

Initial Experiments with a Scalable Machine Learning Based Approach for Downscaling the MOD16A2 Evapotranspiration Product

ABSTRACT

There exist many inequalities related to Groundwater use in India which poses a need to monitor groundwater levels in different rural areas to address and resolve these inequalities. Evapotranspiration constitutes a significant component of the groundwater equation, but traditional datasets that provide evapotranspiration data have sparse spatial and temporal resolution which hampers the accuracy in localised decision making of water-related applications. In this study we employ the machine learning algorithms like random forest to develop downscaling models capable of predicting high-resolution evapotranspiration. The approach leverages remotely sensed data, meteorological variables, and land surface characteristics as input features to capture the complex relationships governing evapotranspiration data. The results demonstrate the efficacy of machine learning to reproduce fine-scale evapotranspiration data which is validated using observational data from multiple geographic locations, representing diverse land use and land cover conditions. The study underscores the potential of machine learning to produce downscaled evapotranspiration maps for PAN-India using google earth engine, which would act as a valuable tool in water-resource management and climate change impact assessments. This research contributes a scalable and adaptable solution to address the growing demand for fine-resolution hydro-climatic information.

CCS CONCEPTS

- Social and professional topics; • Computing methodologies
→ Machine learning;

KEYWORDS

Downsampling, Evapotranspiration, Water-balance, Random Forests, Regression, Hydrology, Landsat

ACM Reference Format:

. 2018. Initial Experiments with a Scalable Machine Learning Based Approach for Downscaling the MOD16A2 Evapotranspiration Product. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (ACM COMPASS 2024)*. ACM, New York, NY, USA, 16 pages. <https://doi.org/XXXXXXX.XXXXXXX>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ACM COMPASS 2024, July 08–11, 2024, New Delhi, India

© 2018 Copyright held by the owner/author(s). Publication rights licensed to ACM. ACM ISBN 978-1-4503-XXXX-X/18/06
<https://doi.org/XXXXXXX.XXXXXXX>

1 INTRODUCTION

Evapotranspiration stands as a key element in the water cycle, and the importance of precisely estimating it is evident. Evapotranspiration is an integral part of the water balance equation that is used to estimate change in groundwater levels in a region, and has various applications like building water balance models [5]. Evaporation (ET) consists of various components: (vegetation transpiration (E_c), soil evaporation (E_s), and canopy interception evaporation (E_i)) [5], [19], component wise analysis can provide a more precise quantification of water losses from different sources, and it can help identify which processes contribute the most to the overall water loss. This is crucial in regions where water scarcity is a concern, and for fields like drought monitoring and water requirement assessment [12]. It has applications in hydroecology [22], and in agriculture for ecosystem and climatic assessments, efficient use of water resources and precise irrigation scheduling operations [11], [12], agricultural water allocation to different places, irrigation effective monitoring of irrigation and efficient irrigation anagement [8] and in designing effective land management and water conservation strategies [12]. Enhancing the precision of evapotranspiration estimation can boost the effectiveness of hydrological models within agricultural settings. This improvement contributes to more comprehensive investigations into the effects of water management interventions, aiming to achieve environmental sustainability while also enhancing agricultural production, household income, and nutrition [7].

However, there are practical challenges to measure actual evapotranspiration at a fine resolution and with higher frequency of acquisition [31]. One of the data products that gives ET for India is MODIS from NASA, which is available at a coarse resolution of 500m pixels, which does not give a detailed information about variations of ET within the study area. At a 500m resolution, individual pixels may contain a mix of land cover types. This can lead to flaws in estimating ET for specific land use land cover types. Finer resolutions of a scale of 30 m can help in distinguishing and characterizing the variations in ET for different land use land cover types accurately. Downsampling ET will also segregate the ET components, because at a lower resolution like 30m, we can assume a single component contributing most to ET of that pixel. Segregation of ET into its components would be useful for optimisation of water management practices, stream modelling and hydro-logical assessments[19]. It will distinguish between soil evaporation and plant transpiration, which will give a nuanced understanding about the different land use land cover types and their effects on ET [37], which in turn will lead to informed decision making in hydrology and climate change [22].

Our work makes the following contributions:

- Implementation of a Random Forest model that can output the downscaled ET at a scale of 30m across all regions of India in Google Earth Engine, which makes it scalable for extensive utilisation for ecological and hydrological models.

- Visualization of the downscaled ET output using GEE's visualization tools can help in spatial and temporal analysis of trends of ET at the region of interest.

To evaluate our downscaling method, we leverage the underlying assumption of scale invariance. If the model demonstrates effectiveness at coarser resolutions during assessment, it can be used for application to finer resolutions. This assumption has been employed in previous publications [14], [6] for the assessment of downscaling techniques. Our uniqueness and points of departure from these papers have been discussed in section 2.2. Our method aims to refine the spatial resolution of ET, capturing local variations in land cover and other factors influencing ET by utilising machine-learning model that captures the equation that underlies between different factors and ET. This is different from the regional calibration methods that involves adjusting the parameters of the hydro-logical model (Modis ET here) to improve its accuracy within a specific geographic region. Regionally calibrating MODIS ET against ground truth measurements for increased accuracy remains an important area for future research.

2 BACKGROUND

2.1 Penman–Monteith method

The FAO Penman–Monteith (PM) method ([1]) can be used to approximate net ET from meteorological data, which was derived by the United Nations Food and Agriculture Organization for modelling reference evapotranspiration ET_0 . The method uses equation 1

$$ET_0 = \frac{0.408\Delta(R_n - G) + \gamma \frac{900}{T+273} u_2 \delta e}{\Delta + \gamma(1 + 0.34u_2)} \quad (1)$$

where ET_0 is the Reference evapotranspiration which measures Water volume evapotranspired (in $m^3 day^{-1}$), Δ represents Rate of change of saturation specific humidity with air temperature (in $Pa K^{-1}$), R_n is the Net irradiance (in $MJ m^{-2} day^{-1}$) which is the external source of energy flux G is Ground heat flux ($MJ m^{-2} day^{-1}$), usually equivalent to zero on a day, T = Air temperature at 2m(K), u_2 is the wind speed at 2m height (in m/s), δe is vapor pressure deficit (in kPa) and γ is Psychrometric constant (where $\gamma = 66 Pa K^{-1}$).

The MOD16 ET algorithm that is used to output the MODIS ET product is based on the Penman-Monteith equation, and is calibrated based on remotely sensed data for various regions [25]. As non-domain experts, our preference was not to undertake these computations ourselves but rather to utilize the most accurate outputs from other models. The main challenge of the MOD16 algorithm that used equation 1 was the unavailability of base data products of input variables at a finer resolution. It involves various complexities in terms of multiple meteorological parameters like air temperatures, relative humidity, wind speed, and solar radiation. A similar algorithm can not be used to produce ET for all over India at a fine resolution, because obtaining a fine resolution dataset of these constituent parameters for all locations in India is practically challenging.

2.2 Related work

Related research in this field provides a comprehensive overview of different methods used to downscale ET, their evolution, innovations and critical gaps. Earlier works on comparing methods to downscale evapotranspiration by [13] evaluated the performance of spatial interpolation techniques i.e., Inverse Distance Weighing (deterministic method), Ordinary and Universal Kriging (geostatistical method) to downscale Reference Evapotranspiration (ET_0), both of which produced low estimation errors on different locations. However a comprehensive comparison of spatial interpolation methods of environmental variables by [20] and [4] has shown higher prediction accuracy in hybrid interpolation models involving machine learning models like Regression and Random Forest, out of which Random Forest shows the highest correlation with ground truth values of Evapotranspiration. Numerous models have been developed over the recent years to estimate regional ET at a higher resolution (using ensemble Kalman Filter in [26], artificial neural networks in [24], support vector machines (SVM) in [29], regression trees in [33] and Bayesian model averaging methods in [36]). The approach of downscaling Evapotranspiration using a hybrid model of synthetic wavelets and support vector machines by [16] showed high accuracy in prediction of ET. [6] modelled a solution to downscale medium resolution MODIS ET using meteorological and Landsat data by SVM and MLR for north-eastern parts of India.

However, the downscaled ET products have not been yet developed for the entire India, in a form that can be used for hydro-logical and agricultural assessments like local water management, irrigation scheduling, and monitoring water management budgets. In the context of our current research, we aim to address the gaps in previous research, by producing a downscaled ET product at 30 meter resolution for whole India, in a usable form on Google Earth Engine that can be used further for hydro-logical purposes. We add few more climatic variables as features to improve the accuracy of the model, which reflects the dependency of ET on meteorological data as in section 2.1. We opt for Random Forest model among the model utilised in previous research, due to its scalability for integration into Google Earth Engine. We use publicly available dataset on Google Earth Engine (GEE) for our study which is openly accessible facilitating its application in subsequent hydro-logical models and related analyses.

3 DATASET

3.1 MODIS

Evapotranspiration value obtained from Terra Moderate Resolution Imaging Spectroradiometer (MODIS) product MOD16A2 version 6.1 ([28]) is used as the target variable. The algorithm used for the MOD16 data product collection is based on the logic of the Penman-Monteith equation. It is an 8-day composite dataset produced at 500 meter (m) pixel resolution, whose value of ET band represents the sum of Evapotranspiration of all eight days within the composite period.

3.2 Landsat

We use the Landsat 8 data from Operational Land Image (OLI) sensor, (courtesy U.S. Geological Survey and NASA). It provides

bands that describe vegetation types, cultural features, biomass, vigor etc. at a resolution of 30 m. Landsat imagery captures 185km swath as a part of its orbit, and it acquires images of the same location on Earth approximately every 16 days, the *average revisit time* ([23]). The Landsat bands were processed to compute 9 indices namely Normalized Difference Vegetation Index (NDVI), Soil-Adjusted Vegetation Index (SAVI), Normalized Difference Built-up Index (NDBI), Normalized Difference Water Index (NDWI), Modified Soil-Adjusted Vegetation Index (MSAVI), Normalized Difference Moisture Index (NDMI), and Normalized Difference Infrared Index for Band 7 (NDIIB7), Surface Albedo, Land Surface Temperature (LST), at 30 m resolution for the training and test areas, calculated as shown in table 1. The indices NDVI, SAVI, MSAVI are associated with vegetation cover and NDBI, NDWI, NDMI and NDIIB7 are correlated with water consumption that is reflected from ET processes ([17]).

3.3 NOAH-GLDAS

The Penman-Monteith equation 1 incorporates various meteorological variables like humidity, air temperature, irradiance, ground heat flux, wind speed and vapor pressure to estimate ET accurately. By using a interplay of these variables as input features, we can downscale ET to a lower resolution. The meteorological variables include Rain precipitation rate ($kgm^{-2}s^{-1}$), Root zone soil moisture (kgm^{-2}), Soil moisture (kgm^{-2}), Canopy Surface Water (kgm^{-2}), Specific humidity, Wind Speed (ms^{-1}), Pressure (Pa), Soil Temperature (K), Base flow Groundwater runoff (kgm^{-2}), Short wave radiation flux (Wm^{-2}), Heat flux (Wm^{-2}) and Air Temperature (K) were extracted from NASA GLDAS version-2 products [2], refer to table 2). It provided meteorological variables at a temporal resolution of 3 hours with a spatial resolution of 27830 meters.

4 METHODOLOGY

Figure 1a represents all the important components our pipeline. As shown, we train our model on the basis of Agro-ecological zones classified as per [10]. These zones provide a meaningful way to characterize specific environmental and agricultural characteristics of different regions. For each zone, we choose all the tiles corresponding to tiles of Landsat 8 satellite image. For all such regions inside the AEZ, we select the dates on which the Landsat and Climatic features both are available. Further, the images are brought to same resolution and flattened to generate the samples as shown in 1b. Finally we merge the datasets of all the neighbouring Agro-eco zones. We call this collection of neighbouring Agro-eco zones as a cluster of AEZ.

4.1 Data Preprocessing

The selected images of GLDAS NOAH and Landsat 8 are filtered to match the dates with MODIS ET images as shown in Figure 2c. Landsat 8 and GLDAS NOAH images whose acquisition date comes within the 8-day composite window of corresponding MODIS images are selected. The ET value obtained from this MODIS image is the estimate of ET for the corresponding dates of Landsat indices and meteorological variables, since MODIS provide ET values 8-day composite values of this window. The selected images are preprocessed for the computation of Landsat derived indices (as in 1) and

Meteorological variables with MODIS ET value added to their corresponding input images. The actual ET provided as input to the random forest model is the $\frac{1}{8}$ th of the value of 8-day composite ET that is provided by MODIS, where the Landsat and GLDAS-NOAH acquisition date lie within the 8-day period. We have conducted all experiments by taking this value as the corresponding ET value for the acquisition day of inputs, considering that the value of ET on each day within the 8-day composite period will be $\frac{1}{8}$ th of the composite value (assuming approximate uniform distribution of ET for the whole 8-day period [18]).

For alignment of input features with ET for training purpose, we choose the spatial characteristics of MODIS image as the common reference system and align input features with it, according to Figures 2a and 2b. To match the spatial resolution of Landsat 8 indices which is originally at 30 meters with the resolution of target MODIS ET with resolution 500 meters, the first step is aggregation. During the aggregation process, the original data is reduced to a smaller number of pixels, where each pixel represents a larger area than the original pixel size. There are three popular aggregation methods namely averaging, median and central-pixel sampling, each of which involves producing a single value over a window size of $n \times n$ which will now correspond to a single pixel (here $n = 500$). Averaging method is considered to be most suitable for aggregating remotely sensed data, because every pixel of the image represents the integrated data of the corresponding area on earth ([3]). So we use average aggregation to upsample the Landsat 8 images by computing the mean value of all 30 meter pixels of Landsat Image, that are located within each 500 meter pixel of MODIS ET image. This is followed by reprojection of the Landsat 8 image to match the Coordinate Reference System of MODIS image. For spatial resolution of meteorological variables which are originally at 27830 meters resolution with target MODIS ET at resolution 500 meters, we resample the GLDAS NOAH images to 500 meter resolution by converting each pixel of 27830 meter to multiple pixels of 500 meter each, then reproject it to match the Coordinate Reference System of MODIS image. For the training process, everything needs to be sampled to the resolution and projection of MODIS image, as shown in figure 1b.

4.2 Train and Test sets

Post temporally and spatially aligning the Landsat 8 and NOAH GLDAS images with MODIS ET as per figure 2, each pixel of the input images are now at a resolution of 500m. Each input image contains the Landsat derived indices, meteorological variables and ET as its bands.

India experiences 3 primary agricultural seasons namely Kharif (Jul-Oct), Rabi (Jan-Feb, Nov-Dec), Zaid (Mar-Jun), and to capture variations in ET across different seasons, we sample training points uniformly from each season of a particular year. We sample 200 pixels from each input image, and as there are roughly six images that satisfied temporal alignment condition per season, we sample 1200 pixels per season for a year. The total number of sampled pixels in terms of Number of Tiles are:

$$1200 \times 3(3 \text{ seasons per year}) \times k (\text{Number of Tiles chosen}) \times l (\text{Number of Zones}) \times 6 (6 \text{ images obtained for 6 years: 2016-21}) = 21600 \times$$

Landsat Index	Formula	Bands	Reference
NDVI	$\frac{NIR - R}{NIR + R}$	NIR: Band 5, R: Band 4	[27]
SAVI	$\frac{NIR - R}{NIR + R + L} (1 + L)$	$L = 0.5$, NIR: Band 5, R: Band 4	[34]
NDBI	$\frac{SWIR - NIR}{SWIR + NIR}$	SWIR: Band 6, NIR: Band 5	[34]
NDWI	$\frac{Green - NIR}{Green + NIR}$	Green: Band 3, NIR: Band 5	[34]
MSAVI	$\frac{2 * NIR + 1 - \sqrt{(2 * NIR + 1)^2 - 8 * (NIR - R)}}{2}$	NIR: Band 5, R: Band 4	[21]
NDMI	$\frac{NIR - SWIR1}{NIR + SWIR1}$	NIR: Band 5, SWIR2: Band 7	[32]
NDIIB7	$\frac{NIR - SWIR2}{NIR + SWIR2}$	NIR: Band 5, SWIR2: Band 7	[9]
Surface Albedo	$((0.356 * B1) + (0.130 * B2) + (0.373 * B3) + (0.085 * B4) + (0.072 * B5) - 0.018) / 1.016$	B1: Band 1, B2: Band 2, B3: Band 3 B4: Band 4, B5: Band 5	[30]
LST	$K_1 * BT + K_2$	BT: Temperature Band 10 $K_1 : 0.00341802, K_2 : 149.0$	[27]

Table 1: Formula for calculating Landsat indices used in the random forest model to estimate ET

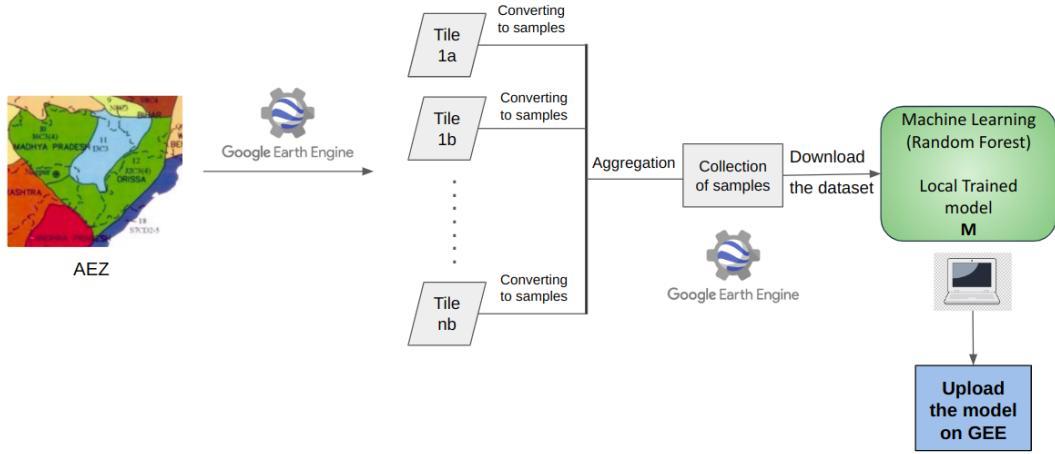
Climatic Variable	Band
Rain precipitation rate	Rainf_tavg
Root zone soil moisture	RootMoist_inst
Soil moisture	SoilMoi0_10cm_inst
Canopy surface water	CanopInt_inst
Average Surface Skin Temperature	AvgSurfT_inst
Specific humidity	Qair_f_inst
Wind speed	Wind_f_inst
Pressure	Psurf_f_inst
Soil Temperature	SoilTMP0_10cm_inst
Base flow groundwater runoff	Qsb_acc
Net Short wave radiation flux	Swnet_tavg
Net Long wave radiation flux	Lwnet_tavg
Heat flux	Qg_tavg
Sensible heat net flux	Qh_tavg
Latent heat net flux	Qle_tavg
Downward short wave radiation flux	SWdown_f_tavg
Air temperature	Tair_f_inst

Table 2: The climatic variables that are used in the Random Forest Model, sourced from NOAH-GLDAS

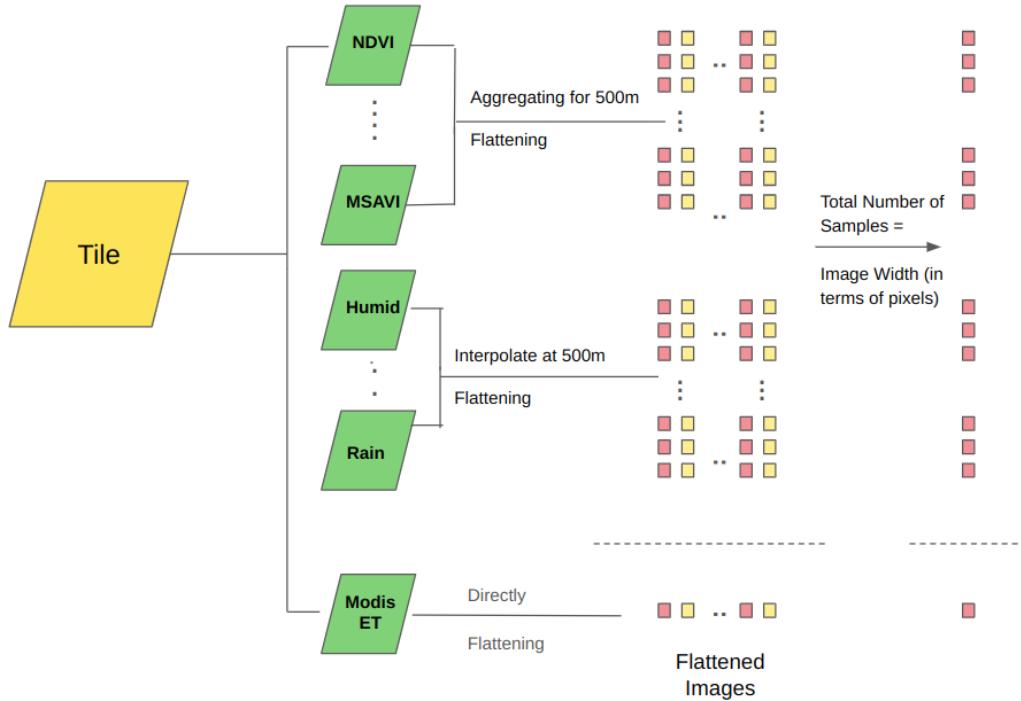
$k \times l$. We assume that tiles chosen from every zone is approximately equal.

We uniformly sample 2 hundred thousand points out of these to constitute the training data to train the model representing the cluster of AEZs. We chose this method of sampling to limit the number of samples in a dataset and to diversify the dataset. Sampling many pixels for a specific date was not useful, as most of them showed similar characteristics.

To validate our model, we harness the scale invariant relationship between remotely sensed input data like Landsat indices and climatic variables and target variable i.e. ET ([14]) as explained in section 1. This directs us to validate our Random Forest model at 500 meter resolution and compare the predicted ET output at 500 meter with the MODIS ET at 500 meter. If the model successfully



(a) Sampling tiles for multiple dates for a particular AEZ. We choose two different tiles for each date (when Landsat is published).



(b) Generating a dataset corresponding to one tile for one particular date

Figure 1: Pipeline for Dataset Extraction and Training the Model

learns the relation between Evapotranspiration (ET) and input features at 500 meters, it will be capable of generating downscaled ET products at 30 meters from the detailed resolution input features. For this purpose, we perform evaluation under different validation-schemes to understand how the variability between training points and validation points impact the model's performance. Let $AEZ_{clusterpart}$ be the AEZ which is involved in training dataset of Machine Learning Model and AEZ_{others} be other zones which are not involved in

training process. We select 2 set of locations for validation. For each of the 2 sets of regions, Landsat 8, NOAH-GLDAS and MODIS-ET images were extracted uniformly for a year different than years chosen for Training from all 3 seasons and aligned temporally as well as spatially (to 500 m resolution and reprojected to MODIS projection) as in section 4.1). Choosing different year would be a stronger criteria of validation as in we can also show that the assumptions we took could be generalized. The schemes are defined

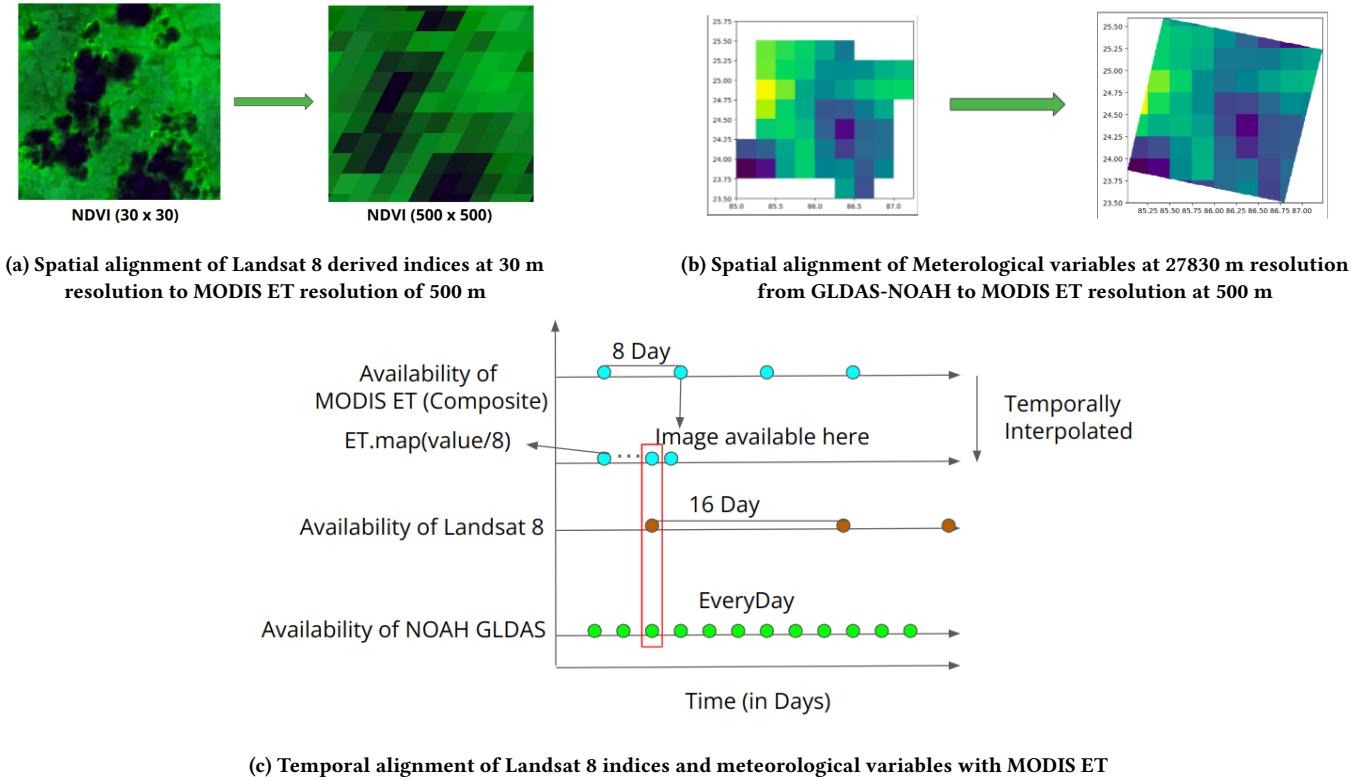


Figure 2: Spatial and Temporal alignments of input features with target variable

as : (1) *Validation Region 1*: Regions of $AEZ_{clusterpart}$ (2) *Validation Region 2*: Region from AEZ_{others} . Validation scheme 2 will provide us a basis for creating models based on different Agro-eco zones. 20000 validation points were uniformly sampled for each of these set of regions, and used to assess the model's prediction accuracy at 500m resolution.

4.3 Model development

We train a random forest model on the collection of training points and store the random forest model corresponding to each cluster. Earlier research (section 2.2) on downscaling ET have shown highest accuracy for using random forest or SVM ([29] and [4]). Additionally, it is also available to be used directly on Google Earth Engine which makes it scalable for incorporation into larger models. Random Forest is an ensemble method, which increases its regularisability. As random forest has inherent parallelism, it can efficiently use GEE's parallel processing capabilities. So, we prefer Random Forest model to other machine learning algorithms. The trained model can be used to predict the ET corresponding to any region belonging to the corresponding cluster on which it is trained. The models are validated for all the validation regions as discussed in section 4.2. We perform grid search on the set of hyperparameters (number of estimators and depth of each estimator in the random forest) and the best performing hyper-parameter tuned random forest model (that is scalable to GEE) is stored for the cluster as an asset in GEE for future use.

5 IMPLEMENTATION

The algorithm is implemented in javascript on Google Earth Engine (GEE) and is available to generate 30 m resolution ET for any region of India. Source code available on github. It outlines the following steps:

- (1) Exporting training data and validation data in the form of CSV files to Google Drive, in Google Earth Engine.
 - (a) Extraction of Landsat 8, NOAH-GLDAS and MODIS-ET images filtered for the specified years on which we want to train. These years are provided as configurable input parameters.
 - (b) Clipping the images for the geometries of primary regions (configurable).
 - (c) Temporally aligning input and MODIS-ET by ensuring that the input data falls within the composite 8-day interval of the MODIS-ET acquisition, achieved by date difference checking.
 - (d) Aggregate to 500 meter resolution (using mean aggregation) and reproject Landsat 8 images to MODIS-ET projection. Resample NOAH-GLDAS images to 500 meter scale and reproject to match MODIS-ET projection.
 - (e) Compute the indices present in table 1, and add the climatic variables in section 3.3 and MODIS-ET values for the corresponding pixels to Landsat images in GEE script.
 - (f) Sample a subset of pixels (as described in 4.2) from the selected Landsat images and export the pixels (in form of tuple of input features and ET values) to a CSV file in Google drive. The filenames are configured as input. Remove the samples

- with any null values. Divide the values of ET by 8 in all the samples as explained in 4.1 for Temporal Alignment.
- (2) Training the Random Forest Model using python locally. Due to limitations of Earth Engine in terms of size of training dataset, hyperparameters tuning and tools for analysis, we have considered the task to be performed locally.
 - (a) The RandomForestRegressor (with tuned hyperparameters) from sklearn is trained on the training data from drive.
 - (3) Export the trained Random Forest Model to GEE as an asset.
 - (a) Serialise the trained RandomForestRegressor to a list of strings (where each string represents a decision tree)
 - (b) Convert the list of strings to a ee.FeatureCollection and export the ee.FeatureCollection to Google Earth Engine as an asset.
 - (4) Use the trained Random Forest Model asset to output 30 meter downscaled ET on GEE
 - (a) Extract Landsat 8 and NOAH-GLDAS images for the required geometry of which we require the downscaled ET product, and filter for the dates on which we want to produce ET.
 - (b) Compute the indices as in table 1
 - (c) Resample NOAH-GLDAS images to 30 meters and reproject it to match Landsat 8 CRS. Add the meteorological variables as in section 3.3 to corresponding pixels of input images.
 - (d) Load the trained model from uploaded asset and use this to inference ET on 30 meter resolution inputs.
 - (e) Visualise 30 meter ET product on GEE map for the specified geometry.

6 RESULTS

In the present study, we show the results of downscaling ET for some major Agro-Eco zones of India which are shown in 3 - AEZ 14 (majorly comprised of Madhya Pradesh (Latitudes 21.6° N to 26.3° N, Longitudes 74.9° E to 82.48° E)), AEZ 15 (majorly comprised of Jharkhand (Latitudes 21.57° N to 25.14° N, Longitudes 83.20°E to 87.58°E), Chhattisgarh (Latitudes 17.55° N to 24.33° N, Longitudes 79.5°E to 84.91°E)), AEZ 17 (Uttar Pradesh (Latitudes 23.52° N to 31.28° N, Longitudes 77.3°E to 84.39°E), Bihar (Latitudes 24.20° N to 27.31° N, Longitudes 83.19°E to 88.17°E)), AEZ 16 (Odisha (Latitudes 17.31° N to 22.31° N, Longitudes 81.31°E to 87.29°E)) and AEZ 1 (Rajasthan (Latitudes 23.30° N to 30.12° N, Longitudes 69.30°E to 78.17°E)). The average rainfall [35] and percentage of Forest cover [15] in these states are also shown in table 3. Our experiments were based on these states to include the diversity in cropping areas, land cover types and rainfall, as these states embody a broad spectrum of climate, land-use variations. We finally show the usefulness of combining multiple AEZs together to form a cluster.

6.1 Single-AEZ Models

We sample points as explained in section 4.2. The results of the individual AEZ-wise random forest models are shown in figures 6 and the RMSE, NRMSE and R2 scores mentioned in table 4

Based on the graphical analysis of plots in 6, it is evident that the random forest model, optimized by hyperparameter tuning, shows favorable performance in the case of points sampled from the zones-1 and 14. However, it performs relatively poorer for values greater than 50 discussed in detail in section 7.2. Root Mean Square Error

(RMSE) and Validation R2 scores, in the instances of points sampled from 15 and notably in the case of points of 16.

We further experimented on this, and plotted the error plots for the random forest model, which was trained on data from the AEZ 16 defined in section 6 in the study area for the year 2016-21, and validation was done on data from AEZ 16 for the year 2022. We plotted the absolute value of error, i.e. $|predicted\ ET - actual\ ET|$ versus the actual ET value, to see the variation of error with the true ET, in figure 4 which shows the Heteroskedasticity in the distribution of error for higher values of ET. To solve this problem, we tried to combine the neighbouring AEZs dataset together in order to decrease the fluctuations in residuals by increasing the sample points with ET values > 50 .

6.2 Multi AEZ Models

Building separate models for each AEZ to predict ET for the whole India can be computationally expensive which may lead to challenges in model management at bigger scales. Combining neighbouring AEZ together can not only decrease the issue of managing different models but can also improve the accuracies by diversifying the dataset. This strategy aims to capture regional patterns while maximizing the utility of available training data.

The experiment involves pooling training data from various neighboring AEZs to construct a comprehensive random forest model. The training dataset now includes data points collected from Bihar, Madhya Pradesh, Jharkhand, Chhattisgarh, and Odisha (thus combining multiple zones) for the years 2016-2021, all situated in close proximity. This pooled training data was used to train the random forest model and the model was validated for points sampled from individual zones in the year 2022. The results on validation data for individual AEZs are mentioned in table 4, and the scatter plots for validation datasets shown in figures 6.

This can prove advantageous, especially in instances where there is a scarcity of training data for certain small zones. As observed from the table 4 and scatter plots in 6, the performance of the multi-AEZ model consisting of points from the neighboring states for the year 2016-21 shows good validation RMSE values and R2 scores when validated on individual states. The accuracies of the model were approximately in line with the performance of individual AEZ wise models while decreasing the number of models to be trained.

6.3 Leave-One-Out Validation

Since the multi-AEZ models gave similar performance to the single-AEZ models but need fewer models to be trained, we proceed with training and evaluating our models Leave-One-Out Validation strategy to rigorously assess the performance of our model across a 6-year temporal span, from the year 2017 to 2022. This validation technique involves systematically withholding one year's data as the validation set while utilizing the remaining years for training the random forest model. By iteratively repeating this process for each individual year, we aim to evaluate the model's robustness and its ability to generalize across diverse temporal contexts. This approach is particularly crucial in our research, as it allows us to capture and understand the model's performance aspects, specific to each year. The temporal variability inherent in the dataset involving remotely-sensed data and climatic variables, spanning six

State	Average Annual rainfall (in mm)	Forest cover (in %)	AEZ
Rajasthan	614.4	4.87	1
Madhya Pradesh	1086.25	25.14	14
Jharkhand	1444.8	29.76	15
Chhattisgarh	1309.6	41.21	15
Uttar Pradesh	914.4	6.15	17
Bihar	1512.7	7.84	17
Odisha	1420.8	33.50	16

Table 3: Average annual rainfall (in 2021) and percentage forest cover (2021) of the chosen states in India

Type of Model	Validation AEZ	RMSE	NRMSE	R2
Single-AEZ	AEZ 1	3.3	0.44	0.86
Single-AEZ	AEZ 14	4.65	0.39	0.87
Single-AEZ	AEZ 16	9.45	0.48	0.70
Single-AEZ	AEZ 15	6.39	0.39	0.82
Single-AEZ	AEZ 17	7.89	0.43	0.65
Multi-AEZ	AEZ 1	3.5	0.40	0.88
Multi-AEZ	AEZ 14	4.62	0.39	0.87
Multi-AEZ	AEZ 16	9.07	0.46	0.70
Multi-AEZ	AEZ 15	6.41	0.41	0.83
Multi-AEZ	AEZ 17	7.01	0.43	0.67

Table 4: RMSE and R2 scores for validation data of the individual AEZ-wise random forest models and Multi-AEZ models trained on 2016-2021 and validated for 2022

AEZ	Validation Year	RMSE	R2
1	2022	3.3	0.86
1	2021	4.21	0.77
1	2020	4.04	0.85
1	2019	3.75	0.84
1	2018	3.44	0.87
1	2017	3.3	0.78
14	2022	4.57	0.87
14	2021	6.72	0.77
14	2020	5.5	0.85
14	2019	5.36	0.82
14	2018	4.52	0.87
14	2017	4.80	0.89

Table 5: RMSE and R2 scores for the Leave-one-out validation experiment performed for data sampled from AEZ 1 and 14, validated on particular years and trained on different years.

consecutive years, necessitates a validation strategy that explicitly considers the influence of each temporal segment on the model's predictive capabilities. This experiment was performed for two zones : AEZ 1 and AEZ 14, and we present the findings obtained from each iteration in Table 5 in the form of RMSE and R2 scores given by the model, and the Validation-Scatter plots are shown in Figures 7 and 8.

For AEZ 1, we obtained an average RMSE of 3.67 and mean R2 score of 0.823, and for AEZ 14 we obtained an average RMSE of 5.245 and mean R2 score of 0.845 which are approximately inline with the validation R2 score for 2022.

6.4 Visualization

As a demonstration of our method, we show the visualisations of the downscaled ET product at 30 meter resolution with the MODIS ET product at 500 meter resolution and the satellite imagery provided by GEE of that area. We can observe the correlation between land cover type and down-scaled 30 m ET in figures 9, 10 which are at different resolutions. In general, the 30 meter output is higher at Forest pixels and lower for built-up regions as expected. The 30 meter outputs clearly distinguishes between the distinctive features of the land cover like forests, crops, roads etc., which aligns with the satellite imagery in Figure 9, where as MODIS ET product does not give us any clear information about the ET distribution within the area of 2 kilometer \times 2 kilometer. This comparison implies that the 30-meter outputs offer advantages in terms of precision and detailed information within smaller regions where

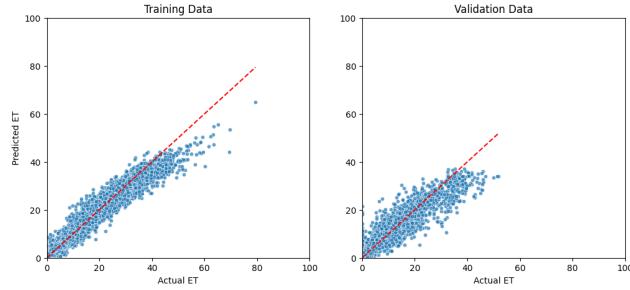
Evapotranspiration values are required for various hydro-ecological analysis, when compared to the MODIS ET product, at a finer spatial resolution.

The 30 meter resolution output of our Random Forest model for larger regions (area approx. 50 km \times 75 km) aligns well with the MODIS ET products at 500 meter resolution, as well as the land cover shown in satellite imagery, as shown in Figure 10. This further validates our Random Forest model, which has been trained on the ET values provided by MODIS dataset, as it shows the variations similar to MODIS ET variations in the regions (Higher MODIS ET values corresponds to greener pixels in the 30 meter output and vice-versa).

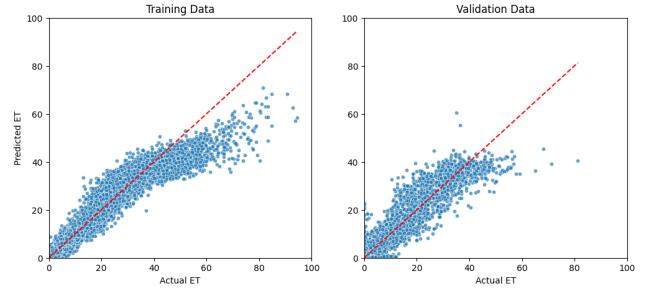
To validate the values of output, we also show time series plots for our Random Forest model's predicted ET for the year 2022, with the MODIS ET value over different small regions of 1 km \times 1 km, in the Figure 11. As observed from the time series plots, the alignment between the Downscaled ET time series and MODIS ET time series shows the accuracy of the random forest model. This suggests that the model is successful in capturing the trends, seasonality, and other patterns present in the Evapotranspiration data.

7 CONCLUSIONS AND FUTURE WORKS

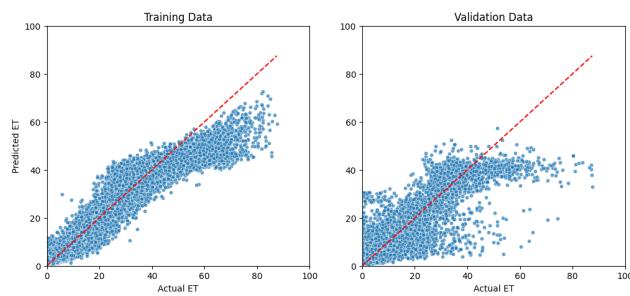
Our paper demonstrates the development and implementation of the Random Forest model for downscaling ET, intended for public



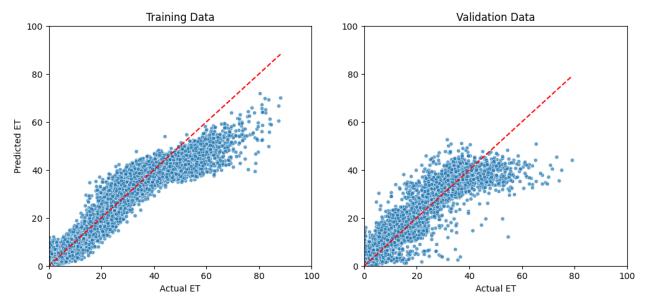
(a) Scatter plots for Train data (Years 2016–2021) and Validation data (Year 2022) for points sampled from AEZ 1. Random Forest regressor shows RMSE = 3.3, NRMSE = 0.44 (on validation data) and R2 score of 0.97 on train data and 0.89 on validation data



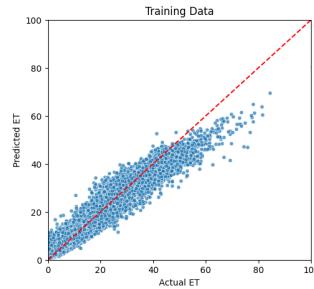
(b) Scatter plots for Train data (Years 2016–2021) and Validation data (Year 2022) for points sampled from AEZ 14. Random Forest regressor shows RMSE = 4.56, NRMSE = 0.39 (on validation data) and R2 score of 0.95 on train data and 0.87 on validation data



(c) Scatter plots for Train data (Years 2016–2021) and Validation data (Year 2022) for points sampled from AEZ 16. Random Forest regressor shows RMSE = 9.45, NRMSE = 0.48 (on validation data) and R2 score of 0.92 on train data and 0.66 on validation data



(d) Scatter plots for Train data (Years 2016–2021) and Validation data (Year 2022) for points sampled from AEZ 15. Random Forest regressor shows RMSE = 6.39, NRMSE = 0.39 (on validation data) and R2 score of 0.93 on train data and 0.82 on validation data



(e) Scatter plots for Train data (Years 2016–2021) and Validation data (Year 2022) for points sampled from AEZ 17. Random Forest regressor shows RMSE = 7.89, NRMSE = 4.43 (on validation data) and R2 score of 0.96 on train data and 0.65 on validation data

Figure 3: Scatter plots for training and validation data for the AEZ-wise random forest models

use on GEE, representing a significant advancement in providing accessible and accurate high-resolution ET estimates for hydro-ecological applications mentioned in 1. The public availability of model on GEE to produce downscaled ET would help in promoting sustainability. We have shown the efficacy of the Random Forest model in downscaling evapotranspiration using Landsat and meteorological indices. It highlights the validation of downscaled ET with different land variations and years, aligns closely with the target ET provided by MODIS and captures the spatial variability of ET within smaller regions. By providing a finer spatial resolution of

ET, it helps in a more detailed understanding of ET patterns within regions. We are attempting some improvements in our method for downscaling ET, described in 7.1 and 7.2.

7.1 Scopes for Interpolation

To solve the problem of coarser resolution of meteorological variables, we tried spatially interpolating them to the resolution of 500 m. Google earth engine provides only 2 options for spatial interpolations: Bicubic and Bilinear interpolation. We tried out both

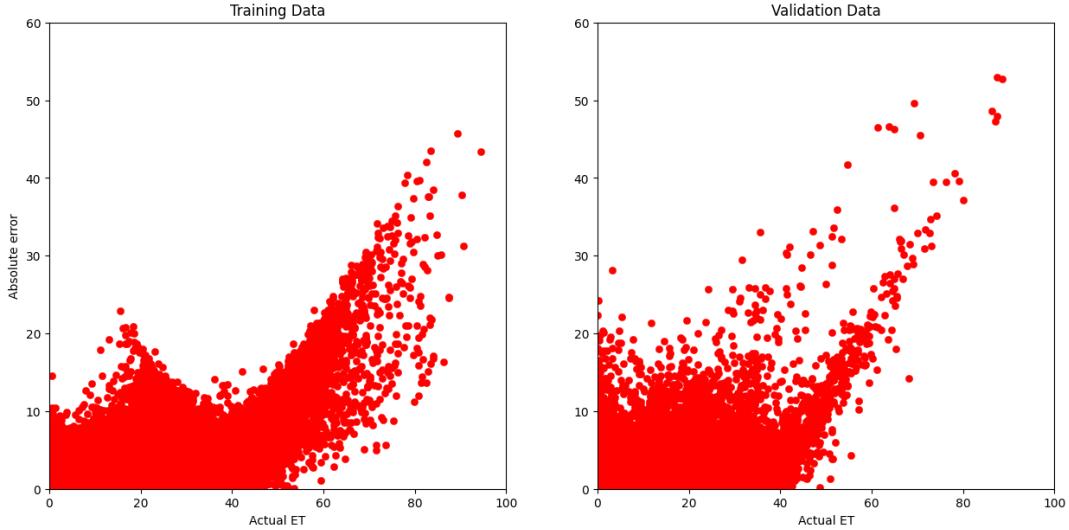


Figure 4: Error plot on Train data from the AEZ 16 (2016-21) and Validation data from the AEZ 16 (2022)

Feature	coef	std err	t	P> t
Constant	295.0797	116.671	2.529	0.011
MSAVI	104.333	131.357	0.794	0.427
NDMI	31.4114	29.832	1.053	0.292
NDVI	-5.44E+06	3.27E+06	-1.661	0.097
NDWI	246.8802	85.267	2.895	0.004
SAVI	3.63E+06	2.18E+06	1.661	0.097
NDBI	-17.1474	30.441	-0.563	0.573
NDIIB7	-77.252	65.953	-1.171	0.241
ALBEDO	0.0017	0.001	3.004	0.003
LST	-0.0322	0.058	-0.558	0.577
HUMID	1749.86	496.181	3.527	0
RAIN	-850.0978	1.72E+04	-0.05	0.96
TEMP	-3.546	1.349	-2.628	0.009
PSURF	-0.0007	0.001	-1.138	0.255
CANOPY	-13.238	9.327	-1.419	0.156
SM	1.1189	0.314	3.568	0
WIND	2.139	1.78	1.202	0.229
ROOT	0.0242	0.022	1.088	0.277
SOILTEMP	0.9162	1.394	0.657	0.511
RUNOFF	9.8863	3.655	2.705	0.007
SW	5.168	44.55	0.116	0.908
QLE	-2.0603	1.85	-1.113	0.266
QH	-1.8832	1.845	-1.021	0.307
QG	-1.5954	1.843	-0.866	0.387
SWNET	-6.2105	53.331	-0.116	0.907
LWNET	2.1169	1.836	1.153	0.249
TAIR	2.0056	1.06	1.891	0.059

(a) OLS regression performed on a subset of points uniformly sampled from training data

Feature	coef	std err	t	P> t
Constant	496.9591	95.011	5.231	0
MSAVI	-194.5039	84.988	-2.289	0.022
NDMI	-20.3992	21.337	-0.956	0.339
NDVI	-5.53E+06	2.07E+06	-2.673	0.008
NDWI	23.4905	66.533	0.353	0.724
SAVI	3.69E+06	1.38E+06	2.673	0.008
NDBI	46.0165	21.622	2.128	0.033
NDIIB7	84.0353	46.929	1.791	0.073
ALBEDO	-0.0006	0	-1.57	0.117
LST	0.0127	0.037	0.346	0.729
HUMID	113.7324	407.537	0.279	0.78
RAIN	-1.97E+04	1.15E+04	-1.715	0.086
TEMP	0.7798	1.079	0.723	0.47
PSURF	0.0006	0	1.439	0.15
CANOPY	11.5222	6.064	1.9	0.057
SM	-0.0098	0.236	-0.041	0.967
WIND	0.5531	1.411	0.392	0.695
ROOT	-0.0153	0.015	-1.021	0.307
SOILTEMP	-0.4629	1.13	-0.41	0.682
RUNOFF	-0.0294	1.818	-0.016	0.987
SW	79.6108	37.662	2.114	0.035
QLE	1.5505	1.557	0.996	0.319
QH	1.5461	1.554	0.995	0.32
QG	1.2561	1.554	0.809	0.419
SWNET	-97.9584	44.952	-2.179	0.029
LWNET	-1.3773	1.549	-0.889	0.374
TAIR	-0.5897	0.898	-0.657	0.512

(b) OLS regression performed on the subset of points of training data for which actual ET value is above or equal to 50

Figure 5: Coefficients of different input features in Ordinary Least Square regression performed on training data

methods and interpolated the meteorological variables to the resolution of 500 m, and trained the random forest model with the interpolated meteorological variables and Landsat variables, but the model did not show any improvement over the model that was

trained on non-interpolated methods. Experimenting with alternative interpolation methods in Google Earth Engine posed another significant challenge, due to their non-availability in GEE. We attempt to explore the interpolation of meteorological variables on Google Earth Engine (GEE) as a means to improve the quality of

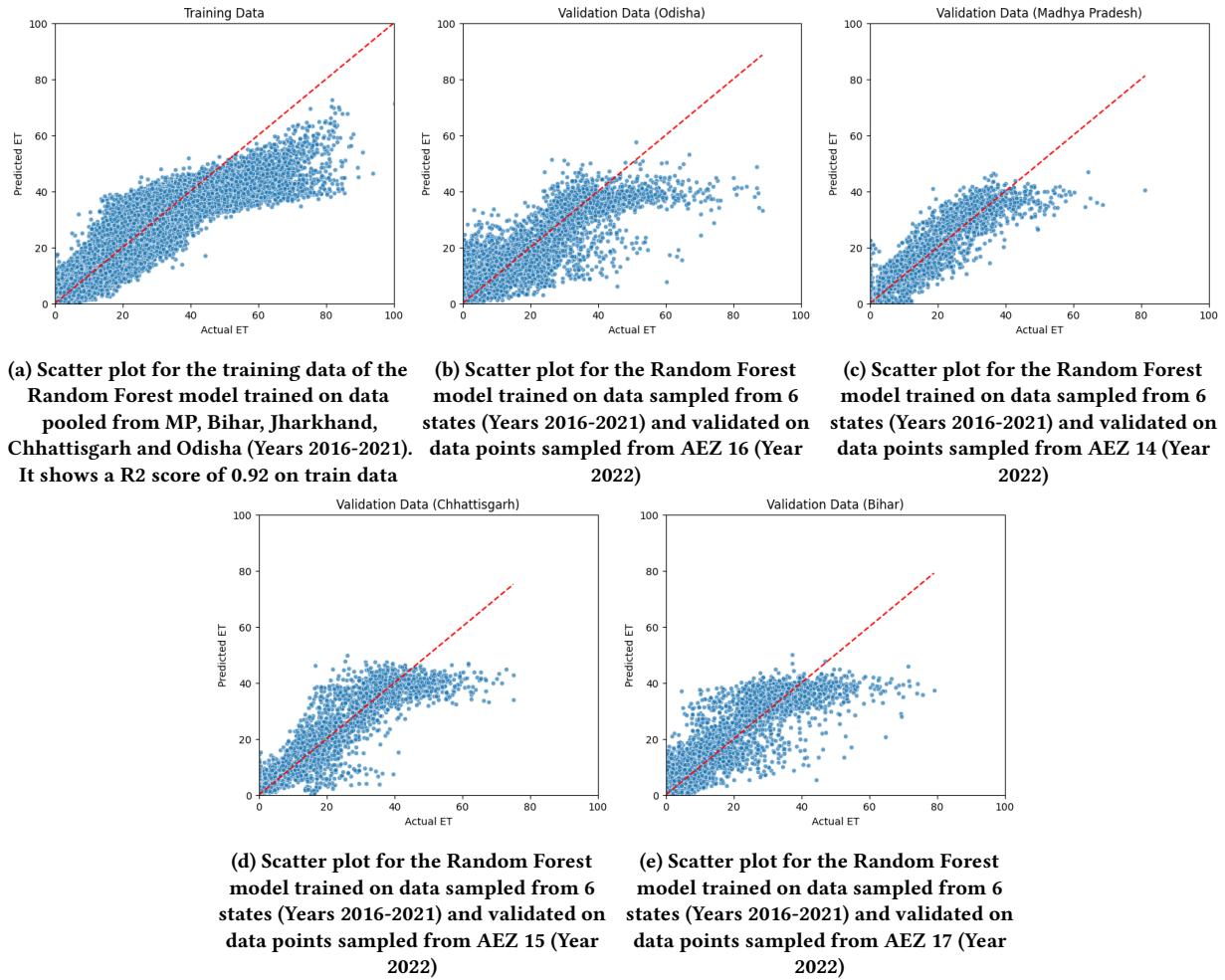


Figure 6: Scatter plots of train data and validation data for the Multi-state model

input data for downscaling Evapotranspiration in future. The interpolated climatic data generated through GEE can then serve as an enhanced input for our downscaling model, potentially enhancing the accuracy and precision of ET predictions, even for higher values, resulting in advancement of our downscaling methodology.

7.2 Low accuracy for higher ET values

The figures 3d and table 4 shows poor RMSE and Validation R2 scores, significantly in the case of points of AEZ 16. This increased errors in predictions accounts to the larger percentage of points sampled in AEZ 16 that have higher ET values. Similar trend is observed in the graph in figure 4, which illustrates that the error is elevated and continues to rise for ET values exceeding approximately 50. This indicates that the random forest model fails to capture the relationship between the input variables and ET for higher values, approximately greater than 50, and this explains the lower validation R2 score and lower RMSE for the model's performance on points sampled from AEZ 16. This might be explained by the observation in OLS coefficients (fig 5), which indicates that

certain meteorological variables exert a more significant impact on predicting evapotranspiration (ET) when ET values are elevated, specifically exceeding 50. Given that meteorological variables are observed at a resolution of approximately 27830 meters, which is notably coarser compared to 500 meters, the random forest model may not be able to accurately capture the relationship between meteorological variables and evapotranspiration (ET). So the higher errors for higher values of ET, in which the meteorological variables have a greater impact. This can explain the reason why the training and validation scatter plots showed greater deviation from 45° line in figure 3c, 3d (points sampled from AEZ 16, AEZ 17 and AEZ 15). We have also observed that often these higher ET values correspond to forest regions. Hence we are also attempting to analyse the model performance on the basis of LULC.

REFERENCES

- [1] Richard Allan, L. Pereira, and Martin Smith. *Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56*. Vol. 56. Jan. 1998.
- [2] H. Beaudoin and M. Rodell. “NASA/GSFC/HSL, GLDAS Noah Land Surface Model L4 3 hourly 0.25 x 0.25 degree V2.1”. In: *Greenbelt, Maryland, USA*,

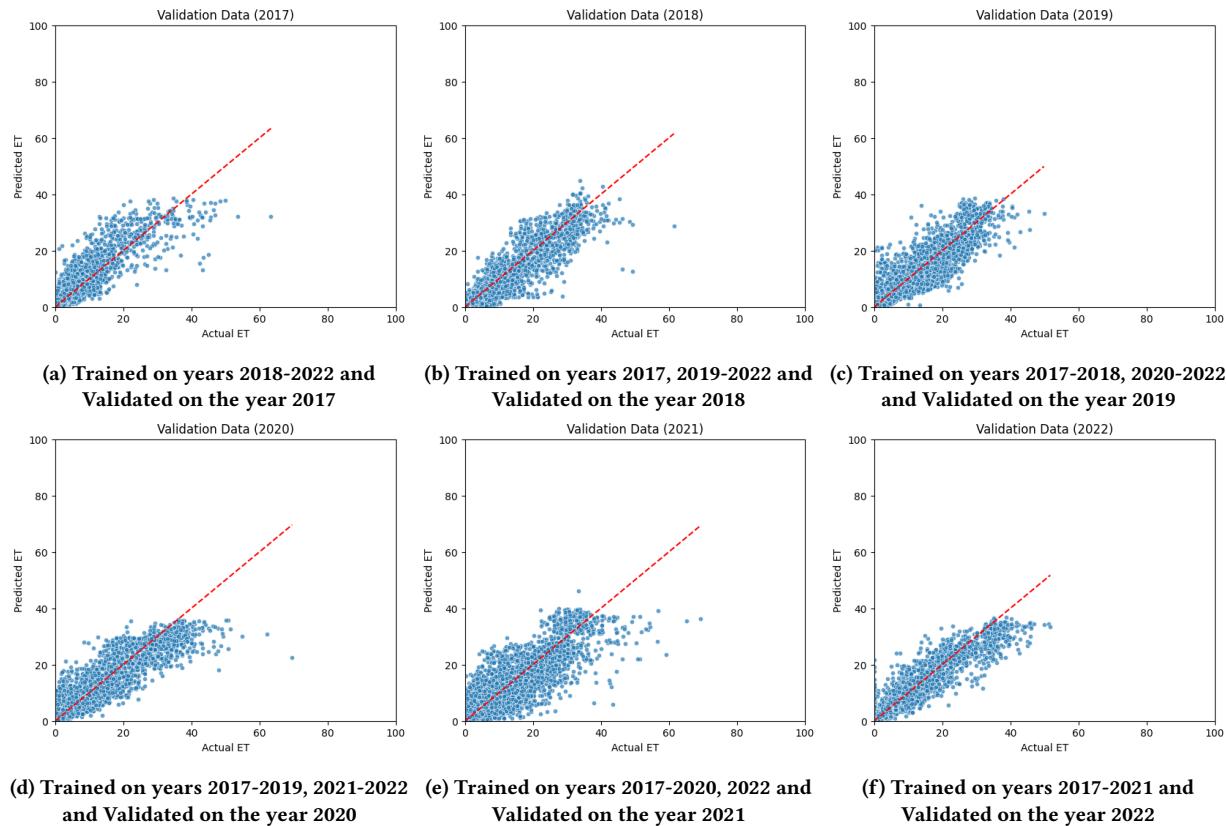


Figure 7: Scatter plots for predicted ET value vs actual ET value for Leave-one-out validation performed for training and validation data both sampled from AEZ 1

- Goddard Earth Sciences Data and Information Services Center (GES DISC) (2020). doi: 10.5067/E7TYRXPJKWOO.
- [3] Ling Bian and Rachel Butler. “Comparing effects of aggregation methods on statistical and spatial properties of simulated spatial data”. In: *Photogrammetric engineering and remote sensing* 65 (1999), pp. 73–84.
- [4] José Da Silva Júnior et al. “Random forest techniques for spatial interpolation of evapotranspiration data from Brazilian’s Northeast”. In: *Computers and Electronics in Agriculture* 166 (Nov. 2019), p. 105017. doi: 10.1016/j.compag.2019.105017.
- [5] T. Davie and NW. Quinn. *Fundamentals of Hydrology*. Routledge fundamentals of physical geography series. Routledge, 2019. isbn: 9780415858700. url: <https://books.google.co.in/books?id=D4b3oAEACAAJ>.
- [6] Arvind Dhaloya et al. “Modeling medium resolution evapotranspiration using downscaling techniques in north-western part of India”. In: *MAUSAM* 74.3 (Jan. 2024), pp. 561–578. doi: 10.54302/mausam.v74i3.5112. url: <https://mausamjournal.indm.gov.in/index.php/MAUSAM/article/view/5112>.
- [7] Yihun T. Dile et al. “Evaluating satellite-based evapotranspiration estimates for hydrological applications in data-scarce regions: A case in Ethiopia”. In: *Science of The Total Environment* 743 (2020), p. 140702. issn: 0048-9697. doi: <https://doi.org/10.1016/j.scitotenv.2020.140702>. url: <https://www.sciencedirect.com/science/article/pii/S0048969720342248>.
- [8] P. Droogers, W.W. Immerzeel, and I.J. Lorite. “Estimating actual irrigation application by remotely sensed evapotranspiration observations”. In: *Agricultural Water Management* 97.9 (2010), pp. 1351–1359. issn: 0378-3774. doi: <https://doi.org/10.1016/j.agwat.2010.03.017>. url: <https://www.sciencedirect.com/science/article/pii/S0378377410001149>.
- [9] I. Aguado E. Chuvieco D. Riaño and D. Cocero. “Estimation of fuel moisture content from multitemporal analysis of Landsat Thematic Mapper reflectance data: Applications in fire danger assessment”. In: *International Journal of Remote Sensing* 23.11 (2002), pp. 2145–2162. doi: 10.1080/01431160110069818. eprint: <https://doi.org/10.1080/01431160110069818>. url: <https://doi.org/10.1080/01431160110069818>.
- [10] KS Gajbhiye and C Mandal. “Agro-ecological zones, their soil resource and cropping systems”. In: *Status of Farm Mechanization in India, cropping systems, status of farm mechanization in India* (2000), pp. 1–32.
- [11] Ikhlas Ghazi, Hamish R. Mackey, and Tareq Al-Ansari. “A Review of Evapotranspiration Measurement Models, Techniques and Methods for Open and Closed Agricultural Field Applications”. In: *Water* 13.18 (2021), issn: 2073-4441. doi: 10.3390/w13182523. url: <https://www.mdpi.com/2073-4441/13/18/2523>.
- [12] M.R. Goyal and E.W. Harmse. *Evapotranspiration: Principles and Applications for Water Management*. Apple Academic Press, 2013. isbn: 9781926895581. url: <https://books.google.co.in/books?id=iVHGAAAAQBAJ>.
- [13] Sanayabi Hodani et al. “Spatial Interpolation of Reference Evapotranspiration in India: Comparison of IDW and Kriging Methods”. In: *Journal of The Institution of Engineers (India): Series A* 98 (Nov 2017), pp. 511–524. doi: 10.1007/s40030-017-0241-z.
- [14] Taehye Hwang et al. “Downscaling real-time vegetation dynamics by fusing multi-temporal MODIS and Landsat NDVI in topographically complex terrain”. In: *Remote Sensing of Environment* 115.10 (2011), pp. 2499–2512. issn: 0034-4257. doi: <https://doi.org/10.1016/j.rse.2011.05.010>. url: <https://www.sciencedirect.com/science/article/pii/S0034425711001829>.
- [15] India State of Forest Report (2021). Tech. rep. Forest Survey of India, Dehradun, 2021.
- [16] Yasir Kaheil et al. “Downscaling and Forecasting of Evapotranspiration Using a Synthetic Model of Wavelets and Support Vector Machines”. In: *Geoscience and Remote Sensing, IEEE Transactions on* 46 (Oct. 2008), pp. 2692–2707. doi: 10.1109/TGRS.2008.919819.
- [17] Yinghai Ke et al. “Downscaling of MODIS One Kilometer Evapotranspiration Using Landsat-8 Data and Machine Learning Approaches”. In: *Remote Sensing* 8.3 (2016). issn: 2072-4292. doi: 10.3390/rs8030215. url: <https://www.mdpi.com/2072-4292/8/3/215>.
- [18] Yinghai Ke et al. “Downscaling of MODIS One Kilometer Evapotranspiration Using Landsat-8 Data and Machine Learning Approaches”. In: *Remote Sensing*

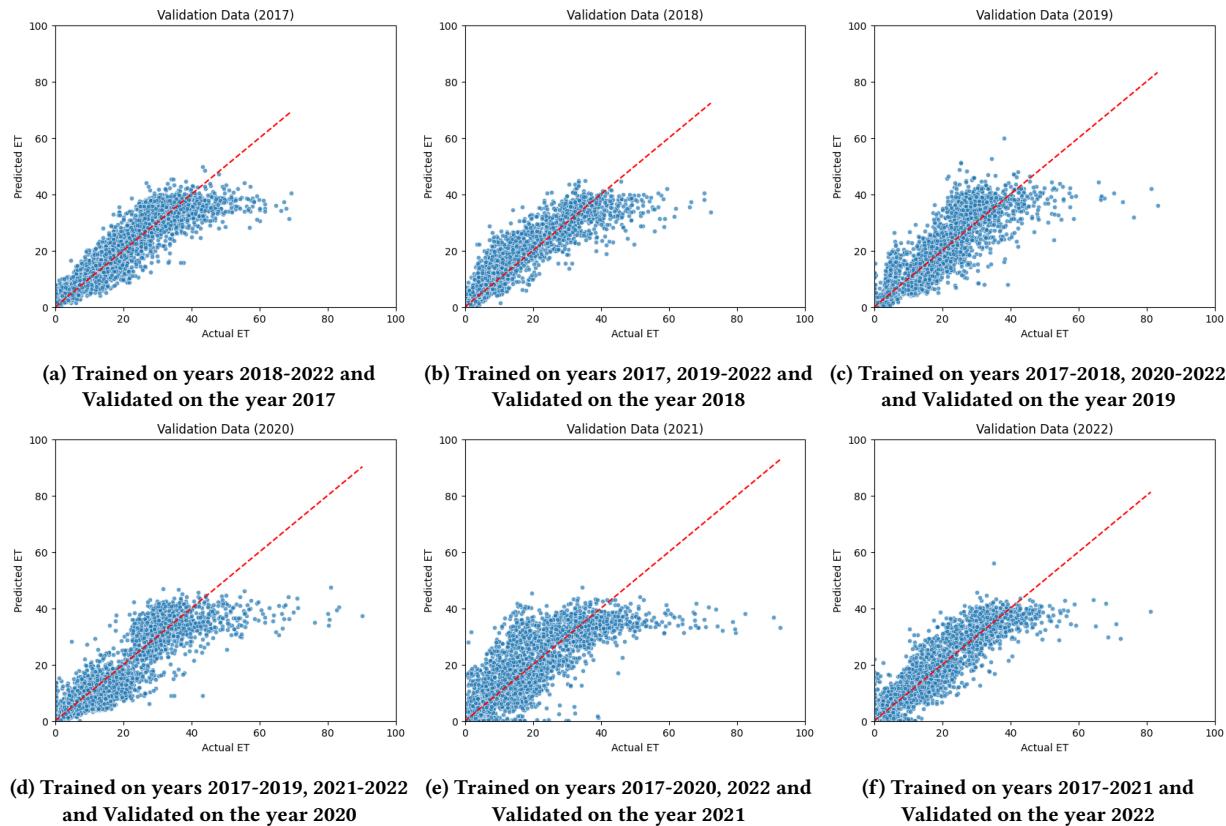


Figure 8: Scatter plots for predicted ET value vs actual ET value for Leave-one-out validation performed for training and validation data both sampled from AEZ 14.

- 8.3 (2016). ISSN: 2072-4292. doi: 10.3390/rs8030215. url: <https://www.mdpi.com/2072-4292/8/3/215>.
- [19] D. Kool et al. “A review of approaches for evapotranspiration partitioning”. In: *Agricultural and Forest Meteorology* 184 (2014), pp. 56–70. ISSN: 0168-1923. doi: <https://doi.org/10.1016/j.agrformet.2013.09.003>. url: <https://www.sciencedirect.com/science/article/pii/S016819231300230X>.
- [20] Jin Li and Andrew D. Heap. “Spatial interpolation methods applied in the environmental sciences: A review”. In: *Environmental Modelling Software* 53 (2014), pp. 173–189. ISSN: 1364-8152. doi: <https://doi.org/10.1016/j.envsoft.2013.12.008>. url: <https://www.sciencedirect.com/science/article/pii/S1364815213003113>.
- [21] Muhammad Usman Liaqat et al. “Evaluation of MODIS and Landsat multiband vegetation indices used for wheat yield estimation in irrigated Indus Basin”. In: *Computers and Electronics in Agriculture* 138 (2017), pp. 39–47. ISSN: 0168-1699. doi: <https://doi.org/10.1016/j.compag.2017.04.006>. url: <https://www.sciencedirect.com/science/article/pii/S0168169916310456>.
- [22] Steven P. Loheide and Steven M. Gorelick. “A local-scale, high-resolution evapotranspiration mapping algorithm (ETMA) with hydroecological applications at riparian meadow restoration sites”. In: *Remote Sensing of Environment* 98.2 (2005), pp. 182–200. ISSN: 0034-4257. doi: <https://doi.org/10.1016/j.rse.2005.07.003>. url: <https://www.sciencedirect.com/science/article/pii/S0034425705002191>.
- [23] Thomas R. Loveland and James R. Irons. “Landsat 8: The plans, the reality, and the legacy”. In: *Remote Sensing of Environment* 185 (2016). Landsat 8 Science Results, pp. 1–6. ISSN: 0034-4257. doi: <https://doi.org/10.1016/j.rse.2016.07.033>. url: <https://www.sciencedirect.com/science/article/pii/S0034425716302905>.
- [24] Xiaoliang Lu and Qianlai Zhuang. “Evaluating evapotranspiration and water-use efficiency of terrestrial ecosystems in the conterminous United States using MODIS and AmeriFlux data”. In: *Remote Sensing of Environment* 114.9 (2010), pp. 1924–1939. ISSN: 0034-4257. doi: <https://doi.org/10.1016/j.rse.2010.04.001>. url: <https://www.sciencedirect.com/science/article/pii/S0034425710001124>.
- [25] Qiaozhen Mu, Maosheng Zhao, and Steven W Running. “MODIS global terrestrial evapotranspiration (ET) product (NASA MOD16A2/A3)”. In: *Algorithm Theoretical Basis Document, Collection 5* (2013), p. 600.
- [26] R.C. Pipunic, J.P. Walker, and A. Western. “Assimilation of remotely sensed data for improved latent and sensible heat flux prediction: A comparative synthetic study”. In: *Remote Sensing of Environment* 112.4 (2008). Remote Sensing Data Assimilation Special Issue, pp. 1295–1305. ISSN: 0034-4257. doi: <https://doi.org/10.1016/j.rse.2007.02.038>. url: <https://www.sciencedirect.com/science/article/pii/S0034425707003215>.
- [27] Nathaniel P. Robinson et al. “A Dynamic Landsat Derived Normalized Difference Vegetation Index (NDVI) Product for the Conterminous United States”. In: *Remote Sensing* 9.8 (2017). ISSN: 2072-4292. doi: 10.3390/rs9080863. url: <https://www.mdpi.com/2072-4292/9/8/863>.
- [28] Q. Mu Running S. and M. Zhao. “MODIS/Terra Net Evapotranspiration 8-Day L4 Global 500m SIN Grid V061”. In: *NASA EOSDIS Land Processes Distributed Active Archive Center* (2021). doi: 10.5067/MODIS/MOD16A2.061.
- [29] NK Shrestha and S Shukla. “Support vector machine based modeling of evapotranspiration using hydro-climatic variables in a sub-tropical environment”. In: *Agricultural and forest meteorology* 200 (2015), pp. 172–184.
- [30] RB Smith. “The heat budget of the earth’s surface deduced from space”. In: *Yale University Center for Earth Observation, New Haven, CT, USA* (2010).
- [31] Alexia Tsoumi et al. “Estimation of Actual Evapotranspiration by Remote Sensing: Application in Thessaly Plain, Greece”. In: *Sensors* 8.6 (2008), pp. 3586–3600. ISSN: 1424-8220. doi: 10.3390/s8063586. url: <https://www.mdpi.com/1424-8220/8/6/3586>.
- [32] Emily Wilson and Steven Sader. “Detection of forest harvest type using multiple dates of Landsat TM imagery”. In: *Remote Sensing of Environment* 80 (June 2002), pp. 385–396. doi: 10.1016/S0034-4257(01)00318-2.
- [33] Jiangzhou Xia et al. “Satellite-Based Analysis of Evapotranspiration and Water Balance in the Grassland Ecosystems of Dryland East Asia”. In: *PLOS ONE* 9.5 (May 2014), pp. 1–11. doi: 10.1371/journal.pone.0097295. url: <https://doi.org/10.1371/journal.pone.0097295>.

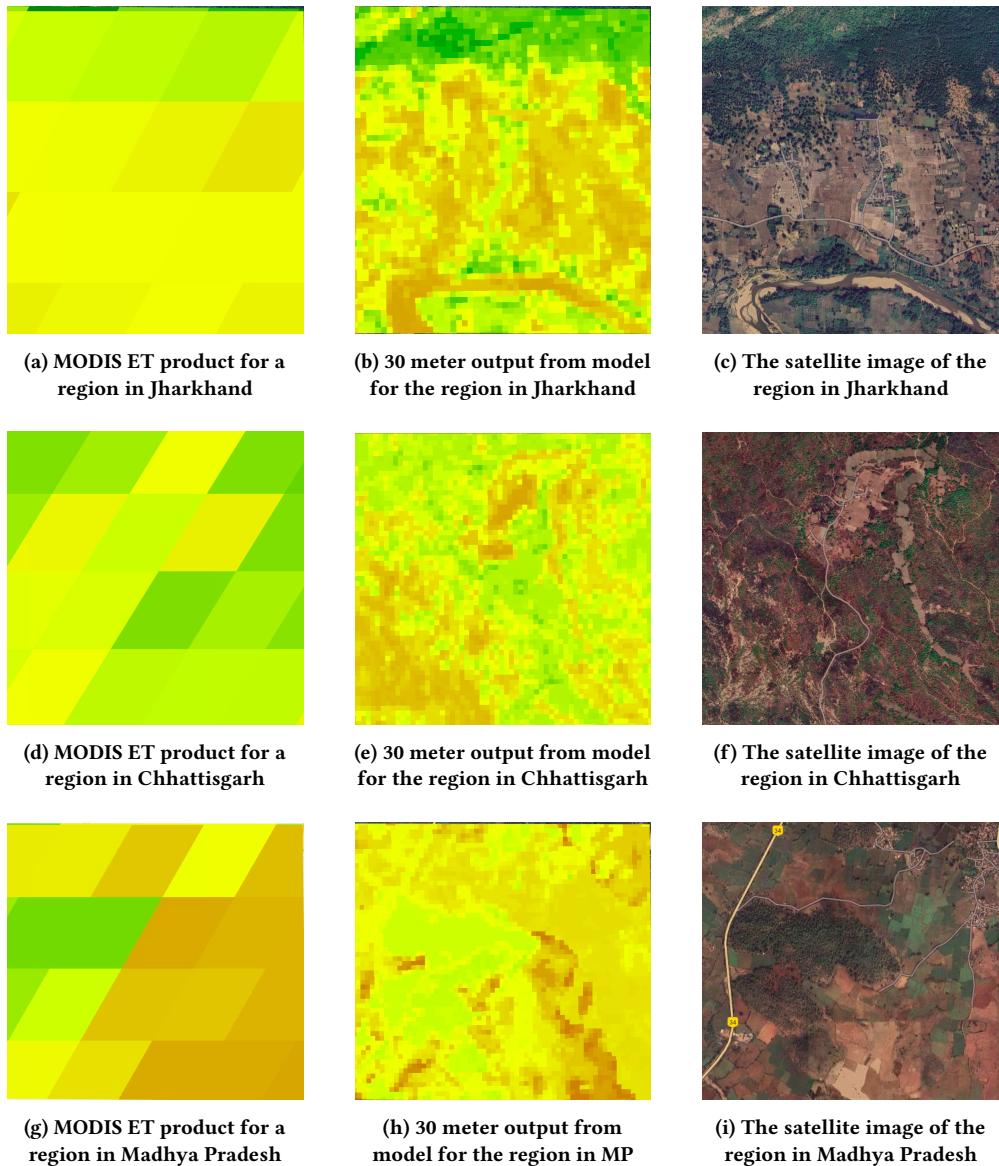
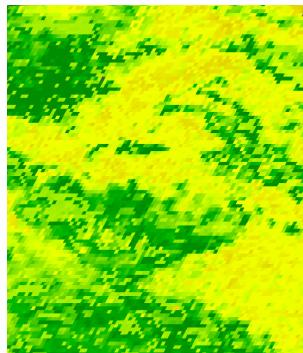
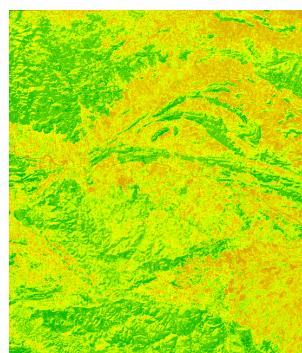


Figure 9: Visualisation of 30 meter resolution ET product predicted by the trained Random Forest Model for regions from Jharkhand, Chhattisgarh and Madhya Pradesh of area 2 km × 2 km against satellite images and MODIS ET products. Green pixels are maximum pixel values: 100 and Red pixels are minimum pixel values: 0

- [34] Hanqiu Xu. "Extraction of Urban Built-up Land Features from Landsat Imagery Using a Thematic-oriented Index Combination Technique". In: *Photogrammetric Engineering Remote Sensing* 73 (Dec. 2007), pp. 1381–1391. doi: 10.14358/PERS.73.12.1381.
- [35] B. P. Yadav et al. "Rainfall Statistics of India – 2021". In: *MoES/IMD/HS/Rainfall Report/02(2022)/60* (2021).
- [36] Yang Yong et al. "Estimating evapotranspiration by coupling Bayesian model averaging methods with machine learning algorithms". In: *Environmental Monitoring and Assessment* 193 (Mar. 2021). doi: 10.1007/s10661-021-08934-1.
- [37] I.A.M. Yunusa, R.R. Walker, and P. Lu. "Evapotranspiration components from energy balance, sapflow and microlysimetry techniques for an irrigated vineyard in inland Australia". In: *Agricultural and Forest Meteorology* 127.1 (2004), pp. 93–107. issn: 0168-1923. doi: <https://doi.org/10.1016/j.agrformet.2004.07.001>. url: <https://www.sciencedirect.com/science/article/pii/S0168192304001662>.



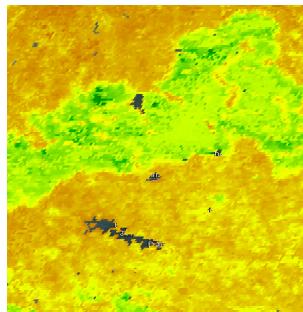
(a) MODIS ET product for a region in Jharkhand



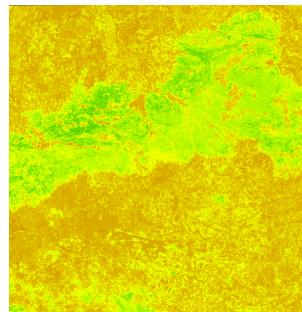
(b) 30 meter output from model for the region in Jharkhand



(c) The satellite image of the region in Jharkhand



(d) MODIS ET product for a region in Chhattisgarh

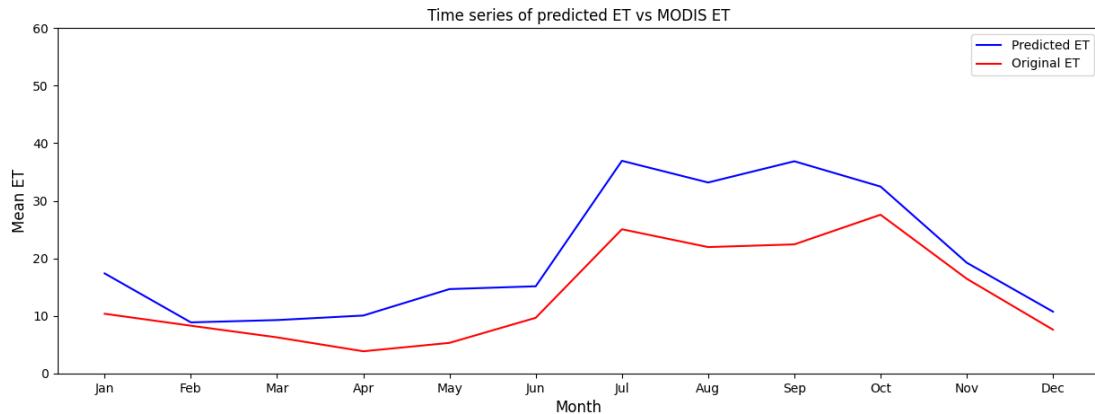


(e) 30 meter output from model for the region in Chhattisgarh

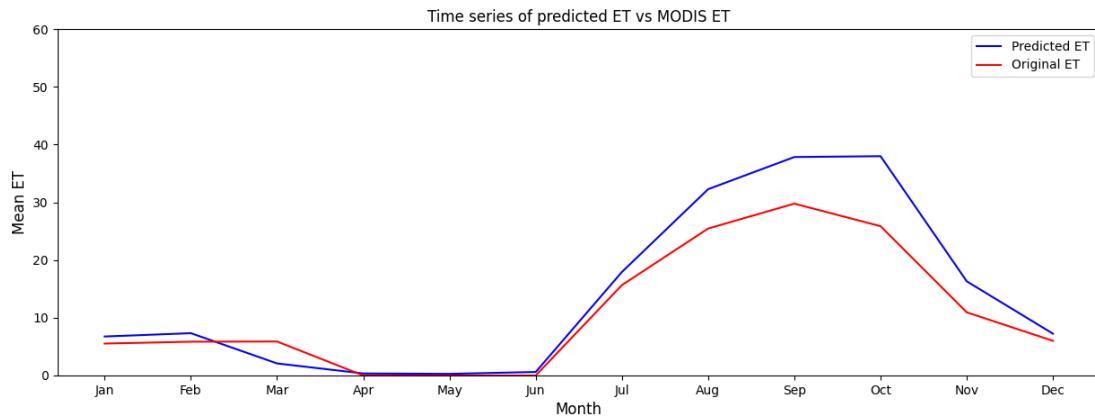


(f) The satellite image of the region in Chhattisgarh

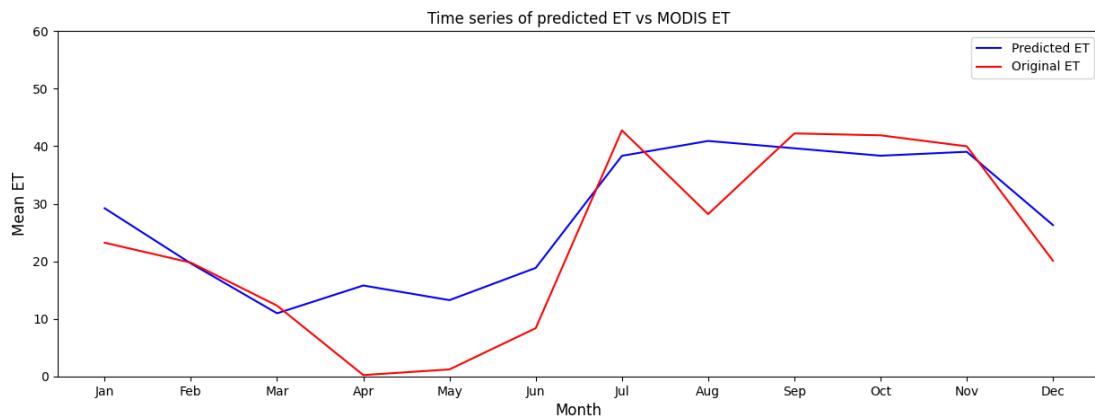
Figure 10: Visualisation of 30 meter resolution ET product predicted by the trained Random Forest Model for regions from Jharkhand and Chhattisgarh of area $50 \text{ km} \times 75 \text{ km}$ against satellite images and MODIS ET products. Green pixels are maximum pixel values: 100 and Red pixels are minimum pixel values: 0



(a) Time series graph of predicted ET and MODIS ET for a 1km × 1km region in Odisha for the year 2022



(b) Time series graph of predicted ET and MODIS ET for a 1km × 1km region consisting of buildup and crops for the year 2022



(c) Time series graph of predicted ET and MODIS-ET for a 1km × 1km region consisting of forests for the year 2022

Figure 11: Time series graphs of Evapotranspiration