eDNA and qPCR survey data

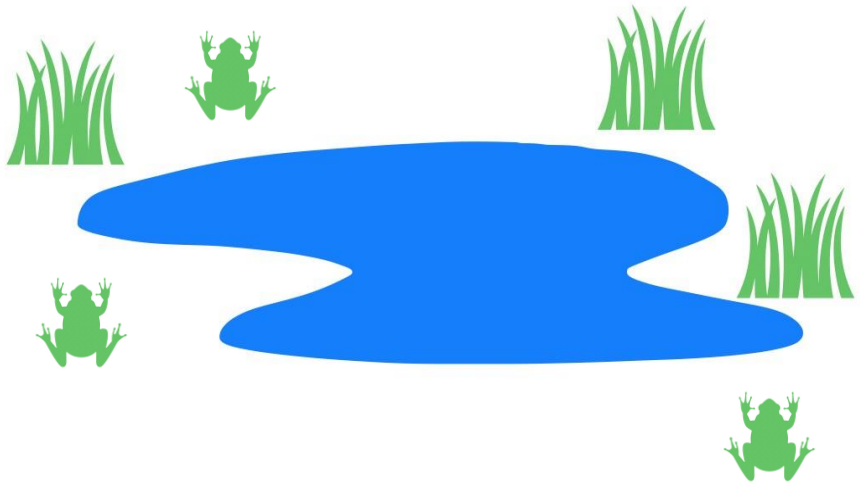
Milly Jones

mlj23@kent.ac.uk

Eleni Matechou, Diana Cole, Alex Diana, Jim Griffin,

Sara Peixoto, Lori Lawson Handley, Andrew Buxton

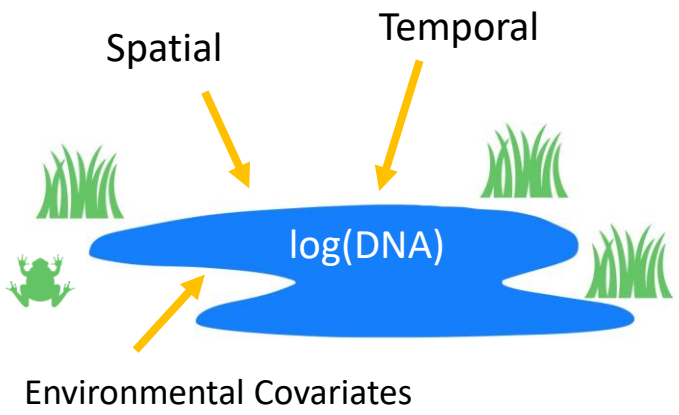
eDNA and qPCR survey data



Terminology:

eDNA: traces of DNA left by organisms in environmental samples such as water and soil

qPCR: Quantitative PCR is a technique used to amplify and quantify a target DNA molecule

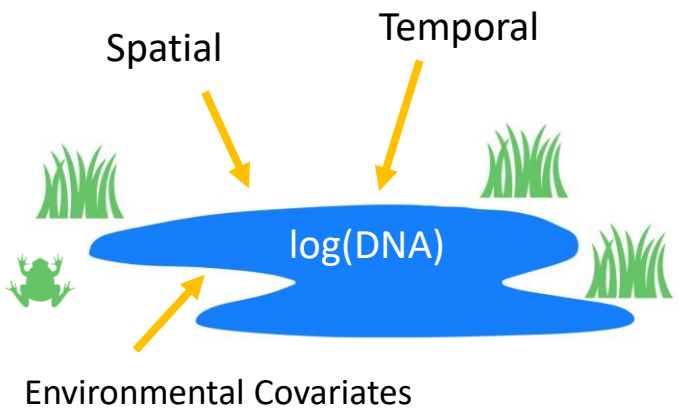


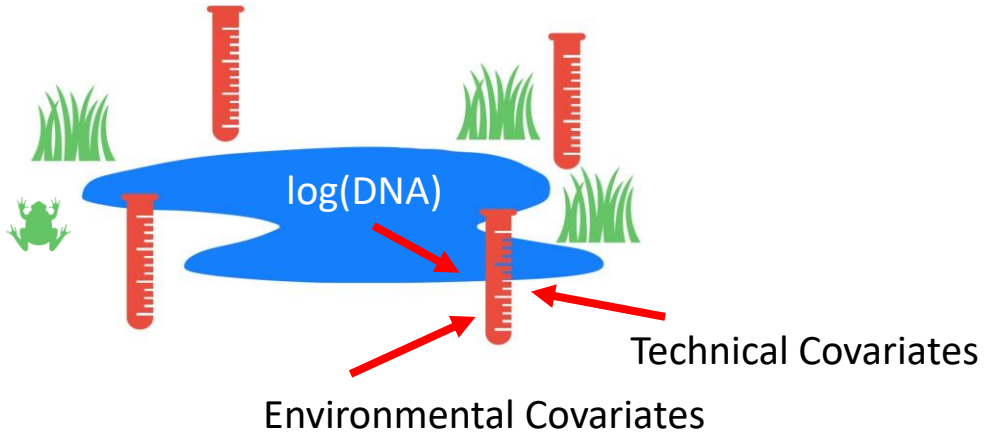
eDNA and qPCR survey data

<u>qPCR survey pipeline</u>

1. DNA availability:

   Across sites $i = 1, \dots, S$ and time-points $t = 1, \dots, T$.





eDNA and qPCR survey data

<u>qPCR survey pipeline</u>

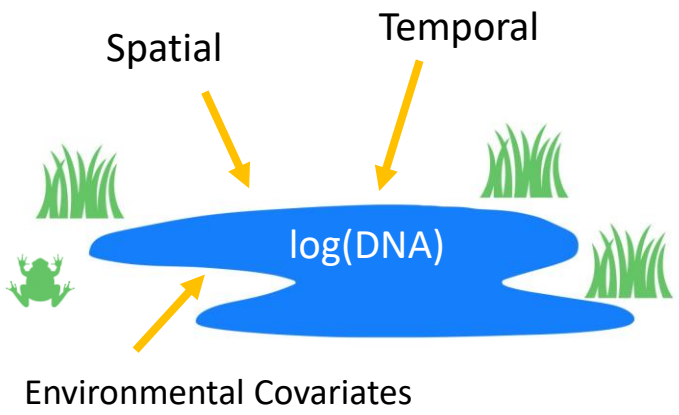
1. DNA availability:

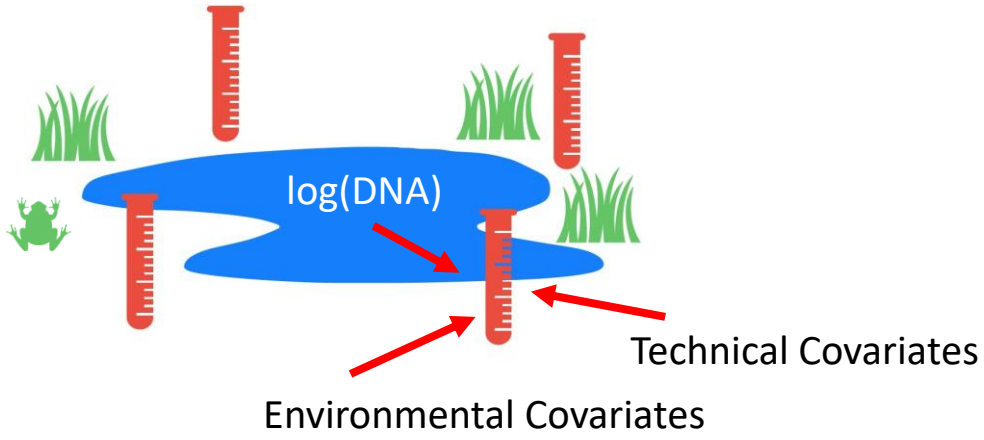
   Across sites $i = 1, \dots, S$ and time-points $t = 1, \dots, T$.

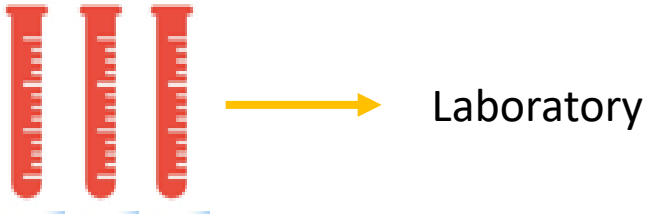
2. DNA collection:

   In samples $m = 1, \dots, M$.

eDNA and qPCR survey data







<u>qPCR survey pipeline</u>

1. DNA availability:

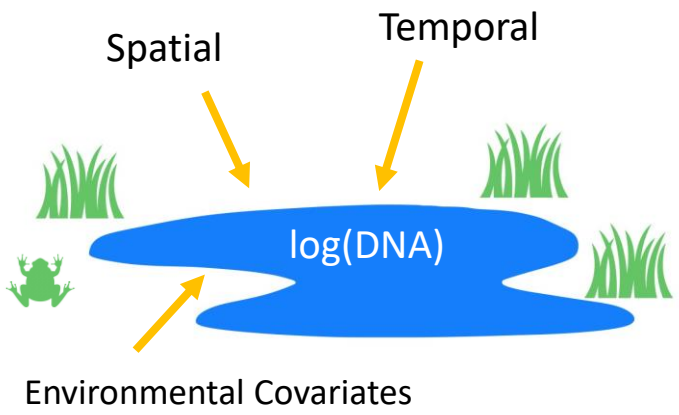
   Across sites $i = 1, \dots, S$ and time-points $t = 1, \dots, T$.

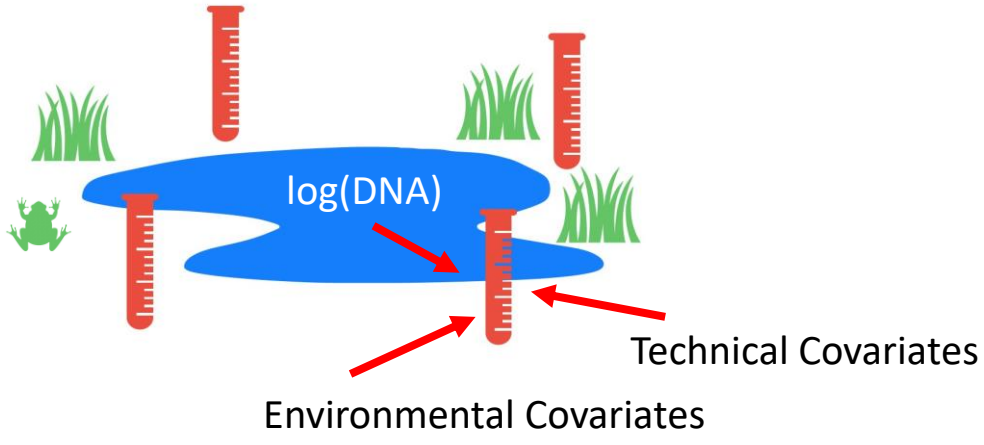
2. DNA collection:

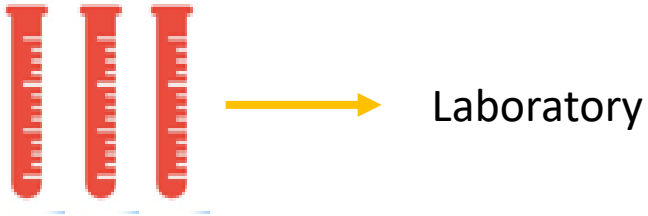
   In samples $m = 1, \dots, M$

3. DNA analysis:

   Model qPCR output on DNA concentration in the sample







eDNA and qPCR survey data

<u>qPCR survey pipeline</u>

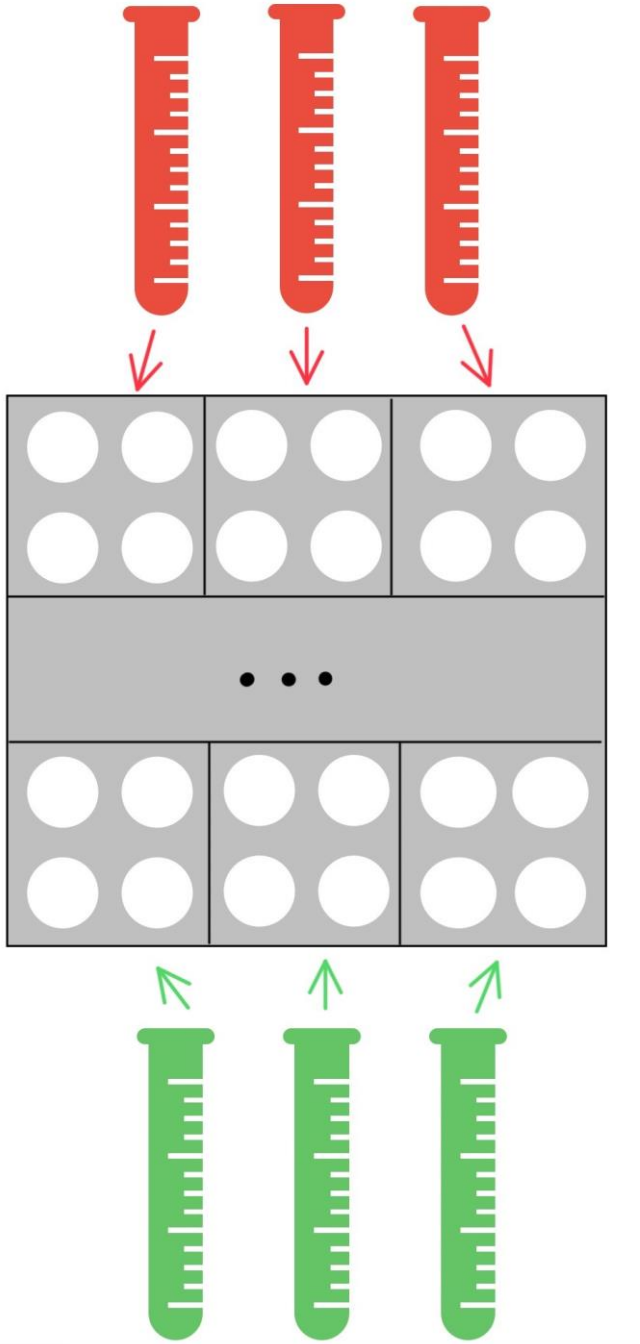
1. DNA availability:
   Across sites $i = 1, \dots, S$ and time-points $t = 1, \dots, T$.
2. DNA collection:
   In samples $m = 1, \dots, M$.
3. DNA analysis:
   Model qPCR output on DNA concentration in the sample

<u>Previous work:</u>

- [Espe et al. 2022] – artemis package
- [Shelton et al. 2022] – spatial model for Pacific Hake

<u>Our focus:</u>

- Modelling contamination and inhibition in the lab
- qPCR output heteroscedasticity



Collected samples

Unknown Concentrations

Plate p

K replicates

per sample

K* replicates

per standard

Standards

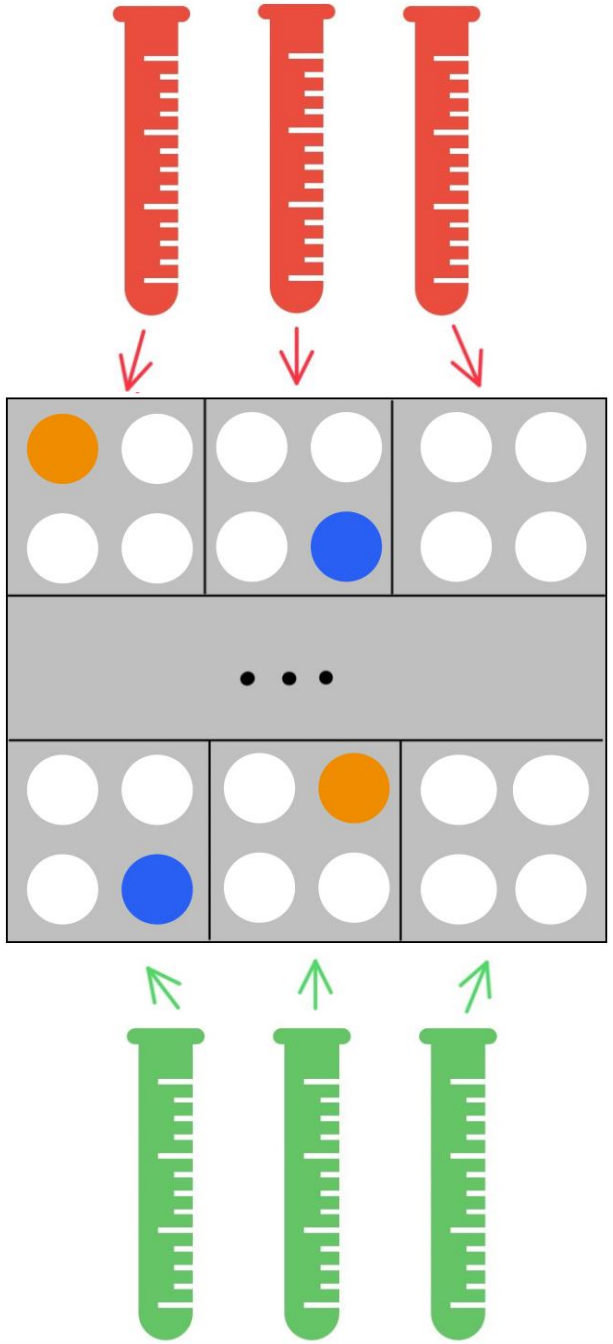
Known Concentrations

DNA analysis

Each plate $p = 1, \dots, P$ contains:

- Samples m each divided into K replicates
- **Standards** (solutions of known DNA concentration) divided into K^* replicates

DNA analysis



M samples

Unknown Concentrations

Plate p

Contaminated

Inhibited

Standards

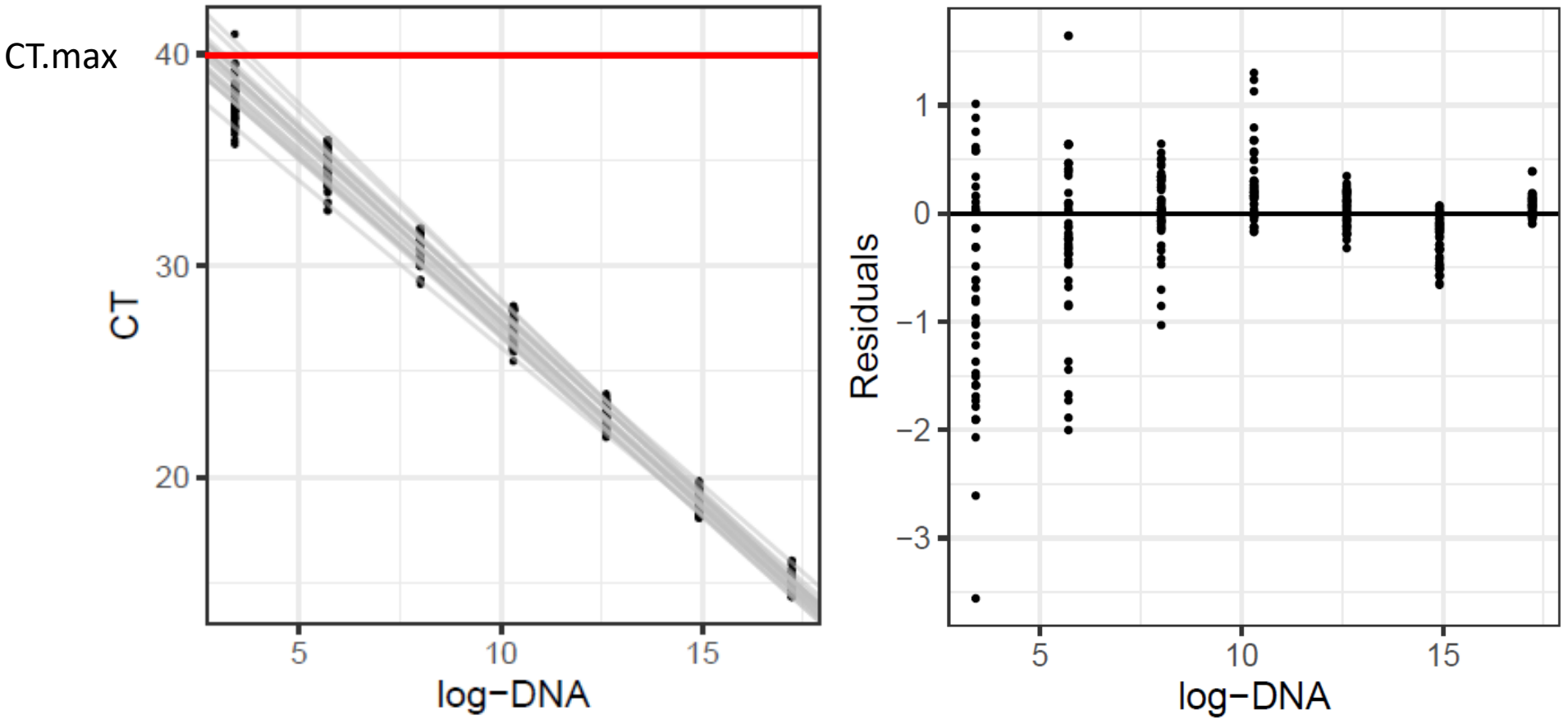
Known Concentrations

Contamination: presence of additional DNA in wells from lab through contaminated equipment or clothing

Inhibition: factors (substances or chemicals) interfering with the PCR process delaying DNA amplification

CT value – cycles to threshold, the (fractional) cycle number at which DNA is detected

DNA analysis



Heteroskedasticity: [Matz et al. 2013]

w_{imt} = DNA concentration

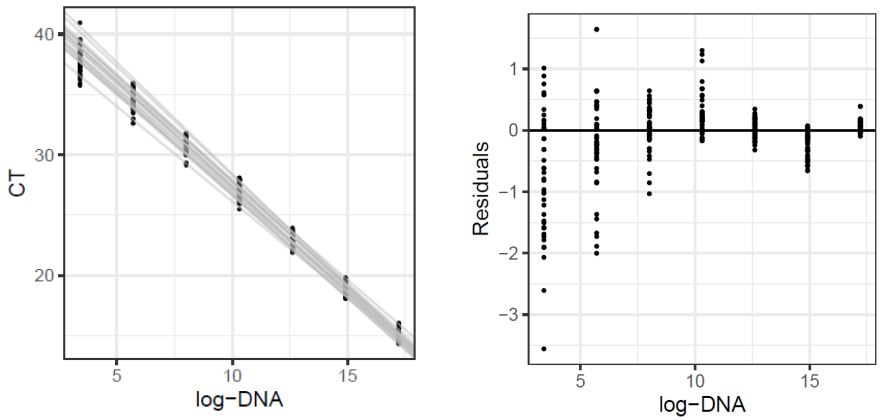
where i is the site, t is the time point,

m is the sample, and k is the replicate.

CT regression line $\mu_{imtk} = \alpha_p^1 + \alpha_p^2 \log(w_{imt}),$

CT heteroscedasticity $\sigma_y^2(w_{imt}) = \exp(a_1 + a_2 \log(w_{imt})),$

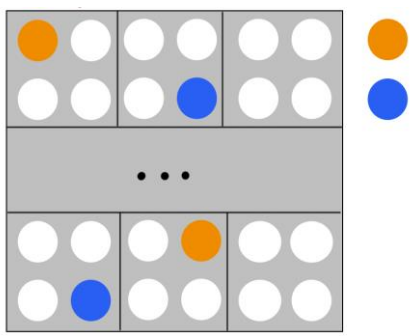
$$\alpha_p^1 \sim \mathrm{N}(\alpha_0^1, \sigma_\alpha^2)$$
$$\alpha_p^2 \sim \mathrm{N}(\alpha_0^2, \sigma_\alpha^2)$$



Status of replicate

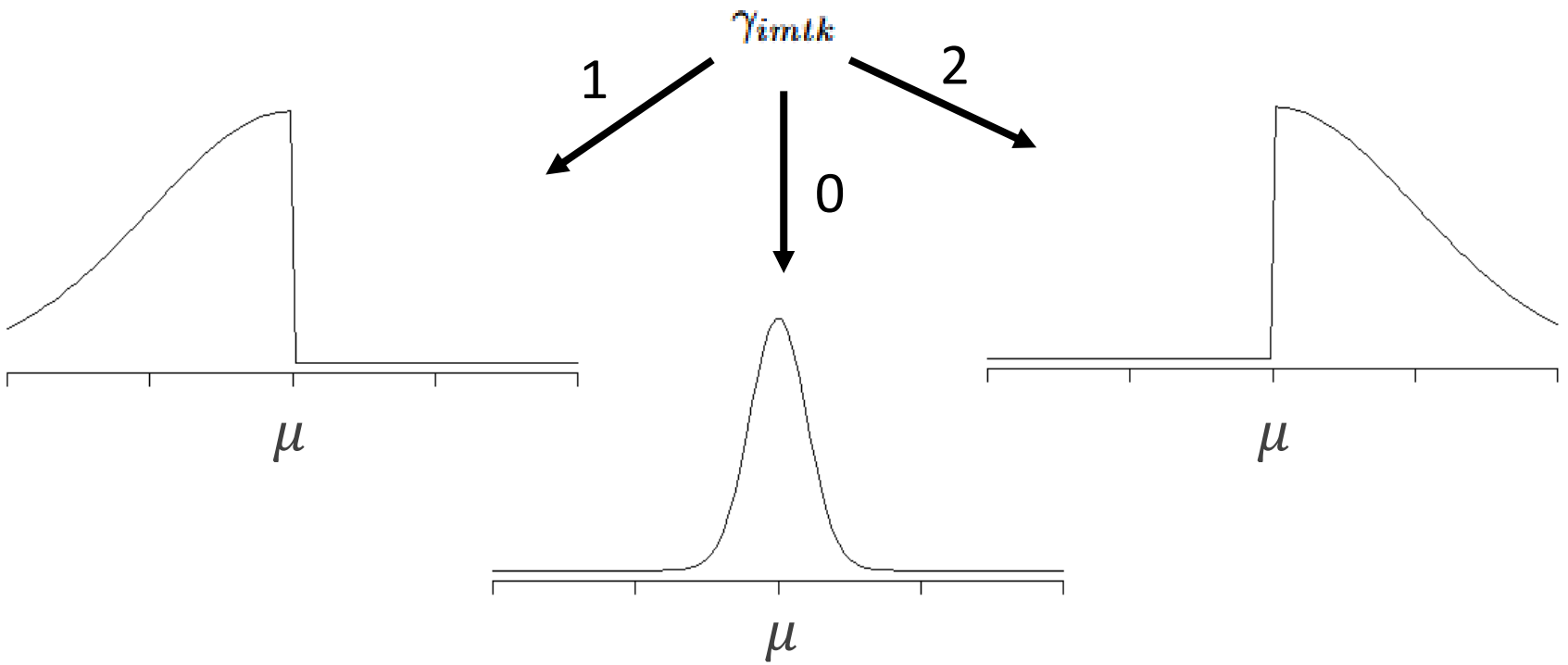
$$\gamma_{imtk} = \begin{cases} 1 & \text{contaminated with probability } p_c, \\ 2 & \text{inhibited with probability } p_h, \\ 0 & \text{neither with probability } 1 - p_c - p_h, \end{cases}$$

$$(p_c, 1 - p_c - p_h, p_h) \sim \mathrm{Dirichlet}(\pi_1, \pi_2, \pi_3)$$



Uncensored CT value

$$\tilde{C}_{imtk} \sim \begin{cases} N(\mu_{imtk}, \sigma_y^2(w_{imt})) & \text{if } \gamma_{imtk} = 0, \\ TN_{0,\mu_{imtk}}(\mu_{imtk}, \sigma_c^2) & \text{if } \gamma_{imtk} = 1, \\ TN_{\mu_{imtk},\infty}(\mu_{imtk}, \sigma_c^2) & \text{if } \gamma_{imtk} = 2, \end{cases}$$



Censored CT value

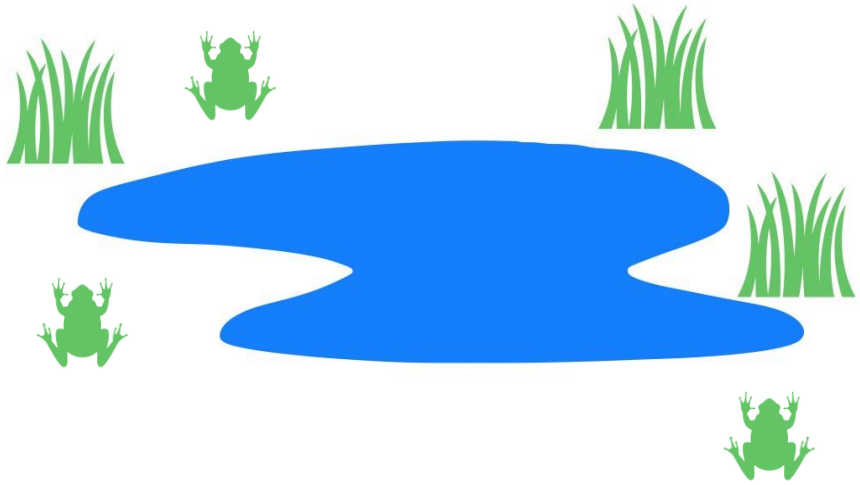
$$C_{imtk} = \begin{cases} \tilde{C}_{imtk} & \text{if } \tilde{C}_{imtk} < \text{CT.max}, \\ \text{NA} & \text{otherwise}, \end{cases}$$

Final Comments

Prior sensitivity analysis on $\alpha_{p,}^{1}$ $\alpha_{p,}^{2}$ a_1, a_2 show these are robust to prior specification.

Ignoring contamination, inhibition, and heteroscedasticity leads to increased bias and uncertainty.

Contamination/inhibition can occur at lab or DNA collection stages. This approach does not account for the collection stage.



References

Espe, M.B., Johnston, M., Blankenship, S.M., Dean, C.A., Bowen, M.D., Schultz, A., Schumer, G.: The artemis package for environmental DNA analysis in R. Environmental DNA 4(3), 523–532 (2022) https://doi.org/10.1002/edn3.277

Shelton, A.O., Ramon-Laca, A., Wells, A., Clemons, J., Chu, D., Feist, B.E., Kelly, R.P., Parker-Stetter, S.L., Thomas, R., Nichols, K.M., Park, L.: Environmental DNA provides quantitative estimates of pacific hake abundance and distribution in the open ocean. Proceedings of the Royal Society B 289(1971), 20212613 (2022) https://doi.org/10.1098/rspb.2021.2613

Matz, M.V., Wright, R.M., Scott, J.G.: No control genes required: Bayesian analysis of qRT-PCR data. PLOS ONE 8(8), 71448 (2013) https://doi.org/10.1371/journal.pone.0071448