Spatial aggregation of ROC and PR curves

Romain Pic[1,2] Zhongwei Zhang[3] Johanna Ziegel[1] Sebastian Engelke[2]

[1]ETH Zurich (Seminar for Statistics) [2]UNIGE (GSEM-RISIS) [3]KIT (Institute of Statistics)

1. Introduction

ROC and Precision–Recall (PR) curves are popular tools to assess the discrimination ability of forecasts for binary events.

- Occurrence/non-occurrence of an event;
- Exceedance of a threshold;
- *Absolute-extreme events* or *relative-extreme events*: exceedance of a fixed or spatially varying threshold.

Aggregating ROC/PR across spatial locations curves provides a comparison of several competing forecasts and summarizes the information. **However, when done improperly, aggregation can lead to misleading conclusions.**

2. Background

- Continuous forecast $X \in \mathbb{R}$, binary observation $Y \in \{0, 1\}$
- $(x_i, y_i), i = 1, \dots, n$, n realizations of the forecast-observation pair (X, Y)

For a decision threshold t_X, the comparison between forecasts and observations is summarized by the counts of the contingency matrix.

		Observation	
		$Y = 1$	$Y = 0$
Forecast	$X > t_X$	$a(t_X)$	$b(t_X)$
	$X \leq t_X$	$c(t_X)$	$d(t_X)$

Table 1. Contingency matrix.

- False alarm rate: $\mathrm{FAR}(t_X) = \frac{b(t_X)}{b(t_X)+d(t_X)}$
- Hit rate or recall: $\mathrm{HR}(t_X) = \mathrm{Re}(t_X) = \frac{a(t_X)}{a(t_X)+c(t_X)}$
- Precision: $\mathrm{Pre}(t_X) = \frac{a(t_X)}{a(t_X)+b(t_X)}$
- Frequency bias (FB): $\mathrm{FB}(t_X) = \frac{a(t_X)+b(t_X)}{a(t_X)+c(t_X)}$

Receiver Operating Characteristic (ROC) curves

- The raw ROC curve is obtained by linearly interpolating points of the form $(\mathrm{FAR}(t_X), \mathrm{HR}(t_X))$, where $t_X \in [-\infty, \infty]$.
- **Dominance:** ROC_A dominates ROC_B if, for all FAR, the HR of ROC_A is greater than or equal to that of ROC_B.
- **Concavity:** ROC curve should be concave for interpretation purposes (non-decreasing conditional event probability $\Pr(Y = 1 \mid X = x)$ w.r.t. x).

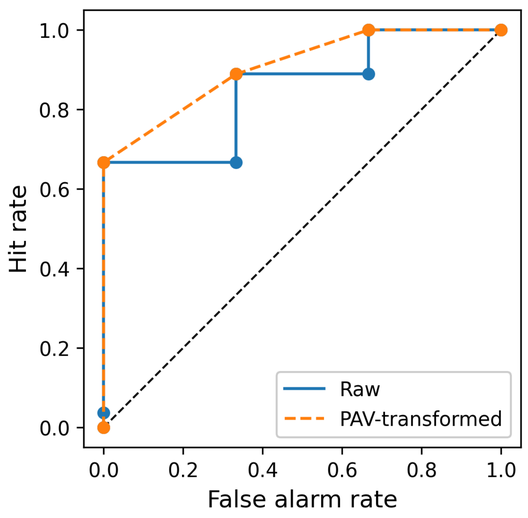

Figure 1. Effect of the PAV algorithm on a ROC curve.

Precision-Recall (PR) curves

- The raw PR curve is obtained by interpolating (**not linearly**) points of the form $(\mathrm{Re}(t_X), \mathrm{Pre}(t_X))$, where $t_X \in [-\infty, \infty]$.
- **Dominance:** PR_A dominates PR_B if, for all recalls, the precision of PR_A is greater than or equal to that of PR_B.
- **Achievability:** The counterpart of the concave hull of a ROC curve is the achievable PR curve [1].

Parameterizations.
A ROC (PR) curve can be obtained from a family of continuous contingency matrices,

$$\left\{ \begin{bmatrix} a(t) & b(t) \\ c(t) & d(t) \end{bmatrix}, t \in \bar{\mathbb{R}} \right\},$$

indexed by $t \in [-\infty, \infty]$ and taking real positive values (allowing for parameterizations beyond those based on t_X or FAR (recall)).

In practice, concave ROC curves and achievable PR curves can be obtained by the Pool-Adjacent-Violators (PAV) algorithm.

Romain Pic

romain.pic@math.ethz.ch

Preprint & Feedback

Equivalence between ROC and PR curves

Theorem [1].
Given a fixed sample size n, there is a one-to-one correspondence between the ROC curve and the PR curve associated with the same forecast-observation pair, except for the point with $\mathrm{Re} = \mathrm{HR} = 0$.

→ Dominance in terms of ROC curves ⇔ dominance in terms of PR curves

→ Concave ROC curve ⇔ achievable PR curve

3. Aggregation and parameterization strategies

Consider forecasts and observations over d locations.

Parameterization strategy

To aggregate multiple curves across locations, one must jointly define the parameters across locations $(t_1, \dots, t_d) \in \bar{\mathbb{R}}^d$, and, moreover, how they jointly vary.

Definition.
A *parameterization strategy* is a mapping $\gamma : [0, 1] \to \bar{\mathbb{R}}^d$ that is non-decreasing in each component, and satisfies the boundary conditions $\gamma(0) = (-\infty, \dots, -\infty)$ and $\gamma(1) = (\infty, \dots, \infty)$.

Examples:

- GraphCast [2]: $\gamma(u)_s = \mathrm{med}_s + \frac{1}{g(u)}(t_{Y,s} - \mathrm{med}_s)$, with med_s the climatological median of the observations at location s and $g(u) \in (0, \infty)$ a gain parameter.
- Location-scale paramaterization [3]: $\gamma(u)_s = t_{Y,s} - g(u) \cdot \sigma_{X,s}$, with $\sigma_{X,s}$ the climatological standard deviation of the forecast at location s and $g(u) \in \bar{\mathbb{R}}$ a gain parameter.
- FB parameterization.
- Parallel lines parameterization.

Aggregating counts

An aggregated contingency matrix is obtained by summing location-wise components, but only allowing for location-wise contingency matrices corresponding to **true counts**.

Warning.
It can lead to cases in which no parameterization strategy can simultaneously preserve dominance and concavity/achievability.

Aggregating interpolated counts

Aggregated contingency matrix, but allowing for **interpolated counts**.

$$\mathrm{FAR}(\gamma(u)) = \frac{\sum_{s=1}^{d} b_s(\gamma(u)_s)}{\sum_{s=1}^{d} b_s(\gamma(u)_s) + d_s(\gamma(u)_s)};$$
$$\mathrm{HR}(\gamma(u)) = \mathrm{Re}(\gamma(u)) = \frac{\sum_{s=1}^{d} a_s(\gamma(u)_s)}{\sum_{s=1}^{d} a_s(\gamma(u)_s) + c_s(\gamma(u)_s)};$$
$$\mathrm{Pre}(\gamma(u)) = \frac{\sum_{s=1}^{d} a_s(\gamma(u)_s)}{\sum_{s=1}^{d} a_s(\gamma(u)_s) + b_s(\gamma(u)_s)}.$$

4. Preservation of dominance

Theorem (ROC curves).
Let A and B be two forecasters. Let γ_A and γ_B be their parameterization strategies such that, for all $s \in \{1, \dots, d\}$ and $u \in [0, 1]$,

$$\begin{cases} \mathrm{FAR}_s^A(\gamma_A(u)_s) \leq \mathrm{FAR}_s^B(\gamma_B(u)_s) \\ \mathrm{HR}_s^A(\gamma_A(u)_s) \geq \mathrm{HR}_s^B(\gamma_B(u)_s) \end{cases}$$

and $\exists u^* \in [0, 1]$ such that one of the inequalities is strict. Then, aggregating interpolated counts preserves dominance in terms of ROC curves.

Theorem (PR curves).
Let γ_A and γ_B satisfy

$$\begin{cases} \mathrm{Re}_s^A(\gamma_A(u)_s) \geq \mathrm{Re}_s^B(\gamma_B(u)_s) \\ \mathrm{Pre}_s^A(\gamma_A(u)_s) \geq \mathrm{Pre}_s^B(\gamma_B(u)_s) \\ \mathrm{FB}_s^A(\gamma_A(u)_s) = \mathrm{FB}_s^B(\gamma_B(u)_s) \end{cases}$$

and $\exists u^* \in [0, 1]$ such that one of the inequalities is strict, for all $s \in \{1, \dots, d\}$. Then, aggregating interpolated counts preserves dominance in terms of PR curves.

→ The FB and parallel lines parameterizations satisfy both conditions.

5. Preservation of concavity/achievability

Theorem.
Let location-wise ROC (PR) be concave (achievable) curves and γ be a parameterization strategy. A necessary and sufficient condition for aggregating interpolated counts to preserve concavity (achievability) is that the aggregated conditional event probability is non-decreasing in the central parameter $u \in [0, 1]$.

→ The FB and parallel lines parameterizations satisfy this condition.

6. Application to WeatherBench 2

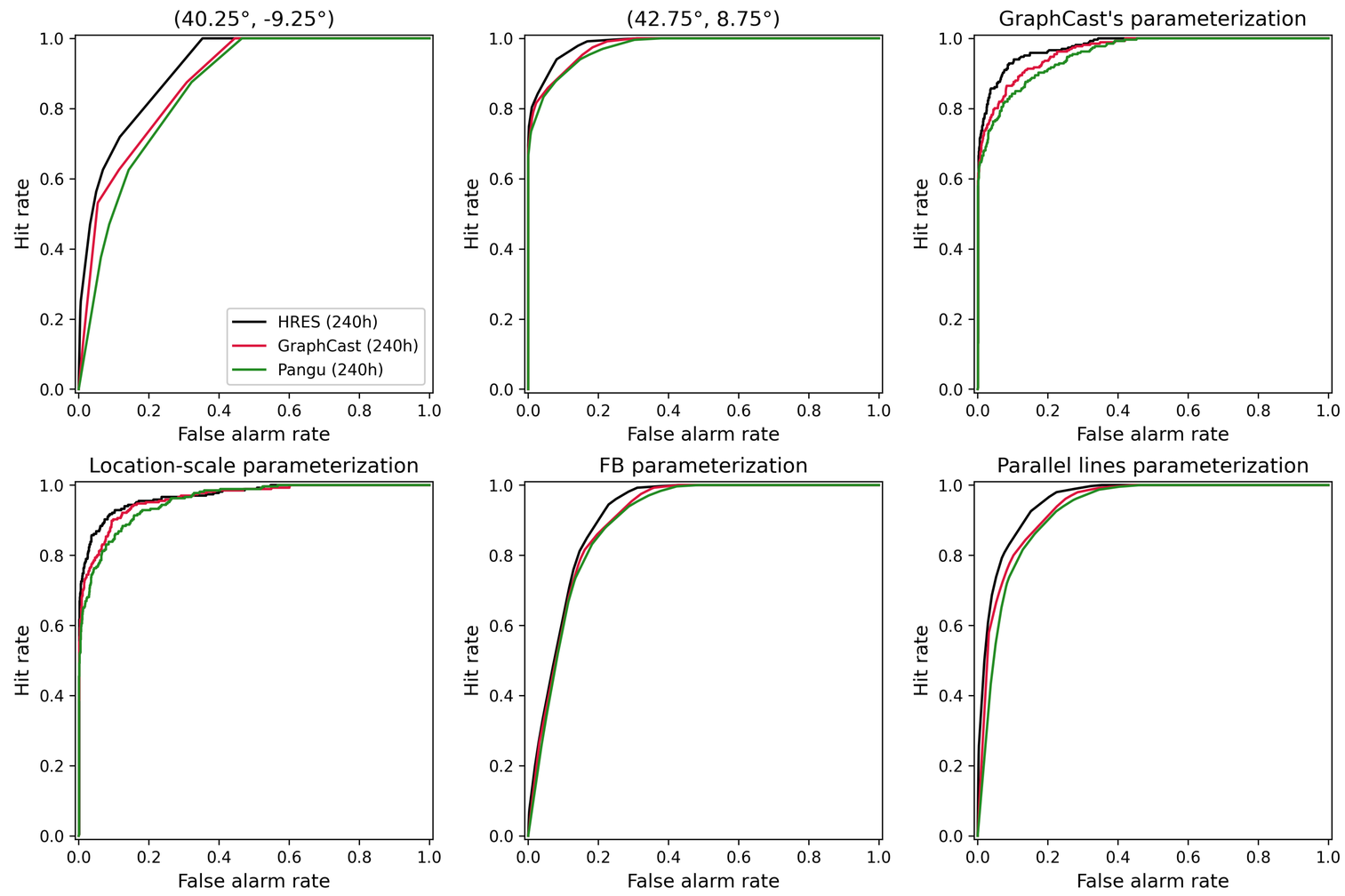

Figure 2. Location-wise ROC curves of HRES, GraphCast, and Pangu with a lead time of 240h and their aggregated ROC curves for multiple parameterization strategies.

Conclusion

- Preservation of dominance and concavity/achievability are desirable properties when aggregating ROC and PR curves.
- Seemingly intuitive parameterization strategies can yield misleading conclusions.
- **Warning** – Preserving dominance or concavity/achievability does not entail the converse.
- **Warning** – Preserving dominance and concavity/achievability is not sufficient to provide desirable information (see FB parameterization).

Recommendations

- Single location: Only concave ROC curves/achievable PR curves should be compared → PAV-transformed forecasts
- Multiple locations: PAV-transformed forecasts, ROC (PR) curves parameterized by FB or parallel lines → aggregating interpolated counts using $\gamma(u)_s = u$.

→ **R. Pic, Z. Zhang, J. Ziegel, S. Engelke. "Spatial aggregation of ROC and Precision-Recall curves" (*in preparation*)**

[1] J. Davis and M. Goadrich. "The relationship between Precision-Recall and ROC curves". In: *Proceedings of the 23rd ICML*. 2006, pp. 233–240.

[2] R. Lam et al. "Learning skillful medium-range global weather forecasting". In: *Science* 382.6677 (2023).

[3] Z. Zhang et al. *Numerical models outperform AI weather forecasts of record-breaking extremes*. 2025. arXiv: 2508.15724.