

# Use the GLM, Luke.

## How the use of proper statistical models can increase statistical power in ecotoxicological experiments.

Eduard Szöcs

August 27, 2014

### 1. Introduction

Not mentioned in Newman (2013).

Sparsely in OECD Guideline.

Canada: 'The concept is quite advanced and as yet is not widely used in environmental toxicology.'

Wang/Riffel vergleichen NP aber kein Wort über GLM.

Brock et al. empfiehlt das sampling zu verbessern um die MDD zu verkleinern, keine Erwähnung von GLMs (damit erhöhen die auch die power, wie man sieht und sind kostenlos)

Schon oft angesprochen (O'Hara 2010, Warton 2010 (arcsine is asinine), ?), aber nicht in der Ökotox aufgenommen.

MDD eigentlich nur ein synonym für power.

Obwohl jeder weiß dass NOEC blöd sind - werden sie weiter genutzt (siehe zb. viele publicationen von Fox). Fox 2012: "A recurring concern expressed about the suggested move to model-based inference is that the derived toxicity measure will be influenced by a host of subjective decisions such as the choice of the mathematical model, its parameterization, and the estimation strategy."

Hoenig and Heisey: No post-hoc power estimation. Equally unhelpful are post hoc estimates of minimum significant difference or detectable ES. Once we have constructed a confidence interval, power calculations yield no additional insights'

### 1.1. Transformation + Normal Model

$$\begin{aligned} \log(Ay_i + 1) &\sim N(\mu_i, \sigma^2) \\ \log(Ay_i + 1) &= \alpha + \beta x_i \\ \text{var}(\log(Ay_i + 1)) &= \sigma^2 \end{aligned} \tag{1}$$

### 1.2. Generalized Linear Model

$$\begin{aligned} y_i &\sim NB(\mu_i, \theta) \\ \log(\mu_i) &= \alpha + \beta x_i \\ \text{var}(y_i) &= \mu_i + \mu_i^2 / \theta \end{aligned} \tag{2}$$

## 2. Methods

Literature review of SETAC Journals -¿ NOEC, GLM, log/arcsine transformation? How often, what is applied?

### 2.1. Simulation scenarios

We simulated data that mimics count data encountered in mesocosm experiments, with five treatments (T1-T5) and one control (C). Mean abundance in treatments T2-T5 was reduced to half of the control treatment and T1 ( $\mu_{T2} = \dots = \mu_{T5} = 0.5\mu_C = 0.5\mu_{T1}$ ). Therefore, the theoretical LOEC is at T2 and the NOEC correspondingly at T1. Counts were drawn from a negative binomial distribution with a fixed  $\theta = 3.91$  for all treatments.

We simulated datasets with different the number of replicates ( $N = \{3, 6, 9, 12\}$ ) and different abundances in control treatments ( $\mu_C = \{2, 4, 8, 16, 32, 64, 128, 256, 512, 1024\}$ ). For each combination we generated 100 datasets.

Each dataset was analysed with different methods/models.

#### 2.1.1. Global test of treatment effect

2 different models were fitted to the data. A model assuming a normal distribution after  $\ln(Ay + 1)$  transforming the response (eqn 1) and a model assuming that the response follows a negative binomial distribution. Treatment effects were in both cases investigated using a Likelihood-Ratio-Test. Moreover, we applied a non-parametric Kruskal-Wallis Rank Sum Test.

1) POWER = ability to detect an effect that is present. their accuracy at 2) (Type I error) : maintaining a significance level of 0.05 when there is no effect to be detected. Maybe check also this.

### **2.1.2. Determination of LOEC/NOEC**

multcomp, nparcomp, pairwise-wilcox. Check p-value adjustment between methods!

Add untransformed

Derzeit,  $\mu_T = 1/2 \mu_C$ . Deshalb nimmt die power ab mit kleiner  $\mu_C$  (die differenz wird kleiner). Wie kann man das umgehen/besser machen?

Andere Daten/Szenarios? Siehe Wang oder Newman.

## **3. Results**

### **3.1. Simulation scenarios**

-¿ bei kleinem  $\mu_C$  im besser weil annahmen nicht passen (checken)?

## **4. Discussion**

Decide between NB and P? - Mean-Var-Plot!

## **5. Conclusion**

### **A. Tipps (Or move up to introduction?) Check distribution**

### **B. TODO**

Read about nparcomp.

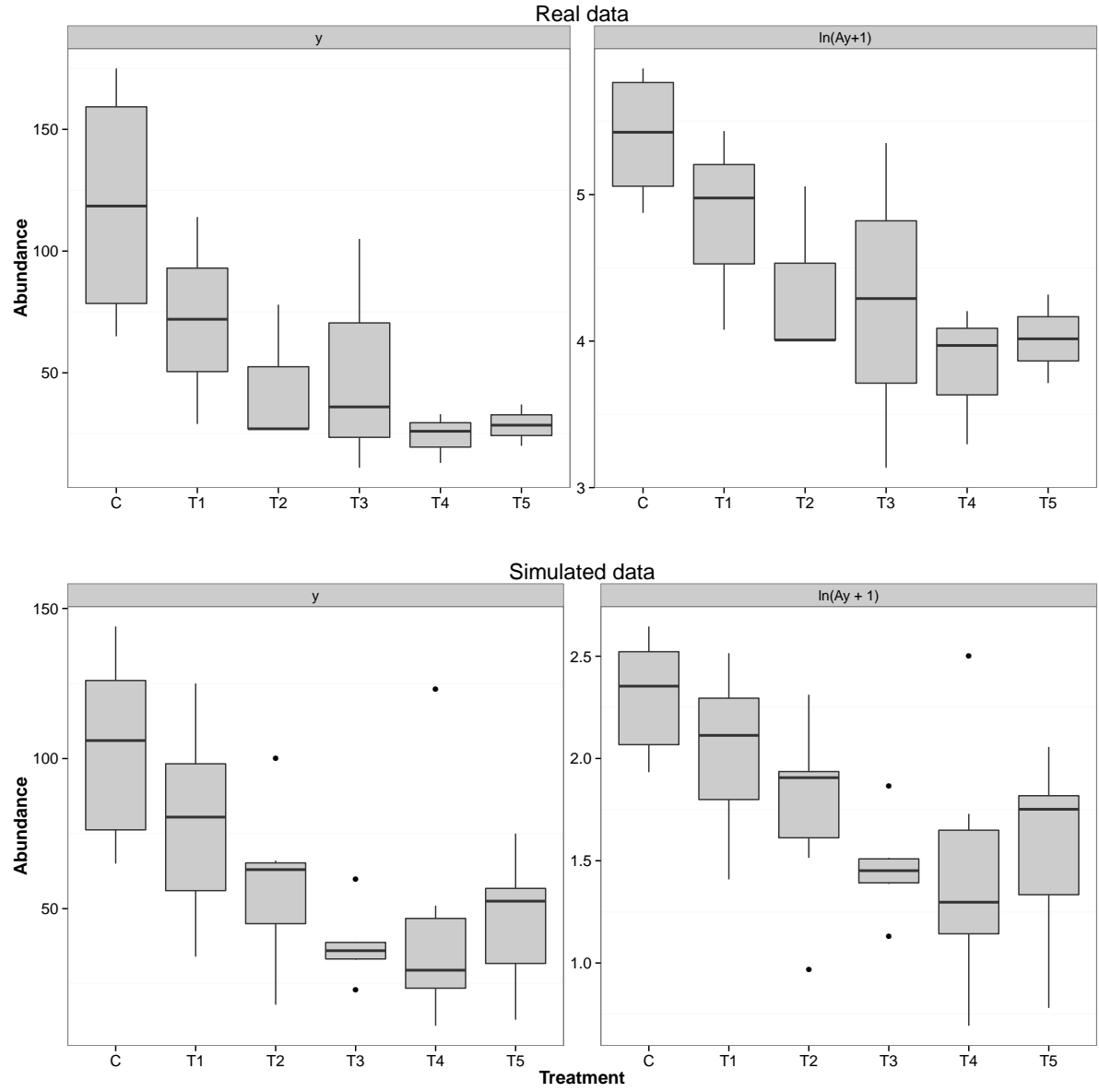


Figure 1: Real data from Brock et al... (top) and one realisation of simulated data (below,  $N = 6$ ,  $\mu_c = 100$ ). Left panel show raw counts, right panel  $\log(A \cdot x + 1)$  transformed counts.

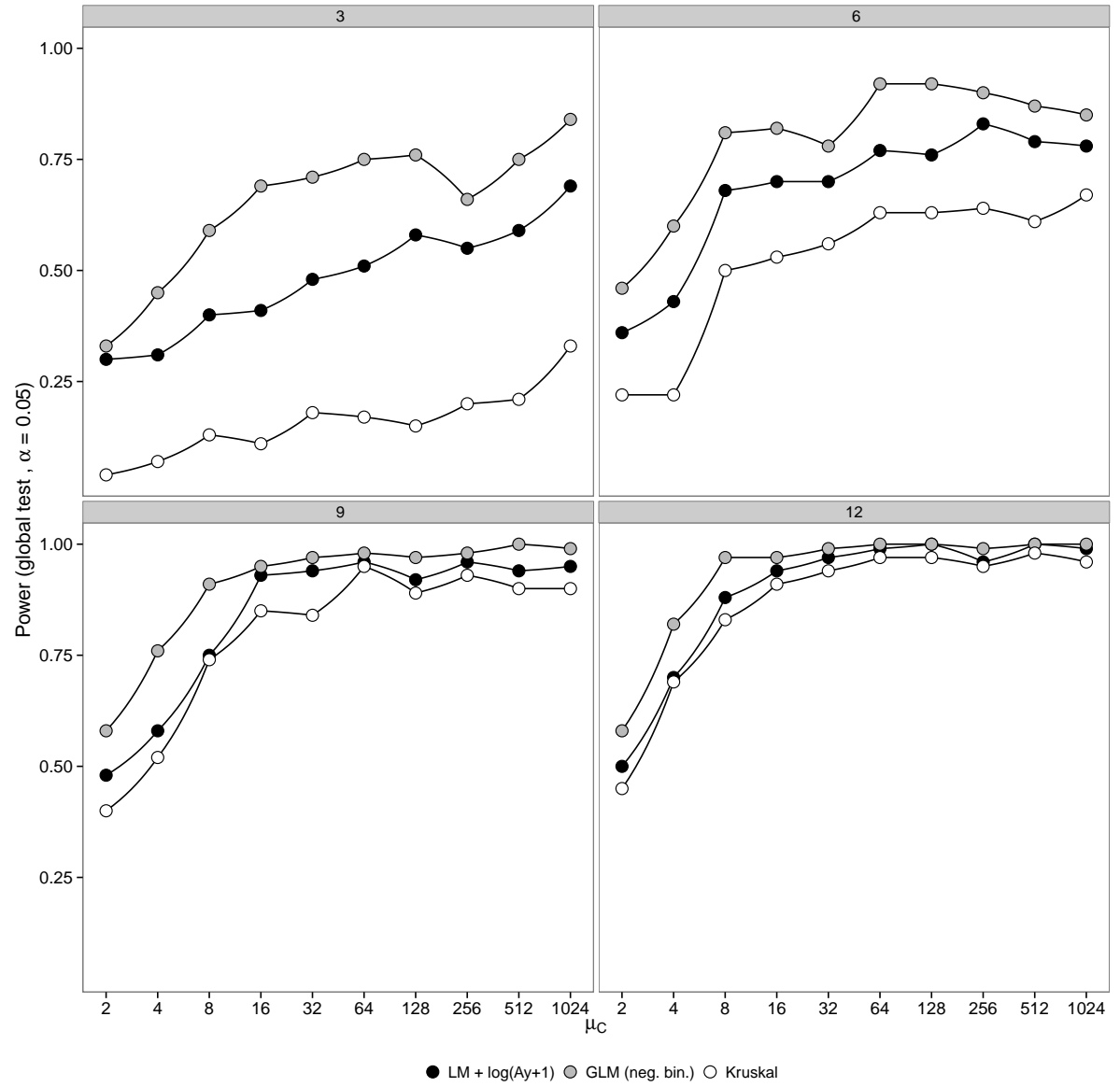


Figure 2: Power.

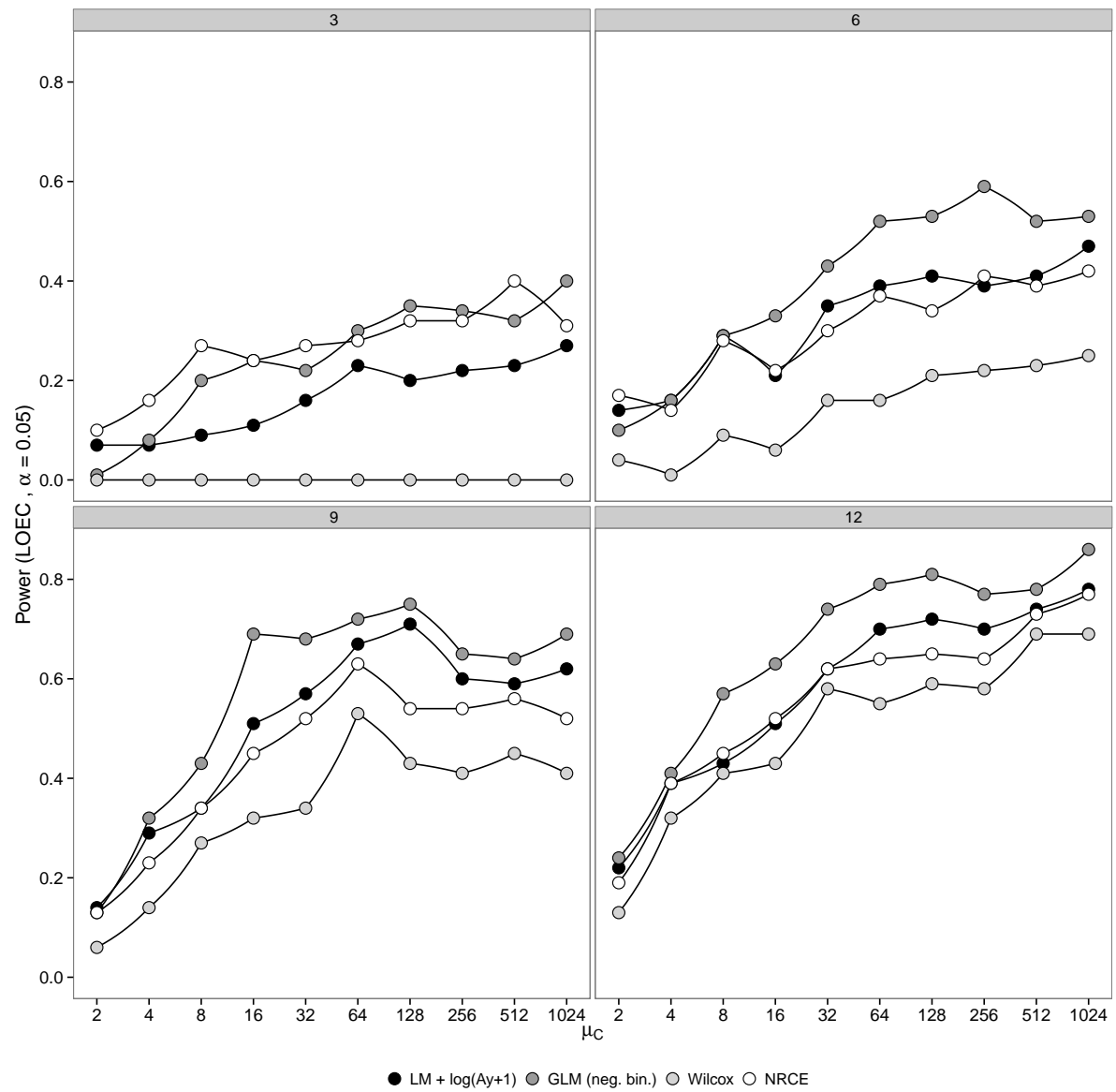


Figure 3: Power to detect LOEC.