Detecting dispersion problems in GLMMs

Melina de Souza Leite

melina.souza-leite@ur.de

Postodoctoral Researcher

University of Regensburg

Ecological data analysis

Increased complexity and flexibility in ecological data modeling:

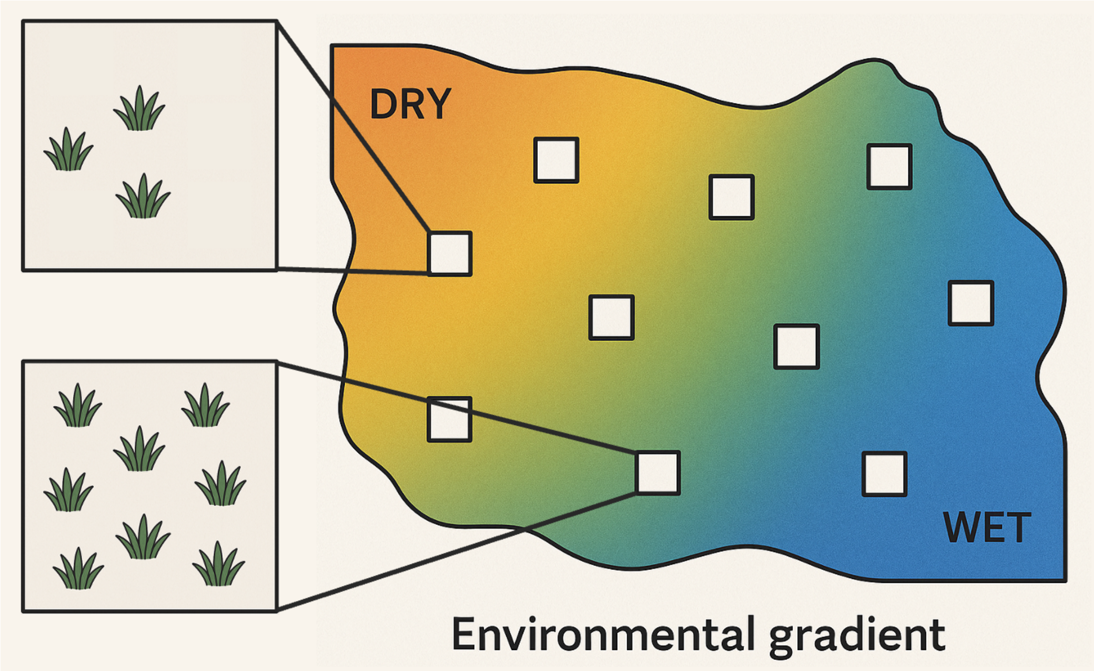
- Generalized linear modes (GLMs)
- Mixture models (e.g. zero-inflated GLMs)
- Hiearchical/Multilevel models, GLMMs

- But still few tools for model diagnostics
- Problem: **failing to check model assumptions**

Can you trust your model?

Dispersion problems in count data

- Example count data:
  - Species richness
  - Abundance of indiviuals
  - Number of success (K) within a number of trials (N)
- Modeling count data, GL(M)M distributions:
  - Poisson
  - Binomial (K/N) proportion



UNDER or OVERDISPERSION

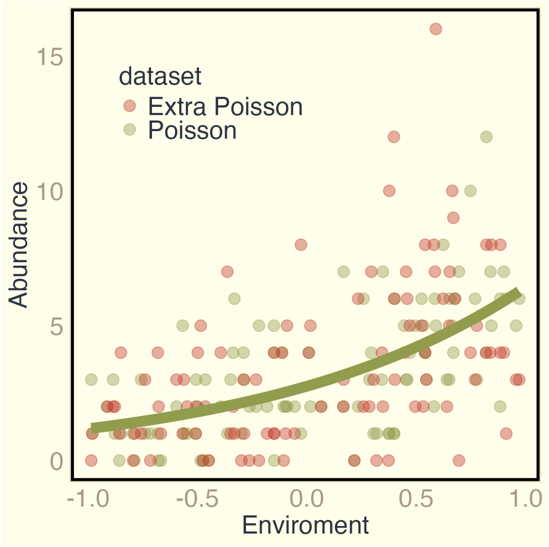
Problem when data has **more or less variability than expected by the distribution** used for modeling:

GOALS

- Aware ecologists of dispersion problems with count data
- Identify and describe the 3 main causes by using model diagnostics tools with `DHARMa`
- Show some modeling solutions for these causes

3 causes of dispersion problems

“Real” overdispersion:



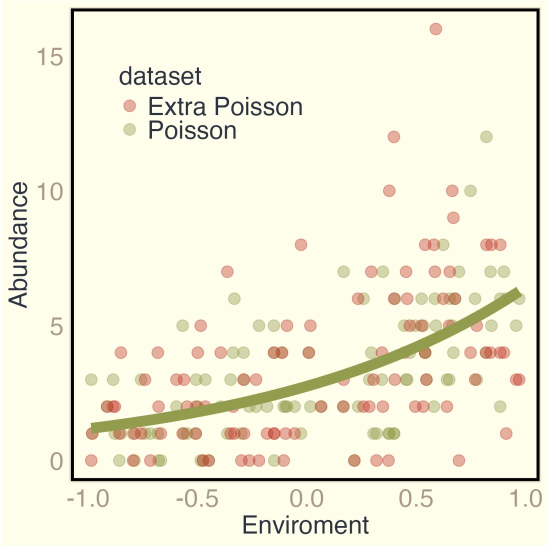
Abundances vary more that expected by the model, in general.

Heteroscedasticity:

Zero-inflation:

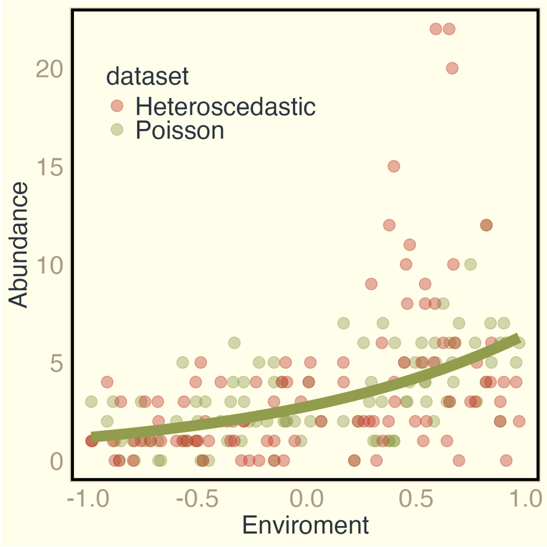
3 causes of dispersion problems

"Real" overdispersion:



Abundances vary more that expected by the model, in general.

Heteroscedasticity:

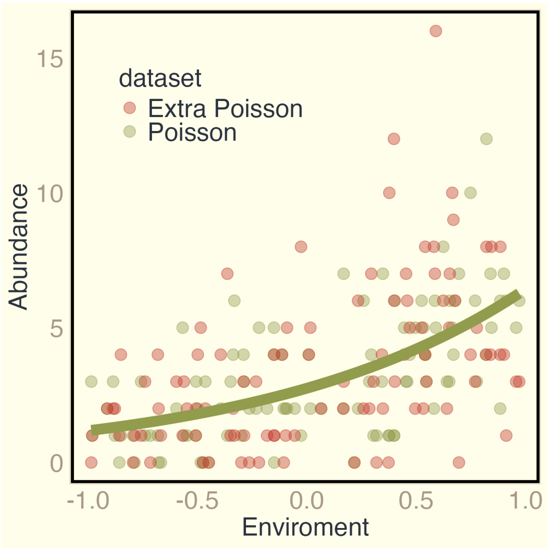


Abundances variation increases with the environmental gradient.

Zero-inflation:

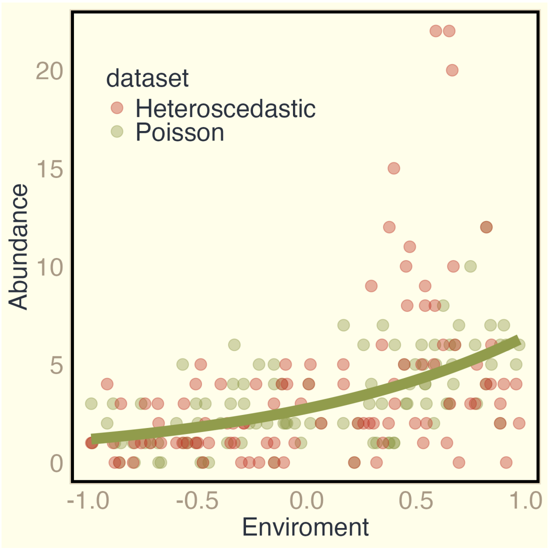
3 causes of dispersion problems

“Real” overdispersion:



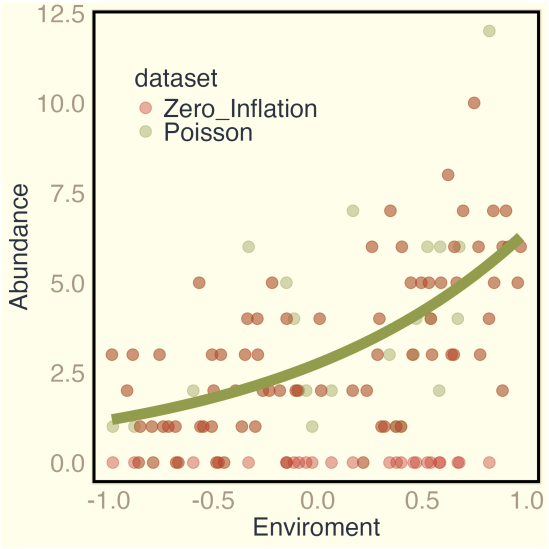
Abundances vary more that expected by the model, in general.

Heteroscedasticity:



Abundances variation increases with the environmental gradient.

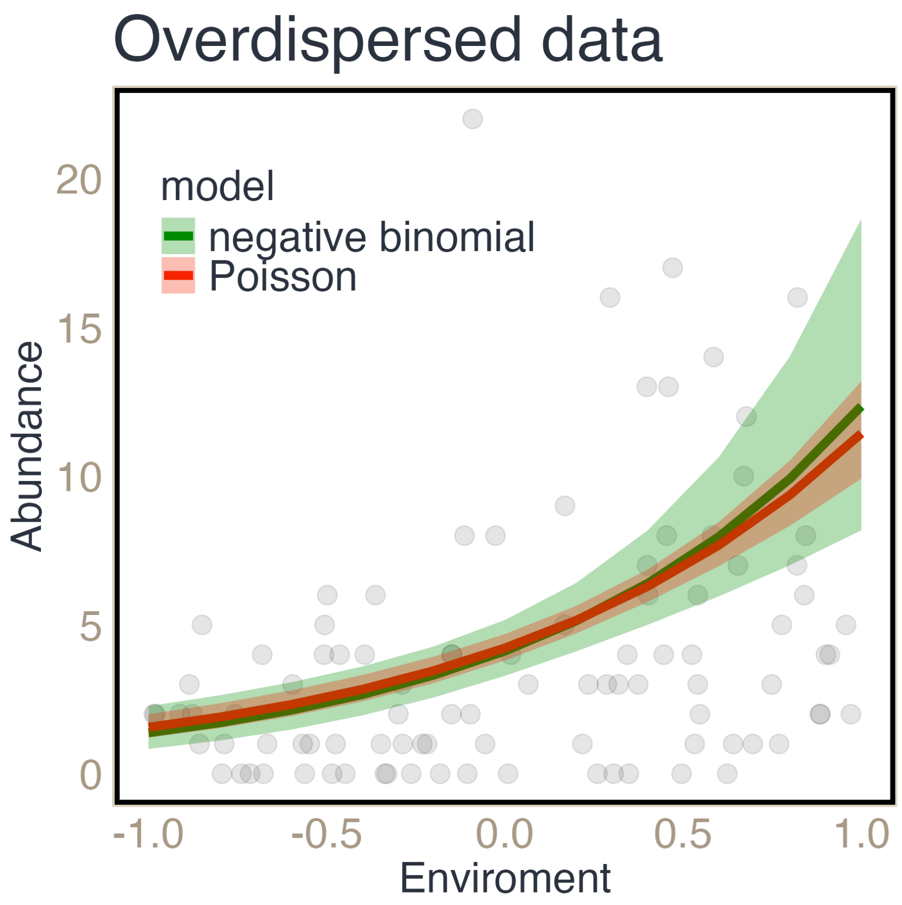
Zero-inflation:

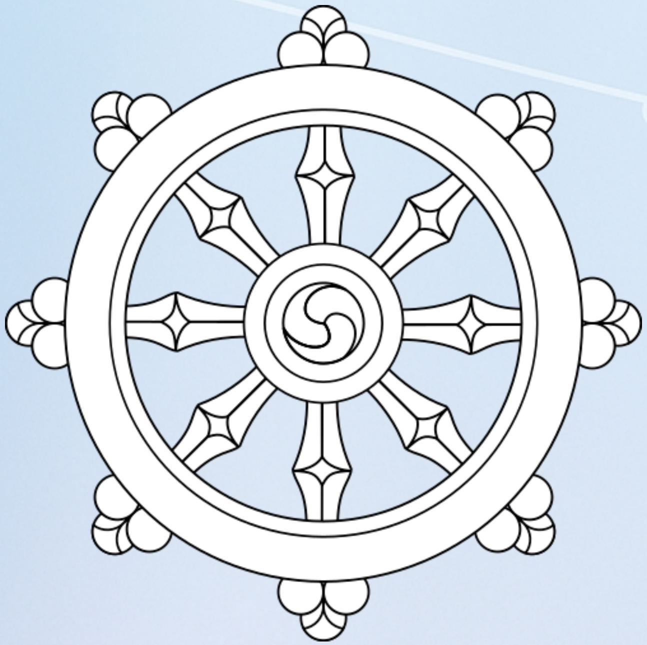


More zero abundances than expected by the model.

Consequences of dispersion problems

- Too small standard error of estimates -> narrower confidence intervals
- Larger chance of type I error: find an effect when it doesn't exist
- Wrong estimates by ignoring other processes (e.g. zero-inflation causes) in your data-generating process.
- Missing the opportunity to learn and get more info from your data. Ecological meanings for modeling/ understanding unexpected variability?



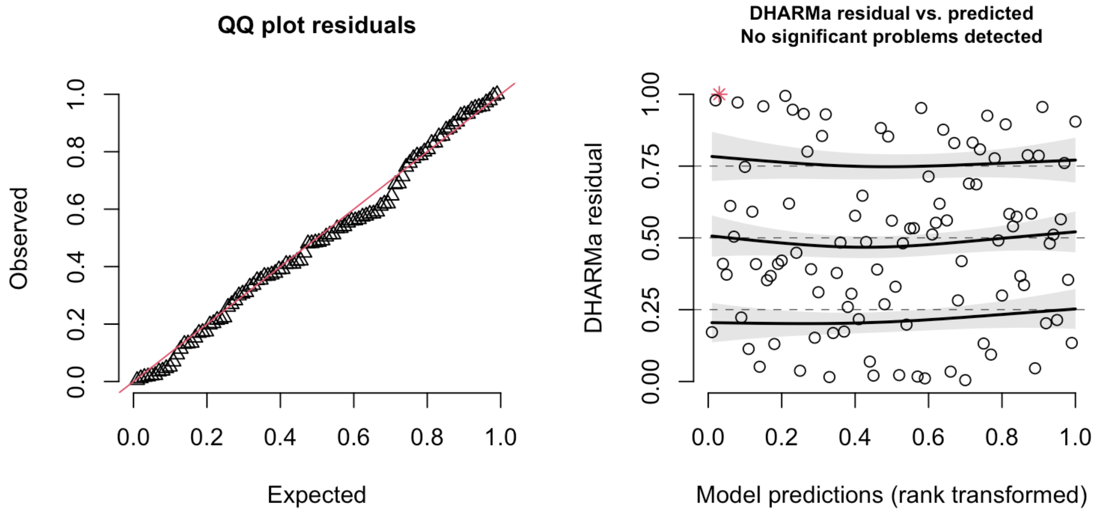


Detecting dispersion problems with DHARMa

Residual diagnostics with `DHARMa`

- Scaled quantile residuals -> Simulating from the model
- Residuals between 0 and 1 for ANY model complexity or distribution
- Interpreted the SAME way:

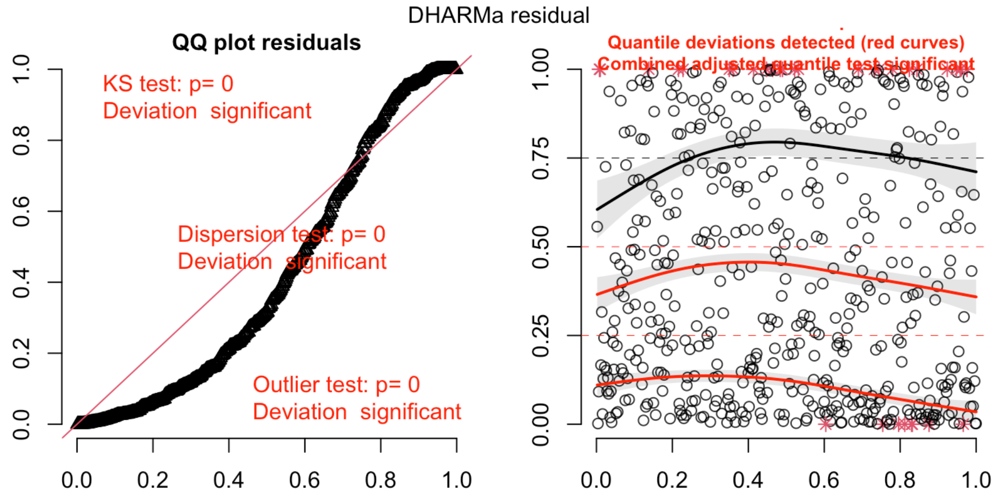
> If your model is correctly specified, i.e. your have the “data-generating process”, scaled quantile residuals will present a uniform “flat” distribution between 0 and 1.



Detecting “real” overdispersion

Wrong model

```r
m <- glmmTMB(observedResponse ~ Environment1 + (1|group),
          family = poisson(), data = overData)
res <- simulateResiduals(m)
plot(res)
testDispersion(res)
```

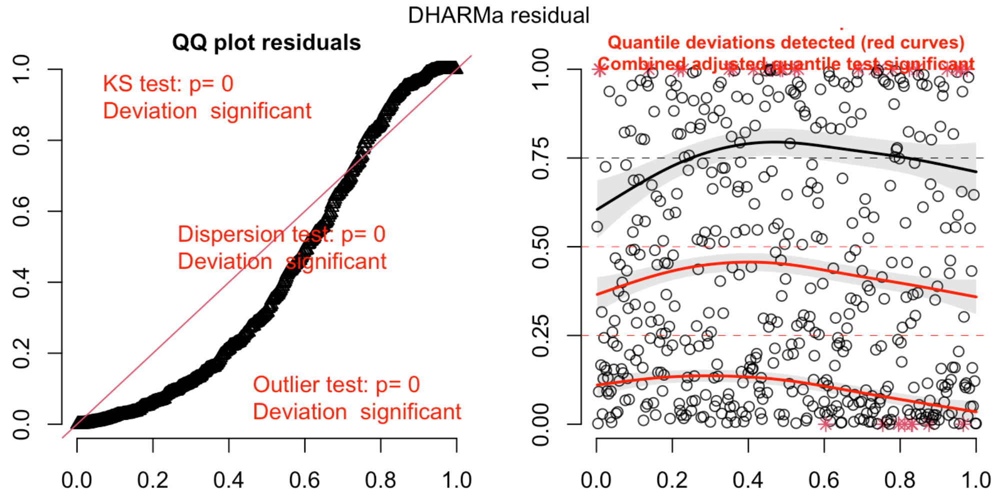


Dispersion = 5.19, p-value = 0.

Detecting "real" overdispersion

Wrong model

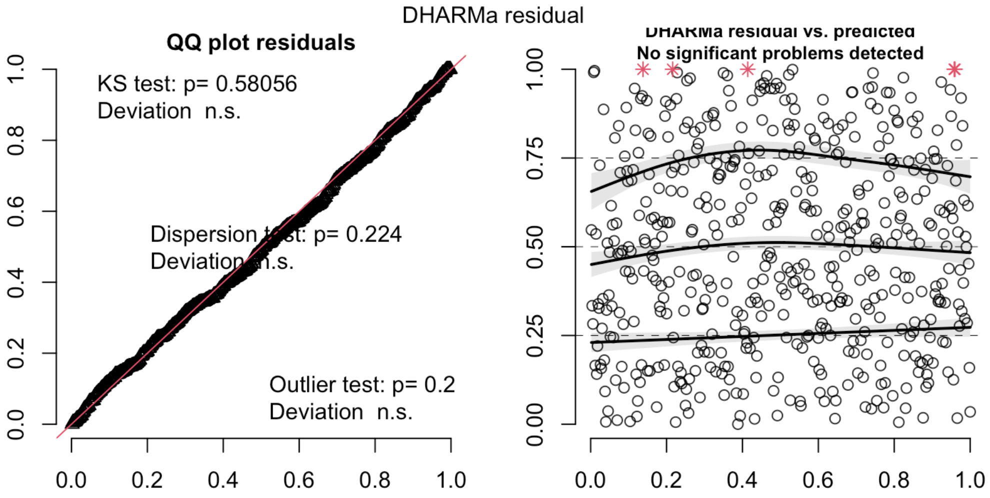
```r
m <- glmmTMB(observedResponse ~ Environment1 + (1|group),
        family = poisson(), data = overData)
res <- simulateResiduals(m)
plot(res)
testDispersion(res)
```



Dispersion = 5.19, p-value = 0.

Solution

```r
m <- glmmTMB(observedResponse ~ Environment1 + (1|group),
        family = nbinom2(), data = overData)
res <- simulateResiduals(m)
plot(res)
testDispersion(res)
```

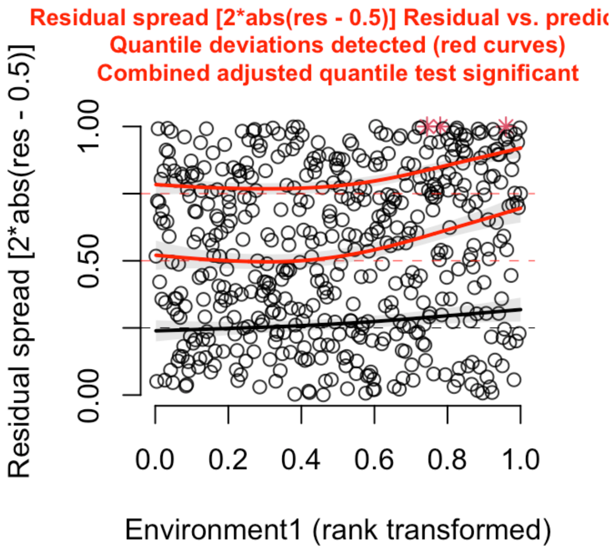


Dispersion = 1.19, p-value = 0.224.

Detecting heteroscedasticity

Wrong model

```r
m <- glmmTMB(observedResponse ~ Environment1 + (1|group),
        family = poisson(), data = overData)
res <- simulateResiduals(m)
plotResiduals(res, form = data$Environment1,
              absoluteDeviation = T)
testDispersion(res)
```

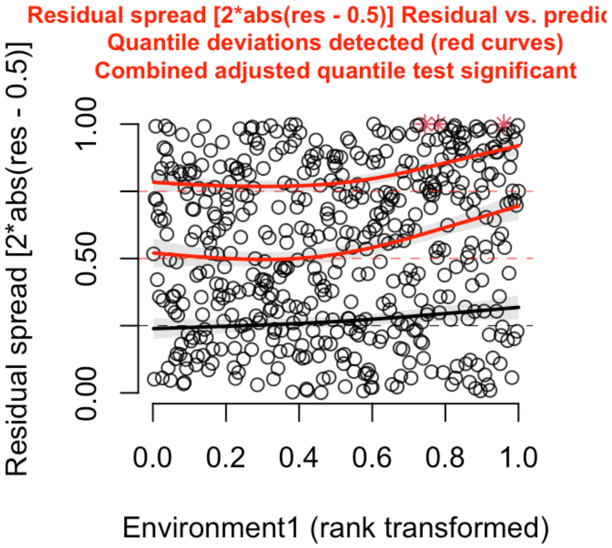


Dispersion = 1.9, p-value = 0.

Detecting heteroscedasticity

Wrong model

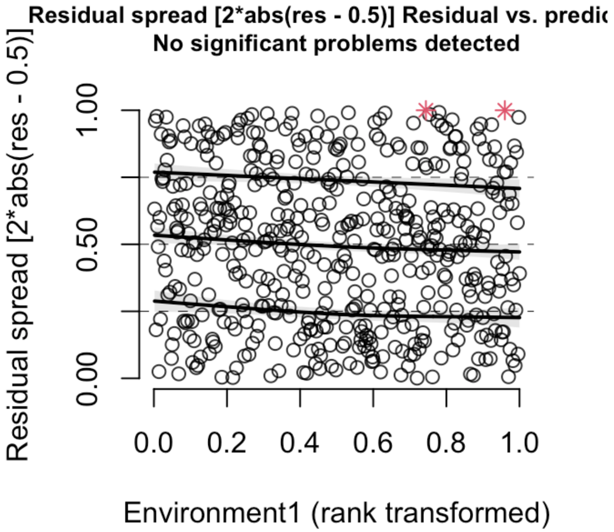
```r
m <- glmmTMB(observedResponse ~ Environment1 + (1|group),
          family = poisson(), data = overData)
res <- simulateResiduals(m)
plotResiduals(res, form = data$Environment1,
              absoluteDeviation = T)
testDispersion(res)
```



Dispersion = 1.9, p-value = 0.

Solution

```r
m <- glmmTMB(observedResponse ~ Environment1 + (1|group),
             dispformula = ~ Environment1, # dispersion formula
        family = nbinom2(), data = data) # but needs negative b
res <- simulateResiduals(m)
plotResiduals(res, form = data$Environment1,
              absoluteDeviation = T)
testDispersion(res)
```

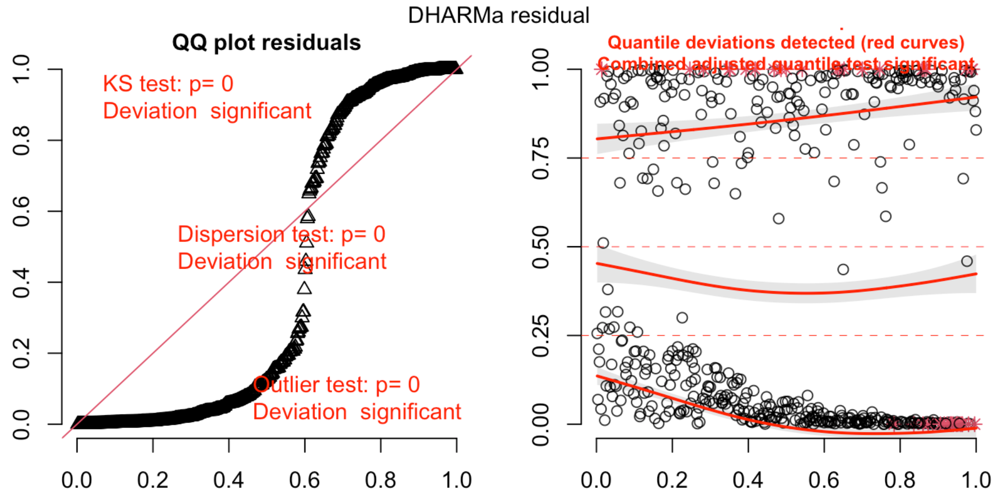


Dispersion = 1.11, p-value = 0.44.

Detecting zero-inflation

Wrong model

```r
m <- glmmTMB(observedResponse ~ Environment1 + (1|group),
        family = poisson(), data = overData)
res <- simulateResiduals(m)
plot(res)
testZeroInflation(res)
```

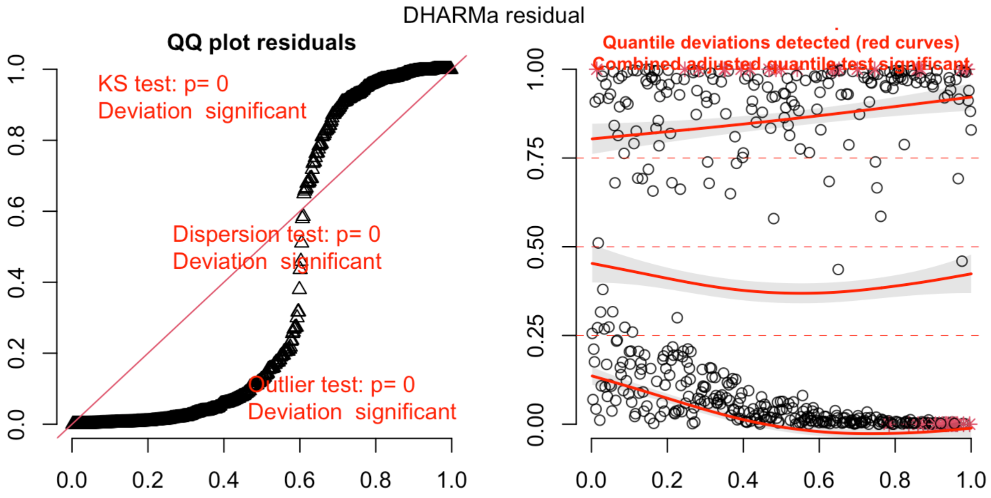


Zero-inflation = 5.16, p-value = 0.

Detecting zero-inflation

Wrong model

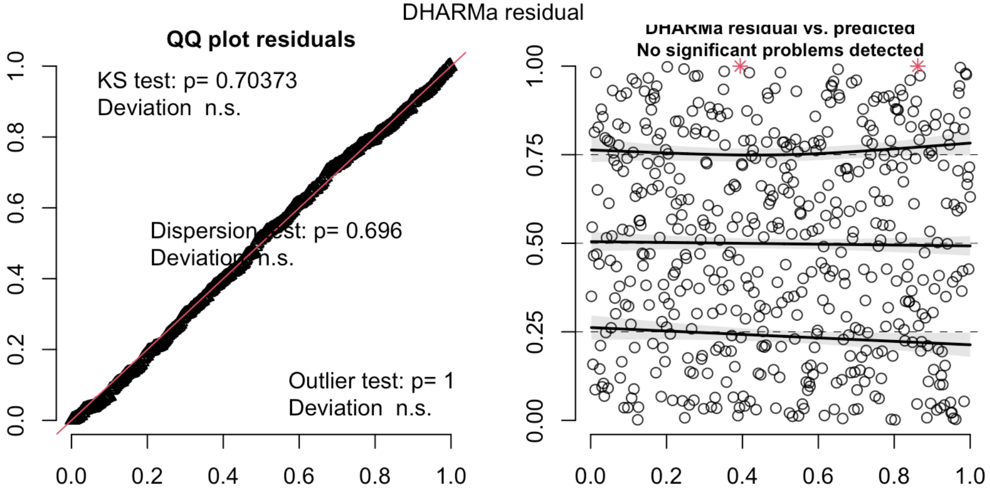
```r
m <- glmmTMB(observedResponse ~ Environment1 + (1|group),
        family = poisson(), data = overData)
res <- simulateResiduals(m)
plot(res)
testZeroInflation(res)
```



Zero-inflation = 5.16, p-value = 0.

Solution

```r
m <- glmmTMB(observedResponse ~ Environment1 + (1|group),
             ziformula = ~ 1,  # zero-inflation formula
         family = poisson(), data = data)
res <- simulateResiduals(m)
plotResiduals(res)
testZeroInflation(res)
```



Zero-inflation = 1, p-value = 1.

Detecting dispersion problems

- Residual patterns alone will not tell you which is the cause of overdispersion. E.g.:
  - ‘Real’ overdispersion will show significant test for zero-inflation, and vice-versa.
  - ‘Real’ overdispersion and zero-inflation may have significant heteroscedasticity.

- Additional check: fit models addressing the potential problems and compare their fit (e.g. AIC, LRT) and residuals diagnostics.

Don’t always assume the most complex/complicated model is the correct one!

Conclusion

- There are many causes of dispersion problems in GLMMs
- Use `DHARMa` residuals tools to detect them
- Address the problem with adequate models, e.g, `glmmTMB`

Take-home message

- Models should ALWAYS be checked: residual diagnostics!

- Avoid an oversimplistic view of dispersion problems

- Detecting and addressing the causes of dispersion problems may also be informative for your system/data.

Comming soon:

Leite et al. *in prep.* **Dispersion tests in GLMMs: a methods comparison and practical guide.**

Thank you!

Vielen Dank!

Acknowledgements to Florian Hartig, Max Pichler, and the Theoretical Ecology Lab group