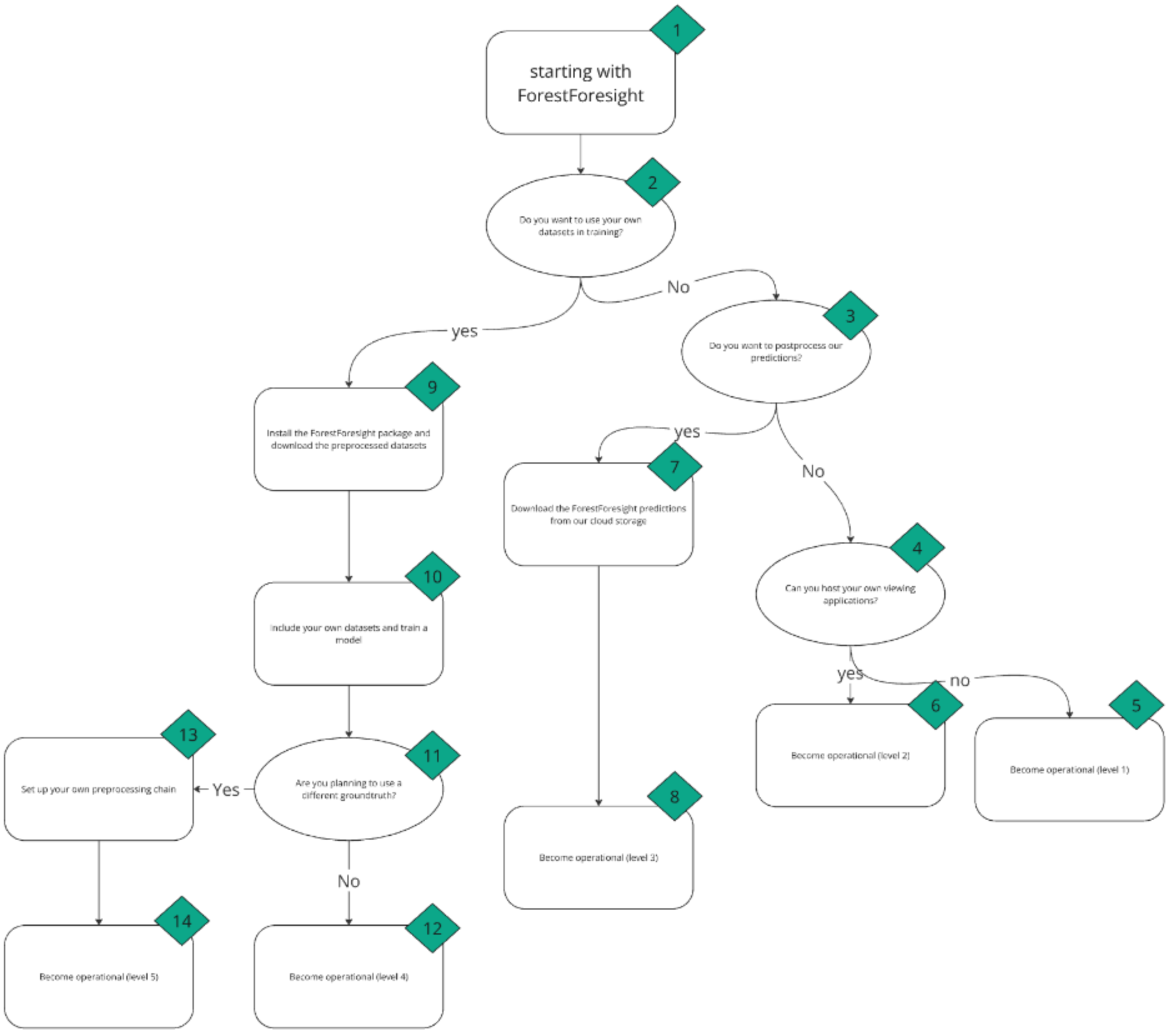
Forest Foresight Manual

This manual provides the user all necessary steps to make ForestForesight operational for their own system. Please start at the overview flowchart to find out which steps you need to take and which level of user you are at any point in time.

For other questions and/or feedback, get in touch through forestforesight@wwf.nl

Overview



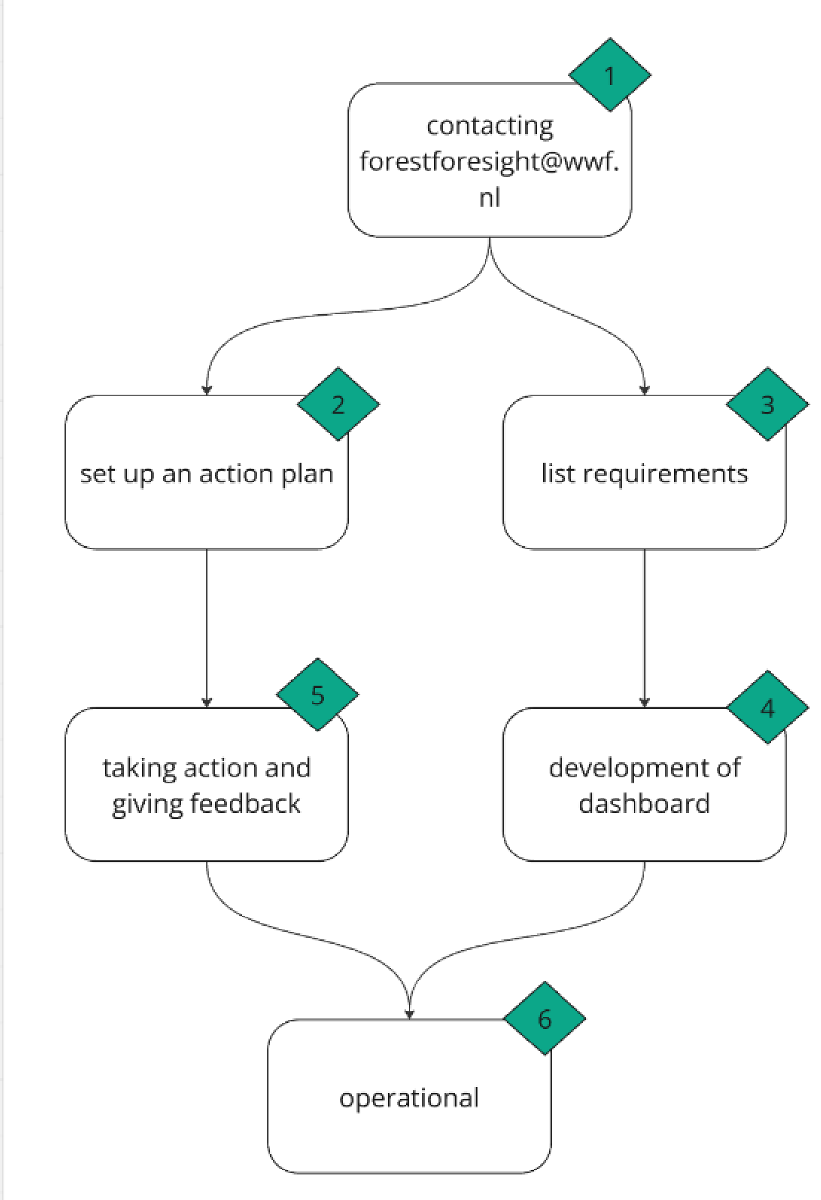
1. Starting with ForestForesight means that eventually you want to use the predictive power of AI to predict where deforestation will take place so that interventions can take place before it has all already taken place. This guide will take you through which method suits your needs and the final solution can be one of 5 options (level 1 from most simple to level 5 for most configurable)
2. Using your own datasets for training means that you have some datasets for which you want to see if it can help to predict deforestation. Before taking this path we want to advise three things:
    a. If the data has a low resolution, spatially or temporally, we know from experience that it will most probably not result in a high increased accuracy.
    b. This step should only be taken if you want to use it to improve the trained model. If you just want to customize the end result then your dataset is

not for training but for postprocessing (and you might end up in level 3 and not 4 or 5)
    c. Data with a high temporal resolution (monthly or more) requires also a large time investment in preprocessing, which should be taken into account.

3. The standard end-result of ForestForesight is a raster dataset of one layer with as value the likelihood that deforestation will happen in that pixel. If you want to do any of the following things, then follow the yes-path.
    a. Convert the data to vector for viewing and interactivity
    b. Add context to predictions (to make the data filterable interactively for instance)
    c. Filter out the predictions before you display them to the users
    d. Display the data on arcgis online, powerbi or other non-GIS specific viewing applications.
4. A viewing application is software like Qgis or ArcGIS. If you have those you can just load in our prediction rasters. If you want to use software like ArcGIS online you can either choose no and we might help with viewing capabilities or go back and use postprocessing to create datasets that can be viewed there.
5. If you do not have viewing applications and don't require customization the ForestForesight team can help you with dashboarding functionality after setting up a cooperative agreement. Get in touch through forestforesight@wwf.nl and follow the chart here
6. In QGis and ArcGIS you can add a live connection to our hosted datasets and have continuously live connection to the most recent deforestation predictions. You can find the instructions here.
7. ForestForesight runs their predictions monthly for all tropical and subtropical areas. You can download the resulting raster datasets from the public repository. Instructions for connecting to our data repository can be found here. After connecting you can go to the folder predictions/{your country code, e.g. BRA for Brazil} and download the data for the latest month or from our historic archive.
8. When you start using our predictions you might want to automate the process somwhat by using the AWS CLI. Instructions can be found here.
9. To start processing (which means training and/or predicting the model with a mix of FF input data and your own) you will need to download and install the ForestForesight package and the preprocessed datasets. Instructions for connecting to the data repository can be found here. Installing the ForestForesight package requires R and instructions can be found here.

10. If you want to include your own datasets for training of the algorithm you can find instructions to do so here.
11. ForestForesight does not do field research for groundtruth because we cover the entire pantropical area on a monthly basis. We assume, even though there are downsides to this, that the integrated alerts of GFW are the “truth” of deforestation on which the model should be trained. However, as a user you can freely choose your own groundtruth datasets. For instructions, follow the link here and specifically the part about dynamic datasets.
12. To become operational on level 4 you will have to update your input datasets and groundtruth datasets on a monthly basis using the ForestForesight repository and where applicable your own. For this you will have to:
    a. Do monthly downloading of input and groundtruth data. For this either use cyberduck according to the instructions from step 7. If you want to sync or automate the process you can do that by using the instructions in step 8. Doing preprocessing is explained in step 13
    b. Periodically update your models. You can train models according to the instructions here
    c. Run the monthly predictions according to the instructions here.
    d. Evaluate your earlier predictions according to the instructions here
13. If you just want to use static datasets you can do so manually and follow the instructions in step 10. If you want to use highly dynamic features like the groundtruth datasets we recommend doing more automation which is explained here
14. After you have set up your preprocessing chain and all the substeps of step 12 you are ready to become operational with your totally personalized FF model.

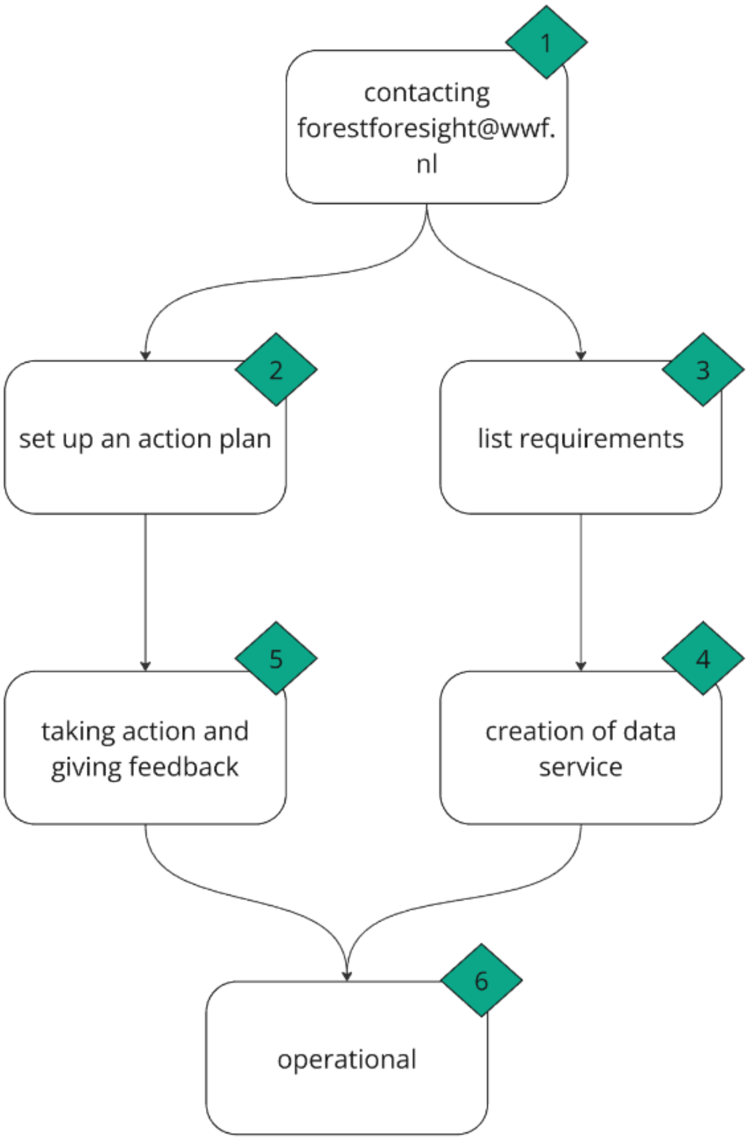
Getting a viewing application (user level 1)



1. First you can get in touch with us about the possibilities of a viewer. You can do so by sending an e-mail to forestforesight@wwf.nl
2. Together we can determine what steps you will take based on our predictions to know what we need for the next step.
3. Based on your needs we will make a list of requirements of what the data should look like and what other datasets you need to be actionable
4. With those needs we will develop the dashboard.

5. With the information of our predictions actions to limit/curb deforestation can be taken and feedback can be given about the dashboard
6. During the operational phase we can work on steps 4 and 5 where needed.

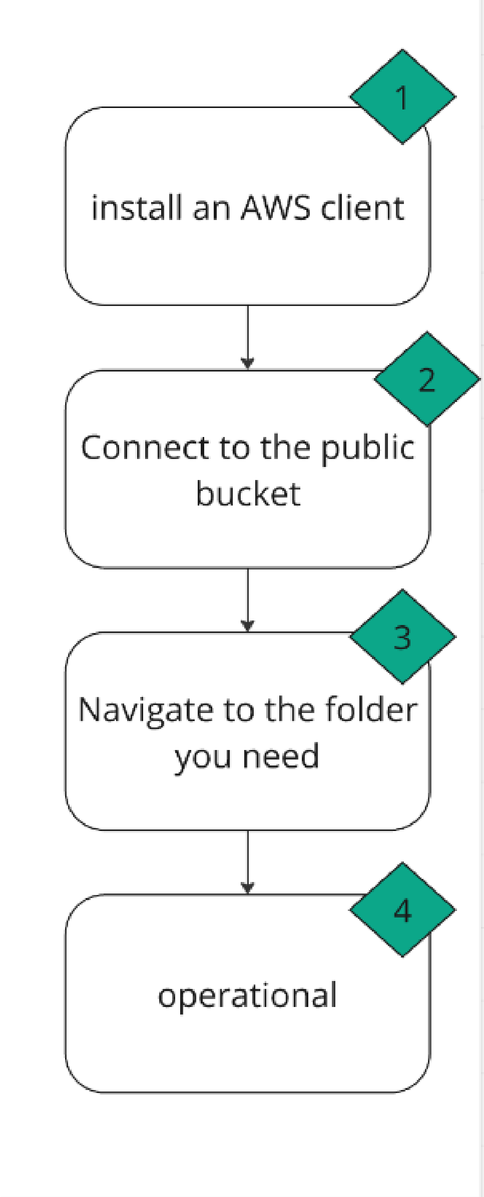
Connecting to ForestForesight dataset services (user level 2)



1. First you can get in touch with us about the possibilities of a data service. You can do so by sending an e-mail to forestforesight@wwf.nl
2. Together we can determine what steps you will take based on our predictions to know what we need for the next step.

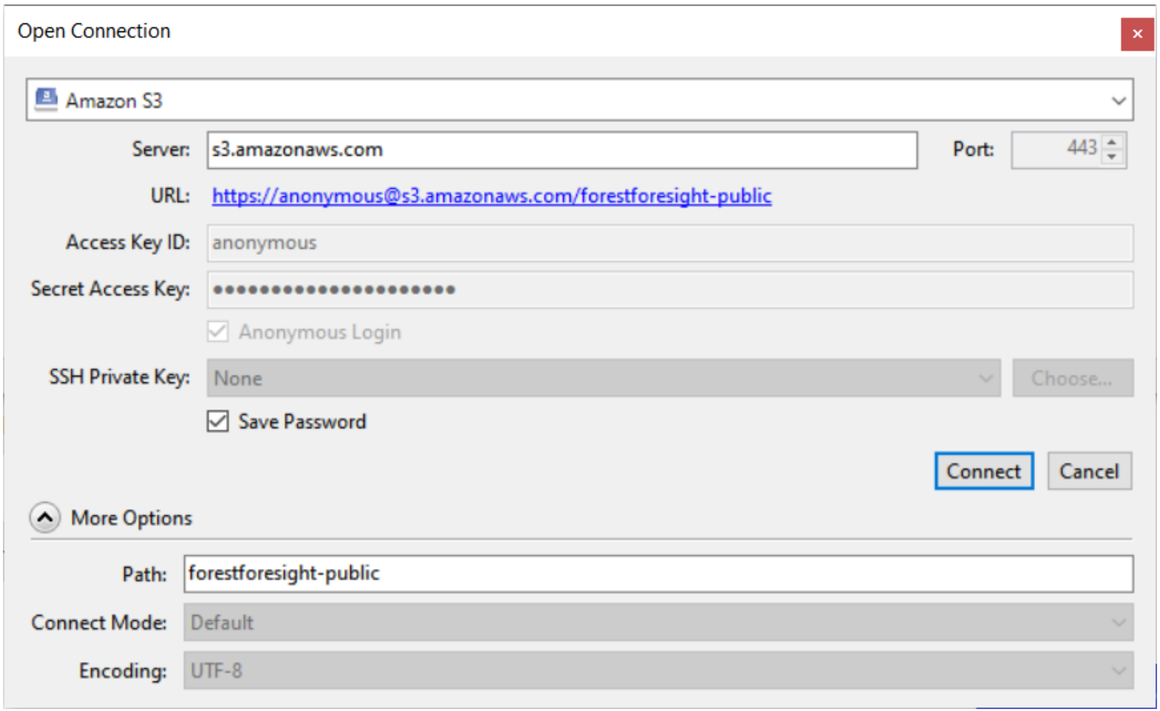
3. Based on your needs we will make a list of requirements of what the data should look like to be actionable.
4. With those needs we will develop data service.
5. With the information of our predictions actions to limit/curb deforestation can be taken and feedback can be given about the data service.
6. During the operational phase we can work on steps 4 and 5 where needed.

Connecting to the ForestForesight data repository



1. There are many FTP clients and other software that can connect to AWS S3 buckets (the buckets we use for data hosting) with an interface. Among them are Cyberduck, Filezilla, S3 browser, WinSCP, CloudBerry and many more. We give in the following steps Cyberduck as an example.
2. To connect to our bucket you need as a bucket name forestforesight-public and you can use anonymous credentials to connect. For cyberduck this looks like

this:



3. You then navigate to the right folder. In the image below you can see the main folders we have. You will probably need the preprocessed (a) data folder to build models yourself or to just get the predictions you go to the Predictions folder (b).
    a. In preprocessed you find tile folders. To know which correspond to the ones you need, have a look [here.](#)
    b. Or the prediction rasters you need the 3 letter iso 3 code of the country to navigate to the right folder. A lookup can be found here.

Attribute	**Preprocessed** Data that goes into the model	**Models** trained models	**Predictions** raster output of FF predictions	**Experimentation** Worthwhile model experimentation	**Accuracy analysis** file with model accuracy
Structure	contains two folders, one for the input data for the model and one for the different groundtruths used to train the model and compare the result. In it there are folders with the GFW-tile identifier (top-left corner of 10x10 degree tiles)	Per country group (grouped by geography and alert type) one folder with one model (the current operational one). The groups can be found in the dataset countries in R-package ForestForesight	Per country (ISO3- country code) a folder containing the predictions per month	No notable structure	A CSV with a unique identifier (linked to dataset degree_polygons in ForestForesight package), Contains the accuracy per known tile per month (in true and false positive and true and false negatives). This file is loaded in our powerbi report
Data type	Geotiffs of 400x400 meters	RDS-files (R-datasets, can be loaded through readRDS in R)	Geotiffs of 400x400 meters	No specific data type	One CSV
update frequency	monthly	irregularly	monthly	irregularly	Monthly

4. Right-click the folders and/or files you want and download to the location you want. For ForestForesight preprocessed data to be used in the ForestForesight package it is absolutely essential that the same data structure is being kept. So

this means a folder called **preprocessed** containing two folders, called **groundtruth** and **input**, with tile folders underneath it.

Setting up the AWS CLI

To connect to the ForestForesight open S3 bucket using the AWS CLI, you don't need to configure AWS credentials. Instead, you can use the `--no-sign-request` option with your AWS CLI commands. Here's how you can do it:

1. Installation
   First, make sure you have the AWS CLI installed on your system. Installation instructions can be found [here.]()
2. Connect to the ForestForesight Bucket

To list the contents of an open S3 bucket:

```
aws s3 ls s3://forestforesight-public/ --no-sign-request
```

To download a file from the bucket:

```
aws s3 cp s3:// forestforesight-public/path/to/file
local/path/to/save --no-sign-request
```

To sync the contents of a bucket to a local directory:

```
aws s3 sync s3://forestforesight-public/ local/directory/ --
no-sign-request
```

To get information about a specific object in the bucket:

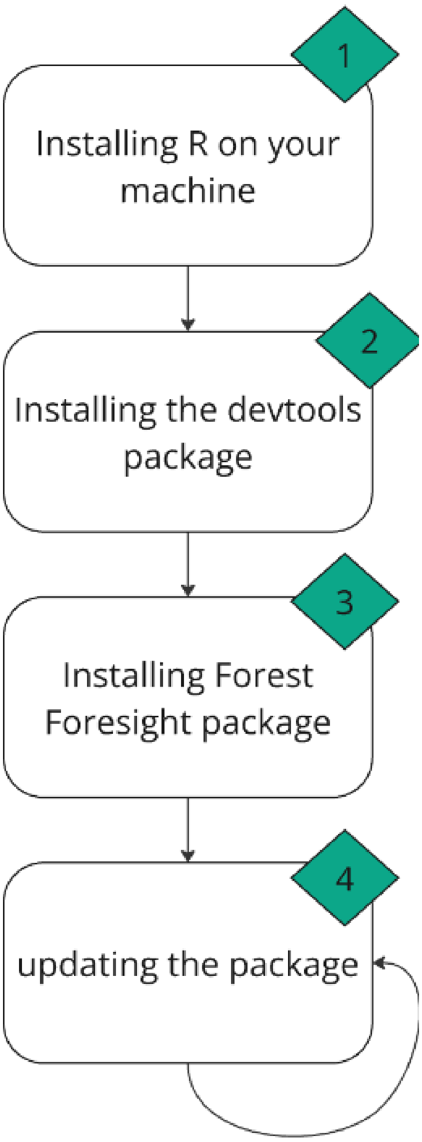
```
aws s3api head-object --bucket bucket-name --key
path/to/object --no-sign-request
```

Note: The `--no-sign-request` option tells the AWS CLI not to sign the request with credentials, which is necessary for accessing the ForestForesight public repository.

We strongly recommend to not sync or copy all the data of ForestForesight to your local computer since it covers the entire world and is in the range of 100-200GB

3. Find the correct folders that you want to sync and download the data you want.

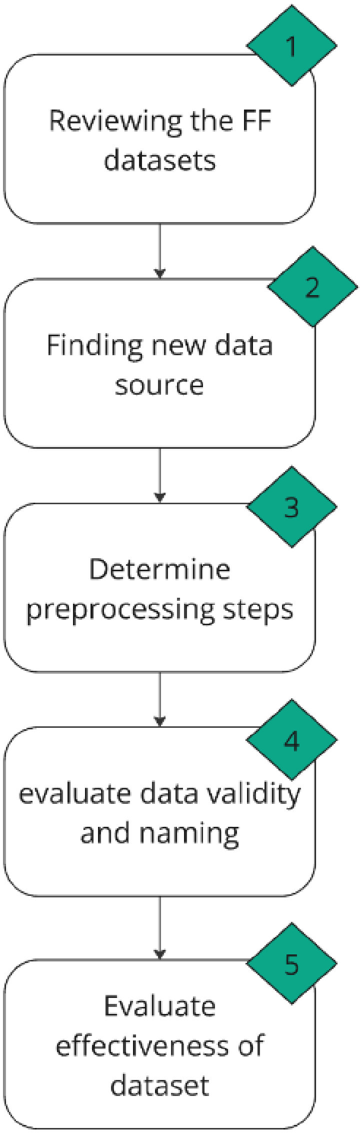
Installing the ForestForesight package



1. First you need R and favorably as an interface Rstudio on your machine. R can be downloaded here, Rstudio can be found here.
2. In the terminal of Rstudio or of R you first need to install the devtools package, which allows you to do step 3 and 4. the line is as follows:
   ```
   install.packages("devtools")
   ```
3. Then run the following line to install ForestForesight:
   ```
   devtools::install_github("jduijvenbode-
   WWF/ForestForesight") #to install the package
   Library(ForestForesight) #to load the package
   ```

4. To know what version you are currently running of ForestForesight you can run the line packageVersion("ForestForesight"). Compare that to the current version of ForestForesight which can be found here. If you are not running the latest version, you can uninstall and reinstall ForestForesight. Uninstalling can be done with the line remove.packages(“ForestForesight”)

Incorporating your own datasets for training



1. Before you decide to include your own datasets it is a good practice to first look at the datasets Forest Foresight already has preprocessed for you. You can find an overview of these datasets here. Things to notice are:
    a. Data quality. FF uses for its global model only global datasets, which might be less recent than your own data (or more of course).
    b. Data preparation. The way FF has preprocessed the data might not match the way you would want to do it. In this case also feel free to contact FF as well to discuss on forestforesight@wwf.nl.

   c. Temporal Relevance. Forest Foresight tries to only incorporate datasets that were already around since 2021 because it gives a larger training period. We have time series for data that changes a lot over time but if you have data that is very valuable but not available since 2021 you might want to opt to include it, resulting in a shorter training period.
2. If the dataset you want to use does not match any that ForestForesight already has you will have to find your own datasets. If they are especially important because they are highly dynamic (updated often) please take attention to the following things:
   a. How hard is the preprocessing? If it is very difficult and you have to do it monthly consider if it is worth your while and so how you can automate the process
   b. Is the resolution of the data suitable compared to predicting on a 400x400 scale? If datasets give you only 3 distinct values for instance across the whole country then it would be less usable.

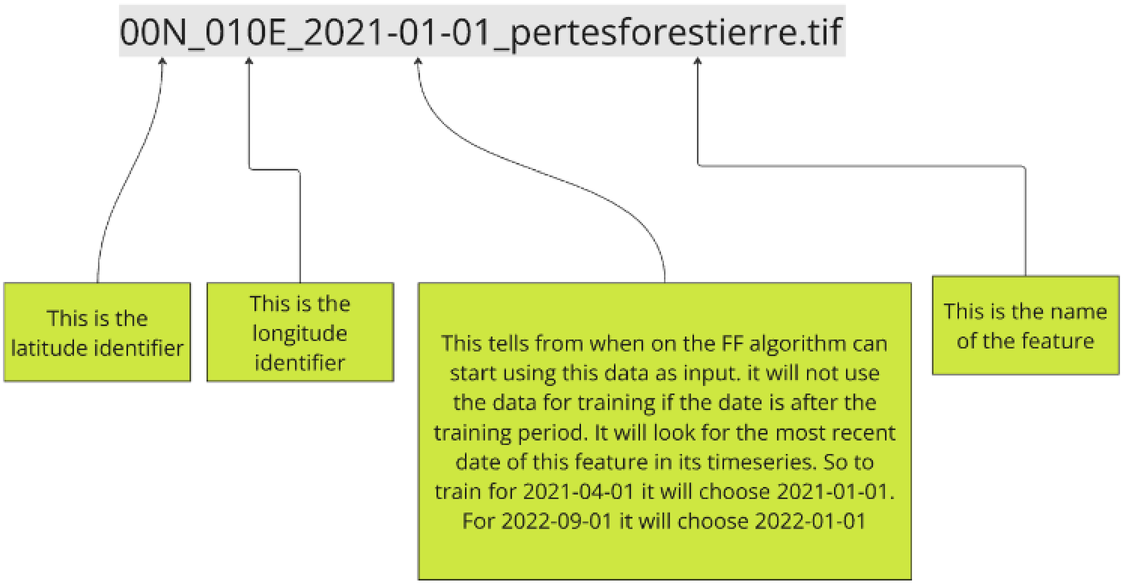
   Other considerations are:

   c. Is the data categorical (meaning that the numbers do not represent a higher or lower value) you have to figure out if they can be converted to non-categorical data. Examples of how to tackle this are:
      i. One-hot encoding, where you make different features of one dataset.
      ii. Finding an average value of deforestation per category. So e.g. if a certain type of forest has a higher chance of deforestation than another, use these relative values instead of class A and B in your dataset.
   d. If the data has a very coarse resolution (of a few meters), how can you resample the data so it still makes sense?
   e. Does it cover the entire area that you are planning to predict for? If not you will have to adjust your training and predicting area to match this dataset.
3. Regardless of the dataset you choose it has to be rebuild into 10x10 degree tiles lat/long on a 0.004x0.004 degree resolution (about 400 meters) with coordinate system WGS84 lat/long and of one band per dataset. Regular steps to be taken for preprocessing are:
   a. Polygon to raster conversion
   b. Distance calculation (for instance roads are converted to distance to roads)
   c. Resampling (from a finer or coarser resolution to the 0.004 degrees)

d. Reprojection (to WGS84 lat long)
e. Multi-band to single band. If your data has multiple bands they need to be split to multiple single-band features
f. Categorical to cardinal data conversion (as explained in point 2c)

We always recommend to use the project and/or rasterize function along with a dataset already from FF in R to make sure that the conversion goes smoothly.

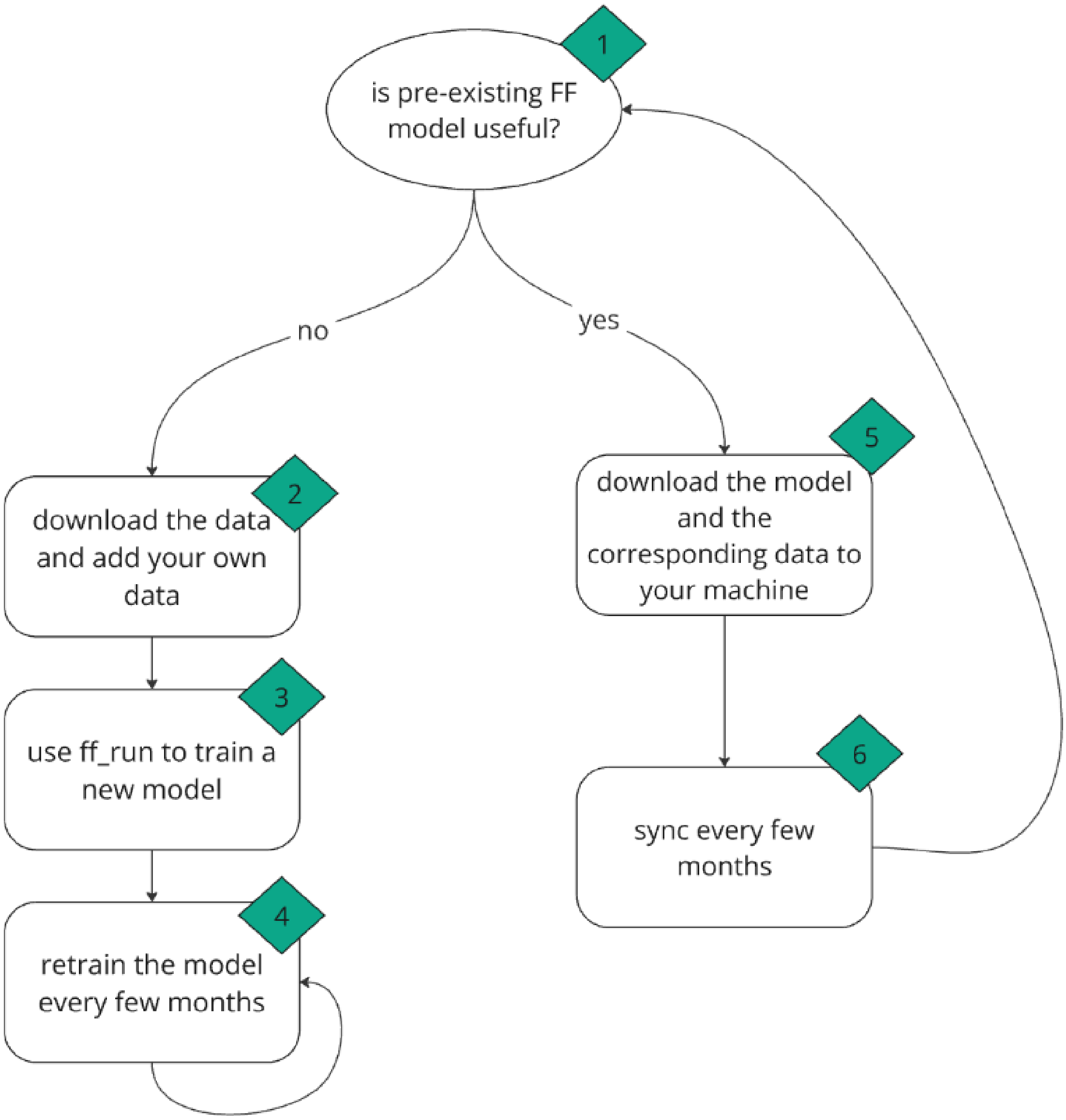
4. When your dataset is prepared it should be stored alongside the already existing datasets of ForestForesight in the preprocessed/input tile folders. The files should follow the naming convention as shown below and should have the geotiff extension.



You can use the function ff_dqc in the forestforesight package to see if your data is adhering to the correct data structure. Use the question mark in r (in this case ?ff_dqc) to see the help text for this function.

5. Be sure to evaluate the effectiveness of your dataset by running the model with and without this dataset (check out the function ff_prep and its parameters to find out how to do so). Comparing the outcoming rasters of ff_predict based on those models is the best way to evaluate your new adapted model. If you find any really valuable features, please let us know on forestforesight@wwf.nl!

Training an FF model



1. Whether the pre-trained models of FF are applicable to you depends on a number of factors. If any of the following situations are applicable to you, you should train your own model:
    a. You add your own features for training
    b. You want to use a different groundtruth dataset
    c. The area you want to process is a lot smaller than the entire country and has distinctively different deforestation drivers

   d. You want to change some hyperparameters (though from experience we have not seen large changes by doing this).

   If this is not the case we recommend using the model that was already built by ForestForesight-NL.
2. Connect to our open data repository and download the data from the tiles of the area you are interested in. For instructions to download the models and data look here.
3. You can now use the ForestForesight package that you already installed (for instructions look here). Make sure the package is loaded by typing library(ForestForesight) . You can then use the function ff_run to load the data into memory, train a model and use that model to make predictions later on. Instructions can be found by using the examples in the help of ff_run (by typing ?ff_run)
4. In ff_run you save the model somewhere and can load it at later dates to process for new months. It is not necessary to retrain every month but it also doesn't hurt so it is up to you to do so. When you add new features we strongly recommend to retrain your model so the features can be included.
5. Connect to our open data repository and download the models from the group that contains the country that you want to process for. For instructions to download the models and data look here. You can find an overview of the groups here to determine which group corresponds to which country. You can find the models under models/{name of group}/{name of group .model and .rda}. These models were prepared by testing many time intervals and evaluating what time interval worked best for these areas. The reason some countries are grouped is because the smaller countries do not have enough groundtruth input to train a model. You have to download both the model file and the rda file. The rda file contains the feature names that were used for the model so that the model knows which were being used as input.
6. ForestForesight-NL updates the models periodically (up to once every month or whenever there are new feature datasets added to the global model). You can either sync or redownload the model you need,

Predicting FF using a built model

Use the function ff_run and a model that was either prebuilt by ForestForesight-NL or by yourself (see here). For instructions on how to use the function ff_run, type ?ff_run in the console and follow the examples and instructions.

Evaluating your FF model

Use the function ff_run and a model that was either prebuilt by ForestForesight-NL or by yourself (see here). Use as prediction dates dates that already have groundtruth data (so at least six months in the past. For these datasets ff_run can analyze the F0.5 score that we use, and you have the option to output a csv where you can see in high spatial and temporal detail the number of True Positives, False Positives and False Negatives. You can analyse this file in Excel, powerBI or any other data analysis software package to know how well your model is performing geographically and over time.