UNIVERSITY of WASHINGTON

rateExtrapolation.py

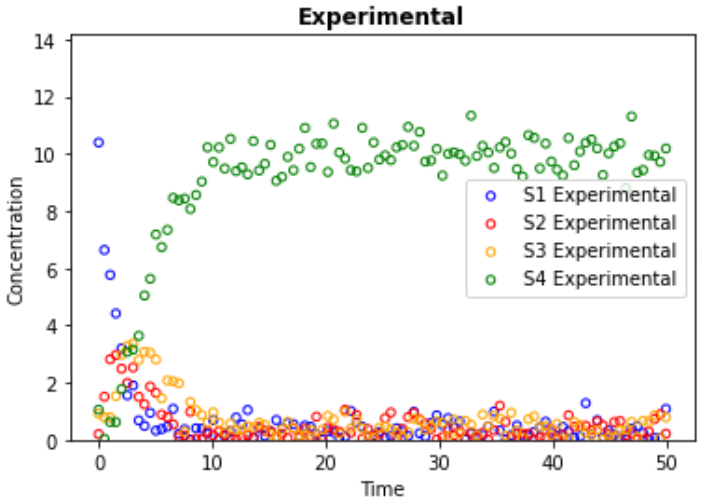
simple reaction rate extrapolation from experimental data

BIOEN537
Sarah Wait
Fall 2021

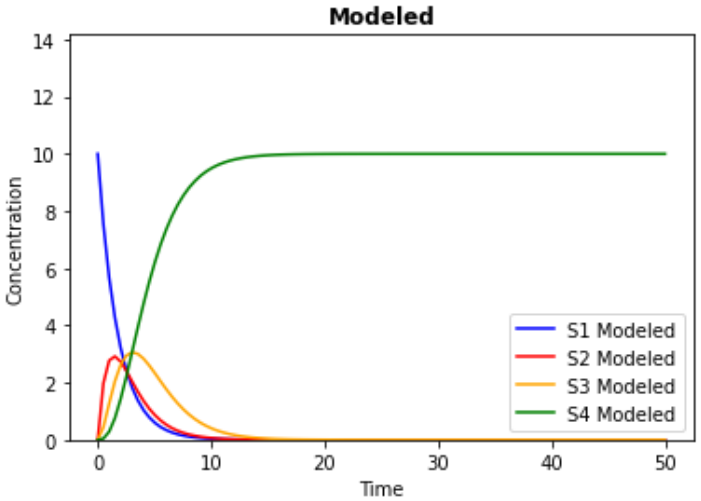
Background: *Empirical data generation and model fitting steps are currently segmented.*

Experimentalists

Computationalists

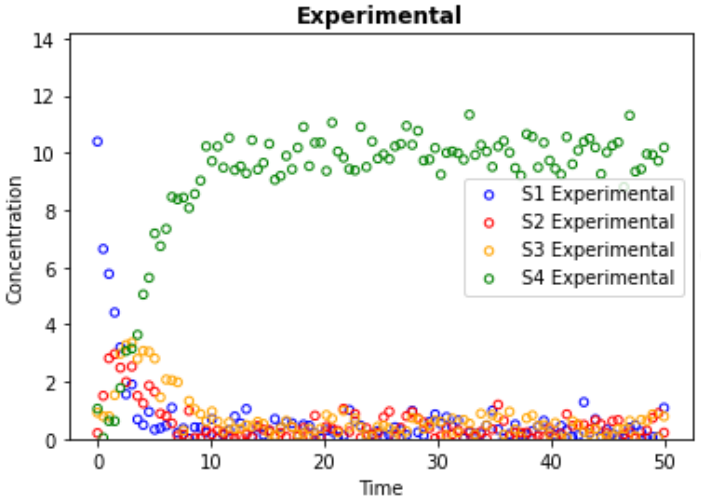


- Parameter Fitting
- Cross Validation
- Summary of Model Fitting



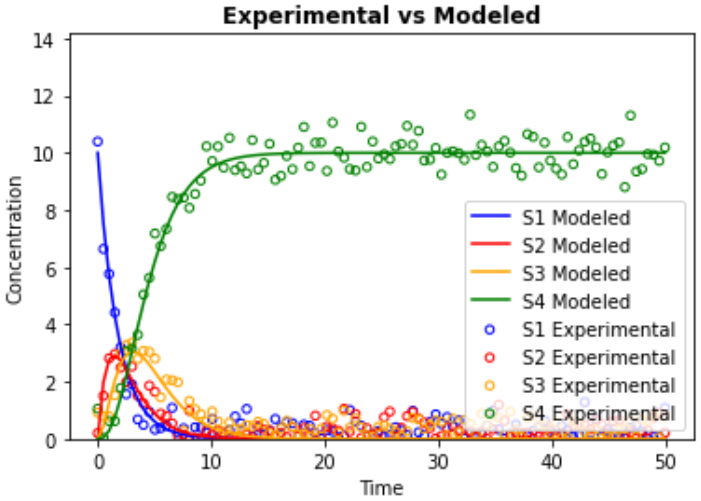
Goal: *Create a simplified function that an experimentalist can utilize to quickly model simple reactions*

Experimentalists



rateExtrapolation.py

- Parameter Fitting
- Cross Validation
- Summary of Model Fitting



Intended Users: *The intended users are experimentalists who have some experience in coding.*

Experimentalists (& computationalists)

Experience coding in Python

Ideally, understands data structures such as float, int, string etc.

Able to import and execute packages in python

e.g. import pandas as pd
& y = rateExtrapolation(data)

Able to write Tellurium Antimony String

e.g. antimony = """
*S1 -> S2; k1*S1;*
*S2 -> S3; k2*S2;*
S1 =10; S2 = 0; S3 = 0;
k1= 0; k2 =0 """

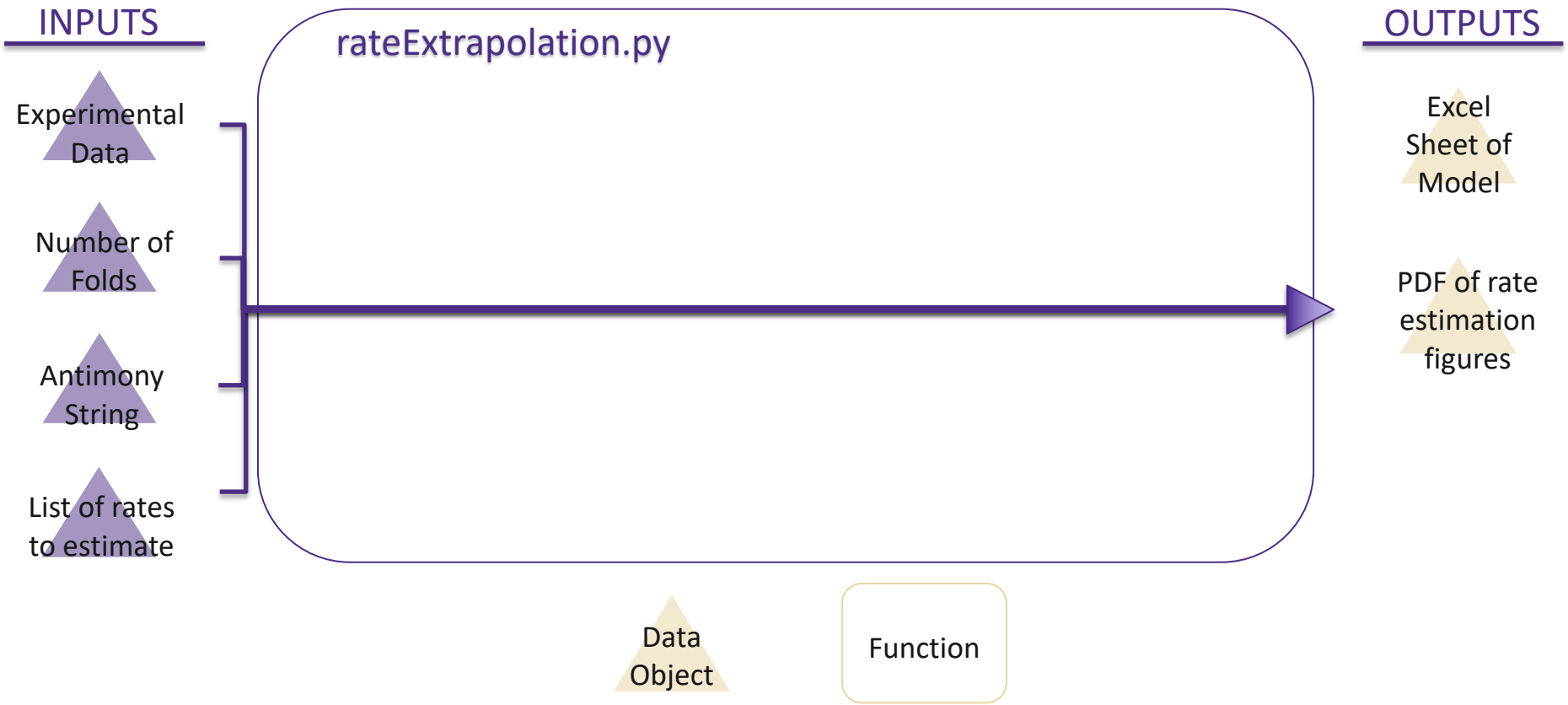
Use cases: *Experimentalist does all steps including data generation, code execution, and analysis of results.*

- User generates time course data for biolgical network of interest

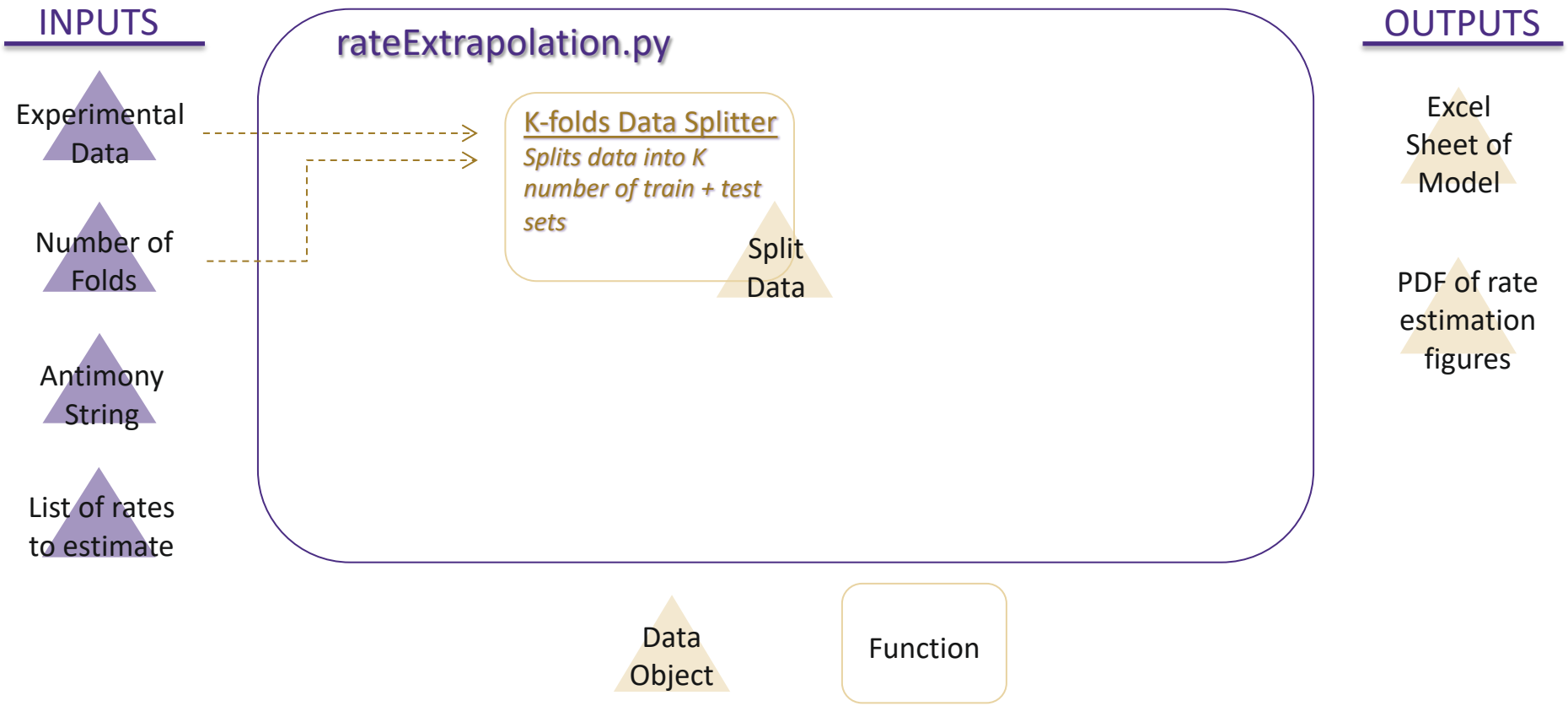
- User opens python notebook
- loads the rateExtrapolation function
- User provides input and executes function

- User reads provided results from the function
- Accepts or rejects the extrapolated rates

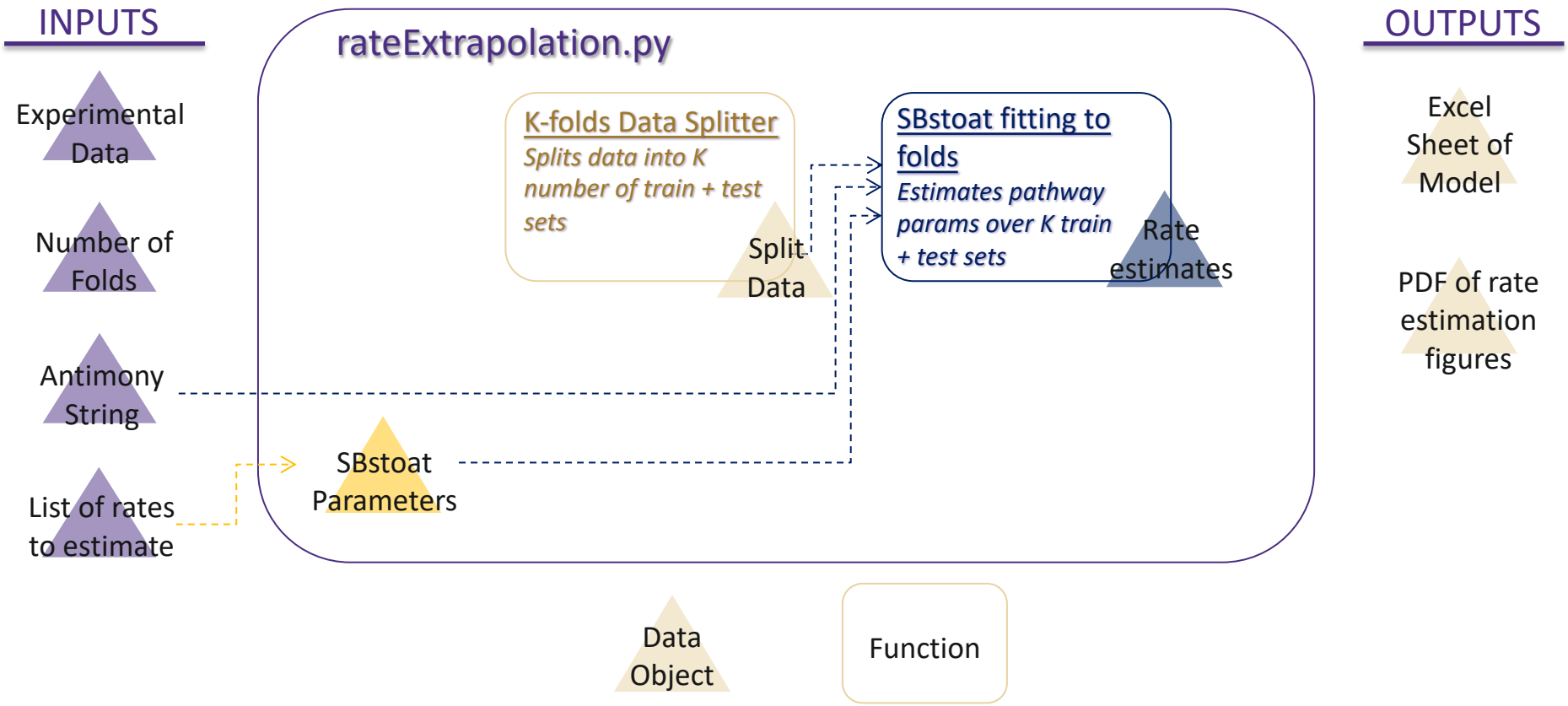
Design


INPUTS
Experimental Data
Number of Folds
Antimony String
List of rates to estimate
rateExtrapolation.py
OUTPUTS
Excel Sheet of Model
PDF of rate estimation figures
Data Object
Function

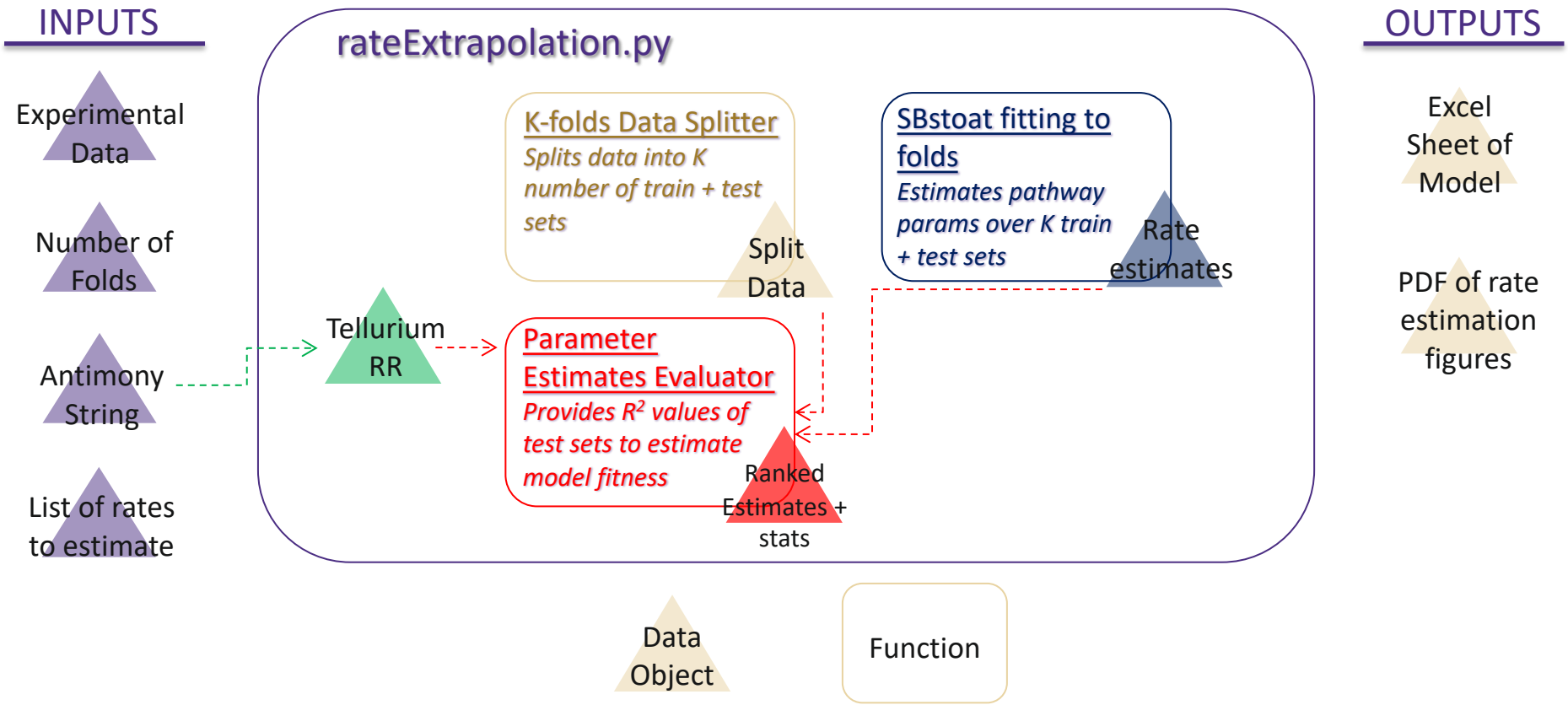
Design


INPUTS
Experimental Data
Number of Folds
Antimony String
List of rates to estimate
rateExtrapolation.py
K-folds Data Splitter
Splits data into K number of train + test sets
Split Data
OUTPUTS
Excel Sheet of Model
PDF of rate estimation figures
Data Object
Function

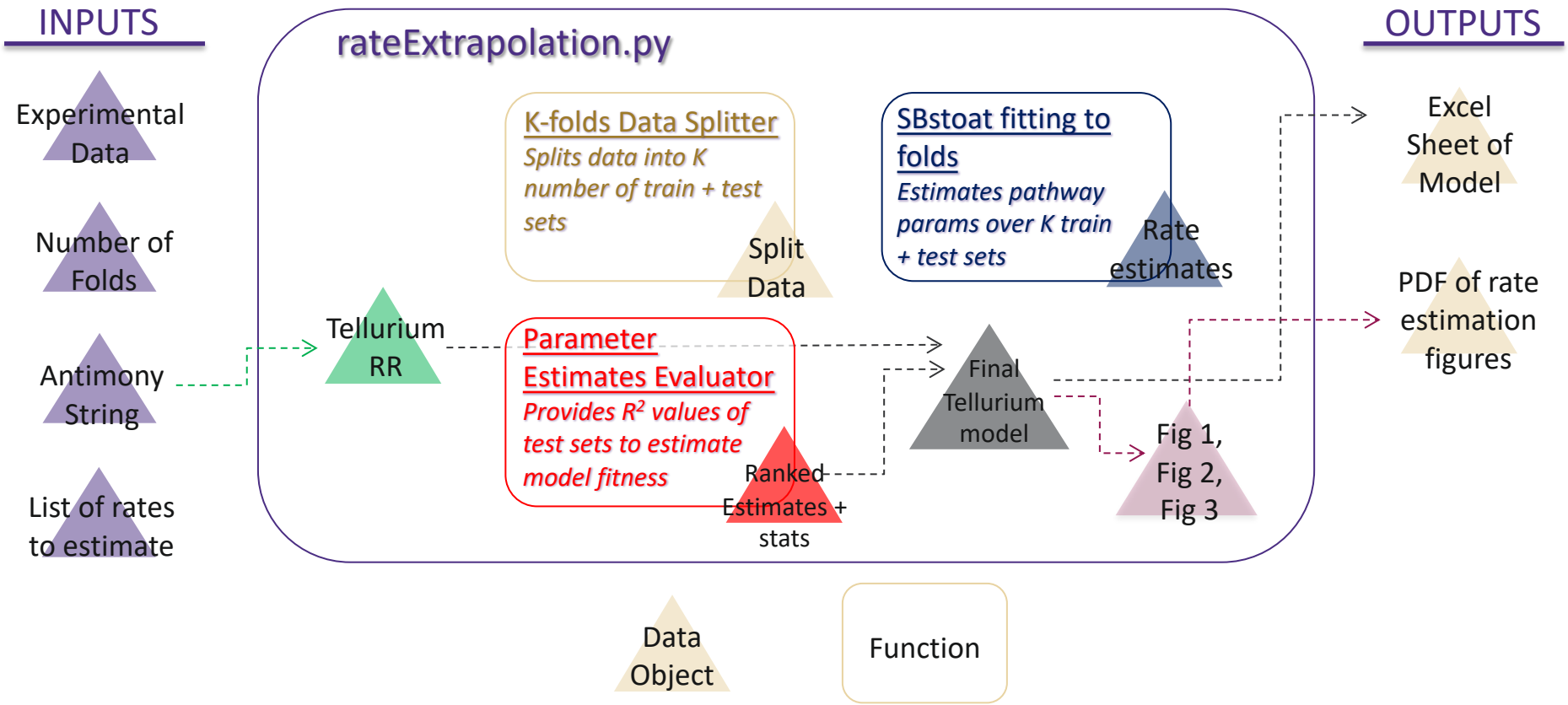
Design


INPUTS
Experimental Data
Number of Folds
Antimony String
List of rates to estimate
rateExtrapolation.py
K-folds Data Splitter
Splits data into K number of train + test sets
Split Data
SBstoat fitting to folds
Estimates pathway params over K train + test sets
Rate estimates
SBstoat Parameters
OUTPUTS
Excel Sheet of Model
PDF of rate estimation figures
Data Object
Function

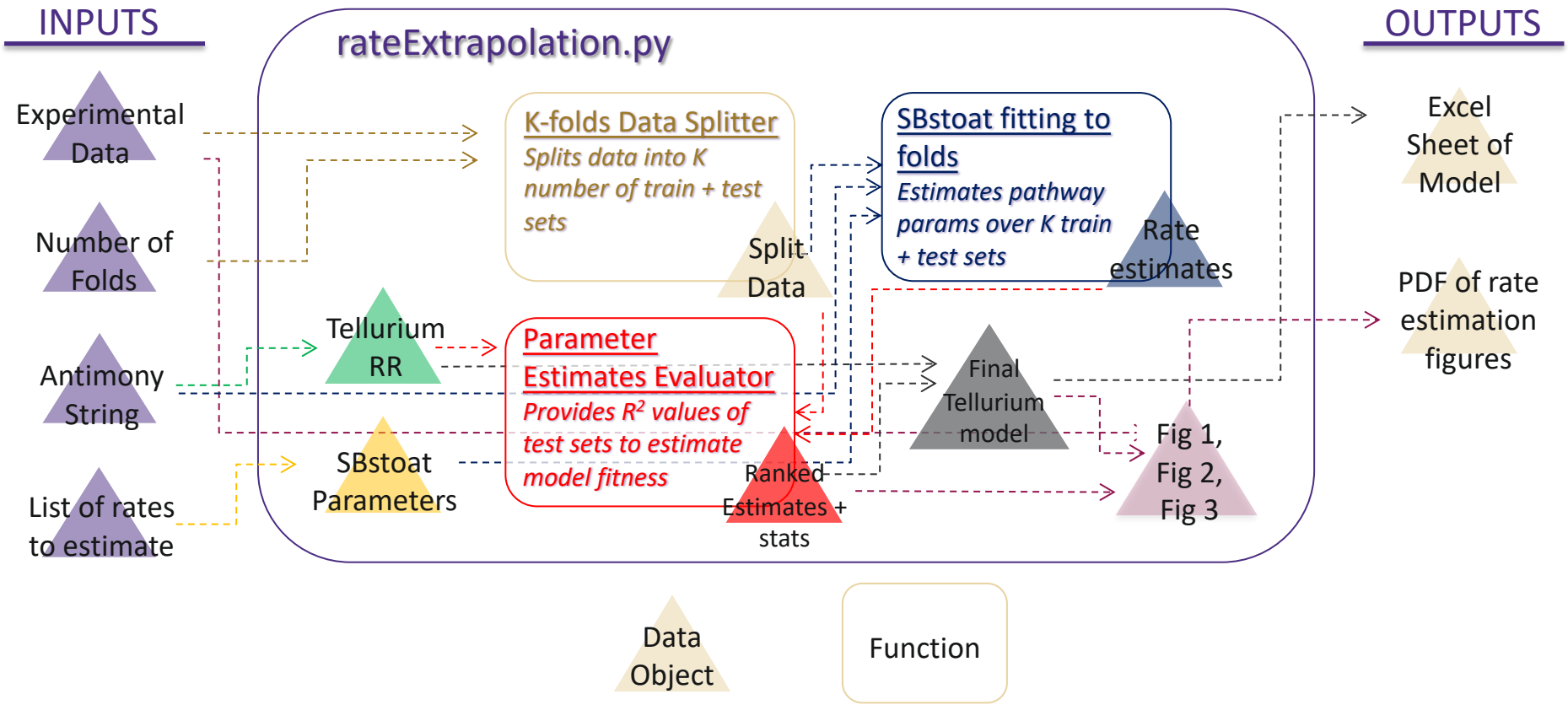
Design

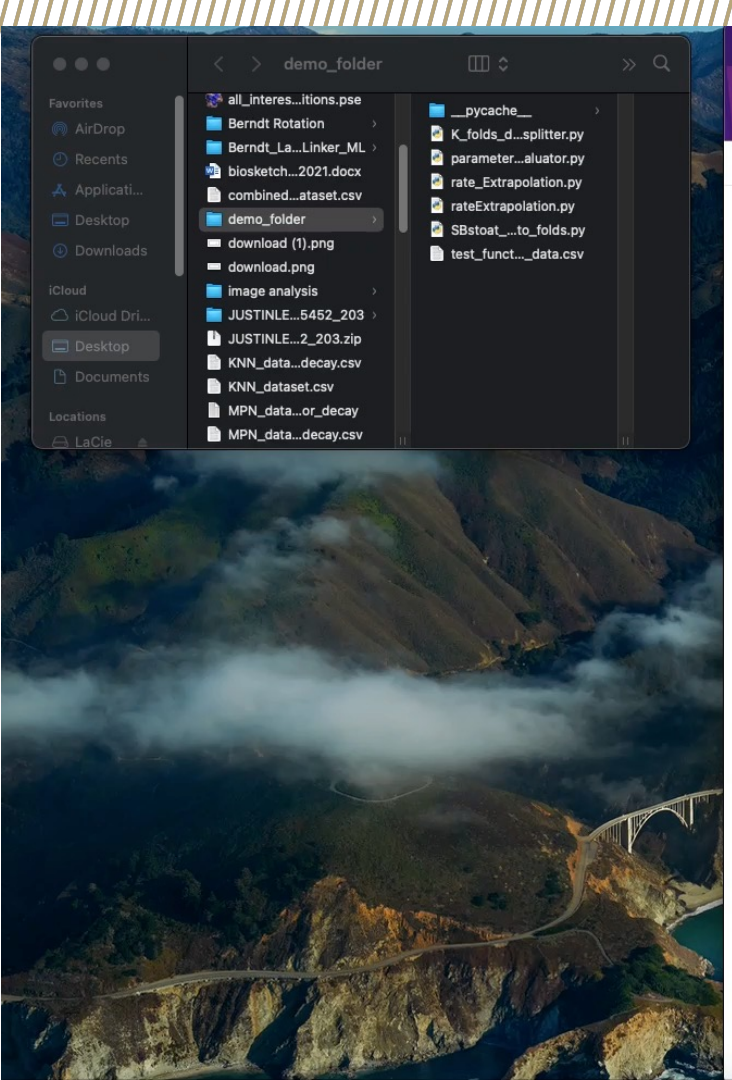

INPUTS
Experimental Data
Number of Folds
Antimony String
List of rates to estimate
rateExtrapolation.py
K-folds Data Splitter
Splits data into K number of train + test sets
Split Data
SBstoat fitting to folds
Estimates pathway params over K train + test sets
Rate estimates
Tellurium RR
Parameter Estimates Evaluator
Provides R^2 values of test sets to estimate model fitness
Ranked Estimates + stats
OUTPUTS
Excel Sheet of Model
PDF of rate estimation figures
Data Object
Function

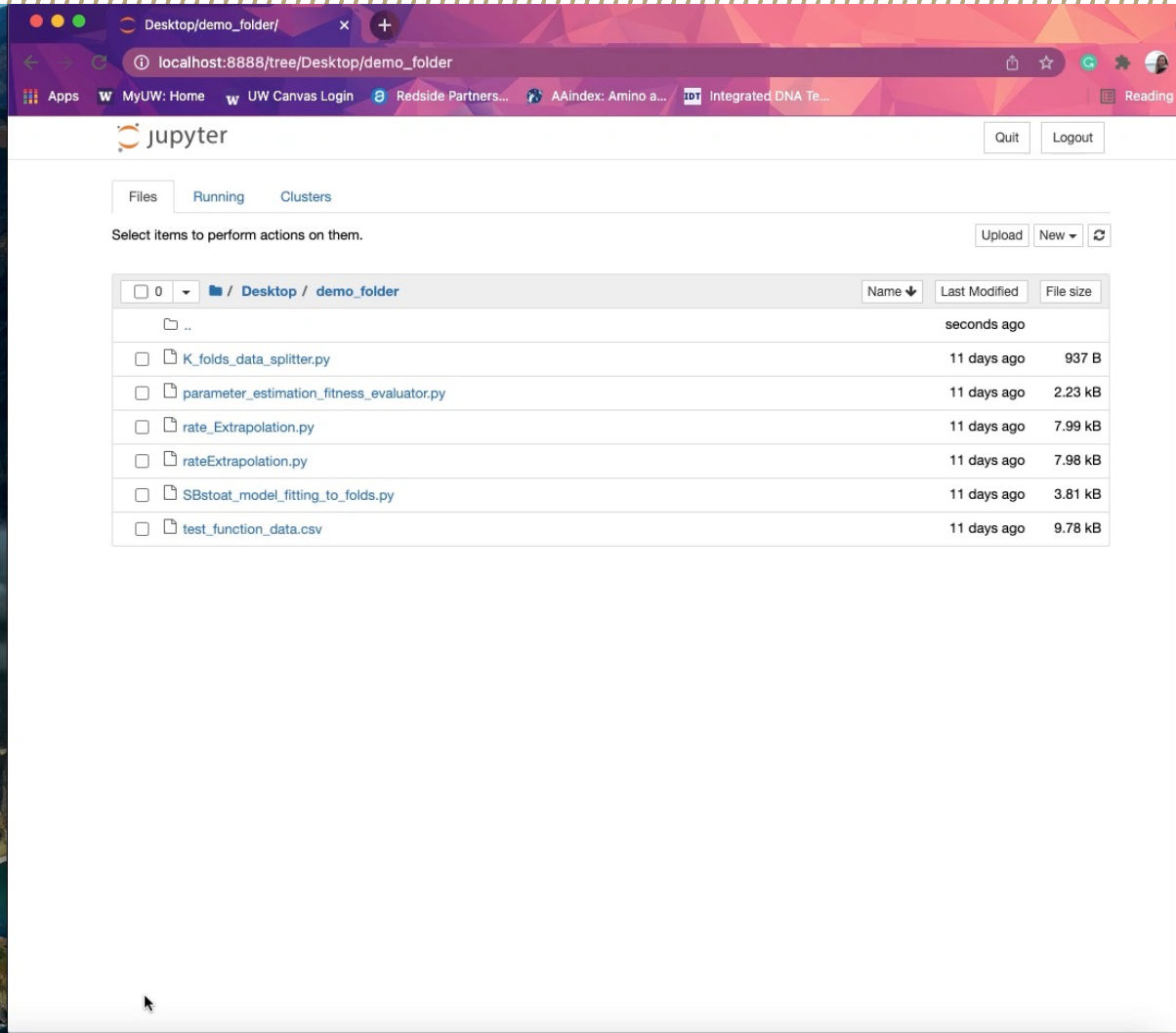
Design


INPUTS
Experimental Data
Number of Folds
Antimony String
List of rates to estimate
rateExtrapolation.py
K-folds Data Splitter
Splits data into K number of train + test sets
Split Data
SBstoat fitting to folds
Estimates pathway params over K train + test sets
Rate estimates
Tellurium RR
Parameter Estimates Evaluator
Provides R^2 values of test sets to estimate model fitness
Ranked Estimates + stats
Final Tellurium model
Fig 1, Fig 2, Fig 3
OUTPUTS
Excel Sheet of Model
PDF of rate estimation figures
Data Object
Function

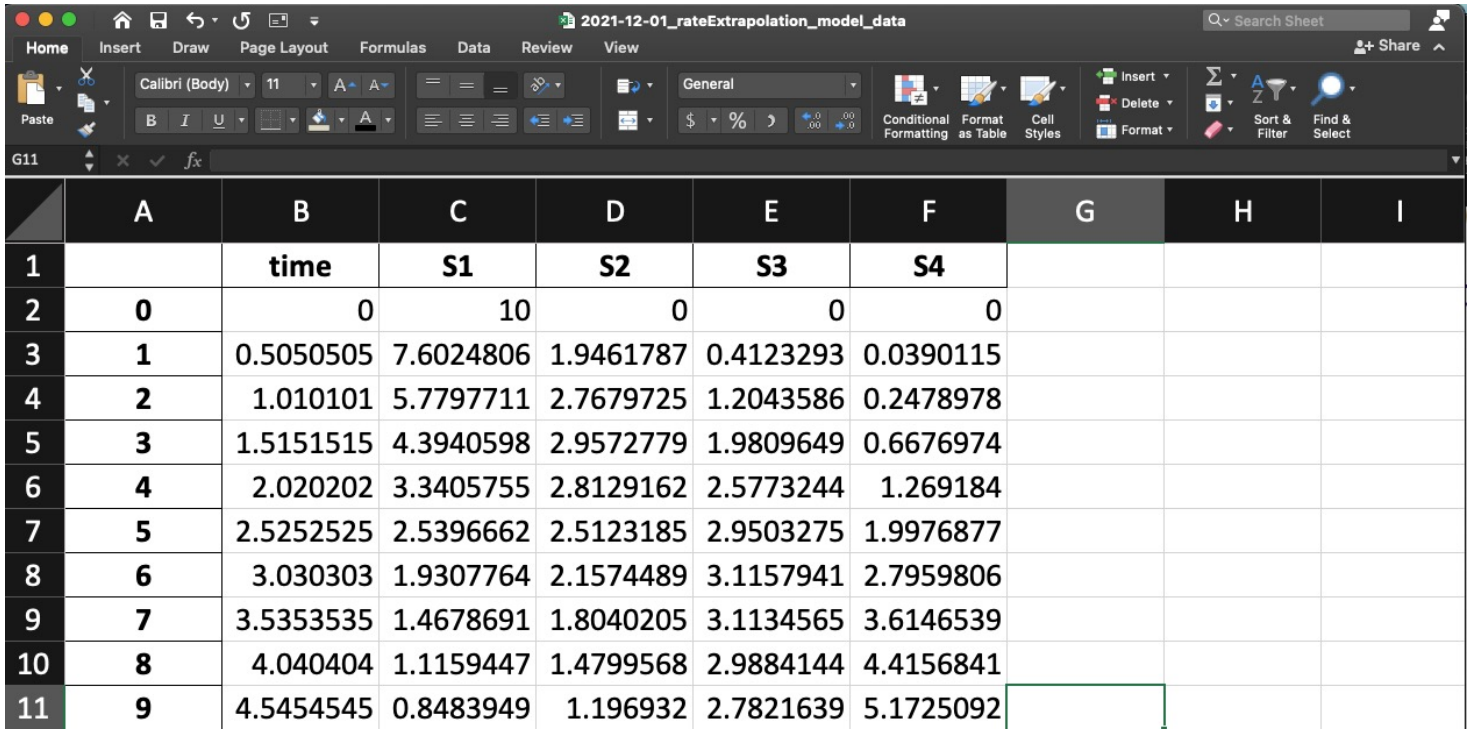
Design


INPUTS
OUTPUTS
rateExtrapolation.py
Experimental Data
Number of Folds
Antimony String
List of rates to estimate
K-folds Data Splitter
Splits data into K number of train + test sets
Split Data
SBstoat fitting to folds
Estimates pathway params over K train + test sets
Rate estimates
Tellurium RR
SBstoat Parameters
Parameter Estimates Evaluator
Provides R^2 values of test sets to estimate model fitness
Ranked Estimates + stats
Final Tellurium model
Fig 1, Fig 2, Fig 3
Excel Sheet of Model
PDF of rate estimation figures
Data Object
Function


demo_folder
Favorites
AirDrop
Recents
Applicati...
Desktop
Downloads
iCloud
iCloud Dri...
Desktop
Documents
Locations
LaCie
all_interes...itions.pse
Berndt Rotation
Berndt_La...Linker_ML
biosketch...2021.docx
combined...ataset.csv
demo_folder
download (1).png
download.png
image analysis
JUSTINLE...5452_203
JUSTINLE...2_203.zip
KNN_data...decay.csv
KNN_dataset.csv
MPN_data...or_decay
MPN_data...decay.csv
__pycache__
K_folds_d...splitter.py
parameter...aluator.py
rate_Extrapolation.py
rateExtrapolation.py
SBstoat_...to_folds.py
test_funct..._data.csv

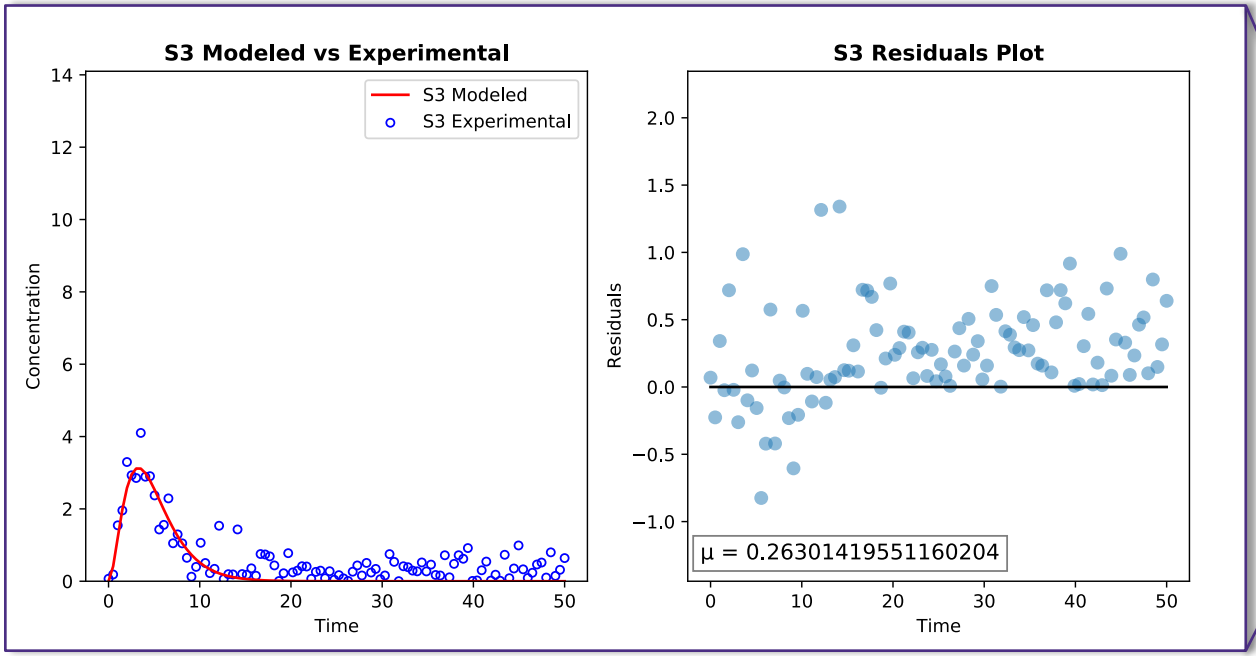

Desktop/demo_folder/
localhost:8888/tree/Desktop/demo_folder
Apps
MyUW: Home
UW Canvas Login
Redside Partners...
AAindex: Amino a...
Integrated DNA Te...
Reading
jupyter
Quit
Logout
Files
Running
Clusters
Select items to perform actions on them.
Upload
New
0
/ Desktop / demo_folder
Name
Last Modified
File size
..
seconds ago
K_folds_data_splitter.py
11 days ago
937 B
parameter_estimation_fitness_evaluator.py
11 days ago
2.23 kB
rate_Extrapolation.py
11 days ago
7.99 kB
rateExtrapolation.py
11 days ago
7.98 kB
SBstoat_model_fitting_to_folds.py
11 days ago
3.81 kB
test_function_data.csv
11 days ago
9.78 kB

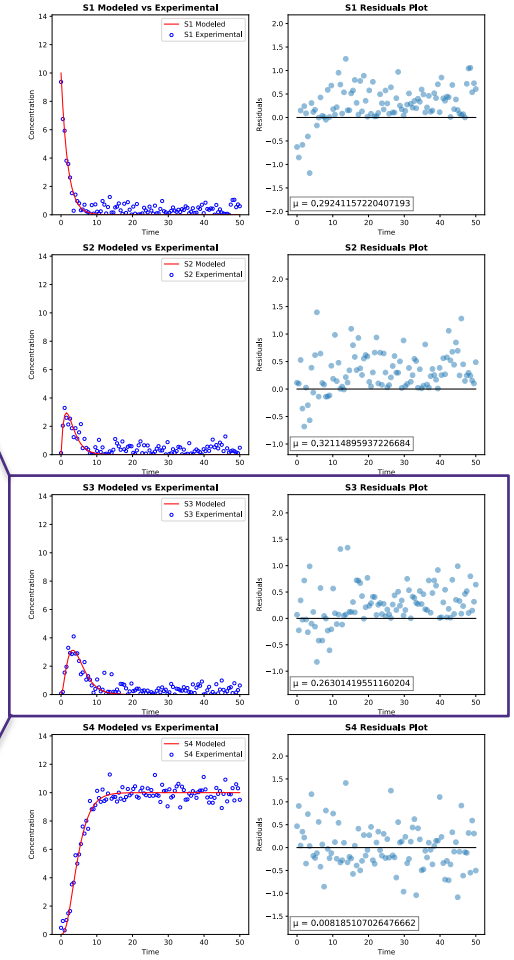
Output Excel File: *Simulated data for all substrates, not just those that were estimated*



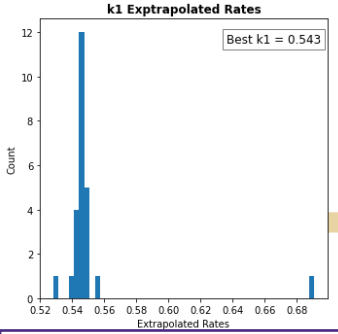
	A	B	C	D	E	F	G	H	I
1		time	S1	S2	S3	S4			
2	0	0	10	0	0	0			
3	1	0.5050505	7.6024806	1.9461787	0.4123293	0.0390115			
4	2	1.010101	5.7797711	2.7679725	1.2043586	0.2478978			
5	3	1.5151515	4.3940598	2.9572779	1.9809649	0.6676974			
6	4	2.020202	3.3405755	2.8129162	2.5773244	1.269184			
7	5	2.5252525	2.5396662	2.5123185	2.9503275	1.9976877			
8	6	3.030303	1.9307764	2.1574489	3.1157941	2.7959806			
9	7	3.5353535	1.4678691	1.8040205	3.1134565	3.6146539			
10	8	4.040404	1.1159447	1.4799568	2.9884144	4.4156841			
11	9	4.5454545	0.8483949	1.196932	2.7821639	5.1725092			

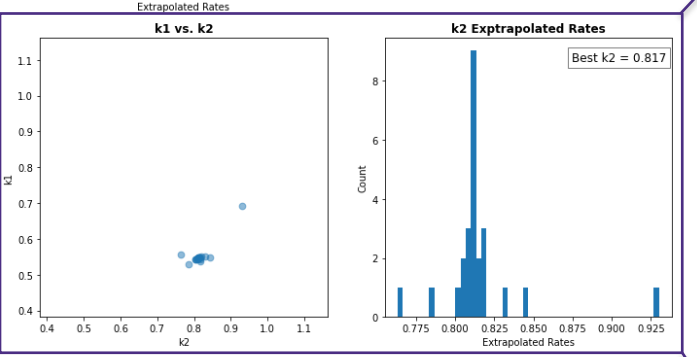
Output Figure #1: *plots of modeled vs experimental data + residuals plot*

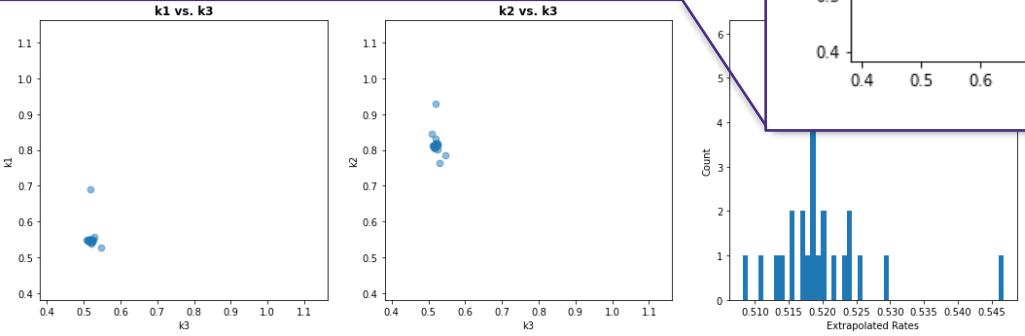


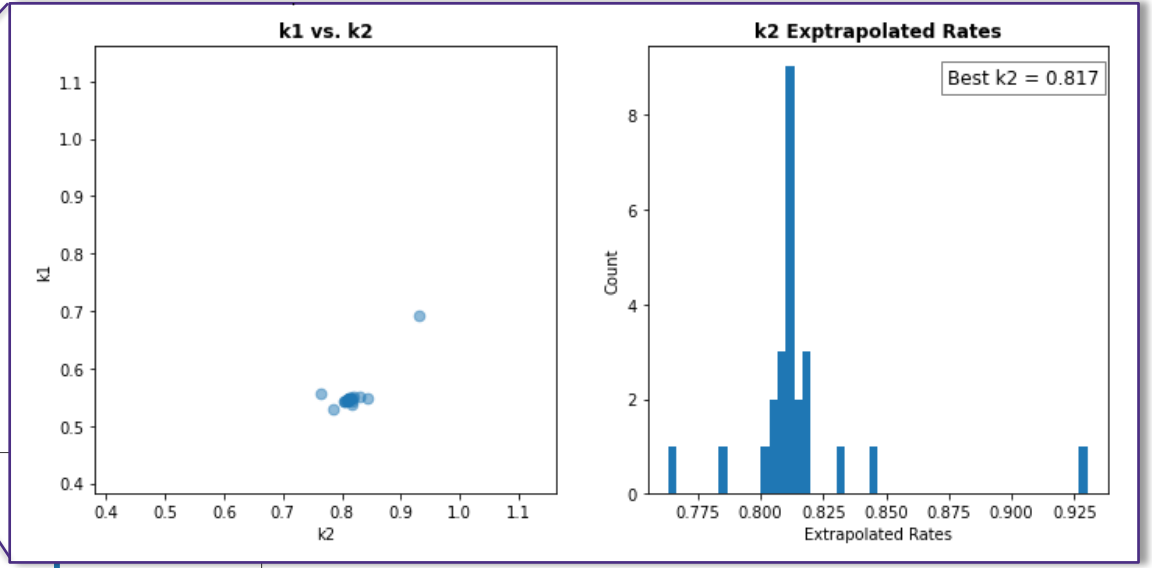


Output Figure #2: *Histograms of all estimations of rates, along with rate analysis*

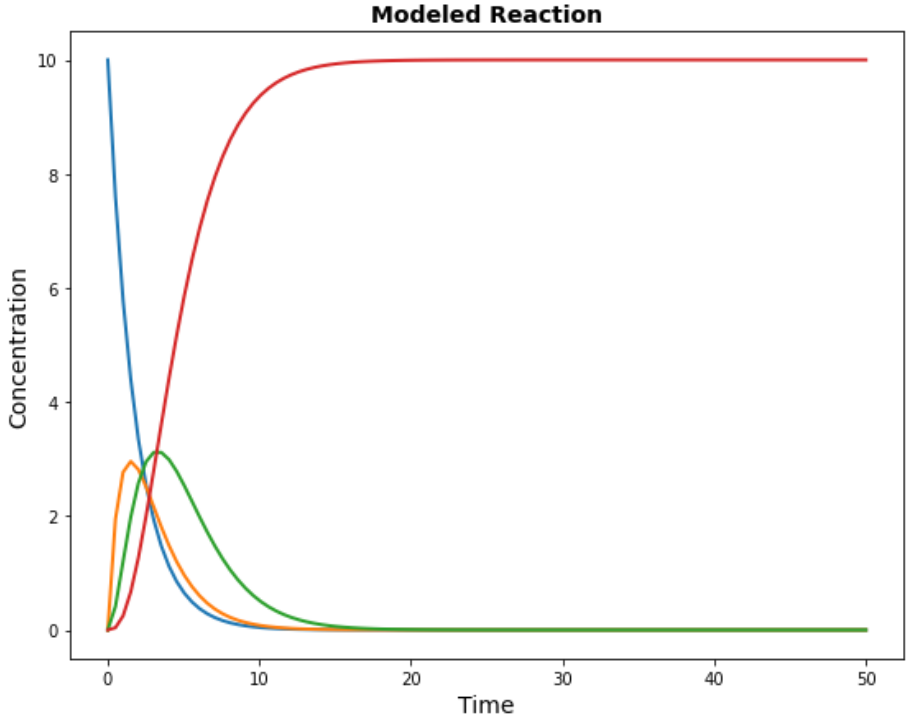


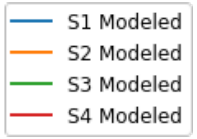






Output Figure #3: *Final summary plot of model, along with estimated parameters and statistics of model fit*





k1	0.543
k2	0.817
k3	0.518
AIC	-593.09
BIC	-581.24
Reduced χ^2	0.21
χ^2	80.68
R^2	0.988

Project Structure

TO GITHUB

https://github.com/sarahwaity/rateExtrapolation

Lessons learned and future work

- Definitely easier to start continual integration practices at the beginning of a project
  - Spent a very long time trying to get tests to pass
- Being explicit in naming not only helps future you but also limits probability of current you running into issues
- Future work:
  - Add more modularity to the function itself, (currently limited to ‘leastsquare’ extrapolation methods etc.)
  - Maybe export the PDF images into a succinct PowerPoint format
  - Be able to do all extrapolations from the command line