Universidad Politécnica de Valencia
Dpto. de Ingeniería Cartográfica Geodesia y Fotogrametría

RESEARCH WORK

Geostatistical approaches for the rainfall interpolation in the Comunidad Valenciana (Spain)

September 2004

Author
Cristina PORTALÉS RICART

Tutors
Josep Eliseu PARDO PASCUAL
Ángel Antonio BALAGUER BESER

INDEX

INDEX 1

ABSTRACT 2

1. INTRODUCTION 2

2. CASE STUDY 3

2.1. Study area 3

2.2. Rainfall data 4

2.3. Secondary variables 6

3. INTERPOLATION PROCEDURES: MATHEMATICAL BACKGROUND 7

3.1. Inverse distance weighted (IDW) 8

3.2. Local polynomial interpolation (LPI) 8

3.3. Multiple regression (MR) 9

3.4. The Kriging estimator 9

3.5. Cokriging 14

4. RESULTS: INTERPOLATED RAINFALL MAPS 16

4.1. IDW and LPI 17

4.2. Multiple Regression 17

4.3. Kriging and Cokriging approaches 20

5. DISCUSSION 22

5.1. Agreement of the models 23

5.2. Observed vs. predicted values 24

5.3. Error maps 25

5.4. Error statistical analysis 25

5.5. Regional analysis 26

6. CONCLUSIONS 27

7. FURTHER WORK 28

8. REFERENCES 29

9. APPENDIX 30

Geostatistical approaches for the rainfall interpolation in the Comunidad Valenciana (Spain)

Cristina Portalés Ricart

ABSTRACT

This report analyses the different approaches currently used to achieve a reliable model for the precipitation resource. The study was carried out for the Comunidad Valenciana, an Spanish Mediterranean coastal region. Monthly rainfall means were available at a total of 207 meteorological stations. Sample data was filtered and grouped into five variables (annual, spring, summer, autumn and winter) and several geostatistical and traditional interpolation methods were tested. To validate the maps, the cross validation method was used and some statistical estimators were computed. Finally, error and regional analysis were performed. The better overall results were achieved for the simple and ordinary Cokriging geostatistical approaches and for the multiple regression procedure with several topographic and geographic variables.

KEY WORDS: precipitation, interpolation, multiple regression, geostatistics.

1. INTRODUCTION

Precipitation is a high valued climatologic resource. Nevertheless, accurate data is only available at certain points, the meteorological stations, and therefore the assessment of water resources is often based on the precipitation spatial modeling. These models can be used, between others, for long-term planning, hydrological forecasting, hydro-chemical modelling and human impact studies [6].

Several interpolation methods have been proposed. Traditionally, simple methods have been used, as for instance linear or quadratic approaches, for which the interpolated values at unsampled locations are derived from the precipitation values at sample points. Recently, improved methods which include secondary topographic and geographic information in the interpolation procedure have been proposed, for example multiple regression or Cokriging approaches [5; 6; 9; 10; 14]. This have quite sense, since the relationship between precipitation and other variables is well described. For instance, it is well known that precipitation generally increases with elevation or with the proximity to a mass of water [5; 10].

Several authors have looked for the most reliable overall model able to describe the precipitation resource. Goovaerts [5] used geostatistical algorithms to include elevation into the interpolation procedure at the South of Portugal; Ninyerola [10] used a lineal regression equation with correctors modeled by Kriging estimators in the area of Catalunya (Spain); Johansson and Chen [6] found a regression model that included the wind variable for Sweden; Marquínez et. al. [9], used a regression model performed with topographic variables for Cantabria (Spain); and Vicente-Serrano et. al. [14], concluded that the best results were achieved by geostatistical methods and a regression model formed by four geographic and topographic variables for the middle Ebro Valley (Spain).

As it can be seen, a valid formula accepted to model the precipitation has not been found, and this seems to depend on the characteristics and extension of the study area as well as on the available topographic variables and the sample data temporal scale (monthly, annual, etc.). In the following sections, the methodology followed to achieve reliable models for the seasonal and annual precipitation variables in the Comunidad Valenciana is described. At the end, some conclusions and proposed further work are given.

2. CASE STUDY

In this section a brief introduction is given about the situation and characteristics of the study area and the rainfall and secondary topographical variables used for the interpolation procedures.

2.1. Study area

The study area is the "Comunidad Valenciana" (figure 1), a Mediterranean coastal region placed on the East of Spain with an area of 23.255 Km^2. The physical geography of the Comunidad Valenciana is quite heterogeneous. It is divided into two sectors: interior and coast. The first one is a mountainous area, integrated in the "Sistama

Ibérico" and the "Cordillera Subbética". The major altitude is reached by the Penyagolosa (1.813 m). The second one is a littoral plain region, principally constituted by smooth beaches and some littoral lakes.

Climatologically, the Comunidad Valenciana is within the dry region of the Iberian Peninsula, with annual precipitation means varying from 400 to 600 mm. In general, it is slightly dryer than the rest of Mediterranean regions placed at similar latitudes, due that it is situated to leeward from the West zonal flow [12].

2.2. Rainfall data

Within this region, a total of 207 rainfall stations of the National Institute of Meteorology (INM) network are placed. In figure 1, the spatial distribution of these stations can be seen. For all these stations, the monthly rainfall means are available for a period varying from 5 to 30 completely years, between 1961-1990, having most of them rainfall data for about 20 years.

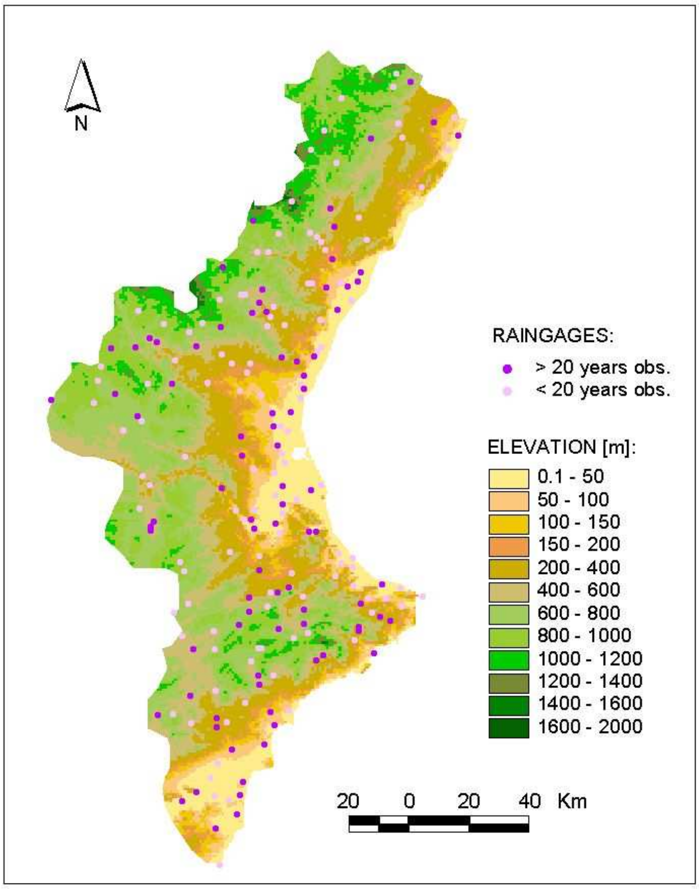


Figure 1. Elevation map of the Comunidad Valenciana together with all the observatories.

Although the World Meteorological Organization (WMO) recommends to use 30 year means, other authors have used shorter series with satisfactory results [9; 10]. In our case, periods varying from 20 to 30 years had to be used due to the lack of data, i.e., there were only 9 observatories with the complete series of 30 years. Of the total of 207 rainfall stations only 86 have means for 20 years or more, which represents the 41% of the rainfall stations. Filtering more than half the data seems to be risky, but it was checked that mixing data from stations having registered rainfall data for only 8 years with others having data for more than 20 years was not plausible, because of the non-normality of data. Furthermore, in our case, the filtering procedure did not influence in the global distribution of the rainfall stations, i.e., there were no (additional) "holes" in their spatial distribution (figure 1).

Monthly rainfall means were grouped into the following variables: ANN (annual), SPR (spring: March, April, May), SUM (summer: June, July, August), AUT (autumn: September, October, November) and WIN (winter: December, January, February). To check if the data followed a normal distribution, the chi-square and the Kolmogorov-Smirnov tests were applied (see analysis reports 1 to 5 in the appendix). The variables ANN, SPR and AUT followed a normal distribution with the 90% or higher confidence. For the variables SUM and WIN, a natural logarithmic transformation was necessary.

In tables 1 and 2, some statistics and correlation coefficients between the rainfall variables can be seen.

	ANN	**SPR**	**SUM**	**AUT**	**WIN**
TOTAL	43005,4	10921,6	5138,5	16404,6	10540,2
MEAN	500,1	127,0	59,8	190,8	122,6
STD. DEV.	137,9	32,8	22,1	56,8	46,1
MAX	832,1	214,2	127,1	317,4	283,3
MIN	236,7	58,5	20,3	87,3	52,7
RANGE	595,4	155,7	106,8	230,1	230,6

Table 1. Statistics for the annual and seasonal rainfall data (86 observatories).

	ANN	**SPR**	**SUM**	**AUT**	**WIN**
ANN	1,00	0,96	0,43	0,95	0,93
SPR		1,00	0,46	0,84	0,90
SUM			1,00	0,29	0,12
AUT				1,00	0,87
WIN					1,00

Table 2. Correlation coefficients between the rainfall variables.

As it can be seen, the SUM variable (highlighted) is a very dry season compared with the other ones and therefore it is low correlated with the other seasons as well as with the annual precipitation. These differences should be remained, as a different behaviour in the interpolation procedures is expected.

2.3. Secondary variables

It is known that some variables derived from the topography and/or geography have some kind of influence in the precipitation. For example, precipitation generally increases with elevation and several authors have introduced them into geostatistical interpolation methods [5; 6; 9; 10; 14]. Other variables having influence are the slope, aspect, distance to the sea, etc. Therefore, taking these variables into account, when possible (see section 3), an improvement of the rainfall prediction is expected. In table 11 (in the appendix) all the tested secondary variables are shown. As it can be seen, some "main" variables (X, Y, distance to the sea, elevation, etc.) were achieved and then additional variables were calculated from multiplication between two of them.

Besides that (last files of table 11), the aspect variable was included in the study in a different manner. The "aspect" is defined as the angle between the North direction and the steepest down-slope direction [1]. This is a special variable because of its circular meaning which makes it not possible to apply it directly, but with some kind of linearisation (for instance a co-sinus transformation). In our case, the normalized scalar product was considered between the aspect direction "A", the direction of the minimal distance to the Mediterranean sea "C" and the cardinal points "N, NE, E, SE" (S, SW, W and NW have a correlation factor of –1 with the previous). Additionally, multiplications between C and the cardinal points were taken into account. The obtained values, that coincide with the cosine of the angle between the two considered vectors, are inside the interval [-1,1].

The mean values of a certain variable (e.g. Z) inside a certain buffer (e.g. circular buffer) were used in order to ensure a wider influence of them. Several tests were made with circular buffers with radii of 1, 2.5, 5 and 10 Km for all the variables. Finally, the buffers of 5 Km radius were selected because of achieving higher correlations with the precipitation variables. Moreover, it was also tested an special buffer consisting on the hillside (with an accumulation area of 5 Km) where an observatory is placed. It was concluded that this last buffer only had sense in the calculation of the aspect variable.

For the extraction of the variables elevation, slope and aspect, a Digital Elevation Model (DEM) covering all the study area was used. This model was available within a grid size of 25x25 m. For some operations, it was rescaled to a 100x100 m grid size. The areas defining the hillsides were automatically achieved by the "VESSANTS" program, a macro developed by the "Departamento de Ingeniería Cartográfica, Geodesia y Fotogrametría" of the "Universidad Politécnica de Valencia", that can run under ArcView GIS.

The spatial analysis as well as all the required map calculations in the process of variable extractions (for instance, the sinus of aspect, buffers around the observatories with a given radius, visual check, etc.) were made by using the software package "ArcView GIS 3.2" of ESRI.

3. INTERPOLATION PROCEDURES: MATHEMATICAL BACKGROUND

Consider the problem of estimating a certain variable (e.g. rainfall) at an unsampled location $\mathbf{u}$. Let $\{ z(\mathbf{u}_\alpha), \alpha = 1,...,n \}$ be the set of (rainfall) data measured at n locations $\mathbf{u}_\alpha$. Consider now the situation where the primary data (rainfall) is supplemented by secondary data (e.g. elevation, slope, etc.) available at all estimation points and related to $(N_v - 1)$ continuous attributes z_i, $\{ z_i(\mathbf{u}_\alpha)$, $\alpha = 1,...,n$, $i = 2,...,N_v \}$. The solution of this problem is not unique, and depends on the interpolation procedure as well as the consideration of secondary information.

In this section, a brief introduction of the different interpolation methods used in this study is given. For a more detailed description see [2; 3; 4; 8]. Some methods are well known and have been traditionally used, for instance, inverse distance weighted (IDW), polynomial functions or simple/multiple regression methods. Geostatistics is a relative new discipline based on a previous statistical data analysis. Within this approach, a Kriging or a Cokriging interpolator is used.

In the following sub-sections, some theoretical background is given for all the considered methods.

3.1. Inverse distance weighted (IDW)

In the IDW approach, the values to be interpolated Z^*_{IDW} are estimated as a linear combination of several surrounding observations, with the weights being inversely proportional to the distance between observations and $\mathbf{u}$ to the power of p (see [5; 7]):

$$Z^*_{IDW} = \frac{\sum_{\alpha=1}^{n(\mathbf{u})} \left[\lambda_\alpha(\mathbf{u}) z(\mathbf{u}_\alpha)\right]}{\sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha(\mathbf{u})}, \text{ with } \lambda_\alpha(\mathbf{u}) = \frac{1}{\left|\mathbf{u} - \mathbf{u}_\alpha\right|^p} \qquad (1)$$

where $n(\mathbf{u})$ is the number of the points at location $\mathbf{u}$ considered for the estimation. The basic idea behind the weighing function, is that observations that are closer in the ground are more similar than that being further away. Hence observations closer to $\mathbf{u}$ will receive a larger weight.

IDW is an exact interpolator, where the maximum and minimum values in the interpolated surface can only occur at sample points [7].

3.2. Local polynomial interpolation (LPI)

In a polynomial interpolation approach, the values being interpolated Z^*_{LPI} are calculated as a function of the ground coordinates (X, Y). The polynomial can be of the form [8]:

$$Z^*_{LPI}(\mathbf{u}) = a_0 + a_1 X(\mathbf{u}) + a_2 Y(\mathbf{u}) + a_3 X^2(\mathbf{u}) + a_4 Y^2(\mathbf{u}) + a_5 X(\mathbf{u}) Y(\mathbf{u}) + \cdots \qquad (2)$$

The a_i coefficients are calculated from the data points coordinates (X_i, Y_i, Z_i) by solving a system of the form:

$$Z_i = a_0 + X_i a_1 + Y_i a_2 + X_i^2 a_3 + Y_i^2 a_4 + X_i Y_i a_5 + \cdots, \quad \text{or} \quad \mathbf{Z}_i = \mathbf{b}_i^T \mathbf{a} \qquad (3)$$

If the number of the data points is greater than that of the coefficients, the system can be solved by a minimum square procedure. Additional weights can be added, as for example $\lambda_\alpha(\mathbf{u}) = \frac{1}{|\mathbf{u}-\mathbf{u}_\alpha|^2}$, where $|\mathbf{u}-\mathbf{u}_\alpha|^2$ is the distance between each data point to the point being interpolated. The system is of the form:

$$\mathbf{a} = (\mathbf{B}^T \boldsymbol{\lambda} \mathbf{B})^{-1} \mathbf{B}^T \boldsymbol{\lambda} \mathbf{z} \quad (4)$$

An important drawback within this interpolation is that when the polynomial reaches a certain degree, it tends to oscillate wildly in the undetermined areas. These oscillations can even increase in the borders of the interpolated area, making it to be only acceptable at the centre of it. To avoid this inconvenience, several polynomials surfaces are calculated, each one within specified overlapping neighbourhoods, making the procedure to be "local".

3.3. Multiple regression (MR)

Within the MR procedure, several independent variables (secondary data) are used to predict one dependent variable (rainfall). The value of a variable Z^*_{MR} at unsampled points is predicted by the following equation (see [5; 14]):

$$Z^*_{MR}(\mathbf{u}) = a_0 + a_1 y_1(\mathbf{u}) + a_2 y_2(\mathbf{u}) + \ldots + a_n y_n(\mathbf{u}) \quad (5)$$

where the $a_0,\ldots,a_n$ are the regression coefficients and $y_1(\mathbf{u}),\ldots,y_n(\mathbf{u})$ are the values of the different independent variables at location $\mathbf{u}$.

A major shortcoming of this type of interpolation, is that the variable being interpolated at a particular grid node $\mathbf{u}$ is derived only from the secondary variables at $\mathbf{u}$, regardless of the records at the neighbouring observatories $\mathbf{u}_\alpha$. Such an approach amounts at assuming that the residual values $r(\mathbf{u}_\alpha) = z(\mathbf{u}_\alpha) - f(y(\mathbf{u}_\alpha))$ are spatially uncorrelated.

3.4. The Kriging estimator

The Kriging interpolation is based on the assumption that the parameter being interpolated can be treated as a regionalized variable. When a variable is distributed in space, it is said to be regionalized. Such a variable varies in a continuous manner from one location to the next one. Therefore, points that are near each other have a certain degree of spatial correlation, meanwhile points that are widely separated can be considered statistically independent.

In the weighting function, instead of the Euclidean distance (e.g. IDW approach), the semivariogram is used as a measure of dissimilarity between observations (see [4]). The experimental semivariogram $\hat{\gamma}(\mathbf{h})$ is computed as half the average squared difference between the components of data pairs:

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} [z(\mathbf{u}_\alpha) - z(\mathbf{u}_\alpha + \mathbf{h})]^2 \quad (6)$$

where $N(\mathbf{h})$ is the number of pairs of data location at a vector $\mathbf{h}$ apart within a certain tolerance angle. The following are the most frequently used basic semivariogram models [3; 4]:

1) **Spherical model** with range a

$$\gamma(h) = Sph\left(\frac{h}{a}\right) = \begin{cases} 1.5\frac{h}{a} - 0.5\left(\frac{h}{a}\right)^3 & if \quad h \le a \\ 1 \quad otherwise \end{cases} \quad (7)$$

2) **Exponential model** with practical range a

$$\gamma(h) = 1 - \exp\left(\frac{-3h}{a}\right) \quad (8)$$

3) **Gaussian model** with practical range a

$$\gamma(h) = 1 - \exp\left(\frac{-3h^2}{a^2}\right) \quad (9)$$

Where the range a is the distance at which the model reaches the sill *(*$\gamma(h)/C$=1 in figure 2); the practical range a is defined as the distance at which the model value is at 95% of the sill.

The election of the model depends on the behaviour of the experimental semivariogram at the origin (figure 2): for a parabolic behaviour, the Gaussian model is the best suited; for a linear behaviour, a spherical or an exponential model can be used.

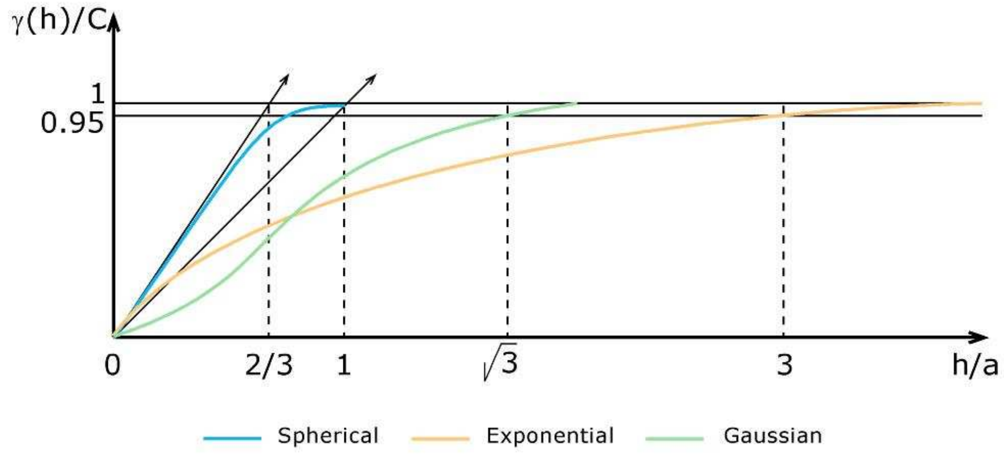


Figure 2. Variogram models (after [3]).

For anisotropic behaviours, the phenomenon can be modeled by fitting more than one semivariogram. A phenomenon is said to be anisotropy when its pattern of spatial variability changes with direction. The next figure shows the variation of the ranges $a_{\alpha i}$ as a function of the direction αi in isotropic and anisotropic models.

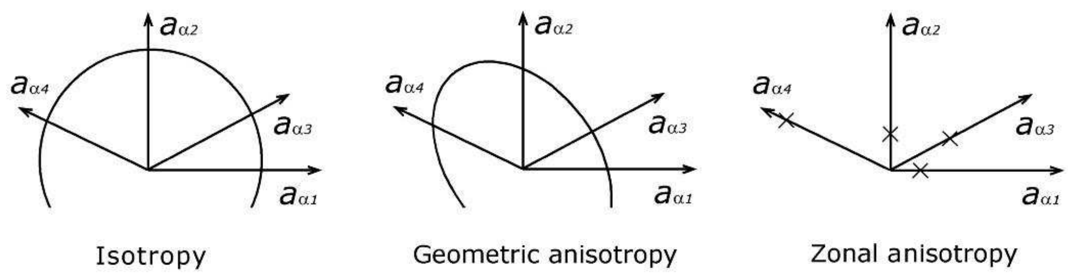


Figure 3. Ranges in case of isotropic and anisotropic models, where α1, α2, α3, α4 are four horizontal directions (after [3]).

There are three possible cases:

1) If the graph can be approximated to a circle of radius *a* for all the horizontal directions αi, the phenomenon can be considered as **isotropic**.
2) If the graph can be approximated by an ellipse (a linear transformation of a circle), the phenomenon is a **geometric anisotropy**. In this case, the range change with the direction.
3) If the graph cannot be fitted to a second-degree curve, a **zonal anisotropy** must be considered. In this case, the sill value does change with the direction.

Within the achievement of the semivariogram model, it should be taken into account that the semivariogram value may not tend to zero when *h* tends to zero. This discontinuity is called the nugget effect and is due to measurement error and/or spatial sources of variation at distances smaller than the shortest sampling interval [4]. The partial sill is defined as the ratio of the nugget discontinuity to the sill value.

Once the semivariogram model is defined, the Kriging interpolation can be performed. All Kriging estimators are variants of the basic linear regression estimator $Z^*(\mathbf{u})$ which is defined as [4]:

$$Z^*(\mathbf{u}) - m(\mathbf{u}) = \sum_{\alpha=1}^{n(u)} \lambda_\alpha(\mathbf{u})\left[Z(\mathbf{u}_\alpha) - m(\mathbf{u}_\alpha)\right] \tag{10}$$

where $n(\mathbf{u})$ is the number of the points at location $\mathbf{u}$ and $\lambda_\alpha(\mathbf{u})$ is the weight assigned to the datum $z(\mathbf{u}_\alpha)$ interpreted as a realization of the random variable $Z(\mathbf{u}_\alpha)$. The values $m(\mathbf{u})$ and $m(\mathbf{u}_\alpha)$ are the expected values of the random variables $Z(\mathbf{u})$ and $Z(\mathbf{u}_\alpha)$.

All Kriging estimators are required to be unbiased and to minimize the error variance, $\sigma_E^2(\mathbf{u}) = Var\{Z^*(\mathbf{u}) - Z(\mathbf{u})\}$, under the constraint that the expected error is zero:

$$E\{Z^*(\mathbf{u}) - Z(\mathbf{u})\} = 0 \tag{11}$$

Each random function is usually decomposed into a residual component and a trend component:

$$Z(\mathbf{u}) = R(\mathbf{u}) + m(\mathbf{u}) \tag{12}$$

The residual component is modeled as a stationary random function with zero mean and covariance function $C_R(\mathbf{h})$:

$$E\{R(\mathbf{u})\} = 0 \tag{13}$$

$$Cov\{R(\mathbf{u}), R(\mathbf{u}+\mathbf{h})\} = E\{R(\mathbf{u}) \cdot R(\mathbf{u}+\mathbf{h})\} = C_R(\mathbf{h}) \tag{14}$$

The expected value of the random variables Z at a certain location $\mathbf{u}$, is the value of the trend component at that location:

$$E\{Z(\mathbf{u})\} = m(\mathbf{u}) \tag{15}$$

Three Kriging variants can be distinguished according to the model considered for the trend $m(\mathbf{u})$ [2; 4]:

a) **Simple Kriging (SK)** considers the mean $m(\mathbf{u})$ to be known and constant through the study area A. The SK estimator can be written as:

$$Z^*_{SK}(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha^{SK}(\mathbf{u})Z(\mathbf{u}_\alpha) + \lambda_m^{SK}(\mathbf{u})m, \quad \text{with} \quad \lambda_m^{SK}(\mathbf{u}) = 1 - \sum_{\alpha=1}^{n(u)} \lambda_\alpha^{SK}(\mathbf{u}) \tag{16}$$

The SK weights are determined such as to minimize the error variance, while ensuring the unbiasedness of the estimator. These weights are estimated by solving the following stationary SK system:

$$\sum_{\beta=1}^{n(\mathbf{u})} \lambda_\beta^{SK}(\mathbf{u})C(\mathbf{u}_\alpha - \mathbf{u}_\beta) = C(\mathbf{u}_\alpha - \mathbf{u}), \qquad \alpha = 1,\ldots,n(u) \tag{17}$$

b) **Ordinary Kriging (OK)** accounts for local fluctuations of the mean by limiting the domain of stationarity of the mean to the local neighbourhood $W(u)$. In this case, the mean is considered to be constant and unknown. The OK estimator is written as:

$$Z^*_{OK}(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha^{OK}(\mathbf{u})Z(\mathbf{u}_\alpha), \quad \text{with} \quad \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha^{OK}(\mathbf{u}) = 1 \tag{18}$$

The OK weights are estimated by solving the following system of equations:

$$\begin{cases} \sum_{\beta=1}^{n(\mathbf{u})} \lambda_\beta^{OK}(\mathbf{u})C(\mathbf{u}_\alpha - \mathbf{u}_\beta) + \mu_{OK}(\mathbf{u}) = C(\mathbf{u}_\alpha - \mathbf{u}) \\ \sum_{\beta=1}^{n(\mathbf{u})} \lambda_\beta^{OK}(\mathbf{u}) = 1 \end{cases} \qquad \alpha = 1,\ldots,n(u) \tag{19}$$

where $\mu_\alpha^{OK}(\mathbf{u})$ is the Lagrange parameter that accounts for the constraint on the weights.

c) **Universal Kriging (UK)**, or Kriging with a trend model, considers that the unknown local mean $m(\mathbf{u}')$ smoothly varies within each local neighbourhood $W(\mathbf{u})$, hence over the entire study area A.

The trend component is modeled as a linear combination of functions of the coordinates:

$$m(\mathbf{u}') = \sum_{k=0}^{K} a_k(\mathbf{u}') f_k(\mathbf{u}') \qquad (20)$$

where $f_k(\mathbf{u}')$ are known functions of the location coordinates and $a_k(\mathbf{u}') \approx a_k$ are constant but unknown parameters $\forall \mathbf{u}' \in W(\mathbf{u})$.

The residual component $R(\mathbf{u})$ is usually modeled as a stationary random function with zero mean and covariance $C_R(\mathbf{h})$. The trend component is usually modeled with low-order (<2) polynomials because it is usually associated with a smoothly varying component of the z-variability.

The UK estimator is written as:

$$Z^*_{UK}(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha^{UK}(\mathbf{u}) Z(\mathbf{u}_\alpha), \quad \text{with} \quad \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha^{UK}(\mathbf{u}) f_k(\mathbf{u}_\alpha) = f_k(\mathbf{u}) \qquad (21)$$

and the UK system is:

$$\begin{cases} \sum_{\beta=1}^{n(\mathbf{u})} \lambda_\beta^{UK}(\mathbf{u}) C_R(\mathbf{u}_\alpha - \mathbf{u}_\beta) + \sum_{k=0}^{K} \mu_k^{UK}(\mathbf{u}) f_k(\mathbf{u}_\alpha) = C_R(\mathbf{u}_\alpha - \mathbf{u}) \\ \sum_{\beta=1}^{n(\mathbf{u})} \lambda_\beta^{UK}(\mathbf{u}) = 1 \\ \sum_{\beta=1}^{n(\mathbf{u})} \lambda_\beta^{UK}(\mathbf{u}) f_k(\mathbf{u}_\beta) = f_k(\mathbf{u}) \end{cases} \qquad \begin{matrix} \alpha = 1,\ldots,n(\mathbf{u}) \\ k = 0,\ldots K \end{matrix} \qquad (22)$$

The case of $K = 0$ is equivalent to the OK.

3.5. Cokriging

In the case of Cokriging interpolation, the primary data $\{ z_1(\mathbf{u}_{\alpha_1}),\ \alpha_1 = 1,\ldots n_1 \}$ are supplemented by secondary data related to $(N_v - 1)$ continuous attributes z_i, $\{ z_i(\mathbf{u}_{\alpha_i})$, $\alpha_i = 1,\ldots,n_i$, $i = 2,\ldots,N_v \}$ at any, possible different, locations. The linear estimator of equation (10) is extended to incorporate such additional information:

$$Z_1^*(\mathbf{u}) - m_1(\mathbf{u}) = \sum_{\alpha_1=1}^{n_1(\mathbf{u})} \lambda_{\alpha_1}(\mathbf{u}) \left[Z_1(\mathbf{u}_{\alpha_1}) - m_1(\mathbf{u}_{\alpha_1}) \right] + \sum_{i=2}^{N_v} \sum_{\alpha_i=1}^{n_i(\mathbf{u})} \lambda_{\alpha_{i1}}(\mathbf{u}) \left[Z_i(\mathbf{u}_{\alpha_i}) - m_i(\mathbf{u}_{\alpha_i}) \right] \qquad (23)$$

where $\lambda_{\alpha_1}(\mathbf{u})$ is the weight assigned to the primary datum $z_1(\mathbf{u}_{\alpha_1})$ and $\lambda_{\alpha_i}(\mathbf{u})$, $i > 1$, is the weight assigned to the secondary datum $z_i(\mathbf{u}_{\alpha_i})$. The quantities $m_1(\mathbf{u})$ and $m_i(\mathbf{u}_{\alpha_i})$ are the expected values for the random variables $Z_1(\mathbf{u})$ and $Z_i(\mathbf{u}_{\alpha_i})$.

All Cokriging estimators are required to be unbiased and to minimize the error variance, $\sigma_E^2(\mathbf{u}) = Var\{Z_1^*(\mathbf{u}) - Z_1(\mathbf{u})\}$ under the constraint that the expected error is zero:

$$E\{Z_1^*(\mathbf{u}) - Z_1(\mathbf{u})\} = 0 \qquad (24)$$

Each random function can be written as the sum of a residual component plus a trend component:

$$Z_i(\mathbf{u}) = R_i(\mathbf{u}) + m_i(\mathbf{u}), \qquad i = 1,\ldots,N_v \qquad (25)$$

The residual component is modeled as a stationary random function with zero mean and covariance function $C_i^R(\mathbf{h})$:

$$E\{R_i(\mathbf{u})\} = 0 \qquad (26)$$

$$Cov\{R_i(\mathbf{u}), R_i(\mathbf{u}+\mathbf{h})\} = E\{R_i(\mathbf{u}) \cdot R_i(\mathbf{u}+\mathbf{h})\} = C_i^R(\mathbf{h}) \qquad (27)$$

The cross covariance is:

$$C_{ij}^R(\mathbf{h}) = Cov\{R_i(\mathbf{u}), R_j(\mathbf{u}+\mathbf{h})\} \qquad (28)$$

According to the trend model $m_i(\mathbf{u})$, three Cokriging variants can be distinguished:

a) **Simple Cokriging (SCK)** considers each local mean $m_i(\mathbf{u})$ known and constant within the study area A. For $(N_v - 1)$ variables, the SCK estimator yields:

$$Z_{SCK}^{(1)*}(\mathbf{u}) - m_1 = \sum_{\alpha_1=1}^{n_1(\mathbf{u})} \lambda_{\alpha_1}^{SCK}(\mathbf{u})\left[Z_1(\mathbf{u}_{\alpha_1}) - m_1\right] + \sum_{i=2}^{N_v}\sum_{\alpha_i=1}^{n_i(\mathbf{u})} \lambda_{\alpha_i}^{SCK}(\mathbf{u})\left[Z_i(\mathbf{u}_{\alpha_i}) - m_i\right] \qquad (29)$$

And the SCK system is:

$$\sum_{j=1}^{N_v}\sum_{\beta_{ji}=1}^{n_j(\mathbf{u})} \lambda_{\beta_j}^{SCK}(\mathbf{u}) C_{ij}(\mathbf{u}_{\alpha_i} - \mathbf{u}_{\beta_j}) = C_{i1}(\mathbf{u}_{\alpha_i} - \mathbf{u}), \qquad \alpha_i = 1,\ldots,n_i(u),\ i = 1,\ldots,N_v \qquad (30)$$

b) **Ordinary Cokriging (OCK)** accounts for local variations of the means by limiting the domain of stationarity of both primary and secondary means (both unknown) to the local neighbourhood $W(u)$. The OCK estimator can be written as:

$$Z_{OCK}^{(1)*}(\mathbf{u}) = \sum_{i=1}^{N_v} \sum_{\alpha_i=1}^{n_i(\mathbf{u})} \lambda_{\alpha_i}^{OCK}(\mathbf{u}) Z_i(\mathbf{u}_{\alpha_i}) \qquad (31)$$

And the OCK system is:

$$\begin{cases} \sum_{j=1}^{N_v} \sum_{\beta_{ji}=1}^{n_j(\mathbf{u})} \lambda_{\beta_j}^{OCK}(\mathbf{u}) C_{ij}(\mathbf{u}_{\alpha_i} - \mathbf{u}_{\beta_j}) + \mu_i^{OCK}(\mathbf{u}) = C_{i1}(\mathbf{u}_{\alpha_i} - \mathbf{u}) \\ \sum_{\beta_i=1}^{n_i} \lambda_{\beta_i}^{OCK}(\mathbf{u}) = \delta_{i1} \end{cases} \quad \begin{array}{l} \alpha_i = 1,\ldots,n_i(\mathbf{u}) \\ i = 1,\ldots,N_v \end{array} \qquad (32)$$

with $\delta_{i1} = 1$ for $i = 1$ and $\delta_{i1} = 0$ otherwise.

c) **Universal Cokriging (UCK)**, or Cokriging with trend models, consists on modeling the trend components as linear combinations of known functions $f_{ki}(\mathbf{u})$ of the spatial coordinates $\mathbf{u}$:

$$m_i(\mathbf{u'}) = \sum_{k=0}^{K} a_{ki}(\mathbf{u'}) f_{ki}(\mathbf{u'}), \quad \text{with } a_{ki}(\mathbf{u'}) \approx a_{ki} \quad \forall \mathbf{u'} \in W(\mathbf{u}) \qquad (33)$$

The trend coefficients $a_{ki}(\mathbf{u'})$ are constant but unknown parameters within each local neighbourhood $W(\mathbf{u})$. The case of $K = 0$ corresponds to the OCK.

The UCK system can be derived from equations (22) and (33).

Note: An important advantage of the Cokriging estimator is that its error variance is always smaller than or equal to that corresponding to the Kriging estimator, which ignores secondary information [4].

4. RESULTS: INTERPOLATED RAINFALL MAPS

In this section, the methodology followed to achieve the previous explained interpolation approaches is related and some remarks regarding to the interpolated maps are highlighted. In the appendix a total of nine “selected” interpolated maps (see the reason in section 5.1.) is shown.

4.1. IDW and LPI

The IDW as well as the LPI processes were achieved with the ArcMap v.8.3 program [7]. Few decisions should be made regarding to the model parameters. In the case of IDW, the power p of the distance in the weighting function of equation (1); in the case of LPI, the order n of the polynomial of equation (2). The neighbourhood should be also specified. Some tests were made and three different models for each interpolation procedure were calculated:

1) IDW with a neighbourhood of 10 points (2 at least):
   - IDW1: IDW with $p=1$
   - IDW2: IDW with $p=2$
   - IDW3: IDW with $p=3$
2) LPI with a neighbourhood of the 50% (43 points):
   - LPI1: LPI with $n=1$
   - LPI2: LPI with $n=2$
   - LPI3: LPI with $n=3$

From the plotted IDW maps, some characteristics stands out:

1) The "hull effect" (circles around sample data) is observed.
2) The transition between areas with different precipitation values is not achieved with smooth curves.

From the LPI, the remarkable features are:

1) In general, the resulting maps are similar to that obtained by the IDW interpolation, but the transition between different precipitation values is smoother and the hull effect is no longer observed.
2) In the SUM map, whose sample data have low standard deviation (table 1) and present a clear SE-NW precipitation gradient, the interpolated surface shows a very smooth behaviour, growing from SE to NW and rejecting local effects. At the NW part, where there is an absence of observatories, the polynomial continues growing and the obtained values overcome the expected ones.

4.2. Multiple Regression

The multiple regression procedure was carried out by using the software "STATGRAPHICS 4.0". To determine which secondary variables should be used in the multiple regression interpolation, the stepwise approach was followed. Within this approach, initially all the 36 secondary variables are considered and then, they are eliminated one by one based on the outset criteria, the P-value. As described in [13], the P-value is a component of the ANOVA (Analysis of Variance) table that serves as a measure of significance. When the P-value is less than 0.05, there is a statistically significant relationship between two variables at the 95 % confidence level.

Within the multiple regression procedure, a coefficient of determination R^2 can be calculated as a measure of the goodness of the model fitting. This factor represents the proportion of the variation of the dependent variable explained by the regression model [9]. The obtained results can be seen in the following two tables, where three different variants/cases were computed.

Rainfall var.	***MR0***	***MR1***	***MR2***
ANN	86.68 %	83.45 % (17)	79.72 % (13)
SPR	87.31 %	81.25 % (13)	79.72 % (12)
Log(SUM)	89.53 %	84.83 % (9)	79.50 % (4)
AUT	87.53 %	86.01 % (20)	80.18 % (13)
Log(WIN)	86.68 %	84.00 % (19)	79.54 % (14)

Table 3. R^2 percentages obtained in the MR process with **MR0**: taking into consideration all the proposed variables (see table 11); **MR1**: a regression with the variables having a P-value<0.05; **MR2**: a regression with the minimal variables needed to obtain a R^2 coefficient rounding the 80 %. In the last two cases, the number of variables are included in brackets.

In the following, the MR0 will no be taken into consideration because most of the great number of variables (36) are not statistically significant at the 90% or higher confidence level.

In the next table, the secondary variables used in the MR1 and MR2 procedures are shown. These variables appear in order according to a major weight in the regression equation (see the regression equations in the analysis reports 6 to 15 in the appendix).

Rainfall variable	*Secondary variables*
ANN	**C_E**, D5000, Y, X, **Z5000_S5000**, **S5000_D5000**, COAST_S5000, **D50002**, X_S5000, COAST_Z5000, X_Z5000, Y_D5000, Y_Z5000, COAST2, X2, Y2, Y_COAST
SPR	**C_E**, D5000, Y, Z5000_S5000, COAST_S5000, X_S5000, COAST_Z5000, Y_D5000, Y_Z5000, COAST2, Y2, X2, Y_COAST
Log(SUM)	**C_E**, **C_NE**, **C_SE**, D5000, COAST, Y_D5000, **COAST2**, **X_COAST**, X_Y
AUT	**C_E**, D5000, S50002, Y, **COAST**, **X**, **S5000_D5000**, **D50002**, COAST_S5000, **X_S5000**, COAST_Z5000, X_Z5000, **COAST_D5000**, Y_Z5000, Y_D5000, COAST2, X_COAST, Y_COAST, X_Y, Y2

Log(WIN)	S5000, **C_N**, D5000, **Z5000**, **COAST**, Y, X, Z5000_S5000, **S5000_D5000**, COAST_S5000, **D50002**, COAST_Z5000, X_Z5000, Y_Z5000, Y_D5000, COAST2, X2, Y_COAST, Y2

Table 4. Secondary variables used in the MR1 rainfall data interpolation; the variables in bold type, are the ones excluded in the MR2 procedure.

Note that six variables appear in all rainfall interpolations (except for SUM): Y, D5000, Y2, COAST2, Y_D5000 and COAST_S5000.

As it can be seen, the variables concerning to scalar products are high weighted, but have also great P-values (see analysis reports 6 to 10) and therefore are eliminated in the MR2 approach (e.g. C_E in SPR), between others. The more simplicity of the MR2 equations do significantly reduce the effort made in the achievement of the secondary variables. Therefore, MR2 could be proposed because of being more efficient. Nevertheless, we will compute the rainfall maps for the MR1 (from now called simply MR), as we have already achieved all the secondary variables.

It is supposed that the variables included in a single regression equation are not high correlated with each other. Nevertheless, this fact was not seen in our case (see table 13 in the appendix). For example, in the Log(SUM) equation, the variables Y_D5000 and D5000 as well as COAST, COAST2, X_COAST and X_Y are high correlated with each other, reaching even a 1.00 correlation factor between COAST and X_COAST.

To compute the maps, a grid with a 1x1 Km step (almost 32.000 grid points) was defined for all the study area for which all the secondary variables appearing in the MR equation were calculated. To achieve these calculations, the MDE was rescaled to a pixel size of 1 Km due to the huge data involved. From the visual check of the resulting MR rainfall maps, some remarks stand out:

1) The strong correlation with the topographic and the geographic variables (compare with figure 1).
2) Local behaviour and more spatial diversity. This is due to the fact that, within the MR approach, the rainfall is derived only from the secondary variables at a single location, regardless of the records at the neighbouring observatories (see section 3.3.).
3) Except for the SUM map, too high rainfall values (specially in the maps for AUT and WIN) do appear in areas where there is both a lack of observatories and high sample data (compare with map of figure 1). In the WIN map a value of 660mm is reached for one of the pixels in the northern area, representing

almost 3 times the maximum value of sample data (see table 1). This can be explained by the spatial situation of the observatories, with only 5 of them situated at an elevation reaching the 800 m height. This makes that the regression equation is not able to well describe the precipitation at elevated areas.

4.3. Kriging and Cokriging approaches

As it was explained in sections 3.4. and 3.5., before achieving a Kriging and/or Cokriging interpolation, a geostatistical analysis of the data should be made in order to compute a model of the theoretical semivariogram. This first step was carried out with the geostatistical program "VARIOWIN 2.2" [11]. Afterwards, the interpolation procedures were carried out with the program "ArcMap 8.3" of ESRI. In the following sections, the methodology and obtained results are described.

4.3.1. Previous geostatistical analysis

To identify the main directions of anisotropies of the spatial behaviour of the rainfall variables, the variogram surfaces were computed (figure 4). Within these graphs, an aligned set of "blue" pixels indicates the directions of maximum continuity (low $\gamma(\mathbf{h})$ values); the direction of minimum continuity is the perpendicular to that of the maximum continuity. In the following table, these main directions and the middle direction between them are shown:

	ANN	**SPR**	**Log(SUM)**	**AUT**	**Log(WIN)**
Max. cont.	135º	100º	45º	150º	135º
Min. cont.	45º	10º	135º	60º	45º
Middle	90º	55º	90º	105º	90º

Table 5. Directions of max./min. continuity and the middle direction between them.

Once these directions were detected, the semivariogram models were calculated by achieving the best possible fit for these three directions. An angular tolerance of 30º was considered. Additionally, an omni-directional model (tolerance = 90º) was also computed. Figures 5 to 9 show the graphical results. As it can be seen, a total of 10 lags were considered, with a lag spacing of 15 Km. This means that the models are computed for a maximum distance of 10x15=150 Km, representing half of the major distance in the total area.

The computed semivariogram models for each one of the rainfall variables were:

Variable	Computed semivariogram
ANN	γ (h) = 5320 + 15580 Gauss.72000 (h)
SPR	γ (h) = 242 + 792 Gauss.83500 (h) + 271.6267 Gauss.16500 (h)
Log(SUM)	γ (h) = 0.006 + 0.09199333 Exp.142500 (h) + 0.03 Exp.142500 (h)
AUT	γ (h) = 704 + 2528 Gauss.76500 (h)
Log(WIN)	γ (h) = 0.026 + 0.124 Gauss.67500 (h)

Table 6. Computed semivariogram models. Where the equations are of the form:
$\gamma(h)$ = nugget effect + (partial sill)$_1$ · Model$_1$.range$_1$(h) + (partial sill)$_2$ · Model$_2$.range$_2$(h).

For the ANN and Log(WIN) variables, geometric anisotropies were considered (with 1.75 and 1.5 factors, respectively). For the SPR and Log(SUM) variables, zonal anisotropies were considered. The AUT variable was considered to be isotropic.

As it can be seen, the semivariograms for all the variables were modeled with Gaussian models except for the SUM variable, for which the sum of two exponential models gave the better fit. The semivariograms were fitted with the objective of capturing the major spatial features, since the objective is not to model all the details of the sample semivariograms [4].

4.3.2. Interpolations

To achieve a Kriging or a Cokriging interpolation in ArcMap, several decisions must be taken (apart from that concerning to the semivariogram model). The explored methodologies were:

1) Ordinary Kriging/Cokriging (OK/OCK)
2) Simple Kriging/Cokriging (SK/SCK)
3) Universal Kriging/Cokriging (UK/UCK)

For some of them, the "order of the trend removal" can be chosen (see [7]). The possible combinations are:

1) OK/OCK: none (n), constant (c), first (f), second (s) and third (t) order.
2) UK/UCK: constant (c), first (f), second (s) and third (t) order.

To distinguish which order for the trend is used, the letter in brackets will be placed after the interpolation procedure. For example, the "OCKf" interpolation, means an ordinary Cokriging interpolation with a first order trend component. A total of 20 combinations are possible (in fact only 19, because "OKn" is equal to "UKc"). Interpolations were achieved for all these combinations.

In a Cokriging interpolation, secondary data is needed. In ArcMap, a maximum of three secondary variables can be used. After some tests with several secondary variables (see table 11) and combinations of them, it was seen that the best results were achieved by considering the predicted values of MR as a "single secondary topographical variable". This have quite sense: the correlation between the rainfall observations and the predicted values by the MR ranges from 0.90 to 0.93; comparing these values with the ones shown in table 12, it can be seen that the MR is the "variable" with the higher correlation coefficient.

From the calculated K/CK rainfall maps (in the appendix), some remarks are noted:

1) Global behaviour; local variations are not appreciated.
2) SK, OK, and UK maps are similar to SCK, OCK and UCK maps, respectively. The difference is that the seconds do show a stronger correlation with the topography (compare with figure 1).
3) In general, the UK maps show a strange behaviour in the upper western area, where the is an absence of observatories and high precipitation values.
4) In the case of SPR maps, the so called "hull effect" is observed. This effect is characterized for the IDW interpolations, but it should be remarked that the difference between IDW and Kriging interpolators is the weighting function. Therefore, the form of the semivariogram model can produce this effect in the interpolated maps.

5. DISCUSSION

From a single visual check of the plotted rainfall maps, the election of the "best" model for each rainfall variable is not an easy work, and some statistics should be computed in order to evaluate and compare them. To check the reliability of the models and select the best one for each rainfall variable, some steps were followed:

1) Some statistical criteria were computed to assess the agreement of the models. In this step, the most relevant models/interpolations were selected.

2) Some graphics with the observed versus predicted values were computed.
3) Error maps were produced and compared with the elevation and slope variables.
4) An error statistical analysis was performed.
5) The study area was divided into eight different regions (according to the terrain elevation). For these regions, the three "best" maps for each rainfall variable were compared.

5.1. Agreement of the models

To validate the obtained maps, the following statistical criteria have been used (from [14]):

Correlation coefficient (r^2)	
Mean bias error (*MBE*)	$MBE = \frac{1}{N}\sum_{i=1}^{N}(P_i - O_i)$
Root-mean-square error (*RMSE*)	$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(P_i - O_i)^2}$
Mean absolute error (*MAE*)	$MAE = \frac{1}{N}\sum_{i=1}^{N}\lvert P_i - O_i\rvert$
Model efficiency (*EF*)	$EF = 1 - \frac{\sum_{i=1}^{N}(P_i - O_i)^2}{\sum_{i=1}^{N}(\overline{O} - O_i)^2}$
Willmott's *D*	$D = 1 - \frac{\sum_{i=1}^{N}(P_i - O_i)^2}{\sum_{i=1}^{N}(\lvert P_i\rvert + \lvert O_i\rvert)^2}$

Table 7. Statistical criteria used to compare the different interpolation techniques. Where N : no. of observations; O : observed value; $\overline{O}$: mean of observed values; P : predicted value.

The correlation coefficient r^2 is a quantity that gives the quality of a least squares fitting to the original data, and therefore can be included as a first calculation of the reliability of the model. Nevertheless, the relationship between r^2 and model performance is not well defined, and the magnitudes of r^2 are not consistently related to the accuracy of prediction. Therefore, some other statistics were used, such as *MBE*, *RMSE* and *MAE*, these two last considered as the “best” overall measures of model performance, as they summarize the mean difference in the units of the observed and predicted data. The difference between them is that *RMSE* places a lot of weight on high errors, meanwhile *MAE* is less sensitive to extreme values. Two other accuracy measurements were calculated: the *EF* and the Willmott’s *D*. A value of *EF* close to zero, indicates that the mean value of the observations is more reliable than the predictions, and therefore the model is not reliable. Willmott’s *D* scales with the magnitude of the variables, retains mean information and does not amplify outliers [14].

Due to the lack of data, the method of cross validation was employed to calculate the predicted values for all the “exact” interpolation methods and for the LPI. In the case of MR, the predicted values were the calculated with the regression equation. Therefore, the results for MR are expected to be a little more optimistic (not so reliable). The computed values for all the methods and climatic variables are shown in table 14 (in the appendix).

As it can be seen from the table, there is no (a significant) improvement for the methods when increasing the order of the polynomials or the trend models. Hence, the methods with the lower order will be “selected” and considered from now until the end of the work. For these selected methods, some graphics were derived to better compare the results (figures 10 to 15). From these figures, some conclusions are derived:

1) In general, the better interpolation methods seem to be: MR, OCKn and SCK.
2) The worse results are achieved by IDW1 and LPI1 methods.
3) In general, the rainfall variable better described by the models is AUT.

5.2. Observed vs. predicted values

The similarity between observed and predicted values is described in figures 16 to 20. The straight line indicates absolute coincidence between both values. Therefore, the more points are concentrated around this line, the more accurate is the model; the more point dispersion, the less accuracy is achieved. The correlation coefficient is shown together with the graphs as a good measure of the predicted-observed values similarity. The visual check of these graphs do corroborate the fact that the more overall reliable models are OCKn, SCK and MR.

5.3. Error maps

To see the spatial distribution of high and/or low errors, some error maps were plotted (see appendix). These errors are compared with the areas of great elevation (>500 m) and/or slope (>10º). From these maps, some conclusions can be made:

1) In general, there are two areas with low errors: the central eastern and the lower part. These both areas are quite dry (except in AUT) and are situated neither in high elevations nor in great slopes.
2) In general, the greater errors occur in the coastal peak, a high elevated area and/or with high slope values.
3) The models for the variable SUM have lots of points with low errors and few with high values. This can be an indicator of a non-normality error distribution.

5.4. Error statistical analysis

If the fitted model is appropriate, the errors should be normally distributed, and should not correlate with each other or with the independent variables [9].

To check the errors normality, some histograms were plotted and the skewness and kurtosis statistical parameters were calculated (figures 21 to 25). The standardized skewness test looks for lack of symmetry in the data. The standardized kurtosis test looks for distributional shape which is either flatter or more peaked than the normal distribution [13]. Values between [-2,2] indicates normality of data. As it can be seen, the better results are achieved for the AUT and Log(WIN) models, whose errors can be considered to be normal distributed for all the models. The SPR models give also good results, as only the errors of the IDW1 interpolation are not inside the desired values for the kurtosis test. The opposite results are given by the Log(SUM) and ANN models. The Log(SUM) errors pass the skewness tests (except for UKf), but the kurtosis test is only valid for the models MR and IDW1. In the case of ANN, errors for the IDW1, OKn, SK and UKf do not follow a normal distribution.

In order to look for systematic patterns between errors and observed/predicted values, the correlation coefficients between errors and observed values as well as between the predicted ones were calculated. The obtained results are seen in the following tables:

	OKn	SK	UKf	OCKn	SCK	UCKc	MR	IDW1	LPI1
ANN	0,25	0,23	0,17	0,17	0,18	0,23	0,27	0,36	0,33
SPR	0,36	0,37	0,32	0,32	0,24	0,37	0,22	0,39	0,40
Log(SUM)	0,06	0,07	-0,04	0,06	0,00	0,04	-0,06	0,00	-0,10
AUT	0,25	0,21	0,24	0,23	0,21	0,25	0,20	0,26	0,27
Log(WIN)	-0,14	-0,19	-0,13	-0,16	-0,12	-0,15	0,02	0,03	-0,04

Table 8. Correlation coefficients between residuals (|P-O|) and observed values.

	OKn	SK	UKf	OCKn	SCK	UCKc	MR	IDW1	LPI1
ANN	0,25	0,20	0,19	0,19	0,21	0,23	0,26	0,38	0,24
SPR	0,32	0,33	0,31	0,32	0,28	0,33	0,29	0,38	0,30
Log(SUM)	0,18	0,17	0,12	0,16	0,12	0,16	0,00	0,05	0,06
AUT	0,20	0,18	0,23	0,16	0,18	0,20	0,09	0,31	0,19
Log(WIN)	-0,12	-0,12	-0,19	-0,13	-0,06	-0,14	0,01	0,12	-0,07

Table 9. Correlation coefficients between residuals (|P-O|) and predicted values.

Low correlation coefficients indicates that systematic patterns between errors and observed values as well as with errors and predicted values do not exist.

5.5. Regional analysis

After the previous analysis, it can be stated that the best overall models are MR, SCK and OCKn (except for SUM, whose models seems to present serious problems). At this point, these three models were selected and compared between them. The aim of this comparison was to look for the most "robust" (of the defined reliable) model, i.e., the most reliable model for some different defined zones or regions was searched. The chosen zone indicator was the elevation and eight different zones were defined:

Zone	***Nr. observatories***	***Min. elevation***	***Max. elevation***	***Mean elevation***
1	11	6,03	31,60	19,85
2	11	34,50	58,71	43,91
3	11	60,83	183,33	102,90
4	11	186,87	326,53	249,09
5	11	373,06	482,26	435,23
6	11	523,21	616,27	558,24
7	11	626,96	708,06	664,01
8	9	719,27	1031,97	830,87

Table 10. Description of the eight different zones.

The selected comparison criteria between models was the MAE coefficient, and graphs with MAE ratios between two interpolation techniques versus the mean elevation of each zone were plotted (figure 26). A value will be above/under the horizontal line (at Y=1) depending which of the MAE coefficients is greater. MAE ratios greater than one indicates that the numerator MAE is greater than that of the denominator, and therefore the model corresponding to the numerator MAE is "less" reliable for that zone; on the contrary, if the ratio is less than one, that model is "more" reliable. From the graphs it can be said that:

1) From the graphs comparing MAE coefficients of SCK and OCKn, it is clearly seen that SCK is in general more reliable for ANN and SUM; in the case of WIN, the better one is the OCKn. In the case of AUT, OCKn is more reliable in low areas whereas SCK works better in elevated areas. Finally, in the case of SPR it is no clear which one is the best because both present better results at four areas, but independently from the elevation.
2) From the rest of the graphs, where the MAE coefficient of MR is compared with that corresponding to the other two models, it is clearly seen that MR is more robust than any of the other models. There are two exceptions: for the SUM variable, SCK seems to be comparable to MR; for the SPR variable, the decision is once more complicated.

6. CONCLUSIONS

After all the developed analysis, some conclusions regarding to the interpolation methods as well as to the precipitation achieved models can be made.

From the evaluated interpolation methods, it can be concluded that:

1) **IDW** and **LPI** methods are not well suited to perform a rainfall interpolation.
2) The **MR** interpolation give the overall best statistical results. Nevertheless, it seems from the visual check of the rainfall maps that the interpolation is not well performed at elevated areas. Therefore, to asses that this method is the best one, it should be tested with additional sample data at those locations.
3) The most reliable **geostatistical** methods are SCK and OCKn, which includes as a secondary variable the results achieved by the MR interpolation. After the regional analysis, it was seen than SCK is more appropriated for ANN and SUM,

meanwhile OCKn is better suited to interpolate the WIN variable. For the variables AUT and SPR, both of them give similar results.

From the precipitation models:

1) The **ANN** variable is well described by the MR, SCK and OCKn methods, being the most reliable MR (with the explained limitations in elevated areas) followed by SCK.
2) The **SPR** variable is better described by the MR, SCK and OCKn methods (being difficult to choose the most reliable one), but it should be taken into account that the "hull effect" is observed for the SCK and OCKn approaches.
3) For the **SUM** variable, only the MR approach, whose errors follow a normal distribution, seems to be reliable.
4) The **AUT** variable is well described by the MR, SCK and OCKn methods. The best statistical results were obtained by MR, but from a visual check of the resulting map, it could be rejected in elevated areas (sample comparison should be made). Both of the geostatistical methods give similar results, being OCKn more reliable in low areas and SCK in elevated areas.
5) The **WIN** variable is well described by the MR, SCK and OCKn methods, being the most reliable MR followed by OCKn. Nevertheless, from a visual check of the MR map, it could be rejected in elevated areas.

7. FURTHER WORK

Further research should investigate in two areas: Firstly, in the acquisition (if possible) of sample data at elevated regions, that would indeed improve the obtained multiple regression equation thus increasing the reliability for all the proposed models. This acquisition would be necessary to asses that the MR is the overall best precipitation model. Secondly, the obtained models could be tested with different temporal scales, for instance monthly or even daily. The analysis of daily sample data would be of special interest for the Comunidad Valenciana in order to predict storms and possible floods.

8. REFERENCES

[1] ArcView GIS : the geographic information system for everyone. - New York : ESRI, cop. 1996.

[2] DEUTSCH, C.V., JOURNEL, A.G., 1998. GSLIB: Geostatistical Software Library and User's Guide. Oxford University Press (New York).

[3] DUTTER, R., 2001. Geostatistics. Manuscript. Vienna University of Technology.

[4] GOOVAERTS, P., 1997. Geostatistics for natural resources evaluation. Oxford University Press (New York).

[5] GOOVAERTS, P., 2000. Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. Journal of Hydrology. 228, 113-129.

[6] JOHANSSON, B., CHEN, D., 2003. The influence of wind and topography on precipitation distribution in Sweden: statistical analysis and modelling. International Journal of Climatology. 23, 1523-1535.

[7] JOHNSTON, K., VER HOEF, J.M., KRIVORUCHKO, K., LUCAS, N., 1997. Using ArcGIS geostatistical analyst : [ArcGIS 8]. Redlands, CA : ESRI.

[8] KRAUS, K., 2000. Photogrammetrie, Band 3: Topographische Informations-systeme. Ed. Dümmler.

[9] MARQUÍNEZ, J., LASTRA, J., GARCÍA, P., 2003. Estimation models for precipitation in mountainous regions: the use of GIS and multivariate analysis. Journal of Hydrology. 270, 1-11.

[10] NINYEROLA, M., 2000. Modelització climàtica mitjançant tècniques SIG i la seva aplicació a l'anàlisi quantitativa de la distribució d'espècies vegetals a l'Espanya peninsular. Tesi doctoral (capítol 3). Universitat Autònoma de Barcelona. Web page: http://magno.uab.es/atles-climatic/cap3_mod_per_a_pdf_tot.pdf (consulted on 22/04/04).

[11] PANNATIER, I., 1996. VARIOWIN. Software for spatial data analysis in 2D. Ed. Springer New York.

[12] PEREZ CUEVA, A. J., 1994. Atlas climático de la comunidad valenciana. Serie: Publicaciones de divulgación técnica. Colección: "Territori". Nr. 4. Ed. Generalitat Valenciana.

[13] Statgraphics plus : [versión 4]. - Standard ed. - Cambridge : Manugistics, cop. 1998.

[14] VICENTE-SERRANO, S.M., SAZ-SÁNCHEZ, M.A., CUADRAT, J.M., 2003. Comparative analysis of interpolation methods in the middle Ebro Valley (Spain): application to annual precipitation and temperature. Climate research. 24, 161-180.

9. APPENDIX

Tables 11 to 14
Figures 4 to 26
Analysis reports 1 to 15
Rainfall maps for ANN, SPR, SUM, AUT and WIN
Error maps for ANN, SPR, SUM, AUT and WIN

Table 11.

Secondary variables

Variables	*Meaning*
X	Longitude
Y	Latitude
COAST	Minimal distance to the Mediterranean Sea
Z5000	Mean elevation within a radius of 5 Km around the observatory
S5000	Mean slope within a radius of 5 Km around the observatory
D5000	Elevation difference within a radius of 5 Km around the observatory
X2	X^2
Y2	Y^2
COAST2	$COAST^2$
Z50002	$Z5000^2$
S50002	$S5000^2$
D50002	$D5000^2$
X_Y	X x Y
X_COAST	X x DCOAST
X_Z5000	X x Z5000
X_S5000	X x S5000
X_D5000	X x D5000
Y_COAST	Y x DCOAST
Y_Z5000	Y x Z5000
Y_S5000	Y x S5000
Y_D5000	Y x D5000
COAST_Z5000	COAST x Z5000
COAST _S5000	COAST x S5000
COAST _D5000	COAST x D5000
D5000_S5000	D5000 x S5000
D5000_D5000	D5000 x D5000
S5000_D5000	S5000 x D5000
A_C	Scalar product between the aspect vector and the vector from the observatory to the direction of minimal distance to the coast
C_N	Scalar product between the vector from the observatory to the direction of minimal distance to the coast and a vector in the N direction
C_NE	Scalar product between the vector from the observatory to the direction of minimal distance to the coast and a vector in the NE direction
C_E	Scalar product between the vector from the observatory to the direction of minimal distance to the coast and a vector in the E direction
C_SE	Scalar product between the vector from the observatory to the direction of minimal distance to the coast and a vector in the SE direction
A_N	Scalar product between the aspect vector and a vector in the N direction
A_NE	Scalar product between the aspect vector and a vector in the NE direction
A_E	Scalar product between the aspect vector and a vector in the E direction
A_SE	Scalar product between the aspect vector and a vector in the SE direction

Table 12.

Correlation coefficients between rainfall and secondary variables.

Values equal or greater than 0.5 are highlighted.

	ANN	*SPR*	*SUM*	*AUT*	*WIN*
X	0,50	0,40	0,04	0,61	0,42
Y	0,40	0,34	0,81	0,39	0,09
COAST	-0,09	0,05	0,39	-0,31	-0,12
Z5000	0,10	0,28	0,50	-0,15	0,04
S5000	0,46	0,57	0,31	0,30	0,44
D5000	0,42	0,56	0,20	0,27	0,44
X2	0,50	0,41	0,06	0,61	0,42
Y2	0,40	0,34	0,81	0,39	0,09
COAST2	-0,13	-0,02	0,34	-0,32	-0,14
Z50002	0,02	0,19	0,49	-0,21	-0,04
S50002	0,47	0,56	0,24	0,34	0,48
D50002	0,40	0,52	0,09	0,28	0,45
X_Y	0,56	0,46	0,32	0,66	0,39
X_COAST	-0,07	0,08	0,41	-0,29	-0,10
X_Z5000	0,13	0,32	0,51	-0,12	0,07
X_S5000	0,49	0,60	0,30	0,33	0,47
X_D5000	0,46	0,58	0,19	0,31	0,47
Y_COAST	-0,09	0,05	0,40	-0,30	-0,12
Y_Z5000	0,10	0,28	0,52	-0,15	0,04
Y_S5000	0,46	0,58	0,33	0,30	0,44
Y_D5000	0,43	0,56	0,22	0,28	0,44
COAST_Z5000	-0,09	0,05	0,44	-0,32	-0,13
COAST_S5000	0,06	0,18	0,41	-0,14	0,04
COAST_D5000	0,07	0,22	0,40	-0,16	0,05
Z5000_S5000	0,29	0,43	0,45	0,07	0,25
Z5000_D5000	0,26	0,42	0,38	0,04	0,23
S5000_D5000	0,45	0,56	0,17	0,32	0,48
A_C	0,23	0,21	0,26	0,26	0,09
A_N	0,09	0,11	-0,19	0,05	0,22
A_NE	0,22	0,23	0,14	0,20	0,19
A_E	0,17	0,17	0,28	0,17	0,04
A_SE	0,08	0,07	0,29	0,10	-0,08
C_N	0,23	0,18	-0,04	0,25	0,27
C_NE	0,12	0,07	0,03	0,15	0,12
C_E	-0,20	-0,22	0,14	-0,17	-0,29
C_SE	-0,30	-0,27	0,10	-0,31	-0,38

Table 13.

Correlation coefficients between the secondary variables.

Values equal or greater than 0.5 are highlighted.

	X	Y	COAST	Z5000	S5000	D5000	X2	Y2	COAST2	Z50002	S50002	D50002	X_Y	X_COAST	X_Z5000	X_S5000	X_D5000	Y_COAST
X	1,00	0,23	-0,79	-0,43	0,01	0,11	1,00	0,23	-0,76	-0,44	0,11	0,17	0,94	-0,78	-0,37	0,08	0,17	-0,79
Y		1,00	0,17	0,20	0,10	-0,04	0,24	1,00	0,18	0,23	0,08	-0,09	0,54	0,18	0,20	0,10	-0,03	0,18
COAST			1,00	0,75	0,28	0,17	-0,78	0,17	0,94	0,73	0,14	0,06	-0,63	1,00	0,71	0,22	0,11	1,00
Z5000				1,00	0,64	0,57	-0,42	0,20	0,64	0,95	0,53	0,48	-0,30	0,77	1,00	0,61	0,54	0,76
S5000					1,00	0,90	0,02	0,09	0,19	0,47	0,96	0,84	0,04	0,30	0,67	1,00	0,89	0,28
D5000						1,00	0,11	-0,04	0,05	0,42	0,85	0,96	0,07	0,19	0,60	0,90	1,00	0,16
X2							1,00	0,24	-0,75	-0,43	0,11	0,17	0,95	-0,77	-0,36	0,08	0,17	-0,78
Y2								1,00	0,17	0,23	0,08	-0,09	0,54	0,18	0,20	0,10	-0,03	0,18
COAST2									1,00	0,65	0,08	-0,03	-0,60	0,93	0,59	0,13	0,00	0,95
Z50002										1,00	0,37	0,34	-0,30	0,74	0,95	0,44	0,38	0,73
S50002											1,00	0,83	0,12	0,15	0,56	0,97	0,85	0,14
D50002												1,00	0,11	0,08	0,51	0,85	0,96	0,05
X_Y													1,00	-0,61	-0,25	0,10	0,13	-0,62
X_COAST														1,00	0,73	0,24	0,13	1,00
X_Z5000															1,00	0,65	0,57	0,71
X_S5000																1,00	0,90	0,22
X_D5000																	1,00	0,11
Y_COAST																		1,00
Y_Z5000																		
Y_S5000																		
Y_D5000																		
COAST_Z5000																		
COAST_S5000																		
COAST_D5000																		
Z5000_S5000																		
Z5000_D5000																		
S5000_D5000																		
A_C																		
A_N																		
A_NE																		
A_E																		
A_SE																		
C_N																		
C_NE																		
C_E																		
C_SE																		

Table 13 (...continuation).

Correlation coefficients between the secondary variables.

Values equal or greater than 0.5 are highlighted.

	Y_Z5000	Y_S5000	Y_D5000	COAST_Z5000	COAST_S5000	COAST_D5000	Z5000_S5000	Z5000_D5000	S5000_D5000
X	-0,42	0,01	0,11	-0,71	-0,61	-0,59	-0,18	-0,14	0,15
Y	0,22	0,12	-0,01	0,22	0,21	0,14	0,19	0,11	0,00
COAST	0,75	0,29	0,17	0,95	0,86	0,85	0,50	0,46	0,10
Z5000	1,00	0,65	0,58	0,84	0,75	0,81	0,87	0,87	0,53
S5000	0,64	1,00	0,90	0,29	0,57	0,57	0,87	0,82	0,94
D5000	0,56	0,89	1,00	0,19	0,43	0,54	0,78	0,83	0,94
X2	-0,42	0,02	0,11	-0,69	-0,60	-0,58	-0,17	-0,14	0,15
Y2	0,22	0,12	-0,01	0,22	0,20	0,14	0,19	0,11	0,00
COAST2	0,65	0,19	0,06	0,93	0,81	0,74	0,42	0,34	0,03
Z50002	0,96	0,48	0,43	0,87	0,69	0,75	0,79	0,80	0,38
S50002	0,53	0,96	0,85	0,17	0,46	0,43	0,84	0,76	0,96
D50002	0,47	0,83	0,96	0,10	0,32	0,44	0,72	0,80	0,95
X_Y	-0,29	0,05	0,09	-0,54	-0,46	-0,46	-0,09	-0,09	0,12
X_COAST	0,77	0,30	0,19	0,94	0,87	0,86	0,52	0,48	0,12
X_Z5000	1,00	0,68	0,61	0,80	0,73	0,79	0,89	0,89	0,57
X_S5000	0,61	1,00	0,91	0,24	0,51	0,51	0,86	0,81	0,95
X_D5000	0,53	0,88	1,00	0,14	0,37	0,48	0,76	0,81	0,94
Y_COAST	0,76	0,29	0,17	0,95	0,87	0,85	0,51	0,46	0,10
Y_Z5000	1,00	0,65	0,57	0,84	0,76	0,81	0,87	0,87	0,53
Y_S5000		1,00	0,90	0,30	0,58	0,57	0,88	0,82	0,94
Y_D5000			1,00	0,20	0,44	0,54	0,79	0,84	0,94
COAST_Z5000				1,00	0,83	0,82	0,59	0,55	0,14
COAST_S5000					1,00	0,94	0,73	0,65	0,41
COAST_D5000						1,00	0,74	0,75	0,45
Z5000_S5000							1,00	0,96	0,82
Z5000_D5000								1,00	0,82
S5000_D5000									1,00
A_C									
A_N									
A_NE									
A_E									
A_SE									
C_N									
C_NE									
C_E									
C_SE									

Table 13 (…continuation).

Correlation coefficients between the secondary variables.

Values equal or greater than 0.5 are highlighted.

	A_C	A_N	A_NE	A_E	A_SE	C_N	C_NE	C_E	C_SE
X	0,31	-0,26	0,14	0,34	0,38	-0,03	-0,20	-0,43	-0,16
Y	0,33	-0,27	0,12	0,31	0,36	-0,01	0,09	0,22	0,11
COAST	-0,13	0,22	0,05	-0,11	-0,20	0,09	0,19	0,26	0,03
Z5000	-0,13	0,14	0,06	-0,04	-0,11	-0,18	-0,16	-0,02	0,16
S5000	-0,02	0,20	0,15	0,02	-0,09	-0,14	-0,20	-0,19	0,04
D5000	-0,01	0,23	0,19	0,03	-0,10	-0,12	-0,22	-0,29	-0,02
X2	0,31	-0,27	0,14	0,34	0,38	-0,03	-0,20	-0,43	-0,17
Y2	0,33	-0,27	0,12	0,32	0,37	-0,01	0,08	0,22	0,11
COAST2	-0,16	0,17	-0,02	-0,14	-0,19	0,10	0,18	0,24	0,02
Z50002	-0,16	0,13	0,02	-0,07	-0,12	-0,18	-0,15	0,02	0,18
S50002	-0,05	0,16	0,08	-0,04	-0,11	-0,15	-0,21	-0,19	0,05
D50002	-0,05	0,22	0,12	-0,04	-0,14	-0,12	-0,23	-0,29	-0,01
X_Y	0,39	-0,32	0,16	0,40	0,45	-0,03	-0,14	-0,29	-0,10
X_COAST	-0,12	0,22	0,06	-0,10	-0,19	0,09	0,19	0,26	0,03
X_Z5000	-0,12	0,13	0,07	-0,03	-0,09	-0,19	-0,18	-0,05	0,16
X_S5000	-0,01	0,18	0,15	0,03	-0,07	-0,14	-0,22	-0,22	0,03
X_D5000	0,00	0,21	0,18	0,04	-0,08	-0,12	-0,23	-0,32	-0,03
Y_COAST	-0,13	0,22	0,05	-0,11	-0,19	0,09	0,19	0,26	0,04
Y_Z5000	-0,13	0,14	0,06	-0,04	-0,10	-0,18	-0,16	-0,01	0,16
Y_S5000	-0,01	0,19	0,15	0,03	-0,08	-0,14	-0,19	-0,18	0,04
Y_D5000	0,00	0,22	0,19	0,04	-0,09	-0,12	-0,22	-0,28	-0,02
COAST_Z5000	-0,16	0,18	0,02	-0,11	-0,17	-0,01	0,06	0,17	0,09
COAST_S5000	-0,07	0,26	0,14	-0,04	-0,16	0,04	0,10	0,14	0,02
COAST_D5000	-0,05	0,29	0,20	-0,01	-0,16	0,04	0,09	0,12	0,01
Z5000_S5000	-0,10	0,15	0,08	-0,03	-0,10	-0,19	-0,21	-0,11	0,13
Z5000_D5000	-0,10	0,19	0,11	-0,02	-0,11	-0,20	-0,24	-0,16	0,12
S5000_D5000	-0,06	0,20	0,10	-0,04	-0,13	-0,15	-0,24	-0,25	0,03
A_C	1,00	-0,29	0,62	0,86	0,76	0,31	0,34	0,16	-0,22
A_N		1,00	0,42	-0,29	-0,73	0,18	0,20	0,11	-0,12
A_NE			1,00	0,75	0,32	0,22	0,19	-0,02	-0,22
A_E				1,00	0,87	0,11	0,05	-0,10	-0,14
A_SE					1,00	-0,02	-0,07	-0,13	-0,04
C_N						1,00	0,91	0,09	-0,90
C_NE							1,00	0,49	-0,63
C_E								1,00	0,36
C_SE									1,00

Table 14.

Statistical values obtained to asses the agreement of the different interpolation methods.

The highlighted methods, are the selected ones to be compared in more detail.

	ANN						***SPR***						**SUM**						**AUT**						**WIN**					
	r^2	**MBE**	**RMSE**	**MAE**	**EF**	**D**	**r^2**	**MBE**	**RMSE**	**MAE**	**EF**	**D**	**r^2**	**MBE**	**RMSE**	**MAE**	**EF**	**D**	**r^2**	**MBE**	**RMSE**	**MAE**	**EF**	**D**	**r^2**	**MBE**	**RMSE**	**MAE**	**EF**	**D**
OKn	0,86	0,93	69,72	50,67	0,74	1,00	0,84	0,19	17,63	13,32	0,71	1,00	0,84	0,09	11,80	7,62	0,71	0,99	0,90	-0,14	24,23	17,59	0,82	1,00	0,88	1,33	22,00	15,43	0,77	0,99
OKc	0,86	1,10	69,56	51,20	0,74	1,00	0,84	0,28	17,78	13,46	0,70	1,00	0,84	0,13	11,88	7,77	0,71	0,99	0,90	0,26	24,39	18,03	0,81	1,00	0,88	1,35	22,07	15,54	0,77	0,99
OKf	0,87	0,93	68,30	49,38	0,75	1,00	0,85	0,41	17,46	13,01	0,71	1,00	0,86	0,58	11,41	7,35	0,73	0,99	0,90	-0,21	24,06	17,82	0,82	1,00	0,87	1,56	22,50	15,42	0,76	0,99
OKs	0,87	2,21	69,08	49,35	0,75	1,00	0,84	0,70	17,61	13,24	0,71	1,00	0,85	0,75	11,70	7,35	0,72	0,99	0,90	0,14	24,18	17,82	0,82	1,00	0,87	1,91	22,91	15,55	0,75	0,99
OKt	0,87	2,54	68,86	49,52	0,75	1,00	0,84	0,65	17,73	13,34	0,70	1,00	0,86	1,19	11,81	7,45	0,71	0,99	0,91	0,15	23,67	17,24	0,82	1,00	0,87	2,00	22,84	15,75	0,75	0,99
SK	0,85	1,39	72,07	52,52	0,72	1,00	0,84	0,36	17,66	13,24	0,71	1,00	0,84	-0,13	12,08	7,66	0,70	0,99	0,90	0,56	24,29	17,92	0,81	1,00	0,88	1,73	21,99	15,45	0,77	0,99
UKc	0,86	0,93	69,72	50,67	0,74	1,00	0,84	0,19	17,63	13,32	0,71	1,00	0,84	0,09	11,80	7,62	0,71	0,99	0,90	-0,14	24,23	17,59	0,82	1,00	0,88	1,33	22,00	15,43	0,77	0,99
UKf	0,87	1,33	67,89	49,05	0,75	1,00	0,84	-0,01	17,45	13,14	0,71	1,00	0,85	0,68	11,55	7,30	0,72	0,99	0,90	-0,10	24,08	17,79	0,82	1,00	0,88	0,30	22,08	15,76	0,77	0,99
UKs	0,86	-0,44	70,09	50,31	0,73	1,00	0,83	0,79	18,19	13,82	0,69	1,00	0,84	-0,20	12,08	7,67	0,70	0,99	0,89	-0,15	25,85	19,35	0,79	1,00	0,87	-0,28	23,01	15,97	0,76	0,99
UKt	0,77	12,75	103,44	67,29	0,42	0,99	0,81	-0,39	19,53	14,13	0,63	0,99	0,71	0,36	17,38	10,14	0,32	0,98	0,80	1,35	40,43	22,90	0,49	0,99	0,86	3,35	26,06	19,07	0,69	0,99
OCKn	0,90	1,57	59,43	44,64	0,81	1,00	0,90	-0,11	14,17	10,77	0,81	1,00	0,88	0,19	10,77	7,41	0,76	0,99	0,92	-0,17	21,65	15,75	0,85	1,00	0,91	1,33	18,68	14,25	0,83	0,99
OCKc	0,88	0,99	66,02	49,33	0,77	1,00	0,88	0,09	15,52	11,80	0,77	1,00	0,85	0,21	11,56	7,60	0,72	0,99	0,91	0,25	23,67	17,47	0,82	1,00	0,89	1,41	20,95	14,94	0,79	0,99
OCKf	0,87	0,94	66,90	48,50	0,76	1,00	0,88	0,29	15,80	11,85	0,77	1,00	0,86	0,64	11,21	7,20	0,74	0,99	0,91	-0,23	23,74	17,50	0,82	1,00	0,88	1,64	21,88	15,01	0,77	0,99
OCKs	0,87	2,29	67,54	47,94	0,76	1,00	0,88	0,53	15,56	11,68	0,77	1,00	0,86	0,86	11,48	7,21	0,73	0,99	0,91	0,17	23,91	17,46	0,82	1,00	0,88	2,01	22,28	15,26	0,76	0,99
OCKt	0,87	2,50	67,72	48,75	0,76	1,00	0,88	0,52	15,39	11,65	0,78	1,00	0,86	1,18	11,80	7,45	0,71	0,99	0,91	0,16	23,48	17,02	0,83	1,00	0,88	2,04	22,37	15,57	0,76	0,99
SCK	0,90	1,22	58,58	44,00	0,82	1,00	0,90	0,11	14,09	10,67	0,81	1,00	0,88	0,37	10,36	7,08	0,78	0,99	0,93	0,19	21,41	15,86	0,86	1,00	0,92	1,49	18,92	14,62	0,83	0,99
UCKc	0,88	0,84	66,09	48,75	0,77	1,00	0,88	0,00	15,37	11,67	0,78	1,00	0,85	0,17	11,44	7,44	0,73	0,99	0,91	-0,16	23,49	16,94	0,83	1,00	0,89	1,39	20,99	14,97	0,79	0,99
UCKf	0,88	1,50	66,40	48,26	0,77	1,00	0,88	0,13	15,41	11,69	0,78	1,00	0,86	0,69	11,51	7,30	0,73	0,99	0,91	-0,01	23,74	17,43	0,82	1,00	0,88	0,39	21,57	15,54	0,78	0,99
UCKs	0,86	-0,41	69,61	50,10	0,74	1,00	0,87	0,61	16,09	12,16	0,75	1,00	0,83	-0,28	12,31	7,77	0,69	0,99	0,89	-0,15	25,88	19,32	0,79	1,00	0,88	-0,26	22,90	15,95	0,76	0,99
UCKt	0,86	13,49	74,51	57,25	0,69	0,99	0,84	-0,03	16,95	12,65	0,71	1,00	0,69	0,51	17,39	10,00	0,28	0,98	0,81	2,21	37,99	21,20	0,54	0,99	0,87	4,32	24,70	18,30	0,73	0,99
MR1	0,91	0,00	55,75	42,42	0,83	1,00	0,90	0,00	14,12	10,42	0,81	1,00	0,92	-0,51	8,88	6,59	0,84	1,00	0,93	0,00	21,11	16,17	0,86	1,00	0,92	-1,39	18,02	13,71	0,85	0,99
MR2	0,89	0,00	61,73	49,00	0,80	1,00	0,89	0,00	14,66	10,66	0,80	1,00	0,90	-0,67	9,66	7,54	0,81	0,99	0,90	0,00	25,13	19,39	0,80	1,00	0,90	-1,66	19,95	15,39	0,81	0,99
IDW1	0,83	2,28	76,52	56,96	0,69	0,99	0,80	0,35	19,68	14,61	0,64	0,99	0,85	-1,01	11,78	8,10	0,71	0,99	0,87	1,57	28,08	21,42	0,75	0,99	0,83	1,37	25,29	18,61	0,70	0,99
IDW2	0,83	4,65	77,83	56,71	0,68	0,99	0,79	1,07	20,16	14,80	0,62	0,99	0,84	-0,69	11,86	7,95	0,71	0,99	0,87	2,17	28,18	20,81	0,75	0,99	0,83	2,09	26,01	18,33	0,68	0,99
IDW3	0,82	6,15	80,79	57,94	0,65	0,99	0,78	1,55	20,89	15,17	0,59	0,99	0,84	-0,52	12,06	8,07	0,70	0,99	0,86	2,68	29,11	21,13	0,73	0,99	0,82	2,44	27,31	18,63	0,65	0,99
LPI1	0,86	-2,44	70,76	53,35	0,73	1,00	0,83	-0,52	18,12	13,77	0,69	1,00	0,87	0,50	10,84	7,57	0,76	0,99	0,90	-0,58	24,81	19,12	0,81	1,00	0,86	-1,03	23,20	17,18	0,74	0,99
LPI2	0,84	-1,43	73,46	55,92	0,71	0,99	0,81	-0,18	18,99	14,13	0,66	0,99	0,87	-0,04	10,78	7,33	0,76	0,99	0,89	-0,76	25,47	19,91	0,80	1,00	0,84	-1,46	24,95	18,25	0,70	0,99
LPI3	0,83	-1,39	75,63	58,48	0,70	0,99	0,80	-0,18	19,75	14,76	0,63	0,99	0,87	0,15	10,85	7,60	0,76	0,99	0,89	-1,26	26,17	20,65	0,79	1,00	0,83	0,40	26,35	19,32	0,67	0,99

Figure 4.

Semivariogram surfaces

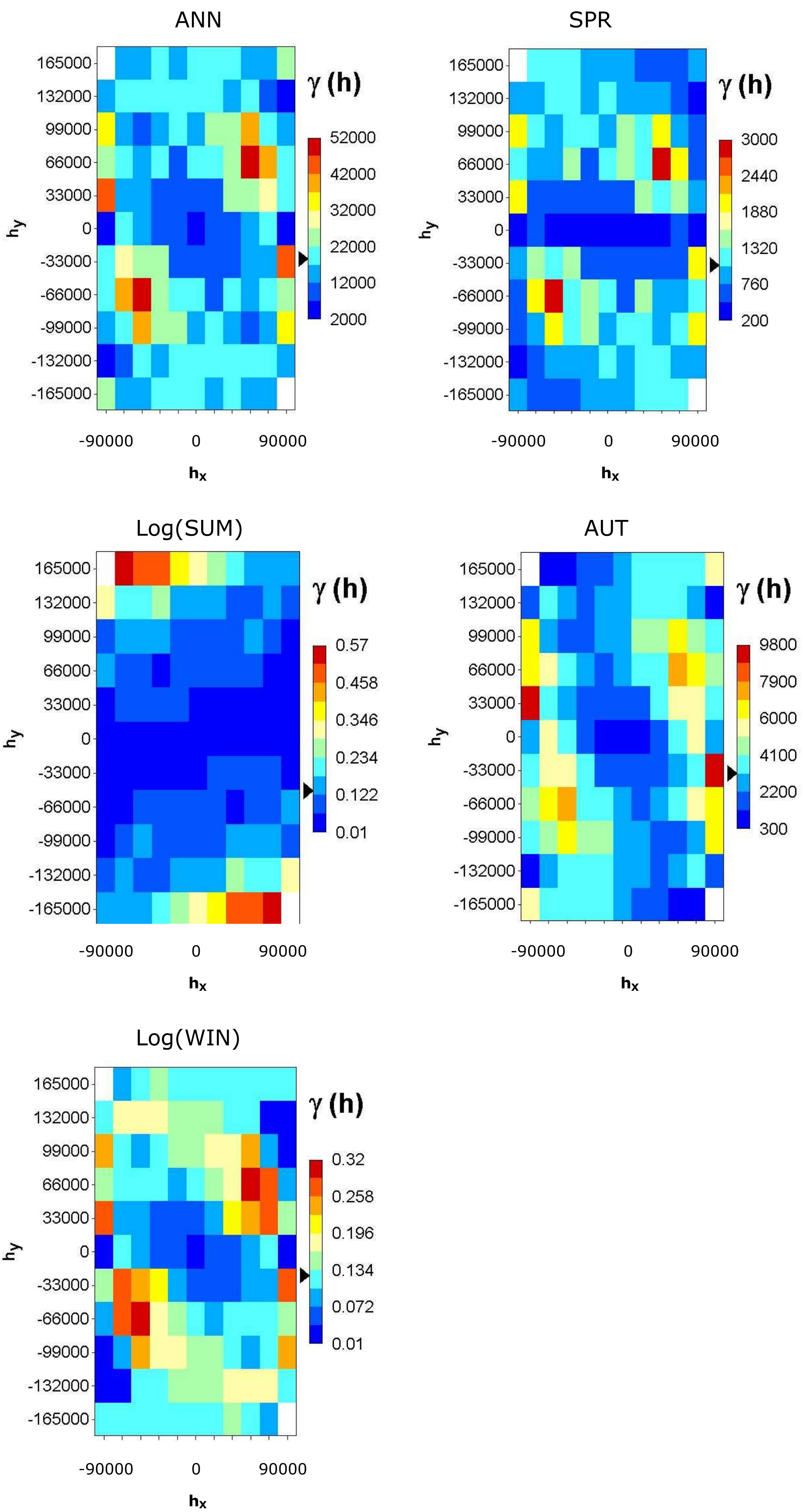


Figure 5.

Semivariogram models for ANN

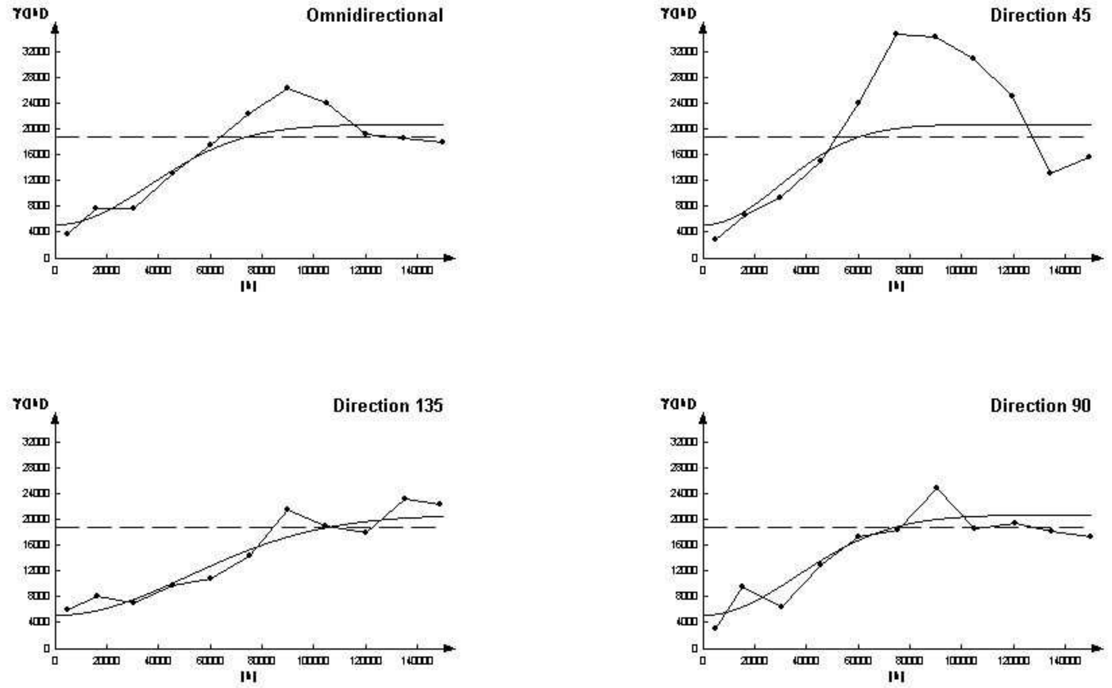


Figure 6.

Semivariogram models for SPR

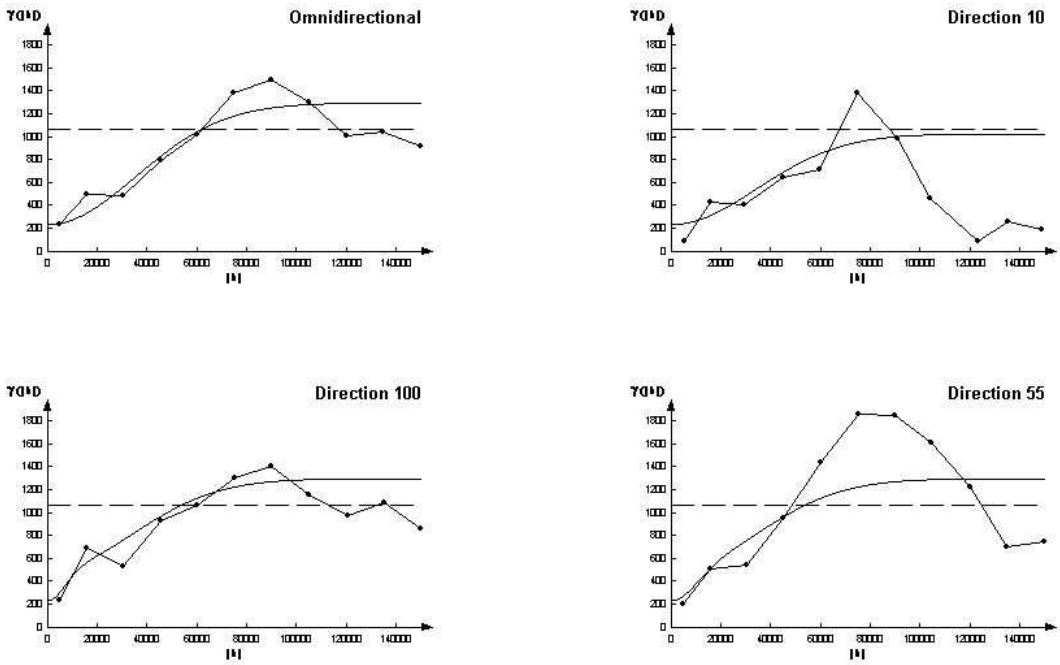


Figure 7.

Semivariogram models for Log(SUM)

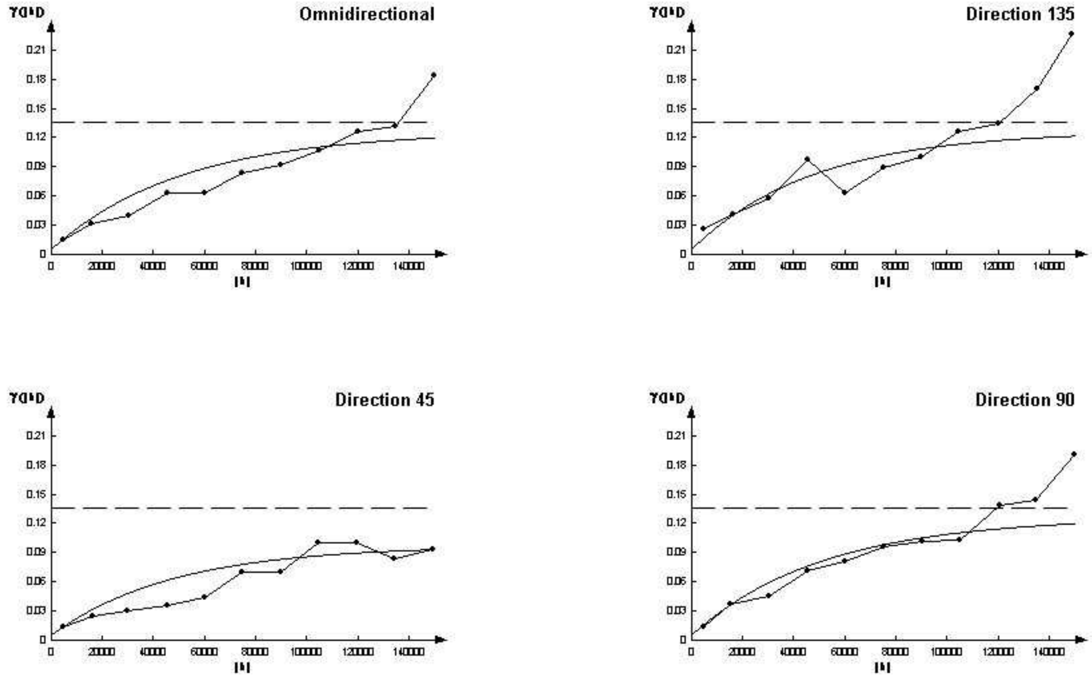


Figure 8.

Semivariogram models for AUT

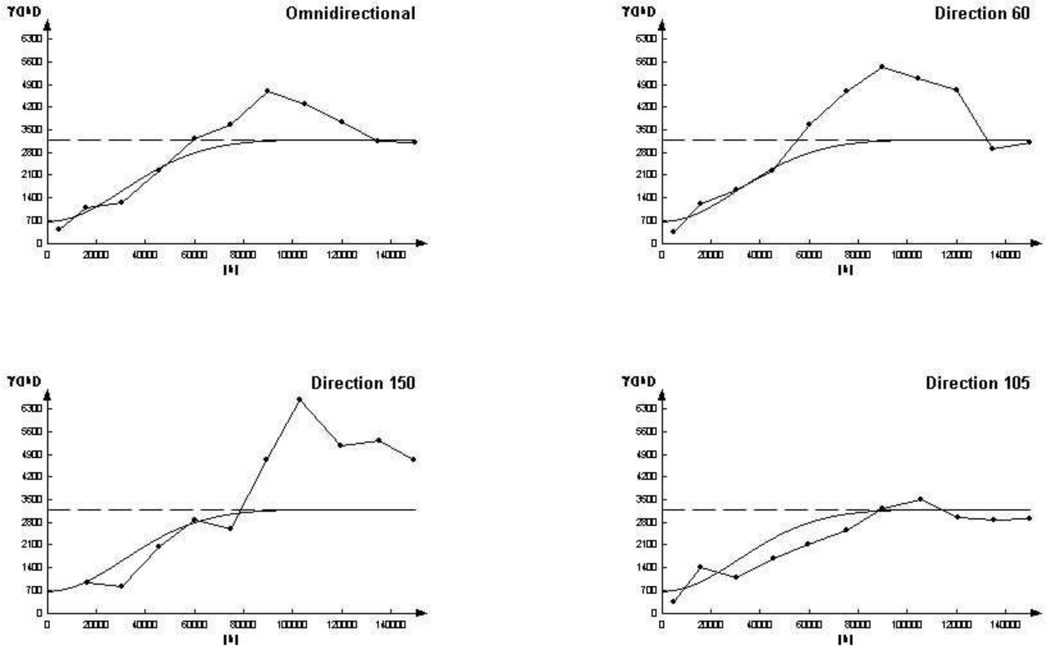


Figure 9.

Semivariogram models for Log(WIN)

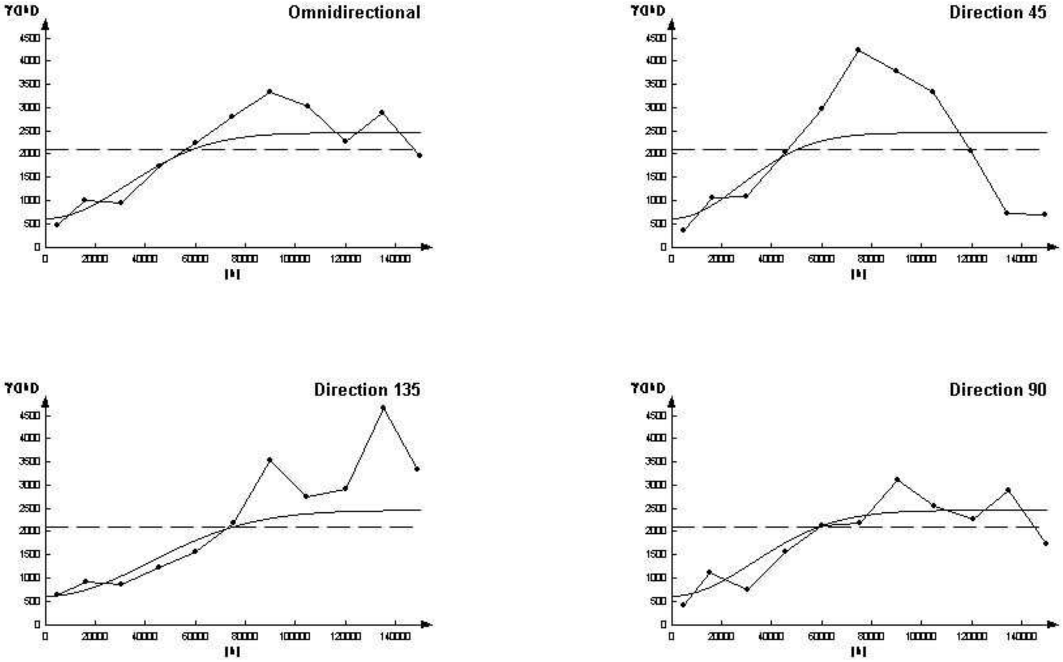
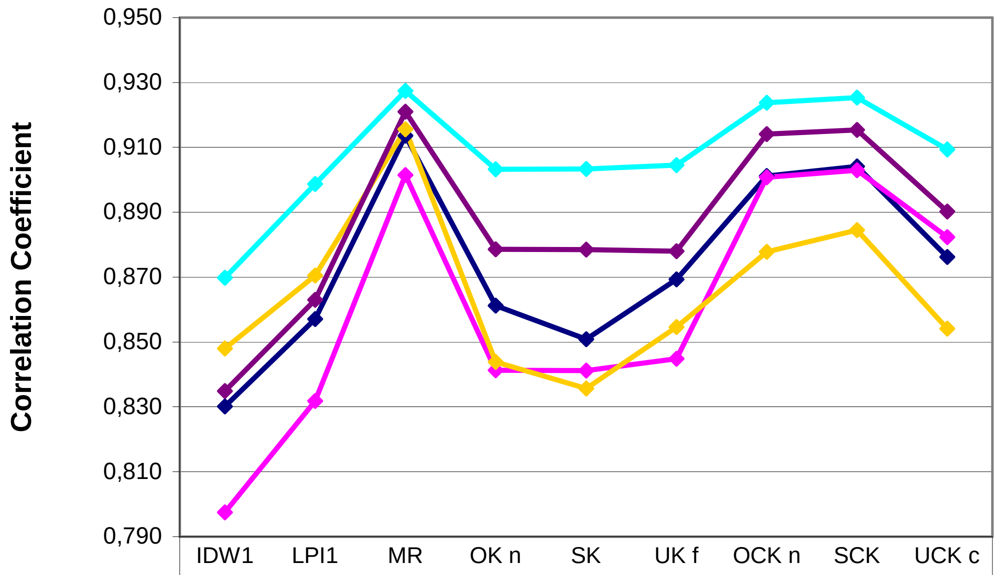


Figure 10.

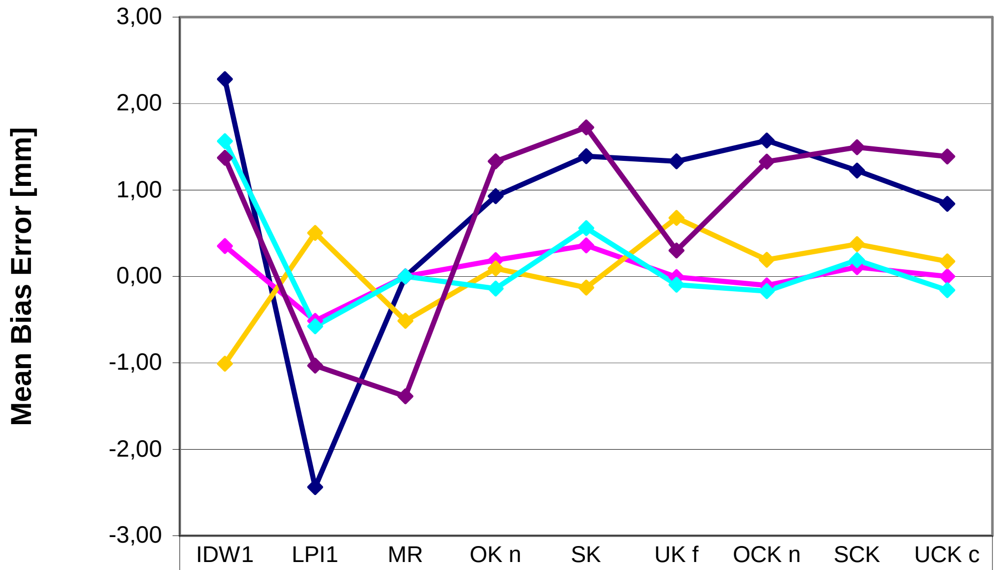
Graph with the correlation coefficient versus the interpolation methods.



	IDW1	LPI1	MR	OK n	SK	UK f	OCK n	SCK	UCK c
ANN	0,830	0,857	0,914	0,861	0,851	0,869	0,901	0,904	0,876
SPR	0,798	0,832	0,901	0,841	0,841	0,845	0,901	0,903	0,882
SUM	0,848	0,870	0,916	0,844	0,836	0,855	0,878	0,884	0,854
AUT	0,870	0,899	0,927	0,903	0,903	0,904	0,924	0,925	0,909
WIN	0,835	0,863	0,921	0,879	0,879	0,878	0,914	0,915	0,890

Figure 11.

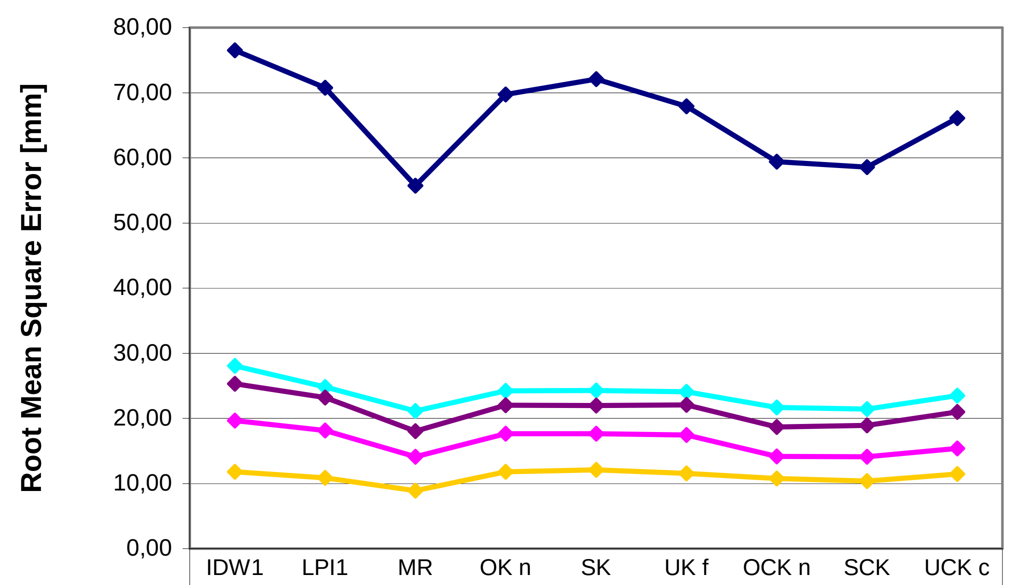
Graph with the mean bias error versus the interpolation methods.



	IDW1	LPI1	MR	OK n	SK	UK f	OCK n	SCK	UCK c
ANN	2,28	-2,44	0,00	0,93	1,39	1,33	1,57	1,22	0,84
SPR	0,35	-0,52	0,00	0,19	0,36	-0,01	-0,11	0,11	0,00
SUM	-1,01	0,50	-0,51	0,09	-0,13	0,68	0,19	0,37	0,17
AUT	1,57	-0,58	0,00	-0,14	0,56	-0,10	-0,17	0,19	-0,16
WIN	1,37	-1,03	-1,39	1,33	1,73	0,30	1,33	1,49	1,39

Figure 12.

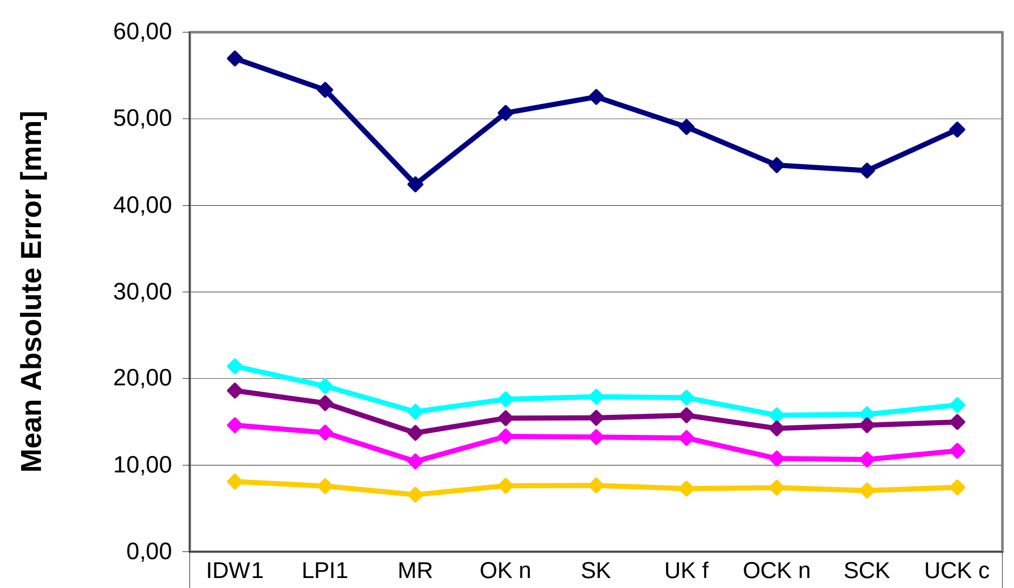
Graph with the root mean square error versus the interpolation methods.



	IDW1	LPI1	MR	OK n	SK	UK f	OCK n	SCK	UCK c
ANN	76,52	70,76	55,75	69,72	72,07	67,89	59,43	58,58	66,09
SPR	19,68	18,12	14,12	17,63	17,66	17,45	14,17	14,09	15,37
SUM	11,78	10,84	8,88	11,80	12,08	11,55	10,77	10,36	11,44
AUT	28,08	24,81	21,11	24,23	24,29	24,08	21,65	21,41	23,49
WIN	25,29	23,20	18,02	22,00	21,99	22,08	18,68	18,92	20,99

Figure 13.

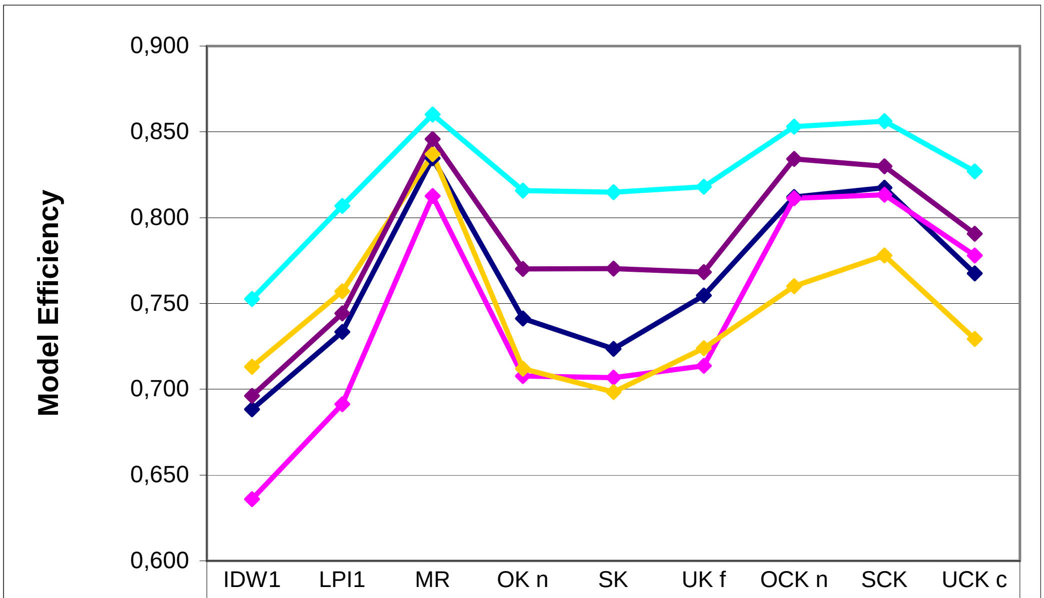
Graph with the mean absolute error versus the interpolation methods.



	IDW1	LPI1	MR	OK n	SK	UK f	OCK n	SCK	UCK c
ANN	56,96	53,35	42,42	50,67	52,52	49,05	44,64	44,00	48,75
SPR	14,61	13,77	10,42	13,32	13,24	13,14	10,77	10,67	11,67
SUM	8,10	7,57	6,59	7,62	7,66	7,30	7,41	7,08	7,44
AUT	21,42	19,12	16,17	17,59	17,92	17,79	15,75	15,86	16,94
WIN	18,61	17,18	13,71	15,43	15,45	15,76	14,25	14,62	14,97

Figure 14.

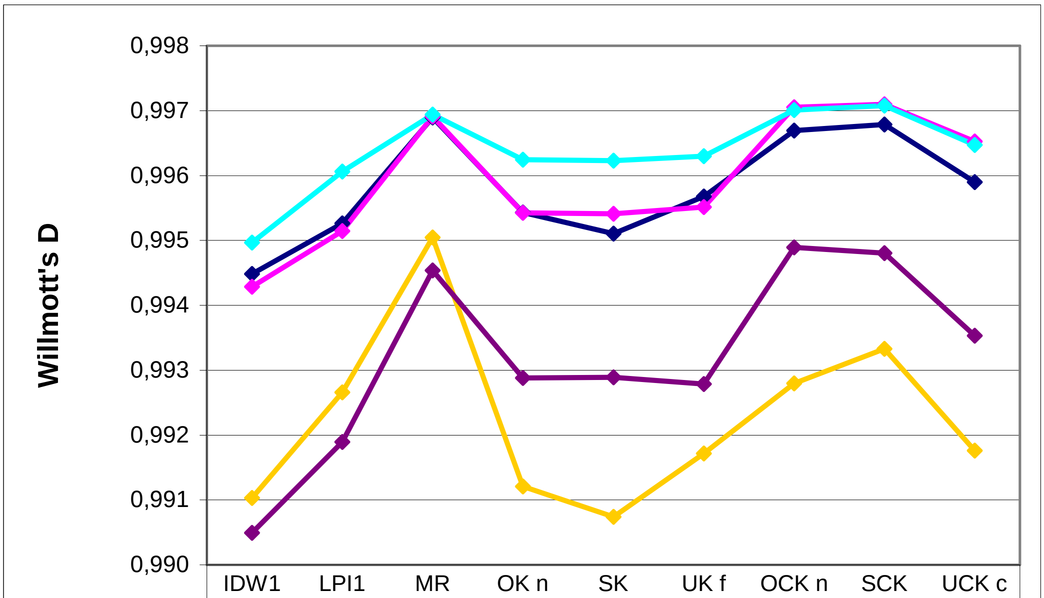
Graph with the model efficiency versus the interpolation methods.



	IDW1	LPI1	MR	OK n	SK	UK f	OCK n	SCK	UCK c
ANN	0,688	0,733	0,835	0,741	0,724	0,755	0,812	0,817	0,768
SPR	0,636	0,691	0,813	0,708	0,707	0,714	0,811	0,813	0,778
SUM	0,713	0,757	0,837	0,712	0,698	0,724	0,760	0,778	0,729
AUT	0,753	0,807	0,860	0,816	0,815	0,818	0,853	0,856	0,827
WIN	0,696	0,744	0,846	0,770	0,770	0,768	0,834	0,830	0,791

Figure 15.

Graph with the Willmott's D versus the interpolation methods.



	IDW1	LPI1	MR	OK n	SK	UK f	OCK n	SCK	UCK c
ANN	0,994	0,995	0,997	0,995	0,995	0,996	0,997	0,997	0,996
SPR	0,994	0,995	0,997	0,995	0,995	0,996	0,997	0,997	0,997
SUM	0,991	0,993	0,995	0,991	0,991	0,992	0,993	0,993	0,992
AUT	0,995	0,996	0,997	0,996	0,996	0,996	0,997	0,997	0,996
WIN	0,990	0,992	0,995	0,993	0,993	0,993	0,995	0,995	0,994

Figure 16.

Predicted versus Observed values of ANN

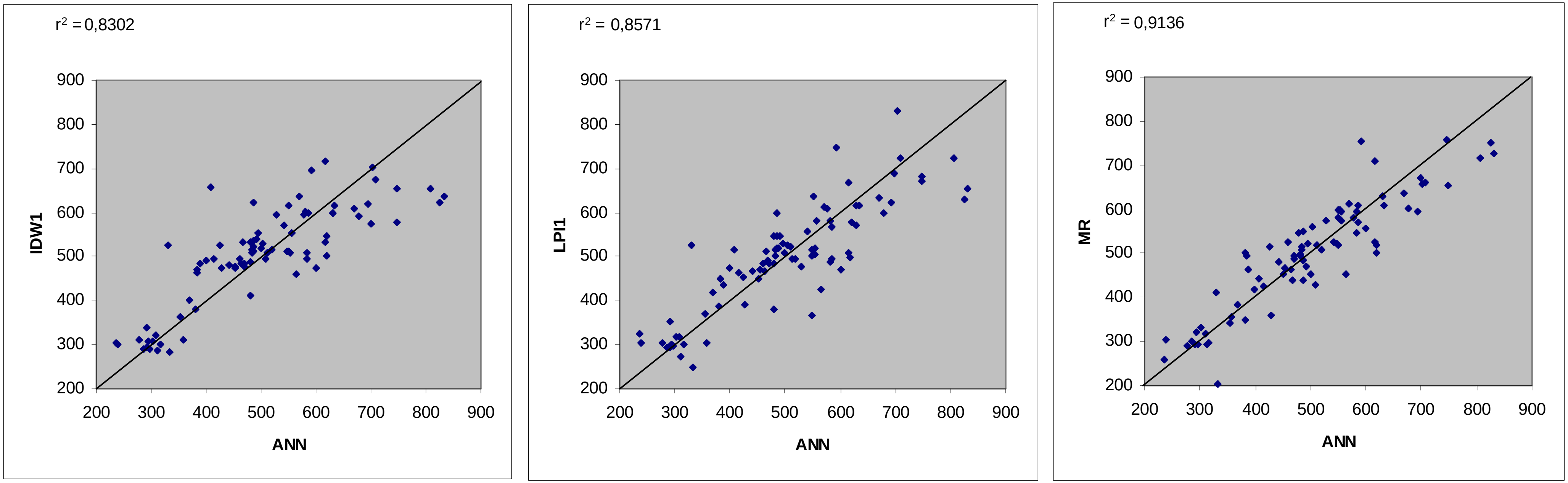


Figure 16. (…continuation)

Predicted versus Observed values of ANN

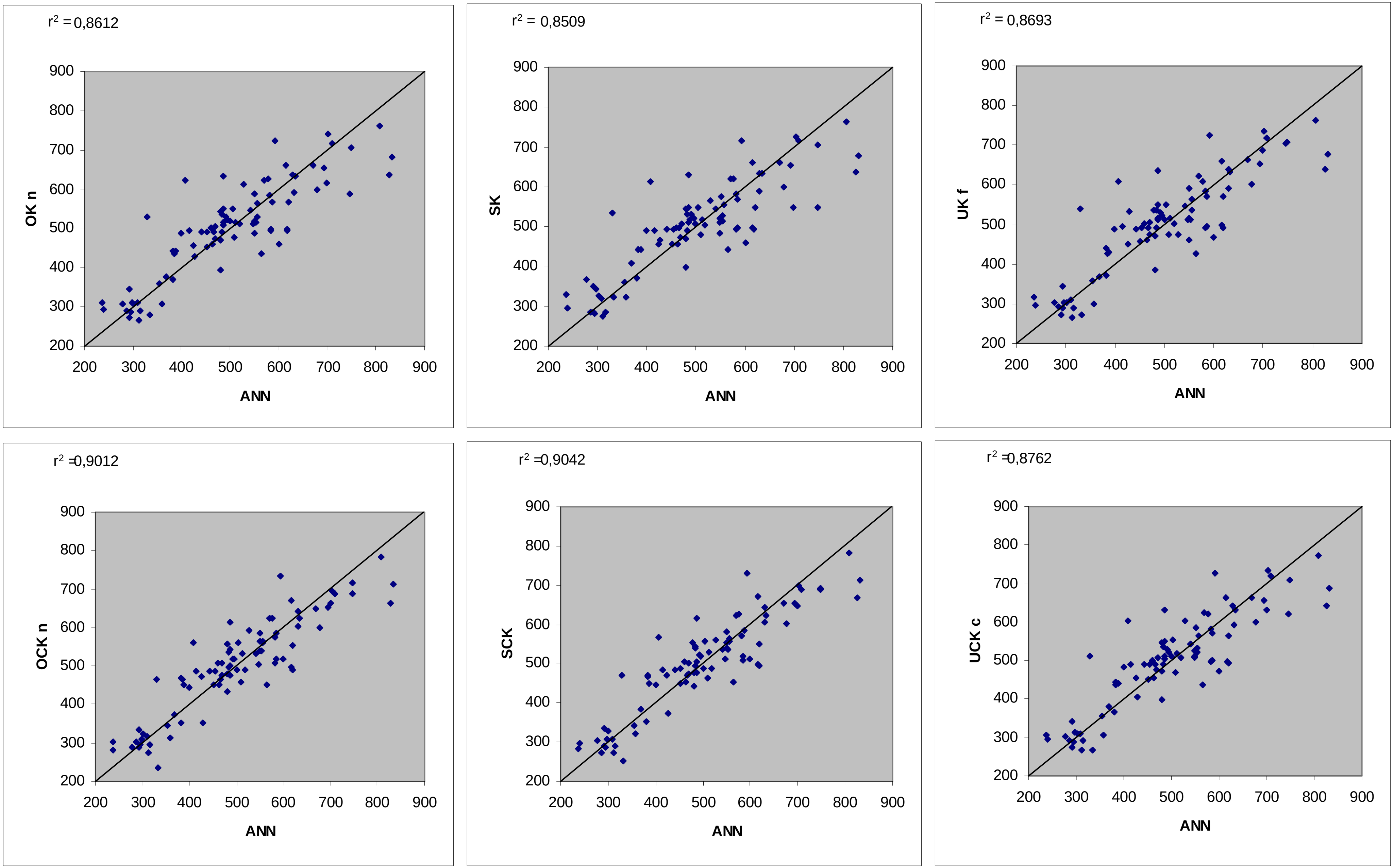


Figure 17.

Predicted versus Observed values of SPR

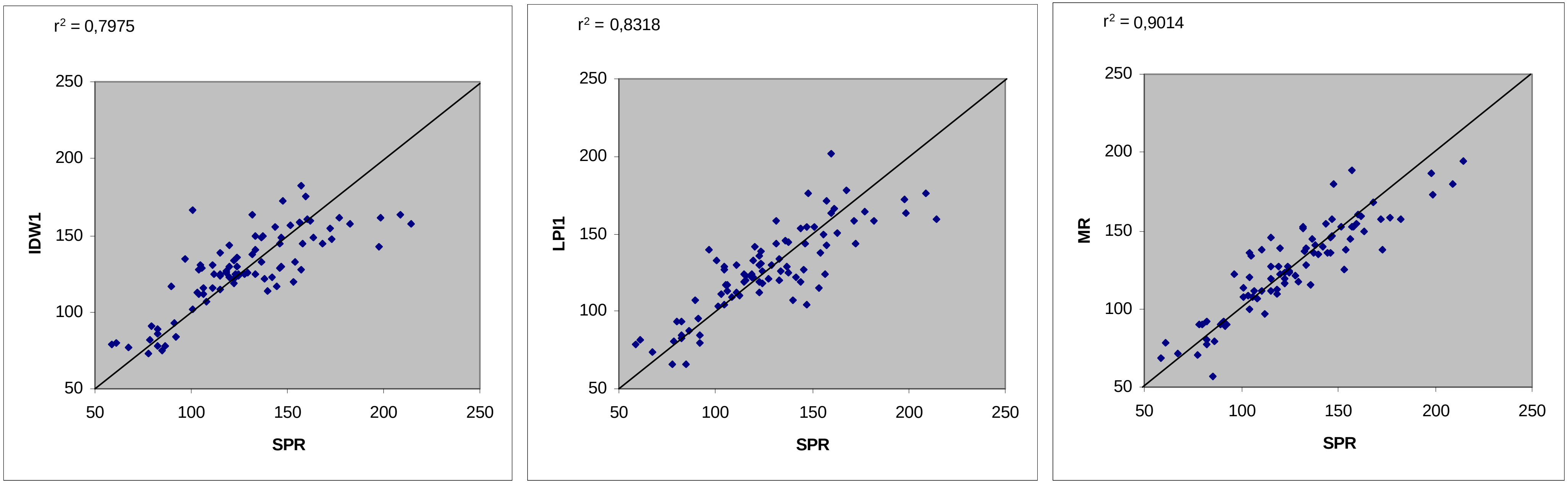


Figure 17. (…continuation)

Predicted versus Observed values of SPR

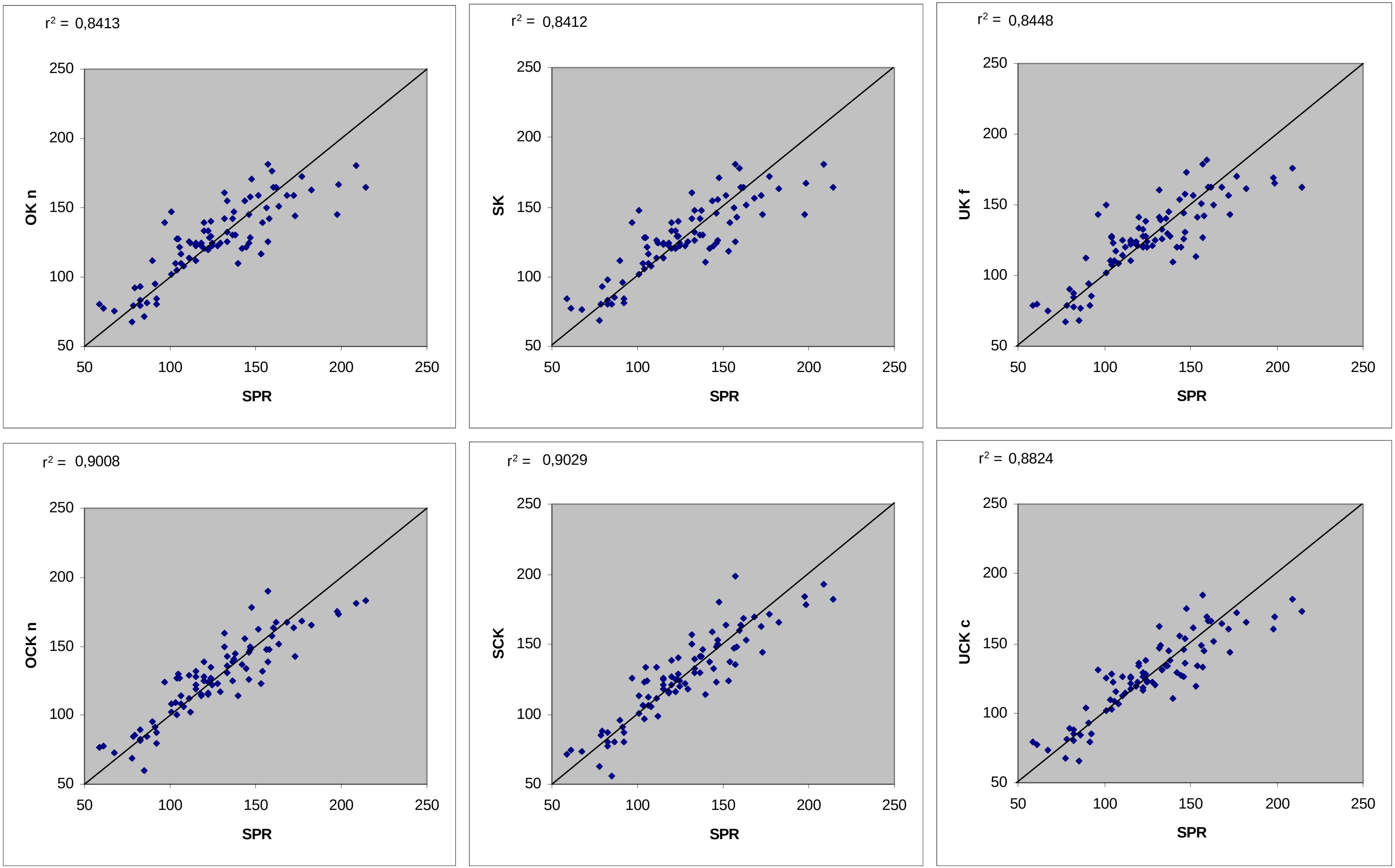
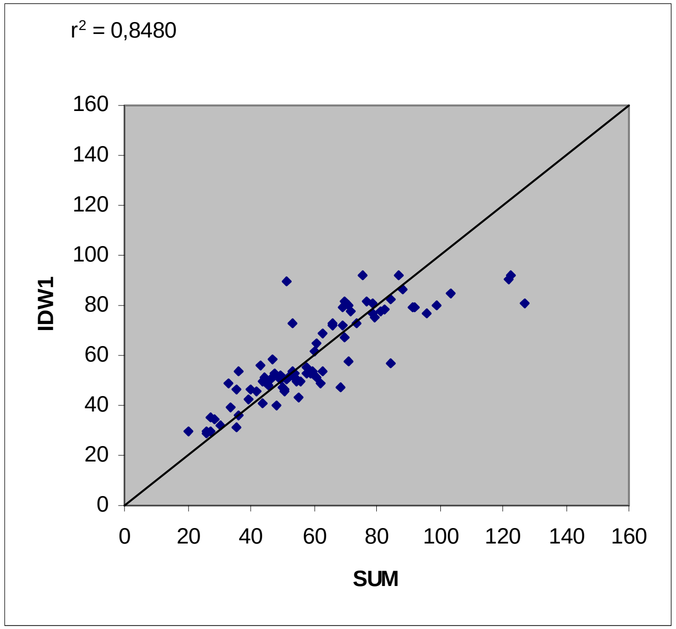
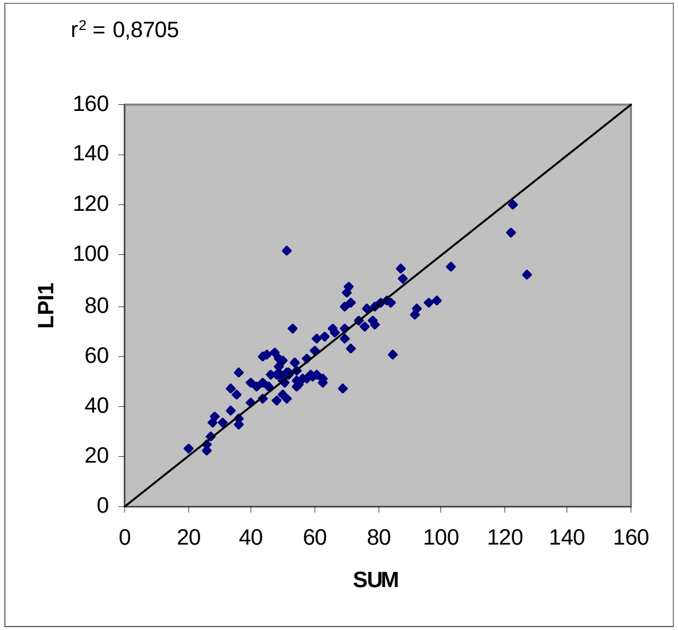


Figure 18.

Predicted versus Observed values of SUM





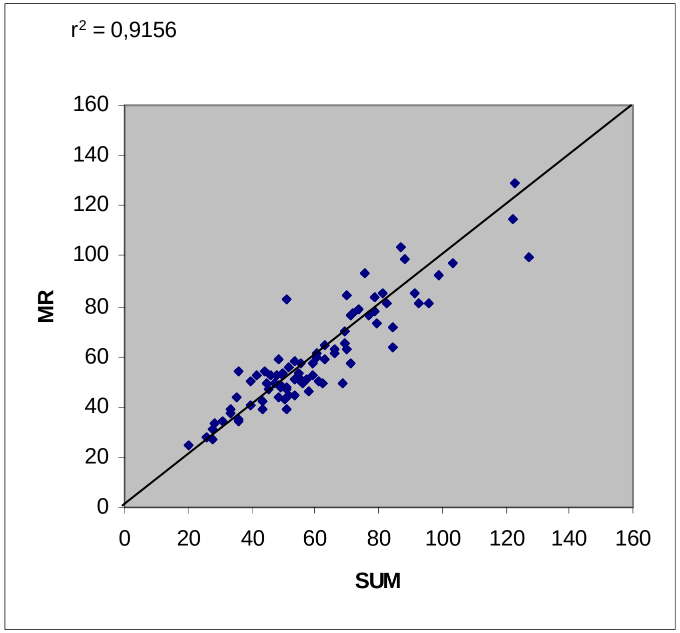


Figure 18. (…continuation)

Predicted versus Observed values of SUM

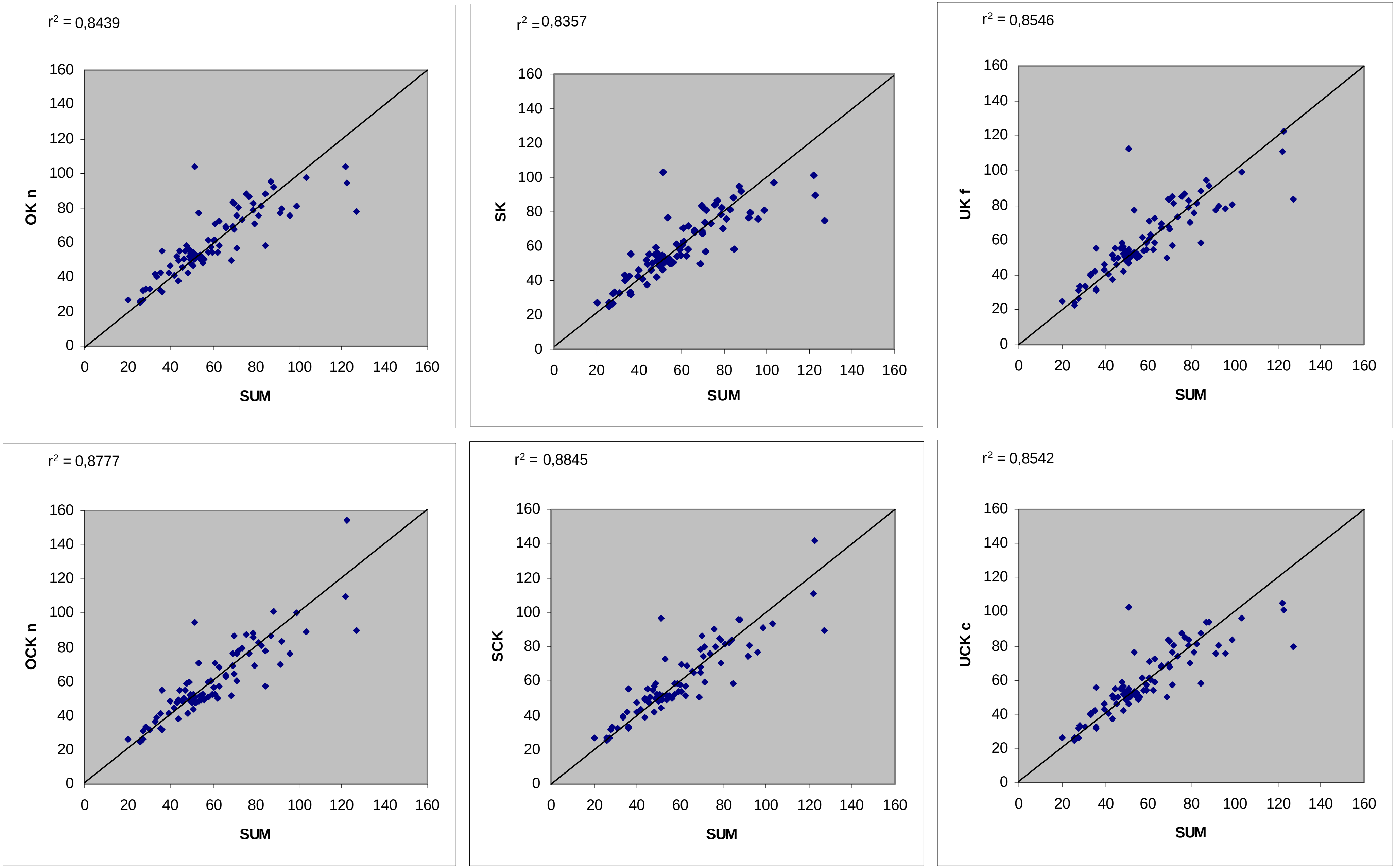


Figure 19.

Predicted versus Observed values of AUT

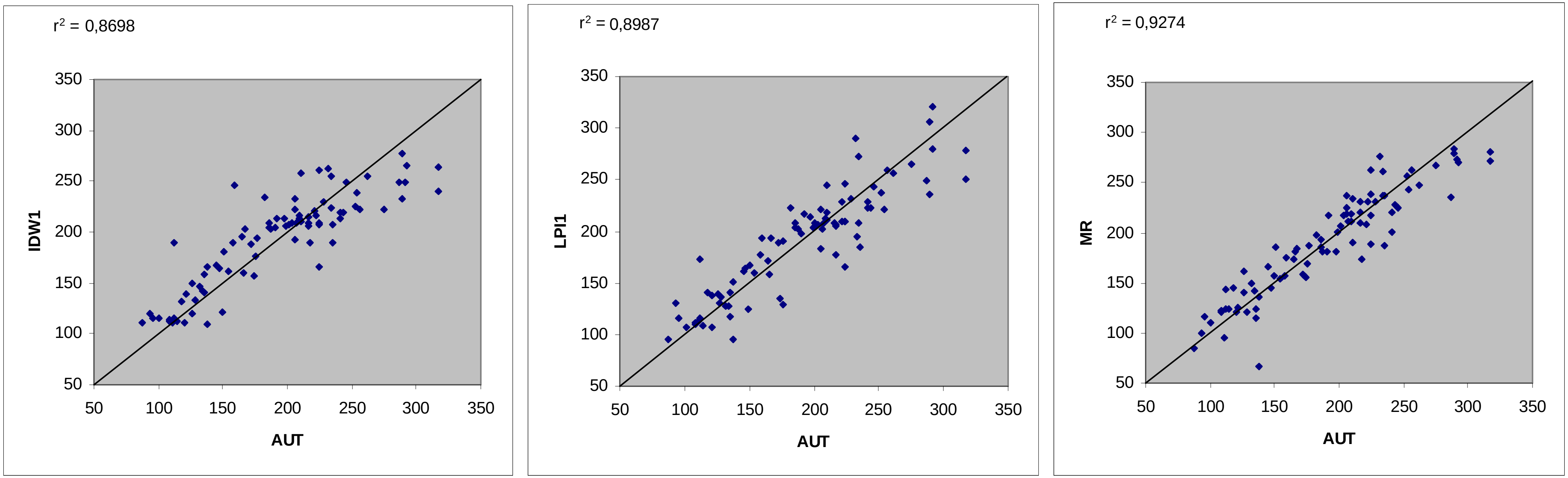


Figure 19. (…continuation)

Predicted versus Observed values of AUT

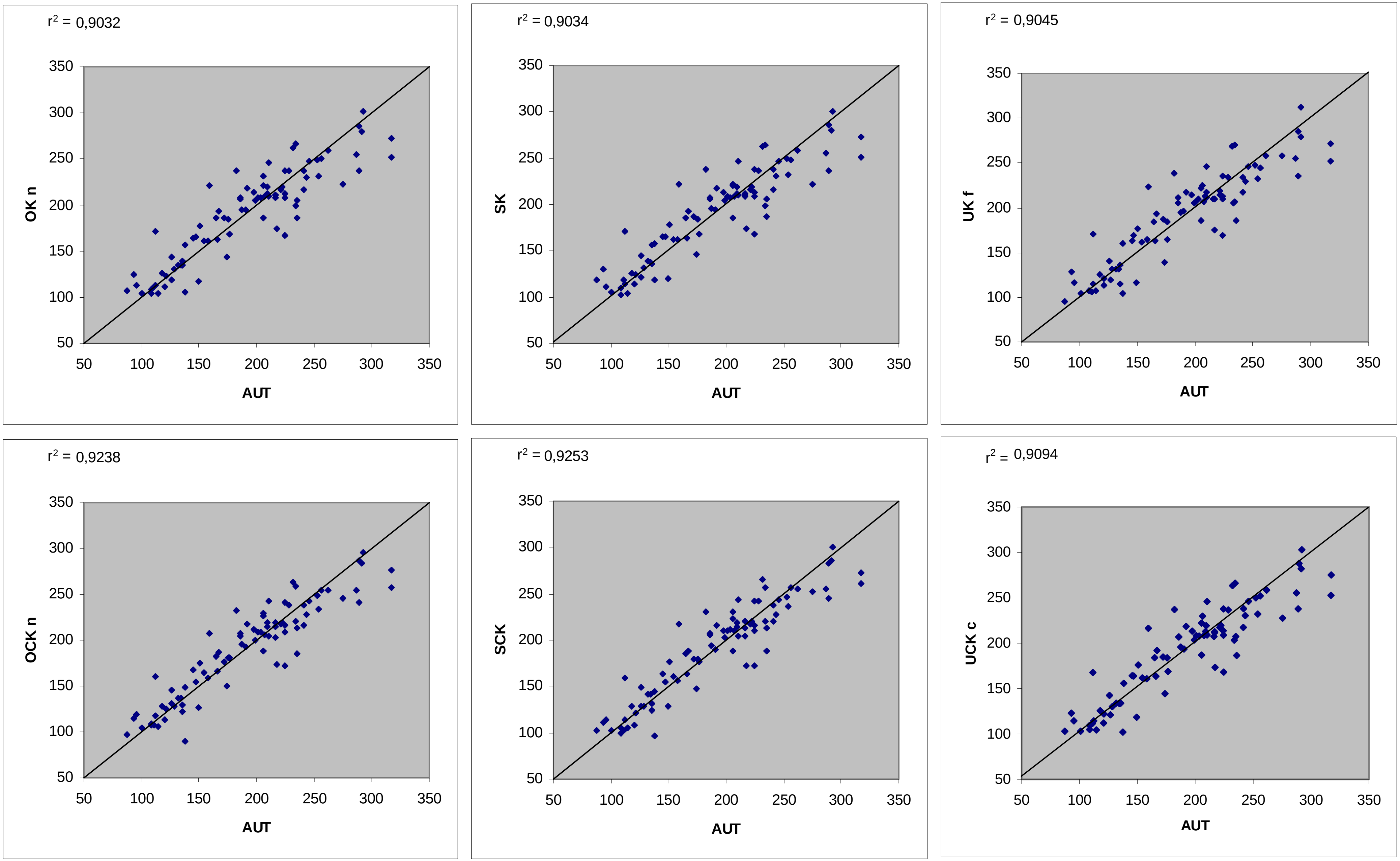


Figure 20.

Predicted versus Observed values of WIN

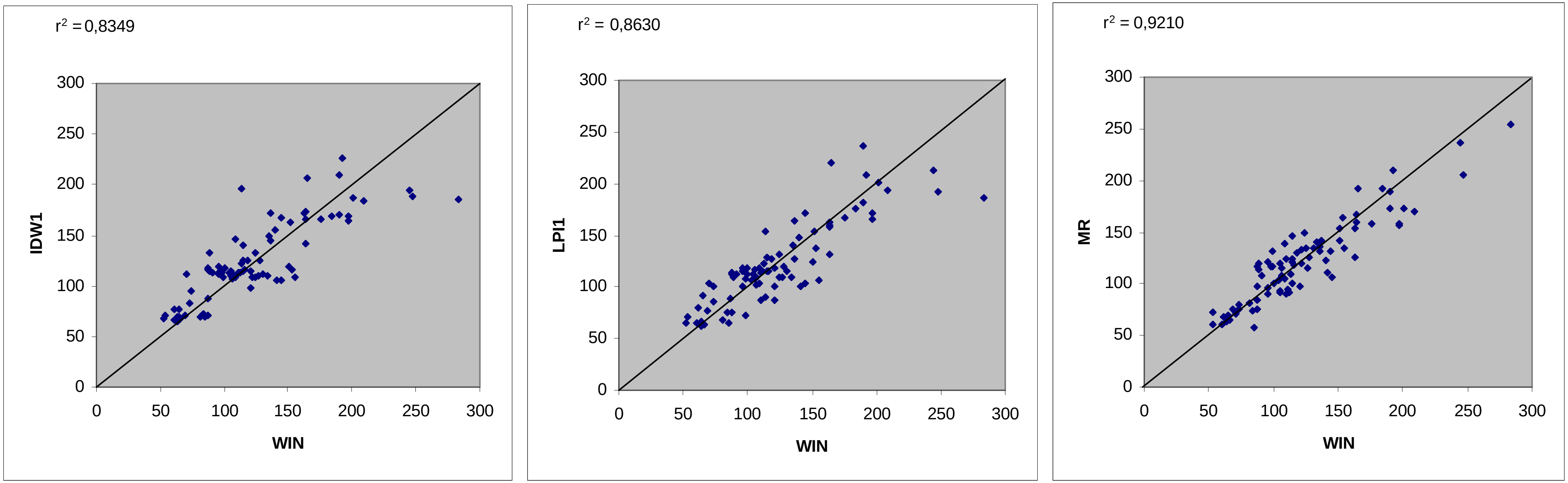


Figure 20. (…continuation)

Predicted versus Observed values of WIN

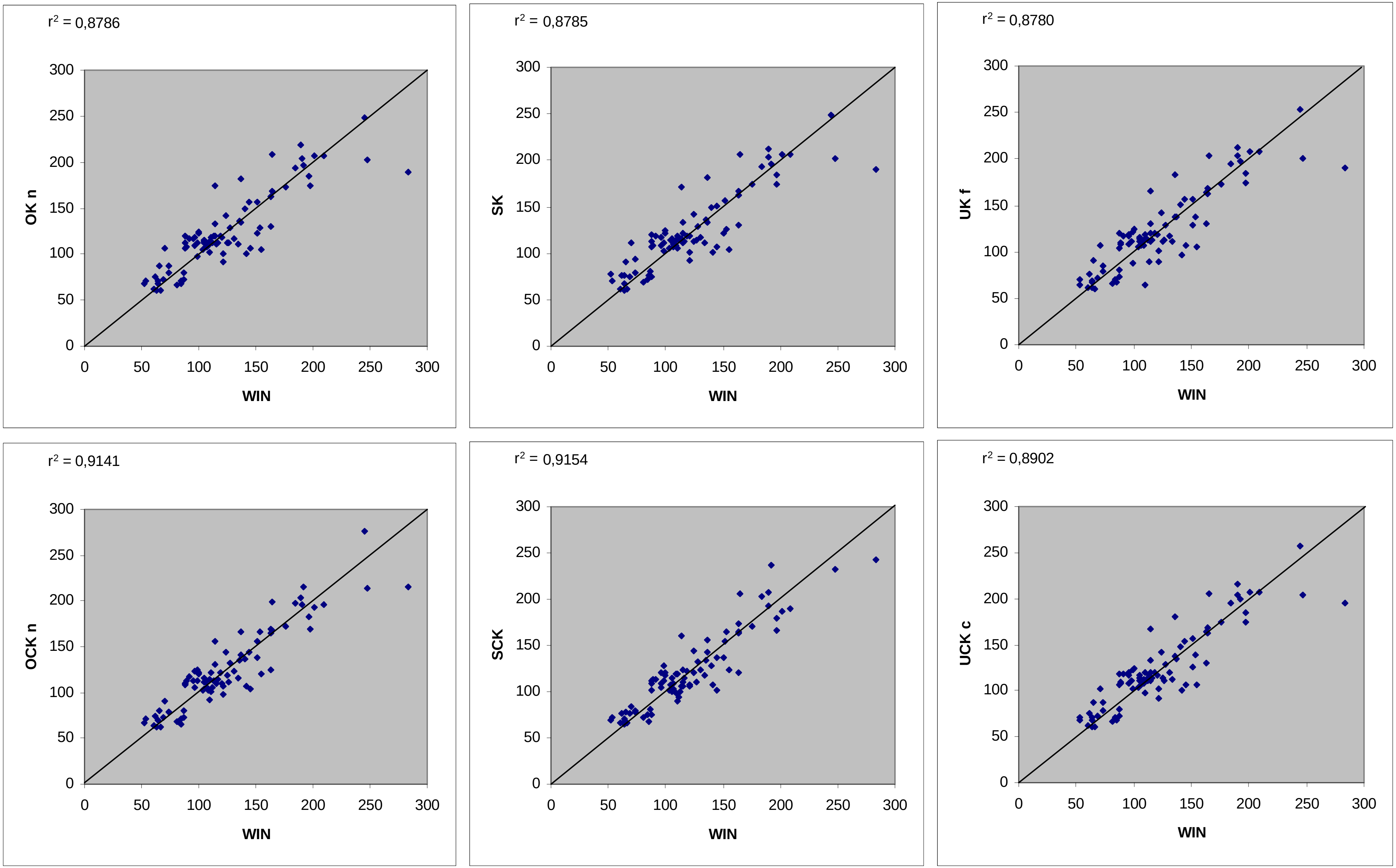


Figure 21.

Error histograms for the variable ANN

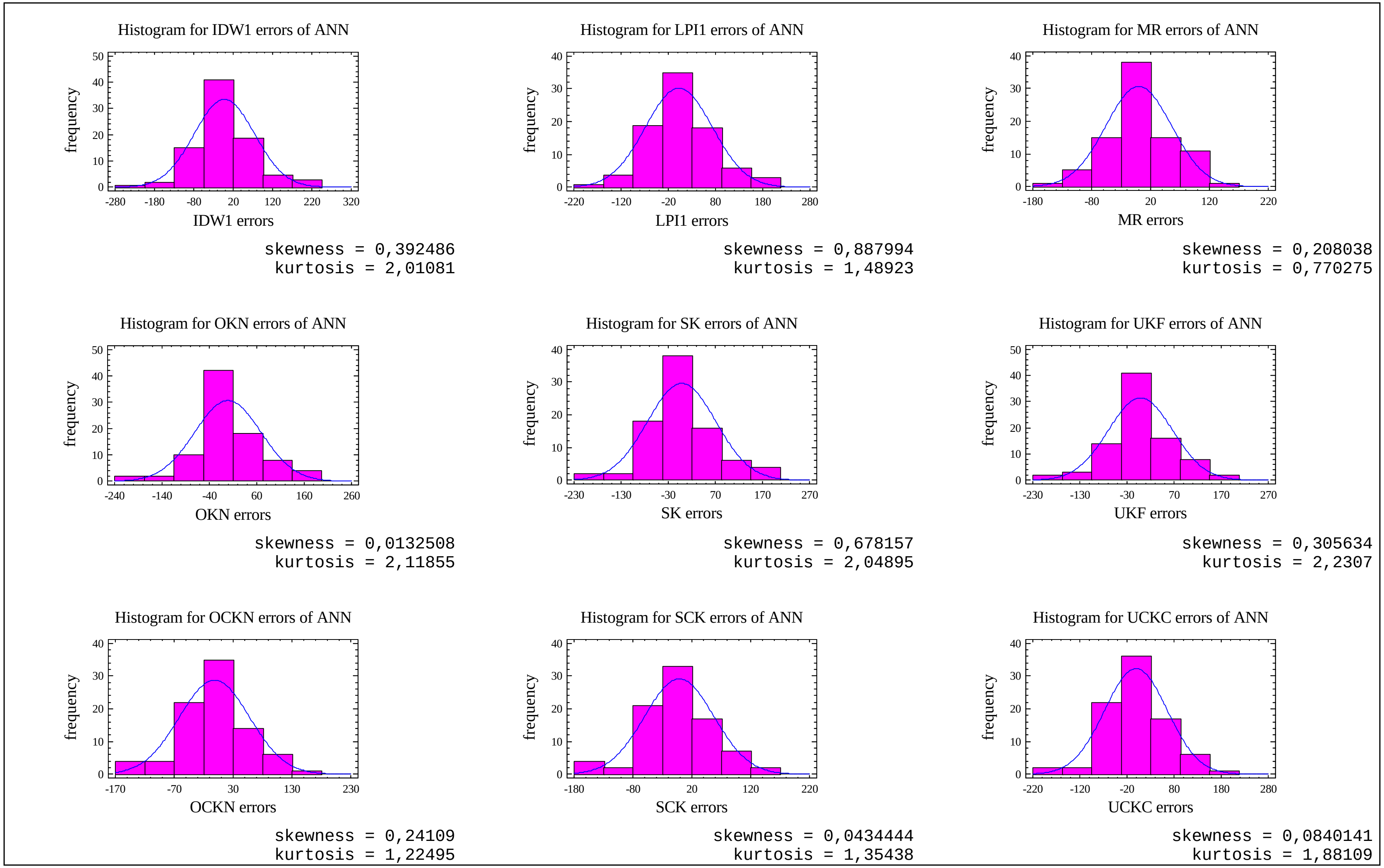


Figure 22.

Error histograms for the variable SPR

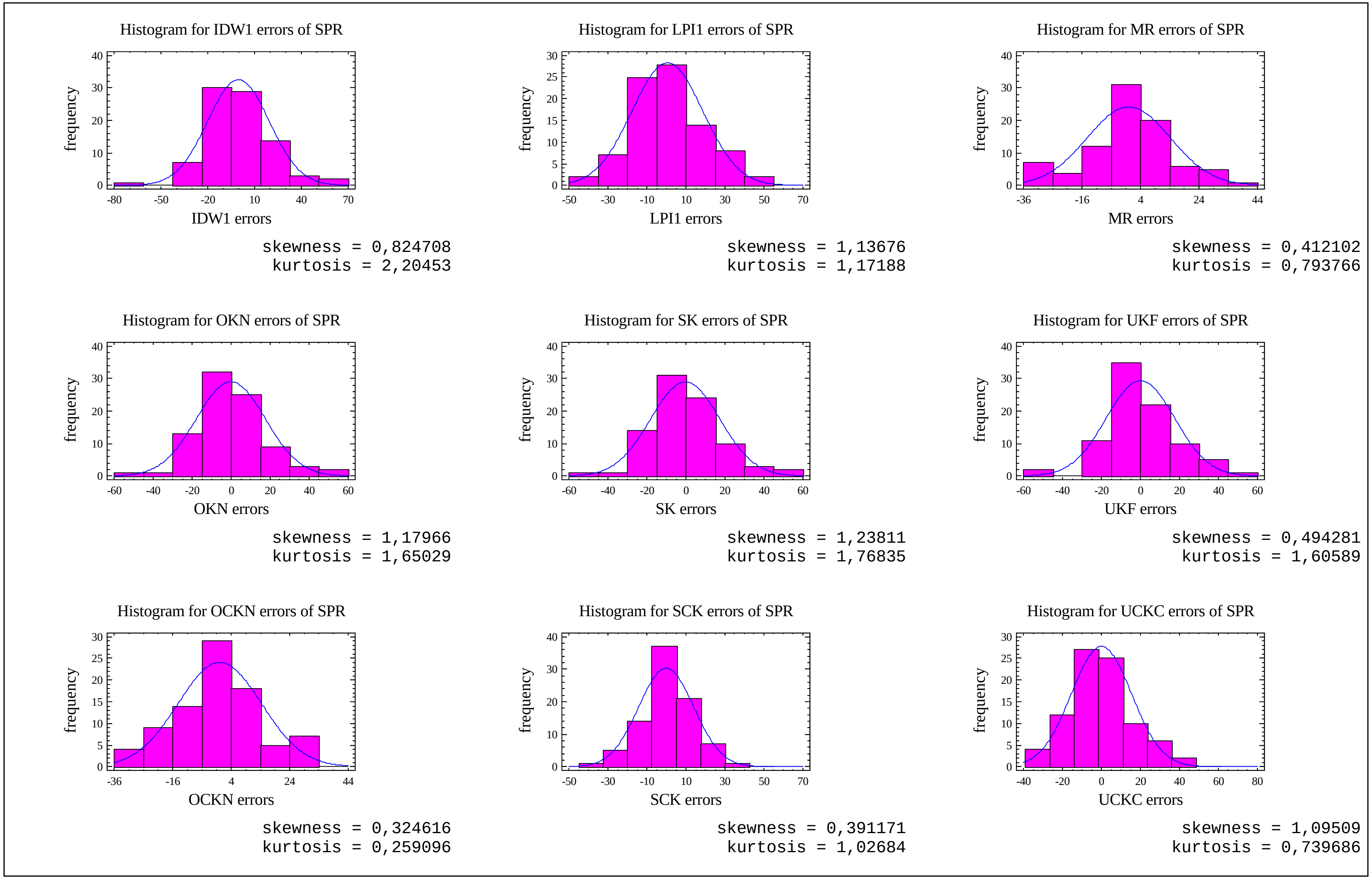


Figure 23.

Error histograms for the variable Log(SUM)

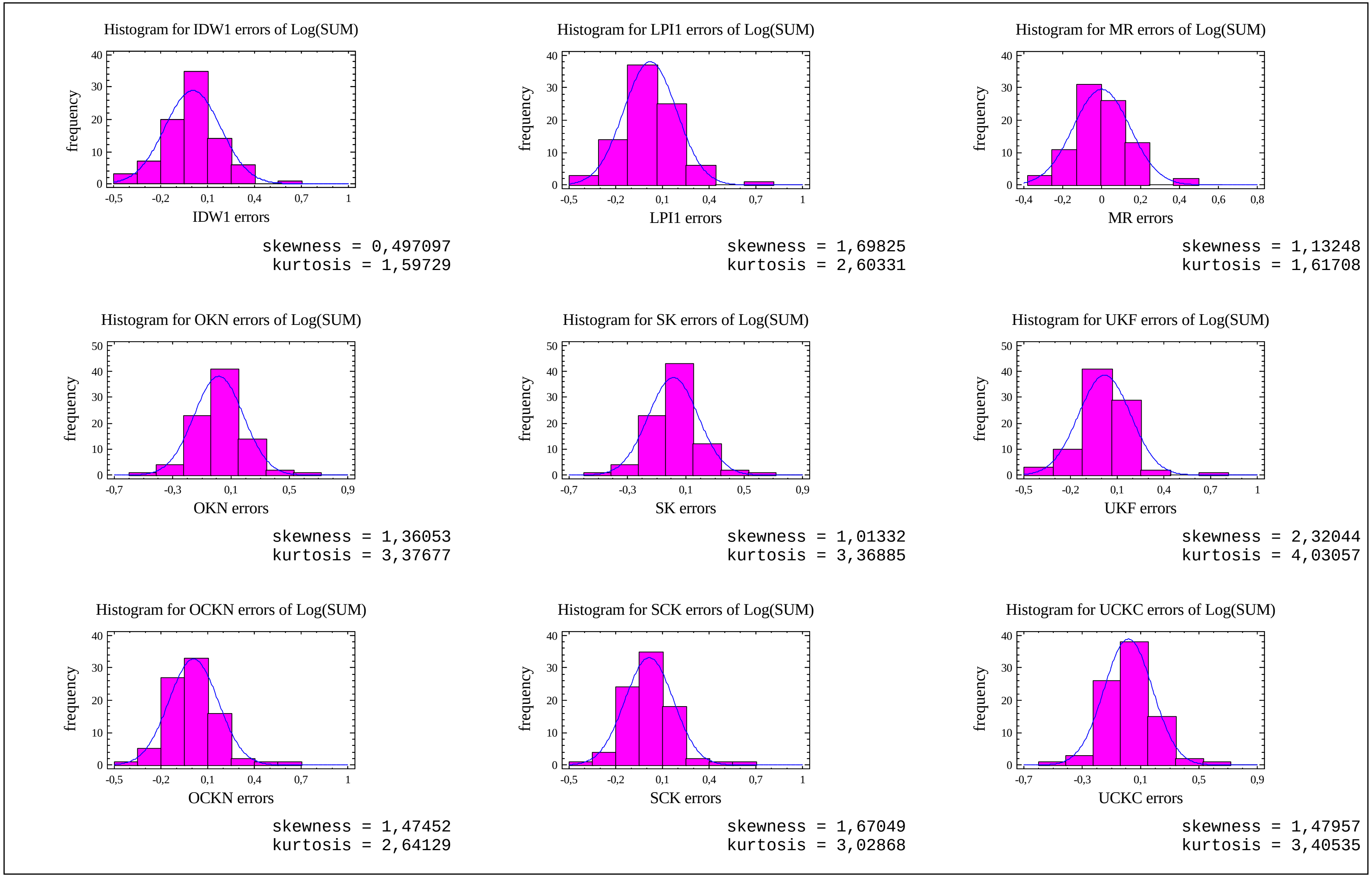


Figure 24.

Error histograms for the variable AUT

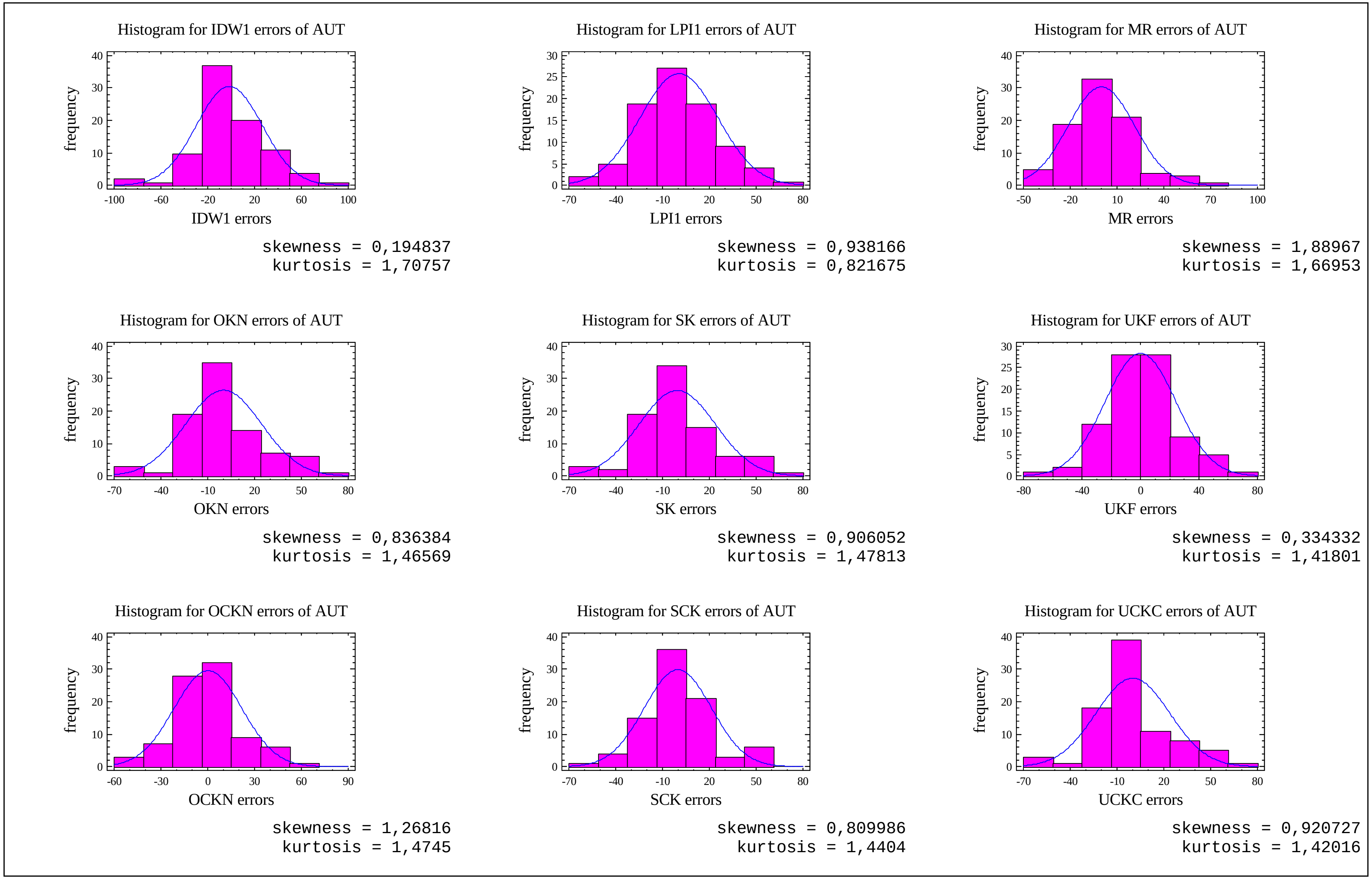
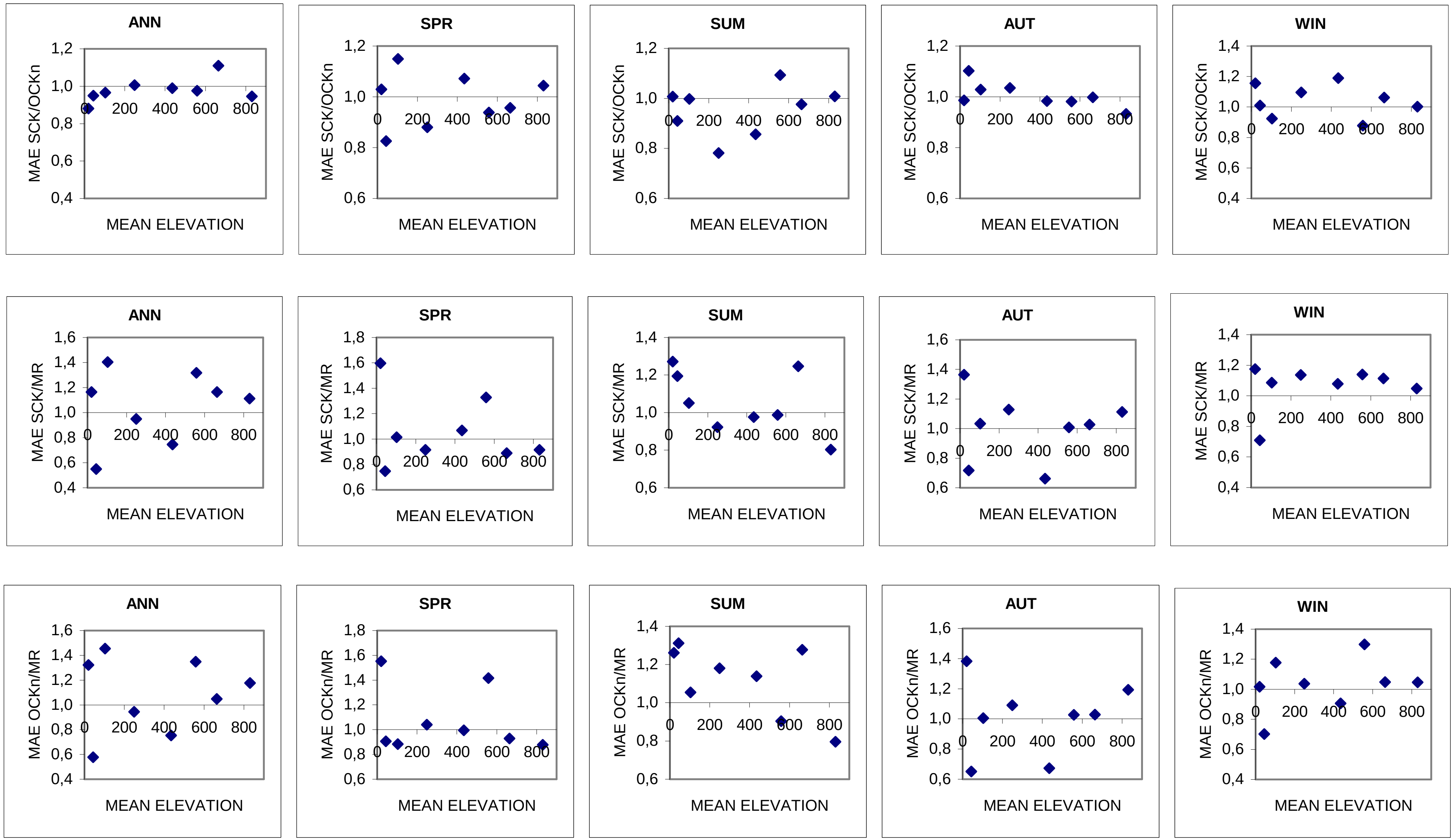


Figure 26.

Ratio "MAE SCK/OCKn", "MAE SCK/MR" and "MAE OCKn/MR" versus the mean elevation of the eight defined zones.



Analysis report 1.

Goodness-of-Fit Tests for ANN

Chi-Square Test

	Lower Limit	Upper Limit	Observed Frequency	Expected Frequency	Chi-Square
	at or below	309,381	10	7,17	1,12
	309,381	366,678	6	7,17	0,19
	366,678	407,067	6	7,17	0,19
	407,067	440,676	4	7,17	1,40
	440,676	471,05	8	7,17	0,10
	471,05	500,063	11	7,17	2,05
	500,063	529,076	6	7,17	0,19
	529,076	559,45	8	7,17	0,10
	559,45	593,059	8	7,17	0,10
	593,059	633,447	8	7,17	0,10
	633,447	690,745	2	7,17	3,72
above	690,745		9	7,17	0,47

Chi-Square = 9,7207 with 9 d.f.   P-Value = 0,373566

Estimated Kolmogorov statistic DPLUS = 0,0496973
Estimated Kolmogorov statistic DMINUS = 0,050014
Estimated overall statistic DN = 0,050014
Approximate P-Value = 0,982538

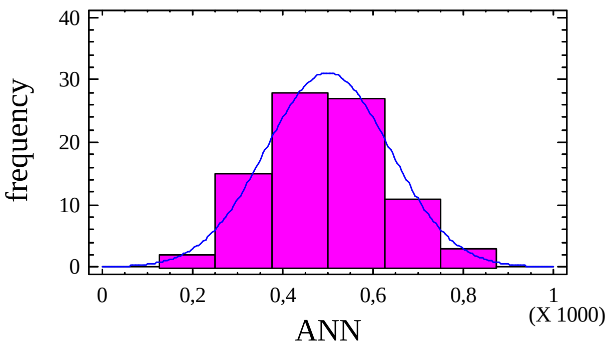
EDF Statistic	Value	Modified Form	P-Value
Kolmogorov-Smirnov D	0,050014	0,467894	>0.10*
Anderson-Darling A^2	0,357104	0,360327	0,4478*

*Indicates that the P-Value has been compared to tables of critical values specially constructed for fitting the currently selected distribution. Other P-values are based on general tables and may be very conservative.

The StatAdvisor

This pane shows the results of tests run to determine whether ANN can be adequately modeled by a normal distribution. The chi-square test divides the range of ANN into nonoverlapping intervals and compares the number of observations in each class to the number expected based on the fitted distribution. The Kolmogorov-Smirnov test computes the maximum distance between the cumulative distribution of ANU and the CDF of the fitted normal distribution. In this case, the maximum distance is 0,050014. The other EDF statistics compare the empirical distribution function to the fitted CDF in different ways.

Since the smallest P-value amongst the tests performed is greater than or equal to 0.10, **we can not reject the idea that ANN comes from a normal distribution with 90% or higher confidence**.



Analysis report 2.

Goodness-of-Fit Tests for SPR

Chi-Square Test

	Lower Limit	Upper Limit	Observed Frequency	Expected Frequency	Chi-Square
	at or below	81,63	6	7,17	0,19
	81,63	95,2617	9	7,17	0,47
	95,2617	104,871	7	7,17	0,00
	104,871	112,866	7	7,17	0,00
	112,866	120,093	8	7,17	0,10
	120,093	126,995	9	7,17	0,47
	126,995	133,898	7	7,17	0,00
	133,898	141,124	5	7,17	0,66
	141,124	149,12	8	7,17	0,10
	149,12	158,729	7	7,17	0,00
	158,729	172,361	6	7,17	0,19
above	172,361		7	7,17	0,00

Chi-Square = 2,1859 with 9 d.f.   P-Value = 0,988175

Estimated Kolmogorov statistic DPLUS = 0,0627799
Estimated Kolmogorov statistic DMINUS = 0,0323508
Estimated overall statistic DN = 0,0627799
Approximate P-Value = 0,886942

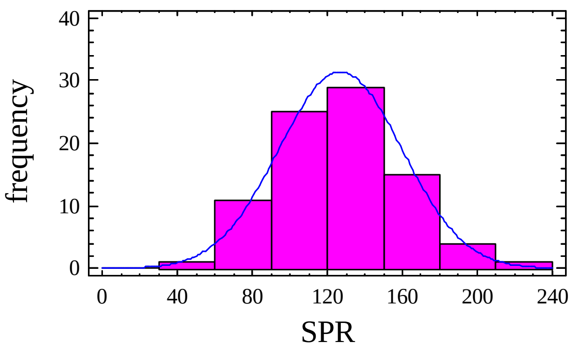
EDF Statistic	Value	Modified Form	P-Value
Kolmogorov-Smirnov D	0,0627799	0,587323	>0.10*
Anderson-Darling A^2	0,224467	0,226493	0,8173*

*Indicates that the P-Value has been compared to tables of critical values specially constructed for fitting the currently selected distribution. Other P-values are based on general tables and may be very conservative.

The StatAdvisor

This pane shows the results of tests run to determine whether SPR can be adequately modeled by a normal distribution. The chi-square test divides the range of SPR into nonoverlapping intervals and compares the number of observations in each class to the number expected based on the fitted distribution. The Kolmogorov-Smirnov test computes the maximum distance between the cumulative distribution of SPR and the CDF of the fitted normal distribution. In this case, the maximum distance is 0,0627799. The other EDF statistics compare the empirical distribution function to the fitted CDF in different ways.

Since the smallest P-value amongst the tests performed is greater than or equal to 0.10, **we can not reject the idea that SPR comes from a normal distribution with 90% or higher confidence**.



Analysis report 3.

Goodness-of-Fit Tests for Log(SUM)

Chi-Square Test

Lower Limit	Upper Limit	Observed Frequency	Expected Frequency	Chi-Square
at or below	3,50768	8	7,17	0,10
3,50768	3,66261	5	7,17	0,66
3,66261	3,77181	4	7,17	1,40
3,77181	3,86268	6	7,17	0,19
3,86268	3,94481	14	7,17	6,52
3,94481	4,02325	8	7,17	0,10
4,02325	4,1017	6	7,17	0,19
4,1017	4,18383	5	7,17	0,66
4,18383	4,2747	10	7,17	1,12
4,2747	4,3839	6	7,17	0,19
4,3839	4,53882	8	7,17	0,10
above 4,53882		6	7,17	0,19

Chi-Square = 11,3954 with 9 d.f.   P-Value = 0,249575

Estimated Kolmogorov statistic DPLUS = 0,0364084
Estimated Kolmogorov statistic DMINUS = 0,0715205
Estimated overall statistic DN = 0,0715205
Approximate P-Value = 0,771205

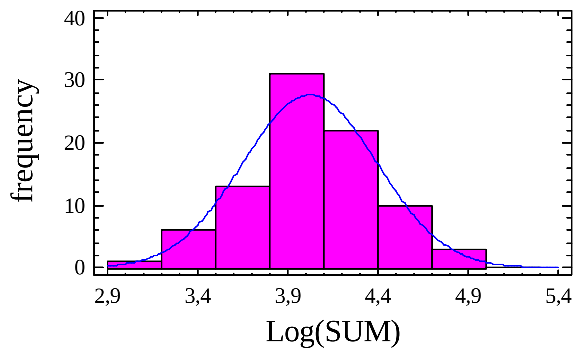
EDF Statistic	Value	Modified Form	P-Value
Kolmogorov-Smirnov D	0,0715205	0,669094	>0.10*
Anderson-Darling A^2	0,341687	0,34477	0,4860*

*Indicates that the P-Value has been compared to tables of critical values specially constructed for fitting the currently selected distribution. Other P-values are based on general tables and may be very conservative.

The StatAdvisor

This pane shows the results of tests run to determine whether Log(SUM) can be adequately modeled by a normal distribution. The chi-square test divides the range of Log(SUM) into nonoverlapping intervals and compares the number of observations in each class to the number expected based on the fitted distribution. The Kolmogorov-Smirnov test computes the maximum distance between the cumulative distribution of Log(SUM) and the CDF of the fitted normal distribution. In this case, the maximum distance is 0,0715205. The other EDF statistics compare the empirical distribution function to the fitted CDF in different ways.

Since the smallest P-value amongst the tests performed is greater than or equal to 0.10, **we can not reject the idea that Log(SUM) comes from a normal distribution with 90% or higher confidence**.



Analysis report 4.

Goodness-of-Fit Tests for AUT

Chi-Square Test

	Lower Limit	Upper Limit	Observed Frequency	Expected Frequency	Chi-Square
	at or below	112,218	9	7,17	0,47
	112,218	135,816	11	7,17	2,05
	135,816	152,45	6	7,17	0,19
	152,45	166,292	5	7,17	0,66
	166,292	178,802	5	7,17	0,66
	178,802	190,751	5	7,17	0,66
	190,751	202,7	4	7,17	1,40
	202,7	215,21	9	7,17	0,47
	215,21	229,052	11	7,17	2,05
	229,052	245,686	8	7,17	0,10
	245,686	269,284	5	7,17	0,66
above	269,284		8	7,17	0,10

Chi-Square = 9,44229 with 9 d.f.   P-Value = 0,397488

Estimated Kolmogorov statistic DPLUS = 0,0789051
Estimated Kolmogorov statistic DMINUS = 0,0669039
Estimated overall statistic DN = 0,0789051
Approximate P-Value = 0,65796

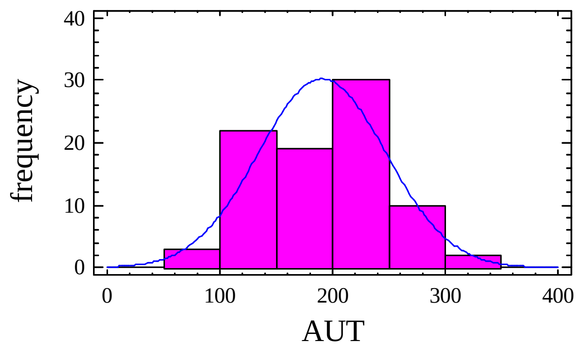
EDF Statistic	Value	Modified Form	P-Value
Kolmogorov-Smirnov D	0,0789051	0,738179	>0.10*
Anderson-Darling A^2	0,580859	0,586102	0,1269*

*Indicates that the P-Value has been compared to tables of critical values specially constructed for fitting the currently selected distribution. Other P-values are based on general tables and may be very conservative.

The StatAdvisor

This pane shows the results of tests run to determine whether AUT can be adequately modeled by a normal distribution. The chi-square test divides the range of AUT into nonoverlapping intervals and compares the number of observations in each class to the number expected based on the fitted distribution. The Kolmogorov-Smirnov test computes the maximum distance between the cumulative distribution of AUT and the CDF of the fitted normal distribution. In this case, the maximum distance is 0,0789051. The other EDF statistics compare the empirical distribution function to the fitted CDF in different ways.

Since the smallest P-value amongst the tests performed is greater than or equal to 0.10, **we can not reject the idea that AUT comes from a normal distribution with 90% or higher confidence**.



Analysis report 5.

Goodness-of-Fit Tests for Log(WIN)

Chi-Square Test

	Lower Limit	Upper Limit	Observed Frequency	Expected Frequency	Chi-Square
	at or below	4,2383	10	7,17	1,12
	4,2383	4,38985	3	7,17	2,42
	4,38985	4,49667	9	7,17	0,47
	4,49667	4,58556	4	7,17	1,40
	4,58556	4,6659	8	7,17	0,10
	4,6659	4,74264	12	7,17	3,26
	4,74264	4,81937	6	7,17	0,19
	4,81937	4,89971	6	7,17	0,19
	4,89971	4,9886	7	7,17	0,00
	4,9886	5,09543	6	7,17	0,19
	5,09543	5,24698	6	7,17	0,19
above	5,24698		9	7,17	0,47

Chi-Square = 9,99964 with 9 d.f.   P-Value = 0,350514

Estimated Kolmogorov statistic DPLUS = 0,0452133
Estimated Kolmogorov statistic DMINUS = 0,0414467
Estimated overall statistic DN = 0,0452133
Approximate P-Value = 0,994643

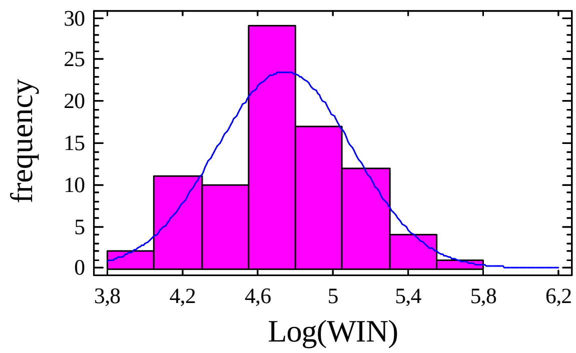
EDF Statistic	Value	Modified Form	P-Value
Kolmogorov-Smirnov D	0,0452133	0,422983	>0.10*
Anderson-Darling A^2	0,24635	0,248574	0,7493*

*Indicates that the P-Value has been compared to tables of critical values specially constructed for fitting the currently selected distribution. Other P-values are based on general tables and may be very conservative.

The StatAdvisor

This pane shows the results of tests run to determine whether Log(WIN) can be adequately modeled by a normal distribution. The chi-square test divides the range of Log(WIN) into nonoverlapping intervals and compares the number of observations in each class to the number expected based on the fitted distribution. The Kolmogorov-Smirnov test computes the maximum distance between the cumulative distribution of Log(WIN) and the CDF of the fitted normal distribution. In this case, the maximum distance is 0,0452133. The other EDF statistics compare the empirical distribution function to the fitted CDF in different ways.

Since the smallest P-value amongst the tests performed is greater than or equal to 0.10, **we can not reject the idea that Log(WIN) comes from a normal distribution with 90% or higher confidence**.



Analysis report 6.

Multiple regression (MR1) analysis for ANN

Dependent variable: ANN

Parameter	Estimate	Standard Error	T Statistic	P-Value
CONSTANT	-270633,0	39279,7	-6,8899	0,0000
X	-0,080805	0,0264062	-3,06007	0,0032
Y	0,138581	0,0212635	6,5173	0,0000
D5000	-11,7718	2,36757	-4,97209	0,0000
X2	5,79469E-8	1,77114E-8	3,27172	0,0017
Y2	-1,60508E-8	2,45777E-9	-6,53063	0,0000
COAST_2	-1,62541E-7	3,78607E-8	-4,29313	0,0001
D5000_2	0,000496298	0,000242424	2,04723	0,0445
X_Z5000	0,00000630851	0,00000288781	2,18453	0,0324
X_S5000	0,0000523116	0,0000159655	3,27654	0,0017
Y_COAST	2,5411E-9	5,74158E-10	4,42579	0,0000
Y_Z5000	-0,000001347	4,97105E-7	-2,70969	0,0085
Y_D5000	0,0000026801	5,40518E-7	4,9584	0,0000
COAST_Z500	0,0000227797	0,00000561039	4,06027	0,0001
COAST_S500	-0,000551525	0,000137573	-4,00897	0,0002
Z5000_S500	0,0455219	0,0153057	2,97418	0,0041
S5000_D500	-0,0431938	0,0188663	-2,28947	0,0252
C_E	103,123	45,2252	2,28022	0,0257

Analysis of Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	1,34856E6	17	79327,1	20,18	0,0000
Residual	267272,0	68	3930,47		
Total (Corr.)	1,61583E6	85			

R-squared = 83,4592 percent
R-squared (adjusted for d.f.) = 79,324 percent
Standard Error of Est. = 62,6935
Mean absolute error = 42,4192
Durbin-Watson statistic = 2,03343

The StatAdvisor

The output shows the results of fitting a multiple linear regression model to describe the relationship between ANN and 17 independent variables. The equation of the fitted model is

ANN = -270633,0 - 0,080805*X + 0,138581*Y - 11,7718*D5000 + 5,79469E-8*X2 - 1,60508E-8*Y2 - 1,62541E-7*COAST_2 + 0,000496298*D5000_2 + 0,00000630851*X_Z5000 + 0,0000523116*X_S5000 + 2,5411E-9*Y_COAST - 0,000001347*Y_Z5000 + 0,0000026801*Y_D5000 + 0,0000227797*COAST_Z500 - 0,000551525*COAST_S500 + 0,0455219*Z5000_S500 - 0,0431938*S5000_D500 + 103,123*C_E

Since the P-value in the ANOVA table is less than 0.01, there is a statistically significant relationship between the variables at the 99% confidence level.

The R-Squared statistic indicates that the model as fitted explains 83,4592% of the variability in ANN. The adjusted R-squared statistic, which is more suitable for comparing models with different numbers of independent variables, is 79,324%. The standard error of the estimate shows the standard deviation of the residuals to be 62,6935. This value can be used to construct prediction limits for new observations by selecting the Reports option from the text menu. The mean absolute error (MAE) of 42,4192 is the average value of the residuals. The Durbin-Watson (DW) statistic tests the residuals to determine if there is any significant correlation based on the order in which they occur in your data file. Since the DW value is greater than 1.4, there is probably not any serious autocorrelation in the residuals.

In determining whether the model can be simplified, notice that the highest P-value on the independent variables is 0,0445, belonging to D5000_2. Since the P-value is less than 0.05, that term is statistically significant at the 95% confidence level. Consequently, you probably don't want to remove any variables from the model.

Analysis report 7.

Multiple regression (MR1) analysis for SPR

Dependent variable: SPR

Parameter	Estimate	Standard Error	T Statistic	P-Value
CONSTANT	-35952,3	7385,28	-4,8681	0,0000
Y	0,016609	0,00342318	4,85191	0,0000
D5000	-1,83015	0,480652	-3,80764	0,0003
X2	1,06739E-9	1,22355E-10	8,72373	0,0000
Y2	-1,94441E-9	3,9763E-10	-4,89001	0,0000
COAST_2	-1,64166E-8	6,57816E-9	-2,49562	0,0149
X_S5000	0,00000404452	0,00000160619	2,51809	0,0140
Y_COAST	8,04909E-10	1,27996E-10	6,28857	0,0000
Y_Z5000	-4,15538E-8	1,21406E-8	-3,4227	0,0010
Y_D5000	4,25669E-7	1,11792E-7	3,80769	0,0003
COAST_Z500	0,00000191704	9,15392E-7	2,09422	0,0398
COAST_S500	-0,000120646	0,0000295372	-4,08455	0,0001
Z5000_S500	0,00682541	0,00257514	2,6505	0,0099
C_E	25,5775	10,8112	2,36583	0,0207

Analysis of Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	74316,2	13	5716,63	24,01	0,0000
Residual	17142,6	72	238,092		
Total (Corr.)	91458,9	85			

R-squared = 81,2564 percent
R-squared (adjusted for d.f.) = 77,8722 percent
Standard Error of Est. = 15,4302
Mean absolute error = 10,4151
Durbin-Watson statistic = 2,30082

The StatAdvisor

The output shows the results of fitting a multiple linear regression model to describe the relationship between SPR and 13 independent variables. The equation of the fitted model is

SPR = -35952,3 + 0,016609*Y - 1,83015*D5000 + 1,06739E-9*X2 - 1,94441E-9*Y2 - 1,64166E-8*COAST_2 + 0,00000404452*X_S5000 + 8,04909E-10*Y_COAST - 4,15538E-8*Y_Z5000 + 4,25669E-7*Y_D5000 + 0,00000191704*COAST_Z500 - 0,000120646*COAST_S500 + 0,00682541*Z5000_S500 + 25,5775*C_E

Since the P-value in the ANOVA table is less than 0.01, there is a statistically significant relationship between the variables at the 99% confidence level.

The R-Squared statistic indicates that the model as fitted explains 81,2564% of the variability in SPR. The adjusted R-squared statistic, which is more suitable for comparing models with different numbers of independent variables, is 77,8722%. The standard error of the estimate shows the standard deviation of the residuals to be 15,4302. This value can be used to construct prediction limits for new observations by selecting the Reports option from the text menu. The mean absolute error (MAE) of 10,4151 is the average value of the residuals. The Durbin-Watson (DW) statistic tests the residuals to determine if there is any significant correlation based on the order in which they occur in your data file. Since the DW value is greater than 1.4, there is probably not any serious autocorrelation in the residuals.

In determining whether the model can be simplified, notice that the highest P-value on the independent variables is 0,0398, belonging to COAST_Z500. Since the P-value is less than 0.05, that term is statistically significant at the 95% confidence level. Consequently, you probably don't want to remove any variables from the model.

Analysis report 8.

Multiple regression (MR1) analysis for Log(SUM)

Dependent variable: Log(SUM)

Parameter	Estimate	Standard Error	T Statistic	P-Value
CONSTANT	-2,98512	0,764849	-3,90289	0,0002
COAST	0,000200605	0,0000513032	3,91018	0,0002
D5000	-0,0124349	0,00280962	-4,42581	0,0000
COAST_2	-3,38442E-10	8,5326E-11	-3,96646	0,0002
X_Y	2,07465E-12	0,0	8,89868	0,0000
X_COAST	-2,42296E-10	6,78878E-11	-3,56906	0,0006
Y_D5000	2,89138E-9	6,52958E-10	4,42812	0,0000
C_NE	-883,28	305,26	-2,89353	0,0050
C_E	1249,18	431,706	2,89358	0,0050
C_SE	-883,21	305,259	-2,89332	0,0050

Analysis of Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	10,0211	9	1,11346	47,23	0,0000
Residual	1,79166	76	0,0235745		
Total (Corr.)	11,8128	85			

R-squared = 84,8328 percent
R-squared (adjusted for d.f.) = 83,0367 percent
Standard Error of Est. = 0,15354
Mean absolute error = 0,112051
Durbin-Watson statistic = 2,03447

The StatAdvisor

The output shows the results of fitting a multiple linear regression model to describe the relationship between Log(SUM) and 9 independent variables. The equation of the fitted model is

Log(SUM) = -2,98512 + 0,000200605*COAST - 0,0124349*D5000 - 3,38442E-10*COAST_2 + 2,07465E-12*X_Y - 2,42296E-10*X_COAST + 2,89138E-9*Y_D5000 - 883,28*C_NE + 1249,18*C_E - 883,21*C_SE

Since the P-value in the ANOVA table is less than 0.01, there is a statistically significant relationship between the variables at the 99% confidence level.

The R-Squared statistic indicates that the model as fitted explains 84,8328% of the variability in Log(SUM). The adjusted R-squared statistic, which is more suitable for comparing models with different numbers of independent variables, is 83,0367%. The standard error of the estimate shows the standard deviation of the residuals to be 0,15354. This value can be used to construct prediction limits for new observations by selecting the Reports option from the text menu. The mean absolute error (MAE) of 0,112051 is the average value of the residuals. The Durbin-Watson (DW) statistic tests the residuals to determine if there is any significant correlation based on the order in which they occur in your data file. Since the DW value is greater than 1.4, there is probably not any serious autocorrelation in the residuals.

In determining whether the model can be simplified, notice that the highest P-value on the independent variables is 0,0050, belonging to C_SE. Since the P-value is less than 0.01, the highest order term is statistically significant at the 99% confidence level. Consequently, you probably don't want to remove any variables from the model.

Analysis report 9.

Multiple regression (MR1) analysis for AUT

Dependent variable: AUT

Parameter	Estimate	Standard Error	T Statistic	P-Value
CONSTANT	-167400,0	24992,5	-6,698	0,0000
X	-0,0787977	0,0277673	-2,83778	0,0061
Y	0,0906509	0,0157834	5,74345	0,0000
COAST	-0,0858838	0,0371391	-2,31249	0,0239
D5000	-3,98643	1,02123	-3,90354	0,0002
Y2	-1,20699E-8	2,33471E-9	-5,16978	0,0000
COAST_2	-1,27747E-7	3,34534E-8	-3,81867	0,0003
S5000_2	0,939207	0,341513	2,75013	0,0077
D5000_2	0,000337216	0,000123124	2,73883	0,0079
X_Y	1,86314E-8	6,38971E-9	2,91584	0,0049
X_COAST	-1,09277E-7	2,8452E-8	-3,84073	0,0003
X_Z5000	0,00000651219	0,00000160246	4,06388	0,0001
X_S5000	0,0000174433	0,00000660046	2,64274	0,0103
Y_COAST	3,87059E-8	9,83387E-9	3,93598	0,0002
Y_Z5000	-0,000001145	2,75945E-7	-4,14939	0,0001
Y_D5000	9,00106E-7	2,3354E-7	3,85418	0,0003
COAST_Z500	0,00000907769	0,00000267072	3,39897	0,0012
COAST_S500	-0,000219527	0,0000733389	-2,99333	0,0039
COAST_D500	0,00000398108	0,00000185905	2,14146	0,0360
S5000_D500	-0,0399968	0,0125571	-3,1852	0,0022
C_E	41,0429	20,0291	2,04916	0,0445

Analysis of Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	235747,0	20	11787,3	19,99	0,0000
Residual	38336,2	65	589,788		
Total (Corr.)	274083,0	85			

R-squared = 86,0129 percent
R-squared (adjusted for d.f.) = 81,7092 percent
Standard Error of Est. = 24,2856
Mean absolute error = 16,1704
Durbin-Watson statistic = 1,81469

The StatAdvisor

The output shows the results of fitting a multiple linear regression model to describe the relationship between AUT and 20 independent variables. The equation of the fitted model is

AUT = -167400,0 - 0,0787977*X + 0,0906509*Y - 0,0858838*COAST - 3,98643*D5000 - 1,20699E-8*Y2 - 1,27747E-7*COAST_2 + 0,939207*S5000_2 + 0,000337216*D5000_2 + 1,86314E-8*X_Y - 1,09277E-7*X_COAST + 0,00000651219*X_Z5000 + 0,0000174433*X_S5000 + 3,87059E-8*Y_COAST - 0,000001145*Y_Z5000 + 9,00106E-7*Y_D5000 + 0,00000907769*COAST_Z500 - 0,000219527*COAST_S500 + 0,00000398108*COAST_D500 - 0,0399968*S5000_D500 + 41,0429*C_E

Since the P-value in the ANOVA table is less than 0.01, there is a statistically significant relationship between the variables at the 99% confidence level.

The R-Squared statistic indicates that the model as fitted explains 86,0129% of the variability in AUT. The adjusted R-squared statistic, which is more suitable for comparing models with different numbers of independent variables, is 81,7092%. The standard error of the estimate shows the standard deviation of the residuals to be 24,2856. This value can be used to construct prediction limits for new observations by selecting the Reports option from the text menu.
The mean absolute error (MAE) of 16,1704 is the average value of the residuals. The Durbin-Watson (DW) statistic tests the residuals to determine if there is any significant correlation based on the order in which they occur in your data file. Since the DW value is greater than 1.4, there is probably not any serious autocorrelation in the residuals.

In determining whether the model can be simplified, notice that the highest P-value on the independent variables is 0,0445, belonging to C_E. Since the P-value is less than 0.05, that term is statistically significant at the 95% confidence level. Consequently, you probably don't want to remove any variables from the model.

Analysis report 10.

Multiple regression (MR1) analysis for Log(WIN)

Dependent variable: Log(WIN)

Parameter	Estimate	Standard Error	T Statistic	P-Value
CONSTANT	-662,571	119,862	-5,52779	0,0000
X	-0,000235139	0,0000763549	-3,07956	0,0030
Y	0,000348092	0,0000642794	5,4153	0,0000
COAST	-0,000370399	0,00014402	-2,57186	0,0124
Z5000	0,0245771	0,0112735	2,18007	0,0328
S5000	0,126993	0,0347966	3,64958	0,0005
D5000	-0,0364857	0,00813294	-4,48616	0,0000
X2	1,64852E-10	5,11595E-11	3,22231	0,0020
Y2	-4,03316E-11	7,41391E-12	-5,43999	0,0000
COAST_2	-5,13358E-10	1,0403E-10	-4,93469	0,0000
D5000_2	0,00000176573	6,82329E-7	2,5878	0,0119
X_Z5000	2,71889E-8	8,41425E-9	3,23129	0,0019
Y_COAST	9,32286E-11	3,35877E-11	2,77568	0,0072
Y_Z5000	-1,13022E-8	3,01568E-9	-3,74781	0,0004
Y_D5000	8,2629E-9	1,855E-9	4,45438	0,0000
COAST_Z500	8,33307E-8	1,59624E-8	5,22045	0,0000
COAST_S500	-0,00000237072	5,02033E-7	-4,72224	0,0000
Z5000_S500	0,000194182	0,0000480462	4,04156	0,0001
S5000_D500	-0,000151714	0,0000529699	-2,86416	0,0056
C_N	0,124824	0,0490546	2,5446	0,0133

Analysis of Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	9,49574	19	0,499776	18,24	0,0000
Residual	1,80798	66	0,0273936		
Total (Corr.)	11,3037	85			

R-squared = 84,0055 percent
R-squared (adjusted for d.f.) = 79,401 percent
Standard Error of Est. = 0,16551
Mean absolute error = 0,112669
Durbin-Watson statistic = 2,03808

The StatAdvisor

The output shows the results of fitting a multiple linear regression model to describe the relationship between Log(WIN) and 19 independent variables. The equation of the fitted model is

Log(WIN) = -662,571 - 0,000235139*X + 0,000348092*Y - 0,000370399*COAST + 0,0245771*Z5000 + 0,126993*S5000 - 0,0364857*D5000 + 1,64852E-10*X2 - 4,03316E-11*Y2 - 5,13358E-10*COAST_2 + 0,00000176573*D5000_2 + 2,71889E-8*X_Z5000 + 9,32286E-11*Y_COAST - 1,13022E-8*Y_Z5000 + 8,2629E-9*Y_D5000 + 8,33307E-8*COAST_Z500 - 0,00000237072*COAST_S500 + 0,000194182*Z5000_S500 - 0,000151714*S5000_D500 + 0,124824*C_N

Since the P-value in the ANOVA table is less than 0.01, there is a statistically significant relationship between the variables at the 99% confidence level.

The R-Squared statistic indicates that the model as fitted explains 84,0055% of the variability in Log(WIN). The adjusted R-squared statistic, which is more suitable for comparing models with different numbers of independent variables, is 79,401%. The standard error of the estimate shows the standard deviation of the residuals to be 0,16551. This value can be used to construct prediction limits for new observations by selecting the Reports option from the text menu.
The mean absolute error (MAE) of 0,112669 is the average value of the residuals. The Durbin-Watson (DW) statistic tests the residuals to determine if there is any significant correlation based on the order in which they occur in your data file. Since the DW value is greater than 1.4, there is probably not any serious autocorrelation in the residuals.

In determining whether the model can be simplified, notice that the highest P-value on the independent variables is 0,0328, belonging to Z5000. Since the P-value is less than 0.05, that term is statistically significant at the 95% confidence level. Consequently, you probably don't want to remove any variables from the model.

Analysis report 11.

Multiple regression (MR2) analysis for ANN

Dependent variable: ANN

Parameter	Estimate	Standard Error	T Statistic	P-Value
CONSTANT	-229152,0	39166,6	-5,85071	0,0000
X	-0,071589	0,0263516	-2,71668	0,0083
Y	0,117719	0,0211636	5,56233	0,0000
D5000	-10,7124	2,34992	-4,55862	0,0000
X2	5,02932E-8	1,7683E-8	2,84416	0,0058
Y2	-1,35827E-8	2,44329E-9	-5,55916	0,0000
COAST_2	-1,25058E-7	3,67879E-8	-3,39944	0,0011
X_Z5000	0,00000968188	0,00000289039	3,34968	0,0013
X_S5000	0,0000247528	0,00000699215	3,54009	0,0007
Y_COAST	1,50963E-9	4,30647E-10	3,50549	0,0008
Y_Z5000	-0,00000179975	5,05313E-7	-3,56166	0,0007
Y_D5000	0,00000249005	5,47522E-7	4,54785	0,0000
COAST_Z500	0,0000214058	0,00000570829	3,74995	0,0004
COAST_S500	-0,000264522	0,0000857661	-3,08422	0,0029

Analysis of Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	1,28816E6	13	99088,9	21,77	0,0000
Residual	327678,0	72	4551,08		
Total (Corr.)	1,61583E6	85			

R-squared = 79,7208 percent
R-squared (adjusted for d.f.) = 76,0593 percent
Standard Error of Est. = 67,4617
Mean absolute error = 49,0007
Durbin-Watson statistic = 1,99479

The StatAdvisor

The output shows the results of fitting a multiple linear regression model to describe the relationship between ANU and 13 independent variables. The equation of the fitted model is

ANN = -229152,0 - 0,071589*X + 0,117719*Y - 10,7124*D5000 + 5,02932E-8*X2 - 1,35827E-8*Y2 - 1,25058E-7*COAST_2 + 0,00000968188*X_Z5000 + 0,0000247528*X_S5000 + 1,50963E-9*Y_COAST - 0,00000179975*Y_Z5000 + 0,00000249005*Y_D5000 + 0,0000214058*COAST_Z500 - 0,000264522*COAST_S500

Since the P-value in the ANOVA table is less than 0.01, there is a statistically significant relationship between the variables at the 99% confidence level.

The R-Squared statistic indicates that the model as fitted explains 79,7208% of the variability in ANN. The adjusted R-squared statistic, which is more suitable for comparing models with different numbers of independent variables, is 76,0593%. The standard error of the estimate shows the standard deviation of the residuals to be 67,4617. This value can be used to construct prediction limits for new observations by selecting the Reports option from the text menu. The mean absolute error (MAE) of 49,0007 is the average value of the residuals. The Durbin-Watson (DW) statistic tests the residuals to determine if there is any significant correlation based on the order in which they occur in your data file. Since the DW value is greater than 1.4, there is probably not any serious autocorrelation in the residuals.

In determining whether the model can be simplified, notice that the highest P-value on the independent variables is 0,0083, belonging to X. Since the P-value is less than 0.01, the highest order term is statistically significant at the 99% confidence level. Consequently, you probably don't want to remove any variables from the model.

Analysis report 12.

Multiple regression (MR2) analysis for SPR

Dependent variable: SPR

Parameter	Estimate	Standard Error	T Statistic	P-Value
CONSTANT	-30731,1	7266,35	-4,22923	0,0001
Y	0,01418	0,00336684	4,21165	0,0001
D5000	-1,63755	0,488396	-3,35291	0,0013
X2	9,04459E-10	1,04268E-10	8,67439	0,0000
Y2	-1,65632E-9	3,90261E-10	-4,24413	0,0001
COAST_Z500	0,00000199429	9,43175E-7	2,11444	0,0379
COAST_2	-1,8526E-8	6,71954E-9	-2,75704	0,0074
X_S5000	0,00000419441	0,0000016547	2,53484	0,0134
Y_COAST	8,521E-10	1,30352E-10	6,53691	0,0000
Y_Z5000	-4,73759E-8	1,22572E-8	-3,86514	0,0002
Y_D5000	3,81403E-7	1,13632E-7	3,35647	0,0013
COAST_S500	-0,000135737	0,0000297345	-4,56498	0,0000
Z5000_S500	0,00798081	0,0026068	3,06153	0,0031

Analysis of Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	72983,6	12	6081,96	24,03	0,0000
Residual	18475,3	73	253,086		
Total (Corr.)	91458,9	85			

R-squared = 79,7993 percent
R-squared (adjusted for d.f.) = 76,4787 percent
Standard Error of Est. = 15,9087
Mean absolute error = 10,6607
Durbin-Watson statistic = 2,32581

The StatAdvisor

The output shows the results of fitting a multiple linear regression model to describe the relationship between SPR and 12 independent variables. The equation of the fitted model is

SPR = -30731,1 + 0,01418*Y - 1,63755*D5000 + 9,04459E-10*X2 - 1,65632E-9*Y2 + 0,00000199429*COAST_Z500 - 1,8526E-8*COAST_2 + 0,00000419441*X_S5000 + 8,521E-10*Y_COAST - 4,73759E-8*Y_Z5000 + 3,81403E-7*Y_D5000 - 0,000135737*COAST_S500 + 0,00798081*Z5000_S500

Since the P-value in the ANOVA table is less than 0.01, there is a statistically significant relationship between the variables at the 99% confidence level.

The R-Squared statistic indicates that the model as fitted explains 79,7993% of the variability in SPR. The adjusted R-squared statistic, which is more suitable for comparing models with different numbers of independent variables, is 76,4787%. The standard error of the estimate shows the standard deviation of the residuals to be 15,9087. This value can be used to construct prediction limits for new observations by selecting the Reports option from the text menu. The mean absolute error (MAE) of 10,6607 is the average value of the residuals. The Durbin-Watson (DW) statistic tests the residuals to determine if there is any significant correlation based on the order in which they occur in your data file. Since the DW value is greater than 1.4, there is probably not any serious autocorrelation in the residuals.

In determining whether the model can be simplified, notice that the highest P-value on the independent variables is 0,0379, belonging to COAST_Z500. Since the P-value is less than 0.05, that term is statistically significant at the 95% confidence level. Consequently, you probably don't want to remove any variables from the model.

Analysis report 13.

Multiple regression (MR2) analysis for Log(SUM)

Dependent variable: Log(SUM)

Parameter	Estimate	Standard Error	T Statistic	P-Value
CONSTANT	-1,26558	0,69599	-1,8184	0,0727
Y_D5000	3,18242E-9	6,74923E-10	4,71523	0,0000
D5000	-0,0137019	0,00289791	-4,7282	0,0000
COAST	0,0000128058	0,00000147382	8,68884	0,0000
X_Y	1,57906E-12	0,0	7,20561	0,0000

Analysis of Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	9,39973	4	2,34993	78,88	0,0000
Residual	2,41305	81	0,0297907		
Total (Corr.)	11,8128	85			

R-squared = 79,5726 percent
R-squared (adjusted for d.f.) = 78,5638 percent
Standard Error of Est. = 0,1726
Mean absolute error = 0,133646
Durbin-Watson statistic = 2,0274

The StatAdvisor

The output shows the results of fitting a multiple linear regression model to describe the relationship between Log(SUM) and 4 independent variables. The equation of the fitted model is

Log(SUM) = -1,26558 + 3,18242E-9*Y_D5000 - 0,0137019*D5000 + 0,0000128058*COAST + 1,57906E-12*X_Y

Since the P-value in the ANOVA table is less than 0.01, there is a statistically significant relationship between the variables at the 99% confidence level.

The R-Squared statistic indicates that the model as fitted explains 79,5726% of the variability in Log(SUM). The adjusted R-squared statistic, which is more suitable for comparing models with different numbers of independent variables, is 78,5638%. The standard error of the estimate shows the standard deviation of the residuals to be 0,1726. This value can be used to construct prediction limits for new observations by selecting the Reports option from the text menu. The mean absolute error (MAE) of 0,133646 is the average value of the residuals. The Durbin-Watson (DW) statistic tests the residuals to determine if there is any significant correlation based on the order in which they occur in your data file. Since the DW value is greater than 1.4, there is probably not any serious autocorrelation in the residuals.

In determining whether the model can be simplified, notice that the highest P-value on the independent variables is 0,0000, belonging to Y_D5000. Since the P-value is less than 0.01, the highest order term is statistically significant at the 99% confidence level. Consequently, you probably don't want to remove any variables from the model.

Analysis report 14.

Multiple regression (MR2) analysis for AUT

Dependent variable: AUT

Parameter	Estimate	Standard Error	T Statistic	P-Value
CONSTANT	-84457,7	12618,1	-6,69336	0,0000
Y	0,038908	0,00584704	6,6543	0,0000
D5000	-3,82883	0,856813	-4,46869	0,0000
Y2	-4,52595E-9	6,79036E-10	-6,66525	0,0000
COAST_2	-6,69212E-8	2,83166E-8	-2,36332	0,0208
S5000_2	0,256202	0,0849752	3,01502	0,0035
X_Y	3,29493E-10	8,80375E-11	3,74265	0,0004
Y_COAST	7,55083E-9	3,98377E-9	1,8954	0,0621
Y_Z5000	-6,9202E-7	2,75157E-7	-2,515	0,0141
Y_D5000	8,92938E-7	1,98898E-7	4,48943	0,0000
COAST_Z500	0,00000646715	0,00000226924	2,84992	0,0057
COAST_S500	-0,0000988261	0,0000357195	-2,76673	0,0072
X_Z5000	0,00000374397	0,00000159929	2,34101	0,0220
X_COAST	-3,98292E-8	2,31241E-8	-1,72242	0,0893

Analysis of Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	219760,0	13	16904,6	22,41	0,0000
Residual	54323,1	72	754,488		
Total (Corr.)	274083,0	85			

R-squared = 80,18 percent
R-squared (adjusted for d.f.) = 76,6014 percent
Standard Error of Est. = 27,4679
Mean absolute error = 19,3855
Durbin-Watson statistic = 2,09528

The StatAdvisor

The output shows the results of fitting a multiple linear regression model to describe the relationship between AUT and 13 independent variables. The equation of the fitted model is

AUT = -84457,7 + 0,038908*Y - 3,82883*D5000 - 4,52595E-9*Y2 - 6,69212E-8*COAST_2 + 0,256202*S5000_2 + 3,29493E-10*X_Y + 7,55083E-9*Y_COAST - 6,9202E-7*Y_Z5000 + 8,92938E-7*Y_D5000 + 0,00000646715*COAST_Z500 - 0,0000988261*COAST_S500 + 0,00000374397*X_Z5000 - 3,98292E-8*X_COAST

Since the P-value in the ANOVA table is less than 0.01, there is a statistically significant relationship between the variables at the 99% confidence level.

The R-Squared statistic indicates that the model as fitted explains 80,18% of the variability in AUT. The adjusted R-squared statistic, which is more suitable for comparing models with different numbers of independent variables, is 76,6014%. The standard error of the estimate shows the standard deviation of the residuals to be 27,4679. This value can be used to construct prediction limits for new observations by selecting the Reports option from the text menu. The mean absolute error (MAE) of 19,3855 is the average value of the residuals. The Durbin-Watson (DW) statistic tests the residuals to determine if there is any significant correlation based on the order in which they occur in your data file. Since the DW value is greater than 1.4, there is probably not any serious autocorrelation in the residuals.

In determining whether the model can be simplified, notice that the highest P-value on the independent variables is 0,0893, belonging to X_COAST. Since the P-value is less than 0.10, that term is statistically significant at the 90% confidence level. Depending on the confidence level at which you wish to work, you may or may not decide to remove X_COAST from the model.

Analysis report 15.

Multiple regression (MR2) analysis for Log(WIN)

Dependent variable: Log(WIN)

Parameter	Estimate	Standard Error	T Statistic	P-Value
CONSTANT	-701,169	104,929	-6,68231	0,0000
X	-0,000137059	0,0000711498	-1,92635	0,0581
Y	0,000348751	0,0000567451	6,14592	0,0000
S5000	0,0404467	0,0145846	2,77325	0,0071
D5000	-0,0266807	0,00623257	-4,28085	0,0001
X2	9,98583E-11	4,76547E-11	2,09545	0,0397
Y2	-4,03944E-11	6,54882E-12	-6,1682	0,0000
COAST_2	-3,55456E-10	9,99967E-11	-3,55468	0,0007
X_Z5000	2,06062E-8	8,19435E-9	2,51468	0,0142
Y_COAST	6,86822E-12	1,64556E-12	4,17378	0,0001
Y_Z5000	-4,25585E-9	1,40479E-9	-3,02952	0,0034
Y_D5000	6,17421E-9	1,45152E-9	4,25362	0,0001
COAST_Z500	6,34305E-8	1,54622E-8	4,10228	0,0001
COAST_S500	-0,00000142706	3,49911E-7	-4,07835	0,0001
Z5000_S500	0,000100578	0,000031121	3,23184	0,0019

Analysis of Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	8,99111	14	0,642222	19,72	0,0000
Residual	2,3126	71	0,0325718		
Total (Corr.)	11,3037	85			

R-squared = 79,5412 percent
R-squared (adjusted for d.f.) = 75,5071 percent
Standard Error of Est. = 0,180477
Mean absolute error = 0,129032
Durbin-Watson statistic = 2,17251

The StatAdvisor

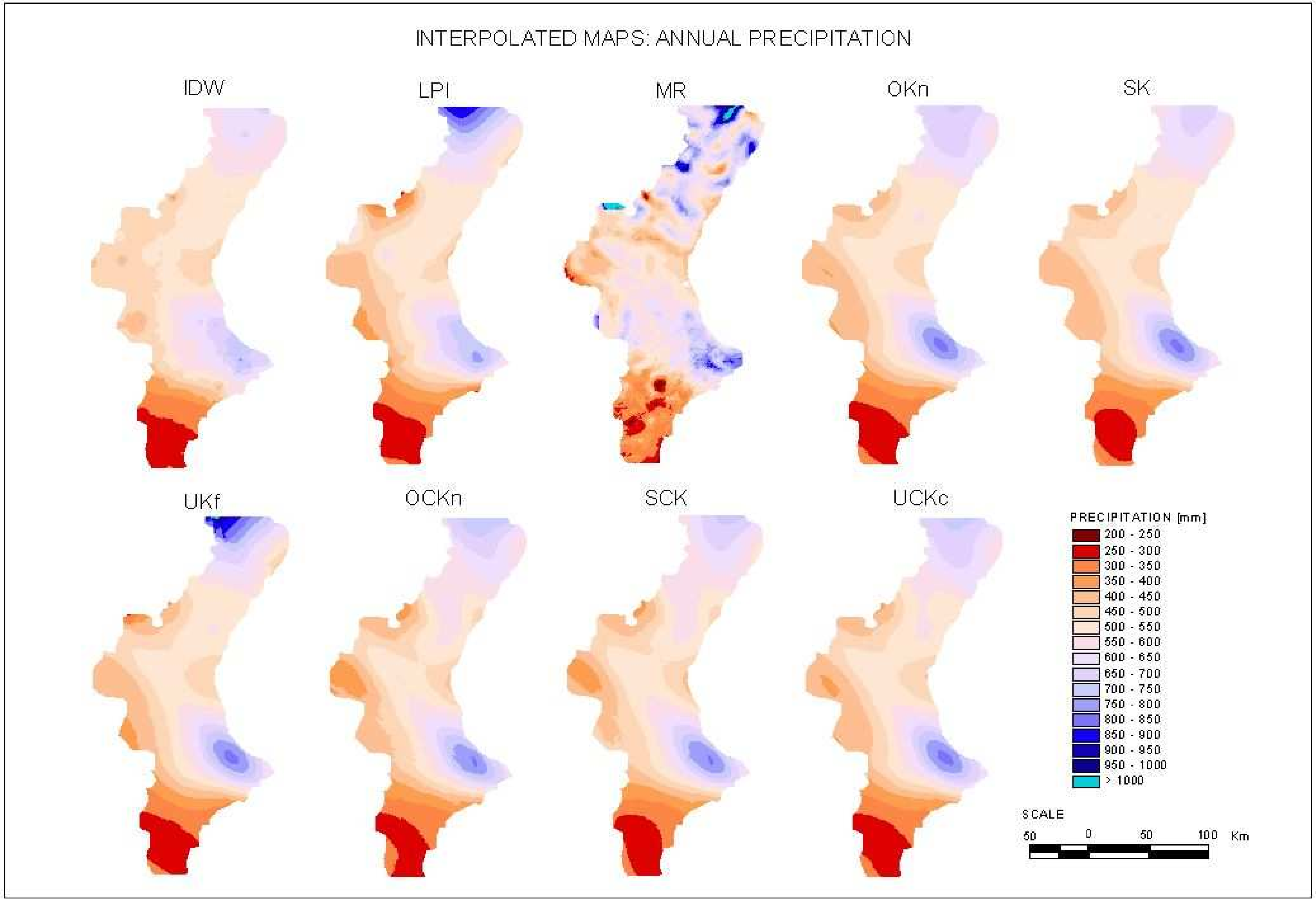
The output shows the results of fitting a multiple linear regression model to describe the relationship between Log(WIN) and 14 independent variables. The equation of the fitted model is

Log(WIN) = -701,169 - 0,000137059*X + 0,000348751*Y + 0,0404467*S5000 - 0,0266807*D5000 + 9,98583E-11*X2 - 4,03944E-11*Y2 - 3,55456E-10*COAST_2 + 2,06062E-8*X_Z5000 + 6,86822E-12*Y_COAST - 4,25585E-9*Y_Z5000 + 6,17421E-9*Y_D5000 + 6,34305E-8*COAST_Z500 - 0,00000142706*COAST_S500 + 0,000100578*Z5000_S500

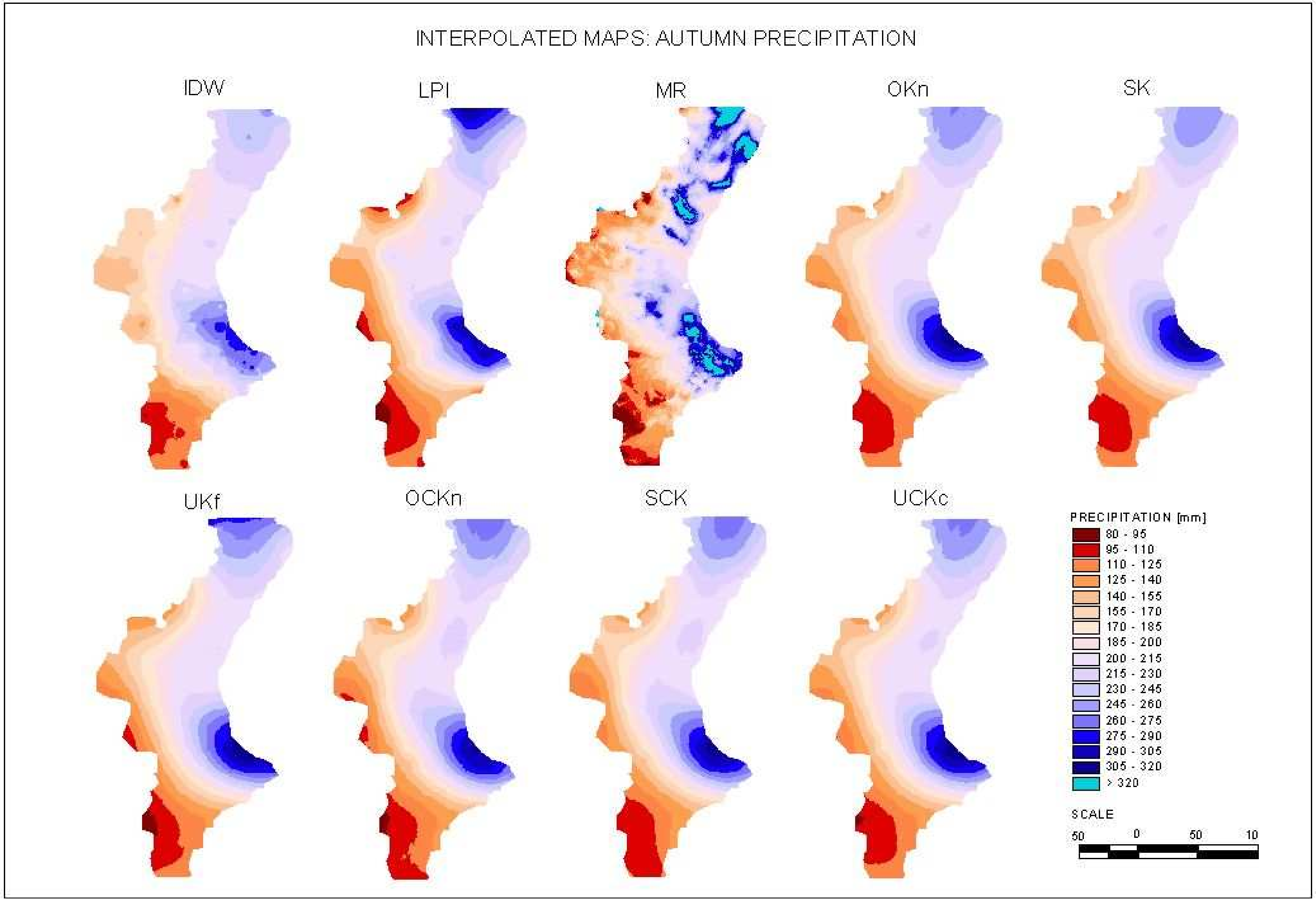
Since the P-value in the ANOVA table is less than 0.01, there is a statistically significant relationship between the variables at the 99% confidence level.

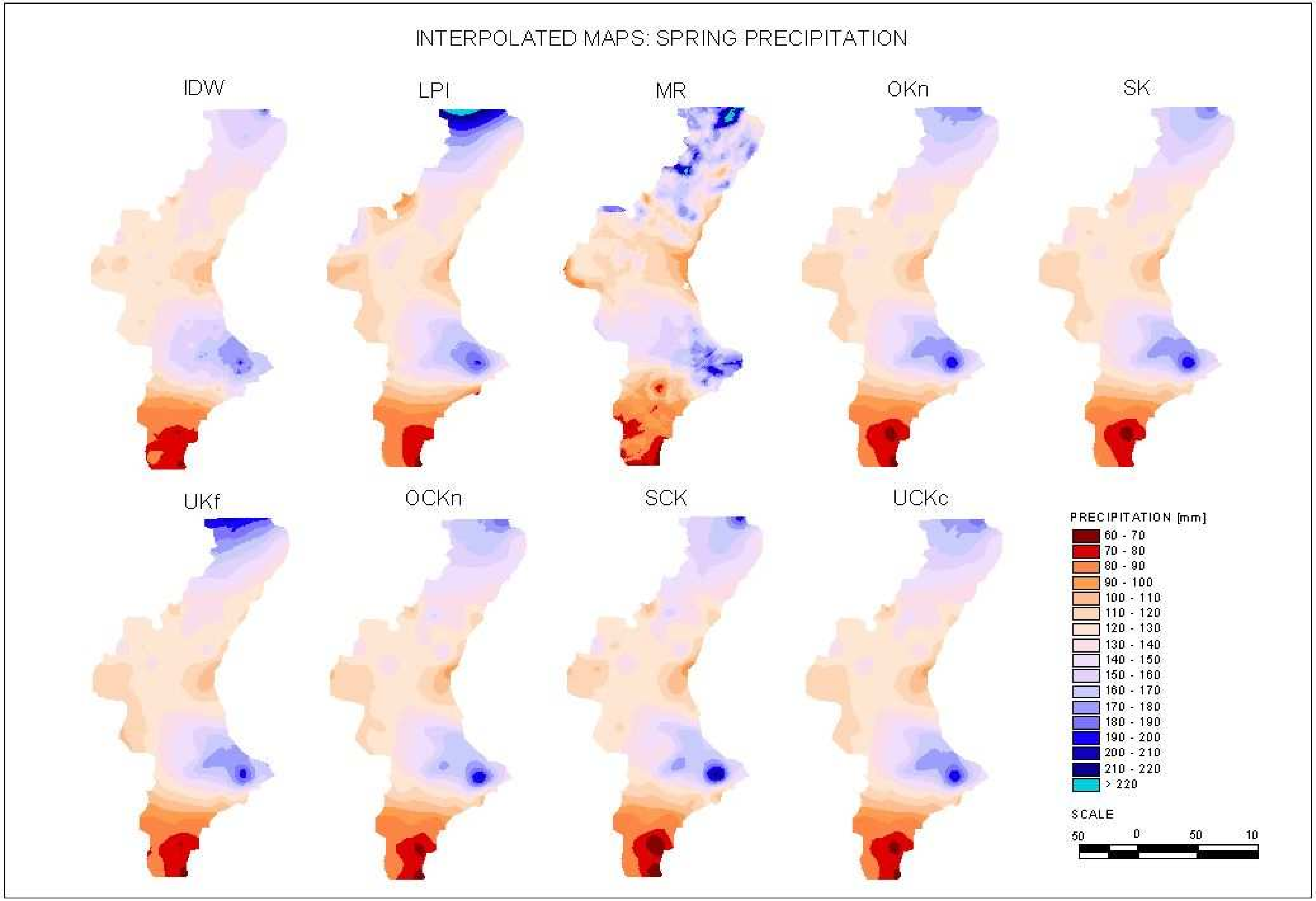
The R-Squared statistic indicates that the model as fitted explains 79,5412% of the variability in Log(WIN). The adjusted R-squared statistic, which is more suitable for comparing models with different numbers of independent variables, is 75,5071%. The standard error of the estimate shows the standard deviation of the residuals to be 0,180477. This value can be used to construct prediction limits for new observations by selecting the Reports option from the text menu. The mean absolute error (MAE) of 0,129032 is the average value of the residuals. The Durbin-Watson (DW) statistic tests the residuals to determine if there is any significant correlation based on the order in which they occur in your data file. Since the DW value is greater than 1.4, there is probably not any serious autocorrelation in the residuals.

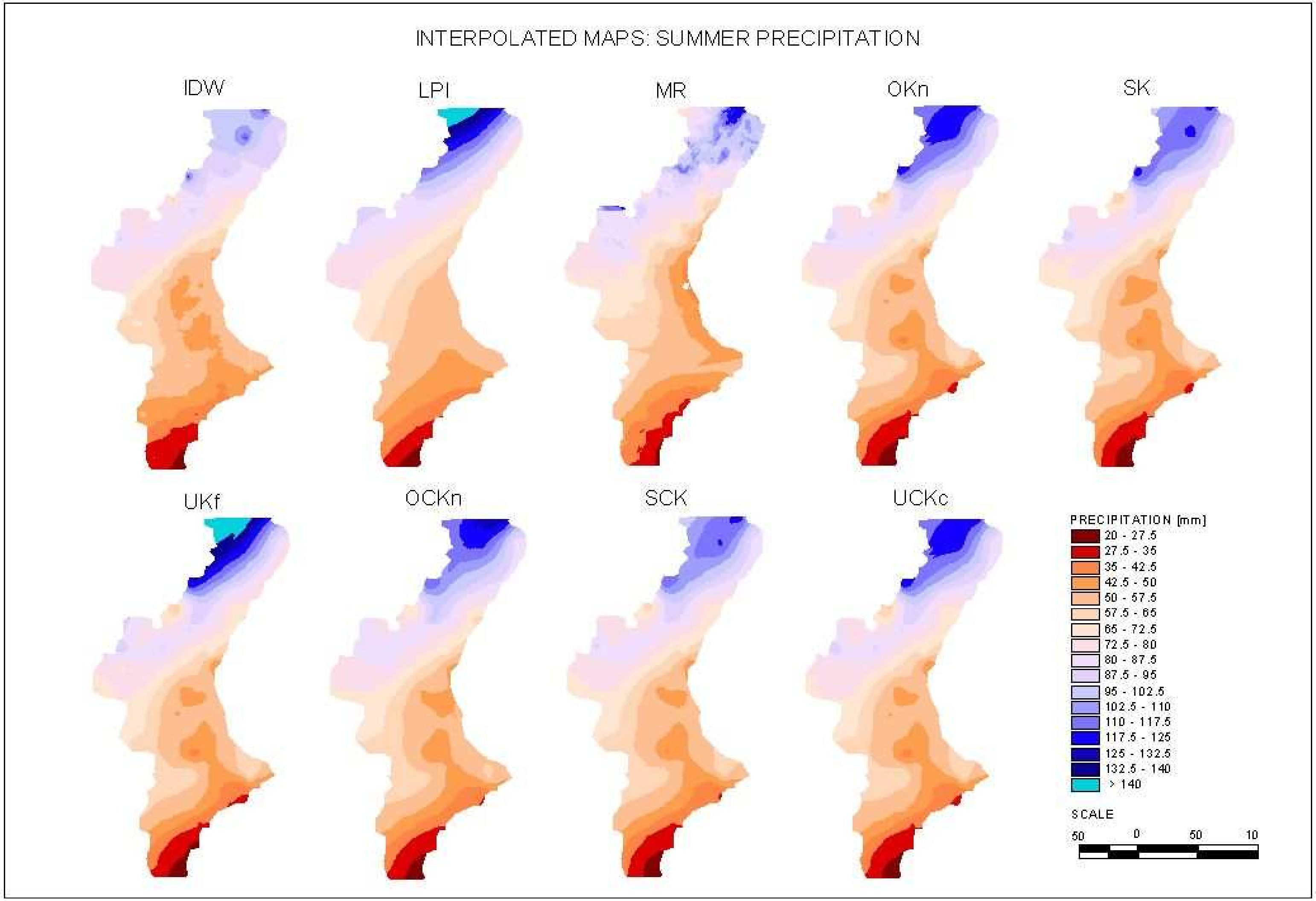
In determining whether the model can be simplified, notice that the highest P-value on the independent variables is 0,0581, belonging to X. Since the P-value is less than 0.10, that term is statistically significant at the 90% confidence level. Depending on the confidence level at which you wish to work, you may or may not decide to remove X from the model.

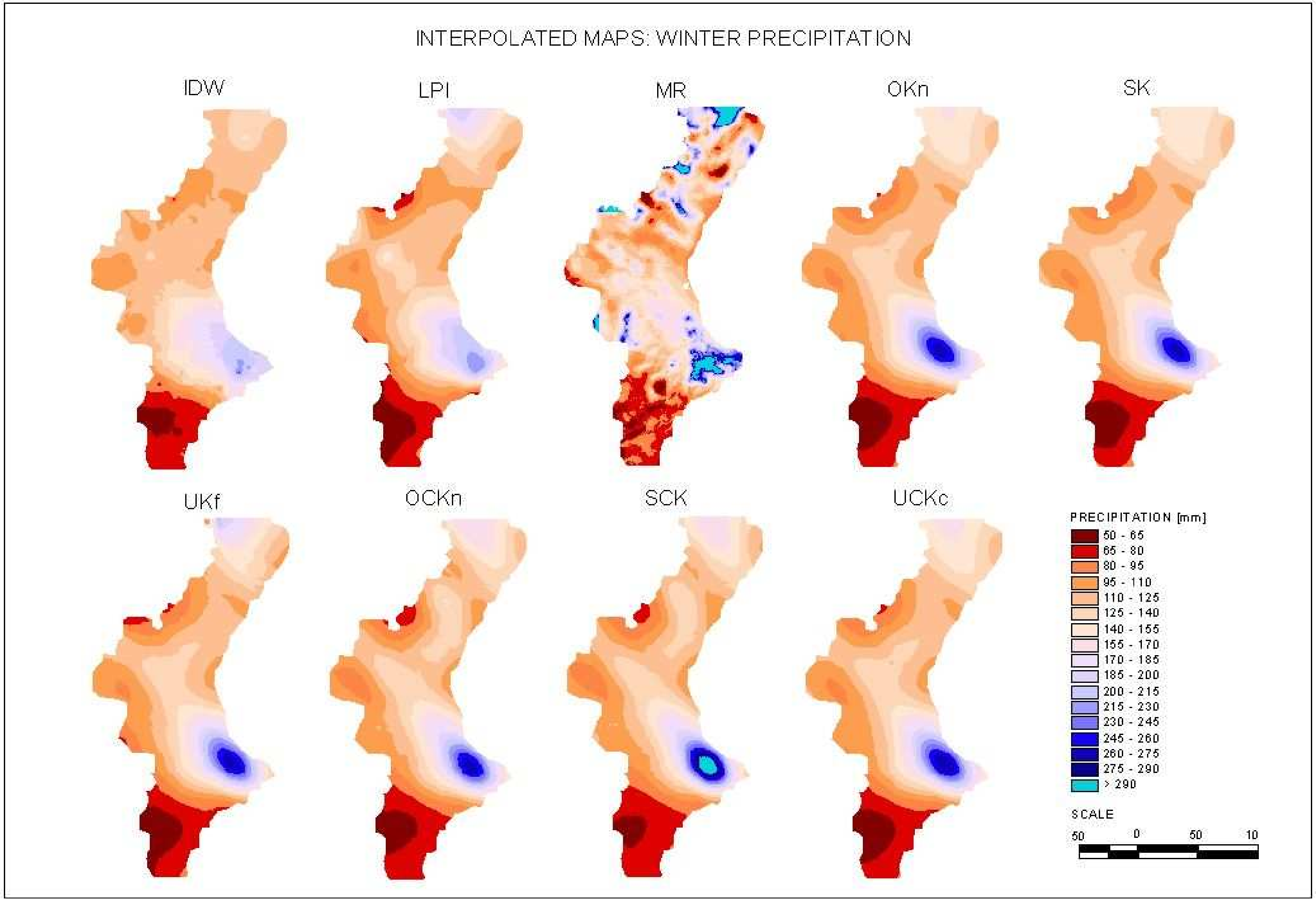


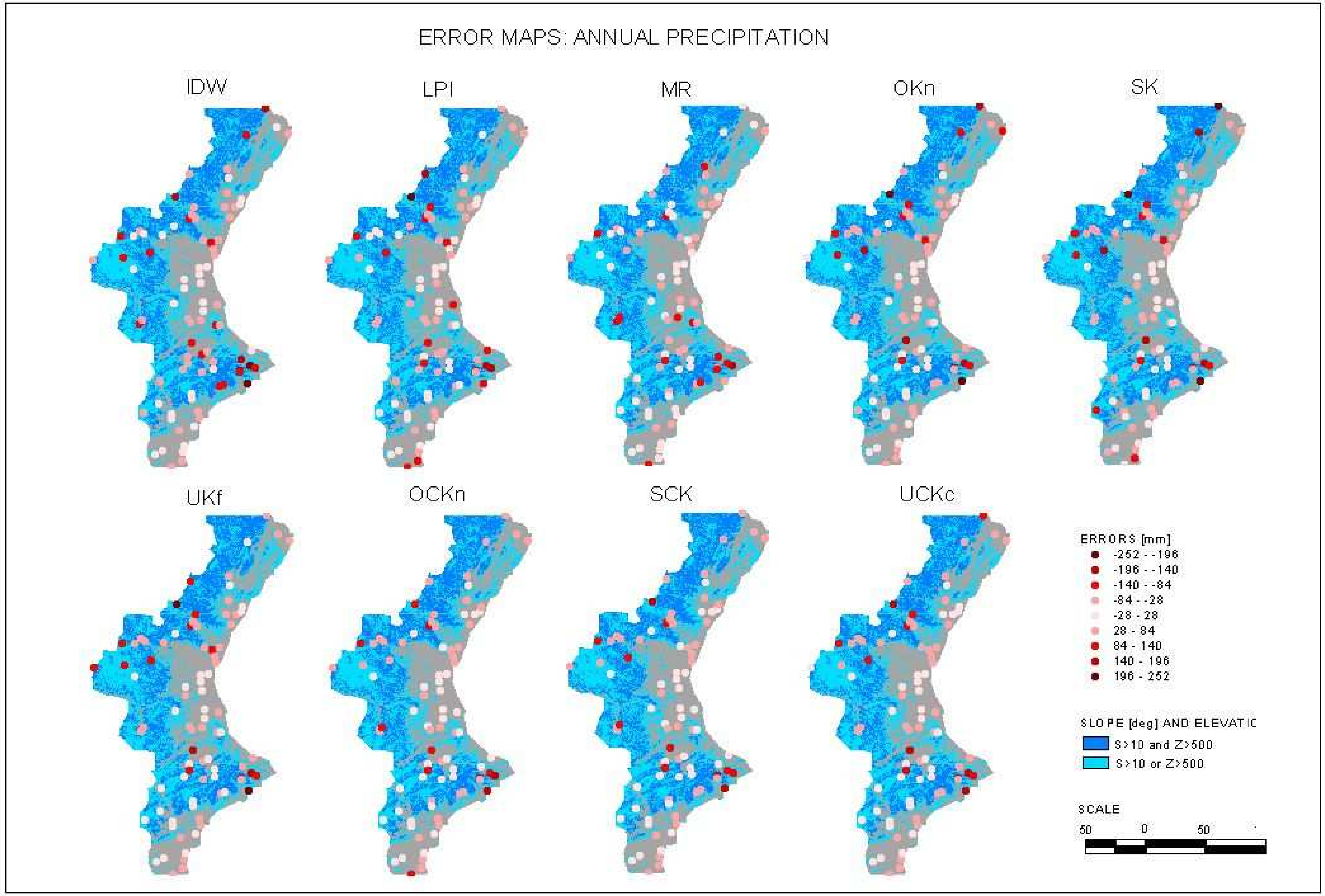
INTERPOLATED MAPS: ANNUAL PRECIPITATION
IDW
LPI
MR
OKn
SK
UKf
OCKn
SCK
UCKc
PRECIPITATION [mm]
200 - 250
250 - 300
300 - 350
350 - 400
400 - 450
450 - 500
500 - 550
550 - 600
600 - 650
650 - 700
700 - 750
750 - 800
800 - 850
850 - 900
900 - 950
950 - 1000
> 1000
SCALE
50
0
50
100
Km

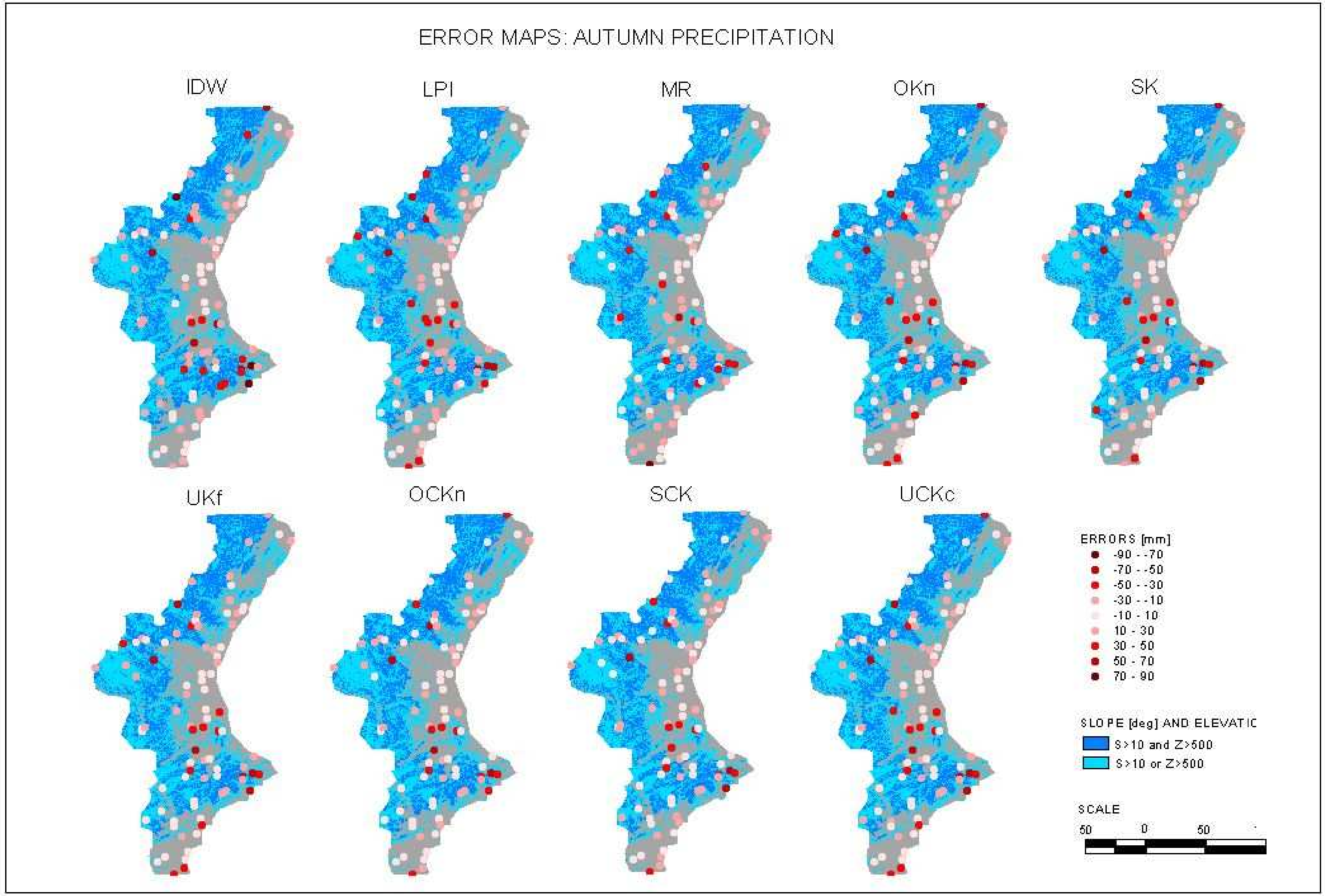

INTERPOLATED MAPS: AUTUMN PRECIPITATION
IDW
LPI
MR
OKn
SK
UKf
OCKn
SCK
UCKc
PRECIPITATION [mm]
80 - 95
95 - 110
110 - 125
125 - 140
140 - 155
155 - 170
170 - 185
185 - 200
200 - 215
215 - 230
230 - 245
245 - 260
260 - 275
275 - 290
290 - 305
305 - 320
> 320
SCALE
50
0
50
10

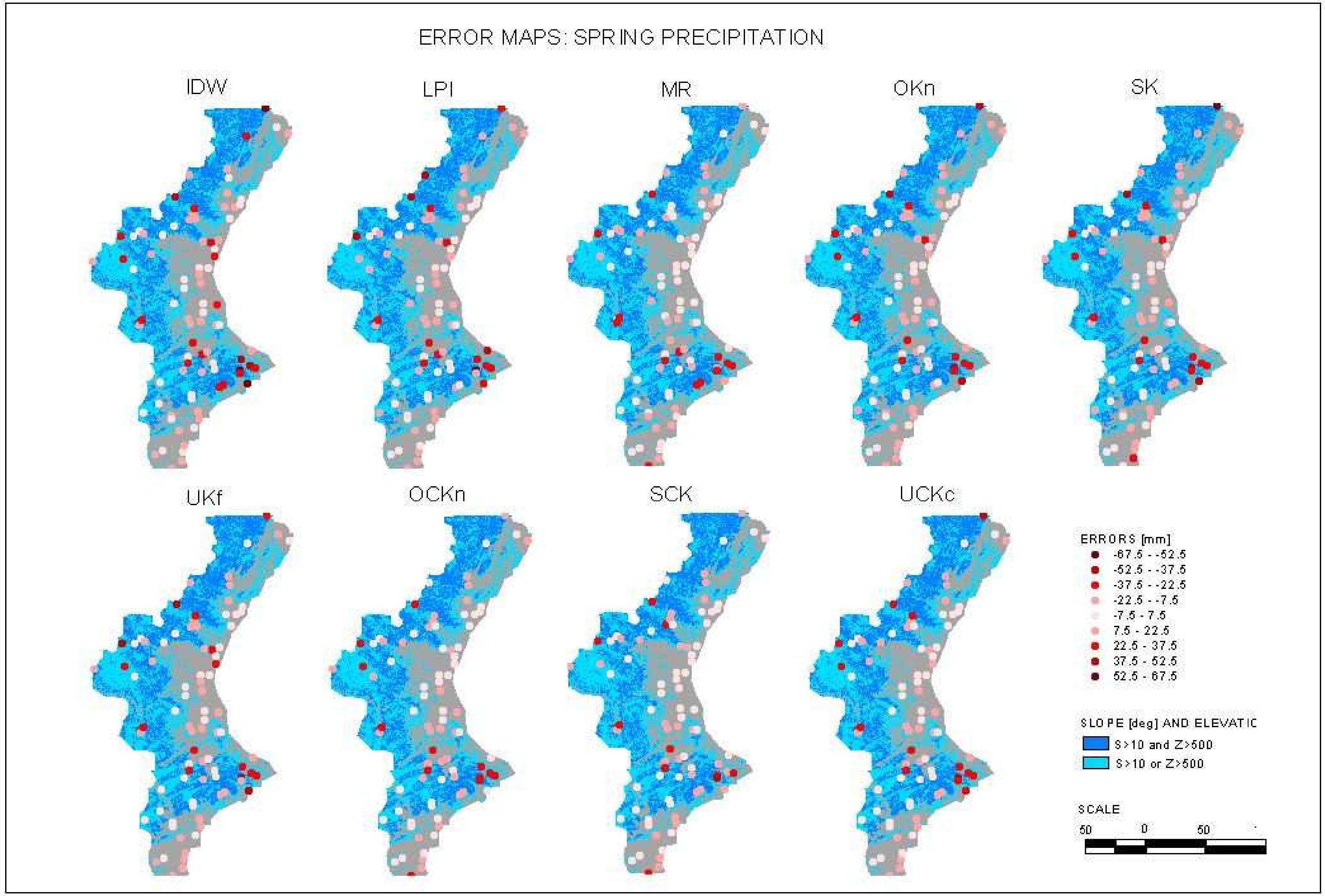

INTERPOLATED MAPS: SPRING PRECIPITATION
IDW
LPI
MR
OKn
SK
UKf
OCKn
SCK
UCKc
PRECIPITATION [mm]
60 - 70
70 - 80
80 - 90
90 - 100
100 - 110
110 - 120
120 - 130
130 - 140
140 - 150
150 - 160
160 - 170
170 - 180
180 - 190
190 - 200
200 - 210
210 - 220
> 220
SCALE
50
0
50
10

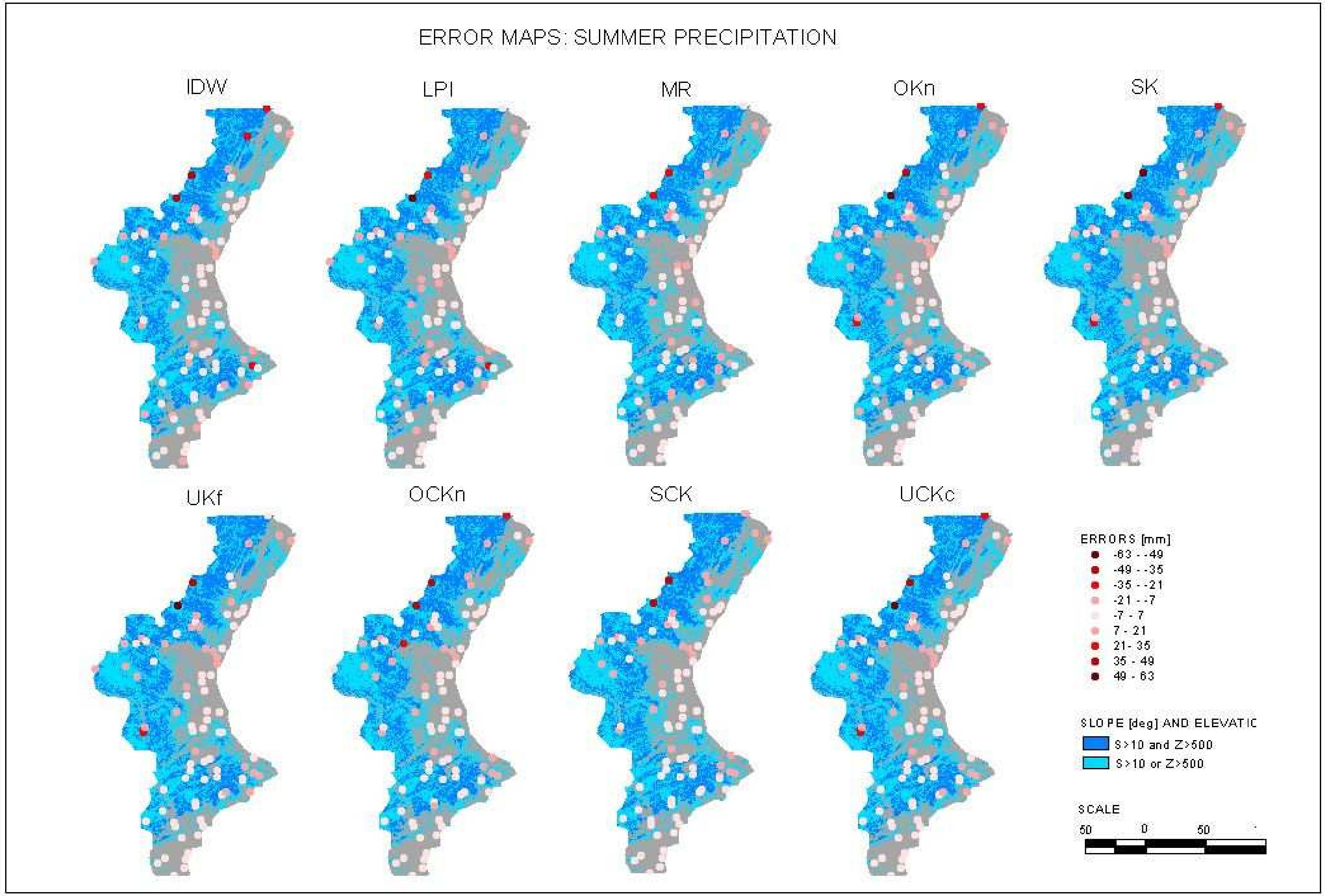

INTERPOLATED MAPS: SUMMER PRECIPITATION
IDW
LPI
MR
OKn
SK
UKf
OCKn
SCK
UCKc
PRECIPITATION [mm]
20 - 27.5
27.5 - 35
35 - 42.5
42.5 - 50
50 - 57.5
57.5 - 65
65 - 72.5
72.5 - 80
80 - 87.5
87.5 - 95
95 - 102.5
102.5 - 110
110 - 117.5
117.5 - 125
125 - 132.5
132.5 - 140
> 140
SCALE
50
0
50
10

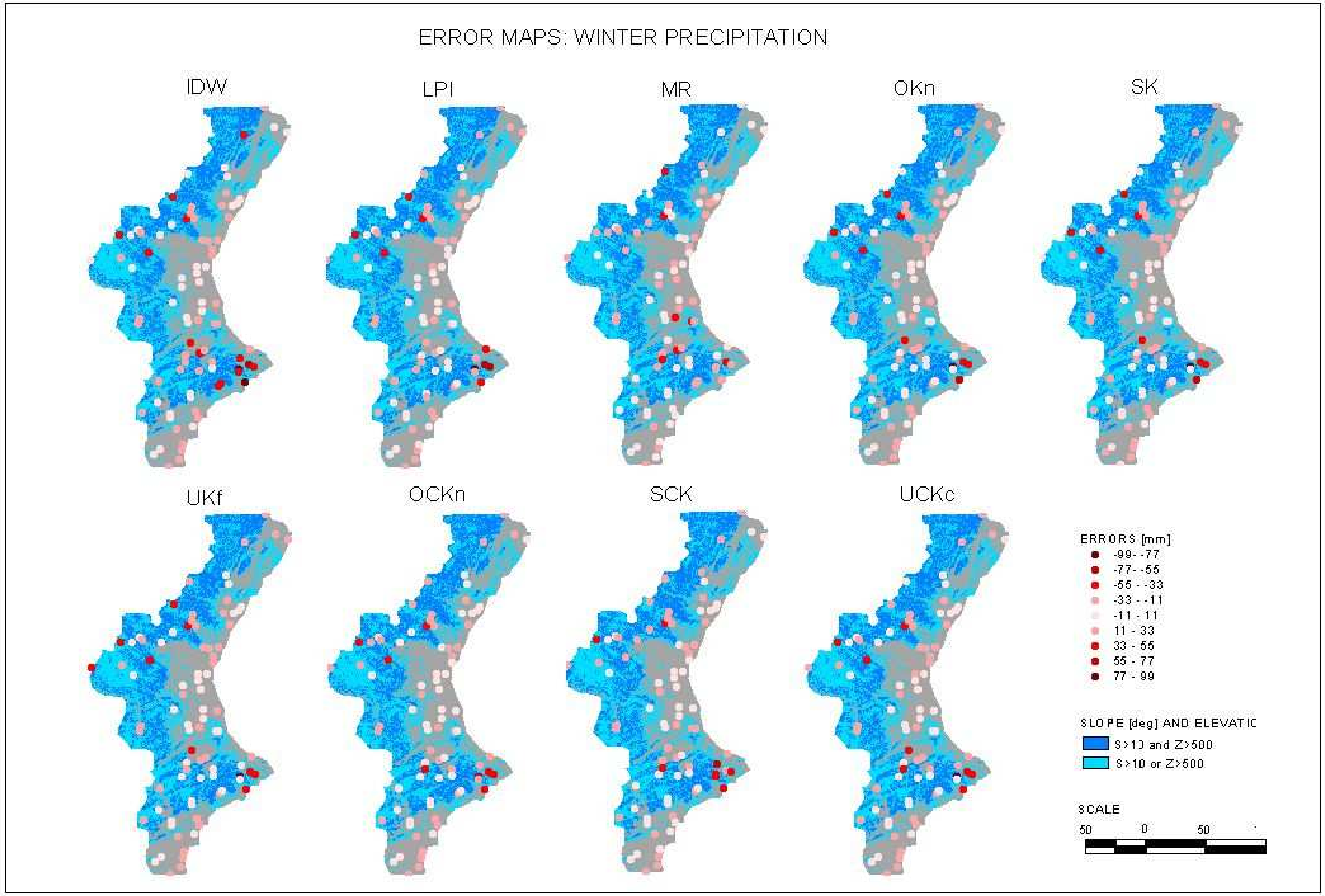

INTERPOLATED MAPS: WINTER PRECIPITATION
IDW
LPI
MR
OKn
SK
UKf
OCKn
SCK
UCKc
PRECIPITATION [mm]
50 - 65
65 - 80
80 - 95
95 - 110
110 - 125
125 - 140
140 - 155
155 - 170
170 - 185
185 - 200
200 - 215
215 - 230
230 - 245
245 - 260
260 - 275
275 - 290
> 290
SCALE
50
0
50
10


ERROR MAPS: ANNUAL PRECIPITATION
IDW
LPI
MR
OKn
SK
UKf
OCKn
SCK
UCKc
ERRORS [mm]
-252 - -196
-196 - -140
-140 - -84
-84 - -28
-28 - 28
28 - 84
84 - 140
140 - 196
196 - 252
SLOPE [deg] AND ELEVATIC
S>10 and Z>500
S>10 or Z>500
SCALE
50
0
50


ERROR MAPS: AUTUMN PRECIPITATION
IDW
LPI
MR
OKn
SK
UKf
OCKn
SCK
UCKc
ERRORS [mm]
-90 - -70
-70 - -50
-50 - -30
-30 - -10
-10 - 10
10 - 30
30 - 50
50 - 70
70 - 90
SLOPE [deg] AND ELEVATIC
S>10 and Z>500
S>10 or Z>500
SCALE
50
0
50


ERROR MAPS: SPRING PRECIPITATION
IDW
LPI
MR
OKn
SK
UKf
OCKn
SCK
UCKc
ERRORS [mm]
-67.5 - -52.5
-52.5 - -37.5
-37.5 - -22.5
-22.5 - -7.5
-7.5 - 7.5
7.5 - 22.5
22.5 - 37.5
37.5 - 52.5
52.5 - 67.5
SLOPE [deg] AND ELEVATIC
S>10 and Z>500
S>10 or Z>500
SCALE
50
0
50


ERROR MAPS: SUMMER PRECIPITATION
IDW
LPI
MR
OKn
SK
UKf
OCKn
SCK
UCKc
ERRORS [mm]
-63 - -49
-49 - -35
-35 - -21
-21 - -7
-7 - 7
7 - 21
21 - 35
35 - 49
49 - 63
SLOPE [deg] AND ELEVATIC
S>10 and Z>500
S>10 or Z>500
SCALE
50
0
50


ERROR MAPS: WINTER PRECIPITATION
IDW
LPI
MR
OKn
SK
UKf
OCKn
SCK
UCKc
ERRORS [mm]
-99 - -77
-77 - -55
-55 - -33
-33 - -11
-11 - 11
11 - 33
33 - 55
55 - 77
77 - 99
SLOPE [deg] AND ELEVATIC
S>10 and Z>500
S>10 or Z>500
SCALE
50
0
50