1. Introduction

Spatial representation of climatic data is a key point in many applications: agriculture management, environment research, transport, sport activities, etc. The impact of climatic data on society increased notably during the last years due to global climate changes. The estimation of the spatial distribution of ambient air temperature (AAT) is therefore necessary but difficult because AAT close to the ground is measured at meteorological stations (at the standard height of 2 m above ground level) which are often far apart (20–30 km). Physical causes for AAT spatial distribution, which usually form spatial gradients, can be statistically described with appropriate tools. Many studies have shown that relief and land cover attributes play an important role in the spatial distribution of AAT (Anquetin et al., 1999; Benichou and Le Breton, 1987; Bolstad et al., 1998; Bootsma, 1976; Joly et al., 2003; Geiger, 1965; Tveito and Førland, 1999). Therefore, the AAT variation as a function of elevation is a strong rule of spatial distribution. In the same way many other rules that explain AAT spatial distribution exist, thus they have to be identified and estimated.

A continuous AAT field can be estimated by one of the several spatial interpolation techniques (Anderson, 2002) applied to AAT measured at meteorological stations. Notably kriging (Courault and Monestiez, 1999; Matheron, 1970), a geostatistic interpolation (usually used in combination with one of the trend determination techniques, is often used for meteorological purposes. However, kriging, inverse distance, splines, etc., perform under expectations when interpolating a variable with low spatial autocorrelation, because all these techniques are based on the assumption that objects that are close to each other are more similar than objects far apart (Kanveski and Maignan, 2004).

The low autocorrelation issue requires other solutions – spatial distribution of AAT can also be explained with environmental parameters/attributes of the meteorological station. These attributes can be used to explain the AAT spatial distribution if a correlation between them and the AAT is statistically significant. This means that the attributes determine the behaviour of the AAT spatial distribution. Therefore, they are first used to generate a rule, which determines the AAT behaviour as regards the explanatory attributes. Later, this rule is applied to the regularly gridded data (which contains the spatial distribution of the attributes) in order to determine the AAT spatial distribution of the entire study area. This method, which is in statistics referred to as multiple regression, can be generally used to determine the relationship between any kind of variables. The rule used to explain the unknown variable with the explanatory variables, is mathematically defined as a polynomial function of explanatory attributes; the evaluated value is usually a linear combination of the explanatory attributes.

The presented study describes an interpolation experience based on the previously stated principles. Therefore, it uses a deductive approach in order to confirm our findings and estimate the general parameters that influence the AAT spatial distribution. The defined spatial climatology requires a database of appropriate data and a set of appropriate tools. The novelty of the presented study, which uses similar methods as for example Brossard et al. (2002) or Joly et al. (2003), is the introduction of remote sensing data as explanatory variables. In the end, the results can be mapped or even used as a model input in some other

1 ZRC SAZU, Slovenia, klemen.zaksek@zrc-sazu.si
2 ThéMA, France, daniel.joly@univ-fcomte.fr
applications. The method was applied to Slovenia using AAT measurements at 20 meteorological stations over the whole year 2005. The theoretical background is described in the second chapter and the case study is presented in third chapter. Finally, the results are discussed in the fourth chapter.

2. The method

The interpolation can begin once the GIS database containing the data able to explain the AAT spatial variation. The following procedure for the determination of AAT spatial distribution consists of six steps (figure 1):

1. computation of the AAT spatial autocorrelation (Moran’s index $I$; if the autocorrelation index is not close to 0, it is better to use a geostatistic or deterministic interpolation than the proposed procedure),
2. systematical computation of Pearson correlation index $r$ between AAT and all the explanatory variables stored in the GIS and then selection of the significantly correlated explanatory attributes,
3. multiple regression using the previous selected variables,
4. evaluation of the results by computing the residuals and cross validation,
5. interpolation of the residuals (by a geostatistic or deterministic interpolation if they are autocorre- lated),
6. mapping the results.

The first two steps are usually preformed smoothly. The problems might arise in step number three (multiple regression), if insufficient measurements are available. Multiple regression should be performed only on those explanatory attributes that are independent and statistically significant. Therefore, the correlations between AAT and explanatory attributes are systematically estimated by the correlation index $r$ (step 2). However, if the number of measurements is small, the computed correlation index might be a consequence of chance. Therefore, one needs to set a threshold that shows which correlation can be considered as a real correlation and not as a coincidence. Only variables that show a significant correlation index (not coincidental) are then kept for the second step, since statistically insignificant data often represents a source of errors. The significance of the correlation index was tested for all explanatory attributes using the equation following a Student $t$-distribution (Lowry, 2007):

$$t = \frac{r \cdot \sqrt{n - 2}}{\sqrt{1 - r^2}}$$

where $n$ is the number of measurements and $t$ is the $t$-distribution value. There are 18 degrees of freedom (to describe a linear relationship) when using 20 uncorrelated measurements, which means that if one is interested in a 95% (90%) confidence interval, the statistics $t$ should be larger than 2.10 (1.73). In this case all correlation indexes larger than 0.38 (0.30) are statistically significant (not coincidental) and the corresponding explanatory attributes can be used within the multiple regression. All the possible combinations of statistically significant explanatory attributes are then tested within a stepwise procedure (Efroymson, 1960). The final model corresponds to the equation that explains the most AAT variance.

No input data is usually left out for validation (step 4) of the deterministic or geostatistic interpolation method; the accuracy is then estimated by the k-fold cross or leave-one-out cross validation (Kanevski and Maignan, 2004). In the presented study the validation was especially problematic because data for only 20 meteorological stations was available (insufficient measurements to make a stable model, because if some measurements are left out of the model calibration in order to use them for the validation, the model can change significantly). Therefore, leave-one-out cross-validation was used; it involved using a single observation from the original sample as validation data, and the remaining observations as training data. This was repeated 20-times per situation so that each observation was used once as validation data – each
measurement is left out once, the model is calibrated on the other 19 measurements and the standard deviation $\sigma$ is computed from the 20 residuals $\text{res}_i$ in order to estimate the result accuracy.

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} \text{res}_i}{n-1}}$$

Furthermore, step number five is interesting because multiple regression can already be the final step in the procedure. However, the results might be enhanced with additional interpolation if the residuals are spatially correlated even if the variable to be interpolated was not autocorrelated at the beginning of the process. On the end, the results are finally mapped using parameters estimated in step 3 and 5.

3. Case study

The proposed procedure was tested in Slovenia which covers approximately 20 000 km$^2$. Its topography features a small coastal strip on the Adriatic Sea, the Alpine region in the north-west, hills in the central part and the Pannonia plain in the east. Due to its position on the crossing of various climatic regions, Slovenia has a number of different climate types. In order to describe this variety one needs various data (the datasets are introduced into the GIS as raster datasets at a 100 m resolution projected to UTM, zone 33N):

1. Environmental Agency of the Republic of Slovenia provided the AAT measurements (ARSO, 2007; measured at the height of 2 m measured at 7:00, 14:00 and 21:00 for the year 2005),
2. Surveying and Mapping Authority of the Republic of Slovenia provided the digital elevation produced from various sources (Podobnikar, 2005; slope, slope N–S direction, aspect, quasi-global solar radiation, topographic roughness and down-up position were derived from elevation data),
3. land cover data and its derivates (distance from the forest, distance from the sea) was prepared by Kokalj and Oštir (2006), and
4. MODIS satellite data averaged to the whole year 2005 (land surface temperature – LST and surface albedo in 1000 m resolution, normalized difference vegetation index – NDVI and enhanced vegetation index – EVI in 250 m resolution) was acquired from NASA (DAAC, 2006).

3.1 Windowing procedure

AAT spatial distribution is influenced by many parameters that might have a micro-local or a regional influence. Therefore, some explanatory variables (elevation, slope, roughness, aspect, solar radiation, down-up position and MODIS NDVI, but not other MODIS data, distance to the forest and to the sea and slope in the N–S direction) were estimated within spatial analyses windows of seven different sizes centred on each meteorological station (table 1). The various sizes of spatial analyses windows made it possible to consider the effect of the surrounding area on the measured AAT at different scale levels.

<table>
<thead>
<tr>
<th>Window number</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size [pixel]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Size [km]</td>
<td>0.3</td>
<td>1.1</td>
<td>2.9</td>
<td>5.1</td>
<td>10.1</td>
<td>15.1</td>
<td>20.1</td>
</tr>
</tbody>
</table>

Tab. 1: Size of spatial analyses windows used during data preparation.

The modelling was applied for each of the available temperature situations. The example for the first set of measurements in the AAT dataset (01. 01. 2005 at 7:00) was used to show how correlation indices be-
between AAT and explanatory variables vary with scale. Figure 1 shows the correlation index \( r \) values for different window sizes for some chosen variables. The elevation effect remains constant from the first to the seventh analysis window (the curve is almost flat), reflecting that the correlation between the elevation and AAT changes gradually in the case study area. This non-scalar pattern was already described by Joly and Brossard (2007). The correlation between AAT and elevation measures the intensity of the vertical thermal gradient. These calculations show that the elevation induces a uniform influence on this gradient (the higher the elevation is, the lower the AAT is).

On the other hand, the remaining explanatory attributes exhibit a strong scale effect. The scale of the explanatory variables indicates the hierarchy of variables that can explain the AAT distribution in a random situation. The optimal window size varies from one situation to another and might differ completely from those in the example presented for 01.01.2005. When the conditions are similar for the entire study area (for example on a clear night), large window sizes are more suitable for the analysis.

![Correlation Index](image.png)

Fig. 1: Correlation index \( r \) for six explanatory variables and the seven window sizes; situation on the 01.01.2005 at 7:00.

Among the available variables introduced in the systematic linear correlation, only four of them were characterised by a significant correlation index and then used within the multiple regression:

1. elevation (window 5; \( r = -0.59 \)),
2. LST (\( r = -0.58 \)),
3. aspect (window 1; \( r = 0.47 \)), and
4. roughness (window 4; \( r = -0.30 \)).

Some variables had also a significant correlation wit AAT but their use did not increase the final accuracy: slope (window 2; \( r = -0.42 \)), down-up position (window 5; \( r = -0.39 \)) and NDVI (window 4; \( r = -0.36 \)). Solar radiation (window 3; \( r = -0.27 \)), slope in the N–S direction (\( r = 0.11 \)), EVI (\( r = 0.11 \)), albedo (\( r = -0.08 \)) and distance to the sea (\( r = 0.00 \)) were not significantly correlated with AAT. The standard deviation of residuals for the 01. 01. 2005 at 7:00 situation equals 2.2 °C.

### 3.2 Standard deviation of residuals

All of the remaining 1095 situations for the year 2005 (365 days, three records per day) are systematically analyzed in the same way as the previous example. The standard deviation of the residuals for the entire
1095 situations equals 1.5 °C. The minimum and the maximum residuals are respectively -10.5 °C and 11.0 °C (perhaps a consequence of gross errors in the measurements). Over 50% of the residuals can be found between -1 °C and 1 °C (figure 2 left). The time in the day has only a small influence on the estimation quality. High residuals occur notably during winter time: from October to March (figure 2 right). The accuracy was for comparison estimated also without using the MODIS images which yielded to the standard deviation of 1.7 °C. The additional data has therefore improved the results.

Fig. 2: Statistical evaluation of the results – histogram of final residuals for 7:00, 14:00 and 21:00 on the left, accuracy according the time of the year and time of the day on the right (365 days; 20 stations).

### 3.3 Explanatory variables significance

Table 2 shows the frequencies of using explanatory variables introduced in the multiple regressions. It reveals that aspect, elevation, and distance from the sea, slope in the N–S direction and solar energy were often selected at 14:00 when AAT is the highest. In opposite, down-high position, roughness, slope, albedo2005, LST2005, NDVI2005 and EVI2005 were selected to explain the AAT distribution usually in morning and evening most likely because of the micro-local scale effects.

Aspect was the most frequently chosen as a significant explanatory variable AAT – more than 1000 situations (91.4%). The second most frequent variable used in the regressions is elevation (79.8%). EVI (MOD13Q1 product) and albedo (MOD43B3 product) with 3.3 and 2.8% frequency were recognized as the least statistically significant parameters, which is surprisingly at least for albedo that influences the amount of the solar radiation to be accepted by the surface.

The other two MODIS datasets played more significant role. The frequency of MODIS NDVI (MOD13Q1) is comparable to other explanatory variables (25.1%). One can also observe that the MODIS and artificial NDVI are well correlated in window 1 due to their similar spatial resolution, thus this combination of data was not used in the study. Furthermore, MODIS LST (MOD11A2 product) is a significant attribute for AAT distribution in 47.9% of all situations. Because of its low spatial resolution (1000 m) it has only a regional influence but it is still the third most frequent explanatory variable.

### 3.4 Interpolated results

Figures 3 and 4 show an example of the interpolated results for January 4 at 7:00 and July 7 at 14:00. The areas in white are nodata areas that occur where the selected explanatory variable is out of range given by
the attribute values on the meteorological stations (in order to prevent extrapolation that can lead to gross errors).

<table>
<thead>
<tr>
<th>Variable</th>
<th>7:00</th>
<th>14:00</th>
<th>21:00</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aspect</td>
<td>91.3</td>
<td>95.6</td>
<td>87.4</td>
<td>91.4</td>
</tr>
<tr>
<td>Elevation</td>
<td>74.1</td>
<td>87.6</td>
<td>77.8</td>
<td>79.8</td>
</tr>
<tr>
<td>Distance from the sea</td>
<td>18.4</td>
<td>43.4</td>
<td>20.3</td>
<td>27.4</td>
</tr>
<tr>
<td>Slope in N–S direction</td>
<td>22.8</td>
<td>33.2</td>
<td>24.1</td>
<td>26.7</td>
</tr>
<tr>
<td>Solar radiation</td>
<td>12.6</td>
<td>22.2</td>
<td>12.1</td>
<td>15.6</td>
</tr>
<tr>
<td>LST2005</td>
<td>61.9</td>
<td>37.8</td>
<td>44.1</td>
<td>47.9</td>
</tr>
<tr>
<td>Down-up position</td>
<td>34.0</td>
<td>13.5</td>
<td>42.5</td>
<td>30.0</td>
</tr>
<tr>
<td>NDVI2005</td>
<td>26.8</td>
<td>15.4</td>
<td>33.2</td>
<td>25.1</td>
</tr>
<tr>
<td>Roughness</td>
<td>21.7</td>
<td>19.4</td>
<td>24.4</td>
<td>21.8</td>
</tr>
<tr>
<td>Slope</td>
<td>14.6</td>
<td>10.7</td>
<td>27.2</td>
<td>17.5</td>
</tr>
<tr>
<td>Albedo2005</td>
<td>2.7</td>
<td>5.0</td>
<td>2.2</td>
<td>3.3</td>
</tr>
<tr>
<td>EVI2005</td>
<td>3.0</td>
<td>1.9</td>
<td>3.3</td>
<td>2.8</td>
</tr>
</tbody>
</table>

Tab. 2: Frequency [%] of explanatory variables introduced in the multiple regression.

4. Conclusion

The results of the presented study, which were not validated with independent data because of the low number of measurements (cross validation was therefore used), are comparable with previous studies. Anderson (2002) tested various interpolation methods in the Phoenix metropolitan area, where the density of measurements was approximately ten-times larger than in the presented study; the best results were obtained by kriging interpolation ($\text{RMSE} = 1.62 \, ^\circ\text{C}$). The presented results are also better than the 24 hour weather forecast (ARSO, 2007), but they cannot be compared to the results of the AAT high resolution interpolation on Svalbard (Joly et al., 2003) where the AAT was determined with the accuracy of $0.3 \, ^\circ\text{C}$ based on 50 measurement stations covering the area of merely 8 km$^2$. In addition to the low density of the meteorological stations, the origin of the rather poor results quality in the case study is mountainous study area which can be characterized also by climatically quasi independent climate valleys (Sevruk and Mieglitz, 2002).

Aspect, elevation and LST are often among the significant attributes that can explain the AAT spatial distribution. Solar radiation, distance from the sea or slope in the N–S direction are also usually significant during daytime in clear-sky conditions. NDVI, EVI, LST, albedo, down-high position and slope are significant in night and by cloudy conditions. Therefore, many attributes are correlated to AAT but the connection between these attributes and the measured AAT is different for different weather situations. Furthermore, the scale has an effect on the results, which makes AAT interpolation a difficult task.

The presented method can be successfully applied for any case study area once GIS database is prepared. It was shown that data from a greater number of meteorological stations is necessary in order to obtain a high accuracy AAT field. A novelty of this study was the inclusion of MODIS data – EVI and surface albedo are not correlated with AAT but LST and NDVI have proved to have a significant influence on AAT and therefore their use improves the overall accuracy for $0.2 \, ^\circ\text{C}$.
Fig. 3: Interpolated AAT field in 100 m spatial resolution; example for 04. 01. 2005 at 7:00). The relief roughness has a large influence on the AAT in the morning as it can explain the thermal inversion.

Fig. 4: Interpolated AAT field in 100 m spatial resolution; example for a cloudy day – 07. 07. 2005 at 14:00). It can be seen that the elevation has a great influence in the case of cloudy weather.
Bibliography


