Development of Climate Analogs for Major Canadian Cities

Final Report

May 2022

Environment and Climate Change Canada
Environnement et Changement climatique Canada

Development of Climate Analogs for Major Canadian Cities.

Final Report
May 2022

AUTHOR(S):
Pascal Bourgault, Ouranos inc.
Travis Logan, Ouranos inc.

COLLABORATORS (by alphabetical order):
Elaine Barrow, Centre canadien des services climatiques, Environnement et Changement climatique Canada
Emilia Diaconescu, Centre canadien des services climatiques, Environnement et Changement climatique Canada
David Huard, Ouranos inc.

Title of the project: Développement d'analogues climatiques pour les grandes villes canadiennes

Number of the project: 706400

Suggested quote: Bourgault, P., Logan, T. (2022). *Development of Climate Analogs for Major Canadian Cities.* Report submitted to Environment and Climate Change Canada, Montréal. 35 p.

ACKNOWLEDGMENTS

This project was made possible thanks to funding from Environment and Climate Change Canada (ECCC) (contract no. 30000729377).

The authors wish to thank Emilia Diaconescu and Elaine Barrow, from ECCC, for their collaboration in this project.

Environment and Climate Change Canada   Environnement et Changement climatique Canada

We acknowledge the World Climate Research Programme, which, through its Working Group on Coupled Modelling, coordinated and promoted CMIP6. We thank the climate modeling groups for producing and making available their model output, the Earth System Grid Federation (ESGF) for archiving the data and providing access, and the multiple funding agencies who support CMIP6 and ESGF.

Data access and analysis environment for this work is provided through the Power Analytics and Visualization for Climate Science (PAVICS) platform (Ouranos and CRIM, 2022). PAVICS is funded through Ouranos, the Computer Research Institute of Montreal (CRIM), CANARIE, the Fonds Vert, the Canadian Foundation for Innovation (CFI), Environment and Climate Change Canada (ECCC) and the Fonds de Recherche du Québec (FRQ).

Generated using Copernicus Climate Change Service Information [2022].

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

CCCS	Canadian Centre for Climate Services
CMIP	Coupled Model Intercomparison Project
CORDEX	Coordinated Regional Downscaling Experiment
DQM	Detrended Quantile Mapping
ECCC	Environment and Climate Change Canada
ECMWF	European Centre for Medium-Range Weather Forecasts
ERA5	ECMWF ReAnalysis
GCM	Global Climate Model
GPW	Gridded Population of the World
NRCan	Natural Resources Canada
UN-WPP	United Nations - World Population Prospects

1. INTRODUCTION

The effects of climate change and the need to implement adaptation strategies in order to minimize potential impacts are more and more evident in Canada and around the world. Despite the existence of numerous scientific evaluations, synthesis reports and other climate change information for projected changes these products are typically produced following a scientific approach providing information at a global or continental scale making them difficult to integrate at a local scale. In 2018, the Government of Canada put in place the Canadian Centre for Climate Services (CCCS) as part of Environment and Climate Change Canada (ECCC) with the goal of providing access to relevant climate data, information and support in order to facilitate comprehension of climate change risks and impacts as well as assist development and implementation of climate change adaptation strategies across Canada.

In relation to the present contract, contemporary spatial climate analogues for Canadian cities could be an effective approach for communicating climate change projections to the public. More specifically the spatial analog approach works by identifying contemporary locations with similar climates (temperature, precipitation, or other climate variables and indices) to those projected by climate models for Canadian cities. As such, abstract projections are translated into clear, real-world examples. The tool developed here allows users to make customized calculations of spatial climate analogues for major Canadian cities offering much more flexibility compared to typical static map products where choice of included variables is generally predetermined for all sites in advance. Note that more intensive testing with potential users is needed to ensure the tool produces pertinent information to decision makers and that this information is interpreted correctly. Development of proper guidance for the use of the tool via user testing is recommended to avoid misuse.

As specified in the Statement of work, the main objective of the proposed work is to develop a tool to allow custom computation and visualization of contemporary climate analogs for major Canadian cities for the CCCS. Ouranos will provide the code developed and results for at least one major city per province and territory, for two climate scenarios covering the 1991-2100 period. The CCCS plans to potentially use the deliverables of this project to develop a Web tool for Canadians to help in the general effort to adapt to climate change.

The contract deliverables include:

- Production of climate indices datasets
    - Climate indices of the reference dataset
    - Climate indices of the CMIP6 bias-adjusted climate scenarios
- Development of the analog-finder code and a visualization prototype
    - Analog-finder Jupyter notebook
    - Visualization prototype
- A technical report describing the datasets, methodology and visualization choices.

2. DATASET PRODUCTION

2.1 POPULATION DENSITY

In a context where analogues will potentially be used in the formulation of adaptation strategies a non-urban analogue for a target city, despite having a very good fit in terms of climatic conditions, could remain of little use if goals include the creation hypotheses with respect to potential future impacts in an urban setting (e.g. urban flooding or heat waves) and possible solutions (e.g. comparison of building codes or green infrastructure). As such, in order to constrain the analogue search to urban areas as similar as possible to the target, a gridded population density data layer was used in order to filter potential analogs candidate locations. As such the analog-finder code only considers points in the reference dataset where the population density is within a range defined around the target's density. As population density information is not provided within the reference dataset, we use the 4th version of the Gridded Population of the World dataset (GPWv4), developed by the Center for International Earth Science Information Network of the Columbia University. This collection of datasets includes very high resolution (30 arcsec) data of population count and density. The selected version is adjusted so that 2015 country totals match the United Nations' 2015 World Population Prospects (UN WPP) (CIESIN, 2018a). GPWv4 uses numerous data sources from 2005 to 2014 and employs methods to extrapolate results in order to give population estimates up to 2020, the year we used in the final deliverable.

As GPWv4 population density data is adjusted to reflect the density only on the land area of the pixel, it cannot be regridded directly to the resolution of the climate reference dataset. We instead started from the population count and used the high-resolution land area fraction also provided by GPWv4 (CIESIN, 2018b) to perform a precise conservative regridding to the reference dataset grid producing a final map of the population density of the land-covered portion of each grid cell. The procedures are described in detail in Appendix A.

Code for this section is found in the Masques.ipynb notebook, in the GitHub code repository.

2.2 TARGET CITIES

While the domain for the analogue search will be all urban areas of Canada and the United States, the client (CCCS) provided the coordinates (latitude and longitude) of the target Canadian cities and towns, considering a fixed and relevant criterion (e.g., location of the City Hall). The list was presented per province and territory as a priority ranking list (up to 5 cities per province and territory), which served as a reference. Initial plans included the analysis of a single city per province and territory as a first step for the project. In the end the number of sites was not an obstacle for the development of the datasets and code, and so all 65 locations were provided (**Table 1**).

When computing the adjusted density map and validating it against public demographic data, a few issues with the city selection were raised and discussed with the CCCS.

A. 4 coastal cities had their city hall location within an "ocean" pixel of the reference dataset, which only has data on land.
B. A few city hall locations did not fall within the cell with the highest population density of their urban area (using our adjusted density).
C. The western part of Newfoundland had no selected location.
D. Many small cities have urban areas significantly smaller than the grid cell size.

Except for case A, these issues do not yield computational problems, but may affect how the results are interpreted or how the tool is used. To address issues A and B, we recomputed the city locations as the center of the land-covered cell with the highest population density within 3x3 square around the specified city hall location. For issue C, Conception Bay, which is quite close to St John's, was replaced by Corner Brook. The only impact of issue D is that the adjusted density on the map is much lower than that reported for the urban area only. No real solution is available without artificially adjusting the density map, which would be quite complex to apply over all of North America.

City data is processed by some sections of the `Masques.ipynb` *notebook and stored in* `geo/cities.nc` *or* `geo/cities.geojson`*, as well as on PAVICS's geoserver.*

Table 1 Selected cities

Province / territory	City	Population	City hall coords		Choice explanation
Alberta	Calgary	1239220	-114,05	51,05	The four largest population centres in Alberta.
	Edmonton	932546	-113,49	53,55	
	Red Deer	100418	-113,81	52,27	
	Lethbridge	92729	-112,83	49,69	
	Grande Prairie	63166	-118,79	55,17	Largest Population centre in North Alberta
British Columbia	Vancouver	631486	-123,12	49,26	Largest population centre in Metro Vancouver district
	Kelowna	127380	-119,50	49,89	Largest population centre in Central Okanagan district
	Abbotsford	141397	-122,33	49,05	Largest population centre in Fraser Valley district
	Victoria	85792	-123,36	48,43	The capital
	Prince George	74003	-122,74	53,91	The largest population centre in Fraser-Fort George district
Manitoba	Winnipeg	705244	-97,14	49,90	The five largest population centres in Manitoba
	Brandon	48859	-99,95	49,84	
	Steinbach	15829	-96,69	49,53	
	Thompson	13678	-97,85	55,75	
	Portage la Prairie	13304	-98,29	49,97	
New Brunswick	Moncton	71889	-64,78	46,09	The five largest population centres in New Brunswick
	Saint John	67575	-66,06	45,27	
	Fredericton	58220	-66,64	45,96	
	Dieppe	25384	-64,75	46,10	
	Miramichi	17537	-65,55	47,00	
Newfoundland and Labrador	St. John's	106172	-52,71	47,56	The three largest Population centres in Newfoundland
	~~Conception Bay South~~	~~24848~~	~~-52,97~~	~~47,51~~	
	Mount Pearl	24284	-52,81	47,52	

	Corner Brook	19806	-57,95	48,95	Largest city on the western side of the Island
	Happy Valley-Goose Bay	7552	-60,33	53,30	Two largest Population centres in Labrador
	Labrador City	7367	-66,91	52,95	
Nova Scotia	Halifax	316701	-63,58	44,65	The five largest population centres of Nova Scotia
	Cape Breton - Sydney	29904	-60,20	46,14	
	Truro	22954	-63,28	45,37	
	New Glasgow	18665	-62,64	45,59	
	Glace Bay	17556	-59,95	46,21	
Ontario	Toronto	2731571	-79,38	43,65	The two largest population centers in Ontario
	Ottawa	934243	-75,69	45,42	
	London	494069	-81,25	42,99	The 6th largest population center in Ontario (eliminated suburbs of Toronto)
	Sudbury	161531	-80,99	46,49	Largest city in Northern Ontario
	Moosonee	1481	-80,61	51,29	The largest population center in the "far northern Ontario"
Prince Edward Island	Charlottetown	44739	-63,13	46,24	The 5 largest population centers in PEI
	Summerside	13814	-63,79	46,39	
	Stratford	9706	-63,09	46,23	
	Cornwall	5348	-63,20	46,24	
	Montague	1961	-62,65	46,17	
Québec	Montréal	1762976	-73,55	45,51	The 4 largest population centers in Quebec (eliminated suburbs or Montréal and Ottawa)
	Québec	538738	-71,21	46,81	
	Sherbrooke	165005	-71,89	45,40	
	Saguenay	144989	-71,06	48,43	
	Chibougamau	7504	-74,37	49,91	Largest population center in North Quebec
Saskatchewan	Saskatoon	246376	-106,66	52,13	The 5 largest population centers in Saskatchewan
	Regina	215106	-104,62	50,45	
	Prince Albert	35926	-105,75	53,20	
	Moose Jaw	33890	-105,53	50,39	
	Swift Current	16604	-107,80	50,29	
Northwest Territories	Yellowknife	19234	-114,38	62,45	The 5 largest population centers in the Northwest Territories
	Hay River	3606	-115,79	60,81	
	Inuvik	3463	-133,73	68,36	
	Fort Smith	2093	-111,88	60,01	
	Behchokǫ̀	1926	-116,05	62,83	
Yukon	Whitehorse	25085	-135,05	60,72	The 5 largest population centers in Yukon
	Dawson City	1375	-139,43	64,06	
	Watson Lake	790	-128,72	60,06	
	Haines Junction	613	-137,51	60,75	
	Carmacks	493	-136,29	62,10	
Nunavut	Iqaluit	7082	-68,52	63,75	The 5 largest population centers in Nunavut
	Rankin Inlet	2441	-92,08	62,81	
	Arviat	2318	-94,06	61,11	
	Baker Lake	1872	-96,06	64,31	
	Cambridge Bay	1619	-105,06	69,12	

2.3 CLIMATE INDICES

Determining climate analogs for a range of plausible future climates at a target location requires 1) the selection of annual climate indices (the climate characteristics on which the analogue is based), 2) an ensemble of climate scenarios for these indices for the target cities' future period that reasonably covers the uncertainty in the projected changes, and 3) reference indices for the potential analogs for the recent-past covering the entire search territory of interest. The chosen annual indices are all derived from daily time series of temperature and precipitation. They are listed in **Table 2**, with the variable name to be found in the datasets, a short definition and the type of bias-adjustment, as explained in section 2.5.4.

> *Climate indices are defined in the* `analog.yml` *file, with French metadata translations (unused in this project for now) in* `analog.fr.json`*. All are based on existing* `xclim` *indicators.*

Table 2 Climate indices

Name		Definition	Adj
Total precipitation	RR	The annual total amount of precipitation (rain and snow combined).	×
Hottest day	TXx	The annual maximum of daily maximum temperature (Tmax).	+
Frost days	FD	Annual number of days with minimum daily temperature below 0°C.	×
Growing season length	GSL	The annual number of days between the last occurrence of Tmin < 0°C in spring and the first occurrence of Tmin < 0°C in fall.	×
Growing degree days (5°C)	GDD5	The annual number of degree days accumulated above a threshold temperature of 5°C.	×
Heating degree days	HDD	The annual number of degree days accumulated below 17°C.	×
Last Spring Frost	LSF	The spring date after which there are no daily minimum temperatures during the growing season less than 0°C (Tmin > 0°C).	×
First Fall Frost	FAF	The first date in the fall (or late summer) on which the daily minimum temperature is less than 0°C (Tmin < 0°C).	×
Growing degree days (10°C)	GDD10	The annual number of degree days accumulated above a threshold temperature of 10°C.	×
Degree days above 0°C	GDD0	The annual number of degree days accumulated above a threshold temperature of 0°C.	×
Mean temperature	TG	Annual average of the daily mean temperature (Tmean).	+
Coldest day	TNn	Annual minimum of daily minimum temperature (Tmin).	+
Days with Tmax > 30°C	TXgt30	Annual number of days with Tmax > 30°C.	×
Days with Tmax > 25°C	TXgt25	Annual number of days with Tmax > 25°C.	×

Tropical nights	TNgt22	Annual number of days with the daily minimum temperature (Tmin) greater than 22°C.	×
Cooling degree days	CDD	The annual number of degree days accumulated above 18°C.	×
Maximum 1-day precipitation	RX1day	The largest annual precipitation total that falls in a single day.	×
Wet days	R20mm	The annual number of days with precipitation > 20 mm.	×
Maximum 5-day precipitation	RX5day	The annual maximum total precipitation that falls over a consecutive 5-day period.	×
Ice days	ID	The annual number of days when the daily maximum temperature (Tmax) is less than 0°C.	×

2.4 CLIMATE INDICES FOR THE REFERENCE DATASET

The chosen reference dataset is ERA5-Land (spatial resolution of ~9 km) over the period 1991-2020 (Muñoz Sabater, 2019). The analog search domain, containing all considered potential analogs, will be Canada and the United States. The choice of this reference was oriented by previous and current work being done at Ouranos for the ESPO-R5 project[1]. For this project, multiple reanalysis datasets were compared with ECCC Adjusted and Homogenized Canadian Climate Data (AHCCD) daily station observations across a large set of criteria. ERA5-land was retained as the reference dataset for the ESPO-R5 project after an evaluation of multiple candidate datasets against observed data for the variables of daily maximum and minimum temperatures, and daily total precipitation for the period 1981-2010. The evaluation criteria included:

1. a comparison of the mean annual cycle,
2. an evaluation of the inter-annual seasonal time series
3. a seasonal evaluation of the quantile bias (5, 25, 50, 75, 95) of the daily values between station data and the various candidate datasets

Summary results of these quantitative comparisons indicate that there is no clear winner for the choice of reference dataset, with results varying by season or criteria[1]. However, ERA5-Land was retained for the analog project as it generally shows good results in the above comparisons while also having a high spatial resolution and availability at least up to 2020.

The ERA5-Land hourly temperature and precipitation data were downloaded from the Copernicus Data Store for the 1991-2020 period, subset over North America before being resampled to a daily frequency.

Climate indices were computed on these daily timeseries using xclim (Logan et al., 2022), before being uploaded onto the PAVICS platform. The adjusted population density map was added to this dataset as an extra invariant variable.

Climate indices are computed with the `calc_indices.py` *script.*

[1] Preliminary results for the ESPO-R5 v1.0 can be consulted at https://github.com/Ouranosinc/ESPO-R and as such are not detailed here

2.5 CLIMATE INDICES OF THE CMIP6 BIAS-ADJUSTED CLIMATE SCENARIOS

The future climate scenarios consist of indices computed for bias-adjusted CMIP6 simulations over 1950-2100 (Eyring et al., 2016) for each of the target cities. The same simulation ensemble is used for all target cities. After discussion, it was decided to use the SSP2-4.5 and SSP5-8.5 emission scenarios. To avoid over representation of individual climate models within the ensemble a single realization per climate model was selected for the project

2.5.1 Selection of the CMIP6 simulations

The objective was to obtain a consistent ensemble that was reasonably small to be usable and help the interpretation, while maximizing the climate variability over the chosen cities and indices. We first gathered all CMIP6 realizations that were available on the Pangeo cloud storage,[2] where the three required variables of daily minimum and maximum temperature (tasmax and tasmin respectively) as well as daily total precipitation (pr) are available. We further only selected those simulations where the pairs of experiments (SSP2-4.5 and SSP5-8.5) as well as the common historical run are available. The data was extracted over the 65 cities by bilinear interpolation. At the moment this step was executed, we found 26 models for a total of 211 realizations that fulfilled all those requirements. 14 of these models only provided a single realization (for each emission scenario). The goal of the next step is thus to choose the most appropriate realization for each of the 12 other multi-realization ensembles. **Table 3** shows the number of realizations from which the selected one was taken for each member of the final ensemble.

The ensemble was reduced by using a combination of principal components analysis (PCA) and the KKZ ensemble reduction method.

1. PCA:
    a. The 20 indices are computed on all members
    b. The mean change is computed as the difference between the mean over 2071-2100 and the one over 1991-2020.
    c. Each delta is averaged over the 65 cities, resulting in (20 indices x 2 SSPs =) 40 criteria.
    d. Assuming each criterion is a dimension, we conduct Principal Component Analysis (PCA) on this ensemble of 157 points in order to reduce it to its 3 principal components.
2. KKZ:
    a. The first selected realization is the one closest to the ensemble centroid (in the 3-dimensional space)
    b. The next realization is the one that has the maximal average distance with all previously selected realizations
    c. All the realizations coming from the same model as the selected one are removed from the ensemble.
    d. Repeat b-c until the ensemble is empty.

The KKZ algorithm is based on Cannon, 2015. As the principal components are standardized at end of the "PCA" part, a simple Euclidean metric is used for computing distances in the "KKZ"

[2] https://pangeo-data.github.io/pangeo-cmip6-cloud/

part. With our original datasets, the 40 criteria reduced to 3 components each explaining 75%, 12% and 4% of the ensemble variance, respectively. The final ensemble has 24 members, shown on
Table 3. The ensemble is ordered so that sub-ensembles can easily be chosen. Each member adds less information than the previous one about the overall spread. Two models[3] were left out of the analysis because of corrupted data on the Pangeo store at the moment this analysis was made.

The selection procedure is coded in the `Simulations CMIP6.ipynb` *notebook. The download is performed by the* `extract_cmip6.py` *script and the index calculation by* `calc_indices.py`. *The final list of members is stored in* `config.yml`. *Parts of the process described here were done in an interactive shell and might be absent from the provided code.*

Table 3 CMIP6 simulations ensemble (ordered)

Institution	Source model	Realization id	N realizations
MIROC	MIROC-ES2L	r1i1p1f2	1
EC-Earth-Consortium	EC-Earth3-Veg-LR	r3i1p1f1	3
CCCma	CanESM5	r13i1p1f1	50
IPSL	IPSL-CM6A-LR	r14i1p1f1	6
MRI	MRI-ESM2-0	r1i1p1f1	5
MPI-M	MPI-ESM1-2-LR	r3i1p1f1	10
CAS	FGOALS-g3	r4i1p1f1	3
NOAA-GFDL	GFDL-CM4	r1i1p1f1	1
MIROC	MIROC6	r1i1p1f1	50
NIMS-KMA	KACE-1-0-G	r1i1p1f1	3
NOAA-GFDL	GFDL-ESM4	r1i1p1f1	1
CMCC	CMCC-ESM2	r1i1p1f1	1
NUIST	NESM3	r1i1p1f1	2
INM	INM-CM4-8	r1i1p1f1	1
CSIRO-ARCCSS	ACCESS-CM2	r3i1p1f1	3
NCC	NorESM2-LM	r1i1p1f1	1
EC-Earth-Consortium	EC-Earth3	r125i1p1f1	57
NCC	NorESM2-MM	r1i1p1f1	1
INM	INM-CM5-0	r1i1p1f1	1
BCC	BCC-CSM2-MR	r1i1p1f1	1
CNRM-CERFACS	CNRM-CM6-1	r1i1p1f2	1
CNRM-CERFACS	CNRM-ESM2-1	r1i1p1f2	1
EC-Earth-Consortium	EC-Earth3-Veg	r4i1p1f1	5
EC-Earth-Consortium	EC-Earth3-CC	r1i1p1f1	1

2.5.2 Description of the post-processing method

The adjustment procedure then uses xclim algorithms to adjust simulation bias following a quantile mapping procedure. In particular, the algorithm used is inspired by the "Detrended Quantile Mapping" method described by Cannon et al., 2015. The procedure is univariate,

[3] AS-RCEC-TaiESM1 and KIOST-ESM.

bipartite, acts differently on the trends and the anomalies and is applied iteratively on each day of the year (grouping) and on each target city.

In the following methods descriptions, Y_r is the reference timeseries from ERA5-Land, X_c is the simulated timeseries over the reference period (1991-2020) and X_s is the full simulated timeseries (1950-2100).

2.5.2.1 Variables

Adjustments are applied separately for each of the 3 variables. Note, adjusting tasmax and tasmin independently can lead to physical inconsistencies in the final data (e.g. cases with tasmin > tasmax) (Agbazo & Grenier, 2020; Thrasher et al., 2012). To ensure a resulting dataset that is as physically consistent as possible, within the limitations of univariate bias-adjustment, we compute the daily temperature range (dtr = tasmax - tasmax) and adjust this variable, in addition to tasmax and pr. A final adjusted tasmin variable is reconstructed after the bias-adjustment.

While tasmax has no physical bounds in practice, this is not the case for pr and dtr where a lower bound of zero exists. As such, the adjustment process explained below exists in two cases: additive and multiplicative. In the latter, it is mathematically impossible for adjusted data to fall below zero yet necessitates special pre-processing steps to avoid division by zero (see details below).

2.5.2.2 Bias-adjustment

The bias adjustment acts independently on each day of the year and each grid point. To make the procedure more robust a window of 31 days around the current day of year is included in the inputs of the calibration (training step). For example, the adjustment for February 1 (day 32) is calibrated using data from January 15 to February 15, over the 30 years of the reference period. For leap years, this would mean that there are 4 times fewer datapoints for the 366th day of the year. To circumvent this issue, we convert all inputs to a "noleap" calendar by dropping data for the 29th of February, except for simulations using the "360 day" calendar. In the latter case, the simulations are untouched but the reference data is converted to that calendar by dropping extra days taken at regular intervals[4].

Detrending

For each day of the year and each grid point, we first compute the averages and "anomalies" of the reference data and the simulations over the reference period, 1991-2020. Depending on the variable, anomalies are either taken additively or multiplicatively:

$$Y_r = \begin{cases} \overline{Y_r} + Y_r', & tasmax \\ \overline{Y_r} \cdot Y_r', & dtr, pr \end{cases}$$

and similarly, for X_c, $\overline{X_c}$ and X_c'.

Instead of a simple moving mean, X_s is detrended with a locally weighted regression (LOESS) (Cleveland, 1979). We chose this method for its slightly heavier weights given at the center of the moving window, reducing impacts of abrupt interannual changes on the trend and anomalies. It also has a more robust handling of the extremities of the timeseries. The LOESS window had a

[4] On a normal year, February 6th, April 20th, July 2nd, September 13th and November 25th are dropped. For a leap year, it is January 31st, April 1st, June 1st, August 1st, September 31st and December 1st.

30-year width and a tricube shape, the local regression was of degree 0 and only one iteration was performed. The detrending was applied on each day of the year but after averaging over the 31-day window, and it yielded the trend $\overline{X_s}$ and the residuals X_s'. Here again, the process can be additive or multiplicative.

Adjustment of the residuals

With $F_{Y_r'}$ and $F_{X_c'}$ the empirical cumulative distribution functions (CDF) of Y_r' and X_c' respectively, an adjustment factor function is first computed:

$$A_+(q) \coloneqq F_{Y_r'}^{-1}(q) - F_{X_c'}^{-1}(q) \qquad\qquad A_\times(q) \coloneqq \frac{F_{Y_r'}^{-1}(q)}{F_{X_c'}^{-1}(q)}$$

Where q is a quantile (in range $[0, 1]$), $A_+(q)$ is the additive function used with tasmax and $A_\times(q)$ the multiplicative one, used with pr and dtr. The CDFs are estimated for each day of the year from the 30 31-day windows. In the implementation, maps of A are saved to disk by sampling q with 50 values, going from 0.01 to 0.99 by steps of 0.02.

The adjustment is then as follows:

$$X_{ba}' = X_s' + A_+\left(F_{X_c'}(X_s')\right) \qquad\qquad X_{ba}' = X_s' \cdot A_\times\left(F_{X_c'}(X_s')\right)$$

Nearest neighbor interpolation is used to map $F_{X_c'}(X_s')$ to the 50 values of q. Constant extrapolation is used for values of X_s' outside the range of X_c'.

Adjustment of the trend

In the training step, a simple scaling or offset factor is computed from the averages:

$$C_+ = \overline{Y_r} - \overline{X_c} \qquad\qquad C_\times = \frac{\overline{Y_r}}{\overline{X_c}}$$

This factor is applied to the trend in the adjustment step:

$$\overline{X_{ba}} = \overline{X_s} + C_+ \qquad\qquad \overline{X_{ba}} = \overline{X_s} \cdot C_\times$$

Final scenario

Finally, the bias-adjusted timeseries for this day of year, grid point and variable is:

$$X_{ba} = \overline{X_{ba}} + X_{ba}' \qquad\qquad X_{ba} = \overline{X_{ba}} \cdot X_{ba}'$$

2.5.2.3 Pre-processing of precipitation

However, the multiplicative mode is prone to division by zero, especially with precipitation where values of 0 are quite common. This problem is avoided by modifying the inputs of the calibration step where the zeros of precipitation are replaced by random values between 0 (excluded) and 0.01 mm/d. The dtr timeseries are not modified since it is almost impossible to have zeros for that variable and the few that appear are dissolved by the aggregations of the calibration step.

As observed by Themeßl et al., (2012), when the model has a higher dry-day frequency than the reference, the calibration step of the quantile mapping adjustment will incorrectly map all dry days to precipitation days, resulting in a wet bias. The frequency adaptation method finds the fraction of "extra" dry days:

$$\Delta P_{dry} = \frac{F_{X_c}(D) - F_{Y_r}(D)}{F_{X_c}(D)}$$

Where D is the dry-day threshold, taken here to be 1 mm/d. This fraction of dry days is transformed into wet days by injecting random values taken in the interval $\left[D, F_{Y_r}^{-1}\left(F_{X_C}(D)\right)\right]$. Both pre-processing functions are applied only on the calibration step inputs (Y_r and X_c) before the division between average and anomalies. As such, only the adjustment factors are impacted by them and there is no explicitly injected precipitation in the final scenarios.

The algorithms described in this section are implemented in `xclim`*. The daily adjustment for this project is done in the* `bias_adjust_daily.py` *script.*

2.5.3 Climate indices

The 20 climate indices are computed over the bias-adjusted daily series of tasmax, tasmin and pr.

Climate indices computation is done in `calc_indices.py`*.*

2.5.4 Self-analog test and second bias-adjustment

As suggested by Grenier et al., (2019), data used as input to the spatial analog analysis should pass the self-analog test, i.e. the spatial analog of a given location over the reference period should be that same location. Because of random variations between the historical climate simulation and the reference data, climates indices over the reference period could be somewhat different and the self-analog test might not pass. In order to correct this, authors of the cited article recommend applying a second bias-adjustment step over the yearly indices.

Here, we checked the self-analogue test by finding the best spatial analog for each timeseries and each climate index. The distance between that “best” location and the target city is then measured. Ideally, both would be the same grid cell and distances would be 0. To decide which second bias-adjustment strategy was best, we counted the number of cities, scenarios and realisations that showed a better self-analog (closer to the target) when using doubly bias-adjusted series than when using the single bias-adjustment version. If the proportion of timeseries that are closer exceeds 50%, then we can say that this second bias-adjustment is beneficial for the self-analog test of that climate index. Remember that, in order to keep it simple, we only computed univariate spatial analogues (considering one index at a time).

While the indices used in the study by Grenier et al. (2019) were adjusted in an additive mode, this is not the case for most of the indices used here, which have a physical bound at 0 and need to be adjusted multiplicatively (see column “Adj” of **Table 2**). Moreover, for many target cities and indices have averages of 0. For example, the number of days with tasmin over 22°C in Iqaluit is 0 until the end of the century for all members of the ensemble, even with the SSP5-8.5. In general, the bias stationarity hypothesis doesn’t hold for most indices included here, even if it could be assumed when adjusting the daily variables.

We decided to generate two versions of the second bias-adjustment and check if the self-analog test results were improved or not. The first method is the same DQM method as used on the daily data, but without the day-of-year moving window. The second one is a simple scaling/offset that only adjusts the mean. Using the same notation as above, but with X representing the climates indices computed over the daily bias-adjusted timeseries:

$$X_{ba} = X_s + (\overline{Y_r} - \overline{X_c}) \qquad\qquad X_{ba} = X_s \cdot \frac{\overline{Y_r}}{\overline{X_c}}$$

The division in the second (multiplicative) equation is highly problematic for all those indices where the average over the reference period is (near-) 0. To avoid division by zero or extremely large adjustment factors, we excluded some timeseries from the adjustment process. For each target city and emission scenario pair, if any of the following conditions was true for any of the 24 realizations, the bias adjustment step was skipped for that pair:

- The mean of Y_r is smaller than 1000x its maximum
- The mean of X_c is smaller than 1000x its maximum
- The standard deviation of X_s over the 2071-2100 horizon is larger than 5 times the one over the 1991-2020 period.

The first two conditions ensure that all indices where the reference period average is near 0 are skipped, in a unit-agnostic way. The last condition is a basic test of the bias stationarity hypothesis: if the distribution completely changes shape with climate change, it is impossible for the simple bias-adjustment methods we are using to improve the results.

With this safety code in place, the second bias adjustment was skipped for a few pairs for the CDD, FAF, GDD10, R20mm and TXgt25 indices while almost all target cities and scenarios were left intact for TNgt22 and TXgt30. **Figure 2-1** Tropical nights (Tmin > 22°C) [days] for the SSP5-8.5 scenario for the city of Charlottetown (PEI). Panel A shows the index data adjusted with the scaling method while B shows the index without a second bias-adjustment.**Figure 2-1** shows an example where the scaling adjustment was applied on data that checked all above conditions (panel A). The final dataset for this case was therefore not bias-adjusted a second time. While the data is available on the server, it is not used in the dashboard, as explained in section 223.1.1.

As a first evaluation of the adequacy of the second bias-adjustment, we plotted figures with the simulated and reference timeseries for each scenario, climate index and city. While this generates a large number of images, it helped in discovering some bugs that were solved with the set of conditions explained above (see **Figure 2-1** for an example). All figures were provided to the CCCS.

Finally, when the self-analog test was computed over these three versions of the climate indices (single, dqm and scaling), no clear improvement could be seen for the doubly adjusted indices. **Figure 2-3** shows the proportion of improved cases for each index and bias-adjustment method. As it can be seen, the improvement depends on the index, but, overall, we found that DQM improved 39% of the self-analogs, while Scaling only improved 16%. In both cases, it seems that the second adjustment worsens the self-analogs more than it improves it, so we recommend sticking to the single adjustment versions. The final dashboard can be run using any of the three versions, but the default choice is to use the single-adjustment indices.

The second bias-adjustment is done with `bias_adjust_annual.py`*. The self-analog test is implemented in* `compute_selfanalogy.py`*. The analysis of this test's results was done interactively and the code couldn't be provided in the Github repository. The figures are generated by the* `fig_indices_comp.py` *script.*

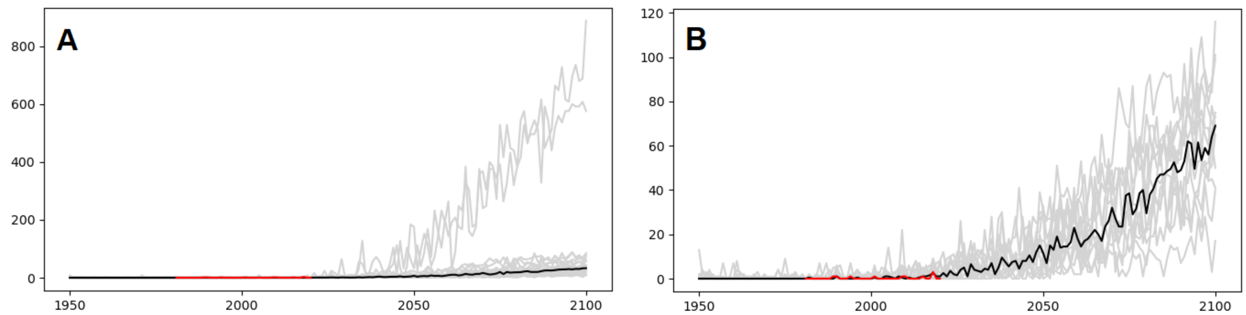

Figure 2-1 Tropical nights (Tmin > 22°C) [days] for the SSP5-8.5 scenario for the city of Charlottetown (PEI). Panel A shows the index data adjusted with the scaling method while B shows the index without a second bias-adjustment. Notice the different Y axes.

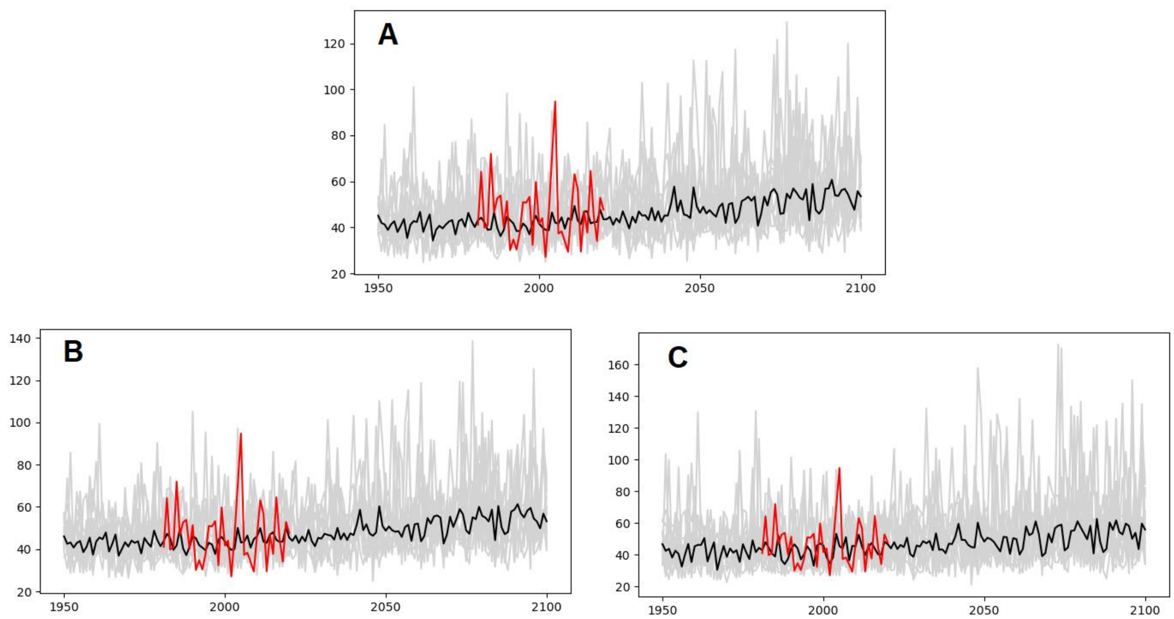

Figure 2-2 Maximum 1-day precipitation amount [mm] for the SSP5-8.5 scenario for the city of Montréal (QC). Panel A shows the index without a second bias-adjustment. Panel B and C show the data for the scaling and DQM methods, respectively. Notice the different Y axes.

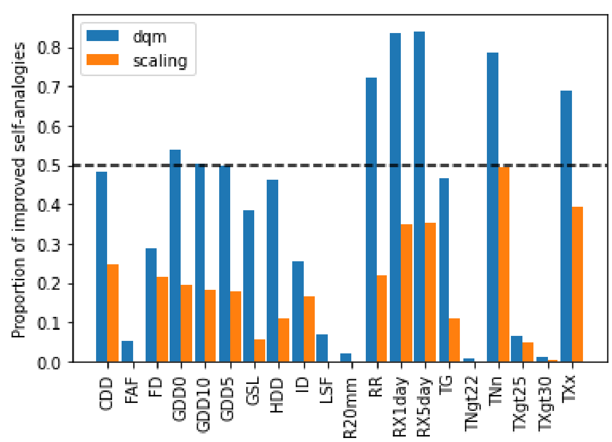

Figure 2-3 Proportion of improved self-analogs when bias-adjusting the yearly indices

3. DEVELOPMENT OF THE ANALOG-FINDER PROPOTYPE

The analog finder code was written in two versions, both including the same overall visualizations and options: a Jupyter notebook and a Panel dashboard. The notebook is meant to be used by users comfortable in the Python language, while the dashboard is a web application with a simplified user interface for controlling the spatial analog search. The following sections describe the analog finding code and the various visualizations implemented.

3.1 FINDING THE BEST SPATIAL ANALOGS

The analog finder enables users to make multiple query options. For a specific query, the user chooses:

- a target city, among the 65 selected (see **Table 1**)
- a 30-year future period of interest, with ending year from 2020 to 2100
- an emission scenario (either SSP2-4.5 or SSP5-8.5)
- up to 4 climate indices (see **Table 2**)
- a factor defining the target range for the population density of the analogs[5]. The lower bound of the range cannot be under 10 people/km².

The following options are controllable in the purely notebook version of the code as well as in the notebook code which launches the dashboard, but are available in the user-interface of the dashboard prototype:

- the climate indices dataset to use (single-adjustment, double-dqm or double-scaling), default is “single-adjustment”.
- the number of realizations N (up to 24 from **Table 3**) to include in the search, default is 12.
- the best analog selecting method (see below), default is “closest-percentile”.
- the thresholds for the analogue quality categories.
- various options for the appearance of the plots and widgets.

For each realization of the ensemble, a dissimilarity metric is used to compare the target future climate to each urban area (ERA5-Land grid tile) of the search domain (with a population density within the requested range), and to select the best analog. Hence, for a specific target city, emission scenario, 30-year period and set of indices, the analog finder identifies as many analogs as there are realizations (N) in the ensemble. The dissimilarity metric chosen for the project is the Zech-Aslan energy statistic (Aslan & Zech, 2002), which was identified by Grenier et al., (2013) as the most suitable (among 6 investigated metrics) for climate analog search.

[5] Example: for a factor of 2, the analog search includes all grid cells of North America with a population density between half and twice the one of the target city.

3.1.1 Determining usable indices

The Zech-Aslan dissimilarity metric involves the computation of the standard deviation of each climate index and will fail in the cases where all the values are the same. As shown above, this happens with some indices in some cities. In the dashboard, each time the user changes the target city, emission scenario or period, the list of available climate indices is updated to remove any cases where any of the following conditions are true:

1. the standard deviation of the index over the reference data is 0
2. the standard deviation of the index over the simulated data over the reference period is 0 for any realization
3. the standard deviation of the index over the simulated data over the target period is 0 for any realization

While the checks are not exactly the same, this is quite similar to the conditions for the second bias-adjustment. Thus, this check removes the timeseries that were not adjusted in the first place, when using the double-dqm and double-scaling data. For example, in some northern cities the TXgt30 index is always zero in the reference period (condition 1). As there is no variation (no standard deviation), we can't compute a dissimilarity score. The checks above will remove the index from the available list when that city is selected. If there were some warm days in the observational records (the reference), even if only one member of the ensemble shows a null standard deviation over the two periods, the index is removed, instead of having an incomplete simulation ensemble (condition 2 and 3).

3.1.2 Interpreting the dissimilarity score

The Zech-Aslan metric doesn't follow a specific distribution. Rather, the distribution of the results depends on that of the inputs, making it impossible to compare scores obtained from different sets of indices. Moreover, the score is a single number, theoretically ranging from $-\infty$ to ∞ (but rarely under 0 in practice); interpreting these raw numbers is not easy. This is in opposition to the metric used by Fitzpatrick & Dunn, (2019) where the score distribution was known, allowing a mapping from scores to percentiles.

In order to compensate for this difficulty, we translate each score into a "quality flag", informing the user about the quality of the analogue. As mentioned by the authors of the metric themselves: *"Rather than [theoretically calculating] these parameters from the moments of the specific φ distributions, we propose to generate the distribution of the test statistic and the quantiles by a Monte Carlo simulation."* (Aslan & Zech, 2002, sec. 2.3). Assuming the distribution of the statistic is specific only to the set of climate indices, we approximated distributions by computing the statistic over random pairs of cells of the reference data, chosen within all North America but only cells with a population density above 10 hab./km². The statistic was computed over 200 000 random pairs, for each of the 6120 possible indices combinations. The percentiles of the approximated distributions were saved to disk.

With this data, we can now translate any given dissimilarity score into an approximated rank of the estimated distribution. For example, if a score maps to the 50th percentile of the distribution, we can interpret this by saying: there is a 50/50 chance of having a better analogue by choosing another random point within North America. This type of comparison seemed good enough for the needs of this project. Only 101 percentiles (0, 1, … 100) of each distribution were saved in the Monte-Carlo process; the rank of any given score is obtained by linear interpolation between those.

The analog finder further simplifies the output by mapping the ranks to quality flags, as detailed in **Table 4**. The bounds of each range can be modified in the notebook, but are fixed for the dashboard user-interface.

Table 4 Analogue quality categories

Color	Description	Percentile range
	Excellent	[0, 1]
	Good	] 1, 5]
	Average	] 5, 10]
	Poor	] 10, 100]

The approximate distributions are computed with the `analog_quantiles.py` *script and stored in a netCDF file available on the THREDDS server of PAVICS.*

3.1.3 Finding the best analog for each simulation

The first version of the prototype selected the "best" analog simply by finding the cell with the lowest dissimilarity score. The fact that the data doesn't pass the self-analog test (see above) was extremely obvious with this method. Moreover, we realized that the dissimilarity fields computed were quite noisy: the "lowest" dissimilarity could be very similar to the target's. The problem was exacerbated by the artificial selection of the population density which creates maps that are sparse in nature. While, we can't fix the self-analogue problem for our choice of indices (see above), we implemented a selection method that allowed smoothing of the results and yielded a spatial analog easier to interpret. In addition to the basic "min" method, the dashboard can be launched with either "closestN" or "closestPer".

- closestN: The N points with the lowest dissimilarity are found and the one closest to the target (in physical distance) is returned as the "best analog". The default value for N is 10.
- closestPer: The rank of the lowest dissimilarity is computed as R_{min}, all points with ranks in the range $[R_{min}, R_{min} + \Delta R]$ are found and the closest (physical distance) one is returned as the "best analog". The default value of ΔR is 1.

The "closestPer" method was chosen as the default. Compared to "closestN", it ensures that the analogues of all points considered for the "best analog" are of similar quality.

3.1.4 Ranking the best analogs

Once the analog finder code is run, we get one best analog location for each realization of the ensemble (12 by default). In the dashboard, the user can switch between the results of each realization through a row of buttons at the top of the page (**Figure 3-1**). Two methods for ranking the realizations were implemented: by score or by representativeness of the ensemble.

The first way simply ranks the realization by the dissimilarity score of the best analog, in ascending order. For the second method, each realization is compared to the ensemble mean through a multivariate Z-score. Considering a space in d dimensions (one for each climate index in the set), $\mathbf{X}$ is the matrix of the d-dimensional points, one for each N realizations and 30 years of the target period. $\mathbf{X}_i$ is the 30-year vector of points of a single realization. With $\langle\ \rangle$ the ensemble average,

$^{-}$ the temporal average, $\| \ \|$ the Euclidean norm and $\sigma(\ )$ the standard deviation across time, our representativeness score S_i is:

$$S_i = \left\| \frac{\overline{\mathbf{X}_i} - \overline{\langle \mathbf{X} \rangle}}{\sigma(\langle \mathbf{X} \rangle)} \right\|$$

This score represents the standardized distance of each realization to the ensemble average. In the final dashboard, the realizations are ranked according to their score S_i.in ascending order. Button color indicates the quality flag category from section 3.1.2, as detailed in **Table 4**.

> *All processes described here are implemented directly in the* `Dashboard.ipynb` *and* `Step_by_step.ipynb` *notebooks. While there shouldn't be any difference between these two and between the Github and PAVICS versions, one can consider the* `Dashboard.ipynb` *notebook provided through* `pbourg`*'s public folder on PAVICS' Jupyterhub as the master version.*

3.2 VISUALIZATION PROTOTYPE

Aside from the technical choices for identifying the best analogs for a specific user query, the issue of results visualization is very important, as this will guide its interpretation. Several key aspects were identified at the beginning of the project. The control panel of the web interface should include sub-panels for: 1) the selected options; 2) a map on which the target city as well as all analog locations are marked, and where the user can highlight one specific analog by cursor movement); 3) climate change uncertainty diagrams that highlight the analog-related underlying climate scenario among the ensemble, for each selected index; and 4) similarity diagrams that show how good the analog is, for each selected index. The last two sub-panels are necessary to guide the user in her/his interpretation of the analogs.

In the final dashboard, all these elements were included. **Figure 3-1** shows the dashboard once a query has been analyzed and spatial analogs were found, and **Figure 3-2** shows one univariate analogy panel expanded. The next sections describe each panel.

The goal of this project was to build a prototype of a web application that could be embedded in the ClimateData.ca portal. As such, several elements may not be in a deployment-ready state and the user interface could benefit from many changes, small or not. In the following, we highlight some elements that we feel are missing. Some we tried to implement, but failed to do so in the time available, others we simply think would be feasible and useful.

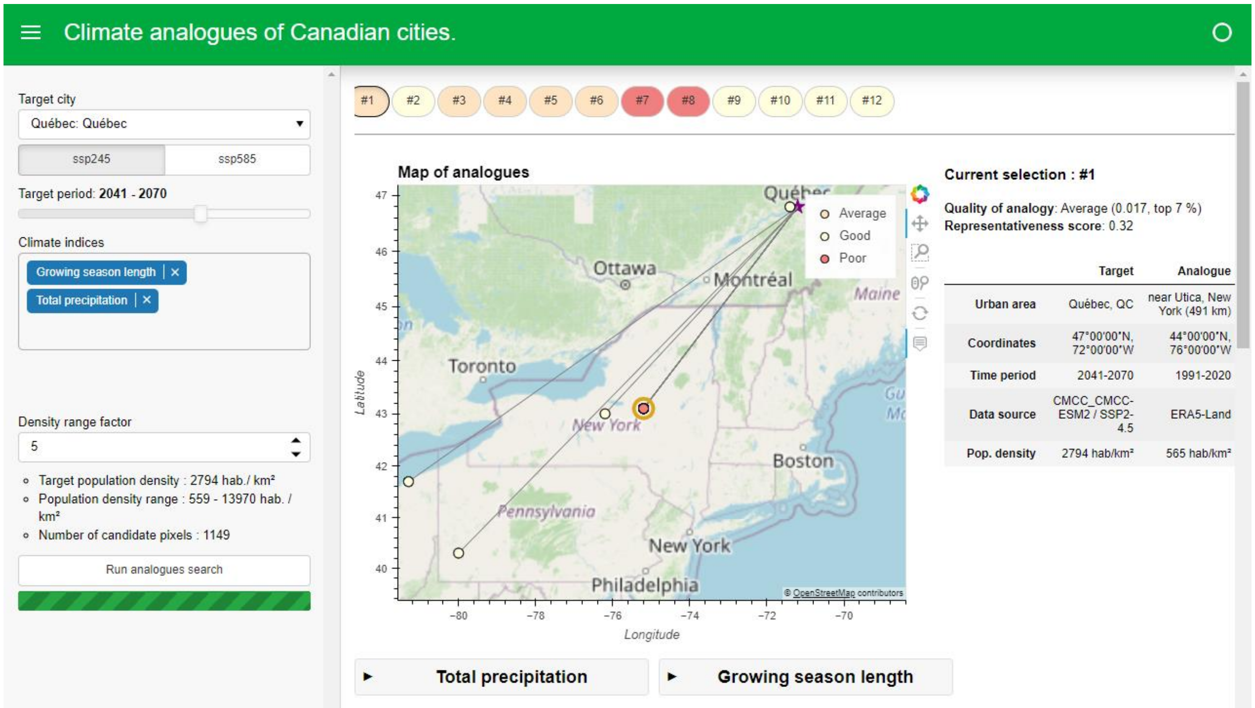

Figure 3-1 Dashboard after an analogue search

3.2.1 Sidebar: Search query builder

All search parameters are available on a sidebar of the dashboard (which can be collapsed). The description of each element was given above. Each time the user changes their selection of target city, scenario or period, the list of available climate indices is updated, removing those considered unusable. A small summary paragraph lists the population density of the currently selected city, as well as the bounds of the range depending on the selected factor. The number of candidate pixels (cells with a density within the range) is given so that the user can estimate the time necessary for the analog search. Once the "run analogues search" button is pressed, the progress bar starts moving.

Missing elements

Like in most of the dashboard, some help text is missing to inform the user about the different parameters and climate indices. A good way to add this would be an info bubble that pops up when the cursor (or finger) hovers over the title of the control widget. A complete description of the app, as an external web page, could also help.

There seems to be a difficulty in linking the computing library (dask) with the widget one (panel) in order to have a progress bar that shows a real percentage of the work done and updates in real time. For now, it only shows that the search is active. Moreover, once the search is done, there is a delay where the progress bar stops moving but the visualisation widgets are being generated.

3.2.2 Realization selector

The top of the main section consists of a row of rounded buttons that allows switching between the different realizations and their associated best analog. As said above, the realizations are ranked (from left to right) according to their representativeness score S_i and the button is coloured according to the quality of the analogue. Each of the 4 categories is mapped to a single colour, the mapping is the same as on the legend of the map below.

Missing elements

This small section could also benefit from some explanations, either directly in the app or as a hover-over text bubble.

3.2.3 Map of analogues and summary table

The map shows all best analogues as a coloured dot linked to the target city, a purple star. The currently selected analogue's dot is circled with a gold ring. The colour of each dot is related to the quality of the analogue and is mapped to one of the 4 categories, as shown in the legend of the map. In many cases, several realizations will find the same best analogs, resulting in many dots superimposed on the map. A hover text bubble was added to the lines and dots so that the user has some way to know when this happens.

The summary table adjacent to the map provides statistics about the currently selected analogue. The quality of the analogue is given through the name of the category, the raw dissimilarity score and the rounded percentile, expressed as the percentage of random pairs with a better (or similar) score (see section 223.1.2). The representativeness score is also given.

The table lists different metadata elements, comparing the selected analogue and the target. The "Urban Area" given for the analogue is taken from a list of populated places derived from a dataset of Natural Earth. The table lists the closest populated place, which may or may not be the main official administrative unit of the grid cell. The "near" keyword was added because of this uncertainty. The full name of the model run that returns this analogue is listed in the target's data source.

Missing elements

Again, information about the different statistics and field listed should be added. As noted in the contract, it would nice to be able to switch realization by clicking the dots in the map, in addition to the button row.

The list of named places used to associate an analogue location with a city name is stored on PAVICS' geoserver (`public:ne_10m_populated_places`*).*

3.2.4 Univariate analogues

Below the map and table, each climate index in the selected set is further analyzed in its own collapsible "card".

The "univariate analogue" refers to the spatial analog analysis re-run on the same data as the best analog, but with a single climate index. It can offer a way to understand which indices of the selected set contribute negatively or positively to the overall dissimilarity score. In addition to this univariate dissimilarity score, the panel gives a description of the index.

The first plot approximates the distribution of the index over the 30 years of each period. It shows the kernel density estimates of the index over the reference period (1991-2020, white) and target period (purple) for the target city (data from the selected realization) and over the reference period for the best analog (data from ERA5-Land, gold). The second plot is unidimensional and shows the averages of the same three distributions as on the first plot. The kernel density estimate plot was not mentioned in the contract, but it seemed to convey more useful information than the comparison of 30-year average values, which was proposed. However, estimating a distribution with only 30 values could potentially result in problems or unusual distribution shapes.

Finally, the third plot shows the full timeseries of all realizations over the target city and the timeseries of the best analog over the reference period. The currently selected realization is highlighted in purple. The target period is delimited with two vertical blue lines and the data from the analog is re-plotted over the reference period so one can visually inspect the difference between this and the target's data.

Figure 3-2 Univariate analogue panel

Missing elements

In addition to improving documentation, some work on the layout of these “cards” could be beneficial.

Currently the distribution plots are created via kernel-density estimation (KDE) whose smoothing algorithms can result in densities for negative climate index values even for those indices where this is not physically possible. Despite this limitation it was felt that the ease of comparison of distribution overlap provided by the KDE plots was more beneficial than not. Future improvements could look into mechanisms for ‘clipping’ KDE density values for climate indices which cannot have values below zero.

> *This section describes the visualizations as implemented in* `Dashboard.ipynb`. *The usage of* `matplotlib` *instead of* `bokeh` *as a plotting backend in* `Step_by_step.ipynb` *yields different results in that notebook. Again, the dashboard provided through PAVICS should be considered the master version is any differences are found with the other versions.*

4. DELIVERY OF THE DATASETS AND CODE

The final products of this project have been delivered: datasets, python code and Jupyter notebooks. Datasets are available on PAVICS' thredds server. The code and notebooks are shared to the CCCS through a private Github repository. The README file of this repository indicates the purpose of each file and their usability state. Details are presented in this report. Finally, the dashboard and the notebook were shared through PAVICS's JuypterHub platform.

Disclaimer

> "*The python code, notebooks and data files associated with this contract are provided AS IS and their usage is at the own risk of its user. Ouranos makes no warranties, expressed or implied, of merchantability and fitness for a particular purpose of the data or code, as well as of non-infringement of rights of others. Ouranos shall not be liable for damages or losses resulting from the use, application, or interpretation of the provided material*".

Datasets (public):
https://pavics.ouranos.ca/twitcher/ows/proxy/thredds/catalog/birdhouse/ouranos/spatial-analogs/catalog.html
This thredds catalog provides:

- `masks.nc` : A collection of time-invariant masks on the ERA5-Land grid.
- `era5-land.ncml` : The climate indices over the reference data and period. Also provides the density mask.
- `cmip6_single.ncml`, `cmip6_scaling.ncml` and `cmip6_dqm.ncml` : The three different flavours of the simulated climate indices.
- `cities.nc` : A netCDF version of the Canadian cities metadata.
- `benchmarks.nc` : The approximate distributions of dissimilarity scores for each possible climate index combination.

Geographical data:
https://pavics.ouranos.ca/geoserver/web/wicket/bookmarkable/org.geoserver.web.demo.MapPreviewPage?3

- `public:ne_10m_populated_places`: For use by the dashboard, the list of named places used by the dashboard to associate each found analogue with a city is provided by a Geoserver instance within the PAVICS ecosystem.
- `analogs:cities`: The Canadian cities data (almost the same as `cities.nc`), but an encoding bug forbids the dashboard from using that version for now. The data is provided in a temporary file beside the dashboard's notebook.

Code (private repository, requires a GitHub account and an invitation):
https://github.com/Ouranosinc/analogues_spatiaux

Dynamic notebooks (read-only, requires a PAVICS account):
https://pavics.ouranos.ca/jupyter/user/pbourg/lab/tree/public/pbourg-public/Analogues%20spatiaux/Dashboard.ipynb

CONCLUSION

The current report, associated code and datasets provide a functional prototype for finding contemporary spatial climate analogues for Canadian cities, identifying locations which currently have similar climatic conditions to those projected in the future for the cities of interest. The prototype enables customized calculations of climate analogues offering a flexible yet straight forward toolset for users to explore potential futures changes in climate.

REFERENCES

Agbazo, M. N., & Grenier, P. (2020). Characterizing and avoiding physical inconsistency generated by the application of univariate quantile mapping on daily minimum and maximum temperatures over Hudson Bay. *International Journal of Climatology*, *40*(8), 3868–3884. https://doi.org/10.1002/joc.6432

Aslan, B., & Zech, G. (2002). *A new class of binning free, multivariate goodness-of-fit tests: the energy tests*.

Cannon, A. J. (2015). Selecting GCM Scenarios that Span the Range of Changes in a Multimodel Ensemble: Application to CMIP5 Climate Extremes Indices. *Journal of Climate*, *28*(3), 1260–1267. https://doi.org/10.1175/JCLI-D-14-00636.1

Cannon, A. J., Sobie, S. R., & Murdock, T. Q. (2015). Bias Correction of GCM Precipitation by Quantile Mapping: How Well Do Methods Preserve Changes in Quantiles and Extremes? *Journal of Climate*, *28*(17), 6938–6959. https://doi.org/10.1175/JCLI-D-14-00754.1

CIESIN, C. for I. E. S. I. N.-C.-C. U. (2018a). *Gridded Population of the World, Version 4 (GPWv4): Land and Water Area, Revision 11*. NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/H4Z60M4Z

CIESIN, C. for I. E. S. I. N.-C.-C. U. (2018b). *Gridded Population of the World, Version 4 (GPWv4): Population Count Adjusted to Match 2015 Revision of UN WPP Country Totals, Revision 11*. NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/H4PN93PB

Cleveland, W. S. (1979). Robust Locally Weighted Regression and Smoothing Scatterplots. *Journal of the American Statistical Association*, *74*(368), 829–836. https://doi.org/10.1080/01621459.1979.10481038

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, *9*(5), 1937–1958. https://doi.org/10.5194/gmd-9-1937-2016

Fitzpatrick, M. C., & Dunn, R. R. (2019). Contemporary climatic analogs for 540 North American urban areas in the late 21st century. *Nature Communications*, *10*(1), 614. https://doi.org/10.1038/s41467-019-08540-3

Grenier, P., Firlej, A., Blondlot, A., Logan, T., & Ricard, M.-P. (2019). The issue of properly ordering climate indices calculation and bias correction before identifying spatial analogs for agricultural applications. *Climate Services*, *16*, 100122. https://doi.org/10.1016/j.cliser.2019.100122

Grenier, P., Parent, A.-C., Huard, D., Anctil, F., & Chaumont, D. (2013). An Assessment of Six Dissimilarity Metrics for Climate Analogs. *Journal of Applied Meteorology and Climatology*, *52*(4), 733–752. https://doi.org/10.1175/JAMC-D-12-0170.1

Logan, T., Bourgault, P., Smith, T. J., Huard, D., Biner, S., Labonté, M.-P., Rondeau-Genesse, G., Fyke, J., Aoun, A., Roy, P., Ehbrecht, C., Caron, D., Stephens, A., Whelan, C., Low, J.-F., Keel, T., Lavoie, J., Tanguy, M., Barnes, C., … Quinn, J. (2022). *Ouranosinc/xclim* (0.35.0). Zenodo. https://zenodo.org/record/6407112

Muñoz Sabater, J. (2019). *ERA5-Land hourly data from 1981 to present.* Copernicus Climate Change Services (C3S) Climate Data Store (CDS). 10.24381/cds.e2161bac

Themeßl, M. J., Gobiet, A., & Heinrich, G. (2012). Empirical-statistical downscaling and error correction of regional climate models and its impact on the climate change signal. *Climatic Change*, *112*(2), 449–468. https://doi.org/10.1007/s10584-011-0224-4

Thrasher, B., Maurer, E. P., McKellar, C., & Duffy, P. B. (2012). Technical Note: Bias correcting climate model simulated daily temperature extremes with quantile mapping. *Hydrology and Earth System Sciences*, *16*(9), 3309–3314. https://doi.org/10.5194/hess-16-3309-2012

APPENDICES

APPENDIX A – REGRIDDING OF THE HIGH-RESOLUTION POPULATION COUNT TOWARDS THE REFERENCE GRID

As the GPWv4 UN WPP-adjusted population density is already calibrated as so to only include population on the land-covered fraction of the grid point, it cannot be easily regridded to the reference grid. We decided to instead regrid the population count and divide it by the land area of the target grid cell in order to get our population density map. But, instead of using the coarse land fraction mask provided with ERA5-land, we used the land area also provided in GPWv4 and aggregated it as well to the reference grid. Figure A1 summarizes this process.

The final result (D_{era} in the figure), represent the population density on the land portion of the ERA5-land grid cell.

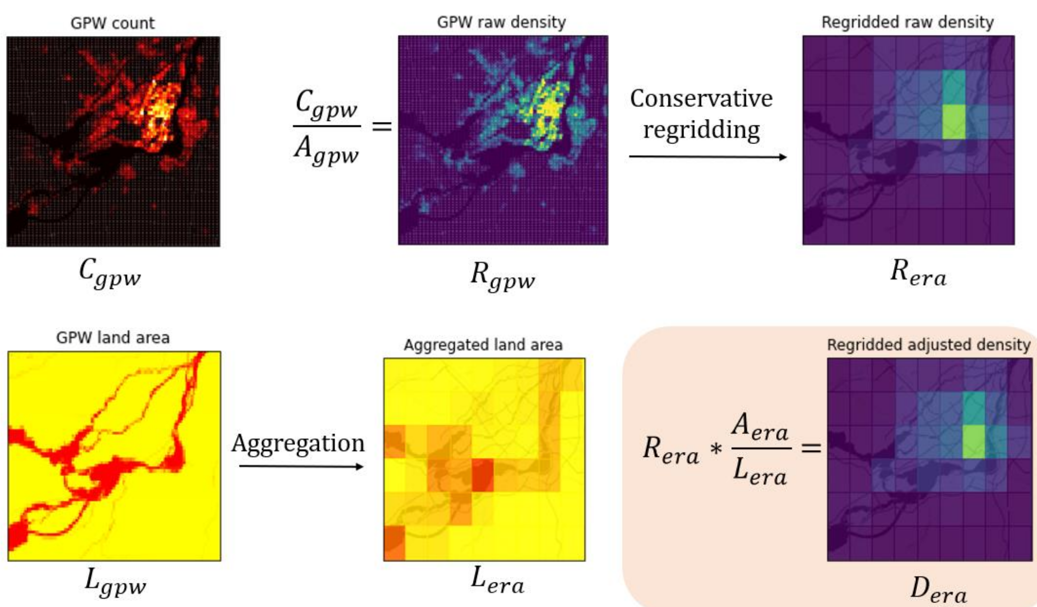

Figure 0-1 Procedure for the regridding of the adjusted population density

Variables of figure 8-1:

- $A_{gpw}, A_{era} =$ Cell area [km²]
- $C_{gpw} =$ Population count [-]
- $R_{gpw}, R_{era} =$ Raw population density [hab / km²]
- $L_{gpw}, L_{era} =$ Land area [km²]
- $D_{era} =$ Land-weighted population density [had / km²]

In the wake of a series of extreme-weather events that highlighted the vulnerability of Quebec communities, a group of progressive-minded scientists and policymakers decided to collaborate on potential solutions; Ouranos is the result.
In 2001, Ouranos was created as a joint initiative of the Québec government, Hydro-Québec and Environment Canada, with the financial support of Valorisation-Recherche-Québec.
Ouranos aims to provide Québec and the rest of Canada with expertise in both climate science and adaptation strategies.
Ouranos engages in focused, practical science in support of sound decisions related to climate change and its impacts. Our approach actively involves an expanded network of researchers, experts, practitioners and policy-makers.

OURANOS

550 SHERBROOKE OUEST, TOUR OUEST, 19e ÉTAGE, MONTRÉAL, QC, CANADA, H3A 1B9 | TÉLÉPHONE 514 282.6464 | TÉLÉCOPIEUR 514 282.7131 WWW.OURANOS.CA