INTERNATIONAL CONFERENCE ON PARALLEL PROCESSING

PhenoTracker: A machine learning model to track grape phenology

ICPP HARVEST 2025

First International Workshop on Applications of HPC and AI in Agriculture

September 10

Nathan Balcarcel[1], **Paola Pesantez-Cabrera[1]**, Kristen Goebel[2], Markus Keller[1], Lav Khot[1], Alan Fern[2], Ananth Kalyanaraman[1]

1 Washington State University

2 Oregon State University

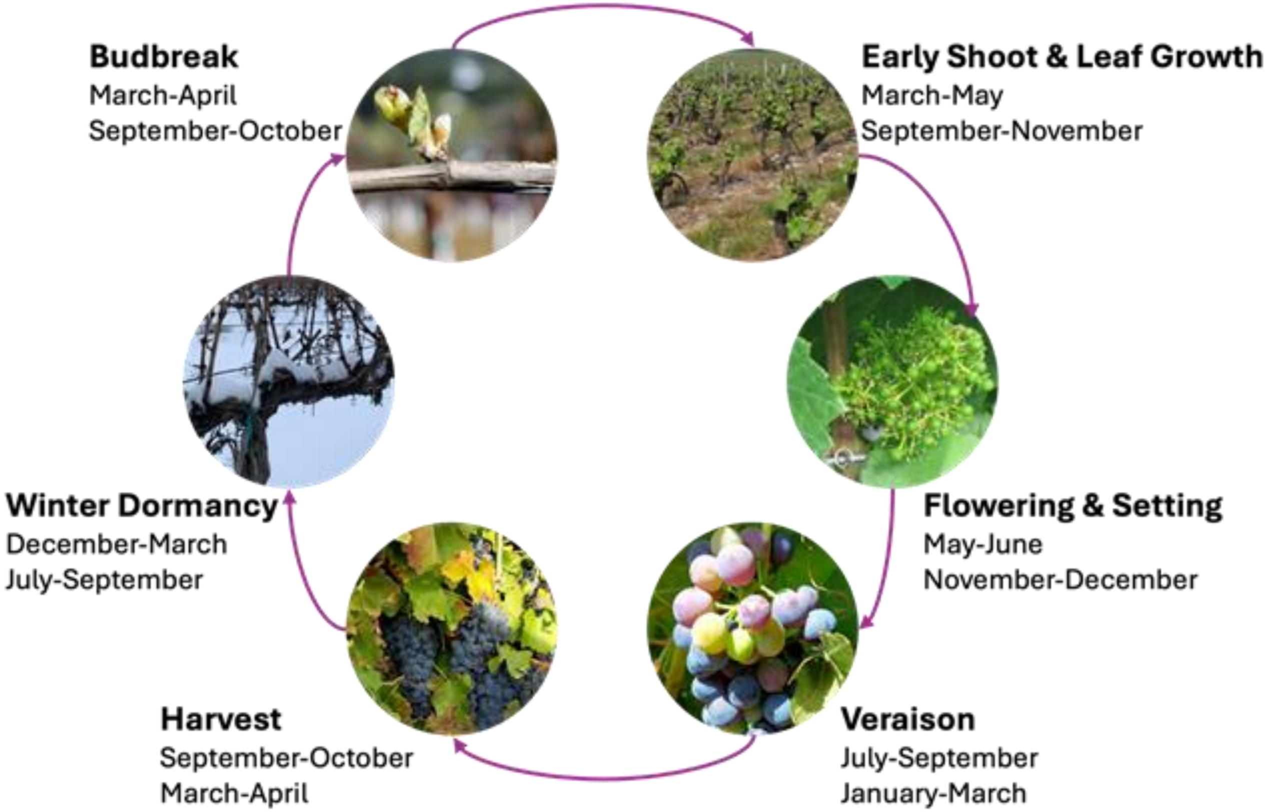
USDA National Institute of Food and Agriculture
United States Department of Agriculture
(funded as part of the National AI Research Institutes program NSF 20-604)

Crop Phenology

- **What's** phenology?
  - Annual developmental cycle
  - Driven by **environmental** and **internal** genotype factors

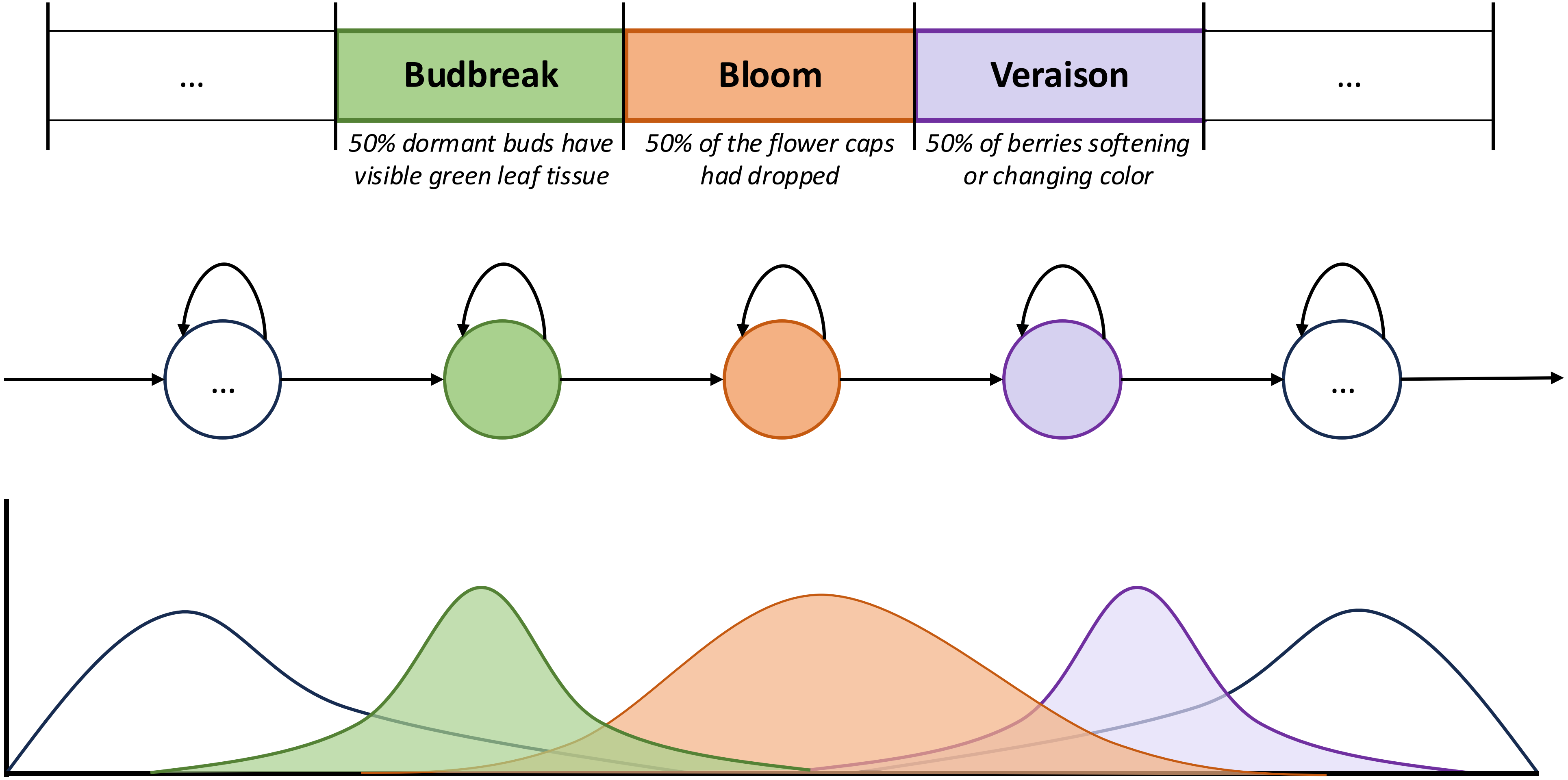
- **Why** predict phenological stages?
  - **Accurate** forecasting ➡ **Supports** timely management and mitigation strategies ➡ **Prevent** crop loss
  - **Goal:** Track developmental progress

- Key predictive **challenges**:
  - **Limited** data
  - Cultivar **variability**
  - Seasonal **dependency**



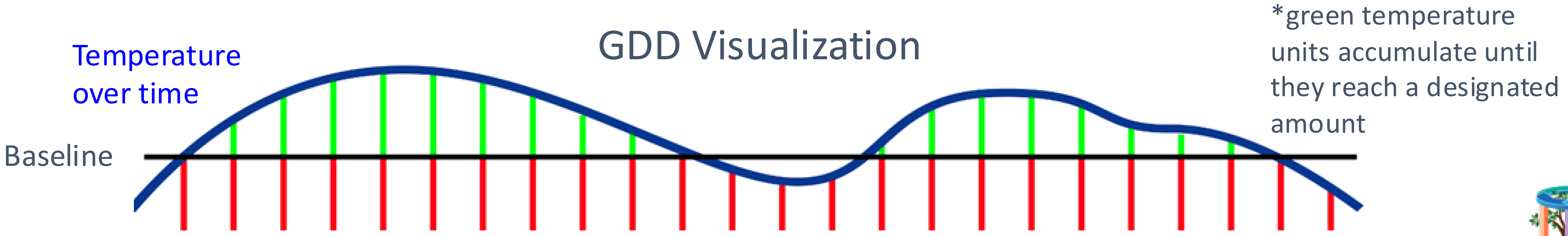
Example of grapevine phenology

Tracking phenological stages: Problem abstraction



Process-Based Approaches

- Past solutions [Parker et al. 2011 and Zapata et al. 2017] focus primarily on growing degree days (GDD)
  - Accumulate heat units above a baseline
  - Use fixed start date
  - Once the sum reaches a certain threshold is when an stage is predicted to take place
  - Only factor in temperature, which paints an incomplete picture
  - Individual cultivars are handled entirely independently, no internal overlap



PhenoTracker Model Overview

Weather Data
- air temperature
- relative humidity
- dew point
- precipitation
- wind speed

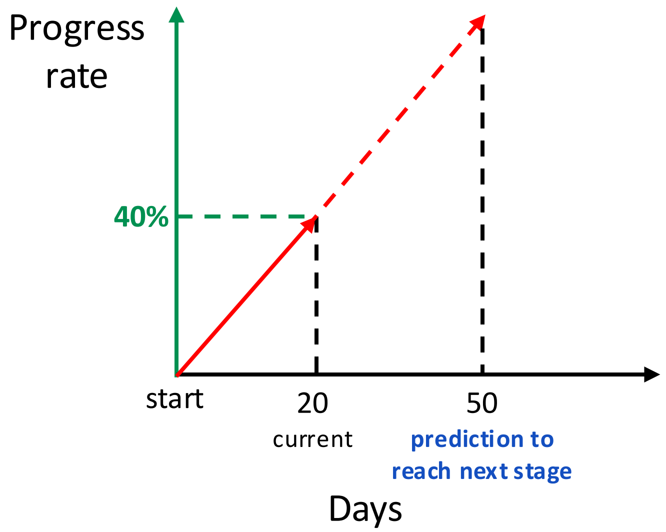
Cultivar ID

Predicted LTE [optional] (from GrapeHardiNET model)

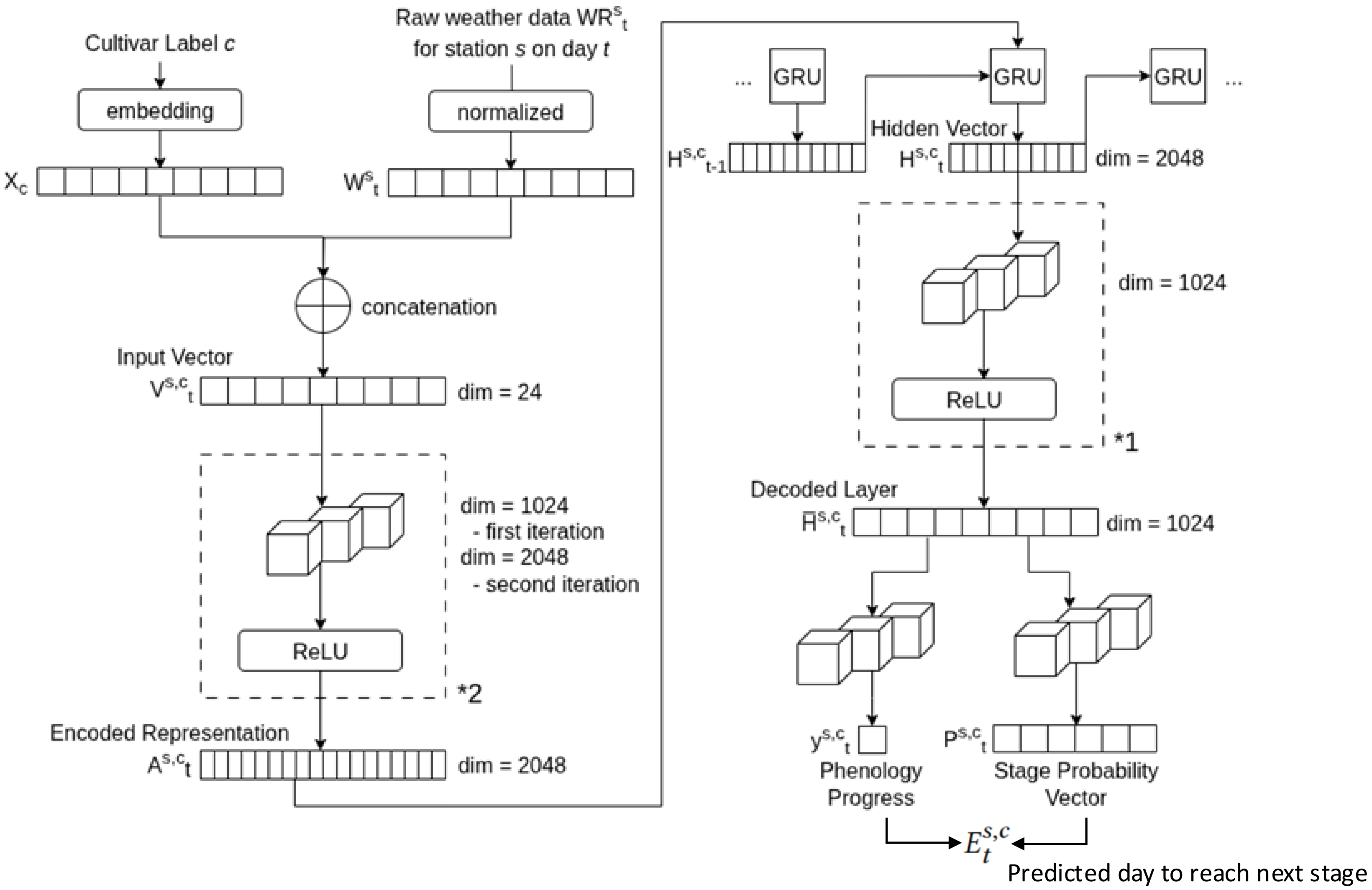
Lowest Chill [optional]

GRU Model

1. **Stage Probability** (most likely stage on any given day)
2. **Phenology Progress** (measured in % of days in current state)



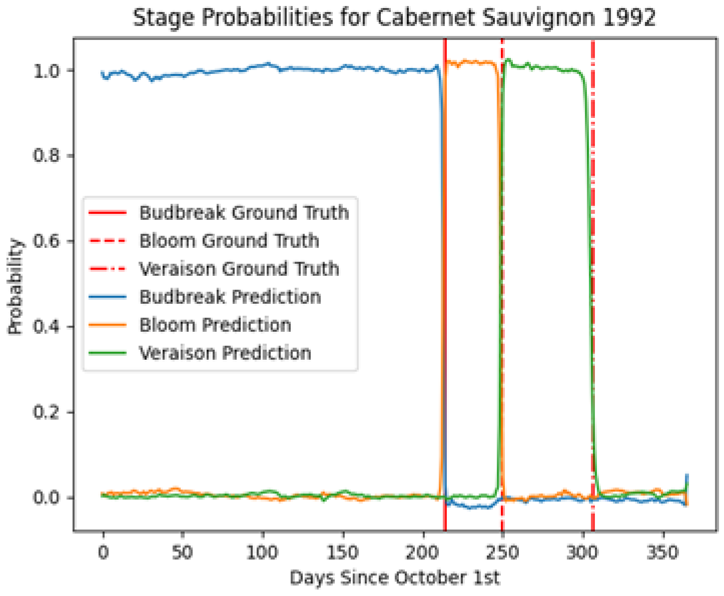
Method: Model Architecture

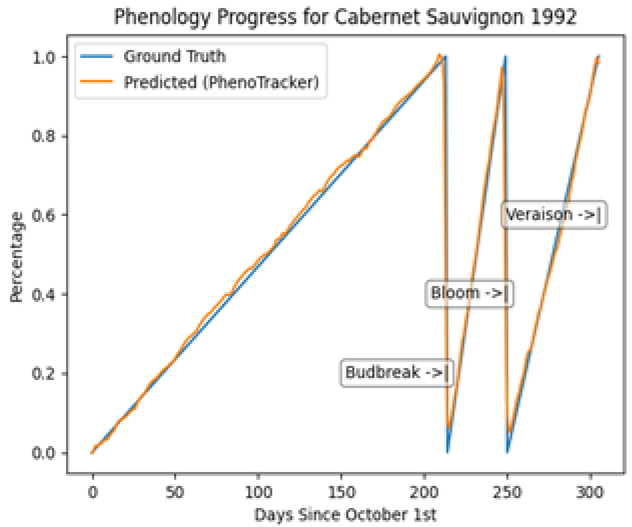


Method: Output Example

- Each season (*year*) is split into stages
  - Each stage represents the time leading up its respective event

- Prediction variables
  - **Stage probability**: one-hot encoding set to 1.0 if corresponding stage is active and 0.0 otherwise
  - **Phenology Progress**: increases linearly from 0.0 to 1.0 from the start of a stage until the end. Predicted in real-time daily
    - **Linearity** for phenology progress is just an assumption





Experimental Setup: Data

- One season starts on October 1st and ends on September 30th

- Phenology data
  - Real-world dataset from the vineyards of the WSU IAREC, Prosser, WA
  - Up to 20 genetically diverse cultivars/genotypes
  - Recorded since 1988

- WSU AgWeatherNet
  - Parker and Zapata model predictions
  - Meteorological/environmental daily data from two weather stations
    - Prosser.NE
    - Roza.2

Cultivar	Start Season	End Season	No. Years Data
Barbera	2015-2016	2022-2023	7
Cabernet Franc	1989-1990	2022-2023	20
Cabernet Sauvignon	1989-1990	2022-2023	21
Chardonnay	1989-1990	2022-2023	25
Chenin Blanc	1990-1991	2022-2023	19
Concord	1991-1992	2020-2021	19
Gewurztraminer	1992-1993	2022-2023	18
Grenache	1992-1993	2022-2023	16
Lemberger	1989-1990	2022-2023	18
Malbec	1989-1990	2022-2023	17
Merlot	1989-1990	2022-2023	25
Mourvedre	2015-2016	2022-2023	7
Nebbiolo	2015-2016	2022-2023	7
Pinot Gris	1992-1993	2022-2023	21
Riesling	1990-1991	2022-2023	19
Sangiovese	2015-2016	2022-2023	6
Sauvignon Blanc	2004-2005	2022-2023	10
Semillon	1990-1991	2022-2023	20
Viognier	2015-2016	2022-2023	8
Zinfandel	1992-1993	2022-2023	16

Seasons used for each cultivar

Experimental Setup: Training & Testing

- Weather data normalized based on their respective minimums and maximums

- For each cultivar with **n seasons**
  - Validation: 2 seasons
  - Training: n-2 seasons

- Model was retrained **9** independent times
  - Choosing a different set of validation seasons each time
  - Leading to **18** validation seasons per cultivar

- Note that each day in the season has its own predicted $E_t^{s,c}$
  - It is simply a date
  - No associated stage
  - Must be used with the stage vector to determine which stage the predicted date is for

Experimental Evaluation

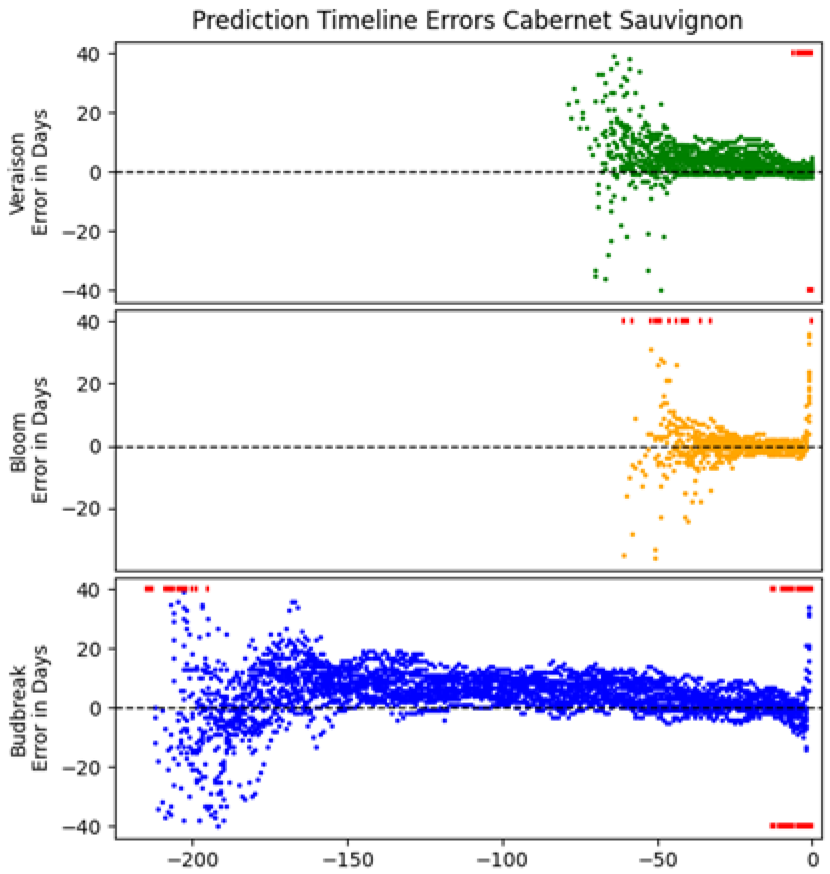
- PhenoTracker was analyzed **individually** as well as **comparatively** using root mean square error (**RMSE**)

- Models were compared by the **error in days** of their predictions for an stage’s date vs the ground-truth

- Results are focused on 4 commonly known cultivars:
    - **Red varieties**: Cabernet Sauvignon, Merlot
    - **White varieties**: Chardonnay, Riesling

Results: Prediction Error

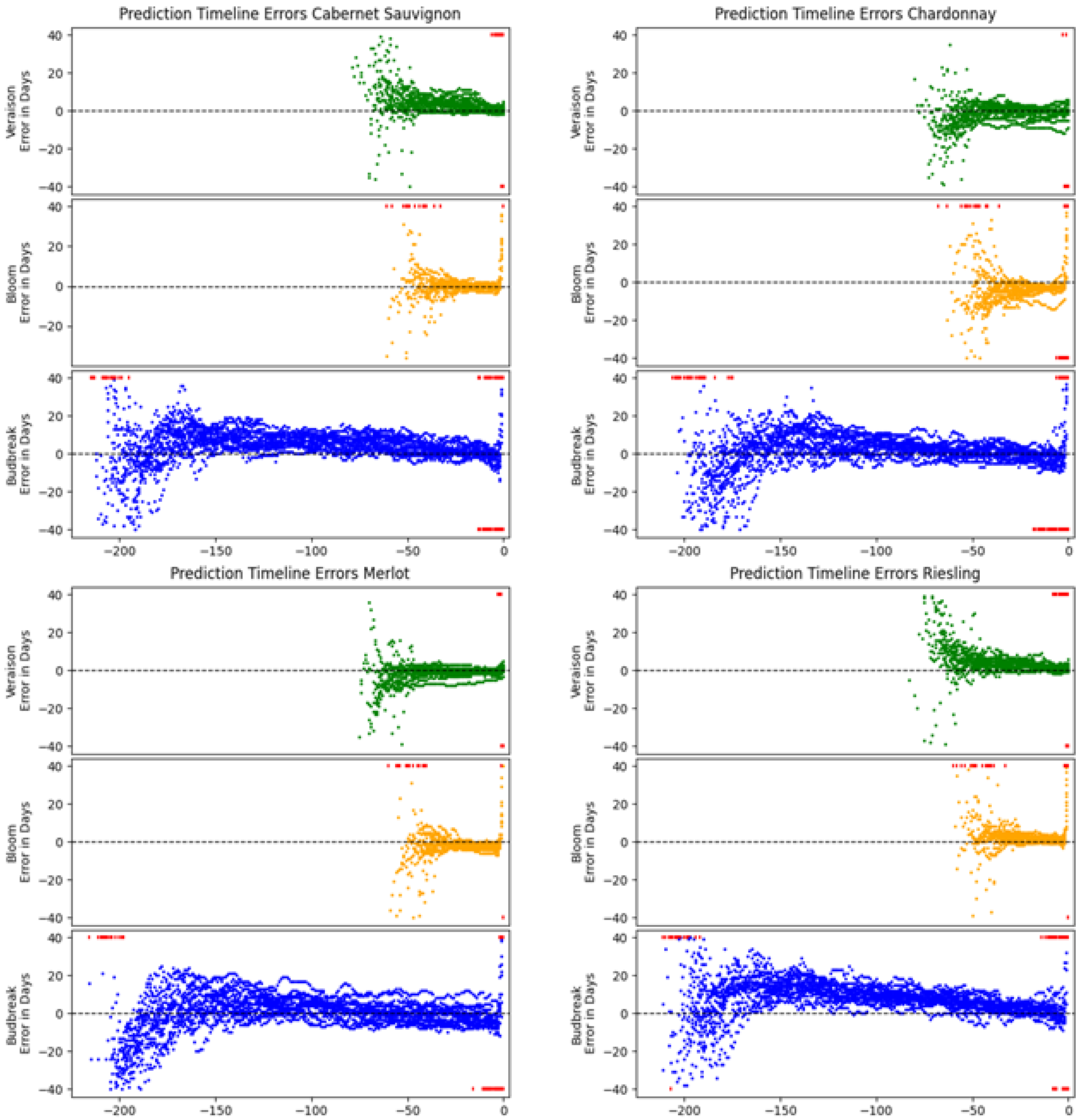
- Certain cultivar/stage combinations appear to be slightly skewed in different ways
- The earlier the prediction, the larger the error. Variance decreases and begins to converge around the true date
- Winter weather provides significant insights to help improve budbreak prediction
- The sooner this precision can be achieved, the earlier the farmer can start planning



Day 0 (far right) is the true day of stage transition

Results: Prediction Error

- Certain cultivar/stage combinations appear to be slightly skewed in different ways
    - **Budbreak**
        - All cultivars tend to overestimate for most of the stage before centering
    - **Bloom**
        - Chardonnay and Merlot tend to underestimate
        - Riesling overestimates
    - **Veraison**
        - Cabernet Sauvignon and Riesling seem to overestimate
        - Merlot seems to underestimate



Day 0 (far right) is the true day of stage transition

Results: RMSE (in days) Comparison

- PhenoTracker significantly **outperforms** the Zapata and Parker models in every case **except** budbreak for Riesling

- PhenoTracker RMSE tends to remain **under 7 days**, showing a reliable amount of variance

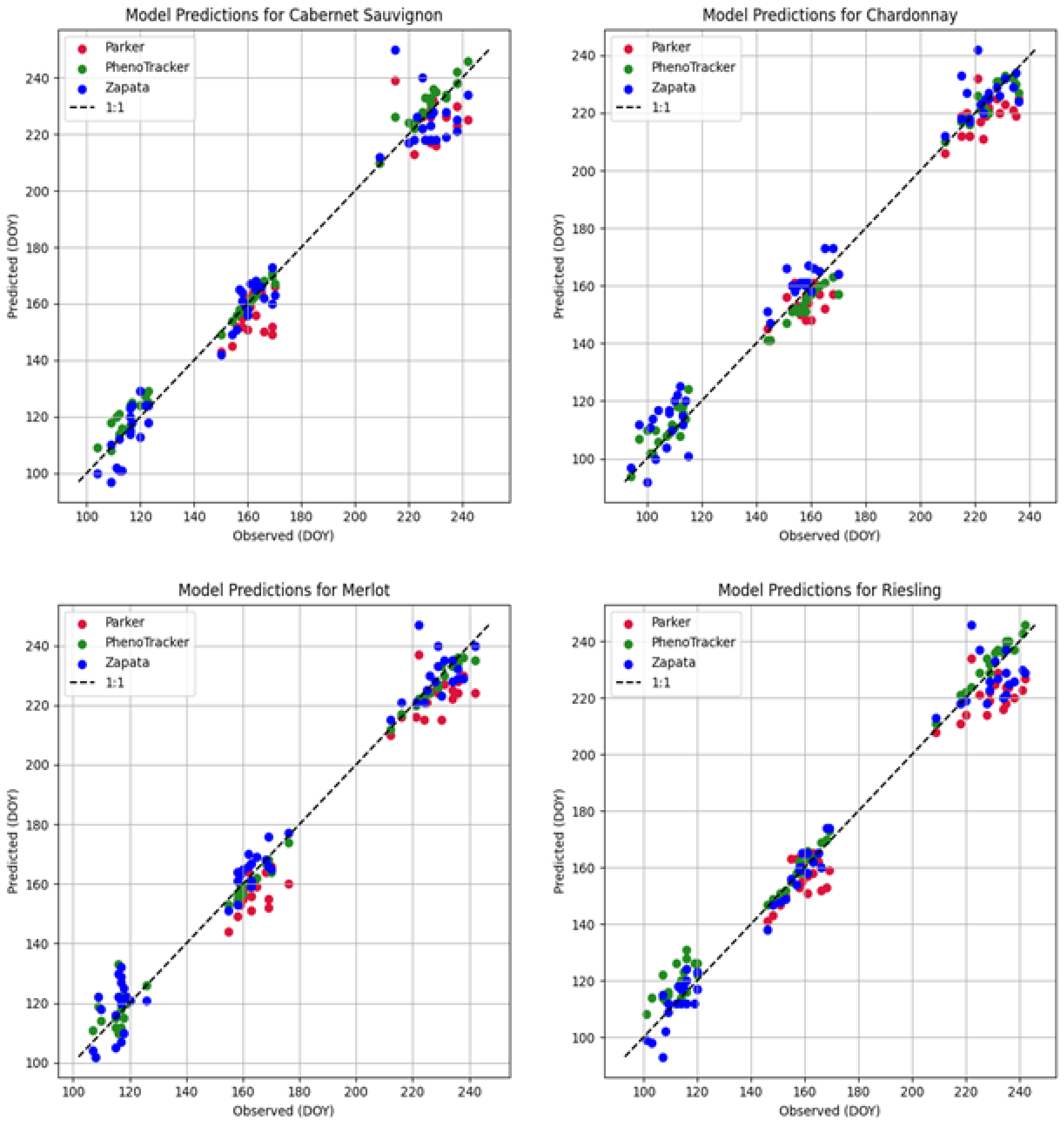
- PhenoTracker provides a daily prediction; hence, we used the **"most-voted day"**

	Cabernet	Chardonnay	Merlot	Riesling
Budbreak				
PhenoTracker	*5.88*	*4.98*	*5.88*	8.61
Zapata	6.61	9.13	8.62	*5.50*
Parker	–	–	–	–
Bloom				
PhenoTracker	*1.22*	*4.80*	*2.61*	*2.11*
Zapata	5.25	6.00	4.47	3.84
Parker	8.55	6.64	8.78	7.11
Veraison				
PhenoTracker	*4.38*	*3.37*	*2.24*	*3.42*
Zapata	12.06	7.75	7.72	10.27
Parker	9.91	8.01	9.01	11.98

Results: Comparison Scatter Plot

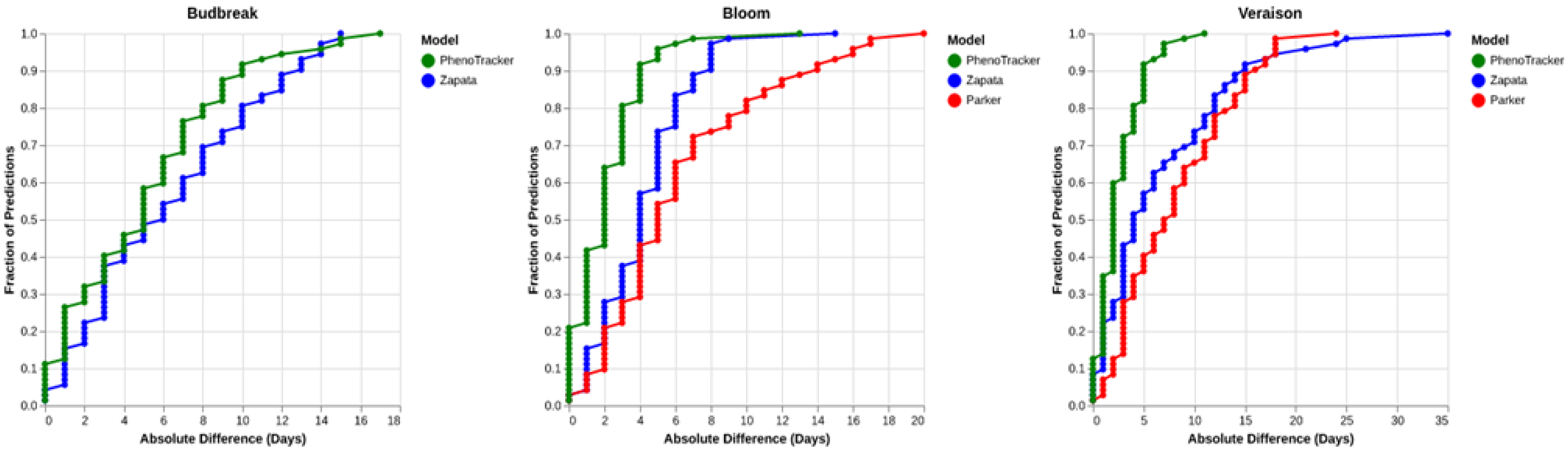
- Each stage across all seasons are plotted
  - Left: Budbreak
  - Center: Bloom
  - Right: Veraison

- PhenoTracker (green) adheres closer to the line than the other models



Results: Performance Profiles

- **X-axis**: number of days a given model deviates from the best performing model
- **Y-axis**: fraction of predictions for which this behavior holds
- The closer a model is to the Y-axis and for a higher fraction of predictions, the better it is
- PhenoTracker is the best overall performing model given all seasons of all cultivars
  - For budbreak, for over 95% of the predictions, PhenoTracker outperforms Zapata



Remarks

- PhenoTracker is **machine learning approach** outperforms traditional GDD-based approaches
- PhenoTracker factors in **more environmental variables** than traditional methods, capturing a more complete picture of grape phenology
- Unified training approach that **addresses several cultivars with one model**
- **Historical patterns can be remembered** and referenced later in the season for decision making
- **Automatically discovers importance of each input feature**, rather than having to be manually discovered

Future Work

1. **Model deployment** on the publicly accessible **WSU AgWeatherNet** website to provide live predictions for future growing seasons

2. **Model refinement** implementing functionalities for finetuning the trained model

3. Integration with the **WOFOSTGym crop simulator**

4. Addition of **other locations** to address generalization

Acknowledgements

- **Keller lab members**, Lynn Mills and Zilia Khaliullina, for their support in collecting grape phenology data

- **WSU AgWeatherNet** for their support on providing environmental data

- This research was supported by USDA NIFA award No. 2021-67021-35344 (AgAID AI Institute).

- The model code is accessible at https://github.com/AgAIDInstitute/PhenoTracker

- **Contact:** Ananth Kalyanaraman (ananth@wsu.edu)