# Categorical soil attribute modeling and mapping: Validation

Soil Security Laboratory

2018

## 1 Model validation of categorical prediction models

The special characteristic of categorical data and its prediction within models, is that a class is either predicted or it is not. For binary variables. the prediction is either a *yes* or *no*, *black* or *white*, *present* or *absent* etc. For multinomial variables, there are more than 2 classes, for example either *black*, *grey*, or *white* etc (which could actually be an ordinal categorical classification, rather than nominal).

There is no in-between; rather discrete entities. Exceptions are that some models do estimate the probability of the existence of a particular class, which will be touched on later. Additionally, there are methods of fuzzy classification which are common in the soil sciences (McBratney and Odeh, 1997), but will not be covered in this section. Discrete categories and models for their prediction require other measures of validation than those that were used for continuous variables. The most important quality measures are described in Congalton (1991) and include:

1. Overall accuracy
2. User's accuracy
3. Producer's accuracy
4. Kappa coefficient of agreement

Using a contrived example, each of these quality measures will be illustrated. We will make a $4 \times 4$ matrix, and call it `con.mat`, and append some column and row names — in this case Australian Soil Classification Order codes. We then populate the matrix with some more-or-less random integers.

```
con.mat <- matrix(c(5, 0, 1, 2, 0, 15, 0, 5, 0, 1, 31, 0, 0, 10, 2, 11), nrow = 4,
    ncol = 4)
rownames(con.mat) <- c("DE", "VE", "CH", "KU")
colnames(con.mat) <- c("DE", "VE", "CH", "KU")
con.mat
```

```
##    DE VE CH KU
```

```
## DE  5  0  0  0
## VE  0 15  1 10
## CH  1  0 31  2
## KU  2  5  0 11
```

`con.mat` takes the form of a confusion matrix, and ones such as this are often the output of a classification model. If we summed each of the columns (using the `colSums` function), we would obtain the total number of observations for each soil class. Having column sums reflecting the number of observations is a widely used convention in classification studies.

```
colSums(con.mat)
```

```
## DE VE CH KU
##  8 20 32 23
```

Similarly, if we summed each of the rows we would retrieve the total number of predictions of each soil class. The predictions could have been made through any sort of model or classification process.

```
rowSums(con.mat)
```

```
## DE VE CH KU
##  5 26 34 18
```

Therefore, the numbers on the diagonal of the matrix will indicate fidelity between the observed class and the subsequent prediction. Numbers on the off-diagonals indicate a mis-classification or error. Overall accuracy is therefore computed by dividing the total correct (i.e., the sum of the diagonal) by the total number of observations (sum of the column sums).

```
ceiling(sum(diag(con.mat))/sum(colSums(con.mat)) * 100)
```

```
## [1] 75
```

Accuracy of individual classes can be computed in a similar manner. However, there is a choice of dividing the number of correct predictions for each class by either the totals (observations or predictions) in the corresponding columns or rows respectively. Traditionally, the total number of correct predictions of a class is divided by the total number of observations of that class (i.e. the column sum). This accuracy measure indicates the probability of an observation being correctly classified and is really a measure of omission error, or the "producer's accuracy". This is because the producer of the model is interested in how well a certain class can be predicted.

```
ceiling(diag(con.mat)/colSums(con.mat) * 100)
```

```
## DE VE CH KU
## 63 75 97 48
```

Alternatively, if the total number of correct predictions of a class is divided by the total number of predictions that were predicted in that category, then this result is a measure of commission error, or "user's accuracy". This measure is indicative of the probability that a prediction on the map actually represents that particular category on the ground or in the field.

```
ceiling(diag(con.mat)/rowSums(con.mat) * 100)
```

```
##  DE  VE  CH  KU
## 100  58  92  62
```

So if we use the `DE` category as an example, the "model" predicts this class correctly 63% of the time, but when it is actually predicted it is correct 100% of the time.

The Kappa coefficient is another statistical measure of the fidelity between observations and predictions of a classification. The calculation is based on the difference between how much agreement is actually present ("observed" agreement) compared to how much agreement would be expected to be present by chance alone ("expected" agreement). The observed agreement is simply the overall accuracy percentage. We may also want to know how different the observed agreement is from the expected agreement. The Kappa coefficient is a measure of this difference, standardized to lie on a -1 to 1 scale, where 1 is perfect agreement, 0 is exactly what would be expected by chance, and negative values indicate agreement less than chance, i.e., potential systematic disagreement between observations and predictions. The Kappa coefficient is defined as:

$$K = \frac{p_o - p_e}{1 - P_e} \tag{1}$$

where $p_o$ is the overall or observed accuracy, and $p_e$ is the expected accuracy, where:

$$p_e = \sum_{i=1}^{n} (\frac{colSum_i}{TO}) \times (\frac{rowSum_i}{TO}) \tag{2}$$

$TO$ is the total number of observations and $n$ is the number of classes. Rather than scripting the above equations, the kappa coefficient together with the other accuracy measures are contained in a function called `goofcat` in the `ithir` package. As we already have a confusion matrix prepared, we can enter it directly into the function as in the script below.

```
goofcat(conf.mat = con.mat, imp = TRUE)
```

```
## $confusion_matrix
##    DE VE CH KU
## DE  5  0  0  0
## VE  0 15  1 10
## CH  1  0 31  2
## KU  2  5  0 11
##
## $overall_accuracy
## [1] 75
##
## $producers_accuracy
## DE VE CH KU
```

```
## 63 75 97 48
##
## $users_accuracy
##  DE  VE  CH  KU
## 100  58  92  62
##
## $kappa
## [1] 0.6389062
```

A rule of thumb as indicated in Landis and Koch (1977) for the interpretation of the Kappa coefficient is:

< Less than chance agreement.
0.01 - 0.20 Slight agreement.
0.21 - 0.40 Fair agreement.
0.41 - 0.60 Moderate agreement.
0.61 - 0.80 Substantial agreement.
0.80 - 0.99 Almost perfect agreement.

# References


Congalton, R. G.
1991. A review of assessing the accuracy of classifcations of remotely sensed data. *Remote Sensing of the Environment*, 37:35–46.

Landis, R. and G. G. Koch
1977. The measurment of observer agreement for categorical data. *Biometrics*, 33:159–174.

McBratney, A. B. and I. O. A. Odeh
1997. Application of fuzzy sets in soil science: Fuzzy logic, fuzzy measurments and fuzzy decisions. *Geoderma*, 77:85–113.