Welcome to Advanced Data Analysis (PHYS 605)

Prof. Claudia Gomes da Rocha

claudia.gomesdarocha@ucalgary.ca

Department of Physics and Astronomy
Faculty of Science, University of Calgary

https://ucalgary.syzygy.ca/

https://intro.syzygy.ca/first-steps/

Web-based Jupyter Notebook (with Python Kernel)

No installation or compilation needed.

You can access it using your UCalgary credentials.

IMPORTANT: DEPENDING ON THE SIMULTANEOUS USE, SYZYGY SERVER CAN CRASH AND IT CAN TAKE A WHILE TO RETURN.

FOR THIS REASON, IF YOU WORK ON SYZYGY ALWAYS BACKUP YOUR WORK LOCALLY AFTER USE.

Sign in with your University of Calgary account

Log in

Terms and conditions / Privacy Policy

Python programming

We can use the U of C Jupyter Notebook platform at
https://ucalgary.syzygy.ca/

NO INSTALLATION REQUIRED!

It is highly recommended to have a Python interpreter installed locally in your computers (with Jupyter Notebook). A possibility is the Anaconda package at

https://www.anaconda.com/distribution/

<u>in case syzygy server is down.</u> Python 3.x version is recommended.

Python programming

We can also use the U of C Jupyter Notebook platform at https://talc.ucalgary.ca

NO INSTALLATION REQUIRED!

But to access that, we need to take care of some steps first. For more information, access https://rcs.ucalgary.ca/TALC_Cluster

To access TALC Jupyter platform for the first time, we need to **accept their terms of use using a command line terminal**…

For Windows OS (version 10 above), type ‘cmd’ on the start/search field to initiate the command prompt terminal. There, type:

> ssh username@talc.ucalgary.ca

A confirmation question will be prompted, and you will need to type ‘yes’ to proceed. Then your password will be asked. Note that as you type the password, **you will not see any characters (not even ‘****’) being displayed so make sure you type it carefully.**

Accept the TALC terms of use. Once you do that, you can type ‘exit’ or ‘logout’ to get out of the terminal.

Open your web browser and type https://talc.ucalgary.ca. Use your UofC credentials to log in. On the ‘username’ field type ONLY the username. You may need to go to the ‘Control Panel’ of the platform and reset the server by clicking on ‘My Server’.

PS: The same steps above are valid for other OS such as macOS and Linux. You just need to start a console terminal that accepts line commands and execute the ssh string.

If you need to continue using TALC **off-campus**, remember that you will need the UofC General VPN FortiClient!

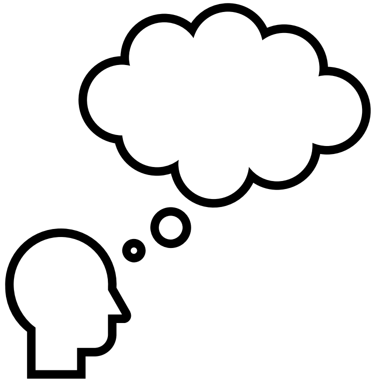
For more information, visit https://www.ucalgary.ca/working-and-learning-home/remote-access-instructions

and set up your VPN **in accordance with the UofC IT guidelines.**

How to make sense out of data?

Classical inference: mathematical activity or process with the purpose of extracting meaningful information from measurements (and data) in order to characterize systems, establish best model parameters, or determine plausibility (truthfulness) of hypotheses and models.

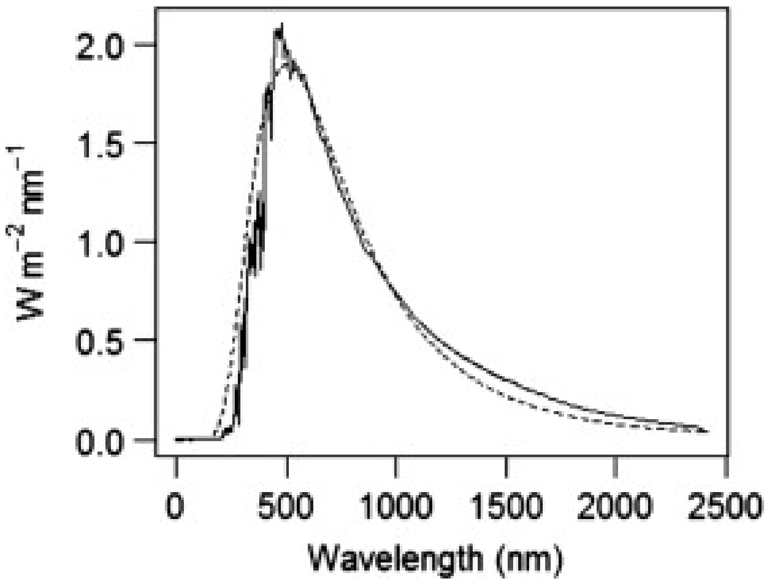
- Optimization (e.g., fitting)
- Estimators (based on the determination of distribution metrics such as moments, and variances with comparison with some model prediction)
- Confidence Intervals and Statistical Tests
- Bayesian Inference



C. A. Pruneau, Data Analysis Techniques for Physical Scientists, Cambridge University Press, 2022.

Curve Fitting and Optimization Problems

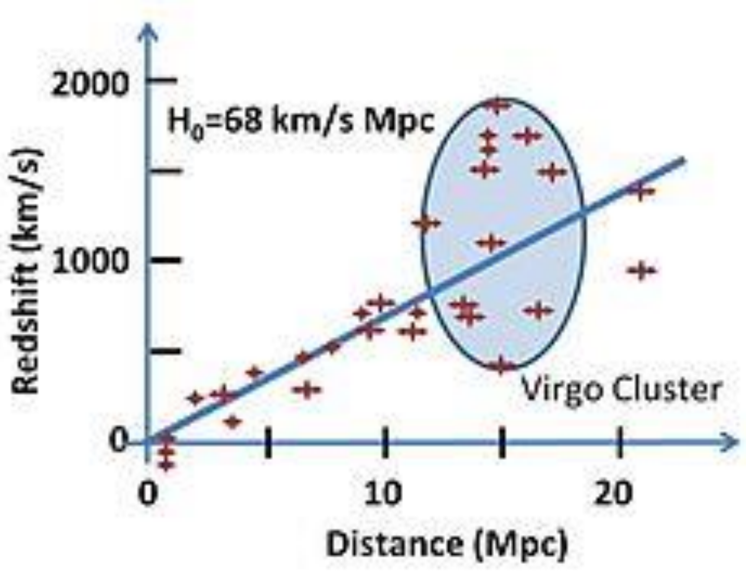
Black-body radiation (Planck's law)



Satellite-observed solar spectrum during the Solar Radiation and Climate Experiment (SORCE) mission (solid curve). Theoretical spectrum from Planck's law (dashed line) at T = 5750 K.

Source: Jinyou Liang, in Chemical Modeling for Air Resources, 2013.

Hubble's law (and the expanding universe)



Keel, W. C. (2007). The Road to Galaxy Formation (2nd ed.). Springer.

$$\boxed{v = H_0 d}$$

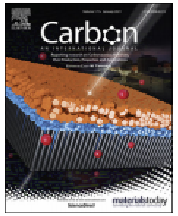
v: Recessional velocity of galaxy
H_0: Hubble constant
d: Distance to galaxy

Carbon 171 (2021) 306–319

Contents lists available at ScienceDirect

Carbon

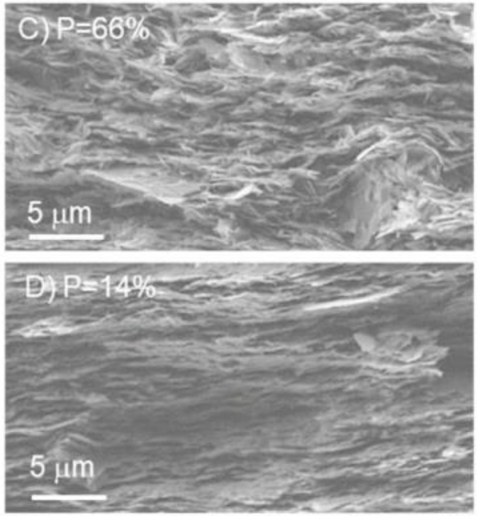
journal homepage: www.elsevier.com/locate/carbon



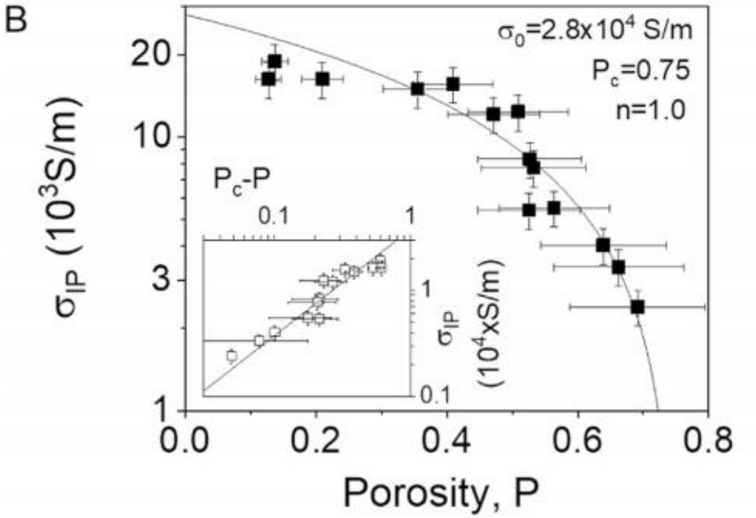
On the relationship between morphology and conductivity in nanosheet networks

Sebastian Barwich [a, b], João Medeiros de Araújo [c], Aran Rafferty [a], Claudia Gomes da Rocha [d], Mauro S. Ferreira [a, b], Jonathan N. Coleman [a, b, *]

[a] Centre for Research on Adaptive Nanostructures and Nanodevices (CRANN) & Advanced Materials and Bioengineering Research (AMBER) Centre, Trinity College Dublin, Dublin 2, Ireland
[b] School of Physics, Trinity College Dublin, Dublin 2, Ireland
[c] Departamento de Física Teórica e Experimental, Universidade Federal Do Rio Grande Do Norte, Natal, RN, 59078-900, Brazil
[d] Department of Physics and Astronomy, University of Calgary, 2500 University Drive NW Calgary, Alberta, T2N 1N4, Canada



Compressed graphene films (to control the conductivity of the films)



In-Plane film conductivity, σ_{IP}, as a function of film porosity, P. Inset: Percolation plot of σ_{IP} vs. $P_c - P$ fitted by percolation theory

$$\sigma = \sigma_0 \left(\frac{P_c - P}{P_c} \right)^n$$

nature COMMUNICATIONS

ARTICLE

Check for updates

https://doi.org/10.1038/s41467-020-20687-y OPEN

Mathematical model of COVID-19 intervention scenarios for São Paulo—Brazil

Osmar Pinto Neto[1,2,3✉], Deanna M. Kennedy[4], José Clark Reis[2], Yiyu Wang[4], Ana Carolina Brisola Brizzi[1,2], Gustavo José Zambrano[2], Joabe Marcos de Souza[2,5], Wellington Pedroso[1,2], Rodrigo Cunha de Mello Pedreiro[1,6,7], Bruno de Matos Brizzi[2], Ellysson Oliveira Abinader[8] & Renato Amaro Zângaro[1,3]

Check for updates

Identifying airborne transmission as the dominant route for the spread of COVID-19

Renyi Zhang[a,b,1], Yixin Li[b], Annie L. Zhang[c], Yuan Wang[d], and Mario J. Molina[e,1]

[a]Department of Atmospheric Sciences, Texas A&M University, College Station, TX 77843; [b]Department of Chemistry, Texas A&M University, College Station, TX 77843; [c]Department of Chemistry, College of Natural Sciences, The University of Texas at Austin, Austin, TX 78712; [d]Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA 91125; and [e]Department of Chemistry and Biochemistry, University of California San Diego, La Jolla, CA 92093

Contributed by Mario J. Molina, May 16, 2020 (sent for review May 14, 2020; reviewed by Manish Shrivastava and Tong Zhu)

www.pnas.org/cgi/doi/10.1073/pnas.2009637117

PLOS ONE

advanced searc

OPEN ACCESS PEER-REVIEWED

RESEARCH ARTICLE

Modeling the impact of public response on the COVID-19 pandemic in Ontario

Brydon Eastman, Cameron Meaney, Michelle Przedborski, Mohammad Kohandel

Published: April 14, 2021 • https://doi.org/10.1371/journal.pone.0249456

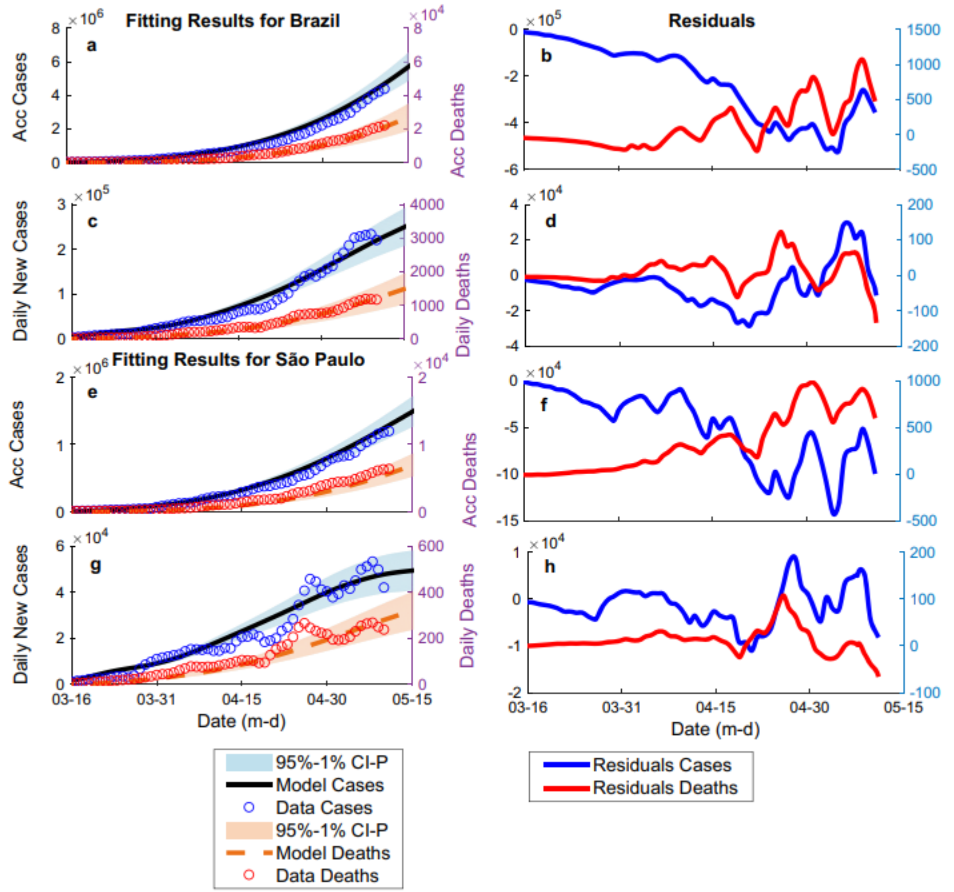


Fig. 1 Fitting results for cases and deaths for Brazil and São Paulo. a Corrected accumulated (Acc) cases and deaths for Brazil. **b** Acc data fitting residuals for Brazil. **c** Corrected daily new cases and deaths for Brazil. **d** Daily data fitting residuals for Brazil. **e** Corrected Acc cases and deaths for São Paulo. **f** Acc data fitting residuals for São Paulo. **g** Corrected daily new cases and deaths for São Paulo. **h** Daily data fitting residuals for São Paulo. Black lines represent the best-fit model expected cases; red dashed lines represent the best-fit model expected deaths. Blue circles are COVID-19 official data cases' data, and red circles are official deaths, both corrected by sub-testing factors. Blue and red shaded regions show confidence intervals of 95%, considering the 2.5 and 97.5% quantiles of the distribution of $n = 300$ uniformly distributed 1% errors or perturbations done to the best-fit model parameters.

Least-Squares Method

Given some set of real observations $\tilde{y}_n$ corresponding to independent variables x_n, we are interested in a model $\hat{y}(x)$ which is a "good fit" to the observations. Most measures of 'fit quality' are based on differences between the model and data **(usually called "residuals")**:

$$r_n = \tilde{y}_n - \hat{y}_n$$

If the model and data are the same, then all residuals will be zero. In practice, there will invariably be some discrepancies between data and model predictions due to:

1. **Data error:** random variations in the observations.
2. **Model error:** systematic differences between our imperfect model $\hat{f}(x)$ and the "true" system function $f(x)$.

Data error ε can be expressed in terms of observations (known) and the underlying true values that we are trying to determine:

$$\tilde{y}_n = y_n + \varepsilon_n$$

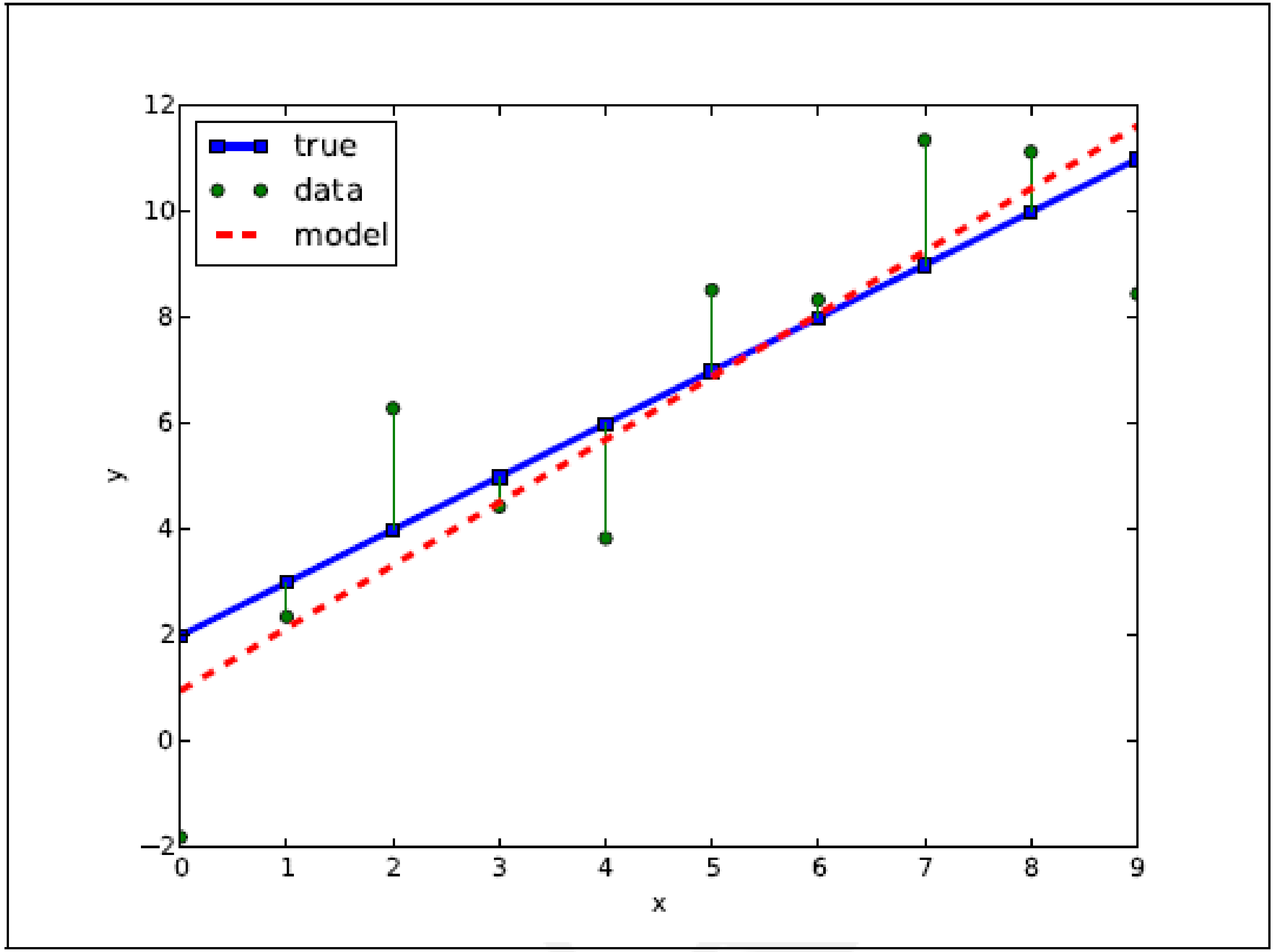


Figure 2.2: Linear process (solid blue) with random errors (green) and best-fit model (dashed red)

Source: Dr. Jackel's PHYS 605 notes