GEM Weather Model Access Portal

Enabling weather model data access with open source python tools.

Joseph Young

40138967

CIVI 7901

Advisor: Dr. Nazemi

1. Abstract

Modern numerical weather predictions can produce sufficiently accurate weather forecasts for virtually all regions of the globe. As climate change drives a higher frequency of extreme weather events, access to weather forecasts is becoming an increasingly crucial tool for individuals and communities to prepare for incoming weather. Environment Canada has developed the Global Environmental Multiscale (GEM) model to forecast weather data in all regions of the globe with a comparatively high level of accuracy (Environment Canada, 2022). However, there are often barriers to access numerical weather prediction data, particularly by individuals in remote and underprivileged regions of the globe with less technological connectivity. Environment Canada's Global Deterministic Prediction System is one of those models that carries barriers to access. This project presents a software product that processes, visualizes and enables greater access to numerical weather predictions from Environment Canada. A usable proof of concept is developed by using open source products and Python code and deployed online. The result is a product that demonstrates a method for the democratization of data access between privileged and under-privileged countries.

2. Introduction

The development of today's numerical weather predictions demonstrates the advancement in scientific knowledge of atmospheric conditions and the processing capabilities of modern computers. Numerical weather predictions are crucial to preparing for weather events, and can save lives by enabling proper preparation in the face of natural disasters. Many governmental organizations develop weather forecast models for global weather data. These weather models provide short term predictions around the entire globe and have increased in accuracy over the past decades. These predictions contain numerical values for common weather variables relevant to human comfort and safety like temperature, humidity, precipitation rate and wind speed.

However, despite the advancements in resolution and accuracy, numerical weather predictions are useless to individuals and groups without technological access to forecast results.

Data provided by these weather models are publicly available, but can often be difficult to access without the necessary knowledge of file transfer protocols, access to technology, and a reliable internet connection. Therefore, there is a need for software tools that ensure democratized access to weather forecasts in developing nations, while also serving researchers as a repository of weather forecast data to be compared with observed data.

This project focuses on providing data access to over the regions of both Canada and Pakistan. These two countries represent a large divide in the level of privilege that citizens experience. Pakistan is currently in the late expanding stage of the demographic transition model, as birth rates slowly decrease (Goujon et al., 2020). However the population of Pakistan continues to grow at a rate twice as fast as other South Asian Countries (Goujon et al., 2020), indicating that the nation is still in the middle of a demographic transition. Nations further along in the demographic transition model typically support higher standards of living and economic opportunities (Reher 2011). By focusing on nations with different levels of privilege, this project seeks to address the imbalance in data accessibility between privileged and under-privileged countries. The potential use cases for data in each of these countries may vary widely. In Canada access to weather model data may be most beneficial to researchers and individuals working on improving weather models. While in Pakistan, weather model data may provide the basic benefit of increased awareness of forecasted weather conditions to individuals, farmers and families.

As the global climate continues to warm due to anthropogenic emissions of greenhouse gasses, extreme rainfall events are likely to increase in intensity and volume. According the 6th Assessment Report from Working Group 1 of the Intergovernmental Panel on Climate Change (IPCC), increases in the water-holding capacity of the atmosphere as a result of increased global temperatures will lead to “robust increases in precipitation extremes” globally, with an increased magnitude of 4-8% per 1 degree celsius of surface warming (Seneviratne et al., 2021).

In the current year of 2022, Pakistan has experienced tragic and historic flooding due to extreme monsoon events, which resulted in the loss of over 1,700 lives and millions more impacted (Fihlani et al., 2022). Climate change is expected to increase the intensity of monsoon weather events over the region of south Asia (Wang et al., 2021) as global temperatures continue to rise, making weather forecast access an important component of climate adaptation in vulnerable regions of the globe.

This paper presents a software product and its potential applications for enabling access to weather forecasts in underserved nations. The technical components of the Python code and the primary processes of the software product are detailed. Additionally, background information is provided on the historical development of numerical weather prediction, the 2022 flooding catastrophe in Pakistan and the gap in technological solutions between the developing and developed nations.

The objective of this project is to produce a generic data processing pipeline and software framework that can make numerical weather model forecasts easily accessible to individual users, particularly in underprivileged regions of the globe. This interactive portal could be deployed to various nations in need of weather data access in the future and displays weather forecast data on geomap visualizations to enable users to assess incoming regional weather. Underprivileged regions like the Middle East are often overlooked in hydrological studies, yet these regions are likely to experience disproportionate impacts of climate change due to the already extreme climate conditions that exist (Lelieveld et al., 2012).

3. Background

3.1 History

Numerical weather predictions were first theorized in the early 1900s by Norwegian physicist Vilhelm Bjerknes, who suggested that weather variables could be quantified and predicted by computing “system[s] of nonlinear partial differential equations” (National Oceanic and Atmospheric Administration, 2021). Over the first half of the 20th century, meteorologists made advances in weather prediction, beginning with Lewis F. Richardson, who proposed creating numerical weather forecasts by integrating the governing equations for atmospheric circulation in 1922 (Kimura, 2002). These early predictions were computed by hand, taking weeks to produce a numerical prediction and were highly inaccurate. Richardson’s first attempt at numerical predictions yielded a surface pressure 100 times greater than realistic values (Kimura, 2002).

The major challenges with the early efforts at numerical weather prediction were due largely to a lack of computing processing power and insufficient data on initial conditions. In particular, Richardson lacked a thorough understanding of the hydrodynamic characteristics of atmospheric circulation (Kimura, 2002) and employed the use of the continuity equation in his prediction of surface pressure (Donner & Schubert, 2011). Max Margules had previously claimed in a 1904 publication that numerical prediction of surface pressure using the continuity equation was infeasible without precise knowledge of initial conditions, which was beyond the current possibilities at the time (Donner & Schubert, 2011)

Later in the first half of the 20th century, researchers developed a better understanding of upper atmosphere conditions. Meteorologists launched the first radiosonde with a weather balloon, equipped with meteorological instruments to measure temperature, pressure, humidity and winds (US Department of Commerce, National Oceanic & Atmospheric Administration, 2021). Electronic computing power developed over the course of the World Wars and suddenly the computational processing required for numerical weather prediction had become a reality. In 1950, with the use of John von Neumann's ENIAC computer, the first one-day numerical weather prediction was carried out, proving the viability of weather forecasting (US Department of Commerce, National Oceanic & Atmospheric Administration, 2021).

In the second half of the 20th century, accuracy of weather modeling increased proportionally to the increase in available computer processing power. Additionally, new prediction techniques like ensemble modeling enabled longer term forecasts with less noise. Ensemble modeling combines the results of multiple models with slightly different initial conditions to produce a smoother prediction result and account for the chaotic and unpredictable nature of atmospheric conditions (US Department of Commerce, National Oceanic & Atmospheric Administration, 2021). As global observational networks have expanded with the use of weather satellites, numerical weather models have continued to improve in accuracy up to the present day. These observation networks greatly improved the ability of meteorologists to initialize weather models. Today in the United States, over 210 million observations are used as inputs into global and regional scale weather models every day (US Department of Commerce, National Oceanic & Atmospheric Administration, 2021).

3.2 In the Context of the 2022 Pakistan Floods

Pakistan has a high range of diversity in its climatic zones, from high elevation arid regions to tropical forests (Siddiqui, 2010). The diversity in landscape and climatic zones causes highly varied temperatures and weather conditions across the nation (Faisal & Sadiq, 2009), along with high fluctuation in daily weather (Faisal & Sadiq, 2009). Temporally, the amount of precipitation varies greatly throughout the year in Pakistan, with 59% of annual rainfall occurring during the monsoon season (Farooqi et al., 2005).

While Pakistan contributes marginally to global greenhouse gas emissions, it is often listed as one of the most vulnerable nations in the face of climate change (Fahad & Wang, 2019). It stands to endure greater natural disasters and loss of life as well as greater economic hardship in the coming decades due to climate change. The Economic and Social Commission for Asia and the Pacific warns that Pakistan will bear the second largest economic burden in all of South and Southwest Asia due to climate change (United Nations Economic and Social Commission for Asia and the Pacific, 2022). With more frequent extreme weather events, agricultural production will decrease. Agriculture plays a significant role in Pakistani citizens' lives. The entire country relies on agriculture for its economic stability and food security, while 48% of citizens already struggle with food security (Azam & Shafique, 2017). Much of Pakistan's population of over 220 million people also reside in low-lying river deltas, which are at a high risk of flooding.

The effects of extreme weather events are on full display in 2022 in Pakistan, where multi-day extreme precipitation events caused flooding in August that affected one third of Pakistan's population (J. S. et al., 2022) and killed over 1,700 individuals (*Pakistan Monsoon Floods 2022 Islamic Relief Pakistan (12 October, 2022) - Pakistan*, n.d.). In the aftermath of the massive flooding, public health officials are warning of a potential increase in water borne and mosquito borne diseases including dengue, malaria and diarrhea (Sarkar, 2022). There have already been tens of thousands of confirmed cases of both dengue and malaria, as stagnant waters provide breeding grounds for disease carrying mosquitoes (Taylor-Robinson, 2022). Major floods occurred in regions of Pakistan where residents were already vulnerable to health risks like malnutrition. One in nine children admitted to health facilities in the flood affected regions exhibit signs of malnutrition (*Devastating Floods in Pakistan*, n.d.). Children and pregnant women are particularly susceptible to the health impacts of the floods. It is estimated that 73,000 women have given birth in the month after the floods (Devi, 2022) and they often struggle to find adequate health facilities to assist with their childbirth (Sarkar, 2022). In addition to all of these public health crises, there is an added threat of a surge in the number of cases of poliovirus, which remains endemic in Pakistan (Taylor-Robinson, 2022).

In addition to loss of life and public health crises, flood disasters also have lasting impacts that can decrease employment opportunities in the affected areas and cause massive economic strain. Worldwide, floods account for one third of the total economic damage due to natural disasters (Li et al., 2021). In Bangladesh, flooding events have led to a decrease in income and economic stability of the residents in flooded regions (Haque & Jahan, 2015). Agricultural output is particularly susceptible to the impacts of floods, which can damage crop yields and render land unfit for agricultural purposes. Thus, a nation heavily reliant on agricultural production and also susceptible to flooding is at a high risk of widespread economic disruption.

The primary driver of the August 2022 floods in Pakistan was extreme precipitation, which caused monthly rainfall greater than two standard deviations from the mean for the 2001-2021 period (J. S. et al., 2022). In addition to the extreme precipitation, glacial melt during the heatwave from March to May may have exacerbated the flooding (J. S. et al., 2022). As global temperatures continue to increase, extreme weather events like the ones witnessed in Pakistan in August 2022 are likely to occur at a higher frequency (Seneviratne et al., 2021).

The combination of high vulnerability to extreme weather and heavy reliance on agriculture underscores the importance of access to weather data and forecasts in Pakistan. For this reason, this project focuses on the region of Pakistan as a proof of concept for the development and deployment of software solutions aimed to make weather data more accessible. With weather data available, citizens may gain advanced knowledge of extreme temperature or precipitation, which could allow individuals to make necessary preparations to ensure their livelihood and survival. During non-extreme weather events, farmers may be better prepared to manage water resources and crop production with insight into the coming environmental conditions.

3.3 Flood prediction and alerts

An important step in mitigating the effects of devastating floods is advanced warning, which could allow individuals and communities to evacuate or protect their assets. An effective flood warning system uses data from multiple sources including hydrologic and weather modeling, hydrologic monitoring systems and radar precipitation techniques (Li et al., 2021). Data from these sources can be combined to depict a detailed assessment of hydrologic conditions in a given region and determine the risk of a flooding event. Other flood warning systems rely more heavily on GIS tools through spatial mapping of runoff coefficients (El Alaoui El Fels et al., 2017). There are also flood warning systems that depend on technological hardware like arduino units deployed in the field to collect measurements and issue alerts when thresholds are hit (Muhd Zain et al., 2020). Many modern flood prediction and warning systems are increasingly relying on machine learning techniques to mimic the mathematical models of physical processes (Mosavi et al., 2018). Throughout all of these different approaches to flood prediction, a common key requirement is access to data in order to create the assessments of flood risk.

In regions without sophisticated flood warning systems, having access to simple weather forecasts can be a step in the right direction towards disaster preparedness. Pakistan is highly vulnerable to intense flooding, yet lacks an advanced flood warning system (Ahmad & Afzal, 2020). Without an organized system to alert residents in remote regions of incoming danger due to weather events, individuals should be enabled to reach the data themselves and equip themselves with the knowledge they need to make decisions in the face of natural disasters.

This software developed in this project is not a flood warning product, but defines a framework of data collection that could be useful in supporting flood risk assessments and warning systems in future development.

3.4 Data Visualization

Today there is a vast and growing amount of captured and predicted data from all sectors of industry, government and society. Pulling usable insight and value out of big data is often a vital but challenging endeavor. Visualization is one tool out of many that can efficiently convert raw data into accessible analysis for end users. Visualization is especially crucial within the science of weather forecasting. Providing intuitive methods to consume the numerical results of weather models is an essential practice to enable public awareness of incoming weather events. One of the most effective ways of conveying weather forecasts is through mapping visualizations. Mapping tools can provide an easy to use solution for the general public to analyze weather data through a web browser (Diehl et al., 2015).

Researchers have designed low-cost methods for pulling weather forecast data into centralized locations to be visualized and consumed by the public (Astsatryan et al., 2018). These mapping visualizations are well suited for conveying weather data to the public, given the inherently spatial nature of weather phenomenon. Mapping tools also represent an accessible method for sharing data. Even users who are less familiar with technology can become quickly familiar with a map, especially if the mapping tool covers their familiar area.

3.5 Access to Technology

There is a global disparity in access to technology hardware and software solutions between the most and least developed countries (‘Utoikamanu, 2018). Citizens of least developed nations do not have the routine access to the technological devices and connectivity, which have become a standard facet of daily life in the most developed nations. There is huge potential for the improvement of the quality of life in underprivileged countries by improving the access to technology.

The UN has announced its efforts to improve the access to technology and warning systems around the globe, citing that one third of the world’s population are not covered by early warning systems for weather events, with the majority of that population residing in the least developed countries (Dickie, 2022). Many academic resources are devoted to improving technology systems in Western countries, while underprivileged countries in other regions of the globe may be overlooked. The development of software tools for countries that lack access to technology has the potential to provide greater benefit by serving a demonstrated need for access to data.

3.6 Datasets

This weather portal product is powered by numerical weather predictions produced by the Canadian Global Deterministic Prediction System (GDPS). The GDPS provides predictions on a 1500 x 751 latitude-longitude grid (Environment Canada, 2022). The highest resolution of the GDPS is 0.24 x 0.24 degrees or roughly 25 square kilometers. For the purpose of this weather product, the highest resolution data available is used for visualization and data downloads.

The Global Deterministic Prediction System is an output of the Global Environmental Multiscale Model (GEM), a global forecast and data assimilation system. The governing equations integrated in the GEM are forced hydrostatic primitive equations (Côté et al., 1998). The temporal discretization of the GEM is semi-lagrangian (Canadian Meteorological Center, n.d.). To approximate the large global system, periodic boundary conditions are employed to define unit cells (Côté et al., 1998). Spatial discretization is cell-integrated horizontally on an Arakawa C grid (Côté et al., 1998). The model also requires that physical phenomena are parameterized, like surface layer effects, gravity wave drag and condensation.

The Global Deterministic Prediction System runs twice daily with initial times at 00 and 12 hours (Wilson, 2017), and typically produces forecasts out to 10 days ahead of initial time. The numerical results of the GDPS are published on the Environment Canada website in GRIB2 format, after the model executes.

Figure 5. Raw GRIB2 Metadata

```
Dimensions:           (latitude: 1201, longitude: 2400)
Coordinates:
    time              datetime64[ns] ...
    step              timedelta64[ns] ...
    heightAboveGround float64 ...
  * latitude          (latitude) float64 -90.0 -89.85 -89.7 ... 89.7 89.85 90.0
  * longitude         (longitude) float64 -180.0 -179.8 -179.7 ... 179.7 179.9
    valid_time        datetime64[ns] ...
Data variables:
    t2m               (latitude, longitude) float32 ...
Attributes:
    GRIB_edition:            2
    GRIB_centre:             cwao
    GRIB_centreDescription:  Canadian Meteorological Service - Montreal
    GRIB_subCentre:          0
    Conventions:             CF-1.7
    institution:             Canadian Meteorological Service - Montreal
    history:                 2022-09-08T16:43 GRIB to CDM+CF via cfgrib-0.9.1...
```

The raw GRIB2 data contains numerical values of a single weather variable at a single time step for all grid cells at the defined resolution. The coordinates of the weather values in the GRIB2 file are time (the initial time of the model run), step (the temporal step in the model run), height above ground, latitude, longitude and valid time. The figure above shows an example of the attributes of a GRIB2 file containing global data for temperature at two meters elevation.

The web portal developed in this project provides daily forecast data for ten days from the current date. This lead time utilizes all of the weather model data that is available from the Canadian Global Environmental Multiscale model. Providing daily results will enable users to have a strong understanding of the variation and trends in weather for the next ten days. Given limitations of the models used, the forecasted results of short term lead times will be more accurate than the forecasted results at the ten day mark. Updated model data is made available from the Canadian government website every 12 hours. The application refreshes its data source and provides the new data, if available, every six hours.

3.7 Variables of Concern

Numerical weather forecast data has the potential to indicate flash weather events from a sample of hydrological variables including the level of soil moisture, the intensity of forecasted precipitation and the duration of forecasted precipitation (Li et al., 2021). This product provides access to common weather variables to demonstrate the proof of concept of a software tool that provides insight into the likelihood of flood conditions and extreme weather events. The variables chosen for this product include precipitation rate, surface temperature at two meters elevation, wind speed and snow depth.

Pakistan's extreme variation in elevation spans from 0m elevation at the Arabian sea to 8611m at the peak of K2. At higher elevations, there is potential for rain on snow events and subsequent floods that occur in the surrounding catchments (Beniston & Stoffel, 2016). Therefore knowledge of the snow depth at the higher elevations can serve as a useful indicator of potential flooding as well as the presence of freshwater snowmelt reservoirs.

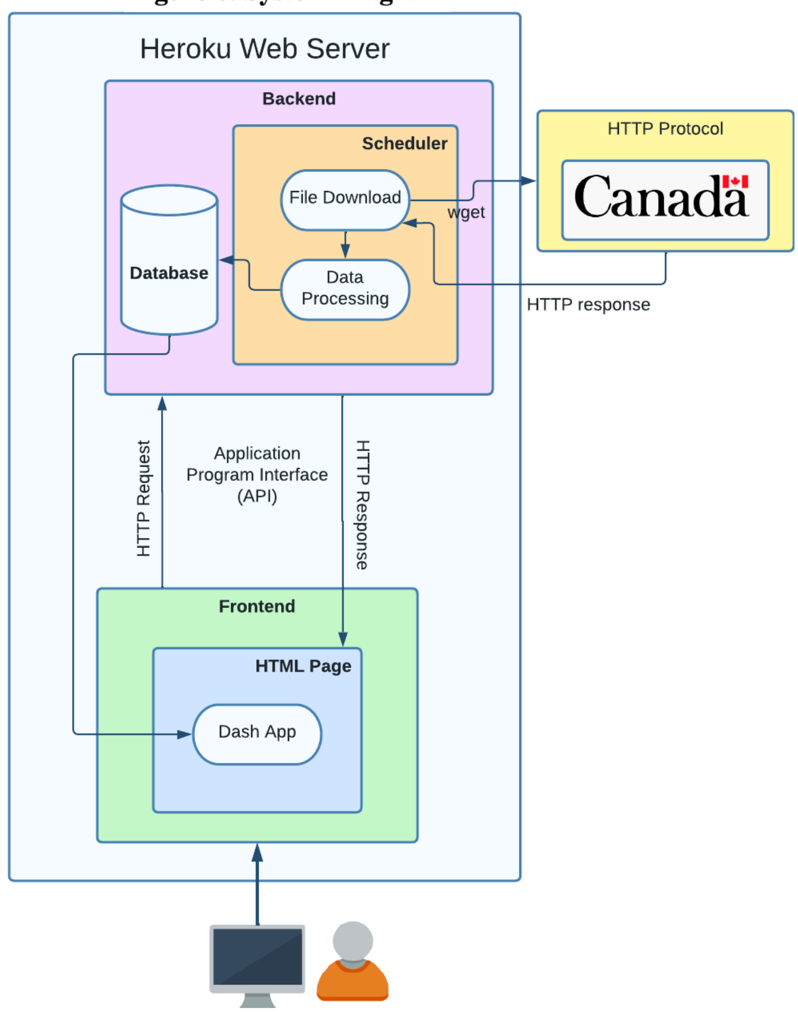
Wind speed can have significant impacts on agricultural output by affecting crop growth (Zhang et al., 2017). Strong winds may have mechanical impacts on plants and damage crop shoots or fell mature plants. However mild winds increase atmospheric circulation and provide more carbon dioxide to the plants, which can facilitate a greater rate of photosynthesis (Gardiner et al., 2016).

Surface temperature provides insight into the likely crop yield of a growing season. Rising global temperatures in general correspond with a decrease in crop yields (Sulser et al., 2021). Not only is temperature an important metric to human comfort, but can also indicate the food security of the agricultural season. Pakistan is at a particular risk of decreased agricultural production as a result of rising temperatures due to climate change (Shakoor et al., 2011). Wheat production decreases with rising temperatures, and the dates for sowing and harvesting may fluctuate depending on the temperature (Shakoor et al., 2011).

Lastly, precipitation rate conveys the intensity of incoming storms which is a strong indicator of flood risk (Camarasa-Belmonte et al., 2015).

4. Methodology

Figure 6. System Diagram



The weather portal developed as part of this project consists of python code and HTML code deployed online onto a web server. The application is composed of both frontend and backend components that fetch data from Environment Canada, process the raw data files, visualize that data in a browser dashboard and allow the download of data extracts to the user's local machine. The major application components are diagrammed here in Figure 6.

To facilitate these features, the Django framework is used to define the application backend, set a URL scheme, handle HTTP requests with Python functions as views and manage HTML templates for the application webpages. Within the backend, schedulers dictate the frequency of requests to the Environment Canada Application Programming Interface (API) to fetch the latest numerical weather forecast data. In this methodology section, an overview will be provided of the logic and tools used in each component of the application,with the full code base available in the appendix.

4.1 Backend

4.1.1 The Common Module

Throughout the backend of the software product, many different files and components require access to the same global variables, sets of data and functions. To reduce redundancy and ensure access to important variables, a Common.py file is defined. This file contains generalized functions to be used in many locations throughout the project. Defining common functions ensures code remains uniform throughout all usages of a given function. When changes to common functions are required, a single update to the source function ripples through imports to the various packages and files in which that function is used, rather than manually updating code in every instance of use, which is prone to leading to errors.

Efforts are made to define as many variables and functions in the Common Module as possible, which enhances the stability of the application by creating more generalized and modular code. Global variables are particularly important to keep generalized in the Common Module, since they dictate many of the data processing steps throughout the application. For example, the dictionary containing the weather variable names is used throughout the application in both the frontend and backend and must stay consistent in all usages. Therefore it is defined once in the Common Module and imported to all other usages.

4.1.2 Raw Data Download

Environment Canada releases its updated Global Deterministic Prediction System forecast data roughly every 12 hours to its HTTP protocol, which is available through a web browser as HTML webpages or through a programmatic file fetch. To create a self-sufficient web portal, this software product makes use of an automated file fetch and wget to execute file transfer protocols over HTTP.

To ensure data is pulled regularly from the Environment Canada API, a scheduler is used to call the file download module. This scheduler executes the file download module every six hours to fetch the latest data that has been released from Environment Canada.

Environment Canada's file transfer protocol requires separate requests for each weather variable at each forecast hour. Therefore, populating the web portal with four weather variables at 15 different forecast hours requires 60 requests to be made to the Environment Canada HTTP protocol. Each of these requests requires a call to a URL with the specific file path for the given variable at the given forecast hour. Thus a programmatic solution is needed to create the list of 60 unique URLs that are required to submit the 60 requests for the data download.

Each URL can be broken down into two separate components. First there is a base file path pointing to the domain of the Environment Canada website, which is uniform for all requests. This base directory is followed by the unique file path for each request, which follows the standard format:

CMC_glb_{variable}_latlon.15x.15_{current_day_utc}{model_run_start}_P{forecast_hour}.grib2

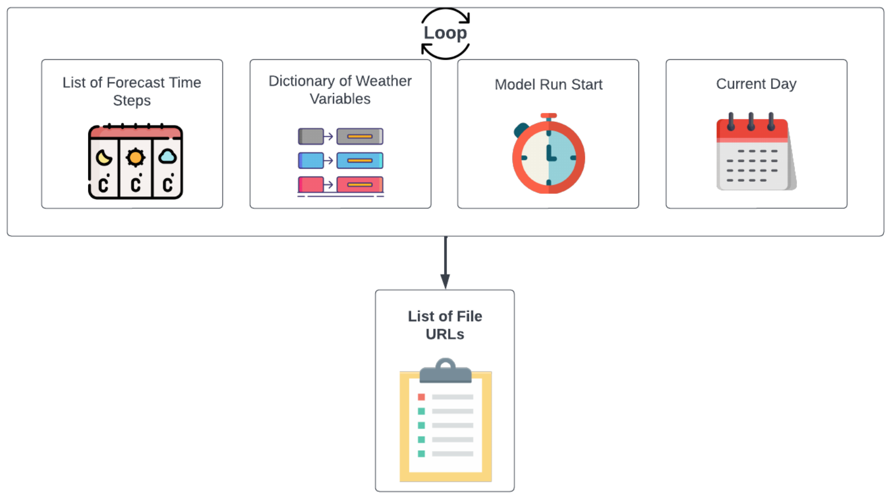
Here the curly brackets denote a variable to be inserted for a unique combination of weather variable and forecast hour. The *variable* label corresponds to the abbreviation for the given weather variable, *current_day_utc* corresponds to today's date in format YYYYMMDD, *model_run_start* corresponds to the initial hour of the model run (either 00 or 12) and *forecast_hour* corresponds to the number of hours in the future from the model run start.

Every six hours, parameters are defined for the unique URLs and a list is programmatically created with all 60 combinations of parameters. "Model run start" is the same for each URL in the same file download execution. "Model run start" is defined by querying the Environment Canada website to find which of the two parameters (00 hour or 12 hour) has the most recent data. Once confirmed, the model run start parameter is fixed for the given batch of file downloads.

The *forecast_hour* parameters are pulled from a global variable in the Common Module. A loop is performed for each value within the list of forecast hours, which defines a URL with all of the remaining parameters for each forecast time step. The *variable* parameter is also pulled from the global variables in the Common Module. A nested loop is performed within the forecast hours loop, over the keys within a dictionary of all weather variables. Therefore, with the two nested loops mentioned above, a URL is created for each variable at each time step. Lastly, the *current_day_utc* parameter is defined by calling the current date from the Datetime python library, and formatting the result in YYYYMMDD format.

Once the functions executing the nested loops have fully completed, a list is returned containing each of the URLs needed to download all weather data at each of the 15 time steps, as shown in Figure 7 here. The next step in the algorithm is to submit the download request to each of the URLs and store the returned data files in a database.

Figure 7. URL List Creation



Each of the 60 requests are submitted using wget and the os python package. The os package is used to execute command line arguments in the terminal from within a python module. With each download request, the returned data is stored in a raw data folder within the project file structure.

4.1.3 Data Storage

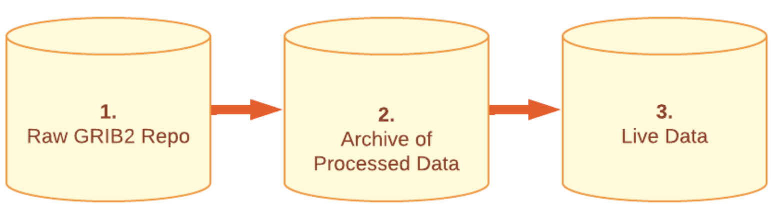


Figure 8. Data Storage Locations

Data is saved in three different locations throughout the data download and data processing steps, as shown here in Figure 8. The first storage location is a repository for raw

GRIB2 data downloaded straight from the Environment Canada website. Files in this repository are organized into subfolders first by country, then weather variable and lastly forecast hour. Raw GRIB2 data is saved to preserve the integrity of the original dataset and for the verification of data processing. Data files accumulated in the raw data folders contain global data, and therefore require large harddrive space. Future deployments of this form of application will require data backups to cloud servers to remove files from the local memory and clear up space.

The second storage location of data files is in the data archive for each country represented in the software product. This folder is populated after raw data is processed from grib2 format to csv format. Whereas raw GRIB2 data has an individual file for each variable at each forecast hour, processed data contains one csv file for each variable with all forecast hours from each data refresh. These files are more efficient in memory since they have been filtered to the country of focus. With each data refresh, new files are appended to the archive directory and a growing database of files accumulates.

The final storage location of weather data is in the live data directory. This directory contains data in the exact same format as the processed data archive. The portal visualizations and download features all point to this file path location as the source of data that powers all application features. With each refresh of data, the live data is overwritten with the most recent files.

4.1.4 Data Processing

A major barrier in accessing a comprehensive weather forecast from the raw prediction values of the Environment Canada Global Environmental Multiscale model is that the raw data files are stored separately by variable and forecast hours. Therefore, if a user is interested in accessing multiple weather variables at multiple different forecast hours, that user must download multiple separate GRIB2 files, have access to software capable of reading GRIB2 files and perform complicated data joins to combine forecast hours and variables into a single file. It is unreasonable to expect an average user to have the skills and time to complete this process whenever they would like to access detailed numerical results from Environment Canada’s weather model. By joining the data prior to visualization, this software product performs those operations programmatically, so the user is free to quickly view and download the data in a friendly format.

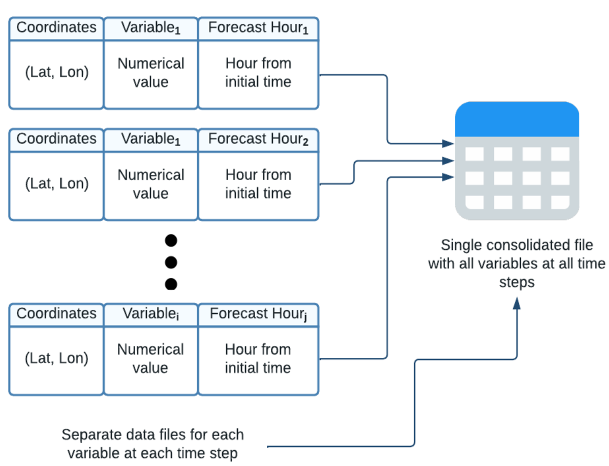
To make the weather data available to the map visualization tools, the data must first be translated from GRIB2 format to a more accessible data format. The Pandas Python package provides a library of functions to analyze, transform and manage tabular data. This software product makes use of Pandas to provide clean tabular data for visualization. The process to convert GRIB2 data to a Pandas dataframe is a multi-step sequence, shown here in Figure 9.

Figure 9. Conversion Steps for GRIB2 Data

The first step is to convert the GRIB2 data to an array. Storing the spatio-temporal GRIB2 data in arrays allows easy conversion to a Pandas dataframe. The Python package Xarray is used to convert the GRIB2 data to an array and subsequently into a dataframe. The resulting dataframe contains the data for a single variable at a single forecast hour. This process is repeated in a loop for all 60 variable and forecast hour combinations. Each of the resulting dataframes is appended to a list to be combined later. The following steps in the data processing sequence handle combining data for each variable into concatenated forecasts.

At this stage in the data processing, the weather data is stored in 60 separate tables for each variable at each forecast hour. Visualizing the data requires joining each of the different data tables on a common key. In this case, the most suitable key is the latitude and longitude coordinates of each grid covered by the numerical weather prediction model. Figure 10 here shows the process of separate tables of weather data coming together to form one comprehensive table. To prepare the data to be joined, the columns containing the latitude and longitude coordinates are set as a multi-column index for each dataframe. Once all dataframe tables have common latitude longitude indexes, a join

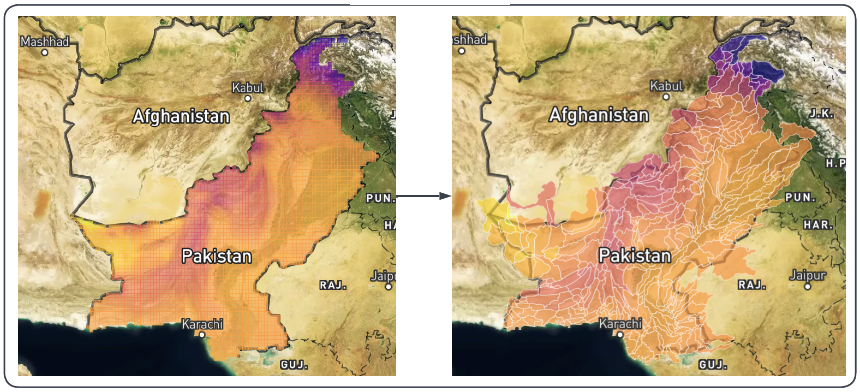
Figure 10. Weather Data Concatenation



is performed to concatenate the individual dataframes together for each variable. The output of this step yields a table containing weather data at all forecast hours for all variables.

4.1.5 Watershed Aggregation

Figure 11.



GRIB2 data supplied by the Environment Canada data portal provides weather forecast data in a grid format of 25km by 25km resolution across the globe. Important insights can be derived from aggregating this data across watersheds, a process visualized here in Figure 11. Watershed level analysis is useful for providing outflow estimates of catchment basins and awareness of water resources for the inhabitants of the watershed (Okolie, 2018). Converting grid measurements to watershed level metrics requires aggregations of the weather variable measurements across the spatial boundaries of the watershed. Open source spatial data is available from HydroSHEDS for data layers on catchment basins and nested sub basins. This data is available as polygons representing boundaries of watersheds. HydroBASINS provides an open data portal for downloading watershed polygons as shapefiles at different hierarchical Pfafstetter levels (Lehner & Grill, 2013).

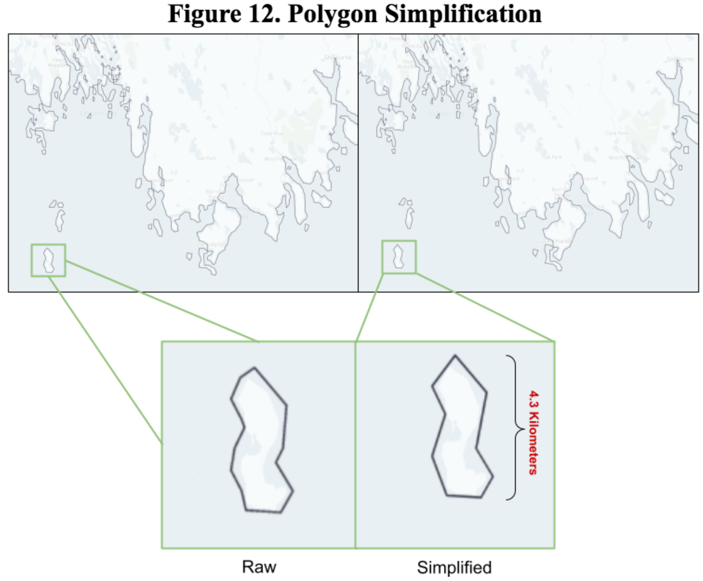
Special attention is paid to the hierarchical level of watershed visualized in this product. The Pfafstetter level used in data aggregations and visualizations is dictated by the resolution of the original source data. At lower granularity Pfafstetter levels, watershed visualizations are inefficient for countries with smaller land areas like Pakistan. For example, at Pfafstetter level 3, the entire land area of Pakistan is divided among only four different catchment basins. Similarly, meaningful analysis is also prevented when the basin granularity is too high. At Pfafstetter level 10, the basin boundaries become too granular and there is not sufficient resolution in the original weather data to aggregate multiple grids across the area of the watershed. Multiple iterations of visualization at different Pfafstetter levels found that Pfafstetter level 7 met the following requirements: Each watershed polygon must be large enough to contain at least two grid samples from the source weather data to be aggregated, and the watershed hierarchy and granularity must not be so low that an insignificant number of watersheds are contained within a country's land area.

This product is intended for analysis across multiple countries. Therefore, the pfafstetter level is dictated by the smallest country included in the visualizations, which in its current state is Pakistan. Future iterations of this product should provide functionality to dynamically select between different Pfafstetter levels of basin and sub basin levels for visualization and aggregation. This feature would be particularly useful for nations with large land areas like Canada, which may provide meaningful analysis from weather aggregations across larger basins at lower Pfafstetter levels.

4.1.6 Simplification of Polygons

Having determined the most appropriate Pfafstetter level of analysis, the shapefiles are downloaded and stored on the same web server that houses the application. The raw shapefiles occupy significant memory on the local server. This causes increased loading times and delays in the user interactions with the portal. To improve the speed of user interactions, the polygon geometry within the shapefiles are simplified to reduce their storage size. To smooth the polylines of watershed boundaries, the Douglas-Peucker Algorithm was used to remove vertices along the watershed boundaries. The remaining vertices are a subset of the original vertices, which prevents the simplified shape from varying greatly from the original (Wu et al., 2003).

Figure 12. Polygon Simplification



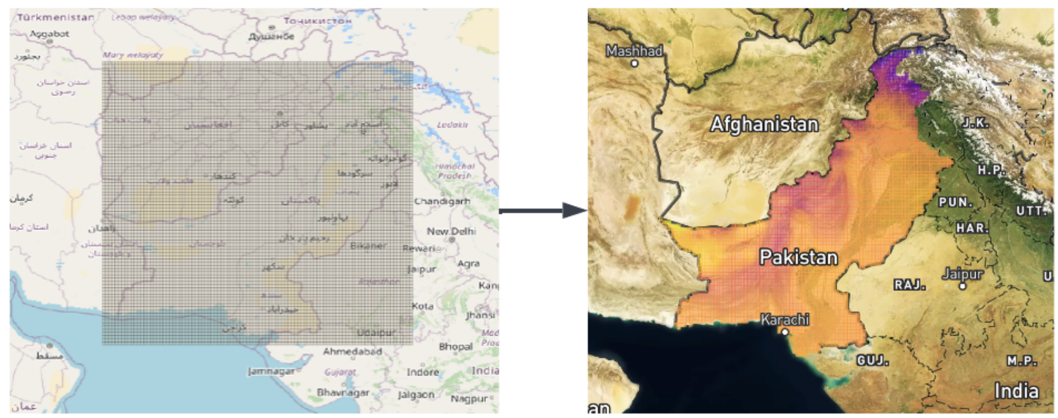
The simplification step takes place after the aggregations have been computed, since the simplified watershed boundaries will vary slightly from the true boundary lines provided by HydroBASIN. Thus simplification slightly affects the visual component of the visualization, but the aggregation calculations are not affected, given that they are executed on the un-simplified true boundaries. Through use of the simplification algorithm, the total file size of the Canadian watershed boundaries is reduced by 77.3%. Simplification greatly increases the speed of interaction within the weather portal, and minimally changes the appearance of the watershed boundaries. With the reduced file sizes, lag times decreased an average of 62% when adjusting the visualization parameters. Figure 12 shows a before and after comparison of polygon simplification on a sample of coastline from Southwest Nova Scotia, with the original unsimplified watershed boundary on the left, and the simplified version on the right.

4.1.7 Grid Geofiles

The raw numerical weather prediction model provides weather data on a grid of latitude and longitude values with a defined resolution. The weather portal created in this project uses the highest resolution weather data available from the Canadian Global Deterministic Prediction System of 25km. To effectively visualize the raw grib2 data, all grid boxes on the map visualizations must present a color-coded representation of the underlying data. To accomplish this, a geofile is required with polygons that represent the model resolution of 25km by 25km cells. A grid cell is created around each grib2 data point, with the center of each grid cell defined to be the latitude and longitude of the grib2 data point.

The number of data points from the grib2 weather data that fall within the boundaries of Pakistan or Canada determine the number of grid boxes and their spatial extent. A spatial join is performed between a geofile of point coordinates from GRIB2 data and a geofile of a grid with 25km resolution that covers the entire extent of the given country. This join results in a polygon layer of a 25km resolution grid, where the center of each grid box falls within the country's borders. This grid is then used to overlay data onto a map in subsequent steps. Figure 13 shows the grid filtering process, from the initial grid on the left to the country-specific grid on the right.

Figure 13. Grid Geofile Processing



4.2 Frontend

4.2.1 Mapping Service

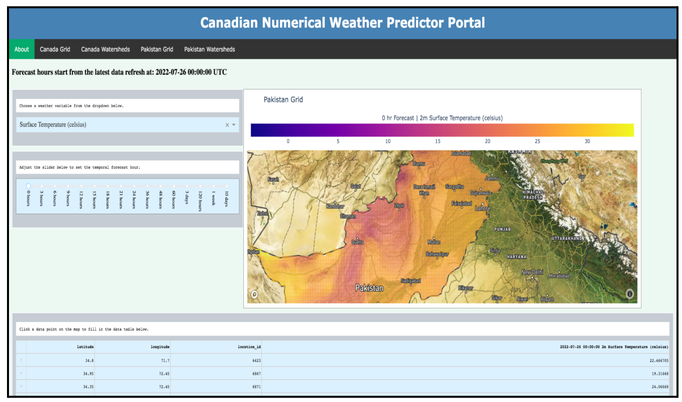
Multiple methods for map visualizations were investigated in the search for a suitable technique to portray weather data on a map in an intuitive manner. The method chosen to visualize weather data must satisfy the core requirements of the project: easy interactivity with the map to select, overlay and download data, flexible customization of the appearance and features, the ability to be embedded in frontend HTML pages and proven methods for deployment within a python application. With these requirements in mind, four different approaches were evaluated.

Kepler.gl is an open source visualization product that specializes in producing interactive maps. It was a strong candidate for this project since maps can be created with python code and embedded into html pages. However, Kepler.gl is limited in its ability to provide robust customization of the map visualizations and does not provide adequate user interaction. In particular, Kepler.gl lacks a distinct separation in privileges between map creator and map user. The visualizations created also could not be catered to user interactions with custom dropdown and slider components. Finally there were no pre-established methods for deploying a python application backend that incorporated dynamic Kepler.gl map visualizations. For those reasons Kepler.gl did not meet the project requirements.

Another potential method for producing geospatial visualizations of weather data is to use the Matplotlib library. While Matplotlib provides an extremely useful library for diverse data visualizations, it lacks sophisticated interaction in a user-facing application. Without methods for fostering advanced interactivity, Matplotlib did not meet the core requirements.

Folium and Leaflet are two additional robust Python libraries which provide mapping tools and data visualization methods. Similar to Kepler.gl and Matplotlib, these libraries fell short in the context of this project due to difficulties packaging dynamic maps in a python deployment. These two libraries also did not support a suitable method of overlaying continuous spatial data on a map.

Figure 14. Frontend User Interface



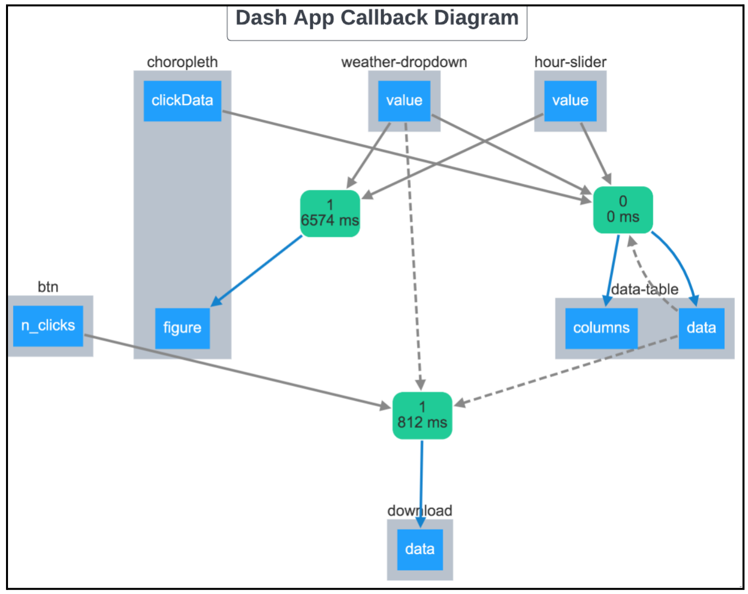
The Plotly graphing library was found to have advanced interactivity features with its visualizations and a robust collection of methods for advanced customization of maps and interaction. Integration with Mapbox allows for high quality mapping tools that support overlay of continuous spatial data. With the integration of Plotly and Dash apps, frontend embedding of Plotly visualizations is highly customizable. Figure 14 shows an example of the dashboard visualizations in this software product. Finally the integration of Django and Dash enabled smooth connectivity between the backend and frontend components of the application and provided a proven method for web deployments. Of the methods investigated, Plotly served the most suitable option for displaying weather data in a highly interactive format in an onlinedeployment.

4.2.2 Visualization and Callbacks

Dash app callbacks facilitate the interactivity of the weather variable dropdown, time step slider, data table and map. Dash callbacks are functions that control inputs and outputs of the HTML components in the Dash app. They enable advanced interaction within the data dashboard and can customize the appearance of the Dash app based on user input. Figure 15 depicts the callback network of the weather portal and the interactions that exist between each HTML component.

Dash app callbacks are defined so that user inputs to one HTML component update the other components. For example, selection of a variable from the weather variable dropdown simultaneously updates the overlay data on the map visualization while also updating the numerical values in the data table to reflect the new variable selection. Similar logic connects the forecast hour slider to the map and datatable. Another example of callback functionality includes the click interactions with the map. When the user clicks a polygon, that data is selected and added to the datatable below the map, which can then be downloaded to the user's local computer.

Figure 15.



Updates to the data table occur in one of two methods. Data is either appended as a new row to the existing data in the datatable, or the pre-existing data is updated to reflect a new user selection. Logic is defined in the Dash app to conditionally update or append data in the data table based on the user interaction. Click interactions on the map will always append data to the table by adding a row with the new grid square or watershed that was clicked, while interactions with the slider and dropdown will always update the data within the table to reflect the new parameters.

When a new variable or forecast hour is selected by the user, the column headers of the datatable update dynamically to reflect the data. The column header for the weather variable includes the date and time of the current forecast. Adjusting the time step slider updates the date and time listed in the data table header to clearly indicate to the user what forecast they are observing.

Dash callbacks also enable the user to download data as a CSV file straight from the weather portal. After selecting locations on the map to add numerical data to the data table, users can click “Download Data” to download the contents of the datatable to their local machine. The CSV served to the user’s local file directory contains time series data of all forecast hours for the given variable at all the selected locations.

4.3 Application Programming Interface

4.3.1 Django

Requests between the frontend and backend of the application are handled within the Django framework. Within the Django backend, the URL scheme is defined for the links and associated webpages within the portal. When an HTTP request is submitted to a URL within the URL scheme, the Django views handle the request and return a defined response. In this application, the HTTP response renders an HTML page from the Django templates directory.

Django templates are used in this product to store the static components of HTML pages, along with HTML code to indicate where the dynamic portions of the webpage will be inserted. In this application, the static HTML components include only the top header and the top navigation bar. The rest of the HTML components are dynamically inserted with a Dash app that contains the map visualizations and interactive controls.

The Django-Dash framework enables smooth integration of Dash apps within a Django backend, and allows Dash content to be dynamically inserted into the Django templates as an iframe. This allows much of the HTML components to be defined within the Dash app as python code.

4.4 Cloud Deployment

4.4.1 Git Version Control System

During development and testing of the weather portal, code changes and version history were tracked by regularly pushing commits with code changes to a remote repository hosted on GitHub. A master branch is maintained with final code and new branches of code were created and merged to master as needed. The version control system provides a backup storage of the code bank for this project as well as a chronological ledger of updates to the application over time. The version control system is also crucial for the deployment of the application to a remote cloud server.

4.4.2 Heroku Platform as a Service (PaaS)

Finished code is deployed to a web server after testing on a local server. To deploy online, Heroku is selected to host the required web server and run the python code for the Django-Dash application. In order to deploy the application onto a Heroku server, there are required files that must first be defined.

One file contains all of the python-package dependencies that the application needs in order to run. These dependencies are defined in the 'requirements.txt' file at the root of the project repository. This file contains the names and version numbers of all of the python packages that a server needs to install in order for the application to run. When the Heroku server is deployed, one of the first actions the server executes is to install the required dependencies using the package management system Pip.

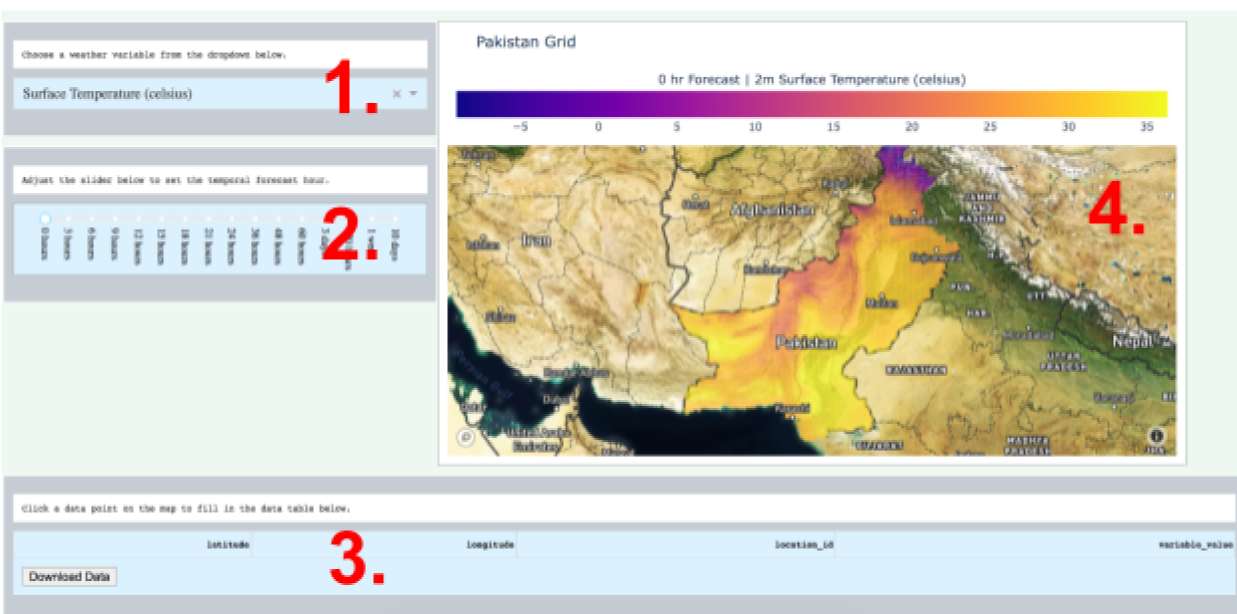
The other Heroku-required file is a Procfile that tells the server what command should be executed to run the app. The Procfile for this application contains a command that attaches the app to the HTTP routing stack of Heroku and enables it to receive web traffic.

When deploying the application code to the web server, Heroku copies the python files from the GitHub version control repository. This is an efficient way to ensure that the most up-to-date code is deployed online, since any changes pushed to GitHub will also be deployed to the Heroku server.

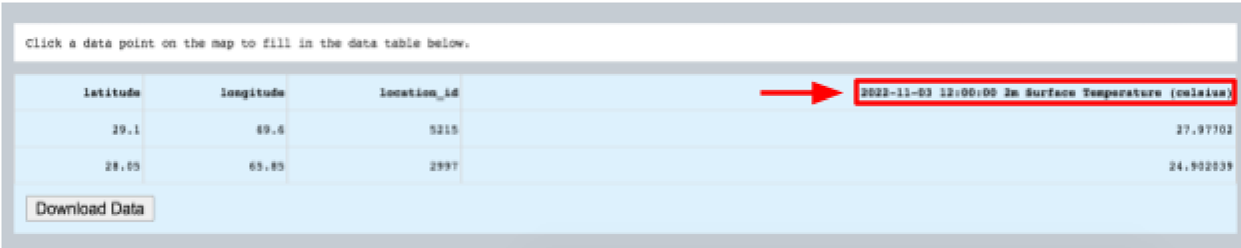
A limitation on the free-tier server currently used for deployment is the memory constraint of 500 megabytes during runtime. This constraint prevents the full application from being deployed online in its entirety, which occupies roughly 700 megabytes of memory. To accommodate this constraint, the application is deployed with only the data visualization dashboards that cover the region of Pakistan, while the dashboards that cover Canada are only available in the offline product.

4.5 Portal Functionality

The primary functionality of the portal revolves around the map visualization referenced by number 4 in the figure below, which covers either the region of Pakistan or Canada. This map portrays the weather data coded by color. The dropdown selector (1) controls which weather metric is overlaid onto the map. Temporal forecast hour is controlled by the hour slider (2) to show the weather data the selected number of hours into the future from the most “model_run_start” parameter of either 00 or 12 hour.. The data table (3) outputs numerical values from the locations the user clicks on the map. Once clicked, either the grid square or watershed polygon will display a highlighted outline to indicate which location has been selected. That location will then be added to the datatable below. Finally, if the user wishes to download the data to their local computer, the ‘Download Data’ button below the datatable will download all forecast hours as a time series for the selected locations within the datatable. To remove a selected area from the datatable, the user must click the polygon on the map again to un-highlight the selection.



When changing the values in either the weather variable dropdown (1) or the forecast hour slider (2), the numerical values in the datatable will automatically update to reflect the new inputs. Additionally, the time and date of the current forecast is indicated in the column title of the weather variable column highlighted in red in the figure below. When the forecast hour slider is adjusted, the column title adjusts as well to display the new forecast date and time.



5. Conclusion

The floods of 2022 are not the first major flooding events to devastate Pakistan in recent history. The monsoon floods of 2010 killed over 1700 people and destroyed or damaged over 1.1 million homes (Kirsch et al., 2012). Researchers have shown that by using numerical weather prediction methods, the 2010 deluge rainfall events were highly predictable 6-8 days in advance (Webster et al., 2011). There is a demonstrated need for better accessibility to weather forecasts and for the presence of alert systems for severe weather events. Developing nations like Pakistan have very limited resources, therefore the expansion of projects like this one could benefit many in the developing world.

The Canadian Global Environmental Multiscale model has been found to provide sufficiently accurate predictions for Middle Eastern regions (Mohammadlou et al., 2021), and given its level of resolution, it is a suitable model for providing access to weather forecasts for individuals. Low-cost tools like the one developed in this project show a promising opportunity for connecting end users to publicly available data. The combination of open source tools incorporated in this software product provide a promising method for providing a data access pipeline through a portal. Incorporation of the Plotly API enables accessible visualization opportunities with a high level of customization..

Providing values of forecasted weather variables has the potential to provide advance warning for individuals who are vulnerable to flooding events. Advanced notice could give increased preparation and evacuation time. Similarly, individuals and families who are reliant on agriculture for their livelihood and sustenance can benefit greatly from having a forecasted awareness of the near-future hydrological conditions in their area. Knowledge of incoming precipitation levels and temperatures, can enable farmers to make better planning decisions, to increase their crop yield and manage their land with data-driven decisions. In the face of increasingly uncertain and dynamic weather due to climate change, people need all the resources they can get to stay aware of their environmental conditions.

6. Final Remarks

6.1 Future improvements

This project provides a framework and data pipeline for the visualization of weather model data, and there are many opportunities to build upon the progress made thus far. Opportunity for the expansion of this project will include a deployment to Pacific island nations highly vulnerable to the rising ocean levels caused by climate change. Further expansion may incorporate data from multiple numerical weather prediction models from other nations, including American models from the National Oceanic and Atmospheric Administration and models from European nations.

Future versions of this project can also include a greater range of weather variables to provide a more complete picture of conditions over the study regions. A particularly useful variable to include is soil moisture, which is another strong indicator of conditions that may lead to a flood. Other variables to incorporate include wind direction. Wind direction may assist farmers to plant wind-buffering crops and provide better knowledge of where to plant crops susceptible to wind exposure. With greater weather data collected and tracked, future versions of this project could seek to support flood warning systems through providing data on indicators related to flood risk.

When aggregating data across watersheds, a useful future functionality will be to dynamically select the Pfafstetter level of watershed hierarchy. This would enable the user to adjust the granularity of the watershed analysis, particularly for countries with large land areas that encompass a wide range of watershed hierarchies.

There are also opportunities for statistical analysis of the data collected to show historical means, extremes and standard deviations. Access to historical data can also be supported in future versions with a query framework to filter and download historical weather forecasts. The performance of the application can be improved through testing with different file storage types and by optimizing the callbacks within the Dash application. The ideal direction of an application of this nature would be to incorporate a global scope of data capture and visualization. This could result in a modular version of this application, capable of covering any nation in the world and providing usable information to the global community.

6.2 Final Words

While the least developed countries in the world contribute only marginally to greenhouse gas emissions, they are likely to experience some of the strongest impacts of climate change, with minimal resources to adapt to and manage natural disasters. In 2019, least developed countries were estimated to have contributed only 1.1% of global carbon dioxide emissions (United Nations Conference on Trade and Development, 2021). Special efforts must be taken to provide support through all available channels to equip underprivileged nations with the resources to withstand climate disasters. This includes technology solutions which focus on providing access to useful data to individuals who live in regions that are technologically isolated and vulnerable to extreme weather. There is a strong obligation from the most developed nations with the highest per capita greenhouse emissions to put forward efforts to increase the resilience of the most vulnerable nations in the face of climate change.

7. References

Ahmad, D., & Afzal, M. (2020). Flood hazards and factors influencing household flood perception and mitigation strategies in Pakistan. *Environmental Science and Pollution Research*, *27*(13), 15375–15387. https://doi.org/10.1007/s11356-020-08057-z

Astsatryan, H., Grogoryan, H., Gyulgyulyan, E., Hakobyan, A., Kocharyan, A., Narsisian, W., Sahakyan, V., Shoukourian, Y., Abrahamyan, R., Petrosyan, Z., & Aligon, J. (2018). Weather data visualization and analytical platform. *Scalable Computing: Practice and Experience*, *19*(2), 79–86. https://doi.org/10.12694/scpe.v19i2.1351

Azam, A., & Shafique, M. (2017). Agriculture in Pakistan and its Impact on Economy―A Review. *International Journal of Advanced Science and Technology*, *103*, 47–60. https://doi.org/10.14257/ijast.2017.103.05

Beniston, M., & Stoffel, M. (2016). Rain-on-snow events, floods and climate change in the Alps: Events may increase with warming up to 4 °C and decrease thereafter. *Science of The Total Environment*, *571*, 228–236. https://doi.org/10.1016/j.scitotenv.2016.07.146

Camarasa-Belmonte, A. M., & Butrón, D. (2015). Estimation of flood risk thresholds in Mediterranean areas using rainfall indicators: Case study of Valencian Region (Spain). Natural Hazards, 78(2), 1243–1266. https://doi.org/10.1007/s11069-015-1769-8

Canadian Meteorological Center. (n.d.). *The Global Environmental Multiscale Model* . Environment Canada. Retrieved September 10, 2022, from https://collaboration.cmc.ec.gc.ca/science/rpn/gef_html_public/INTRODUCTION/gem_intro.html

Côté, J., Gravel, S., Méthot, A., Patoine, A., Roch, M., & Staniforth, A. (1998). The operational CMC–MRB global environmental multiscale (GEM) model. Part I: Design considerations and formulation. *Monthly Weather Review*, *126*(6), 1373–1395. https://doi.org/10.1175/1520-0493(1998)126<1373:TOCMGE>2.0.CO;2

Devastating floods in Pakistan. (n.d.). UNICEF. Retrieved November 6, 2022, from https://www.unicef.org/emergencies/devastating-floods-pakistan-2022

Devi, S. (2022). Pakistan floods: Impact on food security and health systems. *The Lancet*, *400*(10355), 799–800. https://doi.org/10.1016/s0140-6736(22)01732-9

Dickie, G. (2022, March 23). U.N. to roll out global early-warning systems for extreme weather. *Reuters*. https://www.reuters.com/business/environment/un-roll-out-global-early-warning-systems-extreme-weather-2022-03-23/

Diehl, A., Pelorosso, L., Delrieux, C., Saulo, C., Ruiz, J., Gröller, M. E., & Bruckner, S. (2015). Visual analysis of spatio-temporal data: Applications in weather forecasting. *Computer Graphics Forum*, *34*(3), 381–390. https://doi.org/10.1111/cgf.12650

Donner, L., & Schubert, W. (2011). *The development of atmospheric general circulation models: Complexity, synthesis and computation* (pp. 3–15). Cambridge University Press.

El Alaoui El Fels, A., Bachnou, A., & Alaa, N. (2017). Combination of GIS and mathematical modeling to predict floods in semiarid areas: Case of Rheraya watershed (Western High Atlas, Morocco). *Arabian Journal of Geosciences*, *10*(24). https://doi.org/10.1007/s12517-017-3345-x

Environment Canada. (2022, September). *GDPS data in GRIB2 format: 25 km*. https://weather.gc.ca/grib/grib2_glb_25km_e.html

Fahad, S., & Wang, J. (2019). Climate change, vulnerability, and its impacts in rural Pakistan: A review. *Environmental Science and Pollution Research*, *27*(2), 1334–1338.

https://doi.org/10.1007/s11356-019-06878-1

Faisal, N., & Sadiq, N. (2009). Climatic zonation of Pakistan through precipitation-effectiveness index. *Pakistan Journal of Meteorology*, *6*(11), 51–60.

Farooqi, A. B., Khan, A. H., & Mir, H. (2005). Climate change perspective in pakistan. *Pakistan Journal of Meteorology*, *2*(3), 11–21.

Fihlani, P., Fraser, S., & AbdulJalil, Z. (2022, September 6). Pakistan floods: Officials struggle to stop biggest lake overflowing. BBC. Retrieved September 10, 2022, from https://www.bbc.com/news/world-asia-62764224

Gardiner, B., Berry, P., & Moulia, B. (2016). Review: Wind impacts on plant growth, mechanics and damage. *Plant Science*, *245*, 94–118. https://doi.org/10.1016/j.plantsci.2016.01.006

Goujon, A., Wazir, A. & Gailey, N. (2020). Pakistan : un pays de plus de 200 millions d'habitants en retard dans la transition démographique. Population & Sociétés, 576, 1-4. https://doi.org/10.3917/popsoc.576.0001

Haque, A., & Jahan, S. (2015). Impact of flood disasters in Bangladesh: A multi-sector regional analysis. *International Journal of Disaster Risk Reduction*, *13*, 266–275. https://doi.org/10.1016/j.ijdrr.2015.07.001

J. S., N., Kushwaha, A. P., Singh, R., Malik, I., Solanki, H., Chupal, D. S., Dangar, S., Mahto, S. S., Mishra, V., & Vegad, U. (2022). The Pakistan flood of August 2022: Causes and implications. Wiley. http://dx.doi.org/10.1002/essoar.10512560.1

Kimura, R. (2002). Numerical weather prediction. *Journal of Wind Engineering and Industrial Aerodynamics*, *90*(12–15), 1403–1414. https://doi.org/10.1016/s0167-6105(02)00261-1

Kirsch, T. D., Wadhwani, C., Sauer, L., Doocy, S., & Catlett, C. (2012). Impact of the 2010 Pakistan floods on rural and urban populations at six months. *PLoS Currents*.

https://doi.org/10.1371/4fdfb212d2432

Lehner, B., & Grill, G. (2013). Global river hydrography and network routing: Baseline data and new approaches to study the world's large river systems. *Hydrological Processes*, *27*(15), 2171–2186. https://doi.org/10.1002/hyp.9740

Lelieveld, J., Hadjinicolaou, P., Kostopoulou, E., Chenoweth, J., El Maayar, M., Giannakopoulos, C., Hannides, C., Lange, M. A., Tanarhte, M., Tyrlis, E., & Xoplaki, E. (2012). Climate change and impacts in the Eastern Mediterranean and the Middle East. Climatic Change, 114(3–4), 667–687. https://doi.org/10.1007/s10584-012-0418-4

Li, D., Fang, Z. N., & Bedient, P. B. (2021). Flood early warning systems under changing climate and extreme events. In *Climate Change and Extreme Events* (pp. 83–103). Elsevier. http://dx.doi.org/10.1016/b978-0-12-822700-8.00002-0

Mohammadlou, M., Bahremand, A., Princz, D., Kinar, N., & Razavi, S. (2021). *Validation of the global environmental multiscale model (GEM) for Iran*. Research Square Platform LLC. http://dx.doi.org/10.21203/rs.3.rs-216566/v1

Mosavi, A., Ozturk, P., & Chau, K. (2018). Flood prediction using machine learning models: Literature review. *Water*, *10*(11), 1536. https://doi.org/10.3390/w10111536

Muhd Zain, N., Elias, L. S., Paidi, Z., & Othman, M. (2020). Flood Warning and Monitoring System (FWMS) using GSM Technology. *Journal of Computing Research and Innovation*, *5*(1), 7–18. https://doi.org/10.24191/jcrinn.v5i1.158

National Oceanic and Atmospheric Administration. (2021, January 21). The History of Numerical Weather Prediction. NOAA Celebrates 200 Years of Science, Service and Stewardship. Retrieved September 10, 2022, from https://celebrating200years.noaa.gov/foundations/numerical_wx_pred/welcome.html#intr o

Okolie, C. J. (2018, February). Watershed Analysis and its Applications. *Workshop on GIS-Based LRID for NEWMAP Implementation in Akwa-Ibom State*.

Pakistan monsoon floods 2022 islamic relief Pakistan (12 october, 2022) - Pakistan. (n.d.). ReliefWeb. Retrieved November 6, 2022, from https://reliefweb.int/report/pakistan/pakistan-monsoon-floods-2022-islamic-relief-pakistan-12-october-2022

Reher, D. S. (2011). Economic and Social Implications of the Demographic Transition. Population and Development Review, 37, 11–33. http://www.jstor.org/stable/41762397

Sarkar, S. (2022). Pakistan floods pose serious health challenges. BMJ, o2141. https://doi.org/10.1136/bmj.o2141

Seneviratne, S.I., X. Zhang, M. Adnan, W. Badi, C. Dereczynski, A. Di Luca, S. Ghosh, I. Iskandar, J. Kossin, S. Lewis, F. Otto, I. Pinto, M. Satoh, S.M. Vicente-Serrano, M. Wehner, and B. Zhou, 2021: Weather and Climate Extreme Events in a Changing Climate. In Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 1513–1766, doi:10.1017/9781009157896.013.

Shakoor, U., Saboor, A., Ali, I., & Mohsin, A. Q. (2011). Impact of climate change on agriculture: Empirical evidence from arid region. *Pakistan Journal of Agricultural Sciences*, *48*(4), 327–333.

Siddiqui, K. M. (2010). *Asia-Pacific Forestry Sector Outlook Study: Country Report - Pakistan*. Food and Agriculture Organization of the United Nations.

https://www.fao.org/3/W7704E/w7704e03.htm

Sulser, T., Wiebe, K. D., Dunston, S., Cenacchi, N., Nin-Pratt, A., Mason-D'Croz, D., Robertson, R. D., Willenbockel, D., & Rosegrant, M. W. (2021). *Climate Change and hunger: Estimating costs of adaptation in the agrifood system*. International Food Policy Research Institute. http://dx.doi.org/10.2499/9780896294165

Taylor-Robinson, A. (2022). Pakistan floods: Incidence of vector- and water-borne infectious diseases soars. Microbes and Infectious Diseases, 0(0), 0–0. https://doi.org/10.21608/mid.2022.166660.1392

United Nations Conference on Trade and Development. (2021, October). *Smallest footprints, largest impacts: Least developed countries need a just sustainable transition*. United Nations Conference on Trade and Development. https://unctad.org/topic/least-developed-countries/chart-october-2021

United Nations Economic and Social Commission for Asia and the Pacific. (2022). *Pathways to Adaptation and Resilience in South and South-West Asia* (pp. 4–5). UN ESCAP.

US Department of Commerce, National Oceanic & Atmospheric Administration. (2021, January 21). *The history of numerical weather prediction*. https://celebrating200years.noaa.gov/foundations/numerical_wx_pred/welcome.html#early

'Utoikamanu, F. (2018, December). *Closing the Technology Gap in Least Developed Countries*. United Nations Chronicle; United Nations. https://www.un.org/en/chronicle/article/closing-technology-gap-least-developed-countries

Wang, B., Biasutti, M., Byrne, M. P., Castro, C., Chang, C.-P., Cook, K., Fu, R., Grimm, A. M., Ha, K.-J., Hendon, H., Kitoh, A., Krishnan, R., Lee, J.-Y., Li, J., Liu, J., Moise, A., Pascale, S., Roxy, M. K., Seth, A., … Zhou, T. (2021). Monsoons climate change assessment. *Bulletin of the American Meteorological Society*, *102*(1), E1–E19. https://doi.org/10.1175/bams-d-19-0335.1

Webster, P. J., Toma, V. E., & Kim, H.-M. (2011). Were the 2010 Pakistan floods predictable? *Geophysical Research Letters*, *38*(4), n/a-n/a. https://doi.org/10.1029/2010gl046346

Wilson, J. D. (2017, April). *Canada's Global Environmental Multiscale NWP model*. University of Alberta. https://www.eas.ualberta.ca/jdwilson/EAS372_17/eas372_NWP_CMC_2017_6Apr2017.pdf

Wu, S.-T., Marquez, M. R. G., & 16th Brazilian Symposium on Computer Graphics and Image Processing, SIBGRAPI 2003 16 2003 10 12 - 2003 10 15. (2003). A non-self-intersection douglas-peucker algorithm. Brazilian Symposium of Computer Graphic and Image Processing, 2003-january, 60–66. https://doi.org/10.1109/SIBGRA.2003.1240992

Zhang, P., Zhang, J., & Chen, M. (2017). Economic impacts of climate change on agriculture: The importance of additional climatic variables other than temperature and precipitation. *Journal of Environmental Economics and Management*, *83*, 8–31. https://doi.org/10.1016/j.jeem.2016.12.001

8. Appendix

Figure 1. Canadian Watershed Geofile Polygons

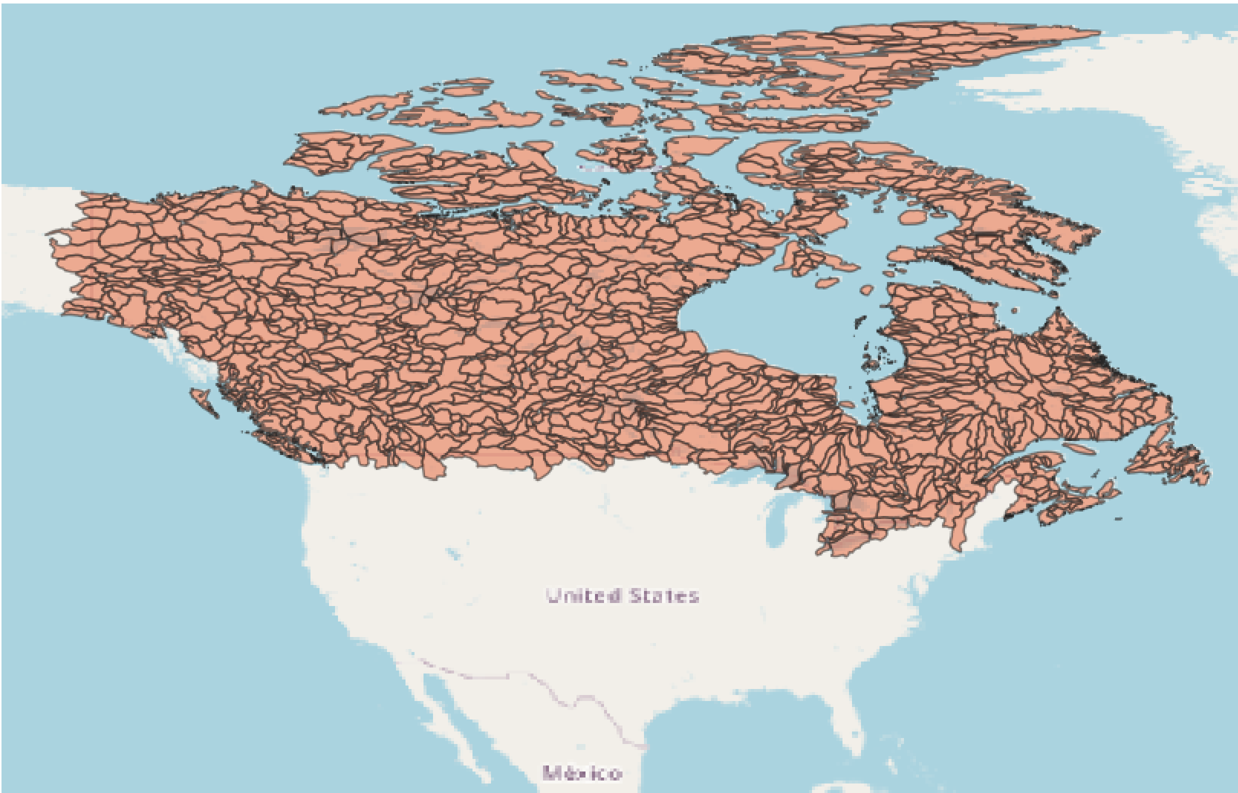


Figure 2. Canadian Grid Geofile Polygons

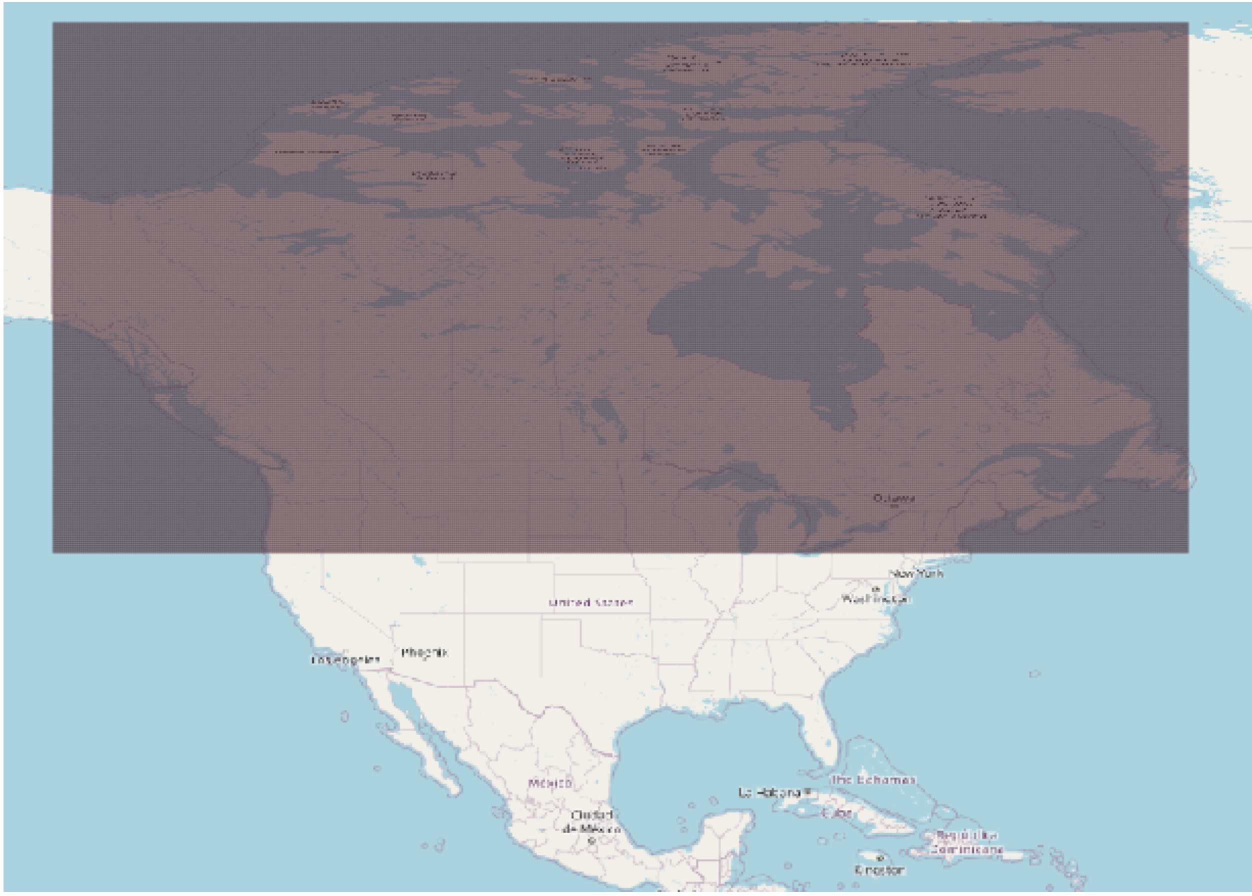


Figure 3. Pakistan Grid Geofile Polygons

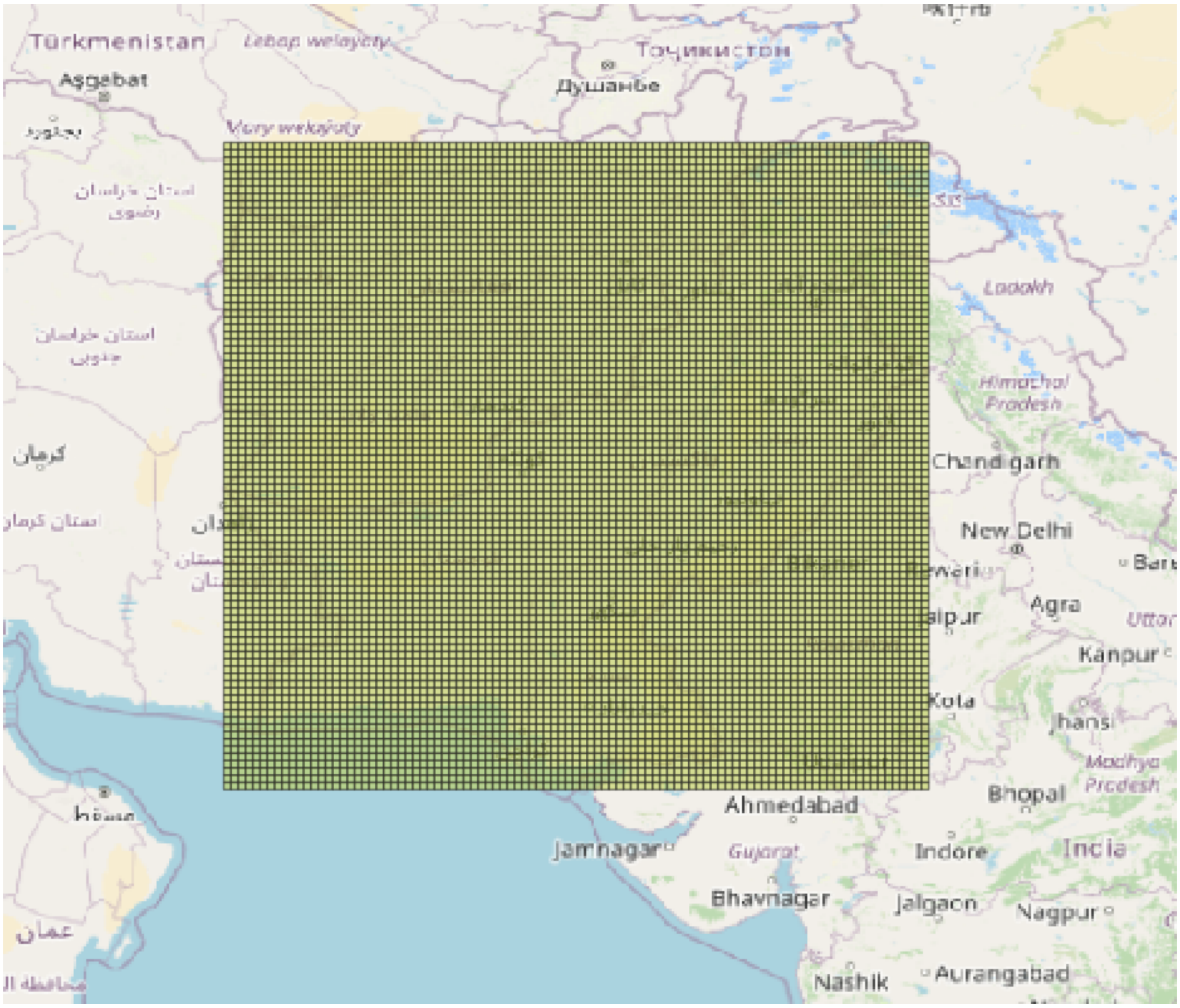


Figure 4. Pakistan Watersheds Geofile Polygons

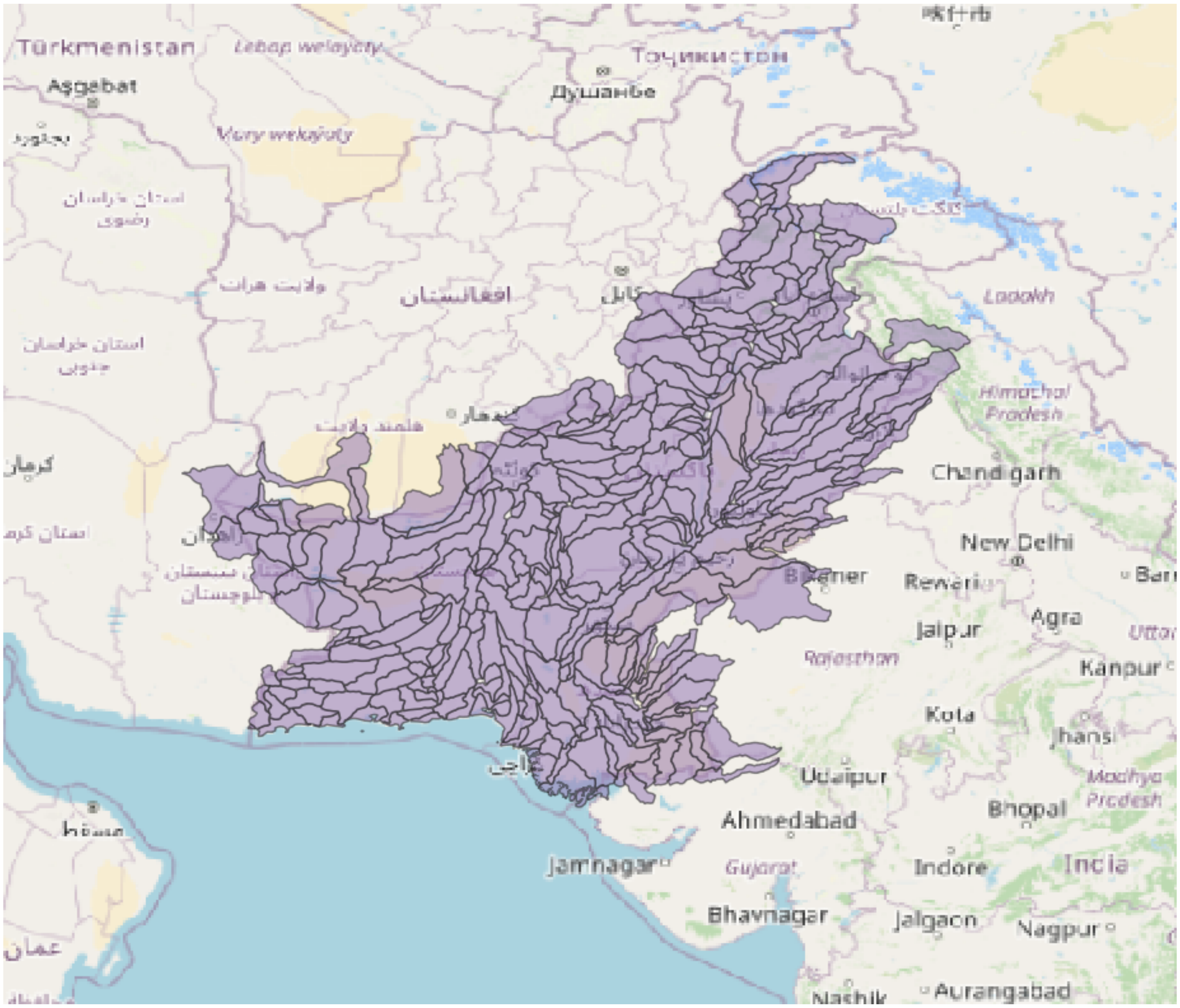


Figure 5. Raw GRIB2 Metadata

```
Dimensions:           (latitude: 1201, longitude: 2400)
Coordinates:
    time              datetime64[ns] ...
    step              timedelta64[ns] ...
    heightAboveGround float64 ...
  * latitude          (latitude) float64 -90.0 -89.85 -89.7 ... 89.7 89.85 90.0
  * longitude         (longitude) float64 -180.0 -179.8 -179.7 ... 179.7 179.9
    valid_time        datetime64[ns] ...
Data variables:
    t2m               (latitude, longitude) float32 ...
Attributes:
    GRIB_edition:            2
    GRIB_centre:             cwao
    GRIB_centreDescription:  Canadian Meteorological Service - Montreal
    GRIB_subCentre:          0
    Conventions:             CF-1.7
    institution:             Canadian Meteorological Service - Montreal
    history:                 2022-09-08T16:43 GRIB to CDM+CF via cfgrib-0.9.1...
```

Figure 6. System Diagram

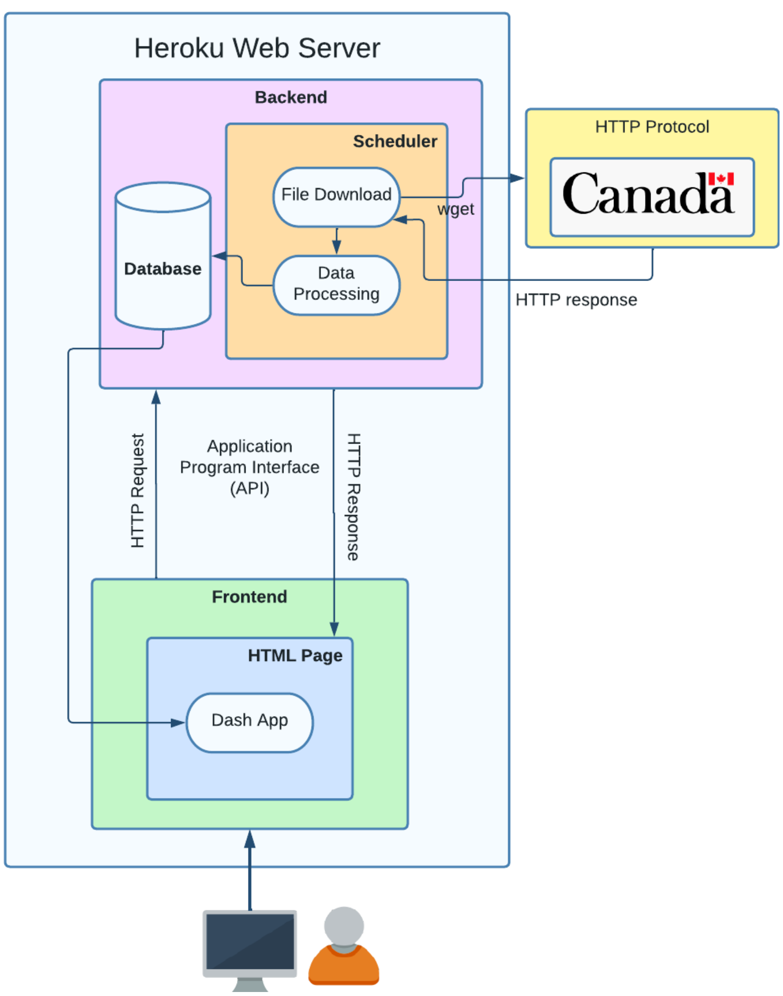


Figure 7. URL List Creation

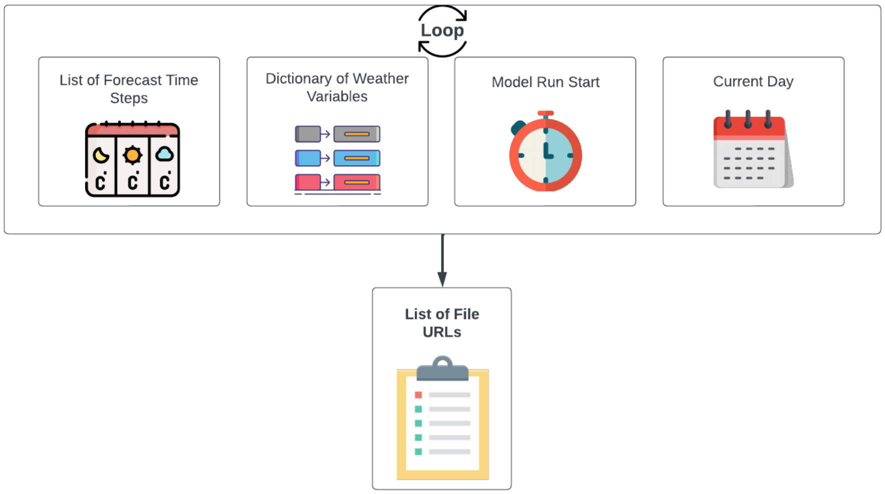


Figure 8. Data Storage Locations

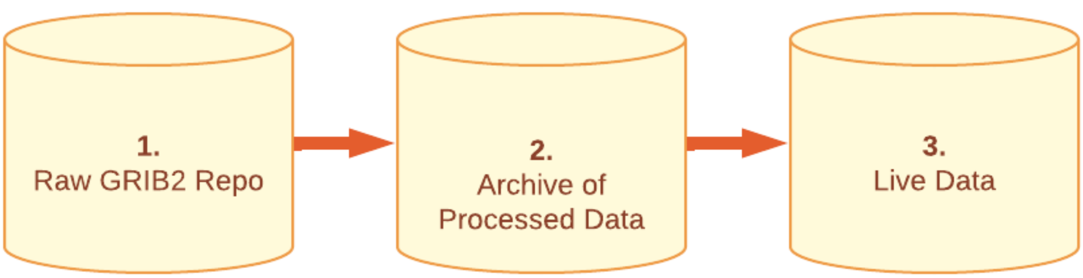


Figure 9. Conversion Steps for GRIB2 Data

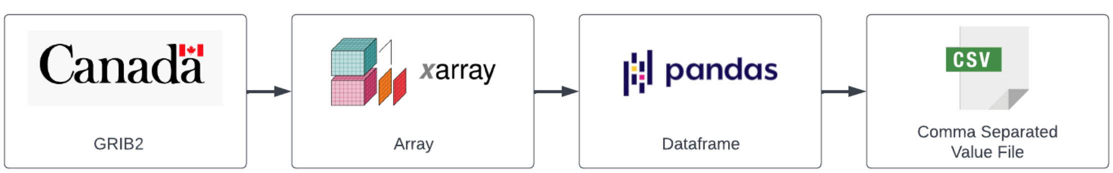


Figure 10. Weather Data Concatenation

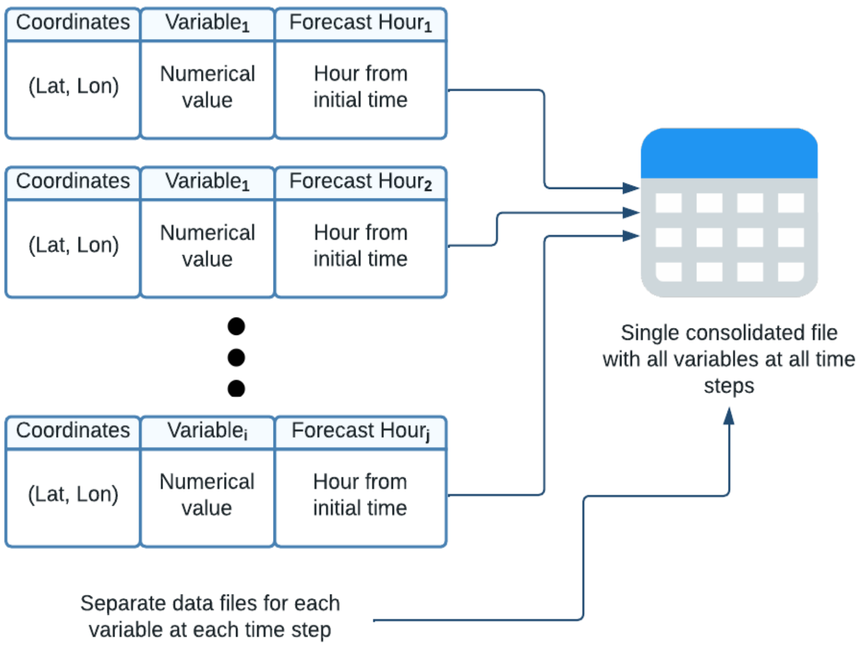


Figure 11. Watershed Aggregation

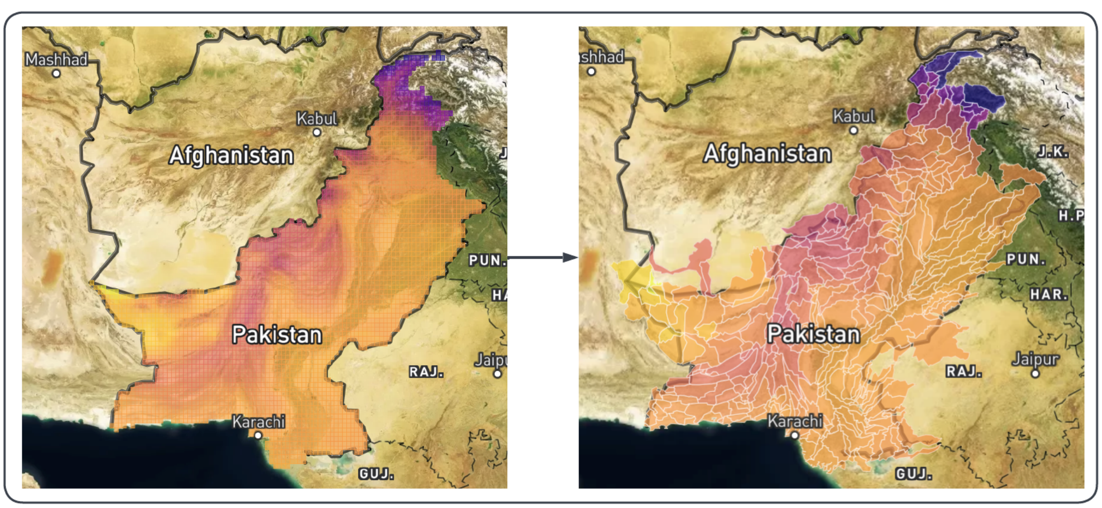


Figure 12. Polygon Simplification

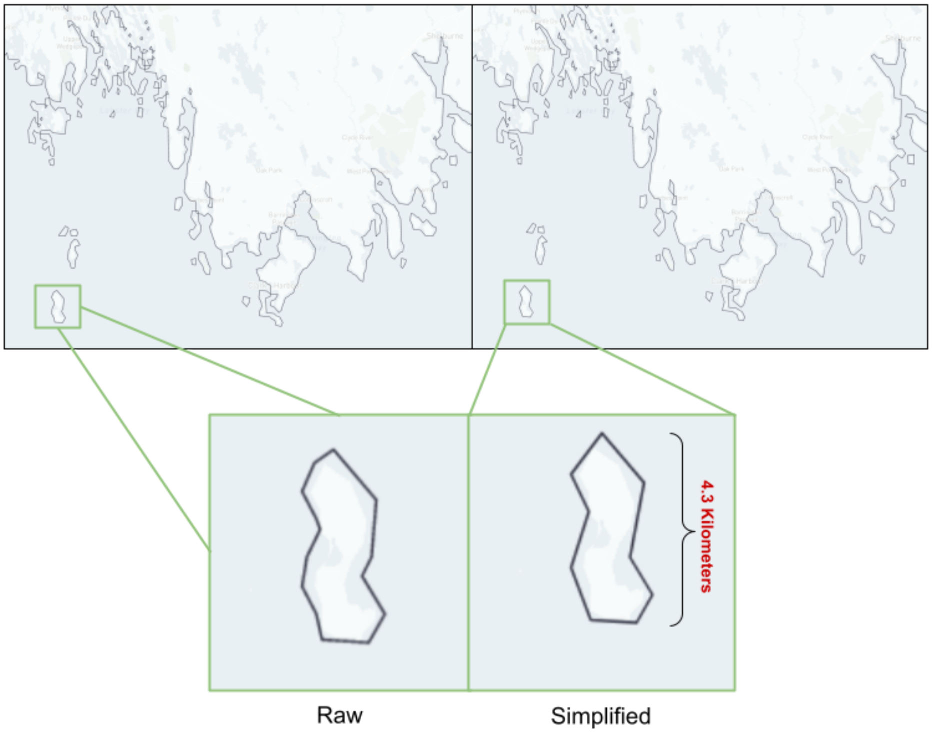


Figure 13. Grid Geofile Processing

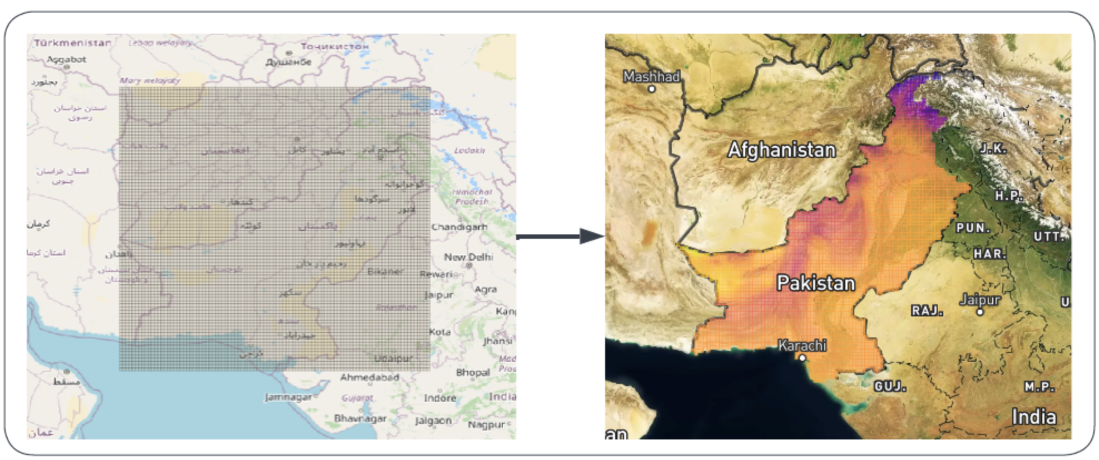


Figure 14. Frontend User Interface

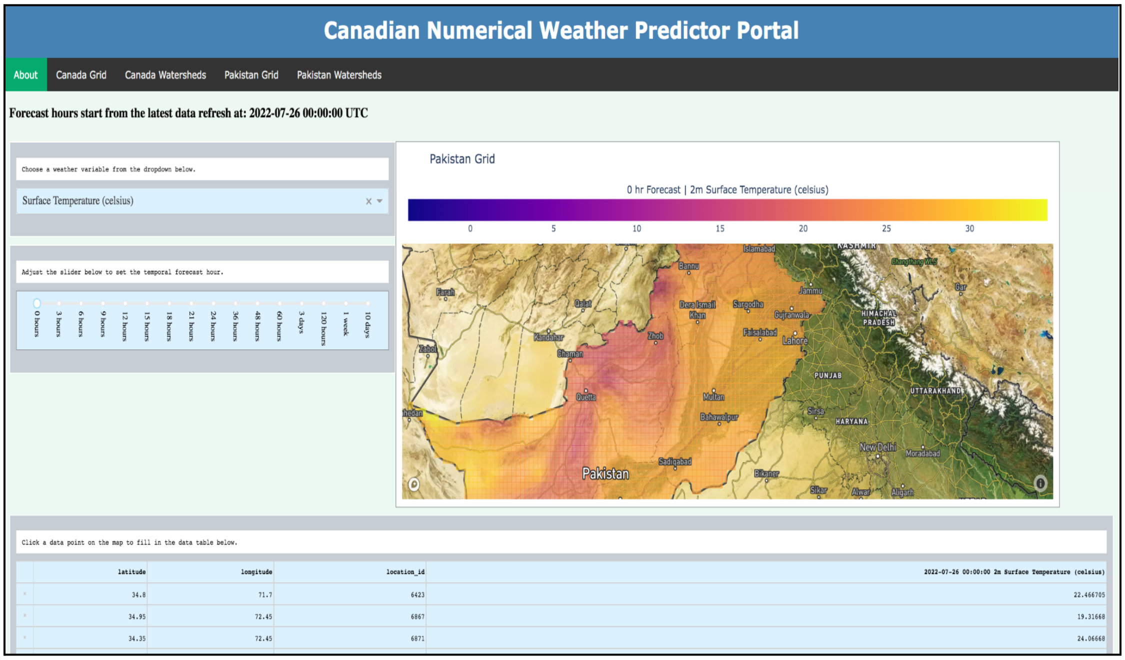


Figure 15. Dash App Callback Diagram