Suitability of North American Regional Reanalysis (NARR) output for hydrologic modelling and analysis in mountainous terrain

Joel W. Trubilowicz,* J. M. Shea, G. Jost and R. D. Moore

1 Department of Geography, University of British Columbia, 1984 West Mall, Vancouver, BC, V6T 1Z2, Canada
2 International Centre for Integrated Mountain Development, GPO Box 3226, Kathmandu, Nepal
3 Generation Resource Management, BC Hydro, 6911 Southpoint Dr., Burnaby, BC, V3N 4X8, Canada
4 Department of Forest Resource Management, University of British Columbia, 225-1984 West Mall, Vancouver, BC, V6T 1Z2, Canada

Abstract:

Meteorological observations at high elevations in mountainous regions are often lacking. One opportunity to fill this data gap is through the use of downscaled output from weather reanalysis models. In this study, we tested the accuracy of downscaled output from the North American Regional Reanalysis (NARR) against high-elevation surface observations at four ridgetop locations in the southern Coast Mountains of British Columbia, Canada. NARR model output was downscaled to the surface observation locations through three-dimensional interpolation for air temperature, vapour pressure and wind speed and two-dimensional interpolation for radiation variables. Accuracy was tested at both the 3-hourly and daily time scales. Air temperature displayed a high level of agreement, especially at the daily scale, with root mean square error (RMSE) values ranging from 0.98 to 1.21 °C across all sites. Vapour pressure downscaling accuracy was also quite high (RMSE of 0.06 to 0.11 hPa) but displayed some site specific bias. Although NARR overestimated wind speed, there were moderate to strong linear relations ($r^2$ from 0.38 to 0.84 for daily means), suggesting that the NARR output could be used as an index and bias-corrected. NARR output reproduced the seasonal cycle for incoming short-wave radiation, with Nash–Sutcliffe model efficiencies ranging from 0.78 to 0.87, but accuracy suffered on days with cloud cover, resulting in a positive bias and RMSE ranged from 42 to 46 Wm$^{-2}$. Although fewer data were available, incoming long-wave radiation from NARR had an RMSE of 19 Wm$^{-2}$ and outperformed common methods for estimating incoming long-wave radiation. NARR air temperature showed potential to assist in hydrologic analysis and modelling during an atmospheric river storm event, which are characterized by warm and wet air masses with atypical vertical temperature gradients. The incorporation of a synthetic NARR air temperature station to better represent the higher freezing levels resulted in increased predicted peak flows, which better match the observed run-off during the event. Copyright © 2016 John Wiley & Sons, Ltd.

KEY WORDS mountain hydrology; reanalysis; downscaling; hydrologic modelling; hydrologic forecasting

Received 31 August 2015; Accepted 19 January 2016

INTRODUCTION

The predictions in ungauged basins decade of 2003–2013 recognized a lack of observations as one of the primary obstacles to increasing our understanding of hydrologic processes (Sivapalan et al., 2003). Great strides have been made in the realms of remote sensing and alternative observation networks over this time period (Hrachowitz et al., 2013), but high-elevation meteorological observations remain one of the major needs for hydrologic modelling and analysis in mountainous environments (Moore et al., 2013). Because of the logistical difficulties and hazards of installation and maintenance, as well as the volatile weather that can affect the ability to collect accurate observations, environmental observation networks are often missing high-elevation data.

In situations in which high-elevation data are lacking, observations from low-elevation stations are typically extrapolated to higher elevations, with orographic effects addressed by application of vertical gradients and other topographic adjustments. Routines and software packages have been developed to estimate weather variables at individual points (e.g. Running et al., 1987) and in gridded spatial fields (e.g. Thornton et al., 1997; Liston and Elder, 2006); however, fundamental assumptions about the vertical gradient of variables must be made in these situations.

Numerous studies have focused on the interpolation/extrapolation of air temperature (e.g. Bolstad et al., 1998; Jarvis and Stuart, 2001; Hasenauer et al., 2003; Garen and
In British Columbia (BC), for example, Stahl *et al.* (2006) found that knowledge of the vertical temperature gradient was the dominant factor in estimating high-elevation temperatures. A major challenge in estimating high-elevation air temperature is that the temperature gradient can vary substantially in space and time, and even within and between individual storm events.

In comparison with air temperature, less research has addressed the extrapolation of other variables. Water vapour content is often assumed to be relatively spatially homogeneous and closely related to the minimum daily air temperature at a site (Running *et al.*, 1987). Kimball *et al.* (1997) estimated water vapour content through an empirical method based on the daily air temperature, mean annual precipitation and potential evapotranspiration. Garen and Marks (2005) distributed vapour pressure by elevation through a combination of upper air soundings and limited surface measurements. Wind speed and direction are highly variable in space and time and closely linked to local land cover and topography (Barry, 2008). Most work on interpolation/extrapolation of wind observations in mountainous areas has involved modelling the effects of the topography on wind speed for the purposes of predicting snow deposition, redistribution and melt (e.g. Winstral and Marks, 2002; Mott and Lehning, 2010). Similar to their approach for vapour pressure, Garen and Marks (2005) used upper air soundings to create 3-hourly wind speed fields for different elevations. Methods for estimating incoming short-wave and long-wave radiation typically require air temperature observations at the high-elevation site. Daily incoming short-wave radiation is often estimated using an approach based on that of Bristow and Campbell (1984), which uses daily air temperature range to estimate atmospheric transmissivity. Bristow and Campbell (1984) noted that the accuracy of this method is dependent on site-specific parameter calibration. Clear-sky long-wave radiation is often estimated using air temperature and vapour pressure at a site using empirical formulae such as that presented by Prata (1996).

An alternative to interpolation/extrapolation from observational networks is to downscale output from atmospheric models, either to infill missing data (Way and Bonnaveventure, 2015) or to drive hydrologic models (e.g. Wilby *et al.*, 2000; Bastola and Misra, 2014), but many atmospheric models are too coarse spatially to represent meteorological variation within mountainous environments at the subregional scale. Regional reanalysis models offer a higher resolution with greater potential in mountainous environments and smaller scales. Fiddes and Gruber (2014) developed the TopoSCALE application to downscale output from the ERA-Interim 0.75° resolution reanalysis model (Dee *et al.*, 2011) and noted that the same techniques could likely be applied to other regional reanalysis products.

With a 32-km (~0.3°) spatial resolution, the North American Regional Reanalysis (NARR) model (Mesinger *et al.*, 2006) is the highest-resolution reanalysis model for North America. NARR temperature output has been found to provide utility for accurately reproducing surface observations in mountainous areas. Holden *et al.* (2015) created a 250-m resolution temperature data set for the US Northern Rocky Mountains using a combination of modelled solar insolation and soil moisture, GIS-based topography and landcover and lapse rates derived from NARR temperature output, and found mean absolute errors <1.4°C when compared with surface observations for daily maximum and minimum temperature. NARR output has been used directly as input to macro-scale hydrologic models (e.g. Woo and Thorne, 2006; Choi *et al.*, 2009), in regional water budget calculations (e.g. Luo *et al.*, 2007; Ruane, 2010; Sheffield *et al.*, 2012), and interpolated and downscaled for use in a glaciation model (Jarosch *et al.*, 2012). Choi *et al.* (2009) and Kesht and Elshorbagy (2011) found that NARR output provided utility for hydrologic purposes, especially when other observations are unavailable. However, those two studies focused on the Canadian Prairies, where the topography is generally low relief. The utility of NARR output for hydrologic purposes in mountainous regions has remained relatively unexplored.

Although the horizontal resolution of NARR output is likely not detailed enough to drive a hydrologic model for small-scale to medium-scale catchments (e.g. up to 1000 km² in area) without ground-based supporting observations in a mountainous region, it might have utility as a supplement to data from observation networks, particularly for extrapolating to higher elevations. Reanalysis output might be expected to perform best on high alpine ridges, where the influence of the boundary layer is at a minimum (Barry, 2008). However, the performance of NARR in replicating observational data at these locations has received little attention because of the lack of high-elevation observations (Jarosch *et al.*, 2012).

In this study, our primary goal is to assess the utility of NARR output for hydrologic analysis and modelling in a mountainous region. We do this by (1) comparing downscaled NARR output to surface observations at automated weather stations (AWS) located at ridgetop locations (Shea and Moore, 2010) and (2) illustrating the potential utility of NARR output as high-altitude intelligence for modelling snow dynamics and streamflow response to an atmospheric river (AR) event (Zhu and Newell, 1998), which are often poorly forecast.
because of incorrectly specified storm freezing levels (Moore, 1993; Rössler et al., 2014).

STUDY AREA

The study was conducted in the southern Coast Mountains of BC, Canada, a rugged, mountainous region with elevations ranging from sea level to greater than 2500 m (Figure 1). Topography is dominated by U-shaped main valleys fed by steep tributary streams, which are typically headed by cirque basins. The region contains substantial ice cover in the form of cirque glaciers, as well as high-elevation ice fields drained by valley glaciers.

The climate is generally moist maritime, with mild temperatures, wet winters and relatively dry summers. Low-elevation areas near sea level generally receive most precipitation in the form of rain, whereas seasonally continuous snow cover with upwards of 2000 mm of snow water equivalent commonly occurs above 1000 m. There is a marked west-east climatic gradient, with strong maritime conditions on the upwind slopes and the coastal divide, grading to drier conditions east of the divide.

Elevations near sea level are dominated by the Coastal Western Hemlock biogeoclimatic zone, shifting first to the Mountain Hemlock zone and then alpine vegetation at higher elevations (Klinka et al., 1991). In the drier, eastern portions of the Coast Mountains, the Coastal Western Hemlock and Mountain Hemlock zones are replaced by the Interior Douglas-fir and Montane Spruce zones, respectively.

METHODS

Weather observations

Meteorological data used in this study were collected from ridgetop AWS to minimize boundary layer effects, particularly those associated with katabatic flow (Shea and Moore, 2010). Station locations are shown in Figure 1. Measured variables include air temperature ($T_a$, °C), relative humidity (RH, %), wind speed (w, m s$^{-1}$) and direction ($\theta$, °), incoming short-wave radiation ($K_{\downarrow}$, Wm$^{-2}$) and incoming long-wave radiation ($L_{\downarrow}$, Wm$^{-2}$) (Table I). Sensors were scanned every second, and averaged values were recorded at 10-min intervals. Data were recorded year-round at the Bridge, Helm and Weart sites, but only during summer at Place. Further details on the instrumentation and collection of surface observations can be found in Shea and Moore (2010).

Vapour pressure ($e_a$, hPa) was calculated as follows:

$$e_a = e_s \frac{RH}{100}$$  \hspace{1cm} (1)

where $RH$ is relative humidity (%) and $e_s$ is the saturation vapour pressure (kPa), computed as in Buck (1981):

$$e_s = \begin{cases} 0.611 \cdot \exp \left( \frac{17.37 \cdot T_a}{T_a + 238.9} \right) & : T_a \geq 0 \\ 0.611 \cdot \exp \left( \frac{22.45 \cdot T_a}{T_a + 272.5} \right) & : T_a < 0 \end{cases}$$  \hspace{1cm} (2)

Figure 1. Study region, showing locations of automated weather stations. The catchment area for BC Hydro’s Daisy Lake Reservoir is shown in orange. North American Regional Reanalysis (NARR) grid points are shown in red. AWS, automated weather station.
NARR products

The NARR model (Mesinger et al., 2006) uses surface, radar and satellite observations as input. Output is gridded at a resolution of 32 km (approximately 0.3°), with a 3-h time step at multiple pressure levels for more than 400 meteorological variables from 1979 to near present. We extracted $T_a$, specific humidity ($q_a$), wind velocity from east–west ($U$) and north–south components ($V$), $K_\downarrow$ and $L_\downarrow$ for comparison with surface observations.

For $T_a$, $q_a$ and wind velocity ($U$ and $V$ components), NARR output is given at multiple pressure levels, with each pressure level having a corresponding dynamic geopotential height. Thus, the interpolation was performed in three dimensions. For each time step, the two nearest heights to the elevation of a site were selected at each of the four nearest NARR centroids (Figure 1), and the desired variable was linearly interpolated to the reference elevation of the site. Following this vertical interpolation, inverse distance weighting of the four nearest NARR centroids was used to interpolate horizontally to the station location. These multilevel variables represent instantaneous measures at a 3-h time step.

North American Regional Reanalysis short-wave and long-wave radiation are only available as mono-level output at the earth’s surface as represented in the NARR digital elevation model. These variables were only interpolated in two dimensions, using inverse distance weighting interpolation from the four nearest points. These mono-level variables represent the mean irradiance over a 3-h time step, with the time stamp at the start of the 3-h interval.

The NARR-specific humidity and $U$ (westerly) and $V$ (northerly) components of the wind vector were post-processed for direct comparison to surface observations. Wind speed was computed from the vector winds with the Pythagorean theorem, and $q_a$ was converted to $e_a$ (kPa):

$$e_a = q_a \frac{P}{0.0378 \cdot q_a + 0.0622}$$

(3)

where $P$ is the air pressure supplied directly from NARR (kPa) at the geopotential height of the station ($m$).

Table I. High-elevation site data

<table>
<thead>
<tr>
<th>Site name</th>
<th>Z (m)</th>
<th>NARR Z (m)</th>
<th>Date range</th>
<th>$T_a$</th>
<th>RH</th>
<th>w</th>
<th>$\theta$</th>
<th>$K_\downarrow$</th>
<th>$L_\downarrow$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bridge</td>
<td>1640</td>
<td>1650</td>
<td>September 2006 to August 2008</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Helm</td>
<td>2192</td>
<td>1248</td>
<td>August 2006 to September 2008</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Place</td>
<td>2075</td>
<td>1519</td>
<td>May 2007 to September 2008</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Weart</td>
<td>2220</td>
<td>1456</td>
<td>July 2006 to September 2008</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

$Z$ is the site elevation (m), and NARR $Z$ is the elevation of the 32-km resolution NARR grid cell in which each site is located.

NARR, North American Regional Reanalysis; RH, relative humidity.

North American Regional Reanalysis variables were extracted using the ‘ncdf’ package (Pierce, 2014) in the R computing language (R Development Core Team, 2015). Data were extracted from the National Oceanic and Atmospheric Administration data server, and all post-processing was performed using R.

Alternative methods for estimating incoming radiation

Some commonly used methods for estimating incoming short-wave and long-wave radiation at remote sites were applied in order to assess whether the NARR output provided any additional skill relative to those methods.

Potential direct incoming solar radiation ($K_\downarrow$) was calculated at each site at both the 3-hourly and daily time steps using standard solar geometry equations from Iqbal (1983) and the potential incoming short-wave solar radiation equation from Hock (1999), which was modified by Jost et al. (2012) as follows:

$$K_\downarrow = I_0 \cdot E_0 \cdot \frac{\tau}{\cos(z)} \cdot \cos(\theta_s)$$

(4)

$$p_0 = e^{\frac{\tau}{\cos(z)}}$$

(5)

where $I_0$ is the solar constant, $E_0$ is the earth orbit eccentricity factor (Iqbal, 1983), $\tau$ is the mean atmospheric clear sky transmissitivity (set at 0.75 as in Hock (1999)), $z$ is the site elevation and $\theta_s$ is the solar zenith angle.

In order to assess the potential benefits of using NARR output over more common methods, incoming long-wave radiation was estimated by the Stefan–Boltzmann law using both an assumed emissivity ($\varepsilon$) of 0.85 and an emissivity calculated using the Prata (1996) equation:

$$\varepsilon = 1 - (1 + \xi) e^{-1.2 + 3.0 \xi^{0.5}}$$

(6)

$$\xi = \frac{e_0}{T_0} - 46.5$$

(7)

where $T_0$ is the surface temperature (K) and $e_0$ is the vapour pressure (hPA).

1http://www.esrl.noaa.gov/psd/thredds/dodsC/Datasets/NARR/
Performance measures

The NARR downscaling skill (and accuracy of alternative estimation methods) was assessed using the mean bias error (MBE), root mean square error (RMSE), the Nash–Sutcliffe model efficiency ($E_m$) (Nash and Sutcliffe, 1970) and the coefficient of determination for a linear regression of NARR versus observed ($r^2$), as follows:

$$\text{MBE} = \frac{1}{n} \sum (y_i - x_i)$$  \hspace{1cm} (8)

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum (y_i - x_i)^2}$$  \hspace{1cm} (9)

$$E_m = 1 - \frac{\sum (y_i - x_i)^2}{\sum (x_i - \bar{x})^2}$$  \hspace{1cm} (10)

$$r^2 = 1 - \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2}$$  \hspace{1cm} (11)

where $y_i$ is the value predicted by NARR and $x_i$ is the observed value at time step $i$, $\bar{x}$ is the mean value of all observations, $n$ is the number of observations, $\bar{y}$ is the mean of all NARR output and $\hat{y}_i$ is the value predicted by a linear model between NARR and observations. We calculated performance measures for both daily mean and instantaneous (3-hourly) downscaled data.

RESULTS

Example time series

Subsets of the downscaled and observed meteorological data at Bridge and Place Glacier sites illustrate the 3-hourly data (Figures 2 and 3). NARR-predicted air temperature follows the observed diel and multi-day variations at both sites. Vapour pressure follows well with observations at the Place Glacier site (Figure 3), but NARR significantly overestimates $e_a$ at the Bridge Glacier site during the first 3 days (Figure 2). Wind speeds from NARR are of similar magnitude to observations at the Bridge Glacier site but do not appear to closely follow individual observations (Figure 2). Downscaled incoming solar radiation tracks observations during most clear days at the Bridge Glacier site, but not during periods with cloud cover (Figure 2). Downscaled incoming long-wave radiation is of similar magnitude to observations at the Place Glacier site but does not appear to correlate well with observations (Figure 3).
Air temperature

There were strong linear relations between observations and NARR downscaled mean daily air temperature at all sites (Figure 4). An over-prediction of 1–2 °C at temperatures below –10 °C can be seen at the Bridge and Weart sites. The seasonality of error is shown in Figure 5. There is no strong seasonal component to the error, but NARR-estimated temperatures at Bridge and Weart tend to be greater than observed in winter. As Place data were only available during summer, an investigation of seasonality was not possible.

The downscaled NARR output had a slight negative bias at the Place site, whereas all other sites had a small, positive MBE. RMSE, $E_m$, and $r^2$ improved in all cases by integrating to daily resolution from 3-h values (Table II).

Vapour pressure

Downscaled values of mean daily $e_a$ are highly correlated with the observations (Figure 6), although more weakly than for $T_a$. Vapour pressure was over-predicted by NARR at the Bridge site, but under-predicted at the Helm site, with low
Figure 6. Comparison of North American Regional Reanalysis (NARR)-derived and observed mean daily vapour pressure. The straight line indicates perfect agreement.

Table II. Performance measures for air temperature at each site

<table>
<thead>
<tr>
<th>Site</th>
<th>MBE (°C)</th>
<th>RMSE (°C)</th>
<th>$E_m$</th>
<th>$r^2$</th>
<th>RMSE (°C)</th>
<th>$E_m$</th>
<th>$r^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bridge</td>
<td>0.17</td>
<td>1.58</td>
<td>0.96</td>
<td>0.96</td>
<td>1.21</td>
<td>0.97</td>
<td>0.98</td>
</tr>
<tr>
<td>Helm</td>
<td>0.43</td>
<td>1.67</td>
<td>0.95</td>
<td>0.95</td>
<td>1.07</td>
<td>0.98</td>
<td>0.98</td>
</tr>
<tr>
<td>Place</td>
<td>−0.13</td>
<td>1.31</td>
<td>0.94</td>
<td>0.94</td>
<td>0.80</td>
<td>0.98</td>
<td>0.98</td>
</tr>
<tr>
<td>Weart</td>
<td>0.47</td>
<td>1.61</td>
<td>0.96</td>
<td>0.96</td>
<td>0.98</td>
<td>0.98</td>
<td>0.99</td>
</tr>
</tbody>
</table>

MBE, mean bias error; RMSE, root mean square error.

Figure 7. Monthly variation of error between North American Regional Reanalysis (NARR)-derived and observed mean daily vapour pressure.
bias for the Place and Weart Glacier sites. Figure 7 reveals this is mostly due to differences in error during summer. Performance measures for vapour pressure (Table III) indicate good downscaling skill, although slightly less strong than for $T_a$. MBE was negative for Helm and positive for all other sites. As with $T_a$, daily integration improved RMSE, $E_m$ and $r^2$ in all cases.

<table>
<thead>
<tr>
<th>Site</th>
<th>MBE (hPa)</th>
<th>RMSE (hPa)</th>
<th>$E_m$</th>
<th>$r^2$</th>
<th>RMSE (hPa)</th>
<th>$E_m$</th>
<th>$r^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bridge</td>
<td>0.09</td>
<td>0.13</td>
<td>0.58</td>
<td>0.84</td>
<td>0.11</td>
<td>0.62</td>
<td>0.89</td>
</tr>
<tr>
<td>Helm</td>
<td>-0.05</td>
<td>0.12</td>
<td>0.69</td>
<td>0.75</td>
<td>0.10</td>
<td>0.77</td>
<td>0.84</td>
</tr>
<tr>
<td>Place</td>
<td>0.02</td>
<td>0.11</td>
<td>0.55</td>
<td>0.60</td>
<td>0.08</td>
<td>0.66</td>
<td>0.70</td>
</tr>
<tr>
<td>Weart</td>
<td>0.02</td>
<td>0.08</td>
<td>0.85</td>
<td>0.86</td>
<td>0.06</td>
<td>0.91</td>
<td>0.92</td>
</tr>
</tbody>
</table>

MBE, mean bias error; RMSE, root mean square error.

Wind speed

Downscaled NARR output consistently and strongly over-predicted wind speed at Helm and Weart glaciers (Figure 8). The over-prediction is most prevalent in the winter season, when the southern Coast Mountains experience frequent mid-latitude storms (Figure 9). The positive bias is most likely due to NARR being
designed to represent the free atmosphere, while surface friction acts to reduce wind speeds measured on the ground.

The bias in the NARR-derived wind speeds is reflected in the performance indices (Table IV). In particular, the negative model efficiencies indicate that the observed mean is a better predictor than the NARR values. RMSE was improved in every case by resampling to daily resolution; however, \( E_m \) was decreased for the Weart site. As expected, MBE was positive in all cases. Although there was a positive bias, there was a reasonable linear relation between NARR output and observed wind speeds, particularly at the Weart Glacier site.

Incoming short-wave radiation

Observed and downscaled \( K_{\downarrow} \) are correlated, although the spread is greater than for \( T_a, e_a \) or wind speed (Figure 10). Performance measures for \( K_{\downarrow} \) indicate large improvements in RMSE when data are integrated to daily resolution (Table V). \( E_m \) decreased when integrated to daily values in all cases, likely because of the removal of the diurnal cycle of short-wave radiation, which is captured well by NARR (Figure 2). MBE was positive for all sites.

The downscaled NARR output outperformed the modified Hock (1999) model, which yielded higher RMSE and lower \( E_m \) and \( r^2 \) (Table VI). The NARR output and the Hock (1999) model had similar patterns of bias among the three sites.

Incoming long-wave radiation

Observed and interpolated long-wave radiation from NARR are poorly correlated (Figure 11), although the period of record was more limited than for the other variables. With only one site and only two summer’s data, it was more difficult to assess the performance of NARR for predicting \( L_{\downarrow} \), or the seasonal variability in errors. Long-wave radiation interpolated from NARR performed better than the empirically estimated alternative (Table VII). Although the Stefan–Boltzmann law with an emissivity of 0.85 yielded a slightly lower MBE than NARR, the NARR output outperformed the alternative approaches on all other performance measures (Table VII).

Table IV. Performance measures for wind speed at each site

<table>
<thead>
<tr>
<th>Site</th>
<th>MBE (m s(^{-1}))</th>
<th>RMSE (m s(^{-1}))</th>
<th>( E_m )</th>
<th>( r^2 )</th>
<th>RMSE (m s(^{-1}))</th>
<th>( E_m )</th>
<th>( r^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bridge</td>
<td>0.75</td>
<td>2.60</td>
<td>-0.52</td>
<td>0.16</td>
<td>1.73</td>
<td>-0.19</td>
<td>0.38</td>
</tr>
<tr>
<td>Helm</td>
<td>2.41</td>
<td>4.39</td>
<td>-1.87</td>
<td>0.34</td>
<td>4.33</td>
<td>-3.37</td>
<td>0.35</td>
</tr>
<tr>
<td>Weart</td>
<td>4.54</td>
<td>5.84</td>
<td>-5.51</td>
<td>0.65</td>
<td>5.49</td>
<td>-7.18</td>
<td>0.84</td>
</tr>
</tbody>
</table>

MBE, mean bias error; RMSE, root mean square error.
are distributed to higher elevations using vertical temperature gradients calibrated by fitting modelled to observed run-off (e.g. Quick and Pipes, 1977; Leavesley and Stannard, 1995). During rain-on-snow conditions associated with ARs, which generate major flood events in the Pacific Northwest of the USA (Neiman et al., 2011) and coastal BC (Spry et al., 2014), environmental lapse rates can be substantially lower than the standard moist adiabatic lapse rate or modelled vertical temperature gradients calibrated for normal conditions, leading to under-prediction of air temperature at higher elevations and incorrect classification of precipitation as snow. As a consequence, the flood response can be dramatically under-predicted (Moore, 1993). In this section, we explore whether the use of NARR air temperatures can improve hydrologic prediction during a major AR event. We hypothesize that the under-prediction of streamflow response during AR events is caused in large part by incorrectly specified vertical temperature gradients and the associated problems with classification of precipitation as snow over substantial portions of the catchment.

**Study catchment and modelling system**

We focus on the 721-km² Cheakamus catchment (with an outlet at Daisy Lake), shown in orange in Figure 1. Elevation ranges from 369 to 2594 m. The Cheakamus catchment is 13% glacierized, and approximately 50% of the total area is above treeline. Although the majority of annual run-off derives from seasonal melting of snow and glacier ice, rain-on-snow events during autumn and winter can generate higher peak flows than those associated with snow and glacier melt.

The Cheakamus watershed has been regulated since 1957 for hydroelectric power generation by BC Hydro, the primary power utility of BC (BC Hydro, 2002). BC Hydro calculates inflows to the Daisy Lake reservoir on a daily basis using a water balance model based on the changes in reservoir levels and the amount of water released from the dam.

The UBC watershed model (UBCWM) (Quick and Pipes, 1977) is the reference hydrologic forecasting model for BC Hydro. UBCWM discretizes a watershed into hydrologic response units within elevation bands. Input data include daily maximum and minimum air temperature and precipitation, both from observations and weather forecasts. Input data are distributed by elevation using a multi-segment approach in which the vertical temperature gradient varies with elevation (with an inflection point at 2000 m), the amount of daily precipitation and the daily temperature range (Quick and Pipes, 1977). Temperature and precipitation gradients are determined through calibration to match the observed hydrograph.

The UBCWM configuration for predicting inflow into Daisy Lake uses a single, synthetic station as input, known as Cheakamus synthetic (CMS). This station is a linear combination of three separate observation locations falling within the watershed and has a representative virtual elevation of 827 m. Along with being a requirement for the operational version of the UBCWM, the use of a single synthetic station and calibrated vertical temperature gradient, instead of three stations separately, streamlines data processing, simplifies situations with

**Table V. Performance measures for K↓ at each site**

<table>
<thead>
<tr>
<th>Site</th>
<th>Obs. (Wm⁻²)</th>
<th>MBE (Wm⁻²)</th>
<th>RMSE (Wm⁻²)</th>
<th>Eᵣ²</th>
<th>r²</th>
<th>RMSE (Wm⁻²)</th>
<th>Eᵣ²</th>
<th>r²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bridge</td>
<td>171</td>
<td>6.3</td>
<td>84.2</td>
<td>0.89</td>
<td>0.89</td>
<td>42.44</td>
<td>0.87</td>
<td>0.87</td>
</tr>
<tr>
<td>Helm</td>
<td>157</td>
<td>23.5</td>
<td>94.6</td>
<td>0.84</td>
<td>0.86</td>
<td>52.1</td>
<td>0.78</td>
<td>0.83</td>
</tr>
<tr>
<td>Weart</td>
<td>168</td>
<td>18.4</td>
<td>92.8</td>
<td>0.86</td>
<td>0.87</td>
<td>46.0</td>
<td>0.82</td>
<td>0.85</td>
</tr>
</tbody>
</table>

MBE, mean bias error; RMSE, root mean square error.

**Table VI. Performance measures for estimation of potential K↓ via the modified Hock (1999) method**

<table>
<thead>
<tr>
<th>Site</th>
<th>Obs. (Wm⁻²)</th>
<th>MBE (Wm⁻²)</th>
<th>RMSE (Wm⁻²)</th>
<th>Eᵣ²</th>
<th>r²</th>
<th>RMSE (Wm⁻²)</th>
<th>Eᵣ²</th>
<th>r²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bridge</td>
<td>171</td>
<td>0.2</td>
<td>109.4</td>
<td>0.82</td>
<td>0.82</td>
<td>62.1</td>
<td>0.71</td>
<td>0.72</td>
</tr>
<tr>
<td>Helm</td>
<td>157</td>
<td>28.0</td>
<td>122.3</td>
<td>0.74</td>
<td>0.80</td>
<td>71.6</td>
<td>0.59</td>
<td>0.83</td>
</tr>
<tr>
<td>Weart</td>
<td>168</td>
<td>18.2</td>
<td>105.1</td>
<td>0.82</td>
<td>0.85</td>
<td>61.1</td>
<td>0.69</td>
<td>0.73</td>
</tr>
</tbody>
</table>

MBE, mean bias error; RMSE, root mean square error.
missing data and provides for a single point for downscaling numerical weather predictions for short-term hydrologic forecasts.

The UBCWM configuration for Cheakamus has a long-term $E_m$ (Nash and Sutcliffe, 1970) of 0.82, and daily forecasts are produced by hydrologic forecasters to improve on this accuracy even further. However, the model often performs poorly during rain-on-snow events, typically under-predicting streamflow response.

Using NARR to estimate vertical air temperature profiles

High-elevation NARR data were used to modify the standard UBCWM lapse method, in order to predict the elevation of the 0°C isotherm, with as little change to the standard method as possible. We produced a second synthetic station at the Helm Glacier AWS site (NARR-Helm; 2912 m) based on NARR output and applied the following rules:

- When there was no inversion and temperatures were below 0°C at CMS, the UBCWM standard temperature gradient was applied with CMS as the base temperature.
- When temperatures were above 0°C at NARR-Helm, or for inversions, when NARR-Helm was warmer than CMS, the UBCWM standard temperature gradient was applied with NARR-Helm as the base temperature.
- For temperatures above 0°C at CMS and below 0°C at NARR-Helm, the standard UBCWM temperature gradient was calculated with a linear interpolation between the two synthetic stations.

Following the estimation of the 0°C isotherms using NARR intelligence, we investigated the potential for NARR to assist in providing a more realistic vertical temperature gradient during an AR event from January 2010 in the Cheakamus watershed. Following Moore (1993), we adjusted the CMS input temperatures (maximum and minimum daily) for the month of January 2010 to cause the UBCWM 0°C isotherms to match the 0°C isotherms predicted by the combination of CMS and the NARR-Helm station. When the two 0°C isotherms were predicted to be greater than 200 m apart, we back-calculated the temperature that would be required at CMS to reproduce the elevation from the NARR/CMS combination. We then ran the UBCWM, without any parameter adjustment or recalibration, for the Cheakamus basin to see how predicted reservoir inflows were altered when the UBCWM was operated using adjusted temperatures for January 2010.

Finally, we calculated the vertical temperature gradients, for the daily maximum temperature, for select times (before, during and after the peak flows were observed in the Cheakamus catchment for the January 2010 event) using the standard UBCWM method and the UBCWM method with temperatures at CMS adjusted to match the 0°C isotherm predicted with intelligence from the NARR-Helm station, as described above. We then extracted the full air temperature profile for the NARR grid point nearest to the centre of the Cheakamus catchment (shown on the southwest of the Cheakamus catchment boundary in Figure 1 during each of the selected days in order to compare temperature profiles produced by NARR output to the two alternative states of temperature profiles that were used within the UBCWM framework).

### Table VII. Performance measures for $L_{\downarrow}$ at the Place site for the three tested methods

<table>
<thead>
<tr>
<th>Method</th>
<th>MBE (Wm$^{-2}$)</th>
<th>RMSE (Wm$^{-2}$)</th>
<th>$E_m$</th>
<th>$r^2$</th>
<th>RMSE (Wm$^{-2}$)</th>
<th>$E_m$</th>
<th>$r^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>NARR</td>
<td>−5.7</td>
<td>30.8</td>
<td>0.15</td>
<td>0.35</td>
<td>19.2</td>
<td>0.47</td>
<td>0.54</td>
</tr>
<tr>
<td>Stefan–Boltzmann</td>
<td>4.3</td>
<td>42.5</td>
<td>−0.61</td>
<td>0.01</td>
<td>37.0</td>
<td>−0.97</td>
<td>0.03</td>
</tr>
<tr>
<td>Prata (1996)</td>
<td>−33.2</td>
<td>51.0</td>
<td>−1.32</td>
<td>0.00</td>
<td>46.8</td>
<td>−2.14</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Mean observed $L_{\downarrow}$ at the Place site was 288 Wm$^{-2}$.

MBE, mean bias error; RMSE, root mean square error; NARR, North American Regional Reanalysis.
Results

During the AR event that occurred in the south coast of BC during early January 2010, approximately 235 mm of precipitation fell over the snow-covered Cheakamus basin with unseasonably high temperatures, producing run-off from both rainfall and snowmelt.

The AR event from 7–15 January 2010 would have been significantly under-forecasted if the watershed model had been run without intervention from BC Hydro’s forecasting team (UBCWM unadjusted, panel 2, Figure 12). The standard UBCWM temperature gradient method produced a freezing level between 800 and 1500 m during this event. As a result, the model predicted snow to be falling at the higher elevations of the watershed, when, in reality, only heavy rains and rain-on-snow melt could cause the peak inflow rate of greater than 200 m$^3$/s, based on past experience.

Prior to the AR event, nearly 40 mm of precipitation was recorded on 1 January with little run-off response (Figure 12). At that time, the standard UBCWM and NARR-Helm freezing levels were in general agreement. When the AR event began a few days later, the calculated freezing levels began to diverge, and run-off was significantly under-predicted by the standard model.

The elevation of the 0 °C isotherm predicted by interpolation between the NARR-Helm and BC Hydro’s standard CMS synthetic station data reached a maximum elevation of 2500 m, but the UBCWM standard temperature gradient produced a maximum freezing level of 1500 m (Figure 12). The standard UBCWM method would never have more than 50% of the watershed above freezing, whereas incorporation of intelligence from NARR showed that there may have been more than
75% of the catchment receiving rainfall during the period of heaviest precipitation.

Early in the AR event (7 January), NARR output indicates the presence of an elevated inversion above 1600 m, which is not well represented by the temperature profile assumed by UBCWM. Adjusting the UBCWM temperature profile to match the NARR-predicted freezing level resulted in temperatures in the lower elevations of the catchment that were substantially higher than both the UBCWM and NARR temperatures (Figure 13, upper panel), causing UBCWM to predict apparently excessive snow melt over a large fraction of the catchment and thus to over-predict reservoir inflows in the earlier part of the event. For 11 January, the wettest day of the AR, the NARR-adjusted temperature profile provides a closer match to the temperature profile from the raw NARR elevation output (Figure 13, middle panel), producing an improved simulation of peak streamflow (Figure 12, panel 2). Later in the event, NARR output indicated a nonlinear temperature profile with an overall warmer air column compared with the UBCWM profile (Figure 13, lower panel).

DISCUSSION

Potential for use of NARR output for hydrology

The NARR products are attractive for hydrologic analysis and modelling because they include all the variables required to drive physically based models of snow and ice melt, evapotranspiration and geocryology. Because of the reasonable precision and low bias, air temperature output from NARR appears to have good potential utility for hydrologic applications at high elevations in mountainous regions. High-elevation air temperature intelligence can provide information about precipitation phase, snow melt and evapotranspiration in locations where no surface observations are available. The use of NARR for air temperature can likely provide more useful information than extrapolation methods that assume static vertical temperature profiles, especially in situations with isothermal or inverted temperature profiles. The RMSE at the daily scale for NARR air temperatures (ranging from 0.98 to 1.21°C) compare favourably with the results from Fiddes and Gruber (2014) who found a daily RMSE of 1.95°C for downscaled climate model output in the Swiss Alps.

Vapour pressure from NARR also appears reasonably accurate and could have potential for use in hydrologic applications in mountainous regions. Vapour pressure was relatively unbiased at Place and Weart glaciers but exhibited a positive bias at the Bridge Glacier site and a negative bias at the Helm Glacier site. This indicates that NARR vapour pressure may exhibit a spatially variable bias within a catchment. Despite this, knowledge of vapour pressure where no observations are available could be potentially useful for quantifying the direction and magnitude of latent heat exchange between the surface and the atmosphere, which can be an important source of energy for snow melt during major rain-on-snow events (Marks et al., 1998).

Downscaled wind speed from NARR has a positive bias, consistent with the fact that observed wind speeds are influenced by surface friction and results from other studies (e.g. Stegall and Zhang, 2012; Abatzoglou, 2013). Fiddes and Gruber (2014) found similar RMSE (2.69 ms⁻¹) but lower bias after applying the topographic corrections from Liston and Sturm (1998) to ERA-Interim reanalysis output. The linear relations between downscaled and observed wind speeds, especially for Weart Glacier (Table IV), indicate that a bias correction could potentially be applied to more accurately downscale the NARR output to surface observations. Another possibility is that a wind record from a temporary gauge could be extended by NARR through calibration.

The Bridge Glacier site is located on a minor ridge below the main ridge crest. Consequently, wind speeds and directions are modified by anabatic and katabatic flow, surface friction and steering of winds along the valley axis. This likely explains the low $r^2$ results at the 3-h time step. Wind speed estimates from NARR are likely only useful as an estimator at the highest elevations in this region.

A mono-level 10-m surface wind output is also available from the NARR model. We tested this output as an alternative to the pressure-level winds. While there was less positive bias, likely because of corrections for surface friction, the $r^2$ values were similar to the results for the NARR pressure-level winds for the Bridge site and worse for the Place and Weart sites. Hence, we argue that, because both products would require bias correction, the downscaled pressure-level winds should be preferred because they yielded higher $r^2$ with observed wind speeds.

On most sunny days, incoming short-wave radiation from NARR appears to be reasonably accurate (Figure 2); however, cloud cover often does not appear to be captured well by NARR (e.g., 3, 4 and 7 July in Figure 2). The positive MBE in summer (Table V) may be due to small-scale convective cloud formation that is common on warm days in the southern Coast Mountains, which may not be captured by the NARR model because of its spatial and temporal resolution. NARR downscaled $K_{down}$ outperforms an estimate of potential incoming solar radiation, but not by a wide margin. Considering the limited accuracy of short-wave radiation from NARR, and particularly its bias, we recommend that caution should be exercised when using NARR short-wave
radiation as input to hydrologic analyses and models. The positive bias of incoming solar radiation is consistent with other studies that found positive biases in NARR incoming short-wave radiation (e.g. Schroeder et al., 2009; Kumar and Merwade, 2011).

Accuracy may be further degraded by the fact that incoming short-wave radiation is only available as mono-level output from NARR and hence may not be representative of the atmospheric transmissivity at a high-elevation site. The NARR mono-level surface elevations are more than 500 m lower in elevation than the true site elevation at the Helm and Weart Glacier sites. At the Bridge Glacier site, the NARR mono-level elevation is within 10 m of the AWS elevation (Table I), and the Bridge Glacier site does display a lower MBE than either Helm or Weart (Table V).

Although fewer data were available to evaluate long-wave radiation, it does appear that NARR downscaled L↓ outperforms common techniques for estimating incoming long-wave radiation, including the use of the Stefan–Boltzman law using both a constant emissivity of 0.85 and the Prata (1996) equation. Further evaluation is necessary, especially during winter periods, to make a more definitive conclusion about the utility of NARR downscaled L↓ for hydrology. NARR downscaled L↓ also performed favourably compared with the results from Fiddes and Gruber (2014), who found an RMSE of 29.4 W m⁻² using the Stefan–Boltzman law modified with a cloud cover component.

**Utility of NARR for hydrologic modelling**

For the January 2010 AR event, using NARR output to define vertical air temperature gradients resulted in a higher snowline elevation and a substantially larger amount of the watershed experiencing rainfall than with the standard temperature gradient. This increase in rainfall caused an increase in predicted reservoir inflows during this event. The incorporation of input temperatures that were adjusted to match the NARR interpolated 0°C isotherm predicted peak reservoir inflows that better matched observed inflows, but the inability of the static temperature gradient parameters of the UBCWM to capture inversion or isothermal conditions (even when the CMS temperatures were adjusted) is still a limitation and will remain so unless dynamic temperature gradients can be implemented within the operational modelling framework. The adjustment of temperatures to match elevated freezing levels during inversion or isothermal conditions resulted in temperatures at lower elevations that were higher than in reality and likely resulted in unrealistic snow melt (panel 1, Figure 13).

Given that operational hydrologic models are highly tuned to produce the most accurate forecasts possible with minimal input data and we adjusted the inputs to the UBCWM without any recalibration, the results of this adjustment do appear promising. The over-prediction of reservoir inflows during the falling limb of the peak event (approximately 15 January in Figure 12) appears similar in shape to the original modelled falling limb but transposed upwards. Hence, the results may be improved with calibration during AR events. The rapid onset of melt-generated runoff when temperatures rise above freezing is another issue known to forecasters using the UBCWM and is primarily due to the lack of cold content tracking within UBCWM.

Although NARR’s temporal latency limits its direct use for near-real-time forecasting, NARR output could be valuable for assisting in model calibration, particularly for the parameters that control run-off generation during rain-on-snow events. Alternatively, switching to an ‘AR’ mode, in which temperature profiles are temporarily adjusted or perhaps set to isothermal, as suggested by Rössler et al. (2014), may assist operational forecasters in producing better run-off forecasts when these meteorologic conditions are expected. In the case of UBCWM, some recoding would be required to incorporate this functionality.

Given the errors in prediction shown in this study, the use of NARR output directly to drive energy balance snow models (through the use of wind speed, vapour pressure and radiation) may not generate accurate predictions, especially during rain-on-snow events. Considering the errors associated with most of the NARR variables, particularly short-wave and long-wave radiation, it is unclear whether their use as input to an energy balance snowmelt model would provide superior performance compared with a temperature index-based model. It would be a useful exercise to compare the relative performance of different complexities of snow models driven by NARR output.

**CONCLUSIONS**

The comparison of NARR output with observations from four high-elevation sites in the southern Coast Mountains of BC demonstrates that air temperature and vapour pressure are predicted with a high level of skill. The air temperature fields, in particular, appear to be of sufficient accuracy to assist in predicting storm freezing levels and are likely to be of value in driving temperature index snowmelt models. Wind speeds at the ridge crest sites were consistently lower than predicted by NARR, as expected, but the reasonable correlation at some sites suggests that the NARR values have potential value as indices, subject to bias correction. Short-wave radiation was predicted more poorly by NARR and may not provide enough of an improvement over standard models of potential incoming short-wave radiation.
to be of use in hydrologic applications. Long-wave radiation requires further testing during winter months, although it does provide increased accuracy over commonly used estimation methods.

Because of a lack of observations, we were not able to test the accuracy of NARR for estimating surface precipitation. Future work should assess the accuracy of NARR precipitation compared with surface observations in mountainous areas. However, testing NARR precipitation output would be more challenging than for variables such as air temperature because (1) precipitation can exhibit greater sub-grid horizontal variability, posing greater uncertainty in downscaling from gridded values to points, and (2) measured precipitation at high-elevation sites can be subject to considerable undercatch because of high wind speed, especially for snow.

Future work should focus on optimal methods for assimilating NARR air temperature into a forecast system. As operational hydrologic models are highly tuned to watersheds, the most effective method for assimilation of NARR output will likely require extensive testing in order to improve run-off forecasts during nonstandard situations while maintaining stability during standard forecasting situations. Although NARR’s temporal availability would not allow it to be used in place of high-altitude observations during real-time forecasting, assimilation and use of NARR air temperature for hindcasting of historic rain-on-snow events could provide guidance for model formulation and calibration and be used to help determine the most useful locations for new AWS installation.

ACKNOWLEDGEMENTS

This research was supported through funding from the Natural Sciences and Engineering Research Council (NSERC) through a discovery grant to R.D. Moore, a PGS-A scholarship to J.M. Shea and an Industrial Postgraduate Scholarship to J.W. Trubilowicz, BC Hydro, the Mitacs Accelerate internship programme and the Canadian Foundation for Climate and Atmospheric Science through its support of the Western Canadian Cryospheric Network. Natalie Stafl, Roger Hudson and Ellen Morgan assisted in field data collection. We would like to thank two anonymous reviewers for their constructive feedback.

REFERENCES
