About myself: *Mikyoung Jun*

- PhD in statistics, University of Chicago (Aug 2005)
  - Advisor: Michael Stein
  - Evaluation of numerical models for air quality (CMAQ)
  - summer intern at EPA, RTP, 2002
  - Space-time asymmetric covariance functions on spheres
- Post-doc at NCAR (Aug 2005 - Dec 2005)
- Assistant Professor of statistics at Texas A& M University (Aug 2005 - present)

Statistical analysis of IPCC climate model biases

Mikyoung Jun

mjun@stat.tamu.edu

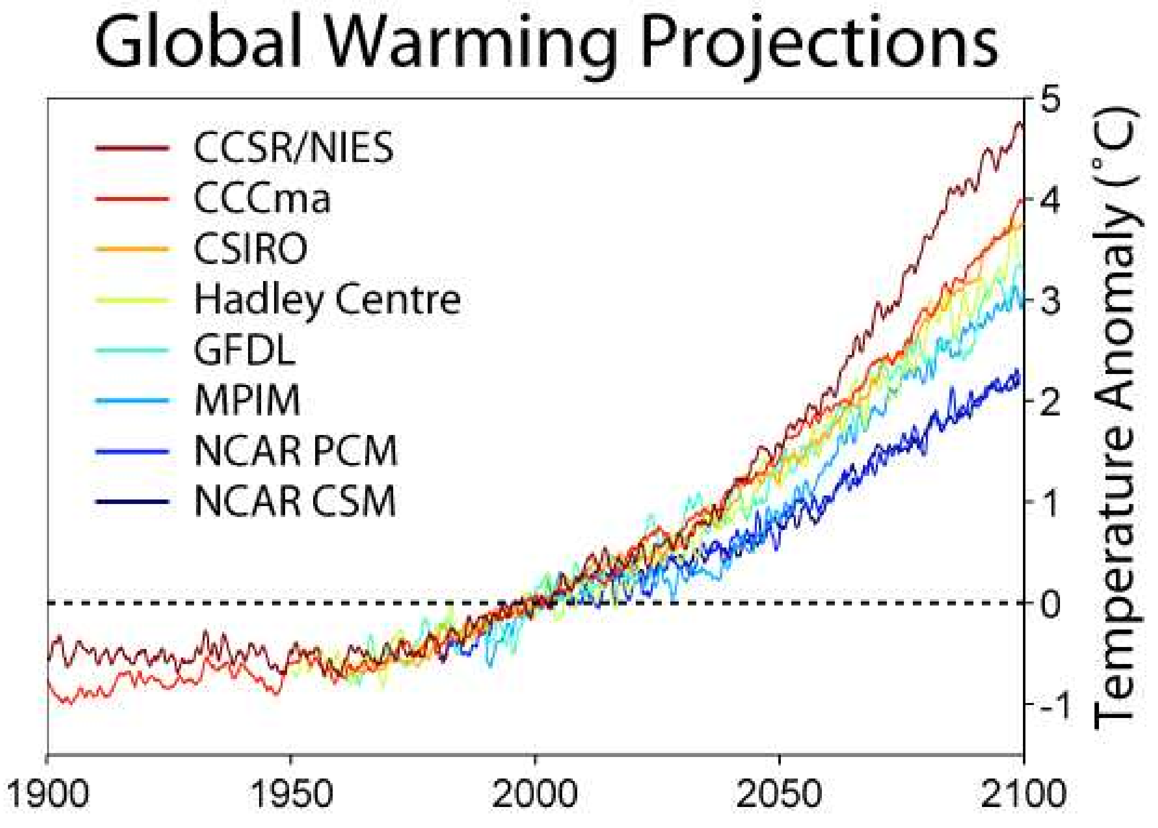
Texas A&M University

This is joint work with Reto Knutti (NCAR/CGD).

IPCC

- The Intergovernmental Panel on Climate Change (IPCC) was established in 1988 by two United Nations organizations, the World Meteorological Organization (WMO) and the United Nations Environment Programme (UNEP) to assess the "risk of human-induced climate change".
- Global climate model: numerical model that gives climate output on grids
- We have more than 20 climate models developed by various organizations over the world
- How accurate are these models?

IPCC climate models



(from Wikipedia)

Questions of interest

- How can we quantify the model bias?
- How are the model biases correlated?
- How can we improve climate estimates through the above information?
- Relationship between model resolution and the model bias?

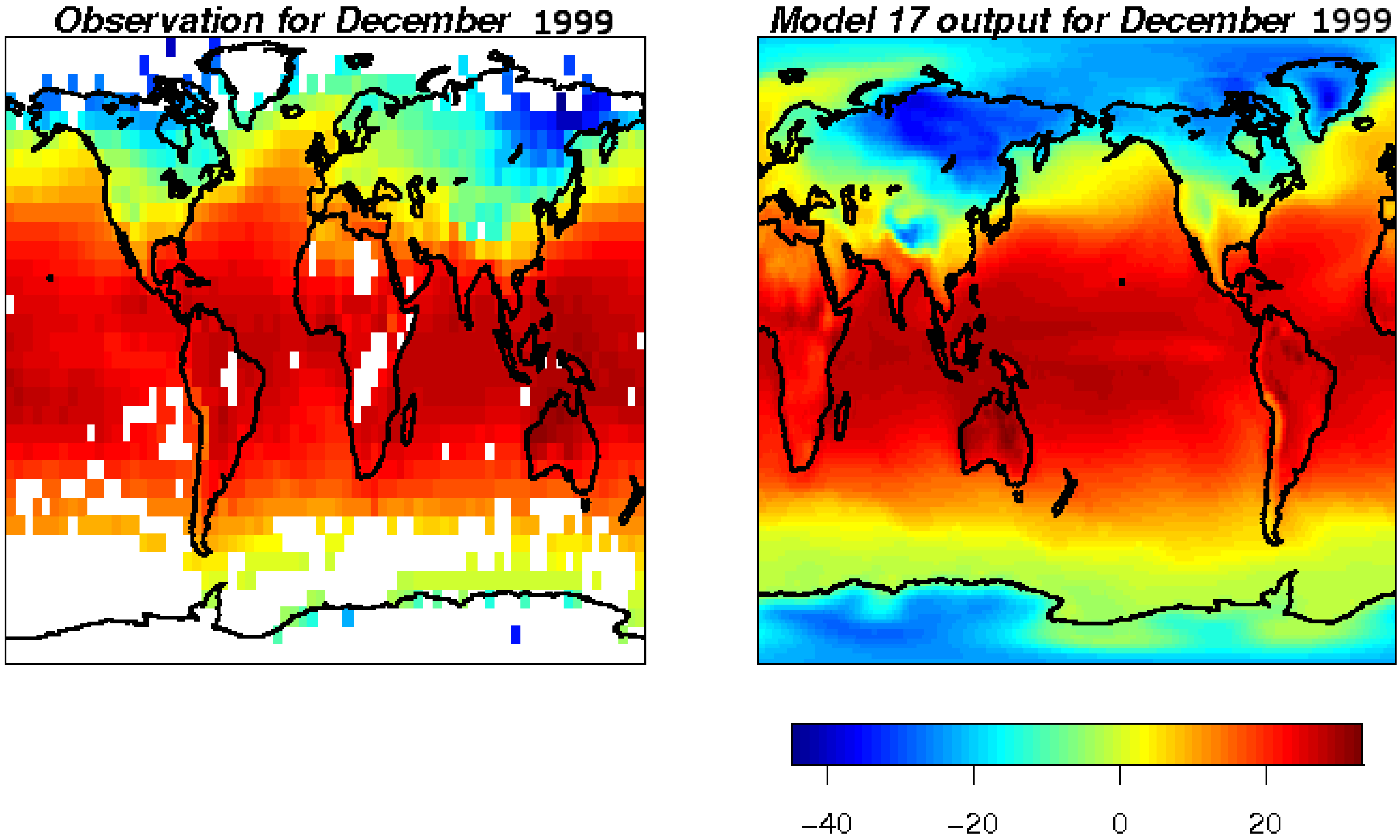
Data

- Monthly averages in 1970-1999 (unit: $^\circ$ C)
- Observations
  - Surface temperature for both land and sea
  - Combined data set from CRU (Climate Research Unit, East Anglia) and the Hadley Centre (UK)
  - A lot of missing data around both poles
- IPCC model outputs
  - 20 models, no missing data

IPCC models

	model	resolution		model	resolution
1	bcc cm1	192×96	11	inmcm3 0	72×45
2	cccma cgcm3 1	96×48	12	ipsl cm4	96×72
3	cnrm cm3	128×64	13	miroc3 2 medres	128×64
4	csiro mk3 0	192×96	14	miub echo g	96×48
5	gfdl cm2 0	144×90	15	mpi echam5	192×96
6	gfdl cm2 1	144×90	16	mri cgcm2 3 2a	128×64
7	giss aom	90×60	17	ncar ccsm3 0	256×128
8	giss model e g	72×46	18	ncar pcm1	128×64
9	giss model e r	72×46	19	ukmo hadcm3	95×73
10	iap fgoals1 0 g	128×60	20	ukmo hadgem1	192×145
	observation	72×36			

Observation vs model output



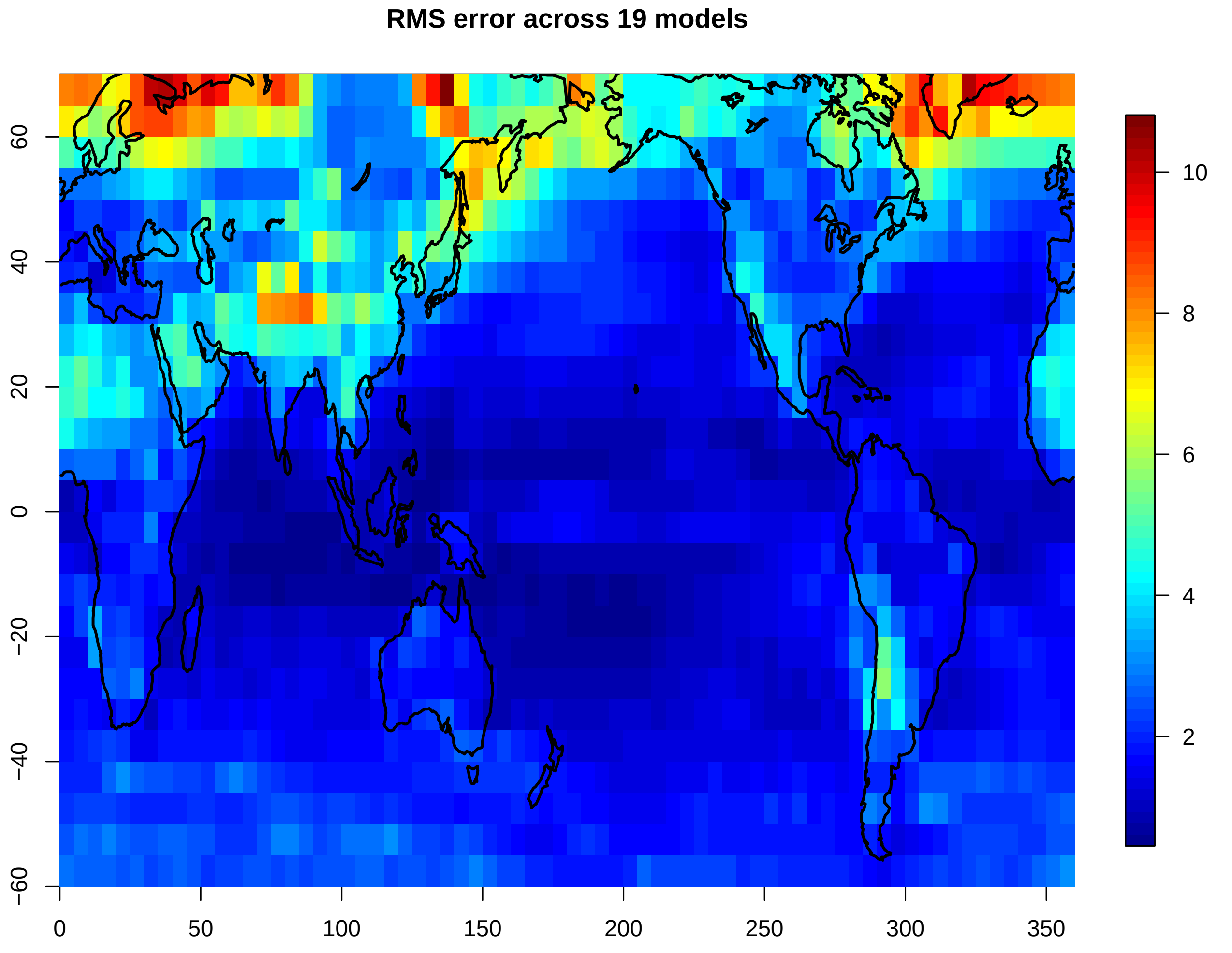
Observation vs model output

- Temporal correlation of less interest: take DJF average
- Due to the missing observations around poles, take latitude range 60° S - 72° N; total 1872 grid cells
- Analysis on the difference between observation and model output (Jun and Stein, 2004)
- Bilinear interpolation of the model outputs to the observation grid
- Naive imputation for a few missing locations

Statistical Model

- $X(\mathbf{s}, t)$: DJF averages of the observation at location $\mathbf{s}$ and year t
- $Y_i(\mathbf{s}, t)$: i^{th} model output
- Difference: $D_i(\mathbf{s}, t) = X(\mathbf{s}, t) - Y_i(\mathbf{s}, t)$
- $D_i(\mathbf{s}, t) = b_i(\mathbf{s}) + u_i(\mathbf{s}, t)$
- Model b_i based on $\bar{D}_i(\mathbf{s}) = \frac{1}{30} \sum_{t=1}^{30} D_i(\mathbf{s}, t)$.

30 year average of differences



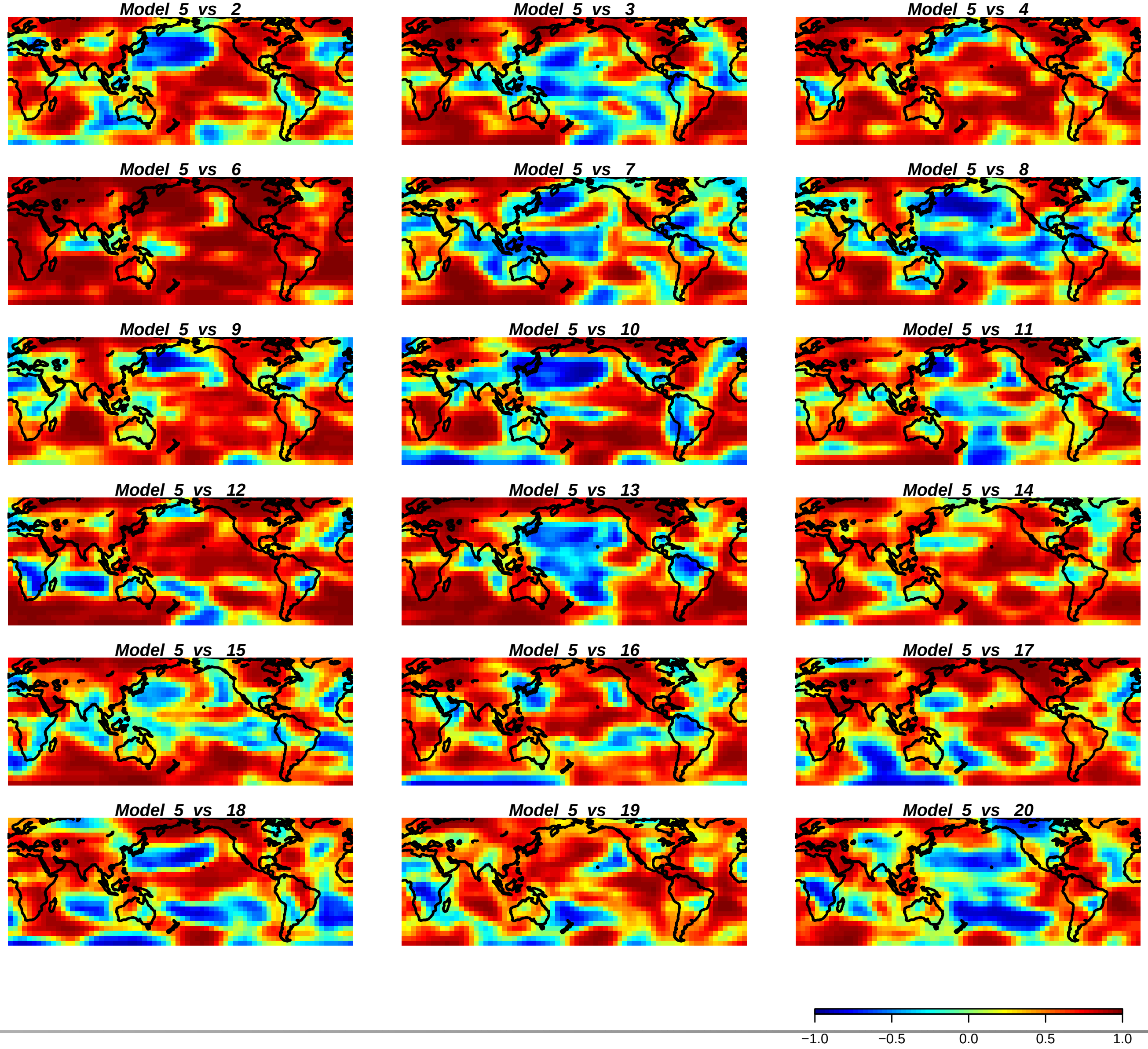
Climate model bias

- $b_i(\mathbf{s}) = \mu_{0i} + \mu_{1i}L(\mathbf{s}) + \mu_{2i}\mathbf{1}_{(\mathbf{s}\in\text{land})} + \mu_{3i}A(\mathbf{s}) + a_i(\mathbf{s})$
- a_i: mean zero Gaussian random field
- Covariance structure of a_i:
  - $a_i(\mathbf{s}) = (\delta_i \mathbf{1}_{(\mathbf{s}\in\text{land})} + 1)Z_i(\mathbf{s})$ ($\delta_i > 0$)
  - Cov$\{Z_i(\mathbf{s}_1), Z_i(\mathbf{s}_2)\} = \alpha_i \mathcal{M}_{\nu_i+1}(d)$ (d: chordal distance between $\mathbf{s}_1$ and $\mathbf{s}_2$)
  - Could apply more complex model for Nonstationarity (Jun and Stein 2005)

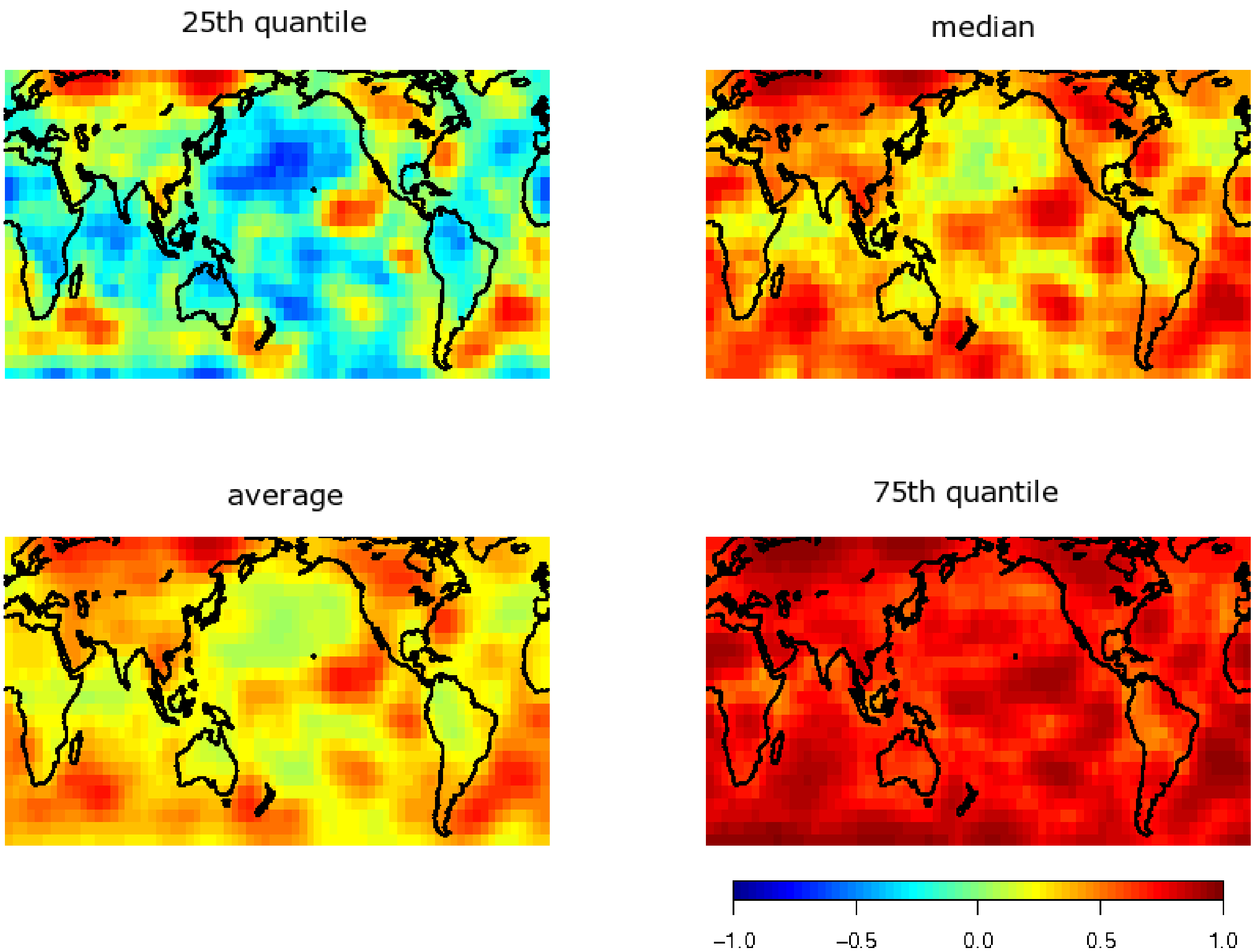
Correlation between model biases

- $\sigma_{ij}(\mathbf{s})$=Cov{$a_i(\mathbf{s})$, $a_j(\mathbf{s})$}
- Apply Gaussian kernel to $\tilde{D}_{ij}(\mathbf{s}) = \tilde{D}_i(\mathbf{s})\tilde{D}_j(\mathbf{s})$, where $\tilde{D}_i(\mathbf{s})$ is “filtered $\bar{D}_i(\mathbf{s})$”
- $r_{ij}(\mathbf{s}) = \frac{\sigma_{ij}}{\sqrt{\sigma ii \sigma jj}}(\mathbf{s})$
  : correlation between i^{th} and j^{th} model biases at location $\mathbf{s}$

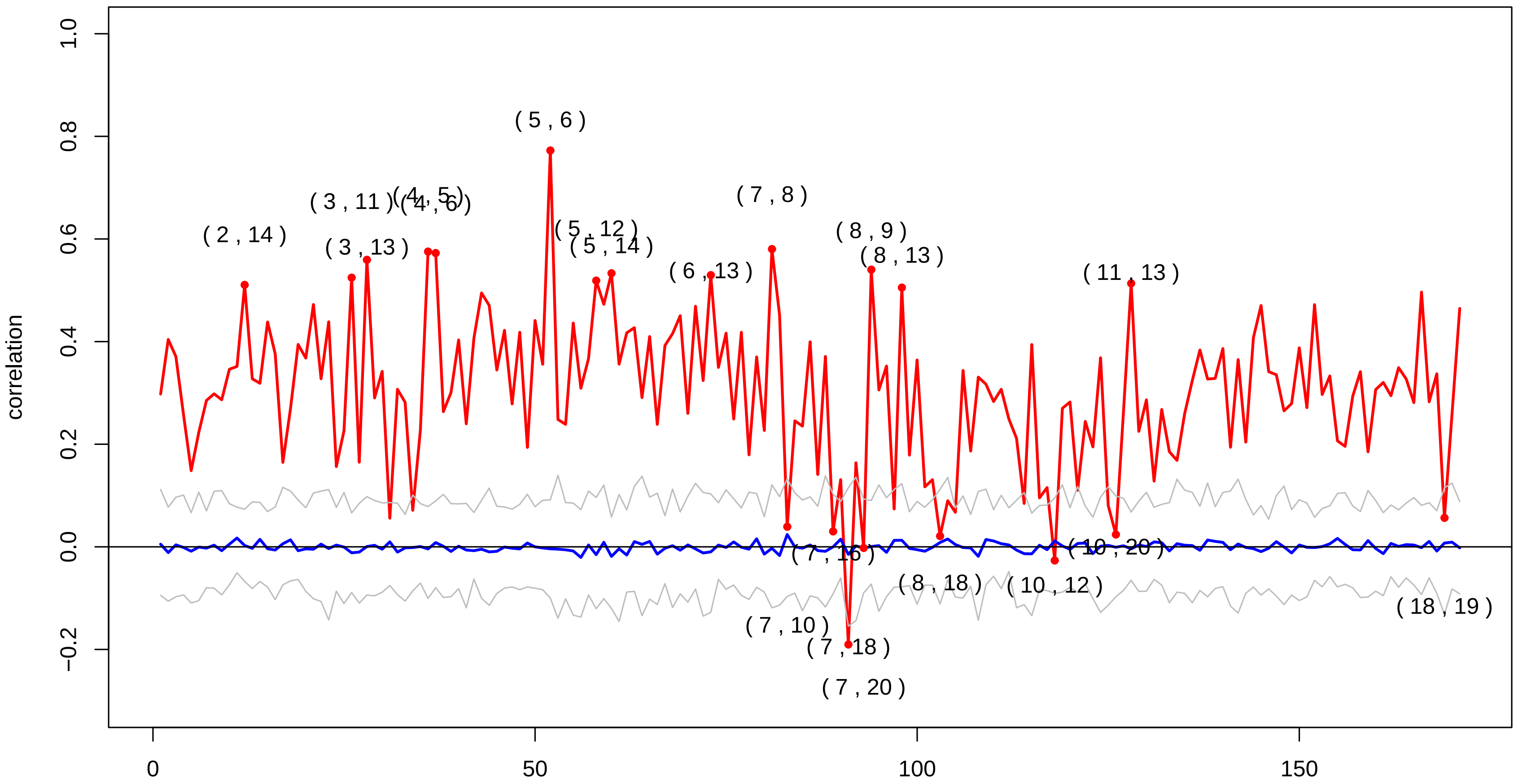
r_{5j}, for $j = 2 \cdots 20$



Summary of r_{ij}

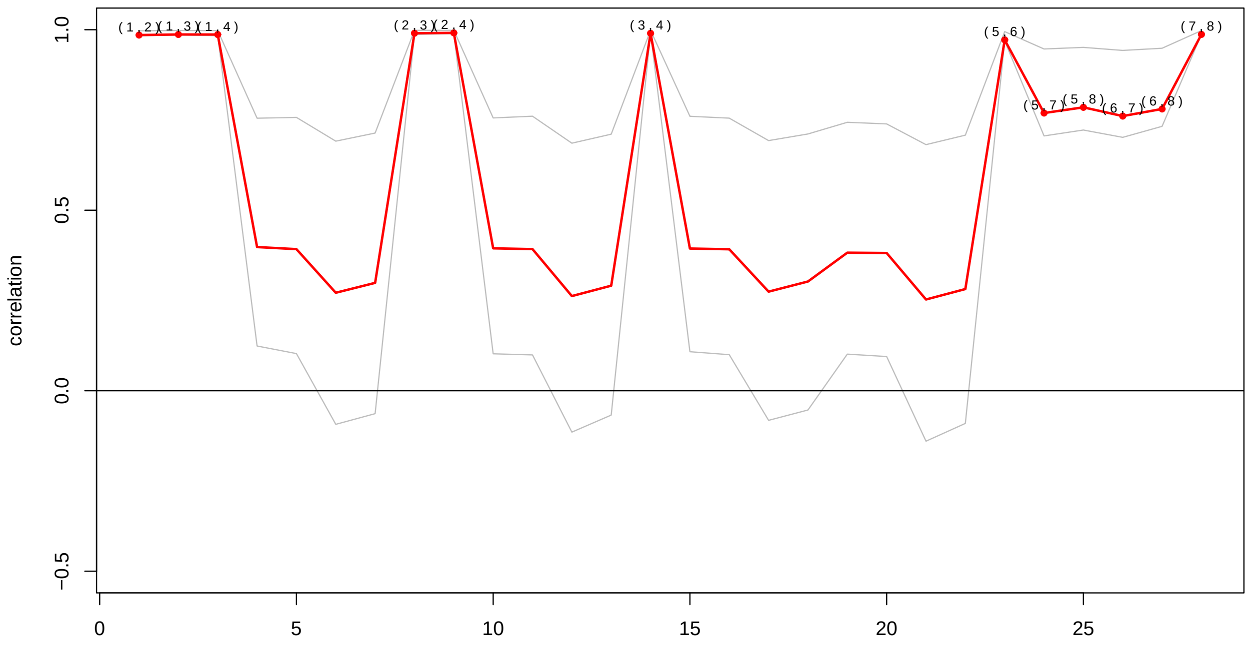


Comparison to simulated data

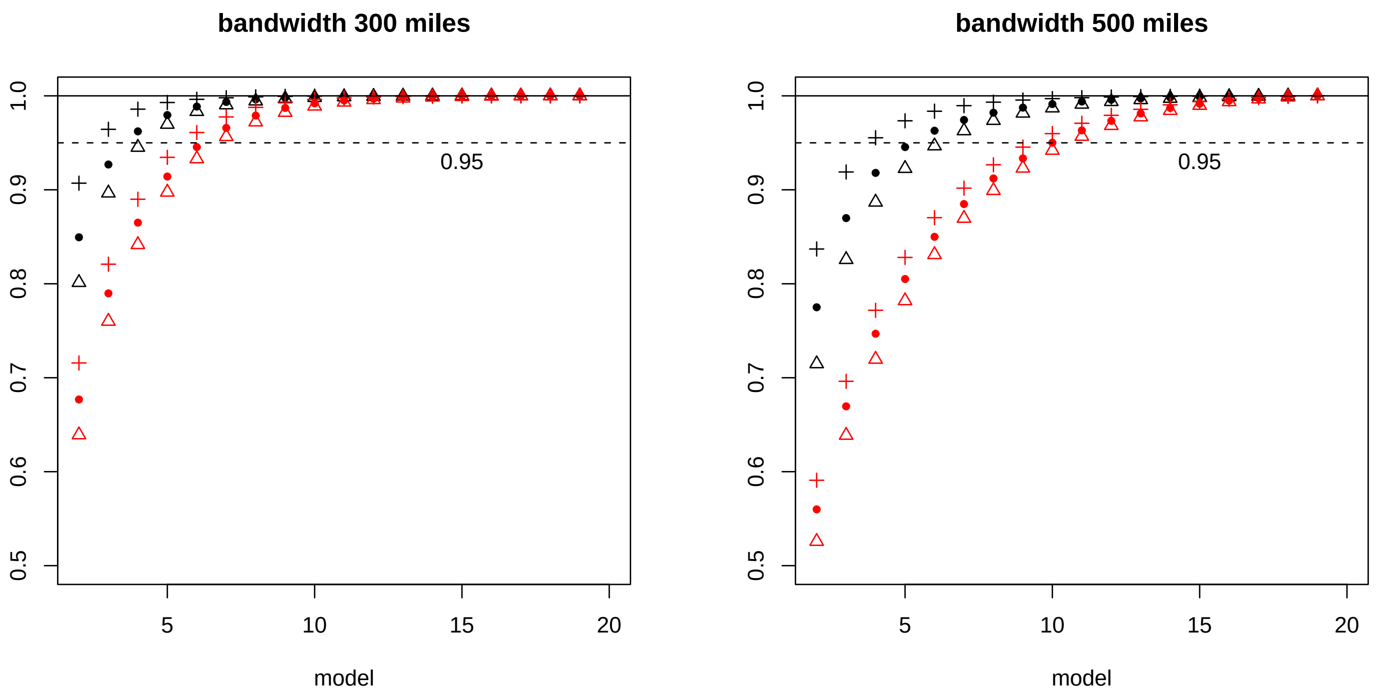


Verification using ensemble runs

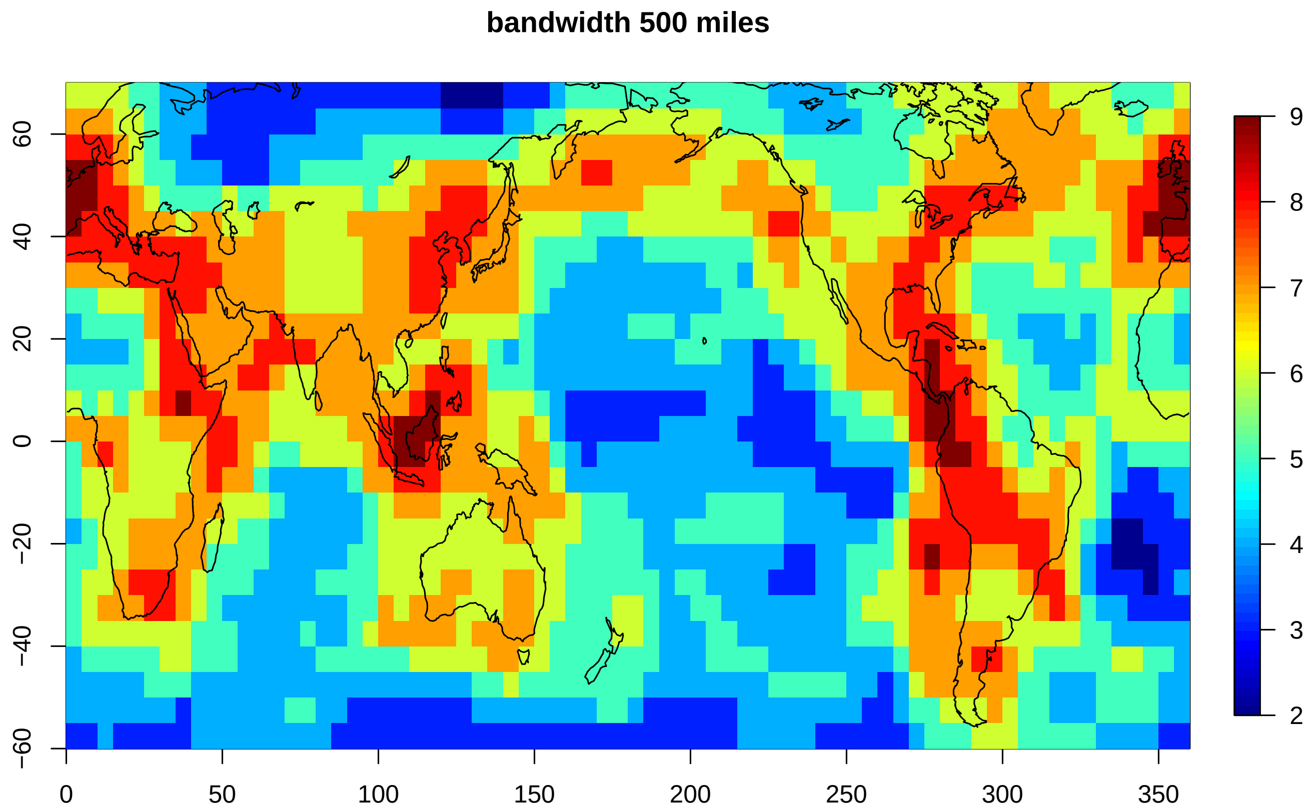
- 4 additional ensemble runs of model 2, 2 for model 5 and 6
- Correlation between ensemble runs above 0.98



Eigenvalues of correlation matrix



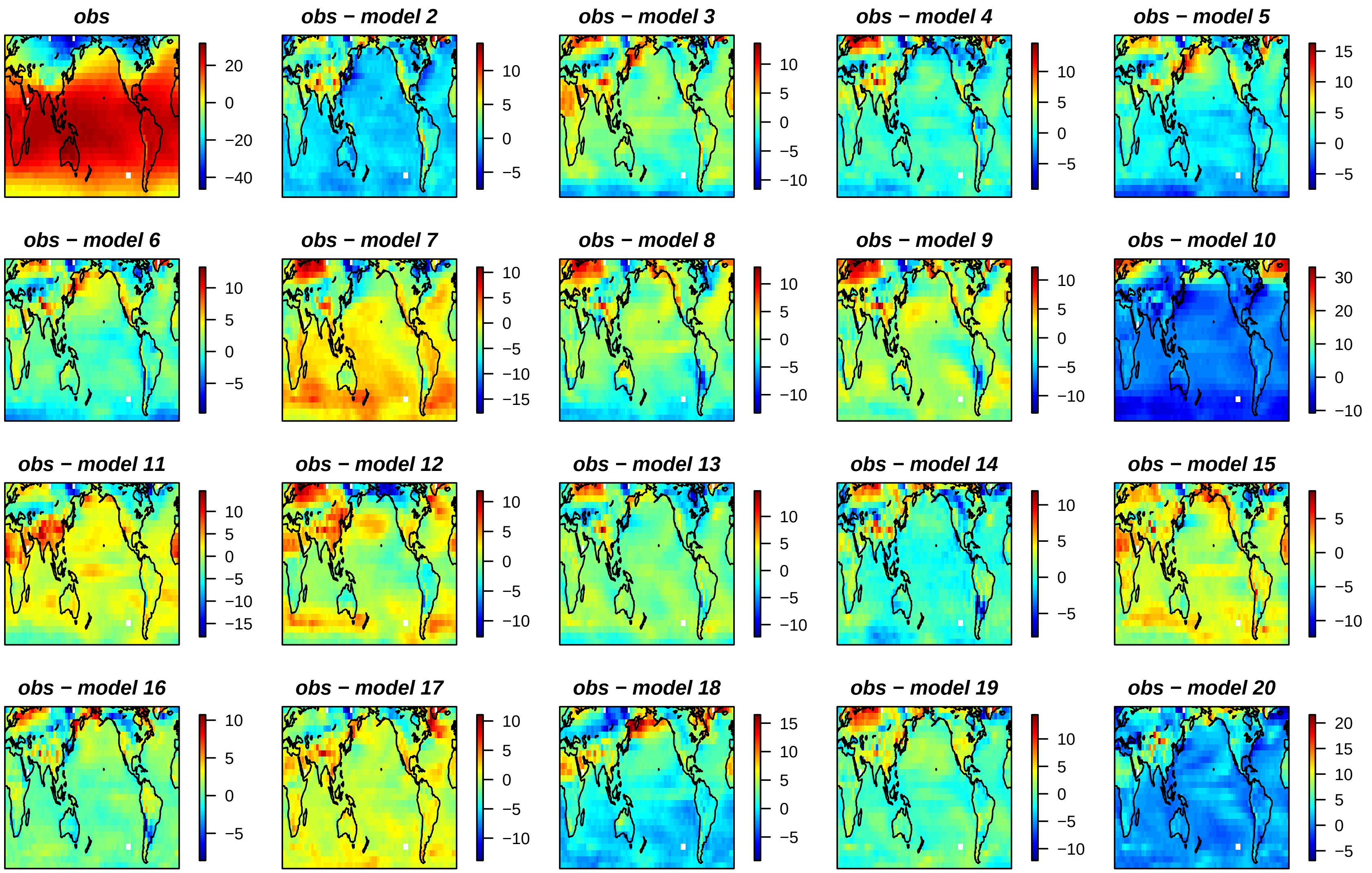
Eigenvalues of correlation matrix



Future direction

- Utilizing r_{ij}'s: classification of the climate models using correlations, weighted mean of climate model outputs
- Cross-covariance models for b_i: need further study of nonparametric covariance estimation for multivariate processes
- Joint space-time models for D_i: could use space-time covariance function in Jun and Stein (2005)
- Model resolution vs bias

30 year average of differences



MLE estimates

Model	μ_{0i}	μ_{1i}	μ_{2i}	μ_{3i}	β_i^{-1}	ν_i	α_i	δ_i
2	-0.12	0.010	-6.1e-04	1.605	0.001	0.44	0.67	1.28
3	0.38	0.029	4.6e-04	0.315	0.001	0.51	0.77	1.18
5	1.25	0.044	5.9e-04	0.486	0.001	0.48	0.73	1.23
6	0.42	0.027	9.2e-04	0.542	0.001	0.52	0.67	1.59
7	-0.48	-0.058	8.5e-04	0.178	0.001	0.32	0.67	1.45
8	-0.69	0.051	1.6e-05	-0.370	0.001	0.42	0.72	1.56
9	0.14	0.017	1.4e-04	0.622	0.001	0.50	0.79	1.31
10	0.25	0.121	-1.8e-03	-0.271	0.001	0.26	0.81	0.66
15	-0.65	-0.014	9.9e-04	0.726	0.001	0.42	0.63	1.50
16	0.67	0.017	-9.8e-04	0.675	0.002	0.76	0.66	1.73
17	0.22	-0.012	4.1e-04	0.361	0.002	0.68	0.70	1.19
18	1.52	0.073	-7.1e-04	-0.093	0.001	0.38	0.75	0.96
19	-0.40	-0.000	7.1e-04	1.799	0.000	0.21	0.48	2.07
20	0.85	0.006	1.6e-04	2.078	0.001	0.41	0.67	1.92

Comparison to simulated data