



HAL
open science

Partage de pratiques et outils pour le calcul scientifique

Marie-Josée Cros, Claire Chabanet, Fabrice Dessaint, David Legland, Serguei Sokol, Herve Richard

► **To cite this version:**

Marie-Josée Cros, Claire Chabanet, Fabrice Dessaint, David Legland, Serguei Sokol, et al.. Partage de pratiques et outils pour le calcul scientifique. Journées Math-Info de l'INRA, Oct 2016, Mallemort, France. hal-01886412

HAL Id: hal-01886412

<https://hal.science/hal-01886412v1>

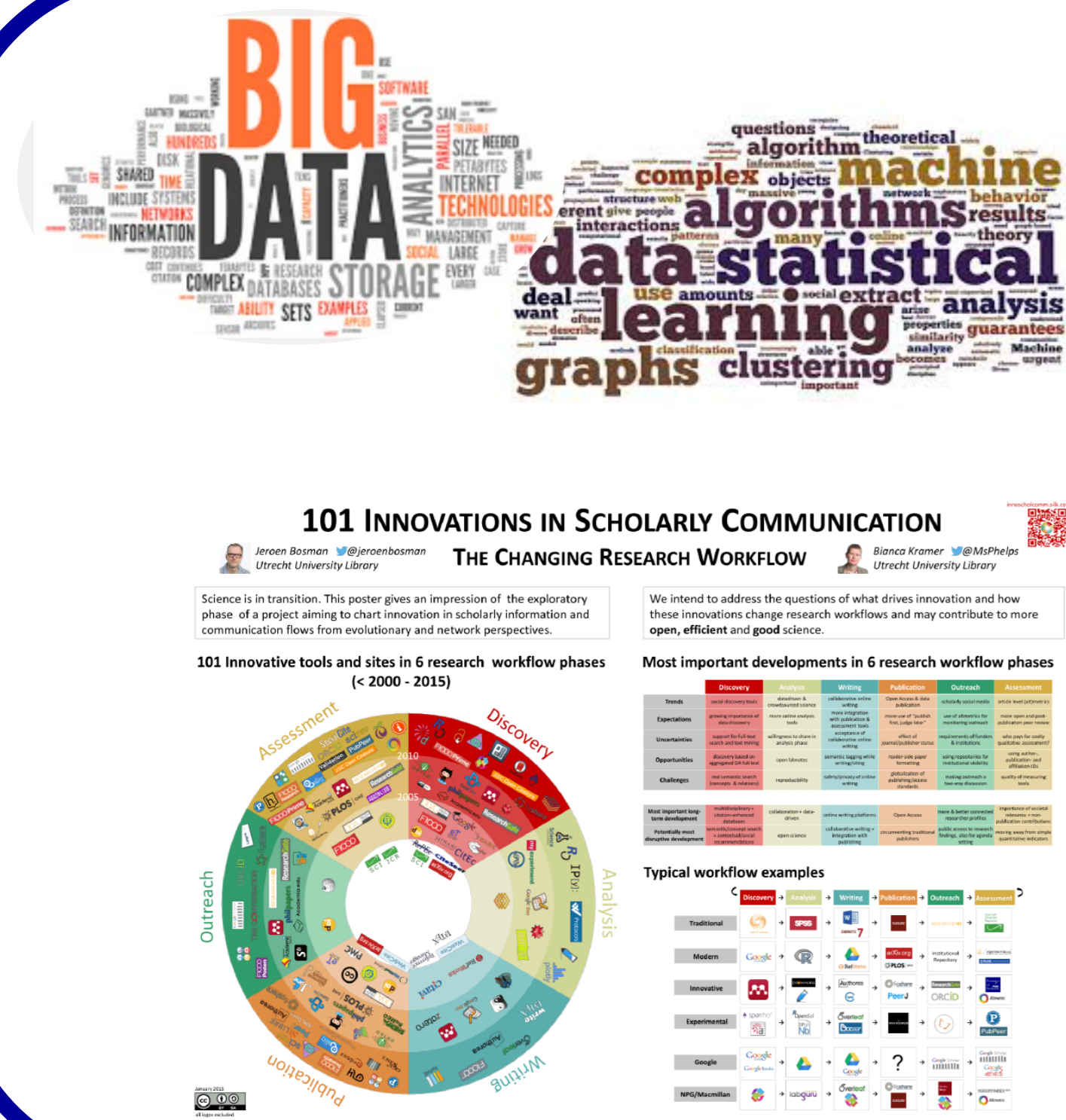
Submitted on 5 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Partage de pratiques et outils pour le calcul scientifique

Groupe Pratiques&Outils du CATI CaSciSDI de l'INRA
pratiques-outils-calcul@listes.inra.fr



CONTEXTE

- Les scientifiques utilisent de plus en plus le calcul scientifique.
- Tous n'ont pas été formés.
- Les pratiques et les outils évoluent.

ENJEUX

Aider les permanents et les non-permanents de l'INRA et de ses partenaires à s'informer sur, se former à, adapter, améliorer les pratiques et outils mis en œuvre pour le calcul scientifique.

Ceci afin d'améliorer la qualité (fiabilité, robustesse, traçabilité...) et la reproductibilité des calculs (effet cumulatif des recherches).

Un site web participatif pour partager des pratiques, des liens, des présentations, des introductions aux outils, un état de l'art ...

<http://wiki.inra.fr/wiki/cascisdi/>

(Le dernier '/' doit être tapé pour un accès sans authentification)

Déterminer la taille des échantillons
notion sous-jacente : puissance d'un test
Claire Chabanet, Fabrice Dessaint
10 mai 2016

1 Objectif
Lors de la planification d'une expérience, il est nécessaire de choisir la taille des échantillons, autrement dit le nombre d'unités expérimentales (individus, placettes, parcelles...) sur lesquelles seront appliqués les différents traitements étudiés. Ce choix est bien évidemment

Conseils pour la saisie des données
Modifié par Claire CHABANET le 2016/03/14 09:02

Outliers ou données extrêmes, comment les détecter?
que faire de ces observations?
Claire Chabanet, Fabrice Dessaint
10 juillet 2015

1 Objectif de la fiche
L'utilité de la rédaction de cette fiche est née de la constatation d'une hétérogénéité des pratiques concernant la façon de gérer les points dits « aberrants ». De la confrontation de la diversité des pratiques, naît une question : « Que faire face à un/des points extrêmes ? »

Conventions de codage pour Matlab
Modifié par David LEGLAND le 2016/07/17 10:40

Résumé de quelques conventions de codage préconisées pour le langage Matlab. On peut aussi consulter le document de Richard Johnson (http://www.datacube.com/downloads/matlab_style_guidelines.pdf), dont ce document est en grande partie inspiré.

Conventions pour les noms
Il est préférable d'employer des noms anglais. Cela permet (1) d'homogénéiser la langue du code et (2) d'éviter les problèmes dus aux accents... On évite les numéros sauf pour des fichiers temporaires ou de données.

Programmation orientée objet sous Matlab
Modifié par Marie-Josée CROS le 2016/09/13 10:38

Cette page contient un guide pour la programmation sous Matlab, et en particulier sur la programmation orientée objet.

Utiliser un gestionnaire de version
Modifié par Gwen MORIN le 2016/04/07 08:55

Qui n'a pas été confronté au problème de s'y retrouver dans différentes versions de code : quelle est la dernière version ? quel est ce qui a été modifié depuis la précédente version ? Aje intègre les modifications de X ? Aje intègre Y de mes collègues ? ...

Versions avec Git
Modifié par Marie-Josée CROS le 2016/04/07 10:43

Si vous êtes sur cette page c'est que vous posez sûrement des questions sur la conformité de Git à vos besoins en gestion de versions de vos fichiers. Pour vous aider à prendre la décision : se lancer dans Git ou pas, nous allons brièvement présenter ce que offre Git et à quel prix. À quel prix ? est au sens figuré du terme car il est gratuit en OpenSource. Il s'agit plutôt de l'investissement de votre temps dans l'apprentissage de Git. Ceci n'est donc pas un « être tenté » sur Git mais une sorte d'aide pour la prise de décision s'y aller ou pas.

1 Utilisation de Git sous RStudio
Modifié par Fabrice Dessaint le 2016/02/15 13:52

Centre open source permettant l'utilisation du gestionnaire de version Git via une interface graphique proposée par l'investissement de développement (IDE) RStudio.

Conditionnement d'une matrice
Modifié par Sylvain ASSON le 2016/07/20 11:03

Motivation
Définitions
Le conditionnement d'une matrice carrée est le produit de la norme de cette matrice avec celle de la norme de son inverse. Si on utilise la norme 2 c'est aussi le rapport entre la plus grande et la plus petite des valeurs absolues des valeurs propres de la matrice. Une matrice non-inversible aura un conditionnement infini.

Recherche Reproductible en calcul scientifique
Modifié par Marie-Josée CROS le 2016/03/25 12:00

Contexte
Le calcul prend une place de plus en plus importante dans la recherche. Un mouvement appelé recherche reproductible se propose d'améliorer la reproductibilité des calculs intervenant dans les recherches afin de permettre de vérifier les résultats, réutiliser les codes. Ce mouvement est lié à un autre appelé Open Research (Open Data, Open Source...).

Régression linéaire sous Matlab
Modifié par David LEGLAND le 2016/07/01 15:09

Contexte
Le logiciel Matlab est très utilisé pour le calcul numérique, mais dès que l'on veut l'utiliser pour répondre à des questions simples, on se heurte assez vite à des difficultés liées à la méconnaissance de l'outil. L'idée de cette page est de présenter, sous la forme de "recette de code", comment utiliser Matlab pour résoudre des problèmes assez classiques de régression, en particulier la régression linéaire.

Plateformes de partage
Modifié par Marie-Josée CROS le 2016/09/20 15:38

Automatiser les tests
Modifié par Marie-Josée CROS le 2016/03/28 15:00

Tests unitaires sous R
Modifié par Marie-Josée CROS le 2015/07/09 11:43

EditeursEnLigne
Modifié par Marie-Josée CROS le 2016/02/27 07:41

Org-mode Emacs
Modifié par Marie-Josée CROS le 2016/01/09 16:26

Ressources extérieures à ce wiki :

- Plateforme Logicielle pour l'Analyse des Signaux et leur Traitement, l'Intégration de données et des Connaissances (PLASTIC) du département Cépia de l'Inra
- Site sur connaissances et pratiques de l'environnement R (R4ciam) du cati CaSciSDI de l'Inra
- Site gestion et partage des données scientifiques de l'Inra

Le site a été initié en 2015 par 8 personnes des départements ALIMH, BAP, CEPIA, MIA, MICA, SPE. Il est ouvert en lecture à tous et pour contributions / commentaires aux personnes ayant un ldap INRA. Le site évolue principalement par itérations lors de périodes définies par le groupe et dans le cadre d'un CATI INRA.