Step 6.2 Identifying misclassifications at the site scale

6.2.3 Computing map of differences

Diane ESPEL

2025-06-20

Contents

1 Objectives

This script aims to quantitatively and spatially compare observed habitat maps with model-predicted habitat maps across multiple classification levels. It generates binary difference maps that highlight spatial discrepancies, enabling the evaluation of prediction performance in ecological mapping tasks.

2 Script explanation

2.1 Clean environment and graphics

```
rm(list = ls())  # Clear all objects from the R environment to start fresh
graphics.off()  # Close all graphics devices (if any plots are open)
```

2.2 Load required packages

```
library(terra)  # For raster and vector spatial data handling
```

2.3 Define Global Variables

Note: A global variable is a variable defined outside of any function. This means the variable is accessible from any part of the code, including inside functions. A global variable retains its value throughout the execution of the R script unless it is explicitly modified in the code.

It is important to define at a minimum:

- the `"District"`: the archipelago of interest (e.g. "CRO" for Crozet archipelago)
- the `"Island"`: the island within the archipelago of interest (e.g. "POS" for Possession island)
- the `"Satellite1"`: the satellite name for multispectral imagery
- the `"Year1"` : the acquisition year of the multispectral imagery
- the `"Res1"` : the spatial resolution of the multispectral imagery
- the `"maxTypoLevel"`: the maximum typology level

```
District = "CRO"  # 3-letter code for archipelago (e.g. Crozet)
Island = "POS"  # 3-letter code for island (e.g. Possession)
Satellite1 = "Pleiades"  # satellite name of multispectral imagery
Year1 = "2022"  # acquisition year of multispectral imagery
Res1 = "50cm"  # spatial resolution of multispectral imagery
maxTypoLevel = 4  # Define maximum typology level
```

2.4 Set working directory

You must define one root directory (`"localscratch"`) that serves as base path for your input and output data. This directory should point to the local environment where:

- input observed map (vector) with new typology is located under `"data/raster/Observed_map/NewTypo"`
- input predict habitat map is located under `"results/Predictions"`
- output maps of differences will be saved under `"results/Difference_maps"`

```
# Base local path (customize to your local environment)
localscratch = paste0("/scratch/despel/CARTOVEGE/")
# localscratch = paste0('your_local_path/')

# Path to open rasterized observed map
open_NewObsMap_raster_path = paste0(localscratch, "data/raster/Observed_map/NewTypo")

# Path to open predict habitat map
open_predictions_path = paste0(localscratch, "results/Predictions")

# Path to save map of differences
save_differences_path = paste0(localscratch, "results/Difference_maps")
```

2.5 Compute binary difference maps

For each modeling strategy (`FLAT` or `HIERARCHICAL`), the script loops over typology levels. For each level, it loads the corresponding predicted (`raster_pred`) and observed (`raster_obs`) raster map. If their spatial extents differ, the observed map is resampled to align with the predicted raster using nearest-neighbor interpolation to ensure accurate pixel-wise comparison.

A difference raster (`diff_raster`) is computed where each pixel value indicates whether the predicted and observed class codes match (`FALSE`) or differ (`TRUE`). The resulting binary raster is both plotted as a PNG (colored yellow for no difference and red for mismatches) and saved as a GeoTIFF.

A summary metric—the proportion of differing cells—is computed and stored in a cumulative results table for each level and model type (`results_summary`) Finally, this summary table provides a clear overview of the spatial agreement between predictions and observations across habitat levels and modeling approaches. This facilitates downstream quantitative analysis of classification accuracy at different thematic resolutions.

```
# Define list of modelling strategy
type_model_list = c("FLAT", "HIERARCHICAL")


# Loop through model types
for (type_model in type_model_list) {

    # type='FLAT' #debug
    print(paste0("Modeling strategy: ", type_model))

    # Initialize a summary table to store the ratio of differences
    results_summary <- data.frame(model = character(), Year = character(), level = character(),
        ratio_diff = numeric())

    # Compute difference map for each level of classification
    for (l in seq(1:maxTypoLevel)) {

        rint(paste0("Processing habitat typology level ", l))

        # Load predicted raster map
        print("Opening smoothed predicted habitat raster")
        LevelFolder = paste0(open_predictions_path, "/", "Hab_L", l)
        FILE1 = paste0(LevelFolder, "/", "Smoothed_final_map_RF_", type_model, "_model_",
            District, "_", Island, "_", Satellite1, "_", Year1, "_", Res1, "_level_",
            l, ".TIF")
        raster_pred = rast(FILE1)
        print(raster_pred)

        # Load rasterized observed map
        print("Opening rasterized observed habitat map")
        FILE2 = paste0(open_NewObsMap_raster_path, "/", "Corrected_observed_map_NewTypo_",
            District, "_", Island, "_", Satellite1, "_", Year1, "_", Res1, "_level_",
            l, "_EPSG32739.TIF")
        raster_obs = rast(FILE2)
        print(raster_obs)

        # Align spatial extent if needed
        if (!identical(ext(raster_obs), ext(raster_pred))) {
            print("Rasters do not have the same extent. Resampling observed raster to match
             ↪  prediction.")
            raster_obs <- resample(raster_obs, raster_pred)
        } else {
            print("Rasters have identical spatial extent.")
        }

        # Compute difference map (TRUE where values differ, FALSE where they
        # match)
        print("Computing raster of differences")
        diff_raster <- raster_obs != raster_pred

        # Release memory
        rm(raster_obs)
        rm(raster_pred)
```

```
        # Plot and save the difference map (.png)
        cols <- c("yellow", "red")  # Yellow = no difference, red = difference
        legend_labels <- c("Pas de différence", "Différence")  #legend

        png(filename = paste0(save_differences_path, "/", "Difference_map_RF_", type_model,
            "_model_", District, "_", Island, "_", Year1, "_", Res1, "_level_", l,
            ".png"))
        p = plot(diff_raster, col = cols, legend = TRUE, legend.labels = legend_labels,
            main = "Raster de Différence entre les habitats observés et les habitats prédits")
        legend("topright", legend = legend_labels, fill = cols)
        print(p)
        dev.off()

        # Save difference map (.tif)
        NOMtiff = paste0(save_differences_path, "/", "Difference_map_RF_", type_model,
            "_model_", District, "_", Island, "_", Year1, "_", Res1, "_level_", l,
            ".TIF")
        writeRaster(diff_raster, NOMtiff, overwrite = T)

        # Calculate the ratio of differing cells
        num_true_cells <- sum(diff_raster[], na.rm = TRUE)
        num_total_cells <- ncell(diff_raster)
        ratio <- num_true_cells/num_total_cells
        cat("Ratio of different cells:", ratio, "\n")  # Print the ratio

        # Append results to summary table
        results_summary <- rbind(results_summary, data.frame(model = paste0(type,
            "_model"), Year = Year1, level = paste0("Hab_L", l), ratio_diff = ratio))

        # Release memory
        rm(diff_raster)

    }  # End of typology level loop

    # Save the summary results to CSV
    # ------------------------------------------------------------

    write.table(results_summary, file = paste0(save_differences_path, "/Difference_ratios_RF_",
        type_model, "_model_", District, "_", Island, "_", Year1, ".csv"), sep = ";",
        dec = ".", row.names = FALSE)

}  # End of model type loop
```