https://doi.org/10.5194/egusphere-2026-4188
Preprint. Discussion started: 4 August 2026

A seamless workflow for wildfire modelers to validate dynamic global vegetation model outputs against remote sensing data

Bikem Ekberzade[1]

[1]Istanbul Technical University, Eurasia Institute of Earth Sciences, Istanbul, Turkey

Correspondence to: Bikem Ekberzade (ekberzade19@itu.edu.tr)

ORCID: 0000-0002-3271-2040

Abstract. Validation of wildfire simulations remains a persistent challenge in dynamic vegetation and Earth system modeling due to structural inconsistencies between model outputs and satellite-derived observation datasets. Researchers often rely on ad hoc, non-reproducible preprocessing workflows to bridge the gap between different projections and resolutions. This study presents a structured, open-source R-based workflow for processing dynamic vegetation model outputs and systematically comparing them with remote sensing burned area products. The workflow provides an end-to-end solution that (i) aggregates burned area from remote sensing data, (ii) harmonizes sinusoidal HDF files into WGS84-based model formats, and (iii) performs regional-scale comparisons across user-defined spatial units. While demonstrated using LPJ-GUESS and MODIS MCD64A1, the workflow employs a model- and dataset-agnostic "Master Grid" approach. This creates a static template to strictly harmonize observational data, thereby eliminating spatial mismatch errors while ensuring reproducibility.

Keywords: burned area analysis, dynamic vegetation model, LPJ-GUESS, MCD64A1, wildfire modelling

Plain language summary: Burned area prediction relies on complex computer models, but checking these models against actual satellite observations often proves to be challenging due to different grid systems. Therefore, scientists usually end up writing custom and hard-to-replicate code to force the data to align. This manuscript details an open-source tool that automatically and accurately aligns model predictions with satellite data. By creating a standardized “Master Grid”, this new workflow eliminates hidden mapping errors and ensures that wildfire researchers can easily and consistently validate their model outputs against real-world fire observations.

https://doi.org/10.5194/egusphere-2026-4188
Preprint. Discussion started: 4 August 2026

1. Introduction

Accurately representing changing wildfire regimes under anthropogenic climate change remains one of the major sources of uncertainty in Earth system modelling. In many fire-prone regions, recent decades have been characterized by increasing fire weather severity, longer fire seasons, greater atmospheric aridity, and more frequent extreme wildfire events, although trends in burned area vary considerably among ecosystems and regions (Abatzoglu et al. 2016; Jones et al. 2022). As climate change continues to alter the spatial and temporal characteristics of fire regimes and expand the climatic envelope conducive to fire, the ability of Dynamic Global Vegetation Models (DGVMs) and Earth System Models (ESMs) to accurately represent these changing dynamics is becoming paramount (Bowman et al., 2020; Harrison et al., 2021). These models are increasingly relied upon by researchers to project future carbon cycle feedbacks and vegetation dynamics; however, their utility is contingent upon robust, systematic evaluation against the historical observational record (Hantson et al. 2016; Lasslop et al. 2020; Ekberzade, 2026a).

Despite the availability of high-quality, long-term satellite records – most notably the MODIS MCD64A1 burned area product – systematic comparison between model simulations and observations remains technically demanding. Differences in spatial resolution, grid structure, coordinate reference systems (e.g., Sinusoidal vs. WGS84; Fig. 1), temporal aggregation and reporting conventions often prevent direct comparison of burned area estimates. These challenges are particularly acute when working with disturbance data, which are highly heterogeneous in space and time and sensitive to "edge effects" during regridding.

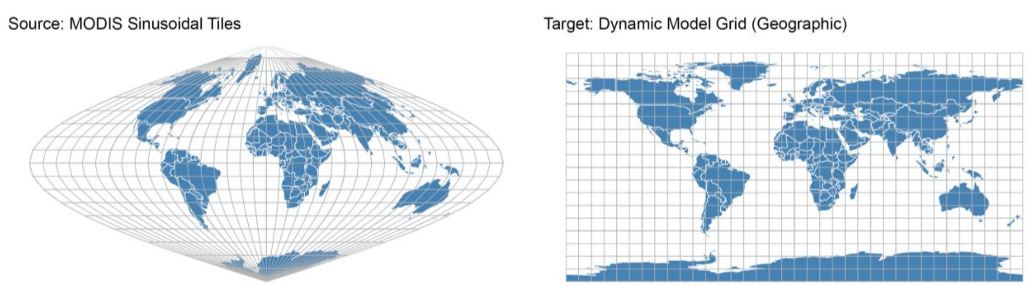

Figure 1. Visual representation of the reprojection challenge. The native sinusoidal projection of MODIS products (Source), vs the equirectangular grid required by DGVMs (Target).

https://doi.org/10.5194/egusphere-2026-4188
Preprint. Discussion started: 4 August 2026

Currently, model assessment efforts frequently rely on ad hoc, case-specific scripts that are difficult to reproduce, extend, or transfer across studies. While major inter-comparison projects (e.g., FireMIP; Hantson et al., 2016) document their processing protocols in detail, the underlying code is rarely published as a reusable tool, forcing subsequent studies to reconstruct these pipelines from scratch. Furthermore, while comprehensive benchmarking systems like ILAMB (Collier et al., 2018) exist, they are often computationally intensive, Python-centric, and optimized for global CMIP-style outputs. This leaves a significant gap for, e.g. the ecological modeling community which predominantly utilizes R (Lai et al. 2019), for a lightweight, transparent, and easily adaptable workflow designed to bridge the technical gap between the complex sinusoidal grids of remote sensing products and custom regional model grids.

This paper addresses these challenges by presenting a transparent and reproducible R-based workflow for the assessment of simulated burned area against satellite-based observations. The workflow is demonstrated using outputs from the dynamic vegetation model LPJ-GUESS-SIMFIRE-BLAZE and the MODIS MCD64A1 burned area product but is designed to be model- and dataset-agnostic. A series of scripts are provided to harmonize spatial and temporal representations and facilitate regional-scale comparisons using user-defined spatial units such as ecological regions or hydrobasins.

The focus of this contribution is methodological rather than evaluative. The example analyses are intended to illustrate the application of the workflow rather than to provide a comprehensive assessment of a particular model configuration or fire regime. The presented code series is organized into three methodological components: (1) *Automated Data Acquisition:* Definition of the study domain via polygon inputs (shapefile or coordinates) and automated identification and download of corresponding MODIS tiles, (2) *Processing and Derivation:* Derivation of pixel-level burned area from the MCD64A1 HDF files, including quality filtering and reprojection from the native sinusoidal grid, into a standard WGS84 coordinate listing, directly comparable to model outputs, (3) *Harmonization and Comparative Analysis:* Construction of a "Master Grid" based on model geometry, aggregation of observational data to this grid, and execution of regional statistical comparisons.

Together, this workflow provides a transparent and reproducible foundation for regional to global validation of wildfire simulations, with particular relevance for researchers working at the interface of dynamic vegetation modeling and Earth observation data in an era of rapidly changing fire regimes.

2. Model and observational data

2.1 LPJ-GUESS dynamic vegetation model

LPJ-GUESS is a process-based dynamic regional-to-global vegetation model (DGVM) that simulates vegetation structure and function through representations of growth, establishment, and mortality processes. The model places particular emphasis on resource competition and disturbance responses (Smith, 2007; Smith et al. 2001) and is widely used to study vegetation dynamics, biogeochemistry, and disturbance–climate interactions. LPJ-GUESS requires gridded climate input data—including temperature, precipitation, solar radiation, wind speed, and relative humidity—along with atmospheric CO_2 concentrations and soil properties. Simulations are conducted at the level of independent stands, each initialized under a “grass-only” condition. While individual plant traits and cohort dynamics determine species persistence, the interactions between climatic variability and disturbance ultimately shape stand-level vegetation composition, biomass, and structure.

Fire disturbance in LPJ-GUESS is simulated using two coupled modules: SIMFIRE and BLAZE. SIMFIRE (Knorr et al., 2014) is a semi-empirical model that calculates the annual burned area fraction of a grid cell. It determines fire probability using a statistical approach driven by fuel continuity (proxied by fAPAR simulated by the vegetation model), fire weather (using the Nesterov Index), and human population density. Subsequently, BLAZE (Rabin et al., 2017) provides a detailed mechanistic treatment of fire behavior and impacts. Using the burned area fraction generated by SIMFIRE to drive stochastic patch selection, BLAZE simulates fireline intensity, combustion completeness, and fire-induced mortality at the cohort level, thereby resolving the biophysical fluxes associated with biomass burning.

2.2 Satellite-derived burned area data

The MODIS MCD64A1 Collection 6 burned area product is a global satellite-derived dataset providing monthly estimates of burned area (BA) at a spatial resolution of 500 m. The MCD64A1 algorithm identifies burned pixels using changes in surface reflectance combined with active fire detections, enabling the

https://doi.org/10.5194/egusphere-2026-4188
Preprint. Discussion started: 4 August 2026

detection of burn scars across a wide range of ecosystems and fire regimes (Giglio et al. 2018). BA in MCD64A1 is reported as the approximate Julian day of burning for each pixel. However, the product is distributed as non-overlapping tiles in a Sinusoidal projection, which must be mosaicked and reprojected to the geographic coordinate systems (e.g., WGS84) standard in Earth system modeling. Additionally, uncertainties related to cloud cover, small fires, and sensor limitations can affect detection accuracy. For model–observation comparisons, these high-resolution, tile-based data must be spatially aggregated to the model grid and temporally harmonized – a process that forms the core of the methodological workflow described in this study.

3. Workflow Description

The workflow is implemented as a modular series of R scripts designed to guide the user from raw data acquisition to final comparative analysis. The process is divided into three distinct stages, utilizing the *terra* package for spatial manipulation, *httr* and *jsonlite* for API interaction, and standard library tools for robust file handling.

3.1 Study Area Definition and Data Acquisition

The first stage establishes the spatial domain and handles the retrieval of remote sensing data. This process addresses the common challenge of identifying which specific MODIS tiles are required to cover an irregular Region of Interest (ROI) and locating the correct version-controlled files. The workflow accommodates flexible input methods to suit different study designs. Users may define their study area by either a bounding box (providing a simple coordinate extent (min/max longitude and latitude), ideal for broad regional sweeps), or a vector polygon (importing a shapefile (.shp) representing the study domain, specific administrative boundaries, hydrobasins, or ecoregions using the *sf* package).

To bridge the gap between geographic coordinates and the MODIS projection, the script projects the user's ROI into the MODIS Sinusoidal coordinate system. It then mathematically intersects this footprint with the global tile grid to generate a precise list of required tile identifiers (e.g., h_{18v04}, h_{19v04}). This ensures that any tile overlapping even a fraction of the study area is identified. For reproducibility, the workflow includes a fully automated download routine (Appendix A). This script queries the NASA Common Metadata Repository (CMR) API to locate files from the latest MODIS Collection (currently 061), verifies

https://doi.org/10.5194/egusphere-2026-4188
Preprint. Discussion started: 4 August 2026

authentication using a user-supplied Earthdata token, and performs integrity checks (e.g., file size verification) during download. Further comments are embedded to the script (Appendix A). If the HDF files are already downloaded during a previous study, the user can skip this stage and simply define the existing folder path in the next stage.

3.2 Processing and Harmonization

The core of the workflow involves transforming the raw, tile-based satellite data into a format compatible with the gridded output of LPJ-GUESS. This is achieved through a two-step process that separates raw data extraction from model specific harmonization. The logic behind this separation is the modular approach to code writing, so that each R file represents a stand-alone application which can be applied independently if required.

3.2.1 Derivation of burned area

First, the downloaded raw MODIS tiles are processed to generate a consistent, pixel-level record of burned area (Appendix B). The script iterates through the downloaded HDF files, extracting the “Burn Date” and “QA” data. It applies a quality mask (defaulting to valid land pixels) and reprojects the data from the native Sinusoidal projection to the geographic WGS84 (EPSG:4326) coordinate system. Crucially, the workflow calculates the exact surface area of every pixel, dynamically accounting for the latitudinal variation in grid cell size inherent in the geographic coordinate system. This ensures that burned area estimates remain physically accurate even at high latitudes where grid distortion is significant. Finally, this stage exports a precise list of burned pixels (latitude, longitude, and area in km^2) for every month and year, providing an intermediate output that is resolution-independent and compatible with standard GIS software.

3.2.2 Grid Harmonization

To ensure spatial consistency during comparison, the workflow utilizes a “Master Grid” approach (Appendix C, Phase 1). Rather than projecting model outputs to match satellite data, the script generates a static “Master Grid” raster based on the unique longitude and latitude coordinates of the LPJ-GUESS input file. The WGS84 pixel list generated in the previous step is then spatially “snapped” to this template, eliminating edge-effect errors during comparison, and preventing the spatial mismatch often introduced when rasterizing point-based model outputs.

3.3 Comparative Analysis and Statistical Evaluation

The final stage utilizes the harmonized dataset to assess model performance against the observational benchmark using a multi-metric statistical framework (Appendix C, Phase 2-4). The processed MODIS dataframe is merged with the LPJ-GUESS outputs based on coordinate keys, allowing for cell-by-cell or regionally aggregated comparisons. To move beyond simple visual comparison, the workflow implements four specific statistical metrics designed to evaluate different aspects of fire regime performance:

1. *Long-term Spatial Agreement:* To assess if the model accurately captures the spatial distribution of fire prone areas, the Spearman rank correlation coefficient (ρ) is calculated between the simulated and observed mean annual burned area across all spatial units. This non-parametric test evaluates the model's ability to rank high- and low-fire areas correctly, minimizing the influence of absolute magnitude biases.
2. *Interannual Variability (IAV):* To evaluate the model's response to climate forcing, the workflow computes the Spearman correlation of the annual burned area time series for each individual spatial unit. This results in a distribution of correlation coefficients, highlighting regions where the model successfully captures fire-climate synchronization.
3. *Peak Fire Year Timing:* Extreme fire years often drive the majority of ecological impacts. The workflow identifies the year of maximum burned area for both the model and observation in every spatial unit and calculates the absolute difference in years. This metric assesses the model's ability to reproduce extreme events driven by specific climatic anomalies.
4. *Fire Dominance and Persistence:* To identify regions that disproportionately contribute to the total fire budget, a dominance index is calculated based on the frequency with which a spatial unit exceeds the 90th percentile of regional burned area. This helps distinguish between regions with chronic, frequent fire activity and those with rare, episodic events.

The workflow allows for these metrics to be aggregated into user-defined functional units (e.g., HydroBASINS or ecoregions) to contextualize model performance within distinct biomes.

4. Workflow Application

https://doi.org/10.5194/egusphere-2026-4188
Preprint. Discussion started: 4 August 2026

The utility of the presented workflow is demonstrated in a study by Ekberzade (2026a), where the results of LPJ-GUESS-SIMFIRE-BLAZE historical run forced with ERA5-land climate reanalysis data are benchmarked against MCD64A1. The model was configured using the parameterizations detailed in Ekberzade et al. (2025), and the simulation was conducted sans-human to isolate climate-driven fire dynamics. The benchmarking against the MCD64A1 dataset (processed via the workflow in Appendix B) covered the entire study area, a region characterized by heavy human interference and highly fragmented natural forests.

4.1 Diagnostic Capabilities

In the reference study, the workflow successfully harmonized the monthly satellite observations to the model's native grid (Appendix C, Phase 1) and generated a suite of diagnostic metrics (Appendix C, Phase 2–4). The automated analysis highlighted three distinct capabilities of the tool:

1. *Spatial Contextualization*: The long-term spatial agreement analysis (as implemented in the workflow's *S_pattern* module in Appendix C, Phase 4) immediately identified a regime-dependent divergence in model performance. The workflow revealed that while the model closely reproduced fire activity in natural forested areas with low fragmentation rates and limited human interference, it missed the fire dynamics in steppe ecosystems which are heavily managed, and where the fire regime is driven by anthropogenic factors. By aggregating pixel-level data into ecoregions, the tool allowed for a rapid assessment of where the model's "fire niche" aligns with reality, moving beyond simple global sums.
2. *Temporal Fidelity:* The interannual variability (IAV) and peak-year metrics provided a standardized way to assess the model's sensitivity to climate forcing. The workflow automatically flagged regions where the simulated fire season was out of phase with observations – specifically in tourism hotspots affected by episodic anthropogenic ignitions – while confirming robust synchronization in climate-driven Mediterranean forests where human activity is relatively limited. This automated flagging significantly reduced the labor required to distinguish model artifacts from genuine forcing responses, and thus provided a diagnostic step that is often manually laborious in ad-hoc comparisons.

https://doi.org/10.5194/egusphere-2026-4188
Preprint. Discussion started: 4 August 2026

3. *Regime Classification:* The fire dominance index successfully separated regions of chronic, fuel-limited fire activity from those driven by episodic, climate-limited events. This classification – generated automatically by the script – provided immediate feedback on whether the underlying driver of simulated fire (fuel availability vs. moisture deficit) matched the observational record.

5. Discussion

The primary contribution of this work is the reduction of technical friction in the validation pipeline. As highlighted in the introduction, the increasing frequency of climate-driven fire events underscores the need for rapid and robust model evaluation. However, comparing dynamic vegetation or Earth system model (ESM) outputs with remote sensing data has often relied on bespoke, project-specific and generally single-use scripts to handle the reprojection of Sinusoidal MODIS tiles and the aggregation of high-resolution pixels to coarse model grids (Hantson et al., 2016; Lasslop et al., 2020; Harrison et al., 2021). This ad-hoc approach is labor-intensive and prone to spatial edge artifacts. By standardizing the Processing (Appendix B) and Harmonization (Appendix C) stages, the presented workflow reduces the time required to move from raw model output to actionable diagnostics from weeks to hours, allowing researchers to focus on the ecological implications of their simulations rather than data engineering.

A key methodological innovation in the presented workflow is the "Master Grid" strategy. Comparisons between model simulations and satellite-derived observations require spatial harmonization between datasets that differ in projection, spatial resolution, and land-mask definition, a process that can introduce interpolation, aggregation, and geometric inconsistencies, particularly along coastlines and heterogeneous landscapes (Warszawski et al., 2014; Lasslop et al., 2020; Harrison et al., 2021). By generating a static template based strictly on the model's unique coordinate list, the workflow harmonizes observational data to the exact geometry of the simulation. This ensures that, in the final analysis, the observed and the simulated burned area fields are represented on an identical spatial geometry, minimizing spatial mismatch artifacts and facilitating robust grid-cell-by-grid-cell statistical analysis.

Additionally, while comprehensive benchmarking frameworks such as ILAMB provide extensive capabilities for evaluating ESMs, they are primarily designed for large-scale model intercomparison projects and are often implemented within Python-based software ecosystems that require substantial infrastructure and

https://doi.org/10.5194/egusphere-2026-4188
Preprint. Discussion started: 4 August 2026

configuration (Collier et al., 2018). Likewise, initiatives such as FireMIP have established standardized experimental protocols for global fire modelling, but the implementation of the associated processing and evaluation workflows is generally left to individual modelling groups (Rabin et al., 2017). The workflow presented here complements these efforts by providing a lightweight, model-agnostic methodology implemented entirely using open-source R packages (e.g., *terra*, *sf*), making standardized model evaluation more accessible to fire and vegetation ecologists as well as regional modelling studies.

Crucially, this design democratizes access to rigorous validation tools. Early-career researchers and domain experts often possess deep knowledge of fire ecology but may lack the specialized software engineering skills required to construct complex spatial data pipelines from scratch. Packaging these operations into a transparent framework aims to empower a broader spectrum of scientists to conduct reproducible model benchmarking (Baker 2016, Peng 2011). This supports the broader goal of the Earth system modeling community to standardize evaluation protocols, ensuring that model improvements are measured against consistent, reproducible baselines rather than methodological artifacts.

While the current implementation relies on the MODIS MCD64A1 product, the modular architecture of the workflow facilitates adaptation to other observational datasets. The Comparative Analysis (Appendix C) is inherently dataset-agnostic, requiring only a standardized coordinate-time-value input. Consequently, the Data Processing module (Appendix B) can be readily adapted to ingest higher-resolution products – such as active fire detections from VIIRS or burn scars from Sentinel-2 – while retaining the same robust "Master Grid" harmonization and statistical benchmarking framework. Future iterations could further expand this utility by incorporating support for sub-daily temporal resolution to validate diurnal fire cycles, thereby extending the tool's applicability to fine-scale fire behavior modeling.

6. Conclusion

Validating dynamic vegetation and fire models against satellite observations is a critical step in establishing confidence in future Earth system projections. However, this process is frequently hindered by the technical barriers of incompatible coordinate systems, differing spatial resolutions, and the lack of standardized processing protocols. The ad-hoc scripts often used to bridge this gap can introduce significant "edge-effect" errors and limit the reproducibility of model inter-comparison studies.

https://doi.org/10.5194/egusphere-2026-4188
Preprint. Discussion started: 4 August 2026

This paper presents a comprehensive, open-source R-based workflow designed to overcome these challenges. By providing a modular solution for (1) automated data acquisition, (2) physically accurate pixel-level derivation of burned area, and (3) rigorous spatial harmonization using a "Master Grid" approach, the workflow streamlines the path from raw data to actionable diagnostics. The utility of this framework is evident in its ability to not only quantify global bias but to spatially contextualize model performance – distinguishing, for example, between climate-driven fire regimes in Mediterranean forests and human-driven dynamics in managed steppes.

Ultimately, by reducing the technical burden associated with data preprocessing, this contribution shifts the research focus from code management to scientific inquiry. The adoption of this standardized protocol will facilitate more frequent and transparent validation efforts, thereby accelerating improvements in the next generation of fire-enabled dynamic vegetation models.

https://doi.org/10.5194/egusphere-2026-4188
Preprint. Discussion started: 4 August 2026

APPENDIX A

```
############################################################
# Data acquisition
############################################################
library(terra)
library(httr)
library(jsonlite)
library(sf)         # Added for optional shapefile handling
#############################################################
#                  USER SETTINGS
#############################################################
# 1. Earthdata Token (Get from https://urs.earthdata.nasa.gov/ -> Generate Token)
# Protip: tokens have expiration dates, so make sure it is still valid.
token <- "YOUR_TOKEN_HERE"

# To avoid ghost writing, a hard directory set is recommended
setwd(“THIS_IS_YOUR_DOWNLOAD_DIRECTORY”)

# 2. Define Input Method: Set ONE of these (leave the other NULL)
# Option A: Path to a shapefile (e.g., "data/my_study_area.shp")
shapefile_path <- NULL

# Option B: Bounding Box c(xmin, xmax, ymin, ymax) - Default: Turkey/East Med, adjust
# for your study area

roi_bounds <- c(26, 45, 36, 42)

# 3. Time Period, adjust accordingly
years <- 2021:2025

```

https://doi.org/10.5194/egusphere-2026-4188
Preprint. Discussion started: 4 August 2026


```
# 4. Output Directory: This section is optional. You can skip it if you used a hard
#directory set, setwd(), at the beginning
out_dir <- file.path(getwd(), "data", "modis_raw")
if (!dir.exists(out_dir)) dir.create(out_dir, recursive = TRUE)

# --- SAFETY CHECK ---
if (token == "YOUR_TOKEN_HERE" || is.null(token)) {
  stop("ERROR: Please replace 'YOUR_TOKEN_HERE' with your actual NASA Earthdata
token.")
}

################################################################
#                  FUNCTIONS
################################################################

# A. Calculate MODIS Tiles from Vector (Shapefile or Bbox)
get_required_tiles <- function(roi_vect) {
  # Project ROI to Sinusoidal (MODIS native projection)
  modis_crs <- "+proj=sinu +lon_0=0 +x_0=0 +y_0=0 +a=6371007.181 +b=6371007.181
+units=m +no_defs"
  roi_sinu <- project(roi_vect, modis_crs)
  e <- ext(roi_sinu)

  # MODIS Sinusoidal Tile Dimensions (approximate meters)
  tile_size_x <- 1111950.5197665554
  tile_size_y <- 1111950.5197665554

  # Calculate indices (H: 0-35, V: 0-17); x_origin is left edge, y_origin is top edge
  x_origin <- -20015109.354
  y_origin <- 10007554.677

```

https://doi.org/10.5194/egusphere-2026-4188
Preprint. Discussion started: 4 August 2026


```
  h_min <- floor((e$xmin - x_origin) / tile_size_x)
  h_max <- floor((e$xmax - x_origin) / tile_size_x)
  v_min <- floor((y_origin - e$ymax) / tile_size_y)
  v_max <- floor((y_origin - e$ymin) / tile_size_y)

  # Clamp to valid ranges
  h_min <- max(0, h_min); h_max <- min(35, h_max)
  v_min <- max(0, v_min); v_max <- min(17, v_max)

  tiles <- expand.grid(h = h_min:h_max, v = v_min:v_max)
  tiles$tile_id <- sprintf("h%02dv%02d", tiles$h, tiles$v)
  return(tiles$tile_id)
}

# B. Query NASA CMR API for Exact URLs (Collection 061)
get_modis_urls <- function(tile_id, start_year, end_year, auth_token) {
  print(paste("Searching for tile:", tile_id, "..."))

  url <- "https://cmr.earthdata.nasa.gov/search/granules.json"

  query_params <- list(
    short_name = "MCD64A1",
    version = "061",     #make sure this is the correct version!
    page_size = 2000,
    sort_key = "-start_date",
    temporal = paste0(start_year, "-01-01T00:00:00Z,", end_year, "-12-31T23:59:59Z"),
    "options[readable_granule_name][pattern]" = "true",
    "readable_granule_name" = paste0("*", tile_id, "*")
  )

```

https://doi.org/10.5194/egusphere-2026-4188
Preprint. Discussion started: 4 August 2026


```
  response <- GET(url, query = query_params, add_headers("Authorization" =
paste("Bearer", auth_token)))

  if (http_error(response)) {
    warning(paste("API Error for tile", tile_id, ":", status_code(response)))
    return(NULL)
  }

  content <- fromJSON(content(response, "text", encoding = "UTF-8"))

  # Check if entries found
  if (length(content$feed$entry$id) == 0) {
    warning(paste("No data found for", tile_id, "(Version 061)"))
    return(NULL)
  }

  entries <- content$feed$entry

  # Extract HDF links
  extract_hdf <- function(links_df) {
    if (is.null(links_df)) return(NA)
    url <- links_df$href[grep("\\.hdf$", links_df$href)]
    url <- url[!grepl("\\.xml$", url)]
    if(length(url) > 0) return(url[1]) else return(NA)
  }

  if("links" %in% names(entries)) {
    file_urls <- sapply(entries$links, extract_hdf)
    return(na.omit(file_urls))
  } else {
    return(NULL)
```

https://doi.org/10.5194/egusphere-2026-4188
Preprint. Discussion started: 4 August 2026


```
  }
}

############################################################
#                  EXECUTION
############################################################

# 1. Prepare ROI Vector
if (!is.null(shapefile_path)) {
  print(paste("Reading shapefile:", shapefile_path))
  roi_vect <- vect(shapefile_path)
} else {
  print("Using defined bounding box.")
  roi_vect <- ext(roi_bounds) %>% vect(crs = "EPSG:4326") %>% as.polygons()
}

# 2. Get Tiles
my_tiles <- get_required_tiles(roi_vect)
print(paste("Required Tiles:", paste(my_tiles, collapse = ", ")))

# 3. Download Loop
for (tile in my_tiles) {
  download_links <- get_modis_urls(tile, min(years), max(years), token)

  if (!is.null(download_links) && length(download_links) > 0) {
    print(paste("Tile", tile, ": Found", length(download_links), "files."))

    for (url in download_links) {
      file_name <- basename(url)
      dest_path <- file.path(out_dir, file_name)
```

https://doi.org/10.5194/egusphere-2026-4188
Preprint. Discussion started: 4 August 2026


```
      if (!file.exists(dest_path)) {
        tryCatch({
          response <- GET(url,
                          write_disk(dest_path, overwrite = TRUE),
                          add_headers("Authorization" = paste("Bearer", token)))

          if (status_code(response) != 200) {
            warning(paste("Failed (HTTP", status_code(response), "):", file_name))
            if (file.exists(dest_path)) file.remove(dest_path)
          } else {
            info <- file.info(dest_path)
            if (info$size < 10000) {
              warning(paste("File too small (auth error?):", file_name))
              file.remove(dest_path)
            } else {
              print(paste("SUCCESS:", file_name))
            }
          }
        }, error = function(e) {
          print(paste("Error downloading:", file_name, "-", e$message))
        })
      }
    }
  } else {
    print(paste("Tile", tile, ": No files found."))
  }
}
```

https://doi.org/10.5194/egusphere-2026-4188
Preprint. Discussion started: 4 August 2026

APPENDIX B.

```
############################################################
#   MCD64A1 Burned Area Workflow
# - Monthly & Annual BA (regional totals)
# - Pixel-level BA with lon/lat for GIS
############################################################
library(terra)
library(sf)
library(lubridate)
library(dplyr)
library(zoo)
library(ggplot2)
library(scales)
library(tidyr)

# --- 1. SETUP -------------------------------------------------
# Set the path to where you downloaded your MODIS tiles
setwd("THIS_IS_YOUR_WORKING_DIRECTORY")

files <- list.files(recursive = TRUE, pattern = "\\.hdf$", full.names = TRUE)

aoi_lonlat <- ext(24, 50, 34, 46)
aoi_vect   <- as.polygons(aoi_lonlat, crs = "EPSG:4326")

# --- 2. METADATA ---------------------------------------------

get_metadata <- function(filename) {
  tag <- sub(".*A([0-9]{7}).*", "\\1", basename(filename))
  year <- as.integer(substr(tag, 1, 4))
  doy  <- as.integer(substr(tag, 5, 7))
```

https://doi.org/10.5194/egusphere-2026-4188
Preprint. Discussion started: 4 August 2026


```
  date <- as.Date(doy - 1, origin = paste0(year, "-01-01"))
  list(file = filename, year = year, month = month(date))
}

files_df <- bind_rows(lapply(files, function(f) as.data.frame(get_metadata(f))))

# --- 3. MONTHLY PROCESSOR ------------------------------------

process_monthly_ba <- function(year, month, files_df, aoi_vect,
                               qa_mode = "conf1+") {

  target_date <- as.Date(paste(year, month, "01", sep = "-"))
  next_month  <- target_date + months(1)

  relevant_files <- files_df %>%
    filter(
      (year == year(target_date) & month == month(target_date)) |
        (year == year(next_month)  & month == month(next_month))
    ) %>%
    pull(file)

  if (length(relevant_files) == 0) return(NULL)

  burn_list <- list()

  for (f in relevant_files) {

    b <- try(rast(f, subds = 1), silent = TRUE)
    if (inherits(b, "try-error")) next

```

https://doi.org/10.5194/egusphere-2026-4188
Preprint. Discussion started: 4 August 2026


```
    if (qa_mode != "none") {
      q <- try(rast(f, subds = 3), silent = TRUE)
      if (inherits(q, "try-error")) next

      qa_conf <- q %% 4

      if (qa_mode == "conf1+") qa_mask <- qa_conf >= 1
      if (qa_mode == "high")   qa_mask <- qa_conf == 3

      b <- mask(b, qa_mask, maskvalues = 0)
    }

    if (global(b, "notNA")[[1]] > 0) {
      burn_list[[f]] <- b
    }
  }

  if (length(burn_list) == 0) return(NULL)

  burn_mosaic <- if (length(burn_list) == 1) {
    burn_list[[1]]
  } else {
    mosaic(sprc(burn_list), fun = "max")
  }

  aoi_modis <- project(aoi_vect, crs(burn_mosaic))

  burn_bin  <- ifel(burn_mosaic > 0, 1, NA)
  burn_crop <- crop(burn_bin, aoi_modis, mask = TRUE, touches = TRUE)

```

https://doi.org/10.5194/egusphere-2026-4188
Preprint. Discussion started: 4 August 2026


```
  area_crop <- cellSize(burn_crop, unit = "km")

  list(burned = burn_crop, area = area_crop)
}

# --- 4. MAIN EXECUTION ---------------------------------------

results_list        <- list()
mcd_pixels_monthly  <- list()   # <<< NEW
freq_map_accum      <- NULL

unique_months <- files_df %>%
  distinct(year, month) %>%
  arrange(year, month)

for (i in 1:nrow(unique_months)) {

  y <- unique_months$year[i]
  m <- unique_months$month[i]

  message("Processing: ", y, "-", sprintf("%02d", m))

  output <- process_monthly_ba(
    y, m, files_df, aoi_vect,
    qa_mode = "conf1+"
  )

  if (!is.null(output)) {

    r_month <- output$burned
```

https://doi.org/10.5194/egusphere-2026-4188
Preprint. Discussion started: 4 August 2026


```
a_month <- output$area

# ---- REGIONAL TOTAL --------------------------------------
burned_km2 <- global(r_month * a_month, "sum", na.rm = TRUE)[1, 1]

results_list[[i]] <- data.frame(
  year = y,
  month = m,
  burned_km2 = burned_km2,
  dataset = "MCD64A1"
)

# ---- PIXEL-LEVEL EXTRACTION (NEW) ------------------------
pixel_r <- r_month * a_month

df_pix <- as.data.frame(pixel_r, xy = TRUE, na.rm = TRUE)

if (nrow(df_pix) > 0) {
  names(df_pix) <- c("lon", "lat", "burned_km2")
  df_pix$year  <- y
  df_pix$month <- m
  mcd_pixels_monthly[[length(mcd_pixels_monthly) + 1]] <- df_pix
}

# ---- FREQUENCY MAP ---------------------------------------
r_binary <- ifel(is.na(r_month), 0, 1)

if (is.null(freq_map_accum)) {
  freq_map_accum <- r_binary
} else {
```

https://doi.org/10.5194/egusphere-2026-4188
Preprint. Discussion started: 4 August 2026


```
      r_binary <- resample(r_binary, freq_map_accum, method = "near")
      freq_map_accum <- freq_map_accum + r_binary
    }
  }
}

# --- 5. FINAL DATAFRAMES -------------------------------------

monthly_BA <- bind_rows(results_list)

mcd_pixels_monthly <- bind_rows(mcd_pixels_monthly)

mcd_pixels_annual <- mcd_pixels_monthly %>%
  group_by(year, lon, lat) %>%
  summarise(
    burned_km2 = sum(burned_km2, na.rm = TRUE),
    .groups = "drop"
  )

# --- 6. SAVE OUTPUTS -----------------------------------------
write.csv(
  monthly_BA,
  "MCD64A1_Monthly_BurnedArea_Final_QA1.csv",
  row.names = FALSE
)
write.csv(
  mcd_pixels_monthly,
  "MCD64A1_Monthly_Pixel_BurnedArea_km2_1.csv",
  row.names = FALSE
)
```

https://doi.org/10.5194/egusphere-2026-4188
Preprint. Discussion started: 4 August 2026


```
write.csv(
  mcd_pixels_annual,
  "MCD64A1_Annual_Pixel_BurnedArea_km2_2.csv",
  row.names = FALSE
)
writeRaster(
  freq_map_accum,
  "MCD64A1_Burn_Frequency_2000_2025_1.tif",
  overwrite = TRUE
)
############################################################
# 7. ARCGIS-READY SPATIO-TEMPORAL EXPORTS (OPTION A)
#    - Reproject MODIS Sinusoidal → WGS84 (EPSG:4326)
#    - Export monthly & annual pixel CSVs with true lon/lat
############################################################

# --- Helper: pixel table → sf → lon/lat --------------------

pixels_to_lonlat <- function(df_pixels, template_raster) {

  # Convert to sf using MODIS projection
  sf_modis <- st_as_sf(
    df_pixels,
    coords = c("lon", "lat"),
    crs = crs(template_raster)
  )

  # Reproject to WGS84
  sf_wgs84 <- st_transform(sf_modis, 4326)

```

https://doi.org/10.5194/egusphere-2026-4188
Preprint. Discussion started: 4 August 2026


```
  # Extract lon / lat
  coords <- st_coordinates(sf_wgs84)

  df_out <- sf_wgs84 %>%
    st_drop_geometry() %>%
    mutate(
      lon = coords[, 1],
      lat = coords[, 2]
    )

  return(df_out)
}

# --- Use frequency raster as projection reference ----------

template_raster <- freq_map_accum

# --- Monthly pixel BA (ArcGIS-ready) ------------------------

mcd_pixels_monthly_ll <- pixels_to_lonlat(
  mcd_pixels_monthly,
  template_raster
)

write.csv(
  mcd_pixels_monthly_ll,
  "MCD64A1_Monthly_Pixel_BurnedArea_km2_WGS84_ArcGIS.csv",
  row.names = FALSE
)
```

https://doi.org/10.5194/egusphere-2026-4188
Preprint. Discussion started: 4 August 2026


```
# --- Annual pixel BA (ArcGIS-ready) -------------------------

mcd_pixels_annual_ll <- pixels_to_lonlat(
  mcd_pixels_annual,
  template_raster
)

write.csv(
  mcd_pixels_annual_ll,
  "MCD64A1_Annual_Pixel_BA_km2_WGS84_ArcGIS.csv",
  row.names = FALSE
)

############################################################
# END OF ARCGIS EXPORT BLOCK
############################################################
```

https://doi.org/10.5194/egusphere-2026-4188
Preprint. Discussion started: 4 August 2026

APPENDIX C.

```
library(terra)
library(sf)
library(dplyr)
library(readr)
library(tidyr)
library(ggplot2)
library(scales)    # For nice axis number formatting
library(patchwork) # For final layout touches
library(tibble)
library(grid)
library(gridExtra)

setwd("THIS_IS_YOUR_WORKING_DIRECTORY")

#
#        📝PHASE 1 - Data Ingest, Master Grid Creation & Total Aggregation
#

########################################################################################
# This section of the code reads in LPJ-GUESS BA, turns it into a master grid, reads
# MCD64A1 BA (converted from raw Burn Date), reads in WWF ecoregions, matches all
# datasets to LPJ master grid, aggregates BA over ecoregions), compares BA in
# observation vs simulation
########################################################################################

##################################################################################
#          LPJ-GUESS ops.   - adjust for your model/output file names/types
##################################################################################

```

https://doi.org/10.5194/egusphere-2026-4188
Preprint. Discussion started: 4 August 2026


```r
# Read BA
lpj_tbl <- read_table("annual_burned_area.out", col_types = cols())
# glimpse(lpj_tbl)

# Filter for the relevant time period - to match MCD64A1
lpj_ba_tbl <- lpj_tbl %>%
  filter(Year >= 2001, Year <= 2024) %>% group_by(Lon, Lat, Year) %>%
  summarise(BurnFrac = sum(BurntFr, na.rm = TRUE), .groups = "drop")

# 2. CREATE MASTER GRID & STATIC LAYER
# Step A: Define the Grid Geometry using UNIQUE coordinates only
unique_grid <- lpj_ba_tbl %>% distinct(Lon, Lat) %>% mutate(val = 1)
lpj_template <- rast(unique_grid, type = "xyz", crs = "EPSG:4326")
writeRaster(lpj_template, "lpj_master_grid_G.tif", overwrite = TRUE)

# Step B: Calculate Cell Area (km²)
cell_area_km2 <- cellSize(lpj_template, unit = "km")

# Step C: Create the "Total Burned Area" Raster for ArcGIS
# Sum the fractions across all years to get the total footprint
lpj_total_frac <- lpj_ba_tbl %>%
  group_by(Lon, Lat) %>%
  summarise(TotalBurnFrac = sum(BurnFrac, na.rm = TRUE), .groups = "drop")

# Convert to Raster and multiply by area
lpj_ba_r <- rast(lpj_total_frac, type = "xyz", crs = "EPSG:4326") * cell_area_km2
names(lpj_ba_r) <- "lpj_ba_km2_total"

# Save for spatial analysis
writeRaster(lpj_ba_r, "lpj_ba_km2_total_2001-2024_G.tif", overwrite = TRUE)
```

https://doi.org/10.5194/egusphere-2026-4188
Preprint. Discussion started: 4 August 2026


```
###############################################################################
#          MCD64A1 ops.
###############################################################################

# Read MODIS BA
obs_tbl <- read_csv("MCD64A1_Annual_Pixel_BA_km2_WGS84_ArcGIS.csv", col_types =
cols())
# glimpse(obs_tbl)

# Snap observations to LPJ grid - spatial points (vector), not raster!
obs_pts <- vect(obs_tbl, geom = c("lon", "lat"), crs  = crs(lpj_template))

# Assign each MODIS pixel to an LPJ cell
obs_pts$lpj_cell <- cellFromXY(lpj_template, crds(obs_pts))

# Aggregate MODIS BA to LPJ grid
modis_lpj_tbl <- obs_pts |>
  as.data.frame() |>
  filter(!is.na(lpj_cell)) |>
  group_by(lpj_cell) |>
  summarise(modis_ba_km2 = sum(burned_km2, na.rm = TRUE))

# Create MODIS-on-LPJ raster
modis_lpj_r <- lpj_template
values(modis_lpj_r) <- 0 # Initialize with 0 - GEMINI
values(modis_lpj_r)[modis_lpj_tbl$lpj_cell] <- modis_lpj_tbl$modis_ba_km2
names(modis_lpj_r) <- "modis_ba_km2_total"

writeRaster(modis_lpj_r, "MODIS_BA_aggregated_to_lpj_total_G.tif", overwrite = TRUE) #
MODIS BA aggregated to LPJ
```

https://doi.org/10.5194/egusphere-2026-4188
Preprint. Discussion started: 4 August 2026


```
################################################################################
#   World terrestrial ecoregions ops. - adjust for your own aggregation unit
################################################################################

# Read ecoregions
eco_sf <- st_read("Ecoregions shp/ecoregions_clipped.shp")
# Re-project to LPJ grid if necessary
eco_sf <- st_transform(eco_sf, crs(lpj_template))

# Rasterize ecoregions onto LPJ grid
eco_r <- rasterize(vect(eco_sf), lpj_template, field = "ECO_ID")

writeRaster(eco_r, "eco_reg_lpj_G.tif", overwrite = TRUE)

################################################################################
#                  Ecoregion Aggregation
################################################################################

# 1.1 Zonal burned area (LPJ) — USE AREA, NOT FRACTION
ba_lpj_eco <- zonal(lpj_ba_r, eco_r, fun = "sum", na.rm = TRUE)
names(ba_lpj_eco) <- c("eco_id", "ba_lpj_km2")

# 1.2 Zonal burned area (MCD64A1) — USE MODIS-ON-LPJ RASTER
ba_obs_eco <- zonal(modis_lpj_r, eco_r, fun = "sum", na.rm = TRUE)
names(ba_obs_eco) <- c("eco_id", "ba_obs_km2")

# 1.3 Ecoregion area (LPJ-cell based, latitude-aware)
eco_area <- zonal(cell_area_km2, eco_r, fun = "sum", na.rm = TRUE)
```

https://doi.org/10.5194/egusphere-2026-4188
Preprint. Discussion started: 4 August 2026


```
names(eco_area) <- c("eco_id", "area_km2")

# Final ecoregion comparison table
eco_ba <- ba_lpj_eco %>%
  left_join(ba_obs_eco, by = "eco_id") %>%
  left_join(eco_area,  by = "eco_id") %>%
  mutate(ba_frac_lpj = ba_lpj_km2 / area_km2,
         ba_frac_obs = ba_obs_km2 / area_km2,
         ba_ratio    = ba_lpj_km2 / ba_obs_km2)

# SANITY CHECK
summary(eco_ba$ba_ratio)

#
#    📝PHASE 2: Time-Series Generation (Annual Loop)
#

######################################################################################
# This section considers LPJ-GUESS BA and MCD64A1 to see how they compare through time
# within each ecoregion. For that it computes time-resolved ecoregion series -->
# ecoregion x year - (unit = km² and fractional area);
# builds a canonical table: ecoregion × year × {LPJ_BA, MODIS_BA, ratios} to drive:
# all diagnostics, all plots, ArcGIS exports and regime-specific interpretation
######################################################################################

eco_year_list <- list()

# Identify common years
lpj_years <- unique(lpj_ba_tbl$Year)
mod_years <- unique(obs_tbl$year)
```

https://doi.org/10.5194/egusphere-2026-4188
Preprint. Discussion started: 4 August 2026


```
years <- intersect(lpj_years, mod_years)

message("Processing Years: ", paste(min(years), "to", max(years)))

for (yy in years) {

  # --- A. LPJ Yearly ---
  # Filter data for this year
  lpj_subset <- lpj_ba_tbl %>% filter(Year == yy)

  # Create raster for this year (Coords are unique here, so type="xyz" is safe)
  r_lpj_frac <- rast(lpj_subset[, c("Lon", "Lat", "BurnFrac")],
                     type = "xyz", crs = crs(lpj_template))

  # Extend to full grid (fills missing pixels with 0)
  r_lpj_frac <- extend(r_lpj_frac, lpj_template, fill = 0)

  # Convert Fraction -> Area
  r_lpj_ba <- r_lpj_frac * cell_area_km2

  # (Optional) Write annual raster if needed for ArcGIS animation
  # writeRaster(r_lpj_ba, paste0("LPJ_BA_", yy, ".tif"), overwrite=TRUE)

  # --- B. MODIS Yearly ---
  obs_pts_sub <- obs_pts[obs_pts$year == yy, ]

  modis_agg <- as.data.frame(obs_pts_sub) %>%
    filter(!is.na(lpj_cell)) %>%
    group_by(lpj_cell) %>%
    summarise(modis_ba_km2 = sum(burned_km2, na.rm = TRUE))
```

https://doi.org/10.5194/egusphere-2026-4188
Preprint. Discussion started: 4 August 2026


```
  r_mod_ba <- lpj_template
  values(r_mod_ba) <- 0
  values(r_mod_ba)[modis_agg$lpj_cell] <- modis_agg$modis_ba_km2

  # --- C. Zonal Stats ---
  lpj_stats <- zonal(r_lpj_ba, eco_r, fun = "sum", na.rm = TRUE)
  mod_stats <- zonal(r_mod_ba, eco_r, fun = "sum", na.rm = TRUE)

  names(lpj_stats) <- c("eco_id", "lpj_ba_km2")
  names(mod_stats) <- c("eco_id", "modis_ba_km2")

  # Combine
  eco_year_list[[as.character(yy)]] <- lpj_stats %>%
    left_join(mod_stats, by = "eco_id") %>%
    left_join(eco_area, by = "eco_id") %>%
    mutate(
      year = as.integer(yy),
      lpj_frac   = lpj_ba_km2 / area_km2,              #area-normalized BA
      modis_frac = modis_ba_km2 / area_km2,            #area-normalized BA
      ba_ratio   = lpj_ba_km2 / modis_ba_km2
    )
}

eco_year_ba <- bind_rows(eco_year_list)
# This CSV is your Phase-2 backbone:
write_csv(eco_year_ba, "Ecoregion_Yearly_BA_LPJ_vs_MODIS.csv")

summary(eco_year_ba$modis_ba_km2)
summary(eco_year_ba$lpj_ba_km2)
```

https://doi.org/10.5194/egusphere-2026-4188
Preprint. Discussion started: 4 August 2026


```
#
#      📝PHASE 3: Analytic Plots (Publication Ready)
#

################################################################################
#           Multi eco-region trajectory plot
################################################################################

unique(eco_year_ba$eco_id) #identify individual eco_ids, for sanity comparisons

# 1. Prepare Data Labels
eco_year_long <- eco_year_ba %>%
  filter(!is.na(modis_ba_km2), !is.na(lpj_ba_km2)) %>%
  pivot_longer(
    cols = c(modis_ba_km2, lpj_ba_km2),
    names_to = "raw_source",
    values_to = "ba_km2"
  ) %>%
  mutate(
    source_label = case_when(
      raw_source == "modis_ba_km2" ~ "MODIS (Observed)",
      raw_source == "lpj_ba_km2"   ~ "LPJ-GUESS (Modeled)",
      TRUE ~ raw_source
    )
  )

# 2. Define Colors
my_colors <- c(
  "MODIS (Observed)"      = "steel blue",
```

https://doi.org/10.5194/egusphere-2026-4188
Preprint. Discussion started: 4 August 2026


```r
  "LPJ-GUESS (Modeled)"    = "#D55E00"
)

# 3. Create Main Plot
p <- ggplot(eco_year_long, aes(x = year, y = ba_km2, color = source_label)) +
  geom_line(linewidth = 0.8, alpha = 0.8) +
  facet_wrap(~ eco_id, ncol = 4, scales = "free_y") +
  scale_color_manual(values = my_colors) +
  scale_y_continuous(labels = label_comma()) +
  labs(x = NULL, y = expression("Burned Area (km"^2*")")) +
  theme_bw(base_size = 14) +
  theme(
    legend.position = c(0.55, 0.02),
    legend.justification = c(0.5, 0),
    legend.direction = "horizontal",
    legend.title = element_blank(),
    legend.background = element_rect(fill = "white", color = "grey70"),
    legend.margin = margin(5, 10, 5, 10),
    strip.background = element_rect(fill = "grey95"),
    strip.text = element_text(face = "bold", size = 14),
    panel.grid.minor = element_blank(),
    axis.text = element_text(color = "black")
  )

# 4. Final Polish with Patchwork
final_plot <- p + plot_annotation(
  title = "Burned Area Trajectories: Model vs. Observation",
  subtitle = paste0("Annual area burned across ecoregions (", min(years), "-",
max(years), ")"),
  caption = "Data Sources: MODIS Collection 6 & LPJ-GUESS Simulation",
  theme = theme(
```

https://doi.org/10.5194/egusphere-2026-4188
Preprint. Discussion started: 4 August 2026


```
    plot.title = element_text(size = 18, face = "bold"),
    plot.subtitle = element_text(size = 16, color = "grey40")
  )
)

# 5. Save
ggsave("LPJ-MODIS_ecoregions_multiplot.png", final_plot, width = 16, height = 12, dpi = 300)

################################################################################
#           E. Fire regimes --- boxplot
################################################################################

# Define Regimes
eco_regime <- tibble(
  eco_id = c(786,678,703,650,662,665,649,646,674,658,692,654,804,
             785,790,791,789,
             661,725,688,727,652,735,695,
             831,820,815,841,817,830,739,812,747),
  regime = c(rep("Forest", 13), rep("Mediterranean", 4),
             rep("Steppe", 7), rep("Desert", 9))
)

# Join and Order Factors
eco_year_regime <- eco_year_ba %>%
  left_join(eco_regime, by = "eco_id") %>%
  mutate(regime = factor(regime, levels = c("Forest", "Mediterranean", "Steppe", "Desert")))

# Create Boxplot
fire_regime_plot <- ggplot(eco_year_regime %>% filter(modis_ba_km2 > 1),
```

https://doi.org/10.5194/egusphere-2026-4188
Preprint. Discussion started: 4 August 2026


```
                                  aes(x = regime, y = lpj_ba_km2 / modis_ba_km2)) +
  geom_boxplot(fill = "#1e52a6", color = "#1e52a6", alpha = 0.7,
               outlier.size = 3, outlier.shape = 21, outlier.fill = "#4287f5") +
  coord_cartesian(ylim = c(0, 80)) +
  labs(x = "", y = expression("LPJ-GUESS / MCD64A1 BA ratio"),
       title = "Comparison of LPJ-GUESS and MCD64A1 BA across fire-regime ecoregions",
       subtitle = "(Ratio = 1 indicates perfect agreement)") +
  theme_minimal() +
  theme(
    axis.text = element_text(size = 14),
    axis.title = element_text(size = 14),
    plot.title = element_text(face = "bold", size = 18),
    plot.subtitle = element_text(size = 16),
    plot.background = element_rect(fill = "white", color = NA)
  )

ggsave("LPJ-MODIS_boxplot_fire_regime.png", fire_regime_plot, width = 10, height = 8,
dpi = 300)

################################################################################
#          F. Ecoregion name - id table  --- Supplementary Figure
################################################################################

# 1. Create the Data (input for the table) - adjust for your own regional aggregation

eco_ref <- tribble(
  ~Type, ~ID, ~Name,
  "Forest", 786, "Anatolian conifer and deciduous mixed forests",
  "Forest", 678, "Rodope montane mixed forests",
  "Forest", 703, "Northern Anatolian conifer and deciduous forests",
  "Forest", 650, "Caucasus mixed forests",
```

https://doi.org/10.5194/egusphere-2026-4188
Preprint. Discussion started: 4 August 2026


```
    "Forest", 662, "Eastern Anatolian deciduous forests",
    "Forest", 665, "Euxine-Colchic broadleaf forests",
    "Forest", 649, "Caspian Hyrcanian mixed forests",
    "Forest", 646, "Balkan mixed forests",
    "Forest", 674, "Pannonian mixed forests",
    "Forest", 658, "Crimean Submediterranean forest complex",
    "Forest", 692, "Carpathian montane forests",
    "Forest", 654, "Central European mixed forests",
    "Forest", 804, "Southern Anatolian montane conifer and deciduous forests",
    "Mediterranean", 785, "Aegean and Western Turkey sclerophyllous/mixed forests",
    "Mediterranean", 790, "Cyprus Mediterranean forests",
    "Mediterranean", 791, "Eastern Mediterranean conifer-broadleaf forests",
    "Mediterranean", 789, "Crete Mediterranean forests",
    "Steppe", 661, "East European forest steppe",
    "Steppe", 725, "Central Anatolian steppe",
    "Steppe", 688, "Zagros Mountains forest steppe",
    "Steppe", 727, "Eastern Anatolian montane steppe",
    "Steppe", 652, "Central Anatolian steppe and woodlands",
    "Steppe", 735, "Pontic steppe",
    "Steppe", 695, "Elburz Range forest steppe",
    "Desert", 831, "North Arabian desert",
    "Desert", 820, "Central Persian desert basins",
    "Desert", 815, "Caspian lowland desert",
    "Desert", 841, "South Iran Nubo-Sindian desert and semi-desert",
    "Desert", 817, "Central Asian northern desert",
    "Desert", 830, "Mesopotamian shrub desert",
    "Desert", 739, "Syrian xeric grasslands and shrublands",
    "Desert", 812, "Azerbaijan shrub desert and steppe",
    "Desert", 747, "Tigris-Euphrates alluvial salt marsh"
  )
```

https://doi.org/10.5194/egusphere-2026-4188
Preprint. Discussion started: 4 August 2026


```
# 2. PLOT THE TABLE
# Optional: add a rectangle for every other row to create "zebra striping"
# for readability
eco_table_plot <- eco_ref %>%
  mutate(y_pos = row_number()) %>%
  ggplot() +

  # A. Zebra Striping (Grey bars for odd rows)
  geom_rect(
    data = . %>% filter(y_pos %% 2 == 1),
    aes(xmin = -0.1, xmax = 1.1, ymin = y_pos - 0.5, ymax = y_pos + 0.5),
    fill = "grey95", inherit.aes = FALSE
  ) +

  # B. Columns
  geom_text(aes(x = 0, y = y_pos, label = ID), hjust = 0, fontface = "bold", size = 3.5) +
  geom_text(aes(x = 0.1, y = y_pos, label = Name), hjust = 0, size = 3.5) +
  geom_text(aes(x = 1, y = y_pos, label = Type, color = Type), hjust = 1, size = 3.5, fontface = "italic") +

  # C. Headers (Manually added at y = 0)
  annotate("text", x = 0, y = 0, label = "ID", fontface = "bold", size = 4, hjust = 0) +
  annotate("text", x = 0.1, y = 0, label = "Ecoregion Name", fontface = "bold", size = 4, hjust = 0) +
  annotate("text", x = 1, y = 0, label = "Vegetation Type", fontface = "bold", size = 4, hjust = 1) +

  # D. Scales & Theme
  scale_y_reverse() + # Flips so row 1 is at top
```

https://doi.org/10.5194/egusphere-2026-4188
Preprint. Discussion started: 4 August 2026


```
  scale_x_continuous(limits = c(-0.1, 1.1)) +
  scale_color_manual(values = c("Forest" = "#082621",
                                "Mediterranean" = "#445EF2",
                                "Steppe" = "#425953",
                                "Desert" = "#A67041")) +
  theme_void() +
  theme(
    legend.position = "none",
    plot.background = element_rect(fill = "white", color = NA),
    plot.margin = margin(20, 20, 20, 20)
  )
# Also display it here to check
eco_table_plot

ggsave("Supplementary_Figure_Table.png", eco_table_plot, width = 8, height = 11, dpi = 300)

#
#     📝PHASE 4: Statistical analysis
#

#  Long-term spatial pattern agreement

# eco_year_ba       <- make sure you already have this
# columns: eco_id, year, lpj_ba_km2, modis_ba_km2, area_km2

eco_mean_skill <- eco_year_ba %>%
  group_by(eco_id) %>%
  summarise(
    lpj_mean   = mean(lpj_ba_km2, na.rm = TRUE),
    modis_mean = mean(modis_ba_km2, na.rm = TRUE),
```

https://doi.org/10.5194/egusphere-2026-4188
Preprint. Discussion started: 4 August 2026


```
    .groups = "drop") %>%
  filter(modis_mean > 1)  # remove fire-free noise

spatial_cor <- cor.test(
  eco_mean_skill$lpj_mean,
  eco_mean_skill$modis_mean,
  method = "spearman")

# Sanity check options
range(eco_mean_skill$modis_mean)
range(eco_mean_skill$lpj_mean)

# Save spatial_cor$estimate and p.value for the upcoming plot
rho <- round(spatial_cor$estimate, 2)
pval <- spatial_cor$p.value

# Does the model burn the right places, even if not the right amounts? (regardless
# of magnitude bias)

x_anno <- max(eco_mean_skill$modis_mean, na.rm = TRUE) * 0.95
y_anno <- min(eco_mean_skill$lpj_mean,   na.rm = TRUE) * 1.2

S_pattern <- ggplot(eco_mean_skill,
       aes(x = modis_mean, y = lpj_mean)) +
  geom_point(size = 3, alpha = 0.75, color = "grey30") +
  geom_smooth(method = "lm", se = FALSE, linetype = "dashed") +
  scale_x_log10() +
  scale_y_log10() +

  annotate("text", x = x_anno, y = y_anno,
```

https://doi.org/10.5194/egusphere-2026-4188
Preprint. Discussion started: 4 August 2026


```
    hjust = 1.05, vjust = -0.5,
    label = paste0("Spearman ρ = ", rho,
                   "\np = ", signif(pval, 2)), size = 4) +
  labs(
    x = "Observed mean annual burned area (km², MODIS)",
    y = "Simulated mean annual burned area (km², LPJ-GUESS)",
    title = "Spatial agreement of long-term burned area across ecoregions"
  ) +
  theme_minimal(base_size = 14) +
  theme(plot.background = element_rect(fill = "white", color = NA))
S_pattern

# Spearman ρ → spatial ranking skill ; Regression line → proportional bias ; Log-log →
# cross-regime comparability

ggsave("Longterm_spatial_pattern_agreement.png", S_pattern, width = 10, height = 5,
dpi = 300)

# Interannual variability skill

eco_interannual_skill <- eco_year_ba %>%
  filter(modis_ba_km2 > 1) %>%  # avoid zeros
  group_by(eco_id) %>%
  summarise(
    r_interannual = cor(
      lpj_ba_km2,
      modis_ba_km2,
      method = "spearman",
      use = "complete.obs"), .groups = "drop")

# Does the model respond to climate variability correctly?
```

https://doi.org/10.5194/egusphere-2026-4188
Preprint. Discussion started: 4 August 2026


```
# Boxplot of ecoregion-wise correlations

S_eco <- ggplot(eco_interannual_skill,
       aes(y = r_interannual)) +
  geom_boxplot(fill = "#1e52a6", color = "#1e52a6", alpha = 0.7,
               width = 0.3, outlier.size = 3, outlier.shape = 21, outlier.fill =
"#4287f5") + #outlier.shape = 16, outlier.alpha = 0.6) +
  # stat_summary(fun = mean,geom = "point",
  #              size = 3) +
  geom_hline(yintercept = 0, linetype = "dashed", color = "grey20") +
  coord_cartesian(ylim = c(-1, 1)) +
  labs(
    y = "Spearman correlation (annual burned area)",
    title = "Interannual burned area variability skill across ecoregions"
  ) +
  theme_minimal(base_size = 14) +
  theme(plot.title = element_text(face = "bold", size = 18),
    plot.background = element_rect(fill = "white", color = NA))
S_eco

ggsave("Spearmancorr.png", S_eco, width = 10, height = 5, dpi = 300)

# Peak fire year agreement

eco_peak_years <- eco_year_ba %>%
  group_by(eco_id) %>%
  summarise(
    peak_year_lpj   = year[which.max(lpj_ba_km2)],
    peak_year_modis = year[which.max(modis_ba_km2)],
    .groups = "drop") %>%
  mutate(peak_year_diff = abs(peak_year_lpj - peak_year_modis))
```

https://doi.org/10.5194/egusphere-2026-4188
Preprint. Discussion started: 4 August 2026


```
summary(eco_peak_years$peak_year_diff)
quantile(eco_peak_years$peak_year_diff, probs = c(0.25, 0.5, 0.75))

# Does the model get the big years right?
# Distribution of peak year differences:
S_peak <- ggplot(eco_peak_years,
       aes(x = peak_year_diff)) +
  geom_histogram(binwidth = 1, boundary = 0, closed = "left", fill =  "#1e52a6", color = "#1e52a6", alpha = 0.8) +

  # quantile(eco_peak_years$peak_year_diff, probs = c(0.25, 0.5, 0.75)):
  geom_vline(xintercept = c(3, 7, 12),
             linetype = c("dashed", "solid", "dashed"), linewidth = 0.8, color = "grey50", alpha = 0.7)+

  labs(x = "Absolute difference in peak fire year (years)",
    y = "Number of ecoregions",
    title = "Agreement in timing of peak fire activity") +

  theme_minimal(base_size = 14) +
  theme(plot.title = element_text(face = "bold", size = 18),
        plot.background = element_rect(fill = "white", color = NA))

S_peak

ggsave("Peakyearcheck.png", S_peak, width = 10, height = 5, dpi = 300)

# Fire persistence / dominance metric
global_thresh <- quantile(eco_year_ba$modis_ba_km2, 0.90, na.rm = TRUE)
```

https://doi.org/10.5194/egusphere-2026-4188
Preprint. Discussion started: 4 August 2026


```r
eco_fire_persistence <- eco_year_ba %>%
  group_by(eco_id) %>%
  summarise(
    fire_dominance = mean(modis_ba_km2 > global_thresh, na.rm = TRUE),
    .groups = "drop"
  )

# Which ecoregions contribute disproportionately to large fires across the study area?
# Global-threshold dominance index:

S_persist <- ggplot(eco_fire_persistence %>%
         arrange(desc(fire_dominance)) %>%
         mutate(eco_id = factor(eco_id, levels = eco_id)),
       aes(x = eco_id, y = fire_dominance)) +
  geom_col() +
  coord_flip() +
  labs(
    x = "Ecoregion",
    y = "Fire dominance index",
    title = "Persistence of high burned area across ecoregions"
  ) +
  theme_minimal(base_size = 13) +
  theme(plot.title = element_text(face = "bold", size = 18),
        plot.background = element_rect(fill = "white", color = NA))
S_persist

ggsave("Firepersistence.png", S_persist, width = 10, height = 5, dpi = 300)

# Faceted interannual time series

```

https://doi.org/10.5194/egusphere-2026-4188
Preprint. Discussion started: 4 August 2026


```
top_ecos <- eco_year_ba %>%
  group_by(eco_id) %>%
  summarise(total_ba = sum(modis_ba_km2, na.rm = TRUE)) %>%
  arrange(desc(total_ba)) %>%
  slice_head(n = 6) %>%
  pull(eco_id)

top_ecos <- c(786, 790, 804, 665, 692, 703) #, 727) #649

S_faceted_eco <- eco_year_ba %>%
  filter(eco_id %in% top_ecos) %>%
  pivot_longer(c(lpj_ba_km2, modis_ba_km2),
               names_to = "source",
               values_to = "ba_km2") %>%
  ggplot(aes(year, ba_km2, color = source), alpha = 0.8) +
  geom_line(linewidth = 1.2) +
  facet_wrap(~ eco_id, scales = "free_y") +
  scale_color_manual(values = c("lpj_ba_km2" = "#D55E00",   # Orange
                                "modis_ba_km2" = "steel blue"), # Blue
                     labels = c("LPJ-GUESS", "MODIS")) +     # Optional: Rename legend
labels
  labs(
    x = "",
    y = "Burned area (km²)",
    color = ""
  ) +
  theme_minimal(base_size = 13) +
  theme(plot.title = element_text(face = "bold", size = 18),
        plot.background = element_rect(fill = "white", color = NA))

S_faceted_eco
```

https://doi.org/10.5194/egusphere-2026-4188
Preprint. Discussion started: 4 August 2026


```
ggsave("Faceted_eco.png", S_faceted_eco, width = 10, height = 5, dpi = 300)

Caspian <- eco_year_ba %>%
  filter(eco_id == 649) %>%
  pivot_longer(c(lpj_ba_km2, modis_ba_km2),
               names_to = "source",
               values_to = "ba_km2") %>%
  ggplot(aes(year, ba_km2, color = source)) +
  geom_line() +
  labs(
    x = "Year",
    y = "Burned area (km²)",
    color = "Dataset"
  ) +
  theme_minimal(base_size = 13) +
  theme(plot.title = element_text(face = "bold", size = 18),
        plot.background = element_rect(fill = "white", color = NA))
Caspian
ggsave("Caspian.png", Caspian, width = 10, height = 5, dpi = 300)

EAnatolian_montane_steppe <- eco_year_ba %>%
  filter(eco_id == 727) %>%
  pivot_longer(c(lpj_ba_km2, modis_ba_km2),
               names_to = "source",
               values_to = "ba_km2") %>%
  ggplot(aes(year, ba_km2, color = source)) +
  geom_line() +
  labs(
    x = "Year",
```

https://doi.org/10.5194/egusphere-2026-4188
Preprint. Discussion started: 4 August 2026


```
    y = "Burned area (km²)",
    color = "Dataset"
  ) +
  theme_minimal(base_size = 13) +
  theme(plot.title = element_text(face = "bold", size = 18),
        plot.background = element_rect(fill = "white", color = NA))
EAnatolian_montane_steppe
ggsave("Eastern_Anatolian_montanesteppe.png", EAnatolian_montane_steppe, width = 10,
height = 5, dpi = 300)
```

Code and data availability

The R scripts used to produce the results in this paper are archived on Zenodo and publicly available at https://doi.org/10.5281/zenodo.18300982 (Ekberzade, 2026b). The MODIS MCD64A1 data is openly available from the NASA Earthdata Search (https://search.earthdata.nasa.gov/).

Author contributions

Bikem Ekberzade: Conceptualization, investigation, writing, methodology, software, validation, formal analysis, data curation, visualization.

Conflicts of interest

The author declares no conflicts of interest.

Financial support

This study has been produced benefiting in part from Grant MGA-2026-47431 by Istanbul Technical University Scientific Research Programme (ITU-BAP). Computing resources used in this work were provided by the National Center for High Performance Computing (UHeM) in Turkey, under Grant Number 1007482020.

References

Abatzoglou, J. T. and Williams, A. P.: Impact of anthropogenic climate change on wildfire across western US forests, Proceedings of the National Academy of Sciences, 113, 11770–11775, https://doi.org/10.1073/pnas.1607171113, 2016.

Baker, M.: 1,500 scientists lift the lid on reproducibility, Nature, 533, 452–454, https://doi.org/10.1038/533452a, 2016.

Bowman, D. M. J. S., Kolden, C. A., Abatzoglou, J. T., Johnston, F. H., Van Der Werf, G. R., and Flannigan, M.: Vegetation fires in the Anthropocene, Nature Reviews Earth & Environment, 1, 500–515, https://doi.org/10.1038/s43017-020-0085-3, 2020.

https://doi.org/10.5194/egusphere-2026-4188
Preprint. Discussion started: 4 August 2026

Collier, N., Hoffman, F. M., Lawrence, D. M., Keppel-Aleks, G., Koven, C. D., Riley, W. J., et al.: The International Land Model Benchmarking (ILAMB) system: Design, theory, and implementation, Journal of Advances in Modeling Earth Systems, 10(11), 2731–2754, https://doi.org/10.1029/2018MS001354, 2018.

Ekberzade, B.: Forecasting the Flame: Simulating Future Wildfire Regimes in the Northeast Mediterranean, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2026-1498, 2026a.

Ekberzade, B.: A seamless workflow for wildfire modelers using LPJ-GUESS and remote sensing data for regional and global validation (Version v1.0.0) [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.18300984, 2026b.

Ekberzade, B., Görüm, T., Karabacak, F., et al.: Up in flames: the human factor behind a megafire in Mediterranean Türkiye, npj Natural Hazards, 2, 65, https://doi.org/10.1038/s44304-025-00120-4, 2025.

Giglio, L., Boschetti, L., Roy, D. P., Humber, M. L., and Justice, C. O.: The Collection 6 MODIS burned area mapping algorithm and product, Remote Sensing of Environment, 217, 72–85, https://doi.org/10.1016/j.rse.2018.08.005, 2018.

Hantson, S., Arneth, A., Harrison, S. P., Kelley, D. I., Prentice, I. C., Rabin, S. S., et al.: The status and challenge of global fire modelling, Biogeosciences, 13, 3359–3375, https://doi.org/10.5194/bg-13-3359-2016, 2016.

Harrison, S. P., Prentice, I. C., Bloomfield, K. J., Dong, N., Forkel, M., Forrest, M., Ningthoujam, R. K., Pellegrini, A., Shen, Y., Baudena, M., Cardoso, A. W., Huss, J. C., Joshi, J., Oliveras, I., Pausas, J. G., and Simpson, K. J.: Understanding and modelling wildfire regimes: an ecological perspective, Environmental Research Letters, 16, 125008, https://doi.org/10.1088/1748-9326/ac39be, 2021.

Jones, M. W., Abatzoglou, J. T., Veraverbeke, S., Andela, N., Lasslop, G., Forkel, M., et al.: Global and regional trends and drivers of fire under climate change, Reviews of Geophysics, 60(3), e2020RG000726, https://doi.org/10.1029/2020RG000726, 2022.

Knorr, W., Kaminski, T., Arneth, A., and Weber, U.: Impact of human population density on fire frequency at the global scale, Biogeosciences, 11(4), 1085–1102, https://doi.org/10.5194/bg-11-1085-2014, 2014.

https://doi.org/10.5194/egusphere-2026-4188
Preprint. Discussion started: 4 August 2026

Lai, J., Lortie, C. J., Muenchen, R. A., Yang, J., and Ma, K.: Evaluating the popularity of R in ecology, Ecosphere, 10(1), e02567, https://doi.org/10.1002/ecs2.2567, 2019.

Lasslop, G., Hantson, S., Harrison, S. P., et al.: Global ecosystems and fire: Multi-model assessment of fire-induced tree-cover and carbon storage reduction, Global Change Biology, 26, 5027–5041. https://doi.org/10.1111/gcb.15160, 2020.

Peng, R. D.: Reproducible research in computational science, Science, 334(6060), 1226–1227, https://doi.org/10.1126/science.1213847, 2011.

Smith, B.: LPJ-GUESS–an ecosystem modelling framework, in: Technical description, edited by: Knorr, W., Department of Physical Geography and Ecosystems Analysis, Lund University, Sweden, 2007.

Smith, B., Prentice, I. C., and Sykes, M. T.: Representation of vegetation dynamics in the modelling of terrestrial ecosystems: comparing two contrasting approaches within European climate space, Global Ecology and Biogeography, 10, 621–637, 2001.

Rabin, S. S., Melton, J. R., Lasslop, G., Bachelet, D., Forrest, M., Hantson, S., et al.: The Fire Modeling Intercomparison Project (FireMIP), phase 1: experimental and analytical protocols with detailed model descriptions, Geoscientific Model Development, 10(3), 1175–1197, https://doi.org/10.5194/gmd-10-1175-2017, 2017.

Warszawski, L., Frieler, K., Huber, V., Piontek, F., Serdeczny, O., and Schewe, J.: The Inter-Sectoral Impact Model Intercomparison Project (ISI–MIP): Project framework, Proceedings of the National Academy of Sciences, 111, 3228–3232, https://doi.org/10.1073/pnas.1312330110, 2013.