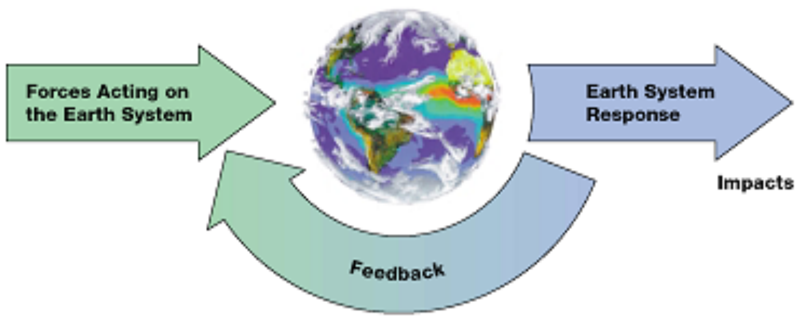
COMPARISON OF THE INFORMATION-THEORETIC METHODS TO ESTIMATE THE INFORMATION FLOW IN A DYNAMICAL SYSTEM

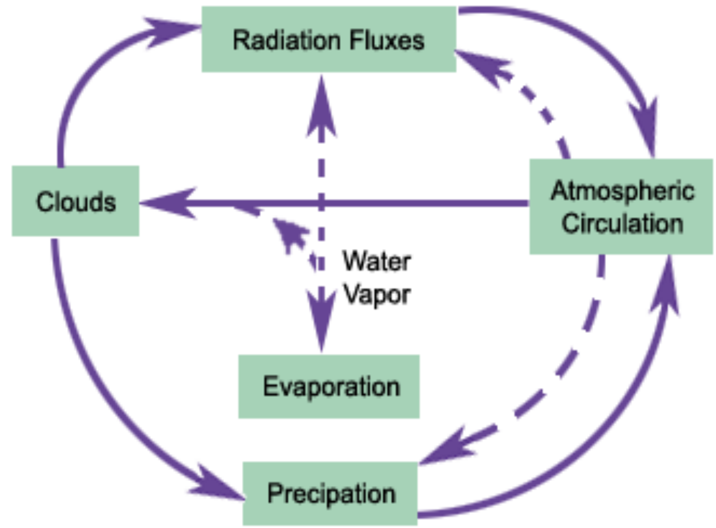
Deniz Gencaga
NOAA-CREST, CCNY, NY, USA
in collaboration with
Kevin H. Knuth and William B. Rossow

Funded by: NASA Cloud Modeling Analysis Initiative (CMAI)

Goal



(a)



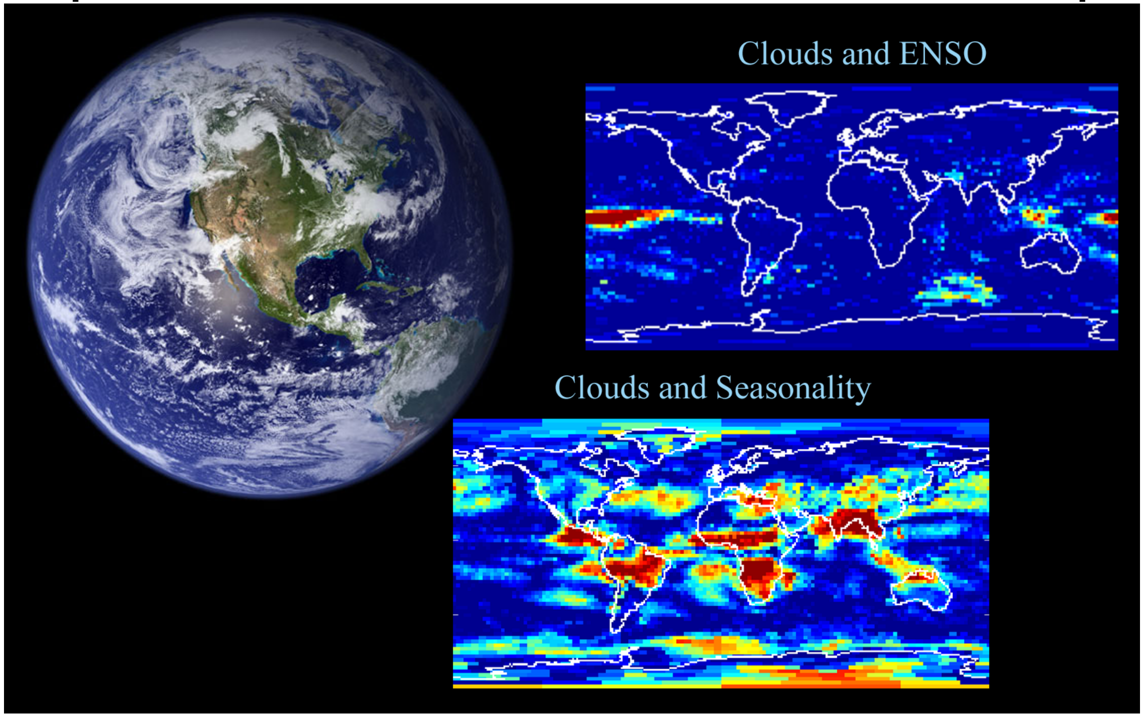
(b)

- Understand Earth climate variability
- Determine and predict the climate's response to both natural and human-induced forcing
- Understand the cause and effect relationships between different climate sub-systems

(a) NASA (b) William Rossow

Relevant Variables (Sub-systems)

- One of the greatest challenges in the area:
  - Which climate variables are related to the phenomena we wish to describe or predict?

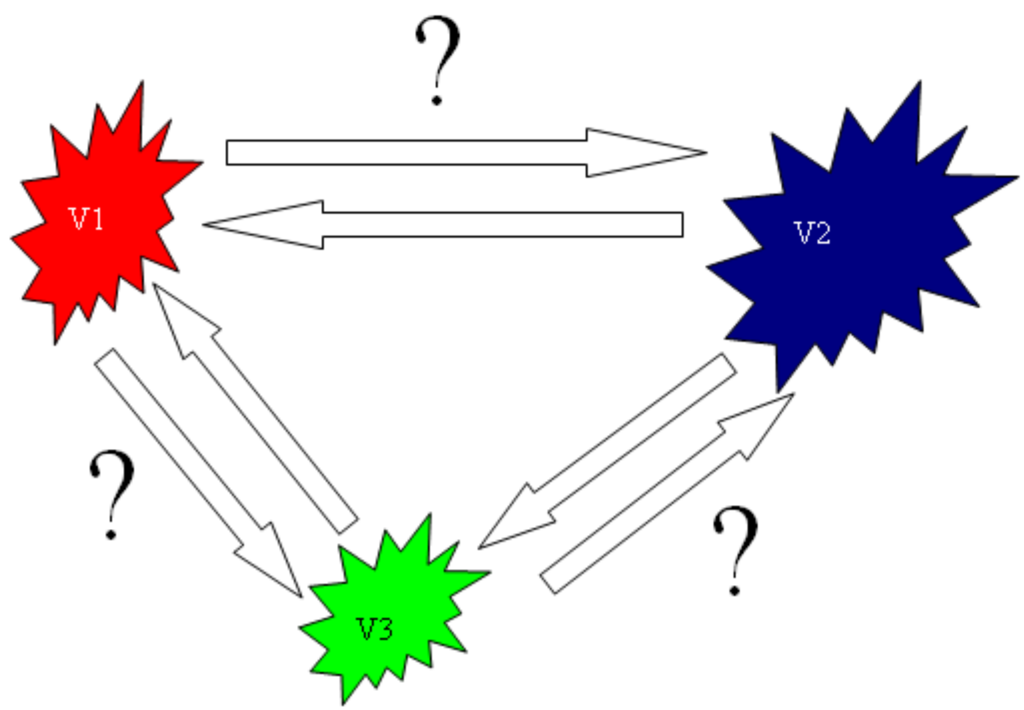


(Top right) A Mutual Information map of Cloud Cover vs. Pacific CTI (Cold Tongue Index), which indexes ENSO (El Nino Southern Oscillation), (Foreground) A Mutual Information map of Cloud Cover vs. Seasonality

Interactions between different variables

OBJECTIVES FOR A BETTER CLIMATE MODELING AND PREDICTION

1) Which variables are related?
2) What is the direction of the information flow (causal interactions)?



Outline

- Objective:
  - Identification of relevant climate variables of a physical phenomenon
  - How can we quantify the interaction between these variables and the physical process?

- Methodology:
  - Information-theoretic approaches
    - Entropy, Mutual Information and Transfer Entropy
  - Estimation of entropy and mutual information from data
    - Piecewise-constant probability density function (pdf) model
    - Continuous probability density function model
  - Estimation of Transfer Entropy methods and their comparison

- Proposed method and demonstrations
- Ongoing research

Information Theory

- To quantify the relevance of a variable to a phenomenon, we use information theory
  - Entropy is used to quantify the amount of information provided by a single variable
  - Mutual information is used to quantify what we learn about one variable from the other

Identifying Causal Interactions

- Another great challenge is to identify the ***direction*** and amount of information flow from one variable to another
    - Data-based approach: Transfer Entropy (Unknown system dynamics) (Schreiber, 2000)
    - Model based approaches (Known system dynamics) (Majda and Harlim, 2007; Liang and Kleeman, 2005)

Estimation of Information-Theoretic Quantities from data

- Entropy is not a property of the data set itself, rather it is a function of the probability density from which the data were sampled

- We propose to estimate the probability density function (pdf) of data first, then estimate the entropy by sampling from pdf

- Estimation through pdf allows us to calculate our uncertainty

(provide error bars → statistical significance of our findings)

Entropy

- Average uncertainty to find the system at a particular state 'x' out of a possible set 'X' is (Shannon) entropy:
- Discrete version: $$H(X) = -\sum_{x \in X} p(x) \log p(x)$$
- Continuous version:

$$h(x) = -\int p(x) \log\left[\frac{p(x)}{m(x)}\right] dx$$

m(x): Lebesgue measure

Entropy

- If the system is described by two parameters, we can define joint entropy which jointly describes two sub-systems:

- Discrete: $$H(X,Y) = -\sum_{x\in X}\sum_{y\in Y} p(x,y)\log p(x,y)$$

- Continuous:

$$h(x,y) = -\int\int p(x,y)\log\left[\frac{p(x,y)}{m(x,y)}\right]dxdy$$

Mutual Information

- The amount of information shared between two sub-systems X and Y:

$$I(X,Y) \quad = \quad H(X)+H(Y)-H(X,Y)$$

- Definition in terms of Kullback Leibler divergence

$$I(X,Y)=\sum_{x\in X}\sum_{y\in Y} p(x,y)\log\frac{p(x,y)}{p(x)p(y)}$$

- Statistical Independence:

$$p(x,y)=p(x)p(y)$$

$$I(X,Y)=\sum_{x\in X}\sum_{y\in Y} p(x,y)\log\frac{p(x,y)}{p(x)p(y)}=\sum_{x\in X}\sum_{y\in Y} p(x,y)\log\frac{p(x)p(y)}{p(x)p(y)}=0$$

OUR GOAL:
Transfer Entropy

- Estimation of the information flow direction between sub-systems using data under the assumption of Markov processes

- Consider two subsystems X and Y, with data in the form of a two time series of measurements

$$X = \{x_1, x_2, \cdots, x_k, x_{k+1}, \cdots, x_n\}$$
$$Y = \{y_1, y_2, \cdots, y_l, y_{l+1}, \cdots, y_n\}$$

then the transfer entropy can be written as

$$T\left(X_{i+1} \middle| \mathbf{X}_i^{(k)}, \mathbf{Y}_j^{(l)}\right) = \sum_{x_{i+1}, \mathbf{x}_i^{(k)}, \mathbf{y}_j^{(l)}} p\left(x_{i+1}, \mathbf{x}_i^{(k)}, \mathbf{y}_j^{(l)}\right) \log \frac{p\left(x_{i+1} \middle| \mathbf{x}_i^{(k)}, \mathbf{y}_j^{(l)}\right)}{p\left(x_{i+1} \middle| \mathbf{x}_i^{(k)}\right)} \qquad \mathbf{x}_i^{(k)} = \left(x_i, \ldots, x_{i-k+1}\right), \mathbf{y}_j^{(l)} = \left(y_j, \ldots, x_{j-l+1}\right)$$

which describes the degree to which information about Y allows one to predict future values of X. This is then a measure of the causal influence that the subsystem Y has on the subsystem X.

Information Flow: Information we learn from the past state of one variable about the current state of the other

TRANSFER ENTROPY (TE)

➢TE is NOT symmetric, thus provides directionality information

$$TE_{X \to Y} = T\left(Y_{i+1} \middle| \mathbf{Y}_\mathbf{i}^{(\mathbf{k})}, \mathbf{X}_\mathbf{j}^{(\mathbf{l})}\right) = \sum_{i=1}^{N} p\left(y_{i+1}, \mathbf{y}_\mathbf{i}^{(\mathbf{k})}, \mathbf{x}_\mathbf{j}^{(\mathbf{l})}\right) \log_2 \frac{p\left(y_{i+1} \middle| \mathbf{y}_\mathbf{i}^{(\mathbf{k})}, \mathbf{x}_\mathbf{j}^{(\mathbf{l})}\right)}{p\left(y_{i+1} \middle| \mathbf{y}_\mathbf{i}^{(\mathbf{k})}\right)}$$

$$TE_{Y \to X} = T\left(X_{i+1} \middle| \mathbf{X}_\mathbf{i}^{(\mathbf{k})}, \mathbf{Y}_\mathbf{j}^{(\mathbf{l})}\right) = \sum_{i=1}^{N} p\left(x_{i+1}, \mathbf{x}_\mathbf{i}^{(\mathbf{k})}, \mathbf{y}_\mathbf{j}^{(\mathbf{l})}\right) \log_2 \frac{p\left(x_{i+1} \middle| \mathbf{x}_\mathbf{i}^{(\mathbf{k})}, \mathbf{y}_\mathbf{j}^{(\mathbf{l})}\right)}{p\left(x_{i+1} \middle| \mathbf{x}_\mathbf{i}^{(\mathbf{k})}\right)}$$

where $x_i(k)=\{x_i,\ldots,x_{i-k+1}\}$ and $y_j(l)=\{y_i,\ldots,y_{i-l+1}\}$ are past states and X and Y are k^{th} and l^{th} order Markov processes

➢There are several methods in the literature, which are not totally clear on choosing *ad hoc* parameters. Thus, we propose using two existing methods and propose our own Bayesian technique and synthesize them.

METHOD1: PIECEWISE-CONSTANT MODEL FOR PROBABILITY DENSITY FUNCTION OF DATA

$$TE_{Y\to X} = T\left(X_{i+1} \mid \mathbf{X}_i^{(k)}, \mathbf{Y}_j^{(l)}\right) = H\left(\mathbf{X}_i^{(k)} \otimes \mathbf{Y}_j^{(l)}\right) - H\left(\mathbf{X}_{i+1}^{(k+1)} \otimes \mathbf{Y}_j^{(l)}\right) + H\left(\mathbf{X}_{i+1}^{(k+1)}\right) - H\left(\mathbf{X}_i^{(k)}\right)$$

where $\otimes$ is used to denote a composite process in higher dimensions.

$$x_i^{(k)} = \left(x_i, \ldots, x_{i-k+1}\right)$$

For k=l=1;

$$TE_{Y\to X} = T\left(X_{i+1} \mid X_i, Y_j\right) = H\left(X_{i+1}, X_i\right) + H\left(X_i, Y_j\right) + H\left(X_{i+1}, X_i, Y_j\right) - H\left(X_i\right)$$

Joint Shannon entropy

$$H(X,Y) = -\sum_{x\in X}\sum_{y\in Y} p(x,y)\log p(x,y)$$

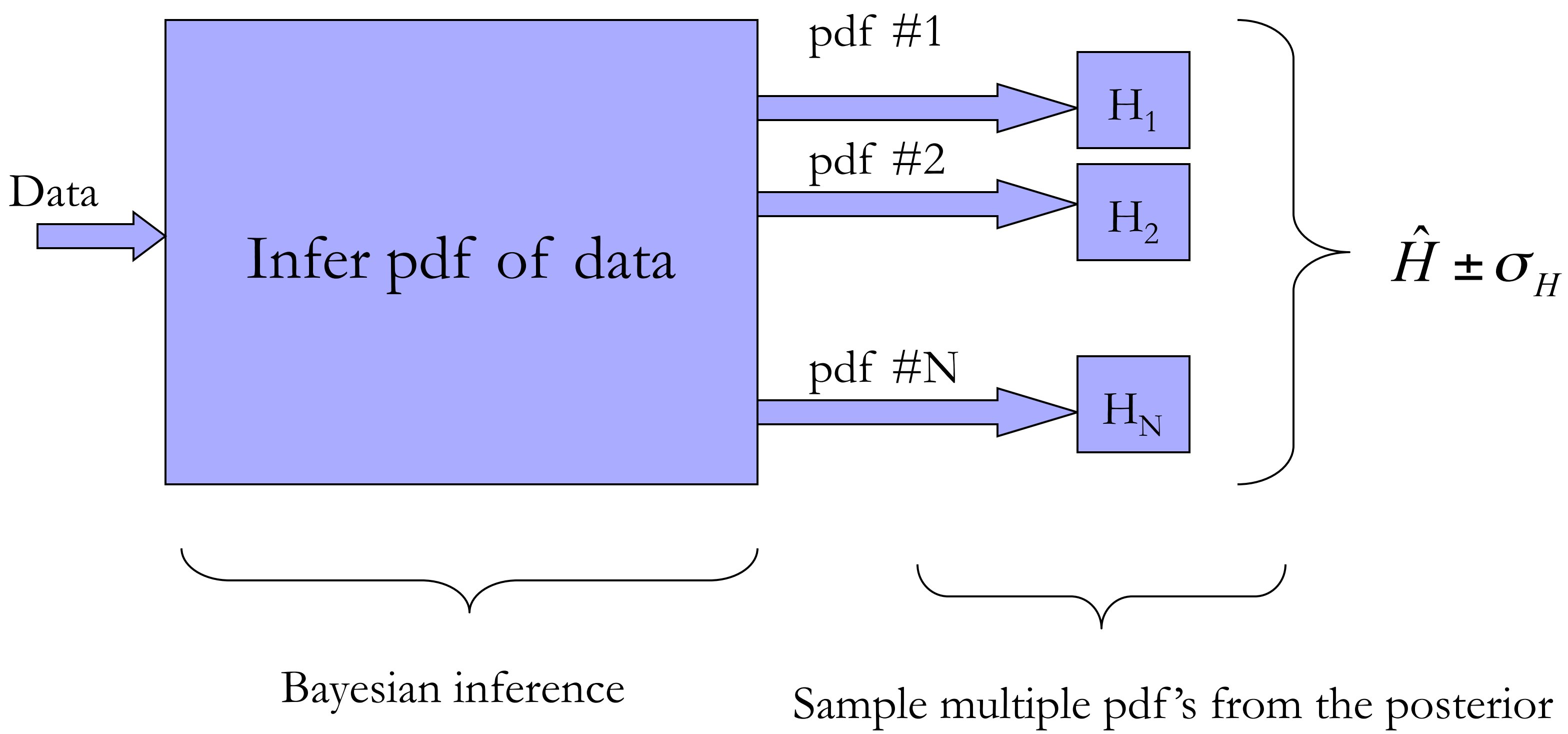
Shannon Entropy

$$H(X) = -\sum_{x\in X} p(x)\log p(x)$$

Estimate these quantities from data appropriately!

Proposal: Piecewise-constant model for pdf, then entropy from pdf

Proposed Methodology: Estimation of entropy from data



Bayesian Modeling of probability density function

- Bayesian modeling:

$$\underbrace{P(model \mid data)}_{\text{Posterior Probability}} = \underbrace{P(model)}_{\text{Prior Probability}} \frac{\overbrace{P(data \mid model)}^{\text{Likelihood}}}{\underbrace{P(data)}_{\text{Evidence}}}$$

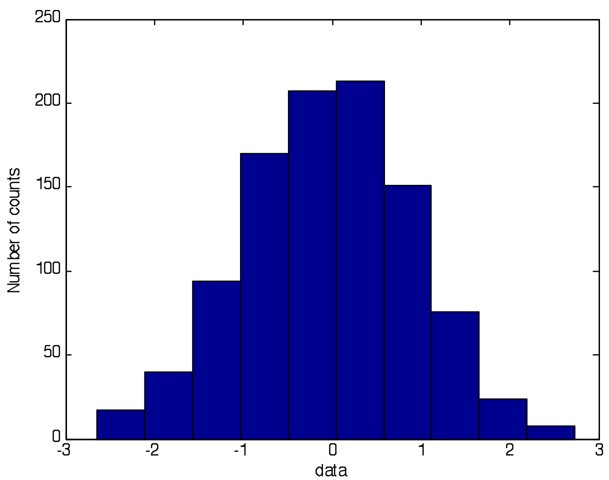
- Update our prior belief with data
- We infer the probability distribution of our model

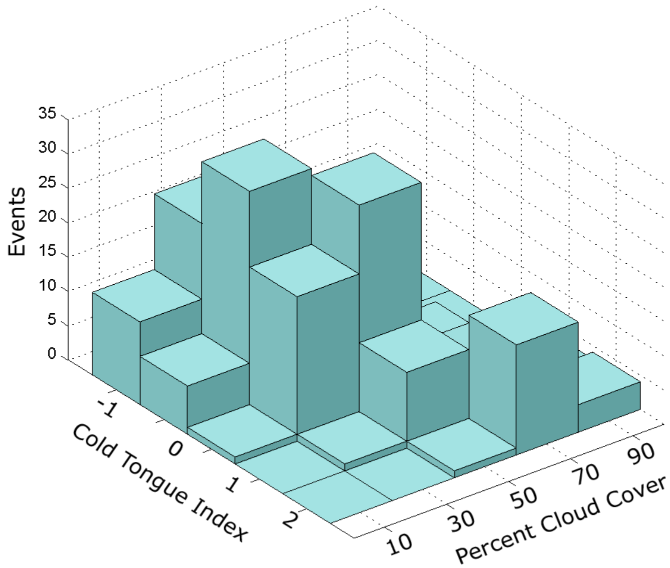
  ↓

  Superior to point estimations
- Having a probability distribution to quantify uncertainties

Piecewise-constant pdf model

- Identification of the optimum number of bins is important to describe the underlying pdf properly (Knuth et al., 2006)
- pdf estimations:



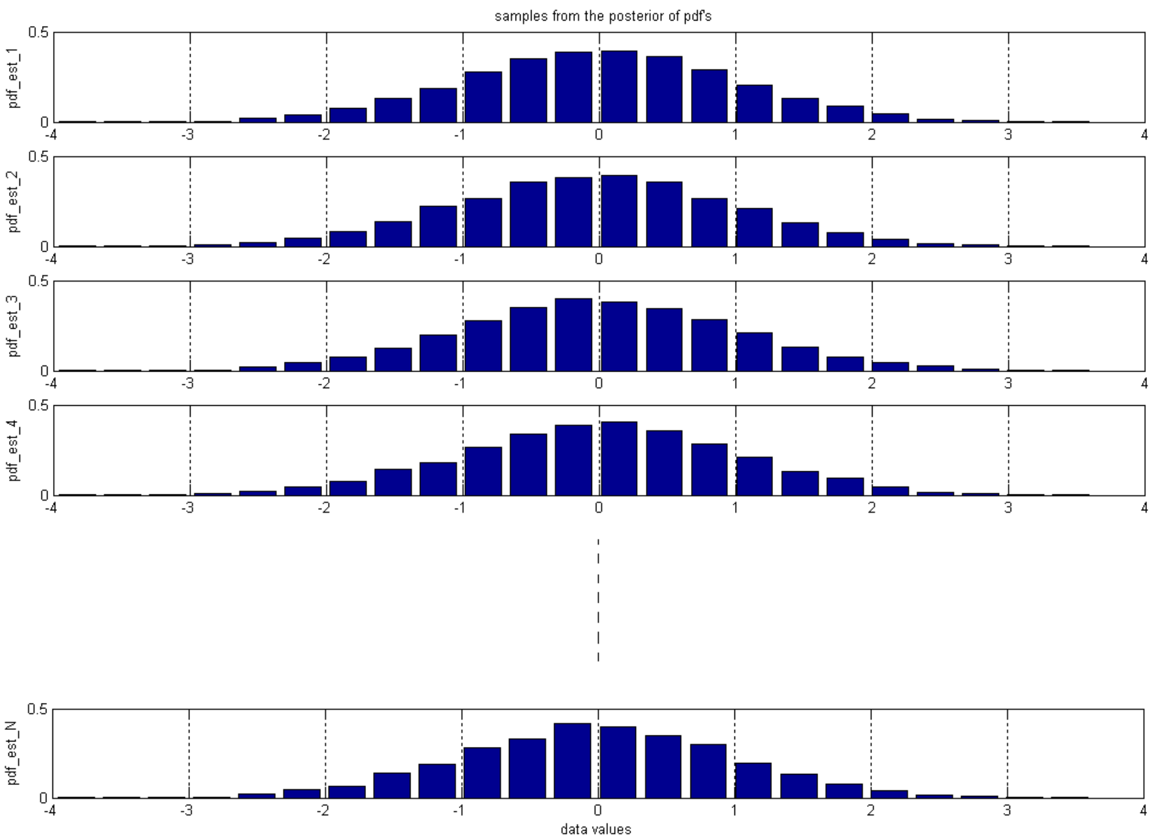


- Entropy calculations:

$$H(X) = -\sum_{x \in X} p(x) \log p(x)$$

$$H(X,Y) = -\sum_{x \in X} \sum_{y \in Y} p(x,y) \log p(x,y)$$

Estimation of entropy from data



$\rightarrow H_1$

$\rightarrow H_2$

$\rightarrow H_3$

$\rightarrow H_4$

$\rightarrow H_N$

$$\hat{H} \pm \sigma_H$$

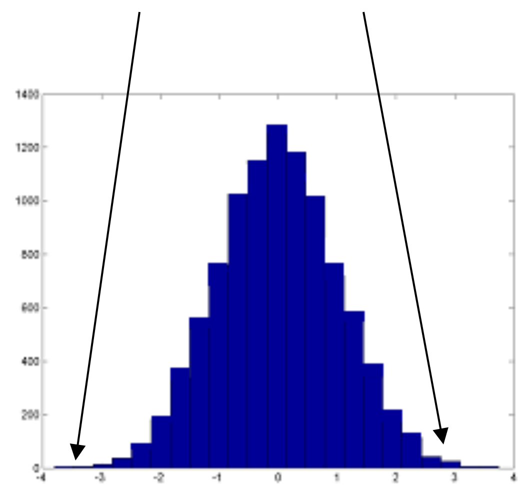
Bayesian modeling of piecewise-constant pdf : Estimation of bin probabilities:

Observed data:

$$\mathbf{d} = \{d_1, d_2, \ldots, d_N\}$$

Parameter of interest:
Bin probabilities

$$\boldsymbol{\pi} = \{\pi_1, \pi_2, \ldots, \pi_{M-1}\}$$



Prior distribution:

$$p(\boldsymbol{\pi}|M) = \frac{\Gamma\left(\frac{M}{2}\right)}{\Gamma\left(\frac{1}{2}\right)^M}\left[\pi_1\pi_2\cdots\pi_{M-1}\left(1-\sum_{i=1}^{M-1}\pi_i\right)\right]^{-1/2}$$

Likelihood function:

$$p(\mathbf{d}|\boldsymbol{\pi}, M) = \left(\frac{M}{V}\right)^N \pi_1^{n_1}\pi_2^{n_2}\cdots\pi_{M-1}^{n_{M-1}}\pi_M^{n_M}$$

Posterior distribution of each bin probability:

$$p(\pi_k|d, M) = \frac{\Gamma\left(N+\frac{M}{2}\right)}{\Gamma\left(n_k+\frac{1}{2}\right)\Gamma\left(N-n_k+\frac{M-1}{2}\right)}\pi_k^{n_k-\frac{1}{2}}(1-\pi_k)^{N-n_k+\frac{M-3}{2}} \rightarrow \langle\pi_k\rangle = \frac{n_k+1/2}{N+M/2}, \quad k = 1,\ldots,M$$

Number of elements in the kth bin

Total number of elements in data set

Number of bins

Finding the optimal bin number

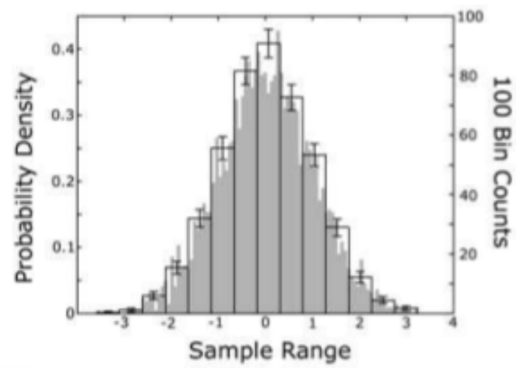
➢Find the joint posterior of bin heights and bin number and maximize this

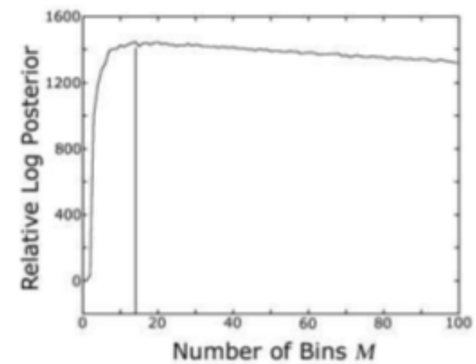
$$p(\pi, M \mid \mathbf{d}) \propto p(\pi)\, p(M)\, p(\mathbf{d} \mid \pi, M)$$

$$p(M \mid \mathbf{d}) \propto \int_{\pi} p(\pi)\, p(M)\, p(\mathbf{d} \mid \pi, M)\, d\pi^{m-1}$$

$$= \left(\frac{M}{V}\right)^{N} \frac{\Gamma\left(\frac{M}{2}\right)}{\Gamma\left(\frac{1}{2}\right)^{M}} \frac{\prod_{k=1}^{M} n_k + \frac{1}{2}}{\Gamma\left(N + \frac{M}{2}\right)}$$

$$\hat{M} = \underset{M}{\arg\max}\left\{\log p(M \mid \mathbf{d})\right\}$$





Finding the optimal bin number: General optBINS

$$p(\pi \mid M) = \frac{\Gamma\left(\sum_{i=1}^{M} \beta\right)}{\prod_{i=1}^{M} \Gamma(\beta)} \left[\pi_1 \pi_2 \ldots \pi_{M-1} \left(1 - \sum_{i=1}^{M-1} \pi_i\right)\right]^{\beta - 1}$$

$$p(M \mid \mathbf{d}) \propto \left(\frac{M}{V}\right)^N \frac{\Gamma(M\beta)}{\Gamma(\beta)^M} \frac{\prod_{k=1}^{M} (n_k + \beta)}{\Gamma(N + M\beta)}$$

$$\langle \pi_k \rangle = \frac{n_k + \beta}{N + M\beta}, \qquad k = 1, \ldots, M$$

Wolpert & Wolf, 1995
(for ***fixed number of bin*** optimizations-sample effect examinations)

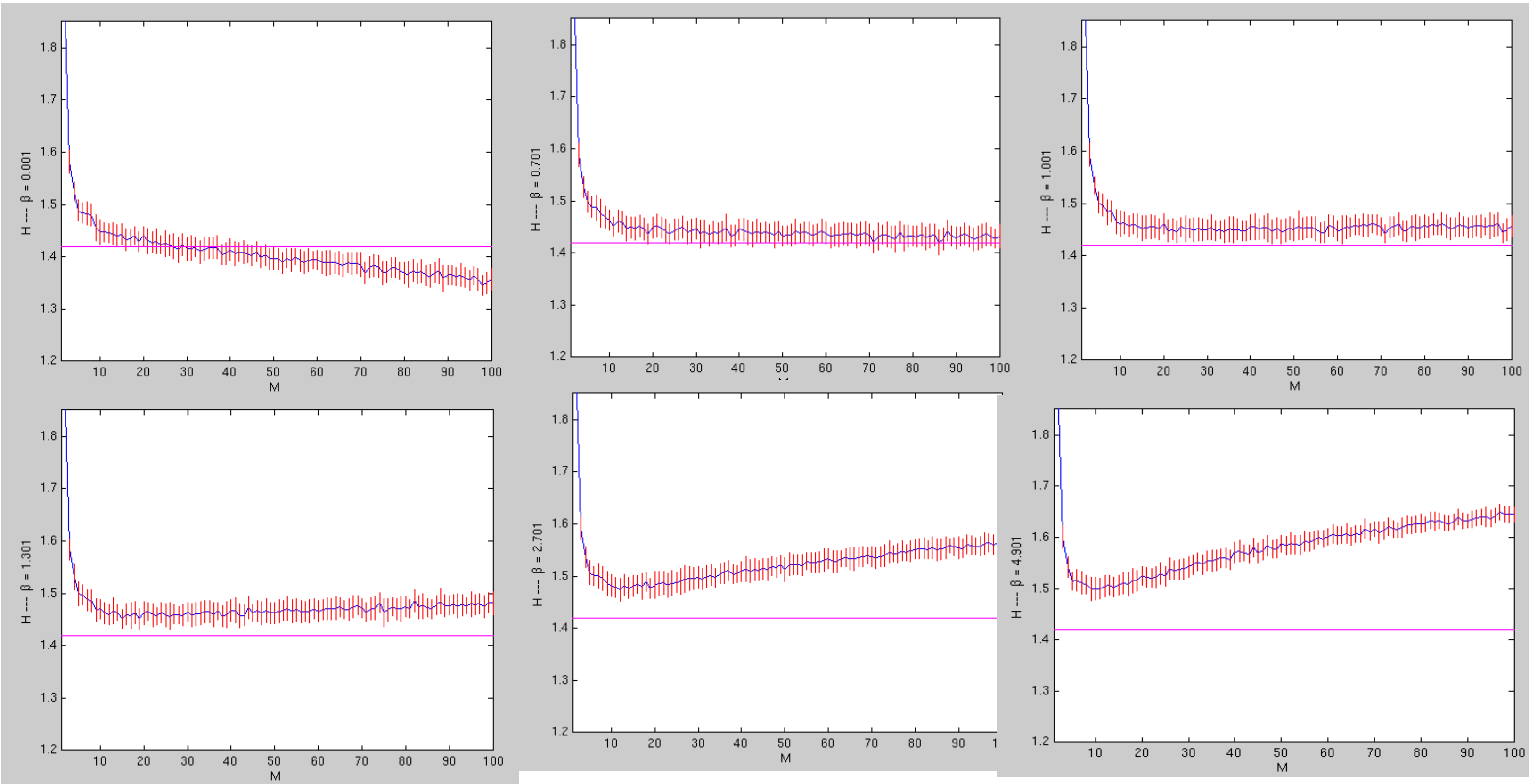
Average bin height
(artificial effect can be decreased by small beta)

$$\hat{M} = \arg\max_{M} \{\log p(M \mid \mathbf{d})\}$$

Optimal bin number

SMALL BETA IS GOOD FOR HIGH DIMENSIONS WHEN BINS ARE EMPTY AND FILLED ARTIFICIALLY

BETA EFFECT



Estimation of the entropy of a Gaussian distributed data using different beta values

H_{true}=1.4189

Piecewise constant model

- We obtain a posterior pdf for each bin probability
- This allows us to sample from this distribution
- We can report our uncertainty in our pdf estimation
- Therefore, we can summarize our uncertainty in entropy estimation by computing its mean and standard deviation from these samples

Disadvantages of piecewise-constant model

- Uniform densities of the bins cause small biases in the entropy estimates
- These biases increase as we go to higher dimensions
- Mutual information and transfer entropy need entropy estimations in higher dimensions, respectively

Advantages of piecewise-constant model

Computationally very fast

Continuous pdf model

- In order to reduce biases arising due to local uniform approximations of histograms, we adopt a continuous (Mixture of Gaussians (MoG)) pdf model:

$$p(x \mid M, A, \mu, \sigma) = \frac{1}{Z(A,\mu,\sigma)} \sum_{k=1}^{M} A_k \exp\left[-\frac{(x-\mu_i)}{2\sigma_i^2}\right]^2$$

- Multivariate version:

$$p(x \mid M, A, \mathbf{m}, \Sigma) = \frac{1}{Z(A,\mathbf{m},\Sigma)} \sum_{k=1}^{M} A_k \exp\left[-\frac{1}{2}[x-\mathbf{m}_i]\Sigma^{-1}[x-\mathbf{m}_i]^T\right]$$

Continuous pdf model

- Estimate the unknown parameters of the MoG model using a Bayesian approach (Nested Sampling, Skilling, 2004)

- Having obtained the pdf models from data, entropy for each model can be computed by integrating h(x) numerically.

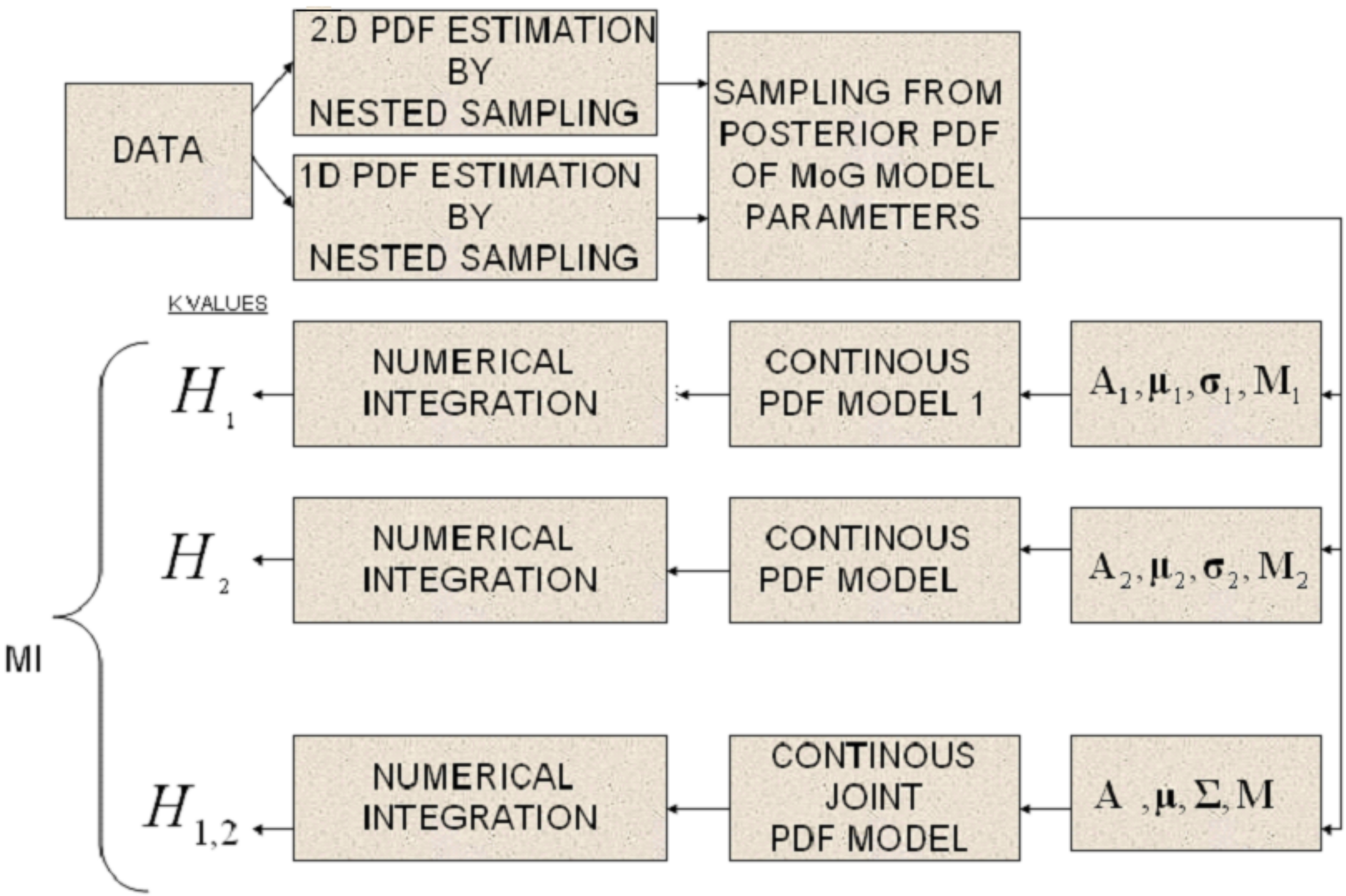
$$h(x) = -\int p(x) \log\left[\frac{p(x)}{m(x)}\right] dx$$

- Weighting each entropy computation by the probability of the model enables us to compute the mean entropy and the error bars.

- To estimate MI between different variables, the marginal and joint H(X,Y) entropies are estimated as described above and used as follows:

$$MI(X,Y) = H(X) + H(Y) - H(X,Y)$$

Estimation of entropies using continuous pdf model



Experiments

Piecewise continuous	***Mixture of Gaussians***

Procedure

1. Choose a pdf model and infer the probability distribution for the model parameters
2. Sample ~60 pdf's from the posterior distribution
3. Compute the entropy of each pdf
4. Find mean and standard deviation of entropy estimates

Experiment 1

1D Gaussian data with N = 1000 data samples (D=1, N=1000)

True entropy value is 1.4189

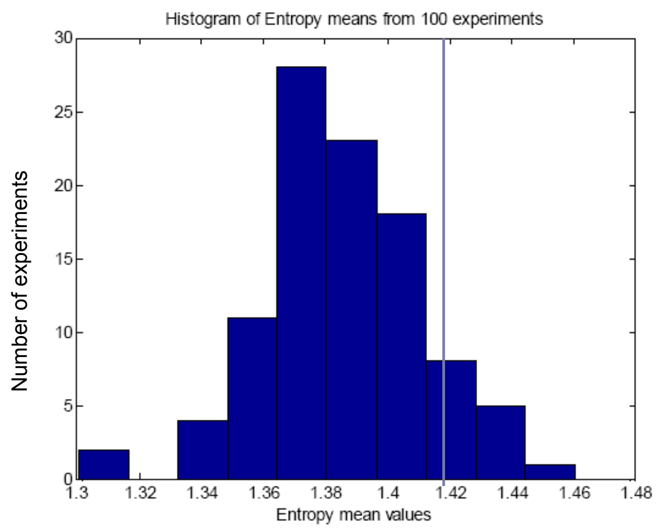
Piecewise-constant: 1.3958 ± 0.0227

Continuous: 1.4148 ± 0.0215

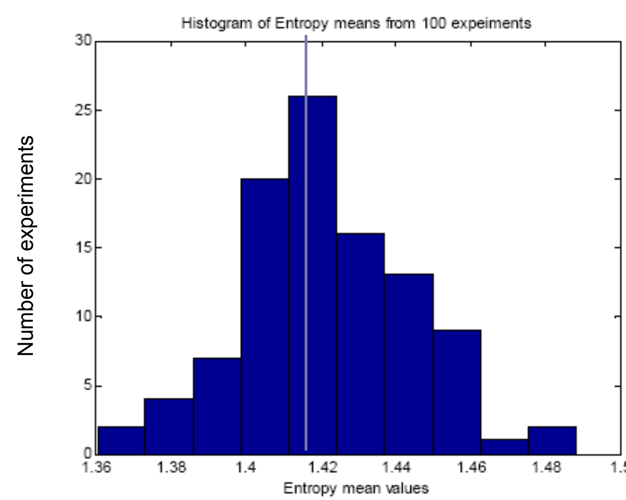
Statistics of 100 experiments

Model	Estimates within σ	Estimates within 2σ
Piecewise-constant	29%	62%
Mixture of Gaussians	68%	96%

Piecewise-constant:



Continuous:



Experiment 2

Results of 100 experiments on a 2D skewed Gaussian
(D=2, N=1000)
True joint entropy = 2.6940

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix} \right)$$

Piecewise constant model

$$h_X = 1.9819 \pm 0.0025$$
$$h_Y = 1.7935 \pm 0.0022$$
$$h_{XY} = 3.0856 \pm 0.0090$$

Continuous model

$$h_X = 1.4499 \pm 0.0214$$
$$h_Y = 1.4222 \pm 0.0216$$
$$h_{XY} = 2.7341 \pm 0.0295$$

Note that the piecewise constant model is extremely biased in the two dimensional case.

Method 2: Adaptive partitioning of the sample space

- On slide (11), MI is calculated as a sum of individual entropies. This brings artifacts. Grassberger has corrections for this but it is not satisfactory.

- Using fixed bin size for the histogram to estimate pdf for individual Shannon entropies in (11) is done a lot in the literature. Fraser and Swinney (1986) and Darbellay and Vajda (1999) shows that it is not effective and propose adaptive bin widths to estimate mutual information.

- Once we estimate mutual information, TE can be estimated as follows (Kaiser & Schreiber, 2002):

$$TE\left(X_{i+1}\middle|\mathbf{X}_{\mathbf{i}}^{(\mathbf{k})},\mathbf{Y}_{\mathbf{j}}^{(\mathbf{l})}\right)=MI\left(\mathbf{X}_{\mathbf{i}}^{(\mathbf{k})}\otimes\mathbf{Y}_{\mathbf{j}}^{(\mathbf{l})},X_{i+1}\right)-MI\left(\mathbf{X}_{\mathbf{i}}^{(\mathbf{k})},X_{i+1}\right)$$

- Here, we utilized the algorithm of Darbellay and Vajda (1999) coded by Petr Tichavsky and openly available on his web site.

Method 3: Kernel Density Estimation (KDE) methods

- We do not divide the data range into certain bins
- A kernel is placed on top of each data point
- Even if rectangular kernel is utilized, the resulting entropy estimation becomes more successful (Prichard & Theiler, 1995)
- TE can be written as follows:

$$TE_{Y\to X} = T\left(X_{i+1}\middle|\mathbf{X}_{\mathbf{i}}^{(\mathbf{k})},\mathbf{Y}_{\mathbf{j}}^{(\mathbf{l})}\right) = \sum_i p\left(x_{i+1},\mathbf{x}_{\mathbf{i}}^{(\mathbf{k})},\mathbf{y}_{\mathbf{i}}^{(\mathbf{l})}\right)\log\frac{p\left(x_{i+1},\mathbf{x}_{\mathbf{i}}^{(\mathbf{k})},\mathbf{y}_{\mathbf{i}}^{(\mathbf{l})}\right)p\left(\mathbf{x}_{\mathbf{i}}^{(\mathbf{k})}\right)}{p\left(x_{i+1},\mathbf{x}_{\mathbf{i}}^{(\mathbf{k})}\right)p\left(\mathbf{x}_{\mathbf{i}}^{(\mathbf{k})},\mathbf{y}_{\mathbf{i}}^{(\mathbf{l})}\right)}$$

- Probabilities are estimated using generalized correlation sums from data as follows:

$$p_\varepsilon\left(x_{i+1},\mathbf{x}_{\mathbf{i}}^{(\mathbf{k})},\mathbf{y}_{\mathbf{i}}^{(\mathbf{l})}\right) \cong \frac{1}{N}\sum_{\substack{j\\ i\neq j}}\Theta\left(\varepsilon - \left\|\begin{matrix} x_{i+1} - x_{j+1} \\ x_i^{(k)} - \mathbf{x}_{\mathbf{j}}^{(\mathbf{k})} \\ y_i^{(l)} - \mathbf{y}_{\mathbf{j}}^{(\mathbf{l})}\end{matrix}\right\|\right) = C(\varepsilon)$$

HOW TO CHOOSE???

$\Theta(x>0)=1;\quad \Theta(x<0)=0$ is the Heaviside function.

radius

Problem: How to pick up the appropriate radius?

- Plot $\log(\varepsilon)$ *vs* $\log C(\varepsilon)$ curve using the Grassberger-Procaccia algorithm and find a linear section if possible
  (we observed the results are too sensitive!)

- Sabesan et al. , 2007 propose using MI of destination signal, then take the k value making its first minimum.

- So, couple of $\log(\varepsilon)$ *vs* $\log C(\varepsilon)$ curves are compared and intersecting regions are found and min. radius is picked

Example: Radius choosing:

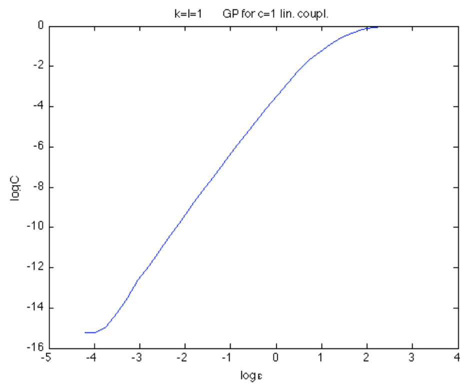
- Bivariate linearly coupled autoregressive signals:

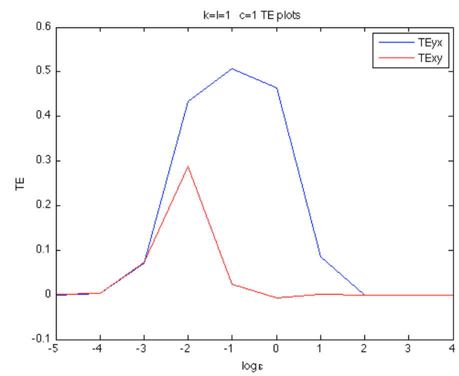
$$y(i+1) = 0.5\,y(i) + n_1(i) \qquad n_1 \sim N(0,1)$$

$$x(i+1) = 0.6\,x(i) + e\,y(i) + n_2(i) \qquad n_2 \sim N(0,1)$$

$$e \in [0.01, 1]$$

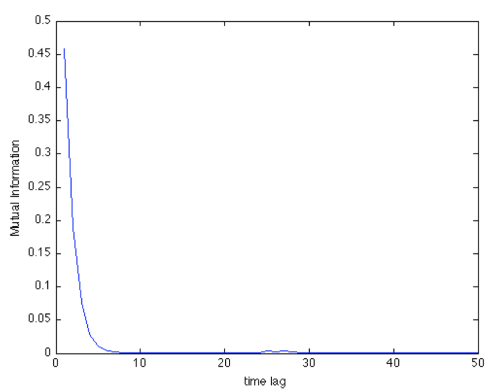
Grassberger Procaccia

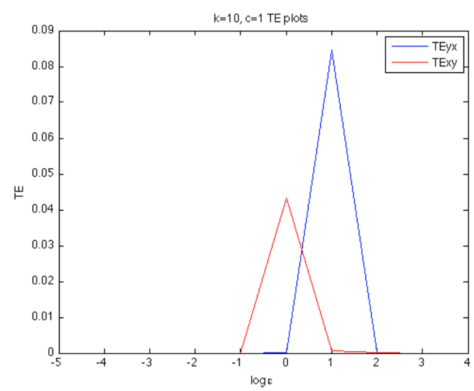




k=l=1

Time-lagged Mutual information



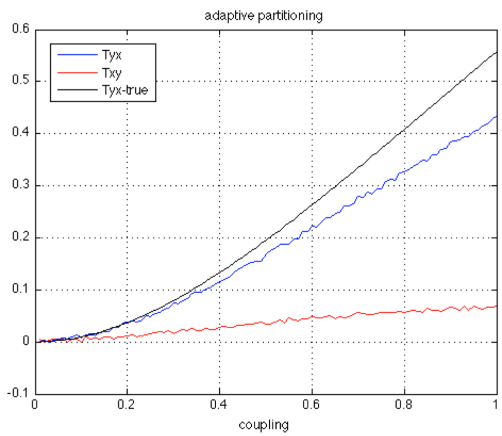


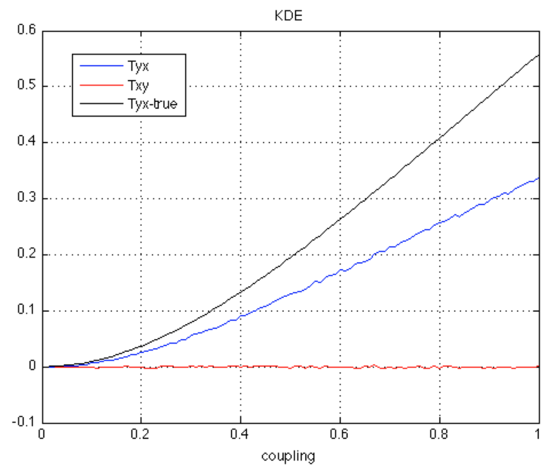
k=10;l=1

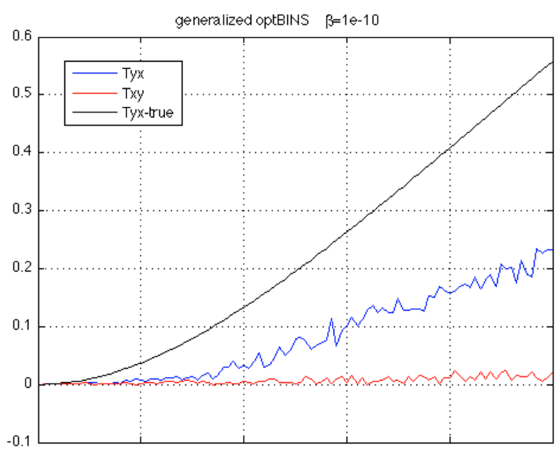
$$\log(\varepsilon) = 0.37$$

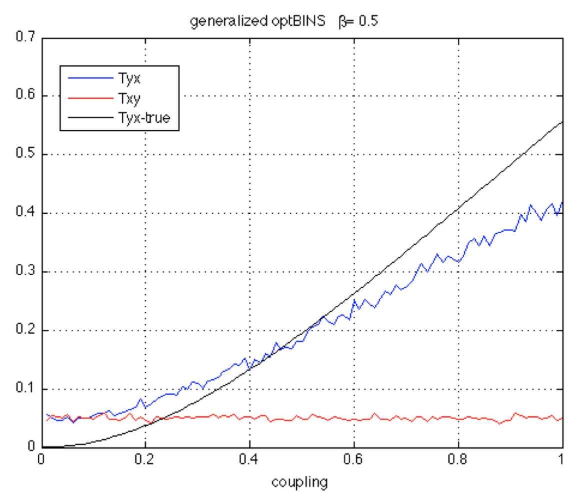
Test of methods on the bivariate coupled case

10 realization average









As long as beta is small directionality is OK, but we have a magnitude problem.

Experiments

- We would like to explore the information flow between the components of a Lorenz system given by the following nonlinear equations:

$$\frac{dx}{dt} = \sigma(y - x)$$

$$\frac{dy}{dt} = -xz + rx - y$$

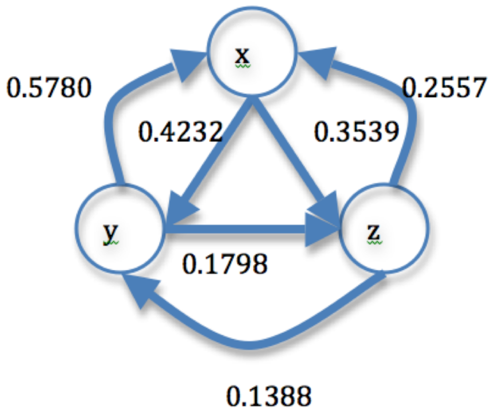
$$\frac{dz}{dt} = xy - bz$$

$\sigma = 10; \quad b = 8/3; \quad r : Rayleigh\ \ number$

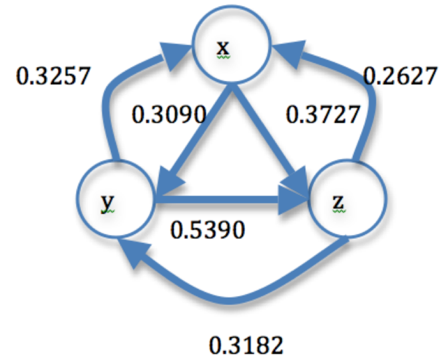
Model of an atmospheric convection roll
x – convective velocity
y – vertical temperature difference
z – mean convective heat flow

Chaotic Regime (r=28)

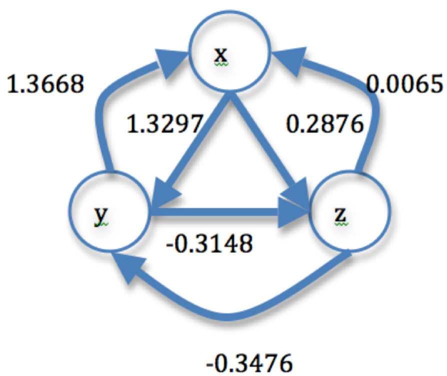
KDE

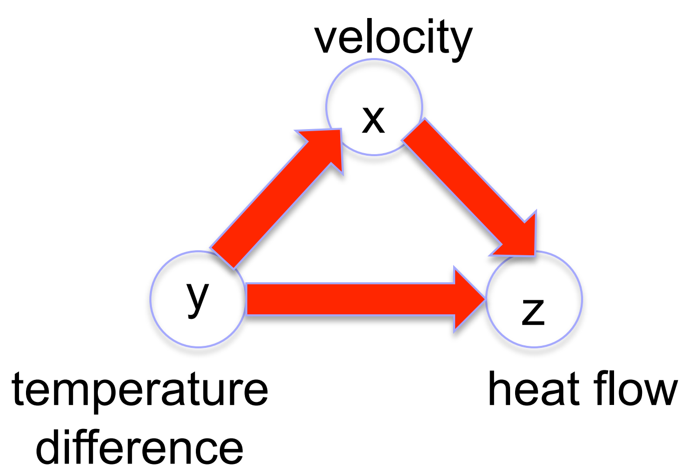


General optBINS



Adaptive partitioning method

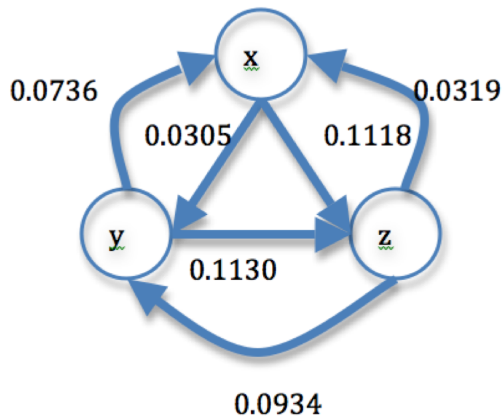




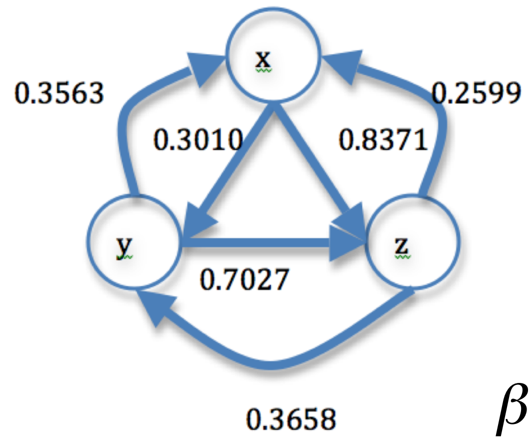
All directions between pairs AGREE!

Subchaotic regime (r=24)

KDE

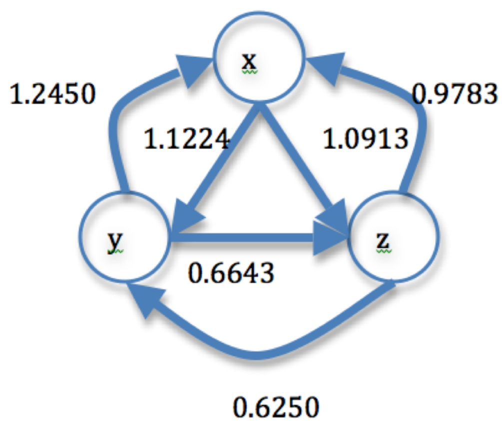


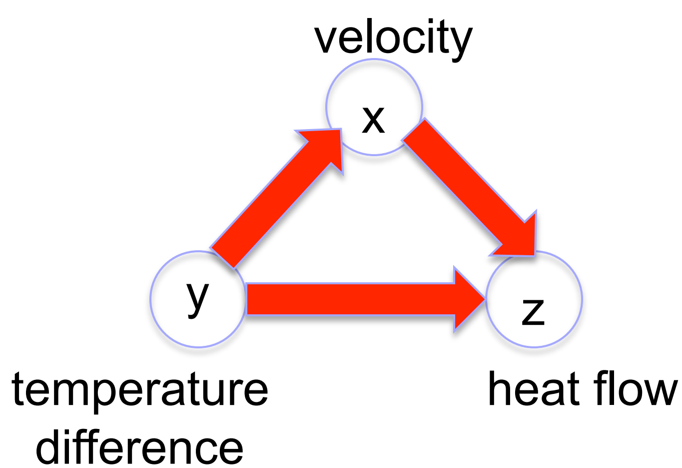
Generalized optBINS



$\beta = 0.1$

Adaptive partitioning method





All directions between pairs AGREE!

Applications to
International Satellite Cloud Climatology (ISCCP) data:
Mutual Information between
Cloud Cover and the Cold Tongue Index

- Mutual Information map between ISCCP percent cloud cover and Cold Tongue Index (CTI), which describes the sea surface temperature anomalies in the eastern equatorial Pacific Ocean (6N-6S, 180-90W) and is indicative of El Nino Southern Oscillation (ENSO) variability.

- Cloud cover data is from ISCCP climate summary product C2 (Schiffer and Rossow, 1983; Rossow and Schiffer, 1999) and CTI data is from T. Mitchell:

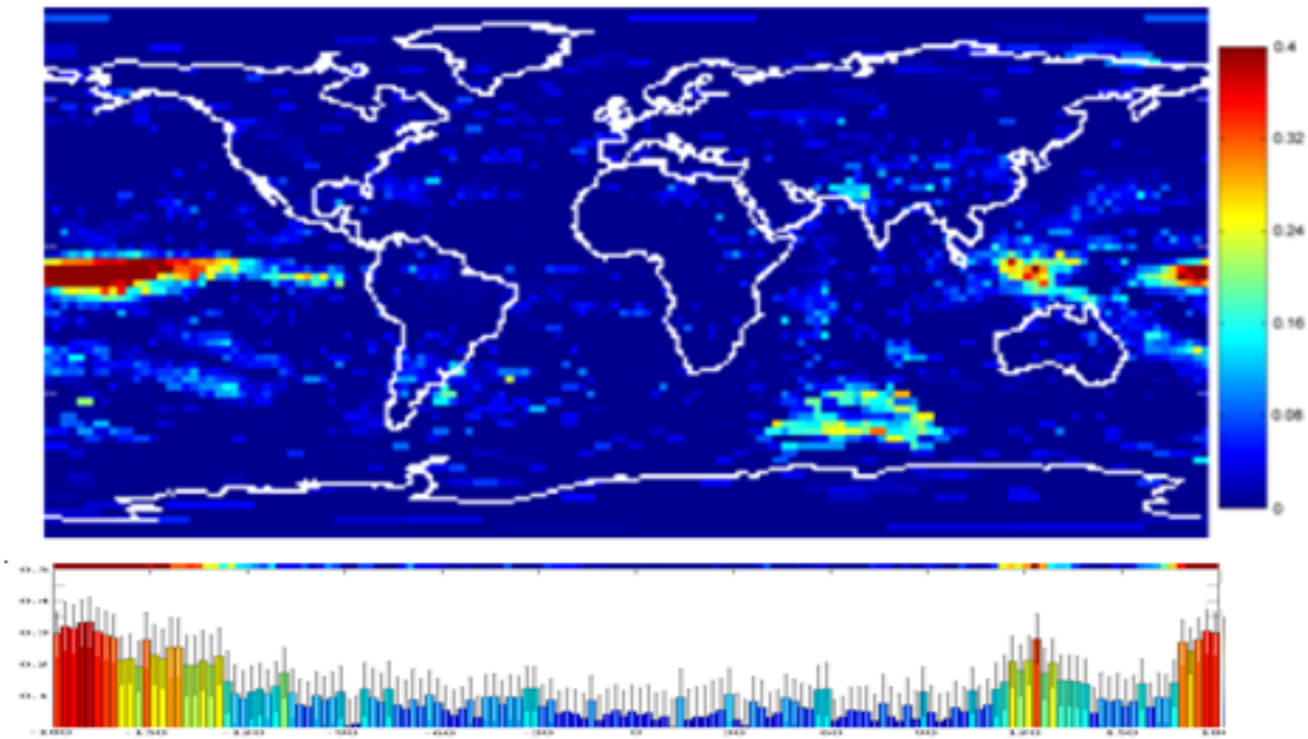
  http://tao.atmos.washington.edu/pacs/additional_analyses/sstanom6n6s18090w.html

Data Description

- Cloud cover data consist of monthly averages of percent cloud cover resulting in a time-series of 198 months of 6596 equal-area pixels each with side length of 280 km. The percent cloud cover data at each pixel can be thought of as a time-series of measurements of subsystem X: X_1, X_2 ,…, X_{6596}

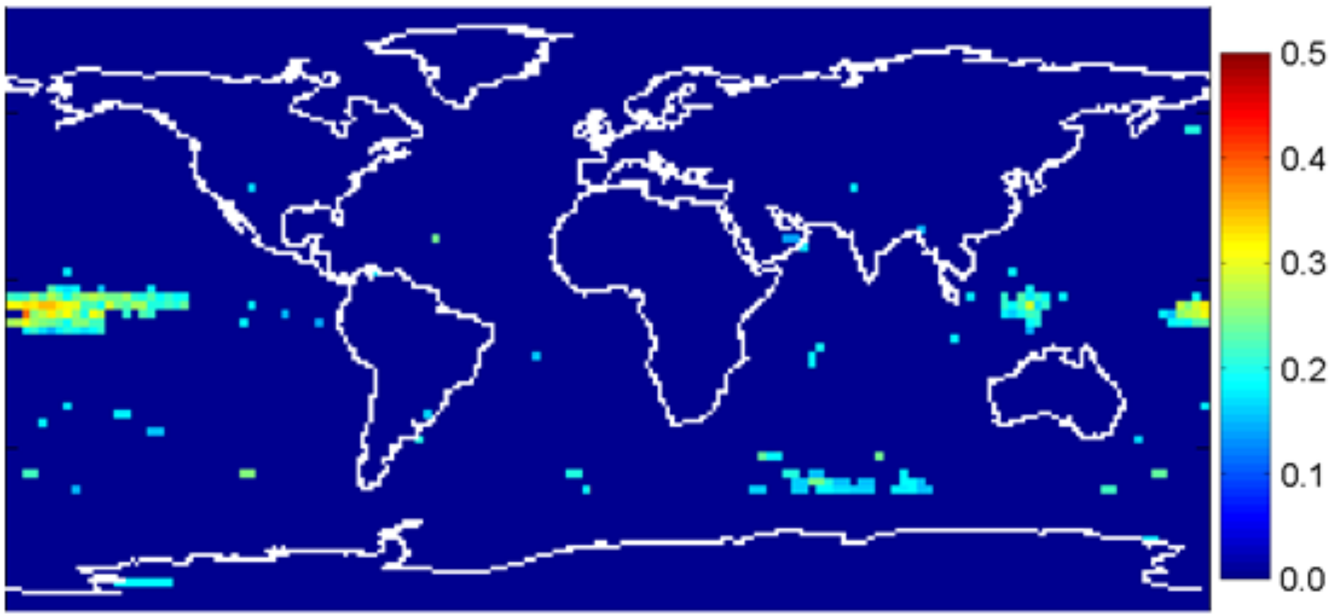
- CTI data consist of the set of 198 monthly values of CTI constitute the sub-system Y

Mutual information between ISCCP cloud cover and CTI
(piecewise-constant pdf method)



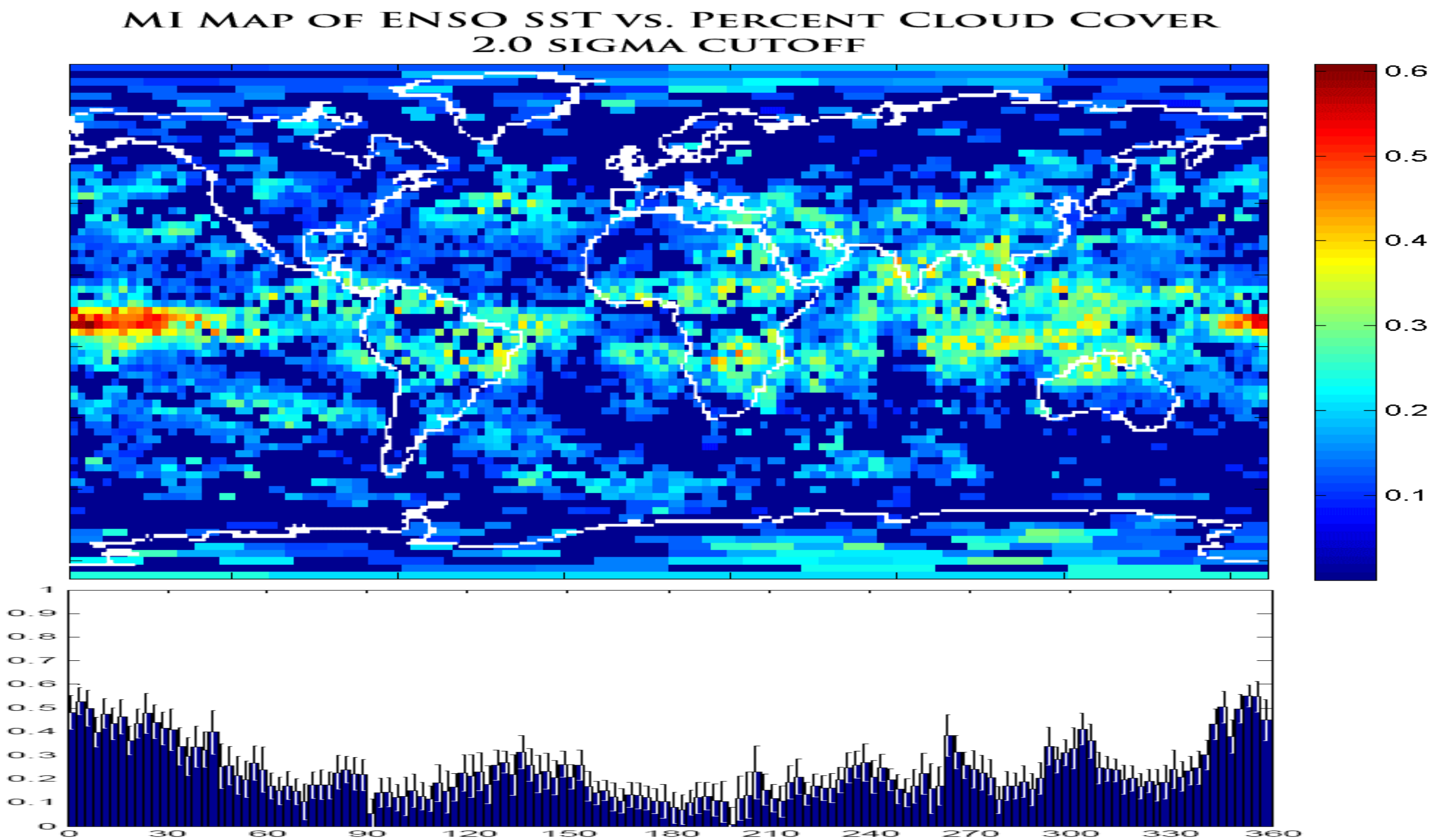
Mutual information values for pixels along the equator with their error bars

Mutual information between ISCCP cloud cover and CTI (piecewise-constant pdf method)



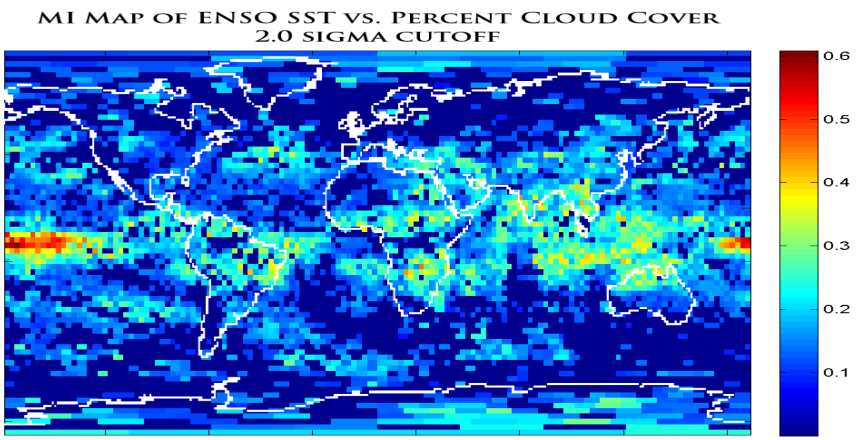
- 2 sigma cut-off (regions having 2 sigma significance and higher)
- Main effect of sea surface temperature on cloud coverage is in the equatorial Pacific, along with and isolated area in Indonesia (satellite coverage artifacts around India)

Mutual information between ISCCP cloud cover and CTI
(MoG pdf model)

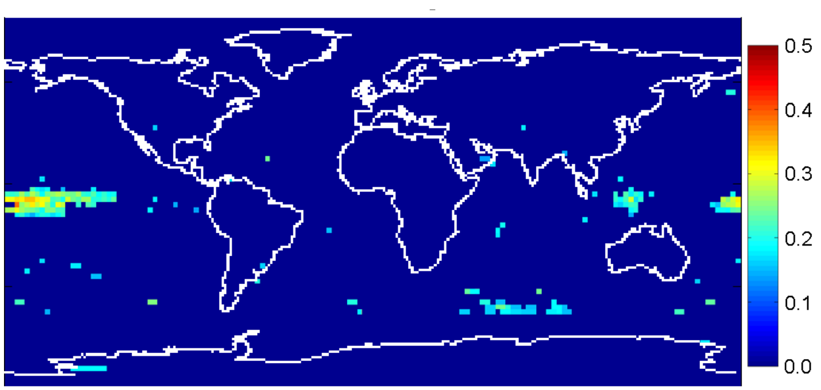


Mutual information between ISCCP cloud cover and CTI

MoG model



Piecewise-constant model

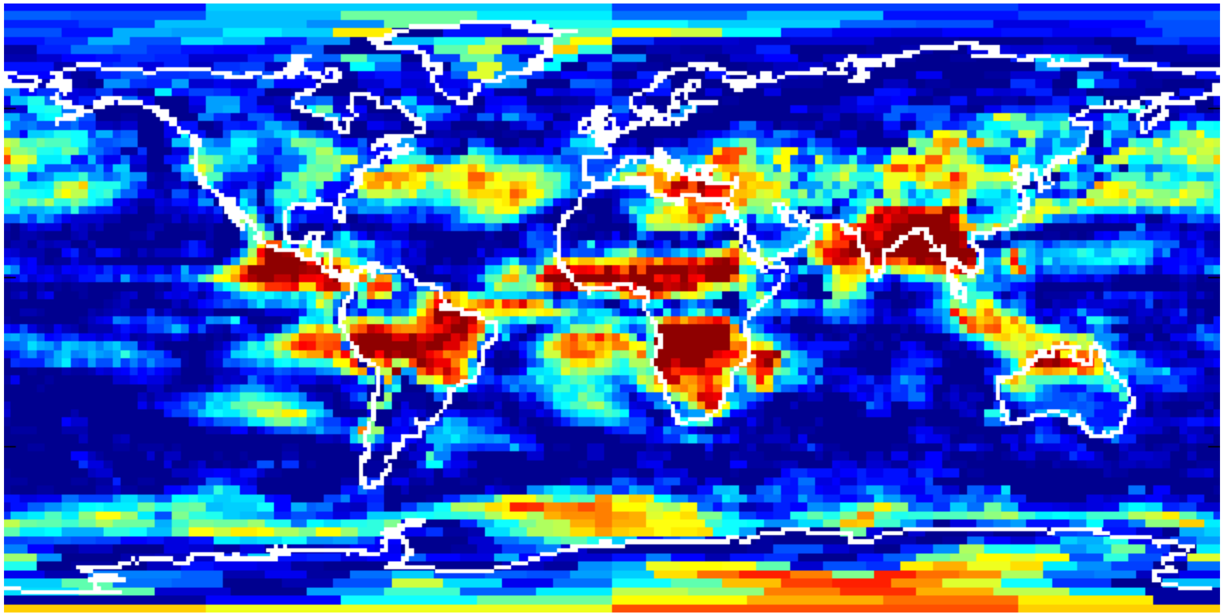


The Mixture of Gaussians model reveals a greater impact of ENSO CTI on global cloud cover.

Not only are the equatorial Pacific and Indonesian areas affected, but ENSO activity affects cloud cover in the inter-tropical convection zones in Africa, India and the eastern coast of South America.

Cloud Cover and Seasonality

Mutual Information between ISCCP percent cloud cover and Seasonality.



This method finds the Inter-Tropical Convection Zones, The Monsoon Regions, the Sea Ice off Antarctica, and cloud cover in the North Atlantic and Pacific.

This figure can be directly compared to the PCA analysis performed by Rossow et al. 1993, J. Climate, 6:2394-2418.

To better understand the relationships between different climate variables we are now working on the application of transfer entropy.

CONCLUSIONS

- We synthesize most commonly utilized TE methods in the literature and use them in combination to analyze the information flow
- Main reason is that every algorithm currently has a point which needs to be fixed. Thus, a combination of methods could help us better understand the problem when they all agree or when any of them disagrees
- In the future, we need a serious method of choosing k and l values which are still not totally well understood in the literature
- Although there are some normalization methods available, a standard method should be brought to avoid scaling problems
- Currently, as all 3 methods work well enough directionwise, we are testing them on atmospheric datasets taken from 16 different satellite

Acknowledgments:

- NASA Cloud Modeling Analysis Initiative (CMAI)

BACKUP SLIDES

Our current research

- We are trying to apply our computational tool to better understand the Madden-Julian Oscillation

- We are looking at possible interactions between different climate variables in the following regions:
  - 65E-95E, 5N-5S
  - 105E-135E,5N-5S
  - 145E-175E,5N-5S

- Currently, we are analyzing vertical winds (Omega 500mb), zonal winds (uwind 875 mb) , MJO indices and different cloud types

Data Description

- Time series of Relative Frequency of Occurence (RFO) for each Weather State (WS) in Tropics

  Weather States
  - WS1 - Deep Cumulus clouds
  - WS2 - Anvil clouds
  - WS3 - Congestus clouds
  - WS4 - Cirrus clouds
  - WS5 - Shallow Cumulus clouds
  - WS6 - Stratocumulus clouds
  - WS7 - Clear sky

- Time series of Vertical Velocity (Omega) in tropics
- MJO Index

Preliminary dependency analysis for MJO

Cloud types vs. MJO index