APUSS

Assessing Infiltration and Exfiltration
on the Performance of Urban Sewer Systems

Contract number : EVK1-CT-2000-00072 APUSS
Project homepage : http://www.insa-lyon.fr/Laboratoires/URGC-HU/apuss

DELIVERABLE 5.2

Integrated model to assign and merge infiltration rates of sub-catchments

December 2004

Torsten Franz and Peter Krebs
Dresden University of Technology, Institute for Urban Water Management, 01062 Dresden
mail: Torsten.Franz2@mailbox.tu-dresden.de

Content

1 Objective

The goal of deliverable 5.2 was to develop a method which allows to transfer the results of in- and exfiltration measurements from sub-catchments to others in order to improve the knowledge about the I/E-situation in a catchment or city, respectively.

Two transfer methods based on statistical procedures were developed. By using these methods the available information can be aggregated on catchment scale.

2 Restrictions

2.1 Ground water information

For applying the transfer methods, information about the ground water situation in the catchment is essential. At least ground water influenced reaches must be identifiable or spatially uniform ground water conditions in relation to the reaches must exist.

2.2 Rates and classes

As a result of the available measuring methods of infiltration and exfiltration only mean rates per sub-catchment or for a relatively long channel distance are known. Thus, the developed transfer methods compare large populations of objects and their mean values. Due to the coarse resolution of the methods a calculation of rates is not reasonable but the definition of classes is necessary. The number of classes depends on the number of sub-catchments with known I/E rates and the range of measured rates.

2.3 Catchment coverage

The developed methods are based on the similarity approach (APUSS, 2004). They assign or compare reaches/sub-catchments with unknown I/E rates to reaches/sub-catchments with known I/E rates. Therefore, sub-catchments with unknown rates must fit the characteristics of sub-catchments with known rates for a succesful assignment. Thus, the coverage of all sub-catchments is not always possible.

2.4 Transferability between catchments

Recent research indicates that regional effects more than local ones might have an overwhelming influence on infiltration (Lucas and Fuchs, 2003). Based on that hypothesis a comparison of different cities at least within one region would be possible.

Figure 1 shows a regression between area-specific infiltration rate and the result of multidimensional scaling (MDS, see APUSS, 2004) for Dresden and Emscher sub-catchments. Significant differences between the cities from different regions can be observed.

Yet, it is not clear whether the differences of Lucas and Fuchs (2003) and Figure 1 are based on natural/regional factors like soil type or on artificial/local factors like building history. Thus, a comparison of cities is not recommended.

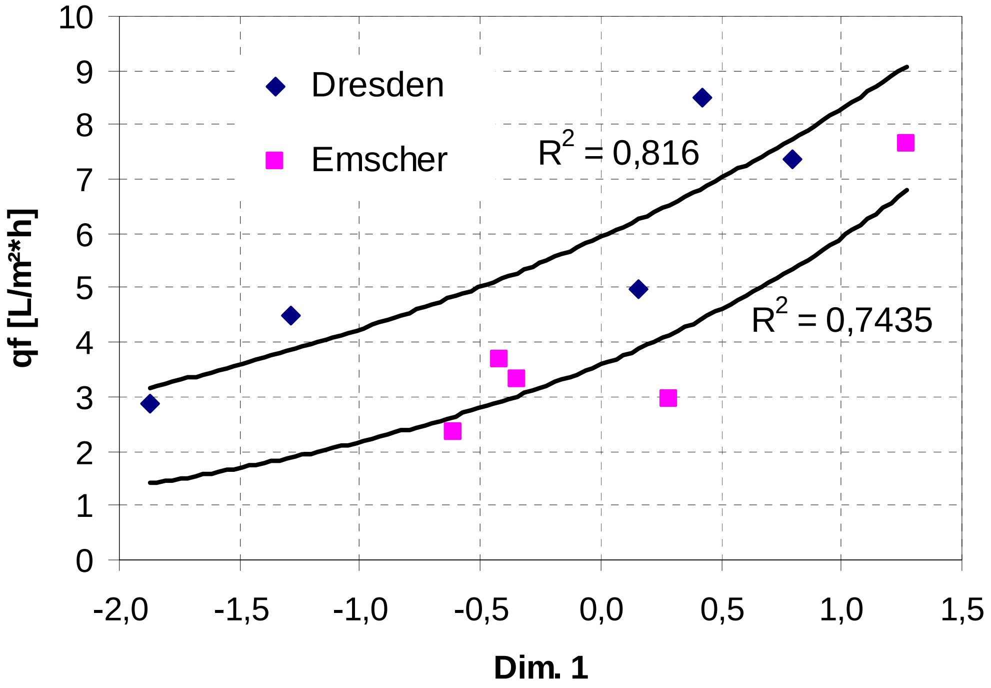


Figure 1: MDS result of Dresden/Emscher data set with groundwater information vs. infiltration rates based on the parameters reach surface, slope, population density, distance to surface waters and buildings

3 Transfer method based on analysis-of-variance

3.1 Statistical background

With analysis-of-variance methods (ANOVA) it can be investigated whether several grouped samples belong to one basic population. For that purpose they analyse variances to verify a significant inter-group difference of mean values. The observed variances are separated into a component based on the coincidental error (i.e. sum of the squares within the groups) and into components based on different mean values. The latter are tested on statistic significance.

ANOVA methods are suitable for non-nominal parameters only. Parameter can be analysed e.g. with one-way ANOVA (metric scale) and Kruskal-Wallis-ANOVA (ordinal scale), respectively (Mueller, 1991).

The connected post-hoc tests and multiple comparison tests determine which means contributed to the effect, i.e. which groups are particularly different from each other. Compared to the similar t-tests, post-hoc comparisons take into account the fact that more than two samples were taken.

3.2 Method

The ANOVA transfer method treats sub-catchments – represented by means and variances – as one coherent unit. The method determines parameter similarities between sub-catchments with unknown I/E class and sub-catchments with known I/E class. In cases of high similarities the I/E class is transferred.

The relevant parameters can be determined according to Franz (2004). For applying the method the parameters must meet the requirements of parametric ANOVA: metric scale, normal distribution and variance homogeneity. For applying multiple comparisons the parameters must be ranked in order to achieve an ordinal scale.

The sub-catchment forms the only categorical predictor variable. The parameters are analysed with post-hoc tests, e.g. the Fisher Least Significant Difference test (Winer *et al.*, 1991), and multiple comparison tests (Siegel and Castellan, 1988), respectively. The number of not rejected tests (i.e. two sub-catchments are determined as belonging to one population for the considered parameter) is summarised. Pairs of sub-catchments with a high sum can be assumed as similar and should have a similar I/E class.

Several calculations show that a change of the significance level within a reasonable range has a low influence on the result.

3.3 Examples

Cases with modelled infiltration rates

Along the lines of APUSS (2004) an infiltration model was implemented for six real data sets containing the parameters date of construction, profile circumference, length, population density, slope and head:

$$Q_{\inf} = \frac{A_S * LENGTH}{\sqrt{DATE\_CONSTR} * (100 * e^{2*POP\_DENS}) * (10 * SLOPE)^2}$$

where A_S = groundwater-influenced pipe surface
LENGTH = reach length
DATE_CONSTR = date of construction
POP_DENS =population density
SLOPE = slope

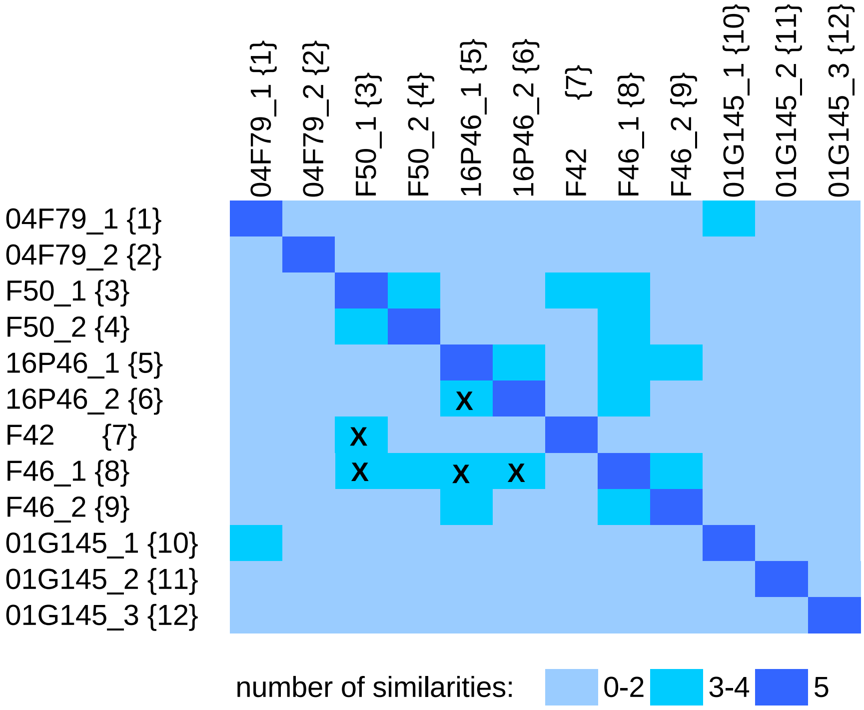
The reach populations were split in order to increase the number of sub-catchments. Regarding the relatively low number of sub-catchments the model-based infiltration rates were equidistantly classified in three classes (Table 1).

Table 1: Sub-catchments with modelled infiltration classes

Sub-catchments	**Origin**	**q_f [L/m²*h]**	**Infiltration class**
04F79_1	Dresden	0.003287	Low
04F79_2	Dresden	0.000858	Low
F50_1	Emscher	0.019093	Low
F50_2	Emscher	0.070007	High
16P46_1	Dresden	0.013085	Low
16P46_2	Dresden	0.003681	Low
F42	Emscher	0.000939	Low
F46_1	Emscher	0.008154	Low
F46_2	Emscher	0.038934	Middle
01G145_1	Dresden	0.056954	High
01G145_2	Dresden	0.034782	Middle
01G145_3	Dresden	0.021446	Low

Figure 2 shows a matrix of the sum of similarities between all paired sub-catchments based on the Least Significant Difference test (LSD). As five parameters were used for the infiltration model, five is the maximum number of similarities. Due to their distribution the minimum number of similarities to be investigated was set to three or 60 %. From the 15 % of pairs identified to be similar only 50 % were correctly assigned (crosses in Figure 2).

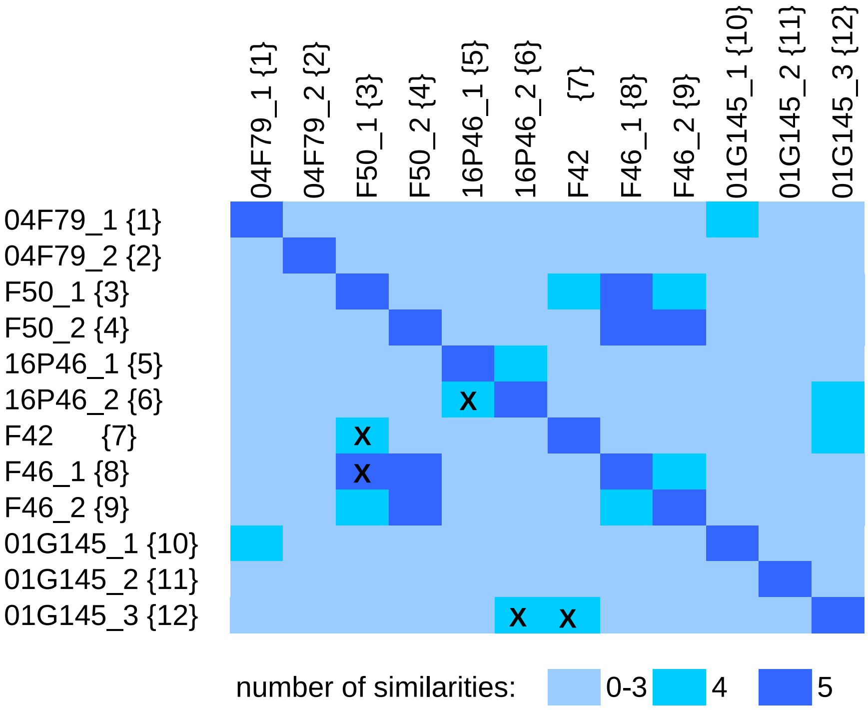
Figure 3 shows a matrix based on multiple comparisons, i.e. the parameters were ranked beforehand. Although this non-parametric test has a lower discriminatory power the results do not differ much. The minimal number of similarities to be investigated was set to four or 80 %. From the 15 % of pairs identified to be similar only 50 % were correctly assigned.



	04F79_1 {1}	04F79_2 {2}	F50_1 {3}	F50_2 {4}	16P46_1 {5}	16P46_2 {6}	F42 {7}	F46_1 {8}	F46_2 {9}	01G145_1 {10}	01G145_2 {11}	01G145_3 {12}
04F79_1 {1}	█	░	░	░	░	░	░	░	░	▒	░	░
04F79_2 {2}	░	█	░	░	░	░	░	░	░	░	░	░
F50_1 {3}	░	░	█	▒	░	░	▒	▒	░	░	░	░
F50_2 {4}	░	░	▒	█	░	░	░	▒	░	░	░	░
16P46_1 {5}	░	░	░	░	█	▒	░	▒	▒	░	░	░
16P46_2 {6}	░	░	░	░	▒ X	█	░	▒	░	░	░	░
F42 {7}	░	░	▒ X	░	░	░	█	░	░	░	░	░
F46_1 {8}	░	░	▒ X	▒	▒ X	▒ X	░	█	▒	░	░	░
F46_2 {9}	░	░	░	░	▒	░	░	▒	█	░	░	░
01G145_1 {10}	▒	░	░	░	░	░	░	░	░	█	░	░
01G145_2 {11}	░	░	░	░	░	░	░	░	░	░	█	░
01G145_3 {12}	░	░	░	░	░	░	░	░	░	░	░	█

number of similarities: ░ 0-2 ▒ 3-4 █ 5

Figure 2: Frequency of similarities between sub-catchments - modelled infiltration classes, parametric tests



	04F79_1 {1}	04F79_2 {2}	F50_1 {3}	F50_2 {4}	16P46_1 {5}	16P46_2 {6}	F42 {7}	F46_1 {8}	F46_2 {9}	01G145_1 {10}	01G145_2 {11}	01G145_3 {12}
04F79_1 {1}	█	░	░	░	░	░	░	░	░	▒	░	░
04F79_2 {2}	░	█	░	░	░	░	░	░	░	░	░	░
F50_1 {3}	░	░	█	░	░	░	▒	█	▒	░	░	░
F50_2 {4}	░	░	░	█	░	░	░	█	█	░	░	░
16P46_1 {5}	░	░	░	░	█	▒	░	░	░	░	░	░
16P46_2 {6}	░	░	░	░	▒ X	█	░	░	░	░	░	▒
F42 {7}	░	░	▒ X	░	░	░	█	░	░	░	░	▒
F46_1 {8}	░	░	█ X	█	░	░	░	█	▒	░	░	░
F46_2 {9}	░	░	▒	█	░	░	░	▒	█	░	░	░
01G145_1 {10}	▒	░	░	░	░	░	░	░	░	█	░	░
01G145_2 {11}	░	░	░	░	░	░	░	░	░	░	█	░
01G145_3 {12}	░	░	░	░	░	▒ X	▒ X	░	░	░	░	█

number of similarities: ░ 0-3 ▒ 4 █ 5

Figure 3: Frequency of similarities between sub-catchments - modelled infiltration classes, non-parametric tests

Cases with measured infiltration rates

The method was applied to the Dresden data set without groundwater information (Franz, 2004). The following parameters were selected for the analysis: date of construction, profile circumference, population density, population-specific length, distance to river Elbe, distance to surface water, distance to storm sewers, distance to drainage, coverage, thickness of cohesive layers and distance to groundwater. Regarding the relatively high number of sub-catchments the model-based infiltration rates were equidistantly classified in five classes (Table 2).

The resulting matrix based on LSD is shown in Figure 4. Due to their distribution the minimal number of similarities to be investigated was set to five. From the remaining 6 % of pairs identified to be similar only 33 % were correctly assigned (crosses), 87 % with an error of one class (circles).

Figure 5 shows a matrix based on multiple comparisons, i.e. the parameters were ranked beforehand. The minimal number of similarities to be investigated was set to seven. From the remaining 6 % of pairs identified to be similar only 40 % were correctly assigned (crosses), 93 % with an error of one class (circles).

Despite the violation of the restriction about groundwater information (chapter 2.1) the results are valid, because spatially uniform ground water conditions in relation to the reaches can be assumed for the pairs identified to be similar.

Table 2: Sub-catchments with real infiltration classes, without ground water information

Sub-catchments	**q_f [L/m²*h]**	**Infiltration class**
01F6	1.20	Low
01G145	0.90	Low
01H34	1.73	Middle
04C16	0.52	Very Low
04F79	0.33	Very Low
08S83	0.52	Very Low
09I1	0.78	Low
15H24	2.07	High
15X54	0.18	Very Low
15Z33	0.73	Very Low
16P46	1.88	High
17Y16	1.67	Middle
29EE27	0.77	Low
ED	1.48	Middle
MA	0.80	Low
OD	1.14	Low
RC	1.56	Middle
SB	0.90	Low
SF	1.02	Low
WE	2.95	Very High
ZWEZG 08K120	2.22	High
ZWEZG 36B13	0.92	Low

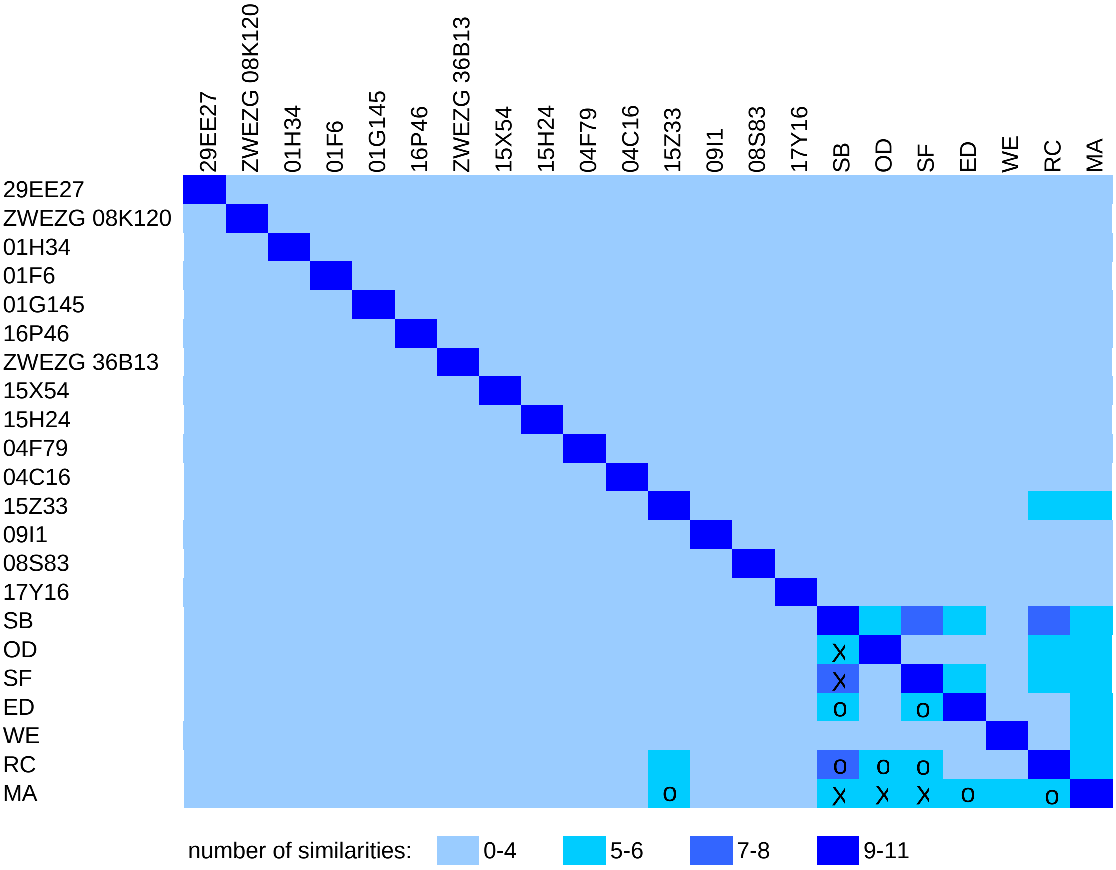


Figure 4: Frequency of similarities between sub-catchments – measured infiltration classes, parametric tests

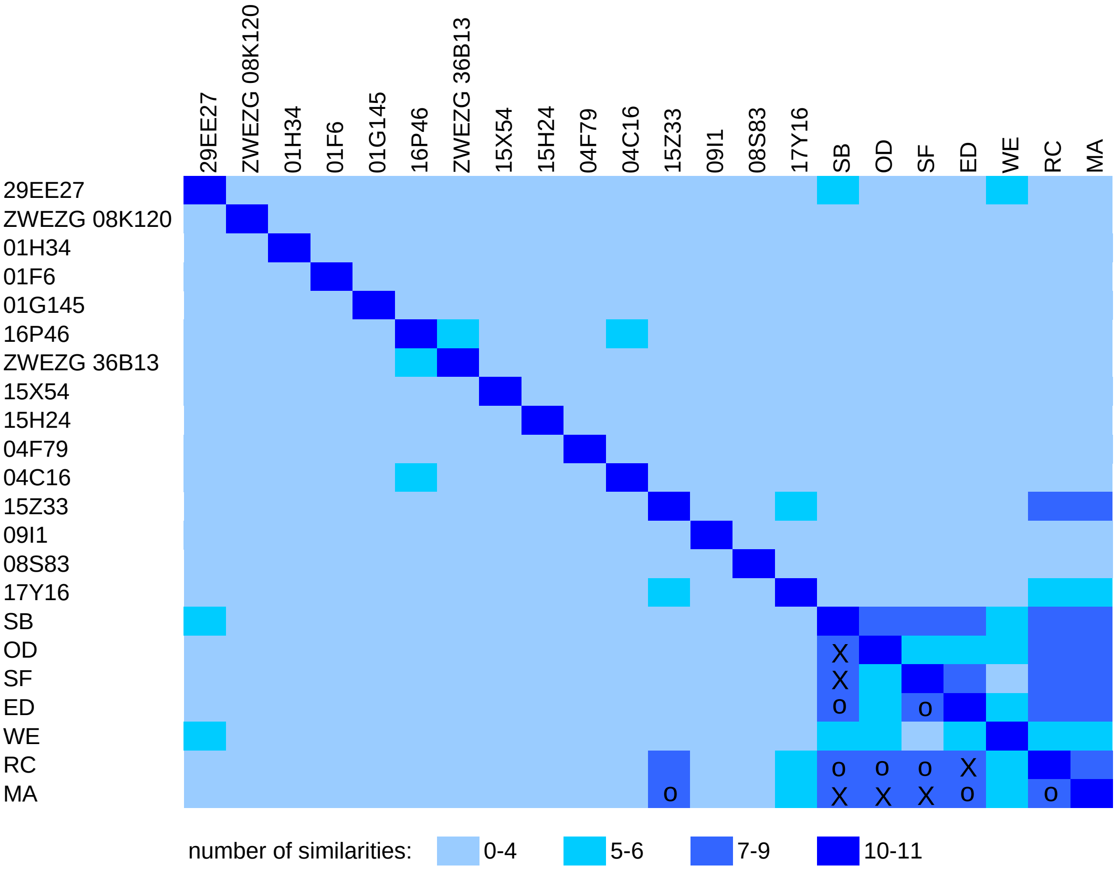


Figure 5: Frequency of similarities between sub-catchments – measured infiltration classes, non-parametric tests

4 Transfer method based on discriminant analysis

4.1 Statistical background

Discriminant analysis is used as either a hypothesis testing or exploratory method. For hypothesis testing it determines which variables discriminate between several naturally occurring groups. For data exploration it groups objects based on their attributes into known classes by using classification functions or analysing the distances between the objects and the groups centroids. A priori probabilities, i.e. unequal group sizes, can be considered. For an overview see Fahrmeir *et al.* (1996) or Backhaus *et al.* (1996).

With the distance concept, an object with an unknown group affiliation is classified to the group where the Mahalanobis distance d_{nm} (Fahrmeir *et al.*, 1996) between object *n* and groups centroid *m* is minimal:

$$d_{nm} = \sqrt{(n-m)^T S^{-1} (n-m)}$$

where S = empirical covariance matrix
n, m = parameter vectors

The distances between the object *n* and all groups *m* = 1,…,M can be represented by classification probabilities *P(m|n)* (Backhaus *et al.* 1996):

$$P(m \mid n) = \frac{\exp\left(-d_{nm}^2 / 2\right) * P_n(m)}{\sum_{m=1}^{M} \exp\left(-d_{nm}^2 / 2\right) * P_n(m)}$$

where $P_n(m)$ = a priori probability for affiliation of object *n* to group *m*

4.2 Method

The discriminant transfer method considers single reaches. The method assigns the reaches of a sub-catchment with unknown I/E class to groups/sub-catchments with known I/E class. The I/E class of the investigated sub-catchment is calculated as a weigthed mean.

Relevant parameters can be determined according to Franz (2004). They must meet the requirements of discriminant analysis: metric scale, normal distribution and variance/covariance homogeneity.

For all reaches *n* = 1,…,N of the sub-catchment with unknown I/E class and all groups *m* = 1,…,M the classification probabilities *P(m|n)* are calculated. For every reach *n* the mean I/E class $c_{I/E,n}$ is determined with

$$c_{I/E,n} = \sum_{m=1}^{M} P(m \mid n) * c_{I/E,m}$$

where $c_{I/E,m}$ = I/E class of sub-catchment *m*

The overall I/E class $c_{I/E,sub\text{-}c}$ for the investigated sub-catchment is calculated as the reach surface weighted mean

$$c_{I/E,sub-c} = \frac{\sum_{n=1}^{N} c_{I/E,n} * A_n}{\sum_{n=1}^{N} A_n}$$

where A = surface of reach n

4.3 Examples

Cases with modelled infiltration rates

The cases described in Table 1 were analysed with the discriminant transfer method. The results are shown in Table 3. The assignment error is calculated with

$$error = \left| c_{I/E,m} - c_{I/E,sub-c} \right|$$

where $m = sub\text{-}c$

Assuming a suitable assignment error of 0.66 class width (equivalent to one third of the range) 50 % of the sub-catchments were correctly assigned. Assuming a softer error of 1.0 class width 75 % of the sub-catchments were correctly assigned.

Table 3: Results for sub-catchments with modelled infiltration classes

Sub-catchments	Infiltration class	Transfer result	Assignment error
04F79_1	1	1.9	0.88
04F79_2	1	1.2	0.17
F50_1	1	1.3	0.26
F50_2	3	1.3	1.66
16P46_1	1	1.9	0.87
16P46_2	1	1.2	0.16
F42	1	1.0	0.05
F46_1	1	2.4	1.42
F46_2	2	2.7	0.69
01G145_1	3	1.8	1.18
01G145_2	2	1.9	0.06
01G145_3	1	1.2	0.20

Cases with measured infiltration rates

The method was applied to the Dresden data set with groundwater information (Franz, 2004). The following parameters were selected for the analysis: date of construction, profile circumference, length, population density, distance to river Elbe, distance to buildings, distance to streets and slope. The measured infiltration rates were equidistantly classified in three classes (Table 4). The results are shown in Table 5.

Assuming a suitable assignment error of 0.66 class width 1 of the 5 sub-catchments were correctly assigned. Assuming a softer error of 1.0 class width 4 of 5 sub-catchments were correctly assigned.

Table 4: Sub-catchments with measured infiltration classes with ground water information

Sub-catchments	q_f [L/m²*h]	Infiltration class
01F6	5.00	Middle
04C16	4.50	High
04F79	2.87	High
16P46	7.35	Low
ZWEZG 36B13	8.52	Low

Table 5: Results for sub-catchments with measured infiltration classes

Sub-catchments	Infiltration class	Transfer result	Assignment error
01F6	2	2.6	0.57
04C16	3	2.1	0.88
04F79	3	2.1	0.93
16P46	1	2.2	1.18
ZWEZG 36B13	1	1.0	0.00

Other cases

Beside the above mentioned cases the following cases were analysed:

- more groups with modelled infiltration rate than described in Table 1
- five instead of three infiltration classes
- mean weighting due to sewer length L_n

$$c_{I/E,sub-c} = \frac{\sum_{n=1}^{N} c_{I/E,n} * L_n}{\sum_{n=1}^{N} L_n}$$

The results of these investigations did not differ significantly from the above illustrated cases.

5 Conclusion

With the proposed application of the similarity approach – the transfer of I/E measurement results from subcatchments to others – it is possible to improve the information about the I/E status of sewer networks. With a suitable error defined as acceptable the transfer of infiltration classes was successful for approximately half of the sub-catchments. Keeping in mind that only few data were available the positive verification of the methods is indicated but not completely satisfying.

Among the two developed methods the ANOVA transfer method is the weaker one. Due to the facts that (i) sub-catchments are treated as one coherent unit and (ii) they are a priori well discriminated (APUSS, 20004) the potential catchment coverage of this method is low. Actually, less then 50 % of the sub-catchments with modelled rates and less then 25 % of the sub-catchments with real rates were assignable to others.

The advantage of the ANOVA transfer method is that it has to comply with less strict assumptions. In comparison to the discriminant transfer method

- ordinal parameters can be included in the analysis
- a test whether sub-catchments with unknown rates fit the characteristics of sub-catchments with known rates is not necessary.

The discriminant transfer method considers single reaches. Hence, it has a greater potential. Keeping in mind that just three infiltration classes were established the method lead to acceptable results. Between 40 % and 80 % of the sub-catchments could be assign correctly.

The similarity approach concludes from known cases to unknown ones. Thus, complete catchment coverage is not possible. Regarding the observed error rates the main problems while applying the two methods are

- What are – beside the non-linearity of I/E processes – the reasons for incorrect assignments?
- How can these misclassifications be identified a priori?

6 Blind alleys

During developing the methods several procedures were investigated which do not lead to results or were not applicable.

Multidimensional scaling

The results of APUSS (2004) and Figure 1 point to a functional relationship between the reduced parameter dim. 1 gained by multidimensional scaling and the specific infiltration rate. This relationship could not be found for cases with modelled I/E rates. Therefore, MDS was not applied for transfer methods.

Convex hull

The restriction that sub-catchments with unknown rates must fit the characteristics of sub-catchments with known rates was proved by a comparison of minimal and maximal values of the parameters.

Assuming the reaches as objects in an *n*-dimensional space where every dimension is related to one parameter (see APUSS, 2004) convex hulls for the respective reach populations can be defined (Preparata and Shamos, 1993). Reaches which are not within one or several hulls would be ignored. Thus, misclassifications can be better avoided.

Up to now, it was not possible to find an algorithm which defines convex hulls for $n > 3$. Therefore, this restriction was not applied yet.

Neural networks

A very comprehensive review about neural networks (NN) can be found in Sarle (1997). Neural networks can be used for classification purposes. Since the discriminate analysis is sufficient and easier to apply, NNs were not implemented.

7 References

APUSS (2004): Classification method to identify sub-catchments for an efficient measurement procedure and transfer of measuring results. Deliverable 5.1.

Backhaus K., Erichson B., Plinke W. and Weiber R. (1996): Multivariate Analysemethoden (Methods of Multivariate Analysis.) Springer: Berlin, Heidelberg, New York. ISBN 3-540-60917-2.

Fahrmeir L., Hamerle A. and Tutz G. (eds.) (1996): Multivariate statistische Verfahren. (Multivariate statistical methods.) Walter de Gruyter: Berlin, New York. ISBN 3-11-013806-9.

Franz T. (2004): Reduction of Parameter Number. Internal report for the APUSS project.

Lucas S. and Fuchs S. (2003): Regionalisierung von Fremdwasserproblemen (Regionalisation of extraneous water problems). *KA Abwasser Abfall*, **50**, 302-307.

Mueller P. H. (ed.) (1991): Wahrscheinlichkeitsrechnung und mathematische Statistik. (Probability calculation and mathematical statistics.) Akademieverlag: Berlin. ISBN 3-05-500608-9.

Preparata F. and Shamos M. (1993): Computational geometry. Springer: Berlin, Heidelberg, New York. ISBN 3-540-96131-3.

Sarle W.S. (ed.) (1997): Neural Network FAQ, periodic posting to the Usenet newsgroup comp.ai.neural-nets, ftp://ftp.sas.com/pub/neural/FAQ.html. 30.12.2002.

Siegel S. and Castellan N. (1988): Nonparametric statistics for the behavioral sciences. 2nd ed. McGraw-Hill: New York.

Winer B., Michels K. and Brown D. (1991): Statistical principals in experimental design. New York: McGraw-Hill.