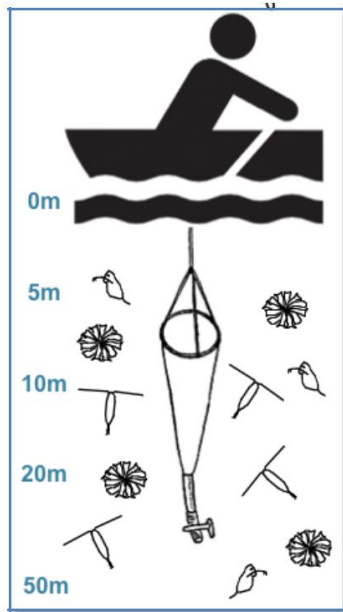


Saisie, manipulation et qualité des données issues des planctons

Source : <http://www.dataone.org/education-modules>



Organisation des données

Contexte : Le plancton est un organisme microscopique qui forme la base de nombreux réseaux trophiques aquatiques et alimente la croissance des poissons et d'autres organismes plus gros. Il est courant de les échantillonner à l'aide d'un filet ou d'un autre contenant qui peut être contrôlé pour recueillir l'eau à certaines profondeurs seulement. On peut voir sur le dessin les différentes façons de saisie : le plancton recueilli à la surface (0 mètre) peut être différent du plancton recueilli à une autre profondeur (par exemple, 10 mètres sous la surface).

(Pour plus d'informations :

<http://en.wikipedia.org/wiki/Phytoplankton> et

<http://en.wikipedia.org/wiki/Zooplankton>)

Les échantillons sont identifiés et comptés au microscope, et leur nombre est généralement exprimé en individus par litre ou par millilitre. Les scientifiques aquatiques prélèvent des échantillons de plancton le jour (p. ex. midi) et la nuit (p. ex. 2 h) parce que le plancton change sa répartition du jour au soir et que les espèces ne changent pas toutes leur répartition de la même façon. Pour plus d'informations, recherchez "*Migration verticale*" sur wikipedia.

Les données

Vous disposez de 3 fichiers de données (fictifs) : pond2010.xlsx, zoop-temp-main.xlsx ; zoop-temp.xlsx. A l'origine, ces trois fichiers devaient tous faire partie de la même étude - les chercheurs voulaient examiner la répartition jour-nuit de deux espèces de zooplancton sur plusieurs années. Le type de zooplancton qu'ils ont étudié est appelé généralement *rotifères*, et plus précisément le type *Conochilus*, dans lequel des groupes de rotifères individuels se regroupent en colonies ([voir http://eol.org/pages/43393/overview](http://eol.org/pages/43393/overview)). Les chercheurs prévoient de répéter cette étude pendant encore plusieurs années.

Tâche 1 : Ouvrir les 3 fichiers et les inspecter. Quels sont les **problèmes** soulevés par la façon dont les données sont actuellement organisées ?

Tâche 2 : Proposer un nouveau système d'organisation des données. Créez une nouvelle feuille de calcul qui peut servir de modèle pour les années ultérieures de collecte de données.

Tâche 3 : Afficher les données sur un graphique (plot) ou bien calculer les min/max d'une colonne pour trouver des anomalies.

Tâche 4 : Suggérer un système qui permet de signaler les données anormales.

Tâche 5 : Suggérer un système qui permet de signaler les données manquantes.

Tâche 6 : Suggérer un système de saisie qui peut prévenir les erreurs.