HMM for precipitation

Pierre Ailliot

National Institute of Water and Atmospheric Research
Victoria University of Wellington (New Zealand)

Peter Thomson

Statistics Research Associates Ltd

Outline

- Basic HMM
  - Model description
  - Limitations
- Extensions
  - HMM with truncated Gaussian distributions
- Conclusion and perspectives

Basic HMM

Model description

- First proposed by Zucchini & Guttorp (1991)
  - Generalized to a nonhomogeneous HMM
    - Hughes et al. (1994)
- <u>Observed process:</u> $Y_t = (Y_t(1),\ldots,Y_t(K))$
  - $Y_t(k) \in R^+$: rainfall during day t at location k
- Existence of "weather type"
  - High pressure systems, frontal systems, …
- …Introduced as a <u>hidden process</u> $S_t \in \{1\ldots Q\}$
  - Common to all locations

Basic HMM

Model description

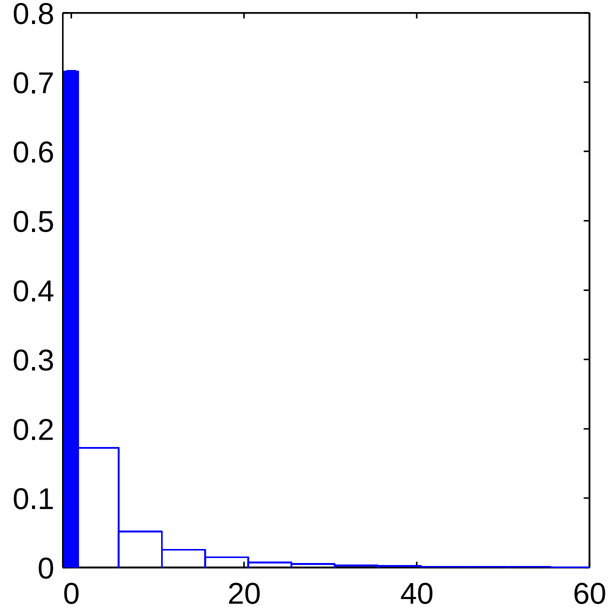
- Conditional independence assumptions
  - Temporal structure (HMM)
    - $P(S_t|Y_1,\ldots,Y_{t-1}, S_1,\ldots,S_{t-1}) = P(S_t|S_{t-1})$
    - $P(Y_t|Y_1,\ldots,Y_{t-1}, S_1,\ldots,S_t) = P(Y_t|S_t)$
  - Dynamics induced only by $\{S_t\}$
  - Spatial structure (conditional independence)
    - $p(Y_t(1),Y_t(2)\ldots,Y_t(K)|S_t)$

      $= p(Y_t(1)|S_t)\, p(Y_t(2)|S_t)\ldots\, p(Y_t(K)|S_t)$
  - Spatial dependence induced only by $\{S_t\}$

Basic HMM

Model description

- Conditional distributions $p(Y_t(k)|S_t=s)$
  - Two components
    - $Y_t\ (k)=0$ if no rainfall occurs
    - $Y_t\ (k)>0$ if a rainfall occurs
  - Mixed discrete-continuous distribution



Marginal distribution of $\{Y_t\ (k)\}$
(Winchmore, New Zealand)

Basic HMM

Model description

- Conditional distributions $p(Y_t(k)|S_t=s)$

$$P(Y_t(k)\in dy|S_t=s)=\begin{cases}1-p_k^{(s)} & \text{if } 0\in dy\\ p_k^{(s)} f(y;\alpha_k^{(s)},\beta_k^{(s)})dy & \text{if } 0\notin dy\end{cases}$$

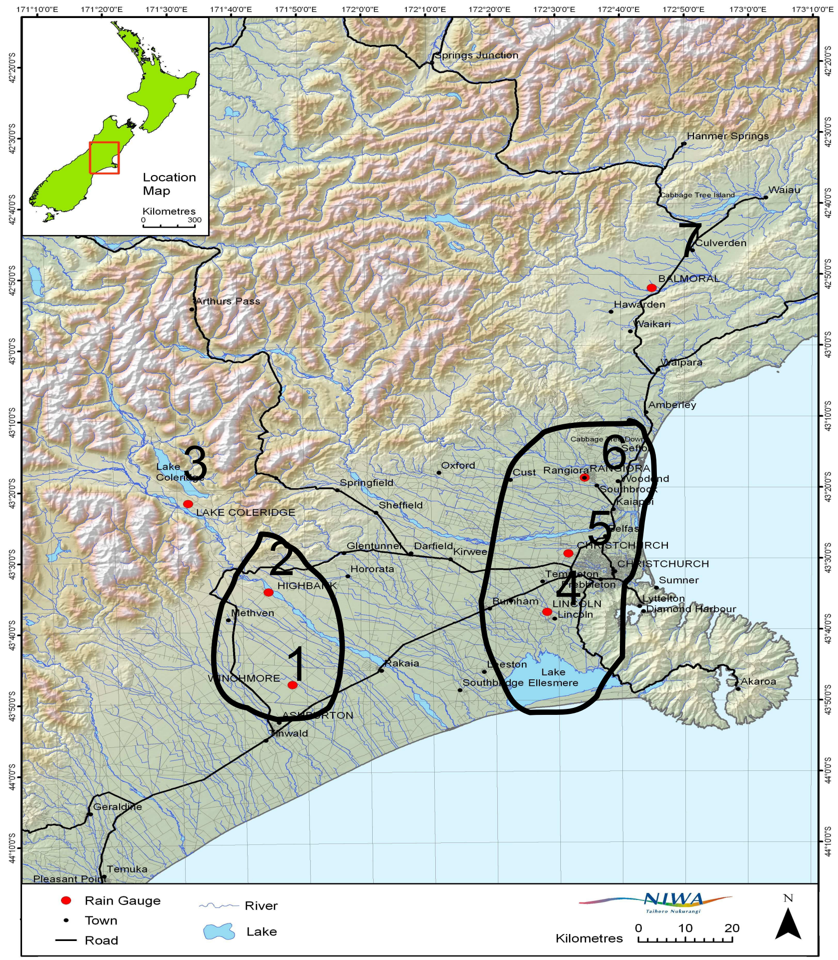
  - $p_k^{(s)}\in[0,1]$, $\alpha_k^{(s)}>0$, $\beta_k^{(s)}>0$
  - $f(y;\alpha,\beta)=\dfrac{1}{\Gamma(\alpha)\beta^{\alpha}}y^{\alpha-1}\exp(-\dfrac{y}{\beta})$

- $3KQ+Q(Q-1)$ parameters

Basic HMM

Data

- Rainfall data in New Zealand
  - Daily rainfall
  - 7 locations
  - 26 years
    - Focus on April



Basic HMM

Parameter estimation

- EM algorithm
- Model selection
  - First selection with AIC and BIC

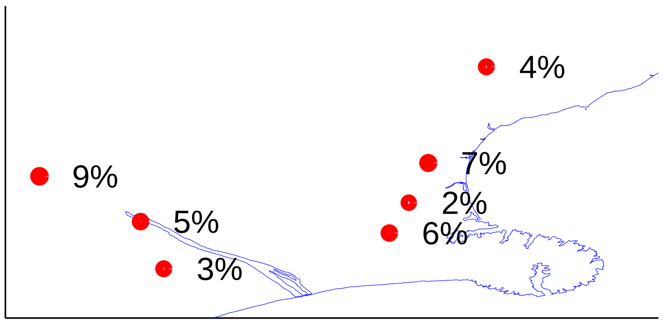
Q	1	2	3	4	5	6
AIC	17404	14317	13436	13213	13144	**12990**
BIC	17502	14523	13760	**13663**	13731	13722

  - Final selection according to
    - Meteorological interpretability
    - Ability to simulate realistic rainfalls
  - Focus on the model with $Q=4$

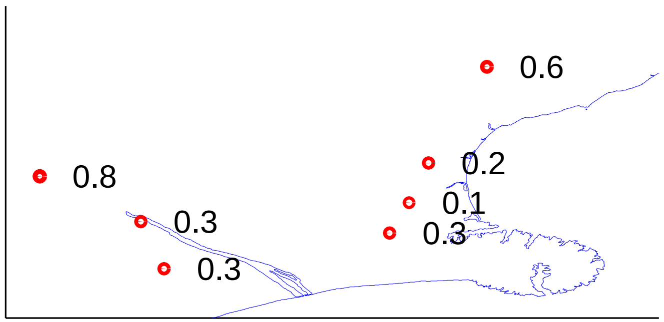
Basic HMM

Meteorological interpretability

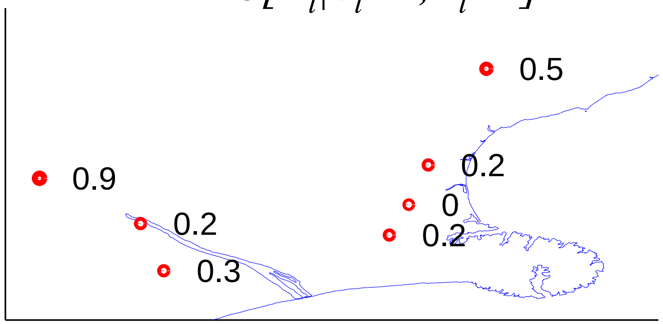
$P[Y_t>0|S_t=1]$



$E[Y_t|S_t=1,Y_t>0]$



$\sigma[Y_t|S_t=1,Y_t>0]$



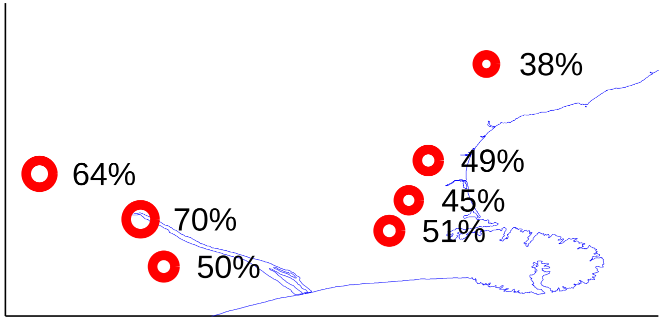
<u>Regime 1</u>

- Low probability of rainfall occurrence
- Low amount, higher at location 3

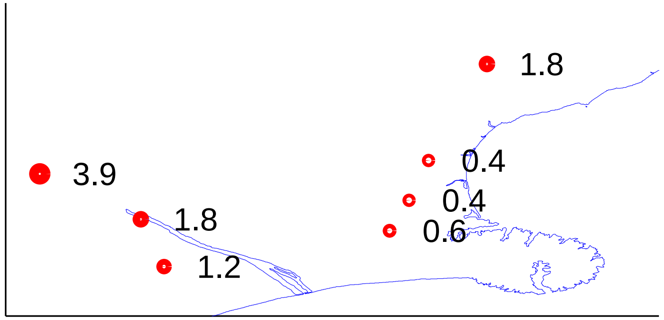
Basic HMM

Meteorological interpretability

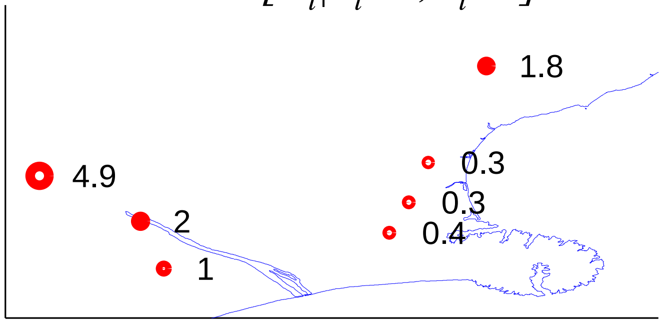
$P[Y_t>0|S_t=2]$



$E[Y_t|S_t=2,Y_t>0]$



$\sigma[Y_t|S_t=2,Y_t>0]$

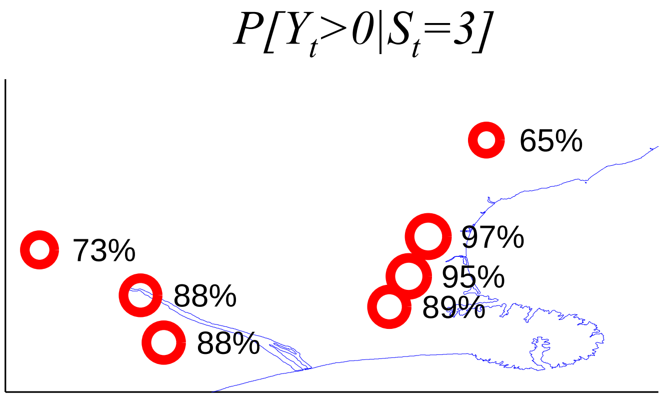


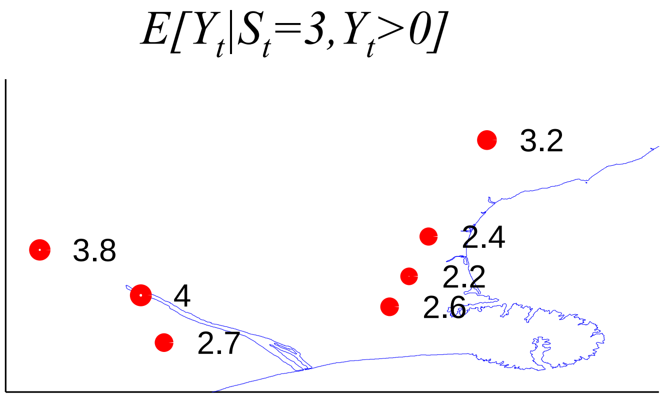
<u>Regime 2</u>

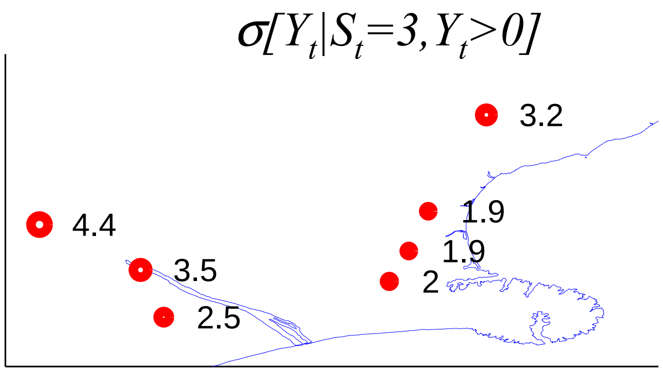
- Moderate probability of rainfall occurrence, higher in the west part
- Moderate amounts in the west part, low in the south-east

Basic HMM

Meteorological interpretability







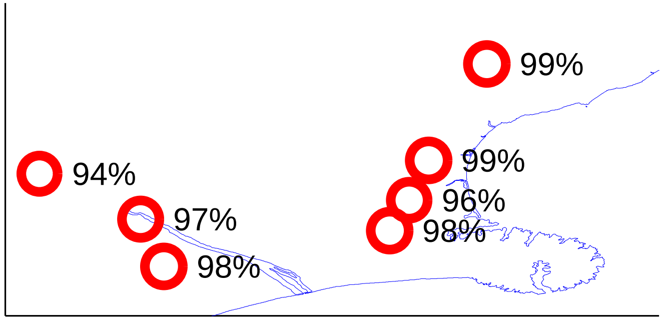
<u>Regime 3</u>

- High probability of rainfall occurrence, higher in the south-east
- Moderate amounts, lower in the south-east

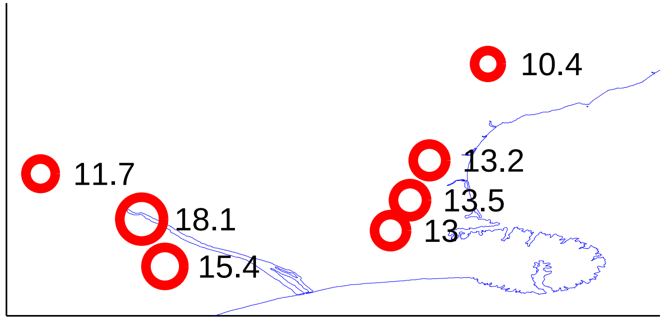
Basic HMM

Meteorological interpretability

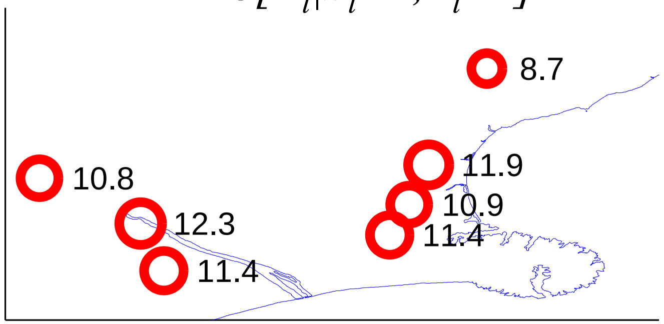
$P[Y_t>0|S_t=4]$



$E[Y_t|S_t=4,Y_t>0]$



$\sigma[Y_t|S_t=4,Y_t>0]$



<u>Regime 4</u>

- High probability of rainfall occurrence
- High amounts

Basic HMM

Meteorological interpretability

- Transition matrix, stationary distribution, mean durations

0.70	0.15	0.09	0.05
0.49	0.18	0.20	0.12
0.35	0.31	0.17	0.16
0.21	0.29	0.25	0.25

0.56
0.20
0.14
0.10

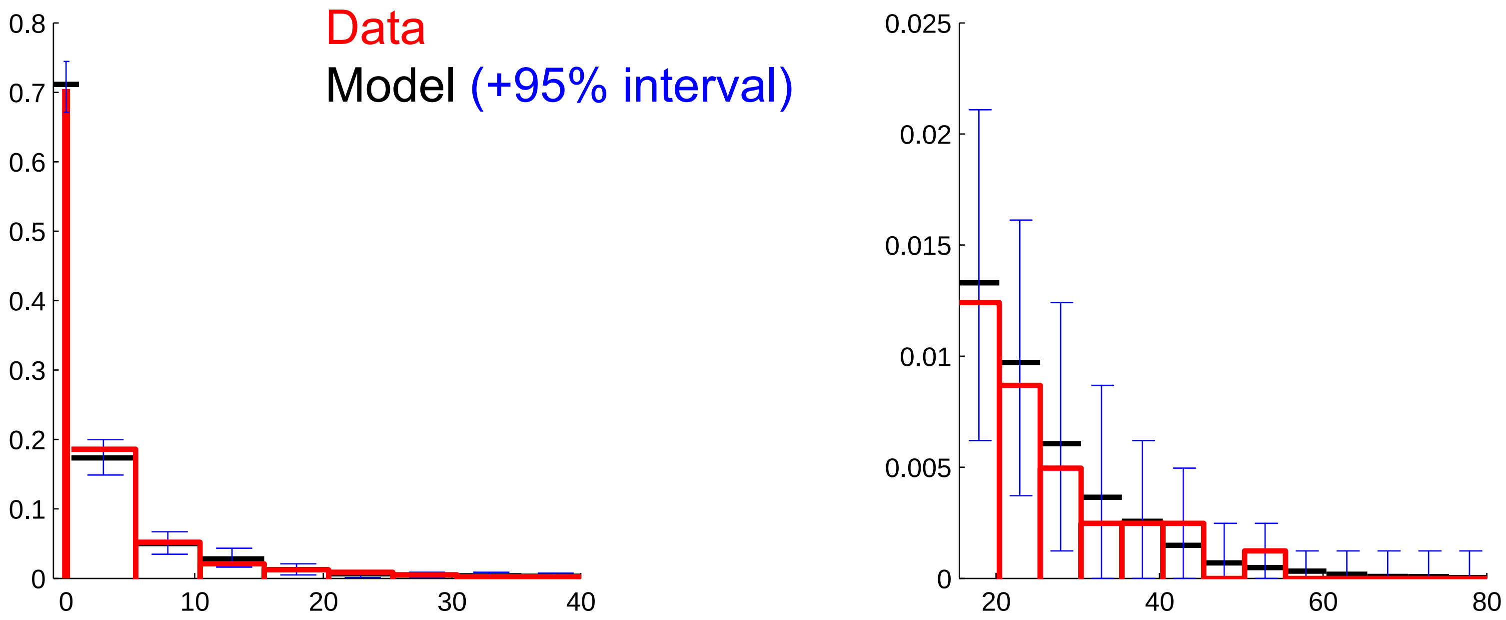
3.33
1.22
1.20
1.33

- Summary:
  - **Regime 1:** dry conditions, long persistence
  - **Regime 2 and 3:** intermediate patterns, regional differences, higher rainfall in regime 3, short persistence
  - **Regime 4:** heavy rainfall
- Similar meteorological interpretation for other datasets

Basic HMM

Realism of simulated sequences

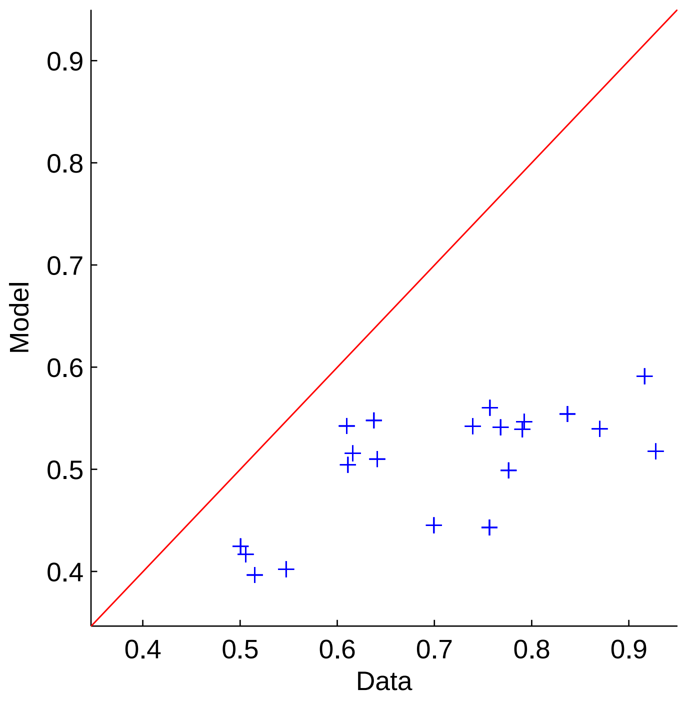
- Univariate marginal distributions (location 1, Winchmore)



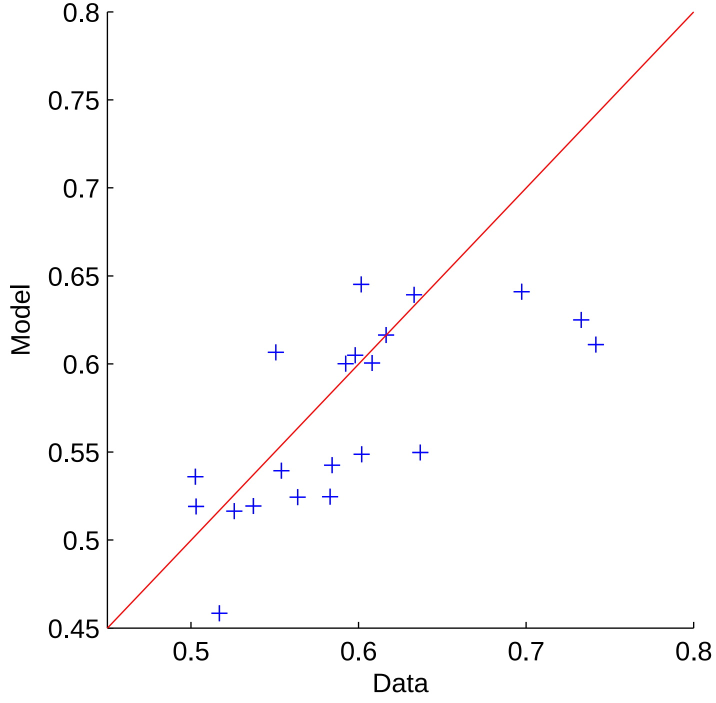
Basic HMM

Realism of simulated sequences

- Spatial pair-wise correlations



Amounts

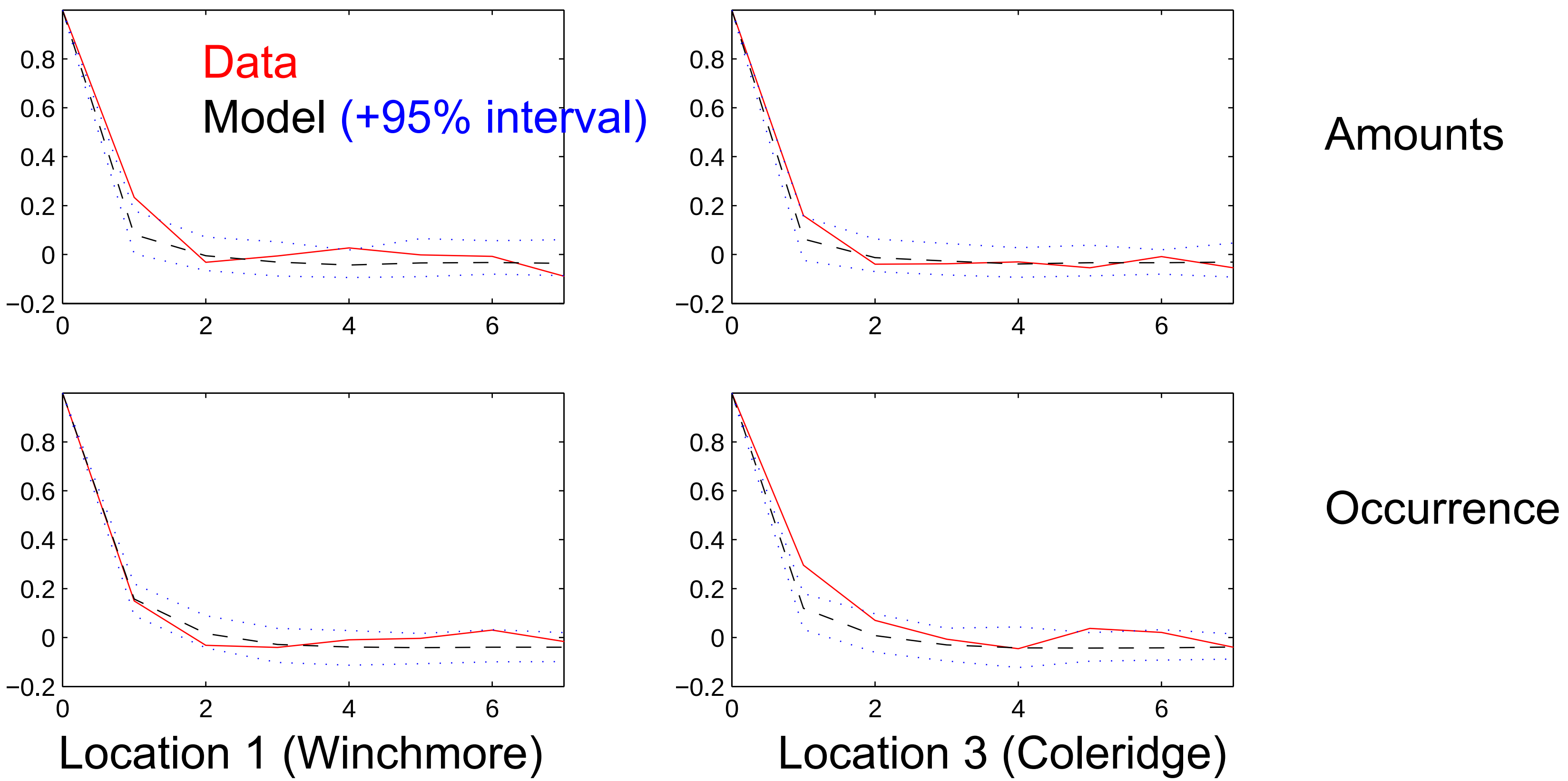


Occurrence

Basic HMM

Realism of simulated sequences

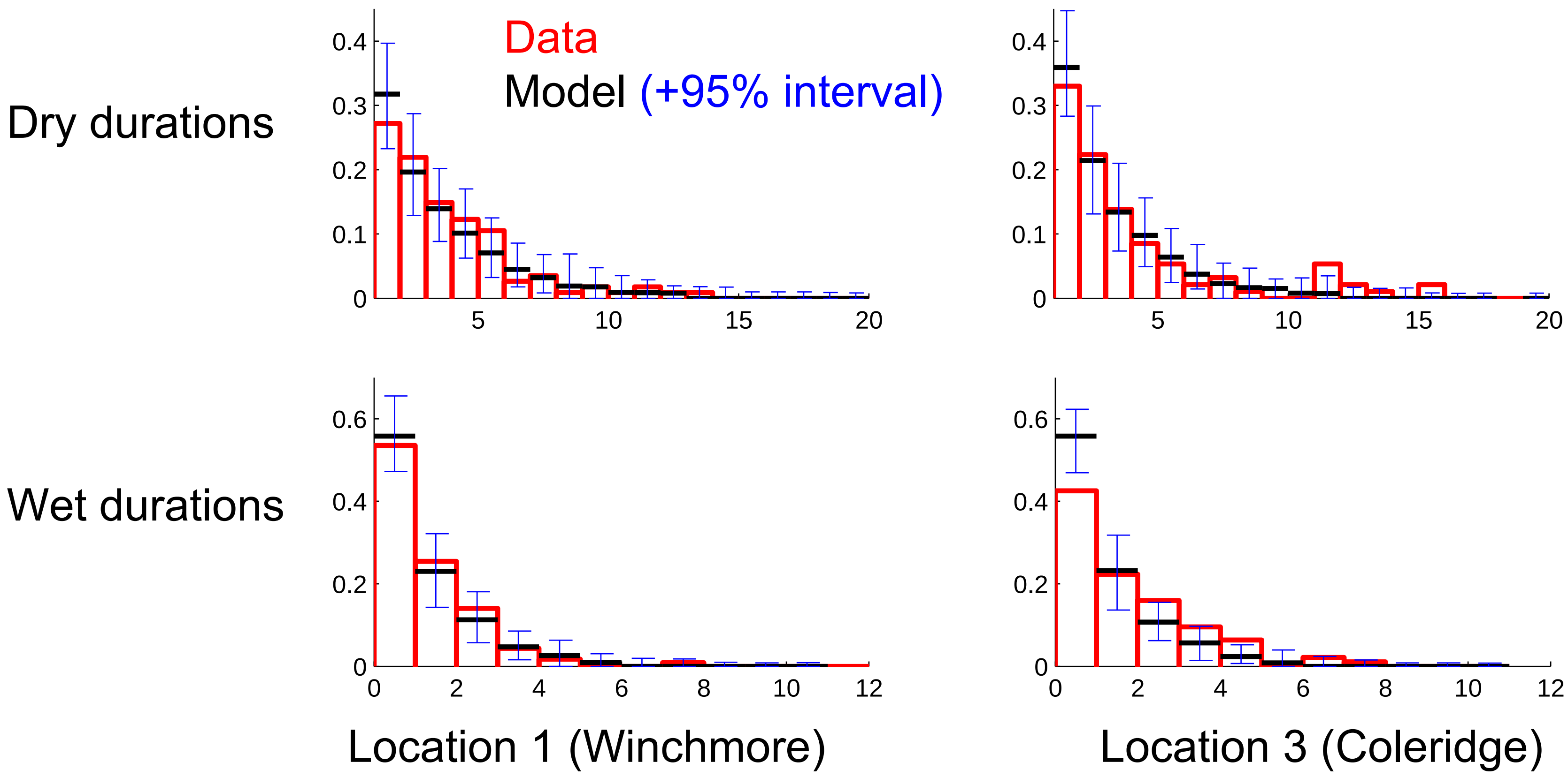
- Autocorrelation functions



Basic HMM

Realism of simulated sequences

- Dry/wet durations



Basic HMM

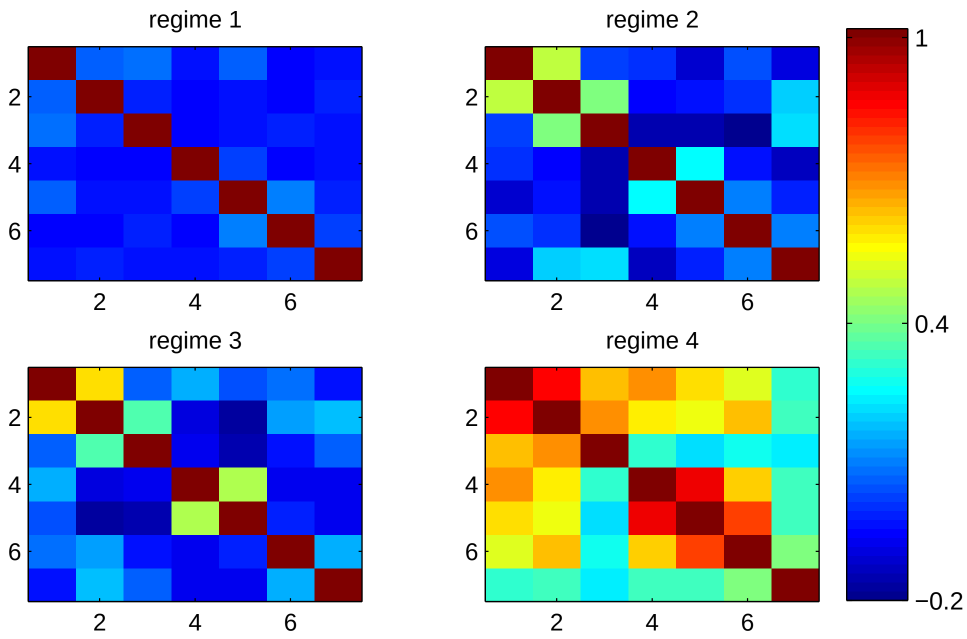
Conclusion

- ☺ Meteorological interpretation
- ☺ Reproduces marginal distribution
- ☹ Fails to reproduce…
  - Spatial structure
  - Dynamics at some locations
- …Need for a more sophisticated model
  - Focus on spatial aspects in the next part

Extensions

- Conditional independence assumption unrealistic

Empirical correlation matrices in the different weather types (identified via the Viterbi algo.)



➤ Residual spatial structure within the weather types

Extensions

- Add spatial structure in the emission probabilities
  - Need model for multivariate mixed discrete-continuous data
- **Markov random fields**
  - For the binary occurrence/non-occurrence process
    - Autologistic model (Hughes et al. (1999))
    - Chow-Liu trees (Kirshner (2005))
  - Generalization to include positive amounts?
- Introducing a **“local” weather type** to relate regional patterns to local rainfall
  - Thompson et al. (2005)
- **Truncated Gaussian random fields**
  - Allcroft et al. (2003) (without Markov switching)

HMM with truncated Gaussian fields

Model description

- If $S_t = s$ then

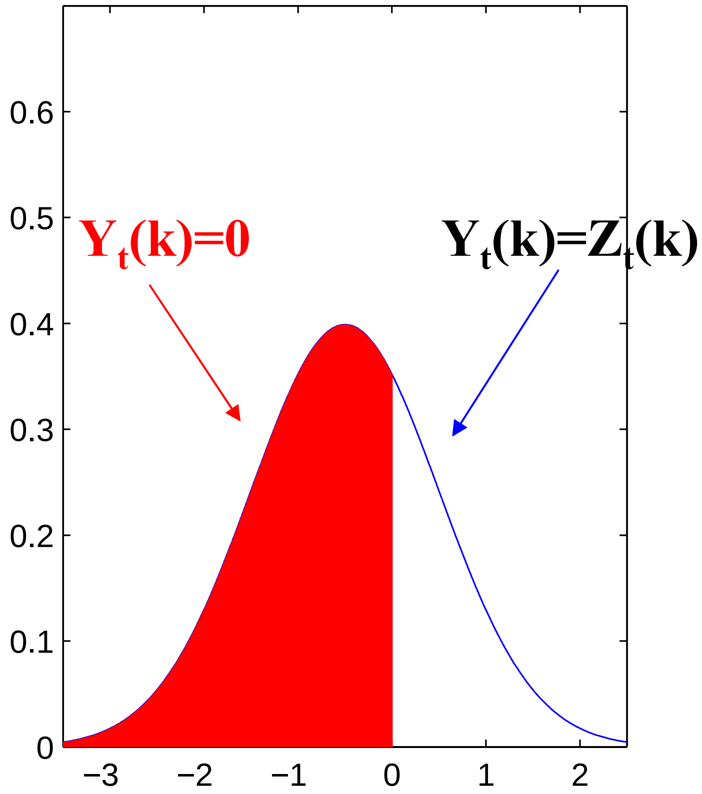
$$Y_t(k) = \max(Z_t(k), 0)$$

with

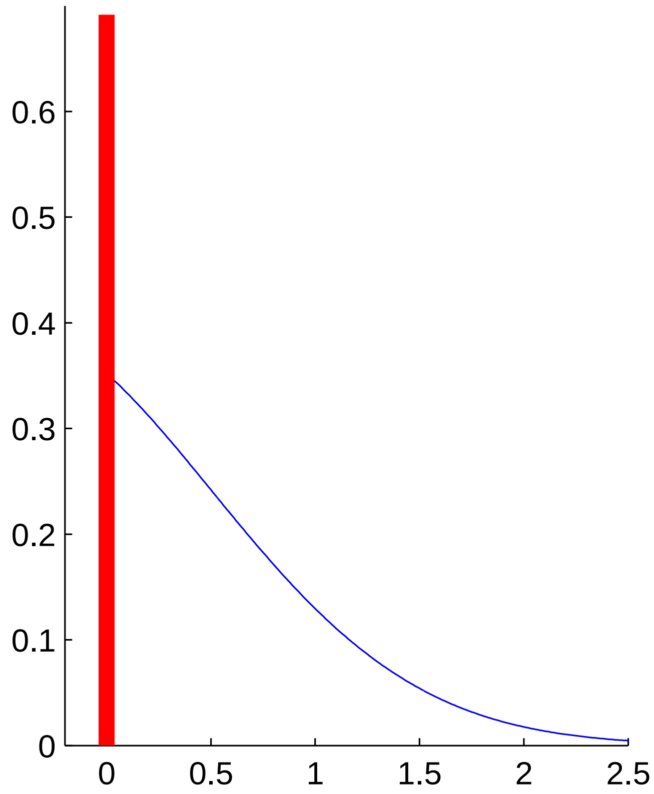
- $Z_t = (Z_t(1), \ldots, Z_t(K))' = \mu^{(s)} + S^{(s)} E_t$
- $\mu^{(s)} \in R^K$, $\Sigma^{(s)} = S^{(s)} S^{(s)}{}' \in R^{K*K}$
- $E_t \sim N(0, I_K)\ i.i.d$

HMM with truncated Gaussian fields

Model description



Conditional distribution of $Z_t(k)$



Conditional distribution of $Y_t(k)$

HMM with truncated Gaussian fields

Model description

- Assumptions on the covariance matrices

$$\Sigma^{(s)} = diag((\sigma_1^{(s)})^2, \ldots, (\sigma_K^{(s)})^2) \quad \text{(HMMCI)}$$

$$\Sigma^{(s)}(i,j) = \sigma_i^{(s)}\sigma_j^{(s)} \exp(-\lambda^{(s)}\, dist\,(x_i, x_j)) \quad \text{(HMMdist)}$$

$$\Sigma^{(s)}(i,j) = \sigma_i^{(s)}\sigma_j^{(s)} \exp(-\lambda_i^{(s)}\lambda_j^{(s)}\, dist\,(x_i, x_j)) \quad \text{(HMMloc)}$$

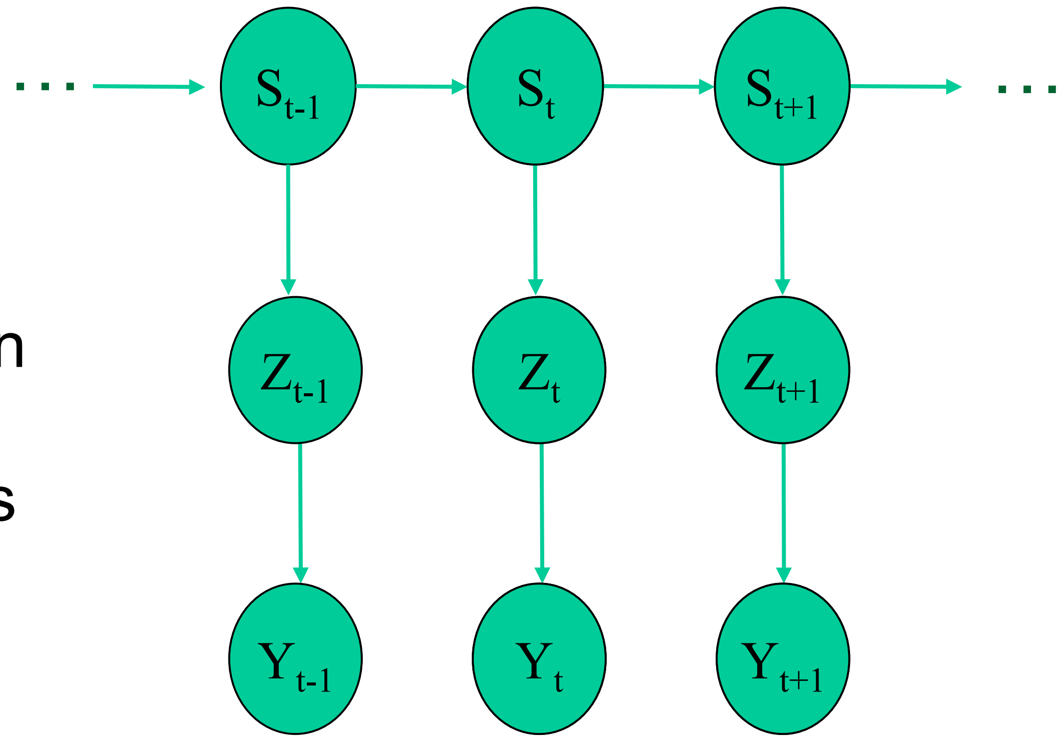
HMM with truncated Gaussian fields

Parameter estimation

Weather type
- Not observed
- Finite values

Conditionally Gaussian
- Partially observed
- Continuous values

Precipitation
- Observed
- Mixed continuous-discrete



HMM with truncated Gaussian fields

Parameter estimation

- E-step (Forward-Backward algorithm)
  - Computation of $p(Y_t=y_t|S_t=s;\Theta)$
  - If $y_t=(0,\ldots,0,y_t(k+1),\ldots y_t(K))$ with $y_t(k+1)>0,\ldots, y_t(K)>0$

$$p(Y_t = y \mid S_t = s;\Theta)$$
$$= \int_{]-\infty,0]^k} f(z_1,\ldots,z_k, y_t(k+1),\ldots, y_t(K);\mu^{(s)},\Sigma^{(s)})dz_1\ldots dz_k$$

- M-step
  - Computation of $E[Z_t|Y_t=y_t,S_t=s;\Theta]$ and $cov(Z_t|Y_t=y_t,S_t=s;\Theta)$
  - …integral expression

HMM with truncated Gaussian fields

Parameter estimation

- Monte Carlo integration
  - If $y_t=(0,\ldots,0,y_t(k+1),\ldots y_t(K))$, then
    - Simulate N samples from the multivariate Gaussian distribution

      $$P(Z(1),\ldots,Z(k)|Z(k+1)=y_t(k+1),\ldots,Z(K)=y_t(K);\mu^{(s)},\Sigma^{(s)})$$

    - Deduce approximate values for the integrals. For ex.:

$$p(Y_t = y \mid S_t = s;\Theta)$$

$$\approx \frac{Nb\, of\ samp.with\, all\, comp. \leq 0}{N} \times p(y_t(k+1),\ldots,y_t(K);\mu^{(s)},\Sigma^{(s)})$$

- MCEM algorithm

HMM with truncated Gaussian fields

	HMMCI (4 states)	HMMdist (4 states)	HMMloc (4 states)	HMMfull (4 states)
loglik	-6837	-6441	-6383	-6311
numpar	68	72	96	152
AIC	13811	13029	12959	**12928**
BIC	14130	**13367**	13409	13641

$$\Sigma^{(s)} = diag((\sigma_1^{(s)})^2, \ldots, (\sigma_K^{(s)})^2)$$

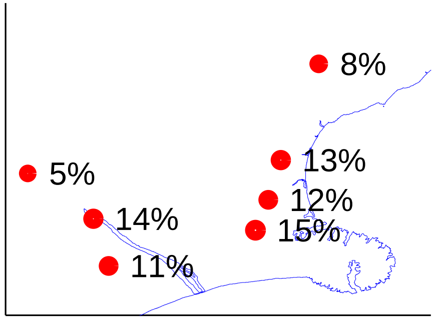
$$\Sigma^{(s)}(i,j) = \sigma_i^{(s)} \sigma_j^{(s)} \exp(-\lambda^{(s)}\, dist(x_i, x_j))$$

$$\Sigma^{(s)}(i,j) = \sigma_i^{(s)} \sigma_j^{(s)} \exp(-\lambda_i^{(s)} \lambda_j^{(s)}\, dist(x_i, x_j))$$

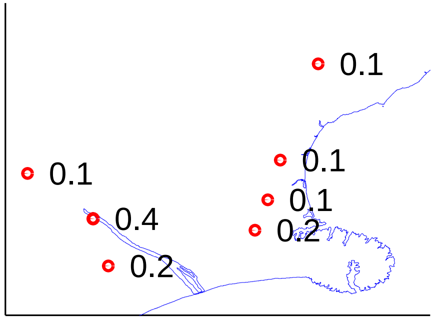
HMM with truncated Gaussian fields

Meteorological interpretability (HMMloc)

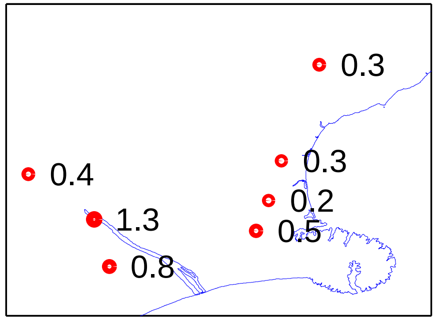
$P[Y_t>0|S_t=1]$



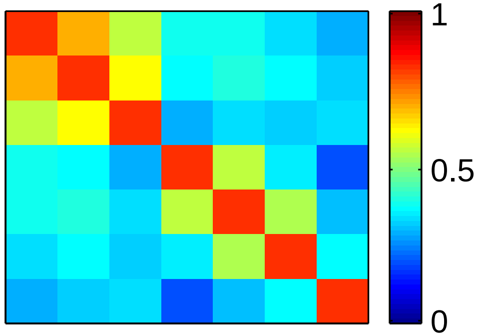
$E[Y_t|S_t=1]$



$\sigma[Y_t|S_t=1]$



$Corr(Y_t|S_t=1)$



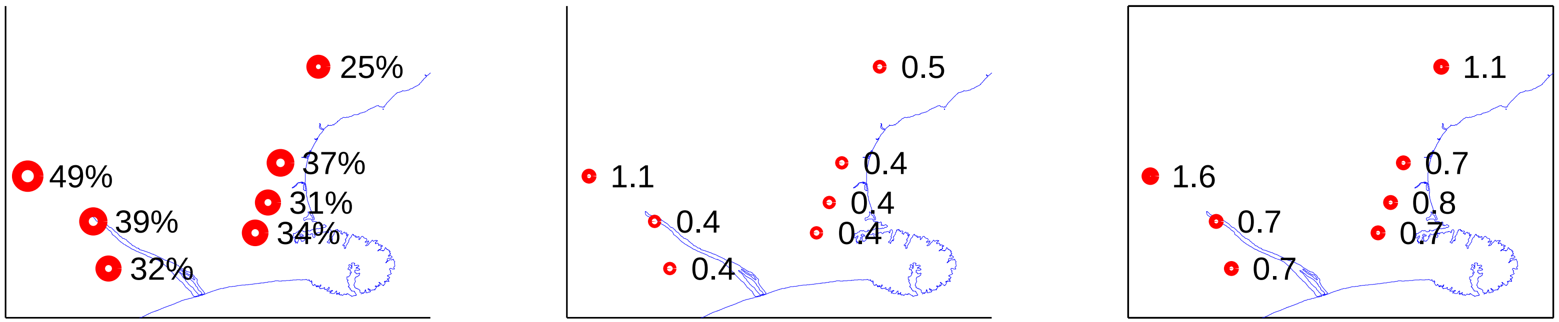
<u>Regime 1</u>

- Low probability of rainfall
- Low amount
- High spatial correlation

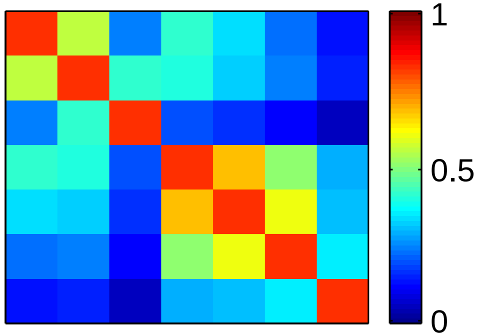
HMM with truncated Gaussian fields

Meteorological interpretability (HMMloc)

$P[Y_t>0|S_t=2]$ $\quad$ $E[Y_t|S_t=2]$ $\quad$ $\sigma[Y_t|S_t=2]$



$Corr(Y_t|S_t=2)$



<u>Regime 2</u>

•Moderate probability of rainfall occurrence, higher at location 3

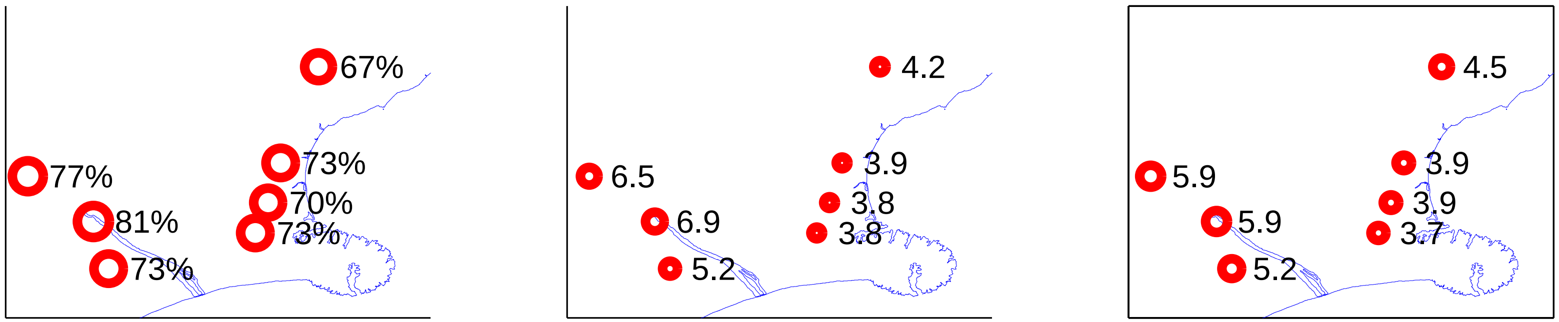
•Low amounts, higher at location 3

•Moderate spatial correlation, low correlation between locations 3 and 7 and other locations

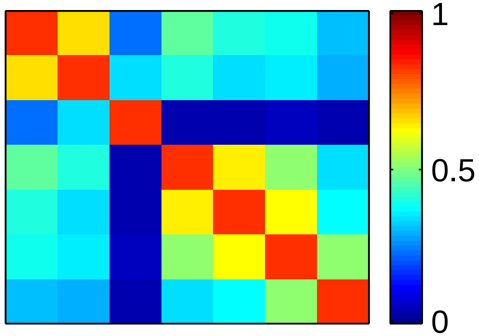
HMM with truncated Gaussian fields

Meteorological interpretability (HMMloc)

$P[Y_t>0|S_t=3]$ $E[Y_t|S_t=3]$ $\sigma[Y_t|S_t=3]$



$Corr(Y_t|S_t=3)$



<u>Regime 3</u>

•Moderate probability of rainfall occurrence

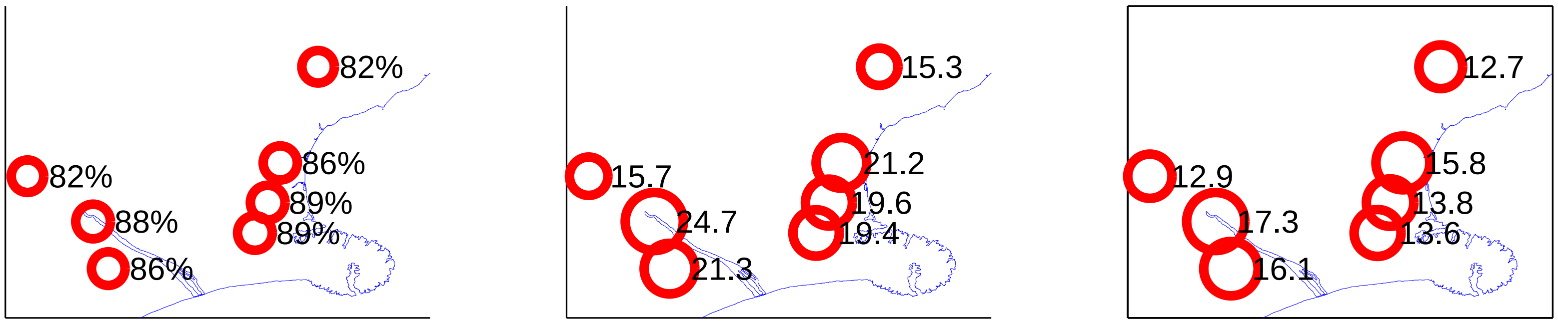
•Moderate amounts, higher in the west

•Moderate spatial correlation, low correlation between location 3 and other locations

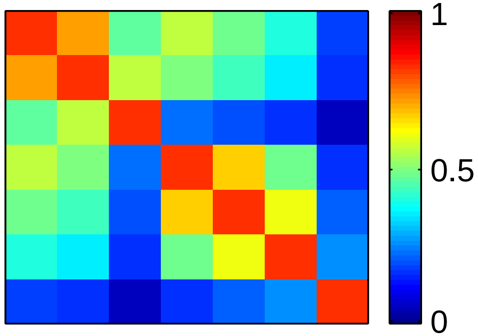
HMM with truncated Gaussian fields

Meteorological interpretability (HMMloc)

$P[Y_t>0|S_t=4]$ $E[Y_t|S_t=4]$ $\sigma[Y_t|S_t=4]$



$Corr(Y_t|S_t=4)$



<u>Regime 4</u>

- •High probability of rainfall occurrence
- •High amount
- •High spatial correlation, except between
  - •location 3 and locations in the east
  - •location 7 and other locations

HMM with truncated Gaussian fields

Meteorological interpretability (HMMloc)

- Transition matrix, stationary distribution , mean durations

0.73	0.18	0.06	0.01
0.29	0.42	0.27	0.03
0.23	0.32	0.36	0.08
0.06	0.42	0.23	0.28

0.48
0.29
0.18
0.05

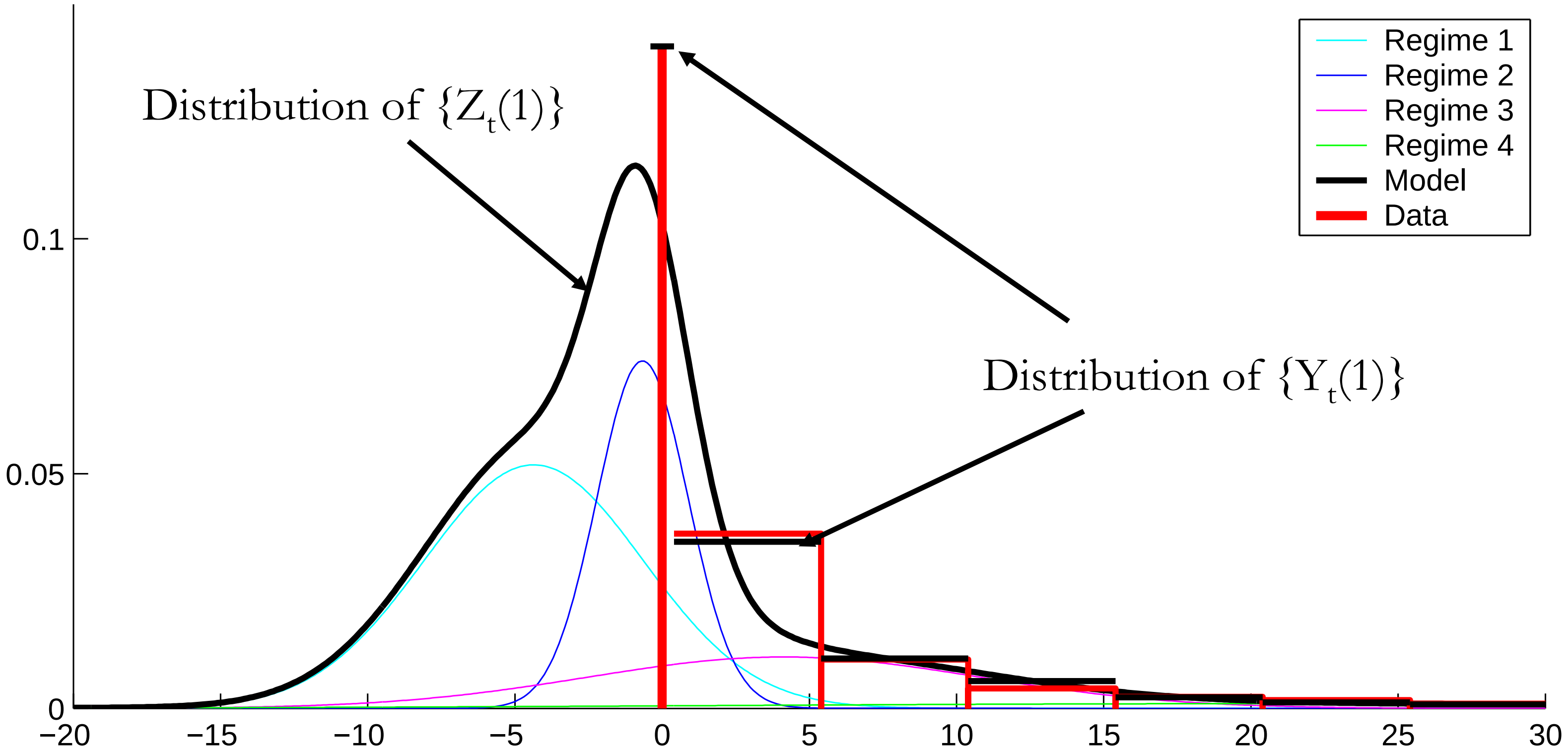
3.78
1.71
1.55
1.40

- Summary:
  - **Regime 1:** low rainfall
  - **Regime 2 and 3:** intermediate patterns, regional differences, higher rainfall in regime 3, short persistence
  - **Regime 4:** high rainfall
  - Location 3 and 7 have specific behaviors

HMM with truncated Gaussian fields

Realism of simulated sequences (HMMloc)

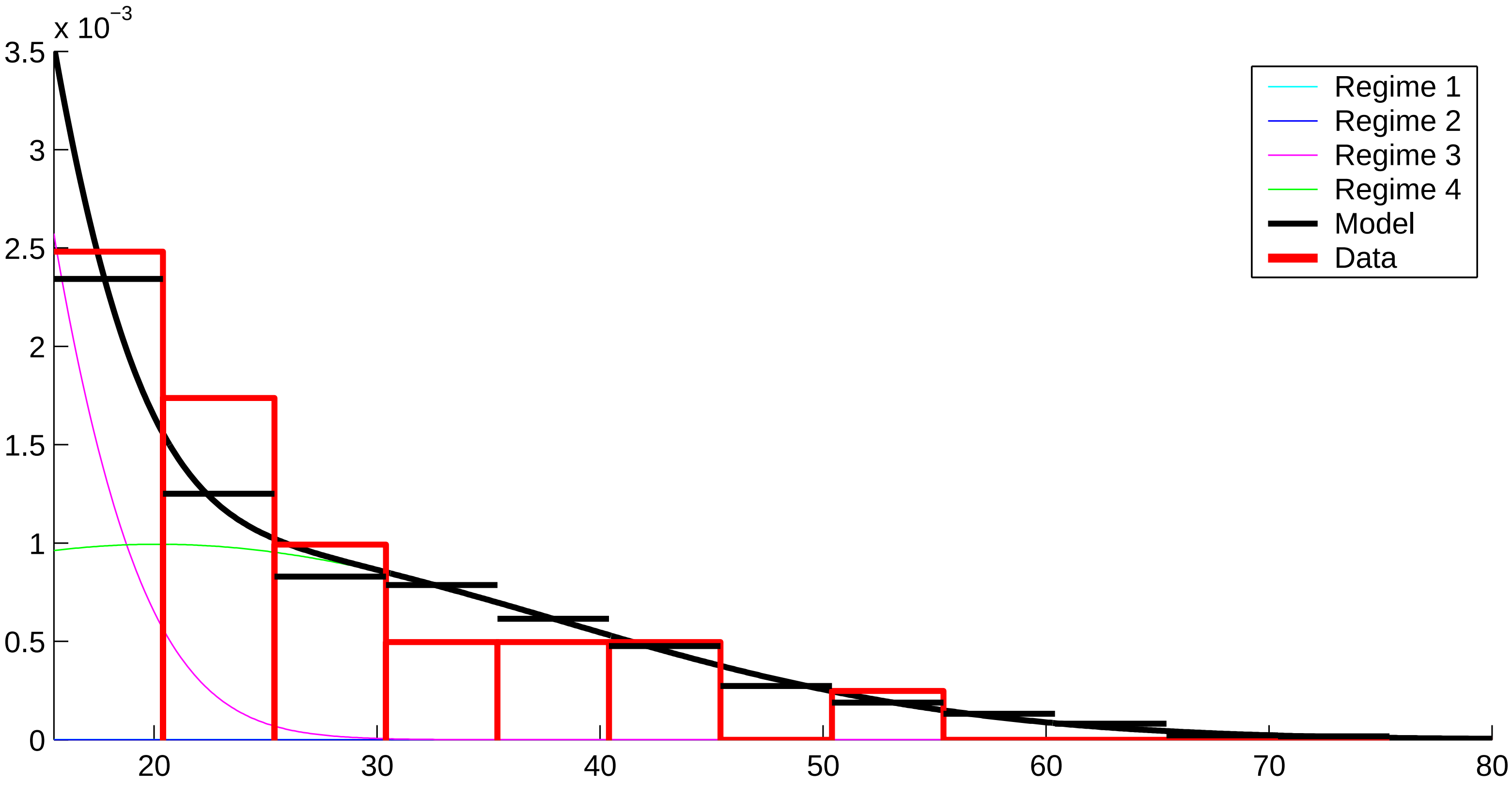
- Marginal distribution (location 1, Winchmore)



HMM with truncated Gaussian fields

Realism of simulated sequences (HMMloc)

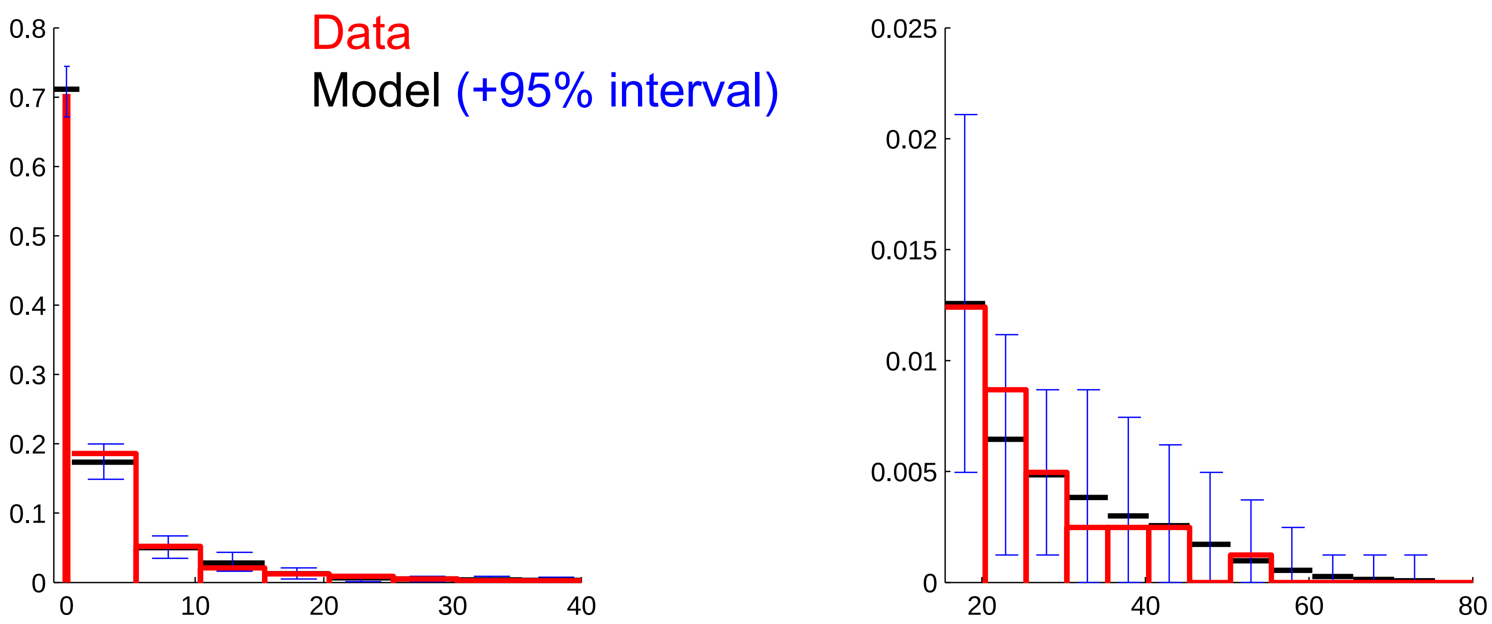
- Marginal distribution (location 1, Winchmore)



HMM with truncated Gaussian fields

Realism of simulated sequences (HMMloc)

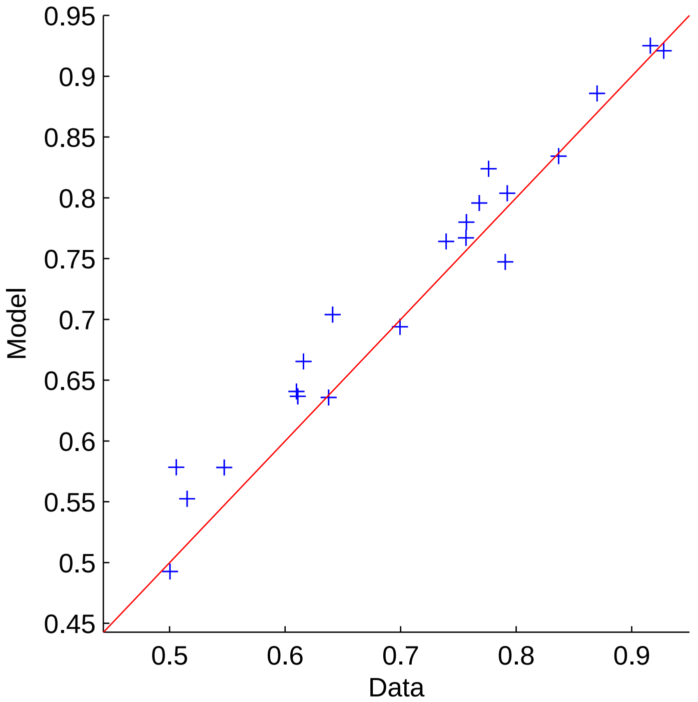
- Marginal distribution (location 1, Winchmore)



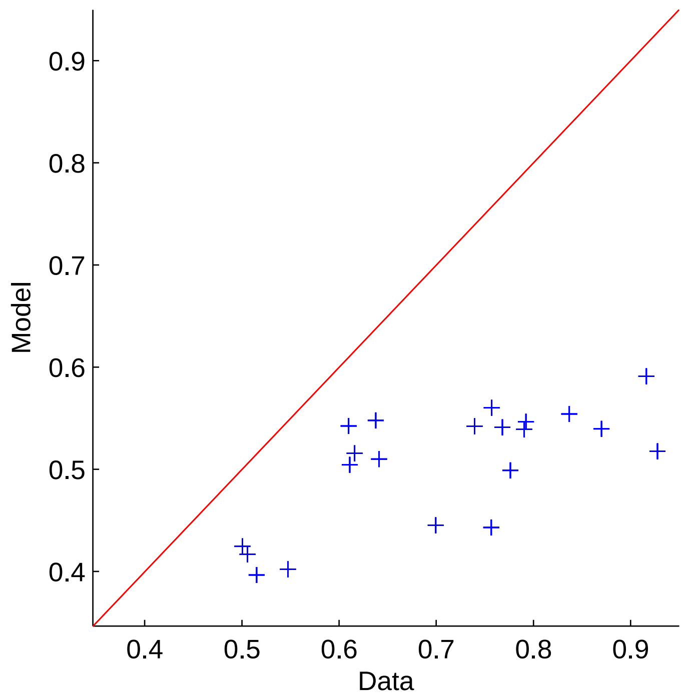
HMM with truncated Gaussian fields

Realism of simulated sequences (HMMloc)

- Pair-wise spatial correlations (amounts)



Model with spatial dependence

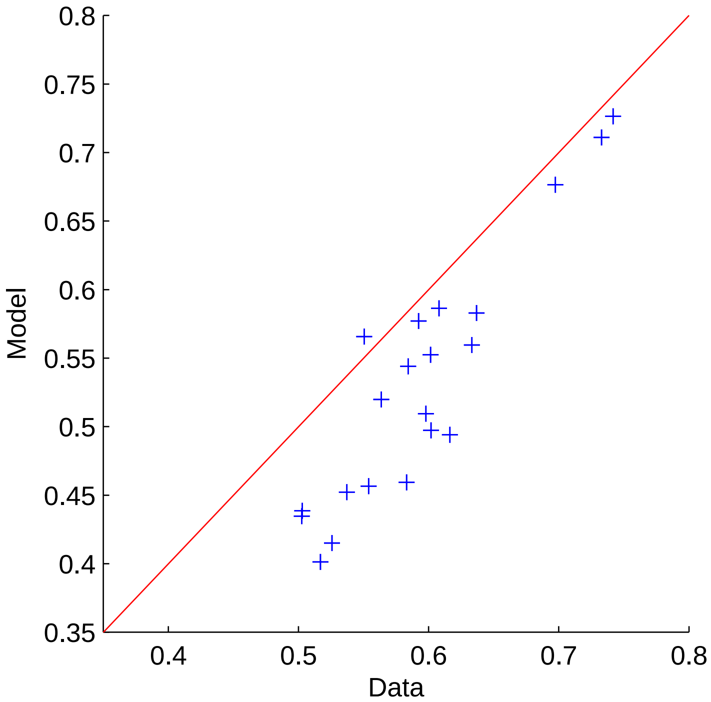


Conditional spatial independence

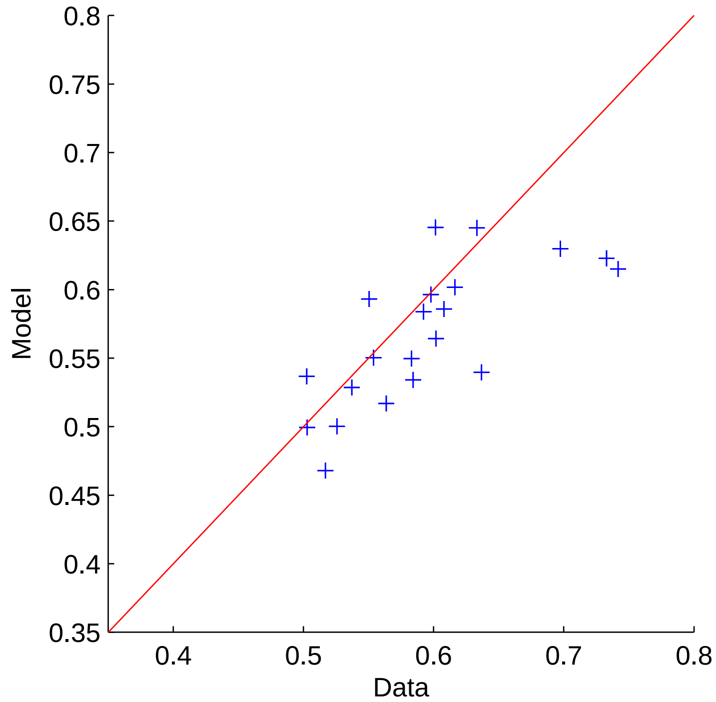
HMM with truncated Gaussian fields

Realism of simulated sequences (HMMloc)

- Pair-wise spatial correlations (occurrence)



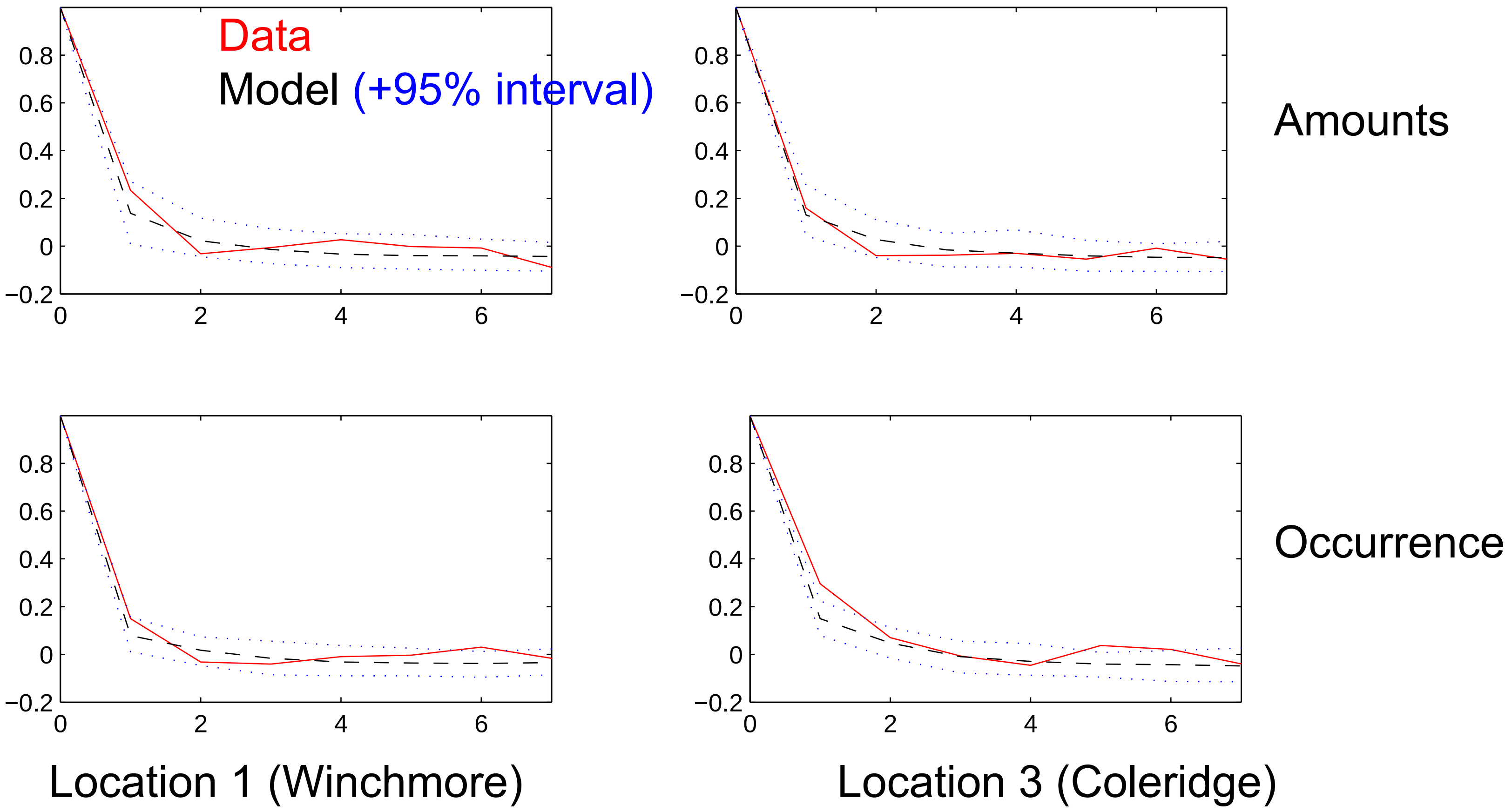
Model with spatial dependence



Conditional spatial independence

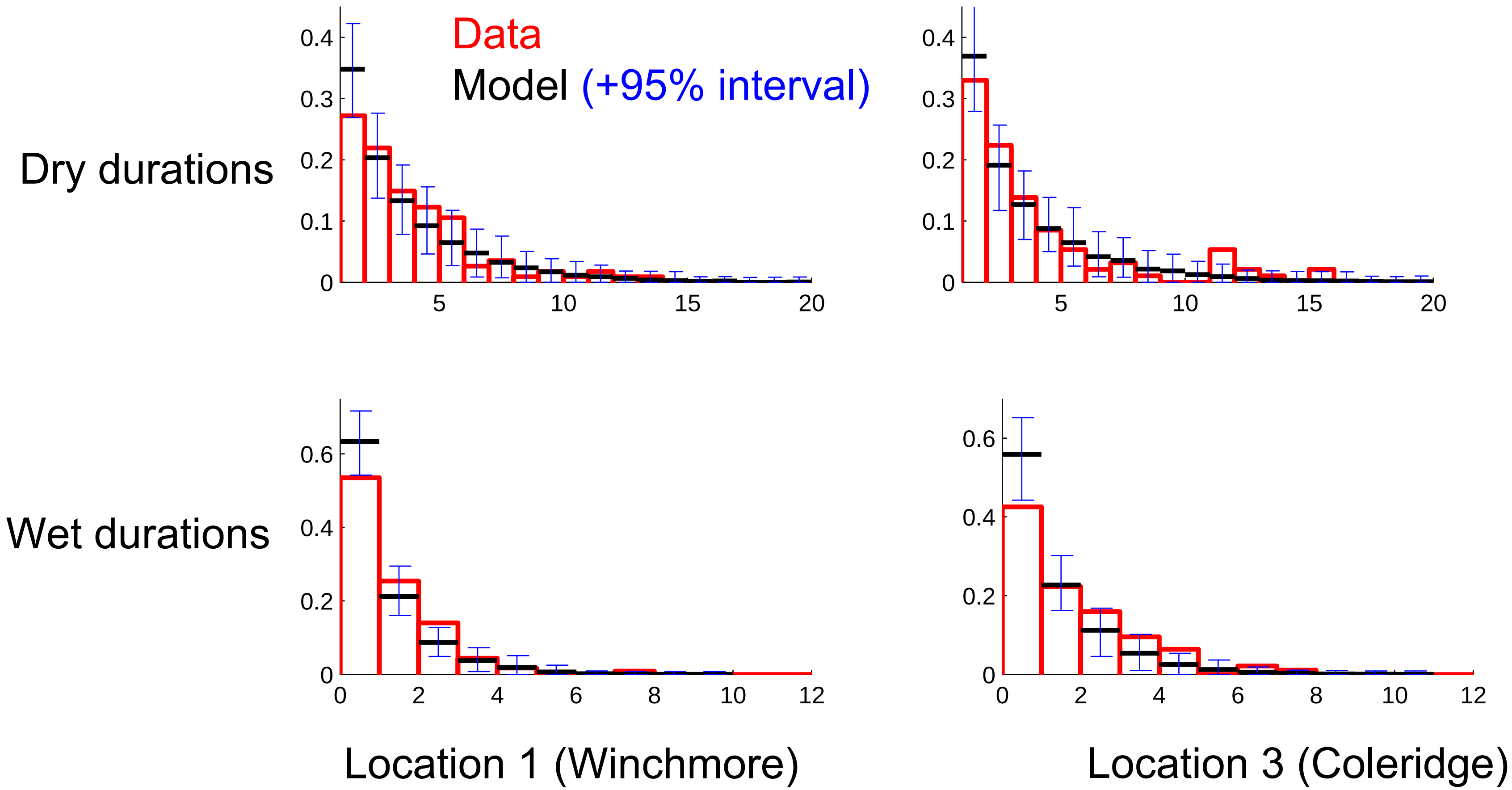
HMM with truncated Gaussian fields

Realism of simulated sequences (HMMloc)



HMM with truncated Gaussian fields

Realism of simulated sequences (HMMloc)



Conclusion & perspectives

- HMM with conditional spatial independence assumption cannot reproduce the spatial structure
- HMM with truncated correlated Gaussian distribution better reproduces the spatial structure
- Dynamics still not well reproduced
  - Add an autoregressive part?
- Explore other possibilities
  - Markov random fields, local weather types,…
- Seasonality, inter-annual variability….

References

- Allcroft D.J. and Glasbey C.A. (2003). A latent Gaussian Markov random--field model for spatiotemporal rainfall disaggregation. Journal of the Royal Statistical Society C, 52, 487--498.
- Hughes JP, Guttorp P. (1994). A class of stochastic models for relating synoptic atmospheric patterns to local hydrologic phenomenon. Water Resources Research, 30, 1535-1546
- Hughes J.P., Guttorp P., Charles S. (1999). A nonhomogeneous. hidden Markov model for precipitation occurrence. Journal of the Royal Statistical Society C, 48,15–30.
- Kirshner S. (2005). Modeling of multivariate time series using hidden Markov models. PhD thesis, University of California.
- Thompson, C.S., Thomson, P.J. and Zheng, X. (2005). A multisite rainfall generation model applied to New Zealand data. NIWA Technical Report 128,National Institute of Water and Atmospheric Research, New Zealand.
- Zucchini W., Guttorp P. (1991). A hidden Markov model for space-time precipitation. Water Resources Research, 27, 1917–1923.