

Resumé

Cet article présente l'intégration de données spatiales dans le logiciel de fouille de données RapidMiner (www.rapid-i.com). RapidMiner est un outil pour **concevoir interactivement des chaînes de traitement**, orientées vers la fouille de données. Dans le cadre de ce travail, nous avons implémenté une extension de RapidMiner pour **traiter des données spatiales**. On bénéficie ainsi d'une **large gamme de méthodes de fouille de données** déjà à disposition pour répondre à une question d'analyse sur des données spatiales. Le fonctionnement de l'extension est illustré par une comparaison de méthodes de classification sur un problème de détection de plante invasive (la Jussie) à partir d'une image aérienne hyperspectrale.

Intégration des données géoréférencées sous forme tabulaire

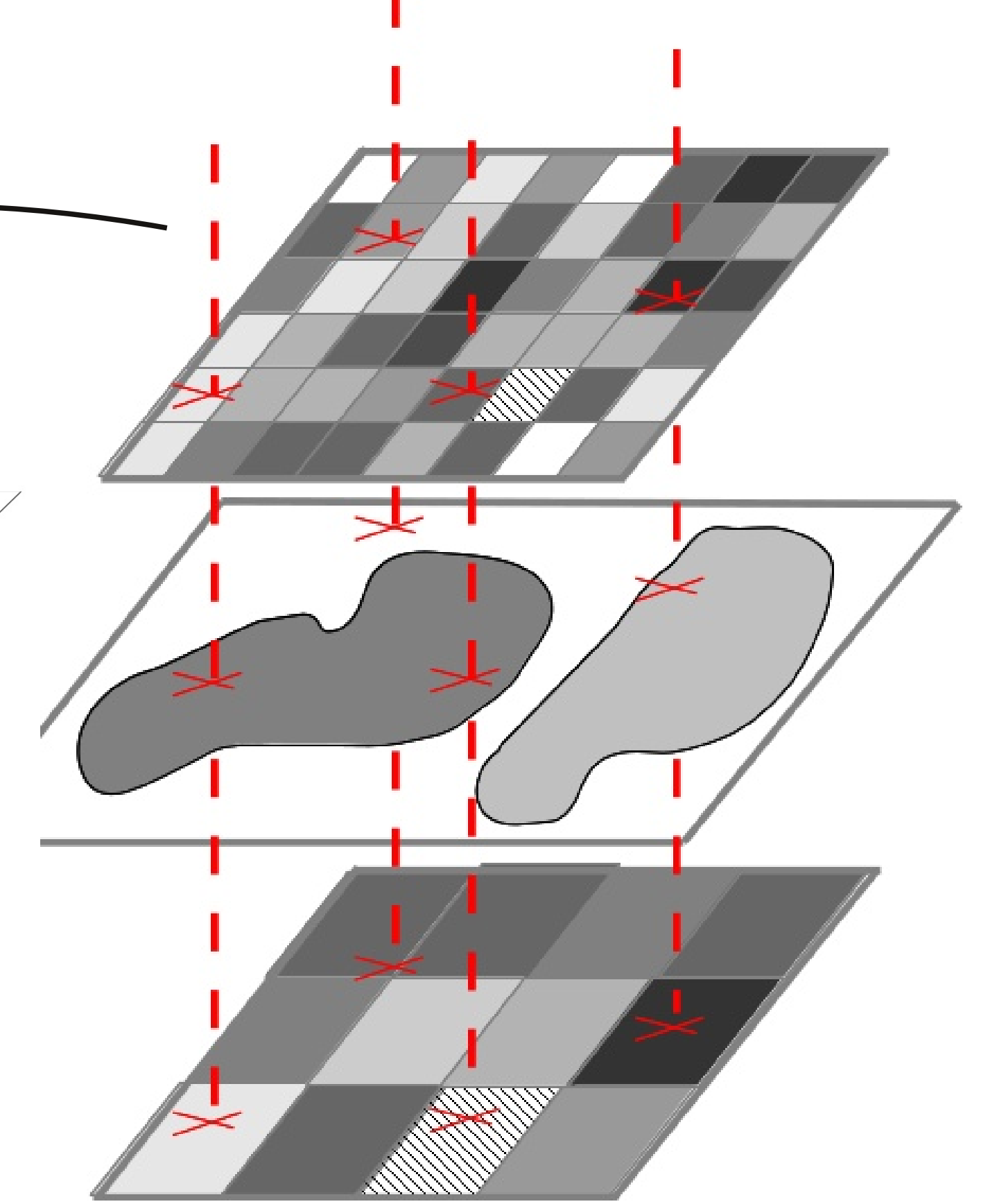
RapidMiner ne peut traiter que des données sous forme tabulaire

- les attributs en colonnes
- les exemples en lignes

Solutions d'échantillonnage proposées:

- l'échantillonnage selon un transect,
- l'échantillonnage régulier,
- l'échantillonnage aléatoire d'une zone rectangulaire,
- l'échantillonnage personnalisé qu'un utilisateur définit par une couche de points.

	X-Coordinate	Y-Coordinate	im1-band1	im1-band2	im1-band3	im2_attr1	im2_attr2	im3-band1	im3-band2	im3-band3
1	0	0,5	0,5	0,4	1	1	0,3	0,4	0,2	
4	1	1,2	0,9	1,4	?	?	2,5	2,1	2,3	
1	4	2,1	2,2	2,2	1	1	0,01	0	0,2	
3	6	7	6,5	5,6	2	2	6,3	6,1	5,9	

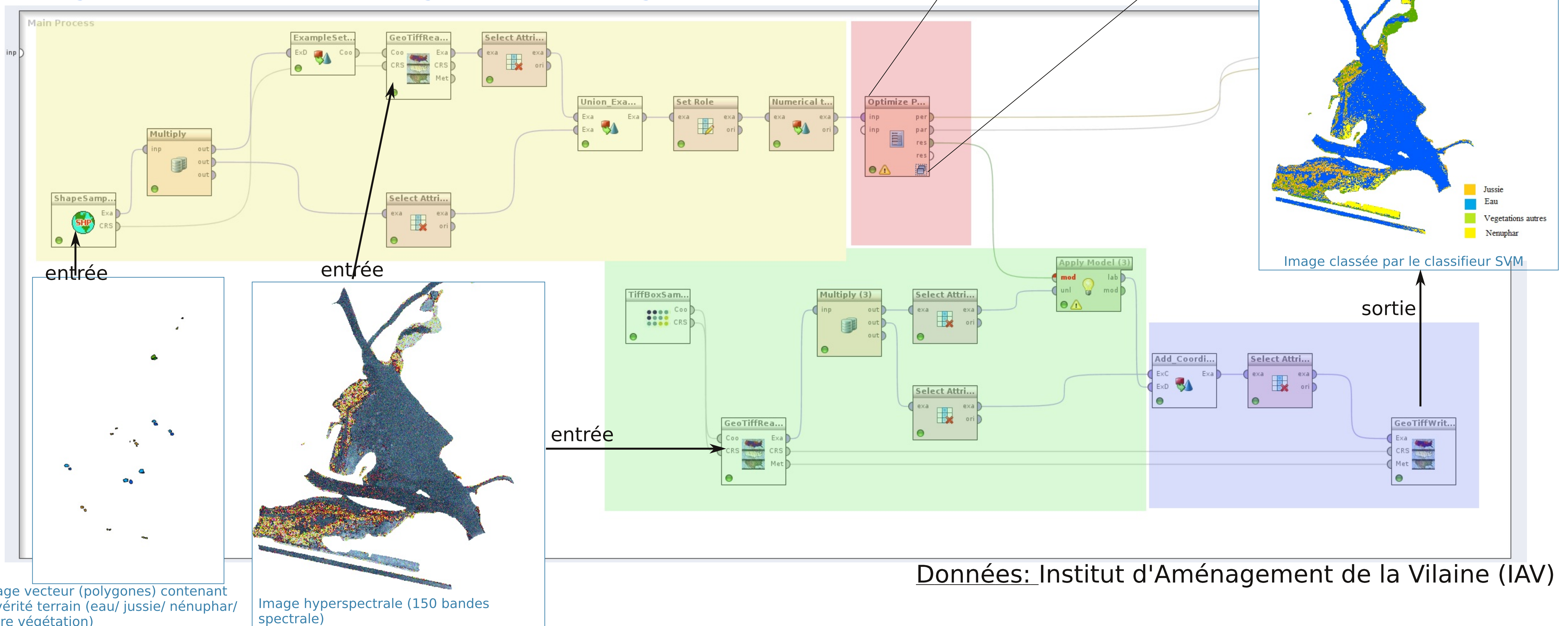
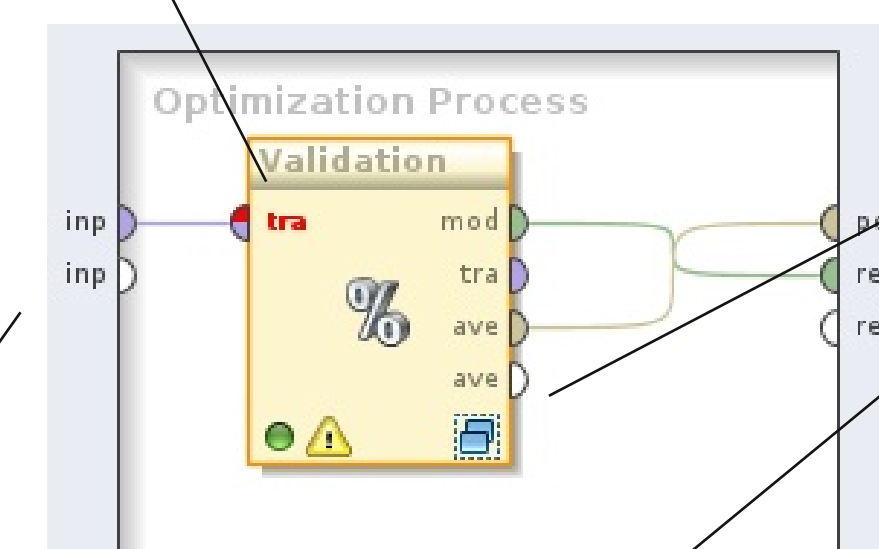
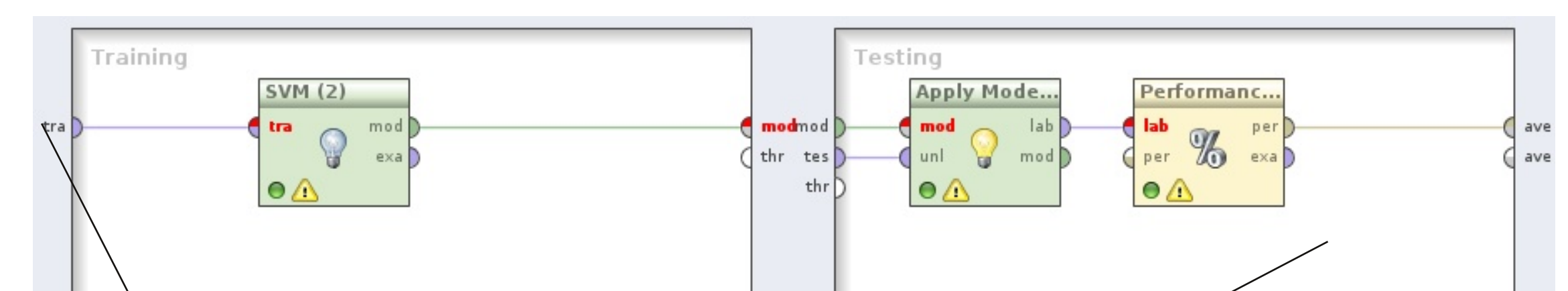


Exemple de chaîne de traitements: détection de la Jussie dans une image hyperspectrale

Objectif: **identifier les paramètres optimaux** d'un classifieur SVM pour faire une classification supervisée d'image hyperspectrale.

Principe (les couleurs correspondent aux différentes parties de la chaîne de traitement ci-dessous):

- On charge des données d'apprentissage :
 - Une image raster : image hyperspectrale à classer
 - Une image vecteur : vérité terrain structurée comme un ensemble de polygones avec une classe
- On entraîne des classifieurs SVM en faisant varier ses paramètres et on évalue ses performances au moyen d'une validation croisée
- Le meilleur modèle est utilisé pour appliquer le classifieur à tous les pixels de l'image
- L'image raster résultante est enregistrée en format geotif



Données: Institut d'Aménagement de la Vilaine (IAV)

Conclusion et Perspectives

- Développement de la première extension RapidMiner pour l'intégration de traitement de données géoréférencées : **outil interactif, sans programmation**, bénéficie de tous les algorithmes d'analyse de données déjà existants dans RapidMiner.

- Outils servant au **prototypage rapide des méthodes d'analyse de données géoréférencées**

- **Expérimentation de l'outil dans un cadre pédagogique** pour découvrir les possibilités des méthodes d'apprentissage automatique et de fouille de données (formation de master en télédétection).

- Seul un sous-ensemble des méthodes de fouille de données spatiales est actuellement intégré à l'outil

Site web du projet
<http://geoimagermp.gforge.inria.fr/>

