Combining spatio-temporal weather and crop data for network-based inference on the international wheat trade

Srishti Vishwakarma, Xin Zhang, Vyacheslav Lyubchich
University of Maryland Center for Environmental Science
{srishti.vishwakarma, xin.zhang, lyubchich}@umces.edu

University of Maryland CENTER FOR ENVIRONMENTAL SCIENCE

supported by
NSF CNS-1739823
NSF IIS-1633355

I. Introduction

- International trade helps with redistributing food across country borders from surplus to deficit regions.
- With respect to food security, it is important to understand factors that influence trade partnerships between countries.
- However, due to the exceptional complexity in multi-source data, only few studies have explored the impact of extreme weather on trade.
- Here, we created a spatially-explicit dataset by combining weather, international crop trade and crop-specific data.
- The study focuses on quantifying the network effects of two novel factors, namely, the extreme weather stress and synchrony of crop yield, on international wheat trade network.

II. Material and methods

1. Covariate development

Table 1: Summary of data used

Data type	Spatial resolution	Temporal resolution
Weather: Total precipitation (m), Avg temperature (K), Max temperature (K), Min temperature (K)	0.75°	2005–2014 (daily)
Crop yield (kg N ha^{-1})	99 countries	2005–2014 (annual)
Bilateral wheat trade (kg and presence/ absence)	99 countries	2005–2014 (annual)
Population-weighted distance	99 countries	-
Contiguity	99 countries	-
Common official language	99 countries	-

- **Yield short-term synchrony**

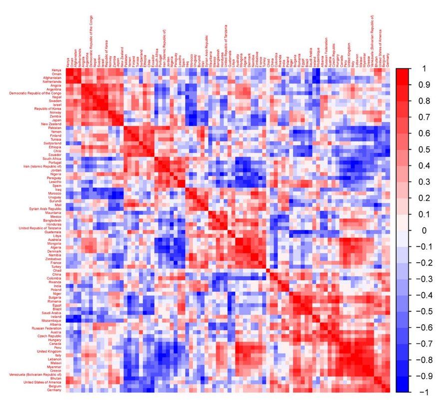

Fig. 1: Pearson correlation of short-term fluctuations (STS) in yield, 2005–2014

- **Stress of extreme weather**

❑ We developed weather indices (night growing degree low, day growing degree low/high, etc.; Zhu et al. 2015) by matching spatial grids of daily weather data (Table 1) with wheat harvested area grid (Monfreda et al. 2008) and gridded crop calendar.
❑ Aligned the direction of weather index with weather stress level.
❑ Selected principal components (PCs) explaining at least 80% of variation in weather indices, 2005–2014 (Fig. 2).
❑ Within each PC, we selected contributing variables using scree plot of contribution in each dimension.

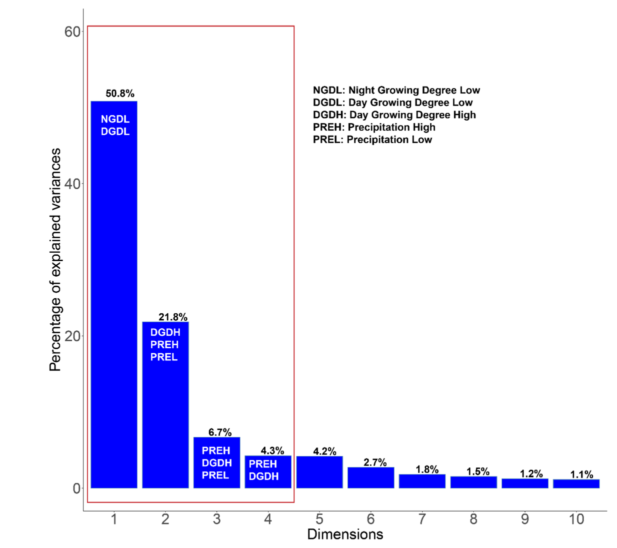

Fig. 2: Scree plot of PCA results. Each bar shows the percentage of explained variance by each dimension and selected variables with the highest contribution in that dimension. The red box shows the selected dimensions, explaining at least 80% of the data variability.

2. Network modeling and testing

- Exponential random graph model (ERGM): Model the distribution of binary adjacency matrix and volume of trade as a function of network statistics and covariates (Hunter et al. 2008).
- Random forest (RF): RF is the machine learning algorithm to fit the relationships by aggregating outputs from multiple data-driven regression trees (Breiman, 2001).
- Cross-validation of ERGM and RF using 100 simulations of random splitting the data into training (5 years) and testing (5 years) sets.

III. Modeling results

Table 2: A summary of results from the tested models; ERGM: Exponential random graph model, RF: Random forest, PCA: Principal component, STS: Short-term synchrony

Response	Model	Significant covariates	Coefficient	Error
Trade (unweighted-directed)	ERGM	PCA (dim 2, 4)	(-1.13E-03, -0.107)	5.80%
		STS	0.154	
		Distance	-1.21E-05	
		Contiguity	2.19	
		Common official language	0.175	
	RF	PCA (dim 1, 2, 3, 4)	Fig. 3	23.40%
		Distance		
		Contiguity		
		Common official language		
Trade (weighted-directed)	ERGM	PCA (dim 1, 2, 3, 4)	(3.06E-04, -5.95E-04, -9.86E-04, 9.90E-04)	1.68
		STS	5.67E-02	
		Distance	-2.64E-05	
		Contiguity	0.461	
		Common official language	-3.29E-02	
	RF	PCA (dim 1, 2, 3, 4)	Fig. 4	0.97
		STS		
		Distance		
		Contiguity		
		Common official language		

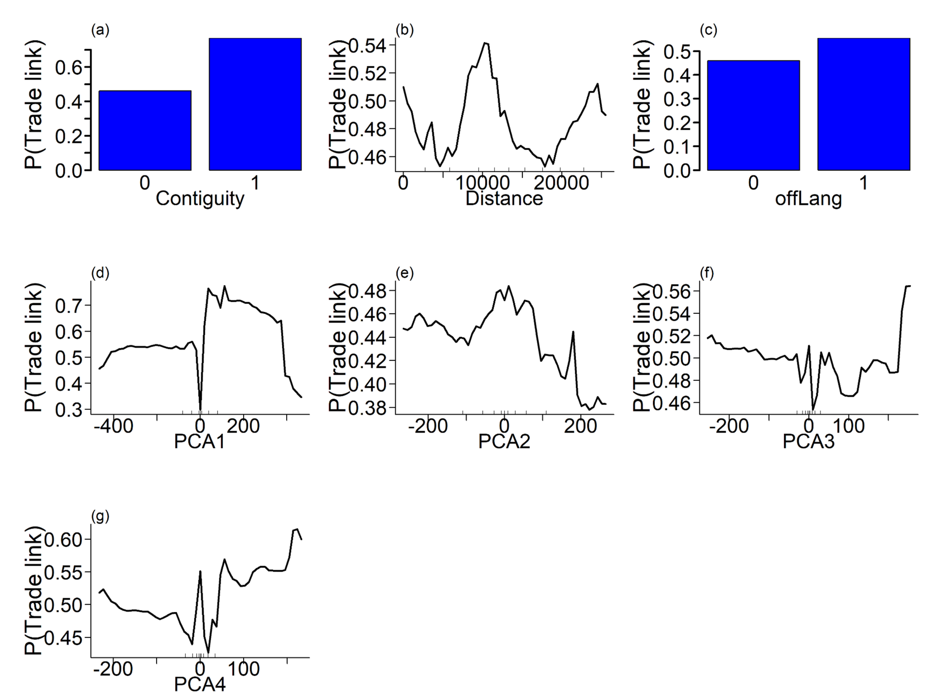

Fig. 3: Partial dependence plots obtained from RF with unweighted trade as the response variable. The x-axes represent the considered covariates, where PCAs are the principal components with numbers 1-4 showing the chosen dimensions, and offLang represents the common official language. The y-axis represents the marginal change in the likelihood of the trade link formation due to changes in the covariates on x-axis

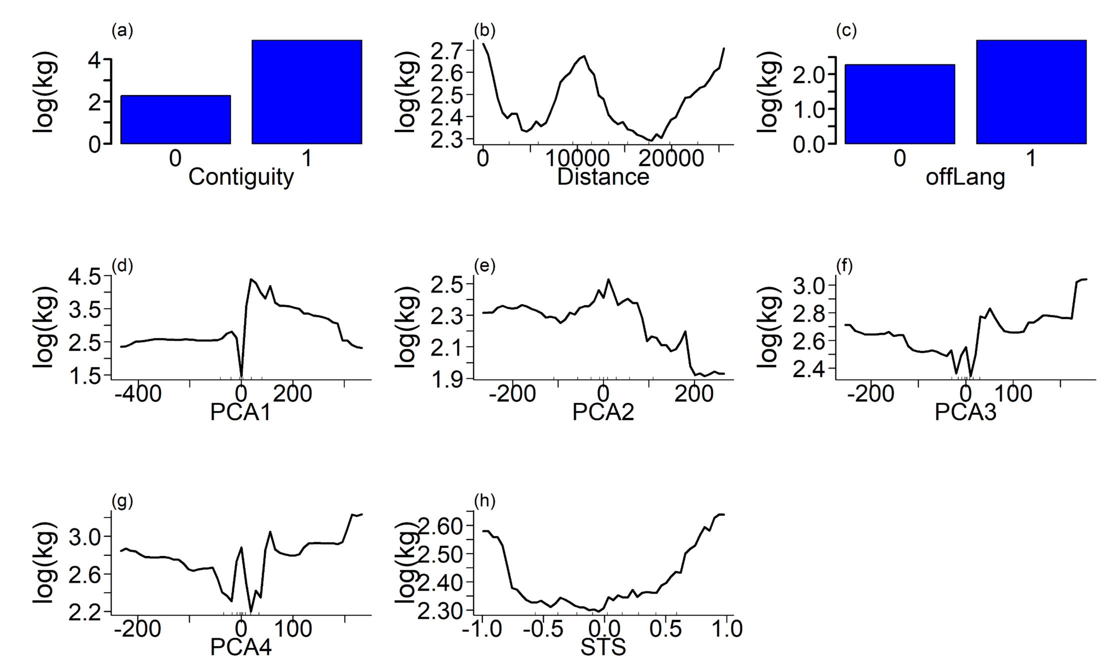

Fig. 4: Partial dependence plots obtained from RF with trade volume as a response variable. The x-axes represent the covariates, where PCAs are the principal components with numbers 1-4 showing the dimensions, and offLang represents the common official language. The y-axis represents the marginal change in the trade volume in log10(kg) due to changes in the covariates on x-axis.

- Countries with higher extreme weather stress (i.e., heat stress) tend to import from countries with less stress (Table 2).
- Countries with more synchronized crop yield tend to have higher chances of trade and higher import volume (Table 2, and Figures 3-4).
- Trade connections are more frequent and trade volumes are higher between countries which are contiguous, closer in distance and use the same official language (Table 2, and Figures 3-4).
- For the unweighted-directed trade network, ERGM performed better than RF (Table 2). In contrast, RF performed better than ERGM for weighted-directed trade network.

IV. Cross-validation

- Misclassification error (%) $= 100 \times \left(1 - n^{-1}(\mathrm{TN} + \mathrm{TP})\right)$ (1)
where TP and TN are true positive and true negative.

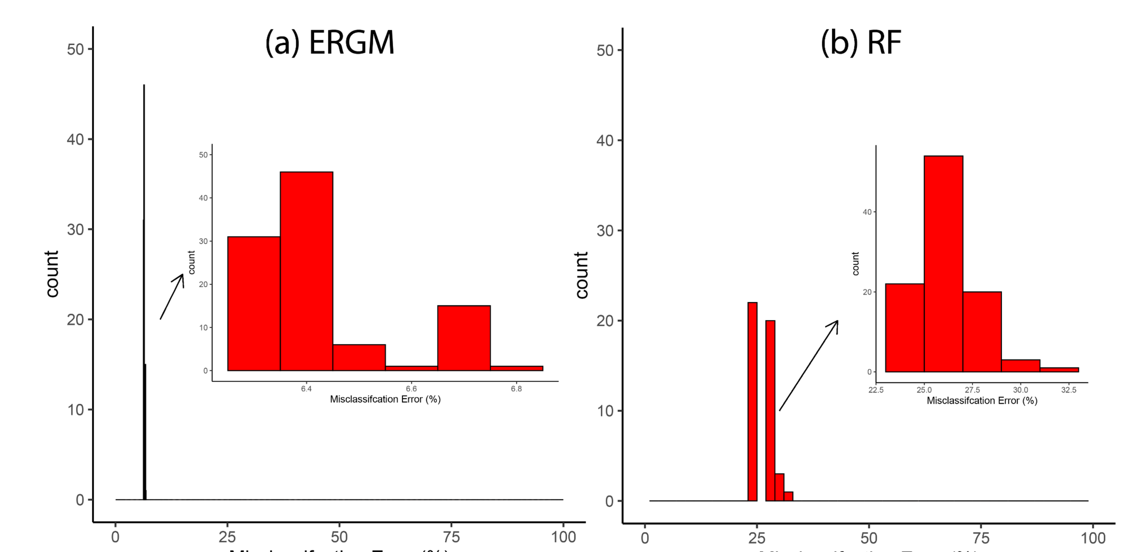

Fig. 5: Histogram of misclassification error (Eq. 1) obtained from (a) ERGM and (b) RF during cross-validation simulations. The cross-validation was comprised of 100 simulations. The arrow points towards a zoomed in version of the histogram.

- Mixed error $= \frac{|\mathrm{OT}-\mathrm{PT}|}{1+\mathrm{OT}}$ (2)
where OT is observed trade volume, and PT is the predicted trade volume.

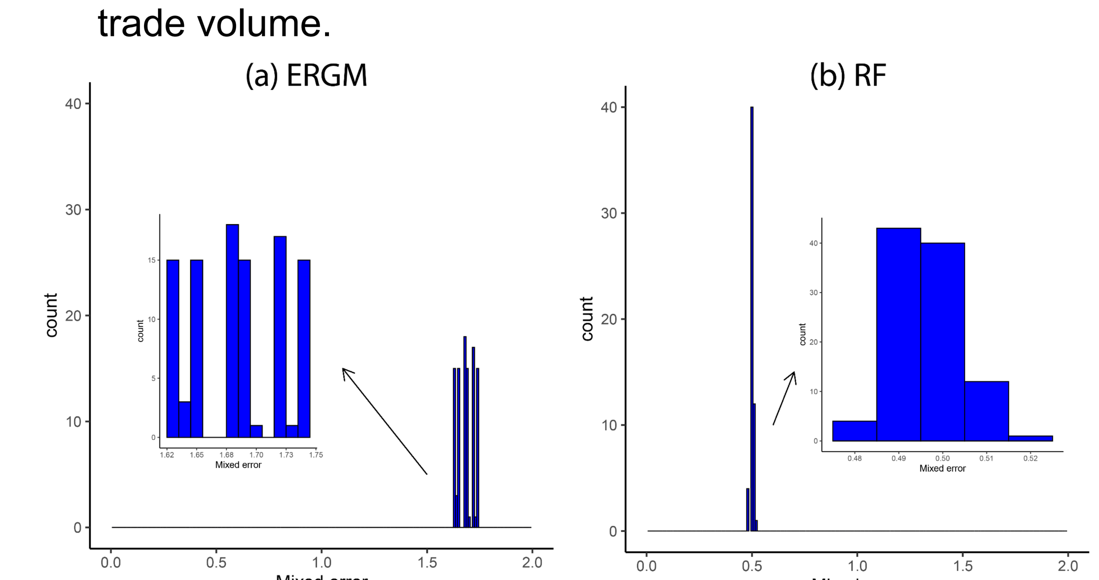

Fig. 6: Histogram of mixed error (Eq. 2) obtained from (a) ERGM and (b) RF during cross-validation simulations. The cross-validation was comprised of 100 simulations. The arrow points towards a zoomed in version of the histogram.

V. Conclusion

- Two novel factors (i.e., extreme weather stress and synchrony of crop yield) are consistently significant in the models tested in this study for trade connections and trade volume.
- More synchronized yield variations between countries are associated with higher likelihood of trade partnership. This represents a systemic risk to the wheat supply, since synchronized yield failure in trade partner countries will likely lead to wheat shortage, which disproportionally affects food security of low-income population and import dependent countries.
- ERGM is more reliable when modeling unweighted-directed network, while RF performs better when the trade network is weighted-directed.
- Further work will be devoted to extending the analysis to other staple crops (e.g., rice, maize, and soybean).

References

Breiman, L. (2001). Random forests. Machine Learning, 45(1):5–32.
Hunter, D. R., M. Handcock, C. T. Butts, S. M. Goodreau, M. Morris, (2008). ergm: A Package to Fit, Simulate and Diagnose Exponential-Family Models for Networks. Journal of Statistical Software 24 3.
Monfreda, C., N. Ramankutty, and J. A. Foley, (2008). Farming the planet: 2. Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000. Global Biogeochem. Cycles, 22, GB1022
Zhu, W., Porth, L., Tan, K.S., 2015. A Credibility-Based Yield Forecasting Model for Crop Reinsurance Pricing and Weather Risk Management. *SSRN* **1**, 1–36 (2015).