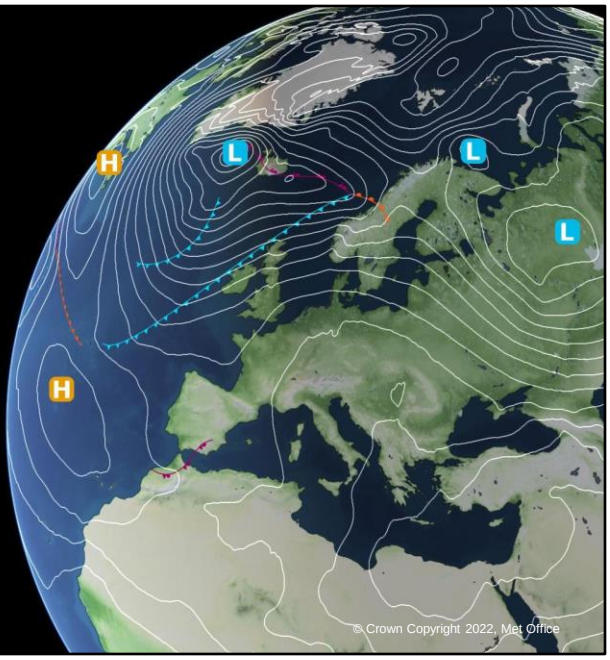
An Introduction to Met Office Data

Nathan Creaser
Harvey South

nathan.creaser@metoffice.gov.uk
harvey.south@metoffice.gov.uk



www.metoffice.gov.uk

Purpose:

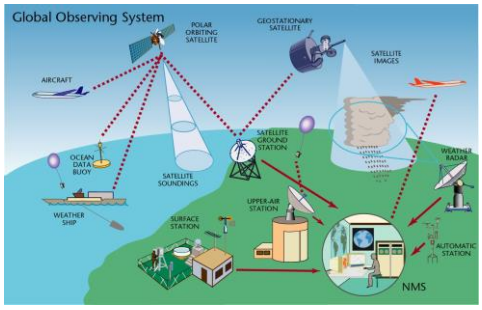
- Insight into Met Office Data
- Awareness of resources for exploring data
- Brief introduction to terms and data surrounding Weather and Climate Science

Met Office

How much data we produce and collect…

- The Met Office data archive system currently stores around 350 Petabytes of data
  - 300 Terabytes are produced per day
  - Produced data expected to grow 12.5x by 2027
  - Archived data to grow to 4 Exabytes by 2032
- New techniques important to capitalise on all this data

Size	Number of bytes
Byte	1
Kilobyte	1,024
Megabyte	1,048,576
Gigabyte	1,073,741,824
Terabyte	1,099,511,627,776
Petabyte	1,125,899,906,842,624
Exabyte	1,152,921,504,606,846,976



"Official" data collection begins in 1914: About Us - Met Office but "Unofficial" data is archived since the early 1700s About our archive collections - Met Office

- Observation collection is processed which feeds into model runs – all is data that needs to be stored

https://blog.metoffice.gov.uk/2017/04/21/big-data-big-challenge/

Met Office

Where to find available Met Office data

CEDA Archive	Datahub	WOW
Centre for Environmental Data Analysis SCIENCE AND TECHNOLOGY FACILITIES COUNCIL NATURAL ENVIRONMENT RESEARCH COUNCIL		

Met Office

Tabular v Gridded

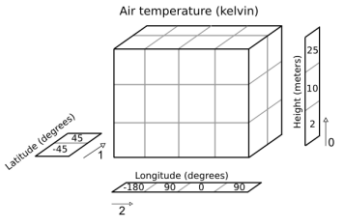
Tabular Data

Values arranged into rows and columns to convey relations between data e.g. .csv

Site Id	Longitude	Latitude	Report Date / Time	Air Temperature	Dew Point	Mean Sea-Level Pressure	Relative Humidity	Wind Speed	Wind Direction	Wind Gust	Visibility
1040	-2.64	51.006	2019-10-01 00:00:00	15.3	13.9	1002.0	91.3	7.0	202.5	NaN	7
1040	-2.64	51.006	2019-10-01 01:00:00	14.6	13.4	1001.0	92.5	6.1	180.0	NaN	7
1040	-2.64	51.006	2019-10-01 02:00:00	15.1	13.8	1000.0	91.9	5.2	180.0	NaN	7
1040	-2.64	51.006	2019-10-01 03:00:00	15.1	13.8	999.0	91.9	4.3	202.5	NaN	7
1040	-2.64	51.006	2019-10-01 04:00:00	15.4	14.1	998.0	92.0	6.1	202.5	NaN	7
...	...	...	...	...	...	...	...	...	...	...	...
7007	-1.33	52.358	2019-07-31 19:00:00	17.8	13.1	1013.0	73.8	8.7	270.0	NaN	9
7007	-1.33	52.358	2019-07-31 20:00:00	16.9	13.2	1014.0	78.7	7.0	270.0	NaN	9
7007	-1.33	52.358	2019-07-31 21:00:00	16.6	13.4	1014.0	81.3	6.1	270.0	NaN	9

Gridded Data

An array of values arranged on an n-dimensional grid



CSV: [Comma-separated values - Wikipedia]

Gridded Data Commonly in netCDF format: [NetCDF - Wikipedia]

Met Office

Tabular data

- In real-world applications, the most common data type is tabular data[1]
- Environmental examples:
  - Site based Weather Observations
  - Site Specific Forecasts
  - Aircraft Sensor data

[1]: Ravid Shwartz-Ziv, Amitai Armon "Tabular data: Deep learning is not all you need"

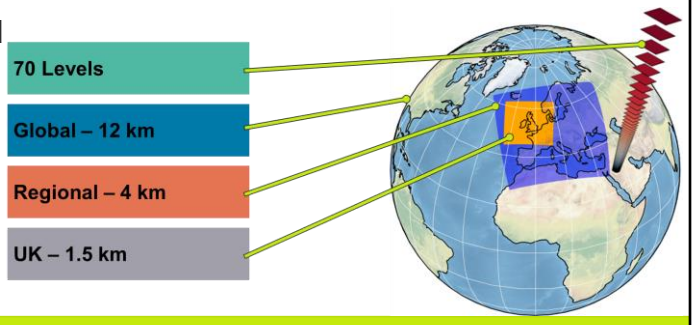
Aircraft e.g: [Dataset Record: MRF A780 JET2000 flight: Airborne atmospheric and chemistry measurements taken on board the Met Office C-130 Hercules aircraft (ceda.ac.uk)]

Microsoft Excel is a popular program to interact with and enter tabular data

Met Office

Gridded Model Data

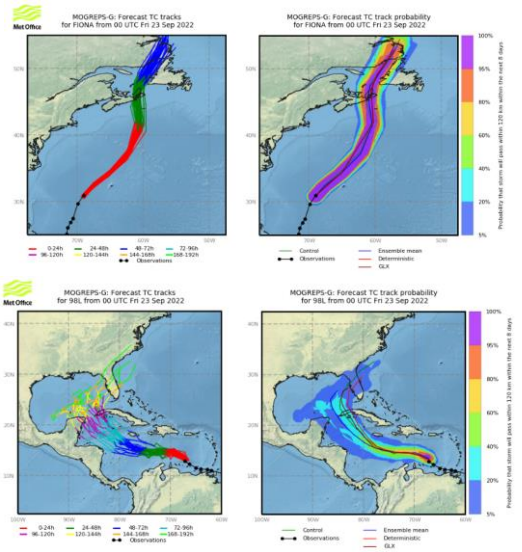
- Resolution
- Single Level vs Multi Level
- Convective v Parameterised
- Climate models
- Uncertainty
- More on: Ensembles



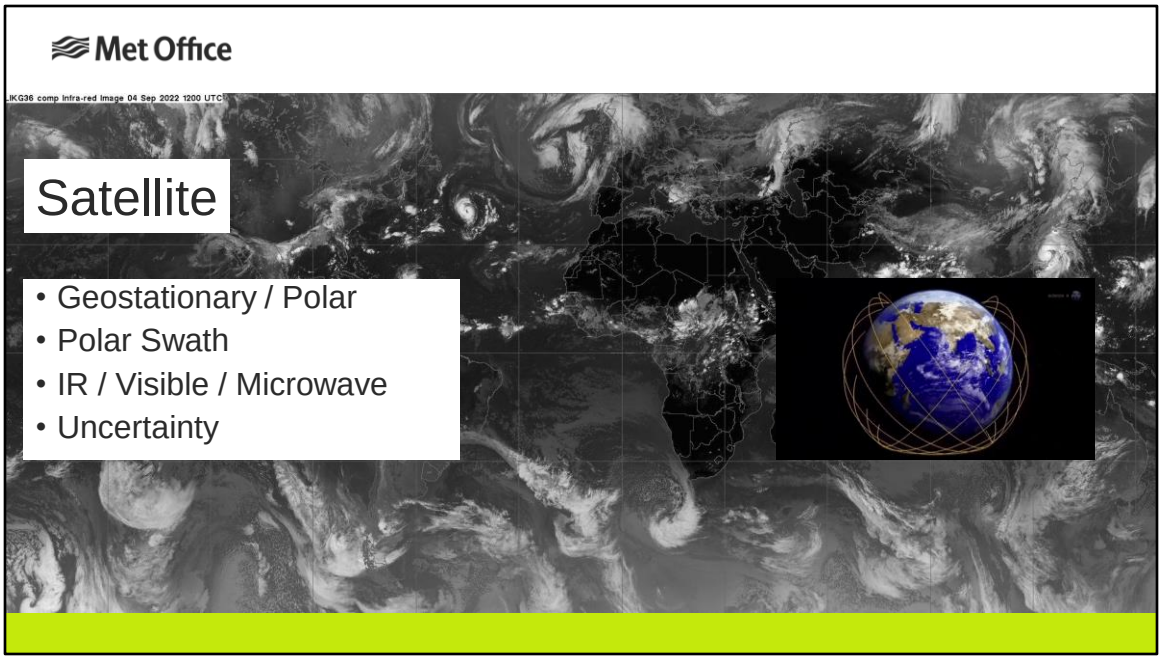
[Key terms Glossary]
[Data - Met Office]

Ensembles

- Different realizations of the same model
- Changing initial conditions changes outcomes
- A consequence of chaos theory
- Spread of ensemble members represent some uncertainty



[What is an ensemble forecast? - Met Office]



Satellite imagery - Met Office
https://www.metoffice.gov.uk/public/weather/world-satellite/#?tab=satImg&map=regionalIR&fcTime=1661882400
https://view.eumetsat.int/
https://earthobservatory.nasa.gov/
https://www.sentinel-hub.com/

Met Office

More Data Categories

- Timescale (Hourly, Daily, Climate Scale etc.)
- Observational vs Simulation
- And of course, many more types/terms of data: Key terms

Metadata

What is Metadata?

Metadata: Data describing data + CF Conventions Home Page (Metadata Conventions for netCDF files)
ML Data for Earth Science: Radiant ML Hub – Radiant Earth Foundation
Lots of Related Data on Open AWS: Registry of Open Data on AWS

Met Office

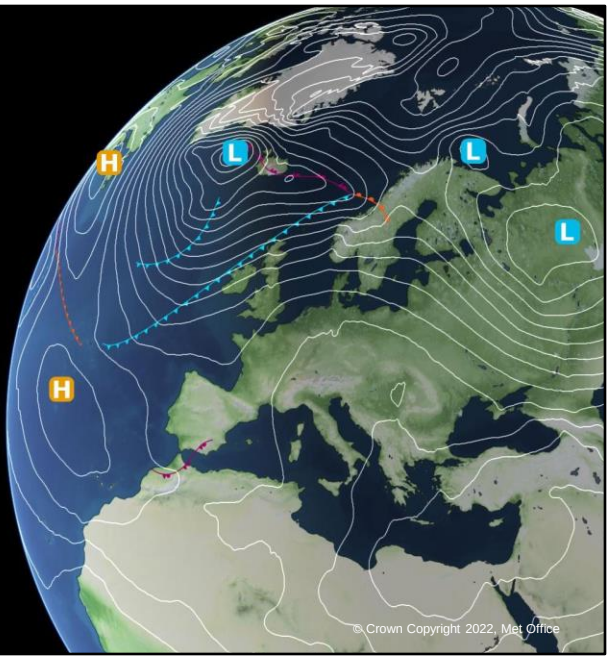
Ethics

- Considerations
  - Uncertainty (In decisions, measurements, and outputs)
  - Effects of decisions made from your projects
  - Licensing
- Implementations
  - Careful Documentation
  - Reading Documentation

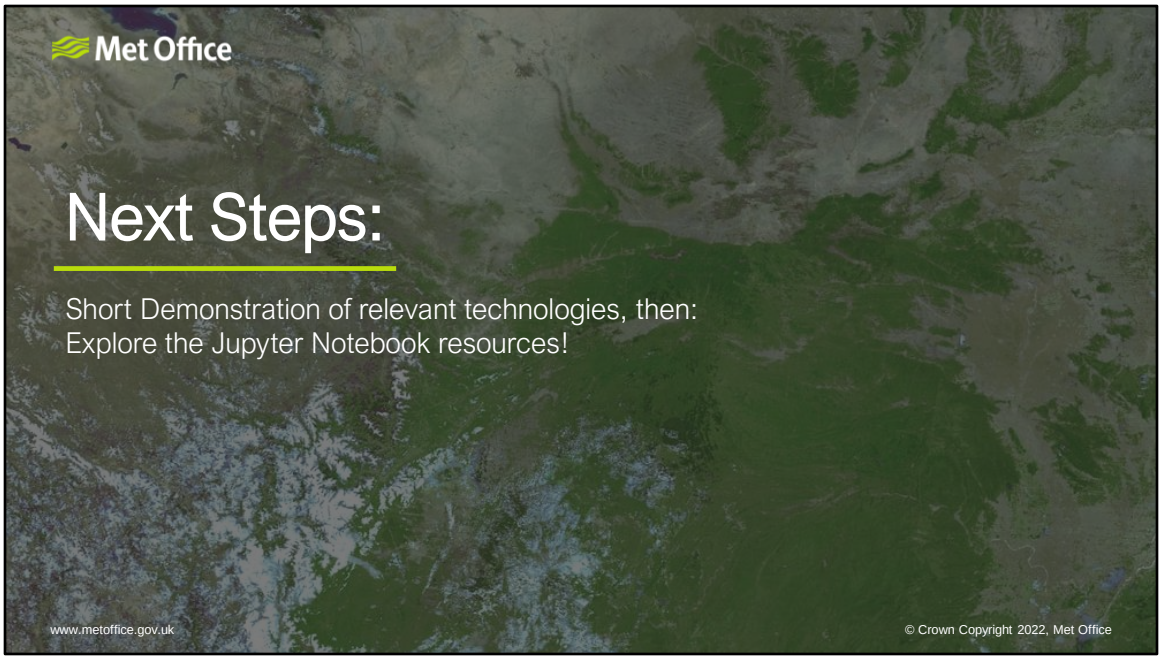
Observations are not infinitely precise – Uncertainty
Models propagate the uncertainty.

Met Office

Any Questions?



www.metoffice.gov.uk



See links above in presentations, or visit [informatics-lab/spaceapps-2022-mo-bootcamp (github.com)]