A non-Gaussian ensemble analysis scheme based on rank histograms

Emmanuel Cosme

CNRS-Université Grenoble Alpes, LGGE, Grenoble

Sammy Metref, Pierre Brasseur

CNRS-Université Grenoble Alpes, LGGE, Grenoble

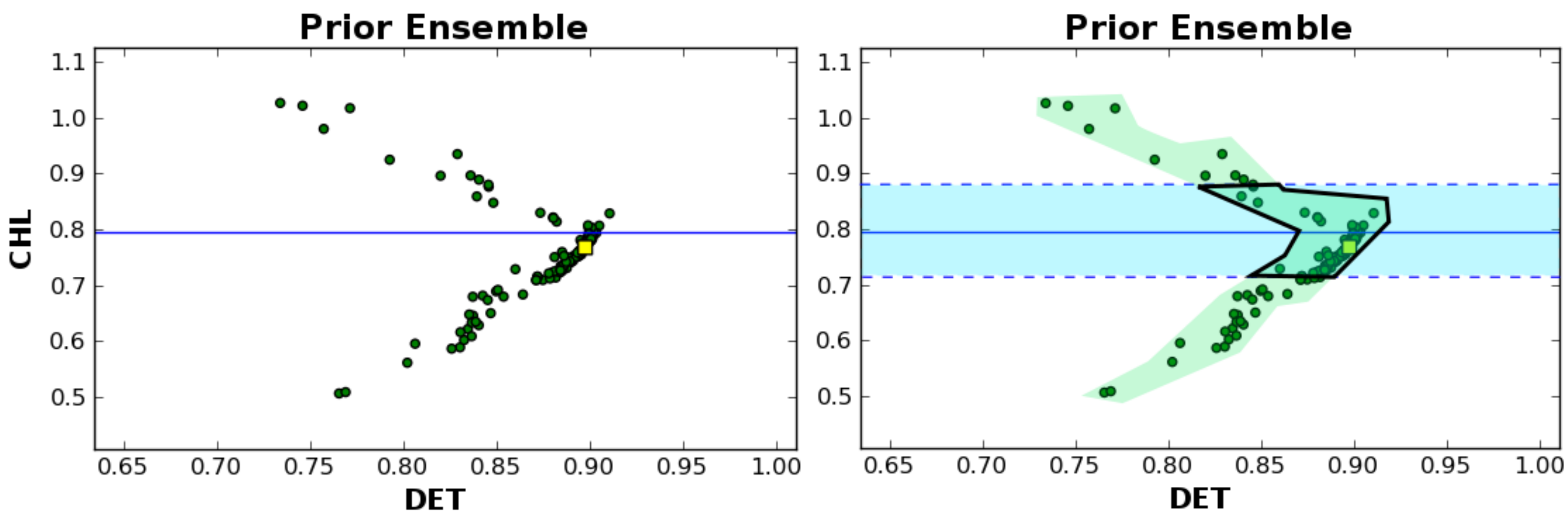
Chris Snyder

NCAR, Boulder, Colorado

EnKF workshop, 23-25th June, 2014, Bergen

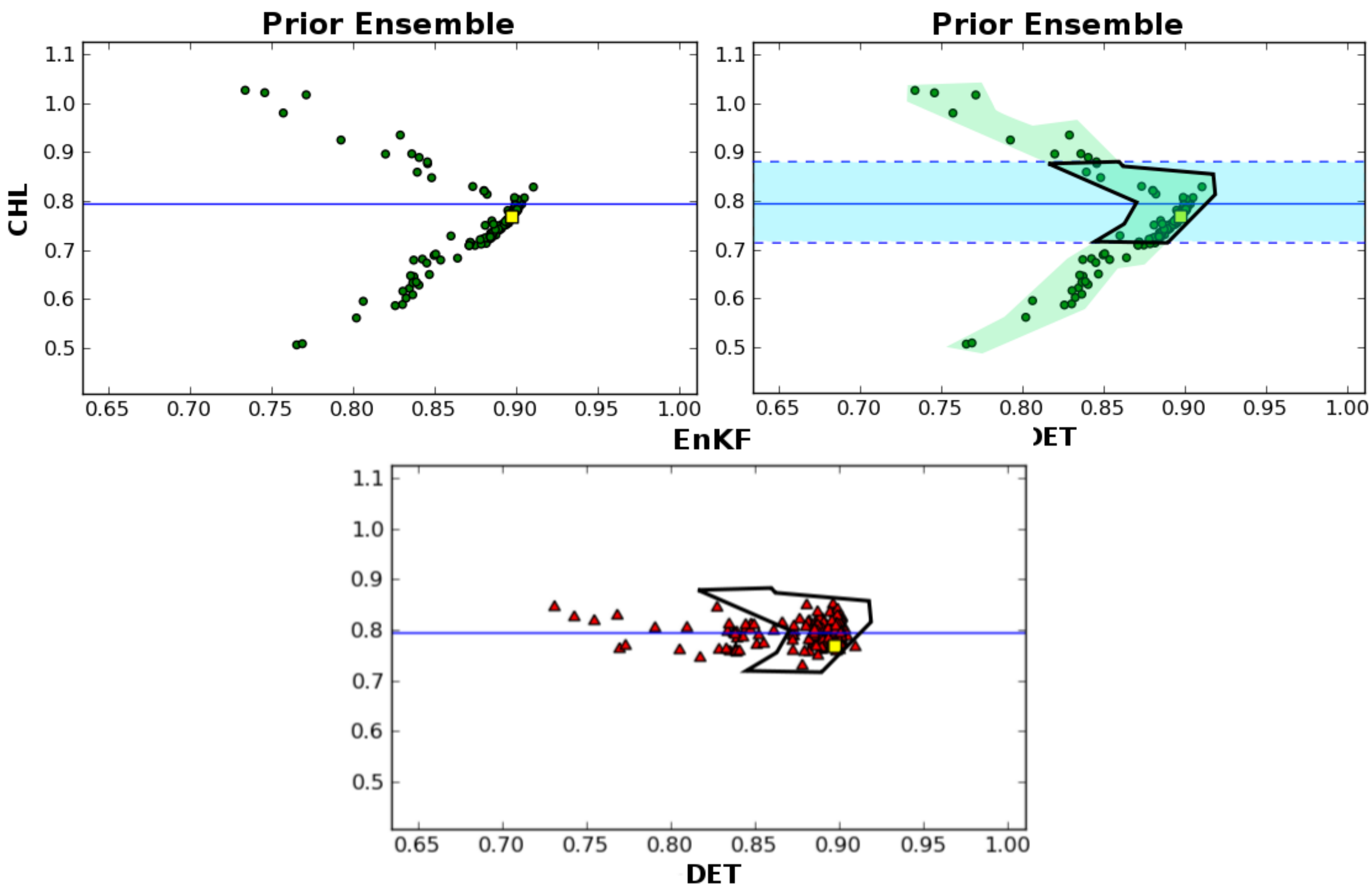
RhôneAlpes Région NCAR NSF European Commission Université Joseph Fourier GRENOBLE cnrs

Introduction



Ensemble of NATL0.25+BGC simulations (Beal et al., 2010), Gulf Stream station (47W/ 40N). chlorophyll (CHL) / detritus (DET).

Introduction



Introduction

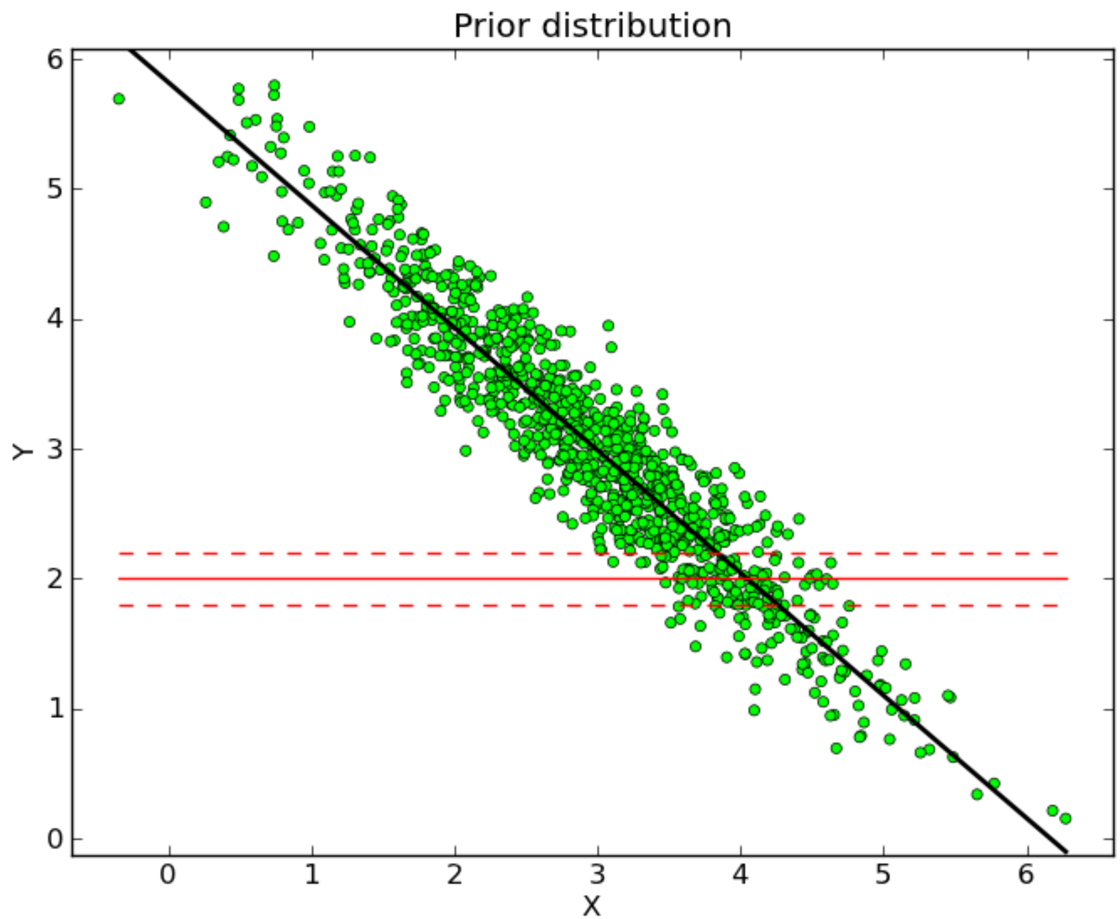
- The analysis step of the serial EnKF can be decomposed in 2 successive operations (Anderson, 2003):
  - Correction of observed variable *z*:

$$\delta z = \frac{\mathrm{Var}(z)}{\mathrm{Var}(z) + \mathrm{Var}(z^o)} \, (z^o - z);$$

  - Corrections of unobserved variables *x* and *y*:

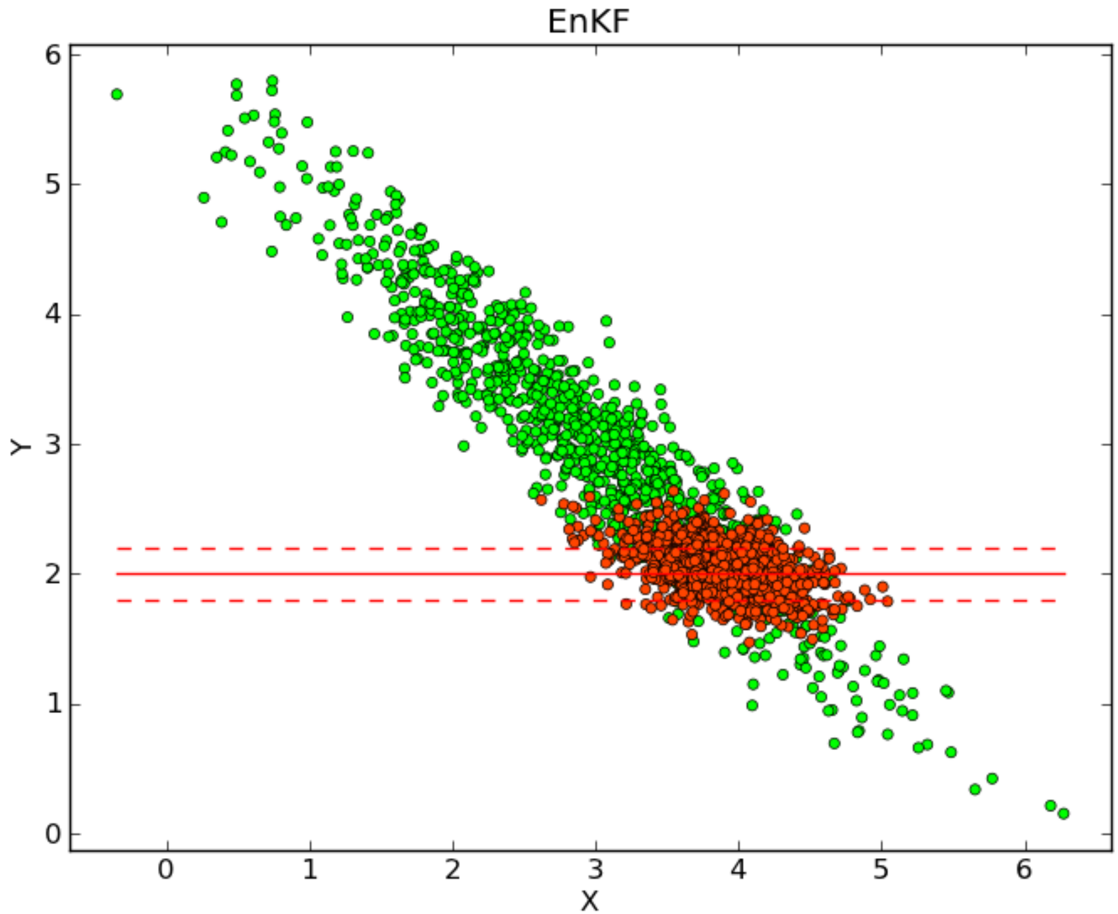
$$\delta x = \frac{\mathrm{Cov}(x,z)}{\mathrm{Var}(z)} \, \delta z, \quad \delta y = \frac{\mathrm{Cov}(y,z)}{\mathrm{Var}(z)} \, \delta z.$$

Introduction

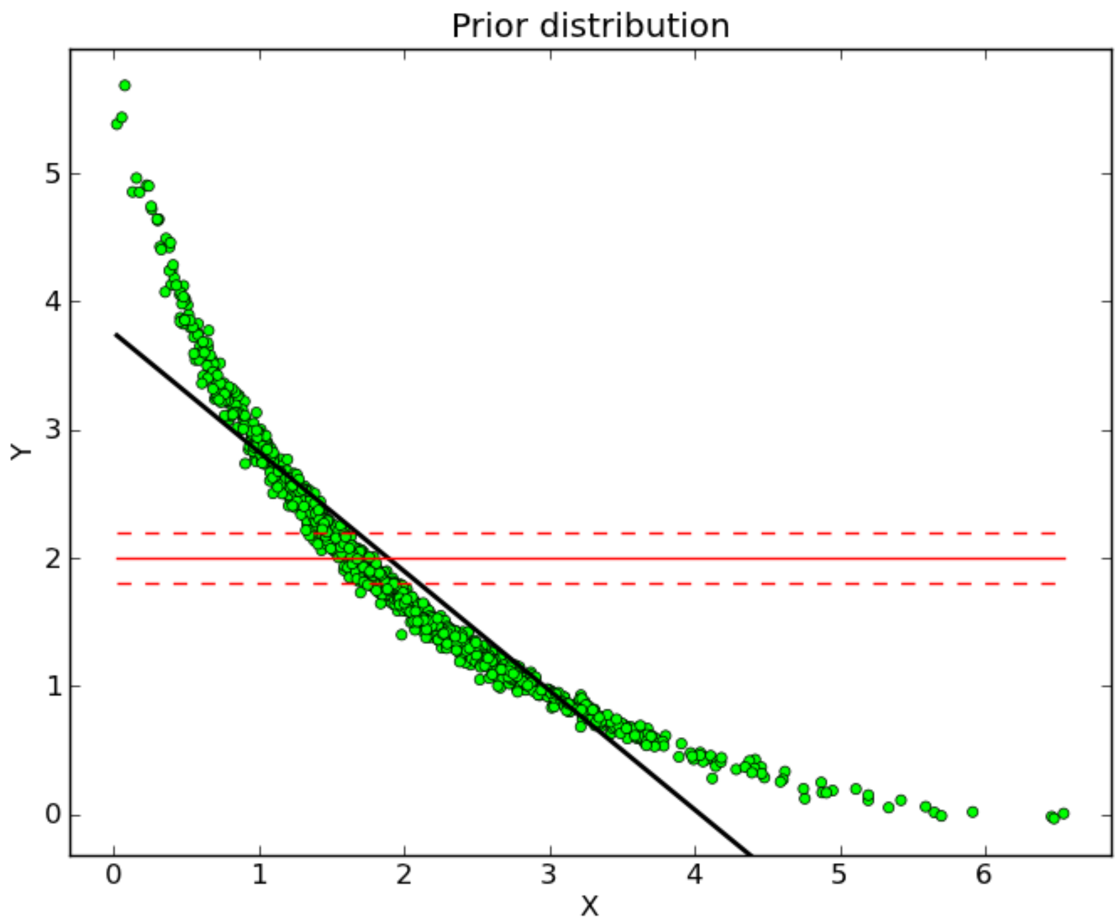


Gaussian distribution

Introduction

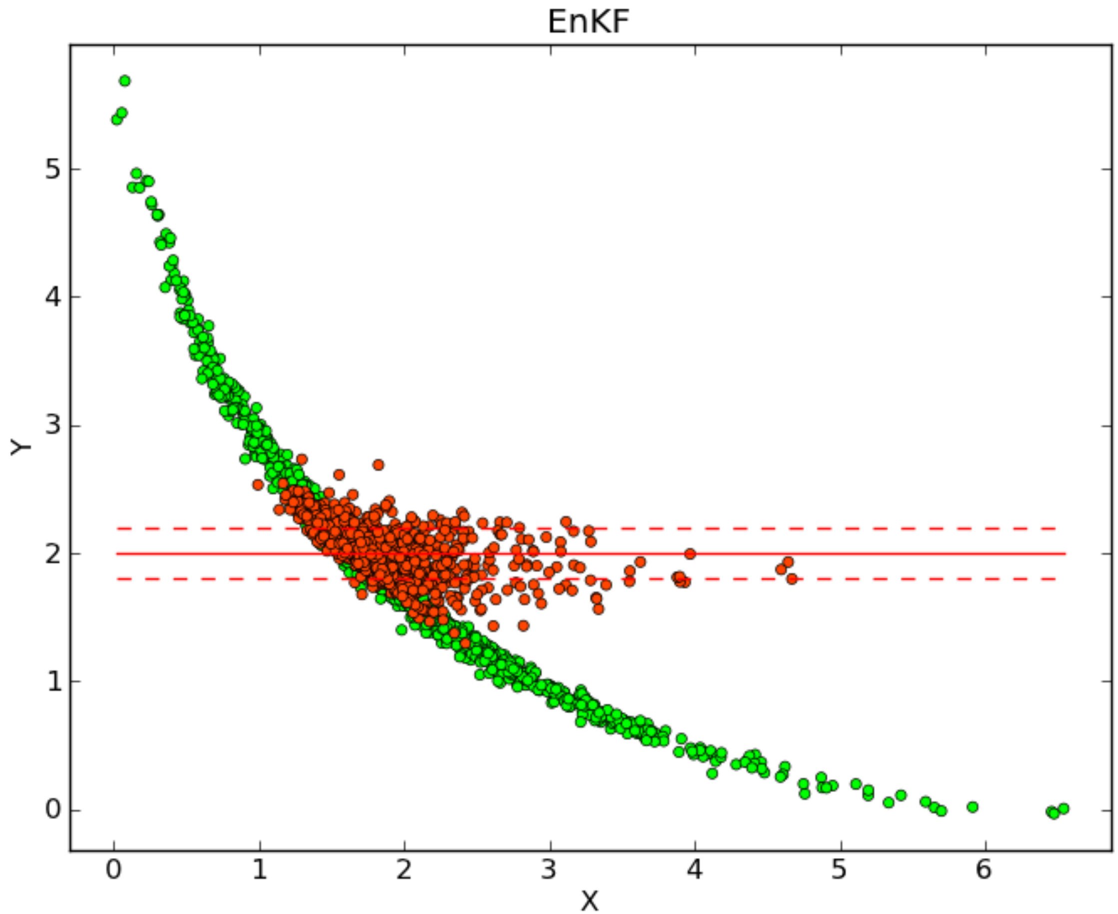


Introduction

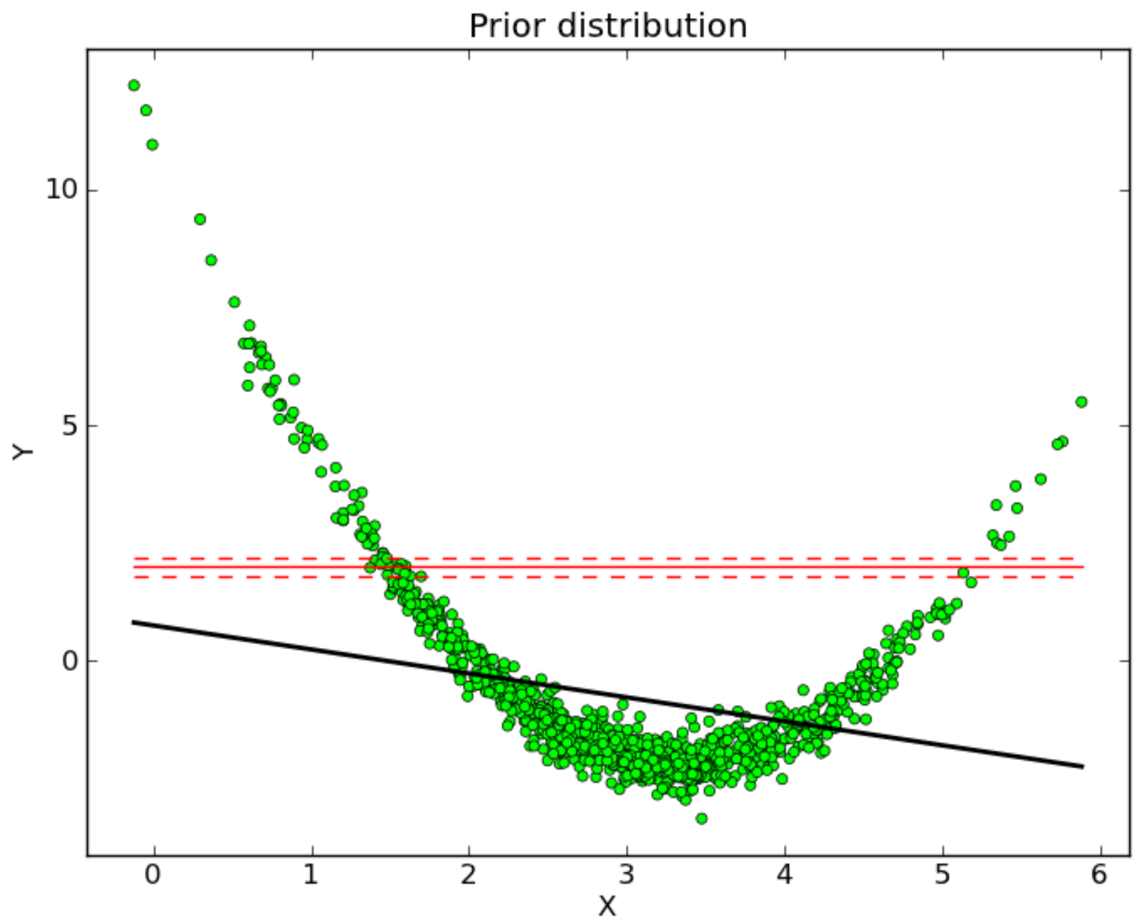


Weakly non-Gaussian distribution

Introduction

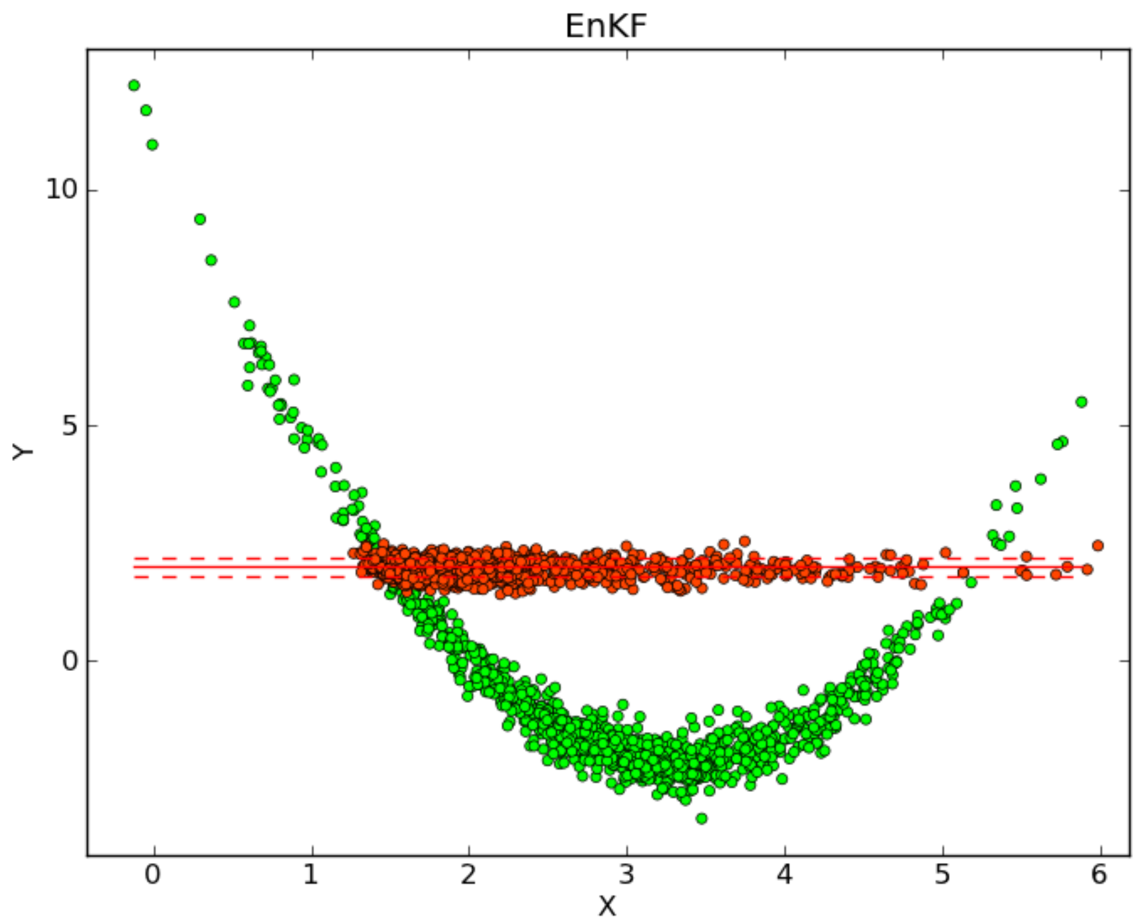


Introduction



Strongly non-Gaussian distribution

Introduction

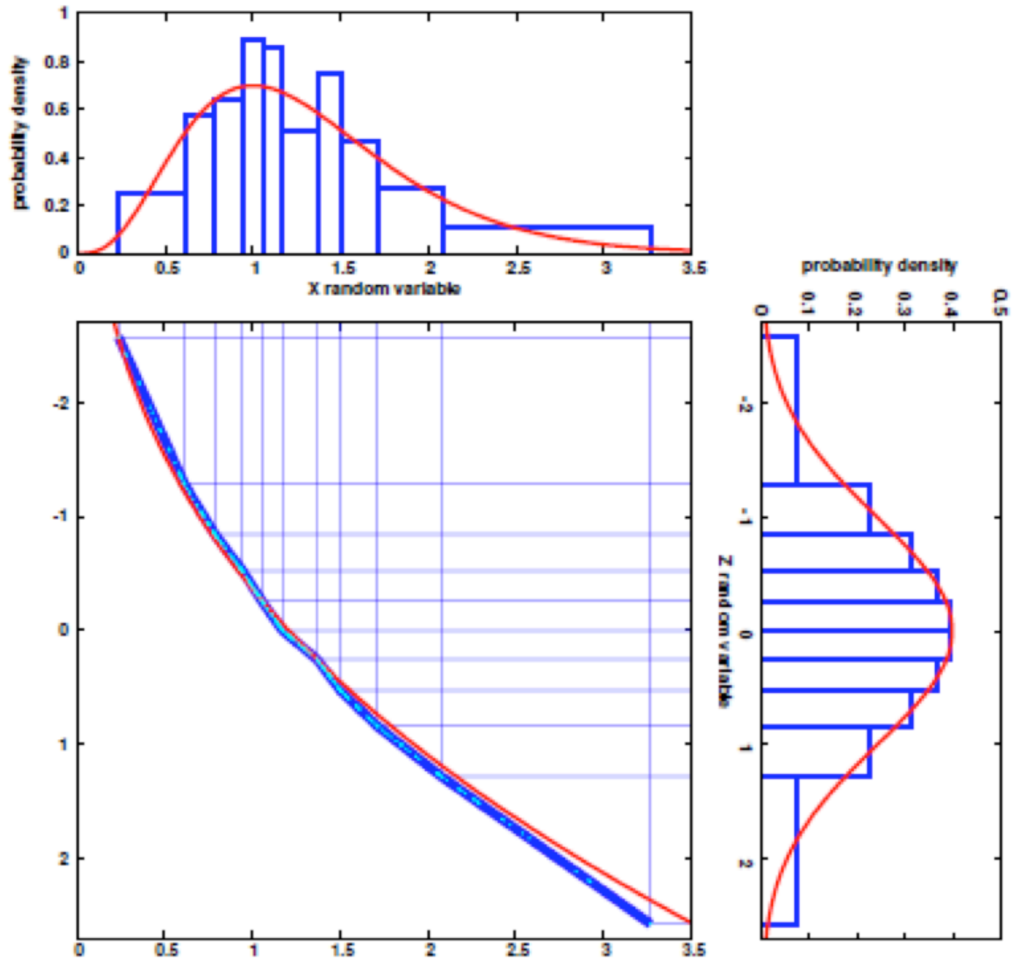


Introduction

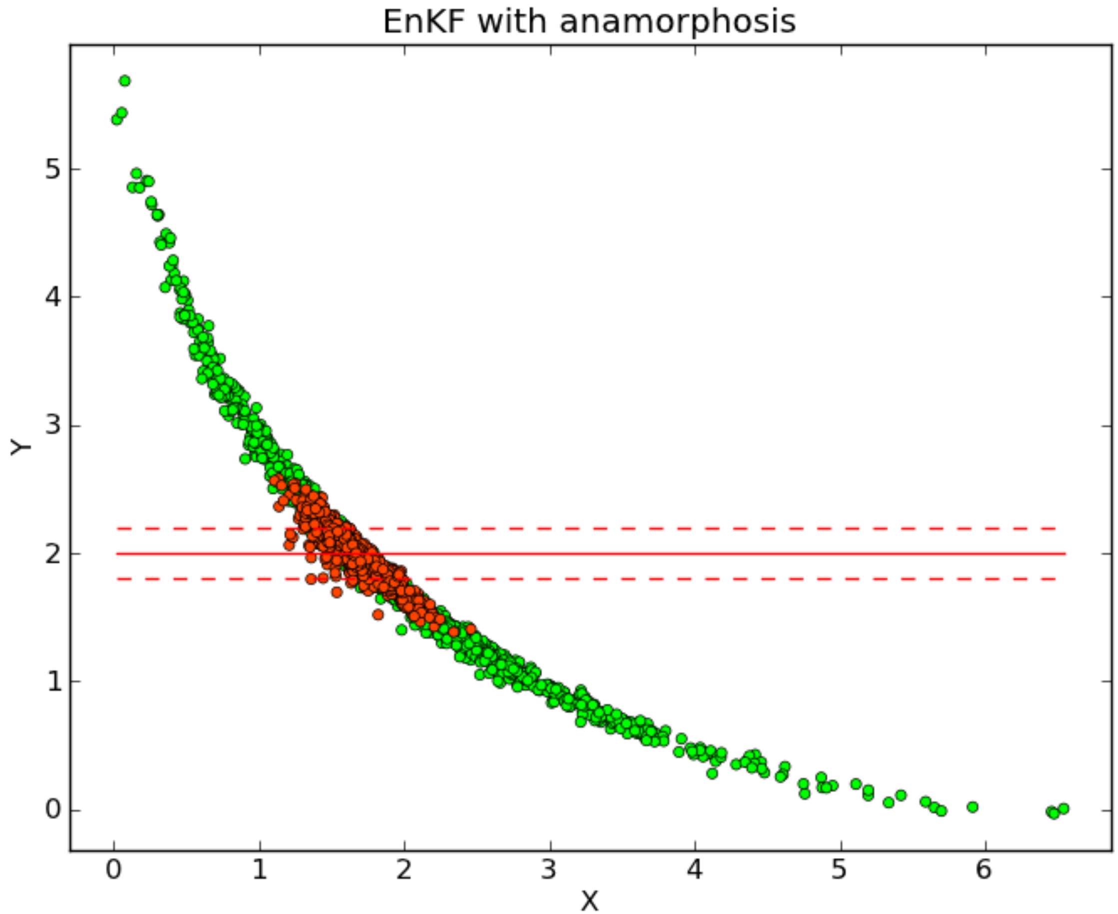
	transform	sampling
Parametric	EnKF (includes ETKF) Truncated-Gaussian EnKF (1)	
Semi-parametric	EnKF with Gaussian anamorphosis (2)	
Non-parametric	ETPF (4)	Particle filters (3)

(1) Lauvernet et al (2009)
(2) Holm et al (2002), Bertino et al (2003); Simon and Bertino (2009); Béal et al (2010); Brankart et al (2012)
(3) Gordon et al (1993); Van Leeuwen et al (2009, 2010), Snyder et al (2008)
(4) Reich (2013)

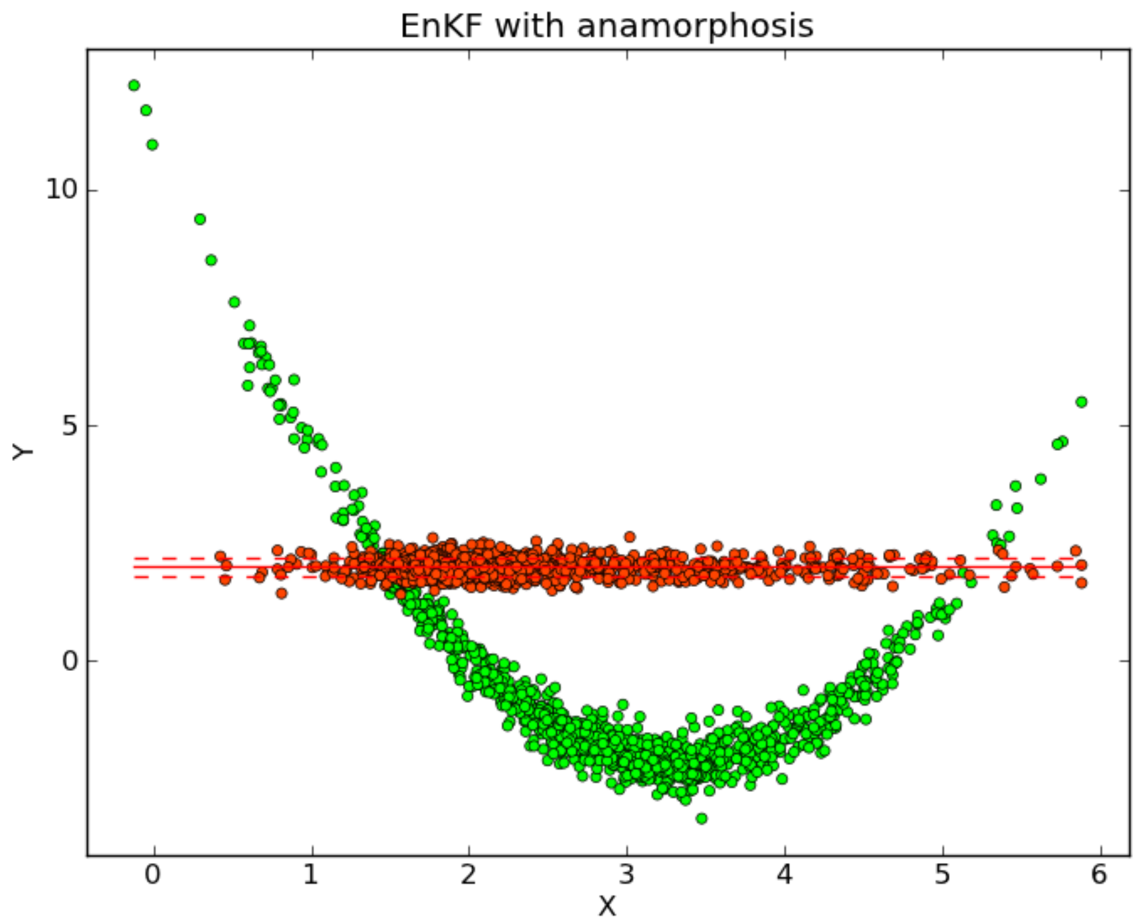
EnKF with Gaussian anamorphosis



EnKF with Gaussian anamorphosis



EnKF with Gaussian anamorphosis



The Rank Histogram Filter (Anderson, 2010)

- The serial EnKF analysis is the parametric, Gaussian implementation of the sequential realization method (Tarantola, 2005, Section 2.3.3). Here with 3 variables:

$$p(x, y, z|z^o) = p(z|z^o)p(x|z, z^o)p(y|x, z, z^o).$$

$$\delta z = \frac{\mathrm{Var}(z)}{\mathrm{Var}(z) + \mathrm{Var}(z^o)}\,(z^o - z);$$

$$\delta x = \frac{\mathrm{Cov}(x, z)}{\mathrm{Var}(z)}\,\delta z, \qquad \delta y = \frac{\mathrm{Cov}(y, z)}{\mathrm{Var}(z)}\,\delta z.$$

The Rank Histogram Filter (Anderson, 2010)

- The serial EnKF analysis is the parametric, Gaussian implementation of the sequential realization method (Tarantola, 2005, Section 2.3.3). Here with 3 variables:

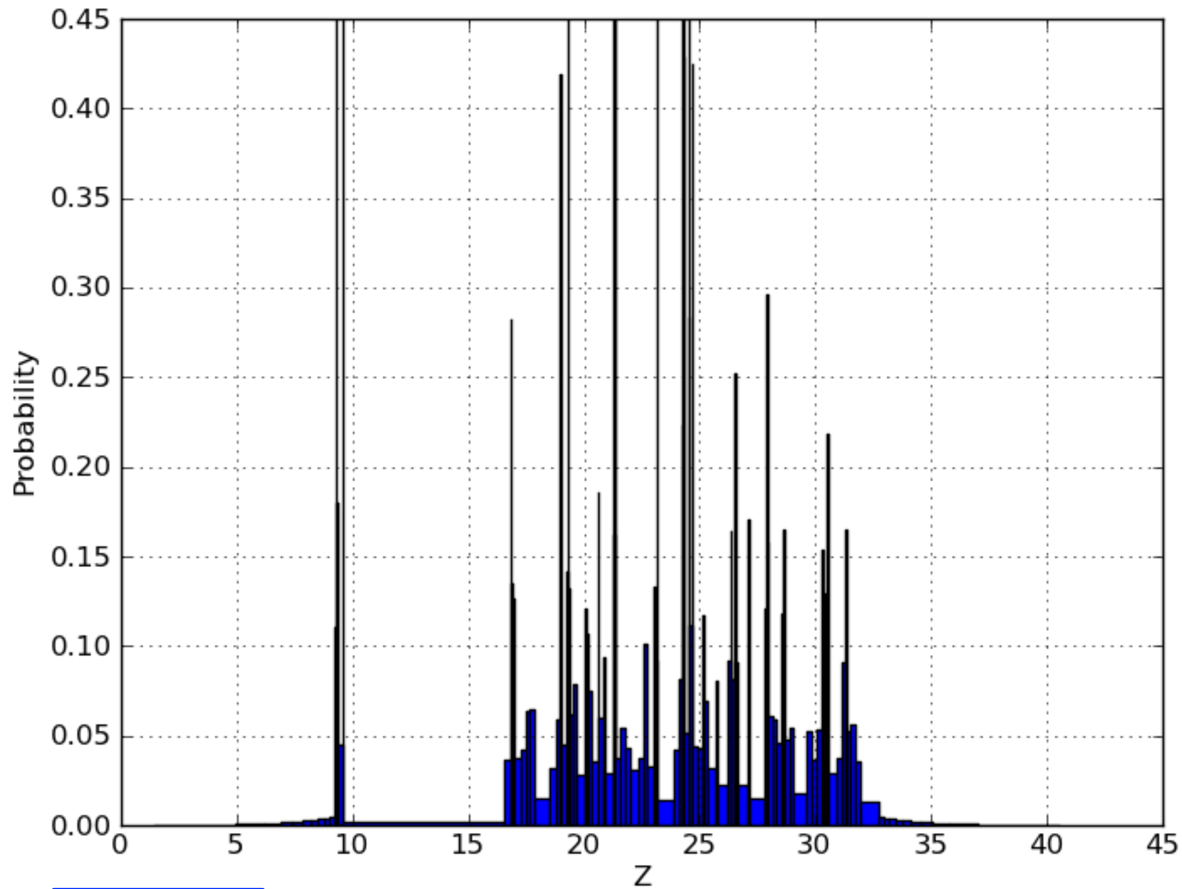
$$p(x, y, z|z^o) = p(z|z^o)p(x|z, z^o)p(y|x, z, z^o).$$

- Anderson (2010) replaces the linear update for *z* with an implementation of Bayes' rule,

$$p(z|z^o) = p(z)p(z^o|z),$$

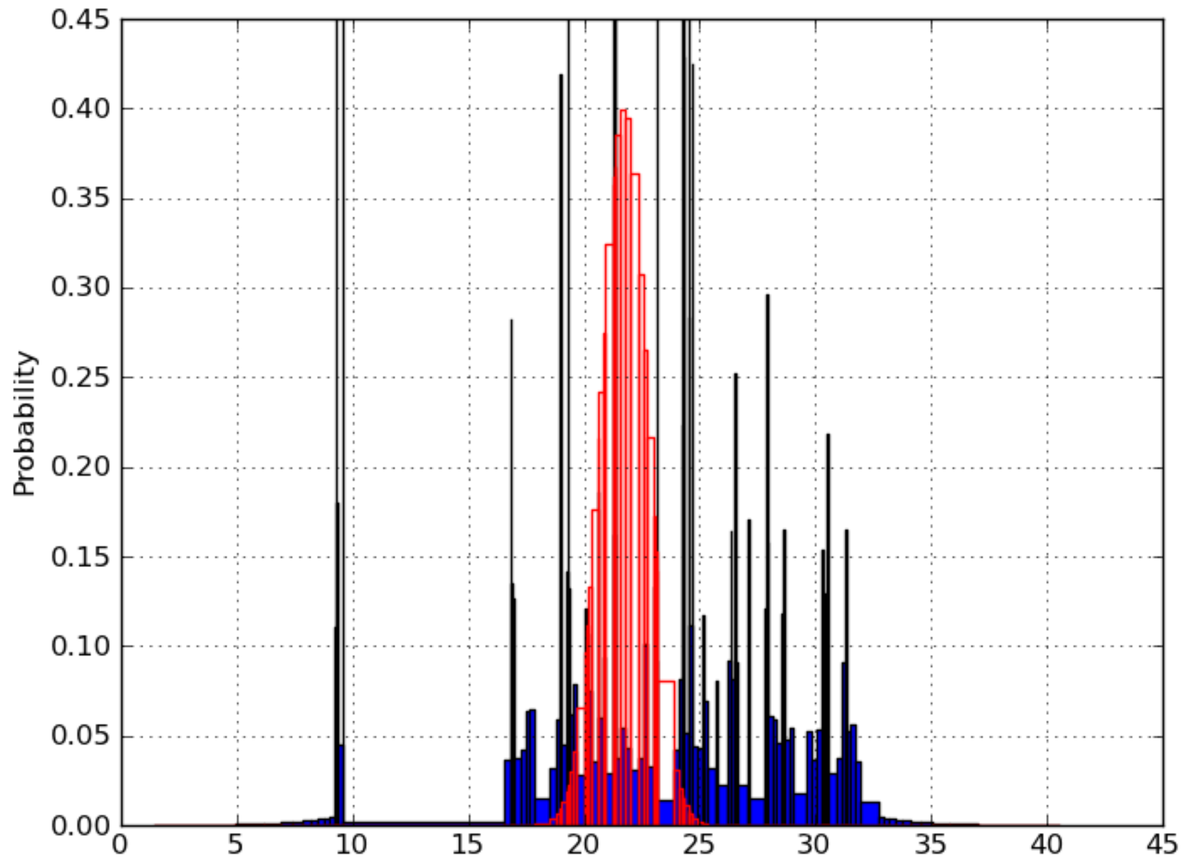
based on rank histograms.

The Rank Histogram Filter (Anderson, 2010)



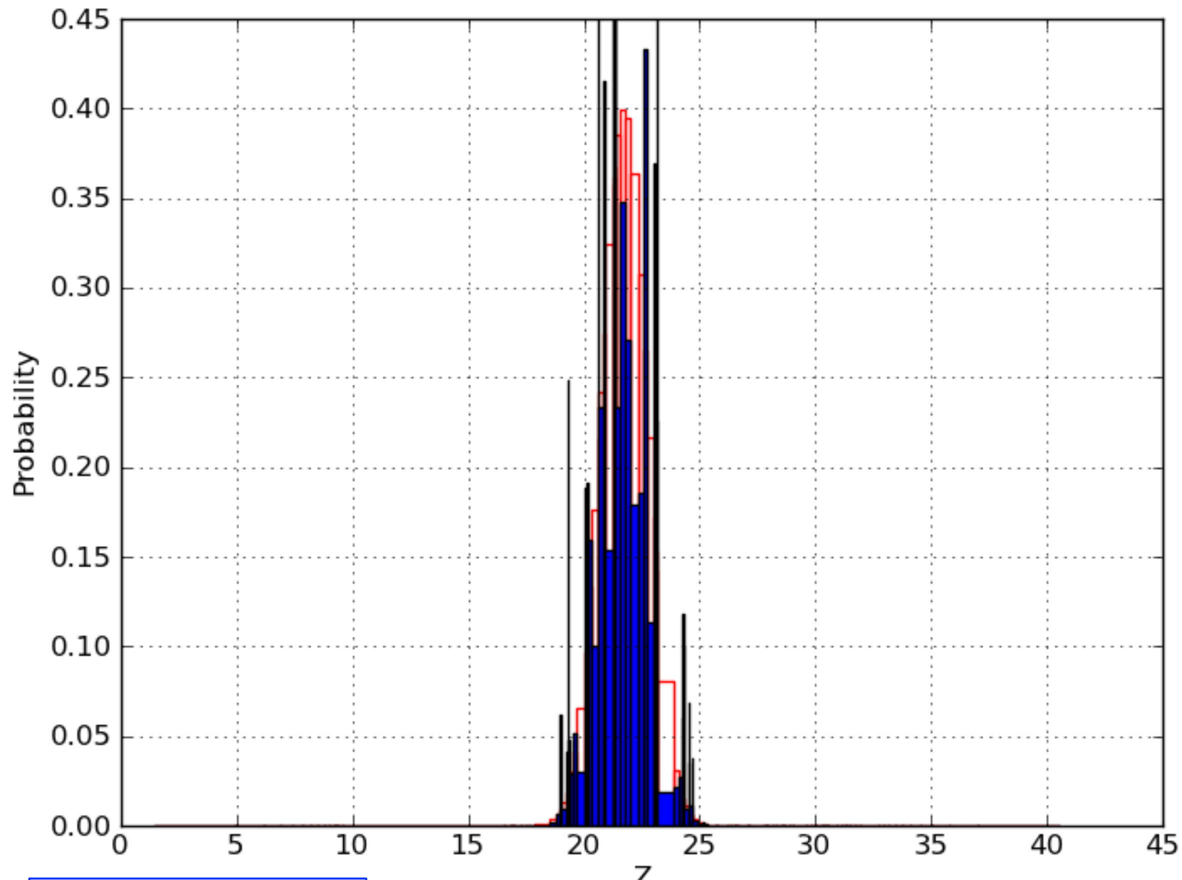
$p(z)$ (mass=1/(N + 1) between 2 consecutive particles)

The Rank Histogram Filter (Anderson, 2010)



$$p(z^o|z)$$

The Rank Histogram Filter (Anderson, 2010)



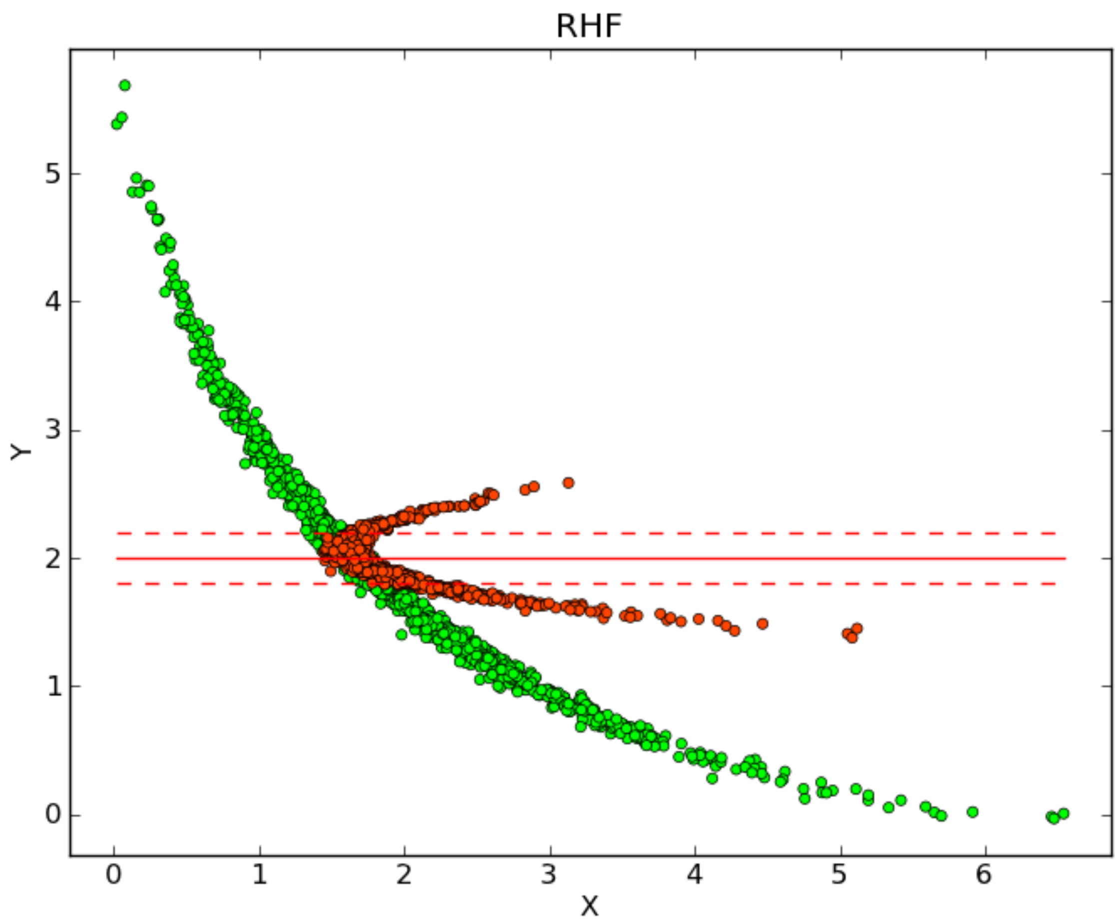
$$p(z|z^o)$$

point-wise product of the previous 2. The analysis z values are sampled by inversion of CDF.

The Rank Histogram Filter (Anderson, 2010)

- And unobserved variables are corrected using linear regressions, as in the EnKF.

The Rank Histogram Filter (Anderson, 2010)



The Rank Histogram Filter (Anderson, 2010)

- Fully non-Gaussian method…
- Robust…
- … for observed variables

Basic idea of the Multivariate Rank Histogram Filter

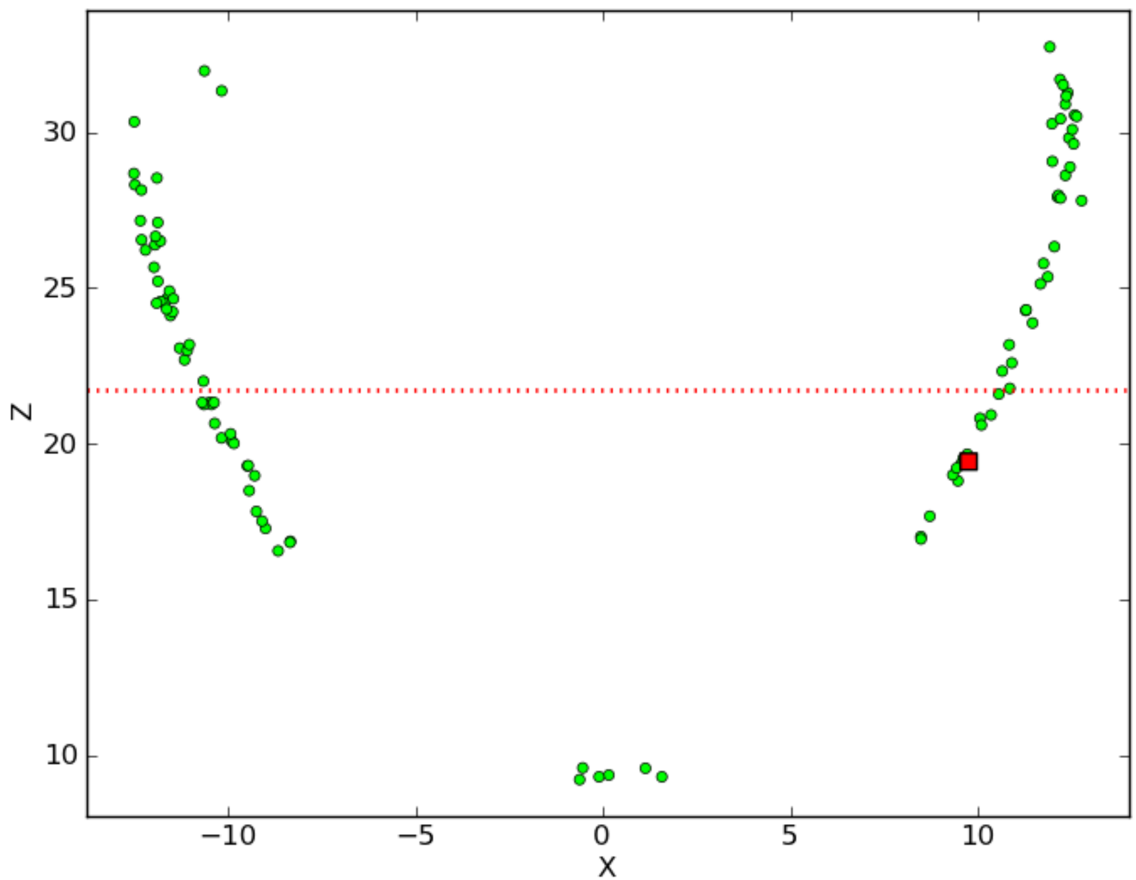
For 2 variables x and z, z is observed by z°.

Knothe-Rosenblatt rearrangement of the joint pdf

$$p(x, z|z^o) = p(z|z^o)p(x|z, z^o)$$

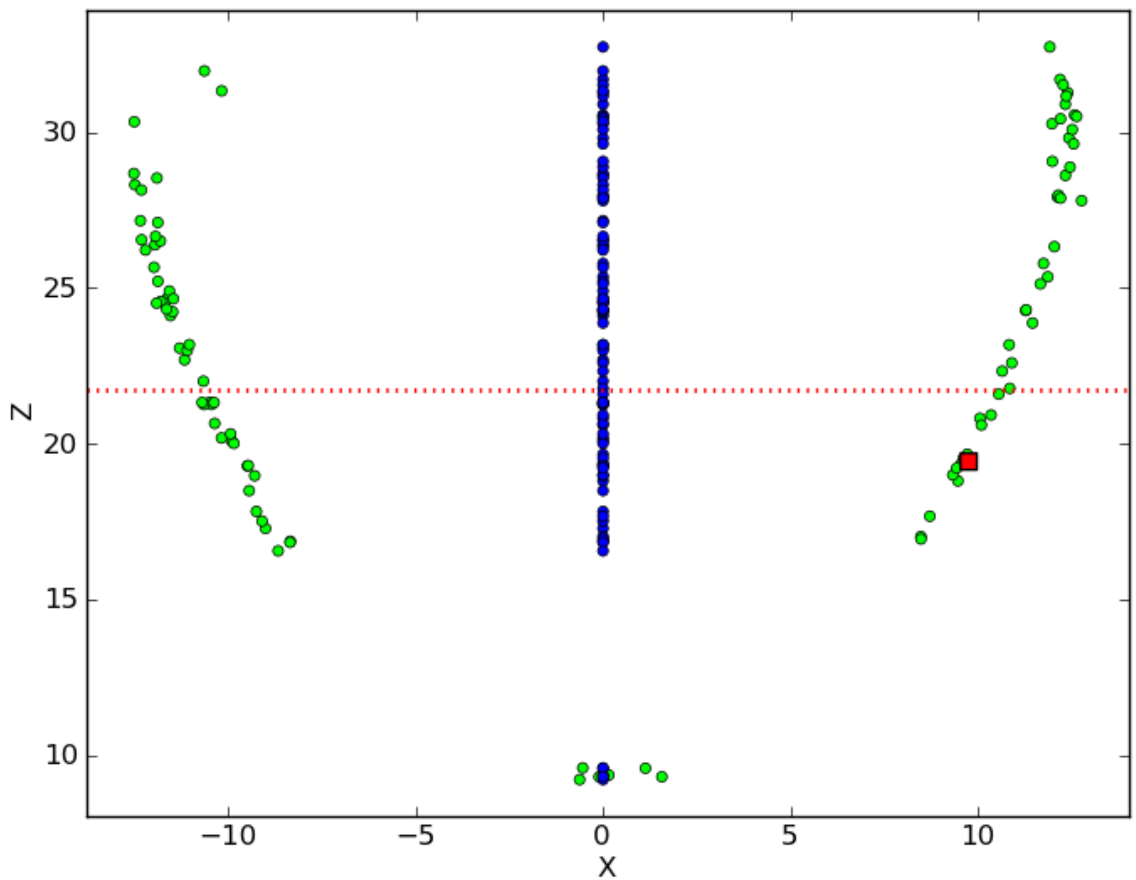
➔ Sequential computation for z and x (as in the EnKF).

The Multivariate Rank Histogram Filter



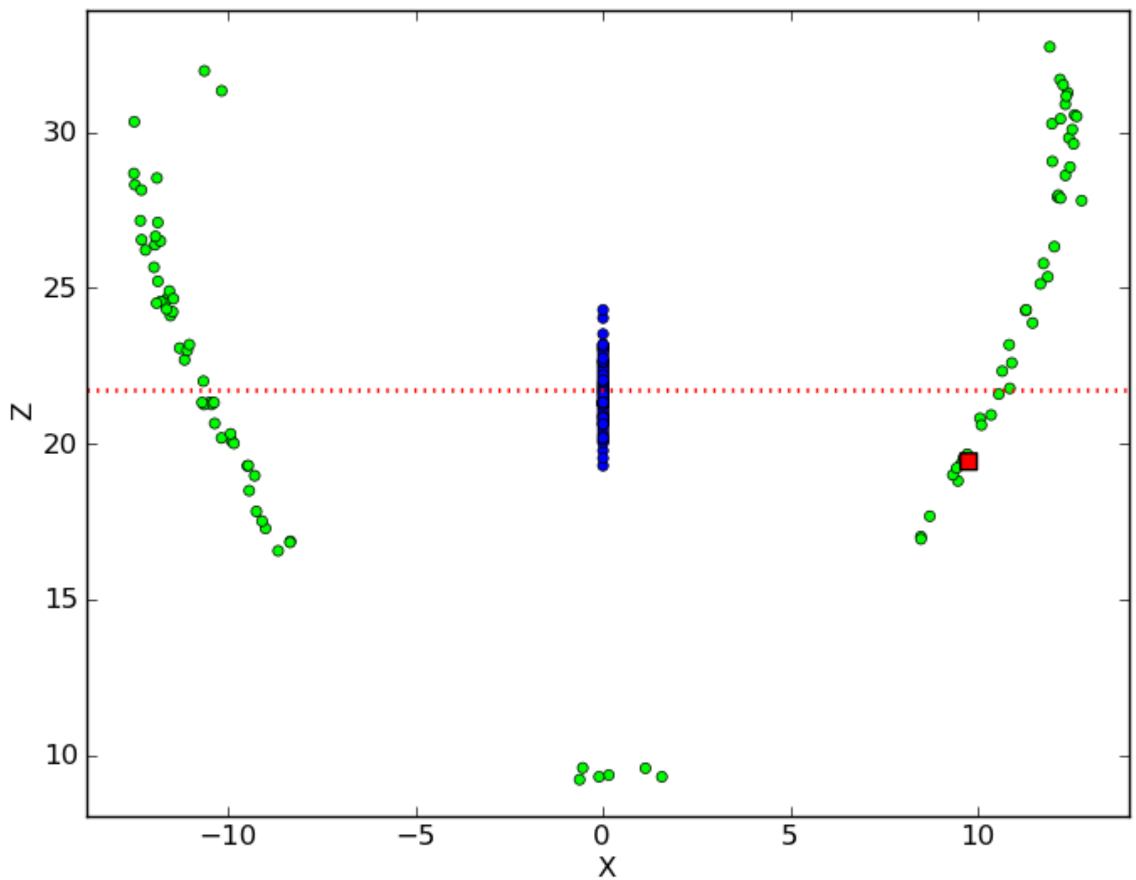
Background ensemble in X − Z plane.
Red dotted line: Z obs. Red square: truth.

The Multivariate Rank Histogram Filter



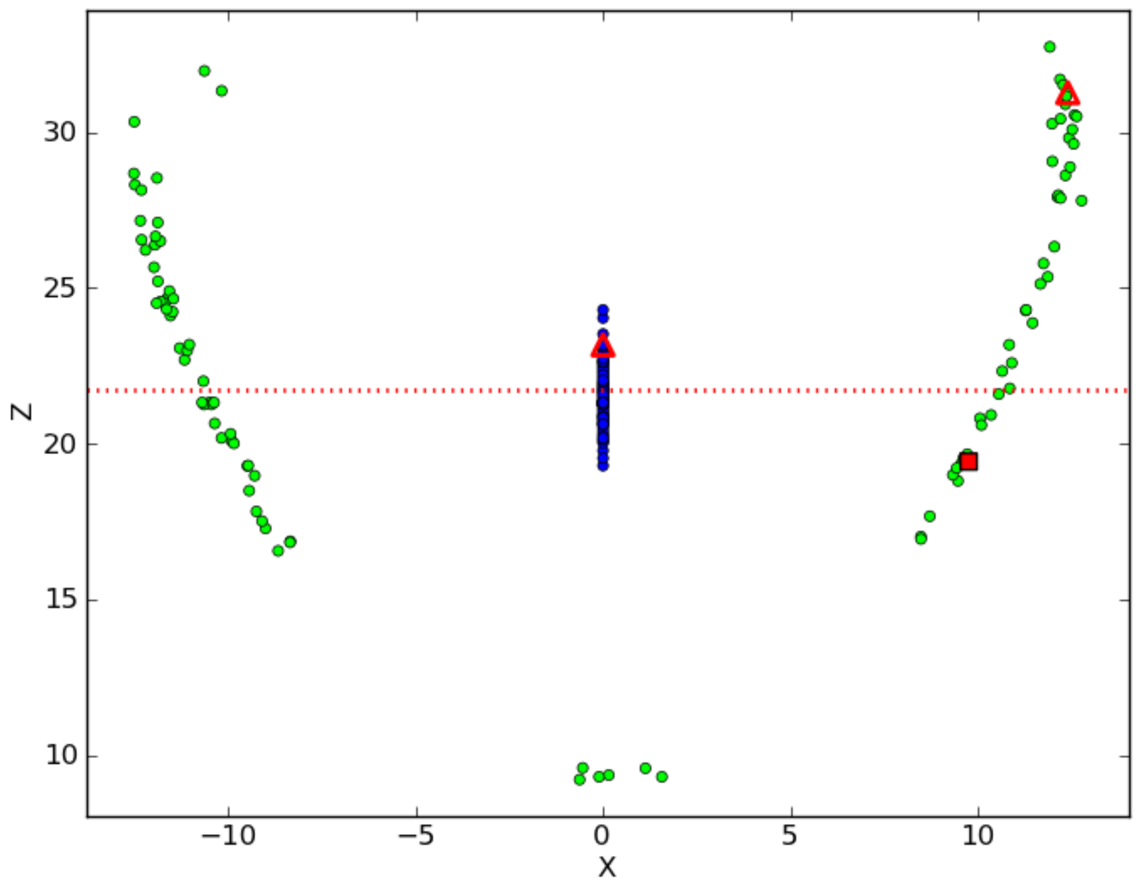
Background Z ensemble for RHF analysis.

The Multivariate Rank Histogram Filter



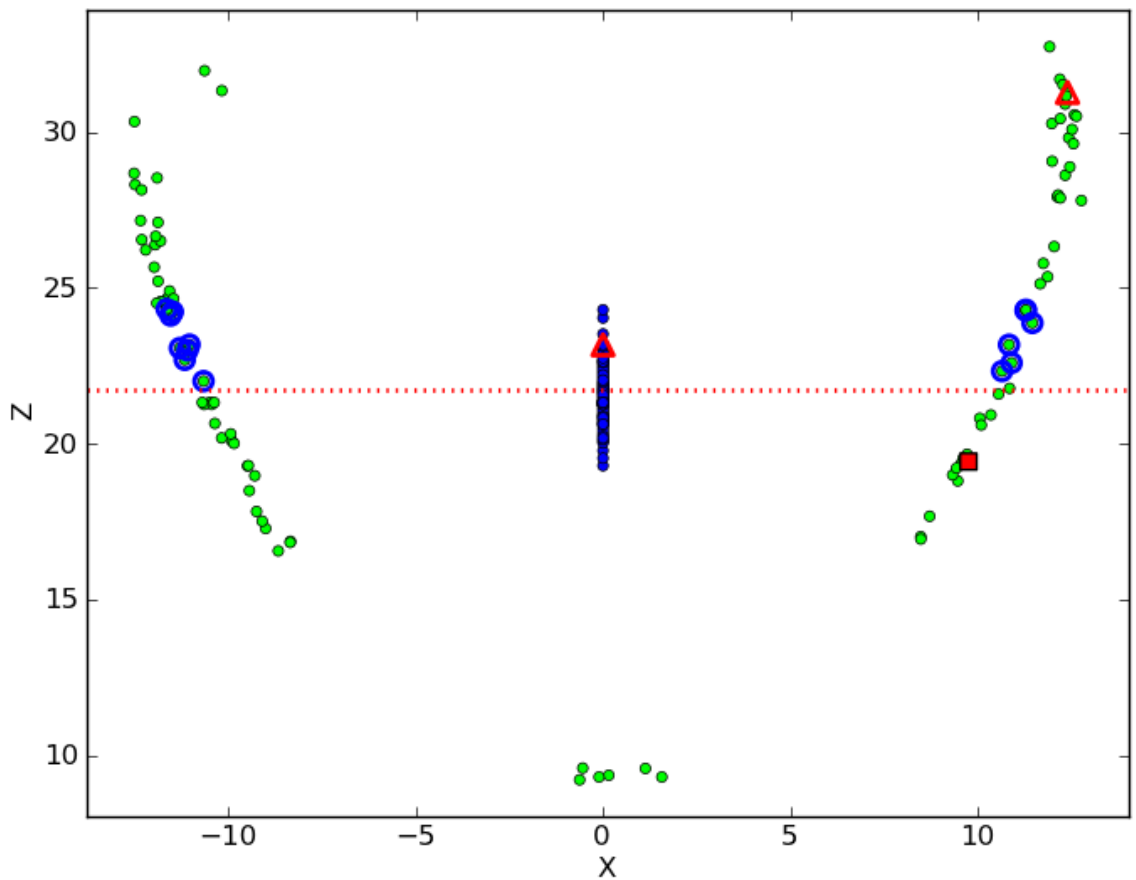
RHF analysis ensemble on Z line.

The Multivariate Rank Histogram Filter



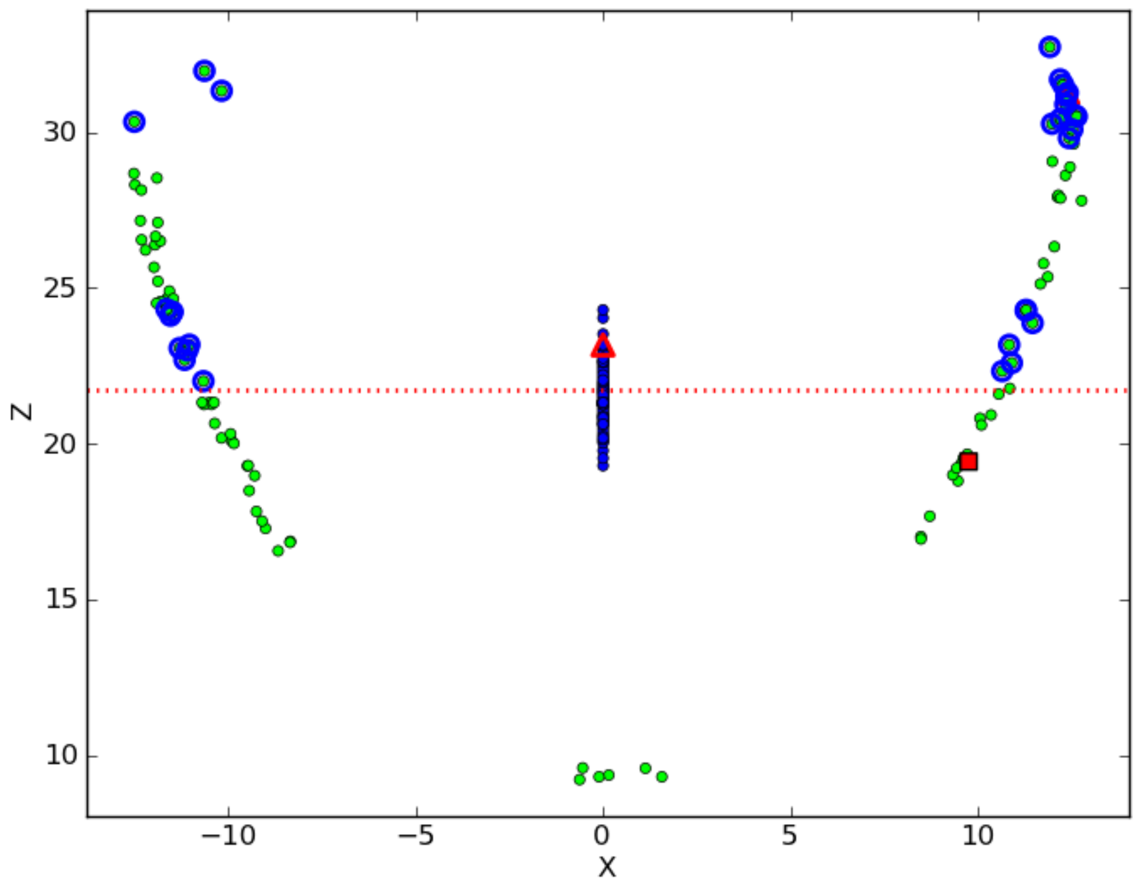
For each particle i, an analyzed value for X must be calculated.

The Multivariate Rank Histogram Filter



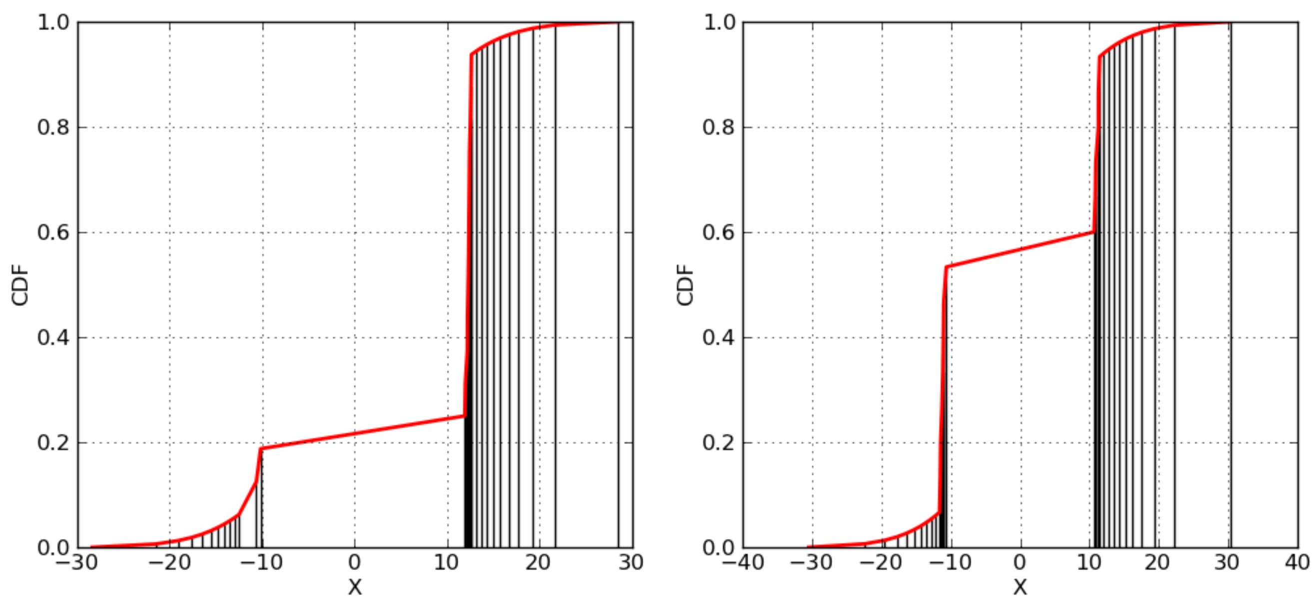
To form $p(X|Z = Z_i^a)$, select particles in the background ensemble. X analysis could be randomly drawn from $p(X|Z = Z_i^a)$.

The Multivariate Rank Histogram Filter



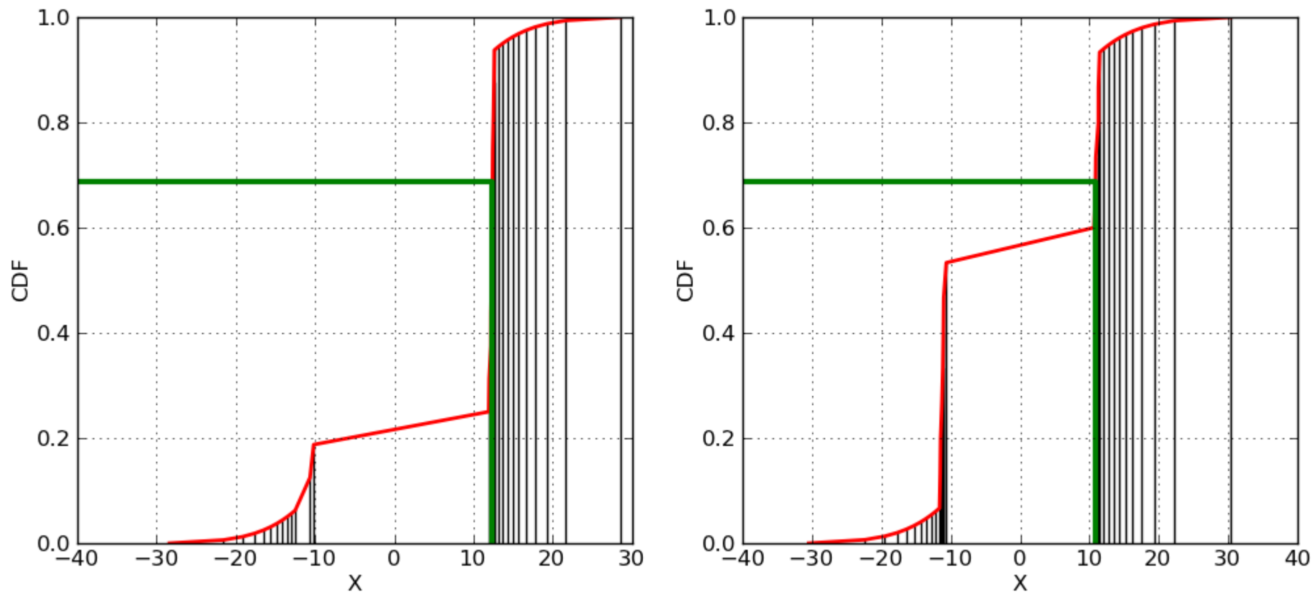
Instead, we select particles to estimate $p(X|Z = Z_i^b)$.

The Multivariate Rank Histogram Filter



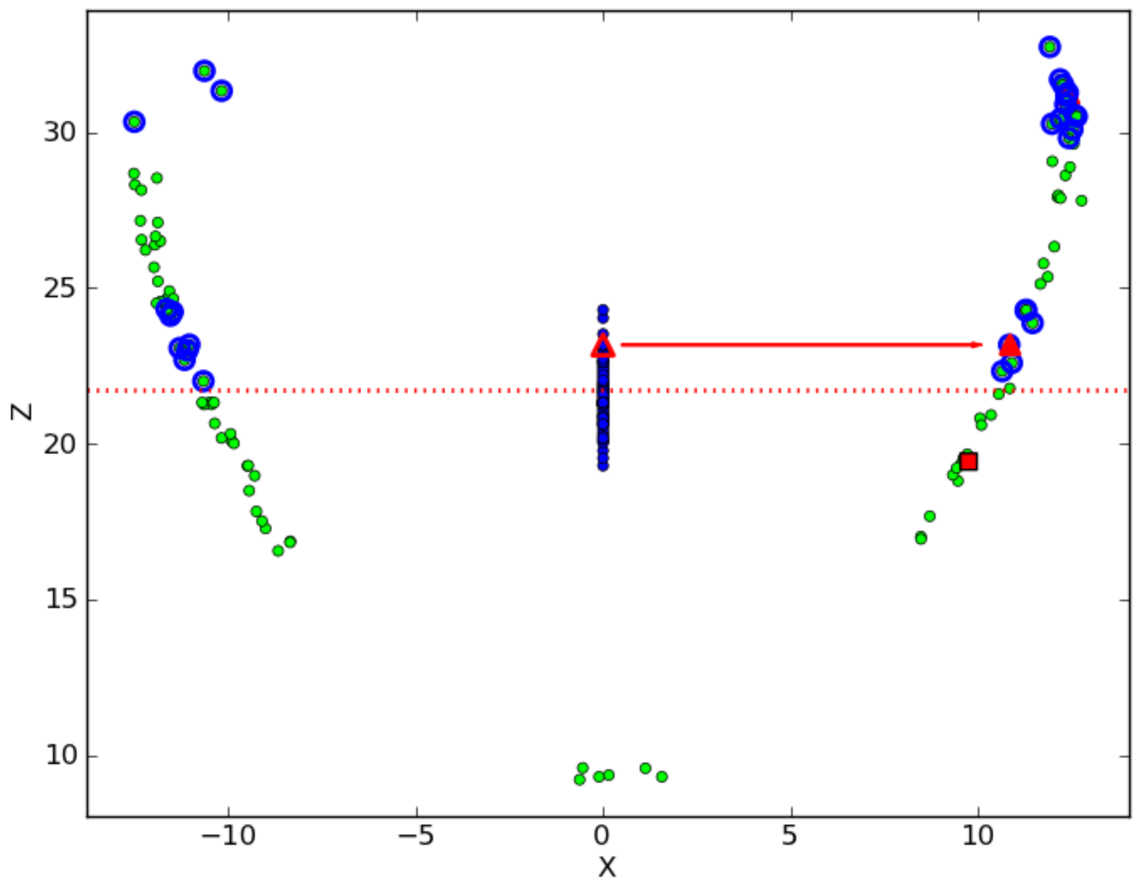
The marginal CDF of $X \mid Z = Z_i^b$ and $X \mid Z = Z_i^a$ are formed.

The Multivariate Rank Histogram Filter



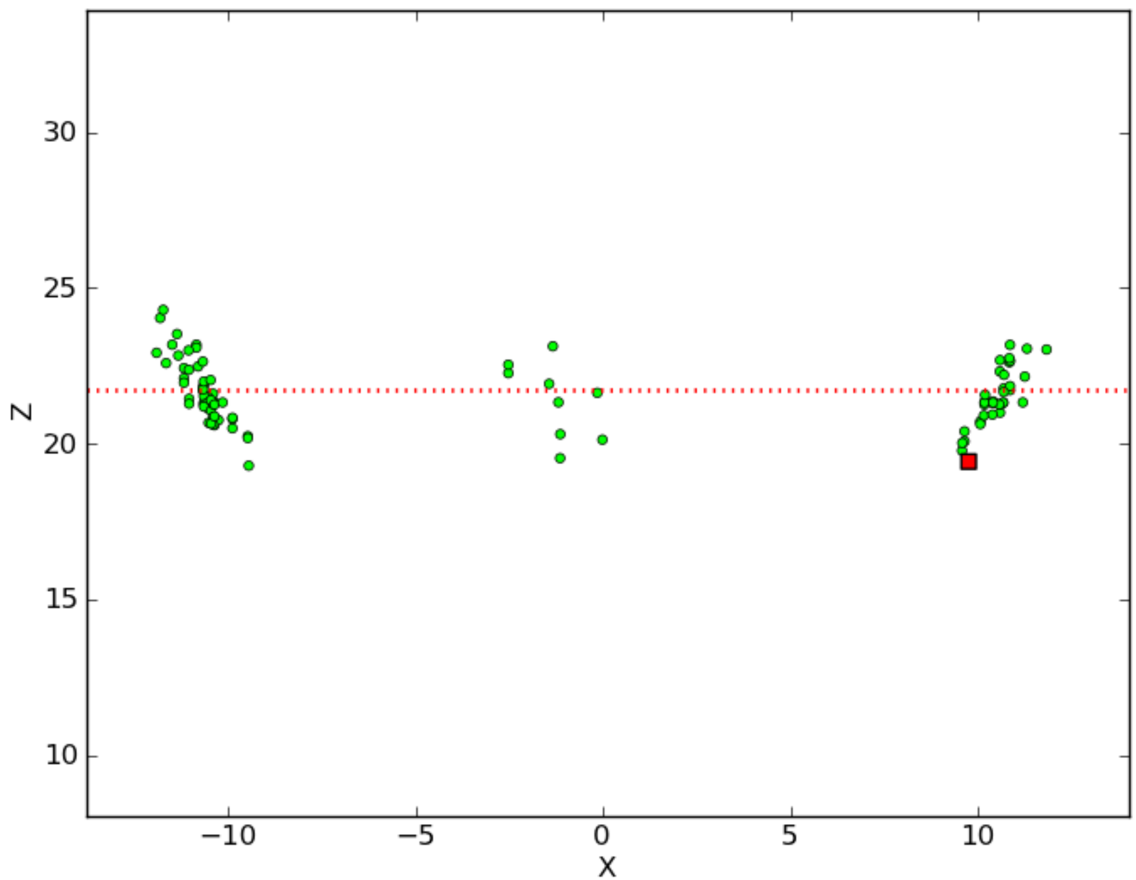
The analysis value for X is obtained by preserving the particle position in the marginal CDFs.

The Multivariate Rank Histogram Filter



This is done for each particle.

The Multivariate Rank Histogram Filter



Analysis ensemble.

The Multivariate Rank Histogram Filter

With 3 variables:

$$p(x, y, z|z^o) = p(z|z^o)p(x|z, z^o)p(y|x, z, z^o)$$

To sample $p(x|z=z^a_i)$, particles are selected based on their distance to z^a_i along z axis.

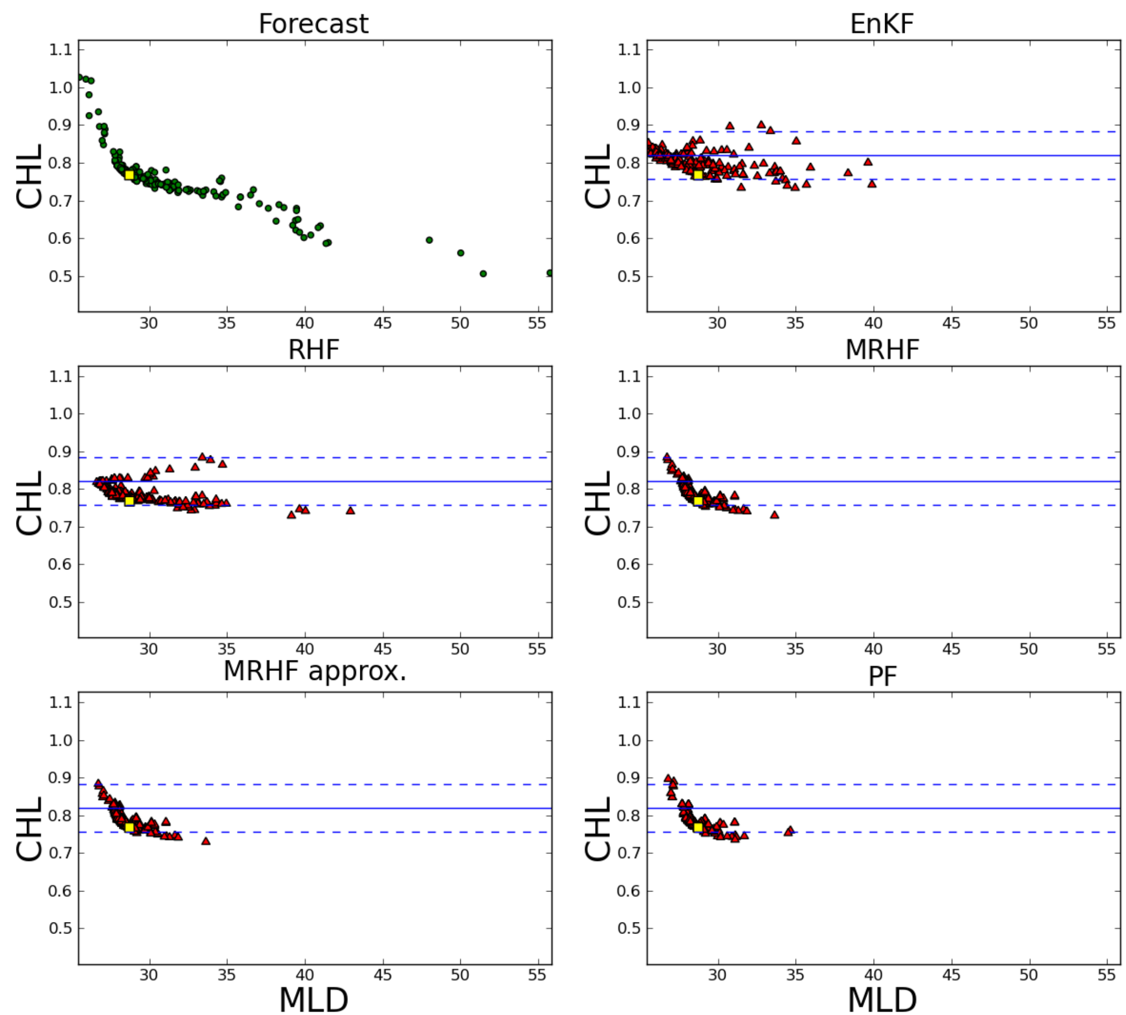
To sample $p(y|x=x^a_i, z=z^a_i)$, particles are selected based on their distance to (x^a_i, z^a_i) in the (x,z) plane.

Analysis illustration

NATL0.25+BGC

Observed: CHL

Unobserved: MLD and DET

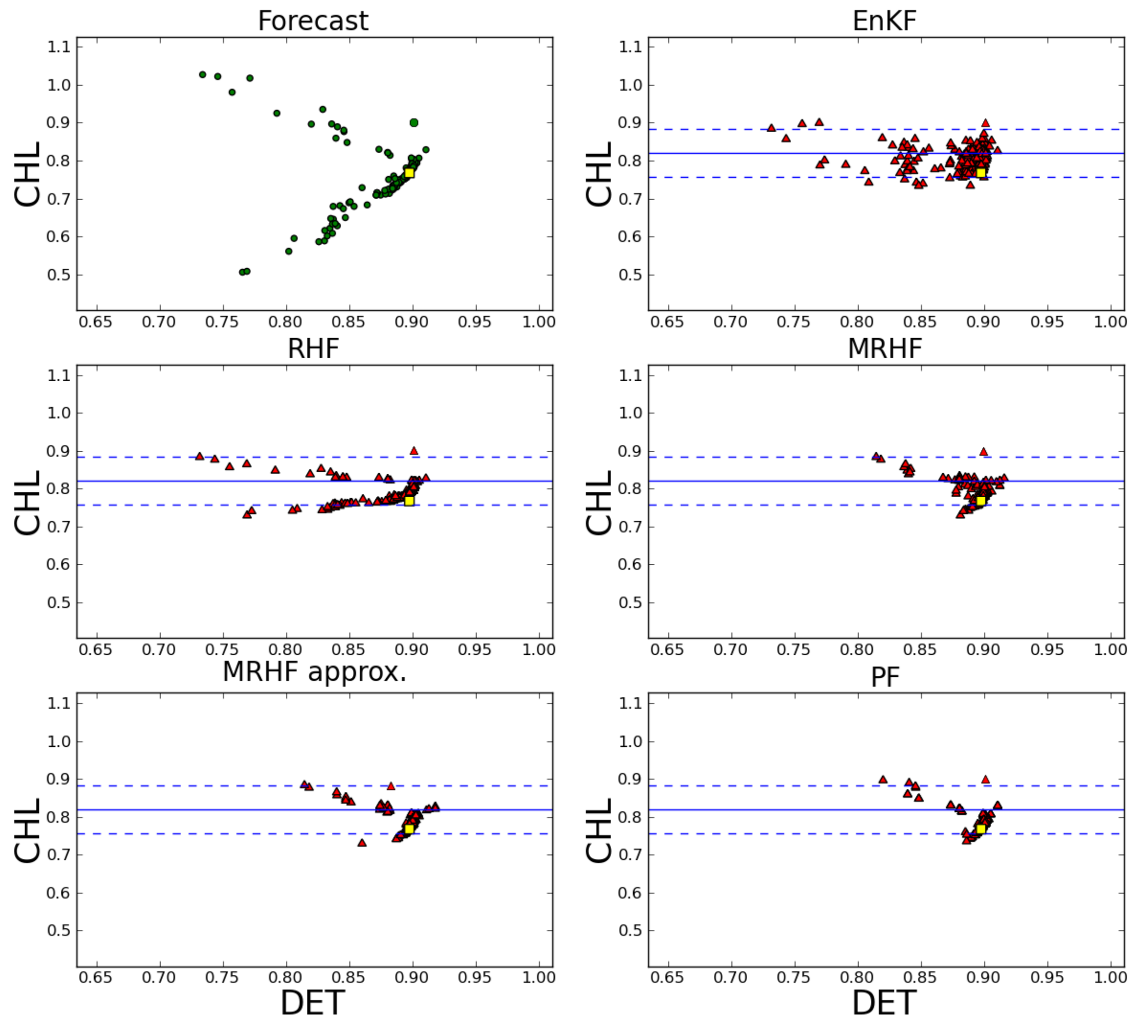


Analysis illustration

NATL0.25+BGC

Observed: CHL

Unobserved: MLD and DET

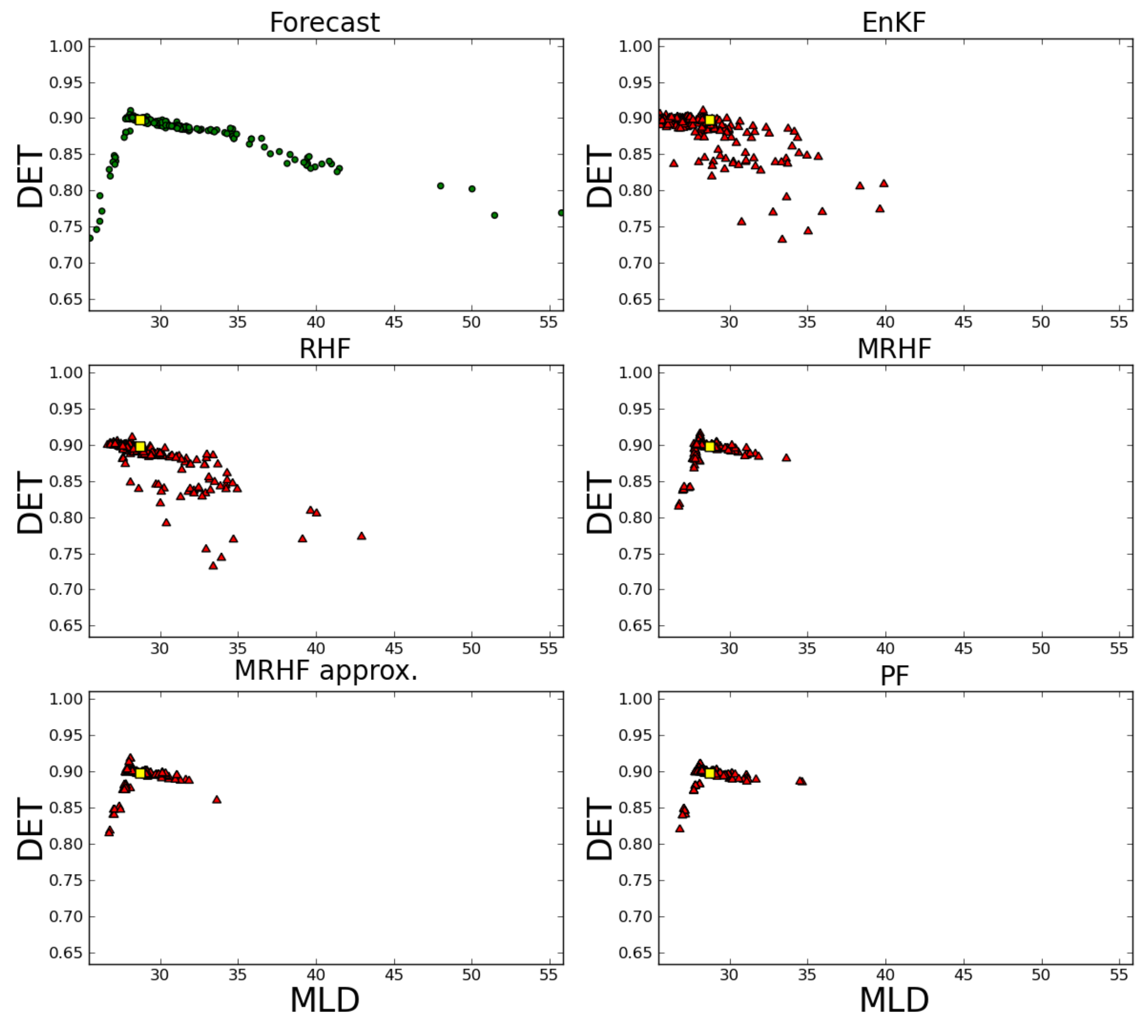


Analysis illustration

NATL0.25+BGC

Observed: CHL

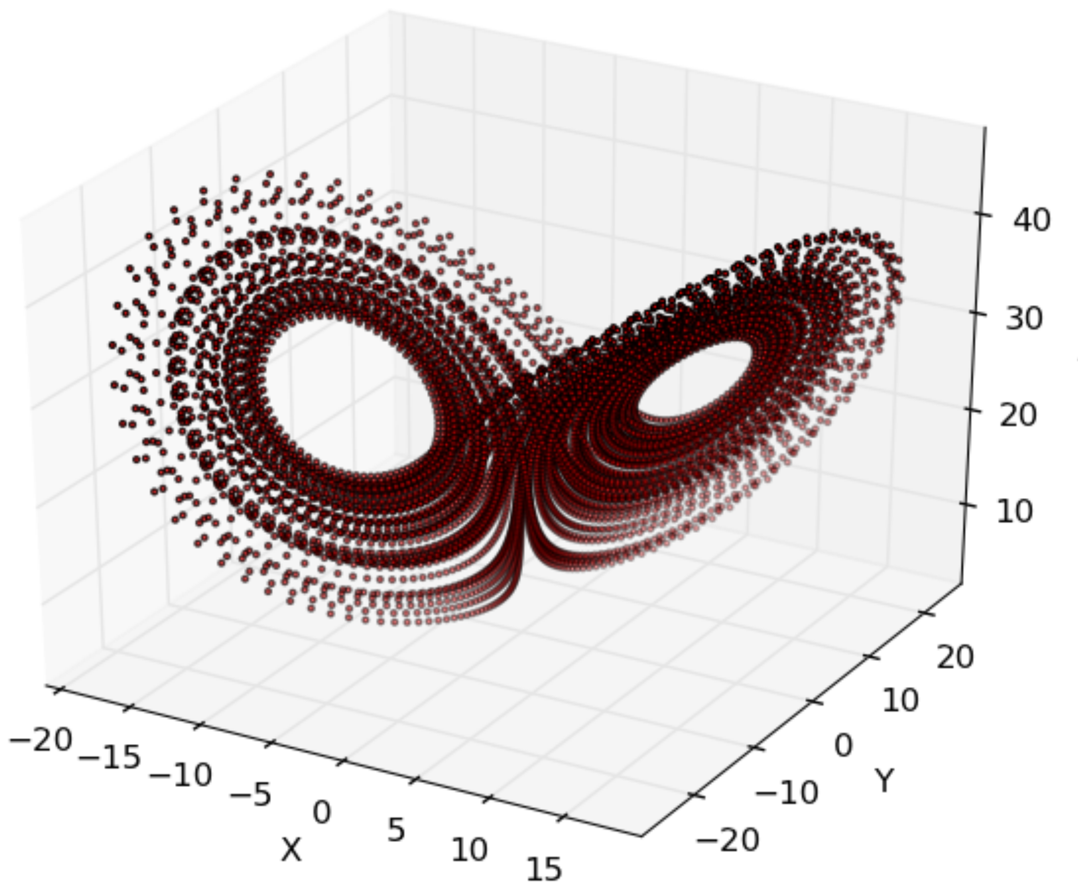
Unobserved: MLD and DET



Experiments with Lorenz 63 system

$$\begin{cases} \dot{x} &= 10(y - x) \\ \dot{y} &= 28x - y - xz \\ \dot{z} &= xy - \frac{8}{3}z \end{cases}$$

Experiments with Lorenz 63 system



Experiments with Lorenz 63 system

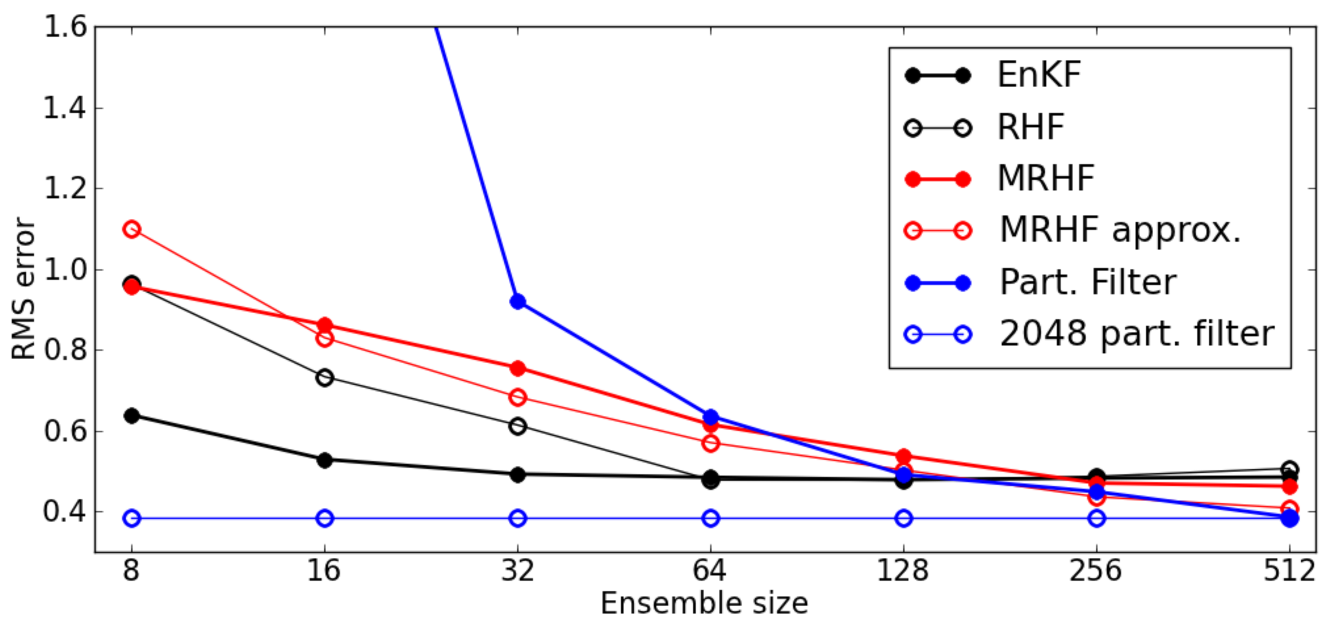
- Experiment 1:
  - X, Y, Z observed (error std=2);
  - Analysis every 10 (weakly), 25 (moderately), and 50 (strongly nonlinear case) time steps.
- Experiment 2:
  - Only Z observed (error std=1);
  - Analysis every 40 time steps.

Experiments with Lorenz 63 system

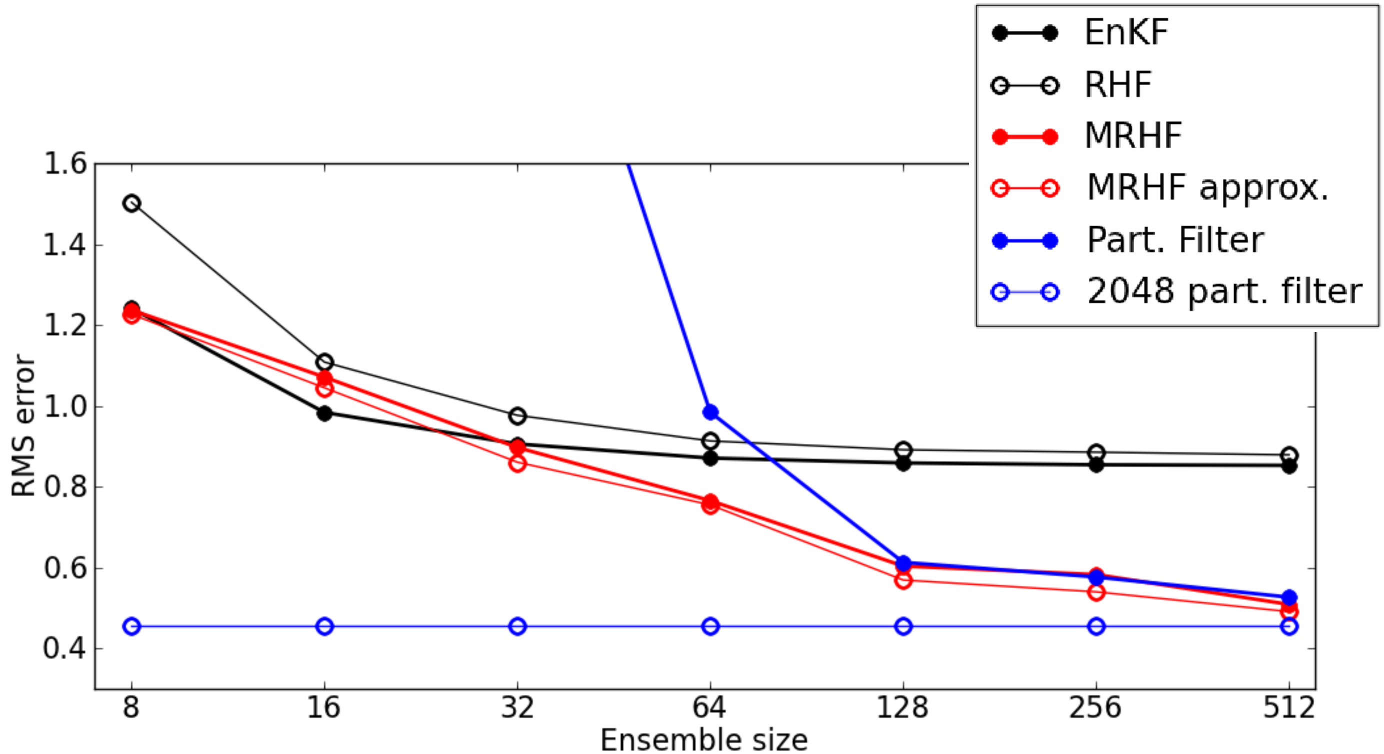
- 10^5 analysis steps;
- Diagnostics: RMS error and Kullback-Leibler divergence (ref: large ensemble SIR filter).

$$d(P,Q) = \int \log \frac{P}{Q} dP.$$

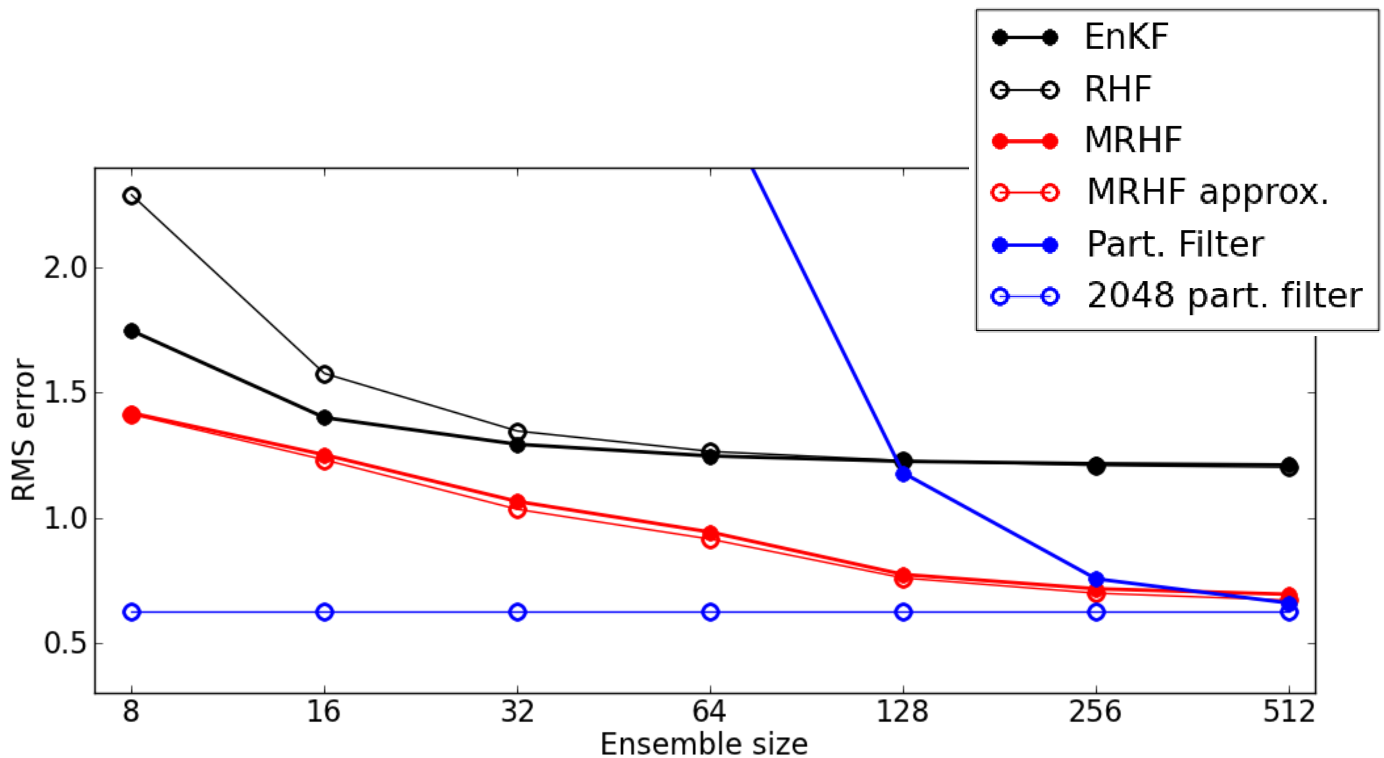
L63, XYZ observed, dt=10
RMS error



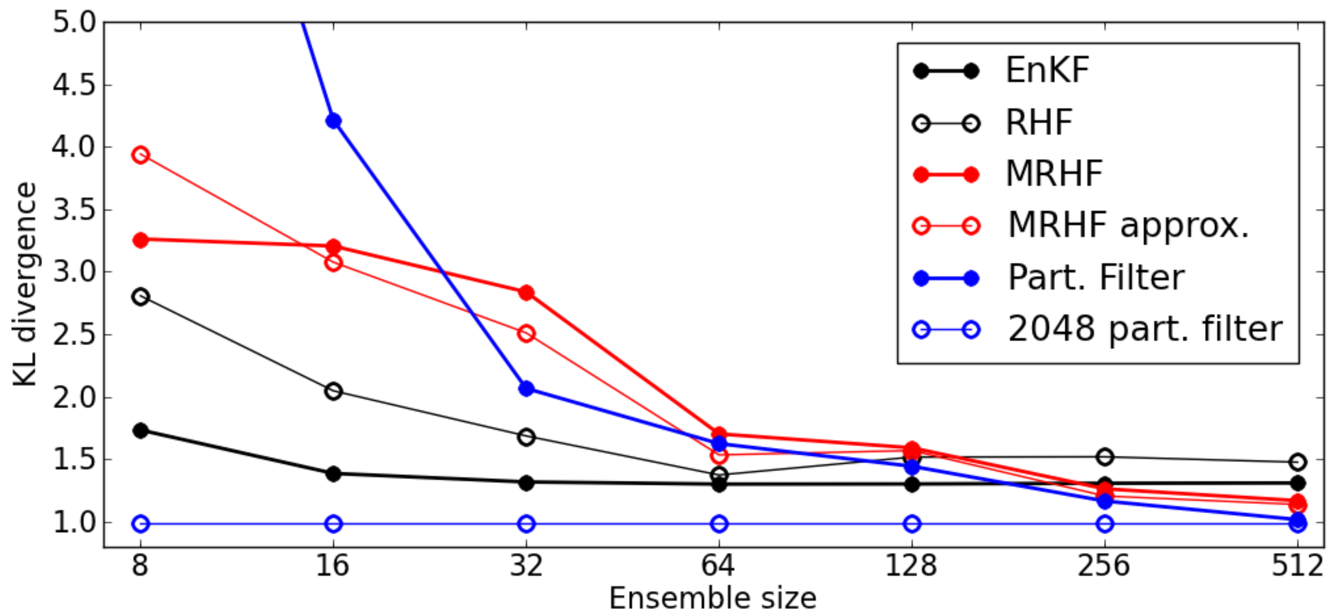
L63, XYZ observed, dt=25
RMS error



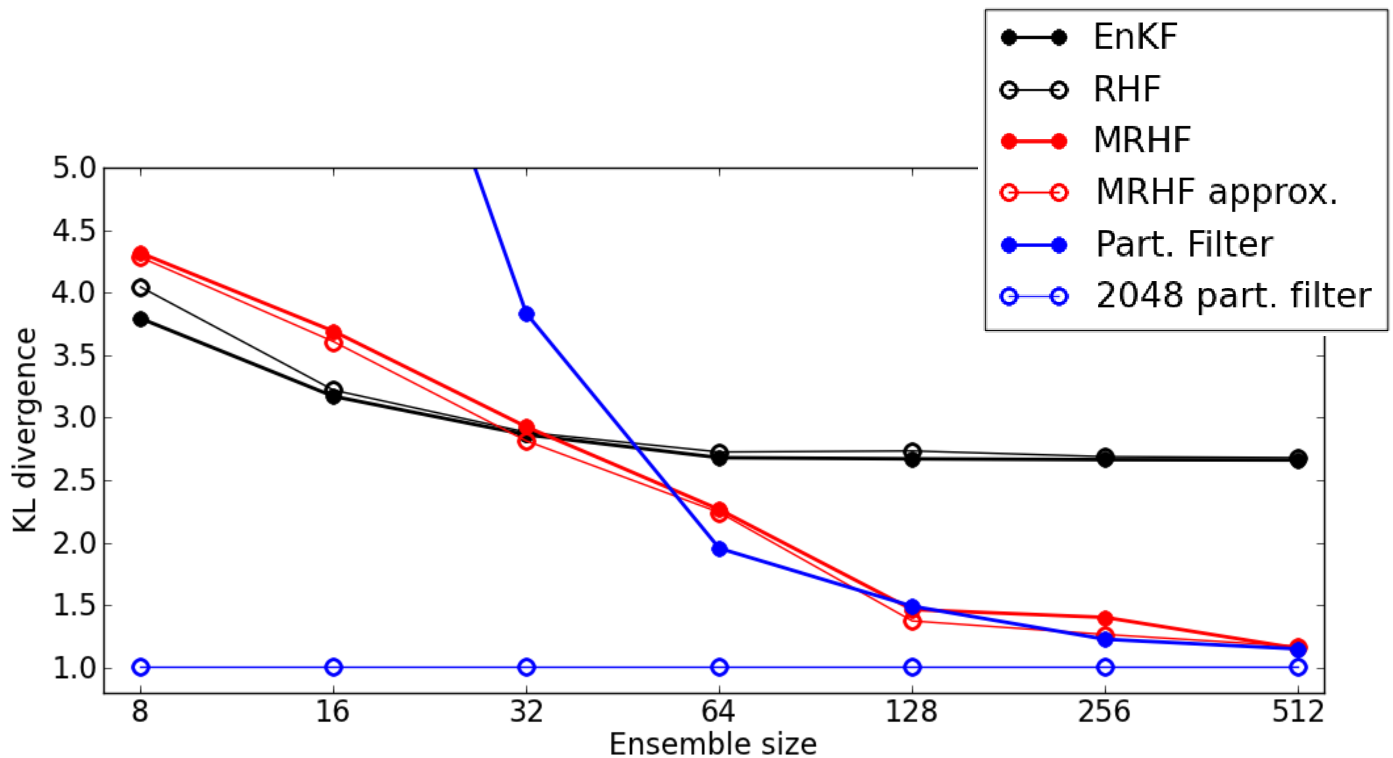
L63, XYZ observed, dt=50
RMS error



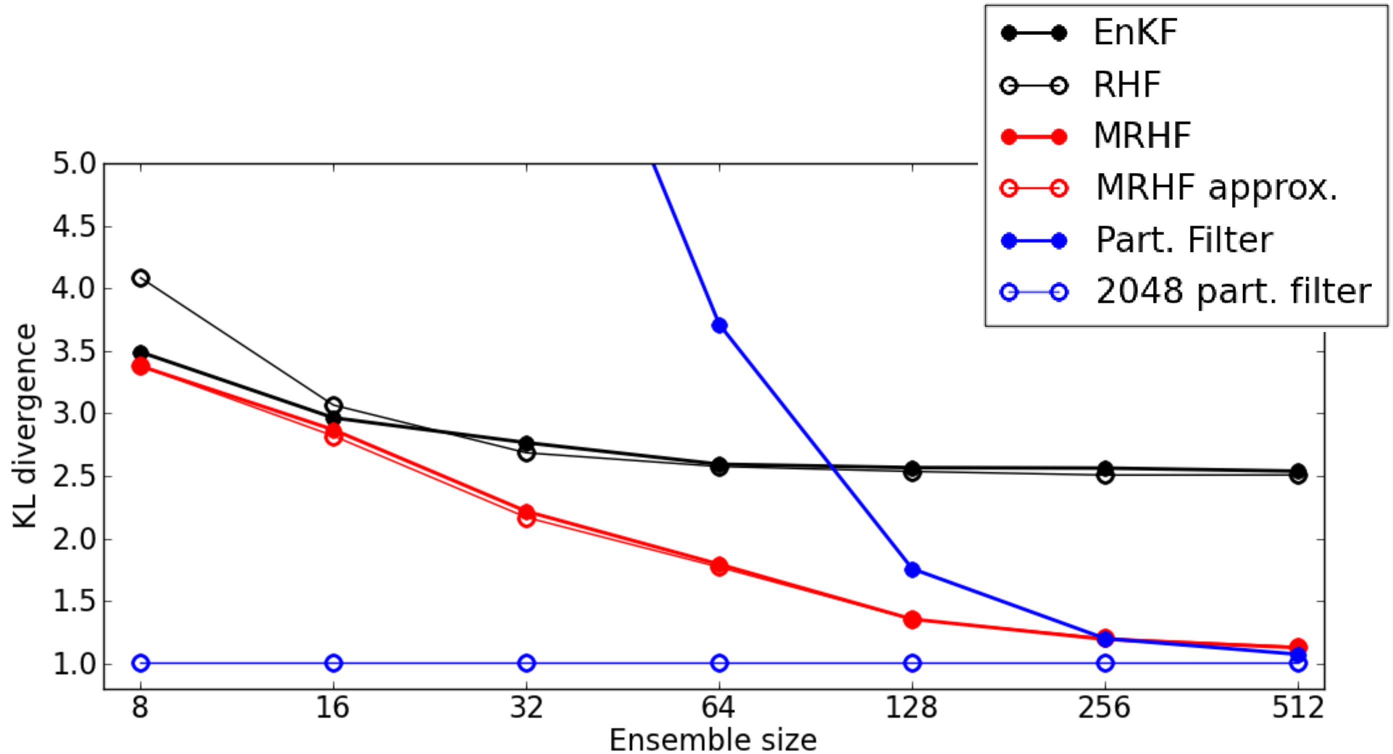
L63, XYZ observed, dt=10
Kullback-Leibler divergence



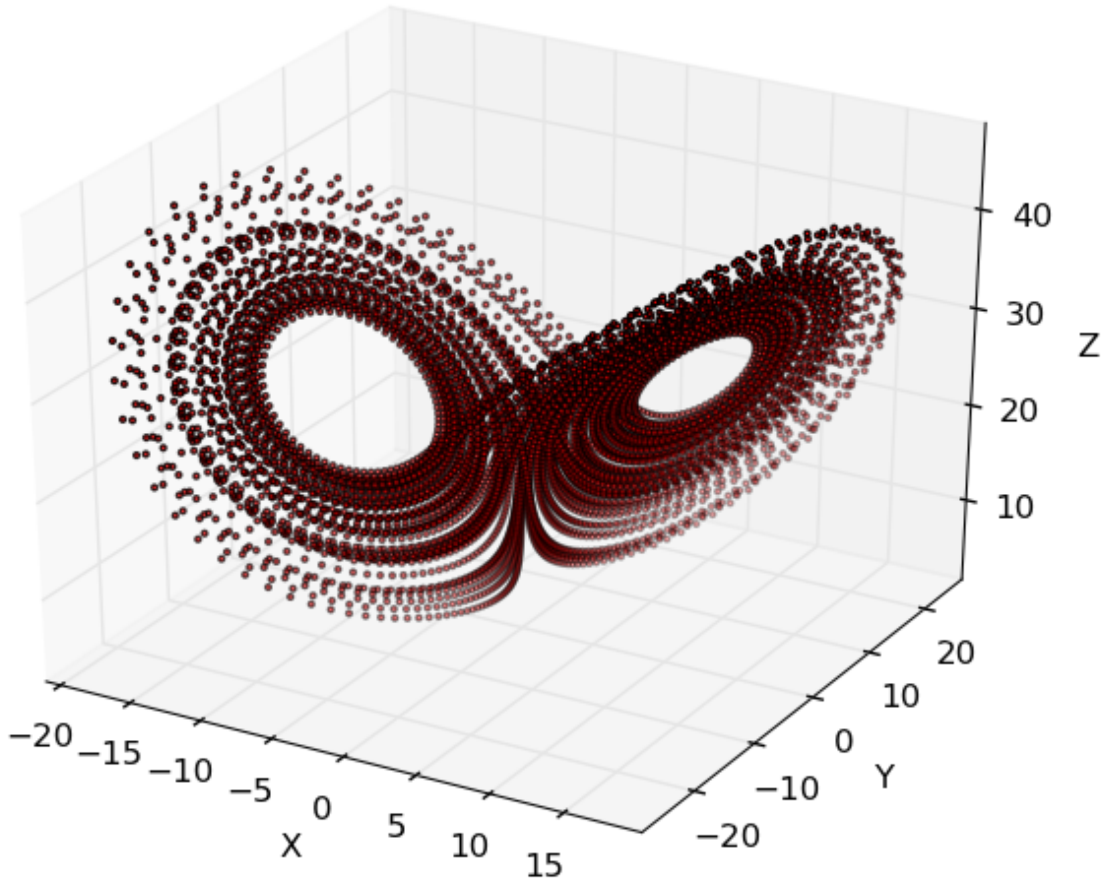
L63, XYZ observed, dt=25 Kullback-Leibler divergence



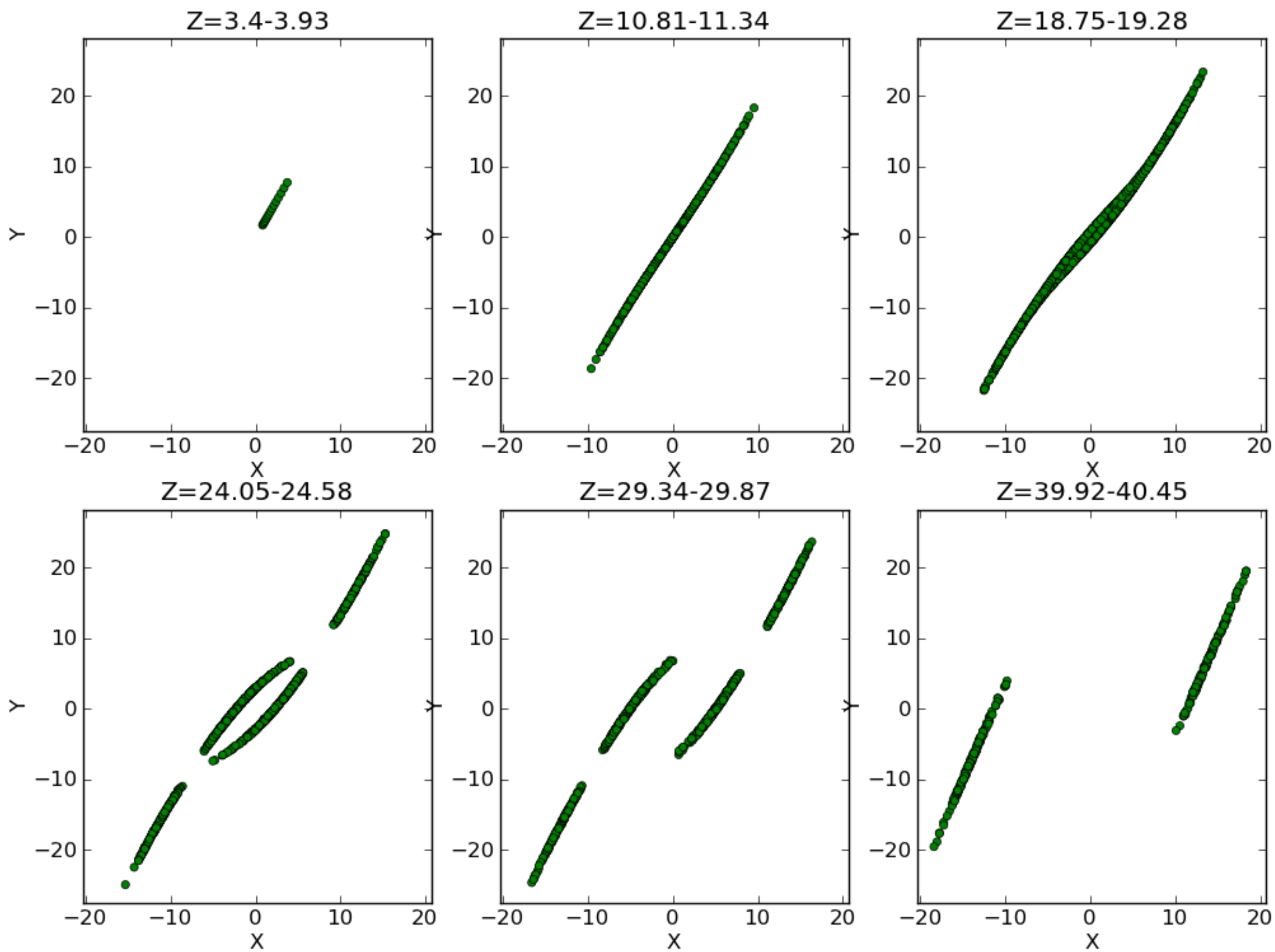
L63, XYZ observed, dt=50 Kullback-Leibler divergence



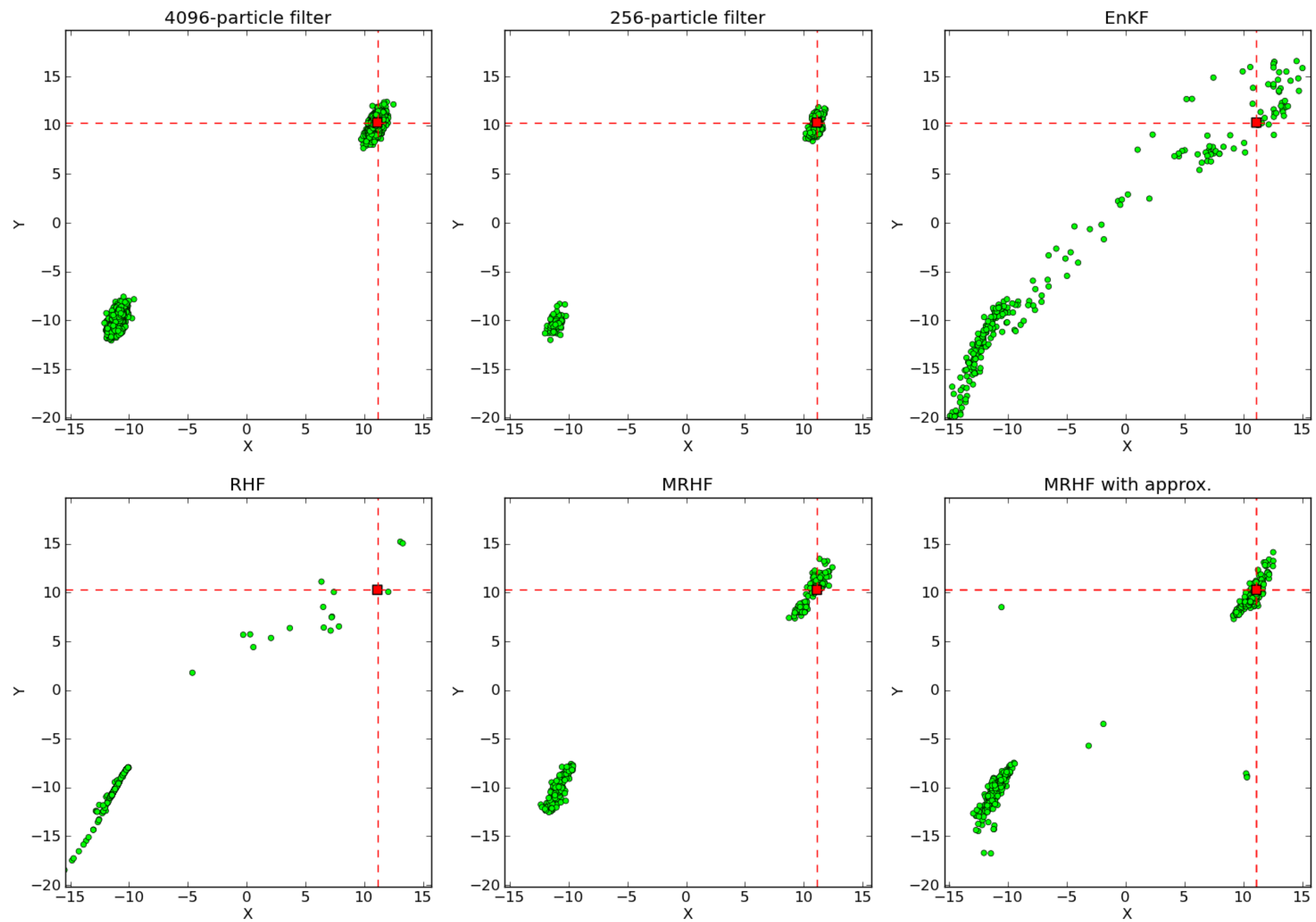
L63, only Z observed



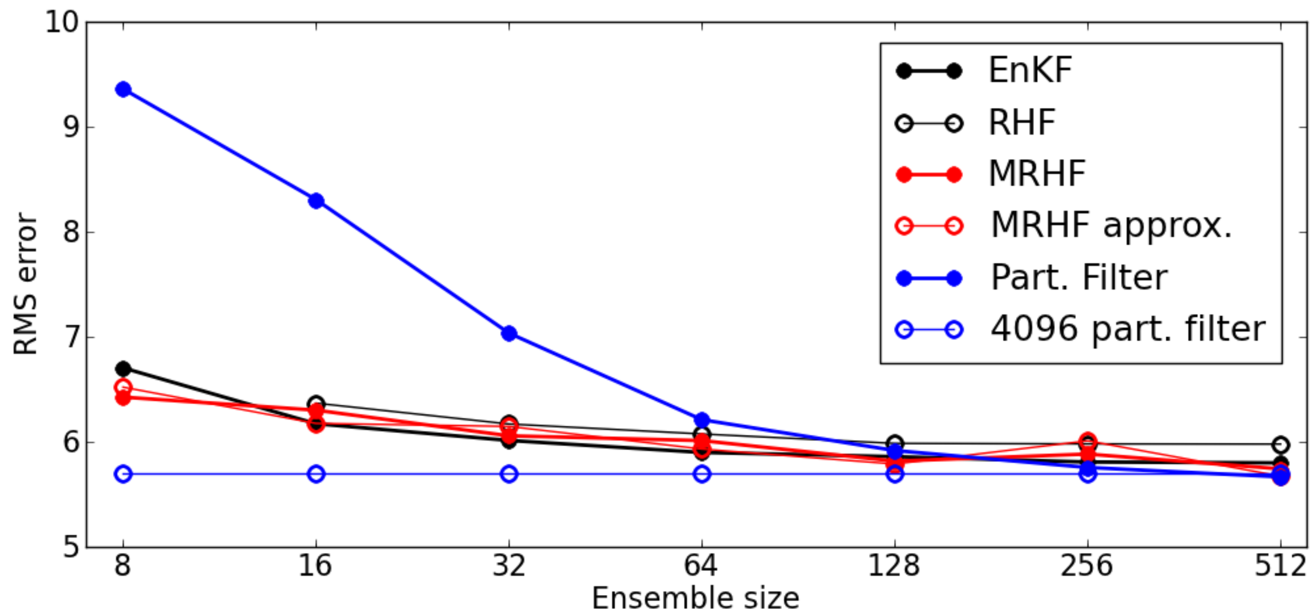
L63, only Z observed



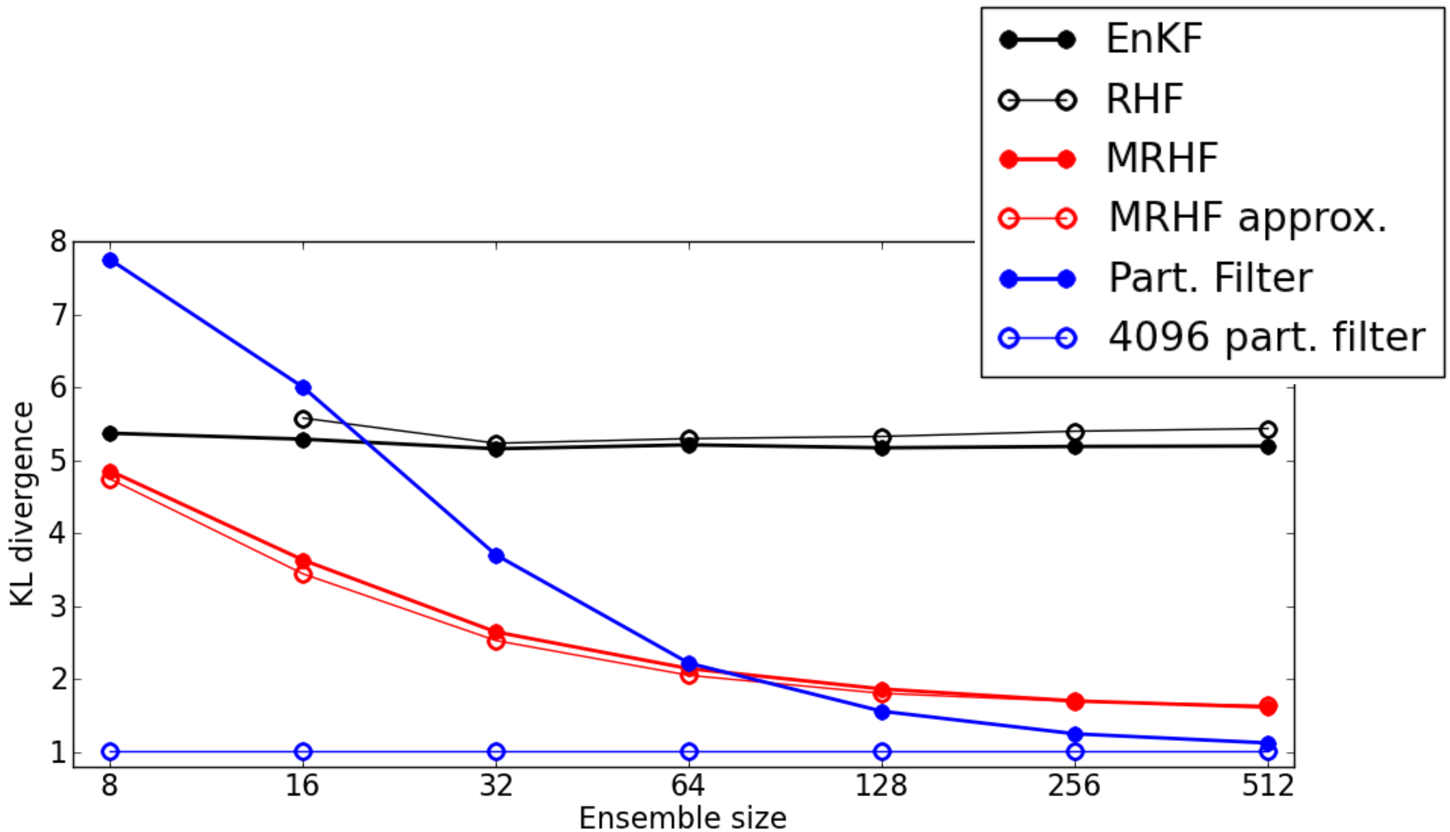
L63, only Z observed Analyses at cycle 364



L63, only Z observed
RMS error



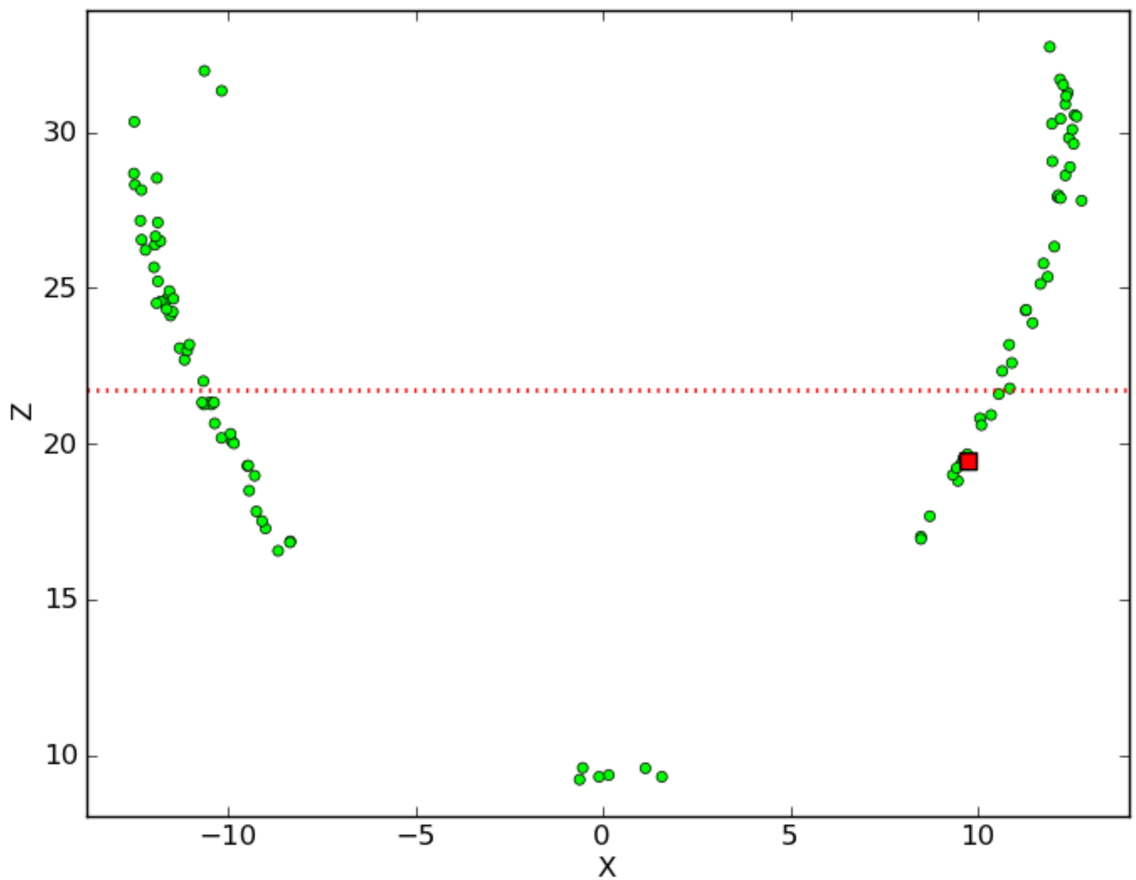
L63, only Z observed
Kullback-Leibler divergence



Limitation

- In its original form, the MRHF is subject to the curse of dimensionality!!

The Multivariate Rank Histogram Filter



Background ensemble in X − Z plane.
Red dotted line: Z obs. Red square: truth.

Limitation

- In its original form, the MRHF is subject to the curse of dimensionality!!
- Need for the mean-field approximation:

$$p(x, y, z|z^o) = p(z|z^o)p(x|z, z^o)p(y|\cancel{x}, z, z^o)$$

Limitation

- In its original form, the MRHF is subject to the curse of dimensionality!!
- Need for the mean-field approximation;
- Extremely expensive in the present form (work in progress)

Last remarks

- The MRHF is one implementation of EnDA based on Optimal transport theory (see S. Reich's papers)
- The MRHF follows the logic of the stochastic EnKF (coupling possible)
- Easy localization of the analysis
- Smoothing straightforward

Introduction

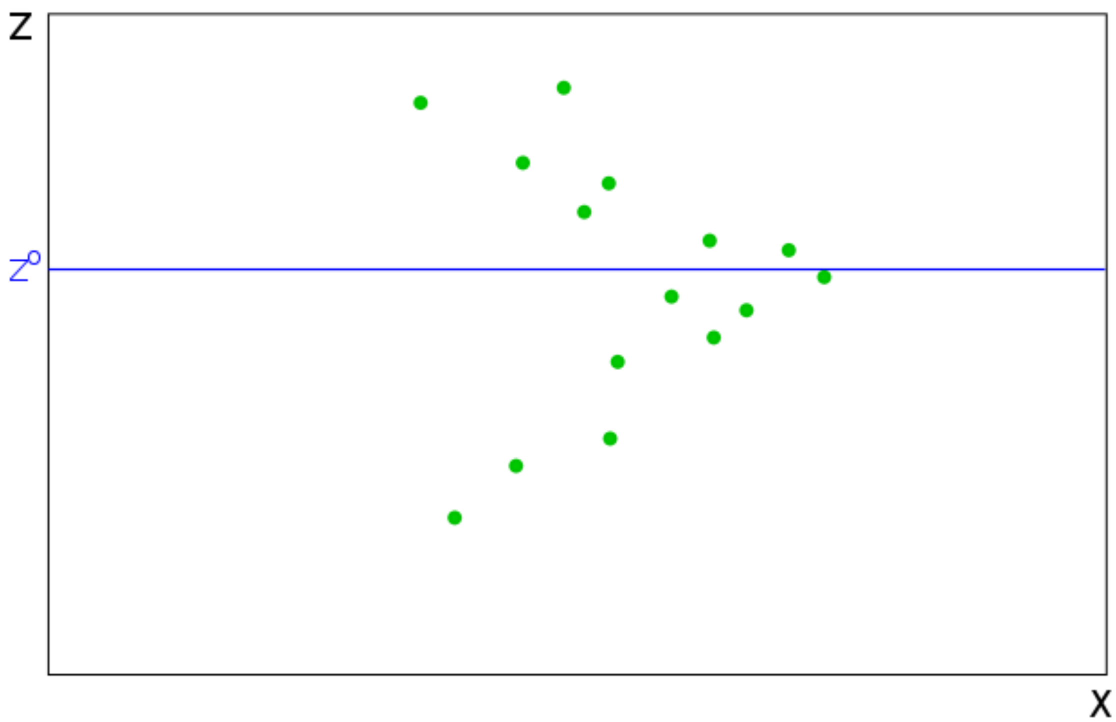
	transform	sampling
Parametric	EnKF (includes ETKF) Truncated-Gaussian EnKF (1)	
Semi-parametric	EnKF with Gaussian anamorphosis (2)	
Non-parametric	ETPF(4) MRHF (5)	Particle filters (3)

(1) Lauvernet et al (2009)
(2) Holm et al (2002), Bertino et al (2003); Simon and Bertino (2009); Béal et al (2010); Brankart et al (2012)
(3) Gordon et al (1993); Van Leeuwen et al (2009, 2010), Snyder et al (2008)
(4) Reich (2013)
(5) Metref et al (in revision)

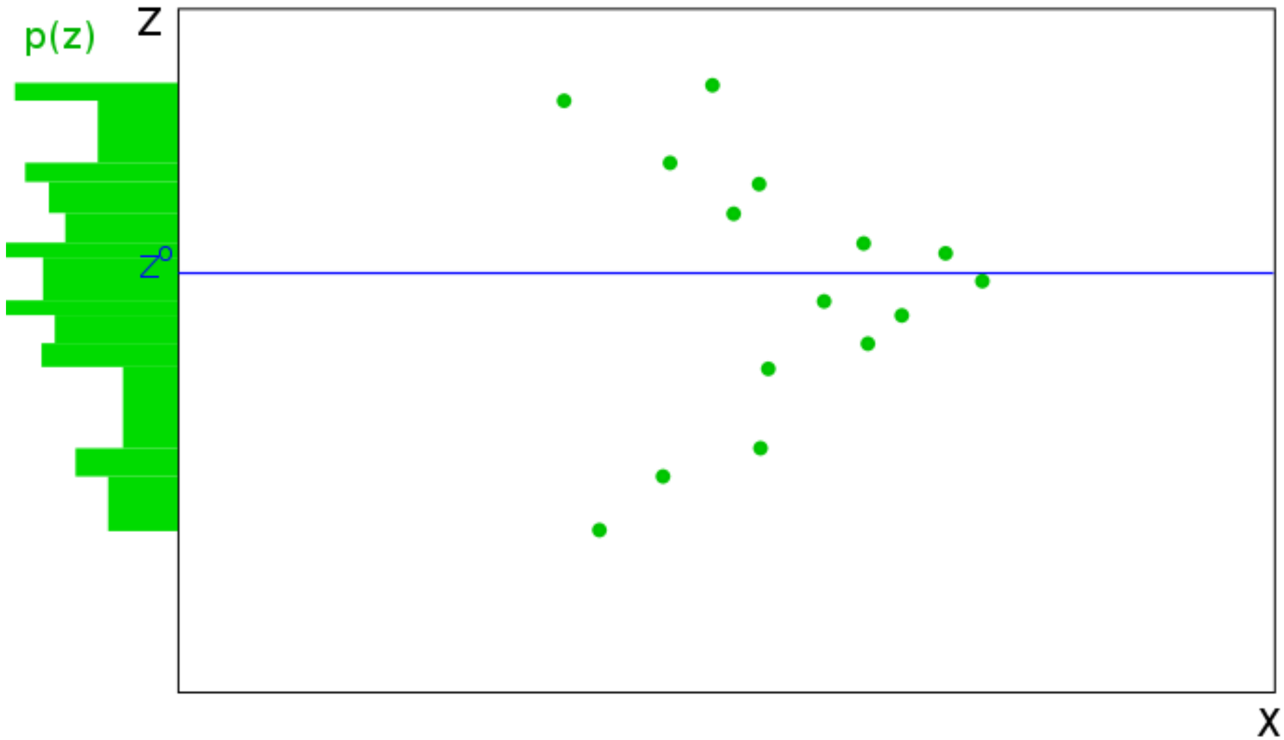
Acknowledgements

This work still under development has been and is being supported by the European Commission (SANGOMA project), CNRS (LEFE/MANU), NCAR, and Region Rhône Alpes.

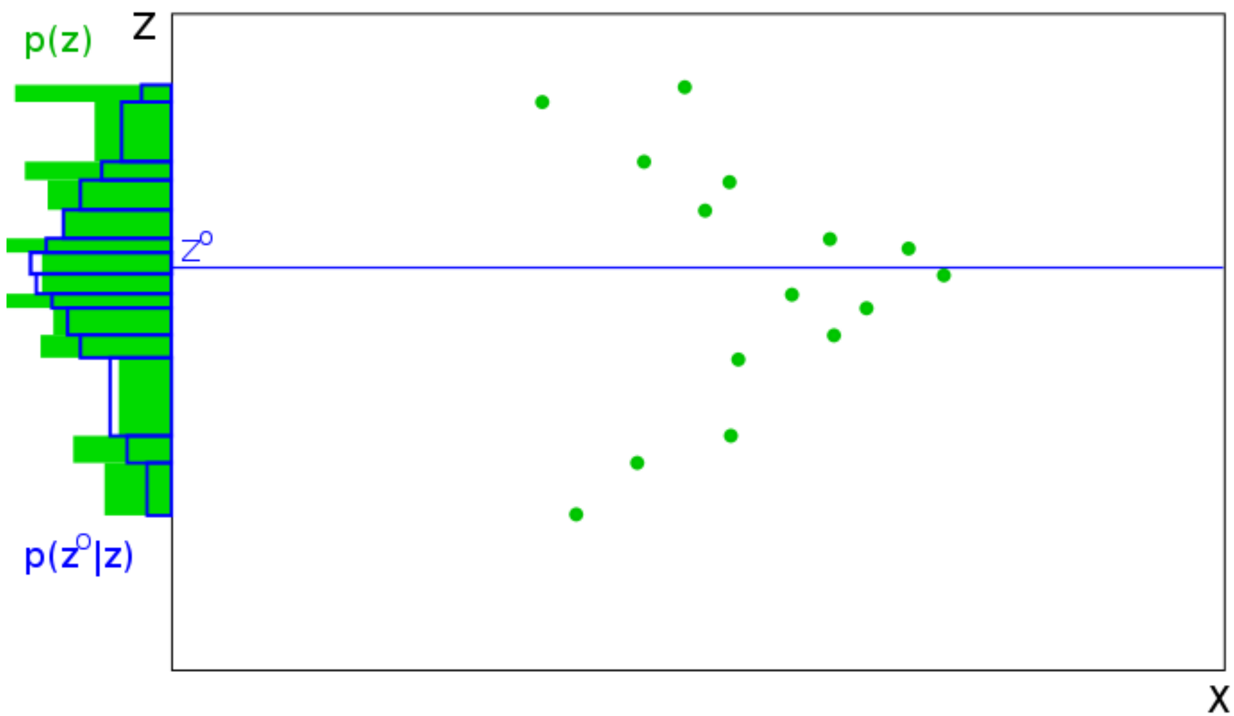
MRHF methodology



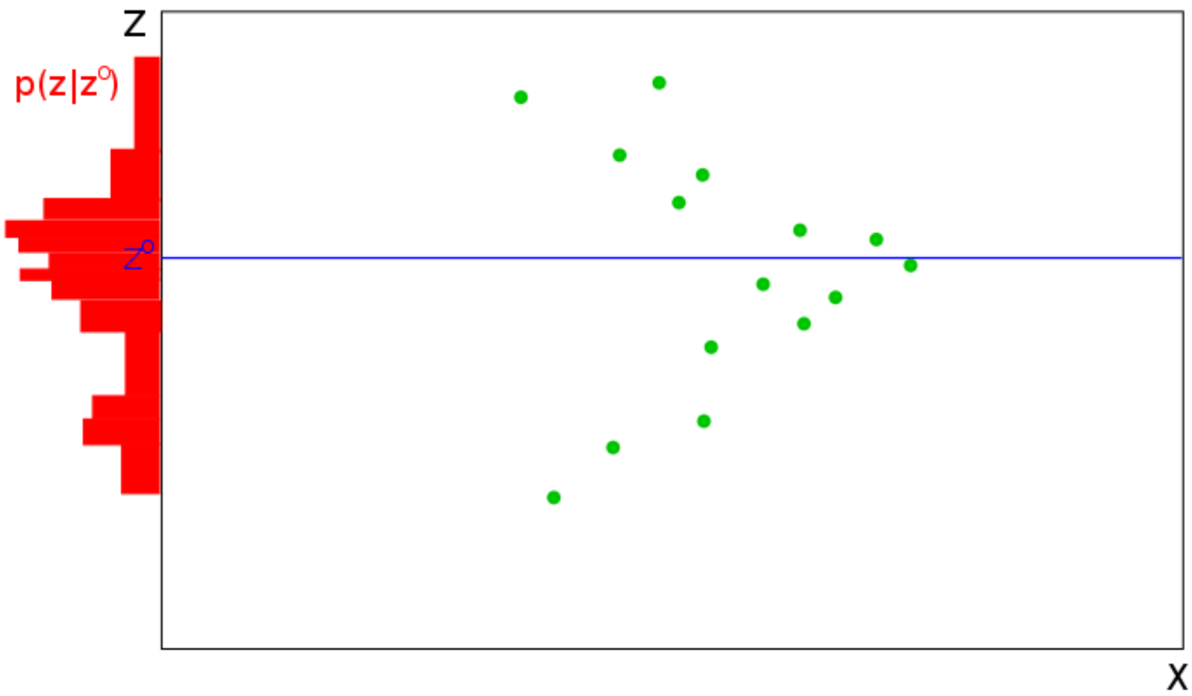
MRHF methodology



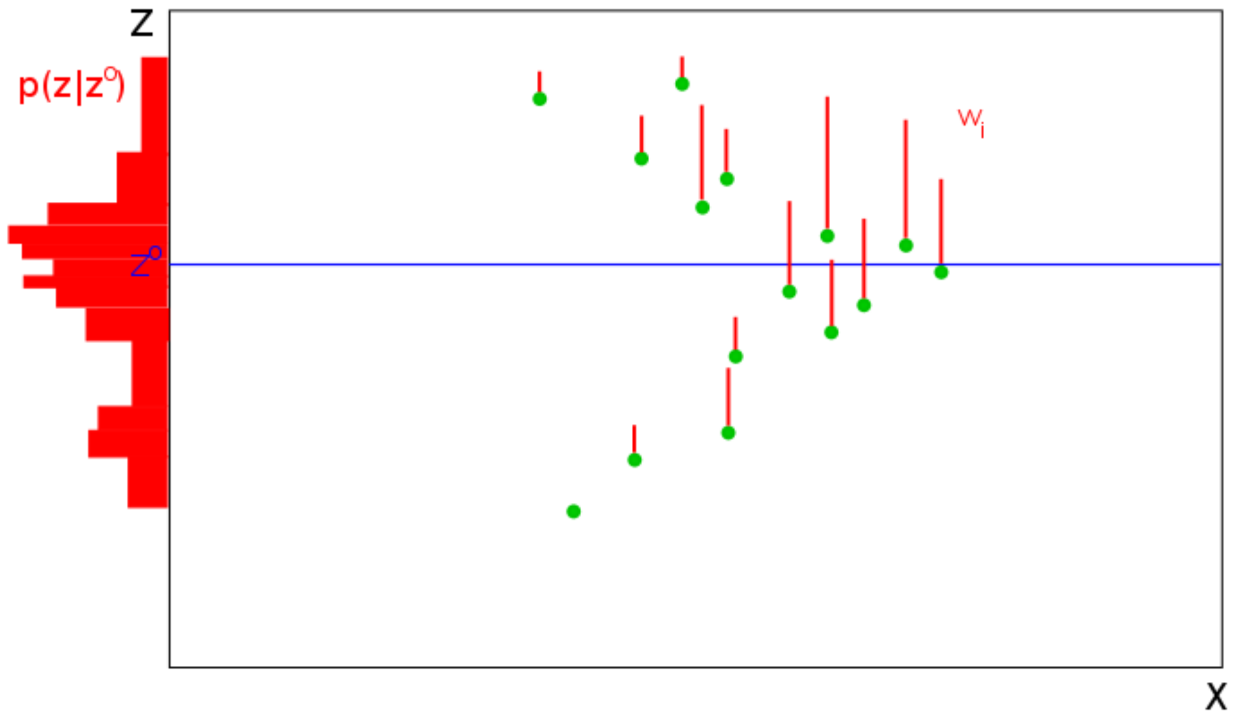
MRHF methodology



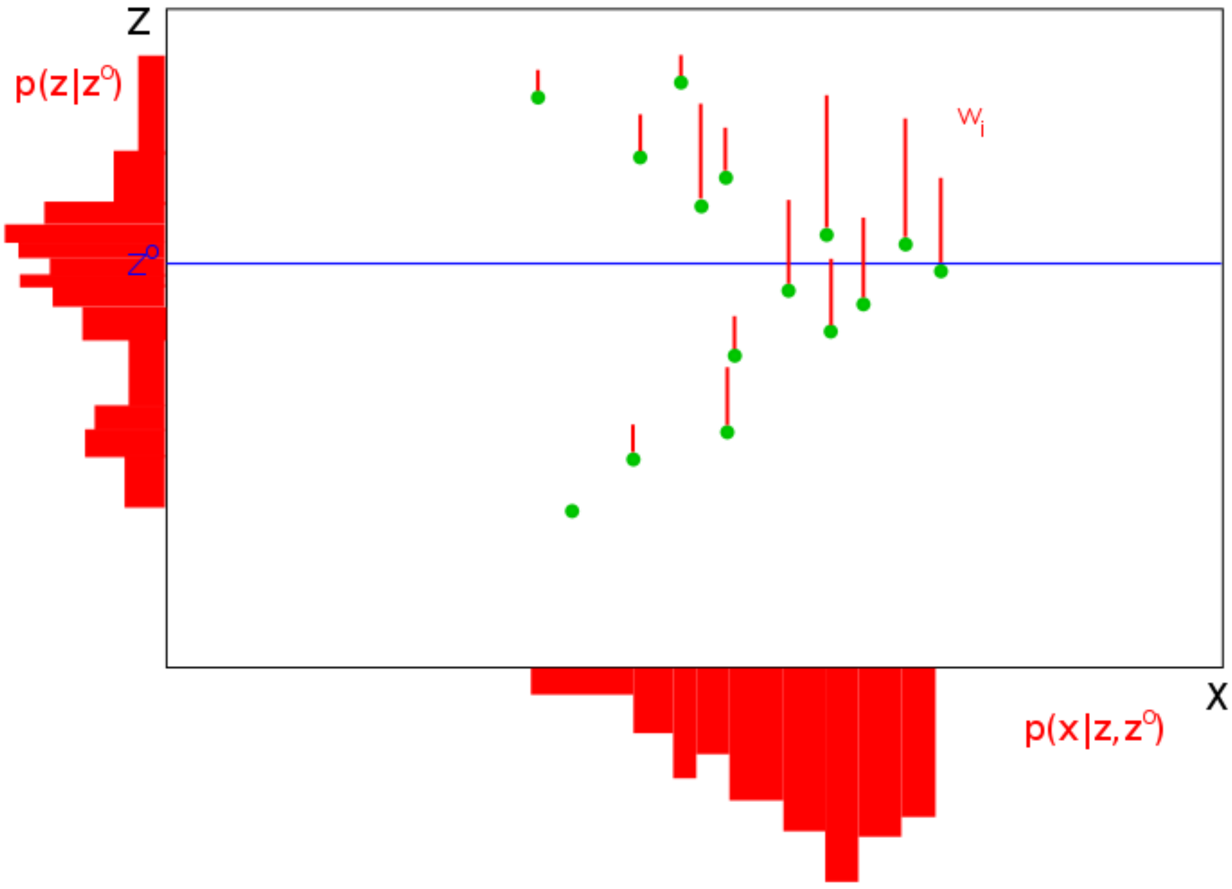
MRHF methodology



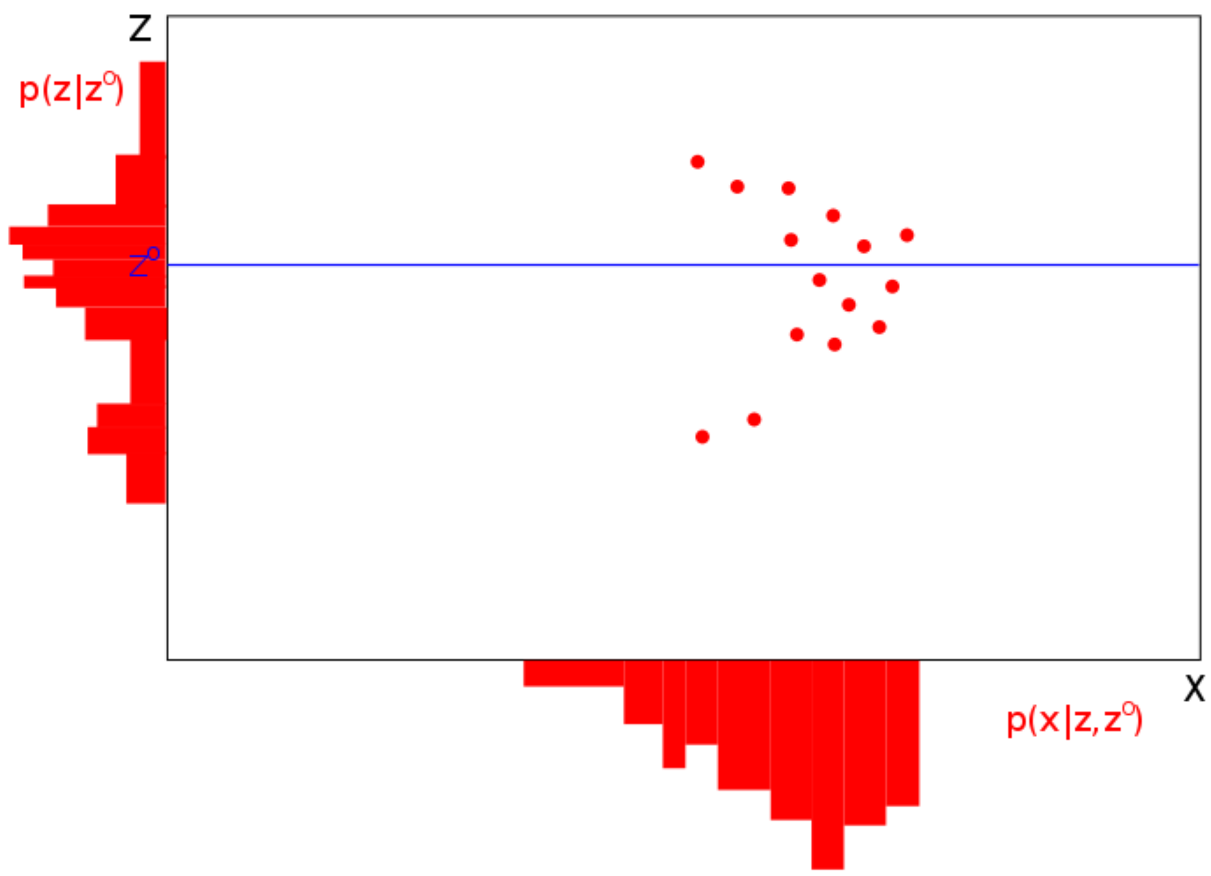
MRHF methodology



MRHF methodology



MRHF methodology



Local least square fit (Anderson, 2003)

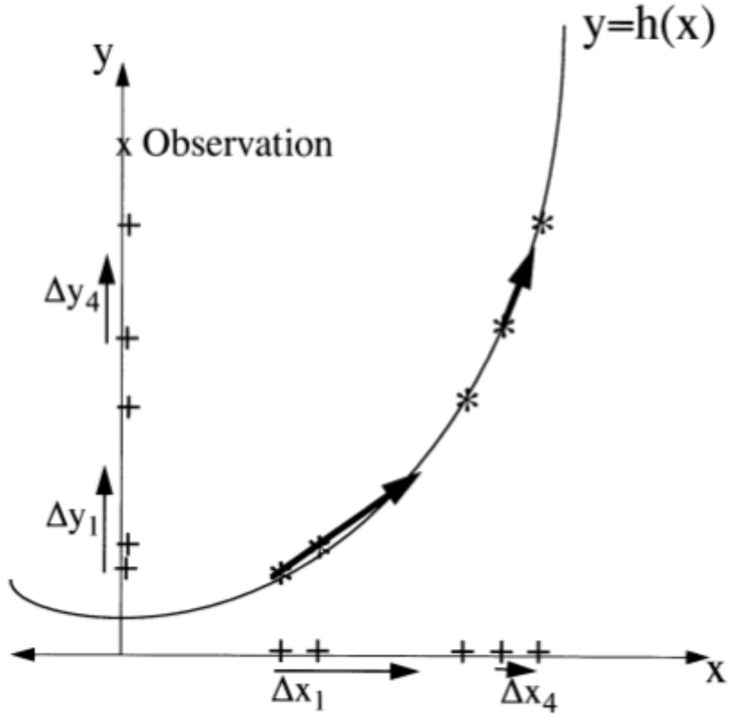


FIG. 2. As in Fig. 1 but showing the application of local least squares fits, in this case using only the nearest neighbor in y, to compute the updates for x given the updates for y. The local updates for the first and fourth ensemble members are shown by the black vectors.

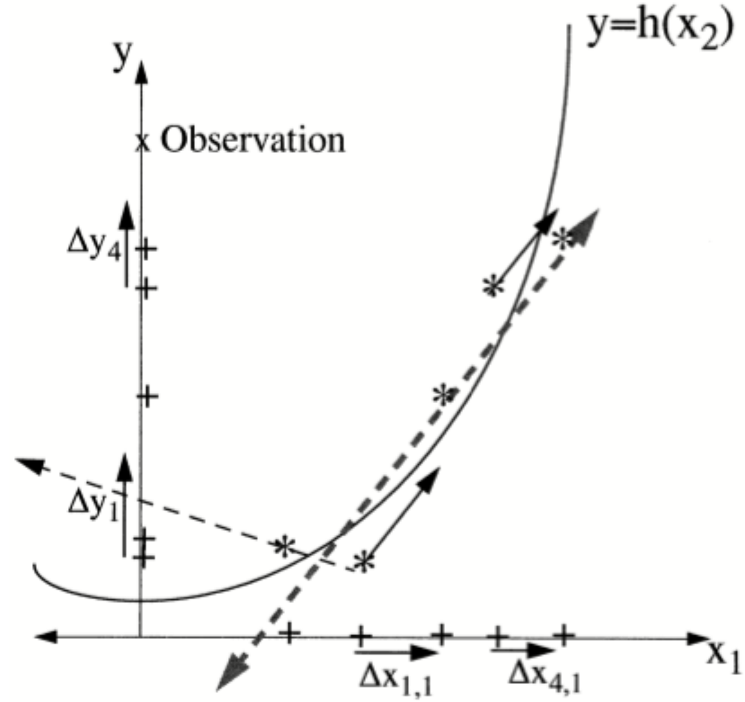


FIG. 3. As in Fig. 1 but now $y = h(x_2)$, where x_2 is a second state variable that is moderately correlated with x_1. The thin dashed vector demonstrates the hazard of using local least squares fits when the observation variable y and the state variable x_1 are not functionally related.