DEVELOPMENT AND COMPARISON OF UNCERTAINTY MEASURES IN THE FRAMEWORK OF A DATA CLASSIFICATION

J. Schiewe [1]

[1] Lab for Geoinformatics and Geovisualization (g2lab), HafenCity University Hamburg, Germany
– jochen.schiewe@hcu-hamburg.de

Commission IV, WG IV/3

KEY WORDS: Uncertainty, Data Classification, Uncertainty Visualization

ABSTRACT:

In the analysis and visualization of spatial information, quite often a data classification is applied. The choice of different methods, together with the choice of a different number of classes, the consideration of open classes and the treatment of outliers, can produce very different results. Hence, it is desirable to quantify the uncertainties that inevitably arise in this process. So far, almost only non-spatial properties have been considered. In addition to an extension of this set of statistical measures, this article also aims to define those which are concerned with the preservation of spatial patterns (e.g., local extreme values) as well as with visual perception. An empirical study will investigate the behavior of all these measures, for example depending on the classification method used or the number of classes. Also, correlations between the uncertainty measures and between the measures and statistical properties of the input data are examined. Finally, is will be shown that the uncertainty measures can not only be used individually or combined for pure evaluation purposes, but also for a-posteriori improvement of classification methods.

1. MOTIVATION

In the analysis and visualization of spatial phenomena, data classification can be helpful in many regards, for example, to reduce the volume of data, to emphasize spatial differences or to facilitate the reading of values. A typical example of data classification arises in the course of the production of choropleth maps for the graphical representation of areal quantities. Various classification methods are available that group the original values according to equidistance, quantiles, natural breaks or other criteria. It is well known that the choice of different methods, together with the choice of a different number of classes, the consideration of open classes and the treatment of outliers, can produce very different results. This leads in the case of choropleth maps to very different visualizations and thus to different cognitive impressions and decisions.

The overall aim of this paper is to quantify the uncertainties that inevitably arise in the generalization process of spatial data classification.

Based on a brief literature review (chapter 2) existing but also supplementary measures are presented that cover the various tasks occurring in the analysis and visualization workflows (chapter 3). Firstly, an extended set of non-spatial measures will be presented. However, these statistical parameters do not satisfy all requirements. Therefore, additional measures are presented that assess the preservation of certain spatial characteristics (e.g., preservation of local extreme values). Furthermore, key figures are presented that describe the visual perception uncertainty. Most of these measures give not only a global rating for a specific classification (i.e., with a predefined method and number of classes), but also specific measures for each individual class.

In chapter 4 an empirical investigation is performed in order to describe the behavior of the uncertainty measures depending on the number classes and used classification methods, the correlations between uncertainty measures as well as differences between uncertainty measures based on various input data characteristics. Chapter 5 demonstrates possible further usages of these measures, either in the context of a multi-criteria analysis, or as starting point for class-specific a-posteriori optimization of a given classification result.

2. PREVIOUS WORK

In the literature, the topic of data classification for cartographic purposes is comprehensively treated - here, reference is made to the textbook by Slocum et al. (2009) as well as the reviews by Cromley & Cromley (1996) or Coulsen (1987).

A number of empirical studies are also concerned with the comparison of methods for addressing typical map-use tasks (e.g., Goldsberry & Battersby 2009, Brewer & Pickle 2002, Mersey, 1990). In addition, interactive tools have been developed to find the "optimal" configuration for a given application; for example, the use of linked views between the data histogram and the choropleth map (Andrienko et al., 2001). While the methods implemented in common software work on a data-driven basis, relatively little was published on the neglect of the spatial context or the use of a task-oriented approach; Armstrong et al. (2003) give an overview.

The description of the uncertainty of a data classification refers almost exclusively to non-spatial properties (e.g., Jenks & Caspall, 1971; Armstrong et al., 2003; Andrienko et al., 2001). Examples of this (e.g., *Tabular Accuracy Index* or *Goodness of Variance Fit*) are picked up in section 3.1. Bregt & Wopereis (1990) investigate the visual assessment in comparison to various complexity measures.

3. UNCERTAINTY MEASURES

There are a number of proposals for statistical measures related to data classification in the literature. In most cases they describe global statistical behavior neglecting uncertainties related to spatial patterns or visual perception. In the following an extended set of measures will be presented that will consist of within-class as well as global descriptions of a classification.

3.1 Non-spatial measures

First, the **within-class homogeneity** can be described, i.e. the variation of values belonging to one specific class. The underlying idea is to keep the variance as low as possible so that all values with one class value are really as similar or identical as the association to one class value implicates. Examples for homogeneity measures are the

- *Tabular Accuracy Index (TAI),* taking the absolute deviations from the class mean value into account (Jenks & Caspall, 1971);
- *Goodness of Variance Fit (GVF),* considering the squared deviations from the class mean (Dent, 1999);
- Entropy approach, applying the logarithmic function to the absolute deviation from the mean (Andrienko et al., 2001).

For all measures the deviation from the class mean values can also be replaced with the median in order to get less influence of outliers. Our empirical testing (see chapter 4) showed significant correlations between these measures so that a reduction to one indicator appears reasonable. For the remainder of this paper we use the *Goodness of Variance Fit* according to Dent (1999) as class homogeneity measure. The GVF_c for a specific class c and GVF_g for the total data set are calculated as follows:

$$GVF_c = 1 - \frac{\sqrt{\frac{\sum_{i=1}^{n_c}(x_i - \bar{x}_c)^2}{n_c - 1}}}{s_g} \tag{1a}$$

$$GVF_g = 1 - \frac{\sqrt{\frac{\sum_{c=1}^{m}\sum_{i=1}^{n_c}(x_i - \bar{x}_c)^2}{n_g - 1}}}{s_g} \tag{1b}$$

where:
x_i : i-th data value
$\bar{x}_c$: mean value within class c
m: number of classes
n_c: number of values in class c
n_g: total number of values in total data set
s_g: standard deviation of values of total data set

An increasing GVF value (with a maximum of +1) corresponds to increasing within-class homogeneity.

Instead of considering the within-class homogeneity it is also possible to evaluate the **between-class heterogeneity**, i.e. the goal is to create well distinguishable classes. Sun et al. (2013) propose a probability-based separability measure for units (e.g., polygons) that are associated to different classes. In the following, a simpler approach will be pursued: It is based on the demand that each value should have a smaller difference to its current class (c) value compared to the difference to the two adjacent classes (c+1; c-1). The respective differences of a value x amount to

$$d_{i.c} = |x_i - \hat{x}_c|$$

$$d_{i.c+1} = |x_i - \hat{x}_{c+1}|$$

$$d_{i.c-1} = |x_i - \hat{x}_{c-1}|$$

$$d_i = MIN(d_{c+1}; d_{c-1}) - d_c$$

where:
$\hat{x}_c$: class value of current classification (average of upper and lower class limit value)
$\hat{x}_{+1}$: class value of subsequent class
$\hat{x}_{-1}$: class value of preceding class

From this a comparison is made between the allocations to the current class and the closer adjacent class. If d is larger than zero the value is placed best in the current class. Finally, the ratio of values that are placed best in the current class to the total number of values can be computed for each class (NN_c) and the entire data set (NN_g) as follows:

$$ct_c = 0$$

$$for\ i = 1..n_c: ct_c = ct_c + 1 \quad if\ d_i \geq 0$$

$$NN_c = \tag{2a}$$

$$ct_g = 0$$

$$for\ i = 1..n_g: ct_g = ct_g + 1 \quad if\ d_i \geq 0$$

$$NN_{gc} = \frac{ct_g}{n_g} \tag{2b}$$

A large value of NN (with a maximum of +1) points to a desired large number of values that are categorized to the current class, which corresponds to a good between-class heterogeneity. By definition, the equidistance method always shows NN values of +1.

In literature, less attention is paid to the deviation of the original values from the corresponding class value (i.e., the average of upper and lower class limits). Even if the within-class homogeneity is very low it can happen that the values are very close to the one or other limit value so that the representation is rather weak. The **within-class matching** values MAT_c for a specific class c and MAT_g for the total data are introduced as follows:

$$MAT_c = 1 - \frac{\sqrt{\frac{\sum_{i=1}^{n_c}(x_i - \hat{x}_c)^2}{n_c - 1}}}{s_g} \tag{3a}$$

$$MAT_g = 1 - \frac{\sqrt{\frac{\sum_{c=1}^{m}\sum_{i=1}^{n_c}(x_i - \hat{x}_c)^2}{n_g - 1}}}{s_g} \tag{3b}$$

where:
$\hat{x}_c$: class value (average of upper and lower class limit value)

Again, a MAT value towards +1 corresponds to a good matching.

In many applications, special attention is paid to the design and expressiveness of border classes. One goal can be the **isolation of global extreme values** in order to make these uniquely visible in choropleth maps. Hence, the *Global Extreme Value index (GEX)*, which is a global measure, looks for the number of elements in the first and last classes as follows:

$$g_{MIN} = \begin{cases} 1 & if \ n_1 = 1 \\ 1 - \dfrac{n_1}{n_g/m} & otherwise \end{cases}$$

$$g_{MAX} = \begin{cases} 1 & if \ n_m = 1 \\ 1 - \dfrac{n_m}{n_g/m} & otherwise \end{cases}$$

$$GEX = {}^1\!/_2 \, (g_{MIN} + g_{MAX}) \tag{4}$$

where:
n_1: number of data values in lowest class
n_m: number of data values in highest class

A large GEX value corresponds to a small number of values in the border classes, with the maximum value of +1 that points to the exclusive appearance of one and only one value in both border classes. Vice versa, a quantile classification produces a (near) equal distribution of values among classes so that GEX will be (close to) zero.

3.2 Spatial measures

The above mentioned classification methods are data-driven, i.e., the intervals are determined solely on the basis of the present frequency distribution of the original values. The spatial context of the underlying data, which is relevant for many applications, is completely neglected when using such divisions. Accordingly, so far there are hardly any measures that describe the uncertainty of a data classification looking for the preservation of spatial properties. One exception is the *Boundary Accuracy Index (BAI)* that evaluates values that are separated by common class boundaries (Armstrong et al., 2003). Hence, further spatial patterns of interest should be considered during the data classification stage. Within the project *aChor* (Schiewe & Chang, 2018) new classification methods are developed that also use measures describing the success rate for preserving specific spatial patterns:

- **Local extreme values** are defined as polygons that show a larger (smaller) value compared to all neighboring polygons. Consequently, the data classification should not aggregate neighboring polygons into the same class. The *Local extreme value preservation rate (LEX)* calculates the ratio of preserved local extreme values after classification compared to the total number of local extreme values (Schiewe, 2017).
- **Hot or Cold spots** are defined as polygons with high (low) values that are surrounded in a certain neighborhood with polygons also showing high (low) values. The hot / cold spot polygons are determined by standard methods such as the Getis Ord index (and derived z-values in conjunction with the application of a corresponding threshold value; Getis & Ord, 1992). From this a binary classification (and labeling) in hot / cold spots as well as other polygons is possible. Similar to the LEX measure, the *Hot/Cold Spot preservation rate (HCS)* is defined by the ratio of preserved Hot/Cold spots in comparison to all detected spots.
- **Edges** are defined as polygon boundaries that show a significant value difference. Because the aforementioned BAI measure shows some disadvantages, the alternative *Edge Preserving Index* (EPI) has been developed (Schiewe, 2016). It compares the preservation (or even, enhancement) of neighboring values compared to resulting class values.

3.3 Visual perception measures

When it comes to visualization of classified data, visual perception is strongly influenced by dominant colors. This can be desired in the case of an actual imbalance of class occupation, or undesired if large regions are perceived more prominent than small regions and the task requires that regions should be perceived equally. Hence, the **global visual balance** over all classes should be quantified. Armstrong et al. (2003) propose a Gini coefficient to express this. Alternatively, for each class the area fraction is calculated (i.e., the total area of all class members divided by the total area of the data set) and compared to an equal area fraction (i.e., the total area divided by the number m of classes). For normalization purposes the sum of all class related differences is related to the worst case, i.e. the situation that only one class covers the entire area (i.e. the absolute difference amounts to 1-1/m) and all remaining classes are empty (i.e., (m-1) · 1/m) – leading to the factor (1/(2-2/m)). With that, the *Global Visual Balance (GVB)* is calculated as follows:

$$GVB = 1 - \frac{1}{2 - \frac{2}{m}} \sum_{c=1}^{m} \left| \frac{\sum_{i=1}^{n_c} A_i}{A} - \frac{1}{m} \right| \tag{5}$$

where:
A: total area of data set
A_i: total area of class

GVB values towards +1 correspond to an equal area distribution within the visualized classification result.

A typical problem with choropleth maps is the **within-class visual imbalance** that is caused by huge area differences within one class ("Russia vs Andorra effect"). For this purpose, only the largest and smallest areas within a class have to be considered. Since this effect is evident especially in border classes (represented by the most intense hues), a weighting towards these classes is performed in the course of determining a global measure. Here, a distinction can still be made between the use of a sequential color scheme (increasing the weight from lowest to highest class) and a bi-polar color scheme (increasing the weight from middle to border classes). In the following formulas, a sequential color scheme as well as an exponential increase is modeled for the *Class Visual Balance* measure for each single class (CVB_c) and the overall data set (CVB_g):

$$CVB_c = 1 - \frac{x_{c.MAX} - x_{c.MIN}}{x_{MAX} - x_{MIN}} \tag{6a}$$

$$CVB_g = \sum_{c=1}^{m} \frac{2^j}{\sum_{k=1}^{m} 2^k} CVB_c \tag{6b}$$

where:
$x_{c.MAX}, x_{c.MIN}$: maximum and minimum value within class c

x_{MAX}, x_{MIN} : maximum and minimum value within entire data set

CVB values towards +1 correspond to low area differences within classes.

4. EMPIRICAL INVESTIGATION OF MEASURES

4.1 Methodology and Data

Based on the aforementioned compilation of the different uncertainty measures an empirical investigation is performed in order to describe the behavior of measures depending on the
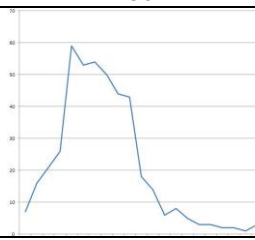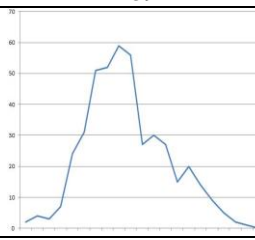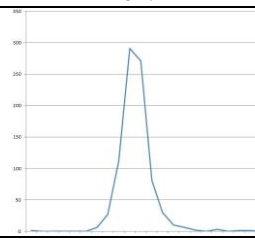
Data set		Rainfall	Average Age	Social Index
Number of polygons (n)		438	439	847
Statistical measures for selected attribute	**Histogram** right axis: attribute value upper axis: frequency			
	RMSE	34,9	1,7	2,9
	MAX-MIN	204,9	9,5	40,7
	RMSE / (MAX-MIN)	0,17	0,18	0,07
	IQR / (MAX-MIN)	0,20	0,25	0,06
Spatial auto-correlation	**Moran Index (p-value)**	0.81 (p < 0.001)	0.59 (p < 0.001)	0.05 (p < 0.001)
Local extreme values	**Absolute number**	167	159	271
	Relative number (to n)	0,38	0,36	0,32

Data sources:
Rainfall: in Germany, per county July 2017, interpolated from: https://www.dwd.de/DE/leistungen/cdcftp/cdcftp.html?nn=16102 (Climate Data Center)
Average age: in Germany, per county in 2011. https://www.destatis.de/DE/Methoden/Zensus_/Downloads/2F_BevoelkerungAlterGeschlecht.html
Social Index: Social monitoring Hamburg 2016. http://suche.transparenz.hamburg.de/dataset/sozialmonitoring-integrierte-stadtteilentwicklung-bericht-2016-anhang

Table 1. (Geo-)statistical properties of used data sets
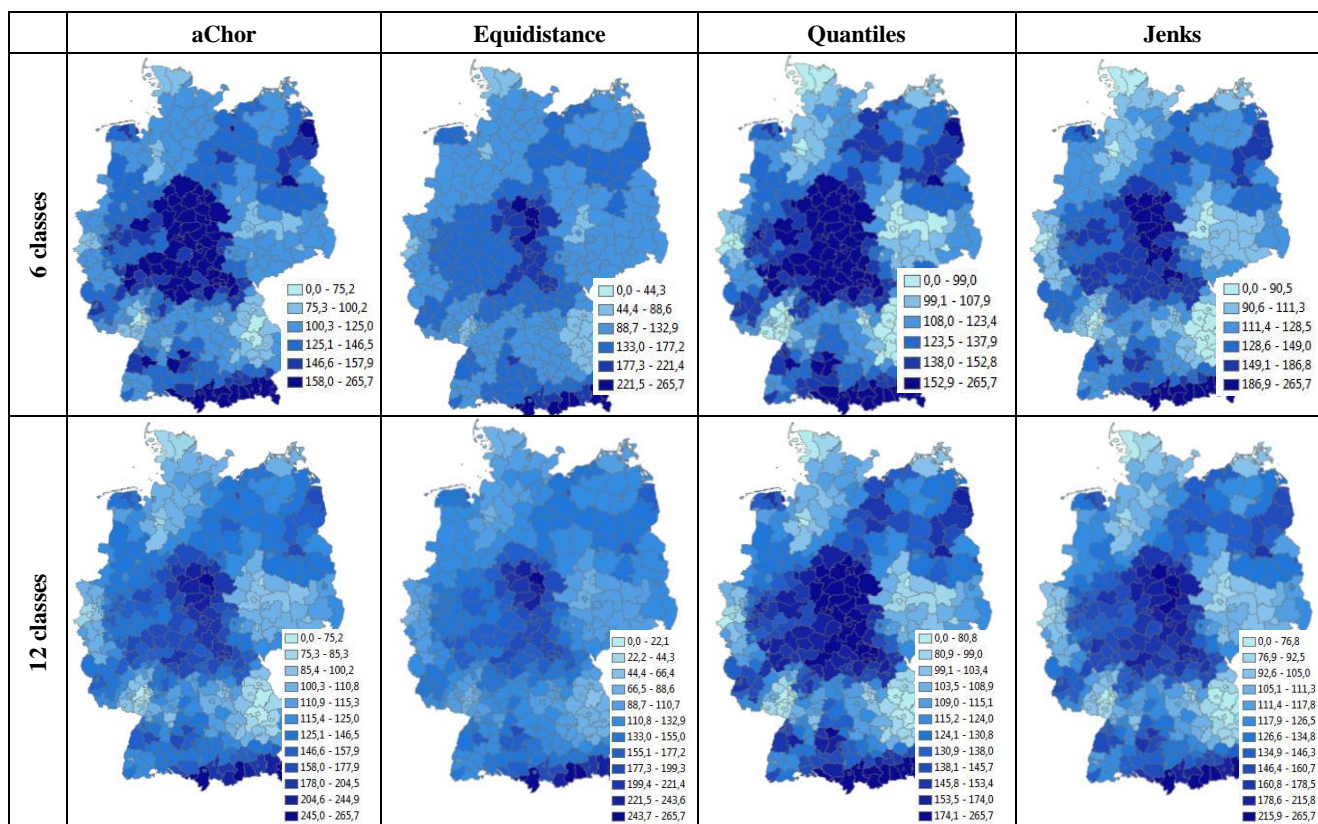(RMSE: root mean square error; MAX: Maximum value; MIN: minimum value; IQR: interquartile range)

Figure. 1. Example choropleth maps for one and the same data set ("Rainfall"), using different classification methods (from left to right) and class numbers (from top to bottom)

number classes and used classification methods as well as the correlations between measures and differences between measures based on various input data characteristics.

In addition to the typical classification methods equidistance, quantiles and natural breaks, also an own method designed for the preservation of local extreme values is used – as a representative for methods that considers spatial properties. This *aChor* method is described in detail in Schiewe (2017). For each method classifications are calculated for 4, 6, 8, 10 and 12 classes, which cover the typical range for choropleth maps taking the visual perception of human users into account (Slocum et al., 2009).

Three data sets with different (geo-)statistical properties are used for the analysis. Table 1 summarizes the respective characteristic values. For illustration purposes, figure 1 shows classifications for the "Rainfall" data set using the four different methods and two selected class numbers.

4.2 Results and Discussion

Figure 2 presents a graphical summary of results – showing the various uncertainty measures for the data sets under investigation.

If one looks at the **relationship between uncertainty measures and the number of classes (m)** the following conclusions can be drawn:

- Within-class homogeneity (GVF_g) increases with the increasing number of classes because the class intervals are getting smaller. This corresponds to the expectation - with the following theoretical extreme case: if m equals n (number of elements), then the variance within each class is zero and GVF_g becomes +1.
- Within-class matching shows the same effect: MAT_g increases with increasing m.
- Between-class heterogeneity (NN_g) increases with nc (theoretically: if m = c, then $NN_g = 1$ – i.e., all values are in their "best" class); however, for the rather small class numbers under investigation (m << n) a strict trend cannot be observed.
- Isolation of global extreme values (GEX) works perfectly if m = n. If m << n, a slightly monotonically increasing trend can be observed (with the exception of the quantiles method, see below).
- Preservation of local extreme values (LEX) gets better with increasing m. For all methods a monotonically increasing trend can be observed.
- Global visual imbalance (GVB) appears rather constant within each method, independently of the class numbers.
- Within-class visual imbalance (CVB_g) shows a monotonically increase. If m increases, also the probability decreases that large value differences appear within a class (again, with the extreme case of m = n: no value difference at all because only one value remains in each of the classes).

All in all, most of the measures show improved properties with increasing number of classes (i.e., increasing measures towards +1). In practice, however, this contradicts to the requirement of a rather small number of classes that is needed to guarantee sufficient visual differentiability.

Secondly, the behavior of each of the **uncertainty measures depending on the data classification method** is analyzed:

- Within-class homogeneity (GVF_g) shows rather similar trend behavior and values (generally, deviations smaller than 0.1) for all methods – with Jenks being always the best method and quantiles always the worst one. In the latter case this can be justified with the absence of an equal distribution of values of all data sets (see histograms in table 1).
- Within-class matching (MAT_g) shows very similar behavior compared to GVF_g.
- Between-class heterogeneity (NN_g) is, by definition, always perfect for equidistance method (i.e., $NN_g = 1$). For the other methods large deviations can be observed for small class numbers (4 and 6), while for larger class numbers deviations are smaller than 0.2. Following equidistance, Jenks gives best values (NN_g larger than 0.9).
- Isolation of global extreme values (GEX) shows strong variations between classification methods: By definition, quantiles (with comparably strong occupation of border classes) shows weak GEX values close to zero. In two out of three examples, equidistance shows best results.
- Preservation of local extreme values (LEX) shows monotonically increasing measures with increasing number of classes for all methods. By purpose, aChor always shows significantly best results (with differences up to 0.3 to all other methods). Quantiles, Jenks and equidistance are always following in this order.
- Global visual imbalance (GVB) appears best with quantiles, which is due to the fact that the data sets do not show huge area differences (no clear "Russia-Andorra" effect). Equidistance always delivers worst results.
- Within-class visual imbalance (CVB_g) is, as expected, always worst with the quantile classification because border classes have relatively strong occupation compared to other methods. Here, equidistance shows best results in all examples.

Summarizing this set of tests, it can be concluded that there is no "perfect" classification method. Instead, a task-oriented selection is needed. With the data sets used here, non-spatial measures have been met well with equidistance, while for preserving spatial patterns (like local extreme values) specific algorithms (like aChor) should be chosen. With respect to visual perception, contradictory results can be observed: By definition, quantiles deliver best results for global and worst results for within-class balance.

For practical purposes, it is desirable to reduce the huge set of uncertainty measures and to avoid dependencies between them, respectively. Hence, investigations are also conducted to evaluate **the correlation between different uncertainty measures**:

- Within-class homogeneity (GVF_g) and within-class matching (MAT_g) show correlation coefficients of (close to or exactly) +1. This is understandable because the mean value of class and the class value have a constant offset.
- High correlations (larger than +0.9) can also be found between GVF_g and LEX, MAT_g and LEX, GVF_g and CVB_g, MAT_g and CVB_g.
- On the other hand, the global visual balance shows variable (and mostly low) correlations with other measures.
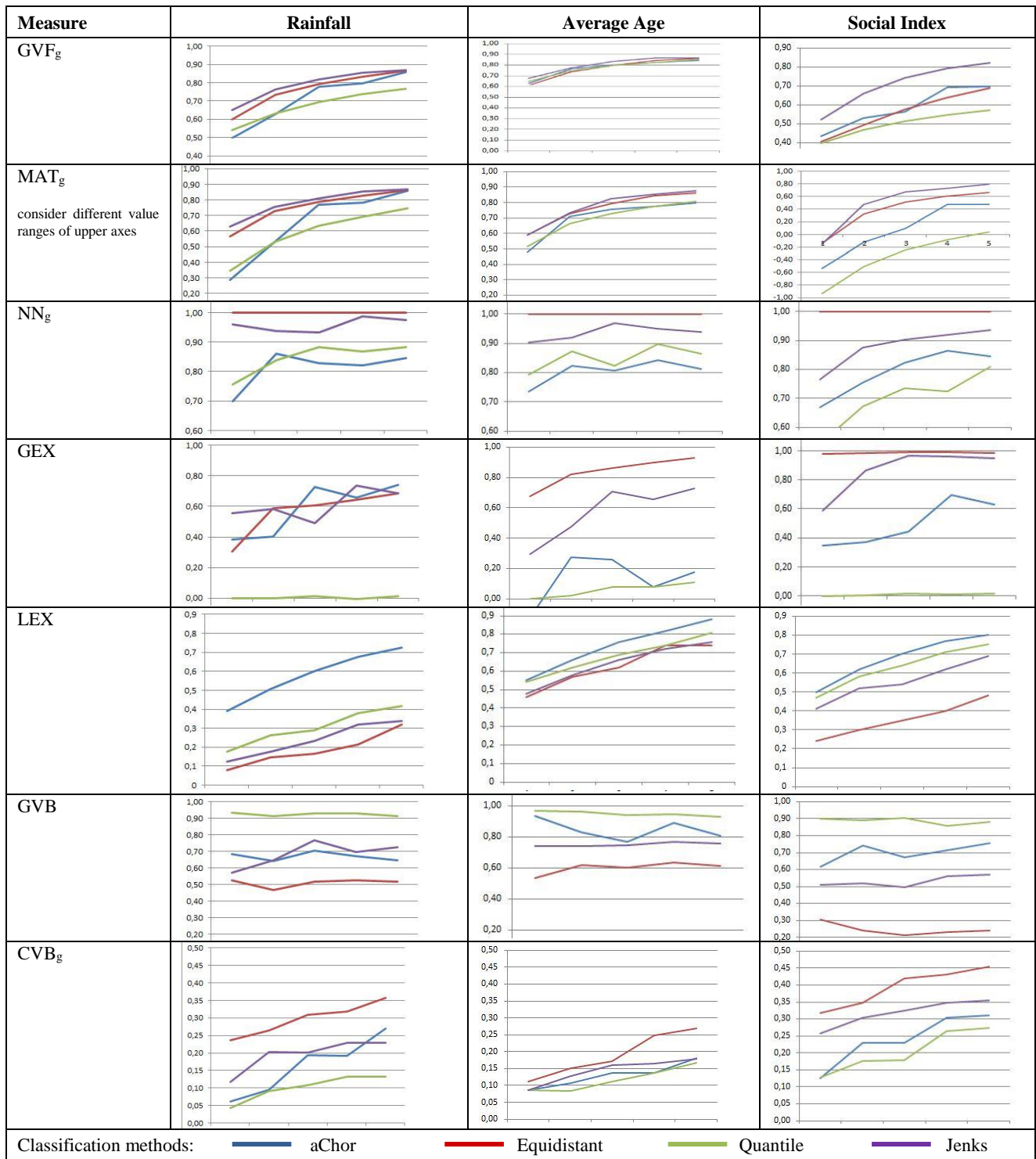
Figure 2: Graphical summary of empirical investigation results: Uncertainty measures (from top to bottom) using different data sets (from left to right). Right axis of each diagram shows number of classes (ranging from 4 to 12, using an interval of 2), upper axis shows uncertainty measure (always with maximum value of +1, representing the "best" case)

From this it can be concluded that general evaluations can reduce the set of non-spatial measures to one or two measures (e.g., GVF_g and NN). However, if a more detailed – and class-specific evaluation and post-processing of the classification is desired (see chapter 5), the whole set of parameters is still of interest.

Also for practical purposes it is desirable to predict uncertainty properties directly from the input data set. For this reason the **differences between uncertainty measures for the three data sets** are calculated and brought into connection with statistical parameters of input data (see table 1).

- Large differences (larger than |-0.5| on average) can be observed between data set "Social Index" and the other two for within-class matching (MAT_g). The significant lower value for "Social Index" can be justified with the smaller RMSE/span width ratio (0.07 against 0.17 or 0.18, resp.). This leads to smaller class intervals in the middle of the data set and broader classes towards to the boundaries. Due the large total number of polygons (n) in this data set these border classes are still well occupied and show very large internal variations that affect the overall measure. A similar effect (but with smaller differences in the order of 0.1 to 0.3 on average) can be observed for within-class homogeneity (GVF_g).
- The aforementioned histogram appearance of "Social Index" (accumulation in the middle, strong reduction towards boundaries) also leads to a better preservation of global extreme values compared to the other two data sets.
- Differences up to |-0.5| are detected for the local extreme value preservation (LEX) between "Rainfall" and the other two data sets. This reduced preservation rate is based on the large RMSE in this data set that leads to a stronger variation of values around local extreme value polygons and thus to a more difficult definition of class breaks that are able to separate all neighboring polygons.

As these described examples show that there are significant and logical dependencies between data set parameters and uncertainty measures. A thorough analysis of input data statistics (as given in table 1) can also speed up the analysis process and help to some extent to find appropriate classification methods (see chapter 5).

5. FURTHER USAGE OF UNCERTAINTY MEASURES

5.1 Multi-criteria analysis

So far, the various uncertainty measures have been treated independently. Apparently, the analysis and selection of classification parameters (especially, method and number of classes) is a multi-criteria process. Using the measures, a linear combination as target function with user-defined weights can be applied. If the weights are normalized (i.e., their sum is set to 1), the overall measure still has a maximum ("best") value of +1. Obviously, setting the weights needs pre-knowledge that is dependent on the application requirements.

The problem of the reduction to an overall criterion is that the individual target values are strictly not interconvertible and the weighting factors are subjective. Therefore, evolutionary or, even more specific, genetic algorithms serve as alternative approaches. For example, Pareto optimization performs a separate optimization for all combinations of weighting factors. The idea is pursued that one criterion is only improved until another is worsened. The uncertainty measures presented here can also be introduced into such a Pareto optimization.

5.2 Optimization of classified data sets

The presented uncertainty measures can not only serve the a-posteriori evaluation of the classifications, but also to improve them. A simple, but brute-force, method is to compute multiple classifications and select the best variant based on the derived uncertainty measures.

Alternatively, a calculated variant can also be optimized by considering the class-specific uncertainty measures in an a-posteriori manner. Figure 3 illustrates the idea: For a given variant, a significant drop in uncertainty (here: NN_c) is observed for the right border class. The shift of the relevant class boundaries and redistribution of the values leads to an improvement of the class-specific and thus also of the global uncertainty measure. If not only one measure is to be taken into account the above-mentioned Pareto principle can again be used for the optimization.

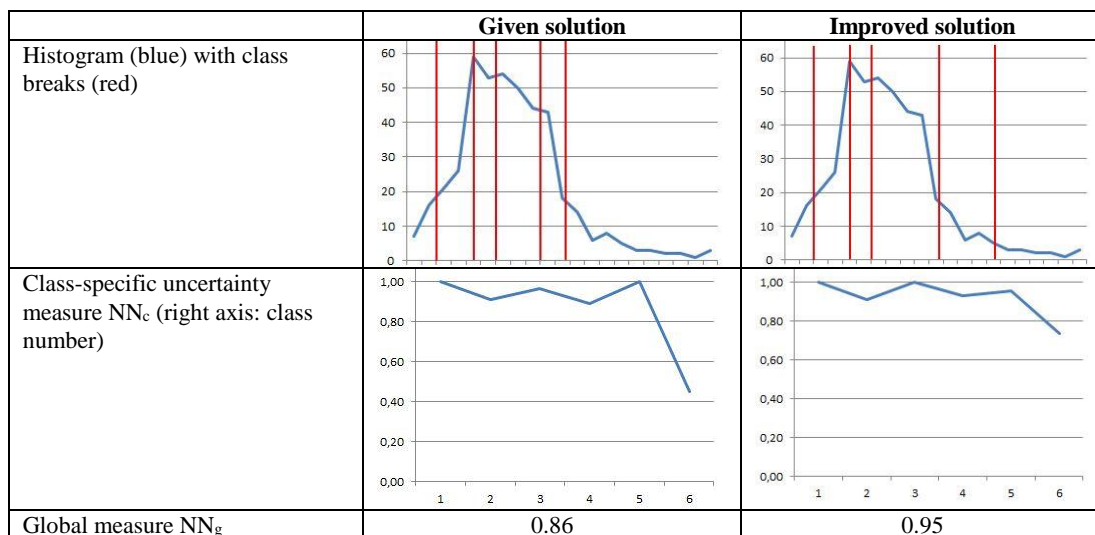|	Given solution	Improved solution
Histogram (blue) with class breaks (red)		
Class-specific uncertainty measure NN_c (right axis: class number)		
Global measure NN_g	0.86	0.95

Fig. 3. Improvement of given solution (here: aChor, 6 classes)
through shifting class breaks (two on the right hand side) and redistribution of values

6. SUMMARY AND FUTURE WORK

This paper aimed to define an extended and task-oriented set of uncertainty measures to quantify generalization effects as part of a data classification process. Not only non-spatial statistical properties were considered, but also the preservation of spatial patterns as well as uncertainties in the course of visual perception. The presented set of measures is not complete - especially for the latter two aspects extensions and further empirical investigations are possible and necessary.

A recommendation for a standardized procedure or selection of uncertainty measures based on the tests carried out here is not possible. It can be expected that due to the variable usages for the classified data, an adapted application-specific selection must take place. For this purpose, user-friendly instructions or interfaces are still missing.

It was also pointed out that uncertainty measures can be used not only for pure evaluation purposes but also for the optimization of existing classification results. Here the class-specific uncertainty measures are very helpful. Future work will address the design and implementation of appropriate evolutionary algorithms that address relevant cases (e.g., optimization of between-class heterogeneity).

ACKNOWLEDGEMENTS

REFERENCES

Andrienko, G., Andrienko, N. & Savinov, A., 2001. Choropleth maps: Classification revisited. Proceedings International Cartographic Conference, 9 p.

Armstrong, M.P., Xiao, N. & Bennett, D.A., 2003. Using Genetic Algorithms to Create Multicriteria Class Intervals for Choropleth Maps. *Annals of the Association of American Geographers*, 91(3): 595-623

Bregt, A.K. & Wopereis, M.C.S., 1990. Comparison of complexity measures for choropleth maps. *The Cartographical Journal*, 27(2): 85-91

Brewer, C.A. & Pickle, L., 2002. Evaluation of Methods for Classifying Epidemiological Data on Choropleth Maps in Series. *Annals of the Association of American Geographers*, 92(4): 662-681

Cromley, R.G., 1996. A comparison of optimal classification strategies for choroplethic displays of spatially aggregated data. *International Journal of Geographical Information Systems*, 10(4) 405-424

Cromley, E.K. & Cromley, R.G., 1996. An analysis of alternative classification scheme for medical atlas mapping. *European Journal of Cancer,* 32A(9) 1551-1559

Coulsen, M.R.C., 1987. In the matter of class intervals for choropleth maps: With particular reference to the work of George Jenks. *Cartographica*, 24(2): 16-39

Dent, B. D., 1999. Cartography: Thematic map design. 5th ed. Boston: McGraw-Hill

Getis, A. & Ord, J.K., 1992. The Analysis of Spatial Association by Use of Distance Statistics, *Geographical Analysis*, 24: 189-206.

Goldsberry, K. & Battersby, S., 2009. Issues of Change Detection in Animated Choropleth Maps. *Cartographica*, 44(3): 201-215

Jenks, G. & Caspall, F., 1971. Error on choroplethic maps: definition, measurement, reduction. *Annals of the Association of American Geographers,* 61: 217-24

Mersey, J.E., 1990. Choropleth Map Design – a Map User Study. *Cartographica*, 27(3): 33-50

Murray, A.T., 2000. Spatially lagged choropleth display. Proceedings of 9th international symposium on spatial data handling, International Geographical Union, Beijing, 10 p.

Schiewe, J., 2016. Preserving attribute value differences of neighboring regions in classified choropleth maps. *International Journal of Cartography,* 2(1): 6-19. DOI: 10.1080/23729333.2016.1184555

Schiewe, J., 2017. Data Classification for Highlighting Polygons with Local Extreme Values in Choropleth Maps. In: Peterson M.P. (ed.) *Advances in Cartography and GIScience*, Lecture Notes in Geoinformation and Cartography, Springer: 449-459

Schiewe, J. & Chang, J., 2018 (accepted): Erhaltung und Betonung räumlicher Muster in Choroplethenkarten. *GIS.Science*

Slocum, T.A., McMaster, R.B., Kessler, F.C. & Howard, H.H., 2009. Thematic Cartography and Geovisualization, 3rd edition. Prentice Hall, 576 p.

Sun, M., Kronenfeld, B.J. & Wong, D.W., 2013. Cartographic techniques for communicating class separability: enhanced choropleth maps of median household income, Iowa. *Journal of Maps*, 9(1): 43-49