

Weather Research Project in Quebec

Using machine learning and social narratives from the past and present to explore changes in vulnerability and resilience over time for weather events

Funded by FRQ Audace grant

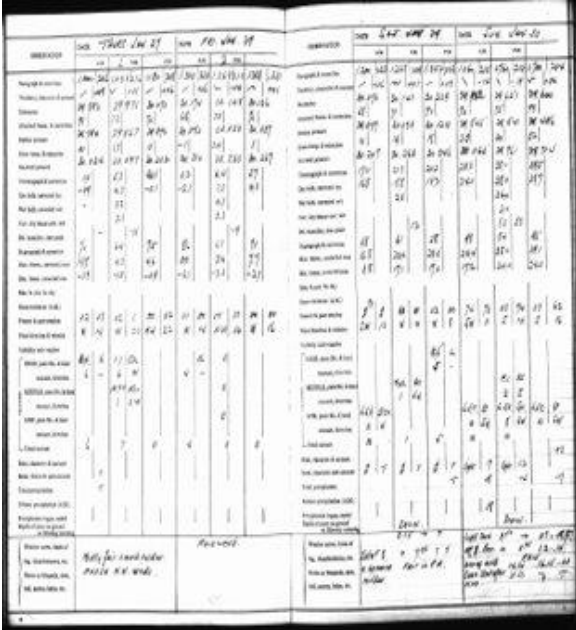
Victoria Slonosky, Renée Sieber, Frédéric Fabry, Gordon Burr, Alyssa Conlon, Yumeng Zhang, Munchen Wang, Antoine Rehberg

SAA Research Forum July 19, 2023



McGill

DRAW (Data Rescue: Archives and Weather)



The image shows an open historical weather logbook with two pages of handwritten data. The pages are filled with columns of numbers and text, representing weather observations over time. The handwriting is in cursive and the paper appears aged. The columns are organized into sections, likely representing different weather parameters or time intervals. The data is dense and covers multiple days.

- Goal:
- To capture historical weather data lost to scientists in analogue archival records through a Citizen Science (crowdsourcing) web application and as a participatory archives project to address the need to rescue data not currently included in contemporary data models to build a more complete view of climate change

Vulnerability and resilience through time

L'ETOILE DU NORD, 2 MAI 1885.

Inondation à Joliette

PANIQUE DANS LA VILLE

LES PONTS EMPORTÉS

LES MOULINS DÉMOLIS

Pertes considérables

DÉTAILS COMPLETS.

Samedi dernier eut lieu la débacle de la rivière L'Assomption. La hauteur de l'eau inspirait des craintes aux habitants de Joliette. l'événement a prouvé que ce n'était pas à tort. Toute la journée de samedi, les nouvelles les plus alarmantes nous arrivaient des localités situées plus haut que la ville sur les rives de L'Assomption. Un moulin situé à deux milles au-dessus de Joliette et appartenant à Mess. Bordeleau a été emporté par les glaces, ainsi que les dépendances contenant tous les animaux du propriétaire. La rivière n'a cessé durant toute la journée, de charroyer des meubles et des débris de charpente. C'était loin d'être rassurant pour la ville.

Les pertes de M. Bordeleau sont évaluées à environ \$7,000.00.

Toute la population de Joliette affoibée de terreur était sur pied; une foule de femmes et d'enfants couraient dans les rues et répandaient partout les sinistres détails du désastre. Jamais, de mémoire d'homme, notre ville n'a été la victime d'un fléau de ce genre; aussi serait-il impossible de peindre l'effroi des habitants et la panique qui s'en suivit jusqu'à une heure avancée de la nuit.

Durant toute la journée de dimanche, la foule des curieux et des intéressés ne cessa de visiter la scène de désolation.

La ville n'ayant plus en communication avec les campagnes dont la rivière nous sépare, la corporation a engagé des bateliers qui feront le service de transport jour et nuit, entre les deux rives, jusqu'à ce qu'un nouveau pont soit construit.

Nous avons pu recueillir de la bouche même des intéressés le montant des pertes occasionnées par l'inondation: M. Geo. Gilmour, \$4,000; M.M. Bordeleau \$7,000; John Grilly propriétaire de la manufacture de papier, 10,000 à \$12,000, la Corporation de Joliette, 8,000 à \$10,000; Kelly et Copping \$2,500; la fonderie de Joliette \$500; la Cie à bois de Joliette \$4,000.

Outre les dégâts, il faut de plus faire entrer dans les pertes causées par la débacle les dom-



JOHN MAHONEY

THE WATERS RISE

Audace FRQ Research Project: On How Vulnerabilities and Resiliencies *change* through time

- Changing societies, technology and infrastructure lead to *different* vulnerabilities and resiliencies
- Communities respond to these challenges at various levels of organization (geographical, governmental, philosophical, expectational)
- Studying the link between disruptive events and the underlying meteorological roots using social narratives and weather data will help us move away from statistical thinking about extreme events towards events that have social impacts

Newspaper articles as a lens for social disruption

Two time periods: 1880-1900, 2000-2020, English and French language newspapers

Study area: Southern Quebec

Focus: disruptive weather events

Challenge: how to define disruptive?

Our workflow

1. Assemble Corpus (Historical/Modern)

1. Identify sources and terms and conditions of data use
2. Determine list of filter terms
3. Extract articles in batches: xmls, txts, tifs and pdfs
4. Curate and store the raw content

2. Pre-process Corpus

1. OCR quality test (could be manual)
2. Automatically clean the articles
3. Document the cleaning processes, which could differ by platform

3. Conduct machine learning classifications

1. Pre-process text
2. Conduct unsupervised classifications
3. Compare results by language, time periods
4. Conduct semi-supervised classification, with boosting of vulnerability, resilience terms
5. Compare results by language, time periods

4. Compare topics to timelines

Topics from #3 are time-agnostic, ALTHOUGH we could look for word clusters in newspaper articles.

We want to get to Steps 3 & 4 but building a corpus for disruptive weather events has been challenging!

1) Choosing Keywords

- a) Selecting applicable terms and phrases for both French and English, historical and modern →
- b) Issues with term retrievability with OCR

2) Obtaining the Articles

- a) Licensing and copyright
- b) Exporting limits and technical limitations (e.g., Internet, storage)
- c) Every platform is different!

3) Turning articles into machine readable text

- a) Can include secondary optical character recognition (OCR) especially for texts, PDFs, XMLs from historical articles (which are themselves .TIFF)
- b) Error analysis of OCR

4) Extracting articles of interest

Keywords

- Weather-related keywords are often used metaphorically
- Goal is to find appropriate terms
 - E.g., Using drought and storm as search terms produce unhelpful results.
- Filter out non-weather articles using combination of words
 - E.g., flood AND swept away

Who's to Blame for Canada's Continued Stanley Cup Drought?

STEVE MACFARLANE 

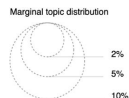
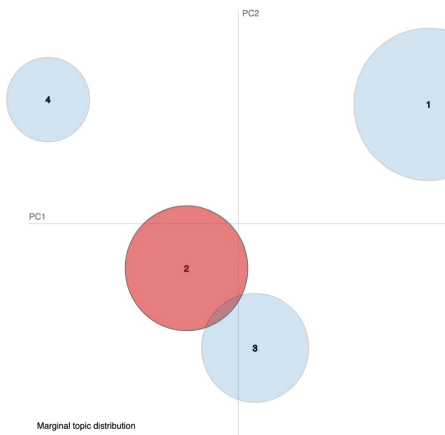
MAY 13, 2015



Using AI/ML to gain insights

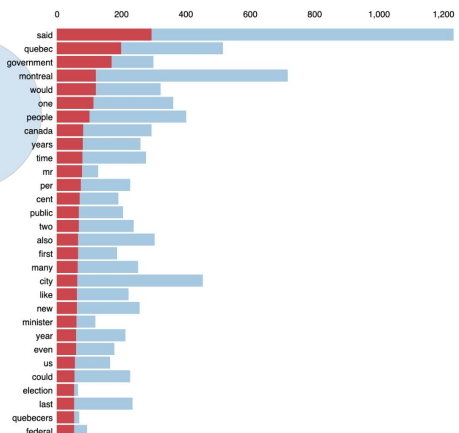
Selected Topic: 0 Previous Topic Next Topic Clear Topic

Intertopic Distance Map (via multidimensional scaling)



Slide to adjust relevance metric:⁽²⁾ $\lambda = 1$

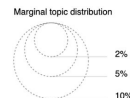
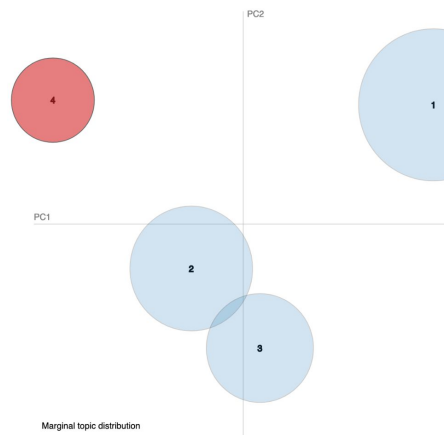
Top-30 Most Relevant Terms for Topic 2 (27% of tokens)



1. $saliency(\text{term } w) = \text{frequency}(w) * [\sum_{t=1}^T p(t|w) * \log(p(t|w)/p(t))]$ for topics t ; see Chuang et. al (2012)
 2. $relevance(\text{term } w | \text{topic } t) = \lambda * p(w|t) + (1 - \lambda) * p(w|t)/p(w)$; see Sievert & Shirley (2014)

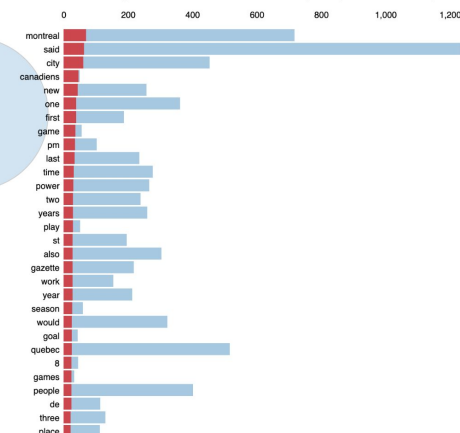
Selected Topic: 0 Previous Topic Next Topic Clear Topic

Intertopic Distance Map (via multidimensional scaling)



Slide to adjust relevance metric:⁽²⁾ $\lambda = 1$

Top-30 Most Relevant Terms for Topic 4 (12.4% of tokens)





1. $saliency(\text{term } w) = \text{frequency}(w) * [\sum_{t=1}^T p(t|w) * \log(p(t|w)/p(t))]$ for topics t ; see Chuang et. al (2012)
 2. $relevance(\text{term } w | \text{topic } t) = \lambda * p(w|t) + (1 - \lambda) * p(w|t)/p(w)$; see Sievert & Shirley (2014)

Unsupervised classification (above Latent Dirichlet Allocation)

Next step: generative AI

Another “mapping” of terms and their clustering

1/208  

id >




434

text >

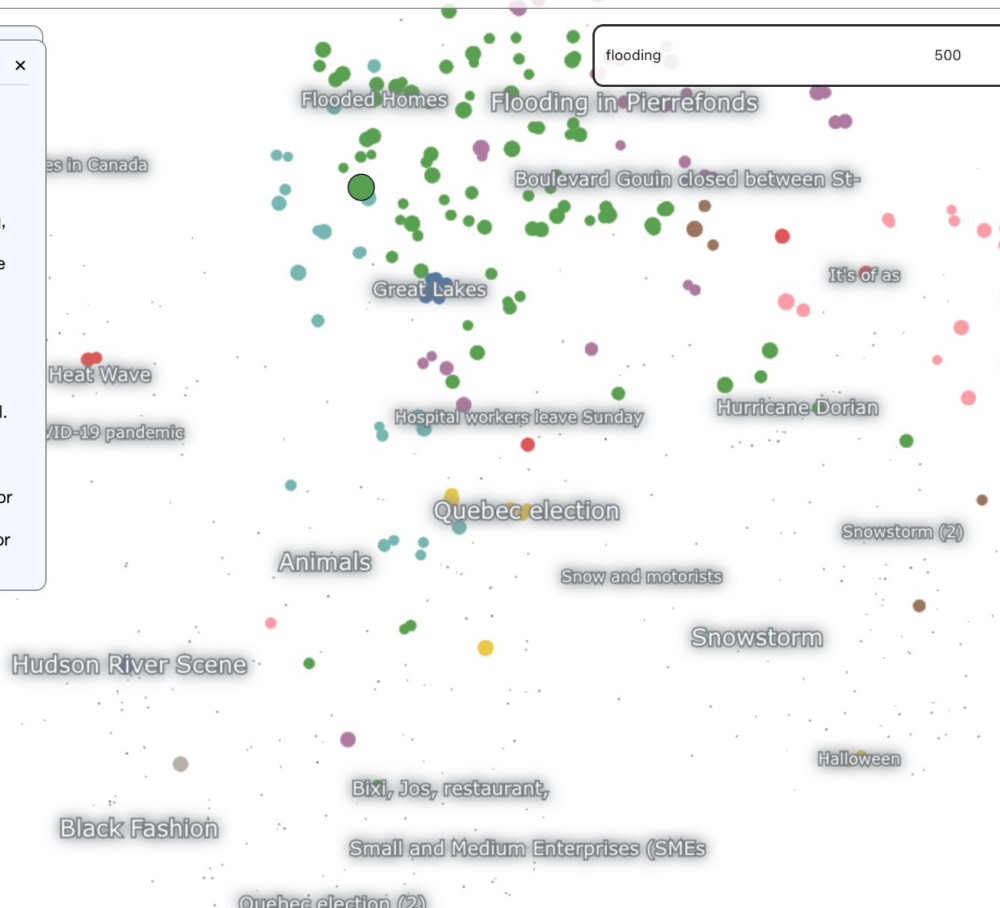
A noted hydrologist is slamming the province's handling of recurrent flooding, saying it has the duty to better manage its water flow, rather than abandon those who have been flooded.

"People who live in flood zones are not culprits, and if they are flooded, it's not just their responsibility, but the responsibility of everyone," said Claude Marche, a retired engineer and hydrologist at the Université de Montréal.

The municipal affairs ministry recently published a map delineating a special intervention zone (SIZ) whereby those who live in that zone won't be covered for extreme damage to their home because of flooding, and instead will be eligible for

English_Modern_topic_model
Ensemble, technologiques,



Building timelines of events from newspapers

- Find days when weather is in the news;
- Select the top N days; investigate the cause(s)
- Goal: Find whether there are surprises, e.g., unexpected events we talk about, or expected events we ignore.



Flooding in Waterton Lakes National Park, June 14, 2022. (Photo: Parks Canada)

By Mario Cabradilla

Alberta releases flood update

Jun 15, 2022 | 6:28 PM

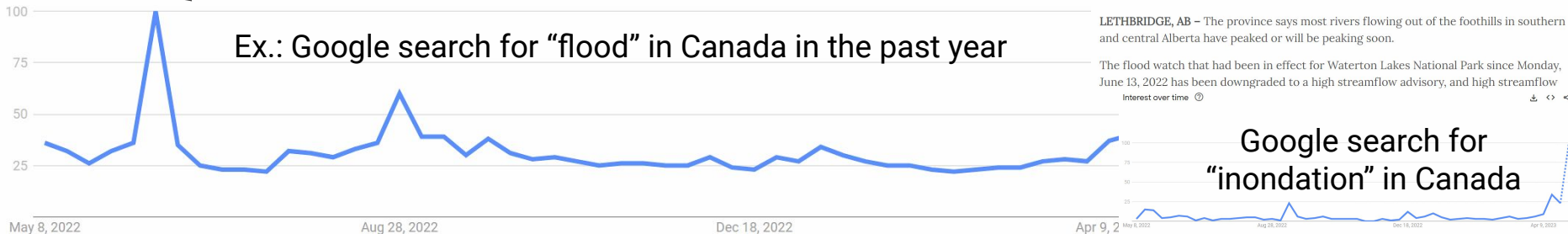
LETHBRIDGE, AB – The province says most rivers flowing out of the foothills in southern and central Alberta have peaked or will be peaking soon.

The flood watch that had been in effect for Waterton Lakes National Park since Monday, June 13, 2022 has been downgraded to a high streamflow advisory, and high streamflow

Interest over time

↕ ↔

Google search for “inondation” in Canada



Preliminary thoughts

- Team of 4 researchers, 5 students, working on different aspects, from keyword searches, cleaning the text to performing AI-based searches. Stay tuned for results next year!
- Surprising/interesting findings until now:
 - Late 19th century Québec very well adapted to snowstorms
 - Weather and transportation always a concern, but it changes form: Mud was bad news in late 19th c., vs snow/ice today
 - Modern Québec still quite well adapted, but more vulnerable due to collective electrification leading to dependency for heat, cooking, transport.
 - Present-day citizens much more vulnerable to ice storms than past, for similar reasons due to collective dependence on electricity
 - Floods are high impact and highly disruptive in both past and present

Acknowledgments

FRQ AUDACE program grant: For additional information on this project: Please contact Victoria Slonosky at victoria.slonosky@mail.mcgill.ca

For information about DRAW contact Gordon Burr at gordon.burr@mcgill.ca

Please go to our website [DRAW](#) and join up. Then click on **Transcribe** and **All pages** to help capture all the remaining Observatory records from 1880-1900 to form the meteorological records dataset to link to the AI results for the Audace Project. Thank you!