The transfer of infiltration measurement results to other sewers by means of discriminant analysis

Torsten Franz and Peter Krebs

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

ABSTRACT

The cost-oriented and sustainable operation of sewer systems requires a comprehensive knowledge about the infiltration situation in the catchment. Owing to the high expenditures for infiltration measurements a reliable transfer of measurement results to other sewer sections would be highly beneficial. Assuming a functional relationship between sewer characteristics and infiltration rates can be identified, such a transfer can be realised by means of classification techniques. In this paper a method is introduced which is based on discriminant analysis and which allows for a transfer of measurement results to similar sub-catchments. The method was applied using two data sets with measured or virtual infiltration rates. It yields acceptable results as a total fraction of 50% to 75% of the investigated sub-catchments was assigned correctly. Furthermore, additional information to assess the results was provided. The quality of the transfer results depends strongly on the homogeneity of the considered sub-catchments. Due to this restriction the practical applicability of the method is restricted. Nevertheless, it might be used as a screening procedure for planning of effective detailed infiltration investigations.

Key words | discriminant analysis, exploratory data analysis, extraneous water, infiltration, measurement, sewer, transfer of result

INTRODUCTION

Extraneous water in sewers might have serious consequences such as increased discharges of waste water to the receiving waters causing pollution and public health risks as well as increased pumping and treatment costs. Operation and maintenance strategies which consider groundwater infiltration as main source of extraneous water require a comprehensive knowledge about the infiltration situation in each individual sub-catchment. Due to the high expenditures of infiltration measurements (APUSS 2004) a reliable transfer of measurement results to comparable sewer sections or sub-catchments would be valuable in order to reduce the measurements efforts significantly. Furthermore, planning instructions (e.g. ATV-A 118 1999) often demand to estimate potential infiltration rates based on measured ones from similar catchments.

Assuming a functional relationship between sewer characteristics and infiltration rates such a transfer can be realised by means of modelling or classification techniques. Since modelling requires calibration (and thus additional measurements), a definite data set and a deeper process-related knowledge, a transfer based on classification (i.e. identifying similar sewers with similar infiltration conditions) is a much simpler and cost-effective approach.

In this paper a method is introduced which is based on the classification technique of discriminant analysis and which allows for a transfer of measuring results to similar sub-catchments. In terms of the functional relationship named above the classification approach has black-box characteristics. Dynamic processes or long-term developments cannot be handled. Thus, because of its moderate dynamics, only infiltration is considered as a source of extraneous water. On the other hand, because of its statistical nature the method has a high degree of freedom against data needs and

can be adapted to nearly every data situation. The results gained with the developed procedure are to be critically evaluated within the boundary conditions of the given data.

METHODS

Discriminant analysis

There are several multivariate classification techniques. As a suitable, easy-to-apply and powerful method the discriminant analysis was chosen, which belongs to the exploratory data analysis methods and evaluates structures in data sets. It is used either as hypothesis testing or as exploratory method. For hypothesis testing, discriminant analysis determines which variables discriminate between several naturally occurring groups, i.e. data is reduced and attributes and variables are ranked. For data exploration items (i.e. reaches) are grouped into classes (i.e. sub-catchments) based on their attributes by using classification procedures. The main difference to cluster analysis whose task is similar is that the groups or clusters are known at the beginning of the analysis.

Computationally, discriminant analysis is similar to analysis of variance and multiple regression analysis (Hill & Lewicki 2006). On the one hand, the method compares matrices of total and within-group variances and covariances by means of multivariate F-tests in order to determine significant differences between the groups. This procedure is identical to multivariate analysis of variance. On the other hand, it determines analogous to multiple regression analysis linear discriminant functions or so-called canonical roots which discriminate the groups by means of a discriminant value $y_j = f(x_1, x_2, \ldots, x_n)$. The functions y_j are a (n_1) dimensional representation of the n-dimensional attribute space. The isoquant $y_j = 0$ is a separation plane between groups.

This is shown in Figure 1. The attributes x_1 and x_2 span a two-dimensional attribute space. Items o_j characterised by the attributes form two groups G_1 and G_2. Distributions of attribute's values and overlaps of the groups are marked by bars. A discrimination of the groups based solely on the attribute x_1 (if $x_{1,o_1} \leq x_{1,\text{crit}}$ then $o_1 \in G_1$, if $x_{1,o_1} \geq x_{1,\text{crit}}$ then $o_1 \in G_2$) yields suboptimal results due to the large overlap of both groups. The discrimination is much better when using the one-dimensional discriminant function y_1, because it minimises the overlapping (checkered bar).
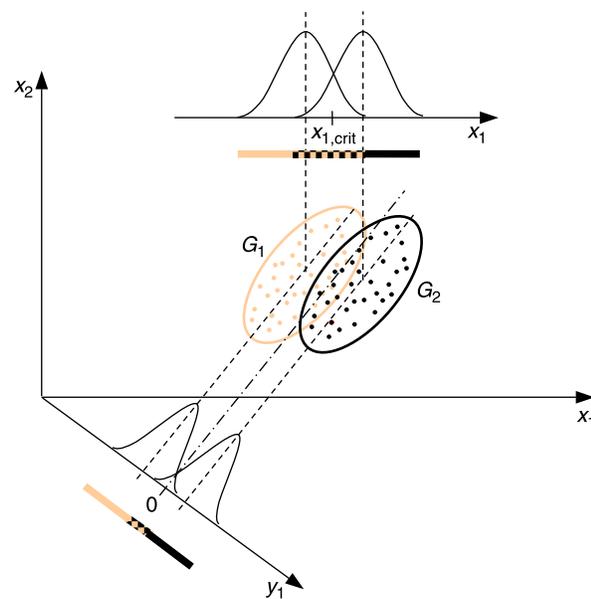

Figure 1 | Discriminant function.

New items can be classified to known groups by means of several concepts: the distance concept using calculated distances between items and group centroids, the definition of classification functions and the probability concept as statistical decision problem. For the development of our method the latter was chosen, because it allows for a classification into multiple groups and thus for potentially subtler results. The classification probabilities of the item o_j are calculated with

$$P(G_i|y_{o_j}) = \frac{e^{\frac{-d(o_j, \bar{o}_i)^2}{2}} P_j(G_i)}{\sum_{k=1}^{n_{\text{group}}} e^{\frac{-d(o_j, \bar{o}_k)^2}{2}} P_j(G_k)} \tag{1}$$

with

$P(G_i|y_{o_j})$ classification probability with which o_j belong to G_i, $\sum_i P(G_i|y_{o_j}) = 1$

$P_j(G_i)$ *a priori* probability for affiliation of o_j to G_i

$d(o_j, \bar{o}_i)$ Mahalanobis distance between o_j and the centroid $\bar{o}_i$ of G_i

Owing to the relationship between discriminant analysis and analysis of variance the following requirements apply: normal distribution of the data, homogeneity of variances and covariances, no correlations between means and variances. Beside this, the attributes which are used to discriminate

between groups must not be completely redundant. A comprehensive explanation and evaluation of the discriminant analysis is given e.g. in Backhaus *et al.* (1994).

Transfer method

Assumptions. Primarily, the transfer method is based on two assumptions:

1. There is a functional relationship between sewer characteristics and infiltration rates. Thus, sewers with similar characteristics have similar infiltration rates (Franz & Krebs 2006).
2. A sufficient number of infiltration-relevant attributes are known. Beside structural data this includes a wider consideration of the structure "sewer" with building techniques, bedding, external loads, investment history, groundwater regime and other factors (O'Reilly *et al.* 1989; Davies 2001; Franz 2007).

Basic concept. The introduced transfer method is a three-stage procedure based on the discriminant analysis. First, the classification probabilities $P(sb-ctm_i|y_{\mathrm{reach}_j})$ of all reaches reach_j of a sub-catchment with an unknown infiltration class $cl_{\mathrm{infiltration}}$ classified to sub-catchments $sb\text{-}ctm_i$

with known infiltration classes are calculated (Figure 2). Thereby, the investigated reaches are classified and assigned, respectively, to sub-catchments with available infiltration measurements. Owing to the classification of individual reaches instead of sub-catchment means the method considers the attribute variations.

Second, the predicted infiltration class $cl_{\mathrm{infiltration,predicted},j}$ for the reach reach_j is determined with the probability weighted mean value

$$cl_{\mathrm{infiltration,\ predicted},j}\underset{\mathrm{reach}_j\notin sb-ctm_i}{=}\sum_{i=1}^{n_{sb-ctm}} P(sb-ctm_i|y_{\mathrm{reach}_j})cl_{\mathrm{infiltration},i} \quad (2)$$

Third, the overall predicted infiltration class $cl_{\mathrm{infiltration,predicted}}$ for the investigated sub-catchment is calculated as the weighted mean. As weighting factor the reach surface $A_{\mathrm{surface},j}$ was chosen (Infraguide 2005).

$$cl_{\mathrm{infiltration,\ predicted}} = \frac{\sum_{j=1}^{n_{\mathrm{reaches}}} cl_{\mathrm{infiltration},j}A_{\mathrm{surface},j}}{\sum_{j=1}^{n_{\mathrm{reaches}}} A_{\mathrm{surface},j}} \quad (3)$$

Data requirements. The data set which is necessary for the transfer of infiltration measurement results contains all
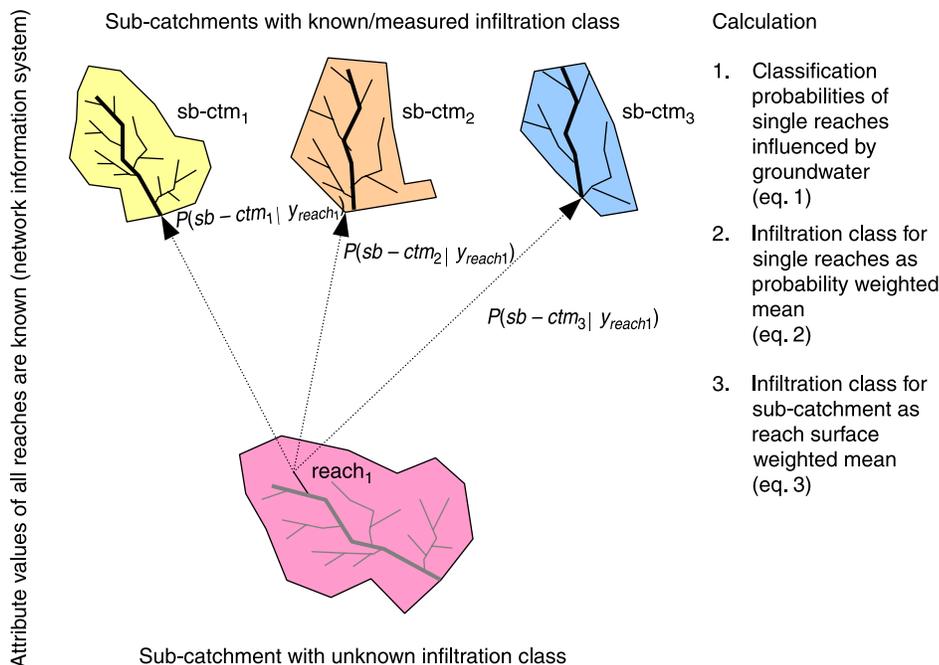

Figure 2 │ Transfer method.

reaches of the considered sub-catchments with their attribute values. Owing to the dependency of the infiltration process from groundwater, only groundwater influenced reaches are considered. With the leakage approach commonly used for infiltration modelling (Gustafsson 2000; Karpf & Krebs 2004), the water head between groundwater table and waste water level or pipe invert might be used as weighting factor. The classification procedure has a high degree of freedom against data needs. Thus, the set of independent attributes and variables, respectively, is not defined beforehand. Every available attribute can be included, if it (1) meets the requirements of the discriminant analysis described above and (2) is relevant for infiltration processes. As a result of this freedom, the result quality will strongly depend on the individual data situation of the catchment.

The data need to be pre-processed. Beside outlier elimination or distribution transformation the attribute values must be normalised. Otherwise the attributes with a large range would have an unintentional weighting. The measured infiltration rates should be classed. Several test calculations showed, that the use of infiltration rates leads to uncertain results. With classes, i.e. an information reduction, this uncertainty can be reduced significantly.

It should be tested whether the considered reaches match the characteristics of sub-catchments to which they are classified in order to avoid extrapolation. An effective method would be the determination of convex hulls of reach populations in the attribute space. The convex hull is the smallest convex shape which encloses a set of points. Reaches which are not within one or several hulls should be ignored during classification and cannot be evaluated. However, the determination of a convex hull in an n-dimensional space with $n > 3$ is extremely complex. Therefore, this procedure

was not applied and a simplified one is proposed. The constraint of characteristic's matching is proved by a comparison of the distributions of all attributes. If the distributions of the respective sub-catchments and the observed attribute values, respectively, overlap sufficiently, the constraint is assumed to be fulfilled.

RESULTS AND DISCUSSION

The transfer method was applied and evaluated on two data sets (Table 1). The data sets contain a number of real sub-catchments from which network structure and a number of reach characteristics were known. The infiltration rates were either measured or calculated with a virtual non-linear infiltration model. Regarding the relatively low number of considered sub-catchments the infiltration rates were classified equidistantly in three classes, separated for both sets (Table 2).

The constraints of discriminant analyses as well as of characteristics matching between sub-catchments with unknown infiltration rates and those with known rates were fulfilled for all cases. For each sub-catchment in a set, the infiltration class was predicted by means of the given classes of the other sub-catchments. The assignment quality and error, respectively, was determined by the difference between the predicted and the actual infiltration class.

The results of the infiltration class transfer are shown in Table 2. The first data set with virtual infiltration rates has a mean assignment error of 0.63 class-width, which is equivalent to approximately one third of the range. Taking 0.3 class-width or 15% of the total range as a tolerable error, 50% of the sub-catchments were correctly assigned.

Table 1 │ Data sets for verification purposes

	Data set 1	Data set 2
Sub-catchments	12 with 93 km total length	5 with 193 km total length
Reach attributes	Year of construction, profile circumference, length, population density, slope	Year of construction, profile circumference, length, population density, population-specific length, distance to surface water, distance to buildings, distance to streets, coverage, thickness of cohesive layers
Groundwater	Virtual	Measured
Infiltration rates	Virtual, per reach	Measured, per sub-catchment

Table 2 │ Results of infiltration class transfer

| Data set 1–virtual infiltration				Data set 2–measured infiltration			
Sub-catchment	Infiltration class [−]	Predicted class [−]	Assignment error [−]	Sub-catchment	Infiltration class [−]	Predicted class [−]	Assignment error [−]
04F79_1	1	1.9	0.88	**Complete data set**			
04F79_1	1	1.2	0.17	01F6	2	3.0	1.00
F50_1	1	1.3	0.26	04C16	3	3.0	0.00
F50_2	3	1.3	1.66	04F79	3	1.0	2.00
16P46_1	1	1.9	0.87	16P46	1	2.7	1.75
16P46_2	1	1.2	0.16	36B13	1	1.0	0.05
F42	1	1.0	0.05	**Reduced data set**			
F46_1	1	2.4	1.42	01F6	2	2.6	0.57
F46_2	2	2.7	0.69	04C16	3	2.1	0.88
01G145_3	3	1.8	1.18	04F79	3	2.1	0.93
01G145_3	2	1.9	0.06	16P46	1	2.2	1.18
01G145_3	1	1.2	0.20	36B13	1	1.0	0.05

Accepting an error of 1.0 class-width, 75% of the sub-catchments were correctly assigned. The second data set with measured infiltration rates has a mean assignment error of 0.96 class-width. With a tolerated assignment error of 0.3 class-width, 40% of the sub-catchments were correctly assigned, with the softer error of 1.0 class-width 60% of the sub-catchments were matching.

The transfer result of data set 2 was improved by using a reduced data set without the attributes population-specific length, coverage and thickness of cohesive layers. The resulting assignment errors have a maximum of 1.18 class width and 80% of the sub-catchments were correctly assigned within an accepted error of 1.0 class width. Apparently, the classification quality of this reduced data set is higher. The reason might be that the attributes which were excluded can be assumed to be redundant.

Regarding the black-box characteristics of the transfer method the results are acceptable. But, they suffer from high attribute variances and thus a suboptimal classification quality which has to be faced in real sub-catchments. A comparison between the assignment errors and the fraction of reaches correctly classified by means of discriminant analysis indicates that a high assignment error might occur parallel to a lower classification quality (Figure 3). This underlines the strong dependency of transfer results from the homogeneity of investigated sub-catchments. Thus, an increase of the total homogeneity, i.e. an *a priori* determi-nation of sub-catchments optimised with regard to their individual homogeneity (Franz & Krebs 2006), is expected to improve significantly the catchment coverage as well as the result reliability.

The significance of the results regarding the applica-bility of the transfer method has to be attenuated because only one sub-catchment was classified per run, i.e. the ratio between classified sub-catchments and sub-catchments to classify was 11:1 (data set 1) and 4:1 (data set 2), respectively. Therefore, the potential assignment quality under practical conditions might be overestimated.
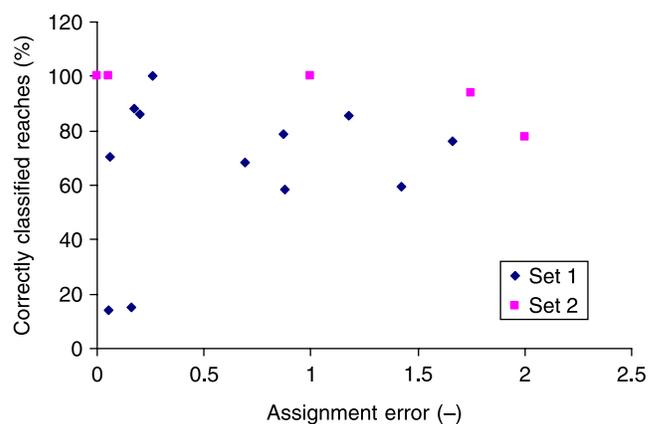
Figure 3 │ Comparison between assignment error and classification quality.

Additionally, the following cases were analysed:

- more sub-catchments with virtual infiltration classes
- more virtual infiltration classes
- weighting of the mean infiltration class due to the sewer length.

The results of these investigations did not differ significantly from the cases illustrated in this paper. As expected, the use of measured infiltration rates instead of infiltration classes did lead to a reduced quality of the results.

The basic approach of the measurement transfer method is promising, as the transfer of infiltration classes succeeded with some tolerance for the majority of the sub-catchments identified to be similar. Nevertheless, the method is not satisfactory for practical purposes. The potential catchment coverage (the "overlapping" of sub-catchment characteristics) as well as the rate of correct assignments are too low. Despite these restrictions, the transfer method might be used as pre-test or screening procedure for planning of effective detailed infiltration investigations. By determining sub-catchments which are not similar to those with known infiltration rates and by comparing classification results of different sub-catchments, it is possible to identify a ranking of further measurements. With this, the measurement effort can be reduced.

CONCLUSION

The introduced method allows for a transfer of infiltration measurement results to other sub-catchments. Such a transfer is recommended at least as a concept during planning and rehabilitation procedures. The method is based on the statistical method of discriminant analysis. It treats single reaches and considers variations within the data sets. Hence, it has a great potential for applicability which is reflected by the verification results. Keeping in mind that just three infiltration classes were established and that the ratio between classified sub-catchments and sub-catchments to classify was high, the method yields acceptable results with a total fraction of 50% to 75% correctly assigned sub-catchments. Furthermore, the method delivers additional information to assess the results.

The quality of the transfer method depends strongly on the individual data situation as well as on the homogeneity of the considered sub-catchments. Thus, a modified measuring point configuration aiming at homogeneous sub-catchments is expected to improve the result reliability as well as the catchment coverage significantly. Furthermore, the knowledge of the groundwater situation seems to be very important. Owing to these restrictions the method is not yet satisfactory for field purposes. Nevertheless, the transfer methods might be used as a screening procedure in the run-up to detailed infiltration investigations in order to reduce the measurement effort or to improve information.

REFERENCES

APUSS 2004 Deliverable 11.2 - Cost-benefit analysis. http://www.insa-lyon.fr/Laboratoires/URGC-HU/apuss/, visited 23.03.05.

ATV 1999 Arbeitsblatt ATV-A 118 – Hydraulische Bemessung und Nachweis von Entwässerungssystemen (Hydraulic dimensioning of sewer systems). ISBN 3-933707-25-0.

Backhaus, K., Erichson, B., Plinke, W. & Weiber, R. 1994 Multivariate Analysemethoden (Methods of Multivariate Analysis), 7th edn. Springer, Berlin, ISBN 3-540-56908-1.

Davies, J. P. 2001 The identification and investigation of the factors associated with rigid sewer pipe deterioration and collapse. Centre for environmental health engineering (CEHE), University of Surrey.

Franz, T. 2007 Spatial classification methods for efficient infiltration measurements and transfer of measuring results. Dissertation, Institut für Siedlungs- und Industriewasserwirtschaft, TU Dresden. Dresdner Bericht 28, ISSN 1615-083X.

Franz, T. & Krebs, P. 2006 Statistical methods towards more efficient infiltration measurements. *Water Sci. Technol.* **54** (6–7), 153–160.

Gustafsson, L. G. 2000 Alternative drainage schemes for reduction of inflow/infiltration – prediction and follow-up of effects with the aid of an integrated sewer/aquifer model. *Proc. 1st International Conference on Urban Drainage via Internet*.

Hill, T. & Lewicki, P. 2006 *Statistics: Methods and Applications*, 1st edition. StatSoft, Tulsa, OK, ISBN 1-884233-59-7.

Infraguide 2005 Inflow/Infiltration Control/Reduction for Wastewater Collection Systems. www.infraguide.ca, visited 10.10.2005.

Karpf, C. & Krebs, P. 2004 Sewers as drainage systems – quantification of groundwater infiltration. *Proc. Novatech 2004*, Lyon/France, 6–10 June 2004.

O'Reilly, M. P., Rosbrook, R. B., Cox, G. C. & McCloskey, A. 1989 *Analysis of defects in 180 km of pipe sewers in Southern Water Authority*. TRRL Research Report 172.