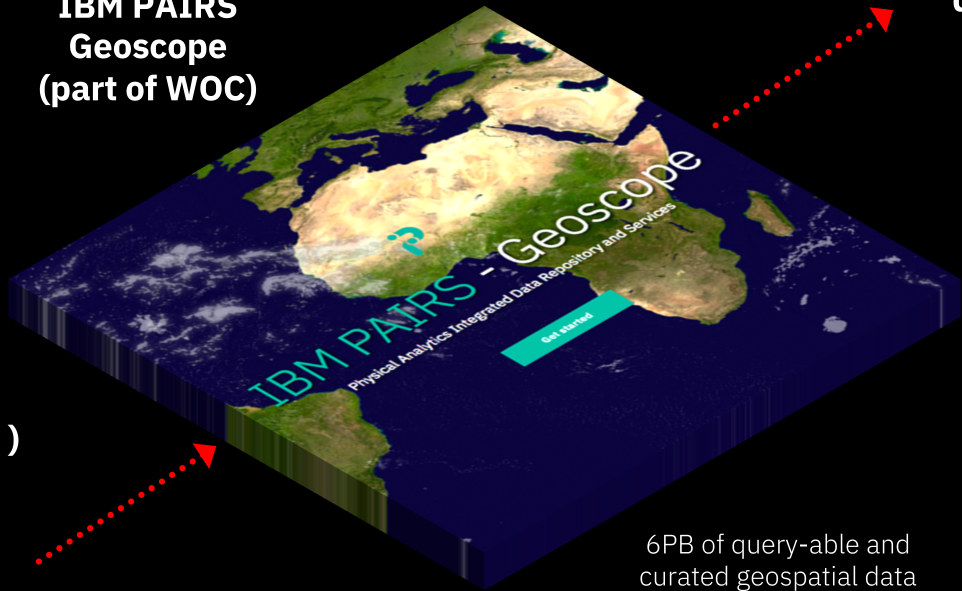
Call for Code Spot Challenge for Wildfires

IBM

Call for Code Spot Challenge for Wildfires data sets have been “extracted” from PAIRS

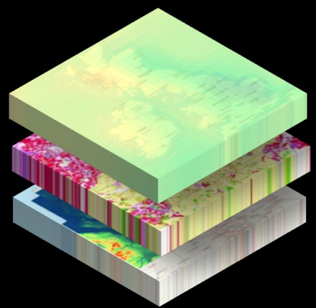
IBM PAIRS Geoscope (part of WOC)

Extract for data science contest



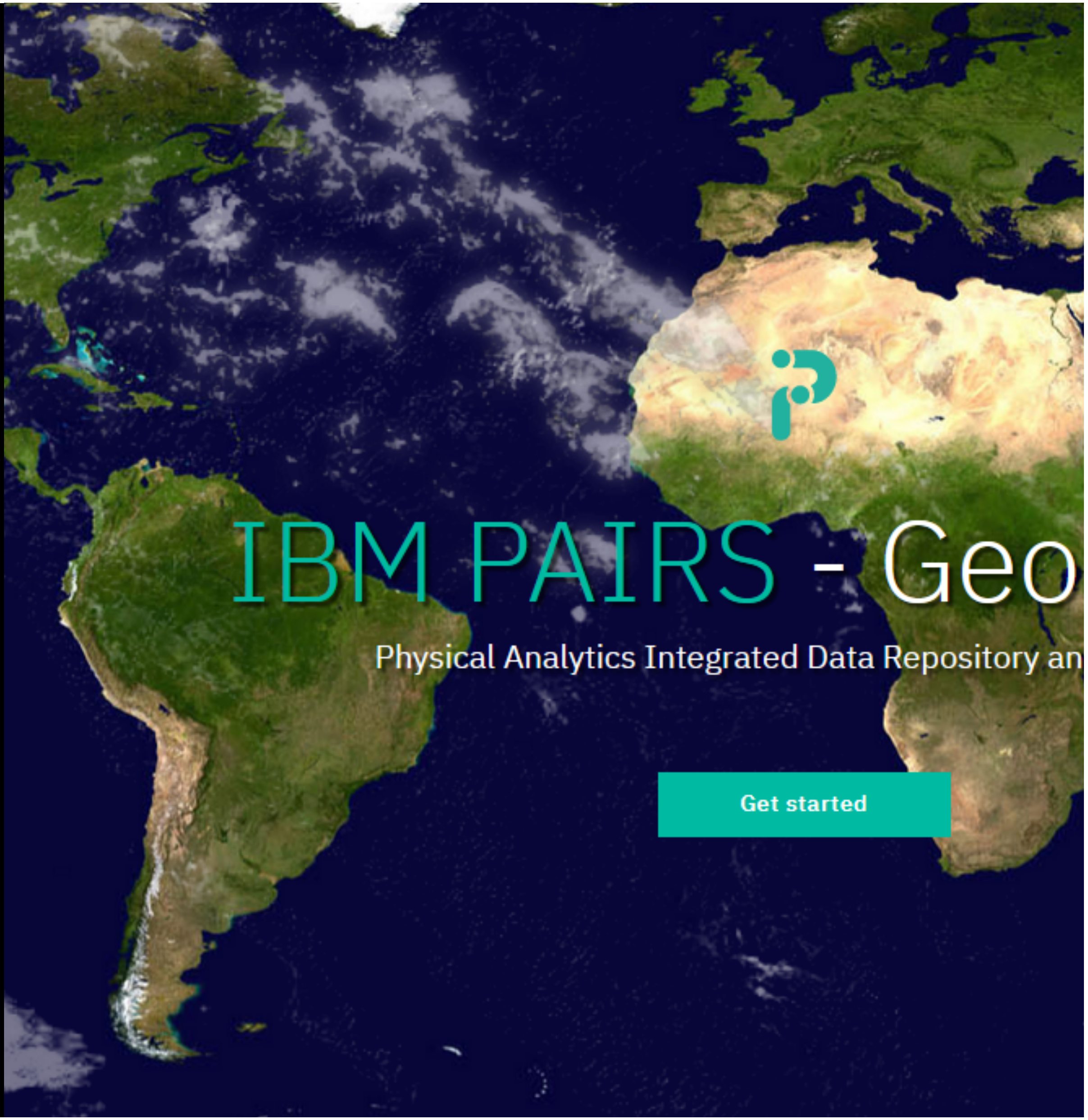
6PB of query-able and curated geospatial data

Raw data (NASA, Copernicus, NOAA etc.)

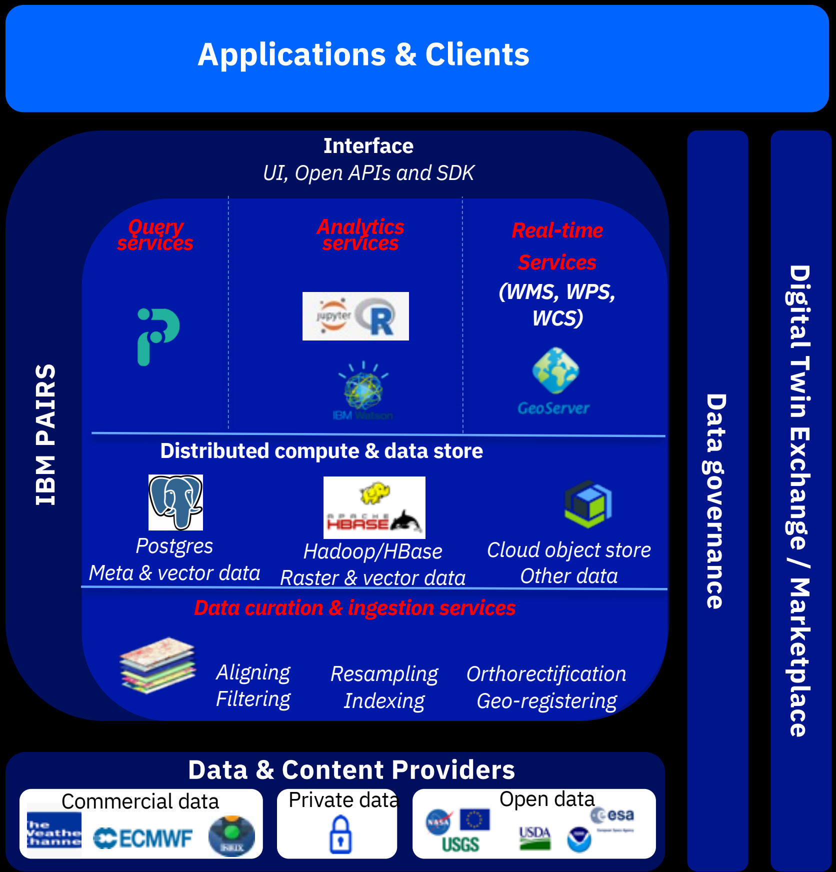


What is PAIRS Geoscope?

A big geospatial-temporal data and analytics platform for rapid development and deployment of solutions



PAIRS Geoscope – Overview



- ✓ Exposed through APIs, SDK and UI as a service
- ✓ Data governance module (manages T&Cs, provenance etc.)
- ✓ Data marketplace via Digital Twin Exchange
- ✓ Scalable geospatial-temporal data analytics platform
- ✓ Batch mode (M/R allows pushing analytics such as UDFs, decision trees into the query)
- ✓ Real-time mode for direct access
- ✓ Set of “starter” models for data scientist and developers
- ✓ Distributed system versus object store or relational databases
- ✓ PetaBytes of data ready to go
- ✓ All data linked in space and time
- ✓ Ingests, curates 10TB+/day of new data
- ✓ Manages data pipelines with major content providers
- ✓ User-enabled data curation/uploads

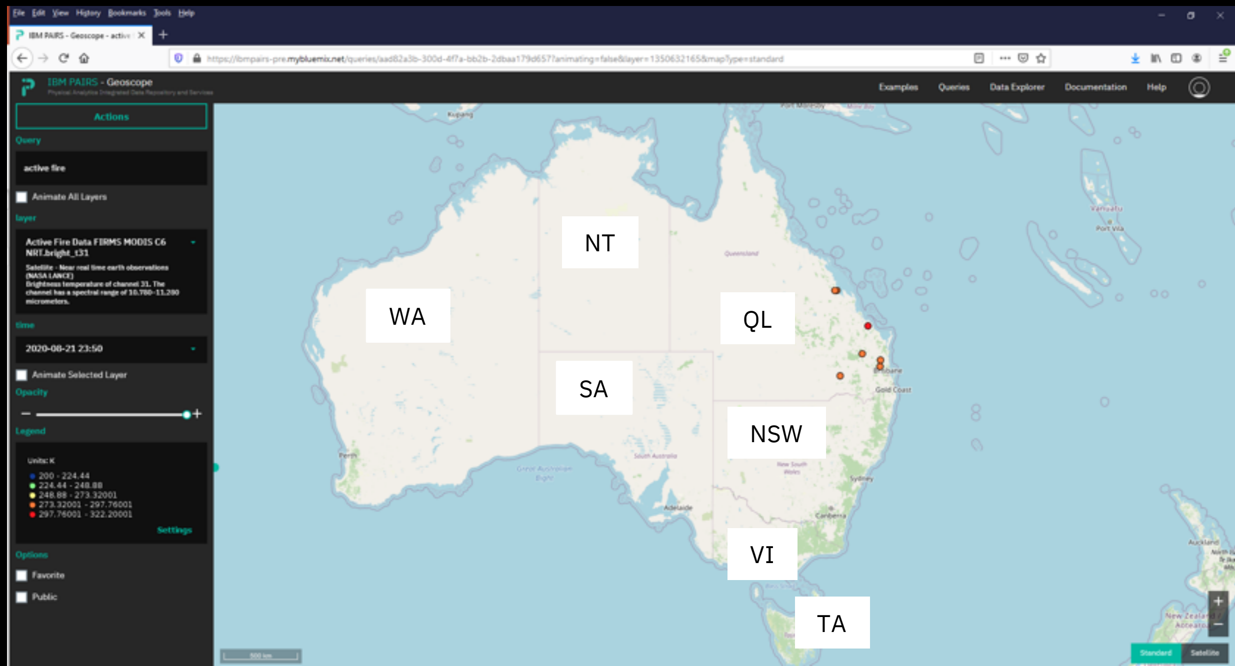
https://ibmpairs.mybluemix.net/

What will be forecasted/predicted in this contest?

Forecast/predict the daily total estimated fire area for 7 states in Australia for Feb. 2021?

NSW=New South Wales*
NT=Northern Territory
QL=Queensland
SA=Southern Australia
TA=Tasmania
VI=Victoria
WA=Western Australia

*excluded the capital region



Why is it important to forecast wildfires?

1. To prepare and respond
2. To understand the root causes
3. To help to mitigate them in the future

What will be submitted?

Date	Region	Total estimated fire area [km^2]
2/1/2021	NSW	15
2/2/2021	NSW	327
…	…	…
2/28/2021	NSW	234
…	…	…
2/1/2021	WA	50
2/2/2021	WA	800
…	…	…
2/28/2021	WA	785

7x28 rows = 196 rows

The data for this competition

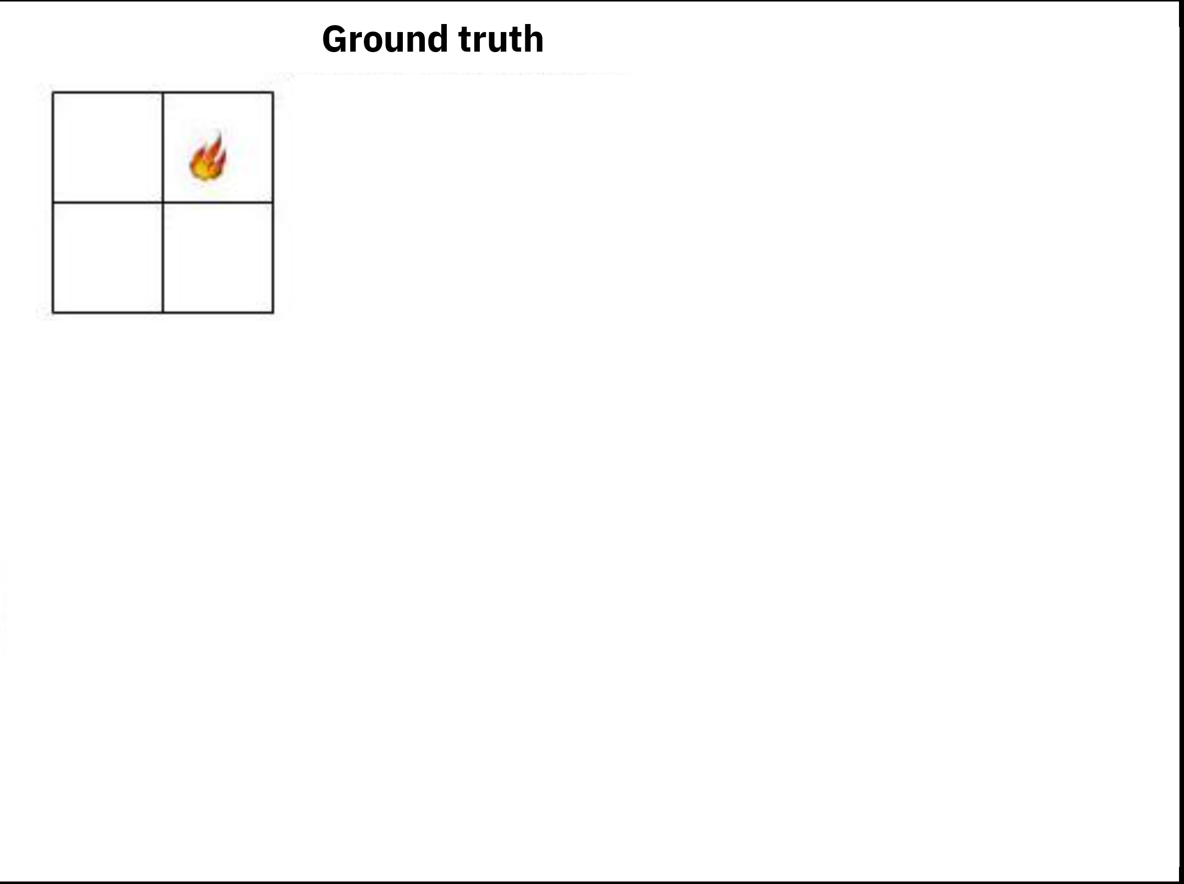
1. Historical Wildfires
2. Historical Weather
3. Historical Weather Forecast
4. Land Class
5. Normalized Vegetation Index

- Readme_Docs.docx V2
- HistorialWeatherForecasts.csv V2
- HistoricalWeather.csv V2
- VegetationIndex.csv
- LandClass.csv V2
- Historial_Wildfires.csv

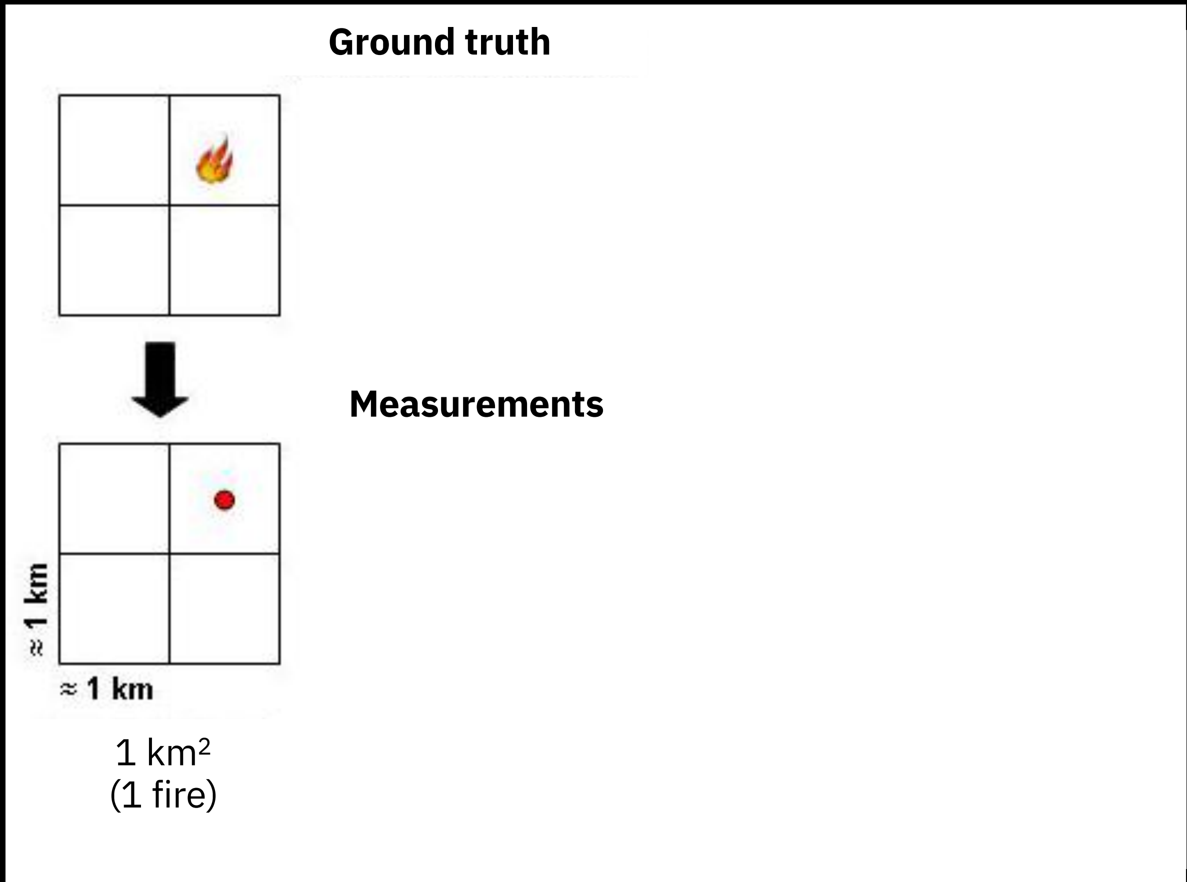
Note that there are many other data sets relevant to wildfire predictions

Historical fire data from a NASA satellite

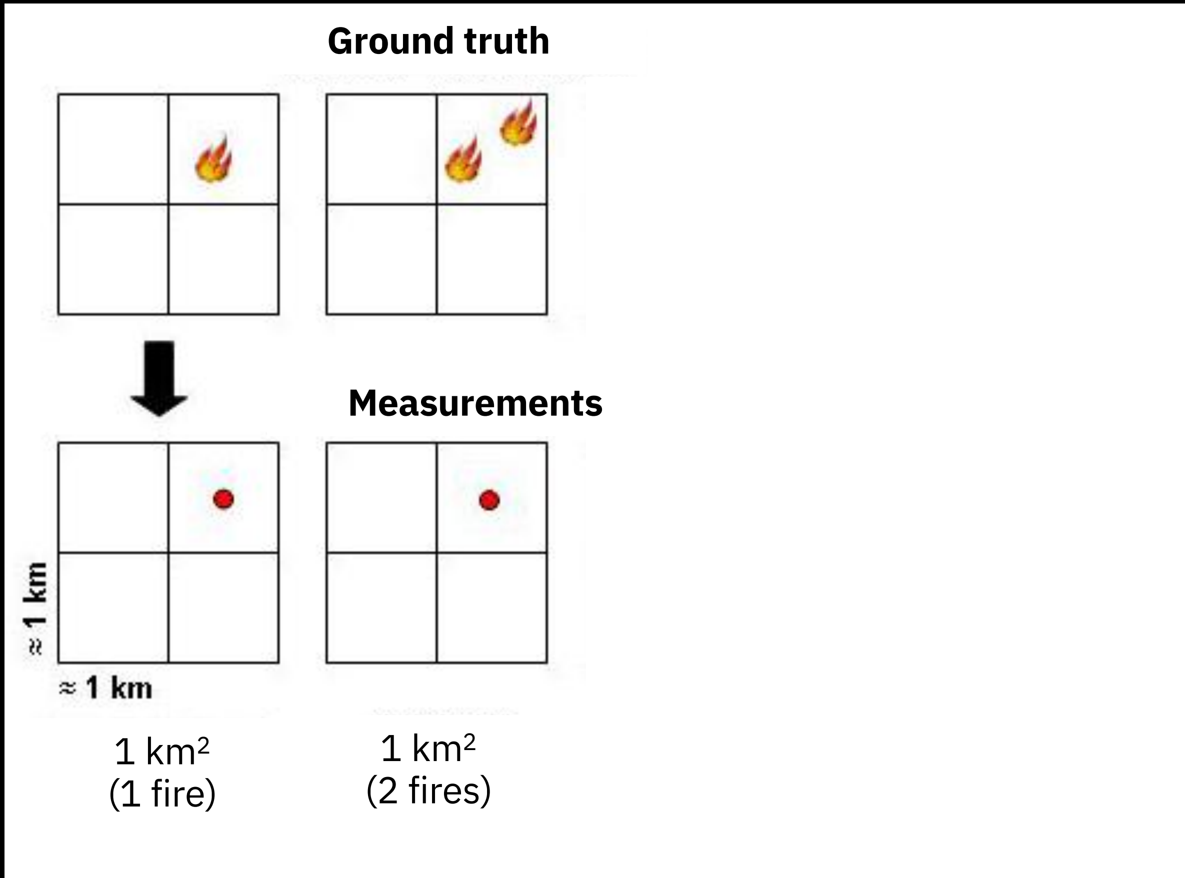
Historical fire data – from hotspots to pixels to estimated areas



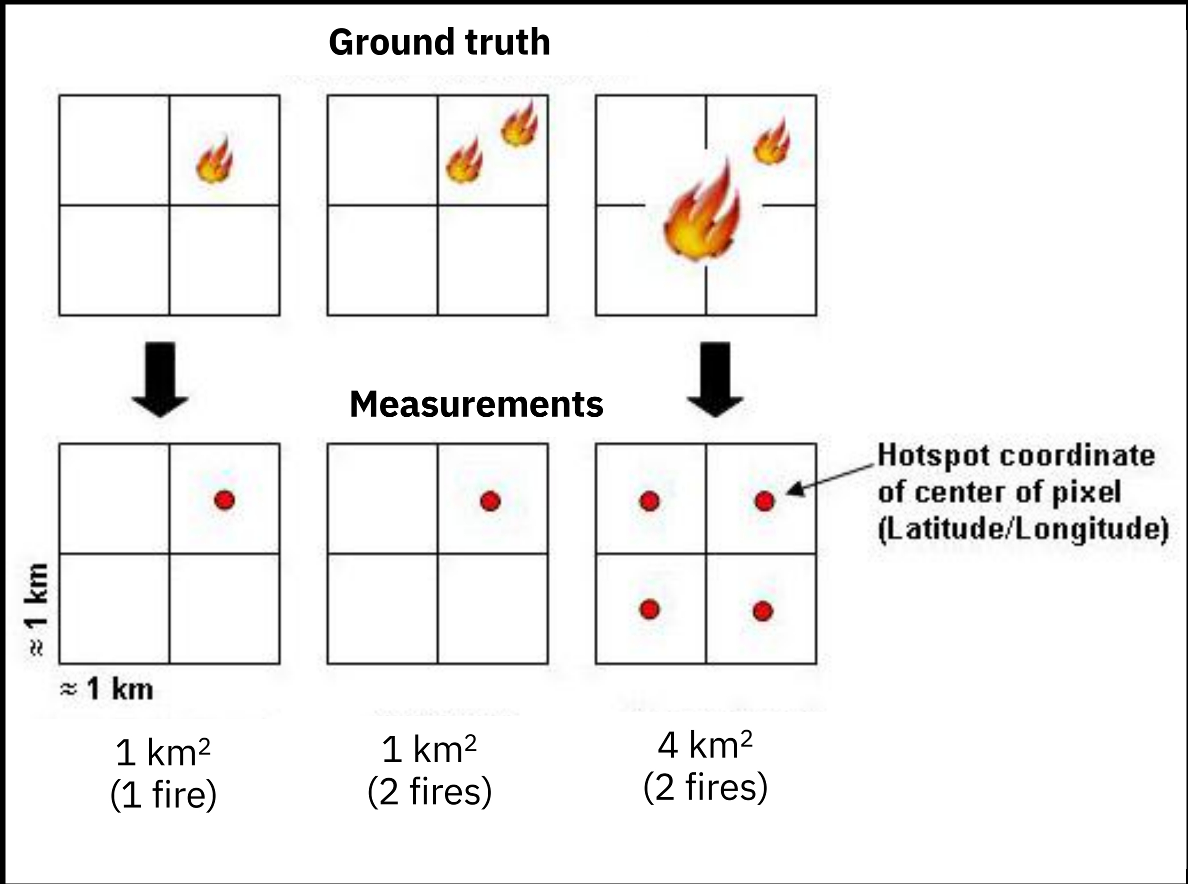
Historical fire data – from hotspots to pixels to estimated areas



Historical fire data – from hotspots to pixels to estimated areas

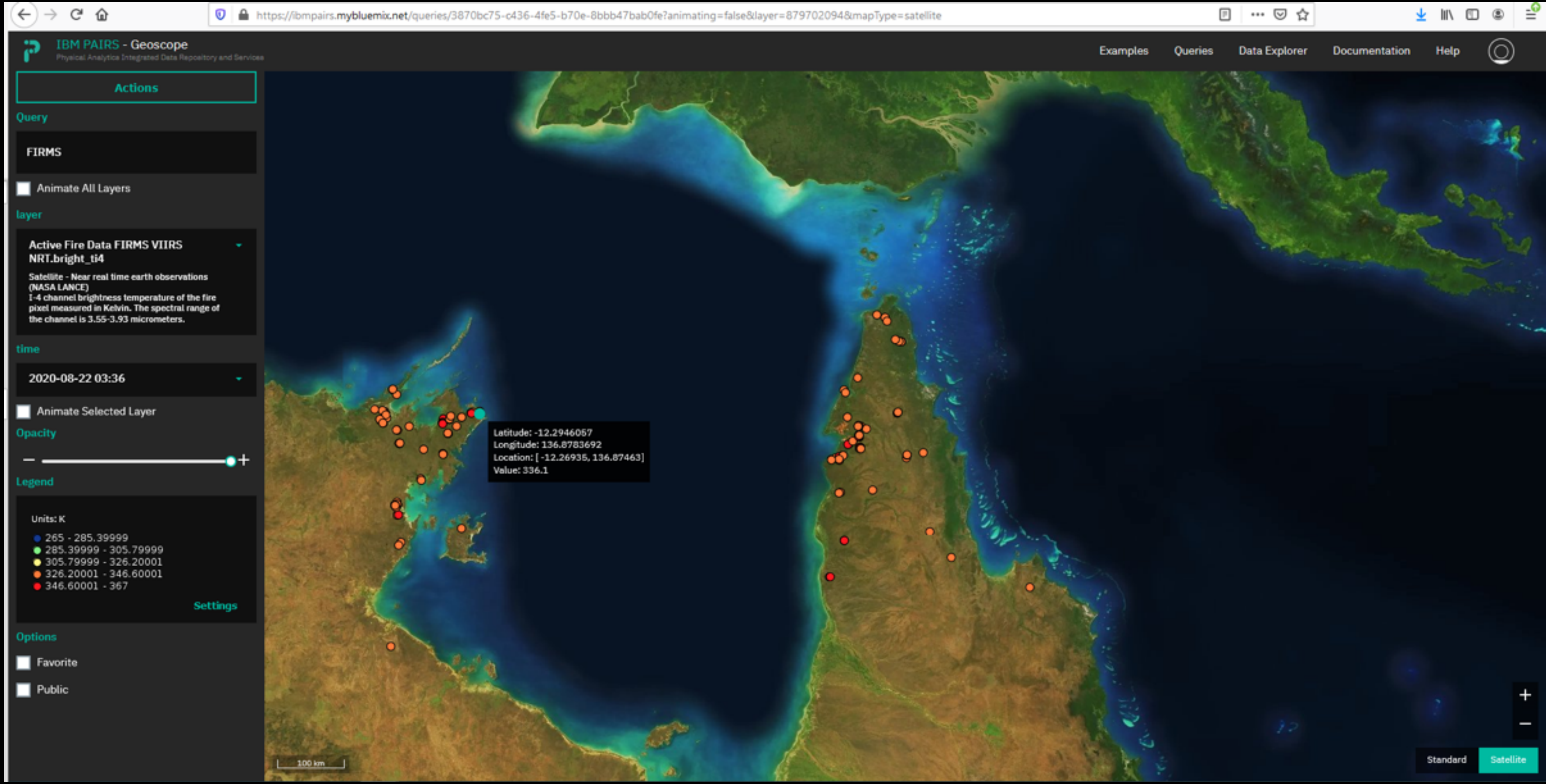


Historical fire data – from hotspots to pixels to estimated areas

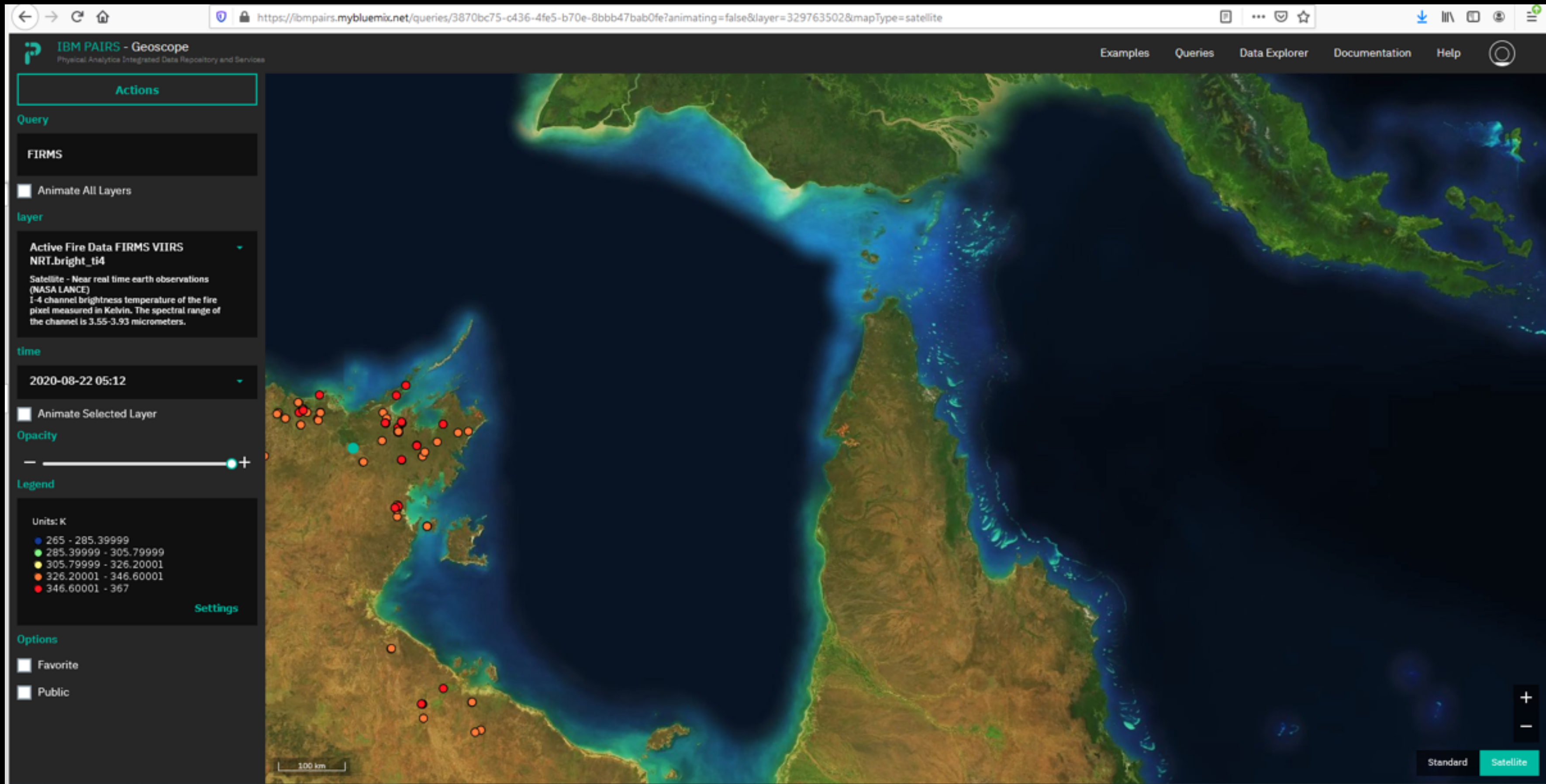


A hotspot plotted using the MODIS thermal anomalies algorithm represents the center of an approximately one-square-kilometer pixel flagged as containing one or more thermal anomalies, which may indicate a fire (upper half of image). The hotspot "location" is the center point of the pixel, which is an approximation of the actual thermal anomaly (lower half of image). Illustration courtesy of NASA FIRMS.

Historical fire data – Raw data



Historical fire data – Raw data



Historical fire data – PAIRS processing

1. Daily averaged
2. Spatially aggregated over 7 regions
3. Confidence of >75%
4. Inferred hotspot type = 0 meaning a presumed vegetation fire
5. Area estimated by multiplying the along scan pixel size by the along track pixel size.
6. Brightness estimated by averaging the means of both the brightness temperature 21 (obtained from channel 21/22) and brightness temperature 31 (obtained from channel 31).

Historical fire data – Columns

1. Region
2. Date
3. Estimated_fire_area [km^2]
4. Mean_estimated_brightness [K]
5. Mean_estimated_fire_radiative_power [MW]
6. Mean_confidence [%]
7. Std_confidence [%]
8. Var_confidence [%]
9. Count
10. Replaced [Y/N]

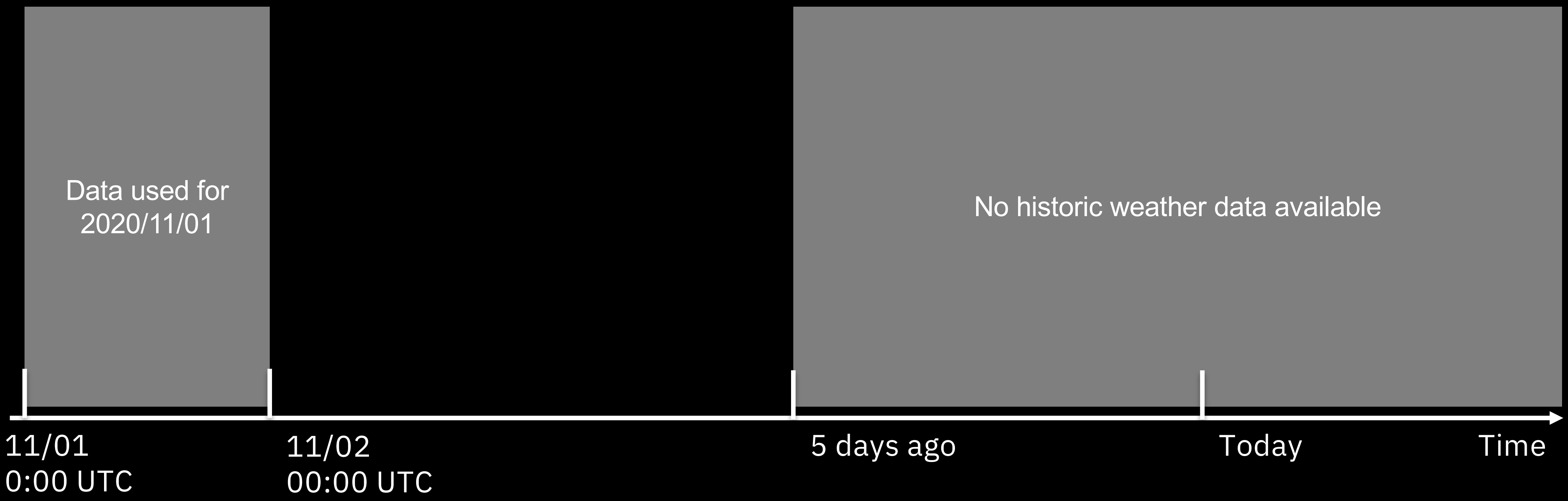
Historical weather data – Overview

ERA5 climate reanalysis data loaded from and processed with IBM PAIRS

Parameter	Unit
Precipitation	mm/day
Relative humidity	%
Soil water content (0-7 cm)	$mm^3\ mm^{-3}$
Solar radiation	MJ/day
Temperature	C
Wind speed	m/s

Spatial aggregates of daily values for each region.

Historical weather data – Time



Historical weather data – Daily aggregates

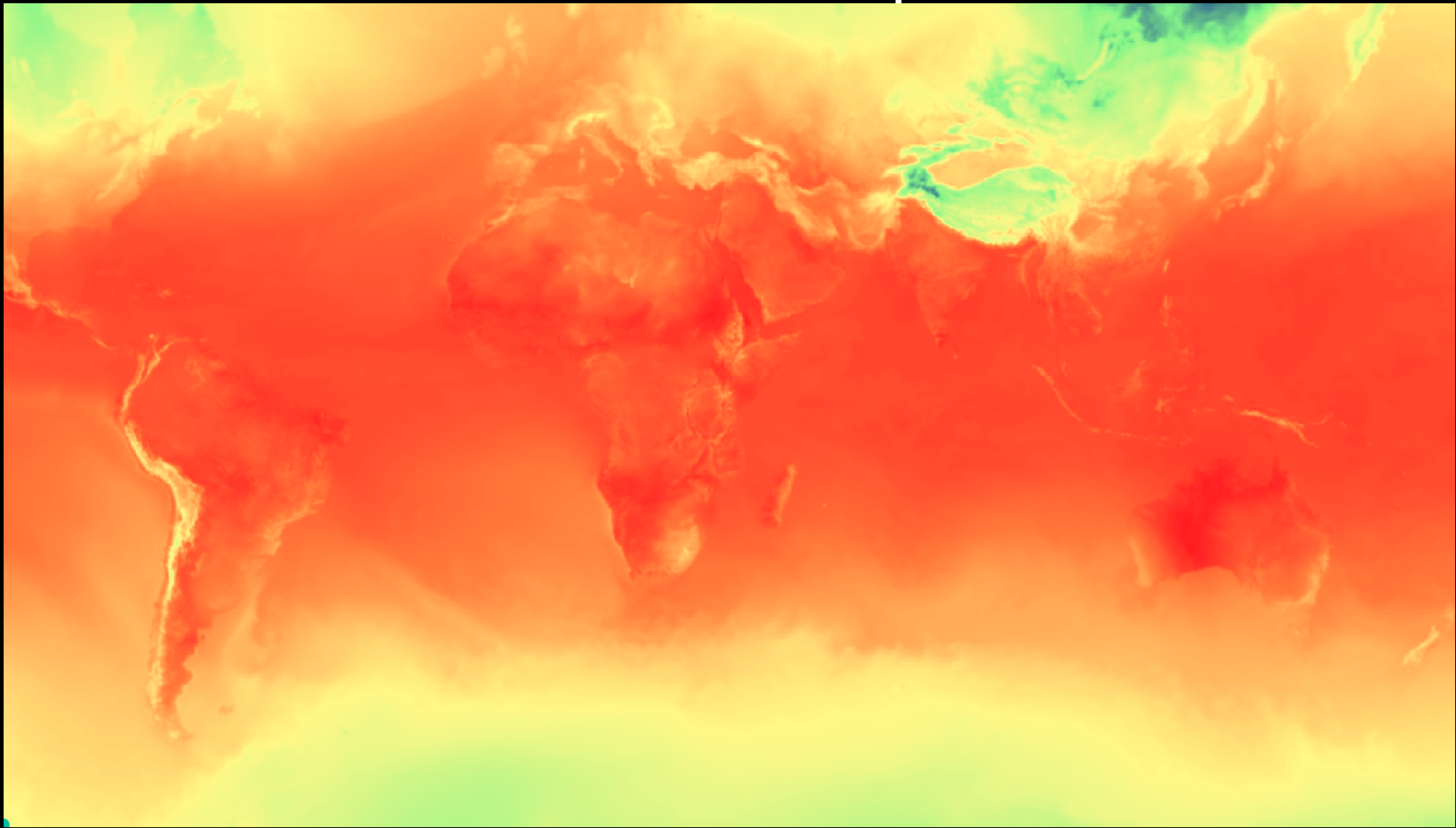
Parameter	**Aggregate**
Precipitation	Sum
Relative humidity	Mean
Soil water content (0-7 cm)	Mean
Solar radiation	Sum
Temperature	Mean
Wind speed	Mean

Daily aggregates from hourly values.

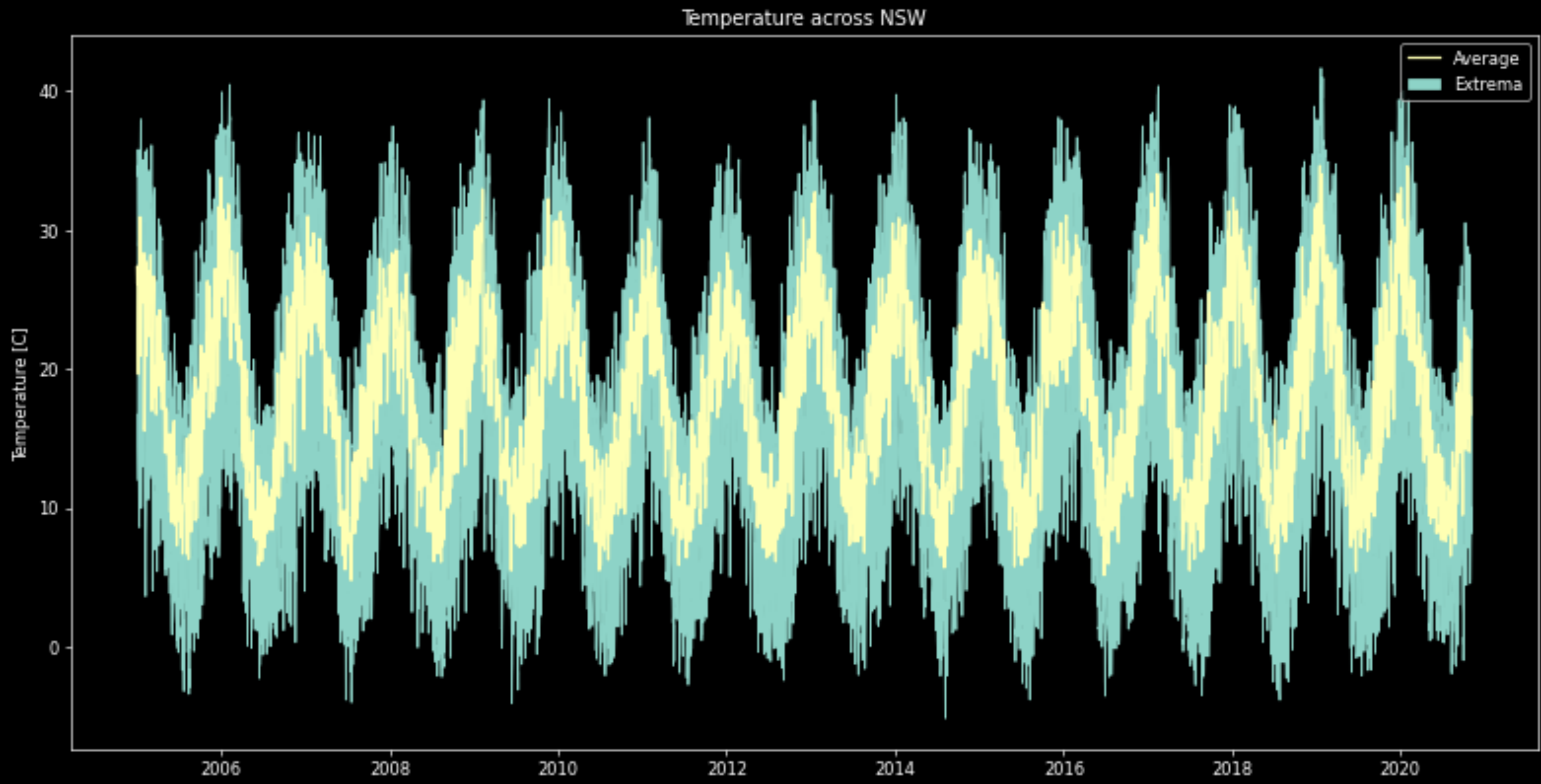
Historical weather data – Data frame

1. Region
2. Date
3. Parameter
4. count() [km2]
5. min(): Minimum value of the spatial aggregation.
6. max(): Maximum value of the spatial aggregation.
7. mean(): Average of the spatial aggregation.
8. variance: 2nd moment of the spatial aggregation.

Historical weather data – Temperature

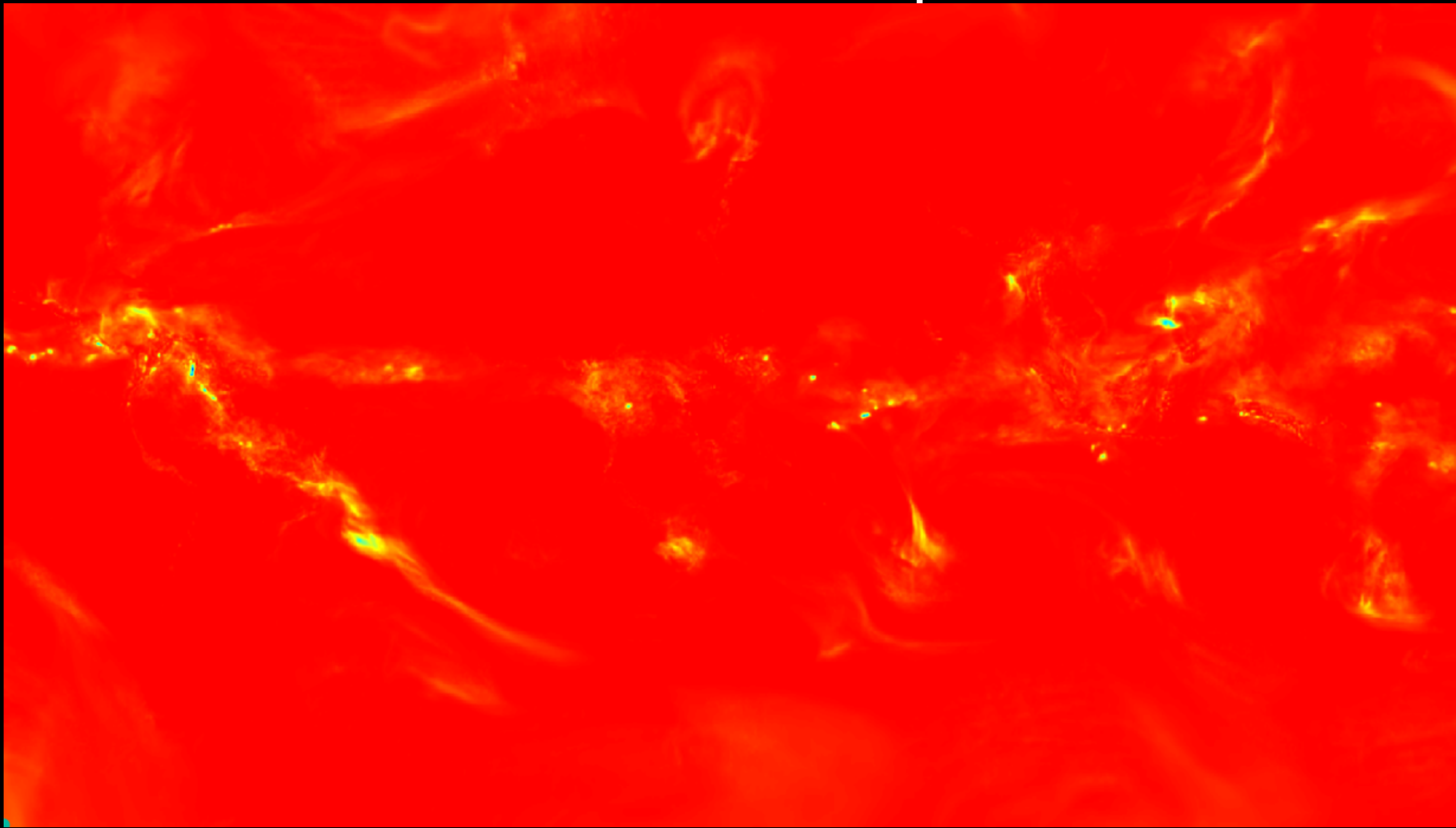


Historical weather data – Temperature

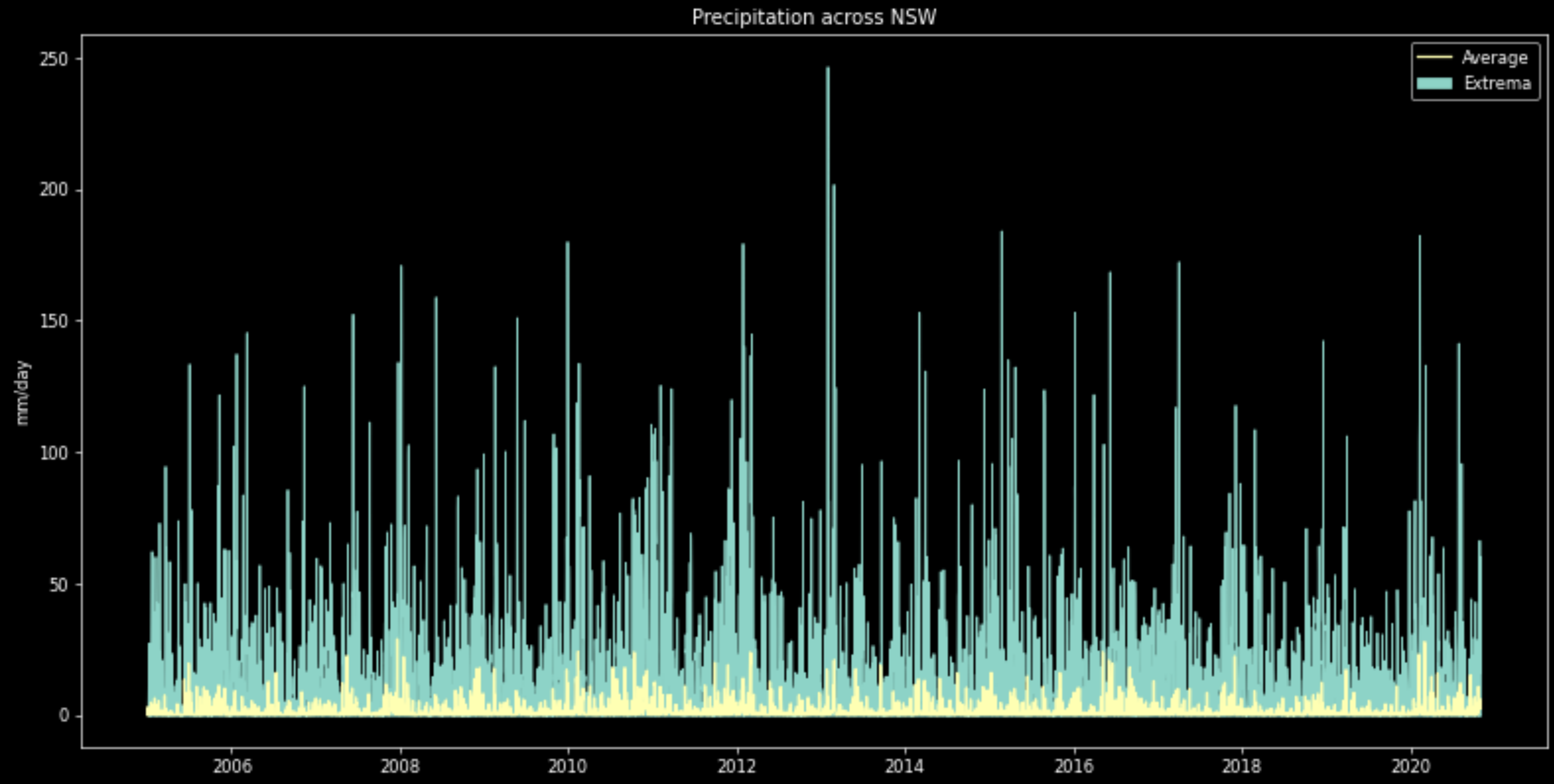


IBM

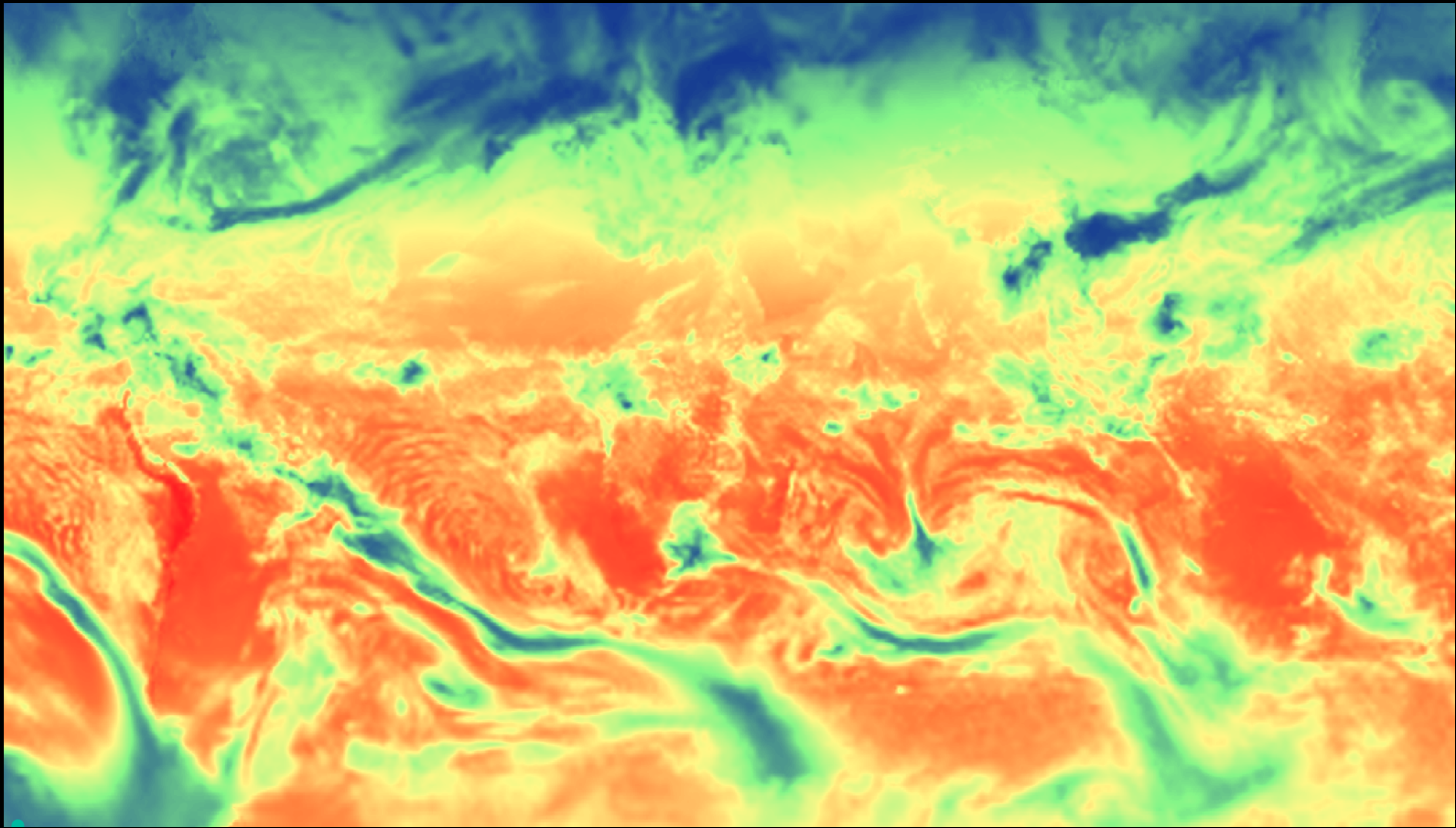
Historical weather data – Precipitation



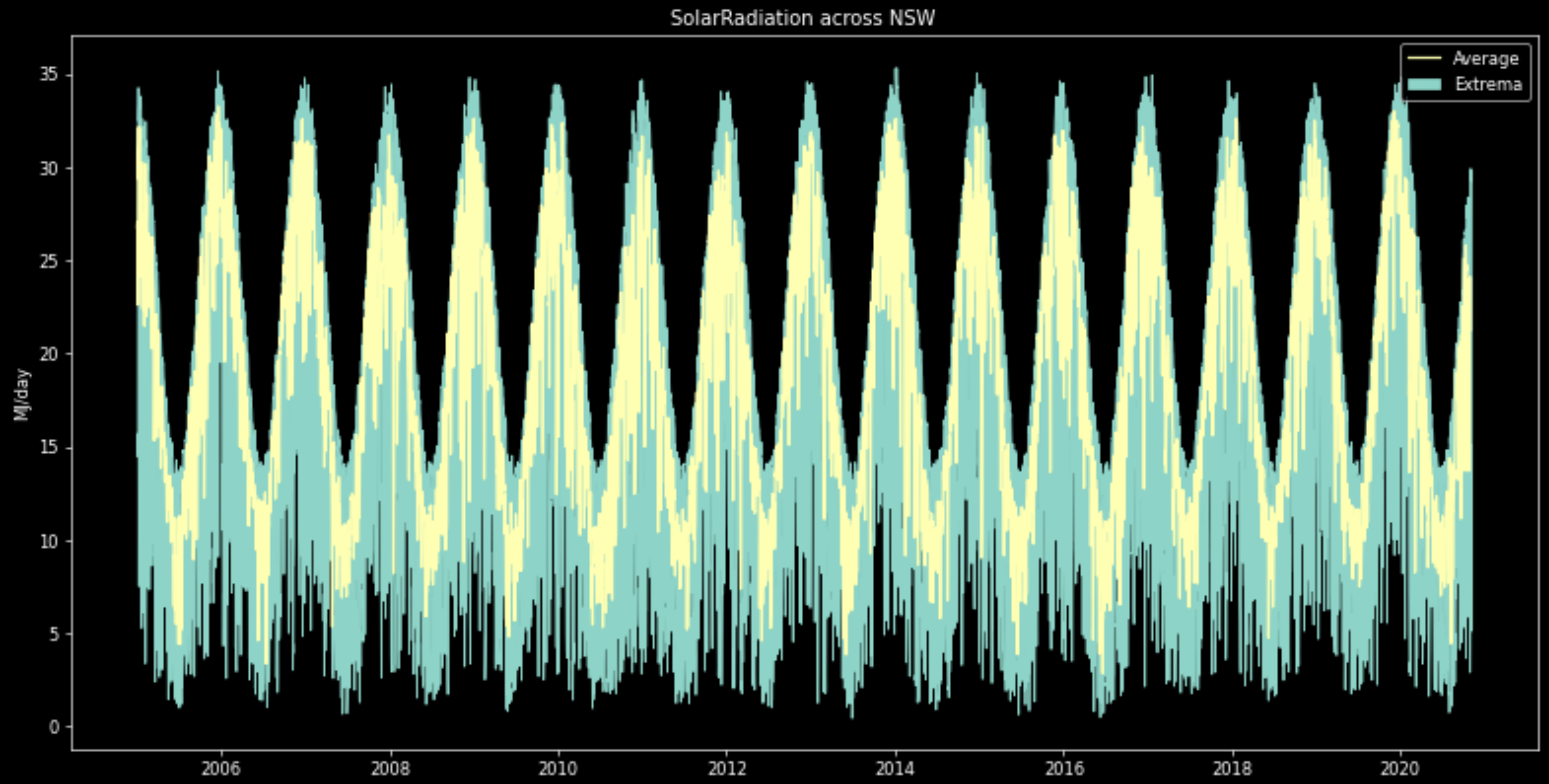
Historical weather data – Daily precipitation



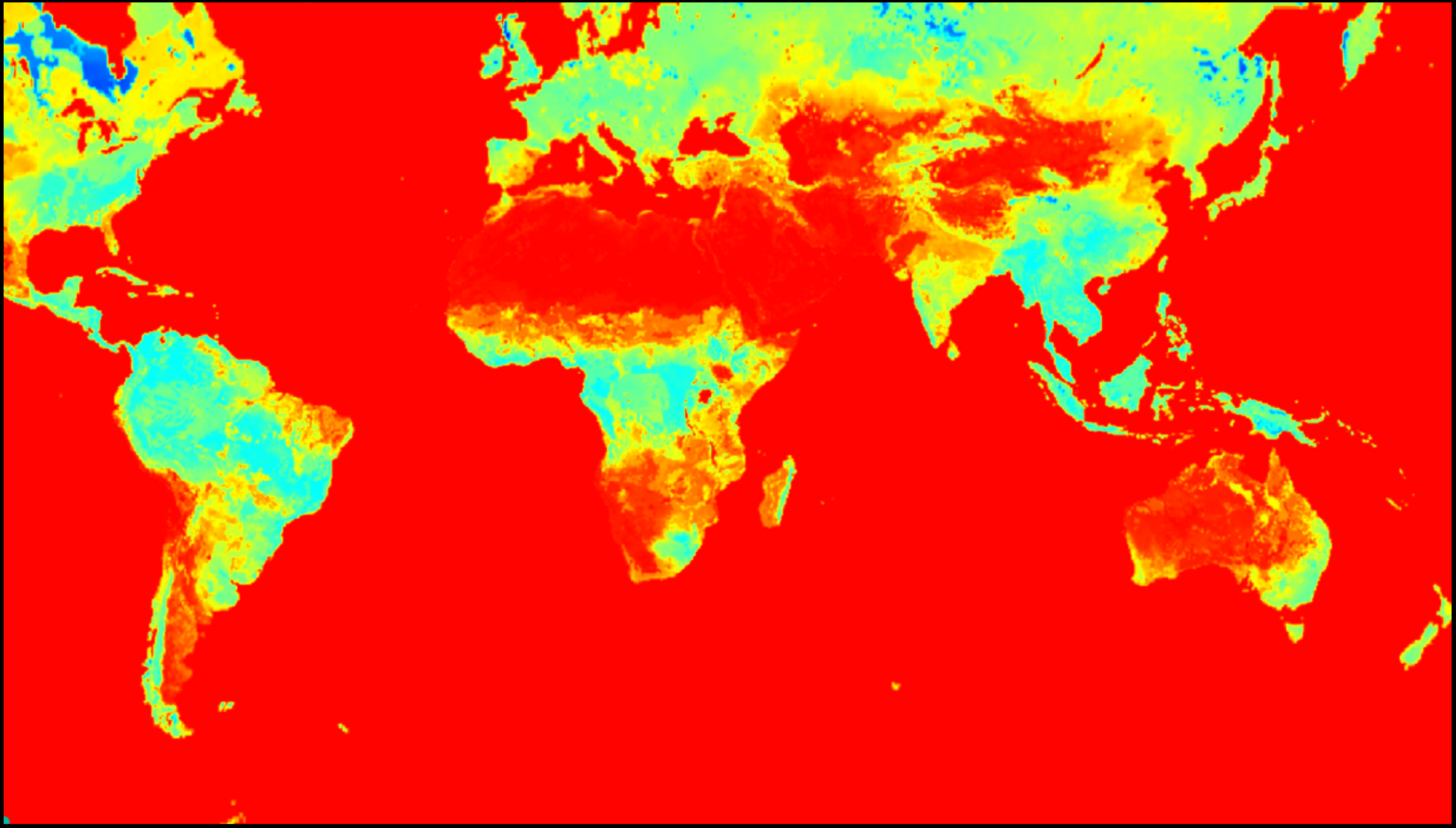
Historical weather data – Solar radiation



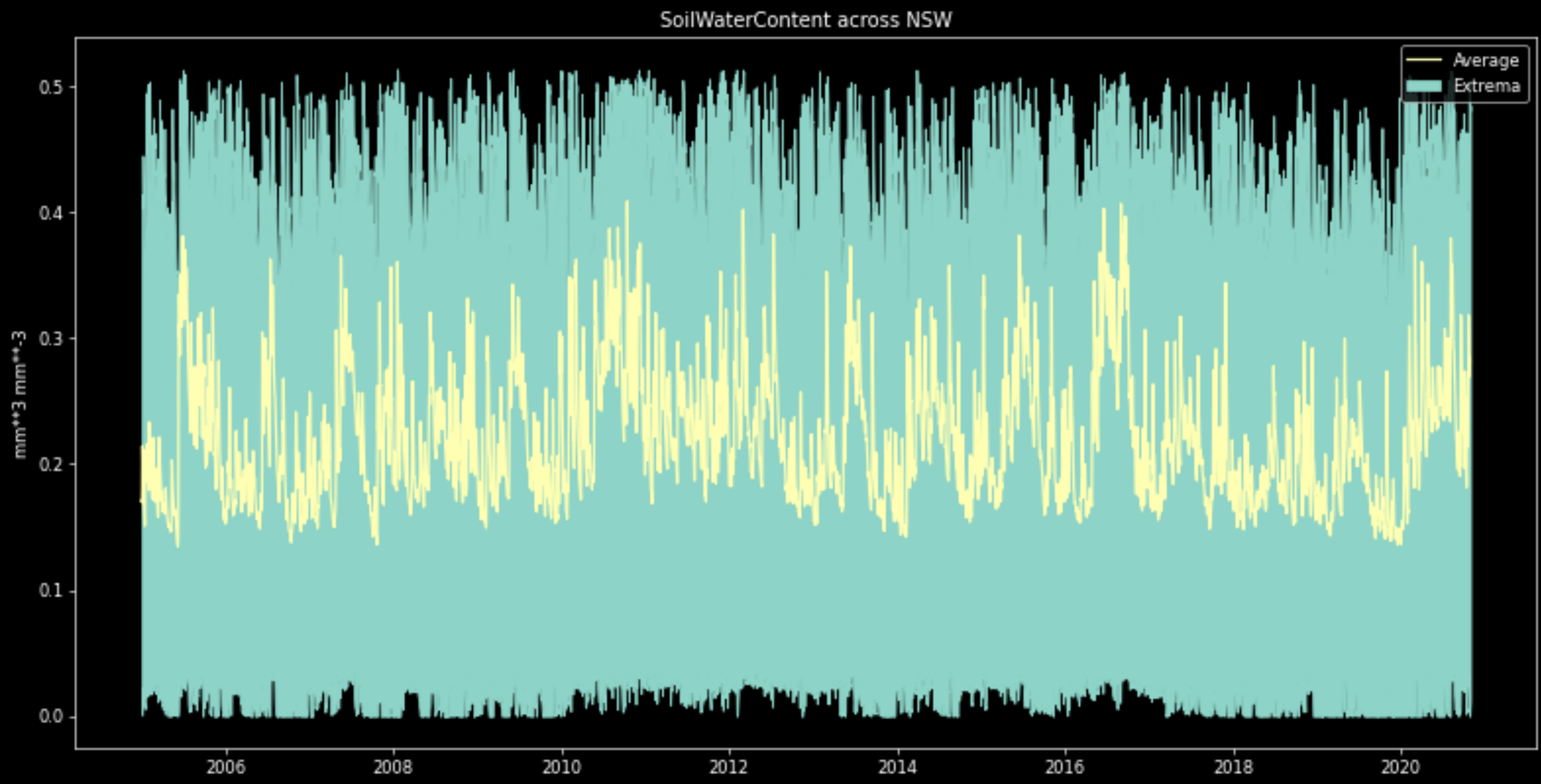
Historical weather data – Solar radiation



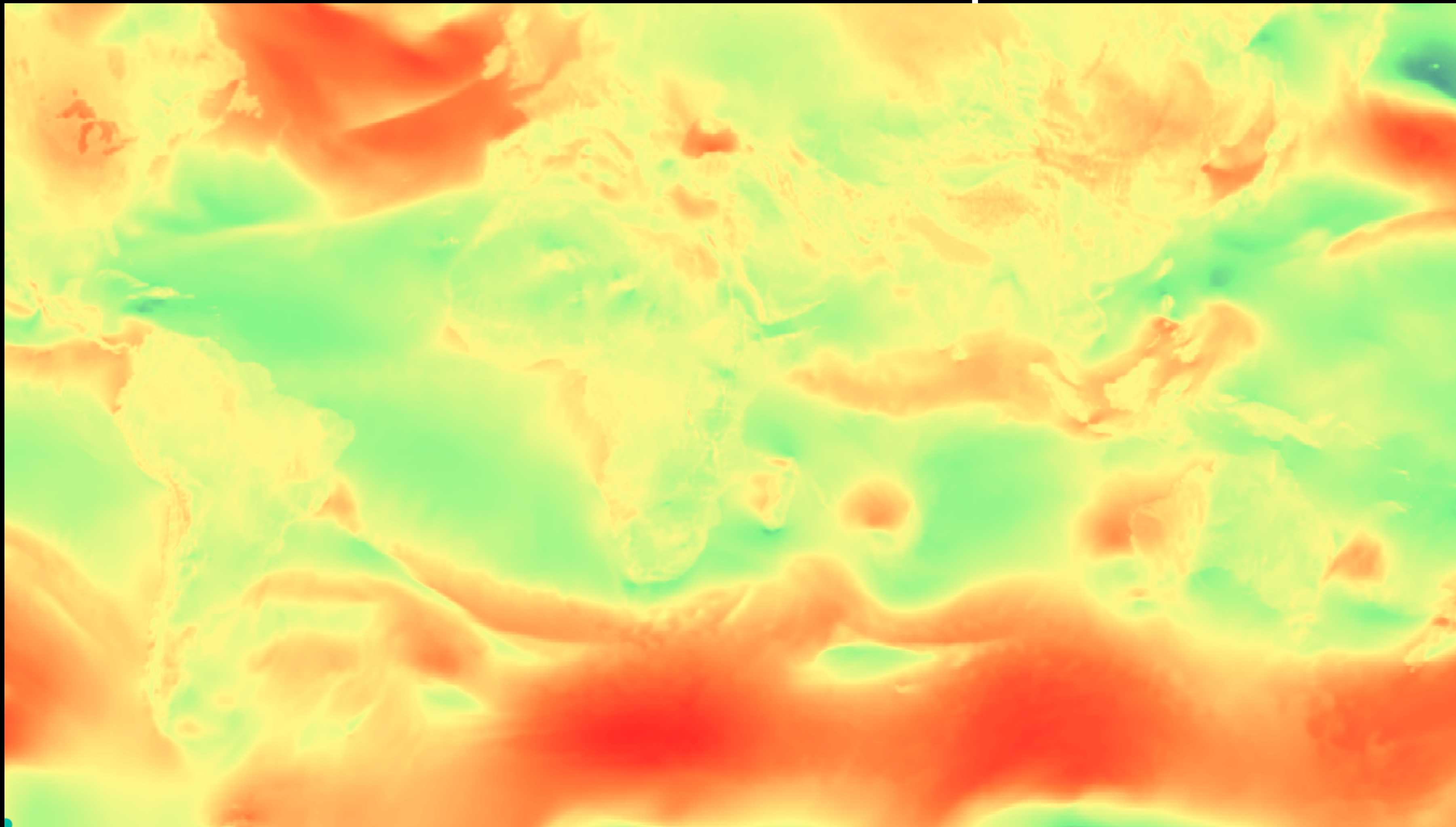
Historical weather data – Soil water content



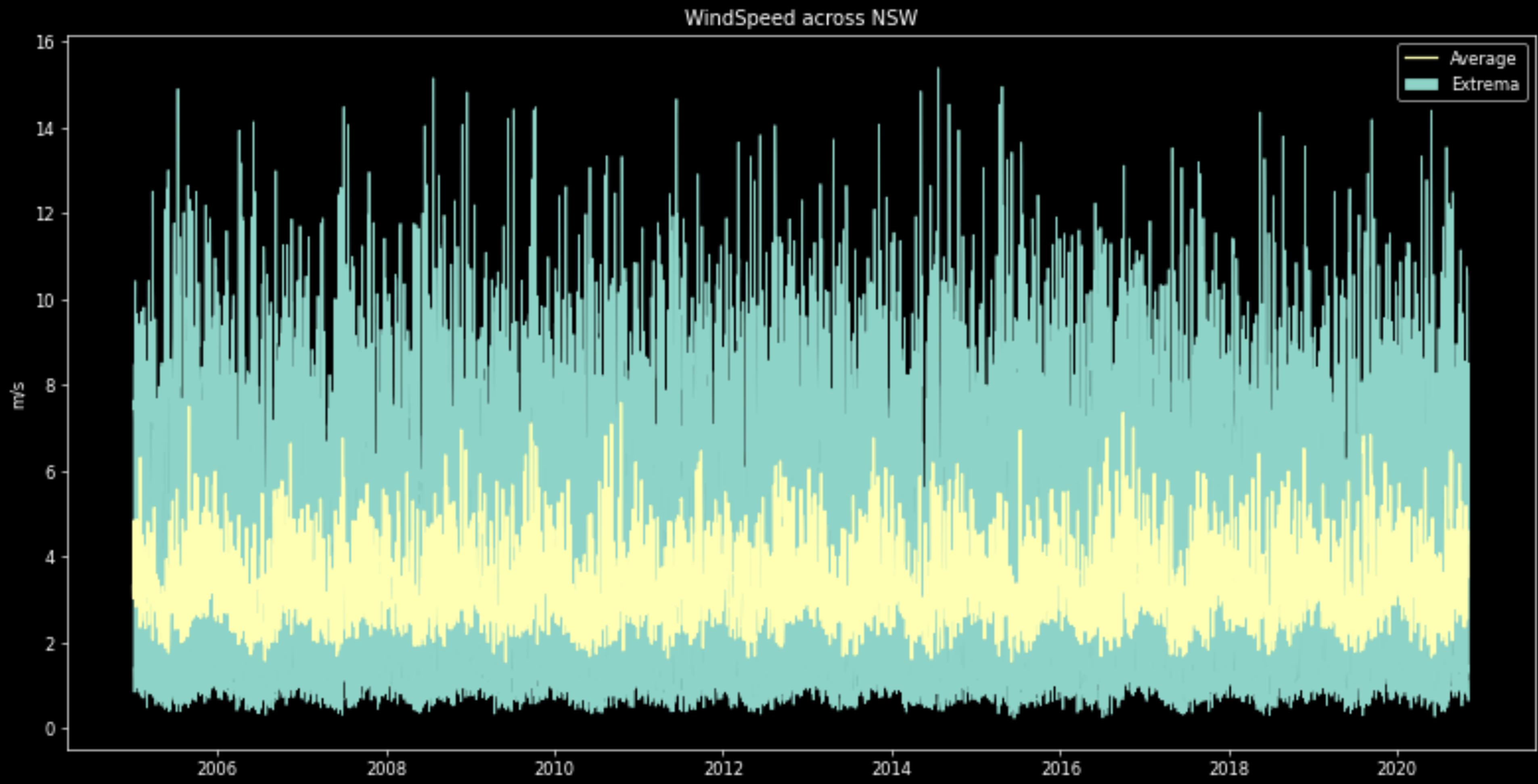
Historical weather data – Soil water content



Historical weather data – Wind speed



Historical weather data – Wind speed



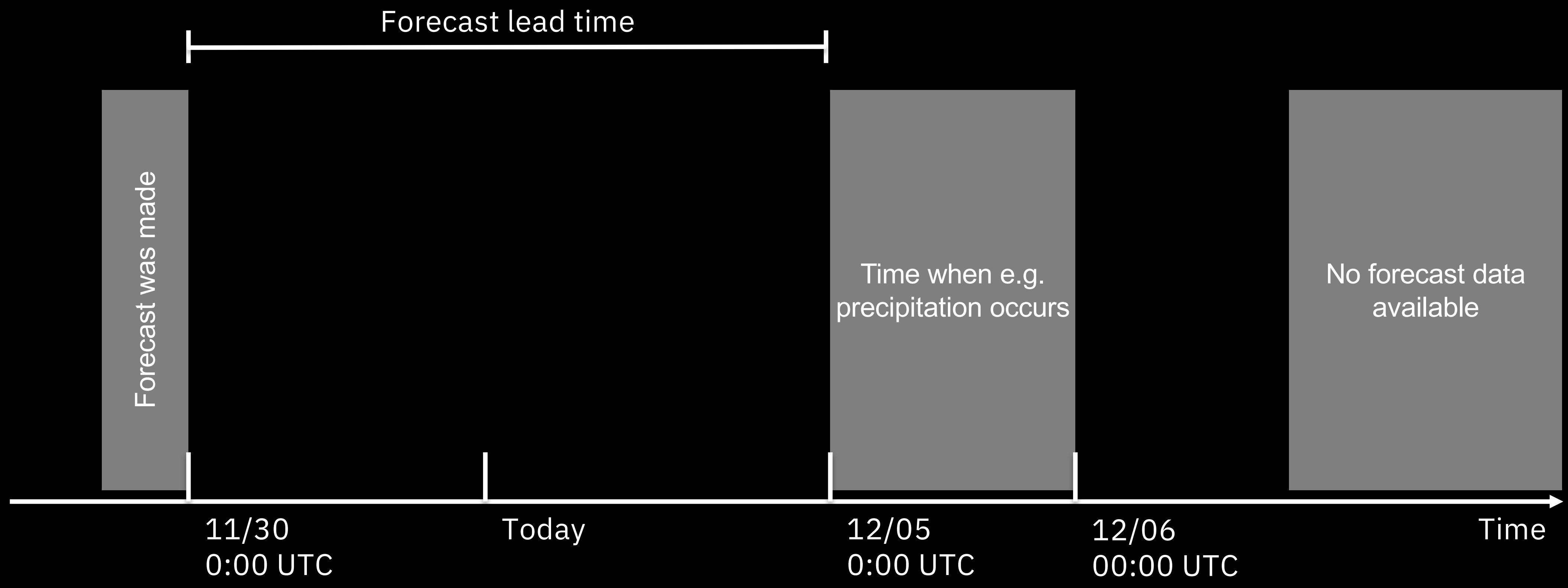
Historical weather forecasts – Overview

GFS weather forecasts loaded from and processed with IBM PAIRS

Parameter	**Unit**
Precipitation	mm/day
Relative humidity	%
Solar radiation	MJ/day
Temperature	C
Wind speed	m/s

Spatial aggregates of daily values for each region.

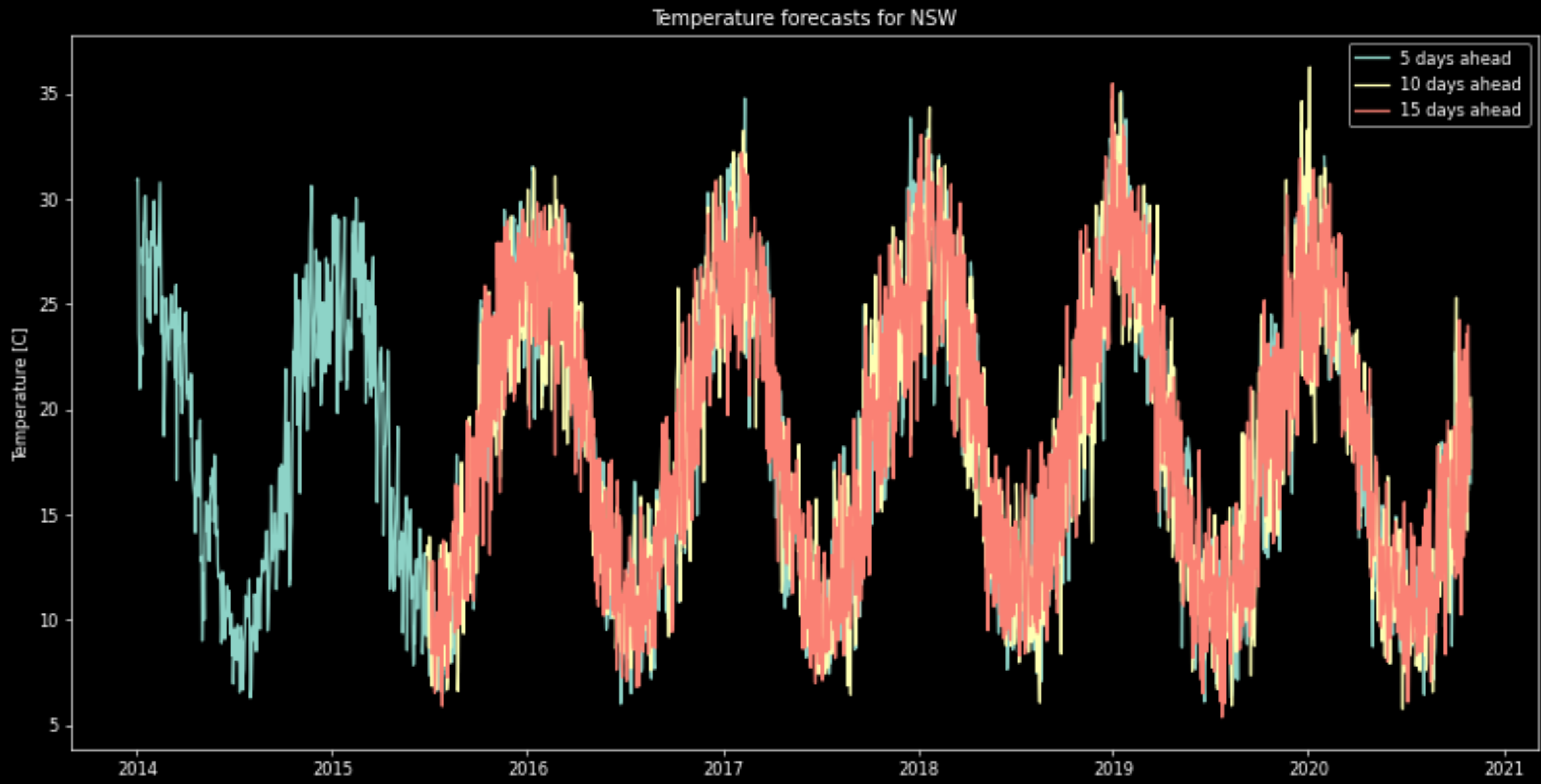
Historical weather forecasts – Time



Historical weather forecasts – Data frame

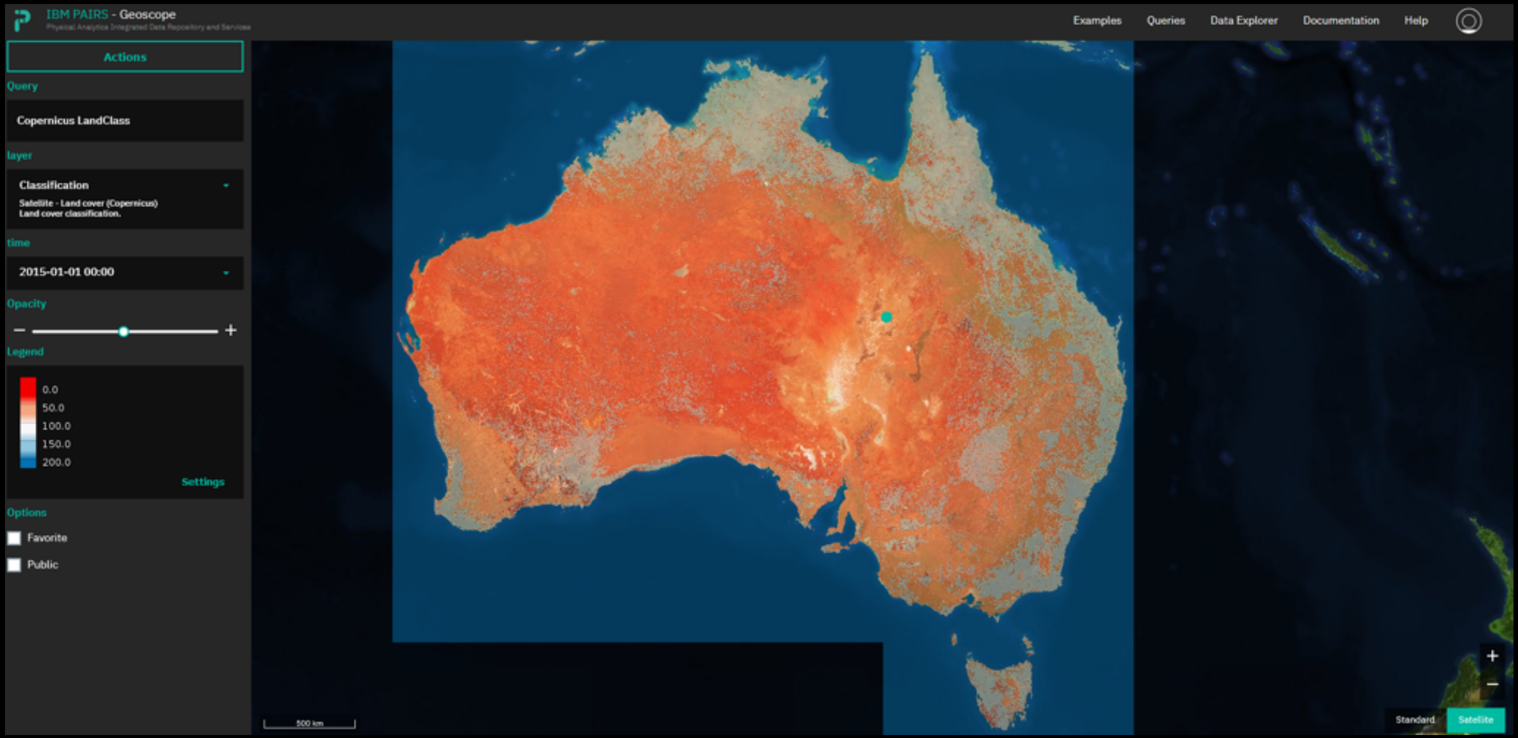
1. Region
2. Date
3. Parameter
4. Lead time [days]
    - 5, 10, 15 days ahead
5. count() [km^2]
6. min(): Minimum value of the spatial aggregation.
7. max(): Maximum value of the spatial aggregation.
8. mean(): Average of the spatial aggregation.
9. variance: 2nd moment of the spatial aggregation.

Historical weather forecasts – Data availability



Land class – PAIRS processing

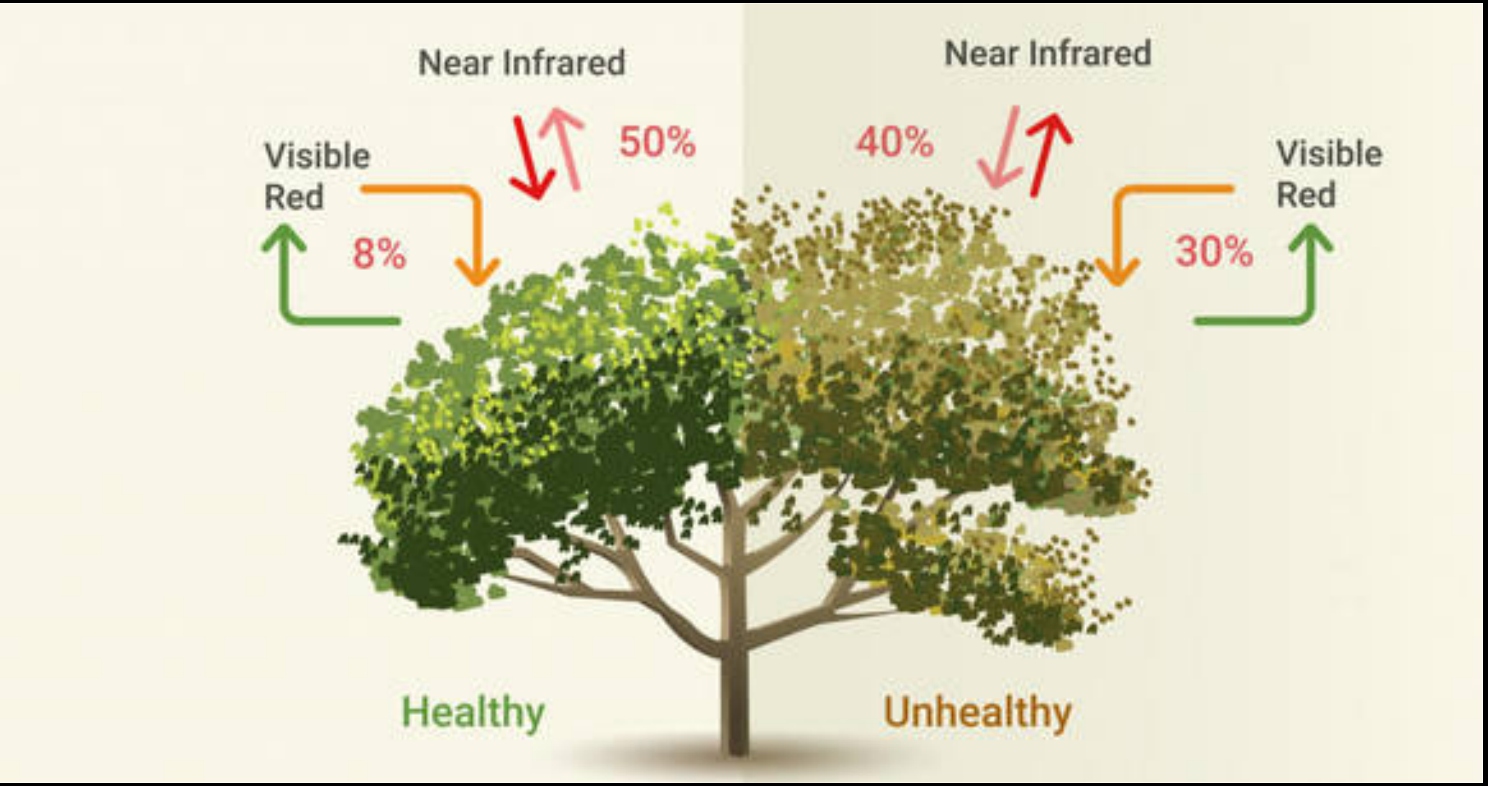
1. Spatially aggregated over 7 regions
2. Data is normalized to 100 [%]



Land class – Columns

1. Region
2. Shrubs [%]
3. Herbaceous vegetation [%]
4. Cultivated and managed vegetation/agriculture (cropland) [%]
5. Urban / built up [%]
6. Bare / sparse vegetation [%]
7. Permanent water bodies [%]
8. Herbaceous wetland [%]
9. Closed forest, evergreen, broad leaf [%]
10. Closed forest, deciduous broad leaf [%]
11. Closed forest, unknown [%]
12. Open forest, evergreen broad leaf [%]
13. Open forest, deciduous broad leaf [%]
14. Open forest, unknown definitions [%]
15. Open sea [%]

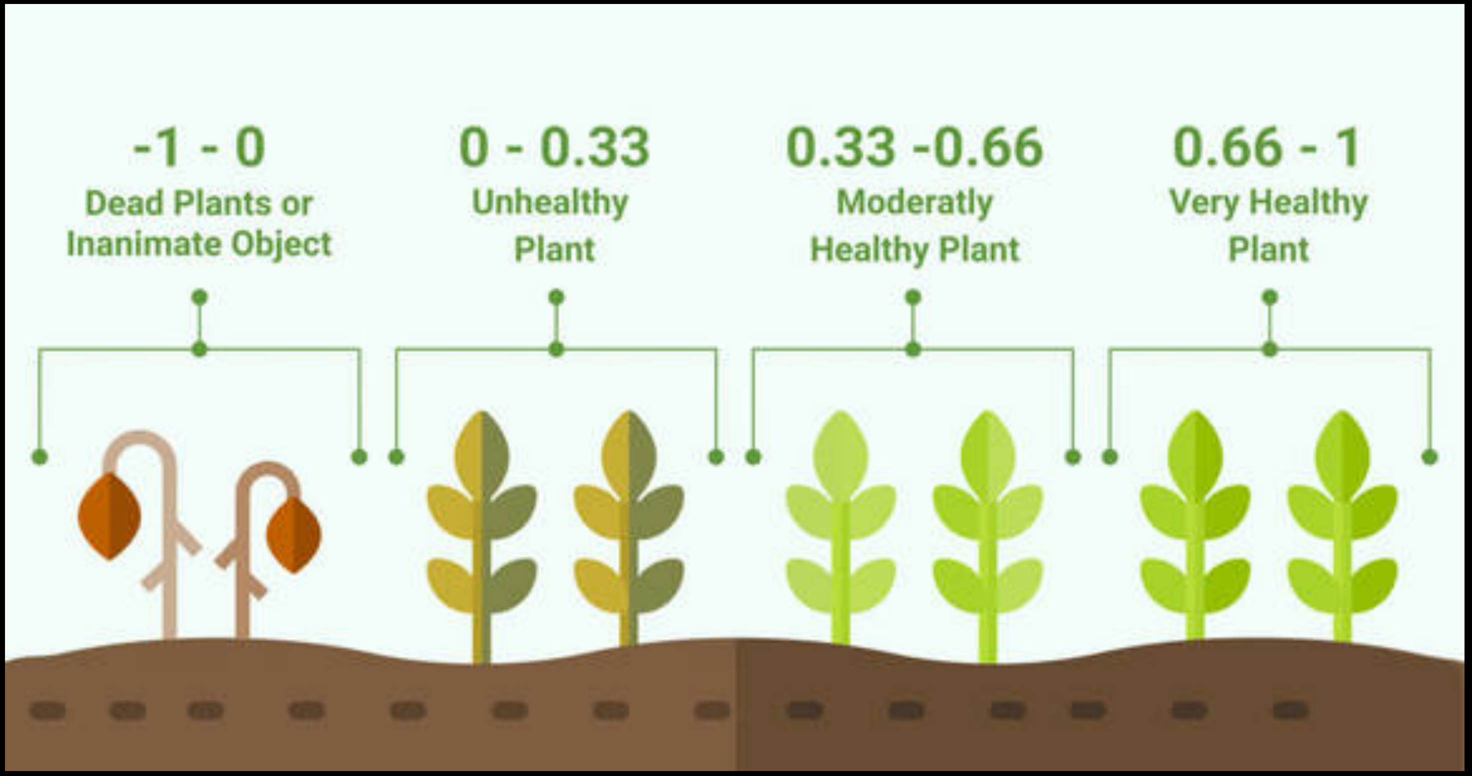
Vegetation index



NDVI = (NIR-Red) / (NIR+Red)

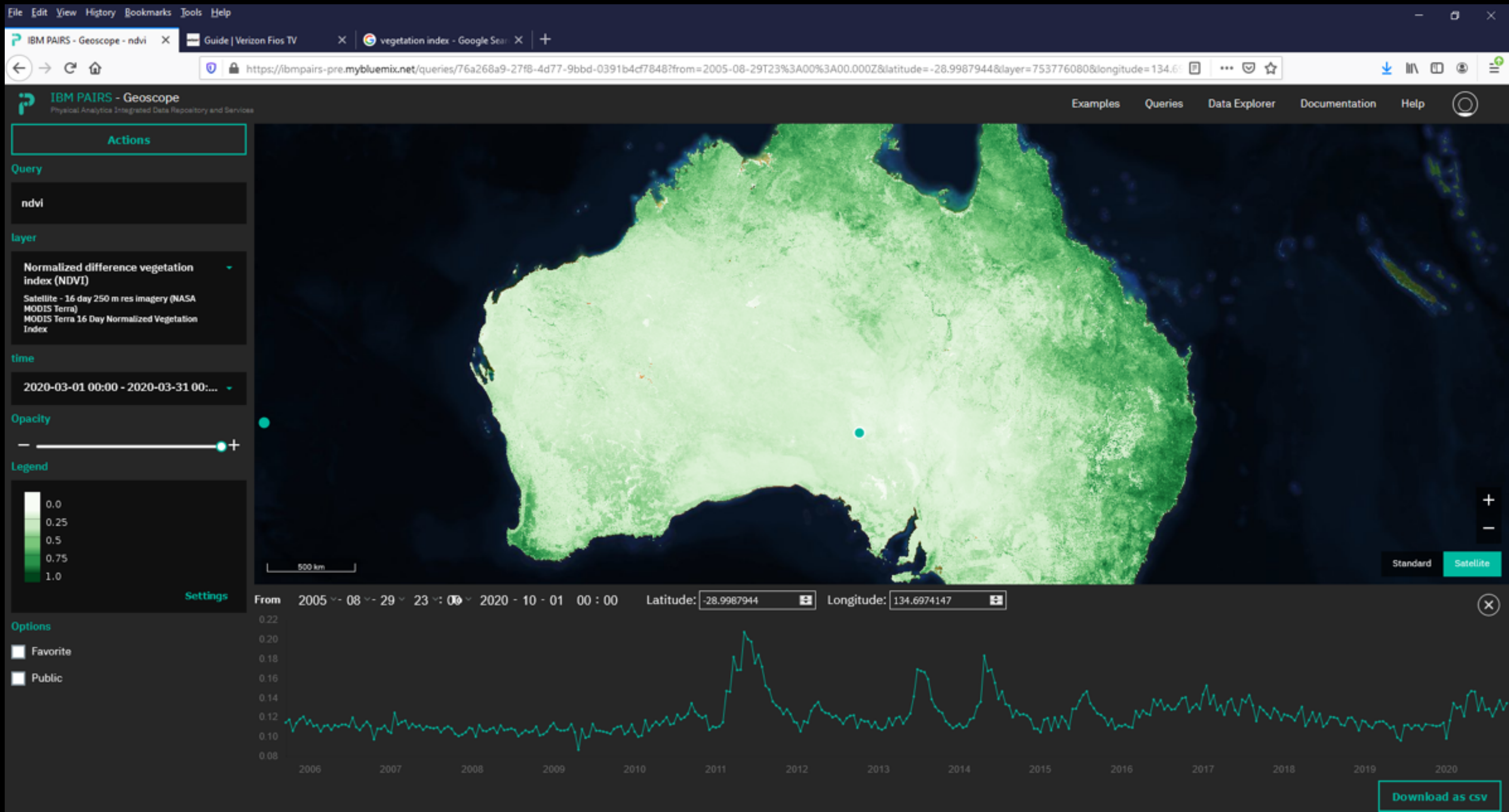
https://eos.com/blog/ndvi-faq-all-you-need-to-know-about-ndvi/

Vegetation index



https://eos.com/blog/ndvi-faq-all-you-need-to-know-about-ndvi/

Vegetation index from a NASA satellite



Vegetation index – PAIRS processing

1. Monthly aggregated
2. Spatially aggregated over 7 regions

Vegetation index – Columns

1. Region
2. Date
3. Vegetation_index_mean
4. Vegetation_index_max
5. Vegetation_index_min
6. Vegetation_index_variance

Vegetation index – Columns

1. Region
2. Date
3. Vegetation_index_mean
4. Vegetation_index_max
5. Vegetation_index_min
6. Vegetation_index_variance