# Meteorology 5340
# Environmental Programming and Statistics

## Contents


1 Introduction ..... 3

a. What this course is about ..... 3

b. Effective Research ..... 3

c. Uncertainty ..... 6

d. Population vs. Sample ..... 8

e. Reducing Dimensionality ..... 9

f. Descriptive vs. Inferential Statistics ..... 11

g. Replicability and Reproducibility ..... 12

h. Putting Statistics to Work ..... 13

i. Navigating Online, Communicating and using your own computing resources ..... 14

j. Check Your Understanding ..... 14

# Viewing Monthly Snowfall at Alta

- https://utahavalanchecenter.org/alta-monthly-snowfall

# Drought Periods from Tree Rings

**Figure 1.1. Summer soil-moisture reconstruction for the western U.S. highlighting recent megadroughts (Williams et al. 2020).**

# Gauging Uncertainty

- Accuracy- difference in response between a standard and instrument in varying environmental conditions. A measure of how close a measurement is to the “true” value

- Precision- how well repeated measurements of some quantity agree with each other. A precise instrument can be inaccurate

# Systematic vs. Random Errors

- Random- that which is not precisely predictable or determinable
- Systematic- errors arising from a consistent response of a measuring device to environmental conditions or faulty characteristics of instrumentation that occurs frequently

**Figure 1.3. A million member sample. The True value is assumed to be -1. The systematic (average) error would be +1.**

# Population vs. Sample

- we never know the entire population of true values as the environmental conditions change in time or space.
- We hope that we choose a sample of observations for analysis such that each element in the population has an equal chance to be selected.
- Sampling issues
  - Trends
  - serial dependence of environmental data
  - model sample tend to be less variable than observed samples

# Selecting a Sample

- Selecting the sample for analysis is a critical aspect of organizing the data and depends on the question to be addressed by the study
- rule of thumb: sample should be large enough to capture the phenomenon of interest many many times
- “Degrees of freedom”: number of independent elements in the sample;
  - usually much smaller than the total number of members in the sample in environmental data sets
- Keeping your powder dry- saving data for an independent sample to evaluate and confirm your results.
- Tendency to assume sample is drawn randomly from the population, when sample grossly underestimates the variability inherent in the population

# Common Goal is to Synthesize and Reduce Dimensionality

- Statistical analysis of environmental data typically involves reducing the dimensionality of the data to a manageable size.
- Which variable(s) do we need to consider?
- Can we consider one variable (univariate analysis) or must we consider multiple variables (multivariate analysis)?
- What time scales are we interested in? Hours, days, months, years? And, what region (local, regional, globally) or level in the vertical (surface, subsurface, upper air)?
- Are the data available on a spatial grid or at specific points?

# Large Dimensionality of Geophysical Data Sets

•Space: x, y, z

•Time: time (t) and forecast time ($t_f$)

•Parameter & Source (temperature, winds, different models, measuring systems, perturbations)

# Air Temperature

# How does the temperature compare to other years?

# Climate

- Statistics often misinterpreted as bookkeeping.
  - What is the warmest temperature on record at Salt Lake City?
  - What is the biggest snow storm at Alta?
- weather and climate:
  - weather- state of the environment
  - climate- aggregate summary of the environment
  - Climate normal: arbitrarily defined reference state: now 1991-2020

# Evolving Climate Normals

# Evolving Climate Normals

# https://www.climate.gov/maps-data

## Link still using 1981-2010 period

June 30 year average

# April 2021 Departure from Old 30Yr Normal

# https://www.ncdc.noaa.gov/sotc/global/202104

## Global Climate Report - April 2021

- Climate Monitoring
- State of the Climate
- Temp, Precip, and Drought
- Climate at a Glance
- Extremes
- Societal Impacts
- Snow and Ice
- Teleconnections
- Monitoring References

Monthly Reports | Summary Info | Monthly Briefings | RSS Feed

Report: Global Climate Report Year: 2021 Month: April Submit

« March 2021

Introduction | Temperature | Precipitation | References

### Additional Resources

Temperature and Precipitation Maps

- April

### Supplemental Material

- 2021 Year-to-Date Temperatures Versus Previous Years
- Global Annual Temperature Rankings Outlook

# Selected Significant Climate Anomalies and Events: April 2021

**GLOBAL AVERAGE TEMPERATURE**
**April 2021 average global land and ocean temperature was the ninth highest for April since global records began in 1880.**

**ARCTIC SEA ICE EXTENT**
The April 2021 Arctic sea ice extent was 5.8% below the 1981–2010 average — the sixth-smallest April extent since satellite records began in 1979.

**ASIA**
April 2021 was Asia's 12th-warmest April on record.

**EUROPE**
Europe had its coldest April since 2003.

**TYPHOON SURIGAE**
Surigae was an extremely powerful cyclone, with the strongest maximum wind speed ever recorded for a storm during the months of January-April anywhere in the world.

**CONTIGUOUS UNITED STATES**
A late-season snowstorm on April 20-27 brought record snow for this time of year across parts of the Ohio Valley and the Northeast. The storm snapped trees and caused power outages and travel problems.

**AFRICA**
Africa's April 2021 temperature departure was the fourth-warmest April on record.

**SOUTH AMERICA**
South America had its seventh-warmest April on record.

**CYCLONE SEROJA**
Cyclone Seroja brought heavy rain and strong winds to parts of Western Australia. Several locations had their highest daily rainfall on record. Seroja also affected Indonesia, causing historic floods and landslides.

**NEW ZEALAND**
This was the fourth-warmest April for New Zealand.

**ANTARCTIC SEA ICE EXTENT**
The Antarctic sea ice extent for April 2021 was 3.4% above average and was the 16th largest for April on record.

Please note: Material provided in this map was compiled from NOAA's State of the Climate Reports. For more information please visit: http://www.ncdc.noaa.gov/sotc

# Land & Ocean Temperature Departure from Average Apr 2021

(with respect to a 1981–2010 base period)

Data Source: NOAAGlobalTemp v5.0.0–20210509

National Centers for Environmental Information
GHCNM v4.0.1.20210507.qfe

Please Note: Gray areas represent missing data
Map Projection: Robinson

# $CO_2$

https://www.esrl.noaa.gov/gmd/obop/mlo/

# CO2 on Campus

# Why use statistics to describe the environment?

- Environment controlled by innumerable factors, which we hope to segregate into a few critical factors from the rest that, for the most part, simply contribute to background noise
- the characteristics of the system include linearly unstable processes such as baroclinic waves that cause growth of small features into larger ones
- the characteristics of the system (dynamics, thermodynamics) are nonlinear and include discrete step functions (i.e., rain/no rain) that can lead to the amplification of small errors into large ones

141229/0000F000 500 hPa Height and Height Anomaly (dam), from GFS

# Degrees of Freedom

- A sample should be large enough to capture the phenomenon of interest many times.
- “Degrees of freedom” refers to the number of independent elements in the sample
- the number of degrees of freedom is usually much smaller than the total number of members in the sample in environmental data sets

# 2 degrees of freedom

https://www.esrl.noaa.gov/gmd/ccgg/trends/full.html

# 3 degrees of freedom: 3 independent values

series/globe/land_ocean/ytd/12/1880-2021

~6 degrees of freedom?

# Steps for Effective Research

- distill a general interest in a subject into a specific question/hypothesis that can be evaluated
- organize the data
- find relationship(s) among the data
- examine the significance of your results
- review thoroughly what you have done and document your analysis and results
- submit your results and study for independent evaluation

Everyone is entitled to his own opinion, but not his own facts.
Daniel Patrick Moynihan
BrainyQuote

# Reproducibility

- The extent to which measurements or observations agree when performed by different individuals defines this important tenet of the scientific method.

CrossMark
click for updates

**Editorials**

**Irreproducible Experimental Results**

**Causes, (Mis)interpretations, and Consequences**

**Joseph Loscalzo, MD, PhD**

+ Author Affiliations

# Reproducible-Science-Curriculum

## Irreproducible Examples

tnabtaf edited this page on Dec 11, 2014 · 18 revisions

Notes on examples of irreproducible research that has caused significant issues. Some of these will become case studies during the course. Initially, these are all from RetractionWatch.

## Papers

- Editorial Expression of Concern: Non-adaptive origins of interactome complexity
- Retraction Watch Univ.: No misconduct, but "poor research practice" in mgt prof's work now subject to 7 retractions
  - "The investigation was severely hindered by the near total unavailability of raw data files and statistical output files for the seven papers under study."
- Retraction Watch Data questions prompt retraction of PLOS ONE cardiovascular paper
  - The authors have indicated that the raw data for Figures 3, 5 and 6 are not

# Reproducibility vs. Replicability

- Replicability is assessed by performing an experiment under exactly the same conditions at different times
- reproducibility is assessed by performing similar, but not identical, experiments at different times, in different locations, and under somewhat different experimental conditions.
- replicability reflects the technical stringency or precision of a specific experiment
- reproducibility reflects the fundamental accuracy of an experimental observation
- a precisely conducted experiment can be inaccurate, and an accurate experiment may be performed imprecisely—especially in biomedicine where many factors can account for irreproducible results

# Statistics and Irreproducibility

- errors in design, data analysis, and data interpretation.
- Bias clearly plays an important role in promoting false-positive results.
- Reasons for bias:
  - lack of experimental balance leading to an impassioned belief in one particular experimental outcome clouding objectivity;
  - perceived pressure to publish for academic advancement or to enhance the likelihood of competing successfully for grant funding[7]; and the lack of appeal of negative (or neutral) studies in most high-impact journals.
- study illustrating the inverse relationship between the scientific hierarchy (physical sciences at the top, social sciences at the bottom, and biological sciences in between) and the publication frequency of “positive” results
- Random and systematic errors contribute as well

# Impact of Rarity on Irreproducibility

Black swan (Taleb)

- outcome is theoretically reproducible, but so rare that a sufficient number of experiments needed to determine its frequency exceeds feasibility
- Owing to its inherent rarity, it is effectively impossible to explain away an irreproducible result on the basis of a rare, singular event for which no supportive evidence can be marshaled
- Examples of irreproducible events in environmental fields?

# What a mess we're in…

- Over the past decade, the number of journal articles worldwide doubled from 1.1 to 1.9 million.
- Of these, 0.5 million articles are published in the field of biomedicine, for an output of 1400 articles per day.
- Given this volume alone, is it any wonder that an increasing number of articles are being published that contain irreproducible results?
- Furthermore, with the growth in search technologies, a greater number of articles is being retracted over time: ~40/year in the late 1990s, ~300 in 2010, and ~400 in 2011.
- The reasons for these retractions include plagiarism, data manipulation (especially in figures), and proven data falsification;
- however, irreproducibility resulting from innocent causes may also be included in this pool of retracted publications without being recognized as such.
- With as many as 50% of all articles listed in PubMed never cited at all, one can conclude either that the work is of minimal significance and not worthy of further pursuit or that it has been pursued and could not be reproduced.
- The extent to which these two explanations account for this statistic has not been (nor cannot easily be) determined.

# Taking Steps

- Nature
- Reproducibility, rigour, transparency and independent verification are cornerstones of the scientific method. Of course, just because a result is reproducible does not make it right, and just because it is not reproducible does not make it wrong.
- The guidelines recommend that journals include in their information for authors their policies for statistical analysis and how they review the statistical accuracy of work under consideration.

nature International weekly journal of science

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video | For A

Archive > Volume 515 > Issue 7525 > Editorial > Article

NATURE | EDITORIAL

## Journals unite for reproducibility

**Consensus on reporting principles aims to improve quality control in biomedical research and encourage public trust in science.**

05 November 2014

PDF | Rights & Permissions

Reproducibility, rigour, transparency and independent verification are cornerstones of the scientific method. Of course, just because a result is reproducible does not make it right, and just because it is not reproducible does not make it wrong. A transparent and rigorous approach, however, will almost always shine a light on issues of reproducibility. This light ensures that science moves forward, through independent verifications as well as the course corrections that come from refutations and the objective examination of the resulting data.

It was with the goal of strengthening such approaches in the biomedical sciences that a group of editors representing more than 30 major journals; representatives from funding agencies; and scientific leaders assembled at the American Association for the Advancement of Science's headquarters in June 2014 to discuss principles and guidelines for preclinical biomedical research. The gathering was convened by the US National Institutes of Health, *Nature* and *Science* (see *Science* **346,** 679; 2014).

**Related stories**

- Metascience could rescue the 'replication crisis'
- A blueprint to boost reproducibility of results
- Code share

More related stories

# What you should be doing

- Read Chapter 1 Notes
- Chapter 1 Check Your Understanding
  - Complete it now- don’t wait!