Frédéric Santos

CONFIGURATION

Along with R 4.0.3 itself [rcoreteam2020\_LanguageEnvironmentStatistical], the R package corrplot [wei2017\_PackageCorrplotVisualization] was used.

library(corrplot)

print(sessionInfo(), locale = FALSE)

R version 4.0.3 (2020-10-10)

Platform: x86\_64-pc-linux-gnu (64-bit)

Running under: Manjaro Linux

Matrix products: default

BLAS: /usr/lib/libopenblasp-r0.3.10.so

LAPACK: /usr/lib/liblapack.so.3.9.0

attached base packages:

[1] stats graphics grDevices utils datasets

[6] methods base

other attached packages:

[1] corrplot\_0.84

loaded via a namespace (and not attached):

[1] compiler\_4.0.3

This document has been written using Org mode 9.4 for Emacs 27.1 [schulte2012\_MultiLanguageComputingEnvironment].

# LOAD DATA FILE

The data file corresponding to the experimental data is imported in R. Here is a summary of this dataframe:

dat <- read.csv("./MeasurementExperiment\_Datas.csv",

header = TRUE,

sep = ";", dec = ",",

stringsAsFactors = TRUE,

na.strings = "", fileEncoding = "utf-8")

summary(dat)

Experiment\_ID Location Material Zone.nbr

XPme1:45 Approximate:90 Glass:90 Min. :1

XPme2:45 Precise :90 Resin:90 1st Qu.:3

XPme3:45 Median :5

XPme4:45 Mean :5

3rd Qu.:7

Max. :9

Measure.Nbr Thickness

Min. :1 Min. :1091

1st Qu.:2 1st Qu.:1093

Median :3 Median :1144

Mean :3 Mean :1148

3rd Qu.:4 3rd Qu.:1203

Max. :5 Max. :1220

Since there are four experimental settings, the dataframe is split into four corresponding subsets:

sp <- split(dat, dat$Experiment\_ID)

# HELPERS

Two R functions will serve as helpers in this study. The first one, based on the R package corrplot (and which is basically a wrapper for the R function corrplot()), will allow for a simple visualization of means or standard deviations of all measurements made on each thin section.

plot\_section <- function(values, which = c("sd", "cv", "mean"),

cl.lim = NULL) {

### values: dataframe of experimental measurements

### which: statistic to be computed (either "sd", "cv" or "mean")

### cl.lim: argument passed to corrplot()

if (which == "sd") {

res <- tapply(values$Thickness,

INDEX = values$Zone.nbr,

FUN = sd)

} else if (which == "cv") {

res <- tapply(values$Thickness,

INDEX = values$Zone.nbr,

FUN = function(x) return(sd(x) / mean(x)))

} else {

res <- tapply(values$Thickness,

INDEX = values$Zone.nbr,

FUN = mean)

}

corrplot(matrix(res, ncol = 3, byrow = TRUE),

is.corr = FALSE, method = "number",

number.digits = 6,

cl.pos = "r",

mar = c(0, 0, 1, 0),

title = paste(values$Material[1],

values$Location[1],

sep = " / "),

cl.lim = cl.lim,

bg = "lightsalmon")

return(res)

}

The second one aims to centre the values of each thin section. Glass and resin thin sections have slightly different thicknesses (around 1090 for glass sections, around 1200 for resin sections); and even within one given thin section, a (moderate) heterogeneity in thickness does exist depending on the location on which the measurement is done. The following R function thus centres thickness values, for each (square) location in each given thin section. I.e., for each given section, the mean thickness is computed among the 5 replicates made on each square location, and this mean value is subtracted for all individual values—absolute values thus become relative deviations to the mean.

center\_section <- function(values) {

moy <- tapply(values$Thickness,

INDEX = values$Zone.nbr,

FUN = mean)

return(values$Thickness - rep(moy, each = 5))

}

For instance, the first 10 values of the section XPme1 are the following:

head(sp$XPme1[, c(1, 4, 6)], 10)

Experiment\_ID Zone.nbr Thickness

1 XPme1 1 1092

2 XPme1 1 1092

3 XPme1 1 1092

4 XPme1 1 1092

5 XPme1 1 1091

6 XPme1 2 1091

7 XPme1 2 1091

8 XPme1 2 1092

9 XPme1 2 1091

10 XPme1 2 1091

and the corresponding “centred” values are as follows:

head(center\_section(sp$XPme1), 10)

1 1 1 1 1 2 2 2 2 2

0.2 0.2 0.2 0.2 -0.8 -0.2 -0.2 0.8 -0.2 -0.2

\pagebreak

# DATA VISUALIZATION

The average thicknesses observed on each square zone of each section are represented in Figure 1

par(mfrow = c(2, 2))

lapply(sp, plot\_section, which = "mean", cl.lim = c(1090, 1220))

Resin sections appear to be slightly thicker. Similarly, the standard deviations of thickness values are represented in Figure 2.

par(mfrow = c(2, 2))

lapply(sp, plot\_section, which = "sd", cl.lim = c(0, 6))

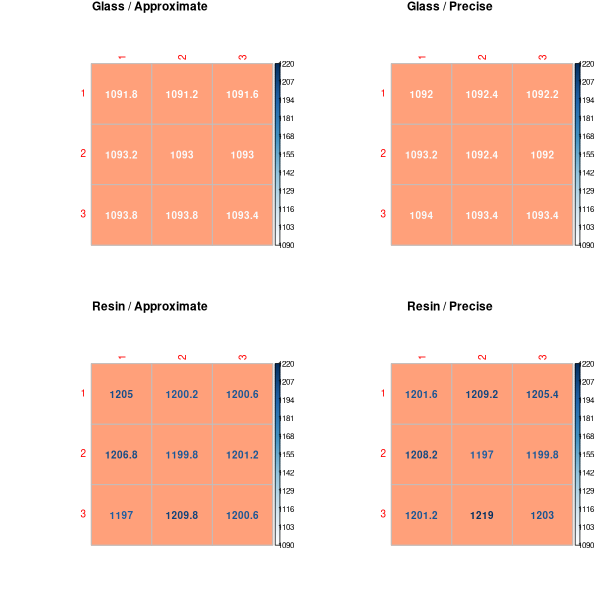


Figure 1: Average thicknesses for each section.

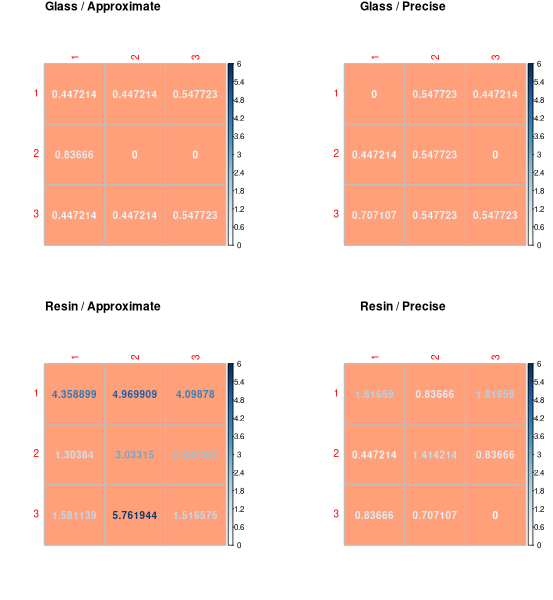


Figure 2: Standard deviations of thickness values.

The observed variability is much higher for the resin section with an approximate measurement protocol.

# COMPARISON OF THICKNESS VARIABILITY

### *Hypotheses*

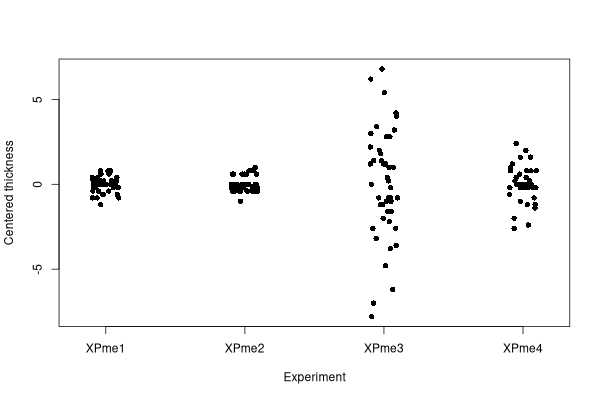
The variability of the $9 \times 5 = 45$ “centred” thickness values[[1]](#footnote-1) can be compared among all four experimental settings. Unlike Figure 2, Figure 3 allows for a global comparison of the thickness variability among thin sections.

values\_per\_slide <- lapply(sp, FUN = center\_section)

stripchart(values\_per\_slide, vertical = TRUE, pch = 16,

method = "jitter", ylab = "Centered thickness",

xlab = "Experiment")



**Figure 3: Stripcharts of the 45 centred thickness values for each thin section. fig-strip**

Once again, the experimental setting XPme3 (“resin / approximate”) shows a much higher variability than all other settings. On the other hand, no clear difference can be observed between XPme1 and XPme2 (“glass / approximate” and “glass / precise” respectively).

## Comparison depending on the measurement protocol.

1. First, the variance of (centred) thickness values between XPme1 and XPme2 are compared, to know whether a “precise” protocol allows for a lower variability than an “approximate” protocol, when performed on glass thin sections.

var.test(center\_section(sp$XPme1), center\_section(sp$XPme2))

F test to compare two variances.

data: center\_section(sp$XPme1) and center\_section(sp$XPme2)

F = 1, num df = 44, denom df = 44, p-value = 1

alternative hypothesis: true ratio of variances is not equal to 1

95 percent confidence interval:

0.549539 1.819707

sample estimates:

ratio of variances

1

The observed variance is identical: a “precise” protocol does not bring any improvement for glass sections.

1. Then, the variance of (centred) thickness values between XPme3 and XPme4 are compared, so as to know whether a “precise” protocol allows for a lower variability than an “approximate” protocol, when performed on resin thin sections.

var.test(center\_section(sp$XPme3), center\_section(sp$XPme4))

F test to compare two variances

data: center\_section(sp$XPme3) and center\_section(sp$XPme4)

F = 10.096, num df = 44, denom df = 44, p-value =

3.225e-12

alternative hypothesis: true ratio of variances is not equal to 1

95 percent confidence interval:

5.548416 18.372656

sample estimates:

ratio of variances

10.09649

Switching from an “approximate” to a “precise” protocol allowed for lowering the variance of the measurements by a factor of 10 (with an associated 95% confidence interval equal to [5.54, 18.37]).

## Comparison between glass and resin

Similar comparisons are performed between glass and resin, while keeping fixed the protocol (“approximate” or “precise”) this time.

1. Comparing XPme4 and XPme2 allows for the comparison of glass and resin for a “precise” protocol:

var.test(center\_section(sp$XPme4), center\_section(sp$XPme2))

F test to compare two variances

data: center\_section(sp$XPme4) and center\_section(sp$XPme2)

F = 5.4286, num df = 44, denom df = 44, p-value =

1.205e-07

alternative hypothesis: true ratio of variances is not equal to 1

95 percent confidence interval:

2.983212 9.878410

sample estimates:

ratio of variances

5.428571

The variance of the measurements are 5 times higher for resin than for glass in “precise” conditions (with an associated 95% confidence interval equal to [3, 9.9]).

1. Comparing XPme3 and XPme1 allows for the comparison of glass and resin for an “approximate” protocol:

var.test(center\_section(sp$XPme3), center\_section(sp$XPme1))

F test to compare two variances

data: center\_section(sp$XPme3) and center\_section(sp$XPme1)

F = 54.81, num df = 44, denom df = 44, p-value <

2.2e-16

alternative hypothesis: true ratio of variances is not equal to 1

95 percent confidence interval:

30.11997 99.73728

sample estimates:

ratio of variances

54.80952

The variance of the measurements are 55 times higher for resin than for glass in “approximate” conditions (with an associated 95% confidence interval equal to [30, 100]).

1. For each thin section, there are 9 square locations and 5 replicates per square. [↑](#footnote-ref-1)