Cherry Blossom Peak Bloom Prediction 2026

Team 5103 — University of Maryland

2026-02-01

Datasets Overview

Competition datasets — bloom DOY for 5 target sites:

Dataset	Location	Records	Span
`kyoto.csv`	Kyoto, Japan	837	812 – 2025
`washingtondc.csv`	Washington DC	106	1921 – 2025
`liestal.csv`	Liestal, Switzerland	132	1895 – 2025
`vancouver.csv`	Vancouver, Canada	4	2022 – 2025
`nyc.csv`	New York City	3	2019 – 2025

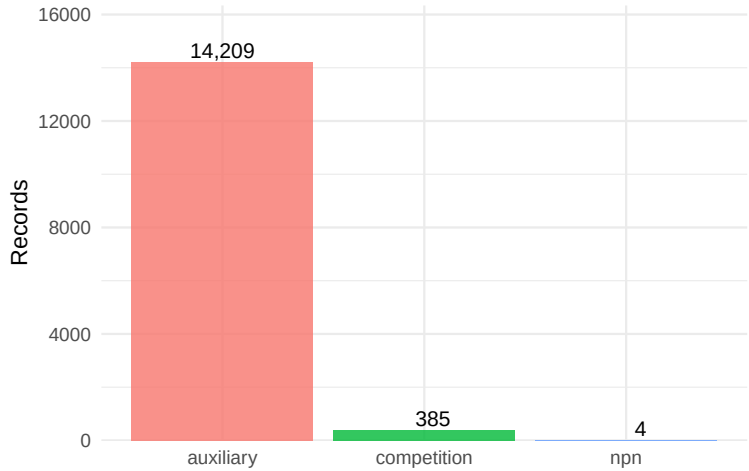
Auxiliary datasets — broaden geographic & temporal coverage:

Dataset	Re
`japan.csv` (regional bloom dates)	
`meteoswiss.csv` (Swiss phenology)	
`south_korea.csv`	

USA-NPN enrichment (NYC):

- Site 32789 (Washington Square Park)
- Species 228 (*Prunus* × *yedoensis*)
- Phenophase 501 (Open flowers)
- 5 extra bloom-year records added

Total training pool: ~14,598 rows



Models & Methodology

Model A — Local Trend (per site)

- Recency-weighted quadratic: `bloom_doy ~ year + year²`
- Weights: $w_i = e^{(i-n)/6}$, half-life 6 yr
- Fallback: linear (2–3 obs) or mean (1 obs)
- Captures site-specific momentum

Model B — Pooled GAM (all sites jointly)

$$\text{DOY} \sim s(\text{year}) + s(\text{lat}, \text{long}) + s(\text{alt}) + s(\text{site\_obs}) + \text{source}$$

- REML estimation on **14,598** records

Ensemble blending (data-driven)

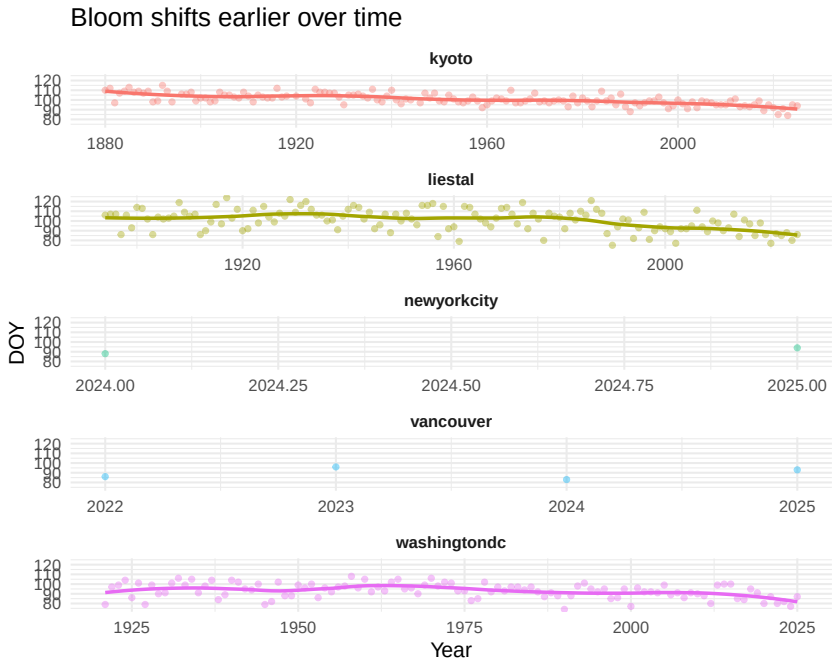
- Rolling-origin backtest (1900 – 2025)
- Inverse-MAE weights from out-of-sample errors:

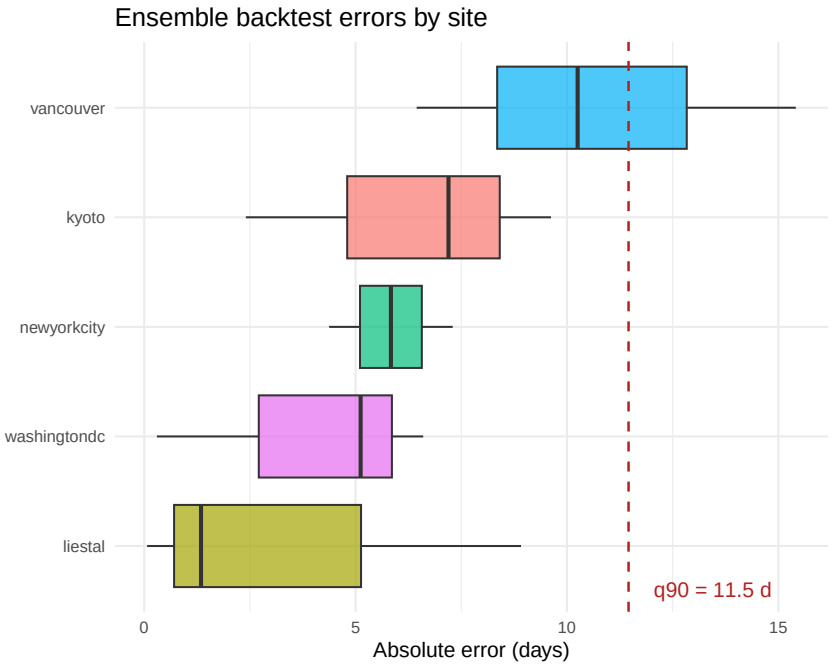
$$w_A = \frac{1/\text{MAE}_A}{1/\text{MAE}_A + 1/\text{MAE}_B}$$

Model	Backtest MAE
Local (A)	7.01 days
GAM (B)	7.21 days
Ensemble	**6.1 days**

Prediction intervals — split-conformal: 90th-percentile of backtest |residuals| per

Backtest Performance & EDA

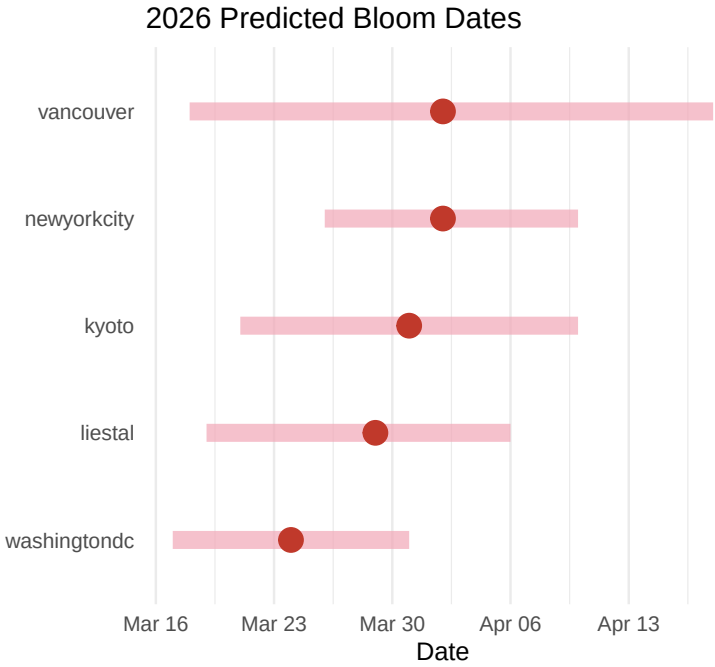




- All 5 sites show a **downward bloom-DOY trend** — climate warming signal
- Ensemble achieves **6.1-day MAE** on rolling held-out years
- R and Python pipelines agree within **1.8 days** on average → blended submission

Final 2026 Predictions

Location	DOY	Date	
kyoto	90	Mar 31	Mar 21
liestal	88	Mar 29	Mar 19
newyorkcity	92	Apr 02	Mar 26
vancouver	92	Apr 02	Mar 18
washingtondc	83	Mar 24	Mar 17



Sum of squared interval widths:
2106

Metric	Value
Ensemble backtest MAE	6.1 days
R vs Python mean gap	1.8 days
Local weight (w_A)	50.7%
GAM weight (w_B)	49.3%

Thank You

Cherry Blossom Peak Bloom Prediction 2026

Team 5103 — University of Maryland

> **Tip**
>
> All code, data, and outputs are publicly available and fully reproducible.

```
quarto render solution.qmd                                    # H
jupyter nbconvert --execute Solution.ipynb --inplace           #
```