Supplement 4 – Modelling fungal infection risk

By Tom R. Olk, Jens Wollebæk and Espen Lydersen

Tom R. Olk is Ph D. candidate at the University of Southeast-Norway

Jens Wollebæk (Ph D) is senior advisor at the Norwegian Agriculture Agency

Espen Lydersen (Ph D) is professor at the University of Southeast-Norway

*Correspondence to*: Tom R. Olk (Tom.Robin.Olk@usn.no)

# Data analysis

The probability of fungal infections in the hand-picked plots was investigated using binomially distributed generalized additive mixed models inferred using the mgcv-package (Wood, 2004; 2011) for additive modelling, and the nlme-package for mixed modelling (Pinheiro et al., 2017) in R (R Core Team, 2017). Treatment was used as a categorical fixed effect, and the locations of the plots were used as random effect. Implementing location as a random effect ensured treating measurements from the same plots as repeated measurements. Plots that were not hand-picked were excluded from this analysis. The presence of fungal infections was only registered at the end of the study in each plot. In addition, infected material was not removed from these plots. According to our definition of fungal infections as at least one egg, larvae or empty eggshell exhibiting visible signs of fungal infection, unpicked plots would have been defined as infected for the rest of the study as soon as one egg was infected. Consequently, observations of fungal infections in unpicked plots did not reflect the risk of contracting a fungal infection within one week, but were highly dependent on the presence of fungal infections during previous observations. Four candidate models were tested:

* M1: only including treatment as fixed effect, evaluating the overall effect of formalin treatment before the eyed stage on the risk of initial fungal infections in plots
* M2: including both treatment and the number of eggs remaining per plot as fixed effects. The number of eggs remaining are a measure of stocking density, which influences the risk of hyphal infections spreading from one egg to another.
* M3: including both treatment and the number of dead eggs in each plot as fixed effects. The number of dead eggs represents the amount of dead organic material that is easily infected, and provides a measure of general egg viability.
* M4: including both treatment and dpf as fixed effects. Dpf is a measurement of time, which represents the developmental stage of the eggs and differences in the abundance of Oomycetes over the course of the experiment.

Each model was controlled for overfitting by estimating the effective sample size (Neff) according to the procedure described in Zuur et al., (2009), and comparing it to a required sample size (NReq), which was calculated as 30 for each linear numerical fixed effect multiplied with the number of individual levels of all factorial fixed effects combined. Residual plots were produced to verify viable models. Approximate confidence intervals were estimated according to the method described by Agresti and Coull, (1998), assessed slightly liberal as the method produces narrower limits than the actual 95 % confidence interval. In addition to the intra class correlation (ICC) estimated through the models, ICC was calculated for fungal infections in treated and non-treated plots separately, using the location of the plot (n = 15) as grouping variable, with eight individual measures per location (k = 8) recorded on different dates. ICCs and confidence intervals were estimated using the ICC-package (Wolak et al., 2012) in R (R Core Team, 2017). The estimation of the confidence intervals was based on the exact confidence limit equation (Searle, 1971).

# Results

The intra class correlation (ICC) for fungal infections among formalin treated plots was estimated to 0.026 (95 % confidence interval: -0.051 – 0.208) based on a within group variance of 0.224, and an among group variance of 0.006. For non-treated plots, the ICC was estimated to 0.008 (95 % confidence interval: -0.062 – 0.178), based on a within group variance of 0.235, and an among group variance of 0.002.

Model M1 exhibited an ICC of 0.16, and a resulting Neff of 71. M1 only included one categorical fixed effect with two levels (Treatment).Thus, Nreq was estimated to be 30, meaning that the sample size requirements were satisfied for M1. The effect of treatment was significant (p < 0.001). The average probability for a hand-picked plot that was not treated with formalin to be infected by fungus was estimated to 0.672 (0.536 – 0.706), while the average probability of a hand-picked and formalin treated plot was estimated to 0.347 (27.1 – 43.9). M1 had an adjusted R2 of 0.072.

Model M2 exhibited an ICC of 0.16, and a resulting Neff of ≈ 71. The model contained one linear numerical fixed effect and one categorical fixed effect, thus Nreq = 60, and M2 satisfied the minimum sample size requirements. Of the fixed effects incorporated in M2, treatment was significant (p < 0.001), while the number of eggs remaining was not significant (p > 0.100). M2 had an adjusted R2 of 0.069.

Model M3 exhibited an ICC of 0.16, with a resulting Neff of ≈ 70. The model contained one linear, numeric fixed effect, and one factorial fixed effect with two levels, resulting in Nreq = 60. Consequently, Model M3 satisfied the minimum sample size criteria. The effect of treatment was significant in this model (p < 0.001). Formalin treated groups without dead eggs had an estimated probability to contract fungal infections of 0.31, while non-treated groups without dead eggs had an estimated probability to get infected of 0.60. The linear effect of the number of dead eggs was only near significant (p = 0.086) and positive. R2 of the model (M3) was estimated to 0.078.

Model M4 exhibited an ICC of 0.29, and a resulting Neff of ≈ 45. The model contained one non-linear numerical fixed effect and one categorical fixed effect with two levels, thus Nreq > 60. Consequently, M4 was over fitted, and did not produce reliable results. The effects of Treatment and dpf were both statistically significant at a level of p < 0.001. M4 exhibited an adjusted R2 of 0.277.

# Discussion

The risk of initial fungal infection in the plots was mainly governed by generic factors, as the ICC was low. Similarities in measurements from the same plot are caused by differences in susceptibility to fungal infections between families, the location and flow pattern within the compartment, the distance of the plot to not hand-picked plots, fungal infections that were not removed by hand-picking, and other similarities. The risk of fungal infections was not severely influenced by any of these factors, meaning that hand-picking is an effective way to remove fungal infections locally. In addition, the infection pressure appears to be similar in all hand-picked plots within one compartment. Consequently, placing all hand-picked plots and the different families in bulk did not affect the results of this study. However, placing hand-picked plots and not hand-picked plots in the same compartment likely raised the infection pressure by spores of Oomycetes.

Formalin was effective at reducing the risk of fungal infections by half in this study (M1; Fig. S1). This is consistent with results published by Bailey and Jeffrey, (1989), who investigated the effect of formalin treatments administered three times weekly over two weeks at concentrations of 250 mg L-1 for 60 minutes on rainbow trout eggs. The extent of the treatment used by Bailey and Jeffrey, (1989) is fairly similar to the treatment used in our study. Stocking density, measured as the number of eggs remaining per plot, did not significantly influence the risk of fungal infection in our study (M2; Fig. S2). Stocking density has previously been outlined as an important factor influencing the occurrence of fungal infections in fish eggs (Post, 1987). However, low stocking densities, and a narrow range of different stocking densities caused by the similar number of eggs in each plot, probably hampered the detection of such an effect in our study. In addition, *Saprolegnia spp.* and probably other oomycetes were found to primarily infect dead eggs by spores, while life eggs were usually infected by hyphae from adjacent infected material (Smith et al., 1985, Kitancharoen and Hatai, 1996, Thoen et al., 2011). Consequently, the amount of eggs in a plot *per se* may not be closely associated to the risk of fungal infections. Contrary, the number of dead eggs in each plot would influence the risk of fungal infections. This hypothesis could not be entirely confirmed in this study (M3; Fig. S3). The number of dead eggs in each plot was positively associated to the risk of fungal infections, but only approaching statistical significance. A larger number of plots would have been necessary to confirm this effect at a significance level of α = 0.05. There were clearly other, more important factors determining the risk of fungal infections in this study, as M1, M2, and M3 all exhibited R2 values of 0.069 – 0.079, meaning that the models explained less than 10 % of the variation present in the data.

Model M4 did not produce reliable results due to overfitting. However, the higher R2-value of 0.277 and the significant non-linear effect of dpf indicated, that either developmental stage or other effects related to time influenced the probability of initial fungal infection of a plot. The model predicted higher probabilities of fungal infections at the onset of the eyed stage (first sampling) in the plots that were not treated with formalin and around hatch regardless of treatments (Fig. S4). The higher initial mortalities in untreated plots were likely a result of the lack of treatment for fungal infections, which allows the infections to disperse to additional plots. This prophylactic effect of the formalin treatment, however, appeared to cease relatively quickly, as confidence intervals for the estimate of the probability of fungal infections began to overlap at ca. 49 dpf. At the time of hatch, more organic debris was present in all plots, as empty egg shells had accumulated for one week. Infected eggs entangled in hyphae that were associated with empty egg shells were frequently observed during hand-picking at this stage. Consequently, the results of Model M4 were consistent with the observations made in this study. However, the statistical significance of the effect of dpf on the probability of initial fungal infections in the plots may also be a result of overfitting the model to random variation within the dataset. This effect has previously been demonstrated even for statistical techniques that are robust to overfitting, and causes models to appear adequate at predicting observations from the dataset that the model is constructed upon (Subramanian and Simon, 2013). Contrary, such models cannot be generalized to other data sets and do not adequately describe general trends. Further studies using a larger number of plots for each treatment would be necessary to identify general trends of the probability of initial fungal infection for each plot. Investigating the risk and extent of fungal infections over time could yield valuable insight, as chemical treatments could be administered specifically at times of higher risks of fungal infections. Adjusting the treatment concentration, duration or frequency to the risk of fungal infection would reduce the overall use of chemicals and cost of the treatment, as well as it may increase the number of eggs surviving until hatch.

# References

Agresti, A. and Coull, B. A.: Approximate is Better than "Exact" for Interval Estimation of Binomial Proportions, Am. Stat.*,* 52, 119 – 126, doi: 10.1080/00031305.1998.10480550, 1998.

Bailey, T. A. and Jeffrey, S. M.: Evaluation of 215 Candidate Fungicides for Use in Fish Culture, in Investigations in Fish Control, U. S. Fish and Wildlife Service, LaCrosse, WI, USA, URL: <https://pubs.usgs.gov/ifc/099/report.pdf>, 1989.

Kitancharoen, K., and Hatai, K.: Experimental Infection of *Saprolegnia* spp. in Rainbow Trout Eggs, Fish Pathol., 31, 49 – 50, doi: [10.3147/jsfp.31.49](https://doi.org/10.3147/jsfp.31.49), 1996.
Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., and R Core Team: Nlme: Linear and Nonlinear Mixed Effects Models (version 3.1-131), 2017.

Post, G.: *Textbook of Fish Health*, T. F. H. Publications, Neptune City, NJ, 1987.

R Core Team: R: A Language and Environment for Statistical Computing (version 3.3.3), R Foundation for Statistical Computing Vienna, AT, 2017.

Searle, S. R.: Linear Models, Wiley, New York, doi: 10.1002/9781118491782, 1971.

Smith, S. N., Armstrong, R. A., Springate, J., and Barker, G.: Infection and Colonization of Trout Eggs by Saprolegniaceae, Trans. Br. Mycol. Soc., 85, 719 – 764, doi: [10.1016/S0007-1536(85)80268-0](https://doi.org/10.1016/S0007-1536%2885%2980268-0), 1985.

Subramanian, J., and Simon, R.: Overfitting in Prediction Models - Is it a Problem only in High Dimensions?, Contemp. Clin. Trials, 36, 636 – 641, doi: 10.1016/j.cct.2013.06.011, 2013.
Thoen, E., Evensen, Ø., and Skaar, I.: Pathogenicity of *Saprolegnia* spp. to Atlantic Salmon, *Salmo salar* L., Eggs, J. Fish Dis., 34, 601 – 608, doi: 10.1111/j.1365-2761.2011.01273.x, 2011.

Wolak, M. E., Fairbairn, D. J., and Paulsen, Y. R.: Guidelines for Estimating Repeatability, Methods Ecol. and Evol., 3, 129 – 137, doi: 10.1111/j.2041-210X.2011.00125.x, 2012.

Wood, S. N.: Stable and Efficient Multiple Smoothing Parameter Estimation for Generalized Additive Models, J. Am. Stat. Assoc., 99, 673 – 686, doi: [10.1198/016214504000000980](https://doi.org/10.1198/016214504000000980), 2004.

Wood, S. N.: Fast Stable Restricted Maximum Likelihood and Marginal Likelihood Estimation of Semiparametric Generalized Linear Models, J. R. Stat. Soc. Series B Stat. Methodol., 73(1), 3 – 36, doi: [10.1111/j.1467-9868.2010.00749.x](https://doi.org/10.1111/j.1467-9868.2010.00749.x), 2011.

Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A., and Smith, G. M.: Mixed Effects Modelling for Nested Data, in: Mixed Effects Models and Extensions in Ecology with R, Springer, New York, NY, USA, 101 – 142, doi: 10.1007/978-0-387-87458-6\_5, 2009.