

Introduction to Reanalysis



Prof Sarah L Dance

s.l.dance@reading.ac.uk



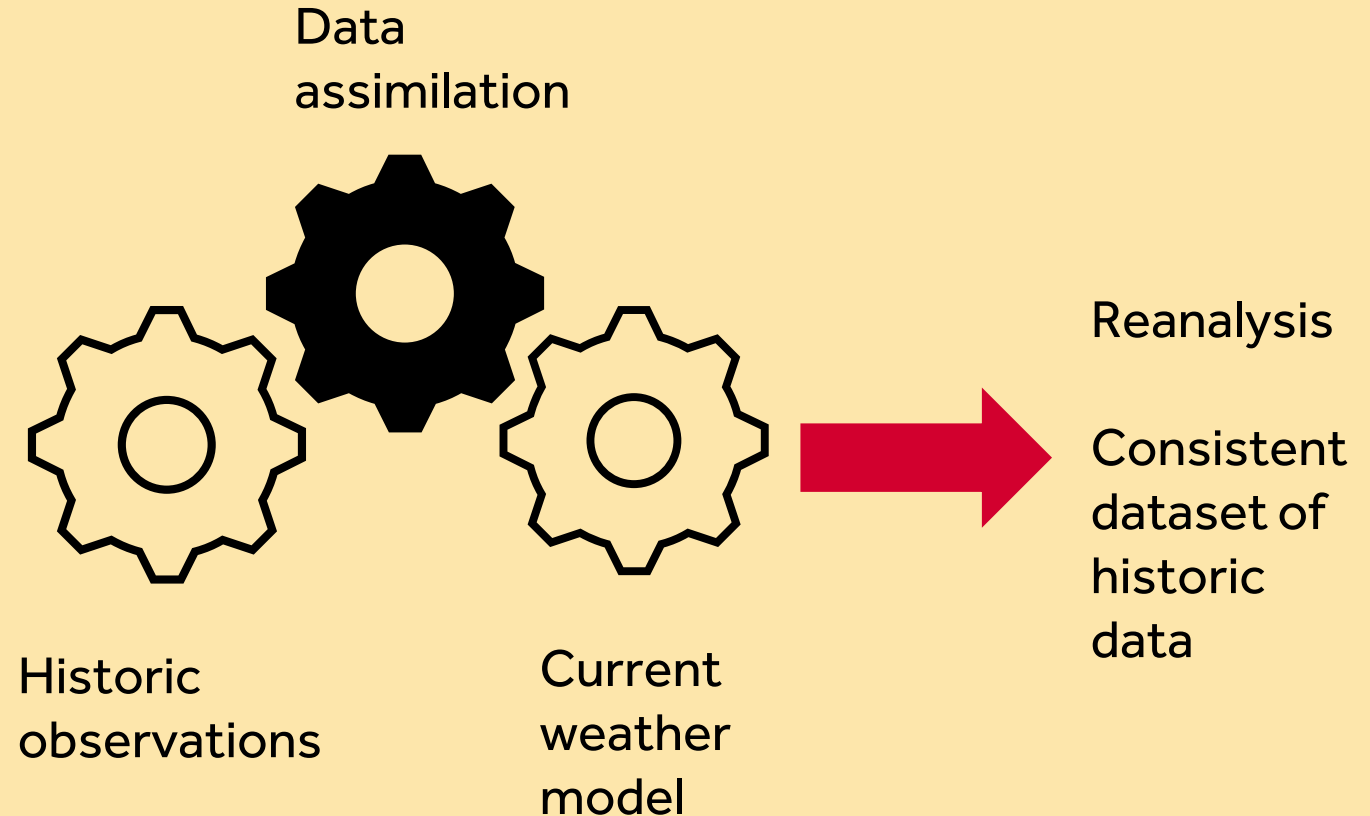
@DrSarahDance

Outline

- **What** is reanalysis?
- Reanalysis **strengths** and **weaknesses**
- Please note that I will talk mostly in the context of weather data for simplicity, but of course you can do this with all components of the earth system!

What is reanalysis?

- A consistent set of historic weather data, at regular time and space intervals
- An optimal combination of the current state-of-the-art weather prediction model with observations taken at the time.
- By historic observations, we mean any type of observational data from as far back as records begin to the present day.

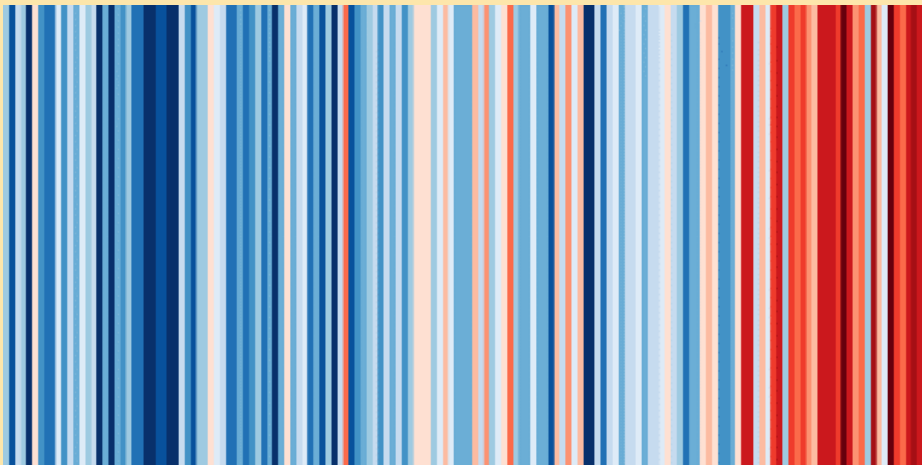


Strengths of weaknesses of reanalyses

- Come from the ingredients used to make them
 - Observations
 - Dynamical model
 - Assimilation system

Observation datasets

- A long term dataset of observations provides important information about a range of features, including short-term and underlying trends as well as specific events
- But data may be spatially or temporally limited, or very uncertain



Warming stripes for Berkshire, showing average annual temperature in Berkshire from 1863-2020

Temporal consistency

For long time-series over decades, data consistency is likely to be an issue

- The observing instruments may change over time
- There may be gaps in the records, and fewer observations as you go back in time
- The position of the “same” observation may move by a few or possibly hundreds of metres
- A single location may change over time as trees grow, buildings come and go and the land-use of the surrounding area may change (e.g., urbanization)



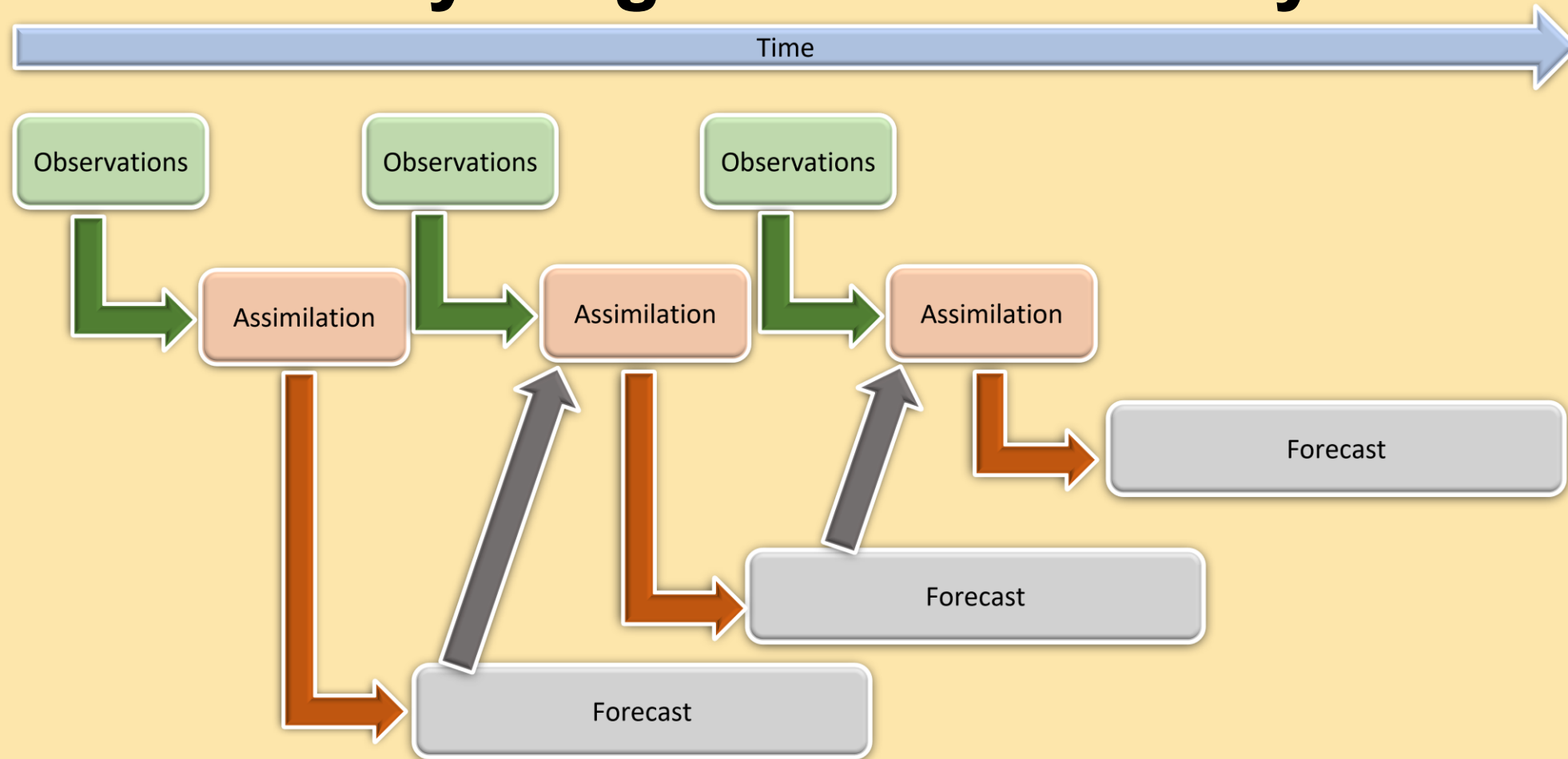
Data latency

- How long does it take for observations to arrive? (aka **data latency**)

Example 1

- The UK Met Office takes hourly observations from around 270 weather stations in near real time
- BUT daily summary observations from around 170 human observers come in at the end of each month

Assimilation cycling and data latency



Reanalysis allows us to assimilate modern observations that arrive too late to be used for forecasting

Example 2 – Data Rescue

Station *Chemulpo* Monat *September* 19 *15*

Datum	Barometer (mm) reduziert auf 0° C., Seehöhe =			Temperatur (Centigrade)						Dunstspannung in mm			Relat. Feuchtigkeit in Prozenten				
	Ya	2p	9p	Trockenes Therm.			Feuchtes Therm.			Ya	2p	9p	Ya	2p	9p		
				Ya	2p	9p	Ya	2p	9p							Max.	Min.
1	54.3	53.9	54.2	23.2	20.4	25.0	22.3	27.0	24.0	31.3	23.0	19.5	24.4	24.6	92	87	92
2	53.6	49.3	43.9	22.4	20.2	19.3	22.0	19.8	18.8	23.1	21.8	19.4	16.9	15.9	96	96	95
3	38.2	43.4	46.4	18.7	21.1	20.8	18.6	20.9	19.9	21.6	18.5	15.9	18.3	16.7	99	98	92
4	49.6	51.6	52.2	21.3	25.3	21.6	20.9	23.5	21.0	25.9	19.0	18.1	20.4	18.1	96	85	95
5	53.1	53.7	54.5	19.8	26.4	21.4	19.7	20.2	20.5	27.4	19.5	16.5	19.2	17.4	96	75	92
6	55.3	54.7	55.0	20.4	26.8	22.1	20.2	23.8	21.2	27.7	20.0	17.5	20.1	18.2	99	77	92
7	57.0	55.7	56.7	21.9	26.4	23.8	21.2	24.0	22.4	29.3	20.8	18.3	19.5	19.3	94	68	88
8	56.7	53.6	52.7	20.7	22.2	20.4	19.0	20.1	20.0	23.9	20.0	15.3	16.2	17.1	84	82	97
9	53.5	53.9	54.6	25.6	22.3	21.8	23.4	21.6	21.2	25.7	20.0	21.3	18.7	18.3	98	94	95
10	53.8	52.5	53.1	20.8	23.4	22.0	20.6	23.0	21.8	23.8	17.0	17.9	20.7	19.3	98	97	98
Summen	525.1	521.7	524.2	212.8	246.5	218.2				257.3	199.6	179.7	194.4	181.9	952	849	936
11	55.3	55.4	56.8	20.3	26.0	21.0	19.8	22.2	20.2	26.2	20.0	16.9	17.6	17.7	95	71	93
12	56.9	54.8	54.4	19.0	24.6	21.6	18.8	21.8	20.0	25.8	18.0	16.0	17.7	16.4	98	77	86
13	53.4	52.2	51.0	21.4	22.4	22.0	21.4	22.3	22.0	23.8	20.6	18.9	20.0	19.7	100	99	100
14	57.1	50.5	51.5	23.8	23.8	22.8	23.4	23.8	22.4	24.2	22.0	21.2	22.0	19.9	94	100	96

Windrichtung nach Beaufort		
Ya	2p	9p
SW	2	SW
SW	2	SW
SW	4	N
SW	3	SW
SW	2	WSW
SW	2	N
SW	2	WSW
SW	2	E
SW	8	NW
SW	2	SW
SW	2	SW
SW	2	SW
SW	7	SW

- Paper records can also be digitized and used for reanalysis
- Image from Chemulpo, South Korea from September 1905, from the DWD archives at Deutscher Wetterdienst
- Such records can't be digitized automatically (e.g., via optical character recognition) as this would result in too many errors.
- Citizen science projects instead

Kaspar et al (2015)
<https://doi.org/10.5194/asr-12-57-2015>

The dynamical model

- Model provides **physical consistency** in space and time
- “Maps with no gaps”
- Spatially and temporally comprehensive (at the resolution of the model output)
- Internally consistent
- Variables changing in a physically credible way
- DA allows for uncertainties to be accounted for objectively



Quiz (in person/poll online)

Match the limitations to the components of the system

Components

Observations

Models

Limitations

A Limited resolution: small details are missed or have to be approximated

B Limited coverage in space or time

C Updated versions will give different results with the same input data

D Can be inconsistent over time

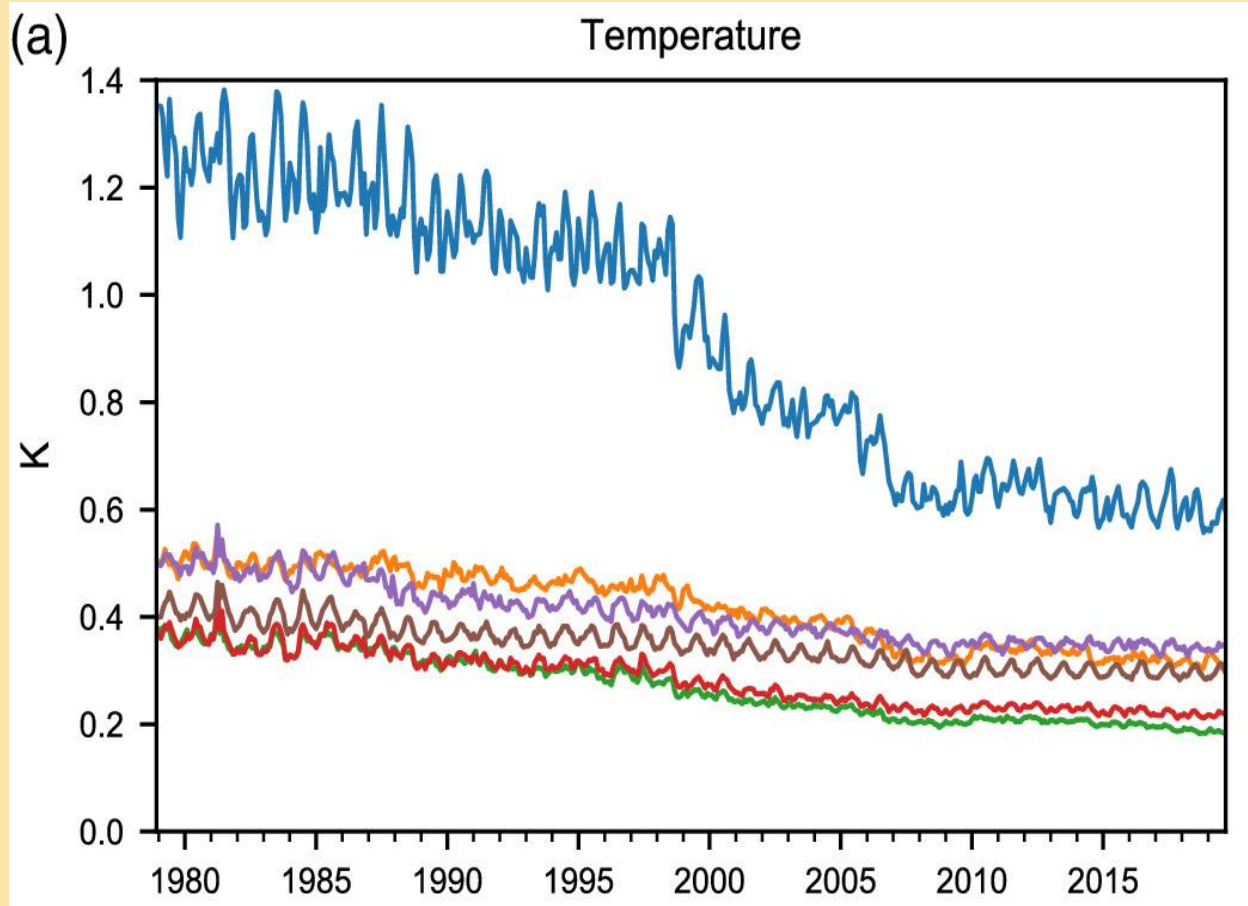
Key concepts and cautions

- Reanalyses are created using an unchanging DA system and model that ingests all of the (QC-ed) observation data available over the reanalysis period
- While the model and DA system are kept constant the **nature of the observation data changes with time**, due to changes in the observing network.

Reanalysis Uncertainty

- DA takes account of the uncertainties in the observation and model data (see other lectures) to produce an improved estimate of the atmospheric system.
- However, the reanalysis data nevertheless contains residual uncertainties.
- Some reanalysis products come with uncertainty metadata
- These are NOT usually a complete, classical metrological estimate of the uncertainty
- Instead they are a by-product from the assimilation, and thus don't take into account any systematic errors that are not accounted for in the DA system

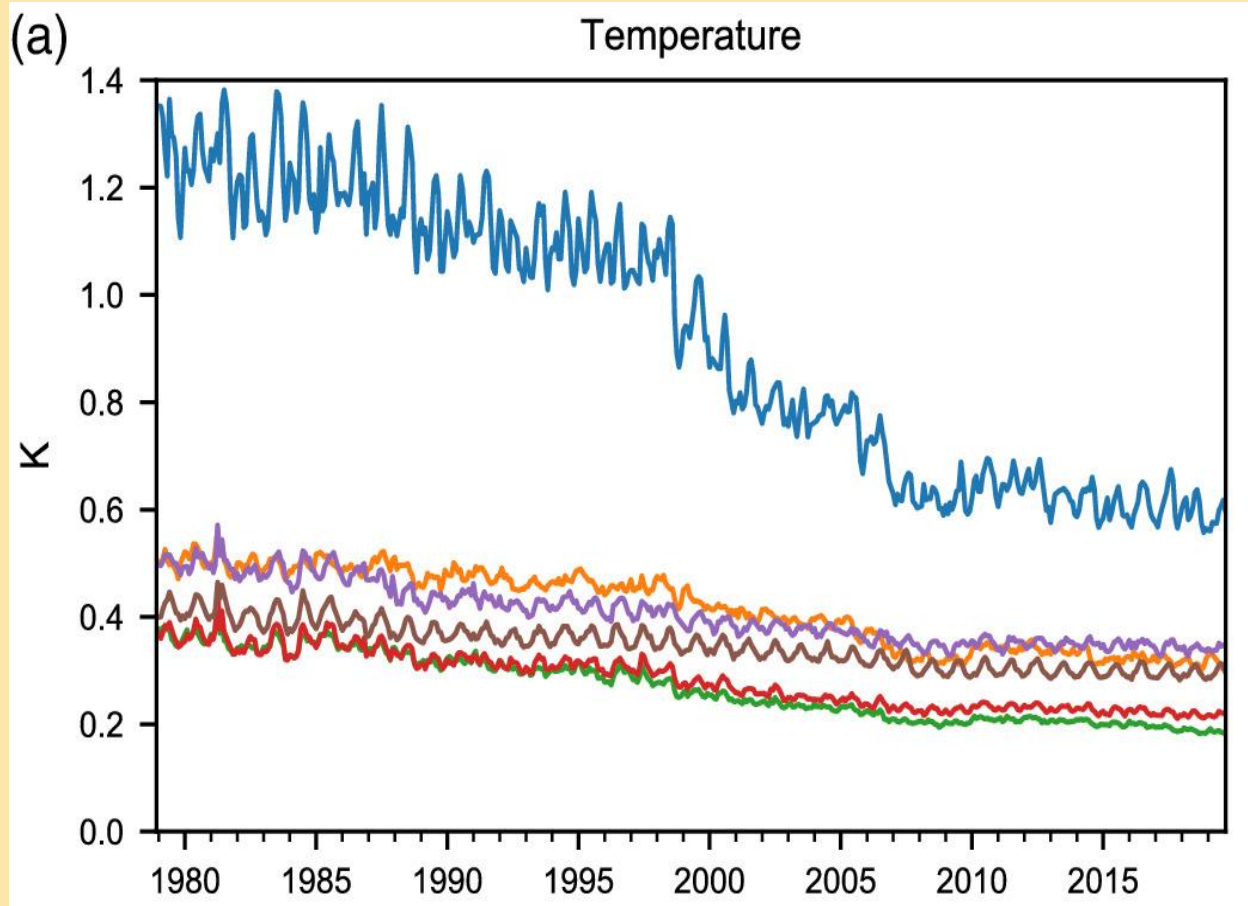
Uncertainty Example



- Time series of monthly and globally averaged ERA5 ensemble spread from 1979 to 2018 at indicated pressure levels for temperature (K).
- Different colours are different heights in the atmosphere
- Large ensemble spread should indicate high uncertainty

from Hersbach, et al. Q J R Meteorol Soc. 2020;
146: 1999– 2049. <https://doi.org/10.1002/qj.3803>

Uncertainty Example



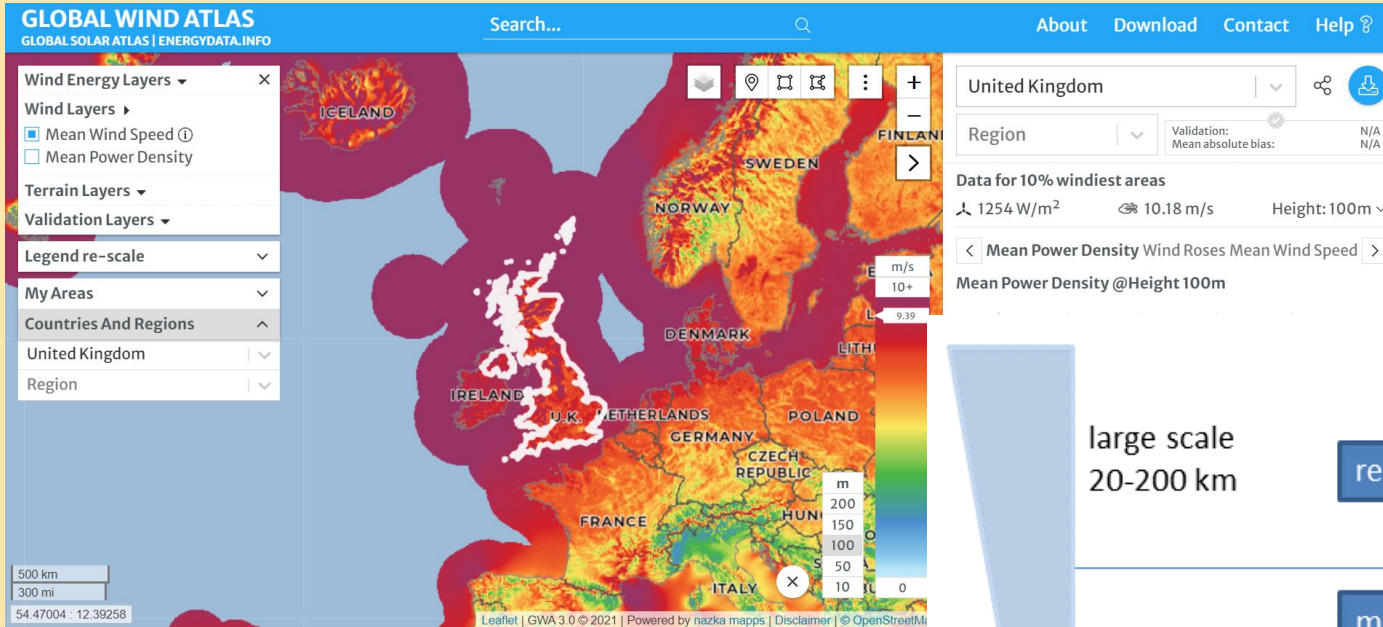
from Hersbach, et al. Q J R Meteorol Soc. 2020;
146: 1999– 2049. <https://doi.org/10.1002/qj.3803>

- Improvements often coincide with introduction of a new observing instrument
- Blue line is ensemble spread in the stratosphere
- Large improvements coincide with the coming of a new satellite instrument (ATOVS) 1998 and growth in numbers of another type of satellite data (GNSS RO) in 2006.

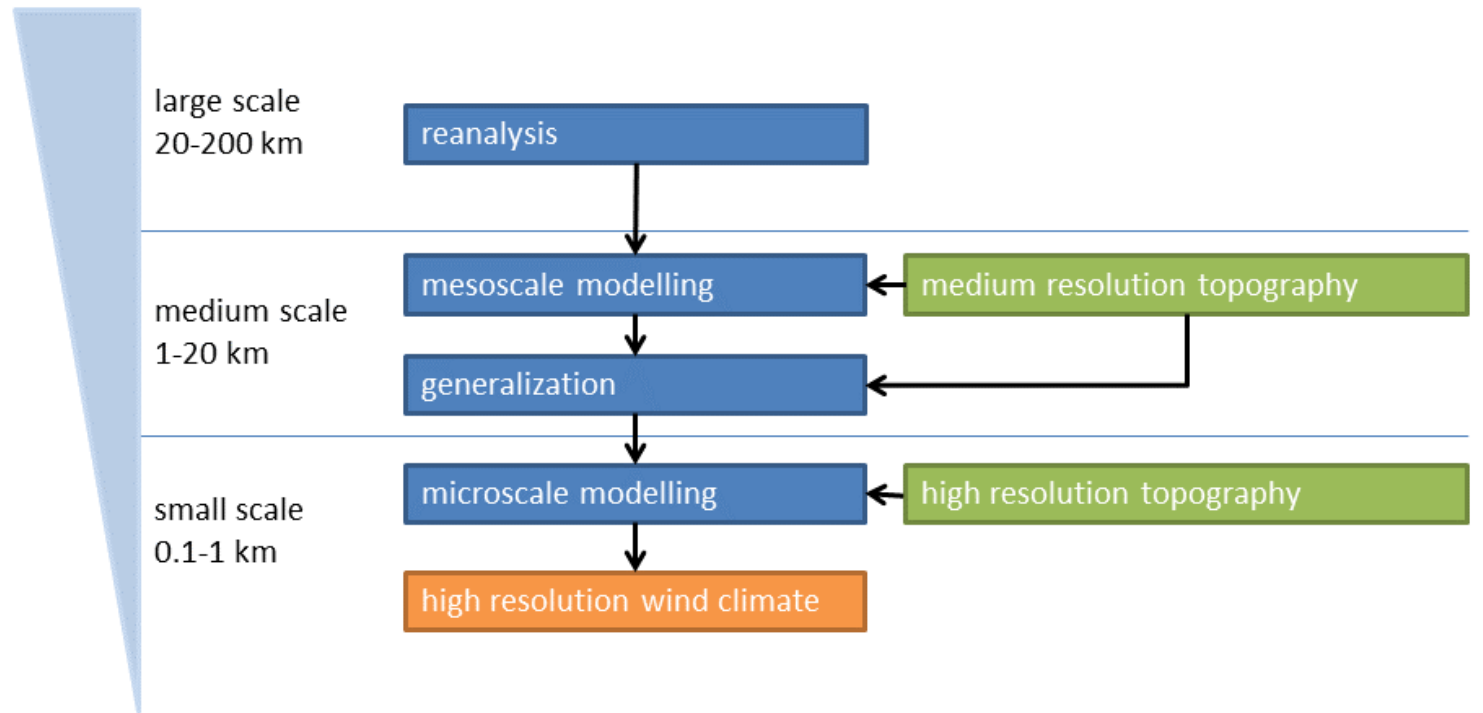
Representative scales

- A typical spatial temporal temporal gridding for global reanalysis is **3 hourly intervals** and **30 kilometres** in the horizontal
- These scales may not resolve the details of the variables of interest required for studies of systems impacted by extreme events
- For studies involving complex terrain such as mountains or coasts it may be better to use the reanalysis to drive a physically consistent downscaling model

Example – using reanalysis to site wind-turbines

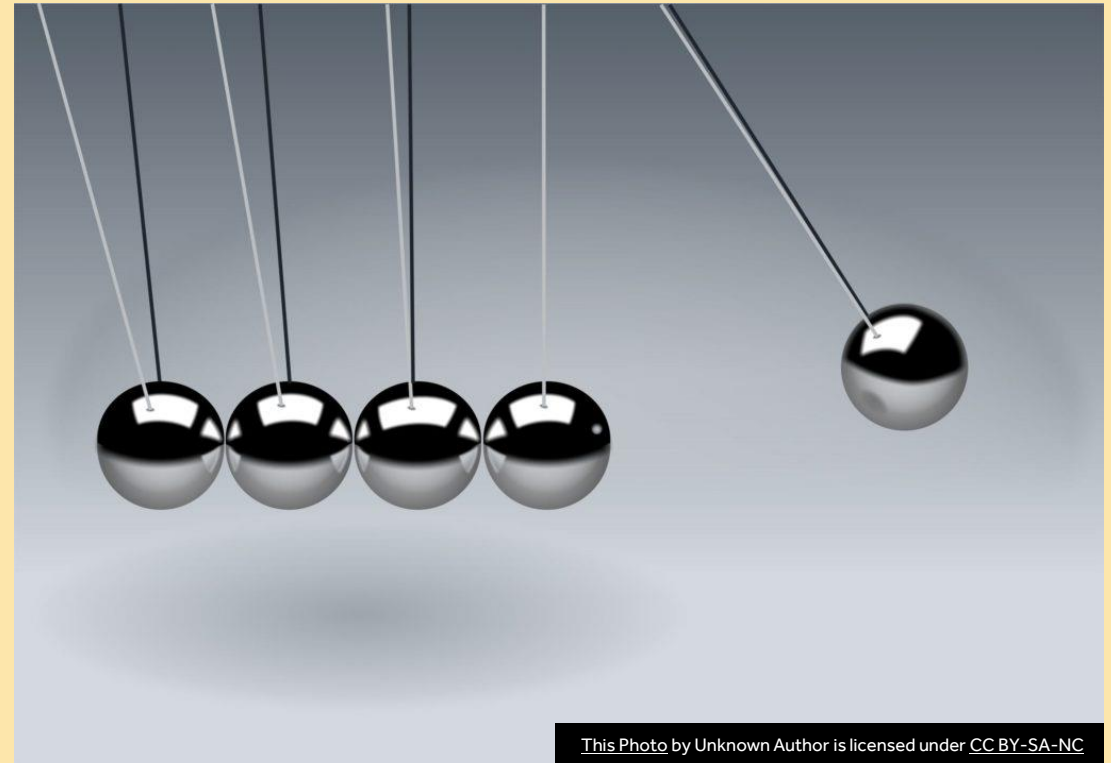


“Global Wind Atlas 3.0, a free, web-based application developed, owned and operated by the Technical University of Denmark (DTU). The Global Wind Atlas 3.0 is released in partnership with the World Bank Group, utilizing data provided by Vortex, using funding provided by the Energy Sector Management Assistance Program (ESMAP). For additional information: <https://globalwindatlas.info>”



Non-conservative system

- Reanalyses do not preserve physically conserved quantities (e.g., energy)
- The assimilation acts to keep the model “on-track”, closed to the observed quantities.
- For example, if a model has a warm bias compared with the real world, then the assimilation will remove energy so that the analysis fits the observations more closely.

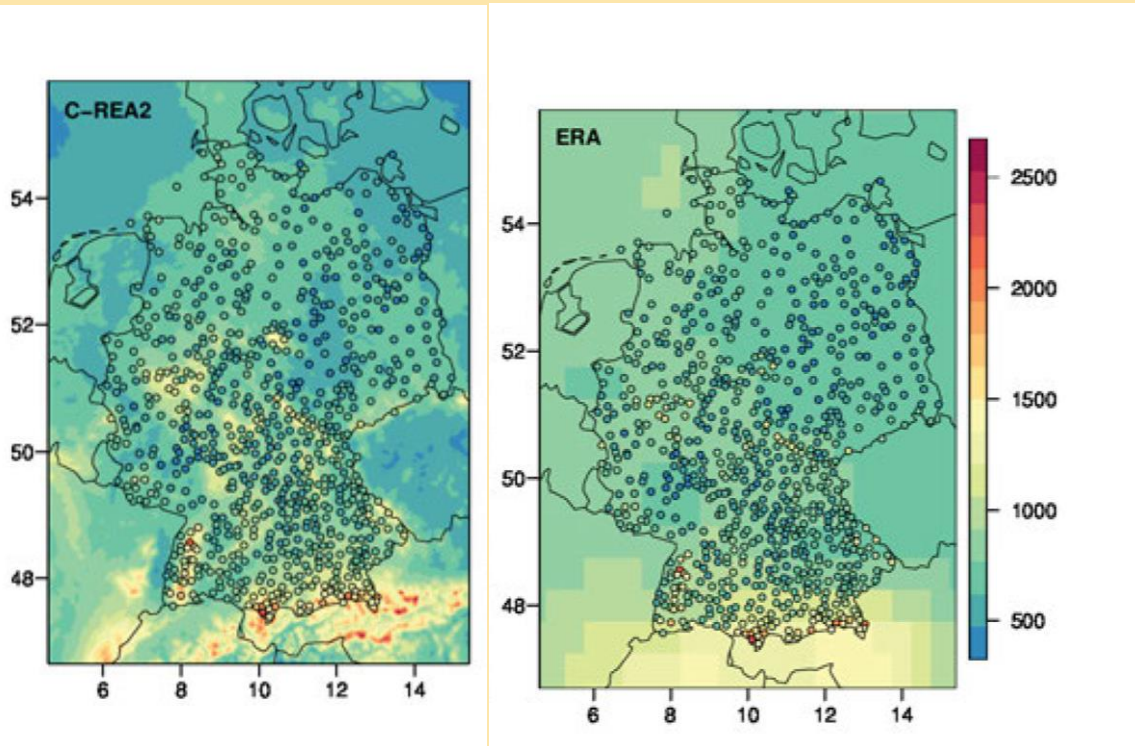




Some variables are more trustworthy than others

- Reanalyses **inherit limitations** from the dynamical model used
- Often models cannot represent the complex process is associated with cloud and precipitation very well
- This makes variables like **pressure** more reliable than variables related to **precipitation**

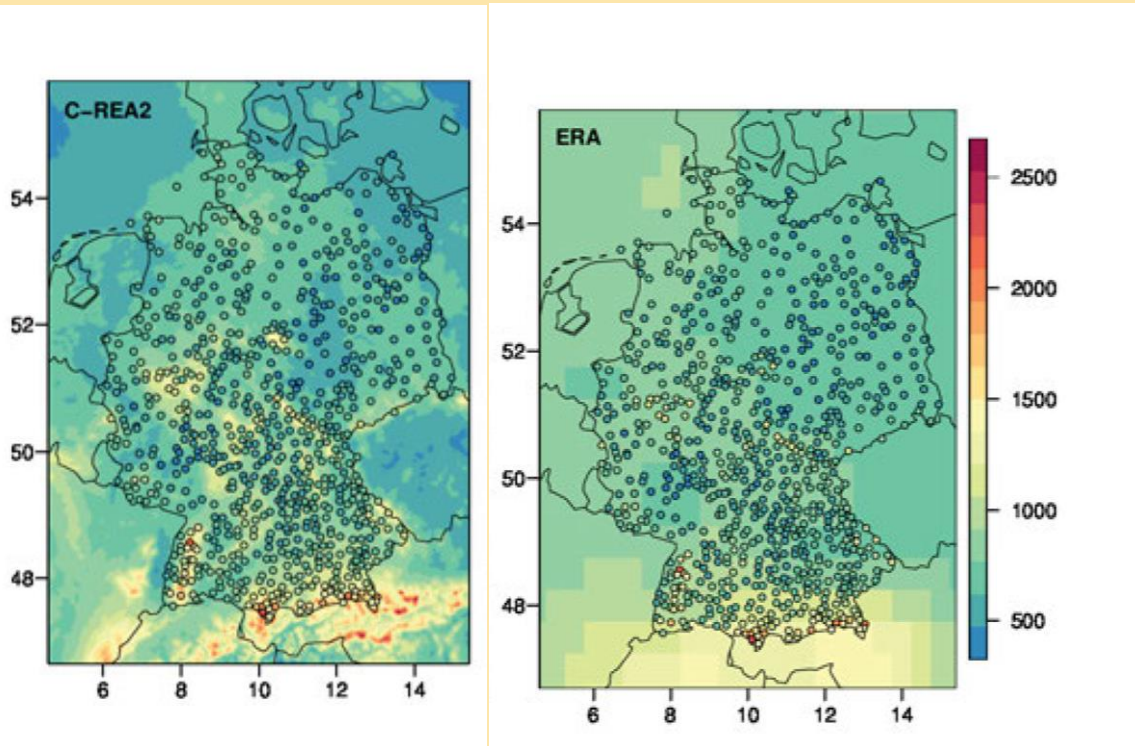
Example - precipitation over Germany



Wahl, et al 2017.
<https://doi.org/10.1127/metz/2017/0824>

- Mean annual precipitation in mm/year over Germany for the period 2007-2013, from a regional reanalysis (C-REA2, 2km grid) and the global reanalysis ERA-Interim (80km grid).
- Independent data from 755 rain gauges is overplotted (small circles).
- The mean annual precipitation received at a rain gauge station between the years 2007 and 2013 varies between 500 and 2300mm/yr.
- The regional reanalysis is on average able to capture this variability, but the ERA-Interim misses the extreme events and only has a range of between 600 and 1400mm/yr.

Example - precipitation over Germany



- This difference in variability between the regional and global reanalyses is in part due to the differences in grid-length between the systems and ability to capture extremes.
-
- Nevertheless they both underestimate the frequency of large events and overestimate the frequency of small events.

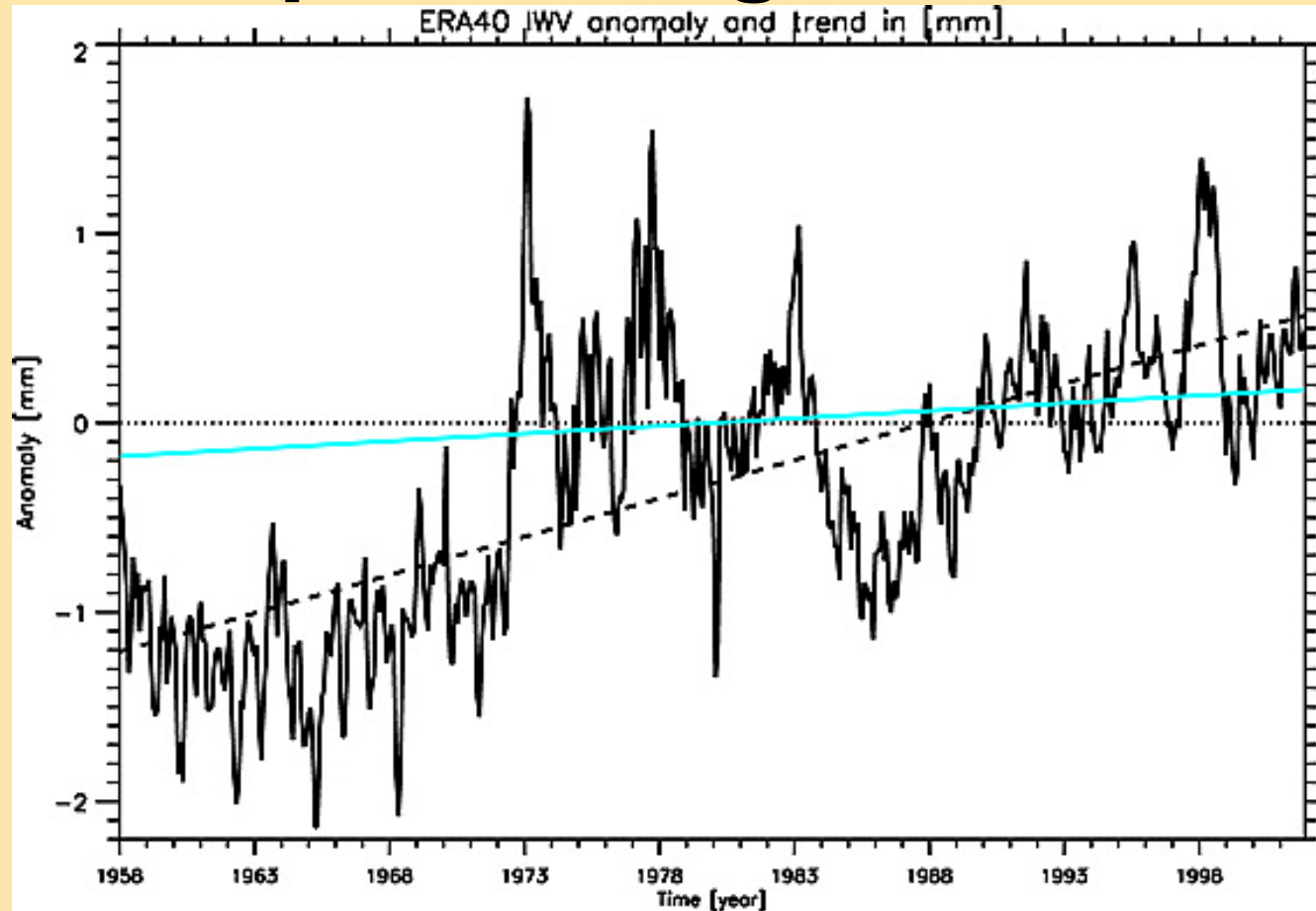
Wahl, et al 2017.

<https://doi.org/10.1127/metz/2017/0824>

Changes in the system with time – beware trends

- While the model used in the reanalysis system is kept constant with time the observing network changes
- Reanalysis developers put significant effort into reducing or removing spurious jumps and trends introduced by
 - changes in the observing system
 - time varying biases in models and observations
- Long term climate signals are inevitably affected in these data and should be viewed with caution

Example – Bengtsson et al (2004)



Bengtsson et al (2004). Can climate trends be calculated from reanalysis data?. *Journal of Geophysical Research: Atmospheres*, 109(D11).

- Vertically integrated water vapour (IWV) anomalies (differences from a reference value) from the ERA40 reanalysis for the period 1958–2001.
- Black line shows the data from ERA-40.
- Dashed line shows an increasing trend calculated from these data.
- Cyan line – trend from reanalysis experiment with no satellite data

Choice of reanalysis products

What time
period?

Which variable
are you
interested in?

Which temporal
and spatial
resolution for
you need?

What do you
want to use it
for?

Satellite data
or not?

Which model and
assimilation
system would
you prefer?

Useful websites

- Comparison of atmospheric reanalyses

<https://reanalyses.org/atmosphere/comparison-table>

- Comparison of ocean reanalyses

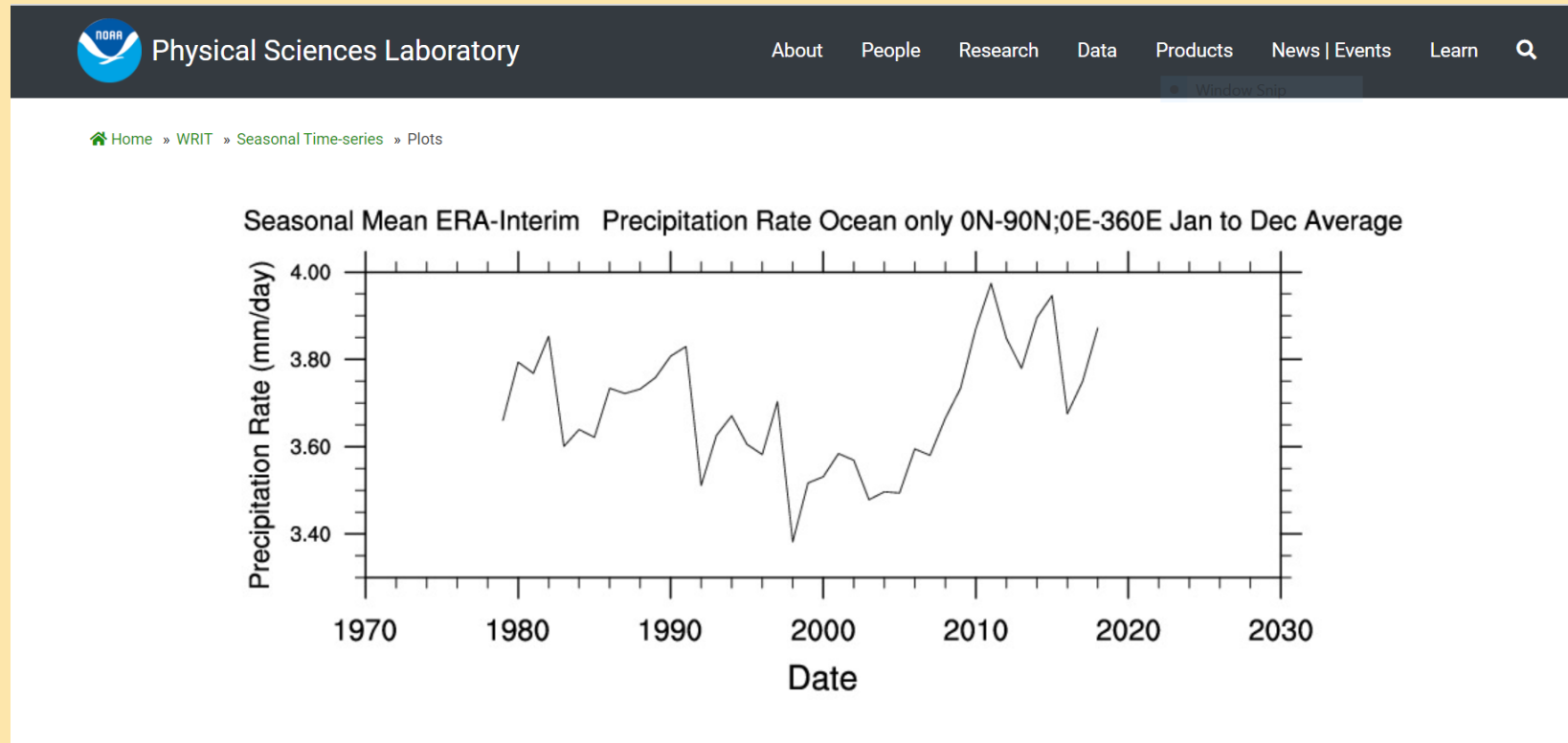
<https://reanalyses.org/observations/ocean-reanalyses-table>

Exercise - if time

1. Visit <https://psl.noaa.gov/data/atmoswrit/timeseries/> - we will use this website to produce a timeseries plot of some reanalysis data.
2. Select the following on the webpage (leave the other settings as default)
 - Dataset 1 – ERA Interim
 - Which variable for D1? – Precipitation Rate
 - Time averaging – Seasonal
 - Land/Ocean Mask options – Ocean
- Then click create plot

Exercise cont

3. The plot should look like this



4. Q What do you think might have happened in 1992?

- Answer: A new observation type was added, namely rain-affected SSM/I satellite radiances. The impact of these satellite observations on the assimilation is most directly felt over the oceans where there are few other in situ observations to constrain the precipitation estimates.
- For more information see Dee, et al. (2011), The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Q.J.R. Meteorol. Soc., 137: 553-597. <https://doi.org/10.1002/qj.828>)

Summary

- Reanalyses are very useful consistent gridded datasets of historic weather data
- But they have inherent uncertainties and should never be treated as “the truth”
 - They inherit the deficiencies of the modelling system
 - They should not be used for trend analysis due to changes in the observing system
 - They are not conservative
 - Uncertainty metadata is a useful qualitative guide, but is not calculated by metrological techniques