

More Individuals Hypothesis

Simulation ideas

Tim Szewczyk

2017 April 12

Background

The More Individuals Hypothesis (MIH) posits that communities with more individuals will contain more species, attempting to link productivity to diversity mechanistically. The empirical support is variable, with many studies failing to find a linear relationship between the number of individuals and the number of species and instead finding stronger support for a unimodal relationship.

One implicit assumption is that the abundances of each species (or perhaps the community as a whole) are relatively invariant across years. That is obviously not the case for individual species, most of which show dramatic & stochastic fluctuations interannually in abundance. Simulations show that this stochasticity inhibits our ability to detect long-term changes in carrying capacity due to climate change; they may do the same to our ability to detect associations between total abundance and species richness for a given community.

These simulations quantify the effects of interannual stochasticity in a theoretical (computerized) environment with population parameters and underlying relationships pre-determined.

Simulation description

There will need to be a lot of simulations to explore across a sufficiently broad parameter space. For each combination of parameters, a community will be simulated, with a Ricker model for each species that incorporates stochasticity into the annual growth rate. This takes the form of:

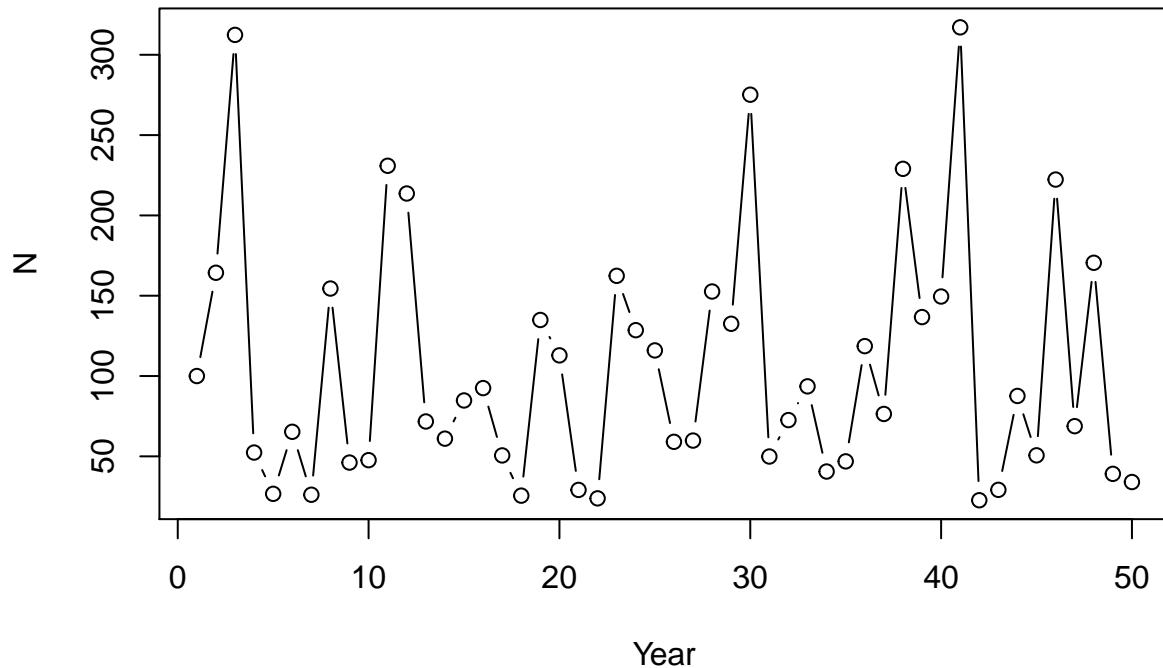
$$N_{t+1} = N_t * e^{r_t}$$

$$r_t \sim \text{Norm}\left(\mu = \bar{r}\left(1 - \frac{N_t}{K}\right), \sigma^2\right)$$

where N_t is the population abundance at time $t = 1$, r_t is the growth rate for the transition from t to $t+1$, \bar{r} is the average log intrinsic growth rate, K is the carrying capacity, and σ^2 is the annual variance in the population growth rate. For example:

```
# set parameters
N <- rep(NA, 50); N[1] <- 100
r <- 1
sd <- sqrt(0.5)
K <- 110

# simulate population for t=50 years
for(t in 2:50) {
  r.t <- rnorm(1, r*(1 - N[t-1]/K), sd)
  N[t] <- N[t-1]*exp(r.t)
}
plot(1:50, N, type="b", xlab="Year")
```



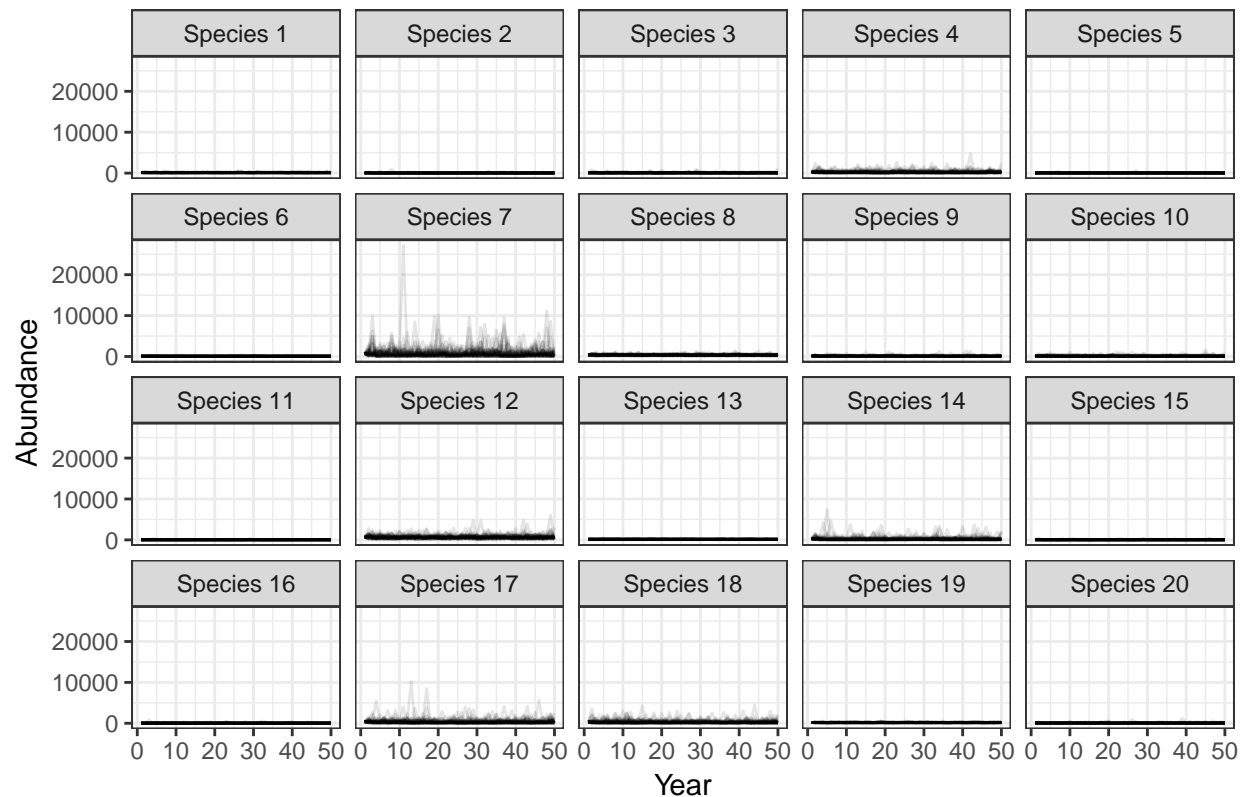
For each community, the abundances of all i species in the community will be simulated using Ricker models for each species in each of t years, with s simulations. The output for each community will be a 3-dimensional array (i, t, s).

```
# set parameters
nSim <- 50
nSpp <- 20
tMax <- 50
K.comm <- nSpp*200
N <- array(dim=c(tMax, nSpp, nSim))
r <- runif(nSpp, min=0.9, max=1.1)
sd <- sqrt(runif(nSpp, min=0.1, max=1))
K <- diff(sort(runif(nSpp, 0, K.comm)))
K <- c(K, K.comm-sum(K))
N[1,,] <- K

# simulation loop
for(s in 1:nSim) {
  N[,s] <- rick.mod.stoch(nSpp, tMax, r, N[,s], K, sd)
}

N.out <- summarize.out(N)
N.out$N.df$Species <- factor(N.out$N.df$Species, labels=paste("Species", 1:nSpp))
ggplot(N.out$N.df, aes(x=as.numeric(Year), group=Simulation, y=N)) +
  geom_line(alpha=0.1) + facet_wrap(~Species) +
  labs(x="Year", y="Abundance") + ggtitle("Abundance dynamics in a sample community")
```

Abundance dynamics in a sample community



One option is to model j communities with different species richnesses, with an explicit underlying positive relationship between S and K_{comm} . The model output could then be compared to look for that relationship. With no stochasticity, every simulation produces the expected linear relationship between S and N_{comm} :

```
# set global structural parameters
nSim <- 25
tMax <- 20
nComm <- 30
minS <- 5
maxS <- 45

# set global population parameters
K.multiplier <- 100 # K.comm = nSpp*K.multiplier -- built-in linear relationship
r.min <- 0.5 # min possible avg growth rate
r.max <- 1.1 # max possible avg growth rate
r <- runif(maxS, min=r.min, max=r.max) # avg growth rate for each species

nSpp <- seq.int(minS, maxS, length.out=nComm)
comm.ls <- vector("list", nComm)

for(j in 1:nComm) {
  # set community parameters
  K.comm <- nSpp[j]*K.multiplier # Community-level carrying capacity
  N <- matrix(nrow=tMax, ncol=nSpp[j]) # Abundance storage
  r.c <- r[1:nSpp[j]]
  K <- diff(c(0, sort(runif(nSpp[j]-1, 0, K.comm)), K.comm)) # Broken stick method for now
  N[1,] <- K
```

```

# simulation loop (parallel)
N.ls <- foreach(s=1:nSim) %dopar% {
  rick.mod.det(nSpp[j], tMax, r.c, N, K)
}
N <- abind(N.ls, along=3)

# store output
comm.ls[[j]] <- summarize.out(N)$N.comm
}

# reformat list into a dataframe
comm.df <- ldply(comm.ls, .id=NULL)

# fit models
out <- fit.mods(comm.df, tMax)

# all species
out$out.all[, .(nBest=sum(BestMod), propBest=round(sum(BestMod)/.N, 3)),
  by=.(Model)]

##           Model nBest propBest
## 1:      Linear    25         1
## 2: Log-Normal     0         0
## 3:   Quadratic     0         0

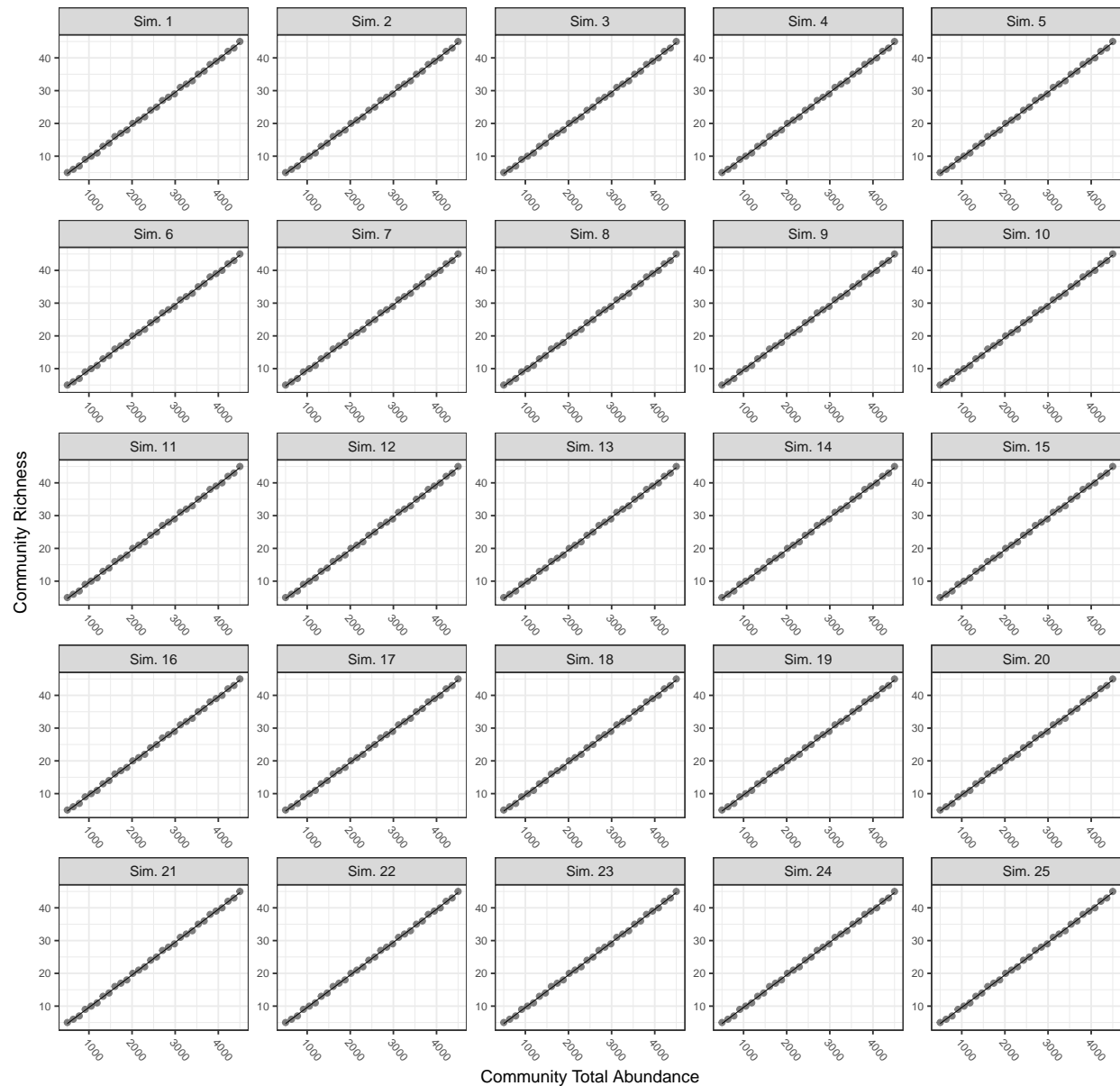
# excluding the most abundant species in each community
out$out.noDom[, .(nBest=sum(BestMod), propBest=round(sum(BestMod)/.N, 3)),
  by=.(Model)]

##           Model nBest propBest
## 1:      Linear    25         1
## 2: Log-Normal     0         0
## 3:   Quadratic     0         0

# plot representative simulations
comm.df$Sim <- factor(comm.df$Simulation, labels=paste("Sim.", 1:nSim))
ggplot(filter(comm.df, as.numeric(Simulation)<82, Year==20),
  aes(x=N.tot, y=S, group=Sim)) +
  facet_wrap(~Sim, scales="free") + geom_point(alpha=0.5) +
  stat_smooth(colour="black", se=FALSE, size=0.25, method="loess", span=1.1) +
  labs(x="Community Total Abundance", y="Community Richness") +
  ggtitle("Community abundance at t=20 with deterministic population dynamics.") +
  theme(axis.text.x=element_text(angle=310, hjust=0),
    axis.text=element_text(size=7))

```

Community abundance at t=20 with deterministic population dynamics.



With stochasticity incorporated into the growth rate for each species, the observed pattern becomes much more variable:

```
# set global structural parameters
nSim <- 800
tMax <- 20
nComm <- 30
minS <- 5
maxS <- 45

# set global population parameters
K.multiplier <- 100 # K.comm = nSpp*K.multiplier -- built-in linear relationship
r.min <- 0.5 # min possible avg growth rate
r.max <- 1.1 # max possible avg growth rate
r <- runif(maxS, min=r.min, max=r.max) # avg growth rate for each species
```

```

s2.min <- 0.01 # min possible species stochasticity
s2.max <- 2 # max possible species stochasticity
sd <- sqrt(runif(maxS, min=s2.min, max=s2.max)) # sd of r for each species

nSpp <- seq.int(minS, maxS, length.out=nComm)
comm.ls <- vector("list", nComm)

for(j in 1:nComm) {
  # set community parameters
  K.comm <- nSpp[j]*K.multiplier # Community-level carrying capacity
  N <- matrix(nrow=tMax, ncol=nSpp[j]) # Abundance storage
  r.c <- r[1:nSpp[j]]
  sd.c <- sd[1:nSpp[j]]
  K <- diff(c(0, sort(runif(nSpp[j]-1, 0, K.comm)), K.comm)) # Broken stick method for now
  N[1,] <- K

  # simulation loop (parallel)
  N.ls <- foreach(s=1:nSim) %dopar% {
    rick.mod.stoch(nSpp[j], tMax, r.c, N, K, sd.c)
  }
  N <- abind(N.ls, along=3)

  # store output
  comm.ls[[j]] <- summarize.out(N)$N.comm
}

# reformat list into a dataframe
comm.df <- ldply(comm.ls, .id=NULL)

# fit models
out <- fit.mods(comm.df, tMax)

# all species
out$out.all[, .(nBest=sum(BestMod), propBest=round(sum(BestMod)/.N, 3)),
  by=.(Model)]

```

```

##           Model nBest propBest
## 1:      Linear   37   0.046
## 2: Log-Normal  305   0.381
## 3:   Quadratic  458   0.572

```

```

# excluding the most abundant species in each community
out$out.noDom[, .(nBest=sum(BestMod), propBest=round(sum(BestMod)/.N, 3)),
  by=.(Model)]

```

```

##           Model nBest propBest
## 1:      Linear   70   0.088
## 2: Log-Normal  110   0.138
## 3:   Quadratic  620   0.775

```

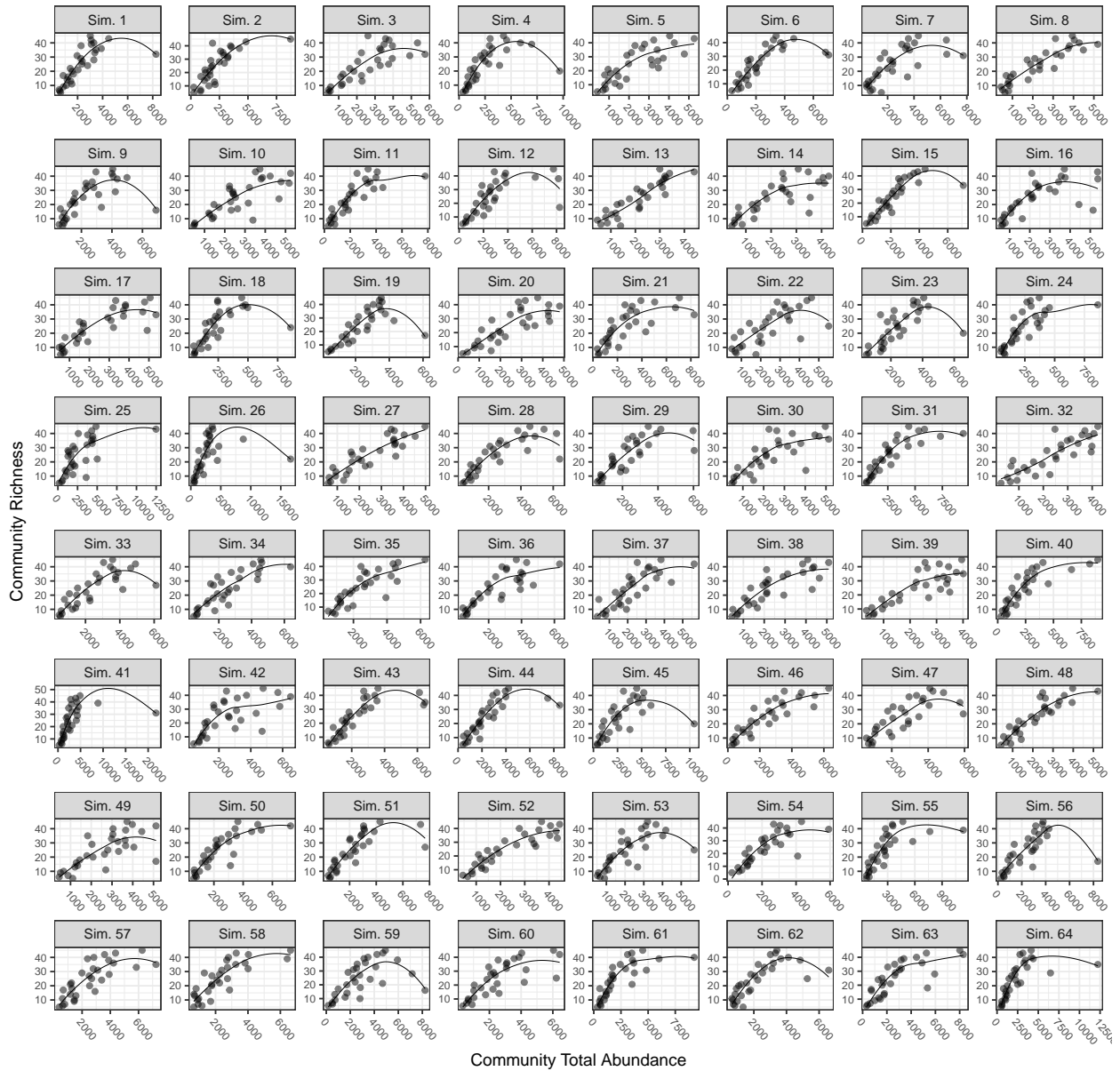
```

# plot representative simulations
comm.df$Sim <- factor(comm.df$Simulation, labels=paste("Sim.", 1:nSim))
ggplot(filter(comm.df, as.numeric(Simulation)<65, Year==20),
  aes(x=N.tot, y=S, group=Sim)) +
  facet_wrap(~Sim, scales="free") + geom_point(alpha=0.5) +

```

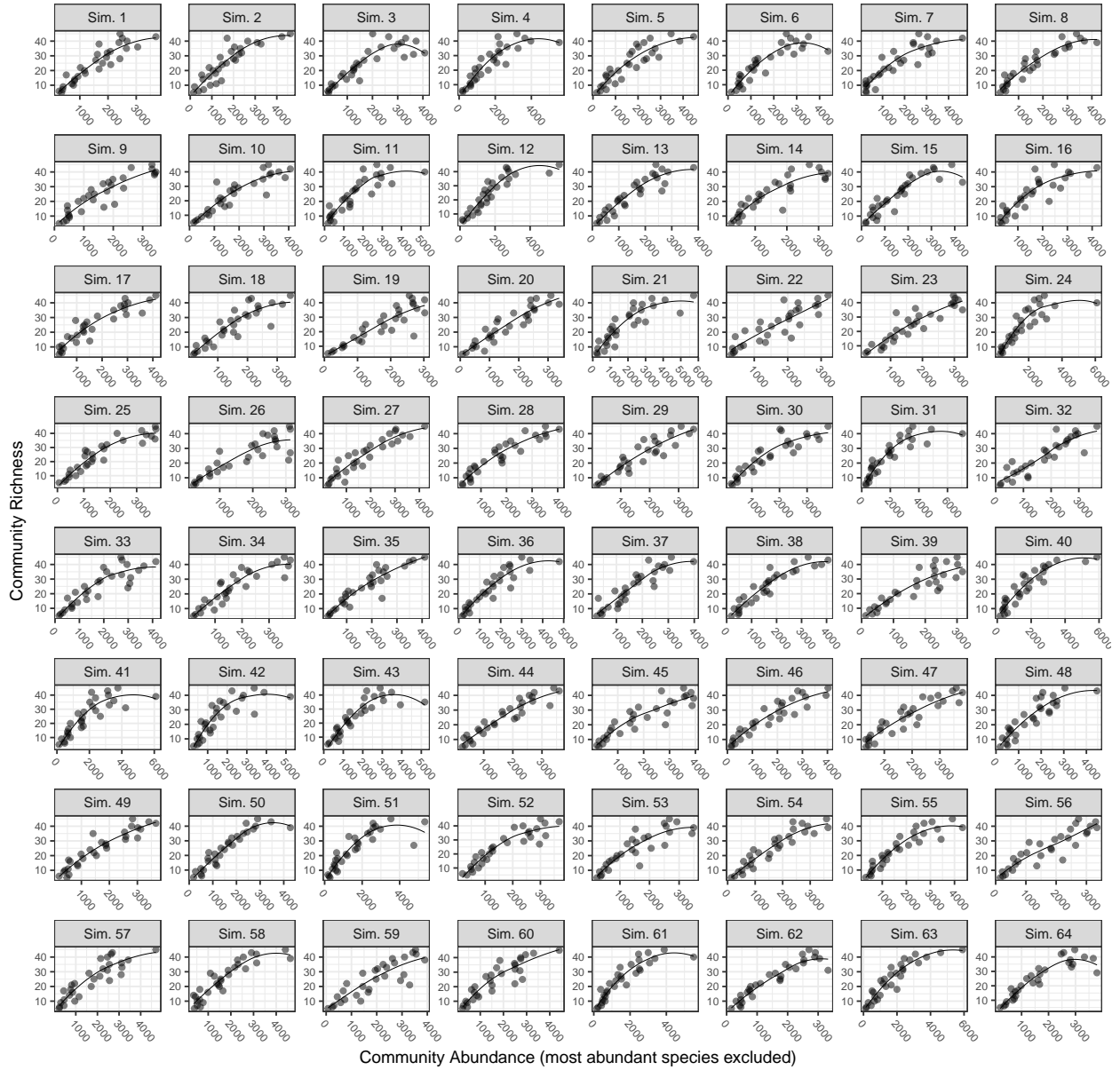
```
stat_smooth(colour="black", se=FALSE, size=0.25, method="loess", span=1.1) +
labs(x="Community Total Abundance", y="Community Richness") +
ggtitle("Community abundance at t=20 with stochastic population dynamics.") +
theme(axis.text.x=element_text(angle=310, hjust=0),
axis.text=element_text(size=7))
```

Community abundance at t=20 with stochastic population dynamics.



```
ggplot(filter(comm.df, as.numeric(Simulation)<65, Year==20),
aes(x=N.noDom, y=S, group=Sim)) +
facet_wrap(~Sim, scales="free") + geom_point(alpha=0.5) +
stat_smooth(colour="black", se=FALSE, size=0.25, method="loess", span=1.1) +
labs(x="Community Abundance (most abundant species excluded)", y="Community Richness") +
ggtitle("Community abundance without most abundant species at t=20 with stochastic population dynamics.") +
theme(axis.text.x=element_text(angle=310, hjust=0),
axis.text=element_text(size=7))
```

Community abundance without most abundant species at $t=20$ with stochastic population dynamics.



For each simulation, we could fit a linear model, quadratic, etc and use AIC to determine the best model. Then, the main output to analyze would be the proportion of simulations best described by each model. The models to evaluate are easily updated in the *fit.mods()* function.

Details that need to be decided:

- Distribution for K (Broken stick, geometric series, Fisher's log-series, Zipf-Mandelbrodt, Metacommunity zero-sum multinomial, Pareto, compound Poisson-gamma, Poisson-lognormal, compound Poisson-exponential, Puyeo's power-bend discrete, power discrete, etc – many options in *sads* package)
- Within a simulation, r_i and σ_i^2 are set for all species, then communities are built from nested subsets of these species, such that all species in community j are also in community $j+1$. However, K_i is determined randomly in each community. So while the population growth parameters for each species are constant across communities, the carrying capacities are not. To model a linear relationship between S and K_{comm} , I think this might be unavoidable.

- Which models to compare. Currently:
 - $S \sim N.\text{comm}$
 - $S \sim \log(N.\text{comm})$
 - $S \sim N.\text{comm} + N.\text{comm}^2$