WaPUG Scottish Meeting 2005

Modern Techniques for Checking Data Quality

Martin Allitt
Richard Allitt Associates Ltd
The Old Sawmill
Copyhold Lane
Lindfield
Haywards Heath
West Sussex
RH16 1XT
Tel: (01444) 451552
martin.allitt@raaltd.co.uk

1. Introduction

Many of the standard specifications for surveys for wastewater modelling were written for the time when surveys were carried out across whole catchments; this was especially the case with manhole surveys. Nowadays, most parts of the UK have been surveyed with generally quite good surveys in most urban areas though with odd pockets of incomplete or missing data.

The requirements for surveys nowadays are predominantly for infill surveys. The specifications for quality control checks are in many cases not appropriate or practical for such infill surveys and as a consequence quality control checks are generally no longer undertaken in the field. This places a greater burden on the modeller to assess whether the quality of the data is adequate – this can be difficult from behind a desk or computer. This paper is intended to explore techniques available with the latest generations of software to carry out these assessments.

2. Why is it important to check quality

With the advances in modelling techniques and the power of the desktop computers that can now be purchased we are building and verifying ever bigger and more complex hydraulic models. These large, complex models rely heavily on the quality of the data that makes up the model. Also, many more activities involved in building hydraulic models are becoming automated with very large volumes of data simply being transferred from corporate databases. The old adage that the '*the quality of a model is only as good as the quality of the data used*' is as true today as it was a decade ago – we can just do things faster now.

3. Manhole Surveys

Accurate manhole data is one of the key items of data that is used in model building and as such it is very important that the data is of a high quality and that no rogue values are input into the model. The "Model Contract Document for Manhole Location Surveys and the Production of Record Maps"[2] states the quality control checks and the procedure that should be followed when supervising a manhole survey. This is really only applicable when carrying out large scale manhole surveys as the quality control requirement is to resurvey 5% of the surveyed manholes within a grid square or tile, and then accept all the manholes in that square if the 5% pass. Nowadays since we seem to carryout small scale or infill manhole surveys to collect missing data rather than blanket surveys of the catchment we are modelling we need to look at more cost effective and reliable methods of quality control checks. Also the demise of the Ordnance Survey benchmarks is making things more difficult.

Digital Terrain Modelling is becoming established more widely and these provide techniques which can be used to check the quality of cover levels from sewer records and manhole surveys. The simplest way to use this technique is to use the cover levels from you existing sewer records and produce a list of the cover levels along with their coordinates as a text file. This text file can then be imported into InfoWorks™ or

InfoNet™ and a TIN ground model can be created. This DTM can then be viewed and any low spots or high peaks will be apparent.

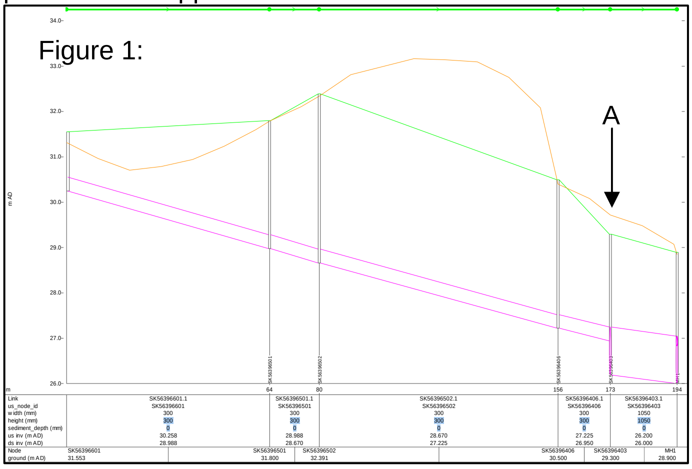


Figure 1:

Another method for checking cover levels is to purchase a DTM and import it into either InfoWorks™ or InfoNet™, produce long sections and see how the cover levels compare to the ground level of the DTM. The example in Figure 1 shows the continuous ground levels from the DTM (brown line) and the ground levels at the nodes (green line). A good correlation can be seen in this example. It might however be worthwhile having the cover level at node "A" checked or it might be worth checking if there is any history of flooding at this location which might support there being a low spot at this node.

It is important when using this approach to take into account the accuracy of the DTM being used. There have been several papers [3, 4] in recent years which have covered the accuracies of various DTM and their sources and it is worth reading these for a better understanding of Digital Terrain Models. It is also worth noting that a good DTM can be used to complete any missing cover levels, such as cover levels at junctions or dummy nodes which cannot be surveyed, in both InfoWorks™ and InfoNet™ there are inferencing routines that can be used to do this.

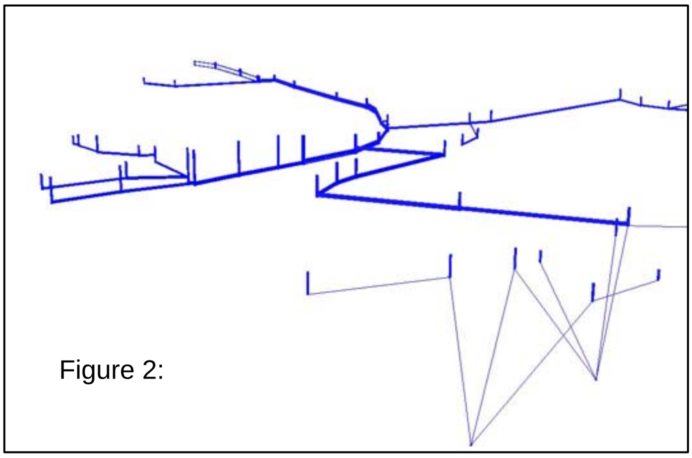


Figure 2:

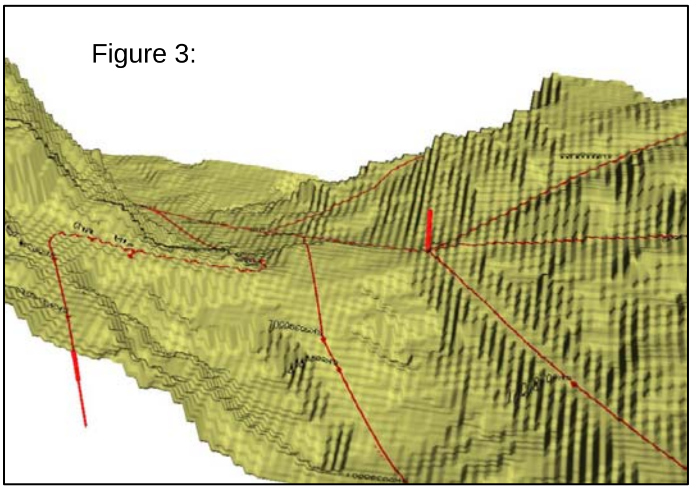


Figure 3:

An alternative way of checking the manhole survey data is to use the new feature in both InfoWorks™ and InfoNet™ which allows the underground network to be viewed in 3D. Figure 2 shows an example of a section of Sewer Records which was imported into InfoWorks™. The network was then displayed in 3D (a DTM is not needed to do this) and it can clearly be seen where the invert levels are wrong (in this case many were zero), by zooming in close enough it is also possible to see where pipe sizes are missing or incorrect. With a high quality DTM, this approach can be used to identify any manholes which stand above the ground level (see example in Figure 3) and also where any pipelines come above ground.

It is with experience of using high quality DTM's that users will become proficient in visualizing where there are errors in the data and the best way of learning is for users to experiment with the 3D capabilities in Infoworks™ and InfoNet™.

It is not just the surveyed data that needs to be checked but care should also be taken when inputting this data into the network models. With the launch of InfoNet™ in recent years there is now a tool which can be used to carry out detailed validation on the data input in real time. For example if all the sewers in a network are less than 1200mm dia then a validation rule can be set only to permit pipe sizes less than 1200mm to be input, this will stop 150mm dia sewers been accidentally input as 1500mm dia. InfoNet™ has very powerful, customisable validation checks and therefore is well suited to carrying out rigorous quality checks on all underground asset data. Another feature of InfoNet™ is that it holds all survey records so that several surveys of the same asset can be compared to see how closely they match.

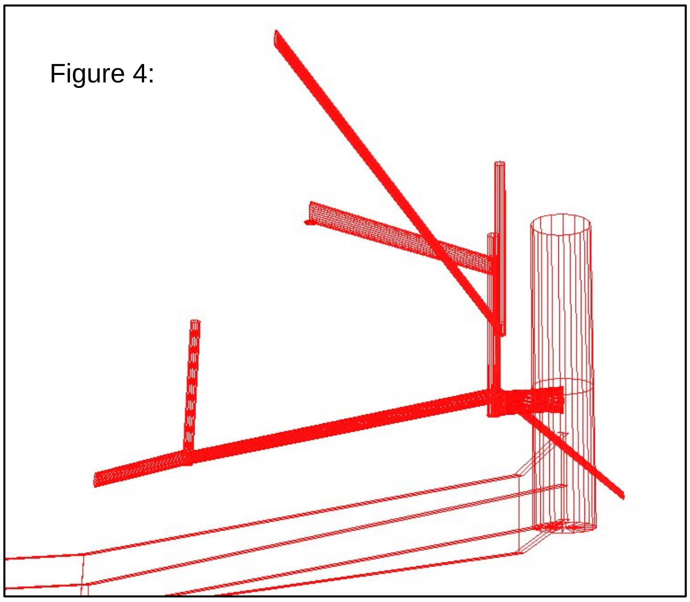
4. Ancillary Surveys

Ancillaries are key components in the sewerage network and they control how the network will operate and clearly incorrect data on ancillaries may significantly affect how the performance of the network is simulated. In much the same way as with manhole surveys the type of quality checking which should be carried out depends on the number of ancillaries being surveyed.

There are a number of checks which should be carried out on site during the survey such as repeating pump drawdown tests until 3 similar values are achieved. In addition there are a number of checks which the modeller can carry out back at the office:-

- Check that the pump rates are realistic by checking that the velocity in the rising main is somewhere near to $1m^2/s$; if it is significantly different from this there is a probability that either the rising main size is incorrect or that the pump rates are incorrect.
- Check the pump rate against the plated capacity on the pumps to see if they are similar. This will also give an idea of how well the pumps are operating.
- Check that the operating range of the pumps is sensible. If the range between switch on and switch off is too small then the situation could arise where the pumps would be operating too frequently and this could cause instabilities in the model.
- Visualize the ancillary in 3D but bear in mind that Infoworks™ and InfoNet™ both show all nodes as circular even though in reality they might be rectangular. Figure 4 shows a 3D view of a pumping station. Storage Nodes in Infoworks™ and InfoNet™ are shown in 3D with square plan shapes and consideration should be given to using 'storage nodes' at pumping stations, CSO's and other complex chambers in order to get a better visualization of them.

Figure 4:



5. Flow Surveys

Flow surveys are expensive and time consuming and it is important that the data collected from a flow survey is of the highest quality and that any sites giving poor quality data are recognised at the earliest opportunity so that monitors can be re-located to alternative sites. The current requirement for flow surveys is to follow the "Model Contract Document for Short Term Sewer Flow Surveys"[1] which states that in-sewer checks should be carried out as part of the data retrieval visits and at the end of the survey to check that the monitors are recording correctly. It is becoming increasingly important that the Modeller is taking a greater role in understanding and checking the quality of the data during the currency of the survey. This involves carrying out more external quality checks on the data. The quality of the flow survey can be broken down in to three sections, the quality of the installation site, the quality of the data at the site and the location and spacing of raingauges.

5.1. Installation

To stand a chance of collecting good quality flow data, it is important that flow monitors are installed at good sites. It is essential that a pre-installation survey is carried out to confirm that the sites chosen in the office will allow good quality data to be collected.

Before the pre-installation survey, time should be spent in the office to understand the catchment and to identify the key areas of the network where flow monitors need to be installed. It is important that the sites chosen in the office are ones that look as if they are suitable, for example they should meet the following criteria:

- The site should be accessible.
- No undue Health & Safety risks (eg traffic)
- Straight through, no sharp bends at manhole.
- No change in pipe size from incoming to outgoing.
- Similar gradients upstream and downstream.
- Similar invert levels for incoming and outgoing.
- Adequate depths of flow and where the depths and velocities are within the measurement envelope of the flow monitor equipment.
- No other significant incoming pipes or rising mains which could cause turbulence.

It is important to have a preferred site for each monitor and ideally at least 2 alternative sites which are also suitable on paper before going into the field and carrying out the pre-installation survey. Then if for some reason the preferred site is not usable or is unsatisfactory, alternative sites have already been identified.

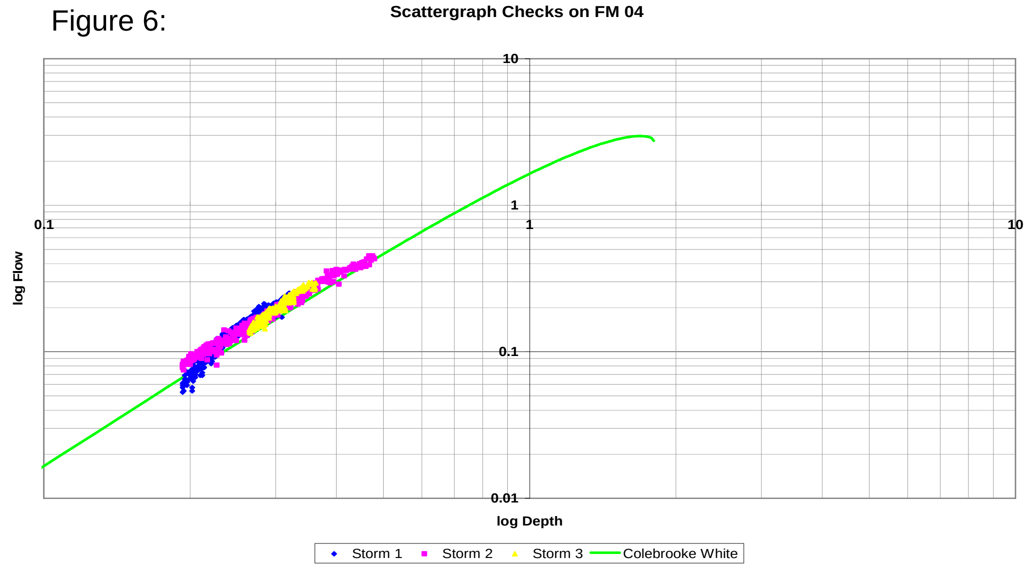
Figure 5:

Figure 5 shows some examples of good and poor monitoring sites.

5.2. Flow Monitor Data

Once the flow monitors have been installed it is advisable to monitor the data from each monitor on a weekly basis to check that good quality data is being recorded. Scattergraphs are a good way of checking that the data being recorded is of good quality and that there is nothing downstream or upstream in the network which is having an adverse effect on the flow regime at the monitoring site. It is worth plotting scattergraphs each week if the flow survey contractor can send you the flow and depth data each week. Scattergraphs can easily be plotted using standard spreadsheet programs but it is worthwhile plotting the data with a log-log scale. You can see from the example in Figure 6 that it is worth plotting the measured data along with the theoretical flow – depth curve (calculated from the Colebrook – White formula). Ideally the data recorded by the installed flow monitor should follow the theoretical line (as in Figure 6).

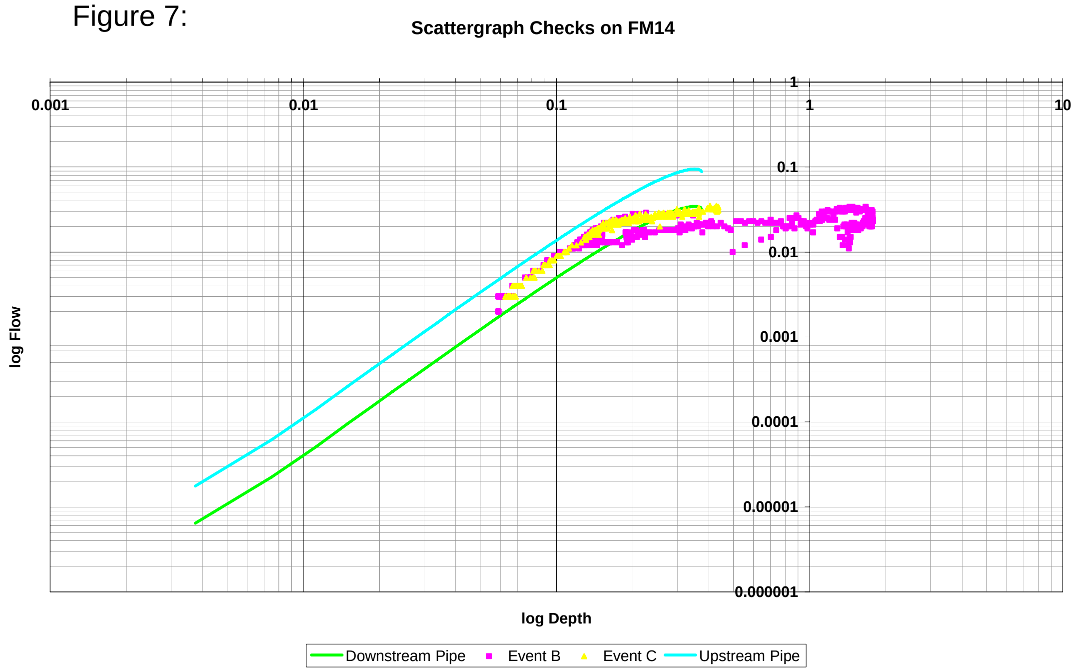
Figure 6:



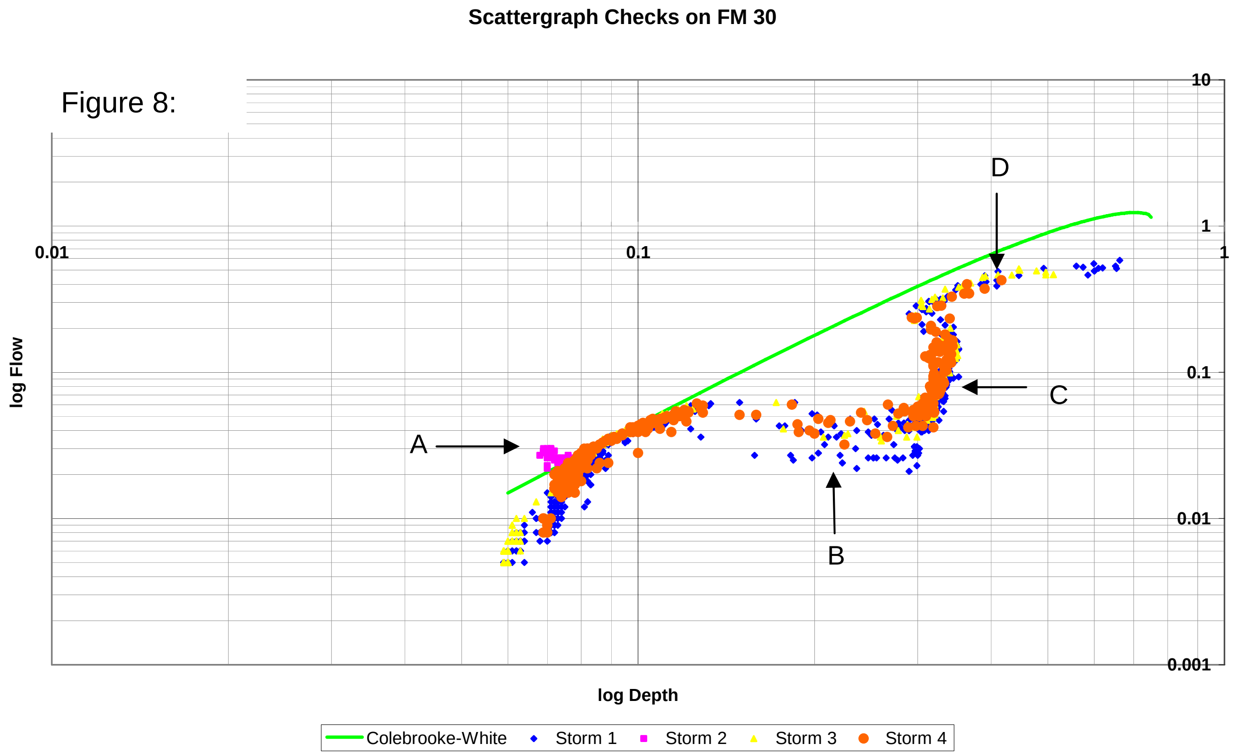
If the downstream pipeline and the upstream pipeline are the same size but have slightly different gradients, the capacities would be slightly different. It is then worthwhile plotting the theoretical line for both pipes and ideally the data should lie between both theoretical lines.

Scattergraphs are very useful tools as they can show lots of information. The scattergraph in Figure 7 shows the data recorded at one monitor site for two verification events. The scattergraph shows that as the depth increases initially there is a linear increase in the flow in the pipeline and that the recorded data fits between the theoretical lines for the upstream pipeline and the downstream pipeline. As the depth increases further the scattergraph shows that the flow no longer increases linearly and that there is virtually no increase in flow. This was caused by a restriction in the pipeline downstream; in this case there was a partial blockage found in the pipeline downstream causing about 80% of the pipe cross sectional area to be lost. After this was identified from the scattergraph a CCTV survey was done and then a jetting exercise to clear the blockage. This was done whilst the flow survey was underway and because of the early intervention subsequent rainfall events (another two were recorded) were with this sewer unrestricted.

Figure 7:



Another example of a scattergraph can be seen in Figure 8. This flow monitor was located just upstream of a CSO and shows 4 stages of flow which can be identified from this scattergraph. Firstly (A) there is a linear increase in depth as the flow increases, and this matches well to the theoretical Colebrooke-White relationship. In this part of the scattergraph up to a depth of about 150mm there are free flow conditions. The next part of the scattergraph shows that the downstream network starts to surcharge and cause backing up to the monitor site; at point (B) it can be seen that the flow has hardly changed but the depth has increased substantially. Once the surcharge level reaches the level of the weir at the CSO at a depth of about 340mm there is a rapid increase in flow for very little increase in depth, this can be seen in part C of the scattergraph. Finally once the overflow has operated it is clear that the capacity of the overflow pipe (rather than the weir)



starts to restrict the flow and at point D the data follows the line for the overflow pipe rather than the main pipe.

Figure 8 shows clearly that just because the data does not follow exactly the line of the theoretical depth-flow curve the data is not automatically of poor quality. The important thing is to understand what the scattergraphs are showing and to act accordingly. In the case of Figure 7 the appropriate action was a CCTV and jetting exercise. In the case of Figure 8 the appropriate action was to ensure that the CSO was modelled correctly and also that the overflow pipe as well as the weir was modelled.

Scattergraphs are very useful tools in understanding the flow in the sewers and with experience, much information can be gained about the flow regimes in the sewerage network up and downstream of the monitor location.

5.3. Rainfall Data

Rainfall data is the hardest item of data on which to carry out quality checks. The “Model Contract Document for Short Term Sewer Flow Surveys”[1] states that the contractor should make sure that each raingauge is operating correctly at each visit and that the data logger records tips in both directions. These checks only check that the raingauge itself is recording correctly. The WaPUG Code of Practice sets out acceptable criteria in terms of rainfall depths, rainfall intensities and variation between adjacent raingauges.

It is getting harder these days to find suitable locations at which raingauges can be installed and as a result it is becoming more difficult to achieve an adequate coverage across a catchment. Therefore it is important that the quality of data that is recorded is of a high standard.

There are only a limited number of checks which a modeller can undertake to check the quality and consistency of rainfall data. One very useful method is to plot graphs of cumulative rainfall both weekly and ongoing for the whole survey period. This should be done throughout the flow survey period so that an early indication can be gained if one raingauge is recording less than the others. Figure 9 shows an example of 3 raingauges but Raingauge No 2 was installed one week earlier than the other two and also Raingauge No 1 was inoperative for a period of nearly one week due to the filter becoming blocked.

Figure 9:

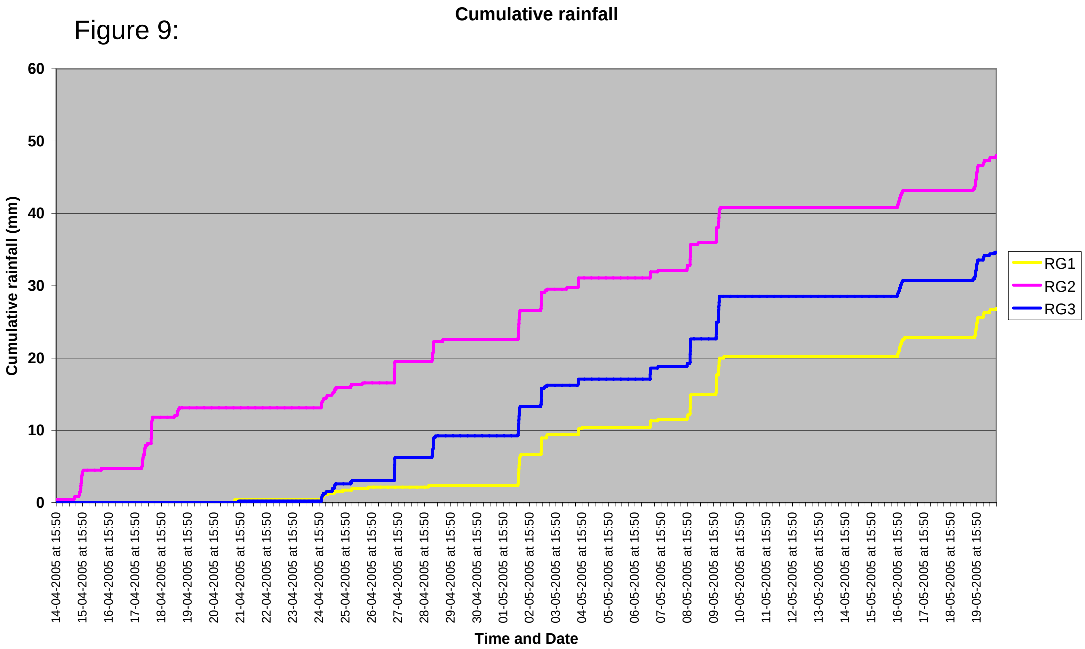


Figure 10:

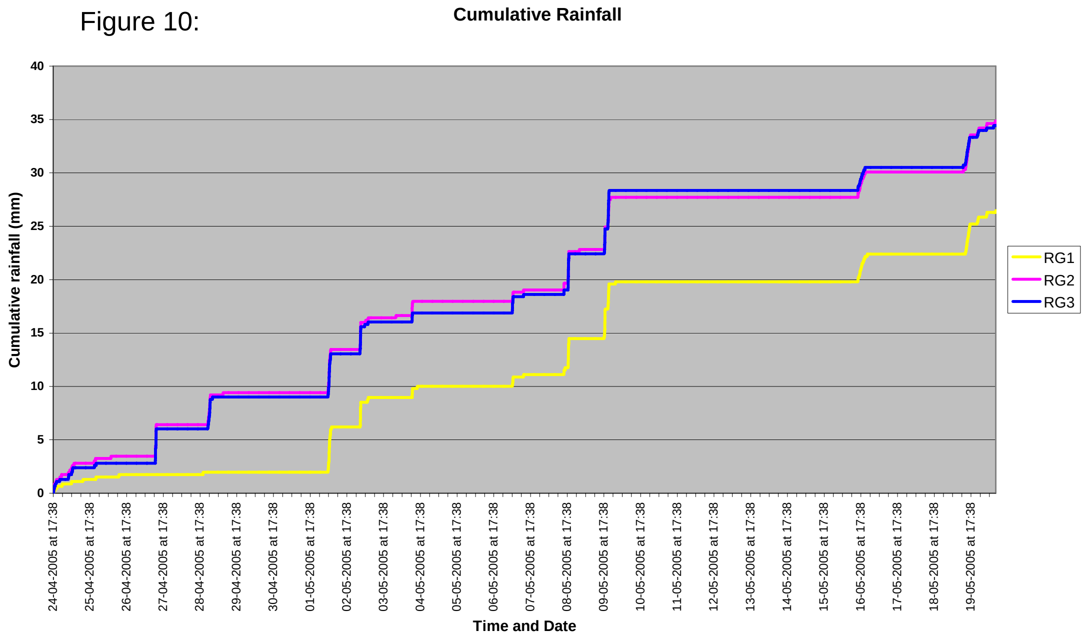


Figure 10 shows the same information but with the first week for Raingauge No 2 removed – it can now be seen how close the measurement is at RG2 & RG3. This technique can become cumbersome with very large surveys and in those situations it is probably best to plot a series of graphs with only adjacent raingauges shown on each graph.

There is scope for better methods of quality checking of rainfall data and maybe there is a role for weather radar in this.

6. Impermeable Area Surveys

Impermeable Areas are another important aspect of model building and it is important that the data is adequately checked. It could perhaps be argued that the verification process is in itself a checking process on the impermeable areas input into the model. However, there can still be significant errors which cancel each other out in verification but do affect how extreme storms are simulated.

The recommended way of checking the quality of the IAS data has always been for the Consultant to go out into the field with the surveyor and resurvey 5% of the study area and check that the results from the first survey match the second survey of those areas. This method is still the best approach for Impermeable Area Studies which are carried out on large areas and whole catchments. When the surveys are more localised infills and where the study area is small it can be meaningless to carry out a 5% check. In these cases it maybe better to consider resurveying a larger area and carrying out meaningful 5% checks. An alternative approach is to ascertain whether other checking methods could be applied to resurveys of selected areas.

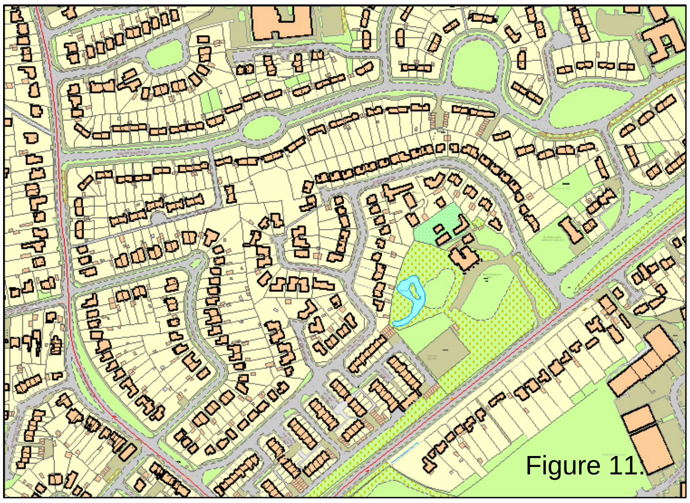


Figure 11.

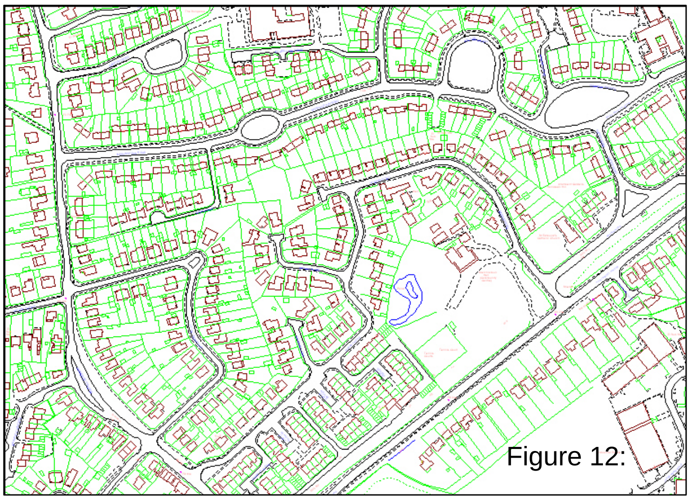


Figure 12:

The possible alternative methods for checking IAS data alongside infield checks can be:

- Mastermap®.
- Aerial Photographs (especially Infrared).
- Shape recognition of housing styles.

The Mastermap® product available from Ordnance Survey is a very useful product and can be used in the office to carry out checks on the IAS data received from the field. An example of Mastermap® can be seen in Figure 11. Mastermap® can be used to check that all the impermeable area in the catchment has been found and allocated to one or other of the drainage system types. The advantage of Mastermap® over Land Line data for these checks is that the topography layer of Mastermap® shows the difference types as closed colour coded polygons. It is clear to see that Mastermap® can gives us a much better understanding of the surface types in the catchment than looking at Landline mapping (as Figure 12); this difference can be seen by comparing Figure 11 and Figure 12. By using Mastermap® it is possible to check that all impermeable areas are accounted for.

Whilst correct allocation of impermeable areas is the principal task in IAS surveys it is becoming more important that the different types of permeable surfaces are identified and where necessary correctly allocated in the model. Mastermap® is a valuable tool in this work but also the use of aerial photography can be of help. High resolution infrared photography can be used as an alternative method for checking the quality of IAS data. Infrared photography clearly shows the areas of vegetation by picking up the chlorophyll in the plants. From this it is clear to see the vegetated permeable areas. The disadvantage of this method is that it is expensive, but on larger catchments it might be worth the money. Figure 13 shows an example of a high resolution aerial photograph of an urban area; clearly showing the large areas of permeable and impermeable surfaces which make up the catchment surface. Figure 14 shows an infrared aerial photograph of the same area as in Figure 13; the areas of vegetation show up clearly as the chlorophyll is detected and in this example is shown in red. It is noticeably easier to distinguish the impermeable and permeable areas in Figure 14 than it is Figure 13.


Figure 13

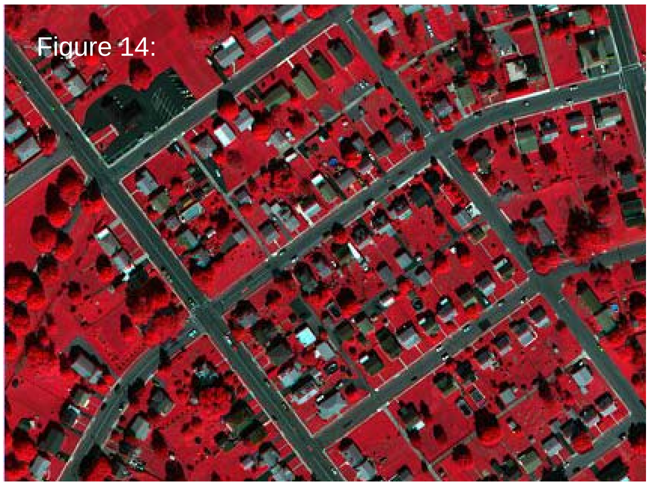


Figure 14:

Another method for checking that Impermeable Areas have been correctly assigned to the different drainage systems might be to use shape recognition. Almost throughout the UK there are certain easily recognised house types which lend themselves to shape recognition. A couple of examples are shown below.

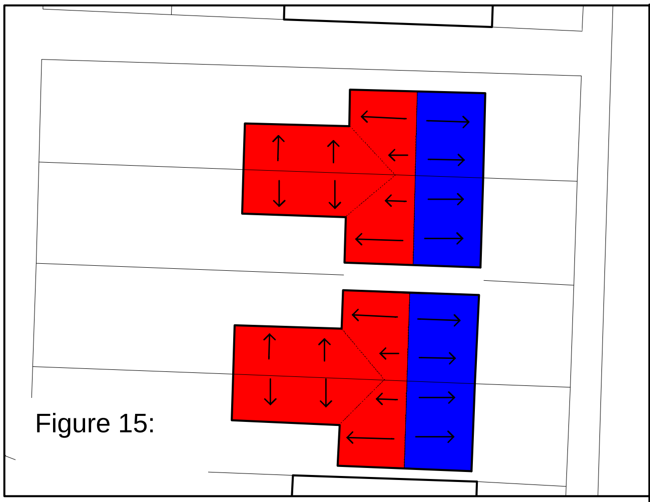


Figure 15:


Figure 16:

The first example shown in Figures 15 & 16 is typical pre-1919 semi-detached houses which have a gable wall at each end, a ridge parallel to the road and a single rainwater pipe at the front which discharges into the highway, frequently with a channel grating across the footway. The front part of the house drains to whatever drainage system serves the road (frequently small SW sewers) whilst the rear roof areas drain to the combined sewers.

The second example shown overleaf in Figures 17 & 18 are typical pre-1919 terraced houses.

Figure 17

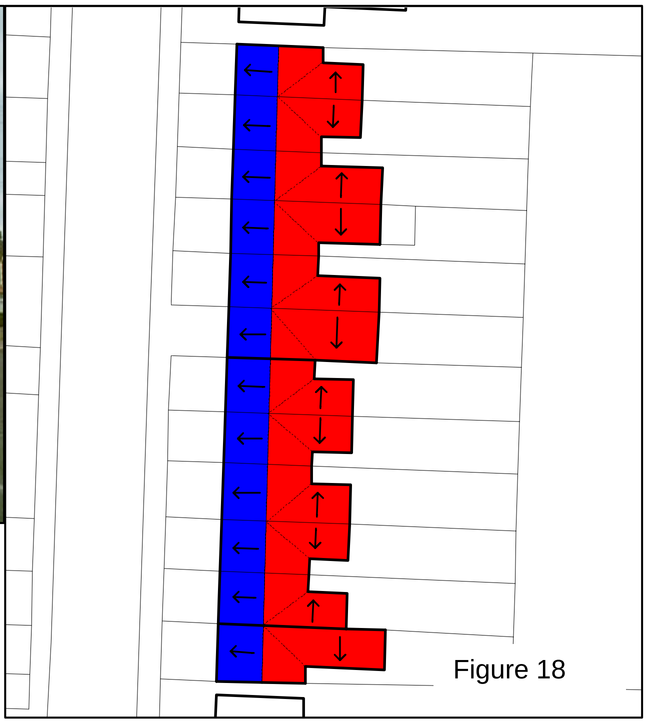


Figure 18

With this house type there can be long rows of terraced houses with typically one rainwater pipe at the front for every two houses and these typically discharge onto the highway and therefore the drainage system type is whatever drains the road. The rear roofs which are inevitably larger drain to the combined system.

These two typical house types are easily identified but there will also be other standard types, perhaps on a regional basis which can be recognised. If an IAS survey is presented which shows houses of these standard house types but with different drainage arrangements it maybe worthwhile repeating the survey on those properties. It is a relatively easy matter for modellers or organisations to build up a library of standard house types in their area which are drained in standard ways.

7. Data Flags

The final aspect in my paper concerns the way in which we demonstrate and record the quality of the various items of data which have been used to develop the model.

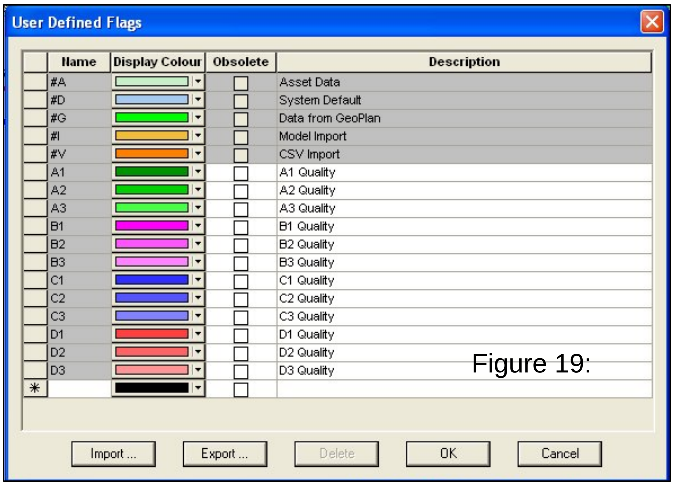


Figure 19:

InfoWorks™ and InfoNet™ uses a system of data flags which the majority of modellers use to record who has input the data or where the data came from (eg initials of modeller or the name of survey contractor). In time as models are used by different organisations or different Engineers these data flags not only become meaningless but become unmanageable.

At Richard Allitt Associates Ltd we have changed to using a system based solely on data quality and method of collection rather than who has input the data. Figure 19 shows the user defined flags used by RAA in both InfoWorks™ and InfoNet™. We have followed the WaPUG Code of Practice for the Hydraulic Modelling of Sewer Systems Version 3.001[5] for data collection levels (Table 3.1). This system has 2 parts to it, the first part is a letter (A, B, C or D) which relates to the method of data collection laid out in the WaPUG COP[5]. For example the letter A would be used to flag data which had come from a survey whereas the letter D would be used to flag data which had come from existing records. The second part of the data flagging method adopted by RAA is a number between 1 and 3, this stands for the quality of that data. This is especially useful in InfoNet where you hold multiple survey records for each asset, and where you might decide that at manhole "X" the quality of the most recent survey is not as good as previous surveys at this location and you wish not to use the data from this survey in your network model. In this system 1 is used for data of high quality and 3 is used for data of low quality.

This is a very simple system of data flags which we would like to see adopted across the industry. We have created a csv file with these data flags and modellers are welcome to download this from our web site at www.raaltd.co.uk

8. References

1 Model Contract Document for Short Term Sewer Flow Surveys.

2 Model Contract Document for Manhole Location Surveys and the Production of Record Maps

3 Hale J, "*Urban Flood Routing…..The Next Step"*, (2003)

4 Allitt M, "Using DTM's to determine locations of permeable runoff into Sewerage Networks", (2004)

5 The WaPUG Code of Practice for the Hydraulic Modelling of Sewer Systems Version 3.001, (2002)