Design and optimisation of monitoring networks in urban drainage

G.J.R. Henckens[*] and F.H.L.R. Clemens**

* Witteveen + Bos Consulting Engineers, P.O. Box 233, 7400 AE Deventer, The Netherlands, e-mail: G.Henckens@witbo.nl (corresponding author, tel 0031 570 697 316, fax 0031 697 344)

** Delft University of Technology, dept. civil eng. sanitary eng, PO Box 5048 Delft, E-mail f.h.l.r.clemens@citg.tudelft.nl

Keywords: *monitoring networks, optimisation, urban drainage, information content*

1. Introduction

Investments in the improvement of an urban drainage systems to abate flooding or environmental problems are normally based on model results holding unknown, but suspectedly, significant deviations[1]. These deviations are due to several sources of errors (see e.g. Clemens & von der Heide, 1999 or Clemens, 2001[a]). An effective method to quantify, and eventually reduce, these deviations between model and measured data is model calibration (see Henckens et al, 2003 and Clemens, 2001[a]). In order to obtain enough information to calibrate a model (i.c. a hydrodynamic model) monitoring results are needed; in the case of a hydrodynamic model water level, discharges and rainfall data are the most fundamental quantities to be monitored as a function of time.

When setting out to design a monitoring network several questions are to be answered;

- How many locations should be monitored (in practice it is impossible to monitor every manhole and conduit)?
- Which locations are most fit for obtaining information?
- What measuring accuracy and frequency are needed?

Basically this can be seen as an constrained optimisation problem; optimise the information content using a limited (due to budget) number of sensors. A method to achieve this was proposed by Clemens (2001[b] and 2002) and has been successfully applied in practice in Loenen (Witteveen+Bos, 2001).

In other practical application though, some limitations of the method applied became apparent. In part this was caused by the introduction of some simplifications in the original algorithm and partly by new insights with respect to optimal sensor location. In practice it turns out that, for instance, an ‘overlap’ in sensor information is a perquisite for a successful model calibration, while the original model was tailored to identify sensor locations leading to the absence of redundant information. Therefore an algorithm is developed with the following characteristics:

- All calculation simplifications will be removed
- A certain preferred mutual overlap´ in sensor information can be set
- A minimal number of monitoring results per location per storm event is obligatory (e.g. at least 30 readings in access of 5 cm water depth per storm).

In this paper the Beekbergen catchment (combined sewer system) is used as example. Figure 1 shows this system, external weirs and pumping stations are indicated.

---

[1] The risk involved using such model results in making decisions for investments is described in some detail by Korving (2001)

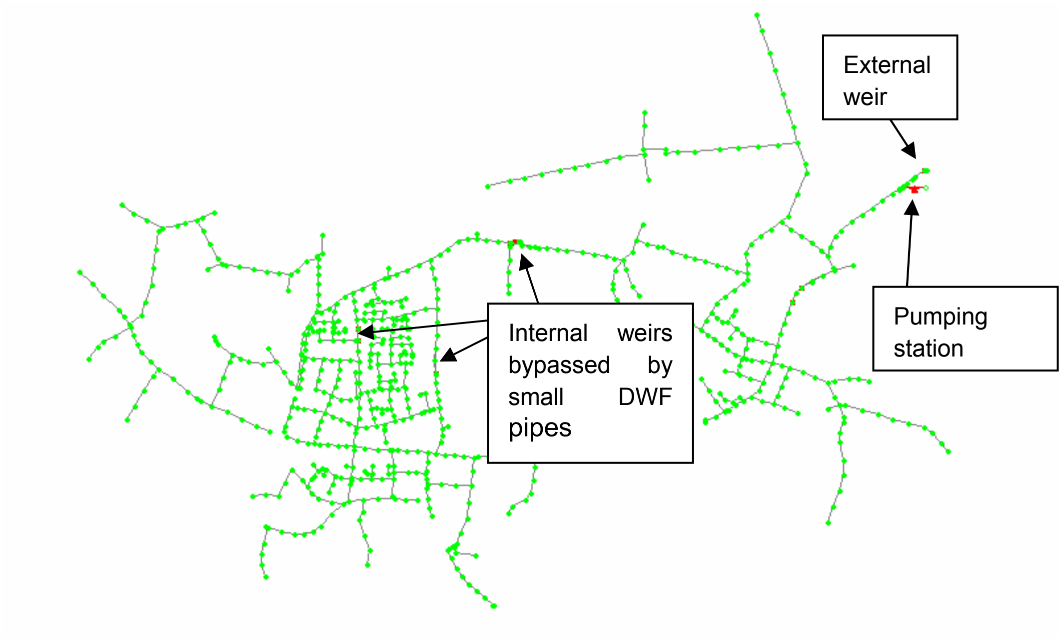

Figure 1: The sewer system of Beekbergen

2. Measurement network optimisation

Network optimisation is based on the potential information content of the various possible monitoring network geometry's. This information content is carried by the Jacobean matrix:

$$\underline{\underline{J}}_h = \begin{bmatrix} \frac{\partial e_{h_1}}{\partial p_1} & . & . & . & \frac{\partial e_{h_1}}{\partial p_n} \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ \frac{\partial e_{h_m}}{\partial p_1} & . & . & . & \frac{\partial e_{h_m}}{\partial p_n} \end{bmatrix}$$

In which:

e_{hi}	a model results (i.c. either discharge or water level) at a certain location at a certain moment
p_i	model parameter no. i

The Jacobean is basically a generalised sensitivity matrix in which the change in model results δe_i (water level, discharge) as a results of a change in the model parameters δp_i is quantified. The relevance of this information is obvious; when the model results show no variation when changing a model parameters it is impossible to obtain information in a monitoring network for that parameter.
The application of the Jacobean in estimating the information content is shown by comparing two methods:

The first method applied is based on making sensitivity analysis for every individual model parameter. To this end model runs are made with variation in one parameter at a time in order to get an estimate of the Jacobean by application of finite differences:

$$\underline{\underline{\hat{J}}}_h = \begin{bmatrix} \frac{\Delta e_{h_1}}{\Delta p_1} & \cdot\ \cdot\ \cdot & \frac{\Delta e_{h_1}}{\Delta p_n} \\ \cdot & \cdot\ \cdot\ \cdot & \cdot \\ \cdot & \cdot\ \cdot\ \cdot & \cdot \\ \cdot & \cdot\ \cdot\ \cdot & \cdot \\ \frac{\Delta e_{h_m}}{\Delta p_1} & \cdot\ \cdot\ \cdot & \frac{\Delta e_{h_m}}{\Delta p_n} \end{bmatrix} \approx \begin{bmatrix} \frac{\partial e_{h_1}}{\partial p_1} & \cdot\ \cdot\ \cdot & \frac{\partial e_{h_1}}{\partial p_n} \\ \cdot & \cdot\ \cdot\ \cdot & \cdot \\ \cdot & \cdot\ \cdot\ \cdot & \cdot \\ \cdot & \cdot\ \cdot\ \cdot & \cdot \\ \frac{\partial e_{h_m}}{\partial p_1} & \cdot\ \cdot\ \cdot & \frac{\partial e_{h_m}}{\partial p_n} \end{bmatrix}$$

In this manner the information that is held by every potential individual monitoring location is quantified as the deviation between model results (either water level or discharge) obtained for different parameter values. In figure 2 the results of such quantification for the sewer system of Beekbergen (municipality of Apeldoorn) are shown. In this picture red indicates a large impact of a change in a weir coefficient (for weir FEO062) while blue implies little to no impact. In Beekbergen seventeen water level sensors have actually been installed, allowing for evaluations of the design by real life data.

The first method does not take into account the added information supplied by the changing of the water levels itself. This information is included in the Jacobean matrix. Clemens (2001, 2002) has proposed a method of using the singular value decomposition of the Jacobean matrix to calculate the optimal measurement locations. In this paper the method is discussed and some improvements are proposed. In figure 3 the calculated information content of the indicated internal weir FEO062 is shown. The Jacobean is calculated according to the same model runs as used for figure 1. As can be seen it is possible to use the singular value decomposition of the Jacobean matrix to provide insight in the influence of a parameter in the sewer system model, and thus it is possible to use it in an optimisation (using for instance a combinatorial evolutionary method (Boomgaard 2001)).

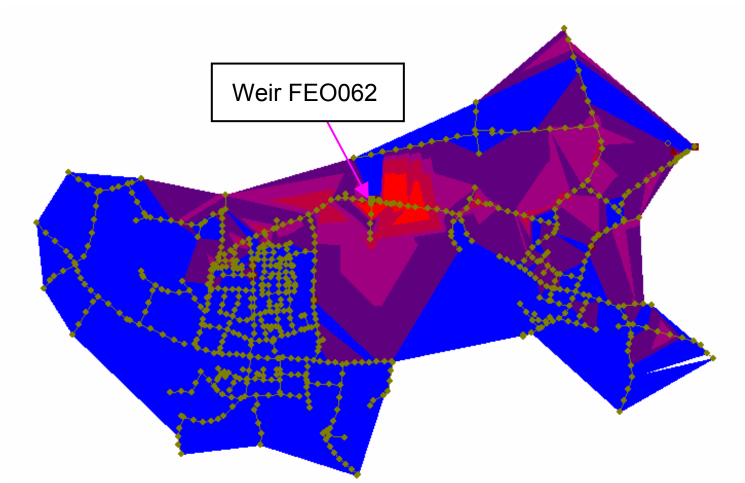

Figure 2: Information content calculated directly from water levels

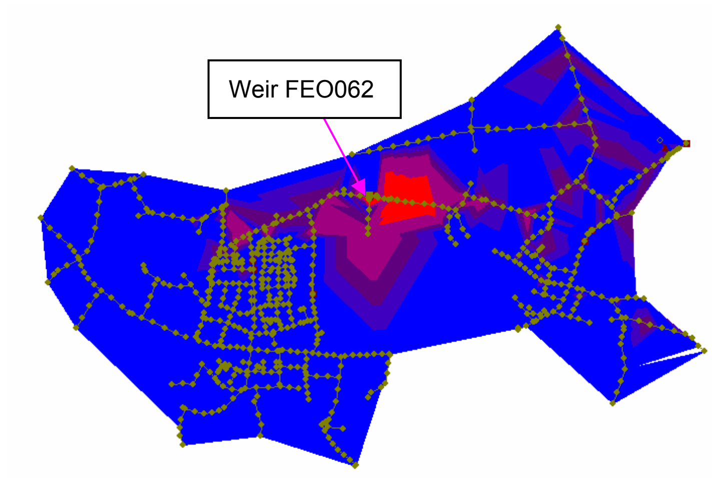

Figure 3: Information content calculated from the Jacobian

While the Jacobean quantifies the information content of the model parameters it does not suffice to calculate an optimal measurement network. Basically the optimisation method can be split in two separate parts (though in the final calculation these two are combined):

1. Calculate the information of the measurement locations
2. Calculate the joint information content by “punishing” the locations for high information correlation (as a measure for redundancy))

In the end the joint information has to suffice for the model calibration.

To illustrate why the total system correlation can not be used as weigh factor, as was once proposed during the project, figure 4 shows a contour plot of both the total information content of all manholes and the total correlation factor of these manholes. The Beekbergen system is, by Dutch standards, very steep. Apparently this causes little redundancy in the information obtained.

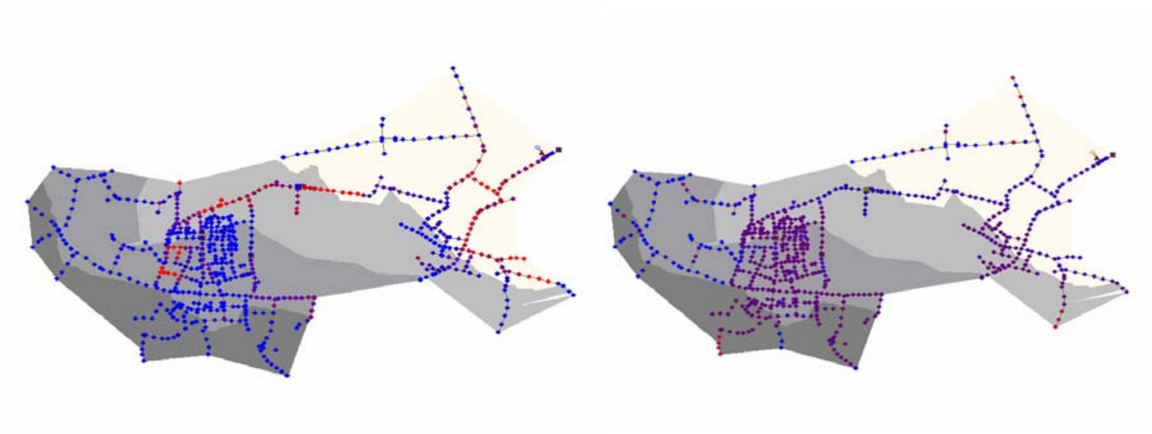

Figure 4: Total information content (left) and total correlation (right) plotted with surface level contours. Red is a high correlation/information content while blue present a low correlation/information content. Dark grey represent high surface levels, light grey low ground levels.

3. Information of the measurement location

When the water level at a location is monitored, information on several model parameters can be acquired. Based on the Jacobean matrix it is possible to extract an information amount per parameter (see figure 5). Since the information content of the network as a whole is calculated also, it is possible to quantify the relative information content per parameter per location. Dividing the information content per parameter by the total information content per parameter and then summing the factors provides insight in the total information content of a location (see formula 1).

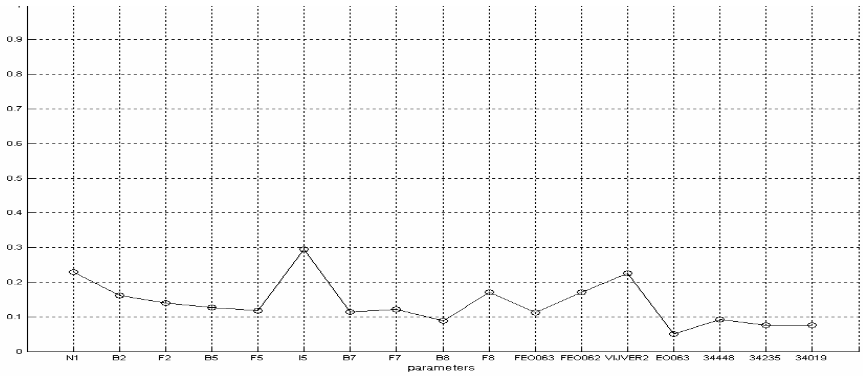

Figure 5: relative information content in 1 manhole. The x-axis shows the various parameters such as the number of inhabitants per hectare impervious surface (N1), the depression storage constants of the various surfaces (Bx), the routing constant (Fx), infiltration (I) and some weir coefficients.

Formula 1:

$$\frac{Pp_1}{Pw_1} + \frac{Pp_2}{Pw_2} + ... + \frac{Pp_n}{Pw_n} = T$$

In which :

P_i = parameter no. i
p = part (1 location)
w = whole (whole network
$_{1,2,\ldots,n}$ = number of parameter
T = total information content

Formula 1 introduces the first optimisation choice; a summation carries the total information content of a measurement location, while the relative importance of the individual parameters depends on the number of parameters. For instance, in the direct vicinity of a weir the information content on that specific weir coefficient will be relatively large. If that is the only parameter used, the preferred measurement location is easily identified. In case 15 parameters are used, that location might carry little information on the other parameters and will therefore possibly be excluded from the monitoring network.

Summation of relative information content leads to an optimal monitoring network which will be able to provide sufficient information to calibrate the model. A risk involved in doing so is that an uneven distribution in the relative information content over the individual parameters may evolve. As a result potentially crucial information (or just commercially interesting information) such as the added bonus of semi-exact weir-coefficients may be lost. Therefore one can argue that the optimisation should be done by optimising the information content for each individual model parameter. This however, would undoubtedly result in a monitoring network holding a large number of sensors and therefore in more expensive solutions.

4. Joint information

Calculating the information content of one location is relatively straightforward, however calculating the joint information content demands some difficult choices. Principal choices about the decorrelation method and the weight factor of the decorrelation factor are needed.

4.1 Decorrelation method

The decorrelation of the original method was unsatisfactory as it "punished" both measurement locations if they had a high correlation instead of just one of the two. For the total information content of a network of (for instance) 3 sensors the decorrelation factor was calculated by formula 2. Using the inverse of the correlation coefficient is a logical assumption due to the nature of the correlation calculation but also has a great impact on the method as a whole (see also paragraph 4.2)

Formula 2:

$$\frac{\left(\frac{1}{abs(C_{11})}+\frac{1}{abs(C_{12})}+\frac{1}{abs(C_{13})}\right)}{\left(\frac{1}{abs(C_{11})}+\frac{1}{abs(C_{12})}+\frac{1}{abs(C_{13})}+\frac{1}{abs(C_{21})}+\frac{1}{abs(C_{22})}+\frac{1}{abs(C_{23})}+\frac{1}{abs(C_{31})}+\frac{1}{abs(C_{32})}+\frac{1}{abs(C_{33})}\right)}=W_1$$

in which:
C_{12} = Correlation factor (0-1) between location 1 and 2
W_1 = total weight factor for location 1

For the new optimisation algorithm a decorrelation was initially chosen where the location with the highest information content is selected (location A), the rest of the locations is "punished" by the inverse of the correlation they have with location A. Then they are ordered again and the next best location is selected. And so forth till the last location, which carries the least information is selected as the best (and last) one.
For 3 sensors the decorrelation factor can be calculated by formula 3, assuming for the sake of convenience that the sensor number (1-3) also signifies the order informationwise.

Formula 3:

$$1 = W_1$$

$$1 \times \frac{1}{C_{21}} = W_2$$

$$1 \times \frac{1}{C_{31}} \times \frac{1}{C_{32}} = W_3$$

Formula 3 has two main drawbacks:

- A certain correlation between sensors is often preferred for inter reference and network security in the case of sensor failure. However correlation is always punished with this formula
- The factor W is always equal or higher than 1. This is not desirable because it interferes with possible analysis later on of the total information content (if this ever will be possible is unknown and shall be researched).

The second drawback can be solved by changing the division 1/*C* in 1-*C.* However this has a direct effect on the magnitude of the weight factor. The first drawback has been briefly discussed in the research goals. A certain overlap can be introduced to at keep at least the option open to sacrify some information for security. In the final algorithm formula 4 is used to calculate the decorrelation factor.

Formula 4

$$1 \times \left(\frac{1 - abs(C_{31})}{1 - O} \right) \times \left(\frac{1 - abs(C_{32})}{1 - O} \right) = W_3$$

In which:
O=overlap

4.2 Weight factor of the decorrelation

The information content derived from the Jacobean matrix is not a constant, weighted factor. It can change significantly by simply changing the rainfall input in the model (see figure 6), as well as vary greatly in pure value. The variation per storm has to be countered by designing networks for various storms.

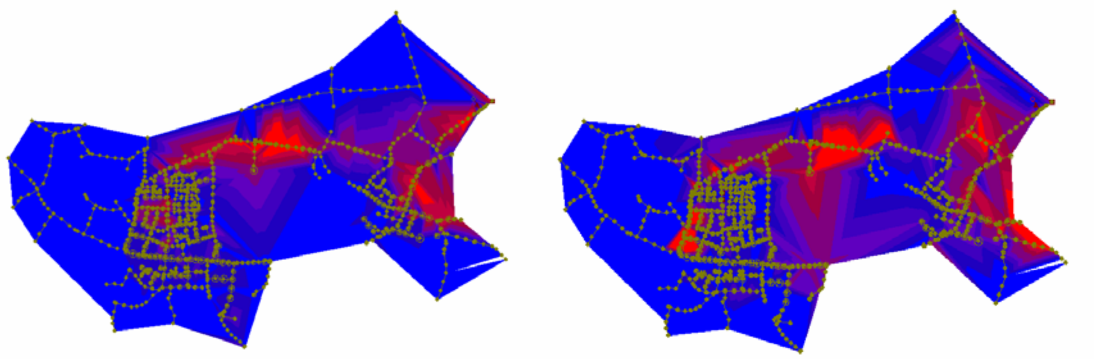

Figure 6: Total information content of the Beekbergen/Lieren manholes during two rainfall events. The rainfall event itself has a great impact on the designed network.

The variation in magnitude of the singular values provides an interesting difficulty. Decorrelation can mean for two locations with a very small correlation that the information content is multiplied by a factor 10^3 which may or may not be in the same order as the singular values of the parameters. Depending on where the decorrelation and the information calculation intertwine this may or may not lead to optimisation problems.

5. The exclusion of manholes

Initially in a monitoring network optimisation all manholes should be considered as potential measuring locations. This in order to get a maximum of information about the system before actually designing the network. After an optimisation using the whole network it will normally become clear that not all “optimal” locations are indeed fit for practical use. In the Loenen project a number of sensors functioned only during part of the storm, resulting in an insufficient amount of data from those sensors to include them in the calibration. The difference between the calculated information content and the actual content can be found in the installation height of the sensors. Even if this is only a few centimetres above the invert level the possible difference between calculated and actual information content can be very significant.

The easiest solution for this problem is the deliberate exclusion of manholes. If the number of time steps in which the calculated water level exceeds a reference is in excess of a given thres-hold (e.g. =20% of the total duration of the storm) the monitoring location can be used, else it should be excluded (see figure 7, red indicates exclusion). A disadvantage of this exclusion method is that locations normally discarded, like manholes in dangerous locations (traffic or otherwise) still have to be excluded manually before applying the algorithm.

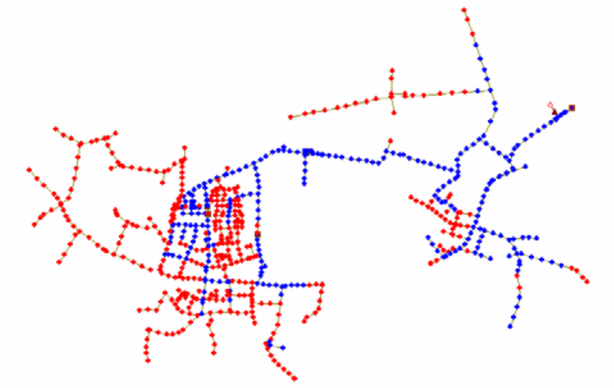

Figure 7: manholes included (blue) and excluded (red) from the optimisation calculated by setting a minimum water level during a minimum time during the rainfall event.

6. Result of the optimisation

The algorithm was used to optimise a network for the sewer system of Beekbergen. The result is figure 8 (next page). It is a completely different network from the network an expert would have designed. In fact in the project itself a different network has been chosen (see figure 9) as at that point in time the research had not cumulated in the proposed algorithm yet.

The discussion can be started whether or not the calculated network is superior to the network designed by an expert. The first apparent difference is that the optimised network has less than half the sensors the expert network has. As all parameters are normalised the total information content can be calculated for both networks. The optimised network gathers about 3 times the amount of information.
However, the optimised network is the ideal network for the design storm (a very short storm with high intensity). Other storm events might result in a completely different set of gauges. The possibility to combine multiple optimisations and the final comparison to an expert network is currently researched. Also the optimisation holds the assumption that the inflow in the system can be sufficiently simulated with the used hydrological model.

7. Using a genetic algorithm

While optimising, one could use randomly created sets of gauges. Although most random generators are influenced by computer speed and often prove to work unsatisfactory this will lead to an optimal result given enough time. To improve optimisation performance Clemens used a genetic algorithm for his measurement network optimisation (Clemens 2000). A year later Boomgaard used the same Jacobean method but now utilised an alternative optimisation technique, simulated annealing. Both algorithms arrived at nearly the same monitoring network and proved to work satisfactory. Both techniques can be used, however decreasing the number of gauges will directly influence the effectiveness of the genetic algorithm because of the decreased genetic pool that is available.

8. Conclusion

Monitoring network optimisation offers certain advantages compared to conventional measurement network design; it is fast (and therefor cheap) and with relatively “safe” results. However, a complete automation is not realised yet. Personal preferences in sensor information overlap and ‘experts judgement’ influences the final result as well as the choice of the design storms used in the optimisation. The improved algorithm is seen as a very powerful tool when designing monitoring networks. The algorithm can be applied for quality monitoring networks also, the only limiting factor is the availability of a good model.

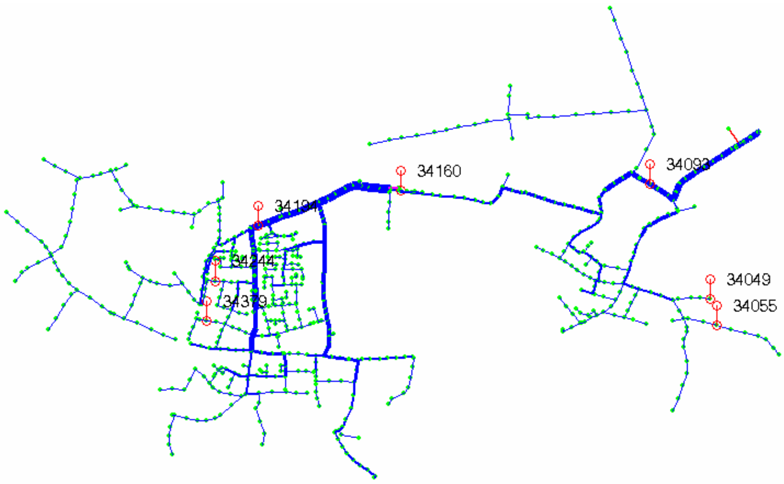

Figure 8: Optimised network

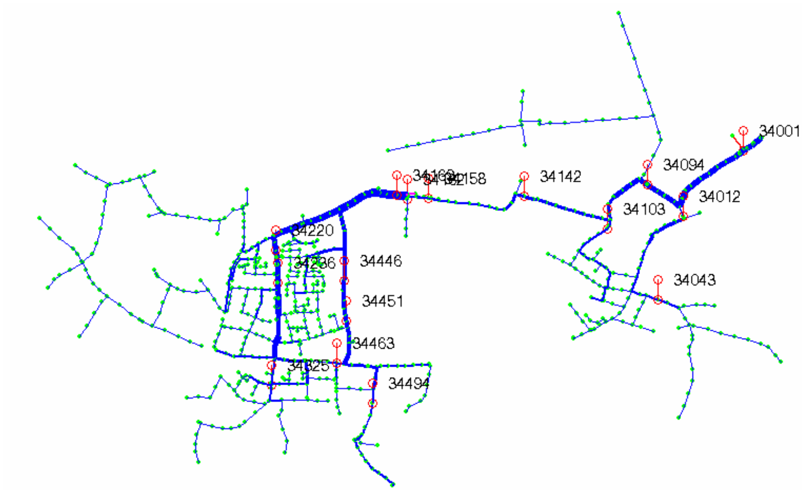

Figure 9: Expert judgement network

References:

Boomgaard, M. E., J.G. Langeveld and F.H.L.R. Clemens (2001*). Wastewater system Optimisation using Genetic Algorithms*. in: Proceedings of the UDM 2001, Orlando, Florida.

Clemens F.H.L.R. (2001[a]), *Hydrodynamic models in urban drainage: application and calibration,* PhD Thesis Delft University of Technology, Delft.

Clemens F.H.L.R. (2001[b]), *A design method for monitoring networks in urban drainage.* In: Proceedings of the UDM 2001, Orlando, Florida

Clemens F.H.L.R. (2002), *Evaluation of a method for the design of monitoring networks in urban drainage.* In: Proceedings of the Ninth International Conference on Urban Drainage, Portland, Oregon.

Henckens. G.J.R, J.G. Langeveld, P.P. van Berkum (2003), *Kalibratie van het hydrodynamische rioleringsmodel van Loenen (In Dutch),* Rioleringswetenschap en techniek, year 3, nr. 11, p. 46, Holapress Wetenschap.

Korving, H (2002). *Influence of model parameter uncertainties on decision making for sewer system management*. In: Proceedings of the Fifth International Conference on Hydroinformatics, Cardiff.

Witteveen+Bos consultants (2001), *Meetnetoptimalisatie Loenen (in Dutch)*, Deventer.