



# QUELS LOGICIELS POUR LA RECHERCHE SCIENTIFIQUE?

## SEMINAIRE MATHS-ÉCO

Christophe Dutang<sup>1</sup>

<sup>1</sup> Institut du Risque et de l'Assurance, Université du Maine, Le Mans, France

22 septembre 2015

Données

○○○

Modélisation

○○○○○  
○○○

Applications

○○○○

Rendu

○○  
○○○○○

Industrialisation

○○○○○  
○○○○○

# OUTLINES

## 1 MANIPULATION DES DONNÉES

- Manipulation et techniques exploratoires

## 2 MODÉLISATION

- Rédaction du modèle et gestion des références
- Résolution formelle ou numérique

## 3 APPLICATIONS NUMÉRIQUES

- Modèles usuels

## 4 COMPTE-RENDU ET SORTIES EN R

- Graphiques et multimédia
- Rapport pdf, poster, web

## 5 INDUSTRIALISATION

- Performance
- Code packaging et travail collaboratif

Données



Modélisation



Applications



Rendu



Industrialisation



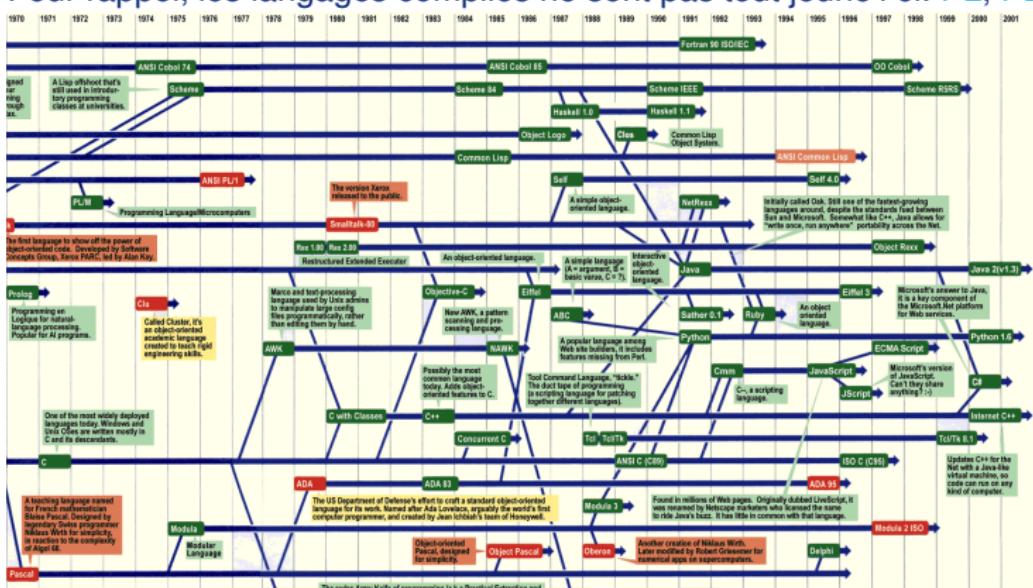
# INTRODUCTION

On se concentre sur 3 logiciels scientifiques

- SAS édité en 1976
- matlab édité en 1984
- R édité en 1993

Sont-ils vieux ?

Pour rappel, les langages compilés ne sont pas tout jeune ! cf. PL, PL



**Données**

●○○

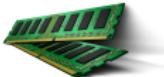
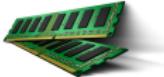
**Modélisation**○○○○○  
○○○**Applications**

○○○○

**Rendu**○○  
○○○○○**Industrialisation**○○○○  
○○○○○

# DONNÉES CHARGEABLES EN MÉMOIRE VIVE

## Fonctionnement (par défaut)

Logiciel	Stockage	Avantage	Défaut
SAS		Volumineux ( $\approx 1$ To)	Lent (40 à 60 Mo/s)
matlab		Rapide (1 à 20 Go/s)	Peu volumineux ( $\approx 8$ Go)
R		Rapide (1 à 20 Go/s)	Peu volumineux ( $\approx 8$ Go)

## Lecture de fichiers standard

Type	SAS	matlab	R
csv, txt	data+infile	fscanf	read.table
Excel	proc import	xlsread	read.xls

Données



Modélisation



Applications



Rendu



Industrialisation



## DONNÉES MASSIVES

Beaucoup de solutions existent...

Type	SAS (natif)	matlab (spécifique)	R (spécifique)
Base SQL	proc import	xlsread	RODBC, DBI
Internet	proc download	webread	download.file
Enorme fichier	data+infile	datastore	ff, bigmemory

Mais beaucoup reste à faire !

Données



Modélisation



Applications



Rendu



Industrialisation



## TECHNIQUES EXPLORATOIRES (USUELLES)

Type	SAS	matlab	R
stat. desc.	proc univariate	mean, median, histfit, probplot	mean, median,...
pp(qq)plot, hist boxplot		boxplot	qqplot, hist boxplot
tout en un	proc univariate		summary, str

Type	SAS	matlab	R
scatter plot	proc sgplot	scatterplot	matplot, persp
star plot	proc gchart		stars
mozaic, bubble plot	proc freq	scatter	mozaicplot, points
pairs plot	proc variogram		pairs

Type	SAS	matlab	R
Conting. table	proc freq		table, xtabs
PCA	proc princomp	pca	princomp
FCA	proc corresp		ca

# LOGICIELS WYSIWYG VS. LOGICIELS DE COMPOSITION

Type	Nom	Force	Faiblesse	Wikibook
WYSIWYG	Word	Rapide prise en main, interface simple	Prix, formule (maths)	oui
WYSIWYG	OpenOffice	Gratuit, proche de word	Interface simpliste	oui
Form + content	L <small>A</small> T <small>E</small> X	Gratuit, complet, extensible par package	Courbe d'apprentissage	oui

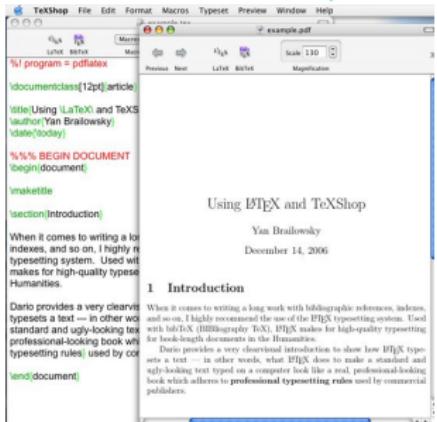
- Word peut être utilisé pour **rédiger** des articles mais inadapté pour produire des livres, des thèses.
- LATEX conçu par D. Knuth et L. Lamport permet de produire des documents de haute qualité (article, thèse, livre,...)



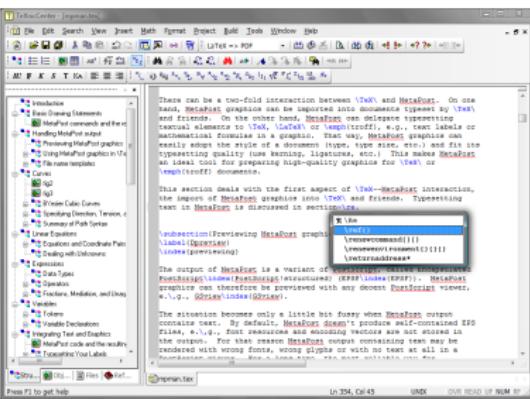
# INTERFACES POUR LOGICIELS DE COMPOSITION

$\text{\LaTeX}$  dispose de plusieurs interfaces

sur mac Texshop



sur windows texniccenter



mais aussi winedit, texmaker,...



# INTERFACES POUR LOGICIELS DE COMPOSITION

...et d'une surcouche entièrement WYSIWYG **lyx** disponible sur mac, windows et linux.

**5.1.6 Operators with Limits**

Sum  $\Sigma$  and integral  $\int$  operators are very often decorated with limits. These limits can be entered in LyX by entering them as you would enter a super- or subscript, directly after the symbol. The sum operator will automatically place its "limits" over and under the symbol in displayed formulas, and on the side in inline formulas. Such as  $\sum_{n=0}^{\infty} \frac{1}{n!} = e$ , versus

$$\sum_{n=0}^{\infty} \frac{1}{n!} = e$$

Integral signs, however, will place the limits on the side in both formula types.

All operators with limits will be automatically re-sized when placed in display mode. The placement of the limits can be changed by placing the cursor directly behind the operator and hitting M-m l or using the menu Edit > Math > Change\_Limits\_Type.

Certain other mathematical expressions have this "moving limits" feature as addition, such as  $\lim_{x \rightarrow \infty} f(x)$ ,

which will place the  $x \rightarrow \infty$  underneath the "lim" in display mode. In inline formulas it looks like this:  $\lim_{x \rightarrow \infty} f(x)$ .

Note that the lim-function was entered as the function macro \lim. Have a look at section [Ref: sub\\_Functions](#) for an explanation of function macros.

**5.1.7 Math Symbols**

## 5.1.6. Operators with Limits

Sum  $\Sigma$  and integral  $\int$  operators are very often decorated with limits. These limits can be entered in LyX by entering them as you would enter a super- or subscript, directly after the symbol. The sum operator will automatically place its "limits" over and under the symbol in displayed formulas, and on the side in inline formulas. Such as  $\sum_{n=0}^{\infty} \frac{1}{n!} = e$ , versus

$$\sum_{n=0}^{\infty} \frac{1}{n!} = e$$

Integral signs, however, will place the limits on the side in both formula types.

All operators with limits will be automatically re-sized when placed in display mode. The placement of the limits can be changed by placing the cursor directly behind the operator and hitting M-m l or using the menu Edit > Math > Change\_Limits\_Type.

Certain other mathematical expressions have this "moving limits" feature as addition, such as

$$\lim_{x \rightarrow \infty} f(x),$$

which will place the  $x \rightarrow \infty$  underneath the "lim" in display mode. In inline formulas it looks like this:  $\lim_{x \rightarrow \infty} f(x)$ .

Note that the lim-function was entered as the function macro \lim. Have a look at section [5.1.9](#) for an explanation of function macros.

Données  
○○○

Modélisation  
○○○●○  
○○○

Applications  
○○○○

Rendu  
○○  
○○○○○

Industrialisation  
○○○○  
○○○○○

## GESTION DES RÉFÉRENCES BIBLIOGRAPHIQUES

Un grand **nombre de logiciels** existent pour gérer une bibliothèque d'articles.  
Beaucoup se basent sur le format bibtex :

```
@book{Ben62,
    title = "L'Oxydation des m\`etaux",
    author = "J. Benard and J. Bardolle and F. Bouillon and M. Cagnet
and J. Moreau and G. Valensi",
    publisher = "Gauthier-Villars",
    year = "1962"
}

@article{Kaku41,
    Author = {S. Kakutani},
    Journal = {Duke Math. J.},
    Number = {3},
    Pages = {457-459},
    Title = {{A generalization of Brouwer's fixed point theorem}},
    Volume = {8},
    Year = {1941}}
```



# JABREF ET BIBDESK

Nom	Système	Fichiers entrée	Fichiers sortie	Base Données	Dépot
bibdesk	mac	bibtex, endnote, medline, ...	bibtex, medline, XML	arXiv, CiteSeer, JSTOR, Gscholar	oui
jabref	mac, win	bibtex, endnote, medline, ...	bibtex, endnote, XML	arXiv, CiteSeer, JSTOR	

JabRef - Fenêtre principale (bibdesk, thesis.bib)

File Edit View BibTeX Tools Web search Design Options Help

Actions New Edit Delete Preview

All Entries Cod

- Ingestion
- Endnotes
- Type spelling
- Salmons
- Modeling
- Articles
- Tracking
- Checkpoint
- Major quality

Search

Search All fields

Incremental

Show results in dialog

Global search

Settings

Article

Required fields Optional fields General Abstract Review BibTeX source

Author: Ericaud, A. and Bedhomme, A. L. and Morel, A.

Title: Optical-properties of diverse phytoplankton species --- Experimental results and theoretical interpretation

Journal: Journal of Plankton Research

Manager

Toggle abbreviation

Year: 1988

Volume: 10

Pages: 851–873

Editor: Bricaud1988

Status: Preferences recorded

thesis.bib

Action New Edit Delete Preview

Keywords

- attitude
- american estimates
- calibration
- computer vision
- context
- dynamics
- DS
- order angles
- order axis
- CPS
- hand-held
- holonomic
- MU
- NO
- PA
- rotation
- kinematics
- multimotor
- observability
- perturbation analysis
- quaternions
- reference frames
- rotations composition
- rotations matrices
- SLAM
- vernier

Search

Search Bibliography

Key Type Date

# 1 Article Review of Attitude Representations Used for Aircraft Attitudes

75 A Article Ericaud, A. and Bedhomme, A. L. and Morel, A. 1988 Optical-properties of diverse phytoplankton species --- Experimental results and theoretical interpretation. *Journal of Plankton Research*, 10(8), pp. 851-873. DOI: 10.1080/01678348808523553

76 A Article Ericaud, A. and Bedhomme, A. L. and Morel, A. 1989 Natural variability of phytoplanktonic algal pigments. *Journal of Plankton Research*, 11(12), pp. 1253-1263. DOI: 10.1080/01678348908523554

77 A Article Ericaud, A. and Bedhomme, A. L. and Morel, A. 1990 Variations of light absorption by suspensions of phytoplankton. *Journal of Plankton Research*, 12(12), pp. 1353-1363. DOI: 10.1080/01678349008523555

78 A Article Ericaud, A. and Bedhomme, A. L. and Morel, A. 1991 Embryology, ethology and ecology of *P*. *brunneum*. *Bracke*, 1991. Perspectives on ecosystem-based app. 1994. *Marine E.*, 1994. DOI: 10.1080/01678349408523556

79 A Article Ericaud, A. and Bedhomme, A. L. and Morel, A. 1992 Toward a metabolic theory of ecology. *Ecology*, 73(12), pp. 3533-3543. DOI: 10.2307/19412100

80 A Article Ericaud, A. and Bedhomme, A. L. and Morel, A. 1993 Larval recruitment of Atlantic cod (*Gadus*). *Cad.*, 2003. *Aquacult.*, 1993. DOI: 10.1080/01678349308523557

81 A Article Ericaud, A. and Bedhomme, A. L. and Morel, A. 1994 The use of behavioural observations in larval fish recruitment models. *Environ Biol Fish*, 1994. DOI: 10.1007/BF00045200

82 A Article Ericaud, A. and Bedhomme, A. L. and Morel, A. 1995 Theoretical and experimental studies on the development of the larval fish *Gadus*. *Environ Biol Fish*, 1995. DOI: 10.1007/BF00045201

83 A Article Ericaud, A. and Bedhomme, A. L. and Morel, A. 1996 The optical properties of phytoplankton. *Environ Biol Fish*, 1996. DOI: 10.1007/BF00045202

84 A Article Ericaud, A. and Bedhomme, A. L. and Morel, A. 1997 The optical properties of phytoplankton. *Environ Biol Fish*, 1997. DOI: 10.1007/BF00045203

85 A Article Ericaud, A. and Bedhomme, A. L. and Morel, A. 1998 The optical properties of phytoplankton. *Environ Biol Fish*, 1998. DOI: 10.1007/BF00045204

86 A Article Ericaud, A. and Bedhomme, A. L. and Morel, A. 1999 The optical properties of phytoplankton. *Environ Biol Fish*, 1999. DOI: 10.1007/BF00045205

87 A Article Ericaud, A. and Bedhomme, A. L. and Morel, A. 2000 The optical properties of phytoplankton. *Environ Biol Fish*, 2000. DOI: 10.1007/BF00045206

88 A Article Ericaud, A. and Bedhomme, A. L. and Morel, A. 2001 The optical properties of phytoplankton. *Environ Biol Fish*, 2001. DOI: 10.1007/BF00045207

89 A Article Ericaud, A. and Bedhomme, A. L. and Morel, A. 2002 The optical properties of phytoplankton. *Environ Biol Fish*, 2002. DOI: 10.1007/BF00045208

90 A Article Ericaud, A. and Bedhomme, A. L. and Morel, A. 2003 The optical properties of phytoplankton. *Environ Biol Fish*, 2003. DOI: 10.1007/BF00045209

91 A Article Ericaud, A. and Bedhomme, A. L. and Morel, A. 2004 The optical properties of phytoplankton. *Environ Biol Fish*, 2004. DOI: 10.1007/BF00045210

92 A Article Ericaud, A. and Bedhomme, A. L. and Morel, A. 2005 The optical properties of phytoplankton. *Environ Biol Fish*, 2005. DOI: 10.1007/BF00045211

93 A Article Ericaud, A. and Bedhomme, A. L. and Morel, A. 2006 The optical properties of phytoplankton. *Environ Biol Fish*, 2006. DOI: 10.1007/BF00045212

94 A Article Ericaud, A. and Bedhomme, A. L. and Morel, A. 2007 The optical properties of phytoplankton. *Environ Biol Fish*, 2007. DOI: 10.1007/BF00045213

95 A Article Ericaud, A. and Bedhomme, A. L. and Morel, A. 2008 The optical properties of phytoplankton. *Environ Biol Fish*, 2008. DOI: 10.1007/BF00045214

96 A Article Ericaud, A. and Bedhomme, A. L. and Morel, A. 2009 The optical properties of phytoplankton. *Environ Biol Fish*, 2009. DOI: 10.1007/BF00045215

97 A Article Ericaud, A. and Bedhomme, A. L. and Morel, A. 2010 The optical properties of phytoplankton. *Environ Biol Fish*, 2010. DOI: 10.1007/BF00045216

98 A Article Ericaud, A. and Bedhomme, A. L. and Morel, A. 2011 The optical properties of phytoplankton. *Environ Biol Fish*, 2011. DOI: 10.1007/BF00045217

99 A Article Ericaud, A. and Bedhomme, A. L. and Morel, A. 2012 The optical properties of phytoplankton. *Environ Biol Fish*, 2012. DOI: 10.1007/BF00045218

100 A Article Ericaud, A. and Bedhomme, A. L. and Morel, A. 2013 The optical properties of phytoplankton. *Environ Biol Fish*, 2013. DOI: 10.1007/BF00045219

[Phillips and Holley, 2001] Phillips, W. and Holley, C. E. (2001). Review of attitude representations used for aircraft kinematics. *Journal of Aircraft*, 38(4):718-727.

[Schaub et al., 1995] Schaub, R., Tokeras, P., and Jenkins, J. L. (1995). Principal rotation representations of proper NaN orthogonal matrices. *International Journal of Engineering Science*, 33(15):2277-2278.

[Shoemake, 1985] Shoemake, K. (1985). Animating rotation with quaternion curves. In *SIGGRAPH '85: Proceedings of the conference on Computer graphics and interactive techniques*, page 245-254. New York, NY, USA: ACM Press.

[Shoemake, 1991] Shoemake, K. (1991). Quaternions. Technical report, Department of Computer and Information Science, University of Pennsylvania, Philadelphia, PA 19104.

Données



Modélisation



Applications



Rendu



Industrialisation



# RÉSOLUTION FORMELLE DU PROBLÈME

## Formulaire

- Abramowitz & Stegun (1972) : [free pdf](#)
- NIST Digital Library of Mathematical Functions : <http://dlmf.nist.gov>

## Logiciels mathématiques

Nom	Système	Calcul formel	Orienté objet	Prix	Appel langage
maple	mac, win	oui	oui	\$2275	C, C#, Fortran
mathematica	mac, win	oui	oui	\$2495 <sup>1</sup>	Java, .NET, C++, Fortran, CUDA, OpenCL, R
sage	mac	oui	oui	gratuit	C, C++, other

1. Solveur en ligne : intégrale, dérivée.

## RÉSOLUTION NUMÉRIQUE - ALGÈBRE LINÉAIRE

La bibliothèque [Lapack](#) fournit une grande variété d'algorithme pour

- la résolution de système linéaire
- les solutions des moindres carrés
- le calcul de valeur propre
- la décomposition en valeur singulière
- les factorisations de matrices (LU, Cholesky, QR, SVD, Schur, generalized Schur)

le tout pour des matrices réelles ou complexes.

Malgré que la bibliothèque soit écrit en Fortran 90, elle est utilisée par

- Matlab
- Octave
- NAG Numerical Libraries
- Maple
- R project
- Scilab
- Ceemple

Données



Modélisation



Applications



Rendu



Industrialisation



## RÉSOLUTION NUMÉRIQUE - OPTIMISATION STANDARD

Algo standard	Type	R	SAS	matlab
Nelder-Mead	derivative free	optim	proc NLMIXED	fminsearch
BFGS	Hessian free	optim	proc NLMIXED	fminsearch
L-BFGS	large scale	optim	proc NLMIXED	

Algo avancé	R	SAS	matlab
opt. sous contrainte	pkg spécifique	proc optmodel	fmincon
opt. sur les entiers	Rsymphony	proc LP	optim. toolbox
algo génétique	rgenoud	proc genetic	pkg spécifique

Voir R, SAS, matlab

Données  
○○○Modélisation  
○○○○○  
○○○**Applications**  
●○○○Rendu  
○○  
○○○○○Industrialisation  
○○○○  
○○○○○

# MODÈLES DE RÉGRESSION ET DE SÉRIES TEMPORELLES

Modèle de régression	R	SAS	matlab
Modèle linéaire	lm	proc glm	fitlm
Modèle linéaire généralisé	glm	proc genmod	fitglm

Séries temporelles	R	SAS	matlab
Autorégressifs AR	ar	proc autoreg	arima
ARIMA	arima	% MA	arima
GARCH	garch	proc autoreg	garch

Données



Modélisation



Applications



Rendu

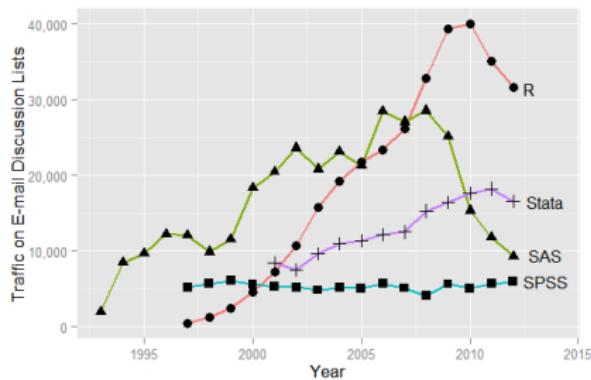
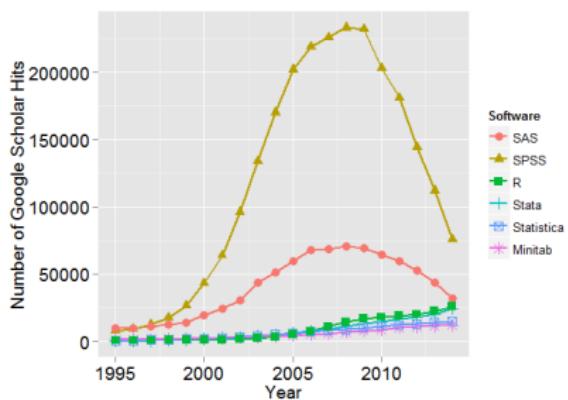


Industrialisation



# CONCENTRONS NOUS SUR R

Popularité de R (<http://r4stats.com/articles/popularity/>)



Nombre de packages dans R cf. [R bloggers](#)

- 2015-08-12 : 7000 packages
- 2013-11-08 : 5000 packages
- 2011-05-12 : 3000 packages
- 2007-04-12 : 1000 packages
- 2003-04-01 : 250 packages

- 2014-10-29 : 6000 packages
- 2012-08-23 : 4000 packages
- 2009-10-04 : 2000 packages
- 2004-10-01 : 500 packages

Données



Modélisation



Applications



Rendu



Industrialisation



# MODÈLES AVANCÉS DANS R

Pour chercher un package, il suffit de consulter les pages **task views**. La liste officielle est

Bayesian ChemPhys ClinicalTrials Cluster DifferentialEquations Distributions Econometrics Environmetrics

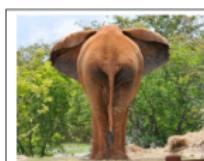
ExperimentalDesign Finance Genetics Graphics HighPerformanceComputing MachineLearning MedicalImaging MetaAnalysis Multivariate

NaturalLanguageProcessing NumericalMathematics OfficialStatistics Optimization Pharmacokinetics Phylogenetics Psychometrics

ReproducibleResearch Robust SocialSciences Spatial SpatioTemporal Survival TimeSeries WebTechnologies gR

## CRAN Task Views

CRAN Task Views are guides to the packages and functions useful for certain disciplines and methodologies. Many long-term R users I know have no idea they exist. As an effort to make them more widely known I thought I'd jazz up the index page. Images are free to use, and got from [SXC](#) stock photo site. Visual puns are mine. Task View links go to the [cran.r-project.org](http://cran.r-project.org) site and not a mirror.



### Bayesian Inference

Applied researchers interested in Bayesian statistics are increasingly attracted to R because of the ease of which one can code algorithms to sample...[\[more\]](#)



### Chemometrics and Computational Physics

Chemometrics and computational physics are concerned with the analysis of data arising in chemistry and physics experiments, as well as the simulation of...[\[more\]](#)



### Clinical Trial Design, Monitoring, and Analysis

This task view gathers information on specific R packages for design, monitoring and analysis of data from clinical trials. It focuses on including...[\[more\]](#)



### Cluster Analysis & Finite Mixture Models

This CRAN Task View contains a list of packages that can be used for finding groups in data and modelling unobserved cross-sectional heterogeneity. Many...[\[more\]](#)



### Probability Distributions

For most of the classical distributions, base R provides probability distribution functions ( $p$ ), density functions ( $d$ ), quantile functions ( $q$ ), and...[\[more\]](#)



### Analysis of Ecological and Environmental Data



### Design of Experiments (DoE) & Analysis of Experimental Data



### Computational Econometrics

# OÙ CHERCHER DE L'AIDE ?

## Docs

- officielle : <https://cran.r-project.org/manuals.html>
- non-officielle : <https://www.r-project.org/other-docs.html>
- recherche : <http://www.rdocumentation.org/>, package SOS

## Listes

- Please Help !! : [r-help@r-project.org](mailto:r-help@r-project.org)
- Développement de R ou des packages : [r-devel@r-project.org](mailto:r-devel@r-project.org) ou [R-package-devel@r-project.org](mailto:R-package-devel@r-project.org)
- Listes spécialisées : [r-sig-finance@r-project.org](mailto:r-sig-finance@r-project.org), [r-sig-insurance@r-project.org](mailto:r-sig-insurance@r-project.org), [r-sig-hpc@r-project.org](mailto:r-sig-hpc@r-project.org)

Données



Modélisation



Applications



Rendu



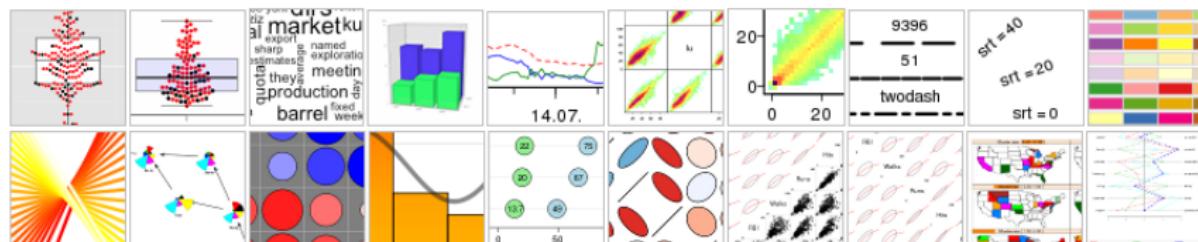
Industrialisation



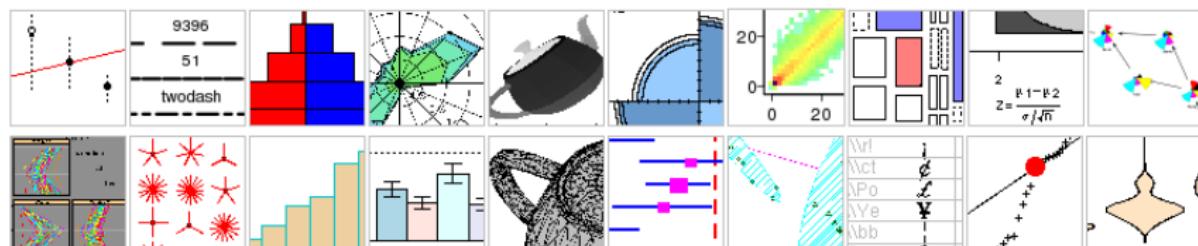
# GRAPHIQUES STATIQUES

Tout type de fichier image png(), jpeg(), bmp() et même pdf().

» Last entries ...



» Random entries



# GRAPHIQUES DYNAMIQUES ET AUTRE MULTIMÉDIA

R permet

- graphiques interactifs avec [SVG](#)
- film avec [animation](#)
- création de son avec [tuneR](#)
- analyse de son avec [seewave](#)

mais aussi

- [LATEX](#) Markup : les packages `reporttools`, `xtable`, `tables` fournissent les outils pour transformer les objets R en [LATEX](#). [exemple xtable, reporttools, et aussi](#)
- HTML Markup : les packages `R2HTML`, `knitr` permet de créer des documents html. [exemple markdown, R2HTML](#)
- Microsoft Formats : fichier words avec `R2wd`, fichier power point avec `R2PPT`, fichier excel (`dataframes2xls`, `WriteXLS`, `xlsReadWrite`).

Données



Modélisation



Applications



Rendu



Industrialisation



## SUR L'UTILITÉ DU PDF

Le [Portable Document Format](#), communément abrégé en PDF, est un langage de description de pages créé par la société Adobe Systems en 1993 disponible sur

- ordinateur : Windows, Mac OS, GNU/Linux,
- téléphone : iOS, Palm OS, Pocket PC, Windows Mobile, Symbian OS, Android,
- serveur : Sun Solaris Sparc, ...

Pour tout logiciel [WYSIWYG](#), il suffit d'imprimer en pdf : via [pdfcreator](#) ou directement le logiciel.

Pour [L<sup>A</sup>T<sub>E</sub>X](#) , c'est la sortie par défaut !

Données



Modélisation



Applications



Rendu



Industrialisation



# PDFS AVEC CODE – SWEAVE = R+LATEX

Voir <https://www.statistik.lmu.de/~leisch/Sweave/>

```

479 in the argument start), as a named list with initial values for each
480 (as they appear in the d, p, q functions).
481 Having defined reasonable starting values\footnote{}
482 The plotdist function can plot any
483 parametric distribution with specified parameter values in argument code
484 help to find correct initial values for the distribution parameters in non-
485 iterative calls if necessary (see the reference
486 manual for examples \ctyp{fitdistplus}).
487 }, %---
488 various distributions can be fitted and graphically
489 compared. On this example, the function \ctyp{cdfcomp} can be used to repre-
490 logsale so as to emphasize discrepancies on the tail of
491 interest while defining an HCS value (Figure-\ref{endo:comp}).
492
493
494 %%% R code
495 <<fendo.echo, echo=TRUE, fig=FALSE>>=
496 data("endosulfan")
497 ATV <- endosulfan$ATV
498 fendo.ln <- fitdist(ATV, "lnorm")
499 library("actuar")
500 fendo.ll <- fitdist(ATV, "llogis", start = list(shape = 1, scale = 500))
501 fendo.P <- fitdist(ATV, "pareto", start = list(shape = 1, scale = 500))
502 fendo.B <- fitdist(ATV, "burr", start = list(shape1 = 0.3, shape2 = 1,
503 rate = 1))
504 cdfcomp(list(fendo.ln, fendo.ll, fendo.P, fendo.B),
505 xlogscale = TRUE, ylogscale = TRUE,
506 legendtext = c("lognormal", "loglogistic", "Pareto", "Burr"))
507
508 % qqcomp(list(fendo.ln, fendo.ll, fendo.P, fendo.B), xlogscale=TRUE, ylog-
509 % scale=TRUE, legendtext = c("lognormal", "loglogistic", "Pareto", "Burr"))
510

```

488:25 [3] (Top Level) :

Console

RStudio: View PDF

~/Documents/recherche-enseignement/cod/

Page: 8 of 23 Automatic Zoom

**Empirical and theoretical CDFs**

CDF

lognormal loglogistic Pareto Burr

data in log scale

Figure 4: CDF plot to compare the fit of four distributions to acute toxicity values of various organisms for the organochlorine pesticide endosulfan (`endosulfan` data set) as provided by the `cdfcomp` function, with CDF values in a logscale so as to emphasize discrepancies on the tail left.

Données



Modélisation



Applications



Rendu



Industrialisation



# PDFS AVEC CODE – R MARKDOWN = R,LATEX,WIKI TYPE

Voir <http://rmarkdown.rstudio.com/>

example.Rmd x

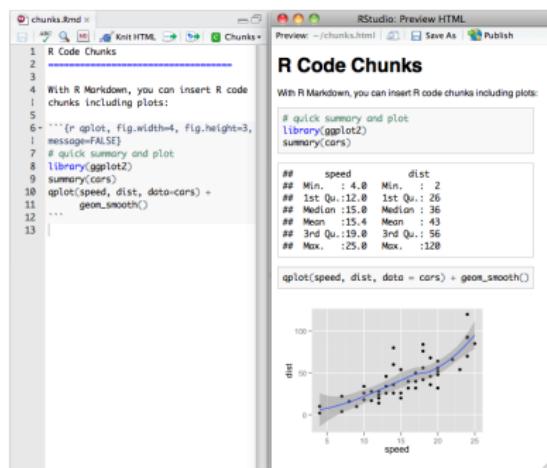
R Studio: Preview HTML

Preview: ~/example.html

```

1 Header 1
2
3 This is an R Markdown document. Markdown is a
| simple formatting syntax for authoring web pages.
4
5 Use an asterisk mark, to provide emphasis such as
| *italics* and **bold**.
6
7 Create lists with a dash:
8 - Item 1
9 - Item 2
10 - Item 3
11
12 You can write 'in-line' code with a back-tick.
13
14 ...
15 Code blocks display
16 with fixed-width font
17 ...
18
19 > Blockquotes are offset
20

```





# POSTER

Une classe poster de **LATEX** permet de réaliser des posters, voir ici

## Global Plasma Modelling of the Venus Express Particle and Magnetic Observations

Riku Järvinen, Esa Kallio, Ilkka Sillanpää, Kajun Liu and Pekka Janhunen

**Terrestrial planets**

Mercury	Venus	Earth	Mars
• No atmosphere	• Characteristics of Earth	• Atmosphere	• Atmosphere
• $1.6 \times 10^{-3} \text{ kg}$			
• $1.6 \times 10^{-3} \text{ kg}$			

**Wind from the Sun**

- Solar wind consists mainly of the proton-electron plasma and the magnetic field of the Sun.
- The solar wind interacts with the Earth's magnetic field.
- When the solar wind encounters planets, different kinds of interactions are formed between the solar wind and the planet's atmosphere.

**Venus-solar wind interaction**

- Reduced by the Venus solar wind interaction model (VSWIM) and the Venus Global Wind Model (VGWM).
- PVO plasma observations (Venus Express).
- PMO magnetic observations (Venus Express).

**Venus Express**

- Launch: November 2005
- All Venus: April 2006
- Orbit insertion: October 2006
- Size:  $1.85 \times 1.7 \times 1.4 \text{ m}$
- Launch mass: 1,270 kg
- Propulsion system: 100 N
- 1 instruments and 2 scientific themes including Venus Express
- Plasma package ASPERA-3 particle instruments
- Magnetometer MAG (3D components of the magnetic field)
- Magnetometer MAG (3D components of the magnetic field)

©ESA 2004. All rights reserved.

**Modélisation**

○○○○○  
○○○

**Applications**

○○○

**Rendu**

○○  
○○○●○○

**Industrialisation**

○○○○  
○○○○○

## Time-dependent quantum transport in the Kadanoff-Baym approach

Hans Marklund<sup>1</sup>, Adnan Imer<sup>2</sup>, Giacomo Bellini<sup>3</sup>, Robert van Leeuwen<sup>1</sup>

<sup>1</sup> Department of Physics, Niels Bohr Institute, Fysik 4014, University of Copenhagen, Denmark, France  
<sup>2</sup> Department of Physics, University of Rome Tor Vergata High Energy Physics Sciences, I-00133 Rome, Italy  
<sup>3</sup> European Theoretical Spectroscopy Facility (ETSF)

**Method**

Hydrodynamic equations:

$$\partial_t \mathbf{U} - \mathbf{A}((\mathbf{E}(t), \mathbf{B}(t)) \cdot \mathbf{U}) + \int d\mathbf{q} \mathcal{L}_{\text{coll}}(\mathbf{U}(t), \mathbf{U}(t)) = \mathcal{L}_{\text{ext}}(\mathbf{E}(t), \mathbf{B}(t))$$

Initial value effects:

$$\mathbf{U}(t) = \mathbf{U}_0 + \int_0^t \mathcal{L}_{\text{coll}}(\mathbf{U}(\tau), \mathbf{U}(\tau)) d\tau + \int_0^t \mathcal{L}_{\text{ext}}(\mathbf{E}(\tau), \mathbf{B}(\tau)) d\tau$$

Effects from the basis (perturbative):

$$\mathbf{U}(t) = \mathbf{U}_0 + \int_0^t \mathcal{L}_{\text{coll}}(\mathbf{U}_0, \mathbf{U}_0) d\tau + \int_0^t \mathcal{L}_{\text{ext}}(\mathbf{E}_0, \mathbf{B}_0) d\tau + \int_0^t \mathcal{L}_{\text{coll}}(\mathbf{U}_0, \mathbf{U}_1) d\tau + \int_0^t \mathcal{L}_{\text{ext}}(\mathbf{E}_0, \mathbf{B}_1) d\tau + \dots$$

Equation of motion for the Green function:

$$i\partial_t G_{\mu\nu} - \mathcal{L}_{\text{coll}}(G_{\mu\nu}, G_{\mu\nu}) + \int d\mathbf{q} \mathcal{L}_{\text{coll}}(G_{\mu\nu}, \mathcal{L}_{\text{coll}}(G_{\mu\nu}, \mathbf{U})) = 0$$

**Model**

Hybrid model:

$$\mathbf{U}(t) = \mathbf{U}_0 + \int_0^t \mathcal{L}_{\text{coll}}(\mathbf{U}_0, \mathbf{U}_0) d\tau + \int_0^t \mathcal{L}_{\text{ext}}(\mathbf{E}_0, \mathbf{B}_0) d\tau + \int_0^t \mathcal{L}_{\text{coll}}(\mathbf{U}_0, \mathbf{U}_1) d\tau + \int_0^t \mathcal{L}_{\text{ext}}(\mathbf{E}_0, \mathbf{B}_1) d\tau + \dots$$

Hybrid modelling is a way to get, treat and solve the kinetic theory and have the electron kinetics

**Quasi-neutral hybrid model**

Lorentz force for ions:  $m_i \frac{d\mathbf{v}_i}{dt} = q_i(\mathbf{E} + \mathbf{v}_i \times \mathbf{B})$

Momentum eq of electrons fluid:  $m_e \frac{d\mathbf{v}_e}{dt} = q_e(\mathbf{E} + \mathbf{v}_e \times \mathbf{B})$

$\mathbf{E} = \mathbf{E}_0 + \mathbf{E}_1 = \mathbf{E}_0 + \mathbf{E}_{\text{coll}}$

Quasi-neutrality condition:  $\rho_e = \sum_i n_i e_i = \text{const}$

Electromagnetism:  $\nabla \times \mathbf{B} = \mu_0 \mathbf{j}$

Total current density:  $\mathbf{j} = \sum_i q_i n_i \mathbf{v}_i$

**Results**

Hybrid model results:

- Hybrid model results provide a powerful kinetic model combining quantum transport and time-dependent classical hydrodynamics.
- Many hydrodynamic models have large effects on steady-state wave and transient spectra.
- Hybrid model results are in quantitative agreement with VSWIM and VGWM.
- The results show that the effects of the time-dependent currents

**Conclusions**

The hybrid model provides a powerful kinetic model combining quantum transport and time-dependent classical hydrodynamics.

Many hydrodynamic models have large effects on steady-state wave and transient spectra.

Hybrid model results are in quantitative agreement with VSWIM and VGWM.

The results show that the effects of the time-dependent currents

**References**

- Hans Marklund, Adnan Imer, Giacomo Bellini and Robert van Leeuwen, Phys Rev Lett 96, 105002 (2006)
- C. Deffner and C. Ott, Phys Rev A 73, 052102 (2006)
- C.-K. Peng, Phys Rev Lett 98, 105004 (2007)
- Y. Guo and R. van Leeuwen, Phys Rev Lett 98, 105005 (2007)

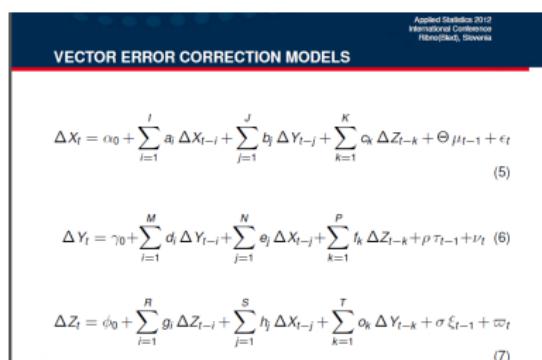
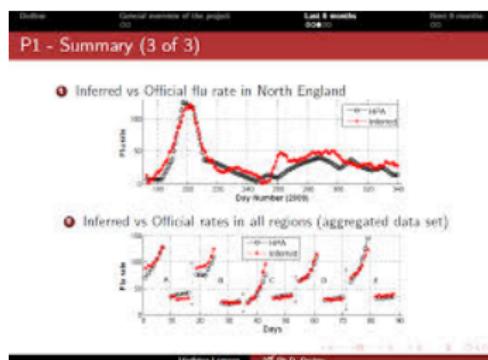
UNIVERSITY OF JYväSKYLÄ

28/39 – Christophe Dutang – Séminaire maths-éco

22/09/2015

# PRÉSENTATION

La classe beamer de L<sup>A</sup>T<sub>E</sub>X permet de réaliser des présentations, voir ici



Données  
○○○

Modélisation  
○○○○○  
○○○

Applications  
○○○○

Rendu  
○○  
○○○○○

Industrialisation  
●○○○  
○○○○○

## SIMULATION

Des simulations démarrent avec de bons générateurs aléatoires. Tous les logiciels sauf excel utilisent l'algorithme Mersenne Twister

Logiciel	R	matlab	SAS	scilab	excel
Algo U(0,1)	MT19937	MT19937	MT19937	MT19937	$\approx$ Wichman-Hill <sup>2</sup>
Période	$2^{19937} - 1$	$2^{19937} - 1$	$2^{19937} - 1$	$2^{19937} - 1$	$6.9536e12$
Init seed	oui	oui	oui	oui	non
Algo N(0,1)	inv. num.	Marsaglia-Tang		Box-Muller	
Parallèle	oui	oui	oui		

- R propose une très grande variété des lois ou de modèles dans des [pkg spécifiques](#).
- matlab propose moins de lois mais plus [par défaut](#).
- [comparatif R et SAS](#)
- Attention, excel n'est pas conçu pour la simulation.

2. [1] prouve qu'excel n'implémente pas WH.

Données



Modélisation



Applications



Rendu



Industrialisation



## PARALLÉLISATION EN R

Nombreuses options existent pour paralléliser le code

- parallélisation explicite : package `parallel` de R mais aussi `Rmpi`, `snow`

temps (sec)	R (base)	parallel 1	parallel 2	parallel 3	parallel 4
10 échant.	0.97	1.03	0.53	0.49	0.40
100 échant.	9.42	10.26	5.08	4.20	3.33
1000 échant.	90.71	101.97	50.01	41.16	32.71

TABLE: Moyenne d'échantillons gaussien

- parallélisation implicite par `pnmath`, ...
- GPUs : `gputools` fournit des outils, `WideLM` des modèles de régression, `gmatrix` des opérations de matrices, mais nécessite une architecture matérielle particulière.

## CODE COMPILE

Quand la parallélisation est impossible ? ⇒ langages compilés

De base, on peut utiliser

- C via un package R : chapitre 6 de [r-exts.pdf](#)
- Fortran via un package R : chapitre 6 de [r-exts.pdf](#)

permet d'accélérer le code, d'utiliser la librairie LAPACK, les versions C des fonctions R (optimisation, fonction spéciales,...)

D'autres packages sont disponibles

- [inline](#) pour utiliser du code C, Fortran, C++ directement dans R.
- [Rcpp](#) pour du code C++.
- [rJava](#) pour du code Java.

matlab et SAS proposent des fonctionnalités similaires pour intégrer du code C, Fortran, C++.

Données



Modélisation



Applications



Rendu



Industrialisation

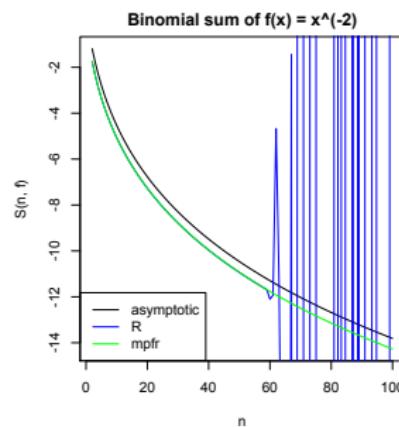
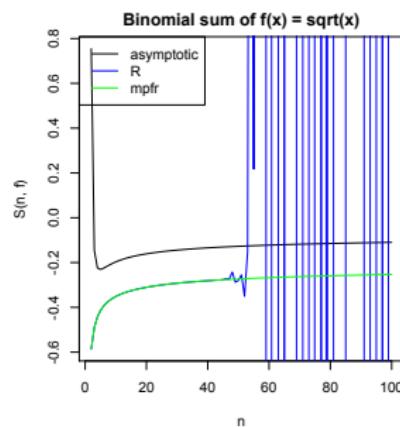


## CALCUL HAUTE PRÉCISION

Comment calculer la somme binomiale alternée par

$$S_n(f) = \sum_{k=n_0}^n C_n^k (-1)^k f(k),$$

où  $C_n^k = \frac{n!}{k!(n-k)!}$  et  $f$  une fonction, par exemple  $\log(\cdot)$ ,  $\sqrt{\cdot}$  ou un polynôme.  
 Utilisez les bibliothèques [GMP](#), [mpfr](#) via le package [Rmpfr](#).



Données

○○○

Modélisation

○○○○○  
○○○

Applications

○○○○

Rendu

○○  
○○○○○

Industrialisation

○○○○○  
●○○○○

## FAIRE UN PACKAGE

### Pourquoi faire ?

- documenter un groupe de fonctionnalités (pour ne pas oublier),
- faciliter la maintenance,
- partager avec d'autres.

### Comment ?

- en R, un modèle est fourni avec `package.skeleton`
  - fichiers DESCRIPTION, NAMESPACE
  - dossiers R, data, inst, man, src, tests, vignettes
- en matlab, [http://fr.mathworks.com/help/matlab/matlab\\_oop/scoping-classes-with-packages.html](http://fr.mathworks.com/help/matlab/matlab_oop/scoping-classes-with-packages.html)
  - "Using a package provides a means to organize classes and functions."
  - et la doc ?

### Les bonnes pratiques

- Ne jamais utiliser d'accent
- Réaliser des tests unitaires
- Bien choisir les noms de fonction et les documenter

## LES SYSTÈMES DE VERSIONING

Pour n'importe quel projet collaboratif, les conflits de version sont ingérables.  
Des solutions existents : **versioning** et **comparaison**

- **svn** fonctionne sur le mode client-serveur, avec :
  - un serveur informatique centralisé et unique où se situent : les fichiers constituant la référence (dépot), un logiciel serveur Subversion ;
  - des postes clients sur lesquels se trouvent : les fichiers recopiés depuis le serveur, éventuellement modifiés localement depuis, un logiciel client SVN (TortoiseSVN,...).
- **git** est une version décentralisé de type svn très performante.
- **des versions payantes existent**

Des plateformes conviviales permettent de les utiliser

- **github** est un service web d'hébergement, gratuit pour des dépôts publics.
- **bitbucket** est un service web d'hébergement, gratuit pour des dépôts publics et privés (dans une certaine mesure).
- **et d'autres**

Données

○○○

Modélisation

○○○○○  
○○○

Applications

○○○○

Rendu

○○  
○○○○○

Industrialisation

○○○○○  
○○●○○

## CONCLUSION

Pour quel prix ?

- R : 0 €
- maple : \$2275 ?
- mathematica : \$2495 ?
- matlab : 2000 € (individuel), 500 € (académique), 35 € (étudiant) + toolbox,
- sage : 0 €
- SAS : \$8,700 ? + packages
- scilab : 0 €

*"Let's not kid ourselves : the most widely used piece of software for statistics is Excel."*

B.D. Ripley, Opening lecture RSS 2002, Plymouth

# CONCLUSION

Pour quel rapidité de calcul ? un exemple tiré de Aruoba & Villaverde (2015)

Average and relative run time (Seconds).

Language	Mac			Windows		
	Version/Compiler	Time	Rel. Time	Version/Compiler	Time	Rel. Time
C++	GCC-4.9.0	0.73	1.00	Visual C++ 2010	0.76	1.00
	Intel C++ 14.0.3	1.00	1.38	Intel C++ 14.0.2	0.90	1.19
	Clang 5.1	1.00	1.38	GCC-4.8.2	1.73	2.29
Fortran	GCC-4.9.0	0.76	1.05	GCC-4.8.1	1.73	2.29
	Intel Fortran 14.0.3	0.95	1.30	Intel Fortran 14.0.2	0.81	1.07
Java	JDK8u5	1.95	2.69	JDK8u5	1.59	2.10
Julia	0.3.7	1.91	2.62	0.3.7	1.80	2.37
Matlab	2014a	7.91	10.88	2014a	6.74	8.92
Python	Pypy 2.2.1	31.90	43.86	Pypy 2.2.1	34.14	45.16
	CPython 2.7.6	195.87	269.31	CPython 2.7.4	117.40	155.31
R	3.1.1, compiled	204.34	280.90	3.1.1, compiled	184.16	243.63
	3.1.1, script	345.55	475.10	3.1.1, script	371.40	491.33
Mathematica	9.0, base	588.57	809.22	9.0, base	473.34	626.19
Matlab, Mex	2014a	1.19	1.64	2014a	0.98	1.29
Rcpp	3.1.1	2.66	3.66	3.1.1	4.09	5.41
Python	Numba 0.13	1.18	1.62	Numba 0.13	1.19	1.57
	Cython	1.03	1.41	Cython	1.88	2.49
Mathematica	9.0, idiomatic	1.67	2.29	9.0, idiomatic	2.22	2.93

Voir aussi [matloff](#), [kaggle](#)

# RÉFÉRENCES

-  McCullough, B. D. (2008), Microsoft excel's 'not the wichmann–hill' random number generators Computational Statistics and Data Analysis 52, 4587-4593. 10, 26
-  Aruoba, S.B., Fernandez-Villaverde, J. (2015), A comparison of programming languages in macroeconomics Journal of Economic Dynamics & Control 58, 265-273.
-  Chang, W. (2013), R graphics cookbook, O'Reilly
-  O' Connor, Brendan (2014) Comparison of data analysis packages.  
<http://brencon.com/blog>
-  Jain, K. (2014) SAS vs. R (vs. Python) – which tool should I learn ?.  
<http://www.analyticsvidhya.com/blog>
-  Wikipedia  
Comparison of statistical packages, Comparison of versioning software, Comparison of reference software
-  CRAN  
site officiel, book list, task view
-  CTAN  
site officiel, book list,
-  Maechler, M. (2014). Rmpfr : R MPFR - Multiple Precision Floating-Point Reliable. R package version 0.5-7

## REGARDER LES SOURCES DE R

- Rendez vous sur <https://cran.r-project.org/src/base/>,
- Choisissez la dernière version.
- Dézipper et chercher le dossier ./src/
  - ./src/library/ contient les packages de base : base, graphics,...
  - ./src/main/ contient les codes C appelés par le package de base
  - ./src/include/ les fichiers header (.h) listant ce qu'on peut appeler depuis un autre package en C.
  - ./src/extral et ./src/appl/ les fichiers C ou fortran de certaines librairies.
  - etc...
- Regardez les doc officielles :  
<https://cran.r-project.org/manuals.html>