Probabilistic Interpretation of Downscaled Forecasts

Vladimir Kryjov

Young-Mi Min

OUTLINE

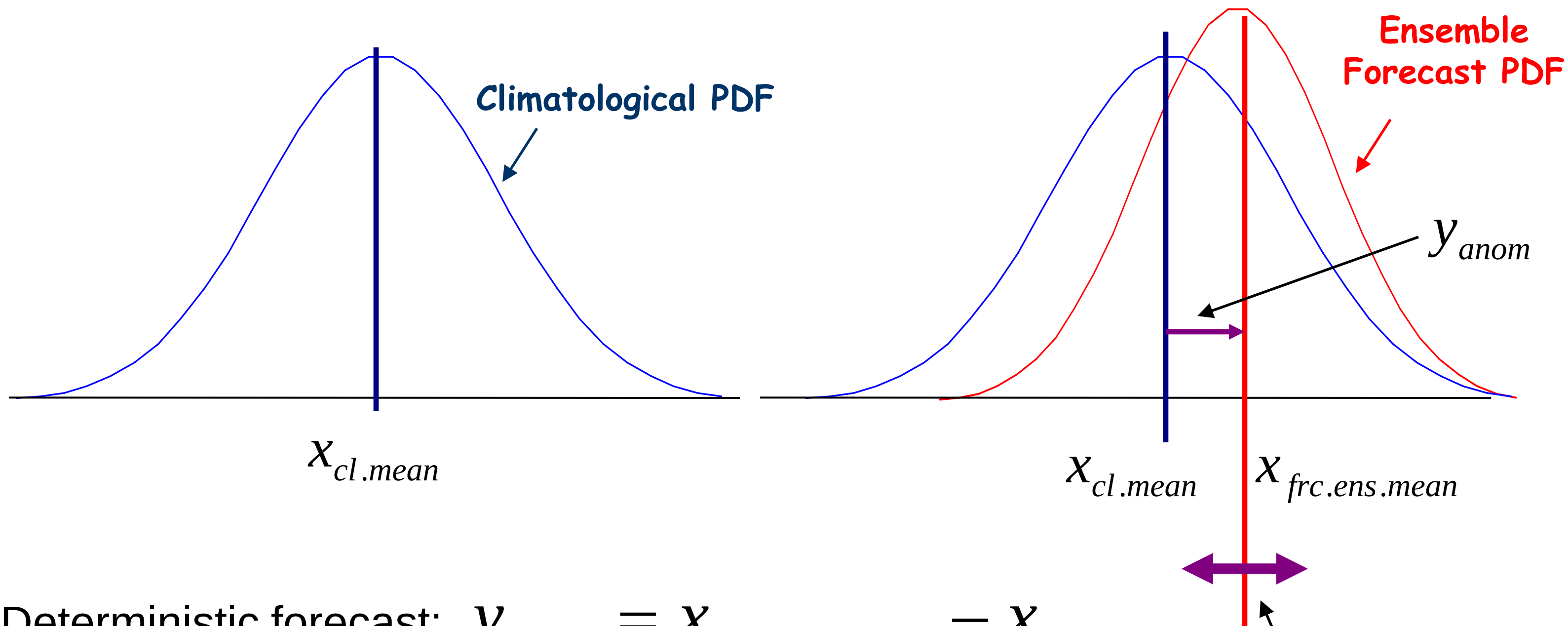
Uncertainty of a Raw Model Ensemble Forecast and its Probabilistic Interpretation

Uncertainty of a Regression Based Downscaled Forecast

Probabilistic Interpretation of the Downscaled Forecast

Uncertainty of a Raw Model Ensemble Forecast and its Probabilistic Interpretation

Deterministic forecast

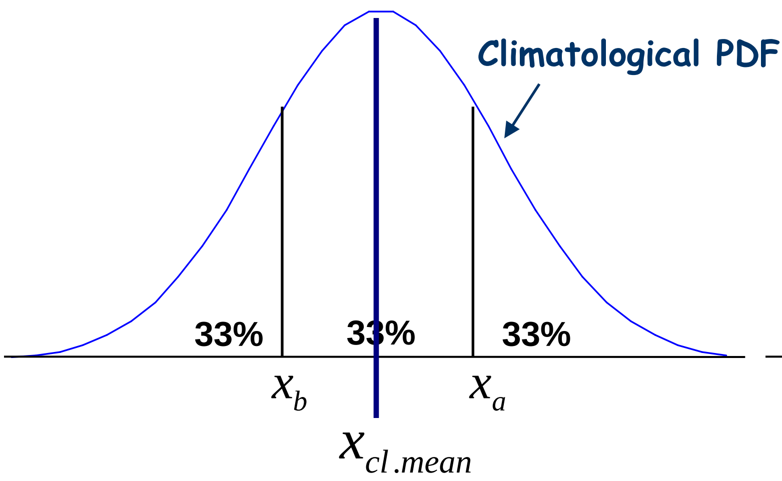


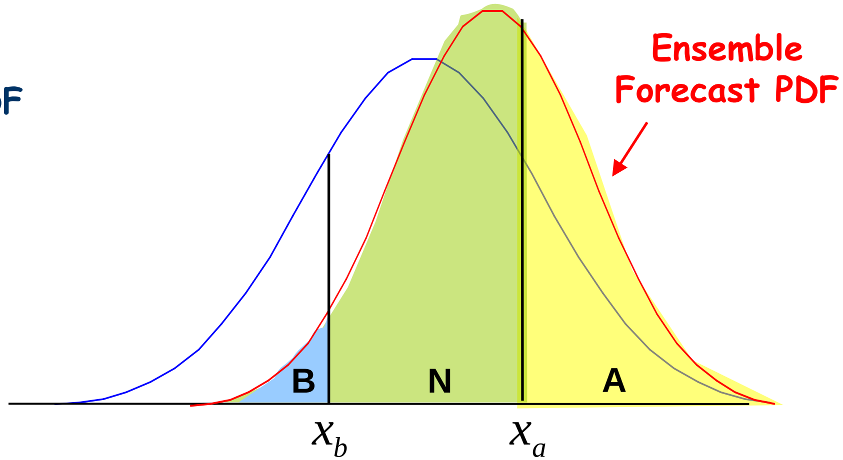
Deterministic forecast: $y_{anom} = x_{frc.ens.mean} - x_{cl.mean}$

Issuing deterministic forecast

we usually "forget" about its uncertainty

Probabilistic forecast





Probabilistic forecast
= probability of each of the categories
(usually - three equiprobable categories)

- **A** : **Probability of Above-normal**
- **N** : **Probability of Near-normal**
- **B** : **Probability of Below-normal**

$$P(BN) = prob[x \le x_b] = \int_{-\infty}^{x_b} f(x)dx$$

$$P(NN) = prob[x_a < x \le x_b] = \int_{-\infty}^{x_a} f(x)dx - P(BN)$$

$$P(AN) = prob[x_a < x] = 1 - P(BN) - P(NN)$$

where, $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{(x-\mu_{frc})^2}{2\sigma^2})$

- with Gaussian Approximation

  - **For the middle/upper tercile boundary**

    mean plus 0.43 times the standard deviation

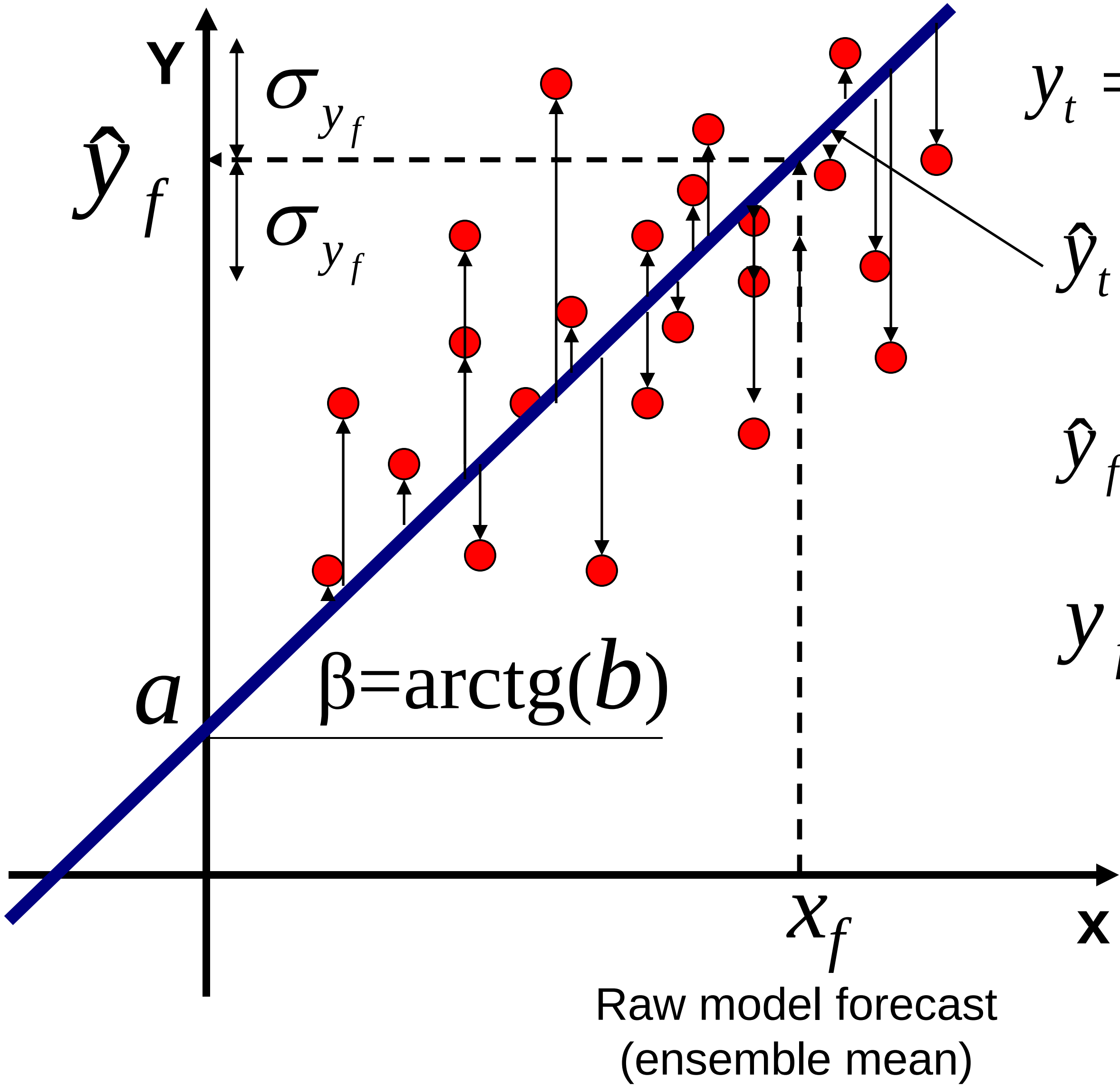
    □ $$x_a = x_{cl.mean} + .43\sigma$$

  - **For the lower/middle tercile boundary**

    mean minus 0.43 times the standard deviation

    □ $$x_b = x_{cl.mean} - .43\sigma$$

Uncertainty of Regression Based Downscaled Forecast

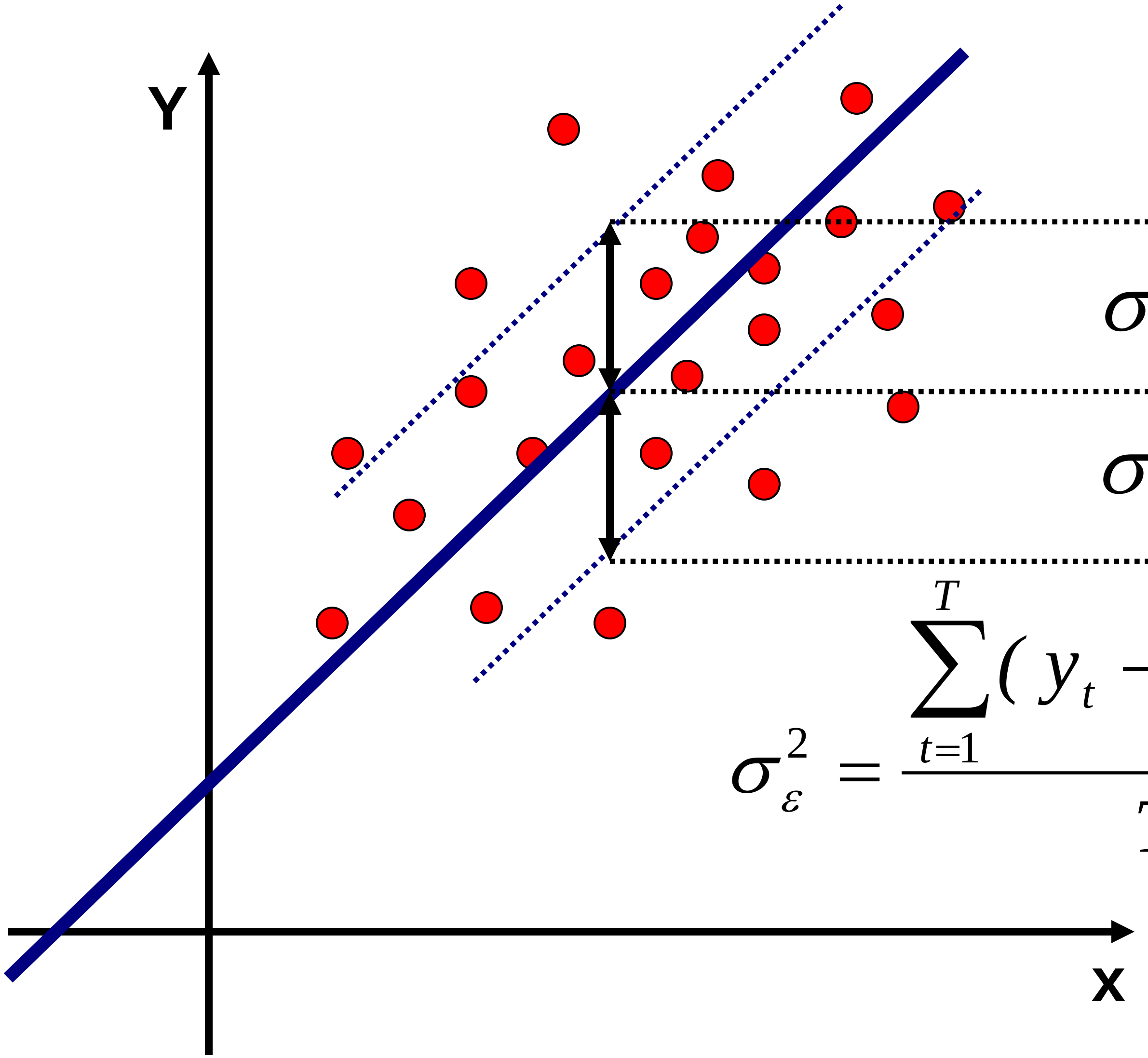


$$y_t = a + bx_t + \varepsilon_t$$

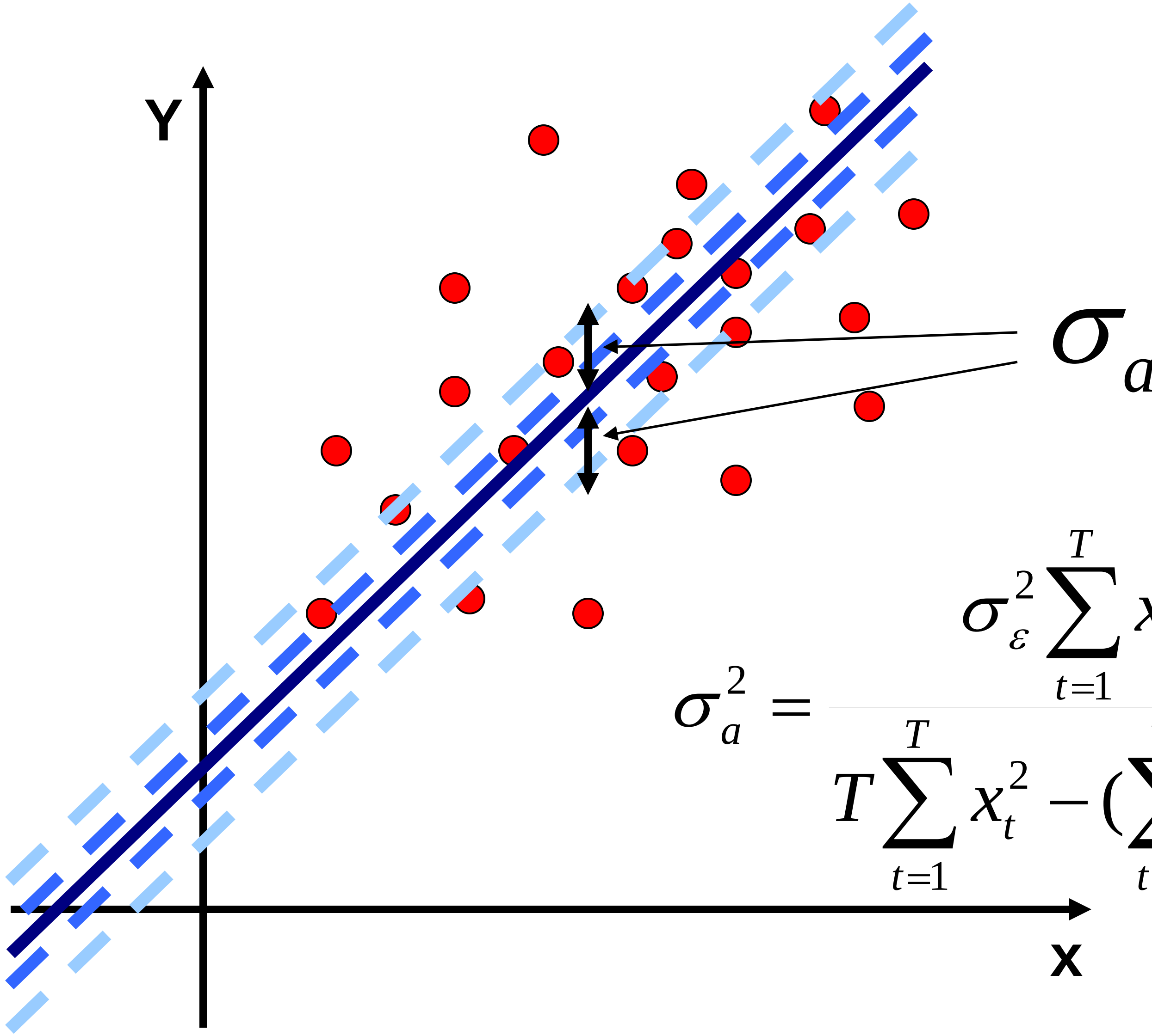
$$\hat{y}_t = a + bx_t$$

$$\hat{y}_f = a + bx_f$$

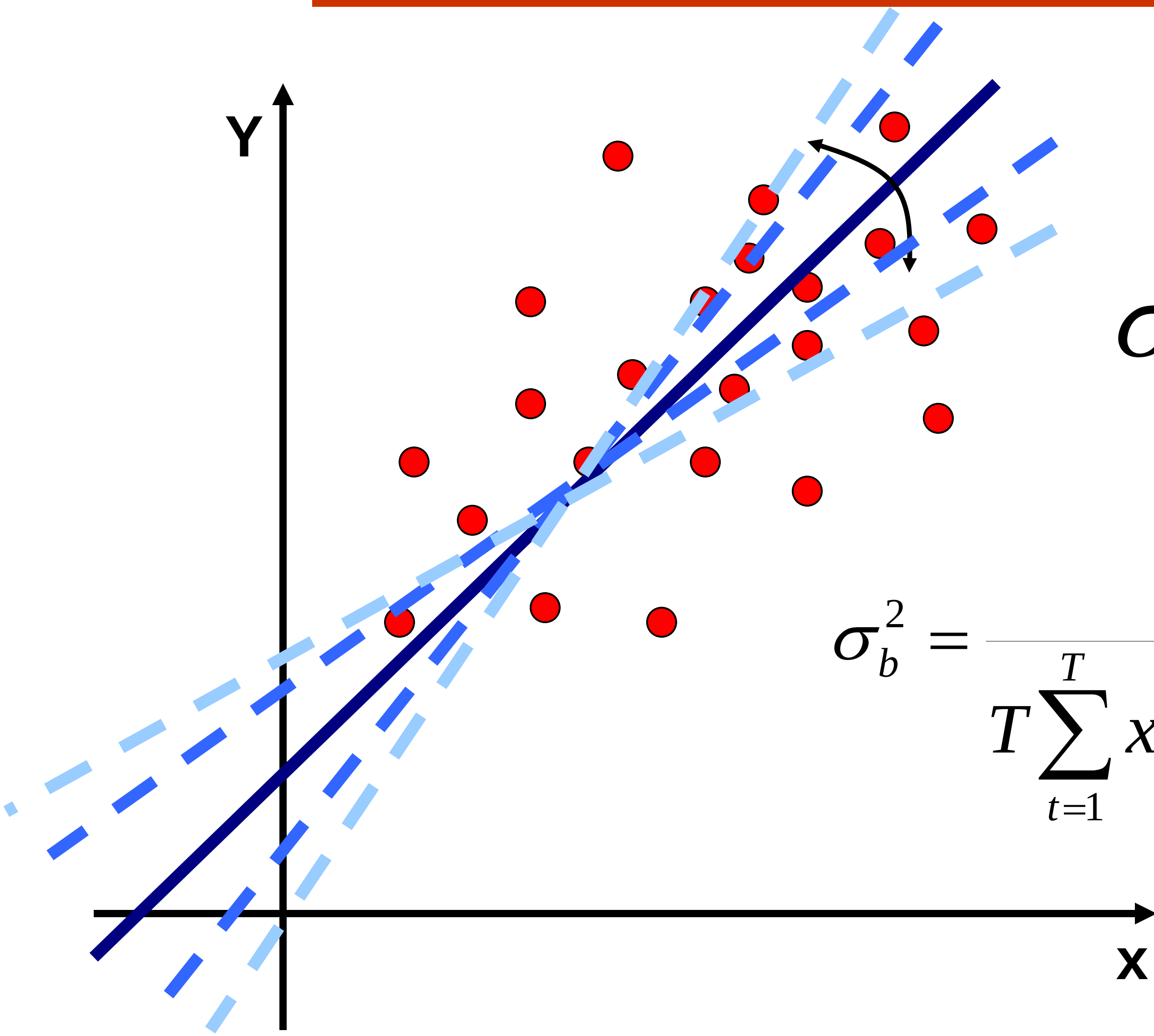
$$y_f = \hat{y}_f \pm \sigma_{y_f}$$



$$\sigma_\varepsilon^2 = \frac{\sum_{t=1}^{T} (y_t - (a + bx_t))^2}{T - 2}$$



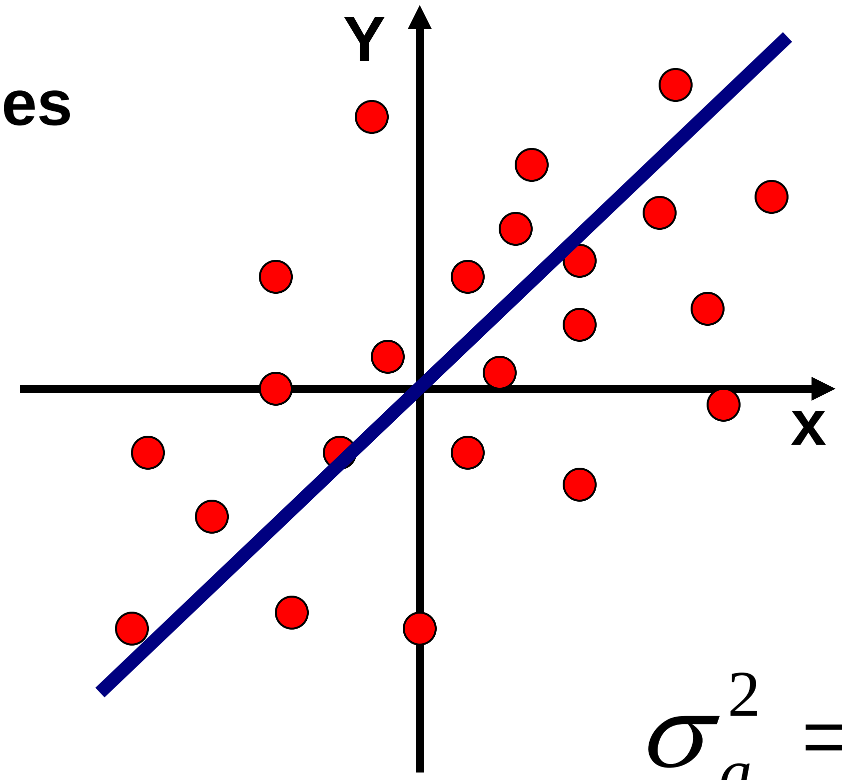
$$\sigma_a^2 = \frac{\sigma_\varepsilon^2 \sum_{t=1}^{T} x_t^2}{T\sum_{t=1}^{T} x_t^2 - \left(\sum_{t=1}^{T} x_t\right)^2}$$



σ_b

$$\sigma_b^2 = \frac{T\sigma_\varepsilon^2}{T\sum_{t=1}^{T} x_t^2 - \left(\sum_{t=1}^{T} x_t\right)^2}$$

For anomalies



$$\sigma_a^2 = \sigma_\varepsilon^2 / T$$

$$\sigma_b^2 = \sigma_\varepsilon^2 / \sum_{t=1}^{T} x_t^2$$

Error (uncertainty) in the forecast by a single predictor

(under assumption that all the errors are independent upon each other)

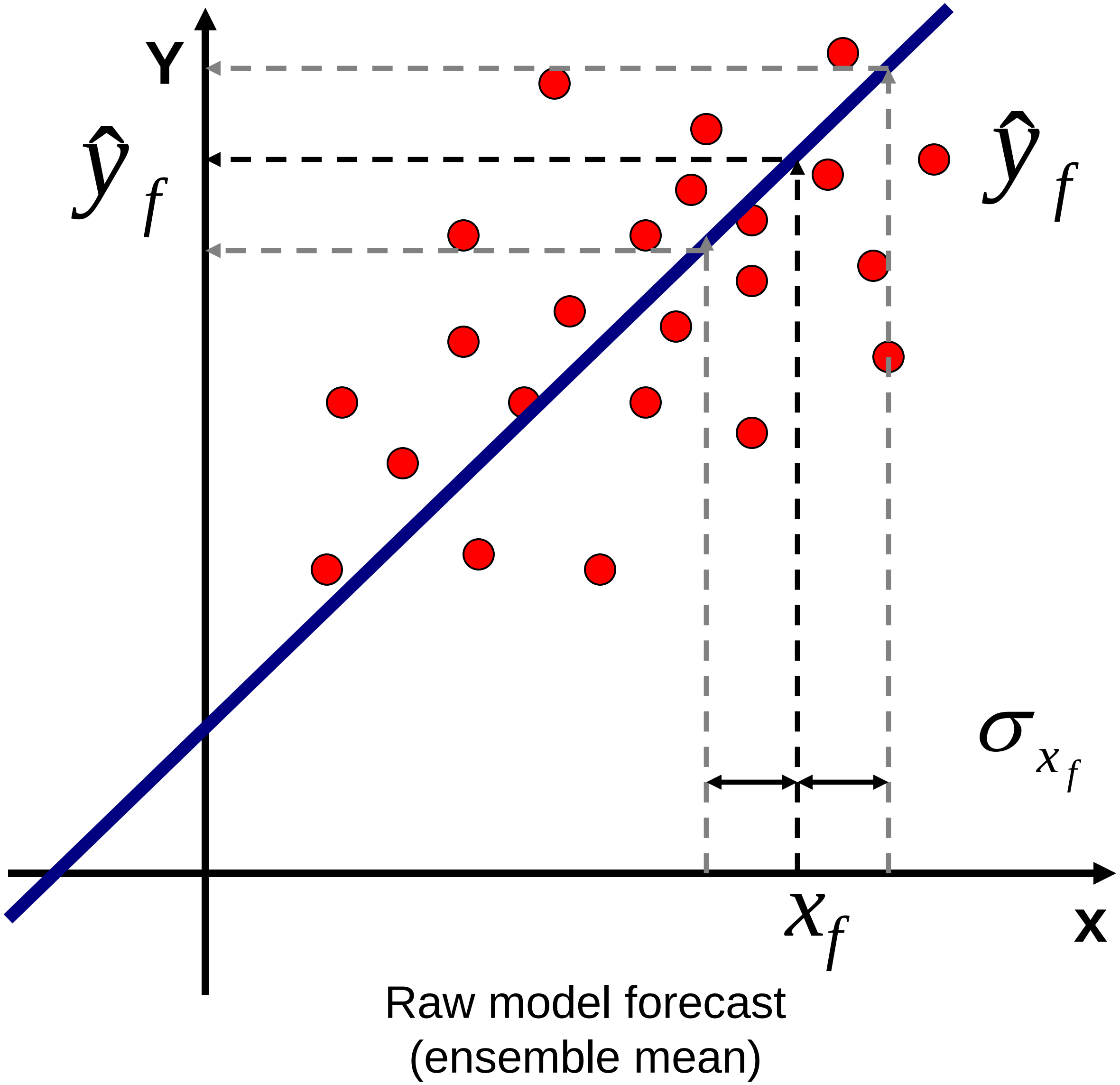
$$\sigma_{y_f}^2 = \sigma_\varepsilon^2 + \sigma_a^2 + \sigma_b^2 x_f^2 + \sigma_{x_f}^2 b^2$$

, with $\sigma_{x_f}^2$ being a variance of predicted ensemble mean:

$$\sigma_{x_f}^2 = \sigma_{xe_f}^2 / n$$

, where $\sigma_{xe_f}^2$ is a predicted ensemble variance,

n is an ensemble size



$$\hat{y}_f = a + b x_f$$

$$\sigma_{x_f} = \frac{\sigma_{xe_f}}{\sqrt{n}}$$

Rough estimates of contribution by each error

based on precipitation and temperature forecasts
for 60 Korea stations

$$\sigma_{y_f}^2 = \sigma_\varepsilon^2 + \sigma_a^2 + \sigma_b^2 x_f^2 + \sigma_{x_f}^2 b^2$$

100% = 30-90% + 1-10% + 1-20% + 5-60%

Probabilistic Interpretation of the Downscaled Forecast

$$y_f = \hat{y}_f \pm \sigma_{y_f}$$ - Deterministic forecast with uncertainty

$\hat{y}_f$ and σ_{y_f} are parameters of Gaussian PDF

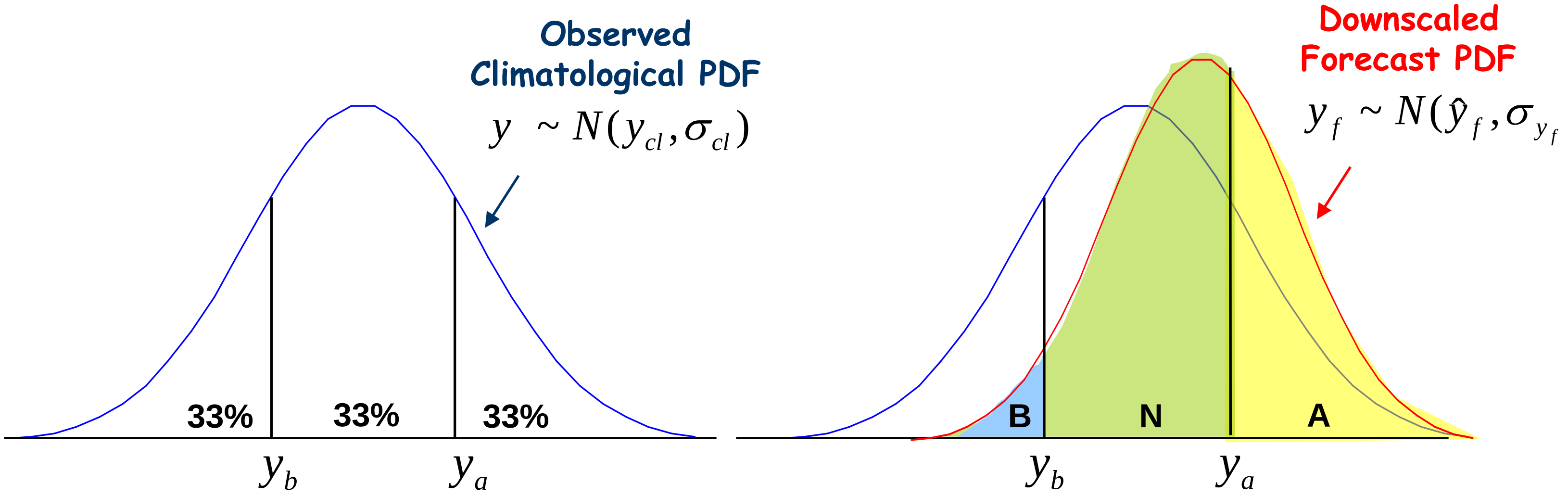
$$y_f \sim N(\hat{y}_f, \sigma_{y_f})$$

– the predictand downscaled forecast PDF.

We know the predictand climatological PDF:

$$y \sim N(y_{cl}, \sigma_{cl})$$

and we can estimate ***P(AN), P(NN), P(BN).***



$$P(BN) = prob[y \le y_b] = \int_{-\infty}^{y_b} f(y)dy$$

$$P(NN) = prob[y_b < y \le y_a] = \int_{-\infty}^{y_a} f(y)dy - P(BN)$$

$$P(AN) = prob[y_a < y] = 1 - P(BN) - P(NN)$$

where $f(y) = \frac{1}{\sigma_{y_f}\sqrt{2\pi}} exp(-\frac{(y - \hat{y}_f)^2}{2\sigma_{y_f}^2})$ is Gaussian PDF

MULTIMODEL COMBINATION

- (weighted) average of individual model forecasts

$$P(E_j) = \frac{1}{\sum_m w_m} \sum_{m=1}^{M} w_m (P(E_j))_m$$

where E_j is an event ***AN*** or ***NN*** or ***BN*** and so $\sum_{j=1}^{3} P(E_j) = 1$

Individual model weights $\boldsymbol{w_m}$ is a subject of separate invesatigation. Reasonable values are proportional to correlation coefficients between predictand and $\boldsymbol{m}$-model predictor.

Thank You