



AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search

<http://ageconsearch.umn.edu>

aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

How to Estimate Crop Yield Densities of Counties with Missing Data

Eunchun Park, Ardian Harri, and Keith Coble



MISSISSIPPI STATE
UNIVERSITY™

Department of Agricultural Economics

NASS county-level corn yield data

Minnesota (57 out of 87 counties from 1955 to 2017)

Missouri (48 out of 114 counties from 1955 to 2017)

Maryland (13 out of 23 counties from 1955 to 2017)

Colorado (2 out of 64 counties from 1963 to 2017)

Montana (1 out of 56 counties from 1963 to 2017)

Data Omission

Quantity

- Level of data omission

Quality

- Recent data omission

Question

How to use a **spatial dependence** of crop yields to **estimate** and **reproduce** the **crop yield distributions** of counties with **missing values** or **no reports at all?**

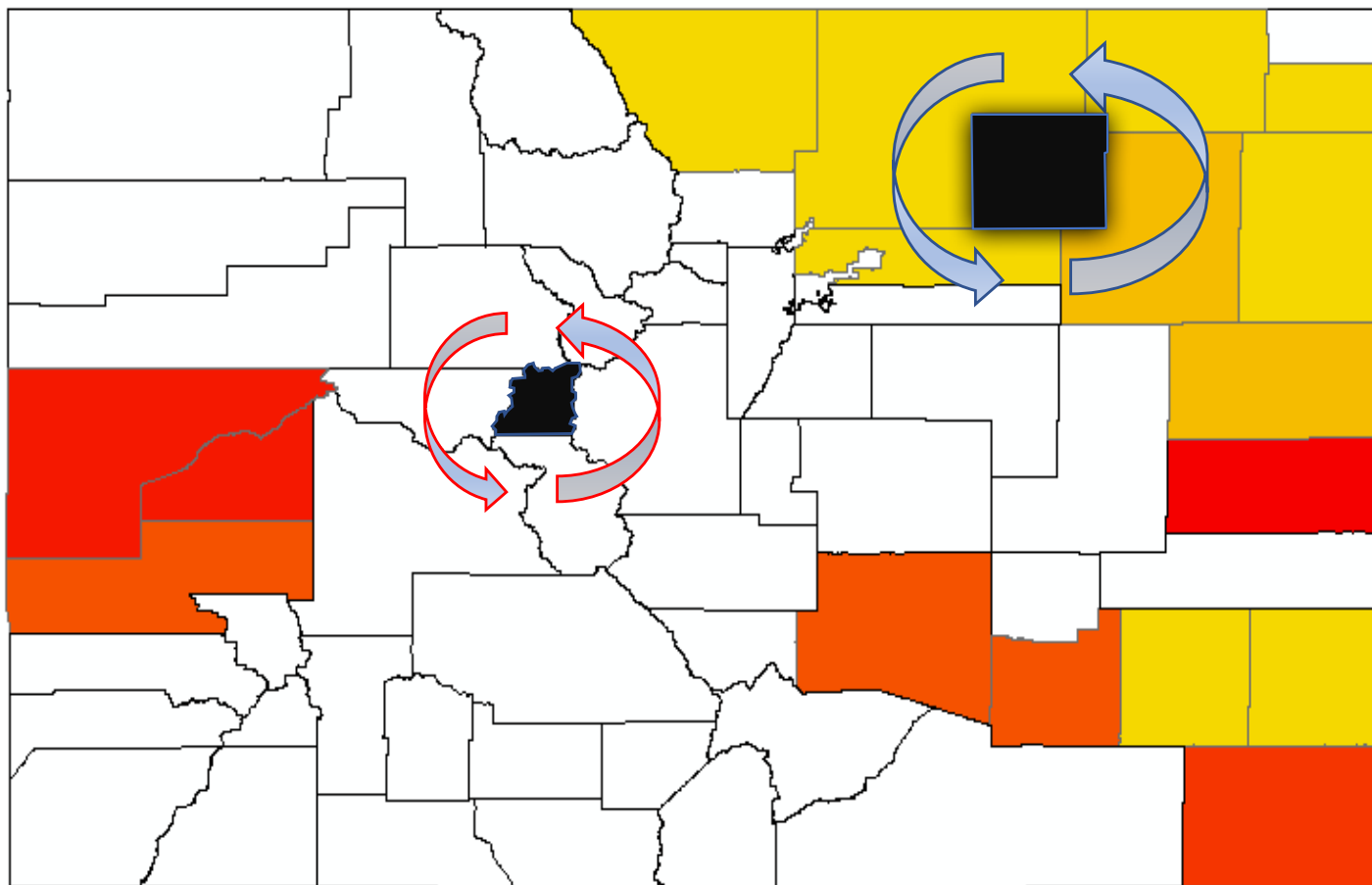
Possible options

1. Using neighboring counties yields

2. Bayesian Model Averaging (**Benchmark**)

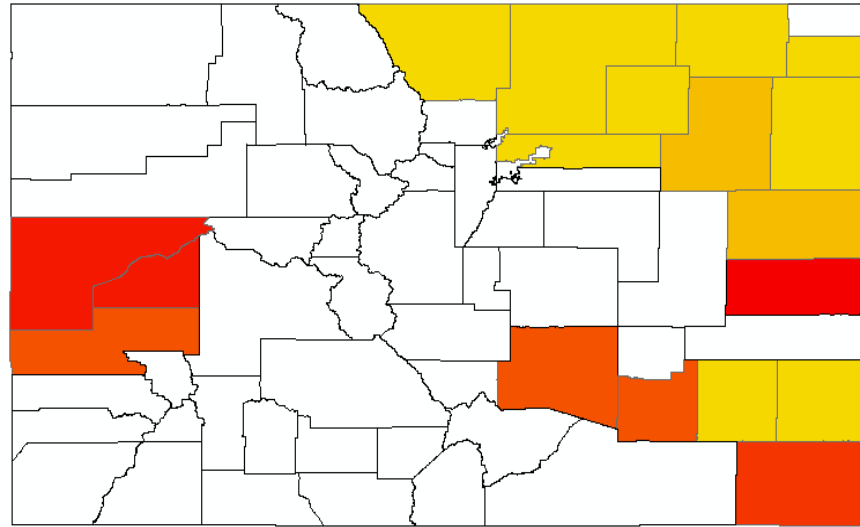
3. Bayesian Spatial Interpolation (**Proposed**)

Neighboring counties yields



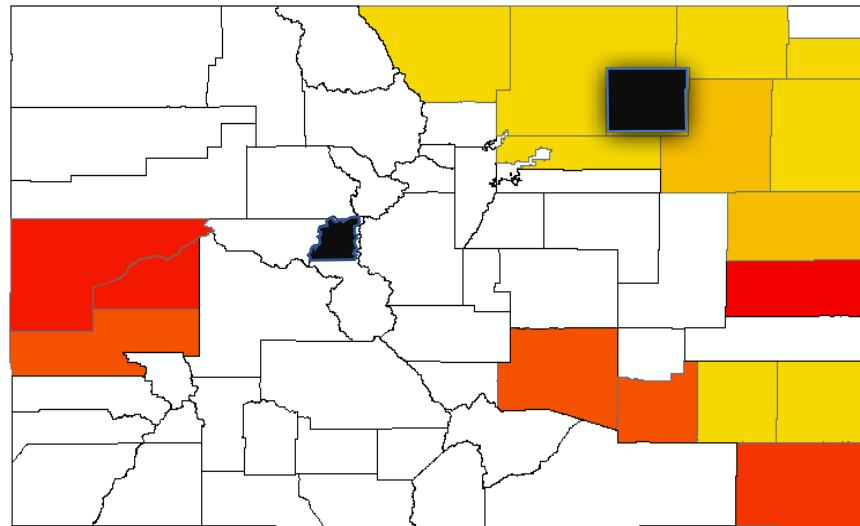
Bayesian Model Averaging (BMA)

Estimate each county's density (ϕ_i) independently by using its own yield records



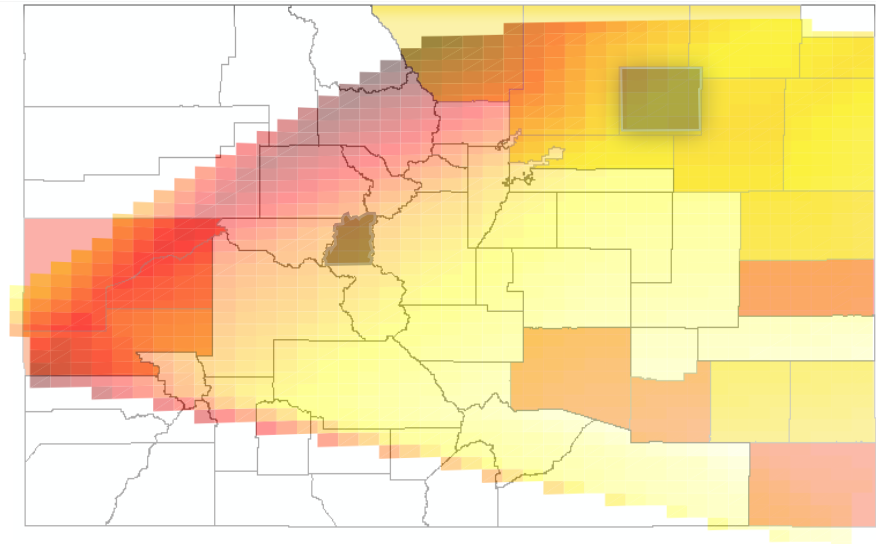
Final density (f_i) is an **weighted average** of **independently** estimated

density ϕ_j , $f_i = \sum_{j=1}^J w_j^i \phi_j$



Bayesian Interpolation Model

Regard county level densities
as a **variation from an
integrated yield density
structure across space**



Bayesian Spatiotemporal Models

- **Parameters** are stochastic and are spatially correlated by **distance** (Gaussian spatial process)
- Provides **site (county) specific density** estimates
- **Strong efficiency gain** even when # of obs is small
(by updating spatial structures)
- Posterior predictive distribution (**time and space**)
 - Unbalanced panel data
 - Recover missing observations

Bayesian Spatiotemporal Models

- y_{it} be the crop yield of county i at year t , and be $y_{it} = \mu_{it} + \varepsilon_{it}$,
- Define a quantile function $Q_{it}(\pi)$ of county i at time t that satisfies the condition $P\{\mu_{it} < Q_{it}\} = \pi \in [0,1]$

$$Q_{it}(\pi) = \sum_{l=1}^L \{\beta_{0i} + \beta_{1i}t + [\gamma_{0il} + \gamma_{1il}t]B_l(\pi)\},$$

$$P(y_{it}) = \sum_{l=1}^L I\{Q_{it}(h_l) \leq y_{it} < Q_{it}(h_{l+1})\}N(\beta_{0i} + \beta_{1i}t, (\gamma_{0il} + \gamma_{1il}t)^2),$$

where $B_l(\pi)$ is l th basis (quantile) function.

Bayesian Spatiotemporal Models

$$P(y_{it}) = \sum_{l=1}^L I\{Q_{it}(h_l) \leq y_{it} < Q_{it}(h_{l+1})\} N(\beta_{0i} + \beta_{1i}t, (\gamma_{0il} + \gamma_{1il}t)^2),$$

$$\text{cov}(\beta_{0i}, \beta_{0j}) = \Sigma_{\beta_0} = \rho_{\beta_0} e^{-D_{ij}/\theta_{\beta_0}}$$

$$\text{cov}(\beta_{1i}, \beta_{1j}) = \Sigma_{\beta_1} = \rho_{\beta_1} e^{-D_{ij}/\theta_{\beta_1}}$$

$$\text{cov}(\gamma_{0il}, \gamma_{0jl}) = \Sigma_{\gamma_{0l}} = \rho_{\gamma_{0l}} e^{-D_{ij}/\theta_{\gamma_{0l}}}$$

$$\text{cov}(\gamma_{1il}, \gamma_{1jl}) = \Sigma_{\gamma_{1l}} = \rho_{\gamma_{1l}} e^{-D_{ij}/\theta_{\gamma_{1l}}}$$

Empirical Application

Empirical application

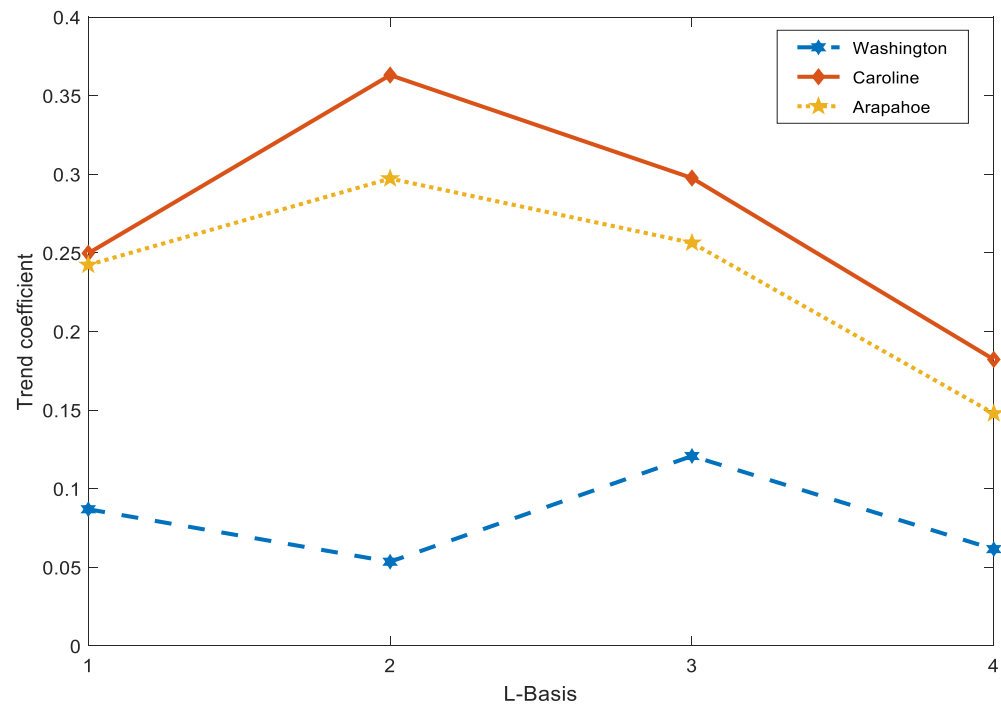
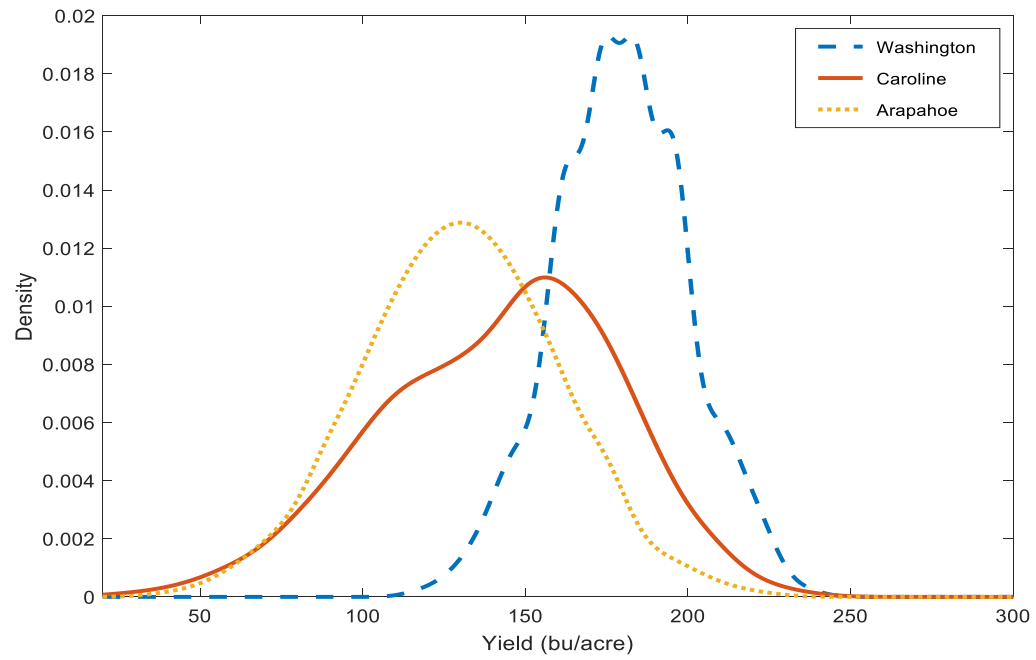
- Iowa
 - **99** counties (1955-2017)
- Maryland
 - **23** counties (1955-2017) : **Small number of counties**
- Colorado
 - **53** counties (1963-2017)

Empirical application

- Iowa
 - **Homogeneous geography**
- Maryland
 - **Heterogeneous geography**
- Colorado
 - **Heterogeneous geography**

Empirical application

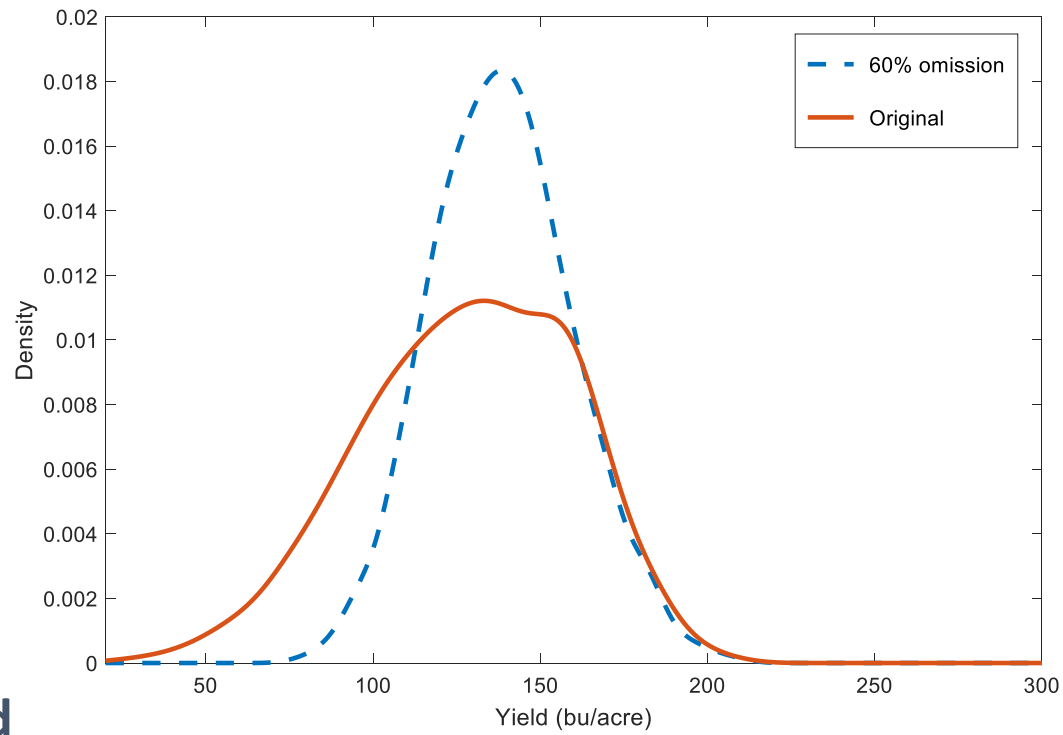
- Iowa
 - **No significant data omission (0.1%)**
- Maryland
 - **No significant data omission (3%)**
- Colorado
 - **Significant data omission (47%)**



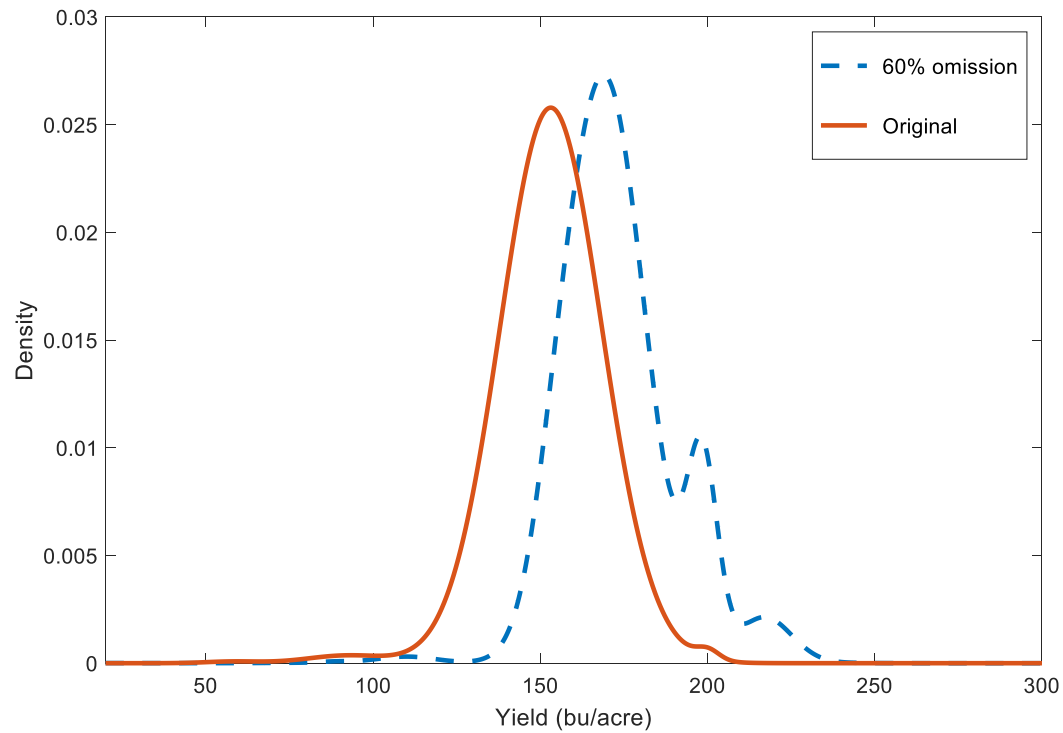
Loss ratio

20,000 posteriors and burn in first 10,000 obs

State / Data omission		60% omission		Original	
Model		BMA	Proposed	BMA	Proposed
Iowa	Mean	4.43	1.42	1.40	1.24
	RMSE	7.02	2.99	2.08	2.15
	Max	21.02	4.49	12.08	3.04
	Min	0.23	0.00	0.23	0.00
Maryland	Mean	4.64	2.18	1.47	0.75
	RMSE	5.99	3.13	1.72	1.12
	Max	57.84	4.20	2.85	1.69
	Min	0.74	0.54	0.54	0.15
Colorado	Mean	-	-	7.23	4.17
	RMSE	-	-	7.80	4.64
	Max	-	-	26.50	9.38
	Min	-	-	0.00	0.00



Allegany, Maryland



Out of sample premium game

- ***Between*** game : BMA vs Proposed model
- ***Within*** game : with vs without data omission

Between game (BMA vs Proposed)

State	Number of Counties	Dataset	<i>p</i> -value
Iowa	99	Original	0.593
		60% omission	0.084
Maryland	23	Original	0.000
		60% omission	0.000
Colorado	42	Original	0.006

Within game (original vs 60% omitted)

State	Model	<i>p</i> -value
Iowa	BMA	0.132
	Proposed	0.313
Maryland	BMA	0.021
	Proposed	0.191

Findings and Discussions

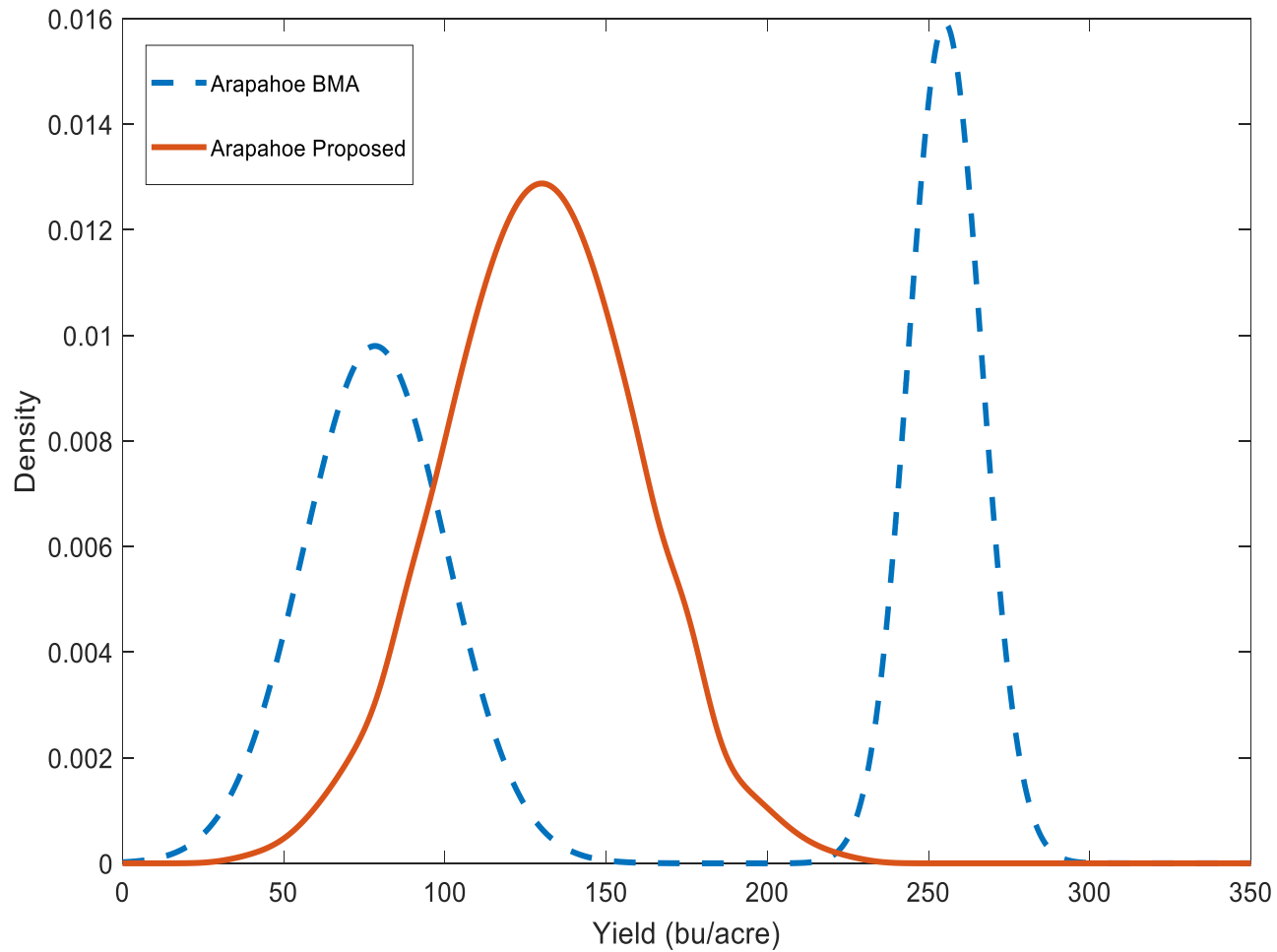
- The data omission problem, in general, **underrates insurance premiums** and results in **larger loss ratios**
- The estimation results remain consistently favorable to the proposed model compared to the BMA model in most of the states with/without data omission, except Iowa with the original dataset
- The proposed model is **superior** to the BMA model when there is a **considerable level** of data omission
- The proposed model tends to be less sensitive to the **quantity** (level) and **quality** (recent data) of the data omission

Findings and Discussions

- **Identify changes** of multiple quantile levels of the crop yield distributions over **space** and **time**
 - Time varying skewness
- Apply the model for **identifying changes** in crop yield distribution over **time** and **space** in response to the changes of other principal factors, such as **climate**, **elevation**, and **soil types**
- Computational complexity

Thank you

Arapahoe, Colorado



- Total 35 years of yield data, but has no data since 2005