

# Detecting dispersion problems in GLMMs

Melina de Souza Leite

[melina.souza-leite@ur.de](mailto:melina.souza-leite@ur.de)

Postdoctoral Researcher

University of Regensburg

# Replicability crisis in Ecology too?

- High false positive rates (type I error) in studies
- A non-intentional cause is **failing in checking model's assumptions**
- Researchers don't check models -> higher chances of false conclusions
- Few tools for model diagnostics of GLMs, GLMMs.

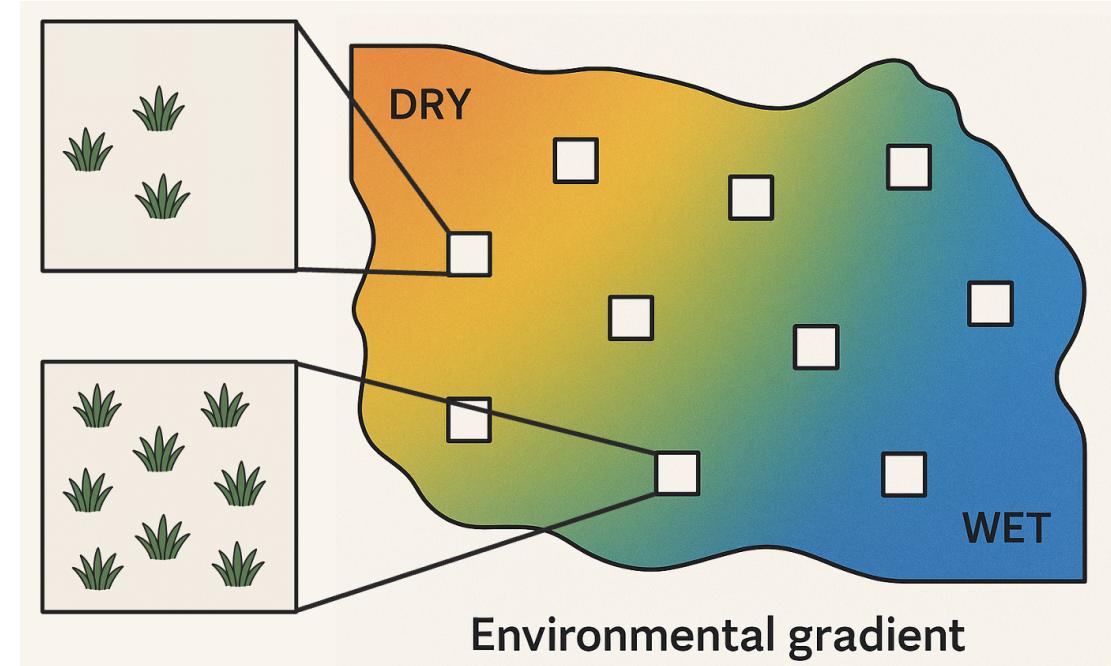


**Can you trust your model?**

# Dispersion problems in count data

- Example count data:
  - Species richness
  - Abundance of individuals
- Modeling with Poisson GLMs/GLMMs

**UNDER or OVERDISPERSION:**



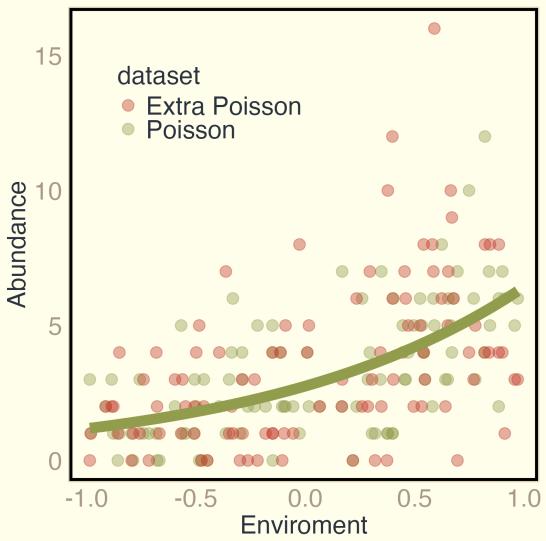
When data has more or less variability than expected by the distribution used for modeling.

# 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:



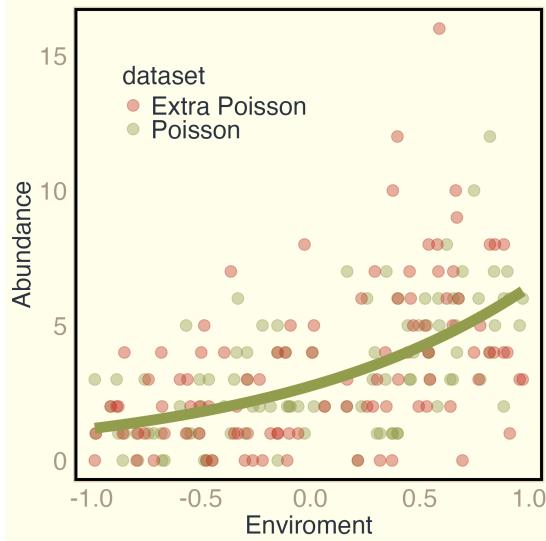
Heteroscedasticity:

Zero-inflation:

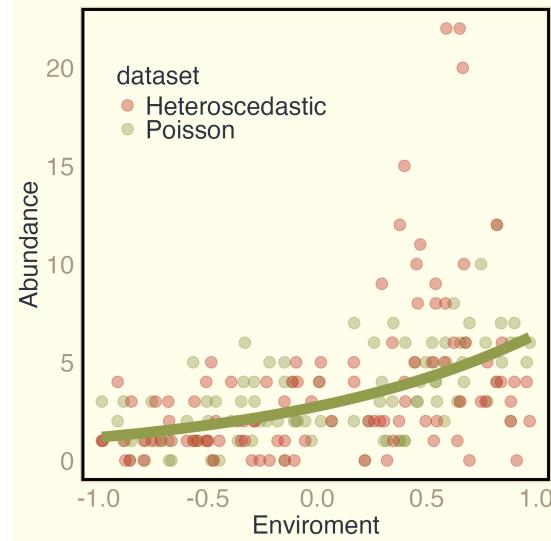
Abundances vary more than expected by the model, in general.

# 3 causes of dispersion problems

“Real” overdispersion:



Heteroscedasticity:



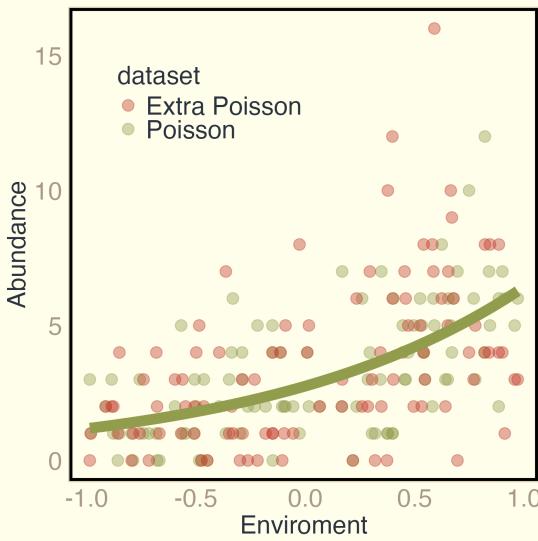
Zero-inflation:

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

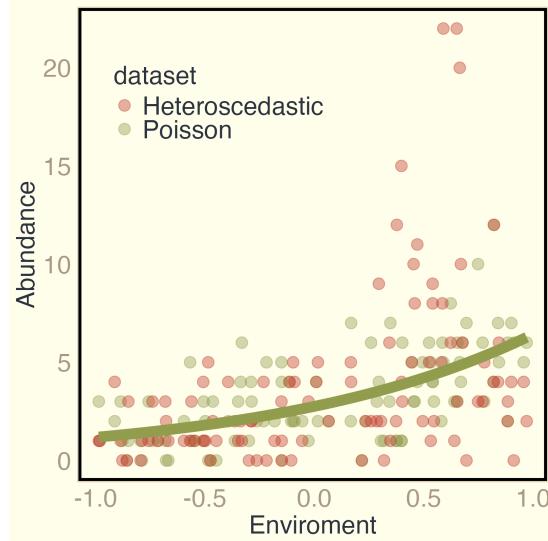
Abundances variation increases with the environmental gradient.

# 3 causes of dispersion problems

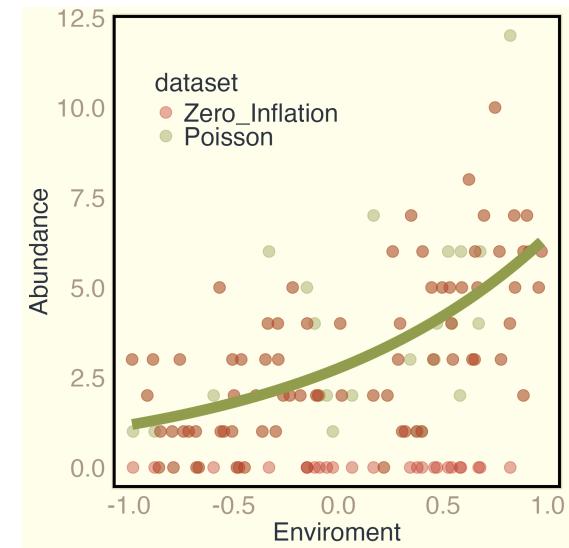
“Real” overdispersion:



Heteroscedasticity:



Zero-inflation:



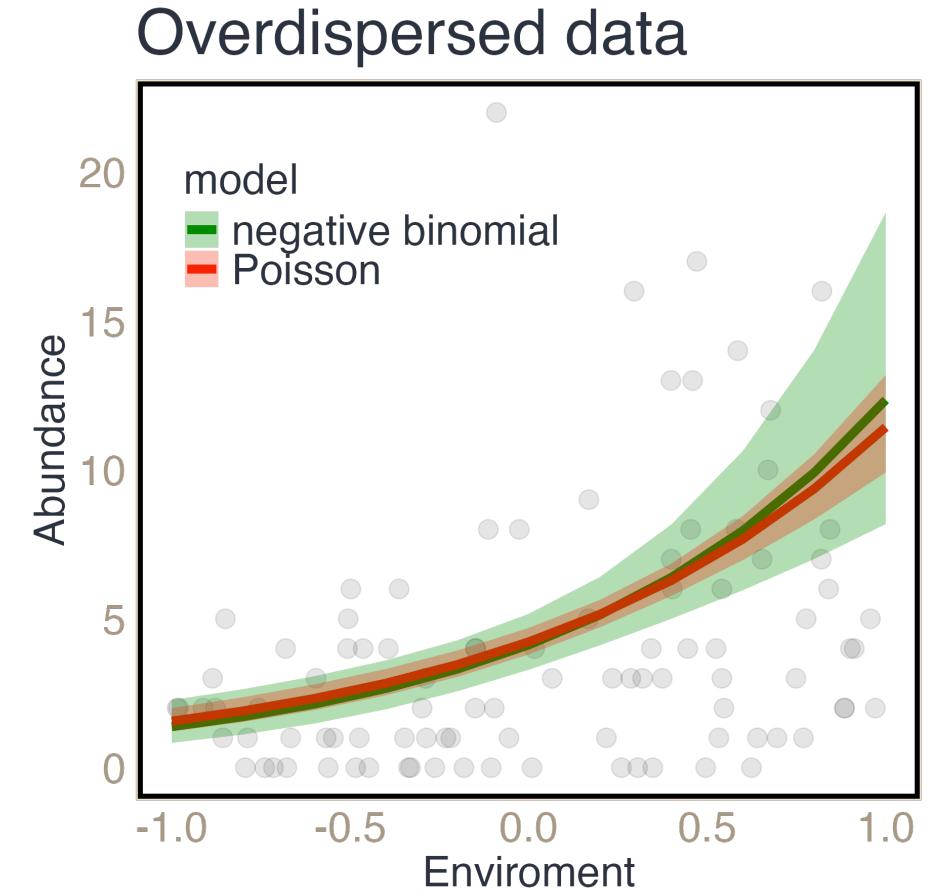
Abundances vary more than expected by the model, in general.

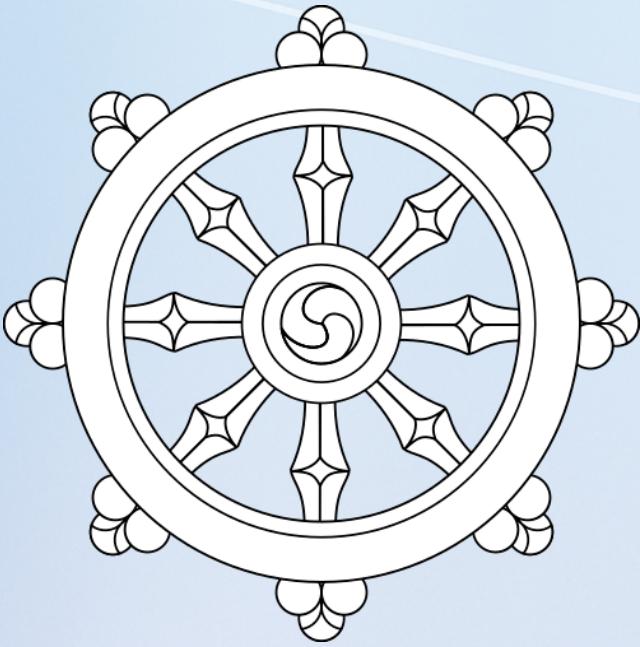
Abundances variation increases with the environmental gradient.

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?



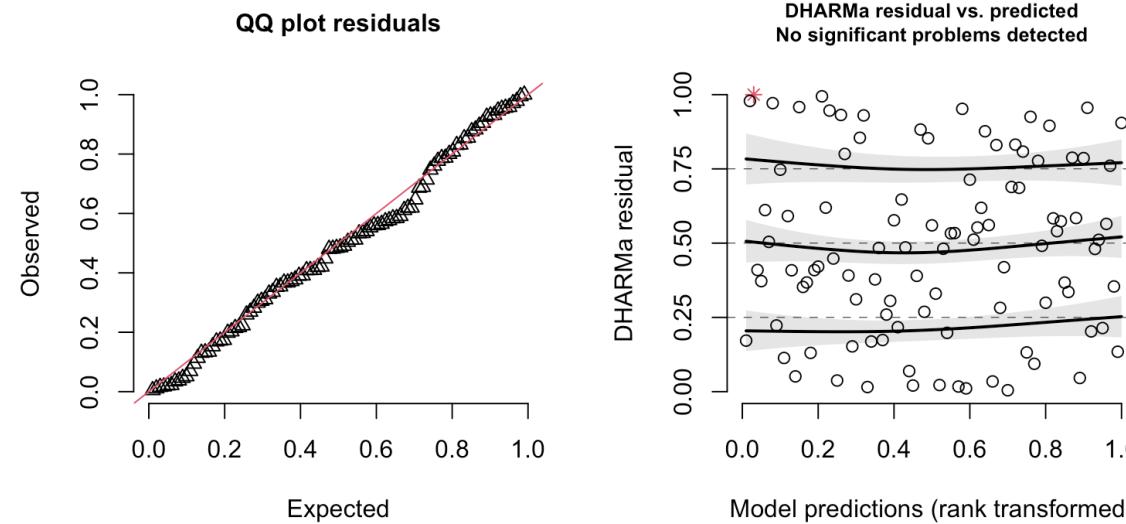


Detecting dispersion problems with DHARMa

# 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:

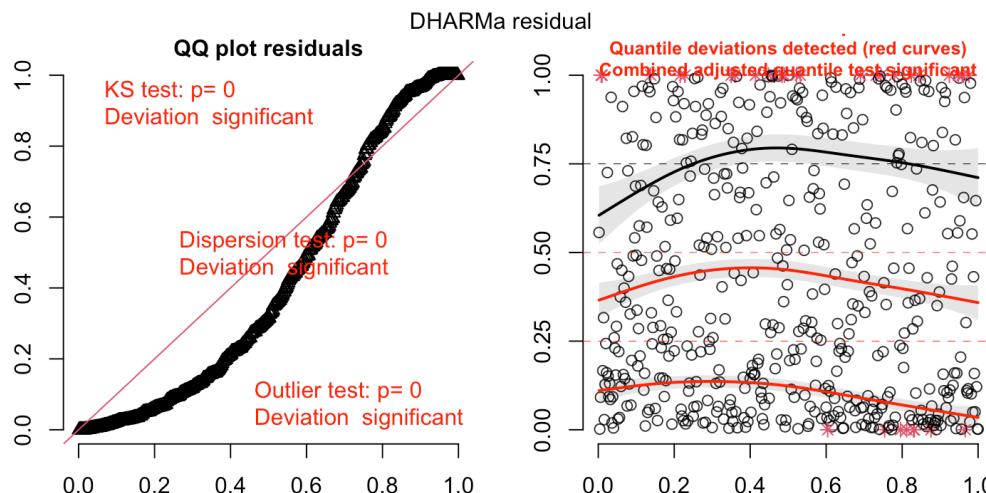
If your model is correctly specified, i.e. you have the “data-generating process”, scaled quantile residuals will present a uniform “flat” distribution between 0 and 1.



# Detecting “real” overdispersion

Wrong model

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

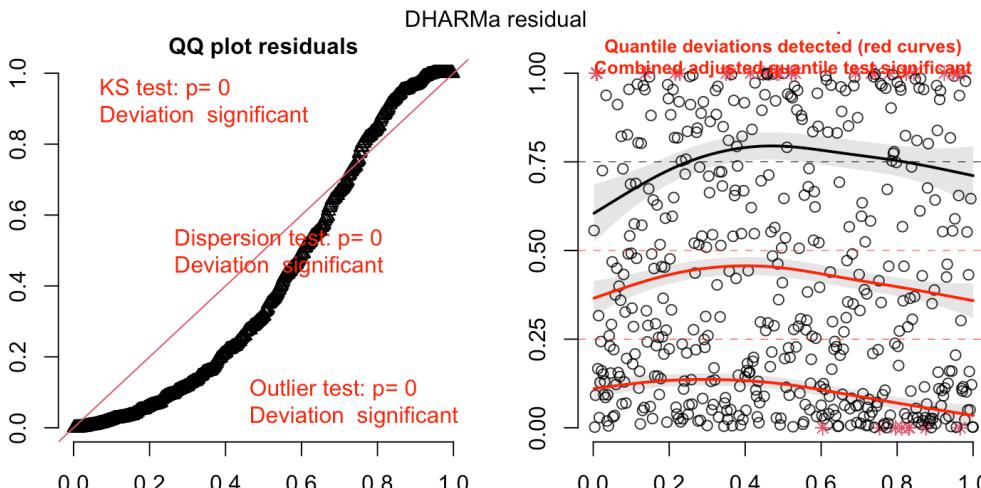


Dispersion = 5.19, p-value = 0.

# Detecting “real” overdispersion

Wrong model

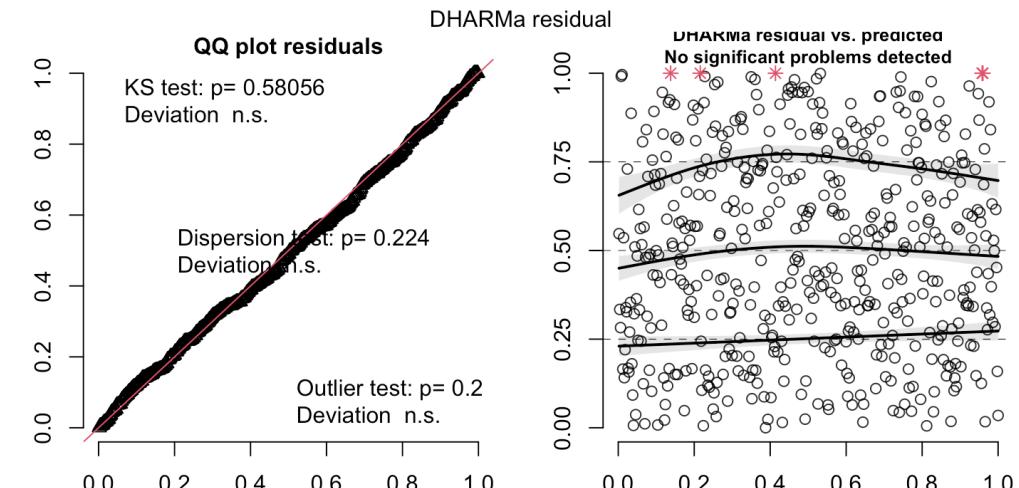
```
1 m <- glmmTMB(observedResponse ~ Environment1 + (1|group),  
2   family = poisson(), data = overData)  
3 res <- simulateResiduals(m)  
4 plot(res)  
5 testDispersion(res)
```



Dispersion = 5.19, p-value = 0.

Solution

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

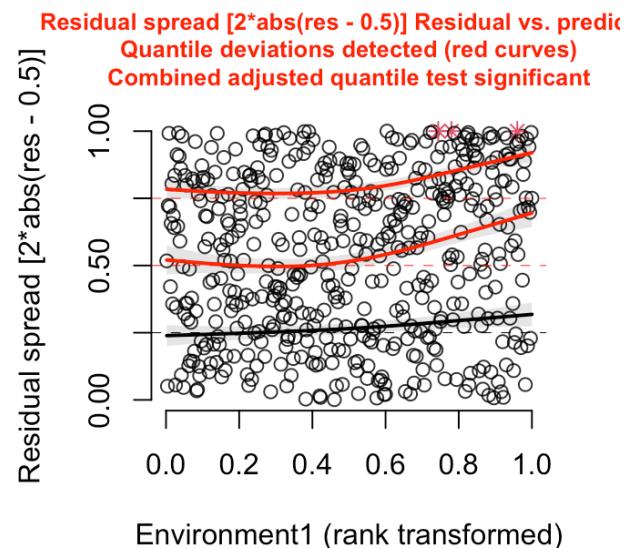


Dispersion = 1.19, p-value = 0.224.

# Detecting heteroscedasticity

Wrong model

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

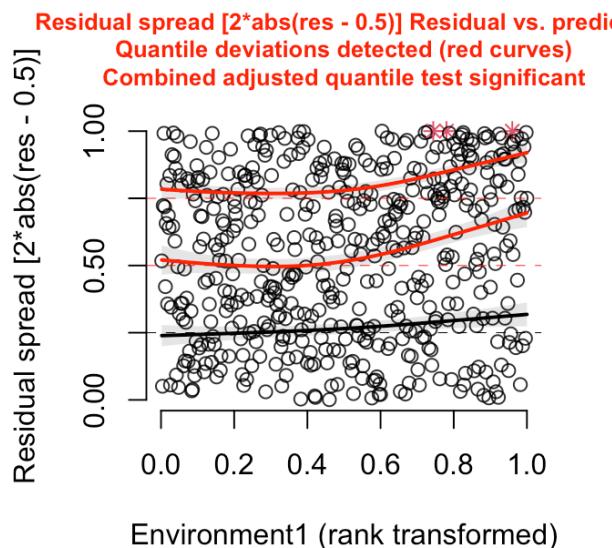


Dispersion = 1.9, p-value = 0.

# Detecting heteroscedasticity

Wrong model

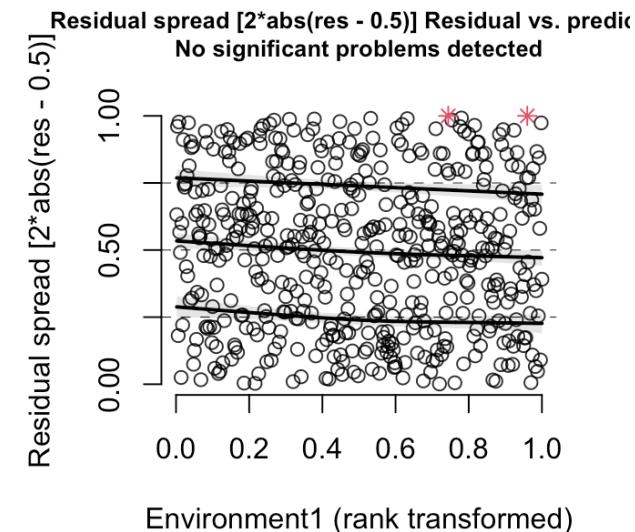
```
1 m <- glmmTMB(observedResponse ~ Environment1 + (1|group),  
2     family = poisson(), data = overData)  
3 res <- simulateResiduals(m)  
4 plotResiduals(res, form = data$Environment1,  
5                 absoluteDeviation = T)  
6 testDispersion(res)
```



Dispersion = 1.9, p-value = 0.

Solution

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

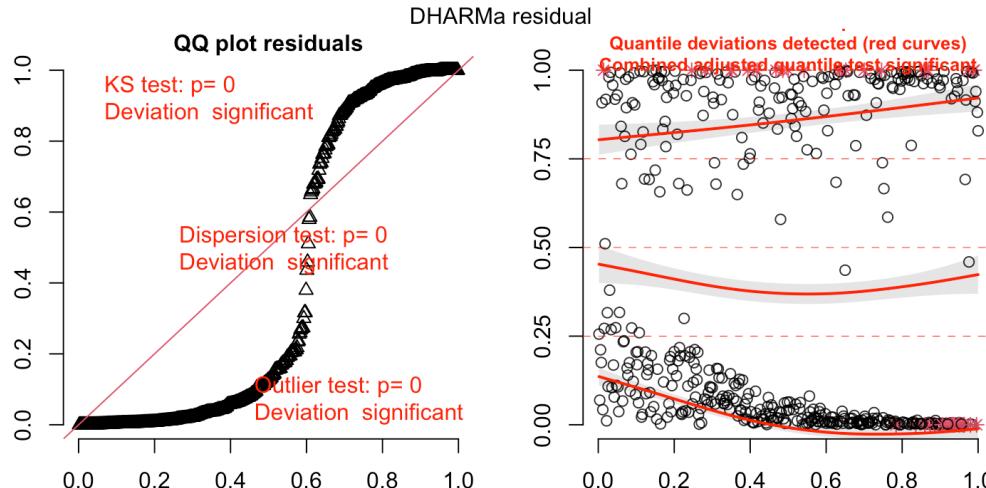


Dispersion = 1.11, p-value = 0.44.

# Detecting zero-inflation

Wrong model

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

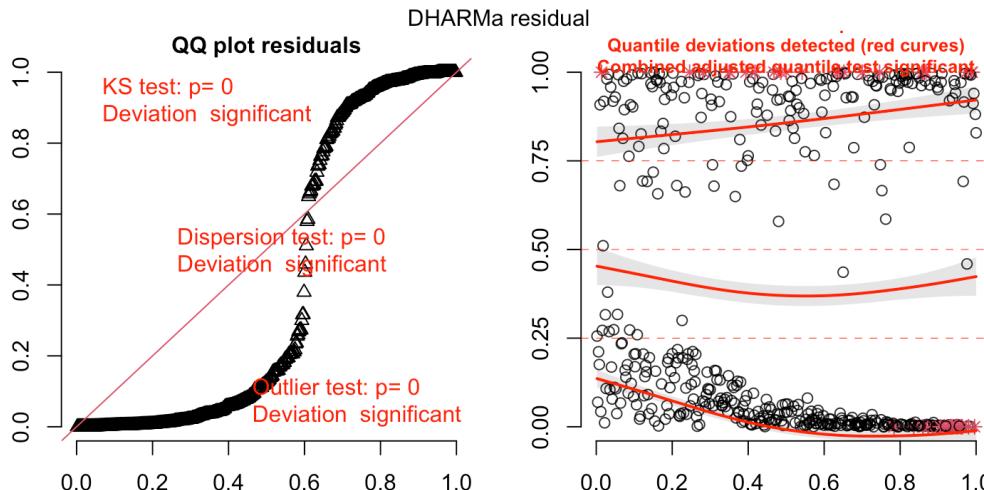


Zero-inflation = 5.16, p-value = 0.

# Detecting zero-inflation

## Wrong model

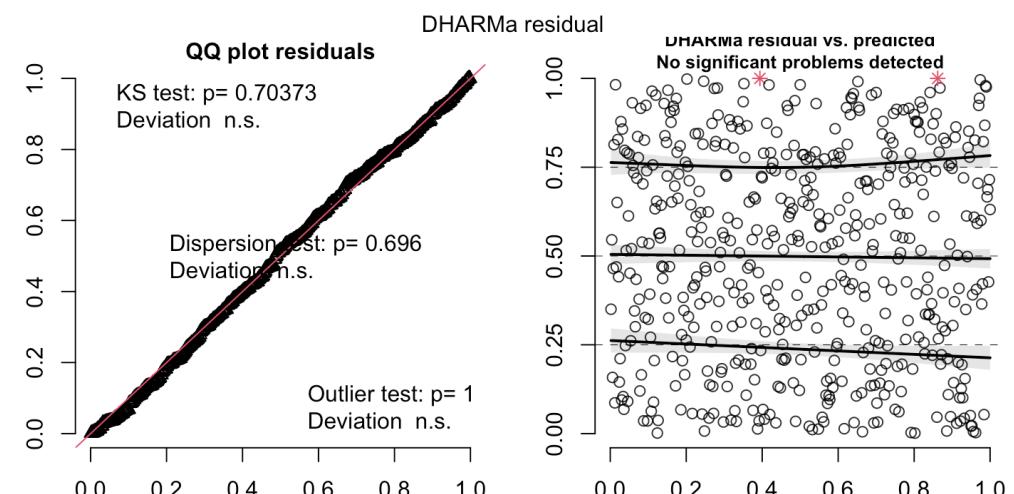
```
1 m <- glmmTMB(observedResponse ~ Environment1 + (1|group),  
2   family = poisson(), data = overData)  
3 res <- simulateResiduals(m)  
4 plot(res)  
5 testZeroInflation(res)
```



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

## Solution

```
1 m <- glmmTMB(observedResponse ~ Environment1 + (1|group),  
2   ziformula = ~ 1, # zero-inflation formula  
3   family = poisson(), data = data)  
4 res <- simulateResiduals(m)  
5 plotResiduals(res)  
6 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