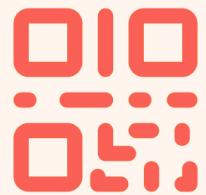




slido



Join at [slido.com](https://www.slido.com)
#3153196

- ① Start presenting to display the joining instructions on this slide.

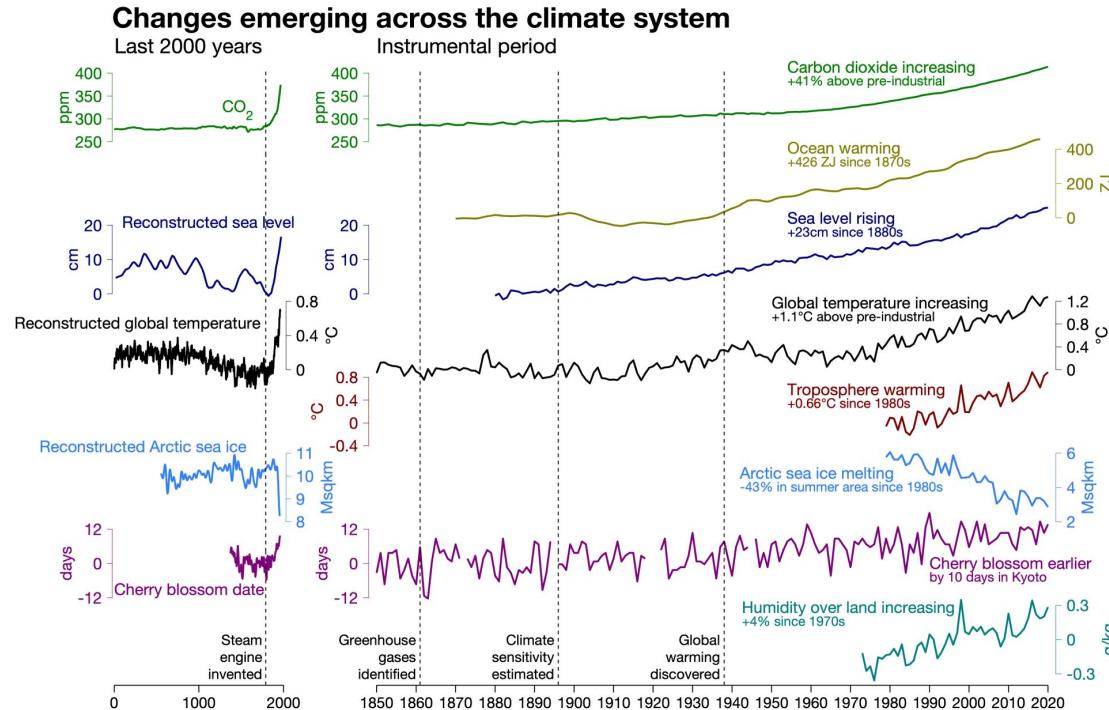


WE
PERSEVERE



Lecture 19

Climate Data Science and XArray



Data 100/Data 200, Spring 2023 @ UC Berkeley

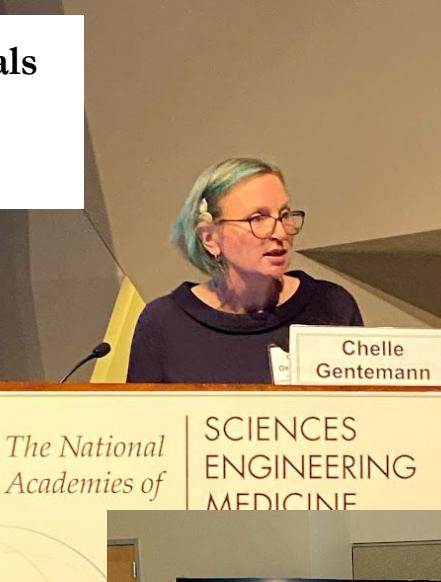
Guest Lecture: Fernando Pérez

Lecture materials by: [Dr. Chelle Gentemann](#), NASA

About the author of these materials

Dr. Chelle Gentemann

More: [@ChelleGentemann](https://twitter.com/ChelleGentemann)



The foundation of climate change: the Greenhouse Effect



who discovered the greenhouse effect



About 20,400,000 results (0.71 seconds)

John Tyndall

Hear this out loud

John Tyndall set the foundation for our modern understanding of the greenhouse effect, climate change, meteorology, and weather. But did he 'discover' it? 160 years ago, on 18 May 1859, the Irish physicist John Tyndall wrote in his journal 'the subject is completely in my hands'. May 17, 2019

Women researchers are cited less than men. Here's why—and what can be done about it

13 OCT 2022 • BY CLAUDIA LOPEZ LLOREDA

Article | Open Access | Published: 22 June 2022

Women are credited less in science than men

Matthew B. Ross, Britta M. Glennon, Raviv Murciano-Goroff, Enrico G. Berkes, Bruce A. Weinberg & Julia I.

Lane

Nature 608, 135–145 (2022) | Cite this article

62k Accesses | 11 Citations | 2335 Altmetric | Metrics



About 13,600,000 results (0.40 seconds)

physicist John Tyndall

Instead of Foote's story, the historical record holds that, in 1859, the Irish physicist John Tyndall discovered the warming effect of the sun's rays on carbon dioxide and water vapor, better known as the greenhouse gas effect. Tyndall made the discovery three years after Foote, and yet he's received the credit. Jul 17, 2019

National Audubon Society
<https://www.audubon.org/news/the-female-s...>

The Female Scientist Who Discovered the Basics of Climate ...

blish with us

382

On the Heat in the Sun's Rays.

ART. XXXI.—*Circumstances affecting the Heat of the Sun's Rays;*
by EUNICE FOOTE.

(Read before the American Association, August 23d, 1856.)

MY investigations have had for their object to determine the different circumstances that affect the thermal action of the rays of light that proceed from the sun.

Several results have been obtained.

First. The action increases with the density of the air, and is diminished as it becomes more rarified.

The experiments were made with an air-pump and two cylindrical receivers of the same size, about four inches in diameter and thirty in length. In each were placed two thermometers, and the air was exhausted from one and condensed in the other. After both had acquired the same temperature they were placed in the sun, side by side, and while the action of the sun's rays rose to 110° in the condensed tube, it attained only 88° in the other. I had no means at hand of measuring the degree of condensation or rarefaction.

The observations taken once in two or three minutes, were as follows:

Exhausted Tube		Condensed Tube.	
In shade.	In sun.	In shade.	In sun.
75	80	75	80
76	82	78	95

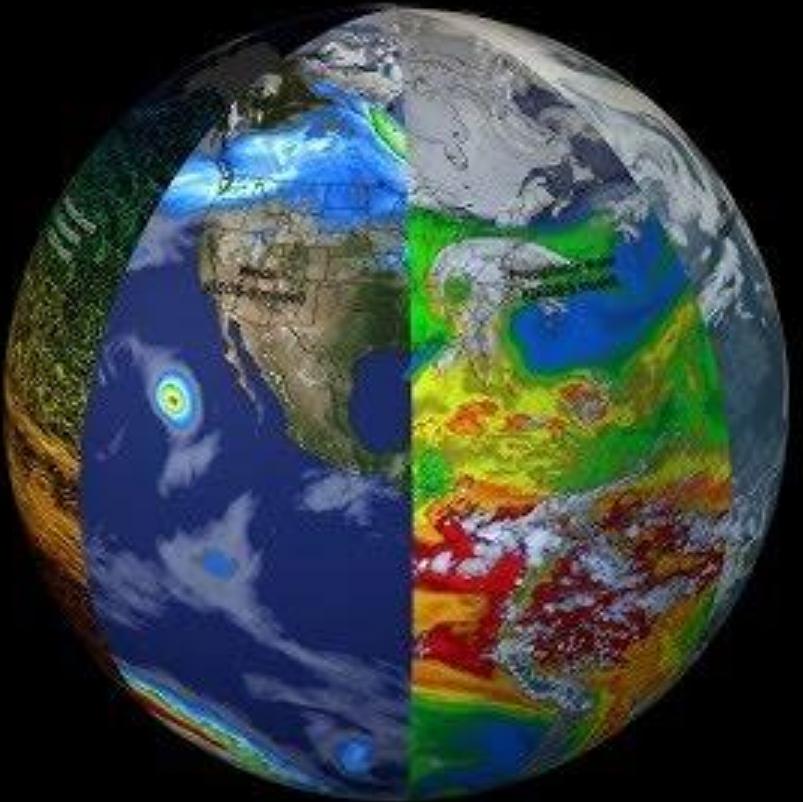


Climate Change

A simplified animation of the greenhouse effect. Credit: NASA/JPL-Caltech

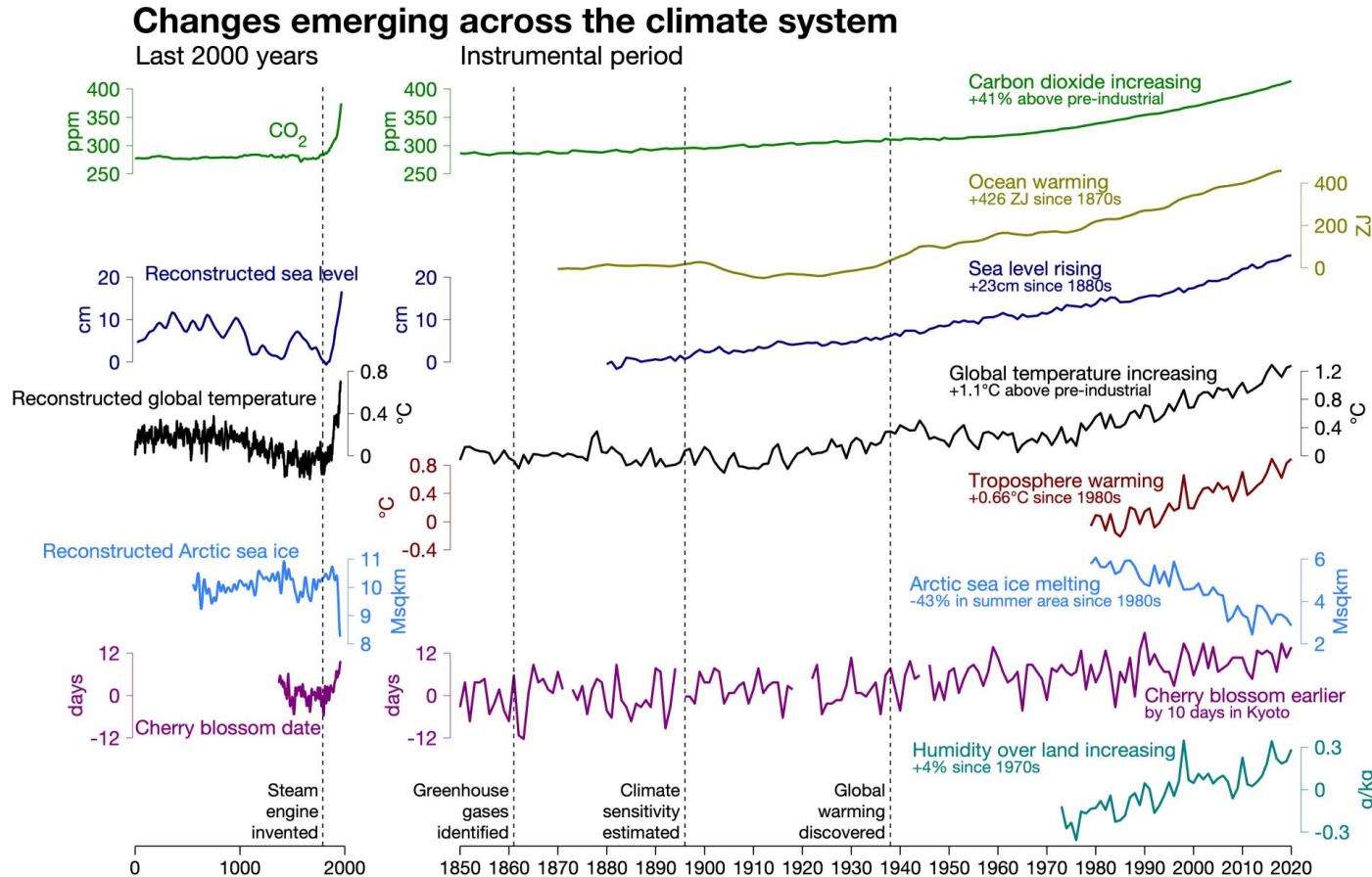


Video credit:
NASA SVS



Video credit:
NASA SVS

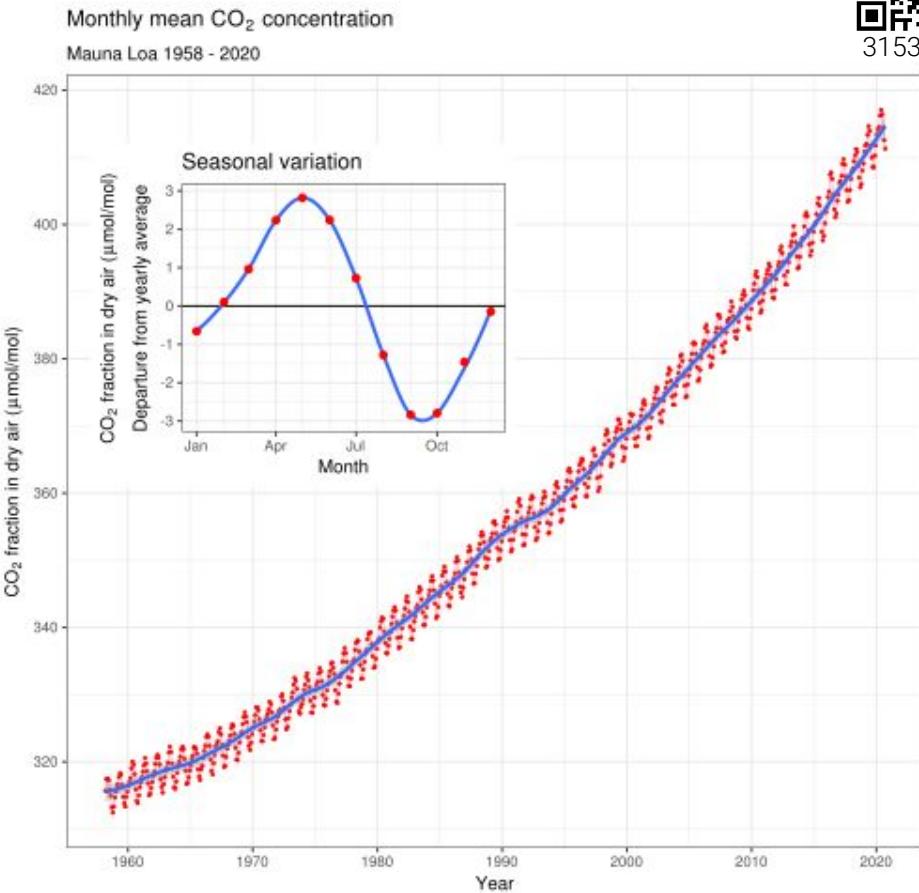
What does the data say about our climate?



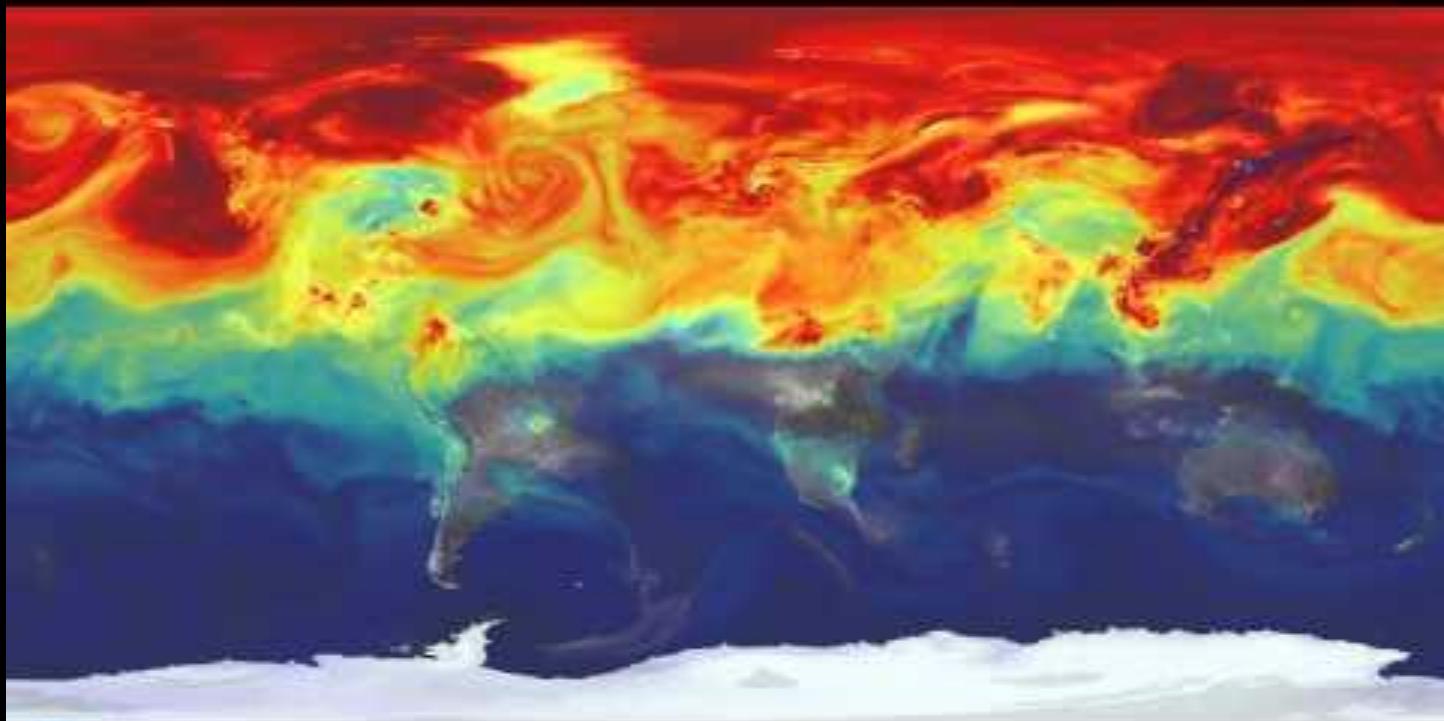
Keeling Curve



- In 1958 Keeling got a grant to begin monitoring CO₂ in Hawaii.
- Roger Revelle (a famous scientist) argued that they just needed a snapshot - CO₂ was too variable - and another snapshot 20 years later to show that CO₂ was increasing
- Keeling advocated for precise measurements over time. By the mid-1960s we had both a measurement of the Earth's breathing and the global increase in CO₂.
- How would you remake this figure?



Data : Dr. Pieter Tans, NOAA/ESRL (www.esrl.noaa.gov/gmd/cogg/trends/) and Dr. Ralph Keeling, Scripps Institution of Oceanography (scrippsc02.ucsd.edu/). Accessed: 2020-10-31



2006 / 05 / 09

Global Modelling and Assimilation Office

Column Abundance (10⁻¹⁰ mol/m²)



Column Depth (ppmv)



Video credit:
NASA SVS

Calculating the Greenhouse Effect



Goal 1: Calculate how the temperature is changing with increasing CO₂.

The Planetary energy balance:

$$\text{Energy (from the Sun) absorbed} = \text{Energy emitted} \quad E_{in} = E_{out}$$

How does CO₂ affects this?

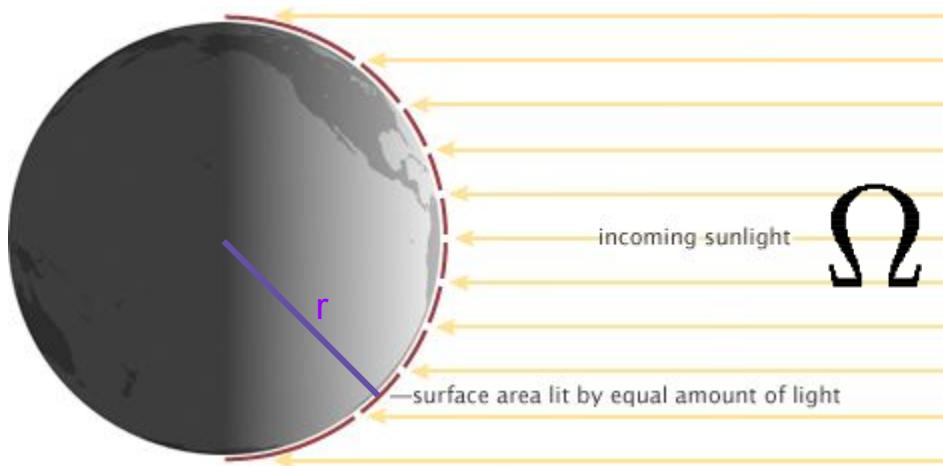
- A) The Sun emits radiation that is absorbed by the Earth (~30% is reflected by clouds, ice/snow, desert, this is the **albedo**)
- B) The Earth emits radiation according to Stephan-Boltzman's Law: the rate that a body emits radiation (per unit area) is directly proportional to the body's absolute temperature to the fourth power (blackbody radiation)
- C) The emitted radiation doesn't all go back into space.....

$$E_{out} = \sigma T^4$$



Calculate energy in

A) The Sun emits radiation Ω absorbed by the Earth



E_{in} Energy in equals the incoming sunlight (W/m^2) multiplied by the area (m^2) to get W

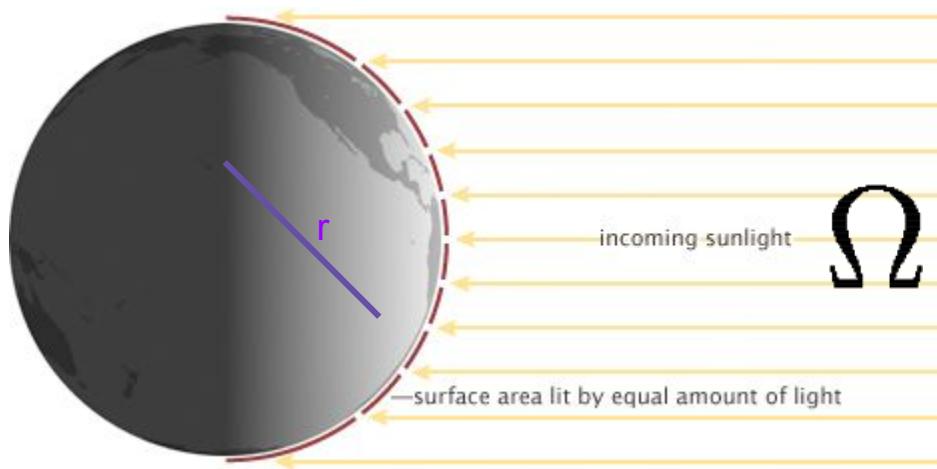
$$E_{in} = \text{Incoming sunlight} \times \text{area}$$

$$E_{in} = \Omega \pi r^2$$

Some of the energy in is reflected by the atmosphere



- A) The Sun emits radiation absorbed by the Earth (some is reflected by the atmosphere)



Albedo = the fraction of radiation reflected back to space by the atmosphere

The amount that gets through:

$$(1 - A)$$

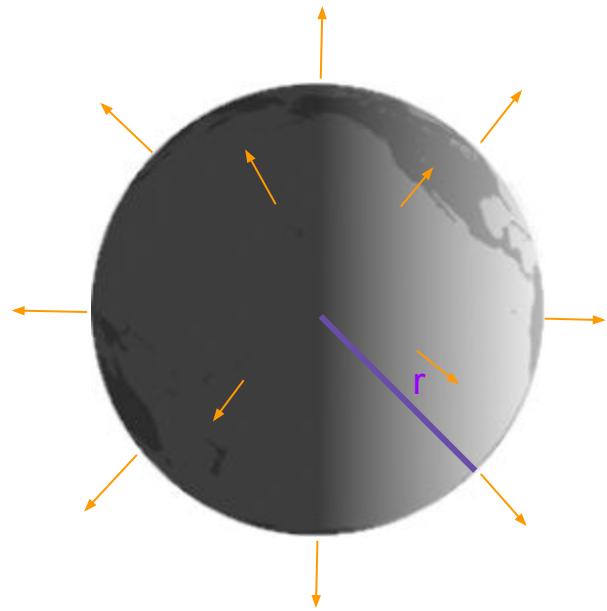
$$E_{in} = \Omega(1 - A)\pi r^2$$

Reflective surface

Calculate Energy out



B) The Earth emits blackbody radiation



$$E_{out}$$

Energy out equals the emitted radiation (W/m^2) multiplied by the area (m^2) to get W

$$E_{out}$$

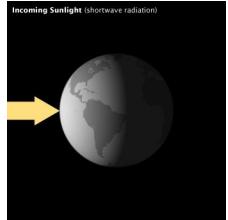
= Outgoing radiation x area

$$E_{out} = \sigma T^4 4\pi r^2$$

Energy in = Energy out

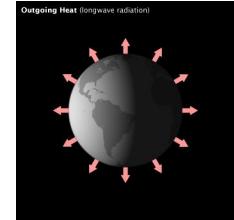


Goal 1: Calculate how the temperature is changing with increasing CO₂



$$E_{in} = \Omega(1 - A)\pi r^2 \quad E_{out} = \sigma T^4 4\pi r^2$$

Reflective surface



$$E_{in} = E_{out} \quad \text{Planetary energy balance}$$

$$\Omega(1 - A)\pi r^2 = \sigma T^4 4\pi r^2$$

$$T = \sqrt[4]{\frac{\Omega(1 - A)}{4\sigma}}$$

Calculating temperature



Goal 1: Calculate how the temperature is changing with increasing CO₂

What is the Earth's temperature?

$$\Omega = 1372 \text{ Wm}^{-2}$$

$$A = 0.3$$

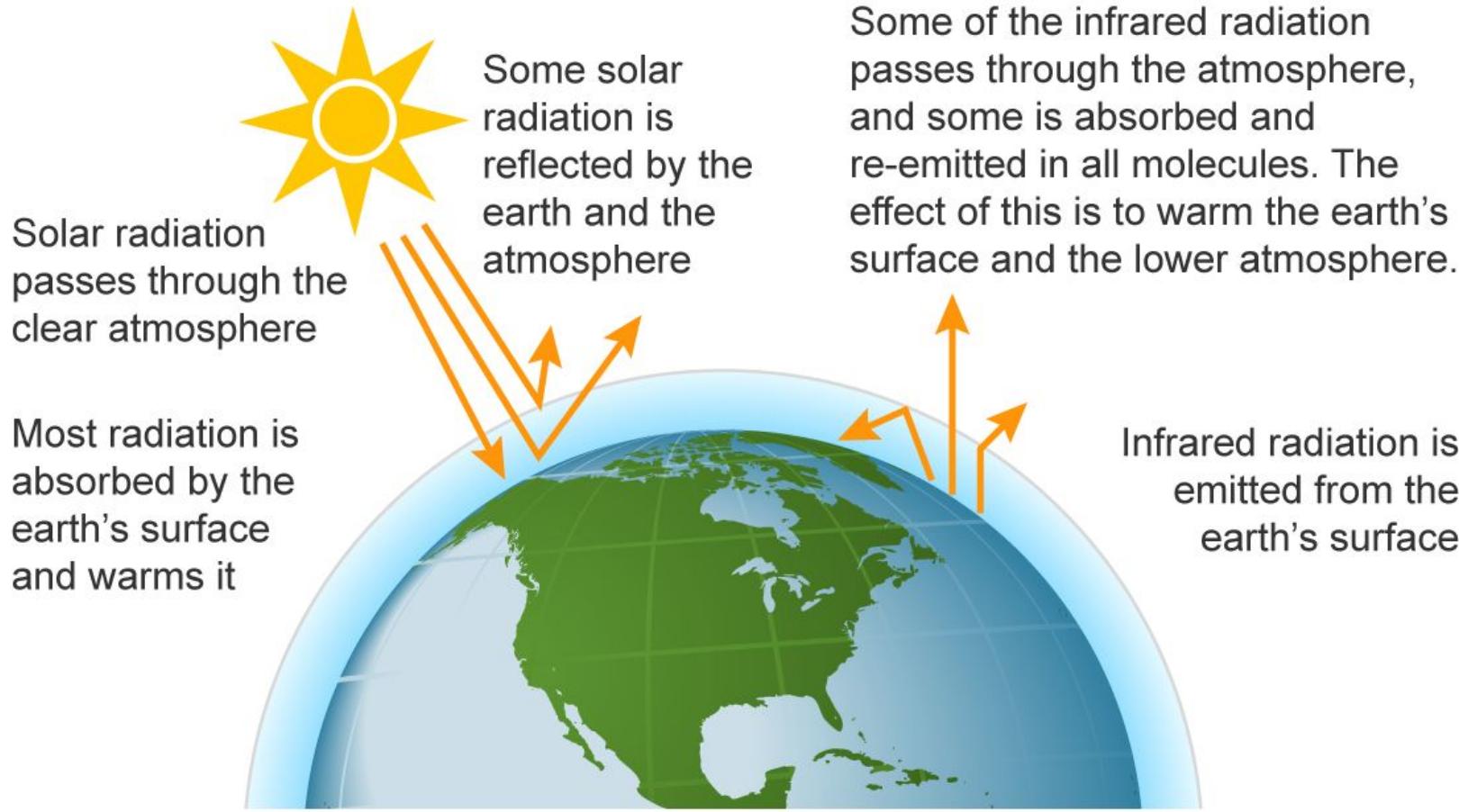
$$\sigma = 5.67 \times 10^{-8} \text{ Wm}^{-2}\text{K}^{-4}$$

$$T = \sqrt[4]{\frac{\Omega(1-A)}{4\sigma}}$$

~255 K ~-16 C ~1 F



The greenhouse effect



Our atmosphere is like a blanket



$$E_{in} = E_{out}$$

Planetary energy balance

$$\Omega(1 - A)\pi r^2 = \sigma T^4 4\pi r^2$$

without an atmosphere

$$\Omega(1 - A)\pi r^2 + \boxed{\Delta E 4\pi r^2} = \sigma T^4 4\pi r^2$$

with an atmosphere

$$\Omega = 1372 \text{ W m}^{-2}$$

$$A = 0.3$$

$$\sigma = 5.67 \times 10^{-8} \text{ W m}^{-2} \text{ K}^{-4}$$

$$T = 288 \text{ K}$$

- Solve for the greenhouse effect!
- What happens to the temperature if we increase the greenhouse effect?
- What happens to the temperature if we decrease/increase the albedo?

Calculating the Temperature dependence on CO₂



Goal 1: Calculate how the temperature is changing with increasing CO₂

CO₂ is ~420 ppm (parts per million - co2.earth/daily-co2)

$$\Omega(1 - A)\pi r^2 + \Delta E 4\pi r^2 = \sigma T^4 4\pi r^2$$

$$\Delta E = 133.26 + 0.044 \times CO_2$$

As CO₂ increases what happens to the temperature?

Plot the results



Use the equation to calculate the increase in temperature with time due to the increase in CO₂

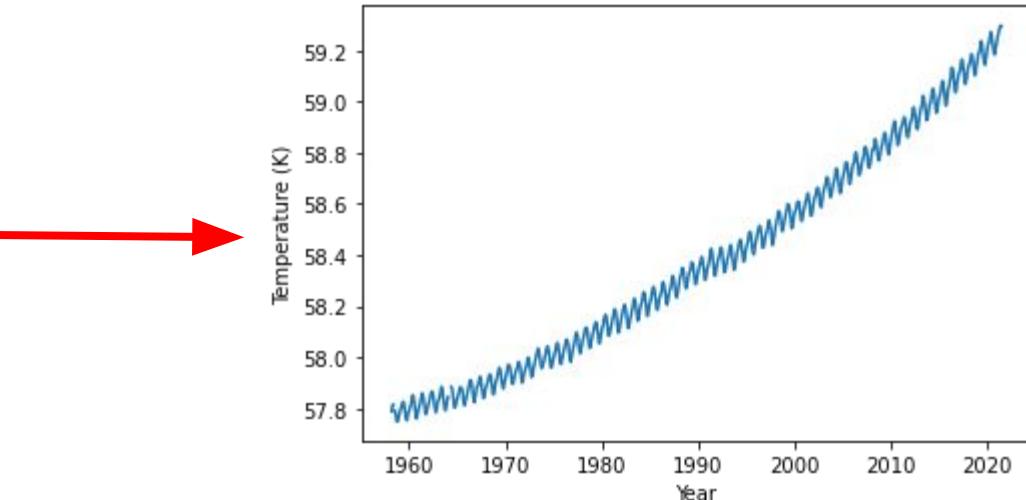
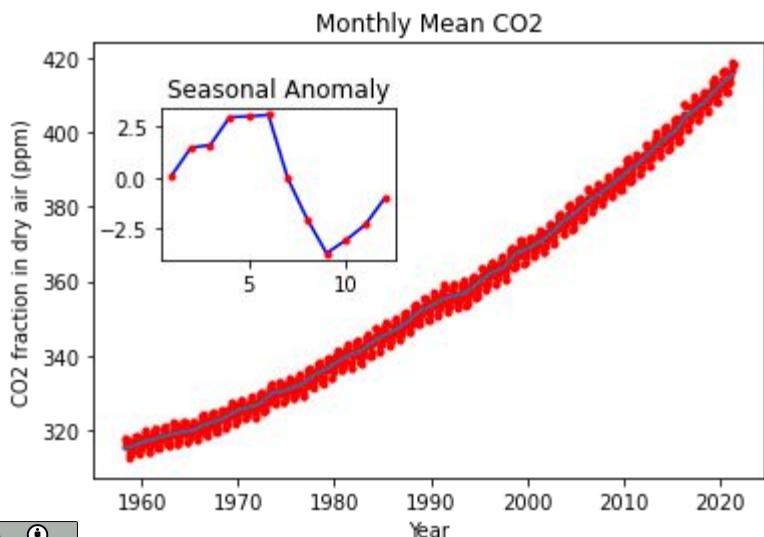
$$\Omega(1 - A)\pi r^2 + \Delta E 4\pi r^2 = \sigma T^4 4\pi r^2$$

$$\Omega = 1372 \text{ Wm}^{-2}$$

$$A = 0.3$$

$$\sigma = 5.67 \times 10^{-8} \text{ Wm}^{-2}\text{K}^{-4}$$

$$\Delta E = 133.26 + 0.044 \times CO_2$$



Global warming is a climate crisis

The equation provides solutions to global warming:

$$\Omega(1 - A)\pi r^2 + \Delta E 4\pi r^2 = \sigma T^4 4\pi r^2$$

Sun-shades in space?

White House is pushing ahead research to cool Earth by reflecting back sunlight

PUBLISHED THU, OCT 13 2022 1:35 PM EDT | UPDATED THU, OCT 13 2022 8:50 PM EDT

Catherine Clifford
@CATCLIFFORD
• CATCLIFFORD

Change the planetary albedo to reflect more radiation to space (increase aerosols, clouds, make surface more reflective)



Real Ice

Real Ice pledges to help Indigenous people obtain re-icing machines that can increase ice thickness and restrict ice melt in Arctic regions. We aim to achieve this by replenishing Arctic ice, using concept tested, wind powered, re-icing machines.

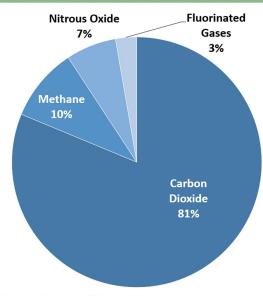
Arctic Ice Project

The most studied ice restoration effort in the world. We are developing a safe, long-term strategy to enhance the Arctic's natural ability to reflect solar radiation out of the atmosphere, increase the Earth's planetary albedo, and slow the rate of global warming.



Reduce greenhouse gases (CO₂, Methane)

Overview of Greenhouse Gas Emissions in 2018



U.S. Environmental Protection Agency (2020). Inventory of U.S. Greenhouse Gas Emissions and Sinks: 1990–2018

Climeworks Turns On the World's Largest Carbon Capture and Storage Plant

The Iceland operation can remove 4,400 tons of CO₂ from the air each year.

By Lloyd Alter | Fact checked by Haley Mast on September 14, 2021 11:34 PM EDT



Published: 25 May 2017

Iron-dumping ocean experiment sparks controversy

Jeff Tollefson

Nature 545, 393–394 (2017) | Cite this article

510 Accesses | 4 Citations | 343 Altmetric | Metrics

CO₂ Capture & Storage

AFFORESTATION AND REFORESTATION

POTENTIAL 1–14 Gt CO₂/year

MATURITY Good to go with opportunities to improve



3153196



3153196

AR6 Climate Change 2021: The Physical Science Basis

Changing by Alisa Singer

"As we witness our planet transforming around us we watch, listen, measure ... respond."

www.environmentalgraphiti.org – 2021 Alisa Singer.



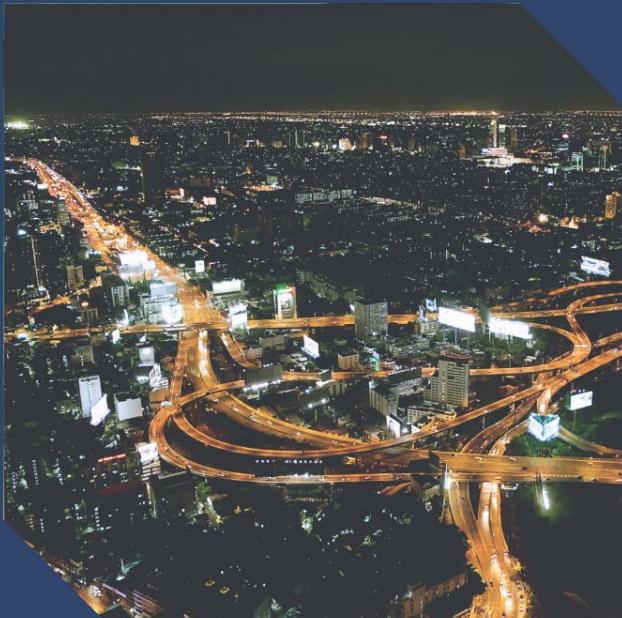
3153196



“Recent changes in the climate are widespread, rapid, and intensifying, and unprecedented in thousands of years.



3153196



[Credit: Peter John Maridabile | Unsplash]

“ Unless there are immediate, rapid, and large-scale reductions in greenhouse gas emissions, limiting warming to 1.5°C will be beyond reach.



3153196



[Credit: Yoda Adaman | Unsplash]

“ It is indisputable that human activities are causing climate change, making extreme climate events, including heat waves, heavy rainfall, and droughts, more frequent and severe.



[Credit: Hong Nguyen | Unsplash]

“ Climate change is already affecting every region on Earth, in multiple ways.

The changes we experience will increase with further warming.



“There’s no going back from some changes in the climate system...

[Credit: Jenn Caselle | UCSB]



3153196



[Credit: Andy Mahoney | NSIDC]

“...However, some changes could be slowed and others could be stopped by limiting warming.

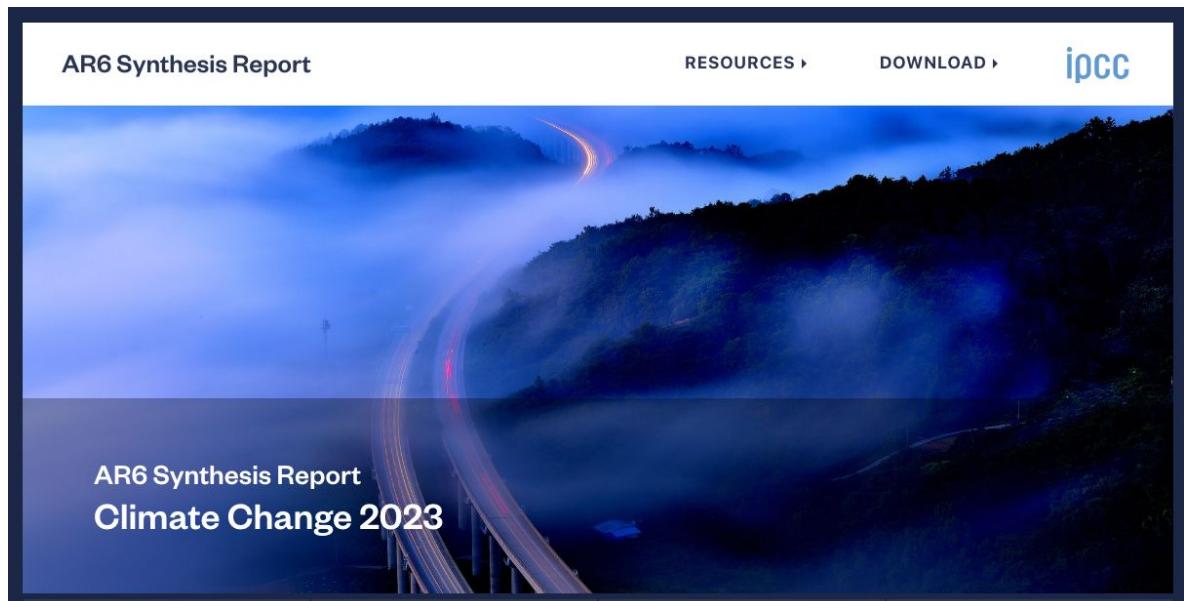


Earth to Hit Critical Warming Threshold by Early 2030s, Climate Panel Says

A new U.N. report says it is still possible to hold global warming to relatively safe levels, but doing so will require global cooperation, billions of dollars and big changes.

[nytimes.com/2023/03/20/climate/global-warming-ipcc-earth.html](https://www.nytimes.com/2023/03/20/climate/global-warming-ipcc-earth.html)

(released yesterday, March 20, 2023)

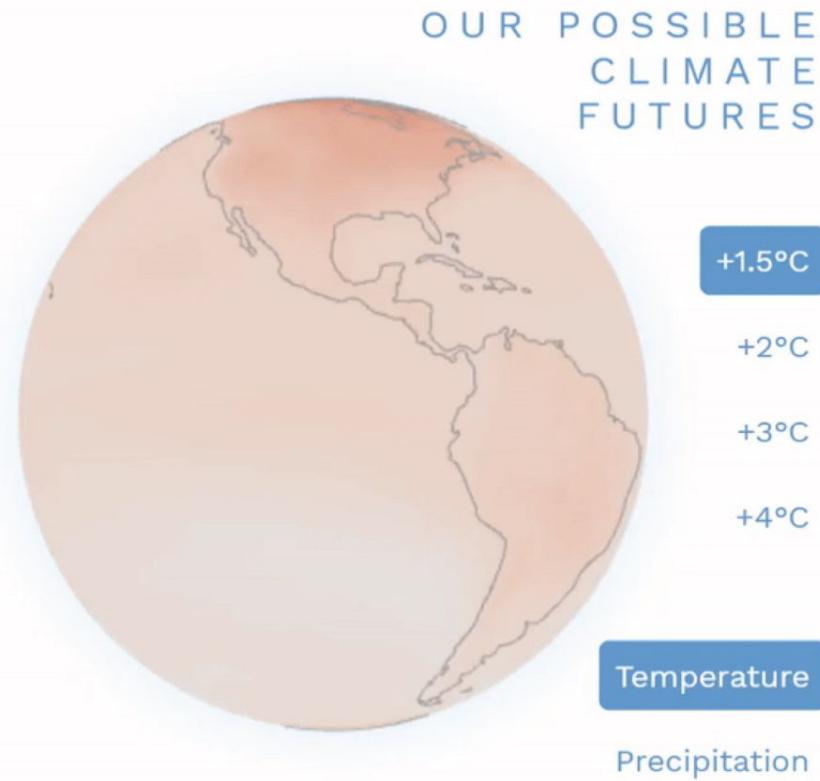


ipcc.ch/report/ar6/syr



3153196

Interactive atlas



<https://interactive-atlas.ipcc.ch/>

#IPCCData

#IPCCAtlas



BY THE NUMBERS

Author Team

234 authors from **65** countries

28% women, **72%** men

30% new to the **IPCC**

Review Process

14,000 scientific publications assessed

78,000+ review comments

46 countries commented on Final Government Distribution



3153196

Let's check our understanding

This equation:

$$\Delta E = 133.26 + 0.044 \times CO_2$$

Shows that all climate change is due to CO₂. If we stop CO₂ emissions, we can stop climate change.

- True
- False



The given equation shows that all climate change is due to CO₂. If we stop CO₂ emissions, we can stop climate change.

- ⓘ Start presenting to display the poll results on this slide.



Image credit: NOAA

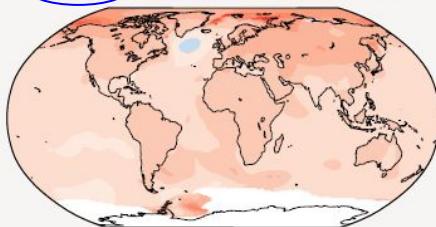
With every increment of global warming, changes get larger



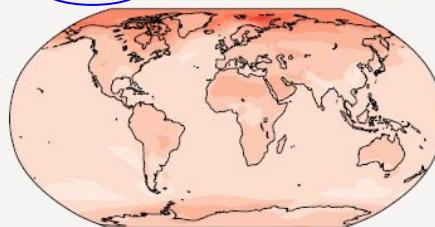
a) Annual mean temperature change ($^{\circ}\text{C}$) at 1 $^{\circ}\text{C}$ global warming

Warming at 1 $^{\circ}\text{C}$ affects all continents and is generally larger over land than over the oceans in both observations and models. Across most regions, observed and simulated patterns are consistent.

Observed change per 1 $^{\circ}\text{C}$ global warming



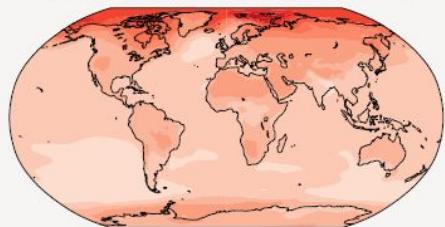
Simulated change at 1 $^{\circ}\text{C}$ global warming



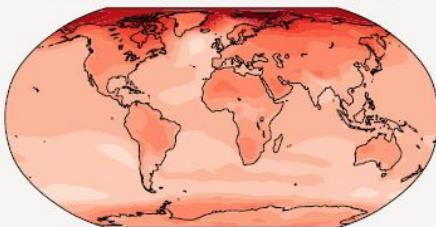
b) Annual mean temperature change ($^{\circ}\text{C}$) relative to 1850-1900

Across warming levels, land areas warm more than oceans, and the Arctic and Antarctica warm more than the tropics.

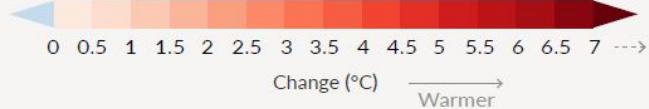
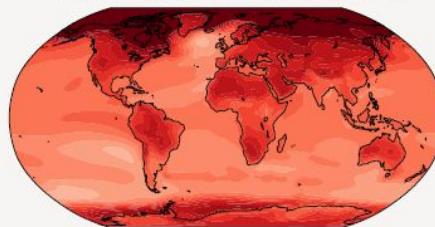
Simulated change at 1.5 $^{\circ}\text{C}$ global warming



Simulated change at 2 $^{\circ}\text{C}$ global warming



Simulated change at 4 $^{\circ}\text{C}$ global warming

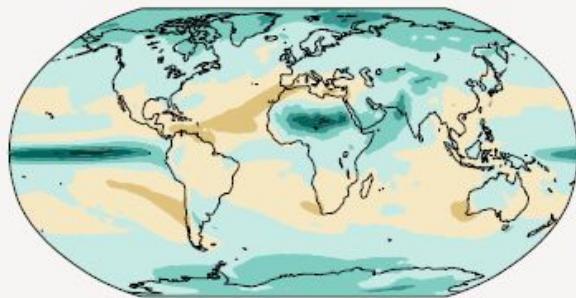


... in precipitation

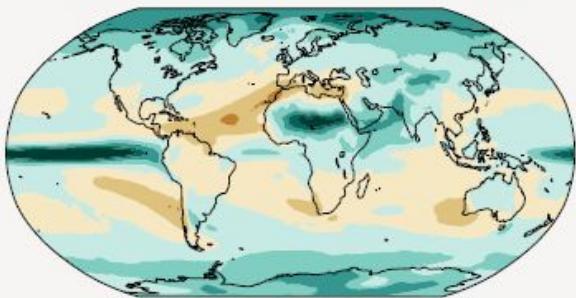


c) Annual mean precipitation change (%) relative to 1850-1900

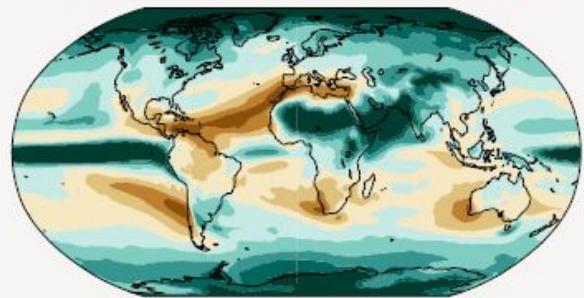
Simulated change at 1.5 °C global warming



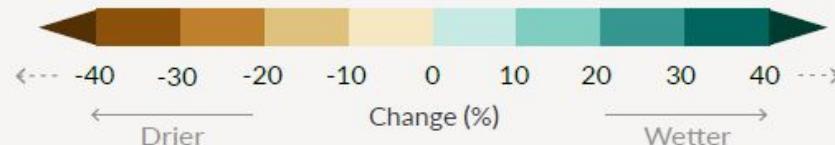
Simulated change at 2 °C global warming



Simulated change at 4 °C global warming



Relatively small absolute changes may appear as large % changes in regions with dry baseline conditions



CC-relationship water vapor - temperature - pressure

<https://www.jbarisk.com/news-blogs/the-physics-of-precipitation-in-a-warming-climate/>

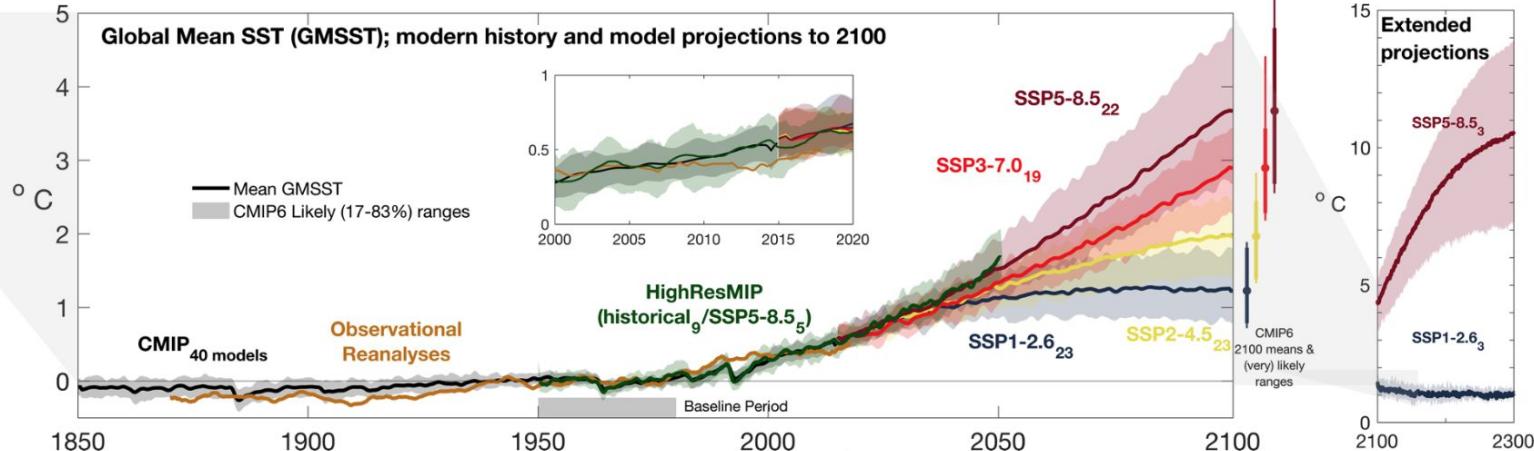
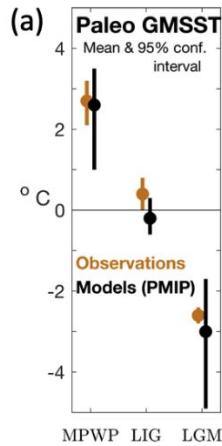
<https://www.ipcc.ch/report/ar6/wg1/#FullReport>

Our trajectory from data and models



Sea Surface Temperature (SST) Anomalies and Maps

Observation-based estimates and CMIP6 multi-model means, biases and projected changes



https://raw.githubusercontent.com/BrodiePearson/IPCC_AR6_Chapter9_Figures/main/Plotting_code_and_data/Fig9_03_SST/Fig9_03_SST.png



Extremes are the new normal

ENVIRONMENT | NEWS

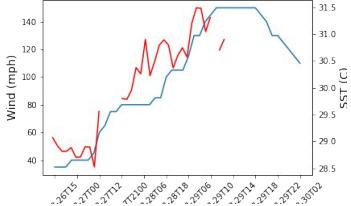
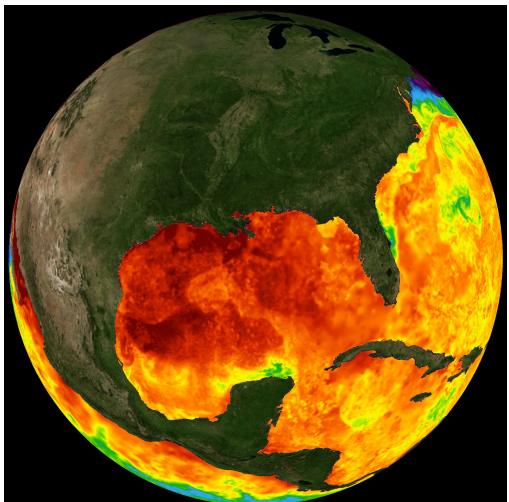
How climate change is fueling hurricanes like Ida

Hurricanes feed off heat, a growing source of fuel in a warming world.

BY SARAH GIBBENS



PUBLISHED AUGUST 31, 2021 • 8 MIN READ



The Creek Fire, in the Sierra National Forest in California, has burned hundreds of thousands of acres. Its spread was fueled by the presence of many dead, super dry trees; climate change contributed to both their death and their dryness.
PHOTOGRAPH BY STUART PALLEY, NATIONAL GEOGRAPHIC

SCIENCE | NEWS

The science connecting wildfires to climate change

A heating-up planet has driven huge increases in wildfire area burned over the past few decades.

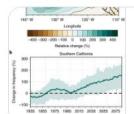
Climate crisis likely creating extreme winter weather events, says report

Arctic change increased chances of tightly spinning winds above North Pole, authors say, boosting chances of extreme weather



 Daniel Swain 
@Weather_West

It is worth noting that this exact situation--an extremely strong atmospheric river bringing brief period of record rainfall in midst of severe and temperature-amplified drought--is what we expect to see in California with #ClimateChange. #CAwater
#CAwater



nature.com
Increasing precipitation volatility in...
Nature Climate Change - California recently experienced a rapid shift from multi-year drought ...

6:08 PM · Oct 24, 2021 · Twitter Web App



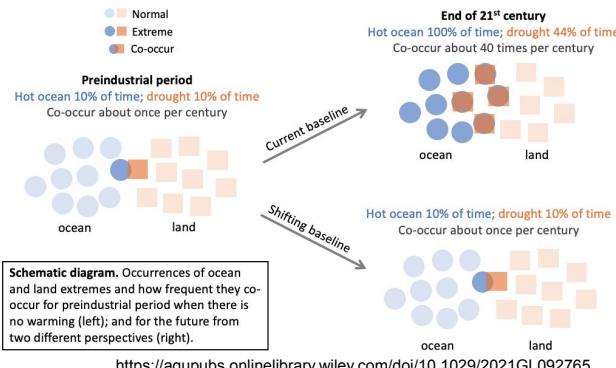
Extremes are the new normal

A decade ago - scientists would argue - we can't attribute any single weather event to climate.

In the last decade, we have all experienced major shifts in our climate through changes in our local weather and scientists have figured out 'climate event attribution'

They look at the probability of the occurrence of an event (eg. a temperature extreme) in models run without human-influence and then compare it to the probability in models run with human-influence.

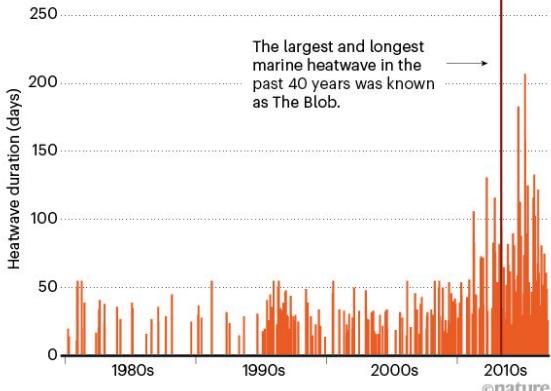
How?



<https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2021GL092765>

FEVERED WATERS

A plot of the 300 largest marine heatwaves between 1981 and 2017 shows that such events are hitting more frequently and lingering for longer.



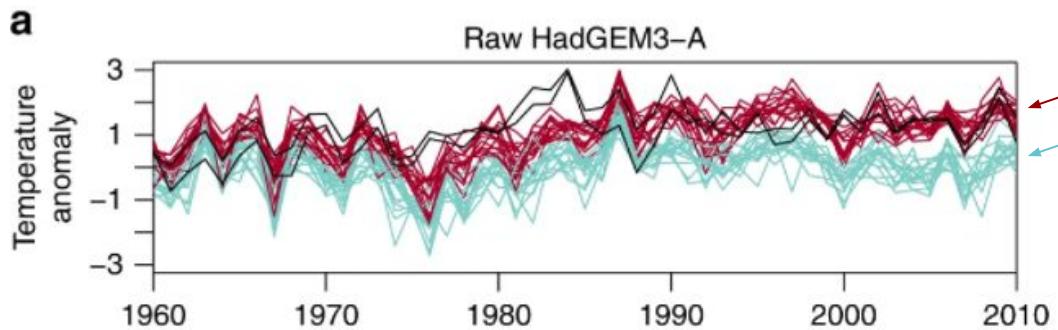
<https://www.nature.com/articles/d41586-021-01142-4>



Extremes are the new normal

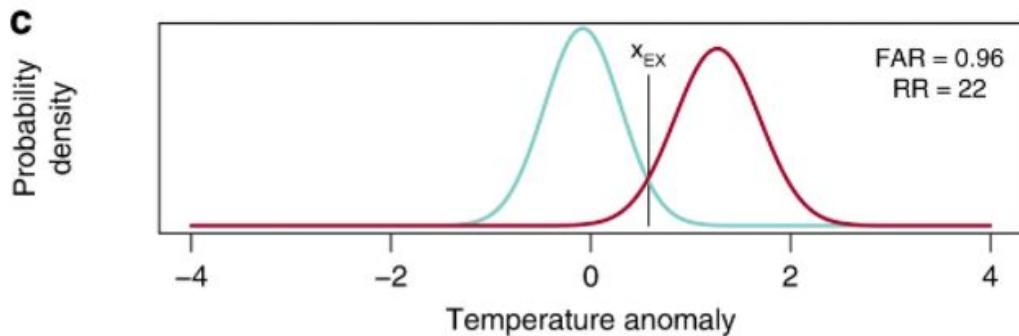


The probability of an event changes



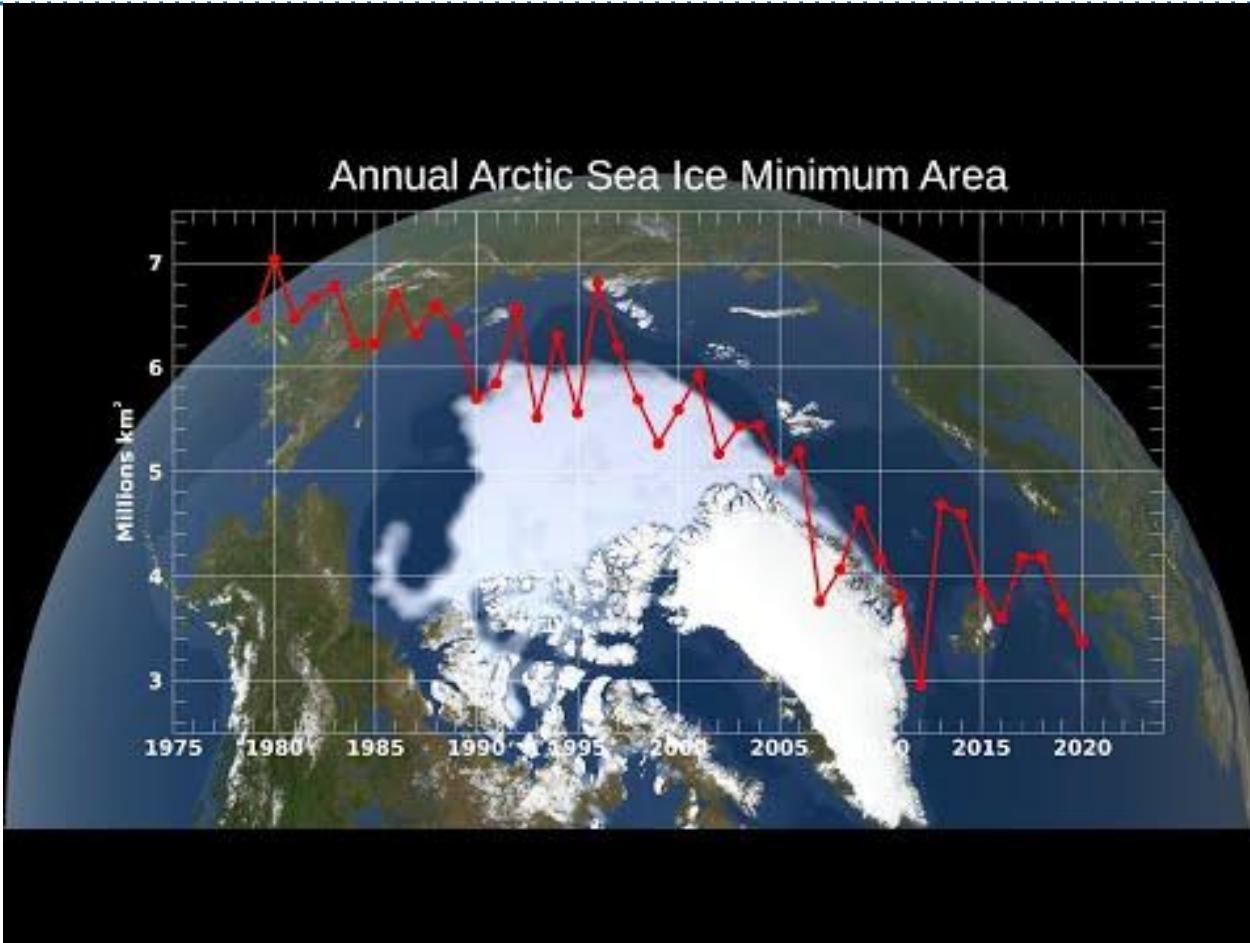
Model runs with all forcings

Model runs with only natural



How likely is it that there will be a 2 degree anomaly?

Positive Feedbacks in the Climate System

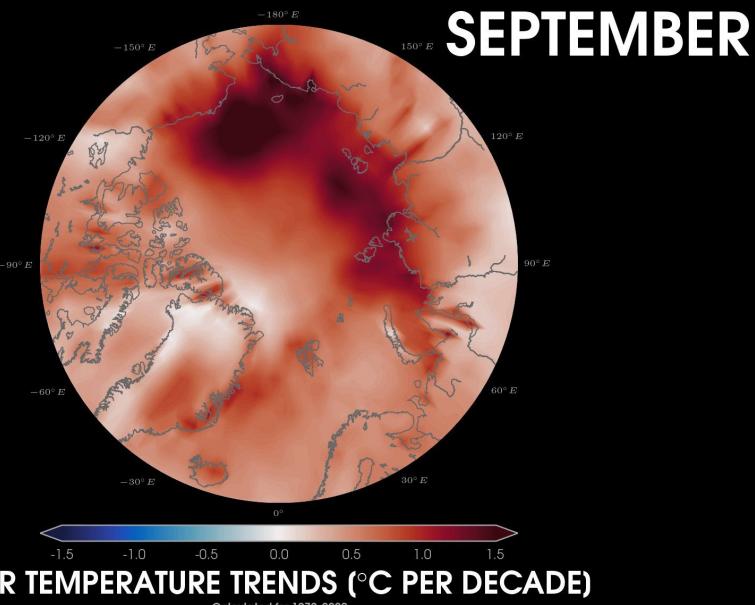


Goal: Positive feedbacks in the climate system

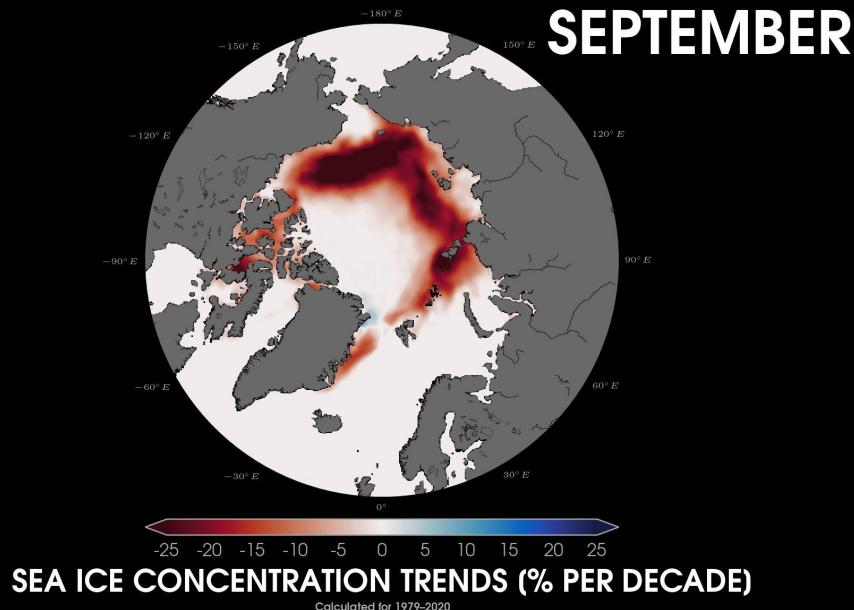


The Earth is a system - a complicated game of dominos

As air temperature increases, sea ice concentration decreases.....



GRAPHIC: Zachary Labe (@ZLabe)
SOURCE: <https://climate.copernicus.eu/>
DATA: Copernicus Climate Change Service (ECMWF ERA5 - 2m T))



GRAPHIC: Zachary Labe (@ZLabe)
SOURCE: <https://nsidc.org/>
DATA: NOAA/NSIDC CDR of Passive Microwave Sea Ice Concentration v4 (1979-2020)

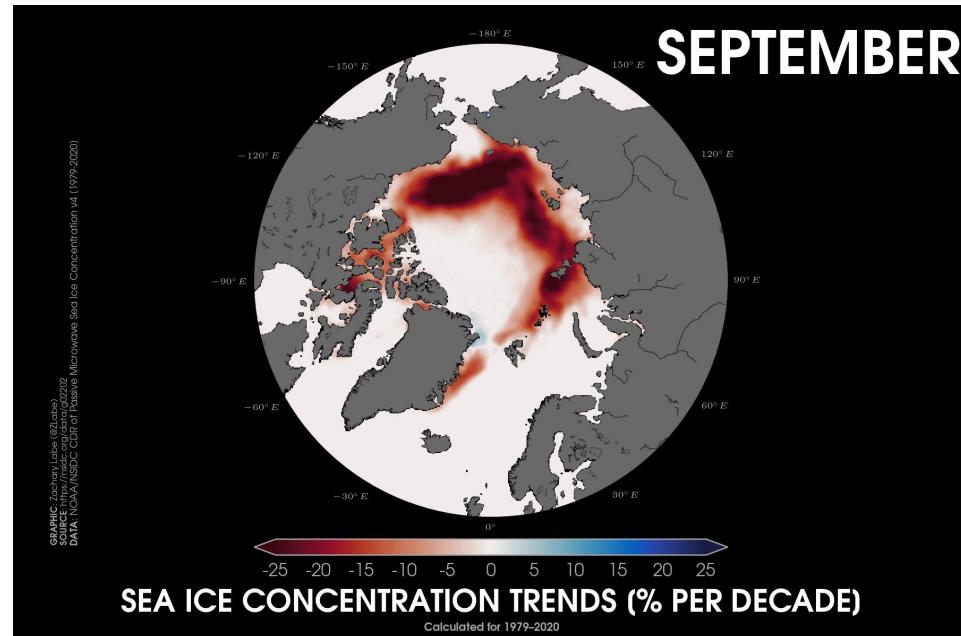
Goal: Arctic Amplification: a positive feedback

A positive feedback amplifies the initial perturbation

reflectivity!

Albedo feedback:

1. The albedo of sea ice is ~0.5-0.7.
Most sunlight is reflected back to space.
2. The albedo of the ocean is 0.04.
Most sunlight is absorbed and warms the seawater.
3. A warmer ocean melts more sea ice.



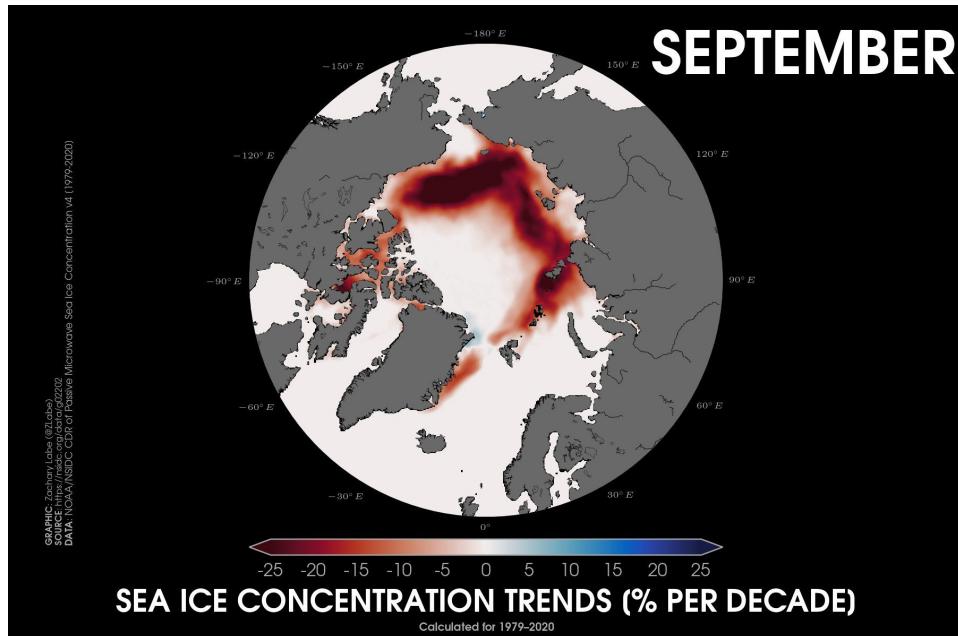
Goal: Arctic Amplification: a positive feedback



A positive feedback amplifies the initial perturbation

Cloud feedback:

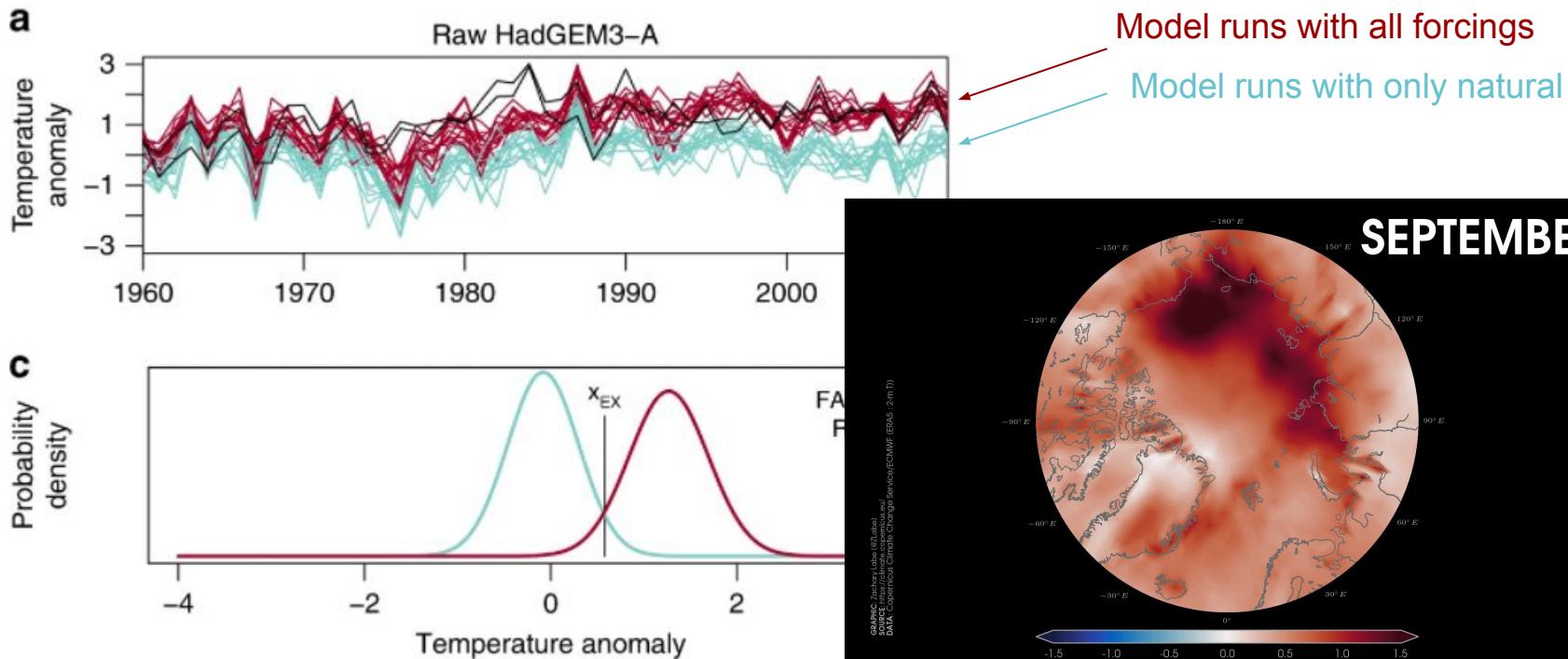
1. As sea ice melts, more ocean is exposed and more moisture is absorbed by the atmosphere
2. More moisture == more clouds
3. Clouds trap longwave radiation, warming the Arctic, melting more sea ice.



Can we look at this ourselves? --- this is global data --- 3D data



..... Latitude, longitude, and time..... How do we handle that?



Learning objective: climate data analysis

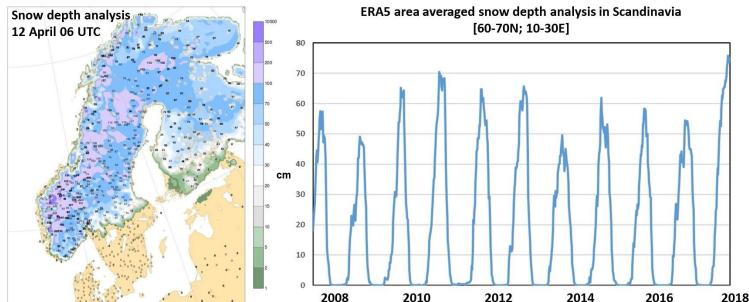


- 3D data (time,latitude,longitude)
- Introduction to Xarray python library
- Xarray Probability density functions
- Xarray linear regression: calculating mean trends
- Xarray global analysis of trends

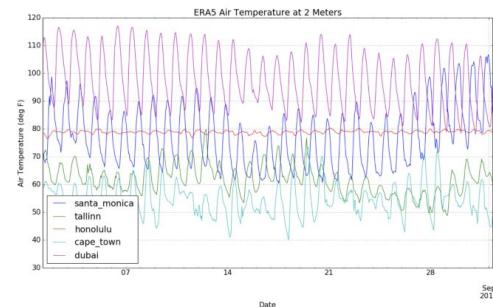
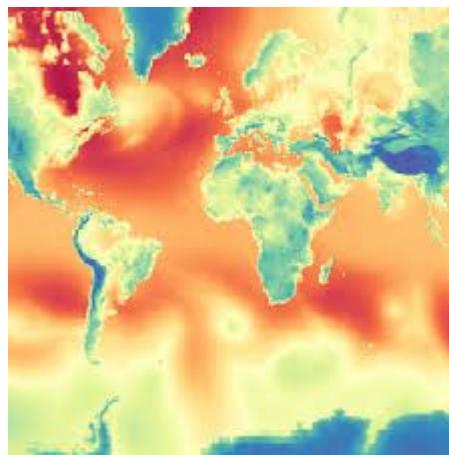
ERA5 - 5th gen ECMWF atmospheric global climate ReAnalyses



- ERA5 combines vast amounts of historical observations into global estimates using advanced modelling and data assimilation systems.
- From 1979 - 2019, hourly estimates of atmospheric, land and oceanic climate variables.
- 30 km global grid, with 137 levels from the surface up to a height of 80 km.



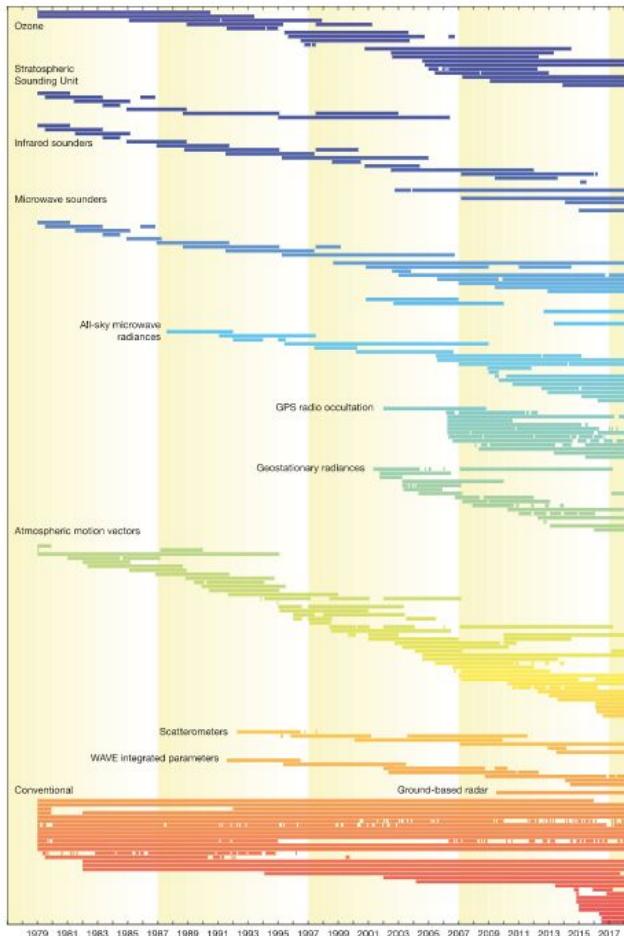
An [ECMWF snow depth analysis](#) for Scandinavia using ERA5 data shows the highest levels in a decade.



<https://medium.com/planet-os/era5-re-analysis-data-on-s3-cee2f2242ae>

Vast amount of data

- The number of observations assimilated in ERA5 has increased from approximately 0.75 million per day on average in 1979 to around 24 million per day by the end of 2018
- ~14,000 GB (14 TB)
- A key dataset used for understanding our weather and climate, but inaccessible to all but a few privileged institutions



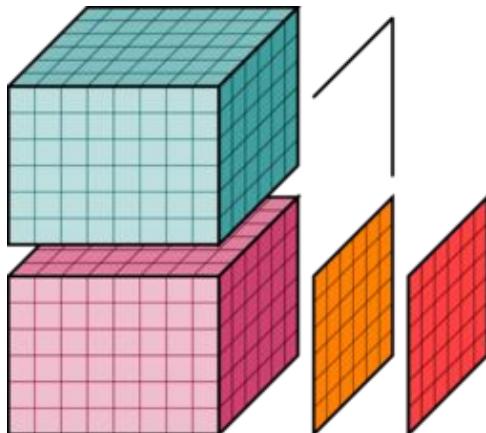
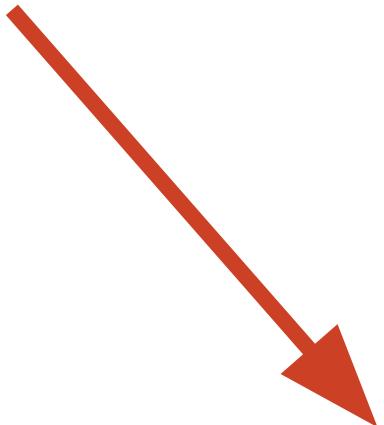
PANGE \ominus

A community platform for Big Data geoscience

[contributors](#) 71 · [discourse](#) 783 users · [chat on gitter](#) · [follow @pangeo_data](#) 3.9k

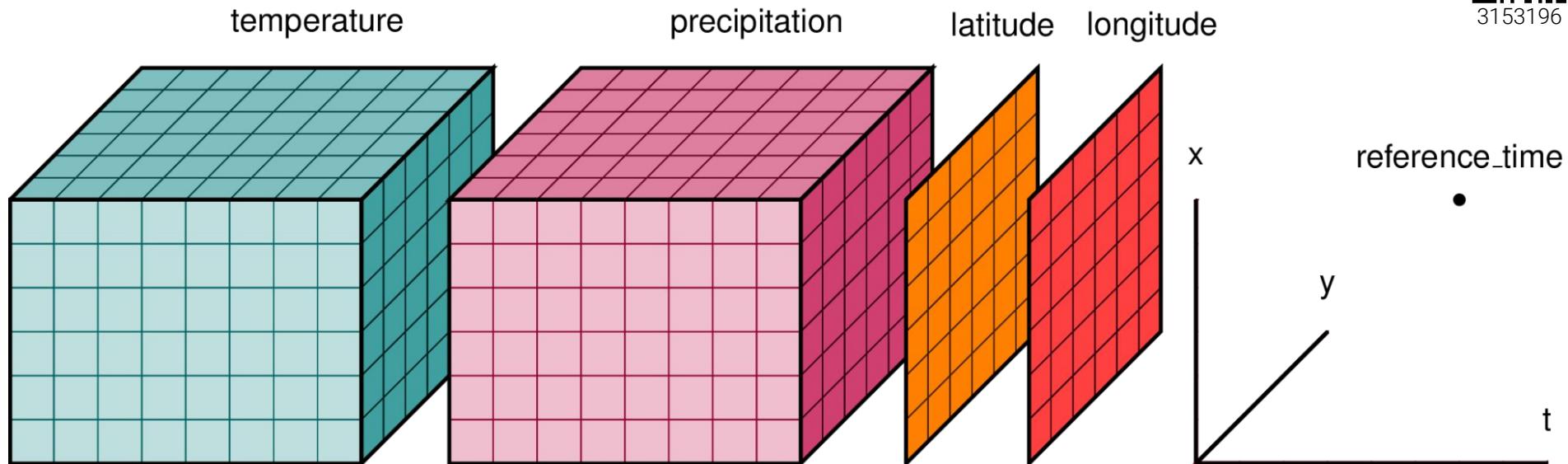
This website contains general information about the Pangeo project. For news and updates about Pangeo, check out our [Medium blog](#) and our [Twitter feed](#). To engage with the Pangeo community, head over to our [Discourse forum](#) or browse our [GitHub repos](#).

Data Frames are not enough: not all data is tabular



xarray

Xarray



'data variables' : temperature and precipitation “data variables”
'coordinate variables': they label the points along the dimensions.

Xarray introduces labels in the form of dimensions, coordinates and attributes on top of raw NumPy-like multidimensional arrays, which allows for a more intuitive, more concise, and less error-prone developer experience.

Xarray: Dataset



Xarray Datasets are essentially groups of DataArrays.

This is **really** valuable when you are looking at datasets that have multidimensional groups of data, for example, temperature, precipitation, cloud cover.

Some Xarray methods can be applied to all that Dataset contains.

For example, you can subset a Dataset and the subset, interpolate, calculate a mean, and it will do this across all the DataArrays the Dataset contains.

ERA5

xarray.Dataset

Dimensions: (time: 504, latitude: 90, longitude: 180)

Coordinates:

time	(time)	datetime64[ns]	1979-01-16T11:30:00 ... 2020-12-...		
latitude	(latitude)	float32	-88.88 -86.88 ... 87.12 89.12		
longitude	(longitude)	float32	0.875 2.875 4.875 ... 356.9 358.9		

Data variables:

air_pressure_at_...	(time, latitude, longitude)	float32	...		
air_temperature_...	(time, latitude, longitude)	float32	...		
air_temperature_...	(time, latitude, longitude)	float32	...		
air_temperature_...	(time, latitude, longitude)	float32	...		
dew_point_temp...	(time, latitude, longitude)	float32	...		
eastward_wind_a...	(time, latitude, longitude)	float32	...		
eastward_wind_a...	(time, latitude, longitude)	float32	...		
integral_wrt_tim...	(time, latitude, longitude)	float32	...		
lwe_thickness_of...	(time, latitude, longitude)	float32	...		
northward_wind...	(time, latitude, longitude)	float32	...		
northward_wind...	(time, latitude, longitude)	float32	...		
precipitation_am...	(time, latitude, longitude)	float32	...		
sea_surface_tem...	(time, latitude, longitude)	float32	...		
snow_density	(time, latitude, longitude)	float32	...		
surface_air_press...	(time, latitude, longitude)	float32	...		

Attributes:

institution :	ECMWF
source :	Reanalysis
title :	ERA5 forecasts

Xarray read in a Dataset



```
import xarray as xr  
ds = xr.open_dataset('./../data/era5_monthly_2deg_aws_v20210920.nc')  
ds
```

xarray.Dataset

Dimensions: (time: 504, latitude: 90, longitude: 180)

Coordinates:

time	(time)	datetime64[ns]	1979-01-16T11:30:00 ... 2020-12-		
latitude	(latitude)	float32	-88.88 -86.88 ... 87.12 89.12		
longitude	(longitude)	float32	0.875 2.875 4.875 ... 356.9 358.9		

Data variables:

air_pressure_at...	(time, latitude, longitude)	float32	...		
air_temperature...	(time, latitude, longitude)	float32	...		
air_temperature...	(time, latitude, longitude)	float32	...		
air_temperature...	(time, latitude, longitude)	float32	...		
dew_point_temp...	(time, latitude, longitude)	float32	...		
eastward_wind_a...	(time, latitude, longitude)	float32	...		
eastward_wind_a...	(time, latitude, longitude)	float32	...		
integral_wrt_tim...	(time, latitude, longitude)	float32	...		
lwe_thickness_of...	(time, latitude, longitude)	float32	...		
northward_wind...	(time, latitude, longitude)	float32	...		
northward_wind...	(time, latitude, longitude)	float32	...		
precipitation_am...	(time, latitude, longitude)	float32	...		
sea_surface_tem...	(time, latitude, longitude)	float32	...		
snow_density	(time, latitude, longitude)	float32	...		
surface_air_press...	(time, latitude, longitude)	float32	...		

Attributes:

institution :	ECMWF
source :	Reanalysis
title :	ERA5 forecasts

- 3D data
- Dimensions: 504x90x180
- Coordinates: time,latitude,longitude
- 15 data variables (in DataArrays)
- Attributes

Explore the data



Xarray allows you easily explore the data

xarray.Dataset

Dimensions: (time: 504, latitude: 90, longitude: 180)

Coordinates:

time	(time)	datetime64[ns]	1979-01-16T11:30:00 ... 2020-12-...		
latitude	(latitude)	float32	-88.88 -86.88 ... 87.12 89.12		
longitude	(longitude)	float32	0.875 2.875 4.875 ... 356.9 358.9		

Data variables:

air_pressure_at...	(time, latitude, longitude)	float32	...		
air_temperature...	(time, latitude, longitude)	float32	...		
air_temperature...	(time, latitude, longitude)	float32	...		
air_temperature...	(time, latitude, longitude)	float32	...		
dew_point_temp...	(time, latitude, longitude)	float32	...		
eastward_wind_a...	(time, latitude, longitude)	float32	...		
eastward_wind_a...	(time, latitude, longitude)	float32	...		
integral_wrt_tim...	(time, latitude, longitude)	float32	...		
lwe_thickness_of...	(time, latitude, longitude)	float32	...		
northward_wind...	(time, latitude, longitude)	float32	...		
northward_wind...	(time, latitude, longitude)	float32	...		
precipitation_am...	(time, latitude, longitude)	float32	...		
sea_surface_tem...	(time, latitude, longitude)	float32	...		
snow_density	(time, latitude, longitude)	float32	...		
surface_air_press...	(time, latitude, longitude)	float32	...		

Attributes:

institution :	ECMWF
source :	Reanalysis
title :	ERA5 forecasts

Look at data attributes

xarray.Dataset

Dimensions: (time: 504, latitude: 90, longitude: 180)

Coordinates:

time	(time)	datetime64[ns]	1979-01-16T11:30:00 ... 2020-12-...		
latitude	(latitude)	float32	-88.88 -86.88 ... 87.12 89.12		
longitude	(longitude)	float32	0.875 2.875 4.875 ... 356.9 358.9		

Data variables:

air_pressure_at...	(time, latitude, longitude)	float32	...		
air_temperature...	(time, latitude, longitude)	float32	...		

long_name : 2 metre temperature
nameCDM : 2_metre_temperature_surface
nameECMWF : 2 metre temperature
product_type : analysis
shortNameECM... : 2t
standard_name : air_temperature
units : K

air_temperature...	(time, latitude, longitude)	float32	...		
air_temperature...	(time, latitude, longitude)	float32	...		

Coordinates versus dimensions



- DataArray objects inside a Dataset may have any number of dimensions but are presumed to share a common coordinate system.
- Coordinates can also have any number of dimensions but denote constant/independent quantities, unlike the varying/dependent quantities that belong in data
- A dimension is just a name of an axis, like ‘time’

```
ds.dims
```

```
Frozen({'time': 504, 'latitude': 90, 'longitude': 180})
```

```
ds.coords
```

Coordinates:

```
* time      (time) datetime64[ns] 1979-01-16T11:30:00 ... 2020-12-16T11:30:00
* latitude   (latitude) float32 -88.88 -86.88 -84.88 ... 85.12 87.12 89.12
* longitude  (longitude) float32 0.875 2.875 4.875 6.875 ... 354.9 356.9 358.9
```

DataArray are data variables in a Dataset



- A dataArray holds a multi-dimensional information
- dataArray objects inside a dataset may have any number of dimensions but are presumed to share a common coordinate system.
- You can explore the data easily using either syntax

```
ds["air_temperature_at_2_metres"]
```

xarray.DataArray 'air_temperature_at_2_metres' (time: 504, latitude: 90, longitude: 180)

[8164800 values with dtype=float32]

▼ Coordinates:

time	(time)	datetime64[ns] 1979-01-16T11:30:00 ... 2020-12-...	
latitude	(latitude)	float32 -88.88 -86.88 ... 87.12 89.12	
longitude	(longitude)	float32 0.875 2.875 4.875 ... 356.9 358.9	

▼ Attributes:

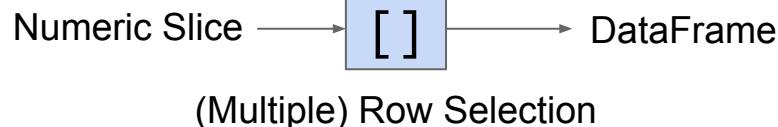
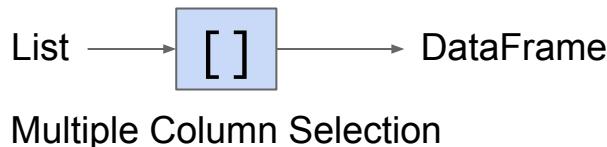
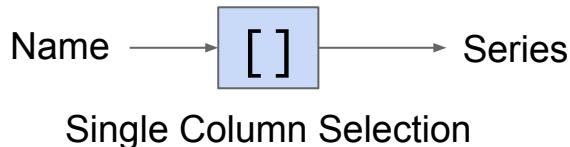
long_name : 2 metre temperature
nameCDM : 2_metre_temperature_surface
nameECMWF : 2 metre temperature
product_type : analysis
shortNameECM... : 2t
standard_name : air_temperature
units : K

```
ds.air_temperature_at_2_metres
```

Review: DataFrame access: [], loc, iloc



[:]: flexible, confusing?



iloc: integer/positional

- Always 0-based, for rows and columns.
- Slices as usual, end-point exclusive.
- Use carefully (error prone).

loc: Labels

- Strings, integers - row/column labels
- Lists - similar, but always return dataframes
- Slices of labels: **end-point inclusive!**
- Boolean arrays: “mask” selection.

New: DataArray access: [], sel, isel



[]: flexible, confusing?

- Only for DataArrays

```
point = ds.air_temperature_at_2_metres[0, 26, 119]
```

isel: integer/positional

- Always 0-based
- Slices as usual, **end-point exclusive**.
- Use carefully (error prone).

```
point = ds.isel(time=0,  
                 latitude=26,  
                 longitude=119)
```

sel: coordinates

- Strings, integers - coordinates*
- Slices: **end-point inclusive!**

```
point = ds.sel(time="1979-01",  
               latitude=37.125,  
               longitude=238.875)
```

▼ Coordinates:

time	(time)	datetime64[ns]	1979-01-16T11:30:00 ... 2020-12-		
latitude	(latitude)	float32	-88.88 -86.88 ... 87.12 89.12		
longitude	(longitude)	float32	0.875 2.875 4.875 ... 356.9 358.9		

Find data at a point or in a region:

```
point = ds.sel(time="1979-01",  
               latitude=37.125,  
               longitude=238.875)
```

```
region = ds.sel(time="1979-01",  
                 latitude=slice(30,40),  
                 longitude=slice(230,250))
```

sel: coordinates

- Strings, integers - coordinates*
- Slices of coordinates: **end-point inclusive!**

Xarray helps you understand your code.



Xarray has all sort of high-level cool tricks built in



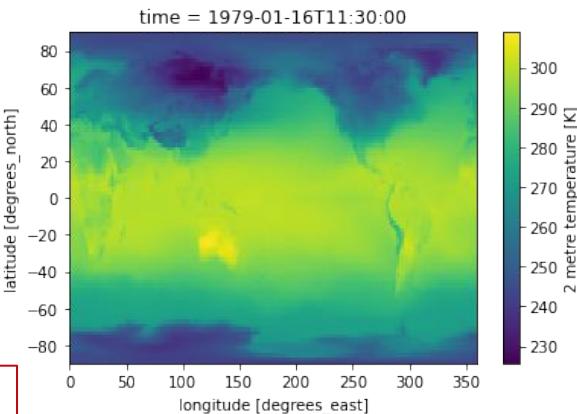
Select data and plot data in one line (using matplotlib)

```
ds.air_temperature_at_2_metres.sel(time="1979-01").plot()
```

Select a variable

Select coordinate

Apply a method()

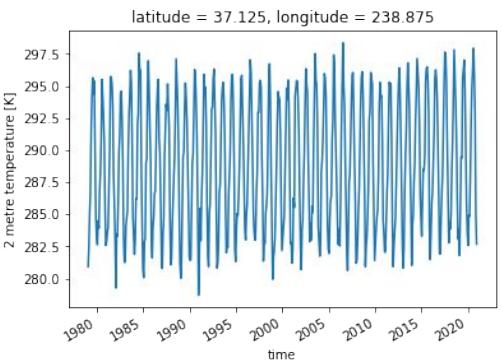


```
ds.air_temperature_at_2_metres.sel(  
    latitude=37.125,  
    longitude=238.875).plot()
```

Select a variable

Select coordinate

Apply a method()



Methods can be called across a dataArray or a Dataset -- LAZY



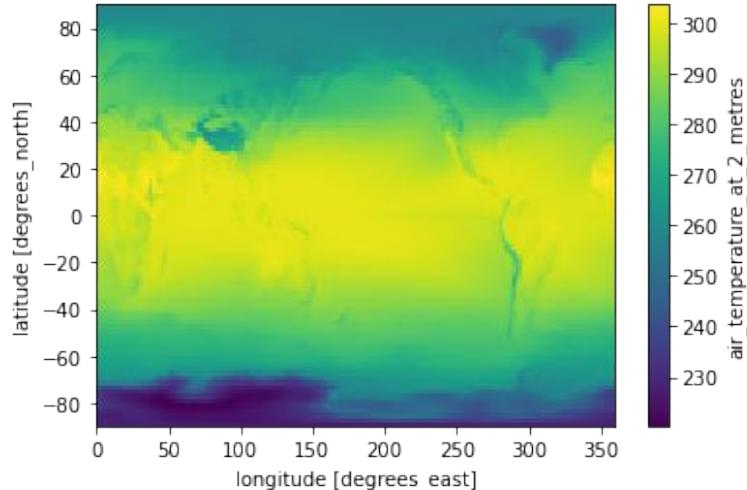
Select data and plot data in one line (using matplotlib)

```
ds.air_temperature_at_2_metres.mean("time").plot()
```

Select a variable

Apply a
method()
across a
coordinate

Apply a
method()



Same thing, but across all variables

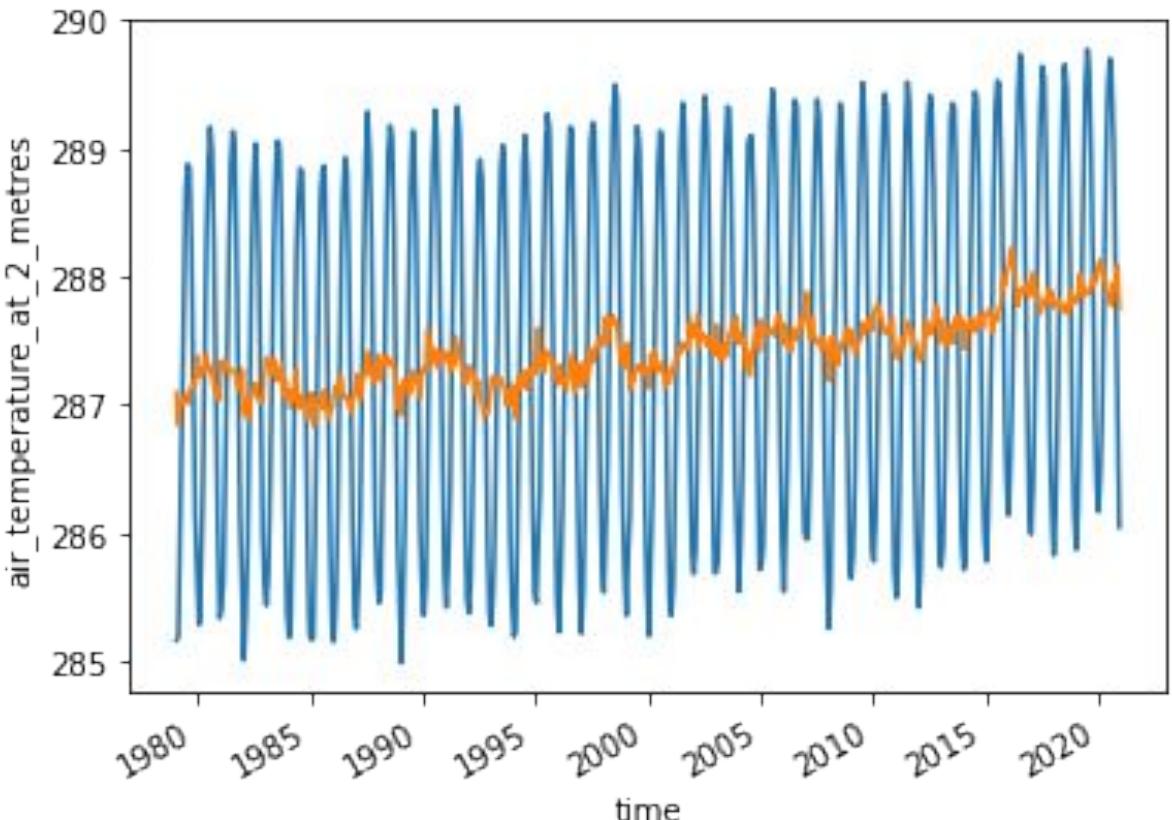
```
mean_map = ds.mean("time")  
mean_map.air_temperature_at_2_metres.plot()
```

Lazy

Plot the global trend in a variable



- Calculate a time series
- Take out the annual cycle
- Plot the trend



Goal: Calculate time series

Xarray has high level methods like `.mean()`, `.std()`, etc.

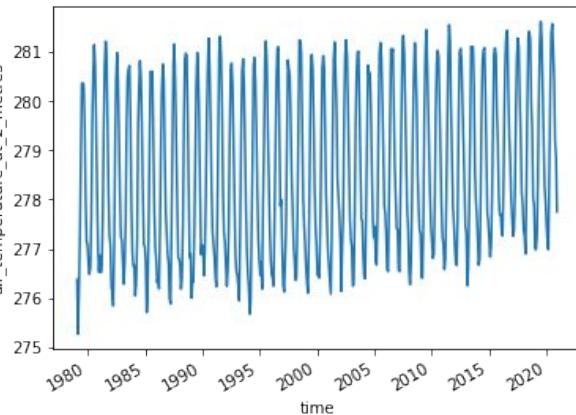
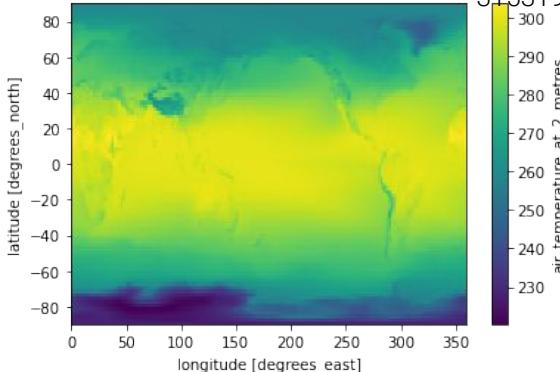
```
ave = ds.mean("time")  
ave.air_temperature_at_2_metres.plot()
```

Take the mean across all time

Take the mean across all locations

```
ave = ds.mean(("latitude", "longitude"))  
ave.air_temperature_at_2_metres.plot()
```

Does that look right?
(279 K ~ 6 °C ~ 43 °F)



The map is flat - but the Earth is not - Gaussian grid



Programs aren't smart - you are - so what went wrong?



Gridded data is nice to work with but what does it represent?

How many grid points are at 90N (the North Pole)?

How many grid points are at the Equator?

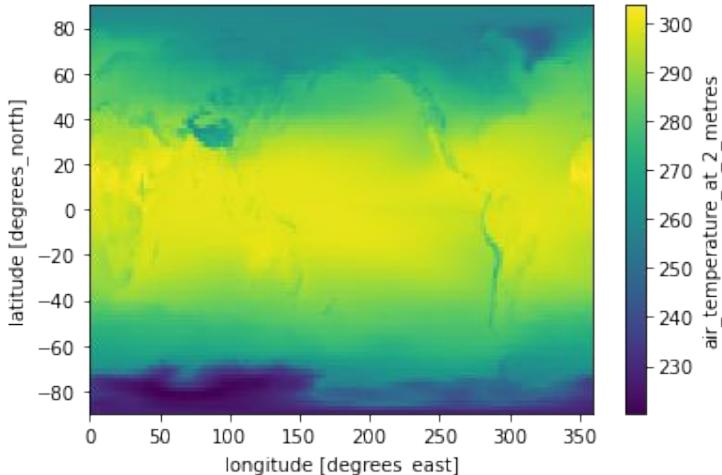
ds.dims

```
Frozen({'time': 504, 'latitude': 90, 'longitude': 180})
```

ds.coords

Coordinates:

```
* time      (time) datetime64[ns] 1979-01-16T11:30:00 ... 2020-12-16T11:30:00
* latitude   (latitude) float32 -88.88 -86.88 -84.88 ... 85.12 87.12 89.12
* longitude  (longitude) float32 0.875 2.875 4.875 6.875 ... 354.9 356.9 358.9
```

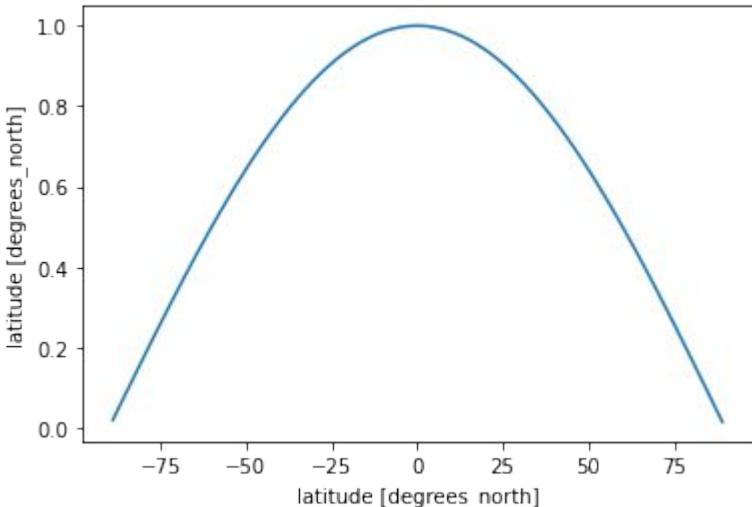


Weigh your data



Xarray provides the ability to weigh your data

```
weights = np.cos(np.deg2rad(ds.latitude))
weights.name = "weights"
weights.plot()
```



Goal: Examine average values - weighted version



Xarray methods like `.weighted()` can be combined with `.mean()`

```
ds_weighted = ds.weighted(weights)
weighted_mean = ds_weighted.mean()
weighted_mean
```

xarray.Dataset

» Dimensions:

» Coordinates: (0)

▼ Data variables:

air_pressure_at...	0	float64	1.011e+05		
air_temperature_...	0	float64	287.4		
air_temperature_...	0	float64	287.5		
air_temperature_...	0	float64	287.2		
dew_point_temp...	0	float64	282.4		
eastward_wind_a...	0	float64	-0.3118		
eastward_wind_a...	0	float64	-0.3675		
integral_wrt_tim...	0	float64	6.769e+05		
lwe_thickness_of...	0	float64	0.3232		
northward_wind...	0	float64	0.1729		
northward_wind...	0	float64	0.1776		
precipitation_am...	0	float64	0.0001195		
sea_surface_tem...	0	float64	291.2		
snow_density	0	float64	111.1		
surface_air_press...	0	float64	9.856e+04		

Does that look right?

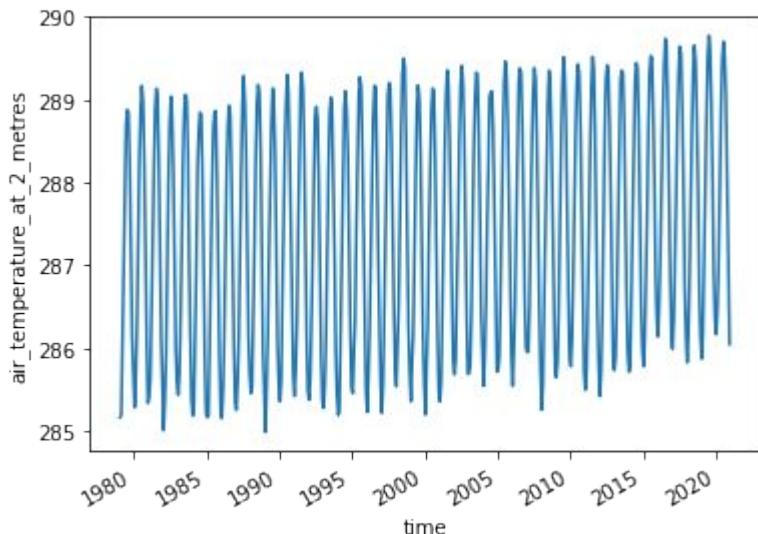
» Attributes: (0)

Goal: Weighted global time series data



You can create means across coordinates: eg. latitude and longitude

```
ds_weighted = ds.weighted(weights)
weighted_mean = ds_weighted.mean(("latitude", "longitude"))
weighted_mean.air_temperature_at_2_metres.plot()
```



288 K ~ 15 °C ~ 59 °F. 

Take out the annual cycle using .groupby()



Use .groupby on a coordinate

pandas.DataFrame.groupby

```
DataFrame.groupby(by=None, axis=0, level=None, as_index=True, sort=True, group_keys=True,  
squeeze=False, observed=False, dropna=True) [source]
```

Group DataFrame using a mapper or by a Series of columns.

A groupby operation involves some combination of splitting the object, applying a function, and combining the results. This can be used to group large amounts of data and compute operations on these groups.

Parameters: `by : mapping, function, label, or list of labels`

Used to determine the groups for the groupby. If `by` is a function, it's called on each value of the object's index. If a dict or Series is passed, the Series or dict VALUES will be used to determine the groups (the Series' values are first aligned; see `.align()` method). If an ndarray is passed, the values are used as-is to determine the groups. A label or list of labels may be passed to group by the columns in `self`. Notice that a tuple is interpreted as a (single) key.

`axis : {0 or 'index', 1 or 'columns'}, default 0`

Split along rows (0) or columns (1).

`level : int, level name, or sequence of such, default None`

If the axis is a MultiIndex (hierarchical), group by a particular level or levels.

`as_index : bool, default True`

For aggregated output, return object with group labels as the index. Only relevant for DataFrame input. `as_index=False` is effectively "SQL-style" grouped output.

`sort : bool, default True`

Sort group keys. Get better performance by turning this off. Note this does not influence the order of observations within each group. Groupby preserves the order of rows within each group.

GroupBy: split-apply-combine

xarray supports "group by" operations with the same API as pandas to implement the split-apply-combine strategy:

- Split your data into multiple independent groups.
- Apply some function to each group.
- Combine your groups back into a single data object.

Group by operations work on both `Dataset` and `DataArray` objects. Most of the examples focus on grouping by a single one-dimensional variable, although support for grouping over a multi-dimensional variable has recently been implemented. Note that for one-dimensional data, it is usually faster to rely on pandas' implementation of the same pipeline.



Resampling and grouped operations

Datetime components couple particularly well with grouped operations (see [GroupBy: split-apply-combine](#)) for analyzing features that repeat over time. Here's how to calculate the mean by time of day:

```
In [23]: ds.groupby("time.hour").mean()  
Out[23]:  
<xarray.Dataset>  
Dimensions:  (hour: 4)  
Coordinates:  
  * hour    (hour) int64 0 6 12 18  
Data variables:  
  foo      (hour) float64 728.0 729.0 730.0 731.0
```

Goal: Calculate annual cycle

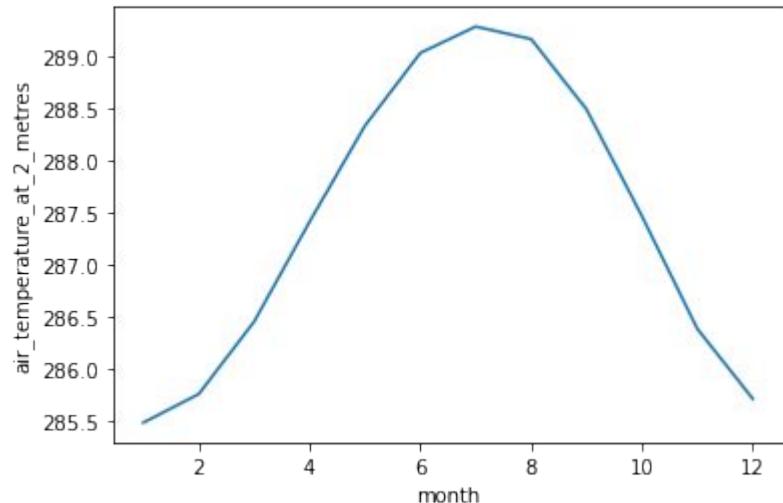


Can use `.groupby & .mean`

```
annual_cycle = weighted_mean.groupby("time.month").mean()  
annual_cycle.air_temperature_at_2_metres.plot()
```

xarray.Dataset

► Dimensions: (month: 12, latitude: 90, longitude: 180)
▼ Coordinates:
latitude (latitude) float32 -88.88 -86.88 ... 87.12 89.12
longitude (longitude) float32 0.875 2.875 4.875 ... 356.9 358.9
month (month) int64 1 2 3 4 5 6 7 8 9 10 11 12
► Data variables: (15)
► Attributes: (0)



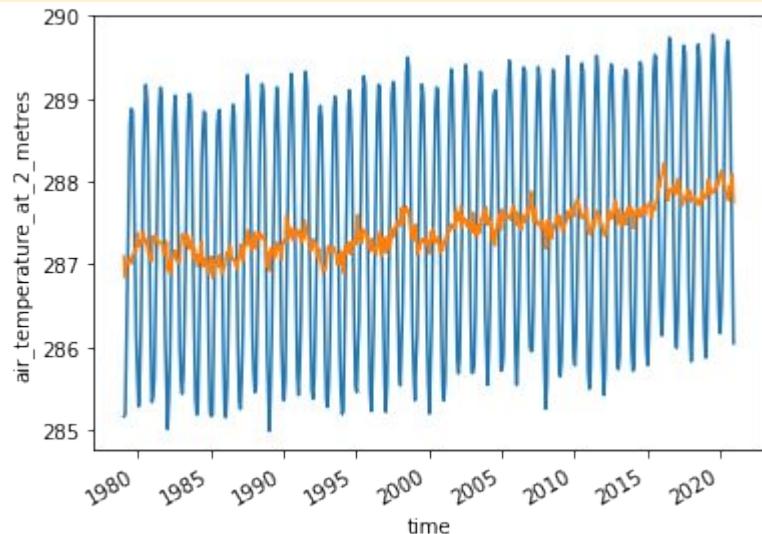
Put it all together and plot the trend



Can use .groupby & .mean

```
weighted_mean = ds_weighted.mean(("latitude", "longitude")) #weighted mean time series
annual_cycle = weighted_mean.groupby("time.month").mean() #calculate annual cycle
annual_cycle += annual_cycle.mean() #add back in the mean value

weighted_trend = weighted_mean.groupby("time.month") - annual_cycle
weighted_mean.air_temperature_at_2_metres.plot()
weighted_trend.air_temperature_at_2_metres.plot()
```





Goal: Are extremes more likely? PDF analysis.

xarray.plot.hist

```
xarray.plot.hist(darray, figsize=None, size=None, aspect=None, ax=None, xincrease=None,  
yincrease=None, xscale=None, yscale=None, xticks=None, yticks=None, xlim=None, ylim=None, **kwargs)  
[source]
```

Histogram of DataArray.

Wraps `matplotlib.pyplot.hist()`.

Plots N -dimensional arrays by first flattening the array.

Parameters

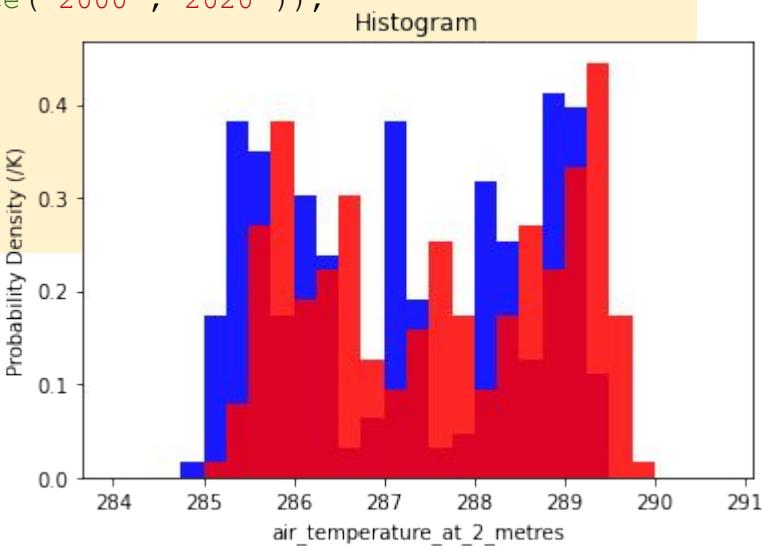
- **darray** (`DataArray`) – Can have any number of dimensions.
- **figsize** (`tuple, optional`) – A tuple (width, height) of the figure in inches. Mutually exclusive with `size` and `ax`.
- **aspect** (`scalar, optional`) – Aspect ratio of plot, so that `aspect * size` gives the `width` in inches. Only used if a `size` is provided.
- **size** (`scalar, optional`) – If provided, create a new figure for the plot with the given size: `height` (in inches) of each plot. See also: `aspect`.
- **ax** (`matplotlib axes object, optional`) – Axes on which to plot. By default, use the current axes. Mutually exclusive with `size` and `figsize`.
- ****kwargs (optional)** – Additional keyword arguments to `matplotlib.pyplot.hist()`.

```
xr.plot.hist(darray,  
             bins=bin_array,  
             density=True,  
             alpha=.9,  
             color="b",  
)
```

ERA5 temperature PDFs



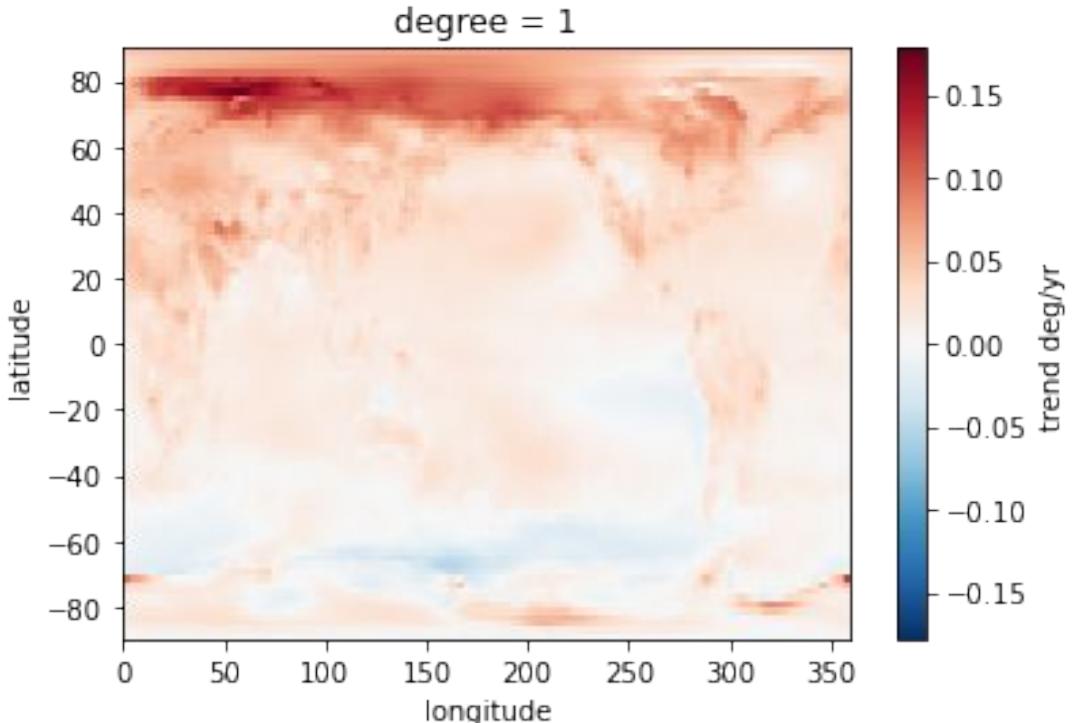
```
bins = np.arange(284, 291)
xr.plot.hist(
    weighted_mean.air_temperature_at_2_metres.sel(time=slice("1980","2000")),
    bins=bins,
    density=True,
    alpha=.9,
    color="b",
)
xr.plot.hist(
    weighted_mean.air_temperature_at_2_metres.sel(time=slice("2000","2020")),
    bins=bins,
    density=True,
    alpha=.85,
    color="r",
)
plt.ylabel("Probability Density (/K)")
```



Goal: Can we plot the trend with our ERA5 data?



```
pfit = ds.air_temperature_at_2_metres.polyfit("time", 1)
pfit.polyfit_coefficients[0] *= 3.154000000101e+16
pfit.polyfit_coefficients[0].plot(cbar_kwarg={ "label": "trend deg/yr" })
```

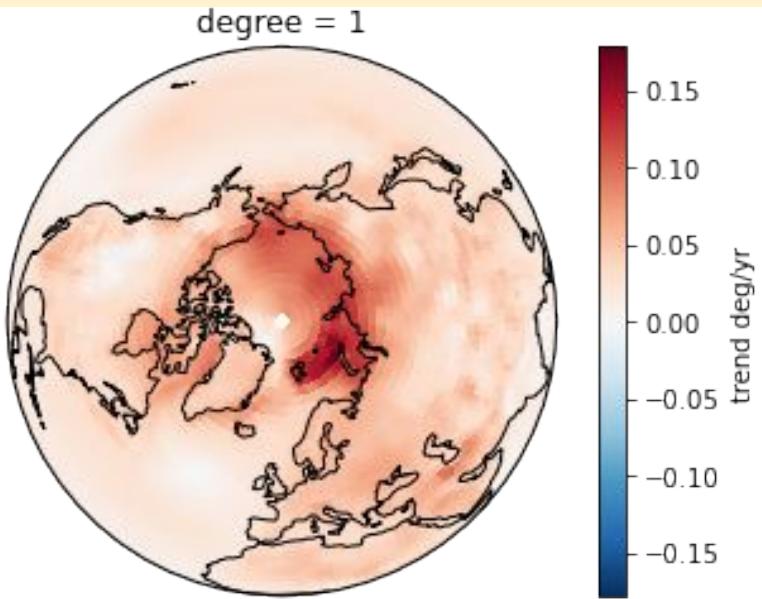


Goal: How about all fancy on a globe?



```
import cartopy.crs as ccrs

p = pfit.polyfit_coefficients[0].plot(
    subplot_kw=dict(projection=ccrs.Orthographic(0, 55), facecolor="gray"),
    transform=ccrs.PlateCarree(central_longitude=0),
    cbar_kwargs={"label": "trend deg/yr"},
)
p.axes.coastlines()
```

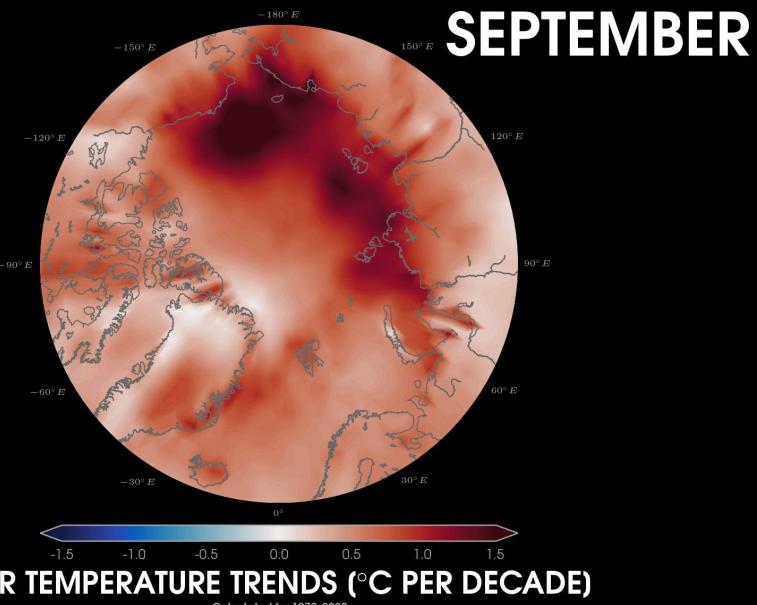


Goal: Positive feedbacks in the climate system

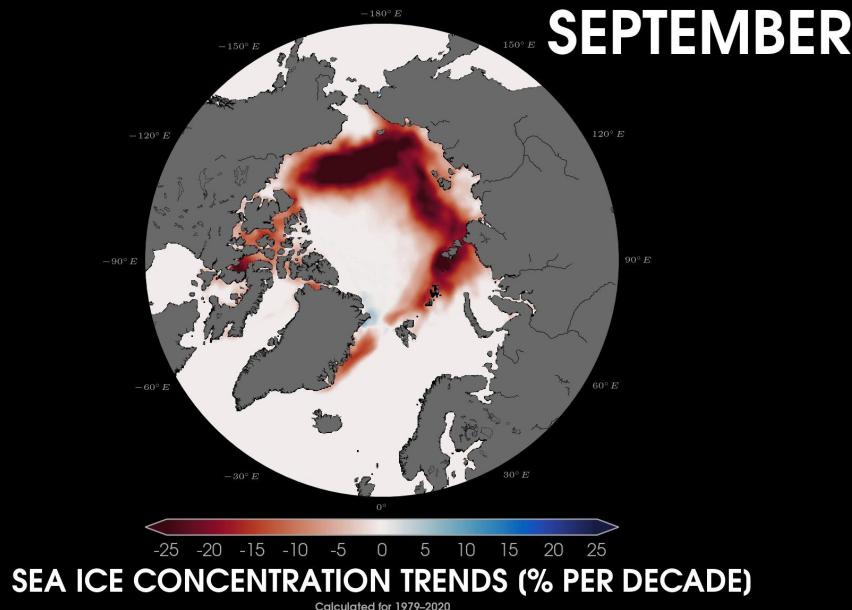


The Earth is a system - a complicated game of dominos

As air temperature increases, sea ice concentration decreases.....



GRAPHIC: Zachary Labe (@ZLabe)
SOURCE: <https://climate.copernicus.eu/>
DATA: Copernicus Climate Change Service (ECMWF ERA5 - 2m T))



GRAPHIC: Zachary Labe (@ZLabe)
SOURCE: <https://nsidc.org/>
DATA: NOAA/NSIDC CDR of Passive Microwave Sea Ice Concentration v4 (1979-2020)

Models and data tell us about our weather and climate



Xarray is a powerful tool to analyze climate data.

I've just given some simple examples here, but really, we need more eyes on the data, more eyes on the models.

The models use 'parameterizations' which are approximations for all sorts of different physical processes. Each parameterization uses coefficients derived from data science - but the experiments may be limited or imperfect. Often parameters are adjusted to compensate for an error, but then end up causing other issues, and we are all still working on this.

We need more of you, more data scientists working with climate and weather scientists to look at this data, helping to find new discoveries, amplify messages about changes to our climate and their impacts, and build machine learning models to replace old parameterizations.

(Channeling Dr. Gentemann) I'm just an oceanographer



I started off like you, I'm just an oceanographer trying to find my way through the data science world. You are all data scientists maybe trying to find something interesting to work on.

Well, we all need your help. We need minds like yours to help solve the problems our and previous generations have caused. We need your voices, backed by solid data science, to mitigate what is going on, to change our trajectory.

I've shown you how to take pandas and leap to another level library, Xarray, you can do tutorials, leap to Scipy, or other libraries. I hope you enjoyed this small tour of Xarray and climate science.

IMPORTANT UPDATE: The Office of STEM Engagement is migrating to a new application system and is NOT accepting applications for the Spring 2022 session until October 1, 2021. We apologize for the inconvenience. Please visit this site again in October to apply for our Spring and Summer 2022 sessions. Thank you for your understanding while we migrate to a new platform.

INTERNS

Being an intern is about space. Interns use their skills to support and impact NASA's mission, such as developing software, writing code, or being part of an amazing team. Interns will work with leading experts in their field to conduct research and mission planning. Applicants for this program must be currently enrolled in college or graduate school.

The application system is being upgraded. Check back on Oct. 1 to apply!
Click banner for more information!

**EXPLORE NASA INTERNSHIPS**

Meet Our Interns
Virtual Career Fair
Learn more about Artemis

**Internships and Other Student Work Opportunities****Internships and Fellowships**

NASA internships, fellowships and scholarships leverage NASA's unique missions and programs to enhance and increase the capability, diversity and size of the nation's future STEM (science, technology, engineering and math) workforce.

NASA's Goddard Space Flight Center offers hundreds of internship opportunities each year across four campuses located at:

- Greenbelt, Maryland
- Wallops Flight Facility, Wallops Island, Virginia
- Goddard Institute For Space Studies, New York City
- Independent Verification and Validation Facility, Fairmont, West Virginia

Internships are available at all levels of education from high school to graduate. Internships provide students with the opportunity to participate in either research or other experiential learning, under the guidance of a mentor at a NASA installation.

Eligibility Requirements

- U.S. citizenship
- GPA: 3.0 on a 4.0 scale
- High school students
 - At least 16 years of age and a current sophomore, junior or senior
- Undergraduate or graduate students
 - At the time the opportunity begins, must be accepted/enrolled full-time in an accredited U.S. college or university

To apply for NASA internships, fellowships and scholarships, visit NASA's OSSI site: <https://intern.nasa.gov>

For additional information: GSFC-Education@mail.nasa.gov

Intern

Discover exciting internships and research opportunities at the leading center for robotic exploration of the solar system.

ABOUT**APPLY****FAQ****ABOUT** ▾**RESOURCES** ▾**LOCATIONS****CHALLENGES****SIGN UP****LOGIN****Challenge****SPACE FOR CHANGE****DETAILS****RESOURCES****TEAMS (14)****EXAMPLE RESOURCES** ▾

Space For Change Challenge Video

FIND OR START A TEAM

to change everything, we need everyone



A NASA OPEN SOURCE SCIENCE INITIATIVE:
TOPS: TRANSFORM TO OPEN SCIENCE

Open Source Science

A Conversation with Leaders at NASA and UC Berkeley



October 26, 2022

2:00pm refreshments, Banatao Auditorium, UC Berkeley

Photo: NASA

News post and video



<https://science.nasa.gov/open-science/>

<https://github.com/nasa/Transform-to-Open-Science>

2023

NASA's Year of Open Science



[TOPS Email List](#)

NASA Transform to Open Science Mission

Dr. Chelle Gentemann, Science Lead

Yvonne Ivey, Equity Lead

Cyndi Hall, Community Coordinator

Isabella Martinez, Content Coordinator

Dr. Yaitza Luna-Cruz, TOPS Program Officer

Dr. Paige Martin, TOPS Program Officer

Kevin Murphy, Chief Science Data Officer SMD

Katie Baynes, Deputy Chief Science Data Officer SMD

Dr. Steve Crawford, Science Policy Officer SMD

Andy Mitchell,

Dr. Elena Steponaitis, SMD Development Program Executive

Amy (Uyen) Truong, Chief Science Data Office Coordinator

Dr. Rachel Paseka, OSSi Program Officer

Dr. J.L. Galache, OSSi Program Officer

Dr. Demitri Muna, OSSi Program Officer

Molly Adams, OSSi Coordinator



[TOPS Website](#)