

---

# **Постпроцессинг Долгосрочных модельных прогнозов**

**В.Н.Крыжов**

**Y.-M. Min**

---

---

## **План**

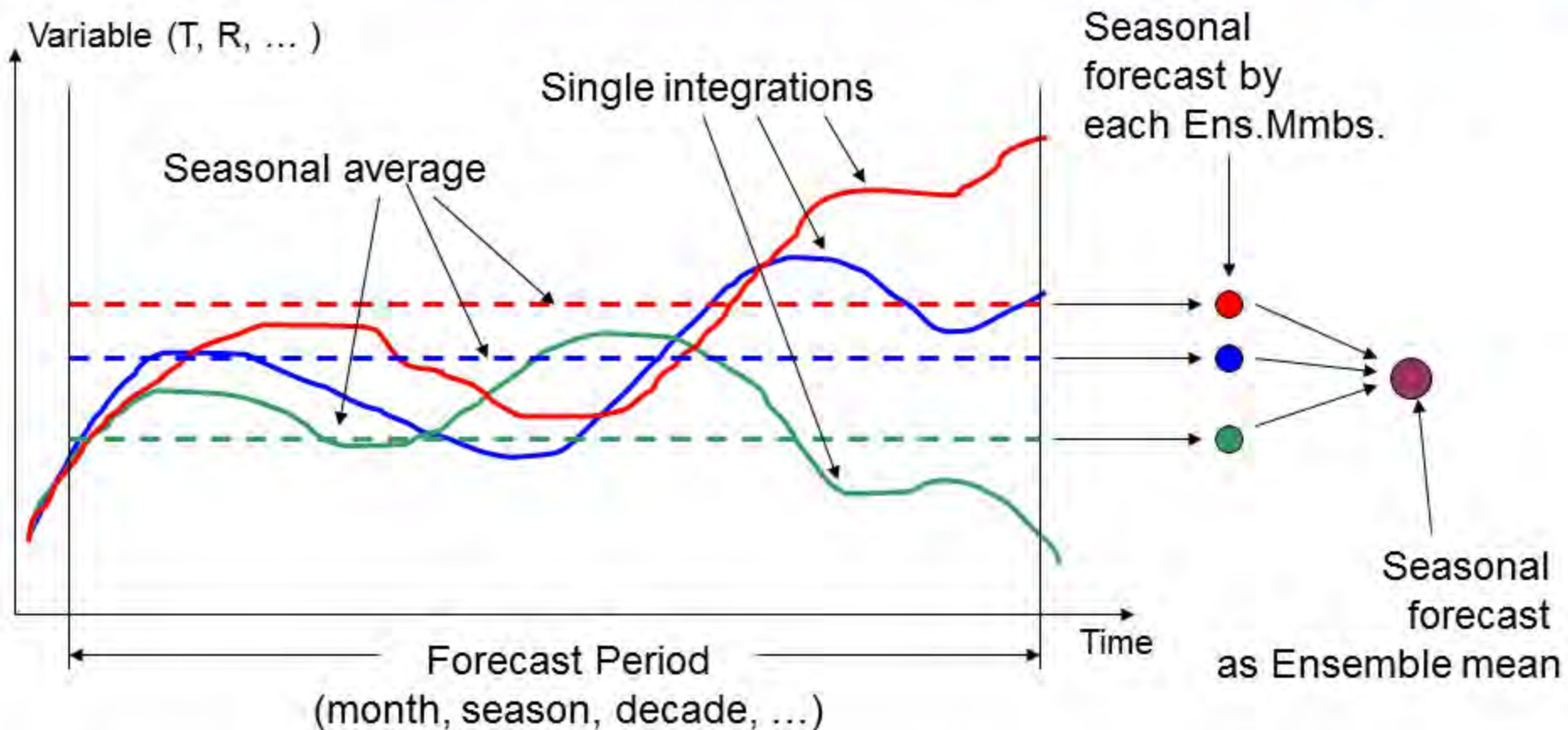
- Долгосрочный модельный прогноз**
  - Ансамблевый прогноз**
  - Мультимодельный прогноз**
-

### Краткосрочный модельный прогноз

- предсказуемость первого рода (Lorentz, 1963):  
модель воспроизводит развитие атмосферных процессов синоптического масштаба, стартуя с начальных условий и сохраняя память о начальных условиях

### Долгосрочный модельный прогноз

- предсказуемость второго рода (Lorentz, 1963):  
модель не помнит начальные условия, она воспроизводит аномалии макромасштабных атмосферных процессов, связанные с аномалиями граничных условий.



**Ансамбль интегрирований и ансамблевый прогноз**

## Долгосрочный модельный прогноз

Устранение «смещения климата» в долгосрочном модельном прогнозе

«Сырой» модельный прогноз можно представить как

$$X_{model} = X + model\_climatology$$

$X$  – аномалия

$model\_climatology$  – модельный климат

– рассчитывается как среднее значение по историческим прогнозам

Наблюдения можно представить как

$$YY = Y + climatology$$

$Y$  – аномалия

$climatology$  – климат – рассчитывается как среднее значение

по наблюдениям за период исторических прогнозов

Смещение климата – это разница между  $model\_climatology$  и  $climatology$

- для определенного периода исторических прогнозов – это константа

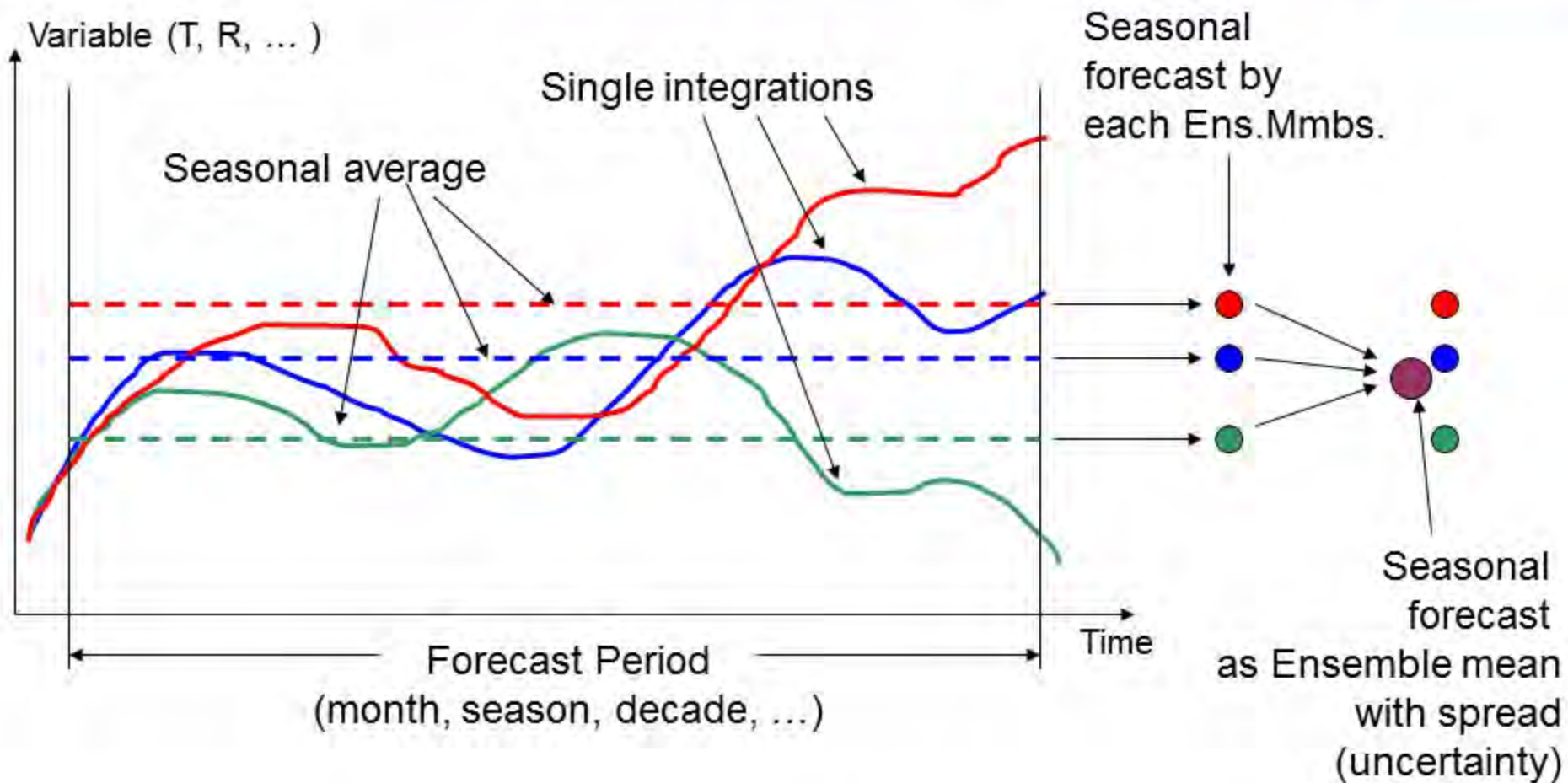
Модельный прогноз с устраненным смещением:

$$XX = X + climatology$$

---

## **Ансамблевый прогноз**

---



**Ансамбль интегрирований и ансамблевый прогноз**

Идея, лежащая в основе конструкции ансамбля, хорошо иллюстрируется разложением единичного модельного прогноза аномалии  $X$  на два компонента (Казначеева и Тростников, 2008, и ссылки):

$$X = yu + err\_n$$

В этом разложении

$yu$  – это часть прогноза модели, зависящая от граничных условий

$err\_n$  – нормально распределенная случайная погрешность – «погодный шум».

Если принять за ансамблевый прогноз среднее ансамбля, то

$$X\_m = \frac{1}{n\_ens} \sum (yu + err\_n)$$

$$err\_n\_ens = \pm \frac{\sigma\_err\_n}{\sqrt{n\_ens}} \quad (68\% \text{ conf.int. } X\_m)$$

$err\_n\_ens \rightarrow 0$  с увеличением  $n\_ens$



echam.pst 97 aso pcp 10 FCSTS run aug97  
 $R = 0.62$ ,  $P = 0.56$ , "INDONESIA": 10S20N ; 95E140E

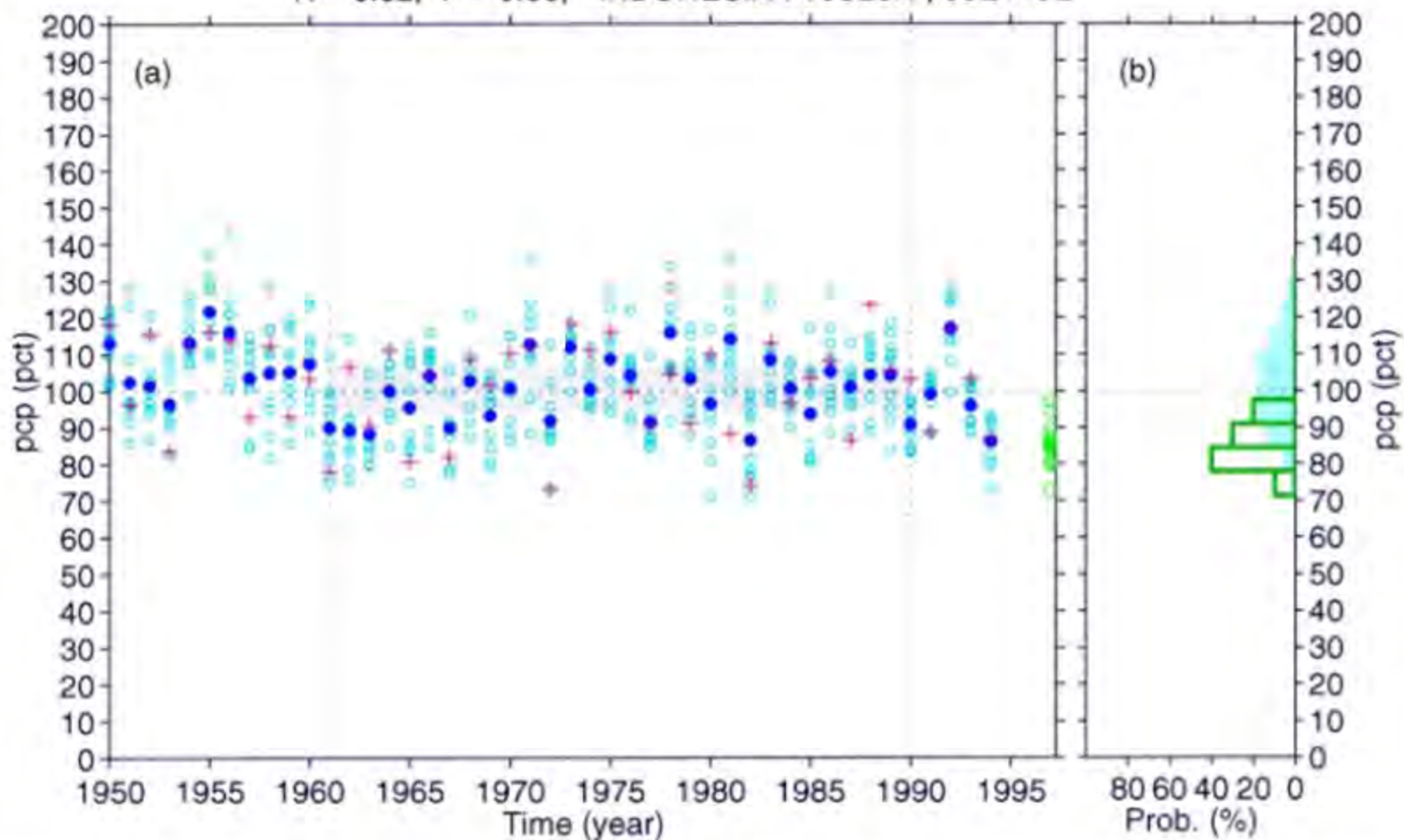


FIG. 6. (a) Historical performance of the ECHAM3 model for the August–October season compared to observations averaged over the “Indonesia” region (10°S–20°N, 95°E–140°E). The open blue circles show the model anomaly for individual ensemble members (expressed as a percentage of long-term mean), solid blue circles show the ensemble mean, and red crosses indicate the observed anomaly. The green circles are for the current forecast, and the solid green circle represents the ensemble mean. The gray-shaded area indicates the range of the near-normal tercile based on the climatological period 1961–90. The numbers at the top of the graph indicate the correlation between the ensemble mean simulation and the observed anomalies ( $R$ ) and the tercile hit score ( $P$ ). (b) Distribution of forecast members for August–October 1997 (open green bars) relative to the climatological distribution (solid blue bars). From Mason et al., 1999

# CATEGORICAL PROBABILISTIC FORECAST

echam.pst 97 aso pcp 10 FCSTS run aug97  
 R = 0.62, P = 0.56, "INDONESIA": 10S20N ; 95E140E

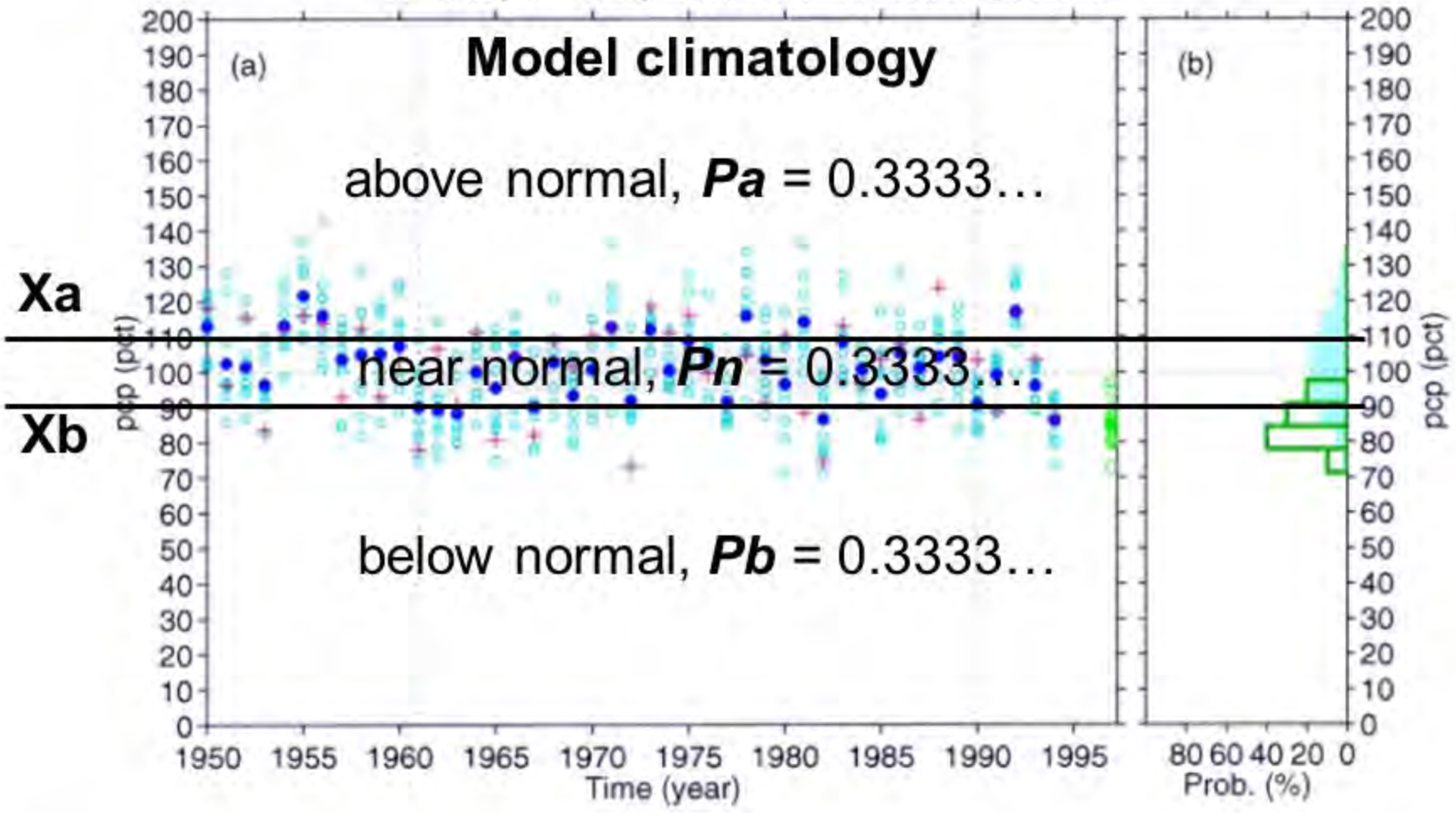


FIG. 6. (a) Historical performance of the ECHAM3 model for the August–October season compared to observations averaged over the “Indonesia” region (10°S–20°N, 95°–140°E). The open blue circles show the model anomaly for individual ensemble members (expressed as a percentage of long-term mean), solid blue circles show the ensemble mean, and red crosses indicate the observed anomaly. The green circles are for the current forecast, and the solid green circle represents the ensemble mean. The gray-shaded area indicates the range of the near-normal tercile based on the climatological period 1961–90. The numbers at the top of the graph indicate the correlation between the ensemble mean simulation and the observed anomalies ( $R$ ) and the tercile hit score ( $P$ ). (b) Distribution of forecast members for August–October 1997 (open green bars) relative to the climatological distribution (solid blue bars). From Mason et al., 1999

# CATEGORICAL PROBABILISTIC FORECAST

echam.pst 97 aso pcp 10 FCSTS run aug97  
 $R = 0.62$ ,  $P = 0.56$ , "INDONESIA": 10S20N ; 95E140E

## Forecast

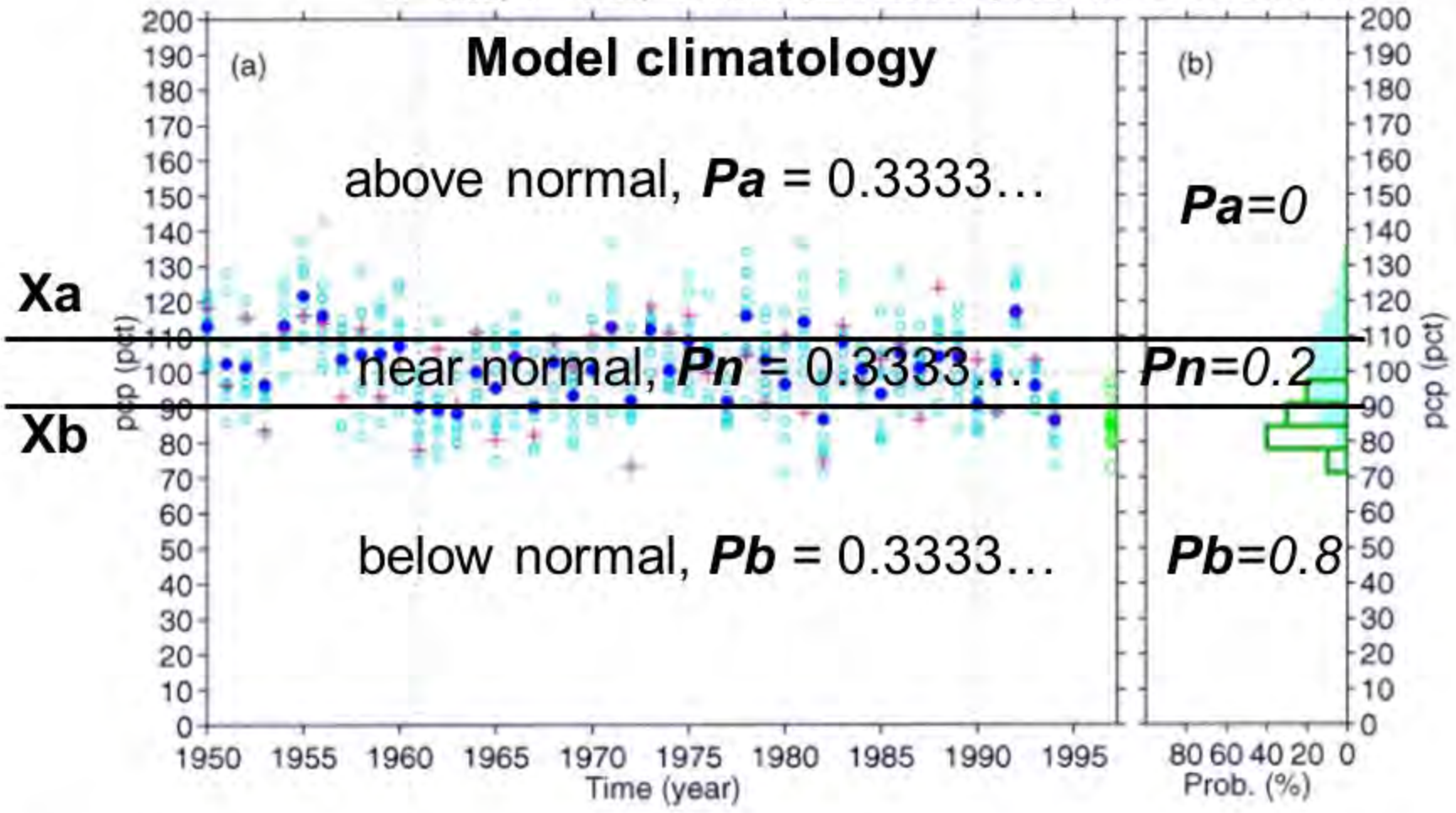


FIG. 6. (a) Historical performance of the ECHAM3 model for the August–October season compared to observations averaged over the “Indonesia” region (10°S–20°N, 95°–140°E). The open blue circles show the model anomaly for individual ensemble members (expressed as a percentage of long-term mean), solid blue circles show the ensemble mean, and red crosses indicate the observed anomaly. The green circles are for the current forecast, and the solid green circle represents the ensemble mean. The gray-shaded area indicates the range of the near-normal tercile based on the climatological period 1961–90. The numbers at the top of the graph indicate the correlation between the ensemble mean simulation and the observed anomalies ( $R$ ) and the tercile hit score ( $P$ ). (b) Distribution of forecast members for August–October 1997 (open green bars) relative to the climatological distribution (solid blue bars). From Mason et al., 1999

## EMPIRICAL PROBABILITY DISTRIBUTION

How to estimate probabilities of the categories below, near, above normal?

$$P_b = m_b/n$$

$$P_n = m_n/n$$

$$P_a = m_a/n$$

where  $n$  is FORECAST sample size;

$m_b$ ,  $m_n$ ,  $m_a$  – numbers of FORECAST ensemble members  $x$  such that

$$\begin{array}{l} x \leq X_b \\ m_b \end{array}$$

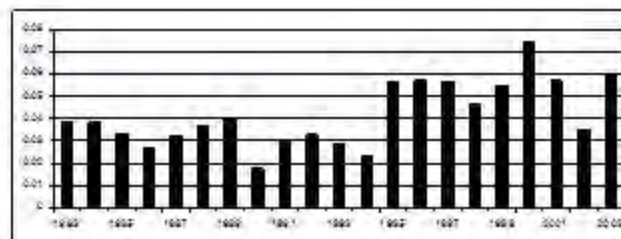
$$\begin{array}{l} X_b < x \leq X_a \\ m_n \end{array}$$

$$\begin{array}{l} X_a < x \\ m_a \end{array}$$

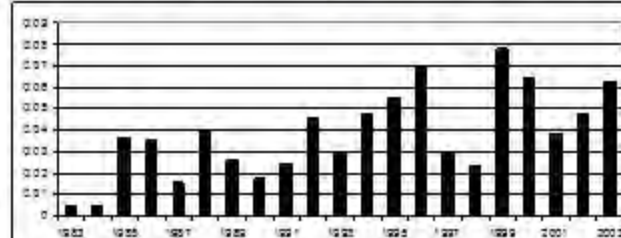
# GAUSSIAN PROBABILITY DISTRIBUTION Vs. EMPIRICAL PROBABILITY DISTRIBUTION

Показатель мастерства вероятностного прогноза в ранжированных категориях для глобального вероятностного прогноза, рассчитанного с использованием нормального распределения, по отношению к прогнозу, рассчитанному на основе эмпирического распределения. T850, зима, 1983 – 2003 гг.

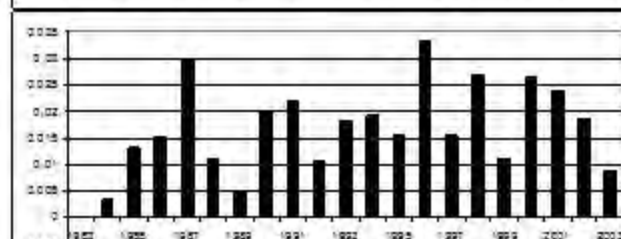
Размер ансамбля  
5



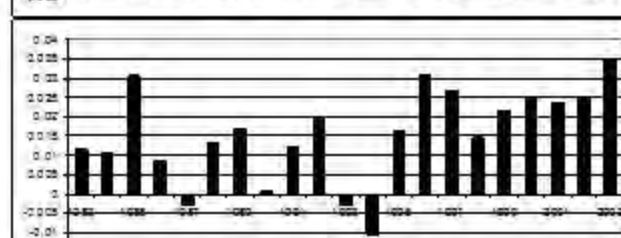
6



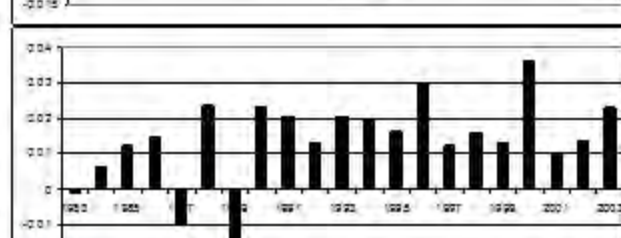
10



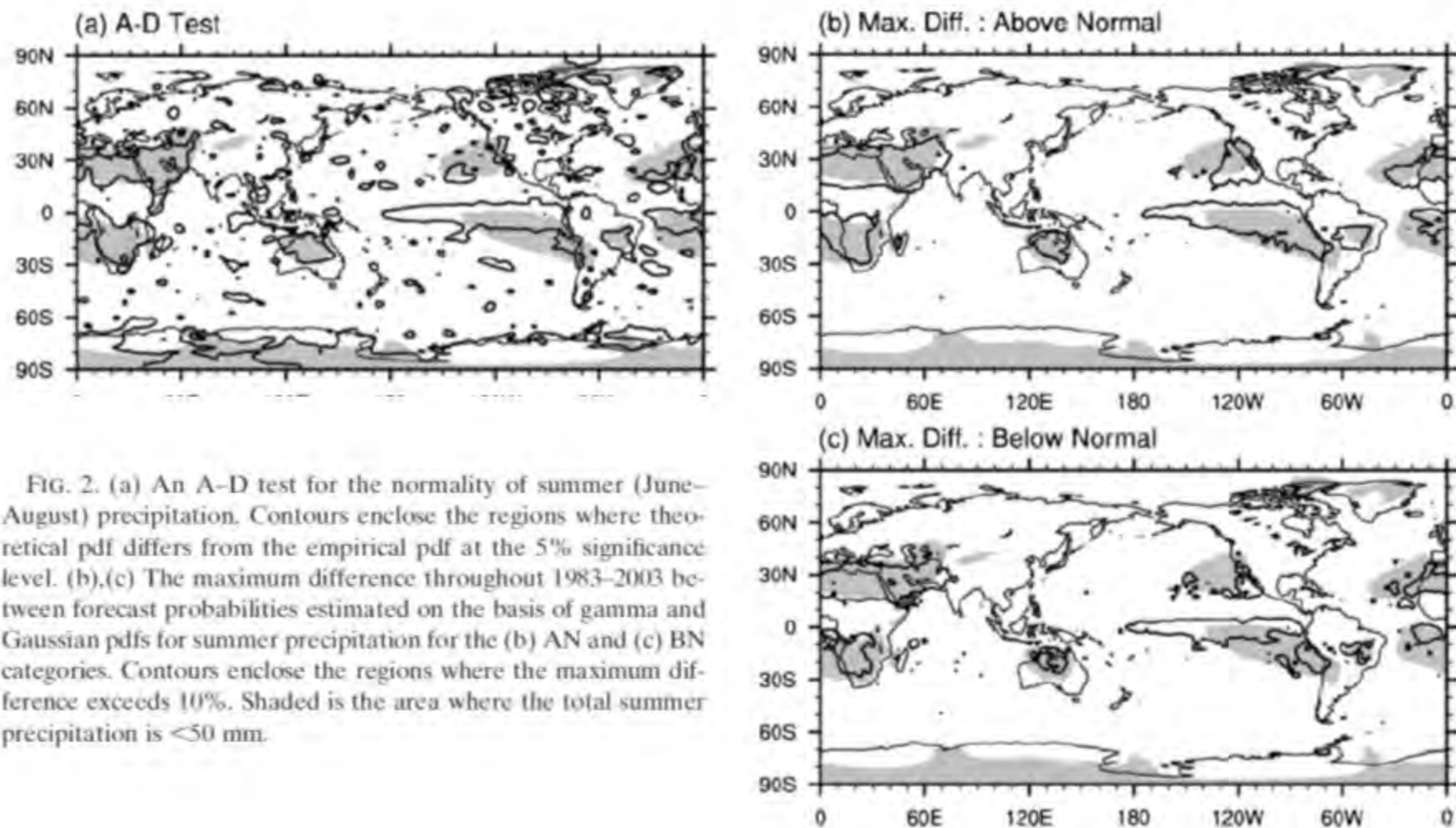
15



20

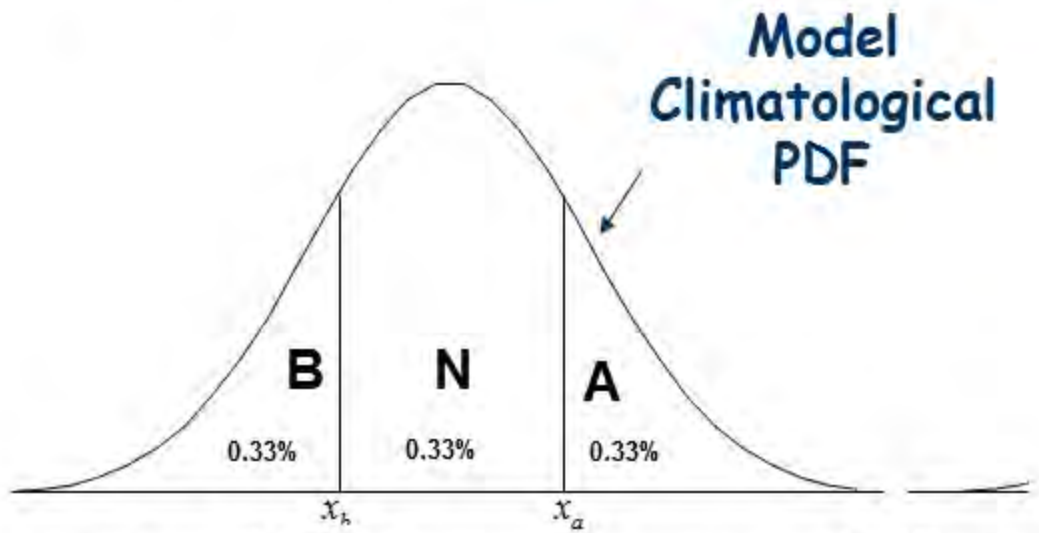


# GAUSSIAN PROBABILITY DISTRIBUTION



From Min et al., 2009

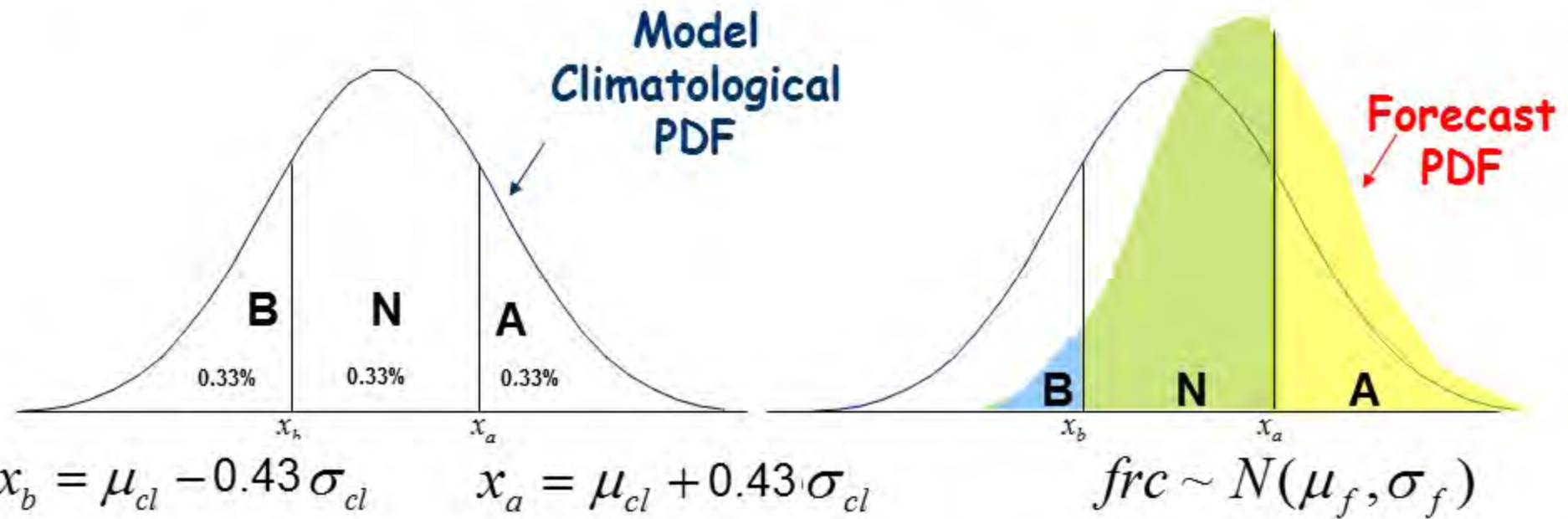
# GAUSSIAN PROBABILITY DISTRIBUTION



$$x_b = \mu_{cl} - 0.43 \sigma_{cl}$$

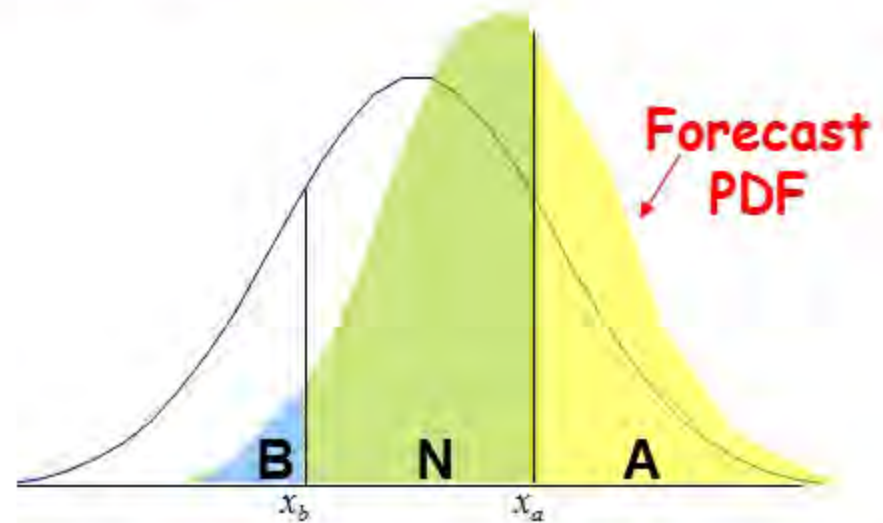
$$x_a = \mu_{cl} + 0.43 \sigma_{cl}$$

# GAUSSIAN PROBABILITY DISTRIBUTION





# GAUSSIAN PROBABILITY DISTRIBUTION



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

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

$$P(AN) = \text{prob}[x_a < x] = 1 - P(BN) - P(NN),$$

Where  $f(x)$  is a predicted Gaussian Probability Density Function:

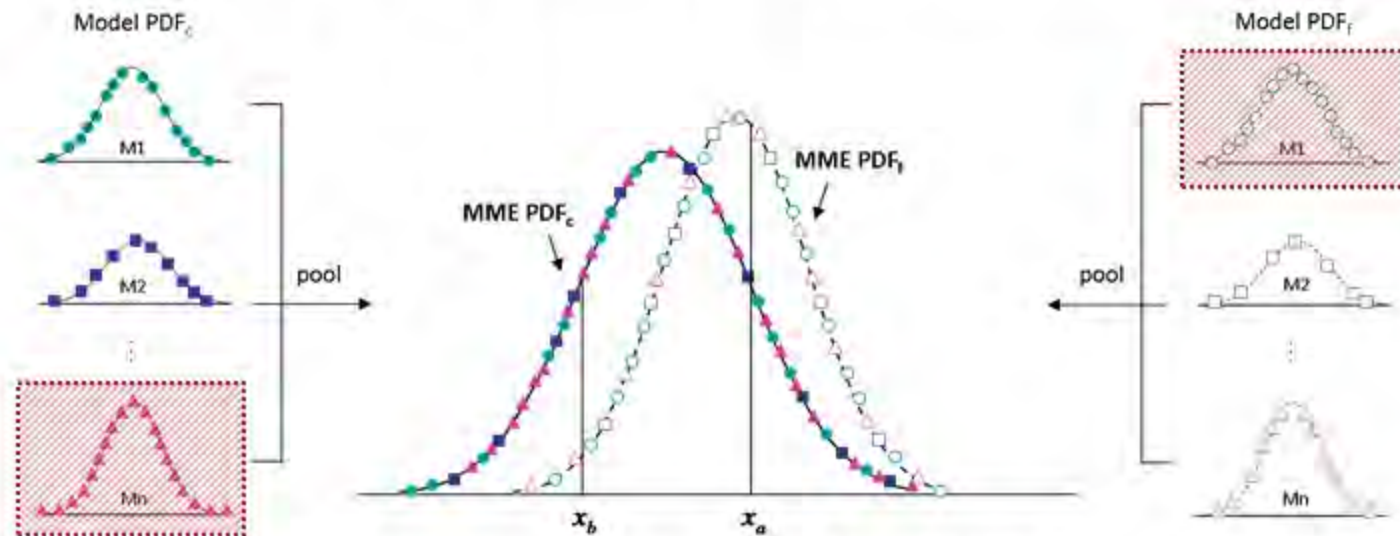
$$f(x) = \frac{1}{\sigma_f \sqrt{2\pi}} \exp\left(-\frac{(x - \mu_f)^2}{2\sigma_f^2}\right)$$



## **Мультимодельный прогноз**



# POOLING – all the ensemble members from all the models are combine in a single sample which is considered as a single model ensemble forecast



a single superensemble with equal weight of each ensemble member  
→ model weights proportional to their ensemble size

- **Strengths**: transparent, clear, large ensemble size of the forecast sample

- **Weakness**: incorrect result if there is an inconsistency between model weights in climatological and forecast PDFs

# Example of multimodel ensemble with inconsistent weights of the models in the hindcast and forecast data sets

**Model 1 (JMA):**

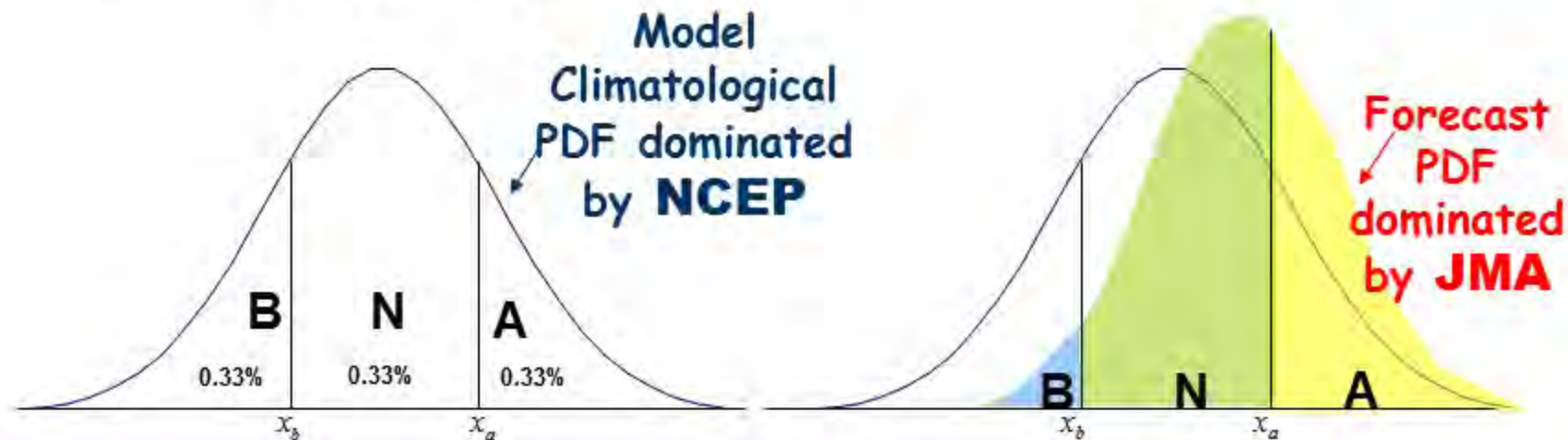
Hindcast - 5 ensemble members,

Forecast - 31 ensemble members

**Model 2 (NCEP):**

Hindcast - 15 ensemble members,

Forecast - 15 ensemble members



---

**The main restriction to the use of pooling:**

It requires **CONSISTENCY** between model weights (ensemble sizes) in the forecast and hindcast data sets.

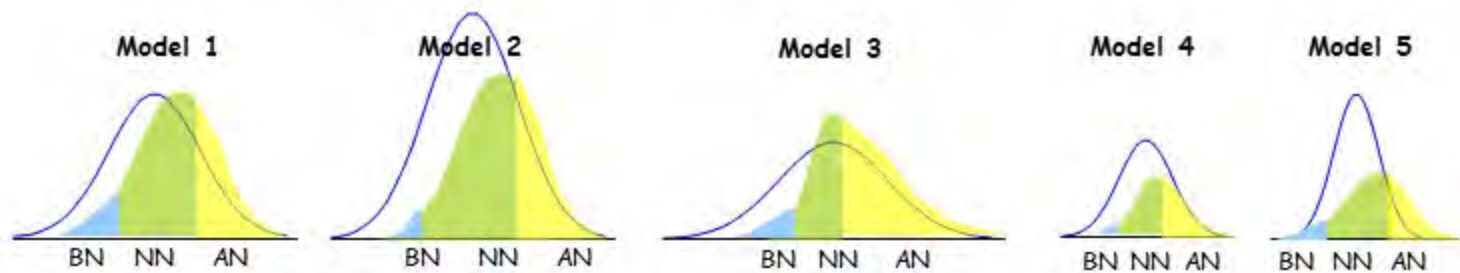
$$n\_hnd_i = k_i * n\_frc_i$$

with  $k_i$  being any positive real number equal for all the participating models

**Otherwise, forecast will be incorrect!**

---

## Total probability formula with unequal weights (general case)



“ Probabilistic Multi-Model Ensemble “

$$P(E) = \sum_{i=1}^M P(\text{Model}_i) \times P(E / \text{Model}_i)$$

model weight

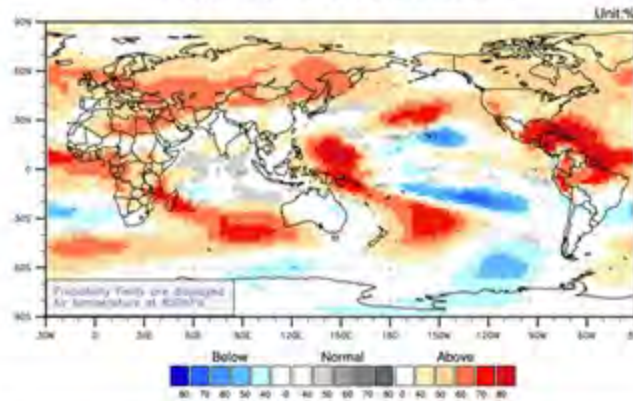
forecast probability of an event

**APCC: model weight is inversely proportional to the random error in probability associated with (sampling) standard error of ensemble mean**

$$P(E) = \frac{1}{\sum_{i=1}^M \sqrt{n_i}} \sum_{i=1}^M \sqrt{n_i} P(E / \text{Model}_i) \rightarrow \text{Model weights proportional to the square root of the ensemble size of } \text{Model}_i$$

## (a) APEC Climate Center

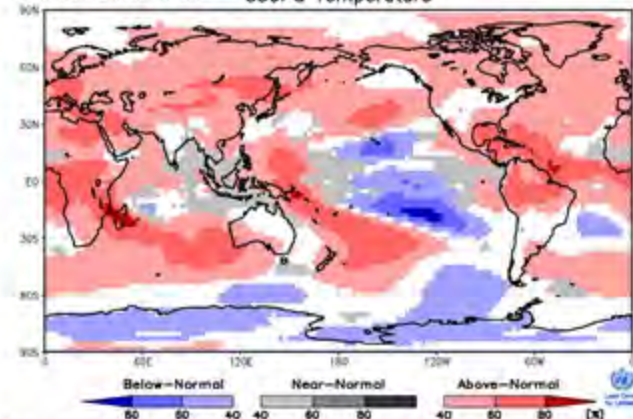
Temperature for July-September 2011



© APEC Climate Center

## (b) WMO Lead Center

Probabilistic Multi-Model Ensemble Forecast  
/GPCR\_sens/GPCR\_westpac/GPC\_nhbourse/GPC\_skye/GPC\_aster/GPC\_noratl\_pac/GPC\_noratl\_atl  
/GPC\_noratl\_pam3/GPC\_noratl\_pam2/GPC\_noratl\_pam1/GPC\_noratl\_pam0  
forecast time=2011:7: 2011:9: 850Pa Temperature



Probabilistic multi-model ensemble prediction maps for 2011JAS (July-August-September) temperature at (a) APEC Climate Center (available online at <http://www.apcc21.org>) and (b) (b) WMO Lead Center (available online at <http://wmoic.org>).

## Что стоит запомнить:

1. Начинайте с оценки и устранения смещения климата.  
Имейте в виду, чем больше смещение, тем больше вероятность ошибочного прогноза даже после устранения смещения
2. В ансамблевом прогнозе смысл имеют только среднее и разброс, но не отдельные члены ансамбля из-за большой случайной ошибки.  
Чем больше ансамбль, тем меньше случайная ошибка среднего
3. При вероятностном прогнозе лучше аппроксимировать ансамблевое распределение вероятностей нормальным распределением
4. При составлении мультимодельного прогноза обратите внимание на пропорциональность размеров модельных ансамблей в исторических и реальных прогнозах
5. Прогностические карты терцильных прогнозов имеют 4 цвета – 3 цвета – вероятности терцильных категорий  
и белый цвет – области неопределенного прогноза





**Thank You**

