Towards a coalescence of

DA and EPS

Creating equally likely

Initial Conditions
Ensemble Members

Properties of an

Ideal ensemble

IDEAL ENSEMBLE

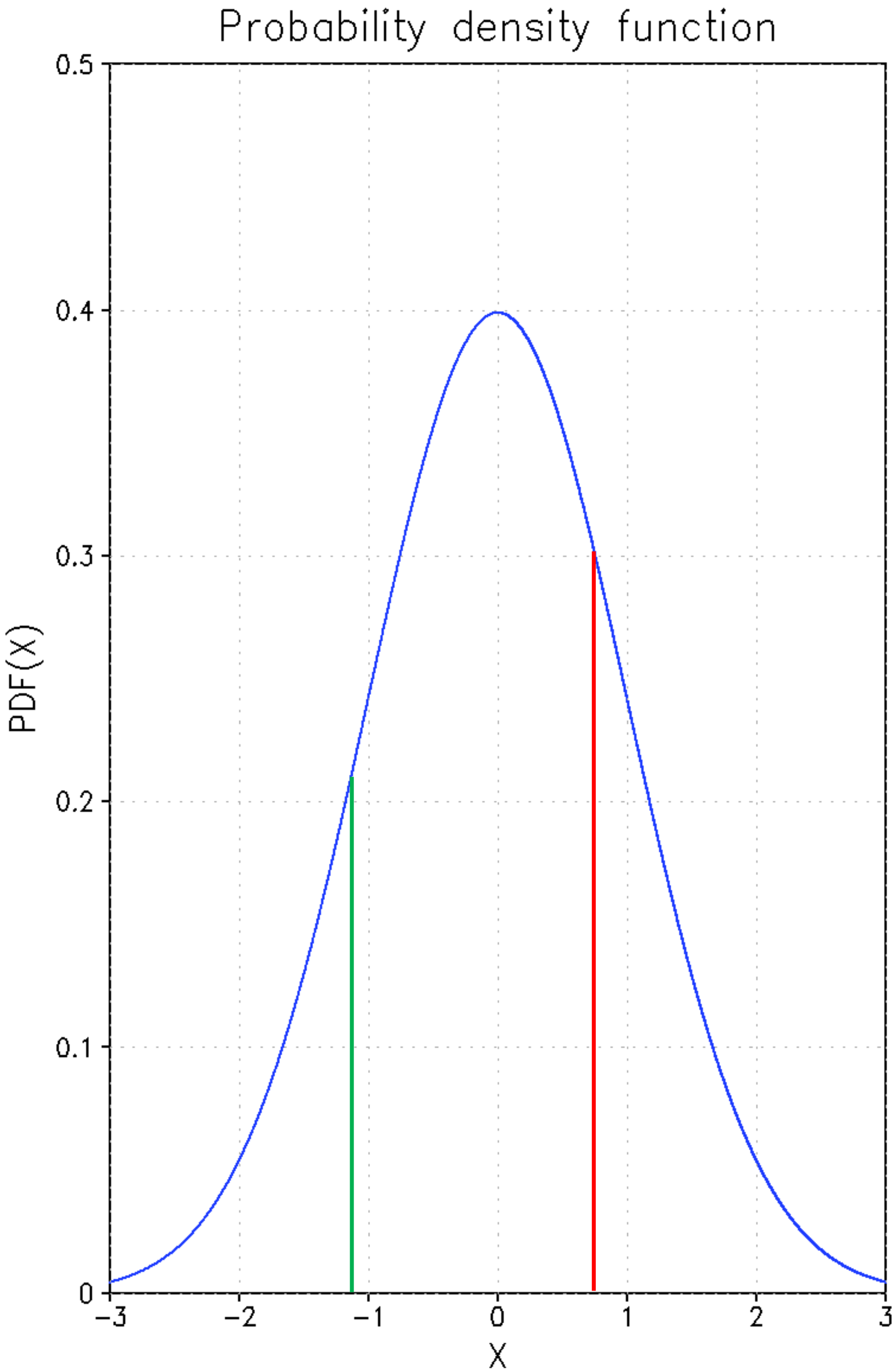
- Infinitely many members
- All members are
  - i.i.d.
  - statistically equal to the atmosphere

which implies that

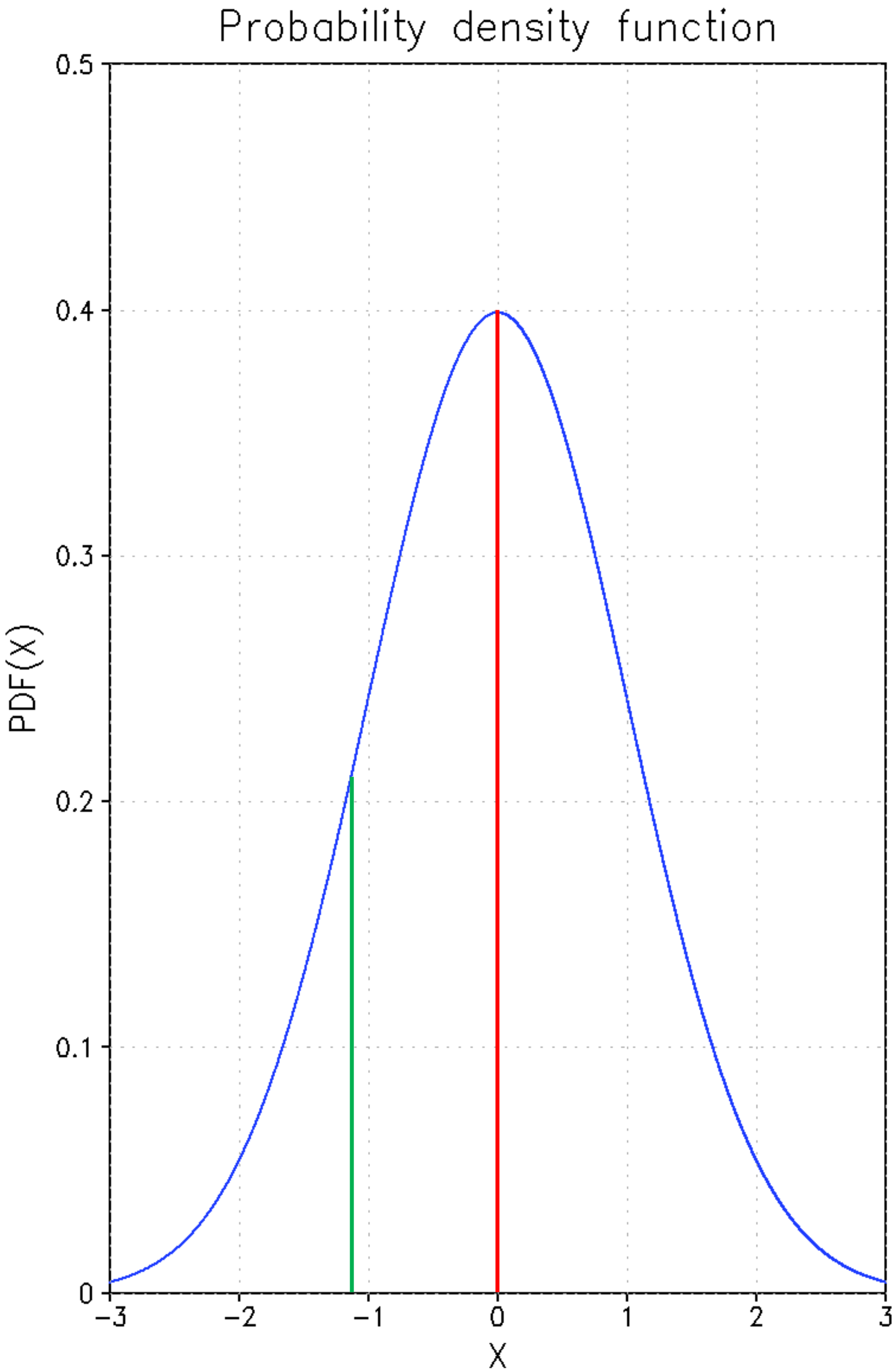
- All members are
  - equally likely to be the truth
  - bias-free and have correct variance
- Skill = Spread

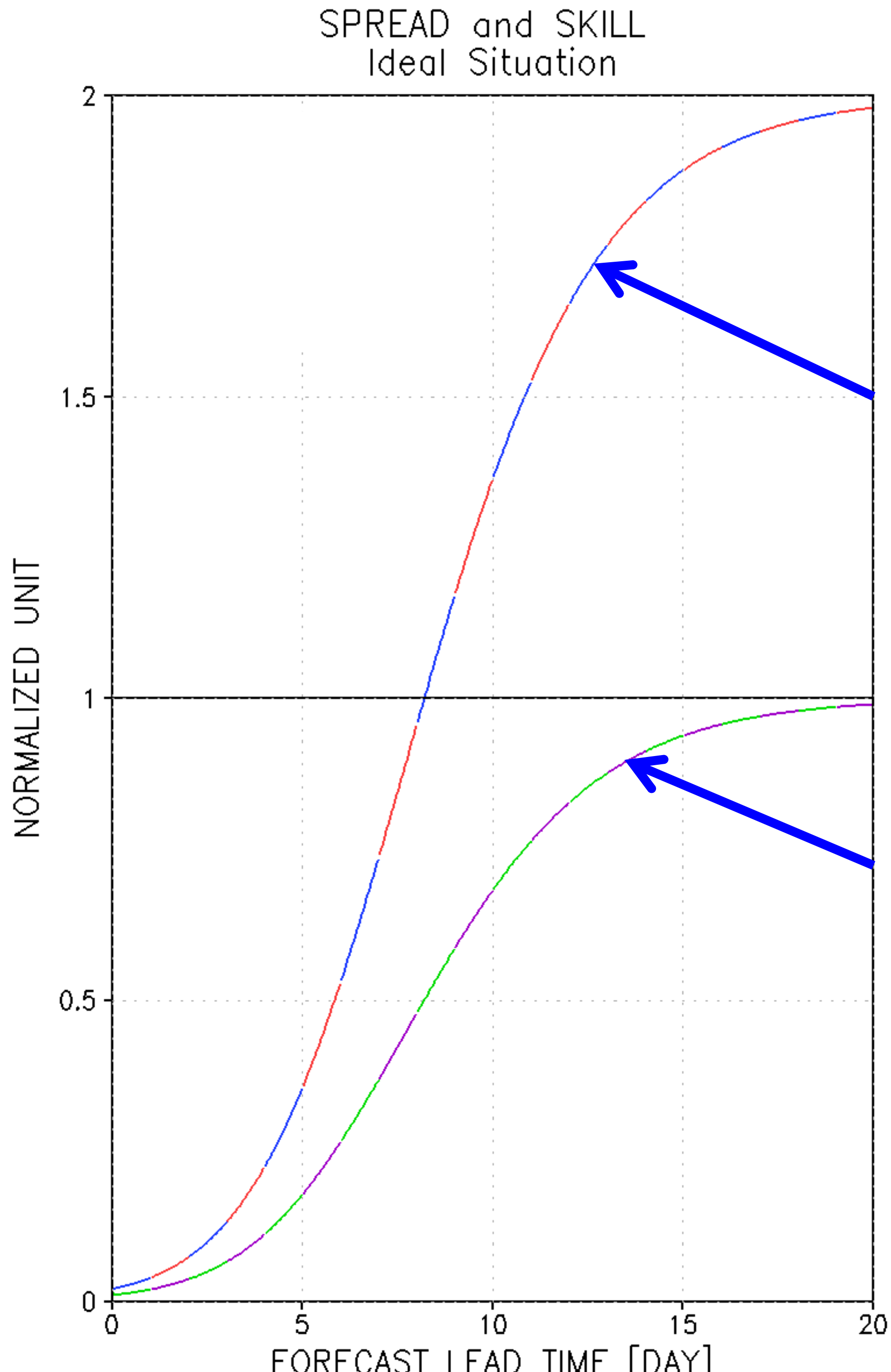
Relationship between the skill of Ensemble Mean and Individual Ensemble Members

$$\mathrm{E}\left[\mathrm{MSE}_{\mathrm{M}}\right] = \frac{1}{2}\mathrm{E}\left[\mathrm{MSE}\right]$$



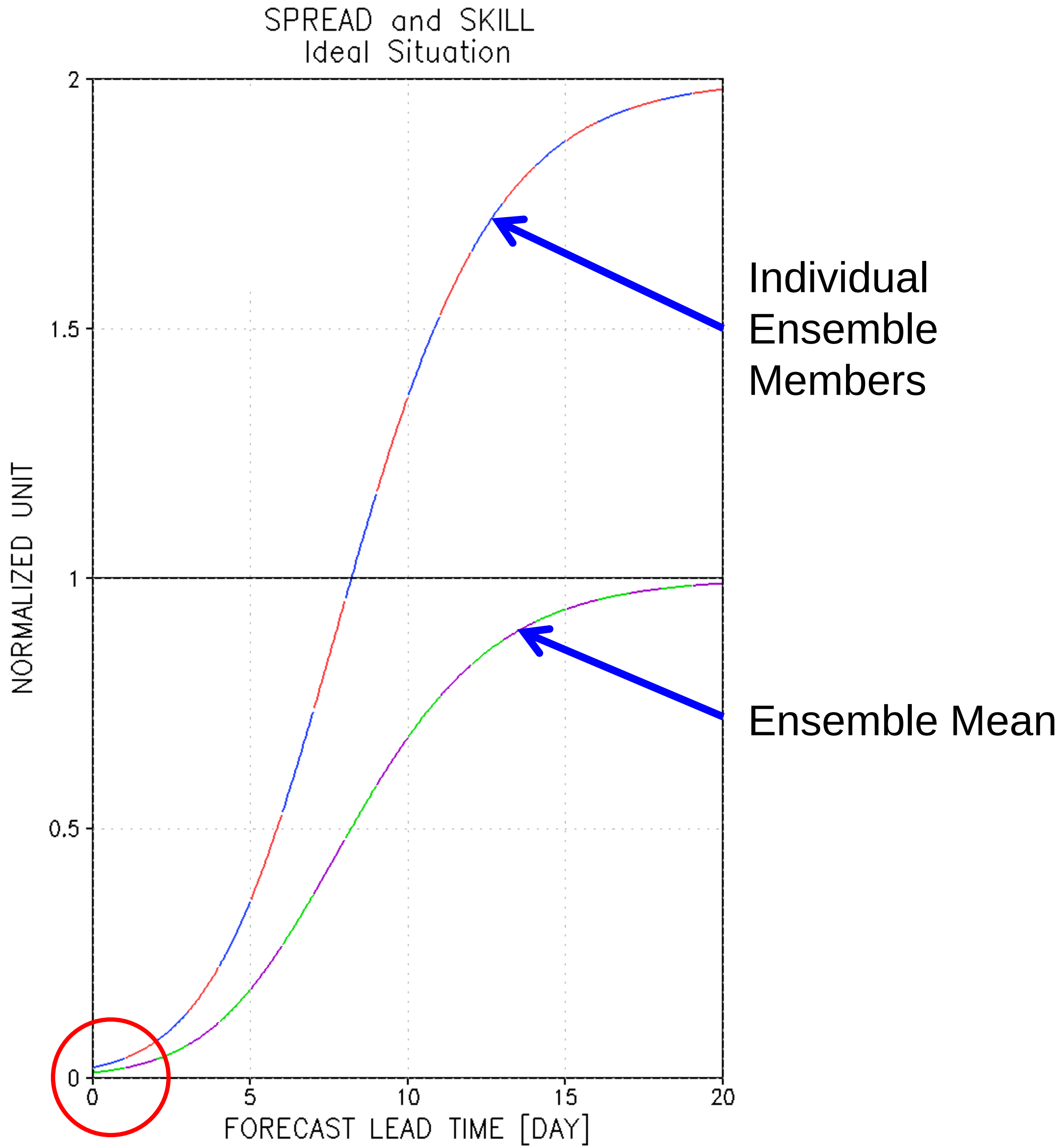
Probability density function
0.5
0.4
0.3
0.2
0.1
0
PDF(X)
−3
−2
−1
0
1
2
3
X


Probability density function
PDF(X)
X
0.5
0.4
0.3
0.2
0.1
0
−3
−2
−1
0
1
2
3

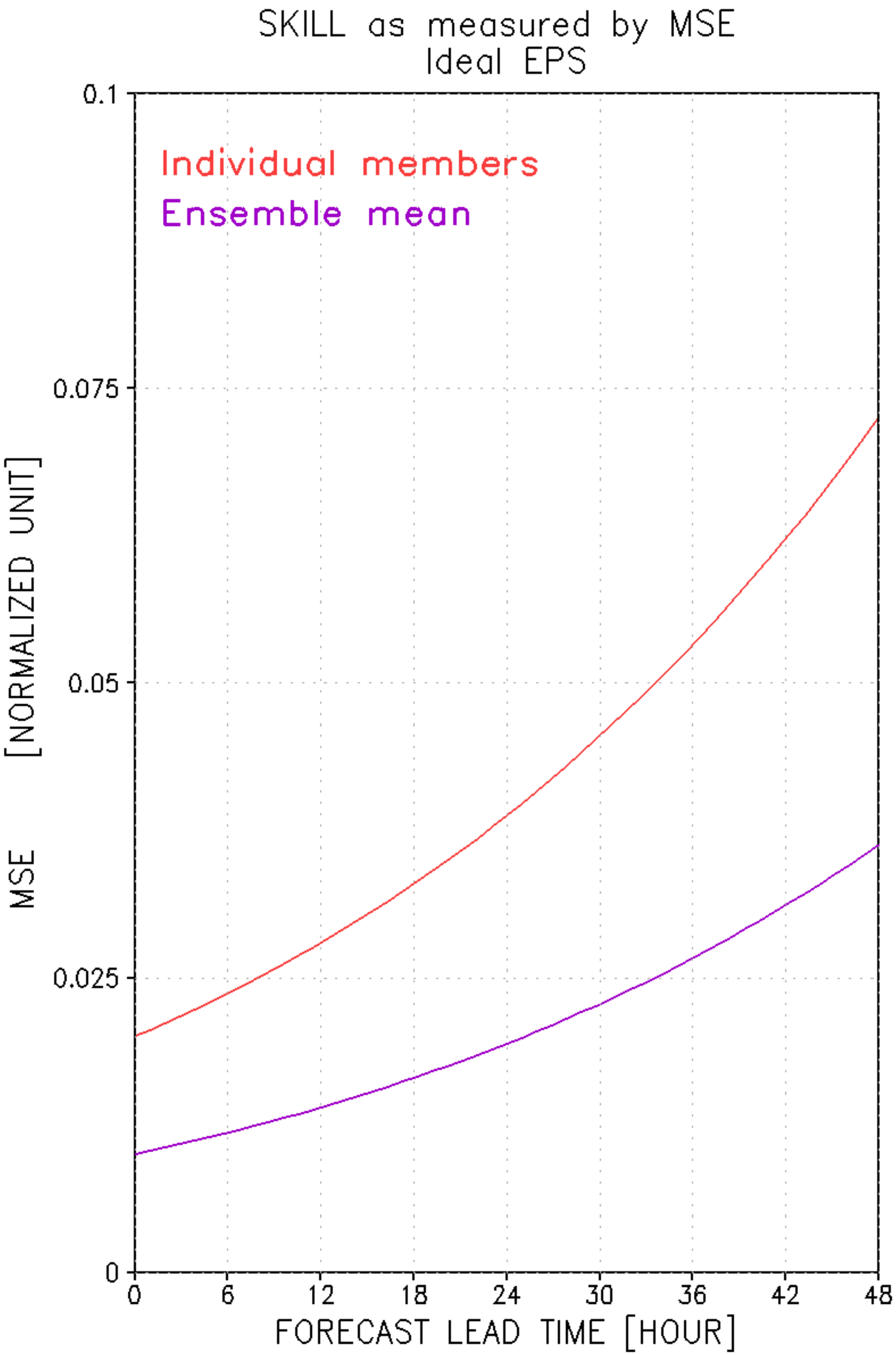


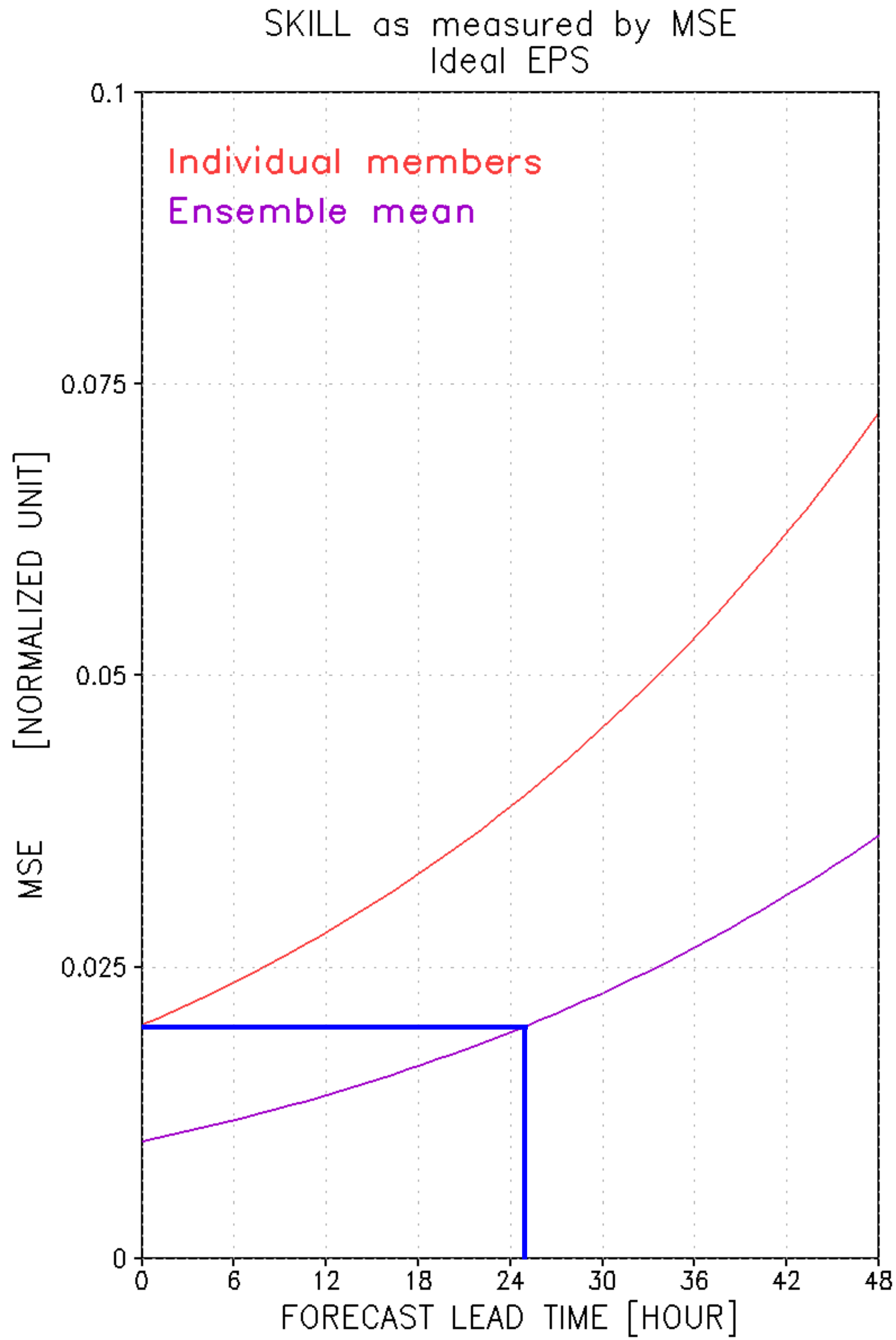
Individual Ensemble Members

Ensemble Mean



SPREAD and SKILL
Ideal Situation
NORMALIZED UNIT
2
1.5
1
0.5
0
0
5
10
15
20
FORECAST LEAD TIME [DAY]
Individual Ensemble Members
Ensemble Mean


SKILL as measured by MSE
Ideal EPS
Individual members
Ensemble mean
0.1
0.075
0.05
0.025
0
MSE [NORMALIZED UNIT]
0
6
12
18
24
30
36
42
48
FORECAST LEAD TIME [HOUR]


SKILL as measured by MSE
Ideal EPS
Individual members
Ensemble mean
MSE [NORMALIZED UNIT]
0.1
0.075
0.05
0.025
0
0
6
12
18
24
30
36
42
48
FORECAST LEAD TIME [HOUR]

Presently used

Paradigm for creating

IC for EPS

Present Paradigm

1. A control analysis is assumed to be the uniquely best estimate of the IC

Present Paradigm

1. A control analysis is assumed to be the uniquely best estimate of the IC

2. An ensemble member is defined as the sum of the control analysis and a perturbation

Present Paradigm

1. A control analysis is assumed to be the uniquely best estimate of the IC

2. An ensemble member is defined as the sum of the control analysis and a perturbation

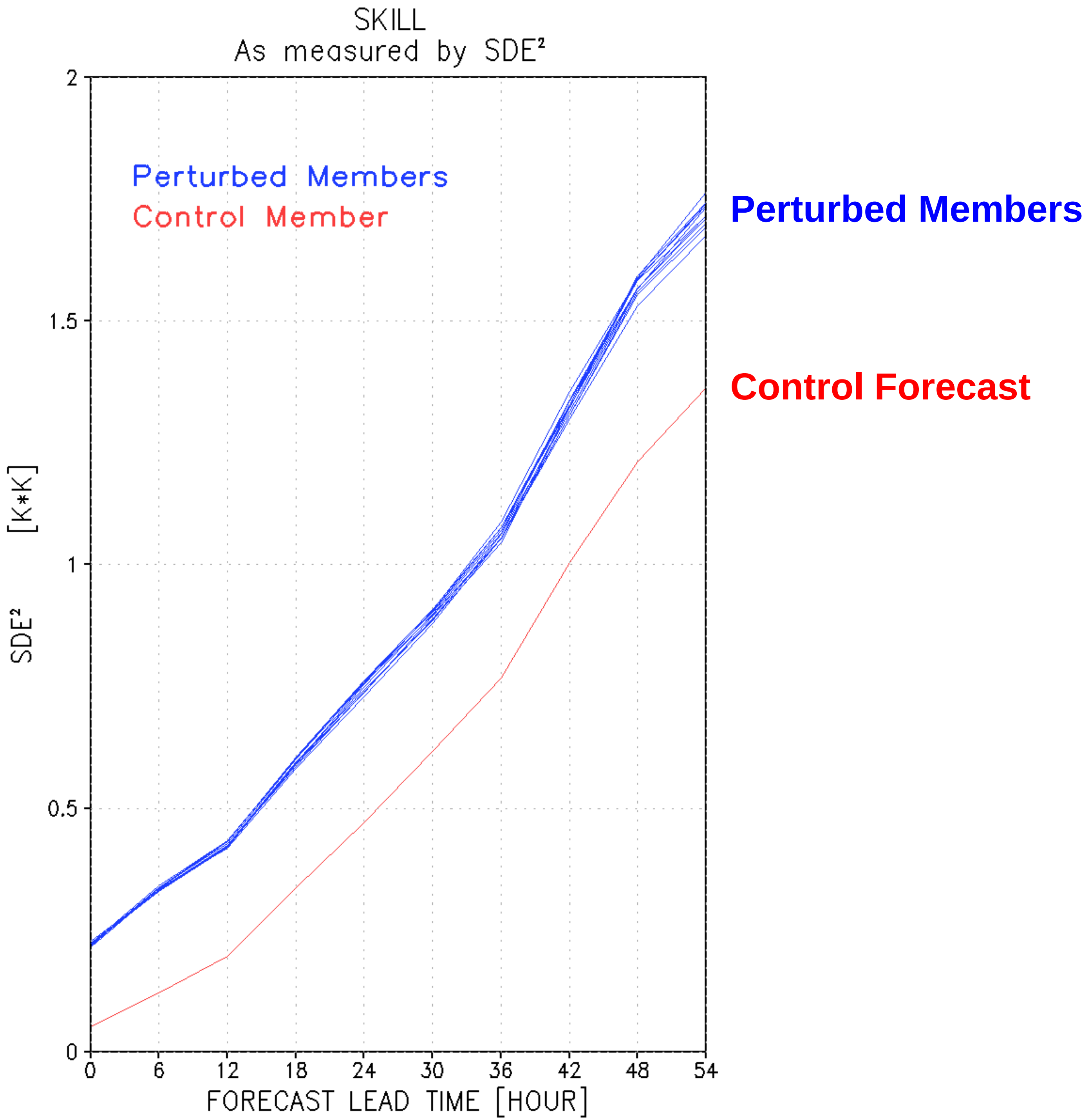
3. The sum of all ensemble perturbations add up to zero – that is – they are centered around the control analysis

Consequence 1

Perturbed ensemble members at the initial time is constructed by adding a perturbation to the control analysis

↓

The Perturbed members are then – by construction – made inferior to the Control Analysis

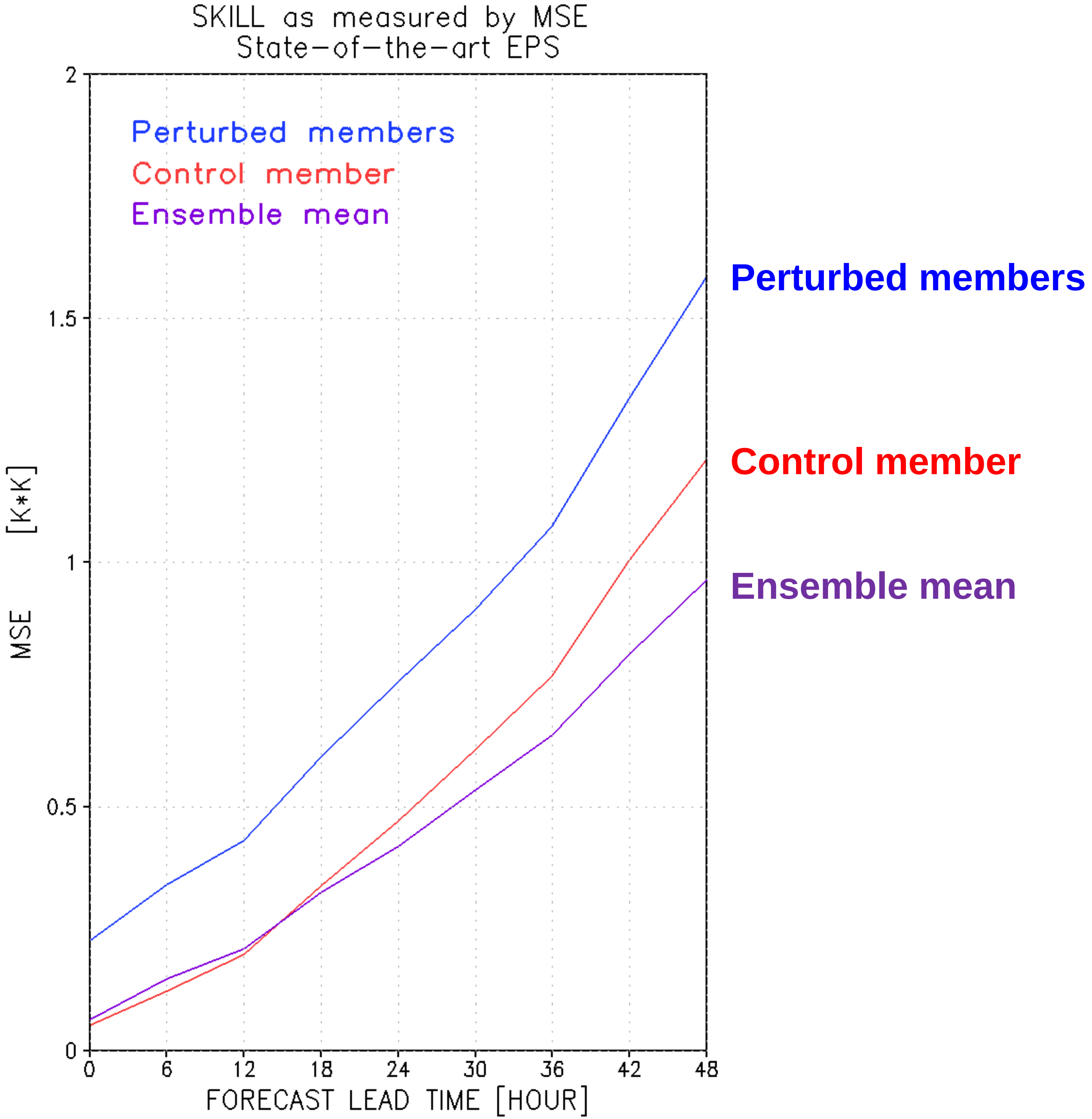

SKILL
As measured by SDE²
Perturbed Members
Control Member
SDE² [K*K]
FORECAST LEAD TIME [HOUR]
0
0.5
1
1.5
2
0
6
12
18
24
30
36
42
48
54
Perturbed Members
Control Forecast

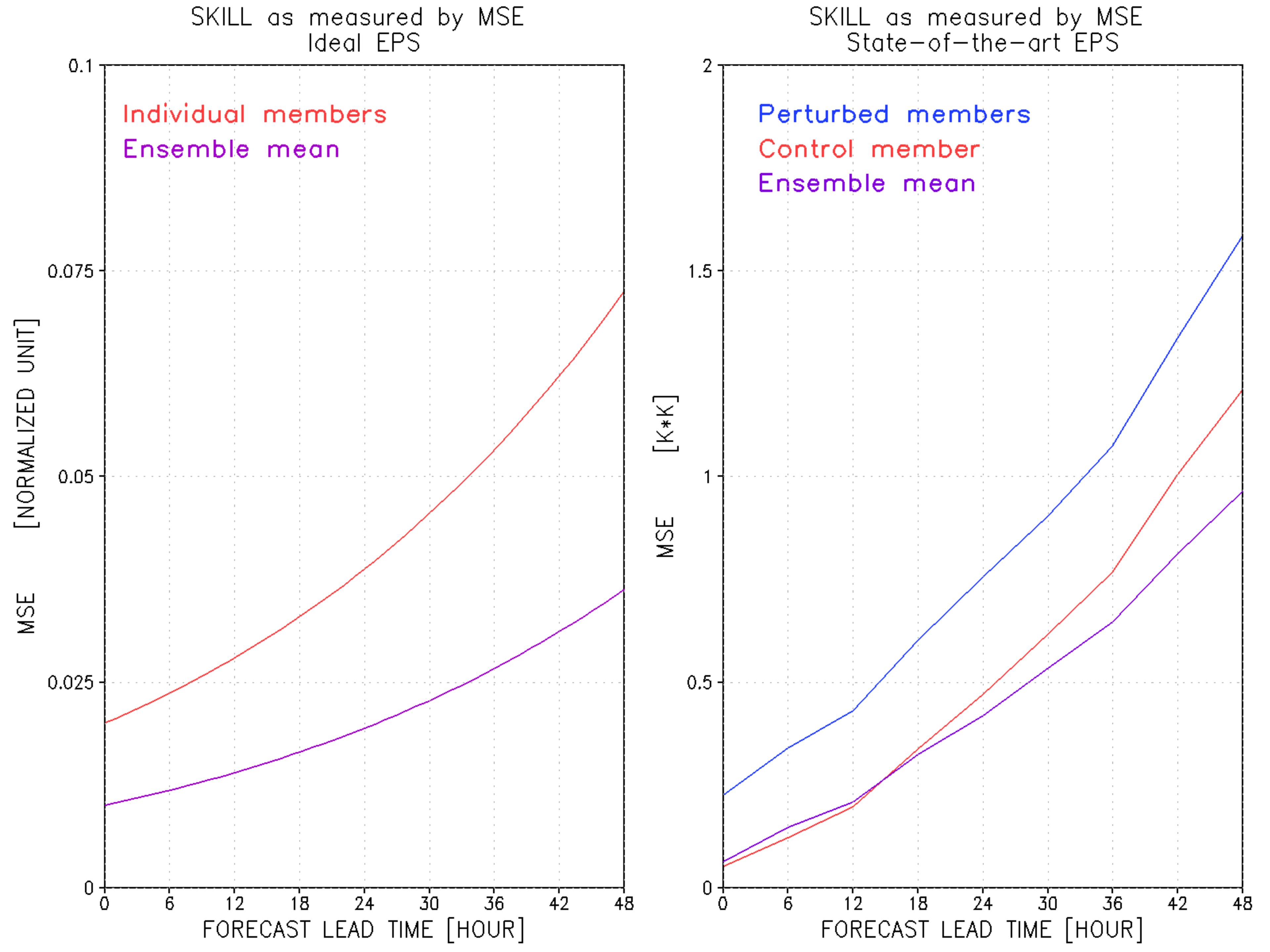
Consequence 2

Centering the perturbed ensemble members around the control analysis

The Ensemble Mean is then – by construction – made equal to the Control Analysis

The error growth for the ensemble mean is almost identical to that of the control forecast at the beginning of the forecast.


SKILL as measured by MSE
State-of-the-art EPS
Perturbed members
Control member
Ensemble mean
MSE [K*K]
0
0.5
1
1.5
2
0
6
12
18
24
30
36
42
48
FORECAST LEAD TIME [HOUR]
Perturbed members
Control member
Ensemble mean

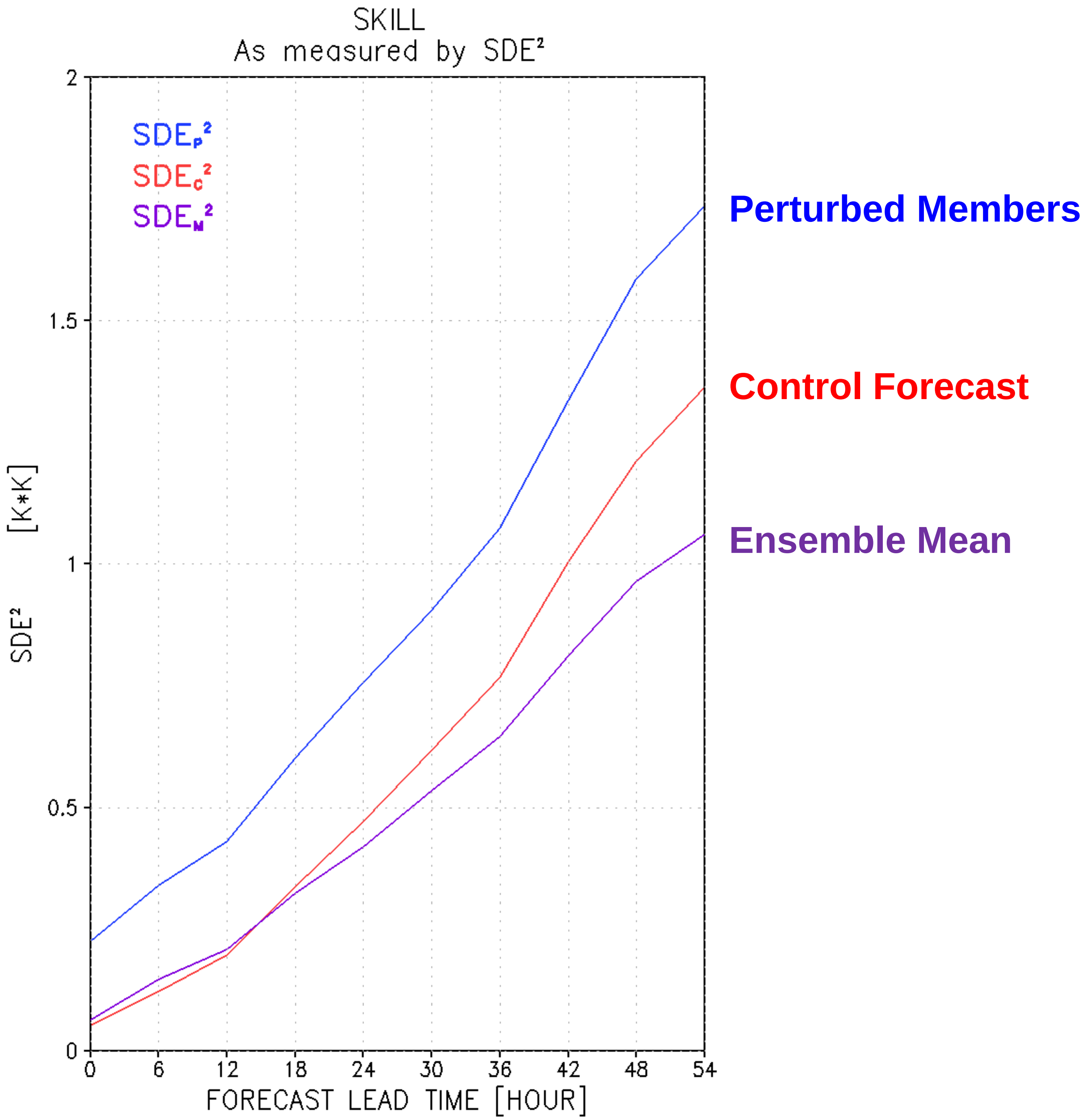

SKILL as measured by MSE
Ideal EPS
Individual members
Ensemble mean
MSE [NORMALIZED UNIT]
0.1
0.075
0.05
0.025
0
0 6 12 18 24 30 36 42 48
FORECAST LEAD TIME [HOUR]
SKILL as measured by MSE
State-of-the-art EPS
Perturbed members
Control member
Ensemble mean
MSE [K*K]
2
1.5
1
0.5
0
0 6 12 18 24 30 36 42 48
FORECAST LEAD TIME [HOUR]

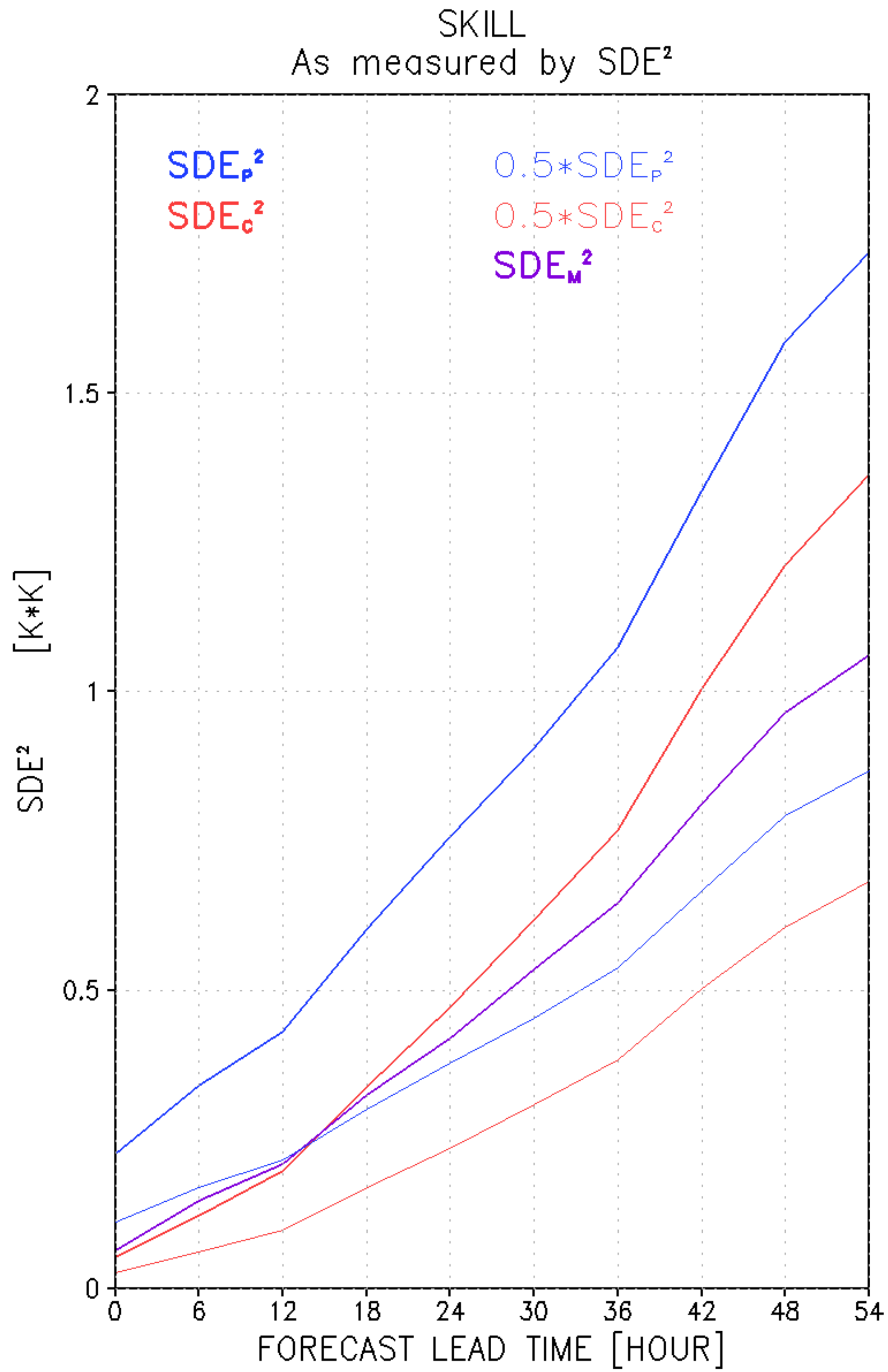
Consequence 3

The potential improvement in the skill of the Ensemble Mean forecast as indicated by

$$E\left[MSE_M\right] = \frac{1}{2} E\left[MSE\right]$$

is severly impeded by the fact that the errors of the perturbed members are so much larger than that of the control forecast


SKILL
As measured by SDE²
SDEP²
SDEC²
SDEM²
SDE² [K*K]
FORECAST LEAD TIME [HOUR]
0
0.5
1
1.5
2
0
6
12
18
24
30
36
42
48
54
Perturbed Members
Control Forecast
Ensemble Mean



Perturbed Members

Control Forecast

Ensemble Mean

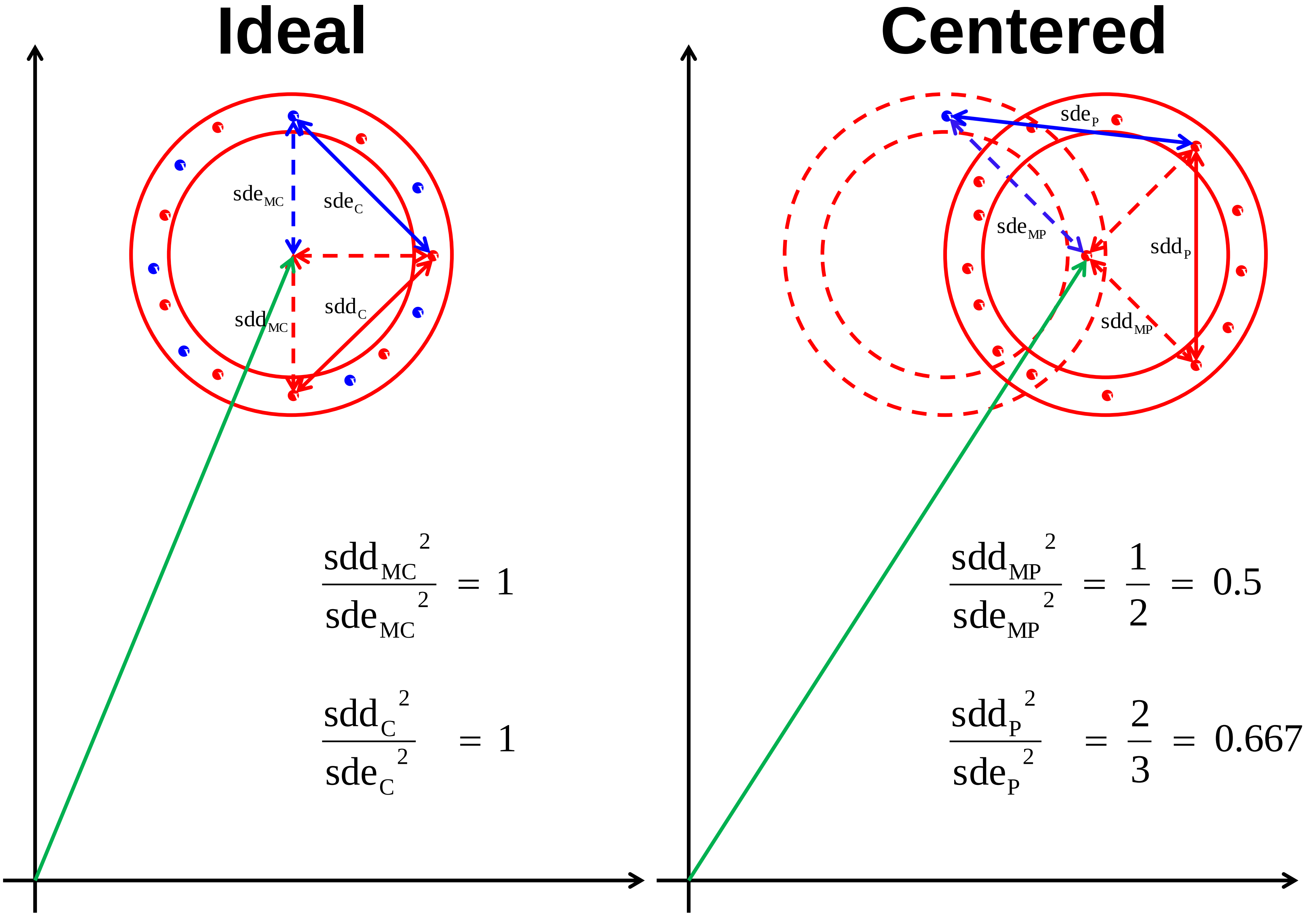
Ideal Ensemble Mean based on Pert Memb

Ideal Ensemble Mean based on Cntr Memb

Consequence 4

Centering the perturbed ensemble members around the control analysis

The Ensemble is thereby – by construction – made under-dispersive



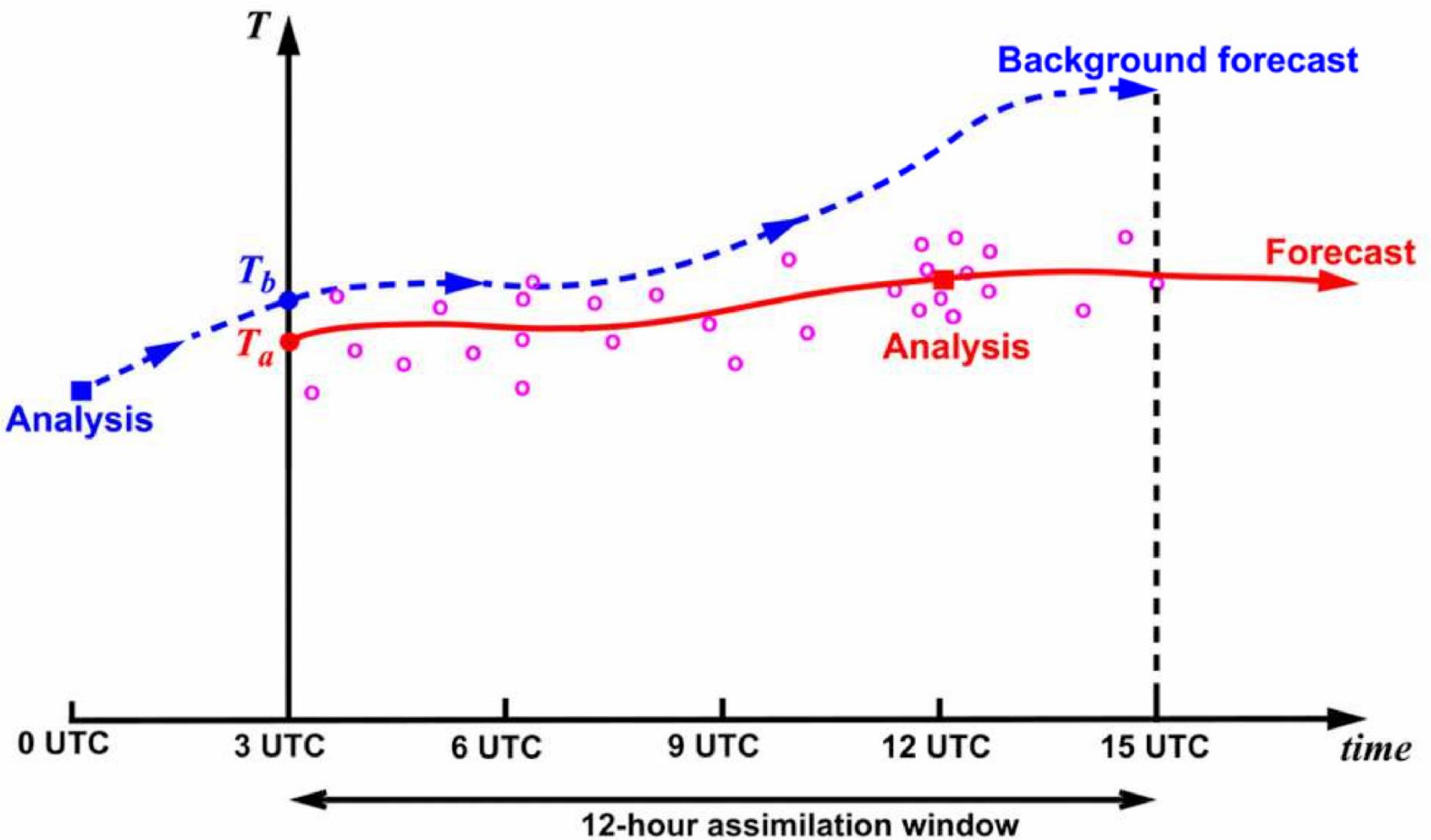
Ideal
Centered
sde_MC
sde_C
sdd_MC
sdd_C
sde_P
sde_MP
sdd_P
sdd_MP
sdd_MC^2 / sde_MC^2 = 1
sdd_C^2 / sde_C^2 = 1
sdd_MP^2 / sde_MP^2 = 1/2 = 0.5
sdd_P^2 / sde_P^2 = 2/3 = 0.667

Conclusion

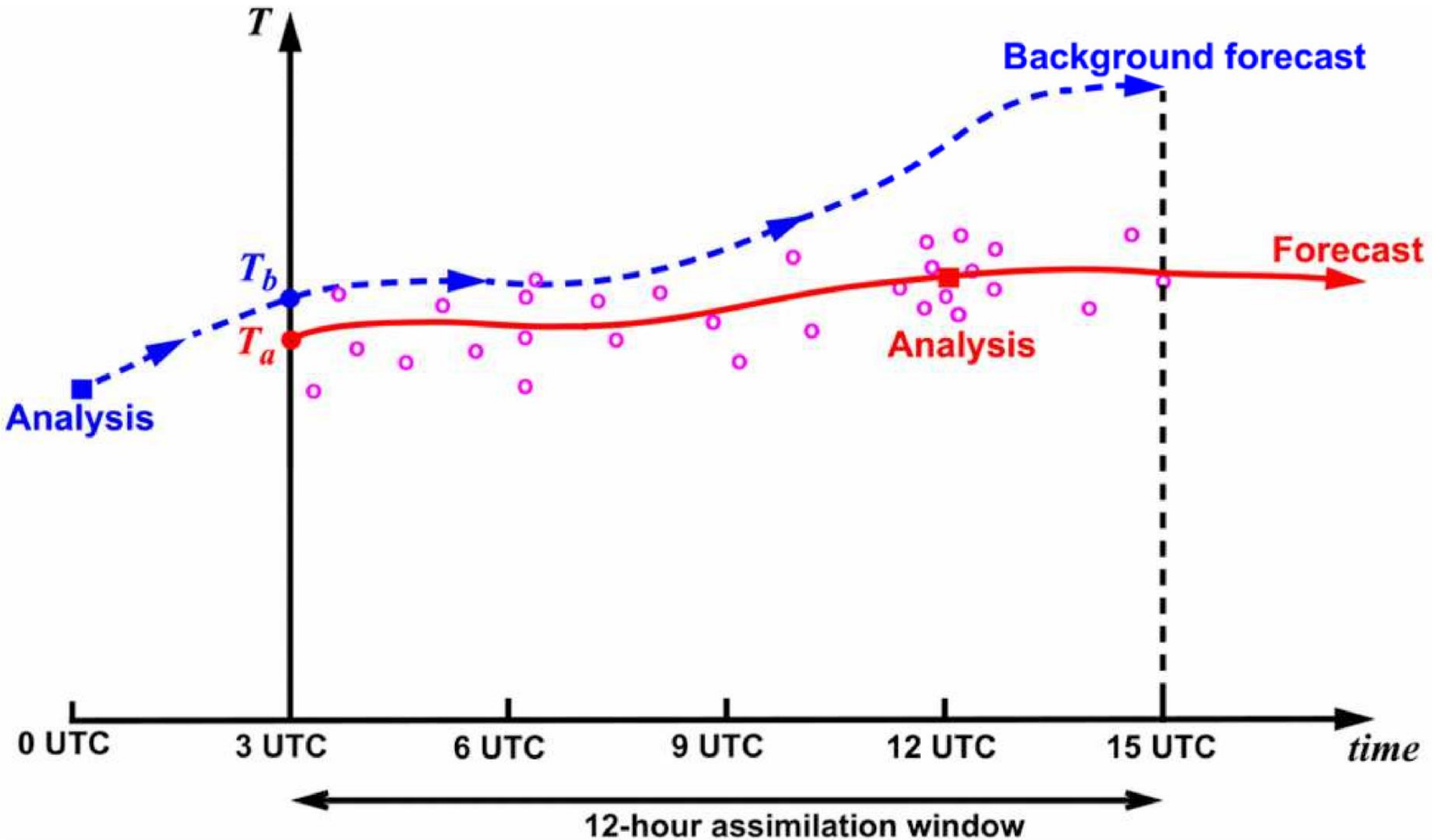
The currently used practice of generating ensemble members at the initial time should be replaced by a method that creates equally likely ensemble members with the same quality as the control analysis.

Proposed
Joint DA-EPS scheme

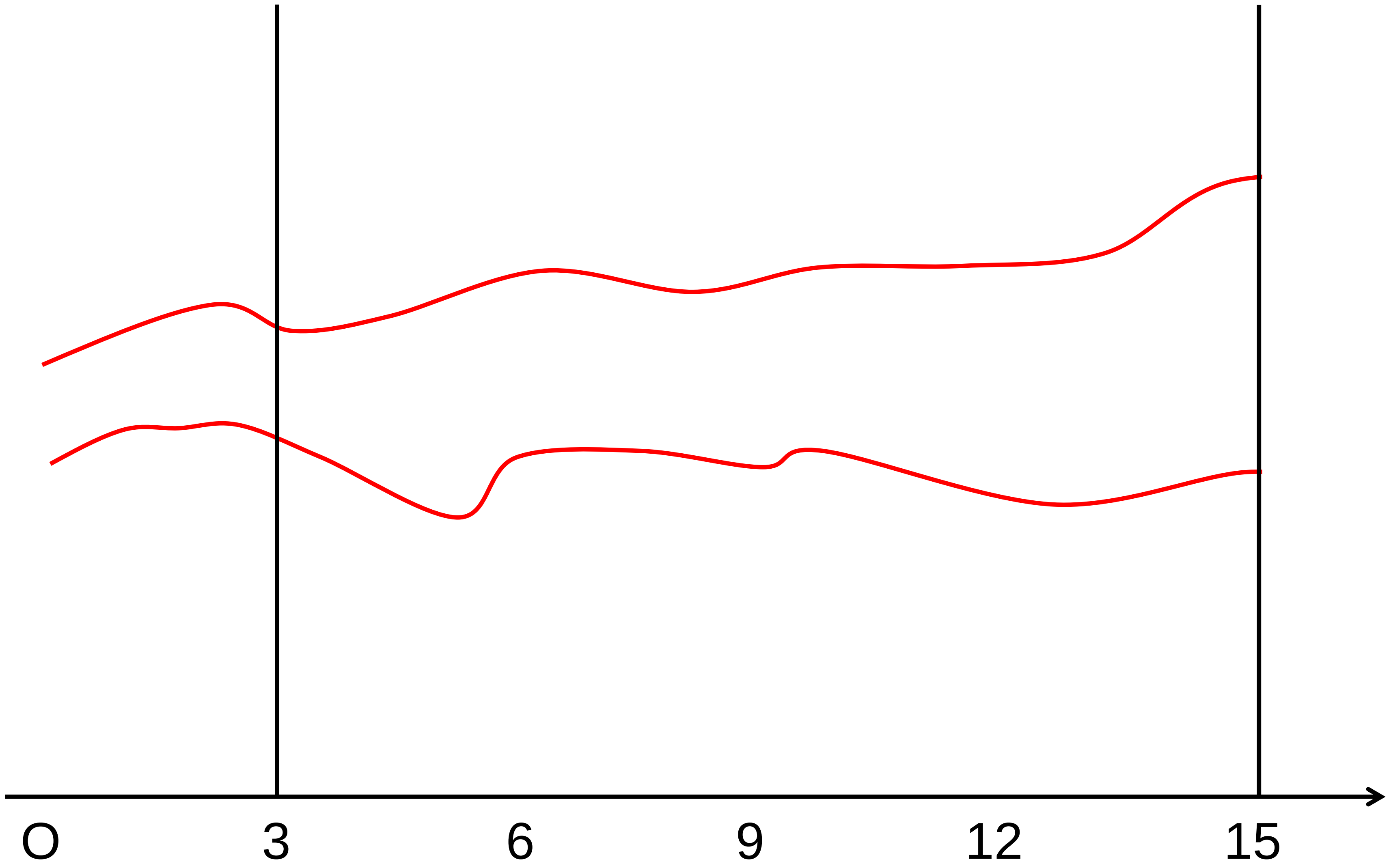
Traditional 4DVAR



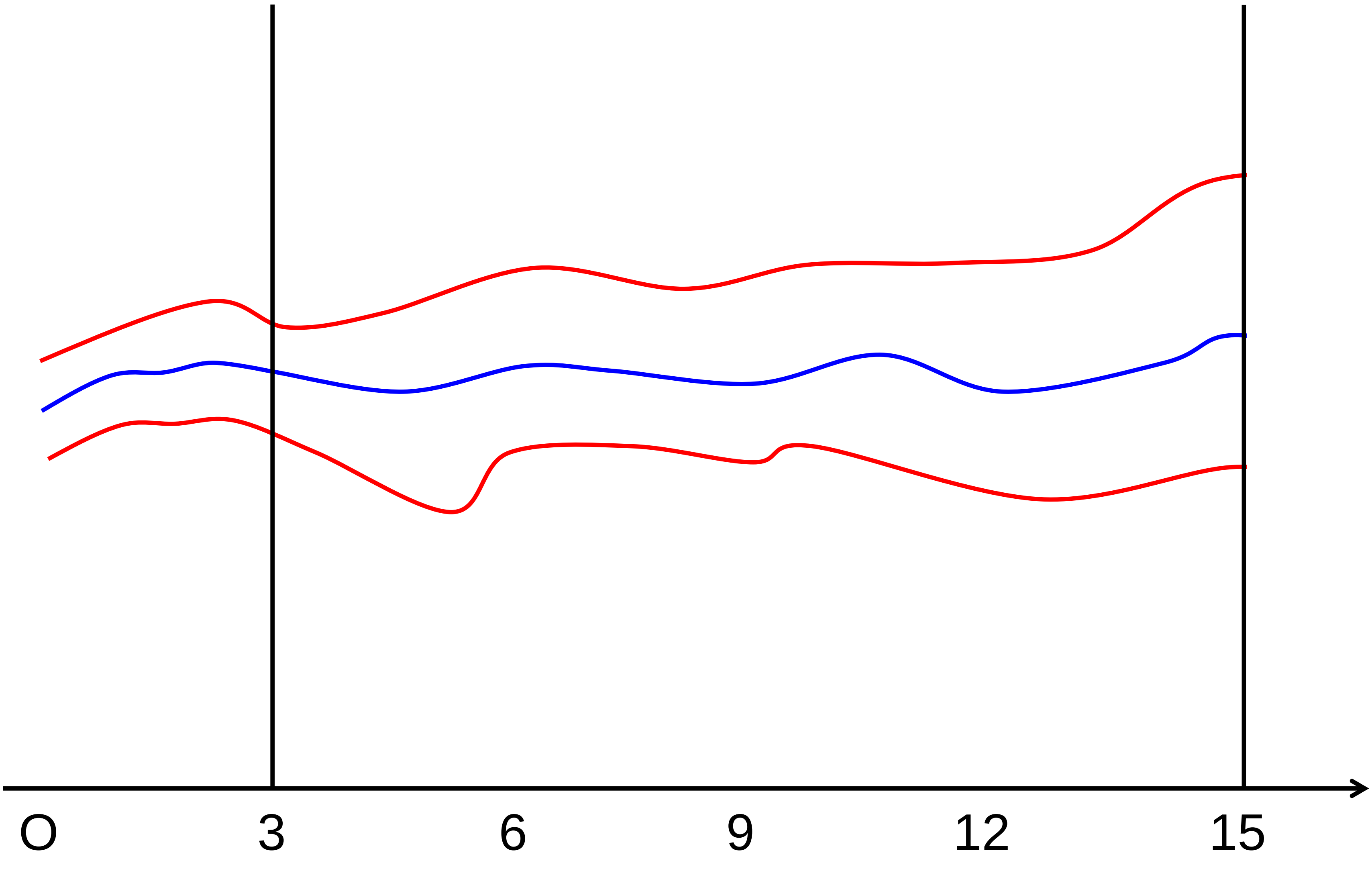
$$J(\mathbf{x}_o) = J_b + J_o = \frac{1}{2}\left(\mathbf{x}_o - \mathbf{x}_\mathbf{o}^\mathbf{b}\right)^T \mathbf{B}^{-1}\left(\mathbf{x}_o - \mathbf{x}_\mathbf{o}^\mathbf{b}\right) + \sum_{i=1}^{I}\frac{1}{2}\left(H_\mathbf{i}\,\mathbf{x}_i - \mathbf{y}_\mathbf{i}\right)^T \mathbf{R}_i^{-1}\left(H_\mathbf{i}\,\mathbf{x}_i - \mathbf{y}_\mathbf{i}\right)$$



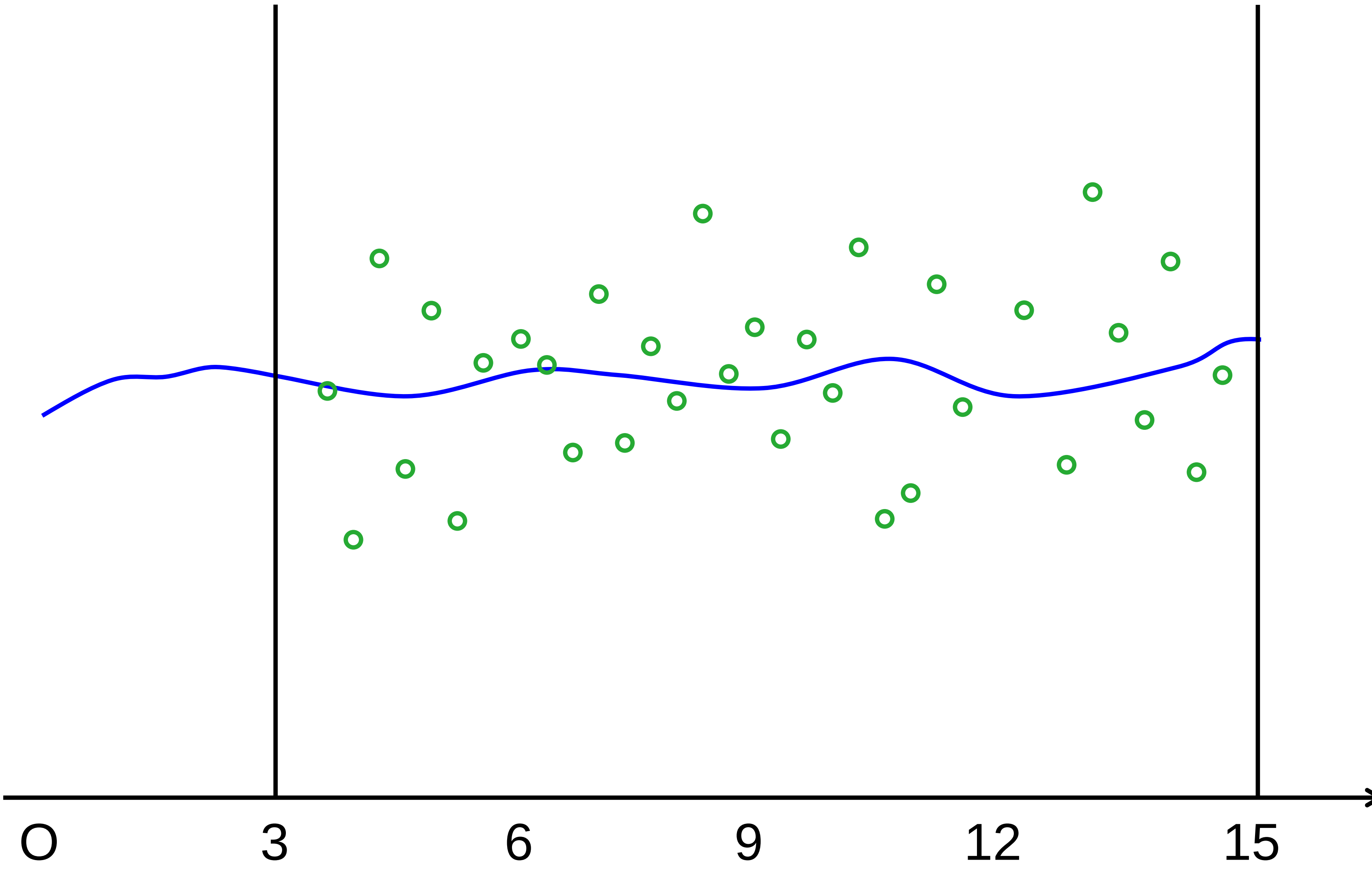
Ensemble Mean 4DVAR



Ensemble Mean 4DVAR



Ensemble Mean 4DVAR



Quantity of available data

$$\text{Number of Observations} = \text{M} \sim 10^5 - 10^6$$

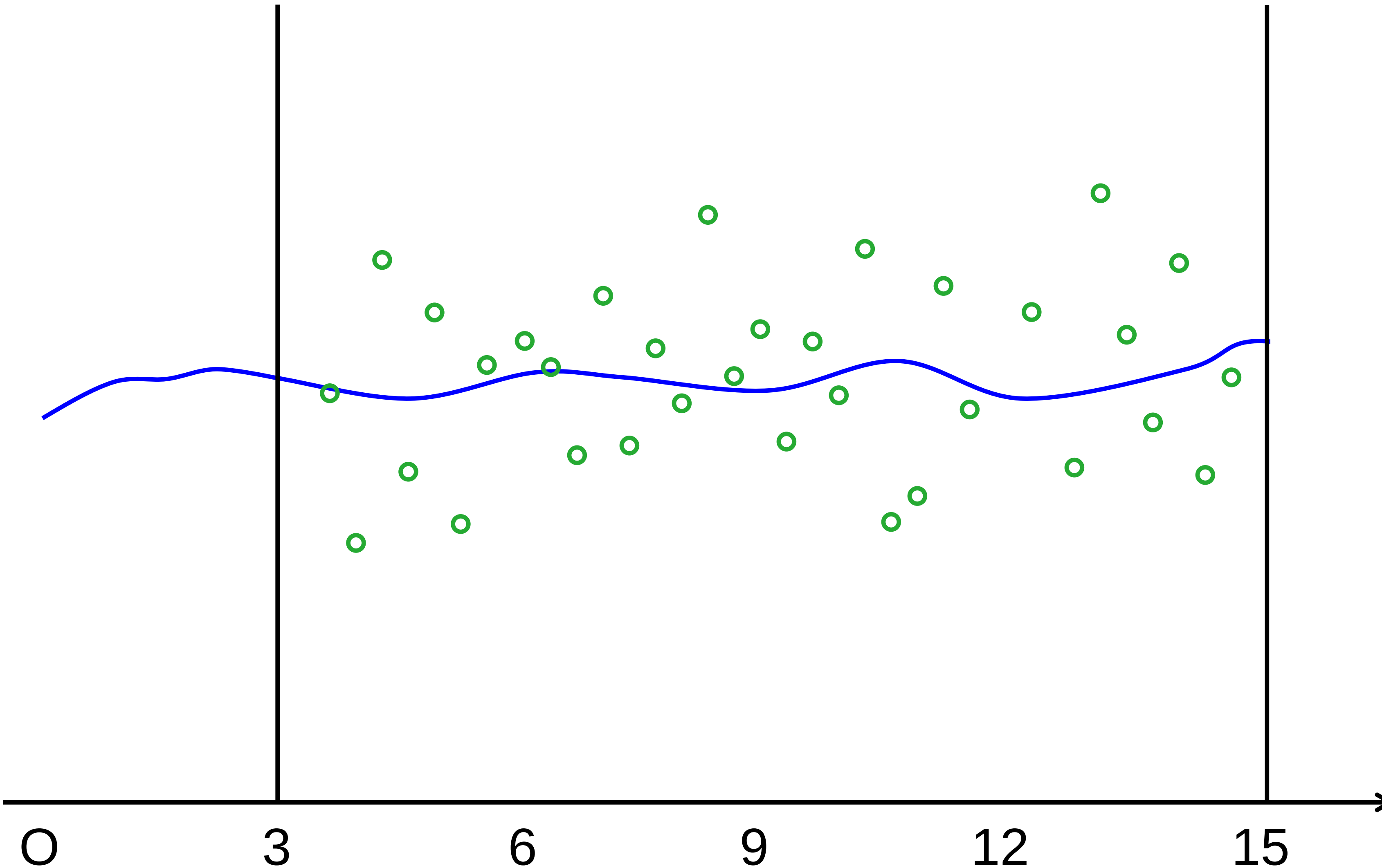
$$\text{Dimension of State Vector} = \text{N} \sim 10^7 - 10^9$$

M << N

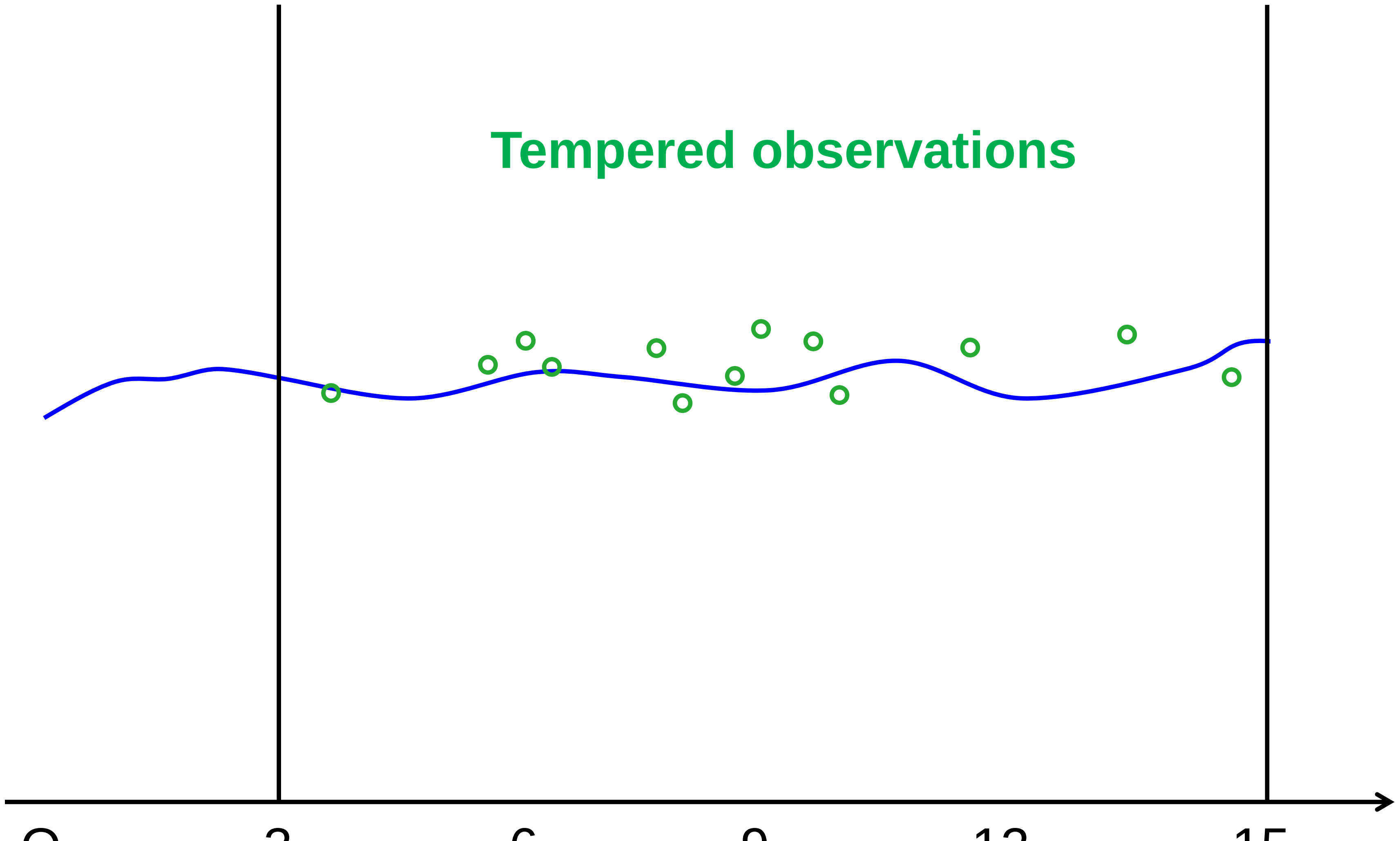
⬇

Only the largest scales are really defined by the available data

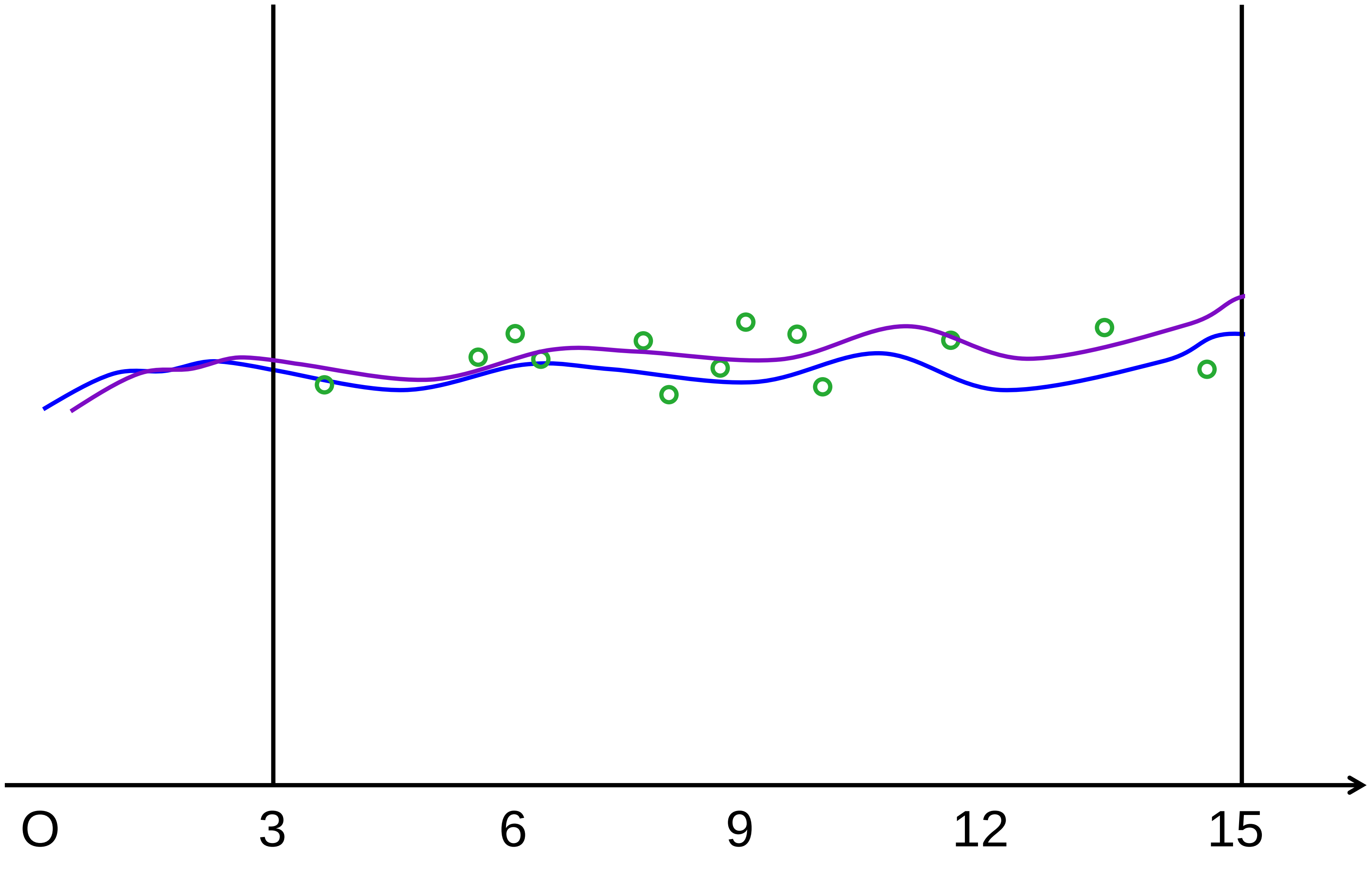
Ensemble Mean 4DVAR



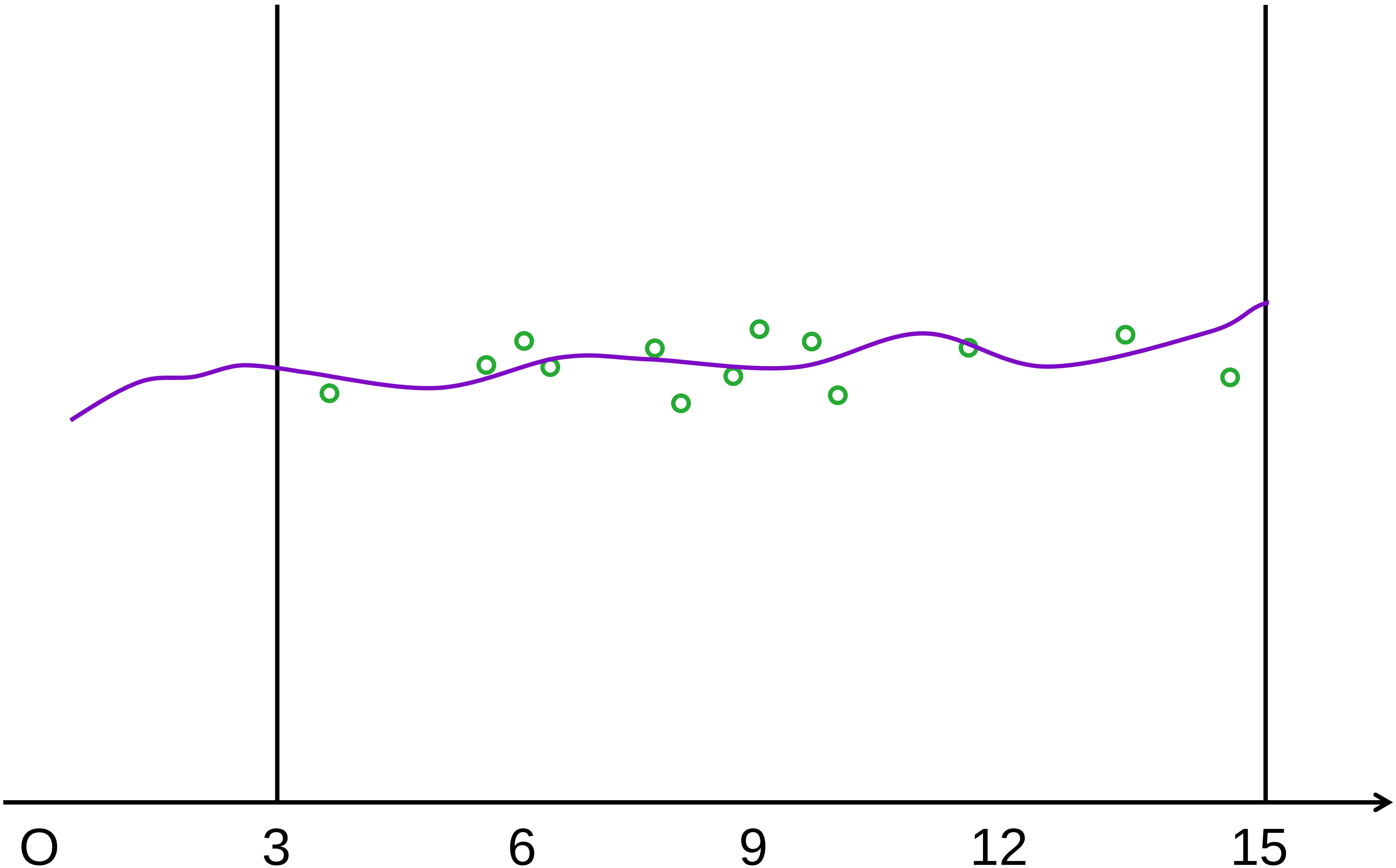
Ensemble Mean 4DVAR



Ensemble Mean 4DVAR



Ensemble Mean 4DVAR

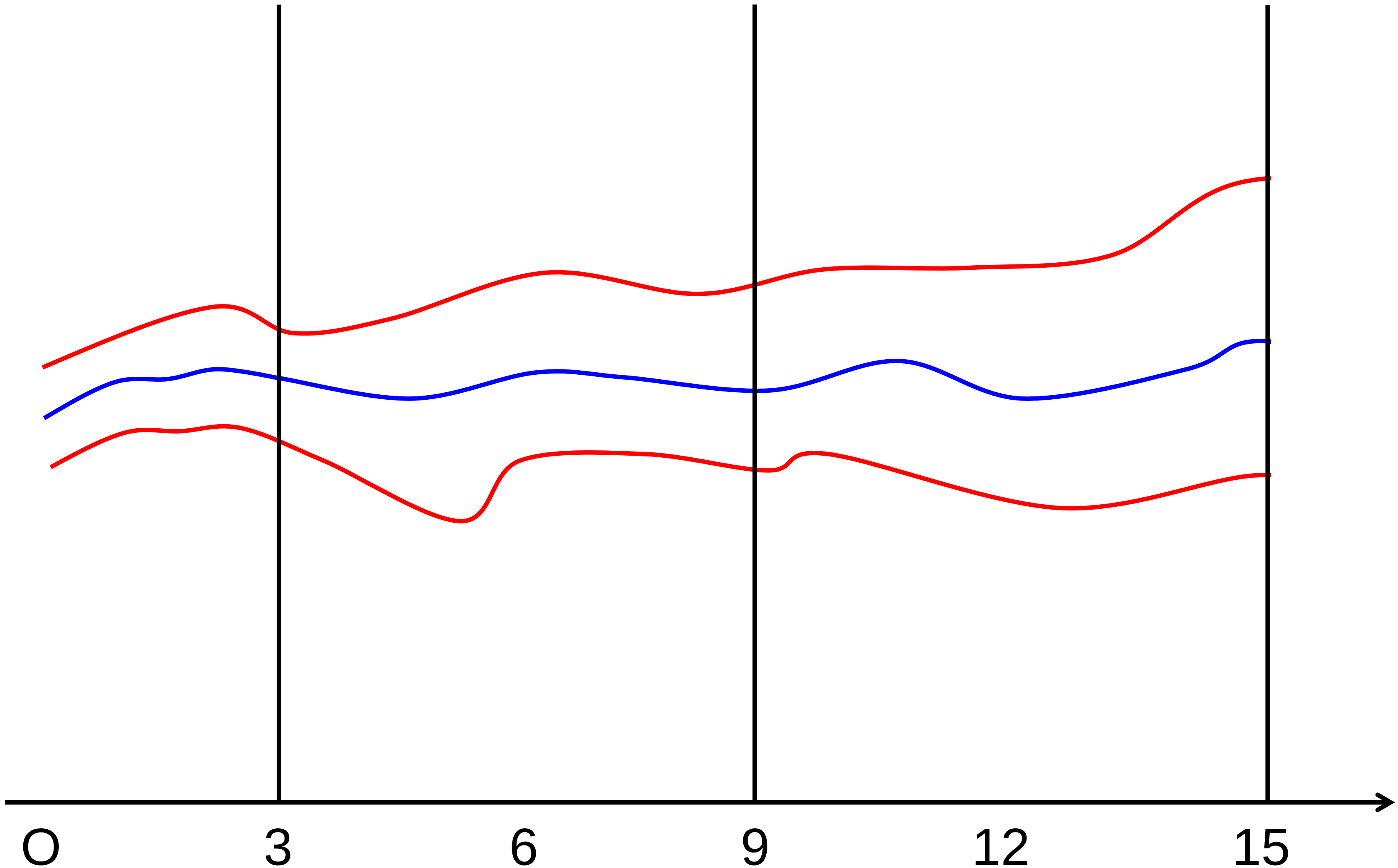


Ensemble Mean 4DVAR

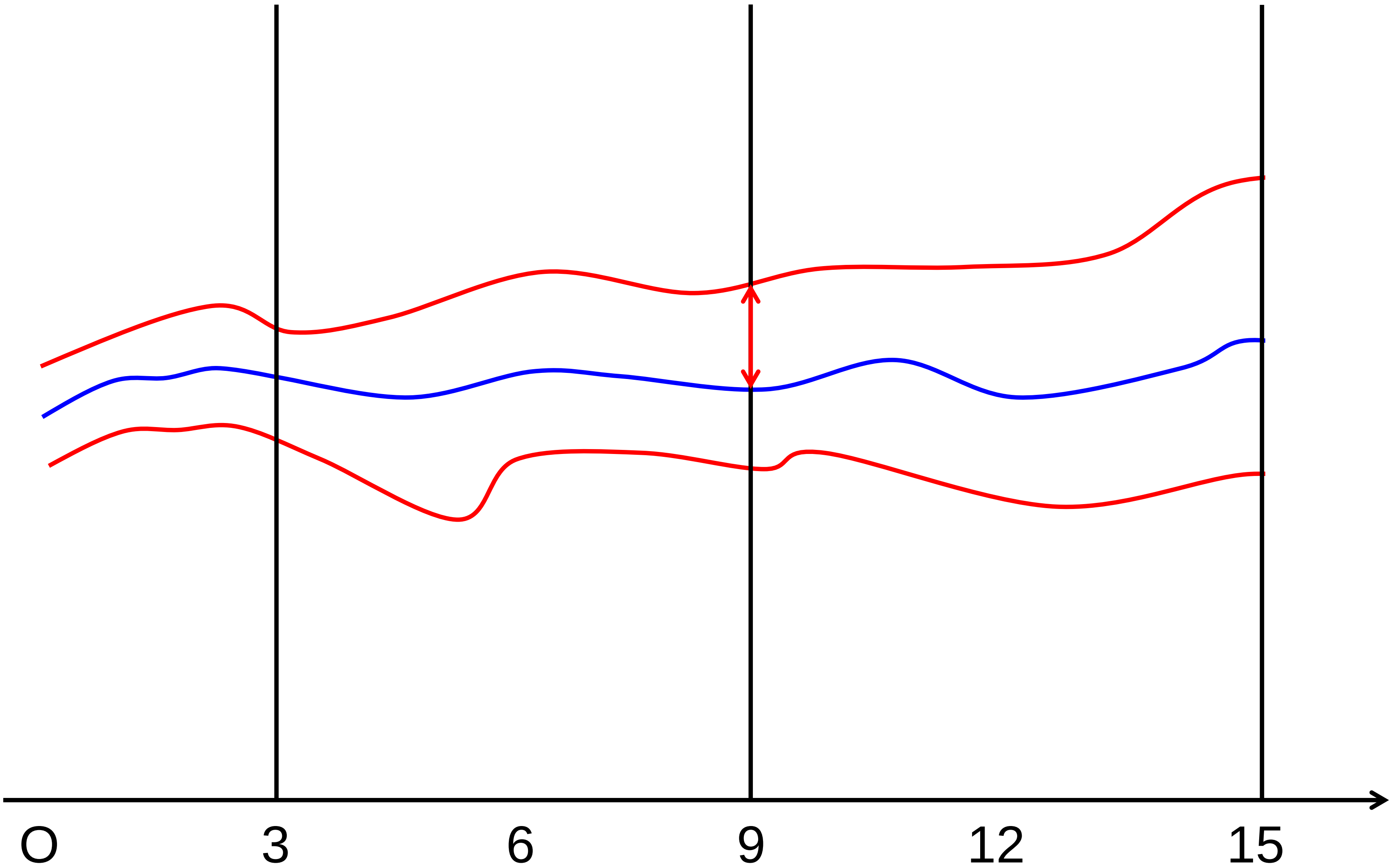
$$J = J_b + J_o = \sum_{i=1}^{I} \left[\frac{1}{2} \left(\mathbf{x}_i - \mathbf{x}_\mathbf{i}^\mathbf{b} \right)^T \mathbf{B}^{-1} \left(\mathbf{x}_i - \mathbf{x}_\mathbf{i}^\mathbf{b} \right) + \frac{1}{2} \left(\mathbf{H}_\mathbf{i} \mathbf{x}_i - \mathbf{y}_\mathbf{i} \right)^T \mathbf{R}_i^{-1} \left(\mathbf{H}_\mathbf{i} \mathbf{x}_i - \mathbf{y}_\mathbf{i} \right) \right]$$

1. The $\mathbf{x}_\mathbf{i}$ represents the **ensemble mean** instead of a control member
2. The $\mathbf{y}_\mathbf{i}$ represents **tempered observations** instead of raw observations
3. All time points $\mathbf{t}_\mathbf{i}$ are used instead of only the initial time point $\mathbf{t}_\mathbf{o}$
4. No need for TL and AD models
5. The MEAN is estimated instead of the MODE

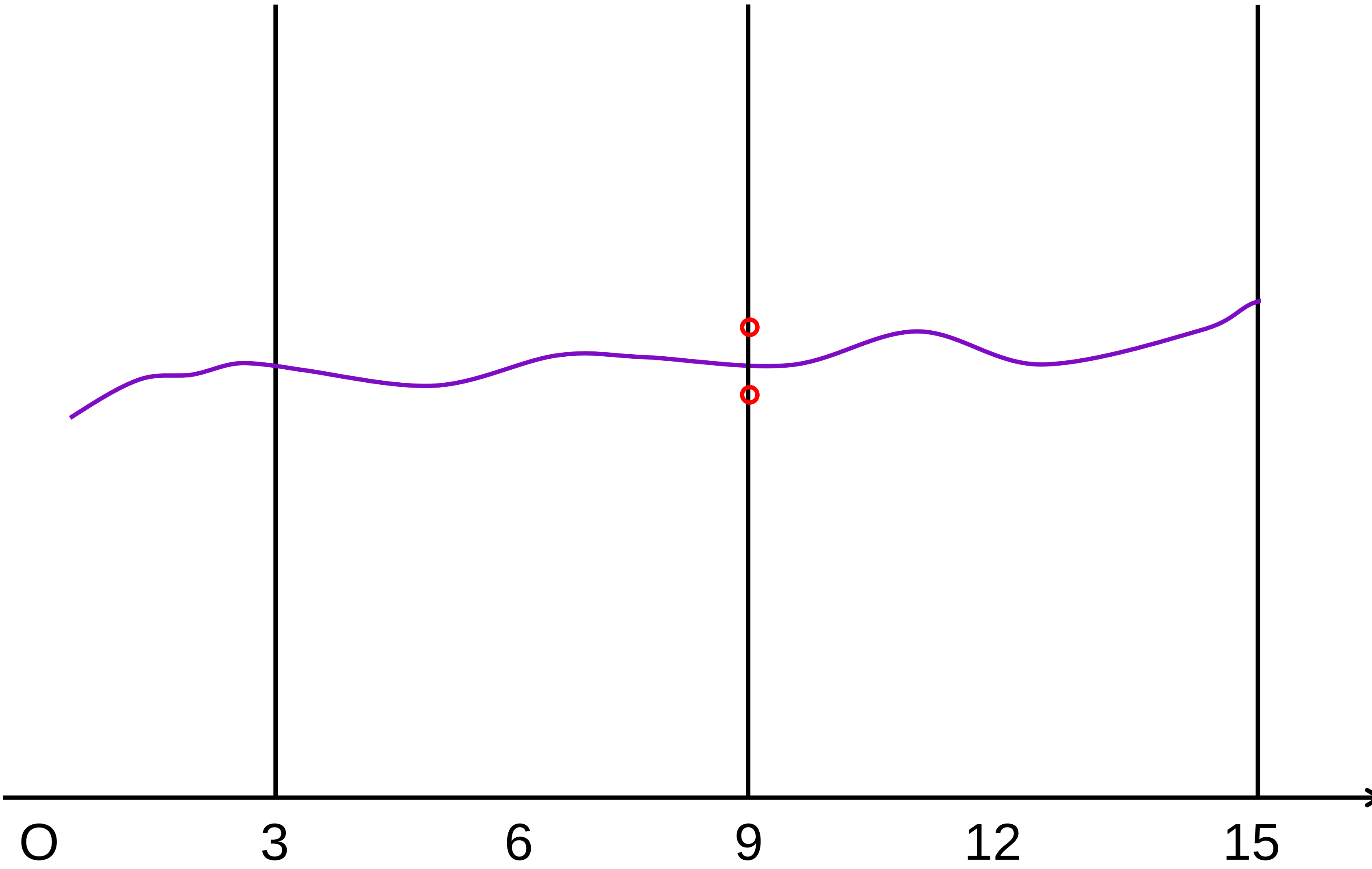
Ensemble Members



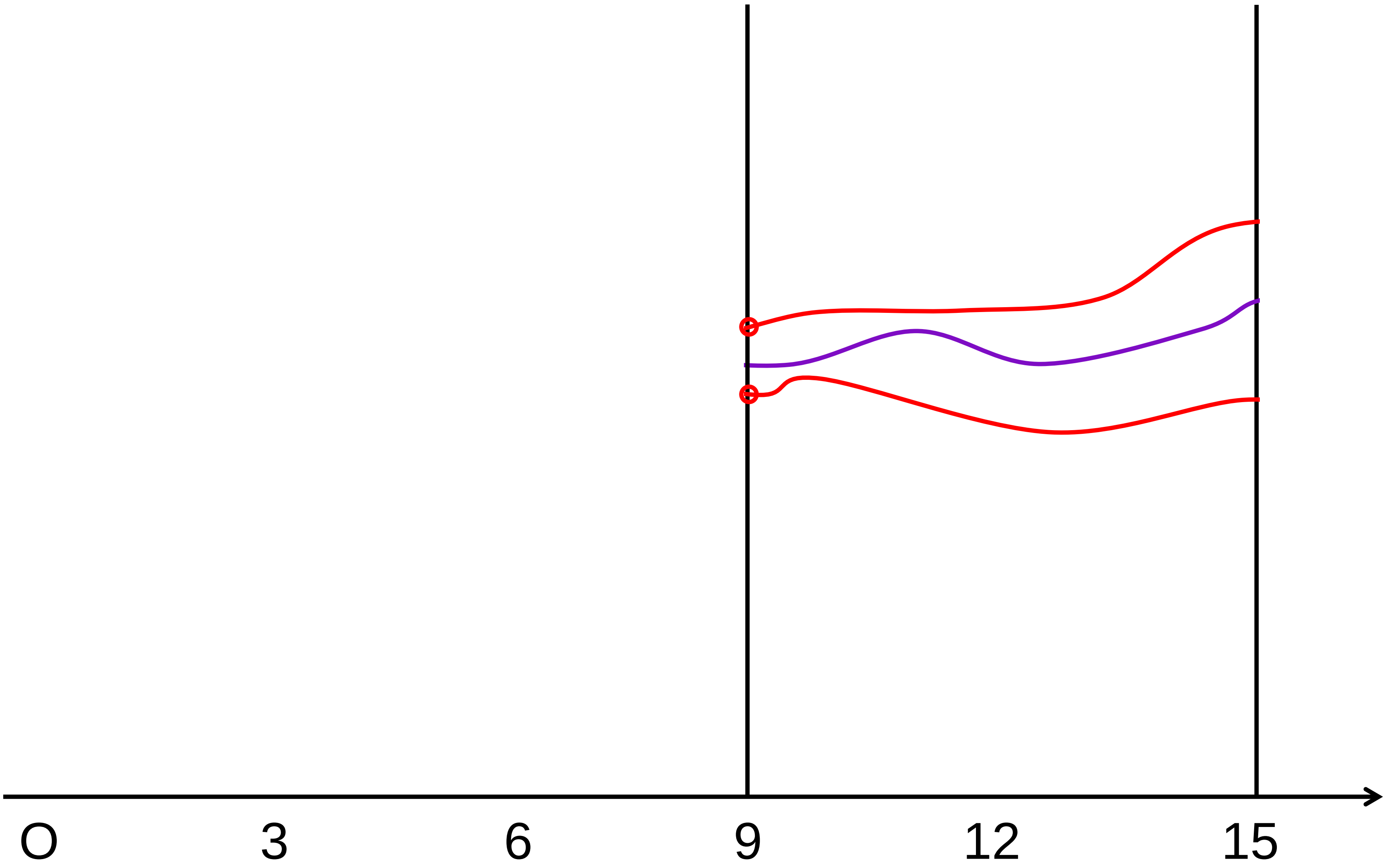
Ensemble Members



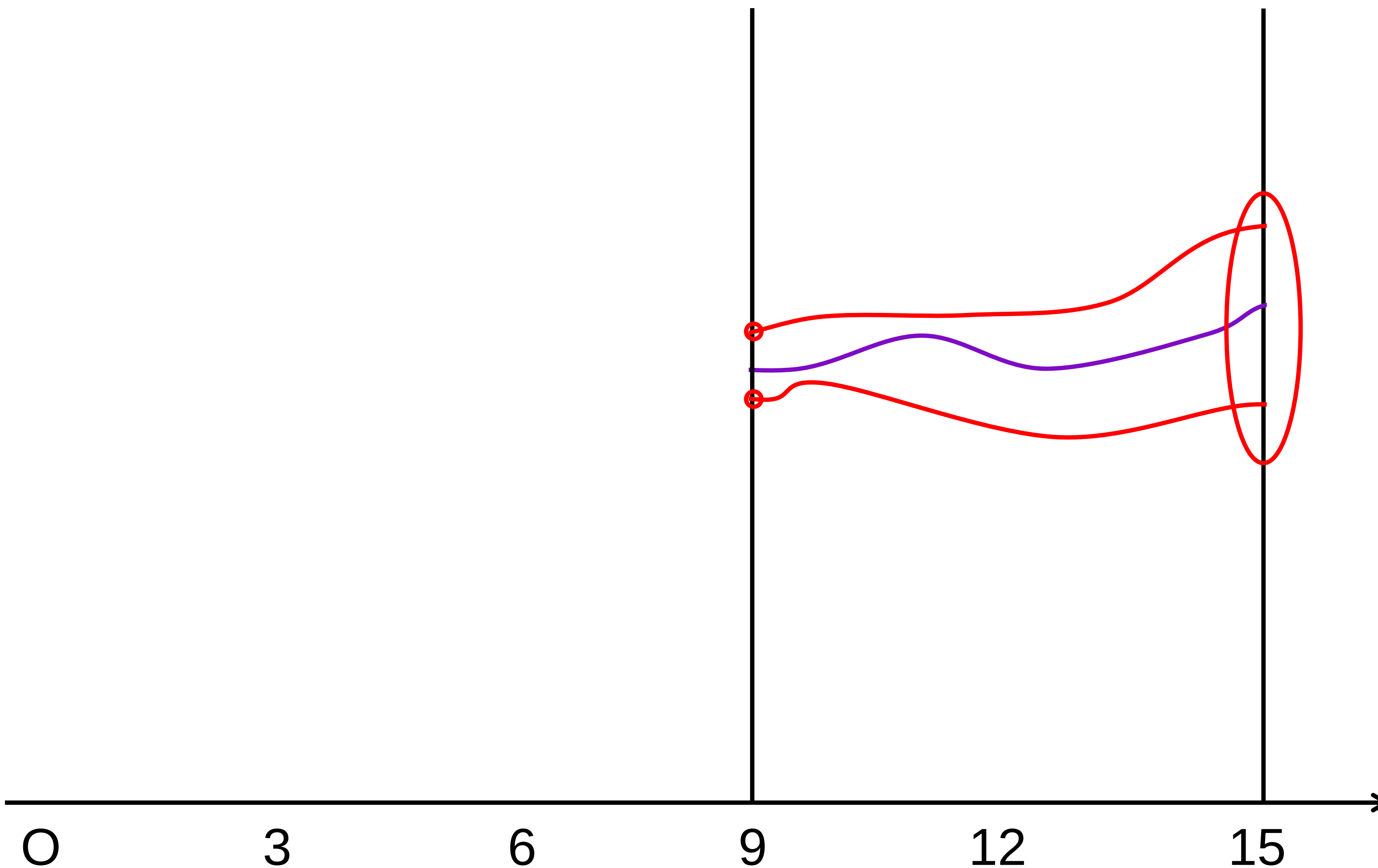
Ensemble Members



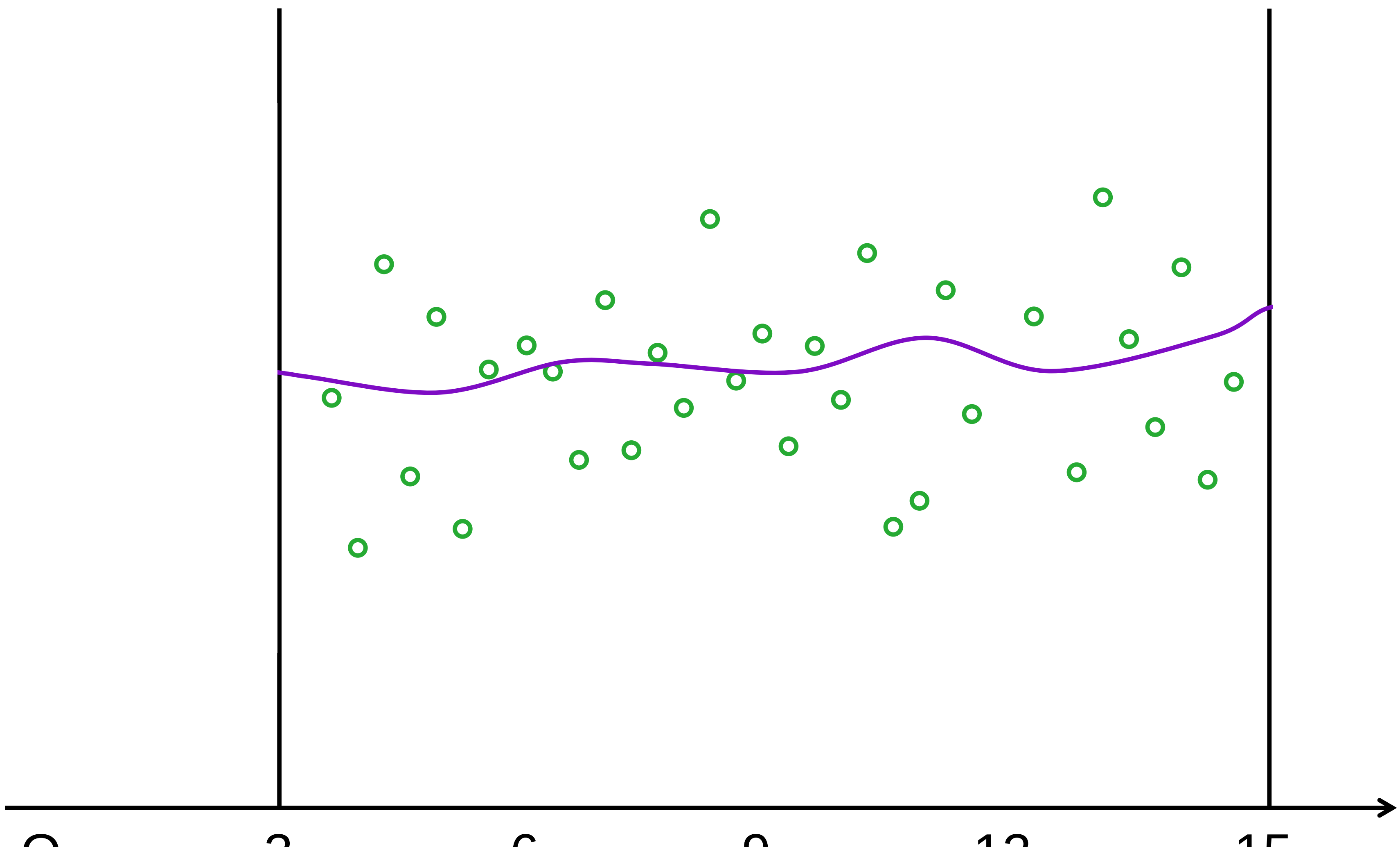
Ensemble Members



Spread



Skill



Quantity of available data

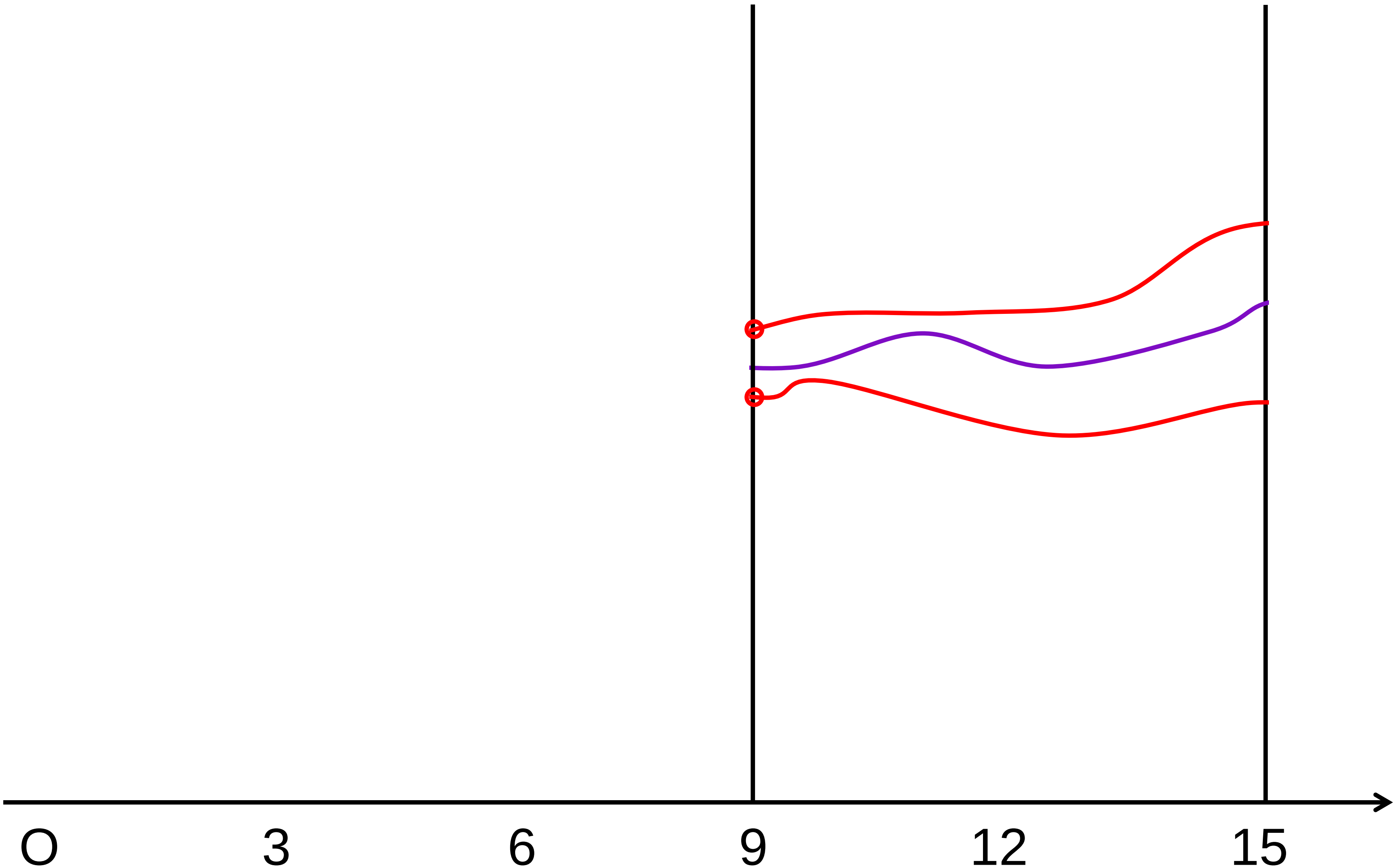
$$\text{Number of Observations} = M \sim 10^{20}$$

$$\text{Dimension of State Vector} = N \sim 10^{15}$$

M >> N

However

Only the largest scales are predictable at the end of the window


O
3
6
9
12
15