Errata Zuur, Saveliev, Ieno (2012).

Zero Inflated Models and Generalized Linear Mixed Models with R

Chapter 6 (corrected in the second print):

Page 193: Change

```
M3 <- zeroinfl(RA ~ fYear + Temperature + fBottomType +
            offset(LogSA),
            dist = "poisson", link = "logit",
            data = Skates)
```

to:

```
M3 <- zeroinfl(RA ~ fYear + Temperature + fBottomType +
            offset(LogSA) | fYear + Temperature + fBottomType,
            dist = "poisson", link = "logit",
            data = Skates)
```

This ensures that the offset is not included in the binomial part of the model. Estimated parameters, standard errors, *z*-values, *p*-values, overdispersion and likelihood ratio test results on pages 194-195 will change slightly.

The same must be done on page 197 with the ZINB, on page 203 with the ZIP, and also on page 218 (although the R code is not shown on this page). Differences between the results obtained by the faulty and corrected code are minimal for all models. (Mainly in digits, except for the intercept in the binomial part, which makes sense)

Chapter 7 (corrected in the second print):

There is a small mistake in the R code in Subsection 7.5.2. The results therefore changed slightly. It is probably easier to copy and paste the modified Subsections 7.5.2 and 7.5.3, see below.

New subsections 7.5.2 and 7.5.3:

7.5.2 ZINB with a residual CAR correlation in R

Code to fit a ZINB with a residual CAR correlation structure is not available in packages such as `mgcv`, `nlme`, `lmer`, and `pscl`. We therefore formulate the ZINB in a Bayesian context and use the package `R2WinBUGS` to fit the model. Fitting ZIP and ZINB models within a Bayesian context was explained in detail in Chapter 4, and CAR GLM was discussed in Chapter 6.

The core of the code for the ZINB with spatial correlation follows. To reduce computing time we use the optimal ZINB model as starting point.

```
Model {
  #Likelihood
  for (i in 1 : N) {
    #Logit part
    W[i] ~ dbern(psim1[i])
    eta.psi[i] <- gamma[1] + gamma[2]*Depth[i] +
```

```
                 gamma[3]*SQDistRck[i] + gamma[4]*Swell[i] +
                 RE.bin[i]
    logit(psi[i]) <- max(-20, min(20, eta.psi[i]))
    psim1[i] <- min(0.99999, max(0.00001,(1 - psi[i])))

    #Negbin part
    p[i] <- size / (size + eff.mu[i])
    ParrotFish[i] ~ dnegbin(p[i], size)
    eta.mu[i]  <- beta[1] + beta[2]*Depth[i] + beta[3]*SST[i]+
                 RE.nb[i]
    log(mu[i]) <- max(-20, min(20, eta.mu[i]))
    eff.mu[i]  <- W[i] * mu[i]

    #CAR stuff
    EPS.mean[i] <- 0
  }

  #Priors
  for (i in 1:4) { gamma[i] ~ dnorm(0, 0.001) }
  for (i in 1:3) { beta[i]  ~ dnorm(0, 0.001) }
  size ~ dunif(0.2, 2)

  #Proper CAR prior distribution for spatial random effects:
  RE.bin[1:N]                                                  ~
car.proper(EPS.mean[],C[],adj[],num[],M[],tau.bin,
                           gamma.bin)
  RE.nb[1:N]                                                   ~
car.proper(EPS.mean[],C[],adj[],num[],M[],tau.nb,
                           gamma.nb)

  #CAR priors:
  tau.bin   ~  dgamma(0.01, 0.01)
  tau.nb    ~  dgamma(0.01, 0.01)
  gamma.min <- min.bound(C[], adj[], num[], M[])
  gamma.max <- max.bound(C[], adj[], num[], M[])
  gamma.bin ~  dunif(gamma.min, gamma.max)
  gamma.nb  ~  dunif(gamma.min, gamma.max)
}
```

The code that is under the `Likelihood` heading is nearly identical to the ZIP code in Kéry (2010). The only difference is that we use a negative binomial distribution, and the terms `RE.bin` and `RE.nb` in the link functions refer to the residuals ζ_i and ε_i, respectively.

Diffuse (non-informative) priors are used for all regression parameters $\alpha, \beta_1, ..., \beta_7, \gamma_0, ..., \gamma_7$. This explains the

```
for (i in 1:n.gamma) { gamma[i]  ~ dnorm(0, 0.001)}
for (i in 1:n.beta)  { beta[i]   ~ dnorm(0, 0.001) }
```

in the code. The WinBUGS function `car.proper` carries out the residual CAR model; its input matrices `C[]` and `M[]` need to be specified before running the MCMC. These are the **C** and **M** matrices in vector format (see Chapter 6). We use a threshold of 10 metres to define neighbouring sites, and set all elements of **C** to 0 for non-neighbouring sites (distance ≥ 10 metres). For neighbouring sites (distance < 10 meters), the elements of **C** are set to the reciprocal of the number of neighbours (see also the GeoBUGS user manual). This means that the CAR correlation will capture small-scale correlation. The diagonal elements of **M** contain the reciprocal of the number of neighbours. This part of the code is difficult, and the relevant sections in the online GeoBUGS user manual are useful.

The variables `gamma.bin` and `gamma.nb` represent the spatial auto-correlation parameters ρ_ξ and ρ_ε, respectively. The variables `prec.bin` and `prec.nb` are the reciprocals of the square root of the variance terms for the residuals in the logistic and log-link function, respectively, σ_ζ and σ_ε. The `gamma.min` and `gamma.max` variables specify the range of the spatial correlation parameters (see also the help file of `car.proper`).

7.5.3 MCMC results

The chief point of interest in the numerical output is whether the spatial auto-regressive parameters ρ_ξ and ρ_ε are different from 0. If these parameters are 0, there is no spatial correlation (at least not the type of correlation that can be modelled with a CAR), in which case we can proceed with the ZINB model presented in Subsection 7.3.3.

For the MCMC settings, we use 3 chains with a burn-in of 5,000 iterations and a thinning rate of 1,000. The number of draws from the posterior distribution is 5,000,000, and the resulting 4,995 stored values are used to create 95% credibility intervals.

The MCMC chains mix well for some parameters, and all parameters pass the tests for stationarity. Posterior distributions for all the parameters follow. We are especially interested in the spatial auto-regressive parameters ρ_ξ and ρ_ε, which are the `gamma.bin` and `gamma.nb`.

```
Inference for Bugs model at "ZINBCAR.txt", fit using WinBUGS,
 3 chains, each with 5e+06 iterations (first 5000 discarded),
 n.thin = 1000, n.sims = 14985 iterations saved

            mean   sd   2.5%    25%    50%    75%  97.5% Rhat n.eff
gamma[1]   -15.6  9.2  -36.9  -21.0  -14.1   -8.7   -2.0    1   610
gamma[2]   -16.2  7.5  -33.0  -20.8  -15.3  -10.7   -4.3    1   550
gamma[3]    52.3 19.6   16.9   38.2   51.3   65.1   93.8    1   650
gamma[4]    13.6  6.6    3.1    8.7   12.8   17.6   28.8    1   500
beta[1]      1.6  0.1    1.4    1.6    1.6    1.7    1.9    1  2100
beta[2]      0.5  0.1    0.2    0.4    0.5    0.6    0.8    1 15000
beta[3]     -0.3  0.1   -0.5   -0.4   -0.3   -0.2   -0.1    1 15000
size         0.5  0.1    0.4    0.4    0.5    0.5    0.6    1  9600
gamma.bin    0.2  0.6   -0.9   -0.3    0.3    0.7    0.9    1  2000
gamma.nb     0.0  0.5   -0.9   -0.4    0.1    0.5    0.9    1 15000
tau.bin      0.6  9.1    0.0    0.0    0.0    0.0    0.5    1   700
tau.nb      38.9 51.8    2.5    8.6   20.0   47.3  185.3    1  4700
deviance  1405.5 23.1 1363.0 1389.0 1404.0 1421.0 1453.0    1  9500

For each parameter, n.eff is a crude measure of effective sample
size, and Rhat is the potential scale reduction factor (at
convergence, Rhat=1).

DIC info (using the rule, pD = var(deviance)/2)
pD = 266.2 and DIC = 1671.7
DIC is an estimate of expected predictive error (lower deviance is
better).
```

Results for the parameters ρ_ξ and ρ_ε indicate that 0 is within the 95% credibility interval, indicating that there is no strong spatial correlation. Note that the dispersion parameter k is similar to the one obtained by model `M5`. Using other values (e.g. 5 metres, 15 metres, 25 metres) for the distance threshold defining neighbouring sites gives similar results.

The bottom line is there is no strong spatial correlation; therefore we continue with the ZINB model presented in Subsection 7.3.3.

End of new text Subsections 7.5.2 and 7.5.3

Chapter 8 (corrected in the second print):

The label for Figure 8.8 was modified to: Figure 8.8. Histogram of …..