Appendix E

Mathematical Formulation of Statistical Analysis

This appendix describes the mathematical formulas used in the present study for the statistical analysis of linear and circular data. The statistics for linear data are defined according to Schneggenburger (1998) and Luo (1995), while those ones related with circular data are defined according to Mardia (1972) and Fischer (1993). The notation used for the formulation is described in Table E-1.

Table E-1: Explanation of the statistic notation

Notation	Description
x	Observed values
y	Model values
θ_x	Observed circular values
θ_y	Model circular values

E.1 Statistics for Linear Data

The calculation of the average value for the observed and model data is defined in equation E.1; with N being the number of data values.

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \qquad \bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i \tag{E.1}$$

The root mean square error (rmse) or average quadratic deviation between the observed and model values is given by the equation E.2.

$$\text{rmse} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - y_i)^2} \tag{E.2}$$

The goodness of fit rule or correlation coefficient, which explains how well the model results approach the observed values, is given by equation E.3. A statistic close to one indicates that linear correlations exist between the model results and the measurements. For a statistic value around zero no correlation between the observed and model values exist.

$$gof = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{N}(y_i - \bar{y})^2}} \qquad (E.3)$$

E.2 Statistics for Circular Data

The calculations of the different statistics according to the formulas described in E.1 cannot be exported for circular data. In the present research a special statistical treatment has been applied to velocity direction data. Thus, the complete set of the equations for the statistics used is given.

Thus, the average direction for a sample of circular data has been calculated using equation E.4.

$$\bar{\theta} = \arctan\left(\frac{\bar{S}}{\bar{C}}\right) \qquad (E.4)$$

With

$$\bar{S} = \frac{1}{N}\sum_{i=1}^{N}\sin(\theta_i)$$

$$\bar{C} = \frac{1}{N}\sum_{i=1}^{N}\cos(\theta_i)$$

According to the length of the vector and its average direction the variance S_o is defined as

$$S_o = 1 - \frac{1}{N}\sum_{i=1}^{N}\cos(\theta_i - \bar{\theta}) = 1 - \bar{R} \qquad (E.5)$$

With $\bar{R} = \sqrt{\left(\bar{C}^2 + \bar{S}^2\right)}$

The variance lies always in the interval [0,1].

For circular values the standard deviation σ is not calculated as the square root of the variance, but according to the expression defined in equation E.6. Thus, the standard deviation value obtained lies in the interval [0,∞].

$$\sigma = \left(-2 \cdot \ln\left(1 - S_o\right)\right)^{\frac{1}{2}} \tag{E.6}$$

The root-mean-square error formula (E.7) for circular values has analogy with the definition given in E.1 for the calculation of rmse of linear data. This represents a measure of the square root of the standard deviation of the difference between observed and calculated directions.

$$\text{rmse} = \sqrt{-2 \cdot \ln\left(\frac{1}{N}\sum_{i=1}^{N}\cos\left(\theta_{x,i} - \theta_{y,i}\right)\right)} \tag{E.7}$$