Graphical Latent Variable Models with Applications to Climate Data

Padhraic Smyth
Department of Computer Science
University of California, Irvine

UCIrvine
UNIVERSITY OF CALIFORNIA, IRVINE

Outline

- **Graphical Models**
  - Bayesian learning
  - Latent variable models
  - Hidden Markov models

- **Applications to Precipitation and Atmospheric Data Sets**
  - Non-homogeneous hidden Markov models for downscaling
  - Grid-based graphical models for ITCZ detection

- **Concluding Comments**

Graphical Models and Bayesian Estimation

Spectrum of Machine Learning Techniques

- **Predictive Modeling**
  - Predict Y given X, e.g., compute P [Y | X]
  - Emphasis on prediction accuracy, not so much on model interpretation

- **Generative Modeling**
  - Model Y and X jointly, i.e., model P(Y, X)
  - Emphasis both on prediction (e.g., for missing data) and interpretation
  - Often use graphical models and Bayesian methods

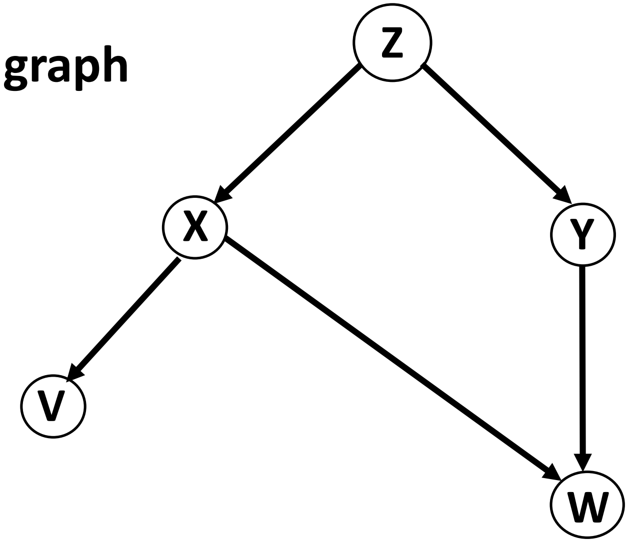
- **Pattern Finding**
  - Clustering, outlier detection, exploratory data analysis
  - Can be used to support predictive and generative modeling

Directed Graphical Models

Represent stochastic dependencies with a directed graph

Nodes = variables

Edges = direct dependencies

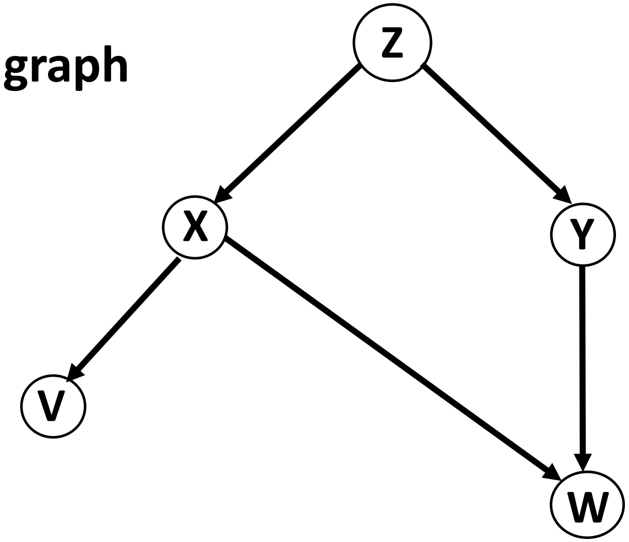


Directed Graphical Models

Represent stochastic dependencies with a directed graph

Nodes = variables

Edges = direct dependencies



Joint distribution

= product of conditional distributions

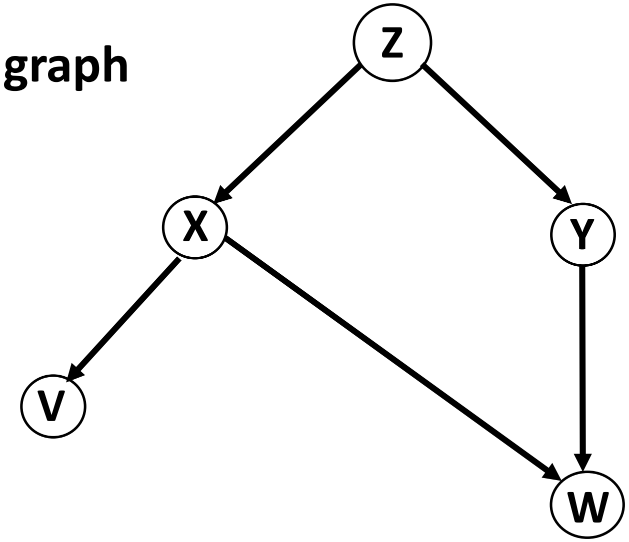
$$P(V, W, X, Y, Z) = P(V|X)P(W|X, Y)P(X|Z)P(Y|Z)P(Z)$$

Directed Graphical Models

Represent stochastic dependencies with a directed graph

Nodes = variables

Edges = direct dependencies



Joint distribution

= product of conditional distributions

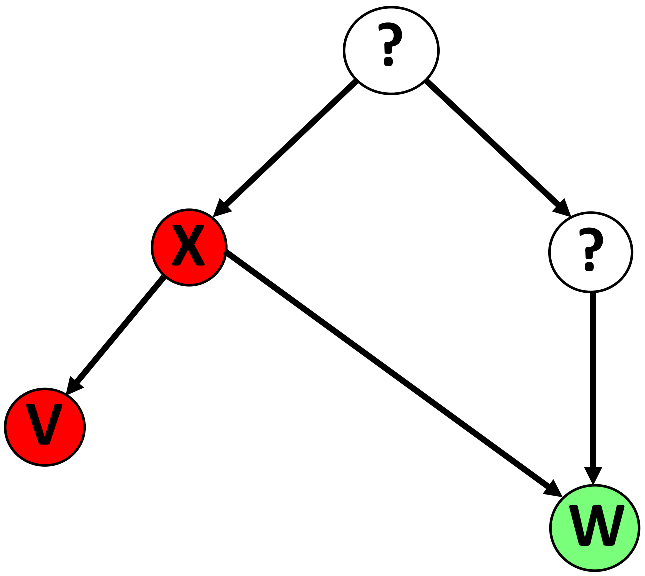
$$P(V,W,X,Y,Z) = P(V|X)P(W|X,Y)P(X|Z)P(Y|Z)P(Z)$$

Why is this useful?

Representational language

Computational algorithms

Computation of Conditional Distributions

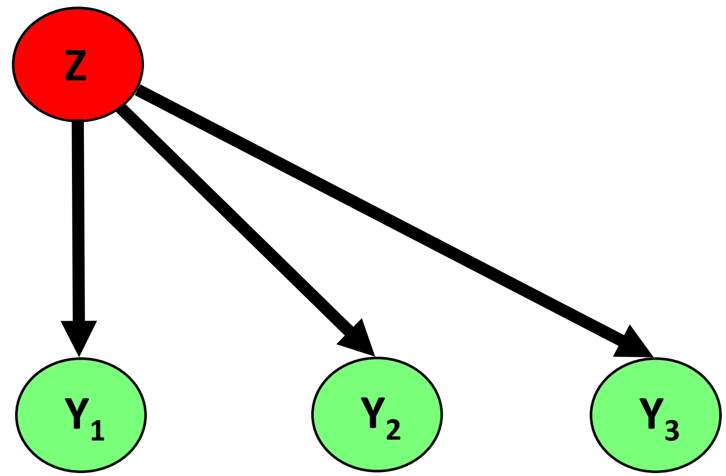


Can use the graph structure + Bayes rule to compute P(Z, Y | W)

Conditional Independence and Hidden Variables

Hidden variable

Observed variable



Latent/Hidden Variable Models

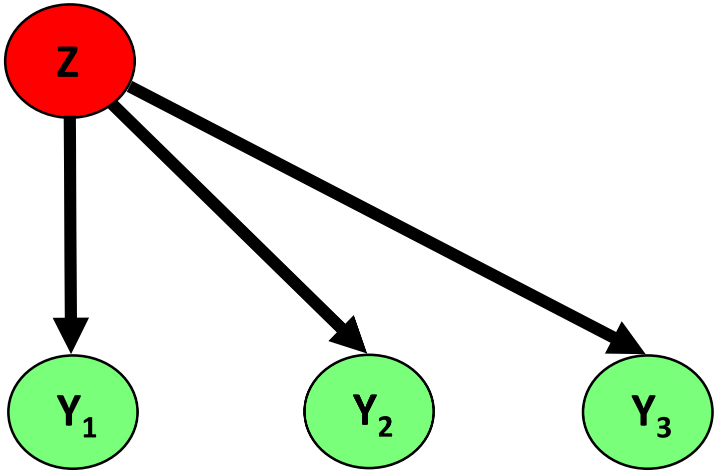
- **Example: modeling high-dimensional y**
  - Introduce latent variables z that simplify P(y)
  - e.g., given latent variables, the y's are approximately conditionally independent

- **Interpretation of latent variables z**
  - Physical interpretation:
    - z is in principle measurable, e.g., Kalman estimation
  - Exploratory interpretation:
    - Discrete z, e.g., clustering of weather states, storm trajectories,
    - Real-valued z, e.g., potential low-dimensional latent mechanisms, e.g., EOFs
  - Agnostic interpretation:
    - z's are useful for approximation and modeling, not necessarily any interpretation

Conditional Independence

Hidden variable

Observed variable

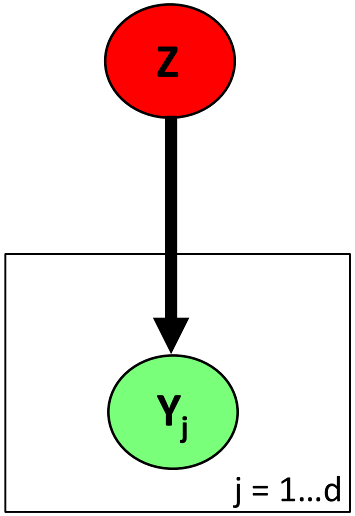


UCIrvine
University of California, Irvine

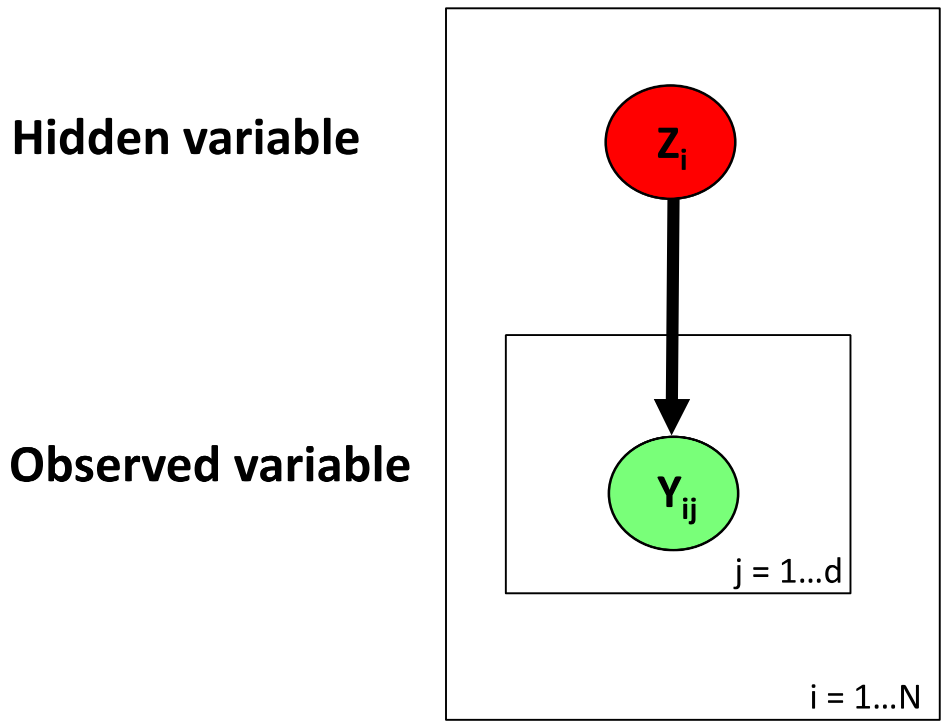
Plate Diagram

Hidden variable

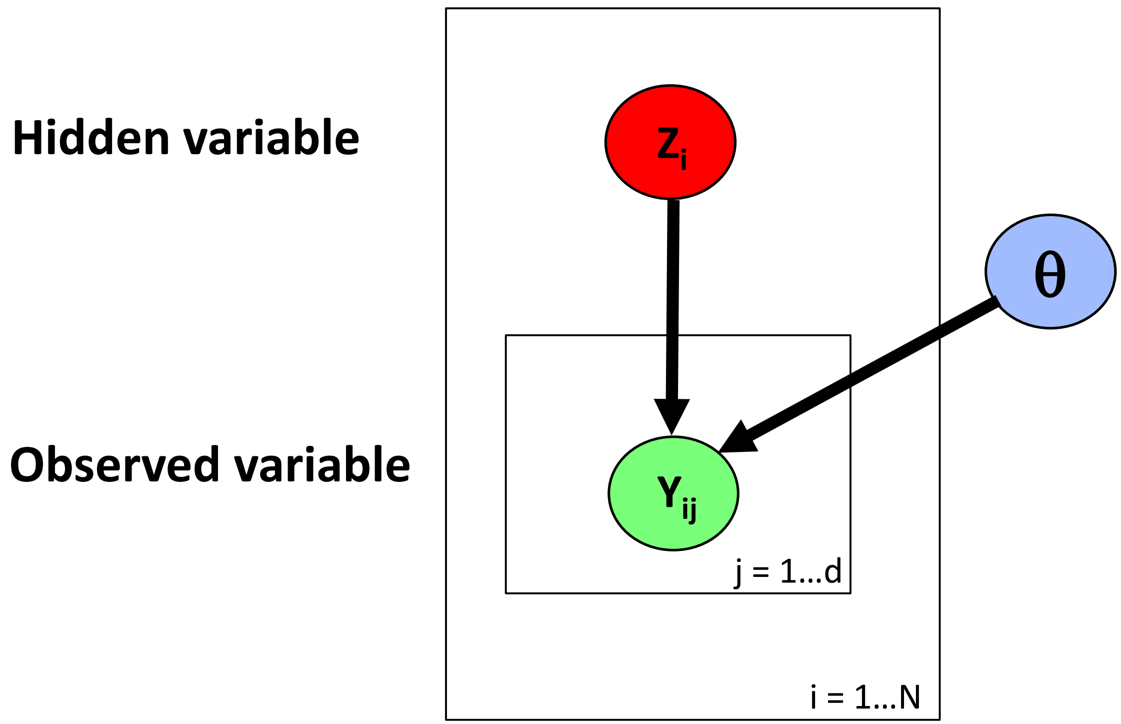
Observed variable



Multiple Observations



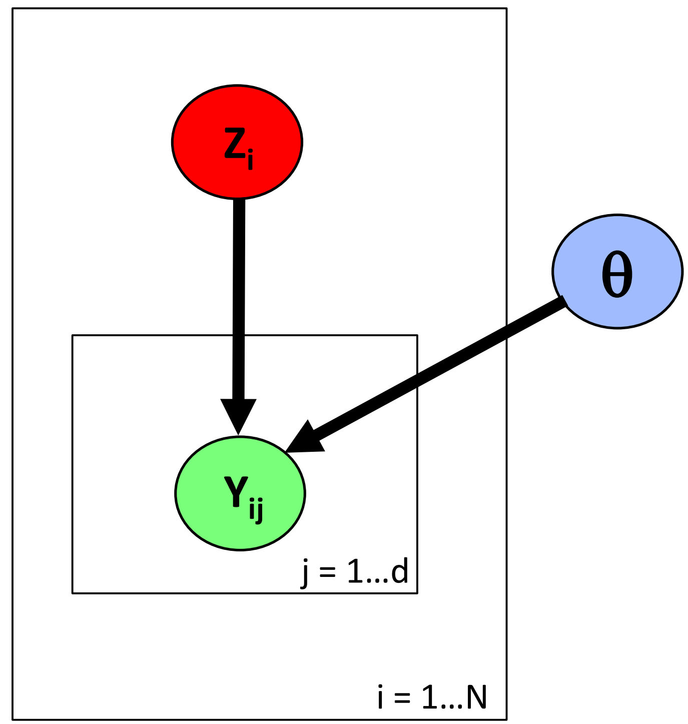
Parameters and Likelihood



Parameters and Likelihood

Hidden variable

Observed variable

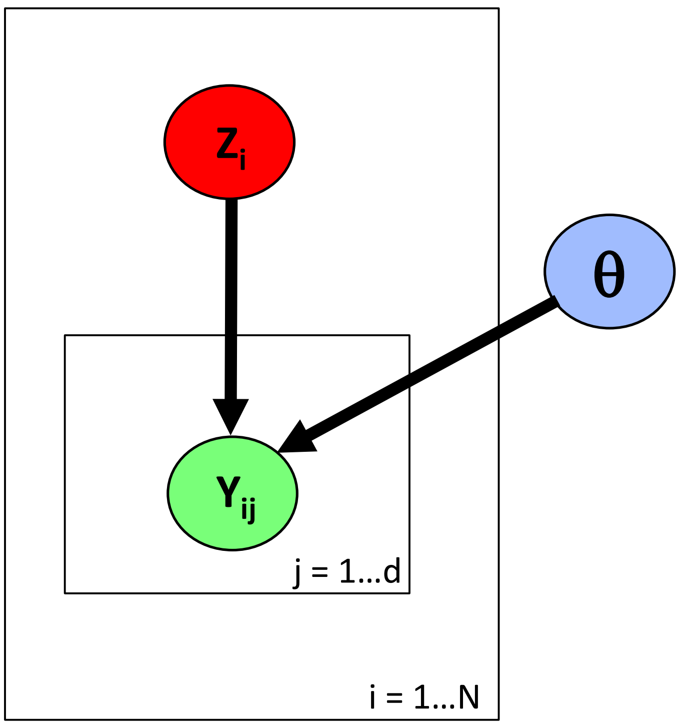


$$L(\theta) = P(Y|\theta, Z) = \prod_{i=1}^{N} \prod_{j=1}^{d} P(y_{ij}|z_i, \theta)$$

Parameters and Likelihood

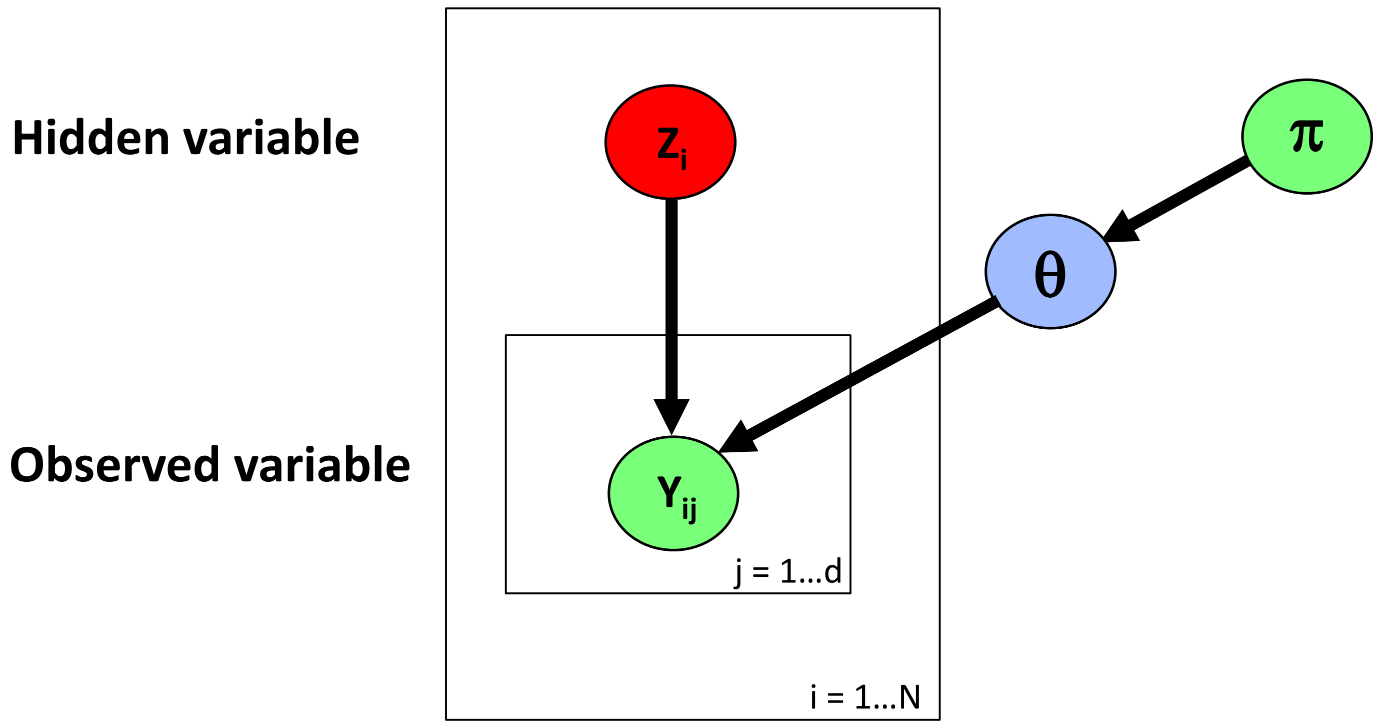
Hidden variable

Observed variable



$$L(\theta) = P(Y|\theta, Z) = \prod_{i=1}^{N} \prod_{j=1}^{d} P(y_{ij}|z_i, \theta)$$

Priors and Bayesian Estimation

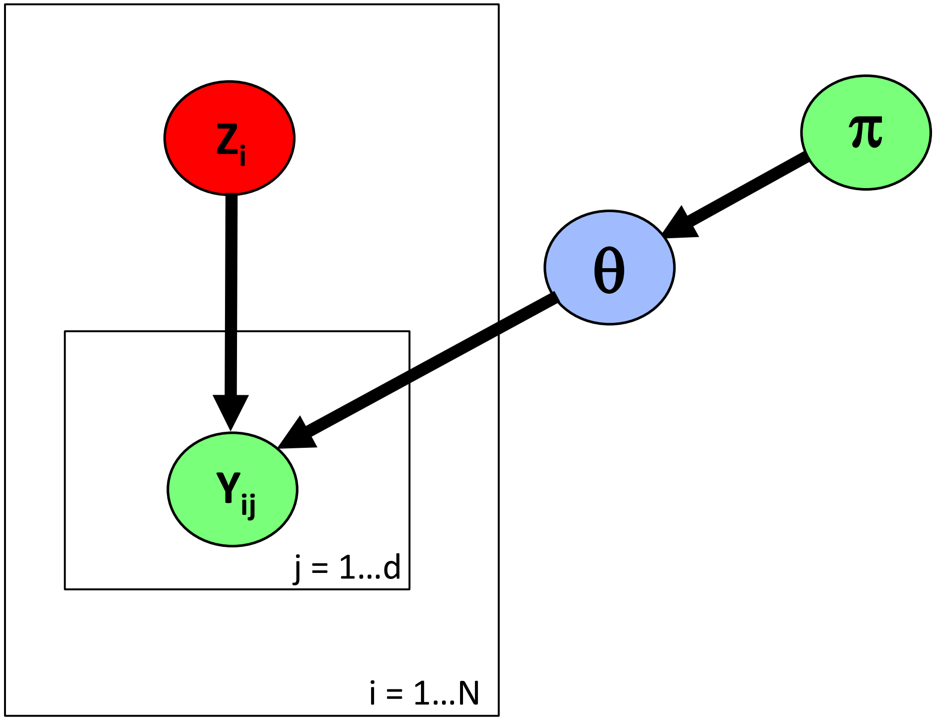


$$P(\theta|Y, Z, \pi) \propto P(Y|\theta, Z)P(\theta|\pi)$$

Priors and Bayesian Estimation

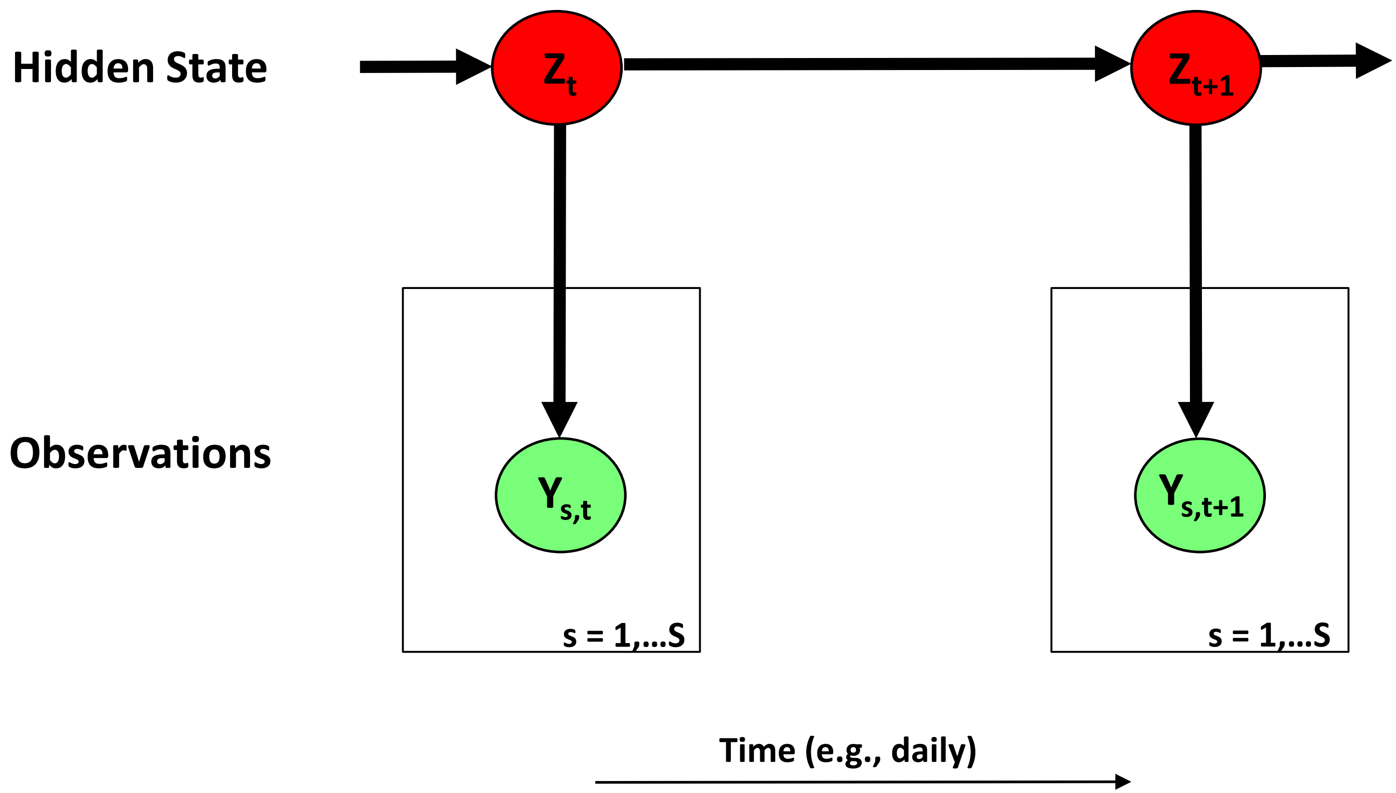
Hidden variable

Observed variable

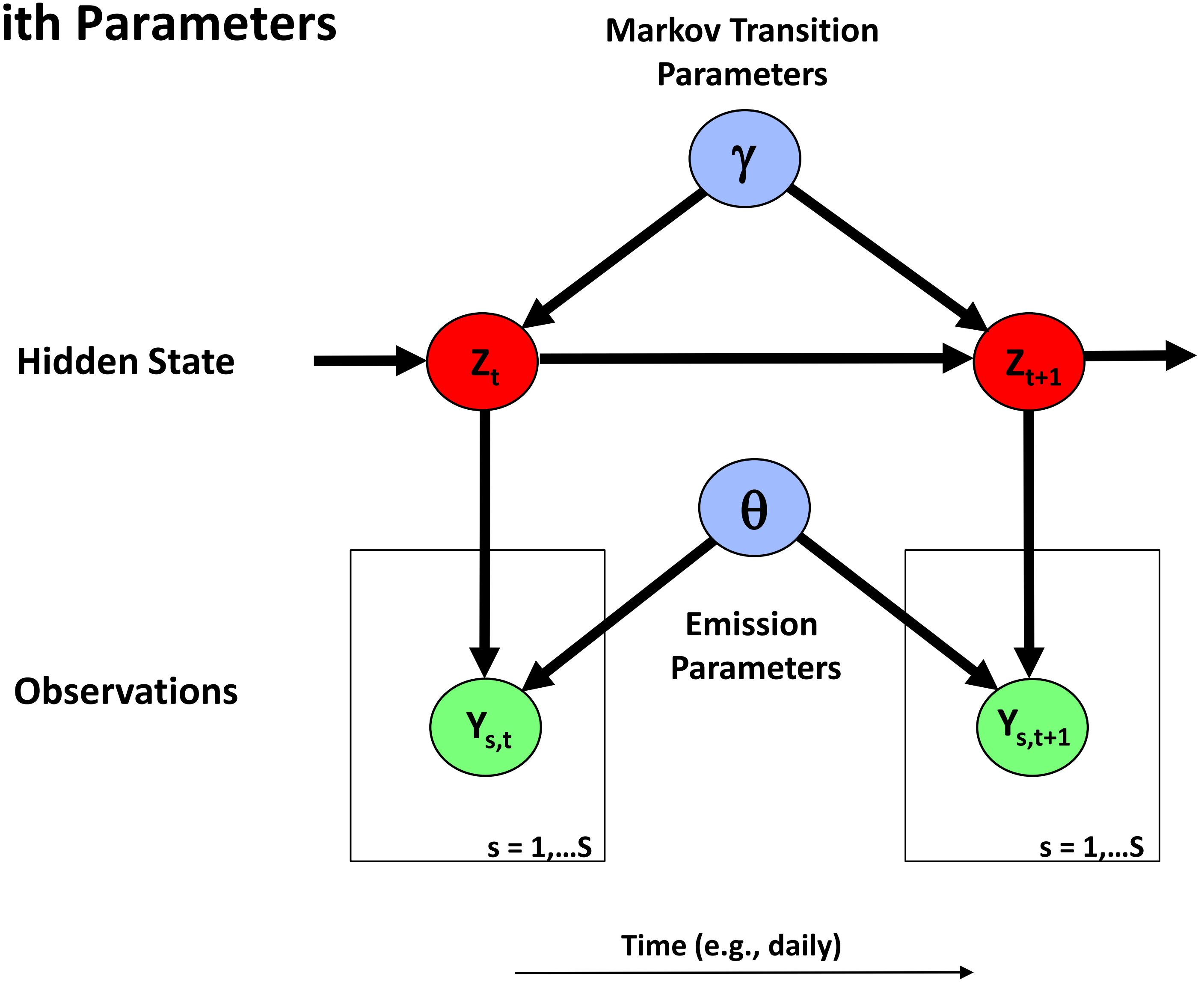


$$P(\theta|Y,\pi) \propto \left(\sum_{Z} P(Y|\theta,Z)P(Z)\right)P(\theta|\pi)$$

Hidden Markov Model

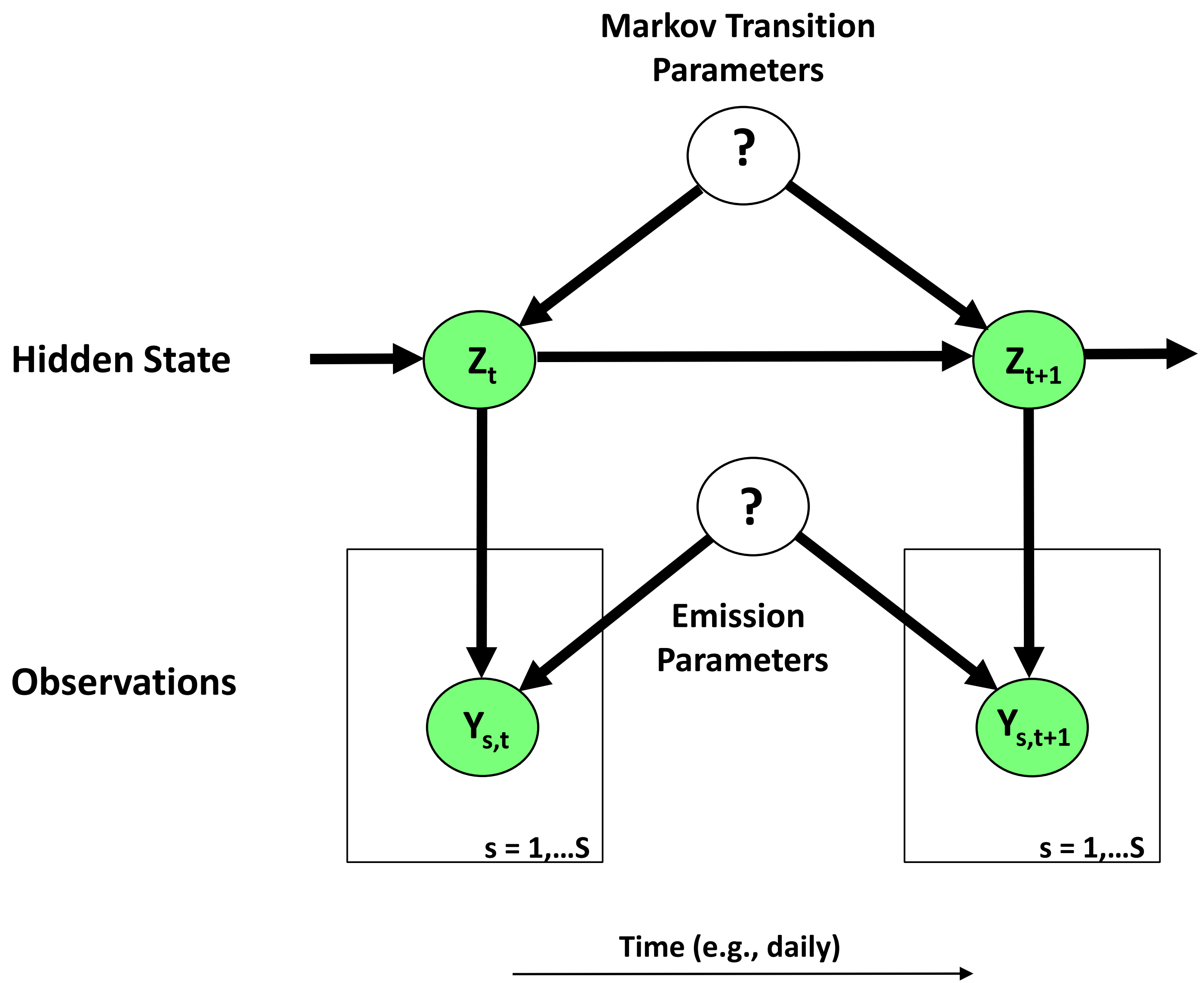


…with Parameters



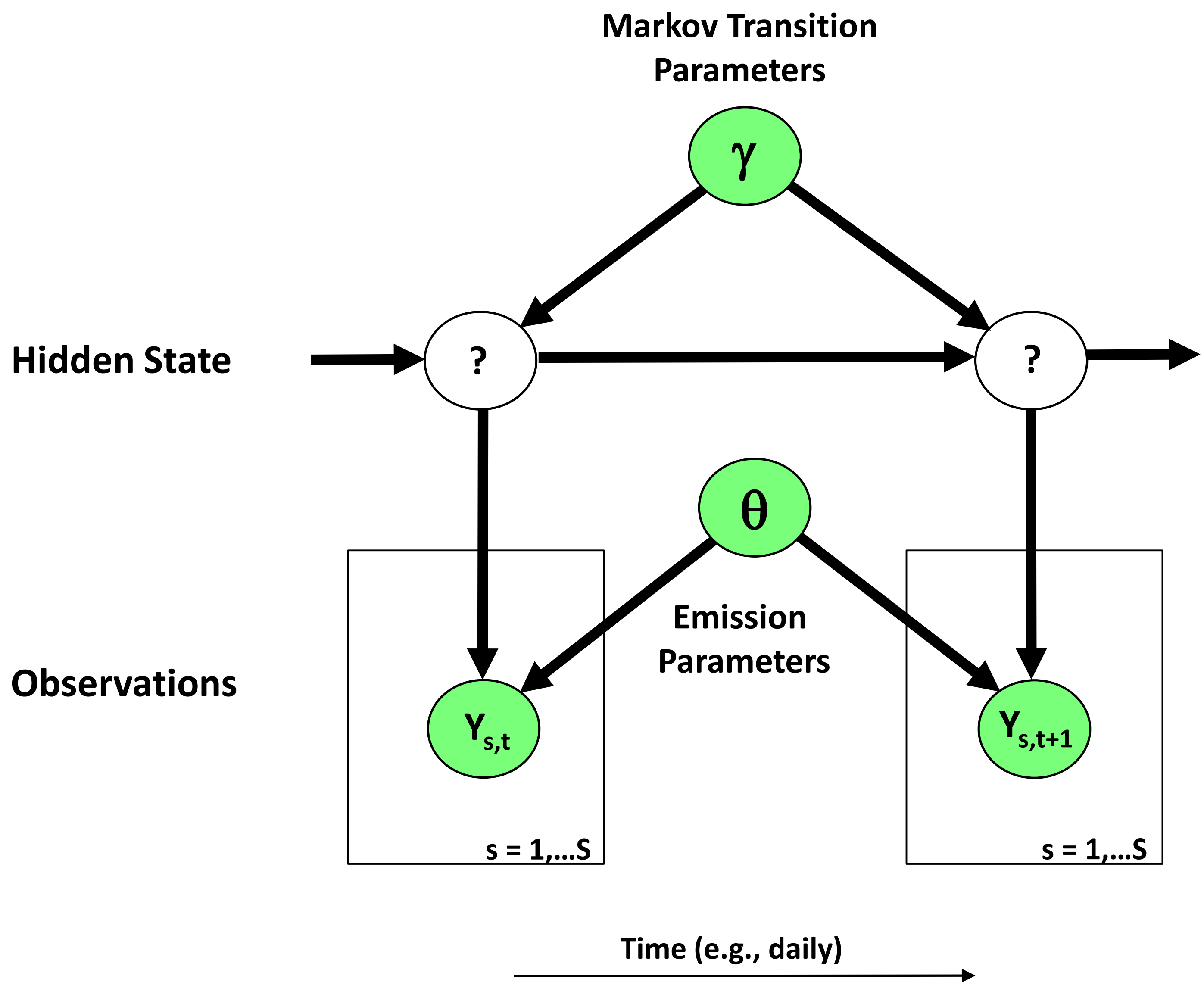
MCMC: Sample Parameters given Hidden States

(Scott, JASA, 2002)

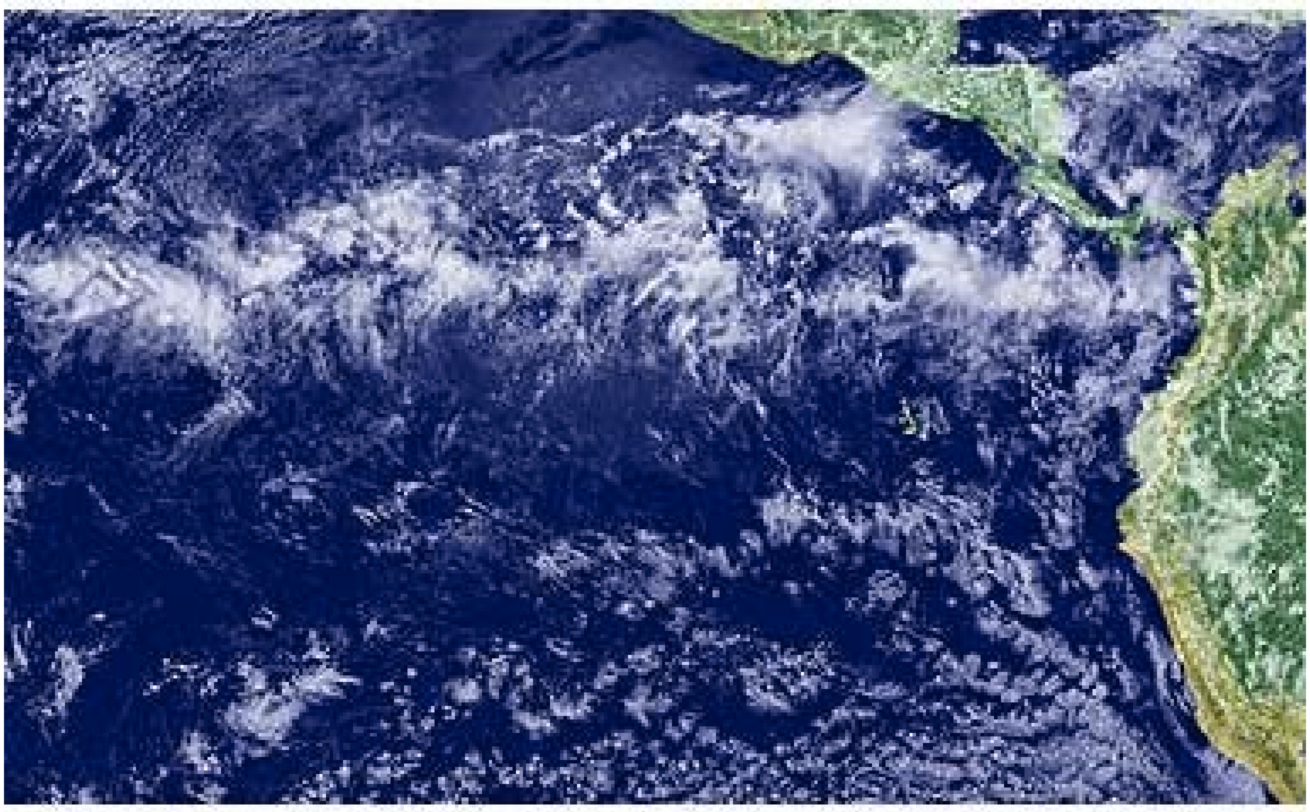


MCMC: Sample Hidden States given Parameters

(Scott, JASA, 2002)



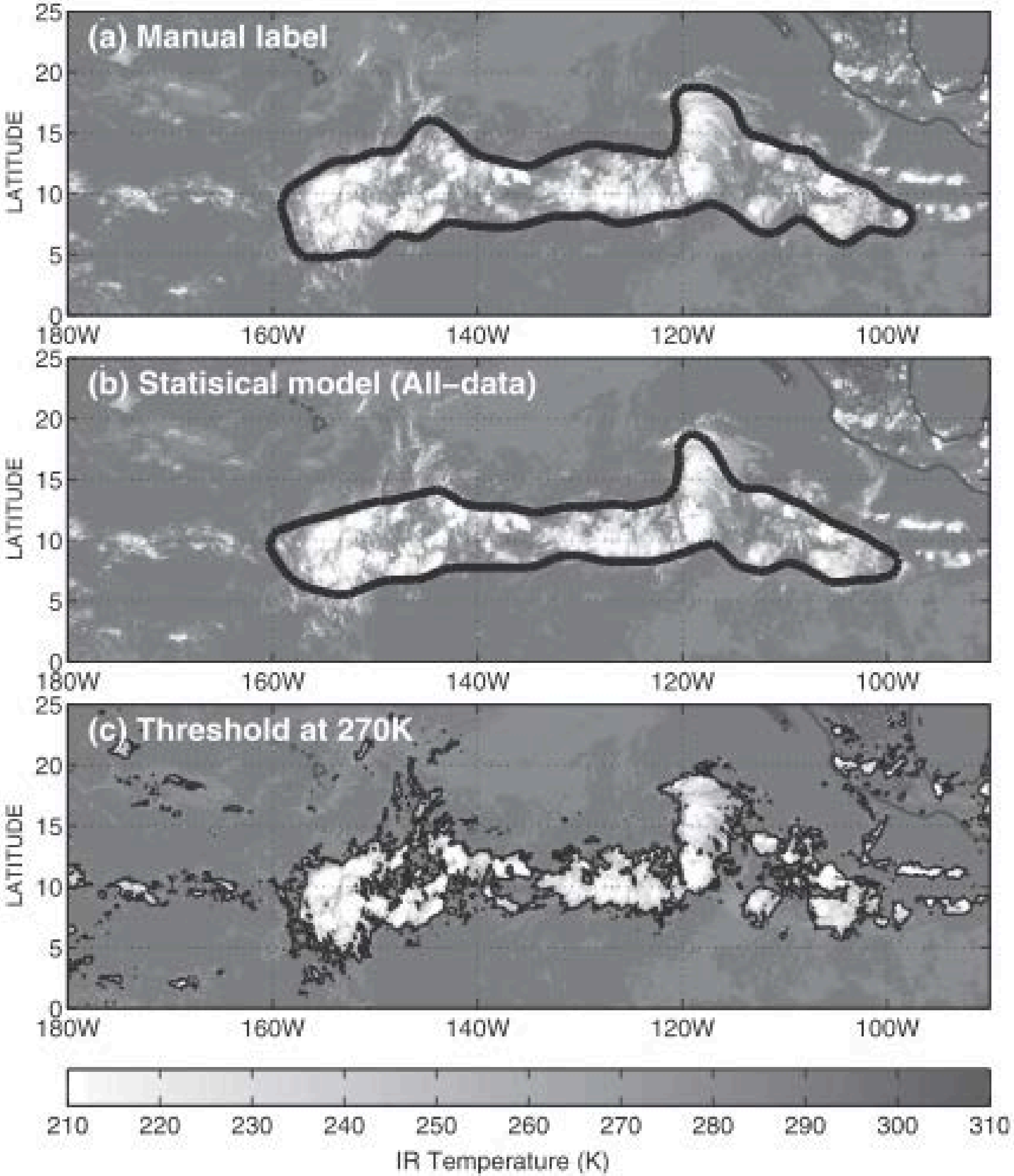
Analyzing the InterTropical Convergence Zone (ITCZ)



Data:
3 hourly infra-red,
1980-2009

24-hourly visible,
1995 to 2008

Problem:
Limited availability of
"data products"



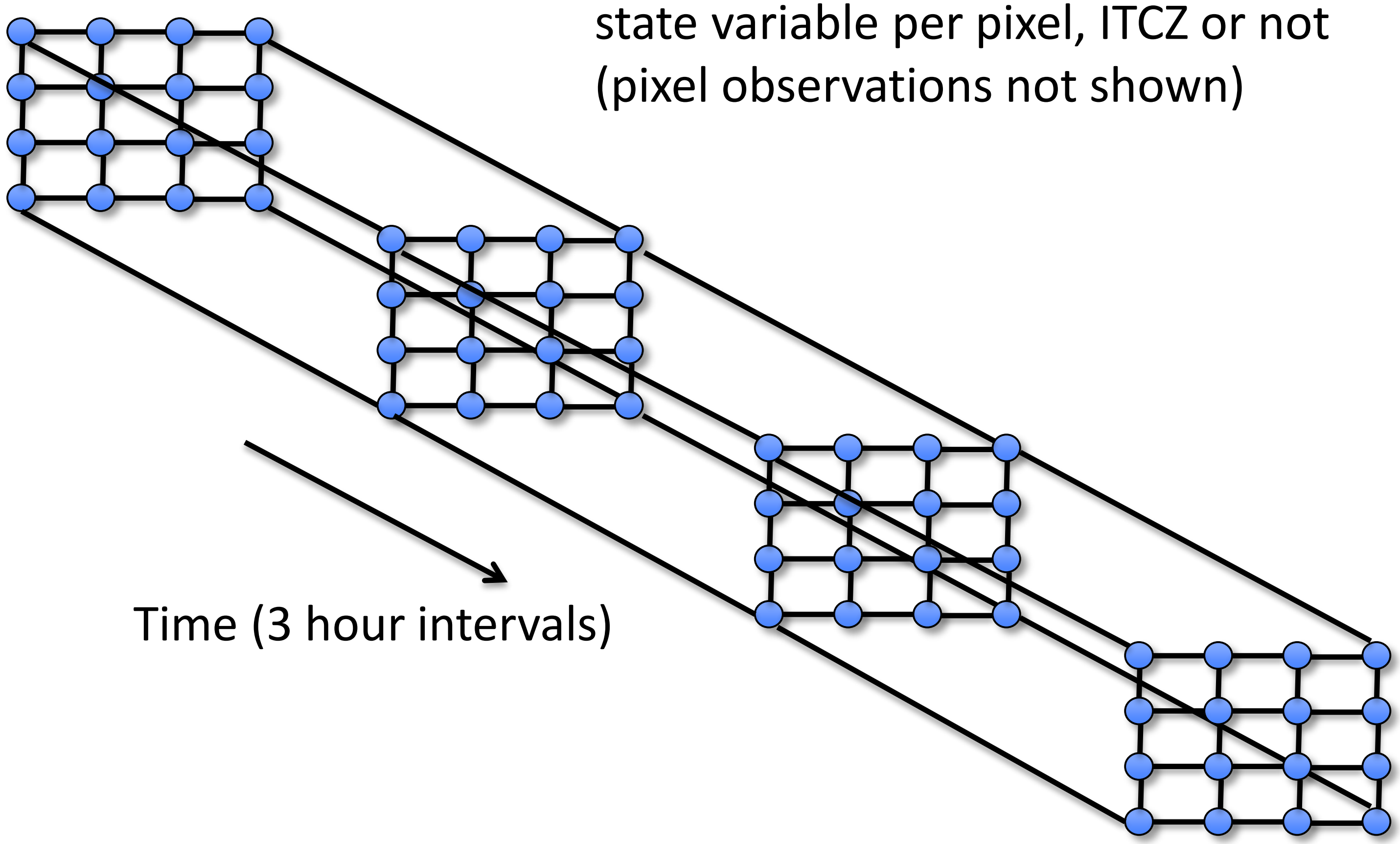
UCIrvine
University of California, Irvine

Example of Results

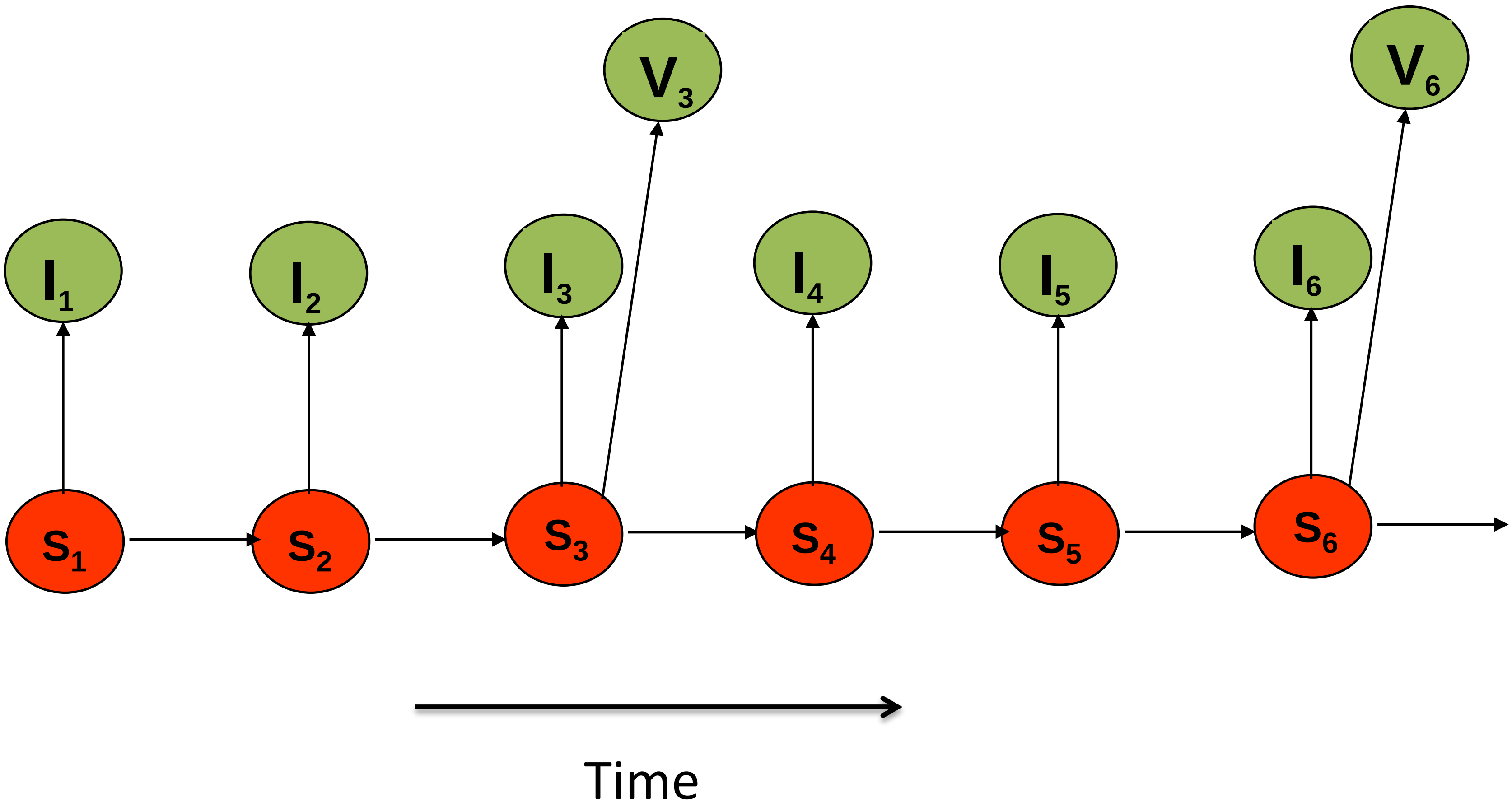
From
Bain et al, *JGR Atmospheres*, 2010
Bain et al, 2011, *Journal of Climate*

Graphical Model for ITCZ Detection

Hidden state structure, one binary hidden state variable per pixel, ITCZ or not (pixel observations not shown)



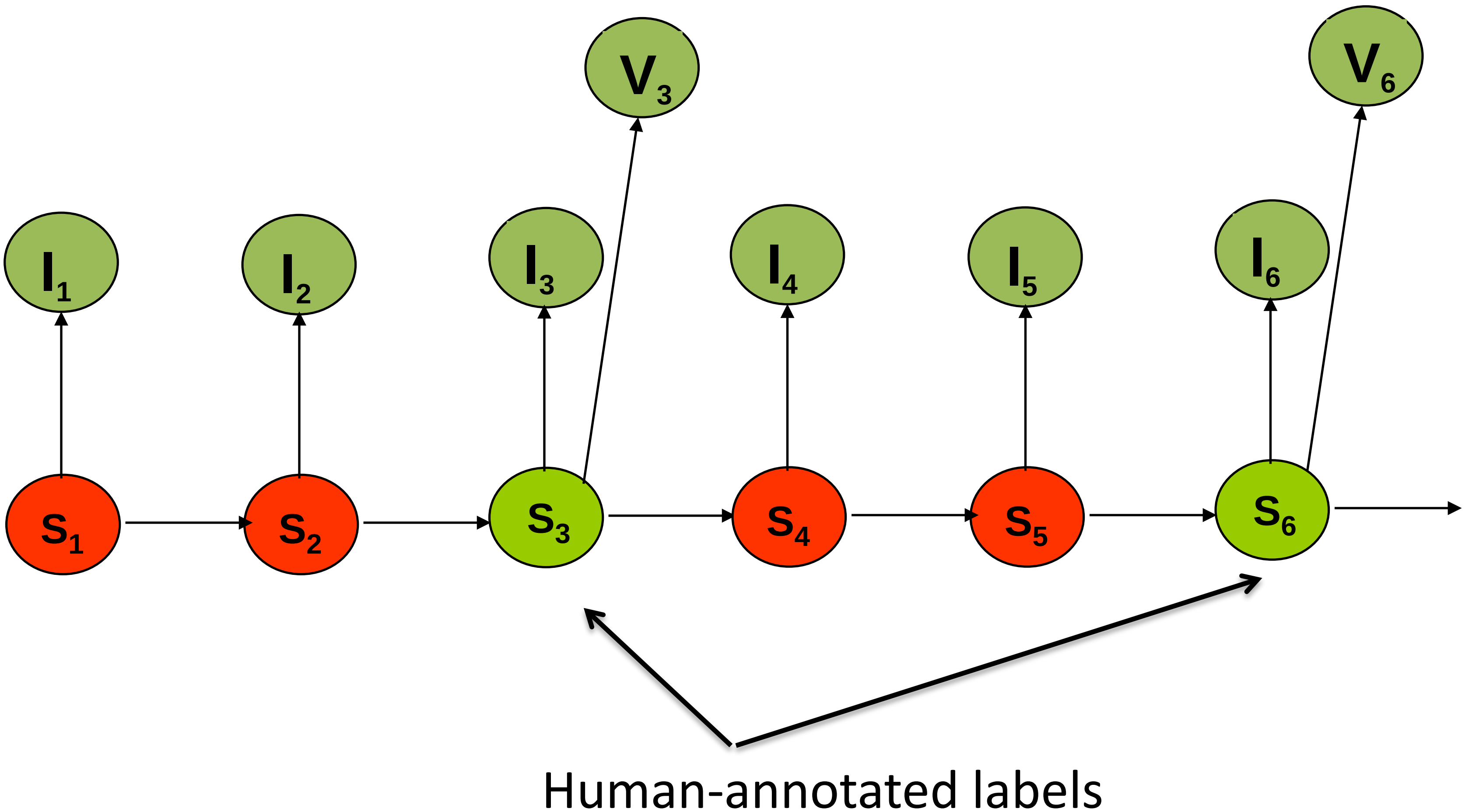
States and Observations per Pixel



UCIrvine
University of California, Irvine

Semi-Supervised Labeling

Henke et al, *Remote Sensing of Environment*, 2012



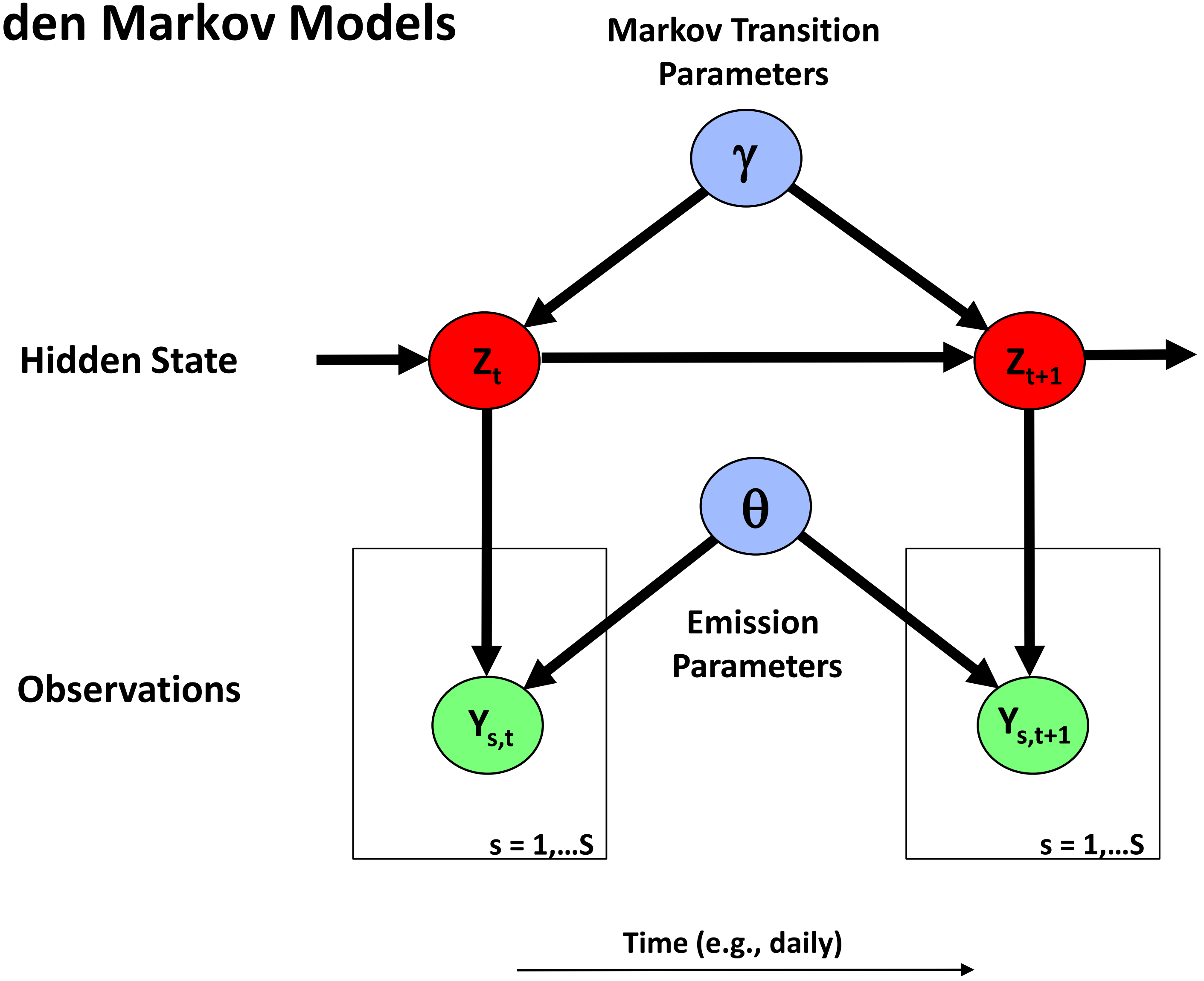
Key Points about Graphical Models

- **Builds probability models via conditional independence assumptions**
- **Graph representation**
  - Nodes represent variables of interest
  - Edges (and absence) encode conditional independence assumptions
- **Computation**
  - Inference corresponds to propagating information on the graph
  - Natural treatment of missing data and latent variables
- **Bayesian view:**
  - Parameters can also be treated as nodes in the graph
  - Parameter estimation corresponds to computation in the graphical model

Applications to Precipitation Data

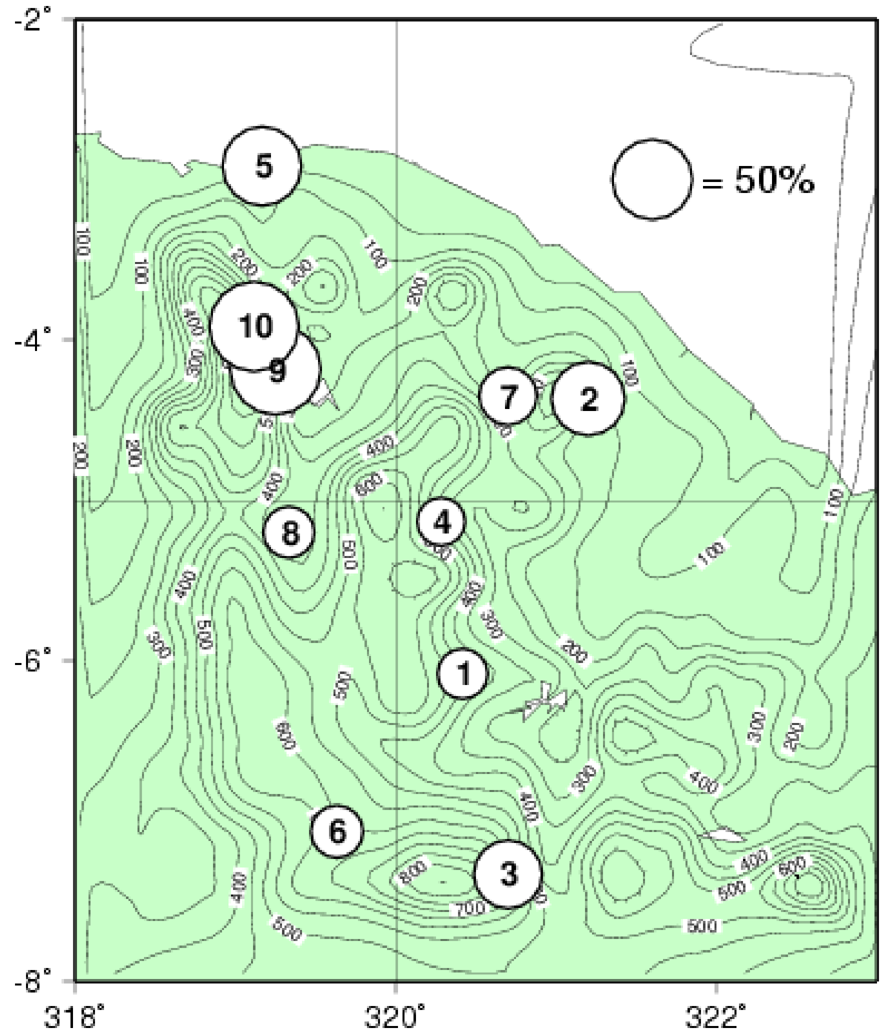
UCIrvine
University of California, Irvine

Hidden Markov Models

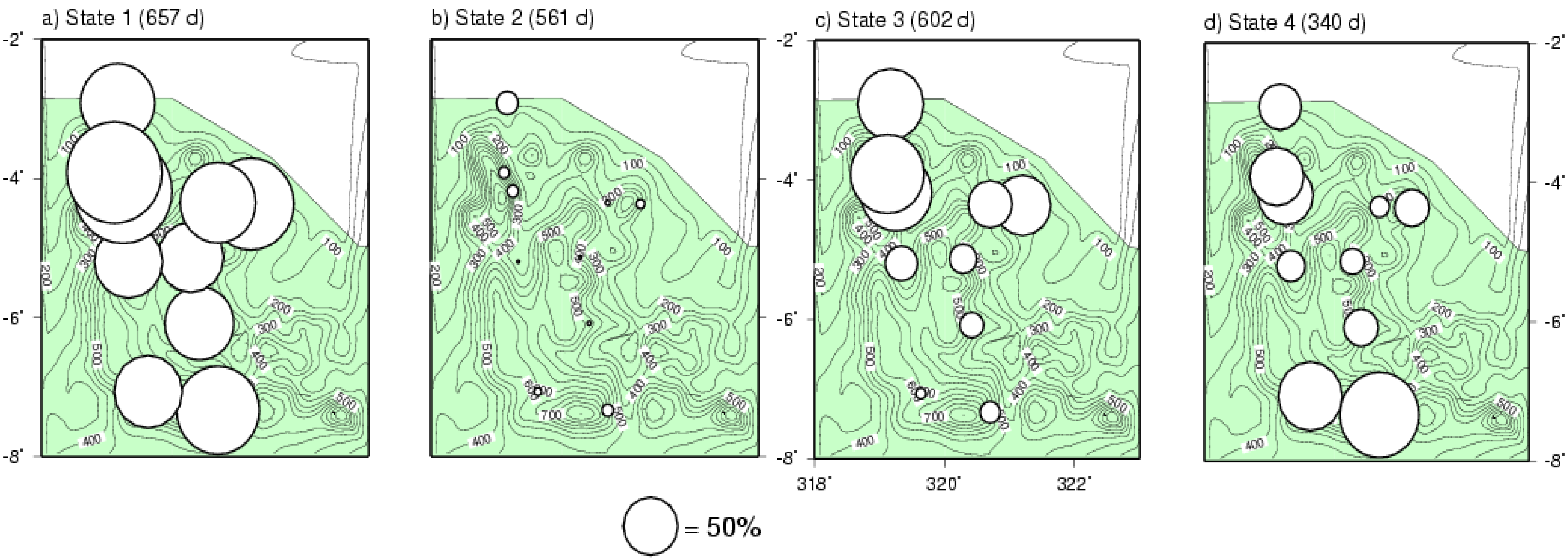


Daily Station Data from Northeast Brazil

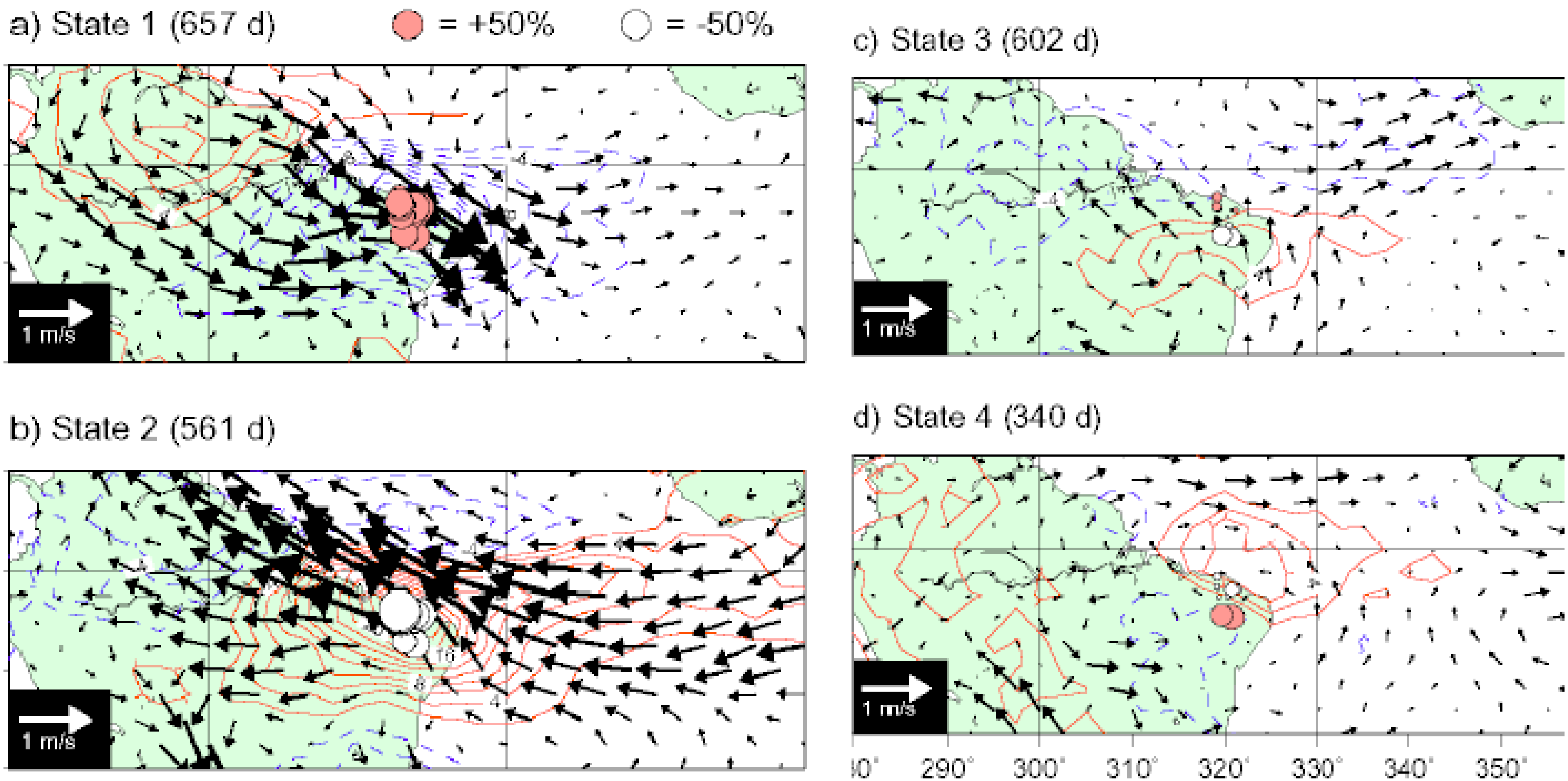
90-day time series
24 years
10 stations



Weather States from the Model



Composites of Wind Fields for each State



Challenges in Precipitation Data Analysis

- **Data quality**
  - Rain gauges: very sparse global coverage
  - Satellites: limited temporal coverage, calibration issues
- **Distributional characteristics**
  - Skewed, non-normal
  - Extremes are very important: but limited data availability for modeling
- **Temporal characteristics**
  - Bursty, seasonal (e.g., monsoon)
  - Interannual variability poorly understood
- **Spatial characteristics**
  - Spatial correlations are not isotropic, depend on many factors
  - Significant regional differences (e.g., tropical versus non-tropical rainfall)

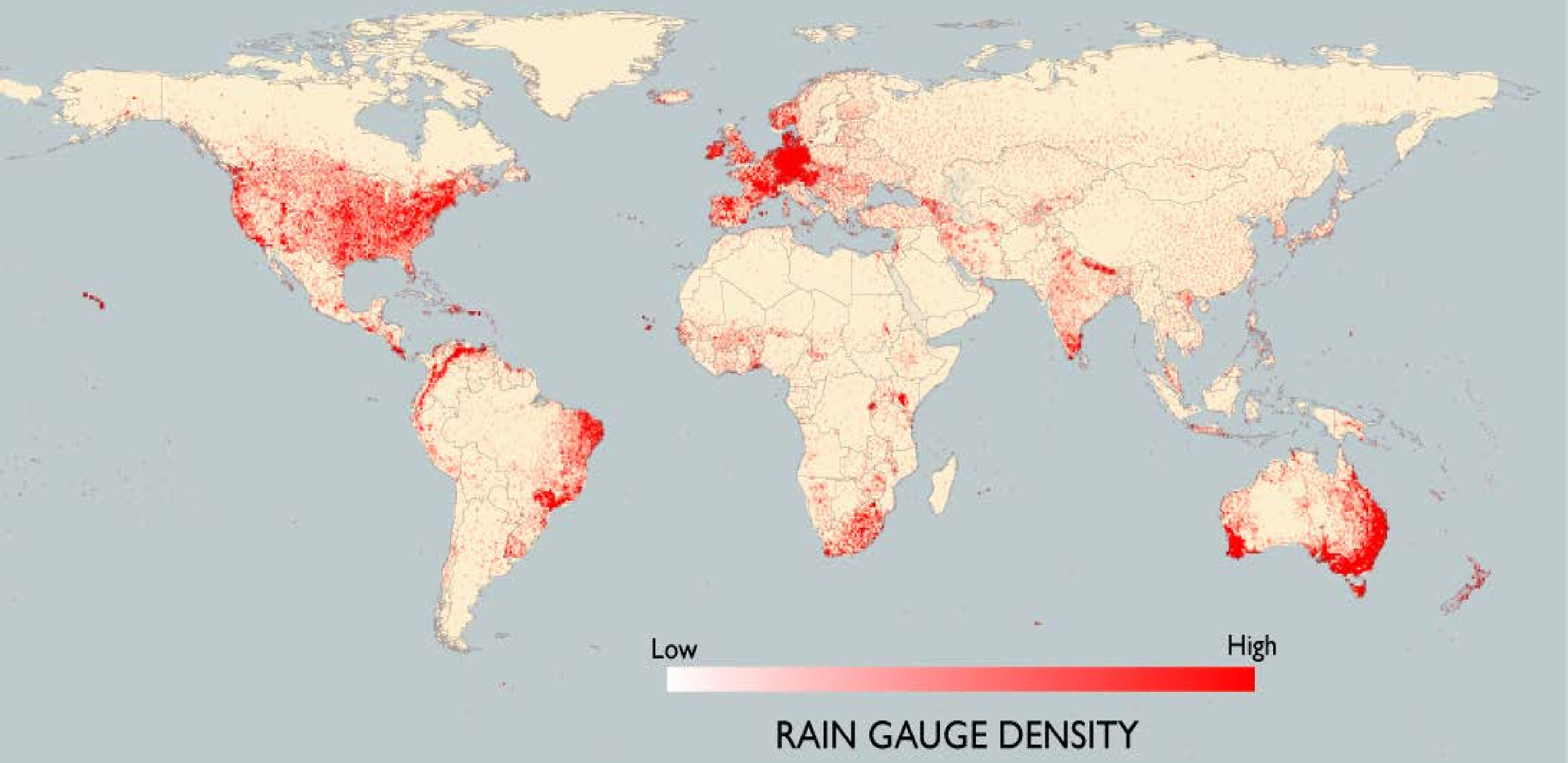


Figure from Lisa Goddard, International Research Institute for Climate and Society

Rain Gauges

Africa: 2,967

Germany: 4,133

Land Area

Africa **11.7** million sq. miles

Germany **0.14**

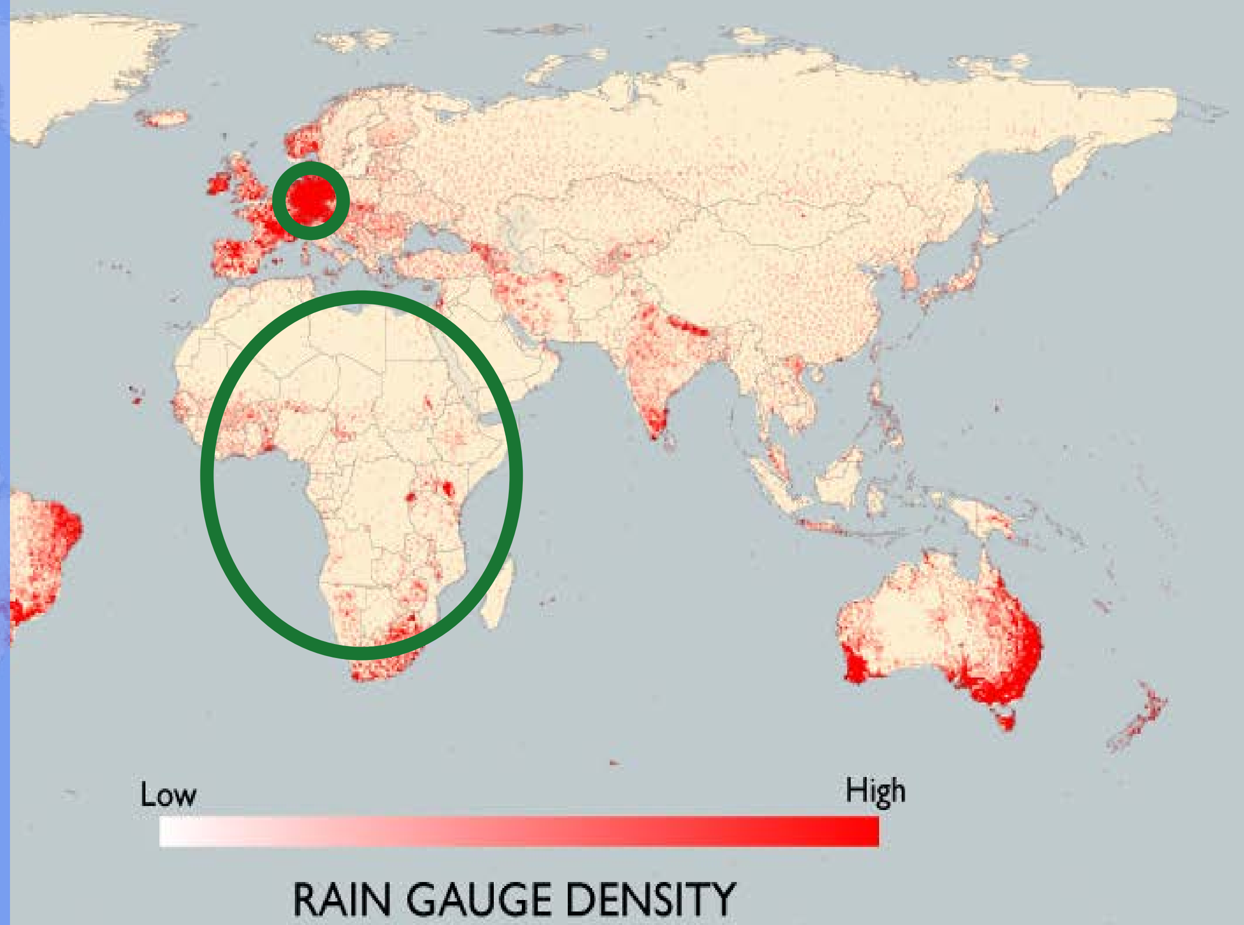
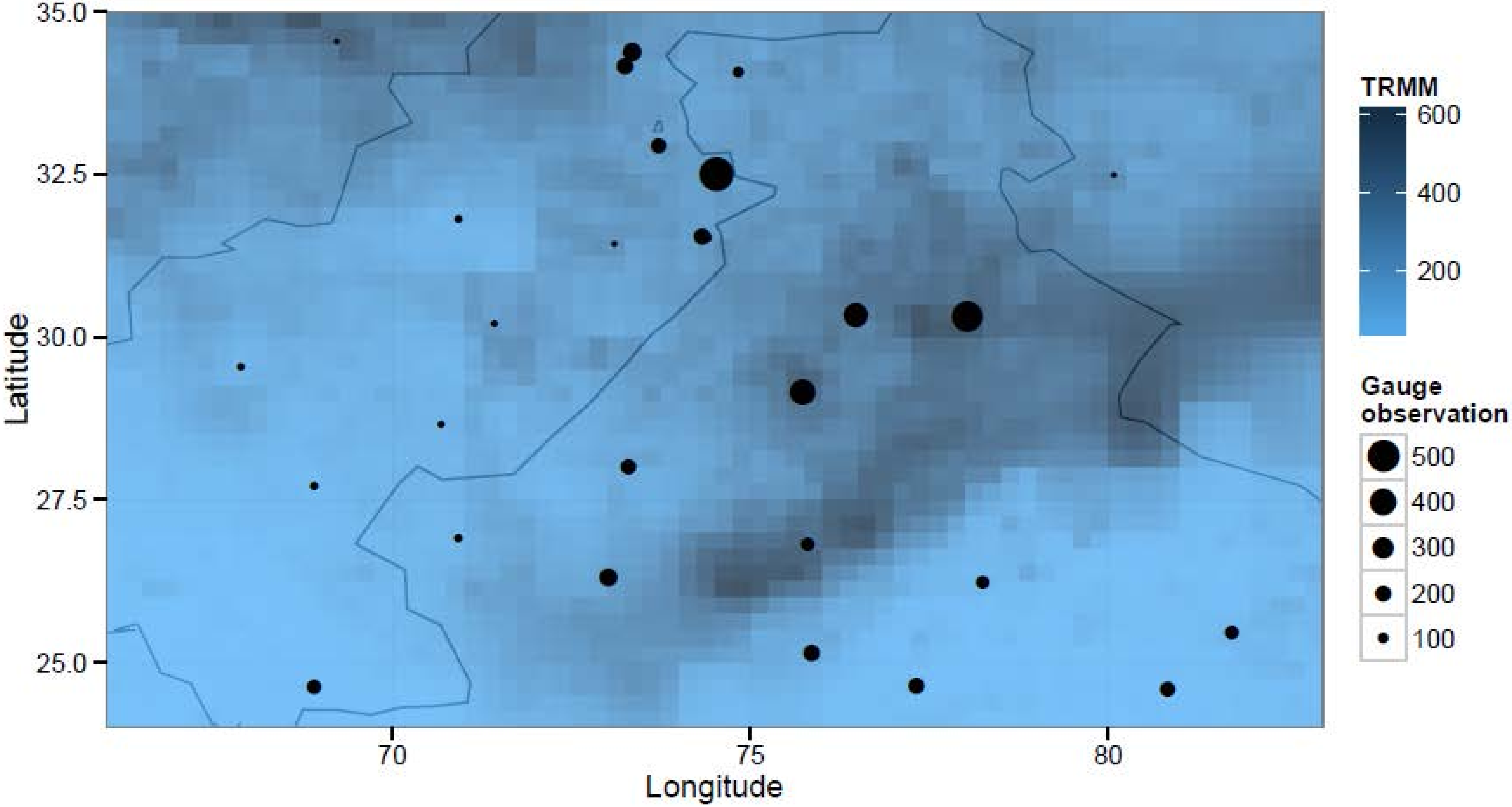
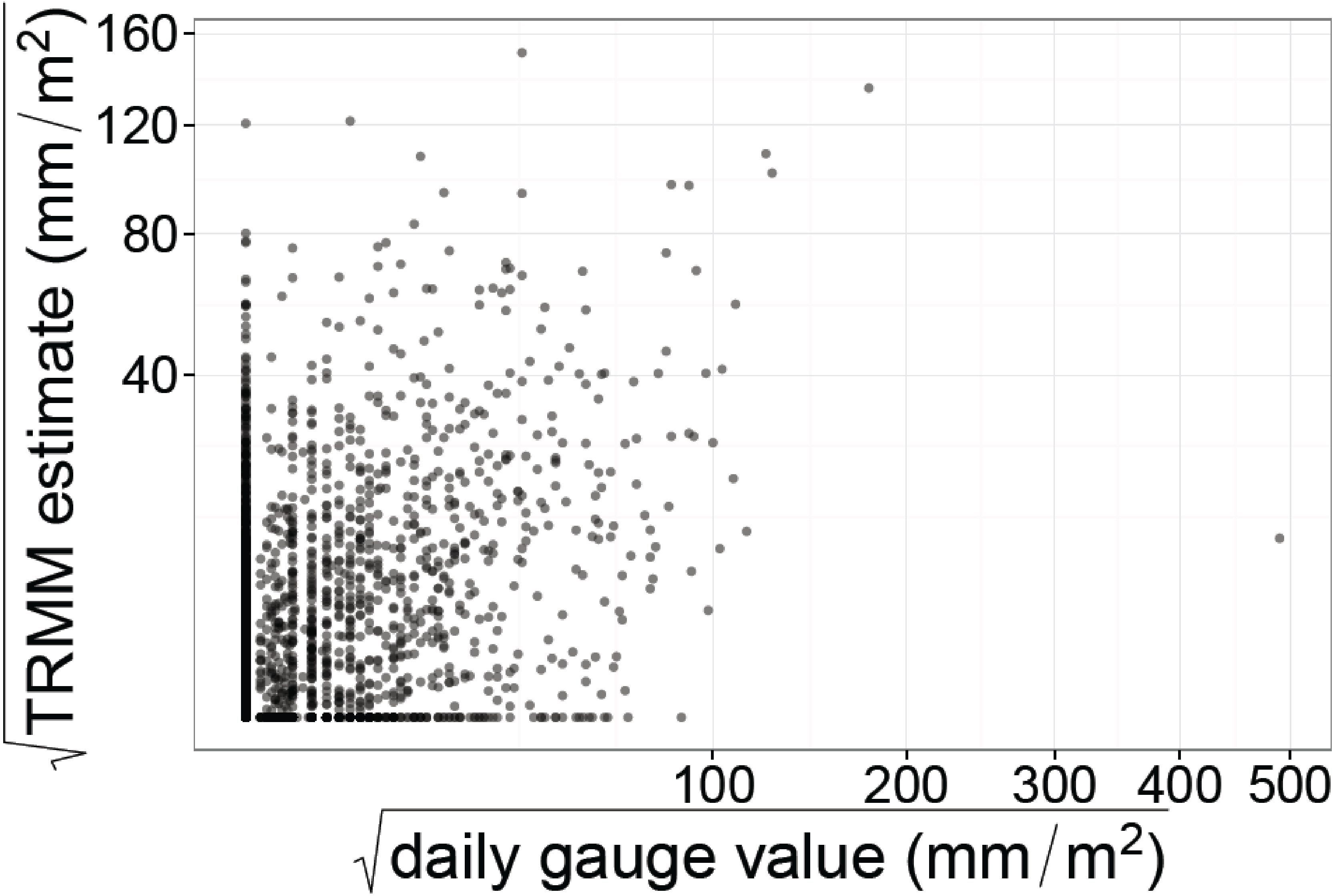


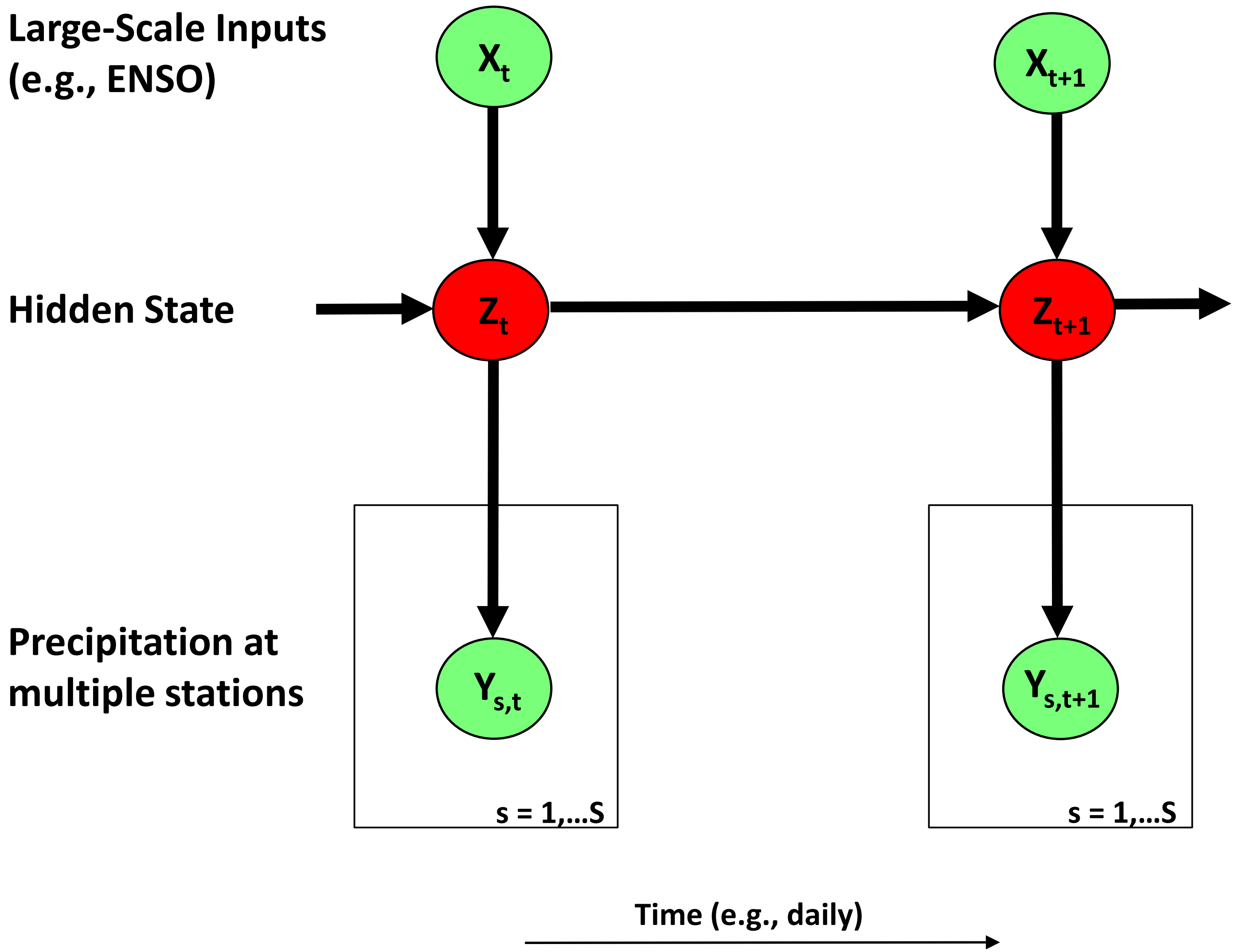
Figure from Lisa Goddard, International Research Institute for Climate and Society



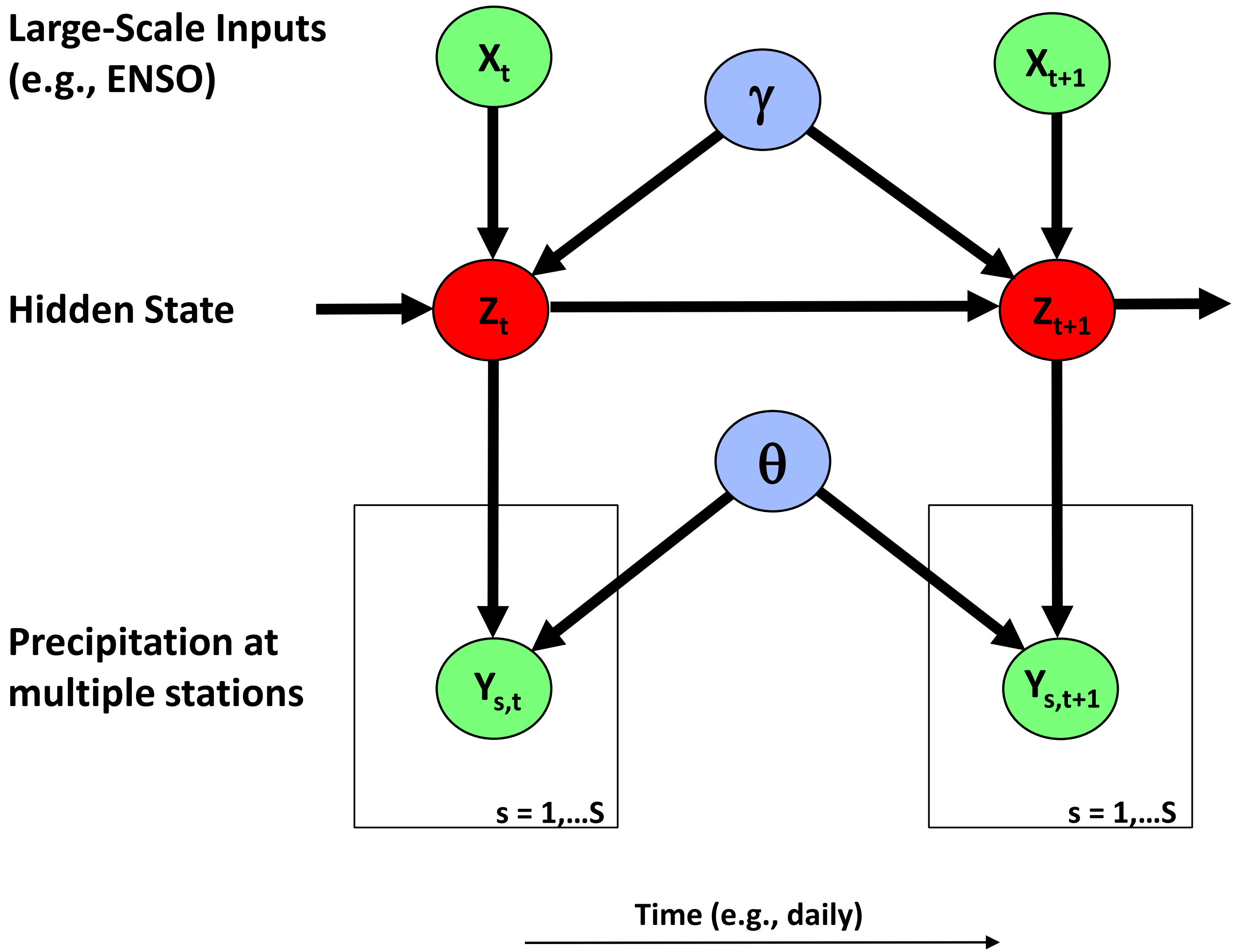
35.0
32.5
30.0
27.5
25.0
Latitude
70
75
80
Longitude
TRMM
600
400
200
Gauge observation
500
400
300
200
100


160
120
80
40
100
200
300
400
500
√TRMM estimate (mm/m²)
√daily gauge value (mm/m²)

Non-Homogeneous Hidden Markov Model for Downscaling



…with Parameters



Non-Homogeneous Hidden Markov Model

- **Transition probabilities vary as a function of exogenous variable x**
  - x = ENSO time-series, wind-shear index, etc
  - Transition probability vector modeled as multinomial logistic function of X

$$P(z_t = j | z_{t-1} = i, \boldsymbol{x}_t, \boldsymbol{\zeta}) = \frac{\exp(\xi_{ij} + \boldsymbol{x}_t \boldsymbol{\rho}_i)}{\sum_{m=1}^{K} \exp(\xi_{im} + \boldsymbol{x}_t \boldsymbol{\rho}_m)}, \qquad 1 \le i, j \le K$$

  - Sampler uses Polya-Gamma latent variable methods (Polson et al., JASA, 2013)

Non-Homogeneous Hidden Markov Model

- **Transition probabilities vary as a function of exogenous variable x**
  - x = ENSO time-series, wind-shear index, etc
  - Transition probability vector modeled as multinomial logistic function of X

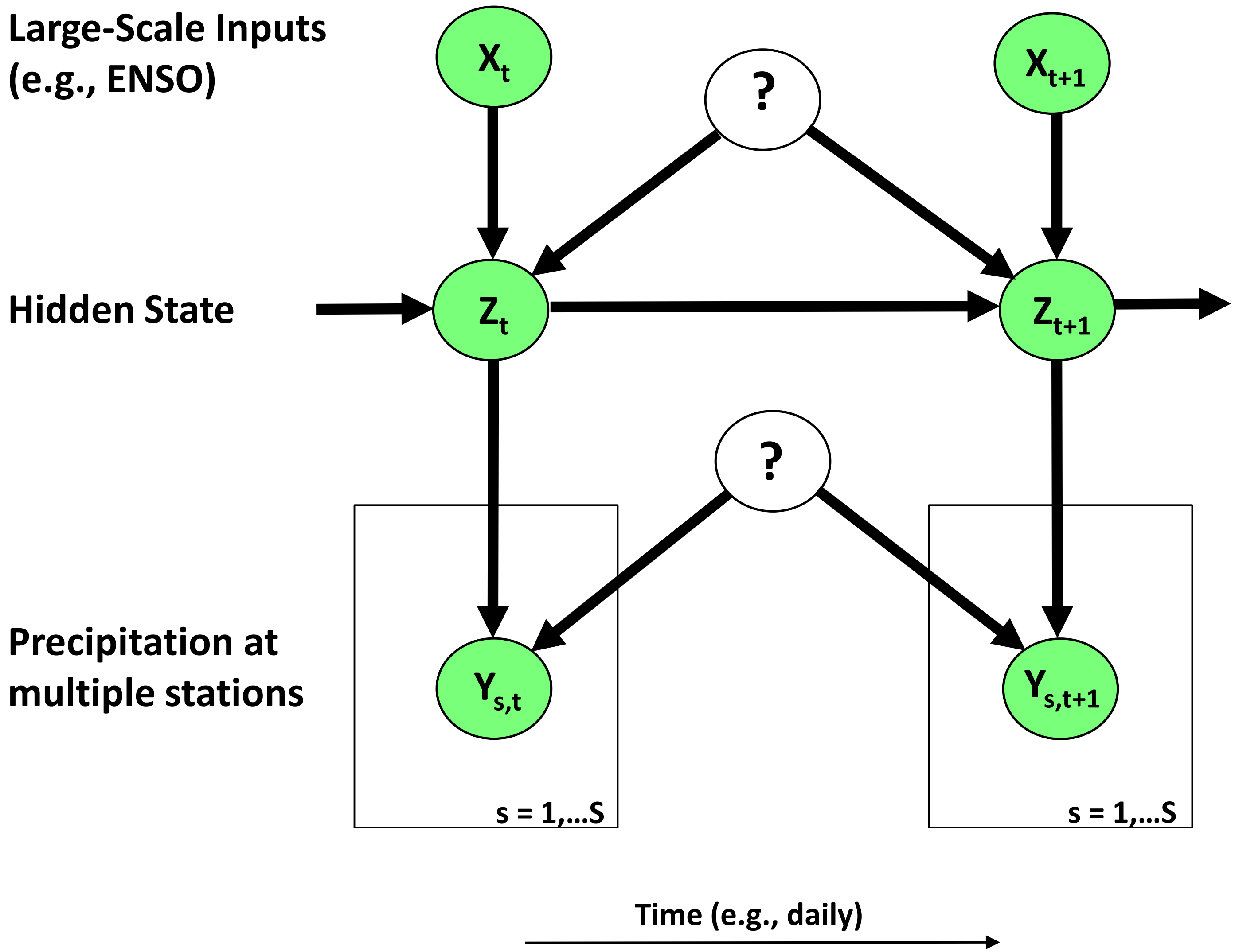
$$P(z_t = j | z_{t-1} = i, \boldsymbol{x}_t, \zeta) = \frac{\exp(\xi_{ij} + \boldsymbol{x}_t \boldsymbol{\rho}_i)}{\sum_{m=1}^{K} \exp(\xi_{im} + \boldsymbol{x}_t \boldsymbol{\rho}_m)}, \qquad 1 \leq i, j \leq K$$

  - Sampler uses Polya-Gamma latent variable methods (Polson et al., JASA, 2013)

- **Emission distributions**
  - Mixture of delta function (no precipitation) + 2 exponentials
  - Conditionally independent, station-specific
  - Mixing weights are functions of additional exogenous variables (via GLM-probit)

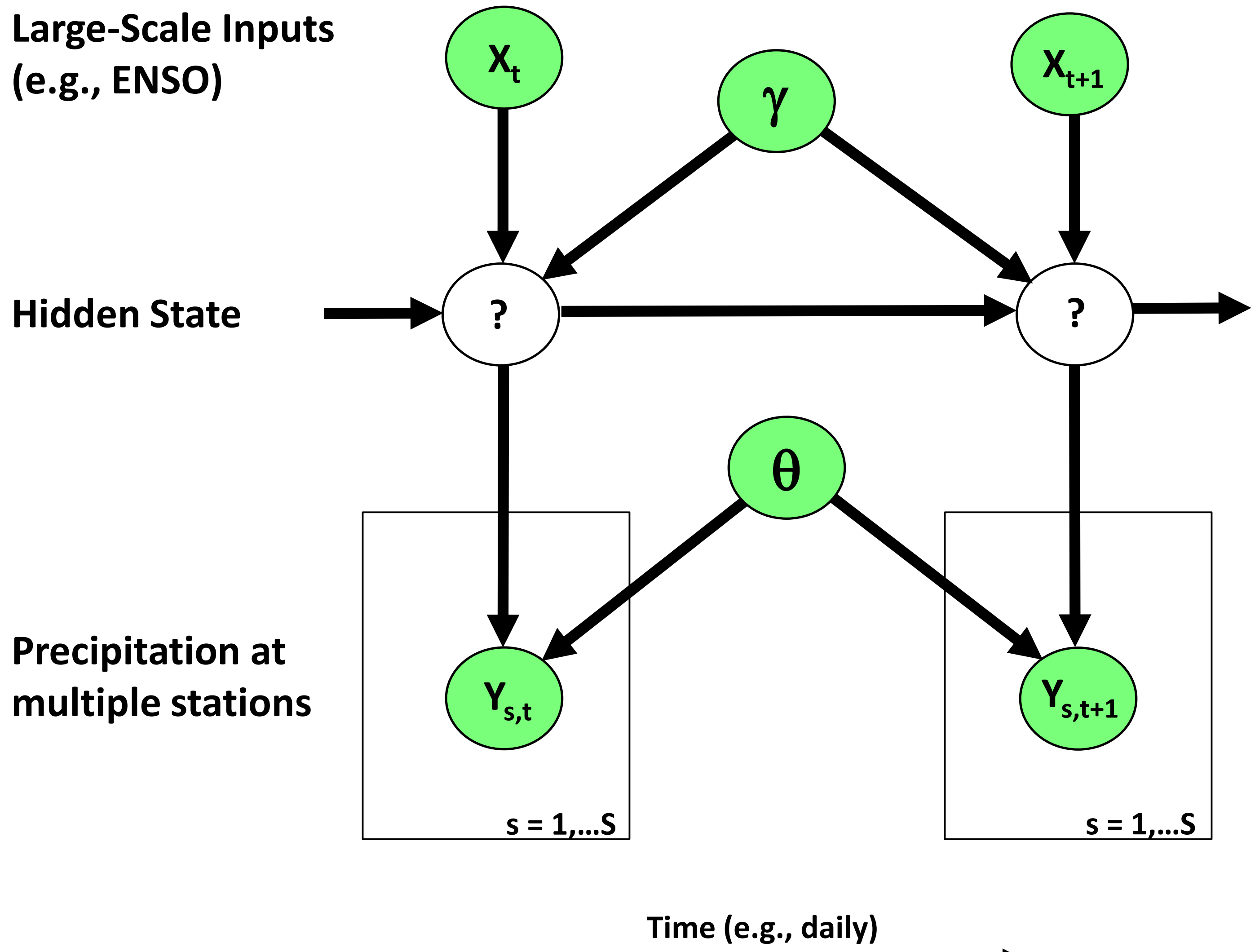
MCMC: Sample Parameters given Hidden States

(R software package *NHMM*, 2014)



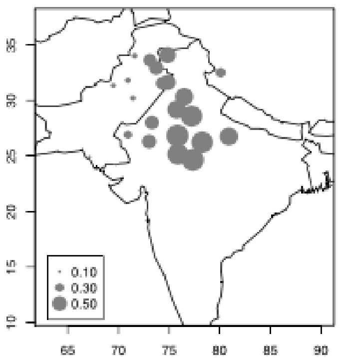
MCMC: Sample Hidden States given Parameters

(R software package *NHMM*, 2014)

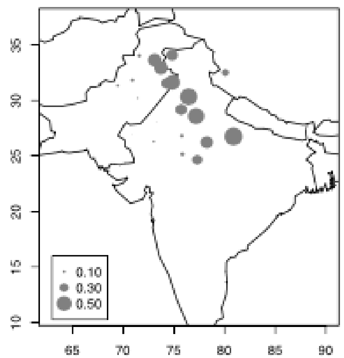


Examples of Inferred States

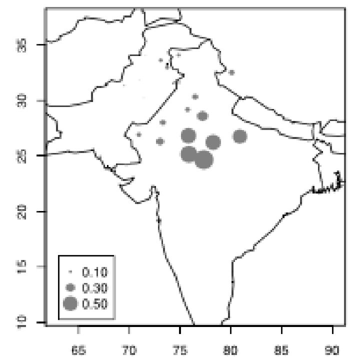
Probability of Occurrence



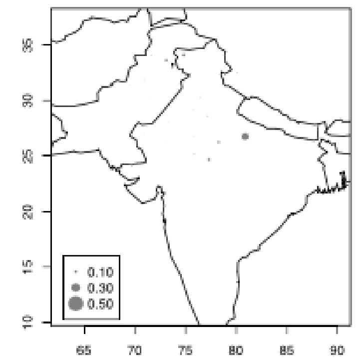
(a) State 1



(b) State 2



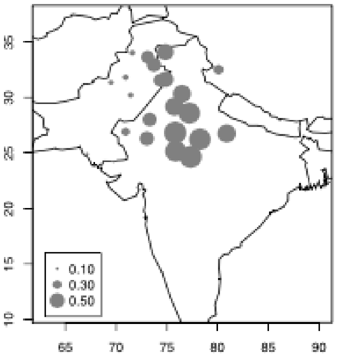
(c) State 3



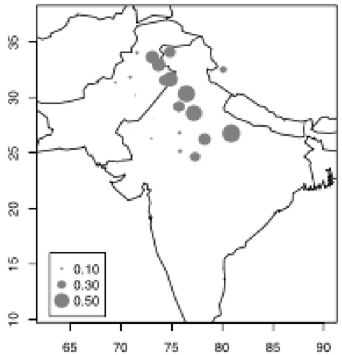
(d) State 4

Examples of Inferred States

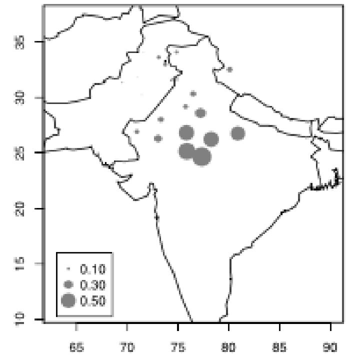
Probability of Occurrence



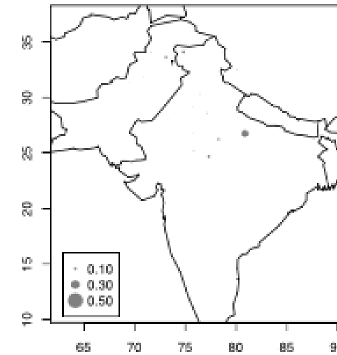
(a) State 1



(b) State 2

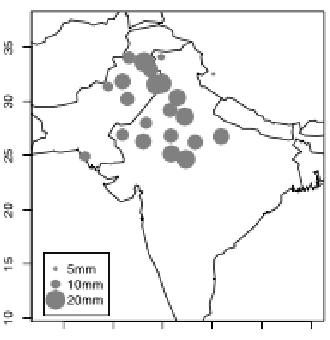


(c) State 3

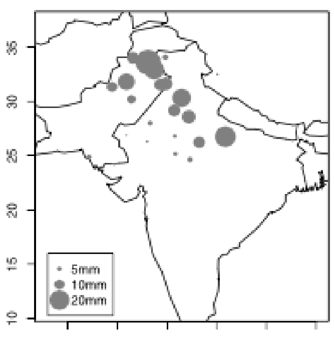


(d) State 4

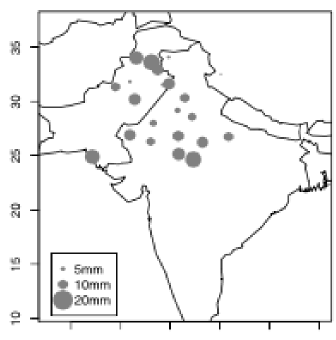
Mean Intensity



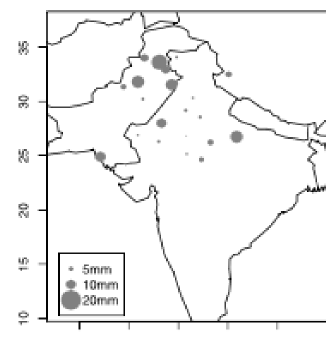
(e) State 1



(f) State 2



(g) State 3

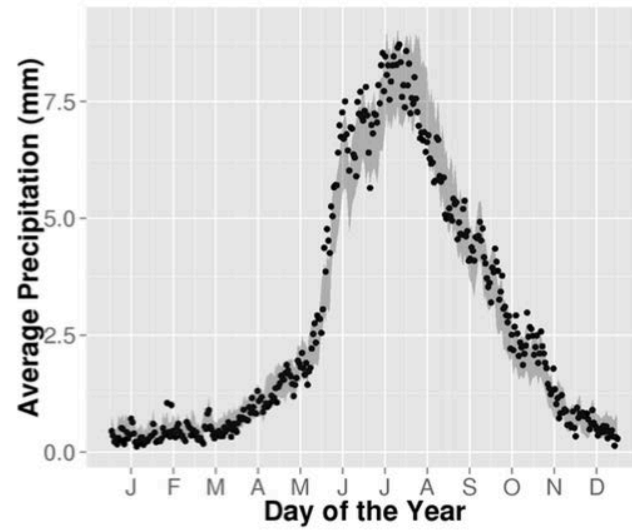


(h) State 4

Simulated and Actual Precipitation

Black dots = observed rainfall

Grey bands = 95% posterior intervals
from 1000 model simulations



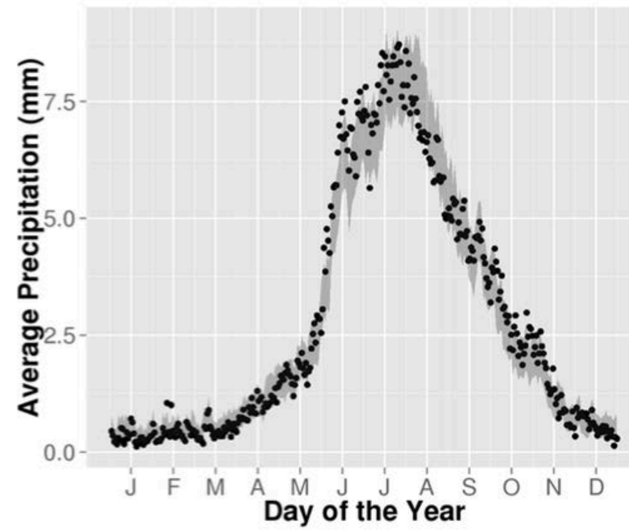
(a) Rainfall

UCIrvine
University of California, Irvine

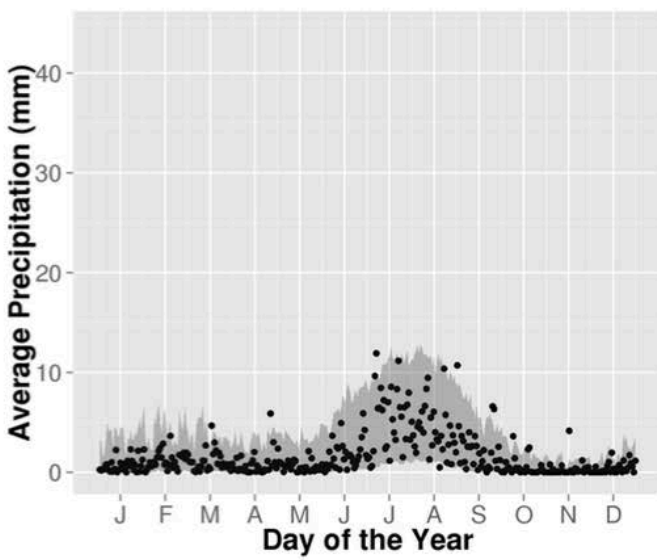
Simulated and Actual Precipitation

Black dots = observed rainfall

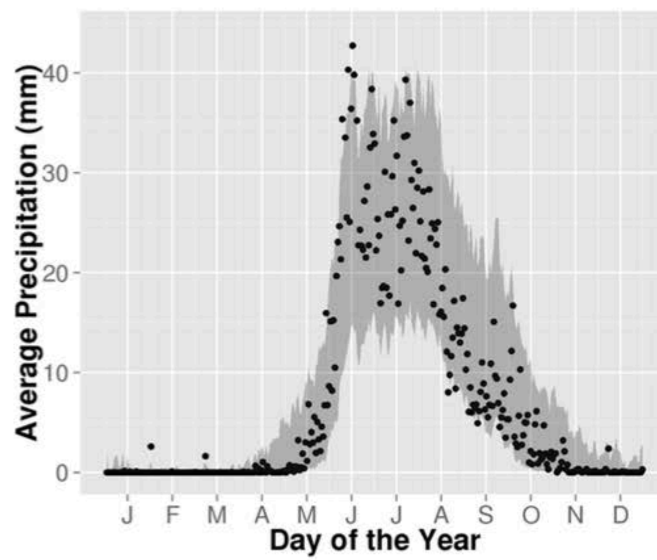
Grey bands = 95% posterior intervals from 1000 model simulations



(a) Rainfall



(b) Station 3: Rainfall

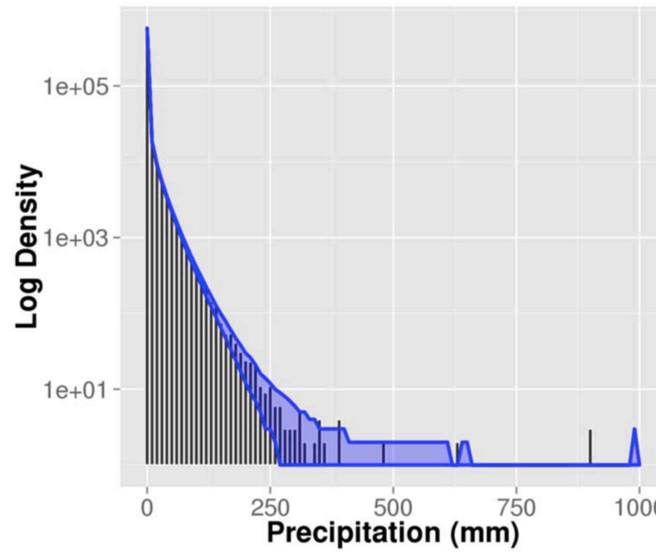


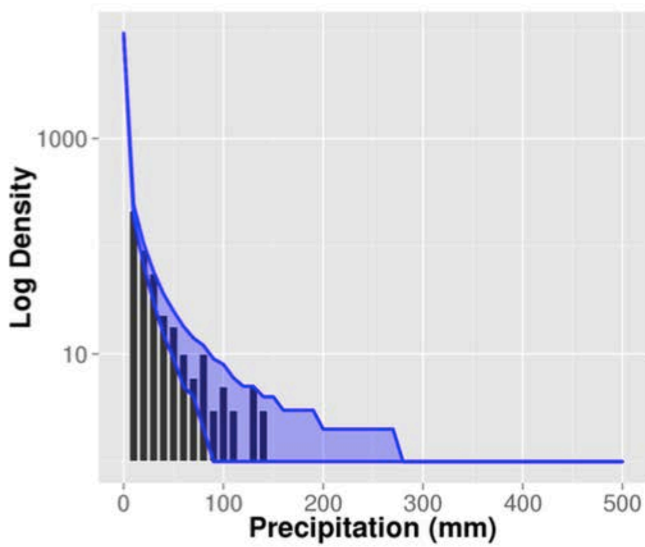
(c) Station 40: Rainfall

Distributional Checks

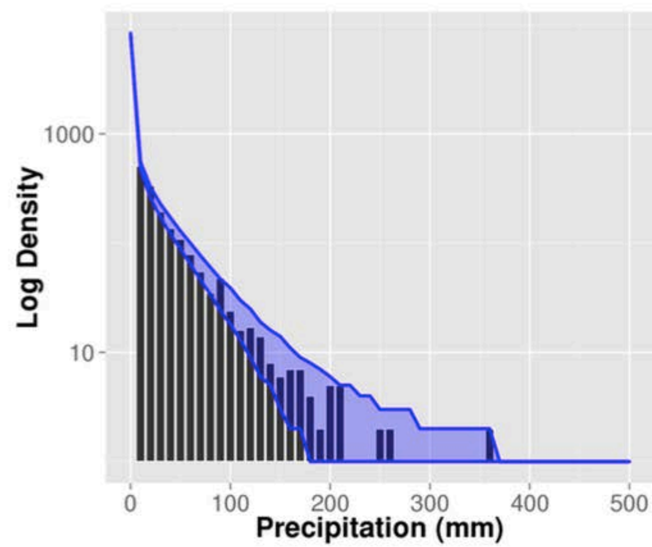
Bars = observed rainfall

Purple bands = 95% posterior intervals
from 1000 model simulations



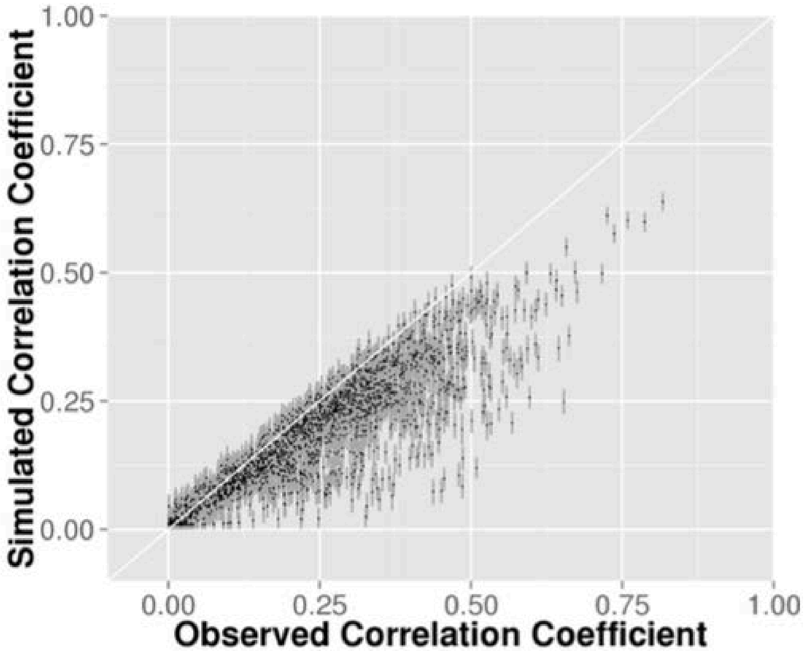


(b) Station 3 Precip.

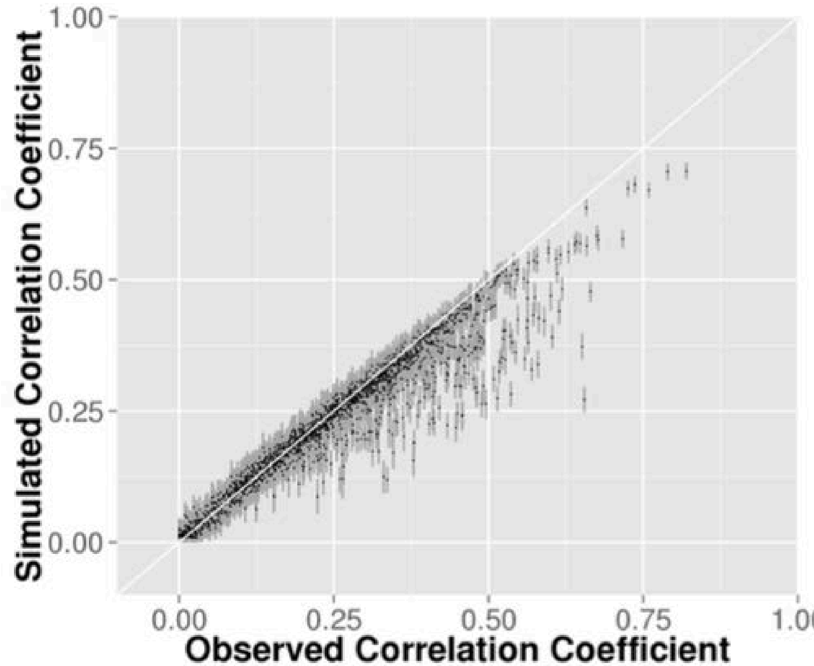


(c) Station 40 Precip.

Observed and Simulated Spatial Correlations



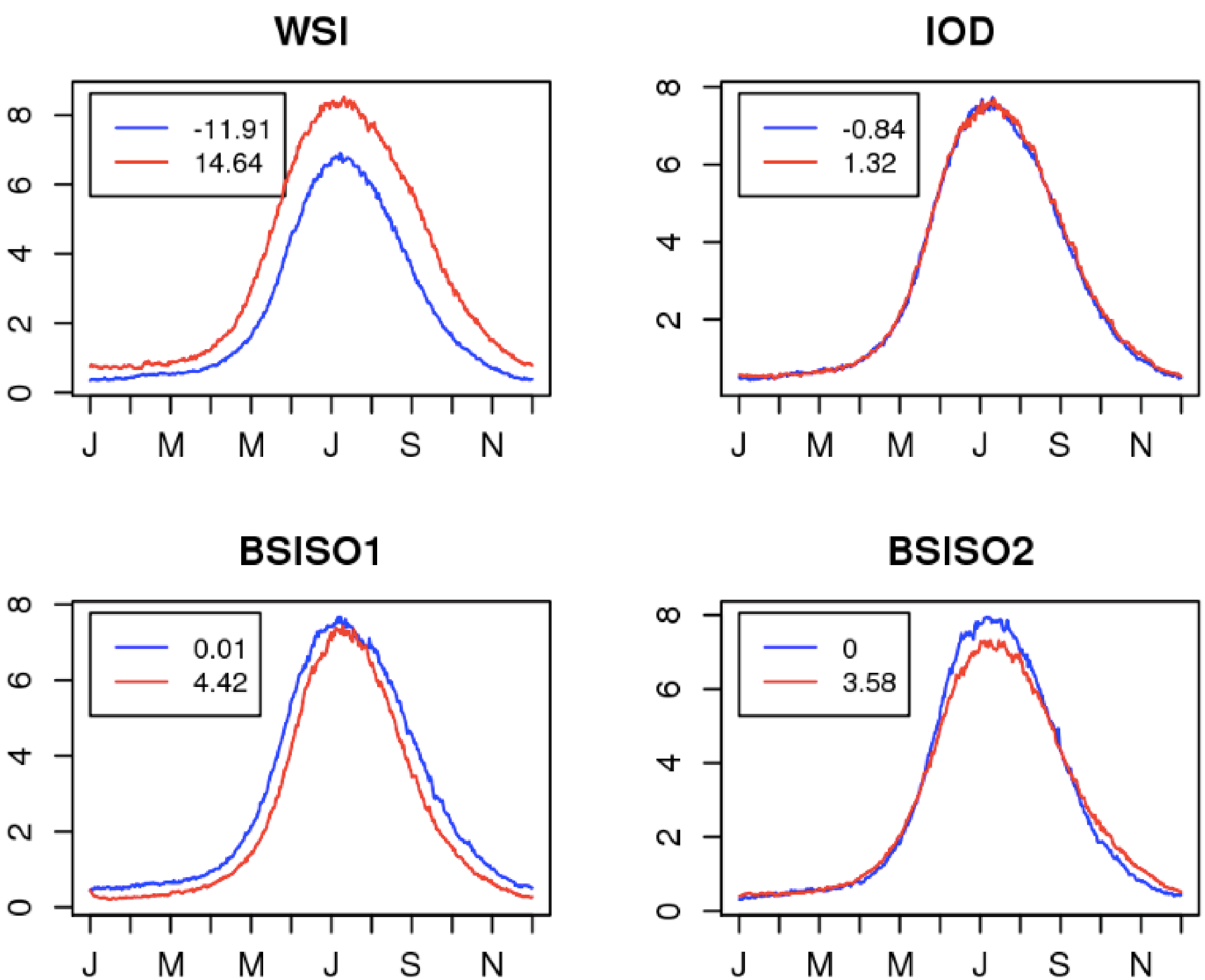
(a) GLM: Precip. amount



(c) mNHMM: Precip. amount

Effect of Input Variables

Red: simulations with variable's max value
Blue: simulations with variable's min value



Concluding Comments

Recent Trends in Machine Learning: Scalability

- **Computation-intensive techniques such as MCMC were limited to relatively small data sets in the past**
  - This is no longer the case….

- **Techniques for improving scalability**
  - Distributed sampling algorithms
  - Approximate sampling methods
    - Variational inference
    - Operating on data subsamples
  - Combining optimization and sampling

- **Focus for climate scientists should be on model structure**
  - Inference should just be a "black box", turn the crank…..

Recent Trends in Machine Learning: Deep Learning

- **Deep learning**
  - neural networks with multiple hidden layers designed for high-dimensional prediction problems
  - Have recently been very successful in image and speech recognition tasks

- **Useful for climate science? There are limitations…**
  - Typically very large amounts of labeled (classification) data
  - Models can lack interpretability…very much “black box prediction”

- **However….**
  - Significant research underway on developing unsupervised/latent variants
  - Useful as feature extractors/detectors? whose outputs are integrated into spatio-temporal models with probabilistic semantics

Issues in Applying Machine Learning to Climate Science

- **Emphasis on black-box predictive modeling**
  - Dominance of applications where retraining the model is cheap and easy

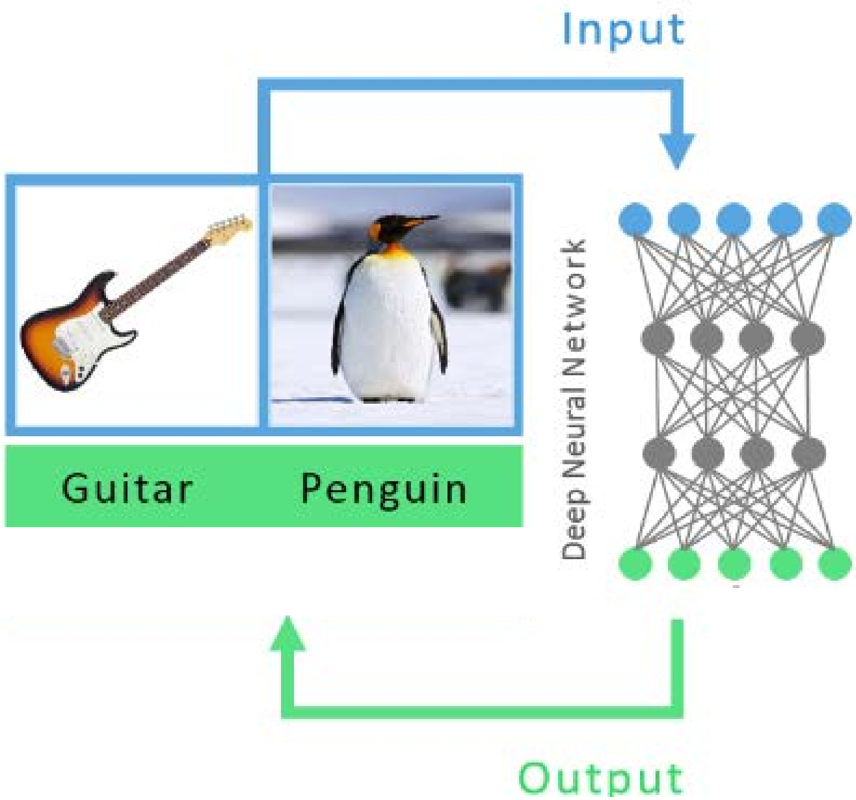
- **Widespread stationarity assumptions**
  - Test data expected to be the same as training data

- **The tyrany of tabula-rasa learning**
  - Using domain knowledge is not "cool"
  - e.g., lack of knowledge of spatial topography in models

- **Calibration**
  - Models don't know what they don't know

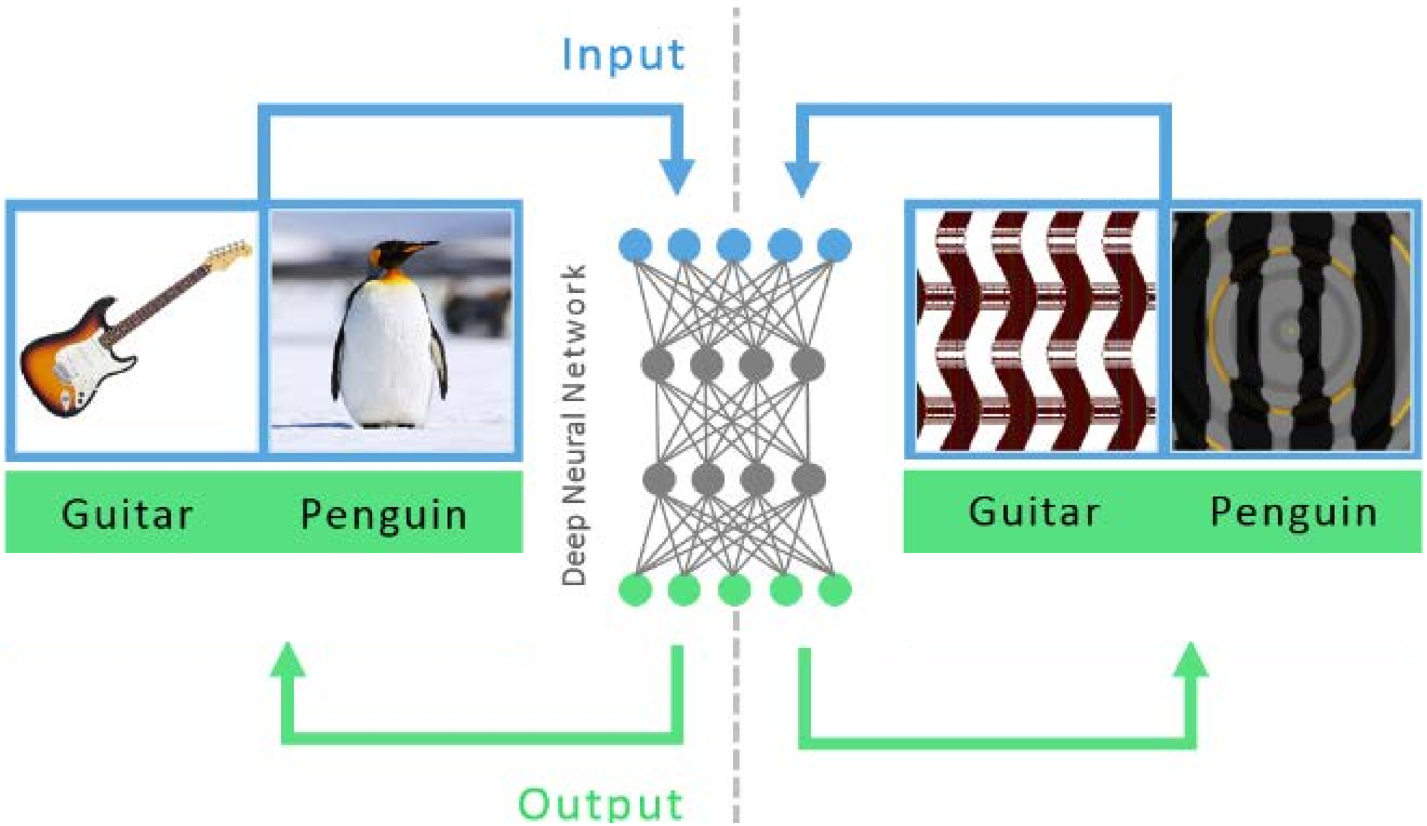
A Deep Neural Network for Image Recognition

From Nguyen, Yosinski, Clune, ArXiv preprint, 2014



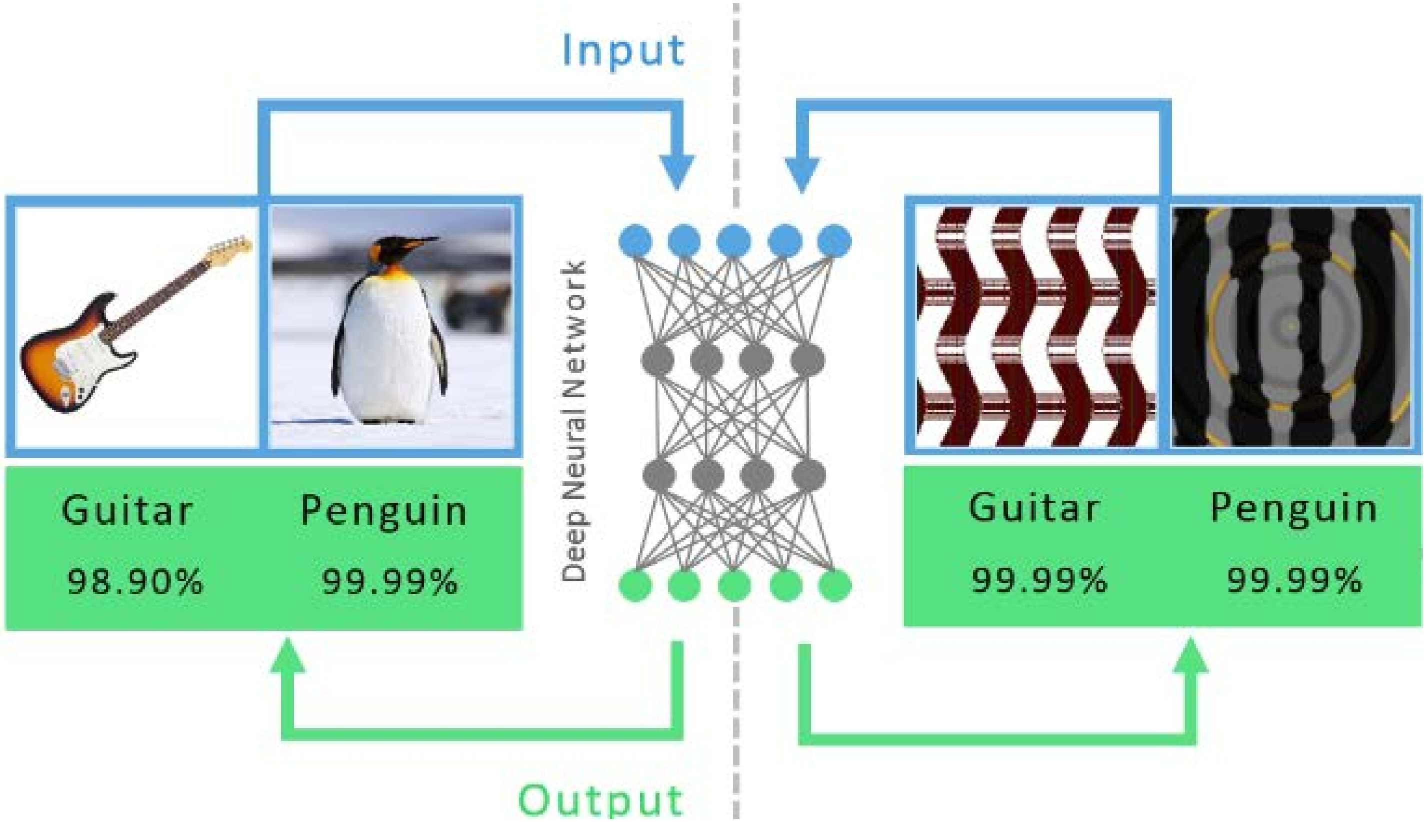
Extrapolation

From Nguyen, Yosinski, Clune, ArXiv preprint, 2014



Lack of Calibration

From Nguyen, Yosinski, Clune, ArXiv preprint, 2014



Summary

- Graphical models provide a broad framework for building complex statistical models
- The machine learning perspective on graphical models puts more emphasis on prediction and scalability (compared to statistics)
- For many climate science problems, graphical models (and Bayesian inference) tend to be a more natural fit than “black box” prediction models
- Advances in machine learning are promising, but its not clear yet what benefit they will provide to climate science

Acknowledgements

Colleagues and Students

Caroline Bain, Daniel Henke, Michael Ghil, Arthur Greene, Tracy Holsclaw, Scott Gaffney, Sergey Kirshner, Gudrun Magnusdottir, Andy Robertson









Funding

National Science Foundation
WHERE DISCOVERIES BEGIN

U.S. DEPARTMENT OF
ENERGY