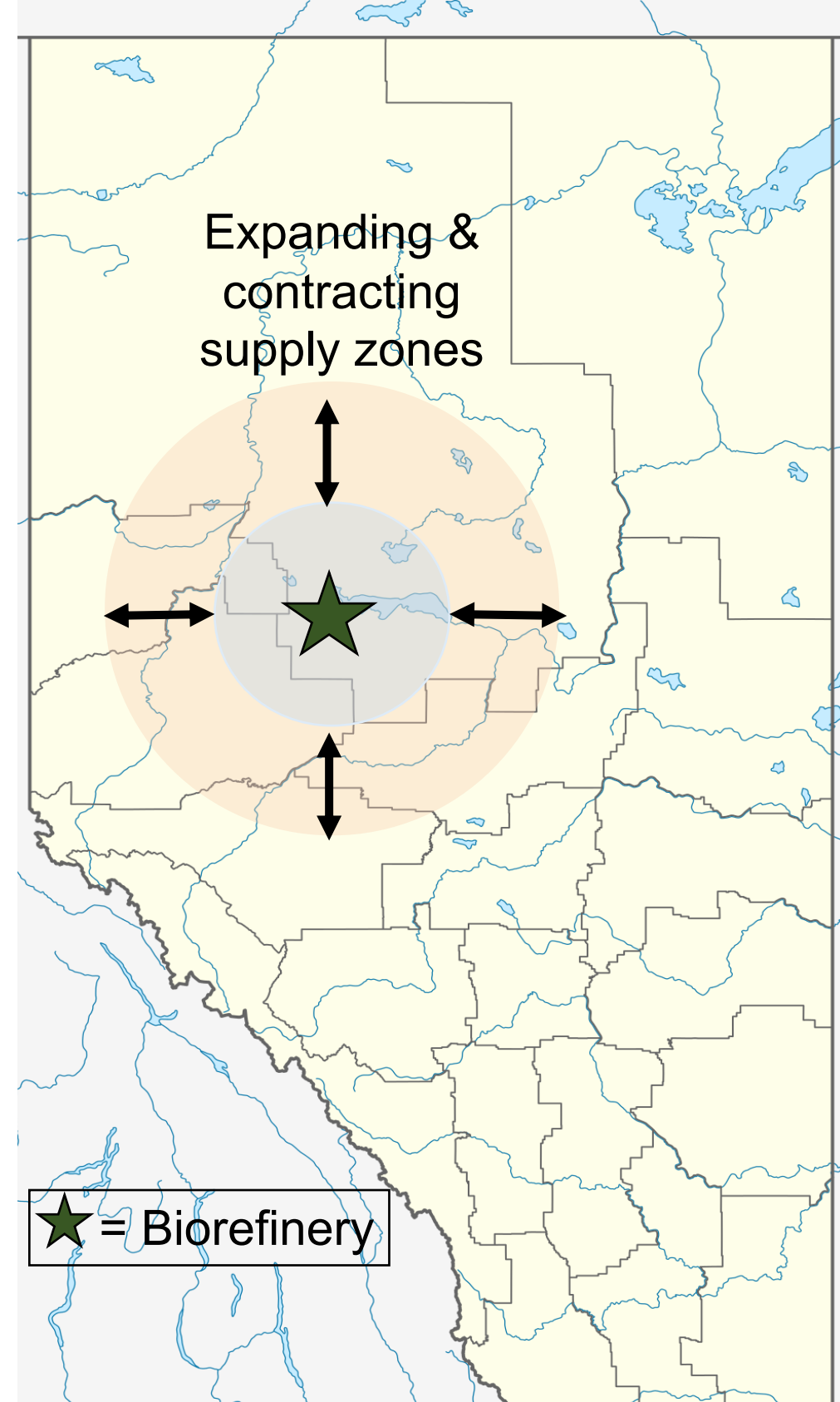


Modeling Variability in Biomass Feedstock Supplies with Limited Data: An Application of Data Clustering

Amy Xu, Grant Hauer, Marty Luckert¹, Feng Qiu²

Large scale production of bioenergy is impacted by the variability of biomass supply

- Emerging biotechnologies need to consider the COSTS and RISKS of obtaining and managing variable sources of bio feedstock supply.
- Crop yields and residues for bio-feedstocks will be variable over time.



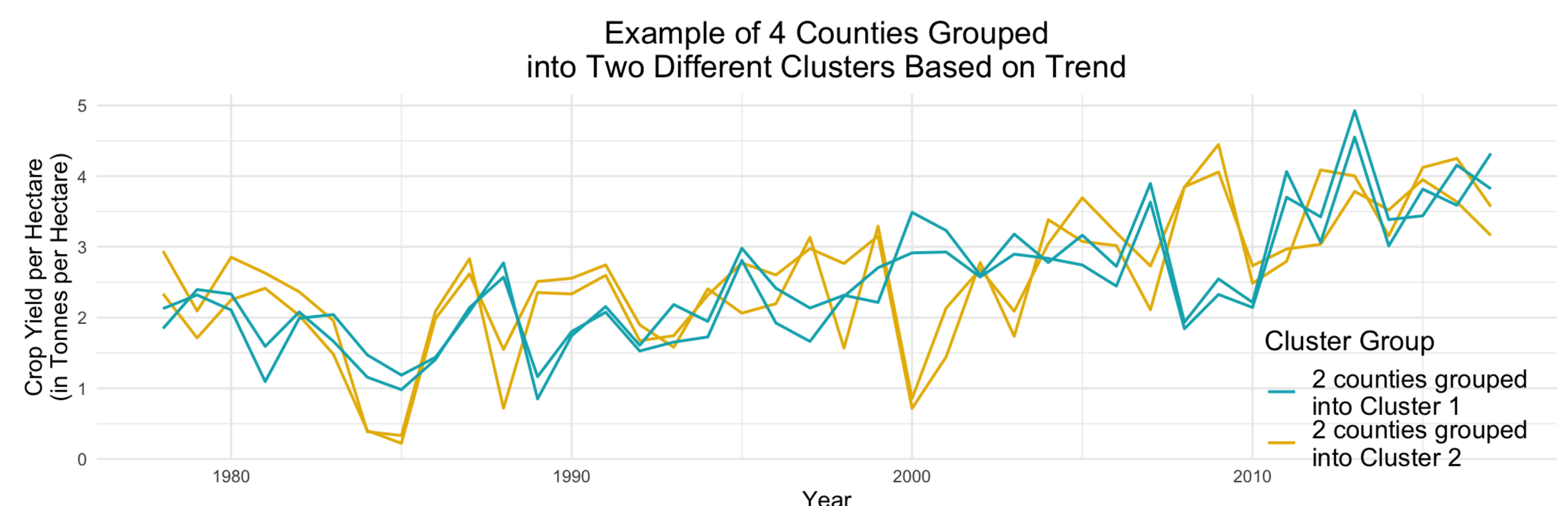
- Supplying a biorefinery with a fixed capacity, in the context of variable feedstock supply, requires an expanding and contracting supply zone over time.
- Accurately modeling crop variability is important.

Accurate modeling of variability is impacted by limited data

- We analyze crop yield variability across the 69 counties in Alberta over 40 years.
- To model variability over location and time across counties, the typical approach is to estimate the sample correlation using traditional statistical approaches.
- However, 40 years of data makes it difficult to consistently and accurately estimate correlations among 69 counties.
- To overcome the problem of limited data, we use data clustering algorithms.

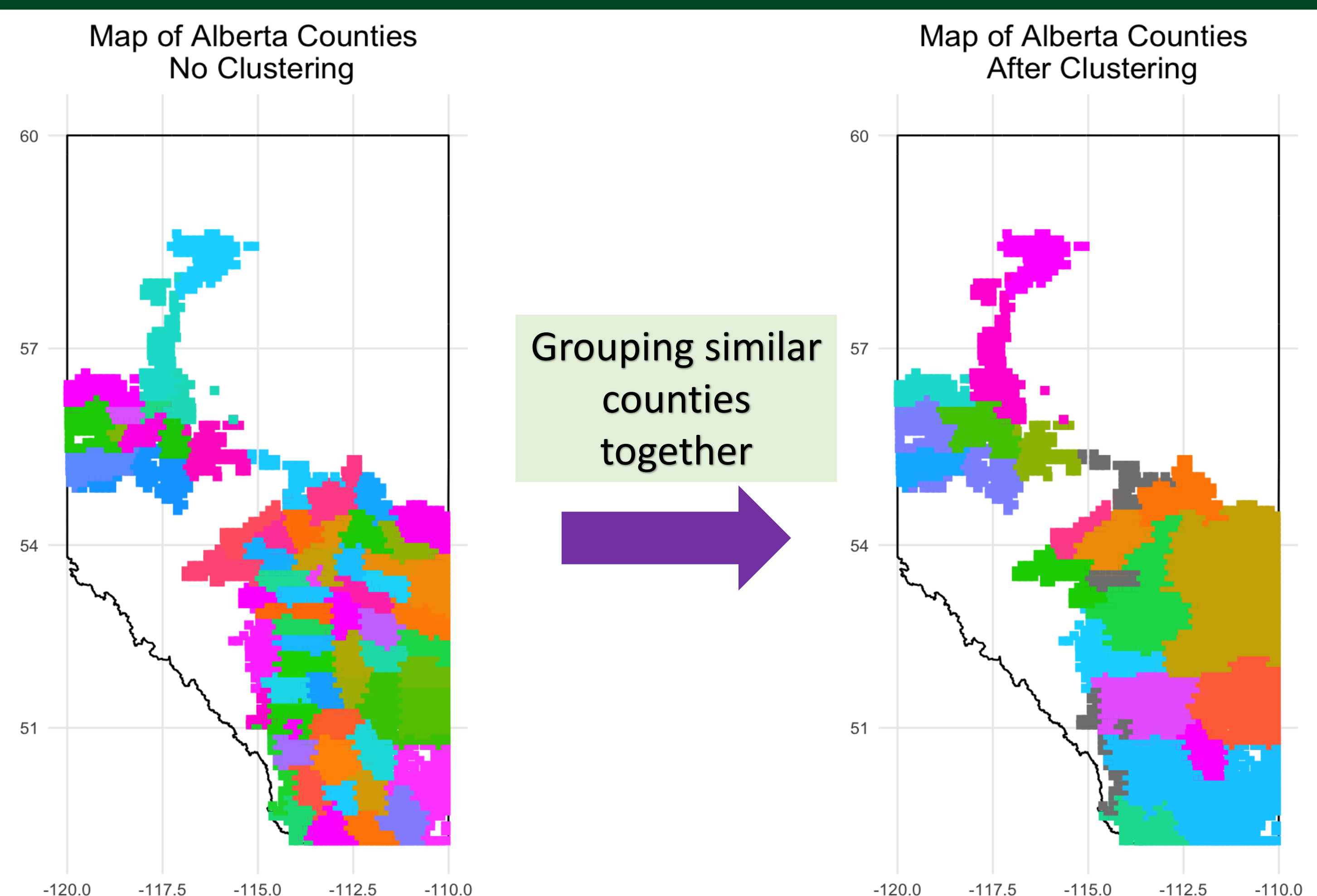
Data clustering groups similar counties

- We group together counties that have similar trends in crop yields over time as measured by hierarchical clustering approaches (see example below).



Grouping similar counties allows for a more consistent estimate of variability

- If we do not apply clustering, variability amongst counties cannot be accurately identified.
- By clustering similar counties, differences amongst dissimilar counties can be more accurately represented.
- We used two distance measurements for the hierarchical clustering:
 1. Average linkage
 2. Complete linkage
- The map on the far right represents the clustered counties obtained after hierarchical clustering with average linkage.
- After clustering, we re-estimated the variability.

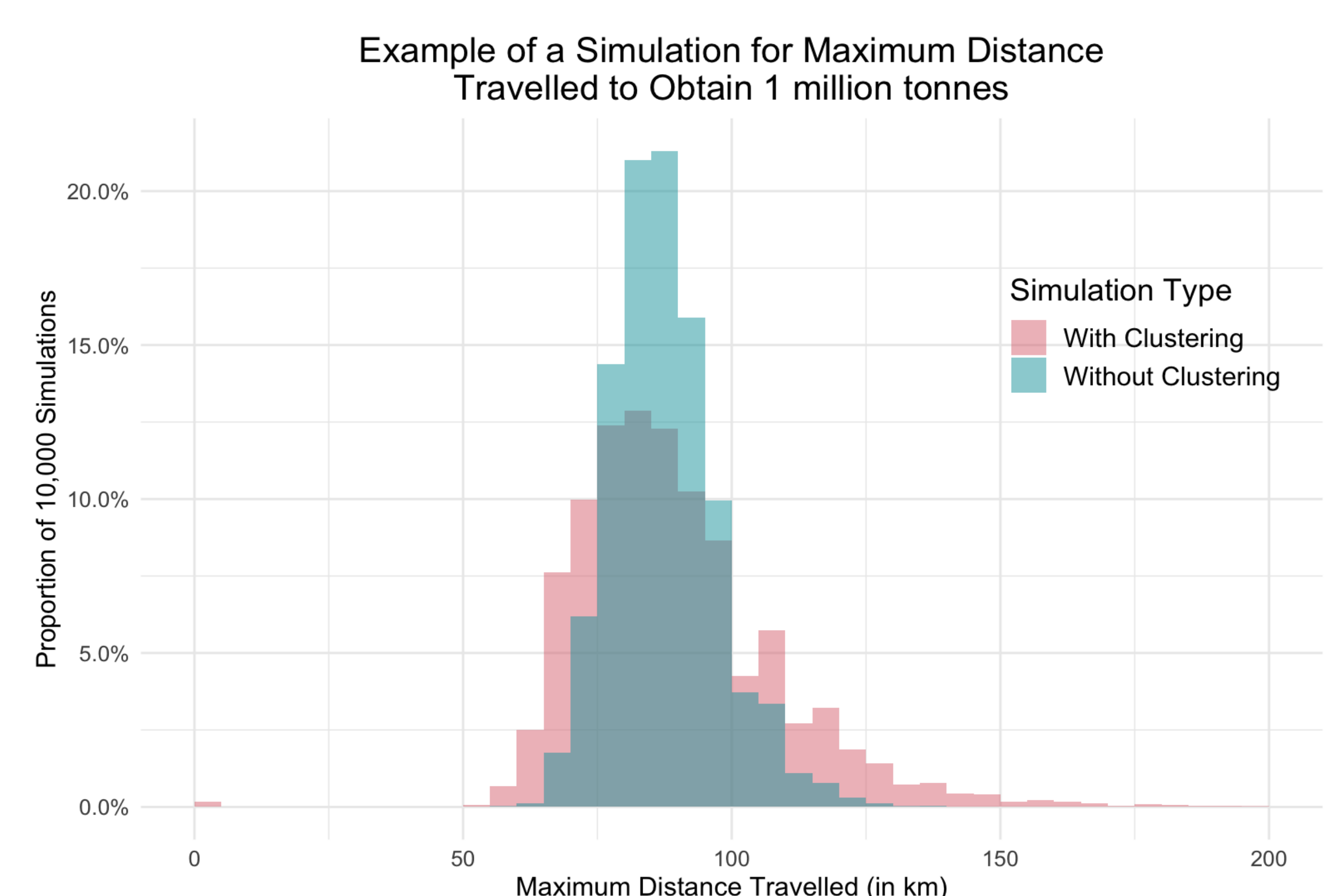


Comparing estimates of supply variability with and without clustering

- We simulate supply for various possible plant locations with a fixed plant capacity of 1 million tonnes per year. Each of the 10,000 simulations resulted in a supply zone with a maximum distance travelled to satisfy the plant capacity.
- An example simulation is shown (see Figure). The distribution curves show that there is higher variability in supply distances after clustering (i.e. a bioenergy plant situated in one location may have to go out further to get the same supply quantity).
- KEY POINT:** Estimating supply variability without clustering could underestimate the variability of biomass availability.

FUTURE WORK

- Investigate different ways of incorporating variation within clusters.
- Expand the model to investigate optimal biorefinery location for multiple plant locations.



¹ Principal Investigator. Email: mluckert@ualberta.ca

² Department of Resource Economics and Environmental Sociology, 515 General Services Building, University of Alberta, Edmonton, AB, T6G 2H1