



**Le logiciel open-source R en génie chimique**  
**Des plans d'expériences au big data**

**patrice.kiener@inmodelia.com**

**Cnam Paris – 5 septembre 2018**

## Historiquement, la chimie et le génie chimique utilisent des logiciels spécialisés :

- Souvent des logiciels commerciaux
- Diagramme de phases, Spectrométrie, Thermodynamique, Mécanique des fluides, Mélangeage, Échange de chaleur, Distillation, etc
- Codés dans les langages de bas niveau : C, C++, Fortran, Delphi
- Avec des interfaces graphiques (⇒ attention portée à l'ergonomie)
- Souvent sous Windows

## Alternative : utiliser des logiciels généralistes avec des packages dédiés :

- Souvent des logiciels gratuits avec code source disponible (open source)
- Perl (30 ans), Python (20 ans), R (20 ans), Julia (10 ans)
- Multi-plateformes : Windows, Mac, Linux, BSD, Solaris
- Avec ou sans interface graphique
- Vu davantage comme du traitement de données que comme du logiciel applicatif
- Qui embrassent un champ très large  $\Rightarrow$  du plan d'expériences au big data

## R est à la fois un langage et un logiciel

<https://www.r-project.org> + <https://cran.univ-paris1.fr>

- Au départ était le logiciel commercial **S** créé par les Bell Labs pour Unix : version 1 en 1976, v2 en 1981, v3 en 1983, v4 en 1995
- 1993 : Ross Ihaka et Robert Gentleman de l'Université d'Auckland créent **R**
- 1995 : Licence GPL
- 1997 : Dépôt **CRAN** (3 miroirs) + R en version 0.49 + 12 packages
- 1999 : Fonctions *install.packages()*, *update.packages()*
- 2000 : Version 1.0 le 29 février, v2.0 en 2004, v3.0 en 2013, v3.5.1 le 2 juillet 2018
- 2004 : Première conférence internationale à Vienne (Autriche)
- Développement actif : environ 3 nouvelles versions par an
- Multi-plateformes 32/64 bits : Windows, Mac, Linux, Solaris, FreeBSD, OpenBSD
- Utilise actuellement gcc 4.9.3. Dans un futur proche gcc 8.1.

**CRAN = Comprehensive R Archive Networks**

**1 serveur central à l'Université Technique de Vienne (Autriche)**  
**95 serveurs miroirs dans le monde (85 universités, 10 entreprises)**



## Accès libre : tout le monde peut développer un package R (Et c'est très facile)

|            |           |            |            |            |             |
|------------|-----------|------------|------------|------------|-------------|
| 1997-04-23 | 12 pkgs   |            | 2013-11-08 | 5000 pkgs  | (+69/mois)  |
| 2002-09-17 | 68 pkgs   | (+1/mois)  | 2014-10-29 | 6000 pkgs  | (+86/mois)  |
| 2003-04-01 | 250 pkgs  | (+24/mois) | 2015-08-12 | 7000 pkgs  | (+106/mois) |
| 2004-10-01 | 500 pkgs  | (+14/mois) | 2016-02-29 | 8000 pkgs  | (+151/mois) |
| 2007-04-12 | 1000 pkgs | (+16/mois) | 2016-08-22 | 9000 pkgs  | (+174/mois) |
| 2009-10-04 | 2000 pkgs | (+34/mois) | 2017-01-27 | 10000 pkgs | (+193/mois) |
| 2011-05-12 | 3000 pkgs | (+52/mois) | 2017-07-04 | 11000 pkgs | (+193/mois) |
| 2012-08-23 | 4000 pkgs | (+65/mois) | 2017-12-15 | 12000 pkgs | (+185/mois) |
|            |           |            | 2017-08-31 | 13001 pkgs | (+118/mois) |

Août 2015 (2018) : 7000 (13000) packages gérés par 4130 (7730) mainteneurs différents.

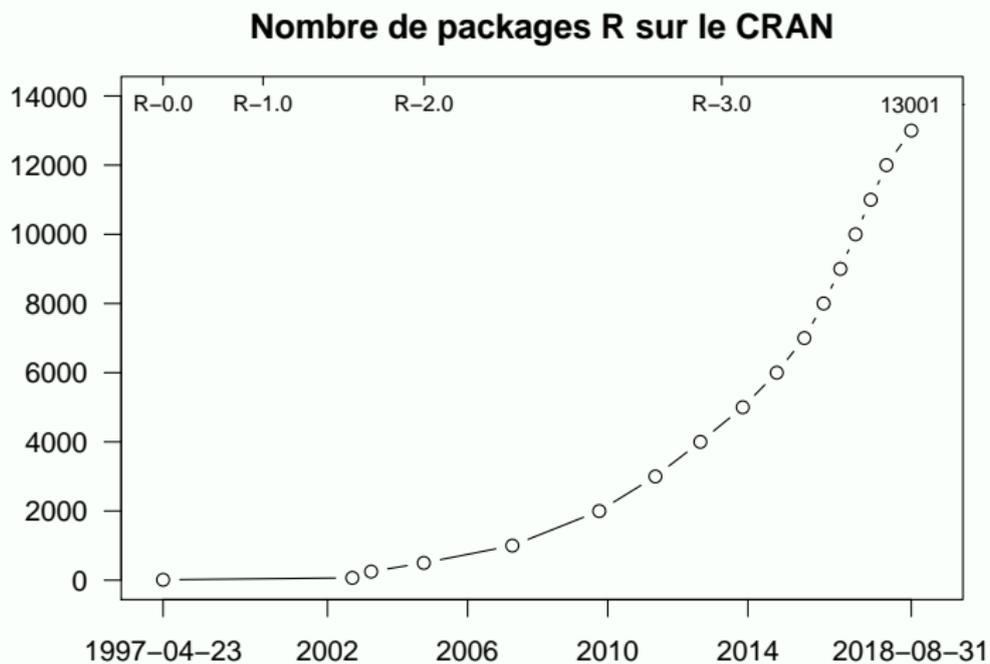
Plus de 300 packages connectent R à des bibliothèques externes libres ou commerciales<sup>1</sup>.

Il existe un autre dépôt, Bioconductor, dédié à la génomique.

---

1. En janvier 2015, Microsoft a racheté la société Revolution Analytics. En août 2016, Microsoft annonçait que SQL Server était totalement connecté en import et en export à R, donnant ainsi aux utilisateurs payants de SQL Server un accès direct à tout l'environnement gratuit R. Voir : <https://docs.microsoft.com/fr-FR/sql/advanced-analytics/r/sql-server-r-services?view=sql-server-2016>

# 13.000 packages R : une croissance quasi-exponentielle



# La page web du CRAN qui présente le package *ggtern*

http://cran.univ-paris1.fr/web/packages/ggtern/index.html

## ggtern: An Extension to 'ggplot2', for the Creation of Ternary Diagrams

Extends the functionality of 'ggplot2', providing the capability to plot ternary diagrams for (subset of) the 'ggplot2' geometries. Additionally, 'ggtern' has implemented several NEW geometries which are unavailable to the standard 'ggplot2' release. For further examples and documentation, please proceed to the 'ggtern' website.

Version: 3.0.0  
Depends: [ggplot2](#) (≥ 3.0.0)  
Imports: [compositions](#) (≥ 1.40-2), [grid](#), [gridExtra](#), [gtable](#) (≥ 0.1.2), [latex2exp](#) (≥ 0.4), [MASS](#), [plyr](#) (≥ 1.8.3), [scales](#) (≥ 0.3.0), [stats](#), [proto](#), [utils](#), [lattice](#)  
Enhances: [sp](#)  
Published: 2018-08-22  
Author: Nicholas Hamilton [aut, cre]  
Maintainer: Nicholas Hamilton <[nick at ggtern.com](mailto:nick at ggtern.com)>  
License: [GPL-2](#) | file [LICENSE](#)  
URL: <http://www.ggtern.com>

NeedsCompilation: no

Materials: [README NEWS](#)

CRAN checks: [ggtern results](#) ⇒

Downloads:

Reference manual: [ggtern.pdf](#)

Package source: [ggtern\\_3.0.0.tar.gz](#)

Windows binaries: r-devel: [ggtern\\_3.0.0.zip](#), r-release: [ggtern\\_3.0.0.zip](#), r-oldrel: [ggtern\\_3.0.0.zip](#)

OS X binaries: r-release: [ggtern\\_3.0.0.tgz](#), r-oldrel: [ggtern\\_3.0.0.tgz](#)

Old sources: [ggtern archive](#)

Reverse dependencies:

Reverse imports: [tricolore](#)

Web cran.univ-paris1.fr/web/checks/check\_results\_ggtern.html

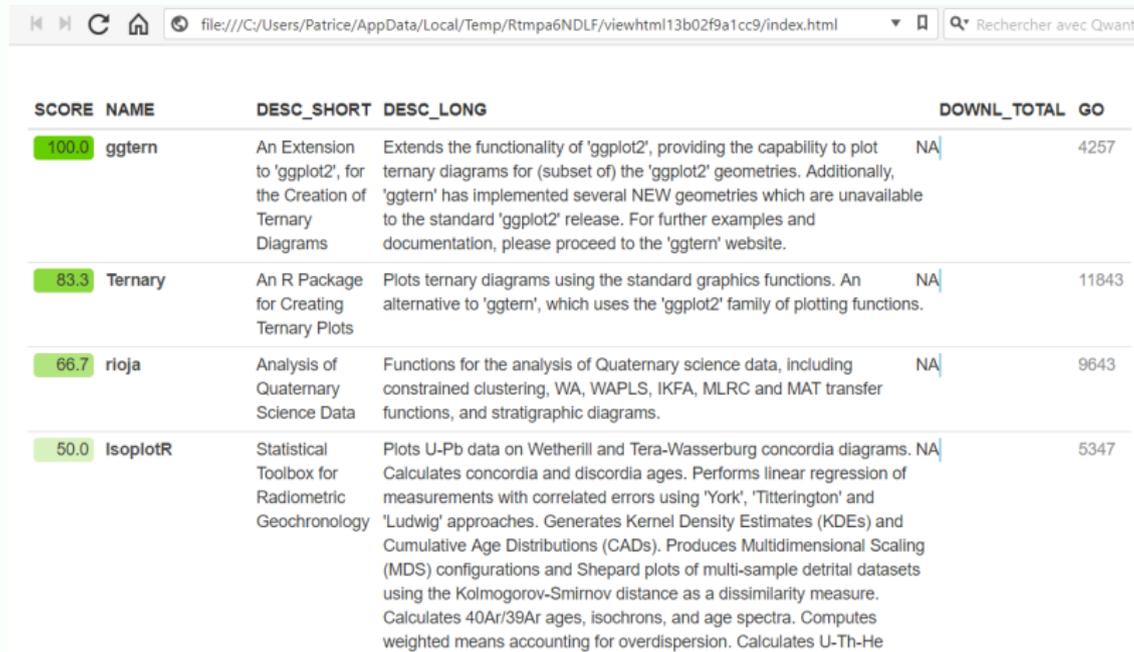
### CRAN Package Check Results for Package *ggtern*

Last updated on 2018-09-03 07:47:11 CEST.

| Flavor                                            | Version | T <sub>install</sub> | T <sub>check</sub> | T <sub>total</sub> | Status | Flags |
|---------------------------------------------------|---------|----------------------|--------------------|--------------------|--------|-------|
| <a href="#">r-devel-linux-x86_64-debian-clang</a> | 3.0.0   | 11.77                | 111.21             | 122.98             | OK     |       |
| <a href="#">r-devel-linux-x86_64-debian-gcc</a>   | 3.0.0   | 10.18                | 87.04              | 97.22              | OK     |       |
| <a href="#">r-devel-linux-x86_64-fedora-clang</a> | 3.0.0   |                      |                    | 147.46             | OK     |       |
| <a href="#">r-devel-linux-x86_64-fedora-gcc</a>   | 3.0.0   |                      |                    | 144.41             | OK     |       |
| <a href="#">r-devel-windows-i386-x86_64</a>       | 3.0.0   | 33.00                | 185.00             | 218.00             | OK     |       |
| <a href="#">r-patched-linux-x86_64</a>            | 3.0.0   | 11.07                | 108.52             | 119.59             | OK     |       |
| <a href="#">r-patched-solaris-x86</a>             | 3.0.0   |                      |                    | 196.10             | OK     |       |
| <a href="#">r-release-linux-x86_64</a>            | 3.0.0   | 13.94                | 109.50             | 123.44             | OK     |       |
| <a href="#">r-release-windows-i386-x86_64</a>     | 3.0.0   | 31.00                | 182.00             | 213.00             | OK     |       |
| <a href="#">r-release-oss-x86_64</a>              | 3.0.0   |                      |                    |                    | OK     |       |
| <a href="#">r-oldrel-windows-i386-x86_64</a>      | 3.0.0   | 12.00                | 154.00             | 166.00             | OK     |       |
| <a href="#">r-oldrel-oss-x86_64</a>               | 3.0.0   |                      |                    |                    | OK     |       |

# La documentation :

`packagefinder::findPackage(c("ternary", "diagram"), "and")`  
retourne une page html des packages



| SCORE | NAME     | DESC_SHORT                                                      | DESC_LONG                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           | DOWNL_TOTAL | GO    |
|-------|----------|-----------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------|-------|
| 100.0 | ggtern   | An Extension to 'ggplot2', for the Creation of Ternary Diagrams | Extends the functionality of 'ggplot2', providing the capability to plot ternary diagrams for (subset of) the 'ggplot2' geometries. Additionally, 'ggtern' has implemented several NEW geometries which are unavailable to the standard 'ggplot2' release. For further examples and documentation, please proceed to the 'ggtern' website.                                                                                                                                                                                                                                                                                          | NA          | 4257  |
| 83.3  | Ternary  | An R Package for Creating Ternary Plots                         | Plots ternary diagrams using the standard graphics functions. An alternative to 'ggtern', which uses the 'ggplot2' family of plotting functions.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    | NA          | 11843 |
| 66.7  | rioja    | Analysis of Quaternary Science Data                             | Functions for the analysis of Quaternary science data, including constrained clustering, WA, WAPLS, IKFA, MLRC and MAT transfer functions, and stratigraphic diagrams.                                                                                                                                                                                                                                                                                                                                                                                                                                                              | NA          | 9643  |
| 50.0  | IsoplotR | Statistical Toolbox for Radiometric Geochronology               | Plots U-Pb data on Wetherill and Tera-Wasserburg concordia diagrams. Calculates concordia and discordia ages. Performs linear regression of measurements with correlated errors using 'York', 'Titterton' and 'Ludwig' approaches. Generates Kernel Density Estimates (KDEs) and Cumulative Age Distributions (CADs). Produces Multidimensional Scaling (MDS) configurations and Shepard plots of multi-sample detrital datasets using the Kolmogorov-Smirnov distance as a dissimilarity measure. Calculates 40Ar/39Ar ages, isochrons, and age spectra. Computes weighted means accounting for overdispersion. Calculates U-Th-He | NA          | 5347  |

# La documentation :

`sos::findFn("ternary diagram")`  
retourne une page html des fonctions

file:///C:/Users/Patrice/AppData/Local/Temp/RtmpGOGQt2/file30b0752633.html

## findFn Results

call: "x <- sos::findFn(string = 'ternary diagram')"

For a summary by package, see: "packageSum(x,...)"

See also: vignette('sos')

| Id | Count | MaxScore | TotalScore | Package         | Function           | Date | Score | Description and Link                                                            |
|----|-------|----------|------------|-----------------|--------------------|------|-------|---------------------------------------------------------------------------------|
| 1  | 7     | 531      | 716        | Compositional   | ternary            | NA   | 531   | <a href="#">Ternary diagram</a>                                                 |
| 2  | 7     | 531      | 716        | Compositional   | 00Index            | NA   | 60    | <a href="#">Compositional Data Analysis</a>                                     |
| 3  | 7     | 531      | 716        | Compositional   | bivt.contour       | NA   | 29    | <a href="#">Contour plot of the t distribution in S^2</a>                       |
| 4  | 7     | 531      | 716        | Compositional   | norm.contour       | NA   | 29    | <a href="#">Contour plot of the normal distribution in S^2</a>                  |
| 5  | 7     | 531      | 716        | Compositional   | skewnorm.contour   | NA   | 29    | <a href="#">Contour plot of the skew skew-normal distribution in S^2</a>        |
| 6  | 7     | 531      | 716        | Compositional   | comp.kerncontour   | NA   | 19    | <a href="#">Contour plot of the kernel density estimate in S^2</a>              |
| 7  | 7     | 531      | 716        | Compositional   | diri.contour       | NA   | 19    | <a href="#">Contour plot of a Dirichlet distribution in S^2</a>                 |
| 8  | 6     | 485      | 1938       | robCompositions | ternaryDiag        | NA   | 485   | <a href="#">Ternary diagram</a>                                                 |
| 9  | 6     | 485      | 1938       | robCompositions | ternaryDiagAbline  | NA   | 443   | <a href="#">Adds a line to a ternary diagram.</a>                               |
| 10 | 6     | 485      | 1938       | robCompositions | ternaryDiagEllipse | NA   | 433   | <a href="#">Adds tolerance ellipses to a ternary diagram.</a>                   |
| 11 | 6     | 485      | 1938       | robCompositions | ternaryDiagPoints  | NA   | 426   | <a href="#">Add points or lines to a given ternary diagram.</a>                 |
| 12 | 6     | 485      | 1938       | robCompositions | 00Index            | NA   | 91    | <a href="#">Robust Estimation for Compositional Data</a>                        |
| 13 | 6     | 485      | 1938       | robCompositions | plot.imp           | NA   | 60    | <a href="#">Plot method for objects of class imp</a>                            |
| 14 | 5     | 841      | 1832       | ggtern          | 00Index            | NA   | 841   | <a href="#">An Extension to 'ggplot2', for the Creation of Ternary Diagrams</a> |
| 15 | 5     | 841      | 1832       | ggtern          | ggtern_package     | NA   | 461   | <a href="#">Ternary Diagrams in R</a>                                           |
| 16 | 5     | 841      | 1832       | ggtern          | theme_rotate       | NA   | 423   | <a href="#">Rotate Ternary Diagram</a>                                          |

## Les « Task Views » : une présentation des packages par thème

(Recensement non exhaustif réalisé par des bénévoles)

<https://www.r-project.org/taskviews.html>

| pkg        | Task view                 | pkg        | Task view                   | pkg        | Task view            |
|------------|---------------------------|------------|-----------------------------|------------|----------------------|
| 135        | Bayesian                  | 27         | Genetics                    | 12         | Pharmacokinetics     |
| <b>93</b>  | <b>ChemPhys</b>           | 42         | Graphics                    | 87         | Phylogenetics        |
| 53         | ClinicalTrials            | <b>99</b>  | <b>HighPerformComputing</b> | 203        | Psychometrics        |
| 109        | Cluster                   | 101        | MachineLearning             | 60         | ReproducibleResearch |
| <b>29</b>  | <b>DifferentialEquat.</b> | 34         | MedicalImaging              | 60         | Robust               |
| 207        | Distributions             | 99         | MetaAnalysis                | 82         | SocialSciences       |
| 140        | Econometrics              | 31         | ModelDeployment             | 176        | Spatial              |
| 105        | Environmetrics            | 121        | Multivariate                | 64         | SpatioTemporal       |
| <b>124</b> | <b>ExperimentalDesign</b> | 55         | NaturalLanguageProces.      | 269        | Survival             |
| 24         | ExtremeValue              | 94         | NumericalMathematics        | <b>257</b> | <b>TimeSeries</b>    |
| 152        | Finance                   | 116        | OfficialStatistics          | 246        | WebTechnologies      |
| 39         | FunctionalData            | <b>125</b> | <b>Optimization</b>         | 36         | gR                   |

3700 packages répertoriés dont **3000** uniques. Très loin des 13.000 packages !

# La task view *ChemPhys* : 93 packages recensés

https://cran.univ-paris1.fr/web/views/ChemPhys.html

CRAN Task View: Chemometrics and Computational Physics

**Maintainer:** Katharine Mullen

**Contact:** [katharine.mullen@stat.ucla.edu](mailto:katharine.mullen@stat.ucla.edu)

**Version:** 2018-08-20

**URL:** <https://CRAN.R-project.org/view=ChemPhys>

Chemometrics and computational physics are concerned with the analysis of data arising in chemistry and physics experiments, as well as the simulation of physico-chemico systems. Many of the functions in base R are useful for these ends.

The book *Chemometrics with R* by Ron Wehrens, ISBN: 978-3-642-17840-5, Springer, 2011, provides an introduction to multivariate statistics in the life sciences, as well as coverage of several specific topics from the area of chemometrics; the examples in the book are possible to reproduce using the package [ChemometricsWithR](#).

The book *Modern Statistical Methods for Astronomy With R Applications* by Eric D. Feigelson and G. Jogesh Babu, ISBN-13: 9780521767279, Cambridge, 2012, provides an introduction to statistics for astronomers and an overview of the foremost methods being used in astrostatistical analysis, illustrated by examples in R.

The book by Kurt Varmuza and Peter Filzmoser, *Introduction to Multivariate Statistical Analysis in Chemometrics*, ISBN 978-1-420-05947-2, CRC Press, 2009, is associated with the package [chemometrics](#).

A special issue of R News with a focus on [R in Chemistry](#) was published in August 2006. A special volume of Journal of Statistical Software (JSS) dedicated to [oscopy and Chemometrics in R](#) was published in January 2007.

[Please let us know](#) if we have omitted something of importance, or if a new package or function should be mentioned here.

## Linear Regression Models

- Linear models can be fitted (via OLS) with `lm()` (from stats). A least squares solution for  $x$  in  $Ax = b$  can also be computed as `qr.coef(qr(A), b)`.
- The package [nls](#) provides a means of constraining  $x$  to non-negative or non-positive values; the package [bvlx](#) allows other bounds on  $x$  to be applied.
- Functions for isotonic regression are available in the package [Iso](#), and are useful to determine the unimodal vector that is closest to a given vector  $x$  under least squares criteria.
- Heteroskedastic linear models can be fit using the `glm()` function of the [nlme](#) package.

## Nonlinear Regression Models

## Nombre de packages contenant les « mots » du génie chimique parmi les 13.000 packages du CRAN

| pkg | Mot                         | pkg | Mot                  | pkg | Mot             |
|-----|-----------------------------|-----|----------------------|-----|-----------------|
| 0   | <b>chemical engineering</b> | 13  | molecule             | 3   | mixing          |
| 1   | transfer function           | 8   | <b>kinetic</b>       | 2   | equilibrium     |
| 1   | heat transfer               | 7   | gas                  | 1   | hydrodynamic    |
| 1   | grain size                  | 7   | pressure             | 1   | liquid          |
| 2   | phase diagram               | 7   | chromatography       | 1   | crystallization |
| 3   | <b>chemical reaction</b>    | 7   | ternary              | 0   | combustion      |
| 8   | process control             | 5   | thermal              | 0   | distillation    |
| 35  | <b>spectrometry</b>         | 5   | geochemical          | 0   | exchanger       |
| 25  | chemical                    | 3   | biochemical          | 0   | oxydation       |
| 16  | chemistry                   | 3   | QSAR                 | 0   | polymerisation  |
| 14  | reaction                    | 3   | pH                   | 0   | separation      |
| 14  | molecular                   | 3   | <b>thermodynamic</b> | 0   | bond graph      |

R est davantage tiré par la génétique, la biochimie, l'environnement, la chimie analytique.

R est bien plus fort sur les équations différentielles.

## Molécules : Accès aux bases de données web, SMILES

- **webchem** : Lecture des bases de données web (CIR, ChemIDPlus, ChemSpider ID, CTS, ETOX, flavornet, PAN, PubChem, PHYSPROP, Wikidata) dans différents formats (CAS, CDK, InChIKey, Molfile, SMILES)
- **rpubchem, iqspr** : Lecture des SMILES de Pubchem, Génération de SMILES
- **rcdk, rcdklibs** : Manipulation des données CDK
- **BMRBr** : Base des données en RMN de de biomolécules (peptides, protéines)
- **isopat** : Base de données et calcul des isotopes de molécules

## QSAR/QSPR, Dynamique moléculaire, Propriétés, Isotopes :

- **PeriodicTable** : la table des éléments périodiques
- **DEMOVA, omd, QSARdata** : des outils autour du QSAR/QSPR
- **modygliani, MDplot** : Dynamique moléculaire
- **ChargeTransport** : Modèles de Marcus-Levich-Jortner et Landau-Zener
- **ArArRedux** : Données sur l'argon Ar36, Ar37, Ar38, Ar39, Ar40.

## pH, Thermodynamique, thermique

- **titrationCurves, AmmoniaConcentration** : pH de solutions, ph de l'ammoniac
- **CHNOSZ** : Équations HKF pour molécules hydrophiles + base de données
- **aiRthermo, humidity, IAPWS95** : Propriétés de l'air, de la vapeur d'eau
- **SorptionAnalysis** : Modèles d'adsorption de Langmuir, Freundlich, Temkin
- **zFactor** : Facteur de compression z des gaz d'hydrocarbures
- **Ritc** : Analyse calorimétrique et courbes d'isothermes
- **Thermimage** : Analyse des images thermiques

## Diagrammes binaires et ternaires :

- **LLSR** : Analyse des systèmes liquide-liquide
- **ggtern, Ternary, trifold** : Diagrammes ternaires
- **rysgran, robCompositions, mixexp** : Taille des grains, Modèles de mélange

## Procédés :

- **biogas, aRpsDCA** : Production de biogaz, Puits de pétrole et gaz
- **renpow** : Énergies renouvelables
- **eChem** : Électrochimie
- **flux, gasfluxes, eixport** : Gaz à effets de serre et polluants atmosphériques
- **AquaEnv, seacarb, marelac** : Eau de mer
- **RadTran** : Transport du Radon dans les milieux poreux 2D

## Modèles cinétiques, Modèles :

- **RxnSim** : Similarité entre deux réactions chimiques
- **kinfit, mkin, EKCMC** : Modèles de cinétique chimique (+ Chaines de Markov)
- **pauwels2014, smfsb, marelac** : Cinétique des systèmes biologiques
- **bioPN, sbioPN** : Cinétique des systèmes chimique et biologiques
- **CAMAN, sybilccFBA, HMR** : Modèles de mélange, modèle de flux
- **ADDT** : Modèle de vieillissement accéléré
- **HomoPolymer** : Modèle de polymérisation

## Contrôle de procédés, Cartes de contrôle :

- **vMAsk** : C. Contrôle par somme cumulée
- **spc4sts** : C. Contrôle pour surface texturée
- **Copula.Markov** : C. Contrôle sur copules et chaînes de Markov
- **EnergyOnlineCPM** : C. Contrôle et fonction caractéristique de vecteur
- **PPQplan** : Un package fourni par [www.Merck.com](http://www.Merck.com)

## Packages disponibles en grand nombre :

- Spectrométrie
- Chromatographie
- Modèles PK/PD
- Équations différentielles
- Plans d'expériences (le point de départ de R)

**Conclusion** : Un inventaire assez hétérogène

## Rappel : les 3 V

- **Volume** : Volume des données > Capacité du disque dur ou de la mémoire vive
- **Vitesse** : Durée du calcul  $\Rightarrow$  Parallélisation
- **Variété** : Diversité des données  $\Rightarrow$  Réduction de dimension

## La task view *HighPerf Computing* : 99 packages recensés



### CRAN Task View: High-Performance and Parallel Computing with R

**Maintainer:** Dirk Eddelbuettel

**Contact:** Dirk.Eddelbuettel at R-project.org

**Version:** 2018-08-15

**URL:** <https://CRAN.R-project.org/view=HighPerformanceComputing>

This CRAN task view contains a list of packages, grouped by topic, that are useful for high-performance computing (HPC) with R. In this context, we are defining 'high-performance computing' rather loosely as just about anything related to pushing R a little further: using compiled code, parallel computing (in both explicit and implicit modes), working with large objects as well as profiling.

Unless otherwise mentioned, all packages presented with hyperlinks are available from CRAN, the Comprehensive R Archive Network.

## Variété

- R gère tous les types de données qui peuvent être regroupés naturellement dans un objet de type *liste*.
- R est excellent pour les données manquantes !!
- R propose de nombreux packages pour la réduction de dimension :  
⇒ données univariées, temporelles, fonctionnelles, spatiales, corrélées, etc...

## Très grands nombres

- Package **bit64** pour les entiers  $\pm 2^{63}$
- Voir les packages **gmp** (Gnu Multiple Precision), **bigIntegerAlgos**
- En double précision, R manipule des nombres jusqu'à  $1.79 \times 10^{308}$
- Le package **Brodingnag**<sup>2</sup> manipule des nombres jusqu'à  $10^{7.8 \times 10^{307}}$

---

2. Désigne l'île des géants dans *Les voyages de Gulliver* (Swift, 1721). Le CRAN accepte les auteurs facétieux, ici le package d'un professeur de l'Université de Cambridge.

## Lecture de gros fichiers .txt ou .csv

- Lecture simple : Package **data.table** (> 100 GB),
- Lecture fractionnée : Packages **ff**, **ffbase**, **LaF**

## Calcul fractionné

- Packages **bigmemory**, **biglm**, **speedglm**, **biglars**

## Lecture et calcul via les bases de données

- Les packages **MonetDB.R** et **sqldf** fournissent des outils de type BDD
- Le package **DBI** donne accès aux bases de données et au calcul via les BDD
- Publicité : Regardez la BDD de la start-up [www.quasardb.net](http://www.quasardb.net) (100 Gbit/s)

## Hadoop, Spark, TensorFlow

- Les packages **HadoopStreaming**, **hive**, **hdf5** donnent accès à Hadoop
- Les packages **SparkR**, **sparklyr**, **spark rsparkling** donnent accès à Spark
- Les packages **tensorflow**, **tfruns**, **tfestimators** donnent accès à TensorFlow

## R intègre aisément des codes C, C++ et Fortran (et vice-versa)

- Soit de manière directe pour le C et le Fortran
- Soit via les packages **Rcpp** et tous ses dérivés dont **RcppParallel**. Plus de 1300 packages utilisent **Rcpp** (première version en 2008) !!
- Le package **RInside** permet d'intégrer du code R dans du code C.

## Calcul parallélisé sur plusieurs cœurs d'une même machine

- Package **parallel** pour Unix (Linux, MacOSX, BSD, Solaris) et Windows qui réalise une synthèse de plusieurs packages plus anciens OS dépendants. Le même code pour tout le monde.

## Calcul parallélisé (implicite, explicite) sur plusieurs machines

- Solution OpenMP : packages **romp**, **Rborist**
- Solutions MPI : packages **snow** + **snowfall**, **Rhpc**, **Rmpi**, **pbdMPI** et ses dérivés
- Avec utilisation de BLAS : package **pbdSLAP**, **RhpcBLASctl**
- Une autre suite de packages **drake**, **future**, **batchtools**, **future.batchtools**

## Calcul sur GPU

- Le package **gcbd** pour comparer BLAS et les résultats obtenus par GPU.
- Le package **OpenCL** pour la programmation GPU OpenCL (cartes AMD).
- Le package **gpuR** qui encapsule les fonctions OpenCL dans la syntaxe traditionnelle de R.
- Des packages applicatifs pour CUDA **cudaBayesreg**, **permGPU** (micro-array).
- Les packages **tensorflow** et **tfEstimators** utilisent TensorFlow.

## Les ordonnanceurs de calcul

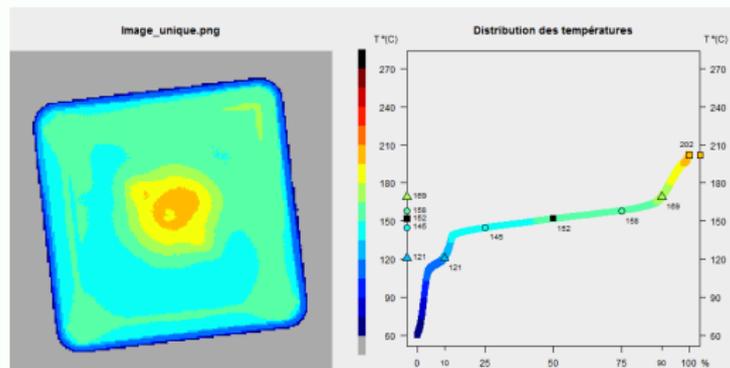
- Packages **rslurm**, **batch**, **BatchJobs**, **BatchEperiments**, **flowr**, **clustermq**

## R propose des solutions élégantes pour la génération de documents

- R ⇒ Markdown ⇒ Word
- R ⇒ Markdown ⇒ (Xe,Lua)Latex ⇒ pdf
- R ⇒ Shiny ⇒ Service web  
en moins de 20' chez EDF Commerce qui a tout automatisé. Voir :  
<https://www.inodelia.com/download/2018-0528-Template-R-shiny-edf.pdf>

## R permet de générer des vidéos

- Voir <http://www.inodelia.com/download/Temperature.mp4>



## Conférences nationales. En France :

---

Les Rencontres R (Bordeaux, Lyon, Montpellier, Grenoble et Toulouse)  
La journée annuelle du MNHN. Le Meetup R-addicts  
Des écoles de printemps

---

## Conférence internationale **useR !** depuis 2004

<https://www.r-project.org/conferences.html>

---

|      |                    |                                                                                       |
|------|--------------------|---------------------------------------------------------------------------------------|
| 2004 | Vienne             |                                                                                       |
| ...  |                    |                                                                                       |
| 2013 | Castilla La Mancha | <a href="http://www.edii.uclm.es/~useR-2013/">http://www.edii.uclm.es/~useR-2013/</a> |
| 2014 | Los Angeles        | <a href="http://user2014.stat.ucla.edu/">http://user2014.stat.ucla.edu/</a>           |
| 2015 | Aalborg            | <a href="http://user2015.math.aau.dk/">http://user2015.math.aau.dk/</a>               |
| 2016 | Stanford           | <a href="http://user2016.org/">http://user2016.org/</a>                               |
| 2017 | Bruxelles          | <a href="http://user2017.brussels/">http://user2017.brussels/</a>                     |
| 2018 | Brisbane           | <a href="https://user2018.r-project.org">https://user2018.r-project.org</a>           |
| 2019 | <b>Toulouse</b>    | <a href="http://www.user2019.fr">http://www.user2019.fr</a>                           |
| 2020 | Boston             |                                                                                       |

---



useR! 2019 @UseR2019\_Conf · 7 juil. 2017



After #useR2017 in Brussel & Brisbane in 2018, we will be hosting #useR2019 in Toulouse. Happy and thrilled by this great honor! #rstats #TY



**Comité d'organisation** : Nathalie Vialaneix (présidente), INRA

- U. Warwick : Heather Turner (vice-présidente du C.O. et présidente du C.S.)
- INRA : Robert Faivre, Élise Maigné, Rémi Servien, Matthias Zytnicki
- IMT : Sébastien Déjean (vp), Xavier Gendre, Pierre Neuvial
- TSE : Anne Ruiz-Gazen (vp), Aurore Archimbaud, Christophe Bontemps, Thibault Laurent



**Merci pour votre attention**

**[patrice.kiener@inmodelia.com](mailto:patrice.kiener@inmodelia.com)**

Tél. : +33.9.53.45.07.38