

# ECOFI: a database of sugar and energy cane field trials

Mathias Christina<sup>1,2</sup>; Maxime Chaput<sup>1,2,3,4</sup>; Jean-François Martiné<sup>1,2</sup> ; Sandrine Auzoux<sup>1,2</sup>

<sup>1</sup> CIRAD, UPR 115 AIDA, Saint-Denis, La Reunion

<sup>2</sup> UPR AIDA, Univ Montpellier, Montpellier, France

<sup>3</sup> CIRAD, UPR 78 Recyclage & Risques, Saint-Denis, La Reunion

<sup>4</sup> UPR Recyclage & Risques, Univ Montpellier, Montpellier, France

\* email: mathias.christina@cirad.fr

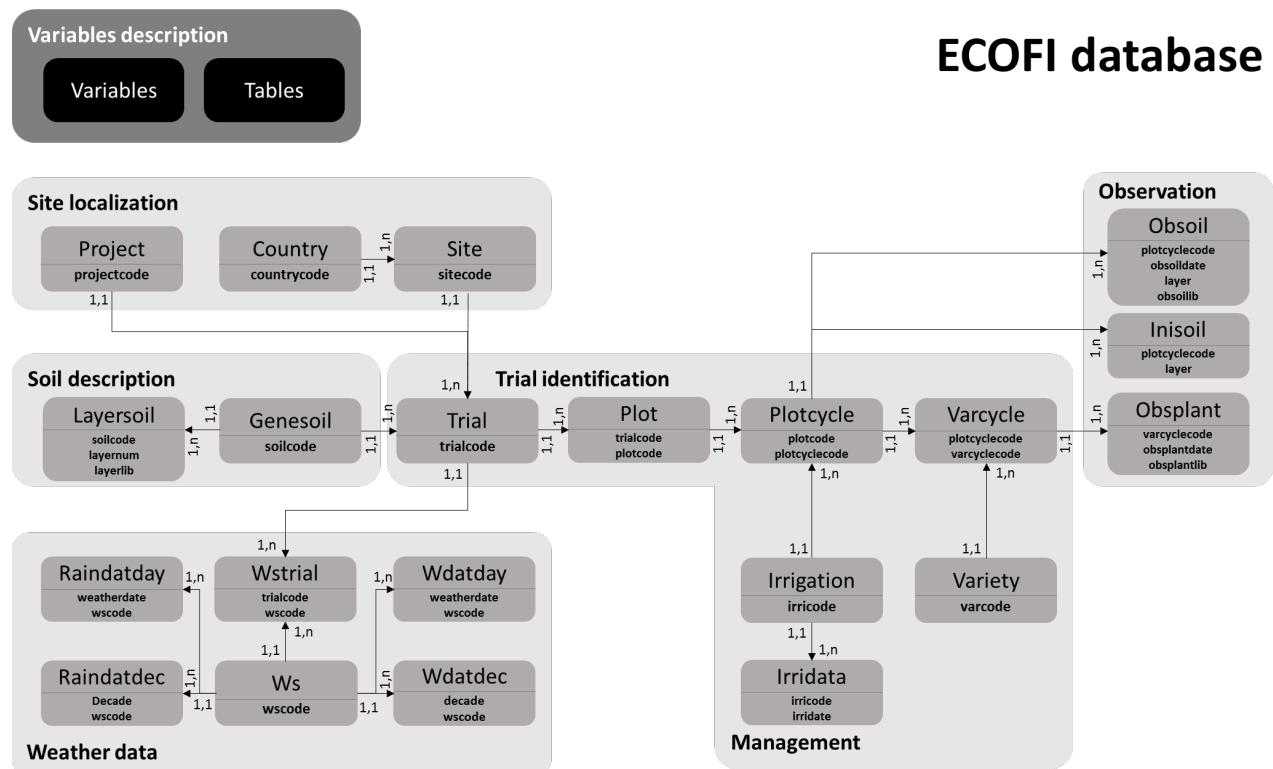
**Abstract:** Agroecological studies on sugarcane dealing with genotype, by environment and by management interactions commonly generate complex datasets. To facilitate the use of these datasets, a relational database, named ECOFI, was designed from the analysis of the content and the structure datasets of multidisciplinary experiments with sugarcane and energy cane. The database described in this paper includes data from 58 trials carried out in 11 countries from 1986 to 2016, including 24 trials in Reunion Islands and 15 in Guadeloupe. Taking into account plots within the trials and crop cycle, it includes 725 crop cycles in total, with 60 different cultivars. The datasets contain data for crop observations (e.g. dry mass), soil (water contents), weather (all essential meteorological parameters) and management (sowing, cultivars and harvest). Additionally the datasets contain metadata to qualify observations. This dataset provides an adequate experimental set to calibrate or validate crop model simulations under genotype x environment interaction.

**Keywords:** agro-ecosystems modeling, crop-weather-management dataset, energy cane, field experiments, sugarcane

**1 BACKGROUND AND ORIGINAL PURPOSE:** This article presents the result of a collaborative effort by a multidisciplinary team of computer scientists, biostatisticians, agronomists, weed experts and crop modelers who have been confronted with data and metadata management issues in agroecological research on sugarcane. Relational databases are widely used in agriculture to manage field experimental data as well as pest and weed management information (Bechini and Stöckle 2007). The ECOFI database (experimental database on ecophysiological and environmental observations) has been developed (Auzoux et al., 2017) to solve the problem of security and persistent storage of data, to standardize the annotation of heterogeneous data, to improve analysis and facilitate access to data for sugarcane modelling in Reunion and in other sugar-producing countries in Africa and Central America. First, the database was built with the aim of providing simulation data to the sugarcane growth model MOSICAS (Martiné et al., 2007). The datasets are divided into several text files which are freely available under [doi:10.18167/DVN1/1GCL8F](https://doi.org/10.18167/DVN1/1GCL8F) and located in the CIRAD's dataverse, an online repository with shared data.

**2 STRUCTURE OF THE ECOFI DATASETS:** Unlike a traditional database that is modeled as a function of the problem, ECOFI database can manage all experimental data collected through agroecological studies that take account of many factors varying in time and space. The database consists of 23 tables linked to each other (Figure 1). The description of the different tables is presented in Table 1 to facilitate data access.

**Figure 1.** An ECOFI global database schema on cane crop agricultural systems showing links between trials, soils, climate and observations. Primary keys are indicated in bold.



### 3 EXPERIMENTAL DATA

**3.1 EXPERIMENTAL SITE:** The ECOFI database describes 63 trials carried out in 11 countries: Reunion Island (28), Guadeloupe (15), Brazil (4), Morocco (3), Australia (2), Zimbabwe (2), USA (2), South Africa (2), Ecuador (1), Senegal (1) and Mauritius (1). Each trial is associated with a localized site (latitude, longitude and altitude), country and project. The different treatments of each trial are defined in the “Plot” table and the crop cycle is defined in “Plotcycle” in case of multiple-year and multiple-variety trials.

**3.2 WEATHER AND SOIL DATA:** Weather and soil data are described in each site. The weather data are given at a daily time step or averaged every 10 days. Available meteorological data are: minimum, mean and maximum temperature; minimum, mean and maximum relative humidity; total and maximum wind speed; global radiation; potential evapotranspiration (Penman-Monteith); rainfall. In some case, some meteorological data are missing when not measured (in particular for relative humidity). The database made it possible to associate two set of rainfall data for a specific site (one in “Wdat” and the other in “Raindat”). The variable units are described in “Tables”. Soil data are described in “Genesoil” and “Soillayer” (defining multiplied soil layers) and limited to variables used for the soil water balance assessment: soil water capacity, field capacity, wilting point, saturation, bulk density.

**Table 1.** Dictionary of tables of the ECOFI database

Table name	Description
<i>Variables</i>	Library of observed variables with the variable code, unit and description in french or english
<i>Project</i>	Project associated with each trial (name, summary and coordinator)
<i>Country</i>	Country code description
<i>Site</i>	Trial localization: altitude, latitude, longitude and country
<i>Trial</i>	Trial description: code, starting and ending dates of the trials, soil code, root depth
<i>Plot</i>	Plot description within the trial: plot code, plot area, block, modality, bud density and row spacing
<i>Plotcycle</i>	Plot cycle code within the trial: crop cycle (plant or ratoon crop), initial soil volume, irrigation code
<i>Varcycle</i>	Plot cycle description: cane variety, starting and ending date of the cycle
<i>Wstrial</i>	Weather station code associated with the trial. A weighted parameter allows to decide if the rain data has to be read in "Wdatday" or "Raindatday"
<i>Ws</i>	Weather station description: name, localization and type of station
<i>Wdatday</i>	Daily weather data: minimum, mean and maximum temperature; minimum, mean and maximum relative humidity; total and maximum wind speed; global radiation; potential evapotranspiration; rainfall
<i>Wdatdec</i>	Average weather data per periods of 10 days
<i>Raindatday</i>	Additional daily precipitation data if obtained from a different source
<i>Raindatdec</i>	Additional average precipitation per periods of 10 days
<i>Irrdata</i>	Irrigated water amount and date of application associated with "Plotcycle"
<i>Irrigation</i>	Irrigation type and efficiency
<i>Variety</i>	Code variety and name
<i>Genesoil</i>	Soil description: code, type of soil, number of soil layers, water capacity, soil depth and p0 parameter for one layer soil water balance model
<i>Layersoil</i>	Soil layer description: bulk density, depth of the layer, humidity at field capacity, saturation and permanent wilting point
<i>Obsplant</i>	Plant observations: date and variable observed. Description of observed variables are described in "Variables"
<i>Obssoil</i>	Soil observation: layer identification, date of observation
<i>Inisoil</i>	Soil water humidity at the beginning of the plot cycle

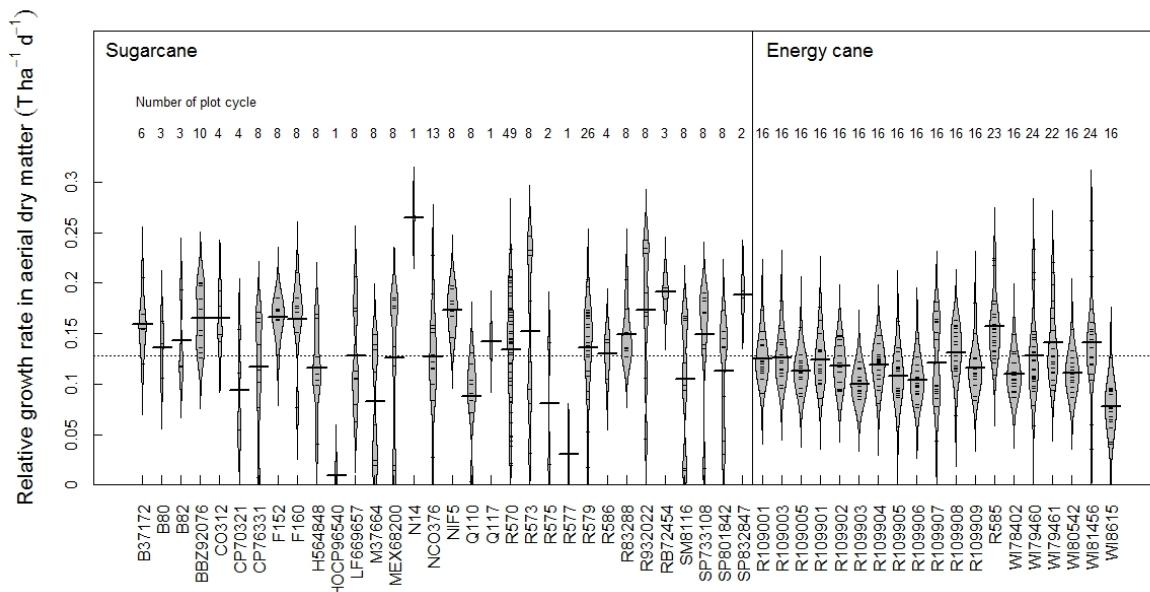
**3.3 MANAGEMENT DATA:** Management information of each trial is detailed in "Plot" (planted area of each plot, number of buds and row spacing), "Plotcycle" (crop cycle: plant or ratoon crops, irrigation management) and "Varcycle" (planted variety, starting and ending dates of the cycle). If not rainfed, the irrigation management are detailed in "Irrdata" and "Irrigation" where daily water amount at specific dates is indicated as well as the efficiency of the irrigation (taking into account water loss). In case the irrigode is "NC", it means that the crop is irrigated but that the irrigation amount and frequency is not available. Two types of irrigation were used in the different trials (sprinkler or drip irrigation). In total, 60 different varieties were used in the different trials (41 sugarcane and 19 energy cane). Nevertheless, the number of crop cycle for each variety is heterogeneous, with a maximum of 95 crop cycle for the R570 variety.

**3.4 PLANT AND SOIL OBSERVATION:** The ECOFI database provides measurements in the different trials at various levels: plot (e.g. soil water content and aboveground dry matter), plant (e.g. aboveground fresh matter per plant), stem (e.g. stem sugar content), leaf (e.g. fresh matter of leaves) or limb level (e.g. width of leaf blades). All measured plant, soil and weather variables are described in the "Variables" table with their respective unit, type and description.

**3.5 ADDING NEW DATA:** In the event that users would like to add observations, the new trial must be created in “Trial” with its location in the table “Site”. The elementary plots within the trial are described in “plot” with potentially different planting densities and row spacing. Finally, “Plotcycle” and “Varcycle” describe the cane cycle and the variety planted. Observations can then be added to “Obsplant” and “Obssoil” after having specifying the new definitions in the metadata table “Variables”, if required. The observed variables are standardized as much as possible using terminology from controlled vocabulary in order to make the data interoperable. If the information is available, the user can also associate soil and climate data with the new dataset.

**4 DATA APPLICATION:** The ECOFI database present various uses for meta-analyze and cropping system models: data exploration, data management and model calibrations. A strong point of the database in data mining is the large number of varieties included. In Figure 2 we present an example with the growth differences in dry mass between varieties obtained from the database. ECOFI could be used to further explore the effects of the interaction between climate and variety, or with management (irrigation, row spacing...). The database has thus been used in the context of projects on sugarcane (e.g. AGMIP) or energy cane (e.g. SYPECAR).

**Figure 2.** Growth rates in aerial dry matter ( $T \text{ ha}^{-1} \text{ d}^{-1}$ ) depending on the cane variety. Growth rate was calculated from the database as the aerial dry matter divided by the number of days since the beginning of the crop cycle. Data are presented using a beanplot with median (black line), normal density (gray background) and populations (small black lines). The horizontal dashed line indicated the overall median. The number of plot cycle available for each variety with aerial dry matter is indicated.



## REFERENCES

- Auzoux, S., Martiné, J.F., Loison, R., Poser, C., Marnotte, P., Goebel, F.R., 2017. "ECOFI: A new generic database to analyse complex agroecological experimentation". International Sugar Journal, 119(1419): 214-218. url: <https://agritrop.cirad.fr/584134>
- Bechini, L., Stöckle, C.O, 2007. "Integration of a cropping systems simulation model and a relational database for simple farm-scale analyses". Agronomy Journal 99:1226-1237. doi: [10.2134/agronj2006.0305](https://doi.org/10.2134/agronj2006.0305)
- Chaput, M., Christina, M., Versini, A., Février, A. and Soulié, J.C., 2019. "Modeling the impact of soil and climatic variability on sugarcane growth response to mineral and organic fertilizer". Poster session presented at: XXX International Sugar Cane Congress ISSCT 2019, September 2-5, Tucuman, Argentina. doi: [10.13140/RG.2.2.16150.50242](https://doi.org/10.13140/RG.2.2.16150.50242)
- Christina, M., Versini, A., Fevrier, A., Mansuy, A., Auzoux, A., 2018. "Modeling sugarcane growth and yield using STICS model, parameterization and applications in complex agro-systems". Poster session presented at: ISSCT Agricultural Engineering, Agronomy and Extension Workshop, September 23-28, Saint-Gilles, La Réunion. doi: [10.13140/RG.2.2.29572.27525](https://doi.org/10.13140/RG.2.2.29572.27525)
- Christina, M., Le Mezo, L., Mezino, M., Barau, L., Tendero, A., Auzoux, S., Todoroff, P., 2019. "Modeling the annual sugarcane yield variability in Reunion Island". Paper presented at: XXX International Sugar Cane Congress ISSCT 2019, September 2-5, Tucuman, Argentina. url: <https://agritrop.cirad.fr/593441/>
- Goebel, F.R., Chopart, J.L., Poser, C., Braconnier, S., Martiné, J.F., Gérardeaux, E., 2017. "Diversifying biomass uses through new cropping systems". In : Sustainable development and tropical agri-chains. Biénabe Estelle (ed.), Rival Alain (ed.), Loeillet Denis (ed.). Dordrecht : Springer, 187-200. ISBN 978-94-024-1015-0. doi: [10.1007/978-94-024-1016-7\\_15](https://doi.org/10.1007/978-94-024-1016-7_15)
- Martiné, J.F. 2007. "Analysis and forecasting of the sucrose content of sugarcane crops during the harvest period in Reunion Island". In: 26th Congress of the International Society of Sugar Cane Technologists, Durban, South Africa, pp. 607–612. url: <https://agritrop.cirad.fr/540854>
- Martiné, J.F., Poser, C., Roussel, C., Gérardeaux, E., Chopart, J.L., 2016. "Modelling energy can growth in contrasting environments: calibration and validation". In : Proceedings of the 29th Congress of the International Society of Sugar Cane Technologists ISSCT Congress 29, Chiang Mai, Thailand. url: <https://agritrop.cirad.fr/582778>.