nature climate change

ARTICLES

https://doi.org/10.1038/s41558-020-0863-4

Identification of local water resource vulnerability to rapid deglaciation in Alberta

Sam Anderson ✉ and Valentina Radić

Global glacier retreat driven by climate change will have major impacts on regional water availability, as many communities rely on glacier runoff for water supply during warm and dry seasons. A community whose water resources are potentially vulnerable is one that sources water from a glacier-fed river where that river is expected to substantially change if glacier contributions become negligible. However, regional assessments identifying which communities' water resources are most vulnerable to such changes are lacking. Here we use observed streamflow measurements, gridded climate data and a database of municipal water sources for communities in Alberta, Canada, to identify the relative importance of glacier runoff at the local scale. In a scenario of negligible glacier runoff, we predict unprecedented streamflow lows at several communities. This approach provides a methodology to identify communities whose water resources may be vulnerable to glacier retreat and would benefit from more-focused research.

Glacier runoff provides a portion of water supply for communities around the world[1–6]. In North America, the coming decades will see rapid and unprecedented changes to glaciers; notably, current models predict that western Canada will lose 70–90% of glacier ice volume by 2100[7,8]. Glacier-melt models typically either are detailed analyses of a single or few glaciers[9–11] (relying on dense observational data, which is unavailable in many basins of interest) or are generalized at larger spatial scales[2,3,12–16] (facing data-availability challenges). Neither approach answers the question of how glacier-fed rivers are different from non-glacier-fed rivers in the same region. If glacier-melt contributions to streamflow were to become negligible, which communities would experience a change in drinking-water supply? Regional- or global-scale analyses of glacier contributions to streamflow imply that the loss of glacier ice will impact water supply at communities[2,8]; however, to our knowledge, no regional scale study explicitly identifies all communities that source water from glacier-fed rivers in the study region and assesses how streamflow is projected to change at these locations. Furthermore, recent studies that address global or regional freshwater stress seldom consider glacier contributions to streamflow and the unique challenges facing these communities[17–21]. Policymakers require detailed analyses on a subregional scale to understand the range of outcomes that face communities within the same political jurisdiction.

Our goal is to identify which communities' water resources are vulnerable to glacier retreat by assessing observed characteristic streamflow patterns and municipal water sources within the whole region. We do not assess the ability of these communities to adapt to changes in water supply; rather, we aim to bridge the current knowledge gap that exists between regional- or global-scale glacier models (which have high uncertainties on the local scale) and the local-scale projections of streamflow that are needed for policymakers to develop adaptive strategies. If glacier runoff is an important constituent of streamflow, then it should be possible to identify fingerprints of glacier runoff by analysing streamflow time series and comparing the characteristic behaviour of glacier-fed rivers to that of non-glacier-fed rivers in the same region. We have developed a method for analysing multivariate and multidisciplinary data in a case study of Alberta, Canada. This method, which combines principal component analysis (PCA), clustering and multivariate linear regression (MLR), is flexible by design to be applicable to other glaciated regions. This method identifies the variables that should be predicted (PCA), visualizes their variability in space and time (clustering) and then predicts the key variables (MLR). The analysis framework is not restricted to the specific locations, variables or time periods presented in this case study, nor is it restricted to specific outcomes of each step in the analysis (for example, number of clusters chosen).

Alberta has a population of over 4 million people, with lowlands that experience a semi-arid climate in the rain shadow of the Rocky Mountains[22]. Alberta is the westernmost of the agriculturally productive Canadian Prairie provinces, characterized by the Rocky Mountains along the southwest border, boreal forests in the north and grassland in the southeast[23]. In lower elevation regions outside of the Rocky Mountains, mean yearly temperature varies from −3 °C in the north to 6 °C in the southeast, while mean cumulative precipitation varies from approximately 350 mm yr^{-1} in the southeast to over 500 mm yr^{-1} in the northeast[24]. Mean cumulative precipitation is much greater in the Rocky Mountains, and although data are more sparse at high elevations, where studies exist it has been measured to exceed 1,600 mm yr^{-1} (refs. [25,26]). Most major rivers are partially glaciated, including the Athabasca, Peace, and North and South Saskatchewan, and these rivers are used for municipal water supply, irrigation and hydropower generation[22,27]. While Alberta is sensitive to drought[22], glacier runoff (which peaks in summer) provides contributions to summer streamflow even after seasonal snowpacks melt[13,15,28].

We use daily discharge data of 194 stream gauge stations that are upstream of dams and have data from 1987 to 2010. Historical streamflow data were obtained through Environment Canada's HYDAT database (see Methods). We use daily near-surface air temperature, precipitation and evaporation from ERA-Interim reanalysis[24]. Glacier locations and areas are extracted from the Randolph Glacier Inventory v.6.0[27]. To our knowledge, no unified database exists that describes the drinking-water source of communities in Alberta (cities, villages, towns and hamlets), so we collect this

Department of Earth, Ocean and Atmospheric Sciences, University of British Columbia, Vancouver, British Columbia, Canada. ✉e-mail: sanderson@eoas.ubc.ca

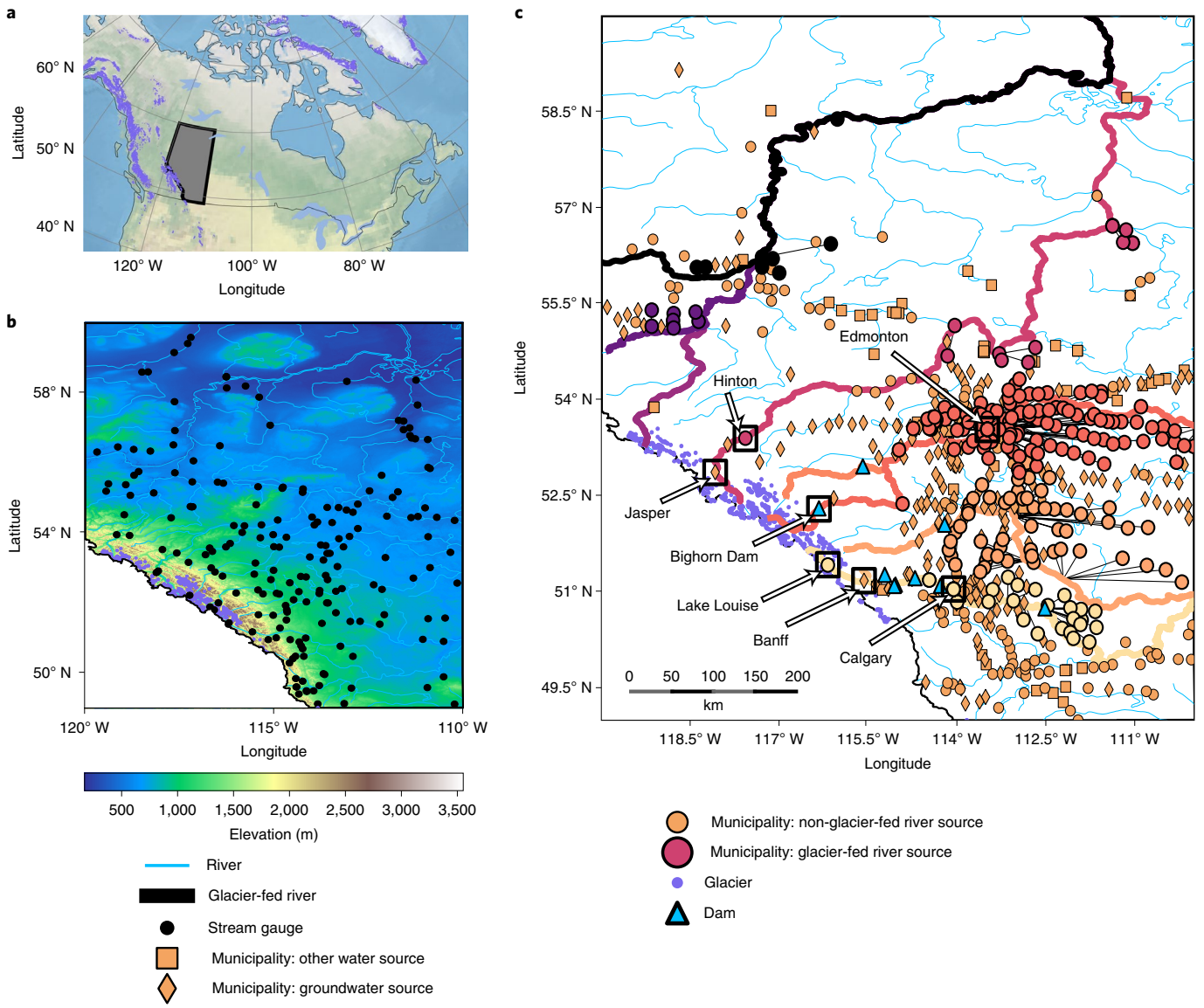

Fig. 1 | Study region of the province of Alberta, Canada. a, Location of Alberta in Canada. **b**, Locations of glaciers and stream gauges in Alberta. The shading corresponds to elevation. **c**, Locations of municipalities, rivers and dams along glacier-fed rivers. The colours of municipalities that source water from glacier-fed rivers correspond to the colours of the glacier-fed rivers from which they intake water supply. Rivers generally flow in an east or northeast direction due to the topographic gradient from the Rockies to the Prairies. Only glaciers in Alberta are plotted in **b** and **c**.

data for 567 communities from municipal websites and through directly contacting municipalities or utility providers. We classify the municipal water source as being from a glacier-fed river, from a non-glacier-fed river, from groundwater or from another source (such as a lake or a combination of sources). The locations of communities, stream gauges, glaciers and glacier-fed rivers are shown in Fig. 1.

Our method consists of the following key steps: (1) identifying signatures of glacier runoff, (2) predicting the spatial distribution of the main modes of inter-annual August streamflow for both glacier-fed and non-glacier-fed rivers and (3) quantifying differences in predicted August streamflow of glacier-fed municipal water intake locations between scenarios with and without glaciers in the watershed. August is investigated because relative to the rest of the year, it has high glacier contributions[2,28,29], low snowmelt contributions[29] and high municipal water demand[22,30]. Details of each step are provided in the Methods, and here we present the main findings of each step.

In step (1), we aim to find how August streamflow is different in glacier-fed rivers as compared with non-glacier-fed rivers in the region. We perform PCA on normalized and smoothed time series of August streamflow for the 194 stations and 24 years of observations (1987–2010). PCA reduces the dimensionality of the problem by finding linearly independent modes of variability that span the input variable space, under the condition that each mode describes the maximum possible amount of variance across the observations. We find that the first two modes (Fig. 2a) describe 98% of the total variance (Supplementary Fig. 1). The first mode describes 88% of the total variance. Principal components of the first mode (PC_1) are highly positively correlated with mean-normalized August streamflow across all stations and years (Supplementary Fig. 2a), meaning that high PC_1 values correspond to high mean-normalized August streamflow. Then, we use self-organizing maps (SOMs)[31] to cluster stream gauges on the basis of their similarity in a three-dimensional space set by PC_1, PC_2 and time; in other words, we group the stream gauges according to how similarly their key August streamflow

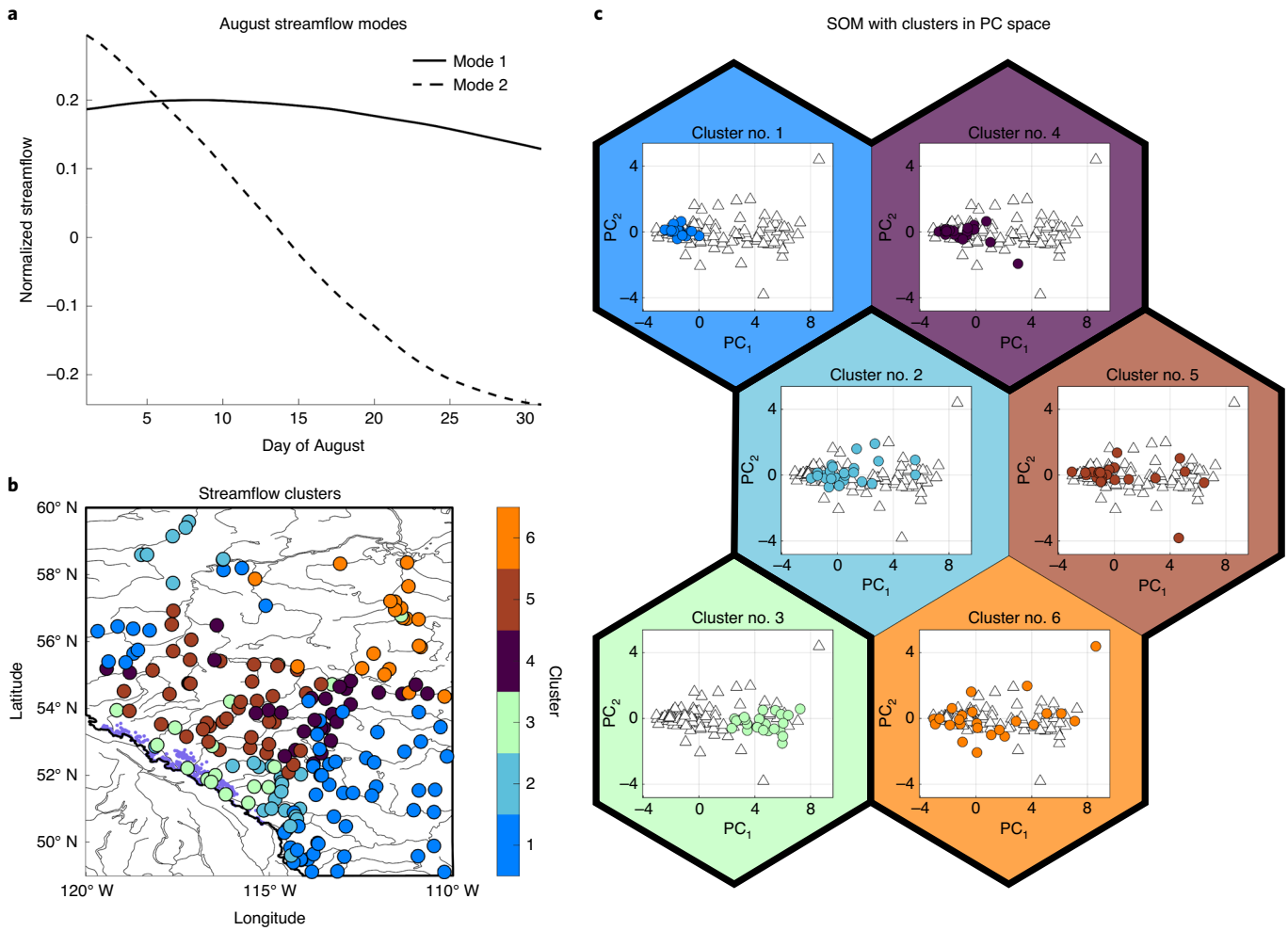

Fig. 2 | Results of PCA and clustering with SOM, showing the first two eigenvectors, the clusters in space and the SOM with clusters in PC space. **a**, Eigenvectors of the first two modes of August streamflow. **b**, The locations of stream gauges coloured according to the cluster to which they belong, as a result of the SOM clustering analysis. Glacier-fed rivers have unique behaviour, demonstrated by light-green stations (cluster no. 3). There is spatial coherence among non-glacier-fed rivers (all other clusters), indicating that nearby non-glacier-fed rivers have similar PC_1 and PC_2 time series. **c**, The final 3×2 SOM. The coloured hexagons are the clusters on the SOM, and the coloured data points are the pattern in PC space of each cluster. Coloured circles represent the data points associated with the given cluster, while uncoloured triangles show all other data points. These six clusters are characterized more generally into three groups, each marked by a thick-line border in the SOM. First, clusters no. 1 and no. 4 primarily occur in the grasslands and parklands natural regions and are characterized by having consistently low PC_1 values, meaning that mean August streamflow is consistently low relative to annual discharge. Next, cluster no. 3 occurs downstream of natural reservoirs (glaciers and lakes) and is characterized by having consistently high PC_1 values, meaning that mean August streamflow is consistently large relative to annual discharge. Finally, clusters no. 2, no. 5 and no. 6 primarily occur in the boreal forest, foothills and non-glaciated parts of the Rocky Mountain natural regions and are all characterized by having a large range of PC_1 values, meaning that mean August streamflow is highly variable relative to annual discharge; in other words, in some years August is a substantial part of annual streamflow, while other years it is not.

patterns fluctuate in time (Fig. 2b,c). We find that glacier-fed rivers (cluster no. 3) are regionally unique, being characterized by relatively high PC_1 and low inter-annual variability of PC_1, meaning that mean August streamflow is consistently high relative to yearly discharge. By contrast, we find that for non-glacier-fed rivers, the mean of PC_1 ($\overline{PC_1}$) is positively correlated with the standard deviation of PC_1 (σ_{PC1}). This indicates that in non-glacier-fed rivers, mean August streamflow can be high on average compared with yearly discharge, but at the cost of high year-to-year variability. The relationship between $\overline{PC_1}$ and σ_{PC1} for glacier-fed and non-glacier-fed rivers is shown in Fig. 3a.

In step (2), we aim to predict the leading mode of streamflow for both glacier-fed and non-glacier-fed rivers. Rather than predict PC_1 for each year, our goal is to model its mean ($\overline{PC_1}$) and its standard deviation (σ_{PC1}) over the whole period, from which we can reconstruct statistical distributions of streamflow. More-detailed derivations of these models are presented in the Methods, and here we review the main results. We find that $\overline{PC_1}$ and σ_{PC1} can be successfully predicted ($R^2=0.64$ and $R^2=0.59$, respectively) for non-glacier-fed rivers using MLR models with the following set of predictors averaged over the whole period: mean annual temperature (T_{year} in K), mean June, July and August (JJA) temperature (T_{JJA} in K), mean JJA precipitation (P_{JJA} in mm), mean JJA evaporation (E_{JJA} in metre water equivalent (m.w.e.)), stream gauge elevation (h in m above sea level (a.s.l.)) and the logarithm of the ratio of minimum seasonal flow to maximum seasonal flow. Spatial distributions of these predictors are shown in Supplementary Fig. 3. By contrast, these MLR models poorly predict $\overline{PC_1}$ and σ_{PC1} for glacier-fed rivers

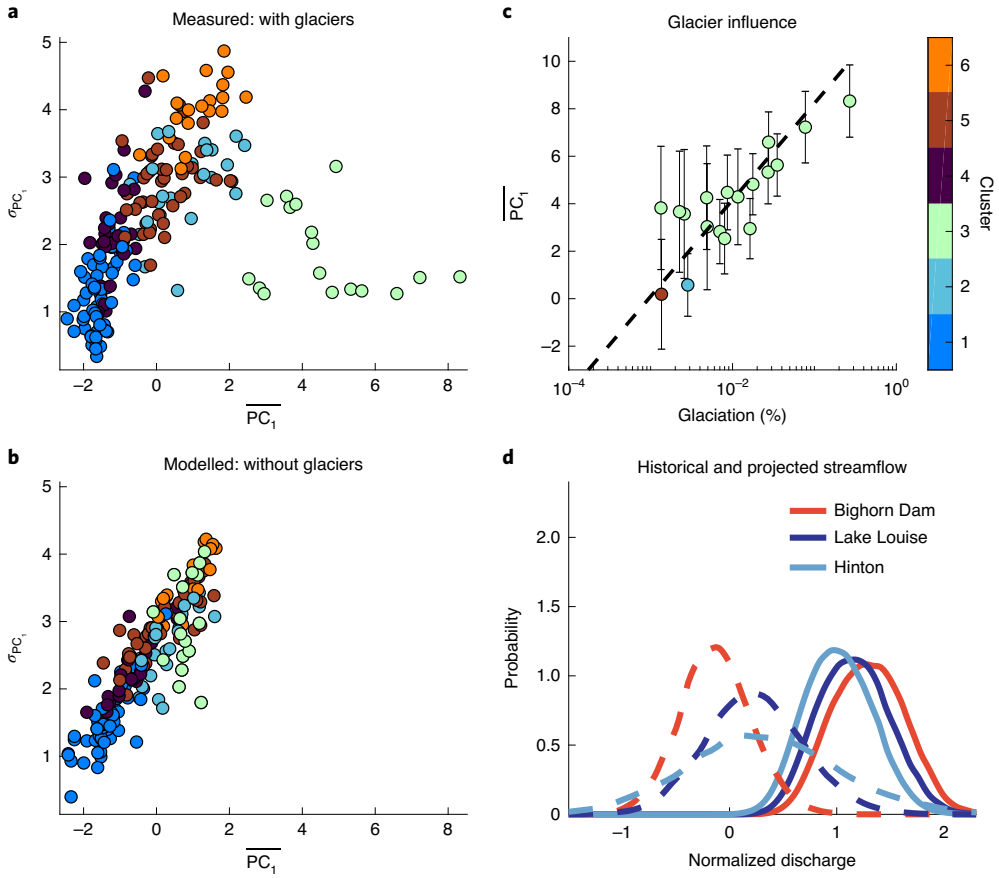

Fig. 3 | The results of the regression models and the projected streamflow at the identified vulnerable communities. a, σ_{PC1} versus $\overline{PC_1}$, coloured according to the clusters in Fig. 2. Increasing $\overline{PC_1}$ is accompanied by increasing σ_{PC1} in all clusters except cluster no. 3. **b**, The results of the MLR models that predict $\overline{PC_1}$ and σ_{PC1}. The models reproduce the positive relationship between $\overline{PC_1}$ and σ_{PC1}. $\overline{PC_1}$ and σ_{PC1} of glacier-fed rivers in this plot (cluster no. 3) describe the projected summer streamflow if glacier runoff was negligible. **c**, $\overline{PC_1}$ versus percentage glaciation for glacier-fed rivers. Error bars are σ_{PC1} ($n=24$). $\overline{PC_1}$ values increase with percentage glaciation of a basin. **d**, Normalized August streamflow probability distributions for the Bighorn Dam, Lake Louise and Hinton. All curves are normalized to have unity area. Solid curves are modelled estimates of the historical streamflow at community intake locations, and dashed curves are modelled estimates of future streamflow at community intake locations if glacier runoff becomes negligible and other factors remain constant.

($R^2=0.26$ and $R^2=0.01$, respectively). We hypothesize that this poor performance is because streamflow is instead driven by glacier runoff, the driving processes of which are not well captured by the comparatively low-resolution climate reanalysis data. Following a set of tests for optimal predictors (see Methods), we find that the leading mode is best predicted by an ordinary linear regression model with the logarithm of percentage glaciation as a predictor ($R^2=0.70$ for $\overline{PC_1}$; Fig. 3c). Overall, the analysis reveals that all non-glacier-fed rivers behave in a similar way to each other (with positively correlated $\overline{PC_1}$ and σ_{PC1}, which can be predicted with MLR models) while glacier-fed rivers have starkly unique behaviour (which can be predicted by the percentage glaciation). We hypothesize that if glacier runoff were to become negligible, glacier-fed rivers would behave similarly to non-glacier-fed rivers, and their future streamflow behaviour could be predicted by the MLR models.

In step (3), we use this hypothesis to identify local-scale water resource vulnerability by predicting streamflow in glacier-fed rivers at the locations of community water intakes when glaciers no longer dominate the streamflow signal. We consider only how streamflow will be different if glacier contributions are negligible, and do not consider how other changes to climate may change streamflow. We focus on communities that source water from glacier-fed rivers (Fig. 1c). We reconstruct their normalized August streamflow for two cases. In the first case, representing the present state with glaciers in the basins, PC_1 is modelled with the ordinary linear regression model using the logarithm of percentage glacierization as the predictor. In the second case, representing no glaciers in the basins, PC_1 is modelled with the MLR models, which work well for non-glacier-fed rivers in the region. Normalized streamflow for both scenarios is then reconstructed from the modelled PC_1 values. Communities with the most-vulnerable water resources are taken to be those that have the largest difference in mean August discharge between the two scenarios.

The results revealed four locations where water resources are most vulnerable if glacier contributions become negligible: the hamlet of Lake Louise, the Bighorn Dam, the town of Rocky Mountain House (downstream of the Bighorn Dam) and the town of Hinton. These locations indicate diverse stakeholders: Lake Louise as part of the Canadian Rocky Mountain Parks UNESCO World Heritage Site; the Bighorn Dam as the largest reservoir in Alberta, which over one million people depend on; and Hinton and Rocky Mountain House, communities of fewer than ten thousand people each. The Bighorn Dam will experience the largest decrease in mean August flow, with a future mean flow that is without historical precedent; however, the variability of flow is not projected to substantially change. By contrast, Hinton is projected to have the smallest change

in mean August flow but a comparatively large increase in variability. Lake Louise is projected to experience a decrease in mean flow and an increase in flow variability that are between those projected for the Bighorn Dam and Hinton. In all cases, days that are below average flow in the future are without historical precedent (Fig. 3d).

We highlight that some mountain communities, markedly Jasper and Banff (shown in Fig. 1c), are situated along glacier-fed rivers but source municipal water from deep underground aquifers. In fact, there are eight communities in total whose water supplies would be vulnerable if they sourced water from a nearby glacier-fed river, but instead they source their water supply from groundwater, shielding them from the effects of glacier retreat in the short term. This result confirms that municipal water source data are important for understanding water resource vulnerability in the region.

Our framework has several key advantages over previous studies. First, our approach starts with streamflow observations and detects the importance of glacier runoff, while well-established approaches in glaciohydrological research start by assessing the inflow processes such as snowmelt and glacier runoff and then validate modelled streamflow against available observations. By doing this inverse approach, we aim to circumvent the problem of limited glacier mass-balance, snowmelt and weather data in mountainous terrain. Second, we highlight the intra-basin heterogeneity of streamflow patterns in glaciated basins and demonstrate that a majority of a basin can remain unperturbed by the loss of glacier ice while acute vulnerable locations can exist in the upper reaches. Huss and Hock used a temperature-index-based melt model to project changes in glacier runoff and the relative importance of glacier runoff to 56 large-scale glacierized drainage basins around the world[2]. They concluded that the Nelson River basin in Alberta has negligible glacier impact; however, this basin-wide approach neglects the acute local impacts that will occur in the upper reaches of the basin (at the Bighorn Dam, identified by our method). Third, we highlight that knowing the water source of communities is key to understanding regional vulnerability to glacier retreat. Naz et al. used glaciohydrological models to study glacier contributions to streamflow in the Bow River at Lake Louise[15]. They concluded that glacier retreat will have important implications for water availability on both local and regional scales. Including municipal water source data, however, revealed that glacier retreat in the headwaters of the Bow River will have important implications for municipal water availability at the hamlet of Lake Louise but may not have important implications for municipal water availability impacts on a regional scale.

It is challenging to calibrate glacier melt models in a region with limited mass-balance observations, such as in western Canada. For example, Clarke et al. noted that simplified degree-day melt models are used in part because there is "insufficient knowledge to justify anything more complex"[8]. In addition, there are longstanding challenges in hydrological modelling, such as equifinality and the incorporation of complex but poorly represented processes such as lateral groundwater flow[32]. Our method is advantageous as it circumvents these challenges. The method alone, however, is not designed to quantify the rate at which streamflow dynamics are expected to change. We suggest that our approach be used to complement modelling studies in two ways: first, the identified glacier runoff signatures in streamflow can serve as reference data for model validation, and second, local-scale municipal water source data should be incorporated to understand the actual vulnerability of human populations to glacier retreat. Future research should investigate community resilience (for example, coping and adaptive capacity) to water resource vulnerability.

We chose to regress onto the mean PCs of the first PCA mode and not a more physical quantity, such as mean August streamflow, in part because PCA modes form a linearly independent basis from which we can reconstruct streamflow. To test how sensitive our results are to this choice, we also regress onto mean-normalized August streamflow instead of $\overline{PC_1}$. This test leads to the same key conclusions and findings as derived with the original regression (Fig. 3): first, that glacier-fed stations have uniquely high mean-normalized flow with relatively low year-to-year variability; second, that mean-normalized August flow increases nonlinearly with percentage glacierization; third, that the same locations are identified as having the most-vulnerable water sources; and fourth, that these locations face varying degrees of vulnerability in terms of a reduction of mean flow and an increase in year-to-year variability. These conclusions are consistent between the two approaches because the PCs of the first mode are highly correlated with mean-normalized August discharge (Supplementary Fig. 2) and the first PCA mode alone explains a sufficient amount of variance for reconstruction (Supplementary Fig. 1). As these two PCA characteristics may not be representative for other case studies, we recommend the use of PCA modes rather than selected physical quantities in the regression models.

We have designed a new way to quantify the importance of glacial meltwater contributions to each river in Alberta, and we provide locally resolved projections of streamflow to identify vulnerable water resources. We identify the most-vulnerable locations in Alberta, including the Bighorn Dam, the largest reservoir in the province, which over one million people depend on. We also show that including local water source data plays a substantial role in determining whether communities' water resources are vulnerable in glacierized basins. Our approach is transferable and relatively simple, requiring streamflow and climate data across a region, and can serve as a template for studies wherever people depend on glacier-fed water sources, such as the Andes, Alps and Himalayas.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41558-020-0863-4.

Received: 21 June 2019; Accepted: 6 July 2020;
Published online: 3 August 2020

References

1. Barnett, T. P., Adam, J. C. & Lettenmaier, D. P. Potential impacts of a warming climate on water availability in snow-dominated regions. *Nature* **438**, 303–309 (2005).
2. Huss, M. & Hock, R. Global-scale hydrological response to future glacier mass loss. *Nat. Clim. Change* **8**, 135–140 (2018).
3. Immerzeel, W. W., van Beek, L. P. H. & Bierkens, M. F. P. Climate change will affect the Asian water towers. *Science* **328**, 1382–1385 (2010).
4. Kaser, G., Grosshauser, M. & Marzeion, B. Contribution potential of glaciers to water availability in different climate regimes. *Proc. Natl Acad. Sci. USA* **107**, 20223–20227 (2010).
5. Pritchard, H. D. Asia's shrinking glaciers protect large populations from drought stress. *Nature* **569**, 649–654 (2019).
6. Milner, A. M. et al. Glacier shrinkage driving global changes in downstream systems. *Proc. Natl Acad. Sci. USA* **114**, 9770–9778 (2017).
7. Marshall, S. J. et al. Glacier water resources on the eastern slopes of the Canadian Rocky Mountains. *Can. Water Resour. J.* **36**, 109–134 (2011).
8. Clarke, G. K. C., Jarosch, A. H., Anslow, F. S., Radić, V. & Menounos, B. Projected deglaciation of western Canada in the twenty-first century. *Nat. Geosci.* **8**, 372–377 (2015).
9. Ebrahimi, S. & Marshall, S. J. Parameterization of incoming longwave radiation at glacier sites in the Canadian Rocky Mountains. *J. Geophys. Res. Atmos.* **120**, 12536–12556 (2015).
10. Fitzpatrick, N., Radić, V. & Menounos, B. Surface energy balance closure and turbulent flux parameterization on a mid-latitude mountain glacier, Purcell Mountains, Canada. *Front. Earth Sci.* **5**, 67 (2017).
11. Gascoin, S. et al. Glacier contribution to streamflow in two headwaters of the Huasco River, Dry Andes of Chile. *Cryosphere* **5**, 1099–1113 (2011).
12. Bliss, A., Hock, R. & Radić, V. Global response of glacier runoff to twenty-first century climate change. *J. Geophys. Res. Earth Surf.* **119**, 717–730 (2014).

13. Comeau, L. E. L., Pietroniro, A. & Demuth, M. N. Glacier contribution to the North and South Saskatchewan Rivers. *Hydrol. Process.* **23**, 2640–2653 (2009).
14. Jost, G., Moore, R. D., Menounos, B. & Wheate, R. Quantifying the contribution of glacier runoff to streamflow in the upper Columbia River Basin, Canada. *Hydrol. Earth Syst. Sci.* **16**, 849–860 (2012).
15. Naz, B. S., Frans, C. D., Clarke, G. K. C., Burns, P. & Lettenmaier, D. P. Modeling the effect of glacier recession on streamflow response using a coupled glacio-hydrological model. *Hydrol. Earth Syst. Sci.* **18**, 787–802 (2014).
16. Soruco, A. et al. Contribution of glacier runoff to water resources of La Paz city, Bolivia (16° S). *Ann. Glaciol.* **56**, 147–154 (2015).
17. Greve, P. et al. Global assessment of water challenges under uncertainty in water scarcity projections. *Nat. Sustain.* **1**, 486–494 (2018).
18. Flörke, M., Schneider, C. & McDonald, R. I. Water competition between cities and agriculture driven by climate change and urban growth. *Nat. Sustain.* **1**, 51–58 (2018).
19. Hoekstra, A. Y. Water scarcity challenges to business. *Nat. Clim. Change* **4**, 318–320 (2014).
20. Farinotti, D., Usselmann, S., Huss, M., Bauder, A. & Funk, M. Runoff evolution in the Swiss Alps: projections for selected high-alpine catchments based on ENSEMBLES scenarios. *Hydrol. Process.* **26**, 1909–1924 (2012).
21. Hagg, W., Hoelzle, M., Wagner, S., Mayr, E. & Klose, Z. Glacier and runoff changes in the Rukhk catchment, upper Amu-Darya basin until 2050. *Glob. Planet. Change* **110**, 62–73 (2013).
22. Schindler, D. W. & Donahue, W. F. An impending water crisis in Canada's western prairie provinces. *Proc. Natl Acad. Sci. USA* **103**, 7210–7216 (2006).
23. Downing, D. & Pettapiece, W. *Natural Regions and Subregions of Alberta* (Natural Regions Committee, 2006).
24. Dee, D. P. et al. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.* **137**, 553–597 (2011).
25. Marshall, S. J. Meltwater run-off from Haig Glacier, Canadian Rocky Mountains, 2002–2013. *Hydrol. Earth Syst. Sci.* **18**, 5181–5200 (2014).
26. Demuth, M. & Keller, R. in *Peyto Glacier: One Century of Science* (eds Demuth, M. et al.) 83–132 (Environment Canada, 2006).
27. RGI Consortium *Randolph Glacier Inventory (RGI)—A Dataset of Global Glacier Outlines* (GLIMS, 2017); https://doi.org/10.7265/N5-RGI-60
28. Bash, E. A. & Marshall, S. J. Estimation of glacial melt contributions to the Bow River, Alberta, Canada, using a radiation–temperature melt model. *Ann. Glaciol.* **55**, 138–152 (2014).
29. Moore, R. D. et al. Glacier change in western North America: influences on hydrology, geomorphic hazards and water quality. *Hydrol. Process.* **23**, 42–61 (2009).
30. *Human Activity and the Environment: Freshwater in Canada. Section 2: Freshwater Supply and Demand* (Statistics Canada, 2017).
31. Kohonen, T. Self-organized formation of topologically correct feature maps. *Biol. Cybern.* **43**, 59–69 (1982).
32. Shen, C. A transdisciplinary review of deep learning research and its relevance for water resources scientists. *Water Resour. Res.* **54**, 8558–8593 (2018).

Methods

Data and pre-processing. Historical streamflow was extracted from the Environment and Climate Change Canada Historical Hydrometric Data (HYDAT) website on September 19, 2018 (https://wateroffice.ec.gc.ca/mainmenu/historical_data_index_e.html). There are 194 stations that have data from 1987 to 2010 and are upstream of dams. Of these, 18 are glacier fed; the rest are non-glacierized catchments sourced in the Rocky Mountains, are tributaries or larger river stems, or are sourced at non-glacier-fed lakes.

The normalization process is shown in Supplementary Fig. 4 and is described here. Each year of daily streamflow is normalized at each station by subtracting that year's mean, dividing by that year's standard deviation, and smoothing each year using a 30-day running mean. 'Year' refers to calendar year (January 1 to December 31). Under this scheme, values of normalized streamflow indicate how substantial each 30-day window of streamflow is as compared with the year's average flow at that station. We then only use days in August from the smoothed normalized streamflow dataset; however, because we use a 30-day window for smoothing, 'August' streamflow comprises observation as early as July 17 and as late as September 14 each year.

Municipal water source data are collected for 567 communities in Alberta. For communities that source water from a glacier-fed river, we also identify where along the glacier-fed river they intake water. In many cases, a single water treatment facility will pull water from a glacier-fed river, treat the water and then distribute the water to multiple different communities. For these communities, the water intake location is taken to be at the intake of the water treatment facility. This is shown by the black lines in Fig. 1c that connect multiple glacier-fed communities to the same point along a glacier-fed river. For communities that source water from a glacier-fed river but are not part of one of these water distribution networks, we assume that the water intake location is at the location on the glacier-fed river that is closest to the community.

Municipal water source data were collected from various online sources such as municipality websites, utility provider websites and regional freshwater reports or by directly contacting municipal governments and utility providers. These data are sometimes provided on subregional scales, such as for a particular subbasin or for communities that all connect to the same water treatment and distribution plant; however, to our knowledge, no provincial-wide inventory of these data exists. The lack of a province-wide publicly available dataset was confirmed by the Alberta WaterPortal Society (personal communication, S. Shane). Our collected water supply data are made publicly available (https://doi.org/10.5281/zenodo.3266447).

Percentage glaciation at a location is defined as the total glacier area in the basin divided by the area of the basin. Glacier areas and outlines are taken from the Randolph Glacier Inventory v.6.0[27]. Glacier runoff pathways are calculated using a D8 flow model[33], starting from the lowest elevation point on the glacier. Total glacier area upstream of a location is calculated by summing the areas of all glaciers that have runoff routed through that location. Total basin area upstream of a location is calculated by multiplying the flow accumulation (from the D8 flow model) by the area of each grid cell. We use a 90 m digital elevation model from the Shuttle Radar Topography Mission[34].

We use daily 2-m air temperature (short name: 't2'; parameter ID: 167), total precipitation (short name: 'tp'; parameter ID: 228) and evaporation (short name: 'e'; parameter ID: 182) from ERA-Interim reanalysis[24], downloaded at 0.75°×0.75° spatial resolution.

PCA and SOM. In step (1) described in the main text, we apply PCA on the normalized and smoothed time series of daily August streamflow for all 194 stations, finding the linearly independent modes that explain the maximum amount of variability. In our case, there are 31 variables (31 days of normalized streamflow values) and 4,656 observations (194 stream gauges, each with 24 years of August streamflow from 1987 to 2010). This means that there are 31 eigenvectors, each with length of 31. The first eigenvector explains the most variance in the dataset, and the second eigenvector explains the second-most variance in the dataset. The PCs of each mode have 4,656 points (one for each observation) and represent how strongly weighted a particular mode is in each observation. So for each stream gauge, the PC_1 time series has length 24, where values describe how strongly weighted the first eigenvector is in each August; $\overline{PC_1}$ is the average PC_1 value over 1987–2010 for each station and represents how strongly weighted the first eigenvector is in each August on average.

The eigenvectors of the first two modes are shown in Fig. 2a, and the percentage of variance explained by each mode is shown in Supplementary Fig. 1. The first mode is described in the main text. The second mode describes 10% of the total variance, and its eigenvector shows a pattern that decreases monotonically throughout the month. This pattern is related to the steepness of August flow, indicating how fast the streamflow is changing over the month. PC_2 is highly negatively correlated (correlation <-0.95) with the slope of a trendline of normalized August streamflow (Supplementary Fig. 2b), meaning that high PC_2 values correspond to streamflow that more rapidly decreases over August.

Next, we cluster the stream gauges, each represented by the following vectors:

$$\mathbf{v_k}(t) = \mathrm{PC}_1^k(t)\hat{\mathbf{i}} + \mathrm{PC}_2^k(t)\hat{\mathbf{j}} \quad (1)$$

where $\mathbf{v_k}(t)$ is the vector of stream gauge k to be clustered, $\mathrm{PC}_1^k(t)$ is the PC_1 time series for stream gauge k, $\mathrm{PC}_2^k(t)$ is the PC_2 time series for stream gauge k, and $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$ are orthonormal unit vectors. All of $\mathbf{v_k}(t)$, $\mathrm{PC}_1^k(t)$ and $\mathrm{PC}_2^k(t)$ are dimensionless and have size 24×1 since there are 24 years of observations.

We use SOMs[31] to cluster the vectors $\mathbf{v_k}(t)$ according to their similarity. SOMs are a type of artificial neural network that uses unsupervised learning to cluster the input data onto a two-dimensional map where similar clusters are placed closer together, while dissimilar ones are placed further apart. Similar to other clustering methods (for example, K-means, hierarchical clustering), the optimal number of clusters (or size of SOM) is challenging to determine as there is no universally applicable metric involved in the decision making[35]. Here, we build upon the method of Unglert et al. to identify the optimal size of SOM[36], and we use a two-dimensional colourmap from Steiger et al[37]. We first map the input data to a large SOM, with size (final number of clusters) calculated from the input dataset length and the ratio of the two largest eigenvalues for the dataset[38] (Supplementary Fig. 5a). We perform PCA on the SOM patterns and build a topology defined as the sum of the squares of the PCs of the first two modes (Supplementary Fig. 5b). We determine the number of characteristic patterns to be the number of local maxima and global minima on this topology. Nodes at local maxima are distinct patterns that are reconstructed by the first two eigenvectors in unique ways, while the node at the global minima represents the pattern that is captured the least by the first two eigenvectors. The number of characteristic patterns and the ratio of the two largest eigenvalues in the dataset is then used to compute the size of the final SOM.

We note that the number of clusters chosen does not directly affect our final results. We use clustering to learn that non-glacier-fed rivers have similar August streamflow patterns as other nearby non-glacier-fed rivers, and that glacier-fed rivers have unique August streamflow patterns as compared with other nearby non-glacier-fed rivers. This result remains true when using a smaller (for example, 3×1 SOM) or larger (for example, 3×3 SOM) number of clusters (Supplementary Fig. 6).

Regression models. We aim to predict the first PCA mode rather than a more physical quantity (such as mean August streamflow) because (1) the first PCA mode describes the maximum amount of variance within the dataset that could be described by any one variable, and (2) in this case, the first mode alone explains enough variance that it alone can be used to reconstruct August streamflow. Instead of prescribing the features of importance, we use PCA to find the characteristic features that can be used to reconstruct our original variables (31 days of August streamflow), and now we aim to predict these characteristic features.

To develop the MLR models, we aimed to select predictors that are physically linked to the driving inflow and outflow processes of streamflow in the region: rain, snowmelt, evaporation and groundwater flow. Here we briefly outline our motivation for each predictor used.

- T_{year}: Colder regions may allow snow storage to persist through summer, while warmer regions may not.
- P_{year}: Regions with high yearly precipitation may have large snowpacks (which may provide water storage through summer), plenty of summer rain, or both.
- T_{JJA}: Regions with high summer temperatures may have less storage of water as ice and snow.
- P_{JJA}: Summer precipitation is a key component of the summer water balance, and we expect stations in wetter regions to have a higher mean flow.
- E_{JJA}: Evaporation is a key outflow process of summer streamflow and even exceeds precipitation in some parts of Alberta[22].
- $\mathrm{Log}[Q_{min}/Q_{max}]$: The ratio of minimum seasonal flow to maximum seasonal flow is a proxy for the importance of groundwater contributions that maintain high flows in the dry season relative to the rest of the year (Supplementary Fig. 7). This ratio spans several orders of magnitude, and so we use the logarithm of it.
- Elevation: Higher elevations tend to receive more precipitation and have lower temperatures, which allows for the generation of a larger snowpack that may persist longer through the summer.
- Stream order and drainage area: Both are determinants of runoff, where streams of higher order or with larger drainage area are expected to have more runoff than lower-order streams or catchments with smaller area in the same watershed[39]. As drainage areas in our dataset span several orders of magnitude, we use the logarithm of drainage area as a potential predictor.

Each predictor is normalized to have zero mean and unity variance. We use stepwise regression to determine the statistically significant predictors used in the MLR models. We only keep predictors with $p<0.01$. We test for multicollinearity by calculating variance inflation factors (VIF), and we keep predictors with $\mathrm{VIF}<5$. If two or more predictors do not meet this criterion, we remove the predictor with the greatest VIF, and then recalculate until all predictors have $\mathrm{VIF}<5$.

We assess the robustness of the stepwise regression approach by using an iterative bootstrapping method to select a random subset of stations for calibration/validation. For each iteration, we make MLR models with all possible combinations of predictors, and we calculate the best-performing model. We find that the predictors that are most often used to make these best-performing models

are those that were selected by stepwise regression. The final predictors are shown in Supplementary Fig. 3 and Equations (2) and (3).

With the remaining predictors, we use a bootstrapping approach to calculate regression coefficients. Distributions of MLR coefficients are shown in Supplementary Fig. 8. The final MLR models used are as follows:

$$\overline{\mathrm{PC_1}} = -0.34 - 0.60\times T_{\mathrm{year}} + 0.96\times P_{\mathrm{JJA}} + 0.87\times T_{\mathrm{JJA}} + 0.44\times E_{\mathrm{JJA}} + 0.41\times \log_{10}\left[\frac{Q_{\mathrm{min}}}{Q_{\mathrm{max}}}\right] \quad (2)$$

$$\sigma_{\mathrm{PC_1}} = 2.49 + 0.52\times P_{\mathrm{JJA}} + 0.26\times E_{\mathrm{JJA}} - 0.68\times h + 0.42\times \log_{10}\left[\frac{Q_{\mathrm{min}}}{Q_{\mathrm{max}}}\right] \quad (3)$$

Root mean squared errors between observed and modelled $\overline{\mathrm{PC_1}}$ and σ_{PC1} are larger by factors of nearly five and two, respectively, when comparing glacier-fed with non-glacier-fed stations. Since the MLR models perform poorly for glacier-fed stations, we analysed other potential drivers of spatial variability for these stations. Under the hypothesis that glaciation is driving the unique $\mathrm{PC_1}$ behaviour, but with only a sparse amount of climatological data at the local (glacier) scale, we consider three potential predictor variables that are available globally[27]: the hypsometric index[40] averaged over contributing glaciers in each basin, the median glacier elevation averaged over contributing glaciers in each basin and the natural logarithm of percentage glaciation of each basin. We find that only the logarithm of percentage glaciation is significant ($p<0.01$) and use linear regression to make the following model:

$$\overline{\mathrm{PC_1}} = 11.58 + 1.74\times \ln(G) \quad (4)$$

where G is the percentage glaciation of a basin (between 0 and 1). This ordinary linear regression model explains 70% of the variability in $\overline{\mathrm{PC_1}}$ for glacier-fed stations (Fig. 3c). Due to worse performance of regression models for estimating σ_{PC1} and since σ_{PC1} has a narrow range for glacier-fed rivers relative to non-glacier-fed rivers (cluster no. 3 in Fig. 2b), we estimate σ_{PC1} as the mean standard deviation of $\mathrm{PC_1}$ of all measured glacier-fed rivers.

As a sensitivity test, we rerun the analysis but regress onto mean (normalized) August streamflow. This test leads to the same key conclusions as when we regress onto $\overline{\mathrm{PC_1}}$ (Fig. 3 and Supplementary Fig. 9). However, there is a slight difference in the interpretation of the final results. When regressing onto $\overline{\mathrm{PC_1}}$, the smoothed, normalized August streamflow is reconstructed by multiplying $\mathrm{PC_1}$ by the first eigenvector. Therefore, the probability density functions (Fig. 3d) represent the probability that any day in the smoothed, normalized August streamflow vector has that value of streamflow. When regressing onto the mean-normalized streamflow directly, we are not reconstructing the entire August vector; instead, we are estimating the mean, rather than each value in the 31-day vector. Therefore, the probability density functions (Supplementary Fig. 9d) represent the probability that the mean-normalized August streamflow has that value, not the probability that any day in the smoothed, normalized August streamflow vector has that value.

Vulnerability assessment. We investigate risk for rivers that are sufficiently glaciated as to expect glacier forcing to be substantial; here, we define 'sufficiently glaciated' as basins that have a percentage glaciation greater than 1%, as all stations above this glaciation are grouped in cluster no. 3 by the SOM. For these stations, we predict $\overline{\mathrm{PC_1}}$ by the ordinary linear regression model.

We reconstruct streamflow using only the first mode as it describes 88% of the total variance. For each location and each scenario ('no glaciers in basin' as calculated by Equations (2) and (3), and 'glaciers in basin' as calculated by Equation (4)), we have modelled values of $\overline{\mathrm{PC_1}}$ and σ_{PC1}. We then iteratively reconstruct streamflow, randomly selecting $\mathrm{PC_1}$ values from a normalized distribution determined by the modelled $\overline{\mathrm{PC_1}}$ and σ_{PC1}. We calculate normalized streamflow for each iteration by:

$$\mathbf{Q}^{j}_{\mathrm{norm},i} = \mathrm{PC}^{j}_{1,i}\times \mathbf{e}_1 + \mathbf{Q}_{\mathrm{mean}} \quad (5)$$

where $\mathbf{Q}^{\mathbf{j}}_{\mathrm{norm,i}}$ is the normalized streamflow of the i-th iteration at location j, $\mathrm{PC}^{j}_{1,i}$ is the $\mathrm{PC_1}$ value drawn during the i-th iteration from the $\mathrm{PC_1}$ distribution modelled at location j, $\mathbf{e_1}$ is the eigenvector of the first mode, and $\mathbf{Q}_{\mathrm{mean}}$ is the mean streamflow vector (length of 31) calculated across all stations and years. We iterate 10,000 times for each location. We then calculate the probability distribution of normalized August streamflow for each location by kernel density estimation, using Gaussian kernels[41].

We calculate the difference between modelled mean streamflow (with the MLR models) and observed mean streamflow for all non-glacier-fed stations. The standard deviation of this difference ($\sigma_{\Delta Q}$) is a measure of the error of the MLR models. For communities that source water from a glacier-fed river (>1% glaciation), we calculate the difference in normalized mean streamflow for the two scenarios (ΔQ_c). We consider the vulnerability to be significant when $\frac{\Delta Q_c}{\sigma_{\Delta Q}} > 2$ (that is, where the predicted change in mean streamflow is more than twice the error from the MLR model), and we find that the communities discussed in the main text have $\frac{\Delta Q_c}{\sigma_{\Delta Q}} > 3$.

Uncertainty analysis. Uncertainty in the PCA analysis is tested by performing PCA and clustering with SOM, but using only non-glacier-fed rivers. We find the eigenvectors, eigenvalues and PCs to be very similar to how they appeared in the analysis of all rivers: for example, the eigenvectors of the first mode have a correlation of 0.98, and the eigenvectors of the second mode have a correlation greater than 0.99. This indicates that the eigenvectors, eigenvalues and PCs discussed in the preceding are not biased by the presence of glacier-fed rivers. This also indicates that the features we find describe the behaviour of all rivers in Alberta and do not only describe the difference between glacier-fed and non-glacier-fed rivers. In this analysis of non-glacier-fed rivers, the SOM method did not produce a cluster that had the characteristic streamflow features previously found in glacier-fed rivers.

We also perform PCA and clustering with SOM, but only using non-nested catchments. We again find the eigenvectors, eigenvalues and PCs to be very similar to their equivalents in the analysis of all rivers, with the eigenvectors of the first and second modes from non-nested streams having correlation >0.99 with the eigenvectors from all streams. The SOM clusters again identify a glacier-fed cluster that is characterized by consistently high $\mathrm{PC_1}$ values, while non-glacier-fed rivers have similar August streamflow patterns as other nearby non-glacier-fed rivers.

Like glaciers, lake systems also act as natural reservoirs. In Fig. 3c, two stream gauges are lightly glaciated (<0.3% glaciation) and do not belong to cluster no. 3 since they have comparatively low $\overline{\mathrm{PC_1}}$ values (indicating negligible glacier contributions). Three stations are lightly glaciated (<0.3% glaciation), do belong to cluster no. 3 and do have comparatively high $\overline{\mathrm{PC_1}}$ values, suggesting that high $\overline{\mathrm{PC_1}}$ values can be achieved with negligible glacier contributions. However, these three rivers are downstream of lake systems, and we suggest that lakes are the cause of the anomalously high $\overline{\mathrm{PC_1}}$ values.

Data availability

All data are publicly available. ERA-Interim reanalysis is available from ECMWF[24]. Glacier inventory is available from the RGI[27]. Topography is available from the SRTM[34]. Streamflow is available from the Environment Canada HYDAT database[42]. Municipal water source data, collected in this study, are available on Zenodo (https://doi.org/10.5281/zenodo.3266447)[43].

Code availability

Code to reproduce the main results in this study is available on GitHub (https://doi.org/10.5281/zenodo.3742162)[44].

References

33. Fairfield, J. & Leymarie, P. Drainage networks from grid digital elevation models. *Water Resour. Res.* **27**, 709–717 (1991).
34. Farr, T. G. et al. The Shuttle Radar Topography Mission. *Rev. Geophys.* **45**, RG2004 (2007).
35. Hsieh, W. W. *Machine Learning Methods in the Environmental Sciences: Neural Networks and Kernels* (Cambridge Univ. Press, 2009).
36. Unglert, K., Radić, V. & Jellinek, A. M. Principal component analysis vs. self-organizing maps combined with hierarchical clustering for pattern recognition in volcano seismic spectra. *J. Volcanol. Geotherm. Res.* **320**, 58–74 (2016).
37. Steiger, M. et al. Explorative analysis of 2D color maps. In *Proc. WSCG 2015 Conference on Computer Graphic, Visualization, and Computer Vision* (eds Gavrilova, M. & Skala, V.) 151–160 (Union Angency, 2015).
38. Vesanto, J., Himberg, J., Alhoniemi, E. & Parhankangas, J. *SOM Toolbox for Matlab 5* Report A57 (Helsinki Univ. Technol., 2000).
39. Strahler, A. N. Quantitative analysis of watershed geomorphology. *EOS* **38**, 913–920 (1957).
40. Jiskoot, H., Curran, C. J., Tessler, D. L. & Shenton, L. R. Changes in Clemenceau Icefield and Chaba Group glaciers, Canada, related to hypsometry, tributary detachment, length–slope and area–aspect relations. *Ann. Glaciol.* **50**, 133–143 (2009).
41. Silverman, B. *Density Estimation for Statistics and Data Analysis* (Chapman & Hall, 1986).
42. *Water Survey of Canada HYDAT Data* (Environment Canada, 2018); https://wateroffice.ec.gc.ca/mainmenu/historical_data_index_e.html
43. Anderson, S. Alberta municipal water supply overview. *Zenodo* https://doi.org/10.5281/zenodo.3266447 (2019).
44. Anderson, S. andersonsam/pca_som_streamflow: first release. *Zenodo* https://doi.org/10.5281/zenodo.3742162 (2020).

Acknowledgements

We thank Environment Canada for providing streamflow data, the European Centre for Medium-range Weather Forecasts (ECMWF) for the ERA-Interim reanalysis data, the Randolph Glacier Inventory consortium for glacier inventory data and the Shuttle Radar Topography Mission for providing topographic data. We also thank M. Jellinek for providing feedback on the manuscript. Funding supporting this study was provided through the Natural Sciences and Engineering Research Council (NSERC) of Canada.

Author contributions

S.A. gathered and processed the data. S.A. developed the analysis and made the figures, both with input from V.R. S.A. and V.R. shared the writing of the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41558-020-0863-4.

Correspondence and requests for materials should be addressed to S.A.

Peer review information *Nature Climate Change* thanks Shawn Marshall and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.