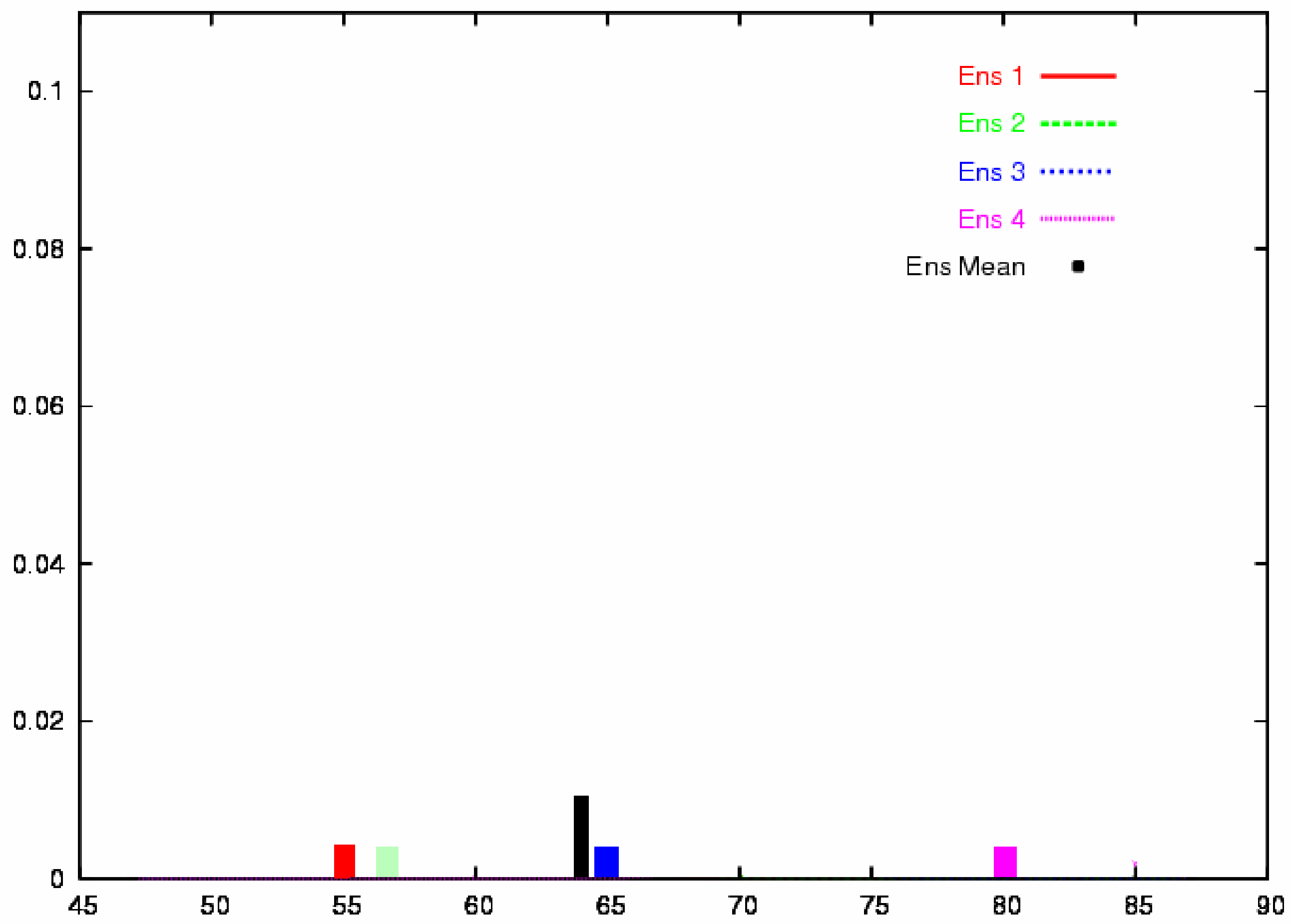
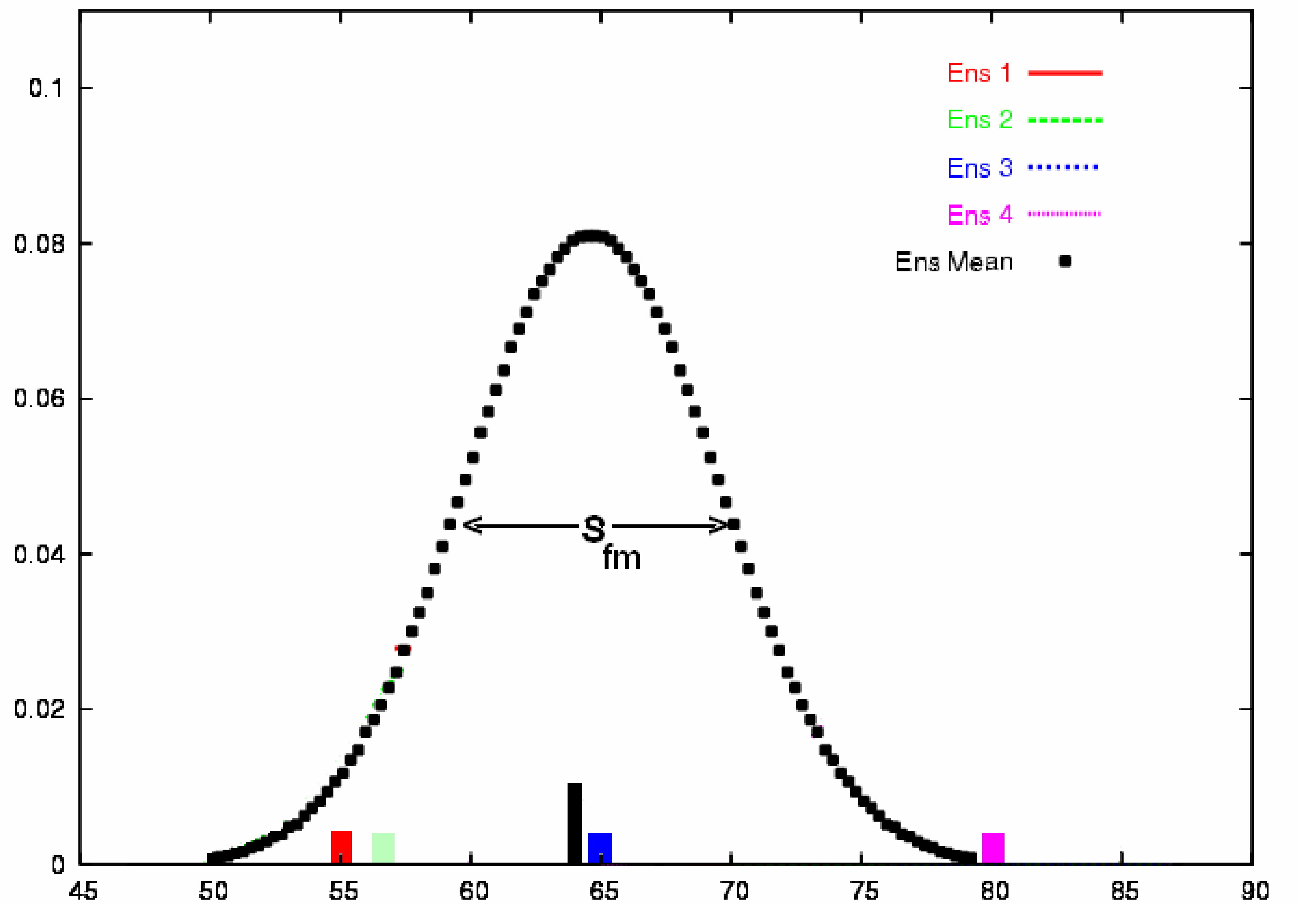
Ensemble Processing

David Unger
Climate Prediction Center

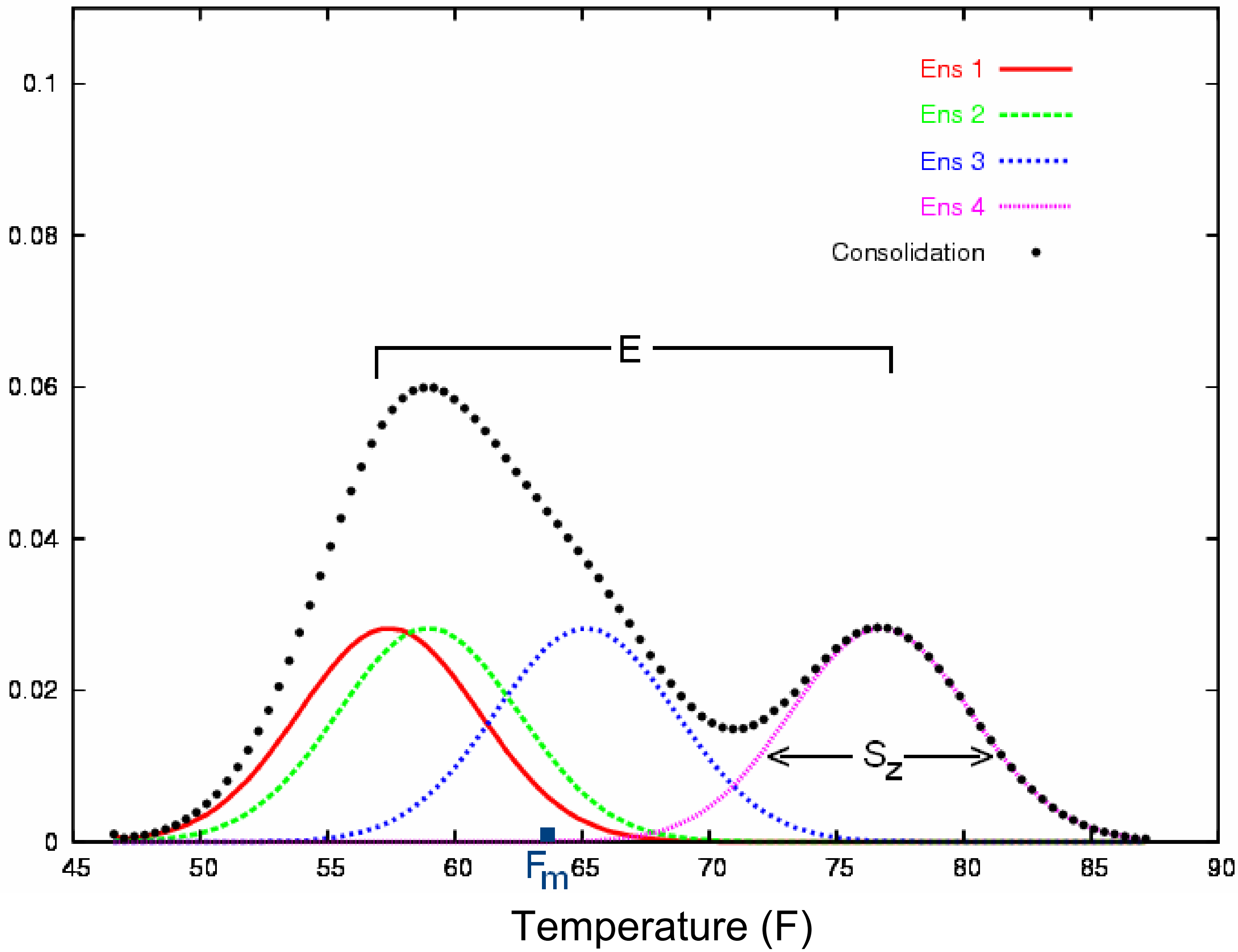
Schematic illustration



Temperature


Ens 1
Ens 2
Ens 3
Ens 4
Ens Mean
S
fm
0
0.02
0.04
0.06
0.08
0.1
45
50
55
60
65
70
75
80
85
90

Schematic example



DEFINITIONS:

$\mathrm{F} = \mathrm{Forecast}$ $\quad$ $\mathrm{C} = \mathrm{Climatology}$ $\quad$ $b = \mathrm{Observation}$

$S =$ Standard deviation of forecast errors

$$S^2 = \overline{(\mathrm{F} - b)}^2$$

$\sigma_{\mathrm{f}} =$ Standard deviation of forecasts

$$\sigma_{\mathrm{f}}^2 = \overline{(\mathrm{F} - \mathrm{C})}^2$$

$$\sigma_{\mathrm{b}}^2 = \overline{(b - \mathrm{C})}^2$$

$E =$ Ensemble Spread

$$E^2 = \sum \frac{(\mathrm{F} - \mathrm{F}_{\mathrm{m}})^2}{\mathrm{N}}$$

Regression=Bias Correction and Standardization

$$Z = \frac{(F - \bar{F})}{(\quad \sigma_{\mathrm{F}})}$$

$$\hat{Z} = RZ$$

$$\hat{F} = \hat{Z}\,\sigma_{\mathrm{C}} + C$$

Regression relationships

$$\sigma_{\mathrm{fm}} = R_{\mathrm{fm}} \sigma_{\mathrm{c}}$$

Standard deviation of forecasts

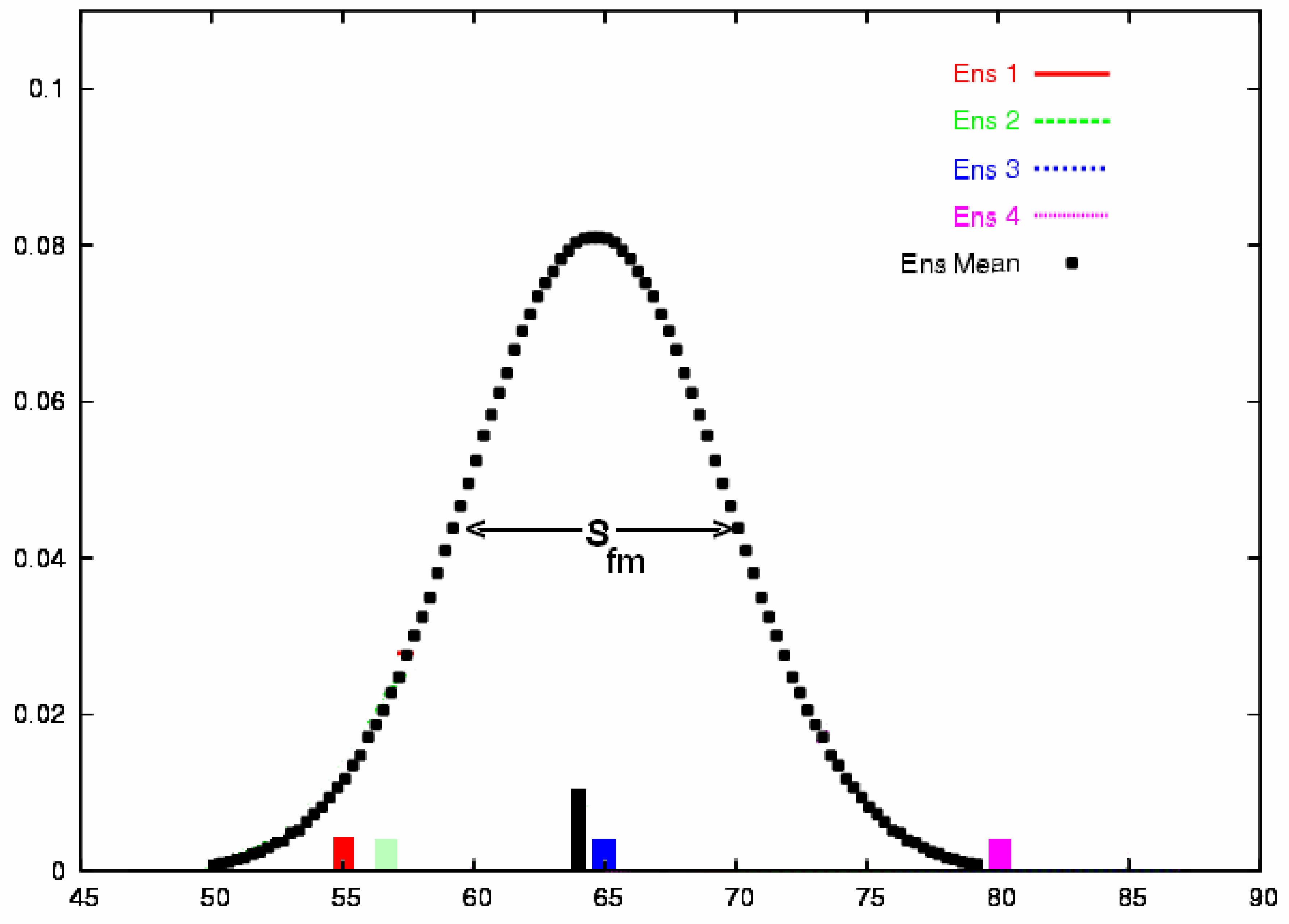
$$S_{\mathrm{fm}} = \sigma_{\mathrm{c}} \sqrt{1 - R_{\mathrm{fm}}^2}$$

Standard deviation of forecast errors

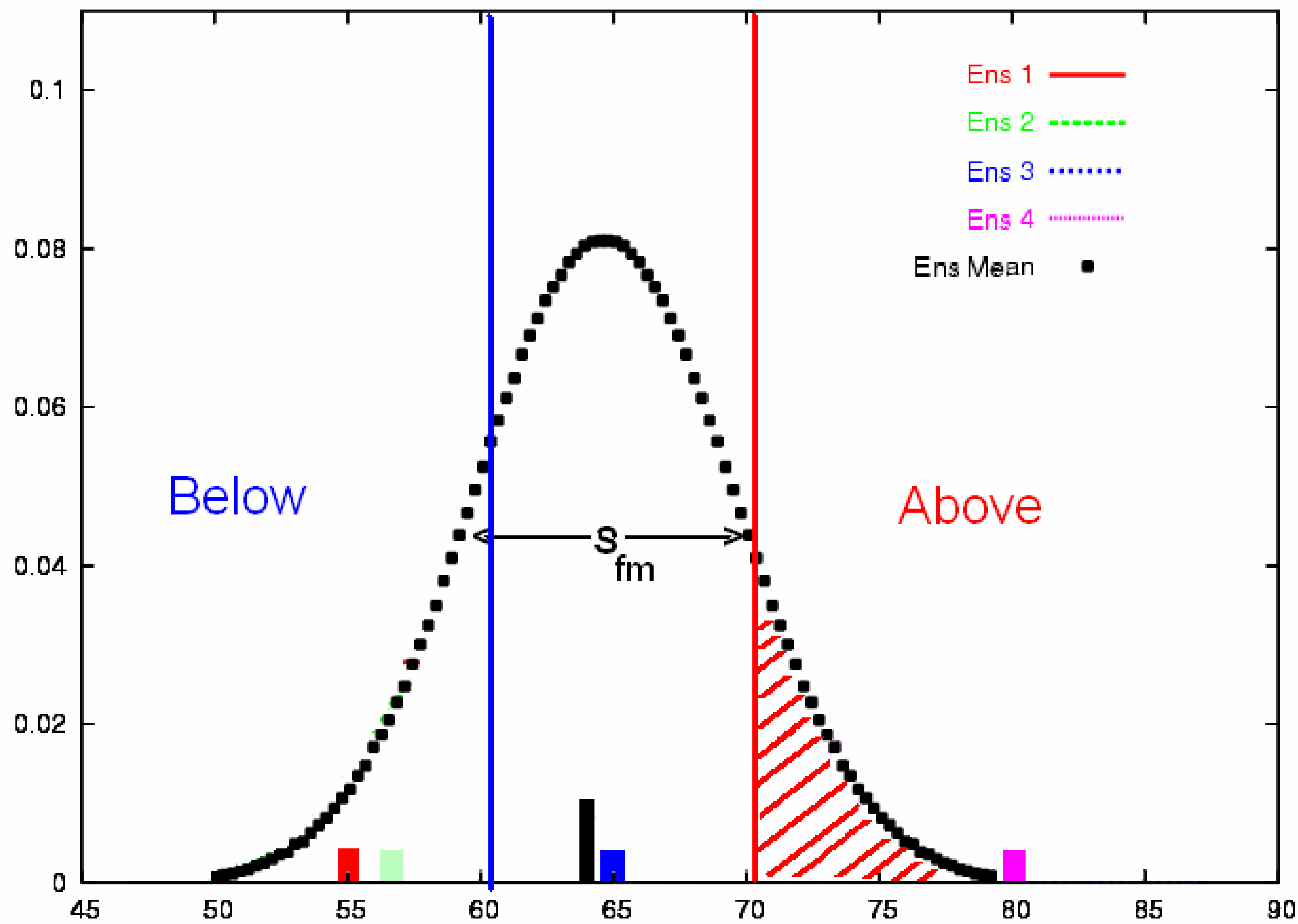
$$\sigma_{\mathrm{fm}}^2 + S_{\mathrm{fm}}^2 = R_{\mathrm{fm}}^2 \sigma_{\mathrm{fm}}^2 + \sigma_{\mathrm{c}}^2 \left(1 - R_{\mathrm{fm}}^2\right)$$

$$\sigma_{\mathrm{fm}}^2 + S_{\mathrm{fm}}^2 = \sigma_{\mathrm{c}}^2$$

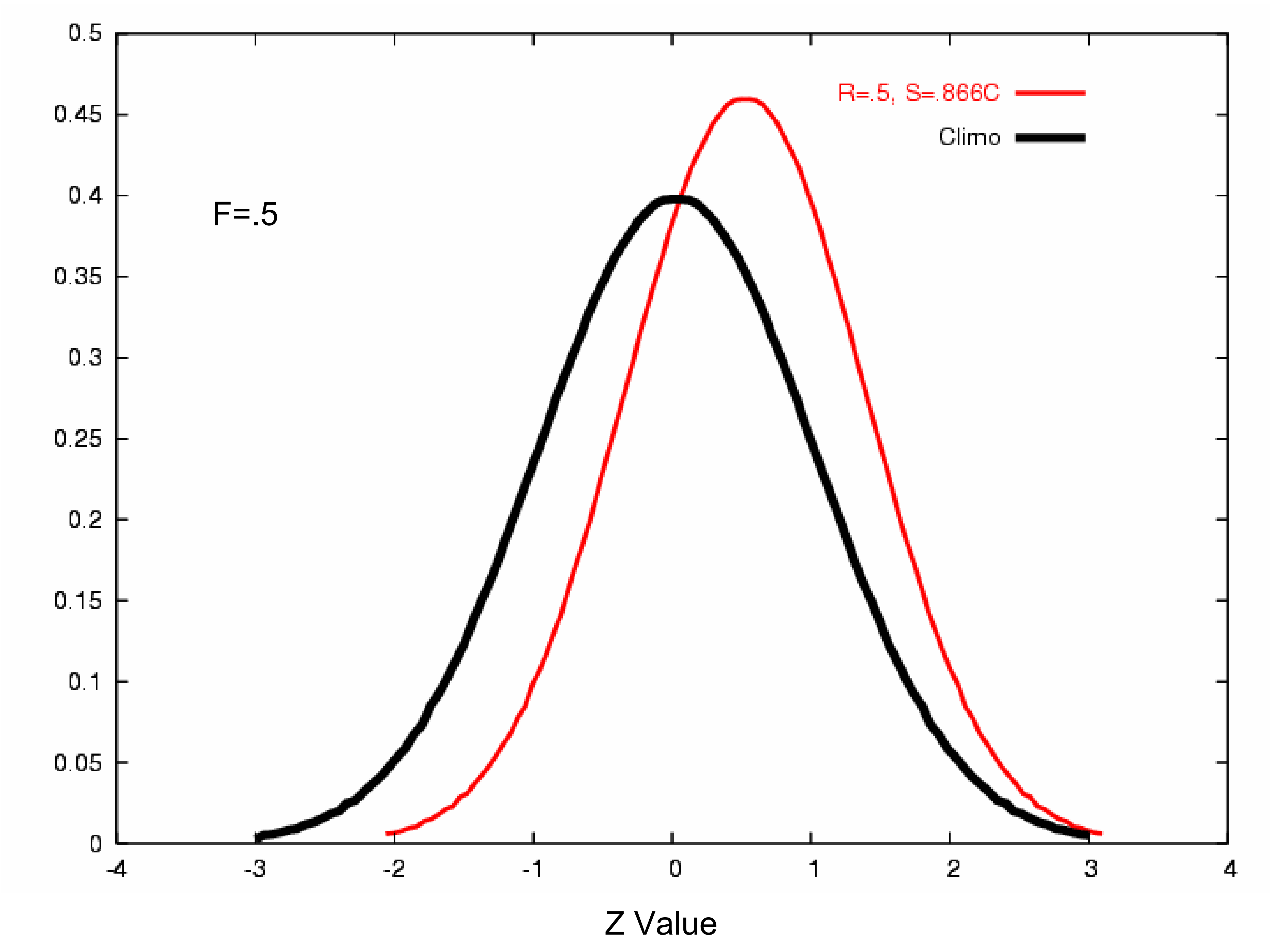
Forecast and error variance
sum to climatological variance



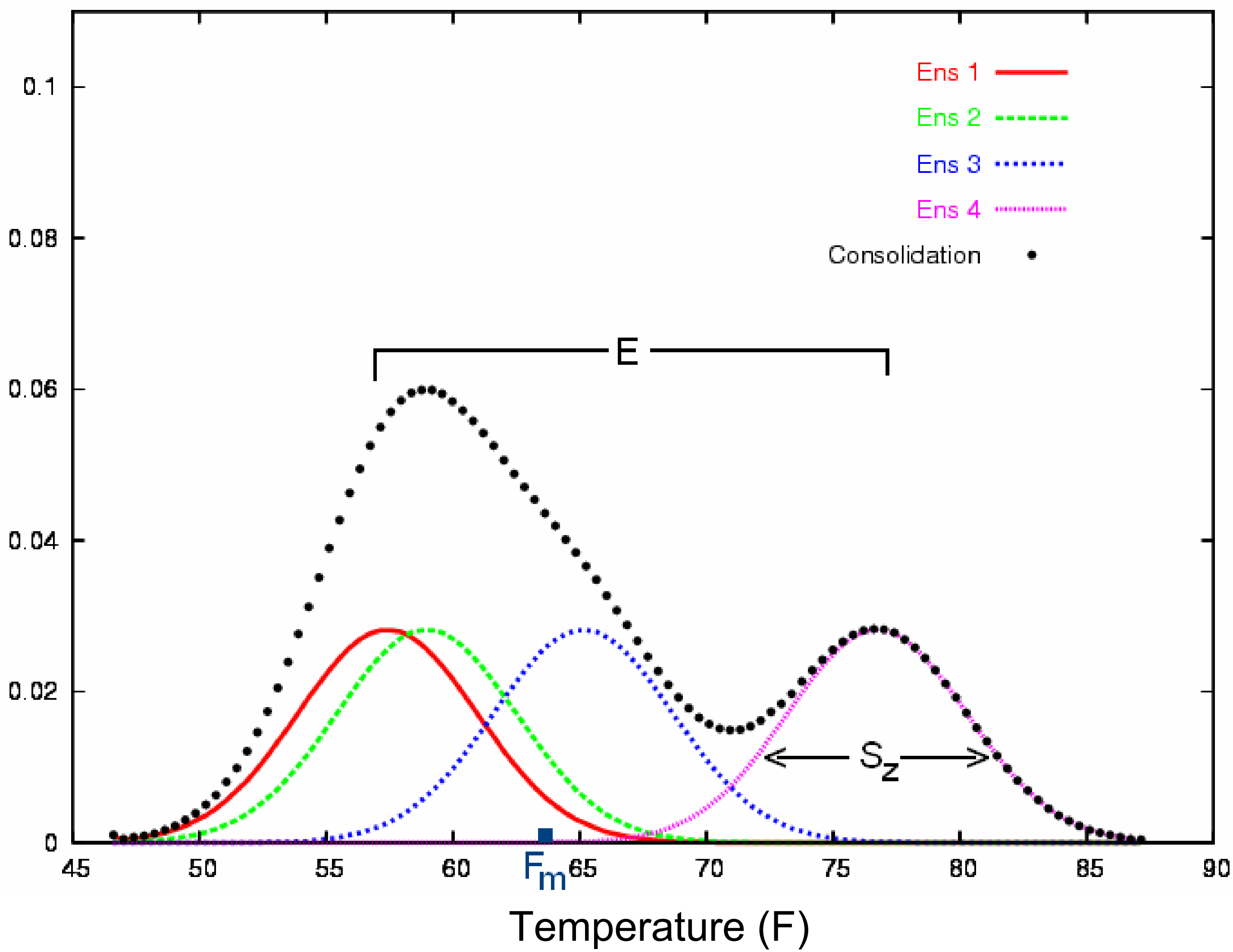
Ens 1
Ens 2
Ens 3
Ens 4
Ens Mean
S
fm
0.1
0.08
0.06
0.04
0.02
0
45
50
55
60
65
70
75
80
85
90



Ens 1
Ens 2
Ens 3
Ens 4
Ens Mean
Below
Above
s_{fm}
0.1
0.08
0.06
0.04
0.02
0
45
50
55
60
65
70
75
80
85
90


R=.5, S=.866C
Climo
F=.5
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
-4
-3
-2
-1
0
1
2
3
4
Z Value

Schematic example



Kernel Math

- Ensemble Spread, E, and S_z, are constrained.

  - Depend on R and R_m

  - Also depends on the standard deviation of f and f_m .

Variance Relationships

- Explained Variance = $R^2\sigma_c^2$
- Unexplained Variance S_{fm}^2 divides into 2 parts:
  1) Ensemble Spread (E^2) (Variable)
  2) Residual Variance (Fixed) (S_z^2)

$S_z^2 \leq S_{fm}^2$: $S_{fm}^2 = S_z^2 + E^2$

Kernel mathematics

$$E^2 = \sum (f - fm)^2$$

$S_z =$ Conditional standard dev.

$$S_f^2 = S_{fm}^2 + E^2$$

$$\sigma_{Total}^2 = \sigma_c^2 = \sigma_f^2 + S_z^2$$

$$S_z^2 + E^2 = S_{fm}^2$$

$$S_z^2 = S_{fm}^2 - E^2$$

Optimum Spread

- Regression gives one candidate for an optimum spread value, E^2
- E^2 depends on the skill difference between the individual ensemble members and the ensemble mean.
- We then can calculate the unexplained variance not accounted for by the spread.

Optimum ensemble spread?

$$S_{\mathrm{f}}^2 = S_{\mathrm{fm}}^2 + E^2$$
$$E^2 = S_{\mathrm{f}}^2 - S_{\mathrm{fm}}^2$$

$$S_{\mathrm{f}} = \sigma_{\mathrm{c}} \sqrt{(1 - R_{\mathrm{f}}^2)}$$

$$S_{\mathrm{fm}} = \sigma_{\mathrm{c}} \sqrt{(1 - R_{\mathrm{fm}}^2)}$$

$$E^2 = \sigma_{\mathrm{c}}^2 (R_{\mathrm{fm}}^2 - R_{\mathrm{f}}^2)$$

Minimum Forecast Error Variance

$$E^2 = \sigma_c^2 (R_{fm}^2 - R_f^2)$$

$$S_z^2 = S_{fm}^2 - E^2$$

$$S_z^2 = \sigma_c^2 (1 - R_{fm}^2) - \sigma_c^2 (R_{fm}^2 - R_f^2)$$

$$S_z^2 = \sigma_c^2 (1 - 2\, R_{fm}^2 + R_f^2)$$

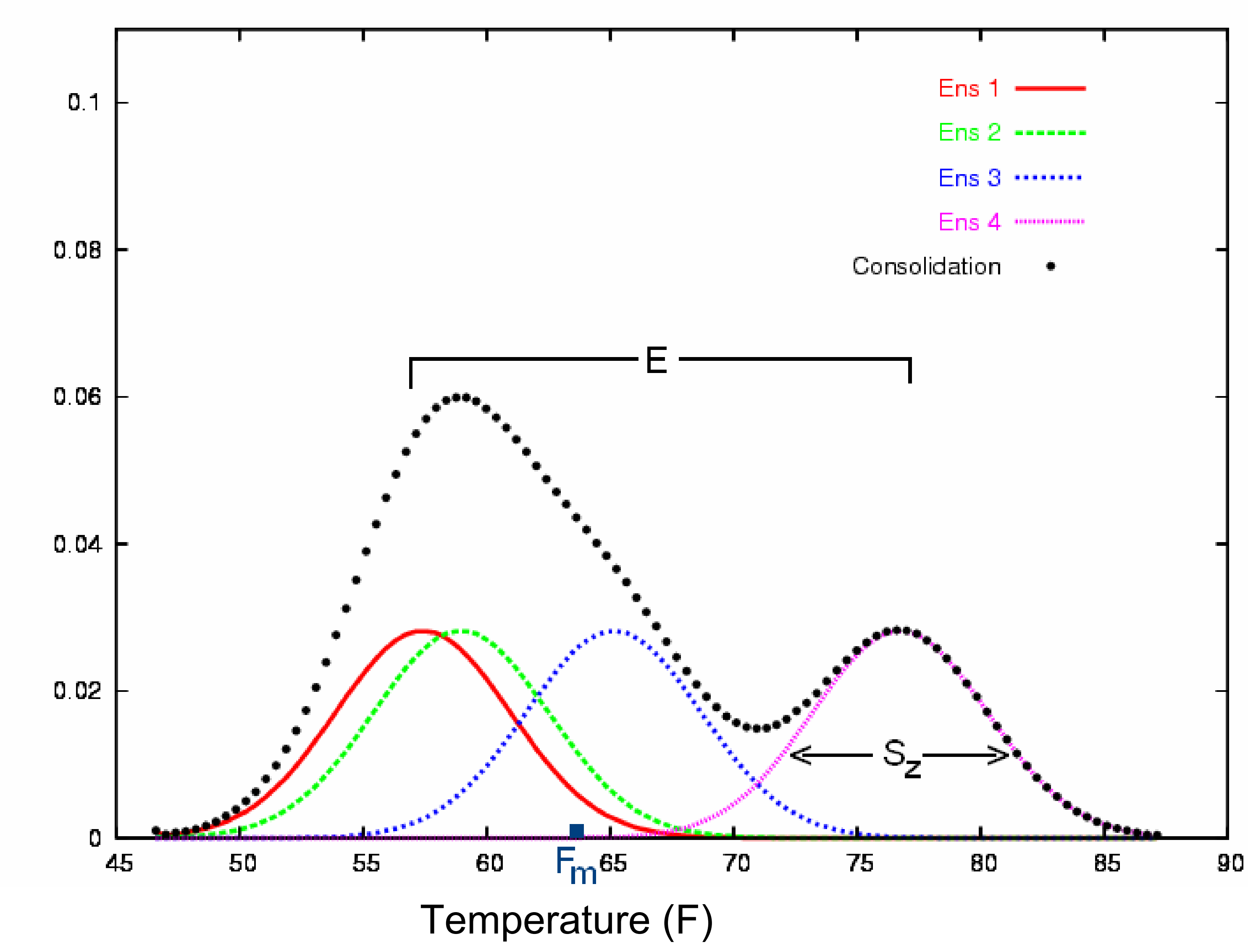
Effective Correlation

$$S_z^2 = \sigma_c^2 (1 - 2 R_{fm}^2 + R_f^2)$$

$$S_z^2 = \sigma_c^2 (1 - R_z^2)$$

$$R_z^2 = 2 R_{fm}^2 - R_f^{\ 2}$$

Schematic example



Forecast computation

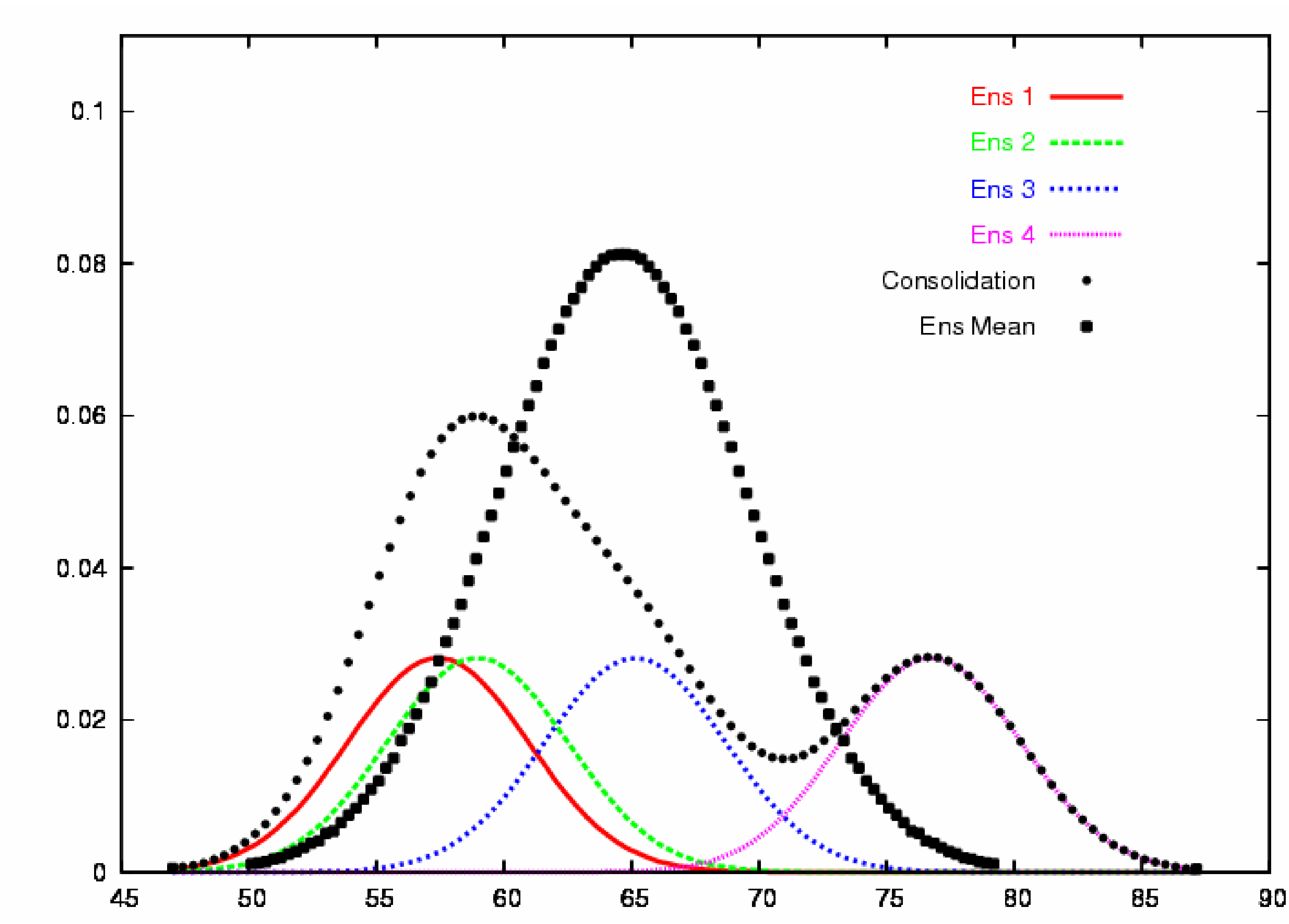
$$Z_{i} = \frac{(F_{i} - \overline{F})}{(\quad \sigma_{F})}$$

$$\hat{Z}_{i} = R_{z} Z$$

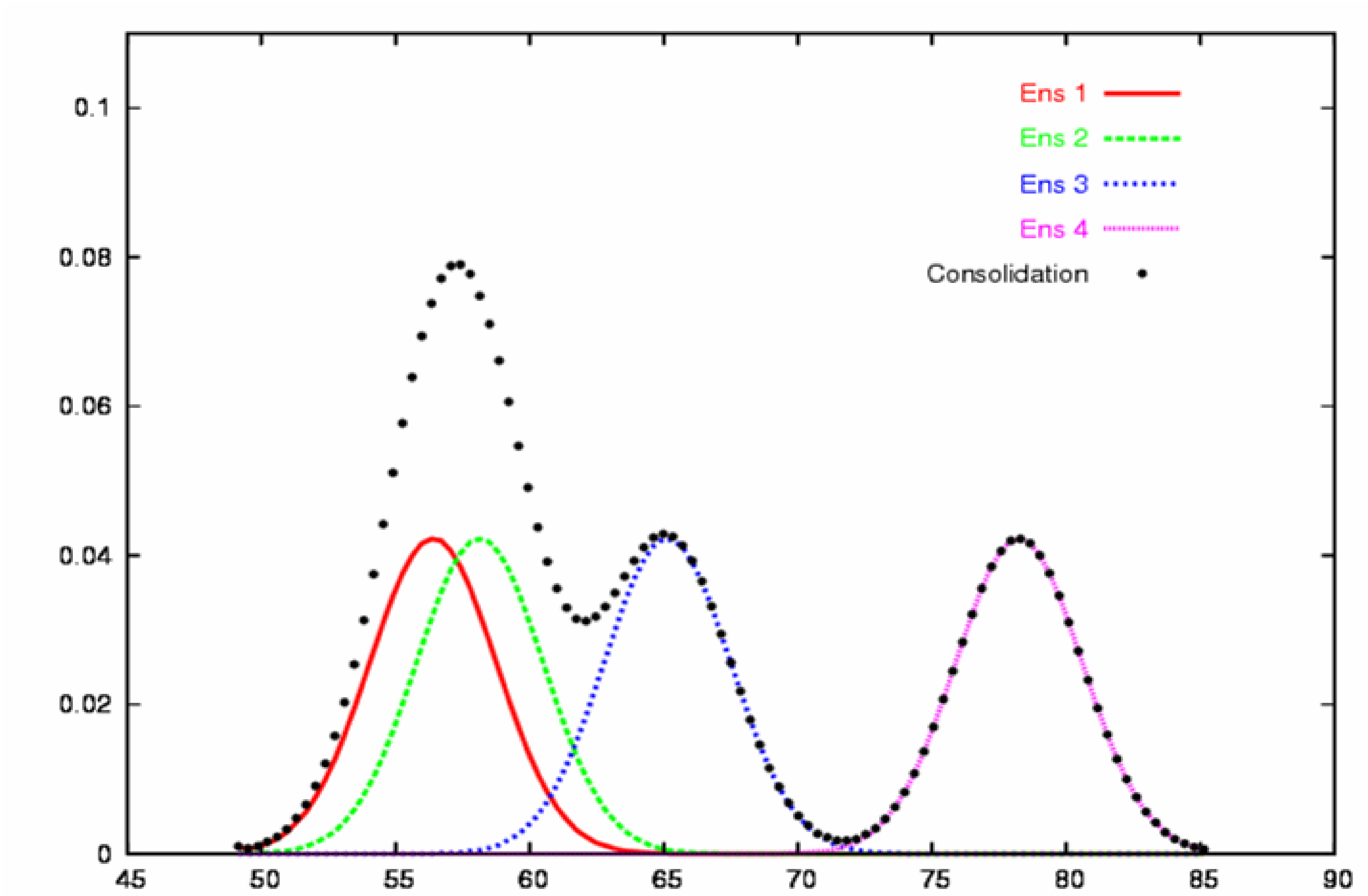
$$\overset{*}{F}_{i} = \hat{Z}_{i} \sigma_{C} + C$$

$$\hat{F}_{i} = (\overset{*}{F}_{i} - F_{m}) \frac{\hat{E}}{\overline{E}} + F_{m}$$

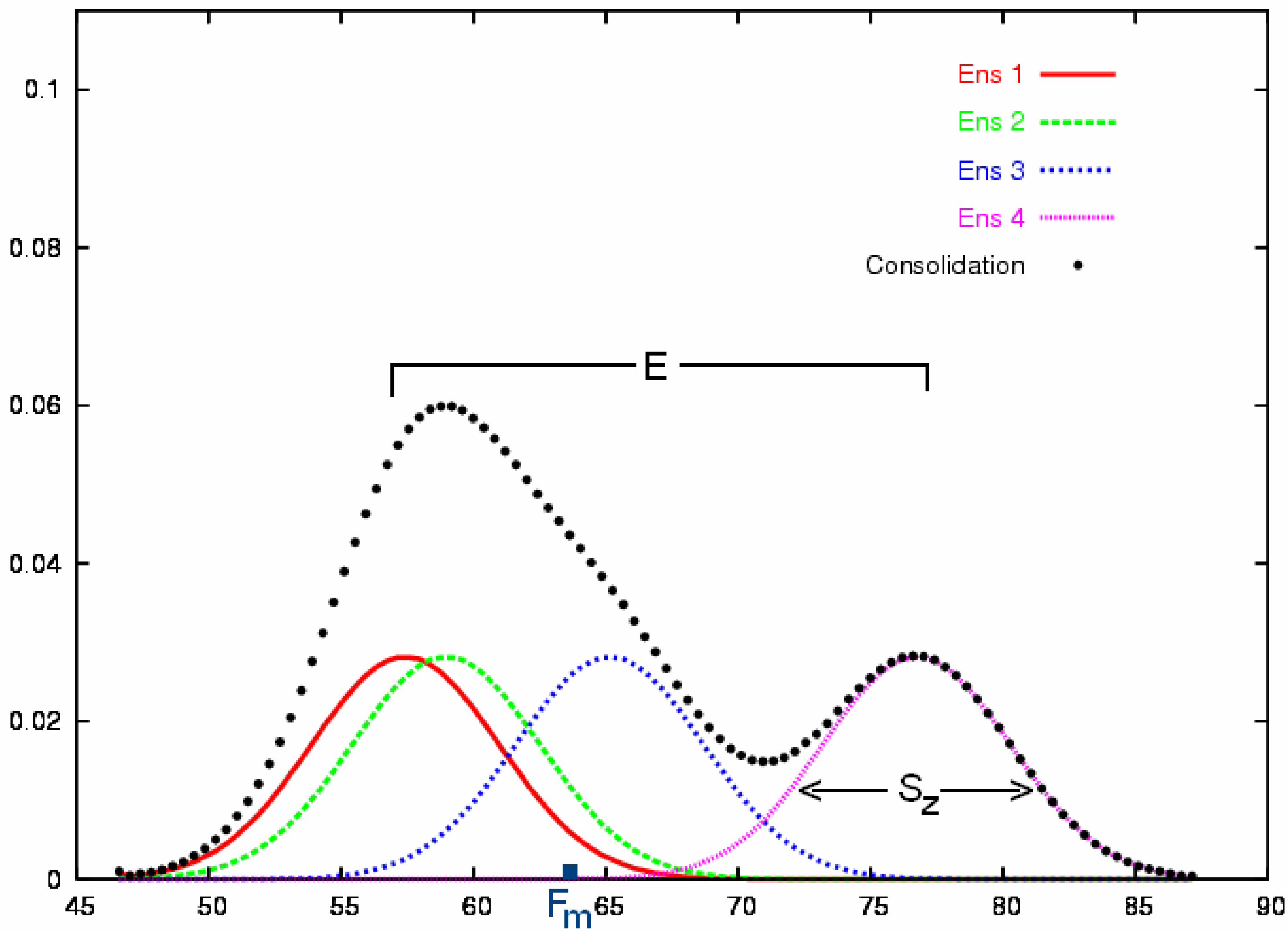
Kernel vs. Mean



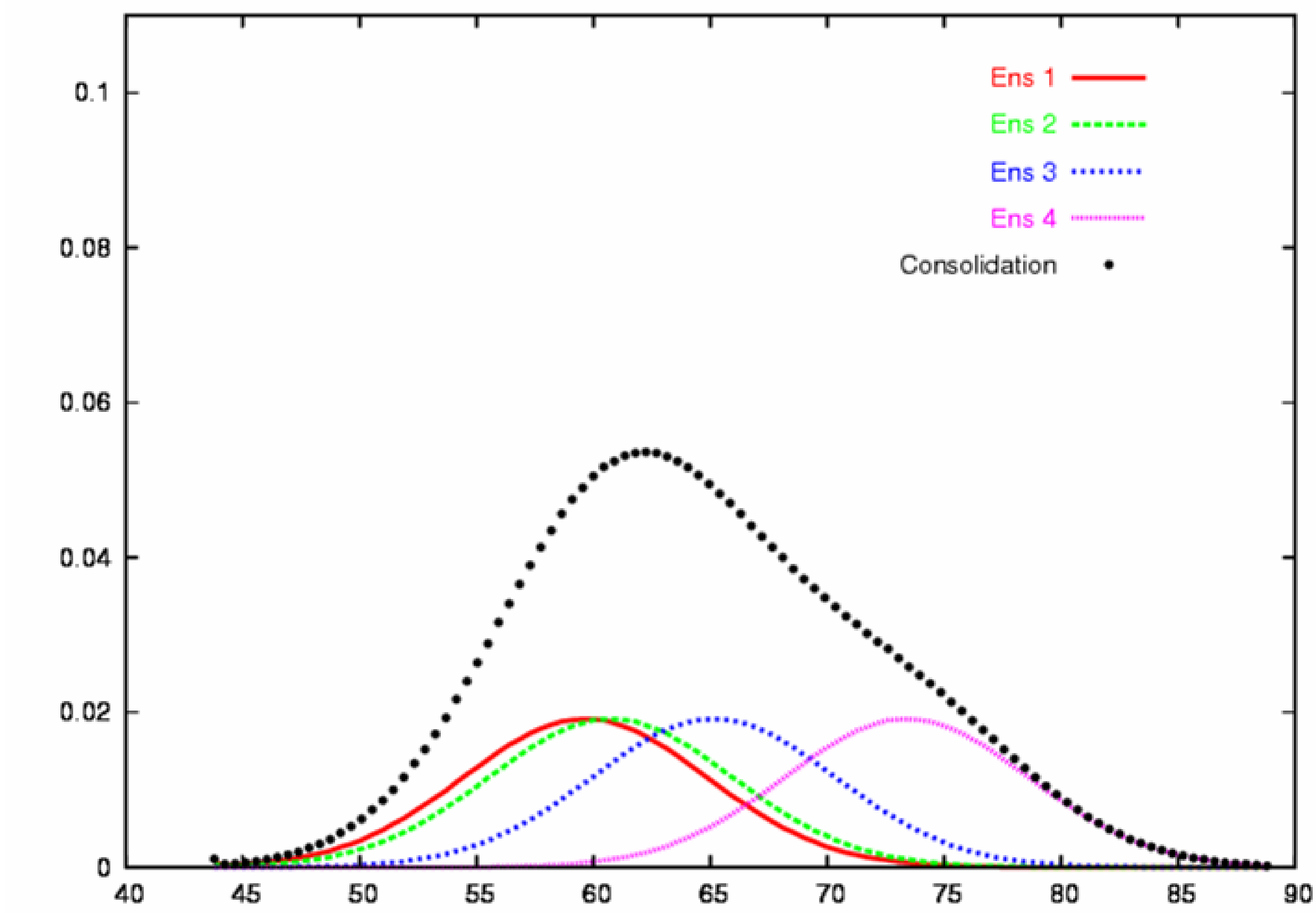
R_z=.97, R_{fm}=.94, R_i=.90



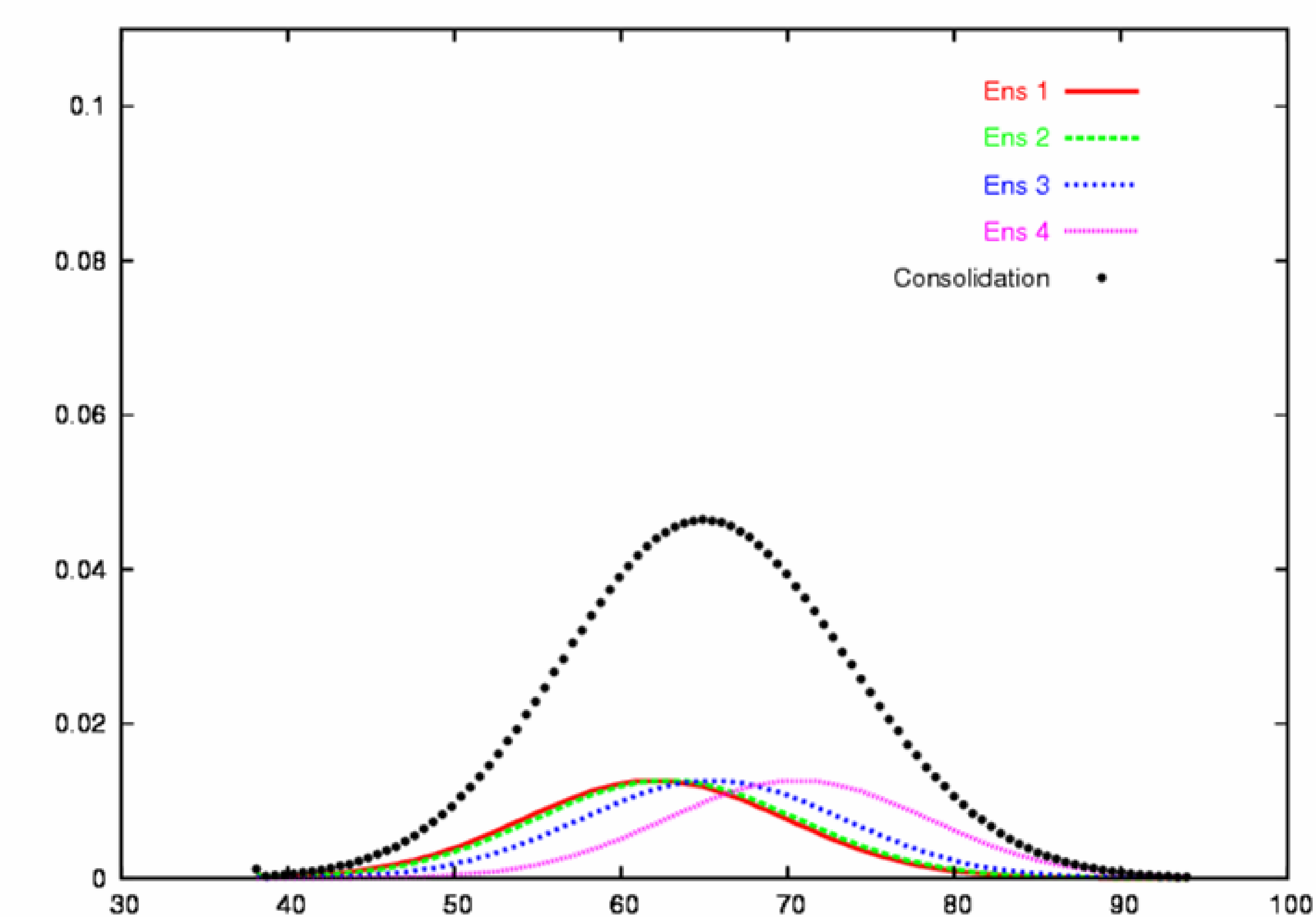
R_z=.93, R_{fm}=.87, R_f=.30



R_z=.85, R_{fm}=.67, R_i=.41



R_z=.62, R_{fm}=.46, R_f=.20



Weighting

$$w_i = \frac{R_i}{(1 - R_i)}$$

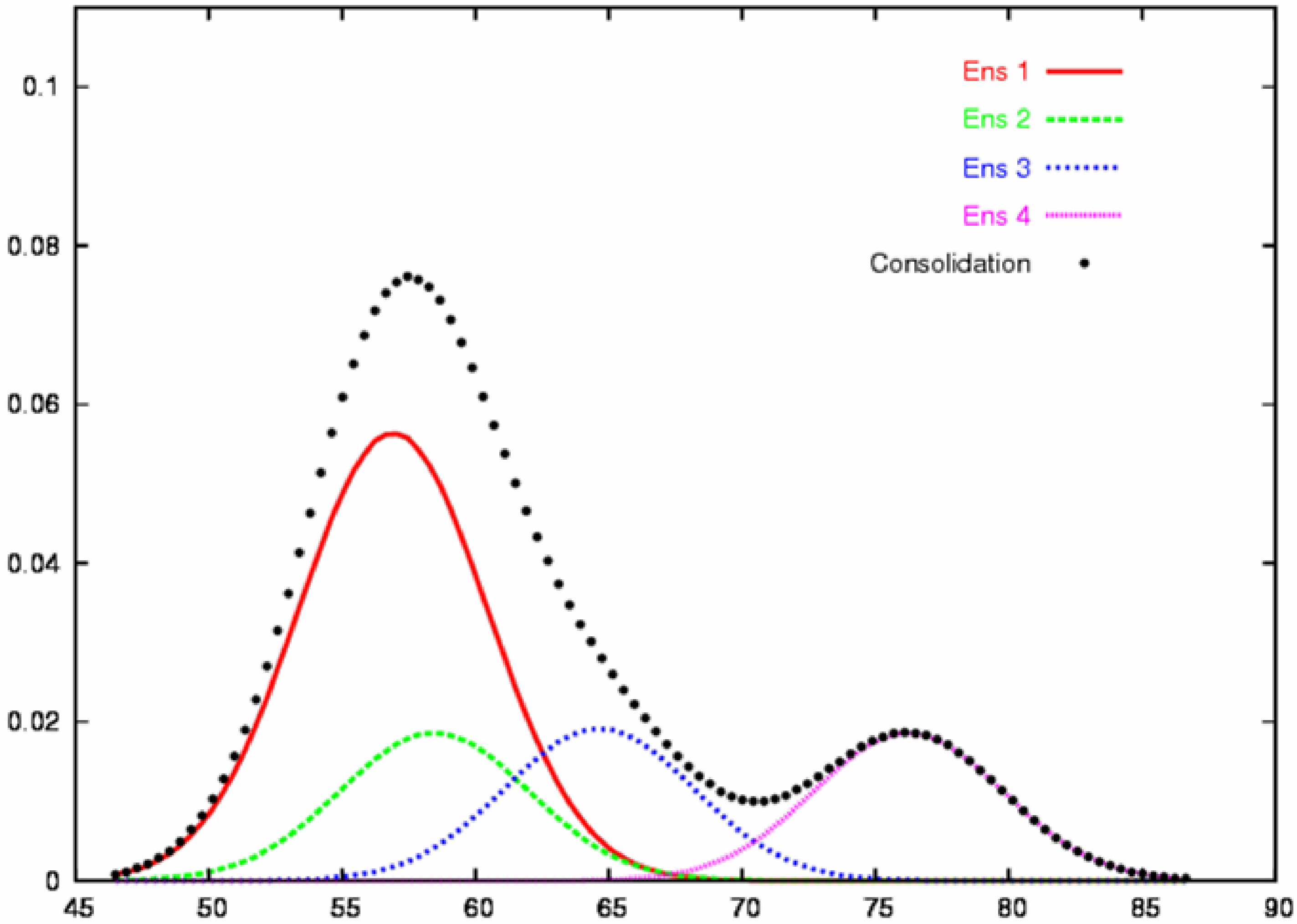
$$wt_i = \frac{w_i}{(\sum (w_i))}$$

$$R = .9: \quad 9 = \frac{.9}{(1 - .9)} \quad , \quad R = .8: \quad 5 = \frac{.8}{(1 - .8)}$$

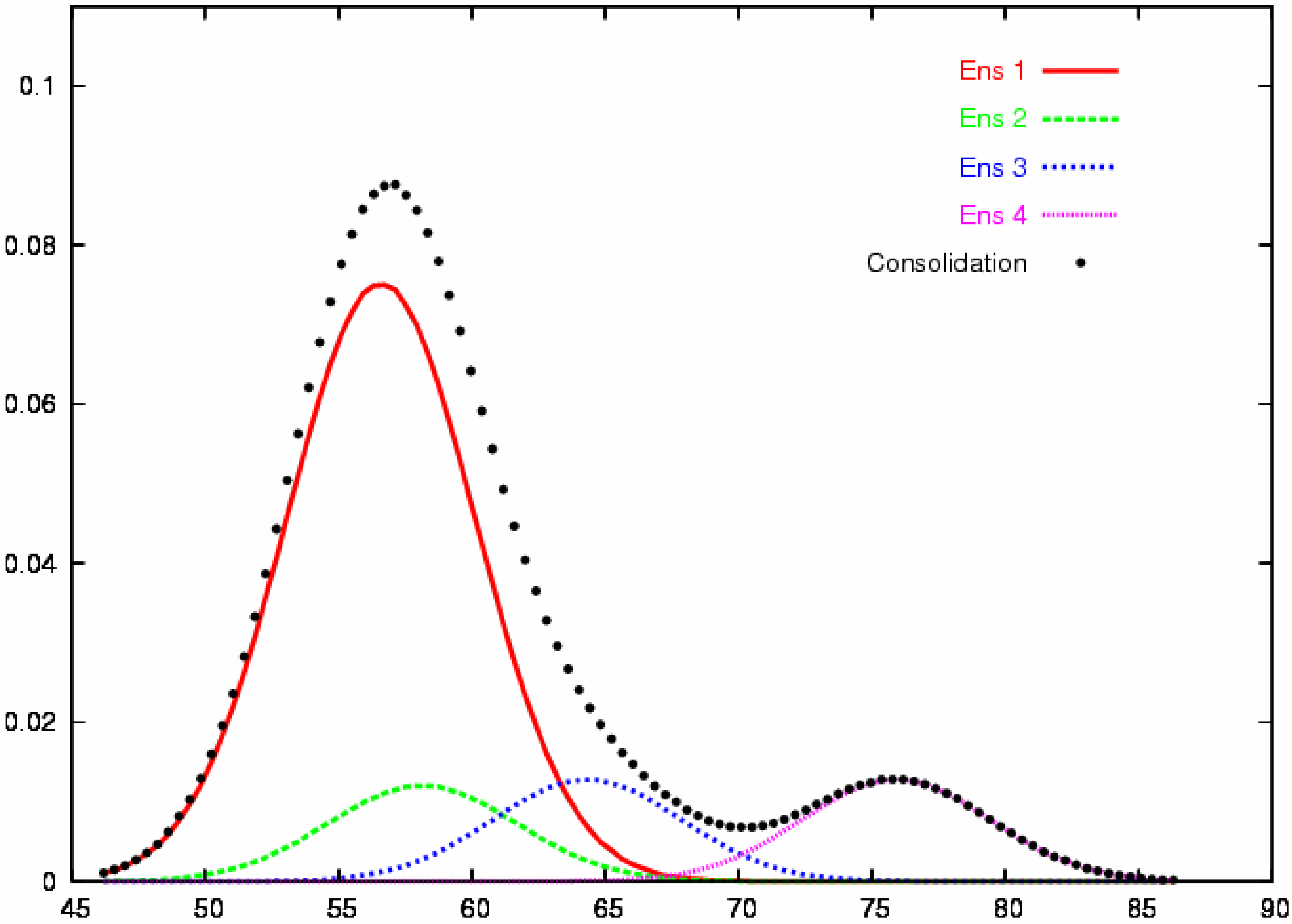
$$9 + 5 = 14$$

$$.64 = \frac{9}{14} \qquad .36 = \frac{5}{14}$$

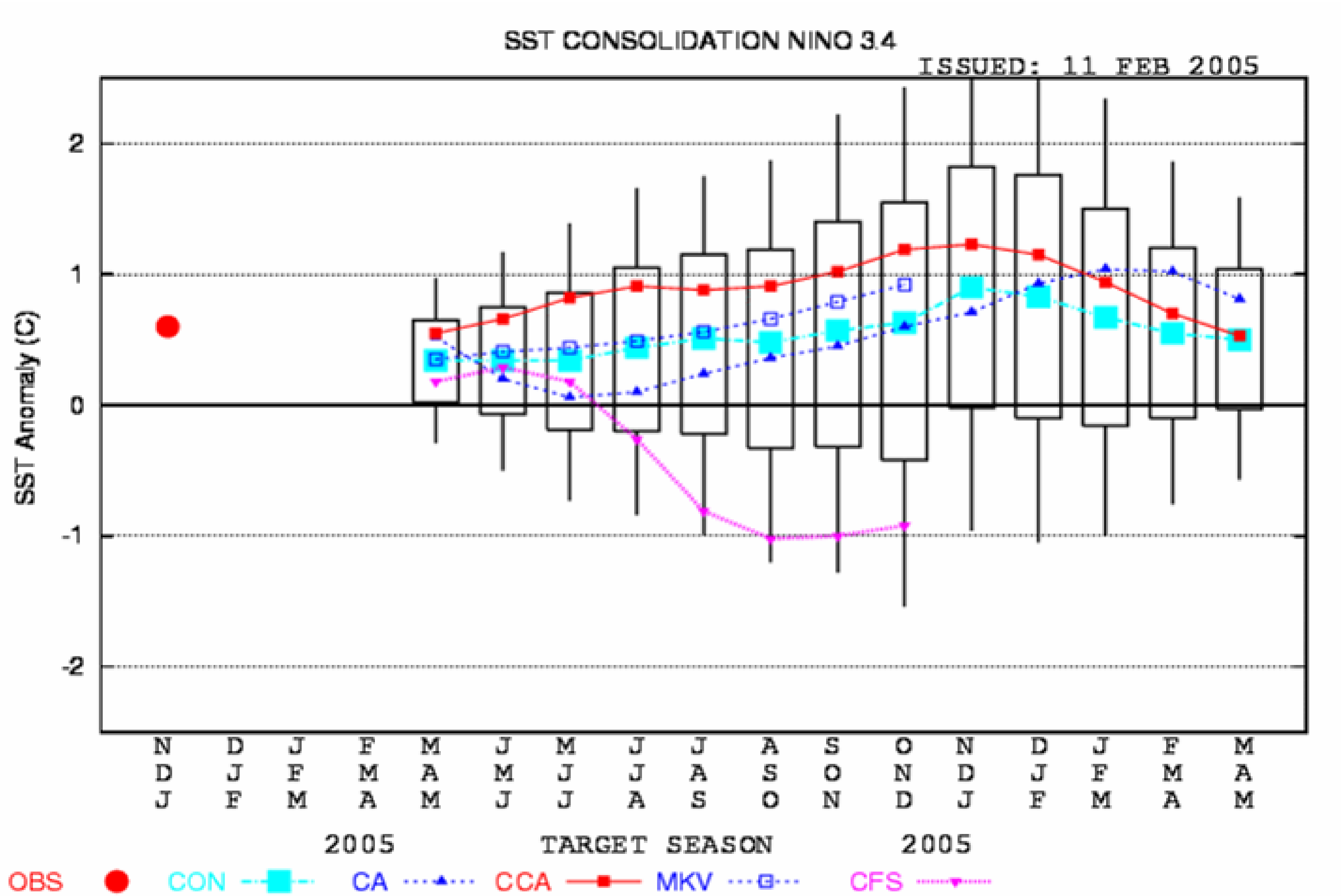
3 members + 1 forecast



3 members half wt. + 1fcst

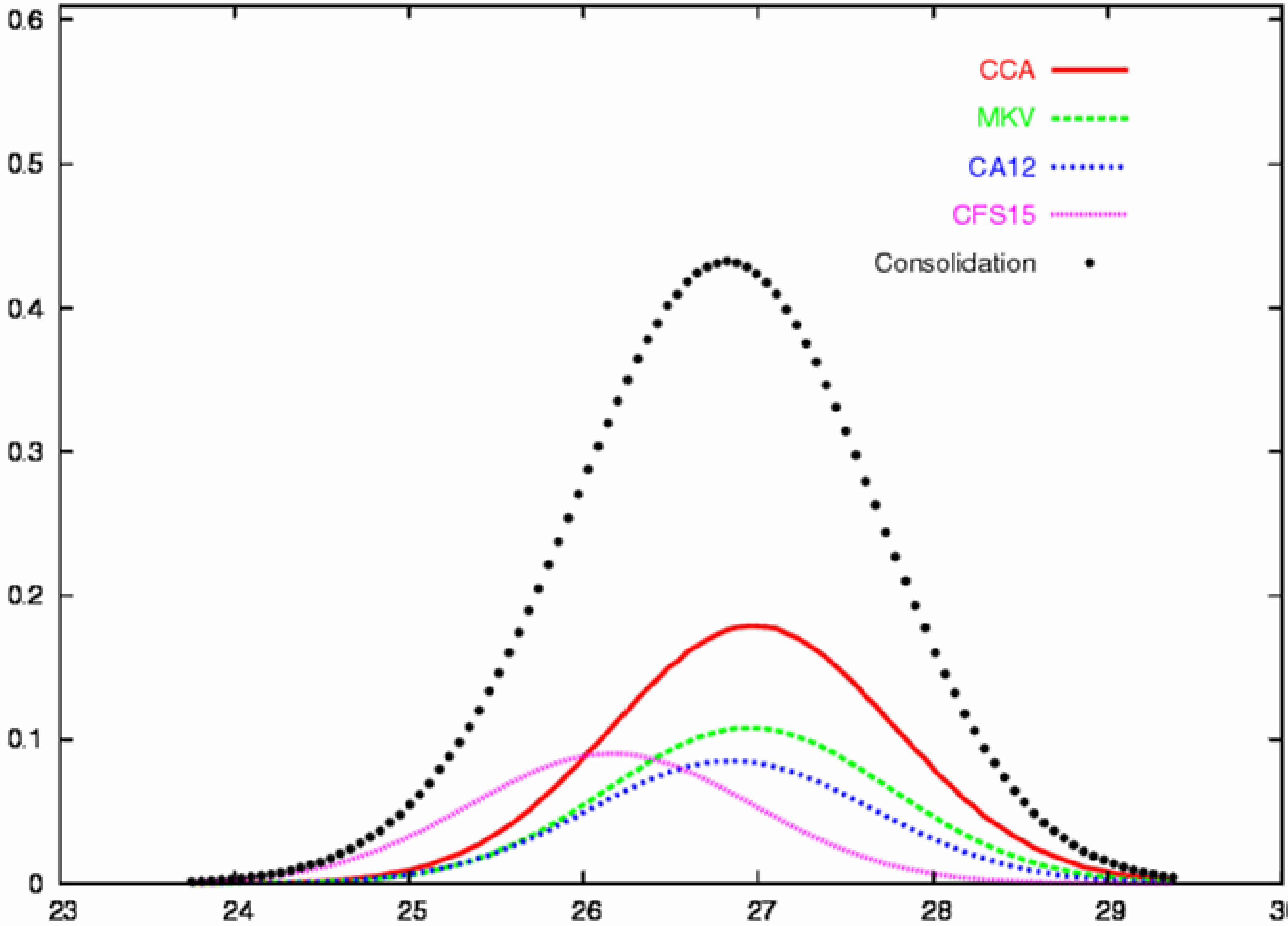


Nino 3.4 Consolidation



SST 3.4 Consolidation

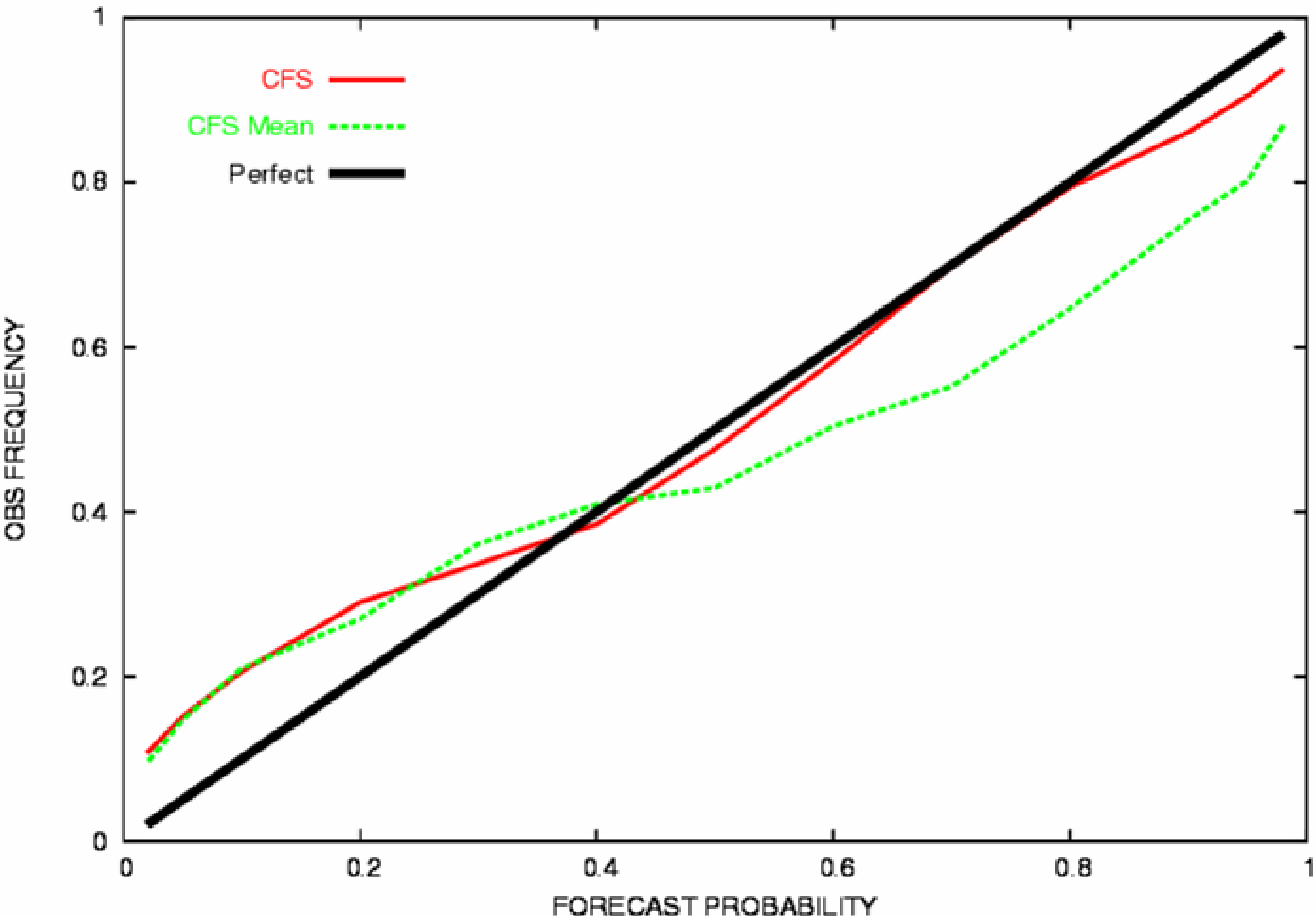
6- Mo. Lead, Rz=.63,Rm=.54,R=.44



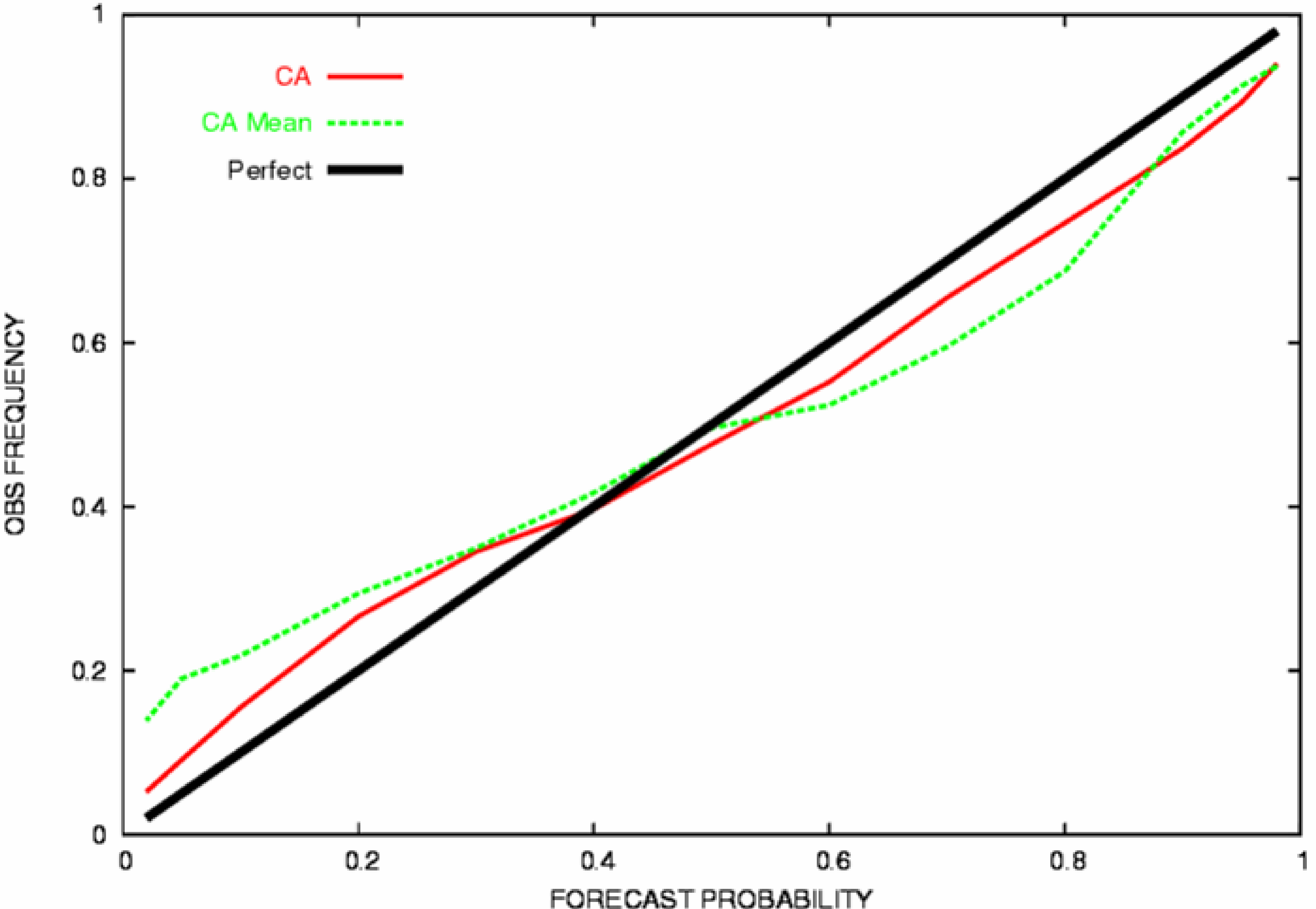
Nino 3.4 SST
6-Mo. Lead, 1981-2003, All Start times

Model	CRPSS	MAE (C)	Bias (C)
CFS	.293	.541	-.058
CFS Mean	.133	.677	.061
CA	.339	.526	.019
CA Mean	.296	.559	-.067
CON	.324	.380	-.016
CON Mean	.349	.366	.022

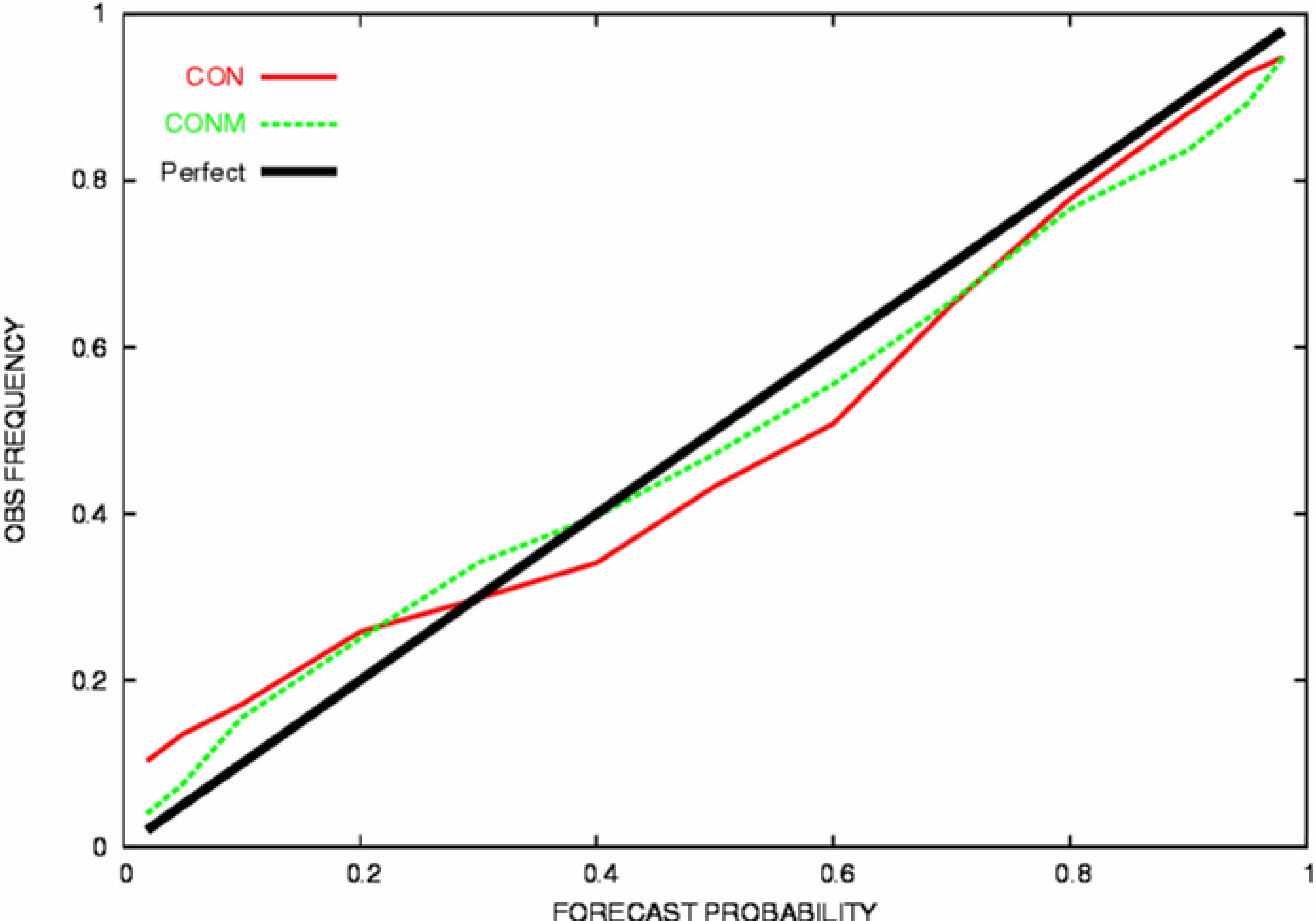
Reliability CFS



Reliability CA



Reliability Consolidation



Requirements

1. Forecast Mean
2. Standard deviation of individual members
3. Standard deviation of individual member errors
4. Standard deviation of ensemble mean
5. Standard deviation of ensemble mean errors
6. Observed mean
7. Standard deviation of observations
8. (Forecast anomaly * observed anomaly)
9. (Ensemble mean anomaly * obs anomaly)
10. Ensemble Spread

Advantages and disadvantages

Kernel Method

Advantages:

- **Uses all ensemble information**
- **One equation set produces forecasts for many thresholds.**
- **Handles irregular distributions.**

Disadvantages

- **Often is very close to a simple ensemble mean**
- **Can’t handle regime dependent bias**

Finish Line