Comparison of Hidden Markov Precipitation Models Using the Continuous Ranked Probability Score

Nick Vasko, Ben Lee, and Murali Haran
Department of Statistics, The Pennsylvania State University, and Network for Sustainable Climate Risk Management

Summary

- Goal: Compare rainfall models and study model selection techniques
- Challenges:
  - Precipitation data have a large proportion of zeros and extreme values
  - May exhibit autocorrelation
  - Building a statistical model with all the above features can be difficult
- Features of Our Best Models:
  - Seasonal Hidden Markov Model (HMM)
  - Gamma or Weibull distributions conditional on the day before
- We recommend Continuous Ranked Probability Score (CRPS) for model choice instead of more popular Bayesian Information Criteria (BIC)

Why is This of Interest

- Climate change will affect precipitation
- Some areas will experience more severe drought, others extreme rainfall
- Understanding precipitation patterns and trends may be useful for public policy and water resource allocation

Models and Data

- Precipitation data obtained from NOAA Climate Data Online database
- Focus here on Warren, PA and Phoenix, AZ
- Hidden Markov Model (HMM)
  - System is a Markov process with hidden unobserved states
  - States are not directly observed
- **Model Approach for Precipitation**
  - HMM has 2 states: 'Rain' / 'No Rain'
    - If 'Rain': Data follows Gamma or Weibull distribution
      - Distribution conditioned on 'Rain' / 'No Rain' day before

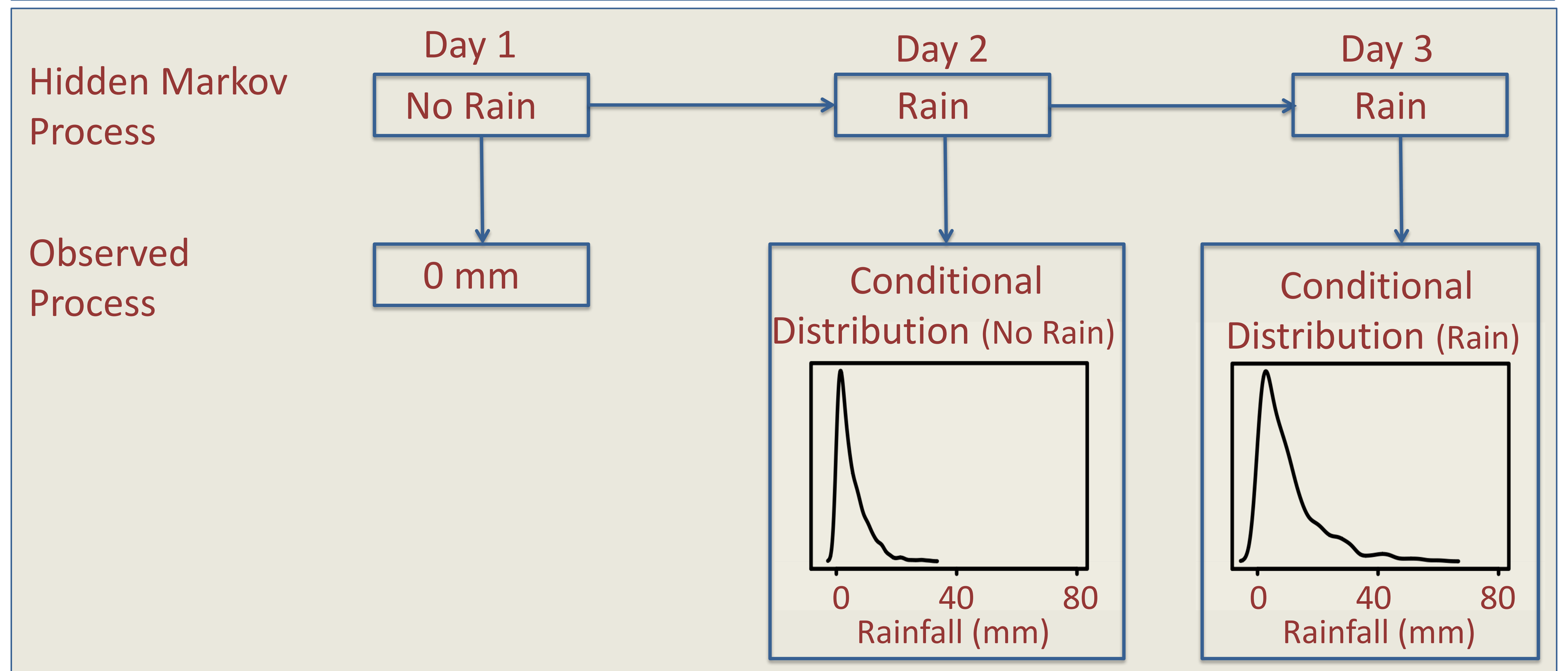

Figure 1. Illustration of Hidden Markov Model and conditional distributions
- Day 2: Smaller values due to 'No Rain' on Day 1
- Day 3: Larger values due to 'Rain' on Day 2

Model Choice

- Continuous Ranked Probability Score (CRPS)
  - Generalization of Absolute error
  - Penalizes bias and incorrect variance of predictive distributions
- Bayesian Information Criteria (BIC)
  - Rewards model fit based on likelihood
  - Penalizes for number of parameters

Results

- Warren Best Model:
  - Seasonal 2-stage (rain/no rain) Hidden Markov Model (HMM)
  - Gamma distribution for amount of rain, given greater than zero
- Phoenix Best Model:
  - Seasonal 2-stage (rain/no rain) HMM with Weibull distribution
  - Distributions conditioned on rain/no rain day before
- Model selection
  - Minimizing CRPS leads to more precise and parsimonious models
  - Minimizing BIC leads to more complicated models

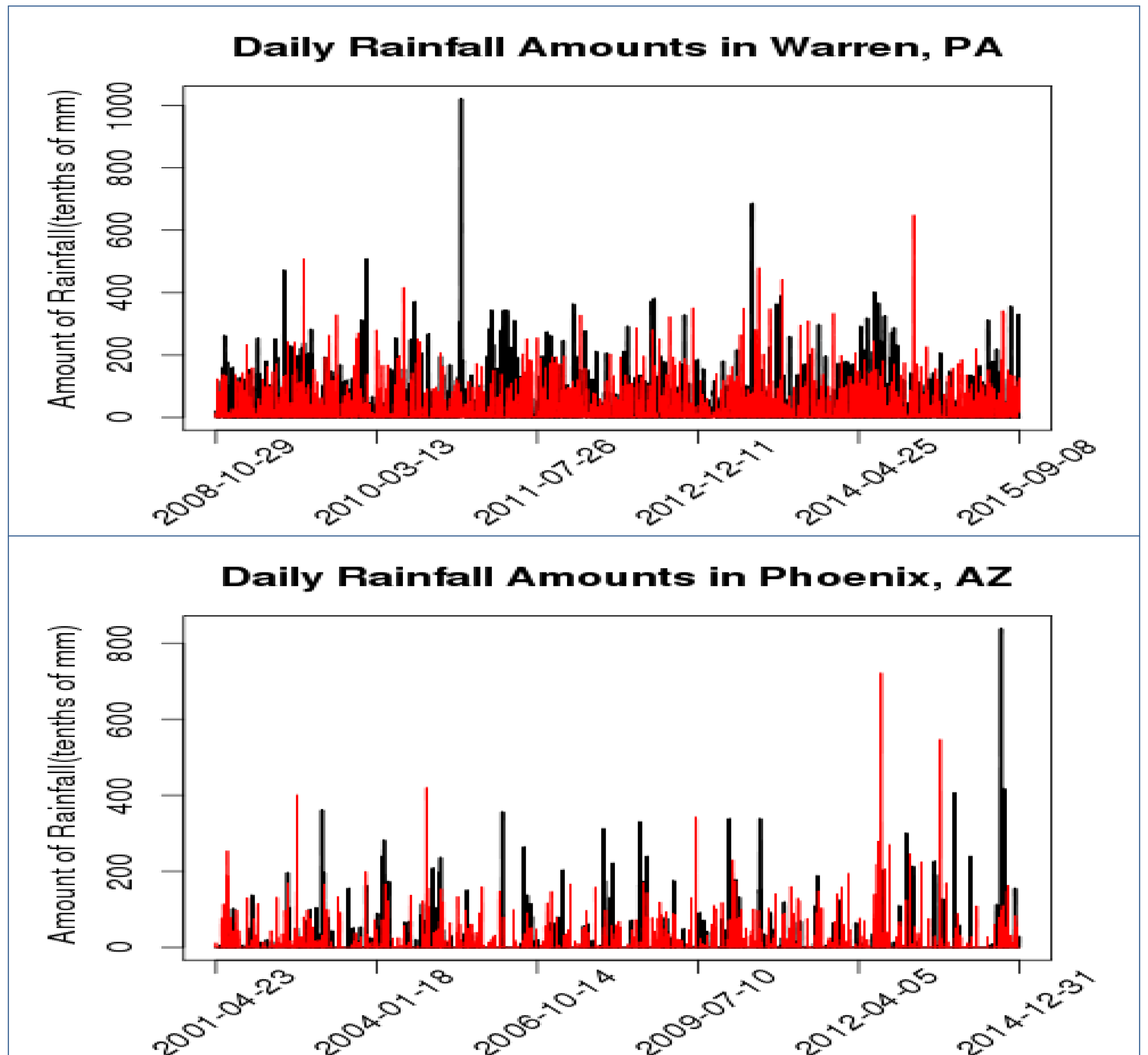

Figure 2. Simulations from the best models (red) capture the zeroes well along with extreme values in the Warren and Phoenix rainfall datasets (black).

Interpretation of Results

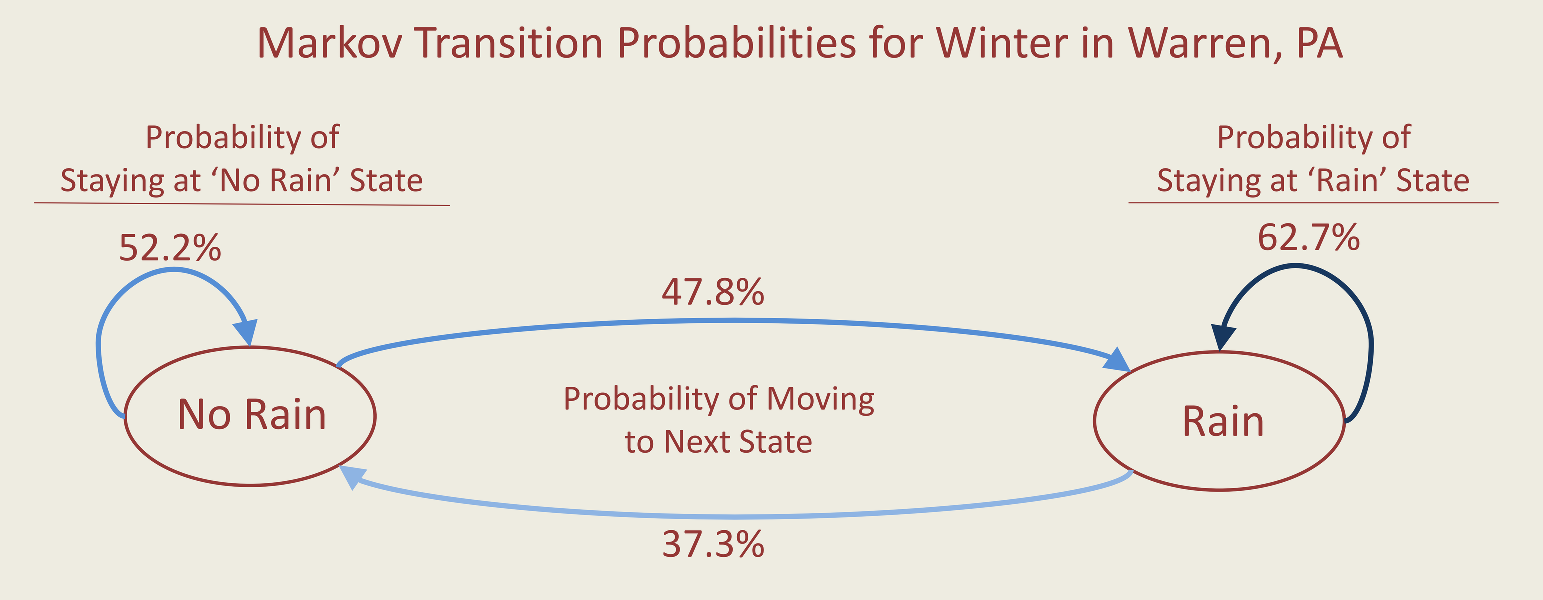

Figure 4. Hidden Markov Model for rainfall in Warren, PA during winter.
More likely that the day following a 'Rain' day it will also Rain', but it is about an equal chance of 'Rain' or 'No Rain' following a 'Rain' day.

Comparison of Best Models: Warren, PA

	CRPS	MSE	BIC
Seasonal 2-stage HMM – Gamma	36.7214	8210	8120
Non-seasonal 2-stage HMM - Gamma	37.1062	8277	7898

Comparison of Best Models: Phoenix, AZ

	CRPS	MSE	BIC
Seasonal 2-stage HMM – Weibull, conditioned on day before	7.1050	1399	8298
Seasonal 2-stage HMM - Weibull	7.1317	1405	8290
Non-seasonal 3-stage HMM – Weibull, Weibull combination	7.1926	1466	7860

Best model by scoring metric

Conclusion and Future Work

- **Conclusions:**
  - Hidden Markov Models with Gamma or Weibull distributions simulated precipitation data well
  - Can improve models by conditioning on rain/no rain the day before
  - CRPS is used as a comparison metric for predictive distributions
    - CRPS compares predictive CDF with observed CDF
    - CRPS leads to models with lower MSE
    - More precise models that better represent observations
- **Future Work:**
  - Add atmospheric variables to the model
  - HMM that changes through time
  - Add spatial data

Contact

Nick Vasko
The Pennsylvania State University,
Network for Sustainable Climate Risk Management,
Department of Statistics
Email: nick7td@gmail.com

Acknowledgements

This work was supported by the National Science Foundation through the Network for Sustainable Climate Risk Management (SCRiM) under NSF cooperative agreement GEO-1240507.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

1. Wilks, D. S. (1999). Interannual variability and extreme-value characteristics of several stochastic daily precipitation models. *Agricultural and Forest Meteorology*, *93*(3), 153-169.
2. Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, *102*(477), 59-378.
3. Grimit et al. (2006). The continuous ranked probability score for circular variables and its application to mesoscale forecast ensemble verification. *Quarterly Journal of the Royal Meteorological Society*, *132*(621C), 2925-2942.
4. Gneiting et al. (2008). Assessing probabilistic forecasts of multivariate quantities, with an application to ensemble predictions of surface winds. *Test*, *17*(2).
5. Ailliot et al. (2015). Stochastic weather generators: an overview of weather type models. *Journal de la Société Française de Statistique*, *156*(1), 101-113.