



The phenor R package for vegetation phenology modelling: An open source toolkit to facilitate model-data integration and promote scientific reproducibility

Koen Hufkens (1) and Andrew D. Richardson (2,3)

(1) INRA, UMR ISPA, Villenave d'Ornon, France (koen.hufkens@gmail.com), (2) School of Informatics, Computing and Cyber Systems, Northern Arizona University, Flagstaff, AZ, USA, (3) Center for Ecosystem Science and Society, Northern Arizona University, Flagstaff, AZ, USA

Ecological data are often scattered across various archives, using different formats. This presents a considerable hurdle to researchers who want to integrate data from different archives. More so, these various sources and standards make reproducibility even more challenging. Here we highlight our phenor R package (Hufkens et al. 2018, *Methods in Ecology and Evolution*) which has overcome these issues by interfacing with various data sources within the context of vegetation phenology modelling.

With vegetation phenology as a first order control on ecosystem productivity, accurate and transparent model predictions of vegetation phenology in a changing climate are key. In order to facilitate easy model comparison and future development of new models the phenor R package assimilates four important phenological records across a variety of ecosystem, plant functional types and spatial scales (e.g. the PhenoCam network, the USA National Phenology Network (USA-NPN), the Pan European Phenology Project (PEP725) and MODIS phenology). These assimilated datasets provide extensive coverage in the US and Europe and results can be easily scaled globally using various gridded meteorological data products (e.g., historical weather data from daymet, climate projections from NASA earth exchange) made accessible through the software.

We provide a worked example for the phenor R package using the recent standardized PhenoCam dataset (Richardson et al. 2018, *Scientific Data*; <http://phenocam.us>) to demonstrate the ease with which a suite of phenological models, using various data sources, can be evaluated and scaled up from sites to regions and biomes, and extrapolated in both forecast and hindcast modes.

This R package provides an easy and consistent way to interface with data providers and facilitate (ecological) research. Unlike unbundled scripts, the R package offers flexibility to respond to future changes in format and content which may be made by data providers. The open source nature of (our) software guarantees scientific reproducibility, transparency and potentially fosters future collaborations, building upon a growing community of developers (e.g., rOpenSci / <http://ropensci.org>).