Юля! Вот посмотри на результаты. Я брал в качестве критерия для отбора проб в тестинг датасет те, у кого Ptros были наиболее близки к 0.2, 0.4, 0.6, 0.8 квантилям распределения Ptros в соответствующем Subset. Сразу скажу, что в качестве границ для отбора не стоит брать Ptros = 0.5 или около того, та как мы выьиваем серединку и от этого все модели сильно едут.

# Regression models parameters

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 | 0.17 | -16.10 | < 0.001 |
| Ptros | 4.6 | 0.30 | 15.45 | < 0.001 |
| Subset(BL) | -0.1 | 0.32 | -0.46 | 0.643 |
| Subset(BH) | 2.3 | 0.21 | 11.00 | < 0.001 |
| Ptros:Subset(BL) | 0.6 | 0.56 | 1.02 | 0.307 |
| Ptros:Subset(BH) | -2.1 | 0.46 | -4.46 | < 0.001 |
| NA | NA | NA | NA | NA |
| (Intercept) | -3.8 | 0.36 | -10.42 | < 0.001 |
| Ptros | 3.4 | 0.57 | 5.96 | < 0.001 |
| Subset(BL) | -0.3 | 0.51 | -0.49 | 0.625 |
| Subset(BH) | 3.1 | 0.42 | 7.31 | < 0.001 |
| Species(*M.trossulus*) | 3.0 | 0.28 | 10.70 | < 0.001 |
| Subset(BL):Species(*M.trossulus*) | 0.7 | 0.46 | 1.43 | 0.152 |
| Subset(BH):Species(*M.trossulus*) | -2.2 | 0.34 | -6.26 | < 0.001 |
| sd\_(Intercept).pop | 0.8 | NA | NA | NA |
| NA | NA | NA | NA | NA |
| (Intercept) | 2.5 | 0.17 | 14.47 | < 0.001 |
| Ptros | -1.6 | 0.29 | -5.63 | < 0.001 |
| Subset(BL) | -0.2 | 0.34 | -0.65 | 0.517 |
| Subset(BH) | -2.5 | 0.21 | -11.63 | < 0.001 |
| Ptros:Subset(BL) | 1.3 | 0.57 | 2.20 | 0.028 |
| Ptros:Subset(BH) | 2.7 | 0.43 | 6.25 | < 0.001 |
| NA | NA | NA | NA | NA |
| (Intercept) | 3.6 | 0.25 | 14.30 | < 0.001 |
| Morph(T) | -4.3 | 0.35 | -12.46 | < 0.001 |
| Ptros | -4.7 | 0.42 | -11.23 | < 0.001 |
| Subset(BL) | 0.2 | 0.31 | 0.55 | 0.58 |
| Subset(BH) | -0.6 | 0.27 | -2.23 | 0.025 |
| Morph(T):Ptros | 9.0 | 0.56 | 16.28 | < 0.001 |
| Morph(T):Subset(BL) | 0.4 | 0.44 | 0.98 | 0.327 |
| Morph(T):Subset(BH) | -0.7 | 0.32 | -2.13 | 0.033 |
| sd\_(Intercept).pop | 0.3 | NA | NA | NA |
| NA | NA | NA | NA | NA |
| (Intercept) | -2.6 | 0.14 | -19.19 | < 0.001 |
| PT | 5.5 | 0.26 | 21.32 | < 0.001 |
| Subset(BL) | 0.2 | 0.16 | 1.08 | 0.278 |
| Subset(BH) | -1.5 | 0.14 | -10.45 | < 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

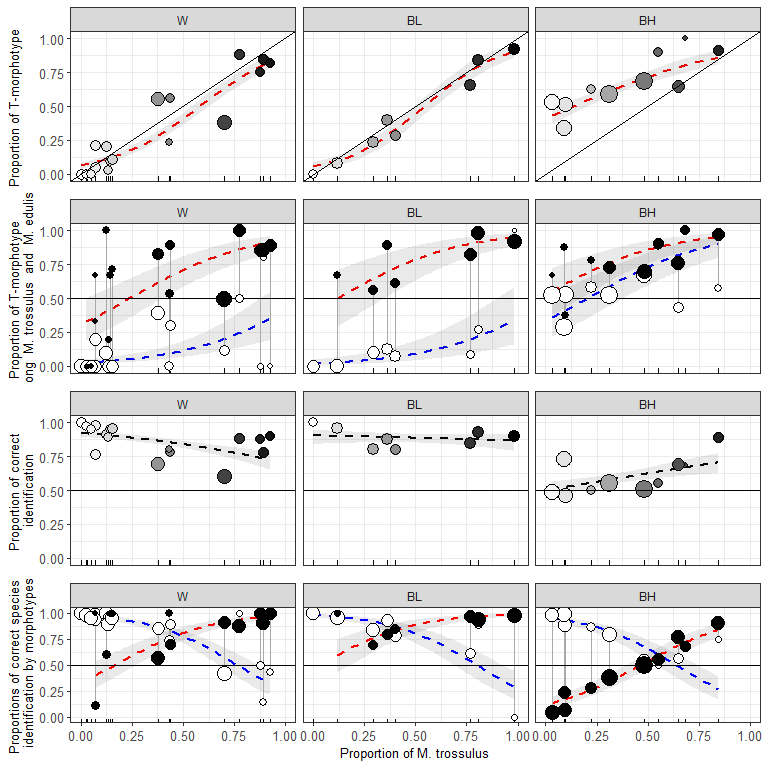


Figure ++. Visualisation of regression models. Initial data are presented as proportions of positive outcome in particular populations. Size of points is proportional to number of mussels in the particular sample. (A) Model1: the fill intensity is proportional to Ptros. (B) Model2: filled points - *M.trossulus* 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

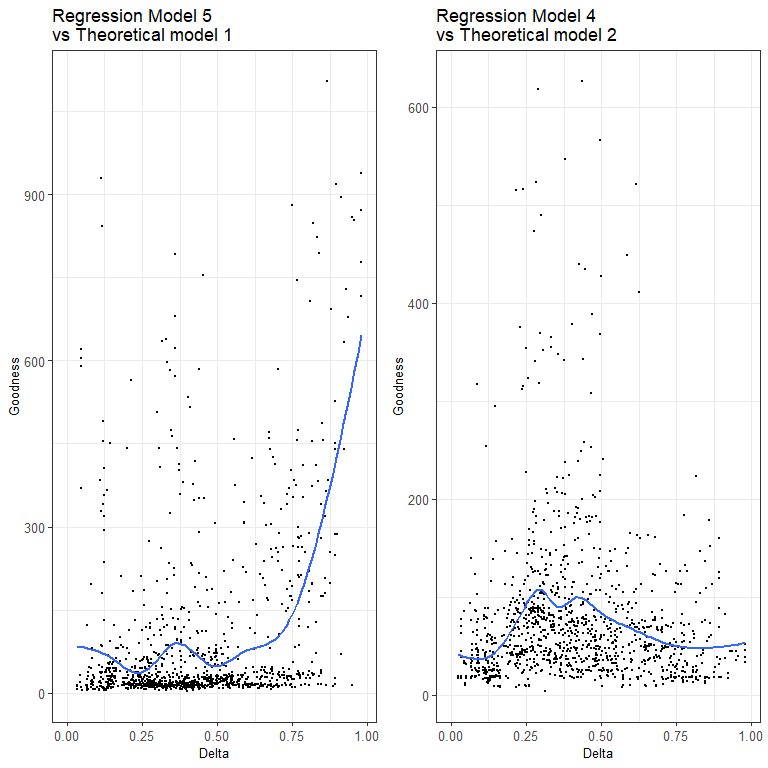


Figure +. Correspondence between regression and theoretical models. Each point corresponds to one of the possible pairs of populations from modelling data set. OX axis represents the differencу in genetic structre for each pair of populations. OY axis represents correspondence between prediction of regression 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).

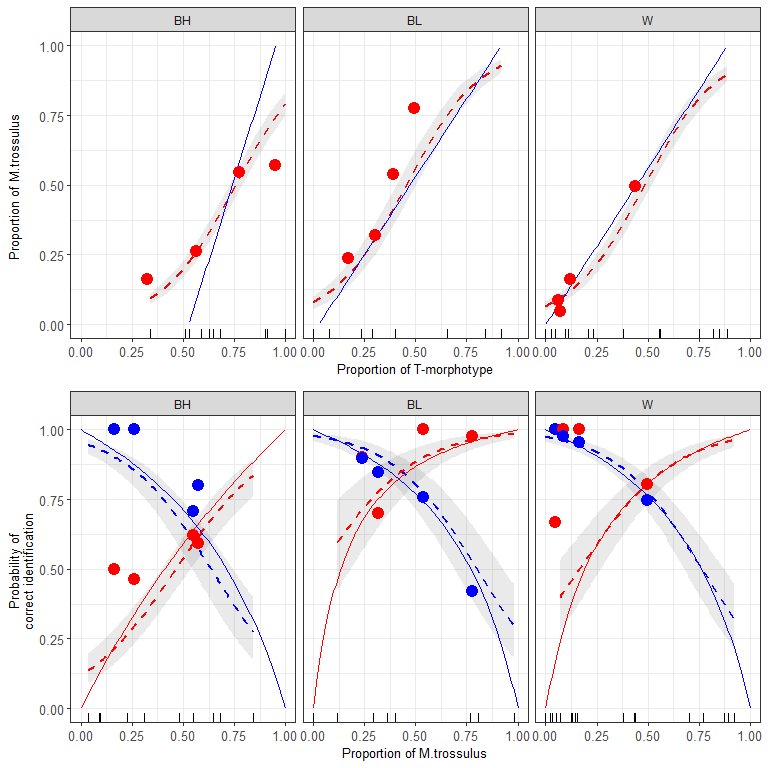


Figure +. Visualisation of regression models and theoretical models. Points represent the testing data set. На этом рисунке пунктирная линия - это эмпирическая регрессионная модель, построенная по modelling dataset. Сплошная линия - это пердсказания соответствующих теоретических моделей, которые построены на основе калибровочных выборок. Калибровочные выборки были отобраны на основе стратегии, описанной в предыдущем разделе. То есть для теоретической модели №1 - это максимально различные по генетическому составу выборки. Для теоретической моедли №2 это выборки имеющие Ptros близкое к 0.5.

Можно привести еще вот такую таблицу, в которой будут характеристики калибровочных выборок. Эти данные можно привести в виде врезки на графиках. Надо подумать.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **pop** | **Subset** | **N\_E** | **N\_T** | **P\_T\_ME** | **P\_T\_MT** |
| NA | NA | NA | NA | NA | NA |
| salnij | W | 3 | 36 | 0.00 | 0.82 |
| umba\_06 | W | 29 | 0 | 0.00 | NA |
| belok2 | BL | 1 | 49 | 0.02 | 0.90 |
| nm | BL | 37 | 0 | 0.00 | NA |
| dz\_lit | BH | 86 | 3 | 0.51 | 0.02 |
| tu\_old | BH | 7 | 38 | 0.09 | 0.82 |
| NA | NA | NA | NA | NA | NA |
| niva\_sl | W | 23 | 18 | 0.17 | 0.39 |
| umba\_pikut | W | 17 | 13 | 0.00 | 0.23 |
| fr | BL | 39 | 16 | 0.07 | 0.16 |
| nm\_last | BL | 32 | 18 | 0.08 | 0.32 |
| ivan\_buy | BH | 7 | 15 | 0.32 | 0.68 |
| vol | BH | 23 | 42 | 0.15 | 0.49 |

А вот про эту визуализацию, которая приведена ниже, я пока не понял что говорить и не уверен, что ее следует приводить.Здесь видно только, что предсказанная вероятнсоть определения обычно ниже, чем наблюдаемая, что, впрочем, видно и на предыдущей картинке.

