Grid Search Methods for Modeling Polymerization Kinetics

Katie Ziebarth
Landis Group, Department of Chemistry, UW-Madison
HTCondor Week, May 2019

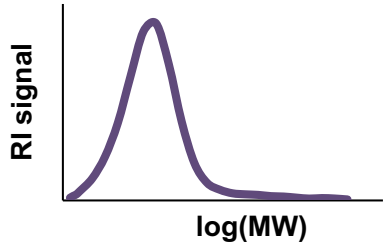
Polymers

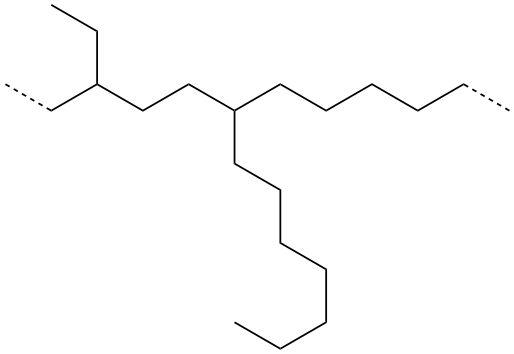
- Consist of long chains of monomers linked together
- >300 million tons of polymers produced per year
- Ubiquitous in modern society

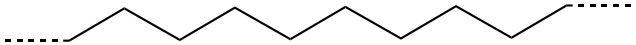
Plastics - the Facts 2017. PlasticsEurope 2018.
Images: https://shop.crayola.com/color-and-draw/markers, https://www.watersandfarr.co.nz/product/imperial-low-density-polyethylene-pe-pipe/, https://www.amazon.com/American-Plastic-Toys-Assorted-Colors/dp/B002RQU3VQ, https://www.uline.com/Cls_35/Food-Service-and-Packaging, https://www.consumerreports.org/new-cars/best-and-worst-new-cars/, https://www.pinterest.com/pin/353321533236396447/

Polymer Properties

- Bulk properties are determined by molecular level features, including
  - Microstructure
  - Molecular weight distribution
  - Nature and sequence of monomers
- Molecular properties are controlled by relative rates of reactions







Images: https://www.newsreck.com/global-linear-low-density-polyethylene-lldpe-market-insights-deap-analysis-2019-2024-dow-exxonmobil-sabic-borealis-nova-chemicals/44071/, https://www.generalkinematics.com/blog/different-types-plastics-recycled/, http://reiloyusa.com/industry-applications/high-density-polyethylene/

Polymerization Kinetics

Kinetic models:

- Consist of a set of rate equations for every species in the system
  - Allows for simulations under any reaction conditions
  - Can predict molecular weight distribution, microstructure, etc.

- Account for the different possible reactions and their relative rates

- Require optimization of certain parameters (rate constants) to best fit the experimental data

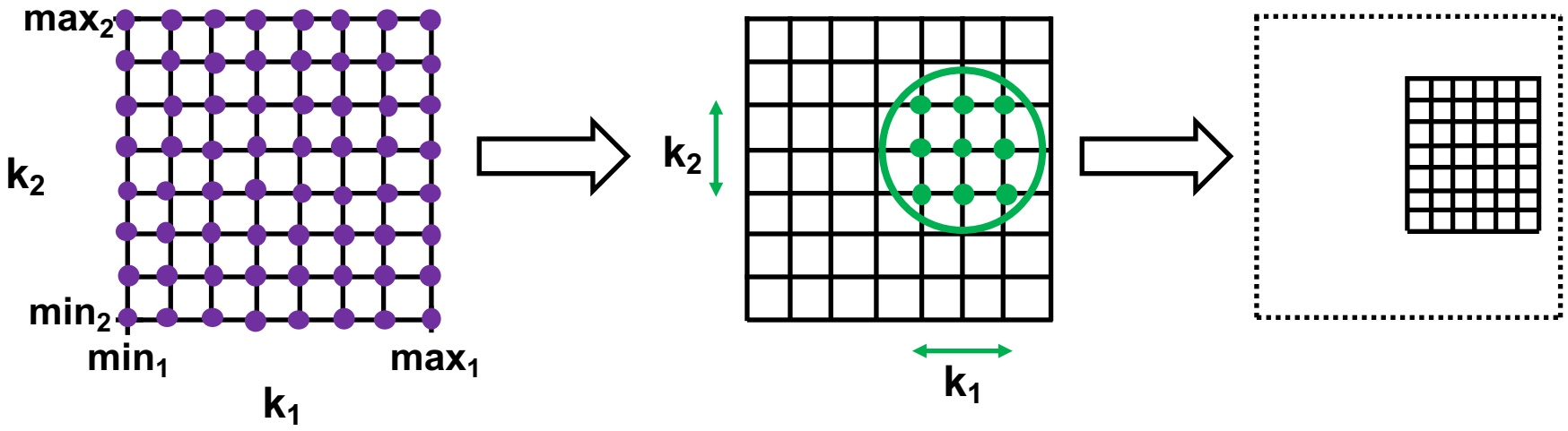
Challenges:

- Typically involve 1000s of ODEs

- Iterative optimization procedures are slow and prone to getting stuck in local minima

Alternative Approach: Grid Searching

- Run simulations for many different combinations of parameters
- Compare simulations to experiment – look for best matches
- Use bootstrapping analysis to examine uncertainty
  - Get distribution of best matches
- Use results to guide selection of an adjusted set of parameters
  - Goal is to have an increasingly fine grid to improve precision of result



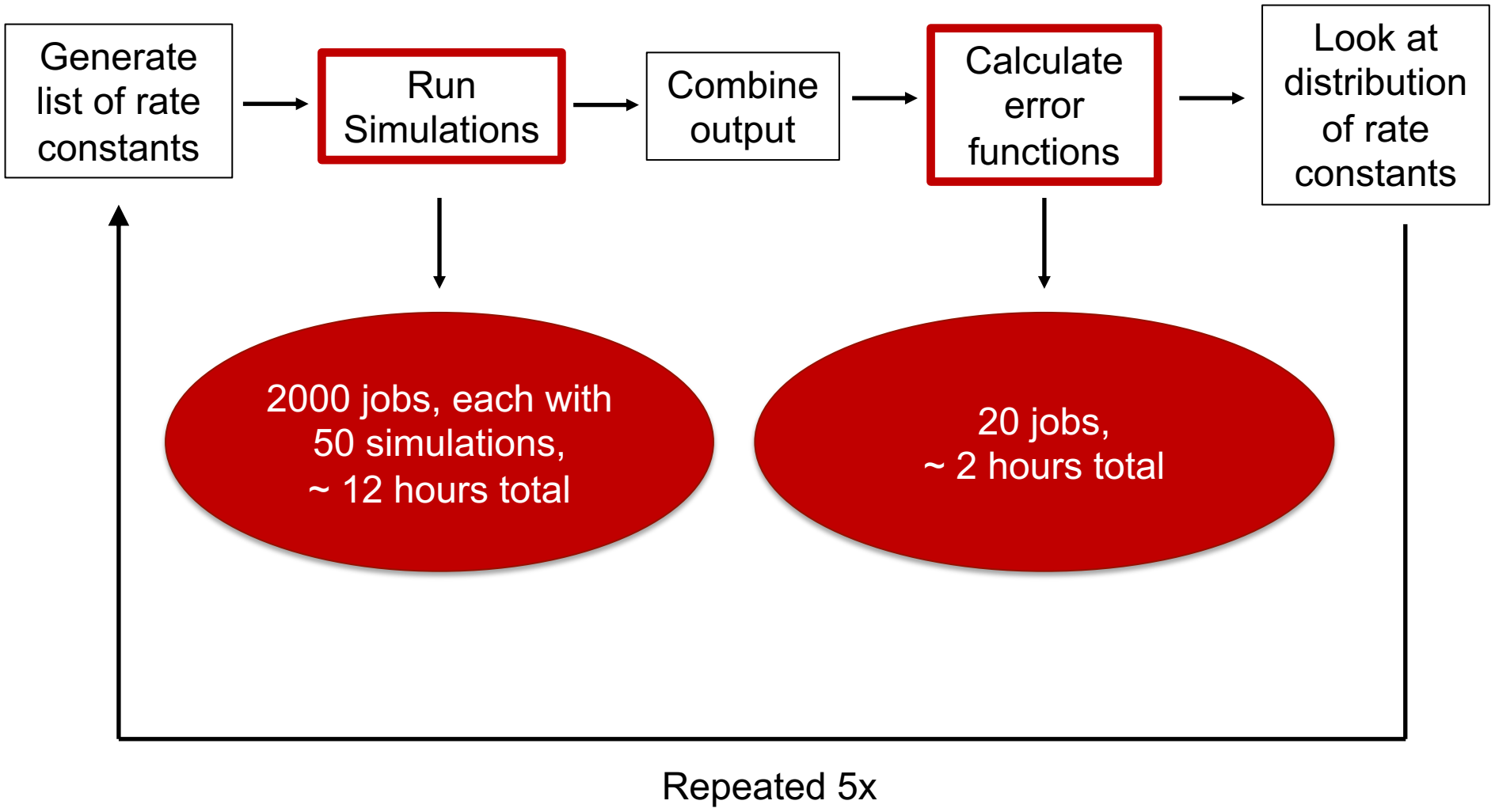
Calculation Set-Up

- Zr-catalyzed 1-hexene polymerization
  - 1658 ODEs
  - 5 parameters to optimize
  - 2443 data points to use for fitting

- Generate kinetic model using SimBiology toolbox in MATLAB
  - Provide chemical species, reactions, and experimental observables

- Perform grid search
  - Consider 10^5 possible combinations of parameters
    - Initially cover range of six orders of magnitude per parameter

Workflow

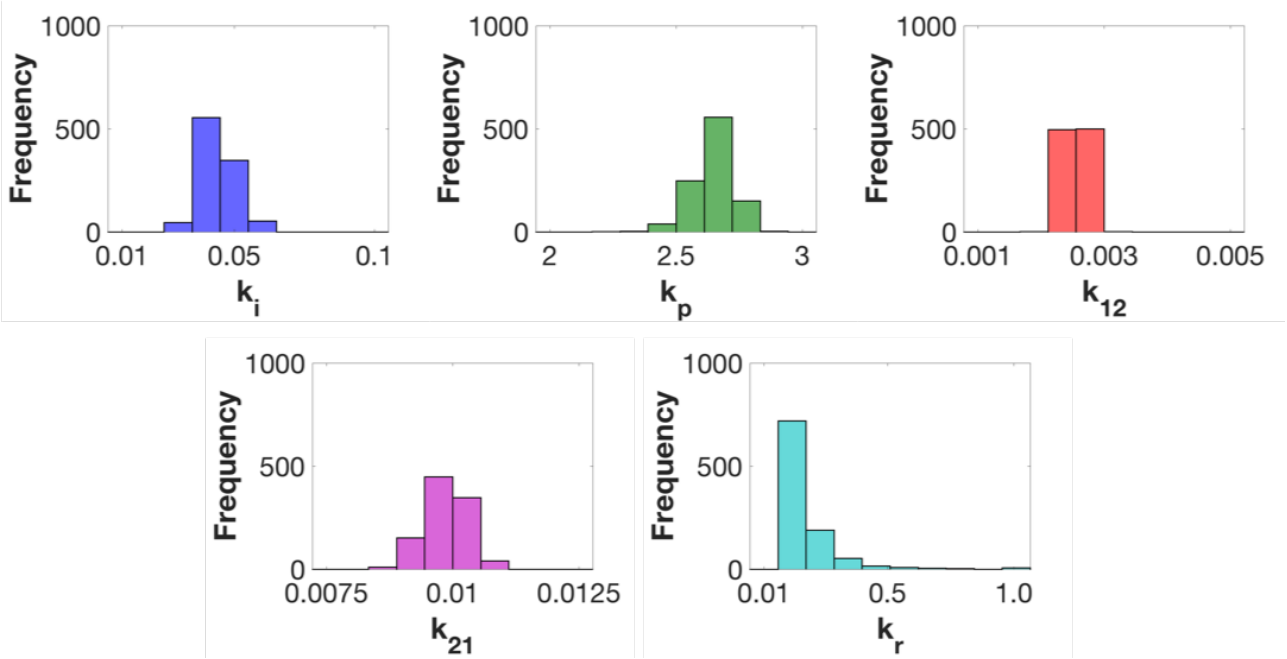

Generate list of rate constants
Run Simulations
Combine output
Calculate error functions
Look at distribution of rate constants
2000 jobs, each with 50 simulations, ~ 12 hours total
20 jobs, ~ 2 hours total
Repeated 5x

HTCondor Submit File

- Submitting multiple jobs:
  - Use process number as input into MATLAB function
- Running MATLAB Calculations:
  - Use MATLAB runtime with compiled functions
- Expanding beyond CHTC:
  - Use UW Grid and Open Science Grid
- Other options:
  - “max_retries = 3” to automatically re-run failed jobs

Modeling Results

Rate constant	Median	95% Confidence Interval	Literature value[1]
k_i	0.04	0.03-0.06	0.029 ± 0.008
k_p	2.6667	2.444-2.7778	2.9 ± 0.3
k_{12}	0.0028	0.0023-0.0028	0.0027 ± 0.0006
k_{21}	0.0097	0.0092-0.0108	0.010 ± 0.0001
k_r	0.1211	0.1211-0.4544	0.11 ± 0.07



1. Nelsen, D. L.; Anding, B. J.; Sawicki, J. L.; Christianson, M. D.; Arriola, D. J.; Landis, C. R. *ACS Catal*. **2016**, *6*, 7398.

Evaluating Computational Efficiency

Power of parallelization:

- 100,000 simulations estimated to take 2-3 months without parallelization
  - Corresponds to one iteration of grid searching process

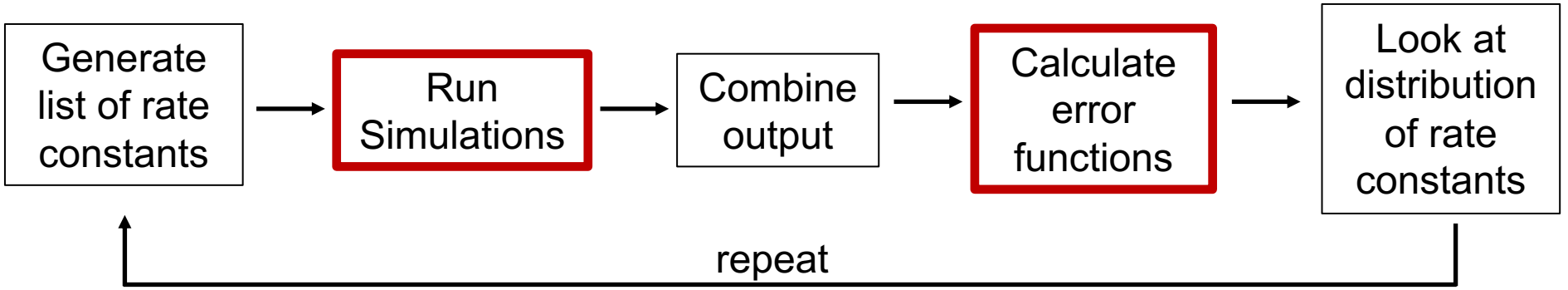
Comparison to optimization:

- ~4 days for 6 iterations of grid searching using HTCondor

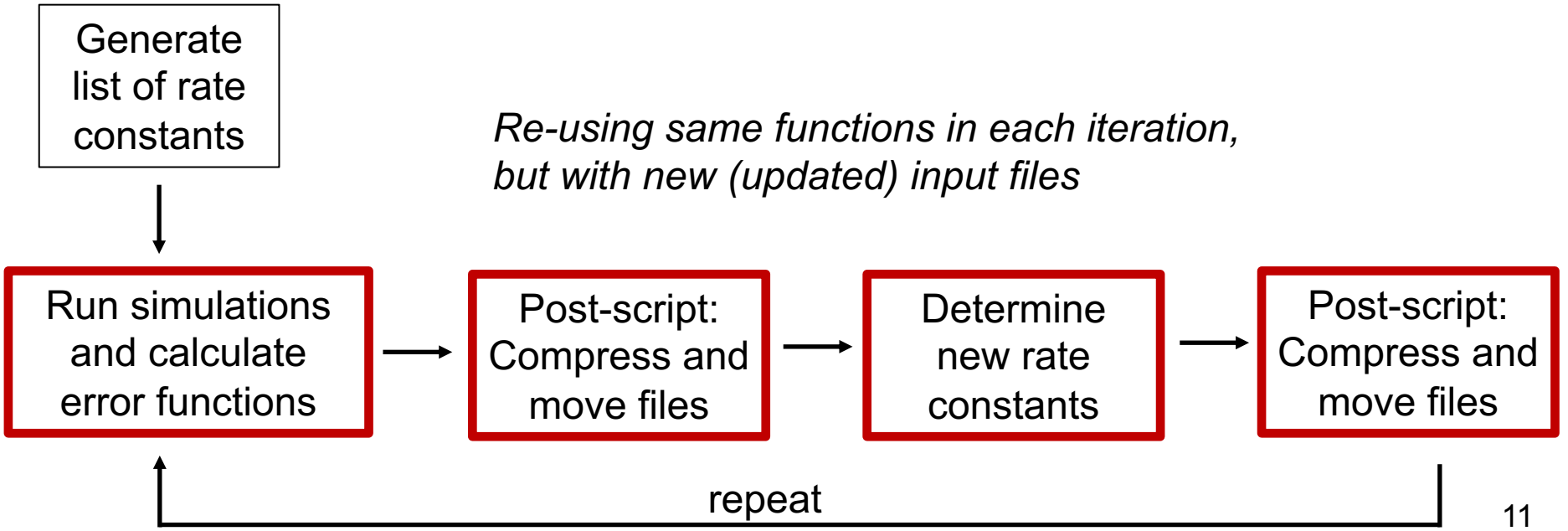
- Ran 10 iterative optimizations with random starting points
  - Using Copasi (kinetic modeling software)
  - Running on department cluster
  - 8/10 finished in >8 days (fastest in 2.5 days)

Improving the Method

Before:



Now:



Re-using same functions in each iteration, but with new (updated) input files

Challenges

- Jobs fail or get “stuck”
  - “max_retries = 3” to automatically re-run failed jobs
  - “condor_ssh_to_job” to check on job
  - “condor_hold” and “condor_release” to restart jobs when necessary

- Ideally automate using DAGMan
  - How do you address “stuck” jobs with DAGMan?
  - Avoid manually checking jobs
  - Slow jobs may be “stuck”, but could also involve stiff system of ODEs

Conclusions and Future Directions

The grid searching method is:

- More computationally efficient than iterative optimization
- More thorough in searching the parameter space than iterative optimization
- Capable of accurately reproducing results from iterative optimization
- Promising as a future kinetic modeling method

The next steps are to:

- Automate the modeling process using DAGMan
- Apply the modeling method to more complex polymerization systems

Acknowledgements

- Professor Clark Landis

- Landis Group Members
  - Eric Cueny
  - Andrew Maza
  - Dr. Tanner McDaniel
  - Megan Nieszala
  - Dr. Nick Staudaher

- CHTC
  - Christina Koch
  - Lauren Michael

Open Science Grid

HTCondor
High Throughput Computing