**Dafnia Magna (SPb)**

**Outliers:** '4s', '4b', '8b', '8s', '28s', '28b', "29s", "29b", '30b', '30s', '24b' (11)

**Test set:** '7s', '13s', '11s', '6b', '15b', '23s',' 19b', '18b', '20s' (9)

**Training set:** 30 samples

PLS results before removing the outliers:

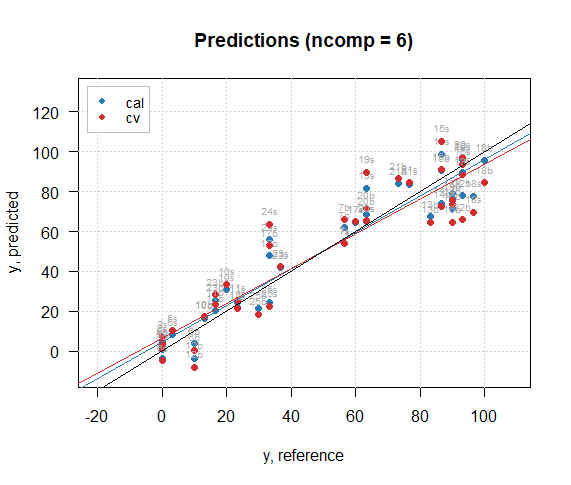
Изображение выглядит как текст

Автоматически созданное описание

After:

Изображение выглядит как текст

Автоматически созданное описание



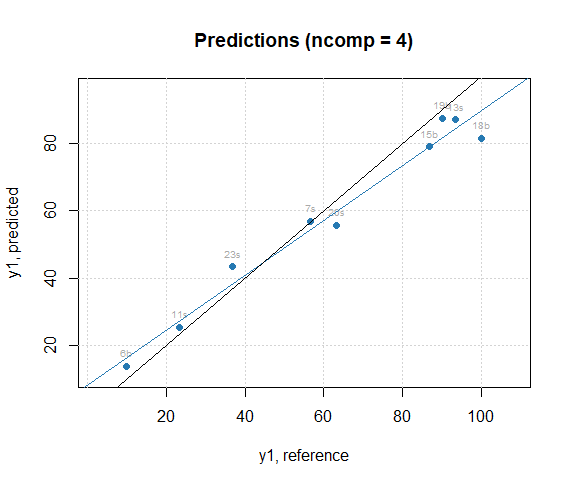
**Results for the test set:**

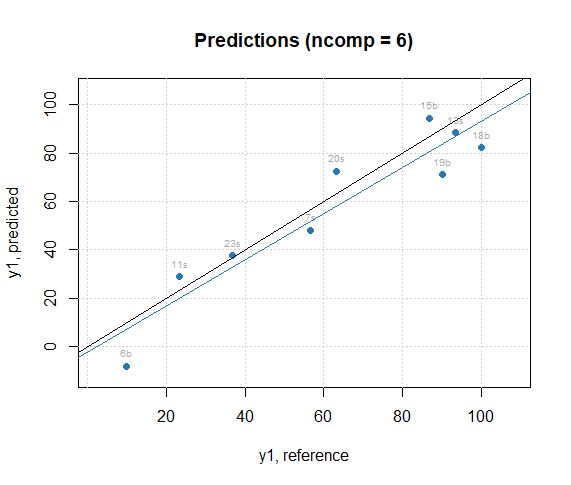
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | PLS | SVM | RF | KRLS |
| RMSEP original data | 7.917 (11.914)\* | 11.392 | 18.16 | 12.823 |
| Параметры | 4 компоненты (6 компонент) | cost: 100  gamma: 1e-04  epsilon: 0.15 | mtry = 5, ntree = 300, node size = 5 | Sigma = ncol, lambda по cv |
| RMSEP processed data | 11.442 (12.916)\*\* | 23.121 | 15.788 | 17.361 |
| Параметры | 2 компоненты (3 компоненты) | Cost: 100  Epsilon: 0.15 gamma: 1e-04 | mtry = 5, ntree = 300, node size = 5 | Sigma = ncol, lambda по cv |

\*по результатам калибровки и cv, алгоритм выбирает 6 компонент, но для тестовых данных ошибка самая малая на 4 компонентах.

Изображение выглядит как текст, квитанция

Автоматически созданное описание





\*\*у новой модели выбирает 3 компоненты, хотя оптимально 5

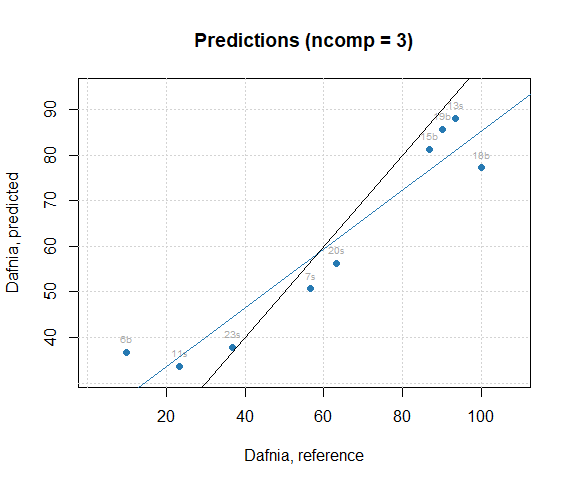
Изображение выглядит как текст

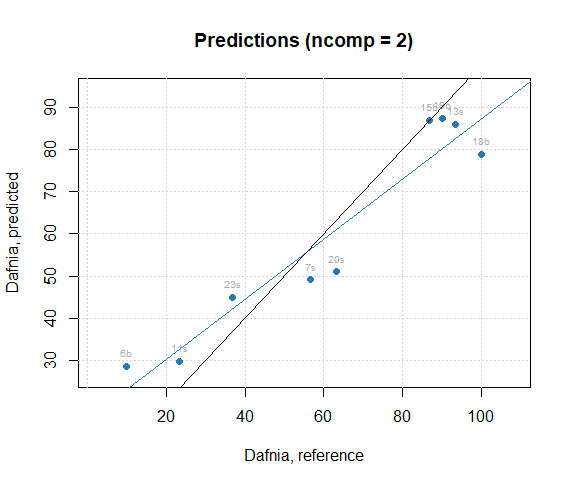
Автоматически созданное описание

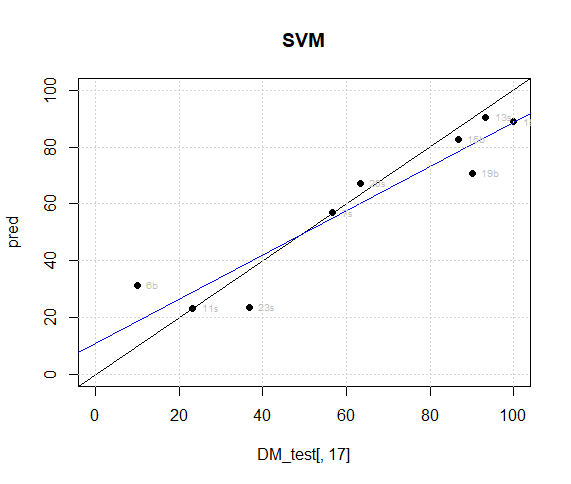
для тестового набора выбирает 3 компоненты, на 2ух ошибка наименьшая

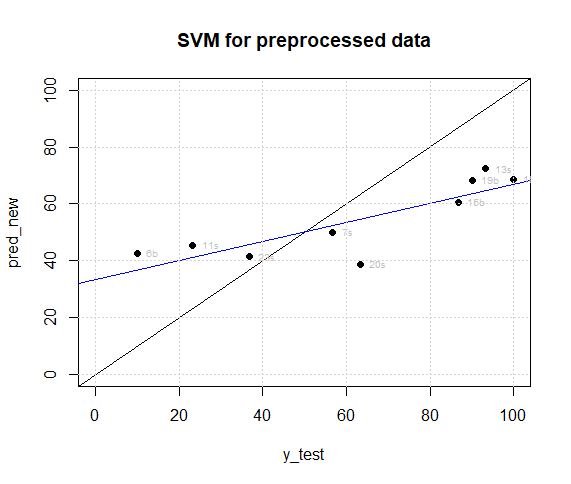
Изображение выглядит как текст, квитанция

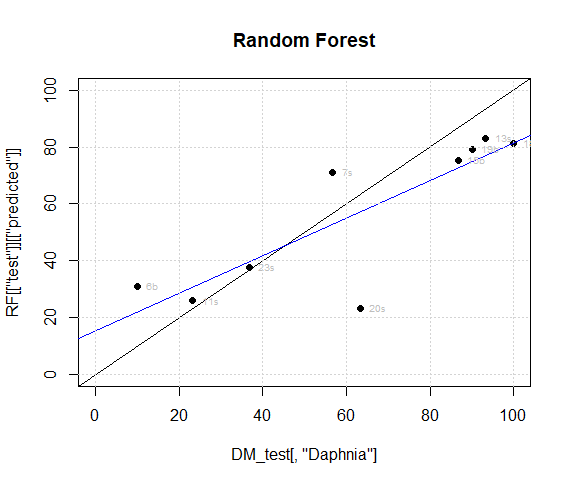
Автоматически созданное описание

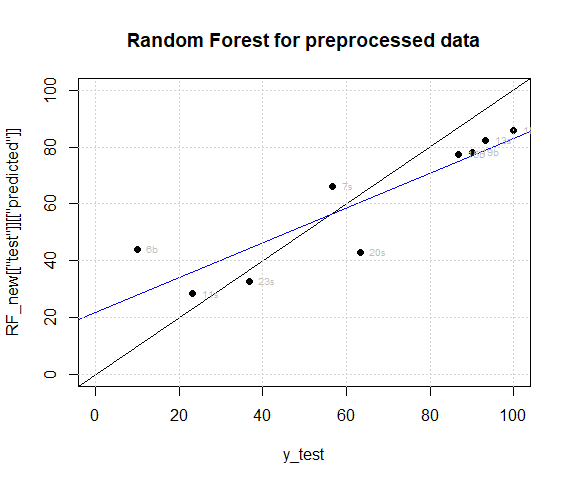


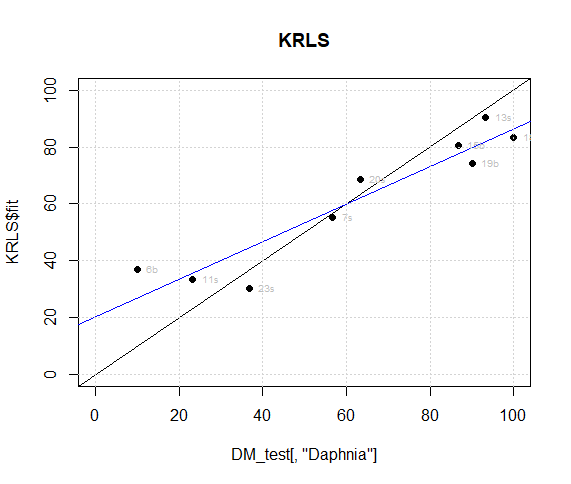


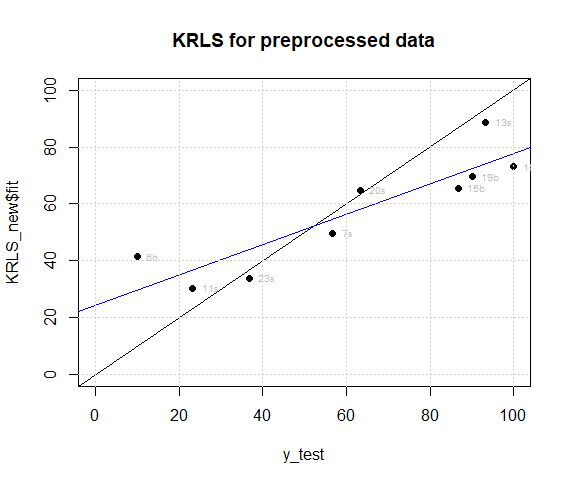




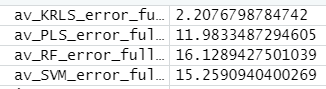








Изображение выглядит как стол

Автоматически созданное описание

**Vibrio Fisheri**

**Outliers:** 'w24', 'w6' (2)

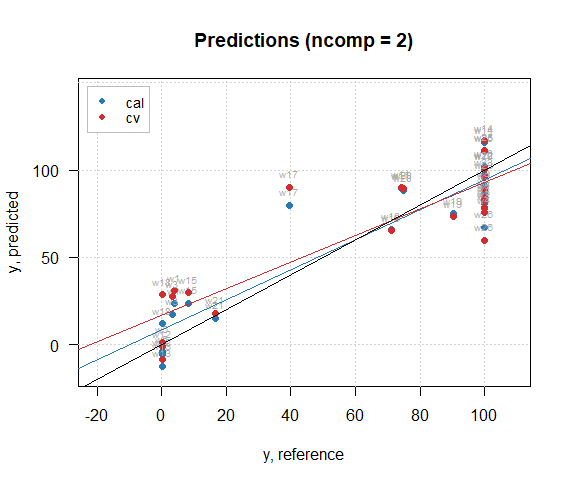
**Test set:** 'w12', 'w16', 'w22', 'w9', 'w8', 'w21' (6)

**Training set:** 18 samples

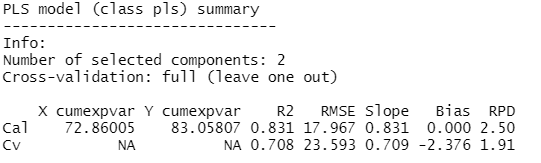
PLS after removing outliers:

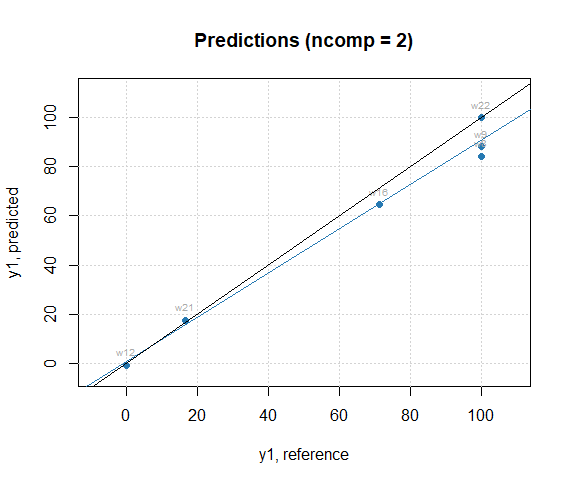
Изображение выглядит как текст

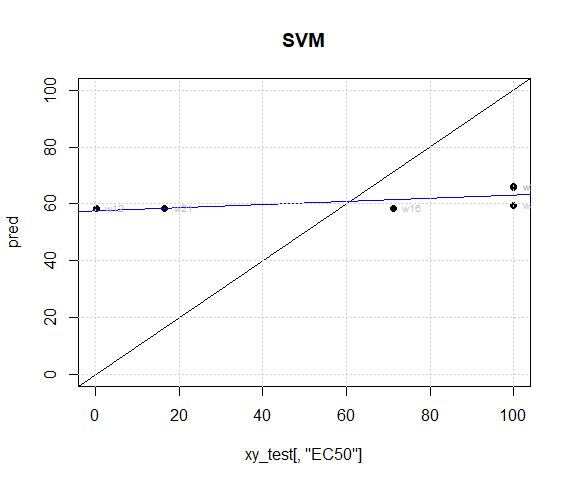
Автоматически созданное описание

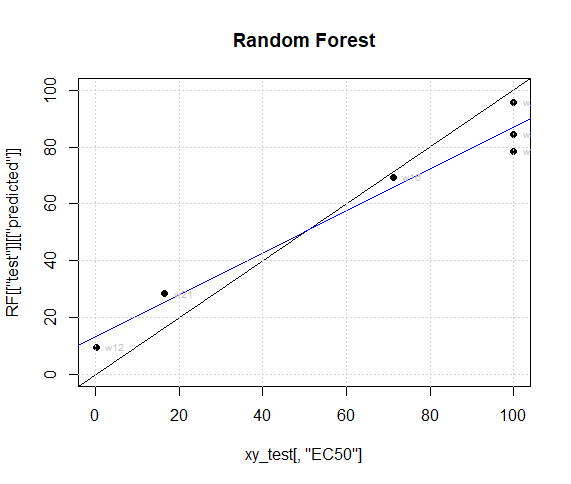


|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | PLS | SVM | Random Forest | KRLS |
| RMSEP | 8.600 | 39.312 | 12.635 | 18.064\* |
| Параметры | 2 comp | Cost: 100  Gamma: 1e-04  Epsilon: 0.01 | mtry = 8, ntree = 500, nodesize = 5 | - |

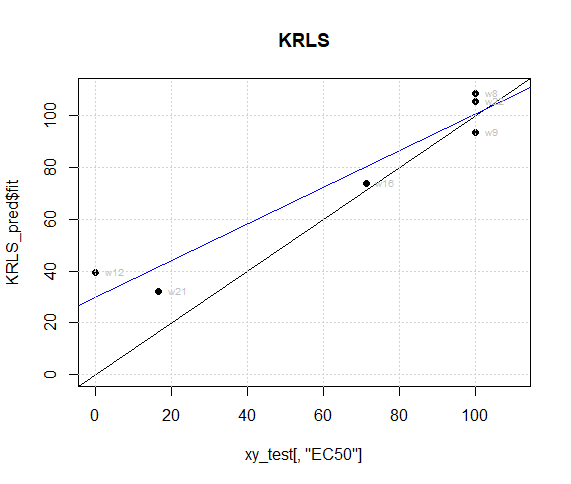




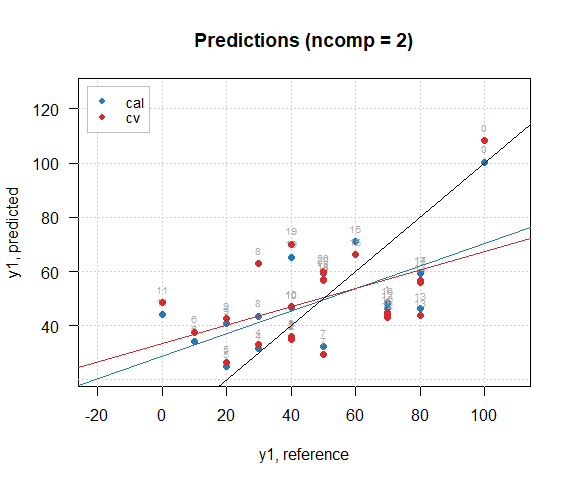




\* у модели очень большой bias при неплохих остальных показателях -- переобучение



**Model data**

Original data:

Изображение выглядит как текст

Автоматически созданное описание

**Outliers:** 11,19, 9, 18 (4)

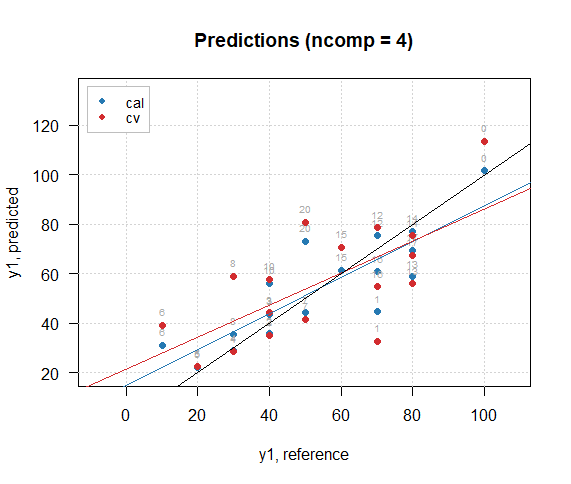
**Test set:** -

**Training set:** -

After removing outliers:

Изображение выглядит как текст

Автоматически созданное описание



Monte carlo cv:

Изображение выглядит как стол

Автоматически созданное описание

Full cv Оставила все:

**Изображение выглядит как стол

Автоматически созданное описание**

**Изображение выглядит как текст

Автоматически созданное описание**

**(!! Там счет с нуля)**

Full cv (without outliers):

Изображение выглядит как текст, стол

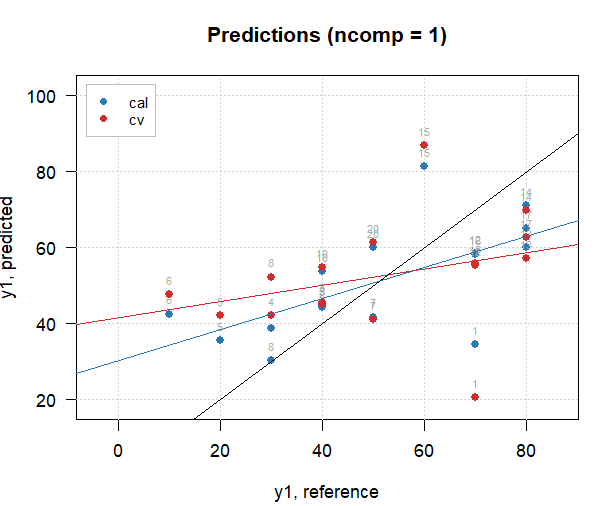
Автоматически созданное описание

После удаления outliers и точки 0:

Изображение выглядит как стол

Автоматически созданное описание

(но ноль не является выбросом)



Оставила все, кроме 0:

Изображение выглядит как текст, стол

Автоматически созданное описание

Оставила все кроме 0 и 11:

Изображение выглядит как текст, стол

Автоматически созданное описание  
**(это уже не очень важно, на PLS cal и cv нормальные рез-ты не дает)**