

# An Introduction to Data Assimilation

Radu Herbei

STAT 8750.06 SSES Research Group

10/26/2015

# What is data assimilation ?

Talagrand (1997)

“...the process through which all the available information is used to estimate as accurately as possible the state of the atmospheric or oceanic flow”.

Bennett (2002)

... interpolating fields at one time, for subsequent use as initial data in a model integration which may even be a genuine forecast”.

Kalnay (2003) :

“... statistical combination of observations and short-range forecasts”.

Definition of DA is “work in progress” (Wikle & Berliner , 2007)

## Simple univariate example

$X$  - unobservable, quantity of interest,  $\sim$  prior  $N(\mu, \sigma^2)$

$Y = (Y_1, \dots, Y_n) = \text{data}$      $Y_i | X=x \sim \text{iid } N(x, \sigma^2)$

posterior  $X | Y \sim N(\text{mean}, \text{variance})$

$$E(X|Y) = \mu + \frac{n\sigma^2}{\sigma^2 + n\sigma^2} (\bar{Y}_n - \mu) = \mu + K (\bar{Y}_n - \mu)$$

$$K = \frac{n\sigma^2}{\sigma^2 + n\sigma^2}$$

$$\text{Var}(X|Y) = (1-K)\sigma^2$$

## Multivariate N-N case

$X \sim N_n(\mu, P)$  where  $\mu, P$  are known

$X = (n \times 1)$  vector

$Y | X=x \sim N_p(Hx, R)$

$Y = (p \times 1)$  vector

$H = (p \times n)$  observation matrix

$R$  = error covariance matrix

posterior  $X | Y \sim N_n(\cdot, \cdot)$

$$\begin{cases} E(X | Y) = \mu + K \underbrace{(Y - H\mu)}_{\text{red}} \\ \text{Var}(X | Y) = (I - K H) P \end{cases}$$

$$K = PH'(R + HPH')^{-1} = \text{gain matrix}$$

## Multivariate N-N case

posterior  $X | Y \sim N_n(\cdot, \cdot)$

$$\begin{cases} E(X | Y) = \mu + K(Y - H\mu) \\ \text{Var}(X | Y) = (I - KH)P \end{cases} \quad K = P H' (R + HPH')^{-1} = \text{gain matrix}$$

$\mu$  = forecast from some deterministic model

$P$  = forecast error

# Kriging

$$X = [ X(s_1), X(s_2), X(s_3) ]^T \quad Y = [ Y(s_2), Y(s_3) ]^T \quad H = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$P = \text{cov}(X) = \begin{bmatrix} C_{11} & C_{12} & C_{13} \\ C_{21} & C_{22} & C_{23} \\ C_{31} & C_{32} & C_{33} \end{bmatrix} \quad R = P^T I$$

$$K = P H^T (R + H P H^T)^{-1} = \begin{bmatrix} C_{12} & C_{12} \\ C_{22} & C_{23} \\ C_{32} & C_{33} \end{bmatrix} \times \left[ R + \begin{pmatrix} C_{22} & C_{23} \\ C_{32} & C_{33} \end{pmatrix} \right]^{-1}$$

## Kriging

Predict  $x(s_i)$  :  $X(s_i) | Y \sim N(\cdot, \cdot)$

$$E(X(s_i) | Y) = \underline{\mu(s_i)} + w_{i2} \underline{(Y(s_2) - \mu(s_2))} + w_{i3} \underline{(Y(s_3) - \mu(s_3))} \quad \text{BLUP}$$

the weights  $w_{i2}, w_{i3}$  can be found from the gain matrix

$$\text{Var}(X(s_i) | Y) = \dots$$

- simple kriging (spatial stats) optimal interpolation (atm/ocean sci.)
- ordinary kriging (constant unknown prior mean)
- etc.

## Variational approach

- One can consider the same problem from a variational (optimization) view

$$\text{minimize } J(x) = (y - Hx)' R^{-1} (y - Hx) + (x - \mu)' P^{-1} (x - \mu)$$

- this will find the mode (mean) of the Bayesian N-N model
- Advantages ? Disadvantages ?

## Sequential methods

$$X_{0:T} = [x_0, \dots, x_T]$$

↳ process

$$Y_{0:T} = [y_0, \dots, y_T]$$

↳ data

posterior :  $p(X_{0:T} | Y_{0:T}) \propto p(X_{0:T}) \cdot p(Y_{0:T} | X_{0:T})$

prior :  $p(X_{0:T}) = p(x_0) \cdot \prod_{t=1}^T p(x_t | x_{t-1})$  (Markov)

transition density

lik :  $p(Y_{0:T} | X_{0:T}) = \prod_{t=0}^T p(y_t | x_t)$

data distribution

## Filtering

assume that  $p(x_{t-1} | y_{1:t-1})$  is available (posterior distr.  $| y_{1:t-1}$ )

goal: find the forecast distribution  $p(x_t | y_{1:t-1})$

$$\begin{aligned} p(x_t | y_{1:t-1}) &= \int p(x_{t-1}, x_t | y_{1:t-1}) dx_{t-1} \\ &= \underbrace{\int p(x_t | x_{t-1})}_{\text{forecast}} \underbrace{p(x_{t-1} | y_{1:t-1})}_{\text{posterior}} dx_{t-1} \end{aligned}$$

## Filtering

analysis step : incorporate obs  $y_t$

$$\underline{p(x_t | y_{1:t})} = p(x_t | y_{1:t-1}, y_t)$$

$$\propto \underline{p(x_t | y_{1:t-1})} \underline{p(y_t | x_t, y_{1:t-1})}$$

$$= \underline{p(x_t | y_{1:t-1})} \cdot \underline{p(y_t | x_t)}$$

iterate for  $t = 1, 2, \dots, T$

## Smoothing

goal: find  $p(x_t | y_{1:T})$   $t=0, 1, 2, \dots, T$

ALL the data

when  $t=T$ , this is the filtering distribution  $p(x_T | y_{1:T})$

in general,

$$p(x_t | y_{1:T}) = \int p(x_t, x_{t+1} | y_{1:T}) dx_{t+1}$$

$$= \int p(x_t | x_{t+1}, y_{1:T}) p(x_{t+1} | y_{1:T}) dx_{t+1}$$

## Smoothing

$$p(x_t | x_{t+1}, y_{1:T}) = p(x_t | x_{t+1}, \underline{y_{1:t}})$$

$$\propto \underbrace{p(x_{t+1} | x_t, y_{1:t})}_{\downarrow} p(x_t | \underline{y_{1:t}}) \\ = p(x_{t+1} | x_t) \underbrace{p(x_t | \underline{y_{1:t}})}_{\text{filter distr.}}$$

forward filtering:  $p(x_1 | y_1), p(x_2 | y_1), p(x_2 | y_{1:2}), p(x_3 | y_{1:2}), p(x_3 | y_{1:3}), \dots$

backward smoothing:  $p(x_T | y_{1:T}), p(x_{T-1} | y_{1:T}), \dots$

## Kalman filter / Kalman smoother

- assume everything is Gaussian

data distribution :  $\underline{Y_t = H_t X_t + \varepsilon_t}$   $\varepsilon_t \sim N(0, R_t)$

process distribution :  $\underline{X_t = M_t X_{t-1} + \eta_t}$   $\eta_t \sim N(0, Q_t)$

$H_t$  = observation operator (linear)

$M_t$  = model operator (linear)

usually  $\varepsilon \perp \eta$

## Kalman filter

$$y_t = H_t x_t + \varepsilon_t \quad \varepsilon_t \sim N(0, R_t)$$

$$\underline{x_t = M_t x_{t-1} + m_t \quad m_t \sim N(0, Q_t)}$$

Filtering:  $p(x_t | y_{1:t-1}) \quad p(x_t | y_{1:t})$  (both Gaussian)

mean:  $x_{t|t-1} = E(x_t | y_{1:t-1})$

$$= E(E(x_t | x_{t-1}) | y_{1:t-1})$$

$$= E(M_t x_{t-1} | y_{1:t-1})$$

$$= M_t x_{t-1|t-1}$$

## Kalman filter

var:

$$P_{t|t-1} = \text{Var}(X_t | Y_{1:t-1})$$

$$= E(\text{Var}(X_t | X_{t-1}) | Y_{1:t-1}) + \text{Var}(E(X_t | X_{t-1}) | Y_{1:t-1})$$

$$= E(Q_t | Y_{1:t-1}) + \text{Var}(M_t X_{t-1} | Y_{1:t-1})$$

$$= Q_t + M_t P_{t-1|t-1} M_t'$$

## Kalman filter

analysis step  $p(x_t | y_{1:t})$

$$x_{t|t} = E(x_t | y_{1:t}) = x_{t|t-1} + k_t (y_t - H_t x_{t|t-1})$$

$$P_{t|t} = \text{Var}(x_t | y_{1:t}) = (I - k_t H_t) P_{t|t-1}$$

$$\begin{aligned} k_t &= P_{t|t-1} H_t' (H_t' P_{t|t-1} H_t + R_t)^{-1} \\ &= \text{Kalman gain} \end{aligned}$$

## Kalman Smoother

- the Kalman filter gives access to

$$\text{the analysis (filter) distr } p(x_t | y_{1:t})$$

$$\text{the forecast distr } p(x_{t+1} | y_{1:t})$$

$t = 1, 2, \dots, T$

- goal : find  $p(x_t | y_{1:T})$  (all the data)

Kalman smoother :  $p(x_t | y_{1:T}) = \text{Gaussian}$

$$x_{t|T} = E(x_t | y_{1:T})$$

$$P_{t|T} = \text{Var}(x_t | y_{1:T})$$

## Kalman Smoother

$$x_{t|T} = E(x_t | Y_{1:T}) \quad P_{t|T} = \text{Var}(x_t | Y_{1:T})$$

$$x_{t|T} = E(x_t | Y_{1:T}) = E(E(x_t | X_{t+1}, Y_{1:t}) | Y_{1:T})$$

$$= x_{t|t} + \gamma_t (x_{t+1|T} - x_{t+1|t})$$

$$P_{t|T} = \text{Var}(x_t | Y_{1:T}) = \dots (\text{similarly})$$

$[x_t | X_{t+1}, Y_{1:t}] \sim \text{Gaussian}$

$$E(x_t | X_{t+1}, Y_{1:t}) = x_{t|t} + \gamma_t (x_{t+1} - M_{t+1} x_{t|t}) \quad \gamma_t = \dots$$

## Monte Carlo sampling

focus on  $p(X_{1:T} | Y_{1:T}) \propto p(X_{1:T}) p(Y_{1:T} | X_{1:T})$

$$\propto \prod p(x_t | x_{t-1}) \times \prod p(y_t | x_t)$$

Importance sampling MC

sample  $X_{1:T}^i \sim g(X_{1:T} | Y_{1:T}) \quad i=1, 2, \dots, M$

calculate weights  $w_i = \frac{p(x_{1:T}^i | y_{1:T}) / g(x_{1:T}^i | y_{1:T})}{\sum_{i=1}^M p(x_{1:T}^i | y_{1:T}) / g(x_{1:T}^i | y_{1:T})}$  ←

approximate  $p(X_{1:T} | Y_{1:T}) \approx \sum_{i=1}^M w_i \delta_{X_{1:T}^i}$

---

## Sequential MC

1.  $p(x_1 | y_1)$

sample  $x_i^i \quad i=1, \dots, M \sim p(x_i) \quad (\text{prior})$

weights

$$w^i = \frac{p(x_i^i | y_1) / p(x_i)}{\sum_i p(x_i^i | y_1) / p(x_i)} = \frac{p(y_1 | x_i^i)}{\sum_i p(y_1 | x_i)}$$

$$p(x_1 | y_1) \approx \sum_i w^i \delta_{x_i^i}$$



## Sequential MC

$$p(x_i | y_i) \approx \sum_i w^i \delta_{x_i}$$

2.  $p(x_{1:2} | y_{1:2}) \propto \underbrace{p(x_1 | y_1)}_{\text{sample } x_1^j} \underbrace{p(x_2 | x_1)}_{j=1, 2, \dots, M} \underbrace{p(y_2 | x_2)}$

sample  $x_1^j \quad j=1, 2, \dots, M \sim p(x_1 | y_1)$

$x_2^j \quad j=1, \dots, M \sim p(x_2 | x_1)$

have  $(x_{1:2}^j) \sim \underbrace{p(x_1 | y_1) p(x_2 | x_1)}$

(Re)Compute weights

$$w^j = \frac{p(y_2 | x_2^j)}{\sum_j p(y_2 | x_2^j)}$$

$$p(x_{1:2} | y_{1:2}) \approx \sum_j w^j \delta_{(x_{1:2})^j}$$

## Sequential MC

$$\begin{aligned} 3. \quad p(x_{1:3} | y_{1:3}) &\propto p(x_{1:2} | y_{1:2}) \cdot p(x_3, y_3 | x_{1:2}, y_{1:2}) \\ &\propto p(x_{1:2} | y_{1:2}) \quad p(x_3 | x_2) \quad p(y_3 | x_3) \end{aligned}$$

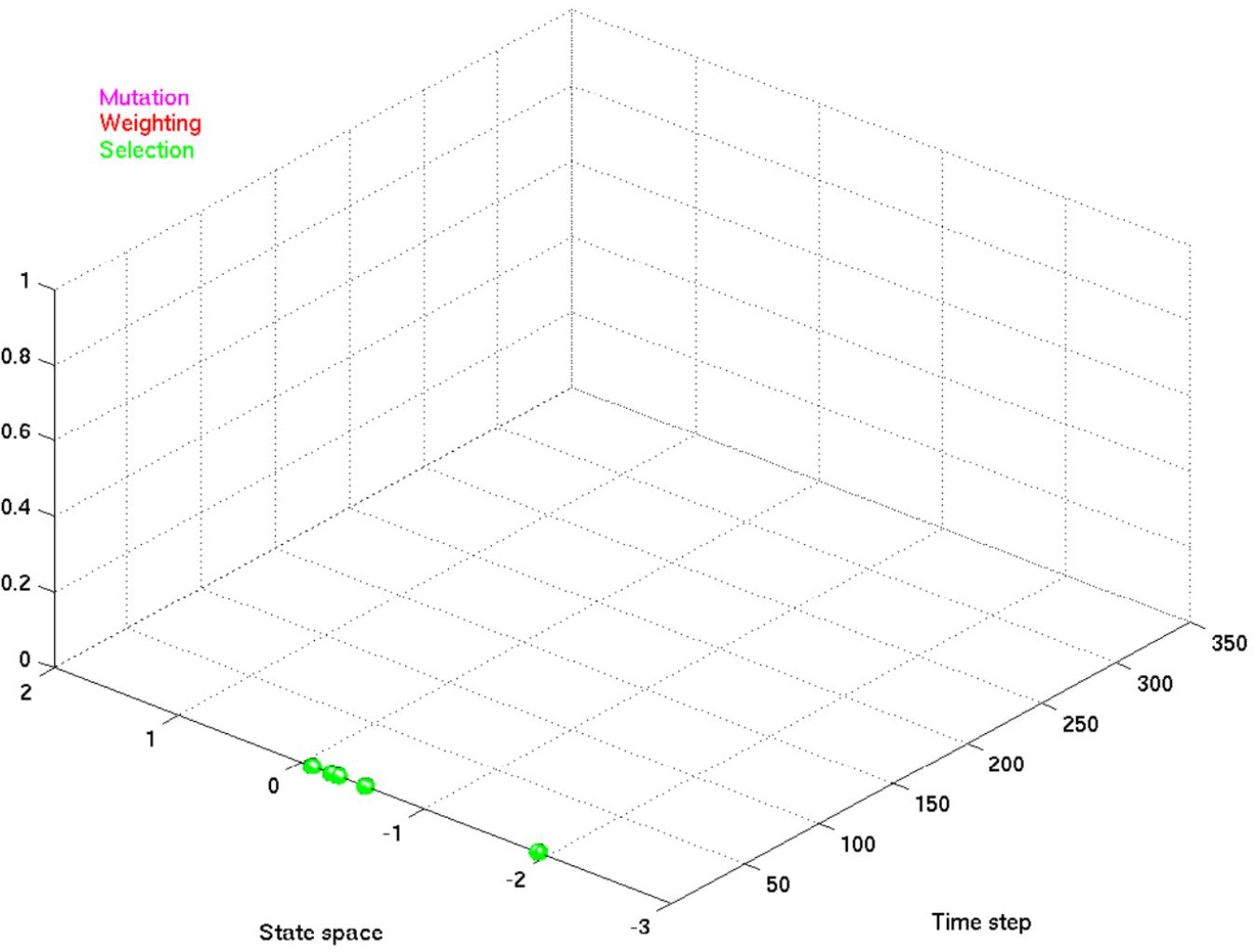
repeat the same procedure

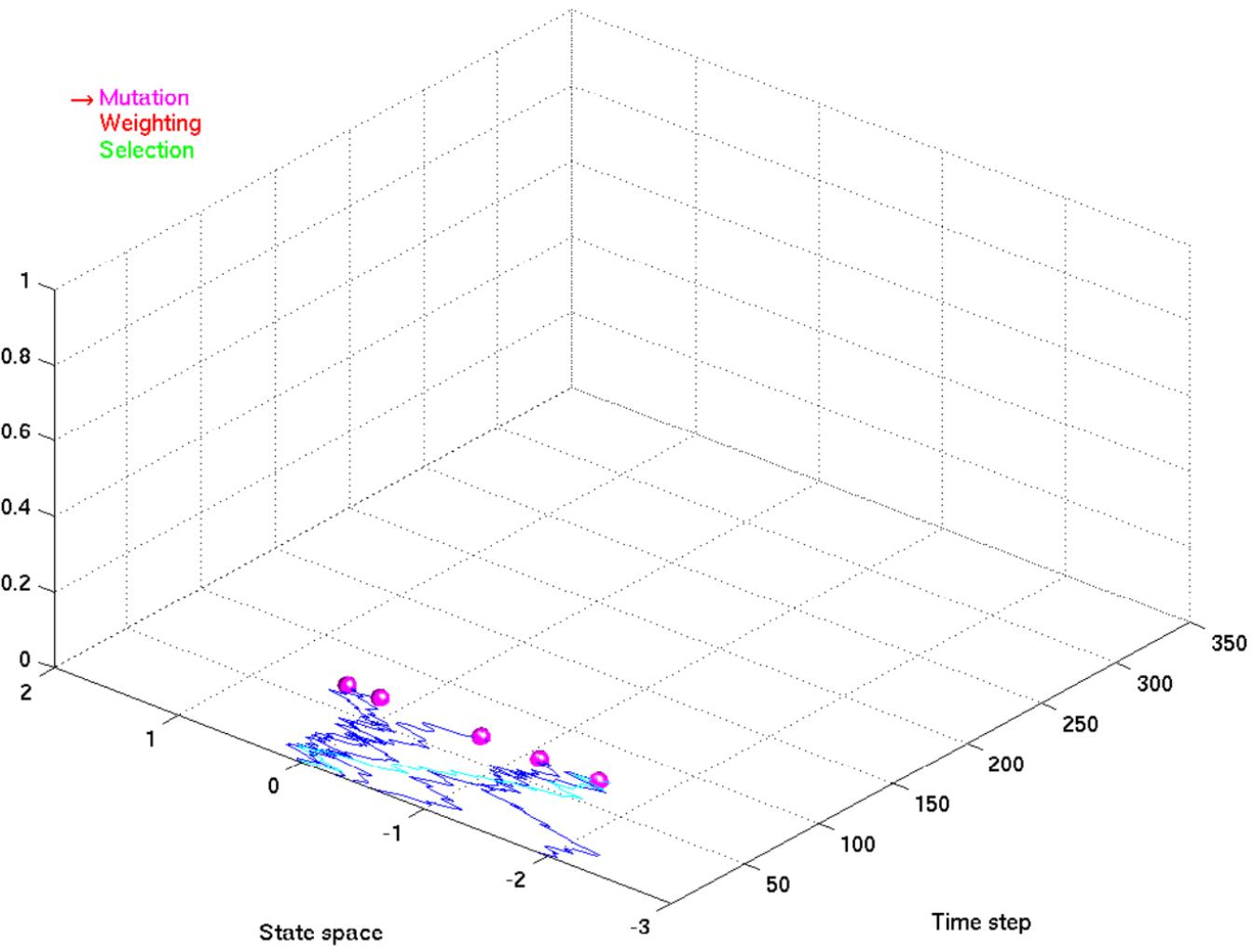
1. resampling from  $p(x_{1:2} | y_{1:2})$

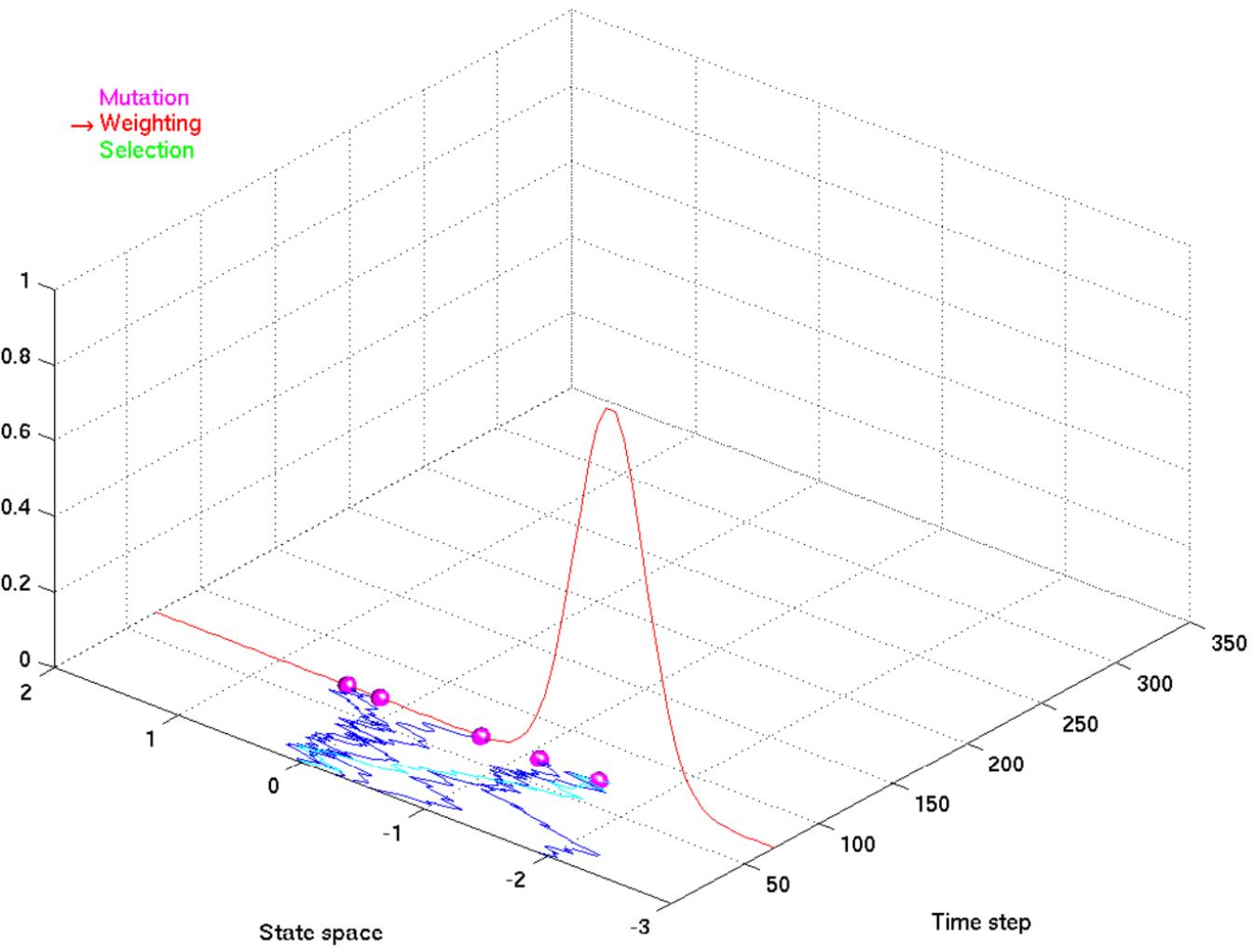
2. propagate using  $p(x_3 | x_2)$

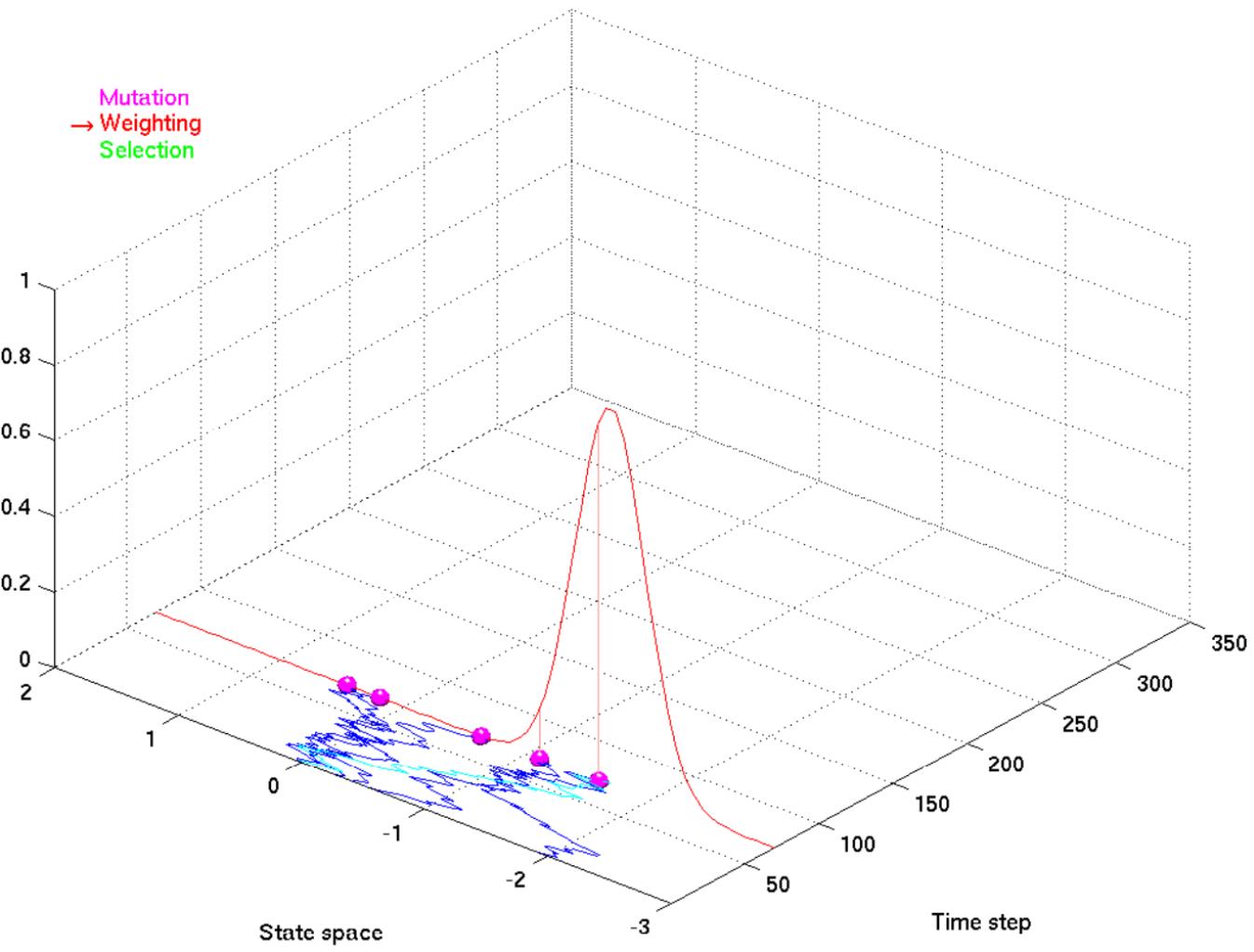
3. weight using  $p(y_3 | x_3)$

Mutation  
Weighting  
Selection

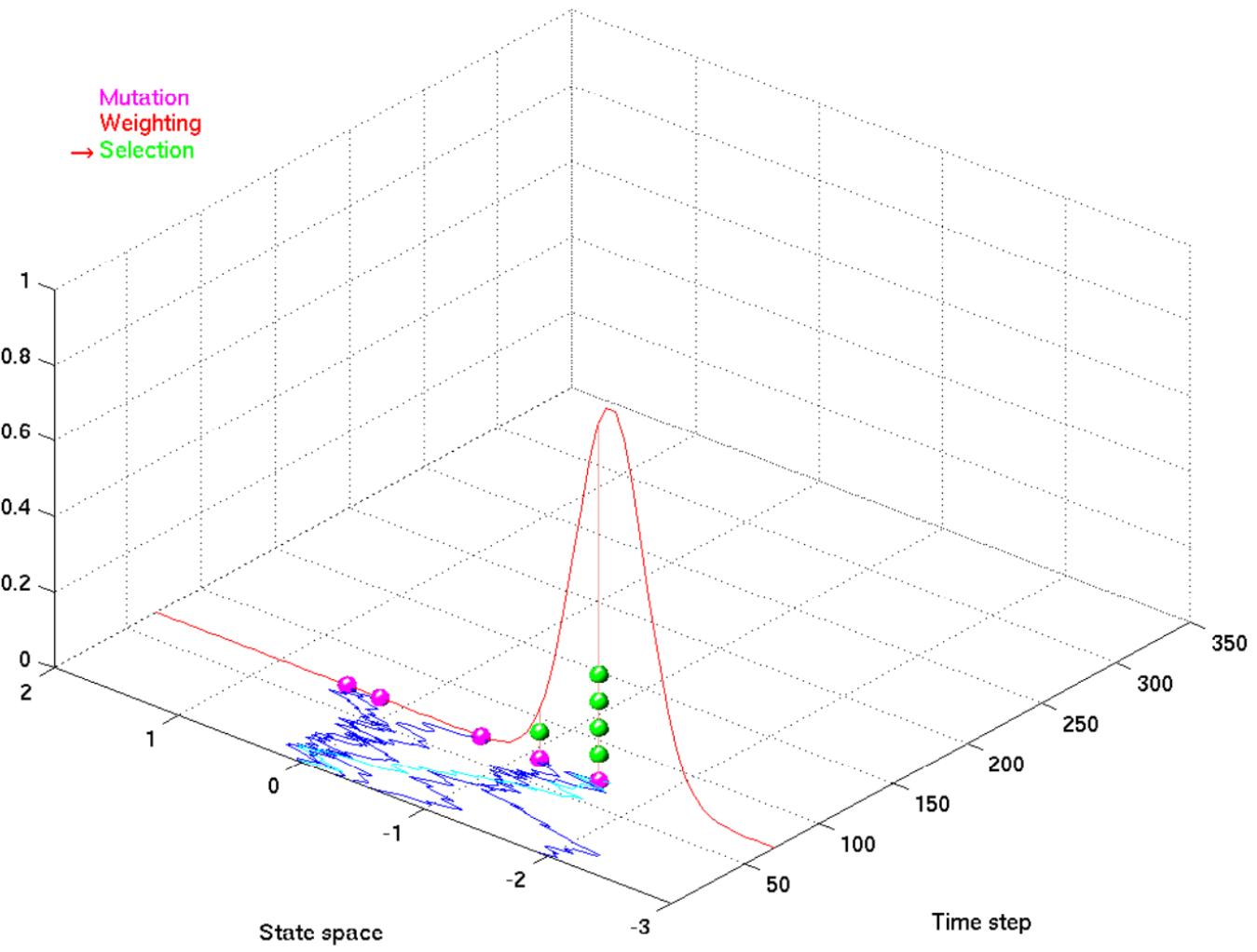




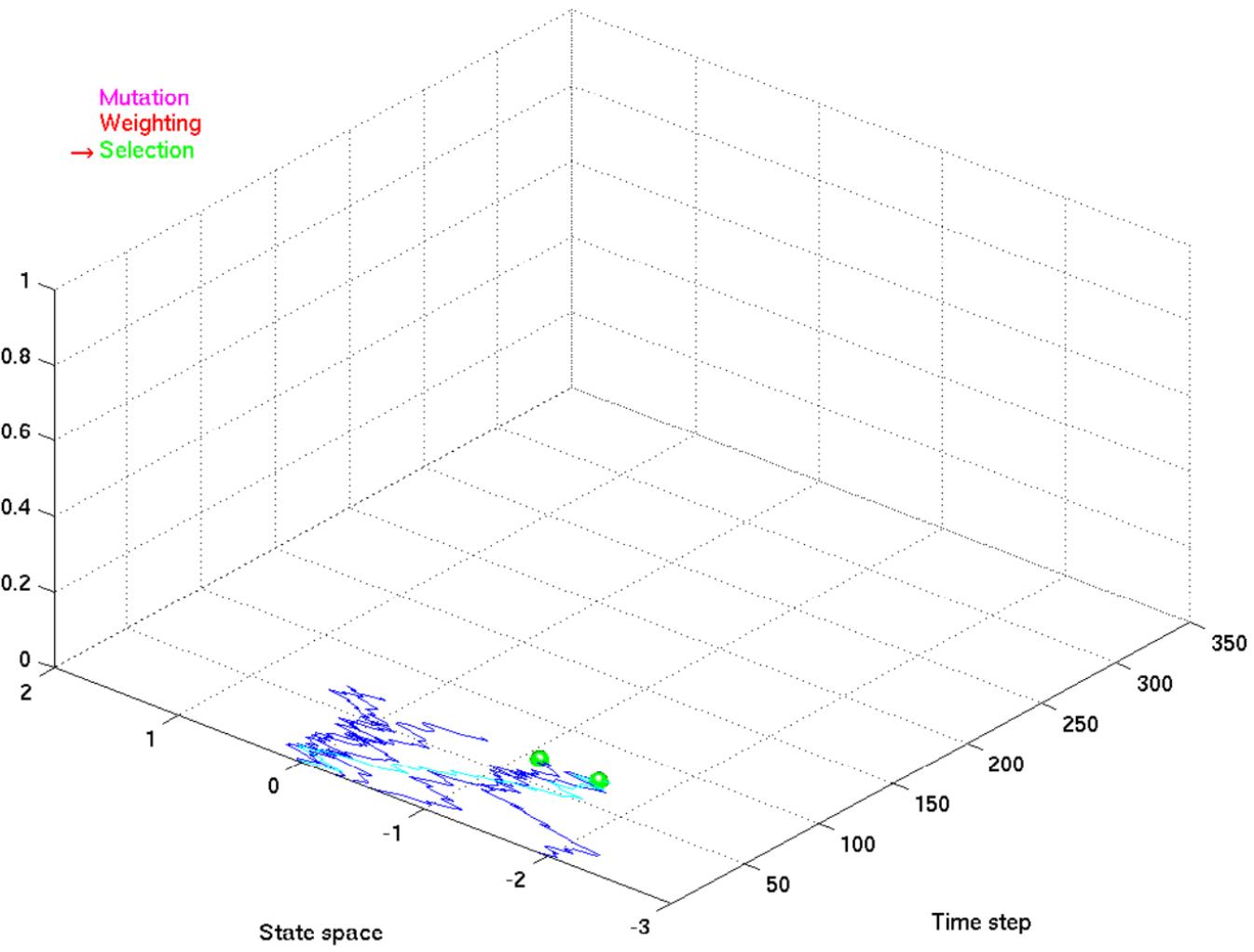


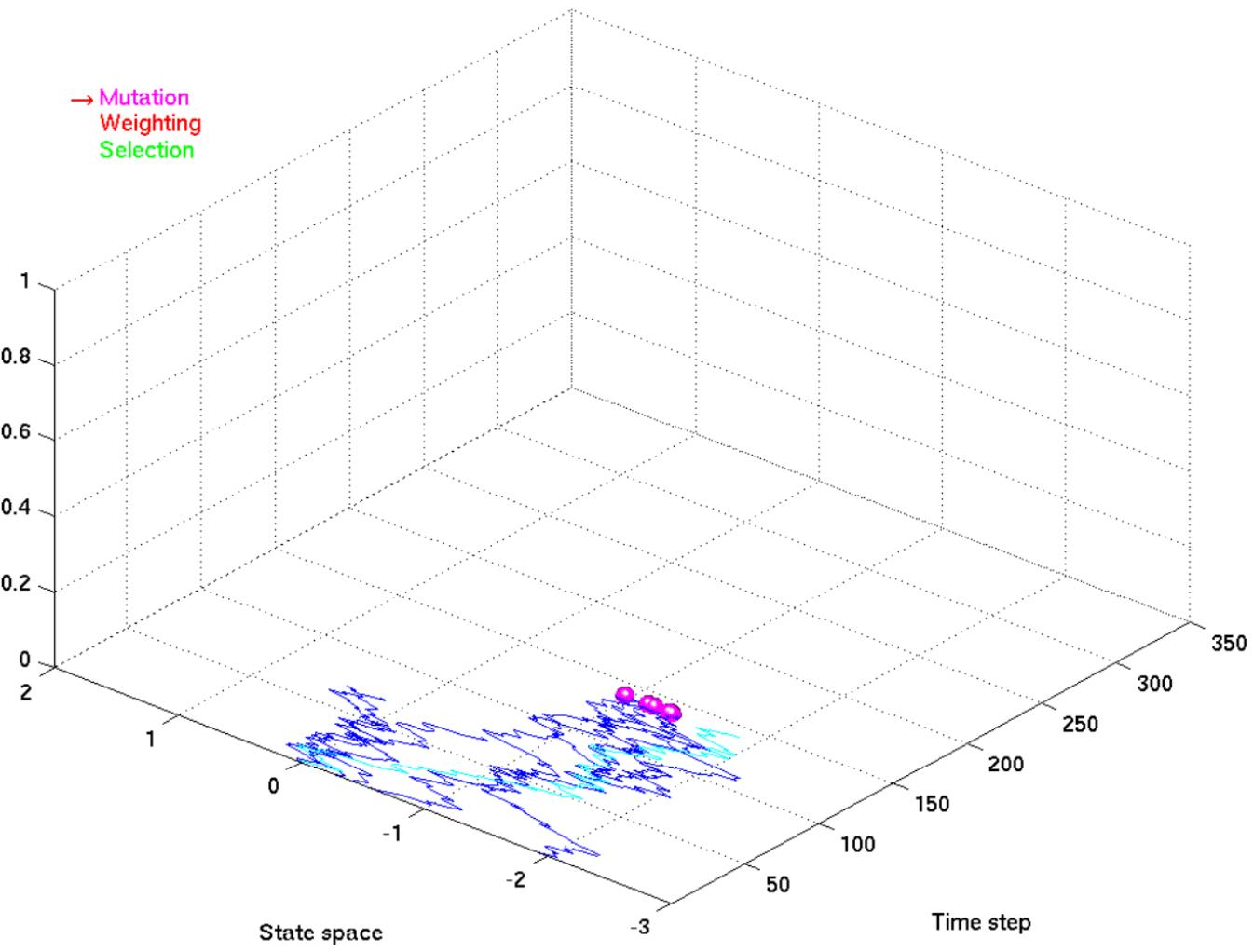


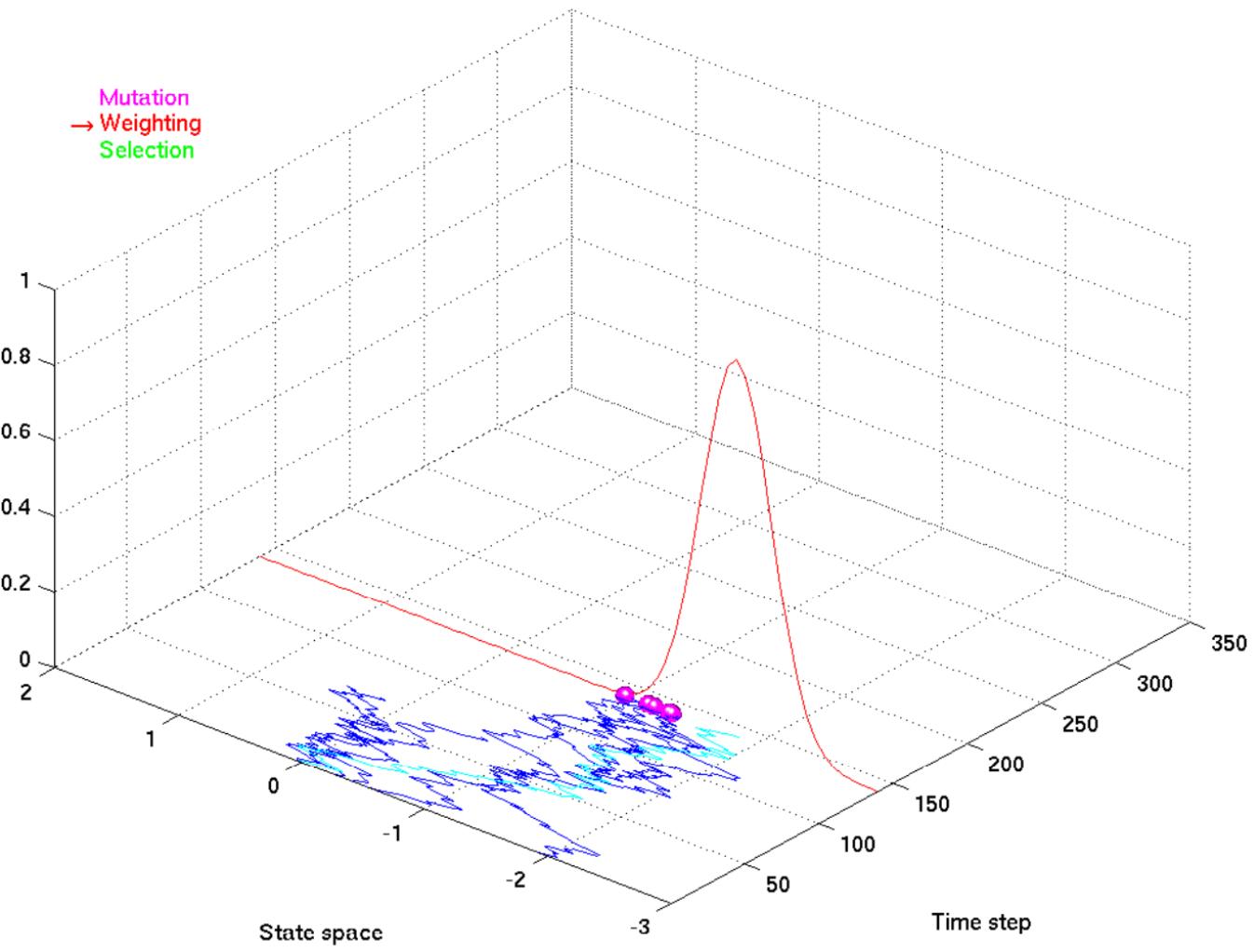
Mutation  
Weighting  
→ Selection

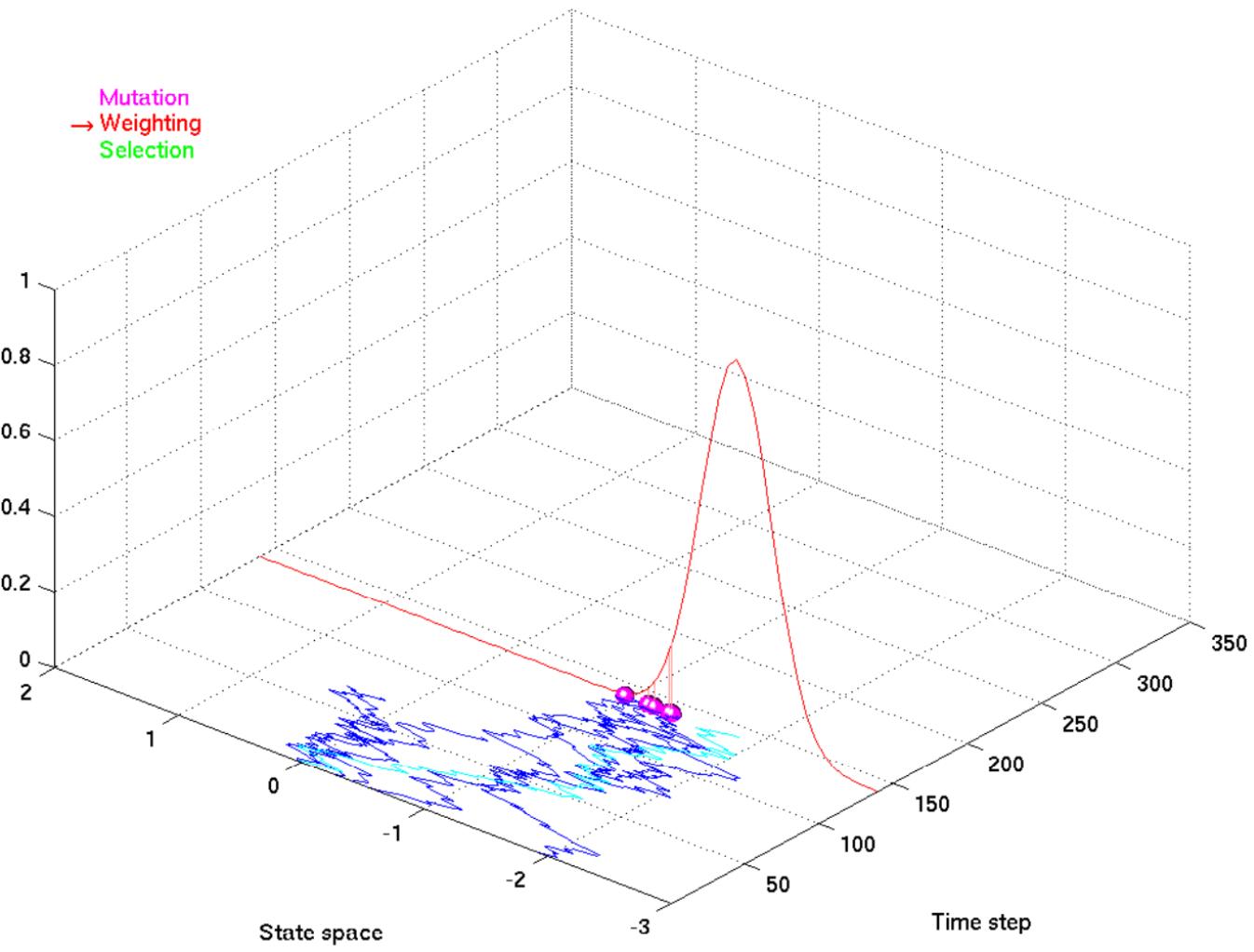


Mutation  
Weighting  
→ Selection

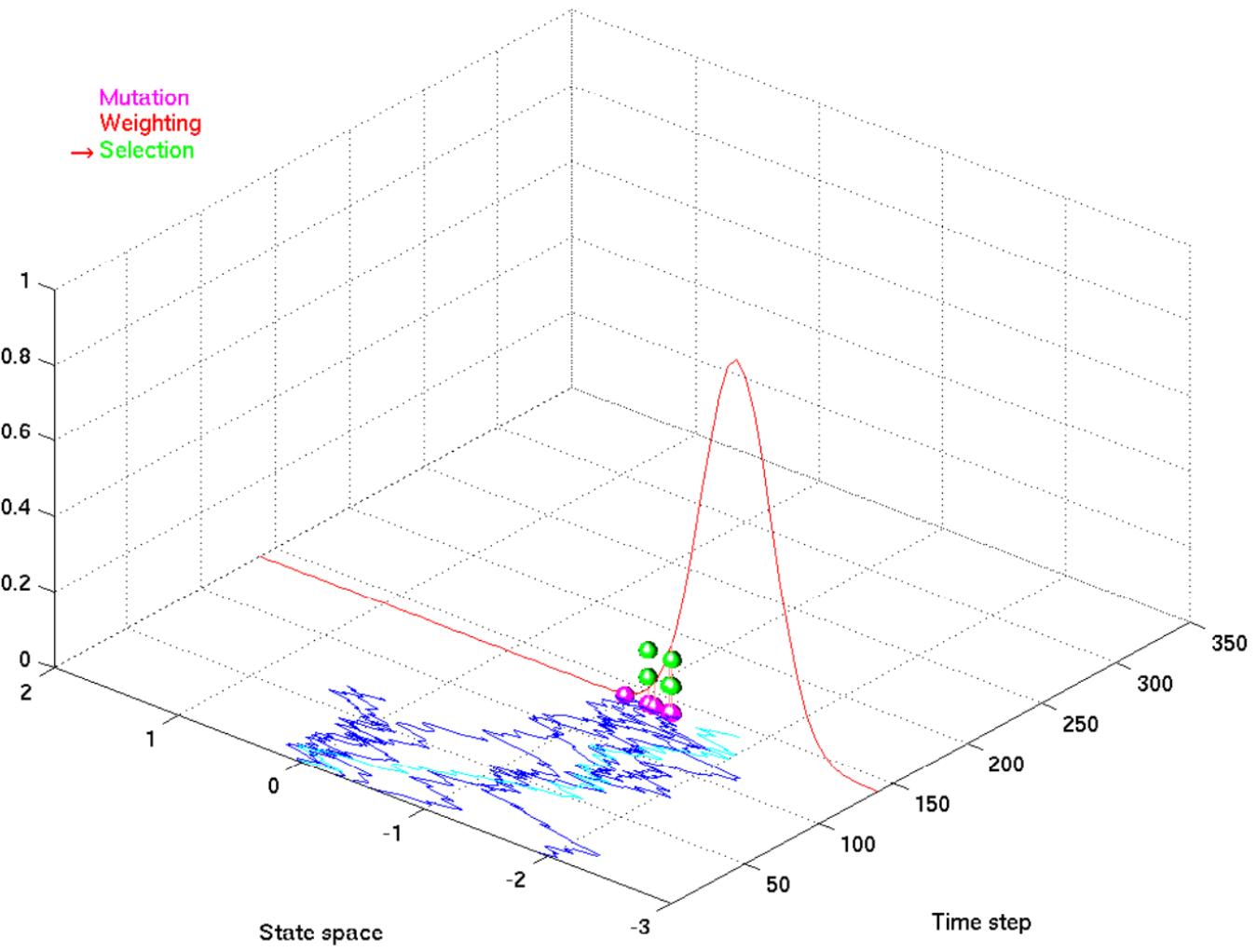




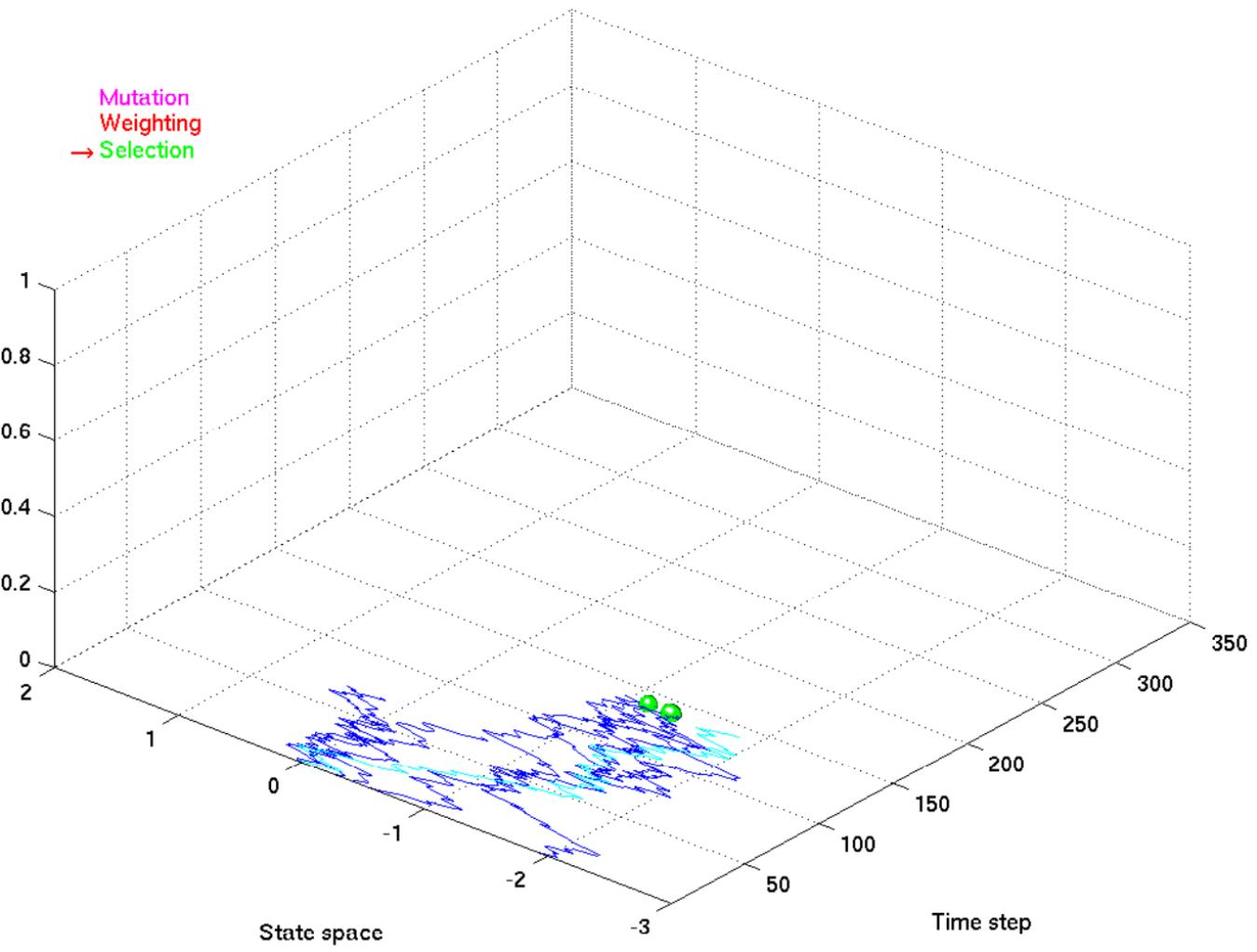




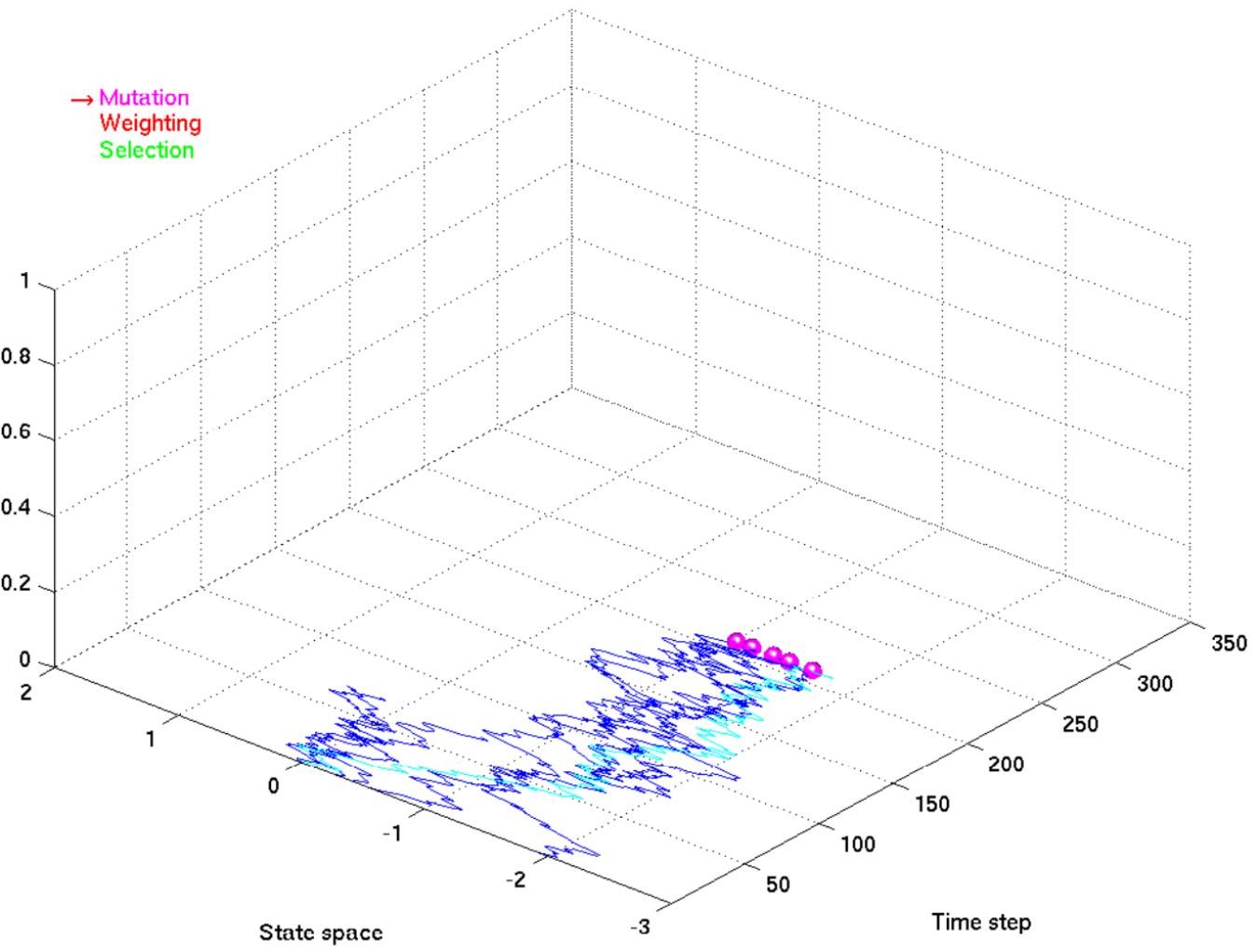
Mutation  
Weighting  
→ Selection

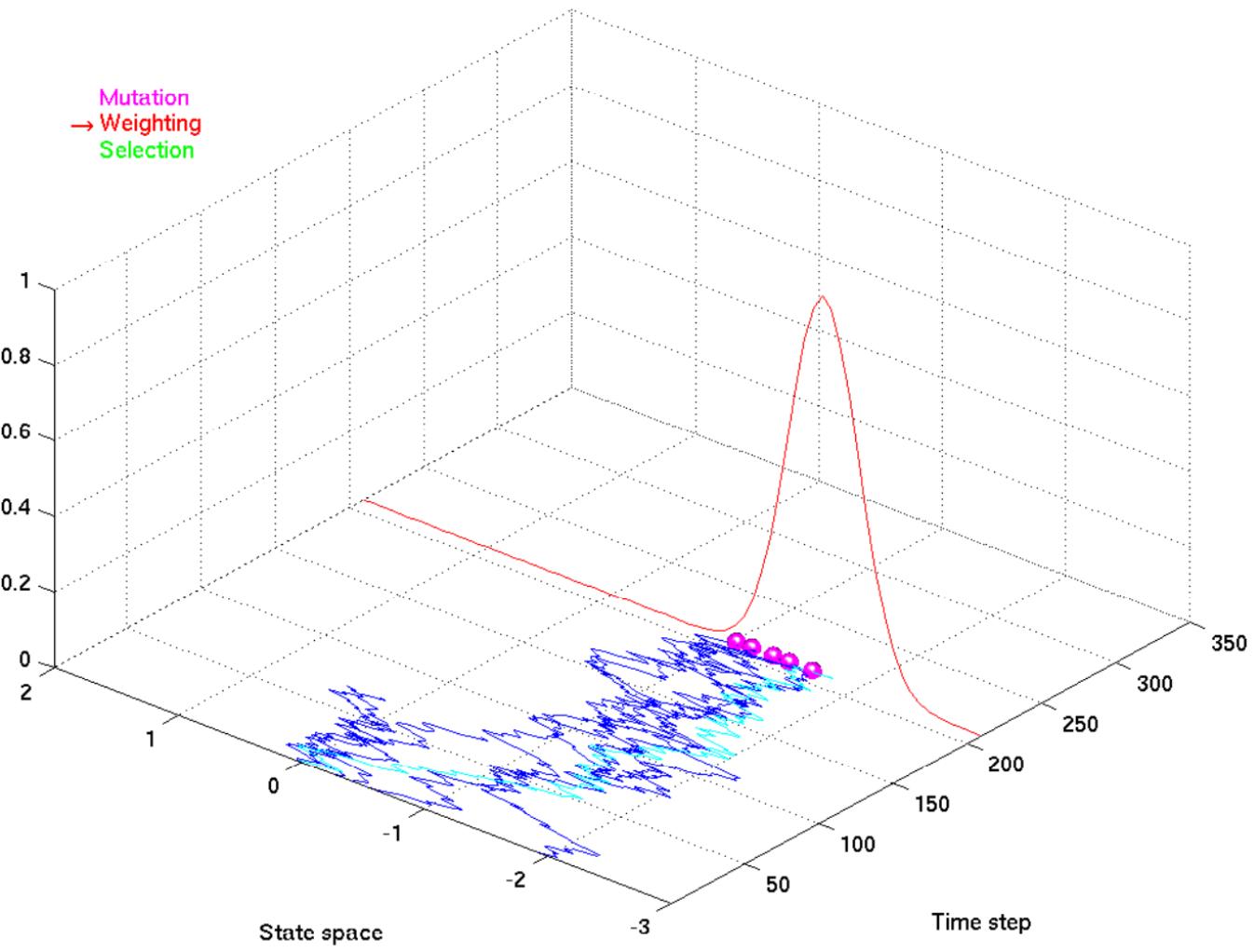


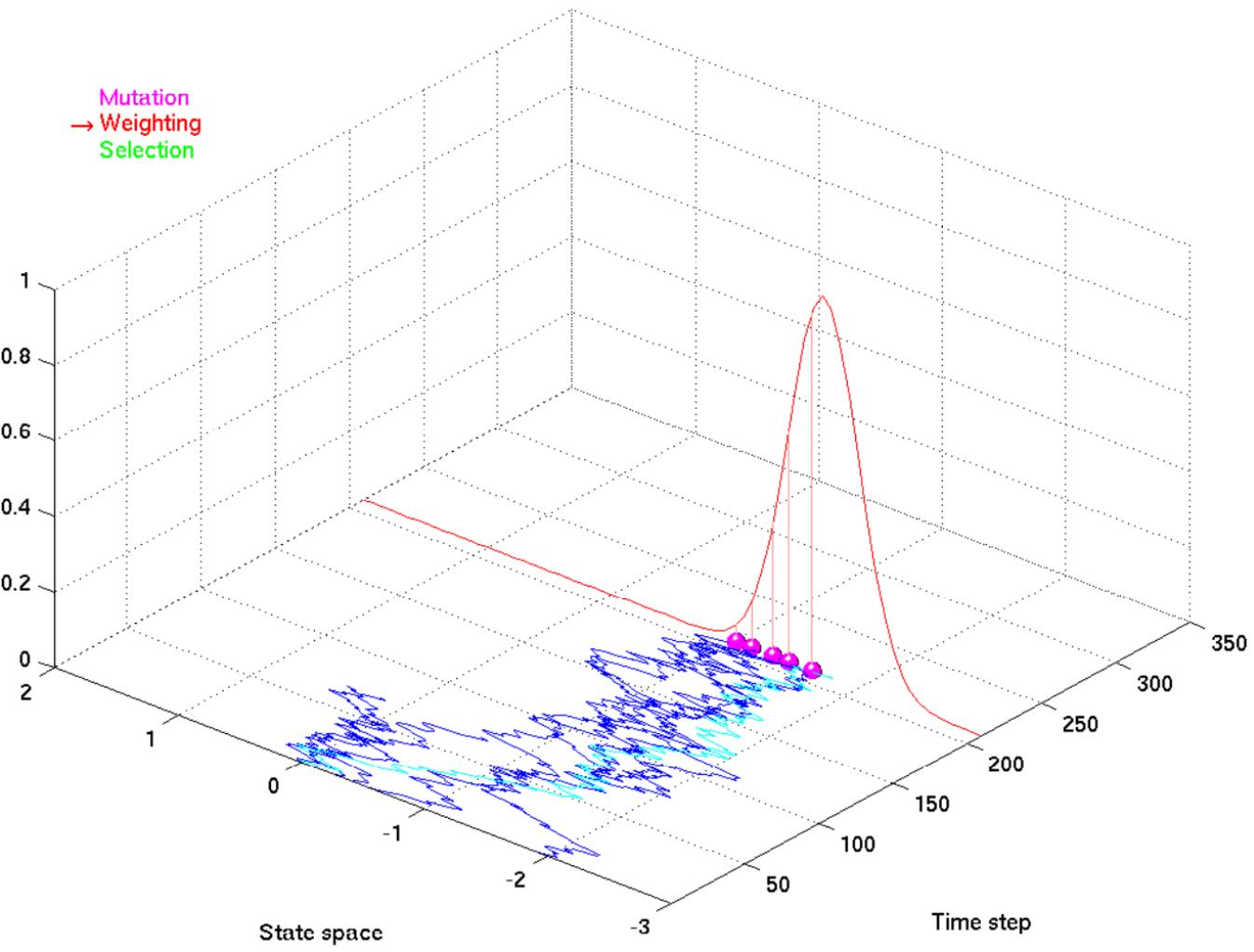
Mutation  
Weighting  
→ Selection



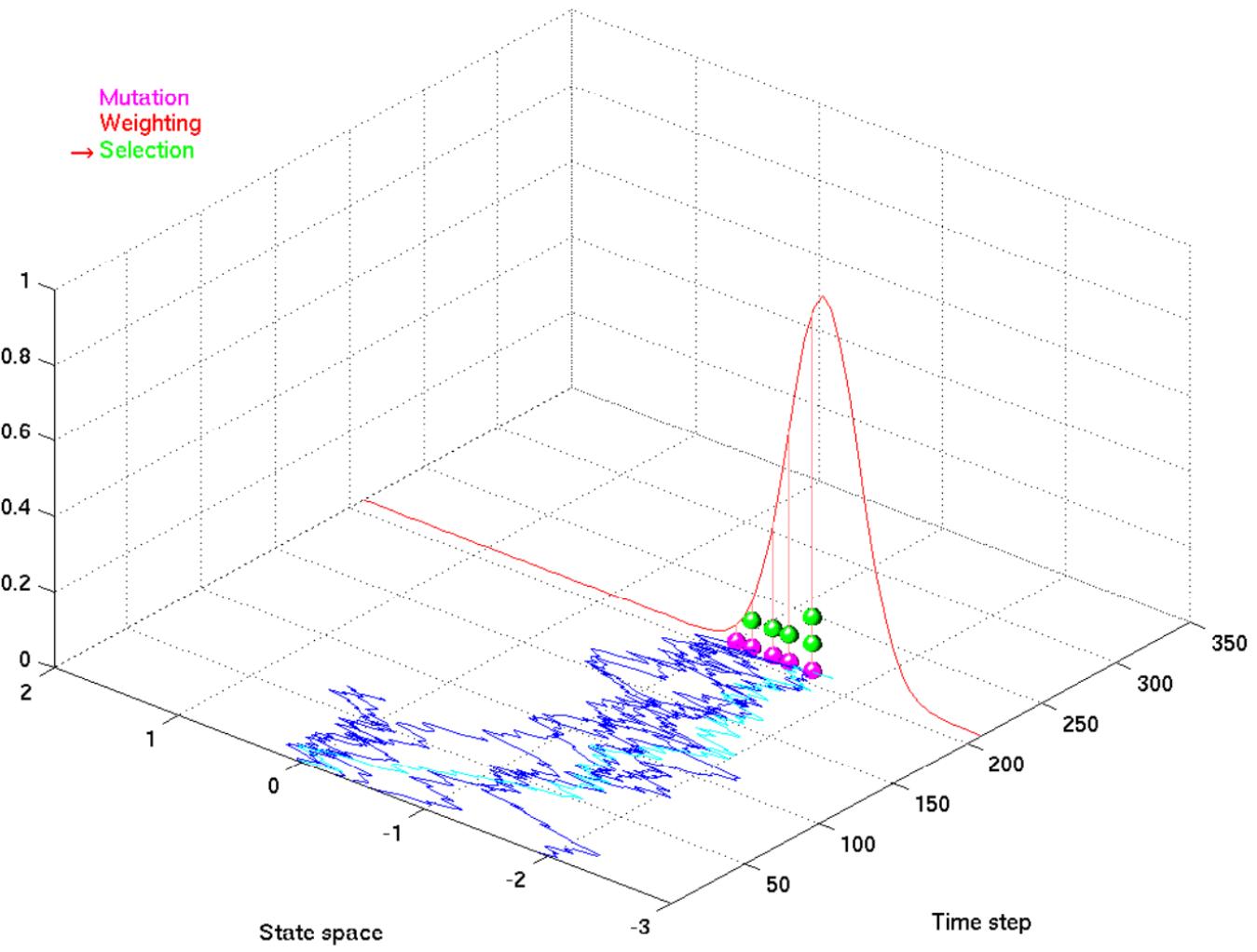
→ Mutation  
Weighting  
Selection



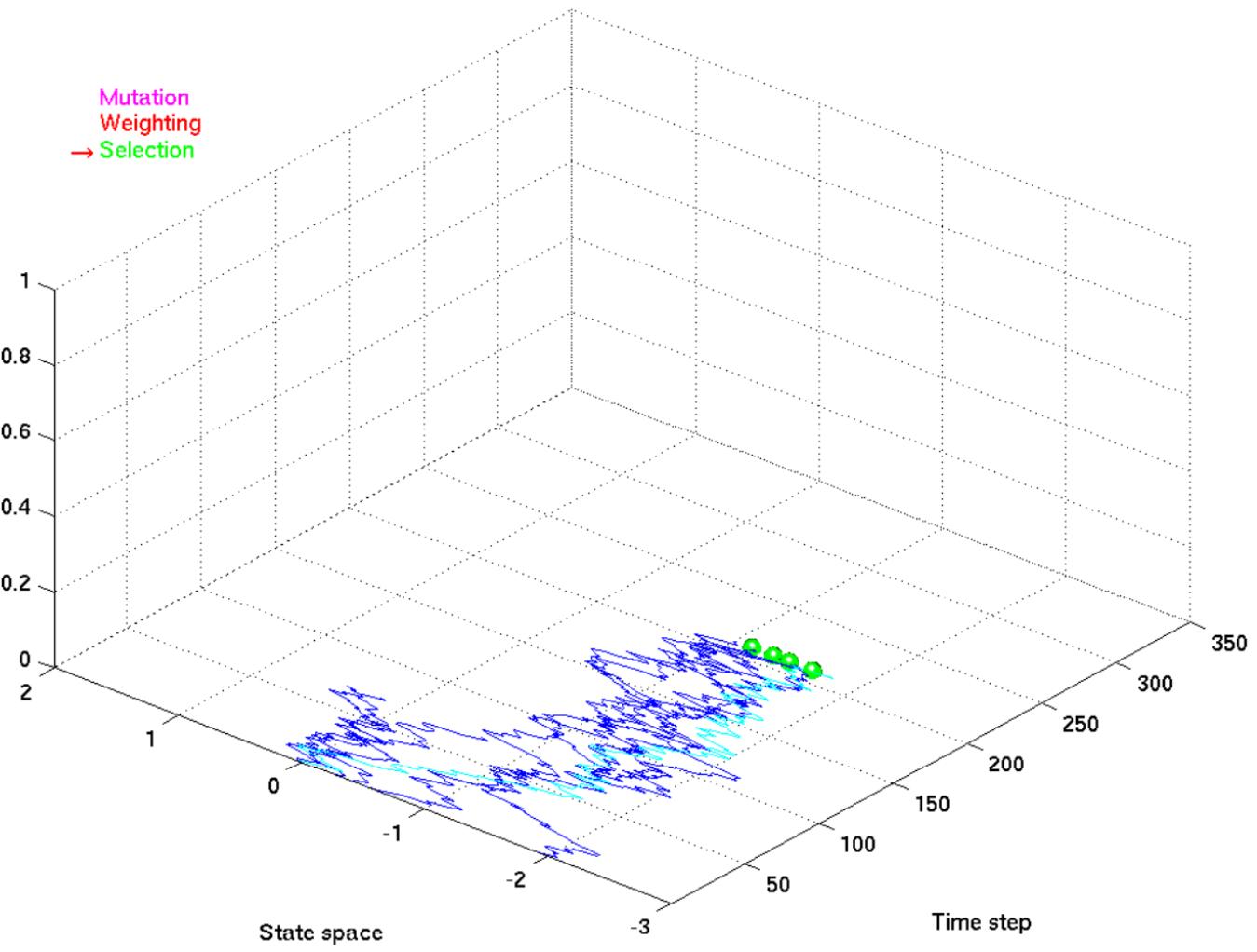


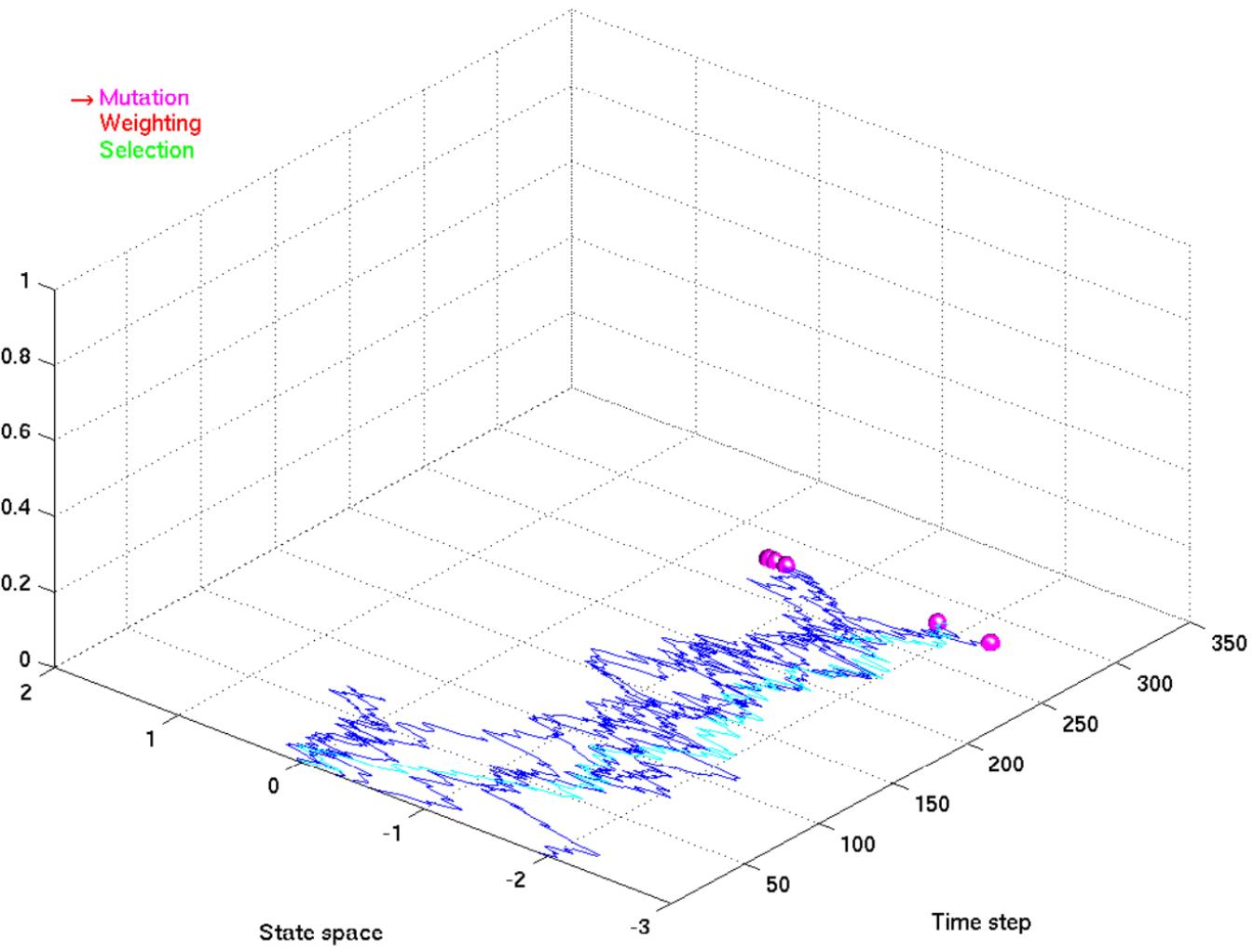


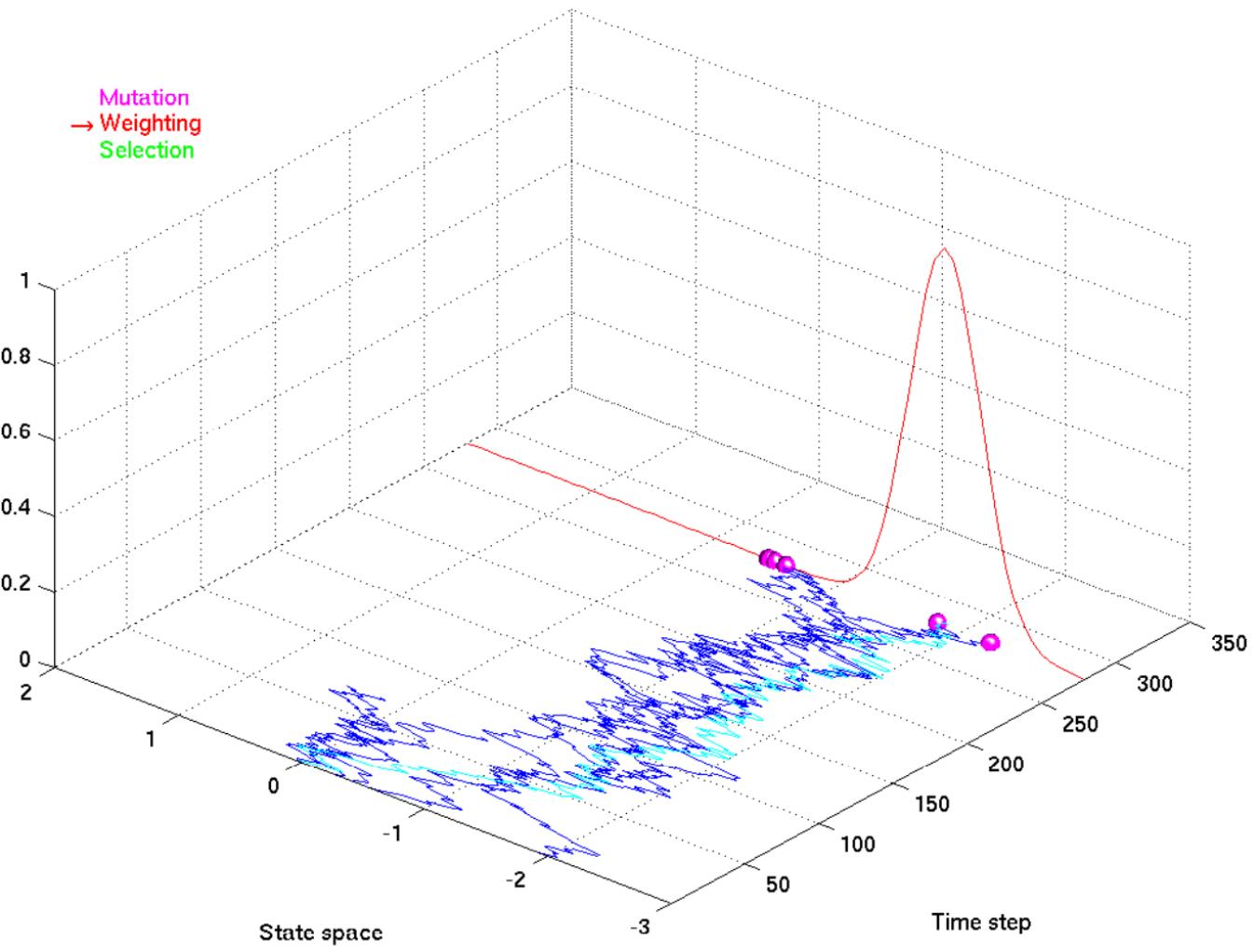
Mutation  
Weighting  
→ Selection

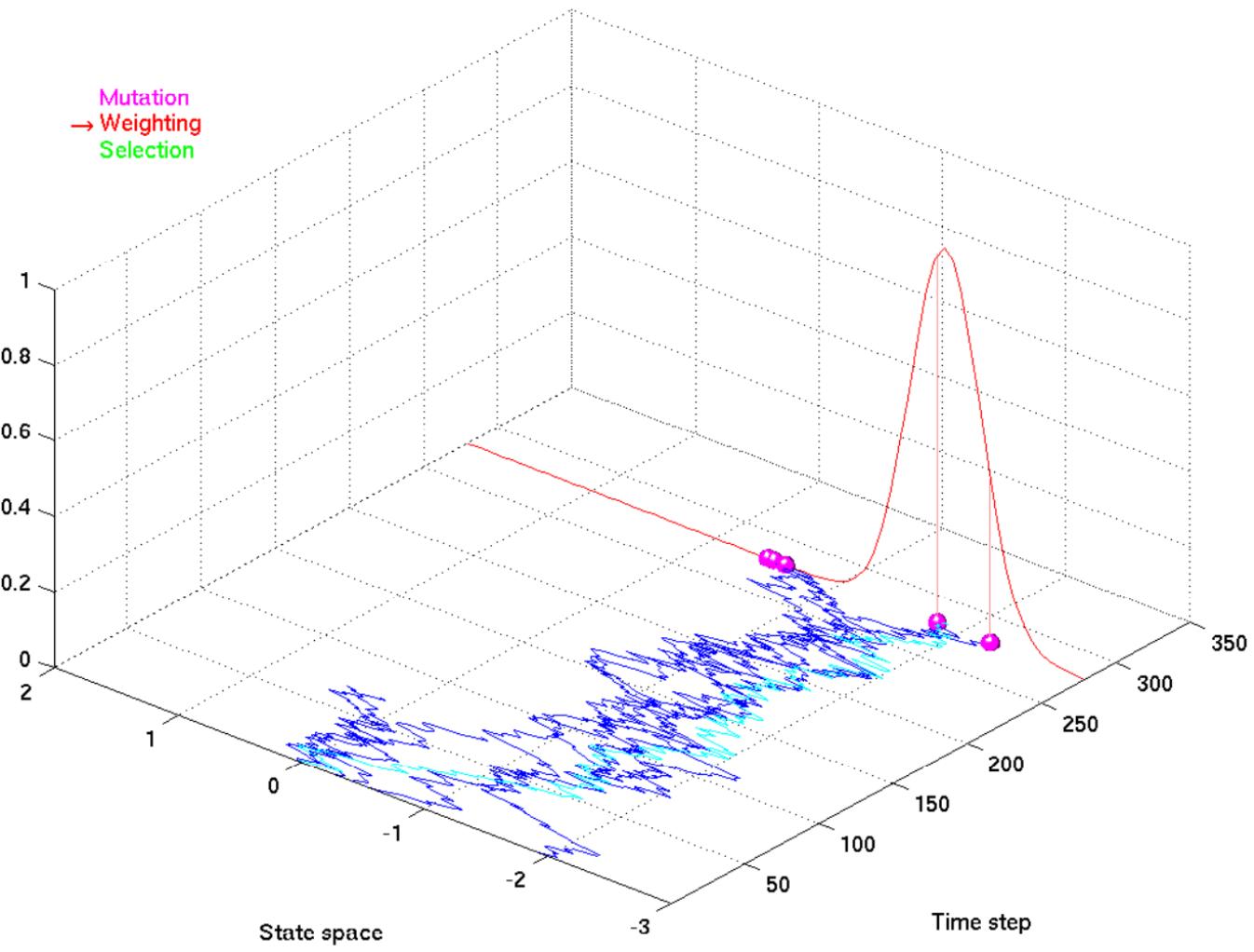


Mutation  
Weighting  
→ Selection

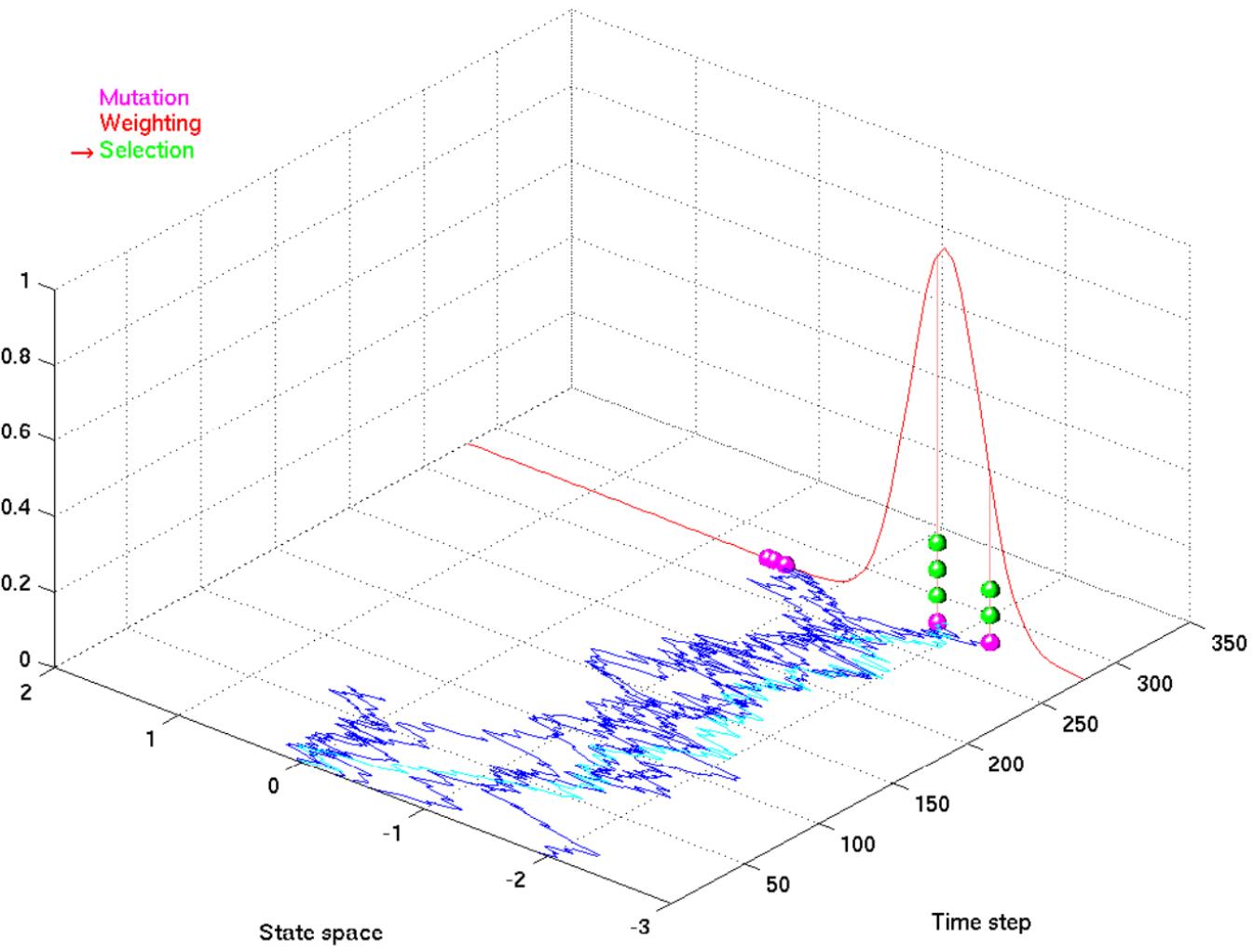




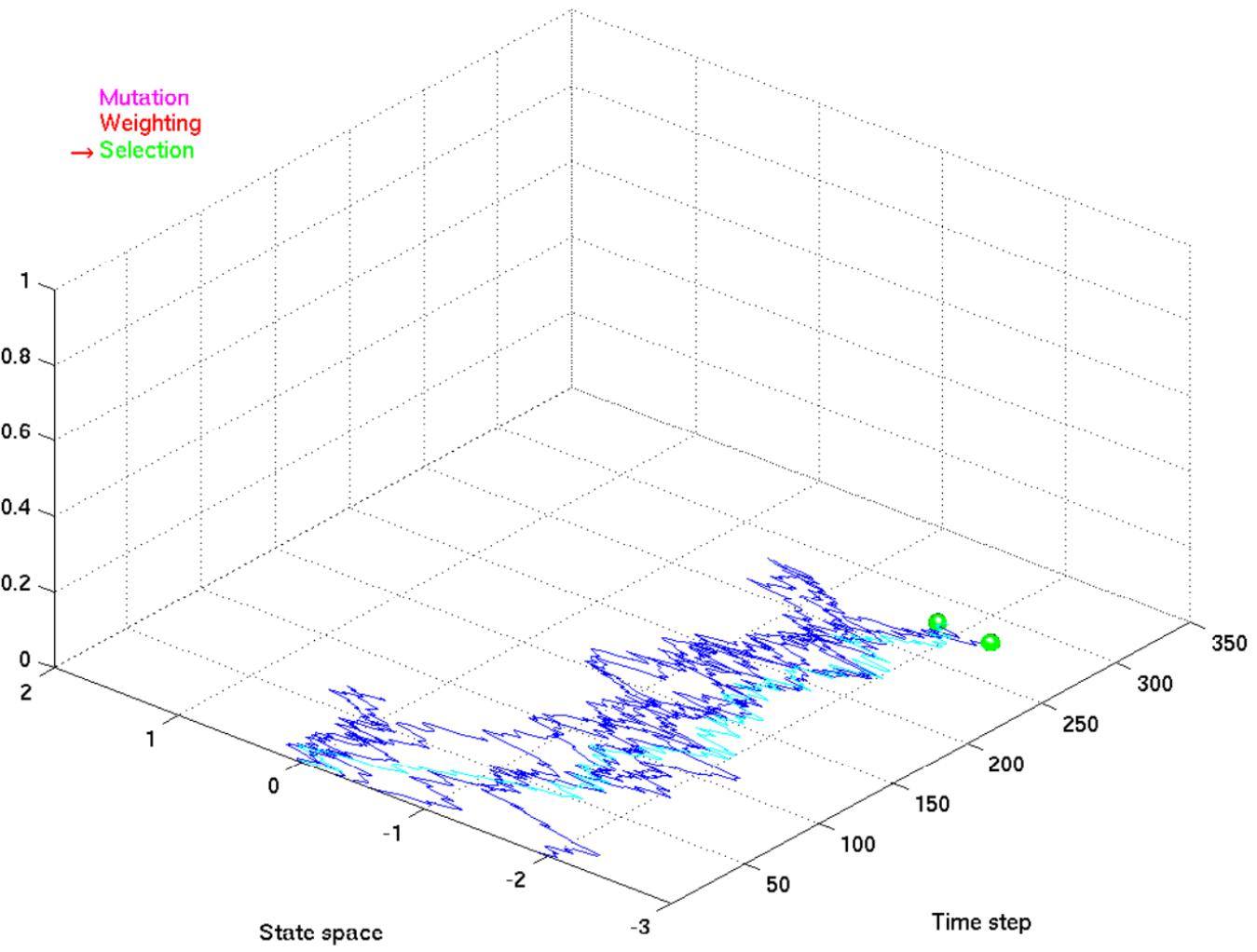




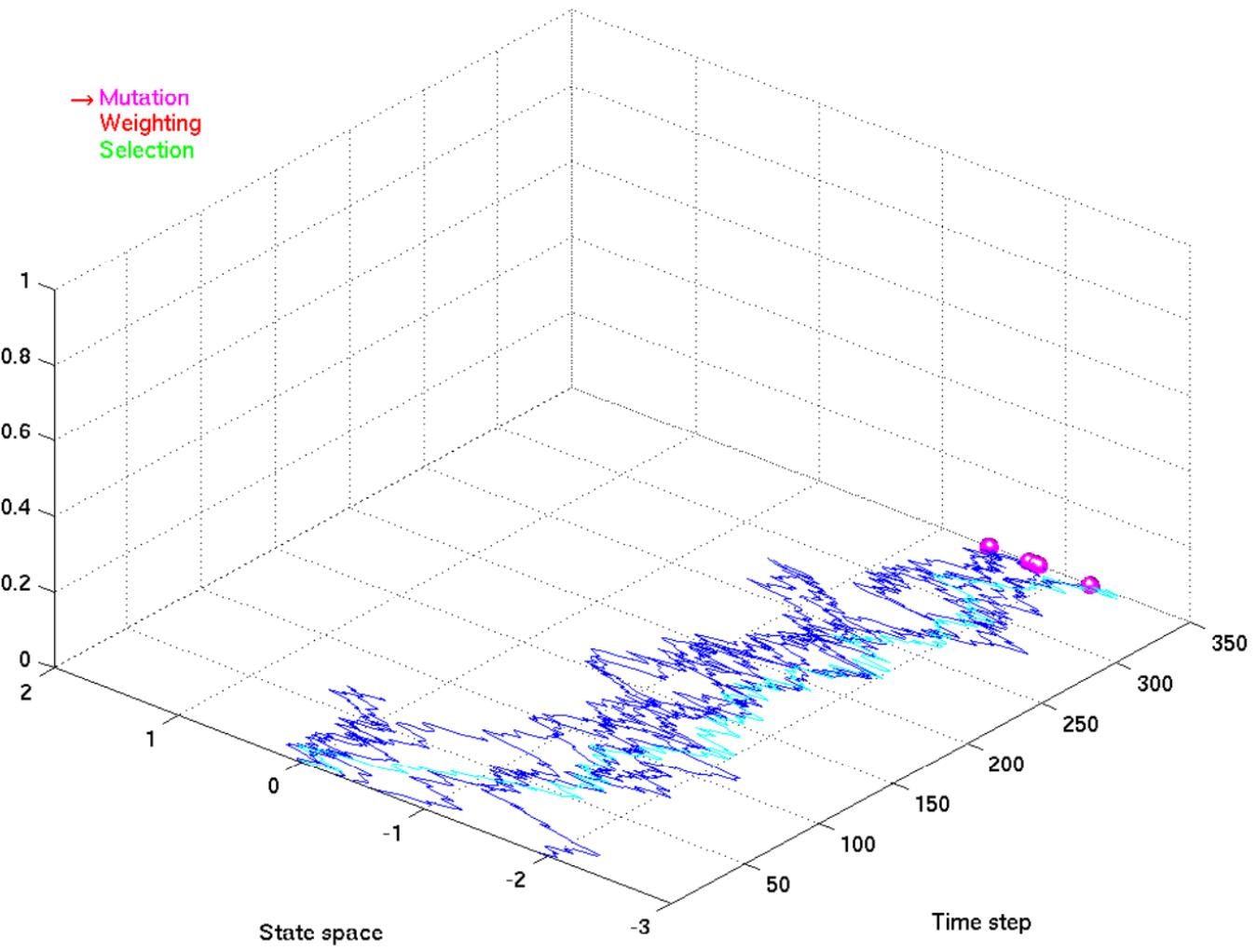
Mutation  
Weighting  
→ Selection

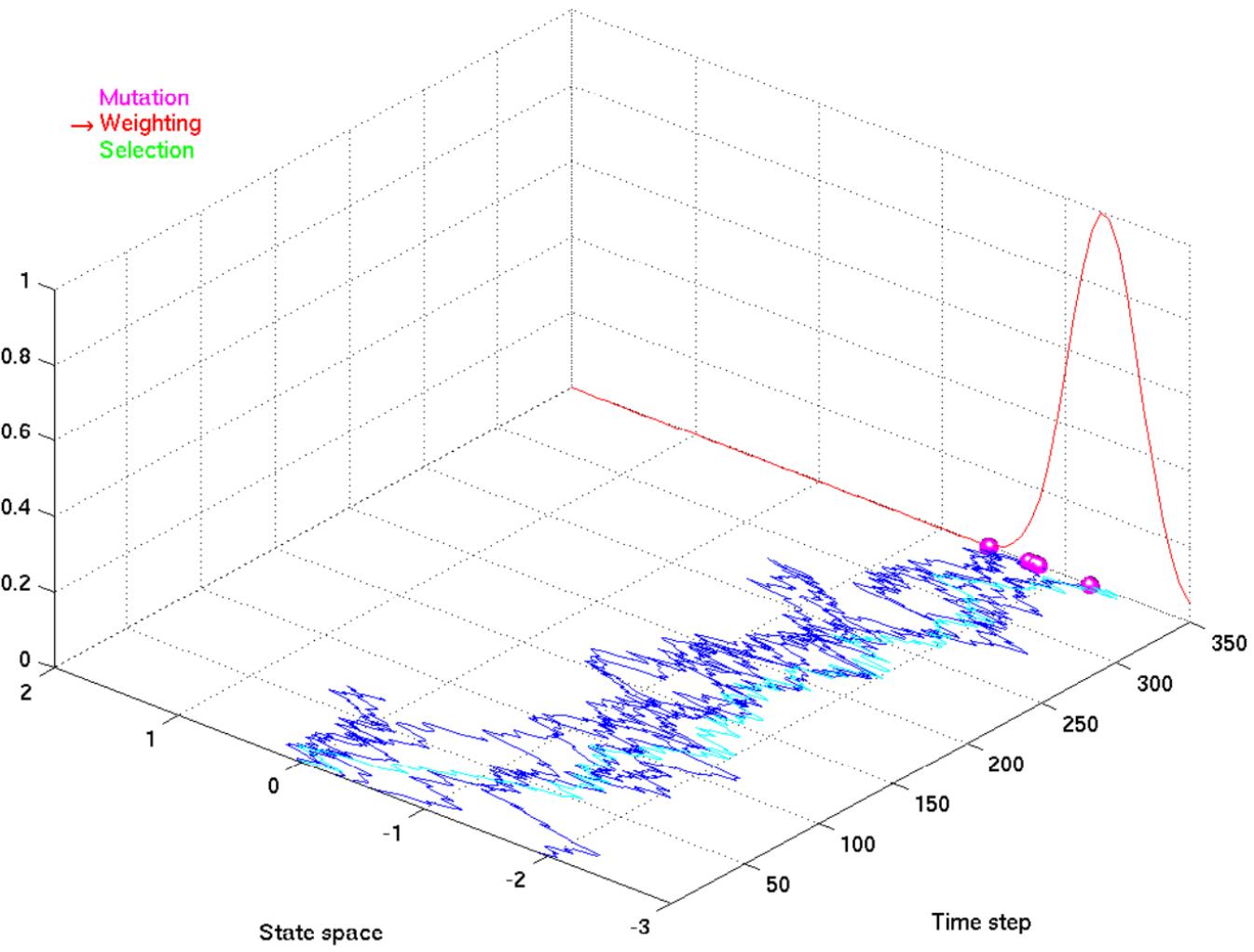


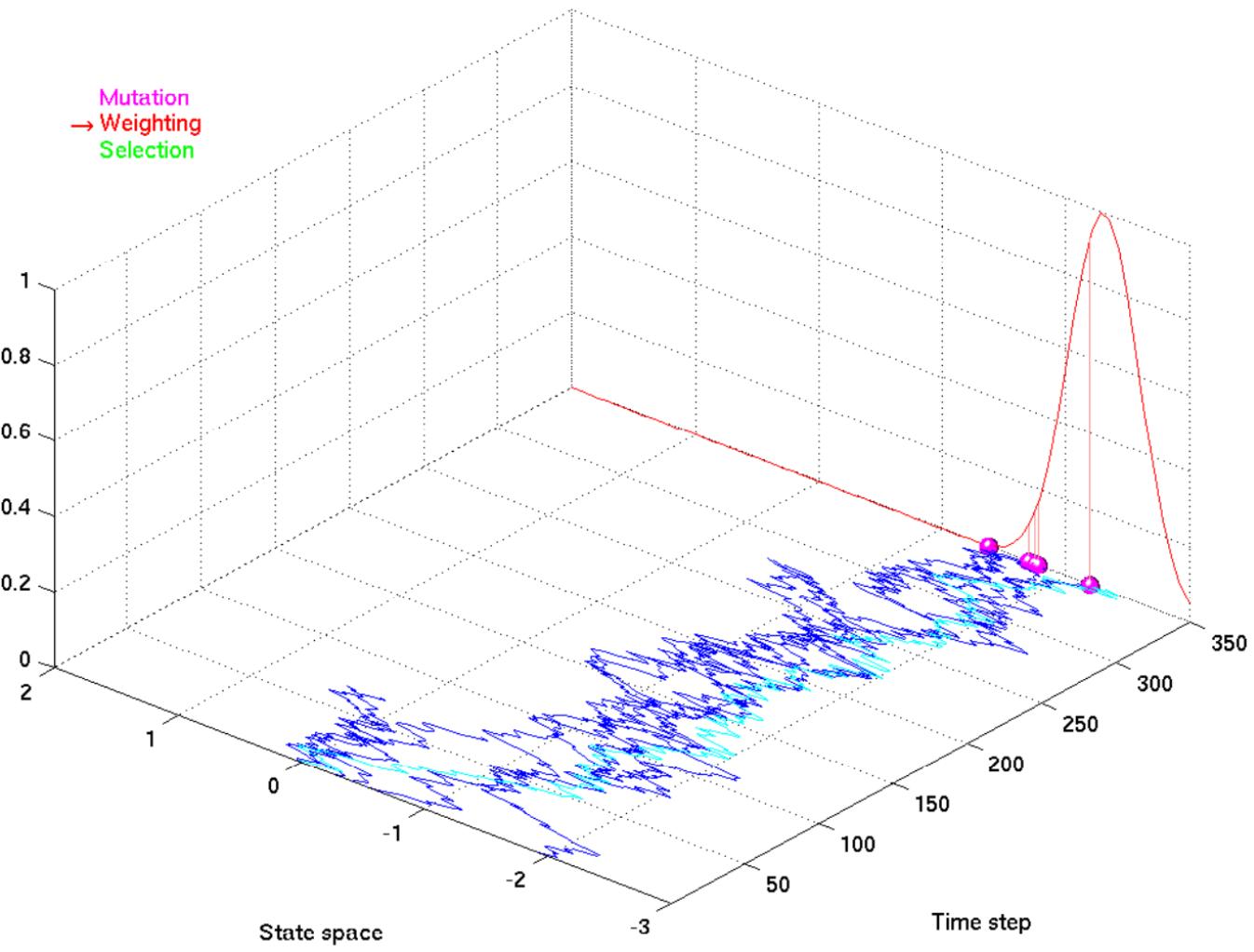
Mutation  
Weighting  
→ Selection



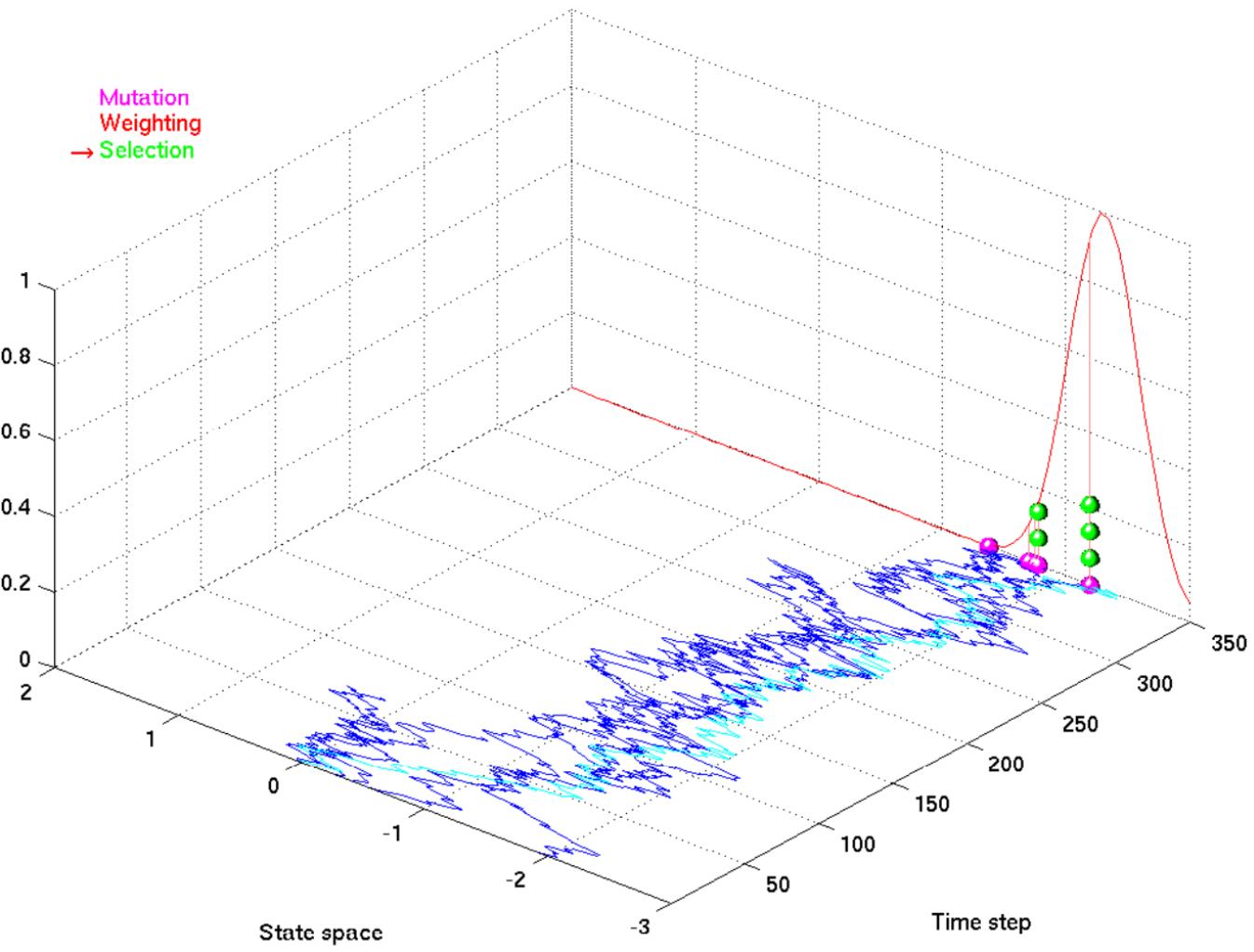
→ Mutation  
Weighting  
Selection





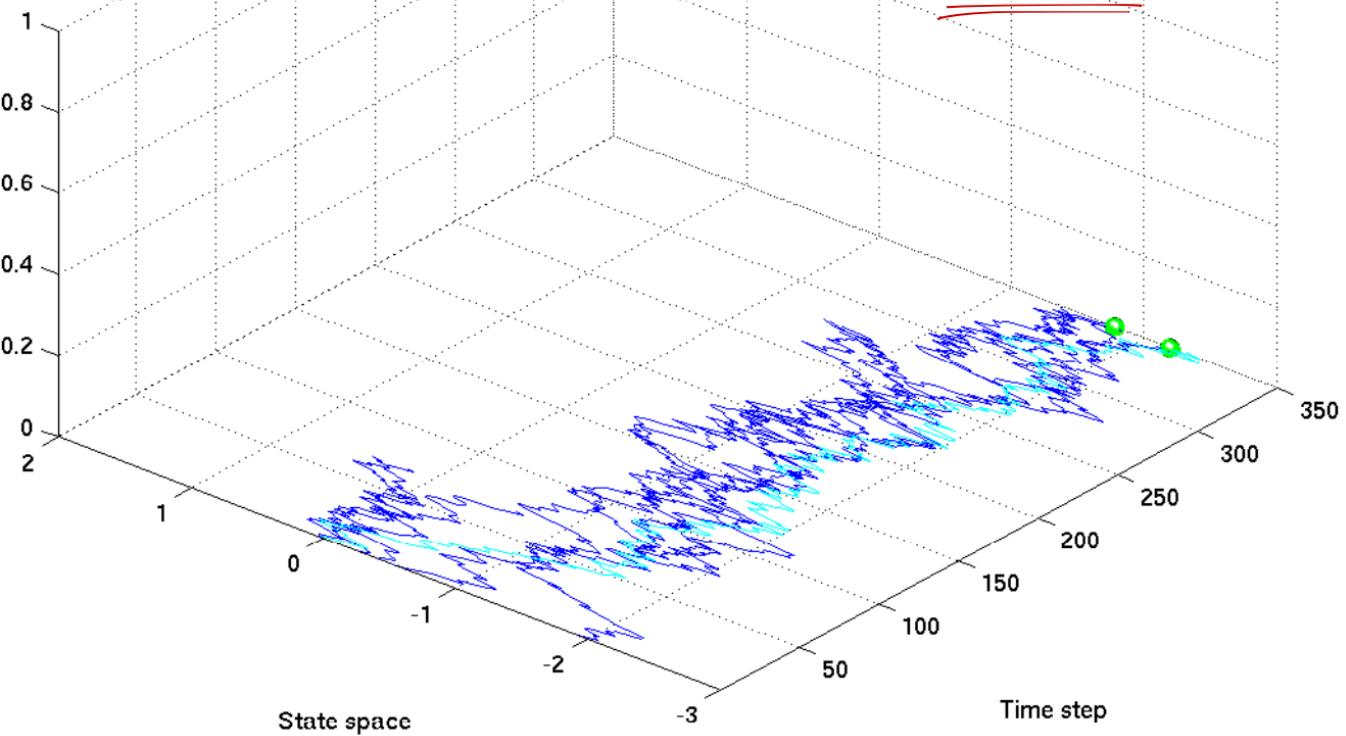


Mutation  
Weighting  
→ Selection



Mutation  
Weighting  
→ Selection

$$p(x_{1:\tau} | y_{1:\tau})$$



## Sequential MC

1.  $p(x_1 | y_1)$

$$p(x_1 | y_1) \approx \sum_i w^i \delta_{x_1^i}$$

weights

$$w^i = \frac{p(x_1^i | y_1) / p(x_1^i)}{\sum_i p(x_1^i | y_1) / p(x_1^i)} = \frac{p(y_1 | x_1^i)}{\sum_i p(y_1 | x_1^i)}$$

un-normalized weights:  $w_i = p(y_1 | x_1^i)$

$$\underline{p(y_1) = \int p(x_1, y_1) dx_1 = \int p(y_1 | x_1) p(x_1) dx_1 = E_{x_1}(p(y_1 | x_1))}$$

$$\approx \frac{1}{N} \sum_{i=1}^N w_i \quad \text{estimate of the marginal likelihood}$$

## Particle MCMC

In general, one can obtain

$$\hat{p}(y_{1:T}) \approx \frac{1}{N} \sum \text{(importance sampling weights)}$$

$\equiv$

(see Andrieu et al. 2010)

Exact MCMC : (MH)

propose  $x_{1:T}^* \sim \hat{p}(x_{1:T} | y_{1:T})$  (sequential MC procedure)

accept with probability

$$\alpha = \min \left[ 1, \frac{\hat{p}(y_{1:T})^*}{\hat{p}(y_{1:T})^{\text{current}}} \right]$$

targets  $p(x_{1:T} | y_{1:T})$  !

(details later 😊)