Thibault Schowing

Th.schow@gmail.com

Résumé

Structure des données fournies, information générale sur le code et les différentes classes utilisées, description du Dataset.

Annexe a - Codebook

Travail de Bachelor 2017

Table des matières

[Données climatiques – extrapolations 2](#_Toc484521684)

[Données climatiques – Librairie 3](#_Toc484521685)

[Données de dégustation de cafés 4](#_Toc484521686)

[Données de qualité de sol 6](#_Toc484521687)

# Données climatiques – extrapolations

Pour les mesures suivantes : **Température** (maximum, minimum, moyenne et DTR) et les **précipitations**, les données ont été extrapolées sur une grande zone du territoire et sont disponible pour chaque mois des années 2011 à 2016. La zone fait environ 300 kilomètres carrés. Les données ont été extrapolées à partir de mesures régulières de différentes stations météorologiques disséminées dans la zone géographique.

|  |  |
| --- | --- |
| Nom de la colonne | Description (valeur par défaut) |
| NCOLS | Nombre de colonnes (720) |
| NROWS | Nombre de lignes (720) |
| XLLCORNER | Longitude, coin inférieur gauche [Degré decimal]  (-77.50042) |
| YLLCORNER | Latitude, coin inférieur gauche [Degré decimal]  3.500417 |
| CELLSIZE | Taille des cellules sur la carte [Degrés decimal]  (0.00416666700000001) |
| NODATA\_value | Valeur « pas de données » (-9999) |
| Datas | Données climatique concernée, tableau de 720 par 720 |

Après lecture d’un des tableau et conversion des données en couleur (RGB 0-255) voici l’image obtenue. Il s’agit ici des températures moyennes du mois de janvier 2011 :

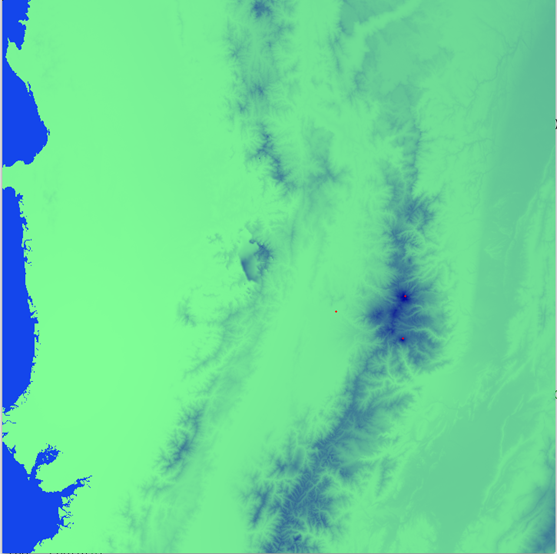


Figure 1 Affichage des points d'un fichier

Les marqueurs rouges ont été utilisés pour confirmer les bons rapports emplacements / indices.

## Données climatiques – Librairie

Les données climatiques sont réparties en 360 fichiers (6 ans x 5 mesures x 12 mois). Une classe Python a été créée comme *Wrapper* autour de ces fichiers afin de pouvoir interagir directement avec des coordonnées GPS (EPSG:4326[[1]](#footnote-1), Degrés décimaux). L’image sur la page précédente a été générée à l’aide de ce Wrapper qui permet d’afficher un des fichiers sous la forme d’une image où les données sont réparties sur une échelle de couleur.

Utilisation :

import Averages as avgs

wrapper = avgs.ClimaticDataWraper(["2011","2012","2013","2014","2015","2016"])

precdata = wrapper.getPrecAverages([x,y])

tmindata = wrapper.getTminAverages([x,y])

tmaxdata = wrapper.getTmaxAverages([x,y])

tmeandata = wrapper.getTmeanAverages([x,y])

dtrdata = wrapper.getDtrAverages([x,y])

# Données des dégustations de cafés

Les données de dégustation ont été reçues sous des formes différentes. Les premières données sont parvenues sous la forme d’un classeur Excel dont seulement 3 entrées sur environ 200 étaient complètes, le reste devant être fournit plus tard. La partie devant fournir les données normalisées n’ayant pas pu réaliser ce travail dans les temps, des données brutes ont été reçues. Différentes étapes ont été nécessaires pour extraire des données cohérentes :

1. Tri des données incomplètes, hors sujet ou vides
2. Rassembler les données cohérentes selon leur types (PDF, Classeurs Excels sous plusieurs formats)
3. Réaliser des scripts d’extraction de données pour chaque type cohérent et former un premier dataset

Une fois le premier Dataset formé, les données ont été envoyées au Comité des Caféiculteurs afin de compléter les données manquantes (compléter les identifiants SICA et les numéros de cartes d’identités, ajout des coordonnées GPS).

Les données comprennent les champs suivants :

|  |  |  |
| --- | --- | --- |
| Champ | Description | Unités ou valeurs |
| Municipio | Municipio dans lequel se trouve la plantation | Catégorique |
| Vereda | Vereda dans laquelle se trouve la plantation | Catégorique |
| Finca | Nom de la ferme | Catégorique |
| SICA | Identifiant unique des cultures de cafés | Nombre entier à 10 digits |
| Variedad | Variété de café cultivé dans la ferme correspondante | Catégorique |
| CENT\_X | Référence spatiales  <http://spatialreference.org/ref/epsg/magna-sirgas-colombia-bogota-zone/> | EPSG:3116[[2]](#footnote-2) |
| CENT\_Y | EPSG:3116 |
| EPSG:4326\_X | Coordonnées utilisées pour accéder aux données climatiques (valide sur la planète entière et pas seulement dans la zone EPSG :3116) | EPSG:4326 |
| EPSG:4326\_Y | EPSG:4326 |
| ASNM | Altitude | Mètres |
| Luminosidad | Description de la luminosité au niveau de l’emplacement |  |
| Date de réception |  |  |
| Date d’analyse |  |  |
| Malla (3x) | Pourcentage de grain compris dans une certaine taille (maille de tamis) |  |
| Défauts (17x) | Notation des défauts physiques des grains | Plus il y a de points moins c’est bien |
| Caractérisation physique | Pergamino  Almendra  AlmendraTotal  AlmendraSana  UV  Olor  Humedad (%)  Merma (%)  Factor de Rendimiento |  |
| Analyse sensorielle | Aroma-Fragancia  Acidez  Cuerpo  Sabor  Sabor residual  Dulzor  Uniformidad  Balance Taza limpia  Puntaje catador  Puntaje Total |  |
| Taza (5x) | Indique si la tasse est « limpla », c’est-à-dire propre (absence de défauts), ou si elle souffre de défauts majeurs. |  |
|  |  |  |

# Données de qualité de sol

Les données de sol étaient aussi très difformes. De nombreux documents ont été fournis mais ne permettaient pas d’avoir des données identiques sur la totalité du territoire étudié. Des *shapes files* disponibles avec un GIS (Geographical Information System) ont permis d’extraire pour une majorité de plantation des caractéristiques de textures, de pH et de taux de matières organiques présentes pour une profondeur allant jusqu’à un mètre. Certaines des zones du GIS n’étaient malheureusement pas disponibles.

|  |  |  |
| --- | --- | --- |
| Champ | Description | Unités ou valeurs |
| Profile | Dénomination du profil de sol |  |
| Unité | Autre dénomination sortie du GIS |  |
| Franco | Variables binaires répétée 3 fois (une fois pour chaque couche, répété si les couches sont identiques jusqu’à 1m)  Franco signifie que le sol à des qualités « parfaites », les autres définissent si la couche tend vers un sol argileux, limoneux, sablonneux ou graillonneux. | 0 ou 1 (Faux ou Vrais) |
| Arcilloso |
| Limoso |
| Arenoso |
| Cascajoso |
| Ph\_avg | Moyenne du PH sur les couches allant jusqu’à 1m | Numérique continue |
| Org\_avg | Moyenne du taux de matière organique jusqu’à 1m | Numérique continue |
|  |  |  |
|  |  |  |

1. <http://spatialreference.org/ref/epsg/wgs-84/> [↑](#footnote-ref-1)
2. <http://spatialreference.org/ref/epsg/magna-sirgas-colombia-bogota-zone/> [↑](#footnote-ref-2)