Big Data Science

Course: 18-788

Patrick McSharry
patrick@mcsharry.net
www.mcsharry.net
Twitter: @patrickmcsharry

Spring 2023

ICT Center of Excellence
Carnegie Mellon University

Syllabus and Agreement

- Please read the syllabus in full to set expectations.
- Assignment 0 is mandatory for the course.
- For Assigment 0, return a signed agreement stating that you will follow the course rules on reading and sending communications, taking responsibility, following instructions, recitations, deadlines, exceptional circumstances, regrade requests, plagiarism, AIV and potential penalties.
- Assignment 0 will test submission skills and provide feedback.
- All assignments will rigidly apply the rules which enable us to efficiently run the course and to be fair.

Communications

- Please read the instructions on communication protocol in the syllabus.
- For course management and waitlists, contact Megan Oliver: mvoliver@andrew.cmu.edu
- Specific TAs will be appointed to lead on communication for different course components.
- For an exceptional cirsumstance, follow the email protocol (instructor, TAs, faculty/staff).
- We cannot guarantee that emails that do not follow the protocols will be processed.

Big Data Science

WEEK 1A

Time Schedule

Week	A	Pittsburgh	Kigali	B	Pittsburgh	Kigali
1	Tues, Mar 14	08:00 ET	14:00 CAT	Thurs, Mar 16	08:00 ET	14:00 CAT
2	Tues, Mar 21	08:00 ET	14:00 CAT	Thurs, Mar 23	08:00 ET	14:00 CAT
3	Tues, Mar 28	08:00 ET	14:00 CAT	Thurs, Mar 30	08:00 ET	14:00 CAT
4	Tues, Apr 04	08:00 ET	14:00 CAT	Thurs, Apr 06	08:00 ET	14:00 CAT
5	Tues, Apr 11	08:00 ET	14:00 CAT	Thurs, Apr 13	08:00 ET	14:00 CAT
6	Tues, Apr 18	08:00 ET	14:00 CAT	Thurs, Apr 20	08:00 ET	14:00 CAT

Course outline

Week	Lecture A	Lecture B
1	Weather & agriculture	Climate change
2	Climate scenarios	Catastrophe models
3	Social trends	Finance
4	Sentiment analysis	Health
5	Telemedicine	Mobile data
6	Data4Dev	Socioeconomic status

Today's Lecture

No.	Activity	Description	Time
1	Challenge	Agricultural productivity	10
2	Discussion	Historical data	10
3	Case study	Weather and tea	10
4	Analysis	Correlations	20
5	Demo	Index insurance	20
6	Q&A	Questions and feedback	10

Agricultural productivity

- Productivity refers to the amount of agricultural output associated with a particular crop (maize, potatoes, tea etc.).
- For a country, this is often measured as a financial amount, $, or % of GDP per annum.
- The economic output depends on both the quantity produced, number of kilograms, and the price per kilogram.
- The price will usually vary according to market conditions based on supply and demand.

Quantity

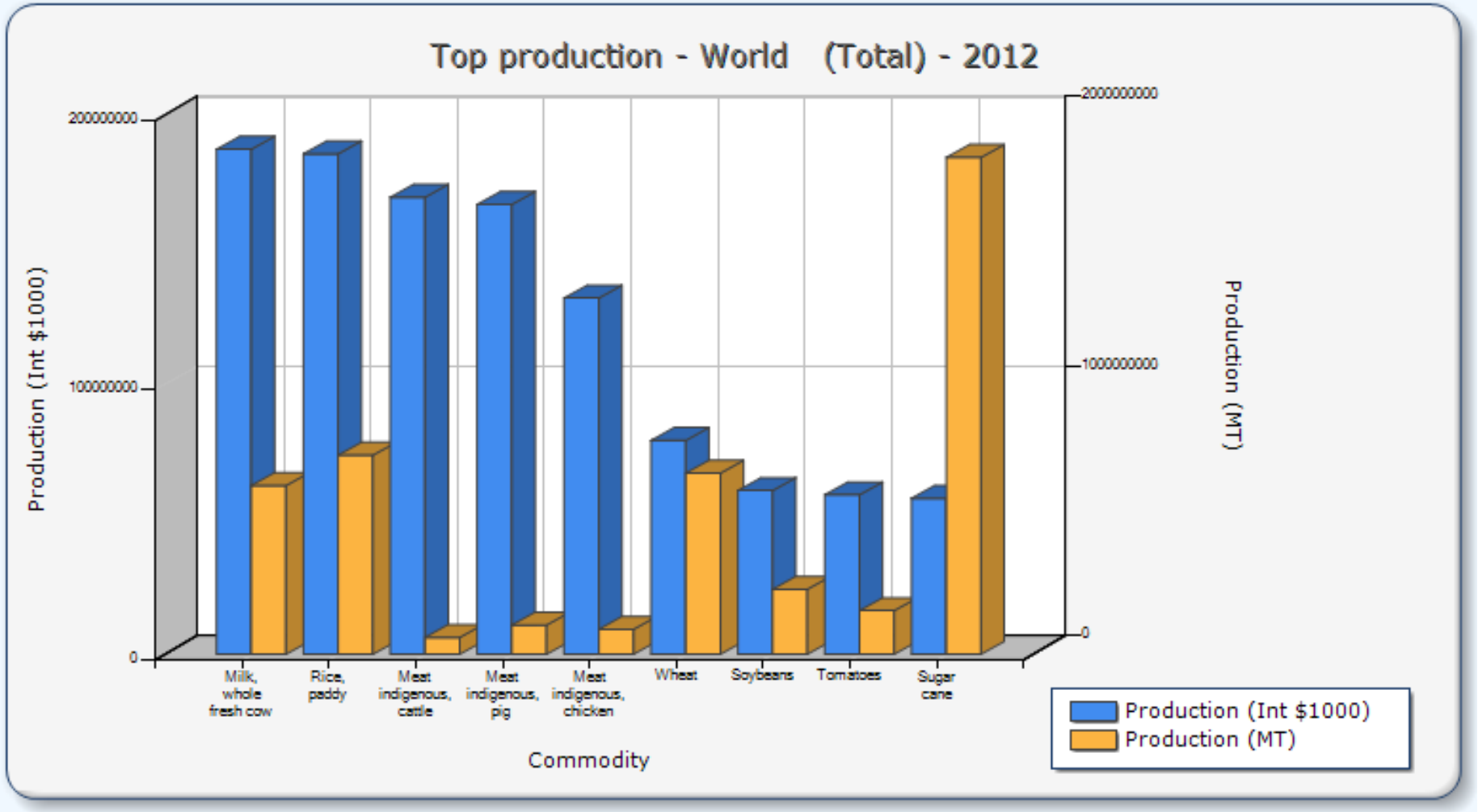
- The quantity is usually measured as the weight of produce, which is obviously more accurate than counting potatoes or green tea leaves.
- However rainfall and humidity can interfere with the weight measurement and is a factor that should be taken into account when comparing productivity.

Most produced foods

- Which food commodity is produced in the largest quantity globally?
  a) Milk
  b) Rice
  c) Sugar cane
  d) Wheat

- **Slido.com #115457**

Production



Source: http://faostat.fao.org

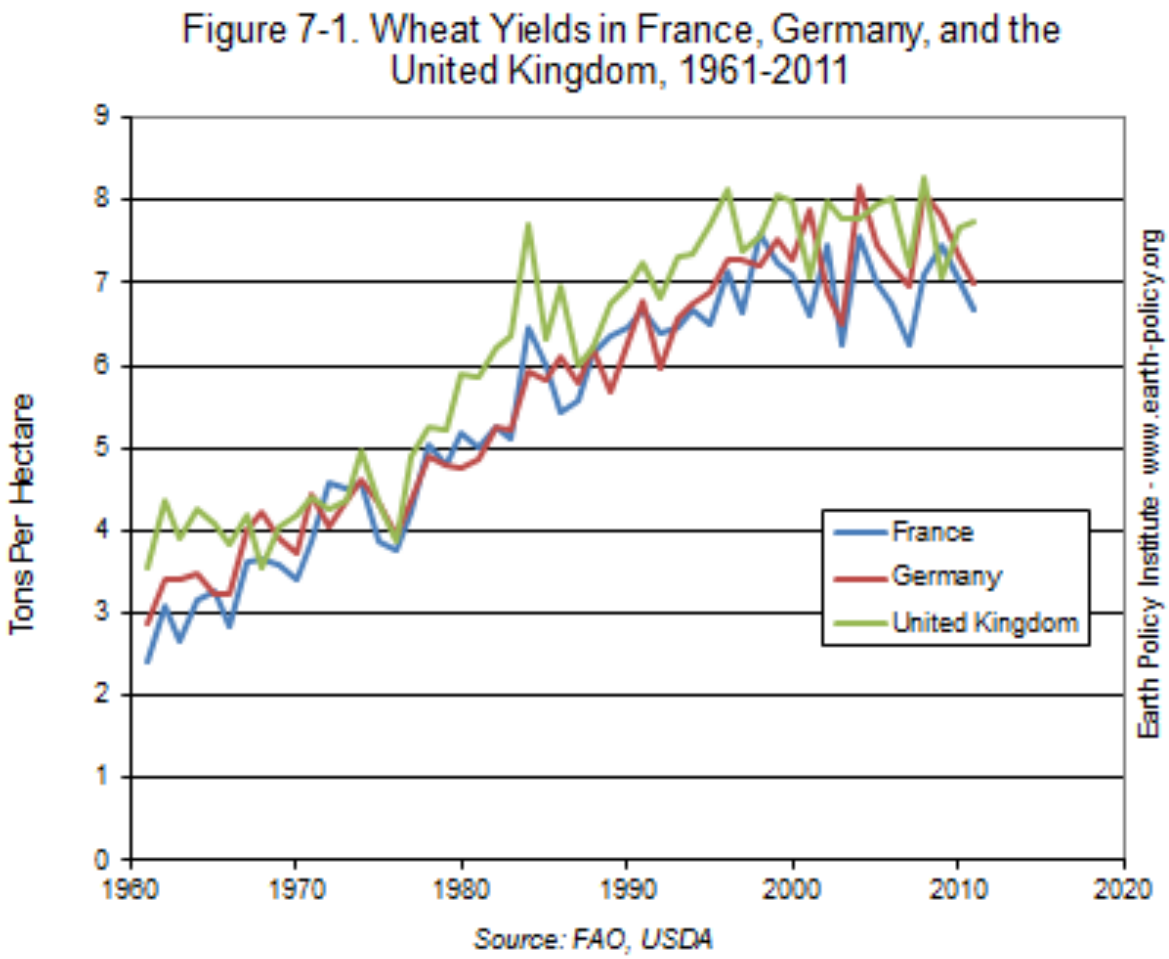
Crop yield

- To make measurements comparable, it is common practice to measure crop yield which is defined as the quantity produced per hectare per year.
- The normalization by area and time duration is important to ensure comparability.
- Yield then tells us how effective a farmer or manager is in producing agricultural goods.

Yield examples

Crop	Yield
Wheat	2.7 – 8.0 tonnes/hectare/annum
Rice	2.2 – 7.5 tonnes/hectare/annum
Coarse grains	1.5 – 9.0 tonnes/hectare/annum
Oilseeds	1.2 – 2.7 tonnes/hectare/annum

Improving yield



Agriculture variables

- Which variables would you collect in order to understand and predict crop productivity?

- **Slido.com #115457**

Important factors in agriculture

- Water and irrigation
- Variety
- Fertilizer
- Pests and pesticides
- Weather extremes
- Climate change

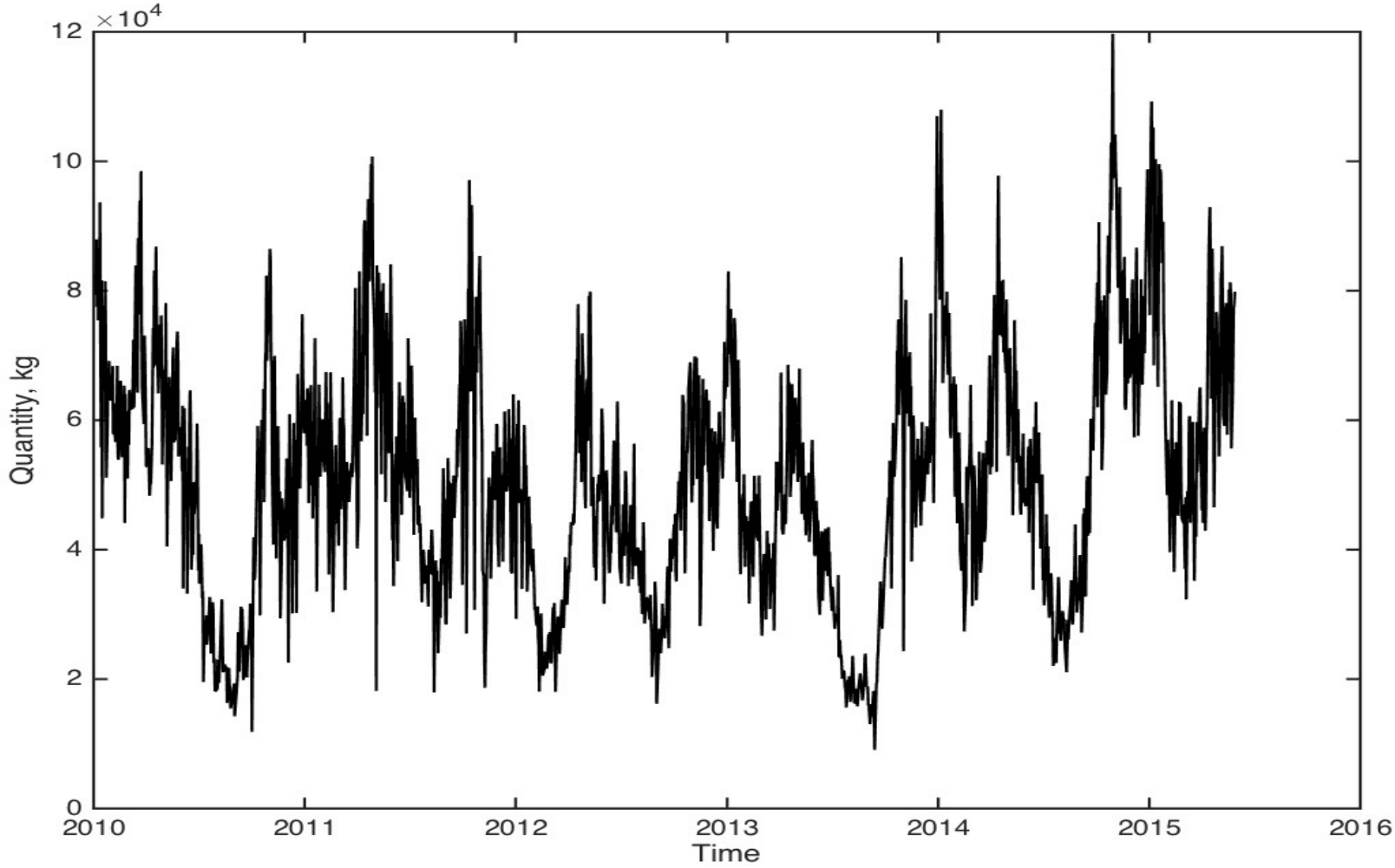
Tea

- Tea is an aromatic beverage commonly prepared by pouring hot or boiling water over cured leaves of the *Camellia sinensis*, an evergreen shrub native to Asia.
- After water, it is the most widely consumed drink in the world.
- Types of tea (pictured above):
  - White: Wilted and unoxidized
  - Green: Unwilted and unoxidized
  - Oolong: Wilted, bruised, and partially oxidized
  - Black: Wilted, sometimes crushed, and fully oxidized (called 'red tea' in China)

Tea in Rwanda

- Agriculture makes up on third of the economy.
- Tea is the second biggest export commodity (after coffee) and revenues were $102 million in 2020.
- There are 25,000 hectares for producing tea at 13 factories, supporting 60,000 households.
- Yields range from 7.5 to 10.5 tonnes/ha/year.

Daily quantity of tea



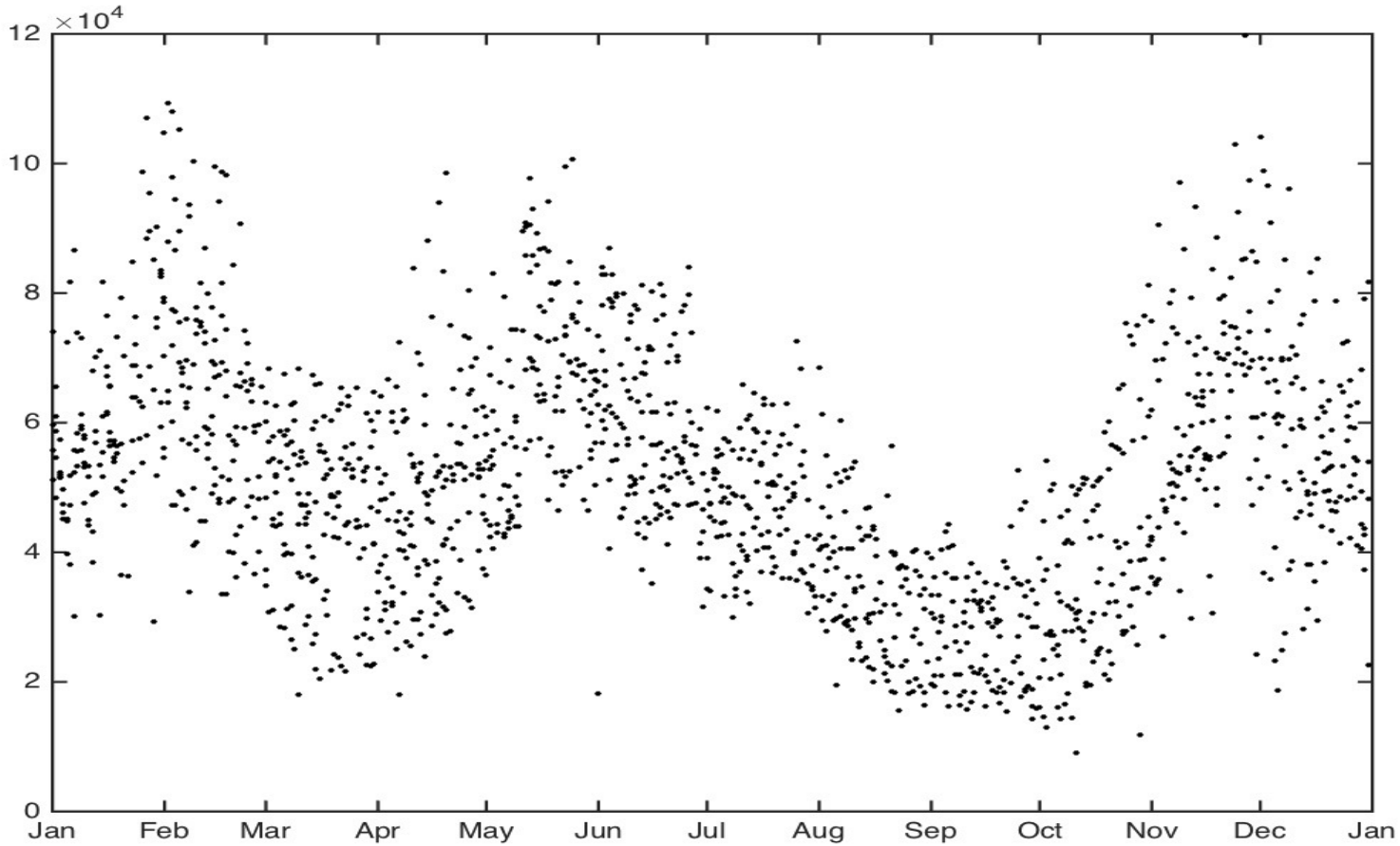
Visualization Quiz

- How can we visually identify intra-annual seasonality in the daily tea quantity time series?

  a) Day of week plot
  b) Day of month plot
  c) Month of year plot
  d) Day of year plot

- **Slido.com #115457**

Quantity of tea by day of year



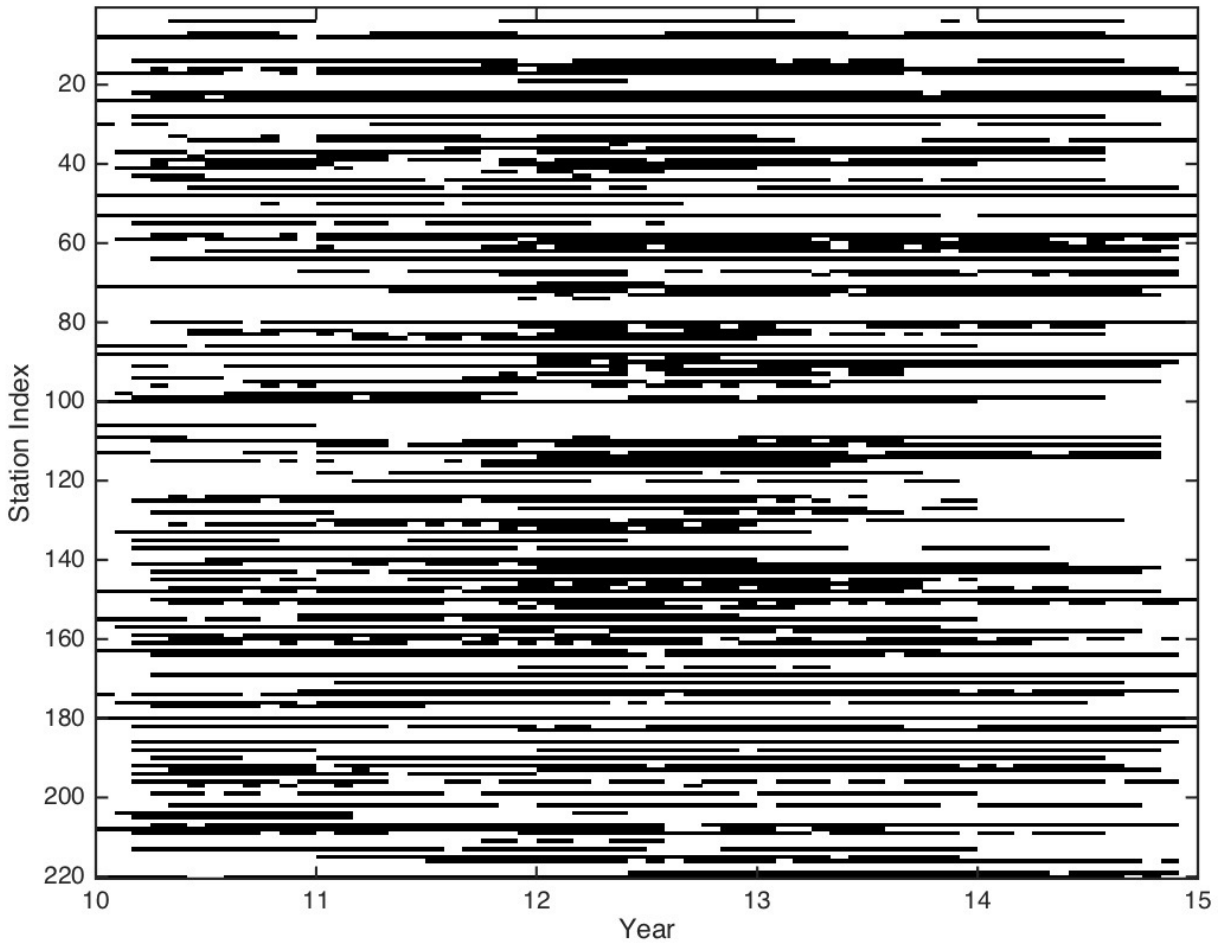
Weather and tea

- Tea productivity, quantity and yield can be measured on a daily basis.
- Tea requires sufficient water and therefore rainfall is an important weather variable.
- Tea does not perform well when faced with hot temperatures and therefore the maximum temperature is an important variable.
- Hot dry conditions are detrimental for tea productivity.

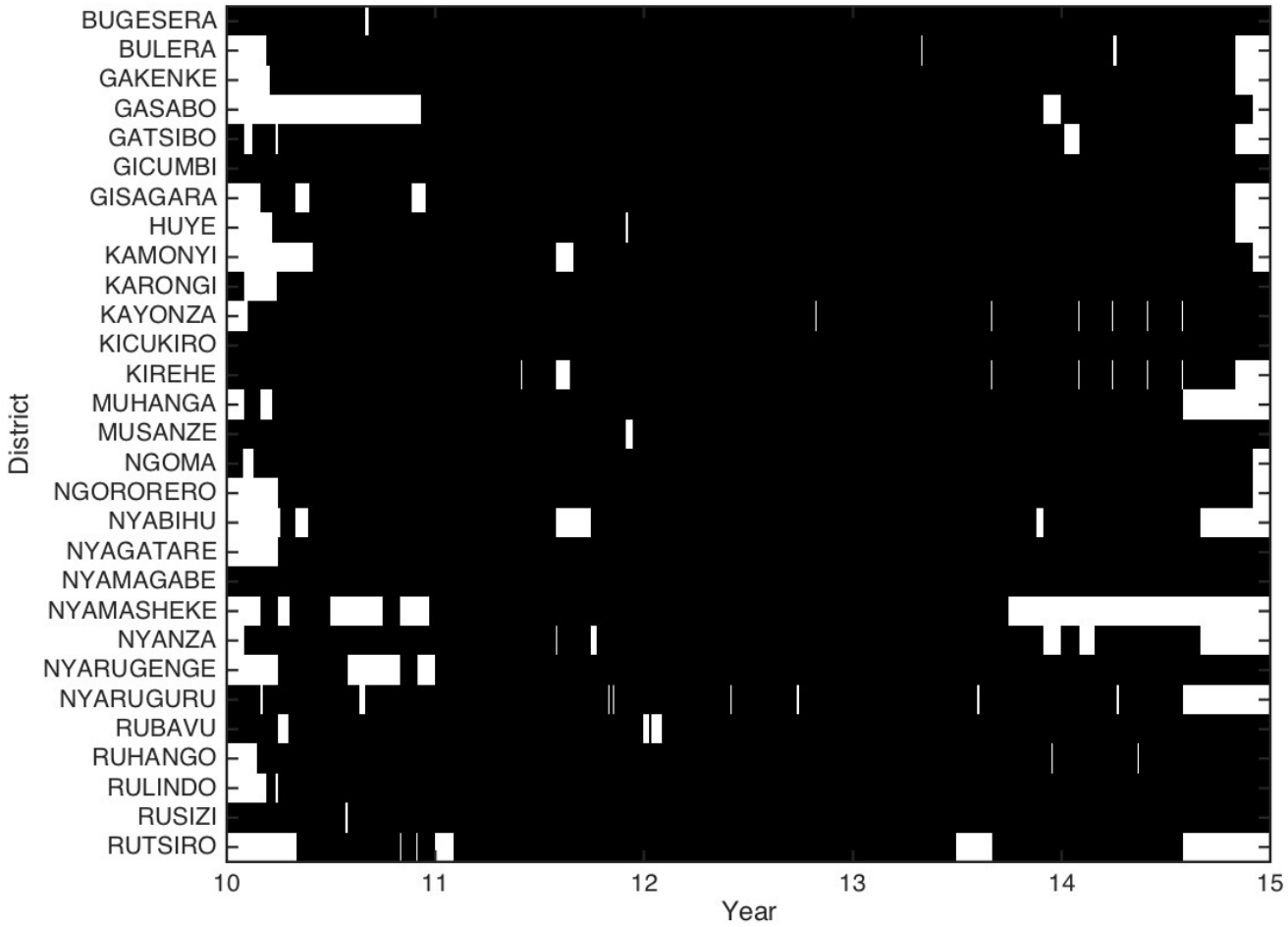
Weather data

- Weather data is often available from both private weather stations at tea factories and nearby Rwanda Meteorological Agency (RMA) stations.
- How reliable is the ground weather data?
- How near does a station need to be in order to obtain accurate observations?

Availability of rainfall data by station



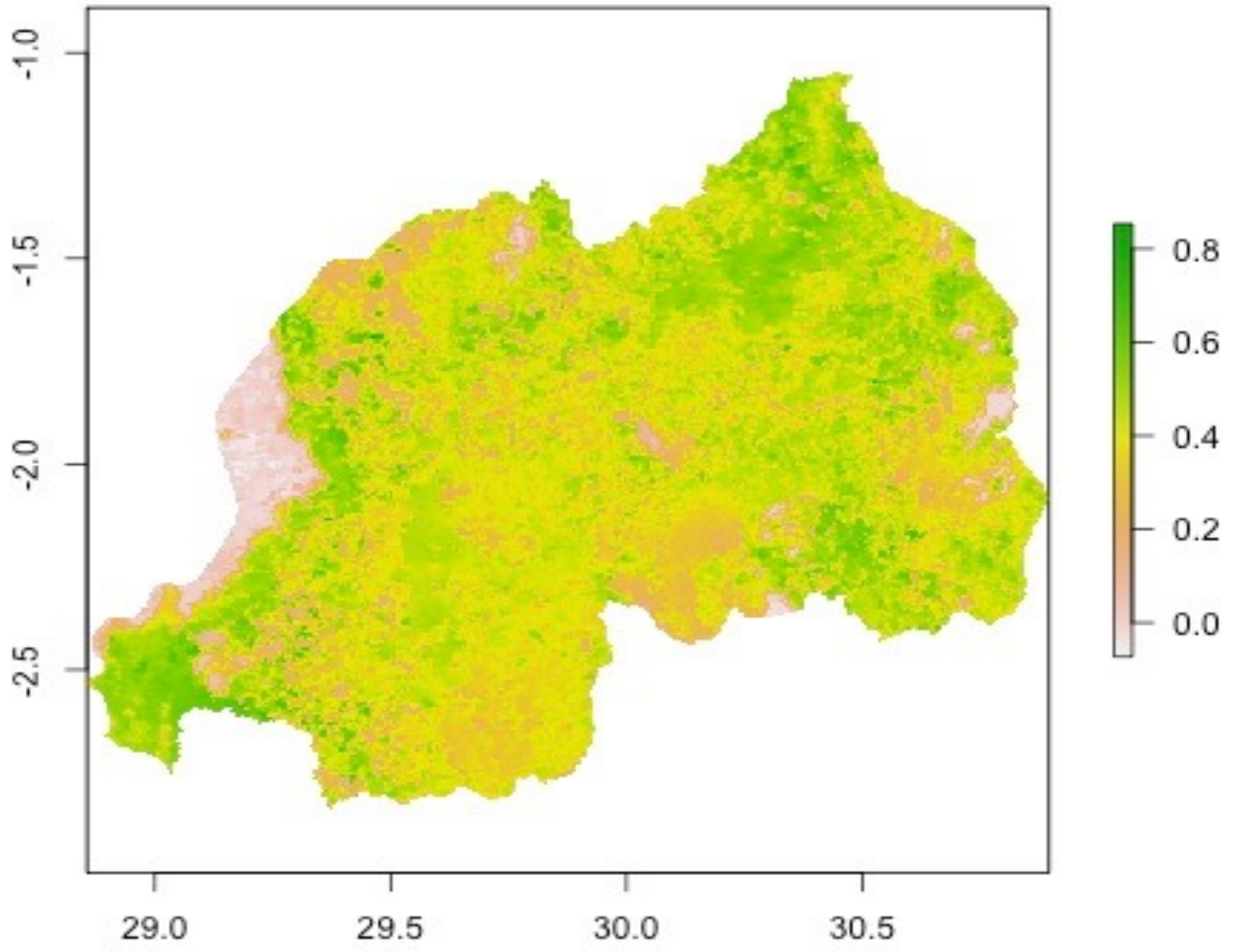
Availability of rainfall data by district



Satellite imagery

- Satellite data can provide estimates in areas far from weather stations, help to verify weather observations, and fill in gaps of missing data in recorded data.
- Additionally, satellites can capture data that would be difficult or impossible to record by hand, such as the Enhanced Vegetation Index (EVI), which is an optimized vegetation index designed to enhance the vegetation signal with improved sensitivity in high biomass regions.
- Other key features include surface and root moisture.
- These supplementary sources can provide further features with explanatory power in predicting yields.
- Although satellite data may add complexity to the quantitative model, it can greatly improve its predictive accuracy.

Satellite Enhanced Vegetation Index



Source: McSharry, Spray, Swartz (2016). Rwanda using MODIS for January 9, 2010

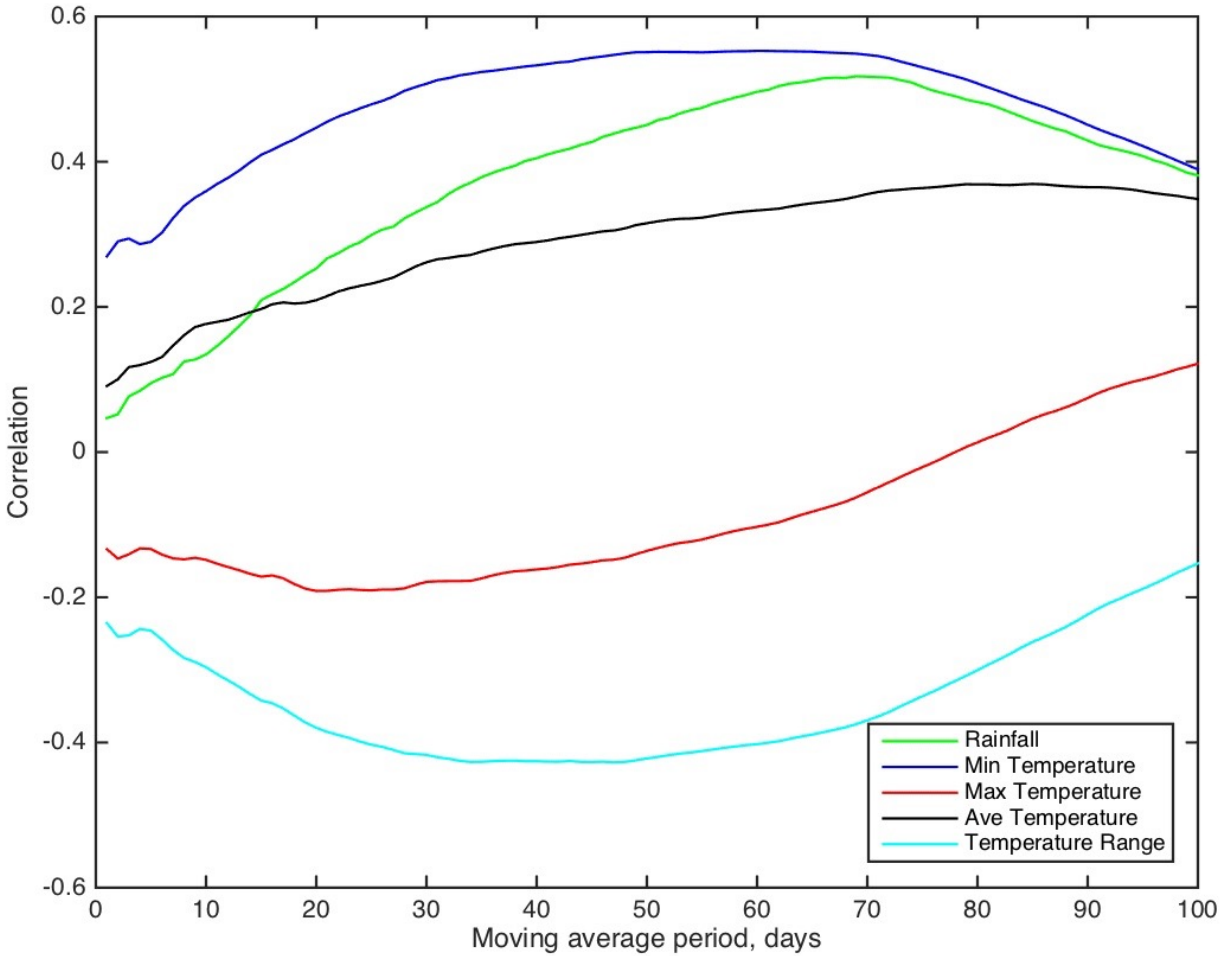
Weather data

- Which of the following provided the most accurate forecasts of tea productivity?

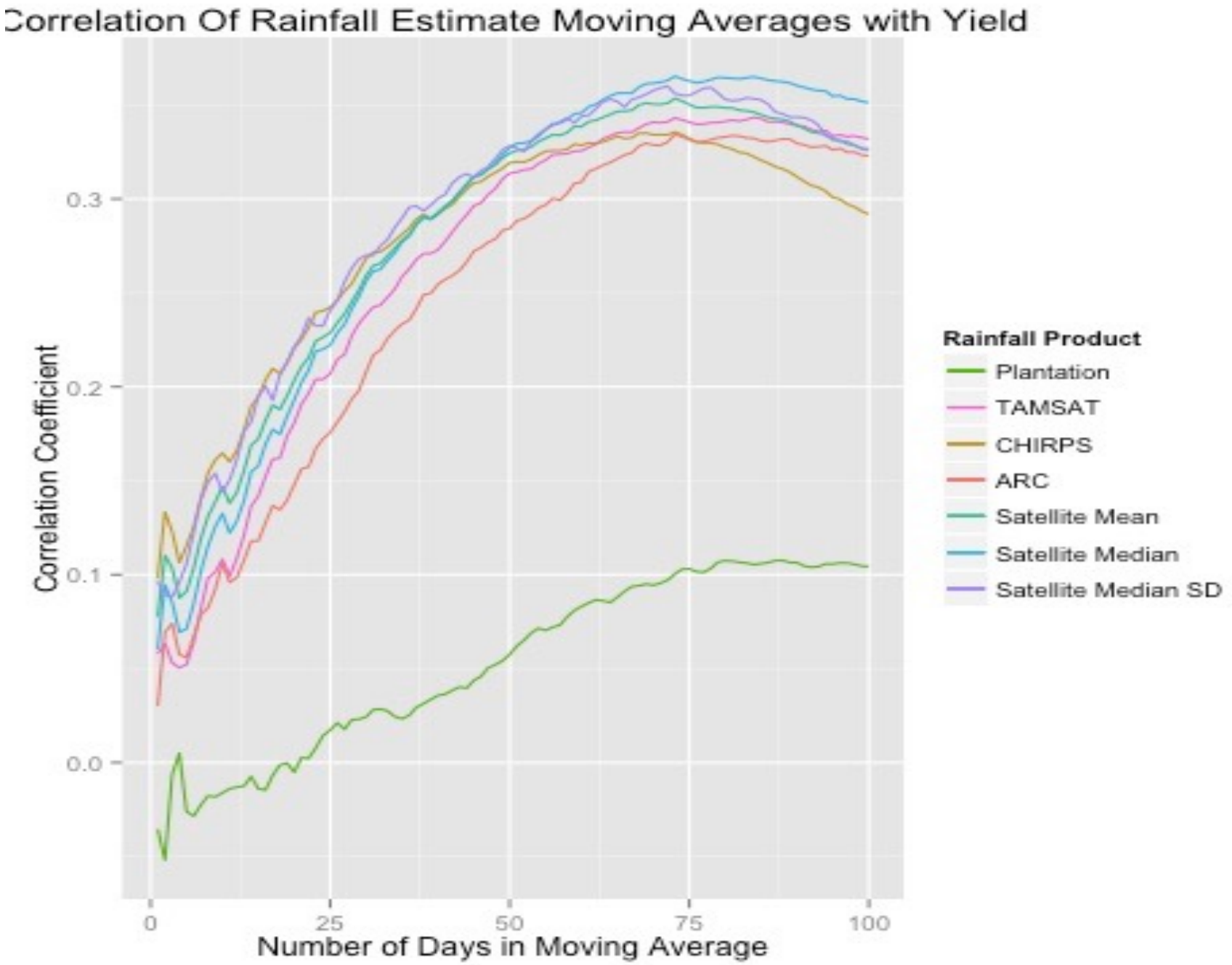
  a) Weather station at tea factory
  b) Weather station at nearest airport
  c) Satellite-based weather data

- **Slido.com #115457**

Correlations



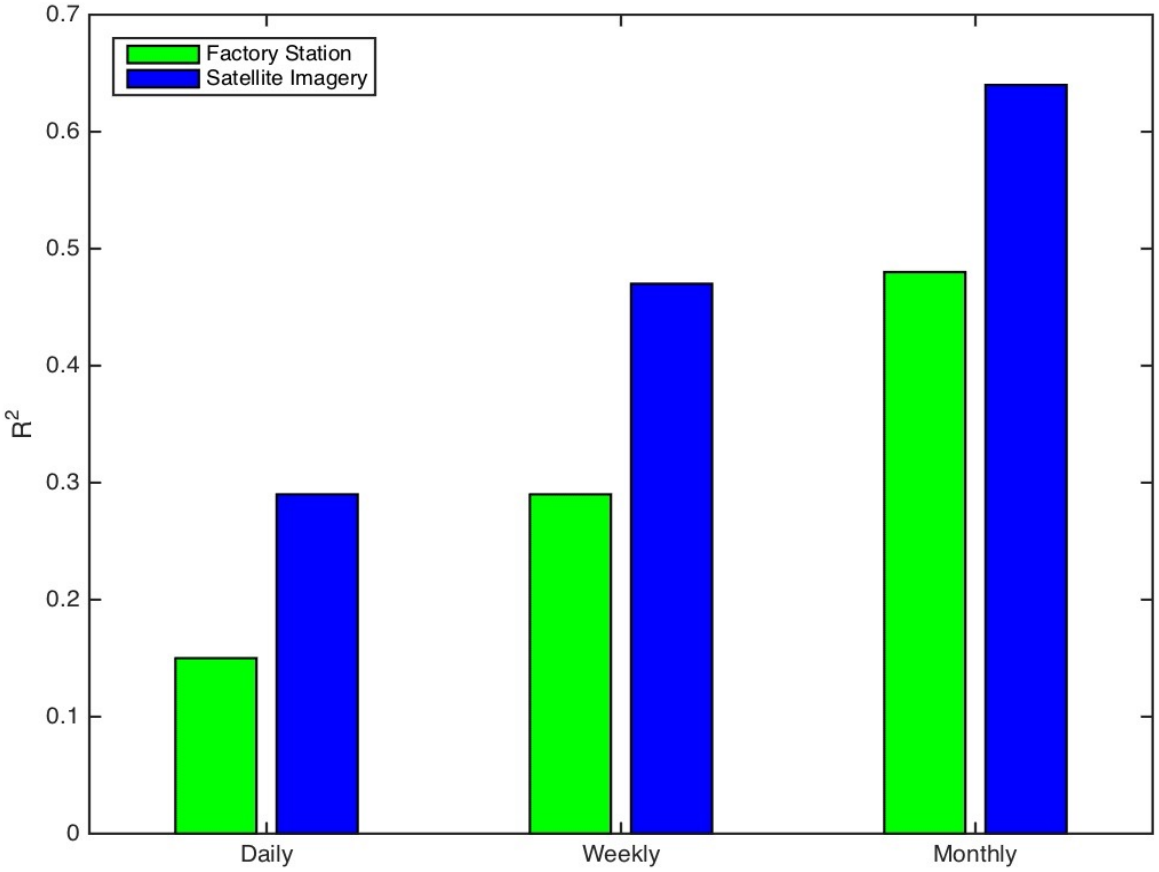
Source: McSharry, Spray, Swartz (2016).


Correlation Of Rainfall Estimate Moving Averages with Yield
Correlation Coefficient
0.3
0.2
0.1
0.0
0
25
50
75
100
Number of Days in Moving Average
Rainfall Product
Plantation
TAMSAT
CHIRPS
ARC
Satellite Mean
Satellite Median
Satellite Median SD

Forecasting tea productivity

- Monthly forecasts based on satellite imagery explained 57% of the variability on average.
- We find satellite imagery outperforms local ground weather stations by 43%.
- Key satellite features were surface moisture content, median rainfall and root zone moisture content.
- Vegetation greenness and the enhanced vegetation index were also relevant.

Forecast Performance



- Forecast performance using satellite imagery is better than factory weather station.
- Forecast performance improves with larger temporal intervals.

Index insurance

- Index insurance products pay out when an index that is correlated with crop performance exceeds a set threshold level.
- The product relies on the quality of the data and the accuracy of the model used in constructing the index.
- Indices are usually based on independently verified weather data from nearby stations or satellite imagery.

Index Insurance (pros and cons)

Advantages of Index Insurance	Disadvantages of Index Insurance
Adverse selection and moral hazard are minimized	Basis risk can arise if an individual insured's loss experience does not correlate with the index payouts
Administrative costs are reduced	There is a lack of high quality weather and yield data in many developing countries
Timeliness of payouts after an event is improved	In most developing countries, there is not an enabling environment for index insurance
	Products are technically complex and yet they have to be packaged in a way that is understandable by most people

Cumulative rainfall index

- A cumulative rainfall index is constructed using

$$X_t = \sum_{n=t-k}^{t} R_n$$

- where R_t is the rainfall on day t and k is the number of days of accumulation.
- It is often the case that monthly intervals are used in calculating the index.

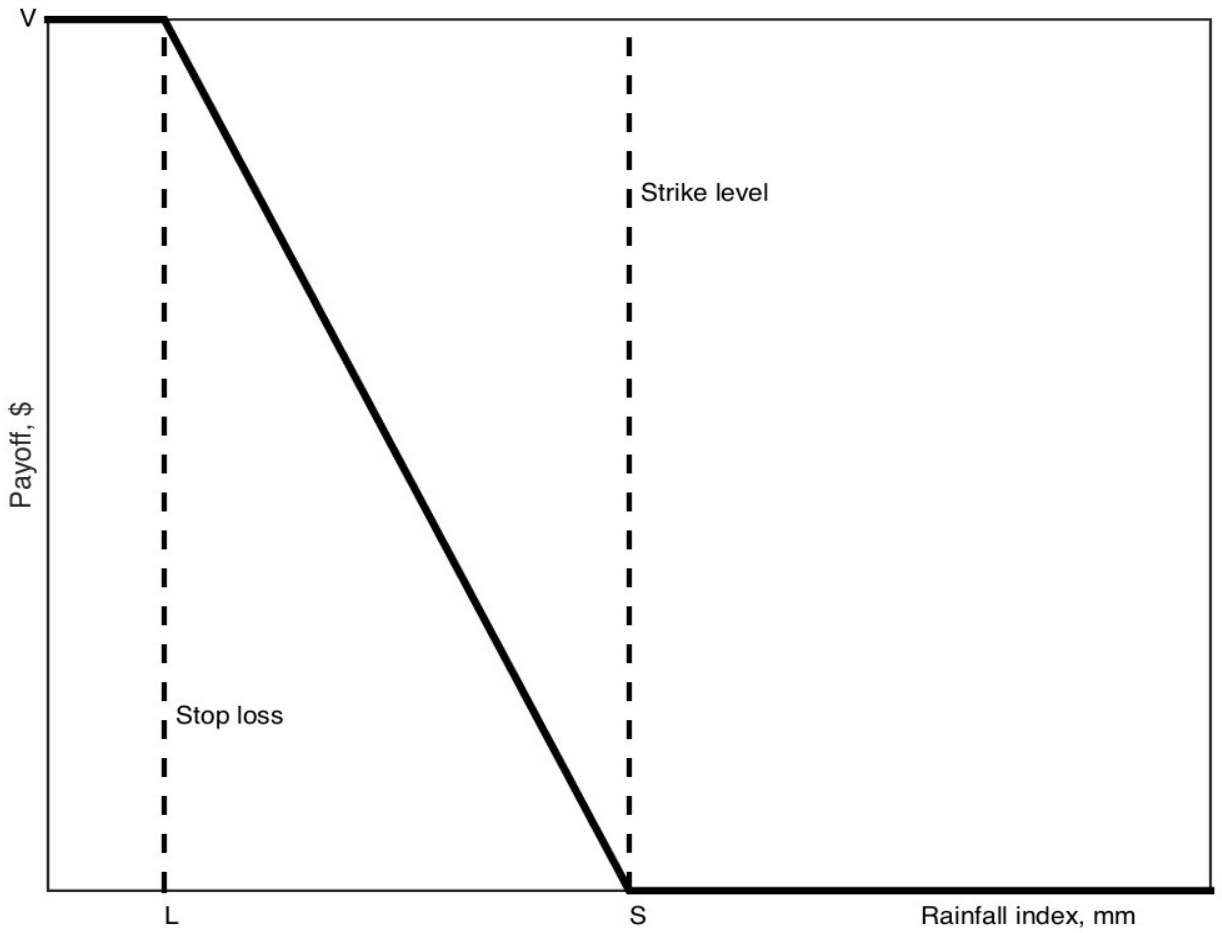
Payoff function

- The insurance payoff, P, also known as a put option, can be expressed as a function of the index X at maturity,

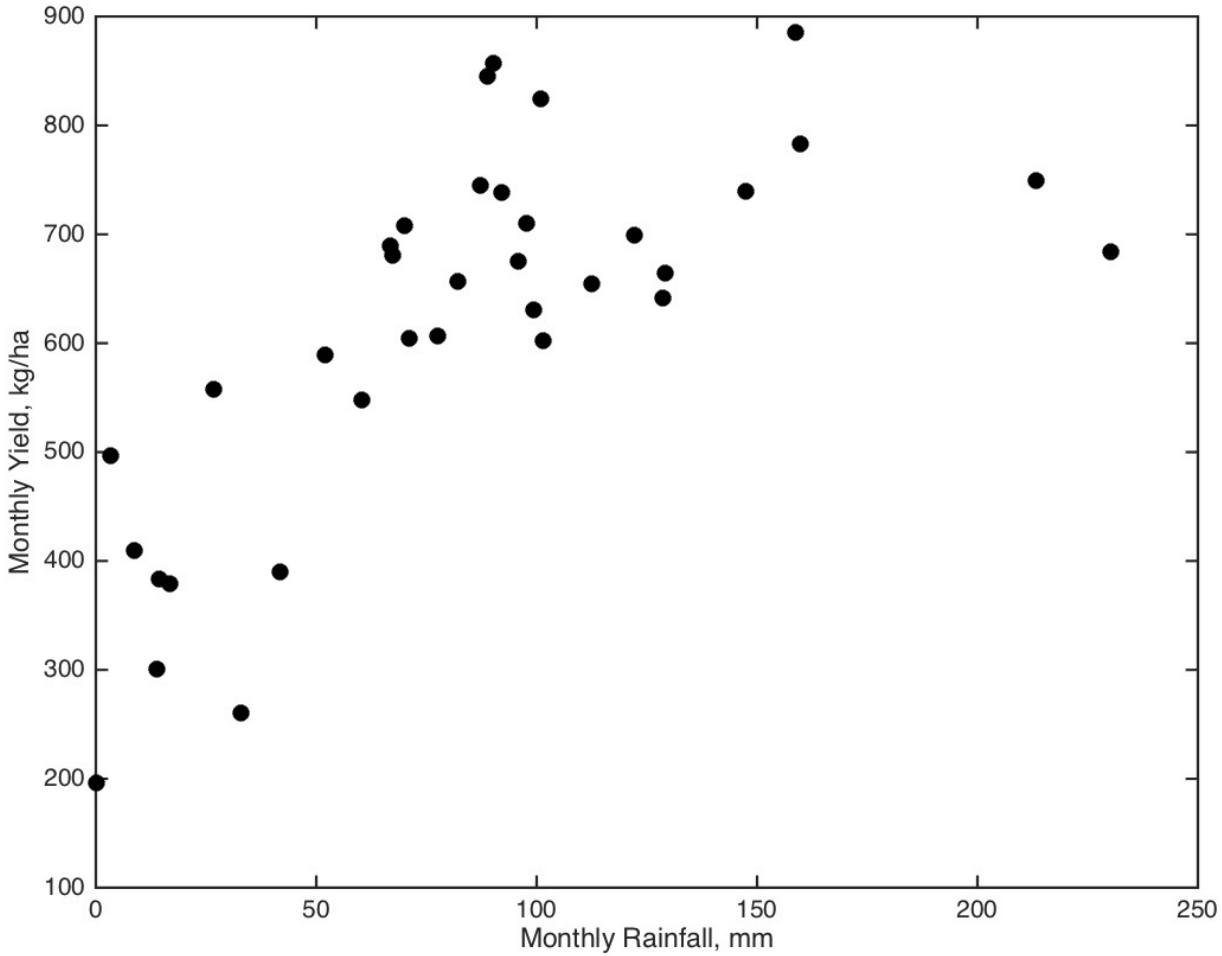
$$P = \begin{cases} V, & X \le L \\ V[1-(X-L)/(S-L)], & L < X < S \\ 0, & X \ge S \end{cases}$$

- where S is the strike level, L is the stop los level and V is the tick size.
- If X is higher than the strike S, there is no payoff but if X falls below S, a payoff is generated.
- A maximum payoff of V occurs when X falls below L.

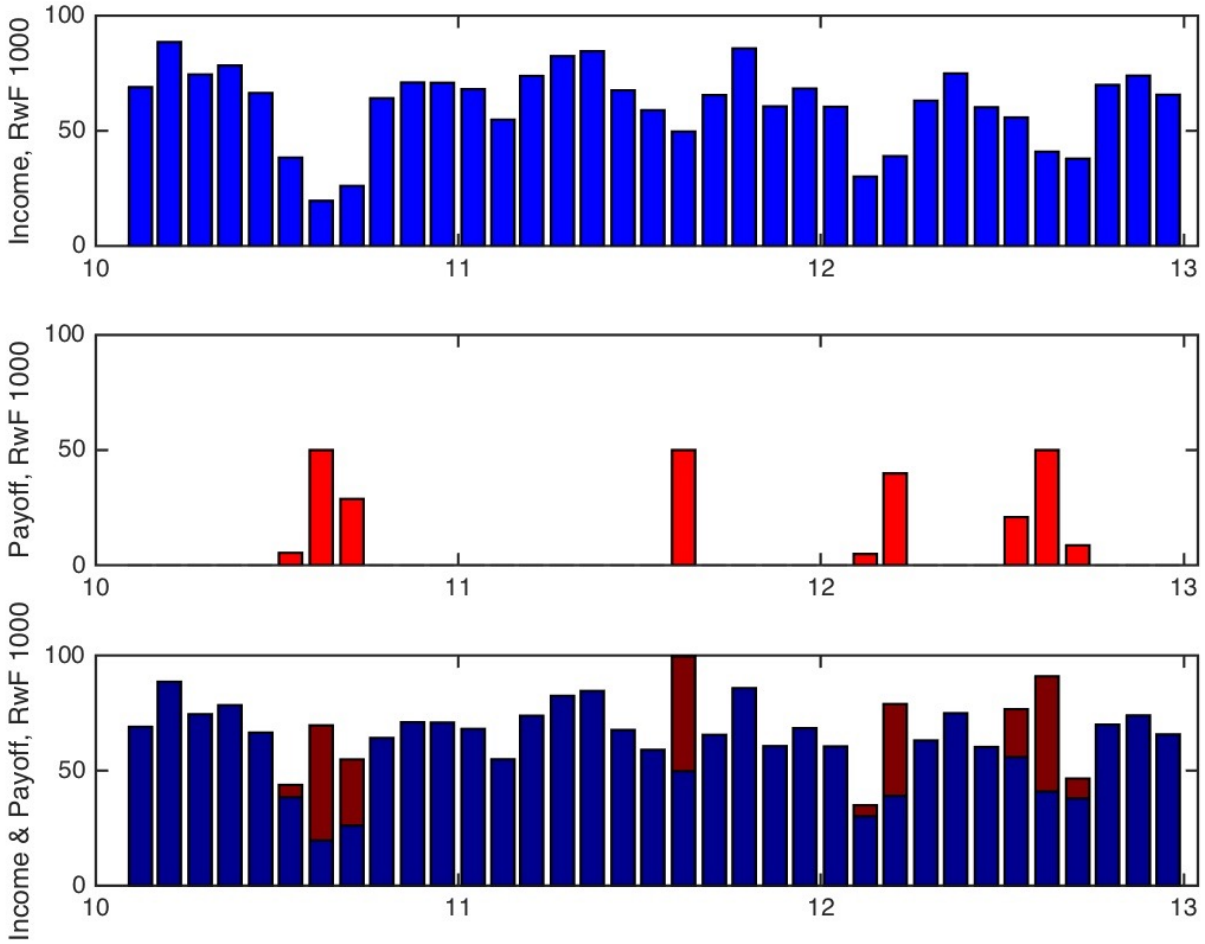
Payoff function



Monthly Yield and rainfall

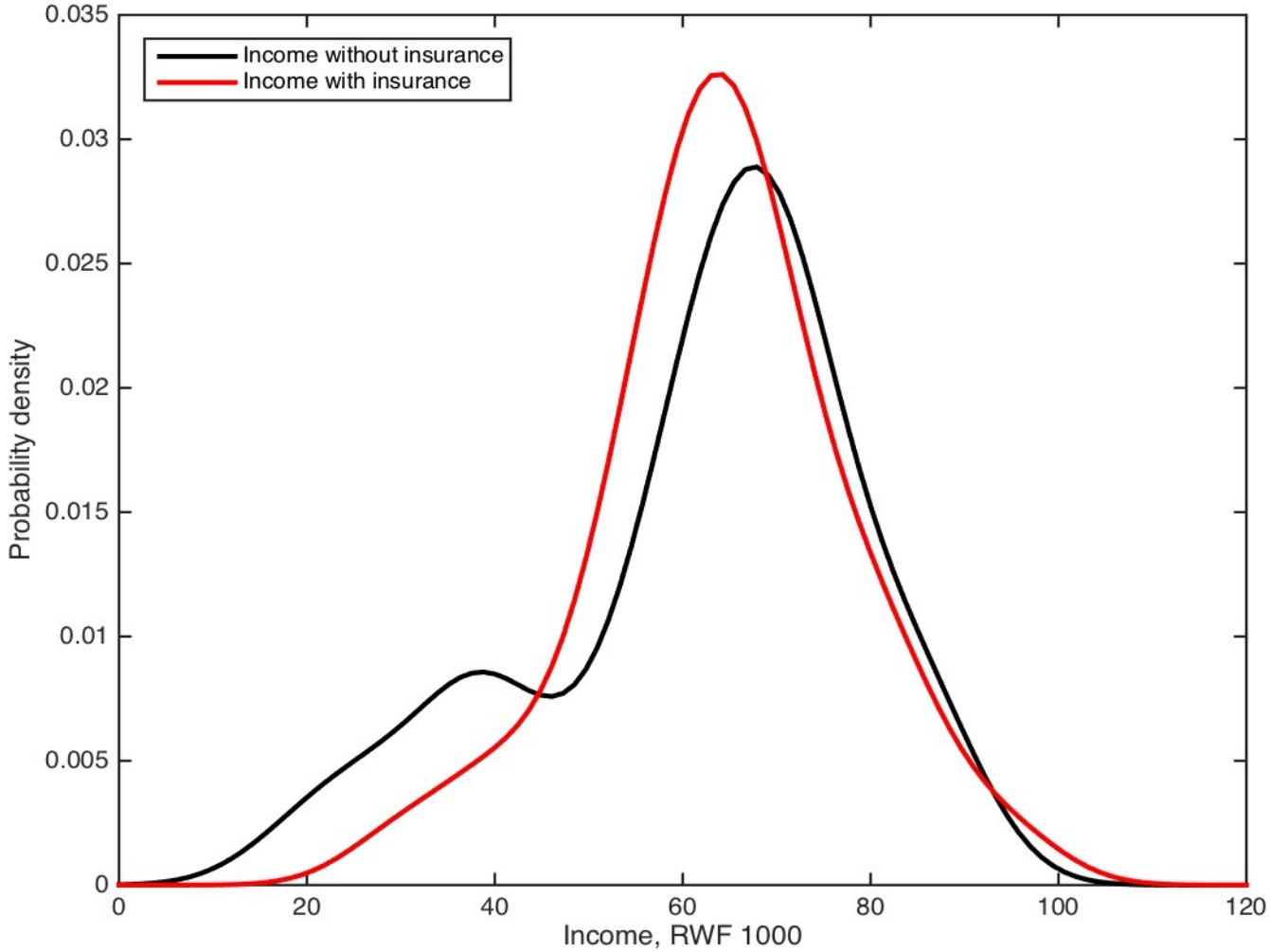


With and without IBWI



Index design: L = 10mm; S = 50mm; V = 50,000 RwF; Price = 100 RwF per kg green leaf

Income distributions



Simulations of an Index insurance product at Mulindi factory increased incomes by 4% and reduces volatility from 28% to 21%.

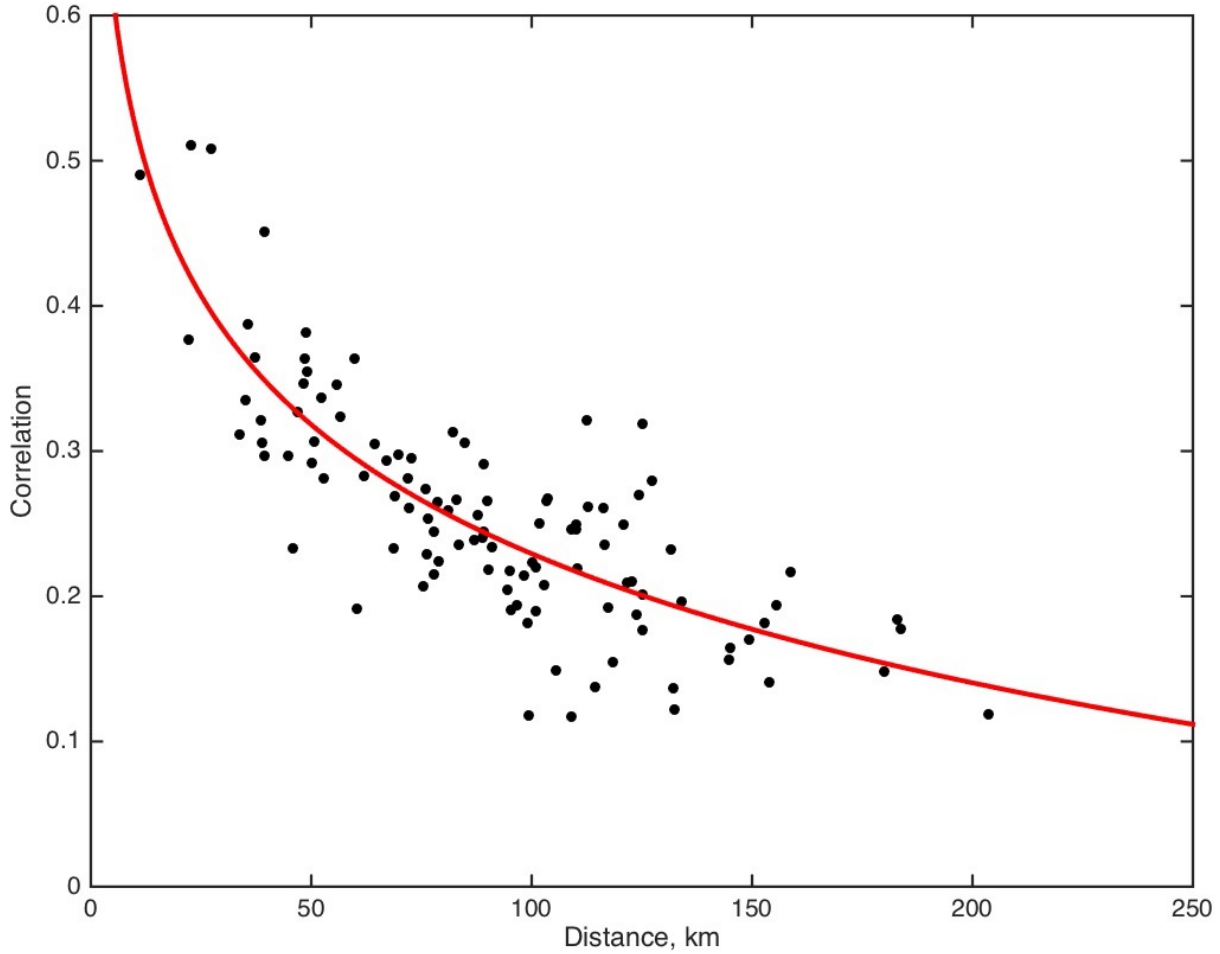
Basis risk

- The design and construction of the index is arguably the most important component of an index based insurance product.
- If the payouts triggered by the index do not represent the actual losses experienced, the product can never be fit for purpose.
- The mismatch between insurance payouts and claims (actual losses) is known as basis risk.

Cause of basis risk

- Basis risk can be caused by a number of factors including geographical issues, poor data quality, short historical records and inadequate index design.
- In essence, basis risk will arise if the quantitative model that underlies the index does not provide sufficient predictive performance.

Rainfall correlation versus distance

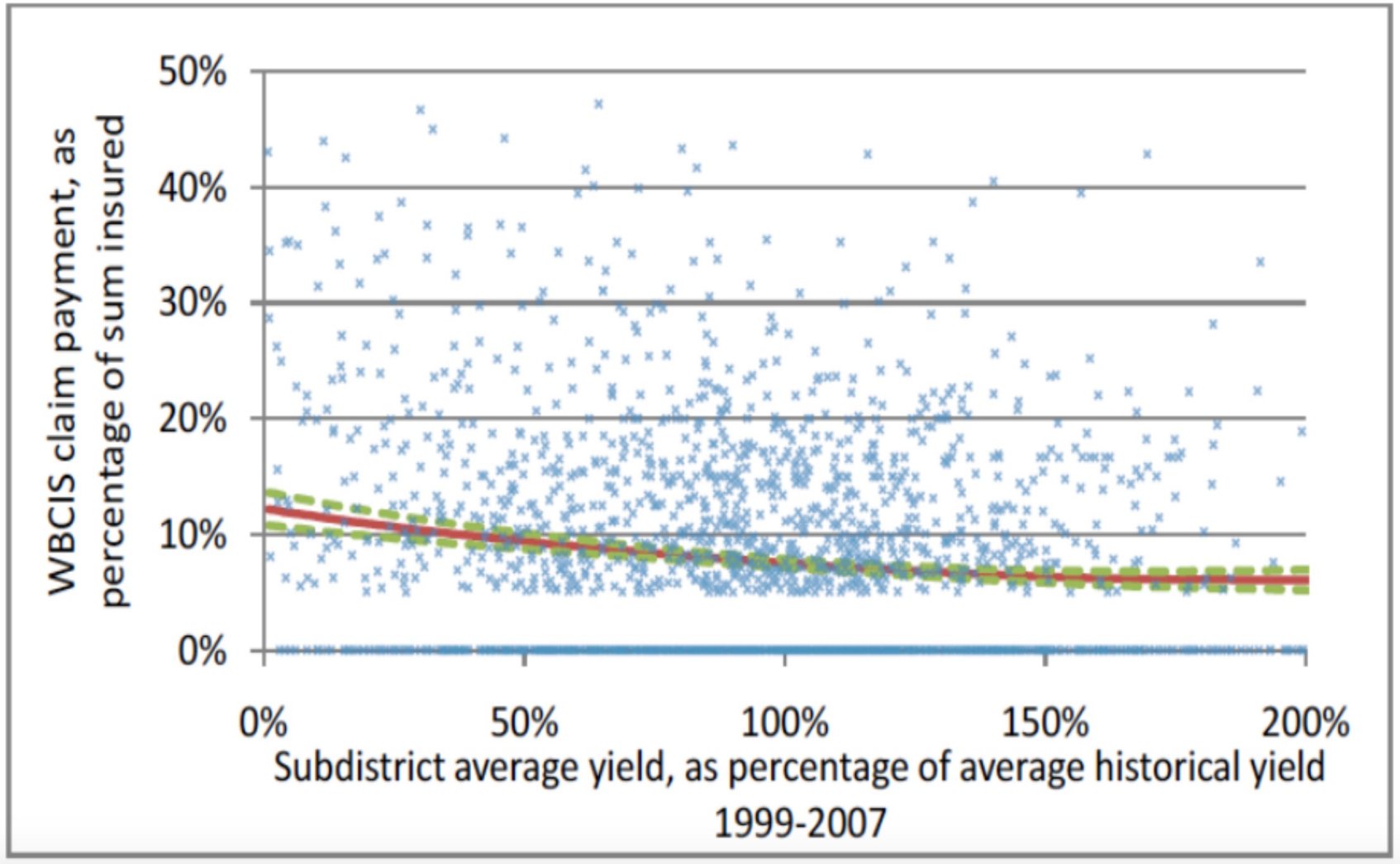


Evidence of microclimates

- There exists a lot of anecdotal evidence of microclimates in Rwanda.
- The land of 1,000 hills is certainly going to be a challenge for observing rainfall with different rainfall patterns on neighbouring hills.
- This represents a substantial cause of basis risk.
- We tested this empirically and constructed estimates of correlation for nearby privately run rain gauges at a tea estate.
- The correlation declines rapidly with short distance indicating that an index product would need to rely on many sensors.

Estate 1	Estate 2	Distance, km	Correlation
Cyohoha	Usine	1.63	0.74
Cyohoha	Rukeri	3.57	0.68
Usine	Rukeri	2.06	0.78

Payments versus Yield



Correlation between yields and claim payment is just -13%.

Source: Clarke et al. (2012) Weather Based Crop Insurance in India. World Bank Policy Research Working Paper No. 5985.

Impact of basis risk

- *“I would be very pleased to stand corrected but my sense is that most unsubsidised weather index insurance programs for smallholder farmers would be tomorrow's mis-selling scandals if developing country regulators were as competent as developed country regulators. Available statistical evidence suggests that these products are essentially just expensive lottery tickets, increasing, not reducing, the vulnerability of farmers.”* - Dr Daniel Clarke, Work Bank.

Explaining basis risk

- Clarke et al (2012) point to three explanations for this low correlation:
- (1) prevalence of pests, disease, wind, flood, frost, hail and localised weather;
- (2) index capturing weather information at the wrong periods; and
- (3) poor historical yield data.

Index insurance for Tea

- Great potential to use index insurance
- Index should be based on satellite imagery and employ sophisticated machine learning techniques
- Multiple satellite products improved the accuracy
- Ground stations suffer from geographical distance, missing data, sensor failure and poor quality.
- Recommend using a third party to evaluate the index and quantify basis risk.

Big Data Science

WEEK 1B

Course outline

Week	Lecture A	Lecture B
1	Weather & agriculture	Climate change
2	Climate scenarios	Catastrophe models
3	Social trends	Finance
4	Sentiment analysis	Health
5	Telemedicine	Mobile data
6	Data4Dev	Socioeconomic status

Assignments

- Process: Piazza > Recitation > TAs > Patrick
- Issues: Classes – Recitations – Early questions
- Research assignments – require critical thinking and exploratory analysis
- Check “Data Inference & Applied Machine Learning” syllabus for assumed prior knowledge

Assignment 1

- Satellite estimates for rainfall and vegetation
- 1. Loading the data
- 2. Visualizing the data
- 3. Summary statistics
- 4. Correlation versus distance (see Lecture 1A)
- 5. Temporal synchronization – we need rainfall and vegetation to be observed on the same time periods for a scatter plot

Assignment 1 - continued

- 6. Feature extraction – how does rainfall affect vegetation. Might there be a delay?
- 7. Feature extraction – is there a dependence on the rainfall over a certain period?
- 8. Is there evidence for a nonlinear model?
- 9. Use cross-validation to select the best feature
- 10. Finally, what about the best model structure to use (linear, nonlinear, nonparametric). Rank models using R^2.

Today’s Lecture

No.	Activity	Description	Time
1	Challenge	Global warming	10
2	Discussion	Open data policy	10
3	Case study	UK Met Office	10
4	Analysis	Trend analysis	20
5	Demo	Climatefrontier.com	20
6	Q&A	Questions and feedback	10

Global warming

- Global warming is now widely accepted as a dangerous threat to society according to the Intergovernmental Panel on Climate Change (IPCC), the scientific intergovernmental body that supports the United Nations Framework Convention on Climate Change (UNFCCC).

- In its Fifth Assessment Report (AR5), the IPPC concluded that **human influence on the climate system is clear** and that it is **extremely likely** (probability greater than 0.95) that **human influence was the dominant cause of global warming** between 1951-2010 (IPCC, 2014).

- Climate model projections summarized in the IPCC report indicate that during the 21st century the global surface temperature is likely to rise a further 0.3 to 1.7 °C for their lowest emissions scenario using stringent mitigation and **2.6 to 4.8 °C** for their highest scenario.

Solving climate change

- What would you do to solve climate change?

- **Slido.com #261096**

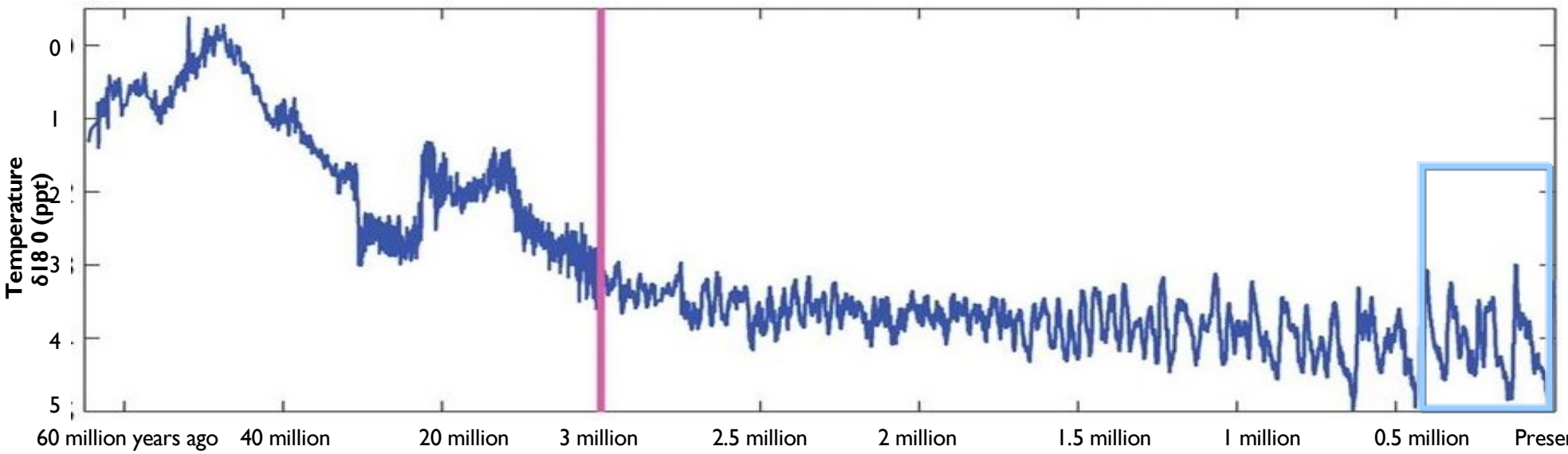
Climate change and the 75% problem

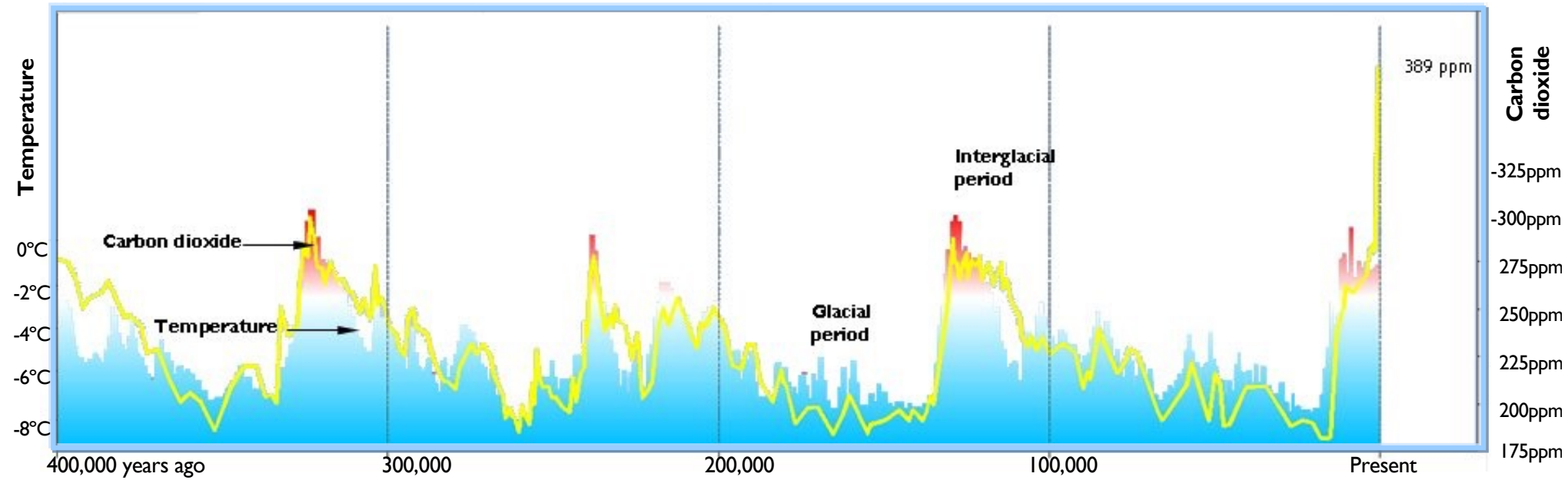
- Making electricity is responsible for only 25% of all greenhouse gas emissions each year.
- To prevent the worst effects of climate change, we need to get to zero net greenhouse gas emissions in every sector of the economy within 50 years—and as the IPCC recently found, we need to be on a path to doing it in the next 10 years. That means dealing with electricity, and the other 75% too.

Sector	Key themes
Electricity (25%)	Renewables, backup, efficient grids
Agriculture (24%)	Cattle, deforestation
Manufacturing (21%)	Plastic, steel, cement
Transportation (14%)	Low emission cars, planes, ships
Buildings (6%)	Airconditioning, insulation

https://www.gatesnotes.com/energy/my-plan-for-fighting-climate-change

OBSERVED GLOBAL TEMPERATURES





Source: Fedorov et al. Science 2006, 312, 1485

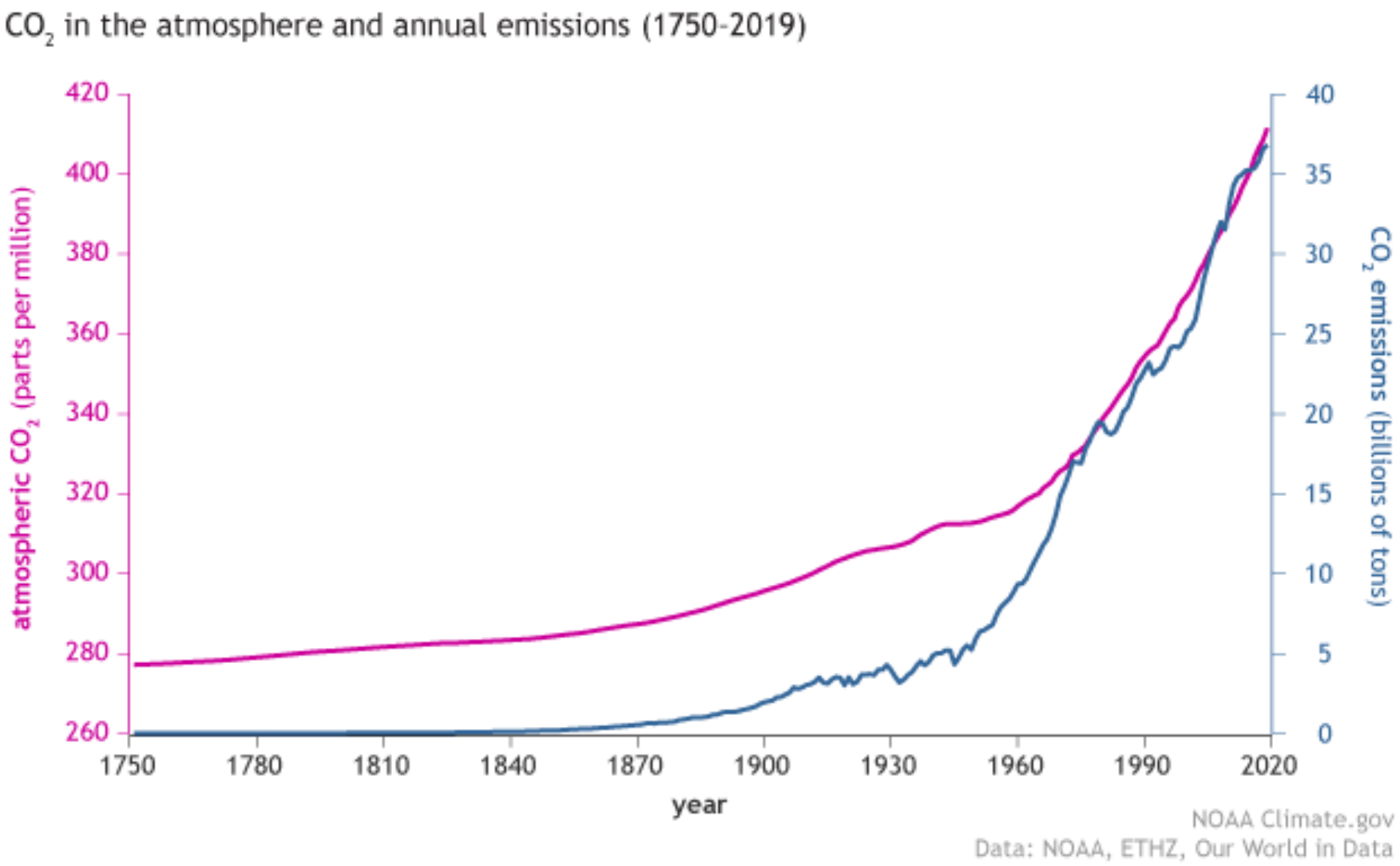
Source: ML Design. From "The Complete Ice Age: How Climate Change Shaped the World" edited by Brian Fagan, Thames & Hudson Ltd., London, 2009

Carbon dioxide

- The amount of carbon dioxide in the atmosphere has increased by **more than 40%** since the start of the Industrial Revolution, from 280 ppm in the mid-18th century to over 412 ppm today.

- Global carbon emissions from fossil fuels have **increased by about 90% since 1970**, with emissions from fossil fuel combustion and industrial processes contributing about 78% of the total greenhouse gas emission increase from 1970 to 2011 (Boden et al., 2015).

CO_2 emissions and atmosphere

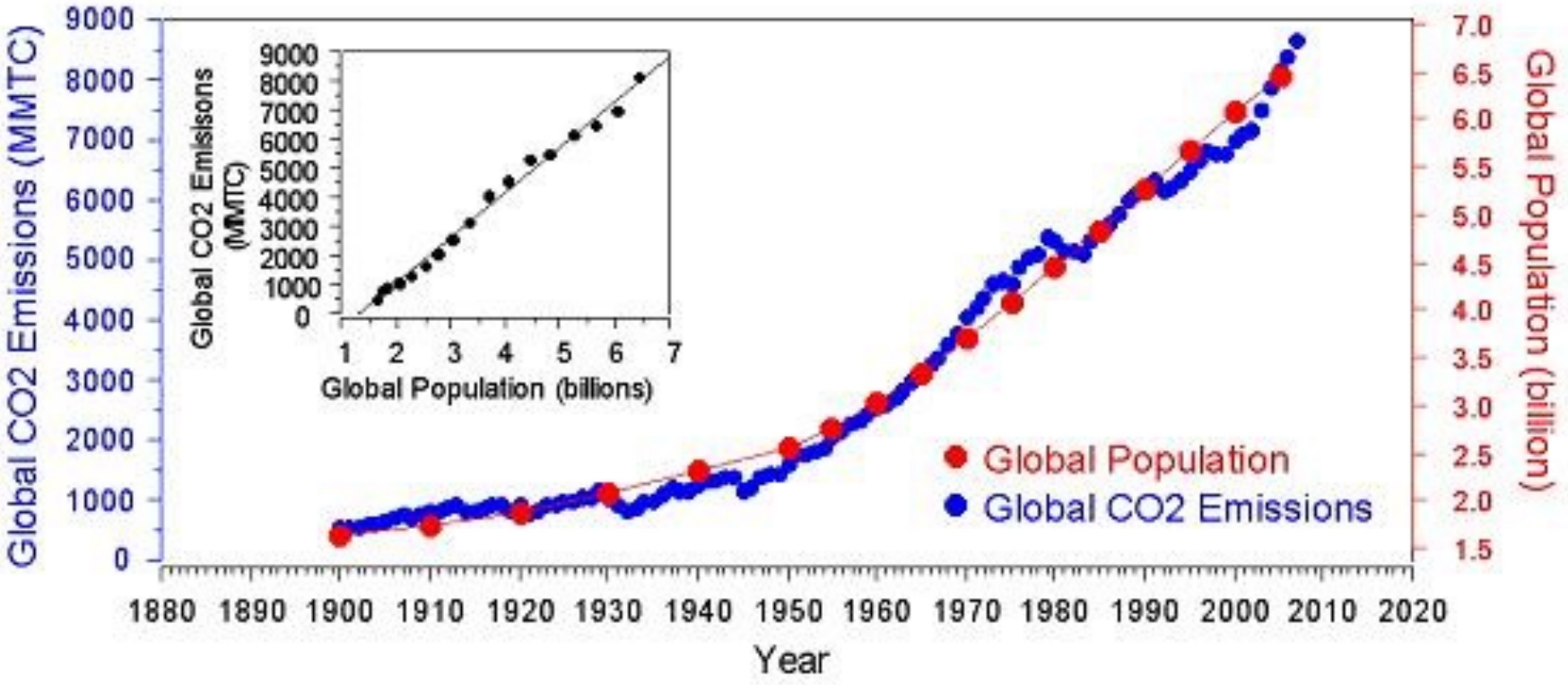


Emissions Poll

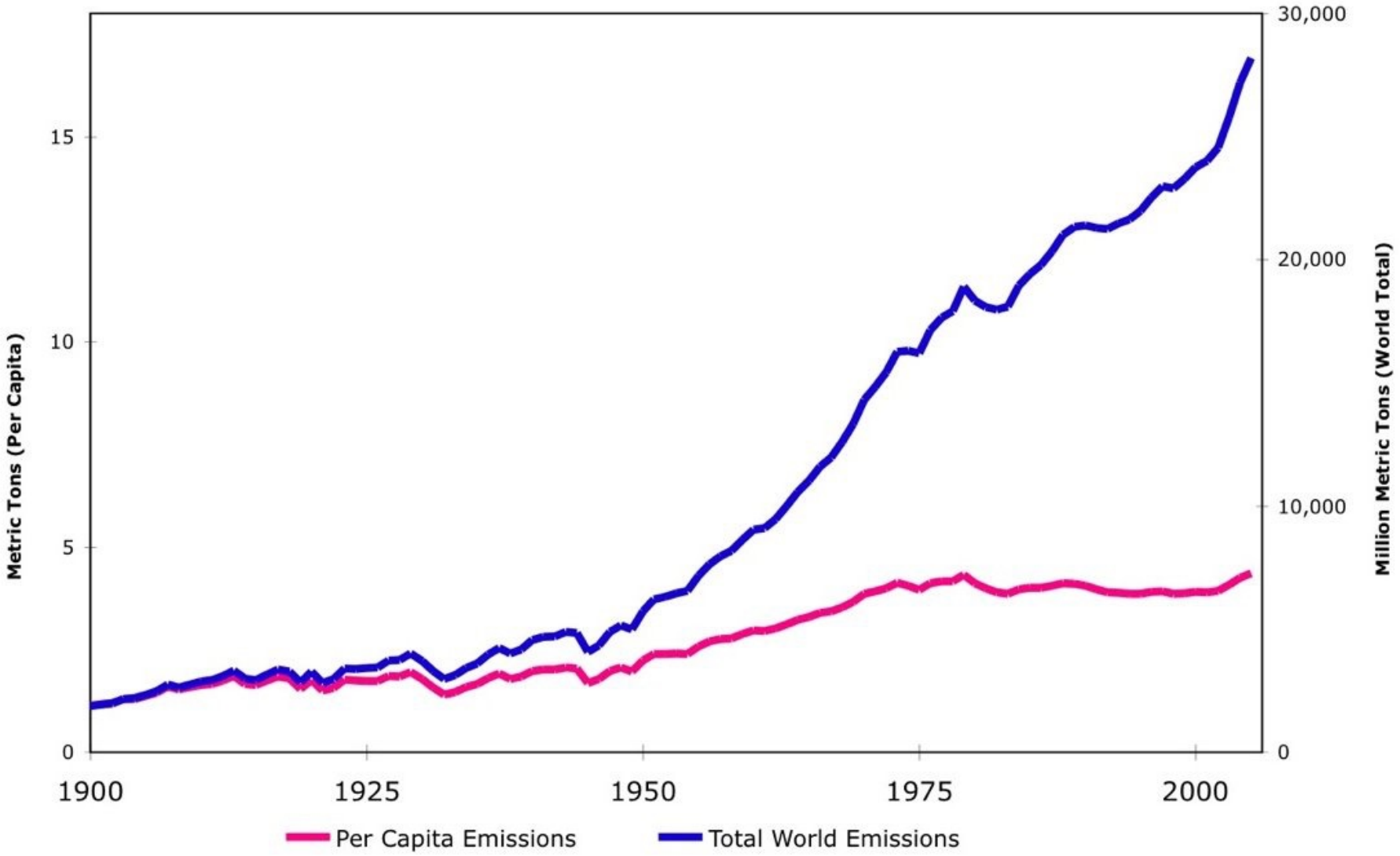
- Are per capita emissions growing faster or slower than total emissions?

- **Slido.com #261096**

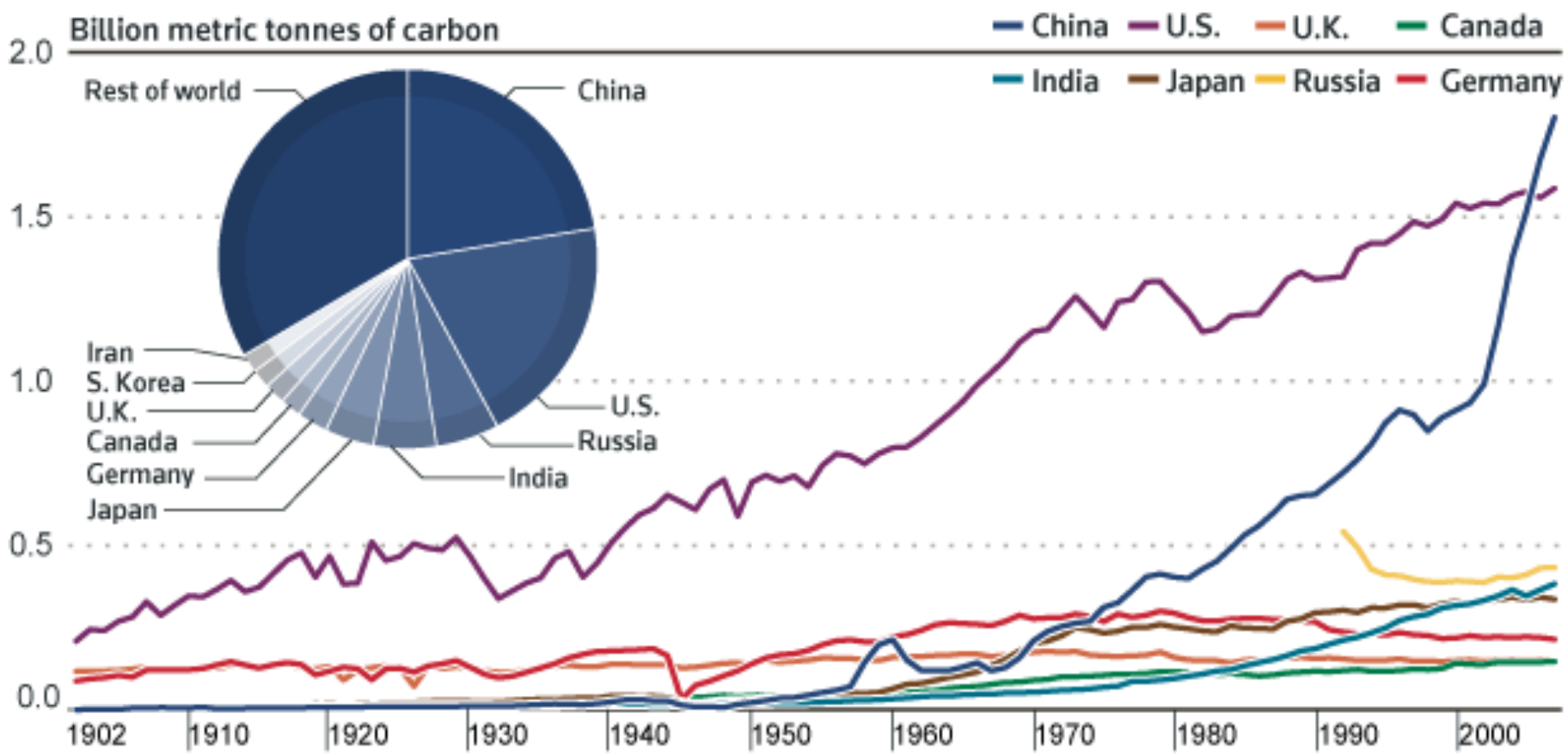
Emissions and population



Per Capita and Total World Carbon Emissions



World carbon emissions



Source: U.S. Dept of Energy's Carbon Dioxide Information Analysis Center (CDIAC)

REUTERS

Reuters graphic/Catherine Trevethan - corrects trillion metric tonnes to billion

02/01/09

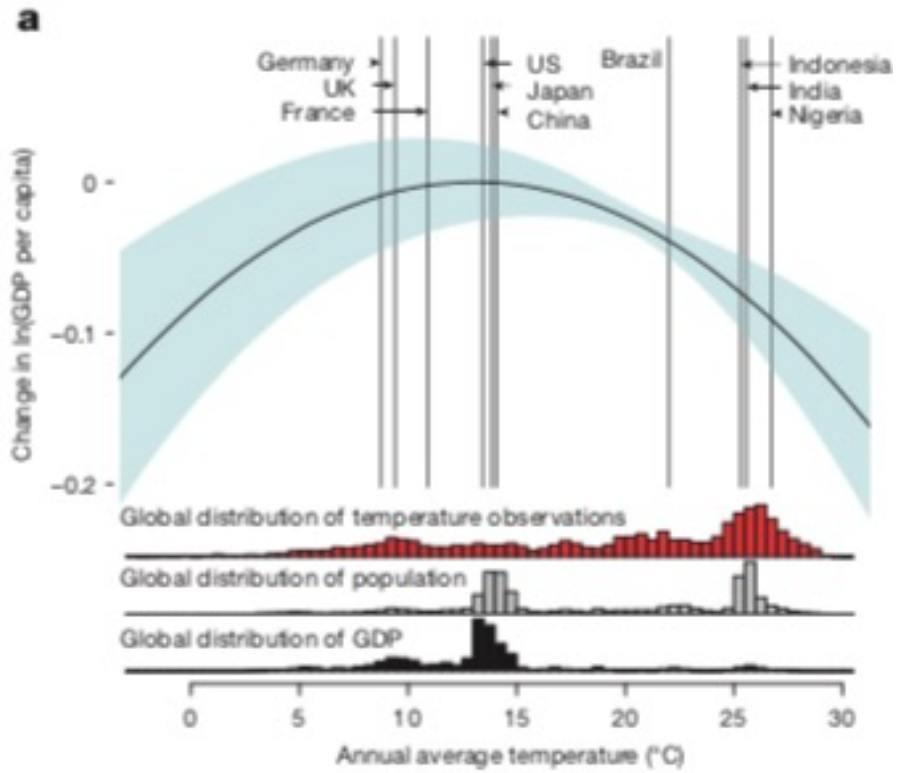
Quiz

- What percentage of countries in the world will see a drop in per capita incomes by 2100 relative to current levels due to climate change?
  a) 5%
  b) 10%
  c) 25%
  d) 75%

Slido.com #261096

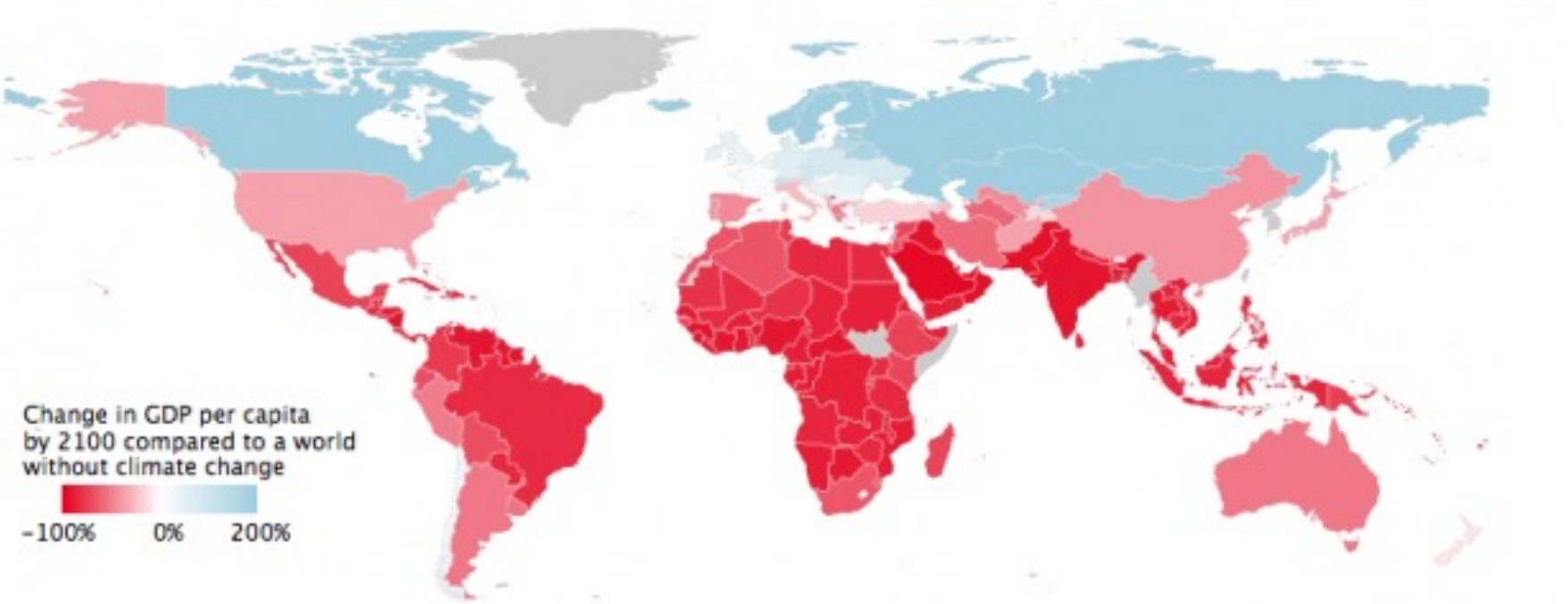
Economic activity and temperature

- Overall economic productivity is non-linear in temperature for all countries, with productivity peaking at an annual average temperature of 13 ^{o}C and declining strongly at higher temperatures.
- Unmitigated warming is expected to reshape the global economy by **reducing average global incomes roughly 23% by 2100** and widening global income inequality.
- **77% of countries in the world** would see a drop in per capita incomes relative to current levels.

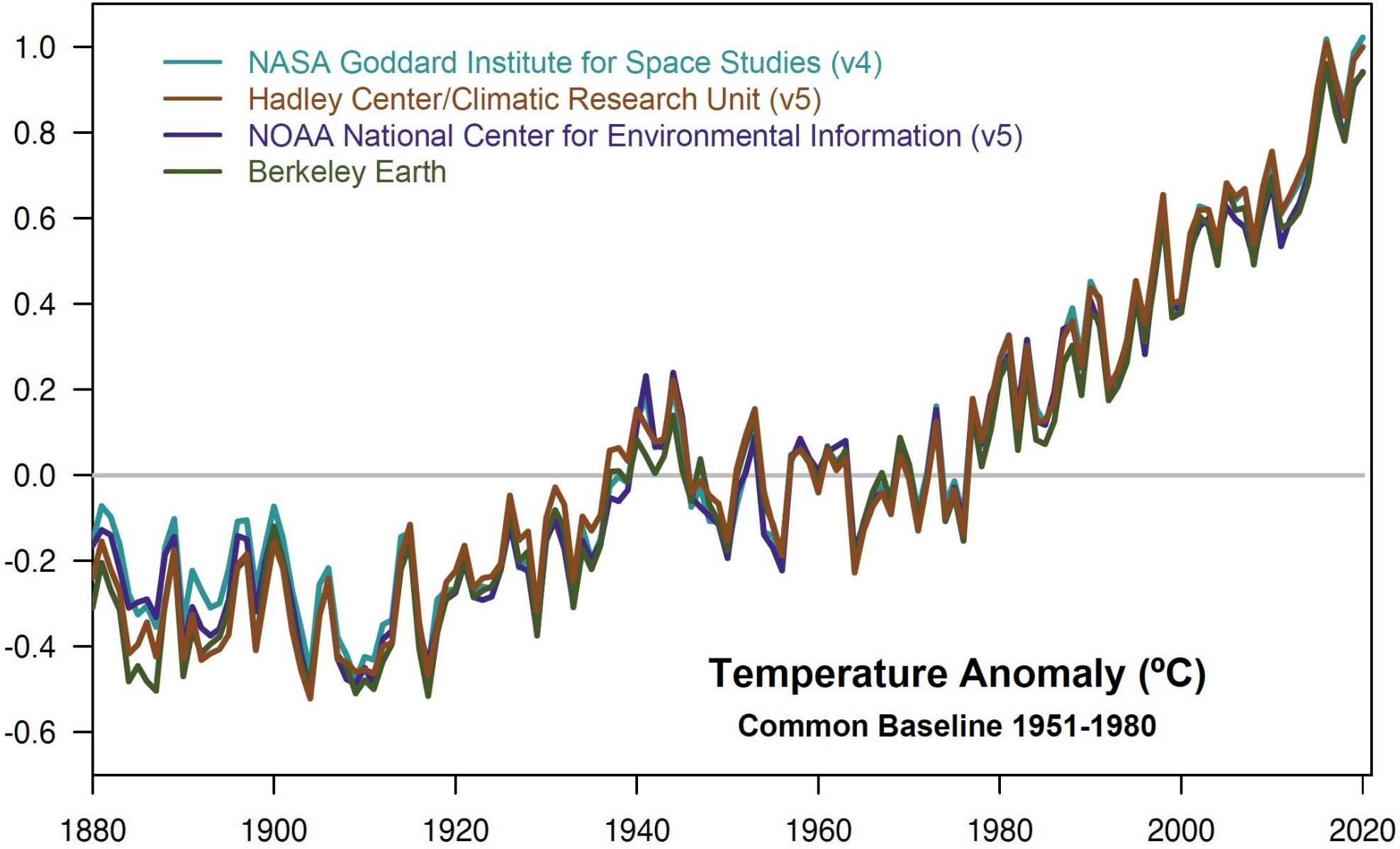


Burke, M., Hsiang, S.M. & Miguel, E. (2015). Global non-linear effect of temperature on economic production. Nature 527, 235–239.

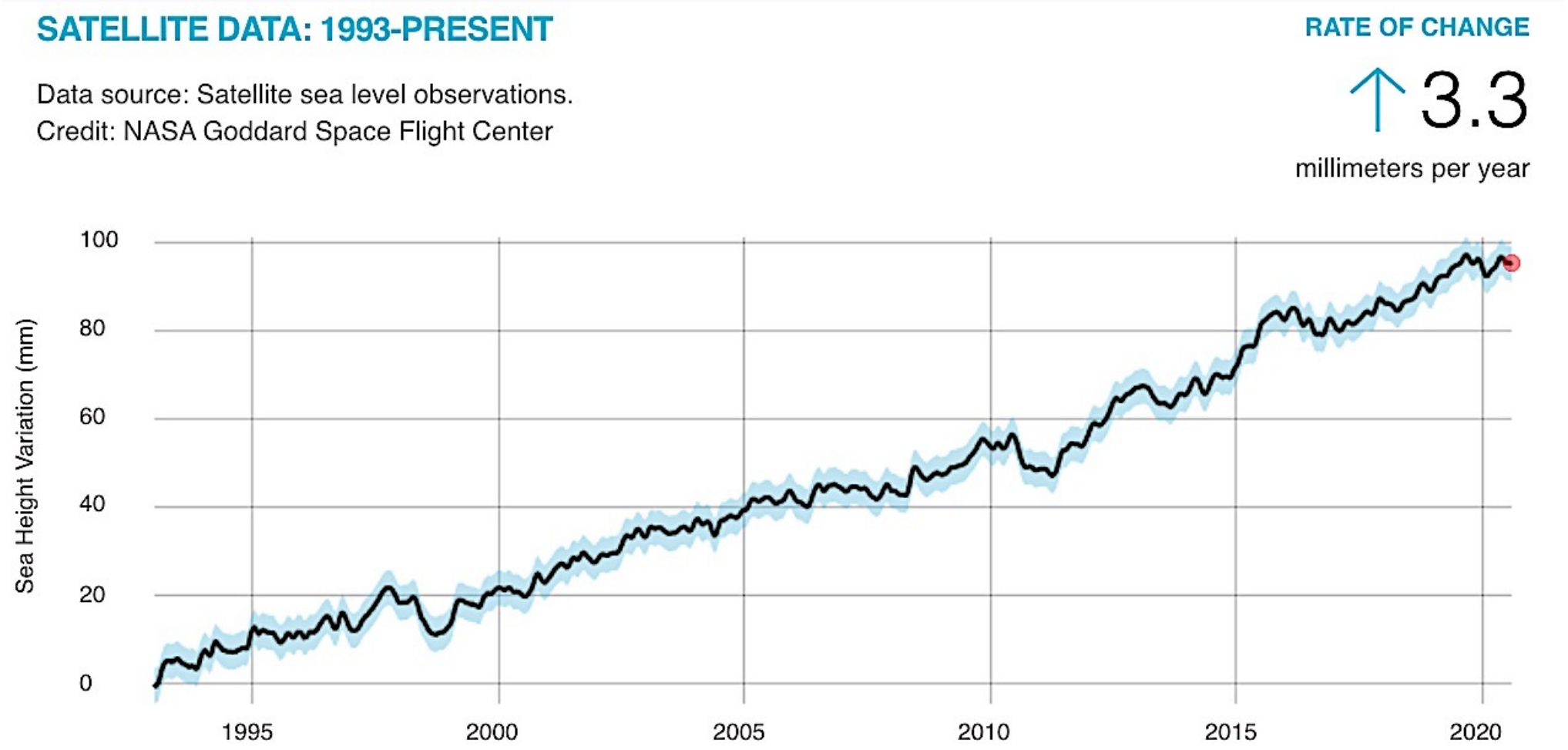
GDP and climate change

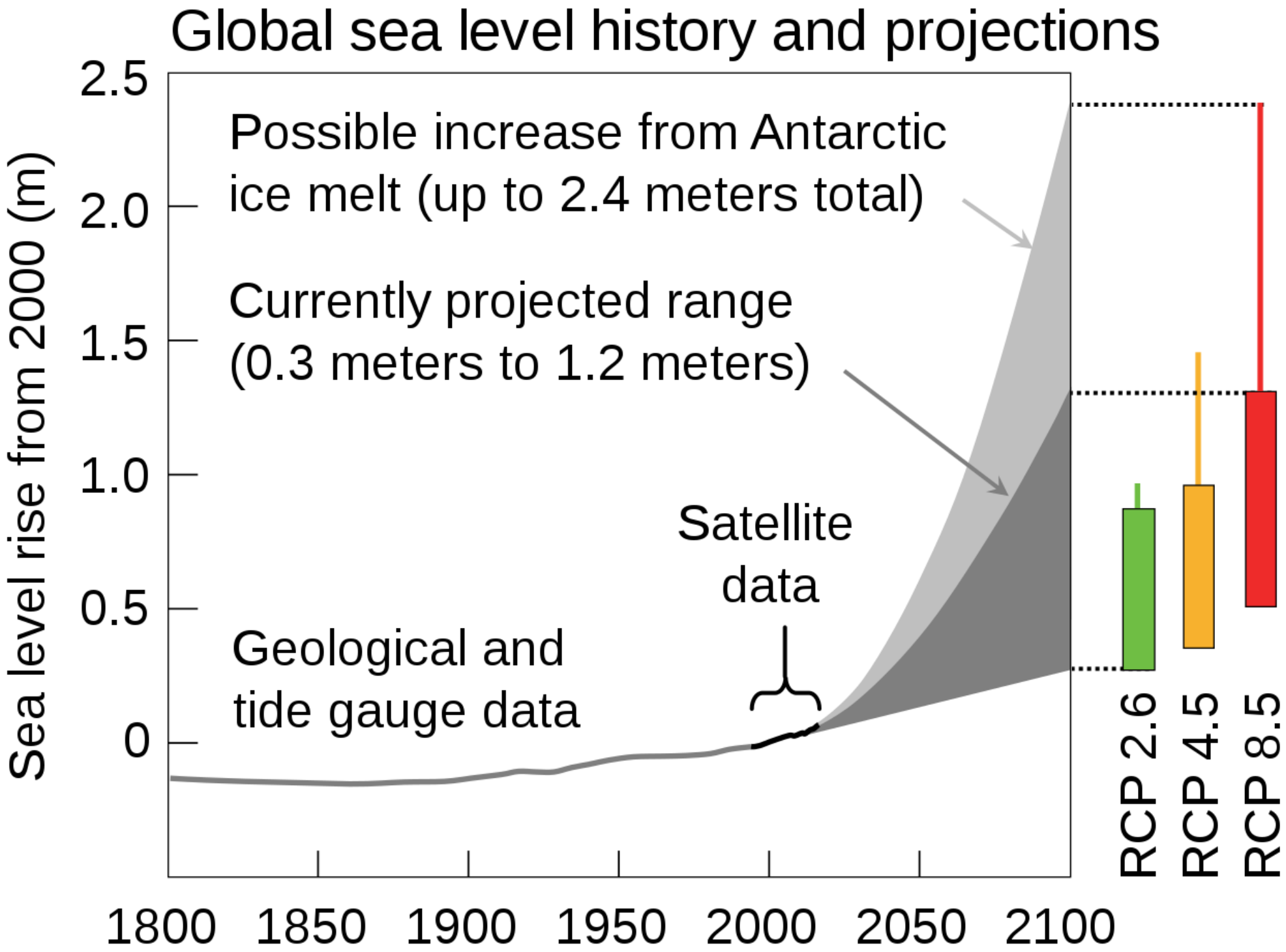


Global Temperature

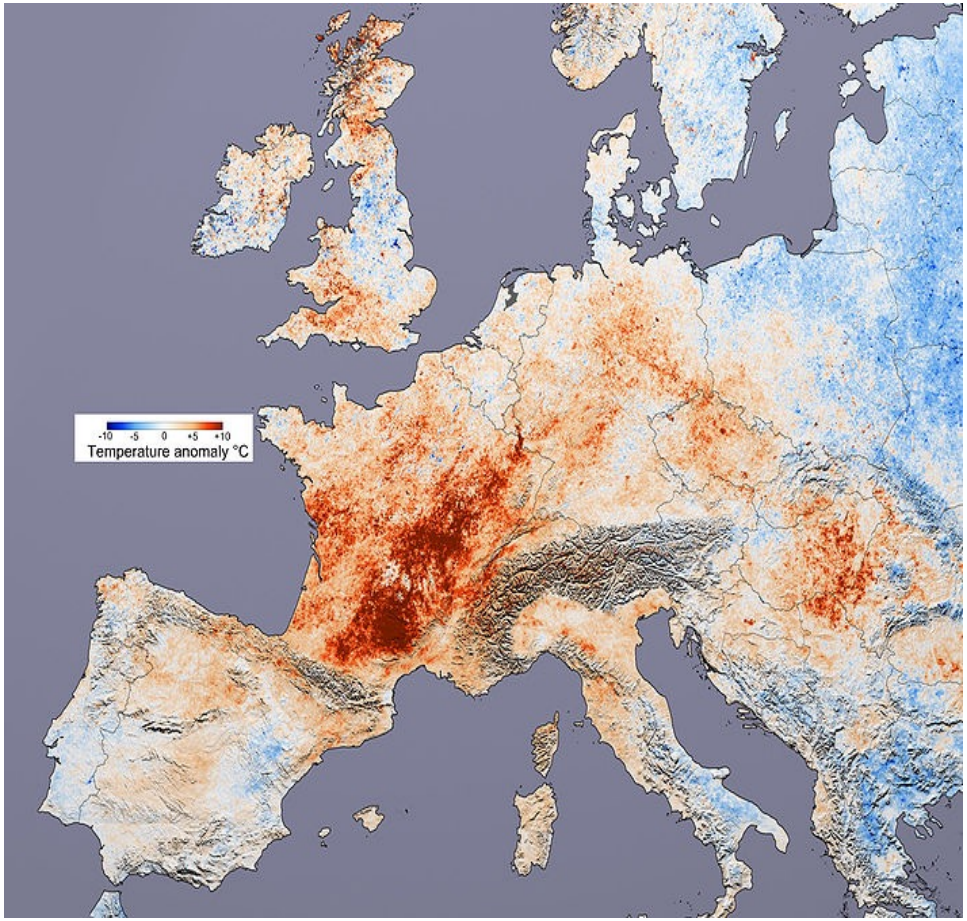


Sea Level Rise




Global sea level history and projections
Sea level rise from 2000 (m)
2.5
2.0
1.5
1.0
0.5
0
Possible increase from Antarctic ice melt (up to 2.4 meters total)
Currently projected range (0.3 meters to 1.2 meters)
Satellite data
Geological and tide gauge data
RCP 2.6
RCP 4.5
RCP 8.5
1800
1850
1900
1950
2000
2050
2100

2003 European Heatwave



Recent studies suggest that the likelihood of heatwaves has increased by a factor of four to nine due to human influence

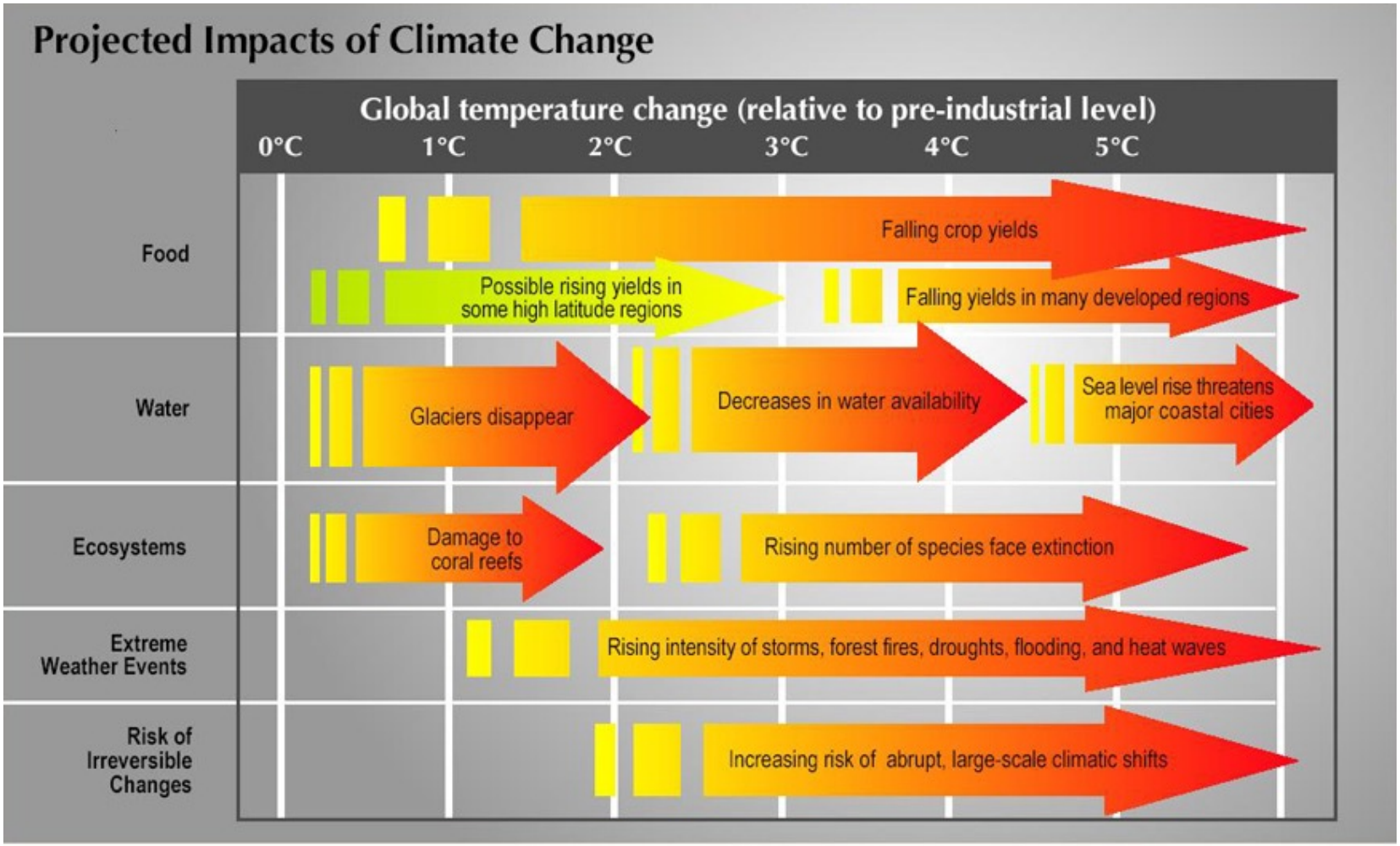
Difference in average temperature (2000, 2001, 2002 and 2004) from 2003, covering the date range of July 20 - August 20

Source: NASA

Stern Review (2006)

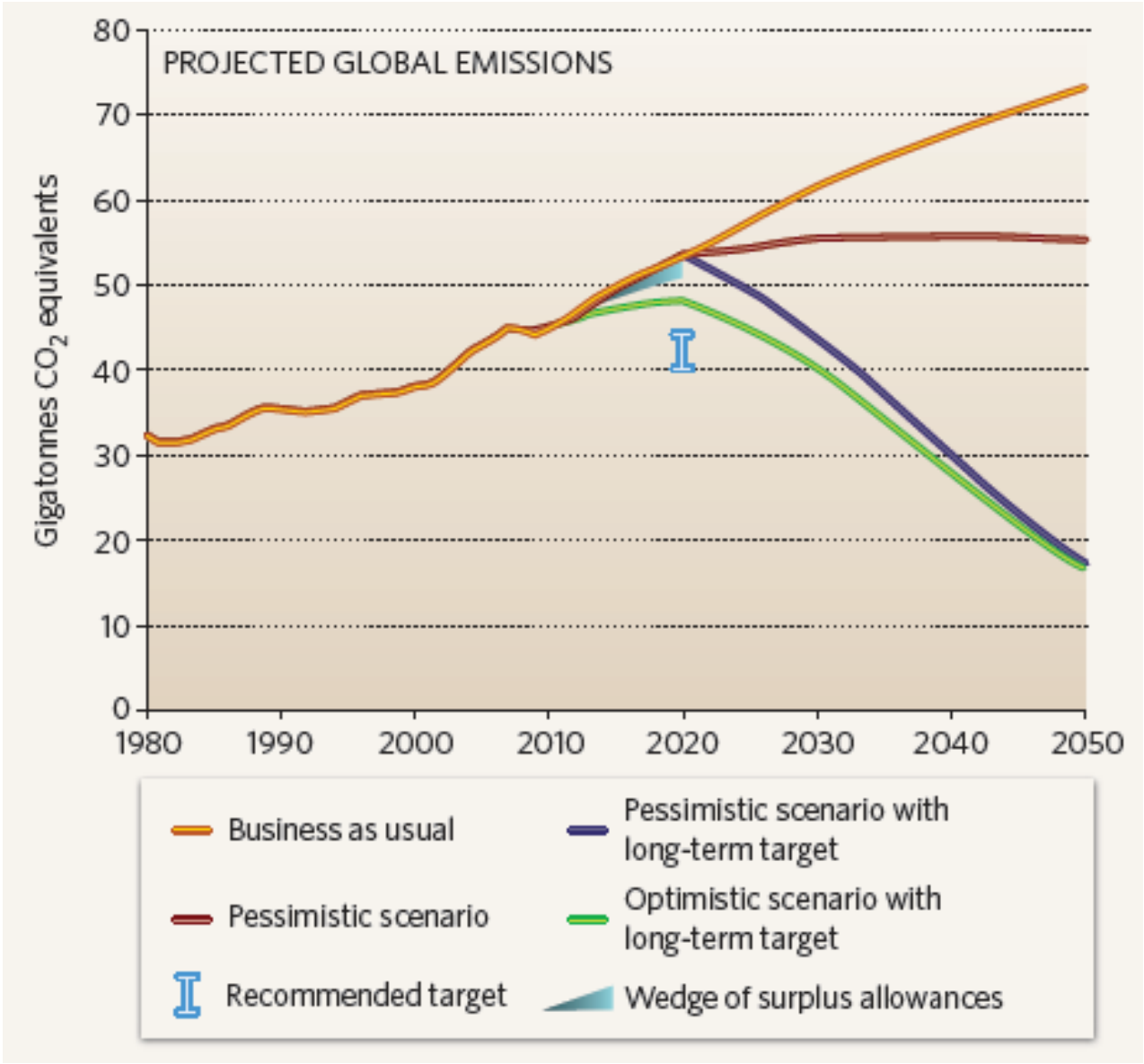
- The Stern Review on The Economics of Climate Change (2006) presented to the British Government.
- Benefits of strong, early action on climate change considerably outweigh the costs.
- One percent of global GDP per annum is required to be invested in order to stabilise emissions [500 - 550 ppm] and avoid the worst effects of climate change.
- Failure to do so could risk global GDP being up to 20% lower than it otherwise might be.
- Stern revised estimate from 1% to 2% in 2008.

Impacts of Climate Change



Source: Stern Review (2006).

Global Emissions Scenarios



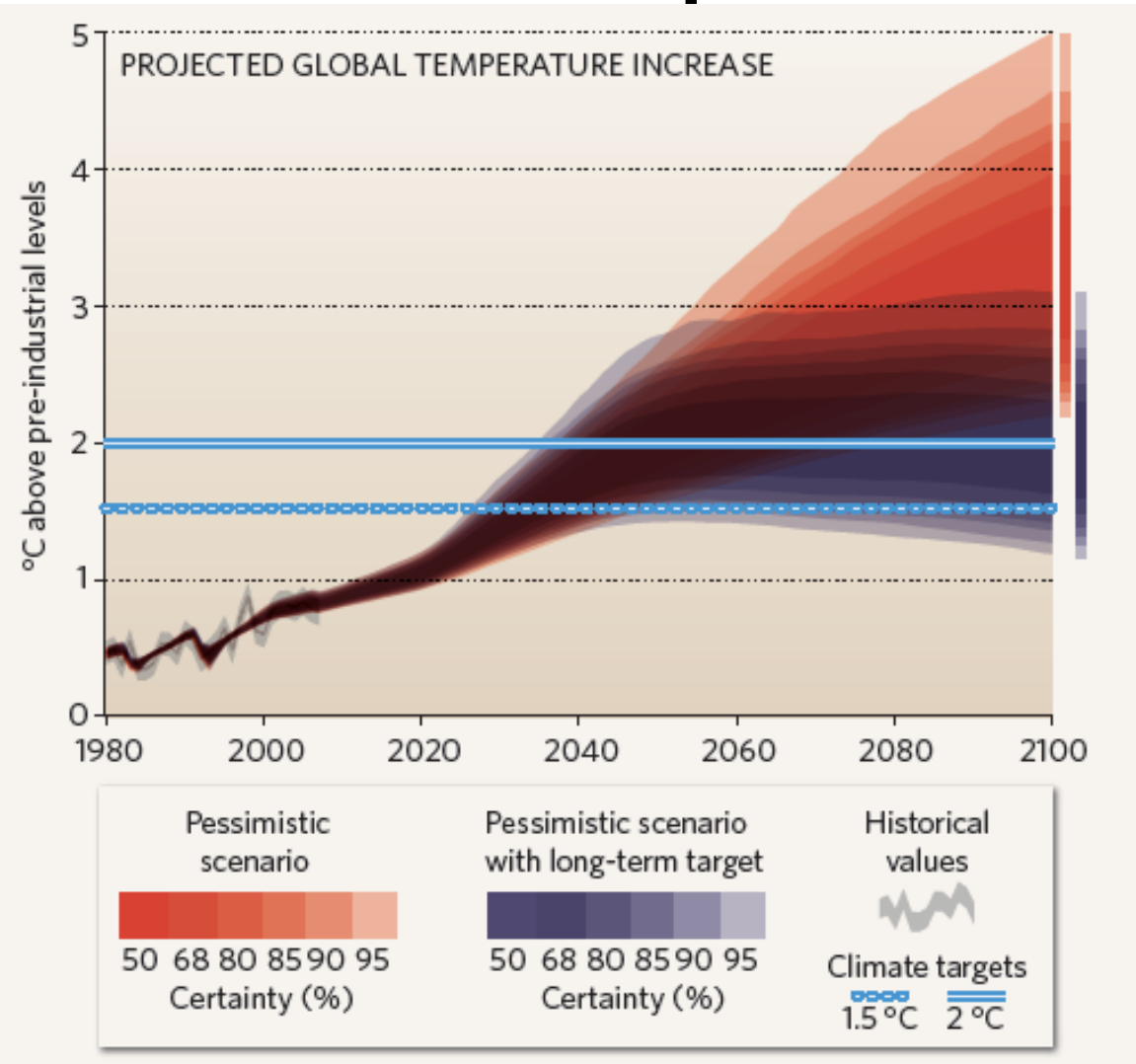
The pessimistic scenario assumes that nations meet only their lowest stated ambitions, and use all surplus allowances and land-use credits.

The optimistic scenario assumes that nations meet their highest stated ambitions, and do not use surplus allowances or land-use credits.

The long-term target is to halve emissions from 1990 levels by 2050.

Source: Rogelj, J., et al. (2010). Copenhagen Accord pledges are paltry. *Nature,* 464, 1127-1128.

Global Temperature Scenarios



Two scenarios:

Red: meet current pledges only

Blue: meet current pledges and go on to achieve a 50% reduction in emissions by 2050

Ambitious long-term emissions reduction goal is needed to maintain temperatures at or below 2° C

Source: Rogelj, J., et al. (2010). Copenhagen Accord pledges are paltry. *Nature,* 464, 1127-1128.

Summary

- Unless emissions are dramatically reduced (4-6% per year), climate change will have high costs for human development, economies and the environment
- Current pledges imply a greater than 50% chance that warming will exceed 3° C by 2100
- If nations agree to halve emissions by 2050, there is still a 50% chance that warming will exceed 2° C and will almost certainly exceed 1.5° C
- Paris COP21 agreed to target 1.5° C
- Solutions depend on behavioural change, community action and innovative technologies

Word Cloud

- What words would you use to define “open data”?

Slido.com #261096

Open data definition

- Open data is data that anyone can access, use or share – www.theodi.org
- Open data is defined as "*Open data is data that can be freely used, re-used and redistributed by anyone - subject only, at most, to the requirement to attribute and sharealike.*" – www.opendatahandbook.org

Open data properties

- **Availability and Access:** the data must be available as a whole and at no more than a reasonable reproduction cost, preferably by downloading over the internet. The data must also be available in a convenient and modifiable form.
- **Re-use and Redistribution:** the data must be provided under terms that permit re-use and redistribution including the intermixing with other datasets.
- **Universal Participation:** everyone must be able to use, re-use and redistribute - there should be no discrimination against fields of endeavour or against persons or groups. For example, 'non-commercial' restrictions that would prevent 'commercial' use, or restrictions of use for certain purposes (e.g. only in education), are not allowed.

Source: www.opendatahandbook.org

Motivation for open data

- When big companies or governments release non-personal data, it enables small businesses, citizens and medical researchers to develop resources which make crucial improvements to their communities.
- It is viewed as important because facilitating access to data allows the private sector to innovate, create jobs and increase economic prosperity.
- However it requires substantial investment and places pressure on those government agencies that are responsible for collecting, cleaning, processing and verifying data.

Open data is not easy

- Many governments are adopting open data programs with varying levels of success.
- Despite the substantial resources required for open data, there is an expectation that making data accessible will drive innovation and strengthen the economy.
- The Open Data Institute (ODI) and Nesta, the UK's innovation charity, claim that open data could strengthen the UK economy with a five to ten fold return on every pound invested.

Benefits of open data

- Open data has proven benefits for our economies and citizens.
- The Shakespeare Review identified £6.8bn of total value in UK public sector data, a report in 2011 estimated that the EU market for public sector data would grow to €40bn per year.
- McKinsey estimated a global market powered by open data from across seven sectors would create between $3tn and $5tn a year.

Reasons to invest in open data

- The Open Data Institute has identified open data-driven UK companies with a combined annual turnover of over £92bn, employing over 500,000 people.
- Transport for London alone has seen a 58:1 return on investment by releasing transport data, in the process helping create global leaders such as Citymapper.
- Denmark has identified a 70:1 return on investment by choosing to publish address data openly.
- In the US, an open data company has sold for $930m and Landsat data create savings of $350m to $436m per year, while at least 84% of American smartphone owners use an application powered by open data every single day.

Rwanda Open Data

- The Government of Rwanda has ratified an Open Data policy.
- This has been renamed the “Data Revolution Policy”.
- The new name highlights the need to access data and also to use big data as a potential means of creating jobs and economic prosperity.

Open Data Motivation

- World Bank and Partnership for Open Data
- Video 1
- Video 2

Facilitating usable data

- There remains a significant gap between the availability of raw data from government websites and the ability to use this data in a research or business context.
- Substantial resources are currently being wasted through the need for individuals to carry out numerous repetitive manual tasks in order to obtain data in an appropriate format.
- By providing suitable data science techniques, we show that it is possible to bring raw data to life by offering summary statistics, extremes, distributions and trends.
- We also offer a user-friendly download portal and application programming interface (API) to facilitate developers.

Climate change and sectors

- Quantification of the economic impact of climate change relies on models to infer the relationship between weather and output.
- This requires long historical records of both weather (climate) and output.
- The temporal resolution should exist at a monthly level, at least, in order to understand the effect of inter-annual seasonality.

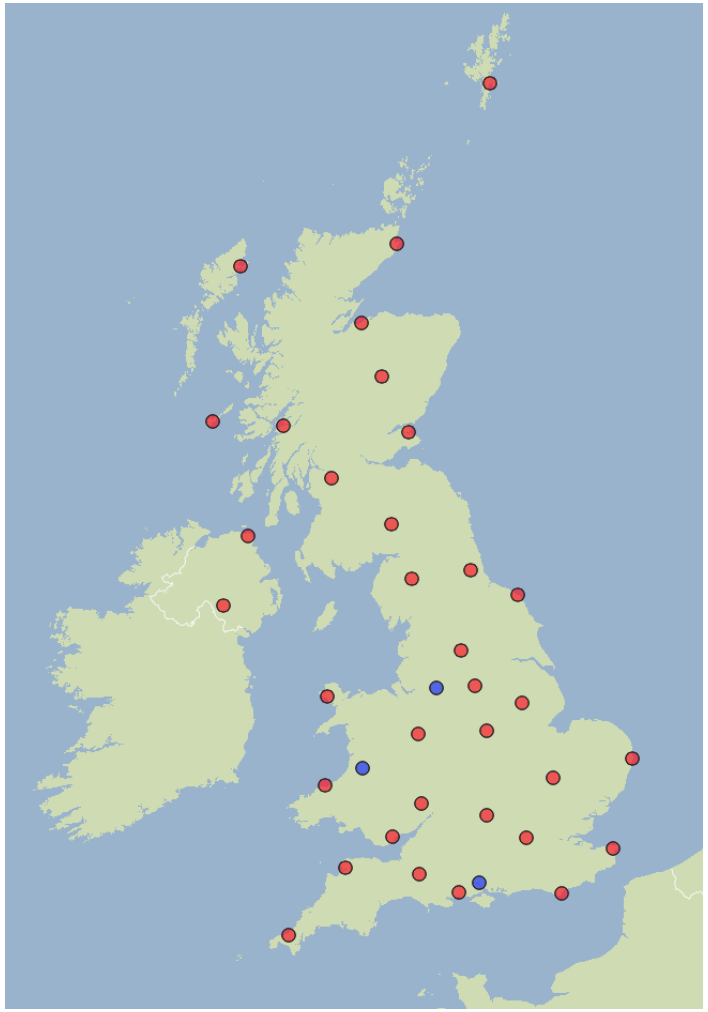
UK Met Office

- The Met Office is the national weather service in the UK. Weather, climate and environmental data have been identified as a key priority for the UK's open data program.
- DataPoint is a service to access freely available Met Office data feeds in a format that is suitable for application developers.
- We focus on long-term historical weather observations at a monthly time scale for 33 stations in the UK.
- The Climate Historic database provides observations on maximum and minimum temperature, rainfall, sunlight and air frost.

UK Met Office Climate Historic

- The UK Met Office has made considerable efforts to make its databases publicly available, following open data policy.
- Available does not mean useful!
- The description of the monthly database of climate data explains how six manual steps are required to extract the data for one station into a comma separated file.
- If you wanted to download all 33 stations, you would have to repeat all six steps for each.

UK climate - Historic station data

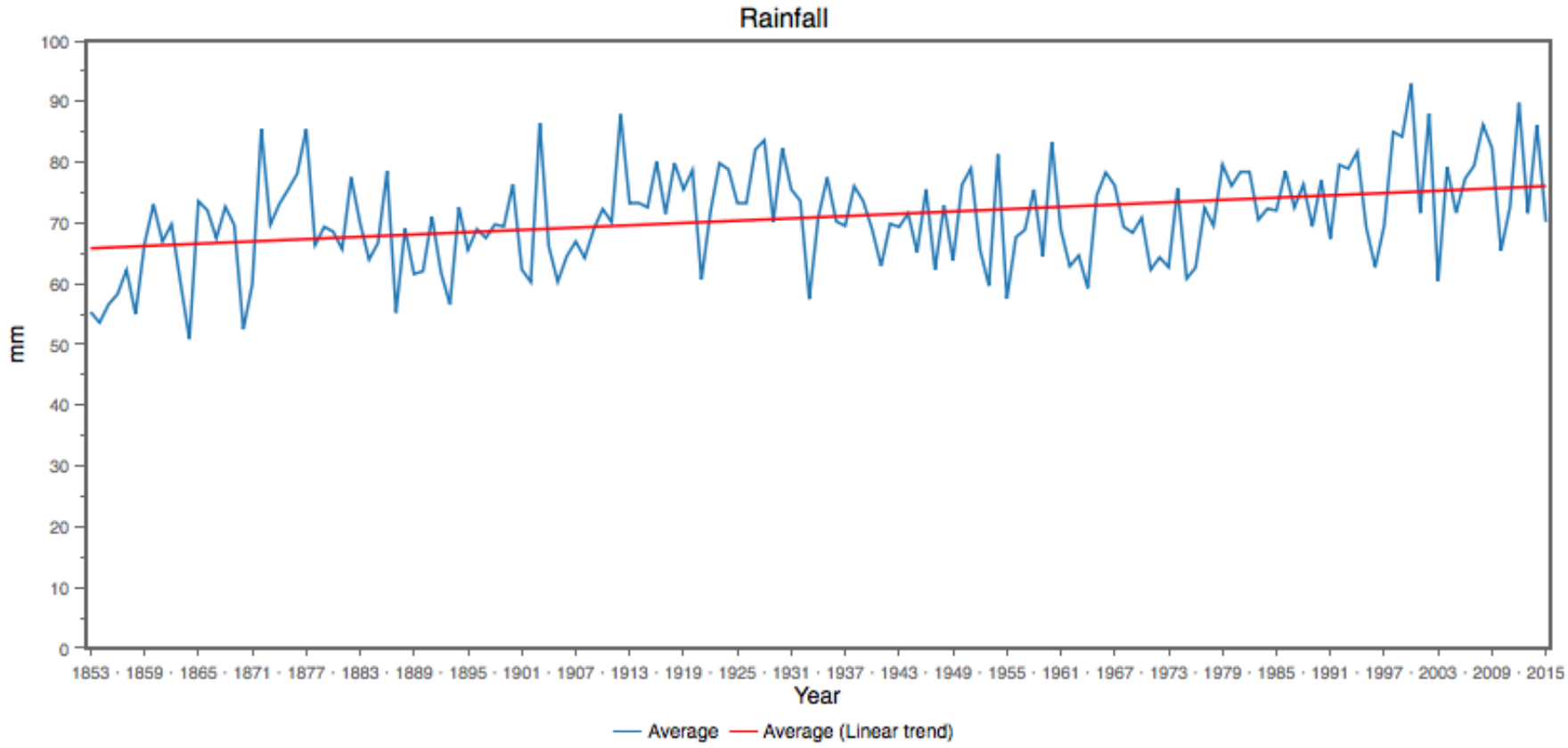


Source: http://www.metoffice.gov.uk/public/weather/climate-historic/#?tab=climateHistoric

Extremes

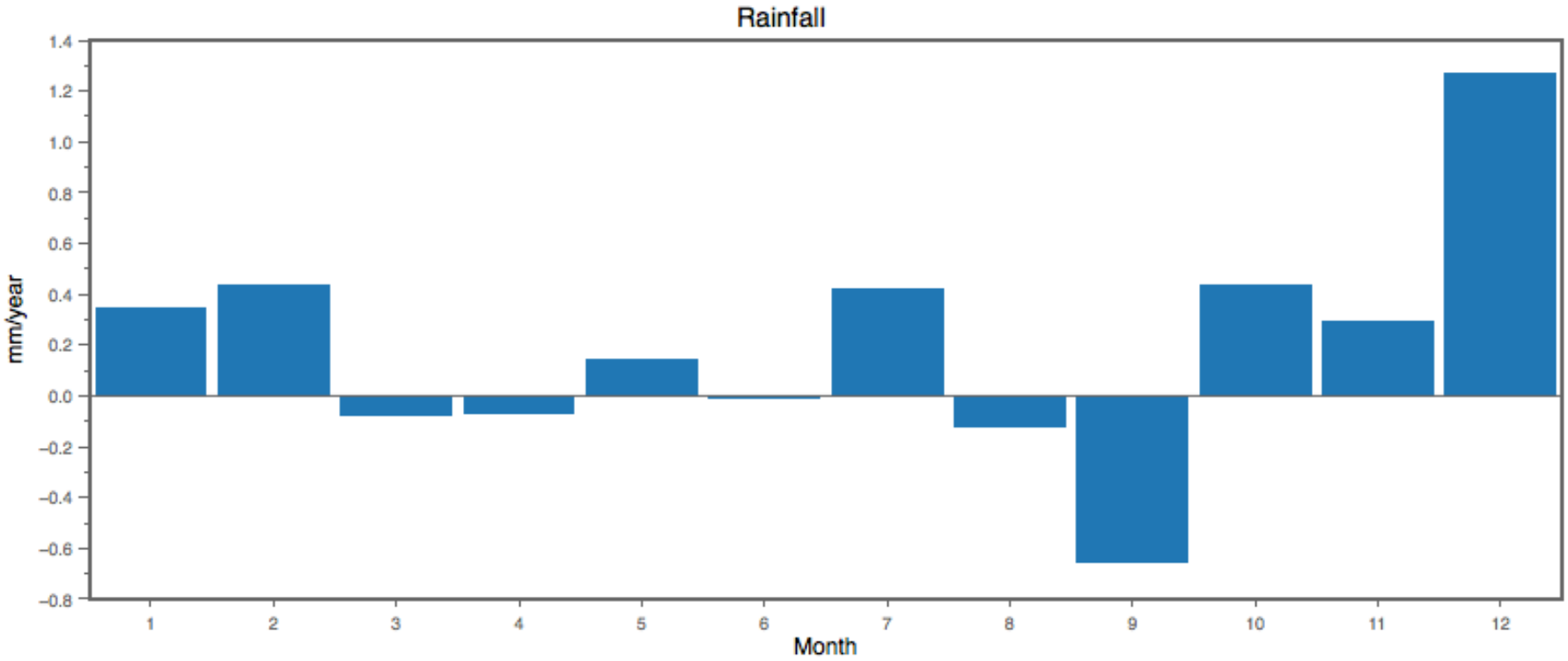
Extreme	Station	Year	Month	Value
Hottest	Cambridge	2006	7	28.3 °C
Coldest	Newtonrigg	2010	12	-7 °C
Wettest	Eskdalemuir	2015	12	568.8 mm
Sunniest	Southampton	1911	7	350.3 hours
Frostiest	Hum	1963	1	30 days

Trends



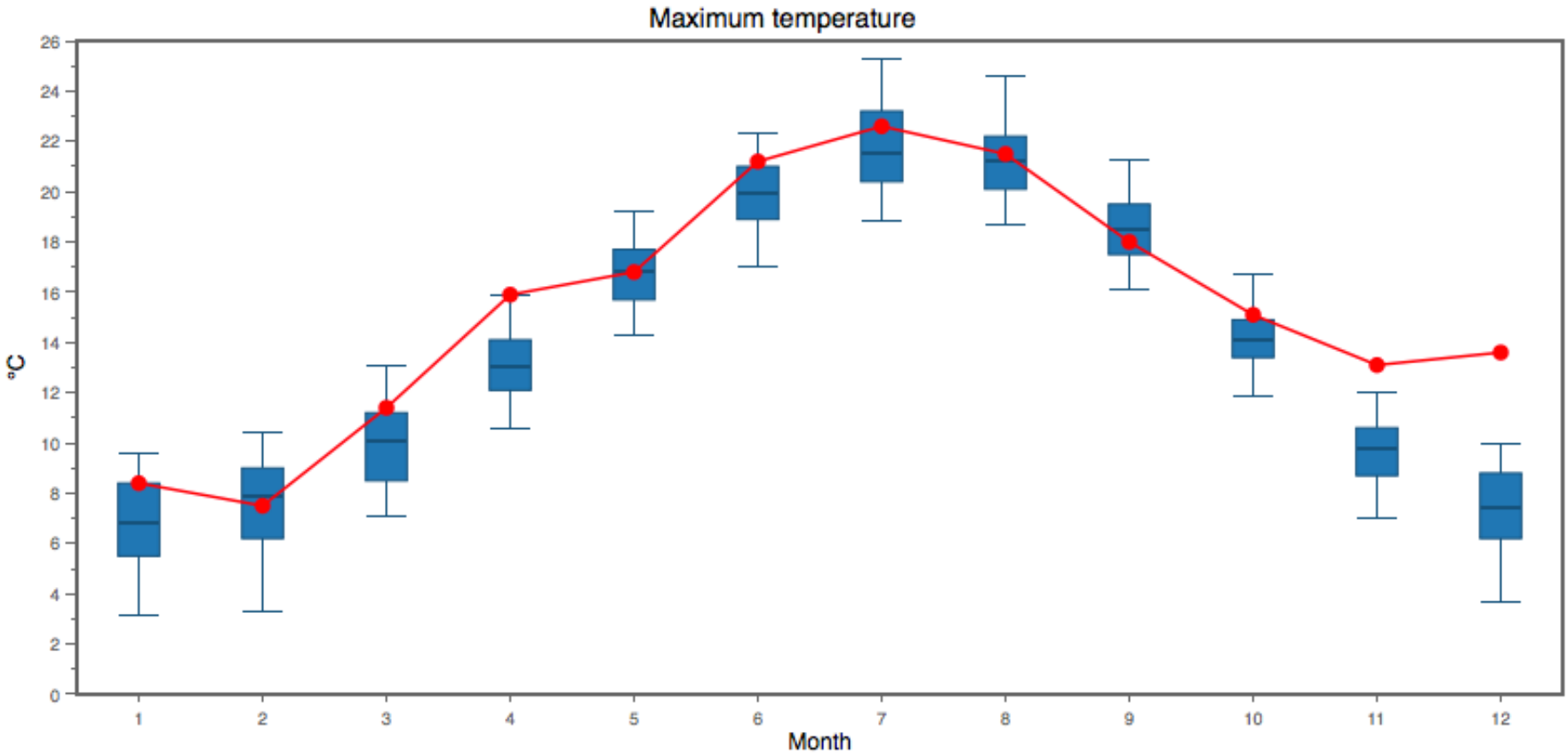
Source: climatefrontier.com (newtonrig station; rainfall)

Trends by month



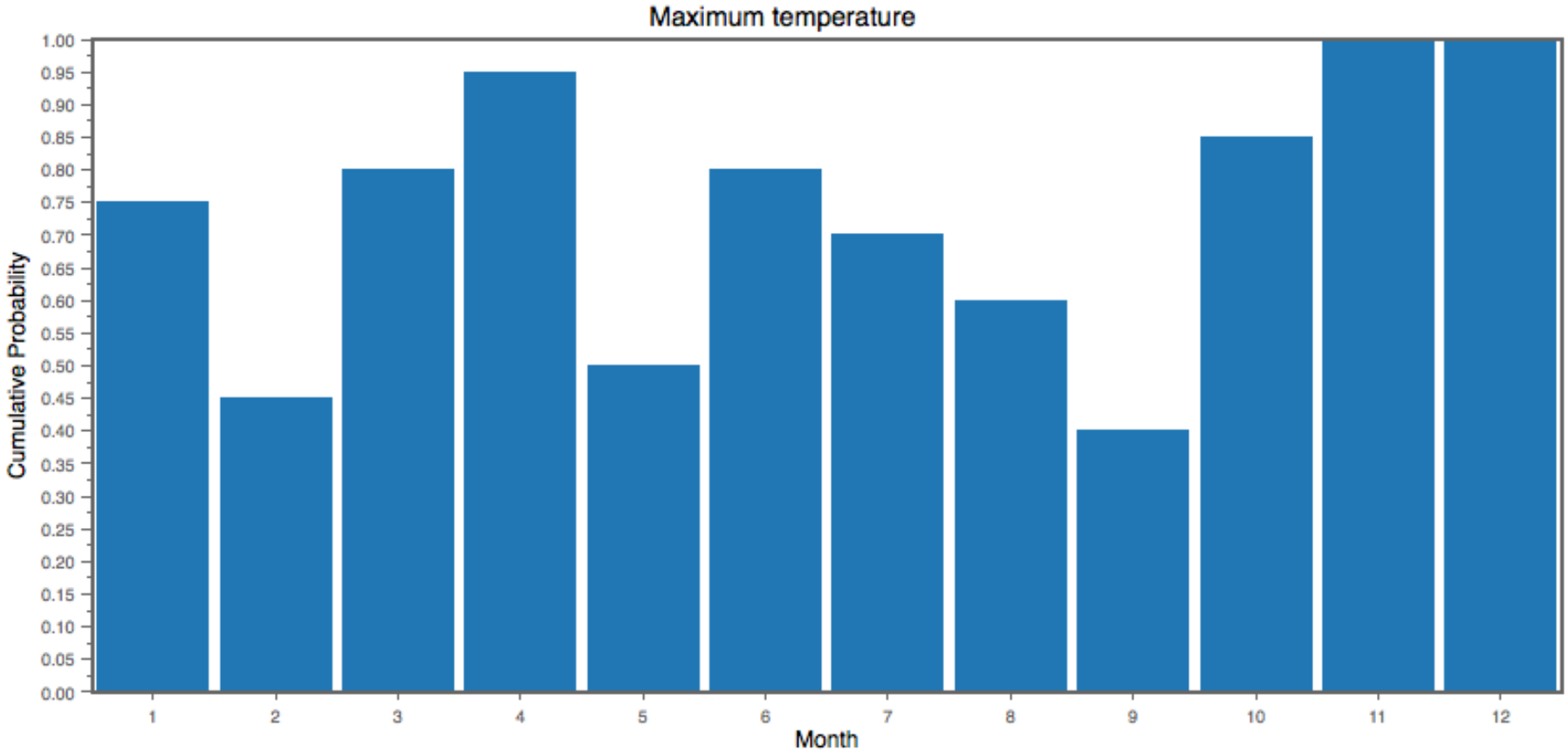
Source: climatefrontier.com (newtonrig station; rainfall)

Oxford Tmax Distributions



Source: climatefrontier.com (oxford station; max temperature)

Recent weather versus climate



Source: climatefrontier.com (newtonrig station; max temperature)

Climatefrontier.com

- Make it possible to access, explore, analyse weather data
- Facilitate selection of data by station and temporal duration
- Ability to download in a machine readable format such as csv file.