# Regression models parameters

NB! Предлагаю убрать слово “Subset” и заменить его на “Region”. Пока оставляю так как было

In all models outcome variables were supposed as binomialy disributed (). Logit was used as a linck function in all models. In all cases full models including all terms and their inteactions were constructed. After the full models were constructed they were simplified accordingly to backward selection protocol (Zuur et al., 2009).

Model1:

Model2:

Model3:

Model4:

Model5:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Terms | Estimate | SE | z-statistic | p-value |
| NA | NA | NA | NA | NA |
| (Intercept) | -2.7 | 0.15 | -18.23 | < 0.001 |
| Ptros | 4.7 | 0.28 | 16.58 | < 0.001 |
| Subset(BL) | 0.3 | 0.29 | 0.92 | 0.359 |
| Subset(BH) | 2.3 | 0.26 | 8.85 | < 0.001 |
| Ptros:Subset(BL) | -0.6 | 0.50 | -1.20 | 0.232 |
| Ptros:Subset(BH) | -1.8 | 0.54 | -3.24 | 0.001 |
| NA | NA | NA | NA | NA |
| (Intercept) | -4.3 | 0.41 | -10.41 | < 0.001 |
| Ptros | 4.8 | 0.87 | 5.58 | < 0.001 |
| Subset(BL) | 0.8 | 0.92 | 0.83 | 0.407 |
| Subset(BH) | 4.3 | 0.95 | 4.57 | < 0.001 |
| Species(*M.trossulus*) | 4.2 | 0.45 | 9.25 | < 0.001 |
| Ptros:Subset(BL) | -2.4 | 1.80 | -1.36 | 0.175 |
| Ptros:Subset(BH) | -3.5 | 1.85 | -1.89 | 0.059 |
| Ptros:Species(*M.trossulus*) | -2.5 | 0.84 | -3.00 | 0.003 |
| Subset(BL):Species(*M.trossulus*) | -1.3 | 0.90 | -1.41 | 0.158 |
| Subset(BH):Species(*M.trossulus*) | -4.9 | 0.76 | -6.40 | < 0.001 |
| Ptros:Subset(BL):Species(*M.trossulus*) | 3.3 | 1.73 | 1.93 | 0.054 |
| Ptros:Subset(BH):Species(*M.trossulus*) | 5.7 | 1.49 | 3.83 | < 0.001 |
| sd\_(Intercept).pop | 0.9 | NA | NA | NA |
| NA | NA | NA | NA | NA |
| (Intercept) | 2.6 | 0.16 | 16.58 | < 0.001 |
| Ptros | -1.9 | 0.28 | -6.82 | < 0.001 |
| Subset(BL) | -0.7 | 0.31 | -2.30 | 0.021 |
| Subset(BH) | -3.1 | 0.26 | -11.86 | < 0.001 |
| Ptros:Subset(BL) | 1.4 | 0.52 | 2.74 | 0.006 |
| Ptros:Subset(BH) | 3.8 | 0.50 | 7.49 | < 0.001 |
| NA | NA | NA | NA | NA |
| (Intercept) | 3.8 | 0.24 | 15.81 | < 0.001 |
| Morph(T) | -3.9 | 0.33 | -11.70 | < 0.001 |
| Ptros | -5.3 | 0.45 | -11.65 | < 0.001 |
| Subset(BL) | -0.5 | 0.43 | -1.18 | 0.239 |
| Subset(BH) | -1.7 | 0.41 | -4.21 | < 0.001 |
| Morph(T):Ptros | 8.3 | 0.57 | 14.58 | < 0.001 |
| Morph(T):Subset(BL) | 0.5 | 0.35 | 1.45 | 0.146 |
| Morph(T):Subset(BH) | -0.8 | 0.28 | -2.87 | 0.004 |
| Ptros:Subset(BL) | 1.0 | 0.75 | 1.29 | 0.196 |
| Ptros:Subset(BH) | 2.5 | 0.75 | 3.37 | 0.001 |
| sd\_(Intercept).pop | 0.1 | NA | NA | NA |
| NA | NA | NA | NA | NA |
| (Intercept) | -2.5 | 0.12 | -20.66 | < 0.001 |
| PT | 5.2 | 0.25 | 20.90 | < 0.001 |
| Subset(BL) | 0.4 | 0.13 | 3.18 | 0.001 |
| Subset(BH) | -1.4 | 0.14 | -9.44 | < 0.001 |

In all models Ptros - proportion of *M.trossulus* in population; Subset - sampling area (*W*, *BL*, *BH*); PT - proportion of T-morphotype in population; Species - mussel genotype (*M.edulis* or *M.trossulus*). *M.edulis* and *W* where used as basic levels for categorical predictors.

## Regression models visualisation

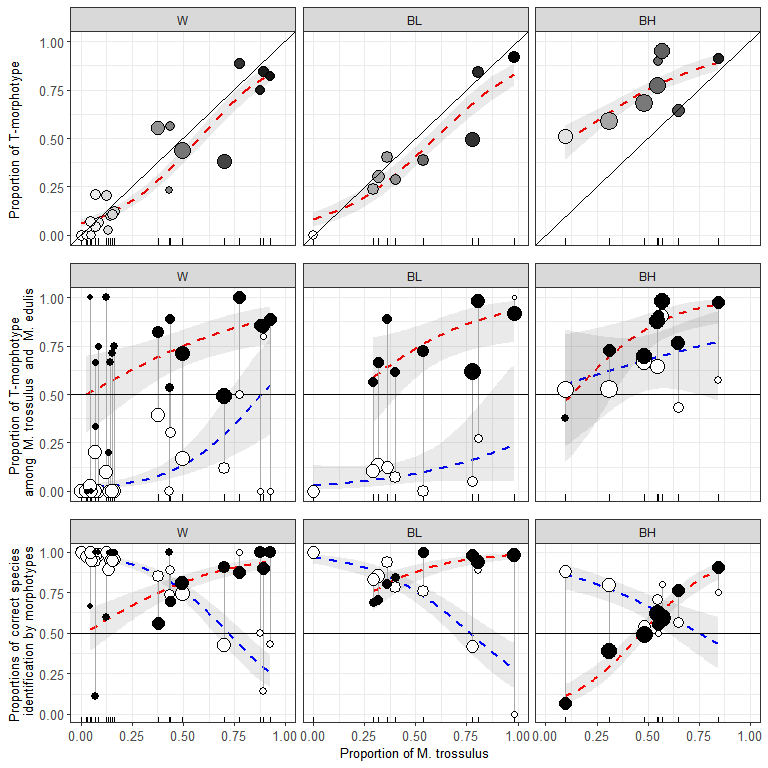


Figure ++. Visualisation of regression models. Initial data are presented as proportions of positive outcome in particular populations. Size of points is proportional to number of mussels in the particular sample. (A) Model1: the fill intensity is proportional to Ptros (**А то этого не видно из значений абсциссы?…**). (B) Model2: filled points - *M.trossulus* with T-morphotype; emty points - *M.edulis* with T-morphotype. (C) Model3: filled points - *M.trossulus* with T-morphotype; empty points - *M.edulis* with E-morphotype.

## Strategy for finding of calibration samples

*Важно!*  
Внимательно посмотрев на результаты подбора моделей (см. таблицу), я понял что никакого статистического основания для объединения W и BL у нас нет. Они везде значимо расходятся (везде BL значимо отличается от базового уровня). Поэтому единственно честное решение рассматривать Subset по-отдельности. Но для поиска стратегии отбора калибровочных выборок можно пойти другим путем. Нам ведь для этого не важно в каком регионе модель работает хорошо, а в каком работает хуже. Нам важно понять какие выборки надо делать, чтобы теоретическая модель была максимально похожа на эмпирическую регрессию. Почему бы в этой ситуации не рассматривать все выборки вместе, без разделения на Subset. То есть мы берем все возможные пары выборок, строим для них бублик и вычисляем на его основе два предсказания: одно для связи Ptros c P\_T (это теоретический аналог для Model5), а второе для связи Pcorrect c Ptros и Morph (теоретический аналог Model4).Ведь оба калькулятора ничего не знают про разность акваторий. Да и мы, стрго говоря, ввели это деление очень условно. Кроме того, поиск стратегии отбора калибровочных выборок - задача чисто техническая, надо ответить на вопрос какой генетической структурой должны обладать две калибровочные выборки, чтобы калькуляторы давали максимально похожий ответ. В такой ситуации вообще не надо делить на акватории. Короче, для этого анализа я ЗА либо полностью объединенные Subset, либо полностью разделенные Subset.

Пока привожу результаты так, как мне кажется лучше.

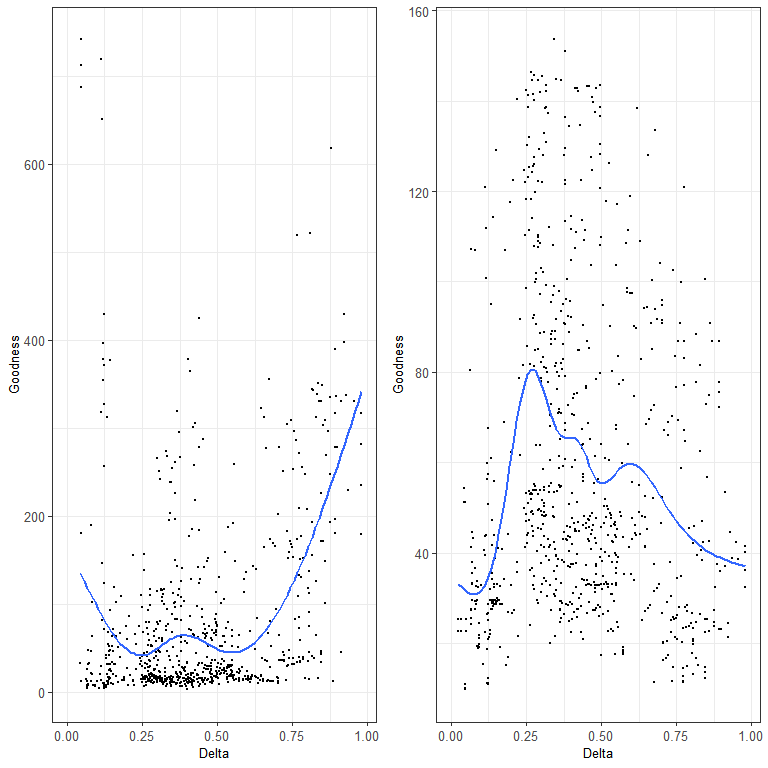


Figure +. Correspondence between regression and theoretical models. Each point corresponds to one of the possible pairs of populations from modelling data set. OX axis represents the differencу in genetic structre for each pair of populations. OY axis represents correspondence between prediction of regression model and theoretical model. Lines represent loess-smoother. (A ) Model 5 describing the dependence of proportion of M.trossulus (Ptros) on proportion of T-morphotype (P\_T) ; (B) Model 4 describing the dependence of probability of correct species identification (Pcorrect) on proportion of M.trossulus (Ptros) and morphotype (Morph).

Отсюда мы выносим два вываода. 1. Для построения теоретической модели для прогноза Ptros по частоте T-морфотипа лучше взять в регионе интереса две калибровочные выборки максимально различающиеся по своей генетической структуре. 2. Для построения теоретической модели, которая позволит оценить вероятность корректного определения вида конкретной мидии по морфотипу необходимо взять калибровочные выборки, в которых доля M.trossulus была бы длизка к 50%.

А во далее я предлагаю показать картинку, на которой будет следующий результат.

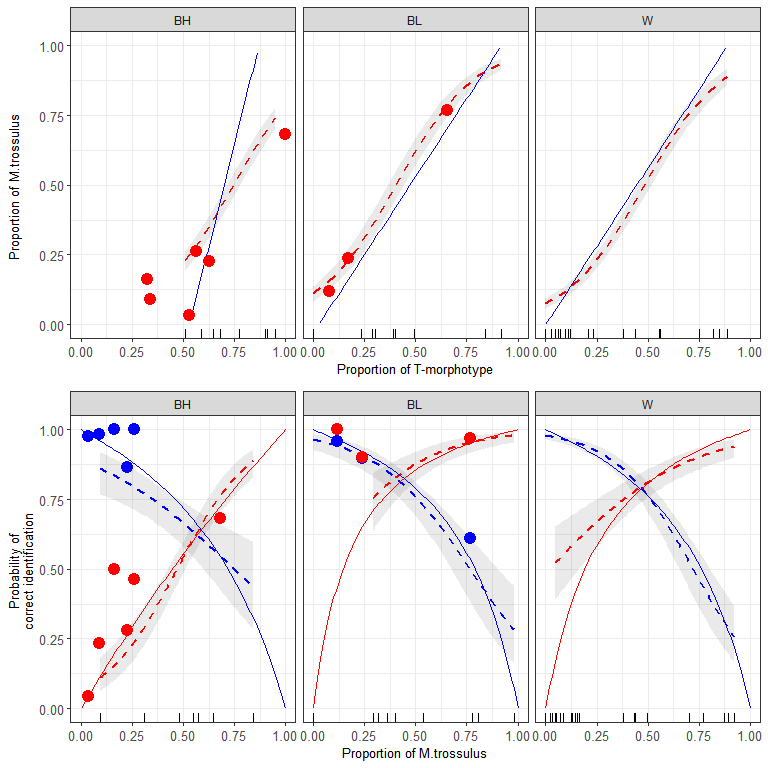


Figure +. Visualisation of regression models and theoretical models. Points represent the testing data set.