

2. Introduction to R for Digital Soil Mapping

Canadian Digital Soil Mapping Workshop, 2022

About Us

Brandon Heung

Dalhousie University

Brandon.Heung@dal.ca



Daniel Saurette

Ontario Ministry of Agri-Food and Rural Affairs

University of Guelph

Daniel.Saurette@ontario.ca

What is R?



R is a free software environment for statistical computing and graphics.

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

- **Free to use**
- Used by majority of statisticians
- R community provides technical support
- Automation of DSM processes
- Sharing of R scripts
- Open source

Installing R

R-3.6.0 for Windows (32/64 bit)

[Download R 3.6.0 for Windows](#) (80 megabytes, 32/64 bit)

[Installation and other instructions](#)

[New features in this version](#)

If you want to double-check that the package you have downloaded matches the package distributed by CRAN, you can compare the [md5sum](#) of the .exe to the [fingerprint](#) on the master server. You will need a version of md5sum for windows: both [graphical](#) and [command line versions](#) are available.

Frequently asked questions

- [Does R run under my version of Windows?](#)
- [How do I update packages in my previous version of R?](#)
- [Should I run 32-bit or 64-bit R?](#)

Please see the [R FAQ](#) for general information about R and the [R Windows FAQ](#) for Windows-specific information.

Other builds

- Patches to this release are incorporated in the [r-patched snapshot build](#).
- A build of the development version (which will eventually become the next major release of R) is available in the [r-devel snapshot build](#).
- [Previous releases](#)

Note to webmasters: A stable link which will redirect to the current Windows binary release is <http://CRAN.MIRROR>bin/windows/base/release.htm>.

Last change: 2019-04-26

Compatible with OS X, Windows, Linux
32-bit & 64-bit options

<http://mirror.its.dal.ca/cran/>

R Packages

Contributed Packages

Available Packages

Currently, the CRAN package repository features 14435 available packages.

[Table of available packages, sorted by date of publication](#)

[Table of available packages, sorted by name](#)

Installation of Packages

Please type `help("INSTALL")` or `help("install.packages")` in R for information on how to install packages from this repository. The manual [R Installation and Administration](#) (also contained in the R base sources) explains the process in detail.

[CRAN Task Views](#) allow you to browse packages by topic and provide tools to automatically install all packages for special areas of interest. Currently, 40 views are available.

Package Check Results

All packages are tested regularly on machines running [Debian GNU/Linux](#), [Fedora](#), OS X, Solaris and Windows.

The results are summarized in the [check summary](#) (some [timings](#) are also available). Additional details for Windows checking and building can be found in the [Windows check summary](#).

Writing Your Own Packages

The manual [Writing R Extensions](#) (also contained in the R base sources) explains how to write new packages and how to contribute them to CRAN.

Repository Policies

The manual [CRAN Repository Policy](#) [\[PDF\]](#) describes the policies in place for the CRAN package repository.

Related Directories

[Archive](#)

Previous versions of the packages listed above, and other packages formerly available.

[Orphaned](#)

Packages with no active maintainer, see the corresponding [README](#).

[bin/windows/contrib](#)

Windows binaries of contributed packages

[bin/macosx/el-capitan/contrib](#)

OS X El Capitan binaries of contributed packages

Each package consists of a suite of **tools** that are used for day-to-day operations in R.

Example:

caret: Machine-learning algorithms

sp: Spatial data analysis

RSAGA: Integration with SAGA-GIS

Each **package** includes a **reference manual** that describes the use of each **tool**.

<https://cran.r-project.org/web/packages/>

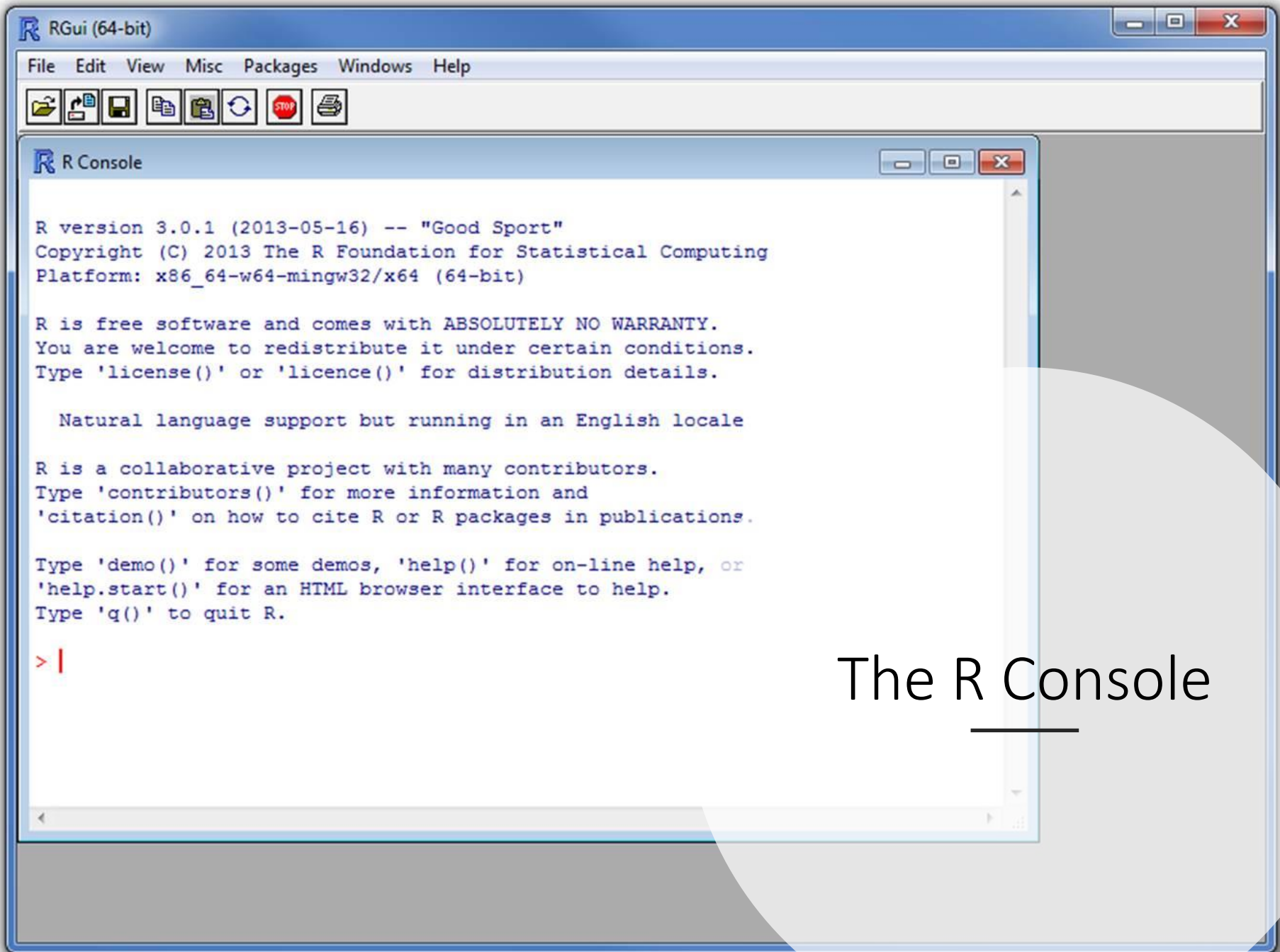
R Scripts

```
1 #####
2 ### Conditioned Latin Hypercube Sampling of Principal Components ###
3 #####
4
5 # STEP 1: Input working directory using setwd("x")
6
7 options( java.parameters = "-Xmx24g" )
8 setwd("E:/2018_Kamloops CLHS/Covariates")
9 setwd("E:/2018_Kamloops CLHS/Covariates/BEC Covariates/IDFDK1_IDFDK1")
10 setwd("E:/2018_Kamloops CLHS/Covariates/BEC Covariates/Reduced Variables")
11
12
13 dir.create("PCA")
14 dir.create("Sample Buffers")
15 dir.create("CLHS")
16
17
18 # STEP 2: Load necessary packages
19
20 library(clhs)
21 library(foreign)
22 library(ggplot2)
23 library(raster)
24 library(rgdal)
25 library(RStoolbox)
26 library(maptools)
27 library(sp)
28 library(spatstat)
29
30 rasterOptions(tmpdir="E:/2018_Kamloops CLHS/tmpdir")
```

Contains a series of command lines that are used to make R do things

Useful for automation and reproducing your workflow

Tons of resources available for you to copy & paste R scripts and modify for your own purposes.



The R Console

R Studio

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Run Source

Environment History Connections

Global Environment

Data

Soil_class	14316 obs. of 11 variables
Soil_PM	14570 obs. of 8 variables
Soil_PM_summary	19 obs. of 7 variables

Files Plots Packages Help Viewer

Zoom Export Publish

Uncertainty Assessment

Uncertainty

Parent Material

C E F FG L LG M O W WG

Script Editor

```
1 library(ggplot2)
2 library(ggpubr)
3 library(Rmisc)
4
5 setwd("E:/Projects/2018_BC Pure Polygon Mapping_Revisions/Entropy Analysis/")
6
7 soil_PM <- read.csv("Validation_PM.csv")
8 soil_class <- read.csv("Validation_Great_Group.csv")
9
10 soil_PM_summary <- summarySE(soil_PM, measurevar="Ent_PM", groupvars=c("obs_PM", "Acc_PM"))
11
12 ggplot(soil_PM_summary, aes(x=obs_PM, y=Ent_PM, fill=Acc_PM)) +
13   geom_bar(position=position_dodge(), stat="identity",
14     colour="black", # use black outlines,
15     size=.3) + # Thinner lines
16   geom_errorbar(aes(ymin=Ent_PM-se, ymax=Ent_PM+se),
17     size=.3, # Thinner lines
18     width=.2,
19     position=position_dodge(.9)) +
20   xlab("Parent Material") +
21   ylab("Uncertainty") +
22   scale_fill_hue(name="Supplement type", # Legend label, use darker colors
23     breaks=c("OJ", "VC"),
24     labels=c("Orange juice", "Ascorbic acid")) +
25   ggtitle("uncertainty Assessment") +
26   scale_y_continuous(breaks=0:20*4) +
27   theme_bw()
28
29 ggplot(soil_PM, aes(Acc_PM, Ent_PM)) +
30   geom_boxplot(aes(color = Acc_PM)) +
31   facet_wrap(~obs_PM) +
32   theme_minimal() +
33   ggtitle("The Effect of Vitamin C on Tooth Growth in Guinea Pigs") +
34   scale_y_continuous(breaks=0:20*4) +
35   theme_bw()
36
37 ylab("Tooth length") +
38 scale_fill_hue(name="Supplement type", # Legend label, use darker colors
39   breaks=c("OJ", "VC"),
40   labels=c("Orange juice", "Ascorbic acid")) +
41 ggtitle("The Effect of Vitamin C on Tooth Growth in Guinea Pigs") +
42 scale_y_continuous(breaks=0:20*4) +
43 theme_bw()
44
45 ggplot(soil_PM_summary, aes(x=obs_PM, y=Ent_PM, fill=Acc_PM)) +
46   geom_bar(position=position_dodge(), stat="identity",
47     colour="black", # use black outlines,
48     size=.3) + # Thinner lines
49   geom_errorbar(aes(ymin=Ent_PM-se, ymax=Ent_PM+se),
50     size=.3, # Thinner lines
51     width=.2,
52     position=position_dodge(.9)) +
53   xlab("Parent Material") +
54   ylab("Uncertainty") +
55   scale_fill_hue(name="Supplement type", # Legend label, use darker colors
56     breaks=c("OJ", "VC"),
57     labels=c("Orange juice", "Ascorbic acid")) +
58   ggtitle("uncertainty Assessment") +
59   scale_y_continuous(breaks=0:20*4) +
60   theme_bw()
61
62 >
```

R Console

Workspace & History

Plot Viewer & File Manager

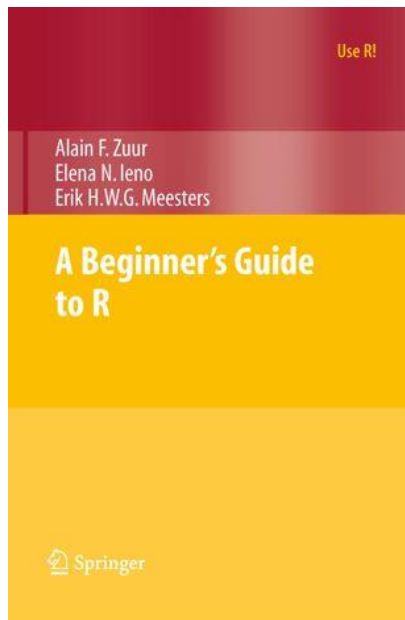
Scripting software that is specifically designed for R

Using R in Digital Soil Mapping

- Management of digital soil information and geospatial data
- Performs (geo-)statistical analyses
 - Conventional statistics
 - Spatial statistics
 - Machine-learning and artificial intelligence
- Integrates with GIS software (SAGA, GRASS GIS, QGIS)
- “Big Data” analytics

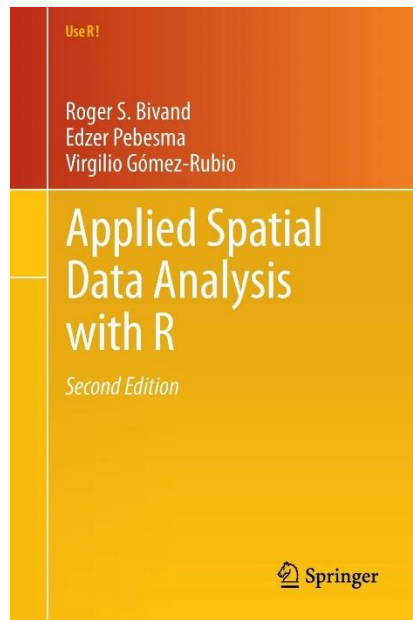


General Resources for R



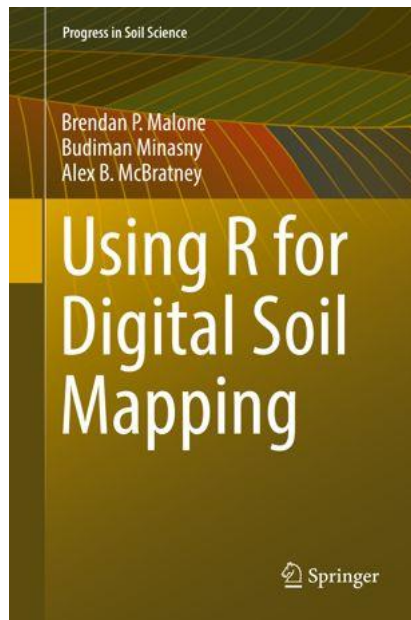
- Many resources for using R at the beginner's level.
- Includes R scripts and detailed walkthroughs

Spatial Data Analysis in R



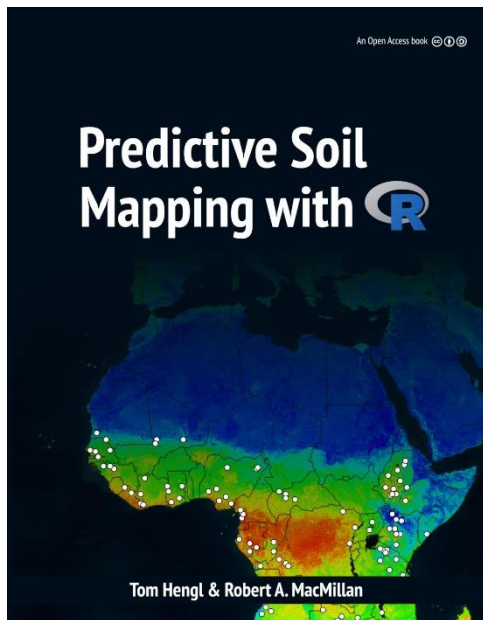
- Handling spatial data in R
- Visualizing spatial data
- Spatial data import and export
- Spatio-temporal data
- Spatial point pattern analysis
- Interpolation & geostatistics
- Modelling areal data

Digital Soil Mapping in R



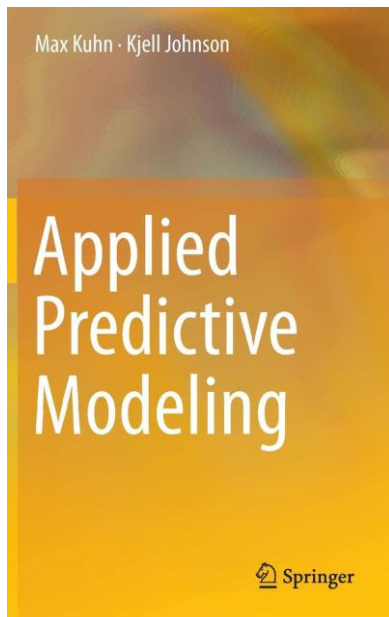
- R literacy for digital soil mapping
- Exploratory data analysis
- Modelling continuous and categorical soil data
- Quantification of prediction uncertainty
- Updating, harmonizing, and disaggregating legacy soil maps
- Digital soil assessments

Digital Soil Mapping in R



- Soil resource inventories
- Software installation
- Soil observations & variables
- Preparation of geospatial layers
- Statistical & machine-learning theory

Predictive Modelling in R

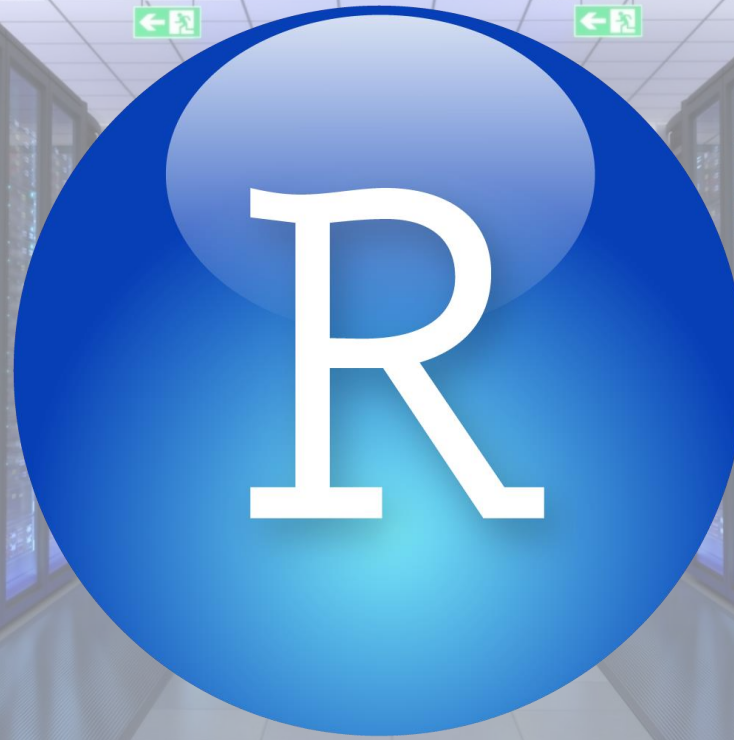


- Data pre-processing
- Model fitting and overfitting
- Regression modelling
- Classification modelling

Module 2: Starting R

- Installation & loading libraries in R
- Setting up your working drive & accessing data
- Basic data exploration





2. Introduction to R for Digital Soil Mapping

Canadian Digital Soil Mapping Workshop, 2022