Worked examples

This is an introduction to the models presented in "A null model for quantifying the geometric effect of habitat subdivision on species diversity". The R code used in the paper are in file 'functions.R' and the aim here is to illustrate their use. The R-code provides a brief explanation of the arguments so I don't go into that too much.

Basically there are two parts to the paper, a validation against empirical data and a theoretical exploration of the implications. I'll give an example in the reverse order here.

Estimating diversity

First we need some data - if we assume a value for the community scaling parameter, we only need the species abundance distribution (SAD). Here I've used data from BCI - that is the Barro Colorado Island forest dynamics plot. I use the 2005 census, with 211845 living stems and 301 species.

Of course, the SAD could be simulated assuming any number of different abundance distributions...

```
load('BCI_SAD.RData')
```

The models are all in 'functions.R'.

```
source("functions.R")
args(predSS.NB)

## function (area, sad, cpar, m, tota = 5e+05)
## NULL
```

The arguments are common to the functions and represent:
–area = sampling grain - in the same units as tota.
–sad = species abundance distribution over the study extent
–cpar = community level scaling parameter at sampling grain 'area'
–tota = extent over which the number of individuals in the species abundance distribution were counted (defaults to 50 ha)

To predict the diversity of a set of samples from we first calculate zeta diversity and then plug it into the formula derived in Hui & McGeoch (2014).

```
z20 <- predSS.NB(area=200, sad=bciSAD, cpar = 0.88, m = 20)
round(z20,2)

##  [1] 32.33 14.26  9.41  7.36  6.26  5.57  5.09  4.74  4.46  4.23  4.04
3.88
## [13]  3.75  3.62  3.52  3.42  3.34  3.26  3.19  3.12
```

We can calculate the total number of species (which is kind of like gamma diversity in a geometric subdivision effect comparison)

```
gam.fn(z20)
```

```
## [1] 141.8386
```

Or the number of species in only a single one of those 20 patches

```
spe.fn(z20)
```

```
## [1] 45.9334
```

Exploring the SLOSS question

So, if we are interested in the relative number of species in one large patch for increasing amounts of subdivision (noting all the model caveats about independence and distance decay described in the paper) all we need is a scaling relationship for c. We'll use the empirical estimate from the data in the paper. Where the c parameter for area a (in m^2) was equal to:

$$c_a = c_{400} * \frac{a}{400}^{0.28}$$

So we can compare the species in, say 10 20 x 20 m quadrats with those in 1 x 8000 m^2 (i.e., same total area), plugging in the correct c value as follows:

```
c.sl = 1.06*(8000/400)^0.28   # single large = 1 x 8000 m2
c.ss = 1.06*(800/400)^0.28   # several small = 10 x 800 m2

sr.sl <- predSS.NB(area=8000, sad=bciSAD, cpar = c.sl, m=1)
sr.ss <- gam.fn(predSS.NB(area=800, sad=bciSAD, cpar = c.ss, m=10))
(sr.ss - sr.sl)/sr.ss
```

```
## [1] 0.06156683
```

So ~6% increase in richness is expected if the single patch was subdivided into 10 independent patches.

Comparing observed and predicted

1. Negative binomial

For comparison with empirical data I'll use BCI again. The object 'calstats' contains various diversity stats, where the values were calculated from 100 repeat samples of 20 randomly positioned 20 x 20 m quadrats.

```
load("calstats.RData")
names(calstats)
```

```
## [1] "zeta"    "spe"    "beta"    "gamma"
```

We have four objects, each giving the mean value from the 100 samples from BCI along with 95% sampling intervals:

–'zeta' is zeta diversity (average number of species shared) in 1, 2, … , 20 samples.
–'spe' is the mean number of species found in only 1 quadrat (single patch endemics)
–'beta' is the mean Sorensen dissimilarity
–'gamma' is the total species richness

We can predict zeta using the non-random shared species equation and then use this to predict the diversity patterns from the zeta components (this is all derived in Hui & McGeoch 2014).

First, we need to estimate the community scaling parameter for the 20 x 20 sampling grain. I used mean species richness at the sampling grain of interest for this.

```
alpha.div = calstats$zeta[1,1]
alpha.div # mean number of spp shared in 1 sample = alpha diversity
```

```
## [1] 49.0575
```

Once we have this value, we can use function fitc.NB() to estimate the parameter. (NB: If you want to get a scaling relationship for c, repeat the sampling and fitting steps at a few sampling grains, then fit Eq 8 in the main text - see below).

```
cpar.fit <- fitc.NB(obs = alpha.div, area = 400, sad = bciSAD, tota= 500000,
low=0, upp=100)
cpar.fit$cpar
```

```
## [1] 1.038374
```

This is the value for the community scaling parameter, c, at this sampling grain. Now we just plug it into fitSS.NB() to estimate zeta diversity in the 20 samples…

```
zeta.est <- predSS.NB(area = 400, sad = bciSAD, cpar = cpar.fit$cpar, m = 20,
tota = 500000)
round(zeta.est,2)
```

```
##  [1] 49.06 25.73 18.15 14.48 12.36 10.99 10.04  9.34  8.81  8.39  8.05
7.76
## [13]  7.52  7.31  7.12  6.96  6.81  6.67  6.55  6.44
```

… comparing with the empirical data:

```
plot(1:20, calstats$zeta[,1], type="l",ylim=c(0,60),ylab= 'Zeta diversity',
xlab="Order of zeta")
lines(1:20, calstats$zeta[,2], lty= 2)
lines(1:20, calstats$zeta[,3], lty= 2)
points(1:20, zeta.est)
legend(15,50,legend=c("Predicted","Observed"),pch=c(1,NA),lty=c(NA,1))
```
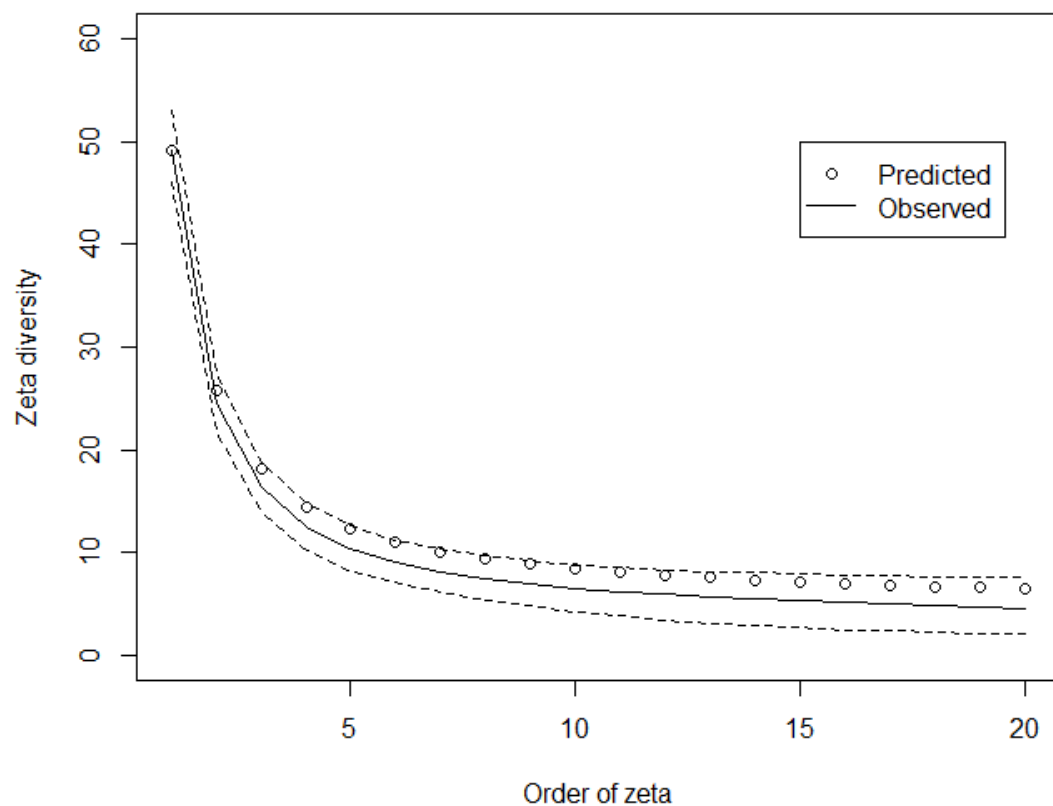
Not perfect, but on the money - it's an approximate model after all. Let's see how it does for the diversity metrics.

First total species richness of the samples:

```
gam.mod <- gam.fn(zeta.est)
gam.obs <- unlist(calstats$gamma)
(gam.mod-gam.obs[1])/gam.obs[1]

##          mean
## -0.02579118

round(calstats$gamma,2)

##        mean hi95 low95
## 20 175.62  188   163
```

Number of species found in a single patch:

```
spe.mod <- spe.fn(zeta.est)
spe.obs <- unlist(calstats$spe)
(spe.mod - spe.obs[1])/spe.obs[1]
```

```
##       mean
## 0.04721336
```

```
round(calstats$spe,2)
```

```
##       mean hi95 low95
## 98% 43.96   56    35
```

And Sorensen dissimilarity:

```
sor.mod <- bd.fn(zeta.est)
sor.obs <- unlist(calstats$beta)
(sor.mod - sor.obs[1])/sor.obs[1]
```

```
##        mean
## -0.03670393
```

```
round(calstats$beta,3)
```

```
##     mean  hi95 low95
## 1 0.494 0.616 0.362
```

One of the things about the null models when compared with real data is that the diversity measures calculated from zeta components seem to be more robust estimates of the empirical values than the zeta values themselves. This seems to remain true, even when the zeta components themselves fell outside the 95% empirical sampling limits (the regularly distributed community in the SI is a good example).

2. Finite negative binomial

For completeness, here's the fitting for the FNB model. It takes a long time and does not seem to provide an improvement that warrants this in my view.

I'm pretty agricultural in my R skills, so I'm sure others will be able to work out a means to speed this up if they so desire... I don't think the improvement in precision is worth it myself.

```
# fit c as with the neg bin
fnb.c <- fitc.FNB(obs = alpha.div, area =400, sad = bciSAD, tota= 500000,
low=0, upp=10)$cpar
round(fnb.c,2)
```

```
## [1] 1.02
```

```
# and predict
zeta.est.fnb <- predSS.FNB(sad = bciSAD, cpar = fnb.c, area = 400, tota =
500000, m = 20)
round(zeta.est.fnb,2)
```

```
##  [1] 49.06 25.81 18.23 14.55 12.42 11.04 10.09  9.39  8.85  8.43  8.09
7.80
## [13]  7.55  7.34  7.16  6.99  6.84  6.71  6.59  6.47
```

How do the FNB and NB shared species estimates compare?

```
plot(1:20, calstats$zeta[,1], type="l",ylim=c(0,60),ylab= 'Zeta diversity',
xlab="Order of zeta")
lines(1:20, calstats$zeta[,2], lty= 2)
lines(1:20, calstats$zeta[,3], lty= 2)
points(1:20, zeta.est)
points(1:20, zeta.est.fnb, col=2, pch=2)
legend(12,50,legend=c("Predicted NB","Predicted FNB", "Observed"),pch=c(1,2,
NA),lty=c(NA,NA, 1), col=c(1,2,1))
```
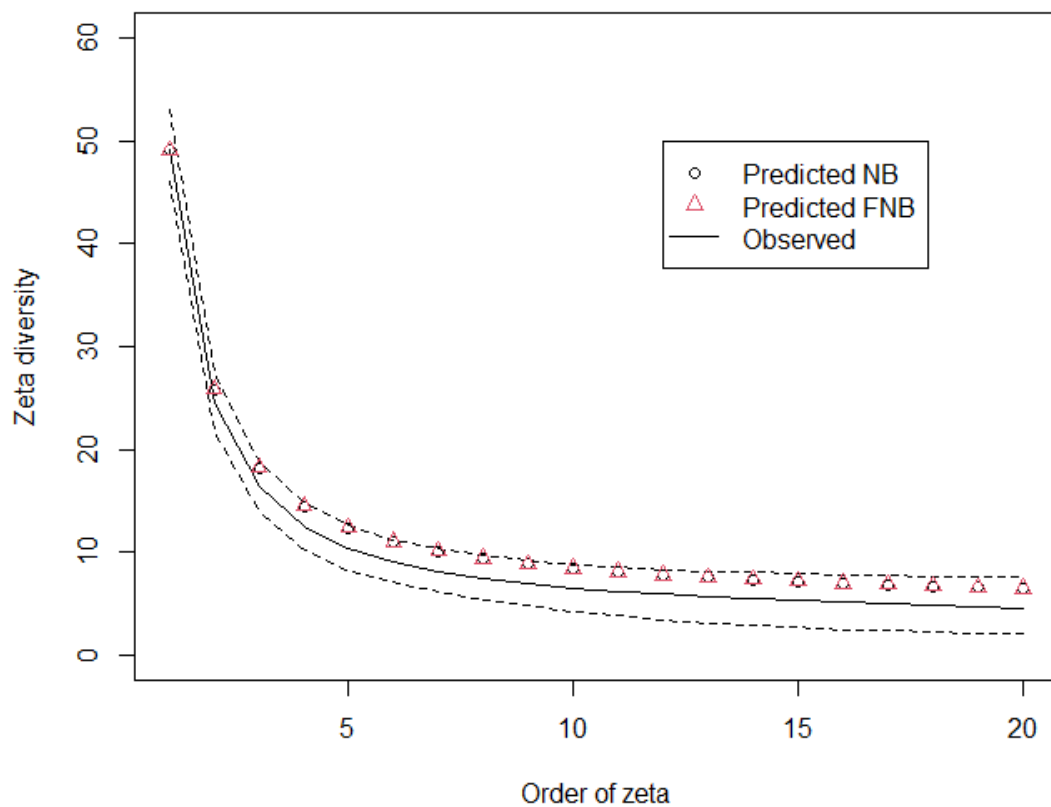


```
plot(zeta.est, zeta.est.fnb);abline(0,1)
```
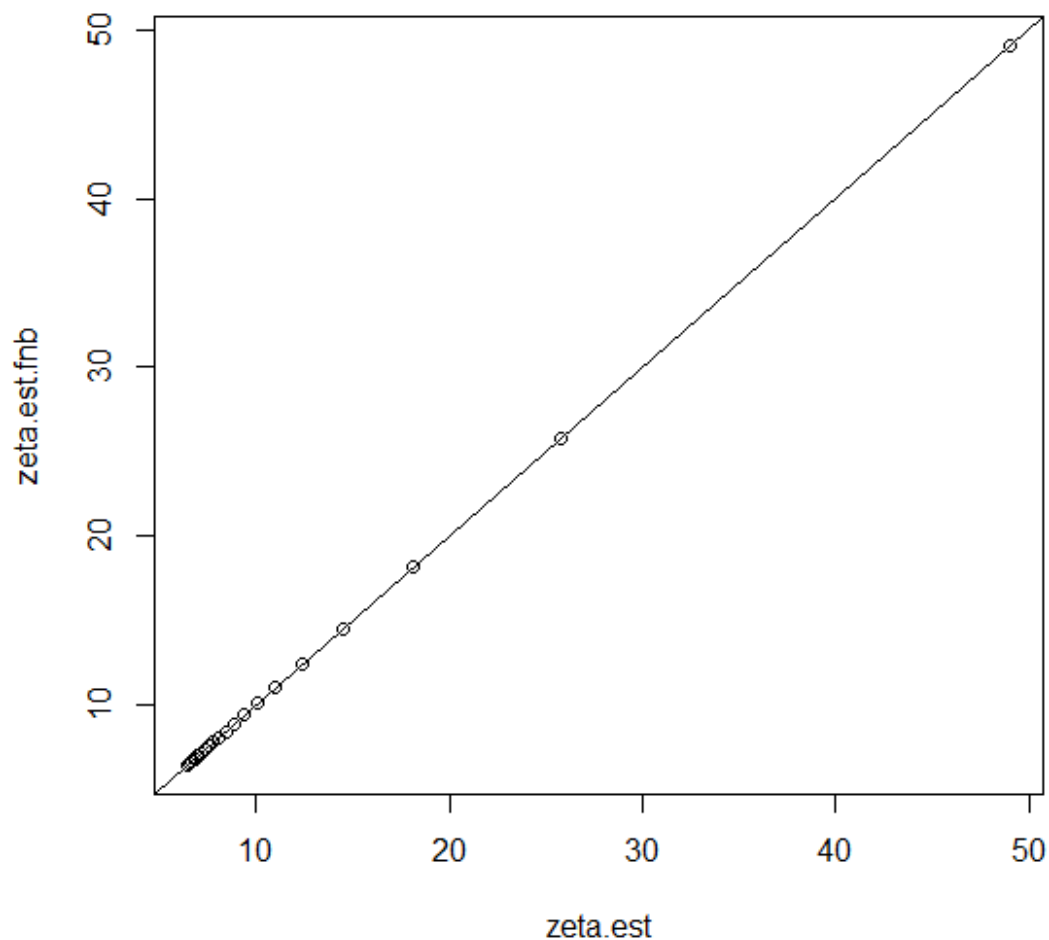
Calculate the scaling relationship

To fit a scaling relationship to a stem-mapped plot (or similar), just need the SAD for the plot and a bunch of species richness estimates at different grain sizes. There are data from BCI with this information in the object 'validationSamples_FNB.RData'.

```
source("functions.R")
load("BCI_SAD.RData")
load("validationSamples_FNB.RData")
names(val.sampsFNB[[1]])
```

```
## [1] "qarea"  "srmod"  "srobs"  "gammod" "gamobs" "spemod" "speobs"
"endmod"
## [9] "endobs" "cest"   "cfit"
```

We'll need quadrat areas (qarea) and observed mean richness for that grain (srobs). Plus, we can compare the fitted c parameter with the original estimate (cest) using the finite NB

(it will be close, but it won't be the same). The same code could be run with the FNB model but it will take a while…

```
x<- val.sampsFNB[[1]] # bci samples
cp <- numeric()
for(i in 1:nrow(x)){
  area = x$qarea[i] # quadrat area (grain size)
  obs = x$srobs[i] # mean richness at this grain
  cp[i] <- fitc.NB(obs = obs, area = area, sad = bciSAD, tota = 5e+05, low =
0, upp = 100)$cpar
  print(c(cp[i], x$cest[i])) # compare with original model
}

## [1] 0.2908141 0.2859559
## [1] 0.5825164 0.5740980
## [1] 0.7998832 0.7875621
## [1] 1.018414 1.001288
## [1] 1.205550 1.182773
## [1] 1.378862 1.349404
## [1] 1.510469 1.473620
## [1] 1.652632 1.607238
## [1] 1.769279 1.714719
## [1] 1.864122 1.799919
```

Now we have the list of areas and the value of the scaling parameter. We just need to fit a power function to predict c at any sampling grain.

```
avec <- x$qarea/x$qarea[4] # use 400 m as base level and convert grain to
relative grain
k0 <- cp[4] # fitted scaling parameter
y <- cp
nly <- nls(y ~ k0*avec^z, start=list(z=0.2))
pry <- predict(nly)
with(x, plot(qarea, cp))
lines(x$qarea, pry)
```
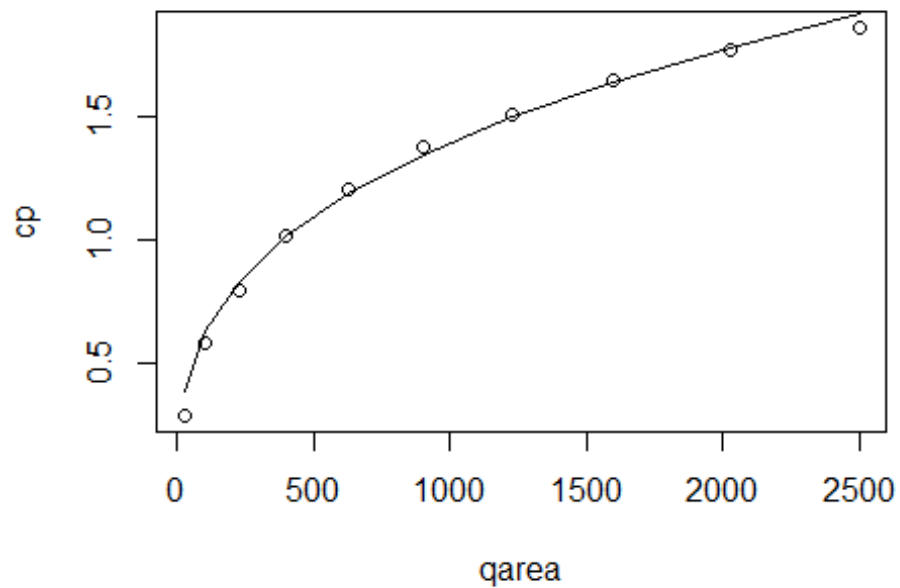
Once we have this relationship, we can explore subdivision of area by adjusting c to the correct grain (as was done in Fig. 5 - see scr_simsSubDiv.R for that code).

Reference

Hui, C. and McGeoch, M. (2014) Zeta diversity as a concept and metric that unifies incidence-based biodiversity patterns American Naturalist 2014 Vol. 184 Issue 5 Pages 684-694