Séminaire de Probabilités et Statistiques

Nathalie Villa-Vialaneix

(INRA)

Multiple kernel learning applied to the integration of Tara oceans datasets

The Tara Oceans expedition [sunagawa et al, 2005] facilitated the study of plankton communities by providing the scientists with ocean metagenomic data combined with environmental measures. During the expedition, 243 seawater samples were collected from 68 locations representing all main oceanic regions at three depth layers : the surface (SRF), the deep chlorophyll maximum (DCM) layer and the mesopelagic (MES) zone. During the presentation, I will describe a method to integrate information provided by different datasets collected during the expedition. The approach uses kernels which are combined in an unsupervised setting for data mining purposes. Additionnaly, tools to help the interpretation of the results are given and shows that well known facts about parts of the datasets are recovered and that new insights on the data are also obtained.

This work is a joint work with Jérôme Mariette.