Taylor & Francis
Taylor & Francis Group

Hydrological Sciences Journal

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/thsj20

Effects of quality controlled measured and re-analysed meteorological data on the performance of water temperature simulations

Amir Sadeghian, Jeff Hudson & Karl-Erich Lindenschmidt

To cite this article: Amir Sadeghian, Jeff Hudson & Karl-Erich Lindenschmidt (2021): Effects of quality controlled measured and re-analysed meteorological data on the performance of water temperature simulations, Hydrological Sciences Journal, DOI: 10.1080/02626667.2021.1994975

To link to this article: https://doi.org/10.1080/02626667.2021.1994975

Published online: 09 Dec 2021.

HYDROLOGICAL SCIENCES JOURNAL
https://doi.org/10.1080/02626667.2021.1994975

Effects of quality controlled measured and re-analysed meteorological data on the performance of water temperature simulations

Amir Sadeghian[a], Jeff Hudson[b] and Karl-Erich Lindenschmidt[a]

[a]Global Institute for Water Security and School of Environment and Sustainability, University of Saskatchewan, Saskatoon, SK, Canada; [b]Global Institute for Water Security and Department of Biology, University of Saskatchewan, Saskatoon, SK, Canada

ABSTRACT
One of the most prominent sources of error and uncertainty in water quality modelling results is the input data. In this study, data from three meteorological databases were used to test the performance of a water temperature model of Lake Diefenbaker: the data from Environment and Climate Change Canada (ECCC) had long-term quality control history (>20 years); the data from the AccuWeather had short-term quality control history (<10 years), and the data from the MeteoBlue database were modelled values. The CE-QUAL-W2 hydrodynamic and water quality model was used for this study. The model was calibrated by adjusting model coefficients controlling the amounts of measured solar radiation and wind that reach the surface of the water. The sensitivity results showed very similar performances, with slightly better performances (root mean square root difference of ± 0.1) with the ECCC data followed by the MeteoBlue data and thereafter by the AccuWeather data.

ARTICLE HISTORY
Received 17 July 2018
Accepted 16 September 2021

EDITOR
A. Castellarin

GUEST EDITOR
C. Cudennec

KEYWORDS
Lake Diefenbaker; hydrodynamic modelling; meteorological data; model calibration

Introduction

Meteorological databases are one of the most important components in many environmental studies, particularly in modelling exercises (Hunt *et al.* 1998, Sheffield *et al.* 2006). They serve as the driving force on model boundaries – air temperature, humidity, wind speed and direction, and solar radiation to name a few. Model outputs can be significantly affected by the quality and spatiotemporal resolution of these datasets (Yang *et al.* 2006, Haberlandt 2007). The reliance on meteorological databases has prompted many federal organizations, e.g. Environment and Climate Change Canada (ECCC), the National Oceanic and Atmospheric Administration (NOAA), and the Swedish Meteorological and Hydrological Institute (SMHI), to introduce strict standards on how data are acquired, controlled and archived to assure data quality and continuity. The ECCC has been collecting meteorological data since 1840 and continuously performs quality control investigations to guarantee high-quality and accurate data (ECCC 2017). The NOAA Quality Controlled Local Climatological Data (QCLCD) project provides hourly, daily, and monthly data for about 1600 US meteorological stations beginning in 2005 (NOAA 2017). SMHI checks observation values with 6-hour predictions and believes that 90% of the errors are captured at this stage (Vejen *et al.* 2002). An investigation of 10 meteorological variables in 726 stations in China showed less than 0.05% inconsistency in data (due to typing errors or incorrect data archiving [e.g. units]), temporally and spatially, for the 1951–2000 study period (Feng *et al.* 2004). As a result, the operational costs for meteorological station installation and maintenance remain very high, limiting the number of available stations, especially those with a long-term quality control such as those mentioned above.

Less than 1% of the globe is covered with measured meteorological data from sparse weather stations, where the data from each station are useful up to a maximum radius of 3 to 12 km (MeteoBlue 2016). These weather stations are also biased towards land surfaces and are unevenly distributed within vast areas, with some areas not having any weather stations at all (Tabios and Salas 1985, Haberlandt 2007). A station close to the study site can provide very high-quality data; otherwise, the data should be used with caution. Another problem arises when the stations do not measure all the variables required for the modelling study. In these cases, data for unmeasured variables can be estimated, for example by using empirical equations (Sentelhas *et al.* 2010). A variable commonly missing in many (e.g. ECCC) stations is sky/cloud cover data (Kassianov *et al.* 2005, Sadeghian *et al.* 2015).

Cloud cover has a substantial impact on the significance of shortwave and longwave radiation that reaches the land and water surface (Cazorla *et al.* 2008). Solar radiation is a critical component for photosynthesis (Aguilera *et al.* 1999, Yamashita *et al.* 2004), hence also for studying eutrophication (Strickland 1958, Goldman 1988, Assemany *et al.* 2015). There are many well-established methods for calculating direct and indirect solar radiation based on location, time of year, air temperature and air moisture, and cloud cover (e.g. De Jong and Stewart 1993, Annear and Wells 2007). However, empirical relations to calculate cloud cover are uncertain and can influence model outcomes (Souza-Echer *et al.* 2006, Cazorla *et al.* 2008). Hence,

CONTACT Amir Sadeghian amir.sadeghian@usask.ca Global Institute for Water Security and School of Environment and Sustainability, University of Saskatchewan, Saskatoon, SK, Canada

generally, all the meteorological variables required for a eutrophication modelling exercise are available, except for cloud cover. A measure for cloud cover would be most beneficial, but is the one parameter that is often missing in a sampling programme.

Measuring meteorological data is accomplished by large national institutions (e.g. ECCC, SMHI, USGS) and small national/private regional institutes that study local climate (Changnon *et al.* 1990, Craft 1999). There are also many large private institutions involved in collecting data from these sources and making universal databases. AccuWeather, with headquarters in the USA, is an excellent example of a database comprised of climate data from numerous sources. These sources range from national organizations such as ECCC and Agriculture and Agri-Food Canada to private local institutions as well as AccuWeather meteorological stations. The company provides hourly measurements and forecasted meteorological data with high accuracy for millions of users worldwide (AccuWeather 2016). Besides direct measurements, some companies forecast and calculate these data using weather models, each using different modelling approaches to simulate the data and incorporate land surface characteristics such as topography, ground cover and surface cover. MeteoBlue, located in Switzerland, is a good example, which since 1984 has provided data on 45 meteorological variables for any location with a resolution of 3–30 km.

In this study, we compare the outputs of a water temperature model using meteorological data from three sources: ECCC, whose data are highly accurate; AccuWeather, which has less than 10 years of quality-controlled data (for the study area used here); and MeteoBlue, whose data stem from model simulations. The primary objective was to quantify the effects of sources of meteorological data on modelled water temperature and, ultimately, water quality results. Based on these results, we can provide guidelines for modellers and researchers for selecting the most suitable climate database. The novelty of this research is classifying climate databases by looking into the effects of the data in each database on a coupled river and reservoir system using a sophisticated physically based two-dimensional (2D) water quality model in contrast to statistical methods and rainfall–runoff models.

The secondary objective in performing this comparative study was to remind the researchers of the possibility of adopting alternative sources and to concentrate on the main objective of the study (in our case building the water quality model) by trusting the work of other scientific groups (meteorology groups in the current case).

Methods

The study area selected for this comparison study is the combined South Saskatchewan River (SSR), Red Deer River (RDR), and Lake Diefenbaker region located in the Canadian prairies. The SSR and RDR are long rivers that originate in the Rocky Mountains and flow through Alberta and Saskatchewan. In Saskatchewan, they merge and flow into Lake Diefenbaker reservoir, 181 km long and 60 m deep, formed by the creation of the Gardiner Dam and Qu'Appelle Dam over the SSR in the 1960s.

A model was required to work with a coupled system of two long rivers and a reservoir with varying depths from a few metres up to 60 m. Hence, a model with at least two dimensions (with averaging over the lateral direction) was required to capture the correct hydrodynamics and variations in the longitudinal and vertical directions. Because of the rivers' and reservoir's narrow widths, the gradients of importance are expected to be in the vertical and horizontal directions, and the third dimension (lateral direction) could be averaged. We used a 2D surface water quality model to evaluate the outcomes of using different meteorological databases. The CE-QUAL-W2 model (Cole and Wells 2015), developed by the US Army Corps of Engineers and maintained by Portland State University, was selected for this study. The CE-QUAL-W2 model has over 40 years of development and has been successfully used for studying hydrodynamic, temperature, nutrient, and sediment transport characteristics of rivers, lakes, and estuaries. The model has also been successfully used on Canadian Prairie reservoirs (e.g. Sadeghian *et al.* 2015, 2017a, 2017b, 2018, Terry *et al.* 2017, 2018, Mi *et al.* 2019).

This study is an extension of a previous study on which the Lake Diefenbaker temperature model was calibrated for the period 2011–2013 (Sadeghian *et al.* 2015), and a basis for parallel studies on effects of climate change (Morales-Marin *et al.* 2021, Akomeah *et al.* 2021), and on downstream macroinvertebrate communities (Carr *et al.* 2019, 2020). In that study, the temperature model was calibrated using four model parameters: light extinction coefficient, wind sheltering coefficient (WSC), solar radiation shading coefficient (SHADE), and inflow water temperature adjustment coefficient. The calibration was done using two methods; with 1000 Monte Carlo runs, and by the combined global–local optimization (PSO + LM) method. Based on temperature model calibration and global sensitivity analysis, the results of temperature simulations had the highest sensitivities with the WSC and the SHADE, which adjust the amount of wind and solar radiation, respectively, applied to the surface of the water. The WSC is the percentage of recorded wind speed from a land station near a water body that is transferred to the water surface. The wind becomes slower when it passes through barriers such as trees and topographic peaks near the edge of the water, and becomes faster when blowing over a lake with a long fetch. Unfortunately, all the meteorological stations used in this study were land stations (located offshore), hence, the use of a WSC was necessary for calibration.

The coefficient SHADE is the amount of shortwave solar radiation that impinges the water surface. In CE-QUAL-W2, 100% SHADE means a clear sky and 0% SHADE means a fully cloudy sky (+ topographic and vegetative). Similar to WSC, SHADE is reduced by vegetation and topographic barriers along the shore. The CE-QUAL-W2 provides a good set of tools to consider the topography around the water body. However, the SSR and Lake Diefenbaker are located in the Canadian prairies, which has a very flat topography. The study area has such a flat topography that most of the watershed areas do not contribute to Lake Diefenbaker and the SSR (Pomeroy *et al.* 2009, Toth *et al.* 2009). As mentioned by Sadeghian *et al.* (2015), about 98% of the inflow to the reservoir originates from the Rocky Mountains in Alberta. Hence, in this model, the shading coefficient was used to account for missing cloud cover, vegetation, and topography collectively.

Sadeghian *et al.* (2015) found that the inflow water temperature, which was calculated from the average weekly air temperature, is the primary source of thermal energy to the reservoir. Lake Diefenbaker is about 181 km long and its boundary starts from Saskatchewan Highway #4. The closet hydrometric stations that measure the water temperature are located in Alberta at Medicine Hat on the SSR (374 river km to Highway #4) and Bindloss on the RDR (218 river km to Highway #4). Hence, the model boundaries were extended to Medicine Hat and Bindloss, where hydrometric stations are located and the water temperature is measured frequently (Fig. 1). Comparing the modelled water temperatures with the few recorded water temperature values at the lake's inlet confirms that the effects of ambient air temperature, solar radiation, and the travel times between the stations and the lake's inlet (1–3 d based on discharge rate) are enough to remove the effects of initial boundary conditions (water temperature) by the time the water arrives at Lake Diefenbaker (Highway #4).

In the model, the rivers and the reservoir are divided into a total of nine interconnected water bodies, as shown by the black lines in Fig. 1. The model reads meteorological data for each water body separately, allowing the use of nine meteorological stations along the river and reservoir. These nine water bodies were selected based on changes in topography, morphology, and climatology. For the water bodies where a meteorological station was not available, the station from a neighbouring water body was used (Table 1). The state variables for meteorological data are air temperature, dew point temperature, wind speed, wind direction, cloud cover, and shortwave solar radiation (Cole and Wells 2015). Although the model can be used regardless of the frequencies of the input variables, hourly or daily values must be used for accurate results. We used hourly data for all the meteorological variables and daily flow data at the inlet, and at the Gardiner

Table 1. List of stations along the South Saskatchewan River and Lake Diefenbaker with hourly meteorological data for the study period (2011–2013) available in the databases. ECCC: Environment and Climate Change Canada; WB: water body.

Stations	AccuWeather	ECCC	MeteoBlue
WB1	Medicine Hat	Medicine Hat	Medicine Hat
WB2			Twin Peaks
WB3	Burstall	Leader Airport	McNeill
WB4	Estuary		Estuary
WB5			Cramersburg
WB6	Beaver Flat		Saskatchewan Landing
WB7		Lucky Lake	Lake Diefenbaker
WB8	Riverhurst		Riverhurst
WB9	Elbow	Elbow	Elbow

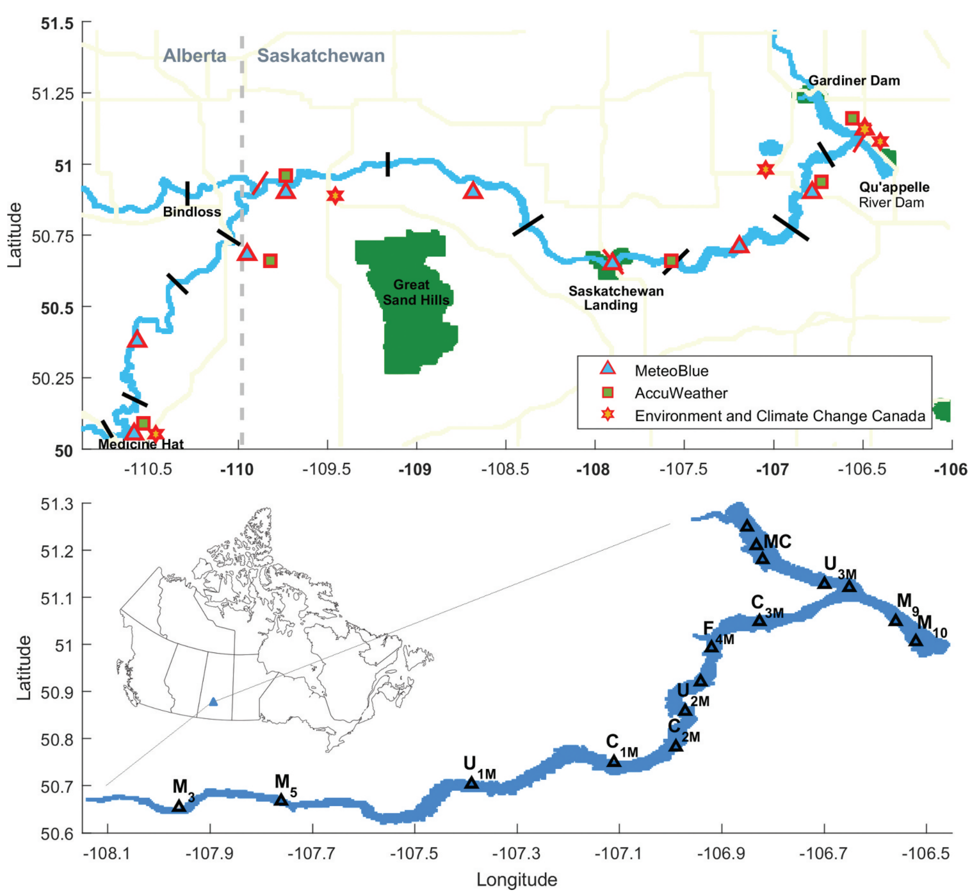

Figure 1. Map of the South Saskatchewan River (SSR), Red Deer River (RDR) and Lake Diefenbaker. The meteorological stations for the three databases are shown in the map on top. The map on the bottom shows the locations of water temperature sampling stations used in model calibration.

and Qu'Appelle dams. Model simulations were started 1 April 2011, assuming isothermal conditions to the reservoir after ice melt and spring turnover, until 31 December 2013.

We ran 160 Monte Carlo runs for each meteorological database (ECCC, AccuWeather, and MeteoBlue) for a total of 480 runs on the University of Saskatchewan High performance computing (HPC)research cluster. There are a total of 96 nodes, each of which has 16 processors, on the research cluster, giving a total of 1536 processors. To use the computing resources more efficiently, the submitted task should be a multiple of 16 (the number of processors in each node). Hence, 160 runs (10 nodes) were used for each station. The 160 random values were the same in all set-ups, with a random selection of two variables – WSC and SHADE uniformity – distributed between the defined ranges. The model outputs were compared with the measured temperature profiles at 16 stations along the reservoir (Hudson and Vandergucht 2015, Sadeghian *et al.* 2015). The root mean square error (RMSE) was used as the metric for model performance:

$$RMSE = \sqrt{\frac{\sum (O - S)^2}{n}} \tag{1}$$

where O is the value of observation in one column, S is the corresponding simulated value, and n is the number of samples. The nine meteorological data files for the water bodies in the model were prepared based on the description presented below.

ECCC database

ECCC has a total of 8732 meteorological stations, as reported on its website. Of these stations, 1441 stations are listed more than once mainly because renovated stations are listed with a new station ID; hence, 7291 stations remain when duplicates are removed (Table 2). The number of working stations in 2013 was a small subset of this, about 1500 stations. Only 59 stations with hourly data intervals are located in Saskatchewan, which is very sparse considering the large size of the province (Table 2). Fifty-nine stations is equivalent to almost one station per 11 035 km^2. However, the quality of the data from these stations is high due to long-term quality control history. For our study area, the accuracy of measured data were reviewed using over 20 years of historical data with outliers and errors removed.

Another key limitation is the lack of cloud cover data. Cloud cover data are critical for calculating shortwave and longwave solar radiation that reaches the water and land surfaces. The model uses one meteorological station for each water body; hence, nine meteorological stations are required. However, there are only four stations within the proximity of the SSR/Lake Diefenbaker region in the ECCC database. Thus, these four stations were also used for the remaining water bodies, according to their proximities.

AccuWeather database

Different meteorological variables in the AccuWeather database stem from different sources, which could be based on a single station, or a combination of several nearby stations owned by governmental or private agencies, or their own stations. If a weather station continues reporting for a long period of time (e.g. 10 years) such that all the metrics can be validated for the time frame for which the data are requested, it is designated a "primary" station; otherwise, it is called "secondary." Thus, the secondary stations are meteorological stations that are missing a robust predicting standard from which to estimate the missing values in case of a device failure. Because all the metrics cannot be validated due to gaps in the data, these stations are also referred to as backup stations by AccuWeather. Most of the AccuWeather meteorological stations used in this study were among the secondary stations; hence their data did not pass any quality control screening assurance. Among the nine water bodies in the model, there were only six stations within the proximity of the SSR/Lake Diefenbaker region in AccuWeather's database. The remaining water bodies used one of these stations according to their proximity.

MeteoBlue database

MeteoBlue calculates and forecasts meteorological data using its own model based on Nonhydrostatic Meso-Scale Modelling (NMM) technology. Different modelling approaches are used

Table 2. Coverage and density of Environment and Climate Change Canada (ECCC) stations per province. The total number of stations is listed as 8732. However, after removing duplicates, 7291 stations remain. Station information downloaded from the ECCC website on 24 July 2016.

Name	Area (km2)		Number of stations			Active stations in 2013		One station for (km2)
	Total	Land	Total	Daily	Hourly	Daily	Hourly	
Alberta	661 848	642 317	1378	1331	294	262	262	2526
British Columbia	944 735	925 186	1653	1633	249	297	152	6215
Manitoba	647 797	553 556	512	507	81	95	57	11 365
New Brunswick	72 908	71 450	192	192	33	32	27	2700
Newfoundland and Labrador	405 212	373 872	280	264	84	69	51	7945
Northwest Territories	1 346 106	1 183 085	141	130	87	60	63	21 367
Nova Scotia	55 284	53 338	278	260	72	50	43	1286
Nunavut	2 093 190	1 936 113	176	149	144	78	84	24 919
Ontario	1 076 395	917 741	1498	1421	256	216	180	5980
Prince Edward Island	5660	5660	43	42	11	17	9	629
Quebec	1 542 056	1 365 128	987	928	209	237	139	11 094
Saskatchewan	651 036	591 670	659	651	85	103	59	11 035
Yukon	482 443	474 391	124	120	37	32	26	18 556
Sum	9 984 670	9 093 507	7921	7628	1642	1548	1152	8667

Table 3. Comparison of measured data (in general) from the meteorological stations from all over the world and MeteoBlue re-analyses data.

	MeteoBlue	Measurement
Spatial resolution	3–30 km	<1 km
Worldwide coverage	100%	<1%
Number of parameters	45	<10
Number of years	30	2–30 (with gaps)
Time intervals	Hourly, 3 hourly, daily	(Hourly), 3 hourly, daily
Completeness (no data)	100%	10–99%
Consistency	100%	Variable

Source: MeteoBlue (2016).

to reflect detailed topography, ground cover, and surface cover characteristics. The company has provided data for 45 meteorological parameters with a spatial resolution of 3–30 km and a temporal resolution of one hour (hourly data) worldwide since 1984, without any gaps or missing data (MeteoBlue 2016) (Table 3). The reader is referred to <http://content.meteoblue.com/en/verified-quality/verification> for verification and accuracy control information. Since MeteoBlue data are modelled, all the water bodies were assigned a station in the CE-QUAL-W2 model.

Results and discussion

In many cases, the data in ECCC and AccuWeather databases have similar values for the meteorological stations at the same location by a water body [e.g. at Medicine Hat and Elbow (Table 1)]. Hence, the ECCC stations are among the main sources in the AccuWeather database for available variables. For example, the values recorded for the station at Elbow (water body 9) have almost the exact same statistics for ECCC and AccuWeather (Table 4) for the available variables. The ECCC has all the variables required as input data for the CE-QUAL-W2 model except for cloud cover and shortwave solar radiation data. The shortwave solar radiation can be calculated internally by the model using the latitude of the water body; hence, the cloud cover is really the only missing data. MeteoBlue, on average, overestimates air temperature by 2°C, dew point temperature by 1.6°C, wind speed by 1.1 m/s, and solar radiation by 61.9 Wm^{-2}, and underestimates the cloud cover about 23% (2.3 out of 10) (Table 4 and Figs 2–7) compared with ECCC and AccuWeather. In terms of annual peaks, MeteoBlue has a maximum air temperature of 36.4°C (about 2.5°C warmer than of ECCC and AccuWeather), a maximum wind speed of 23.2 m/s (about 5.8 m/s faster than of ECCC and AccuWeather), and a maximum shortwave radiation of 968 Wm^{-2} (about 94 Wm^{-2} higher than of ECCC and AccuWeather). Hence, the model is predicted to get more heat and show stronger mixing (by wind) using MeteoBlue data compared with ECCC and AccuWeather data.

The sensitivity analysis results show that all three databases were able to produce acceptable model performances using the 160 Monte Carlo simulations, and using the two calibrating coefficients (WSC and SHADE) (Table 5 and Figs 8 and 9). In terms of model performance, the RMSE has optimum values of 1.40, 1.08, and 1.13 with AccuWeather; 1.27, 0.91, and 1.22 with ECCC, and 1.21, 1.12, and 1.20 with MeteoBlue for 2011, 2012, and 2013, respectively. Hence, the best performance was obtained using MeteoBlue in 2011, ECCC in 2012, and AccuWeather in 2013.

It was expected that the ECCC data would yield the lowest performance, followed by MeteoBlue. The main reason was that ECCC does not have data in close proximity to the study region compared with the other two databases. Although ECCC has fewer stations (four), and these stations are several kilometres away from our research sites, the flat topography around Lake Diefenbaker meant that the distance between the stations and sites did not lead to large errors. Second, the lack of cloud cover data could successfully be compensated for using the SHADE coefficient. Similarly, MeteoBlue overestimated almost all the variables compared with measured values in ECCC and AccuWeather databases. However, calibrating SHADE and WSC coefficients seems to correct the values imposed on the lake's surface.

It is worth mentioning that the SHADE coefficient only compensates for the effect of cloud cover on shortwave solar radiation and does not affect the longwave radiation calculations. However, in the case of Lake Diefenbaker, a large portion of thermal energy comes from the SSR, according to Sadeghian *et al.* (2015). In that study, the proportion of the

Table 4. Statistics on the meteorological variables in the three databases for the station at Elbow (water body 9) for 2011–2013. ECCC: Environment and Climate Change Canada; TAIR: air temperature (°C); TDEW: dew point temperature (°C); PHI: wind direction (rad); CLOUD: cloud cover (0–10); Solar: shortwave solar radiation (Wm^2 s^{-1}).

		AccuWeather			ECCC			MeteoBlue		
		2011	2012	2013	2011	2012	2013	2011	2012	2013
TAIR	Min	−35.0	−33.9	−34.4	−35.1	−34.4	−34.6	−30.6	−27.9	−34.1
	Max	32.8	32.8	33.9	33.0	32.7	34.0	35.4	35.3	36.4
	Average	3.4	3.9	2.0	3.4	3.9	2.0	5.1	5.9	4.1
TDEW	Min	−38.9	−37.8	−38.9	−38.9	−38.4	−38.7	−33.8	−31.1	−37.6
	Max	22.8	22.2	20.6	23.2	22.0	20.3	20.5	22.0	18.7
	Average	−2.1	−1.8	−3.3	−2.1	−1.8	−3.3	−0.6	0.1	−1.9
WIND	Min	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Max	14.8	16.5	17.4	15.0	16.4	17.5	23.2	20.3	20.7
	Average	4.6	4.5	4.5	4.6	4.5	4.5	5.7	5.7	5.5
PHI	Min	0.0	0.0	0.0	0.2	0.2	0.2	0.0	0.0	0.0
	Max	6.3	6.3	6.5	6.3	6.3	6.3	6.3	6.3	6.3
	Average	3.5	3.5	3.4	3.5	3.5	3.4	3.6	3.6	3.5
CLOUD	Min	0.0	0.0	0.0				0.0	0.0	0.0
	Max	10.0	10.0	10.0				10.0	10.0	10.0
	Average	6.6	6.8	7.0				4.2	4.3	4.8
Solar	Min	0.0	0.0	0.0				0.0	0.0	0.0
	Max	871.0	874.0	855.0				968.0	867.8	869.5
	Average	114.5	111.4	110.2				189.1	162.6	170.0

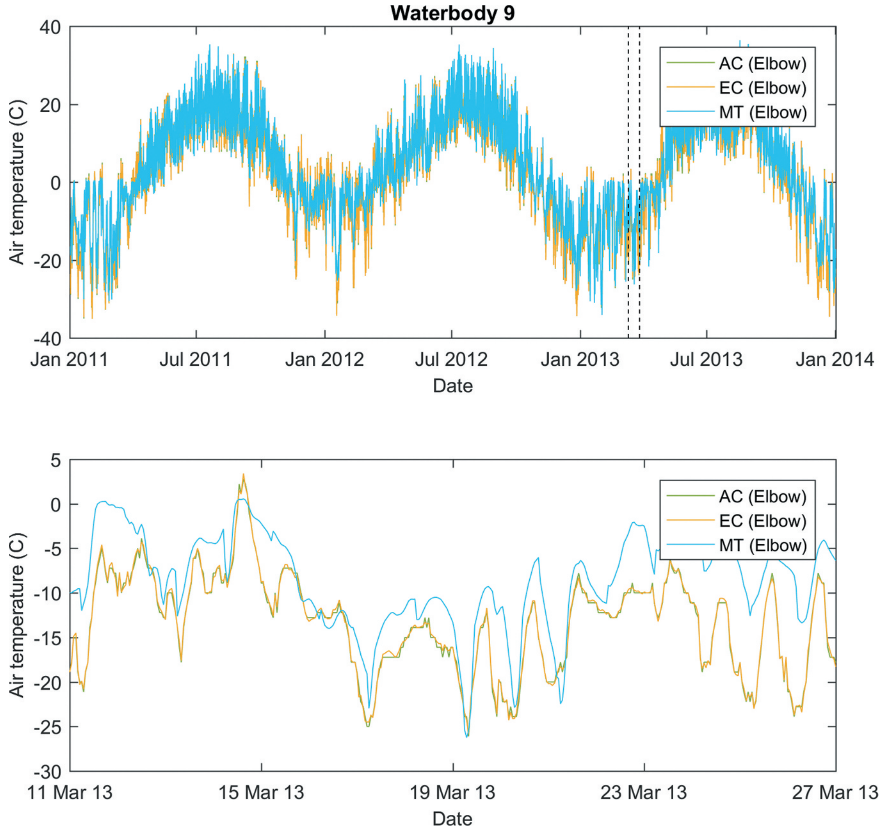

Figure 2. Time series comparison of air temperature in the three databases for the station at Elbow (water body 9) for 2011–2013. The plot at the bottom is enlarged for the selected period. AC: AccuWeather; EC: Environment and Climate Change Canada; MT: MeteoBlue.

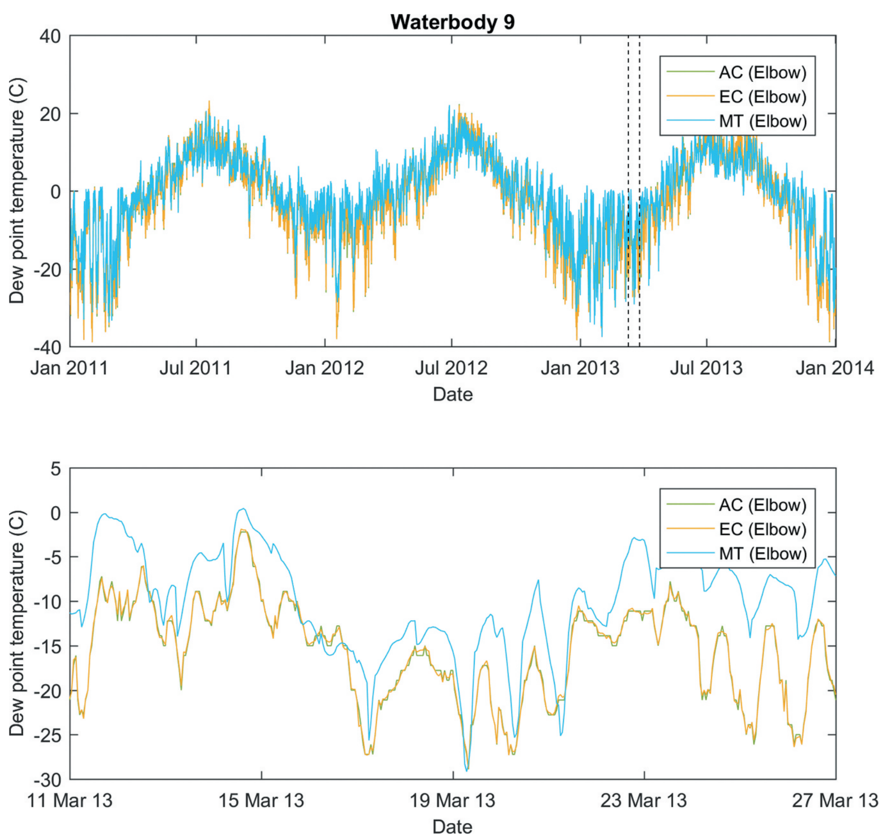

Figure 3. Time series comparison of dew point temperature in the three databases for the station at Elbow (water body 9) for 2011–2013. The plot at the bottom is enlarged for the selected period. AC: AccuWeather; EC: Environment and Climate Change Canada; MT: MeteoBlue.

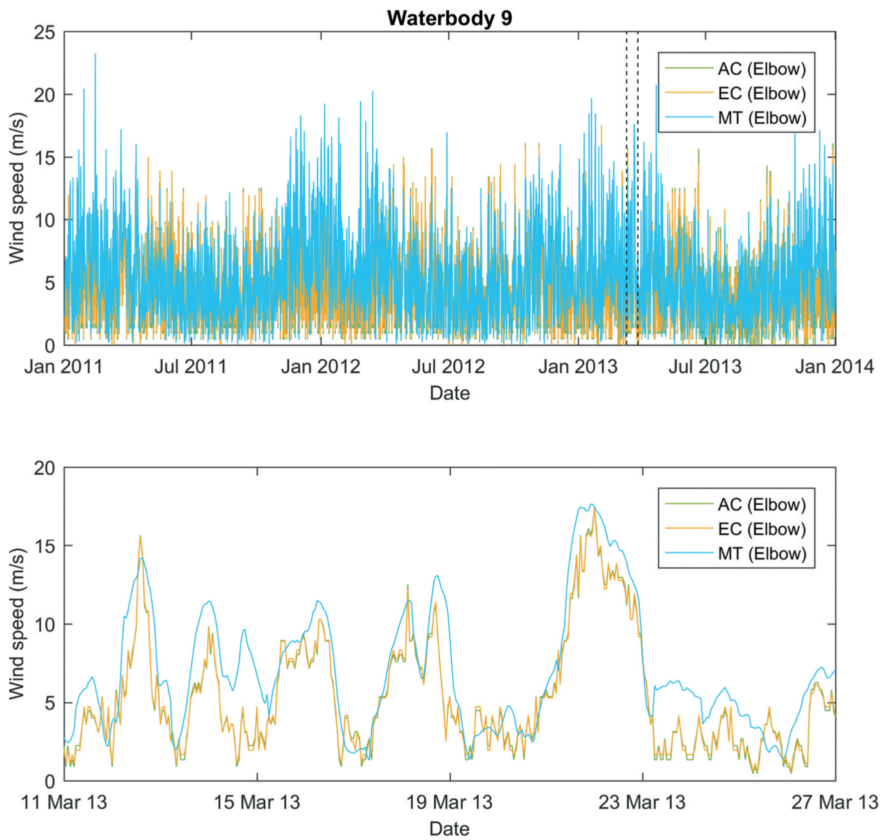

Figure 4. Time series comparison of wind speed in the three databases for the station at Elbow (water body 9) for 2011–2013. The plot at the bottom is enlarged for the selected period. AC: AccuWeather; EC: Environment and Climate Change Canada; MT: MeteoBlue.

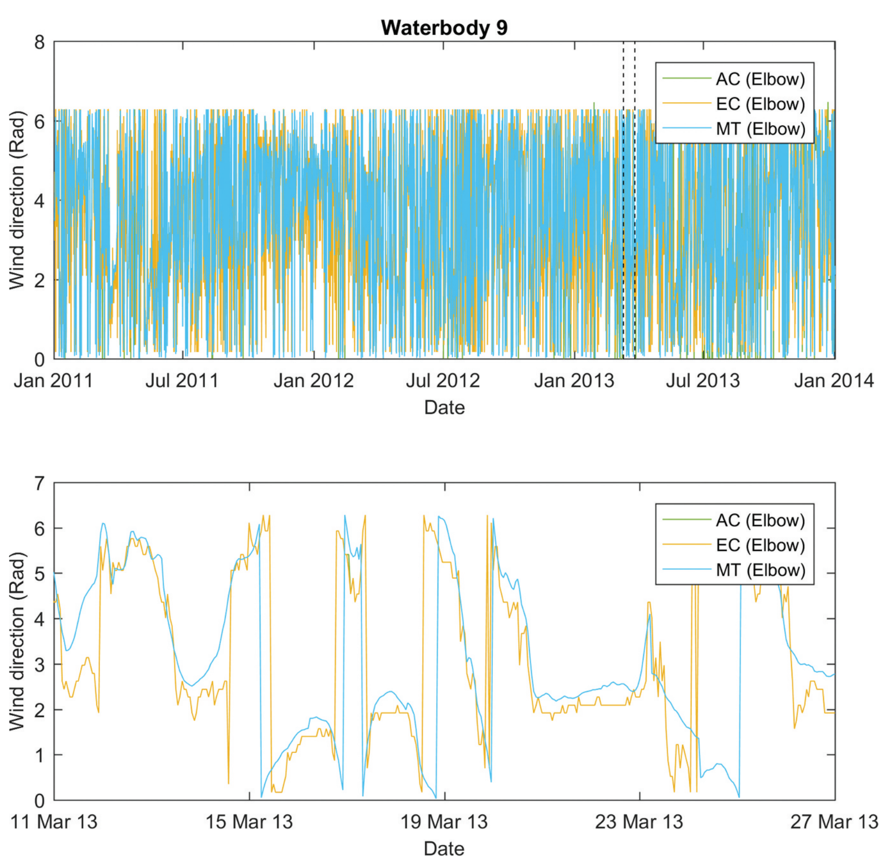

Figure 5. Time series comparison of wind direction in the three databases for the station at Elbow (water body 9) for 2011–2013. The plot at the bottom is enlarged for the selected period. AC: AccuWeather; EC: Environment and Climate Change Canada; MT: MeteoBlue.

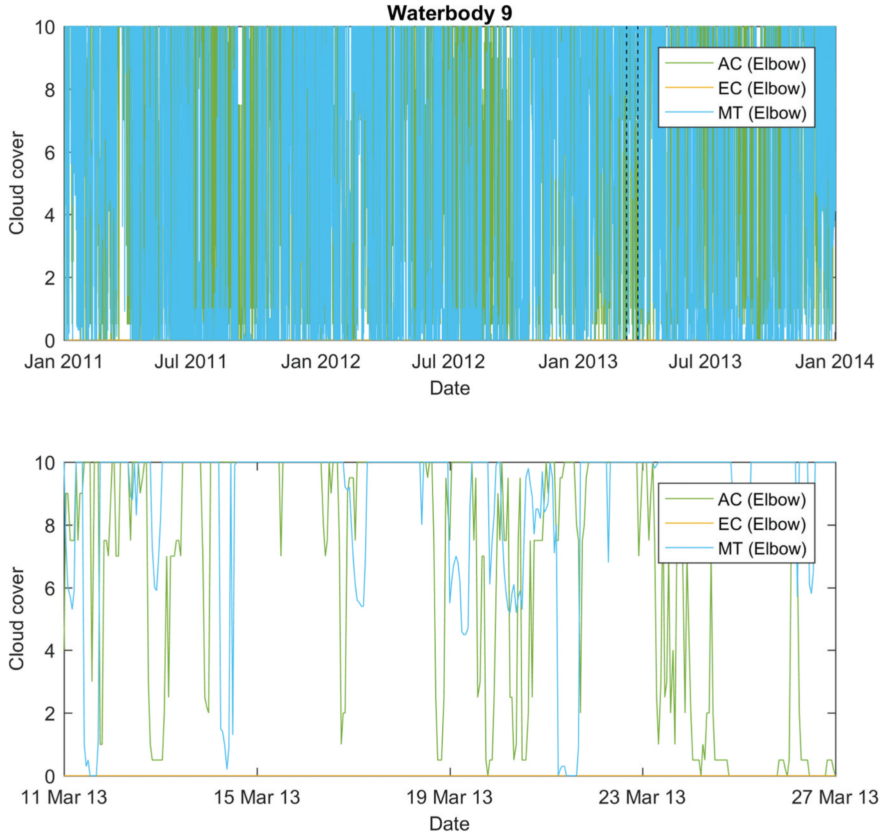

Figure 6. Time series comparison of cloud cover in the three databases for the station at Elbow (water body 9) for 2011–2013. The plot at the bottom is enlarged for the selected period. AC: AccuWeather; EC: Environment and Climate Change Canada; MT: MeteoBlue.

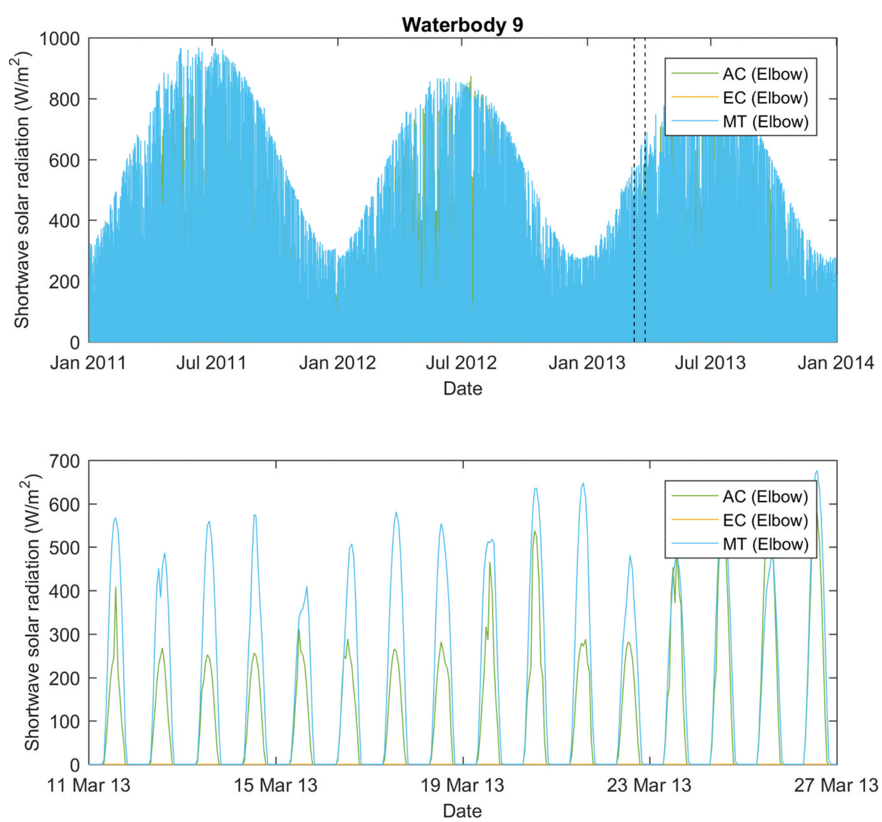

Figure 7. Time series comparison of shortwave solar radiation in the three databases for the station at Elbow (water body 9) for 2011–2013. The plot at the bottom is enlarged for the selected period. AC: AccuWeather; EC: Environment and Climate Change Canada; MT: MeteoBlue.

Table 5. Model performance results based on root mean square error (RMSE) for different stations during 2011–2013. ECCC: Environment and Climate Change Canada.

Year	AccuWeather	ECCC	MeteoBlue
2011	1.40	1.27	**1.21**
2012	1.08	**0.91**	1.12
2013	**1.13**	1.22	1.20

thermal budget from each thermal source to the lake was quantified; it showed a considerable portion of the thermal budget is the thermal energy from the inflow water. Hence, although the SHADE does not affect the longwave radiation, our models were able to perform well due to the hydrological characteristics of the prairies, with many clear-sky days (i.e. shortwave radiation dominating the longwave radiation).

The three databases produced very good model performances, but with slightly different parameter settings. Both the WSC and SHADE parameters show sensitivity (RMSEs between 1 and 5°C) to the databases, which varies for different years. The WSC has optimum values of 0.72, 0.92, and 0.91 with AccuWeather, 0.86, 1.00, and 0.96 with ECCC, and 0.88, 1.09, and 0.97 with MeteoBlue for 2011, 2012, and 2013, respectively. The SHADE has optimum performances with 0.90, 1.00, and 0.92 with AccuWeather, 0.63, 0.71, and 0.60 with ECCC, and 0.61, 0.82, and 0.65 with MeteoBlue for the same years. It is worth mentioning that a value of 1 (100%) for these coefficients means that there is no need for adjustment. Thus, AccuWeather data for WSC (wind) provide the best results without any adjustment, while MeteoBlue data require some calibration. For wind speed data, MeteoBlue performs slightly better than AccuWeather for the whole 2011–2013 period.

Errors in model performance with the MeteoBlue data decreased when the WSC values were greater than 1 in 2012 (Fig. 8). A value greater than 1 means the recorded wind speed from the meteorological station needs to be intensified when it is applied to the reservoir. The primary reason is the fetch effect, which is a well-understood concept when using land station data on a water body (e.g. Gulliver and Stefan 1986, Condie and Webster 1997, McJannet *et al.* 2012). According to the yearly averages (Table 4), MeteoBlue has even higher wind speed values (1.1 m/s on average) compared to the other databases. Hence, the main reason for higher WSC values could be inaccuracies in wind direction calculations, and consequently smaller fetch, in MeteoBlue.

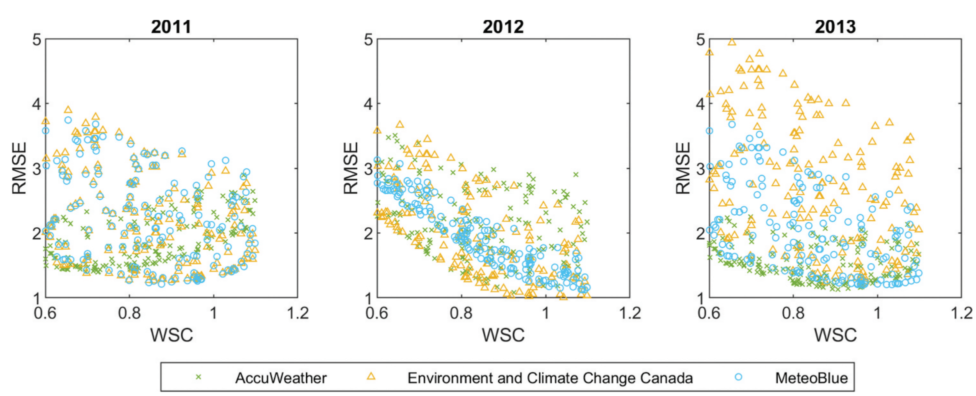

Figure 8. Sensitivity analysis of Lake Diefenbaker temperature model performance using the three meteorological databases, for each study year. The figure shows the effects of wind sheltering coefficient (WSC) on model performance for all the measurements based on the best run for each meteorological station for 2011–2013.

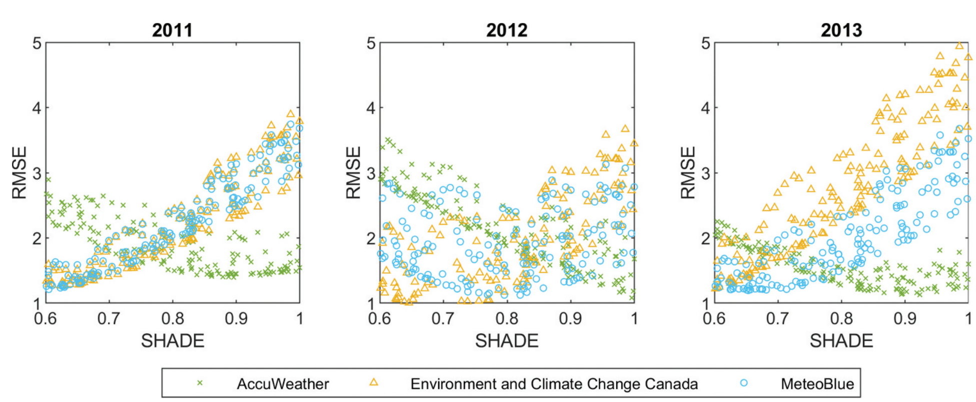

Figure 9. Sensitivity analysis of Lake Diefenbaker temperature model performance based on the three meteorological databases. The figure shows the effects of solar radiation shading (SHADE) on model performance for all the measurements based on the best run for each meteorological station for 2011–2013. RMSE: root mean square error.

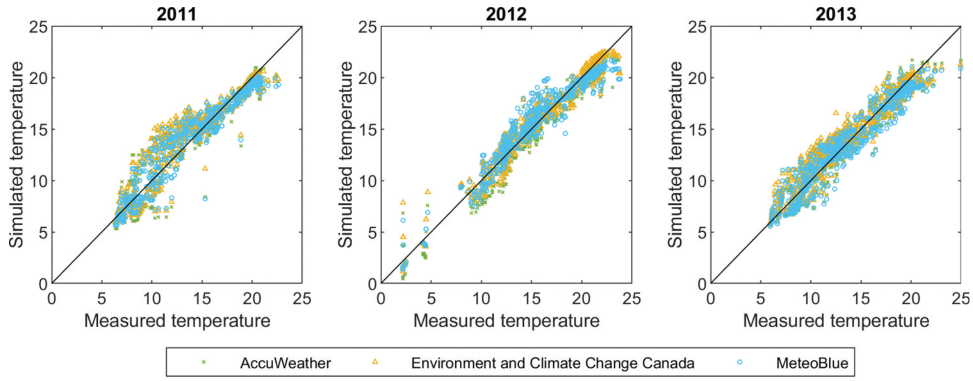

Figure 10. Sensitivity analysis of Lake Diefenbaker temperature model performance based on the three meteorological databases. The figure shows the overall model performance for all the measurements based on the best run for each meteorological station for 2011–2013 by using Wind sheltering coefficient (WSC) and solar radiation shading coefficient (SHADE)as calibrating parameters.

In contrast to the WSC, model errors decrease with MeteoBlue and ECCC data when smaller values are used for the SHADE coefficient (Fig. 9). Small SHADE values indicate that the amount of solar radiation that reaches the surface of the water in the river and reservoir should be decreased. As mentioned before, cloud cover data are an important variable, which is absent in the ECCC database. Also, in MeteoBlue stations, the cloud cover is underestimated while the shortwave radiation is overestimated. Therefore, the small SHADE coefficient compensates for cloudy days which are missing from the ECCC, and adjusts the inaccuracies in cloud cover and solar radiation in MeteoBlue.

Model performances with the data from the AccuWeather database are the best, with no need for adjustments except for WSC in 2011. Discovering this point is a rewarding achievement because we now know we can use the AccuWeather data for historical modelling when there are no data available for model calibration and validation with higher confidence.

Based on RMSE values, the re-analysed data in the MeteoBlue database may produce slightly better results compared with the actual measured data in the AccuWeather database but with more parameter adjustments. The best results with the MeteoBlue data for WSC are similar to those obtained with the AccuWeather data when the calculated wind speed values are used without any correction. But the model outcomes based on the SHADE coefficient show that the results match better with the observed temperature values when the solar radiation input is used with slight alterations. The reason for this may be uncertainties in cloud cover estimation in the MeteoBlue database while, in AccuWeather database, the cloud cover data are actual measurements.

Figure 10 compares all field measurements and simulated values for the best model performances based on each database for each year. The recorded water temperature data used for model calibration are from field observations from late spring until early fall. Hence, the measurements seldom approach 0°C. The majority of observations range between 10°C and a maximum water temperature of 25°C in summer. The graphs show that the simulated values from all three databases plotted against measured values overlap each other very well, with similar deviation patterns from the straight line. These deviations are higher in 2011 followed by 2013 (Fig. 10).

Temperature profiles for the best performances of each database show that the model was able to emulate the stratification at different places along the reservoir (Fig. 11 and Appendix Fig. A1). Vertical water temperature gradients are high during the stratification period. Hence, correctly calculating the time of stratification and the depth and thickness of the thermocline significantly affects the model performance. This is the main reason for the strong influence of an accurate prediction of meteorological forcing data on the model performance, especially where the effects of inflow water are reduced. Thermocline thickness predictions were accurate in model runs using the data (see the Appendix, Fig. A1). The vertical profiles for all three databases almost overlay each other completely in most locations and most times. Thus, the modelled data from the MeteoBlue database produced results as accurate as those from the AccuWeather and the ECCC database.

Another important issue for selecting a database is ease of access to the data. ECCC data are available through the ECCC website free of charge and are the easiest to obtain. In contrast, there are processing fees for obtaining the data from the other two databases. MeteoBlue provides data free of charge for educational purposes, and we were able to obtain the data in less than 24 hours. AccuWeather data are most expensive, and it took almost a month to obtain the data.

In selecting a good climate database, it is important that the database has all the required variables for the specific study. ECCC has high-quality databases; however, a limited number of meteorological stations and an absence of cloud cover data can produce large errors in

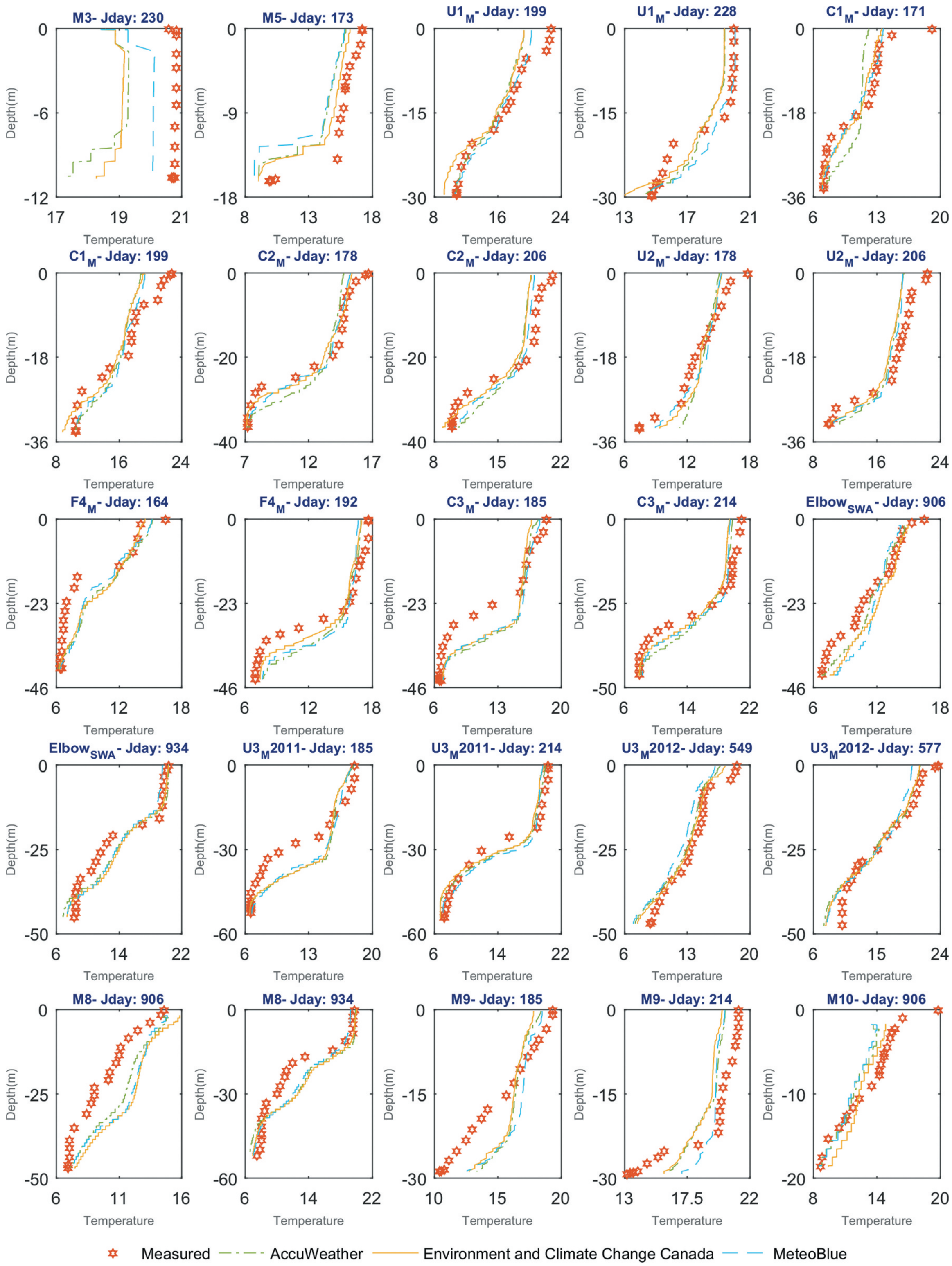

Figure 11. Comparing the temperature profiles from the results of the three meteorological databases with the measured water sampling data. The results are presented from the most upstream station to the most downstream station for observations made during 2011–2013.

the absence of measured data to calibrate the model. AccuWeather has more stations and variables, but the data are expensive and takes weeks to receive. We also found the MeteoBlue database a very good option because the data produce good results, are free of charge, and are readily obtained.

Conclusions

We used three different climate databases to test the performance of a water quality model: ECCC data with long-term quality control standards, AccuWeather data with less than 10 years of quality control data, and modelled MeteoBlue data. We used the climate data from these databases to run the CE-QUAL-W2 water quality model for Lake Diefenbaker in Saskatchewan, Canada. The results show very good model performances with all three databases, but with more reliance on calibration in ECCC and MeteoBlue. The main reasons are a lack of cloud cover data and the lower number of stations in ECCC databases, and the overestimation of solar radiation and underestimation of cloud cover data in MeteoBlue. Cloud cover influences the heat budget by affecting shortwave and longwave solar radiation.

MeteoBlue data, which are modelled data, produced small errors but required adjustments to some parameters. These results are a validation of the MeteoBlue modelling algorithms as well. Easiness and quick access to the database and support of academic projects by providing free access to those data can be a motivator for considering the use of MeteoBlue data in many environmental studies in future.

AccuWeather data are expensive and time consuming to obtain but provide the best results with little need for adjustment of parameters. Considering the length and cost of a study, these fees and the time to obtain data may be justifiable in many large research programmes. The database can be used for modelling the locations or time frames when observations are not available to calibrate model parameters. Additionally, AccuWeather data can significantly reduce the field monitoring costs with lower data requirements for model verification.

When selecting a meteorological database, it is important that the database has all the required variables. ECCC has high-quality databases; however, the lack of cloud cover data demands model parameter calibration to avoid errors in the CE-QUAL-W2 water quality model. Also, the limited number of meteorological stations is a drawback. Depending on the proximity of the study site to stations available in the database and the variables required for the study, the ECCC data can be used with confidence.

Acknowledgements

This work was financially supported by the Canada Excellence Research Chair in Water Security through the Global Institute for Water Security. We thank Environment and Climate Change Canada for providing the meteorological data and the hydrometric data. We are grateful to www.meteoblue.com for providing the meteorological data. Thanks to the Limnology Laboratory at the University of Saskatchewan for providing the water temperature data. Thanks to the Department of Geography and Planning at the University of Saskatchewan for providing the bathymetry data.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Global Institute for Water Security, University of Saskatchewan.

ORCID

Amir Sadeghian http://orcid.org/0000-0003-4157-0176
Jeff Hudson http://orcid.org/0000-0002-0256-3240
Karl-Erich Lindenschmidt http://orcid.org/0000-0001-6348-2295

References

AccuWeather, 2016. AccuWeather exceeds 9.5 billion requests for global data every day, setting new record in big data demand. Available from: http://www.accuweather.com/en/press/43009943, [Accessed 7 December 2016].

Aguilera, J., *et al.*, 1999. Effects of solar radiation on growth, photosynthesis and respiration of marine macroalgae from the Arctic. *Marine Ecology Progress Series*, 191, 109–119. doi:10.3354/meps191109

Akomeah, E., *et al.*, 2021. The impacts of changing climate and streamflow on nutrient speciation in a large Prairie reservoir. *Environmental Management*, 288, 112262.

Annear, R.L. and Wells, S.A., 2007. A comparison of five models for estimating clear-sky solar radiation. *Water Resources Research*, 43 (10), 10. doi:10.1029/2006WR005055.

Assemany, P.P., *et al.*, 2015. Algae/bacteria consortium in high rate ponds: influence of solar radiation on the phytoplankton community. *Ecological Engineering*, 77, 154–162. doi:10.1016/j.ecoleng.2015.01.026

Carr, M., *et al.*, 2019. Modelling the possible impacts of climate change on the thermal regime and macroinvertebrate species of a regulated prairie river. *Ecohydrology*, 12 (6), e2102. doi:10.1002/eco.2102.

Carr, M.K., *et al.*, 2020. Impacts of varying dam outflow elevations on water temperature, dissolved oxygen, and nutrient distributions in a large prairie reservoir. *Environmental Engineering Science*, 37 (1), 78–97. doi:10.1089/ees.2019.0146.

Cazorla, A., Olmo, F., and Alados-Arboledas, L., 2008. Development of a sky imager for cloud cover assessment. *JOSA A*, 25 (1), 29–39. doi:10.1364/JOSAA.25.000029.

Changnon, S.A., Lamb, P.J., and Hubbard, K.G., 1990. Regional climate centers: new institutions for climate services and climate-impact research. *Bulletin of the American Meteorological Society*, 71 (4), 527–537. doi:10.1175/1520-0477(1990)071<0527:RCCNIF>2.0.CO;2.

Cole, T.M. and Wells, S.A., 2015. *CE-QUAL-W2: a two-dimensional, laterally averaged, hydrodynamic and water quality model*. 3.72 edition. Portland, OR: Department of Civil and Environmental Engineering, Portland State University.

Condie, S.A. and Webster, I.T., 1997. The influence of wind stress, temperature, and humidity gradients on evaporation from reservoirs. *Water Resources Research*, 33 (12), 2813–2822. doi:10.1029/97WR02405.

Craft, E.D., 1999. Private weather organizations and the founding of the United States Weather Bureau. *The Journal of Economic History*, 59 (4), 1063–1071. doi:10.1017/S0022050700024141.

De Jong, R. and Stewart, D.W., 1993. Estimating global solar radiation from common meteorological observations in western Canada. *Canadian Journal of Plant Science*, 73 (2), 509–518. doi:10.4141/cjps93-068.

Environment and Climate Change Canada, 2017. Quality of historical climate data. Available from: http://climate.weather.gc.ca/climate/data/data/quality/e.html [Accessed 28 February 2017].

Feng, S., Hu, Q., and Qian, W., 2004. Quality control of daily meteorological data in China, 1951–2000: a new dataset. *International Journal of Climatology*, 24 (7), 853–870. doi:10.1002/joc.1047.

Goldman, C.R., 1988. Primary productivity, nutrients, and transparency during the early onset of eutrophication in ultra-oligotrophic Lake Tahoe, California-Nevada. *Limnology and Oceanography*, 33 (6), 1321–1333. doi:10.4319/lo.1988.33.6.1321.

Gulliver, J.S. and Stefan, H.G., 1986. Wind function for a sheltered stream. *Journal of Environmental Engineering*, 112 (2), 387–399. doi:10.1061/(ASCE)0733-9372(1986)112:2(387).

Haberlandt, U., 2007. Geostatistical interpolation of hourly precipitation from rain gauges and radar for a large-scale extreme rainfall event. *Journal of Hydrology*, 332 (1–2), 144–157. doi:10.1016/j.jhydrol.2006.06.028.

Hudson, J.J. and Vandergucht, D.M., 2015. Spatial and temporal patterns in physical properties and dissolved oxygen in Lake Diefenbaker, a large reservoir on the Canadian Prairies. *Journal of Great Lakes Research*, 41, 22–33. doi:10.1016/j.jglr.2015.06.007

Hunt, L., Kuchar, L., and Swanton, C., 1998. Estimation of solar radiation for use in crop modelling. *Agricultural and Forest Meteorology*, 91 (3–4), 293–300. doi:10.1016/S0168-1923(98)00055-0.

Kassianov, E., Long, C.N., and Ovtchinnikov, M., 2005. Cloud sky cover versus cloud fraction: whole-sky simulations and observations. *Journal of Applied Meteorology*, 44 (1), 86–98. doi:10.1175/JAM-2184.1.

McJannet, D.L., Webster, I.T., and Cook, F.J., 2012. An area-dependent wind function for estimating open water evaporation using land-based meteorological data. *Environmental Modelling and Software*, 31, 76–83. doi:10.1016/j.envsoft.2011.11.017

MeteoBlue, 2016. Weather simulation data – what's that? Available from: https://content.meteoblue.com/en/products/time-dimensions/history/weather-simulation-data [Accessed 7 December 2016].

Mi, C., *et al.*, 2019. Variable withdrawal elevations as a management tool to counter the effects of climate warming in Germany's largest drinking water reservoir. *Environmental Sciences Europe*, 31 (1), 19. doi:10.1186/s12302-019-0202-4.

Morales-Marin, L.A., Carr, M., Sadeghian, A., and Lindenschmidt, K.E. 2021. Climate change effects on the thermal stratification of Lake Diefenbaker, a large multi-purpose reservoir. *Canadian Water Resources Journal/Revue Canadienne Des Ressources Hydriques*, 46 (1–2) , 1–16.

NOAA, 2017. Quality controlled local climatological data (qclcd). Available from: https://www.ncdc.noaa.gov/data-access/land-based-station-data/land-based-datasets/quality-controlled-local-climatological-data-qclcd [Accessed 28 February 2017].

Pomeroy, J.W., Fang, X., and Williams, B., 2009. *Impacts of climate change on Saskatchewan's water resources*. Centre for Hydrology. Report No. 6. Saskatoon, SK, Canada: University of Saskatchewan.

Sadeghian, A., *et al.*, 2015. Lake Diefenbaker temperature model. *Journal of Great Lakes Research*, 41, 8–21. doi:10.1016/j.jglr.2015.10.002

Sadeghian, A., *et al.*, 2017b. Sediment plume model—a comparison between use of measured turbidity data and satellite images for model calibration. *Environmental Science and Pollution Research*, 24 (24), 19583–19598. doi:10.1007/s11356-017-9616-y.

Sadeghian, A., *et al.*, 2018. Improving in-lake water quality modeling using variable chlorophyll a/algal biomass ratio. *Environmental Modelling and Software*, 101, 73–85. doi:10.1016/j.envsoft.2017.12.009

Sadeghian, A., De Boer, D., and Lindenschmidt, K.-E., 2017a. Sedimentation and erosion in Lake Diefenbaker, Canada: solutions for shoreline retreat monitoring. *Environmental Monitoring and Assessment*, 189 (10), 507. doi:10.1007/s10661-017-6217-7.

Sentelhas, P.C., Gillespie, T.J., and Santos, E.A., 2010. Evaluation of FAO Penman–Monteith and alternative methods for estimating reference evapotranspiration with missing data in Southern Ontario, Canada. *Agricultural Water Management*, 97 (5), 635–644. doi:10.1016/j.agwat.2009.12.001.

Sheffield, J., Goteti, G., and Wood, E.F., 2006. Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling. *Journal of Climate*, 19 (13), 3088–3111. doi:10.1175/JCLI3790.1.

Souza-Echer, M., *et al.*, 2006. A simple method for the assessment of the cloud cover state in high-latitude regions by a ground-based digital camera. *Journal of Atmospheric and Oceanic Technology*, 23 (3), 437–447. doi:10.1175/JTECH1833.1.

Strickland, J.D.H., 1958. Solar radiation penetrating the ocean. a review of requirements, data and methods of measurement, with particular reference to photosynthetic productivity. *Journal of the Fisheries Research Board of Canada*, 15 (3), 453–493. doi:10.1139/f58-022.

Tabios, G.Q. and Salas, J.D., 1985. A comparative analysis of techniques for spatial interpolation of precipitation. *JAWRA Journal of the American Water Resources Association*, 21 (3), 365–380. doi:10.1111/j.1752-1688.1985.tb00147.x.

Terry, J., *et al.*, 2018. Challenges of modelling water quality in a shallow prairie lake with seasonal ice cover. *Ecological Modelling*, 384, 43–52. doi:10.1016/j.ecolmodel.2018.06.002

Terry, J., Sadeghian, A., and Lindenschmidt, K.-E., 2017. Modelling dissolved oxygen/sediment oxygen demand under ice in a shallow eutrophic prairie reservoir. *Water*, 9 (2), 131. doi:10.3390/w9020131.

Toth, B., *et al.*, 2009. The natural characteristics of the South Saskatchewan River Basin: climate, geography and hydrology. *Prairie Forum*, 34 (1), 95–127.

Vejen, F., *et al.*, 2002. Quality control of meteorological observations automatic methods used in the Nordic countries. *Climate Report*, 8, 1–102 .

Yamashita, M., Yoshimura, M., and Nakashizuka, T., 2004. Cloud cover estimation using multitemporal hemisphere imageries. *International Archives of Photogrammetry, Remote Sensing and Spatial Information*, 35, 826–829.

Yang, K., Koike, T., and Ye, B., 2006. Improving estimation of hourly, daily, and monthly solar radiation by importing global data sets. *Agricultural and Forest Meteorology*, 137 (1–2), 43–55. doi:10.1016/j.agrformet.2006.02.001.

Appendix

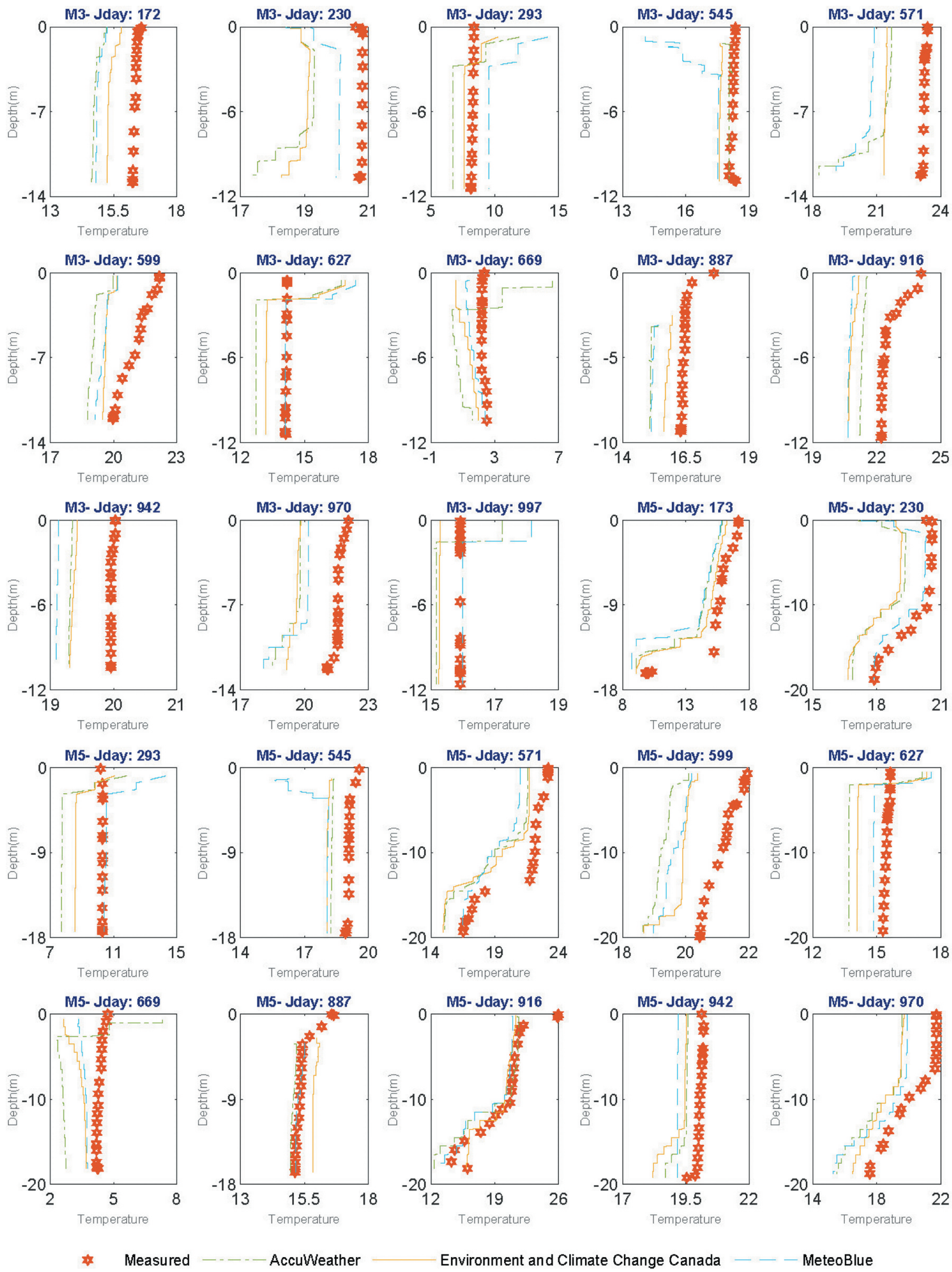

Figure A1. Temperature profiles for the best performances of the Environment and Climate Change Canada, AccuWeather and MeteoBlue databases among all the temperature profiles recorded in Lake Diefenbaker during 2011–2013.

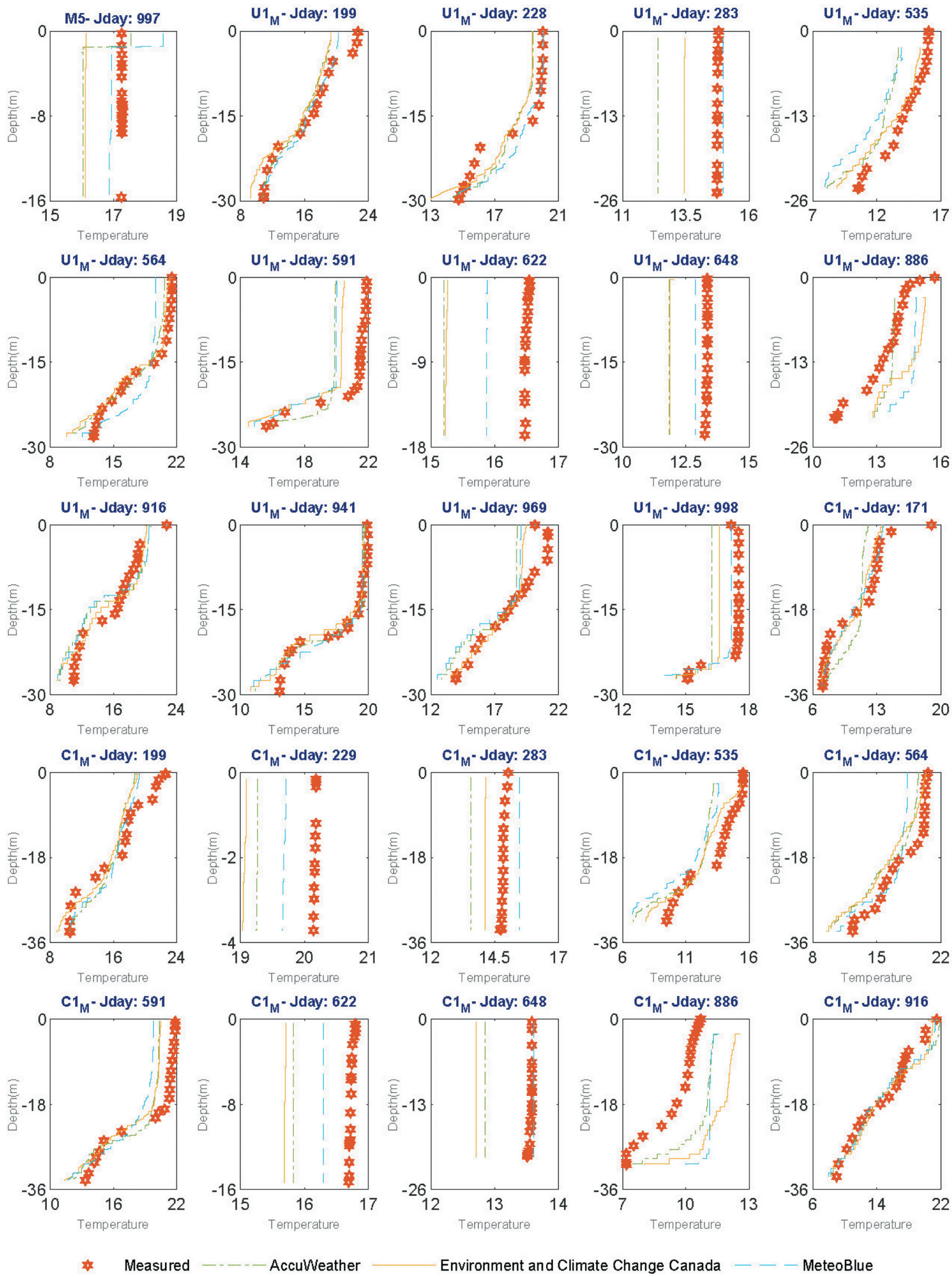

Figure A1. (Continued).

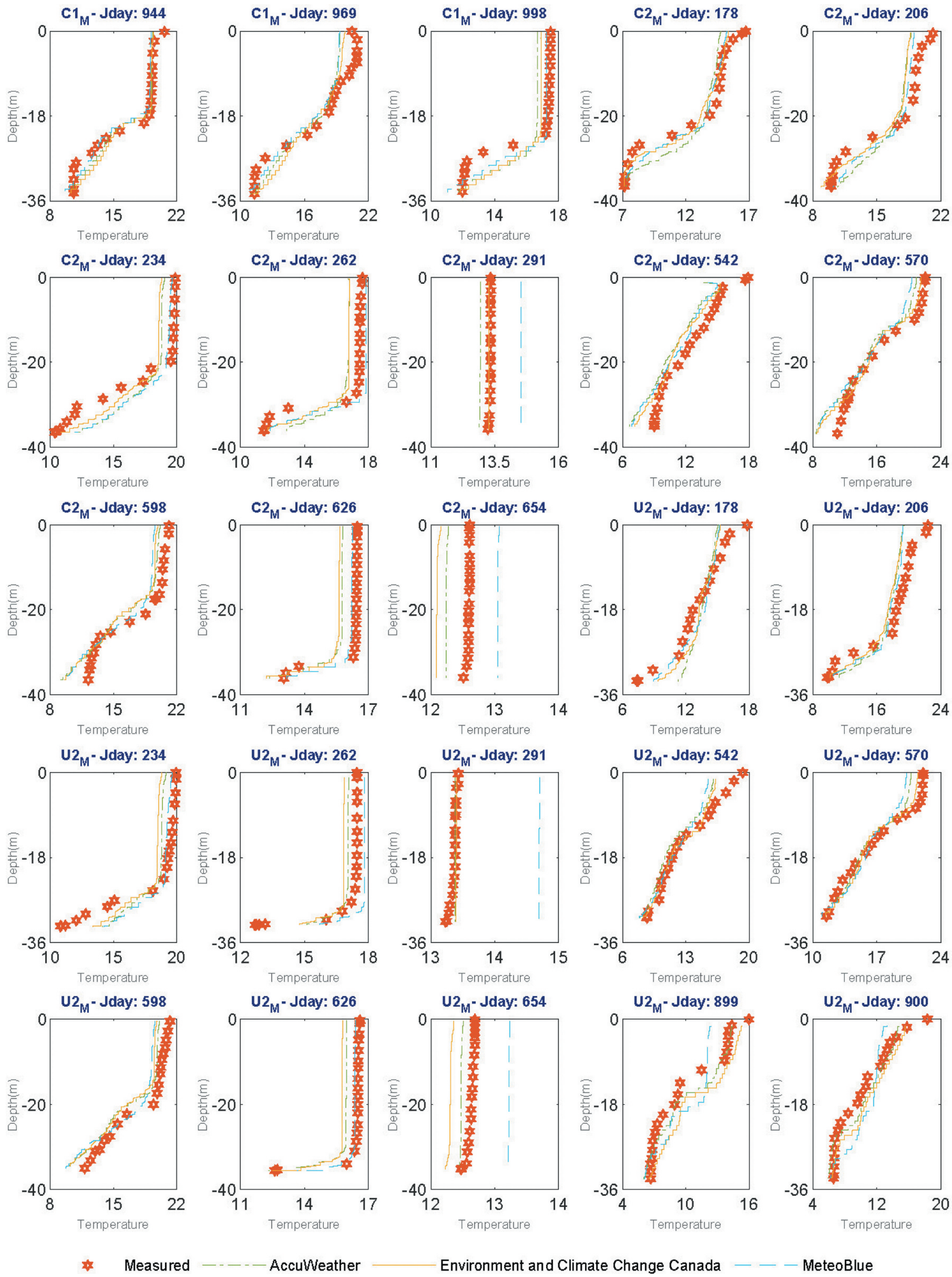

Figure A1. (Continued).

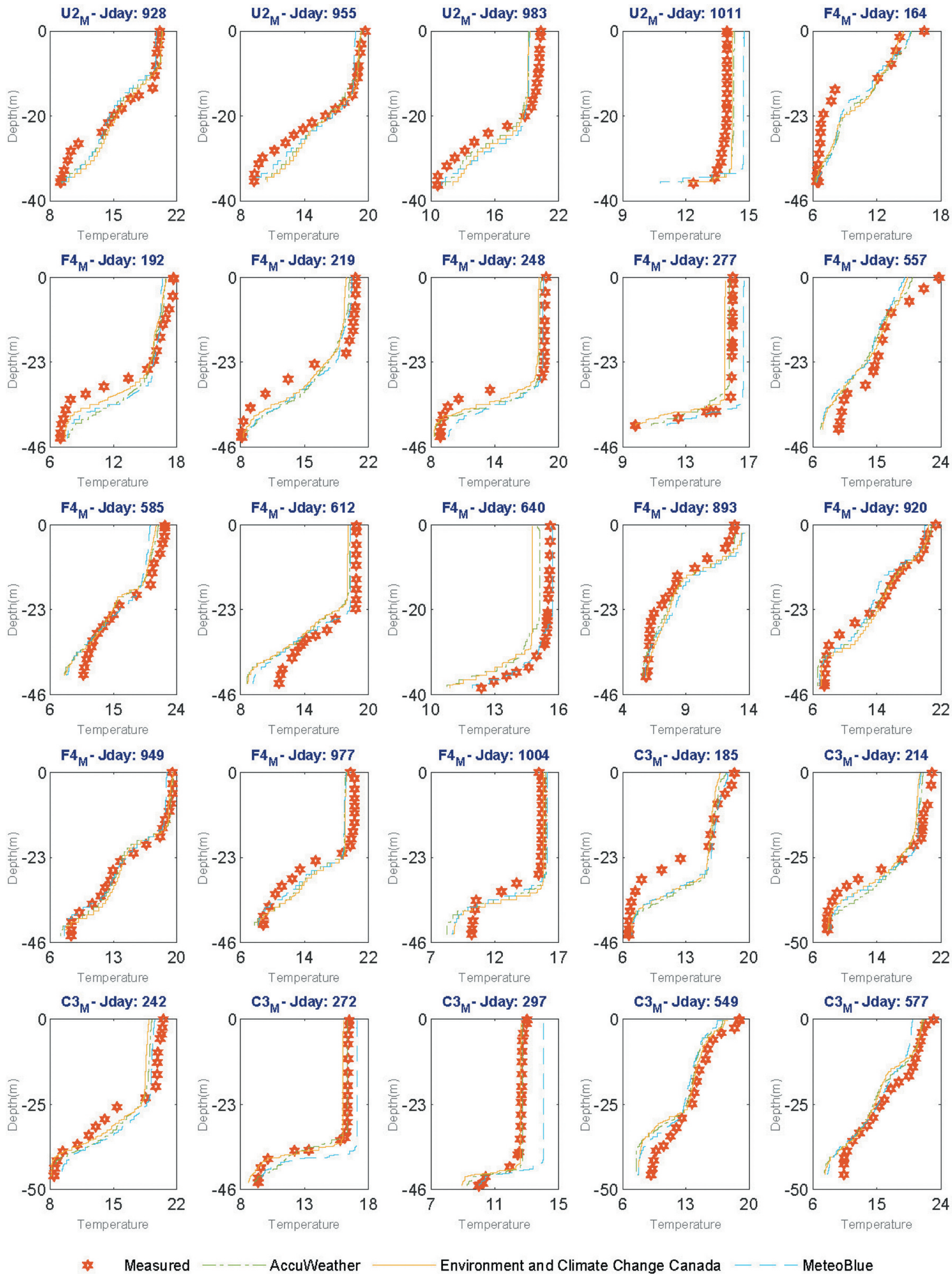

Figure A1. (Continued).

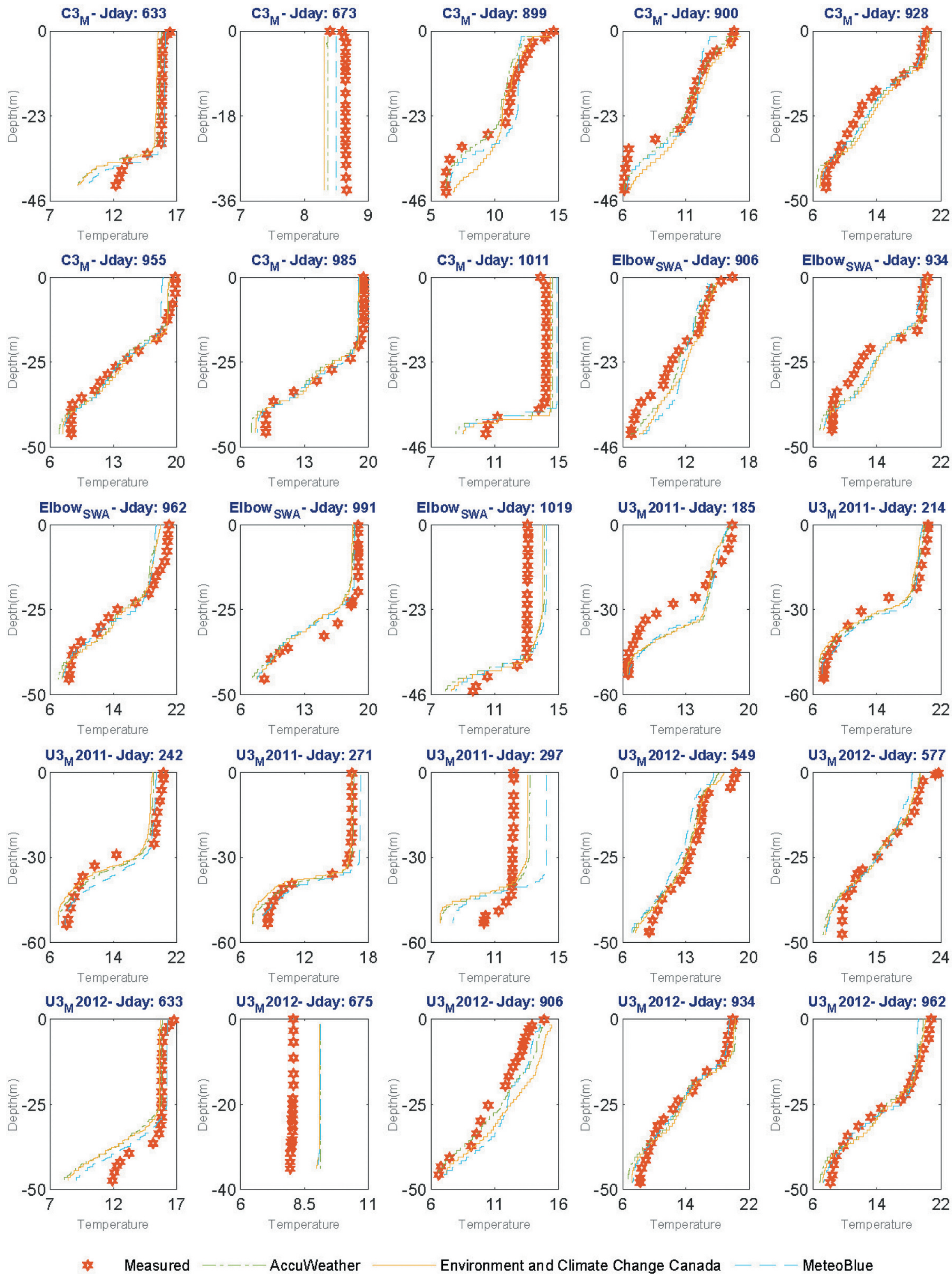

Figure A1. (Continued).

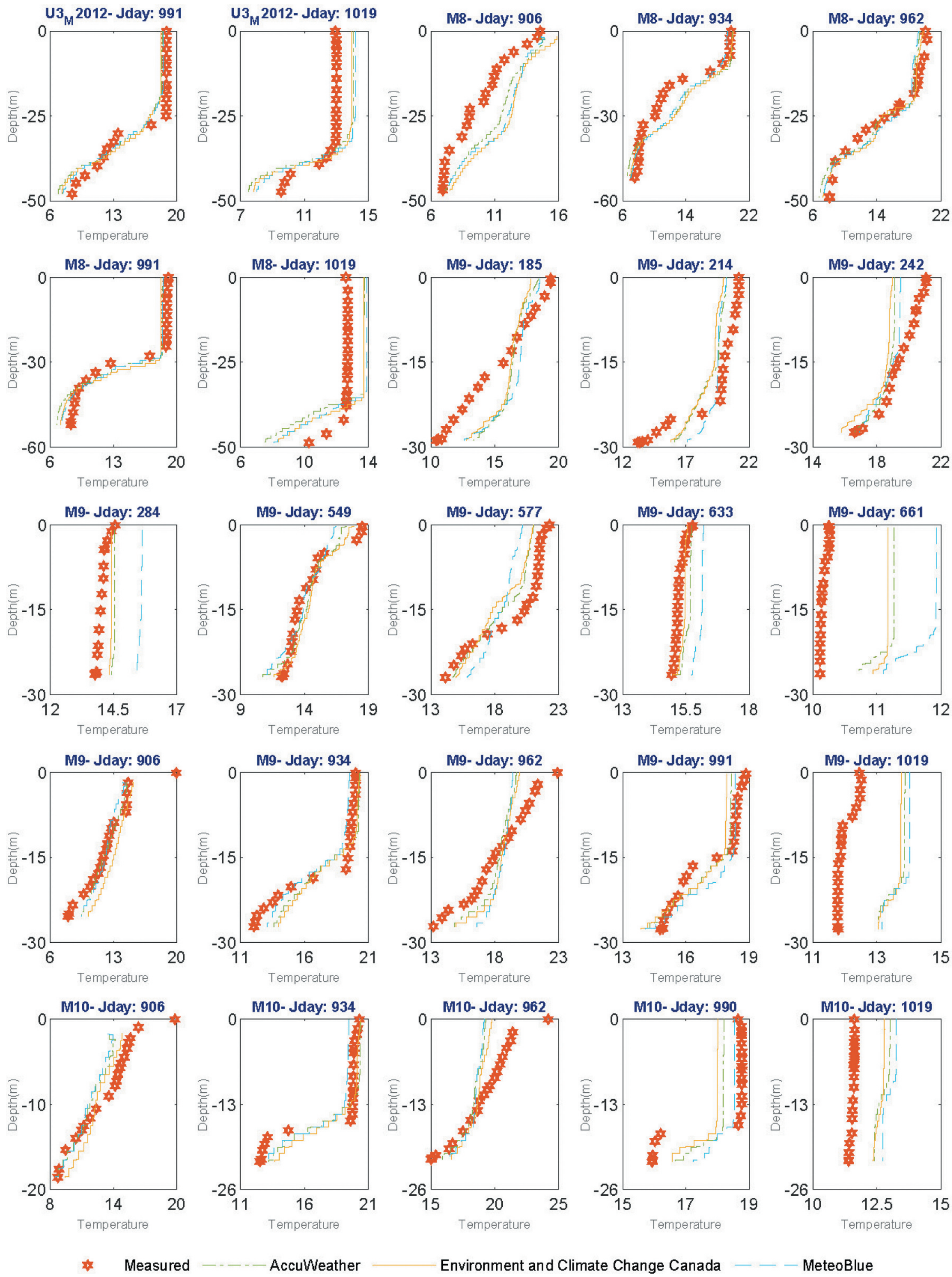

Figure A1. (Continued).