Datasets description

2. Digitalis

The Digitalis database was created in the year 2000 in order to manage the ecological and forestry geographic information layers produced in the framework of the research activities of the forest ecology team. Its objective is to centralize, organize, and make available original spatialized data, over large areas and at the finest possible spatial resolution. These data mainly concern the spatial distribution and, if possible, the evolution in time of the main ecological conditions and of the forest resource.

2.1. pH

The pH map of the A horizon of French forest soils was developed from the bioindicator character of forest plants. It was produced using the bioindicator values for France from the EcoPlant database (Gégout et al. 2005) and floristic surveys (Drapier and Cluzeau 2001). The bioindication made it possible to calculate, for each species, the indicator value which represents the pH value for which the frequency of the species is maximum (Gégout et al. 2003). Using all the species present in a floristic survey, the pH of the plot can be predicted. This technique has been applied to the National Forest Inventory (IGN) surveys, and the predicted values on the forest plots were then interpolated over France (Coudun et al. 2006; Coudun and Gégout 2007). The bioindicator character of forest plant species with respect to pH was first calculated from 3835 plots of the EcoPlant database with a floristic inventory and a laboratory measurement of water pH of the organo-mineral A horizon of the soil. The surface pH of 104,375 NFI plots inventoried between 1989 and 2004 was then estimated by averaging the indicator values of the species present on each site. A kriging interpolation method was used to estimate the values between plots and create a surface pH map of French forest soils with a resolution of 1 km^2. The quality of the map was evaluated from 261 plots systematically sampled in France and not used to calculate the indicator values or to elaborate the pH map. The mean difference between the map pH and the measured pH on these plots is equal to 0.81 and the square of the correlation coefficient between predicted and measured pH values (R^2) reaches 0.58.

2.2. Temperatures (T)

Average monthly temperature values in forest areas (°C), calculated for winter (Dec, Jan, Feb), summer (June, July, Aug.) and the whole year, for the reference period 1961-1990 (Ninyerola et al. 2000; Richard 2011; Bertrand et al. 2011). The 50-m resolution maps were produced from a set of 237 Météo-France measurement stations, covering the period 1961-2010. Temperatures were spatialized using statistical models developed with topographic (altitude, exposure, solar radiation, ...), geographic (distance to different ocean masses) and dominant land use variables, whose spatial distribution is known in a relatively accurate way. The residuals (the difference between the measured and modeled values) were then interpolated and added to the map resulting from the model. Due to the inclusion of land cover in the model, the mapped temperatures correspond to the conditions observed at the top of the canopy in a forest environment, which are cooler than in an open environment. These data are derived from models with inaccuracies that are not homogeneous in space. At the scale of France, a validation was made for the period 1996-2007 from measurements on 493 stations, we obtain an R^2 of 0.93 on the annual data with an RMSE of 0.54°C (for values varying mainly between -10 and +16°C). The worst model is obtained in June ($R^2 = 0.91$) and the best in March ($R^2 = 0.94$).

2.3. Precipitation (P)

Cumulative precipitation value (mm) for the summer (June, July, Aug.) or the whole year, averaged over the period 1961-1990 (Ninyerola et al. 2000; Richard 2011). The maps were produced from a set of 435 Météo-France measurement stations. Precipitation values were mapped by 1-km pixels according to techniques described by Ninyerola et al (2000). Statistical models were developed using variables whose spatial distribution is well known, characterizing the topography (degree of exposure to prevailing winds, slope, etc.) and geographical position (distance to different ocean masses). The residuals were then interpolated and summed to the model map. The values were summed for the periods considered. These modeled data present inaccuracies not homogeneous in space. At the scale of France, a validation was made for the period 1996-2007 from measurements on 471 stations, we obtain an R^2 of 0.78 on the annual data with an RMSE of 136 mm (for values varying mainly between 300 and 2400 mm). The worst model is obtained in December ($R^2 = 0.66$) and the best in July ($R^2 = 0.85$).

2.4. Solar radiation (SR)

Solar radiation (in J/cm^2) is calculated using the Helios model. This model combines the effects of direct, diffuse and reflected radiation (Lebourgeois and Piedallu 2005; Piedallu and Gégout 2008; Richard 2011). It takes into account topographical effects (slope, exposure, mask effects) and large-scale effects (latitude and cloud cover), which vary with the date. The values provided are monthly averages for the summer (June, July, Aug.) or the whole year, calculated for the period 1971-2000. Solar radiation was simulated under clear skies for each month using the IGN BD alti® digital terrain model, at a 50-m resolution, then aggregated by 1 km cells. The cloud cover data were modeled from a dataset collected from 87 meteorological stations and integrated into the radiation model. These data are derived from models with non-homogeneous inaccuracies in space. At the scale of France, a validation on 88 measurement stations of Météo-France gives an R^2 of 0.82 with an RMSE of 227 MJ/m^2 (for a data varying mainly between 2500 and 6800 MJ/m^2). The best predictions are obtained in winter ($R^2 = 0.89$ for January and December), the lowest values being obtained in summer. We have an R^2 of 0.82 for March, 0.62 for June and 0.73 for September.

2.5. Evapotranspiration (ET)

Average monthly value of potential evapotranspiration for the summer (June, July, Aug.) or the whole year, calculated over the period 1961-1990 (ET, in mm). ET represents the theoretical water demand under conditions of non-limiting water availability, combining soil and canopy evaporation and vegetation transpiration. The calculation of ET was performed for each month using the Turc formula (Turc 1955, 1961; Lebourgeois and Piedallu 2005), which combines temperature and solar radiation values. Solar radiation values were calculated using the Helios model (Piedallu and Gégout 2007, 2008) and forest temperature data were spatialized using methods combining modeling with topographic (elevation, exposure, solar radiation), geographic (distance to different ocean masses), and dominant land use covariates (Bertrand et al. 2011). The map has a resolution of 50 m. These data are derived from models with non-homogeneous inaccuracies in space. At the scale of France, there is no validation of the ET but a validation of the elementary variables constituting them, the temperatures and the solar radiation.

2.6. Climatic water budget (CWB)

Average monthly value of the climatic water budget for the summer (June, July, Aug.) or the whole year, calculated over the period 1961-1990 (in mm of water) at 50-m resolution. It represents the amount of rainfall available to plants, once evaporation and transpiration needs have been met. It is calculated by the difference between precipitation (P) and potential evapotranspiration (ET) estimated according to the Turc method (Turc 1961). The elementary components (precipitation, temperature and solar radiation needed to calculate the Turc ET) are modeled according to a methodology specific to each of them (Piedallu and Gégout 2007, 2008), for each month of the year, before being assembled for the period under consideration and validated with an independent data set. Each component of the climate water budget has been validated separately with an independent data set, in order to determine the quality of the predictions at the scale of France. The calculation of the climatic water budget is a very simplified estimate of the water available for plants, which does not take into account many parameters, the most important of which is the soil's capacity to store water. As a result, water deficits are exaggerated compared to reality. On the other hand, the basic data are derived from models with non-homogeneous inaccuracies in space, which can accumulate during their assembly.

2.7. Maximum available water capacity (MAWC)

The maximum available water capacity (MAWC) represents the maximum quantity of water that a soil can contain. It is conditioned by many parameters such as texture, root exploration depth, pebble load and bulk density. It is estimated from soil surveys carried out in the field, usually from a pit. This parameter is an input component to the calculation of the available water capacity (AWC), which represents the water actually available in the soil. The map of MAWC of forest soils was produced using data from the IGN forest inventory (Drapier and Cluzeau 2001). Thus 100307 plots were used for the construction of the map, 20596 for its validation and 3762 to test the relevance of the indices produced to predict the growth of some species (Piedallu et al. 2011). The MAWC was estimated on those plots for which soil descriptions exist which are limited to the first meter and can be dissociated into two horizons maximum. The MAWC was calculated from texture data using Al Majou's pedotransfer classes (Al Majou et al. 2008), and an estimate of prospectable soil volume that was made using depth, pebble load, and rock outcrop data. Spatialization of this information was done by modeling using geological and topographical variables, and adding interpolated residuals. The quality of the map was evaluated using a set of 20595 plots. We obtain an R^2 of 0.35 and a mean estimation error of 34 mm for a value varying between 0 and 150 mm. This relatively low value of R^2 is partly explained by scale problems, the map prediction being provided by 500-m pixels, whereas the validation plot gives information at the pit scale, knowing that the MAWC can vary greatly locally.

2.8. Available water capacity (AWC)

The available water capacity (AWC expressed in mm) is an index derived from the calculation of the soil water balance that represents the quantity of water available to plants for a given period (Thornthwaite, 1955). It is mainly dependent on precipitation inputs, soil capacity to store water, vegetation cover and soil evaporation. This amount of water available to plants changes during the year according to water inputs and outputs, within the limit of the maximum available water capacity. The available water capacity was calculated monthly for the average values of the period 1961-1990 using the Thornthwaite formula (Thorthwaite and Mather 1955). The computation of AWC requires combining for each month temperature and solar radiation data in order to calculate the evapotranspiration with the Turc formula (Turc 1961), as well as precipitation and maximum available water capacity values in order to estimate the available water stock. The models for these different data use a methodology specific to each of them (Piedallu and Gégout 2007, 2008; Bertrand et al. 2011; Piedallu et al. 2013). The climate models (rainfall, temperature, solar radiation) have the characteristic of trying to take into account as much as possible the topographic effects, including local ones. The MAWC model incorporates soil depth, texture, and pebble load. Maps were produced for each of the variables using these models, at a 50-m resolution for temperature and solar radiation, at a 500-m resolution for MAWC and at a kilometer resolution for precipitation. Water budgets were calculated at a 50-m resolution, and then the values were aggregated to the kilometer resolution. The values provided correspond to monthly averages for the summer (June, July, Aug.) or the whole year, calculated for the period 1961-1990 at a 50-m resolution. The data provided were calculated using models with spatially inhomogeneous inaccuracies. It is not possible to validate the available water capacity due to the lack of independent measurements.

3. GISSOL

The soil quality measurement Network (RMQS) is a national program for the evaluation and long-term monitoring of French soil quality (Jolivet et al. 2018). This network is based on the monitoring of 2240 sites representative of French soils and their occupations, distributed over the entire French territory (metropolitan France and overseas) according to a systematic grid of 16 km sides. The sites cover various occupations (field crops, permanent grasslands, forests, vineyards and orchards, less anthropized environments, urban parks). Physical, chemical and biological properties of the soils are measured on each site with a periodicity of about fifteen years. These analyses are associated with the search for explanatory factors for the spatial and temporal variability of soil properties (biophysical variables, sources of contamination, history of occupation and management practices at each site). The first sampling campaign in mainland France took place from 2000 to 2009. This campaign, which focused on soil contamination, made it possible to map the main soil parameters (28 variables) as well as the levels of 12 trace metals (TMEs) in total or partial extraction and 70 persistent organic pollutants.

We used the results of analyses of 2146 sites in mainland France, implemented during the first RMQS campaign (2000-2009) (RMQS 2019). The analyses were performed on composite samples taken by auger according to 2 sampling layers (0-30 cm or thickness of the cultivated soil layer called surface composite or composite 1 and the underlying layer up to 50 cm, called sub-surface composite or composite 2). Each composite sample was made up of a mixture of 25 individual samples taken from a 400 m² sampling area using a stratified random sampling design. A third layer of composite samples could be collected in forest or grassland, from the forest floor (corresponding to the O soil horizons) when these horizons were sufficiently thick (at least 1 cm) and continuous over the sample area. In accordance with the legal framework currently in force, the dataset provided includes the coordinates of the center of the grid cell, called centroid coordinates or theoretical coordinates.

4. Geology

We used the geological map at 1:50,000 scale in vector format from the Geological and Mining Research Bureau (Bureau de Recherches Géologiques et Minières BRGM) (BGRM 2017). We selected the fields: lithology, land/sea, geochemistry, system, age of the geological stratum, nature of the stratum, geochemistry, and simplified lithology.

5. IGN

5.1. BD ALTI®

We used the gridded digital terrain model at 25 m BD ALTI® version 2 from the National Geographic Institute (IGN) to obtain the features: altitude, aspect, slope, hillshade and roughness (IGN 2017). The aspect, the slope, the hillshade and the roughness were computed with the corresponding QGIS functions and the variable values Z factor=1, azimuth (horizontal angle)=300 and vertical angle=40.

5.2. BD TOPO®

The BD TOPO® is a 3D vectorial description (structured in objects) of the elements of the territory and its infrastructures, of metric precision, exploitable at scales going from 1 : 2 000 to 1 : 50 000 produced by the National Geographic Institute (IGN) (IGN 2022). We used the following data:

- Public forest: In France, a distinction is made between private and public forests. These public forests are managed by the National Forest Office (ONF). These forests can be owned by the state or by the city.

- Park or nature reserve: Park or nature reserve enclosure where certain activities are governed by specific regulations. It is broken down into national park, regional nature park, nature reserve, national hunting and fauna reserve, each with its own regulations.
- Roads: All roads for automobiles, pedestrians, bicycles or animals, paved or unpaved, are included. Public and private roads are not distinguished. Using QGIS we performed proximity analysis and obtained a 10-m raster with the distance between the pixel and the nearest road.
- Population: For each city, we extracted the administrative center and the population. Using QGIS and the location of the administrative center with the associated population, we performed an inverse distance weighting interpolation with a distance coefficient of 2 at a 1-km resolution.

6. References

Al Majou H, Bruand A, Duval O, et al (2008) Prediction of soil water retention properties after stratification by combining texture, bulk density and the type of horizon. Soil Use Manag 24:383–391. https://doi.org/10.1111/j.1475-2743.2008.00180.x

Bertrand R, Lenoir J, Piedallu C, et al (2011) Changes in plant community composition lag behind climate warming in lowland forests. Nature 479:517–520. https://doi.org/10.1038/nature10548

BGRM (2017) Infoterre

Coudun C, Gégout J-C (2007) Quantitative prediction of the distribution and abundance of Vaccinium myrtillus with climatic and edaphic factors. J Veg Sci 18:517–524. https://doi.org/10.1111/j.1654-1103.2007.tb02566.x

Coudun C, Gégout J-C, Piedallu C, Rameau J-C (2006) Soil nutritional factors improve models of plant species distribution: an illustration with Acer campestre (L.) in France. J Biogeogr 33:1750–1763. https://doi.org/10.1111/j.1365-2699.2005.01443.x

Drapier J, Cluzeau C (2001) La Base de données écologiques de l'IFN. Rev For Fr 53:365. https://doi.org/10.4267/2042/5251

Gégout J-C, Coudun C, Bailly G, Jabiol B (2005) EcoPlant: A forest site database linking floristic data with soil and climate variables. J Veg Sci 16:257–260. https://doi.org/10.1111/j.1654-1103.2005.tb02363.x

Gégout J-C, Hervé J-C, Houllier F, Pierrat J-C (2003) Prediction of forest soil nutrient status using vegetation. J Veg Sci 14:55–62. https://doi.org/10.1111/j.1654-1103.2003.tb02127.x

IGN (2017) BD ALTI® version 2

IGN (2022) BD TOPO®

Jolivet C, Boulonne L, Ratié C (2018) Manuel du Réseau de Mesures de la Qualité des Sols. Unité InfoSol, INRA, Orléans, France

Lebourgeois F, Piedallu C (2005) Appréhender le niveau de sécheresse dans le cadre des études stationnelles et de la gestion forestière à partir d'indices bioclimatiques. Rev For Fr 57:331. https://doi.org/10.4267/2042/5055

Ninyerola M, Pons X, Roure JM (2000) A methodological approach of climatological modelling of air temperature and precipitation through GIS techniques. Int J Climatol 20:1823–1841. https://doi.org/10.1002/1097-0088(20001130)20:14<1823::AID-JOC566>3.0.CO;2-B

Piedallu C, Gégout J (2008) Efficient assessment of topographic solar radiation to improve plant distribution models. Agric For Meteorol 148:1696–1706. https://doi.org/10.1016/j.agrformet.2008.06.001

Piedallu C, Gégout J-C (2007) Multiscale computation of solar radiation for predictive vegetation modelling. Ann For Sci 64:899–909. https://doi.org/10.1051/forest:2007072

Piedallu C, Gégout J-C, Bruand A, Seynave I (2011) Mapping soil water holding capacity over large areas to predict potential production of forest stands. Geoderma 160:355–366. https://doi.org/10.1016/j.geoderma.2010.10.004

Piedallu C, Gégout J-C, Perez V, Lebourgeois F (2013) Soil water balance performs better than

climatic water variables in tree species distribution modelling. Glob Ecol Biogeogr 22:470–482. https://doi.org/10.1111/geb.12012
Richard J-B (2011) Caractérisation de la contrainte hydrique des sols à l'aide de cartes numériques pour prendre en compte les effets potentiels du changement climatique dans les catalogues de stations forestières : Applications aux plateaux calcaires de Lorraine, Champagne-Ardenne et Bourgogne. AgroParisTech
RMQS (2019) GISSOL
Thorthwaite CW, Mather RJ (1955) The Water Balance Publication. In: Climatology. Drexel Institute of Technology, Laboratory of Climatology, Centerton, New Jersey
Turc L (1955) Le bilan d'eau des sols : relations entre les précipitations, l'évaporation et l'écoulement. Journ Hydraul 3:36–44
Turc L (1961) Evaluation des besoins en eau d'irrigation et évaporation potentielle. Ann Agron 12:13–49