User guide to phenology dataset for STATS 605 2023 Fall

Prepared by Yiluan Song

1. Background

Phenology is the timing of recurring biological events, such as flowering and leafing. Phenology is shifting during ongoing global changes, not only causing ecological repercussions but also directly impacting our life, through changing crop yield and allergy seasons. It is important to advance the modeling of phenology for two reasons: to gain mechanistic knowledge about long-term impacts of climate change and to deliver accurate short-term forecasting to inform decision making.

2. Challenge

Despite the rapid expansion of phenological data, there is a methodological gap in integrating different types of data. Phenology models (*1*) typically rely on continuous data from remote sensing or discrete data from field observations, but rarely using both. These two types of data are correlated and complement each other, but they measure different properties and have distinct error structures (*2*). Existing attempts to assimilate both types of data with one model, such as the continuous development model (*3*), have not been widely applied. An integrative modeling approach is much needed to assimilate various data streams to collectively infer the mechanisms of phenology and make predictions.

3. Datasets

We provide colocated datasets at two sites in the United States: Harvard Forest (HARV) and San Joaquin Experimental Range (SJER) (*metadata.csv*). HARV has deciduous forests in a temperate climate and SJER has open oak woodlands in a Mediterranean climate. This means that the phenology at two sites may be drastically different, with trees greening up in the winter at SJER. At each site, there are tagged individual plants that are monitored for their phenology (*metadata.csv*).

We provide two types of <u>discrete response variables</u>, including phenophase status and phenophase intensity (*discrete.csv*) (*4*, *5*). Field technicians visually assess the phenophase (life cycle events) of tagged individual plants at varying frequency, more frequent when phenophases are more likely to occur. Phenophase status is a binary variable, indicating whether or not a phenophase is occurring; phenophase intensity is an ordinal variable, quantifying the degree to which a given phenophase is being expressed.

We provide <u>continuous response variables</u>, namely enhanced vegetation index (EVI), on two spatial resolutions (*continuous_3m.csv*, *continuous_500m.csv*). We can think of EVI as an indicator of photosynthetic activity of an area derived from remote sensing data, related to but different from phenophases assessed by humans. 3-m EVI (*6*) were retrieved for pixels corresponding to each tagged individual plant, whereas 500-m EVI (*7*) were retrieved for two pixels corresponding to the two sites.

We provide some commonly used weather variables (*weather.csv*) at the two sites to serve as possible <u>predictors</u> (*8*), but please feel free to use any publicly available data for predictors.

4. Usage

We tagged each observation with their intended use (training or validation). We would like you to use all types of training data to develop a coherent phenology model, and then make in-sample and out-of-sample predictions of all types, including phenophase status, phenophase intensity, 3-m EVI, and 500-m EVI. Interpretability on the effect of weather variables on phenology is strongly favored.

5. References

1. K. Hufkens, D. Basler, T. Milliman, E. K. Melaas, A. D. Richardson, An integrated phenology modelling framework in r. *Methods Ecol. Evol.* **9**, 1276–1285 (2018).
2. A. Donnelly, R. Yu, K. Jones, M. Belitz, B. Li, K. Duffy, X. Zhang, J. Wang, B. Seyednasrollah, K. L. Gerst, D. Li, Y. Kaddoura, K. Zhu, J. Morisette, C. Ramey, K. Smith, Exploring discrepancies between in situ phenology and remotely derived phenometrics at NEON sites. *Ecosphere*. **13**, e3912 (2022).
3. J. S. Clark, C. Salk, J. Melillo, J. Mohan, Tree phenology responses to winter chilling, spring warming, at north and south range limits. *Funct. Ecol.* **28**, 1344–1355 (2014).
4. National Ecological Observatory Network, Data Product DP1.10055.001, Plant phenology observations (2020).
5. S. C. Elmendorf, K. D. Jones, B. I. Cook, J. M. Diez, C. A. F. Enquist, R. A. Hufft, M. O. Jones, S. J. Mazer, A. J. Miller‐Rushing, D. J. P. Moore, M. D. Schwartz, J. F. Weltzin, The plant phenology monitoring design for The National Ecological Observatory Network. *Ecosphere*. **7**, e01303 (2016).
6. Planet Team, *Planet Application Program Interface: In Space for Life on Earth* (San Francisco, CA, 2017; https://api.planet.com).
7. K. Didan, A. Barreto, VIIRS/NPP Vegetation Indices 16-Day L3 Global 500m SIN Grid V001 [Data set] (2018), (available at https://doi.org/10.5067/VIIRS/VNP13A1.001).
8. P. E. Thornton, M. M. Thornton, B. W. Mayer, Y. Wei, R. Devarakonda, R. S. Vose, R. B. Cook, Daymet: Daily Surface Weather Data on a 1-km Grid for North America, Version 3. *ORNL DAAC* (2016), doi:https://doi.org/10.3334/ORNLDAAC/1328.