

# How to use the R code that accompanies “Improved threshold diagnostic plots for extreme value analyses”

Paul Northrop and Coleman

January 23, 2014

This note explains how to use the R in the file `NorthropColeman2014.R` to reproduce some of the plots in Northrop and Coleman (2014).

## 1 Required software

R (R Development Core Team, 2010) is a free software environment for statistical computing and graphics. It can be installed from [www.http://cran.r-project.org/](http://cran.r-project.org/).

Our code is largely stand-alone but the package `lattice` (Sarkar, 2010) is required to produce figure 7. Our main function `score.fitrange` uses code that implements Grimshaw (1993) and has an option to use a (modified version of) the function `gpd.fit` from Coles and Stephenson (2009).

The R commands used below are given in the file `NorthropColeman2014.R`.

## 2 Getting started

Install the R package and (if figure 7 is required) install `lattice`. Then we

- set the working directory to one containing the file `NorthropColeman2014.fns`
- read in the functions contained in `NorthropColeman2014.fns` using the `source` function.

```
> setwd("C:/Users/paulnorthrop/Documents/UGPROJ/CLAIRE/PAPER R/SWEAVE")
> source("NorthropColeman2014.fns")
```

A brief description of the arguments to `score.fitrange`. Only the first two arguments (`raw.data` and `u`) are required.

```
score.fitrange <- function(raw.data,u,GP.fit="Grimshaw",do.LRT=FALSE,size=NULL,
                           my.xlab=NULL,xi.tol=1e-3){

# raw.data      : raw data
# u             : m-vector of thresholds (smallest to largest)
# GP.fit        : function used to fit GP model:
# Using 1D optimisation ...
#               "1D.nlm" : using nlm()
#               "1D.optim" : using optim()
#               "Grimshaw" : code from Grimshaw (1993)
# [Grimshaw, S. D. (1993) Computing Maximum Likelihood Estimates for the Generalized
# Pareto Distribution, Technometrics, Vol. 35, No. 2, 185-191.]
# Using 2D optimisation ...
#               "ismev" : (algebraic gradient added) version of gpd.fit() from ismev
```

```
#
# The choice of GP.fit should make no difference but we have kept these options
# just in case. The results in Northrop and Coleman (2014) used "1D.nlm".
# Subsequently we discovered the Grimshaw code, which is probably superior.
# Therefore, we have made "Grimshaw" the default.
#
# do.LRT      : if TRUE do LR test (in addition to the score test)
# size        : level at which a horizontal line is drawn on multiple threshold plot
# my.xlab     : (optional) x-axis label
# xi.tol      : if the absolute value of xi1.hat is less than xi.tol use linear interpolation
#              to evaluate score vectors, expected Fisher information matrices, Hessians
```

### 3 Analyses of Nidd data (section 4.1)

We set a vector of thresholds from  $65 \text{ m}^3 \text{ s}^{-1}$  to  $120 \text{ m}^3 \text{ s}^{-1}$  at intervals of  $5 \text{ m}^3 \text{ s}^{-1}$ . First we perform score test only, then both the score test and likelihood ratio tests.

```
> u <- c(seq(from = 65, to = 115, by = 5), 120)
> res.nidd.5 <- score.fitrange(nidd.thresh, u)
> res.nidd.5.LRT <- score.fitrange(nidd.thresh, u, do.LRT = T)
```

Look at a summary of the output list object. The variable names should be self-explanatory, apart from \$nllh, which is the negated log-likelihood at the (restricted) MLE.

```
> summary(res.nidd.5.LRT)
```

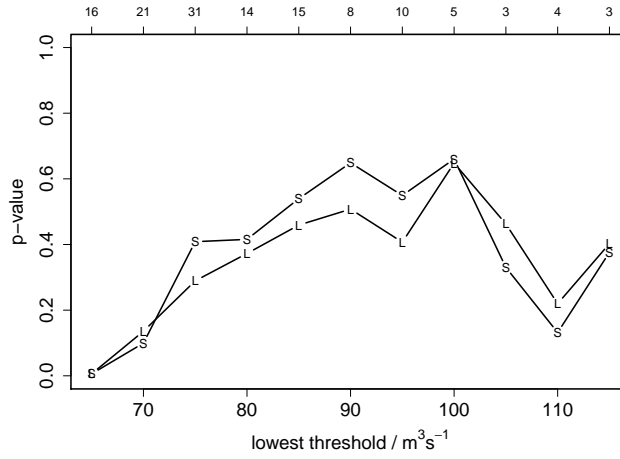
	Length	Class	Mode
thresh	12	-none-	numeric
nexc	12	-none-	numeric
n.between	12	-none-	numeric
xi.mle	12	-none-	numeric
sigma.mle	12	-none-	numeric
nllh	12	-none-	numeric
df	11	-none-	numeric
e.test.stats	11	-none-	numeric
e.p.values	11	-none-	numeric
LRT.p.values	11	-none-	numeric
LRT.test.stats	11	-none-	numeric
u	11	-none-	numeric

Produce figure 4a (excluding the lines based on simulation).

```
> par(mar = c(4.2, 4.2, 2.2, 1), lwd = 2, cex.lab = 1.5, cex.axis = 1.5)
> my.xlab <- expression(paste("lowest threshold / ", m^3 * s^-1))
> plot(res.nidd.5.LRT$u, res.nidd.5.LRT$e.p.values, type = "b", lty = 1, pch = "S",
+       ylim = c(0, 1), xlab = my.xlab, ylab = "p-value")
> lines(res.nidd.5.LRT$u, res.nidd.5.LRT$LRT.p.values, type = "b", lty = 1, pch = "L")
> lines(res.nidd.5.LRT$u, res.nidd.5.LRT$e.mult.mult, type = "l", lty = 2, pch = "")
> lines(res.nidd.5.LRT$u, res.nidd.5.LRT$l.mult.mult, type = "l", lty = 4, pch = "")
> axis(3, at = res.nidd.5.LRT$u, labels = res.nidd.5.LRT$n.between[1:length(res.nidd.5.LRT$u)],
+       cex.axis = 1)
```

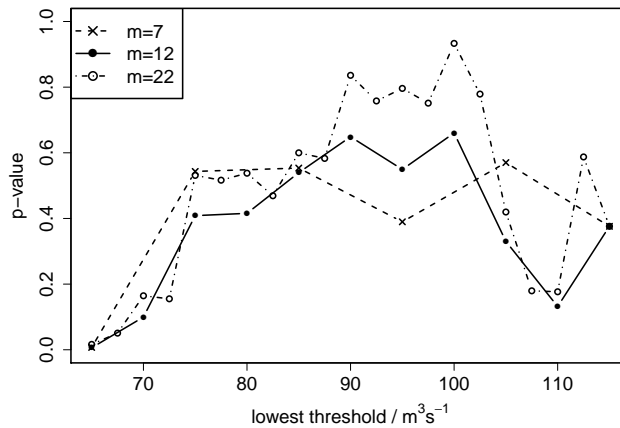
Perform the score test only, with fewer thresholds than more thresholds.

```
> u <- c(seq(from = 65, to = 115, by = 10), 120)
> res.nidd.10 <- score.fitrange(nidd.thresh, u)
> u <- c(seq(from = 65, to = 115, by = 2.5), 120)
> res.nidd.2.5 <- score.fitrange(nidd.thresh, u)
```



Produce figure 4b (excluding the lines based on simulation).

```
> par(mar = c(4.2, 4.2, 2.2, 1), lwd = 2, cex.lab = 1.5, cex.axis = 1.5)
> my.xlab <- expression(paste("lowest threshold / ", m^3 * s^-1))
> plot(res.nidd.5$u, res.nidd.5$e.p.values, type = "b", ylim = c(0, 1), xlab = my.xlab,
+      ylab = "p-value", pch = 16, lty = 1)
> lines(res.nidd.10$u, res.nidd.10$e.p.values, type = "b", pch = 4, lty = 2)
> lines(res.nidd.2.5$u, res.nidd.2.5$e.p.values, type = "b", pch = 1, lty = 4)
> legend("topleft", legend = c("m=7", "m=12", "m=22"), lty = c(2, 1, 4), pch = c(4,
+      16, 1), cex = 1.5)
```



## References

- Coles, S. and A. Stephenson (2009). *ismev: An Introduction to Statistical Modeling of Extreme Values*. R package version 1.34. Original S functions by Stuart Coles and R port and R documentation files by Alec Stephenson.
- Grimshaw, S. D. (1993). Computing maximum likelihood estimates for the generalized Pareto distribution. *Technometrics* 35(2), 185–191.
- Northrop, P. J. and C. L. Coleman (2014). Improved threshold diagnostic plots for extreme value analyses. *Extremes*. To appear.
- R Development Core Team (2010). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.
- Sarkar, D. (2010). *lattice: Lattice Graphics*. R package version 0.18-5.