Climate and Weather Data

STEG Virtual Course on “Data in Macro Development”

Jonathan Colmer
University of Virginia and CEPR

May 31st, 2024

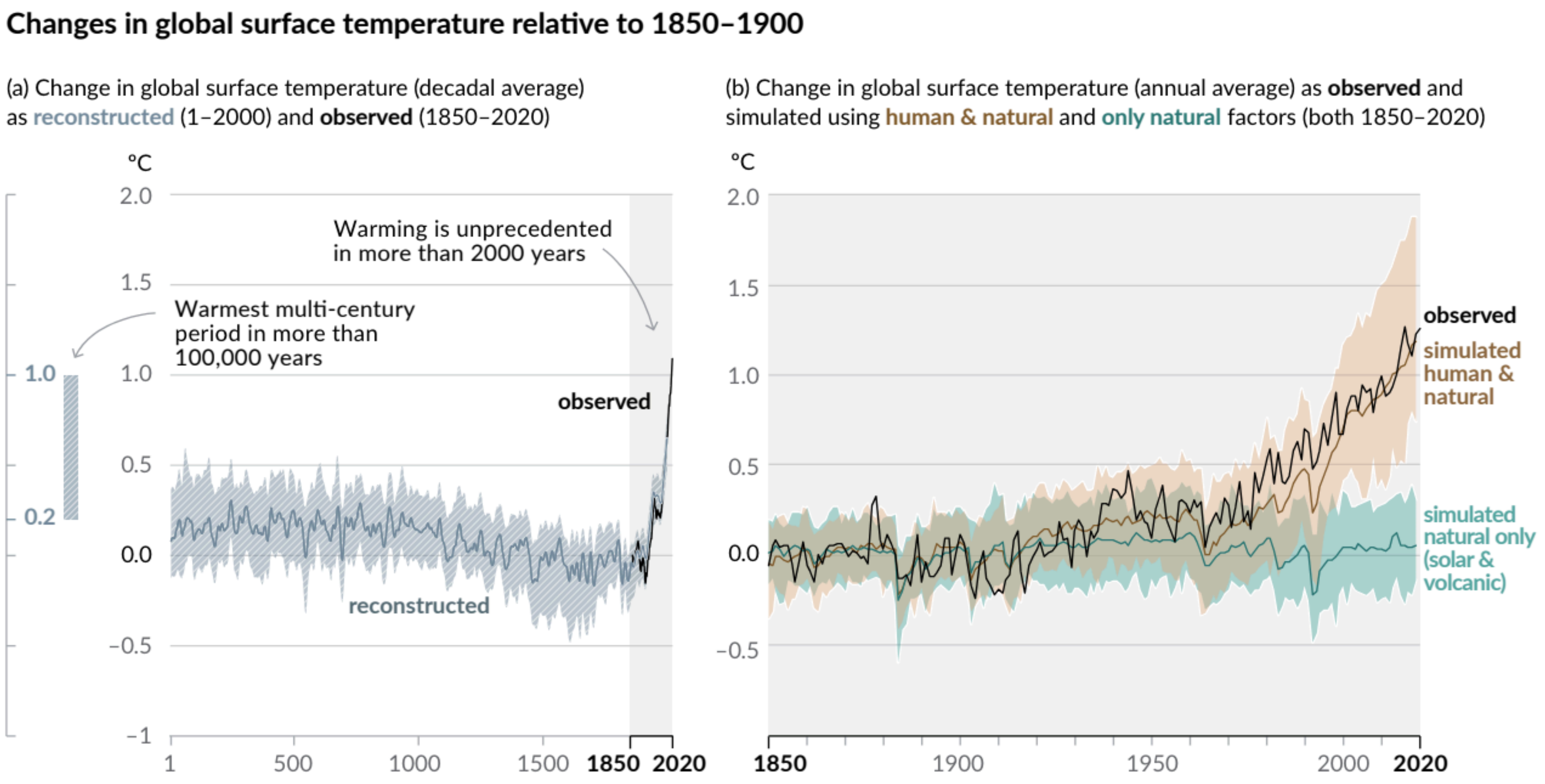
ENVIRONMENTAL INEQUALITY LAB

UNIVERSITY of VIRGINIA

STEG
STRUCTURAL TRANSFORMATION AND ECONOMIC GROWTH

CEPR

Human influence has warmed the climate at an unprecedented rate.



Source: IPCC (2023)

A Global Challenge that is Experienced Locally

↑ Extreme Heat

↑ Intense Rainfall

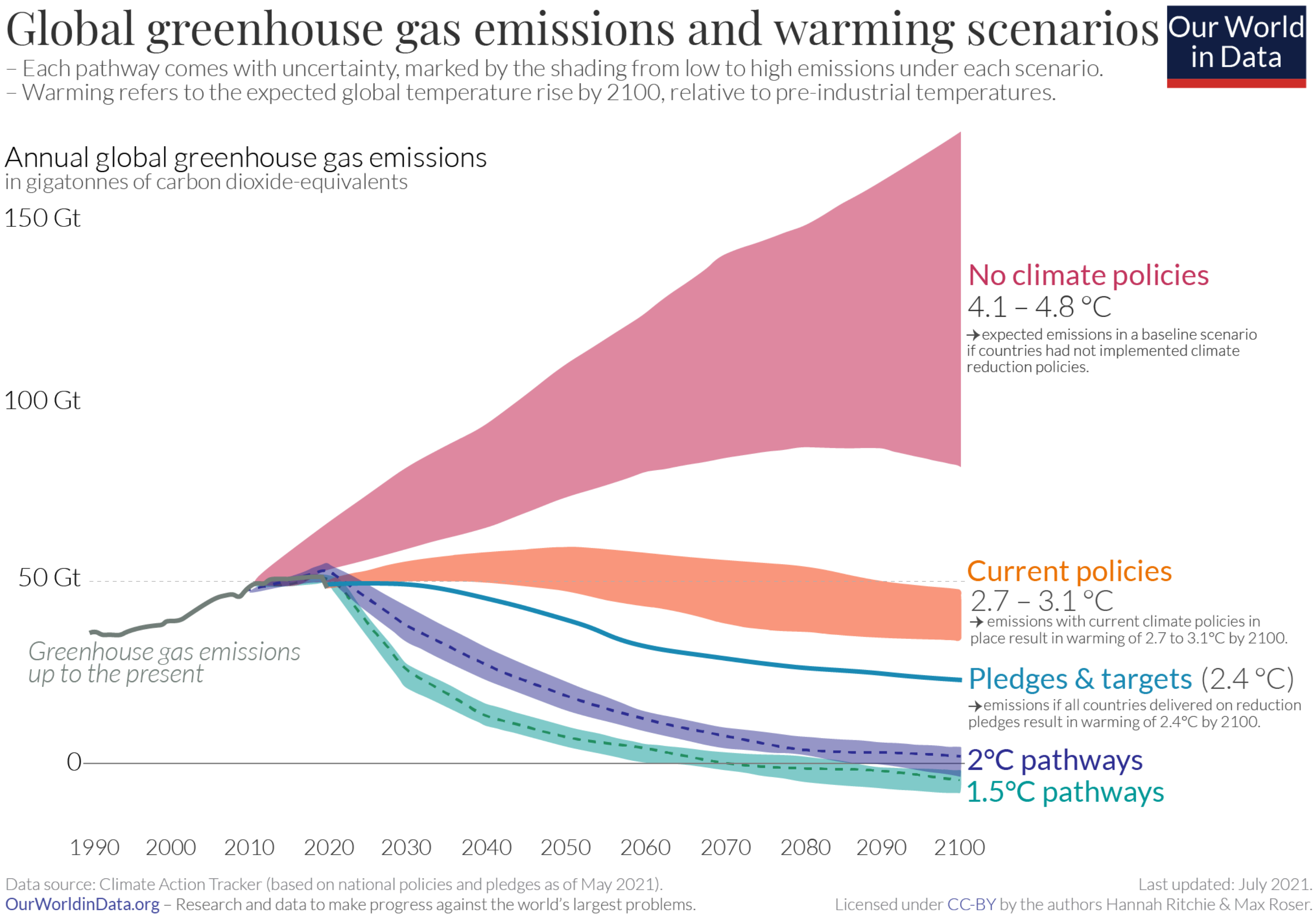
↑ Wildfires

↑ Droughts

↑ Intense Storms

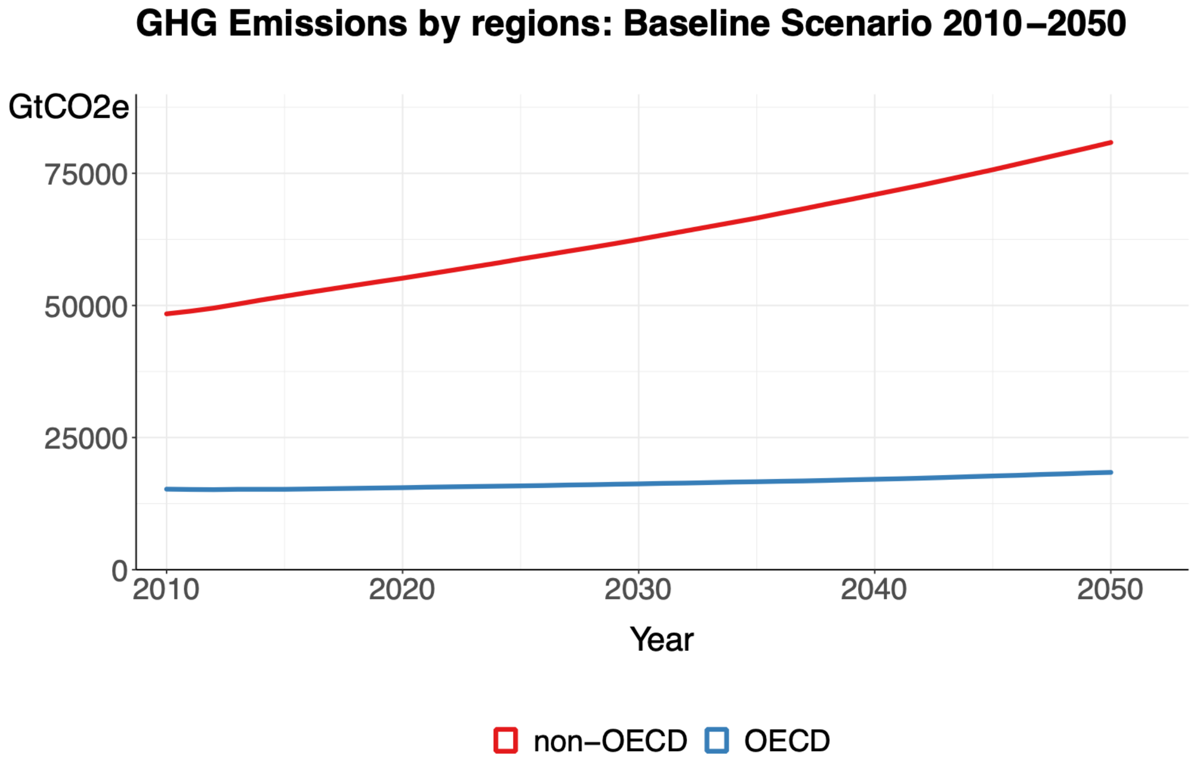
↑ Relative Sea Level Rise

The Challenge



A Changing Composition

- Developed economies got us into this mess.
- LMICs:
  - currently generate 2/3 of emissions
  - account for almost all emissions growth
  - are rightly focused on alleviating poverty and increasing relative living standards.
- Emissions will continue to increase unless decoupled from growth.
- **Fundamental Question:** How do we balance growth and the externalities from growth?



Source: OECD Environmental Outlook Baseline

Decarbonization and Growth

- Low-carbon “frontier” growth opportunities largely will come from developed economies — a “new industrial revolution”. (Acemoglu et al., 2012; Aghion et al., 2016; Van Reenen et al. 2020; Stern and Valero, 2021)
- No guarantees (Besley and Persson, 2023).
- To what degree will LMICs “catch up”?
    - We need to understand constraints to financing, adoption, transfer, and integration.
    - It doesn’t matter where emissions reductions occur (Glennerster and Jayachandran, 2024)

How will climate change affect growth and development?

$$Y_{it} = A_{it}F(L_{it}, K_{it})$$

- Clear channels through which climate change could affect labor, capital, and efficiency.
- More scope for growth effects in LMICs.
- Less scope (imo) for growth effects in developed economies.
- Understanding opportunities for adaptation is critical, but not free.

What Do We Know?

BREAD-IGC Virtual PhD Course on Environmental Economics, September/November 2023

Past Webinar • Online • From 14 Sep 2023 at 16:00 to 16 Nov 2023 at 17:30 • Sustainable Growth, Energy and Climate change

Economic impact of climate change

9 Nov 2023 at 16:00

Michael Greenstone
Milton Friedman Professor of Economics, University of Chicago

Sea level rise

16 Nov 2023 at 16:00

Clare Balboni
Assistant Professor, London School of Economics and Political Science

Allan Hsiao
Assistant Professor, Economics and Public Affairs, Princeton University

Climate adaptation

28 Sep 2023 at 16:00

Esther Duflo
Abdul Latif Jameel Professor of Poverty Alleviation and Development Economics, Massachusetts Institute of Technology

Renewables

5 Oct 2023 at 16:00

John Van Reenen
Ronald Coase School Professor, London School of Economics and Political Science

Mar Reguant
Professor of Economics, Northwestern University

Climate migration

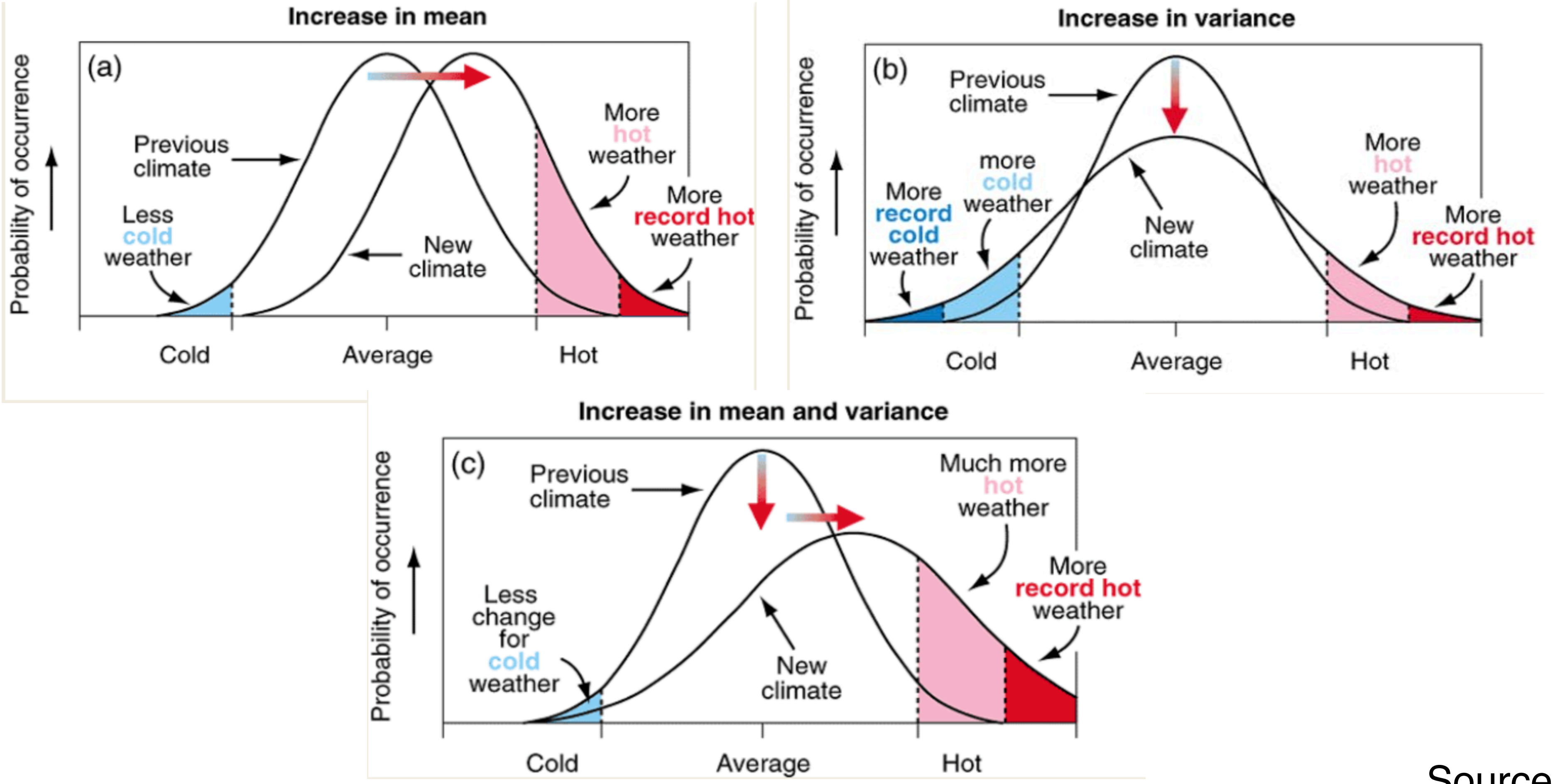
19 Oct 2023 at 16:00

Gharad Bryan
Associate Professor, London School of Economics and Political Science (LSE)

Melanie Morten
Assistant Professor, Stanford University

Weather vs. Climate

"Climate is what you expect, weather is what you get."— Mark Twain



Source: IPCC (2001)

A Brief History of Meteorology

- **650 BC** The Babylonians tried to predict short-term weather changes based on the appearance of clouds and optical phenomena
- By **300BC** Chinese astronomers had developed a calendar dividing the year into 24 festivals (each associated with a different type of weather).
- In **340BC** Aristotle wrote *Meteorologica*, which provided various theories about the formation of weather. Many of these theories were erroneous, but it took until the 17th century to realize this.
- (**16th-19th Ce**) Many advances in our understanding with the invention of the thermometer, hygrometer, and barometer, etc.
- Modern advances in computation and satellites has led to an explosion of data and massive improvements in weather forecasting accuracy.
- **Core motivations:** Agriculture, navigation, military operations, public health, scientific curiosity.

Development Economists Have Been Talking about the Weather for a Long Time

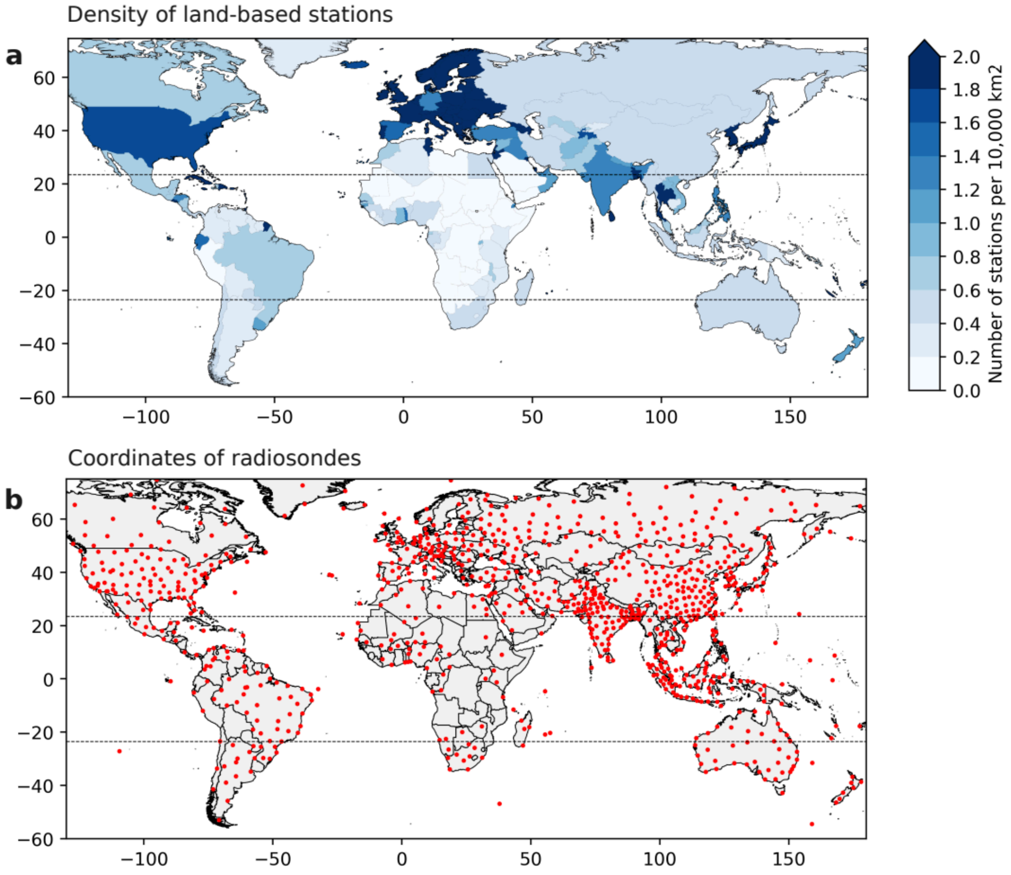
- Unpredictable weather is a dominant risk for those living in rural parts of developing countries.
- Key questions:
  - How do individuals manage risk?
  - How are expectations formed?
  - What are the barriers to mitigating risk?
  - How does the economic and policy environment help/hinder?
  - What are the aggregate consequences of risk?

An Exciting Time for Macro-Development

- Many of the remaining important questions are macro.
    - How do we aggregate micro-estimates?
    - What about general equilibrium effects?
    - Short-run vs. long-run elasticities?
    - How are expectations about climate change formed?
- Much of the existing literature is “model free”.
- We need more structure to answer these questions.
- Important to match historical aggregate patterns and trends.
- Opportunities for micro-to-macro (Donovan 2021; Buera et al., 2023)
- Learning about the transition path is more useful at this point.

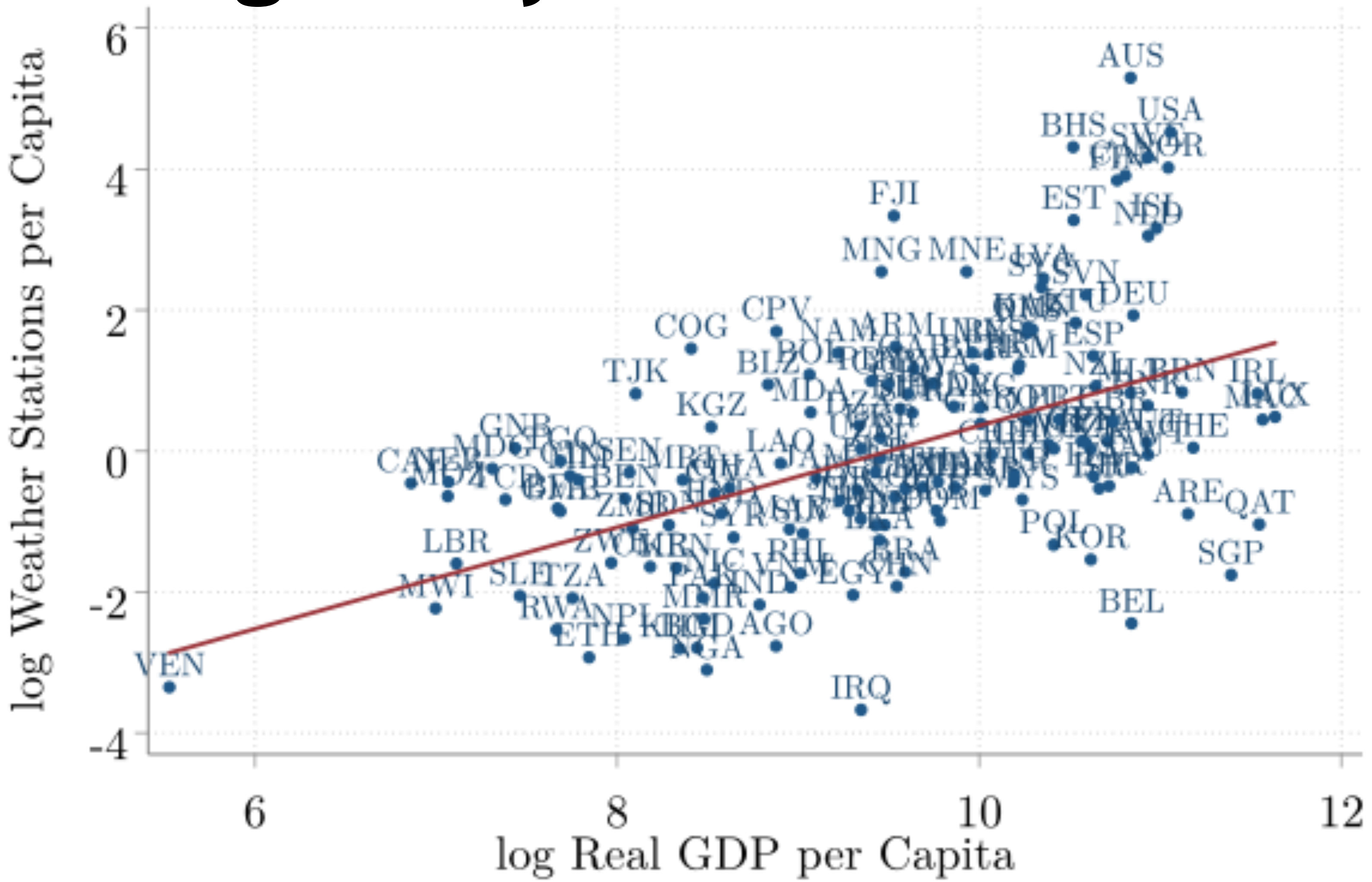
Weather Station Data

- Access to weather data is not equal.
  - Variation in distribution of stations, quality of data, and accessibility across and within countries.
  - There is an urban bias
  - Some countries have good historical data, but it’s not accessible outside of national meteorological agencies.
  - Conflict results in large gaps in the record, e.g., the Rwandan genocide resulted in the loss of 15 years of data.



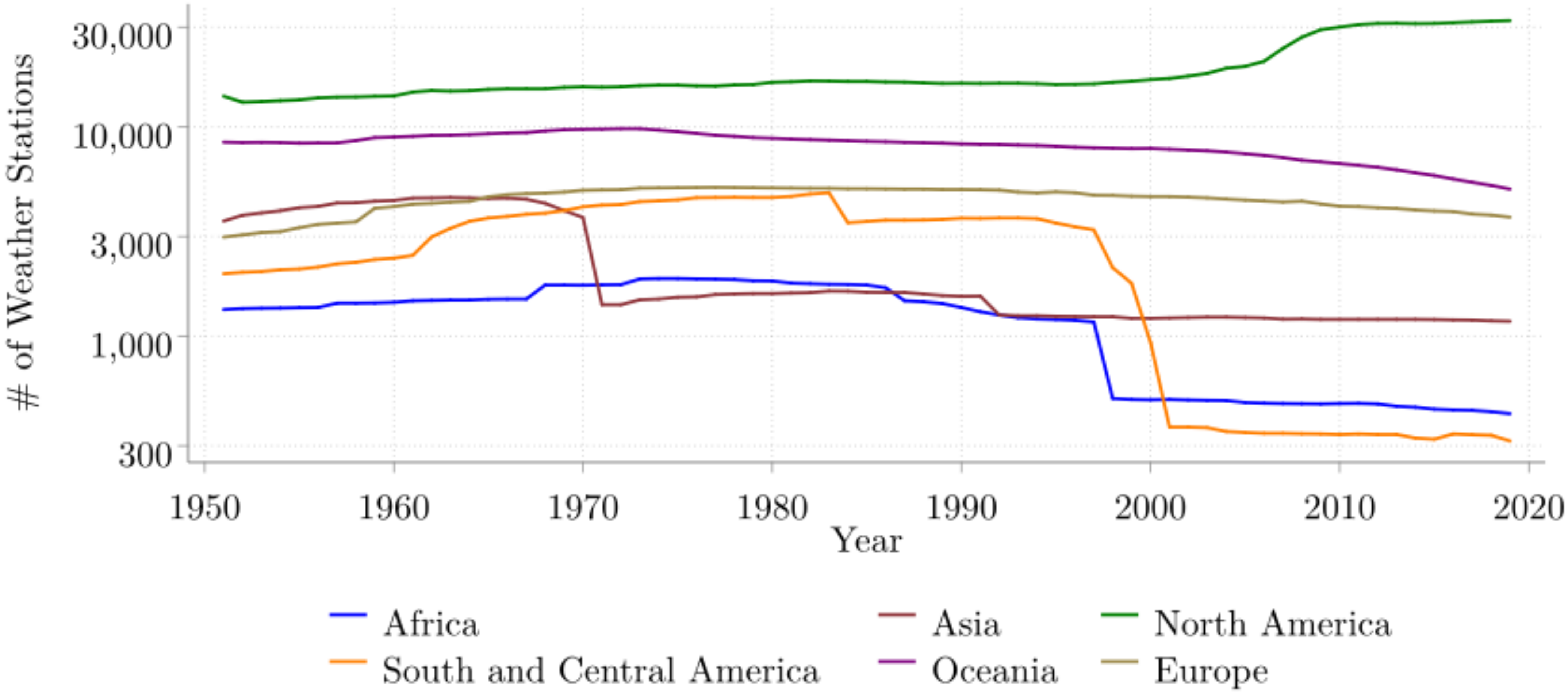
Source: Shrader and Linsenmeier (2023)

Endogeneity across Countries



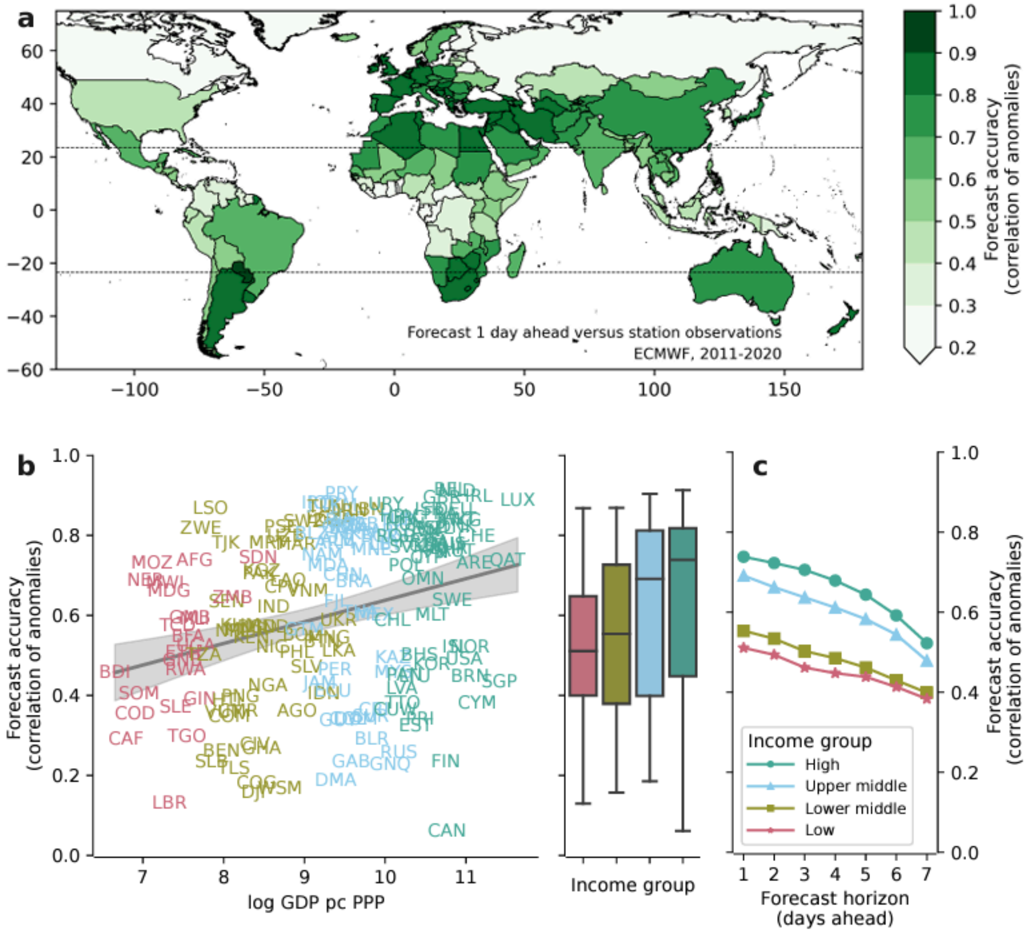
Source: Own calculations using GCHN and PWT data

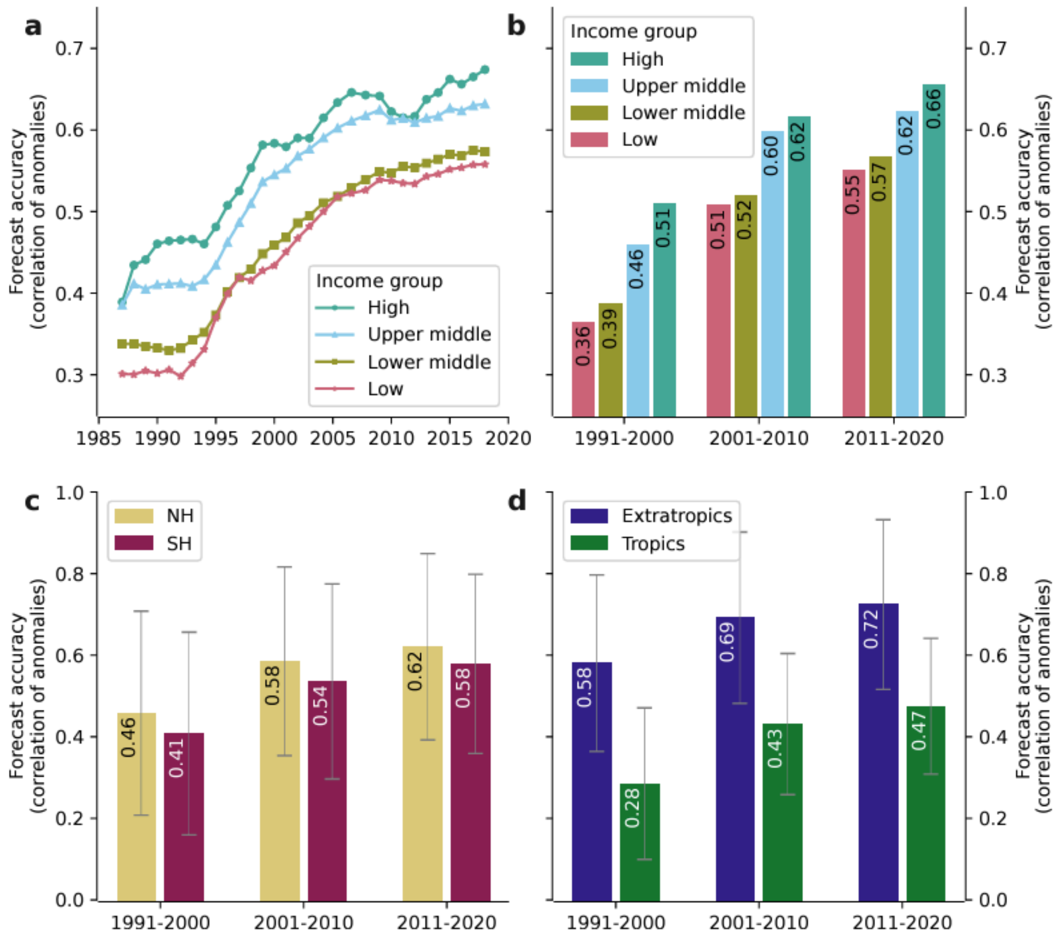
Endogeneity over Time



Source: Own calculations using GCHN and PWT data

Weather Forecasting Accuracy

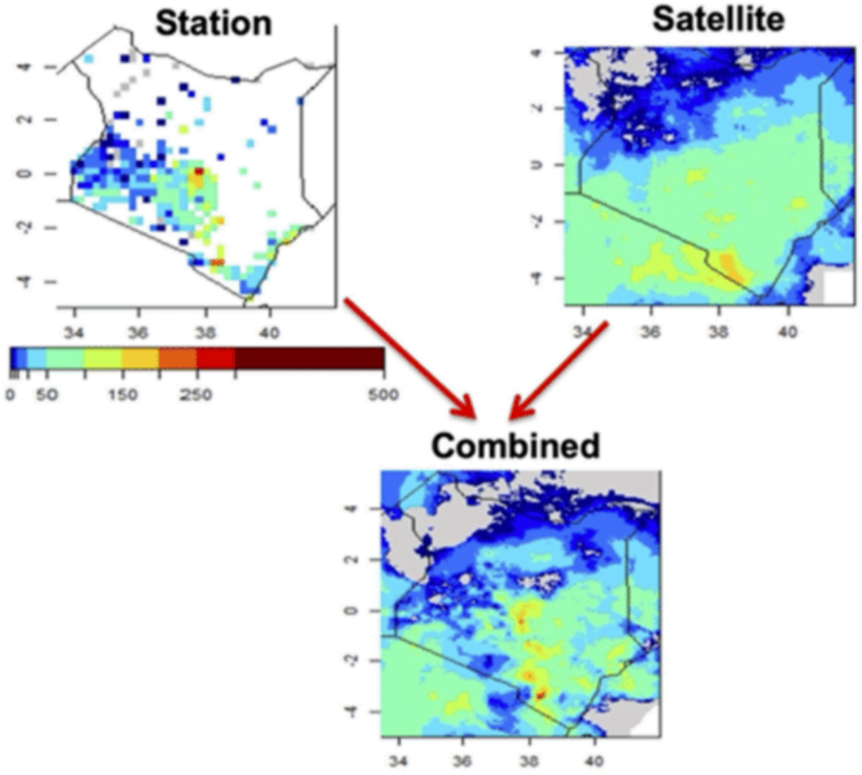




Source: Linsenmeier and Shrader (2024)

Satellite Data

- Satellite data alleviates the issues with observational stations.
- However, measurements are largely based on radiance.
- Radiance ≠ Precipitation or Temperature.
- Lots of assumptions (but physical ones).
- Be careful aware of prediction error/model uncertainty (Proctor, Carleton and Sum, 2023)
- Best practice = combination of satellite and in-situ observations



Source: Dinku et al., (2022)

Reanalysis Data

- Reanalysis data combines in-situ observations, satellite data, and climate models.
  - Provides a consistent best-estimate of weather realizations over time and space.
  - Available for every hour at multiple atmospheric levels.
- Reanalysis fills in the missing jigsaw pieces using structural models based on the laws of physics.
- Limitations are well understood; biases are well known (although identified relative to in-situ observations).
- More physically reasonable for places with low station density, but potentially less accurate in places with in-situ measurements.

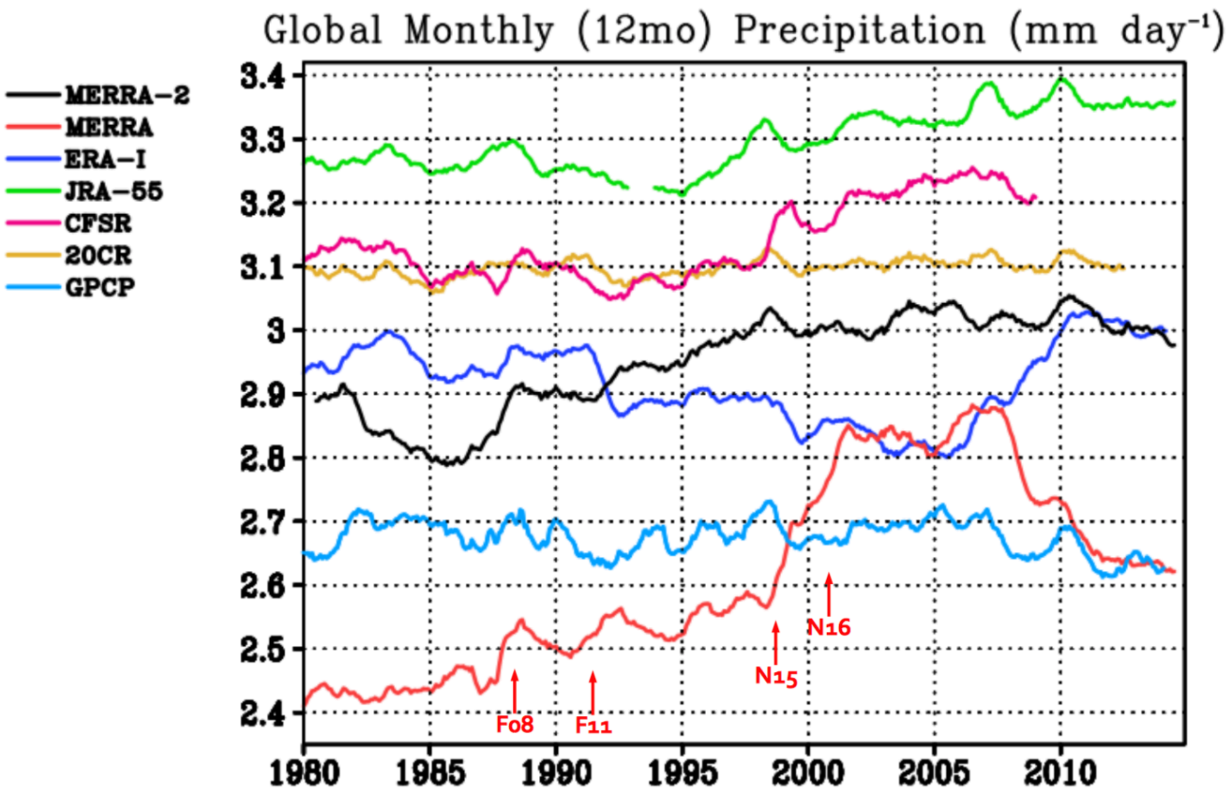
Source: Copernicus ECMWF

Working with Weather and Climate Data

- There is no universally correct weather or climate data product.
- Auffhammer et al. (2013) provides an excellent discussion of the trade-offs.
- More often than not, you will work with gridded data:
    - Interpolated datasets (UDEL, BEST, PRISM) vs. reanalysis data (ERA5)
    - Interpolated data is simple and works best in places with good station density.
- Research what data products are most appropriate for your region and question of interest.
    - Right variables? Suitable resolution? Are biases reasonable for variables and region of interest?

Hydrological Variables vs. Temperature

- Precipitation:
  - Highly heterogeneous over time and space
  - Very difficult to measure accurately.
- Temperature:
  - Relatively uniform over time and space
  - Can be interpolated with reasonable confidence.
- **Take-home:** There is way more uncertainty about hydrological variables.



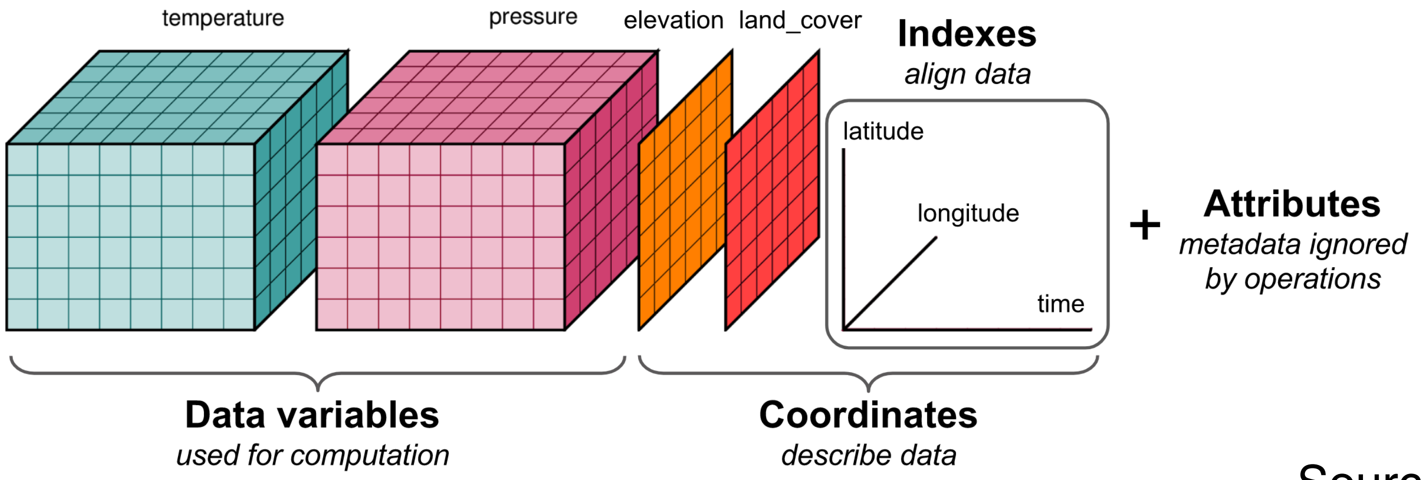
Data from Bosilovich et al. (2015). Gridded data products disagree on average global monthly precipitation by up to 40% and aren't always consistent!

Processing Weather and Climate Data

- There are usually three types of input you'll work with,
  1. **Climatic data:** gridded data providing information on temperature, precipitation, vectors of wind speed, surface pressure, etc.
  2. **Shapefiles:** polygons of the administrative boundaries of interest, e.g., country, state, districts, counties, etc.
  3. **Secondary Weights:** gridded data on population, land use, etc.

Multidimensional Data

- Climatic data is often stored as multidimensional arrays and saved as **NetCDF** files.
  - NetCDF files are “self-describing”.
  - Metadata provides detailed information about the data’s structure and content.
  - Data are indexed by latitude, longitude, atmospheric level, and time.
  - Variables are then assigned to these dimensions, e.g., for a given elevation, hourly temperature is stacked at each latitude and longitude over time.



Source: Xarray Community

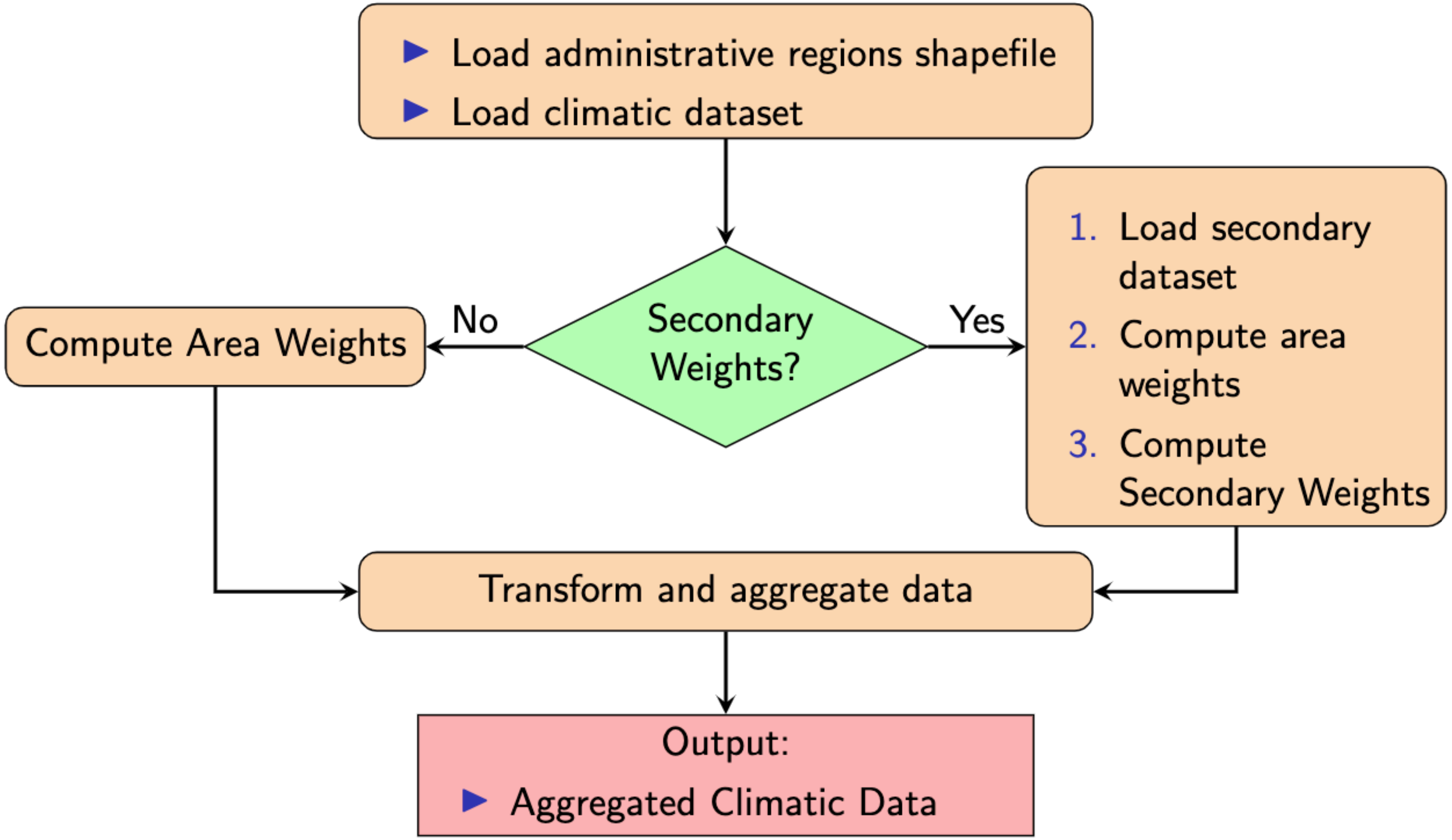
The Choice Set

- For any given research project, you need to decide:
  - What type of weather and climate you want to work with.
  - What temporal scale are you working at (daily, monthly, annual, multi-year).
    - Often constrained by economic data.
  - What region and level of spatial aggregation you are studying
  - Whether you are going to use spatial weights (if aggregating)
  - Whether you are going to aggregate-before-transforming the data or transform-before-aggregating
    - Often, you’ll want to ”transform-before-aggregating”

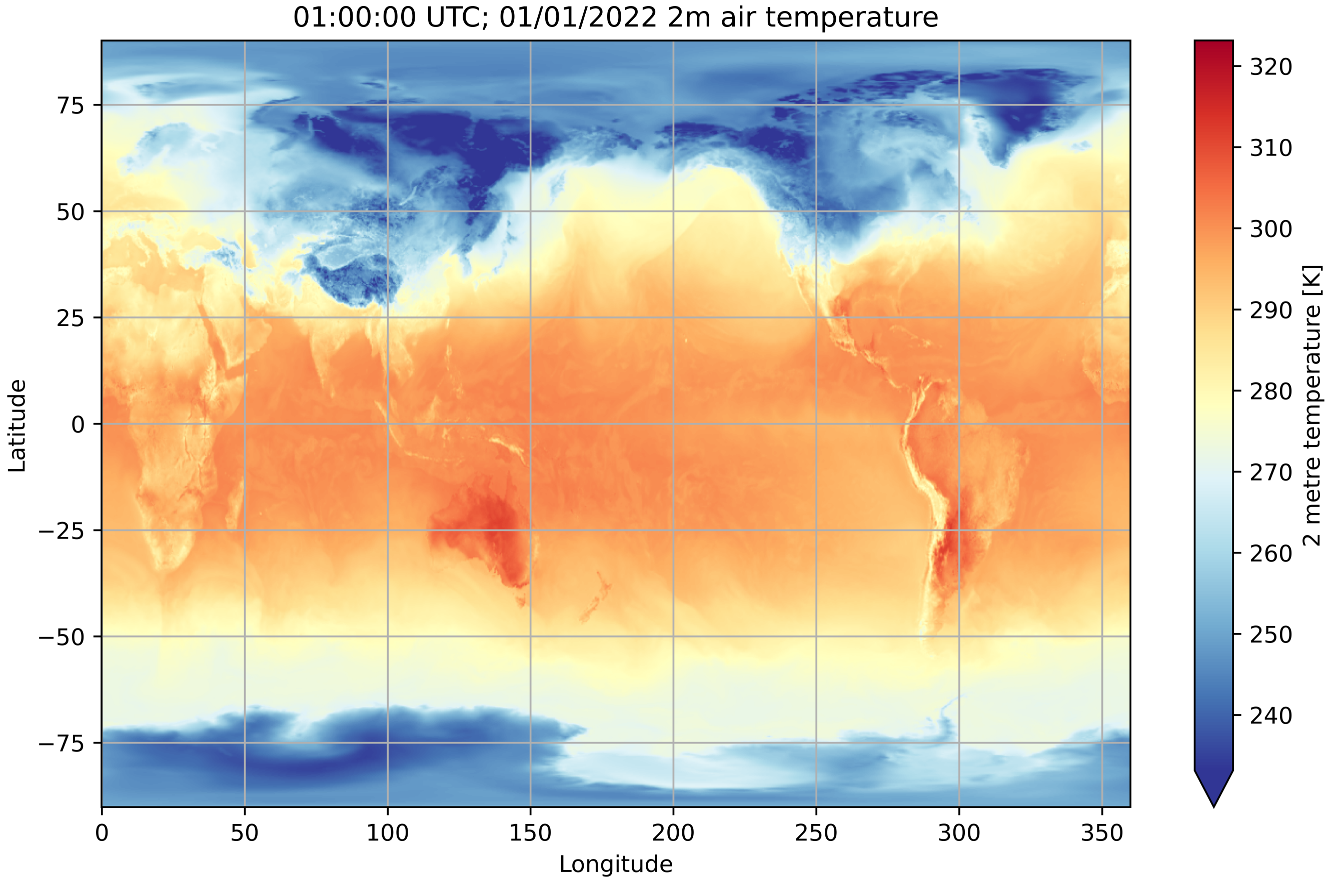
A New Tool: `AggFly`

- A new python package designed by the amazing Dylan Hogan (Columbia University, PhD *forthcoming*)
  - Aggregates climatic data over different spatial and temporal scales.
  - Efficiently automates computational and memory-intensive geospatial operations (a common barrier to entry).
  - Allows for secondary weights.
  - Constructs a wide variety of different functional forms (helpful if transforming-before-aggregating)
  - Kicks out “ready-to-use” aggregated climatic data.
  - Documentation and examples coming out soon.
- Tamma Carleton (UCSB) has put together a similar package in R `(stagg)`

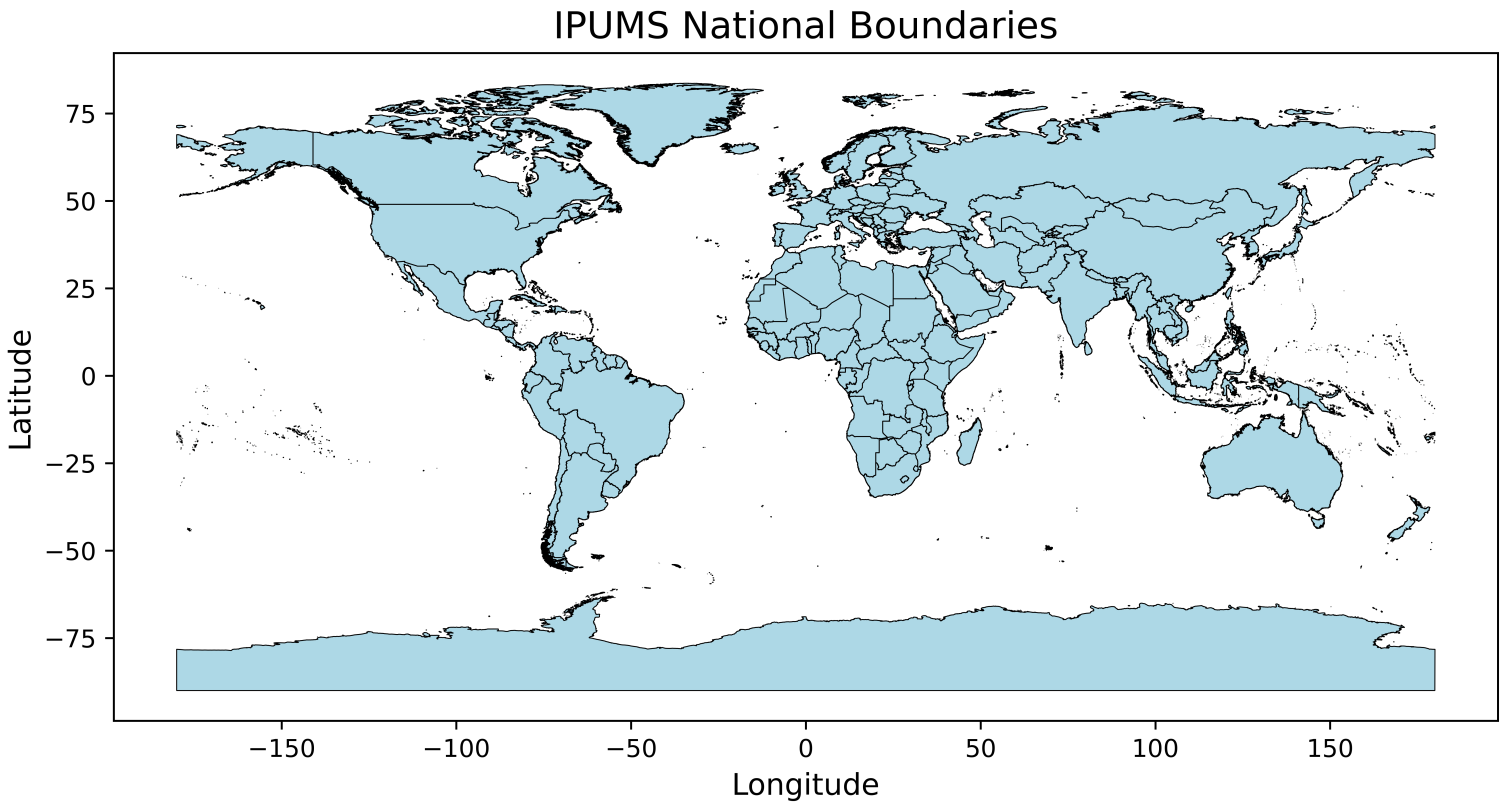
Workflow


Load administrative regions shapefile
Load climatic dataset
Secondary Weights?
No
Yes
Compute Area Weights
1. Load secondary dataset
2. Compute area weights
3. Compute Secondary Weights
Transform and aggregate data
Output:
Aggregated Climatic Data

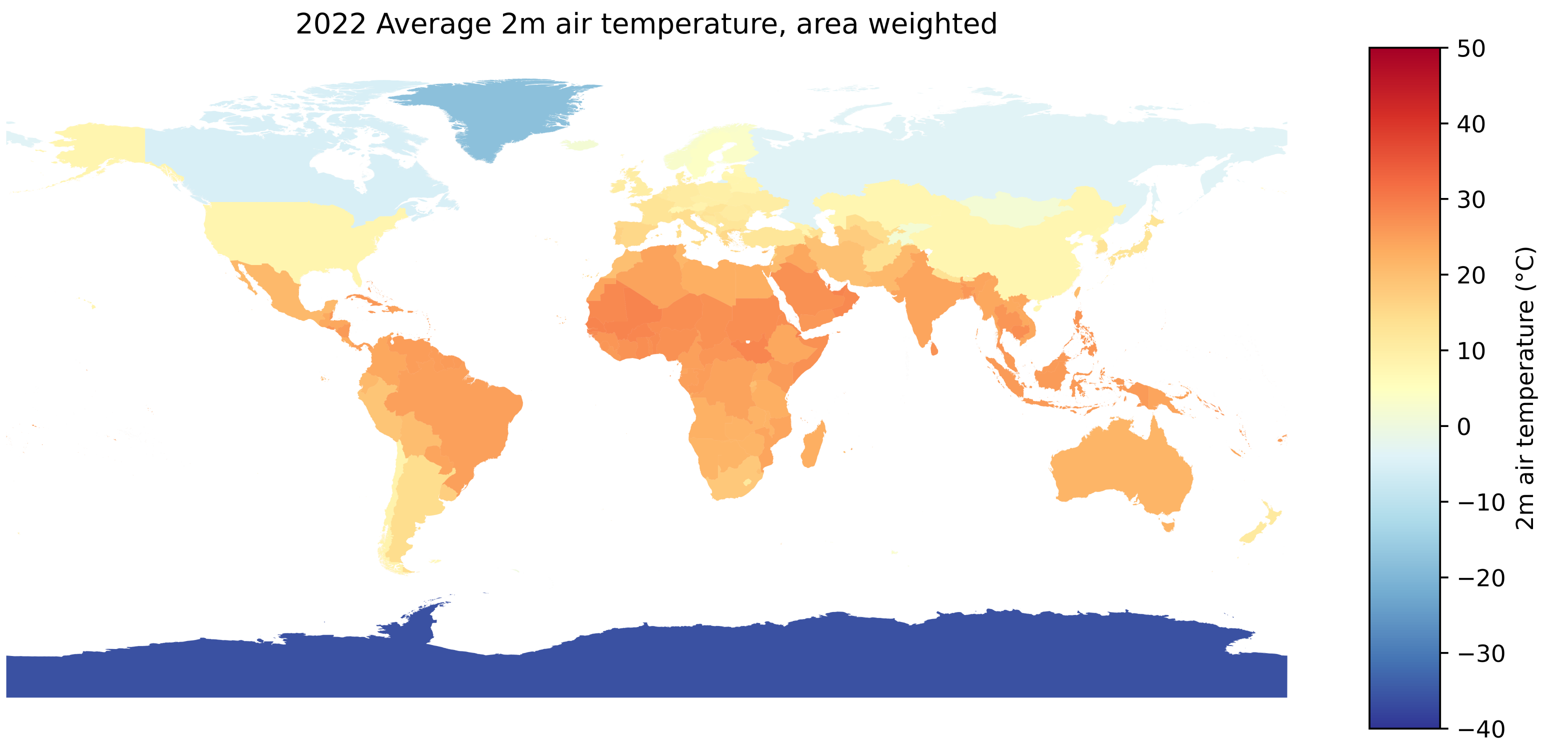
Gridded Weather data



+ Administrative Boundary Shapefiles

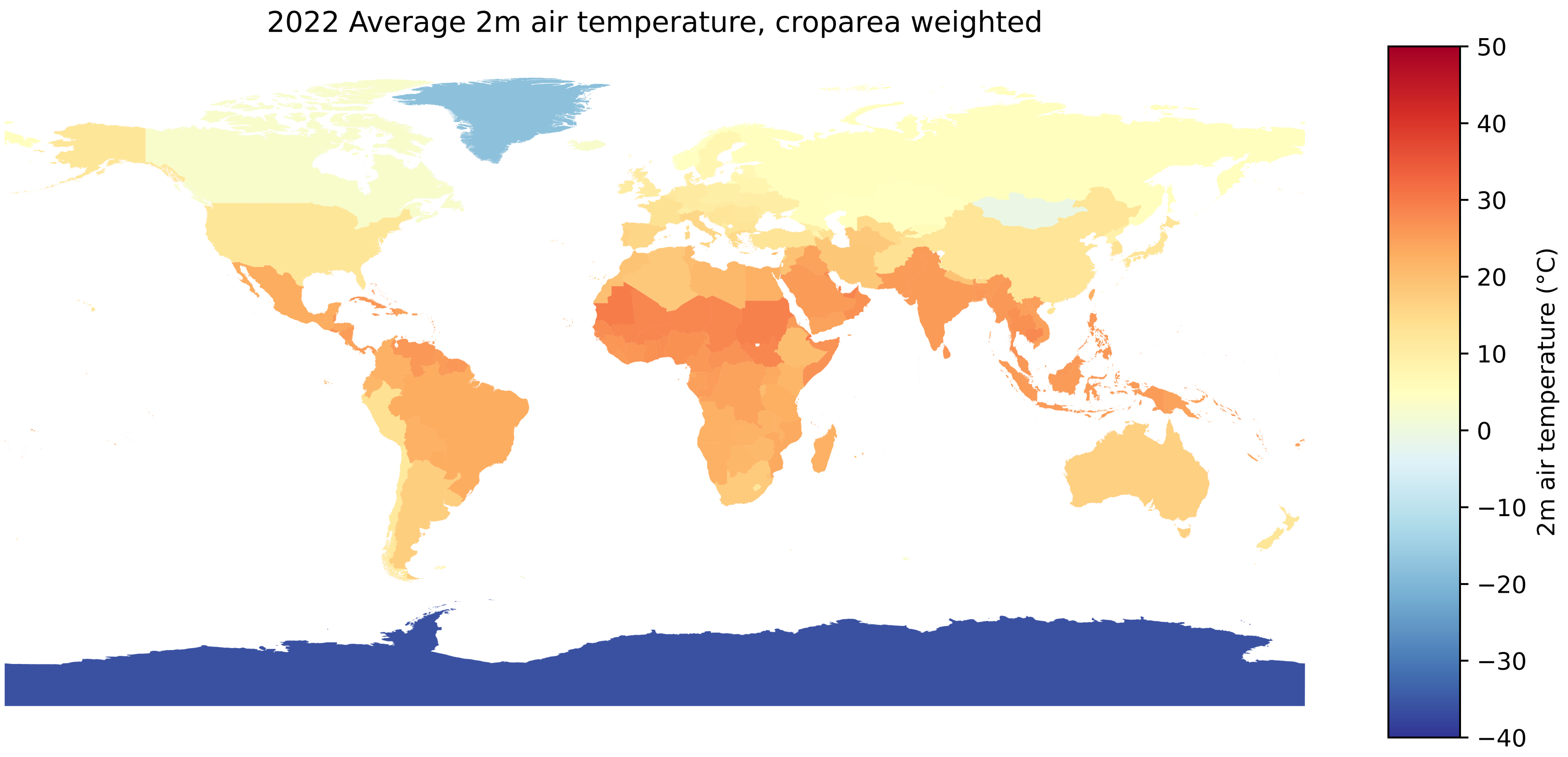


= Area-Weighted Weather Data



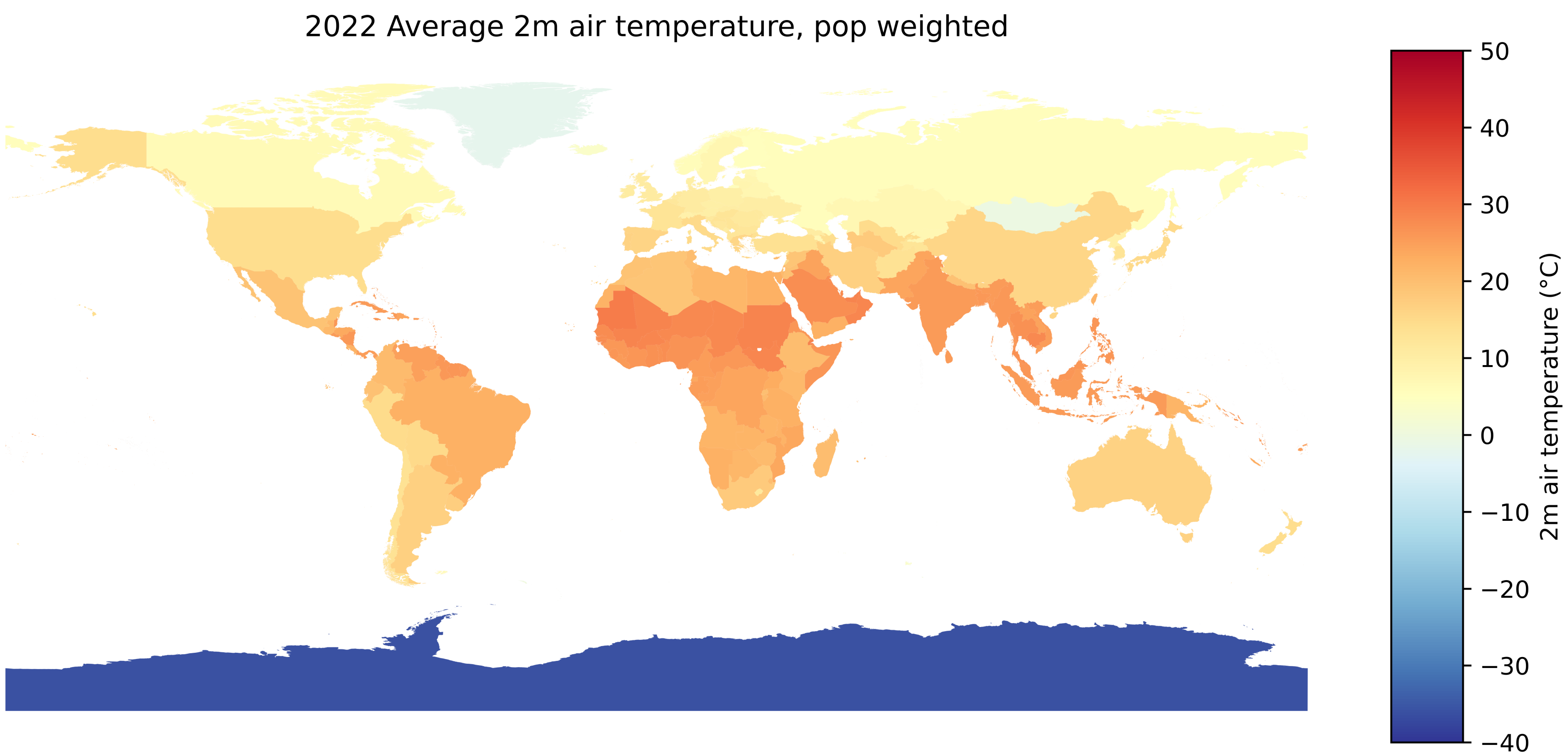
Data: ERA5 from the Copernicus Climate Data Store

= Crop Area-Weighted Weather Data



Data: ERA5 from the Copernicus Climate Data Store

= Population-Weighted Weather Data



Data: ERA5 from the Copernicus Climate Data Store

Weights

- Area weights:
  - `AggFly` uses area weights as standard.
  - Accounts for differences in the sizes of grid cells and intersecting boundaries (partial coverage).
- Secondary weights:
  - Weights based on specific variables like population or crop coverage
  - Useful for studies focusing on health, or agricultural productivity, etc.
  - Population weights example: cells with higher population contribute more that the computation of average temperature.

Current Output

- `AggFly` currently constructs aggregated climatic variables such as:
  - **Mean**: e.g. mean temperature over specified time period.
  - **Sum:** e.g. total precipitation over specified time period.
  - **Min:** e.g. minimum temperature over specified time period.
  - **Max:** e.g. maximum temperature over specified time period.
  - **Polynomials:** e.g. the sum of polynomials of daily temperature up to the n^{th} degree.
  - **Degree Days:** e.g. the sum of daily temperatures exceeding a base threshold.
  - **Bins:** e.g. the number of days that average temperature was in a a given 3 degree Celsius interval in each year.

Tropical Cyclones

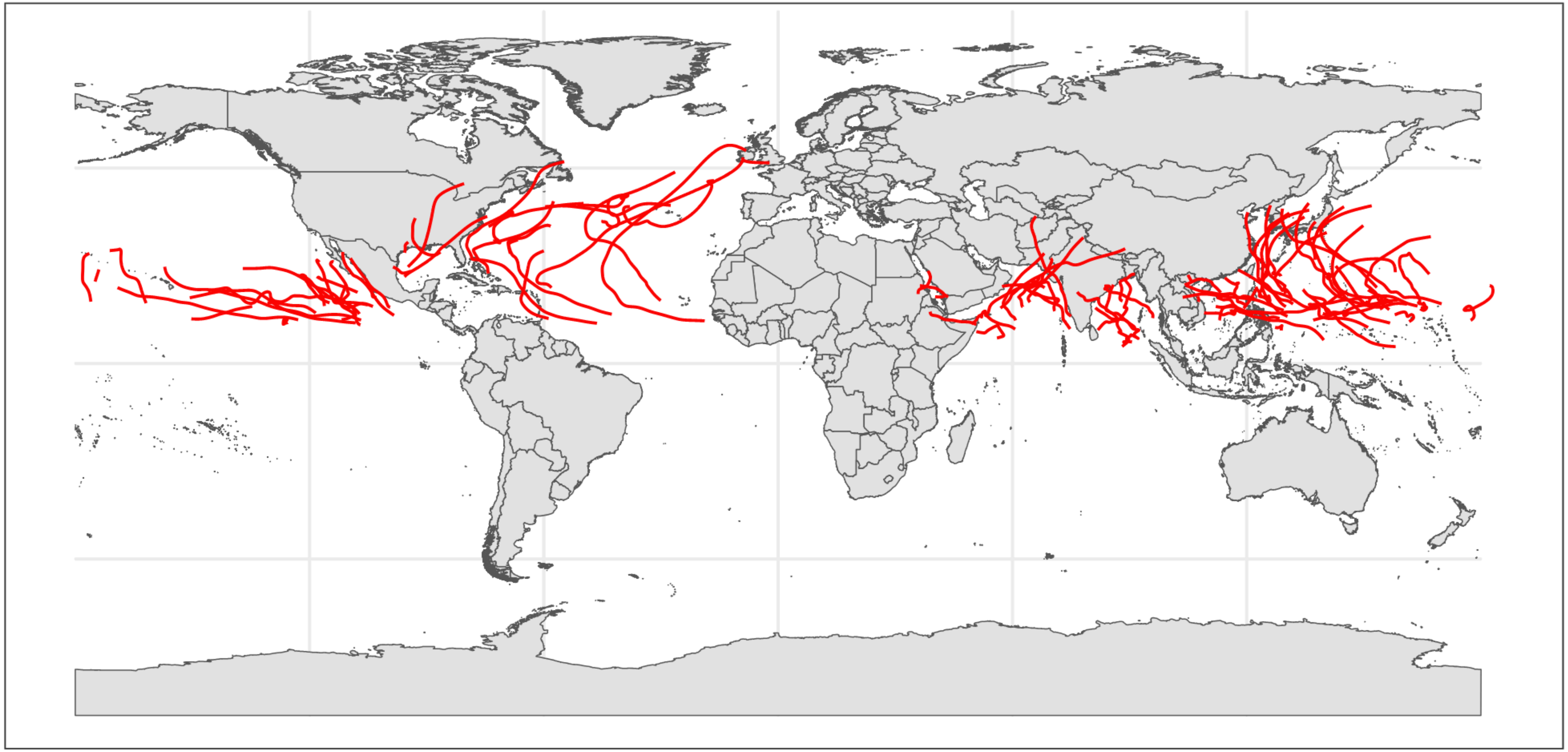
- ~35% of the global population is affected by tropical cyclones (Hurricanes, tropical cyclones, typhoons) (Hsiang and Jina, 2014).
- Tropical cyclones: large storms that form over the oceans and cause physical damage and loss of life via intense winds, heavy rainfall, and ocean surges.
- It's important to construct physical indexes of disaster exposure.
  - Self-reported damages are endogenous (concerns about coverage, quality, and amount of damage itself, etc.)
  - There is lots of nice natural random variation to exploit in terms of formation, cyclone path, and intensity.

Historical Tropical Cyclone Data

- NOAA's International Best Track Archive for Climate Stewardship (IBTrACS) database provides the most comprehensive collection of tropical storms globally.
- Provides information on path of the storms':
    - minimum central surface air pressure
    - maximum sustained wind speeds

at 6 hour intervals.

IBTrACS storm paths (2019)



Wind Field Models

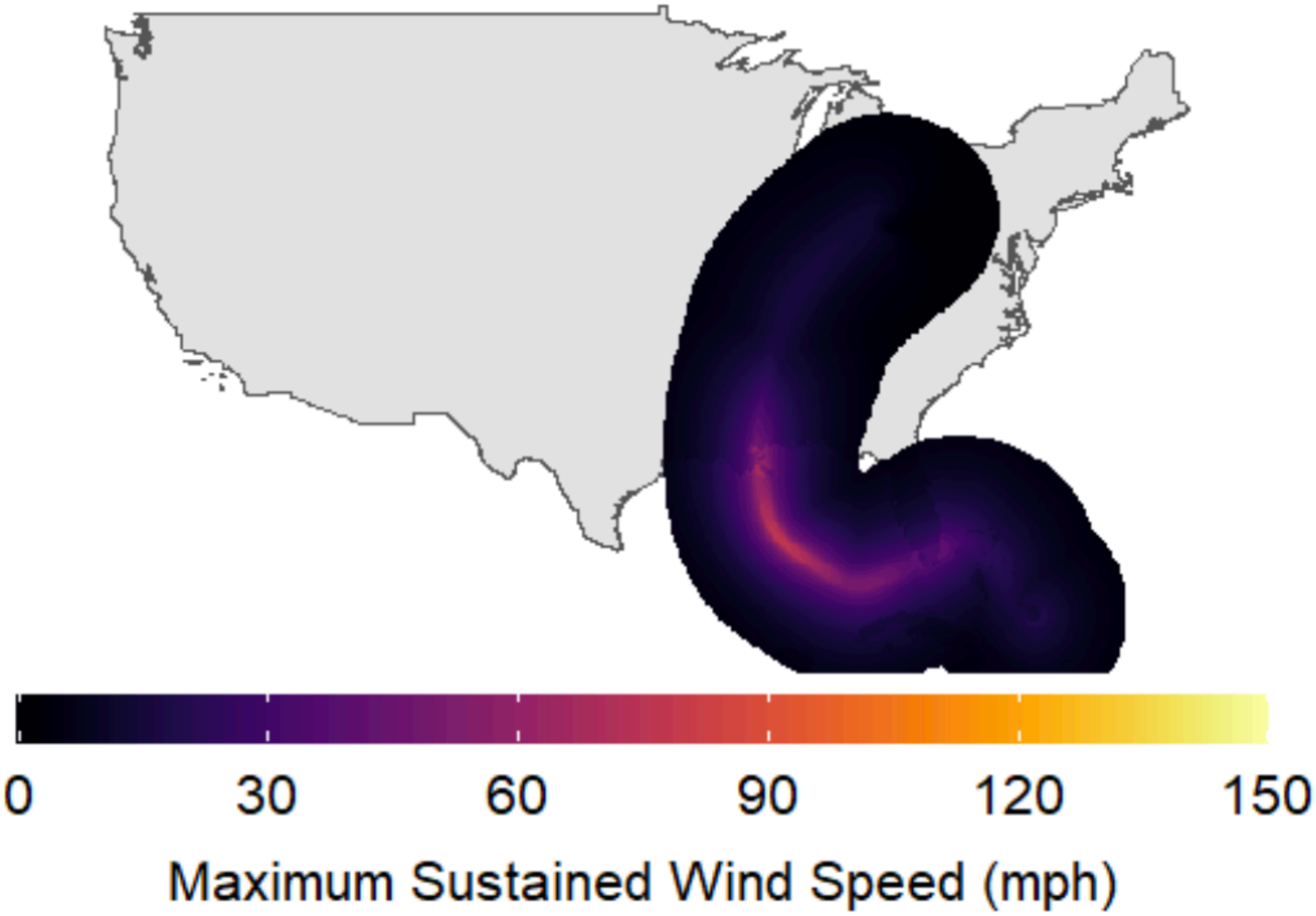
- Storm tracks aren’t sufficient.
- Tropical cyclones can have meaningful effects hundreds of km away from the center of the storm.
- To account for this we need to combine information from storm tracks with a wind field model.
    - e.g. the `stormwindmodel` package in R.
- Provides estimates of maximum sustained windspeed relative to the hurricane center.

The Willoughby Model

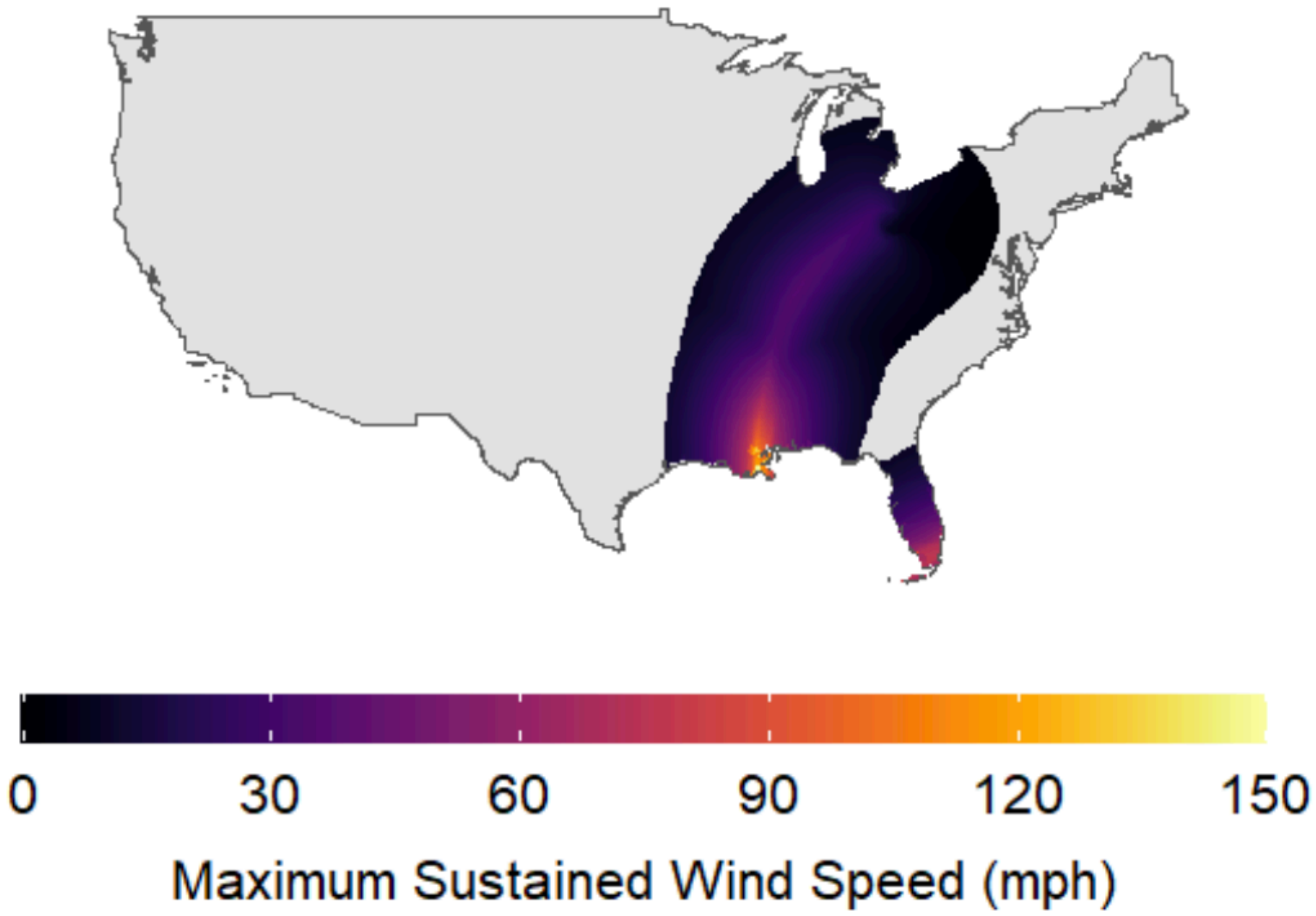
- A parametric model of radial and azimuthal wind profiles.
  - **Radial wind profile:** how does wind speed change as you move away from the center
  - **Azimuthal wind profile:** how does wind speed change as you move around, holding distance fixed (influenced by earth's rotation, storm motion, etc.)
- Provides a simplified yet accurate representation of the wind field in hurricanes.
- Key components:
  - Maximum windspeed.
  - Radius of maximum winds.
  - Shape parameters that constrain how quickly windspeed decay as you move away from the center.

Hurricane Katrina (2005)

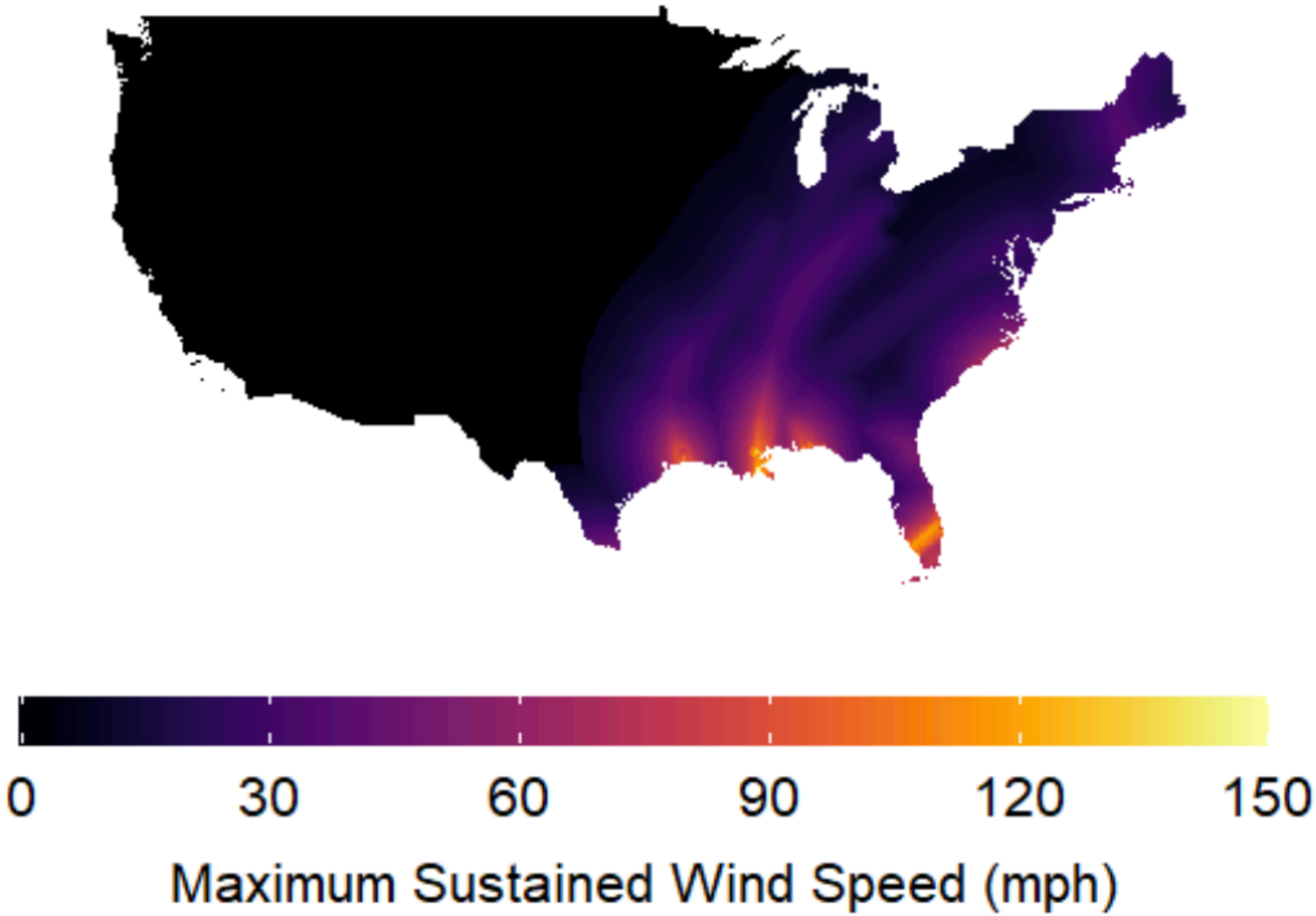
Hurricane Katrina (2005)



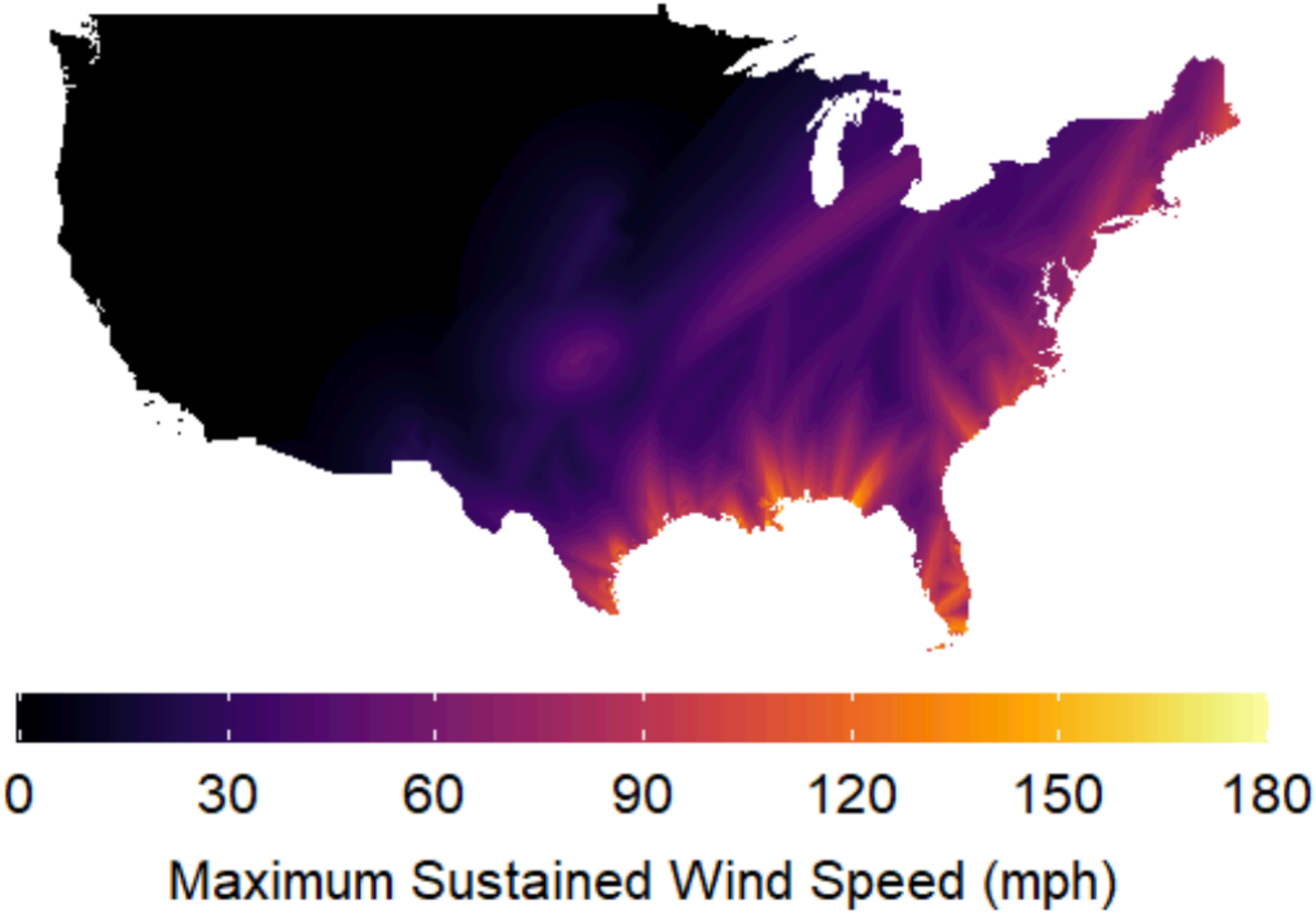
Hurricane Katrina (2005)



All Hurricanes (2005)



All Hurricanes (1969-2019)

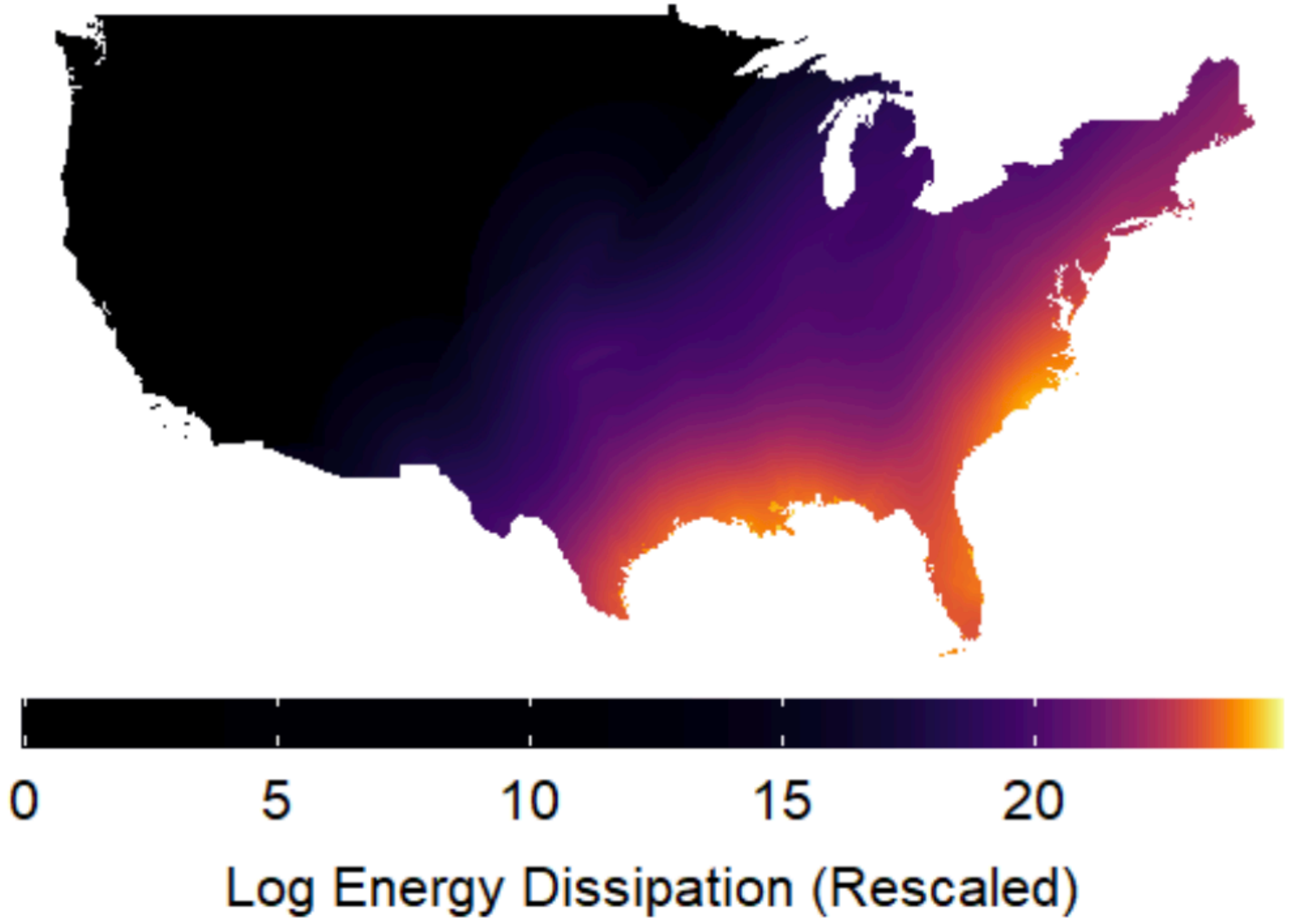


An Alternative Measure of Exposure

- Max windspeed doesn't account for the amount of energy dissipated:
  - Slower moving storms are more damaging because locations are exposed to high windspeed for longer.
  - Multiple storms vs. max across storms.
- Storm damage increases following a power rule (approximately) equal to the cube of max wind speed (Emmanuel, 2005)
- Energy dissipated (m^3/s^2) can be calculated by summing "energy" over time within a cell.

$$\text{Energy Dissipation}_i = \sum_t (V_{\text{max}}(t))^3 * \Delta t$$

"Energy" — All Hurricanes (1969-2019)



Conclusions

- Weather and climate are important for decision-making and economic outcomes, especially in LMICs.
- Vast improvements in measurement and availability of data.
- It’s important to understand the data you are using:
    - Measurement can be endogenous even when physical realizations aren’t
    - Non-linearities are important.
    - Think carefully about your context and what data would be most appropriate.
    - Understand strengths and limitations.
    - Be clear about trade-offs.
    - New tools will make it increasingly easier to process data