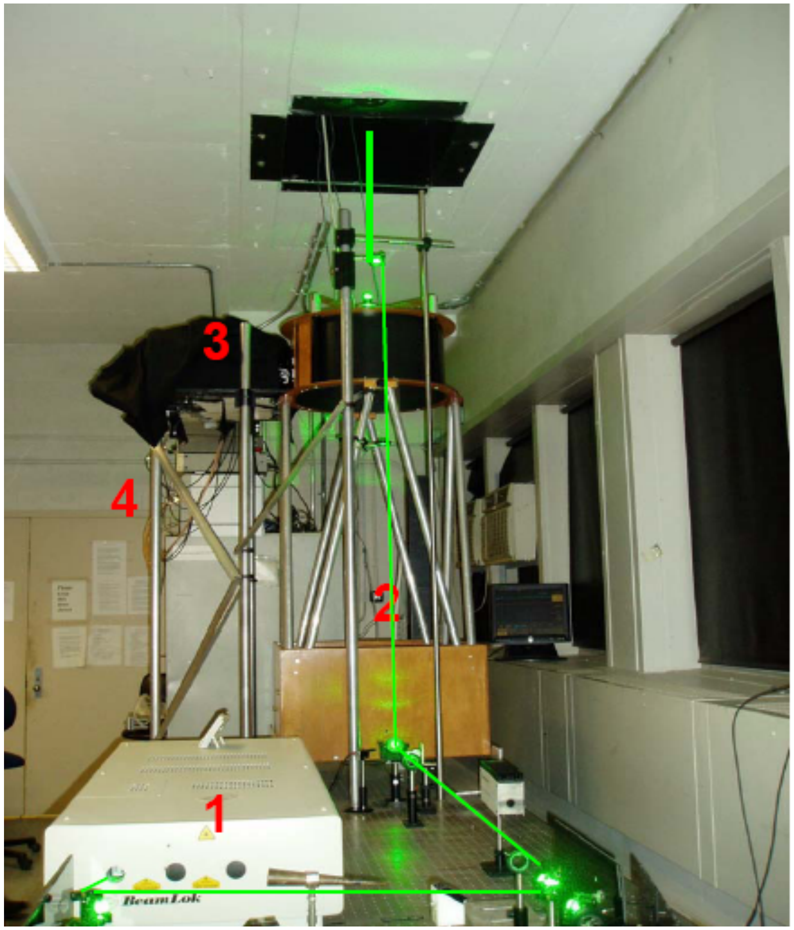
Stochastic Monte Carlo methods for non-linear statistical inverse problems

Benjamin R. Herman
Department of Electrical Engineering
City College of New York
herman@ee.ccny.cuny.edu

Overview

- Aerosol lidar system
- Retrieval of aerosol properties from lidar measurements
- Metropolis-Hastings algorithm for uncertainty analysis
- Techniques for its efficient implementation

CCNY Aerosol/Water vapor lidar



Transmits light at 1064, 532, and 355nm

Detects signals at transmitted wavelengths as well as two wavelengths shifted from 355nm by Raman scattering from N2 and H2O

1: Three wavelength pulsed Nd:YAG Laser
2: Newtonian Telescope
3: Receiver optics and photo-detectors
4: Licel transient recorder, analog and photon counting

Optical coefficients and lidar equations

Elastic lidar equation $P(R,\lambda_r=\lambda_t)=\frac{E\,C\,O(R)}{R^2}\cdot\left(\beta_a(R,\lambda_t)+\beta_m(R,\lambda_t)\right)\cdot T_m{}^2(R,\lambda_t)\cdot T_a{}^2(R,\lambda_t)$

Raman lidar equation $P(R,\lambda_r)=\frac{E\,C\,O(R)}{R^2}\cdot\beta_{N_2,\mathrm{Ram}}(R,\lambda_t,\lambda_r)\cdot T_m(R,\lambda_t)\cdot T_m(R,\lambda_r)\cdot T_a(R,\lambda_t)\cdot T_a(R,\lambda_r)$

$$T_{a|m}(R,\lambda)=\exp\left(-\int_0^R \alpha_{a|m}(R',\lambda)\mathrm{d}R'\right)$$

R - Range

$P(R,\lambda_r)$ - Optical signal

λ_t - Transmitted wavelength

λ_r - Received wavelength

E - Laser pulse energy

C - System calibration coefficient

$O(R)$ - Geometric overlap function

$\beta_{a|m}$ - Elastic backscatter coefficient (aerosol or molecular)

$\beta_{N_2,\mathrm{Ram}}$ - Nitrogen Raman backscatter coefficient

$T_{a|m}$ - Optical transmission

$\alpha_{a|m}$ - Extinction coefficient

Multi-mode aerosol model and optical coefficients

Aerosol is modeled as being composed of different constituents, each with a size distribution and index of refraction

Optical coefficient dependency

$$y_n(\boldsymbol{\theta}) = \sum_{i=1}^{M} V_i \int_0^\infty n(r;\bar{r}_i,\sigma_i) K_n(r;m_i)\,\mathrm{d}r$$

Aerosol state vector

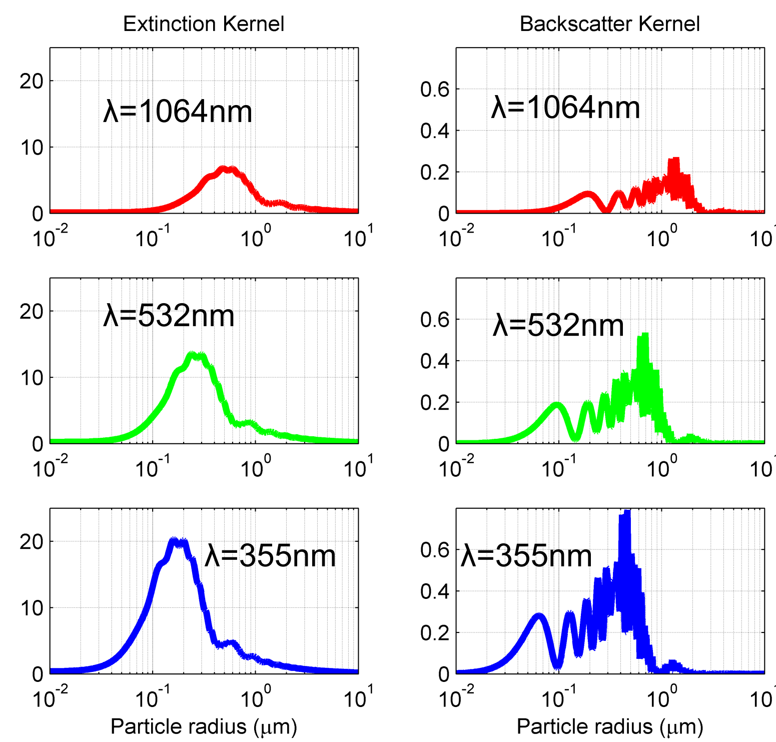
$$\boldsymbol{\theta} = (V_1,\bar{r}_1,\sigma_1,m_1 \cdots, V_M,\bar{r}_M,\sigma_M,m_M)$$

Volume concentration

Median Radius

Geometric standard deviation

Refractive index



Uncertainty analysis

Uncertainty is expressed in the form of a posterior probability density function (PDF)

$$f_{\text{posterior}}(\boldsymbol{\theta} \mid \hat{\mathbf{y}}) = \frac{f_{\text{likelihood}}(\hat{\mathbf{y}} \mid \boldsymbol{\theta}) f_{\text{prior}}(\boldsymbol{\theta})}{f_{\text{marginal}}(\hat{\mathbf{y}})}$$

$$f_{\text{likelihood}}(\hat{\mathbf{y}}, \boldsymbol{\theta}) = f_{\boldsymbol{\varepsilon}|\boldsymbol{\theta}}(\hat{\mathbf{y}} - \mathbf{y}(\boldsymbol{\theta}) \mid \boldsymbol{\theta})$$

$$\hat{\mathbf{y}} = \mathbf{y}(\boldsymbol{\theta}) + \boldsymbol{\varepsilon}$$

$\hat{\mathbf{y}}$ Estimated optical coefficients (retrieval results from lidar signal processing)

$\boldsymbol{\varepsilon}$ Estimation error

How can uncertainty in the high dimension aerosol model be understood to assess uncertainty of aerosol properties?

Stochastic Monte Carlo

Sequence of Random Variables Θ_n *with PDFs* f_n

Markov Chain

$$f_{n|n-1,n-2,\cdots}(\theta_n \mid \theta_{n-1}, \theta_{n-2}, \ldots) = f_{n|n-1}(\theta_n \mid \theta_{n-1})$$

$$f_{n+1|n} = f_{n|n-1} = K$$

$$f_{n+1}(\phi) = \int K(\phi,\theta) f_n(\theta) d\theta$$

Stationary distributions $f_\infty(\phi) = \int K(\phi,\theta) f_\infty(\theta) d\theta$

Monte Carlo simulation

Begin with Θ_0 *and generate subsequent* Θ *according to some* K

Metropolis-Hastings Algorithm

Designed so that a target distribution $\pi(\theta)$ *is stationary*

1) Generate candidate Φ given current outcome Θ from a random number generator congruent with a proposal distribution whose PDF is $q(\varphi|\theta)$

2) Accept Φ as the next outcome with probability, $\alpha(\Phi,\Theta) = \min\left\{1, \dfrac{\pi(\Phi)q(\Theta\mid\Phi)}{\pi(\Theta)q(\Phi\mid\Theta)}\right\}$

3) If Φ is not accepted then Θ is retained as the next outcome

Initial distributions $f_0(\theta)$ *will converge to* $\pi(\theta)$ *with only lose requirements of the proposal distribution. Its choice affects the rate of convergence but generally not the stationary distribution.*

Types of Proposal Distributions

- Random walk generators
- Independent generators

Implementation

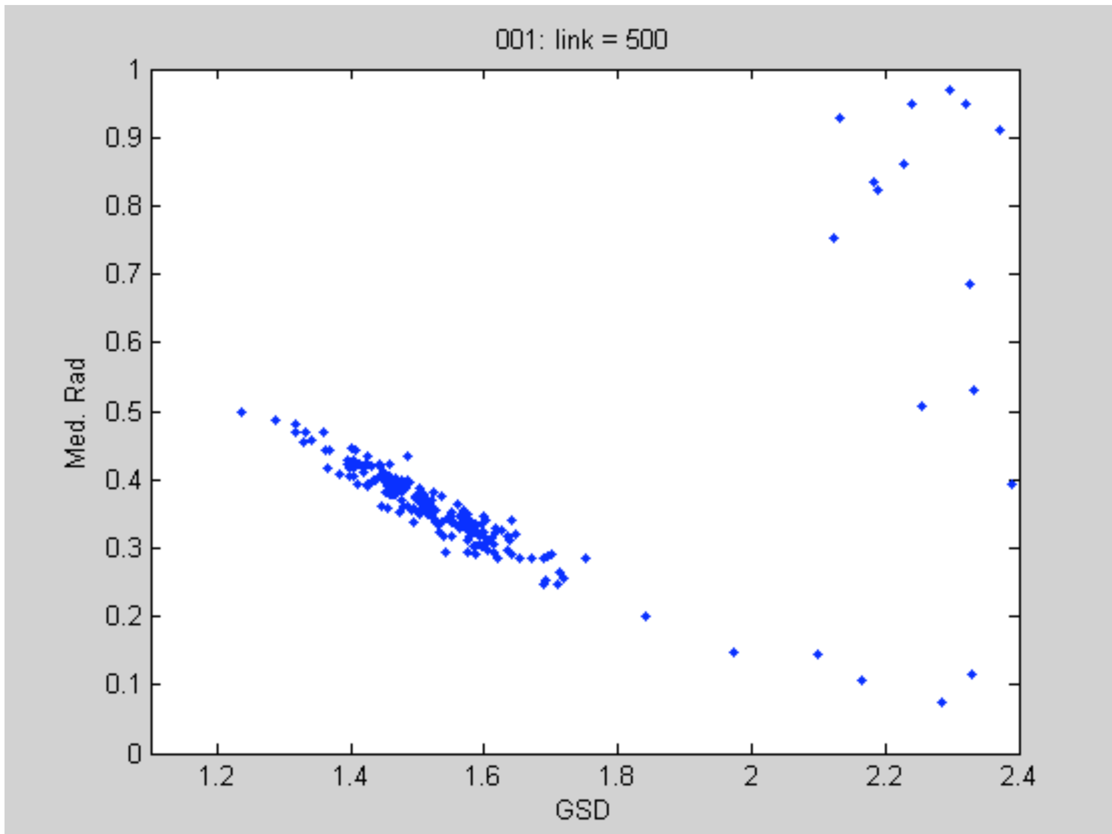
- *Ensemble of parallel chains starting from an initial distribution*
- *Random walk generation for first several links*
- *Ensemble is analyzed partway along the chain, clusters are determined and used to form a hybrid proposal distribution*

Hybrid proposal distribution

Randomly choose walk generation or generation from a distribution corresponding to one of the clusters, thereby allowing teleportation between regions. This aspect is essential for dealing with multiple local maxima in the posterior PDF.

Ensemble evolution animation

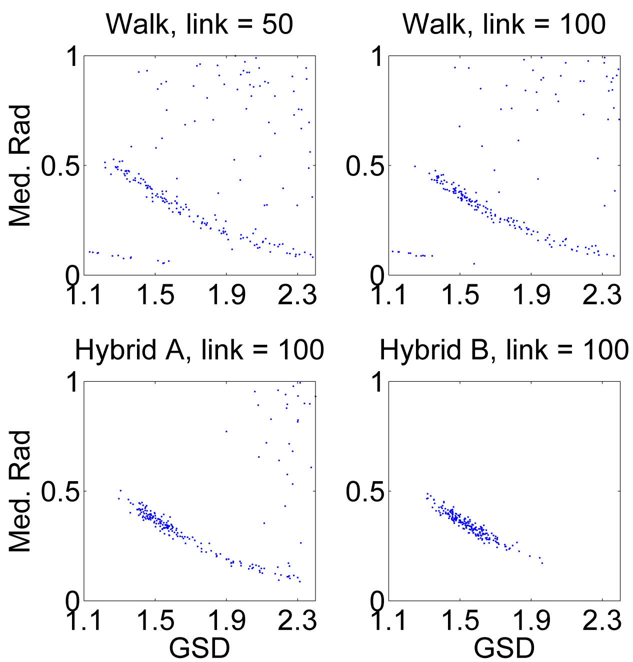
Example in a reduced single mode aerosol size distribution
The target distribution PDF is only a function of median radius and geometric standard deviation



Speeding convergence by initial resampling

Ensemble members are not getting represented by clusters

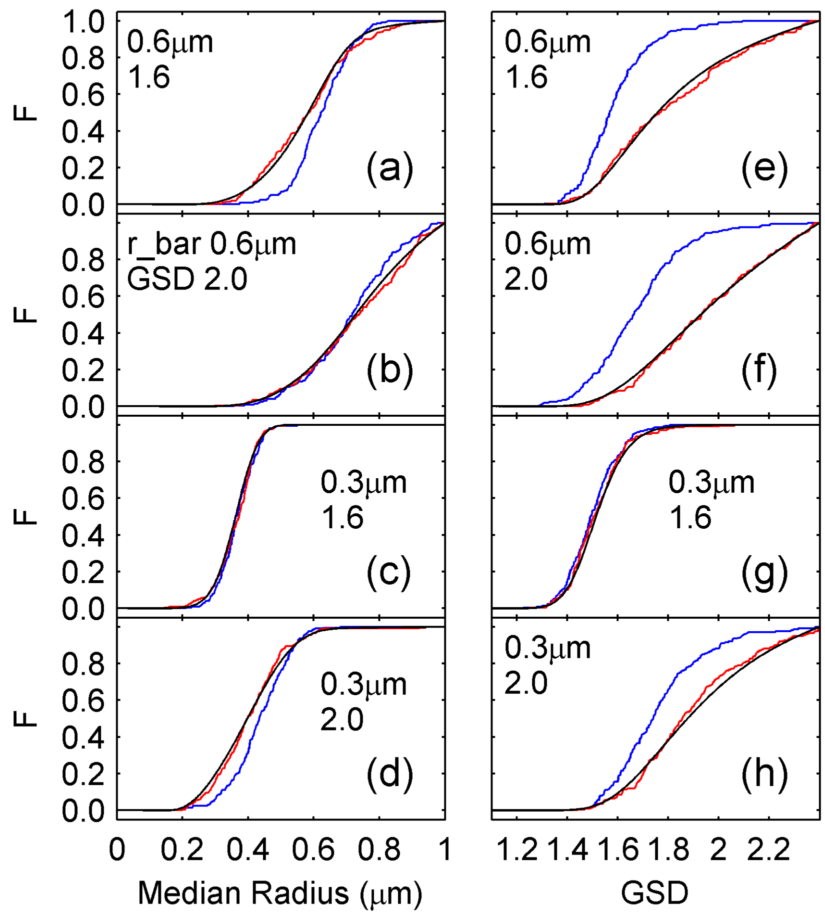
- *Results in tendency not to teleport*



Hybrid B also includes a small probability of sampling from the initial distribution

$$\alpha = \min\left\{ 1, \quad \frac{\pi(\Phi)\left(p_w q_w(\Theta \mid \Phi) + \sum_{i=1}^{N_c} p_i q_i(\Theta) \right)}{\pi(\Theta)\left(p_w q_w(\Phi \mid \Theta) + \sum_{i=1}^{N_c} p_i q_i(\Phi) \right)} \right\}$$

Comparison with forward Monte Carlo analysis



Numerically integrated posterior cumulative distribution functions (CDFs) (Black) and CDFs estimated from the simple forward Monte Carlo method applied using the *maximum a posteriori* inverse (Blue) and the Ensemble Metropolis-Hastings algorithm (Red).

Extension to bi-modal particle size distributions

Speeding convergence by fitting proposal PDF to the target PDF

Multivariate Gaussian PDF

$$\ln(g(\boldsymbol{\theta})) = -\tfrac{1}{2}(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}})^T \mathbf{S}^{-1}(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}) - c$$

$\mathbf{S}$: Covariance matrix

$\bar{\boldsymbol{\theta}}$: Mean vector

$$\bar{\boldsymbol{\theta}} = \boldsymbol{\theta} + \mathbf{S}\nabla_{\boldsymbol{\theta}} \ln(g(\boldsymbol{\theta}))$$

First derivative fitting

$$\left(\mathbf{S}^{-1}\right)_{ij} = -\partial_i \partial_j \ln(g)$$

Second derivative fitting

$$\mathbf{S} = \mathbf{V}\begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_N \end{bmatrix}\mathbf{V}^T$$

Eigenvalue conditioning

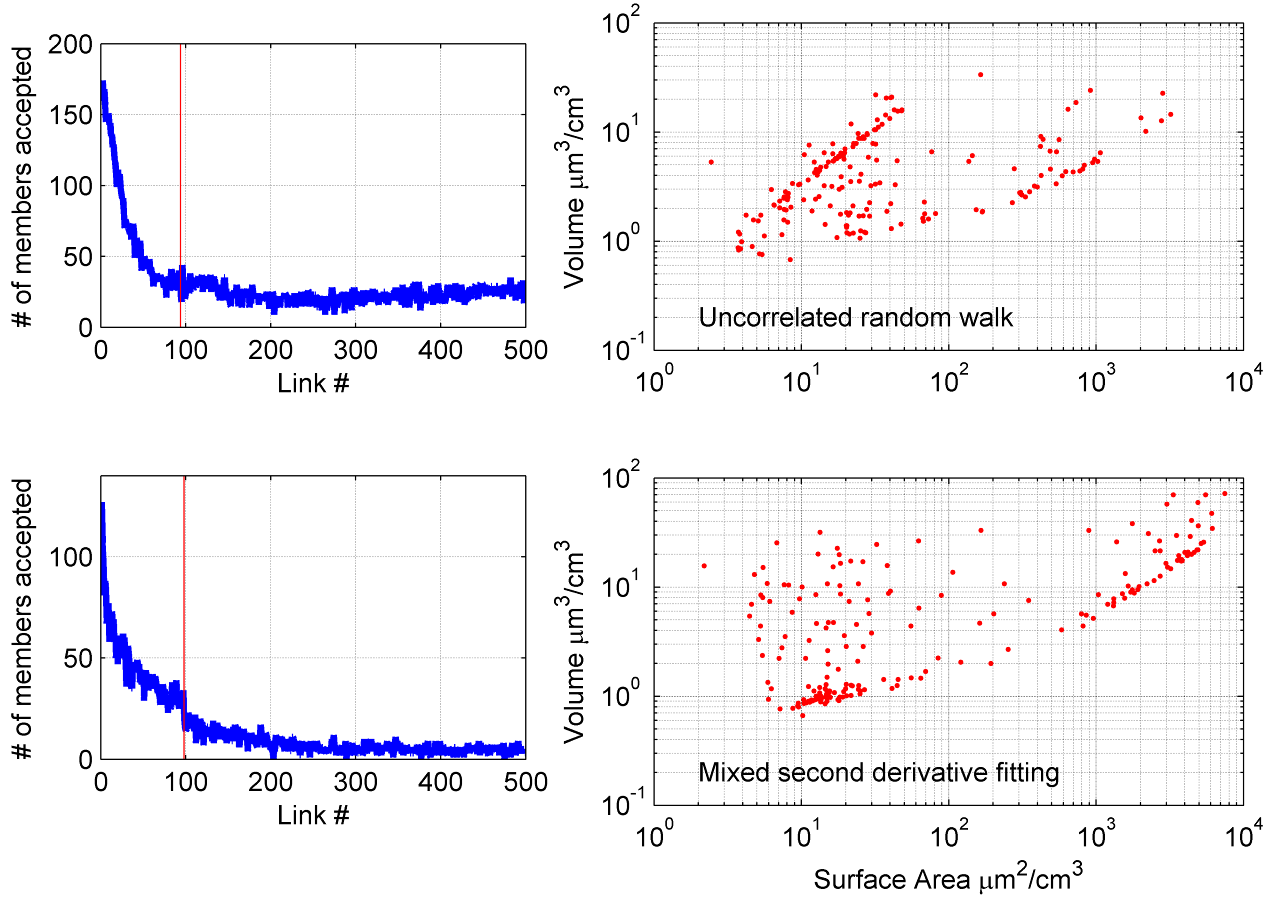
S must be positive definite to be a valid covariance matrix. This is not guaranteed if the target distribution is not Gaussian, and so it is remedied through eigenvector decomposition. Negative Eigenvalues are set to an upper limit and **S** is recomposed.

Implementations

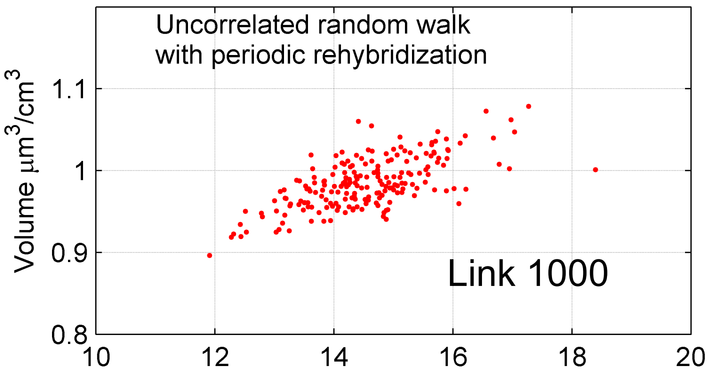
Acceptance rate is used for determining at which link to cluster the ensemble to form the hybrid proposal distribution

Full second order derivative fitting compared with uncorrelated partially fitted random walk generation

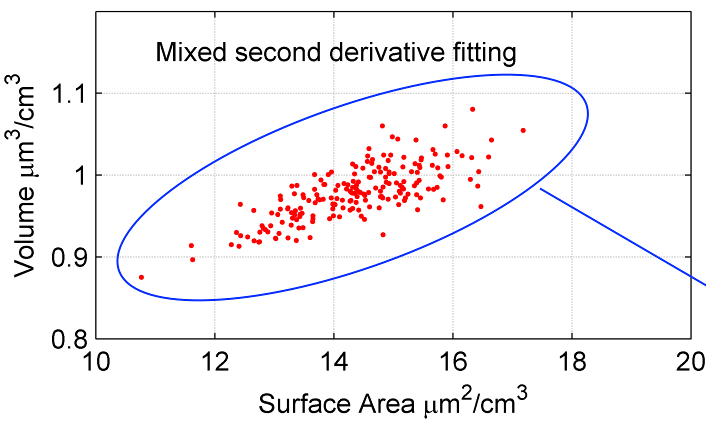
Higher acceptance probability with uncorrelated random walk is due to scaling down the fitted variance



Periodic re-hybridization to speed convergence

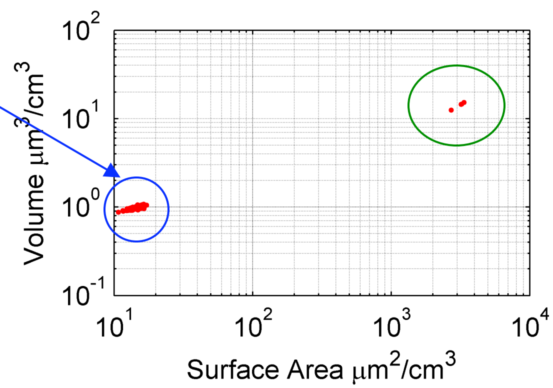


All ensemble members
had been together
since ~link 500



Hybridization only once

After 10000 links there still remain some
members around a local maximum



Things to think about

- How to choose when to form the multi-component independent generator from clusters
- Can the independent generator be periodically reformulated indefinitely while still maintaining convergence to the target distribution
- Which clustering method works best
- Is there an optimal compromise between random walk with uncorrelated variance versus complete fitting of all mixed derivatives