Climate Data and information Cheat Sheet

There is a plethora of climate data available but knowing what scenarios and timeframes to use and where to get the data are needless barriers to entry. This QR code is a link to a climate data cheat sheet that should give a robust start to your climate data journey. Please feel free to disseminate it and reach out with any questions you may have.

do Climate

What data do I need?

Before you look at the data you need you need to establish what you want to achieve.

- Long term change over time – Annual or better absolute or anomaly point or average data over at least 30 years
- Seasonal shift in rainfall – monthly or better point data over 30 years
- Extreme event analysis – daily or better point consecutive data for 30 years with minimal data gaps
- Gradient over a country – spatial data at an averaged time over a 30-year period with a spatial resolution of 10km to 50km depending on the area size.
- Gradient over a province – spatial data at 5km-10km (be careful with data that goes any lower than this)

Analysis Tools

- For spatial analysis free software like QGIS or google earth works well, though they do have some problems using datasets like NetCDF files. Arc is expensive but works well.
- For timeseries data the simplest solutions are to use excel or google sheets, but packages like R and python have excellent climate data libraries.
- Packages such as R and Python that can do both statistical and spatial analysis. Other tools that are more niche you can use MATLAB or NOAA Ferret. These options are likely more advanced than most people may need.

Essential reading

- USAID Climate links provides summaries at a glance of a country's climate context. This is a good starting point when working in a new country and need basic climate information.
- WorldBank Climate Change Knowledge Portal is an excellent resource that provides comprehensive country or provincial level summary data as well as links to risk profiles. There is a lot presented where and it's worth looking around. Data downloads can be done with the hamburger button on each map/graph.

Rapid results

- IPCC interactive atlas provides global standard to get to grips with likely changes in many future variables. The graphs are highly customizable but can be bit over whelming.
- Climate research Unit Google earth interface is a very useful and simple way to get historical monthly data through a familiar interface. The data can be downloaded into excel and used for historical seasonal and long-term trend data.
- Global Drought Monitor is an interactive map that uses the SPEI (Standardised Precipitation-Evapotranspiration Index) to show the severity of observed global drought.

In-depth assessment

- CSAG Climate information Portal is a custom-made climate data portal that seeks to give not only powerful climate data but also more contextual information to the user making it a great place to start researching. The data is a little outdated but is still likely robust.
- Global historical weather and climate data is a great resource to get station data with daily temperature and rainfall available. This data is not always complete so use with caution.
- Climate information Data Access Platform provides an interactive map allowing you to get model outputs at a sub national level for many variables like the standard temperature and rainfall but also aridity, soil moisture, water discharge and runoff. The outputs are highly customisable, but using bias adjusted ensemble will likely yield the best result

Advanced analysis

- WorldClim provides spatial dataset with many rainfall and temperature variables at very high spatial resolution (I would use the 5 minute – 8km data) and has historical and future (2021-2040, 241-2060, 2061-2080, 2081-2100) projected CMIP6 data
- Copernicus CMIP6 climate projections is the raw climate data that can be used with various advanced analysis techniques. This is daily or monthly data from 1850-2300 for all SSP scenario and for all the major variables. This is by no means the easiest data to use but is the raw data from which all the other CMIP6 data is based.

Climate Data and information Cheat Sheet

Climate change, like many other disciplines, has a ton of jargon as assumes a lot of everyone in the discussion. In other settings this is not a problem, but climate change is so wide reaching and urgent that everyone needs to be on the same page with climate change discussions. Hopefully, this guide this serves as an insight into these essential discussions.

Climate risk assessment use indexes for Hazard (which estimates the change in the climate such as more frequent drought), Exposure (this is the presences of assets that may be impacted such as farms or livestock density), Vulnerability (covers factors that may make one area more susceptible to the impacts than another such as reliance on rainfall vs reticulation for farming, diversity of livelihoods mitigating impact).

Rainfall modelling still has uncertainty, particularly for convective (summer afternoon storm) rainfall as these systems tend to be smaller and are sometimes poorly simulated in models or models rely on parameterisation (using formulas at very small scale rather than using weather processes) to present these events.

Downscaling uses known climatic forcings like mountains, ocean, and/or landcover to simulate the climate within the climate model grid cell. For instance, Durban and Pietermaritzburg are only ~60km apart and may therefore fall within the same climate grid cell and have a single projected climate value. Downscaling will consider the fact the Durban is next to the ocean and has a low elevation while Pietermaritzburg is at ~800m elevation to calculate more applicable climate values. This is standard, but there is a limitation to this and models that are claiming downscaling down to 1km resolution or less should be used with caution.

Resolution is either the spatial scale (number of km² in box forming a grid cell) or the temporal scale (the timestep, 1 hour, 1 day, 1 month, 1 year, decadal) of the climate model domain (the spatial area covered by the model). The resolution of a model is based on the requirements of assessment. Flood modelling will require 1 hour or better temporal resolution at ~1km or better spatial resolution rainfall model over a focused domain like the city of Johannesburg. Drought modelling will require monthly assessment at 50km resolution over Southern Africa for example.

RCP and SSP are very similar and represent the old CMIP5 (RCP: Representative Concentration Pathway) and new CMIP6 (RCP: Representative Concentration Pathway). RCP: Representative Concentration Pathway while SSPs provide the socioeconomic context for RCPs. Current climate funds are suggesting that work use RCP4.5 or SSP2-45 as their base.

Ensembles are a clustering of models to better represent a climate. Each model has its own strength and weaknesses and will never perfectly simulate the climate perfectly. Using an ensemble of models will give a more robust assessment. The selection of models forming an ensemble can be determined through validation.

Anomalies are a departure from the expected or typical climate conditions in a particular region or over a specific period. Climate anomalies can manifest as deviations from long-term averages, unusual weather patterns, extreme conditions, or shifts in climate-related phenomena.

Calibration of climate data is testing for and adjusting the model data based on known station data. This is essential for all assessments as models sometimes are incorrect on the magnitude of climate variables. This is also part of climate model validation which tests the accuracy of climate models against station data in terms of seasonal magnitudes, long term trends, etc.

Climate window is 30 years because years on year variability is a natural occurrence but once a climate has been occurring for 30 years, it's considered the normal. Historical climate window is currently 1990-2020 and projected is 2020-2050.

Data quality often refers to where data is limited, has errors, or is generally incomplete. This can lead to inadequate analysis. To combat this, look for strange anomalies in timeseries data or abrupt jumps in spatial data. This can result from changing measuring units or reading errors. Where possible, exclude the obviously wrong data and use an ensemble of data to make up for weaknesses in any one dataset.

Reanalysis data is historical modelled data that has some station or observation data input into the model to force the climate to match that area. Areas that don't have station data are then simulated such that the areas with data match between the simulation and the data. This us very useful in data poor areas but does come with some uncertainty.

Remote sensed data is historical data that was derived or estimated from indirect measurements such as satellites. This is tested against known station measurements to calibrate. This is also very useful in data poor areas