Stochastic subgrid scale modeling

Daan Crommelin

Centrum Wiskunde & Informatica, Amsterdam

joint work with Eric Vanden-Eijnden, New York University

The **parameterization problem**:

how to incorporate the effect of unresolved
(subgrid scale) processes on resolved model variables

A well-known problem in atmosphere-ocean science

Example:

Convection processes and cloud formation: scales < 1 km

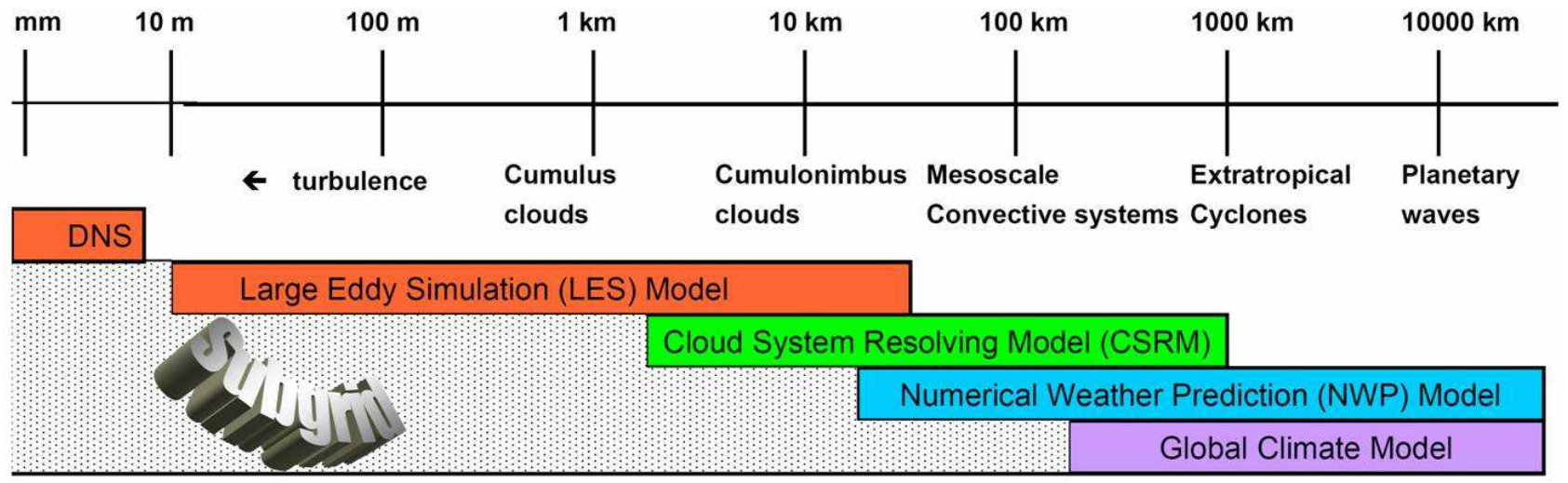
Weather forecasting models: grid scale $\sim 10 - 40$ km

Climate models: grid scale $\gtrsim 100$ km

Other parameterizations:

radiation transfer, boundary layer turbulence, surface fluxes, subgrid orography, gravity wave drag, ...

Spatial scales in atmospheric flow



Courtesy of Pier Siebesma, KNMI

Model resolution

Courtesy of Frank Selten, KNMI

Stochastic parameterization

- For given macrostate, range of subgrid scale responses instead of just one
- Subgrid scales not always diffusive or damping ("backscatter")
- Underdispersive ensembles

Incomplete model $\frac{d}{dt}X = \dot{X} = F(X, B)$

$X(t)$: state vector

$B(t)$: fluxes / subgrid scale stresses / ...

For example: $\dot{X}_k = f(X_k, X_{k\pm1}) + B_k$

k: grid point index

Parameterization / closure: need model for B

- Construct deterministic function $B = B(X)$ (systematic derivation / physical arguments / ad hoc / data driven).
- Computational approach: $B = B(Y),\ \dot{Y} = \varepsilon^{-1} g(X, Y)$. To reduce computational cost, use short simulations of Y with X fixed to produce B. Relies on $\varepsilon \ll 1$.
- Model B as stochastic process conditional on X. Data-driven: infer process from (X, B) data.

Data-driven stochastic parameterization:

Assumptions: (i) B is a Markov process
(ii) the process is conditional on X

Use conditional transition probability
$P(B(t+\Delta t) \mid B(t), X(t))$,
inferred from (X, B) data, to evolve B in time

Variations on a theme:
e.g. $P(B(t+\Delta t) \mid B(t), X(t), \dot{X}(t))$, etc.

(C. and Vanden-Eijnden 2008)

For variables on a lattice, $X = (X_1, X_2, ..., X_K)$ and $B = (B_1, ..., B_K)$:

Use $P(B_k(t + \Delta t) \mid B_k(t), X_k(t))$ instead of $P(B(t + \Delta t) \mid B(t), X(t))$ for tractability

i.e., replace K-dim process B by collection of 1-dim. processes B_k

Algorithm:

Given $X(t), B(t)$
(i) integrate $\dot{X} = F(X, B)$ from t to $t + \Delta t$ with B fixed at $B(t)$
(ii) obtain $B(t + \Delta t)$ by sampling from distribution
$P(B_k(t + \Delta t) \mid B_k(t), X_k(t))$ for each k separately

Practical implementation: small Markov chains for B_k

$\rightarrow$ Hybrid stochastic-deterministic model,
ODEs for X coupled to Markov chain(s) for B.

Test case: the Lorenz ’96 model

Nonlinear, forced-dissipative model on a 1-dim. periodic lattice.

Often used for parameterization and predictability studies in atmospheric science.

$$\begin{cases} \dot{X}_k = f(X_k, X_{k\pm1}, X_{k-2}) + B_k \\ B = B(Y) \\ \dot{Y} = \varepsilon^{-1} G(Y, X) \end{cases}$$

with $X \in \mathbb{R}^K, \;\; Y \in \mathbb{R}^{K \times J}$

X: "large scale" variables

Y: "small scale" variables

model parameters:

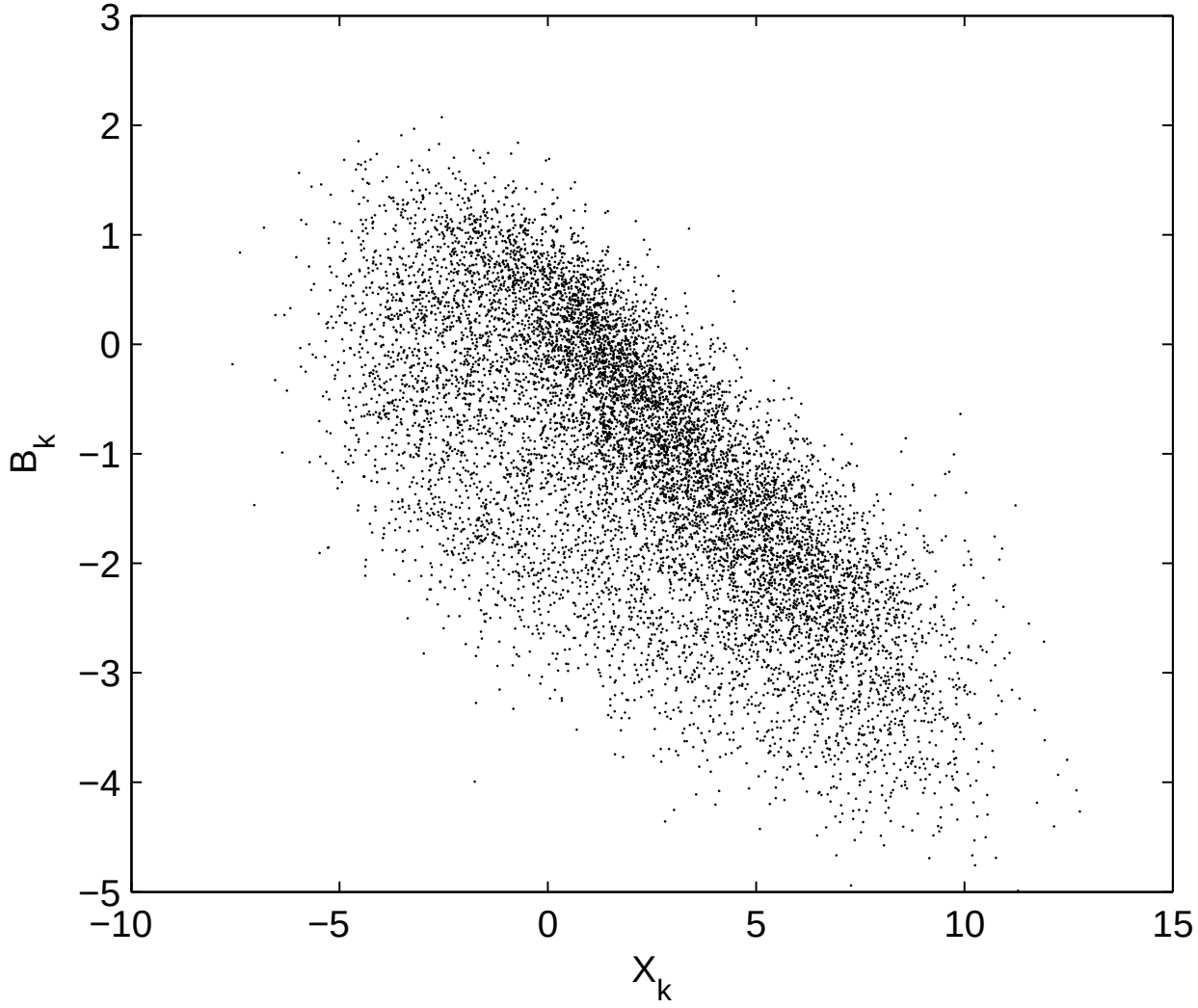
$\varepsilon = 0.5$ (i.e., no real time scale separation),

$K = 18, \quad J = 20$

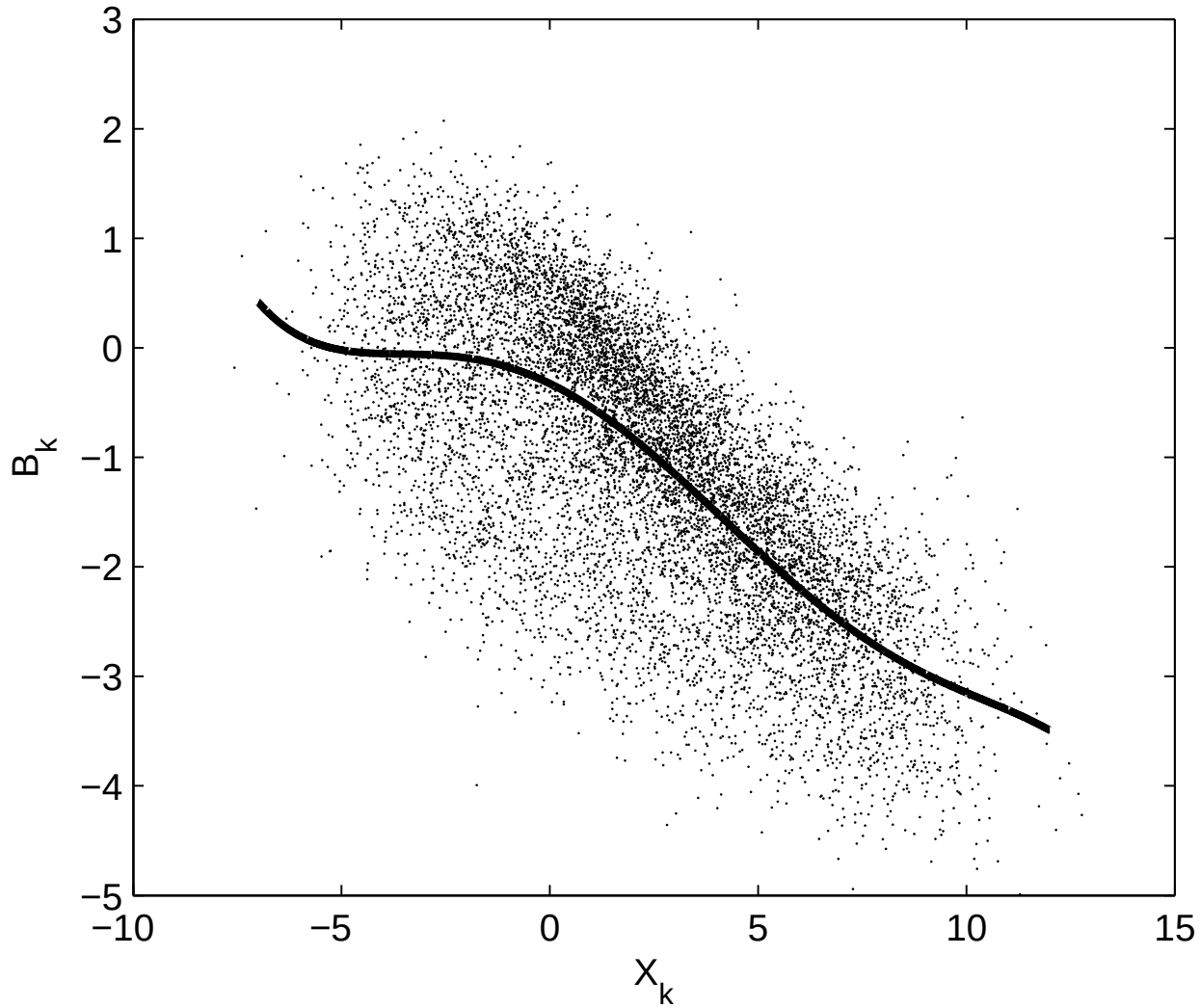
Aim:

construct closed model for X by parameterizing B

B_k versus X_k (statistically identical for all k)

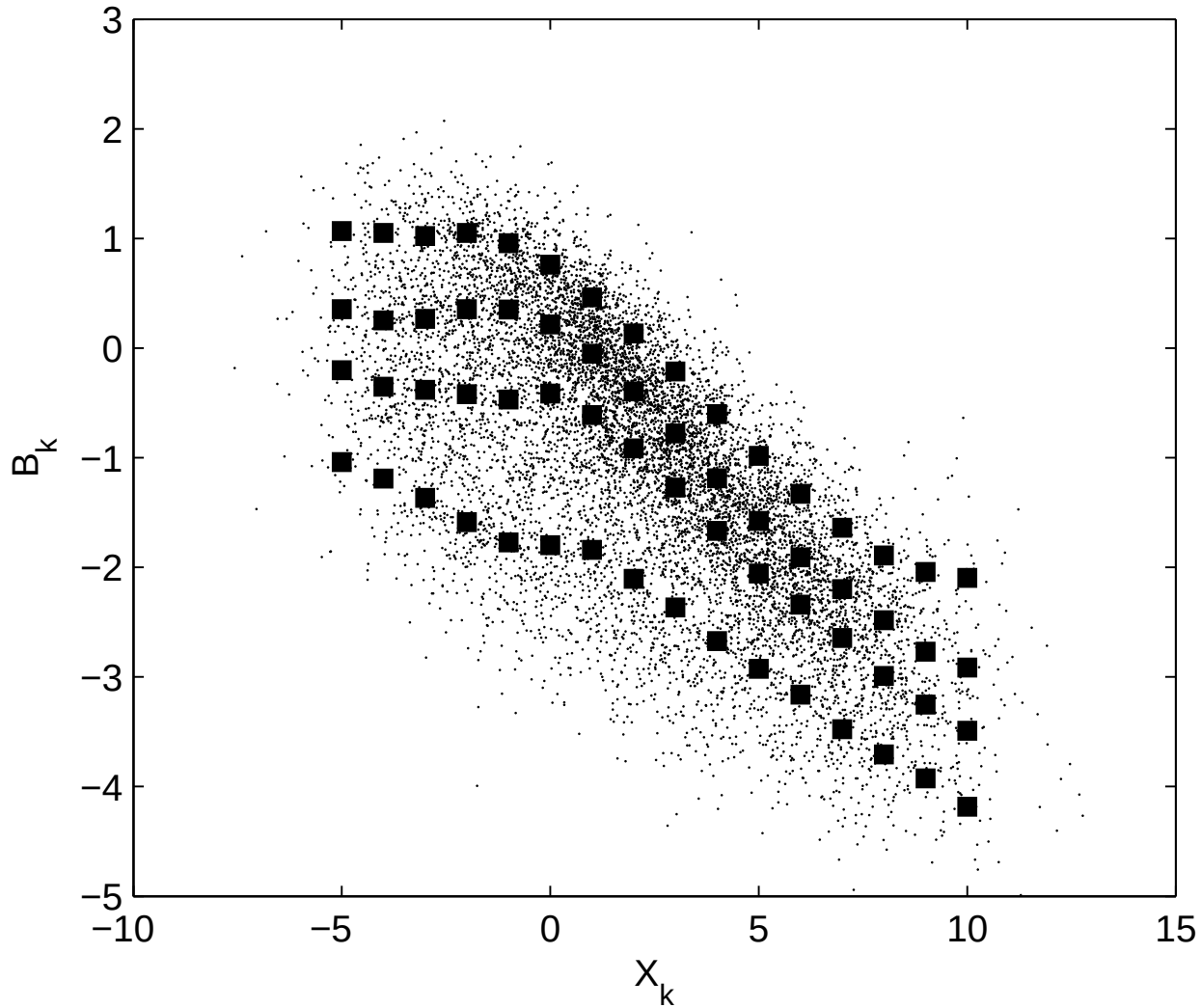


Deterministic: fit curve through (X_k, B_k) data



Stochastic: B_k jumps between ■'s,

using $\sim \; P(B_k(t+\Delta t) \mid B_k(t), X_k(t), \dot{X}_k(t))$



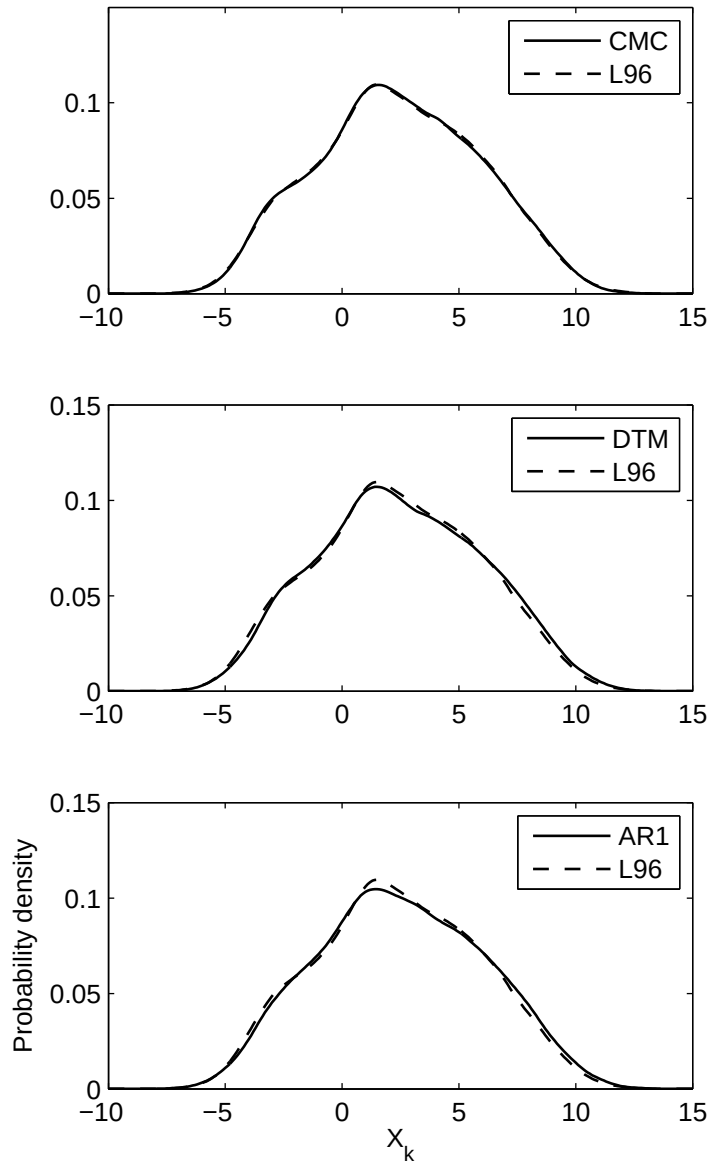
Comparison:

CMC: Conditional Markov chain scheme

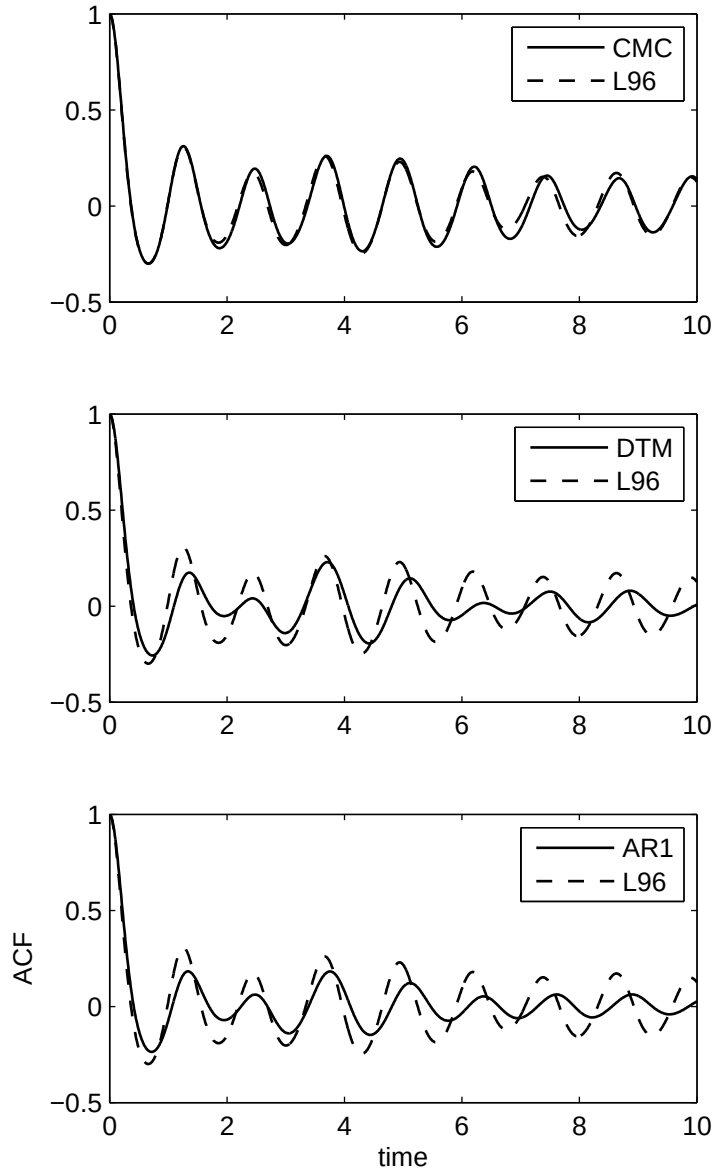
DTM: deterministic, $B_k(t) = g(X_k(t))$, curve fit for $g(X_k)$

AR1: $B_k(t) = g(X_k(t)) + \xi(t)$ with $\xi(t)$ an AR(1) process fitted to timeseries of $B_k(t) - g(X_k(t))$

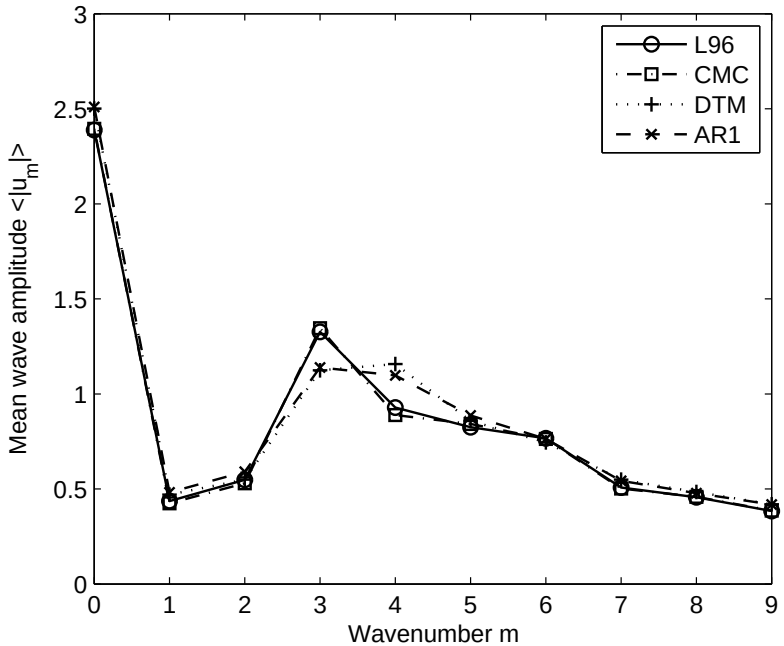
Probability density functions

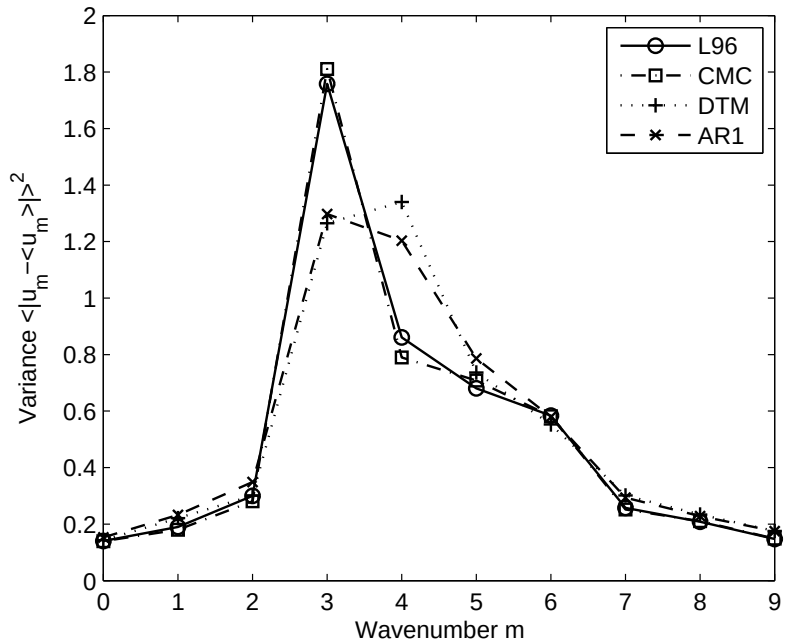


Autocorrelation functions

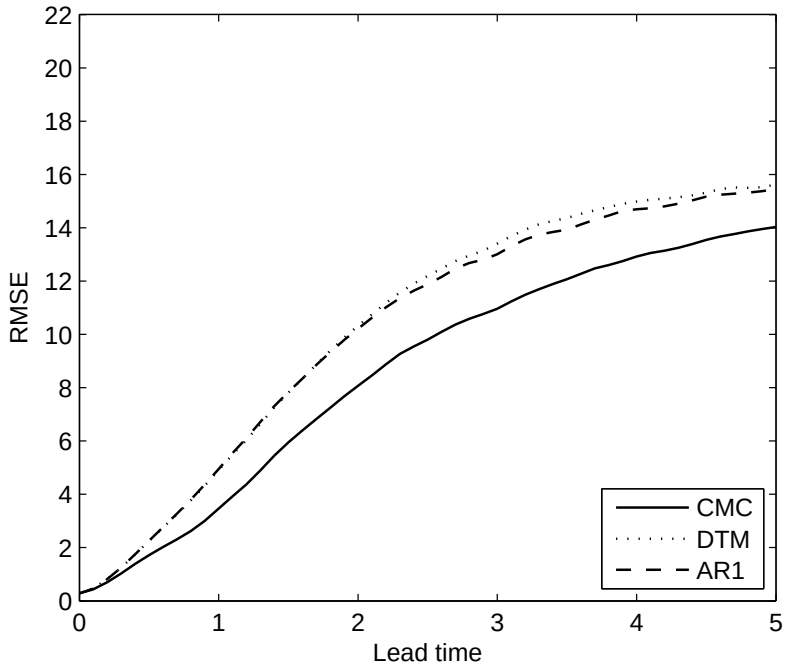


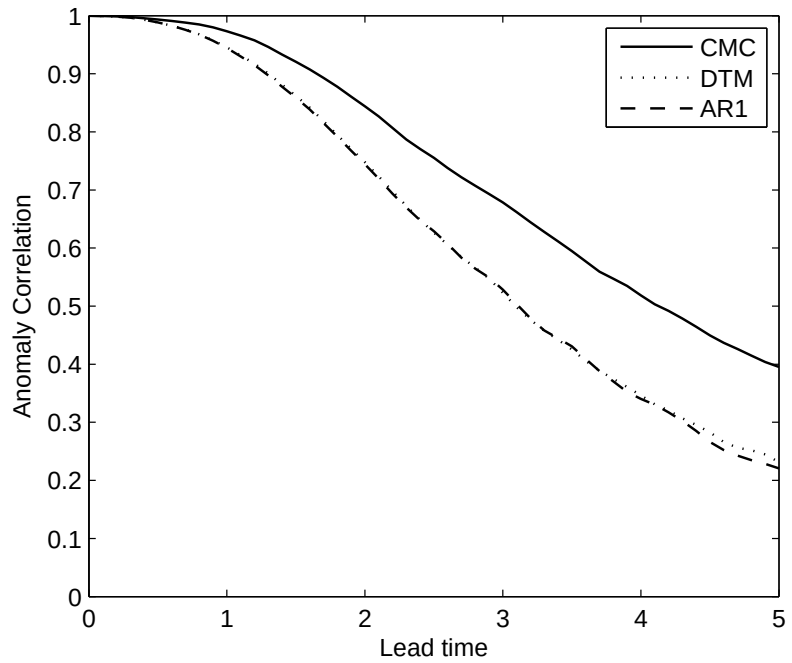
Wave amplitudes and variances ($\mathbf{X} \xrightarrow{FFT} \mathbf{U}$)



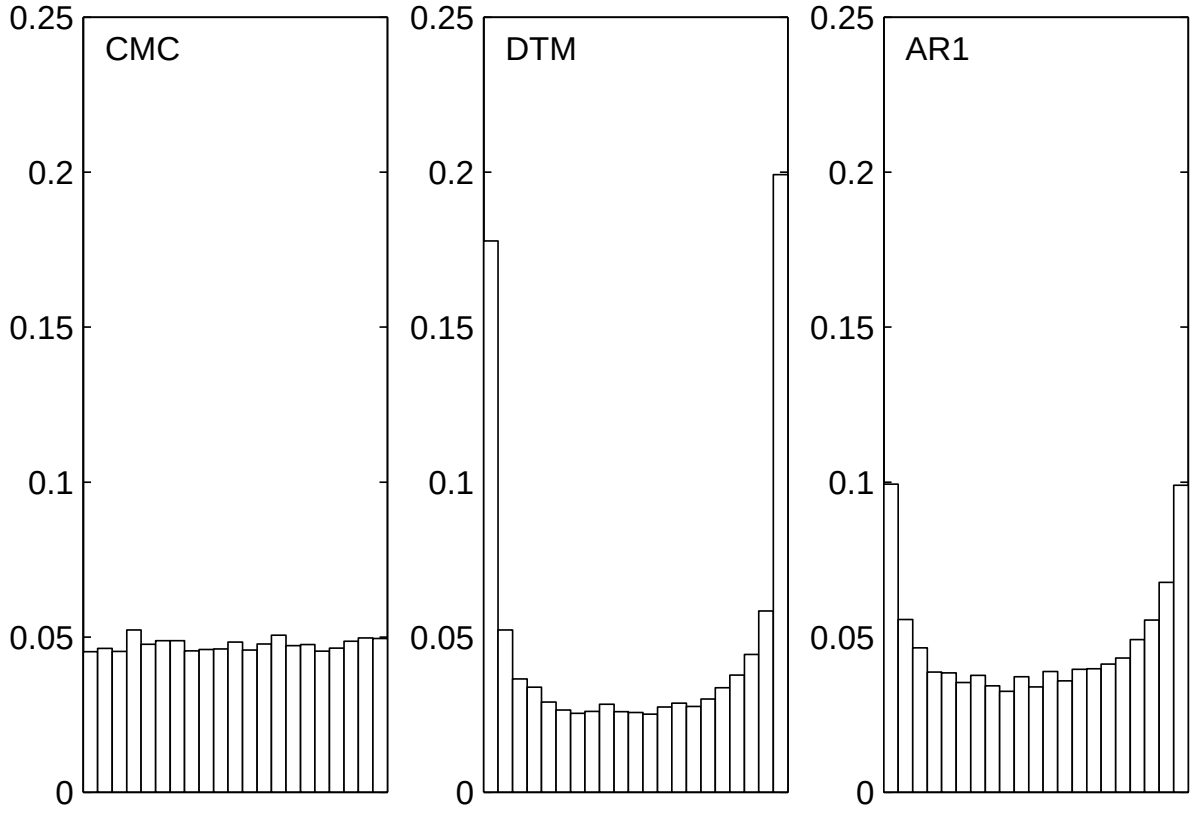


Ensemble integrations (5 members): RMSE and anomaly correlation





Ensemble integrations (20 members): Rank histogram



Outlook: Parameterization of atmospheric convection

Joint project with KNMI and TU Delft (Selten, Siebesma, Jonker)

Why convection / clouds ?
"Cloud feedbacks remain the largest source of uncertainty in climate sensitivity estimates" (IPCC 4th Assessment Report, 2007)

From L96 to realistic convection

L96: at each gridpoint k, scalar X_k and B_k

Convection: at each k, 5 functions X_k (vertical profiles for velocities, temp. and humidity) and B_k (turbulent fluxes)

Data: generated with Large Eddy Simulation (LES)