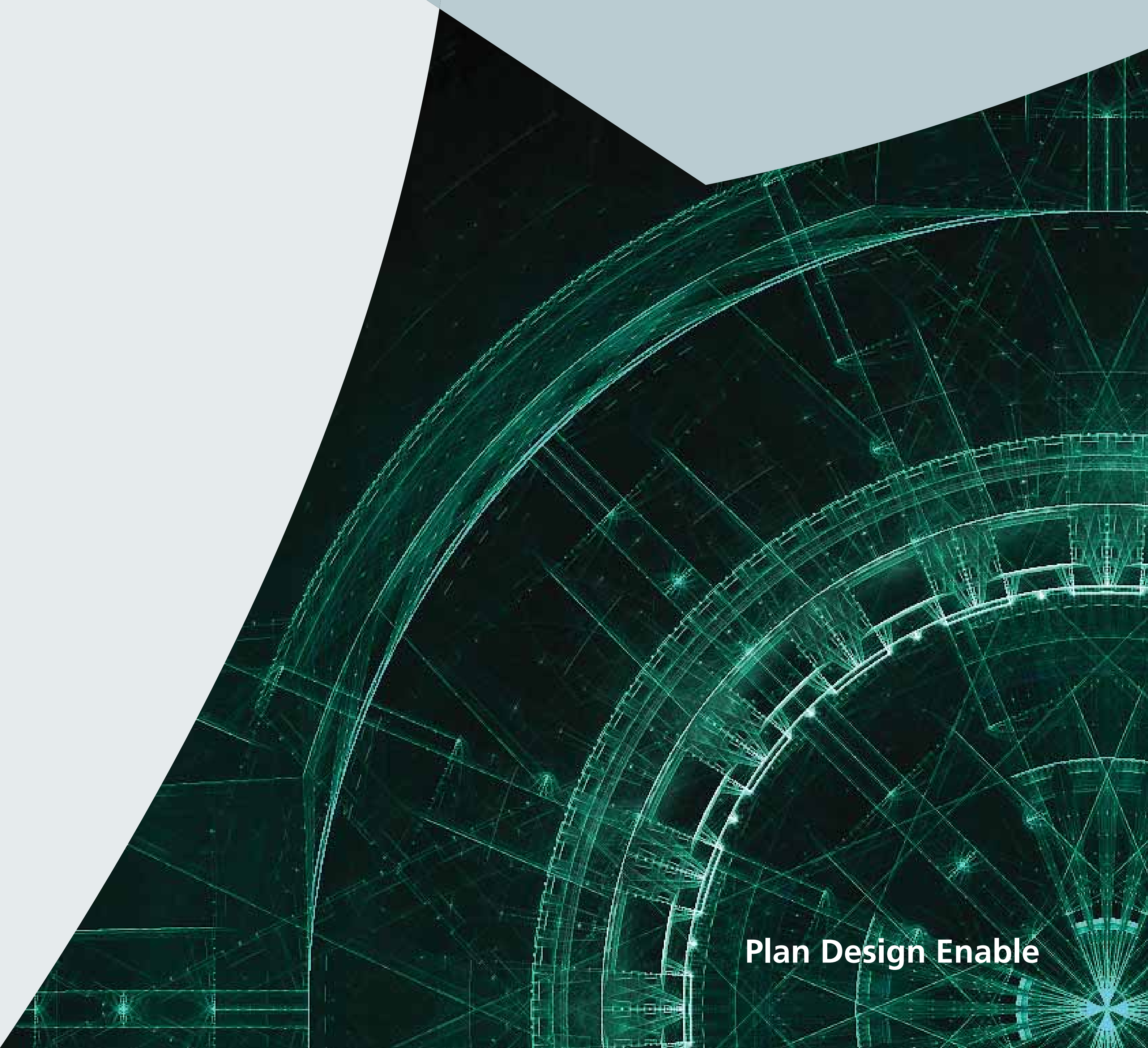
03

ATKINS

Technical Journal

Papers 034 - 050



Plan Design Enable

FOREWORD

Welcome to the third edition of the Atkins Technical Journal.

Since publication of the second Technical Journal, interest in contributing has been high across the Atkins Group and we have received considerable praise for the exceptionally high standard of papers from clients and staff alike. This is testament to the excellent technical expertise of staff that Atkins' continues to value and develop. The third edition covers an even broader spectrum of papers from the technical disciplines across Atkins.

In the current global economic climate, it is essential that technical excellence remains at the core of Atkins' service offering. The contents of this third Technical Journal demonstrates that we are continuing to achieve this. Increasingly, the papers presented mirror the agendas of the Technical Networks that Atkins has successfully established in recent months and we look forward to continuing promotion of the successes of these disciplines both internally, to capture the valuable lessons we have learned, and externally to our valued clients.

I hope you enjoy the selection of Technical papers included in this edition.

Chris Hendy
Chair of Technology Board

Transport Solutions
Highways & Transportation

Technical Journal 3

Papers 034 - 050

CONTENTS

The maintenance of the main expansion joints on the Forth Road Bridge

David Timby

Principal Bridge Engineer
Highways & Transportation
Atkins

Barry Colford

Chief Engineer & Bridge Master

Forth Estuary Transport Authority

Katy Brown

Senior Engineer

Highways & Transportation
Atkins

Abstract

This paper discusses the life, maintenance, proposed fail safe measures and eventual replacement of the roller shutter joints of the Forth Road Bridge.

The bridge spans the Firth of Forth linking the towns of North and South Queensferry, approximately 9 miles (15km) to the west of Edinburgh (see Figure 1).

Construction of the bridge was started in 1958 and work was completed with the official opening in 1964. The bridge forms a vital link in Scotland's strategic road network saving a travelling distance for road traffic of about 28 miles. Traffic usage in the first year of opening was over 4 million vehicles and this has grown steadily to over 24 million in 2006. The bridge is operated and maintained by the Forth Estuary Transport Authority (FETA). In 2001 the bridge gained 'Category A' listed structure status reflecting its national importance.

Bridge construction description

The main structure (shown in Figure 2) is a 3-span suspension bridge with a central span of 1006 metres (3300 feet) and side spans of 408 metres (1338 feet). On both approaches to the bridge there are multi-span viaducts.

The road over the bridge comprises a pair of two lane carriageways, both 7.3 metres (24 feet) wide. The carriageways are flanked by combined footway and cycle paths 4.57 metres (15 feet) wide. The overall width of the structure is 33 metres (108 feet).

The deck to the main bridge comprises a series of steel trusses spaced at 9.144 metre (30 feet) centres, spanning transversely and hung by the top chords of the trusses from vertical hangers of the suspension cables. Between the steel trusses are two longitudinal stiffening trusses which run the length of the suspended deck. Each carriageway and footway/cycleway is a separate discrete construction above the trusses. For the main span, the deck comprises longitudinal steel stringer beams and 12.7mm (0.5 inch) thick steel deck plates. The side spans also have longitudinal steel girders but with a reinforced concrete deck slab.

Figure 1 - Location plan

The main cables are 600mm (2 feet) nominal diameter and are each made up of 11618 galvanized wires, each 4.98mm (0.2 inches) in diameter, that transfer the loads from the deck to the main and side towers and also down to the north and south anchorages. The trusses are linked to the main cables by 192 sets of steel wire rope hangers.

The main towers are of steel cellular construction rising some 156 metres (512 feet) above river level and are formed from fabricated steel boxes joined by cover plates.

The legs of each tower are connected by cross members at the top and just below deck level and by diagonal stiffened box bracing above and below the deck.

The decks on the approach viaducts comprise a pair of longitudinal steel box girders supporting a series of transversely spanning steel girders. The transverse girders cantilever out from the box girders to support the parapets and verge construction. Over the transverse beams is a reinforced concrete deck slab.

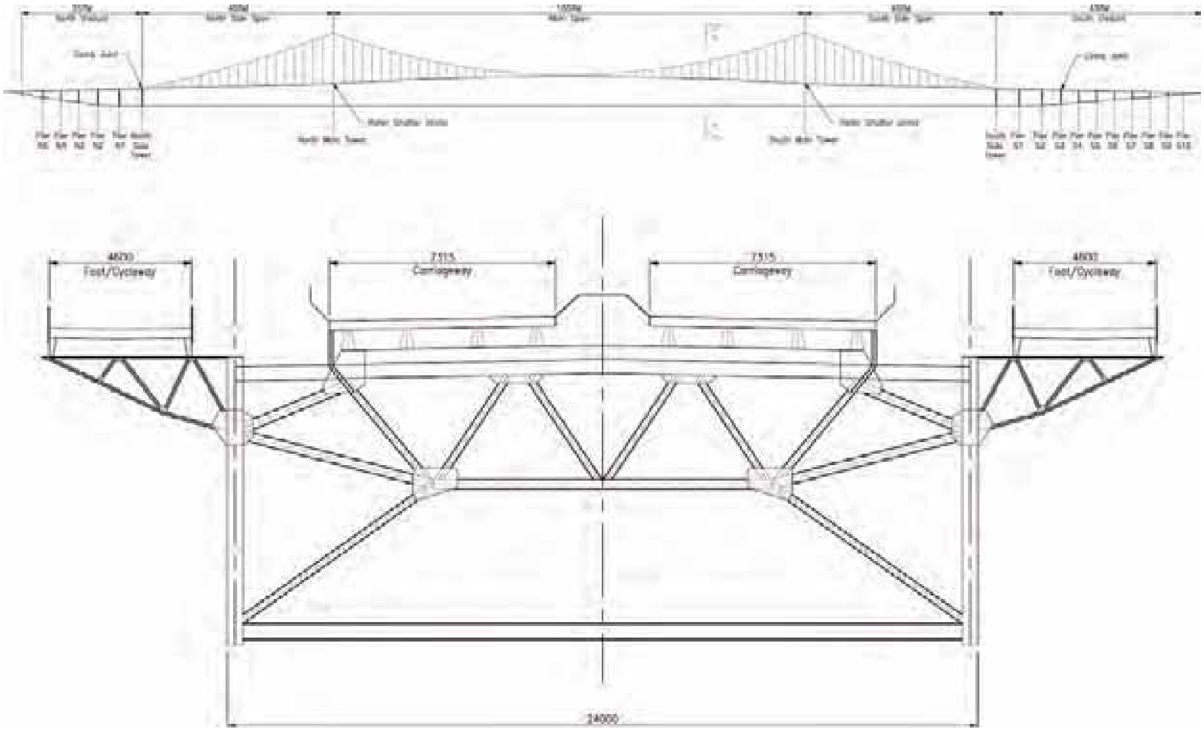


Figure 2 - Elevation and typical cross section of main deck

Bridge deck expansion joints

The deck expansion joints allow for the movement of the bridge structure. Movements on a suspension bridge, such as the Forth Road Bridge, occur as a result of expansion and contraction of the deck and cables from temperature effects and in response to vehicle and wind loading. Such movement occurs in a combination of vertical and horizontal directions as well as in rotation about these axes.

In each carriageway, at both main suspension towers, there are two joints to accommodate movement in the main centre span of the bridge and each of the side spans. These joints are of the roller shutter type. There are further joints in both carriageways on the approach viaducts comprising of interlocking combs (or fingers). In addition, at each carriageway joint, there are also corresponding joints in both footway/cycleways. These comprise steel plates that slide over each other.

Figure 3 - Aerial image looking south

Roller shutter joints

The roller shutter joints are original to the construction of the bridge. This type of joint is able to accommodate the large movements that occur in a bridge of this type. Information from the 'as built' drawings indicate that the joints were designed to have a movement capacity of +810mm/-920mm (total 1730mm, 68 inches) in the main span and +150/-260mm (total 410mm, 16 inches) in the side span.

The deck expansion joints comprise a series of six units per carriageway. Each unit has effectively two movement joints, one for each (side and main) span. As shown in Figure 5, each individual joint comprises a shuttle (or anchor or bridge) plate which is articulated on one (deck) side of the joint. This plate effectively spans over the structural gap of the joint itself. Attached to the opposite edge of the shuttle plate are a series of link plates all connected together by hinge mechanisms to form a train. The shuttle and link plate trains slide, via discrete feet, over the curved top flanges of support beams as the structural gap moves. The other side of the joint is a tongue plate. This tongue plate is supported on a cross beam and also rests on the link plate train to form a smooth running surface for traffic.

Figure 4 - Roller shutter joint installation (early 1960s)

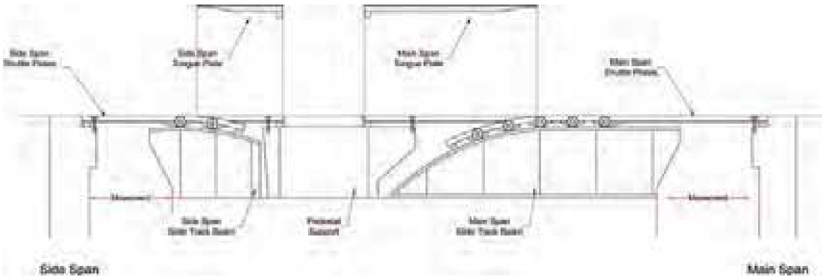


Figure 5 - Long section of the joint

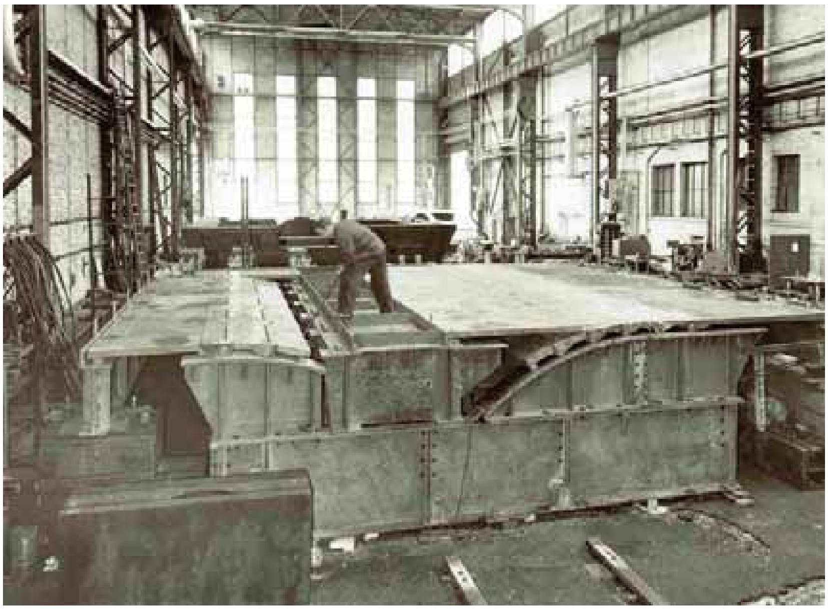


Figure 6 - Trial erection of roller shutter joint

Maintenance history

Roller shutter joints of this type have a typical design life, before major maintenance, of 20 to 30 years. The joints in the Forth Road Bridge are now over 45 years old but they are subject to regular inspection and maintenance by FETA's maintenance team. In 1975 an inspection and overhaul was undertaken which reported that the joints were generally performing well but that evidence of wear was becoming apparent. FETA have also undertaken further inspection and maintenance work. This has included:

- Re-profiling of the top flanges of the plate train track beams and plate train feet to try and rectify the effects of uneven wear
- Welding 'keeper plates' on the sides of the hinges to prevent the hinge pins from migrating outwards
- Replacing shuttle and tongue plate holding down pins and springs
- Lifting sample plate train and tongue plates to undertake a detailed inspection of the condition of the sliding surfaces and wear to hinge pins.

The recent sample inspections by FETA have shown excessive wear in the plate train track beams and in the hinges and feet of the link plates. The effect of this wear is that the joint is now becoming increasingly noisy and play in the mechanisms and wear surfaces will cause increasing damage to the joints. Ultimately the joints will break up with safety implications for the travelling public and/or parts of a joint will seize causing excessive loading on the bridge structure. In addition to the wear in the joint itself the traffic surfaces of the steel plates have also become excessively worn. This has resulted in the wear of the leading edge of the tongue plate. In addition, the anti-skid pattern in the traffic face of the main link plates has worn away. See Figure 7 to Figure 11.

Despite the increasing wear the joints are generally continuing to perform well and reliably in service but, with the overall wear, are now becoming an increasing maintenance liability.

Images of wear:

Figure 7 - Wear in top flange of track beams

Figure 8 - Wear in hinge pin hole and misalignment of the pin

Figure 9 - Misalignment between adjacent shuttle plates

Figure 10 - Misalignment of tongue plates

Figure 11 - Opening and closing of gaps in the plate train (viewed from below)

Feasibility study and design work

In response, FETA invited tenders to undertake a feasibility study and then provide the necessary design work for the repair or refurbishment of the deck joints. In April 2007, following a quality/cost tender process, Atkins was appointed to undertake the work.

Atkins' brief for the feasibility study was to consider options for the repair, refurbishment or replacement of not only the roller shutter joints but also the approach viaduct and footway/cycleway joints. In developing the options many factors and constraints that would affect the works including health and safety, traffic management, programme, access, resources, procurement, future maintenance and budget were considered. Some of the key points are noted below.

Traffic management

A high priority of FETA is to minimise delays to users of the bridge whilst still protecting the health and safety of the users and the FETA maintenance team. Traffic flows over the bridge exceed the theoretical capacity of the carriageway almost daily and therefore any restrictions would inevitably cause significant traffic problems on the road network in the region. Conversely, any works on the other local strategic routes need to be co-ordinated with the Forth Road Bridge works. There are also various external factors which affect the volume of traffic flow over the bridge, such as the improved weather in summer months and large public events like the Edinburgh Festival and the Open Golf Championship.

Programme

To minimise the risk of disruption to the work, and to reduce the potential safety impact on the workforce, it would be preferable to undertake the work in the spring/summer months when the weather is more likely to be suitable. The time period to undertake the work therefore needs to be planned in advance. Allowance would need to be made for unexpected/unforeseen difficulties that may cause time over runs. In addition, some of the work operations, such as painting or lifting using cranes are particularly weather sensitive.

Access

Access to the roller shutter joints under the bridge deck is via walkways from the footway/cycleway. The existing access walkways do not allow all parts of the joints to be easily seen and when undertaking a detailed inspection FETA has to supplement the existing walkways with temporary scaffolding. Access from above the bridge deck would require traffic management with the attendant issues discussed above.

Resources

Repair, refurbishment or replacement of the joints would require the use of specialist materials and labour, as well as lifting equipment. Such resources may not be immediately available and would need to be planned for in advance. The works would also be expensive and finance may need to be spread over more than one financial year.

Procurement

The number of contractors who would be willing and able to undertake the work would be limited. The bridge is in a harsh and relatively high risk environment and the type of work is specialist in nature. To ensure the contractor chosen to do the work can make available the right resources a reasonable mobilisation period would be required.

Consideration would need to be given to the type of contract and clauses within. Work is likely to be disrupted by poor weather conditions and disruptions on the road network may prevent the installation of traffic management. FETA would need to decide who would carry the risk of such disruption.

Options for joints

For the roller shutter joints five options were considered. These were:

(i) Continuing with the current maintenance regime (i.e. effectively a 'do minimum' option)

(ii) Full refurbishment of the joints

(iii) Full replacement of the joints with a similar type

(iv) Full replacement of the joints with a supported finger comb type joint

(v) Full replacement of the joints with an elastomeric type with incorporated steel girders.

For the approach viaduct comb joints four options were considered these being:

(i) Continuing with the current maintenance regime

(ii) Completely refurbish the existing joints

(iii) Completely replace the existing joint with a new comb joint

(iv) Completely replace the existing joint with a multi-element type.

Finally, for the foot/cycleway joints, just two options were considered these being:

(i) Continuing with the current maintenance regime

(ii) Completely refurbish the existing joints.

Recommendations of feasibility report

Atkins' feasibility report made recommendations for each of the joint types. For the roller shutter joints it was recommended that the joints were replaced with new roller shutter joints of a similar design. However, the opportunity was identified to use improved steels and manufacturing techniques that were not available at the time of the original construction. This should enable the joints to be manufactured to tight tolerances and allow a more robust link plate hinge mechanism. Other advantages of reusing this type of joint include using a tried and tested system that involves very little modification to the surrounding supporting steelwork. This option would also likely require the least time for traffic management measures on the carriageways. Full replacement was also favoured by FETA on grounds of safety as this reduced the amount of work on site.

To enable future inspections and maintenance to be undertaken more efficiently the report also recommended that access walkways were improved as part of the works.

For the approach viaduct comb joints it was also recommended that these joints were fully replaced with a joint of a similar type. Refurbishment would have resulted in most of the joint being replaced anyway without the advantage of using improved materials and designs. For the sliding joints in the footway/cycleway it was recommended that these were fully refurbished.

Actions following the feasibility report

Design work

The next steps were to further develop each of the recommended options into detailed designs, specifications and contracts. In addition, the works would require advance planning, programming and publicity. Although the same type of joint was being proposed a full design review and Category II check of each component was undertaken to ensure structural adequacy particularly as vehicle loadings had significantly increased since the original construction.

Planning work

As mentioned previously, it was considered preferable to undertake the work in the summer months when the weather was more likely to be suitable and daylight hours were longest. The time periods to undertake each section of work needed to be planned in advance. Time allowance also had to be made for any unexpected or unforeseen difficulties that could cause time over runs.

It was recognised that carriageway closures would be essential but the preferred time to undertake the site works coincided with the time of year for the highest traffic flows. The works therefore had the potential to be the most disruptive to the road network that Scotland had ever seen.

As with the original construction much of the fabrication work could be done off site and this would have the benefit of ensuring that each whole joint unit could be set up accurately.

It was estimated that closure of the first carriageway would be necessary for about eight weeks, but it was hoped that, with the benefit of experience savings in time could be made for the work on the second carriageway.

However, it was recognised that where roller shutter expansion joints had been replaced on bridges elsewhere longer periods of closures were required. With consideration for the time of year and the time required to undertake the work it was determined that the most suitable time period for the works was April to June, before the peak traffic flow months of July and August.

Recent work for the Main Cable Replacement/Augmentation Study had shown that the notional road user delay costs of closing one carriageway on the bridge is around £650,000 per weekday. These costs do not take into account the costs to the wider business community which are of a higher order of magnitude than the user delay costs. For an eight week carriageway closure the costs and associated disruption was very significant, so further work was carried out to try and find an innovative way to determine if such a closure period could be reduced.

Figure 12 – Virtual Reality image of temporary bridge

Provision of temporary bridges

Given the implications of the works FETA engaged consulting engineers Flint & Neill to undertake a peer review of the feasibility and design work undertaken. Flint and Neill were in agreement with the proposals recommended by Atkins.

Temporary bridging over the roller shutter joints to reduce traffic delays had been considered at the feasibility stage but was discounted mainly because it limited access to the joints themselves. However, this option was revived following a suggestion from Flint & Neill that "a temporary bridge construction (Mabey QuickBridge) had been used on the E4/E20 Essingeleden in Stockholm, a main ring road which takes the E4/E20 around the west side of the city centre this had been used on another bridge." After a series of meetings between FETA's project team, Flint & Neill and Atkins, with preliminary design work by Atkins, and check by Faber Maunsell, it was determined that it was feasible to erect similar temporary bridges over the joints on the Forth Road Bridge. Erection and dismantling of the temporary bridges would still require temporary closures of a carriageway but following discussions with potential suppliers it appeared feasible that such bridges could be erected over a long weekend. At the completion of the works on the first carriageway the temporary bridges would then be transferred onto the second carriageway.

The temporary bridges, including approach ramps, would be about 80 metres in length. The temporary bridges could only accommodate two narrow carriageway lanes between the existing bridge parapets.

The existing bridge deck would also have to be strengthened locally to allow for the loadings from the temporary bridge and even then it would be necessary to limit the maximum vehicle weight on the temporary bridges to 3.5 tonnes. The design would also have to permit the movement of the main bridge. The headroom under the cross bracing of the bridge towers restricted the deck height of the temporary bridge and therefore the available room underneath for working.

Other factors considered included a concern that the raised level of the temporary deck would make vehicles more susceptible to cross winds, the containment level of the temporary bridge parapets would be difficult to design for, and the vertical alignment of the carriageway would reduce the throughput of vehicles. All these factors meant that a 30 mph speed restriction would need to be imposed and that vehicle weight and speed would need to be closely policed.

The restrictions on the use of the bridge by heavy goods vehicles would apply only to one carriageway at a time. There would be no restrictions on the opposite carriageway. Heavy goods vehicles make up approximately 6% of the traffic on the bridge and the split northbound and southbound is approximately 50/50. Therefore, although the temporary bridges would restrict heavy goods vehicles from using one carriageway for a period, which would have an effect on industry, only around 3% of the traffic would be affected by the restriction. The fact that most traffic could still cross meant that the diversion routes would not be as congested.

The FETA Board accepted the recommendation that the user delay benefits and the significant reduction in the level of disruption to the public outweighed the additional costs and restrictions of the temporary bridges. The temporary bridges were therefore included in the project.

Adopting the temporary bridges led to a review of the proposal to replace the joints in the approach viaducts. Atkins concluded that, although the viaduct joints were the originals, their range of movement was less than the roller shutter joints so they had suffered less wear giving no immediate urgency for replacement. In addition, the consequence and mode of failure of these smaller joints was less than for the roller shutter joints. It was now recommended that the viaduct joints continue to be monitored and inspected and their replacement programmed to be carried out in the future.

Tenders for the joint replacement works

Following a pre-qualification phase, tender information was issued to three contractors in July 2008 and tenders were returned in September 2008. The tender was based on a quality/cost ratio of 80/20 for the following work:

- The replacement of all roller shutter joints with a similar joint of the same movement range but with improved details
- The replacement of all adjacent sliding plate joints within the cycleways/footways with a similar joint
- Provision of new permanent access platforms to allow improved inspection and maintenance of the roller shutter joints below the carriageway. The design of these elements was to be the contractor's responsibility
- The provision of two temporary bridges to allow vehicles under 3500kg to use the carriageway while the joints are being replaced. The design of these elements was also to be the contractor's responsibility
- The strengthening of the deck and truss to accommodate the loading from the temporary bridges.

The pre-tender estimate for the Works was approximately £8.7m. Including the design, site supervision costs and ancillary works the total cost of the project was brought to approximately £10.3 million. Following an analysis of the tenders the submission recommended for acceptance on the basis of being most economically advantageous was for a sum of £13.8m, approximately £5m above the estimate. The main difference between the estimate and the returned tender was in preliminary items and the temporary bridging. These items proved difficult to evaluate because of their unique nature and high contractual risk. Based on the returned tenders the cost of the temporary bridging was over £6m. The contract, therefore, could not be awarded until the issue of funding was resolved and so FETA began to explore ways of bridging the funding gap.

Figure 13 – Plate train removed

Announcement of the Forth Replacement Crossing

While funding was trying to be resolved the Scottish Government announced details of their Strategic Transport Projects Review which sets priorities for funding over the coming 10 year period. Within that announcement were further details relating to the Forth Replacement Crossing stating that it would be in place by 2016. Following the Minister's announcement and the commitment to a definite programme for the construction of the Forth Replacement Crossing a review of the project was carried out to determine whether the replacement of the joints could be deferred until 2016.

The advantages in deferring the work until 2016 would be substantial. It would be possible to close one carriageway at a time and carry out the work with less disruption as traffic could be diverted to the replacement bridge. This, therefore, removed the need to provide the temporary bridging and would result in a saving of over £6m. However, the delay would mean that the existing joints would need to be maintained for longer.

Review of the project

Failure Mode and Effect Analysis

It was known that the joints had reached, or were reaching, the end of their service life and there were concerns over their reliability and the consequences of failure in service. The brief for the review, therefore, focused on the safety of bridge users and potential disruption to traffic should a joint fail. It was also clear that if, following the review, FETA determined that replacement of the joints could not be delayed then action on solving the funding shortfall would recommence immediately.

The object of the review was to determine not only potential failure modes but also whether or not it was possible to put safeguarding measures in place.

The review was carried out using Failure Mode and Effect Analysis (FMEA) techniques by FETA and Atkins, with Flint & Neill continuing to act as peer reviewer, see table 1 for extract of FMEA assessment. The FMEA was suggested by FETA, and although it is more commonly used in the aeronautical and process industries rather than in construction, it was considered to be the best means of identifying the likelihood and consequences of failure of each of the components that make up the joints. Once the risk and consequence of failure of individual components had been identified an analysis to determine what mitigation measures could be put in place to reduce the risk to an acceptable level was then carried out.

To assist in the above one plate train, including shuttle plate and tongue plate was removed, see Figure 13, by FETA staff for inspection during a weekend closure in January 2009. This inspection provided further information to allow benchmarking of the condition of the joints to be established.

STRUCTURES

A workshop for the FMEA, with both consultants and FETA staff, was also carried out during that weekend to enable a decision to be reached as soon as possible.

To enable identification of which component has the highest failure risk factor a Risk Priority Number (RPN) was determined. The highest risk components will have the highest RPN. The RPN is derived from three criteria: severity, occurrence and detection, with each based on a 10 point scale. Severity is defined as the consequence of failure should it occur, i.e. traffic accident and/or carriageway closure. Occurrence is defined as the probability or frequency of the component failure occurring and detection is defined as the probability of the failure being detected before the impact of the effect is realised.

A high RPN does not necessarily mean that a component has a high probability of failure. The component could have a relatively low risk but the effect of the failure is severe or the possibility of detection prior to failure is low. Consideration was also given to the possibility that failure of an individual component may not lead to a failure of the joint in itself but that it may lead to a 'domino effect' (or failure tree) by overstressing other components which could then lead to joint failure.

Mitigation measures

The review team was in agreement with the earlier study that the joints had reached the end of their service life. However, it was recognised that the decision to replace the joints had been taken without the certainty of the timescale for the Forth Replacement Crossing. Given this certainty the review team concluded that it would be possible to delay the replacement of the joints until 2016 subject to the following:

- The current inspection and monitoring regime should be increased significantly
- The installation of the permanent access system which formed part of the original replacement contract should be installed as soon as practicable
- The shuttle and tongue plate holding down pins and springs should be replaced
- Temporary failsafe devices should be installed as a precaution in case of failure of components with high risk
- The decision to defer the joint replacement should be reviewed annually or following any significant component failure or adverse inspection finding.

Increasing the current inspection and monitoring regime is seen as the key factor in ensuring the continued safety and reliability of the joints. This relies on the in-house skills and experience of FETA staff which has been built up over the years. The inspection work would include lifting out each of the 48 plate trains of the joints over the period between now and 2016. The experience gained in January 2009 led to the conclusion that this work could be achieved during overnight weekend carriageway closures with minimal disruption to traffic.

The temporary failsafe measures will involve installing restraint straps and stop end blocks that would prevent loss of the plate trains and shuttle plates that would lead to large gaps being created in the carriageway. It was recognised that failure could potentially occur because of defects which are not immediately obvious during the inspections.

The failsafe measures and increased inspections would lead to additional costs but these are estimated to be between £150,000 and £250,000 as a total between now and 2016. These costs are relatively small in comparison to the cost saving of providing the temporary bridging (over £6m). The cost of the installation of the improved access system (approximately £600,000) is seen as neutral as it would be included in the joint replacement works.

Residual risks

Even with the above measures there are other residual risks that have been identified. These include risks to operational service levels, unforeseen risks that future movements and articulations that cannot be predicted cause failure, and that the Forth Replacement Crossing does not open in 2016.

In extending the service life of the joints the risk that a component fails obviously increases, however, the mitigation measures of inspections, failsafe devices and replacing pins and springs will minimise that risk to an acceptable level. A failure event is likely to result in current operating service levels being reduced as remedial works would necessitate temporary closures of a carriageway. The duration of such carriageway closures depends on the nature of the failure but they would be expected to be measured in days rather than weeks.

The recent inspection of the joints confirmed that, because of wear, they are not behaving in a manner that was envisaged when they were designed. Given this, it is not possible to predict with any real accuracy future unknown movements and articulations of the joints as they increasingly continue to wear. Therefore, there is a risk that such unknown future movements will result in the need to replace a major component such as a train prior to 2016. This would be very difficult to do without replacing other adjacent trains and would likely result in the need to replace the whole joint, which in turn would likely lead to the decision to replace all of the joints. It is expected that the recommended increase in the scope and frequency of the inspections will enable adequate, advance warning of the need to replace the joints before traffic restrictions would be required. However, funding would have to be arranged with the Scottish Government and a new contract to replace the joints would need to be awarded with the required temporary works to minimise disruption. It is estimated that it would take between 15 and 18 months to complete this work if it was found to be necessary.

The final main residual risk is that there is a delay to the opening of the Forth Replacement Crossing in 2016. As FETA does not have ownership of that risk this is something that should be monitored. A review of the decision to delay the replacement of the joints should also be carried out if there is any significant delay to the Forth Replacement Crossing project.

Table 1 - Extract from FMEA table

Line	Comp No	Component and function	Potential Failure Mode	Potential Effect of Failure	Potential Causes of Failure	Current Controls, Prevention	Current Controls, Detection
1	1	Shuttle Plate horizontal thrust block-attached to plate.	Loss of horizontal restraint of plate train.	Plate train becomes free and could fall into joint.	Weld failure from fatigue.	None.	6 monthly inspections.
2	1	"	"	"	Overloading of thrust block on shuttle plate (where wear between the feet and the track beams cause extra resistance).	None.	6 monthly inspections.
3	17	"	"	"	General corrosion.	None.	6 monthly inspections.
4	17	Shuttle Plate horizontal thrust block-attached to support.	Loss of horizontal restraint of plate train.	Plate train becomes free and could fall into joint.	Weld failure from fatigue.	None.	6 monthly inspections.
5	17	"	"	"	Overloading of thrust block attached to support (where wear between the feet and the track beams cause extra resistance).	None.	6 monthly inspections.
6	17	"	"	"	General corrosion.	None.	6 monthly inspections.
6a	17	"	"	"	Impact loading due to lack of fit.	None.	6 monthly inspections.
7	18	"	"	"	Overloading of thrust block support (where wear between the feet and the track beams cause extra resistance). Local failure of the top flange/cracking around block within supporting steelwork.	None.	6 monthly inspections.
8	2	Vertical bearing to Shuttle Plates-attached to plates.	Loss of vertical restraint of plate train.	Shuttle plate can rotate upwards about opposite bearing and protrude into carriageway.	Weld failure from fatigue.	None.	6 monthly inspections.

Initial Assessment							Resulting assessment with recommended action in place					
Severity							Severity					
Economic	Perception	Overall	Occurrence	Detection	RPN	Recommended Action	Economic	Perception	Overall	Occurrence	Detection	RPN
8	10	10	6	8	480	Close examination of plate train/tongue plate Access walkways Test welds End Stops Longitudinal restraint Anchored end	5	5	5	6	5	150
8	10	10	5	9	450	Close examination of plate train/tongue plate Access walkways End Stops Longitudinal restraint Anchored end	5	5	5	5	7	175
8	10	10	2	7	140	Close examination of plate train/tongue plate Access walkways End Stops Longitudinal restraint Anchored end	5	5	5	2	5	50
8	10	10	5	8	400	Close examination of plate train/tongue plate Access walkways End Stops Longitudinal restraint Anchored end	5	5	5	6	5	150
8	10	10	4	8	320	Close examination of plate train/tongue plate Access walkways End Stops Longitudinal restraint Anchored end	5	5	5	5	6	150
8	10	10	2	5	100	Close examination of plate train/tongue plate Access walkways End Stops Longitudinal restraint Anchored end	5	5	5	2	6	60
8	8	8	2	7	112	Close examination of plate train/tongue plate Access walkways End Stops Longitudinal restraint Anchored end	5	5	5	4	6	120
8	10	10	2	8	160	Close examination of plate train/tongue plate Access walkways End Stops Longitudinal restraint Anchored end	5	5	5	2	6	60
6	6	6	6	7	252	Close examination of plate train/tongue plate Access walkways End Stops Longitudinal restraint Anchored end	5	7	7	6	5	210

The maintenance of the main expansion joints on the Forth Road Bridge

Summary

The main expansion joints on the Forth Road Bridge have been in service since the bridge opened in 1964. The joints are of the rolling plate type whereby a series of plates slide over fixed curved girders.

The joints are now considered to have reached the end of their service life and a decision was taken to replace them. It was deemed unacceptable to close traffic lanes for the period required for the replacement work to be carried out; therefore, an innovative scheme to construct temporary bridges over the joints to allow traffic to pass over the works was devised. Tender prices for the contract were much higher than estimated due in the main to the high cost of the temporary bridges.

While additional funding was being sought an announcement by the Scottish Government for the Forth Replacement Crossing to be open in 2016 led to a review of the project.

The review was to determine if the service life of the joints could be extended until the replacement bridge opened. Delaying the replacement of the joints avoids the need for temporary bridges and therefore leads to a direct saving of some £6m, as well as indirect savings in delay costs.

The review was carried out using Failure Mode and Effect Analysis (FMEA) techniques. This analysis was considered to be the best means of identifying the likelihood and consequences of failure of the various components that make up the joints.

Once the risk and consequence of failure of individual components had been identified an analysis to determine what mitigation measures could be put in place to reduce the risk to an acceptable level was then carried out.

The FMEA review was a useful method of identifying where the risk of failure lies and enabled attention to be focused on where appropriate mitigation measures should be. The review concluded that the service life of the joints could be extended by increasing inspection and monitoring and by installing failsafe devices without reducing the current high operational safety levels on the bridge.

However, it was recognised that there was a risk of a reduction in operational service levels which had to be balanced against the potential cost savings. It is now expected that the replacement of the joints can be deferred until 2016.

Predicting the Post-limit Softening response of structural materials by implementing Finite Element analysis

Fung Tong
Graduate Civil Engineer
Rail
Atkins

Abstract

This paper reports on the proposed finite element-based material model, known as the Post-limit Softening Material (PSM) model, which was developed to simulate the softening responses of steel and concrete materials. The PSM model is able to overcome the renowned computational numerical instabilities, due to negative stiffness, to capture the complete stress distribution of the materials.

Several validation cases were presented which includes examples from previous work and from recent developments of the model. Preliminary results have generally shown good agreement. Independent case studies were considered due to lack of available test data. While this model could not be regarded as fully developed it serves to stimulate further interest in softening analysis in the structural field.

Introduction

Limit state design has proved to be adequate and successful in its implementation into structural engineering designs. This design method, however, provides no indication of the softening response and possible failure predicament that is to follow. Two reasons why softening, a phenomena characterised by the decrease in stress with increasing deformation, cannot be ignored: firstly, there is obvious evidences that the softening phenomena happens in laboratory testing; and secondly, from an engineering point of view, the softening response leads to complete failure and thus 'warns' us of the imminent failure mechanism should ultimate limit state be exceeded.

It is generally acknowledged that the failure mechanism (i.e. softening, fracture etc.) of structural materials has gained much research interest among scholars in the recent times (Xiao, et al., 2005) (Komori, 2002) (Belnoue, et al., 2007) (Caballero, et al., 2005) (Rots, 2001).

The implementation of finite element (FE) analysis for this purpose, however, has been complicated by the inability of the FE computational algorithms to comprehend matrix non-positivity, characterized by the negative slope in the materials' stress-strain relations (Bangash, 2001). Analysis will either abort due to un-convergence or continue to follow the hardening paths, depending on the adopted material models. Therefore, an attempt was made to overcome this limitation which lead to the development of the proposed Post-limit Softening Material (PSM) model and hence opened the door for FE softening analysis. The ANSYS finite element software was used as the platform at which PSM operates.

A modeling technique was proposed in the previous study (Tong, 2008) (Xiao, et al., 2008) to investigate the tension softening behavior of typical metallic materials which included steel and copper. This technique was incorporated into the PSM model and was further enhanced to also allow for the post-limit investigation of cementitious materials.

In the present study, it was used to capture the softening response of typical plain concrete specimens under uniaxial tensile loading.

Although concrete is weak in tension, (codes of practice for design, such as BS8110 and BS5400 ignore tensile strength), the knowledge of it is essential, especially in controlling crack propagation in cementitious composites during thermal movement and shrinkage. It is foreseen that further research will allow designers to optimise the amount of reinforcement required by benefiting from the extra inherent strength, leading to a more sustainable design.

The constitutive theory of the Post-limit Softening Material model

The proposed PSM model has been developed based on credible laboratory observations on material behavior under cyclic loading. When the peak stress of each cyclic loop is connected they form a near identical hardening and softening path as a single monotonic test would follow (Karihallo, 2001).

This principle of increasing (and decreasing) stresses with increasing load cycles was adopted in this model, along with a series of material property update procedures to simulate the softening behavior.

The stress - strain relationships of metallic and concrete materials are characterized respectively by a 7-order polynomial function and the double-exponential (double-e) model (Barr, et al., 2003) as follows;

$$= C_i \quad ^i \qquad \text{(Equation 1)}$$

$$= c_1(e^{-c_2} - e^{-c_3 e}) \qquad \text{(Equation 2)}$$

where, - stress (N/mm²); - strain (dimensionless); *Ci* - polynomial constants (dimensionless) and 0 < *i* < 7; c_1, c_2 and c_3 - double-e constants (dimensionless). See (Tong, 2008) and (Barr, et al., 2003) for the determination of parameters.

When the ultimate tensile strength is attained, the material softens and hence the capacity to withstand load will decrease. Therefore an optimization procedure, employing the bisection method was developed (Tong, 2008) to determine the corresponding reduced load. This procedure utilizes the modified Voce hardening function (Voce, 1955) to update the material properties at each subsequent softening point. This function is given by;

$$= k + R \quad ^{pl} + R \ (1 - e^{-b \ ^{pl}}) \qquad \text{(Equation 3)}$$

where; k - elastic limit (N/mm²); R - modified constant (N/mm²), (originally the threshold stress in Voce's relation); ^{pl} - equivalent plastic strain (dimensionless); R - asymptotic stress (N/mm²); and b - Voce's parameter (dimensionless).

By employing such approach of updating the material property sets at each subsequent step, the analysis solution could avoid un-convergence due to negative stiffness.

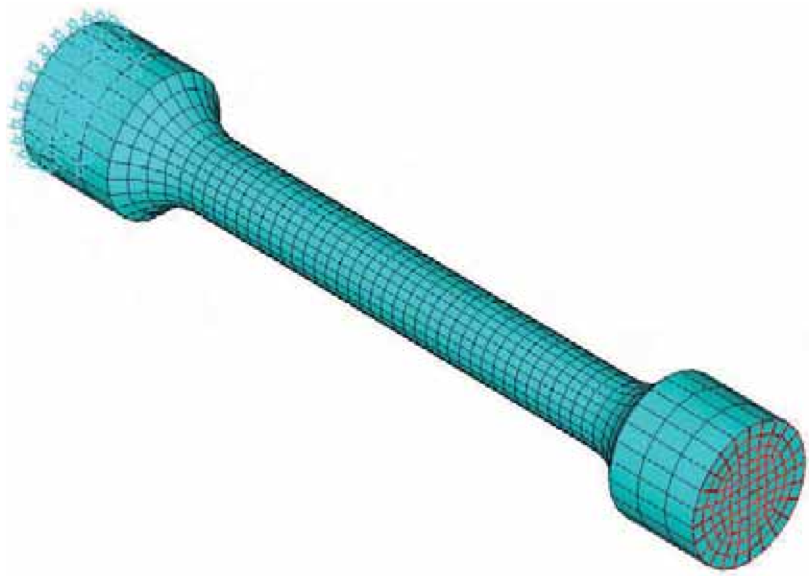


Figure 1 - The element mesh of G1X1A with applied load and boundary condition(s)

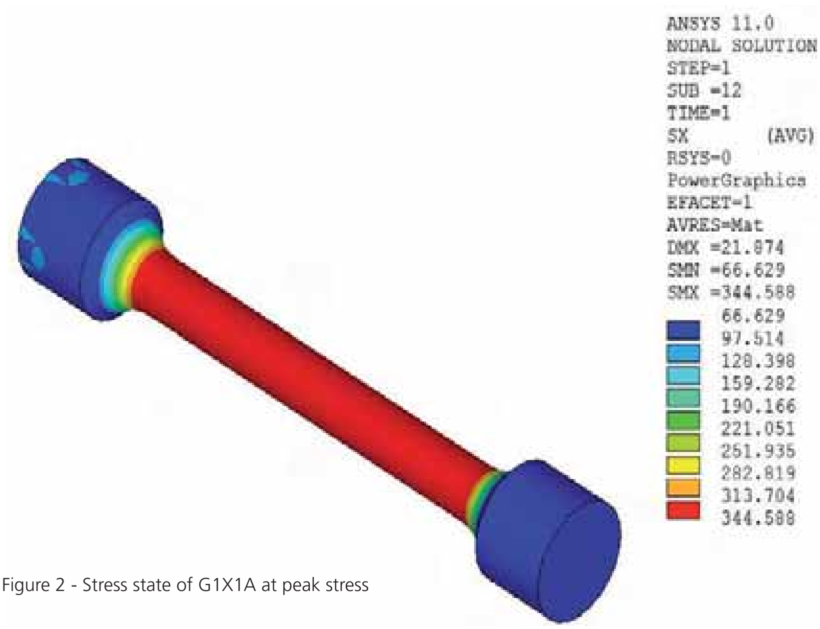


Figure 2 - Stress state of G1X1A at peak stress

Table 1. Specification of steel G1X1A and DP800

Specimens	Circular solid steel; G1X1A	Dual phase steel strip; DP800
Tensile Yield Stress, $_y$	125 MPa	500 MPa
Ultimate Strength, $_u$	250 MPa	780 MPa
Dimension	35 x φ5 mm	40 x 20 x 5 mm
References	(Barret, 1999)	(Xin, 2005)

Table 2 - Specification of prismatic concrete 2HB-1 and young concrete Mix B

Specimens	Throated prismatic concrete; 2HB-1	Young concrete; Mix B
Compressive strength, fcu	31.41 MPa	15.77 MPa
Tensile Strength, fct	2.819 MPa	3.21 MPa
Dimensions	100 x 100 x 100 x 210/70 x 70 mm	350 x 100 x 20 mm
References	(Guo, et al., 1987)	(Jin, et al., 2000)

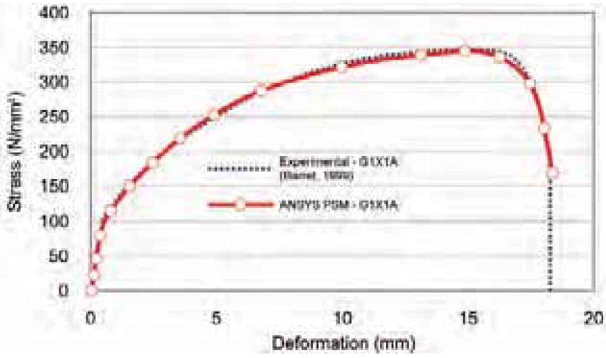


Figure 3 - Stress - deformation response of G1X1A

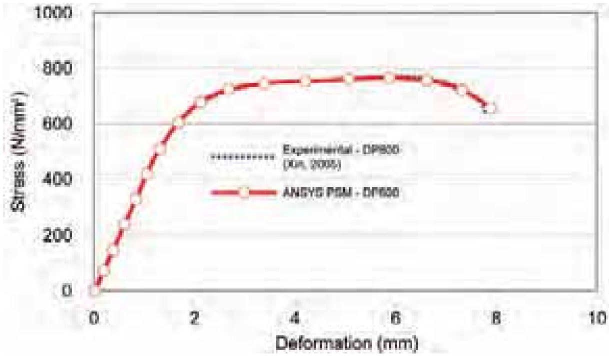


Figure 4 - Stress - deformation response of DP800

The Finite Element model

The finite element (FE) models, consisting of prismatic, dumbbell-like and plate geometries, were modelled by adopting 3-dimensional solid elements. The settings of the boundary conditions were such that they resemble the fixing conditions in the experimental tensile tests; one end being constrained in the x, y and z directions while pressure (surface) loads were applied at the other end. Larger load values were applied at the start of the solutions and the bisection subroutine were called to determine the maximum load at each loadstep. The material properties at each softening stage were updated automatically at intervals. The linear elastic properties, i.e. Young's (elastic) modulus and Poisson's ratio remain consistent and unchanged throughout the analysis. Each analysis was carried out in a continuous solution, from elastic to plastic hardening and to plastic softening.

Validation of experimental test cases

Several validation cases were presented to determine the efficiency of the PSM model and to provide confidence of its capabilities. Two examples on steel, extracted from previous work (Tong, 2008) (Xiao, et al., 2008), and further two examples on concrete from recent work were presented. The deformations (elongations) were measured between nodes corresponding to the location where the gauges were placed during the laboratory tests.

Table 1 tabulates the specifications of the FE models for the steel specimens. The sources from which the reference curves were obtained are also stated.

The FE models and the applied boundary conditions are shown in Figures 1 and 4. The corresponding stress - deformation responses of specimens G1X1A and DP800, as obtained from the analysis, were plotted shown in Figures 3 and 4. These are shown as the solid lines. The dotted (experimental) reference curves were also included as comparison.

The averaged nodal stress states at peak and fracture were also captured and shown in Figures 2,3,9 and 10. For the reader's reference, SMN and SMX denote the minimum and maximum stress respectively. DMX is the maximum displacement between the ends of the FE model and does not correspond to the displacement of the measuring gauges. Same applies to Figures 12, 13, 15 and 16.

Table 2 tabulates the specifications of the FE models for the two concrete specimens.

The FE models and the applied boundary conditions for the concrete specimens are shown in Figures 11 and 14. The PSM's results for the throated prismatic 2HB-1 and young concrete Mix B specimens, and the references curves are shown in Figures 5 and 6. The averaged nodal stress states at peak and fracture were captured, as shown in Figures 12, 13, 15 and 16.

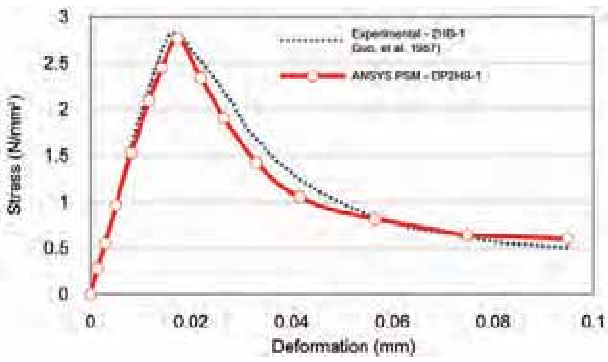


Figure 5 - Stress - deformation response of 2HB-1

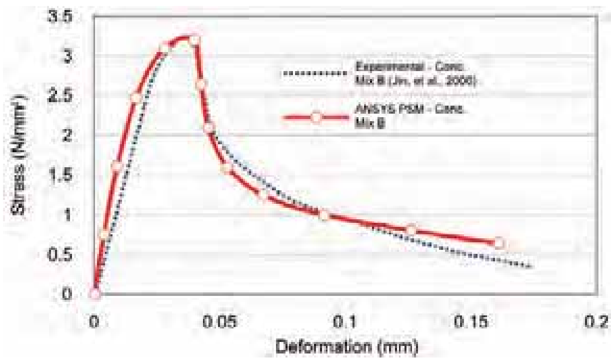


Figure 6 - Stress - deformation response of concrete Mix B

STRUCTURES

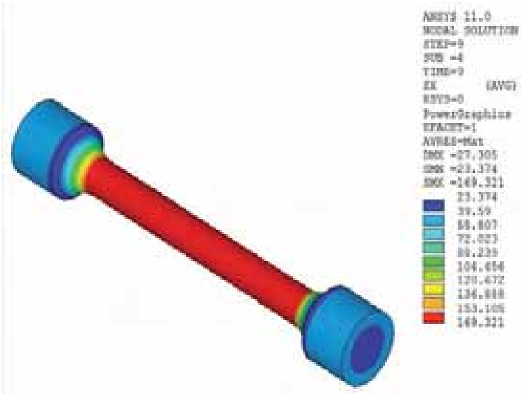


Figure 7 - Stress state of G1X1A at fracture stress

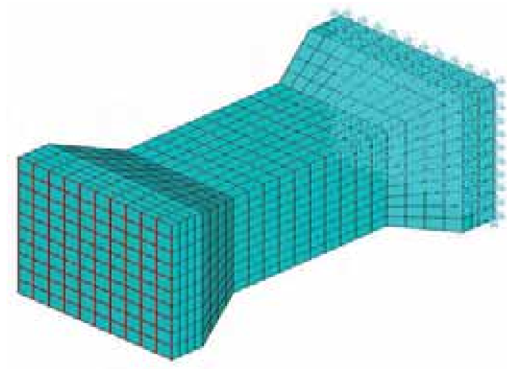


Figure 11 - The element mesh of 2HB-1 with applied load and boundary condition(s)

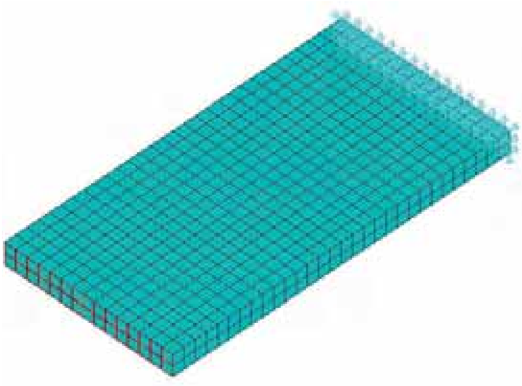


Figure 8 - The element mesh of DP800 with applied load and boundary condition(s)

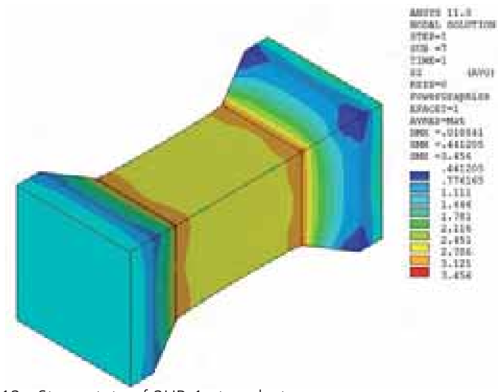


Figure 12 - Stress state of 2HB-1 at peak stress

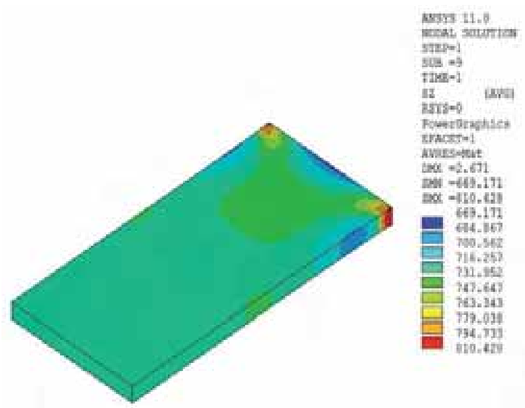


Figure 9 - Stress state of DP800 at peak stress

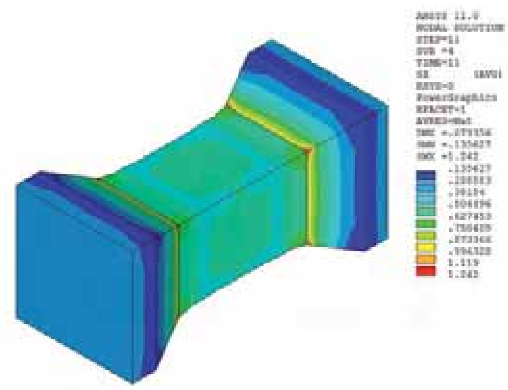


Figure 13 - Stress state of 2HB-1 at fracture stress

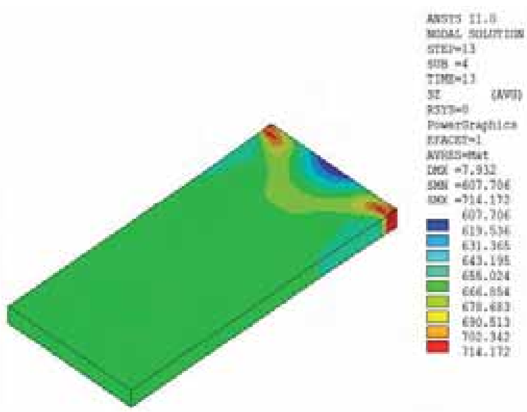


Figure 10 - Stress state of DP800 at fracture stress

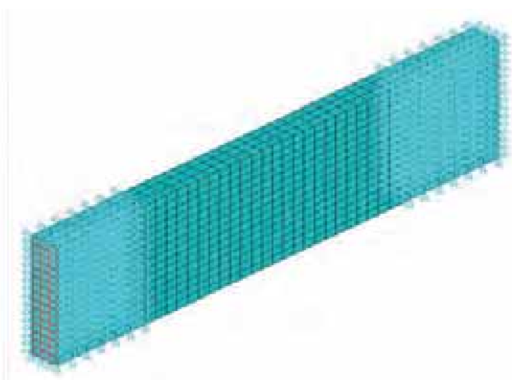


Figure 14 - The element mesh of Mix B concrete with applied load and boundary condition(s)

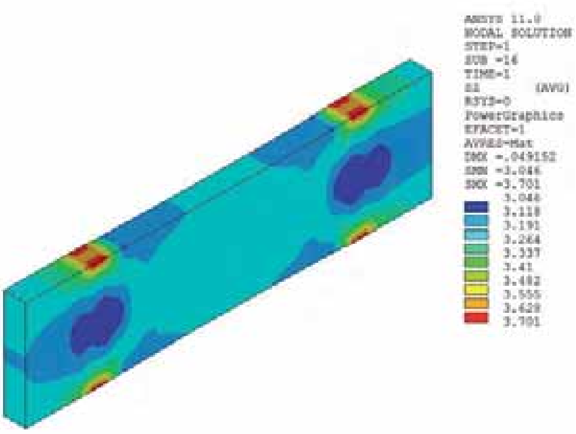


Figure 15 - Stress state of Mix B concrete at peak stress

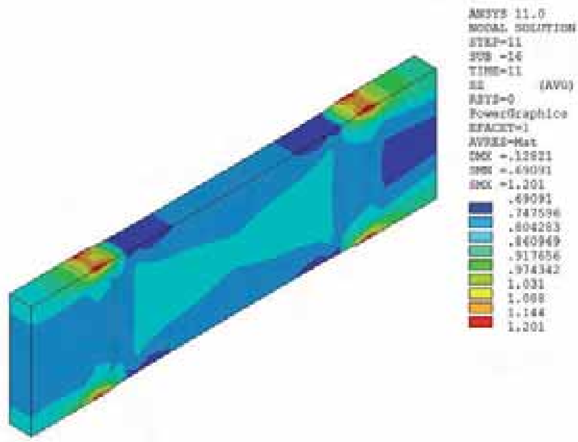


Figure 16 - Stress state of Mix B concrete at fracture stress

Discussions

It can be observed, by qualitative basis, that excellent (steel cases) and good (concrete cases) agreements between PSM results and the reference curves were attained with maximum differences of approximately 3.5% (G1X1A), 1.6% (DP800), 12.5% (2HB-1) and 19.8% (concrete Mix B) respectively.

The high accuracy as attained for the steel specimens owes to the high polynomial order of Eq. 1. It is, however, considered unsuitable for long term development of this equation due to the analytical nature of the softening parameters C_i. These parameters do not have a physical meaning but rather computed through curve fitting (Tong, 2008). On the other hand, the parameters in Eq. 2 for the concrete test cases were governed by the elementary material properties, e.g. elastic limit, ultimate strength etc, giving a more sensible solution for the softening analysis.

The results from these four cases have indeed evidently demonstrated that the softening (stress - deformation) responses of typical structural materials beyond the limit point could be captured with the proposed PSM model without encountering numerical instabilities.

Still not regarded as fully developed in the present time, the PSM model has great prospects to further enhance its capability. Therefore, the following has been identified for further developments of the model;

(1) Consideration for localisation of deformation. This is especially significant for ductile materials whereby some extent of the necking phenomenon will be expected; similarly, capturing cracking pattern in concrete

(2) Considerations for various loading conditions, e.g. flexural, punching shear, torsional failure, etc.

(3) Parametric considerations; size effects, boundary condition effects, prediction of crack growth in concrete etc.

This list is not exhaustive but should provide a good lead towards the enhancement of the model. It should be noted that large database of experimental test data will be required for the behavioral investigation of these structural materials. The validation cases carried out in the present study were case dependant, due to the lack of available test data.

Conclusions

Finite Element analysis beyond the peak stress has been limited by numerical instabilities, resulting from the negative stiffness which is characterized by the descending slope in the materials' stress-strain relations. An attempt was made to overcome this limitation which gave birth to the proposed PSM model. Several validation cases were performed and the results have demonstrated its capability of predicting the stress distribution by achieving good agreement with experimental data. Further developments of the current material model are required to boost its applicability to wide range of FE softening analysis. Presently, the PSM model has been adopted in uniaxial tension test on typical structural materials and has yet to be implemented to multiaxial structural engineering problems. It is hoped that the current study could serve to stimulate further research interest towards structural softening analysis.

References

1. Bangash M. Y.H. Manual of Numerical Methods in Concrete. Book. Thomas Telford Ltd.: London, 2001.
2. Barr B. and Lee M. K. Modelling the strain-softening behaviour of plain concrete using a double-exponential model. Magazine of Concrete Research. Thomas Telford Ltd., 2003; 4:55; pp. 343-353.
3. Barret Z. Uniaxial Tensile Test - Material Testing. Technical Report, University of Wales Swansea, Swansea 1999.
4. Belnoue J. P., Nguyen G. D. and Korsunsky A M. A One-Dimensional Nonlocal Damage-Plasticity Model for Ductile Materials. International Journal of Fracture. Springer Netherlands, 2007; 1:144; pp. 53-60.
5. Caballero A., Carol I. and Lopez C. M. 3D Meso-Structural Fracture Analysis of Concrete Under Uniaxial Tension and Compression. Anales De Mecanica De La Fractura. Barcelona; 2005; 22; pp. 581-586.
6. Chun L., Knutzen P. and Shen C. Cyclic Load Testing of Steel Bars. WWW -classes.usc.edu/engr/ce/334/PPT-5.ppt . - 2001.
7. Guo Zhen-hai and Zhang Xiu-qin Investigation of Complete Stress - Deformation Curves for Concrete in Tension. ACI Materials Journal. Detroit : 1987; 4:84; pp. 278-285.
8. Jiao H. and Zhao X.L. Material Ductility of Very High Strength (VHS) Circular Steel Tubes in Tension. Thin-Walled Structures; 2001; 11:39; pp. 887-906.
9. Jin X. and Li Z. Investigation on Mechanical Properties of Young Concrete. Materials and Structures. RILEM, 2000; 10: 33; pp. 627-633.
10. Karihallo Bhushan L. Fracture Mechanics & Structural Concrete. Book; ed. Kong F. K. and Evans R. H.. - London : Longman Group Ltd., 2001.
11. Komori K. Simulation of Tensile Test by Node Separation Method. Journal of Material Processing Technology. Japan : Elsevier Science, 2002; 125-126; pp. 608-612.
12. Neville A. M. Properties of Concrete (4th edition). Book. New York : John Wiley & Sons, Inc., 1963..
13. Rots J. G. Sequentially Linear Continuum Model for Concrete Fracture. Fracture Mechanics of Concrete Structures; ed. Borst R. de [et al.]; Lisse : A.A. Balkema, 2001; pp. 831-839.
14. Tong F. M. Nonlinear Finite Element Simulation of Non-local Tension Softening for High Strength Steel Materials. M.Phil Thesis. Swansea University, Swansea; 2008.
15. Voce E. A Practical Strain-Hardening Function. Metallurgica. 1955; Col. 51; pp. 219-220.
16. Xiao R. Y. and Chin C. S. Nonlinear Finite Element Modelling of the Tension Softening of Conventional and Fibrous Cementitious Composites. 13th UK National Conference of the Association of Computational Mechanics in Engineering. Sheffield; 2005; pp. 103-106.
17. Xiao R. Y., Tong F. M. and Chin C. S. Nonlinear Finite Element Simulation of Non-local Tension Softening for High Strength Steel Material. Symposium of Tubular Structures. Shanghai; 2008.
18. Xin Y. Optimisation of the Microstructure and Mechanical Properties of DP800 Strip Steel. Thesis. University of Wales Swansea, Swansea; 2005.

Design of collar beam for corrugated steel culvert

Neal Smith

Senior Assistant Engineer
Highways & Transportation
Atkins

Abstract

In accordance with BD12/01 (Highways Agency DMRB 2.2.6), end treatments to corrugated steel culverts have to be designed to support the face edges of the steel where:

- The skew angle of the structure exceeds 15°
- The bevel of the square end exceeds 2:1
- The cut end supports highway loading.

One suitable form of end treatment is a reinforced concrete collar beam. This should provide adequate support to the corrugated steel structure at its weakest location – the cut ends, where it is unable to act in ring compression. The design of Smallways North Bridge revealed that no specific guidance exists for the design of such collar beams, neither in standards nor in literature published by suppliers, of corrugated steel culverts. The purpose of this paper is therefore to outline one potential means of designing such elements.

Structure background

Smallways North Bridge was designed as part of the Highways Agency's A66 North East Package A carriageway widening project. The corrugated steel culvert carries the new eastbound carriageway of the A66 Trunk Road. It has a clear skew span of 7.953m, a skew angle of 25°, a bevel of 2:1 and a clear height of 4.346m.

Structural global analysis

A structural analysis package was used to model the collar beam with beam elements along the centreline of its section.

The model used in the analysis of Smallways North Bridge is shown in Figures 1 to 3. These Figures illustrate the skew and bevel of the collar beam.

Vertical and horizontal spring supports were used to model the ground supporting the structure, in accordance with the guidance given in Bridge Deck Behaviour by Hambly[2].

Vertical spring supports were provided to each joint on the underside of the model and a single lateral support was applied to the bottom of the model. Horizontal spring supports in the longitudinal direction of the culvert were applied at each joint below proposed ground level.

A single support was used at the top of the model (as shown in Figures 1 to 3) where the corrugated steel is able to act in ring compression.

The areas of steel unable to support themselves in ring compression, i.e. past the point of closed ring cross section, were assumed to be supported solely by the collar beam.

Therefore, all loads (loads per metre length) acting on those areas of steel were calculated and applied as point loads to the collar beam model shown above at its joints.

The collar beam was designed to resist the horizontal earth pressures acting on the "walls" of the corrugated steel, where ring compression cannot be achieved, and also self weight.

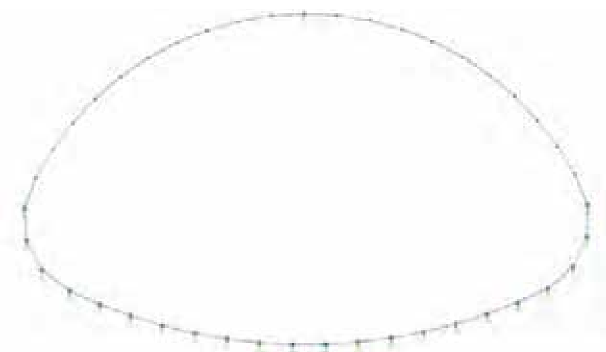


Figure 1 - End elevation of multi-radii analysis model

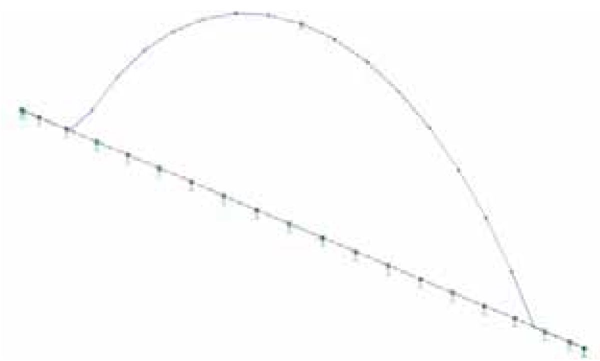


Figure 2 - Plan view of analysis model showing bevel and skew of culvert

STRUCTURES

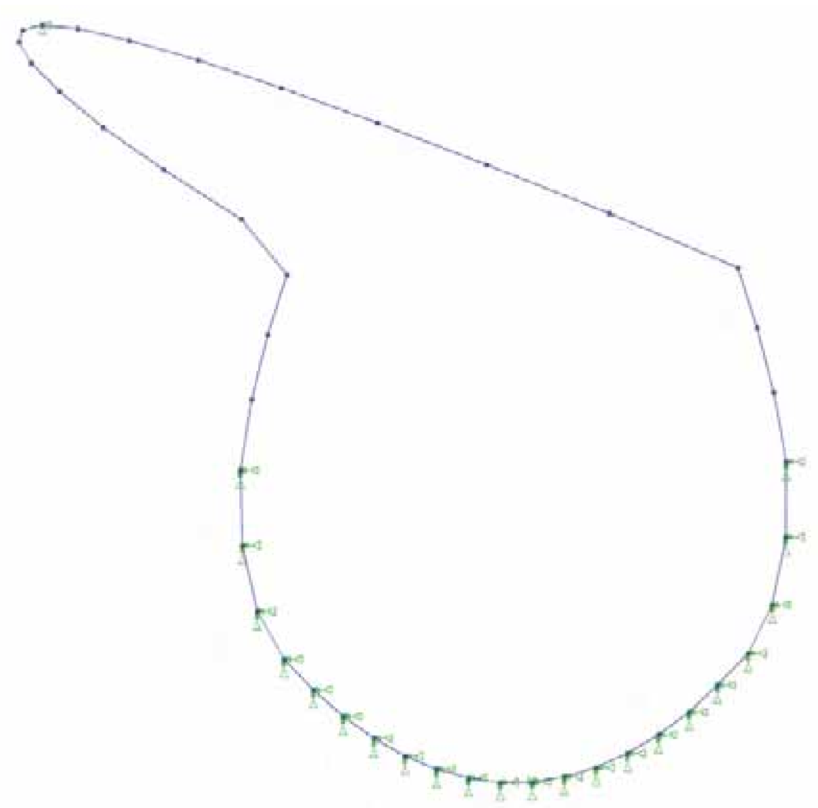


Figure 3 - Side elevation of analysis model showing bevel and skew

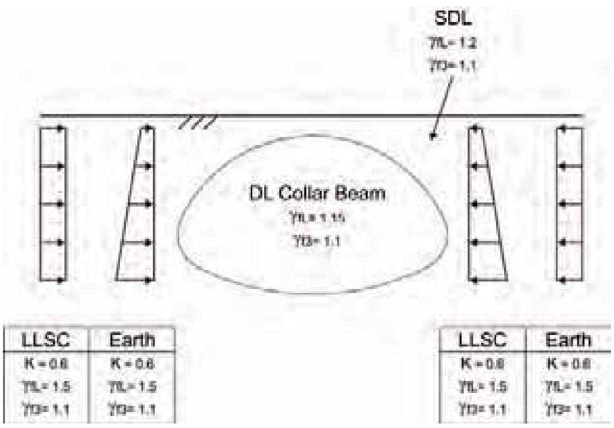


Figure 4 - Maximum vertical load with maximum horizontal load

The areas of steel requiring support were determined first. Support is offered to these areas by the collar beam and also by the adjacent steel acting in ring compression. To be conservative it was assumed that these areas were supported by the collar beam alone.

The loads acting on these areas were then calculated and applied to the collar beam model. Earth pressures were applied to the model along with dead loads and superimposed dead loads in the same manner as for the design of a buried culvert to BD 31/01[3] (DMRB 2.2.12), using load factors given in Table 3.1 and 4.1 of BD 12/01[1]. Maximum and minimum horizontal earth pressures were calculated using k × × D, where k is the earth pressure coefficient (to be taken as 0.2 and 0.6 for minimum and maximum horizontal loadcases respectively), is the unit weight of the fill material and D is the depth of fill material.

Similarly a horizontal live load surcharge, p_{sc} was applied to represent construction vehicles compacting the fill behind the structure, given by $p_{sc} = k \times v_{sc}$ where v_{sc} is the vertical surcharge pressure of the construction vehicle, given in BD 31/01[3] as 10 kN/m^2. These horizontal loads were applied as point loads at joints also.

Load combinations were used to determine maximum bending moments and shear forces:

- Maximum vertical loads + maximum horizontal loads (see Figure 4)
- Maximum vertical loads + minimum horizontal loads (see Figure 5)
- Minimum vertical loads + maximum horizontal loads (see Figure 6)

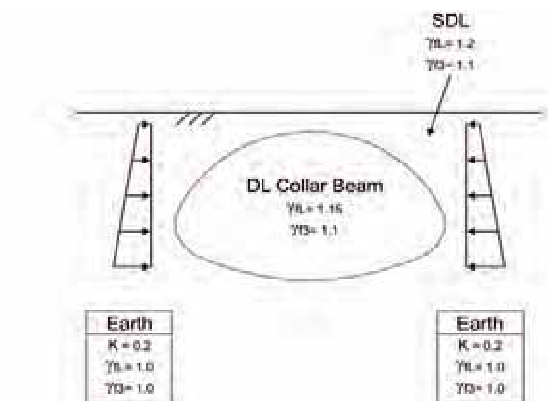


Figure 5 - Maximum vertical load with minimum horizontal load

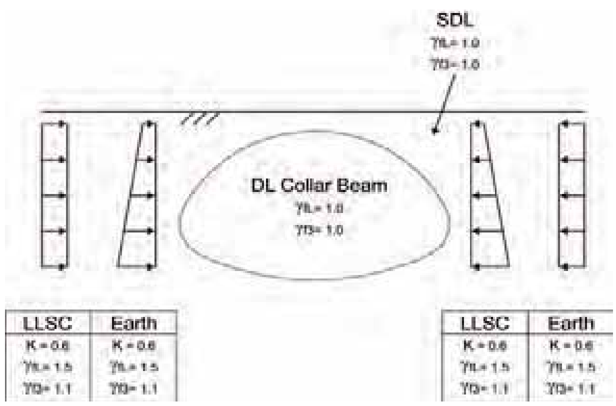


Figure 6 - Minimum vertical load with maximum horizontal load

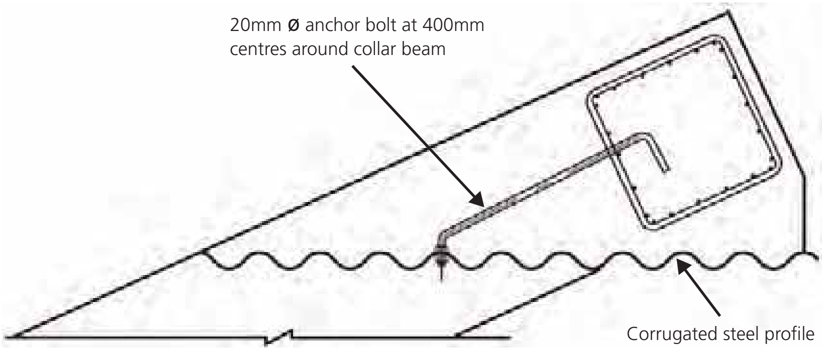


Figure 7 - Collar beam fixings

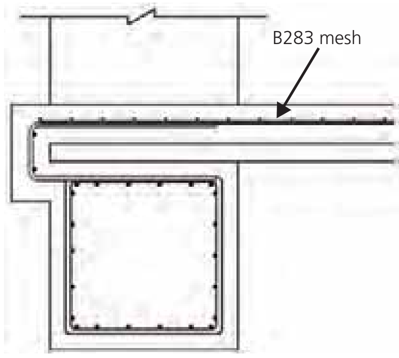


Figure 8 - Collar beam reinforcement lapping with reinforcement for invert protection paving

Figure 9 - Fixing of reinforcement to bottom half of collar beam

Figure 10 - Fixing of reinforcement to top half of collar beam

Figure 11 - Scaffold arrangement for access and for propping of corrugated steel during concrete pour

Figure 12 - Collar beam concrete in place

Figure 13 - Finished collar beam

Section design

Once the load effects were obtained from the model the reinforced concrete section was designed in accordance with the requirements of BS 5400-Part 4[4].

The arrangement of steel reinforcement in the collar beam was taken to be constant throughout its length. The amount of longitudinal steel required for bending was determined at the peak bending moment positions.

A check of the additional longitudinal steel required for shear was then carried out at the peak shear force positions to make sure that the total amount of reinforcement required did not exceed that area of steel required for bending at the peak bending moment position. The amount of longitudinal steel required for torsion was then checked. Closed links were provided to resist torsion as per the requirements of BS 5400-Part 4[4].

For a complex shape of collar beam, being skewed and bevelled, it was considered sensible to use B12 reinforcing bars. This would make the bars easier to fix on site as they could be bent slightly to suit.

The reinforced concrete collar beam was also checked for crack widths in accordance with BS 5400-Part 4[4] and early thermal cracking in accordance with BD 28/87[5] (DMRB 1.3.14).

Construction

The collar beam was fixed to the corrugated steel using 20mm diameter anchor bolts at 400mm spacings around the structure (see Figure 7).

A reinforced concrete invert protection paving system was incorporated into the design to meet the requirements of BD 12/01[1]. The fabric mesh reinforcement from the invert protection paving was lapped with the collar beam reinforcement, as shown in Figure 8.

The construction of Smallways North Bridge is shown in Figures 9 to 13. Figure 9 illustrates the fixing of reinforcement to the bottom half of the collar beam, while Figure 10 shows fixing of reinforcement to the top half. Figure 11 illustrates the use of scaffolding and temporary propping required during the pouring of concrete. Figures 12 and 13 show the finishing of the concrete and completed collar beam respectively.

During the construction of Smallways North Bridge, Smallways Beck was diverted around the new structure by using water pumping equipment, as can be seen in Figure 9.

Conclusions

A simple design procedure for reinforced concrete collar beams has been proposed. Where a reinforced concrete collar beam is required in accordance with BD 12/01[1], it may be designed as follows:

(i) Produce line beam model representing centreline of collar beam

(ii) Use vertical spring supports to underside of collar beam and horizontal spring supports in the longitudinal direction of the culvert at those locations where the collar beam is below ground

(iii) Apply loads to the model: dead loads superimposed dead loads, earth pressures and live load surcharges to represent construction vehicles backfilling and compacting the material behind the culvert walls. Loads should be applied as those combinations shown in Figures 4 to 6

(iv) Provide sufficient steel to resist the load effects of bending, shear and torsion, obtained from the model

(v) Carry out SLS checks for crack widths and early thermal cracking.

References

1. HIGHWAYS AGENCY. Design Manual for Roads and Bridges. BD 12/01 – Design of Corrugated Steel Buried Structures with Spans Greater than 0.9 Metres and up to 8.0 Metres. Highways Agency, London, 2001.
2. HAMBLY E. C. Bridge Deck Behaviour. Second Edition. E & F N Spon, 1991.
3. HIGHWAYS AGENCY. Design Manual for Roads and Bridges. BD 31/01 – Design of Buried Concrete Box and Portal Frame Structures. Highways Agency, London, 2001.
4. BRITISH STANDARDS INSTITUTION. Steel, Concrete and Composite Bridges, Part 4 – Code of Practice for Concrete Bridges. BSI, London, 1984, BS 5400.
5. HIGHWAYS AGENCY. Design Manual for Roads and Bridges. BD 28/87 – Early Thermal Cracking of Concrete (Incorporating Amendment No. 1 Dated August 1989). Highways Agency, London, 1987.

Lift-over crossings as a solution to tram-ted ground-borne vibration and re-radiated

James P Talbot

Principal Vibration Engineer
Design & Engineering
Atkins

Abstract

The operation of tramways close to sensitive buildings can lead to concerns over ground-borne vibration and re-radiated noise. Vibration generated at the wheel-rail interface propagates through the track structure, through the ground and into buildings, where it may cause disturbance as perceptible vibration and/or re-radiated noise.

This paper presents work undertaken to solve a re-radiated noise problem within a UK concert hall. The hall in question is situated alongside a tramway that includes a crossover between two rail tracks. Initial measurements established the dominance of re-radiated noise over airborne noise. Simultaneous noise and vibration measurements were then used to establish the relative significance of the impulsive vibration generated at the various rail discontinuities of the crossover, compared with the essentially continuous vibration due to wheel/rail roughness. The results led to the selection of a new 'lift-over' crossing, together with an improved design of switch, as the basis for solving the problem.

The paper includes descriptions of the experimental methods, together with a summary of the results. The new crossover design is described and the results of the commissioning measurements are presented as a final demonstration of the new hardware's performance.

Introduction

Tramways are one of the most significant sources of ground-borne vibration in our cities[1-3]. Vibration generated at the wheel-rail interface propagates through the track structure, through the ground and into buildings, where it may cause elements of the building structure to vibrate. This vibration can be felt by a building's occupants and is known as perceptible vibration when the level is such that the comfort of the occupants is adversely affected.

Structural vibration also radiates sound and this can be significant within the audio frequency range, approximately 25Hz and above. Re-radiated noise (also termed structure-borne or ground-borne noise) describes vibration, originally radiated through the ground and into a building, which is then re-radiated as airborne noise [4].

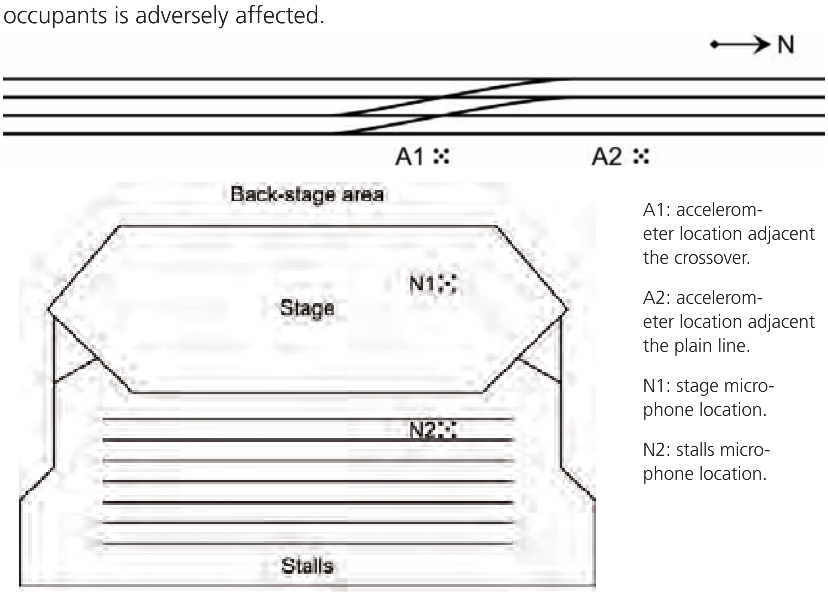


Figure 1 - Schematic diagram of the concert hall in plan, showing the location of the adjacent crossover and the locations at which noise and vibration measurements were made

The result is an audible low-frequency 'rumble' which, depending on the radiation efficiency of the particular structure, is usually most noticeable in the frequency range from 50Hz to 125Hz.

This paper is concerned with the diagnosis and solution of an intrusive re-radiated noise problem within the auditorium of a UK concert hall.

Overview of the problem

The concert hall in question is situated alongside a tramway that includes a crossover between two rail tracks. The crossover lies approximately 8m away from the rear façade of the concert hall, which separates the tramway from the back-stage space of the auditorium, as illustrated in Figure 1.

The hall was built to a high standard approximately 20 years ago but this was before the tramway was conceived and no special measures were taken to limit the effects of ground-borne vibration. Since the construction of the tramway the hall suffered significant disturbance due to the operation of the trams. Although the level of perceptible vibration is low the level of re-radiated noise was significant.

The individual tram pass-bys were clearly perceptible as low-frequency 'rumbles', both on the stage and in the auditorium, and the noise was intrusive for both the orchestra and the audience.

Overview of the project

The auditorium is well isolated acoustically and standard sound-insulation measurements, made using the global loudspeaker method[5], showed that the level of airborne noise from the trams was insignificant. Having formally confirmed that ground-borne noise was the dominant cause of the disturbance, the work presented here focused on diagnosing the source of the noise, followed by the design of a replacement crossover to mitigate the problem.

Diagnosis of the re-radiated noise source

There are primarily two sources of ground-borne vibration, and hence re-radiated noise, associated with tramways: the inherent roughness of the wheels and rails; and discontinuities in the rails, such as those found at conventional track crossovers. This section describes the investigatory measurements that were made to establish the relative significance of the impulsive vibration generated at the various rail discontinuities of the crossover compared with the essentially continuous vibration due to wheel/rail roughness.

The purpose of the measurements was to record a series of continuous noise and vibration time-histories as trams made controlled pass-bys from the nominally straight section of plain line, over the crossover and beyond. Tram speeds of 10 kph and 20 kph (the typical service speed) were considered and these were held constant over both the plain line and the crossover.

Noise and vibration measurements

Vibration measurement locations were established by the side of the concert hall, adjacent to the centre of the crossover and adjacent to the preceding section of plain line see Figure 1. In both cases, the stand-off distance of the measurement location from the centre of the southbound track was 3m.

The same type of accelerometer was used to measure the vertical vibration of the ground at both locations. The accelerometers were mounted on heavy steel blocks, which provided adequate coupling to the ground over the frequency range of interest.

Noise measurement locations were established inside the concert hall, on the stage and in the front seats of the stalls see Figure 1. These are the closest locations to the trams at which a musician and member of the audience may be seated. In both cases, the microphone was positioned approximately at the head-height of a seated person. The same type of sound level meter was used for both locations.

The output signals from the accelerometers, along with the linear outputs of the sound level meters, were recorded simultaneously by a common data acquisition unit. Subsequent data processing was undertaken using the Matlab technical computing software[6].

Typical time-history results

Figure 2 presents some typical results for southbound trams travelling at 20 kph. The re-radiated noise results are presented in terms of the A-weighted sound pressure level, which is commonly used to characterise environmental noise [1]. It is based on the A-weighting curve, which filters the noise to account for the non-linearity of human hearing, and characterises the noise measured across the whole of the audible frequency range in terms of a single overall level. The time averaging employed is that corresponding to the 'slow' time constant.

The corresponding lineside vibration is presented in terms of the acceleration level of the ground surface.

As the tram approaches the crossover, and gets closer to the concert hall, both the noise and the vibration levels gradually increase. The transition from the essentially continuous roughness-induced vibration to the impulsive vibration generated at the rail discontinuities is clear approximately 12s into the recording. Once on the crossover, the individual wheel-rail impacts are clearly visible in the vibration time-history, and these result in significantly higher noise levels in the hall.

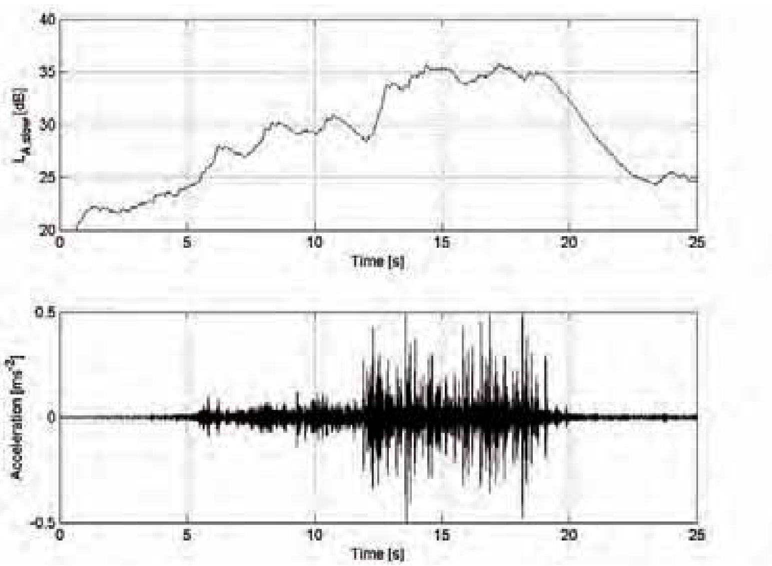


Figure 2 - Typical variation in the A-weighted sound pressure level with time, as measured on the concert hall stage, together with the associated lineside vibration measured adjacent the crossover. Southbound tram at 20kph

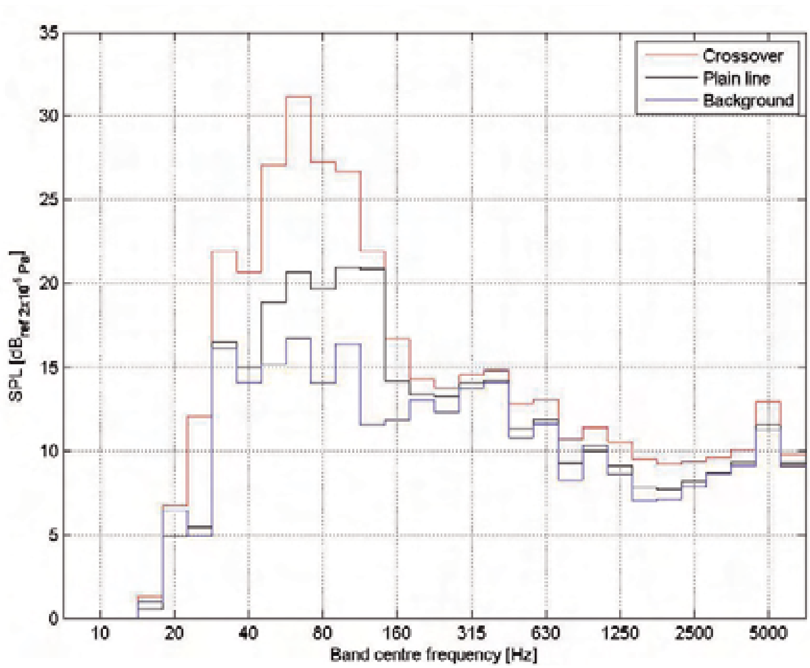


Figure 3 - comparison of the mean A-weighted noise spectrum due to trams travelling on the crossover with that due to trams on the preceding plain line. The background spectrum (in the absence of any trams) is also plotted. Stage measurement location; southbound trams at 20kph

Noise spectra for the crossover and plain-line sections

In addition to a single overall level, noise levels are often described in terms of A-weighted third-octave band spectra. These use the same A-weighting as the LA metric but characterise the noise in terms of its distribution with frequency, which is divided into bands of one third of an octave. Third-octave spectra enable the dominant frequency components of the noise to be established.

Due to the sensitivity of rail vehicle dynamics to a large number of variables (speed, bogie stiffness, wheel-rail interface parameters, etc.), significant variations are observed between nominally identical pass-bys. In this case, re-radiated noise levels are observed to vary by up to 1.5dB(A) for the same tram travelling on the same line at the same speed. Mean noise levels should therefore be considered wherever possible, using a large dataset to obtain statistically significant results. Throughout this project, between 10 and 18 pass-bys were considered for each particular pass-by condition.

Figure 3 compares the mean spectrum of the re-radiated noise due to trams travelling at 20kph on the crossover with that due to trams on the plain line.

The spectra are calculated by sectioning the noise time-histories and processing separately the data acquired with the tram on the crossover and that acquired with the tram on the plain line immediately preceding it.

At a tram speed of 20kph, the re-radiated noise levels on the stage due to trams travelling on the plain line exceed the background level by up to 10dB(A) in any one third-octave band. In contrast, the levels due to trams traversing the crossover exceed the background by up to 16dB(A). In general, the noise levels significantly exceed the background level over the frequency range from approximately 25Hz to 160Hz, with the peak level occurring in either the 63Hz or 80Hz bands. It is this peak in the noise spectrum that aids the perception of the trams in the concert hall.

Overall noise levels

By analysing the overall noise levels (such as those plotted in Figure 2) and comparing the peak level with that generated by the tram on the plain line, just before the first wheel impacts, it was possible to estimate the expected reduction in re-radiated noise in the event that the crossover was removed or made to behave effectively as plain line. These levels are summarised in Table 1.

For trams travelling at 20kph, the data indicate that removal of the crossover would result in a reduction in overall noise levels of approximately 6dB(A). At 10kph the reduction is approximately 4dB(A). Both reductions are significant in that they would be clearly noticeable – changes in level of over 3dB(A) are typically discernible by the human ear.

Identification of vibration triggers

The measurements presented above provide clear evidence that impulsive vibration generated at the rail discontinuities of the crossover was the dominant source of the re-radiated noise. It was also important to understand the nature of these vibration 'triggers', so that informed consideration could be given to possible remedial measures.

A comprehensive track survey was undertaken, together with more detailed analysis of the vibration time-history data. This enabled various features of the time-histories to be correlated with identified features of the track hardware. It was clear that the same transients were evident in different time-histories, indicating that the mechanisms of vibration generation were repeatable between trams and that the same track features were responsible during each tram pass-by.

Table 1 - Comparison of mean peak noise levels ($L_{Amax,slow}$) associated with trams traversing the crossover with those due to trams travelling on the preceding plain line

Tram speed [kph]	Peak level (crossover) [dB(A)]		Level on plain line [dB(A)]		Difference [dB(A)]	
	Stage	Stalls	Stage	Stalls	Stage	Stalls
10	31.4	30.1	27.6	26.0	3.8	4.1
20	35.7	32.6	30.1	26.6	5.6	6.0

Switch, Mate and Frog, Unbroken Main Line Construction

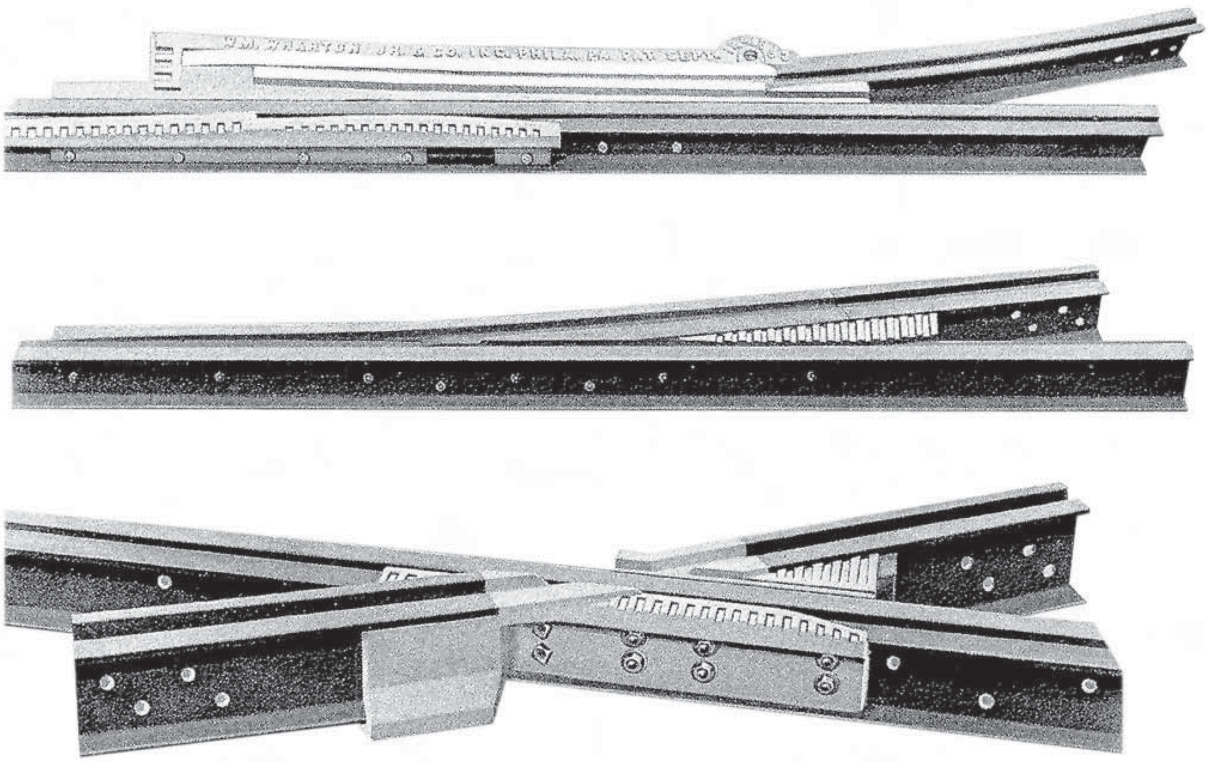


Figure 4 - The Wharton Unbroken Main Line Construction, one of the earliest lift-over crossings, together with its modern equivalent. The latter indicates some wear of the main rail head

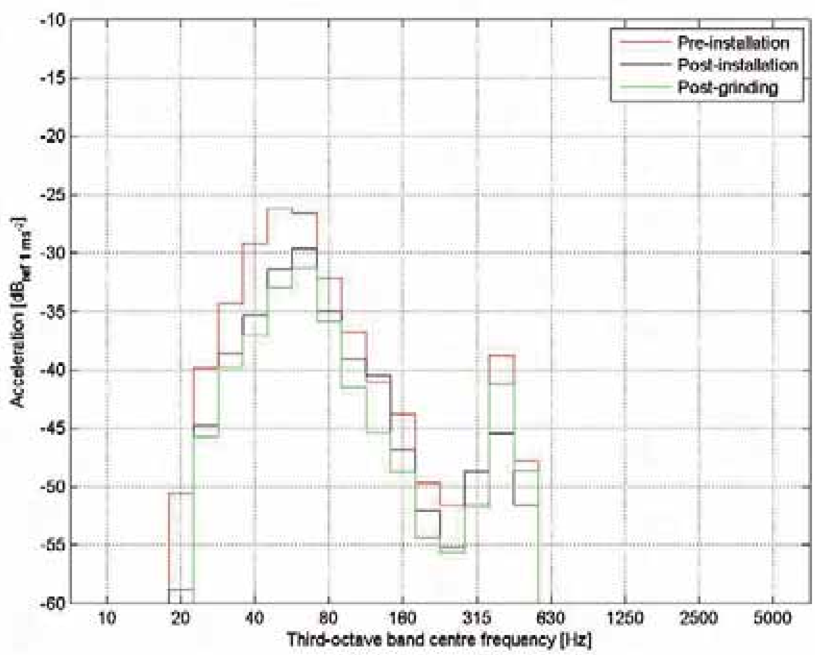


Figure 5 - Mean acceleration spectra measured adjacent the crossing before and after the installation of the new hardware, and after the subsequent rail grinding, due to southbound trams travelling on the crossover at 20kph

Knowing the mean speed of the trams as they traversed the crossover, and the spacing of their 6 axles, it was possible to calculate the times at which individual axles passed any one location. These times were then used to establish which transients in the vibration time-histories were due to a common feature of the track. This enabled the most significant vibration triggers to be identified and ranked according to the associated levels of peak ground acceleration. The most significant triggers were all associated with the rail crossings and switches of the crossover.

Design of remedial solution

If practicable, any remedial measures should address a noise and vibration problem at its source. In this case, the measurements presented provide clear evidence that the primary cause of the re-radiated noise in the concert hall was wheel-rail impacts at the crossings and switches of the crossover.

Following a review of various options, including relocating the crossover and providing a new floating-slab vibration isolation system, a new track form known as a lift-over crossing was selected as the basis of the remedial solution. Lift-over crossings are not a new concept.

One of the earliest examples was known as the Wharton Unbroken Main Line Construction, patented in 1893 by William Wharton, Jr., & Co., Inc. in the United States. Figure 4 illustrates the design principle, together with its modern equivalent. Wharton described the crossing as follows.

Where a switch is needed, but only occasionally used, it gives most decided advantages. The main track is entirely smooth and unbroken, and when the curve is not used there is no wear and tear on the switch.

The switch is provided with the peculiarly shaped tongue made of manganese steel. When set for the curve its inclined end raises the car wheel over the head of the main rail, the guard on the tongue deflects and guides the wheel into the curve.

Wharton's description of the crossing's benefits still applies today. In the main direction, the crossing behaves effectively as plain line, with no wheel-rail impacts. In the turnout direction, a ramp within the groove of the rail transfers the running contact of the wheel from the tread to the flange, which then rides over the head of the main rail. The disadvantage of this design, as suggested in Figure 4, is that there is the potential for excessive wear of the main rail head and the development of an alternative vibration trigger. However, where the turnout is used infrequently, the potential for vibration reduction is clear.

As an 'emergency crossover', the crossover discussed here is indeed used infrequently in the turnout direction. It was therefore considered to be an ideal candidate for a modern lift-over crossing, which, in replicating plain-line running in the main direction, has the potential to realise the anticipated noise reductions outlined. The switches of the original crossing also acted as significant vibration triggers. New close-tolerance switches were therefore also selected as part of the replacement crossover.

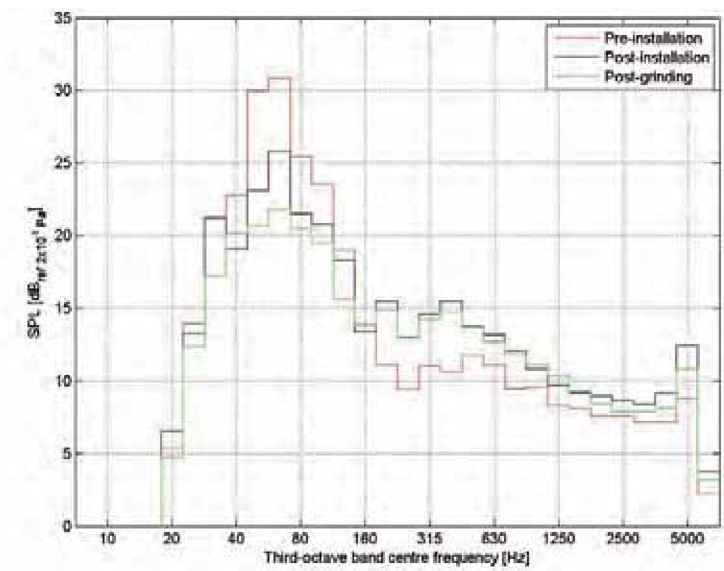


Figure 6 - Mean A-weighted noise spectra calculated from the stage measurement location data before and after the installation of the new hardware, and after the subsequent rail grinding, due to southbound trams travelling on the crossover at 20kph

Table 2 - Summary of mean peak noise levels ($L_{Amax,slow}$) measured before and after the installation of the new crossover

Tram pass-by condition	Peak level [dB(A)]	
	Stage	Stalls
S/B @ 20 kph – pre-installation	35.9	33.4
S/B @ 20 kph – post-installation	32.3	30.7
S/B @ 20 kph – post-grinding	29.9	28.5
Overall reduction	6.0	4.9
N/B @ 20 kph – pre-installation	37.6	34.8
N/B @ 20 kph – post-installation	34.0	31.6
N/B @ 20 kph – post-grinding	31.3	29.0
Overall reduction	6.3	5.8

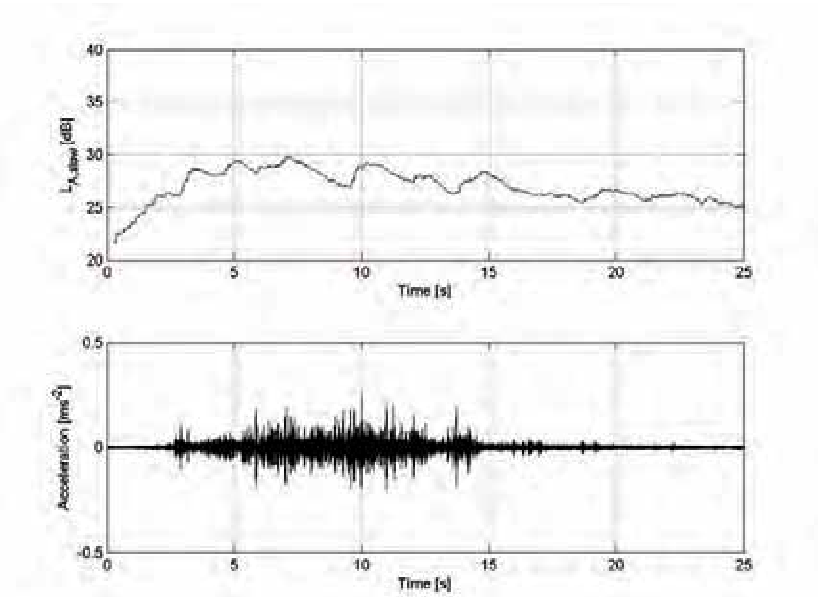


Figure 7 - Typical post-grinding variation in the A-weighted sound pressure level with time, as measured on the concert hall stage, together with the associated lineside vibration measured adjacent the crossover. Southbound tram at 20kph

Confirmation of performance

This section summarises the results of the final commissioning measurements. The same method was used as for the investigatory measurements, based on a series of simultaneous measurements of lineside vibration and re-radiated noise.

The measurements were made in three stages. Some additional 'pre-installation' measurements were made to get the best representation of the crossover before the installation of the new hardware.

These were followed by 'post-installation' measurements, approximately two months after the installation but before the final rail grinding.

The final measurements are referred to here as 'post-grinding'; they follow the completion of the remedial work in so far as noise and vibration control associated with the crossover itself is concerned. Ongoing work concerns the development of an optimum tram wheel profile, which is part of a longer-term study.

The final post-grinding dataset comprises 18 southbound and 11 northbound trams at a speed of 20kph ±10%. The results presented here focus on southbound tram pass-bys.

Summary of vibration results

Figure 5 summarises the lineside vibration results in terms of the mean third-octave band spectra measured adjacent the crossing. The post-grinding results with those of the pre- and post-installation measurements are compared. The gradual improvement in vibration levels is clear, with significant reductions evident in all of the frequency bands of concern (25Hz to 160Hz). Reductions of between 3.7dB and 7.3dB are evident in the dominant frequency bands between 40Hz and 80Hz. Of these reductions, up to 3dB may be associated with the rail grinding.

Vibration levels in the 400Hz band are almost as significant now as they were prior to the installation. However, these higher-frequency vibrations are not transmitted effectively to the concert hall and are not a cause for concern with respect to the re-radiated noise.

Summary of noise results

Figure 6 plots the mean noise spectra measured on the concert hall stage. The final results are again compared with those from the pre- and post-installation measurements. Note that, during the pre-installation measurements, the air conditioning in the hall was switched off and this leads to lower background noise levels at mid-frequencies (above 250Hz), although it does not influence the range of concern.

The noise levels also exhibit a gradual improvement, with final reductions of between 2.5dB(A) and 9.3dB(A) evident in the dominant frequency bands between 40Hz and 80Hz. As with the lineside vibration, a significant proportion of the overall improvement is due to the final rail grinding, with up to 6dB(A) difference in the dominant frequency bands between the post-installation and post-grinding measurements.

As well as reducing in magnitude, the peak in the noise spectra has also become broader. This is likely to contribute to the reduced perceptibility of the tram noise above the background.

Table 2 summarises the overall noise results in terms of the mean levels calculated across the tram pass-bys within each group. The results indicate reductions of between 4.9dB(A) and 6.3dB(A) due to the installation of the new crossover, with the greatest reductions achieved on the stage. Of these reductions, up to 2.7dB(A) may be associated with the rail grinding.

These reductions are significant in that they correspond to a reduction in the power of the noise source of between 68% and 77%. This is reflected in the perceived noise levels in the concert hall, which are significantly lower subjectively than those measured prior to the works.

An overall appreciation of how noise and vibration levels have changed as a result of the new crossover may be gained by considering Figure 7. This plots some typical stage measurement results from southbound trams, along with the corresponding lineside vibration. The improvement is clear when comparing these results with those of Figure 2.

Conclusions

This paper has illustrated how a measurement-based approach has successfully diagnosed and guided the solution of an intrusive re-radiated noise problem within the auditorium of a UK concert hall. Simultaneous measurements of sound pressure time-histories and the associated ground-borne vibration, together with a comprehensive track survey, provided clear evidence that the primary cause of the problem was wheel-rail impacts at the crossings and switches of the adjacent tramway crossover.

On the basis of these measurements, and following a review of various options, a replacement crossover was designed based on a new lift-over crossing and close-tolerance switches. This has resulted in reductions of between 4.9dB(A) and 6.3dB(A) in the overall noise levels in the concert hall, in line with the anticipated noise reductions attributed to replicating plain-line running. These results are reflected in the perceived noise levels, which are significantly lower subjectively than those measured prior to the works and are now likely to be regarded as unintrusive above the background noise by the majority of people.

It is worth noting that, of the reductions achieved, up to 2.7dB(A) may be associated with the final rail grinding. This result illustrates the importance of maintaining the condition of the wheel-rail interface. Ongoing work concerns the development of an optimum tram wheel profile, which is part of a longer-term study aimed at maintaining, and perhaps improving, the results reported here.

Acknowledgements

The author wishes to acknowledge Mr Arthur Durham of Atkins Rail, together with our colleagues in the partner organisations of this project for their contributions to this work.

References

1. Handbook of Noise and Vibration Control, M.J. Crocker, John Wiley and Sons, New York, 2007.
2. H. Kuppelwieser , A tool for predicting vibration and structure-borne noise emissions caused by railways. Journal of Sound and Vibration, Vol. 193, No. 1 (1996), pp. 261-267.
3. J. Lang, Ground-borne vibrations caused by trams, and control measures. Journal of Sound and Vibration, Vol. 120, No.2 (1988), pp. 407-412.
4. E. Vadillo et al., Subjective reaction to structurally radiated sound from underground railways: field results. Journal of Sound and Vibration, Vol. 193, No.1 (1996), pp. 65-74.
5. BS EN ISO 140-5, Measurement of sound insulation in buildings and of building elements – Part 5: Field measurements of airborne sound insulation of façade elements and facades, British Standards Organisation (1998).
6. MATLAB technical computing software, R2008a, The Mathworks.

Civil enforcement - threats and opportunities

Robin Cordell

Principal Engineer Enforcement

Intelligent Transport Systems

Atkins

Abstract

Traffic management regulations in the UK have been undergoing a significant transformation over the last four years. Principal among the changes is the transfer of certain types of traffic enforcement from the Criminal to the Civil domain. This brings both threats and opportunities to local highways authorities. The principal opportunity will be the ability to focus enforcement activities on parts of the network where poor traffic discipline is causing congestion. The main threat comes from the new equipment standard that is expected to result in a profusion of new enforcement equipment and methods leading to inappropriate choices by local highways authorities. This paper outlines the strategy for a fundamental change in the market in the UK that includes parallel development of policy, technology, supply base and regulation. Progress to date is outlined and future plans are described.

Introduction

With the introduction of the Traffic Management Act in 2004 (TMA2004), it was envisaged that local highways authorities in England and Wales would take responsibility for efficient management. Until now, the tools, in the form of the supporting Statutory Instruments (executive legislation that gives effect to framework laws such as the TMA2004), were not in place. In 2008, the first of these, covering parking, was implemented (coming into force in March 2009) and it is proposed that a second covering moving traffic contraventions, are still being considered by the Department for Transport (DfT). It is, however, unlikely that this will happen before the next UK general election.

Given these regulatory tools and appropriate equipment, the post of the "Traffic Manager" within the local highways authority should finally be able to address congestion on their networks. What's more, how the intervention is targeted should be determined solely on traffic management grounds and not dependent upon local policing priorities.

This opportunity to control road user discipline at locations determined by local congestion oriented targets could mean that congestion hotspots will be able to have 'softer' interventions before resorting to hard (and therefore expensive) blacktop modifications.

As yet, there is not widespread experience of these measures in England or Wales. They are likely to include the bus gate projects in a number of local authorities and the Transport for London (TfL) Digital Traffic Enforcement System (DTES) project.

To sound a note of caution, there is a significant risk associated with this development, which if not handled carefully could seriously compromise both the evidential and public acceptability of these tools. The principal risk concerns the inexpert application of technology resulting in poorly presented cases discrediting not only the individual scheme but also the whole national roll out of the civil enforcement strategy.

This risk arises because the initiative concerns the introduction of new technology, operating under a new legislative framework with an equipment approval process that has not yet been tested legally. Then this change is being imposed on perhaps one of the most conservative of institutions (quite correctly in this instance) - the British justice system. To mitigate this, work is already underway, primarily in London, where the new technology is being tested under an existing legal framework. Shortly, a pilot of the new technology combined with the new equipment approval process will be commencing before a full roll out in late 2008/early 2009.

Helping the traffic manager

In the introduction, we concentrated on the issues relating to the technology and the legislation and their effect on the roll out of the new strategy. In this section, we will try to explain what effect these new measures will have on the role of the traffic manager. Fundamentally, it will leave them free to focus on outcomes.

Department for Transport 'guidance' is that enforcement should not be seen as a revenue stream by local highways authorities. So in this context, achieving compliance (of road users with the posted regulations) is the issue. Road users have to be able to see (even if indirectly) some benefit to themselves of the enforcement that is taking place.

In the ongoing work being carried out by TfL and a number of London Boroughs under their own specific legislation, the effects of enforcement have been generally positive with some areas of concern. In particular with bus lanes, contravention rates fell by 90% after just 6 months of enforcement operations (Patrick Troy and Jim Lewis: ITS-UK Enforcement Interest Group annual conference 2005). With Yellow Box Junctions, contravention rates also fell, but by a smaller amount, after a similar enforcement duration (Steve Smith: Association of London Government Enforcement Taskforce Conference 2006).

Unfortunately with Parking, there seems to be no clear trend as a result of enforcement with anecdotal evidence indicating that whilst some classes of road users become more compliant, other classes become significantly less compliant. This might mean that an alternative treatment (other than the financial only penalty) may be needed such as vehicle removal / immobilisation for persistent offenders.

Undoubtedly, this new initiative will help the traffic manager to keep major arterial routes moving. However, to ensure long term benefit, public acceptance will be a key issue in determining what tools to use, where to use them and how they should be applied to the specific environment that the Traffic Manager is responsible for.

The financial challenge

As this is a very new market with only a small number of early adopters, the barrier to entry is still very high. In particular, the up front cost of the back office is in the $500k - $1m range because they are currently bespoke modifications to existing speed and red light enforcement packages.

For a typical local highways authority, with an initial need for 5 to 10 enforcement cameras, this is just too significant an up-front investment. If, instead of each local authority having their own back offices, one local authority were to buy the back office and 'rent' this out as a service to their neighbours, then it may be possible. It must be borne in mind, however, that this still requires one of the local authorities to find what is a very significant sum of money at a time when capital budgets are under more pressure than ever.

If this cannot be easily funded in the public sector, can the private sector help? Well the answer here is a definite maybe! This is clearly an opportunity where a reasonable financial model can be generated for a supplier to offer a lease back type service. However, as mentioned earlier in this paper, there is still significant risk associated with both the judicial acceptability of the technology and processes and the public reaction to enforcement of traffic violations that are currently commonplace and only enforced if you happen to contravene in front of a police officer. If the public is not 'on side' and a concerted evasion campaign is waged by road users, then the financial case begins to look very weak.

The bottom line, is that this particular set of tools are expensive and unless the local highways authority is responsible for a large and strategically/politically important road network (e.g. London, Birmingham and Manchester), it is unlikely that they will buy the full suite. Fundamentally, to achieve road user compliance with traffic regulations costs money and for early adopters, that cost can be substantial. Some of the individual tools will soon be available for a much more reasonable cost and most local authorities will find a measure of benefit in using that limited subset.

Getting it all to work

Simply buying all of the tools (even if the local highways authority had the money) is not the end of the problems. Before any enforcement system can be used in the UK it has to be certified as fit for use by either the Secretary of State for the Home Office (under the existing criminal regime) or the Secretary of State for Transport (for the new civil enforcement regime). In the case of the criminal regime, this involves demonstrating that the system will operate to provide proof that a contravention has taken place beyond all reasonable doubt. This is very onerous requirement and thus requires significant effort to be expended by both the system stakeholders and Secretary of State's representatives. Recent changes have meant that policing traffic contraventions that have little impact on road user safety are now a very low priority for the Home Office and it is unlikely that any new systems will be approve by this route. That leaves the civil enforcement regime as the likely method to be employed by local highways authorities and manufacturer to get systems and components approved. Unfortunately, this is currently unproven and the amount of input necessary to achieve certification is unknown. It is also likely to require expert assistance to prepare the necessary documentation expertise that is likely to be available within the manufacturing community but is unlikely to be present to any degree within the local highways authority community.

Once the systems are in place and certified ready to go, the final barrier to successful enforcement is the need to ensure that the appropriate Traffic Regulation Orders (legal documents that define the restrictions on road use at a particular location) are in place, that signage is not only correct but also in accordance with the Traffic Regulation Order (TRO) and finally that any road markings are correct and in accordance with the TRO. Only if all three elements in this ladder are correct in full, can enforcement take place at the desired location. It is worth pointing out at here that there are many groups on the Internet that offer much practical advice to motorists sent a penalty charge notice regarding how to appeal and what to look for in signs, lines and orders that might render them invalid.

End note

Since this paper was written, a number of civil traffic enforcement schemes have been implemented and a great deal of experience in the processing of certification applications has been gained. The current certification guidelines have been found to be satisfactory for all applications submitted to date but work continues to review the guidelines to ensure that they remain applicable to new technology. The author of this paper has been retained by the Department for Transport as an expert advisor in the drafting of the next issue of the guidance.

Copies of the guidance and other enforcement resources are available on the DfT TMA2004 portal at

http://www.dft.gov.uk/pgr/roads/tpm/tmaportal/tmafeatures/tmapart6.

A flexible approach to motorway control

Keith McCabe

Principal Consultant

Intelligent Transport Systems

Atkins

Andy Riley

Senior Engineer

Intelligent Transport Systems

Atkins

Abstract

This paper describes recent work undertaken by the UK Highways Agency to consolidate the experience gained from the implementation of Controlled Motorway and Ramp Metering Schemes, in addition to the innovative Active Traffic Management Pilot Scheme. A flexible approach to motorway control is described that can address a particular problem at a specific location at a given time via combined and coherent use of available traffic control measures based on Intelligent Transport Systems (ITS). The form of control will be based on the available physical infrastructure (e.g. gantry or post mounted signals) used in different ways at different times. The traffic problem may change over time, which requires flexibility in the control algorithms to cope with changing circumstances.

Introduction

The aim of this paper is to provide a summary of the progress on the development of variable speed limits in the UK. The paper provides a short review of the results from the Controlled Motorways (CM) pilot scheme on M25 and considers proposals for possible enhancements. Results of a theoretical study into combining variable speed limits with ramp metering are also presented.

There are three main reasons to control motorway traffic using mandatory or advisory variable speed limits and/or ramp metering, namely:

- To improve efficiency – reducing congestion and journey times
- To improve safety – reduce the number and severity of accidents
- To produce environmental benefits – reduce emissions.

A positive result in all three areas is the aim, however, under particular circumstances it may be sensible to prioritise one form of benefit over another (e.g. a reduction in speed to reduce accidents and emissions may result in increased journey times). The aim of this paper is to discuss the relationship between particular control interventions and the impact on efficiency, safety and environment.

Controlled Motorways on M25

The Controlled Motorways pilot scheme on the M25 began in 1995 and a business case for this system was completed in 2002[1] The original main aims of the scheme were to provide a smooth traffic flow, to improve journey times, journey time reliability and lane utilisation and to reduce the incidence of stop-start driving and the stress of driving. Subsequent notable achievements also included the development of proven technology and the reduction in environmental problems (i.e. noise and pollution).

During the monitoring of CM there was an improvement in journey times on the clockwise carriageway in the morning peak period. The observed behaviour was attributed to drivers becoming familiar with the CM and driving more smoothly to prevent flow breakdown. Other areas of the M25 did not show as much improvement in journey time performance because congestion had increased, which was attributed to yearly increases in flow levels (more road users).

The number of shockwaves decreased between 1995 and 2002, with a reduction from a typical seven shockwaves per morning rush hour down to a typical five. This observed behaviour could be attributed to the smooth driving behaviour or improvements to the control system over this period. It contributes to the improvements in journey time and queuing despite the increased throughput.

Evidence of safety improvements were demonstrated by the studies of injury accidents. The variable speed limit system on the M25, known as controlled motorways, has resulted in steadier and less stressful journeys and thus reduced the number and severity of accidents, "injury" accidents by 10% and "damage only" accidents by 30%.

Emissions have decreased overall by between 2% and 8%. The weekday traffic noise adjacent to the scheme has reduced by 0.7 decibels.

Lane utilisation and headway distribution had been improved. Lane utilisation became more balanced making better use of the road space, with a reduction in the number of very short headways (pre 1995).

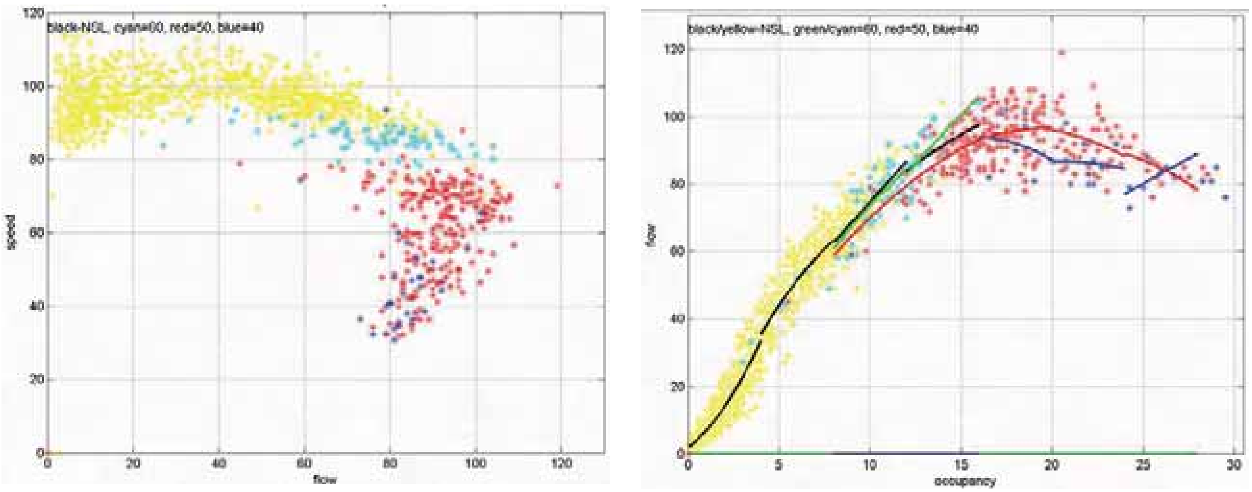


Figure 1 - M42 study mandatory speed limits - speed flow diagram and flow occupancy diagrams

The aim to reduce the stress of driving appears to have been achieved, with the primary road users indicating a positive response to the CM scheme. The driver opinion survey carried out in the first year of operation indicated that over half the respondents noticed an overall improvement; two thirds would like to see the scheme extended to other stretches of motorway. The survey indicated very clearly that the automatic speed cameras were essential to ensure drivers' compliance with speed limits.

Controlled Motorway on M42

A study was undertaken to investigate the effect of mandatory Variable Speed Limits (VSL) on motorway throughput and traffic flow efficiency using traffic data from the M42 motorway Active Traffic Management section (Junctions 3A to 7) as well as to determine possible proposals for improvement of the VSL control algorithm used on M42 as well as in other UK motorways. Speed limits are activated/modified/de-activated when flow and/or speed measurements cross pre-set thresholds - full details can be found in (2).

A mandatory 3 lane VSL has been in operation in the aforementioned stretch of M42 since late 2005. Threshold values have been fine tuned over a period since the mandatory speed limits went operational, to ensure safe operation and consistent signals to motorists. The major tuning activities were completed in January 2006.

Traffic data (i.e. flow/occupancy/ speed measurements as well as speed limit records) were available for this study from two periods: 2005 period traffic data corresponding to the VSL control algorithm application before the major fine tuning and 2006 period traffic data when the control algorithm was operating using the fine-tuned threshold values.

Validation of the traffic data was performed to remove faulty data or data that corresponded to low traffic demand, incidents of significant time duration. Visual inspection as well as curve fitting methods were used for the purposes of the evaluation analysis.

The signals were activated significantly less frequently during the period before the major fine tuning, hence the data set contained many non-VSL data at dense and critical traffic conditions than those included in the 2006 period.

A main focus of this work was on verifying the impact on the shape of the flow-occupancy diagram, see Figure 1, (or, equivalently, of the speed-flow diagram); more specifically VSL are thought to:

(i) Lower the slope of the flow-occupancy diagram at under critical occupancies

(ii) Shift the critical occupancy (at which maximum flow occurs) to higher values and possibly increase the flow capacity.

Notice that, if (ii) is correct, then VSL control (if properly operated) can have the effect of allowing a higher throughput at critical or slightly overcritical traffic conditions as compared to the no control case, even if the VSL affected flow capacity does not actually increase. On the other hand, if (i) is correct, then VSL could increase journey times for less congested traffic conditions if applied with less appropriate threshold values.

Main findings

Speed limits - when applied at less congested occupancies - decrease the slope of the flow-occupancy diagram (see (i) above). Moreover, the smaller the imposed speed limit, the larger the decrease in the slope of the flow-occupancy diagram.

Importantly, the VSL-affected flow-occupancy curve indeed crosses (at least for some VSL) the non-VSL curve, shifting the critical occupancy to higher values in the flow-occupancy diagram (see (ii) above). The major cross points were found to lie on or slightly beyond the non-VSL critical occupancy.

Regarding the potential increase of flow capacity, the data analysis was rather inconclusive, as at some locations a slight increase is indeed visible while at other locations no increase can be observed.

On "rainy" days, the non-VSL flow capacity and the critical speed are reduced by some 10%.

The same quantities were observed to change quite substantially from day to day (even for the same location) without any obvious reason (stochastic effects). In contrast, the non-VSL critical occupancy was found to be rather insensitive to weather conditions and stochastic effects.

The speed limit of 50mph was found to be the main contributor to modified aggregate traffic flow behaviour that could be exploited towards more efficient traffic flow. A speed limit of 60mph was found to have a rather moderate impact; while a speed limit of 40mph is used at high occupancies in the interest of traffic safety rather than traffic flow efficiency.

In the data period before the major fine tuning, the speed limits were activated either "too early" or "too late", leading to cases where either speed limits were imposed although traffic was under critical; or speed limits were imposed after congestion had occurred, hence having little effect on congestion avoidance. In contrast, the behaviour of the VSL control algorithm after the major fine tuning is more robust, preserving similar behaviour at different days.

Algorithm issues

The algorithm currently used for mandatory VSL is based on a rule-based method designed to harmonise traffic speeds and to reduce the severity of congestion. The control system responds to one minute traffic speed/flow data measured by the MIDAS loop detectors. Speed limits are reduced if the measured flow goes above a pre-defined Rising Flow Threshold or the measured speed goes below a pre-defined Falling Speed Threshold. Likewise speed limits are increased if the measured flow goes below a pre-defined Falling Flow Threshold and the measured speed goes above a pre-defined Rising Speed Threshold. The speed threshold is essentially used to ensure that the signals remain set when flow drop in "stop & go" driving conditions. On this basis the algorithm is based on a practical method.

Whether such an algorithm would be successful in preventing traffic breakdown largely depends on the methodology used to determine Rising/Falling Speed and Flow Thresholds.

The methodology originally used for the determination of speed/flow thresholds for the M25 CM was to apply speed limits at times when average traffic speeds were approaching these speeds anyway with a high likelihood of flow breakdown occurring. This was based on the assumption that the reduction in actual traffic speeds due to the speed limits is likely to be small; hence if there is an increase in delay to drivers, this is also going to be small. On this basis, the initial algorithm was based on a homogenisation approach.

The M25 CM has been subject to continuous monitoring and fine tuning over recent years. As a result, there have been a number of changes to the algorithm parameters which may have positively affected the impacts of the algorithm. The monitoring results suggest that the system has produced positive impacts on flow breakdown, journey time reliability, safety and air pollution but neutral or negative impacts on journey time and throughput.

The review of the algorithms indicated that limitations can be attributed to the current algorithm threshold setting constraints, location dependencies and general complexity of combined parameters. In addition, MIDAS capability is generally underused, but there are limitations at multiple MIDAS sites. There are also high speed limit settings in heavy congestion:

- The algorithm does not allow the use of separate thresholds for different time of day (e.g. am/pm), day of the week, different weather or lighting conditions, seasonal variations or bank holidays
- The thresholds apply to the total flow across the carriageway lanes - separate thresholds per lane are not available.

The algorithm parameters have to reflect a wide range of factors involved in forming different traffic situations e.g. road geometry (i.e. number of carriageway lanes, gradient, distance between the junctions) and traffic characteristics (e.g. traffic composition, lane utilisation, lane changing frequency, proportion of merging and diverging traffic).

It is very difficult to reflect the combined effects all these factors in a limited number of control parameters (e.g. speed/flow thresholds). As the tuning of the system is done manually, it is a great challenge to correlate all thresholds and algorithm parameters in this way.

No provisions are currently made for the use of lower than 40mph speed limits for incident and congestion management; severely congested areas are largely uncontrolled, therefore. Customers often get frustrated when speed limits of 40mph or higher are posted when they are actually doing much lower speeds, despite the fact the speed limits represent the maximum speed at which they should travel. Any improvement in this respect may contribute to improved driver acceptance of the system.

Although the VSL control algorithm behaves more robustly and efficiently after the major fine-tuning, it still faces some limitations mainly due to the sensitivity of the appropriate thresholds to weather and stochastic traffic conditions. Thus, the VSL control algorithm in several cases does not impose the appropriate speed limits although the traffic state has entered the unstable overcritical occupancy-flow region, thus leading to strong flow variations.

The remaining drawbacks of the VSL algorithm are mainly due to the use of absolute threshold values of flows or speeds for VSL activation (e.g. VSL set to 60mph if measured flow higher than a flow threshold); while in reality the appropriate critical thresholds may change from day-to-day even under the same weather and light conditions.

Possible modifications to the algorithm

There is significant potential for the improvement of the algorithms in the longer term by changing the MIDAS Site Data to allow the use of different thresholds for different weather and lighting conditions. This would also require linking the MIDAS Subsystem to the Meteorological (MET) Subsystem to send the alerts related to the changes in weather/lighting conditions to the MIDAS Outstations. For this purpose, the MIDAS Site Data should also be modified to allow for using different thresholds for different weather/lighting conditions.

In view of changing real absolute thresholds, a more robust and efficient algorithm might result if the decisions on VSL-activation are based on estimated slopes in the flow-occupancy diagram. In other words, while, e.g. the motorway's flow capacity value may vary from day to day, the fact that flow capacity values occur at zero slope does not change.

An adaptive algorithm - based on the algorithm for critical occupancy estimation developed in a previous study[3] - has been preliminarily tested for slope estimation using the M42 traffic data. The algorithm parameters were chosen to be the same for all motorway locations, and the preliminary tests indicate that the adaptive algorithm may be more robust than the current VSL control algorithm even after the major fine tuning. More precisely, the adaptive algorithm seems to avoid situations where speed limits are imposed "too early" or "too late". Quite importantly, the adaptive algorithm is expected to avoid the time and effort consuming process of fine tuning the location dependent VSL control algorithm thresholds.

As a longer term development, the Highways Agency is looking towards a network oriented traffic control model for the management of incidents and congestion on a motorway corridor or network. For control systems, this would offer features such as coordination, prediction, integration with other systems, control by other parameters (e.g. environmental) and macro level area traffic management. This model is called the Network ATM Supervisory Subsystem (NASS)[4].

Network oriented traffic control has several advantages compared to local control. For example, the capacity of a route is the same as the capacity of its most constrained point. Solving a single local traffic problem can only result, therefore, in the vehicles travelling faster to downstream congestion. The same number of vehicles have to pass the downstream bottleneck (with a given capacity). In such a case, the average travel time on the network level will be determined by the capacity of the downstream congestion point. A network oriented approach would take this into account and, if possible, would resolve both congested areas.

Using network oriented traffic control, it would be possible to integrate the control system with other motorway traffic management measures such as Ramp Metering (RM), Mainline Metering, Strategic Traffic Management, etc. The synergistic effects of the integrated systems would increase the overall benefits of the HA Traffic Management Systems.

Multiple objectives could be achieved using network oriented traffic control (for example in pollution and noise control as well as for incident and congestion management).

The inflow to a congested area could also be limited to a level that is less than outflow of the area using network oriented traffic control. The overall delays on the network could be maintained below the non-controlled situation. This is a fast and effective approach to the alleviation of severe congestion and to prevent secondary breakdowns.

Theoretical design for combining CM with RM

An examination of the theoretical impact of the two systems was carried out to determine the likely impacts of operating CM and RM on the same section of motorway. Research into two previously conducted studies was considered. These studies simulated the use of control strategies of Model Predictive Control (MPC)[5] and Stackelberg Game Theory[6]. Whilst it is not possible to directly compare the two strategies against each other (as they were tested under different conditions), the review showed definite theoretical benefits in terms of journey time.

Combining the systems was shown to be better than using either RM or CM singularly. Hence it is considered that linking the control systems could be beneficial.

Complementary RM and CM

The existing operation of RM and CM could act in a complementary way when deployed on the same section of motorway flow breakdown is delayed and platoon management can be achieved.

In the period when the flow is reaching capacity the operation of RM delays flow breakdown allowing the upstream speeds to remain at 50mph for longer. The benefits of increased throughput due to RM and higher harmonised upstream speeds due to CM are potentially enhanced by operating both control measures on the same section of motorway.

In the period after flow breakdown RM improves the merge behaviour by producing small platoons of traffic, at the same time, CM protects the back of any queue that has formed (setting 40mph using HIOCC), the resultant impact on safety of both control measures operating on the same section has the potential to provide enhanced safety benefits.

Interaction of RM and CM

There are two situations where the interaction between the two control systems on the same section of motorway needs to be carefully considered:

(1) When the speeds are rising as a result of RM holding back the slip road traffic, there is the potential to interact with a purely flow based signal setting (60 and 50mph) on CM. The interaction between the two systems in this area of operation using the existing control algorithms requires careful consideration as there is the potential either to delay the flow recovery process or have inappropriate speeds for the traffic conditions.

(2) The potential impact of setting mandatory 60/50/40mph speed limits on the value of critical occupancy also requires attention. Difficulties may arise from different values of critical occupancy when different speed limits are set. Potentially this could lead to inefficient operation of RM.

Support of CM by RM

RM could assist the current operation of CM when the CM system is setting 60/50mph in congested and heavily congested periods. RM could potentially control the flow to complement the setting of 50mph and 60mph signals on the main line. This could result in an improved speed harmonisation allowing a higher level of service (higher main line speed/ shorter journey time) to be achieved for longer, with only short delays to slip road traffic. This mode of operation would require sufficient acceleration distance from the stop line to the merge to enable the vehicle to reach merge speeds of between 50 and 60mph. In a number of circumstances this may require works to extend the length of the slip road. In such a circumstance a site specific business case would be required prior to undertaking the work.

Support of RM by CM

CM could assist the operation of RM when the slip road queue is reaching the slip road storage capacity. In theory setting a 30mph signal upstream of the merge area could reduce the mainline flow to a level that would allow more slip road traffic to join without exceeding the downstream capacity. This is currently a modelled theoretical concept that in theory produces benefits in terms of allowing the downstream flow to run at capacity for longer without the bottleneck causing flow breakdown. When this has been modelled a significant reduction in the number of main line shockwaves has been noted, with resulting accident and journey time reliability benefits.

Overview of combined RM/CM operation

Overall the benefits of CM and RM that were demonstrated in the pilots are achievable when the two control systems operate on the same section of motorway. The areas of concern that needs to be examined are during flow recovery and when inappropriate speeds are set on the main line. The first of these problems may be overcome with the enhancements to CM because the problem is most likely to occur with a CM algorithm based purely on flow level settings. The second could be addressed through the use of an adaptive system for critical occupancy for the RM (possibly AD-ALINEA). Careful selection of a test site may enable an early demonstration of the overall benefits of the two systems operating on the same section of road.

The RM and CM could be enhanced to improve combined operation. RM could improve the operation of CM at 60 and 50mph. This would require an enhancement to the RM control system and careful site selection (to allow vehicles to achieve merge speeds of greater than 50mph). A link from CM to RM to let the RM system know when the 60mph and 50mph limits are in operation would be a useful feature of this system. A carefully selected trial site where the benefits of the enhancement could be isolated would be required prior to a wider implementation of this enhancement. CM could improve RM operation by managing the flow at capacity and after flow breakdown by setting 30mph limits on gantry signals. This would require a fully integrated control system that could be a subset of the NASS control system using the principles in Stackleberg Game or Model Predictive Control systems. This type of system could be developed to "lock in the benefits of motorway widening" and possibly form part of a broader access management/ motorway control system.

Conclusions

Studies into the operation of controlled motorways have shown that potentially benefits can be achieved with reduced congestion, reduced accidents and reduced emissions. However, it should be noted that the current methodology of using speed and flow thresholds requires substantial fine tuning to achieve efficient results. A system that automatically fine tunes parameter settings potentially gives long term efficiency gains.

A fully integrated system using variable speed limits linked to ramp metering/access management potentially can manage a motorway to address a range of local priorities (congestion, safety or environmental). Reducing speeds to improve safety and produce environmental benefits can potentially increase journey times. However, the safety and environmental benefits can be enhanced by linking to ramp metering/access management to maintain constant speeds and improve journey time reliability.

Overall the principle of using variable speed limits and ramp metering to improve efficiency, safety and environmental impacts has been theoretically proven. To maximise the benefits of implementations the automation of parameter setting should result in more consistent outcomes.

Acknowledgements

The work described in this paper is sponsored by the Highways Agency. The views expressed in this paper are those of the authors and do not necessarily reflect those of the Highway Agency.

The Technical University of Crete Dynamic Systems and Simulation Laboratory undertook the analysis of the data on variable speed limits on the M42 on behalf of the Highways Agency.

References

1. Rees, T; Harbord, B; Dixon, C; Abou-Rahme, N. Speed Control and Incident Detection on the M25 Controlled Motorway (Summary of Results 1995-2002), Report, Transport Research Laboratory, PPR033.
2. Markos Papageorgiou, Elias Kosmatopoulos, Matt Protopapas, Ioannis Papamichail (2006), "Evaluation of the effects of variable speed limits on motorway traffic using M42 traffic data, DSSL Technical University of Crete Internal Report 2005-25.
3. Kosmatopoulos, E., Papageorgiou, M., Monolis, D., Hayden, J., Higginson, R., McCabe, K., Rayman, N. (2006) Real Time estimation of critical occupancy for maximum motorway throughput, Transportation Research Record, in press
4. NASS fact sheets #1 - #3, Available from IPL Ltd.
5. Hegyi, A. (2004). Model Predictive Control for Integrating Traffic Control Measures. TRAIL Thesis Series T2004/2, The Netherlands, TRAIL Research School.
6. Zhenlong, L. (2005) Optimal Coordination of Variable Speed and Ramp Metering based on Stackelberg Game. 12th World Congress on Intelligent Transport Systems. Nov. 2005, San Francisco, USA.

Integrated urban traffic management and control strategies

Gary Hui

Engineer

Intelligent Transport Systems

Atkins

Abstract

The introduction of the Traffic Management Act 2004 (TMA2004) set out Government's vision to tackle congestion and disruption on the UK road network. The Act places a duty on local authorities to ensure the expeditious movement of traffic on their road networks and the interface with surrounding authorities.

In Essex, EssexITS (a joint partnership of Essex County Council, Atkins and Siemens), has taken a hands-on approach to assist the County Council in carrying out its responsibility under TMA2004. EssexITS has developed a number of Integrated Urban Traffic Management and Control (UTMC) strategies to manage Essex's road network, tackle congestion and proactively prevent or reduce the effect of congestion.

UTMC systems are designed to allow the different technologies to be used within single platform management systems, enabling communication and interaction of different applications and systems. EssexITS has deployed a number of UTMC systems in the last decade, together with the implementation of the UTMC Common Database and have developed innovative traffic control strategies using a combination of UTMC technologies.

This paper presents a case study on the practical use of various UTMC systems and integrated strategies to tackle congestion. It looks into a particular congestion hotspot in Chelmsford, the Army and Navy roundabout, on the major route linking Chelmsford town centre to the A12 and A130. The traffic flow across this roundabout is often disrupted causing delay to traffic going into and leaving Chelmsford town centre. The case study looks into how a combination of UTMC systems works together to deliver the most appropriate traffic control at this site.

Background of the site and the scheme

The site

The Army and Navy Roundabout is one of the most important interchanges in Chelmsford, it is situated south east of Chelmsford town centre see Figure 1. The roundabout connects the town centre with major destinations around Essex:

- The A1060 Parkway connects the roundabout with the town centre
- The A138 Chelmer Road is connected with the A12 junction 19 for Colchester, Ipswich and the coastline of Suffolk and Norfolk
- The A1114 Essex Yeomanry Way connects Chelmsford with the A414 to Maldon and the A130 to Basildon, Southend and Canvey Island. The A1114 also connects to junction 17 and 18 of the A12. Although not the natural route for traffic from/to the A12, it is often used as an alternative route

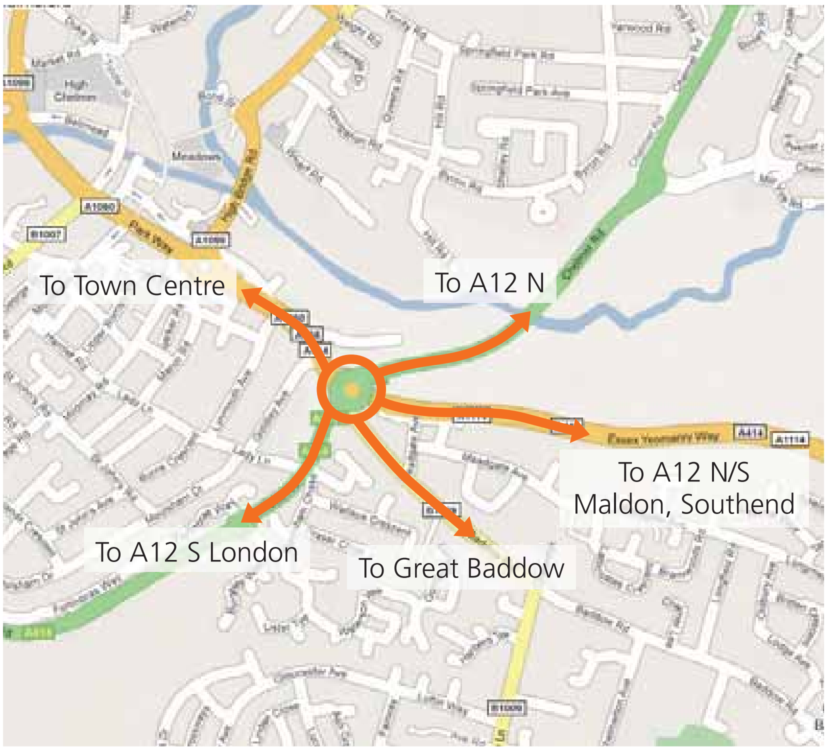


Figure 1 - The Army & Navy Roundabout

Figure 2 - Aerial photo of the Army and Navy Roundabout

- The A414 Princess Road links to the south of Chelmsford and A12 for Brentwood, M25 and London
- The B1009 Baddow Road connects to Great Baddow, a residential area near the outskirts of Chelmsford.

There is a single lane one-way flyover connecting Parkway with Essex Yeomanry Way, allowing a dedicated straight ahead movement above the roundabout. The one-way flyover operates in tidal mode, opening 'inbound' allowing commuters from Essex Yeomanry Way to travel into the town centre in the morning and 'outbound' in the afternoon and the evening.

Congestion hotspot

Due to the location of the Army and Navy roundabout peak hours are normally heavily congested. The morning peak traffic flows into the roundabout from Essex Yeomanry Way towards the town centre. Minor traffic flow from Chelmer Road has priority over traffic not using the flyover . This causes tailbacks and more than doubles the journey time for vehicles. The reverse happens during the evening peak hours.

The dominant traffic flow on Parkway is helped by the flyover which has been changed to 'outbound' for Essex Yeomanry Way and the dedicated left turn lane for Chelmer Road, however, it is still usually congestion during evening peaks. Any long tailback on Parkway will have a severe knock on effect on traffic flow in the town centre causing other junctions and roundabouts to exit block.

There is also a large flow of traffic on Essex Yeomanry Way towards the roundabout in the evening, and because the flyover is used by vehicles leaving Chelmsford in the afternoon and evening, Essex Yeomanry Way is also always congested in the evening.

Congestion scheme

The 'Essex Works' campaign has set amongst its many objectives, to tackle traffic congestion. As a result, EssexITS has initiated a number of congestion schemes, one of them being at the Army and Navy roundabout.

The objective of this scheme is to tackle congestion and proactively prevents traffic accessing Parkway and Essex Yeomanry Way, allowing more efficient flow of traffic from Essex Yeomanry Way, and preventing tailbacks on Parkway causing congestion to spread to other junctions and roundabout.

To achieve this, EssexITS have installed part-time traffic signal on Parkway circulatory and the Chelmer Road circulatory as a part of an integrated UTMC control strategy for this site see Figure 3.

Integrated UTMC strategy

The integrated UTMC strategy for the Army and Navy Roundabout was to be rolled out in stages:

- Stage 1 – Implementation of part-time traffic signals
- Stage 2 – Timetable switch on of part-time traffic signals
- Stage 3 – UTMC Integration (Manual dynamic switch on)
- Stage 4 – Automated UTMC strategy (Automated dynamic switch on)
- Stage 5 – Pollution Control Strategy

Stages 1 and 2 - part-time traffic signal and operations

The first stage of the roll out was to install traffic signals on the Parkway circulatory node and the Chelmer Road circulatory node of the roundabout, allowing a certain level of control of the roundabout. Both sets of signals operated in part-time mode.

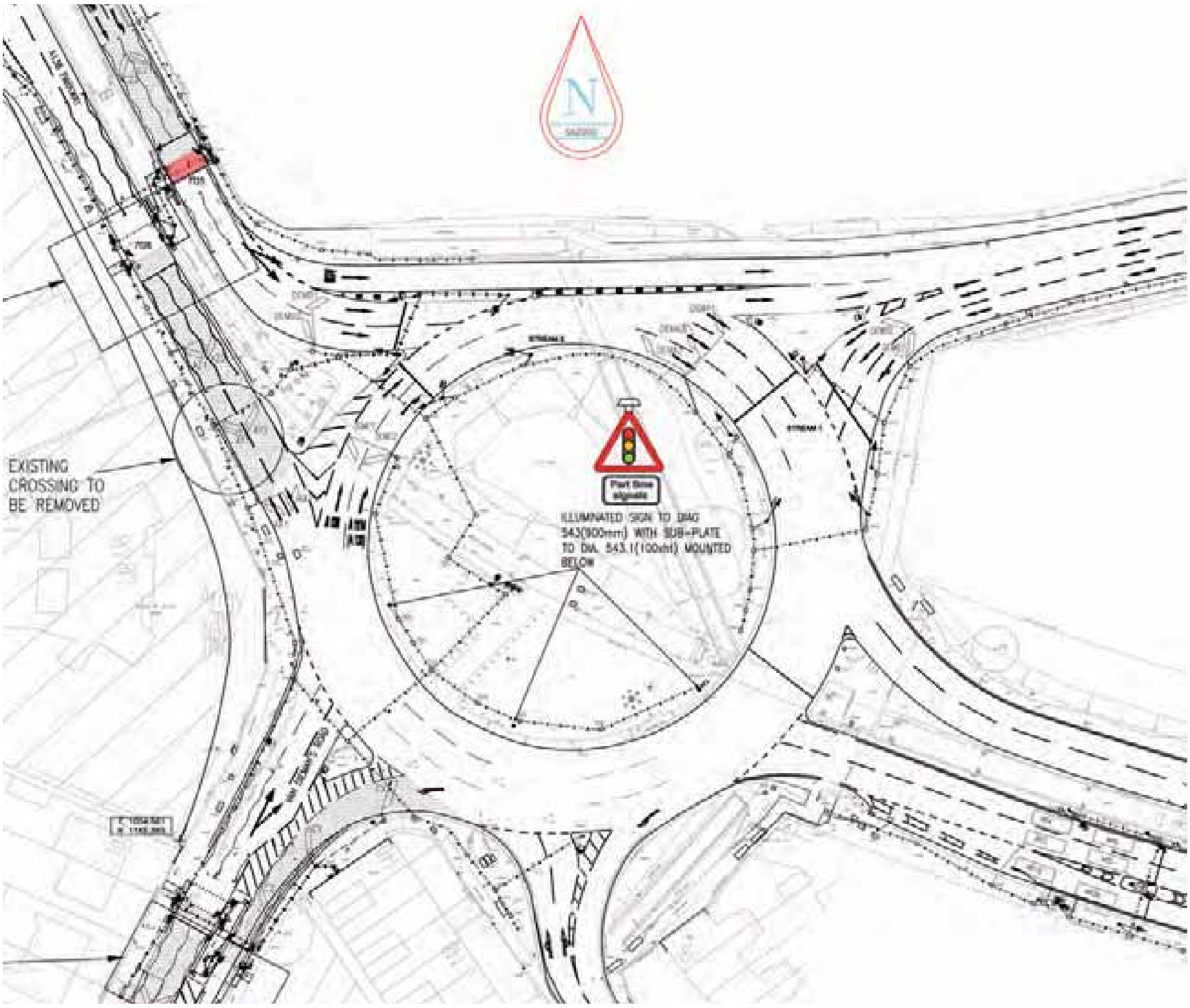


Figure 3 - Signal approval drawing for the Army and Navy congestion scheme and description of operation

The Parkway circulatory traffic signals will allow the clearance of Parkway traffic when congestion is severe. Although the operation of the traffic at a roundabout might introduce inefficiency to its operation, the part-time traffic signal can be switched on only when it has a beneficial effect to the traffic and its adjacent network.

The traffic signal on this node was only be switched on when congestion is severe on Parkway causing tailback into the town centre.

This requires an Urban Traffic Control (UTC) engineer to run a cast in UTC for the switch on of the signal when required.

The primary function of the Chelmer Road circulatory node traffic signals is to create opportunities for traffic on Essex Yeomanry Way to enter the roundabout during ‘inter-greens’.

The diagram in Figure 4 below demonstrates the operation.

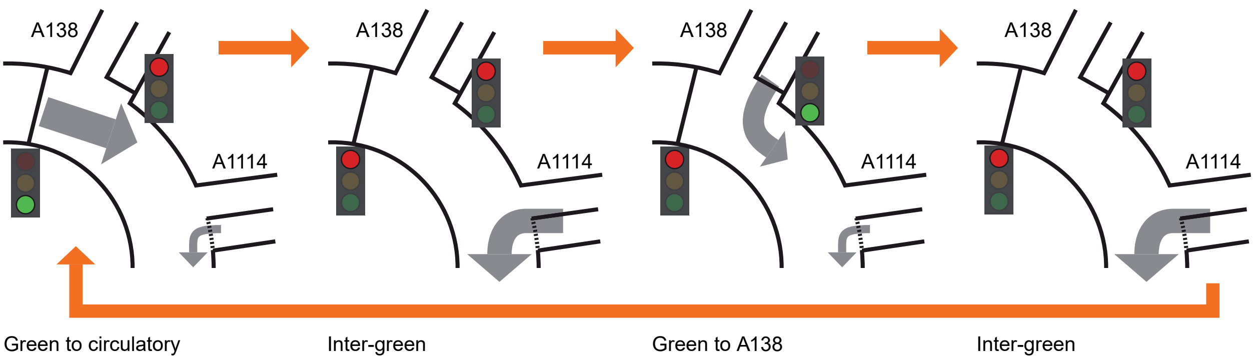


Figure 4 - Clearance of traffic on Essex Yeomanry Way during inter-green periods

Figure 5 - CCTV images showing effects of part-time signal operation

The traffic signal also has a consolidating effect on traffic flow on Chelmer Road, instead of ‘drip feeding’ traffic onto the circulatory, they are compacted and released as a solid platoon when the signal is green. This allows better flow from Essex Yeomanry Way onto the circulatory.

The Chelmer Road circulatory traffic signal node is timetabled to be switched on during the evening peak from 15:45 to 18:00 to ease congestion on Essex Yeomanry Way. The effects of the signal can be seen from the CCTV images in Figure 5 and journey time graph in Figure 6.

Stage 3 – UTMC Integration (manual dynamic switch)

UTMC systems and strategies

The introduction of the Siemens UTMC common database, “COMET”, which links up with several UTMC compatible systems opens a vast range of new applications and strategies for traffic control. The link up of COMET with various UTMC systems enables the interactions between systems and integration of control strategies. An illustration of the EssexITS UTMC setup and example of such UTMC strategies can be found in Figure 7.

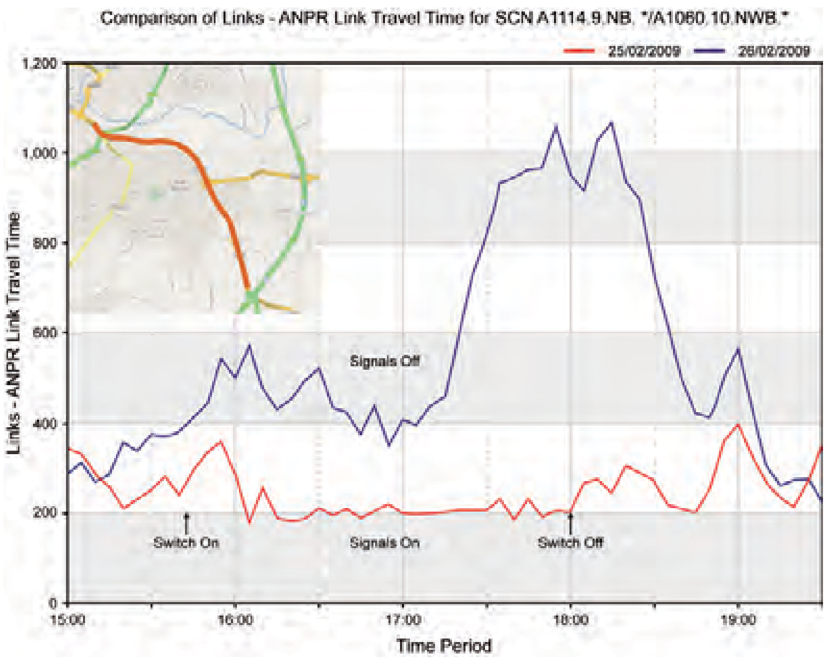


Figure 6 - Comparison of journey time on the Essex Yeomanry Way from north of junction with A12 to Parkway (signal on vs signal off)

UTMC strategy example 1: Galleys Roundabout part time signal switch on by video detection

The Galleys roundabout part time signal is designed to ease congestion on the A120 during the evening peak hours on weekdays and at noon during weekends or holidays. The part time signal uses an innovative automatic on/off switch. A video detection system (using digital video processing) was installed on the A120 approach approximately 250m from the Galleys roundabout to visually detect queuing traffic. This system, Traficon, then passes on the real-time data from site to COMET. A number of automation logics are defined in COMET for the on/off switch of these signals, when the real-time traffic data from Traficon meets any pre-defined conditions. When this happens, COMET interacts with UTC for the on/off switch of the part-time signals.

COMET will also interact with SieSpace (a variable message sign and car park guidance control system) to display messages on the variable message sign (VMS) informing drivers of the operation of the part time signals.

As traffic patterns change throughout the year the automation strategy ensures the signals are on, whenever required. Regular revalidation of the signal switch on/off times is no longer required and the signal operation is always responsive to the real-time conditions on site, including any exceptional condition such as incidents or accidents. The integration of three UTMC systems (UTC, Video detection and VMS) ensures a high level dynamic control of traffic which was not possible before.

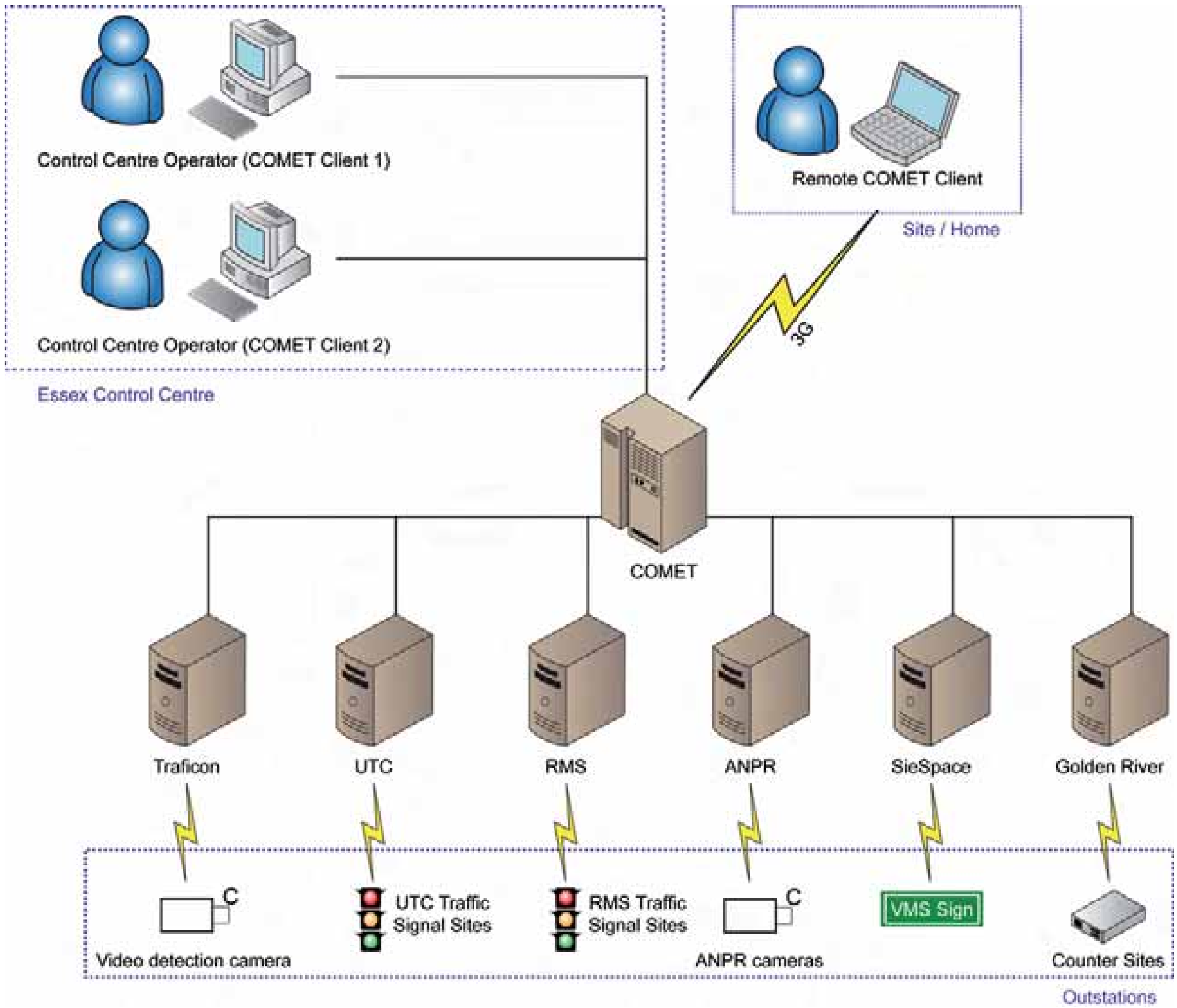


Figure 7 - Schematic diagram showing the UTMC network setup in Essex

UTMC strategy example 2: Waterhouse Lane SCOOT automation

Split Cycle Offset Optimisation Technique (SCOOT) is a centralised computer automated method of controlling traffic and is designed to implement synchronised control for multiple traffic signal sites to optimise traffic flow through a region. Under low traffic flow condition SCOOT can be unresponsive and drivers can experience delays in a SCOOT region.

The Waterhouse Lane SCOOT scheme automates the switching of control modes of traffic signals, allowing drivers to benefit from the most efficient method of control in the region at anytime.

Strategies are defined in COMET to react to flow data for the region from UTC, and when certain pre-defined conditions defined in COMET are met, COMET interacts with UTC and switches the mode of control of the traffic signal region from vehicle actuation (VA) to SCOOT or vice versa. The switching of control mode of traffic signal is traditionally done via timetables and the automation of such ensures the most efficient mode of control is select based on real-time traffic condition and will not require constant re-validation of timetable entries.

Manual Dynamic Switch On of Part-Time Signals

As discussed earlier, the use of part-time signals is only beneficial under certain traffic condition and is traditionally switched on/off via pre defined entries in the signal controller or UTC timetable.

As traffic patterns differ from day to day, reacting to temporary changes (incident/accidents) or permanent (road widening, new traffic signals etc.) changes to the road network, the operation of a fixed timetabled part-time signal is unresponsive and can be inefficient. It also required regular re-validation of the timetable to ensure the traffic signals operate at the optimal times.

For dynamic switch on operation the traffic signal is put under UTC (or Remote Monitoring System – RMS) control where it 'listens' to UTC commands for the switch on/off. Two 'casts' are setup in UTC, one that switches the signal on and the other that switches it off and these pre defined 'casts' can be directly actioned by an UTC engineer running the 'cast' command on the UTC terminal.

It is inappropriate for a UTC engineer to constantly monitor traffic flow and manage the switch on of the traffic signals or for any control centre operators to have direct engagement with the UTC system. This is where the COMET UTMC common database bridges the gap in providing an easy to use operator interface for various UTMC systems including UTC with pre-defined access levels see Figure 8. Corresponding strategies are defined in COMET for the switch on and off of the traffic signals, these can be action via the a symbol on the COMET map interface see Figure 9. When actioned, COMET will interact with UTC and action the corresponding "casts" in UTC for the switch on/off.

With the help of CCTV images, operators at Essex Traffic Control Centre (ETCC) can remotely switch on/off the signals, via COMET client terminals situated in the ETCC, on the Parkway circulatory node when traffic is queuing back into town and on the Chelmer Road circulatory node when congestion on Essex Yeomanry Way occurs.

Manual Dynamic Switch On with VMS control

The signal switch on/off strategies in COMET are also linked in with a VMS situated on the Parkway approach. A message is displayed informing drivers when the part time signals are in operation when the switch on strategy is actioned. When this happens, COMET also sends a pre defined command to SieSpace, another UTMC system that controls VMS, car park counters and car park guidance VMS. This strategy integrates the control of two completely different traffic systems and operates them via a single control strategy see Figure 10.

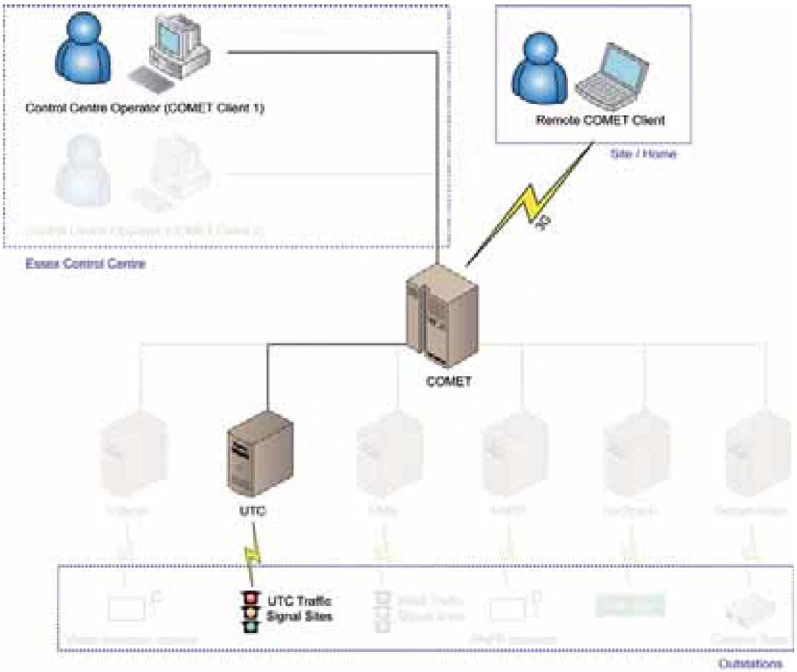


Figure 8 - COMET providing the interface to UTC for operators

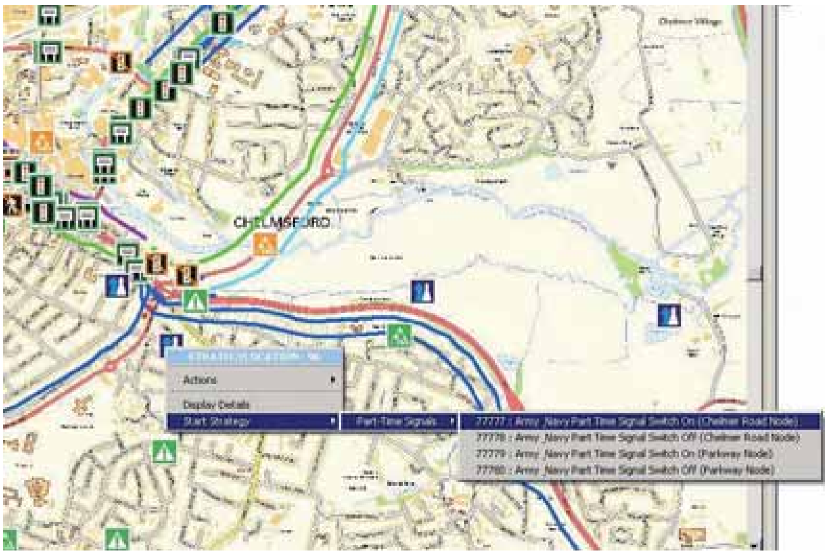


Figure 9 - COMET map interface signal switch on/off strategies

Stage 4 - Automated Dynamic Switch On

Images from CCTV can often be deceiving and manual strategies rely on control centre operators' manual interventions in a busy and multi-tasking environment. This is often affected by operator's perception and judgment which relies on attention, discipline and skill of individual operators, imprecise actioning of strategies and human errors are sometime unavoidable.

In order for precise and unbiased activation of UTMC strategies, EssexITS has taken an extra step to automate UTMC strategies.

Site condition based activation

During the installation of the traffic signal site, additional inductive loop detectors were installed on Parkway and Chelmer Road, these are linked to the traffic signal controller. These detectors do not have any direct input to the signal controller's operation and are only relayed to UTC. There are also existing inductive loop detectors on the Essex Yeomanry Way approach, these are connected into another set of traffic signals (Essex Yeomanry Way bus gate) and traffic flow data is also relayed to UTC.

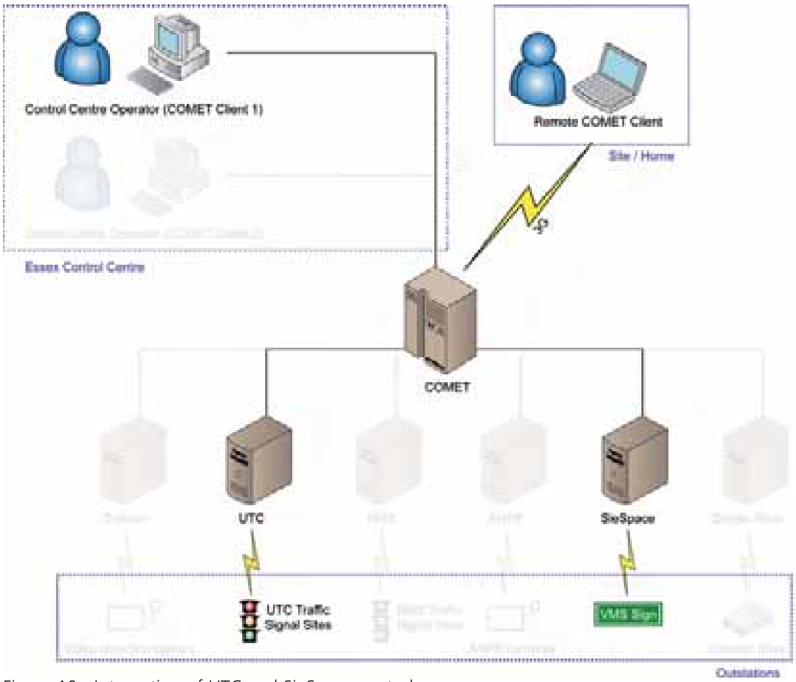


Figure 10 - Integration of UTC and SieSpace control

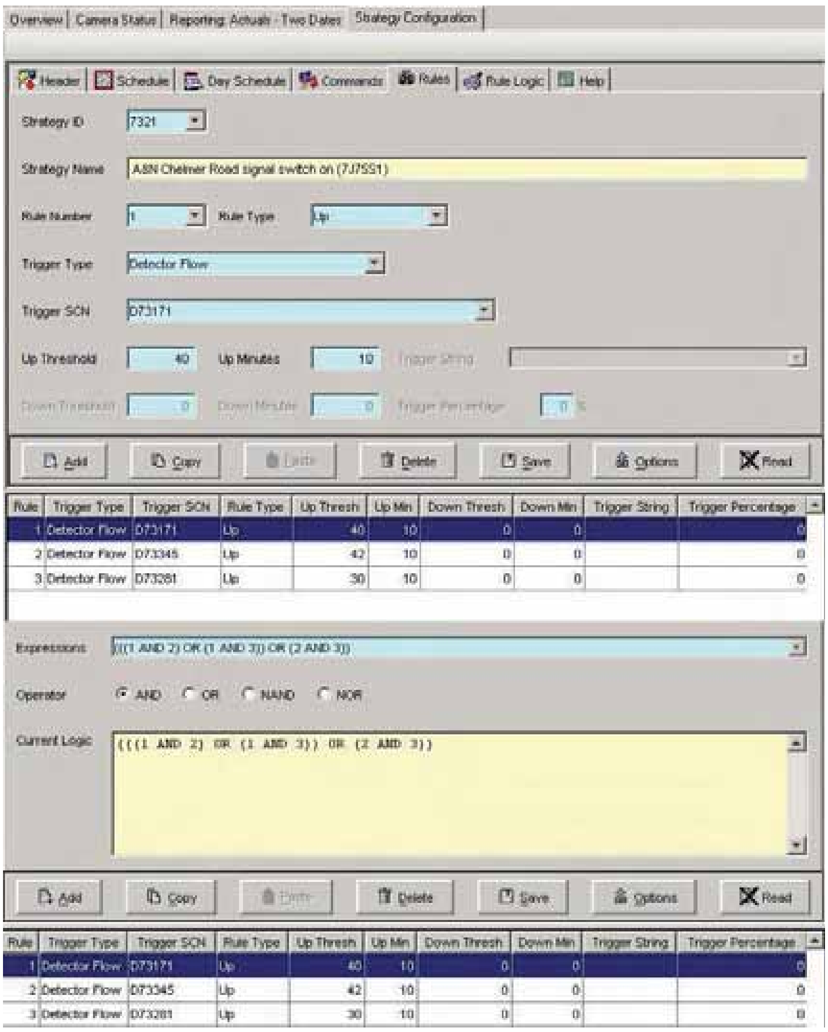


Figure 11 - UTMC strategy with a list of commands to action when activated

When this raw traffic flow data from the loop detectors is transmitted into UTC, it is analysed and converted into flow and congestion figures. Flow is the rate of vehicle movement across the detector, while congestion is the percentage of congested time (when a vehicle is present on a detector for 4 seconds). These data are then forwarded to COMET and are stored in the UTMC database.

An understanding of how the flow and congestion data reacts to traffic conditions is required. For example, the decrease in flow and increase of congestion following a high flow state suggests the start of a flow breakdown and queuing on site. The relation of this data to traffic condition could slightly differ between sites, but can be further understood by the study of historic data together with data from other systems such as CCTV and the journey time measurement system.

With the knowledge of the relation between these data parameters to the true traffic conditions on site, trigger points can then be carefully selected for the activation of strategies see Figure 11. For example, if the congestion is above 80% for more than 10 minutes on Parkway, and the flow is below 20 minutes, this suggests that there is serious congestion on site which could start tailing back into the town centre, spreading the congestion into other areas. When this condition occurs, the automated strategy will be activated and COMET will action a list of pre defined commands, turning on the Parkway/Circulatory signals via UTC and setting the approaching VMS sign on via SieSpace see Figure 12.

The Chelmer Road/circulatory traffic signal node is automated in a similar way. When flow on Essex Yeomanry Way reaches a certain level, the signal will switch on to deal with the heavy flow, but since this is likely to have a negative effect on Chelmer Road there is another rule in the strategy that prevents the signal from switching on, or off, when congestion is excessive on Chelmer Road.

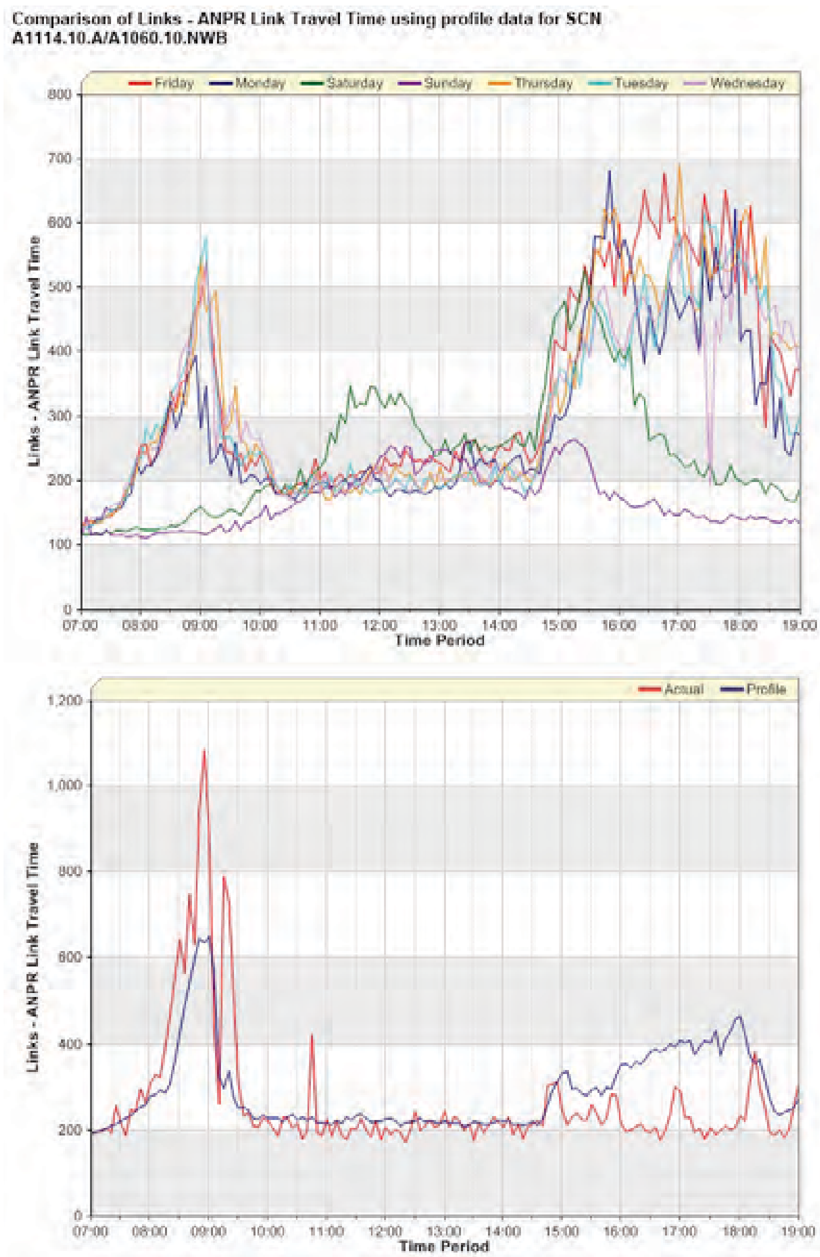


Figure 12 - Figure showing journey time profile data for different day types

The automation of these strategies means that human intervention is not normally required in the management of the dynamic switch on/off of the signal.

This enables more reliable and robust management of the signals switch on/off, however, the effectiveness of this automated system depends on having the appropriate triggers and conditioning logics defined and will often require a trial period to optimise.

Real-time performance based control

Traditionally, it has been difficult to quantify or measure the performance of signal control operation using available detection technology. Technical data such as flow, congestion, occupancy, headway and speed can be converted into performance indicators but this often requires complex modelling and is often difficult to generate in real-time.

A more simple indicator of traffic performance that can be generated in real time, with the current UTMC setup in Essex is "vehicle seconds".

This is the journey time difference from the profile (benchmarked) journey time multiplied by the number of vehicles making that journey. The indicators allow the 'effect' of the traffic signal operation to be quantified in numeric terms. This enables performance based strategies that offers extra efficiency and effectiveness over the condition based (flow and congestion data) strategies. To understand this, we need to first understand how profiling works.

Profiling is a function in COMET that allows five minute slot profiles to be built up using real-time data, with a weighted moving average. For example, on Monday at 09:30, a journey time of 450 seconds is recorded, this will then be averaged (10% weighted) with Monday 09:30 profile journey time. Profiling happens every 5 minutes, therefore individual profiles exist for a Monday 09:30, Monday 09:35, Monday 09:45 etc. Profile is also based on day types, therefore the profile journey time data for Monday 09:30 will be different to the one for Tuesday 09:30, Sunday 09:30 etc (Figure 12). Profiling allows the build up of historic data in a way that enables direct comparison with real time data. Profile data is often understood as "what the data should be generally under normal conditions, at a certain time, on a certain day type".

The comparison of real time journey time with profile data allows a measurement of journey time against normal average conditions and is useful in quantifying effects of new projects such as part-time signal installation, congestion plans and events such as incidents and accidents. Figure 12 shows a journey time on the Essex Yeomanry Way approach into town on the day when the part time signals were first switched on, showing the journey time reduction on the Essex Yeomanry Way approach.

EssexITS has deployed more than 80 Automatic Number Plate Recognition (ANPR) sites around the county over the last few years for measurement of journey times on major inter-urban and radial routes.

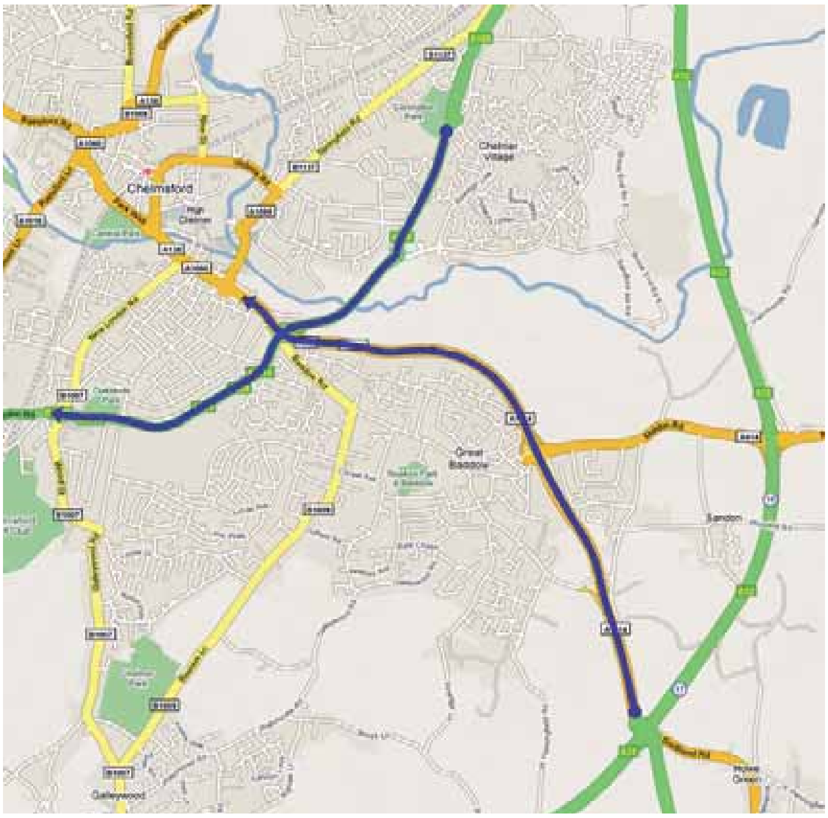


Figure 13 - Figure showing journey time routes

With the Army & Navy roundabout being one of the most important junctions in Chelmsford, journey time measurement is available on Chelmer Road and Essex Yeomanry Way across the roundabout see Figure 13. With the Army and Navy roundabout being the only major roundabout across these journey time routes any changes to the roundabout operation (e.g. part time signal operation) will be reflected in a change of the journey time.

When the Essex Yeomanry Way circulatory part-time signals are switched on, one would expect a decrease of journey time on the Essex Yeomanry Way route and an increase on the Chelmer Road route. But how exactly do we know when any negative effect on Chelmer Road surpasses the benefits that are being yielded on Essex Yeomanry Way? The "vehicle seconds" performance indication will allow a direct numeric comparison of the performance.

Journey time is a measurement of the how quickly traffic is flowing along the route but it is important that journey time data is used together with flow data - the resulting product is "vehicle seconds". A simple example can demonstrate the concept of "vehicle seconds".

Imagine the operation of the traffic signals increases journey times for approach A by 10 seconds and decrease the journey time for approach B by 3 seconds, one may argue that it is overall not beneficial to operate these signals.

However, if there are 20 vehicles suffering from the 10 seconds delays on approach A (200 vehicle seconds) and a 100 vehicles benefiting from a 3 seconds journey time reduction (-300 vehicle seconds), there is an overall improvement of -100 vehicle seconds justifying the operation of the part-time traffic signal.

This real-time model of traffic performance economy can be applied to UTMC strategies. In the case of the Chelmer Road circulatory traffic signals, a direct logical comparison of the "vehicle seconds" on Essex Yeomanry Way and Chelmer Road can be used to justify the switch on/off.

This is not possible in the current version of COMET as the complexity of this strategy exceeds the level supported by the strategy, triggering and conditioning setup in COMET. However, improvements to the strategy management engine in COMET in the near future will make this possible.

Stage 5 – Automated pollution control strategies

EssexITS is proposing further automated strategies at the Army and Navy roundabout, not to tackle congestion but to control the level of pollution. The junction is within an Air Quality Management Area (AQMA) and as such it is essential for Essex County Council to reduce the pollution level at this location.

Emissions from vehicles are especially poor in start-stop traffic, the use of traffic signals on queuing approaches to the roundabout could reduce vehicle emissions by compacting traffic into platoons and transporting them in batches, by running long green/red timings at the traffic signal on the roundabout. This will reduce the amount of start-stop movement and smooth the travel approaching to the roundabout; however drivers might experience longer waiting periods for the red signal due to the long green stage timings on other approaches.

EssexITS is proposing to install pollution monitors on approaches to the Army and Navy roundabout. These monitors will be UTMC compatible and will be able to be linked into COMET. These pollution monitors will enable pollution clearance strategies to be developed and triggered based on real time pollution levels on site. Any pollution clearance strategies can be operated in parallel with congestion strategies they may, however, conflict and, as such, strategy priorities will need to be defined.

Conclusion

The implementation of the UTMC framework in Essex allowed different traffic technologies to communicate and to be used within a single platform. Together with the introduction of the UTMC common database, different traffic system with different purposes can interact enabling new and innovative network management applications.

Unlike most other traffic systems such as UTC, the UTMC common database offers a user friendly graphical interface for traffic operators, enabling control centre operators to indirectly engage with traffic systems that are traditionally only used by traffic engineers.

Simple strategies can be predefined in the common database for functions such as the display of legends on VMS or the switch on/off of traffic signal via the click of a button. More complex strategies involve logics where a vast range of real time data from various UTMC systems can be used as triggers, whether it is flow data from a UTC site, ANPR journey time data, pollution data or car park occupancy etc. The resulting control from a UTMC strategy can also exist in many forms, whether it is the display of messages on VMS, traffic signal timings, plans or mode of traffic control, change of car park status etc. As such, any real-time data from any UTMC systems can have an influential effect on the operation of other UTMC systems. For example, traffic signal can now use real time journey time data, pollution levels, and data from other traffic signal sites as inputs that enable the optimal control to be implemented on site.

As demonstrated in this case study of the Army and Navy Roundabout congestion scheme, the linkage of the UTC, ANPR-JTMS, SieSpace with COMET has opened up vast opportunities for innovative network management strategies for the optimisation of network performance. Many of these strategies are developed to give a co ordinated high level control across different traffic control platforms based on real time traffic condition and performance on site.

UTMC has opened up a new chapter for the industry, and innovative UTMC strategies are rewriting conventional methods and tactics used in traffic signal control and network management.

Stochastic model for strategic assessment of road maintenance

Seósamh Benedict Costello

Senior Consultant

Highways & Transportation
Atkins

Martin Somerville Snaith

Professor of Highway Engineering, Civil Engineering

University of Birmingham
UK

Henry Godfrey Rupiny Kerali

Lead Highway Engineer

The World Bank
USA

Lila Vassiliki Tachtsi

Chief Engineer

Highways & Transportation
Atkins

José Joaquín Ortiz-García

Group Engineer

Highways & Transportation
Atkins

Abstract

This paper describes the development of a stochastic model for the strategic assessment of road maintenance funding and policy decisions. The algorithms used for simulating pavement deterioration utilise Markov processes, resulting in a transition probability matrix that defines defect progression as opposed to the more familiar regression-type model popular with engineers. The model is designed to be used by road maintenance managers and senior administrators for planning medium to long term maintenance investment requirements for local and national road networks. Finally, the application of the model is demonstrated with a case study from central Europe.

Background

This paper describes the results of research carried out at the University of Birmingham over a period of some 10 years in the development of a stochastic model for the strategic assessment of road maintenance funding and policy decisions. The original research[1] was carried out for the Network General and Maintenance Division of the Department of Transport under contract to the Transport Research Laboratory (TRL). This work resulted in a prototype computer model: the TRL Visual Condition Model for Road Networks, also known as the road NETwork COndition Model (NETCOM). More recently[2] the concept was further developed as part of an Engineering and Physical Sciences Research Council (EPSRC) LINK programme into the more advanced prototype STRATegic planning model called STRAT-2.

Objectives of the research

The aim of the research was to develop a stochastic model able to assess strategic funding policies by predicting the effects of various funding and remedial periodic works scenarios on the medium to long term performance of road networks. In particular the model has been developed to:

(i) study the effect of changes in network-level maintenance funding on the condition of a road network

(ii) study the effect of changes in maintenance policies on budget requirements and road network condition

(iii) estimate the total maintenance budget requirement of a road network in future years, thereby identifying future peaks in the requirement for maintenance.

Homogeneous sub-networks

From the outset it was recognised that although all roads deteriorate at differing rates, roads with similar pavement structure and traffic loading, within the same geographic location, tend to exhibit a similar performance. Consequently, it is helpful to categorise the road network of a country or region into homogeneous sub-networks in order to obtain an indication of overall performance. Typical examples of such homogeneous sub-networks in the UK include sections of the motorway network, the trunk road network, and the classified or unclassified road networks in either rural or urban areas. These could be further categorised by the consideration of, for example, climatic region, subgrade soil characteristics or traffic loading. The model then analyses each homogeneous subnetwork individually, and the greater the homogeneity, or specificity, of the network achieved the greater will be the accuracy of the modelling.

Strategic planning

Pavement management is a complex process requiring the asset, in this case the road network, to be managed as a business, as distinct from the regrettably too frequent piecemeal rectification of fully failed sections. As an aid to developing a business framework for pavement management Robinson et al.[3] have classified the various functions of the management process into four distinct components: planning, programming, preparation and operations.

Of particular interest in this study is the planning function. Planning is primarily concerned with the development of longterm, or strategic, estimates of road maintenance expenditure and road condition forecasts under various budgetary scenarios. This involves the analysis of the road network as a whole. However, it is recognised that the road network may be more easily dealt with as a series of sub-networks, each with its own characteristics of traffic, environment, construction and the like. In this paper it is these sub-networks that are modelled independently, and it is recognised that within any one of these sub-networks there will be some variation in the condition that will be reported.

The strategic planning of works is a vital high-level function for pavement management systems (PMS). The defect progression modelling necessary to allow the prediction of future budgetary needs may be done with either deterministic or stochastic (probabilistic) models, with the former being the preferred choice of the majority of system designers. However, this paper addresses the latter as it is believed that road condition, particularly at the network level, is probabilistic in nature.

Markov theory applied to pavement deterioration modelling

Essentially, a stochastic process is a collection of random variables. In statistical terms, a random variable is one of the possible outcomes of an experiment, together with its associated probability of occurrence.

When dealing with pavement deterioration the requirement may be to determine the condition at a particular location on the road network. The random variable would then be the pavement condition at that location coupled with the associated probability of finding it in that condition. The main components of the stochastic process are, therefore, states (condition) and transition probabilities. The transition probabilities specify the likelihood that the pavement will move from one state to another. However, at the network level, the probability that a pavement is in a given condition is interpreted as the expected proportion of pavements in that condition, thereby allowing the proportion of the network expected to be in a certain condition to be calculated:

The Markov prediction model is a specific type of stochastic process and is governed by three 'restrictions'. It is possible to show that the Markov process may be used in the determination of pavement deterioration as it approximates to these three restrictions as follows:

(i) The stochastic process should be discrete in time. Although pavement deterioration is 'continuous' in time, it is possible to consider it as being discrete in time, as it is common to analyse road network condition at specific points in time, usually annually.

(ii) The stochastic process should have a countable or finite state space. Although the state space (that is, the theoretical number of possible outcomes) is infinite in pavement deterioration, this is overcome by expressing the range of possible outcomes as a set of discrete states. That is, the state space is defined as a relatively small number of fixed bands of condition for the particular defect under consideration.

(iii) The stochastic process should satisfy the 'Markov property', where this means that the future state of the process depends on its present, but not past, state. In pavement deterioration it is assumed that the Markov property holds.[1]

Furthermore, by applying the Markov model over a series of years, it is possible to predict the proportion of the network in any state in any future year. Such a series of Markov predictions is called a Markov chain.

In addition to these three restrictions, a discrete-time Markov chain is said to be 'stationary', or homogeneous, in time if the probability of going from one state to another is independent of the time at which the step is being made[4]. In this case it is considered that the road network will deteriorate following the transition probabilities of a single transition matrix. However, if the pattern of deterioration of a particular road network is likely to change at a certain point in time, t, the deterioration process may be modelled by a 'nonstationary' chain. This implies the use of a different transition matrix before and after t. In this case, the distribution of condition at t will become the starting distribution for the second chain, which will operate with a different transition matrix. This type of arrangement may be performed as many times as required, thereby allowing account to be taken, for example, of changing traffic patterns.

Current network condition

The current year network condition provides the base year scenario for both engineering and financial planning in subsequent years. Consequently, the estimate of that condition is important, as of course, are the parameters that are used to describe it. Using Markov theory the initial state of any process may be described by a starting vector, $a_0 = (\ _1, \ _2, \ldots, \ _n)$. Using the analogy of pavement deterioration the starting vector indicates the current condition of the network, defined as the proportion of the network lying in each band of condition. In the model the starting vector is represented by the current condition distribution derived from a road condition database. Table 1 provides an example of such a distribution for wheel track rutting. It will be noted that the defects are divided into bands of condition, also known as severity bands. The severity band limits, upper and lower, and the proportion of the network currently in each of the severity bands are also displayed.

Table 1 - Distributions of condition for wheel track rutting

Severity band	Lower limit: mm	Upper limit: mm	Current network proportion	Post-treatment proportion
1	≥0	<1	0.40	0.48
2	≥1	<5	0.40	0.40
3	≥5	<10	0.07	0.07
4	≥10	<15	0.05	0.05
5	≥15	<20	0.05	0.00
6	≥20	–	0.03	0.00

To illustrate the point, if the data in Table 1 are used then the starting vector for wheel track rutting would be a_0 = (0.40, 0.40, 0.07, 0.05, 0.05, 0.03), where, to satisfy Markov theory, the sum of all ${}_i$ should be equal to 1, and all entries should be positive values.

Deterioration modelling

To model pavement deterioration with time, it is necessary to establish a transition probability matrix (TPM), denoted by P. The general form of P is given by

(1)
$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix}$$

This matrix contains all the information necessary to model the movement of the process among the condition states. The transition probabilities, p_{ij}, indicate the probability of the portion of the network in condition i moving to condition j in one 'duty cycle'. A duty cycle, in pavement deterioration, refers to the degradation in condition due to the damaging effects of one year's traffic and environment. Similar to the starting vector, for every TPM the sum of the entries in each row should be equal to 1, and all entries should be non-negative.

In matrix notation, the probability distribution of the states of the process at a specific elapsed time in years, say t = 1, is therefore given by

(2)
$$a_1 = a_0 P$$

Or, more generally, the probability distribution of the states of the process at any time t may be calculated by

(3)
$$a_t = a_0 P^t$$

The deterioration can therefore be modelled using equation (3), where at is the distribution of condition at time t, ${}_0$ is the distribution of condition at time 0 (that is, the starting vector), and Pt is the TPM raised to the power of t, the elapsed time in years. Two more conditions apply to the process when used to simulate pavement deterioration. First, p_{ij} = 0 for i . j, signifying the general belief that roads cannot improve in condition without

(4)
$$P = \begin{bmatrix} p_{11} & p_{12} & p_{13} & \cdots & p_{1n} \\ 0 & p_{22} & p_{23} & \cdots & p_{2n} \\ 0 & 0 & p_{33} & \cdots & p_{3n} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

first receiving treatment. Second, p_{nn} = 1, signifying a holding state whereby roads that have reached their worst condition cannot deteriorate further. Consequently, in pavement deterioration the general form of the transition matrix P is denoted by

The modelling process can be demonstrated by an example. The distribution of current condition for wheel track rutting displayed in Table 1 is taken as the starting vector, ${}_0$.

Table 2 contains probabilities for wheel track rutting, and is taken as the TPM, P. Then, to simulate one year's degradation in condition due to the damaging effects of traffic and the environment, the two matrices are multiplied as per equation (2). This is represented below:

(5)
$$a_0P = [0.40 \quad 0.40 \quad 0.07 \quad 0.05 \quad 0.05 \quad 0.03] \times \begin{bmatrix} 0.93 & 0.04 & 0.02 & 0.01 & 0 & 0 \\ 0 & 0.90 & 0.06 & 0.03 & 0.01 & 0 \\ 0 & 0 & 0.85 & 0.11 & 0.02 & 0.02 \\ 0 & 0 & 0 & 0.86 & 0.11 & 0.03 \\ 0 & 0 & 0 & 0 & 0.80 & 0.20 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

with the resulting distribution of condition, ${}_1$, represented by the product of the two matrices as follows

(6)
$$a_1 = [0.37 \quad 0.38 \quad 0.09 \quad 0.07 \quad 0.05 \quad 0.04]$$

A further restriction allowing condition to deteriorate by no more than one severity band in any one year is commonly used in pavement deterioration modelling. The TPM is then denoted by

(7)
$$P = \begin{bmatrix} p_{11} & p_{12} & 0 & \cdots & 0 \\ 0 & p_{22} & p_{23} & \cdots & 0 \\ 0 & 0 & p_{33} & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

The Markov prediction model described above has been used in a number of PMS to predict network deterioration. Prominent among these are the Highway Investment Programming System[5] MicroPAVER[6] and the Network Optimisation System[7]

Table 2 - Transition probability matrix for wheel track rutting

Current Severity band	Severity band in the following year						Total
	1	2	3	4	5	6	
1	0.93	0.04	0.02	0.01	0.00	0.00	1
2		0.90	0.06	0.03	0.01	0.00	1
3			0.85	0.11	0.02	0.02	1
4				0.86	0.11	0.03	1
5					0.80	0.20	1
6						1.00	1

Table 3 - Intervention levels for wheel track rutting

Defect	Operator	Intervention value: mm	Treatment
Wheel track rutting		15	Overlay
Wheel track rutting		20	Reconstruction

Maintenance renewals

The above scenario, however, assumes that there is no maintenance carried out on the network from year to year. In practice, maintenance is carried out on a yearly basis as dictated by the intervention levels set from maintenance standards. The selection of the severity bands for the starting vector has to take account of these intervention levels. To illustrate the point, refer to the example maintenance standards for wheel track rutting contained in Table 3. Clearly, the severity band upper and lower limits in Table 1 were defined appropriately, as the lower limit in severity band 5 corresponds with the intervention level for overlay in Table 3. Likewise the lower limit in severity band 6 corresponds with the intervention level for reconstruction. Based on these maintenance standards it can be seen that 5% of the network requires overlaying and 3% of the network requires reconstruction. Assuming sufficient funds are available, the above treatments would be carried out to rectify the proportions of the network with wheel track rutting greater than or equal to 15mm. When a remedial treatment is applied it is assumed to correct the causative defect completely. This is simulated by returning the proportions treated to severity band 1. The post-treatment scenario is displayed in the final column of Table 1. In practice, multiple defects occur on the network, resulting in various remedial measures. Precedence rules are therefore required to satisfy the logical framework of the computer model.

Consequently, treatments are prioritised, with the most effective treatment receiving priority 1 and the rest following in descending order. The example in Table 4, adapted from Kerali and Snaith,1 shows that reconstruction will be the first treatment to be applied on the road network, followed by overlay and then surface dressing. Within each treatment the defect priority determines the order in which defects are treated in the modelling process.

Multiple occurrence of defects

The above procedure, however, assumes that all defects occur independently of each other. Referring to the example in Table 4, it is assumed that road lengths with a low residual life do not necessarily have high rutting or low skid resistance. In practice, however, the more expensive maintenance activities applied to eliminate a severe defect will treat other lower-order defects occurring on the same road length. In the model it is assumed that all defect severity levels are affected by the applied treatment with the exception of those that would trigger higher-order treatments.1 Again using the example in Table 4 to illustrate the point, if reconstruction were applied to a length of road to extend its residual life then all other defects on the same road length would also be rectified and hence eliminated from further consideration. To cater for this in the model a facility exists that allows the user to specify what the program should do when there is a joint occurrence of defects on a road length.

Table 4 - Treatment and defect priorities

Treatment	Treatment priority	Defect	Defect priority	Intervention
Reconstruction	1	Residual life Rutting	1 2	< 0 years ≥ 20mm
Overlay	2	Residual life Rutting	1 2	≥ 0 years but < 10 years ≥ 15mm but < 20mm
Surface dress	3	Skid resistance	1	< 0.35srv

The example in Table 5 defines the percentage of the road network with a given primary defect also having other secondary defects. For example, if 3% of the network is reconstructed because residual life is less than 0 years then, referring to Table 5, 30% of this (i.e. 0.9% of the network) will also have rutting and hence will also be treated. In this particular example the distribution for rutting is adjusted so that the percentage of the network in severity band 1 (that is, perfect condition) is increased by 0.9%. The other severity bands, in total, are decreased by the same amount, each severity band being decreased in proportion to the percentage of the network in the severity band. However, if the treatment had been overlay, which, according to the example in Table 4, does not treat rutting greater than 20mm, then only the distributions of rutting severities below the 20mm intervention level would be treated. It should be noted that the sum of joint occurrence of defects in Table 5 could exceed 100%: this is explained by the fact that more than two types of defect may simultaneously occur on the same part of the road network. The example in Table 5 is a two-dimensional representation of an otherwise complex multi-dimensional matrix.

Ideally the multiple occurrence should be specified for each severity band of the primary defect related to each and every severity band of the secondary defects. Assuming each of the defects in Table 5 had five bands each then the multi-dimensional version of Table 5 would require 150 entries compared with the existing six entries, thereby making it impractical for the user to populate.

Budget allocation

In the above example it was assumed that a sufficient budget was available to treat all defects on the network. However, in practice this is rarely the case. Consequently, maintenance budget limits can be specified within the model. Maintenance treatments are effected according to the priority order specified until the relevant budget is exhausted.

This can be carried out either through a manual iterative process whereby the user maintains complete control over the process or by an automated process based on pre-defined rules built into the system. If the required budget exceeds the budget available in a particular year then a shortfall in the budget occurs, and parts of the network remain untreated.

Analytical framework

The processing involved in the model is broken down into a simulation of the effects of maintenance treatments applied in each year followed by a calculation of the annual progression of defect distributions using equation (2). This iterative process is outlined in Figure 1 and detailed in the steps below.

(i) The base-year defect distributions or starting vectors, ${}_0$, are derived from current condition data stored in the road database.

(ii) The proportion of each network requiring maintenance is then calculated based on the defect distributions and the specified maintenance standards, or intervention levels. When these are aggregated for the network this figure is referred to as the total maintenance need.

(iii) The percentage of each network that will actually receive treatment is calculated based on the specified maintenance budget limits. The actual budget spent in maintaining the network is referred to as the maintenance expenditure.

(iv) The effect that the maintenance will have on the condition of each network is calculated having allowed for the multiple occurrence of defects.

(v) Finally, the condition of each network in the following year is predicted using the TPM. The above process is carried out for each year of the analysis until the end of the analysis period is reached.

Table 5 - Multiple occurrance of defects

Primary defect	Secondary defects		
	Residual life	**Rutting**	**Skid resistance**
Residual life	-	30%	20%
Rutting	20%	-	10%
Skid resistance	5%	5%	-

Case study

The above model has been used in a number of design, build, finance and operate (DBFO) concessions in the UK; however, because of the competitive nature of the Private Finance Initiative (PFI) the details of these trials remain confidential. Instead, a further application in Slovenia, funded by the European Bank for Reconstruction and Development (EBRD), called the Slovenia Private Roads Maintenance Project[8] is presented. The overall aim of the feasibility stage of this project was to ascertain whether undertaking all or some part of the maintenance of the state road network would offer better value for money if procured under a series of public–private partnership (PPP) concession arrangements as opposed to the current procurement method.

To aid in this process the study included the definition of the extent and condition of the road network as a starting-point, followed by an analysis of its deterioration and maintenance effects under a number of different maintenance strategies. This resulted in a preferred long-term rehabilitation and maintenance expenditure profile, and it is this that is presented here. Further details of the case study can be found in Ortiz-García[9].

Homogeneous sub-networks

The Slovenian State Road Network, over 5000km in total, was first divided into five candidate concession areas. Within each concession area a homogeneous sub-network was defined for each combination of traffic group and climatic zone. Roads were categorised into one of five traffic groups, depending on the total number of 80kN standard axles predicted to traffic the pavement during the 20-year analysis period. The ranges for each traffic group are presented in Table 6. Three distinct climatic zones exist in Slovenia: sub-Mediterranean, moderate continental, and mountainous.

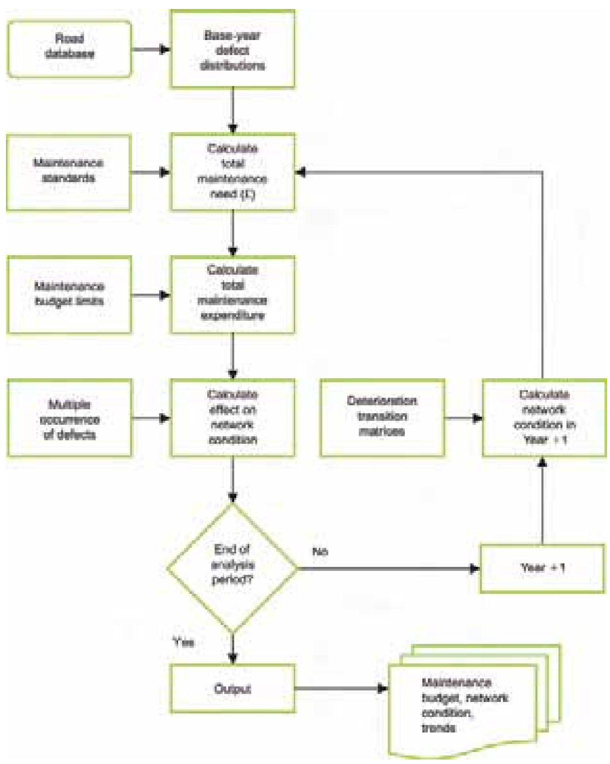


Figure 1 - Analytical framework

HIGHWAYS

Table 6 - Traffic classification

Traffic group	Range: equivalent standard axles of 80kN
Very Heavy	> 7,000,000
Heavy	2,000,000 - 7,000,000
Medium	700,000 - 2,000,000
Light	200,000 - 700,000
Very Light	< 200,000

Table 7 - Climate zones

Climatic zone	Average temperature of the coldest month °C	Average temperature of the warmest month °C	Average annual precipitation mm
Sub-Mediterranean	0 - 4	20 - 22	1200 - 1700
Moderate continental	-3 - 0	15 - 20	800 - 2800
Moutainous	< -3	~ 10	1100 - 3500

Table 8 - Boundary MSI values based on traffic

Traffic volume (AADT)	Condition band				
	Very good (VG)	Good (G)	Fair (F)	Poor (P)	Very poor (VP)
< 1000	< 0.9	0.9 - 1.6	1.6 - 2.1	2.1 - 3.3	> 3.3
1000 - 2000	< 0.8	0.8 - 1.5	1.5 - 2.0	2.0 - 3.2	> 3.2
2000 - 5000	< 0.7	0.7 - 1.4	1.4 - 1.9	1.9 - 3.1	> 3.1
5000 - 10,000	< 0.6	0.6 - 1.3	1.3 - 1.8	1.8 - 3.0	> 3.0
10,000 - 20,000	< 0.5	0.5 - 1.2	1.2 - 1.7	1.7 - 2.9	> 2.9
> 20,000	< 0.4	0.4 - 1.1	1.1 - 1.6	1.6 - 2.8	> 2.8

Table 9 - Example TPM

From\To	VG	G	F	P	VP
VG	0.80	0.20			
G		0.86	0.14		
F			0.93	0.07	
P				0.96	0.04
VP					1

Table 10 - Transition probabilities

Matrix	$P_{vg\text{-}vg}$	$P_{g\text{-}g}$	$P_{f\text{-}f}$	$P_{p\text{-}p}$
TPM1	0.80	0.86	0.93	0.96
TPM2	0.76	0.83	0.93	0.95
TPM3	0.72	0.80	0.88	0.94
TPM4	0.85	0.90	0.95	0.97

Pavements in these zones are subject to different ranges of temperature and varying levels of rainfall and snowfall, as shown in Table 7. The resulting five by three matrix has 15 cells, each one a distinct homogeneous sub-network with a unique combination of traffic loading and climate.

Current network condition

Condition information was available for each section in the network in terms of the Modified Swiss Index (MSI)[10] which is a composite condition index based on the extent and severity of visually collected pavement condition parameters.

MSI values were grouped into severity bands as shown in Table 8. It was therefore possible to determine condition distributions for each homogeneous sub-network. In the modelling process outlined above, these equate to $_0$, the starting vector.

Deterioration modelling

The next stage of the process required the definition of a transition probability matrix (TPM) for each homogeneous subnetwork. The lack of available historic data on pavement deterioration resulted in the development of such models through panel discussions with local pavement engineers. An example of one particular transition matrix, TPM1, may be seen in Table 9. In the modelling process outlined above, the TPM is denoted by P.

In total, four TPMs were developed, as shown in Table 10. The transition probabilities p_{ii} indicate the probability of a road length remaining in condition band i in any one year. The probabilities of moving from condition band i to condition band j (next worse) in any one year are calculated as p_{ij} = 1 - p_{ii}. As an example, for TPM1 the probability that roads currently in 'very good' condition will remain in 'very good' condition ($p_{vg\text{-}vg}$) after one year's degradation due to the damaging effects of traffic and the environment is 0.80. Consequently, the probability that roads currently in 'very good' condition will degrade to 'good' condition ($p_{vg\text{-}g}$) after one year's degradation due to the damaging effects of traffic and the environment is 0.20.

The allocation of TPMs to homogeneous sub-networks is shown in Table 11. It was Traffic group Range: equivalent standard axles of 80kN therefore possible to model deterioration for each homogeneous sub-network using equation (2).

Maintenance renewals

The acceptable maintenance standard or intervention level chosen required roads to be kept above the 'fair' condition level. Consequently, maintenance renewals refer to all activities required to bring roads in 'very poor', 'poor' or 'fair' condition to 'very good' condition. Maintenance renewal costs used in the modelling were based on generic activities required to treat a road that has fallen into 'very poor', 'poor', or 'fair' condition, bringing it back to 'very good' condition. These costs may be found in Table 12.

Table 11 - Allocation of TPMs to homogeneous sub-networks

Climatic zone	**Traffic group**				
	Very heavy	**Heavy**	**Medium**	**Light**	**Very light**
Sub-Mediterranean	TPM1	TPM1	TPM1	TPM4	TPM4
Moderate continental	TPM2	TPM2	TPM1	TPM1	TPM1
Mountainous	TPM3	TPM3	TPM2	TPM2	TPM2

Table 12 - Maintenance renewals unit costs: €/m²

Traffic group	**Pavement condition before rehabilitation**		
	Very poor	**Poor**	**Fair**
Very heavy	29.81	16.31	10.77
Heavy	19.99	13.93	9.88
Medium	14.82	12.56	8.99
Light	13.93	10.77	8.99
Very light	9.64	9.88	8.99

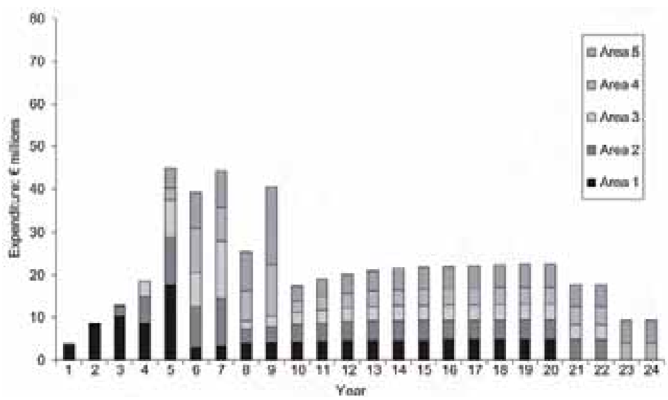


Figure 2 - Maintenance expenditure

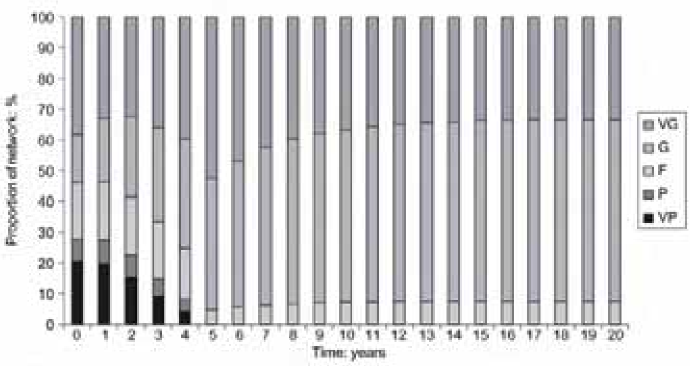


Figure 3 - Progression of network condition

Strategy generation

The chosen strategy aimed at prioritising maintenance expenditure according to the importance of the road. Importance was assumed to be defined by the traffic group. Consequently, in the model, roads carrying 'very heavy' traffic were treated first, followed by roads carrying 'heavy' traffic, and so on.

Strategy outcome

The expenditure profile obtained with this strategy, as shown in Figure 2, shows a gradual increase in expenditure during the first five years up to a level of about €40 million per year (all concession areas), for five years. Expenditure then reaches a level of about €20 million per year for the next 11 years, and then decreases gradually during the last four. In this strategy the most important roads are treated first, which will contribute to a favourable public perception of the concessions scheme. Should this strategy be adopted by the concessionaire the progression of improvement in road condition would follow the pattern shown in Figure 3.

Discussion and conclusions

The stochastic model, developed as part of this research, provides a suitable framework to assist maintenance managers and senior administrators in predicting the medium to longterm performance of road networks. Such a framework allows for the strategic assessment of road maintenance funding and policy decisions. Where changes to funding levels are proposed, the effect of those changes on the future condition of the road network can be modelled. Likewise, any proposed changes to maintenance policy can be assessed by modelling the effect of these changes on future budget requirements and road network condition. Alternatively, based on current funding levels and existing maintenance policies, the model provides a framework to estimate the total maintenance budget requirement of the road network in future years. Finally, the applicability of the model has been demonstrated with a case study from Central Europe.

Acknowledgements

The authors wish to thank the TRL, EPSRC and the LINK programme industry sponsors, namely WS Atkins Consultants Ltd and Highway Management Services Ltd, for their financial support down through the years in making this research possible.

REFERENCES

1. KERALI H. R. and SNAITH M. S. NETCOM: The TRL Visual Condition Model for Road Networks. Transport Research Laboratory, Crowthorne, 1992, Contractor Report 321.
2. COSTELLO S. B. The Development of an Integrated Strategic Planning Tool for Road Maintenance Funding. PhD thesis, The University of Birmingham, 2001.
3. ROBINSON R., SNAITH M. S. and DANIELSON U. Road Maintenance Management: Concepts and Systems. Macmillan, Basingstoke, 1998.
4. ISAACSON D. L. and MADSEN R. W. Markov Chains: Theory and Applications. Wiley, New York, 1976.
5. THOMPSON P. D., NEUMANN L. A., MIETTINEN M. and TALVITIE A. A micro-computer Markov dynamic programming system for pavement management in Finland. Proceedings of the 2nd North American Conference on Managing Pavements, Toronto, 1987, 2, 2.241–2.252.
6. SHAHIN M. Y. and KOHN S. D. Overview of the ‘PAVER’ Pavement Management System. Construction Engineering Research Laboratory, Champaign, 1982, Technical Manuscript M-310, USA-CERL.
7. GOLABI G., KULKARNI R. and WAY G. A statewide pavement management system. Interfaces, 1982, 12, No. 6, 5–21.
8. DRSC. Slovenia Private Road Maintenance Project: Strategy Report. WS Atkins Consultants Limited, Birmingham, 2001.
9. ORTIZ-GARCÍA J. J. Long-term strategic planning of pavement maintenance in privately financed highway network concessions: Slovenia case study. Proceedings of the 22nd PIARC World Roads Congress, Durban, 2003.
10. BOLE D., MILJEVIČ J. and GREGORC C. State road network rehabilitation models application to Slovenia. Proceedings of the 2nd European Road Research Conference, Brussels, 1999.

The identification of maintenance hotspots on the South Wales Trunk Road Network

David Wightman
Group Engineer
Highway Asset Management Group
Highways & Transportation
Atkins

Abstract

This paper describes a study undertaken by Atkins in 2008 on behalf of the South Wales Trunk Road Agency (SWTRA), and in conjunction with the SWTRA Asset Management Section. The primary purpose of the study was the development of a methodology for the identification of condition hotspots based on network-level survey data.

As part of its remit to manage, maintain and improve the strategic motorway and trunk road network in South Wales, SWTRA are required to identify and then develop candidate maintenance schemes as part of the Value Management (VM) process.

A robust means of identifying future maintenance schemes is critical both to SWTRA, and to its various stakeholders, including the Welsh Assembly Government (WAG). From SWTRA's perspective, scheme identification drives the forward works programme and consequently the maintenance budget. From WAG's perspective, a robust process helps to ensure that funds are targeted where maintenance need is greatest, and to ensure that its various strategic objectives (safety, journey reliability, environmental impact, level of service, and value for money, etc.) are met.

The WAG network is surveyed and inspected on a systematic and regular basis. Surveys are conducted which measure surface defects, structural adequacy and skidding resistance. The data resulting from these surveys provides a potentially sound basis for the identification of schemes (as part of the forward work programme). SWTRA perceived a potential benefit in utilising the available data in this manner, seeing it as means of making best use of available resources.

The aim of this study, led by the author, was to develop a robust means of using the collected survey data for scheme identification and prioritisation. The Atkins project team, supported by colleagues at SWTRA, have developed a computerised toolkit which takes survey data extracted from the Welsh Highways Information System (WHIS), applies various phases of data manipulation/analysis, and identifies a series of "areas of interest" (or hotspots) for further investigation, and eventual scheme development. This paper describes the methodologies adopted and the outcomes of the work.

Methodology overview

The basis of the approach adopted for this study, was the application of a hotspot identification toolkit called Schema. This toolkit has been developed by the author (with the assistance of colleagues from Atkins) over the course of a number of similar studies. The key elements of the Schema approach are as follows:

Multi-Criteria analysis (MCA)

MCA techniques are used to replicate the decision-making process used by pavement engineers when identifying maintenance hotspots. The Schema MCA framework is configured to use the range of factors (or criteria) typically used in such decisions. When applied to any given section of road, the MCA generates a score based on the relevant values for the selected criteria.

If the MCA is run for the entire network, a score is produced for each subsection; this can then be used to rank the entire network in terms of maintenance need.

Encapsulation of engineering judgement

By assigning user-defined weightings to each of the criteria used in the MCA, the relative importance of the various criteria on the decision-making process can be modelled. A pair-wise comparison exercise is undertaken with local engineers in order to ascertain the appropriate weighting values.

Data quality assessment

Various aspects of data quality are assessed including completeness, currency, and reliability. A by-product of this process is that various useful network statistics are generated (including condition banding, etc). In addition, any anomalies in survey data are identified (including outlier values, assignment to invalid Cross-Sectional Position - XSP - etc).

Cluster analysis

Clustering techniques are used to identify groups of hotspots in close geographical proximity. The intention is to identify sites with strong evidence of maintenance need. These sites (or "areas of interest") are targeted for further investigation, with some going on to form candidate VM schemes.

Automation

The above steps are largely automated, having been encapsulated in a software toolkit (Schema). The complete process involves many phases and can involve many thousands of individual calculations. As well as significantly speeding up this process, automation minimises the risk of human error and helps to ensure objectivity in the decision making process.

These are the key generic features of the Schema toolkit, and are used in all applications. The sections that follow describe how the toolkit was applied for the SWTRA study.

Identification of decision criteria

Discussions were held to determine which indicators (or criteria) should be used for hotspot identification. The primary aim was to establish those factors which are currently used by Area Engineers and Route Stewards as part of existing scheme identification. There is often a trade-off in such situations between the desire to replicate the current decision making process as closely as possible, and the practical matter of data availability (i.e. there is little value in using an indicator which is not readily available, or which has poor network coverage).

In the case of SWTRA the vast majority of available data is stored in the Welsh Highway Information System (WHIS). WHIS is based upon the widely used WDM Pavement Management System (PMS) and includes the data obtained from TRACS, Deflectograph and SCRIM surveys, as well as traffic, accident and network referencing data.

Based on initial conversations with the Agency, and as a result of experience from past studies, the following criteria were initially selected as the basis for scheme identification:

- Maximum rut depth
- LPV 3m
- SCRIM difference
- LPV 10m
- Texture depth
- LPV 30m
- Route hierarch
- Residual life
- Network disruption (expressed in terms of traffic)
- Accidents

Dynamic segmentation

The MCA module requires that the host network is partitioned into a series of homogeneous subsections. In this instance, a homogeneous subsection is defined as a length of road for which the various indicators (i.e. TRACS, SCRIM, Deflectograph, traffic, etc) remain constant.

The primary difficulty here is the fact that the reporting intervals for the various data-sets are not aligned. Whereas the data resulting from the TRACS and Deflectograph surveys are stored in WHIS at regular 100m intervals, SCRIM data is stored at intervals that reflect site category length, as defined in HD28/04[2]. These are very rarely at regular 100m intervals.

The SWTRA data-set was partitioned using a form of dynamic segmentation sometimes referred to as the "Highest Common Denominator" approach. Using this approach, the network is divided into a series of homogeneous subsections where the length of any given subsection is the longest possible length for which all relevant characteristics are constant (or homogeneous). This process resulted in 18,681 homogeneous subsections, covering a total of 1,128 lane kilometres.

Handling gaps in data

The hotspot identification process is heavily reliant on the completeness (and quality) of the source data. An analysis was conducted to establish data completeness for the selected MCA criteria. Table 1 shows the data coverage for the CL1 and CR1 XSPs (i.e. the nearside lane).

As can be seen, coverage for CL1 and CR1 was deemed to be either Very Good or Excellent for all data-sets. In the experience of the project team, WHIS data coverage is at least as good as that of comparable agencies.

As coverage for other XSPs is less complete (as is traditionally the case), it was agreed that analysis would be based on CL1/CR1 data only.

Although data coverage was found to be excellent, of the 18,681 homogeneous subsections resulting from the preceding step, 5,053 (29%) of these had values missing for one or more of the analysis criteria (put another way, 4.18% of values were missing). Although the proportion of missing data is relatively small, the MCA framework does not tolerate missing data, and requires values to be supplied for all criteria. Therefore, some approach to handling incomplete data is required.

In order to retain as full a data-set as possible, whilst ensuring that any default values used were as reasonable as possible, it was decided that any gaps would be populated using route average values. To facilitate this approach, route averages were calculated for the entire network, for each of the selected MCA criteria. Prior to the use of the base data-set in the MCA analysis, all data gaps were then populated using these average values.

Other aspects of data quality

Data completeness is only one aspect of data quality. The Schema toolkit includes a number of facilities for investigating other aspects of data quality:

Data currency

The age of data can impact on its relevance to the analysis process. An analysis of data currency was therefore undertaken. As with completeness, data currency was found to be generally excellent.

Table 1 - CL1/CR1 Data Coverage

Criteria	Coverage	Excellent > 95%	Very Good 80-95%	Good 65-80%	Poor 50-65%	Very Poor < 50%
Rutting	96.37%	x				
Texture	96.37%	x				
LPV	96.37%	x				
SCRIM	98.38%	x				
Residual life	80.00%		x			
Traffic	99.95%	x				

Outliers

A small number of instances were encountered where values in the base data-set were deemed to be outside reasonable limits[1]. In particular 2.85% of LPV10m, and 10.1% of LPV30m data was deemed to be outside reasonable bounds. In such cases, the offending values were replaced with route average values. (NB. both null values and out-of-range values were excluded from the calculation of route averages). In addition, a number of instances were encountered where data was assigned to an XSP which was invalid for the relevant road type. Any subsections thus affected were excluded from the analysis.

Network statistics

As part of the data quality analysis process described above, a number of useful network statistics are calculated (including condition banding, various network length breakdowns, etc). These statistics are displayed on the Schema Dashboard.

Weighting acquisition

The next step in the process was to determine weightings for each of the decision-making criteria identified above. The resultant weightings are used to reflect the relative importance of the various criteria on the decision making process. In addition to the initial choice of criteria, the application of weightings is the primary means by which engineering judgment is reflected in the hotspot identification process. In order to establish appropriate weightings, a series of meetings was held with experienced engineers. Experience of similar studies has shown that engineering decisions are often based on instinct. The engineer is not always aware of (or able to relate) the various factors involved in his decision making process, nor the relative importance attached to each. Furthermore, different engineers will often have differing views on the relative importance of the various criteria. So it proved in this case. Figure 1 illustrates the range of different weighting scenarios which resulted from this initial consultation (NB. the term weighting scenario is used to refer to a particular combination of weighting values):

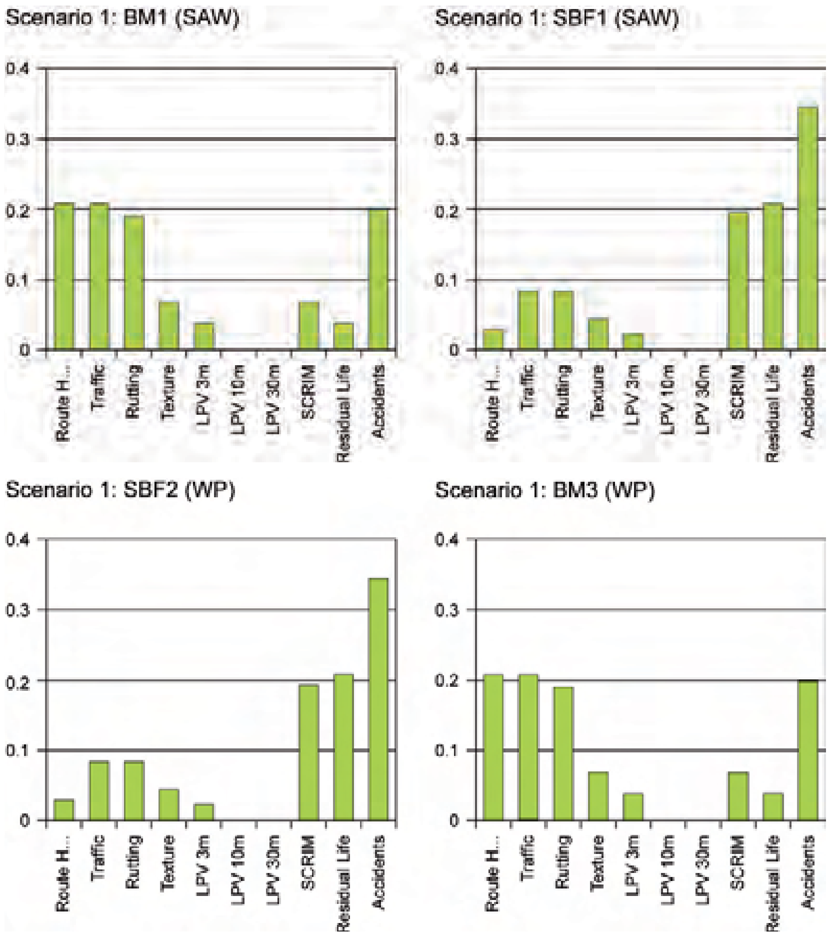


Figure 1 - Initial range of weighting scenarios

There are a number of common ways in which MCA weightings are elicited. For the SWTRA study, it was decided that the Analytical Hierarchy Process (AHP) would be used. The AHP approach requires the domain-expert to specify the relative importance of the various criteria in terms of a hierarchy. The hierarchy can be visualised as a tree, like that depicted in Figure 3, with the goal on the left hand side, and the criteria and supporting data used to arrive at the goal on the right.

Using this process, a number of initial weighting scenarios were established. A trial MCA run was then conducted for each scenario in order to test its "reasonableness". A further workshop was held at which the results of these runs were examined, with the objective of testing how closely the results of each run reflected engineering judgement/experience. As a result of this exercise it was decided that the following criteria should be excluded from subsequent analyses:

Route hierarchy/traffic

It was felt that these indicators unnecessarily skewed results in favour of heavily trafficked routes, particularly the M4. In addition it was felt that inclusion of both indicators introduced an element of double-counting.

Accidents

It was decided that accident data should be excluded since a separate accident analysis exercise was already in existence (for identifying safety schemes).

LPV10m/LPV30m

It was decided that the LPV3m measure provided the best indicator of maintenance need. Furthermore, inclusion of all three indicators might result in double-counting. Exclusion of the 10m and 30m measures also circumvented the reliability problems identified above.

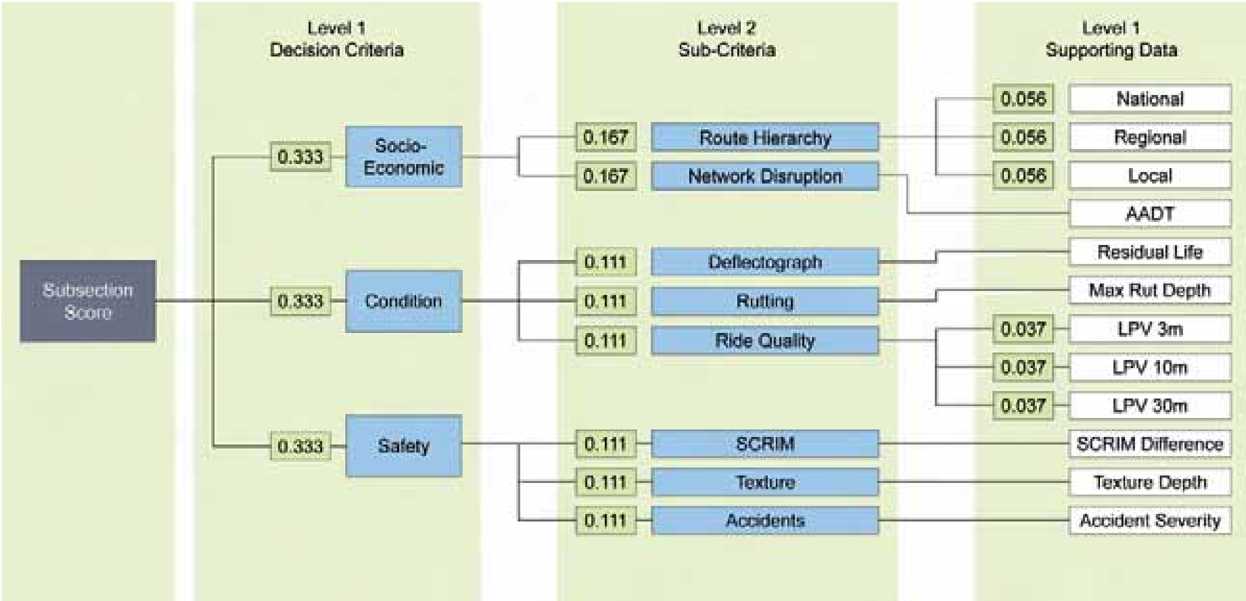


Figure 2 - Sample AHP decision tree

In addition to the exclusion of the above criteria, revised weightings were attached to the remaining criteria, resulting in a final unified weighting scenario as illustrated in Figure 3.

Multi Criteria Analysis

At this point, the various pre-requisites for MCA were now in place: suitable base data had been obtained; a set of homogeneous subsections had been derived; the resulting data had been analysed and cleansed as appropriate; MCA criteria had been established and a final set of weightings obtained. MCA was now conducted on the resultant data-set.

Various forms of MCA are commonly used. The Schema toolkit incorporates three which have been judged to be most suitable for hotspot identification, namely: Weighted Product (WP), Simple Additive Weighting (SAW), and Technique for Order Preference by Similarity to Ideal Solution (TOPSIS). For the purposes of the SWTRA study, the MCA was conducted using both the WP and SAW methods to produce two separate rankings. A weighted average of these was then calculated to produce a final ranking.

In this ranking, all 18,681 subsections are ordered according to their MCA score. This score is determined using the subsection's values for the various MCA criteria, and the weightings elicited from engineers.

The subsections at the top of the ranking (positions 1, 2, 3, etc.) are deemed to have the worst condition; those at the bottom of the ranking (positions 18,679, 18,680, 18,681) are deemed to have the best condition. Figure 4 shows sample output from the MCA process:

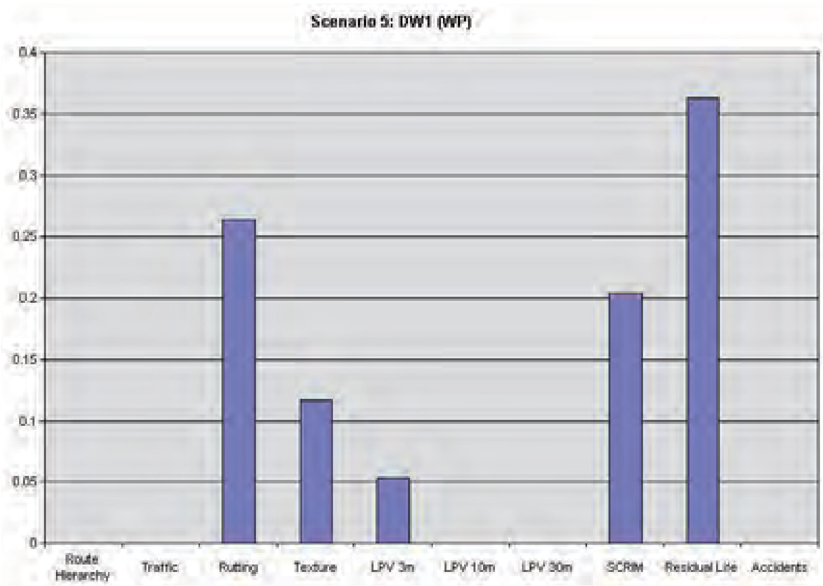


Figure 3 - Final weighting scenario

Cluster Analysis

Cluster Analysis techniques were applied to the ranked list resulting from the preceding step, in an attempt to identify areas of the network with greatest maintenance need (i.e. hotspots). The Schema toolkit incorporates a generic cluster analysis algorithm which can be configured to suit the needs of a particular project.

ID	Road	Section	XSP	Section Type	Start	End	Length	Ranking 1	Ranking 2	Ranking 3	Ranking 4	Ranking 5	Ranking 6	Avg Ranking
9826	A470	1047002004	CL1	ONEWAY 2-LANE CARRIAGEWAY	50	100	50	1	42	1	38	6	77	27.5
9824	A470	1047002004	CL1	ONEWAY 2-LANE CARRIAGEWAY	0	40	40	2	43	2	41	7	78	28.8
9825	A470	1047002004	CL1	ONEWAY 2-LANE CARRIAGEWAY	40	50	10	3	44	3	48	8	79	30.8
15175	M4	1000416088	CL1	DUAL 2-LANE	300	370	70	5	1	361	1	1	1	61.7
16103	M4	1000430006	CL1	SLIP 2-LANE CARRIAGEWAY	100	200	100	80	38	11	4	82	163	63.0
15176	M4	1000416088	CL1	DUAL 2-LANE	370	400	30	6	2	398	2	2	2	68.7
6178	A449	1044905006	CL1	DUAL 2-LANE	100	140	40	12	222	7	67	76	42	71.0
9863	A470	1047004008	CL1	SLIP 2-LANE CARRIAGEWAY	300	330	30	41	70	240	52	37	16	76.0
9844	A470	1047002005	CL1	ONEWAY 2-LANE CARRIAGEWAY	400	420	20	4	73	4	279	9	110	79.8
16102	M4	1000430006	CL1	SLIP 2-LANE CARRIAGEWAY	0	100	100	101	41	32	7	89	233	97.2
10553	A470	1047009087	CL1	SLIP 2-LANE CARRIAGEWAY	10	100	90	27	6	546	3	10	3	99.2
10552	A470	1047009087	CL1	SLIP 2-LANE CARRIAGEWAY	0	10	10	31	7	501	5	11	4	106.5
16125	M4	1000430011	CL1	SLIP 2-LANE CARRIAGEWAY	480	500	20	168	47	17	6	128	322	114.7
16124	M4	1000430011	CL1	SLIP 2-LANE CARRIAGEWAY	460	480	20	188	48	26	11	129	323	120.8
16123	M4	1000430011	CL1	SLIP 2-LANE CARRIAGEWAY	400	460	60	189	49	27	12	130	324	121.8
14939	M4	1000414006	CL1	DUAL 2-LANE	200	300	100	33	72	19	27	538	387	179.3
16122	M4	1000430011	CL1	SLIP 2-LANE CARRIAGEWAY	360	400	40	417	58	135	23	179	378	198.3
14909	M4	1000410096	CL1	DUAL 2-LANE	500	600	100	465	24	500	16	49	171	204.2
16121	M4	1000430011	CL1	SLIP 2-LANE CARRIAGEWAY	300	360	60	433	63	155	25	181	379	206.0
9839	A470	1047002005	CL1	ONEWAY 2-LANE CARRIAGEWAY	0	100	100	34	391	10	217	85	579	219.3
84	A40	1004002020	CL1	DUAL 2-LANE	900	920	20	28	567	8	526	131	101	226.8
60	A40	1004002005	CL1	ROUNDABOUT (2-LANE)	0	100	100	40	492	63	692	80	39	234.3
16265	M4	1000434009	CL1	SLIP 2-LANE CARRIAGEWAY	440	490	50	87	52	1220	55	30	9	242.2
16264	M4	1000434009	CL1	SLIP 2-LANE CARRIAGEWAY	400	440	40	97	59	1315	77	31	10	264.8
16119	M4	1000430011	CL1	SLIP 2-LANE CARRIAGEWAY	240	260	20	530	66	420	33	197	429	279.2
16118	M4	1000430011	CL1	SLIP 2-LANE CARRIAGEWAY	200	240	40	540	67	438	43	198	430	286.0
16120	M4	1000430011	CL1	SLIP 2-LANE CARRIAGEWAY	260	300	40	541	68	439	44	199	431	287.0
9908	A470	1047003010	CL1	SLIP 2-LANE CARRIAGEWAY	0	30	30	596	133	122	32	337	595	302.5
9910	A470	1047003010	CL1	SLIP 2-LANE CARRIAGEWAY	70	100	30	600	139	128	37	338	596	306.3
9909	A470	1047003010	CL1	SLIP 2-LANE CARRIAGEWAY	30	70	40	608	146	137	42	340	597	311.7
16207	M4	1000432011	CL1	SLIP 2-LANE CARRIAGEWAY	430	440	10	296	269	33	70	564	708	323.3
16206	M4	1000432011	CL1	SLIP 2-LANE CARRIAGEWAY	400	430	30	301	274	36	73	565	709	326.3
15122	M4	1000416012	CL1	DUAL 2-LANE	1100	1150	50	569	87	346	50	342	564	326.3
15123	M4	1000416012	CL1	DUAL 2-LANE	1150	1200	50	570	88	347	51	343	565	327.3
62	A40	1004002005	CL1	ROUNDABOUT (2-LANE)	190	200	10	622	136	541	64	322	294	329.8
61	A40	1004002005	CL1	ROUNDABOUT (2-LANE)	100	190	90	623	137	543	66	323	295	331.2
16208	M4	1000432011	CL1	SLIP 2-LANE CARRIAGEWAY	440	470	30	306	285	40	84	567	714	332.7
16209	M4	1000432011	CL1	SLIP 2-LANE CARRIAGEWAY	470	490	20	307	287	42	88	568	715	334.5
14936	M4	1000414005	CL1	DUAL 2-LANE	300	380	80	44	357	22	102	992	601	336.3
14893	M4	1000410092	CL1	DUAL 2-LANE	300	400	100	88	177	54	72	1122	546	343.2

Figure 4 - Sample MCA outputs

In this case the clustering algorithm was constrained such that it only identified clusters which lay on a single SWTRA road. Clusters which spanned multiple roads were therefore not considered.

The identified clusters were plotted on a map of the SWTRA network. A sample cluster plot is shown in Figure 5.

It should be noted that the sites resulting from the cluster analysis were not put forward as candidate schemes. The cluster analysis is merely used as a first pass attempt at grouping those hotspots in closest geographic proximity. The resultant sites are then used to guide subsequent site visits by engineers (as described below). Ultimately, engineering judgement is used to determine scheme locations and extents.

Site investigations

The top 16 sites resulting from the Cluster Analysis process were combined with 12 sites identified by the Area teams using traditional means (defect reports, drive-by inspections, local knowledge, etc.). All 28 sites were then visited by local staff together with an Atkins pavement engineer. In each case either a driven or a walked inspection was conducted.

The Schema software was used to generate a Condition Summary Sheet (CSS) for each site. The CSS is intended to aid defect location and in due course to target appropriate treatment.

Figure 5 - Indicative Cluster Plot using Sample Data

HIGHWAYS

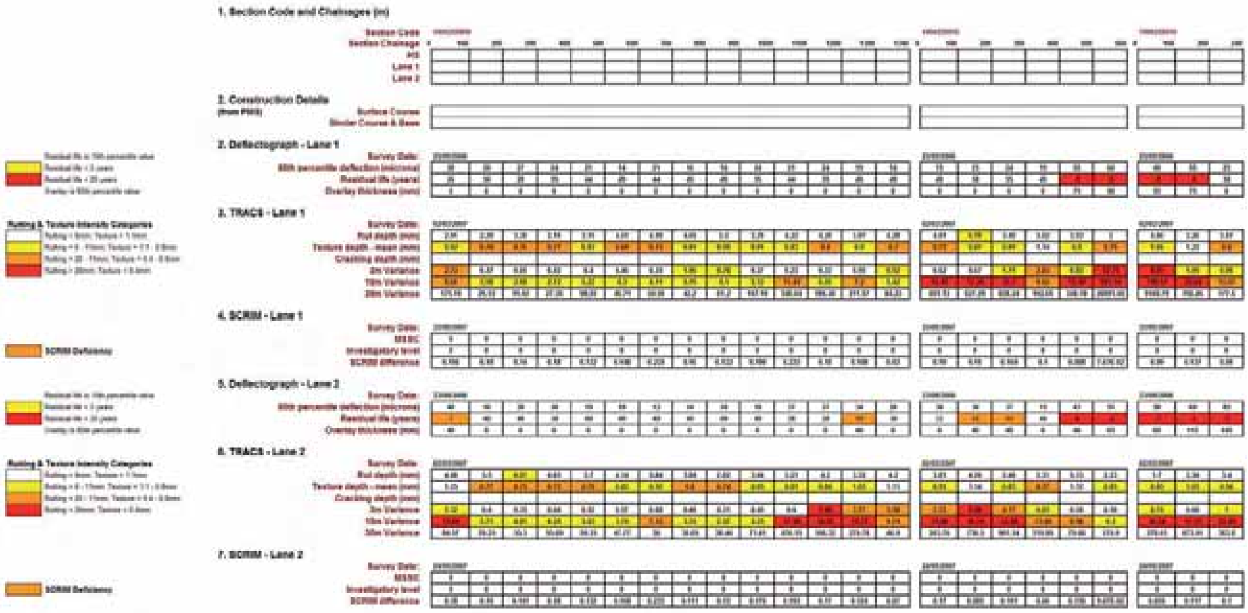


Figure 6 - Sample condition summary sheet

Outcomes

Following the site visits, a workshop was held between SWTRA and Atkins project staff. Of the 28 sites inspected, 8 MCA-identified sites were discarded as it was found that maintenance work had been undertaken since the most recent survey. Two further candidate sites were excluded: one was found to be part of a major forthcoming maintenance scheme; the other was identified as requiring early intervention and was fast-tracked as a Stage 2 scheme.

Of the remaining 18 sites, the 10 deemed to have greatest maintenance need were designated as Stage 1 - Year 1 schemes; the remaining 8 were designated as Stage 1 - Year 2 schemes.

Of the 10 sites designated as Stage 1 schemes:

- 5 were identified by both MCA and the Area teams
- 3 were identified solely by MCA
- 2 were identified solely by the Area teams

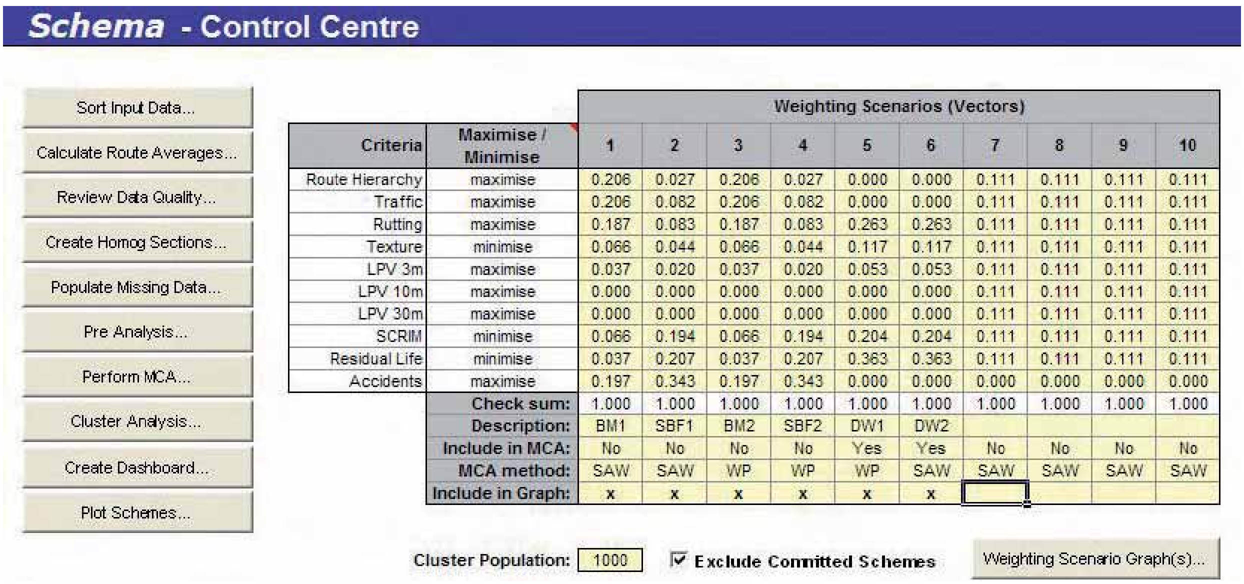
Eight of the worst 10 sites had identified by the MCA. In the case of the final two sites, it was found that the relevant WHIS survey data did not reflect site conditions.

Visual assessment of these sites confirmed the maintenance need.

In conclusion, the use of Schema toolkit for hotspot identification was judged to be successful by all parties. The application of the Schema toolkit was found to have the following benefits in this study:

Speed/cost effectiveness

The core analysis can be undertaken relatively quickly (in a matter of days), and considerably more quickly than previous manual methods. The process is therefore very cost effective.



Criteria	Maximise / Minimise	1	2	3	4	5	6	7	8	9	10
Route Hierarchy	maximise	0.206	0.027	0.206	0.027	0.000	0.000	0.111	0.111	0.111	0.111
Traffic	maximise	0.206	0.082	0.206	0.082	0.000	0.000	0.111	0.111	0.111	0.111
Rutting	maximise	0.187	0.083	0.187	0.083	0.263	0.263	0.111	0.111	0.111	0.111
Texture	minimise	0.066	0.044	0.066	0.044	0.117	0.117	0.111	0.111	0.111	0.111
LPV 3m	maximise	0.037	0.020	0.037	0.020	0.053	0.053	0.111	0.111	0.111	0.111
LPV 10m	maximise	0.000	0.000	0.000	0.000	0.000	0.000	0.111	0.111	0.111	0.111
LPV 30m	maximise	0.000	0.000	0.000	0.000	0.000	0.000	0.111	0.111	0.111	0.111
SCRIM	minimise	0.066	0.194	0.066	0.194	0.204	0.204	0.111	0.111	0.111	0.111
Residual Life	minimise	0.037	0.207	0.037	0.207	0.363	0.363	0.111	0.111	0.111	0.111
Accidents	maximise	0.197	0.343	0.197	0.343	0.000	0.000	0.000	0.000	0.000	0.000
	Check sum:	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Description:	BM1	SBF1	BM2	SBF2	DW1	DW2				
	Include in MCA:	No	No	No	No	Yes	Yes	No	No	No	No
	MCA method:	SAW	SAW	WP	WP	WP	SAW	SAW	SAW	SAW	SAW
	Include in Graph:	x	x	x	x	x	x				

Figure 7 - Main screen from Schema user interface

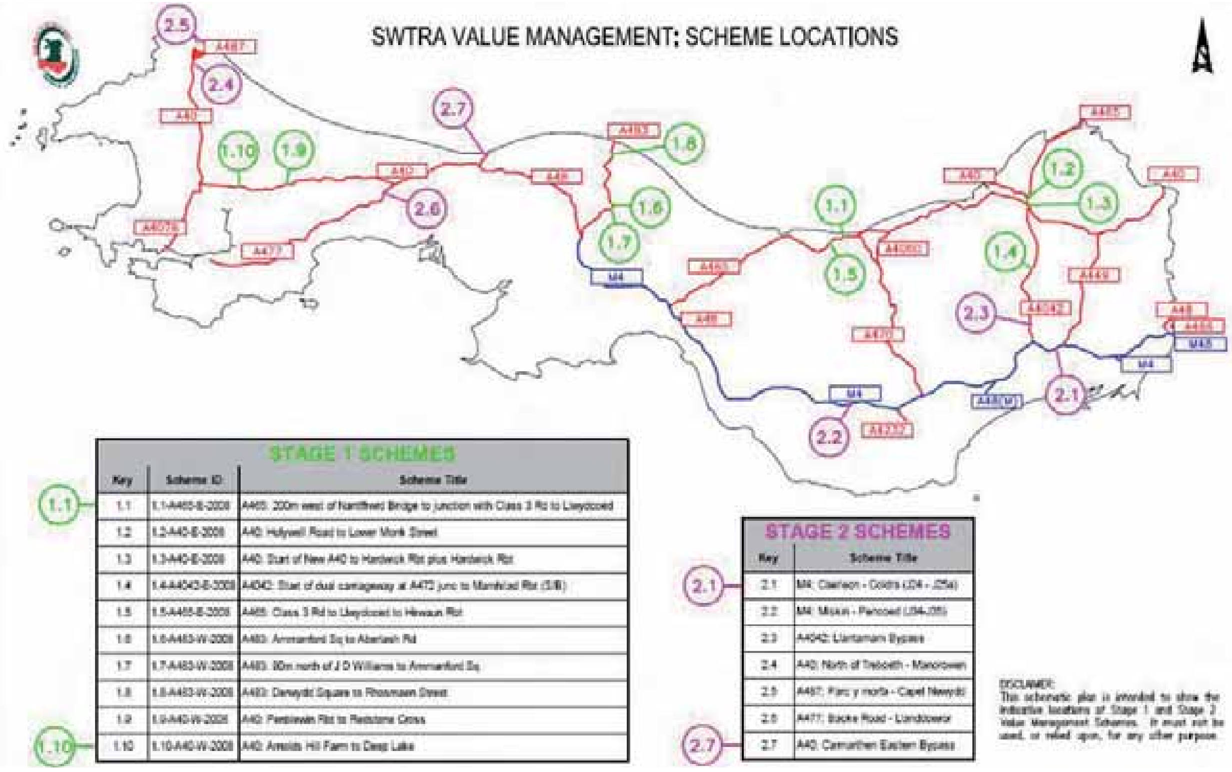


Figure 8 - Indicative plot of proposed Stage 1 & 2 schemes

Robustness

The potential for user introduced errors is minimised, helping to ensure more robust/accurate outputs.

Knowledge encapsulation

A large part of the analytical expertise required to undertake the process is encapsulated in the Schema software. As a result, specialist analytical skills are not needed in order to conduct the analysis. Similarly, much of the engineering knowledge involved in scheme identification is also encapsulated.

Objectivity

The automation of the process reduces the opportunities for manual intervention, and thereby ensures objectivity.

Transparency

The process by which schemes are identified and eventually developed is both open and defendable. A clear link is demonstrated between collected/reported condition data and eventual maintenance work.

Compliments existing practices

The process fits well with existing practices/methods, and thus far cross-correlation is good (i.e. the MCA approach appears to replicate the traditional decision-making process well).

Use of data

Significant cost is involved in the collection and storage of TRACS, SCRIM and Deflectograph data. The use of these data sets for hot spot identification further justifies their collection. In addition, survey data is subjected to a degree of scrutiny that would otherwise not be the case. Errors are identified, and ultimately addressed, that would otherwise go unnoticed.

Added value

In addition to its main function of hotspot identification, the work described in this paper has provided a number of additional benefits to SWTRA including: data quality assessment, network statistics, condition summary sheets etc.

It should be emphasised that the methodology developed during this study is still regarded as a work-in-progress by the Atkins and SWTRA teams.

Although the intention is to incorporate the methodology as an integral part of the annual SWTRA VM process, the project team are keen to further enhance and refine the methodology as we move forward. At the time of writing, the second application of the approach has just commenced.

Acknowledgements

I would like to acknowledge the support of a number of colleagues during the course of the work on which this paper is based. In particular, I would like to thank Steve Finnie, Tom Brownlee, and Chris Walsh of the Highway Asset Management Group at Atkins, as well as Chris Nelson and Alan Brandon of the Asset Management Section at SWTRA.

References

1. HIGHWAYS AGENCY. Traffic-Speed Condition Surveys (TRACS): Revised Assessment Criteria, Interim Advice Note (IAN) 42/05, Highways Agency, UK, 2005.
2. HIGHWAYS AGENCY. Skidding Resistance, HD28/04, Design Manual for Roads and Bridges (DMRB, Volume 7, Section 3, Part 1, Highways Agency, UK, 2004
3. HIGHWAYS AGENCY. Skidding Resistance, HD36/06, Design Manual for Roads and Bridges (DMRB), Highways Agency, UK, 2006
4. HIGHWAYS AGENCY. Structural Assessment Methods, HD29/94, Design Manual for Roads and Bridges, (DMRB), Volume 7, Section 3, Part 2, Highways Agency, UK, 2001
5. SAATY, T. L. The Analytic Hierarchy Process. Pittsburgh, RWS Publications, 1990.
6. YOON K. P. and HWANG C. L. Multiple Attribute Decision Making: An Introduction. Sage, Thousand Oaks, CA, 1995.

Lighting system at the Conwy Tunnel

Wyn Roberts

Senior M & E Engineer

Traffic Wales
Conwy, UK

Abstract

Recent failures of tunnel lighting systems within the Conwy tunnel in North Wales have realised the fact that there is a need for review of current standards relating to materials used in road tunnel mechanical installations.

This paper seeks to identify the cause and analyse the reasons for the mechanical failures and identifies the need to consider more fully the specification of materials in tunnel lighting and other related installations with due regard to the extreme environment to which they are exposed.

Introduction

Background

The Conwy tunnel is an immersed tube tunnel of 1090 metres in length and was opened in 1991 following a five year construction program. It forms part of the Trans European Road Network which is designated as E22 that runs from Holyhead to Ishim in Russia. The tunnel was designed by consulting engineers R.Travers Morgan in association with Christiani & Nielsen. The installed Mechanical and Electrical systems were designed by Mott MacDonald Consultants in association with EIWHS as the M & E contractors.

The tunnel luminaires and associated suspension brackets were supplied by GEC (Street Lighting) Ltd whilst base lighting dimming controls were designed and provided by Helvar Ltd.

Tunnel lighting is affected by a system that is designed and installed to maintain a continuous adequate lighting level and to minimise any optical shock on entering or leaving the tunnel. Each tunnel bore is divided into six zones for the purpose of the lighting installation and are configured as follows:

- Threshold Entry Zone 110m long
- 1st Transition Zone 60m long
- 2nd Transition Zone 80m long
- 3rd Transition Zone 80m long
- Interior Zone 679m long
- Exit Zone 80m long

Two principal types of luminaire are used throughout the tunnel; single tube 58W fluorescent for base lighting and double tube 58W fluorescent and double 400W, 250W, 150W high pressure sodium discharge lamps for boost lighting at the tunnel bore threshold and exit zones. 1795 luminaires in each bore are mounted in up to three rows along the tunnel haunches on adjustable suspension brackets. Base lighting levels are controlled by a three stage dimming system that allows fluorescent type luminaires to be operated at 30%, 60% and 100% illumination. Boost lighting is operated on a switched basis with provides five additional stages of lighting.

Lighting control is affected through switched outputs from a control PLC in response to photometer readings. Stage 1 lighting is controlled via a solar clock which activates night time illumination between a period 30 minutes before sunset and 30 minutes after sunrise. The remaining stages respond to the outside luminance values at set thresholds.

Each luminaire type is constructed of an extruded HE9 silicon magnesium aluminium alloy carcass with die cast LM6M silicon aluminium alloy end caps that allows for access to the internal components and provides a termination point for power cables. Internal components consist of a gear tray, reflector, lamp and associated terminal blocks and cabling. The aluminium carcass and end caps are protected by a 65 micron thick epoxy powder coating for additional corrosion protection and have an ingress protection rating (IP) of IP66. Each luminaire is suspended from adjustable stainless steel support brackets and fixings which in turn are secured to the high level tunnel steelwork. The general arrangement is shown in Figure 1.

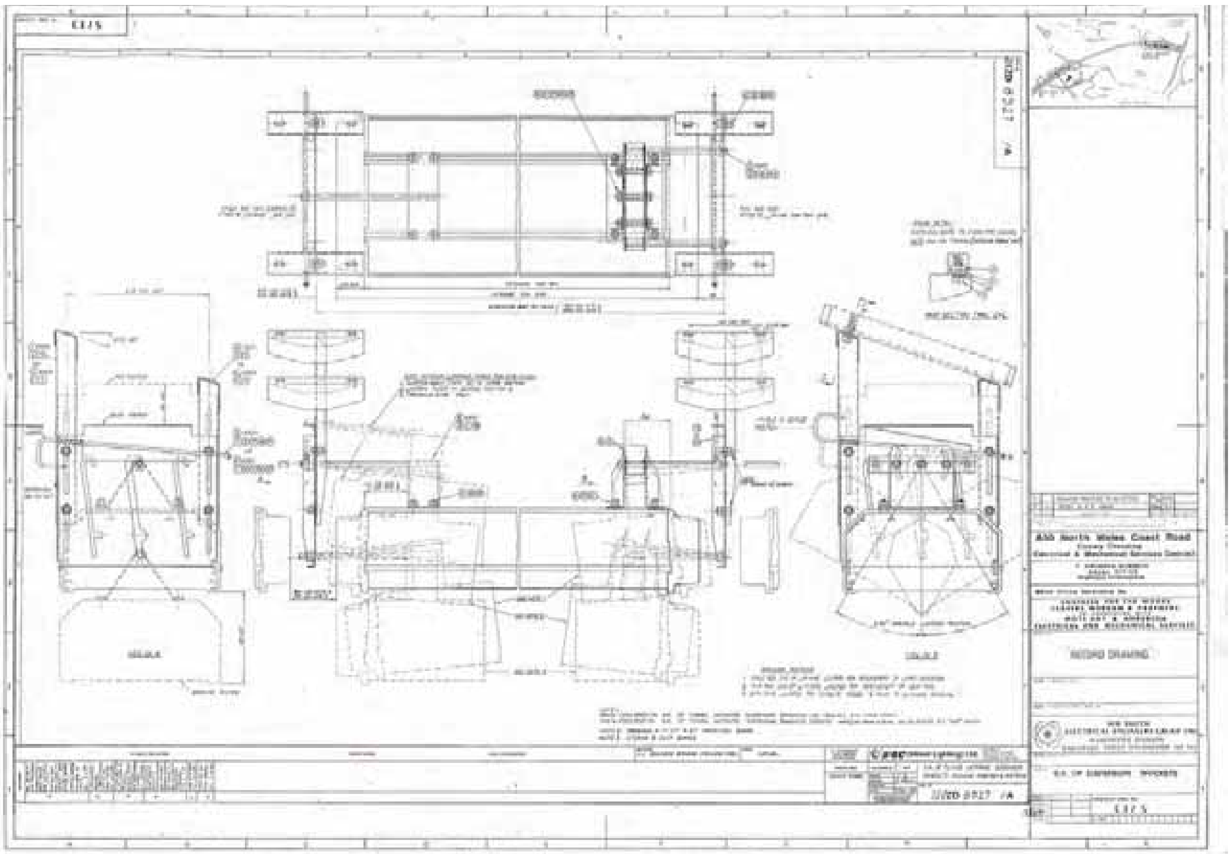


Figure 1 - General arrangement of assembly

The expected serviceable life of tunnel luminaires and associated control gear designed to this type of construction is expected to be in the region of 18 years (ref. BD78/99) where serviceable life is defined as the period of duty after which replacement, rather than continued use, is anticipated to be justifiable on an economic or operational basis, either due to a greater risk of failure or to an unacceptable increase in unreliability, operating and maintenance costs. Precise estimates of service life are not possible, since external influences (such as traffic changes, legislation changes, obsolescence, technology advances etc) may have a significant impact on the achievable life of tunnel equipment. The service life figure is considered as guidance for typical values of current UK road tunnel equipment and values need to be adjusted according to local conditions, better or worse than the average.

Over the last 24 month period it has become evident that failure rates associated with the tunnel luminaires have increased both in terms of the internal electrical equipment and external mechanical components. These have manifested themselves by way of increasing corrosion of the aluminium housing leading to mechanical failure and the loss of light output due to control gear problems.

The planned maintenance regime in place caters for integrity inspections to be carried out on the luminaires and associated support brackety on a 3 monthly basis and has identified the increasing problems related to the latter.

Recent Events

From handover to the operating authority in 1991 it has been observed that water leaks emanating from cracks in the tunnel roof structure have been impinging on the top of a number of luminaires which affected the supporting arrangement by way of attack from the seawater. A regime of fitting stainless steel drip trays to the tunnel roof was instigated in 1994 in an attempt to channel away any such leakage from the affected light fittings. Alternative drip tray designs were fitted directly onto the luminaires support brackets to reduce the incidence of drilling the tunnel structure. This activity continued until a programme of crack sealing was instigated by the Trunk Road Agency Structures Group which appeared to reduce dramatically the occurrences of water impingement onto the fittings.

Two recent instances have been observed by maintenance personnel transiting the tunnel whereby two luminaires, one in each tunnel bore appeared to be hanging off one of their supporting brackets. The luminaires, however, remained attached to the structure by the remaining bracket. On each occasion operators at the North Wales Traffic Management Centre (NWTMC) informed the Regional Maintenance Contractor (RMC) and North Wales Trunk Road Agency (NWTRA) Route Steward who arranged police rolling blocks in order to access the tunnel bores for removal of the affected luminaires. The rolling blocks were supplemented by the use of the tunnel VMSS, approach EMS and traffic signals where applicable. High level access was gained via a lift platform and it was clear on each occasion that one supporting arm that is secured to the luminaire body had failed.

Figure 2

Figure 3

Figure 4

The affected units were removed by the RMC and the tunnel vacated safely. Replacement units are to be fitted at the next available maintenance closure.

The luminaires were removed to the Welsh Assembly Government (WAG) stores where inspection of the units took place and photographic images taken for record purposes (see Figure 2 - Figure 4). On inspection it was evident that corrosion between the luminaire bodies and the stainless steel bolts securing the support arms to the carcass had taken place causing the bracket to detach from the body. This is attributed to galvanic corrosion due to the presence of two dissimilar metals in contact, in this instance aluminium and stainless steel.

In addition to the above failures this report seeks to highlight the problems associated with increasing electrical component failures associated with the tunnel luminaires that are a cause for concern. Historical data indicates that increasing failures of the lighting system supports the need to upgrade/ replace the existing installation.

Technical review

Failure Modes

Electrolytic/Galvanic Corrosion

Galvanic corrosion is an electrochemical process in which one metal corrodes preferentially when in electrical contact with a different type of metal and when both metals are immersed in an electrolyte. When two or more different sorts of metal come into contact in the presence of an electrolyte a galvanic couple is set up due to the fact that different metals have different electrode potentials. The electrolyte provides a means for ion migration whereby metallic ions can move from the anode to the cathode. This leads to the anodic metal corroding more quickly than it otherwise would and corrosion of the cathodic metal is retarded even to the point of stopping.

Metals (including alloys) can be arranged in a galvanic series representing the potential they develop in the presence of an electrolyte against a standard reference electrode.

The relative position of two metals on such a series gives a good indication of which metal is more likely to corrode more quickly. When two metals are submerged in an electrolyte, while electrically connected, the less noble (base) will experience galvanic corrosion. The rate of corrosion is determined by the electrolyte and the difference in nobility. The difference can be measured as a difference in voltage potential. In effect the further apart the metals are in the series the greater the extent of potential corrosion.

The following is the galvanic series for a similar environment.

(NOBLE)

- Graphite
- Palladium
- Platinum
- Gold
- Silver
- Titanium
- Stainless steel (316 passive)
- Stainless Steel (304 passive)
- Silicon bronze
- Stainless Steel (316 active)
- Monel 400
- Phosphor bronze
- Admiralty brass
- Cupronickel
- Molybdenum
- Red brass
- Brass plating
- Yellow brass
- Naval brass 464
- Uranium 8% Mo
- Niobium 1% Zr
- Tungsten
- Stainless Steel (304 active)
- Tantalum
- Chromium plating
- Nickel (passive)
- Copper
- Nickel (active)
- Cast iron
- Steel
- Lead
- Tin
- Indium
- Aluminum
- Uranium (pure)
- Cadmium
- Beryllium
- Zinc plating (see galvanization)
- Magnesium

(BASE)

It can be seen from the series that there is a significant difference between the nobility of stainless steel and aluminium. In the presence of a humid atmosphere where additional elements such as sulphur, carbon, sodium chloride etc are present as is the case of tunnels, the rate of electrolytic corrosion will be accelerated. Table 1 gives a good indication of where the existing construction materials for the luminaires are positioned with regard to their suitability in such environments.

In most surroundings - inside or in the open air - aluminium has a very high resistance to corrosion. One reason for this is that aluminium is covered spontaneously by a thin but effective coating of oxide which protects the aluminium from further oxidation. The aluminium oxide that is formed is impermeable and is integral to the base metal. If the coating of oxide is damaged mechanically this coating will be renewed immediately. The coating of oxide is a main reason for the improved corrosion characteristics of aluminium. The coating stays stable for pH-values between 4 and 9. Generally alloys with more than 0.5% copper have a worse resistance to corrosion and they should not be used in unprotected, strong chlorine surroundings.

The most frequent types of corrosion are;

- Galvanic Corrosion
- Pitting
- Crevice Corrosion.

A small surface of the cathode and a large surface of the anode lead to a low corrosion. In reverse case the aluminium will be attacked quickly. In most combinations with other metals, aluminium is the less noble. Aluminium is, therefore, of greater risk of galvanic corrosion than that of the other constructive materials.

These dangers of galvanic corrosion only exist in metallic contact with more noble metals (or other conductors of electrons, for example graphite) and then when an electrolyte with good conductivity exists between the metals. The occurrence of galvanic corrosion will be encouraged because of a disadvantageous construction of the extrusion.

Galvanic corrosion does not happen in dry surroundings; however, contrary to this, in surroundings with chloride the risk of galvanic corrosion always exists, for example near the coastline. However, galvanic corrosion can be prevented if certain steps are taken; one being by providing electrical isolation between the metals, where the insulation has to interrupt the metallic contact completely.

In large constructions where electrical insulation is difficult it is possible to use an electrolyte insulator between the two metals; for example by painting to interrupt the connection. Often it is beneficial to conceal the surface of the cathodes (that of the nobler metal); however, another possibility is the installation of an intermediate layer.

Another method of protection is cathodic protection. Cathodic protection can be achieved in two ways. Often anodes, consisting of less noble metal can be in contact with the surface of Aluminium that has to be protected. Within this process the less noble metal is sacrificed (it is corroding) and is therefore called a sacrificial anode. Often Zinc or Magnesium anodes are used for protecting aluminium. Cathodic protection can be reached by an exterior constant-potential supply and by connecting an aluminium object to the negative pole.

Pitting is a frequent type of corrosion of Aluminium and occurs due to the presence of an electrolyte. The corrosion is seen as very small pits and only penetrates a small way into the surface when induced in air; however, in water and earth larger pits can occur. As residues of corrosion often cover the small pits the attacked places in the surface of Aluminium are rarely visible. Small pits are mainly a problem of appearance and have no negative influence on the temper; however, treatment of the surface, anodising and painting can prevent this corrosion.

Table 1 - Compatibility Chart

Metal Corroding / Contact metal	Magnesium and alloys	Zinc and alloys	Aluminium and alloys	Cadmium	Steel-carbon	cast iron	Stainless steels	Lead,tin and alloys	Nickel	Brasses, nickel,silvers	Copper	Bronzes,cupro nickels	Nickel copper alloys	Nickel-chrome-molybdenum alloys,titanium silver,graphite,gold,platinum
Magnesium and alloys		X	X	X	X	X	X	X	X	X	X	X	X	X
Zinc and alloys			X	X	X	X	X	X	X	X	X	X	X	X
Aluminium and alloys				X	X	X	X	X	X	X	X	X	X	X
Cadmium					X	X	X	X	X	X	X	X	X	X
Steel-carbon				X		X	X	X	X	X	X	X	X	X
Cast iron					X		X	X	X	X	X	X	X	X
Stainless steels					X	X		X	X	X	X	X	X	X
Lead,tin and alloys							X		X	X	X	X	X	X
Nickel										X	X	X	X	X
Brasses, nickel, silvers						X	X	X	X		X	X	X	X
Copper							X	X	X			X	X	X
Bronzes,cupro nickels													X	X
Nickel copper alloys														X

Crevice corrosion can arise in small crevices filled with fluid; however, this rarely happens in extruded constructions and this risk is increased in marine atmospheres. It can happen during transport and storage that water gathers in the crevices between Aluminium surfaces opposite to each other and this can lead to corrosion on the surface (water spots).

Methods of reducing corrosion resulting from galvanic corrosion include:

- Selection of materials that are close together in the galvanic series for the relevant environment
- Insulation of metals from each other
- Exclusion of electrolyte from around the bimetallic junction;
- Painting of both metals where possible – if impractical paint the most noble metal
- Provision of additional corrosion allowance on the less noble metal
- Application of compatible metal or sacrificial metal coatings.

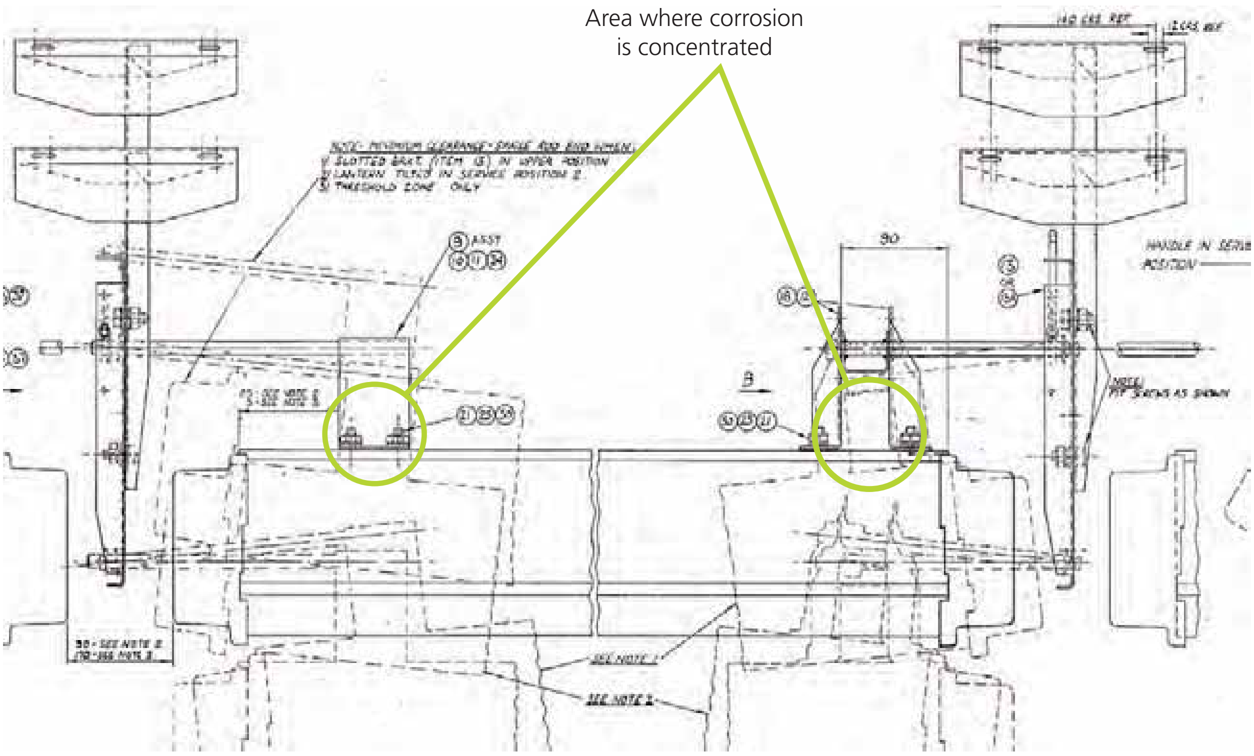


Figure 5 - General arrangement of luminaire support

It is evident from inspection of the failed units that none of the above methods were implemented as part of the original design.

Luminaire Control gear failure

The luminaire control gear associated with each unit consists of a single gear tray on which the ballasts, igniters, power factor correction capacitors, fuse links and associated wiring is located. Other electrical components of the luminaire consist of a cable distribution tray, lamp holders, and reflector and terminal blocks for cable terminations.

Failure modes are mainly limited to ballast units which control current flow. A lighting ballast is necessary to operate discharge lights because they have negative resistance, meaning they are unable to regulate the amount of current that passes through them. An electromagnetic lighting ballast uses electromagnetic induction to provide the starting and operating voltages of a gas discharge light.

Failures will mainly be attributed to the re-striking effect during lamp failure which puts additional stress on the ballast unit.

Under normal circumstances where lighting units can be accessed readily then failure rates can be reduced by changing a failing lamp.

However, as access is only available to maintain the lighting system within the tunnel bore on a 3 month basis then it becomes impracticable to carry out reactive maintenance on an individual basis.

Failure rates increase during the summer period due to the increased demand on the boost lighting - whereas rates decrease during periods of low boost lighting level demands such as is the case during the winter periods.

The available data shows that failure rates attributed to the base lighting dimming system are very low and whilst the equipment associated with the dimming controls are no longer supportable by the manufacturer the low failure rates do not cause any immediate concern as there are an adequate level of spares held by WAG within the North Wales stores. Additionally, repairs and maintenance support to the equipment are carried out by specialist contractors (Electricly Ltd) engaged by the RMC.

Analysis of failures

Corrosion

The luminaires consist of an extruded aluminium body (silicon magnesium aluminium alloy grade HE9 TE) to which are mounted two 316 grade stainless steel mounting brackets that are secured together with eight M6 stainless steel bolts, nuts and washers. The securing bolt head is located within a channel formed as part of the aluminium extrusion of the luminaire body and clamps both the bracket and luminaire together. The arrangement is shown in Figure 5. Whilst the aluminium is powder coated for additional protection the action of tightening the nut and bolt arrangement inevitably damages that protective barrier exposing the metal and ultimately results in the dissimilar metals achieving contact. In the presence of an electrolyte this sets up the electrolytic cell and allows the less noble metal (aluminium in this case) to corrode.

The bolt head sitting within the channel has approximately 2.5-3.0mm of contact area on each side of the hexagonal cross section of the bolt head as can be seen in Figure 6.

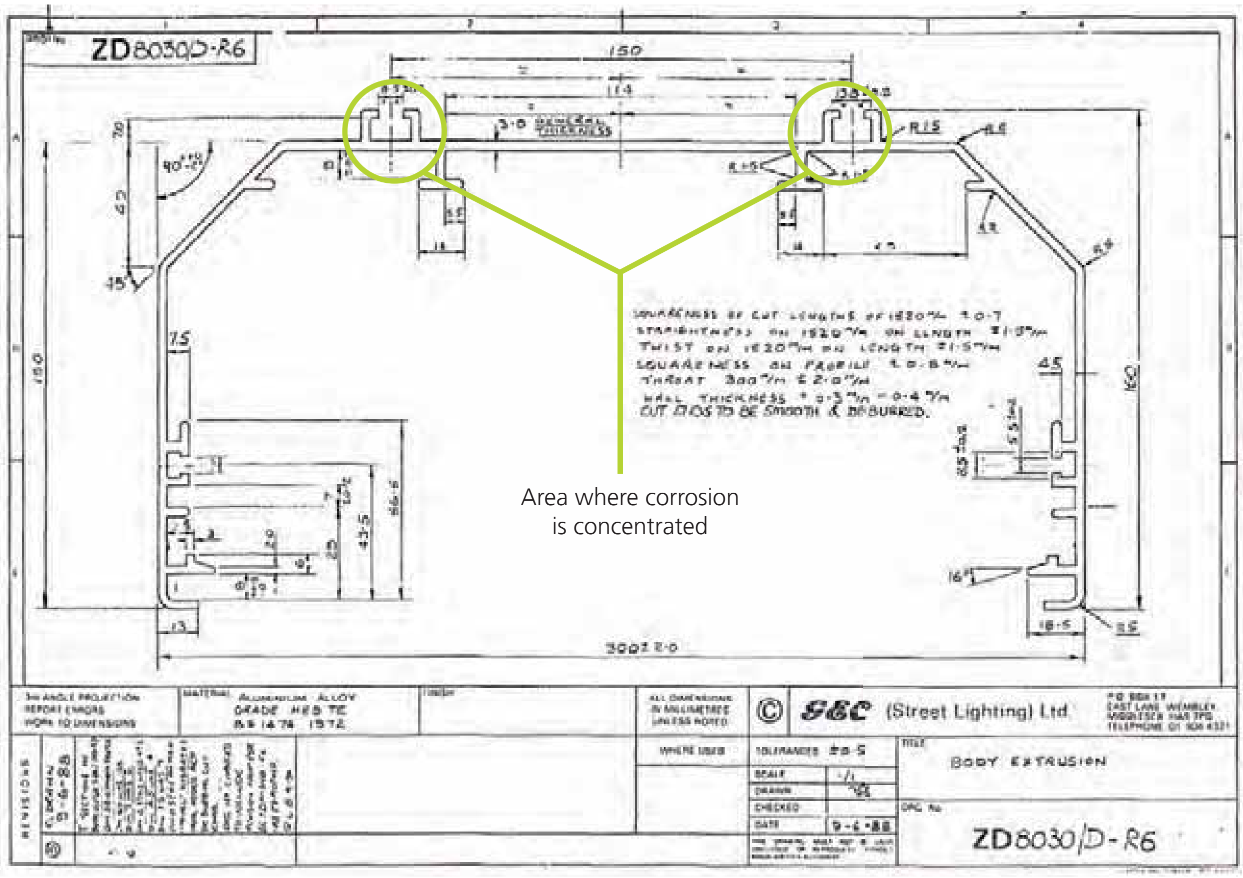


Figure 6 – Cross Section of luminaire carcass

This does not allow for much load distribution over the contact area and corrosion will rapidly degrade the aluminium to the extent that the bolt head will loose any securing effect resulting in the bracket and body parting company with resultant failure as was observed on the occasions of the 1st and 21st August 2008.

Over the period that the lighting system has been in service, there has inevitably been a general degradation over time of the powder coating, again leading to the ideal conditions for corrosion to take place. This is particularly evident on the attachment points located adjacent the end caps securing clips.

As would be expected the majority of bracket failures have been located at the heaviest end of the luminaire, i.e. the end that houses the gear tray. This is particularly the case for the SON type discharge lamps as the gear trays weigh approximately 15kg.

Aluminium has a good durability against many chemicals and in this instance the grade of aluminium alloy used is HE9 TE to BS1474: 1972 designated as a 6063 series alloy. This exhibits good corrosion characteristics as opposed to the lower series alloys. However, low or high pH-values (less than 4 and more than 9) can lead to dissolution of the coating of the oxide and to a rapid corrosion of the Aluminium. Inorganic acids and strong alkaline solutions can then easily attack Aluminium. Recent studies carried out in the A55 tunnels showed that tunnel washing chemicals in dilution with water had pH values in the region of 13 and appears to be a key contributory factor to the high rate of corrosion experienced. Electrolyte conductivity along with elevated temperatures also plays a major part in the degree to which corrosion increases. In this instance sodium chloride and sulphuric acid no doubt exists due to the presence of airborne salt and vehicle pollutants respectively. The aluminium extrusion is subject to elevated temperatures; particularly the high power SON types which can reach up to 60-700 C.

Whilst general 3 monthly inspections of the high level steelwork are carried out as part of the M & E routine planned maintenance, close inspection of the fixing arrangements of the brackets to the luminaires are difficult due to the design of the fixing arrangements. As a matter of course where extreme corrosion can be seen, then replacement takes place but it is very difficult to make an assessment of the amount of corrosion evident without complete disassembly of the supports.

Gear failure

Over a period of approximately 24 months the failure rates of components of the luminaire gear trays have increased considerably; the main component subject to failure being the ballast units in the high pressure sodium (SON) luminaires. Whilst the failures cannot be attributed directly to a single problem, the age and service life of the system is clearly a major factor.

The SONs are re-lamped on a bulk basis every 5 years and it is possible that lamp service life may be contributing to the failure of the gear tray components.

On each occasion of lamp failure the control gear may be burning out through the continual re-striking of the lamp for more than a few hours. Unfortunately the luminaires are inaccessible for periods between routine tunnel closures to address such problems; therefore, immediate lamp replacement on failure is not possible or cost effective.

Statistics

Failure rates

Data collated from historical maintenance records indicate the different failure rates of the lighting system reported to date. Clearly as shown in Figures 7 and 8 on the following page the data indicates an increase in the problems attributed to gear tray failures with a rising trend from 1997 onwards. This would appear to be in accordance with the well known maintenance "bathtub" failure model whereby increased rates of failure are seen at installation and end of the service life of the equipment. Whilst the same model cannot be directly applied to the corrosion problems it is evident that over the service life of the luminaires to date that the number of failures has accumulated over time and that we are now seeing a period where increased failures will become evident. Looking at the fault data it would appear that a number of the units failing due to corrosion were or are located at sites within the bore where water leakage from the tunnel ceiling is evident. These may be attributed to new or recent leakage points; however, data is not readily available to support the fact.

The number of failures due to corrosion over the past 12 years is:

Eastbound bore
48 - which equates to 2.7% of all units

Westbound bore
27 - which equates to 1.5% of all units

Rationalising these figures gives an average failure rate per annum of 0.2% of all fittings in the Eastbound bore and 0.125% in the Westbound bore.

The number of gear tray failures over the past 12 years is:

Eastbound bore
240 - which equates to 13.4% of all units

Westbound bore
223 - which equates to12.4% of all units

Rationalising these figures gives an average failure rate per annum of 1.1% of all fittings in the Eastbound bore and 1.0% in the Westbound bore.

As average values these figures do not indicate a significant rate of failure; however, failure rates are increasing over the period and indicate a trend of increasing rates of failure in terms of continued service.

The graphs opposite clearly show an increasing trend of failure due to gear tray problems both for the eastbound and westbound bores whilst a slight increase is seen in the trend of luminaire failure in terms of the corrosion problem.

Maintenance costs

Available data indicates that the total number of reported faults for the North Wales tunnels for the period April 2007 - March 2008 was 56 with attributed faults costs for tunnel lighting of £44k of which an estimated £30k is directly attributable to Conwy.

Options

Short term remedial options

Options

High level inspection of the tunnel supporting steelwork is carried out at every 3 monthly routine planned tunnel closure as part of the M & E maintenance requirements whereby the integrity of high level steelwork within the tunnel is checked.

Options available to address any further failures of the supporting brackets would be limited to those that could be carried out within a planned maintenance closure. This would mean supplementing the bracket support arrangement with a form of banding possibly affected through the use of heavy duty plastic or metallic cable tie wraps.

A band secured around the ends of each fitting would provide a secure anchorage for a second band which would be attached to the luminaires support bracketry. Further failures as seen recently would be averted due to the additional support. It is recommended that heavy duty plastic bands be utilised as metallic elements may react further with the existing installation.

Due to the high level of heat generated by each luminaire, particularly the higher power SON type discharge lamps, any banding utilised would have to be temperature rated for such service. Additionally, the effects of the light output of the luminaires on the bands would have to be considered and selected accordingly.

Gear tray failures will continue to occur due to various reasons; however, both tunnel bores boost lighting luminaires were re-lamped in 2007 and therefore a reduction should be seen in the number of gear failures associated with the high pressure sodium (SON) luminaires.

Costs

It is envisaged that an 8 man team utilising 4 access platforms could complete the task for each bore within a normal planned closure period of two nights. An additional extra night may be required but is dependant on the progress made. The figures below indicate an estimate of costs based on engineering judgement for implementing the temporary solution on a per bore basis.

Labour	£5,000
Plant	£2,500
Materials	£2,000
TOTAL	**£9,500 (per bore)**

Timescales

The works could be carried out during the normal planned tunnel closures and could be affected at the next available opportunity. The available tunnel closure dates for 2008 are:

Conwy Westbound – 9th September 2008 and 25th November 2008

Conwy Eastbound – 14th October 2008 (2009 closure dates yet to be advised)

HIGHWAYS

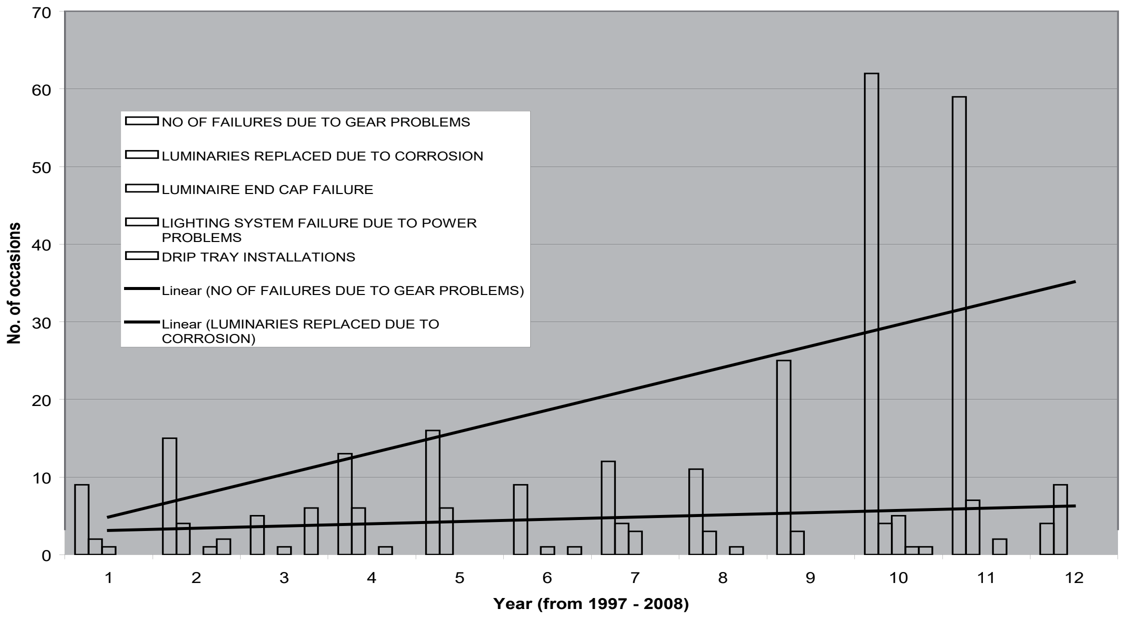


Figure 7 - Conwy Eastbound bore tunnel lighting failure rates

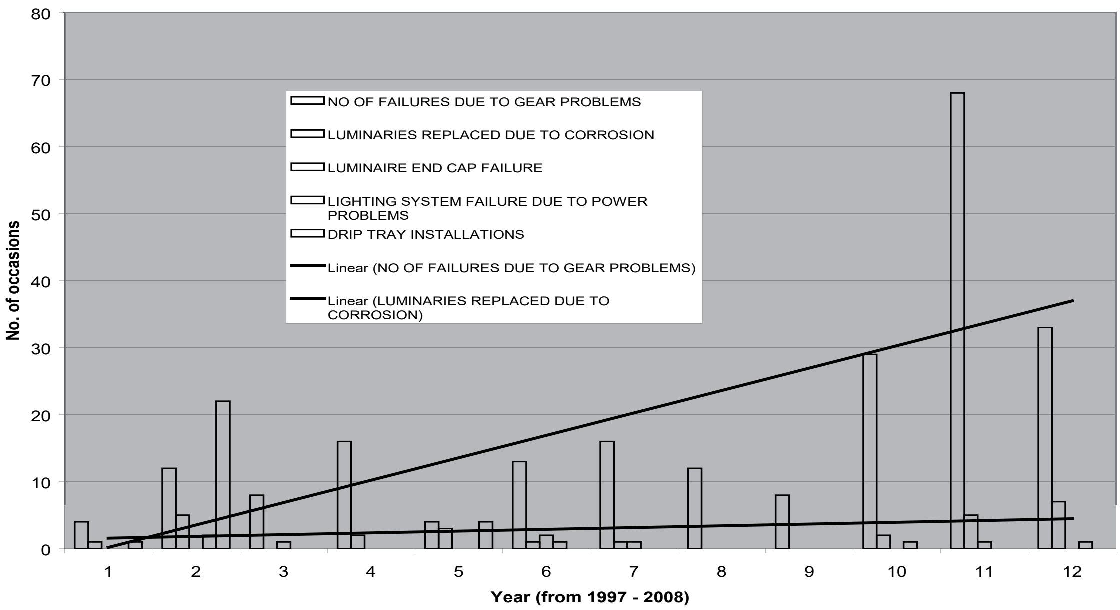


Figure 8 - Conwy Westbound bore tunnel lighting failure rates

Future proofing

Clearly this would only be seen as a temporary arrangement until such time as the Client decides as to the way forward given available options.

Concerns

The act of fitting a banding system must be seen as a temporary solution requiring that any future maintenance activity involving the dropping of the unit to its maintenance position will necessitate the banding to be removed and ultimately replaced.

Therefore, the banding activity would have to be repeated every time a lamp and / or gear tray replacement is carried out and any programmed bulk lamp changes in which case the above costs will be repeated.

The next due dates for tunnel bulk lamp changes are:

Eastbound and Westbound bores base lighting - early 2010 and then at 2 yearly intervals.

Eastbound and Westbound bores boost lighting - early 2012 and then at 5 yearly intervals.

Tunnel base lighting forms 68% of the installed lighting in each bore with the remaining 32% being boost lighting. Therefore, unless the system is upgraded the activity suggested above would have to be repeated at the stated intervals with the associated costs in addition to those associated with the lamp change activity.

Longer term technical solution

Options

The aim of this report is to comment on the current condition of the lighting system and to provide practical solutions with estimates for maintenance replacement. Whilst the aim of this report does not seek to address major improvements in either the design or operation of the system, recent advances in tunnel lighting technology and control means that various options to upgrade are available.

In terms of direct maintenance replacement luminaires then the options are fairly limited particularly taking into account the fact that use should probably be made of the existing supporting steelwork and bracketry which are structurally sound. If that is the case then direct replacements are available; however, serious consideration should be given to units manufactured or designed to alleviate the problems experienced with galvanic corrosion. In effect either a 316 grade stainless steel fitting (metallic units conduct the internally generated heat) or a similar aluminium alloy (as fitted) with improved bracket fixing details in terms of a change in material to ensure isolation of the fixings and extrusion should be considered.

Costs

A conservative estimated cost for direct replacement units utilising the existing supporting steelwork would be:

Material costs (luminaires)
= £2,000,000 (per bore)

Material costs (other items)
= £50,000 (per bore)

Labour (installation)
= £100,000 (per bore)

Plant
= £50,000 (per bore)

TOTAL
= £2,200,000 (per bore)

Timescales

The estimated costs above have been based on a working team of eight, two man teams with associated working platforms.

The installation period would take, given that the works could be carried out during normal two night tunnel closures, ten closure periods (a total of 2.5 years assuming concurrent installations in each bore). Clearly, a concentrated period whereby a tunnel bore could be shut would significantly reduce the overall installation activity to the order of 3 – 6 weeks.

Life expectancy

Given that a new installation designed to include for luminaires constructed of materials resilient to the aggressive atmosphere within the tunnel there is no reason not to expect a reasonable service life. Figures quoted by manufacturers for available luminaire designs vary from 18 -25 years.

Summary

It is evident that the system is rapidly approaching the end of its service life both in terms of mechanical and electrical serviceability as demonstrated by the increasing failures highlighted within this report. Guidance is available within BD78/99 as to the expected service life of such an installation and the data supporting the “bathtub” failure model supports the fact that the expected service life of approximately16 years has been reached and a period of increasing failure rates is now likely to ensue. It is not evident as to the basis of the derivation of the service life quoted in BD78/99 but factors such as design; installation methods, ambient conditions and maintenance must surely be prominent. Manufacturers’ guidance on service life available in the Operation and Maintenance documentation suggest replacement periods of 15 years for the mechanical elements and 10 years for electrical components.

Whilst the design of the installed GEC luminaire incorporates two channels as part of the aluminium extrusion body to cater for the use of standard fixings to allow the supporting brackets to be fitted there seems to have been a lack of forethought given during the design as to the effect of using dissimilar metals in such an installation.

The use of insulating materials in an attempt to reduce the contact areas or the use of less aggressive materials (e.g. magnesium as a substitute for stainless steel) does not appear to have been considered. This coupled with a damp aggressive atmosphere would appear to be the major factor in the corrosion of the extrusion.

It can be assumed that the electrolytic corrosion between the aluminium alloy carcasses and the stainless steel fixings is widespread throughout the tunnel to varying degrees but ascertaining the true extent of the corrosion has historically been difficult due to the orientation of the units and the bracket fixing details. At this point in time retrospective use of a modified fixing detail utilising either different materials or insulators should not be considered as a solution as the re-working required would be extremely expensive and time consuming. The extent of the degradation of the extrusion would probably limit any viable practical refurbishment.

Direct replacement of the existing luminaires with modern designs should be considered to future proof the installation. Whilst due consideration should be given to emerging technology within the tunnel lighting field the utilisation of existing luminaire bracketry and the use of standard replacement luminaires (albeit of different materials or fixing design) would reduce overall installation costs. In consideration of the latter, then care should be exercised in the selection of construction materials of replacement units so as not to compromise the designed loading subjected to the tunnel steelwork and ultimately the structure. Aluminium luminaires continue to be manufactured and installed in tunnels but the use of less aggressive materials are widespread. If aluminium were to be the choice of material for the extrusion then a likely replacement material for the stainless fixings would be magnesium or cadmium.

In terms of the electrical elements of the system, fault statistics support the fact that the existing Helvar base lighting dimming controls continues to give relatively trouble free service with very few faults.

However, supportability of the equipment has been suspended by the manufacturer which will inevitably impact on the continued future service of the system. Unless the manufacturer can provide direct replacement equipment to upgrade then the following options remain available to ensure continued service of the system:

(1) Consider the re-design of the base lighting to allow switched staging rather than dimmed.

(2) Replace existing dimming system with upgraded equipment.

(3) Persevere with the existing dimming system.

Option 1) would require major re-design of the system and would inevitably result in substantial costs.

Option 2) would be relatively easy to implement assuming that the existing operational philosophy is applied.

Adequate spares are held for the system and the future operational service could be extended beyond its normal serviceable life given that a good level of routine maintenance continues.

The increasing failures attributable to the components that make up the lighting gear trays particularly for the high pressure sodium boost lighting lies within the fact that all electrical and mechanical components have a finite life and it is clear in this instance that the compounding effects of individual failures are causing the luminaires to fail. As referred to previously in this report a large number of failures may be attributed to early failure of SON lamps, subsequently causing the ballasts to fail due to the action of re-striking. In order to address the problem it is not practical to replace lamps on an individual failure basis due to access. The boost lighting running hours are within the quoted service life; however, the action of switching the units on and off reduces that life considerably as opposed to dimmable units. In an attempt to lessen premature failures consideration may be given to reducing the re-lamp frequency, but this would however, incur increased maintenance costs. For example, a bulk re-lamp of the boost lighting will cost in the order of £15k per bore. Alternatively, implementation of a dimmable system should be considered.

In order to ensure compliance with new Tunnel Regulations and European Directives in respect of tunnel design, operations, maintenance and safety a compliance survey carried out by external consultants on behalf of NWTRA has identified the need to extend the existing threshold lighting zones along with the provision of contra flow lighting. Given that any modifications carried out to ensure compliance with the regulations would have to integrate with the existing lighting system it would appear reasonable to address both requirements with a complete lighting upgrade. Some of the costs associated with a system upgrade could be recouped via the scrap value of the aluminium extrusions and end caps to an estimated value of £20k based on current market value of scrap aluminium.

Whilst the content of this report is limited to addressing the failures associated with the Conwy tunnel lighting installation, available fault data does not support the fact that the corrosion problems are inherent to the same degree in the Penmaenbach and Pen y Clip tunnels. However, increasing failure rates are becoming evident in relation to luminaire electrical components at those sites. The luminaire designs for the Penmaenbach and Pen y Clip tunnels are of comparable specification to the Conwy units albeit manufactured by Thorn and Phillips respectively and while failures are increasing slowly they do not exhibit the same rates of failure as Conwy. This is contrary to what would be expected given the exposed nature of the tunnel bores at Penmaenbach and Pen y Clip. Ultimately replacement of the lighting systems at the latter tunnels will have to take place, but, given available data do not pose an immediate cause for concern at present.

References

1. Oldfield,J.W; Electrochemical Theory of Galvanic Corrosion; 1988; pp 5-22.
2. National Physical Laboratory; Guide to Good Practice in Corrosion Control; 1982.
3. Diamant,R.M.E; Applied Chemistry for Engineers;1972.

Better practice in supplying water to the poor in global Public Private Partnerships

Julian Jacobs
Business Analysis
Water & Environment
Atkins

Dr Richard Franceys
Senior Lecturer
Cranfield University

Abstract

Amidst the general controversy surrounding Public Private Partnerships in lower-income countries there has been little consideration of the successes and failures of the international private operators in serving the urban poor, a subset of the overall task in any concession contract.

This paper reports on research undertaken over a number of years in target communities, completed recently through private operator head office interviews, investigating the techniques and approaches used to serve the slums with clean water.

Following the demise of the PPP 'experiment' and with the present billion slum dwellers projected to increase to two billion over the next generation, there is a clear need for these 'better practice' experiences to be made accessible, and used, by the public sector operators for whom serving the poor is the only real justification for their existence – the rich always find a way to access what they need.

Introduction

This paper documents the policies and practices of multi-national water companies in order to investigate and draw out better practices, ideally 'best practices' though that is yet to be proven, in serving the poor in lower-income, 'developing', countries. At its peak, the global water privatisation market accounted for about 5% of the urban population in lower-income countries. The focus was almost exclusively on the urban context and has been dominated by the three French operators, Suez, Veolia and Saur[1] with varying inputs from English operators such as United Utilities, Thames Water, Anglian Water and Severn Trent Water.

The reason for involving the private sector, in what had been a near universal public service (for those able to access the pipe network service that is, recognising the major role of the private 'alternative' providers to the poor), was generally believed to be access to private finance and operational skills. Expertise in serving slums was not seen as significant initially.

However, the financing justification diminished as the international operators, as all businesses have to, chose to access the cheapest finance available which in this case was either retained earnings or the multilateral financing institutions. The multilaterals were doubly pleased. They had an apparently more effective operator to lend to, with a greater hope of a direct pay-back of their loans. They also had clients who in many cases learnt to use the same language that, in public, justifies their very existence: 'IBRD & IDA: Working for a World Free of Poverty'.

There is no simple or precise definition of what constitutes 'the poor' based on a single variable such as income or place of residence. The poor are no more a homogeneous group than the middle classes or football supporters. It is more appropriate, instead of a definition, to set out some attributes which are associated with the poor[2,3,4] as this also highlights some of the challenges for water utilities in serving them: no access or illegal access to mainstream piped water supply; poor-quality and/or insufficient water resources; residing in informal or unplanned settlements; high density housing; difficult topography; lack of infrastructure (e.g. roads, postal delivery service); lack of legal status and security of land tenure; irregular income and/or wide range of levels of income. Understanding the need for any customer focused organisation to segment its target audience in order to meet their needs, Franceys & Gerlach[5] recognise five useful categories, the vulnerable lower middle-income, the developing poor, the coping poor, the very poor and the destitute. The understanding is that with appropriate service approaches it is possible to incorporate the developing, coping and very poor into networked service coverage, delivering cleaner water at a massively reduced cost to the poor consumers.

The motivation of private companies to work with and serve the poor is often called into question[6,7,8] though there are a number of valid reasons why they would do so. These include contractual obligations to achieve specified service coverage, thereby including the poor, the profit motive as the poor can be such a substantial part of the urban population (typically between 30% and 60%), a sign of competence in what had been perceived to be

Figure 1 - The challenge of the peri-urban areas, La Paz, Bolivia (Photo: Franceys)

a potentially lucrative market in lower-income countries, corporate social responsibility, professional responsibility, optimisation providing surpluses that can be re-directed to new connections, and various other political or financial motives.

Background

There is a paucity of academic research identifying best practice in serving the poor in global water Public Private Partnerships. Furthermore, of the practice that has been identified, nearly half of the literature is normative in approach, proposing what should be done rather than based on documentation of actual experience.

It appears that the rush to condemn privatisation as a policy has created a barrier to investigate or discuss individual practices of private companies, possibly through fear of being seen to legitimise PPPs. Fuest & Haffner[9] put it thus: "After this analysis [criticising PPP contracts failing to meet the needs of the poor], the consideration of technical solutions may be perceived as merely addressing a proximate cause of a problem."

The research undertaken for this study has included data collection focused on content analysis, review and analysis of over 100 documents from four principal sources: academic search engines, websites of specific institutions, grey literature provided by water companies and references of articles and books to trace other sources.

Interviews were carried out with professionals from a range of backgrounds: three based in three global water companies, Suez Environment, Veolia Environnement and Severn Trent International, two in a national company that had been privatised, a regulator, an NGO engaged with the private company and an academic researcher, all of the latter based in Bolivia. The interviewees were selected because of their first-hand experience of global water privatisations. These recent interviews (2007) have been supplemented by information derived from a variety of site visits over a number of years, to communities and service providers in La Paz – El Alto (2004 & 2007), Buenos Aires (1999 & 2004), Santiago (2004), Manila (2002), Jakarta (2003) and Johannesburg (2007), including discussions with slum dwellers, in conversation and in formal focus groups.

Trial and error

The research clearly demonstrates that despite the impression created by various institutions, political leaders and the water companies themselves, the international operators were unsure how to meet the challenge of serving the poor when PSP was first placed on the agenda[10]. Although Brailowsky et al[11] highlight the importance of "the social know-how" of the private sector, it was by experimentation, experience and trial and error over the last two decades that better practice has emerged[12]. The Buenos Aires concession is a case in point: five years of diagnostics, three years of pilot projects and a third stage of wide scale implementation that commenced eleven years into the contract[13]. While Suez has recently been more vocal in publicising the success of its technological service differentiation there is a real danger that the knowledge that has been acquired is not valued and that the 'baby is being thrown out with the bath water' as the contract was terminated. The key question is whether the applicability and replicability of better practice identified herein is of value to public water providers.

The marketing approach

Public sector agencies are often captured by the interests of the producers of the service rather than being driven by the needs of their customers. The adoption of a marketing-based approach on the other hand[14] entails identifying a group's special needs (the poor in this case) and designing services to meet them better and in a financially sustainable way. Traditionally the 'product' of water suppliers has been considered to be drinking water but it also necessarily encompasses the hardware firstly by which the water is delivered and by which customers can pay and also the software, the processes by which often 'informal' communities are empowered to achieve networked water supply.

From the interviews with the international operators the private water companies definitely understand this need and, from the information gathered, have placed considerable emphasis on broadening the range of technological options in use in lower-income countries as well as in developing the processes through which the particular needs of the poor might be met.

The key drivers that determine their choice of technology are:

- the level of investment that is available (for example whether capital investment is to be met solely by users or subsidised by the State/international institutions);
- the technical feasibility (including condition of the network, its extension capacity, ability to deliver good quality drinking water);
- social acceptability (meeting needs, willingness rather than ability to pay, improved facilitation of payment);
- potential for partnering;
- and contractual obligations (flexibility in standards).

Examples of these elements have been summarised in Table 1 into a typology of water supply and sanitation services for serving the poor which have been used in recent PPPs.

Figure 2 - Piped water through remote meters replacing potentially contaminated groundwater in a slum in Manila, the Philippines (Photo: Franceys)

Network solutions

In addition to conventional pipe networks, surface level pipes have been used as a response to the population density and geographical conditions in, for example, Gabon and Manila[15]. Similarly autonomous, small water distribution systems with boreholes and gravity tanks, which can be connected to the city network at a later date, have been developed in Manaus. In Manila, tertiary networks have been established whereby households make their own connection using flexible hose to meter banks. Inspections are carried out by the water utility to ensure quality assurance[16]. This technology meets the challenge of the lack of space, scattered and haphazard layout of dwellings and lack of reliable statistics and population mobility which mitigate against formal planning.

The operators describe how surface level pipes, in locations which do not face the threat of frost, have the advantage of ease of construction as well as of maintenance with leakage detection, rather obviously, being much faster. Avoiding the need for trenching also avoids disruption to existing drainage channels, significantly reducing any costs involved in reinstatement.

Surface level pipes also make it easier for the community to be involved in construction, reducing costs and thereby reducing connection charges - though operating within government pricing structures these reductions have not always been passed on directly. Working with the community, through formal groups as well as informal, the operators have also been able to involve customers in preventing illegal connections, made even more obvious through above ground pipe networks.

The promotion of alternative technology has been aided by contractual design for example in Manila where the concessionaires had the choice of offering standposts in low-income areas instead of private connections, permitting third party provision and flexible technological standards to meet their coverage obligations as well as the assurance of a moratorium on demolition of the slum areas by the public authorities[17]. In the La Paz and El Alto concession, the contract was explicitly designed to expand services to the poor and permitted a social tariff, access to micro-credit and derogation for use of low-cost technology[18]. Meanwhile Veolia recommends exclusivity be limited to areas already supplied by the network, so that small private operators (in partnership or not with the official private operator) can play a role in water distribution in unserved areas.

Pricing solutions

Pre-paid meters have been installed in Johannesburg (Soweto) as a payment mechanism for consumption over and above the free water allocation following consultation and general agreement (though not by all). The electronic pre-paid meters in South Africa have been developed to recognise a household's free 6m^3 per month. Yard taps have also been installed in South Africa which are a half-way house between standpipes and household connections, the latter of which requires significant additional expense in terms of internal plumbing and fittings which has proved a problematic additional cost for example in El Alto. The service may be free (flow limiters may be included) or billed. Meanwhile, standpipes have been found not to be a popular choice but they are recognised as a temporary solution to give an improved service in a short space of time and have been used in Soweto[19] and Morocco[20]. Payment methods vary from free water to pre-payment or delegating responsibility as a commercial operation.

Sanitation solutions

There is a range of on-site sanitation options that is being used such as VIP latrines in Johannesburg[19], septic tanks and small local treatment plants in Manila[21] or semi-collective systems where a group of houses is connected to one large septic tank which is emptied regularly by the water utility in Casablanca.

Condominial sewerage in La Paz and El Alto is often cited as a success story[22]. This is understandable as the project embraced the use of alternative, cheaper technology and involved the participation of the community in its construction. However caution is required because interviewees have cited the system as a failure in El Alto. They put this down to a mixture of the low consumption rate and inability to look after the systems. One interviewee stated that the system functioned effectively in La Paz, where the conditions are different (gravity, higher consumption rate, higher education level among customers) though no other evidence has been found to corroborate his assessment.

Table 1 - Typology of water supply and sanitation services for serving the poor

Product	Options	Example
Conventional technology	Conventional water supply Conventional sewerage network	Morocco, Buenos Aires, Jakarta Morocco
Adapted or appropriate Technology	**Water supply** Standpipes	Morocco, Niger, Rabat-Sale
	Surface level pipes	Gabon, Manila
	Smaller diameter pipes	Cordoba, Ivory Coast, Manila, Senegal
	Autonomous, small water distribution systems	Manaus
	Shallow and deep wells	Manila
	Bulk delivery of water – tertiary network	Manila, Cartagena
	Group taps	Manila
	Yard taps	Johannesburg, El Alto
	Group/street/remote meters	Buenos Aires, Manila
	Sanitation	
	Condominial sewerage	El Alto-La Paz
	On-site sanitation. Options include VIP latrines, septic tanks, semi-collective systems	Johannesburg, Manila & Casablanca respectively
Advanced technology	Membrane filtration technology	Durban
	Advanced leakage control technology	Casablanca, Senegal
	GIS mapping	Niger
Payment technology	**Free** Flow limiters installed	 Johannesburg
	Payment	
	Coin-operated standpipes	Ivory Coast
	Pre-paid meters for individual connections	Johannesburg
	Metering programme (collective or individual metering)	Aguas Calientes, Burkino Faso, Cartagena

Others were however reluctant to condemn the technology, arguing that it was simply inappropriate in this location. Their sentiments are reflected in both Veolia[16] which sets out the advantages and drawbacks of the condominial system and states (without citing specific examples) that the results vary from zero to total success, as well as the infamous declaration of the then director of Aguas del Illamani, Arnaud Bazire, who described the population of El Alto as "the worst client imaginable" and "the worst consumer in the world" for their low consumption of water, for some households at the level of 1m^3 per month[23].

Overall, it appears that sanitation remains the poor relation of water supply in many cases as the connection rate is slower despite the rhetoric. This can be explained by two factors: conventional sewerage is expensive and poorly adapted to the conditions of slums and poor areas[24] and there is a relative lack of innovation to date in this area[25].

Partnership working

Evidence from contracts for example in Buenos Aires[26], Senegal[1], KwaZulu-Natal[16], Manila and Manaus[27] demonstrate the value and necessity of establishing partnerships between the private company, the public authorities, the NGO/community sector and the communities themselves (sometimes dubbed 'tripartite partnerships'[28]). See Table 1.

The benefits of the approach adapted from 13 are:

Commercial benefits	▶	New clients incorporated
Security	▶	No incidents / Theft minimised
Complaints	▶	Decrease in work complaints
Institutional benefits	▶	Public authorities involved
Clients	▶	Engender sense of ownership
NGOs involvement	▶	Part of the solution not problem or obstacle
Financial savings	▶	Participation in construction
Communication benefits	▶	Promote transparency & understanding

The relationship in low-income areas differs in this approach. It is not only with the individual customer (the case in conventional customer relations) but usually with the community as a whole because the challenge for the utility is to expand coverage in a way that is acceptable and affordable to the majority of households as well as ensuring economies of scale. It is therefore necessary to engage customers and potential customers to build up accurate data about them, which is generally lacking[1], as well as working with the community to formulate their demand as many are unaware of the options that are available, making the population "a partner in the process and not the object of a project"[26].

In some locations, the concept of community participation has also been formalised to promote local economic development. Instead of using one-off community labourers, training and employment have been generated in activities such as trench digging and community facilitation: in Soweto on average 1,200 people were employed at any one time from low income communities[19]; in Buenos Aires, worker co-operatives were established as part of the Agua y Trabajo programme; in Cordoba local labour were employed as part of a municipality initiative[29].

Dedicated unit

There is also a strong evidence of the need to establish a separate, dedicated unit within a water utility to serve the poor. Dealing with low-income neighbourhoods in lower-income countries is not a core competence of Western water companies. Fundamentally the socio-economic status of communities in La Paz or Casablanca is significantly different from the challenges faced by disadvantaged communities in their home countries. The establishment of a special unit or programme provides a momentum for the utility to tackle the challenges that lie ahead and encourages the utility to be pro-active rather than reactive. For example in Buenos Aires, Manila and Casablanca, the utility has been pro-active with the municipality on the question of land tenure. The approach must focus on how to integrate low-income customers into the mainstream activities of the rest of the company while recognising the specific challenges that serving them entails.

Furthermore, as problems rarely occur in isolation, the international operators have been able to draw upon their worldwide experience in the group to adapt a solution.

Dedicated units within the international operators at home office level have been necessary for diffusion of good practice

There are initiatives to promote best practice amongst public utilities[8], particularly now through the public operator partnership approach, but this is often a donor-led process. In Bolivia, according to the regulator, there was no interest from the public utilities to learn from the alternative practices of the private utility operating due to a mixture of "arrogance …pride…belief that the private sector has nothing to offer…".

Training

The subject of pro-poor training appears to be largely absent from documentation but the challenge faced within the water utility to persuade its staff that the poor are viable clients should not be underestimated[10, 11]. "The first obstacle is incredulity, the staff do not think that the poor can or want to pay". Brailowsky explained: "A common problem in a public and private utility…but possibly accentuated in a public one… is that when the utility employee is confronted by customers he adopts a certain attitude...becoming arrogant, pretentious and very technical in his speech. It is even worse when employees deal with the low income neighbourhoods as generally the public utility has little or no relationship with low income neighbourhoods."

The approach employed to overcome this perception focuses on training and experience. It is necessary to literally open up the minds of the employees, challenge their prejudices and lay down simple rules for example the value of respecting a commitment "….because the foundation for good working relations is based on trust". The next step is to equip employees with the tools and support to serve low-income communities[13, 16] based on a high level of professionalism and "…going against a charity or voluntary approach of the poor neighbourhoods projects". However good the training, it is ultimately only once staff undertake concrete projects that they become convinced: "the incredulity vanishes when we have results, not before".

Training must not be limited only to the specialist unit as it is important that the majority of the staff in the utility understand how serving low-income communities fits into the business. Aguas Argentinas for example provided training to District heads, the technical and sales contacts, maintenance teams, the customer service staff in each District and Regional Directors[13]. It is also a vital cog at community level "…so that [the community] can assume the devolved responsibility that the management model is going to give to them".

Capacity-building may be carried out by the utility, e.g. Buenos Aires[26], an NGO, e.g. Manaus[27] or the municipality e.g. Cordoba[29].

Affordability

Affordability is essential to create a sustainable service for the poor. Pricing policies generally fall under the domain of the public authorities. This is not to say that the utility does not exert considerable influence. This refers not only to cost of consumption, but also the connection fee[30, 16] and the terms of payment.

There are many combinations that promote affordability in terms of connection fees and consumption charges. They may include subsidies between taxpayers and customers (e.g. Buenos Aires, Cordoba, Morocco, Senegal, Chile, Jakarta, Guinea), a special charge or tax on the water bill which funds network expansion (e.g. Buenos Aires, Ivory Coast), international grants and loans that subsidise socially assisted connections (e.g. Niger, Burkino Faso, Guyana, Senegal), cross-subsidisation between through differentiated tariffs (e.g. South Africa, Ivory Coast, Morocco, Gabon, Niger, Buenos Aires, Jakarta, Senegal) or between large towns and isolated areas (e.g. Gabon, Niger, Guyana), and inter-service cross-subsidisation for example where electricity revenues partially fund investment or consumption (e.g. Gabon, Morocco). There have also been examples of community participation in the construction of the network as a means to lower connection cost (e.g. El Alto, Buenos Aires) though interviewees noted that this practice has provided grounds for subsequent conflict.

The obstacle is often less about the price than the terms of payment. Income in low-income communities is irregular and cash-in-hand but because facilitating payment is a sine qua non for private companies to maximise income, they have sought to adapt payment options and facilities to local circumstances. Thus, connection fees, which traditionally have ranged from an up-front payment to instalments spread over one year[(24)] have been spread over periods ranging from two up to 10 years (often interest free) in El Alto, Morocco, Gabon, Manaus and Manila.

In Tangier, mobile agencies visit peri-urban areas, facilitating payment, saving customers' time and money on travel and enabling the company to provide a more uniform level of service for all customers; in Rabat-Salé customers can pay bills in small shops in their neighbourhood seven days a week, nearly 24 hours a day. The lack of infrastructure in poor communities such as no street names or postal service has also led to the water company seeking new methods to deliver bills. In Buenos Aires and Manila, this has meant devolving responsibility to a neighbourhood association to facilitate payment[21]. This also has the added benefit of providing local employment, while in Durban meter readers are tasked with delivering bills. This type of adaptation to local realities, which involves high levels of co-operation with the community, also underlines the importance of building up the capacity of local people through strong communication, training and other follow-up tools.

Postscript

It is interesting to note that while many of the contracts with the multi-national private companies were terminated prematurely:

- EPSAS in La Paz is planning to continue the policy of community labour and is carrying out willingness to pay studies in El Alto, a policy which, it is said, would have never happened pre-privatisation
- Regulatory staff in Buenos Aires acknowledge that despite its challenges, there was real value in the participation of the community as well as the use of alternative technology (even suggesting expanding these experiences to incorporate other technologies) and makes the case for responding to the "informed demand" of the population
- the new owners of the company operating the Manila East Zone concession (Maynilad Water) appear to have carried on with many of the same practices and technologies promoted by its predecessor.
- the mutual learning, perhaps through competition, between the two private water companies in Manila has led to significant increases in piped connections for the poor.

Essential ingredients

There are several ingredients that stand out above all as linking practices together. Firstly, political will of the public authorities is a prerequisite to bring about change. Thus in Morocco much of the impetus reportedly came from an initiative of King Mohammed VI to speed up significantly the connection programme; in South Africa the Government made coverage a central plank of its policy and has been open-minded and practical in its pursuit of solutions[19];

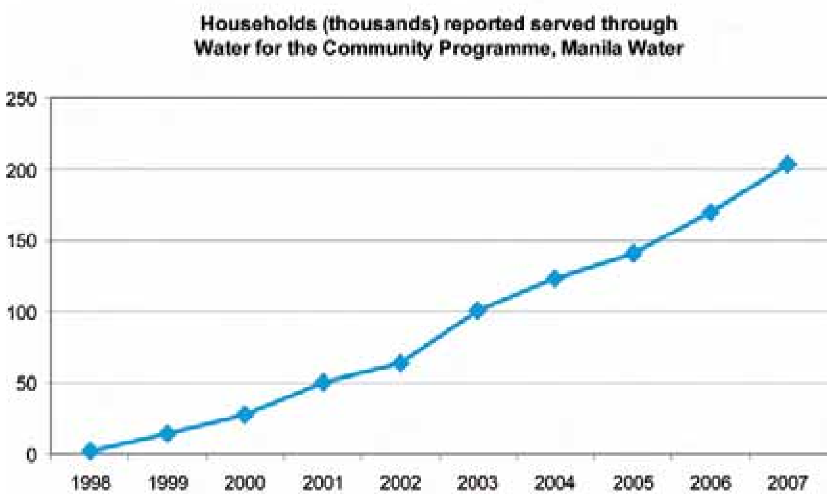


Figure 3 - Poor households reported served through Manila Water's 'Water for the Community' Programme. (Authors' analysis of Manila Water Annual Reports)

in Argentina the Agua y Trabajo programme can be traced back to a request (and funding) from the president[11, 31]; in Chile the State took responsibility for funding and organising a subsidised tariff for low income customers[32]; in Senegal the Government made service both accessible and affordable to the poor, recognising the importance of subsidising access not just consumption[1].

It was originally envisaged that private sector participation would provide the means to bring about change by challenging vested interests where political will was not present[7]. However, the evidence shows that the most successful examples of better practice are where there is partnership. As several operators have learnt to their cost, they might attempt to challenge these interests but they are deeply entrenched and the companies are unlikely to emerge as the winners.

The other essential ingredients are flexibility, adaptability and innovation. The typology of water and sanitation services demonstrated that there is no one size fits all technological or management prescription either between or just as importantly within contracts. In Buenos Aires and Manila, Suez experimented with different models of management (co-operative action, communitarian management and private delegation) as well as technological differentiation (conventional water and sanitation supply, above ground networks, collective, individual and/or remotely located metering, on site sanitation).

There is also considerable hypocrisy in the public-private debate in terms of the negativity with which service differentiation by private utilities is treated. International NGOs have been pursuing this policy for many years in rural and peri-urban areas in lower-income countries and the virtues of promoting appropriate technology as opposed to importing Western solutions are often extolled. While the context is different, the issue of affordability and service differentiation has also been applied in the United Kingdom with regard to on-site sanitation. It is common practice to install septic tanks in isolated communities where economies of scale do not exist to allow conventional sewerage services. Similarly the WaterSure programme and special needs registers differentiate water companies' charging and responses to customers with particular needs.

It is questionable whether any of the technological or partnership practices are really new. However the significant 'scaling-up' and sustainability of service provision to the slums through a comprehensive, technically and socially adaptive manner is new. Brailowsky[33] describes social engineering as "a new profession" but while it may be new for the corporate sector, this type of activity has a long history in the NGO world. Similarly, much of the technological practices that have been implemented are copied or borrowed from NGOs and small-scale private operators[2]. Yet it does not matter ultimately if the private sector practices are original or not: the key point is that they have merit if they address the problem and can be replicated on a large-scale at lower cost, particularly in the context of an overall improved service provider, necessary for sustainability.

Implications for Policy and Practice

Casarin et al[34] criticised a concession for turning an access problem into one of affordability. But they appear to disregard a challenge that cannot be ignored however unpalatable it may be: water supply and sanitation are expensive. If the State (through loans, international aid or taxpayers' revenue) does not meet this cost, then it falls on the customers (or most likely, a combination of the two). The alternative is the status quo and the poor remain unserved. This is where the lessons contained within this research and others are of value as they demonstrate how a utility that treats the poor as customers and designs technological solutions that are appropriate to their needs and means can meet the challenge of serving this section of society.

It is worth reiterating that the objective has not been to justify privatisation as a policy but to identify and highlight practices from global water privatisations that can serve as a model to copy or adapt. The beneficiaries may be other international operators, the domestic private sector or public utilities supporting Biswas's[35] call for more "South-South transfer between lower-income countries" rather than transposing conventional developed countries' solutions. Some suggest that international bodies could be the means for making such a transfer process happen…", but there is the danger that these bodies are perceived to be promoting the privatisation agenda[6, 8]. It is therefore necessary that the repository of private sector 'better' practice is seen as 'policy' neutral, to be mined for the benefit of all.

References

Unless otherwise referenced, all quotes and information are taken from fieldwork interviews with particular thanks to Brailowsky, A., Langton, S., Mathys, A., and Renard, N. in Europe and Chacon, C., Claros Rojas, F., Fernandez, J., Irusta, I. and Poupeau, F. in Bolivia. The financial support of WSUP (WSUP.com) is acknowledged in facilitating the interviews in Paris.

1. Brocklehurst, C, Janssens, J (2004) Innovative Contracts, Sound Relationships: Urban Water Sector Reform in Senegal Discussion Paper Series Paper No.1, Water Supply and Sanitation Sector Board, World Bank, Washington
2. Allen, A, Dávila, J, Hofmann, P (2006) The peri-urban water poor: citizens or consumers? Environment & Urbanization, IIED, Vol. 18, No.2, pp333–351
3. Kariuki, M., (2002) Water and sanitation utilities and the urban poor Waterlines Vol. 21 No. 2 October
4. Addo-Yobo, F, Njiru, C (2006) Role of consumer behaviour studies in improving water supply delivery to the urban poor Water Policy IWA Publishing Vol. 8, pp111–126
5. Franceys, R., Gerlach, E., (2008) Regulating water and sanitation for the poor, Earthscan, London

6. Barlow, M., and T. Clarke (2002) Blue Gold: The Battle against Corporate Theft of the World's Water, Stoddart, Toronto
7. Calaguas, B (2003) Private Sector Participation (Unpublished paper) WSSCC (WWW Document) www.wsscc.org/vision21/draft/index.html (Accessed 13th June 2007)
8. Hall, D (2001) Water in public hands (Unpublished paper) PSIRU, University of Greenwich, London
9. Fuest, V, Haffner, S (2007) PPP – policies, practices and problems in Ghana's urban water supply, Water Policy Vol.9 pp169–192
10. Botton, S (2007) Mettre en place une démarche d'ingénierie sociale: Suez et les quartiers défavorisés de Buenos Aires LATTS, Ecole Nationale des Ponts et Chaussées, Paris
11. Brailowsky, A, Botton, S, Matthieussent, S (2004) The real obstacles to universal access to the water service in developing nations (Unpublished paper) Paris
12. Pigram, J (2001) Opportunities and Constraints in the Transfer of Water Technology and Experience between Countries and Regions, International Journal of Water Resources Development, Vol. 17, No.4, pp563 – 579
13. Suez (2001) Integrated management for low income neighbourhoods (CD Rom) Water and Sanitation for All, Aguas Argentinas & Suez, Paris
14. Sansom, K., Franceys, R., Kayaga S., Njiru C., Coates, S., and Chary, S., (2004) Serving All Urban Consumers: a marketing approach to water services in low- and middle-income countries, WEDC, Loughborough University, 2004
15. Nickson, A, Franceys, R, (2003) Tapping the Market: The challenge of institutional reform in the urban water sector, Palgrave Macmillan, London
16. Veolia, UNESCO, Solidarité Eau (2004) Water, sanitation and sustainable development – the challenge of cities in developing countries (Published report) Co-production of UNESCO, Veolia Water, Programme Solidarité-Eau, M I Imprimerie, Paris
17. Rosenthal, S, Alexander, I (2003) Private sector participation and the poor: realizing the full potential of transactions in the water sector International Journal of Regulation and Governance Vol. 3, Issue 1, pp33-58
18. Budds, J & McGranahan, G (2003) Privatization and the provision of urban water and sanitation in Africa, Asia and Latin America, IIED, London
19. Jowam (2006) Jowam Contract Review (Unpublished internal report) Jowam, Johannesburg
20. Veolia (2006) Expertise and a commitment to sustainable development – 2005 Overview (Corporate publication), Veolia Water, Paris
21. Inocencio, A, Cristina, C (2001) Public-Private-Community Partnerships in Management and Delivery of Water to Urban Poor: The Case of Metro Manila Discussion Paper Series No. 2001-18, Philippine Institute for Development Studies, Makati City
22. Komives, K (2001) Designing pro-poor water and sewer concessions: early lessons from Bolivia, Water Policy, Vol. 3, pp61-79
23. Crespo, C Campanini, O (2006) Conflicto de Aguas del Illimani: Practicas regulatorias de servicios de agua potable y su impacto en la probreza, La Paz in 32
24. McIntosh, A (2003) Asian water supplies – Reaching the urban poor Asian Development Bank, IWA Publishing, London
25. WUP (2003) Better water and sanitation for the urban poor: Good practice from Sub-Saharan Africa, Water Utility Partnership, EU & WSP, Kenya
26. Aguas Argentinas (2004) Développement Durable - Modèle de Participation à la Gestion: San Isidro (Company publication) Aguas Argentinas, Buenos Aires
27. Aguas do Amazonas (2005) L'eau, facteur de développement social - L'exemple de Manaus au Brasil (Company publication) Aguas do Amazonas, Manaus
28. Franceys, R., Weitz, A. (2003) Public-private community partnerships in infrastructure for the poor Journal of International Development, Vol.15, pp1083–1098
29. Nickson, A (2001b) The Cordoba water concession in Argentina, University of Birmingham/ GHK Working Paper, No. 442 05, GHK International Publishing, London
30. Franceys, R., Kayaga, S. (2004) Charging to enter the water shop? The costs of urban water connections for the poor (Conference paper) 30th WEDC International Conference, Vientiane
31. Suez (2005) Global skills for the environment: Community participation in developing countries (Corporate publication) Suez, Paris
32. Orellana Halkyer, R Yañez, N (2006) Servicios de agua potable y pobreza (Draft document) IDRC & Agua Sustentable, La Paz
33. Brailowsky (2007) The impossibility of working according to a logic of common interests: A transverse approach to the La Paz and Buenos Aires contracts (Unpublished report) Suez, Paris
34. Casarin, A, Delfino, J, Delfino, M (2007) Failures in water reform: Lessons from the Buenos Aires's concession (Draft Copy) Utilities Policy pp1-14
35. Biswas, A (1998) A pragmatic approach to integrated water resources management (Unpublished communication) cited in 12.

Unattended environmental noise measurements - a can of worms?

Mike Wright
Associate
Acoustics Noise & Vibration
Atkins

Abstract

Seen by some as a 'solution' to obtain longer term data without the costs of manning, there is always the risk of leading everyone to the wrong conclusions. The procedure for determining ambient noise levels around (say) proposed wind farm developments are well known - you simply need to measure over a long period of time under a reasonably representative range of conditions.

This paper firstly considers a real situation where there are concerns over the validity of noise measurements and no matter how accurate these may be, illustrates how the lack of attendance could lead to considerable examination at say, a public inquiry. Whilst this paper may not be able answers to all potential situations that could arise during the period of unattended measurements, it does offer suggestions that would enable some answers to be given to strengthen your evidence.

A can of worms?

An example

Firstly consider a situation where there are concerns over a set of unmanned noise measurements. Some questions arose at the long Bathside Bay Container Terminal public inquiry in 2004 where the Principal acoustic consultant, Bernard Postlethwaite giving evidence for Harwich International Port, was examined by Timothy Straker QC for Hutchinson Ports. An illustration from the transcript is presented below:

'**Straker**: *If we pause there for one moment, please. There these are unattended noise measurements, so there is no one, so to speak, sitting there all the time whilst the measurements are being made?*

Postlethwaite: *That is right.*

Straker: *Can you help the Inquiry as to the mechanism which is employed to ensure that the measurements that you read from these unattended noise measurements are properly reflective of the noise which is being recorded and there is no interference from any outside source?*

Postlethwaite: *Yes. The unattended noise measurements clearly are, as we say, unattended. Therefore, one can never be certain there is no interference from outside noise sources. However, it is normally the case by midnight to 4 o'clock in the morning there is very little chance that there is are other things happening perhaps locally that might affect things.*

Now, you can often look at the whole range of noise parameters and see whether or not just the measurements you are interested in, like the L_{A90} or the L_{Aeq}, look strange, but there may be, for example, a sudden peak of noise which is recommended by, say, a high value in the maximum noise level, which may indicate that there is a spurious event happened which is not relevant to the more continuous background noise level. The first two tables are actually the unattended measurements and these only relate to the period between midnight and 4 o'clock. For the Felixstowe location that was in fact on the balcony of the Martello Tower. Clearly, these are not going to be affected by people at that time in the morning -- seagulls perhaps occasionally. But, generally, you can have a good idea as to whether or not the levels are stable by looking at the differences between, for example, the L_{A90} value and the Leq. You can see, for example, for, say, a south-westerly wind, we have in the top table is an L_{A90} of 50.3 and an L_{Aeq} is 52.2. There is a very small difference between those two parameters. This tells me that noise level is fairly steady and it has not been affected by discrete isolated events.

I suspect you were referring more to the attended noise measurements, where one does have to be careful in taking noise measurements. Certainly there, where one has taken, say, short samples of noise, you would be very careful not to include extraneous events that perhaps were not relevant to the measurements which you are trying to make. And you can pause out measurements when you are actually taking them to stop that contribution to your measurement sample.'

The above demonstrates the difficulty in substaintiating the validity of unattended measurements. This is something that is not always possible, especially if the expert giving the evidence was not the person actually out there getting the data.

Figure 1 - Typical equipment available in the late 60's and early 70's

Other work on good practice

Aspects of instrument accuracy and calibration are not discussed here and it is hoped that this paper will help to augment the well produced advice in the Good Practice Guide by Nicholas Craven and Geoff Kerry[1]. This was the result of a project at Salford University supported by the Department of Trade and Industry which showed that errors can be up to ±6dB). Further work is being presently undertaken by Hoar Lea[2]. This is expected to give "guidance on how procedural uncertainty in the measurement of both background and specific environmental noise may be understood and reduced".

Road traffic noise surveys

Calculation of road traffic noise

Calculation of Road Traffic Noise[3] describes a way to undertake 18-hour noise measurements. It is simple, straightforward and the method ensures that measurements meet a particular set of conditions. At the time the original 1975 edition was written, the use of unattended measurements was still a distant dream. Whilst it was possible to measure over 18 hour periods, the kit available, as illustrated in Figure 1, could record but not process any data. The data processing was done in the laboratory and was a time consuming process requiring continuous human intervention.

However, this combination of analogue equipment did allow all the manned measurement samples to be carefully checked during recording and again during the analysis which was generally played back through a speaker. Unwanted sounds could easily be edited out and the results over the necessarily limited sample periods to represent 18 one-hour periods or the shortened procedure that were taken could be well documented. By following the prescribed procedures, one could be assured of a reasonably accurate representation of conditions at that time.

'Going unattended'

The requirements of DMRB 11:3:7[4] and the Department for Transport's WebTAG[5] has placed an increasing need to investigate noise in quiet areas that are too far from a road to make meaningful calculations. In these circumstances measurements taken over long periods are necessary in order to obtain adequate data under representative conditions.

With the modern logging instruments now available, unattended 18-hour measurements are commonplace in road traffic noise surveys. They are considered the routine approach and are often extended to record 'around the clock' so that night-time data can be recorded. Loggers are often left to gather data over several days. However, the quantity of data may not be a match for quality.

Let us firstly consider what can influence the readings:

- Wind and weather
- Variable effects of traffic congestion
- Intermittent short-term effects of traffic diversions
- Interim effects of road works, traffic management
- Irregular effects from other local sources of noise
- Variable nature of intervening ground conditions during period
- Electrical corona discharge effects
- The possibility of interference with equipment.

Wind and Weather Conditions

Defra dismissed the possibility of using measurements in the process of undertaking the UK Noise mapping undertakings on grounds of the various effects of weather; expressly:

- Effects on the source of noise – wet roads, airport operations
- Effects on equipment – electronic noise in the microphone
- Noise from the weather itself – wind noise, rain noise and action on trees and buildings

It concluded that weather conditions can impose some constraints on the number of days (or nights) when measured noise levels can be relied on.

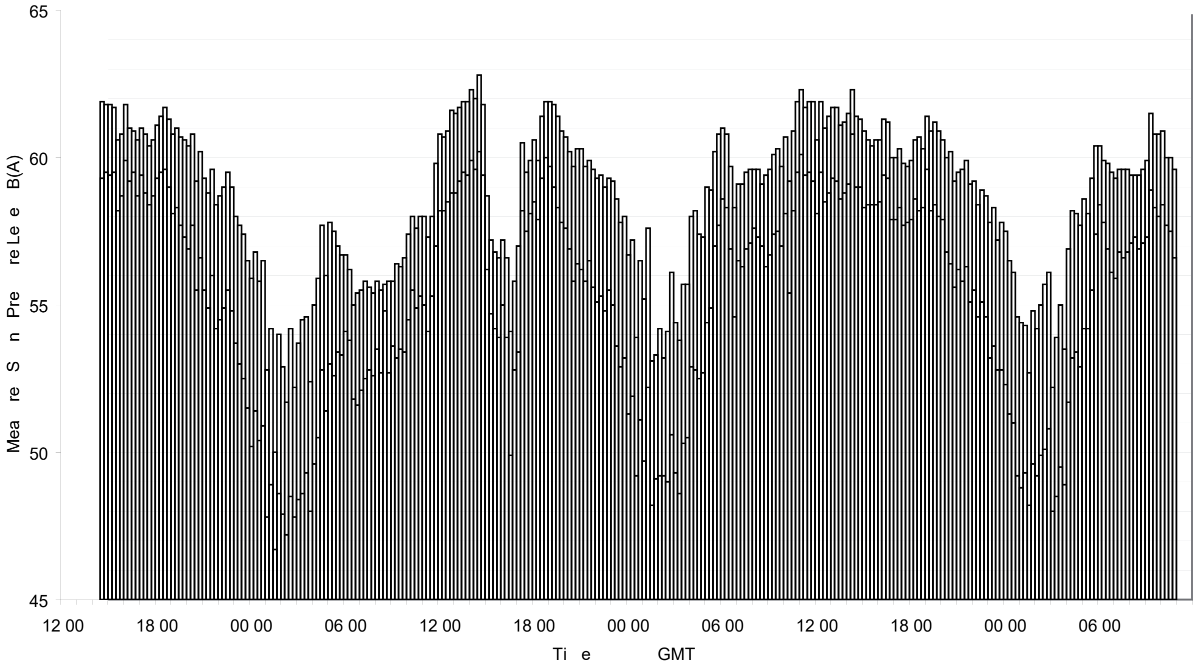


Figure 2 - Measured Noise Levels

However, by using a logging meteorological station set to run at the site or at a representative position in the area where a number of simultaneous measurements are made the effects of weather conditions on the data can be determined. Careful attention is required in setting up to ensure the weather station is properly calibrated and the wind vane compass correctly aligned. It is then possible to demonstrate that measurements would comply with paragraph 39.2 of CRTN.

This therefore provides a very strong argument to use such equipment as a means to validate all data that has been obtained from outdoor unmanned noise measurements.

The user should be aware of certain limitations. The weather station will not be able to tell you when the ground was wet, the microphone covered in ice, or when thunder was heard. For the latter, there is a good website look up data from the Isle of Wight weather station[6] which maps the position of lightning strikes in the UK.

For other meteorological information, as local conditions can vary, it is not recommended that publicly available data is used unless the noise measurements are taking place very close to the meteorological station in question.

Traffic Congestion

With today's crowded roads, traffic congestion can be a major factor in determining noise levels. CRTN essentially works on the premise that traffic is free-flowing and significant differences can rise if that is not the case.

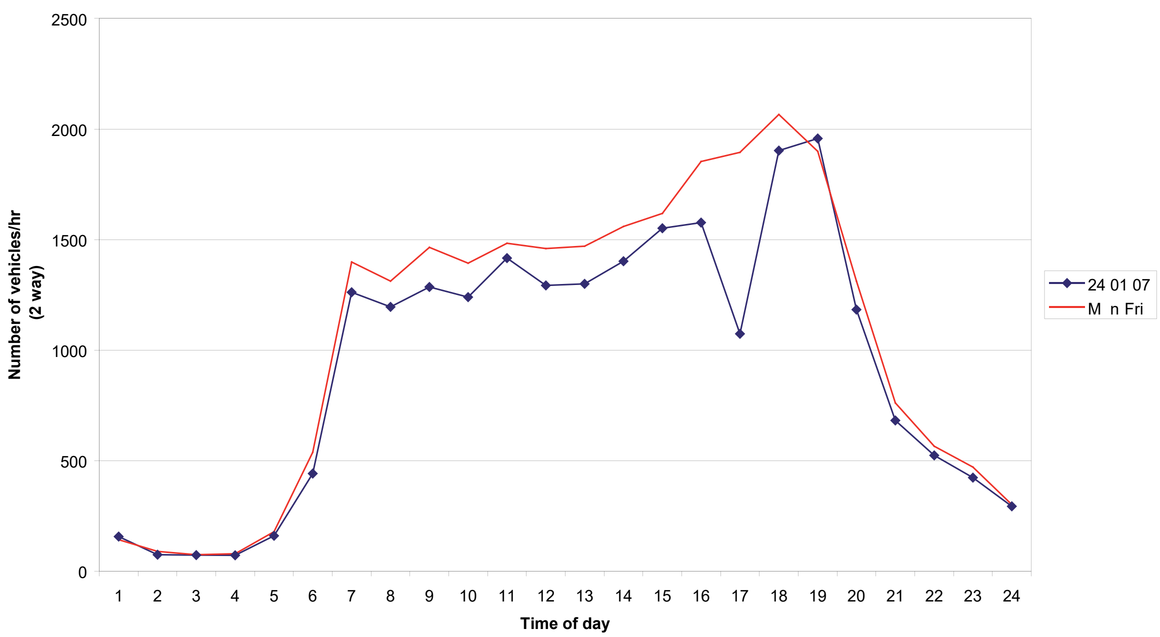


Figure 3 – Traffic Flow data

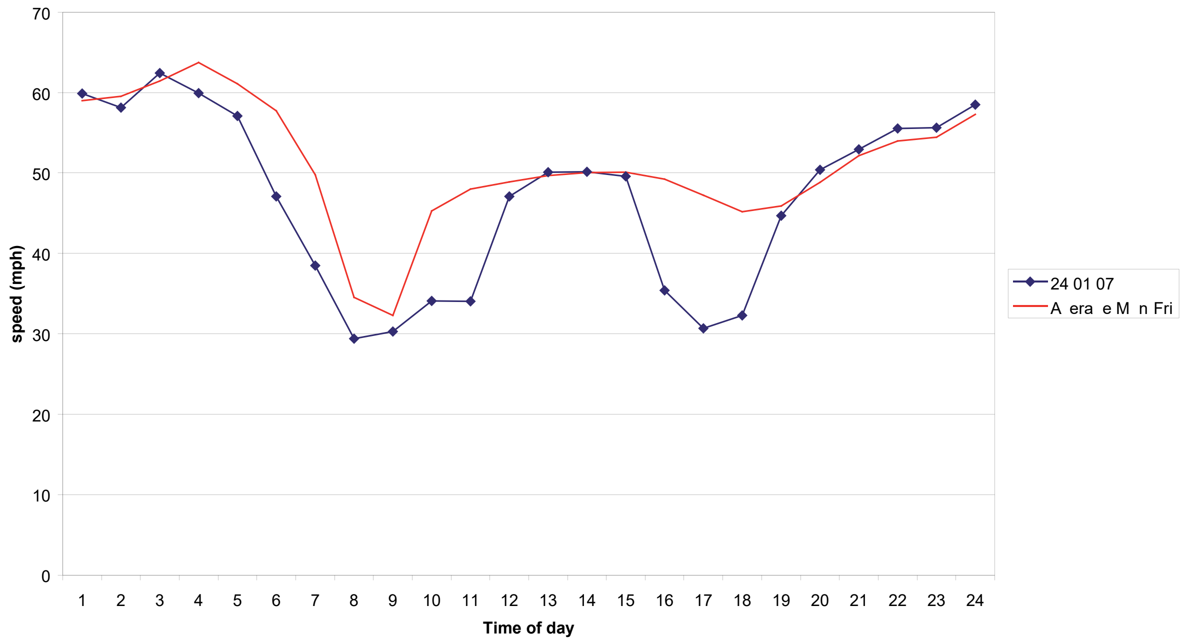


Figure 4 – Speed data

Consider the following data taken some 70m back from a major trunk road in Surrey during last winter. It will clearly be seen that the two complete days of data are markedly different. In this case, and unknown to the person undertaking the measurements, heavy snowfall had affected traffic speeds in the morning and more so in the early afternoon as shown. However, a rapid thaw came later leading to a return to normal conditions.

However other effects such as fresh lying snow and ice or slush covered roads were present which also challenge the validity of the data.

By reference to publicly available traffic data at a nearby counter run for the Highways Agency for their Traffic Information Database[7],it was possible to show what really happened. Figure 3 illustrates the effect on traffic flows on 24 January compared to the average for the 5 weekdays. During most of the daytime flows were lower than usual with a very pronounced dip in the late afternoon.

Figure 4 shows an even more dramatic effect on speeds. This casts much more light on the effects on traffic noise. It should be added that the effect of lying snow, wet roads and wind would have compounded to produce overall levels that might be markedly different from those that would be obtained if measurement conditions were in accordance with CRTN.

The Highways Agency traffic database only covers the UK backbone of trunk roads under the management of the Highways Agency. Where this is not available, it is possible that local authority or other data might be available. Where it is not possible to get 'off the shelf' data to assist in validating surveys, there is a strong argument to arrange for such data to be obtained to validate your data using the appropriate equipment. Such equipment can be hired and managed by a number of appropriate companies.

Short-term effects of traffic diversions

Diversions may be set up as a part of road works or as a result of accidents or other incidents. Whilst the effects may be rather obvious if the diversion directly concerns the road where measurements are taking place, diversions on other roads can cause atypical conditions to occur.

Table 1 - Attenuation per doubling of distance

Site and (Ground Cover)	Attenuation dB per doubling of distance	
	L_{A10}	L_{A90}
Site 1 (Grass)	5.4	2.6
Site 2 (Deciduous woods)	6.2	1.5
Site 3 (Rhododendron, bracken & Birch)	6.4	1.5
Site 4 (Rhododendron and Pines)	6.8	3.1
Site 5 (Gorse and Brambles)	6.5	4.6
Site 6 (Dense Spruce)	7.4	3.2

Part of the task of validating your measurements will be to establish whether any diversions were in place in the general area.

Effects of road works, traffic management

Effects of nearby roadworks could be significant and may occur at any time. However, inquiries to the appropriate highway authorities before starting your survey may help avoid spurious data. Traffic management measures could include temporary speed limits, lane marking or other restrictions. A reduction from 70mph to 40mph as in the case of many motorway traffic management programmes during maintenance usually gives a 3dB reduction. A total ban on heavy vehicles on town roads could have similar effect.

Figure 5 - Microphone array for directional monitoring

Irregular effects from other local sources of noise

The effects of other local sources can become dominant at certain times of the day and may not always be recognisable from a time history of successive 1-hour loggings. However, they may become evident if much shorter periods are logged. With the dramatic reduction in the costs of data storage, one answer may to greatly increase the resolution of the data - perhaps by giving periods as short as 1-minute.

This approach could be enhanced if one produced a simultaneous real-time digital recording to (say) MP3 quality so that post-survey investigative work is possible.

Even with its 'lossy' compression algorithm that is designed to greatly reduce the amount of data required, this would still be more than sufficient to allow post measurement processing which would allow selection of acceptable 1-minute samples in order to remove the offending periods of measurements.

The following are common factors that could yield unexpectedly high levels of noise at certain times of the day especially in areas where noise levels are generally low:

- Intermittent industrial and agricultural operations should be investigated before the survey and appropriate actions taken to document events
- Early morning bird song, levels can reach 60dB in some instances, the timing of this can be partly related to the timing of the start of astronomical twilight
- Shift working patterns should be determined in advance of the survey. Events such as milk and postal rounds may start early in the morning
- Intermittent and variable effects of waves, river weirs, canal locks, garden water features etc should be considered

Variable nature of intervening ground conditions

CRTN acknowledges the effects of acoustic impedance due to different ground surfaces and takes a necessarily simplistic approach in its calculation procedure. In reality, research has shown that mechanisms giving rise to 'ground attenuation' are very complicated. The complexity increases with distance and the effects are known to cause considerable differences arising between calculated and measured noise levels.

Whilst such effects are of little consequence at locations that are close to a noise source, any assessment of quiet areas set well back (say more than 300m from a major road or 100m from a minor road is likely to require some measurements to take place. However, the nature of the intervening ground can vary considerably over a period of time and the following factors contribute to the overall effects:

- Leaf cover, height, density and nature of undergrowth
- How wet the ground and vegetation are
- Effects of frozen ground and/or snow.

The variable effects of vegetation have been known for many years and many papers have been published on the effects from pioneering work by Eyring as early as 1946[8] through to more recent papers such as TRL RR238[9]. This demonstrated that foliage was important at reducing higher frequency noise whilst lower frequencies are attenuated by the absorbing qualities of the ground which may be enhanced by plant roots and leaf litter. This report also describes variability with different types of vegetation as summarised in Table 1, showing attenuation with distance in summer conditions.

It is reasonable to suppose that leaf cover, density and nature of undergrowth would vary according to the seasons of the year and effects of human activities. This may lead to the effects of acoustic impedance being less in winter than in summer. It is therefore important that photographs are taken to record conditions.

How wet or dry the ground and vegetation are should be noted and reference should be made to rainfall records made during the survey.

Evidence of frozen ground can be noted from temperature records that show freezing conditions for a period of time.

The effect of snowfall may be more difficult to record and this could markedly increase acoustic impedance when fresh snowfall takes place This is illustrated in Table 1 of Nord 2000[14]. By the following example flow resistivities for different ground coverings:

- Very soft snow or moss-like 12.5kNsm^{-4}
- Normal uncompacted ground forest floors, pasture fields 200kNsm^{-4}
- Hard surfaces – dense asphalt, concrete, water 20,000kNsm^{-4}

The effect of the snow can rapidly reduce flow resisistivities temporarily to a level that is less than would be expected over uncompacted covered ground. As it turns to hard ice or water, this can increase rapidly.

The effect of snow cover on the microphone assembly could also be significant.

If measurements cannot be avoided at the time, it is recommended that steps are in place to monitor conditions from time to time when it is considered that snowfall and/ or freezing conditions are possible. The use of remote time-lapse recording cameras synchronised to the measurement times could be considered provided that there are no other reasons to preclude their use.

Time-lapse recording camera techniques could also assist in determining whether road surfaces are wet or dry, another variable that can arise during unmanned measurements. Such may not be apparent from the weather records obtained from a synchronized self-recording meteorological station due to the additional and variable time that surfaces can take to dry out. It is known that increases of 3 to 5dB are not uncommon when a road surface is wet. However, much can depend upon the type of surface which can have variable effects under different conditions.

Electrical corona discharge effects

Noise can arise from an overhead power lines and insulators under certain conditions. This can be a crackling or 100Hz hum. It is produced by a phenomenon known as 'corona discharge' and can occur during damp weather conditions or during long dry spells when airborne debris sticks to the lines. The latter clears quickly when rain occurs. Reference to the meteorological records may not provide clues here. However, it is recommended that any noise measurements made where there are high voltage transmission lines or substations nearby should be treated with caution under conditions of rain or high humidity or where there are clear measurable effects on the noise at the end of a dry spell.

Human interference with equipment

Incidences of people interfering with equipment are not unknown. This may involve people creating additional sound so as to increase background noise. However, where people are aware that there may be a future increase in noise, people may attempt to move or cover the microphone during the period when they have been informed that it would be unattended.

It has also been known for inadequately secured microphone stands to be blown down and get re-erected by helpful residents. The use of real-time recording or remote time-lapse recording cameras could be considered. However, there may be privacy issues at stake and other methods may need to be considered including the use of sensors to detect movement. Before and after photographs showing the exact location of the equipment and unannounced interim visits should be considered.

Do you know where the sound is coming from?

By using more than one microphone, it is possible to determine where a sound is coming from. The small time differences between the sound arriving at each microphone provide the key to achieve this.

However, the use of an array of microphones is not a new technology. This has been used in the music recording and film industry for over 60 years with the "Decca Tree" being one well-known example. Indeed, Atkins were using microphone arrays on fixed noise monitoring installations 20 years ago. These were not portable and used the best technology of the time and were known then as the "acoustic telescope". However, the processing and data storage required to do this for longer term environmental measurements with portable equipment has only recently been a realistic and affordable possibility. The advent of fast data processing and large amounts of non-volatile memory such as the example in Figure 5, the "Barn Owl - Directional Noise Monitoring" from SoundScience has made it possible to detect sources of noise as well as the level of noise measured.

Other environmental noise

Taking account of noise near airports

Many of the above steps should be considered during the measurements especially the use of automatic recording meteorological stations. Additional data on conditions to the flight paths of the aircraft should also be considered. However it is advised that an approach should be made to the Civil Aviation Authority with a view to obtaining a time history of operations based upon records from air traffic controllers. It may also be possible to get a general idea of the direction of take-off and landing by reference to the wind conditions observed during the survey.

If CAA records cannot be readily obtained and it is considered that aircraft noise could be a significant factor, the use of real time recordings should be considered.

Noise near railways

Again, most of the above will apply subject to CRN[10]. It would be necessary to consider discounting any measurements that take place when temperatures fall bellow freezing as there are risks that the track bed would be frozen.

Noise from industry and BS4142[11]

BS 4142 states "Measured levels shall be considered valid only if they exceed readings on the measuring instrument owing to the above influences (wind, heavy rain and electrical interference) by at least 10dB." Where it is necessary to undertake manned measurements, it would only be possible to ensure this if a real time recording were taken. The periods of wind and heavy rain should be noted from the simultaneous meteorological data which should be recorded close to the noise measurement equipment to account for very local effects. Precautions to prevent exposure to electrical interference should be taken before measurements commence.

Wind farms

Many of the points discussed above will apply. However, the guidance given in ETSU-R-97[12] gives specific instructions that would mean that simultaneous anemometer data from a 10m high mast would be required and it would be necessary to identify periods of 'heavy rain'. The latter is undefined, however, it would be prudent to resort to some real-time recordings to help identify such.

The planning document PPS22[13] prescribes the use of ETSU-R-97 for assessing noise impact of wind farms.

Conclusions

The purpose of this paper is to increase awareness of the pitfalls associated with unmanned recorded noise data. It has not attempted to provide a comprehensive list. However what it is attempting to illustrate is the wide range of additional (non-acoustic) data that is required in order to address some of the most commonly encountered effects. Whilst data logging sound analysers are now available that one could have only dreamt about 30 or 40 years ago, many users are not aware that other technology is available for assisting in the validation process. In summary, the following should be considered before setting out to collect any noise data. Firstly, do you need a long period of unattended measurements? If so:

- A logging meteorological station should be mandatory for ALL unattended noise measurements. This should be able at least to record wind speed and direction, temperature, humidity and rainfall rate with reasonable accuracy. Avoid data from fixed Meteorological stations unless they are very close by
- Where it is considered that traffic congestion may be a variable factor affecting noise, data should be obtained from Highways Agency sources or with appropriate equipment
- Research potential noise generating activities on site including potential intermittent sources. Consider the use of real-time MP3 recording to identify potentially spurious data
- Consider using time-lapse recording cameras to identify ground conditions, road wetness, snow cover and other factors that may be taking place over the time of the survey
- Consider the use of microphone arrays so as to determine where the sound is coming from
- Make periodic and unannounced check visits to the site.

Above all, carry out a thorough inspection of all your data before you leave the site so as to determine whether you need to extend or repeat the survey to cover questionable periods.

References

1. A Good Practice Guide on the Sources and Magnitude of Uncertainty Arising in the Practical Measurement of Environmental Noise – Nicholas J Craven and Geoff Kerry (2001)
2. Project 2.2 'Environmental Noise' of the Department of Trade and Industry's National Measurement System Acoustical Metrology Programme http://www.hla-projects.co.uk/nms22/nmsproject22.asp.
3. Department of Transport / Welsh Office: Calculation of Road Traffic Noise (1975 and 1988)
4. Design Manual for Roads and Bridges, Volume 11, Section 3, Part 7 "Traffic Noise and Vibration": Highways Agency 1994.
5. Department for Transport - Transport Analysis Guidance (TAG) Unit 3.3.2 The Noise Sub-Objective (2006) http://www.webtag.org.uk/
6. The Isle of Wight Weather Station http://www.isleofwightweather.co.uk/live_storm_data.htm
7. Highways Agency Traffic Information Database6 http://www.trads2.co.uk
8. C.F. Eyring – Jungle Acoustics – Journal of the Acoustical Society of America 18(2) pp 257-270 (1946)
9. L. Huddart - The sse of vegetation for traffic noise screening TRL RR238 (1990)
10. Department for Transport - Calculation of Railway Noise (1996)
11. BS 4142: 1997: Method for Rating industrial noise affecting mixed residential and industrial areas.
12. The Assessment and Rating of Noise from Wind Turbines: The working group on noise from wind turbines: DTI 1997.
13. Planning Policy Statement 22: Renewable Energy – Office of the Deputy Prime Minister
14. Nordic Noise Group - Nord 2000 – Comprehensive Outdoor Sound Propagation Model Part 1 – Propagation in an Atmosphere without significant Refraction (2000-2001)

Proposed changes to the UK reservoir safety legislation to incorporate a risk based approach and the problems the UK faces in the future

Dr A K Hughes
Director of Dams
Water & Envrionment
Atkins

Abstract

This paper describes the way in which the UK is moving in respect of reservoir safety and outlines the legislative changes that are currently being drafted. It also identifies issues which the author thinks are important for the future of dam engineers around the world.

Introduction

As we approach 2009 the UK has moved and continues to move towards significant changes in the way that it manages the problems of reservoir safety. Some of these changes are perceived to be good, some are not so good. This paper explores the various initiatives and proposals that are currently in place and sets the scene for comparisons with the rest of the world.

Scope of the problem

Dam building in the UK really started with the industrial revolution and the canal building era, further extended by the construction of landscaping lakes associated with the construction of country houses and their estates (Figure 1). The failure of Bilberry causing 52 deaths, and Dale Dyke, both in the mid 1800's led to the first piece of legislation – which appears to be targeted towards dams – the Water Works Clauses Act of 1863. However it was the failure of dams in 1925 (Coedty and Eigian in North Wales was where 15 lost their lives and Skelmorlie where 5 lost their lives) (Figure 2) which brought about the first piece of 'purpose built legislation' – the Reservoirs (Safety Provisions) Act 1930 (known as the '1930 Act'), which introduced the concept of the 'Panel Engineer' – an engineer with a requisite level of experience, appointed by the Secretary of State, to a panel (a list) of engineers who are independent of the owner and who carry out thorough, independent reviews of the safety of the dam, at regular intervals – no more than every 10 years.

International incidents and accidents including Vaijont and Malpasset caused the UK to revisit the '1930 Act' and to see if any improvements were necessary. This was undertaken in 1966 and twelve recommendations were made, most of which were included in the current legislation, 'The Reservoirs Act 1975' (the '1975 Act'). This was enacted in 1987 and has served us well until fairly recently.

The legislation as it stands applies to some 2,100 dams in England and Wales and an additional 680 in Scotland. The criteria which dictates whether a dam is subject to the Act or not are the retained volume measured above the level of the adjoining ground – the limit is 25,000 cubic metres. This volume is the capacity of one of the dams which failed in 1925 – 5 million gallons.

Thus a 'large raised reservoir' is a reservoir which holds more than 25,000 cubic metres above the level of the natural ground adjacent to the reservoir. The majority of dams in the UK are earthfill embankment dams (more than 80%) of moderate height – (in the range of 5-15 metres) and with an average age of more than 110 years.

Ownership of the dams is vested in more than 700 disparate owners and thus the condition, engineering support systems, and monitoring and surveillance with an Enforcement Authority can vary greatly.

The 1975 Act brought in the requirement to "register" the reservoirs and for the owner, to employ a Supervising Engineer – a trained engineer who would visit at least once a year to look for change and the start of a failure process, and the role of the Enforcement Authority which had the power to enforce, and to cause an owner to employ engineers and to carry out any work that was recommended as necessary. Unfortunately this enforcement role was vested in more than 170 different local authorities and county councils who carry out the role with varying degrees of success.

Proposed changes to the UK reservoir safety legislation to incorporate a risk based approach and the problems the UK faces in the future

Figure 1 - Example of lake landscaping on country estates

As the UK starts to build statistics on incidents and accidents, it is apparent that although, we have had no failures involving loss of life since 1925, we have had some near misses with evacuations of some communities and it is almost the norm to have of the order of 4 serious incidents or accidents each year, which require an emergency draindown of the reservoir.

Recent changes and initiatives

Move to a single Enforcement Authority

Whilst it was considered that the main parts of the 1975 Act were working well, it was evident, as early as the mid 1990's, that by having more than 130 enforcement authorities in England and Wales, and another 40 or so in Scotland, the Enforcement Authority role was being carried out with varying degrees of success. Unfortunately the majority did not carry out the role primarily because they were not funded properly.

A report by the author of this paper to Government in the mid 1990's suggested the adoption of a single Enforcement Authority for the UK.

The Water Act 2003 introduced the single enforcement authority role, but because of the devolution of power to the constituent countries of the United Kingdom, only to England and Wales. The situation in Scotland remains the same with more than 42 different enforcement authorities. The Enforcement Authority in England and Wales is the Environment Agency (EA) – a Government financed organisation formed to "care for the environment". The EA have set up an 'Enforcement Team' in their office in Exeter. It's role is to ensure the owner appoints Supervising Engineers and Inspecting Engineers, and others such as Construction Engineers or 'Qualified Civil Engineers' at the appropriate times and that they carry out, in an appropriate timescale, any works required in the interests of safety. It is a role where they try to persuade an owner to behave/ follow the letter of the law, rather than use their enforcement powers.

They have successfully taken to court some owners and won their case which has resulted in the owners being fined, as well as having to pay for the breach of the Act being corrected. However, the EA has no inhouse engineering skills and all work has to be done under the control of a Panel Engineer.

The adoption of a single Enforcement Authority has resulted in a much more consistent and efficient approval to the regulation of reservoir safety in the United Kingdom. Some have been 'caught' by the enforcement system and the majority are of the opinion that the change has been a change for the good. The level of non compliance with the Act has fallen drastically and there is a consistent, and firm but fair approach to enforcement. Scotland is currently looking at adopting a similar approach by adopting a single enforcement authority and is seeking the views of the owners at the moment. All indications are that a single enforcement authority will be adopted in Scotland.

Figure 2 - Skelmorlie

Build the Register

It is the duty of the enforcement authority (in whose area the reservoir is situated) to form a register of all large raised reservoirs within the meaning of the Act. This was not always done with the thoroughness that was demanded. In this situation it is unlikely that an owner with a large raised reservoir is going to volunteer information unless forced to. And so for many years, some as much as 22 years, many reservoirs avoided being subject to the 1975 Act.

An initiative by the EA called the 'potentials' project set out to identify 'large raised reservoirs' which did not appear on their register (assembled by combining the registers formed by their authorities that they succeeded). This was done in association with the UK's Mapping Organisation – the Ordnance Survey and satellite imagery.

More than 400 bodies of water with the potential to have a volume which made them a "large raised reservoir" were identified and then visited and from this more than 110 "reservoirs" were added to the register.

Incident Register

In any sphere of engineering or it could be said aspect of life, it is important to learn from our mistakes. Indeed it can be argued that the most valuable lessons can be learned from analysis of mistakes. In this respect many technical papers are published each year and from time to time publications such as "Lessons from Dam Incidents" (ASICE 1988) are produced. These provide invaluable data which seek to inform and educate engineers to prevent similar incidents and accidents.

The EA has set up a voluntary incident reporting system which seeks to gather information – from incidents and accidents which can then be analysed/published to educate the profession.

The owner can give the information himself or an engineer can be appointed by the EA to visit the owner to gather the information. The information can be given in such a way that the owner and the dam need not be identified.

A number of incidents and accidents have been reported and a number of data sheets eg; on the performance of masonry spillways, have been issued to Panel Engineers and owners. However, the system is voluntary and it is known that a large number of incidents have not been reported.

Future initiatives are likely to make the reporting of incidents and accidents mandatory. This will also allow incidents and accidents to be the subject of additional data sheets or indeed another publication in the form of "Lessons from Dam Incidents". Again it is believed that the profession will generally welcome the introduction of a mandatory incident reporting system, as long as the anonymity of the owner and his reservoir can be maintained.

Flood Plans

As part of the Water Act 2003 the Secretary of State can direct of an owner to produce a Flood Plan. This has been defined, within a document entitled 'An Interim Guide to Emergency Planning', as being an 'on site plan' produced by the owner, a dam breach and inundation - mapping exercise again produced by the owner, and an 'offsite' or emergency response plan produced by the emergency services – via in the UK organisations made up of the emergency services and utility companies etc called the Local Resilience Forums (LRF's).

The Interim Guide was not well received by the emergency services and the LRF's and so Atkins was appointed to revisit and rewrite the contents, which is currently being done with the close cooperation of the LRF's with trials of the offsite and onsite plans in selected areas.

One of the difficulties experienced in revisiting the Interim Guide was how to decide to what reservoirs Flood Plans should be applied. In July 2007, an incident occurred at Ulley Dam near Rotherham. (Figure 3)

There was no Flood Plan at the time of the incident and it appears that the dam had been wrongly 'categorised' in terms of the consequence of failure. Currently impounding dams are classified in terms of consequence of failure when deciding the design flood according to Table 1.

Proposed changes to the UK reservoir safety legislation to incorporate a risk based approach and the problems the UK faces in the future

Figure 3 - Ulley Dam

In the case of Ulley, it would appear that the dam was incorrectly classified by the Panel Engineer and hence two problems have to be faced in rewriting the Guide - firstly which category of dams should be subject to Flood Plans, and secondly are the current categories correct? One of the criticisms of the Interim Guide was the likely cost to the owner of producing the Flood Plans. New Legislation in the UK must be shown, after public consultation, and after a 'regulatory impact assessment', not to be too onerous on those to which it applies. It was anticipated that Flood Plans could cost as much as £15,000 (A $32,000) to produce and this did not include digital terrain modelling data which is needed to produce the inundation maps which could cost another £20,000 - £30,000 (A $43,000 - 64,000) per reservoir

Currently a trial is being undertaken to try to develop what I have christened the 'Quick & Dry Method' which seeks to find a method whereby a breach and the subsequent inundation mapping can be modelled, with sufficient accuracy for the end users, defined as emergency planning purposes.

It is hoped that sufficient monies will be made available from the Government and that the digital terrain modelling data for all dams in the UK will be released to enable the breach to be modelled and inundation maps be produced for all dams subject to the Act.

In this way, in just over a year's time, it is hoped that all 2,100 dams in the UK subject to the Act, will have Flood Plans which will improve the UK's capability (the owner and the emergency services) to respond to an emergency. The dam breach model and the associated mapping are likely to be given, free of charge, to the owner which clearly will be more acceptable to the owner and the modelling will be specified in such a way so as to allow more detailed modelling/reductions of uncertainty etc, so that the maps can be used for categorisation/ identification of consequence of failure and spacial planning.

Changes and initiatives for the future

The biggest changes and initiatives for the future have gathered importance and momentum since the publication of "The Pitt Report". Pitt endorsed the production of Flood Plans and their publication so that the public can be warned and informed. Following the tragedy of September 11th 2001 the security services have restricted/stopped publication of Flood Plans. It is hoped that within a reasonable time Flood Plans will be published and the public educated and informed as to what to do if there is an emergency at the dam which is upstream of where they live or work.

Change in Legislation - A new Act?

Small dams

In the floods of July 2007 it is known that perhaps as many as 35 small dams (ie those with a capacity of less than 25,000 cubic metres) failed. Mercifully no-one was killed but extensive damage was done.

Table 1 – Categories of dam according to Floods & Reservoir Safety: A Emergency Guide

Category	Description
A	A dam where a breach could endanger lives in a community
B	A dam where a breach (i) could endanger lives not in a community or (ii) could result in extensive damage
C	Where a breach would pose negligible risk to life and cause limited damage
D	Special cases where no loss of life can be foreseen as a result of a breach and very limited additional flood damage would be caused

These incidents reinforced the view that some of us in the profession had held for some time, that small dams in the UK, not subject to legislation, are often more hazardous than some dams currently subject to legislation.

There exist perhaps as many as 10,000 dams, built for the industrial revolution, whose capacity is far less than 25,000 cubic metres, which should they fail could cause loss of life, environmental and economic loss, or, as identified by a number of incidents in 2007. They can often pose a threat to critical national infrastructure such as motorways, treatment works, gas mains, electrical sub-stations etc as well as a potential loss of life. Equally there are many dams currently inspected and supervised whose volume is well in excess of 25,000 cubic metres and may have a volume of many hundreds of thousands of cubic metres which, should they fail would not pose any threat to life and cause no damage - these, it is felt, should not be subject to the legislation.

The UK is thus currently exploring a move towards legislation based on the consequence of failure and not on volume. The challenge faced is to replace a criterion that is very simple and which easily establishes whether a reservoir is subject to the Act or not, with one which is likely to be of necessity more complex perhaps using a number of factors including volume, height of dam, type of dam, population at risk, and even type of owner. If it is too complex it is likely not to work and there will be endless disputes as to whether a reservoir is subject to the legislation or not. If it is too simple some reservoirs which ought to be subject to legislation might be missed again.

This is the issue being faced at the moment. Associated with this problem is the identification of all bodies of water to be considered and it is being suggested that this could be as low as 5,000 cubic metres before an exercise is undertaken to reduce that number to the reservoirs which should be subject to the Act - which might be as many as 4,000 to 5,000.

This task in itself is fraught with problems namely:

- first one has to identify all the bodies of water
- secondly one has to find all the owners
- thirdly one runs the risk of those who do not want to be subject to the legislation trying to breach their dam causing environmental damage
- fourthly one might get unsafe situations arising as owners try to breach their dams or reduce their height/volume without obtaining engineering advice.

All in all a worthwhile and important initiative but one which must be fully considered and planned, and the subject of an extensive communication exercise before it is embarked upon.

Charging regime

Following the principle of "the Pollutor Pays", practised by the Environment Agency, there is a suggestion that a charging regime will be implemented whereby owners will have to pay an annual licensing fee to the Enforcement Authority.

This charge seeks to cover the cost of the administration of the Reservoirs Act but many feel this is a retrograde step when it is increasingly difficult to get owners to maintain and to monitor their dams – a charge at this time gives exactly the wrong message.

For some institutions which are partially funded by the Government, such as British Waterways, it means monies will be paid by Government who will then require some of that money to be repaid in the form of a licence fee. For the small private owner or a fishing club, it could well send them into bankruptcy or close the club leaving the dam to fall into disrepair. For the large water undertakings, they will be able to apply to Government for permission to increasing the charges for water and wastewater treatment to their customers, and therefore the general public will pay. This initiative is not well thought out and will be unwelcome and badly received.

Others

There are a number of other initiatives planned of a minor nature which seek to close some loopholes with the current legislation and these are generally needed and will be found acceptable.

Other issues for the future

The UK, in the opinion of the author, is still in a position where many companies do not "value", their assets properly and they operate within a financial timescale that is too small. As a consequence we are seeing:

- a reduction in maintenance levels and quality
- a reduction in manpower with the necessary engineering skills to maintain our dams, to supervise our dams and to repair our dams
- a tendency to reduce 'revenue' spending in favour of 'capital spending' because the sources of the monies are different and some sources have repercussions on the financial performance of the business
- a desire to rely more heavily on instrumentation rather than use manpower on the ground

- the engagement of contractors and consultants who might not be the most appropriate or who may not have dam engineering skills because they are the cheapest.
- a lack of succession planning, knowledge transfer and training of staff for the future

Many of these issues I have identified before and they result from the short term nature of how businesses are run at the moment and a lack of planning for the future.

The lack of maintenance and the use of inappropriate contractors has led and will lead to a number of incidents and perhaps even failure at some point in the future. It is often only the "Panel" system in the UK which seeks to prevent a problem occurring and Panel Engineers will have to get more 'forceful' in what they recommend in the future, if they feel the situation is getting worse. If a failure does occur I hope that the investigation will examine not only what went wrong from an engineering point of view but perhaps more importantly it will look more deeply at levels of maintenance, financial contracts and procurement strategies and hopefully park the blame for these issues where they should belong – which hopefully will not be an engineer.

The lack of trained personnel on dams, where it is not driven by a desire to reduce salary budgets, is a much larger problem but one which must be dealt with now to prevent a situation which could be critical in a relatively short period of time. In my opinion the employers and the current group of dam engineering 'elders' must start to consider the problems of the future and think about knowledge transfer, succession planning and training. We must:

- encourage the young
- generate passion about what we do
- improve our professional image and status
- generate an interest in dam engineering whether it is associated with new dams or the fascinating work associated with the maintenance and upgrading of existing dams
- provide high quality, focussed training opportunities
- share and pass on to future generations the knowledge that we hold; and
- we must use techniques such as shadowing, mentoring, coaching etc. If we don't, we face problems in the future

Conclusion

This paper seeks to set the scene as to where the UK currently is with regard to reservoir safety and in particular its legislation and where it is likely to move to. Most of the changes will be seen to be beneficial; and some will be easy to achieve, others not so easy. It is also clear that we continue to face problems associated with the maintenance of our dams and the human resources of the future. These two issues, in my opinion, are the most serious for the immediate future.

The comments and views expressed in this paper are those of the author and not necessarily those of Atkins or any body that the author works for or is associated with.

References

1. HMSO The Reservoirs Act 1975 HMSO, 1975
2. DETR The Guide to the Reservoirs Act, Thomas Telford 2000
3. ICE Floods and Reservoir Safety; An Engineering Guide, 3rd Edition, Thomas Telford, 1996
4. ASICE Lessons from Dam Incidents, 1988, ISBN 087262661X
5. DEFRA An Interim Guide to Emergency Planning, 2007

Forward looking infra red camera turret for the Japanese coastguard

John Blackburn

Defence, Aerospace and Communications
Atkins Nedtech
Netherlands

Abstract

The challenge set by this project was to produce a design for the synchronised deployment and retraction of the forward looking infrared (FLIR) camera turret for a Gulfstream 550 aircraft. The modification was required to support the Japanese Coastguard's work patrolling its coastal borders.

Atkins prides itself on its ability to plan design and enable all aspects of its clients' projects and this paper describes how the Atkins Nedtech team achieved this in designing, testing and supporting production of the camera turret.

Introduction

The ability to work across a variety of disciplines, and from initial concept to in-service, was essential to delivering the forward looking infrared (FLIR) camera turret for the Gulfstream 550 aircraft. Atkins' contribution covered trade study, conceptual design and kinematics, detailed design, identification and management of suppliers, safety analysis and engagement with relevant authorities see Figure 1, and test support; demonstrating the Atkins ability to 'Plan Design Enable' on all aspects of our clients' projects.

Design and operation of the FLIR turret

The FLIR turret, when retracted, is located within a fairing, which itself forms part of a radome mounted onto the underside of the fuselage. The opening in the fairing is closed off by means of a movable door. The FLIR Turret is deployed, by means of an electro-mechanical actuator, to a position below the lower surface of the fairing. The door is automatically moved to its open position inside the fairing by a linkage mechanism, which synchronises the door and FLIR Turret movement in such a way that only one actuator is required.

The FLIR Turret is attached to a support arm, which pivots on the main bracket. The main bracket is mounted onto provisions on the fuselage and the door is attached to hinge arms, which pivot on the main bracket. In the deployed position extra stability is achieved by a rod, which automatically stows on retraction of the mechanism.

Plan

The initial concepts for the FLIR camera turret were developed around single actuator, dual actuator and door designs and was framed by client requirements. In particular, the following requirements were made:

- Fokker Services defined the possible locations of the FLIR
- Information concerning the FLIR was based on drawings prepared by FLIR Systems inc., USA
- Load cases (air loads, manoeuvre, crash etc)
- Drag load = 2KN limit (3KN ultimate) applied at approx 175 mm (6.9") to FLIR support
- The radome (fairing) shape definition received from Fokker Services
- Hydraulics and electrics data received from Gulfstream
- Electro-mechanical actuation was to be used
- Signalling was to be either by means of proximity switches or micro switches
- Target mass of fairing and FLIR was yet to be determined at this time

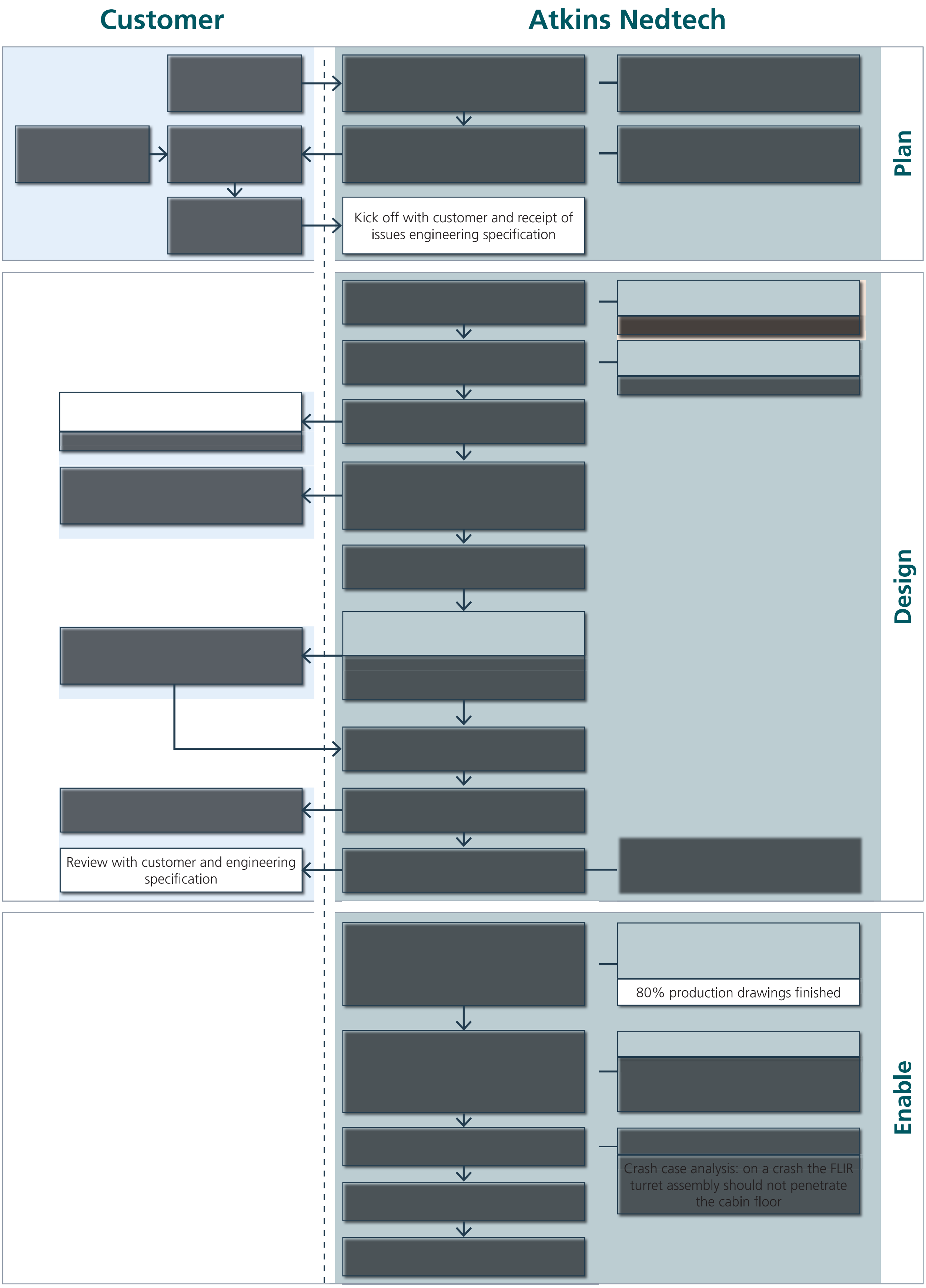


Figure 1 - The Plan Design Enable Process

AEROSPACE

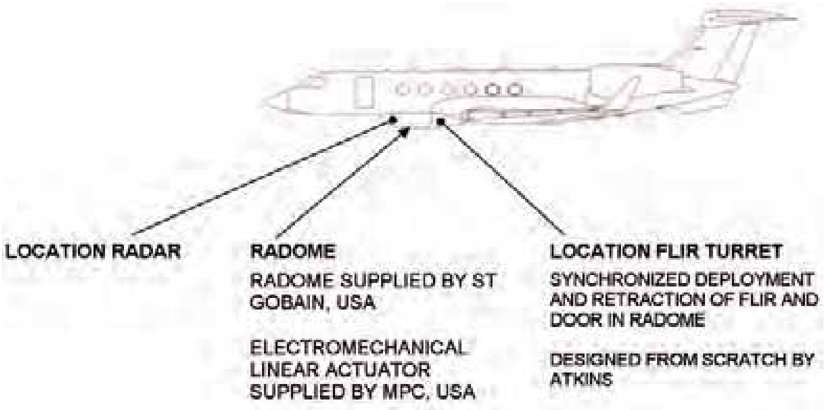


Figure 2 - Outline of Gulfstream 550 showing position of FLIR Turret

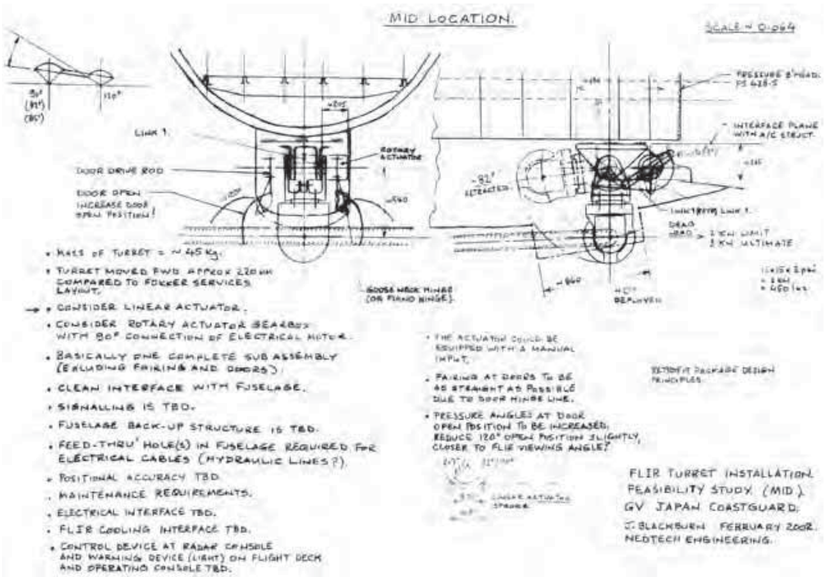


Figure 3 - FLIR Turret Installation (mid) Design Feasibility No.1

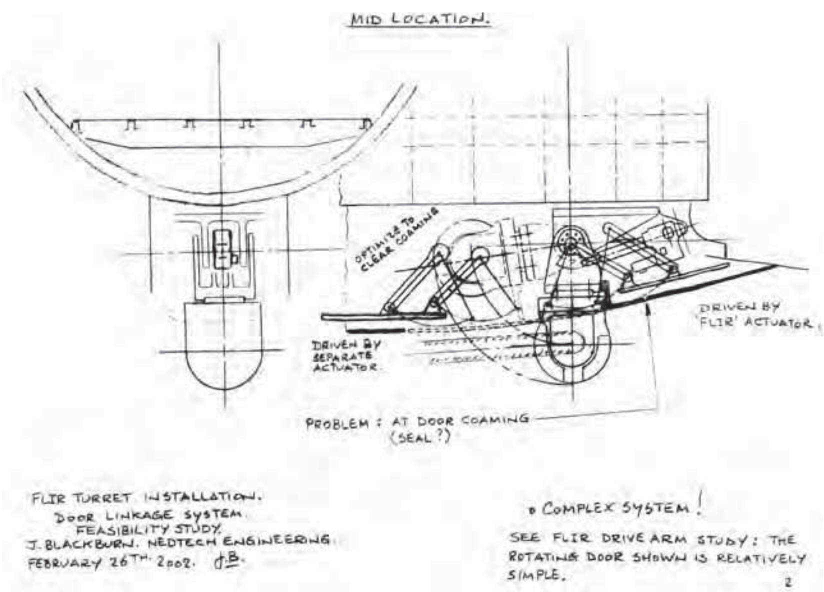


Figure 4 - FLIR Turret Installation (mid) Design Feasibility No.2

Figure 2 shows an outline of Gulfstream 550 with the position of the FLIR turret

Feasibility Study

Five design concepts (Figures 3 - 7) were explored, with each subjected to an assessment of its advantages and disadvantages.

It was important to consider each of the following when examining the feasibility of the designs:

- Manufacturing requirements and availability of components
- Interface with the fuselage and impact on structural integrity
- Interface with the fairing and associated drag
- Weight
- Potential for mechanism failure
- Maintenance requirements
- Link to control mechanism on the aircraft's console
- Overall cost
- A maintenance free design.

Certain configurations around outward opening doors were discounted on advice from the customer. This was due to the loads and high aircraft speeds subsequently creating too high a drag impact on the airflow around the fairing. The customer also discounted any concepts that used a forward location for the assembly, as the location was too far away from the radar and not structurally acceptable.

FLIR Turret Installation (mid) Design

Feasibility No 1: Rotary actuator and lever with link. Outward opening doors.

- FLIR and drive system is basically one sub assembly
- Drive by means of a rotary actuator and lever with link
- Outward opening doors driven by a relatively simple linkage mechanism
- Pressure angles might be critical
- Outward opening doors not accepted by Gulfstream
- Rotary actuator is electro-mechanical

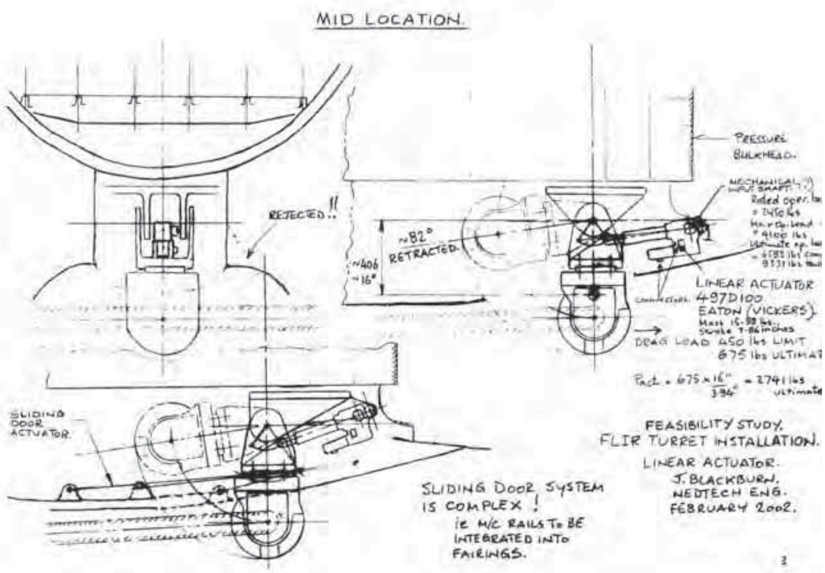


Figure 5 - FLIR Turret Installation (mid) Design Feasibility No.3

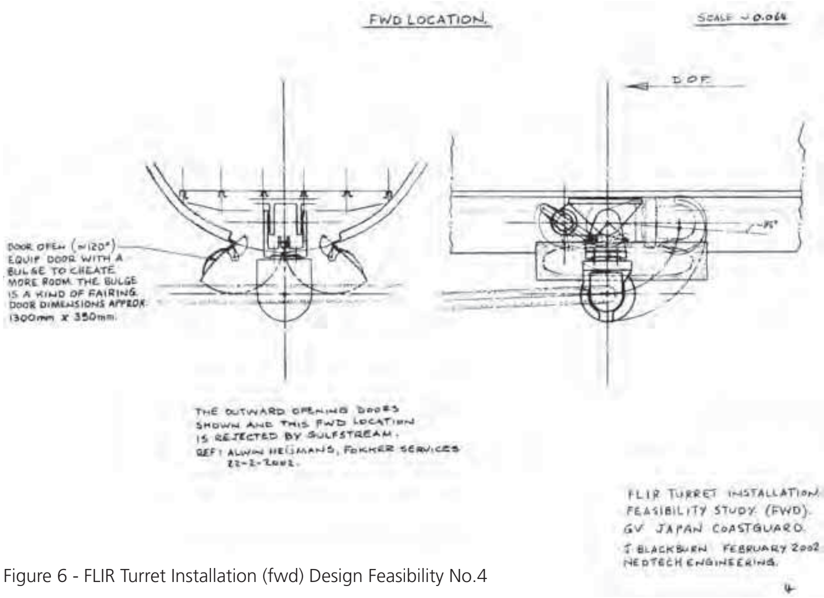


Figure 6 - FLIR Turret Installation (fwd) Design Feasibility No.4

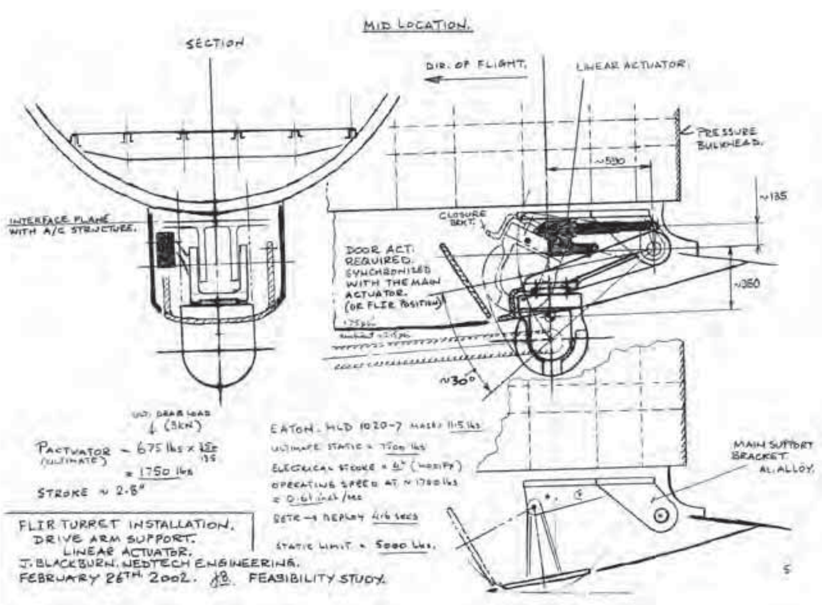


Figure 7 - FLIR Turret Installation (mid) Design Feasibility No.5

FLIR Turret Installation (mid) Design Feasibility No 2: Linear actuator. Inward opening doors.

- FLIR and drive system is basically one sub assembly
- Linear actuator is electro-mechanical
- Complex door drive system
  - Doors are inward opening (pivot points only)
  - Aft door is automatically driven through a separate linkage system
  - Fwd door is driven by an actuator through a separate linkage system
- Complex coaming design to create satisfactory sealing and aerodynamic smoothness (gaps and mismatches)
- Large opening in fairing

FLIR Turret Installation (mid) Design Feasibility No 3: Linear actuator. Inward opening door.

- FLIR and drive system is basically one sub assembly
- Sliding door on rails
- Rails are complex machinings
- Driven through a separate actuator (drive system and doors could be part of the fairing assembly)
- Adequate synchronisation required
- Complex rail interface with fairing
- Large opening in fairing

FLIR Turret Installation (fwd) Design Feasibility No 4: Rotary actuator and lever with link.

- FLIR and drive system is basically one sub assembly
- FLIR installation and outward opening doors not accepted by Gulfstream at this forward location, ref. Fokker Services 22-02-2002
- Extremely large opening in fuselage
- Activities stopped

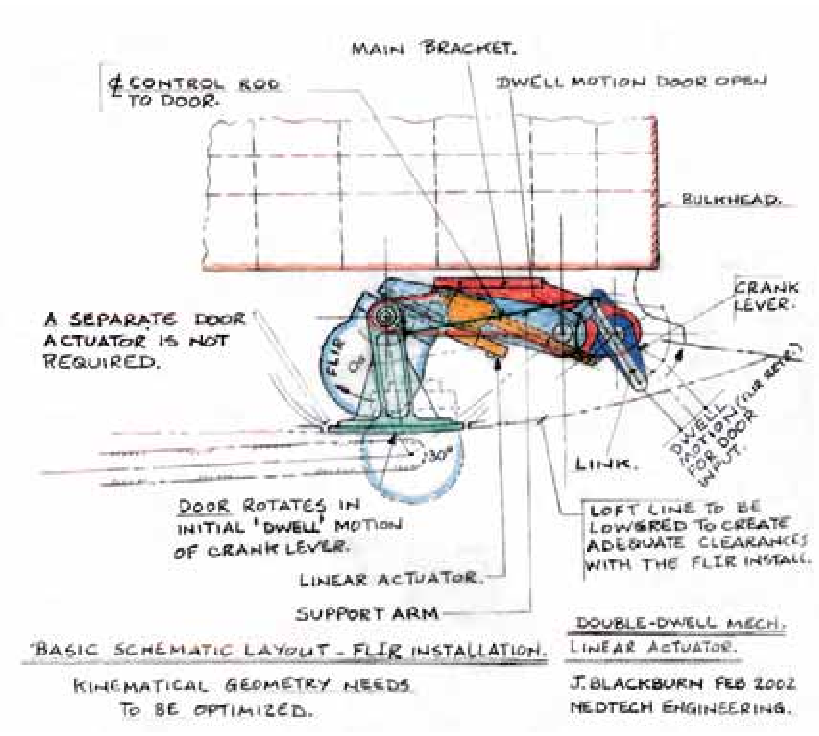


Figure 8 - FLIR Turret Baseline Design

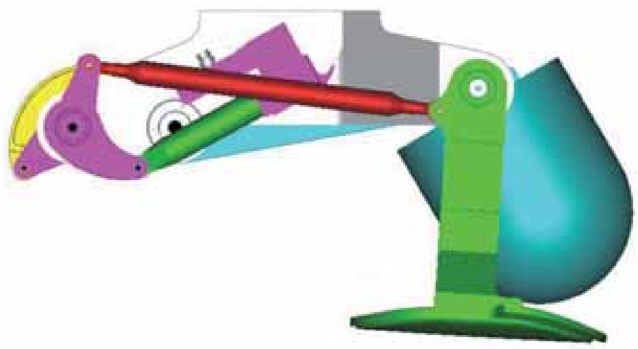


Figure 9 - FLIR Turret Installation (mid) Baseline Design

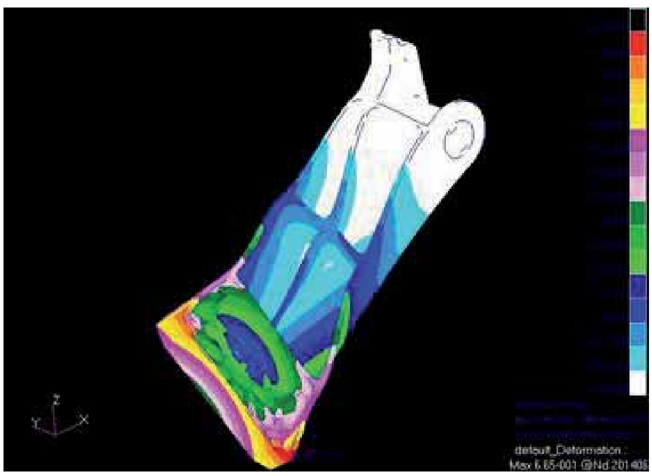


Figure 10 - FLIR Support Arm FEM Model

FLIR Turret Installation (mid) Design Feasibility No 5: Linear actuator. Inward opening (rotating) door.

- FLIR and drive system is basically one sub assembly
- FLIR is attached to an inclined support arm and driven by means of a linear actuator (electro-mechanical)
- Drive by means of a rotary actuator possible
- The support arm is hinged onto a machined main support bracket which interfaces with provisions on the fuselage
- The opening in the fairing is much smaller than the concepts 1 to 4
- Simple rotating door driven by means of a separate actuator (synchronisation required)
- Door interface at coaming on fairing is relatively simple. Should be satisfactory on aerodynamic smoothness (gaps and mismatches)
- The relatively simple and light fairing can easily be removed
- Hinge moment due to drag is lower with this concept

Feasibility study results

From the feasibility study the Atkins Nedtech team developed the following baseline design - figure 8. This design only required one actuator for the FLIR deployment/retraction and automatic door operation. Two dwell mechanisms were introduced in order to allow the correct sequencing of the door and FLIR motions.

The installation now comprised of a single sub-assembly that could be completed offline on the bench, this included the support, the drive mechanism and the door. This is important for installation, ease of assembly and time.

Design

The process that followed involved the optimisation of Kinematic Geometry with respect to weight and smooth operation and to ensure the start and finish points of the movement to deliver the required positioning of the FLIR.

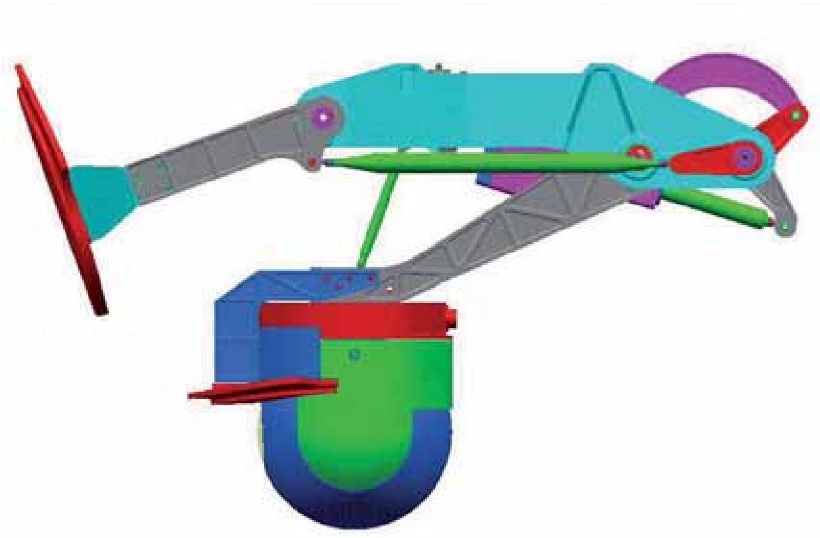


Figure 11 - FLIR Turret Assembly side view

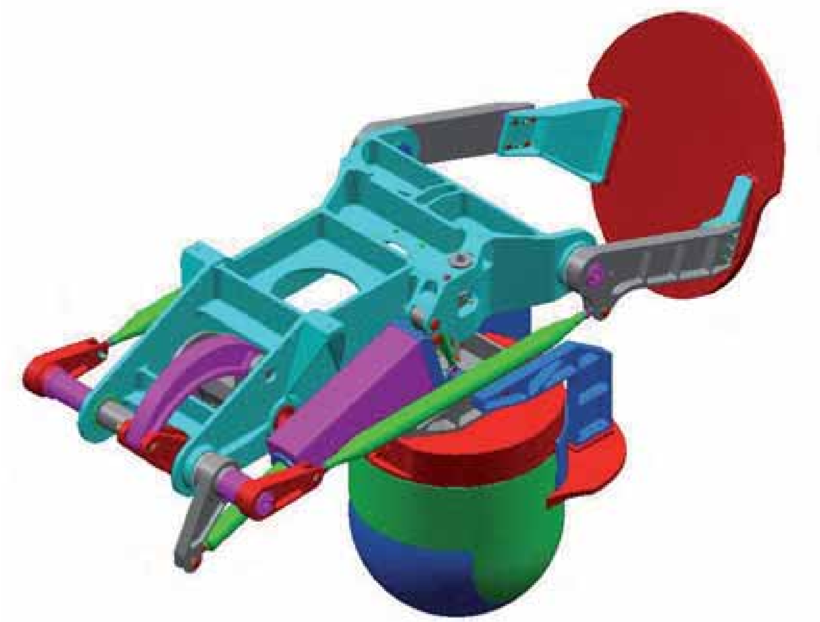


Figure 12 - FLIR Turret Assembly ISO view

Other work that took place at this stage involved scoping out the preliminary design requirements for the actuator which needed to consider the length of stroke required, the loads that the actuator would be subjected to during its life and brake requirements. This was then followed by preliminary sizing and a subsequently trade study of the actuation system itself which culminated in the decision to use an electrical system and not hydraulic.

At this point the baseline design was complete and reviewed with the customer.

With the baseline design in place we engaged with the radome supplier St Gobain. This allowed us work within the radome design envelope, the aperture size for the door and the subsequent sealing arrangement between the door and the radome. The baseline design could now be updated to include the interface data see Figure 9.

The next part of the process was to engage with a suitable Actuator supplier. A series of potential suppliers were identified and contacted. From the information received a selection process identified MPC as a potential supplier partner. A period of consultation was initiated to develop a detailed specification for the actuator.

With all of the design teams in place we optimised the design around the radome, door, sealing and actuator configuration.

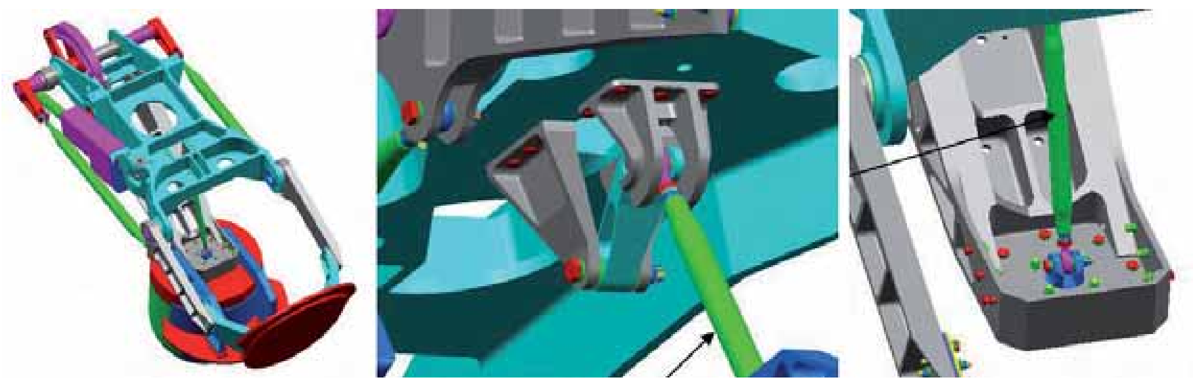


Figure 13 - FLIR Turret Assembly further ISO views

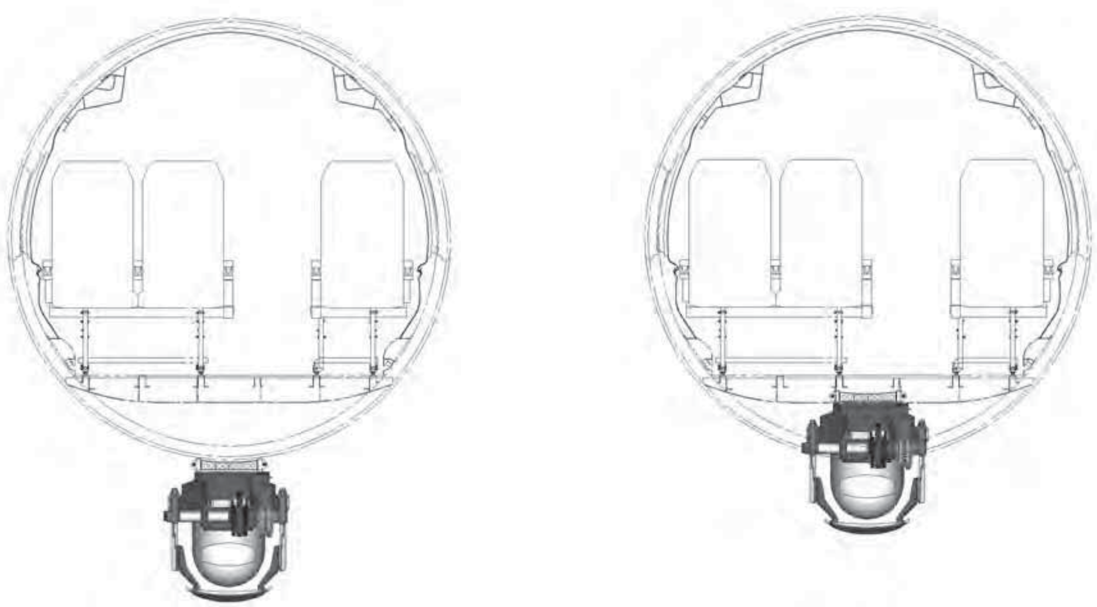


Figure 14 - Final position of FLIR Turret Assembly after upward translation of 12” (306mm)

Figure 15 - FLIR Turret Assembly installed onto aircraft undergoing testing

Figure 16 - In service

Enable

Critical Design Review (CDR)

The design and analysis was presented to all of the stakeholders; this included the OEM (Fokker Services), the customer (Gulfstream), the operator (Japanese Coast Guard), FLIR supplier and FLIR electronics design house.

This CDR identified a new and more stringent stiffness requirement from the radar supplier ‘Thales’. This centred around the FLIR’s alignment with the search radar.

At this point we had produced approximately 80% of the production drawings.

From the CDR outcome we created a local redesign in order to take into account the new stiffness requirement. This required implementation of a new mechanism by the addition of a foldable rod as can be seen in Figure 13.

With the new design change it was now necessary to perform some additional analysis FEM – finite element model and evaluation see Figure 10. This was completed alongside ongoing stress certification activities.

Figure 17 - In service

Drawing release

Production drawings were released in order to allow manufacturing of the parts to start.

With the interface with manufacture now up and running, FMEA's (failure mode and effects analysis) could be completed. This allowed manufacturing and process risks to be evaluated and mitigation plans to be identified. (Further model views Figure 11 – 13.)

Crash case analysis

Mounting an external FLIR camera raised important safety considerations. These, in turn, required Atkins Nedtech to demonstrate that the design complied with international airworthiness requirements.

A complete crash case analysis was performed. The team identified the possibility that the FLIR camera mechanism might, in the event of heavy landing or a crash, impact the structural integrity of cabin floor or even penetrate the cabin. This might restrict passengers from exiting or even inflict injuries.

After performing the study, analysis showed that the results of a 'wheels up landing' were compliant with the requirements of FAR25.561 and FAR25.562. This is shown in Figure 13. A movement in the vertical axis of 12" or 306mm did not cause penetration into the aircraft cabin.

Manufacture and testing

With the design now complete manufacture of the parts and subsequent assembly and integration onto the Aircraft could take place, this included local testing of the system (including the mechanism) see Figure 15.

In service

The design process is now complete and the aircraft with FLIR Turret assembly installed and integrated in fully operational order, meeting all requirements and is delivered to the customer.

The project was delivered on-time and within budget. After approximately five years in-service operation there have been no operational problems experienced by the customer.

A successful ending to a very interesting project.

The use of virtual reality as a tool in planning the deployment of ITS

Paul Grayston

Technical Director

Intelligent Transport Systems

Atkins

Hugh Woods

Senior Group Engineer

Transport Solutions

Atkins

Abstract

The "artist's impression" is often used as a persuasive tool in promoting infrastructure projects such as bridges and buildings. Unfortunately for many public facing ITS projects, static pictures do not give a tangible look and feel for decision makers and end users. Consequently, support for deployment might not be given, or it might not sufficiently meet user needs. This paper describes how VRML (Virtual Reality Modelling Language) has been used successfully to produce interactive demonstrations of ITS technology in advance of deployment or further research, even influencing the design of a major multi-modal interchange.

Introduction

In recent years many ITS technologies have reached the market. Transport authorities have become willing to invest in back office systems, control rooms and on-street infrastructure. Packages of transport investment include ITS as a component: city centre plans include parking guidance and VMS; public transport interchanges have real-time information; inter-urban highways have extensive information and monitoring.

Although, or perhaps because, ITS is relatively inexpensive in comparison with major civil engineering works, pre-implementation evaluation of where to deploy ITS devices is comparatively rare. Budgets are allocated to ITS but it may not always be an ITS specialist who advises on what to install and where.

More established technology such as motorway VMS may have design guidance drawn from experience. Some items such as passenger information systems may have specifications for their layout. However, many innovative technologies, especially those that convey information, can be difficult to envisage in-situ.

When the planned deployment is of a relatively large scale, the difficulty is multiplied. The consequences of poor planning or explanation by the designers and promoters may mean that the wrong ITS is installed or it is installed in the wrong place. Also, users may not understand/don't use the ITS or, worst of all, ITS is dropped as a component of a major scheme because it cannot be understood!

In several recent projects Atkins has used virtual reality (VR) modelling as a low cost method of demonstrating different types of potential ITS deployment at the preliminary or outline design stages. This paper describes the methodology and versatility of the approach. Examples are given of VR models where infrastructure designs have been altered as a result of early consideration of the ITS component and how the all-important stakeholder buy-in has been influenced by the process.

What does VR offer ITS?

VR has been used in applications such as driving simulators where the test user is surrounded by projected images and their reactions are monitored. PC processing speeds and graphical capabilities has advanced to the point where VR examinations can be made at the desktop, even across the internet. In other industries and sectors, such as manufacturing and medical, visualisation is a common tool.

There are many pressures for implementing ITS. Some come from a "technology push" perspective where it is known that the actual technology works (though this may be as pre-production models). Some come from a desire by an institutional (possibly political) stakeholder for a technical solution to a problem, a "surely computers can solve this" approach. Most well designed schemes are the result of acting upon end user needs, though these can often be vague, and have technical or institutional issues to overcome e.g. who owns data, rail solutions not fitting bus environments.

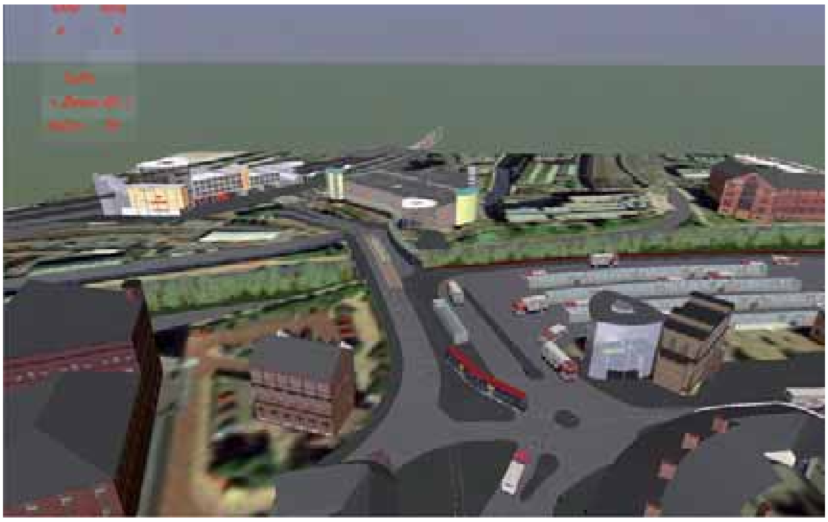


Figure 1 - Aerial View of Wolverhampton Interchange VR model – Note: all vehicles move!

In almost every case there will be an idea of where the ITS might be physically positioned, but this will be subject to survey, sight lines may be difficult to discern for instance. By using a properly scaled VR model the complete effect can be viewed. When it comes to considering the needs of different types of users (e.g. elderly and disabled, visitors, regular users) VR can be used to test their requirements. VR also allows non-ITS components such as fixed information signs and road markings to be incorporated giving the full context of the environment in which the technology is planned.

Creating the VR model

Using VRML97

The VRML97 (ISO Standard 14772) modelling language has been used to create the models. It defines the geometry, topography and appearance of objects and how they can interact on-screen. This allows vehicles to move about, information screens and indicators to change and clocks to tick away in real or accelerated time.

VRML97 was designed for use on the internet. The files can be very compact for fast downloading. Many proprietary software products can export VRML but the results tend to be very face-heavy. Atkins has, therefore, chosen to develop in-house tools and components for performance and functionality.

Importing objects and traffic models into the model

Proprietary and in-house software tools enable VRML to import geometry from design packages, such as AutoCAD™ and Bentley MX™, and to clad this with digital images including photographs.

Other data is captured from internet sites, databases and spreadsheets. Using manufacturers' datasheets, photographs and style guides a library of objects can be created that can be located anywhere within the model. In this way, for example, road markings and sign faces can conform to national standards. A wide range of road and rail vehicles can "drive" correctly to their wheelbases and articulation.

Vehicle paths and speed profiles can be choreographed, or alternatively simulated from traffic modelling packages such as VISSIM™ or Paramics™.

Importing traffic simulations into the virtual world adds the third dimension: levels, gradients and crossfalls are applied. It integrates the traffic with all the other disciplines while maintaining total freedom of navigation and choice of scheme details, lighting conditions etc. At present the traffic simulation tools output fixed-path movies only. VR is not a substitute for good traffic modelling though!

VR maintains a real time clock which can be used to control a wide range of continuous and intermittent actions and responses to events. All modes of traffic can be given realistic speeds, accelerations and stopped periods. All types of variable message signs can be controlled, for example traffic signals, flashing lights and clocks. Passenger information systems can involve a huge amount of text on different types of display equipment. VRML provides the means to build each type as a component that updates as new information is received, and a timetable component to issue data to them all. Time in the virtual world can be scaled up or down so that an hour's timetable can be shown in a minute if required.

Figure 2 - During the construction of the model it was possible for the clients to compare LED (left) with Plasma (right) types of information displays

VIRTUAL REALITY

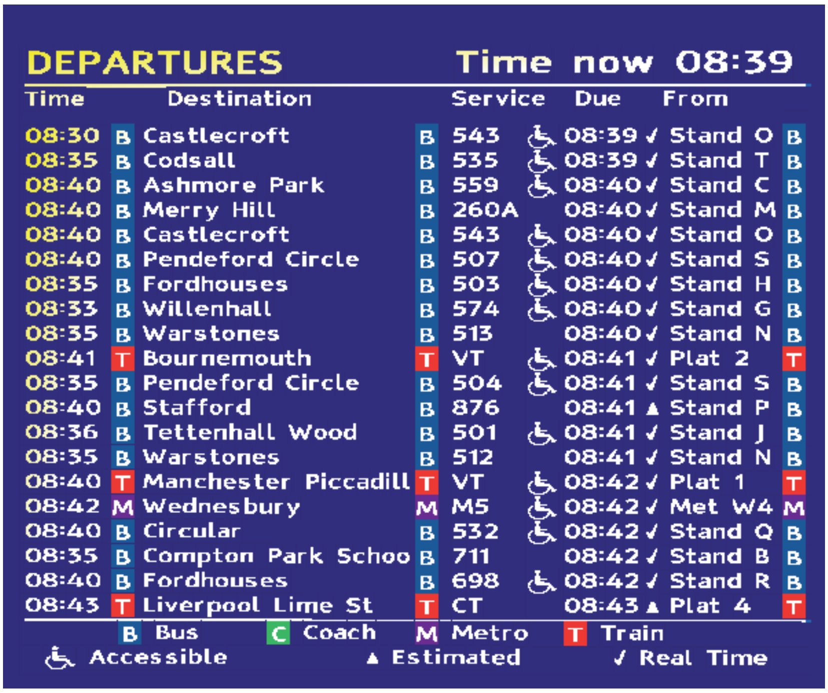


Figure 3 - Screenshot of a "Virtual" multimodal indicator screen within the interchange.

Viewing the VR model

VRML files can be viewed using a "plug-in" to a web-browser. The effect is similar to a computer game: by moving the mouse the user can roam at will around the VR model.

This puts the users in control, with the choice of walking on the ground, flying over it or selecting one of the fixed or moving viewpoints. They can inspect different design or route options or construction phases. They can experience night or fog. They can be in a vehicle or watch it negotiate a junction. They can see an avatar (a controllable humanoid character) interacting with the model as a vivid check on the scale of the planned infrastructure.

Benefits of using VR

The greatest benefits for using VR for ITS have been seen when the models are created soon after the "user need definition" phase of a project. At this stage the "What if?" questions are fresh in the minds of the stakeholders and can be conceptualised and revised if necessary using:

- 3D modelling with interaction in a real time environment for the user. Different conditions can be simulated, hotspots can be used to bring up interactive screens
- Intelligent components such as indicators can be changed to different sizes, types, fonts etc
- Models can be shared over networks, intranets or the internet
- Objects in the virtual world can link to other web pages, drawings and documents as a 3D information system
- Objectors fears can be allayed or clarified for quicker resolution
- VR can give real savings in time and cost and influence the design of civil engineering and structures.

Experiences of using VR for ITS

The perception of the stakeholder must be managed. The model will not have Playstation™ levels of functionality and graphics resolution!

Demonstrations of early versions of a VR model, where developed components may not yet be in the correct positions or be unrendered and represented by "wireframes" may have a negative impact if not explained well. The finishing touches are generally quite quick to implement, but where there are choices (size, manufacturer type, colour etc.) they are best to be done whilst the model is being built.

When the final model is produced it is essential to ensure that a fast PC fitted with a capable graphics card is used to demonstrate for maximum effect. First demonstrations should be by someone who is experienced at navigating through the model quickly as it requires practice to master.

A good use for VR models of ITS is by way of "effectiveness tests". Before construction stakeholders are invited to devise a variety of scenarios that they would like to review within the finished model. Ideally, there should be several variants or ways to satisfy these scenarios. The actual effectiveness of the test can be easily captured in avi movie format which can easily be incorporated into presentations.

Seeing a design in 3D with recognisable buildings and roads can be an inspiring experience for stakeholders. As the possibilities of the model become clearer, requests for additional features often result!

Project examples

WIP-PAT – Wolverhampton Interchange

The Wolverhampton Interchange is an ambitious regeneration project that will create a memorable gateway for users of the city's transport infrastructure. The project, led by Wolverhampton City Council (WCC) and Centro (the West Midlands public transport body), will replace existing rail and bus station architecture, remodel local roads, improve transport links, prepare for the Midland Metro extension and maintain a navigable canal waterway through the area.

Before work began on site a VR model was constructed as part of the work to define the integrated information strategy. Various types of information indicator, kiosk information points and wi-fi hotspots were located within a conceptual design of the proposed interchange.

The model was used successfully to test a variety of user scenarios developed by the project steering group.

Concepts from best practice inspection visits were incorporated and have influenced the overall architectural design.

Figure 4 - A282 VMS and illuminated road studs screen shot

Notably, the VR model has resulted in an additional indoor information point being incorporated in the masterplan for the interchange.

It was considered essential that real services to real places should be shown on the indicators to make the model convincing. Actual timetable information was fed to the signs in accelerated real time giving a true feel of realism. The VR modelling also allowed different layouts of information to be trialled, giving useful guidance on what users were most comfortable with and understood easiest.

A282 VMS Scheme - Variable Message Signs and Illuminated Road Studs Scheme at Junction 1A (Dartford Tunnel) on the M25 Motorway.

The M25 orbital motorway becomes A282 for a short length on the Kent side of the Dartford Tunnel. At Junction 1A northbound the peak traffic demand is too high for the single lane off-slip. An option proposed to the Highways Agency by Atkins was to widen the slip and to dedicate the inside lane to turning traffic during peak periods. This would be achieved by Variable Message gantry signs and illuminated red LED road studs.

The VR model developed to evaluate the option included user interaction controls to simulate:

- Environmental conditions of fog and/or night with street lighting
- Peak or off-peak state of the VMS gantries and the illuminated road studs
- Individual control over the groups of studs as they would be wired – to simulate power supply failures
- Play/Pause/Reset driver's eye view fixed path animations.

Speed influencing road studs

As part of a commission to manage part of a highway network Atkins undertook research into the psychological effect of using pulsating road studs to demonstrate how speed limit adherence might be helped by "moving" the lit road studs at the speed limit. The idea is that traffic would drive at or about the speed of the "movement" of studs, and would know if they were exceeding the speed limit because they had overtaken the lights.

The future

Atkins has been using VR on a range of projects since the 1990s. The ISO Standard VRML97 is in the process of evolving to new open standards for real time communication known as X3D (see www.web3d.org). It is possible to envisage that soon designs can be created and shown on the web using VR. There is even the possibility of using VR as part of ITS applications, for instance in control rooms.

VR for designing ITS provision is well suited to major development projects where architects and civil engineering are the main components, as shown by the Wolverhampton project. It is also useful for short conceptual "will this work" commissions such as the road stud projects.

It is clear from our experience that well designed models can be persuasive to stakeholders and result in better design decisions being made.

Fracture propagation of CO_2 pipelines

Andrew Cosham
Consultant
Atkins Boreas

Robert J Eiber
Robert J Eiber Consultant Inc.
USA

Abstract

The fourth report from the Intergovernmental Panel on Climate Change states that "Warming of the climate system is unequivocal…" It further states that there is a "very high confidence that the global average net effect of human activities since 1750 has been one of warming". One of the proposed technologies that may play a role in the transition to a low-carbon economy is carbon dioxide capture and storage (CCS). The widespread adoption of CCS will require the transportation of the CO_2 from where it is captured to where it is to be stored. Pipelines can be expected to play a significant role in the required transportation infrastructure.

The transportation of CO_2 by long-distance transmission pipeline is an established technology; there are examples of CO_2 pipelines in USA, Europe and Africa. The design and operation of a CO_2 pipeline is more complicated than a typical hydrocarbon pipeline, because of the highly non-linear thermodynamic properties of CO_2 and because it is normally transported in a pipeline as a dense phase fluid. There are number of issues to be considered. Furthermore, CO_2 captured from fossil fuel power stations may contain different proportions and/or types of impurities from those found in the sources of natural or anthropogenic CO_2 transported in the existing CO_2 pipelines.

Fracture propagation control is one such issue that requires careful consideration in the design of a CO_2 pipeline. CO_2 pipelines may be more susceptible to long running ductile fractures than hydrocarbon gas pipelines. The need to prevent such propagating fractures imposes either a minimum required toughness (in terms of the Charpy V-notch impact energy) or a requirement for mechanical crack arrestors. Indeed, fracture propagation control has implications for the diameter, wall thickness and grade of the pipeline, in addition to the CVN impact energy of the line pipe steel, because in some situations the requirement for fracture propagation control will dictate the design of a CO_2 pipeline.

The issues surrounding fracture propagation control in a CO_2 pipeline are illustrated through the means of two simple design examples: a 24in (609.6mm) diameter pipeline with a design pressure of 100barg, and a 18in (457.2mm) diameter pipeline with a design pressure of 180barg. It has been shown that fracture propagation control in a CO_2 pipeline can be addressed relatively simply. Some care is required because the trends observed in CO_2 pipelines are not the same as those in natural gas pipelines, and the required toughness to arrest a ductile fracture may be very sensitive to small changes in the design parameters. Nevertheless, provided that fracture control is considered early in design, any constraints on the design can be identified and, in principle, resolved without too much difficulty. It is important not to forget that transportation is an implicit, and essential, part of CCS.

Introduction

The fourth report from the Intergovernmental Panel on Climate Change (IPCC)[1] states that "Warming of the climate system is unequivocal, as is now evident from observations of increases in global average air and ocean temperatures, widespread melting of snow and ice, and rising global average sea level". It further states that there is "very high confidence that the global average net effect of human activities since 1750 has been one of warming".

Provided that action is taken soon to reduce emissions of 'greenhouse gases', the potentially severe effects of climate change can be avoided, without excessive cost [2,3]

Carbon dioxide capture and storage (CCS) is one of the technologies that has been proposed to reduce emissions of carbon dioxide (CO_2) to the atmosphere from fossil fuel power stations. An IEA (International Energy Agency) study estimated that the widespread adoption of CCS technologies could contribute approximately 20% of the reduction in emissions required to reduce projected emissions in 2050 to their 2003 levels (although, in comparison, energy efficiency measures could contribute approximately 45%)[4].

A second factor that has raised the profile of CCS amongst governments is the need to ensure security of energy supply. Clean-coal technology is seen as having the potential to make a significant proportion of a diverse low-carbon energy mix.

Figure 1 - Fracture control

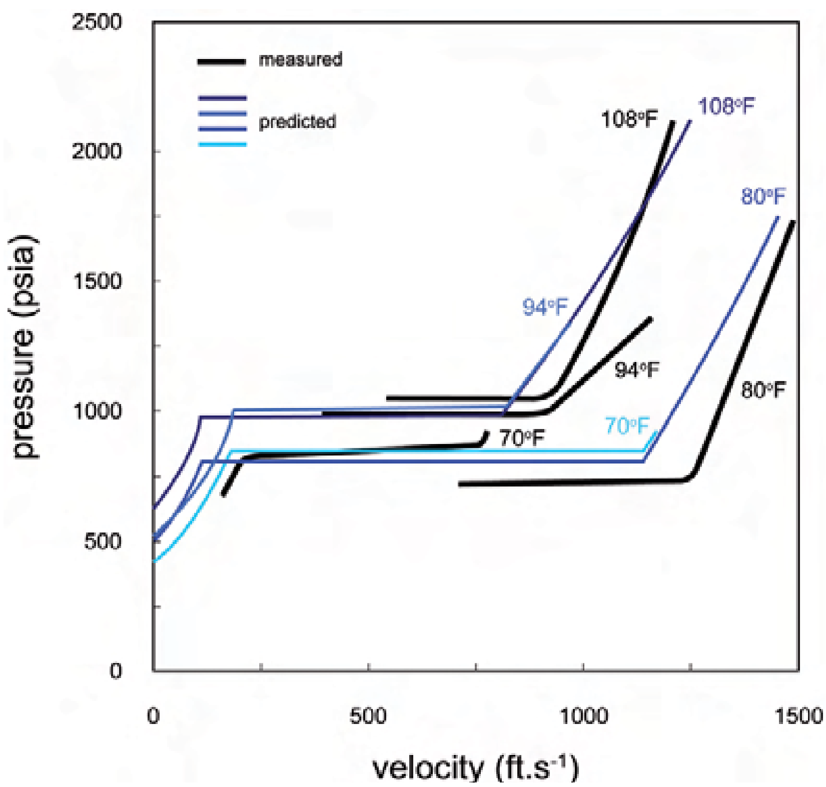


Figure 2 - Experimental and theoretical decompression curves for CO_2 (after Maxey (1986))

Table 1 - Differences between existing and new CO2 pipelines

Existing CO_2 pipelines	New CO_2 pipelines
EOR	CCS
nearly pure CO_2 from dome fields	impurities, depending on the capture method
remote, unpopulated areas	populated areas
'static' demand	fluctuation demand (due to load factors)
	higher throughput ?

Coal is one of the most available sources of energy in the USA, and similarly in the UK and other parts of Europe, not to mention the rest of the world. A clean-coal power station would incorporate CCS.

The UK government, through the Department for Business, Enterprise & Regulatory Reform (BERR) has organised a competition to develop a commercial-scale coal-fired plant capable of demonstrating the full range of CCS technologies, the CCS Demonstration Competition now falls under the remit of the Department for Energy and Climate Change (DECC).

The project envisages the construction of 300-400 megawatt plant, capable of capturing up to 90% of its CO_2 emissions. A number of countries around the world, including Australia, USA and Norway, also have government supported projects to develop commercial-scale CCS power stations. A 30MW pilot plant at the Schwarze Pumpe power station in Germany, demonstrating carbon capture and storage, opened in September 2008[5,6].

However, initially the CO_2 will be transported to the storage site by road tanker, not pipeline. In simple terms, there are two types of carbon capture technology: pre-combustion and post-combustion, with various methods of implementing either technology. The composition of the 'captured' CO_2 will depend on the process used to capture it. Once the CO_2 is captured it needs to be transported to where it is to be stored. Pipelines can be expected to play a significant role in the required transportation infrastructure[7]. Transportation is an essential part of carbon capture and storage, but sometimes appears to be something of a Cinderella subject.

The transportation of CO_2 by long-distance transmission pipeline is an established technology. The design and operation of a CO_2 pipeline is more complicated than a typical hydrocarbon pipeline[7,8]. One of the issues that needs to be considered is fracture control, and specifically fracture propagation control.

Fracture control is concerned with designing a pipeline with a high tolerance to defects introduced during manufacturing, construction and service; and preventing, or minimising the length of, long running fractures. CO_2 pipelines are potentially more susceptible to long running fractures than conventional natural gas pipelines.

Existing CO_2 pipelines transport CO_2 from CO_2 dome fields and plants processing gas from reservoirs with a high proportion of CO_2. 'Captured' CO_2 may have a different composition.

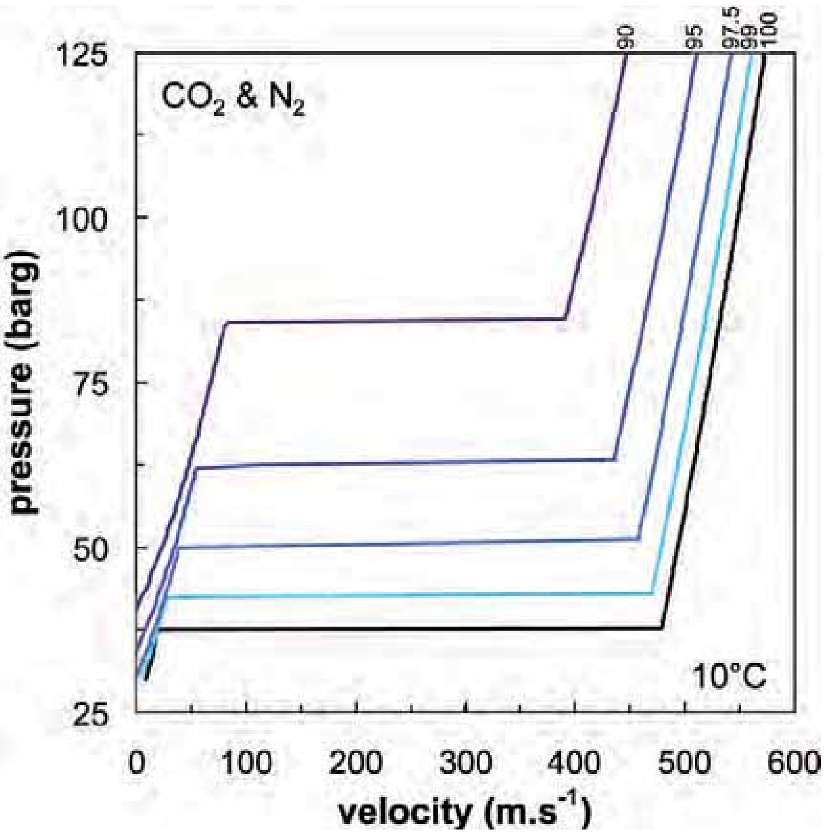


Figure 3 - Theoretical decompression curves for mixtures of CO_2 & N_2

The composition of the CO_2, i.e. the type and proportion of the impurities in the CO_2, may have a significant effect on the susceptibility of the pipeline to a running fracture, in addition to its effect on hydraulic design[9].

In this paper the issues surrounding fracture propagation control in a CO_2 pipeline are illustrated through the means of two simple design examples.

Fracture control

Fracture control is an important consideration in the design of a pipeline. A fracture control plan for a pipeline will consider two issues see Figure 1:

(1) Fracture initiation control

(2) Fracture propagation control.

A propagating (or running) fracture will result in the loss of many lengths of line pipe, and hence is undesirable.

Fracture propagation control is achieved by ensuring that the toughness of the line pipe steel is sufficiently high to arrest propagating fractures. Fracture propagation control needs to be considered in pipelines conveying gaseous fluids, two phase fluids, dense phase fluids, or liquids with high vapour pressures. Propagating fractures are described as either brittle or ductile.

Brittle propagating fractures are prevented by ensuring that the line pipe steel is operating on the 'upper shelf'. Ductile propagating fractures are prevented by specifying a minimum toughness to ensure that a ductile fracture will arrest; or, if the required toughness is too high, by using mechanical crack arrestors.

Line pipe specifications and pipeline design codes specify toughness requirements in terms of the minimum shear area as measured in a Drop Weight Tear Test (DWTT) to address the 'upper shelf' requirement[10-13]. Brittle fracture propagation is not an issue in modern line pipe steel.

A ductile fracture will not propagate if there is insufficient energy in the system to overcome the resistance to propagation. The resistance to a running fracture can be characterised by the Charpy-V notch (CVN) impact energy of the line pipe steel - although the relationship between CVN and fracture resistance becomes non-linear at high impact energies (when the full size impact energy exceeds approximately 100J).

The driving force for a running fracture is the internal pressure. If the fluid in the pipeline decompresses slowly, i.e. high pressures at low decompression wave velocities (as is the case for CO_2, see Figure 2), then a higher toughness is required to arrest the running fracture.

The Battelle Two-Curve Model, widely used in fracture control studies, expresses the resistance and driving force in terms of the fracture and gas decompression wave velocities [11,12]. For CO_2 it can be shown that fracture propagation control can be conservatively simplified to determining the toughness required to ensure that the 'arrest pressure' is greater than the 'saturation pressure' [14-18]. The 'arrest pressure' can be determined using part of the TCM. The 'saturation pressure' can be determined from a phase diagram (or a gas decompression program), given the initial pressure and temperature.

The transportation of carbon dioxide by pipeline

CO_2 is transported in pipelines over long distances as a dense phase fluid, for operational and economic reasons. The typical range of operating pressures and temperatures of CO_2 pipelines are: 1,250psig (86.2barg) to 2,220psig (153barg), 40°F (4°C) to 100°F (38°C)[8,19].

CO_2 pipelines are susceptible to propagating ductile fractures because the CO_2 is transported in the dense phase. It is a high vapour pressure liquid. At high pressures, supercritical CO_2 behaves as a liquid, and has a liquid-like density, but it yields a very large volume of gas when its pressure is lowered[20].

Fracture propagation control requires careful consideration in the design of a CO_2 pipeline, along with a number of other issues such as hydraulics and corrosion control[8].

The transportation of CO_2 by long distance transmission pipeline is an established technology. The Canyon Reef Carriers pipeline system in West Texas was commissioned in 1972[7]. There are over 2,500km of CO_2 pipelines in the USA and Canada for enhanced oil recovery (EOR) projects[7], with other pipelines in the Netherlands, Turkey, North

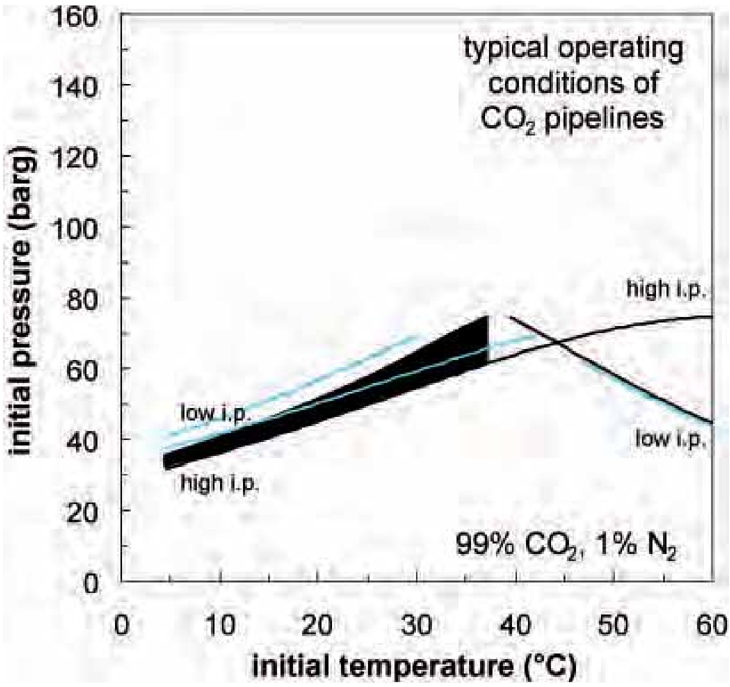


Figure 4 - The effect of initial conditions on the saturation pressure of a mixture of CO_2 & N_2

Africa and Norway (the latter, the Snøhvit pipeline, being the world's first offshore CO_2 pipeline[21]). The source of the CO_2 transported in these pipelines is either natural or anthropogenic (i.e. man-made), although none of the anthropogenic sources are (yet) 'captured' CO_2 from fossil fuel power stations.

In a wider context, it is also worth noting that there are differences between the existing CO_2 pipelines, which were (with the odd exception) constructed for the purposes of EOR, and the new CO_2 pipelines that will be constructed as part of the required transportation infrastructure for CCS. Some of these differences are summarised in Table 1. The implications arising from these differences will need to be addressed.

The fact that long distance, high-pressure CO_2 pipelines have been designed, constructed and operated successfully for many years indicates that the issues associated with the design and operation of CO_2 pipelines can be addressed.

Several CO_2 pipelines in USA have mechanical crack arrestors installed at regular intervals along their length, because line pipe with a sufficiently high toughness was not available when the pipelines were constructed[7,8,22,23]. Fitting crack arrestors is expensive; retro-fitting them to existing pipelines is even more so.

The renewed interest in CO_2 pipelines, both new and the change of use of existing pipelines from their current service to CO_2 service, means that it is informative to look again at the issue of fracture propagation control in CO_2 pipelines.

The decompression characteristics of carbon dioxide and its implications for fracture control

CO_2 exhibits highly non-linear thermodynamic properties, and it departs significantly from ideal gas behaviour as the pressure increases. The critical point of CO_2 is at a pressure of 73.77bar (1,070psi) and a temperature of 31°C (88°F). The presence of impurities, such as methane or hydrogen, can have a significant effect on the behaviour of the fluid[9,18].

The decompression characteristics of a fluid have a significant effect on the toughness required to arrest a running ductile fracture. It is the decompression characteristics of CO_2 that mean that fracture propagation control requires careful consideration.

CO_2 is normally transported as a dense phase fluid. Consider a rupture in a CO_2 pipeline. The CO_2 initially decompresses rapidly as a liquid. The decompression path then crosses the phase boundary, and the now two phase fluid decompresses much more slowly. Experimentally determined decompression curves for CO_2 are illustrated in Figure 2. The discontinuity in the decompression curve occurs when the decompression path crosses the phase boundary; the pressure at which it crosses the phase boundary is the saturation pressure.

The decompression in a pipeline following a rupture can be approximated as an isentropic process. GASDECOM is a program for calculating the decompression curve for mixtures of hydrocarbons [11,12]. It is based on the Benedict-Webb-Rubin-Starling (BWRS) equation of state, with modified constants known to give accurate estimates of isentropic decompression behaviour. Figure 2 compares the measured decompression curves with theoretical predictions using a modified version of GASDECOM. The agreement between the experimental and theoretical decompression curves is relatively good. Consequently, a reasonable estimate of the saturation pressure, for given initial conditions, can be obtained assuming an isentropic decompression and a phase boundary described by the BWRS equation of state.

The saturation pressure is key to determining the toughness required to arrest a propagating ductile fracture in a CO_2 pipeline. Factors that increase the saturation pressure will increase the arrest toughness.

The presence of impurities has a significant effect on the saturation pressure. Impurities such as hydrogen, nitrogen and methane will increase the saturation pressure[18]. Theoretical decompression curves for mixtures of CO_2 and N_2, from 0% N_2 to 10% N_2, are shown in Figure 3[18], illustrating the significant increase in the saturation pressure as the proportion of N2 increases. This increase in the saturation pressure will significantly increase the arrest toughness (as discussed further in the example below).

The initial pressure and temperature of the fluid also have a significant effect on the saturation pressure. Theoretical predictions of the saturation pressure for 100% CO_2, and 99% CO_2 and N_2, for a range

ENERGY

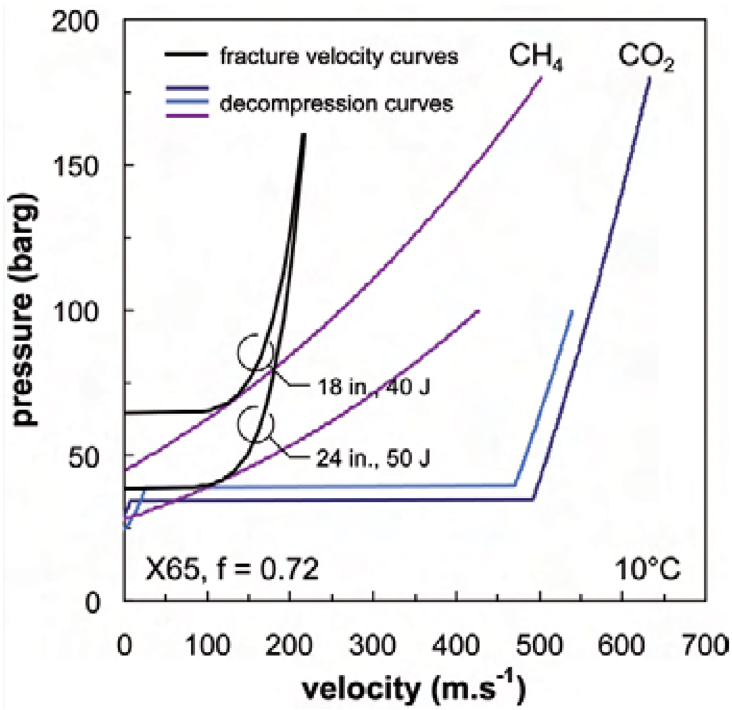


Figure 5 - Theoretical decompression curves for pure CO_2 and pure CH_4, and fracture velocity curves for the 18 and 24in pipelines

of initial pressures and temperatures are shown in Figure 4[18]. Considering the typical operating pressure and temperature range of CO_2 pipelines, increasing the initial temperature and/or decreasing the initial pressure will increase the saturation pressure.

Considering the above, it is clear that the composition of 'captured' CO_2 and the pipeline operating conditions need to be well defined at the early stages of the design, so that the implications for achieving fracture propagation control can be addressed.

An example

The issues associated with achieving fracture propagation control in a CO_2 pipeline are illustrated through two examples:

- 24in (609.6mm) diameter pipeline with a design pressure of 100barg
- 18in (457.2mm) diameter pipeline with a design pressure of 180barg.

The above pipeline diameters and design pressures are representative of what might be required to transport the CO_2 produced by a 1,600MW coal fired power station over a distance of approximately 200km. A power station of this size would produce something of the order of 8million tonnes per year of CO_2.

In both cases, the line pipe grade is taken to be API 5L X65 and the design factor is 0.72. The wall thicknesses are calculated accordingly. The wall thicknesses of the 24in and 18in diameter pipelines are 9.45mm and 12.76mm, respectively (giving diameter to wall thickness ratios of 64.5 and 35.8, respectively).

In practice, a standard API 5L wall thickness would be adopted and depending upon the pipeline design code (e.g. PD 8010-1,2 : 2004 or ASME B31.4[24,25,26]) wall thickness manufacturing tolerances may also need to be considered. In addition, the design factor may not dictate the minimum required wall thickness (e.g. resistance to external interference in onshore pipelines, collapse and stability in offshore pipelines). For simplicity, none of these issues is considered here. The implication of these issues is that the wall thickness will tend to be slightly (or significantly) greater than the minimum required to satisfy the limit on the design factor.

An increase in the wall thickness is beneficial from a fracture control perspective because it reduces the arrest toughness (see below).

It is assumed that the line pipe steel is operating on the upper shelf, i.e. 85% shear area in a DWTT at the minimum pipeline operating temperature .

It is assumed that the pipelines are onshore, although this has little significance for the purposes of these examples. In some design codes and standards the requirements is expressed in terms of the minimum design temperature, and in others in terms of the minimum operating temperature. In most cases the difference is not significant. It is more conservative to use the minimum design temperature.

All references to Charpy V-notch (CVN) impact energy in the following refer to upper shelf values (i.e. 100% shear) measured using full size (1/1) specimens tested at the minimum pipeline operating temperature. In a line pipe specification, the required impact energy may be expressed as the minimum of three test results or the average of three. The implications of this issue are not considered here.

A Methane (natural gas) Pipeline

Firstly, it is instructive to consider the requirements for fracture propagation control if the above two pipelines were transporting methane (CH_4).

There are a number of different methods that could be used to estimate the required Charpy V-notch impact energy to arrest a running ductile fracture (i.e. the arrest toughness). The EPRG Recommendation for crack arrest toughness for high strength line pipe steels[13] would be the simplest approach. The Battelle Short Formula (SF)[11,12], as recommended in ASME B31.8[27], is slightly more complicated.

Table 2 - Required full size CVN impact energy for a CH4 pipeline

	24 in, 100barg	18 in, 180barg
EPRG	40 J	40 J
SF	53 J	53 J
TCM	50 J	40 J

The most accurate, but also the most complicated approach, would be the Battelle Two Curve Model (TCM)[11,12]. It is important to emphasise that the TCM would not normally be used for a CH_4 pipeline (and, in any case, the SF is an approximation to the TCM). It is considered here because its use illustrates the implications of the different decompression curves for CH_4 and CO_2.

Table 2 gives the required toughness for the two pipelines calculated using the EPRG recommendations, the SF and the TCM. The SF is conservative with respect to the TCM, as would be expected. The EPRG recommendations give the lowest required toughness. The EPRG recommendations for X70 and below are based on 0.75 times the AISI formula, and take into account the statistical distribution of the CVN impact energy in an actual line pipe supply (mid-1990s data).

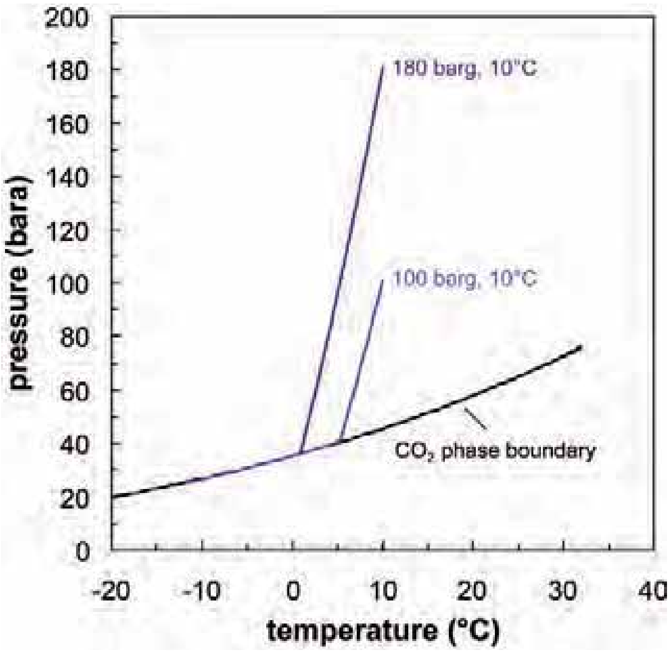


Figure 6 - Isentropic decompression paths

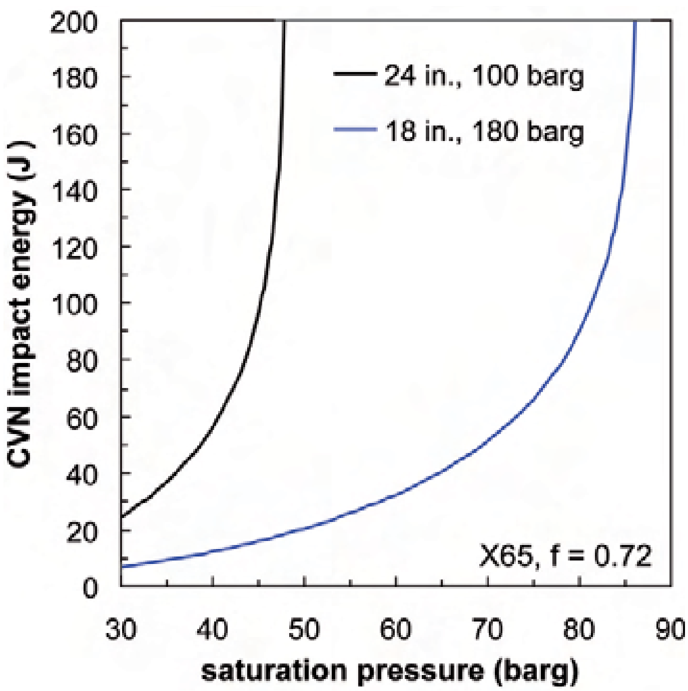


Figure 7 - Effect of saturation pressure on the CVN impact energy required to achieve fracture propagation control in the 18in and 24in pipelines

The minimum toughness specified by the EPRG recommendations ensures that 50% of the line pipe will meet the required toughness. The SF and TCM are simply formulae for calculating the arrest toughness. The results of the three different criteria are broadly comparable. Modern line pipe would easily exceed the required toughness.

Further insight into the underlying behaviour is given in the results of the TCM. Figure 5 shows the theoretical decompression curves for pure methane, based on an initial temperature of 10°C and initial pressures of 100 and 180barg, and the fracture velocity curves for the 18in and 24in diameter pipelines, based on the minimum arrest toughness (and hence the respective decompression and fracture velocity curves intersect at a tangent). The decompression curves are characteristic of the decompression of a gaseous fluid in the gaseous phase. The decompression curve for an initial pressure of 180barg is more severe than that for an initial pressure of 100barg (at any given decompression wave velocity, the decompression pressure is higher). All other factors being equal, increasing the initial pressure increases the arrest toughness. The decompression curve is only part of the picture; it depends only on the initial pressure and temperature and the fluid composition. The fracture velocity curves illustrate the effect of pipeline geometry and grade. The 18in diameter pipeline has a higher resistance to a ductile fracture than the 24in pipeline because the diameter to wall thickness ratio is smaller (for the same toughness, at any given fracture velocity, the required driving pressure is higher) The smaller diameter to wall thickness ratio has a second effect. At any given decompression wave velocity, the hoop stress in the 18in pipeline is lower than that in the 24in pipeline, even though the pressure is higher.

The higher fracture resistance of the 18 in. pipeline more than offsets the higher driving force implied by the decompression curve. Consequently, a lower arrest toughness is predicted for the 18in180 barg pipeline than the 24in 100 barg pipeline.

ENERGY

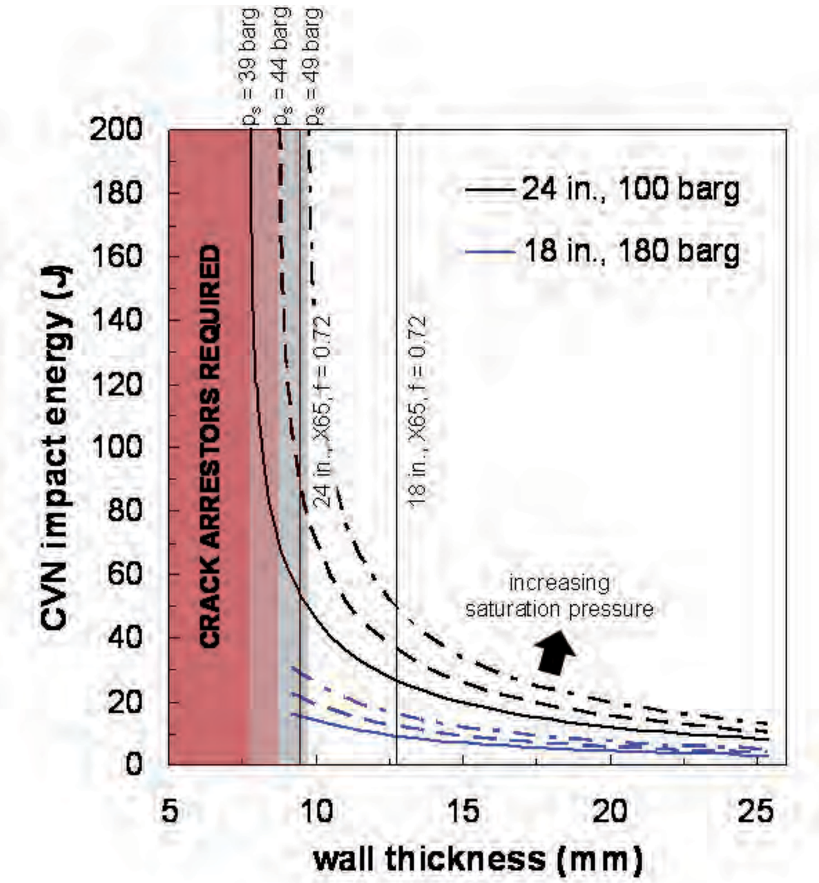


Figure 8 - Effect of wall thickness on the CVN impact energy required to achieve fracture propagation control in the 18in and 24in pipelines

The following general trends for a CH4 pipeline can be identified:

- The higher the initial pressure, the more severe is the decompression curve
- The lower the initial temperature, the more severe is the decompression curve
- The smaller the diameter to wall thickness ratio, the lower is the arrest toughness
- The lower the design factor, the lower is the arrest toughness.

A Carbon Dioxide Pipeline

Consider now the same 18in and 24in diameter pipelines, but transporting carbon dioxide (CO_2) rather than methane.

Figure 5 shows the theoretical decompression curves for pure carbon dioxide, based on an initial temperature of 10°C and initial pressures of 100 and 180barg. The decompression curves are very different from those of methane decompressing from the same initial conditions.

The decompression curves are characteristic of the decompression of a dense phase fluid. The discontinuity, or plateau, in the decompression curve occurs when the decompression path crosses from the single phase region (liquid) into the two phase region (liquid-gas). The marked difference between the decompression curves for methane and carbon dioxide illustrated in Figure 5, clearly demonstrates why it is only necessary to consider the saturation pressure when determining the required arrest toughness in a CO_2 pipeline. The full decompression curve is not required. The isentropic decompression paths corresponding to the decompression curves are shown in Figure 6. The saturation pressure is the pressure at which the decompression path intersects the phase boundary. The saturation pressure for an initial pressure of 180barg is lower than that for an initial pressure of 100barg (the estimated saturation pressures are approximately 35barg and 39barg, respectively).

The decompression curve for an initial pressure of 180barg is less severe than that for an initial pressure of 100barg - the opposite of what was observed above.

In the case of a CO_2 pipeline, the toughness required to arrest a running ductile fracture can be estimated though consideration of the arrest pressure and the saturation pressure (as previously discussed). The saturation pressure follows from the isentropic decompression path, as indicated in Figure 6. The calculation of the saturation pressure is simpler than the calculation of the decompression curve, and similarly the calculation of the arrest pressure is simpler than the calculation of the fracture velocity curve[17].

Table 3 gives the required toughness for the two pipelines calculated in this manner (the TCM would give identical results). Comparing the results for a CO_2 and a CH_4 pipeline, the arrest toughness in the 24in pipeline is slightly higher (although not significantly) when the contents are CO_2, and higher in the 18in pipeline when the contents are CH_4. The differences in the arrest toughness follows from the decompression characteristics of the two fluids. The results for the 18in pipeline are, in fact, somewhat artificial because it is likely that there would be insufficient energy in the system to sustain a running fracture at the saturation pressure, because the hoop stress is low (the hoop stress at the saturation pressure is less than 20% SMYS). It is important to note that the arrest toughness for the 18in pipeline is not determined by the design pressure; in Table 3 it is assumed that the minimum operating pressure is 100barg, and it is this minimum pressure that determines the required toughness (see below).

For CH_4, the higher initial pressure results in a more severe decompression curve. Conversely, for CO_2 pipeline, it is the lower initial pressure.

From this observation, it follows that fracture propagation control is easier to achieve in a CO_2 pipeline that has a high design pressure. This is illustrated in the decompression paths given in Figure 6.

ENERGY

Table 3 - Required full size CVN impact energy for a CO_2 pipeline and a CH_4 pipeline

	24in 100 barg	18in 180 barg
Carbon Dioxide (CO_2)	53J	12J (9J)
Methane (CH_4)	50J	40J

Note: 1. The arrest toughness for the 18in diameter CO_2 pipeline is quoted for an initial pressure of 100barg and, in brackets, an initial pressure of 180barg.

The difference between the saturation pressures for the two initial pressures is small; it is less than 10% of the difference between the initial pressures. The design factor of the 18in and 24in pipelines is 0.72, i.e. the hoop stress is 72% SMYS. The driving force for a running fracture is directly related to the hoop stress. For an initial pressure of 180barg, the saturation pressure is approximately 0.2 times the initial pressure; the hoop stress is less than 20% SMYS. For an initial pressure of 100barg, the saturation pressure is approximately 0.4 times the initial pressure; the hoop stress is approximately 30% SMYS. Consequently, a lower arrest toughness is predicted for the 18in 180barg pipeline than the 24in 100barg pipeline (see Table 3).

The higher hoop stress at the saturation pressure in the case of the 24in. pipeline also has implications for the significance of the effect of impurities on the arrest pressure, see below.

Having established the arrest toughness for the 18in and 24in. diameter pipelines, the sensitivity of this toughness to changes in the wall thickness and the saturation pressure is investigated.

The arrest toughness increases and tends to infinity as the saturation pressure increases, see Figure 7. The limiting saturation pressure, above which mechanical crack arrestors would be required, depends on the diameter, wall thickness and grade. It is the flow stress dependent arrest pressure[17]. Increasing the initial temperature and/or decreasing the initial pressure would increase the saturation pressure. The addition of impurities would change the saturation pressure; some will increase it and others would decrease it[18].

The increase in the arrest toughness is greater for the 24in pipeline than the 18in pipeline because the saturation pressure at the assumed initial conditions is closer to the flow stress dependent arrest pressure.

The arrest toughness decreases as the wall thickness increases, see Figure 8. Increasing the wall thickness is equivalent to reducing the design factor (given that the diameter, grade and design pressure remain unchanged). The arrest toughness increases and tends to infinity as the wall thickness decreases. The limiting wall thickness, below which mechanical crack arrestors would be required, depends on the saturation pressure, diameter and grade (as above, it is related to the flow stress dependent condition). It may or may not correspond to a realistic design case, given that pipeline design codes place a limit on the design factor (e.g. 0.72).

The trends in the results in Figure 7 and Figure 8 show that fracture propagation control is achievable in both the 18in and 24in diameter pipelines, but that the 24in case is more sensitive to changes in the design conditions. Figure 5 illustrates why fracture propagation control is an issue in CO_2 pipelines, but also that it does not always follow that it is more of an issue than in CH_4 pipelines – it depends on the initial conditions and the composition.

Considering the 24in pipeline, the estimated saturation pressure for initial conditions of 100barg and 10°C is approximately 39barg, and the arrest toughness is approximately 53J. A 5bar increase in the saturation pressure increases the arrest toughness to approximately 89 J. A 10bar increase means that mechanical crack arrestors are required (irrespective of the toughness of the line pipe steel), see Figure 8. The design is constrained by the requirements for fracture propagation control, not the design factor. Increasing the wall thickness to 12.4mm (giving a design factor of 0.55) reduces the arrest toughness to approximately 53J. To illustrate the effect of impurities on the saturation pressure, for initial conditions of 100 barg and 10°C, the addition of 1% nitrogen (i.e. a mixture of 99% CO_2 and 1% N_2) would increase the saturation pressure by 5bar, and the addition of 2.5% would increase it by approximately 13bar.

The following general trends for a CO_2 pipeline can be identified:

- The lower is the initial pressure, the higher is the arrest toughness
- The higher is the initial temperature, the higher is the arrest toughness
- The smaller the diameter to wall thickness ratio, the lower is the arrest toughness
- The lower the design factor, the lower is the arrest toughness.

It also follows that there are some combinations of design parameters (diameter, wall thickness grade and design pressure) for which it is not possible to achieve fracture propagation control without the use of mechanical crack arrestors. Mechanical crack arrestors would be needed if the required toughness is too high (i.e. line pipe of the required toughness is not available or too expensive), or if it is simply impossible to arrest a fracture irrespective of the toughness. There is a second consideration when the required toughness is high. The existing models may overestimate the ductile fracture resistance implied by very high CVN impact energies. However, simple modifications to the design (e.g. increasing the wall thickness) will also solve the problem. Consequently, it is important to consider fracture propagation control early in design, before difficult to reverse decisions have been made.

The effect of the initial pressure and temperature on the arrest toughness in a carbon dioxide pipeline are the exact opposite of what is observed in a methane pipeline (and indeed in most, if not all, natural gas pipelines conveying lean or rich gas). This has one very significant implication. When developing a facture control plan for a natural gas pipeline, the limiting condition corresponds to the maximum operating pressure and the minimum operating temperature. The maximum operating pressure will be less than or equal to the design pressure, and it is conservative to use the design pressure.

ENERGY

In other words, the limiting condition is well defined. This is not the case for a carbon dioxide pipeline. When developing a facture control plan for a carbon dioxide pipeline, the limiting condition corresponds to the minimum operating pressure and the maximum operating temperature.

The minimum operating pressure may not be well defined. It is also a conceptual issue because in design it is normally the maximum pressure that defines the worst case.

Through the examples of the 18in 180barg and 24in 100barg pipelines it has been shown that fracture propagation control in a CO_2 pipeline can be addressed relatively simply. There are significant differences between CO_2 and CH_4 pipelines that mean that fracture propagation control is more of an issue in a CO_2 pipeline, but it does not always follow that the arrest toughness will be higher in the CO_2 pipeline than for an equivalent CH_4 pipeline.

The issue of fracture propagation in CO_2 pipelines tends to favour pipelines with a small diameter to wall thickness ratio and large wall thickness (the two are related), low grade, low design factor and a high design pressure, or some combination thereof. The availability of modern, high toughness line pipe steel reduces the significance of some of these trends.

The information required to assess the significance of fracture propagation control will be available at the conceptual stage of design (e.g. pipeline diameter, operating conditions and composition). The calculations are relatively straightforward.

Therefore, it will be simple to define a design envelope in which fracture propagation control can be achieved without the use of mechanical crack arrestors, or, conversely, identify this as an issue early in design. This would be useful in defining the specification for the CO_2, e.g. is it necessary to remove impurities (and if so, to what level) when the CO_2 is captured.

There are a number of underlying issues that have not been considered in detail here, e.g. the range of applicability of the underlying models, the effect of impurities that might be found in 'typical' captured CO_2, and experimental validation of the methods, but the principles have been demonstrated.

Conclusions

(1) Fracture propagation control is an issue for a CO_2 pipeline, but it is readily addressed using the methods that the pipeline industry has developed over the years. Some care is required because the trends observed in CO_2 pipelines are not the same as those in natural gas pipelines, and the required toughness to arrest a ductile fracture may be very sensitive to small changes in the design parameters (e.g. pipeline geometry or fluid composition). Nevertheless, provided that fracture control is considered early in design, any constraints on the design can be identified and, in principle, resolved without too much difficulty.

(2) Fracture propagation control in a CO_2 pipeline can be achieved through consideration of the arrest pressure and the saturation pressure. The required calculations are therefore much simpler than those required for the Two Curve Model.

(3) Impurities such as methane, nitrogen and hydrogen, will increase the saturation pressure and hence the toughness required to arrest a ductile fracture.

(4) The limiting case for fracture propagation control in a CO_2 pipeline is the lowest pressure and highest temperature within the operating envelope.

(5) In some situations, the requirement for fracture propagation control will dictate the design of a CO_2 pipeline.

References

1. BERNSTEIN,L. et al, Climate Change 2007: Synthesis Report Summary for Policymakers, An Assessment of the Intergovernmental Panel on Climate Change, Fourth Assessment Report, Intergovernmental Panel on Climate Change, November 2007.
2. BARKER,T. et al.; Summary for Policy Makers, Working Group III contribution to the Intergovernmental Panel on Climate Change Fourth Assessment Report Climate Change 2007: Mitigation of Climate Change, IPCC WG III 4AR, Intergovernmental Panel on Climate Change, May 2007.
3. STERN,N.; The Economics of Climate Change: The Stern Review, Cambridge University Press, Cambridge, 2007. (also HM Treasury, 2006 www.hm-treasury.gov.uk)
4. PHILIBERT,C.; Technology Penetration and Capital Stock Turnover: Lessons from IEA Scenario Analysis, International Energy Agency, Organisation for Economic Co-operation and Development, COM/ENV/EPOC/IEA/SLT(2007)4, May 2007.
5. HARRABIN,R.; Germany leads 'clean coal' pilot, BBC, 3 September 2008.
news.bbc.co.uk/1/hi/sci/tech/7584151.stm
www.vattenfall.com/www/co2_en/co2_en/index.jsp

6. DOCTOR,R., PALMER,A., COLEMAN,D., DAVISON,J., HENDRIKS,C., KAARSTAD,O., OZAKI,M. and AUSTELL,M.; Chapter 4 Transport of CO_2, IPCC Special Report on Carbon Dioxide Capture and Storage, R. Pichs-Madruga, S. Timashev, Eds., Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, 2005.
7. MOHITPOUR,M., GOLSHAN,H., MURRAY,A.; Pipeline Design and Construction, A Practical Approach, ASME Press, New York, 2000.
8. SEEVAM,P.N., RACE,J.M., DOWNIE,M.J. and HOPKINS, P.; Transporting the Next Generation of CO_2 for Carbon, Capture and Storage: The Impact of Impurities on Supercritical CO_2 Pipelines, Paper No.: IPC2008-64063, Proceedings of the 7th International Pipeline Conference, IPC 2008, Calgary, Alberta, Canada, September 30 - October 03, 2008.
9. ANON.; Specification for Line Pipe, Exploration and Production Department, API Specification 5L, American Petroleum Institute, Forty Second Edition, 2000.
10. EIBER,R.J., BUBENIK,T.A., MAXEY,W.A.; Fracture Control Technology for Natural Gas Pipelines, Final Report to Line Pipe Research Supervisory Committee of the Pipeline Research Committee of the American Gas Association, Project PR-3-9113, NG-18 Report No. 208, Battelle, December 1993.
11. EIBER,R.J., BUBENIK,T.A.; Fracture Control Plan Methodology, Paper 8, Eighth Symposium on Line Pipe Research, Pipeline Research Committee of the American Gas Association, Catalogue No. L51680, Houston, Texas, USA, September 1993.
12. RE.G., PISTONE,V., VOGT,G., DEMOFONTI,G., and JONES,D.G.; EPRG Recommendation for Crack Arrest Toughness for High Strength Line Pipe Steels, Paper 2, Proceedings of the 8th Symposium on Line Pipe Research, American Gas Association, Houston, Texas, 26-29 September 1993, pp. 2-1-2-13. (also 3R International, 34 Jahrgang, Heft 10-11/1995, p. 607-611)
13. KING,G.G.; Design of Carbon Dioxide Pipelines, Energy-Sources Technology Conference and Exhibition, Houston, Texas, USA, January 18-22, 1981.
14. MAXEY,W.A.; Long Shear Fractures in CO_2 Lines Controlled by Regulating Saturation, Arrest Pressures, Oil and Gas Journal, 1986, p. 44-46.
15. ROTHWELL,A.B.; Fracture Control in Natural Gas and CO_2 Pipelines, Conference on Microalloyed HSLA Steels, ASM International, 1988, p. 95-108.
16. COSHAM,A., EIBER,R.J.; Fracture Control in Carbon Dioxide Pipelines, Transmission of CO_2, H2, and biogas: exploring new uses for natural gas pipelines, Global Pipeline Monthly and Clarion Technical Conferences, Amsterdam, The Netherlands, May 2007.
17. COSHAM,A., EIBER,R.J.; Fracture Control in Carbon Dioxide Pipelines – The Effect of Impurities, Paper No.: IPC2008-64346, Proceedings of the 7th International Pipeline Conference, IPC 2008, Calgary, Alberta, Canada, September 30 - October 03, 2008.
18. FARRIS,C.B.; Unusual Design Factors for Supercritical CO_2 Pipelines, Energy Process, Vol. 3, No. 3, September 1983, p. 150-158.
19. WADSELEY,M.W., ROTHWELL,A.B.; Fracture Control for Pipelines Carrying Other Gases – HVPL, CO_2 and Others, Paper 9, Proceedings of the International Seminar on Fracture Control in Gas Pipelines, WTIA/APIA/CRC-MWJ Seminar, Sydney, Australia, 3 June 1997.
20. KOEIJER,G., BORCH,J.H. JAKOBSEN,J. and HAFNER,A.; Construction of a CO_2 Pipeline Test Rig for R&D and Operator Training, Transmission of CO_2, H2, and biogas: exploring new uses for natural gas pipelines, Global Pipeline Monthly and Clarion Technical Conferences, Amsterdam, The Netherlands, May 2007.
21. MARSILI,D.L., STEVICK,G.R.; Reducing the Risk of Ductile Fracture on the Canyon Reef Carriers CO_2 Pipeline, SPE20646, 65th Annual Technical Conference and Exhibition of the Society of Petroleum Engineers, New Orleans, USA, September 23-26, 1990.
22. BARRY,D.W.; Design of Cortez CO_2 System Detailed, Oil and Gas Journal, 1985, p. 96-102.
23. ANON.; Code of practice for pipelines - Part 1: Steel pipelines on land, PD 8010-1 : 2004, British Standards Institution, London, UK, 2004.
24. ANON.; Code of practice for pipelines - Part 2: Subsea pipelines, PD 8010-2 : 2004, British Standards Institution, London, UK, 2004.
25. ANON., Pipeline Transportation Systems for Liquid Hydrocarbons and Other Liquids, ASME Code For Pressure Piping, B31, ASME B31.4 – 2002 Edition (Revision of ASME B31.4 – 1998), American Society of Mechanical Engineers, New York, NY, USA, 2002.
26. ANON; Gas Transmission and Distribution Systems, ASME B31.8-2003, American Society of Mechanical Engineers, New York, NY, USA, March 2004.

Signalling at the St Pancras CTRL Terminal

Mark O'Neill

Senior Signalling Design Manager

Rail

Atkins

Abstract

The new St Pancras CTRL (Channel Tunnel Rail Link) station consists of nine platforms, six for international Eurostar trains and three for regional domestic services from Kent (there are four platforms for the Midland Main Line but there is no through running between the CTRL and Network Rail parts of the station). Trains approaching St Pancras leave the single bore London Tunnels approximately one kilometre from the platform ends. In this last kilometre there are connections to the nearby Network Rail main lines (Midland Main Line, North London Line (NLL) /West Coast Main Line & East Coast Main Line (ECML)) and three approaches into the station (UP, DOWN & RELIEF). All lines are fully bi-directional.

On leaving the London Tunnels, the cab signalling system TVM 430 switches over to conventional UK lineside colour light signalling with the KVB intermittent ATP system. Colour light signalling is necessary in this last kilometre due to the Network Rail interfaces, the need for routing information and because short block sections are required to meet the specified operational headway/journey time (which cannot be achieved within the low speed codes available to the TVM system due to the speed profile of the layout).

Dependability approach

Eurostar has to compete against the low cost airlines. To do that, the passenger perception should not be that of a typical train service, with late, crowded trains, frequent failures and constant excuses. For the service to succeed, we must destroy this perception. Where we as engineers can help is in the reliability of the infrastructure and in particular the signalling. To achieve the challenging RAMS targets (Reliability, Availability, Maintainability and Safety), the integrated Rail Link Engineering (RLE) team focused on three areas,

(1) Designing the station layout and associated infrastructure to minimise the impact of failures on the operation of the train service

(2) Using proven equipment and system configurations and devising methods to mitigate against their known failure modes

(3) Looking at the maintenance response to those failures and designing maintainability, fault tolerance and failure response into the system.

These three areas will now be examined in more detail, but will focus on the signalling, as this paper is written from the signal engineer's perspective.

The station layout

Dependability is proportional to simplicity, therefore a simple layout that met the specified routing criteria was an essential first step. This was difficult within the constrained space available, but three main criteria were adopted, grade separation of all junctions, standard turnouts (i.e. no fixed crossings, single/ double slips or specially designed turnouts), and three independent roads into the station throat.

The infrastructure relating to each of the three roads was also separated (i.e. Overhead Line Equipment (OLE) supply sectioning, cables, cable routes, interlockings and trackside signalling equipment), so that a failure in one area, or even two areas, did not close down the station and still allowed signalled moves to continue.

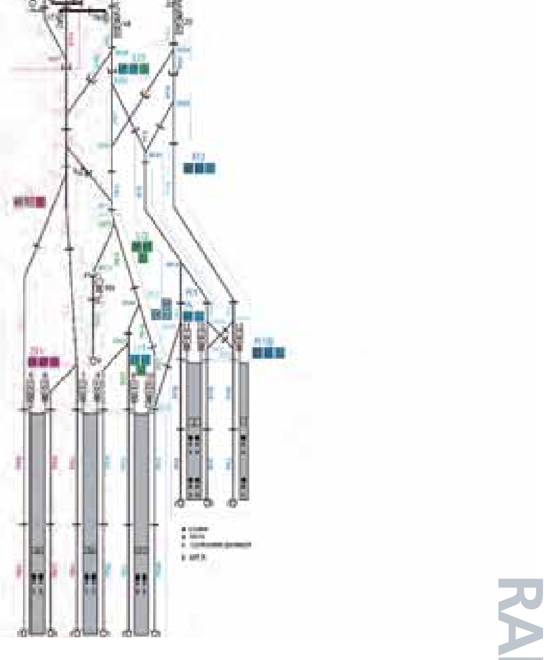


Figure 1 - Schematic signalling layout sketch

RAIL

Figure 2 - Aerial view of St Pancras station

The signalling equipment and configuration

UK colour light signalling has been around for some time and the problems and failure characteristics are well known.

To overcome these common faults and increase the reliability of the signalling at St Pancras, the following enhancements were provided:

System architecture

A number of interlocking options and system configurations were considered for St Pancras, including the Interlocking & Train Control System (ITCS) with the latest trackside modules developed by the CSEE Company for the Spanish ERTMS High Speed Line project (similar to SSI architecture). However, the simplest and cheapest option was using the existing ITCS with a simple relay interface to drive trackside equipment, as this had already been proven on CTRL Section 1, including the control of the small number of colour light signals on the Ashford Chords and Waterloo Connection. Using long tail cables simplified the relay interface circuits for the signals and meant the majority of equipment could be driven directly from the Signalling Equipment Room (SER). However, as UK point machines were used, they needed a nearby location case due to the volt drop problem. The High Voltage Impulse (HVI) track circuits used also needed a location case, as the maximum feeding distance was 500m.

Location cases

To simplify the design, 4 types of location case were identified;

- Points: A single case feeding four point machines
- Track circuit (no power): A single case feeding 4 HVI tracks with power from adjacent point or 2 track circuit case
- Track circuit (with power): A single case feeding 2 HVI tracks with buffer stop & interrupter feeds
- Disconnection: A single case providing a termination point for main cables before feeding individual Engineering Zone of Protection (EZP), Local Release Switch (LR) and Manual Operation of points Devices (MOD) units.

All of the location cases used in St Pancras was from one of these basic designs, with only the function name and sometimes cable allocation changing for each case type.

Points

The latest High Performance Switch System (HPSS) point machine has been used at St Pancras, which has a number of in-built reliability features, including condition monitoring. Permanent-way faults can cause detection failures, so integration between the point machine and the permanent-way layout has been the main goal of RLE permanent-way and signal engineers. Consequently, design meetings between RLE, the permanent-way and the point machine manufacturer have taken place to ensure an optimum design.

Track circuits

The more reliable and simple single rail HVI track circuit has been used in St Pancras, with duplicated tail cables and protected routes for each track circuit.

Signals

Fibre optic signal technology has been chosen for St Pancras. Auxiliary bulbs are provided for all signals (including route indicators, banners and subsidiary signals) due to the double bulb arrangement built into the new fibre optic units. A changeover EKR relay switches in the auxiliary bulb if the primary one fails. A more reliable halogen bulb has been selected but the ability to change over to the LED bulbs being developed is something RLE have been looking at. All signal units are also fitted with a Lamp Proving Relay (ECR) relay, even where there is no requirement to prove a route indication or subsidiary.

Figure 3 - St Pancras Station extension and throat (under construction)

Figure 4 - Engineering Zone of Protection (EZP) Module

All ECR and EKR inputs are input to the Local Maintenance Equipment system. In addition, the signal will step down to a lower aspect when the back-up bulb also fails (i.e. G -> YY, YY -> Y, Y -> R). The signal in rear will only go to red when the back-up red lamp fails. As the bulbs are located at the bottom of the signal in a dedicated light box, maintenance access is simplified.

A modular design approach was also used for the design of the fibre optic light box, with each box being able to hold up to four modules. Three module types were developed:

- Lamp Unit: A unit capable of feeding four indications (either signal aspects, route indications or position light)
- AWS unit: A unit capable of feeding an AWS electro and suppressor magnet
- Termination Unit: A unit capable of providing through terminations for splitting of cables.

Each fibre optic signal consisted of a mixture of these units depending on the number of indications shown, the AWS requirements and the cable arrangement.

Cables

SNCF standard shielded and armoured multicore cables and armoured power cables have been used in the St Pancras area. The armouring on the cable also gives additional protection (particularly with the large number of rodents from the nearby canal). Surprisingly, the standard SNCF cables were an almost identical cost to UK multicores and the short length of the St Pancras area meant that cabling costs were not a significant part of the equipment budget.

With standard location cases and a modular arrangement in the fibre optic signals, a major part of the design effort was in the cable and cable core allocation. Standard allocation rules were used that enabled a computer macro to be developed to produce the cable core allocations.

Power supply

The power to the signalling equipment room will use two independent and separate power supply feeders, backed up by a N+1 Uninterruptible Power Supply (UPS). Each location area has a direct power feed rather than a long feeder arrangement typical in the UK. The worst case location area was used for calculating the size of the power cable, and this same size cable was then used throughout. This simplified procurement and fitted with the standard location case design philosophy.

Though this increased the amount of power cable required, there was a reduction in design costs as well as mitigating the impact of any power fault.

To reduce the size and cost of the UPS battery back-up the point movements will be staggered by 0.5 seconds when the UPS is being used. In addition, a separate transformer/ rectifier has been used for each point end (motor control and detection), to avoid common earth faults.

Local Release of signalling controls

A CTRL principle introduced from the high speed line is the Local Release (LR) switch. This is a switch located at the trackside near to the points. Each LR switch covers the track circuits over the points, and can be used to release the route locking when a track circuit failure occurs. This is because any track circuit failure will dead lock the points or maintain the route locking over them.

Therefore, the technician operates the LR switch in co-operation with the signaller. This releases the sub-route locking and allows the points to be manually moved. An auxiliary route can then be set to authorise the train to pass the signal.

Manual Operation of points Device

Manual Operation of points Devices (MOD) are basically point key units that are provided at each point location. Their function is to ensure that the train crew/technician can only move the correct point end specified by the signaller.

After withdrawing the point key, which cuts the point detection, giving added protection, the train crew/ technician uses the key to operate the points motor manually. After the points have been moved and the key replaced, movements can then take place over the point end concerned. As each key is allocated to a particular point end, the risk of miscommunication and the wrong point end being moved is reduced.

The local release, point key and auxiliary signal are all principles that have been introduced to CTRL to keep the traffic moving during signalling failures, as releases and manual operation of points can be undertaken by train crew as well as technicians. An example of these units is shown in Figure 4.

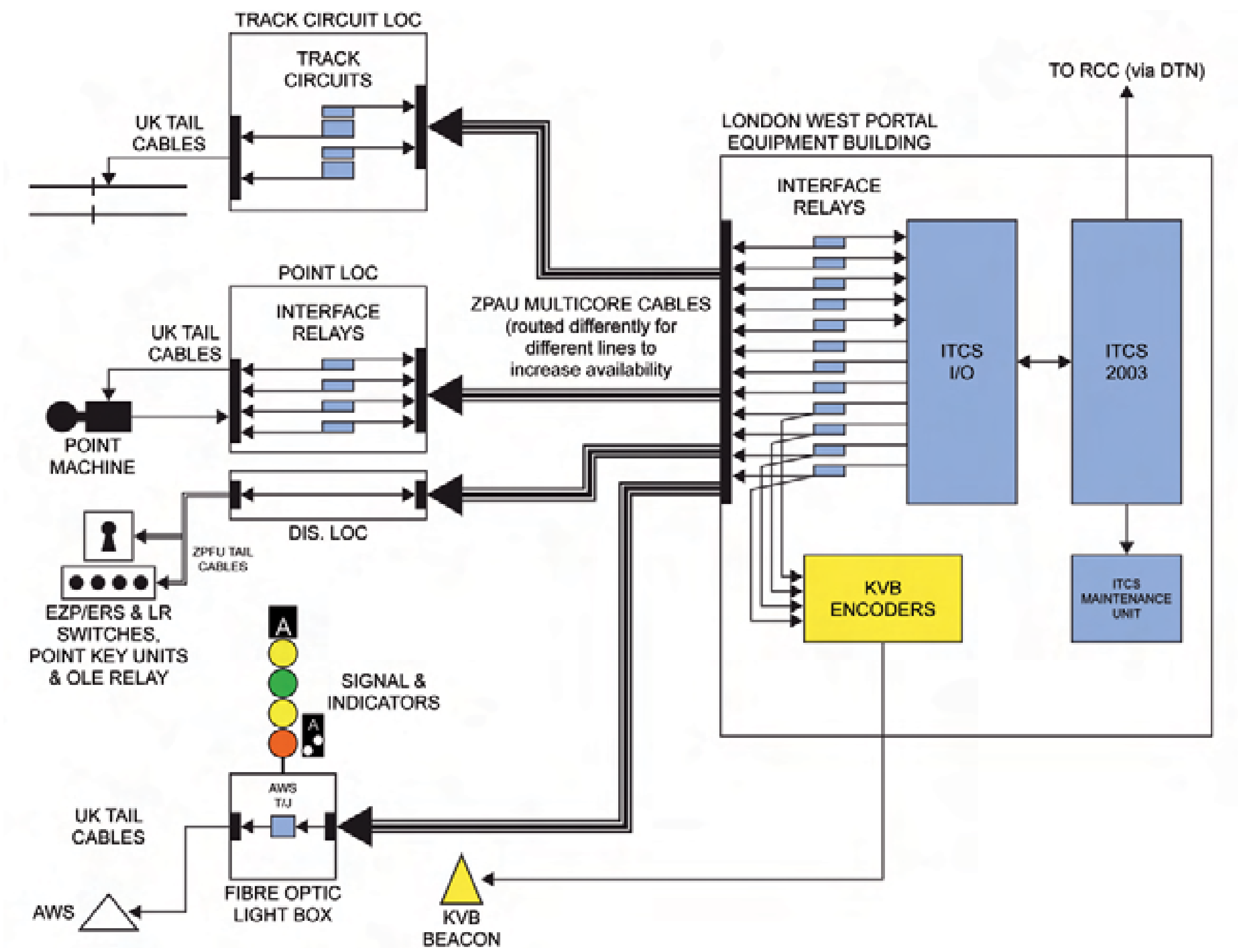


Figure 5 - ITCS Interlocking Architecture

Maintenance response

The ITCS interlocking used at St Pancras has a centralised architecture, with trackside equipment being fed via a relay interface see Figure 3. The ITCS has a dedicated maintenance terminal, which provides specific technicians controls, monitors the status of the interlocking, pinpoints specific faults, stores data logging information, etc.

Safe access and walkways have been provided to all parts of the layout, with strategically placed EZP possession switches to ensure safe access to the track if work is required during train operations. This is another principle adopted from the French high speed line. However, the EZP switches at St Pancras can also be used by staff to replace signals to danger in an emergency.

The Fibre Optic signal has been specially designed for RLE so that staff access to the signal structure is not needed.

The bulbs and equipment are housed in a small location cupboard that is integral to the signal base, and the signal tilts over to allow cleaning. However, St Pancras is closed at night, and any maintenance is planned to occur during this period (via a dedicated possession system).

A maintenance siding and strategically placed storage areas have also been provided so that signalling, OLE, permanent-way spares and point components can be stored and quickly changed when necessary.

Project management philosophy

By having a number of engineers and operations experts within the team RLE took on the role of an informed buyer as well project manager. This allowed the RLE team to develop an integrated design that met the requirements of the project client brief.

This base design could then be discussed with the chosen contractor to evaluate improvements and obtain contractor buy-in. This allowed a firm scope to be produced, which is essential for successful contract management and keeping control of costs.

The RLE set up also had the main discipline engineers and management personnel in close proximity (i.e. permanant-way, OLE, civils, operations, architects, contract management, etc), which assisted in cross discipline co-ordination, particularly when dealing with the evaluation of design changes. Also having the client located in the same building was a further advantage.

From the author's own perspective of working on other large projects, four RLE practices made a personal impression and had a major impact on the way the project was managed. These are as follows:

1. Operations and safety team

This team was lead by SNCF operators, who had a significant input into the scheme development process. This followed French practice where the scheme plan is jointly developed alongside the operations team. Various signal positions, overlaps and routing options were considered and modelled by the operations team, including the TVM interface, in order to obtain a scheme that satisfied the train service requirements (i.e. headways, journey times, train performance, timetable options, degraded mode situations, alternative routes, etc) and was safe to operate (i.e. overlaps to match the ATP system, train stopping locations where passengers could evacuate, emergency replacement facilities, etc).

This team also reviewed the OLE sectioning and permanant-way layouts to ensure train operations and safety were being considered in these disciplines. The author has come across similar teams in UK projects (i.e. like the previous Railtrack Safety and Standards managers) but none that have had such a direct project role, being involved in the design throughout the development phase and ensuring that the engineering will be fit for the train service being planned.

2. Systems integration schematics

The systems integration team produced a number of schematics that helped identify both the interfaces between systems and between contractors (i.e. telecoms, signalling, SCADA, power supply, etc). This allowed the interfaces to be logged in a database and systems interface specifications to be developed where needed. This also ensured that the relevant contractors were aware of who was responsible for what at each of the interfaces identified. Again, the author has seen similar schematics used before, but not as an integral component of identifying, allocating and tracking interface issues.

3. Virtual Reality modelling

A VR model was specifically developed by RLE to undertake preliminary signal sighting. This model also helped the positioning of trackside equipment, which proved essential in the complex and constricted environment at St Pancras.

4. Site construction packages

The St Pancras layout had to be ‘shoe horned’ into a very constrained site in a central London location, with much of the available spare land required for development. Consequently, the space available for equipment near the track was severely restricted, with many conflicts between the competing engineering disciplines (i.e. drainage catchpits, undertrack crossings, location areas, signal and OLE mast bases, walking routes, cable routes, maintenance storage areas and various trackside equipment). In order to avoid these conflicts and co-ordinate the site construction, RLE developed General Arrangement drawings to precisely locate all trackside equipment.

For each piece of trackside construction required, RLE then produced a site construction package, showing General Arrangement drawing extracts, cross section drawings and siting forms to ensure site construction was closely co-ordinated with the design. This is an area where the author has seen mistakes in the past, with site construction not being co-ordinated between disciplines or contractors. However, this RLE practice made sure the contractor was aware of what had to be constructed and where, and allowed close monitoring of the work. One innovation that came from this process was the development of a combined 700mm cable trough/walking route. It was recognised that the main cable routes followed the main walkways, and since most people walk on the cable route anyway, it was decided to combine the two and thus achieve savings in installation costs that were greater than the development cost of the new troughing route.

Conclusion

Of all the CTRL areas, St Pancras was one of the most involved and challenging to signal, with a complicated layout, interfaces to three Network Rail lines, a busy service and demanding RAMs targets. RLE have therefore taken the opportunity to build dependability into the system from the start. The approach taken has focused on the end product, the train service and how to keep the trains running when stopped by a fault and then keeping them running while the fault is rectified. This has been achieved by using a mixture of,

- Proven equipment and configurations from the UK and France
- Recently introduced equipment designed to overcome known deficiencies
- Introducing new signalling principles based on proven degraded mode practices.

In addition, the St Pancras area was challenging to manage due to the number and variety of contracts and contractors involved. The RLE team set up a number of processes to co-ordinate and supervise these contracts, with a particular focus on scope definition and system integration.

ATKINS

Printed on 9Lives Offset, part of the family of 'born again' papers, manufactured with National Association of Paper Merchants Certification (NAPM) using approved 100% recycled fibre pulp, 9lives Offset offers the ultimate in environmental management. The manufacturing mill and printer have also been accredited with the internationally recognised standard ISO 14001.