COLUMBIA UNIVERSITY | MAILMAN SCHOOL of PUBLIC HEALTH

Elizabeth A. Gibson
Environmental Health Sciences
Mailman School of Public Health
Columbia University

Bayesian non-parametric non-negative matrix factorization for identifying patterns in environmental mixtures

PRIME Program Meeting

October 14, 2020

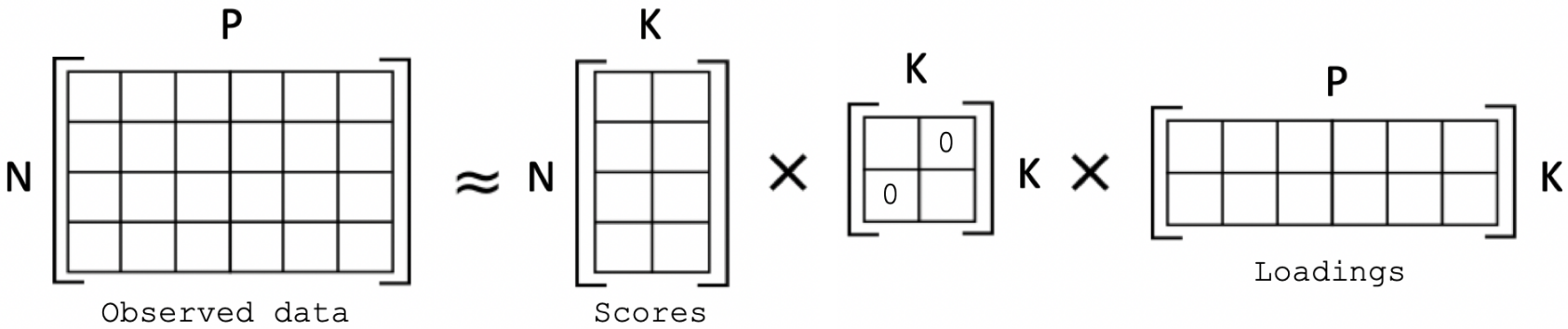
Existing exposure pattern recognition methods

- Choice of k patterns subjective
- $+/-$ values not easily interpretable
- Orthogonality constraint not realistic
- No measure of uncertainty

$\Longrightarrow$ Proposed solution:

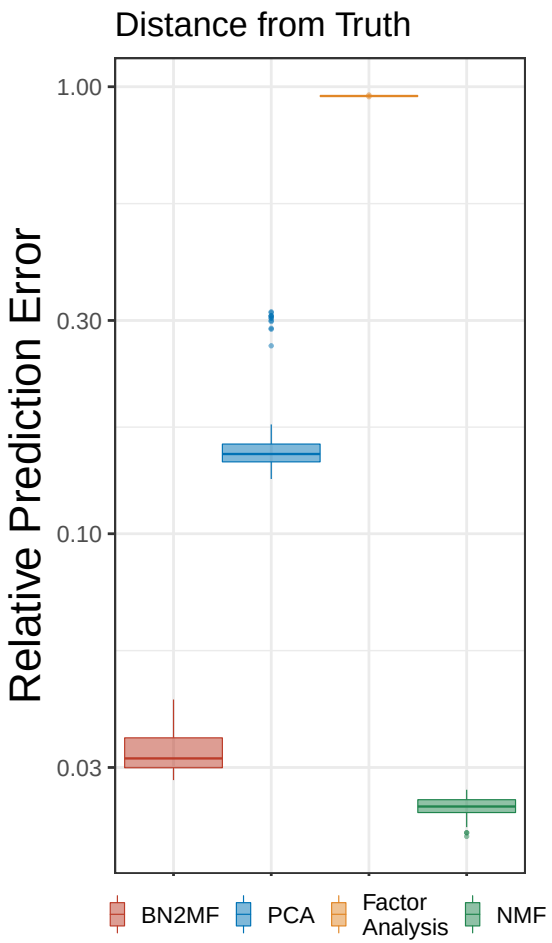
Bayesian non-parametric non-negative matrix factorization

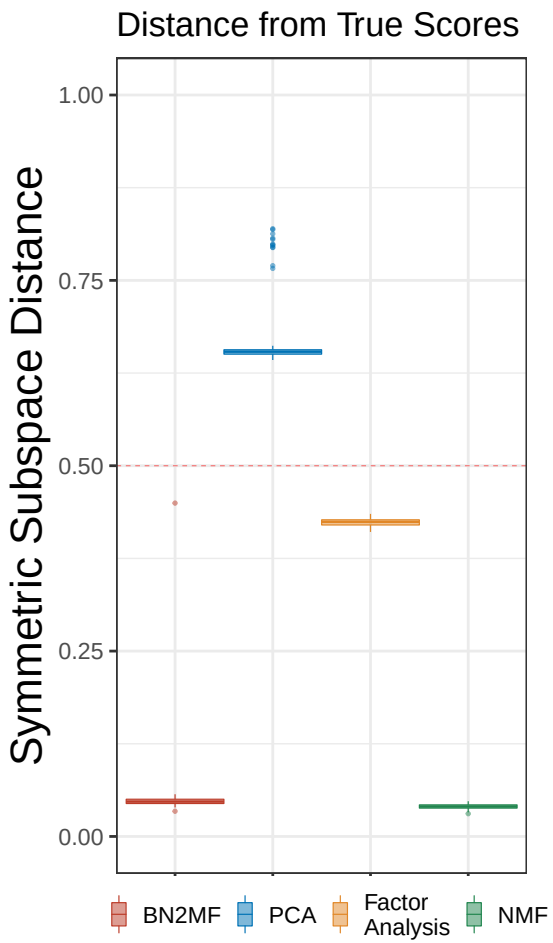
Bayesian non-parametric non-negative matrix factorization (BN^{2}MF)

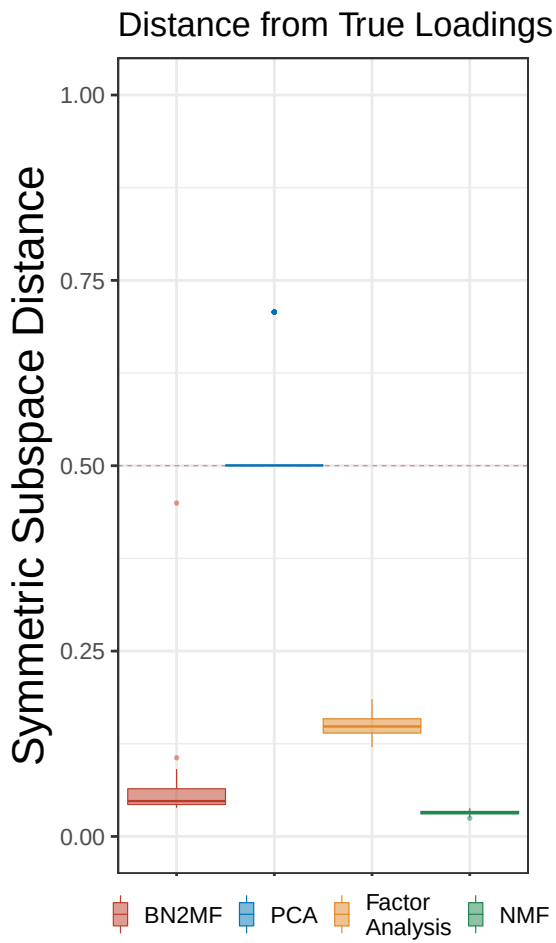


- ✓ Non-negative continuous Gamma priors
  - ✓ Sparse prior on patterns estimates number
- ✓ No orthogonality constraint
- ✓ Variational confidence intervals

Results: simulation study







Conclusion

- Increased interpretability due to:
  - Parts-based (additive) **non-negative** representation of multi-pollutant mixtures
  - Absence of orthogonality constraint on loadings and scores
- ✓ Non-parametric prior on k helps with model selection
- ✓ Bayesian framework allows uncertainty propagation
- Application to real environmental data can identify sensible patterns

Acknowledgements

Columbia University:

Julie Herbstman
Environmental Health

Marianthi-Anna Kioumourtzoglou
Environmental Health

Jeff Goldsmith
Biostatistics

John Paisley
Electrical Engineering

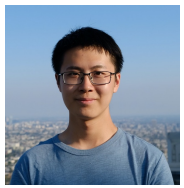
makLab

Mike He
Maggie Li
Robbie Parks
Sebastian Rowland
Jenni Shearston
Rachel Tao

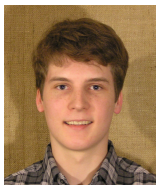
PRIME Team:

John Wright
Electrical Engineering

Yanelli Núñez
Environmental Health



Jingkai Yan
Electrical Engineering



Robert Colgan
Computer Science

Lawrence Chillrud
Computer Science

Supported by:
NIEHS F31 ES030263 &
PRIME R01 ES028805

Contact info:

e.a.gibson@columbia.edu
www.lizzy.codes