° W, 65.250° N	146.250° W, 66.250° N
Kevo	27.000° E, 69.750° N	27.250° E, 69.750° N	26.250° E, 68.750° N
Tiksi	129.000° E, 71.250° N	128.750° E, 71.750° N	128.750° E, 71.250° N
Yakutsk	129.750° E, 62.250° N	129.750° E, 62.250° N	128.750° E, 61.250° N

Each value was set to the corresponding UTC time and then linearly interpolated with the next 6-hourly data to produce data with a 30 min resolution (T_{Lv0}^{30min}).

4.2.3 Specific humidity

Specific humidity for the Lv0 data (q_{Lv0}) was determined by a two-step procedure. First, relative humidity was calculated using the 6-hourly ERA-Interim dew points ($DewT_{ERA}^{6h}$) and air temperatures. Second, 30 min specific humidity was calculated from the previously obtained 6-hourly relative humidity data and the 2 m air temperature calculated in Sect. 4.2.2.

$$RH_{ERA}^{6h} = q \left(DewT_{ERA}^{6h} \right) / q_s \left(T_{ERA}^{6h} \right), \quad (2a)$$

$$q_{Lv0}^{30min} = q_s \left(T_{Lv0}^{30min} \right) \times RH_{ERA}^{6h}, \quad (2b)$$

where $q_s(T)$ denotes saturation specific humidity at temperature T .

4.2.4 10 m wind speed and direction

The 6-hourly wind speed and direction, calculated from zonal and meridional wind of ERA-Interim, was set as the Lv0 data at the corresponding UTC time, and repeated 11 times to create 30 min interval data in the 6 h range.

4.2.5 Total precipitation

The 6-hourly precipitation amounts (P_{Lv0}) were calculated from 6-hourly ERA-Interim precipitation, and a correction was applied in such a way that the monthly amount was proportional to the GPCP precipitation data:

$$P_{Lv0}^{6h} = P_{ERA}^{6h} \times \left(P_{GPCP}^{month} \right) / \sum_{month} \left(P_{ERA}^{6h} \right). \quad (3)$$

To determine how to distribute the 6 h precipitation data among the 30 min time slots, we conducted a preliminary analysis using the following three strategies and evaluated their influence on the simulated energy and water budget keeping other factors and variables fixed: (a) assign the entire amount to the first time slot (i.e. the entire 6 h precipitation amount falls in the first 30 min of the 6 h range); (b) assign the precipitation amounts in ratios of eight, four, two, and

one to the first four time slots (i.e. the first 2 h); and (c) assign equal amounts to the 12 time slots in the range. The results showed a distorted energy and water cycle for (c), while results were reasonable for the first two cases (not shown). Based on this outcome, we used strategy (a) and assigned the entire precipitation amount to the first time slot of every 6 h range.

4.2.6 Downward short-wave radiation

Short-wave radiation (SW), calculated as a theoretical value for latitude, longitude, day, and time for a particular site and day at 30 min intervals, was corrected using daily average cloudiness (Cl_{day}) as defined in the following equation:

$$SW_{Lv0}^{30min} = SW_{theory}(\text{lat, long, day, time}) \times Cl_{day},$$

$$Cl_{day} = \frac{\sum_{day} (SW_{ERA}^{6h})}{\sum_{day} (SW_{theory}(\text{lat, long, day, time}))}. \quad (4)$$

Comparison with observed values shows that thus calculated SW coincides well with observations in terms of daily averages, implying that the total incident solar energy is well represented. Figure 3 shows an example of an annual time series for daily maximum, average, and minimum of 30 min SW for Tiksi.

4.2.7 Downward long-wave radiation

ERA-Interim 6-hourly long-wave radiation data were set as the Lv0 data at the corresponding UTC time and repeated 11 times to create 30 min interval data for each 6 h range.

4.2.8 Surface pressure

ERA-Interim 6-hourly pressure data were set as the Lv0 data at the corresponding UTC time and repeated 11 times to create 30 min interval data.

4.3 Construction of level 1 data

4.3.1 General

For the Lv1 data set (Saito et al., 2014b), we applied a site-specific correction using observational data. The principle of Lv1 data construction is to derive rules to modify base data (ERA-Interim) to make them as similar to the observed values as possible. The formulations for statistical correction

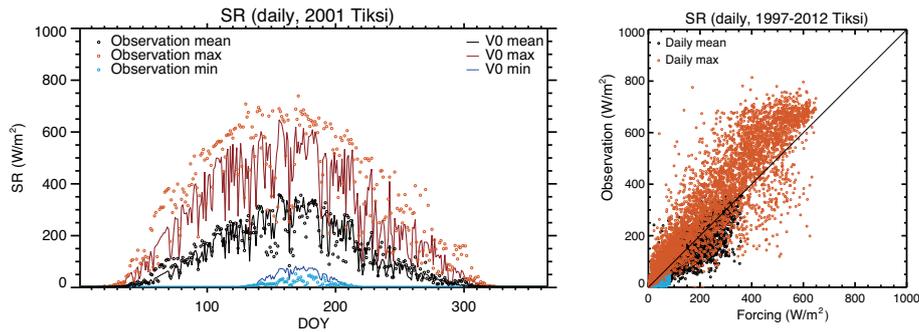


Figure 3. Comparison between level 0 short-wave radiation and observed values. Example shows 1-year time series of daily mean, minima, and maxima during 2001 (left), and scatter plot for all daily data (right) at Tiksi.

were established through comparison between observations and climate data (that is, ERA-Interim, CRU, and GPCP) for the respective observational period at each of the four sites (Table 1, Fig. 2). Observational periods are summarized in Table 1, and visualized by grey bars in the middle panel, below “GRENE-TEA site observation”, of Fig. 2. The Lv0 data were readily available with no or minor modifications for some variables, whereas site-specific treatment and adjustment were needed for other variables. Comparisons between observed and produced data, and the difference between Lv0 and Lv1 data are shown for selected sites and variables in Figs. 4–6. The overall quality of fit for all variables is shown in Fig. 7, which shows gradient a of the fitted line of daily mean observations y against Lv0, x ($y = ax + b$), and Pearson’s correlation coefficient r for x and y . The more a and r approach unity, the better the fit. Note that the order of independent (abscissa, x) and dependent (ordinate, y) variables of the fitting line (exemplified in Figs. 3b and 4c–f) is unconventional because a choice was made to show the necessary modification or corrections made to Lv0 to construct site-corrected Lv1.

4.3.2 2 m air temperature

Comparison with observations shows that the Lv0 temperature data agree well with respect to daily temperature averages, so that Eq. (1) is valid for creation of Lv1 data. The Lv0 data, however, show a smaller daily temperature range than the observed data, mainly because the Lv0 data were produced by interpolation of four points per day. Therefore, we employed two-step evaluations for the following targets (i.e. daily temperature range correction as shown by Eq. (5) and monthly average values correction as shown by Eq. (6), similar to Eq. (1) to produce Lv1 temperature data. The climate data (CD: CRU and/or ERA-Interim) that were used for the diurnal correction differed from site to site and were chosen based on the reproducibility of the observed data (not shown). The height of near-surface air temperature measurements at each site is summarized in Table 1. We chose observations at the height closest to 2 m but did not convert the

observation values for height because the reference height for air temperature is different among models.

First, the daily temperature range (dtr) was corrected for the ERA-Interim 6-hourly data, employing the CRU diurnal monthly temperature range for Fairbanks ($dtr_{CD} = dtr_{CRU}^{month}$) but daily ERA-Interim data (namely, 2 m temperature maximum and minimum; $dtr_{CD} = (T_{max}^{day} - T_{min}^{day})_{ERA}$) for the other three sites (cf. Eq. 5 of Sheffield et al., 2006, and Eq. 6 of Weedon et al., 2010):

$$T_{dtr_corrected}^{6h} = \left(T_{ERA}^{6h} \right) \Big|_{daily} + \left(T_{ERA}^{6h} - \left(T_{ERA}^{6h} \right) \Big|_{daily} \right) \times \left(dtr_{CD} / dtr_{ERA}^{6h} \right). \quad (5)$$

Second, monthly mean values were adjusted using CRU mean temperature ($^{\circ}C$), as shown for the creation of Lv0 data in Sect. 4.2.2, for Fairbanks, Kevo, and Yakutsk:

$$T_{Lv1}^{6h} = T_{dtr_corrected}^{6h} + \left(T_{CRU}^{month} + 273.15 \right) - \left(T_{ERA}^{6h} \right) \Big|_{month}. \quad (6)$$

We did not adjust Tiksi air temperature with CRU data, because the CRU values showed poorer reproducibility of the Tiksi site observations than the ERA-Interim monthly mean values (see Table 3), mostly due to the sparse network of stations in the region. However, the ERA-Interim values at the closest grid point to Tiksi show a more modest continentality (warmer summers and cooler winters) than the observed values at the coastal site. Therefore, a seasonal correction for Lv1 data ($Lv1 = T_{Lv1}^{6h}$) was applied to Lv0' ($= T_{dtr_corrected}^{6h}$) as follows (see also Fig. 4a):

$$Lv1 = 0.90 \times Lv0' + 0.1, \quad (\text{November–March}), \quad (7a)$$

$$Lv1 = Lv0'. \quad (\text{April–October}). \quad (7b)$$

Finally, these 6-hourly values (Lv1) were linearly interpolated with the next 6-hourly corrected data to produce data at 30 min resolution for all sites. Improvement from fitting the

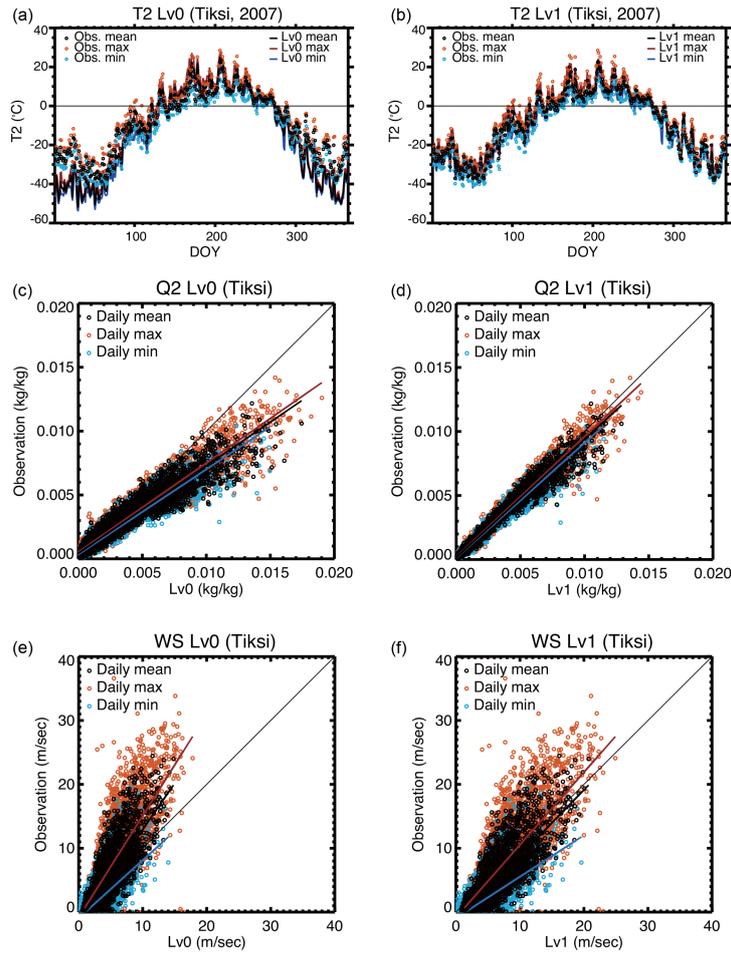


Figure 4. Comparison of GTMIP data with observed values for air temperature (time series in 2007), specific humidity (shown as a scatter plot for all daily data), and wind speed (shown as a scatter plot for all daily data). Examples show the results for level 0 (left column: **a**, **c** and **e**) and level 1 data (right column: **b**, **d**, and **f**) at Tiksi.

produced values to the observational data (Fig. 7a–c) is most apparent for the seasonal correction at Tiksi. For other sites, Lv0 data were already close to the observational data.

4.3.3 Specific humidity

Values of Lv1 specific humidity were determined using relative humidity values derived from ERA-Interim (Eq. 2a) and surface air temperature determined in Sect. 4.3.2. Figure 4b shows an example for differences in the relationship between Lv0 and Lv1 with observational data.

$$q_{Lv1}^{30min} = q_s \left(T_{Lv1}^{30min} \right) \times RH_{ERA}^{6h} \quad (8)$$

Improvement at Tiksi is a reflection of improvement in air temperatures as illustrated in Fig. 7d, which shows the daily mean for specific humidity.

4.3.4 10 m wind speed and direction

We determined 6-hourly wind speed and direction data from U wind and V wind of the ERA-Interim data set and converted the data to the corresponding UTC time. We defined these data as Lv0. Then, we applied a statistical correction based on the relationship between the observed values and Lv0 at Kevo and Tiksi (see Fig. 4c). The observational periods used for the statistical correction were 1999–2001 and 1997–2012 for Kevo and Tiksi, respectively. The equation for correction of the Lv1 data (Lv1) at each site can be written as follows:

$$Lv1 = 0.6 \times Lv0, \quad (\text{Kevo}), \quad (9a)$$

$$Lv1 = 1.4 \times Lv0 \quad (\text{Tiksi}). \quad (9b)$$

No correction of Lv0 wind speed was applied to Fairbanks and Yakutsk for the following reasons (Fig. 7k–m): Yakutsk already showed a good match between observed and Lv0 data when outliers and data collected during equipment malfunctions were removed, while Fairbanks showed such a

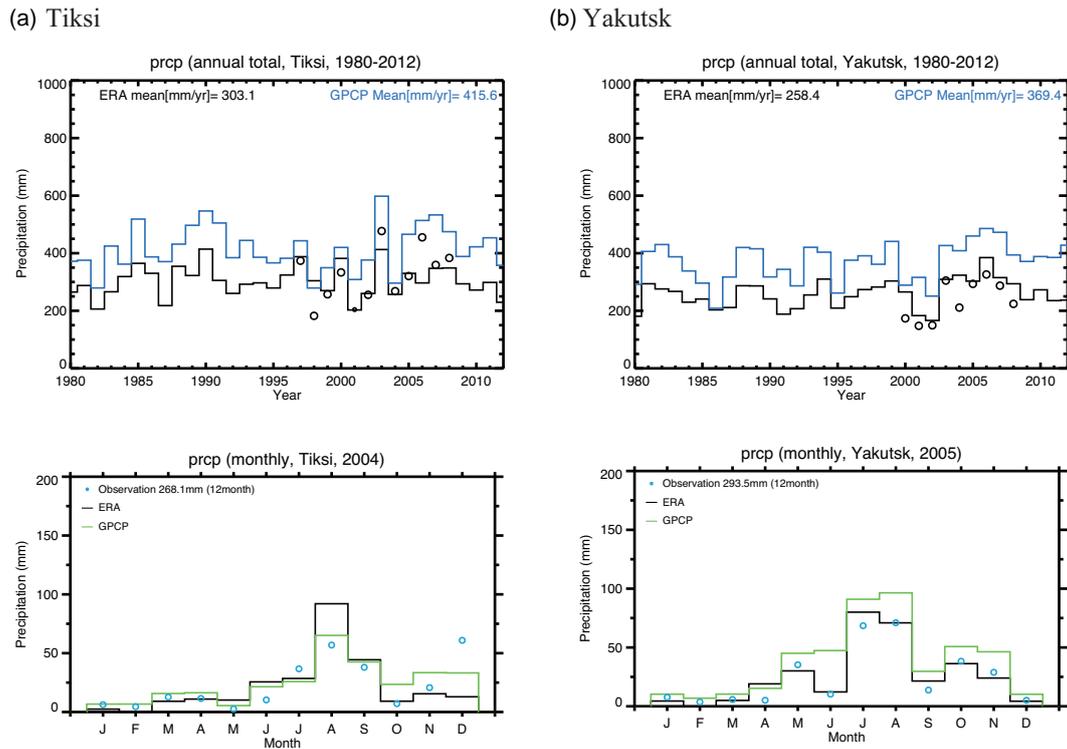


Figure 5. Comparison of annual (upper) and monthly (lower) precipitation amounts between ERA-Interim, GPCP, and observations (nearby stations). Examples show results for (a) Tiksi (left) and (b) Yakutsk (right). Annual precipitation is shown for period 1980–2012, and the monthly values are taken from 2004 for Tiksi and from 2005 for Yakutsk for exemplary purpose. Observations are shown as open circles, ERA data in black lines, and GPCP data in green lines. The annual value for 2001 for Tiksi is shown by a small filled circle for contrast because of exceptionally frequent missing data (151 days) in the year.

poor match between the two data sets that we decided that no statistic correction would have been appropriate. No correction of Lv0 was applied to the wind direction data at all four sites. The heights of the wind measurements at each site are summarized in Table 1. We chose the observations at the height closest to 10 m above ground (Kevo, Tiksi), or canopy (Fairbanks, Yakutsk), but did not convert the observation values for height because the reference height is different among different models.

4.3.5 Total precipitation

Precipitation is the most challenging part of data set construction but can have the largest impact on the simulated results. Reanalysis data (in the case of this study, ERA-Interim), which are virtually a model output, generally contain large uncertainties in precipitation and produce highly averaged and smoothed values because of a 6 h output interval. On the other hand, site observation data are also not always reliable, mainly due to instrumental problems. Another issue with site observations is short observation periods. For these reasons, the precipitation data used to constrain the Lv1 data were those collected at nearby weather stations, which we assumed may not necessarily correlate well on hourly or

daily scales but will show similar variations in precipitation amounts on a longer timescale (month or longer). Analysed precipitation products (in the case of this study, GPCP) have difficulties in high-latitude regions due to a scarcity of available data and poor satellite coverage. With these shortcomings of the data sources in mind, we created site-fitted precipitation data by evaluating the reproducibility of precipitation amounts and variations on a monthly and annual basis with respect to data from nearby weather stations.

For Fairbanks, we decided to use the Lv0 data as Lv1 data (Lv1) rather than attempting an unreliable correction, mostly because the two source data sets did not show any systematic relationship or good correlation with monthly observations, which were made at the international airport, located about 50 km from the site. For Kevo, comparison between the annual total of observed precipitation and that of the ERA-Interim data set showed a good inter-annual correlation, albeit with a quasi-constant underestimation. Since ERA-Interim data also showed a good correlation with intra-annual (monthly) variability, we used the ERA-Interim as the basic data set Lv0', that is, without any monthly correction by GPCP. The resulting correction equation is

$$\text{Lv1} = 0.76 \times \text{Lv0}' \quad (10)$$

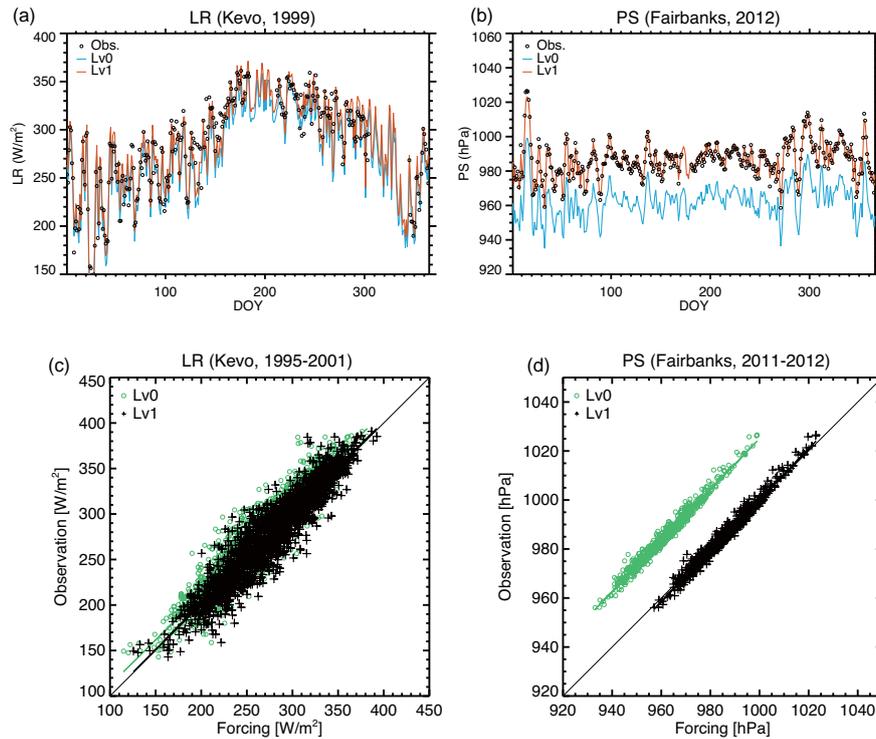


Figure 6. Comparison of GTMIP forcing data and observed values for (a) long-wave radiation (LR) of Kevo (in 1999), and (b) surface pressure (PS) of Fairbanks (in 2011) on a daily basis. Level 0 data are shown in blue and level 1 data in red. Correlation between the forcing data sets and observed values for these examples are also shown for (c) LR of Kevo (in 1995–2001), and (d) PS of Fairbanks (in 2011–2012). Level 0 data are shown in green and level 1 data in black.

For Tiksi, GPCP provided better monthly variations while ERA-Interim data showed better annual total and inter-annual variations. Therefore, we applied a correction to Lv0 using the ratio of annual precipitation of the ERA-Interim data to the annual precipitation of GPCP (Fig. 5a). In contrast, as the ERA-Interim 6-hourly total precipitation amounts agreed well with the observational data both on a monthly and annual scale for Yakutsk, they were directly used as the 6-hourly values without GPCP correction (Fig. 5b).

Finally, the 6-hourly precipitation amounts were assigned to the first time slot of every 6 h range to produce a data set with 30 min intervals. Note that the correlation of daily values is not particularly high (Fig. 7f) relative to that of other variables shown in Fig. 7. This is a natural result since fitting is done only on a monthly and annual basis. The overall improvement is, however, already visible as Lv1 is closer to the 1 : 1 gradient line (i.e. $a = 1$) than Lv0.

4.3.6 Downward short-wave radiation

As stated in Sect. 4.2.6, downward short-wave radiation is computed from the theoretical instantaneous radiation value decreased by daily average cloudiness, which corresponds well to the daily mean incident energy at Earth’s surface.

Some discrepancies were found between Lv0 and the observational data with regard to daily maximum (Fig. 7g) and/or minimum (Fig. 7i). Possible causes are short-time cloud cover or absence of clouds, shading by forested areas (as is the case for the minimum SW in Kevo), or equipment malfunction. As it is difficult to generalize these situations in a statistical formula, no correction of Lv0 was applied for downward short-wave radiation data for all four sites. Figure 7i does not show data for Fairbanks and Yakutsk since the two stations lie south of the Arctic Circle, and hence the daily minimum is always zero.

4.3.7 Downward long-wave radiation

Comparison between observed values and the ERA-Interim output (equivalently, Lv0) revealed that the difference between the two data sets can largely be accounted for by emission differences of the air layer due to altitude differences between the observation sites and the closest ERA-Interim grid points (see Fig. 6a and c as an example for Kevo). The resulting equations for creating Lv1 data, Lv1, are as follows:

$$\text{Lv1} = 1.01 \times \text{Lv0} + 15 \text{ (Fairbanks)}, \quad (11a)$$

$$\text{Lv1} = 1.00 \times \text{Lv0} + 11 \text{ (Kevo)}, \quad (11b)$$

$$\text{Lv1} = 0.85 \times \text{Lv0} + 50 \text{ (Tiksi)}, \quad (11c)$$

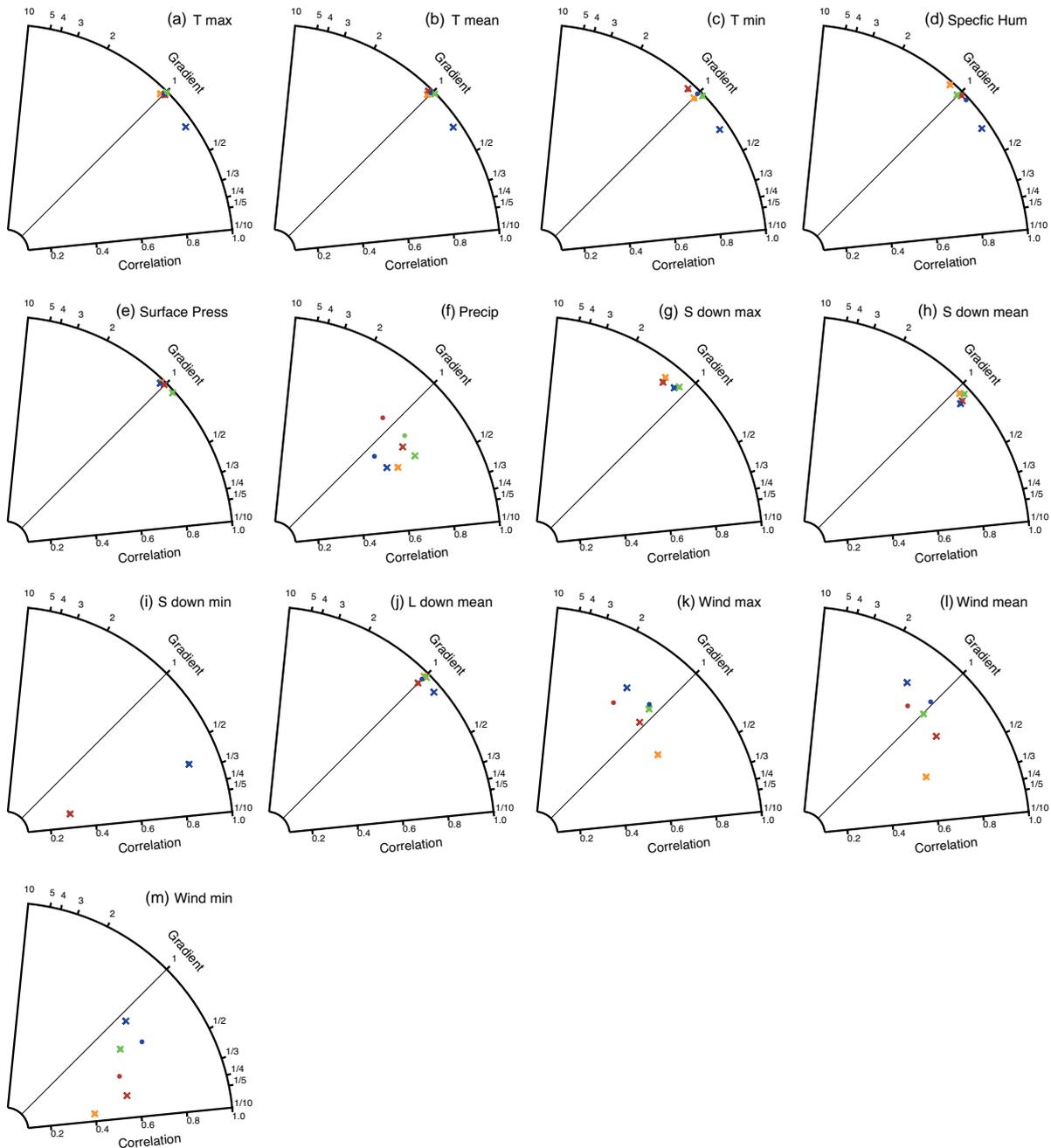


Figure 7. Correlation diagrams for GTMIP data and daily site observation data showing Pearson’s correlation coefficient and gradient of the fitted line for (a) maximum air temperature, (b) mean air temperature, (c) minimum air temperature, (d) specific humidity, (e) surface pressure, (g) maximum downward short-wave radiation, (h) mean downward short-wave radiation, (i) minimum downward short-wave radiation, (j) downward long-wave radiation, (k) maximum wind speed, (l) mean wind speed, and (m) minimum wind speed. Lv0 data are shown as crosses and Lv1 data as dots. Sites are shown by colours, i.e. Fairbanks (orange), Kevo (red), Tiksi (blue), and Yakutsk (green).

$$Lv1 = 0.98 \times Lv0 + 17 \text{ (Yakutsk)}. \quad (11d)$$

Figure 7j shows a marked improvement for Tiksi. This improvement is mostly manifested by the location’s gradient (0.85), which reflects a seasonal difference due to the con-

tinental character of the reanalysis value at the grid point closest to Tiksi against more coastal characteristics of the Tiksi observational site. Since diurnal cycles in downward long-wave and surface pressure are not as prominent as in downward short-wave radiation or air temperature, only daily means are shown in Fig. 7.

4.3.8 Surface pressure

Similar to downward long-wave radiation, differences in surface pressure between the observed and reanalysis data (i.e. Lv0 data, Lv0) are largely correlated with altitude differences (see Fig. 6b and d as an example for Fairbanks). The resulting correction equations are as follows:

$$Lv1 = Lv0 + 24 \text{ (Fairbanks)}, \quad (12a)$$

$$Lv1 = Lv0 + 21 \text{ (Kevo)}, \quad (12b)$$

$$Lv1 = Lv0 + 19 \text{ (Tiksi)}, \quad (12c)$$

$$Lv1 = Lv0 - 5 \text{ (Yakutsk)}. \quad (12d)$$

Since changes in the intercept of the fitting lines in Eq. (11) do not change the gradient and correlation coefficient, Fig. 7e does not show any difference between Lv0 and Lv1.

4.4 Additional data

In addition to the two model driving data sets described in the previous section, we have prepared the following two data sets for standardization and convenience for model integration and spin-up: photosynthetically active radiation (fPAR) and leaf area index (LAI) data sets (Saito et al., 2014c), and a 20-year detrended meteorological driving data set (Saito et al., 2015). The former data set was derived from Moderate Resolution Imaging Spectroradiometer (MODIS) satellite data (MOD15A2, MYD15A2) according to algorithms developed and validated by Sasai et al. (2011). Monthly averaged time series were processed for each of the six GRENE-TEA sites from the closest grid point values.

The latter detrended driving data set, produced by removing monthly long-term trends from the Lv1 data for the period of 1980–1999, was provided for spin-up to include initial soil carbon conditions comparable to other biogeochemical models without being affected by warming trends (Saito et al., 2015).

5 Discussion and outlook

Concepts and principles of a correction method for creating a forcing data set from climate product data (e.g. ERA-Interim, CRU, and GPCP) and site observations were established through a comparison of the two data sources over the observational period as described in this article. The set of observation-based, site-specific data thus produced can be expected to produce a more realistic time series than the climate product data themselves and will allow for a more meaningful comparison to observed target variables at a given site (e.g. evapotranspiration, snowpack, ground temperature and soil moisture, and gross primary products).

Once correction rules (formulae) are defined, a time series can be extrapolated to a period of no observation. Not only can these data sets (Lv0 and Lv1, in this case) be used as common driving data for model intercomparison projects to

examine inter-site and inter-model variations with regard to performance and uncertainty, they are also useful for investigating long-term ecoclimatic variations by analysing the time series per se that incorporates the knowledge of on-going observations and by examining and testing the capability of reproduction and hindcast through model simulations driven by the time series. Naturally, the available time period is still limited by the time covered by the product-based data, and the efficacy of the rules used to create the product-based data should be thoroughly tested when these data sets are applied.

Another target of this model intercomparison activity was to facilitate collaboration between modelling and field scientists. Such collaboration can be difficult and is often insufficient due to discrepancies between science targets and approaches, terminology, and different prioritization during data acquisition among the two groups (for example, field scientists tend to seek for unusual data in their area of investigation, while modellers often need typical data for an area; cf. Saito et al., 2014d).

Considering the aforementioned discrepancies, the GTMIP study was conducted by a team of modelling and field scientists, which was especially important when the quality of observed data used for Lv1 creation was questioned (e.g. outliers, missing values, representativeness of the site and nearby station data). Similarly, when conducting model performance checks using observational data, which is an important and unavoidable process for model development, field scientists joined together to discuss issues with regard to modelled and observed data, capability, and limits of models depending on the extent and complexity of the implemented processes, often concluding that a precise observation does not necessarily coincide with a “correct” model output.

The detailed protocol of GTMIP Stage 1 (site simulations) is described in our companion paper (Miyazaki et al., 2015).

Data availability

All data presented in this paper, as well as an overall outline of the GRENE-TEA model intercomparison project (GTMIP), are available from the Arctic Data archive System (ADS) website (doi:10.17592/001.2015093001 or https://ads.nipr.ac.jp/gtmip/gtmip.html) (Sueyoshi et al., 2015). Included are the metadata of each data set, read-me files and a concise explanation of the data creation procedure.

Acknowledgements. This study was supported by the GRENE Arctic Climate Change Research Project, Ministry of Education, Culture, Sports, Science and Technology, Japan.

Edited by: G. König-Langlo

References

- Adler, R. F., Huffman, G. J., Chang, A., Ferraro, R., Chang, A., Ferraro, R., Xie, P. P., Janowiak, J., Rudolf, B., Scheneider, U., Curtis, S., Bolvin, D., Gruber, A., Susskind, J., Arkin, P., and Nelkin, E.: The Version-2 Global Precipitation Climatology Project (GPCP) Monthly Precipitation Analysis (1979–Present), *J. Hydrometeorol.*, 4, 1147–1167, 2003.
- Dee, D. P., Uppalaa, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, I., Kallberg, P., Kohler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thepaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, *Q. J. Roy. Meteor. Soc.*, 137, 553–597, 2011.
- Harris, I., Jones, P. D., Osborn, T. J., and Lister, D. H.: Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 Dataset, *Int. J. Climatol.*, 34, 623–642, 2014.
- Iwahana, G., Takano, S., Petrov, R. E., Tei, S., Shingubara, R., Maximov, T. C., Fedrov, A. N., Desyatkin, A. R., Nikolaev, A. N., Desyatkin, R. V., and Sugimoto, A.: Geocryological characteristics of the upper permafrost in a tundraforest transition of the Indigirka River Valley, Russia, *Polar Sci.*, 8, 96–113, doi:10.1016/j.polar.2014.01.005, 2014.
- Kodama, Y., Sato, N., Yabuki, H., Ishii, Y., Nomura, M., and Ohata, T.: Wind direction dependency of water and energy fluxes and synoptic conditions over a tundra near Tiksi, Siberia, *Hydrol. Process.*, 21, 2028–2037, 2007.
- Kotani, A., Kononov, A. V., Ohta, T., and Maximov, T. C.: Temporal variations in the linkage between the net ecosystem exchange of water vapour and CO₂ over boreal forests in eastern Siberia, *Ecohydrol.*, 7, 209–225, doi:10.1002/eco.1449, 2013.
- Lopez, M. L., Saito, H., Kobayashi, K., Shirota, T., Iwahana, G., Maximov, T. C., and Fukuda, M.: Interannual environmental-soil thawing rate variation and its control on transpiration from Larix cajanderi, Central Yakutia, Eastern Siberia, *J. Hydrol.*, 338, 251–260, doi:10.1016/j.jhydrol.2007.02.039, 2007.
- Miyazaki, S., Saito, K., Mori, J., Yamazaki, T., Ise, T., Arakida, H., Hajima, T., Iijima, Y., Machiya, H., Sueyoshi, T., Yabuki, H., Burke, E. J., Hosaka, M., Ichii, K., Ikawa, H., Ito, A., Kotani, A., Matsuura, Y., Niwano, M., Nitta, T., Oishi, R., Ohta, T., Park, H., Sasai, T., Sato, A., Sato, H., Sugimoto, A., Suzuki, R., Tanaka, K., Yamaguchi, S., and Yoshimura, K.: The GRENE-TEA model intercomparison project (GTMIP): overview and experiment protocol for Stage 1, *Geosci. Model Dev.*, 8, 2841–2856, doi:10.5194/gmd-8-2841-2015, 2015.
- Nakai, Y., Matsuura, Y., Kajimoto, T., Abaimov, A. P., Yamamoto, S., and Zyryanova, O. A.: Eddy covariance CO₂ flux above a Gmelin larch forest in continuous permafrost of central Siberia during a growing season, *Theor. Appl. Climatol.*, 93, 133–147, doi:10.1007/s00704-007-0337-x, 2008.
- Nakai, T., Kim, Y., Busey, R. C., Suzuki, R., Nagai, S., Kobayashi, H., Park, H., Sugiura, K., and Ito, A.: Characteristics of evapotranspiration from a permafrost black spruce forest in interior Alaska, *Polar Sci.*, 7, 136–148, 2013.
- Ohta, T., Hiyama, T., Tanaka, H., Kuwada, T., Maximov, T. C., Ohata, T., and Fukushima, Y.: Seasonal variation in the energy and water exchanges above and below a larch forest in Eastern Siberia, *Hydrol. Process.*, 15, 1459–1476, 2001.
- Ohta, T., Maximov, T. C., Dolman, A. J., Nakai, T., van der Molen, M. K., Kononov, A. V., Maximov, A. P., Hiyama, T., Iijima, Y., Moors, E. J., Tanaka, H., Toba, T., and Yabuki, H.: Interannual variation of water balance and summer evapotranspiration in an Eastern Siberian larch forest over a 7-year period (1998–2006), *Agr. Forest Meteorol.*, 148, 1941–1953, 2008.
- Ohta, T., Kotani, A., Iijima, Y., Maximov, T. C., Ito, S., Hanamura, M., Kononov, A. V., and Maximov, A. P.: Effects of waterlogging on water and carbon dioxide fluxes and environmental variables in a Siberian larch forest, 1998–2011, *Agr. Forest Meteorol.*, 188, 64–75, 2014.
- Saito, K., Miyazaki, S., Mori, J., Ise, T., Arakida, H., Sueyoshi, T., Hajima, T., Iijima, Y., Yamazaki, T., and Sugimoto, A.: GTMIP meteorological driving dataset for the GRENE-TEA observation sites (level 0.2), 0.20, Arctic Data archive System (ADS), Japan, available at: <https://ads.nipr.ac.jp/dataset/A20141009-005> (last access: 28 April 2015), 2014a.
- Saito, K., Miyazaki, S., Mori, J., Ise, T., Arakida, H., Suzuki, R., Sato, A., Iijima, Y., Yabuki, H., Iijima, Y., Sueyoshi, T., Hajima, T., Sato, H., Yamazaki, T., and Sugimoto, A.: GTMIP meteorological driving dataset for the GRENE-TEA observation sites (level 1.0), 1.00, Arctic Data archive System (ADS), Japan, available at: <https://ads.nipr.ac.jp/dataset/A20141009-006> (last access: 28 April 2015), 2014b.
- Saito, K., Sasai, T., Miyazaki, S., Mori, J., Ise, T., Arakida, H., Sueyoshi, T., Hajima, T., Iijima, Y., Yamazaki, T., and Sugimoto, A.: GTMIP fraction of photosynthetically active radiation (fPAR) and leaf area index (LAI) for the GRENE-TEA observation sites (level 1.0), 1.00, Arctic Data archive System (ADS), Japan, available at: <https://ads.nipr.ac.jp/dataset/A20141009-007> (last access: 28 April 2015), 2014c.
- Saito, K., Miyazaki, S., Hajima, T., and Sueyoshi, T.: Climate and Environmental variations in the Northern terrestrial cryosphere, *Kisho-Kenkyu Note (Meteorological Research Note)*, 230, 196–211, 2014d. (in Japanese)
- Saito, K., Miyazaki, S., Mori, J., Ise, T., Arakida, H., Suzuki, R., Sato, A., Iijima, Y., Yabuki, H., Iijima, Y., Sueyoshi, T., Hajima, T., Sato, H., Yamazaki, T., and Sugimoto, A.: GTMIP meteorological driving dataset for the GRENE-TEA observation sites (20-year detrended), 1.00, Arctic Data archive System (ADS), Japan, available at: <https://ads.nipr.ac.jp/dataset/A20150205-001> (last access: 28 April 2015), 2015.
- Sasai, T., Saigusa, N., Nasahara, K. N., Ito, A., Hashimoto, H., Nemani, R. R., Hirata, R., Ichii, K., Takagi, K., Saitoh, T. M., Ohta, T., Murakami, K., Yamaguchi, Y., and Oikawa, T.: Satellite-driven estimation of terrestrial carbon flux over Far East Asia with 1-km grid resolution, *Remote Sens. Environ.*, 115, 1758–1771, doi:10.1016/j.rse.2011.03.007, 2011.
- Sato, A., Kubota, H., Matsuda, M., and Sugiura, K.: Seasonal variation of heat exchange in the boreal forest of Finnish Lapland, *Second Wadati Conference, on Global Change and the Polar Climate*, extended abstracts, 228–230, 2001.
- Sheffield, J., Goteti, G., and Wood, E. F.: Development of a 50-yr high-resolution global dataset of meteorological forcings for land surface modeling, *J. Climate*, 19, 3088–3111, 2006.

- Sueyoshi, T., Saito, K., Miyazaki, S., Mori, J., Ise, T., Arakida, H., Rikie Suzuki, R., Sato, A., Iijima, Y., Yabuki, H., Ikawa, H., Ohta, T., Kotani, A., Hajima, T., Sato, H., Yamazaki, T., and Sugimoto, A.: The GRENE-TEA Model Intercomparison Project (GTMIP) stage 1 forcing dataset, 1.00, Arctic Data archive System (ADS), Japan, doi:10.17592/001.2015093001, 2015.
- Vuichard, N. and Papale, D.: Filling the gaps in meteorological continuous data measured at FLUXNET sites with ERA-Interim reanalysis, *Earth Syst. Sci. Data*, 7, 157–171, doi:10.5194/essd-7-157-2015, 2015.
- Watanabe, K., Mizoguchi, M., Kiyosawa, H., and Kodama Y.: Properties and horizons of active layer soils in tundra at Tiksi, Siberia, *Journal of Japan Society of Hydrology and Water Resources*, 13, 9–16, 2000 (in Japanese with English abstract).
- Weedon, G. P., Gomes, S., Viterbo, P., Österle, H., Adam, J. C., Bellouin, N., Boucher, O., and Best, M.: The WATCH forcing data 1958–2001: A meteorological forcing dataset for land surface and hydrological models, WATCH Tech. Rep. 22, available at: <http://www.eu-watch.org/publications/technical-reports> (last access: 20 April 2015), 41 pp., 2010.
- Weedon, G. P., Balsamo, G., Bellouin, N., Gomes, S., Best, M. J., and Viterbo, P.: The WFDEI meteorological forcing data set: WATCH Forcing Data methodology applied to ERA-Interim reanalysis data, *Water Resour. Res.*, 50, 7505–7514, doi:10.1002/2014WR015638, 2014.
- Yabuki, H., Park, H., Kawamoto, H., Suzuki, R., Razuvaev, V. N., Bulygina, O. N., and Ohata, T.: Baseline Meteorological Data in Siberia (BMDS) Version 5.0, RIGC, JAMSTEC, Yokosuka, Japan, distributed by CrDAP, Digital media, available at: <http://www.jamstec.go.jp/acdap/Summary.action?selectFile=A20110307-001> (last access: 12 January 2016), 2011.