Using GRACE to derive corrections to precipitation data sets and improve modelled snow mass at high latitudes

Emma L. Robinson[1] and Douglas B. Clark[1]

[1]UK Centre for Ecology & Hydrology, Maclean Building, Benson Lane, Crowmarsh Gifford, Wallingford OX10 8BB

Correspondence: Emma L. Robinson (emrobi@ceh.ac.uk)

Abstract. The amount of lying snow calculated by a land surface model depends in part on the amount of snowfall in the meteorological data that are used to drive the model. We show that commonly-used data sets differ in the amount of snowfall, and more generally precipitation, over four large Arctic basins. An independent estimate of the cold season precipitation is obtained by combining water balance information from the Gravity Recovery and Climate Experiment (GRACE) with estimates of evaporation and river discharge, and is generally higher than that estimated by four commonly-used meteorological data sets. We use the Joint UK Land Environment Simulator (JULES) land surface model to calculate the snow water equivalent (SWE) over the four basins. The modelled seasonal maximum SWE is 38% less than observation-based estimates on average and the modelled basin discharge is significantly underestimated, consistent with the lack of snowfall. We use the GRACE-derived estimate of precipitation to define per-basin scale factors that are applied to the driving data and increase the amount of cold season precipitation by 28% on average. In turn this increases the modelled seasonal maximum SWE by 30%, although this is still underestimated compared to observations by 19% on average. A correction for undercatch of precipitation by gauges is compared with the the GRACE-derived correction. Undercatch correction increases the amount of cold season precipitation by 23% on average, which indicates that some, but not all, of the underestimation can be removed by implementing existing undercatch correction algorithms. However, even undercatch-corrected data sets contain less precipitation than the GRACE-derived estimate in some regions, and it is likely that there are other biases that are not currently accounted for in gridded meteorological data sets. This study shows that revised estimates of precipitation can lead to improved modelling of SWE, but much more modest improvements are found in modelled river discharge. By providing methods to better define the precipitation inputs to the system, the current study paves the way for subsequent work on key hydrological processes in high-latitude basins.

1 Introduction

Seasonal snow cover is an important part of the hydrological cycle over a large part of the Northern Hemisphere, with approximately 45 million km^2 of the land (excluding Greenland) covered by snow at the seasonal maximum (Mudryk et al., 2014). Snow plays an important role in the energy, water and biogeochemical cycles of these areas. The accumulation of water in the snow pack until the spring thaw dominates the seasonal cycle of runoff and river flow in many northern river basins (Grabs et al., 2000) and the physical properties of the snow pack, such as its high albedo, can lead to strong links between the snow pack and the overlying atmosphere (Cohen and Rind, 1991; Mote, 2008).

Many studies have documented changes in seasonal snow cover in recent decades. The extent of Northern Hemisphere snow cover decreased through most of the last century, with a larger rate of decrease since 1970 (Brown and Robinson, 2011). Correlation between changes in spring snow extent and air temperature (Brown and Robinson, 2011) reveal the effects of the snow-albedo feedback (Fernandes et al., 2009). Observed changes in snow water equivalent (SWE) show more variation

5   between regions. The seasonal maximum SWE has increased over northern Eurasia over 1966–2010 (Bulygina et al., 2011) but springtime SWE over the mountains of western North America has decreased over 1960–2002 (Mote, 2006).

Overall the picture is one of considerable variation depending on the years and season of analysis, locations and snow measure (Brown and Mote, 2009). However, declines in snow cover over recent decades are more common than increases, particularly at warmer or lower locations (see Fig. 4.21 in Vaughan et al., 2013).

10   These analyses of historical snow cover have used a variety of data sources, principally ground-based point observations of precipitation and snow cover, and products derived from satellite data, each of which has strengths and weaknesses. Direct measurements of snowfall using gauges and of characteristics of the snow pack (e.g. SWE) are difficult to scale up to large areas for reasons including the high spatial variability in these quantities and the sparse network of observing stations (e.g. Liston and Hiemstra, 2011). Inconsistent trends in precipitation and runoff in large northern catchments confirm that these data

15   are of uncertain quality (Berezovskaya et al., 2004; Pavelsky and Smith, 2006). Satellite data can be used for recent decades, particularly to determine the presence or absence of snow cover. Passive microwave data from satellites can be used to assess SWE (e.g. GlobSnow; Takala et al., 2011), but the need for an algorithm to convert the measured radiance to SWE, coupled with the requirements for supporting data such as ground-based observations of meteorology, introduces uncertainty into the final estimates.

20   An alternative to these observation-based methods is to use a land surface model (LSM) to describe and understand the evolution of the snow pack (e.g. Liston and Hiemstra, 2011). There are uncertainties associated with LSMs, and they require large amounts of input data, but they have the advantage of also being able to estimate future changes in snow cover as they are also used as the land surface components of the climate models that are used to study future climates. Simulations using climate models indicate that Northern Hemisphere snow cover extent will decrease in the future (Brutel-Vuilmet et al., 2013).

25   Although warming will reduce the length of the snow-cover season, precipitation is predicted to increase over many mid- to high-latitude areas, meaning the impact of climate change on the amount and distribution of SWE may differ between regions (Räisänen, 2008). Any such changes in snowfall and snow cover will impact the wider hydrological cycle in these areas.

If we are to have confidence in these analyses of historical and possible future conditions from LSMs, we must first establish that they can accurately represent the snow pack and related processes. To this end community intercomparison experiments

30   have quantified the performance of land surface and other models in a variety of settings, some with a focus on snow processes and others looking at the wider hydrological cycle (e.g. Bowling et al., 2003; Essery et al., 2009; Haddeland et al., 2011). A recurring theme across many of these has been the uncertainty that stems from the input precipitation data, particularly in regions with seasonal snow cover.

This theme is also evident in snow simulations that use the Joint UK Land Environment Simulator (JULES) land surface

35   model (Best et al., 2011; Clark et al., 2011). Burke et al. (2013) noted that seasonal maximum SWE in pan-Arctic simulations

of JULES was often underestimated in comparison to GlobSnow. This is consistent with the fact that the snowfall specified in the meteorological data that are used to drive JULES and other LSMs is often considerably less than the SWE estimated by GlobSnow (Hancock et al., 2014). Ménard et al. (2015) suggest that the largest uncertainty in estimates of SWE from JULES arises from uncertainty in the input precipitation data.

5   A larger body of research concerns the ability of models to represent other aspects of land hydrology, particularly runoff and river flow, because measurements of river flow are generally considered more accurate than those of catchment-average precipitation and snow pack (Troy et al., 2011). Again a recurring finding of this work is the need to supply the models with high quality precipitation data and the uncertainty in these data (Decharme and Douville, 2006; Fiedler and Döll, 2007; Tian et al., 2007; Biemans et al., 2009; Wisser et al., 2010; Weiland et al., 2015; Islam and Déry, 2017; Casson et al., 2018).

10   Furthermore, precipitation data are not only uncertain but possibly systematically biased low with studies suggesting that insufficient modelled river flow in high latitude catchments is a result of insufficient precipitation in the driving meteorology (Tian et al., 2007; Biemans et al., 2009; Alkama et al., 2010). It should be noted that these model biases are not limited to land surface models (those that can be coupled to atmospheric models) but are also shown by global hydrology models, which typically focus on water resources and lateral fluxes (Haddeland et al., 2011).

15   Estimates of precipitation in areas with significant amounts of snow are made particularly difficult by the sparsity of the gauge network in many locations, the difficulty of obtaining an areal average in mountainous terrain, and the need to correct gauge measurements for undercatch and losses (Serreze et al., 2003; Adam and Lettenmaier, 2003; Adam et al., 2006). Correcting for undercatch requires information on (at least) the equipment used, wind speed, and the phase of precipitation, and is itself a difficult and uncertain task (Serreze et al., 2003). Adam and Lettenmaier (2003) derived monthly correction factors for
20   wind-induced undercatch and wetting loss that resulted in a 16% increase in DJF precipitation at the global scale, with much higher values in some regions. Wind-induced undercatch of solid precipitation was the largest source of bias. For stations north of 45° N, Yang et al. (2005) calculated monthly correction factors of 80–120% of winter precipitation winter due to snowfall and higher wind speeds. Correcting for orographic effects resulted in 6% higher global precipitation, 20% in orographically-influenced areas, using a water balance approach (Adam et al., 2006). Although these correction factors are substantial, they
25   are themselves uncertain and often only apply to a particular time period and data set, so cannot necessarily be used generally. Undercatch correction is not routinely to all driving data sets, and some which do not include any correction factors are widely used (e.g. CRUNCEP Viovy, 2014).

Given the difficulties associated with gauge-based data, other products use information from satellite-based earth observation or from atmospheric reanalyses, either alone or in combination with gauge data. However these approaches involve their
30   own uncertainties, such as how to relate the measured radiance to a precipitation rate, and the data sets are also subject to considerable uncertainty (Stephens and Kummerow, 2007; Beck et al., 2017).

Given that areal precipitation is difficult to estimate directly using the aforementioned methods, an alternative approach instead seeks to infer precipitation from analysis of the total water storage (TWS) of the land surface as estimated by the Gravity Recovery and Climate Experiment (GRACE) mission (Swenson, 2012). GRACE uses gravimetry to assess the changing mass
35   of the surface of the Earth, which can be interpreted as changes in water storage, revealing the dynamics of TWS at the

surface on a monthly time scale. While the monthly changes in TWS can be observed, GRACE does not indicate what fluxes or components are the cause of the changes. However, by using the GRACE TWS anomalies, combined with observed or modelled estimates of river flow and gridded evaporation fluxes, it is possible to estimate the areal precipitation (Swenson, 2010).

5   Previous studies have used GRACE TWS to estimate precipitation inputs in boreal regions (Swenson, 2010; Seo et al., 2010; Behrangi et al., 2016) and cold mountainous basins (Behrangi et al., 2017). These have identified deficiencies or uncertainties in traditional estimates of precipitation, but with magnitude that varies between products, locations and times. Swenson (2010) compared GRACE-derived precipitation with two existing precipitation data sets and found a varied picture, with more variability in North America, while many precipitation products are significantly lower than the GRACE-derived estimate in
10   cold mountainous endorheic basins. Behrangi et al. (2016) compared several raw reanalysis outputs, as well as some satellite and gauge based products, and only found deficiencies in gauge-based data sets in some regions. These studies have all used LSM or reanalysis output to provide at least some of the ancillary data required to calculate precipitation from GRACE TWS. Since, in this study, the intention is to investigate how changes to input precipitation can result in changes to the LSM output, independent estimates of evapotranspiration and river flow have been used in preference to LSM output.

15   The remainder of this study investigates the hypothesis that many precipitation data sets that are used to drive land surface models contain insufficient cold season precipitation in mid- to high-latitude land areas, leading to underestimation of SWE and river flow. Cold-season precipitation data are assessed against an estimate based on GRACE TWS data combined with independent estimates of evaporation and basin discharge. The GRACE-derived precipitation estimates are used to rescale the cold season precipitation in the data sets used to drive the JULES model, and the impacts on SWE and river discharge are
20   investigated. This is also compared to the effects of undercatch correction.

In Sect. 2 the data sets used to carry out this study are described and in Sect. 3 the set up of the JULES model and the experiments carried out are defined. The method by which precipitation is calculated from GRACE TWS is described in Sect. 4. The JULES model runs are described and evaluated in Sect. 5 and the implications are discussed in Sect. 6.

2   Data sets

25   All analyses in this paper are carried out using data on a regular latitude-longitude grid at $1°$ resolution. Data with other resolutions are first regridded to this common grid with a modified version of SCRIP (Jones, 1998), using the conservative remapping method normalised by destination area fraction (Jones, 1999). This study focuses on the four major Arctic basins — the Ob, Yenisei, Lena and Mackenzie — which together account for approximately 68% of the total discharge into the Arctic Ocean (Grabs et al., 2000) and are marked by O, Y, L and M respectively in Fig. 1. The basins are defined using the
30   $1°$ TRIP river network data of Oki and Sud (1998). For consistency with the observations of river flow (see Sect. 2.3), only the part of each basin that drains to the flow gauging station was considered when calculating spatial averages of observed or modelled grids. The data sets used all vary in their temporal coverage; for consistency in analysis a "common overlap period" of 2002 – 2010 inclusive was used, during which time all of the driving data sets and the data sets used for the water balance

precipitation estimates were available. In order to capture full annual cycles of snow accumulation and melt, annual averages were calculated from September 1st of one year to August 31st of the next, so the common overlap period contains eight full years for analysis. All comparisons were carried out using data from this common overlap period, and averaged over the defined basins, unless explicitly noted otherwise.

2.1 Meteorological data

A key requirement of land surface models is a time series of meteorological data. To this end, several "driving" data sets have been created that represent realistic global climate for the recent past. These are generally based on global meteorological reanalyses which have been bias-corrected to observational data sets. They provide a full set of the meteorological variables required to run an LSM, including air temperature, pressure and humidity, wind speed, incoming long- and shortwave radiation, as well as the precipitation with which this study is concerned.

Four meteorological data sets were used to drive the JULES model in this study. These are summarised in Table 1 and described below, with a focus on the precipitation data. The data were available at 0.5° resolution unless stated otherwise. Basin mean monthly climatologies of all of the driving variables can be seen in Figs. S1 and S2.

The CRUNCEP v4 data set (CRUNCEP; Viovy, 2014) uses the CRU TS3.21 monthly data (Harris et al., 2013), including gauge-based precipitation, to bias correct the NCEP/NCAR Reanalysis 1 (Kalnay et al., 1996). The reanalysis is spatially interpolated from its original 2.5° resolution to the required 0.5° resolution with no adjustment for grid box elevation. No undercatch correction is applied to the gauge data and the total precipitation is provided (rather than separate estimates of rainfall and snowfall).

The Princeton Global Forcing data v2 (PGF; Sheffield et al., 2006) is based on the NCEP/NCAR Reanalysis 1 with precipitation totals scaled to CRU TS monthly values and further statistical corrections applied to correct the number of rain days and to to scale to the final 1° and 3-hourly resolutions. Again total precipitation is provided. The reanalysis is spatially interpolated, including a correction for differences in grid box elevation. The version used in this study (v2) was *not* corrected for undercatch (Justin Sheffield, personal communication, 2018) contrary to the description of the original data set in Sheffield et al. (2006).

Two variants of the WATCH Forcing Data methodology applied to ERA-Interim Data (WFDEI; Weedon et al., 2014) were used, differing only in the precipitation data used for bias correction: WFDEI-CRU uses CRU TS 3.1.01 until 2009 and CRU TS 3.21 from 2010 onwards (Harris et al., 2013; Trenberth et al., 2013), while WFDEI-GPCC uses GPCC v5 until 2009 and v6 from 2010 onwards(Schneider et al., 2014) and includes many more stations than CRU. Both variants apply undercatch corrections (Adam and Lettenmaier, 2003) and use the ERA-Interim reanalyses (Dee et al., 2011). Variables other than precipitation are bias-corrected to CRU TS 3.1 until 2009 and CRU TS 3.21 from 2010 onwards in both versions of WFDEI. The reanalysis was spatially interpolated from its native N128 resolution (around 0.7° at the equator) to 0.5° resolution, including adjustment for differences in grid box elevation. Rainfall and snowfall were diagnosed separately in the reanalysis, with the fraction adjusted appropriately where differences in grid box elevation resulted in inappropriate precipitation phase.

All of these data sets have been widely used to drive land surface models in previous studies. All of them scale precipitation using data from CRU or GPCC, while an important difference is that only two (WFDEI-CRU and WFDEI-GPCC) include

an undercatch correction. The undercatch correction implemented in Adam and Lettenmaier (2003) resulted in an increase in global precipitation of 12%, with an increase of 95% in some boreal regions. Thus this can be an important source of underestimation of precipitation. Although the driving data sets are primarily used by land surface modellers, they include information from more widely-used precipitation data sets, and their basin averaged annual precipitation values are similar to

5 those from a range of other data sets (Fig. S3).

2.2 GRACE

The twin GRACE satellites made detailed measurements of changes in Earth's gravity field from April 2002 – January 2017. GRACE-Tellus used these to provide estimates of the TWS anomaly relative to the baseline mean over the years 2004–2009 (Swenson and Wahr, 2006; Swenson, 2012; Landerer and Swenson, 2012). GRACE may be used to study the dynamics of

10 the water cycle on a monthly time step. Although it may not be used to estimate absolute stores, it gives an estimate of the integrated water fluxes into and out of a region.

Three methods have been developed to solve the gravity fields using spherical harmonics (CSR, GFZ, JPL). These are then used to calculate the gridded TWS, so that three products are available. As is recommended by Sakumura et al. (2014), this study uses the ensemble average of the three solutions, which effectively reduces the uncertainty in the GRACE product.

15 The gravimetry measured by the GRACE satellites is converted to a TWS anomaly, making use of model output. Corrections for changes in atmospheric pressure are based on ECMWF analysis data, and corrections for post-glacial rebound are also modelled (Wahr et al., 1998). Destriping, Gaussian and degree 60 filters are applied to the gridded data. In order to correct for signal attenuation introduced by the filters applied when calculating the grids from the spherical harmonics, the three TWS land grids must be multiplied by a gridded gain factor, which is derived by applying the same filters to the output from the

20 Community Land Model 4 (CLM4; Landerer and Swenson, 2012).

Although the filters remove a significant component of spatially correlated and random errors, the gridded GRACE data are not independent of their neighbours. Uncertainties in the GRACE TWS data are generally highest at low latitudes, and lower towards the poles. In the boreal regions in this study, estimated measurement error is around 15 mm (Landerer and Swenson, 2012). The application of the gridded gain factor means that the gridded GRACE land data cannot be used to study long-term

25 trends (Landerer and Swenson, 2012).

The data used in this study were the GRACE RL05 gridded land TWS spherical harmonic solutions (Swenson, 2012; NASA, 2017). The data are available for the period April 2002 to January 2017. This study uses the fourteen years from September 2002 to August 2016, so that analyses are carried out over entire cold seasons.

2.3 GRDC river discharge

30 Monthly river discharge data were take from the Global Runoff Data Centre Reference Dataset (GRDC, 2014). For each basin in the current study, the station closest to the river mouth (as defined in the routing grid, see Sect. 3) is used to define basin discharge. In this study, basin discharge is quoted in units of mm, for consistency with the other variables. This is calculated by dividing the total observed discharge by the area of the 1° TRIP river network that drains to the relevant gauging station (Oki

and Sud, 1998). There are missing observations at each station, and some stations ceased contributing data before the end of the study period. Months with missing data were filled using a mean-monthly climatology, calculated using all available data at the station. Data availability varies between basins, with observations ending between 2003 and 2012. A summary of the available and filled data is shown in Table S1.

2.4 GLEAM v3.1a Evapotranspiration

The Global Land Evaporation Amsterdam Model v3.1a (GLEAM) provides estimates of evapotranspiration on a 0.25° resolution global grid, for the years 1980–2016. This is a modelled product based on reanalysis and observational data (Miralles et al., 2011; Martens et al., 2017). The algorithm uses the Priestley-Taylor equation to calculate actual evapotranspiration using net surface radiation and near-surface air temperature. It calculates interception using the Gash model and uses an adaptation of the Priestley-Taylor equation to calculate evaporation from open water and sublimation from ice and snow. Inputs include air temperature and surface radiation from reanalysis and a gauge-based gridded precipitation product, along with satellite products for SWE, vegetation properties and lightning and soil moisture. There is a version, v3.1b, which uses satellite products for air temperature and surface radiation, but this is only available for latitudes between 50° S and 50° N and was not suitable for this study.

Although it is not a direct observation, GLEAM is a well-established estimate of evapotranspiration, derived from observations as well as reanalysis products. It has been validated against soil moisture and flux measurements globally, with the GLEAM v3.1a having an unbiased root mean square difference of 0.7 mm d^{-1} compared with in situ observations (Martens et al., 2017). While evapotranspiration is low during the cold season, it is still important to use a robust estimate in order to understand the water balance. Evapotranspiration is difficult to observe and any gridded product is by necessity derived from other observed (or modelled) quantities (Mu et al., 2007; Mueller et al., 2013).

2.5 Snow water equivalent

2.5.1 GlobSnow

The European Space Agency (ESA) Data User Element (DUE) funded GlobSnow project (Luojus et al., 2013; Takala et al., 2011) provides SWE data for the years 1979–2016. This study uses GlobSnow v2.0. It is derived from a combination of satellite microwave radiometer data (SMMR, SSM/I and SSMIS sensors) and ECMWF weather station observations. A semi-empirical snow emission model is used to convert passive microwave emissions to SWE. A data assimilation scheme is then applied to the weather station and radiometer-derived SWE observations, to produce a map of the Northern Hemisphere in Equal-Area Scalable Earth Grid (EASE-Grid) format, at nominal 25 km resolution. Pixels are masked in areas of open water, and in areas where there are mountains.

For this study, the Northern Hemisphere data were regridded onto the analysis grid. Since the original data are masked where mountains are present, in the regridded data pixels were masked where at least 50% of the contributing pixels were mountain. This mostly affects western North America and some regions in Central Asia, masking 16% of the area of the Yenisei, 13% of

the Ob, 9% of the Lena and 14% of the Mackenzie. In the Eurasian basins the masked regions are generally in the south where the SWE is low, although in the Mackenzie the mountains extend to the north of the basin. The GlobSnow mountain mask is used for all basin averages of SWE in this study, both modelled and observed, so that comparisons are made over consistent regions.

2.5.2 CMC Daily Snow Depth Analysis

The Canadian Meteorological Centre Daily Snow Depth Analysis Data (CMC; Brown and Brasnett, 2010) are an assimilation of a simple snow accumulation and melt model, using temperatures and precipitation from the CMC Global Environmental Multiscale (GEM) forecast model, and daily snow depth data from the World Meteorological Organization (WMO) information system. From snow depth the SWE is estimated using snow density which is dependent on month and climate class. The data are provided as a Northern Hemisphere polar stereographic grid, with a nominal resolution of 24 km and are available for the years 1998–2012.

2.5.3 Comparison of SWE products

Figure 1 (a) and (b) show the mean seasonal maximum SWE for GlobSnow and CMC over the common overlap period (2002–2010), with the season defined to be 1st September – 31st August. Seasonal maximum SWE is important for the hydrology of Arctic river basins, as it is closely related to the amount of water that is released by the spring melt, which dominates the annual cycle of basin discharge in these regions. Many aspects of the large scale distribution are similar in both products and the main maxima are found in similar locations (when not covered by the GlobSnow mountain mask).

Basin average SWE was calculated by averaging over each basin, after having applied the GlobSnow mountain mask. The seasonality of basin average monthly maximum SWE, seen in Fig. 1 (c), is slightly different between GlobSnow and CMC, with GlobSnow tending to have a faster accumulation of snow in the earlier part of the winter, while CMC has a more constant accumulation throughout the winter. However, the seasonal maxima are consistent between the two data sets, although the GlobSnow maximum is slightly earlier in the Lena.

This comparison suggests that the GlobSnow and CMC basin average estimates are remarkably similar for the areas and time periods considered, particularly for estimates of seasonal maxima. A more comprehensive comparison of hemisphere-wide estimates by Mudryk et al. (2015) showed that GlobSnow tended to give larger and earlier seasonal maxima than CMC over the whole Northern Hemisphere, but this does include regions that are not being considered in the current study.

3 The JULES land surface model

This study uses the Joint UK Land Environment Simulator vn4.9 (JULES; Best et al., 2011; Clark et al., 2011) which is the land surface component of the Hadley Centre climate model but is here used in offline mode driven by near-surface meteorology. We provide brief details here and refer the reader to the above references for a fuller description.

JULES calculates the exchanges of radiation, heat, water and carbon between the land surface and the atmosphere. As employed here each grid box contains a mixture of nine surface types (five vegetation types and four non-vegetated surfaces), for each of which the surface fluxes are calculated independently. Subsurface fluxes of moisture and heat are calculated using a layered soil model, with a single soil column in each grid box. Supersaturation of soil moisture is avoided by moving the extra water to lower soil layers. The snow pack is modelled using a multi-layer snow scheme that employs a variable number of layers depending on the depth of snow. Each snow layer has a prognostic temperature, density, grain size, and solid and liquid water contents. Snowfall in vegetated areas with needle-leaf tree cover can be partitioned between the vegetation canopy and the underlying ground (Essery et al., 2003; Essery and Clark, 2003). The intercepted snow leaves the canopy though sublimation and wind-speed-dependent unloading of melting snow. The surface albedo is affected by the evolving grain size of the snow surface and the extent to which the vegetation is buried by snow. In this study we employ a TOPMODEL-based parameterisation of surface runoff (Beven and Kirkby, 1979). River routing is carried out as a post-processing step, using Total Runoff Integrating Pathways on a 1° river routing grid (TRIP; Oki and Sud, 1998).

The ability of JULES to represent snow and other high-latitude phenomena has been evaluated in several studies including those of Burke et al. (2013), Hancock et al. (2014), Chadburn et al. (2015), Ekici et al. (2015) and Ménard et al. (2015). In particular in SnowMIP2 JULES was found to be one of the best models for simulations of open sites, albeit it was thus not as skilful at forest sites (Rutter et al., 2009).

For this study, JULES is run at the native resolution of each driving data set (0.5°, except when using the PGF data which is 1.0° resolution). Meteorological data are provided every 6 hours (CRUNCEP) or 3 hours (all other data) and JULES interpolates these onto a 30 minute time step. The precipitation rate is kept constant over the data time step (3 or 6 hours). The model output is regridded to 1.0° resolution for analysis.

The WFDEI-CRU and WFDEI-GPCC data sets contain separate rainfall and snowfall fields that can be used directly by JULES, but the CRUNCEP and PGF data sets only provide total precipitation. For these, JULES assumes that the precipitation is snowfall when the near-surface air temperature is less than or equal to 274 K, while at higher temperatures the precipitation is assumed to be entirely rainfall. Hancock et al. (2014) suggest that, as long as the start and end dates of accumulation are correct, the modelled SWE is relatively insensitive to the choice of threshold.

The distribution of soil properties and land types was based on data from the IGBP Global Soil Task Force (Global Soil Data Task, 2000), gridded to the resolution of each driving data set. Vertical fluxes of soil water follow Darcy's law, using hydraulic characteristics calculated after Brooks and Corey (1964). Runoff was generated using the TOPMODEL formulation, with topographic index regridded from the high-resolution HydroSHEDS data (Marthews et al., 2015). The model was spun up using the first five years of each data set four times, for a total of 20 years, then run for several decades starting in 1970 (CRUNCEP and PGF) or 1979 (both WFDEI variants) and running to the end of the available data.

3.1 Experiments

Three experiments using JULES were performed:

CTL Control runs, driven with the original meteorological data sets.

GRC Runs in which the total cold season precipitation is scaled to match that derived from GRACE, as described in Sect. 4.1 below. The scale factors, which vary by basin and by run, can be seen in Table 2. These factors are applied constantly to both rain and snow for the months October to February inclusive. In the rest of the year the precipitation is unchanged.

5 **UCC** Runs for which wind-based undercatch correction (Adam and Lettenmaier, 2003) is applied to the precipitation (CRUN-CEP and PGF only). This uses the same catch ratios as were used in the creation of WFDEI (Weedon et al., 2011), one for each month for rainfall and snowfall separately. The correction is applied throughout the whole year. The ratios are available at $0.5°$ resolution, so were re-gridded to $1°$ resolution for PGF.

The full set up for each run is available as a version-controlled Rose suite, through the Met Office repository: https://code.
10 metoffice.gov.uk/trac/roses-u. Full details of the suite number and revision for each run is given in Table S2.

3.2 Mean bias error

To evaluate the runs, we calculate the mean bias error (MBE) of the annual maximum SWE in each basin (M_b) to be

$$M_b = \frac{1}{n_y} \sum_{i=1}^{n_y} S_{i,b} - s_{i,b},\tag{1}$$

where $S_{i,b}$ is the modelled SWE in basin b at timestep i, $s_{i,b}$ is the corresponding observed SWE and n_y is the number of years
15 of overlap between model and observations. To obtain the average over basins, M_S, the area-weighted mean of the basins is calculated,

$$M_S = \frac{\sum_{b=1}^{n_b} A_b M_b}{\sum_{b=1}^{n_b} A_b}\tag{2}$$

where A_b is the total area of basin b.

The variance of the bias error in each basin is

20 $$\sigma_b^2 = \frac{1}{n_y - 1} \sum_{i=1}^{ny} (S_{i,b} - s_{i,b})^2,\tag{3}$$

and the combined variance is

$$\sigma_S^2 = \frac{(n-1) \sum_{b=1}^{n_b} A_b \sigma_b^2}{(\sum_{b=1}^{n_b} A_b) - 1}.\tag{4}$$

4 Precipitation corrections

4.1 GRACE-derived precipitation estimates

25 A water balance approach (Swenson, 2010) was used to estimate the monthly precipitation (P; mm month^{-1}) at a basin scale:

$$P = \frac{dS_{\text{tot}}}{dt} + E + Q_{\text{net}}\tag{5}$$

where dS_{tot}/dt (mm month^{-1}) is the change in TWS (S_{tot}; mm), E (mm month^{-1}) is the total evaporation flux, including sublimation, and Q_{net} (mm month^{-1}) is net runoff. At the annual or monthly scale, this can be calculated to be

$$\int_{t_1}^{t_2} P(t)dt = S_{\text{tot}}(t_2) - S_{\text{tot}}(t_1) + \int_{t_1}^{t_2} (E(t) + Q_{\text{net}}) \, dt, \tag{6}$$

where t_1 and t_2 are the start and end of each accumulation period respectively. TWS encompasses all water storage, including soil moisture, ground water, water in wetlands, lakes and rivers, and water stored as snow in the snow pack. Evaporation includes transpiration, evaporation from soil surfaces, evaporation from intercepted water and other open water (rivers and lakes), as well as sublimation from frozen surfaces. As the water balance is calculated over whole basins there is no incoming runoff and the net runoff is equal to the basin discharge.

The change in TWS is calculated by differencing the GRACE anomalies between months, and averaging over the basin. Monthly evaporation is provided by GLEAM and averaged over the basin, while the basin discharge is obtained from the GRDC measurement station closest to the basin outflow.

The water balance is calculated for each month in the cold season (October to February inclusive). This is defined based on two metrics: a) the change in GRACE TWS must be positive and b) the evaporation must be small. There is inter-annual variability in the length of this season, so the period October to February was chosen as a conservative accumulation period that would apply over all years and basins. During this season snowfall is the dominant precipitation type, the accumulating snow-pack is the dominant change to TWS, and the runoff and evaporation fluxes are relatively small (Fig. S4). This minimises the effect of uncertainty in evaporation and basin discharge products, which could be more significant outside of the cold season (Swenson, 2010; Seo et al., 2010).

4.2 Comparison of precipitation estimates

Figure 2 shows the long-term mean monthly accumulated total precipitation during the cold season (October–February) for each of the four driving data sets and for the GRACE-derived precipitation estimates (all averaged over the common overlap period). The differences between data sets vary between the catchments, but in general the driving data sets accumulate less precipitation than the GRACE estimate. The CRUNCEP and PGF data sets have lower precipitation than the WFDEI data through the whole cold season in all basins, but are very similar to each other as they were both bias corrected to CRU data without undercatch correction. The WFDEI data sets generally have higher precipitation, and are much closer to the GRACE-derived estimate, even indicating a larger accumulation by December in the Lena and the Ob. The final accumulation is similar to the GRACE-derived estimate in three basins for WFDEI-GPCC and two basins for WFDEI-CRU . Overall the implication is that the driving data sets often do not contain enough precipitation to close the water budget in these basins in the cold season, particularly for the CRUNCEP and PGF data.

4.3 Calculation of precipitation scale factors

In order to appropriately increase the amount of precipitation in the driving data sets, a scale factor was calculated for each data set in each study basin. The scale factor between GRACE and each driving data set was found by calculating the ratio of cold season accumulated precipitation derived from GRACE to that in the driving data. As the driving data cover a longer time period than GRACE, to ensure that the difference is not due to inter-annual variability or long-term climate trends each ratio was calculated using only data from the overlapping period of the relevant data set with GRACE. This period starts in 2002 for all of the data sets, and runs to the end of each data set. The scale factors for each basin can be seen in Table 2.

The relative scaling between basins is consistent for each data set, with the Yenisei requiring the largest correction (between 24% and 55% increase), and the Ob requiring the least (between a 2% decrease and 37% increase). There is a striking difference between the WFDEI data sets, which require increases up to 27% (and one decrease), and the PGF and CRUNCEP data sets which require much larger increases of between 35% and 55%.

The GRC runs of JULES were carried out by scaling both snowfall and rainfall by these factors during the cold season only for all years of the run (not just the GRACE period). Snow and rain that falls outside of this season were unchanged. This design ensured that each run received the same amount of cold season precipitation as indicated by GRACE, when averaged over the basin and over the period of overlap between the driving data set and GRACE; the temporal and spatial variability of precipitation still varied between runs.

5 JULES model runs

5.1 Control runs (CTL)

The simulated monthly maximum SWE for each basin can be seen in Fig. 3. In many cases JULES accumulates snow more slowly than the observational estimates, particularly later in the accumulation season, and the SWE is less sharply peaked than the observations. This is possibly due to particular biases in the driving data (perhaps missing smaller events towards the end of the accumulation season) or in how the model responds (perhaps starting to melt too early).

Figure 4 shows the MBE for seasonal maximum SWE averaged over all four basins with respect to both GlobSnow and CMC. In general the magnitude of the MBE is largest for PGF, and smallest for WFDEI-GPCC (Fig. S5 shows the MBE for each basin). The largest SWE deficit is found when using the PGF data in the Lena basin, for which the bias is -57% compared to GlobSnow and -56% compared to CMC. The smallest deficit is found in the Mackenzie using the WFDEI-GPCC driving data, which has a MBE of -17% compared with both GlobSnow and CMC. Maps of the modelled SWE show broadly similar distributions of snow to the observations (Fig. S6). The difference between the modelled and observed SWE (eg. GlobSnow in Fig. S7) show that, although there are a few regions of higher SWE in JULES outside of the studied basins, there is a deficit of modelled lying snow across most of the Northern Hemisphere. The general tendency for JULES to underestimate SWE is consistent with the indication of insufficient cold season precipitation (Fig. 2), and both analyses also agree that the WFDEI

runs are generally closer to the observations. However it is clear that, even in the cases where the precipitation is close to the GRACE-derived estimate, the modelled maximum SWE is still lower than observed.

The annual discharge in JULES is also severely underestimated compared with GRDC observations, as seen in Fig. 5 which shows the basin discharge averaged over all four basins. All of the runs demonstrate a substantial dry bias in comparison with GRDC values. CRUNCEP is particularly poor, with an 80–90% deficit in annual mean basin discharge, while the other runs underestimate basin discharge by between 52–57% on average. This difference between CRUNCEP and the other data sets is consistent across all the basins. The deficit is particularly pronounced during the spring peak (Fig. S9) which in nature is strongly driven by snow melt. In JULES, the peak in the modelled river flow is broader and occurs later than the observations, which may indicate deficiencies in the modelled hydrology as well as a lack of melting snow. However, at the annual scale the bias may also be affected by warm-season rainfall.

5.2 Runs with scaled precipitation (GRC)

Using the GRACE-scaled precipitation to drive JULES gives improved SWE representation overall, with significant increases in monthly and seasonal maximum SWE (Fig. 6). The modelled seasonal cycle is much more similar to observations, particularly in the Yenisei and Mackenzie, and early in the season for the Lena and Ob. In the Ob, the CRUNCEP GRC run is much more similar to the observations than any of the other models, while in the Mackenzie it now overestimates SWE. In the Lena, the seasonal peak is still not attained with any data set. Overall the largest increases in SWE are found for the CRUNCEP and PGF runs, consistent with the larger scale factors used (Table 2). The annual maximum SWE is increased by 30% averaged across all basins and runs, with the CRUNCEP and PGF runs having an average increase of 48% and the WFDEI runs having a smaller average increase of 12%. Overall, the magnitude of the MBE is decreased for all driving data sets (Fig. 4). Although there is an increase in SWE over the extent of each basin (Fig. S8), the effect on MBE varies between driving data sets and basins: the widespread underestimation of the CTL runs is replaced by a more nuanced picture, including some cases with modest overestimation of SWE (Fig. S5). The modelled estimates of maximum SWE in the Yenisei and Mackenzie are now clustered around the observational estimates, while a low bias is still found in most runs for the Lena and Ob. The biggest improvement is seen in CRUNCEP runs, which now only has an MBE of -5% overall.

Basin discharge is higher in the scaled runs than the control runs, although the increase is relatively small (27% overall) compared to the general dry bias (Fig. 5). The improvement in representation of basin discharge varies significantly between data sets, from 8% in WFDEI-GPCC to 57% in CRUNCEP, which still has the lowest basin discharge (the highest absolute increase is 27 mm/yr in PGF, while the lowest is 7 mm/yr in WFDEI-GPCC). Although the basin discharge increases overall, the increase in SWE and the resultant increase in spring snow melt do not change the timing of the peak basin discharge in the models. The modelled estimates of basin discharge are still considerably lower than the GRDC estimates, with CRUNCEP underestimating by 74% and the others between 39–52%. This implies that factors other than cold season precipitation and snow accumulation are contributing to the lack of river flow. In particular the low bias and late peak in discharge for the Yenisei and Mackenzie persists (Fig. S9) despite the peak SWE now being close to the observational estimates.

5.3 Runs with undercatch correction (UCC)

As discussed above, a possible explanation for underestimation of cold season precipitation in observation-based data sets is undercatch. In an attempt to account for this effect two of the data sets — WFDEI-CRU and WFDEI-GPCC — already implement undercatch correction. The two data sets that do not apply a correction required the largest scaling factors to match the GRACE-derived estimates. To investigate the importance of the undercatch correction in comparison to the GRACE corrections, the catch ratios calculated by Adam and Lettenmaier (2003) were used to correct the CRUNCEP and PGF data on a monthly basis. Both rainfall and snowfall were adjusted, using separate catch ratios, for all months of the year, and used to force JULES in the UCC runs.

The undercatch correction increases the precipitation across the whole year, but here we focus on the cold season precipitation increase. The increase is large, but less than indicated by the GRACE precipitation scale factors. In the Ob, undercatch correction increases the amount of cold season precipitation by 26% over the whole basin, compared to an overall increase of 35% in the GRC runs for CRUNCEP and PGF. However, in the other basins the increase in precipitation due to undercatch correction is between 20–25% and the increase based on GRACE precipitation is between 45–55%. Thus, the undercatch correction accounts for between 30–75% of the cold season deficit that is indicated by the comparison with the GRACE-derived precipitation estimates.

The improvement to the estimate of SWE is clear, although it is not as large as in the GRC runs (Fig. 4). On average the two undercatch corrected driving data sets used for the UCC runs have maximum SWE increased by 34%, compared to an increase of 48% when the same two data sets were scaled by the GRACE scale factors. In contrast, the increases in annual basin discharge were much more similar between the GRC and UCC runs. In the Ob, the undercatch correction even resulted in higher basin discharge than the GRACE scaling, despite the GRACE scaling resulting in higher total annual precipitation; this is because in the UCC runs the precipitation is increased across the whole year, contributing to increased river flow in warm months as well as the spring snow melt, whereas the GRC runs only have increased precipitation in the cold season.

Along with the fact that GRACE indicates that the WFDEI runs also under-represent precipitation in at least some of these basins, this suggests that although undercatch correction can account for some of the deficit in precipitation, it cannot account for all of it.

6 Discussion

This study has shown that gauge- or reanalysis-based estimates of cold season precipitation in boreal basins can be significantly lower than is suggested by a water-balance approach using GRACE TWS data. By scaling the driving meteorological data sets, we can significantly improve the representation of lying snow in a land surface model. The deficits seen in precipitation are large but similar in magnitude to those calculated by Finney et al. (2012) who calculated precipitation corrections of between 36 and 70%, based on a comparison of modelled and observed SWE, and by Behrangi et al. (2017) who used GRACE to evaluate precipitation data sets in endorrheic cold mountainous basins, and found that some precipitation data sets only captured between

10 and 60 % of the GRACE-based precipitation estimates. In this study we used time-invariant corrections, but this could be extended to allow time-varying corrections, say on an annual basis.

The GRACE-based estimates provide a means to account for measurement errors in the gauge data that are used to bias-correct reanalyses. This study suggests that undercatch is an important source of error, but in many cases the application of

5 an undercatch correction does not remove the bias entirely. Further sources of error in the gauge-based estimates include spatial variability that is missed by the gauge network. While the GRACE TWS method used here circumvents many of the limitations of the gauge-based correction, it introduces uncertainties from other terms in the water budget (e.g. evaporation). It is also inherently large scale, so improves the basin-averaged SWE, but may not result in improved SWE when looking at the local scale.

10 Undercatch correction, on the other hand, can be calculated at a local or grid box scale and can, in theory at least, more easily take account of changing meteorological conditions down to the time scale of individual storms. Furthermore it is easy to implement and is attractive as it addresses known deficiencies in the observations. It can also be used to correct historical data sets that pre-date GRACE. However the correction is uncertain, including dependency on the type of gauge used in each area, and again requires further inputs, each with uncertainties. A further approach, not studied here, is to use estimates of

15 SWE, either from ground measurements or remote sensing, to estimate snowfall.

This study of cold season processes shows that the undercatch correction is equivalent to a substantial fraction of the GRACE-derived correction, suggesting that for gauge-based data sets the undercatch correction can be considered a minimum requirement that should be applied whenever possible. However, this varies by region, implying that there are different reasons for the deficit in precipitation in different regions. The CRUNCEP and PGF data sets do not include an undercatch

20 correction but have been widely used with land surface models; our results suggest that any aspects of those studies that are potentially sensitive to snowfall, such as high-latitude hydrological analyses, should be regarded as particularly uncertain. Undercatch is a known issue for rainfall as well as snowfall and is considered in undercatch correction algorithms Adam and Lettenmaier (2003); Adam et al. (2006). The annual water balance calculated from GRDC, GLEAM and GRACE suggests that there is insufficient precipitation input in each of these basins to support the observed river flow (Fig. S10). Therefore

25 undercatch errors potentially also affect summertime precipitation.

Figure 8 shows that the increase in seasonal maximum SWE is proportional to the increase in seasonal snowfall between experiments, with a very strong correlation ($r^2 = 0.98$) between the two. The gradient of 0.76 implies that not all of the increase in snowfall manifests as an increase in SWE, which is mainly due to increased sublimation and a small increase in snow melt, but it does confirm that correcting snowfall is a direct and reliable approach with which to target errors in simulated SWE. This

30 correlation also implies that the approach of Finney et al. (2012) to increase driving snowfall based on the required increase in SWE is reasonable.

The use of corrected precipitation improved results in most catchments and with most driving data sets, often substantially. In contrast the larger dry bias in modelled basin discharge was found in all simulations, with only modest improvements from corrected precipitation. This difference is a result of the much closer links between point snowfall and SWE, whereas modelled

35 discharge can be viewed as the net effect of many hydrological processes acting across the catchment. A striking illustration

of this is the large difference between the CRUNCEP and PGF river discharge, despite very similar precipitation inputs (both having been bias-corrected to CRU precipitation). The larger river flow in PGF is balanced by a smaller evapotranspiration flux (Fig. S11), while evapotranspiration in CRUNCEP tends to be closer to the WFDEI-based estimates. In turn this can be related to differences in other aspects of the meteorological forcing (Figs. S1 and S2); in particular, PGF has much higher specific

5   humidity, particularly in the summer, than the other data sets. In spring and summer PGF humidity is 18% – 87% higher than in CRUNCEP, which suppresses evapotranspiration, which is 10% – 17% lower in PGF for the CTL runs. (This can be contrasted with the air temperature, which is nearly identical in PGF and CRUNCEP, implying that the data sets employed different methods to reconstruct humidity.) This is consistent with the fact that evapotranspiration in the PGF runs is generally less than estimated by GLEAM, while the other runs tend to be closer to GLEAM, although there is considerable uncertainty

10   in the GLEAM product. Although the focus of this study is on correction of precipitation products, which is shown to improve modelled SWE, it is clear that other meteorological variables are also important, particularly for other aspects of modelled hydrology.

The modelled runoff ratio (Q_{net}/P) is much lower on average for CRUNCEP than for runs with the other data sets. When combined with the already relatively low precipitation in the CTL run, this means the CRUNCEP data set generates particularly

15   small values of discharge in all basins — overall the CRUNCEP basin discharge is 16% of observed in the CTL run and 25% of observed in the GRC and UCC runs. This is related to the longer time step of the data set (6 hours rather than 3 hours), within which the precipitation rate is constant in the JULES runs; the lower intensity of precipitation tends result in larger interception loss and higher interception loss reduces water available for runoff. For CRUNCEP, the interception is the largest component of evapotranspiration (interception is 35% of total evapotranspiration in the CTL run, whereas transpiration is 28%), while

20   the other driving data sets have lower interception losses (22% – 24% for CTL), with transpiration as the largest component instead (35% – 38% of total evapotranspiration for CTL).

The extra precipitation input in the GRC runs largely appears as increased evaporation in CRUNCEP runs, whereas it is largely converted to runoff for the other data sets (Fig. S11). This means that, even with a significant increase in precipitation in the CRUNCEP runs, the basin discharge is still extremely underestimated.

25   The seasonal maximum SWE tends to be biased low even when precipitation inputs are corrected. Assuming that both the precipitation and SWE observations are now correct, this points to model structural uncertainties as a possible remaining source of bias. One possibility is that the modelled sublimation rate is too high. Sublimation in the CTL runs, averaged over all basins, ranged from 37% of snowfall (PGF) to 44% (CRUNCEP). It is very difficult to measure sublimation, particularly over large areas, and most estimates tend to be based on water balance methods or models (Liston and Sturm, 2004). Estimates at a range

30   of scales indicate that 10 – 50% of snowfall can be lost through sublimation, with substantial variation depending on land cover and meteorological conditions (e.g. Liston and Sturm, 2004; Brun et al., 2013; Casson et al., 2018). For the Mackenzie basin sublimation was estimated as 29 mm (7% of annual precipitation; Déry and Yau, 2002), while numerical modelling by Yang et al. (2010) suggested that approximately 24% of annual snowfall in the area $50 – 70°$ N was lost as sublimation. Thus the evidence suggests that sublimation might be rather high in JULES, albeit the evidence base is itself rather uncertain.

35   Further investigation, including ensembles of LSMs and GHMs, and point-scale runs in data-rich environments to examine

the simulated water budget across a range of land covers such as in SnowMIP2 (Rutter et al., 2009), will allow us to identify the role of model uncertainty and increase our confidence in the ability of JULES and other models to correctly simulate sublimation and other key processes.

However, errors in modelled sublimation and indeed all cold season processes are insufficient to fully account for the dry bias in annual discharge. It is also possible that rainfall outside of the cold season is underestimated in the driving data. In all of these basins there is more precipitation in summer (when it is more likely to be rain) than in winter, suggesting that correction of warm season rainfall could potentially have a larger impact on the annual water balance. A sensitivity test, applying the cold-season scaling factors to the whole year, suggests that this can double the modelled basin discharge, significantly decreasing the underestimation (Fig. S12). However, there are clear signs that meteorological inputs are not the only source of error, and that there are fundamental deficiencies in the model's representation of runoff generation processes: even a good estimate of peak SWE does not result in a good representation of the spring discharge peak (Fig. S9). It is likely that the parameterisation of infiltration into partly-frozen ground and related runoff generation processes are not well represented in JULES. Previous work has shown that alternative descriptions of frozen soil can improve the modelled runoff peak (Finney et al., 2012).

7 Conclusions

There is a substantial body of literature on the intercomparison of global precipitation data sets, with a lesser focus on the particular issues found at high latitudes where much of the precipitation falls as snow. There is an ongoing need to compare these precipitation products and to ensure that the best meteorological data are made available as inputs to land surface modelling. This study has focussed on precipitation but the model results clearly indicate that other variables, such as humidity, are also important.

Land surface modellers should continue to critically evaluate the meteorological data they use, and ideally run a model using a variety of data sets. The extent to which results are sensitive to the choice of meteorological data will vary; for some analyses there might be relatively little sensitivity, but by and large this can only be determined through the use of an ensemble of runs forced by alternative data. Meteorological data should also be evaluated in combination with products that describe related parts of the hydrological system, such as GRACE TWS, estimates of SWE based on remote sensing, and river discharge as used in this study. Although each product comes with its own uncertainty, and a range of alternative and potentially conflicting data sets is often available, the combination of estimates across different parts of the hydrological system can provide extra insights, particularly if a model shows consistent biases across several components.

Even with the scaled precipitation data sets, there are still remaining biases in the model output, which vary across basin and data set. Using scaled precipitation data sets would allow further identification of biases due to model uncertainty.

This study shows that at a basin scale the cold season precipitation in four data sets that are commonly used to drive land surface models is low compared with estimates derived from GRACE TWS. This leads to consistent and large errors in the SWE and basin discharge calculated by JULES, which are also low compared to observations. By increasing the precipitation in JULES to match the GRACE estimates the modelled SWE is substantially improved, although river discharge is still low —

likely because of a combination of poor modelling of runoff processes during the spring melt and possible underestimation of summertime precipitation. By providing methods to better define the precipitation inputs to the system, the current study paves the way for subsequent work on key hydrological processes.

Code availability. The model runs in this manuscript were carried out using JULES version 4.9, with modifications to allow scaling of the precipitation input. This is available through the JULES repository and is located at https://code.metoffice.gov.uk/trac/jules/browser/main/ branches/test/emmarobinson/vn4.9_arctic_scaling?rev=9936 (registration required). The runs were carried out using Rose (http://metomi. github.io/rose), the control files are available through the Met Office rose repository (https://code.metoffice.gov.uk/trac/roses-u), and details are in Table S2. All other analysis code, including river routing and post-processing, available on request.

Author contributions. DBC designed the study. ELR carried out model runs and analysis. Both wrote the manuscript.

Competing interests. The authors declare no competing interests.

Acknowledgements. We would like to thank the editor and the reviewers for their valuable feedback and comments. This work was supported by the Natural Environment Research Council award number NE/H000224/1 and award number NE/S017380/1 as part of the Hydro-JULES program. GRACE land data are available at http://grace.jpl.nasa.gov, supported by the NASA MEaSUREs Program.

References

Adam, J. C. and Lettenmaier, D. P.: Adjustment of global gridded precipitation for systematic bias, Journal of Geophysical Research: Atmospheres, 108, https://doi.org/10.1029/2002JD002499, 2003.

Adam, J. C., Clark, E. A., Lettenmaier, D. P., and Wood, E. F.: Correction of Global Precipitation Products for Orographic Effects, Journal of Climate, 19, 15–38, https://doi.org/10.1175/JCLI3604.1, 2006.

Alkama, R., Decharme, B., Douville, H., Becker, M., Cazenave, A., Sheffield, J., Voldoire, A., Tyteca, S., and Le Moigne, P.: Global Evaluation of the ISBA-TRIP Continental Hydrological System. Part I: Comparison to GRACE Terrestrial Water Storage Estimates and In Situ River Discharges, Journal of Hydrometeorology, 11, 583–600, https://doi.org/10.1175/2010JHM1211.1, 2010.

Beck, H. E., Vergopolan, N., Pan, M., Levizzani, V., van Dijk, A. I. J. M., Weedon, G. P., Brocca, L., Pappenberger, F., Huffman, G. J., and Wood, E. F.: Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling, Hydrology and Earth System Sciences, 21, 6201–6217, https://doi.org/10.5194/hess-21-6201-2017, 2017.

Behrangi, A., Christensen, M., Richardson, M., Lebsock, M., Stephens, G., Huffman, G. J., Bolvin, D., Adler, R. F., Gardner, A., Lambrigtsen, B., and Fetzer, E.: Status of high-latitude precipitation estimates from observations and reanalyses, Journal of Geophysical Research: Atmospheres, 121, 4468–4486, https://doi.org/10.1002/2015JD024546, 2016.

Behrangi, A., Gardner, A. S., Reager, J. T., and Fisher, J. B.: Using GRACE to constrain precipitation amount over cold mountainous basins, Geophysical Research Letters, 44, 219–227, https://doi.org/10.1002/2016GL071832, 2017.

Berezovskaya, S., Yang, D., and Kane, D. L.: Compatibility analysis of precipitation and runoff trends over the large Siberian watersheds, Geophysical Research Letters, 31, https://doi.org/10.1029/2004GL021277, 2004.

Best, M. J., Pryor, M., Clark, D. B., Rooney, G. G., Essery, R. L. H., Ménard, C. B., Edwards, J. M., Hendry, M. A., Porson, A., Gedney, N., Mercadp, L. M., Sitch, S., Blyth, E., Boucher, O., Cox, P. M., Grimmond, C. S. B., and Harding, R. J.: The Joint UK Land Environment Simulator (JULES), model description – Part 1: Energy and water fluxes, Geoscientific Model Development, 4, 677–699, https://doi.org/10.5194/gmd-4-677-2011, 2011.

Beven, K. J. and Kirkby, M. J.: A physically based, variable contributing area model of basin hydrology, Hydrological Sciences Bulletin, 24, 43–69, https://doi.org/10.1080/02626667909491834, 1979.

Biemans, H., Hutjes, R. W. A., Kabat, P., Strengers, B. J., Gerten, D., and Rost, S.: Effects of Precipitation Uncertainty on Discharge Calculations for Main River Basins, Journal of Hydrometeorology, 10, 1011–1025, https://doi.org/10.1175/2008JHM1067.1, 2009.

Bowling, L. C., Lettenmaier, D. P., Nijssen, B., Graham, L., Clark, D. B., Maayar, M. E., Essery, R., Goers, S., Gusev, Y. M., Habets, F., van den Hurk, B., Jin, J., Kahan, D., Lohmann, D., Ma, X., Mahanama, S., Mocko, D., Nasonova, O., Niu, G.-Y., Samuelsson, P., Shmakin, A. B., Takata, K., Verseghy, D., Viterbo, P., Xia, Y., Xue, Y., and Yang, Z.-L.: Simulation of high-latitude hydrological processes in the Torne-Kalix basin: PILPS Phase 2(e): 1: Experiment description and summary intercomparisons, Global and Planetary Change, 38, 1 – 30, https://doi.org/10.1016/S0921-8181(03)00003-1, project for Intercomparison of Land-surface Parameterization Schemes, Phase 2(e), 2003.

Brooks, R. H. and Corey, A. T.: Hydraulic properties of porous media, Hydrological Papers 3, Colorado State University, Fort Collins, 1964.

Brown, R. and Brasnett, B.: Canadian Meteorological Centre (CMC) Daily Snow Depth Analysis Data, Version 1, https://doi.org/10.5067/W9FOYWH0EQZ3, NASA National Snow and Ice Data Center Distributed Active Archive Center. Accessed 2012-07-02, 2010.

Brown, R. D. and Mote, P. W.: The Response of Northern Hemisphere Snow Cover to a Changing Climate, Journal of Climate, 22, 2124–2145, https://doi.org/10.1175/2008JCLI2665.1, 2009.

Brown, R. D. and Robinson, D. A.: Northern Hemisphere spring snow cover variability and change over 1922–2010 including an assessment of uncertainty, The Cryosphere, 5, 219–229, https://doi.org/10.5194/tc-5-219-2011, 2011.

5   Brun, E., Vionnet, V., Boone, A., Decharme, B., Peings, Y., Valette, R., Karbou, F., and Morin, S.: Simulation of Northern Eurasian Local Snow Depth, Mass, and Density Using a Detailed Snowpack Model and Meteorological Reanalyses, Journal of Hydrometeorology, 14, 203–219, https://doi.org/10.1175/JHM-D-12-012.1, 2013.

Brutel-Vuilmet, C., Ménégoz, M., and Krinner, G.: An analysis of present and future seasonal Northern Hemisphere land snow cover simulated by CMIP5 coupled climate models, The Cryosphere, 7, 67–80, https://doi.org/10.5194/tc-7-67-2013, 2013.

10   Bulygina, O. N., Groisman, P. Y., Razuvaev, V. N., and Korshunova, N. N.: Changes in snow cover characteristics over Northern Eurasia since 1966, Environmental Research Letters, 6, 045 204, 2011.

Burke, E. J., Dankers, R., Jones, C. D., and Wiltshire, A. J.: A retrospective analysis of pan Arctic permafrost using the JULES land surface model, Climate Dynamics, 41, 1025–1038, https://doi.org/10.1007/s00382-012-1648-x, 2013.

Casson, D. R., Werner, M., Weerts, A., and Solomatine, D.: Global re-analysis datasets to improve hydrological assessment and snow water

15   equivalent estimation in a sub-Arctic watershed, Hydrology and Earth System Sciences, 22, 4685–4697, https://doi.org/10.5194/hess-22-4685-2018, 2018.

Chadburn, S., Burke, E., Essery, R., Boike, J., Langer, M., Heikenfeld, M., Cox, P., and Friedlingstein, P.: An improved representation of physical permafrost dynamics in the JULES land-surface model, Geoscientific Model Development, 8, 1493–1508, https://doi.org/10.5194/gmd-8-1493-2015, 2015.

20   Clark, D. B., Mercado, L. M., Sitch, S., Jones, C. D., Gedney, N., Best, M. J., Pryor, M., Rooney, G. G., Essery, R. L. H., Blyth, E., Boucher, O., Harding, R. J., Huntingford, C., and Cox, P. M.: The Joint UK Land Environment Simulator (JULES), model description – Part 2: Carbon fluxes and vegetation dynamics, Geoscientific Model Development, 4, 701–722, https://doi.org/10.5194/gmd-4-701-2011, 2011.

Cohen, J. and Rind, D.: The Effect of Snow Cover on the Climate, Journal of Climate, 4, 689–706, https://doi.org/10.1175/1520-0442(1991)004<0689:TEOSCO>2.0.CO;2, 1991.

25   Decharme, B. and Douville, H.: Uncertainties in the GSWP-2 precipitation forcing and their impacts on regional and global hydrological simulations, Climate Dynamics, 27, 695–713, https://doi.org/10.1007/s00382-006-0160-6, 2006.

Dee, D. P., Uppala, 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, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz,

30   B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Quarterly Journal of the Royal Meteorological Society, 137, 553–597, https://doi.org/10.1002/qj.828, 2011.

Déry, S. J. and Yau, M. K.: Large-scale mass balance effects of blowing snow and surface sublimation, Journal of Geophysical Research: Atmospheres, 107, ACL 8–1–ACL 8–17, https://doi.org/10.1029/2001JD001251, 2002.

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

Essery, R. and Clark, D. B.: Developments in the MOSES 2 land-surface model for PILPS 2e, Global and Planetary Change, 38, 161 – 164, https://doi.org/https://doi.org/10.1016/S0921-8181(03)00026-2, project for Intercomparison of Land-surface Parameterization Schemes, Phase 2(e), 2003.

Essery, R., Pomeroy, J., Parviainen, J., and Storck, P.: Sublimation of Snow from Coniferous Forests in a Climate Model, Journal of Climate, 16, 1855–1864, https://doi.org/10.1175/1520-0442(2003)016<1855:SOSFCF>2.0.CO;2, 2003.

Essery, R., Rutter, N., Pomeroy, J., Baxter, R., Stähli, M., Gustafsson, D., Barr, A., Bartlett, P., and Elder, K.: SNOWMIP2: An Evaluation of Forest Snow Process Simulations, Bulletin of the American Meteorological Society, 90, 1120–1136, https://doi.org/10.1175/2009BAMS2629.1, 2009.

Fernandes, R., Zhao, H., Wang, X., Key, J., Qu, X., and Hall, A.: Controls on Northern Hemisphere snow albedo feedback quantified using satellite Earth observations, Geophysical Research Letters, 36, https://doi.org/10.1029/2009GL040057, 2009.

Fiedler, K. and Döll, P.: Global modelling of continental water storage changes — sensitivity to different climate data sets, Advances in Geosciences, 11, 63–68, https://doi.org/10.5194/adgeo-11-63-2007, 2007.

Finney, D. L., Blyth, E., and Ellis, R.: Improved modelling of Siberian river flow through the use of an alternative frozen soil hydrology scheme in a land surface model, The Cryosphere, 6, 859–870, https://doi.org/10.5194/tc-6-859-2012, 2012.

Global Soil Data Task: Global Soil Data Products CD-ROM (IGBP-DIS). CD-ROM., http://www.daac.ornl.gov, International Geosphere-Biosphere Programme, Data and Information System, Potsdam, Germany. Available from Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, U.S.A., 2000.

Grabs, W. E., Portmann, F., and de Couet, T.: Discharge Observation Networks in Arctic Regions: Computation of the River Runoff into the Arctic Ocean, Its Seasonality and Variability, pp. 249–267, Springer Netherlands, Dordrecht, https://doi.org/10.1007/978-94-011-4132-1_11, 2000.

GRDC: The Global Runoff Data Centre, https://www.bafg.de/GRDC, 56068, Koblenz, Germany. Accessed: 2014-09-01, 2014.

Haddeland, I., Clark, D. B., Franssen, W., Ludwig, F., Voß, F., Arnell, N. W., Bertrand, N., Best, M., Folwell, S., Gerten, D., Gomes, S., Gosling, S. N., Hagemann, S., Hanasaki, N., Harding, R., Heinke, J., Kabat, P., Koirala, S., Oki, T., Polcher, J., Stacke, T., Viterbo, P., Weedon, G. P., and Yeh, P.: Multimodel Estimate of the Global Terrestrial Water Balance: Setup and First Results, Journal of Hydrometeorology, 12, 869–884, https://doi.org/10.1175/2011jhm1324.1, 2011.

Hancock, S., Huntley, B., Ellis, R., and Baxter, R.: Biases in Reanalysis Snowfall Found by Comparing the JULES Land Surface Model to GlobSnow, Journal of Climate, 27, 624–632, https://doi.org/10.1175/JCLI-D-13-00382.1, 2014.

Harris, I., Jones, P., Osborn, T., and Lister, D.: Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 Dataset, International Journal of Climatology, 34, 623–642, https://doi.org/10.1002/joc.3711, 2013.

Islam, S. U. and Déry, S. J.: Evaluating uncertainties in modelling the snow hydrology of the Fraser River Basin, British Columbia, Canada, Hydrology and Earth System Sciences, 21, 1827–1847, https://doi.org/10.5194/hess-21-1827-2017, 2017.

Jones, P. W.: SCRIP: A Spherical Coordinate Remapping and Interpolation Package v1.4, https://github.com/SCRIP-Project/SCRIP, 1998.

Jones, P. W.: First- and Second-Order Conservative Remapping Schemes for Grids in Spherical Coordinates, Monthly Weather Review, 127, 2204–2210, https://doi.org/10.1175/1520-0493(1999)127<2204:FASOCR>2.0.CO;2, 1999.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R., and Joseph, D.: The NCEP/NCAR 40-Year Reanalysis Project, Bulletin of the American Meteorological Society, 77, 437–472, https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2, 1996.

Landerer, F. W. and Swenson, S. C.: Accuracy of scaled GRACE terrestrial water storage estimates, Water Resources Research, 48, https://doi.org/10.1029/2011wr011453, 2012.

Liston, G. and Sturm, M.: The role of winter sublimation in the Arctic moisture budget, Nordic Hydrology, 35, 325–334, https://doi.org/10.2166/nh.2004.0024, 2004.

5   Liston, G. E. and Hiemstra, C. A.: The Changing Cryosphere: Pan-Arctic Snow Trends (1979–2009), Journal of Climate, 24, 5691–5712, https://doi.org/10.1175/JCLI-D-11-00081.1, 2011.

Luojus, K., Pulliainen, J., and the GlobSnow Consortiuum: Global Snow Monitoring for Climate Research: Snow Water Equivalent, Finnish Meteorological Institute. Accessed: 2017-06-16, 2013.

Martens, B., Miralles, D. G., Lievens, H., van der Schalie, R., de Jeu, R. A. M., Fernández-Prieto, D., Beck, H. E., Dorigo, W. A., and

10   Verhoest, N. E. C.: GLEAM v3: satellite-based land evaporation and root-zone soil moisture, Geoscientific Model Development, 10, 1903–1925, https://doi.org/10.5194/gmd-10-1903-2017, 2017.

Marthews, T., Dadson, S., Lehner, B., Abele, S., and Gedney, N.: High-resolution global topographic index values, https://doi.org/10.5285/6b0c4358-2bf3-4924-aa8f-793d468b92be, NERC Environmental Information Data Centre. Accessed: 2017-12-01, 2015.

15   Ménard, C. B., Ikonen, J., Rautiainen, K., Aurela, M., Arslan, A. N., and Pulliainen, J.: Effects of meteorological and ancillary data, temporal averaging and evaluation methods on model performance and uncertainty in a land surface model, Journal of Hydrometeorology, https://doi.org/10.1175/JHM-D-15-0013.1, 2015.

Miralles, D. G., Holmes, T. R. H., De Jeu, R. A. M., Gash, J. H., Meesters, A. G. C. A., and Dolman, A. J.: Global land-surface evaporation estimated from satellite-based observations, Hydrology and Earth System Sciences, 15, 453–469, https://doi.org/10.5194/hess-15-453-

20   2011, 2011.

Mote, P. W.: Climate-Driven Variability and Trends in Mountain Snowpack in Western North America, Journal of Climate, 19, 6209–6220, https://doi.org/10.1175/JCLI3971.1, 2006.

Mote, T. L.: On the Role of Snow Cover in Depressing Air Temperature, Journal of Applied Meteorology and Climatology, 47, 2008–2022, https://doi.org/10.1175/2007JAMC1823.1, 2008.

25   Mu, Q., Heinsch, F. A., Zhao, M., and Running, S. W.: Development of a global evapotranspiration algorithm based on MODIS and global meteorology data, Remote Sensing of Environment, 111, 519 – 536, https://doi.org/10.1016/j.rse.2007.04.015, 2007.

Mudryk, L. R., Kushner, P. J., and Derksen, C.: Interpreting observed northern hemisphere snow trends with large ensembles of climate simulations, Climate Dynamics, 43, 345–359, https://doi.org/10.1007/s00382-013-1954-y, 2014.

Mudryk, L. R., Derksen, C., Kushner, P. J., and Brown, R.: Characterization of Northern Hemisphere Snow Water Equivalent Datasets,

30   1981-2010, J. Climate, p. 150807114710000, https://doi.org/10.1175/jcli-d-15-0229.1, 2015.

Mueller, B., Hirschi, M., Jimenez, C., Ciais, P., Dirmeyer, P. A., Dolman, A. J., Fisher, J. B., Jung, M., Ludwig, F., Maignan, F., Miralles, D. G., McCabe, M. F., Reichstein, M., Sheffield, J., Wang, K., Wood, E. F., Zhang, Y., and Seneviratne, S. I.: Benchmark products for land evapotranspiration: LandFlux-EVAL multi-data set synthesis, Hydrology and Earth System Sciences, 17, 3707–3720, https://doi.org/10.5194/hess-17-3707-2013, 2013.

35   NASA: GRACE Tellus Level-3 gridded land fields, http://grace.jpl.nasa.gov, accessed 2018-01-10, 2017.

Oki, T. and Sud, Y. C.: Design of Total Runoff Integrating Pathways (TRIP)—A Global River Channel Network, Earth Interactions, 2, 1–37, https://doi.org/10.1175/1087-3562(1998)002<0001:DOTRIP>2.3.CO;2, 1998.

Pavelsky, T. M. and Smith, L. C.: Intercomparison of four global precipitation data sets and their correlation with increased Eurasian river discharge to the Arctic Ocean, Journal of Geophysical Research: Atmospheres, 111, https://doi.org/10.1029/2006JD007230, 2006.

Räisänen, J.: Warmer climate: less or more snow?, Climate Dynamics, 30, 307–319, https://doi.org/10.1007/s00382-007-0289-y, 2008.

Rutter, N., Essery, R., Pomeroy, J., Altimir, N., Andreadis, K., Baker, I., Barr, A., Bartlett, P., Boone, A., Deng, H., Douville, H., Dutra, E., Elder, K., Ellis, C., Feng, X., Gelfan, A., Goodbody, A., Gusev, Y., Gustafsson, D., Hellström, R., Hirabayashi, Y., Hirota, T., Jonas, T., Koren, V., Kuragina, A., Lettenmaier, D., Li, W.-P., Luce, C., Martin, E., Nasonova, O., Pumpanen, J., Pyles, R. D., Samuelsson, P., Sandells, M., Schädler, G., Shmakin, A., Smirnova, T. G., Stähli, M., Stöckli, R., Strasser, U., Su, H., Suzuki, K., Takata, K., Tanaka, K., Thompson, E., Vesala, T., Viterbo, P., Wiltshire, A., Xia, K., Xue, Y., and Yamazaki, T.: Evaluation of forest snow processes models (SnowMIP2), Journal of Geophysical Research: Atmospheres, 114, https://doi.org/10.1029/2008JD011063, 2009.

Sakumura, C., Bettadpur, S., and Bruinsma, S.: Ensemble prediction and intercomparison analysis of GRACE time-variable gravity field models, Geophys. Res. Lett., 41, 1389–1397, https://doi.org/10.1002/2013gl058632, 2014.

Schneider, U., Becker, A., Finger, P., Meyer-Christoffer, A., Ziese, M., and Rudolf, B.: GPCC's new land surface precipitation climatology based on quality-controlled in situ data and its role in quantifying the global water cycle, Theoretical and Applied Climatology, 115, 15–40, https://doi.org/10.1007/s00704-013-0860-x, 2014.

Seo, K.-W., Ryu, D., Kim, B.-M., Waliser, D. E., Tian, B., and Eom, J.: GRACE and AMSR-E-based estimates of winter-season solid precipitation accumulation in the Arctic drainage region, Journal of Geophysical Research: Atmospheres, 115, https://doi.org/10.1029/2009JD013504, 2010.

Serreze, M. C., Clark, M. P., and Bromwich, D. H.: Monitoring Precipitation over the Arctic Terrestrial Drainage System: Data Requirements, Shortcomings, and Applications of Atmospheric Reanalysis, Journal of Hydrometeorology, 4, 387–407, https://doi.org/10.1175/1525-7541(2003)4<387:MPOTAT>2.0.CO;2, 2003.

Sheffield, J., Goteti, G., and Wood, E. F.: Development of a 50-Year High-Resolution Global Dataset of Meteorological Forcings for Land Surface Modeling, Journal of Climate, 19, 3088–3111, https://doi.org/10.1175/JCLI3790.1, 2006.

Stephens, G. L. and Kummerow, C. D.: The Remote Sensing of Clouds and Precipitation from Space: A Review, Journal of the Atmospheric Sciences, 64, 3742–3765, https://doi.org/10.1175/2006JAS2375.1, 2007.

Swenson, S.: Assessing High-Latitude Winter Precipitation from Global Precipitation Analyses Using GRACE, Journal of Hydrometeorology, 11, 405–420, https://doi.org/10.1175/2009jhm1194.1, 2010.

Swenson, S. and Wahr, J.: Post-processing removal of correlated errors in GRACE data, Geophys. Res. Lett., 33, https://doi.org/10.1029/2005gl025285, 2006.

Swenson, S. C.: GRACE monthly land water mass grids NETCDF RELEASE 5.0., https://doi.org/10.5067/TELND-NC005, PO.DAAC. Accessed: 2018-01-10, 2012.

Takala, M., Luojus, K., Pulliainen, J., Derksen, C., Lemmetyinen, J., Kärnä, J.-P., Koskinen, J., and Bojkov, B.: Estimating northern hemisphere snow water equivalent for climate research through assimilation of space-borne radiometer data and ground-based measurements, Remote Sensing of Environment, 115, 3517–3529, https://doi.org/10.1016/j.rse.2011.08.014, 2011.

Tian, X., Dai, A., Yang, D., and Xie, Z.: Effects of precipitation-bias corrections on surface hydrology over northern latitudes, Journal of Geophysical Research: Atmospheres, 112, https://doi.org/10.1029/2007JD008420, 2007.

Trenberth, K. E., Dai, A., van der Schrier, G., Jones, P. D., Barichivich, J., Briffa, K. R., and Sheffield, J.: Global warming and changes in drought, Nature Climate Change, 4, 17–22, https://doi.org/10.1038/nclimate2067, 2013.

Troy, T. J., Sheffield, J., and Wood, E. F.: Estimation of the Terrestrial Water Budget over Northern Eurasia through the Use of Multiple Data Sources, Journal of Climate, 24, 3272–3293, https://doi.org/10.1175/2011JCLI3936.1, 2011.

Vaughan, D. G., Comiso, J. C., Allison, I., Carrasco, J., Kaser, G., Kwok, R., Mote, P. W., Murray, T., Paul, F., Ren, J., Rignot, E., Solomina, O., Steffen, K., and Zhang, T.: Observations: Cryosphere, pp. 317–382, Cambridge University Press, https://doi.org/10.1017/CBO9781107415324.012, 2013.

Viovy, N.: CRUNCEP v4 (CRU TS 3.21), https://dods.extra.cea.fr/store/p529viov/cruncep/V4_1901_2012, accessed: 2014-07-11, 2014.

Wahr, J., Molenaar, M., and Bryan, F.: Time variability of the Earth's gravity field: Hydrological and oceanic effects and their possible detection using GRACE, Journal of Geophysical Research: Solid Earth, 103, 30 205–30 229, https://doi.org/10.1029/98JB02844, 1998.

Weedon, G. P., Gomes, S., Viterbo, P., Shuttleworth, W. J., Blyth, E., Österle, H., Adam, J. C., Bellouin, N., Boucher, O., and Best, M.: Creation of the WATCH Forcing Data and Its Use to Assess Global and Regional Reference Crop Evaporation over Land during the Twentieth Century, Journal of Hydrometeorology, 12, 823–848, https://doi.org/10.1175/2011JHM1369.1, 2011.

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 Resources Research, 50, 7505–7514, https://doi.org/10.1002/2014WR015638, 2014.

Weiland, F. C. S., Vrugt, J. A., van Beek, R. L. P., Weerts, A. H., and Bierkens, M. F.: Significant uncertainty in global scale hydrological modeling from precipitation data errors, Journal of Hydrology, 529, 1095 – 1115, https://doi.org/10.1016/j.jhydrol.2015.08.061, 2015.

Wisser, D., Fekete, B. M., Vörösmarty, C. J., and Schumann, A. H.: Reconstructing 20th century global hydrography: a contribution to the Global Terrestrial Network- Hydrology (GTN-H), Hydrology and Earth System Sciences, 14, 1–24, https://doi.org/10.5194/hess-14-1-2010, 2010.

Yang, D., Kane, D., Zhang, Z., Legates, D., and Goodison, B.: Bias corrections of long-term (1973–2004) daily precipitation data over the northern regions, Geophysical Research Letters, 32, https://doi.org/10.1029/2005GL024057, 2005.

Yang, J., Yau, M. K., Fang, X., and Pomeroy, J. W.: A triple-moment blowing snow-atmospheric model and its application in computing the seasonal wintertime snow mass budget, Hydrology and Earth System Sciences, 14, 1063–1079, https://doi.org/10.5194/hess-14-1063-2010, 2010.
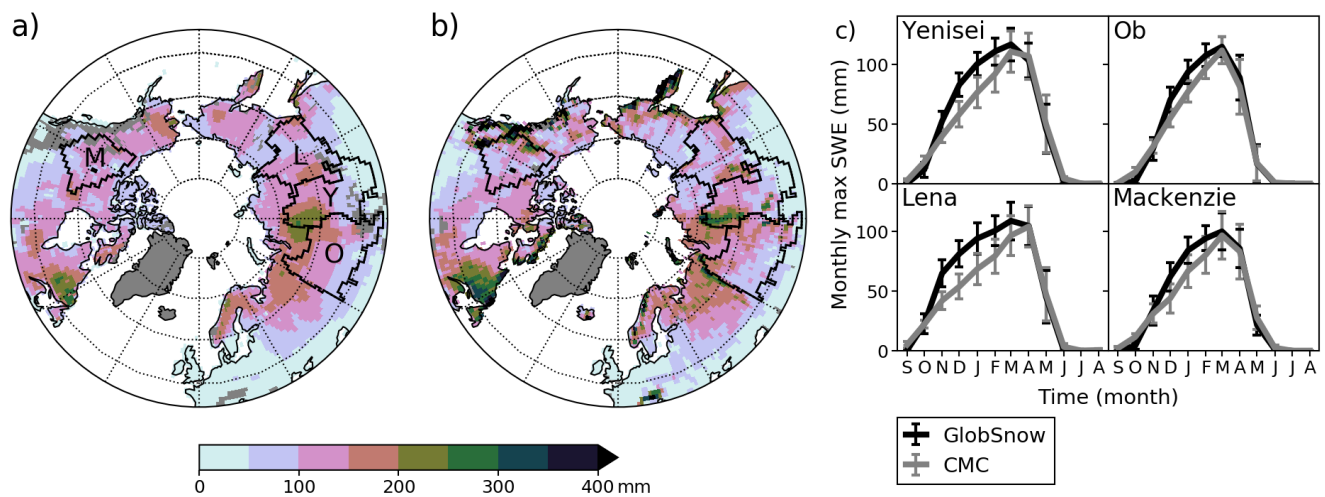
Figure 1. Seasonal maximum SWE (mm) for (a) GlobSnow and (b) CMC for the common overlap period, 2002–2010. The basins used in this study are labelled with their initial (*Y*enisei, *O*b, *L*ena, *M*ackenzie). Grey shows regions that are masked. In both data sets Greenland is masked. In addition, GlobSnow pixels are masked where they are more than 50% "mountain". (c) Mean-monthly climatology averaged over the basins (masked using the GlobSnow mountain mask) for the common overlap period, 2002–2010. Error bars show the inter-annual variability, defined as the standard deviation between years.
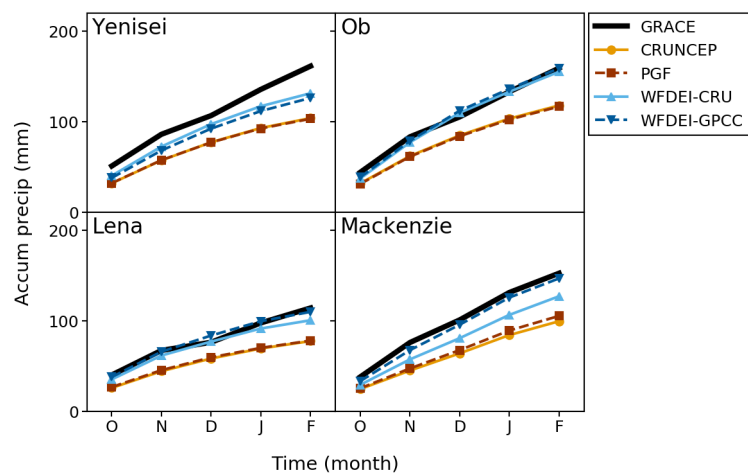
Figure 2. Mean monthly accumulated cold season precipitation in October–February inclusive, averaged over the common overlap period, 2002–2010, as derived from GRACE TWS (Sect. 4) and from the driving data sets (Sect. 2.1).
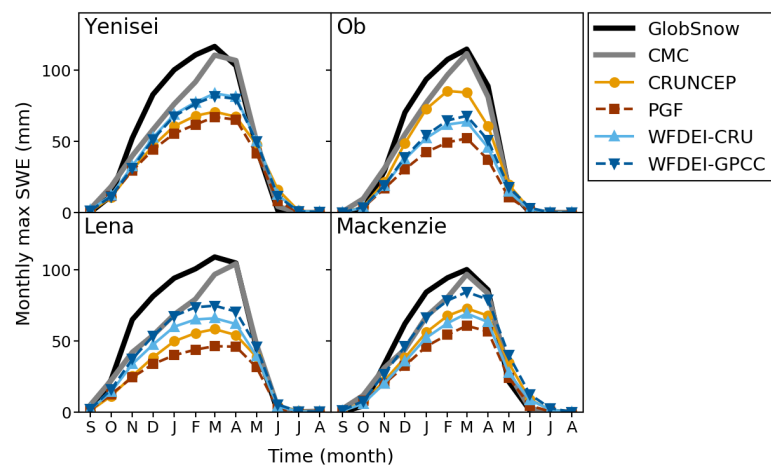
Figure 3. Climatology of monthly maximum SWE (mm) from observations and from the CTL model runs averaged over the common overlap period, 2002–2010.
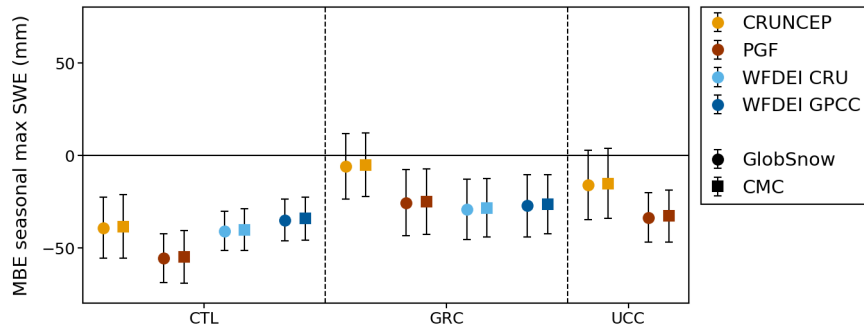
Figure 4. Mean bias error (MBE) of seasonal maximum SWE (mm), averaged over all basins (Eq. 2). Calculated over the common overlap period, 2002–2010. The error bars show the basin averaged standard deviation, σ_S (Eq. 4).
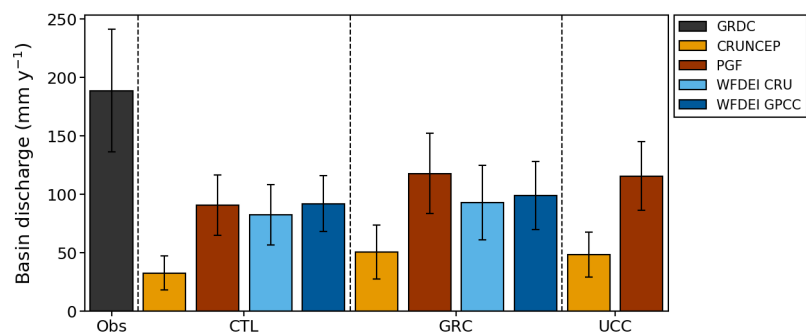
Figure 5. Mean annual observed and modelled basin discharge (mm y^{-1}) for the common overlap period, 2002–2010.
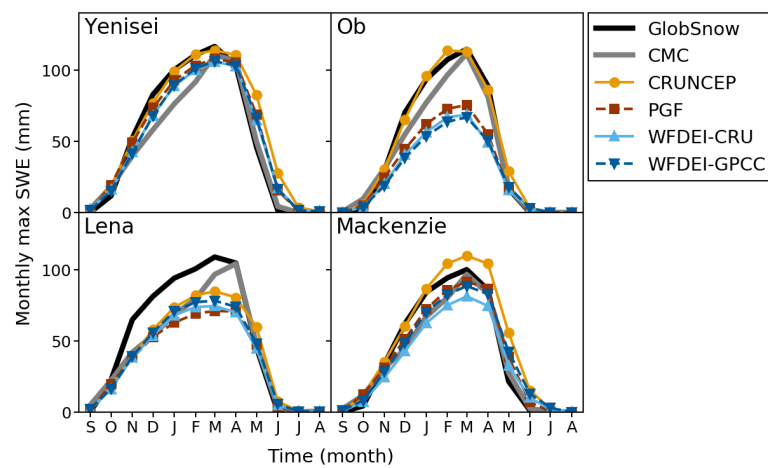
Figure 6. Climatology of monthly maximum SWE (mm) from observations and from the GRC model runs for the common overlap period, 2002–2010.
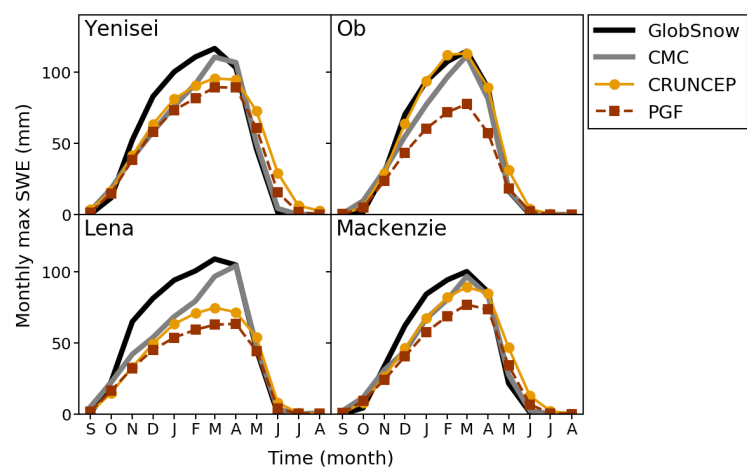
Figure 7. Climatology of monthly maximum SWE (mm) from observations and from the UCC model runs for the common overlap period, 2002–2010.
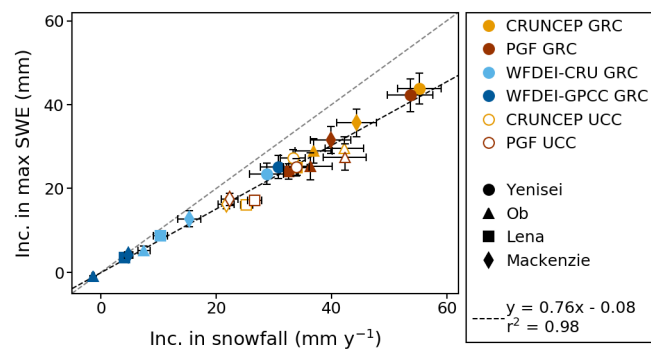
Figure 8. Increase in annual maximum SWE (mm) from CTL to GRC and UCC runs compared to increase in annual snowfall (mm y^{-1}), averaged over all basins (using the GlobSnow mountain mask for snowfall as well as SWE) for each driving data set. This is averaged over the whole of each JULES run (see Table 1).

Table 1. Summary of meteorological driving data.

Driving data	Years avail. (used)	Temporal resolution	Spatial resolution	Reanalysis	Source of precipitation	Precip. variables	Undercatch correction
CRUNCEP v4 (Viovy, 2014)	1901–2010 (1970–2010)	6 hour	0.5°	NCEP/NCAR Reanalysis Project 1	CRU TS 3.21	Total	No
PGF v2 (Sheffield et al., 2006)	1948–2012 (1970–2012)	3 hour	1°	NCEP/NCAR Reanalysis Project 1	CRU TS	Total	No
WFDEI-CRU (Weedon et al., 2014)	1979–2016	3 hour	0.5°	ERA-Interim	CRU TS 3.101 and 3.21	Rain, snow	Yes
WFDEI-GPCC (Weedon et al., 2014)	1979–2013	3 hour	0.5°	ERA-Interim	GPCC v5 and 6	Rain, snow	Yes

Table 2. Precipitation scale factors, calculated as the ratio of GRACE-derived to data set precipitation, using all years of overlap between each driving data set and GRACE.

Driving data	Yenisei	Ob	Lena	Mackenzie	Mean
CRUNCEP	1.55	1.35	1.47	1.53	1.47
PGF	1.55	1.37	1.55	1.54	1.49
WFDEI-CRU	1.24	1.06	1.11	1.16	1.14
WFDEI-GPCC	1.27	0.98	1.04	1.05	1.08