Title: Environmental correlates of tree reproductive phenology in a tropical state of India - insights from a citizen science project

Authors: Krishna Anujan[1,2]*, Jacob Mardian[3], Carina Luo[3], Ramraj R[3], SeasonWatch Citizen Scientist Network[1], Hana Tasic[3], Nadia Akseer[3], Geetha Ramaswami[1]

[1] SeasonWatch, Nature Conservation Foundation, 1311, “Amritha”, 12th Main, Vijayanagar 1st Stage, Mysore 570017, India

[2] Department of Ecology, Evolution and Environmental Biology, Columbia University in the City of New York, 1110 Amsterdam Ave, New York, NY 10025, United States

[3] Modern Scientist Global, 110 James Street, Suite 101, St. Catharines, Ontario L2R 7E8, Canada

* =corresponding author

Abstract:

Tropical tree reproductive phenology is sensitive to changing climate, but inter-individual and interannual variability at the regional scale is poorly understood. While large-scale and long-term datasets of environmental variables are available, reproductive phenology needs to be measured in-site, limiting the spatiotemporal scales of the data. We leveraged a unique dataset assembled by SeasonWatch, a citizen-science phenology monitoring programme in India to assess the environmental correlates of fruiting and flowering in three ubiquitous and economically important tree species - jackfruit, mango and tamarind - in the south-western Indian state of Kerala. Using 165006 observations spread over 19596 individual trees and over 9 years, and temperature and rainfall predictors for the fortnight preceding each observation, accessed using the ERA5-LAND dataset, we modelled the environmental correlates of reproductive phenophases - flower and fruit occurrence - in these species using two statistical approaches - machine learning and generalised linear mixed models. In these complementary approaches, we used a leave-one-year-out approach to cross-validate the

model against observations. Models of phenophase presence for all species had higher predictive power than models of phenophase intensity. We found strong influences of multiple temperature and rainfall variables on phenophase presence - soil moisture and maximum temperature had high importance values under machine learning models while the number of consecutive dry days also had high effect sizes in the generalised linear mixed models. The effect of time-varying environmental factors like total precipitation and consecutive dry days were also modified by the static predictors like elevation, aspect, and urbanisation. Taken together, our results show the pervasive influence of climate on tropical tree reproductive phenology and extent of variability among years and individuals. We also demonstrate the potential and limitations of citizen-science observations in making and testing predictions at scale for predictive climate science in tropical landscapes.

Keywords: citizen science, tropical phenology, climate, machine learning

Introduction:

Tree reproductive phenology is altered by a changing climate, affecting species interactions and ecosystem functions and services (Cleland et al., 2007; Mora et al., 2015). Flowering and fruiting of trees are triggered by environmental cues like rainfall events, dry spells and photoperiods, and therefore are sensitive to a changing environment (Chang-Yang et al., 2016; Chapman et al., 2005, 2018; Pau et al., 2018). Change in timing such as early or delayed onset, and persistence and magnitude of a phenophase, of any species can change flower and fruit abundance in the landscape and influence interactions with other species like pollinators, frugivores and seed dispersers (CaraDonna et al., 2014; Morellato et al., 2016; Visser & Both, 2005). Consequently, for wild populations, changes in phenology can influence regeneration dynamics and population persistence. Changing phenology can also have socio-cultural impacts, through altering food production and creating mismatches in harvest as well as cultural events (Mora et al., 2015; Morellato et al., 2016). While there is a growing understanding of landscape-level phenological shifts in temperate regions, large gaps remain in ecological understanding and predictions for shifts in tropical trees.

Phenology in tropical trees is affected by seasonal changes in environmental factors distinct from temperate landscapes (Abernethy et al., 2018; Pau et al., 2011). Much of large-scale phenological studies have been recorded in temperate landscapes where seasons are distinct and correspond to starkly discernible plant responses like spring “green-up” or synchronised flowering (Elmendorf et al., 2019; Li et al., 2021). On the other hand, in tropical landscapes where trees grow and reproduce year-round, relevant environmental cues for phenological change are more diffused, need not coincide temporally, and could be different across species that co-occur in the same landscape (Abernethy et al., 2018; Chen et al., 2018; Parmesan, 2007). As climate change in the tropics has resulted in altered rainfall and temperature regimes as well as extreme events, tropical environments are less predictable across the years (Mora et al., 2013). To understand how these changes could affect tropical

tree reproductive phenology, including phenophases such as the timing and intensity of fruiting or flowering, an important first step is to understand the baseline variability in phenology.

While the onset of a phenophase is commonly measured, the persistence and intensity of the phenophase are also likely to be altered by a changing climate and these measures are poorly understood. The occurrence of the phenophase, which encompasses the life-cycle of the phenophase, can be an indicator of the net availability of flowers and fruits in the landscape. The onset of tree reproductive phenology is triggered by the release of florigen hormones but can be triggered by a variety of environmental cues in different species (CaraDonna et al., 2014; Li et al., 2021; Satake et al., 2013). A variety of environmental cues that link to temperature, precipitation and solar radiation; cold days, dry spells, total accumulated rainfall and soil moisture can not only affect the onset, but also the persistence of a phenophase (Brearley et al., 2007; Diez et al., 2014; Pau et al., 2013; Wright & Calderón, 2018). Interannual rainfall variability - high rainfall years and drought years including cycles like the El Niño Southern Oscillation and the Indian Ocean Dipole - could result in interannual variability in tree phenophase onset and intensity (Chang-Yang et al., 2016; Chapman et al., 2018; Dunham et al., 2018; Medway, 1972; Pau et al., 2018; Zimmerman et al., 2007). Successful pollination of flowers would lead to subsequent onset of fruiting that could also be affected by rainfall and temperature, as well as flowering (Dunham et al., 2018; Tutin & Fernandez, 1993). Therefore, understanding flowering and fruiting phenology in aggregate requires longitudinal data on environmental correlates and the life-cycle of the phenophase.

Environmental cues could affect phenology with a lag time, or cumulatively over a longer timeframe, and affect individuals differently, necessitating large-scale and long-term data. An exemplar of a long-term phenology dataset is from Japan where the flowering of cherry blossoms (*Prunus jamasakura*) has been documented since the 9th century creating strong

baselines to assess contemporary change (Allen et al., 2014; Aono & Kazui, 2008). In tropical species, at least six years of data is necessary to accurately detect phenology in regularly flowering and fruiting species, with much more for noisy species (Bush et al., 2017). Reproductive phenology could also be affected by static factors - like elevation and the aspect of the slope - that influence water availability, as well as the degree of urbanisation. Even within the same year (similar climatic cues), trees in North America, Europe and China are likely to flower earlier in cities than in surrounding non-urban habitats, likely due to heat island effects (Neil & Wu, 2006). Therefore, besides long-term studies, high sampling effort is required to capture intraspecific variability in response to local habitat. A recent review of fruiting phenology studies from the Neotropics showed that across 218 datasets assessed, the majority of datasets lasted 2 years or less, and only 10.4% monitored > 15 individuals per species (Mendoza et al., 2017). In the absence of robust datasets, potentially high intraspecific variation in reproductive phenology makes it difficult to parametrise models that span spatial scales of environmental change.

Citizen science allows access to longitudinal data across large spatial scales that is difficult to achieve through other sampling techniques. Citizen science data has led to novel understanding of phenological patterns in other taxa - citizen observations of migratory bird populations across North America have been useful in documenting phenological shifts, spatial differences in sensitivities to changing climate, as well as disentangling drivers of change (Haas et al., 2022; Hurlbert & Liang, 2012; Kelly et al., 2016). While these data may have different strengths when compared to expert observations, focussing on common species that are easy to identify can potentially minimise errors while allowing for large spatial coverage. SeasonWatch, a citizen science phenology monitoring programme has been collecting data on several phenophases on common species across the Indian subcontinent (Ramaswami et al., 2019, 2021). Each week, observers record fruiting, flowering and other phenophases on specific trees consistently using uniform protocol that records both presence and intensity of a phenophase (Fig 1a). This dataset, with substantial

contributions since 2014, provides a unique opportunity to build baselines for phenology in the Indian subcontinent. In the past, some analyses of these data have revealed patterns of change in at least one species - *Cassia fistula* - over the time period of observation (Ramaswami et al., 2021; Ramaswami & Quader, 2018). These common species, spread across geography can add to phenological understanding beyond focussed, localised studies.

We used large-scale (165006 observations across 19596 individual trees), long-term (9 years) directly-observed phenological data and matched environmental correlates using machine learning and mixed models to understand links between environment and phenology and to aid future hypotheses on climate-induced phenological shifts. We leveraged 9 years of consistent phenology data across the southwestern Indian state of Kerala collected through SeasonWatch data and remotely sensed environmental variables. We focussed on jackfruit (*Artocarpus heterophyllus*), mango (*Mangifera indica*) and tamarind (*Tamarindus indica*) as the species with most recorded trees and most number of observations across the state. Using co-located reproductive phenophase (flowering and fruiting) and environmental data (patterns of rainfall, temperature and soil moisture), we ask what are the environmental correlates of (i) changing phenophase intensity (ii) phenophase occurrence in these species? (iii) How does the environmental influence on phenophases change with static habitat variables? We expected temperature and precipitation variables to influence the occurrence of reproductive phenophase, but the specific correlates to be different across the assessed species. We also expected local habitat factors to alter the effect of these variables on reproductive phenology.

Methods:

Study Area and Species

We analyzed phenology and environmental data from the state of Kerala in the southwest coast of India. Bounded between 8° 18’ N, 12° 48’ N, 74° 52’ E and 72° 22’ E, Kerala is along the Arabian Sea coast with a tropical monsoon climate. The total population of the state is over 3 million, spread over an area of 38,863 km^2. The state receives both the southwest monsoon and the northeast monsoon in a year with an average of over 3000 mm annual rainfall. Across all SeasonWatch data (SeasonWatch Citizen Scientist Network, 2022), 263,288 observations across species and categories, and 93.23% of observations of top 3 species are from Kerala. Moreover, these data span a large range of the existing environmental variability in the state in elevation, lat/long and distance to coast. This information density and spread, especially for repeat observations, makes it an ideal dataset to answer these questions.

We analyzed reproductive phenology observations on three of the most frequently observed species in the SeasonWatch dataset - *Artocarpus heterophyllus* (henceforth “jackfruit”), *Mangifera indica* (henceforth “mango”) and *Tamarindus indica* (henceforth “tamarind”) from January 2014 - July 2022. The fruiting and flowering phases of these common trees and distinctly observable with the naked eye, lending confidence to community data. The jackfruit tree is evergreen and monoecious with minute, yellowish-green flowers and green fruits with yellow to orange fleshy perianths inside (Kishen, 2005; Sasidharan, 2006). Mango trees have polygamous flowers that are yellowish-green and a fleshy fruit that is green when unripe and yellowish-red when ripe. Tamarind trees have small, bisexual flowers that are yellow with reddish-pink dots. The fruit is a pod about 10-15 cm long, with a hard wall and fleshy mesocarp inside that are both green when unripe and brown when ripe. All three trees are grown across the state in orchards and homes and the fruits are commercially and culturally important. Between 2012 and 2019, India has on average produced 1713000

Metric tonnes (MT) of jackfruit, 14551900 MT of mango and 201330 MT of tamarind (https://ceicdata.com). In Kerala households, all three species are cooking staples and mango and jackfruit also feature in culturally important festivals like Vishu, a harvest festival celebrated in mid-April.

SeasonWatch Method

Observations to SeasonWatch comprise estimates of the presence and quantity of a given phenophase. Observers register one or more trees from a list of suggested species with the project and record phenology on a weekly basis. The standard protocol for reporting a phenophase involves dividing the canopy of a tree into thirds, reporting a phenophase as 'none' if it is not observed, as 'few' if a phenophase is observed only in 1/3rd of the canopy, and as 'many' if a phenophase is observed on more than 1/3rd of the canopy (Fig 1a). The intensity of a phenophase, is thus, on an ordinal scale with 'few' representing a quantity less than 'many'. Observers are children or adults, participating from diverse habitats such as school campuses, urban and rural residential areas, and forests. As in other citizen science projects, there is likely to be inter-observer variability in reporting phenophases as well as bias towards observing familiar species with conspicuous flowers or fruits. At present, the project minimises biases and errors at the training stage. SeasonWatch trains citizen scientists through workshops where step-by-step guidance is provided on the observation process along with practical sessions where possible. One-on-one follow-up, and troubleshooting is also provided to ensure consistency across observers. In Kerala, these training workshops are held at the beginning of the school year in June/July, across 41 education districts, and follow up is done over communication channels such as WhatsApp, email, and social media.

Environmental datasets

To generate a set of predictors for fruiting and flowering, we summarised environmental datasets from various sources within the state boundary of Kerala. We used daily

measurements of precipitation, temperature, solar radiation and soil moisture for the entire soil column from the ERA5-Land dataset at the 11.1km spatial resolution. Using Google Earth Engine, we summarised these layers to arrive at total precipitation, number of consecutive dry days, minimum temperature, maximum temperature, mean temperature, solar radiation and mean soil moisture for the fortnight preceding each observation. We calculated these values also at monthly and quarterly timescales and found that they were highly correlated to the fortnightly value and hence used this finest temporal resolution for the analysis.

We also calculated static spatial predictors at each tree location that might potentially act as modifiers for the temporal predictors. We used an SRTM layer at the 30 m resolution to calculate elevation, slope and aspect of the tree location. Further, we used the NASA nightlights layer from 2016 as a proxy for the degree of urbanisation (Sutton et al., 2001). We chose this metric as the ambient light, one of the proxies for urbanisation, also has a direct ecological relevance on the physiology of plant life. Analyses were performed using R version 4.1.3 and associated packages *plyr*, *ordinal* (R Core Team, 2022; Wickham, 2020).

Data cleaning, validation and analysis

The SeasonWatch dataset was cleaned and processed to maximise inclusion of data points. To validate the geographic location of the data points, we included all the points that were already in Kerala. From the points that had no location information, we excluded those that were single observations of trees, because of no alternate way to fill these. For the rest of the unconfirmed observations, we filled them with lat/longs from the same tree if available, or the average lat/long from other trees reported by the same user. Finally, if no trees by the user had accurate lat/long values and if they were by users who can be located through Google maps (schools or other institutions), we used a geocoding software on Google Sheets, Awesome Table Geocode, to fill these values. We then manually validated these with help from the field coordinator for Kerala. We excluded observations that could not be

filled or verified by these methods. Using these methods, we assembled a final dataset of 165006 observations across 19596 trees; 7927 trees of jackfruit, 9919 of mango and 1750 of tamarind across the state of Kerala (Fig 1).

Participation in SeasonWatch is a completely voluntary activity, and observers may miss one or more weeks in making consecutive observations. To maximise our predictive power and the spatiotemporal scale of the dataset, we used all available observations. Since missing data may affect estimates of start and end dates of phenophase occurrence, we did not model these aspects of tree phenology in this study. We individually modelled the ecological predictors of fruiting and flowering in the three species, as these species are physiologically and evolutionarily distinct and there is no reason to assume common drivers of phenophases. The SeasonWatch raw data for fruiting and flowering are ordinal with “none”, “few” and “many” levels possible for each observation. Along with treating these categories as is, we also recoded this as a binary variable by combining the “few” and “many” categories as “yes” and “none” as “no”. We assumed that despite training, the categorising of “few” and “many” could have some user-based subjectivity which could be resolved through this binary variable. Using both machine learning methods and generalized linear mixed model, we modelled ordinal and binary fruiting and flowering status (“phenophase”) of each of the species. We decided to use two modelling techniques because 1) individual models have their own assumptions and biases that can be partially overcome using multiple lines of evidence. 2) Importantly, interpreting variable importance in machine learning models is subject to equifinality - the potential for arriving at the same prediction results through completely different approaches. This can be partially offset by using multiple lines of evidence and seeking areas of agreement. Analyses were performed using R version 4.1.3, Python 3.9 and ArcGIS Pro version 3.0.2.

<u>Machine Learning models:</u> We used random forest models and catboost models to respectively classify the binary and ordinal phenophase variables based on the predictors.

Random Forest classifier is a supervised learning algorithm that uses an ensemble of decision trees using majority voting to classify the presence/absence of phenophases. CatBoost is a machine learning algorithm that uses gradient boosting on decision trees and is available as an open source library.

phenophase ~ precipitation+maximum temperature+minimum temperature+mean temperature+ dry days+soil moisture+solar radiation+elevation+slope+aspect+urbanisation

We ran these models in python using the packages *sklearn* and *catboost*. Tree-based methods are fairly insensitive to collinearity among predictors, allowing us to compare the full-range of temperature and precipitation variables simultaneously. To test the fit of these models, we ran models by leaving the data from one year out of the modelling in each iteration and tested the model predictions against the observations from this excluded year (leave-one-out-cross-validation; henceforth LOOCV). We tested the model accuracy using F1 scores, as it is a better indicator of the distribution of correct and incorrect misclassifications by taking the harmonic mean of precision and recall compared to the more common model accuracy scores (Grandini et al., 2020; Hossin & Sulaiman, 2015). We calculated variable importance for each model using the Mean Decrease in Accuracy (MDA) metric which removes a predictor from the model, re-trains the model to make predictions, and evaluates the decrease in accuracy. We report diagnostics and variable importance of the ensemble model.

<u>Generalised linear mixed models:</u> To model interactions and the modification of one variable by another, we also modelled the observed phenophase of the three species using these generalised linear mixed models.

*phenophase ~ precipitation temperature + dry days+soil moisture+elevation*precipitation+aspect*soil moisture+urbanisation*dry days+(1|Tree ID)*

Where precipitation, temperature, dry days, soil moisture are temporal predictors for the fortnight before the observation and elevation, aspect and urbanisation are static spatial predictors tied to the tree location. The tree ID is treated as the random effect to account for repeated measures on the same tree. We centred all predictor variables at 0 and scaled it to a distribution with standard deviation 1 for ease of comparing the effect sizes. We ran proportion odds ordinal models using the package *ordinal* for the ordinal variables and logistic regressions using the package *lme4* for the binary variable specification (Bates et al., 2015). To test the predictive power of the model, we iteratively fit models leaving each year out (for 9 years) and testing predictions for this year against actual observations, similar to the method of Leave-One-Out Cross Validation (LOOCV). We report the average value of the coefficient and the average value of standard error of the estimate across these models. We compared the predictive power of our models using mean square error and Pearson's correlation coefficient with the predicted values against the actual observations.

<u>Hotspot analysis:</u> Spatiotemporal patterns in environmental variables can be affected by a changing climate. A hot spot/cold spot is a location which has high/low value for an environmental variable, like temperature, precipitation or soil moisture, and is surrounded by other areas with high/low values as well. With changing patterns in climate, these can shift in location and new hotspots and coldspots can emerge over time. Using fortnightly aggregated environmental variables from 2014-2022, we ran emerging hotspot analysis in ArcGIS Pro to identify statistically significant hot spots and their trends over time for mean temperature, maximum temperature, minimum temperature, mean soil moisture, total precipitation, total solar radiation across Kerala. We used a combination of two statistical measures 1) the Getis-Ord Gi* statistic to identify hot spots and cold spots of climate measurements for each time step (i.e., fortnightly), and 2) the Mann Kendall trend test to examine how hot spots and cold spots have evolved over time. We also ran these analyses to understand spatiotemporal patterns in SeasonWatch data contributions across the state.

Results:

Machine learning models with spatiotemporal predictors were able to classify flowering and fruiting phenophases for the three species with substantial precision. The accuracy of a binary model by random chance would be 50%, leading to an F1 score of 0.5, while of an ordinal model would be 0.33 - models with higher F1 scores than randomly expected have predictive value. Random forest models were able to predict the presence/absence of flowering in jackfruit, mango, and tamarind trees with F1 scores of 0.67, 0.65, and 0.45, respectively. The presence/absence of fruiting was predicted with F1 scores of 0.76, 0.71 and 0.77. These performance metrics describe the ability of the models to predict tree phenophases based on environmental variables alone. Using the catboost algorithm, the ordinal models had a slightly worse fit compared to the binary models. Intensity of the phenophase for jackfruit, mango and tamarind were predicted with F1 scores of 0.58, 0.50 and 0.43 for flowering and 0.55, 0.53 and 0.49 for fruiting using the CatBoost classifier for 500 decision trees each.

Relative importance of environmental variables

Across species, temperature ranked as an important correlate of fruiting and flowering phenophases among temporal predictors. In random forest model ensembles of fruiting and flowering as binary variables, the mean importance of each of the factors, calculated using MDA, was different across species (Fig 2). However, maximum temperature in the previous fortnight was the most important correlate for flowering and fruiting of jackfruit and mango. The flowering of tamarind was also strongly influenced by temperature, but the highest ranking predictor was the minimum temperature in the previous fortnight. For the fruiting of tamarind, maximum temperature was outranked only by mean soil moisture (Fig 2e). Total precipitation in the previous fortnight was also an important environmental correlate for the flowering of jackfruit and mango. When phenophase intensity was considered, soil moisture and maximum temperature were both important; when flowering and fruiting intensity was

classified using catboost, both soil moisture and maximum temperature in the previous fortnight often had the highest rank (Fig 2).

The number of consecutive dry days in the previous fortnight, however, had some of the lowest importance ranks across species for these models. While simultaneous comparison of spatial and temporal predictors is difficult, we focussed on comparing only among the temporal predictors considered. Among the temporal predictors considered, the number of consecutive dry days in the previous fortnight was consistently the least important predictor for the flowering and fruiting phenophases of the species considered. Specifically, this was lowest ranked for all models except for the flowering of mango, where the lowest rank was for mean temperature in the previous fortnight, followed by the number of consecutive dry days. Across models of intensity as well, the lowest rank of importance was most often held by the number of consecutive dry days, or mean temperature followed by consecutive dry days.

Accounting for repeated measures within the same species

While accounting both for repeated measures across species and interactions between variables using a generalised linear mixed model (GLMM), maximum temperature continued to have the strongest effect on the fruiting of mango and jackfruit (Fig 3). However, the results for tamarind were less conclusive with smaller effect sizes across environmental correlates. These models were built using 4685 trees, 6220 trees and 1224 trees each of jackfruit, mango and tamarind respectively. Similarly for ordinal models with the proportion odds assumption, the relative effects of each of the environmental correlates for each species remained consistent in trend and magnitude.

Similar to the machine learning models, binary GLMMs were substantially better fit than ordinal models in predicting the phenophase. Pearson’s correlations between the predicted and observed values of flowering intensity for jackfruit, mango and tamarind for one year of

the data that was withheld from modelling were on average 0.14, 0.27 and 0.02, and MSE was 0.95, 0.91, 0.79. For models of fruiting, Pearson's correlations for the year left out were of jackfruit, mango and tamarind 0.29, 0.41 and 0.04, and MSE was 0.95, 0.81, 1.69, suggesting poor fit. In comparison, binary model predictions and observations were much more correlated at 0.50, 0.45 and 0.34 for flowering and 0.46, 0.55 and 0.37 for fruiting of jackfruit, mango and tamarind respectively. MSE for flowering was 0.19, 0.19 and 0.20 and for fruiting was 0.19, 0.16, 0.22. Therefore, we chose to analyse and report interactions between variables using only binary models.

Interactions between variables

The effects of most temporal predictors on fruiting and flowering phenophases were modified by spatial factors (Fig 4). We considered three sets of interactions in the models; precipitation with elevation, soil moisture with aspect and the number of consecutive dry days with urbanisation. When the full models were considered (without leaving any year out), among these, the interaction between the number of consecutive dry days and urbanisation was significant for the fruiting of all species as well as flowering of jackfruit and tamarind (not significant only for mango flowering). The aspect of the location significantly modified the effect of soil moisture on flowering of jackfruit and tamarind and the fruiting of mango and tamarind. Elevation was not a significant modifier of the effect of precipitation on the flowering of any species, and was significant only for the fruiting of jackfruit.

Spatial factors modified the effects of temporal variables on fruiting and flowering phenophases for the three species considered (Fig 4). On average, precipitation increased the probability of fruiting, but the influence of precipitation on phenology was stronger in low elevation areas and weaker with elevation gain. At low precipitation values, trees at high elevation are more likely to fruit than those in low lying areas, but under high precipitation, trees across the elevation range have similar high values (Fig 4). The effect of aspect on soil moisture diverged across the range of observed values - phenophases of the three species

under wet and dry regions responded differently to aspect. Notably, in tamarind trees, the effect of soil moisture on fruiting reversed with aspect; in locations with relatively low aspect (east facing slopes) trees had increased probability of fruiting with increased soil moisture while under higher aspect (west-facing slopes) had decreased probability of fruiting with soil moisture. Finally, the degree of urbanisation was species-specific. Overall, across species where the effect was significant, trees in more urban areas had higher probability of fruiting and flowering at any environmental condition, but more so under conditions where environmental factors increased the probability. Specifically, flowering probability increased with increased number of dry days in the previous fortnight in jackfruit trees, but this effect was amplified by urbanisation. In tamarind trees, flowering probability increased with decreasing number of dry days, and urbanisation amplified this effect as well. Contrary to the patterns seen in the ML analysis, adding a static variable revealed significant effects of the number of consecutive dry days.

Spatiotemporal patterns in environmental variables

Locations across Kerala are experiencing heterogenous change. Emerging hotspot analysis revealed spatiotemporal patterns in maximum temperature, precipitation etc in the time period when the citizen science data was collected. Aggregated across the state, none of the environmental variables had a significant overall directional trend of change from 2014-2022. However, fortnightly mean, maximum and minimum temperature had persistent hotspots in coastal locations, while inland sites towards the east were persistent cold spots, potentially due to higher elevation. On the other hand, across Kerala, from 2014 to 2022, most locations were oscillating hotspots for total precipitation, variably statistically significant hotspots or coldspots across the years. Finally, mean soil moisture had persistent hotspots in the south centre of the state.

Discussion:

Altered tree phenology patterns under a changing climate can potentially affect human-nature relationships by affecting economic and cultural connections with trees. Although altered temperature and precipitation patterns are expected to affect the phenology of tropical tree species, its variability across large spatial scales is rarely quantified and poorly understood. Combining a unique citizen science dataset from India, unparalleled in scale for tropical trees, with remotely sensed environmental variables, we show that fruiting and flowering in three common tropical species is strongly influenced by temperature and precipitation in the preceding fortnight. Moreover, we show that the influence of these temporally varying environmental factors are modified by the tree habitat in relation with elevation, slope and urbanisation. Finally, for widespread species, existing phenological information tends to be from a few specific locations, making it difficult to measure intraspecific variability across space and make population-level predictions at the regional scale. Our results establish critical baselines for the tree reproductive phenology at the regional scale, where climate predictions currently exist. Our study is the first of its kind that leverages community or citizen science for tropical tree phenology, and so our results also demonstrate the challenges and opportunities towards predictive climate science without compromising fine-scale resolution.

Patterns of temperature and precipitation predicted fruiting and flowering in tropical trees. Across the state of Kerala, the probability of fruiting and flowering of three common trees - jackfruit, mango and tamarind - were influenced by temperature and precipitation in the previous fortnight. Specifically, jackfruit flowering was most influenced by maximum temperature in the previous fortnight, while tamarind was influenced by minimum temperature (Fig 2). Our study adds to literature on the influence of temperature on tropical tree reproductive phenology (Babweteera et al., 2018; Cook, Wolkovich, & Parmesan, 2012; Cook, Wolkovich, Davies, et al., 2012; Mendoza et al., 2017). On the other hand, the

probability of mango flowering was most influenced by soil moisture, with low soil moisture leading to higher probabilities (Fig 2). Previous studies in other tropical landscapes have documented the effect of interannual variability in rainfall on flowering phenology (Dunham et al., 2018). Our results show that besides total precipitation, soil moisture and the number of consecutive dry days were important predictors for observable flowering phenophase in this study (Fig 3). For fruiting as well, maximum temperature, mean soil moisture and consecutive dry days were important predictors (Fig 2 and 3). Taken together, these results suggest that tropical trees may be responding to nuanced weather changes as triggers for reproductive phenophases. Our study shows slightly different influences of certain variables between the machine learning and mixed model approaches, which could be partially because of the way in which the variables were derived. For example, the number of consecutive dry days is sensitive to trace precipitation amounts which makes it a noisier variable. Another contributor to the discrepancy could be that machine learning models did not explicitly account for interactions between the variables.

Static variables modify the responses of reproductive phenology to temperature and precipitation in drier conditions (Fig 4). Combining location information on citizen-observed trees with temporal environmental predictors, we show that specific static variables like elevation, aspect and urbanisation index can alter the influence of time-varying environmental factors on reproductive phenology. Few variables, like consecutive dry days, that did not show up as highly ranked in variable importance in the machine learning models had high influences in the generalised linear mixed models with interactions, suggesting that accounting for interactions can bring out the ecological importance of time-varying environmental variables (Fig 2, 3). Our results show that the divergence among spatial locations is higher under drier environmental conditions; at low total precipitation and high number of dry days (Fig 4). A longitudinal study on forest trees monitored in Uganda showed that the location of the tree, specifically in relation to light, increased the probability of flowering and fruiting (Babweteera et al., 2018). Moreover tropical tree leaf exchange and

reproductive phenology are known to be altered in urban habitats (Kabano et al., 2020; Neil & Wu, 2006). Our study systematically records the extent of intraspecific variation in reproductive phenology in three widespread species, along three habitat axes. By assessing multiple modifiers simultaneously, our study makes novel contributions to predictive modelling of tree phenology in tropical habitats where multiple global and local drivers of change co-occur.

Our results suggest high intraspecific variation in the fruiting and flowering of jackfruit, mango and tamarind trees in Kerala. Beyond spatial correlates that were controlled for, tree random effect was significant in the models for all three species, suggesting unexplained variation among the individual trees considered. Environmental correlates that correspond to temperature and rainfall patterns had effects on the species, but the models overall explained only a moderate amount of variation and predicted test year data with 50-60% accuracy. We speculate that this variation could be linked to species ecology; jackfruit, mango and tamarind are all fleshy fruits with animal-dependent pollination and seed-dispersal (Table 1). In contrast with wind-dispersed species, the timing of fruiting in animal-dispersed species may not be tightly linked to environmental conditions, potentially explaining the large variation observed in the study (Cortés-Flores et al., 2019; Schaik et al., 1993). Another contributor could be the spatial scale of environmental data; assessed variables are at the kilometre spatial scale whereas fine-scale heterogeneity in temperature and rainfall can influence trees (Scheffers et al., 2017). Finally, since these are cultivated species, the differences in varieties among individuals has a known influence on phenological variation, but as this is a variable not recorded in SeasonWatch, our study could not test the extent of this effect (Makhmale et al., 2016; Singh et al., 2021).

The onset of reproductive phenophase is a commonly studied response, while we chose to measure the occurrence of a phenophase. In some temperate landscapes, onset of leaf-out phenology can be remotely sensed and therefore measured at the large scale (Czernecki et

al., 2018). However, measurement of reproductive phenology, especially for species that have inconspicuous flowers and fruits, requires on-ground measurements and is extremely difficult to organise at the large-scale. Citizen science may be one of the few ways to acquire such data, and we have leveraged such a unique dataset compiled by SeasonWatch to predict the occurrence of reproductive phenophase (Abernethy et al., 2018; McDonough MacKenzie et al., 2020; Ramaswami et al., 2021). Although shifts in onset are also possible to model with such observations, because of potential errors associated with such data, including missing observations, true onset is difficult to estimate. We found that restricting the dataset to only consecutive observations drastically reduced the sample size and explanatory power. However, as the dataset continues to grow, so does the opportunity to address such focussed questions that need specialised data.

Our analysis of citizen science phenology information across nine years in a state shows that presence/absence of a phenophase is predicted much more accurately than intensity of reproductive phenophases. While our study covers nine years of observations, more than the minimum recommended 6 years for tropical species, we would like to note that given the geographic variation, this time series may not be sufficiently long. However, this dataset is, to the best of our knowledge, unprecedented in spatial and temporal scale for reproductive phenology in tropical tree species. The ecological mechanisms behind increasing intensity of phenophases upon onset as well as their periodicity is poorly understood. However, interannual variability in flowering and fruiting intensity is reported across tropical tree species, specifically linked to interannual variability in rainfall and temperature (Brearley et al., 2007; Dunham et al., 2018; Mendoza et al., 2017). They could be linked to energy reserves in the growing season, responding to environmental correlates beyond the fortnightly time window that we considered (Babweteera et al., 2018). More trivially, while citizen science data is powerful to make inferences over large scales, the lack of predictive power for phenophase intensity is potentially indicative of variability in observers while recording fine-scale information, increasing noise (Bush et al., 2018). To test the influence of

observer bias, we also modelled the generalised linear mixed models with User ID as a random effect and found significant variation among users (Table S1 and S2). Estimating user variability and incorporating it into future models can potentially improve predictions.

Broadly, we demonstrate the potential of citizen science observations in improving the state of our knowledge in climate science. Kerala, the region analysed, is projected to move into a temperature and precipitation regime that has no analog globally (Dahinden et al., 2017). This makes it crucial to forecast and mitigate biotic responses to specific environmental correlates as well as combined effects. However, climate change predictions and action for the tropics are still largely at coarse spatial scales, making local forecasting and mitigation difficult (Dessai et al., 2009; Shepherd & Sobel, 2020). Moreover, phenology of even ubiquitous species have rarely been recorded systematically, hindering the potential to make large-scale predictions of climate impacts of tree-associated ecosystem services (McDonough MacKenzie et al., 2020; Piao et al., 2019). Our study demonstrates the potential of citizen science in filling these gaps and providing fine-resolution predictions at the large-scale. Besides quantitative information, several citizen scientists have also raised interesting anecdotal observations and cultural knowledge about the fruiting and flowering of these species at SeasonWatch meetings. An important step forward would be to integrate these diverse knowledge systems to generate ground-up predictions for change at the local scale.

In summary, our results show the influence of temperature and precipitation on the flowering and fruiting phenology of three tropical tree species. We also show that the influence of temporally varying environmental factors are modified by habitat factors like terrain and degree of urbanisation. As spatiotemporal patterns in temperature and precipitation are altered by global climate change in tropical landscapes, forecasting phenological change is crucial. Across the Indian subcontinent, precipitation and temperature variability has increased in the past century and is projected to continue to increase (Dahinden et al., 2017;

Power et al., 2013; Zhang & Fueglistaler, 2019). Our analysis of patterns over the last decade showed spatial variation in hotspots and coldspots for temperature and precipitation and their change over time. An important future direction for phenological studies would be to understand intraspecific variation across space and forecast phenological responses in these novel climates.Our study that leverages a unique and extensive citizen science phenology dataset demonstrates the advantages of using these methods to expand the spatial and temporal resolution of climate predictions, including in novel climates.

Acknowledgements:

The authors would like to acknowledge the SeasonWatch team for support during this work, especially Mohammed Nizar, who helped in the manual validation of parts of the data. We also thank Karthik T, Anand M O, Suhel Quader and the NCF EPE team for discussions on data management, analyses and interpretations. SW is grateful for the support of the Wipro Foundation for their continued efforts.

Data accessibility statement:

All the data used in this manuscript are publicly accessible. The code used for accessing the geospatial data is archived in Google Earth Engine. The modified datasets, code for analysis and figures have been archived in GitHub and the repository will be made public upon acceptance.

Research ethics:

The study did not deal with human subjects or data with personal identifiers. For SeasonWatch data, we followed ethical protocol that was agreed on by the SeasonWatch Citizen Scientist Network.

Author contributions:

Funding for this work was acquired by GR and NA, while the study was conceptualised by GR and KA. Phenology data was collected by the SeasonWatch Citizen Scientist Network. KA cleaned and validated the phenology data while JM accessed and collated the geospatial data. KA, JM, CL and RR performed the analysis led by KA and supervised by GR and NA, with administrative support from HT. KA visualised the data and wrote the first draft; KA, JM and GR edited the manuscript with support from all authors.

Funding sources:

The analysis was funded by the Booster Programme at the Patrick J McGovern Foundation. SeasonWatch data collection was supported by the Wipro Foundation.

References

https://doi.org/10.1111/btp.12355

SeasonWatch Citizen Scientist Network (2022) – Data on tree phenology. Data type – 01/01/2014 – 31/07/2022, Kerala, *Artocarpus heterophyllus*, *Mangifera indica*, *Tamarindus indica*, SeasonWatch, India. Data accessed on 01/08/2022

Abernethy, K., Bush, E. R., Forget, P.-M., Mendoza, I., & Morellato, L. P. C. (2018). Current issues in tropical phenology: A synthesis. *Biotropica*, *50*(3), 477–482. https://doi.org/10.1111/btp.12558

Allen, J. M., Terres, M. A., Katsuki, T., Iwamoto, K., Kobori, H., Higuchi, H., Primack, R. B., Wilson, A. M., Gelfand, A., & Silander, J. A. (2014). Modeling daily flowering probabilities: Expected impact of climate change on Japanese cherry phenology. *Global Change Biology*, *20*(4), 1251–1263. https://doi.org/10.1111/gcb.12364

Aono, Y., & Kazui, K. (2008). Phenological data series of cherry tree flowering in Kyoto, Japan, and its application to reconstruction of springtime temperatures since the 9th century. *International Journal of Climatology*, *28*(7), 905–914. https://doi.org/10.1002/joc.1594

Babweteera, F., Plumptre, A. J., Adamescu, G. S., Shoo, L. P., Beale, C. M., Reynolds, V., Nyeko, P., & Muhanguzi, G. (2018). The ecology of tree reproduction in an African medium altitude rain forest. *Biotropica*, *50*(3), 405–417. https://doi.org/10.1111/btp.12563

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Brearley, F. Q., Proctor, J., J. Proctor, Laszlo Nagy, Nagy, L. L., Nagy, L., Dalrymple, G., & Voysey, B. C. (2007). Reproductive phenology over a 10 year period in a lowland evergreen rain forest of central Borneo. *Journal of Ecology*, *95*(4), 828–839. https://doi.org/10.1111/j.1365-2745.2007.01258.x

Bush, E. R., Abernethy, K., Jeffery, K. J., Tutin, C. E. G., White, L. J. T., Dimoto, E., Dikangadissi, J.-T., Jump, A. S., & Bunnefeld, N. (2017). Fourier analysis to detect phenological cycles using long□term tropical field data and simulations. *Methods in Ecology and Evolution*, *8*(5), 530–540. https://doi.org/10.1111/2041-210x.12704

Bush, E. R., Bunnefeld, N., Dimoto, E., Dikangadissi, J.-T., Jeffery, K., Tutin, C., White, L., & Abernethy, K. A. (2018). Towards effective monitoring of tropical phenology: Maximizing returns and reducing uncertainty in long-term studies. *Biotropica*, *50*(3), 455–464. https://doi.org/10.1111/btp.12543

CaraDonna, P. J., Iler, A. M., & Inouye, D. W. (2014). Shifts in flowering phenology reshape a subalpine plant community. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(13), 4916–4921. https://doi.org/10.1073/pnas.1323073111

Chang-Yang, C.-H., Sun, I.-F., Tsai, C. H., Lu, C.-L., & Hsieh, C.-F. (2016). ENSO and frost codetermine decade-long temporal variation in flower and seed production in a subtropical rain forest. *Journal of Ecology*, *104*(1), 44–54. https://doi.org/10.1111/1365-2745.12481

Chapman, C. A., Chapman, L. J., Struhsaker, T. T., Zanne, A. E., Clark, C. J., Clark, C. J., & Poulsen, J. R. (2005). A long-term evaluation of fruiting phenology: Importance of climate change. *Journal of Tropical Ecology*, *21*(1), 31–45. https://doi.org/10.1017/s0266467404001993

Chapman, C. A., Valenta, K., Bonnell, T. R., Brown, K. A., & Chapman, L. J. (2018). Solar radiation and ENSO predict fruiting phenology patterns in a 15□year record from Kibale National Park, Uganda. *Biotropica*, *50*(3), 384–395. https://doi.org/10.1111/btp.12559

Chen, Y.-Y., Satake, A., Sun, I.-F., Kosugi, Y., Tani, M., Numata, S., Hubbell, S. P., Fletcher, C., Supardi, N. N., & Wright, S. J. (2018). Species-specific flowering cues among general flowering Shorea species at the Pasoh Research Forest, Malaysia. *Journal of Ecology*, *106*(2), 586–598. https://doi.org/10.1111/1365-2745.12836

Cleland, E. E., Chuine, I., Menzel, A., Mooney, H. A., & Schwartz, M. D. (2007). Shifting plant phenology in response to global change. *Trends in Ecology and Evolution*, *22*(7), 357–365. https://doi.org/10.1016/j.tree.2007.04.003

Cook, B. I., Wolkovich, E. M., Davies, T. J., Ault, T. R., Julio L. Betancourt, Julio L. Betancourt, Betancourt, J. L., Allen, J. M., Bolmgren, K., Cleland, E. E., Crimmins, T. M., Kraft, N. J. B., Lancaster, L. T., Mazer, S. J., McCabe, G. J., McGill, B. J., Parmesan, C., Pau, S., Regetz, J., … Travers, S. E. (2012). Sensitivity of Spring Phenology to Warming Across Temporal and Spatial Climate Gradients in Two Independent Databases. *Ecosystems*, *15*(8), 1283–1294. https://doi.org/10.1007/s10021-012-9584-5

Cook, B. I., Wolkovich, E. M., & Parmesan, C. (2012). Divergent responses to spring and winter warming drive community level flowering trends. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(23), 9000–9005. https://doi.org/10.1073/pnas.1118364109

Cortés-Flores, J., Cornejo-Tenorio, G., Urrea-Galeano, L. A., Andresen, E., González-Rodríguez, A., & Ibarra-Manríquez, G. (2019). Phylogeny, fruit traits, and ecological correlates of fruiting phenology in a Neotropical dry forest. *Oecologia*, *189*(1), 159–169. https://doi.org/10.1007/s00442-018-4295-z

Czernecki, B., Nowosad, J., & Jabłońska, K. (2018). Machine learning modeling of plant phenology based on coupling satellite and gridded meteorological dataset. *International Journal of Biometeorology*, *62*(7), 1297–1309. https://doi.org/10.1007/s00484-018-1534-2

Dahinden, F., Fischer, E. M., & Knutti, R. (2017). Future local climate unlike currently observed anywhere. *Environmental Research Letters*, *12*(8), 084004. https://doi.org/10.1088/1748-9326/aa75d7

Dessai, S., Hulme, M., Lempert, R., & Pielke, R. (2009). Do We Need Better Predictions to Adapt to a Changing Climate? *Eos, Transactions American Geophysical Union*, *90*(13), 111–112. https://doi.org/10.1029/2009EO130003

Diez, J. M., Ibáñez, I., Silander, J. A., Primack, R., Higuchi, H., Kobori, H., Sen, A., & James, T. Y. (2014). Beyond seasonal climate: Statistical estimation of phenological responses to weather. *Ecological Applications*, *24*(7), 1793–1802. https://doi.org/10.1890/13-1533.1

Dunham, A. E., Razafindratsima, O. H., Rakotonirina, P., & Wright, P. C. (2018). Fruiting phenology is linked to rainfall variability in a tropical rain forest. *Biotropica*, *50*(3), 396–404. https://doi.org/10.1111/btp.12564

Elmendorf, S. C., Crimmins, T. M., Gerst, K. L., & Weltzin, J. F. (2019). Time to branch out? Application of hierarchical survival models in plant phenology. *Agricultural and Forest Meteorology*, *279*, 107694. https://doi.org/10.1016/j.agrformet.2019.107694

Grandini, M., Bagli, E., & Visani, G. (2020). *Metrics for Multi-Class Classification: An Overview*. https://doi.org/10.48550/ARXIV.2008.05756

Haas, E., La Sorte, F. A., Hanna M. McCaslin, Belotti, M. C., & Horton, K. G. (2022). The correlation between <scp>eBird</scp> community science and weather surveillance radar based estimates of migration phenology. *Global Ecology and Biogeography*. https://doi.org/10.1111/geb.13567

Hossin, M., & Sulaiman, M. N. (2015). A Review on Evaluation Metrics for Data Classification Evaluations. *International Journal of Data Mining & Knowledge Management Process*, *5*(2), 01–11. https://doi.org/10.5121/ijdkp.2015.5201

Hurlbert, A. H., & Liang, Z. (2012). Spatiotemporal variation in avian migration phenology: Citizen science reveals effects of climate change. *PLOS ONE*, *7*(2), 31662. https://doi.org/10.1371/journal.pone.0031662

Kabano, P., Harris, A., & Lindley, S. (2020). Sensitivity of Canopy Phenology to Local Urban Environmental Characteristics in a Tropical City. *Ecosystems*, *24*(5), 1–15. https://doi.org/10.1007/s10021-020-00571-y

Kelly, J. F., Horton, K. G., Stepanian, P. M., de Beurs, K. M., Fagin, T. D., Bridge, E. S., & Chilson, P. B. (2016). Novel measures of continental-scale avian migration phenology related to proximate environmental cues. *Ecosphere*, *7*(9).

https://doi.org/10.1002/ecs2.1434

Kishen, P. (2005). *Trees of Delhi: A Field Guide*. Dorling Kindersley. https://books.google.com/books?hl=en&lr=&id=vw9IDLLkWtEC&oi=fnd&pg=PA24&dq=trees+of+delhi&ots=y1GIyxevI3&sig=j6J7no_ucGZCbJzaKt4JZZCBbVo#v=onepage&q=trees%20of%20delhi&f=false

Li, W., Xin, Q., Zhou, X., Zhang, Z., & Ruan, Y. (2021). Comparisons of numerical phenology models and machine learning methods on predicting the spring onset of natural vegetation across the Northern Hemisphere. *Ecological Indicators*, *131*, 108126. https://doi.org/10.1016/j.ecolind.2021.108126

Makhmale, S., Bhutada, P., Yadav, L., & Yadav, B. K. (2016). Impact of climate change on phenology of Mango– The case study. *Ecology Environment & Conservation*.

McDonough MacKenzie, C., Gallinat, A. S., & Zipf, L. (2020). Low□cost observations and experiments return a high value in plant phenology research. *Applications in Plant Sciences*, *8*(4). https://doi.org/10.1002/aps3.11338

Medway, Lord. (1972). Phenology of a tropical rain forest in Malaya. *Biological Journal of The Linnean Society*, *4*(2), 117–146. https://doi.org/10.1111/j.1095-8312.1972.tb00692.x

Mendoza, I., Peres, C. A., & Morellato, L. P. C. (2017). Continental-scale patterns and climatic drivers of fruiting phenology: A quantitative Neotropical review. *Global and Planetary Change*, *148*, 227–241. https://doi.org/10.1016/j.gloplacha.2016.12.001

Mora, C., Caldwell, I. R., McBryde, E. S., Caldwell, J. M., Fisher, M. R., Genco, B. M., & Running, S. W. (2015). Suitable Days for Plant Growth Disappear under Projected Climate Change: Potential Human and Biotic Vulnerability. *PLOS Biology*, *13*(6), 1. https://doi.org/10.1371/journal.pbio.1002167

Mora, C., Frazier, A. G., Longman, R. J., Dacks, R., Walton, M. M., Tong, E. J., Sanchez, J. J., Kaiser, L. R., Stender, Y. O., Anderson, J. M., Ambrosino, C. M., Fernandez-Silva, I., Giuseffi, L. M., & Giambelluca, T. W. (2013). The projected timing of climate departure from recent variability. *Nature*, *502*(7470), 183–187.

https://doi.org/10.1038/nature12540

Morellato, L. P. C., Alberton, B., Alvarado, S. T., Borges, B. D., Buisson, E., Camargo, M. G. G., Cancian, L. F., Carstensen, D. W., Escobar, D. F. E., Leite, P. T. P., Mendoza, I., Rocha, N. M. W. B., Soares, N. C., Silva, T. S. F., Staggemeier, V. G., Streher, A. S., da Cunha Vargas, B., & Peres, C. A. (2016). Linking plant phenology to conservation biology. *Biological Conservation*, *195*, 60–72. https://doi.org/10.1016/j.biocon.2015.12.033

Neil, K., & Wu, J. (2006). Effects of urbanization on plant flowering phenology: A review. *Urban Ecosystems*, *9*(3), 243–257. https://doi.org/10.1007/s11252-006-9354-2

Parmesan, C. (2007). Influences of species, latitudes and methodologies on estimates of phenological response to global warming. *Global Change Biology*, *13*(9), 1860–1872. https://doi.org/10.1111/j.1365-2486.2007.01404.x

Pau, S., Okamoto, D. K., Calderón, O., & Wright, S. J. (2018). Long-term increases in tropical flowering activity across growth forms in response to rising CO2 and climate change. *Global Change Biology*, *24*(5), 2105–2116. https://doi.org/10.1111/gcb.14004

Pau, S., Wolkovich, E. M., Cook, B. I., Davies, T. J., Kraft, N. J. B., Bolmgren, K., Betancourt, J. L., & Cleland, E. E. (2011). Predicting phenology by integrating ecology, evolution and climate science. *Global Change Biology*, *17*(12), 3633–3643. https://doi.org/10.1111/j.1365-2486.2011.02515.x

Pau, S., Wolkovich, E. M., Cook, B. I., Nytch, C. J., Regetz, J., Zimmerman, J. K., & Wright, S. J. (2013). Clouds and temperature drive dynamic changes in tropical flower production. *Nature Climate Change*, *3*(9), 838–842. https://doi.org/10.1038/nclimate1934

Piao, S., Liu, Q., Chen, A., Janssens, I. A., Fu, Y., Dai, J., Liu, L., Lian, X., Shen, M., & Zhu, X. (2019). Plant phenology and global climate change: Current progresses and challenges. *Global Change Biology*, *25*(6), 1922–1940. https://doi.org/10.1111/gcb.14619

Power, S. B., Delage, F., Chung, C. T. Y., Kociuba, G., Kevin Keay, & Keay, K. (2013). Robust twenty-first-century projections of El Niño and related precipitation variability. *Nature*, *502*(7472), 541–545. https://doi.org/10.1038/nature12580

R Core Team. (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Ramaswami, G., Datta, A., Reddy, A., & Quader, S. (2019). Tracking Phenology in the Tropics and in India: The Impacts of Climate Change. In *The Implications of Climate Change for Indian Biodiversity: An Overview* (pp. 45–59).

Ramaswami, G., & Quader, S. (2018, June 26). The Case of the Confusing Kanikonna Trees. *The Wire*. https://thewire.in/environment/the-case-of-the-confusing-kanikonna-trees

Ramaswami, G., Sidhu, S., & Quader, S. (2021). Using citizen science to build baseline data on tropical tree phenology. *CURRENT SCIENCE*, *121*(11).

Sasidharan, N. (2006). *ILLUSTRATED MANUAL ON TREE FLORA OF KERALA SUPPLEMENTED WITH COMPUTER-AIDED IDENTIFICATION* (No. 282). Kerala Forest Research Institute.

Satake, A., Kawagoe, T., Saburi, Y., Chiba, Y., Gen Sakurai, & Kudoh, H. (2013). Forecasting flowering phenology under climate warming by modelling the regulatory dynamics of flowering-time genes. *Nature Communications*, *4*(1), 2303–2303. https://doi.org/10.1038/ncomms3303

Schaik, C. P. van, Terborgh, J. W., & Wright, S. J. (1993). The Phenology of Tropical Forests: Adaptive Significance and Consequences for Primary Consumers. *Annual Review of Ecology and Systematics*, *2*, 353–377.

Scheffers, B. R., Edwards, D. P., Macdonald, S. L., Senior, R. A., Andriamahohatra, L. R., Roslan, N., Rogers, A. M., Haugaasen, T., Wright, P., & Williams, S. E. (2017). Extreme thermal heterogeneity in structurally complex tropical rain forests. *Biotropica*, *49*(1), 35–44. https://doi.org/10.1111/btp.12355

Shepherd, T. G., & Sobel, A. H. (2020). Localness in Climate Change. *Comparative Studies*

of South Asia, Africa and the Middle East, *40*(1), 7–16. https://doi.org/10.1215/1089201X-8185983

Singh, S. N., Sengupta, S., Kundu, M., Sahay, S., & Kumar, D. (2021). Phenological characterization of different mango genotype under Sabour condition. *The Pharma Innovation Journal*.

Sutton, P., Roberts, D., Elvidge, C., & Baugh, K. (2001). Census from Heaven: An estimate of the global human population using night-time satellite imagery. *International Journal of Remote Sensing*, *22*(16), 3061–3076. https://doi.org/10.1080/01431160010007015

Tutin, C. E. G., & Fernandez, M. (1993). Relationships between minimum temperature and fruit production in some tropical forest trees in Gabon. *Journal of Tropical Ecology*, *9*(2), 241–248. https://doi.org/10.1017/s0266467400007239

Visser, M. E., & Both, C. (2005). Shifts in phenology due to global climate change: The need for a yardstick. *Proceedings of The Royal Society B: Biological Sciences*, *272*(1581), 2561–2569. https://doi.org/10.1098/rspb.2005.3356

Wickham, H. (2020). *plyr: Tools for Splitting, Applying and Combining Data*. https://CRAN.R-project.org/package=plyr

Wright, S. J., & Calderón, O. (2018). Solar irradiance as the proximate cue for flowering in a tropical moist forest. *Biotropica*, *50*(3), 374–383. https://doi.org/10.1111/btp.12522

Zhang, Y., & Fueglistaler, S. (2019). Mechanism for Increasing Tropical Rainfall Unevenness With Global Warming. *Geophysical Research Letters*, *46*(24), 14836–14843. https://doi.org/10.1029/2019GL086058

Zimmerman, J. K., Wright, S. J., Calderón, O., Pagan, M. A., & Paton, S. R. (2007). Flowering and fruiting phenologies of seasonal and aseasonal neotropical forests: The role of annual changes in irradiance. *Journal of Tropical Ecology*, *23*(2), 231–251. https://doi.org/10.1017/s0266467406003890

Tables

Table 1: Species descriptions.

Taxonomic description, appearance of tree and reproductive parts, phenology in Kerala and description of the SeasonWatch dataset for jackfruit, mango and tamarind

	Jackfruit	**Mango**	**Tamarind**
Accepted Scientific name	*Artocarpus heterophyllus*	*Mangifera indica*	*Tamarindus indica*
Name in Malayalam	Pilavu/Plavu (tree), chakka (fruit)	Maavu/Moochi (tree), manga (fruit)	Puli, Valan puli, Kolpuli
Family	Moraceae	Anacardiaceae	Fabaceae
Description (Trees of Delhi)	A handsome evergreen tree with glossy leaves up to 20 m tall in ideal conditions. Cultivated for its misshapen fruits - largest edible fruit in the world	Semi-evergreen trees with long, glossy leaves that can grow up to 35 m in favourable sites. Bark is grey-brown and rough with shallow cracks.	Large, handsome, long-lived tree with a squat trunk and a shady crown. Introduced to India from E Africa, but virtually indigenous. The bark is dark grey, rough with shallow fissures
Flower description	Male and female separate, on the	Mango flowers are tiny, yellowish-green,	Small, with 3 unequal-sized yellow

	same tree. Male flowers tiny, in clusters emerging from new branchlets; female flowers densely crowded on old branches or directly from the trunk	strongly scented, in huge, branched clusters. Most flowers are male and the rest bisexual.	petals veined with red; flower cup has 4 creamy segments
Fruit description	Gigantic, lumpy, barrel- or pear-shaped; pollinated by small flies and beetles	Fruit is smooth-skinned, waxy and in various colours depending on the variety. Flesh is fibrous or pulpy and the stone flattened, kidney-shaped	Beanlike pod bulging over the seeds, up to 20 cm long; green at first, ripening cinnamon-brown. Young pod is covered with a downy felt.
Common varieties in Kerala	Varikka, koozha, etc	Muvandan, chandrakkaran, etc	
Uses	Timber, fruit	Fruit	Fruit, timber
Fruiting and flowering months in Kerala	November to April	January to May	September to April
SeasonWatch dataset			

Total number of observations	72097	68696	24213
Total number of trees	7927	9919	1750

Table 2: Variables, data sources and transformations

Datasets	**Data Source**	**Transformations used**	**Variables derived**
Flowering observations	SeasonWatch		Flowering binary and ordinal
Fruiting observations	SeasonWatch		Fruiting binary and ordinal
Precipitation	ERA5-Land	Sum	Total precipitation, DryDays
Temperature	ERA5-Land	Max, min, mean	Max temp, min temp, mean temp
Soil Moisture	ERA5-Land	Mean (of total column depth at each time step), Mean (hourly -> Daily -> Fortnightly)	Mean SM

Solar radiation	ERA5-Land	Sum (hourly -> daily -> Fortnightly)	
elevation	SRTM	terrain functions	Elevation, slope, aspect
urbanisation	NASA nightlights		urbanisation

Figures

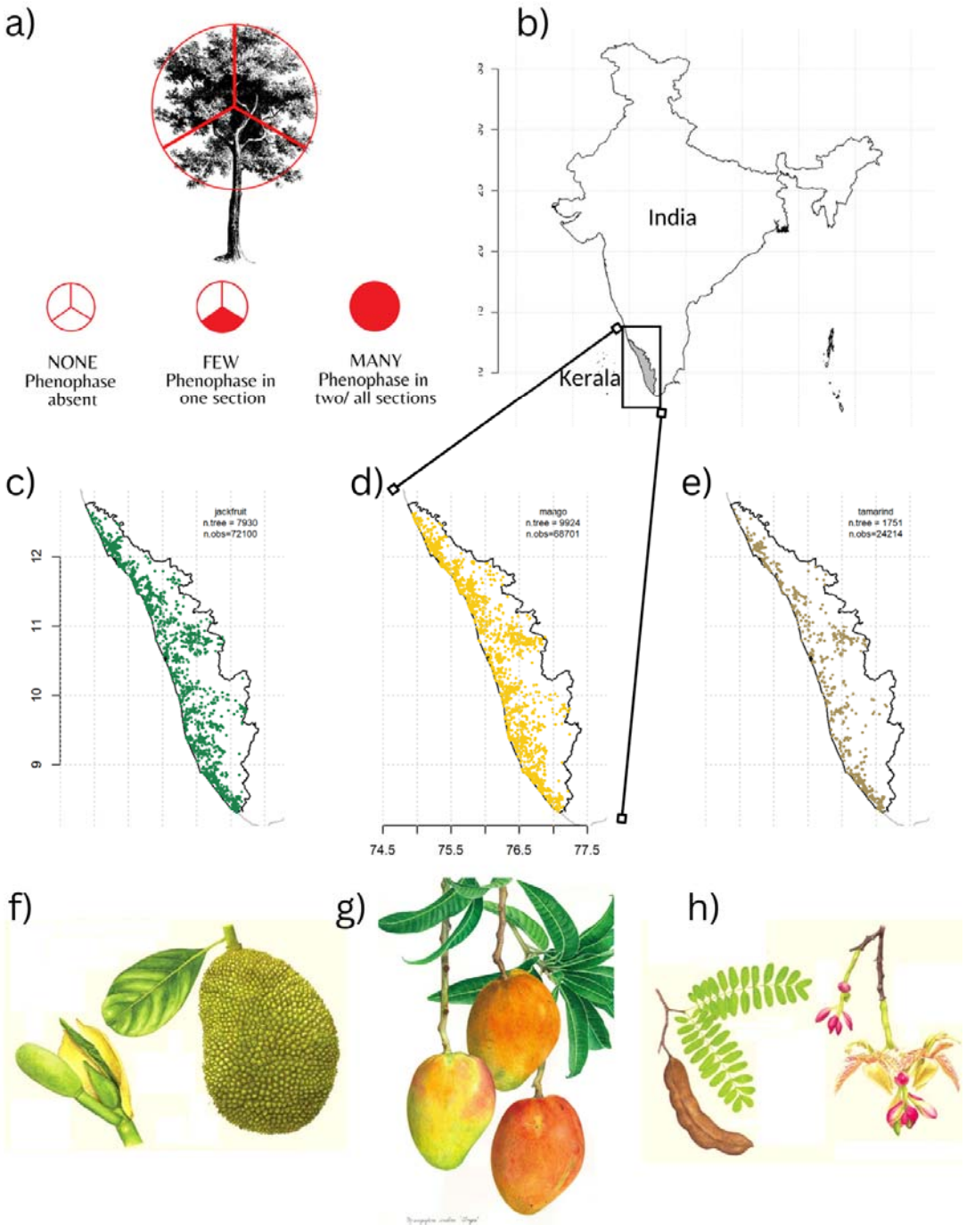
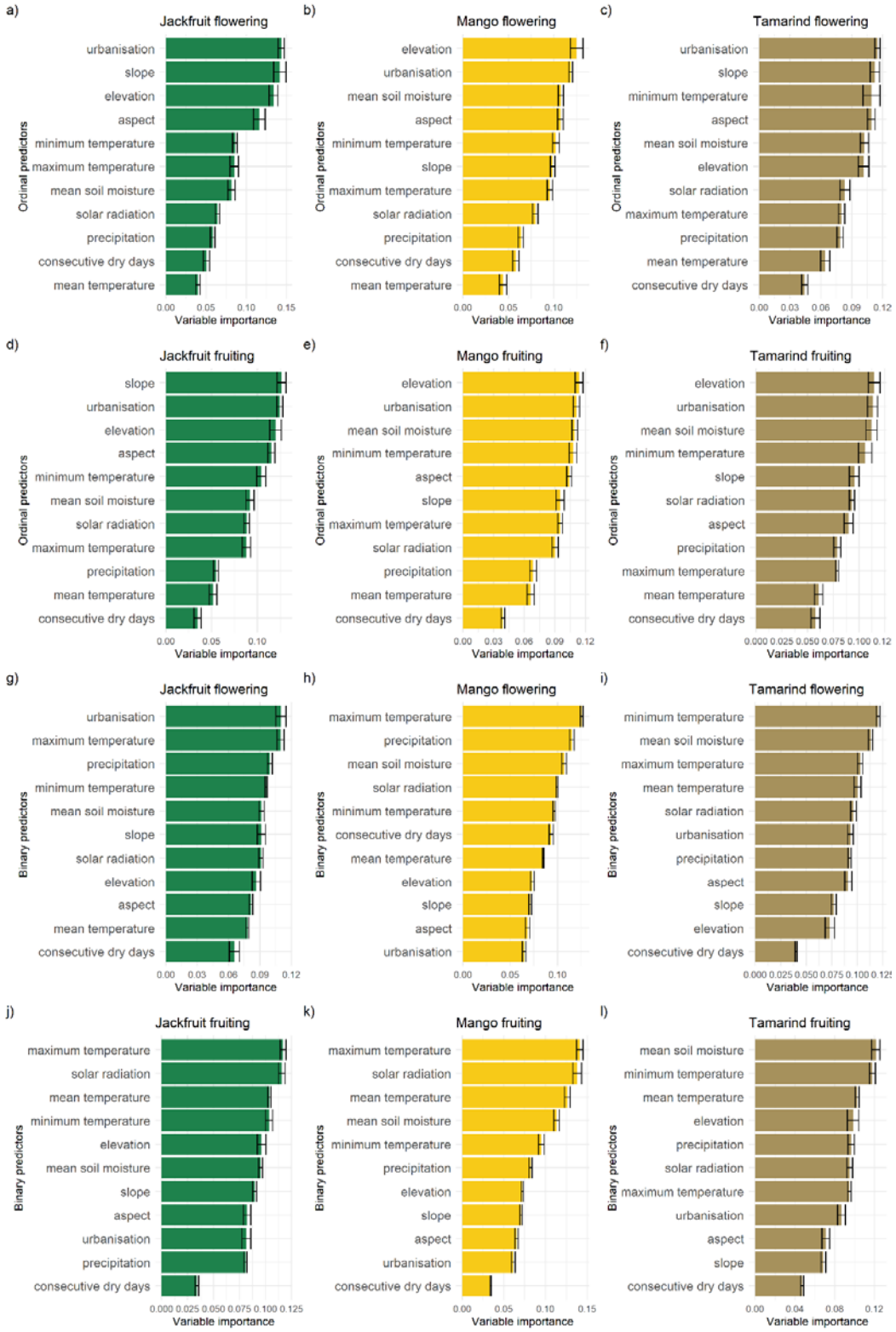


Figure 1: SeasonWatch protocol, species and map of observations.

a) SeasonWatch observation protocol for tree phenology - phenophase quantities ‘none’, ‘few’, ‘many’ roughly correspond to the presence of a phenophase in 0, 30%, and >30% of

the tree canopy b) Outline of India with the state of Kerala marked in grey. c), d) and e) - map of observations of jackfruit, mango and tamarind trees across Kerala, considered in this study. f), g) and h) illustrate flowering and fruiting phenophases in jackfruit, mango and tamarind respectively. Image credits: a) Sayee Girdhari, f) Neelam Modi for SeasonWatch, g) fairchildgarden.org, h) Neelam Modi for SeasonWatch


a)
Jackfruit flowering
urbanisation
slope
elevation
aspect
minimum temperature
maximum temperature
mean soil moisture
solar radiation
precipitation
consecutive dry days
mean temperature
Ordinal predictors
0.00 0.05 0.10 0.15
Variable importance
b)
Mango flowering
elevation
urbanisation
mean soil moisture
aspect
minimum temperature
slope
maximum temperature
solar radiation
precipitation
consecutive dry days
mean temperature
Ordinal predictors
0.00 0.05 0.10
Variable importance
c)
Tamarind flowering
urbanisation
slope
minimum temperature
aspect
mean soil moisture
elevation
solar radiation
maximum temperature
precipitation
mean temperature
consecutive dry days
Ordinal predictors
0.00 0.03 0.06 0.09 0.12
Variable importance
d)
Jackfruit fruiting
slope
urbanisation
elevation
aspect
minimum temperature
mean soil moisture
solar radiation
maximum temperature
precipitation
mean temperature
consecutive dry days
Ordinal predictors
0.00 0.05 0.10
Variable importance
e)
Mango fruiting
elevation
urbanisation
mean soil moisture
minimum temperature
aspect
slope
maximum temperature
solar radiation
precipitation
mean temperature
consecutive dry days
Ordinal predictors
0.00 0.03 0.06 0.09 0.12
Variable importance
f)
Tamarind fruiting
elevation
urbanisation
mean soil moisture
minimum temperature
slope
solar radiation
aspect
precipitation
maximum temperature
mean temperature
consecutive dry days
Ordinal predictors
0.000 0.025 0.050 0.075 0.100 0.12
Variable importance
g)
Jackfruit flowering
urbanisation
maximum temperature
precipitation
minimum temperature
mean soil moisture
slope
solar radiation
elevation
aspect
mean temperature
consecutive dry days
Binary predictors
0.00 0.03 0.06 0.09 0.12
Variable importance
h)
Mango flowering
maximum temperature
precipitation
mean soil moisture
solar radiation
minimum temperature
consecutive dry days
mean temperature
elevation
slope
aspect
urbanisation
Binary predictors
0.00 0.05 0.10
Variable importance
i)
Tamarind flowering
minimum temperature
mean soil moisture
maximum temperature
mean temperature
solar radiation
urbanisation
precipitation
aspect
slope
elevation
consecutive dry days
Binary predictors
0.000 0.025 0.050 0.075 0.100 0.125
Variable importance
j)
Jackfruit fruiting
maximum temperature
solar radiation
mean temperature
minimum temperature
elevation
mean soil moisture
slope
aspect
urbanisation
precipitation
consecutive dry days
Binary predictors
0.000 0.025 0.050 0.075 0.100 0.125
Variable importance
k)
Mango fruiting
maximum temperature
solar radiation
mean temperature
mean soil moisture
minimum temperature
precipitation
elevation
slope
aspect
urbanisation
consecutive dry days
Binary predictors
0.00 0.05 0.10 0.15
Variable importance
l)
Tamarind fruiting
mean soil moisture
minimum temperature
mean temperature
elevation
precipitation
solar radiation
maximum temperature
urbanisation
aspect
slope
consecutive dry days
Binary predictors
0.00 0.04 0.08 0.12
Variable importance

833

Figure 2: Variable importance as calculated using the MDA method from machine learning models of fruiting and flowering occurrence and intensity.

a), b), c) represent MDA measures for flowering intensity and d), e), f) for fruiting intensity of jackfruit, mango and tamarind respectively as modelled using the CatBoot algorithm. g), h), i) represent MDA measures for flowering intensity and j), k), l) for fruiting intensity of jackfruit, mango and tamarind respectively as modelled using Random Forest models. Variable importance has been calculated as an ensemble average from LOOCV models that leave one year out each and variables are ordered in their decreasing order of mean importance.

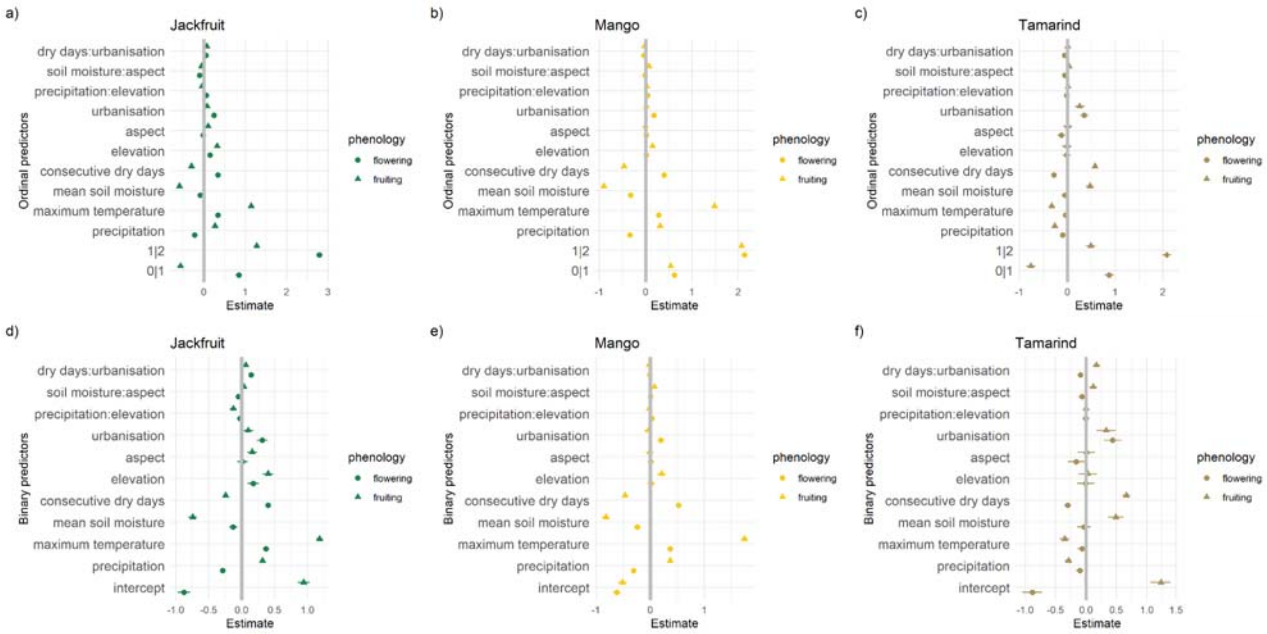
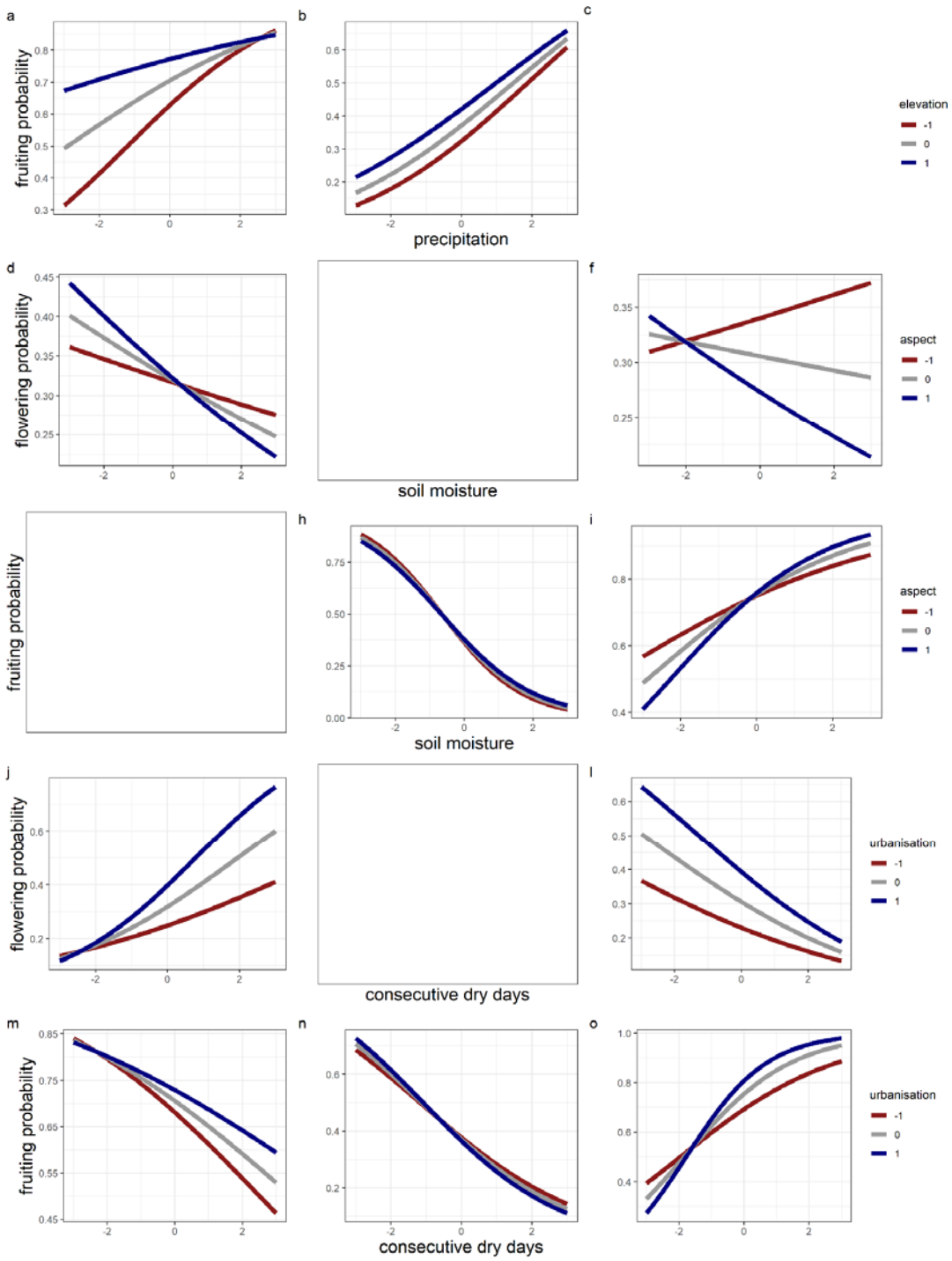


Figure 3: Mean effect sizes of ordinal and binary models with leave-one-year-out cross validation. Top row shows effect sizes of all main effects for the ordinal model; bottom row for the binary model. Shapes indicate models for flowering and fruiting and the grey line is the zero effect line.


a
b
c
fruiting probability
precipitation
elevation
-1
0
1
d
f
flowering probability
soil moisture
aspect
-1
0
1
fruiting probability
h
i
soil moisture
aspect
-1
0
1
j
l
flowering probability
consecutive dry days
urbanisation
-1
0
1
m
n
o
fruiting probability
consecutive dry days
urbanisation
-1
0
1

851

Figure 4 - Predicted responses of flowering and fruiting phenology of jackfruit, mango and tamarind trees (left to right) to interactions between static and time-varying environmental predictors.

Mean and interaction effects derived from generalised linear mixed models of each species using binary response variables. Flowering and fruiting probability are each represented as a function of each time-varying environmental variable at three levels of the static variable - mean value and mean+/- sd for the range of observed values, with the low value represented as red, mean as grey and high as red. Interactions that are not significant are represented as blank plots. One row of plots - effect of precipitation x elevation on flowering probability - was omitted because the interaction effect was not significant for all three species.

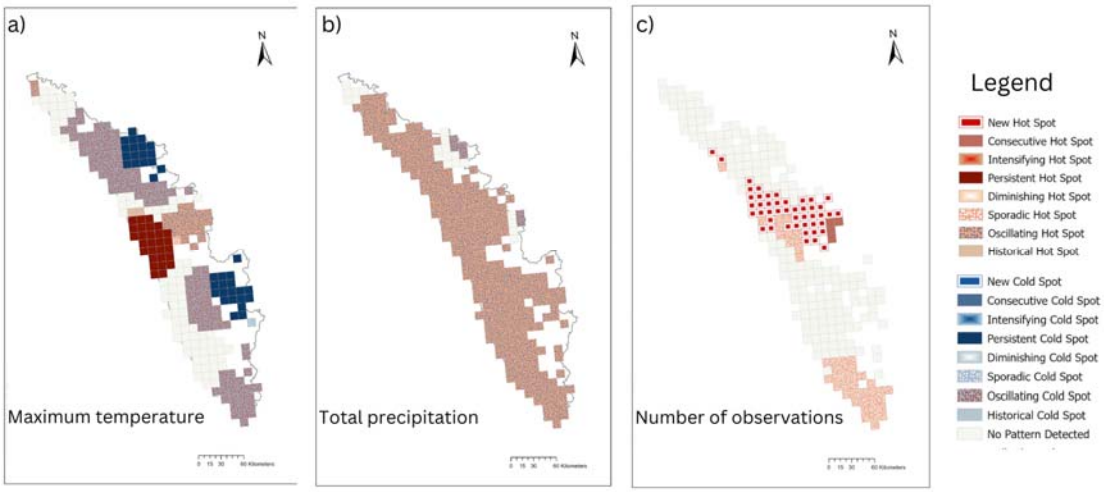


Figure 5 - Hotspot analysis of environmental data and citizen science observations using spatiotemporal data from 2014-2022.

Hotspot analyses for a) maximum temperature within a fortnight, b) total precipitation within the fortnight and c) total number of citizen-science phenology observations recorded in

SeasonWatch in a fortnight. Observations have been aggregated at the 11 km x 11 km pixel size to match environmental data resolution.