Transport for Greater Manchester

Modelling Pedestrian Stages Better:

A Probabilistic Approach to the Determination of Demand Dependent Pedestrian Stage Frequencies

Abstract

New all-red pedestrian stages are usually modelled in LinSig based on an assumption – such as appearing every cycle or every other cycle. A more sophisticated approach is required to provide more accurate modelling. This paper outlines how a Poisson distribution can be used to predict the appearance frequency of demand dependent pedestrian stages, using pedestrian count data. It describes how this approach can be combined with the use of bonus greens in LinSig to model demand dependent pedestrian stages appearing at the predicted rate. This methodology allows for a more refined approach when compared with traditional modelling assumptions. The paper includes evidence showing the difference between predicted and observed frequencies. Various sites across Greater Manchester have been part of this study and the data from these sites is used to support the above. The outcome is an approach to modelling that can be applied by all users of LinSig.

Introduction

Junctions are usually modelled in order to assess their efficiency and determine how any proposed changes/upgrades would impact this efficiency. The results from the model often form part of the basis on which schemes are either approved or rejected. Accordingly, it is critical that the results from a model reflect reality as accurately as possible.

The approach adopted when modelling demand dependent stages can have a major impact on the results of the model. If it is simply assumed that all demand dependent stages occur in every cycle this can lead to significant efficiency reductions at a junction, potentially leading to the rejection of a scheme that would have worked in reality. On the other hand, if the appearance frequency of demand dependent stages is significantly underestimated then this can lead to results which suggest a junction operates more efficiently that it would in reality.

Where a record of the appearance frequency of a stage exists, such as from a UTC system, it is possible to use bonus greens in LinSig to simulate this recorded appearance frequency (bonus greens will be touched upon again later in this paper). However, in many scenarios a recorded appearance frequency is not available, either because no record exists or because the modelled junction is a proposal and hence does not currently exist.

There is a push to upgrade existing signalised junctions to add signalised pedestrian crossings. In these cases, a new all-red stage will often be added to the junction to enable the pedestrian crossings to run. This is a scenario in which there is no record of the appearance frequency of the new pedestrian stage. Typically, in these instances, that pedestrian stage will be modelled as either appearing in every cycle or in every other cycle. However, this is a very crude approach, which can easily fall foul to the issue of over/under-estimating junction efficiency. This paper will explore a way of using pedestrian count data to estimate an appearance frequency for the pedestrian stage.

The Theory

A Poisson distribution can be used to estimate the appearance frequency of the pedestrian stage. The Poisson distribution is a probability function which predicts the likelihood of a given number of events occurring in a fixed interval of time, assuming that the events occur at a constant rate and independently of each other. In the case of pedestrian stages, the event is represented by a pedestrian arriving at the crossing, and the period of time is the cycle time minus the time for which the pedestrian stage is running.

The Poisson distribution relies on several assumptions:

1. It assumes that all of the pedestrians arrive alone (i.e. there are no pedestrians choosing to walk together).
2. It assumes that all pedestrians will place a demand (i.e. no pedestrians will simply cross on red without pushing the button).
3. It assumes that the arrival of the next pedestrian is not related to the arrival of the current pedestrian (i.e. there are no reasons why there would be sudden groups of pedestrians unintentionally walking together).

Naturally, these three assumptions do not hold for pedestrians at crossings in reality. However, in all three cases, using these assumptions would lead to a conservative, and hence desirable outcome, as outlined below:

1. If some pedestrians were to walk with other pedestrians, this means that they could effectively be treated as one pedestrian only, and hence the number of events in reality would be lower than the predicted number of events. Accordingly, this is a conservative approach.
2. If some pedestrians do not place a demand, then they can effectively be ignored. So, again, this means the number of events in reality would be lower than the predicted number of events, and as such this is a conservative approach. Note, if pedestrians arrive, place a demand, and then cross on red, this still counts as an event, because the pedestrian stage would still appear (assuming no kerbside detectors).
3. If there were a reason why pedestrians might suddenly arrive in groups (say for instance the site is near a train station so whenever a train arrives there is a big group of pedestrians) this again effectively means that the number of events in reality would be lower and as such this is a conservative approach.

The Poisson distribution is defined as follows:

$$P_k = \frac{\lambda^k e^{-\lambda}}{k!} \tag{1}$$

Where P_k is the probability of k pedestrians arriving in a given time period, and λ is the average number of pedestrians expected to arrive in the same time period.

When trying to predict the appearance of a pedestrian stage we are interested in the probability of at least one pedestrian appearing over the course of one cycle (minus any pedestrians that arrive while the pedestrian stage is on green or during the preceding interstage). The probability of at least one pedestrian arriving is the same as 1 minus the probability of no pedestrians arriving.

$$P_{at\ least\ one\ pedestrian} = 1 - P_0 \tag{2}$$

Hence, the probability of at least one pedestrian arriving is:

$$P_{at\ least\ one\ pedestrian} = 1 - \frac{\lambda^0 e^{-\lambda}}{0!} = 1 - \frac{1 \times e^{-\lambda}}{1} = 1 - e^{-\lambda} \tag{3}$$

Accordingly, there are three pieces of information necessary to predict the appearance of a pedestrian stage. Firstly, pedestrian count data giving us the average number of pedestrians expected to appear over a period of time. Secondly, the cycle time at which the junction will be running. And thirdly, the length of the pedestrian stage (this needs to include the preceding intergreen and the green time). This allows us to express λ as:

$$\lambda = \frac{d(T - t)}{3600} \tag{4}$$

Where d is the number of pedestrians arriving in an hour, T is the cycle time in seconds, and t is the length of the pedestrian stage in seconds (including the preceding intergreen).

The final step in completing the calculation is to realise that this process is iterative. If the calculation is performed once and the appearance frequency of the pedestrian stage is less than 100% then the length of the pedestrian stage needs to be scaled by the same percentage. The equation converges quickly and hence should only take three iterations before converging on an answer. Combining all of this produces the final equation:

$$P = f(P) = 1 - e^{-d(T - tP)/3600} \tag{5}$$

Where P is the probability of the pedestrian stage appearing.

For example, assume a 3-stage site operates a 60 second cycle time, with a 16 second pedestrian stage (including the preceding intergreen), and two vehicle stages, each with a preceding intergreen of 6 seconds and a green time of 16 seconds. Also assume a peak pedestrian count of 100 pedestrians per hour. Applying Equation 5, the probability of the pedestrian stage appearing in any given cycle is:

$P_0 = 1 - e^{-100(60-16)/3600} = 0.71$

$P_1 = 1 - e^{-100(60-16\times0.71)/3600} = 0.74$

$P_2 = 1 - e^{-100(60-16\times0.74)/3600} = 0.74$

As demonstrated the equation converges within 3 iterations and the pedestrian stage is expected to appear in 74% of the cycles.

Traditionally, in a scenario like the one above, for modelling purposes it would either have been assumed that the pedestrian stage appears in every cycle, leading to a significant underestimation of junction efficiency, or that the pedestrian stage appears in every other cycle leading to a significant overestimation. The next table demonstrates how much of an impact these assumptions have on junction efficiency in this example. For the purpose of calculating the figures in the table below, it has been assumed that each PCU takes 2 seconds to cross the stop line.

Table 1: Comparison of pedestrian stage appearance frequencies and the impact on junction efficiency.

Modelled Pedestrian Stage Appearance Frequency	Every other cycle	74% of cycles	Every cycle
Total vehicle green time over one hour (minutes)	40	36	32
Estimate of additional/lost capacity (PCUs)	**120**	0	**-120**
Estimate of additional/lost capacity (%)	**10%**	0	**-13%**

Comparison of the Theory with Real-Life

The results from five sites across Greater Manchester are presented in this paper. The sites were chosen based on several factors:

- Recent pedestrian survey data.
- A demand dependent all-red pedestrian stage.
- Stage appearance frequency data for the same time period as the pedestrian survey data.

In addition to the above, the sites included in this paper were chosen as they did not have very high pedestrian demands (leading to a scenario in which the pedestrian stage appears in every cycle), nor did they have very low pedestrian demands (leading to a scenario in which the pedestrian stage hardly ever appears. The next table summarises the five sites.

Table 2: Summary of the sites presented in this paper.

Reference Number	Address	Location	Survey Date
0079	Oldham Rd / Collyhurst St / Varley St, Manchester	Residential area on the outskirts of Manchester City	31.01.2024
0383	Droylsden Rd / Kershaw Ln, Droylsden, Tameside	Residential area on the outskirts of Ashton-under-Lyne	30.01.2024
0628	Oldham Rd / Mersey Rd North, Hollinwood, Oldham	Residential/commercial area just inside the M60	31.01.2024
1121	Rochdale Rd / Livesey St / Bromley St, Collyhurst, Manchester	Edge of a residential/industrial area just outside Manchester's inner ring road	07.02.2024
1157	Rochdale Rd / Lathbury Rd, Harpurhey, Manchester	Residential area just outside a park	07.02.2024

Refer to Appendix A for site drawings of the 5 sites.

For each site, a graph has been plotted showing the actual appearance frequency of the pedestrian stage as recorded by the UTC-UX system, versus the appearance frequency as predicted using Equation 5. Each data point represents a full hour, centred around the plotted time.

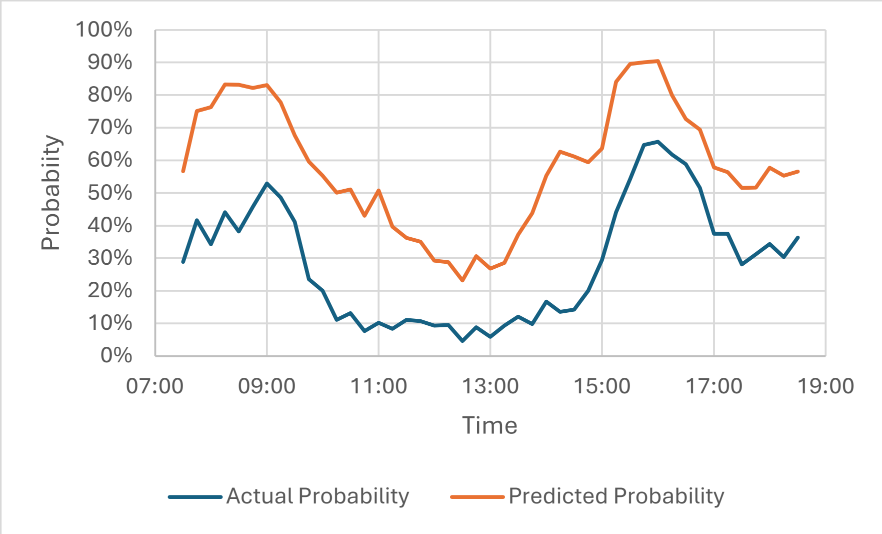

Figure 1: Probabilities for site 0079 on the 31.01.2024.

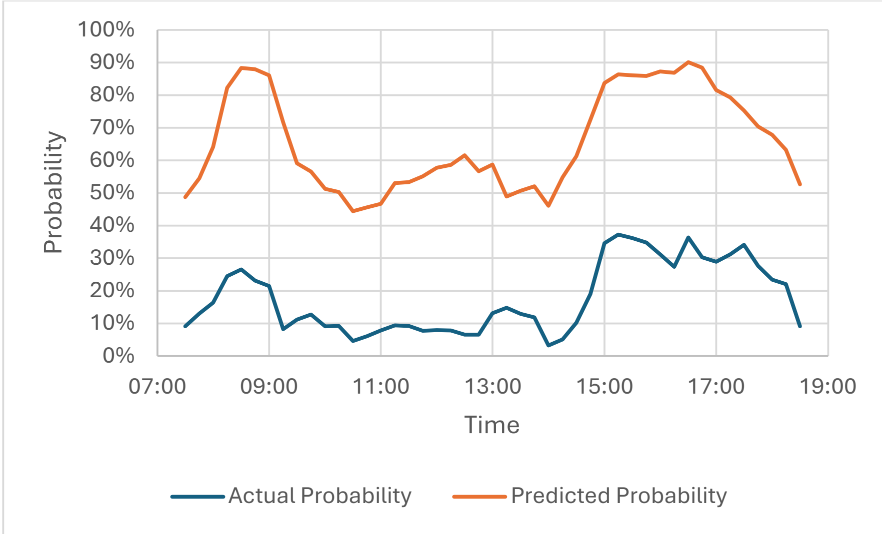

Figure 2: Probabilities for site 0383 on the 30.01.2024.

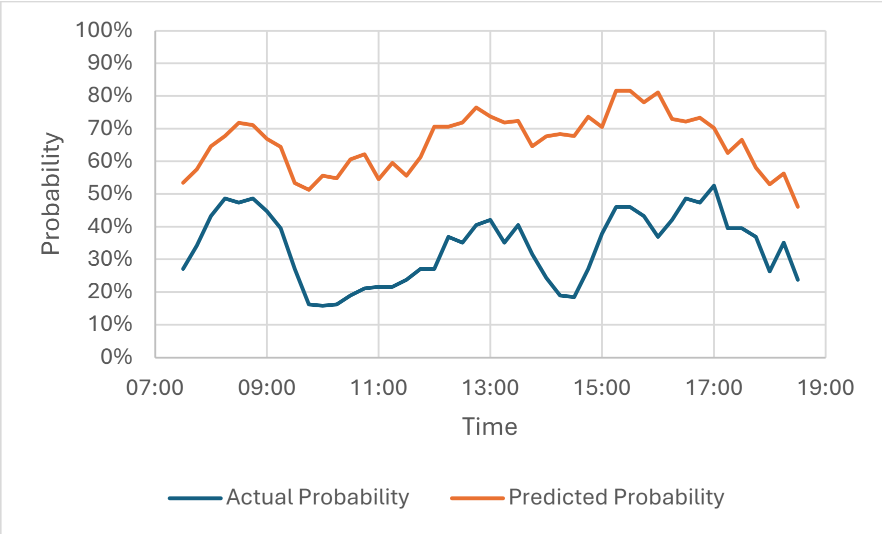

Figure 3: Probabilities for site 0628 on the 31.01.2024.

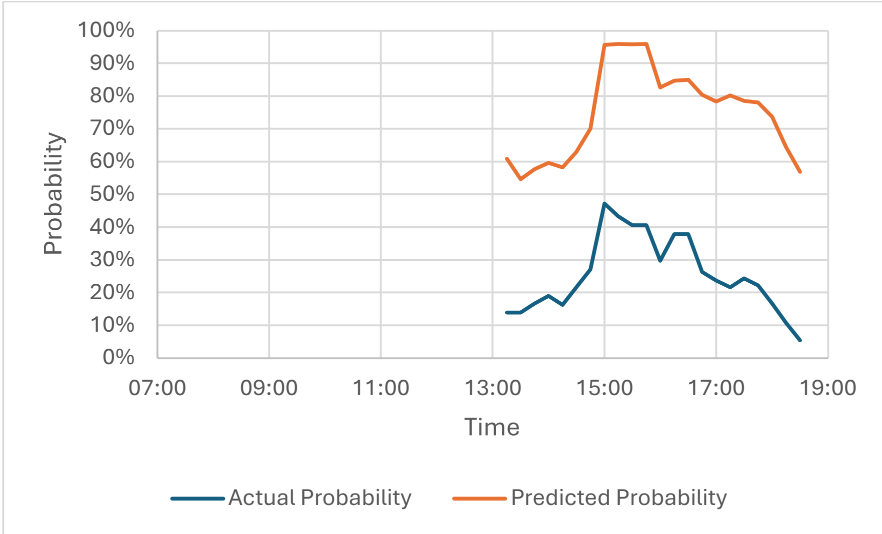

Figure 4: Probabilities for site 1121 on the 07.02.2024 (note, there was no UTC-UX data available until 13:00).

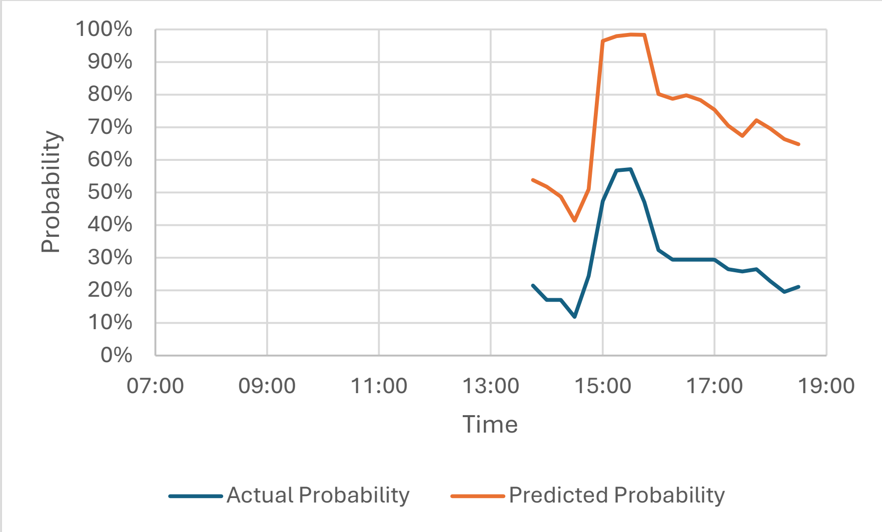

Figure 5: Probabilities for site 1157 on the 07.02.2024 (note, there was no UTC-UX data available until 13:15).

Several observations can be made based on these graphs. The first of these is that the shape of the predicted profiles matches with the actual profiles very well. This demonstrates that (as expected) the appearance frequency of the pedestrian stage does indeed depend primarily on the pedestrian count data.

The second observation is that in all cases the predicted probability is higher than the actual probability. This is a positive result as it proves that the approach is conservative, as discussed earlier in this paper.

However, it also becomes apparent that the approach often significantly overestimates the appearance frequency. This is caused by the assumptions of a Poisson distribution not being met, as discussed earlier in this paper. It shows that a significant number of people arrive in groups, and it shows that a significant number of people will cross on red without placing a demand. To improve upon the predicted probability, it would be necessary to account for these two factors.

Both of these factors are very variable, as shown by the fact that the level of overestimation is variable in the graphs. What percentage of people arrive in groups, and groups of what size, depends largely on the location. Sites near a school or sizeable public transport facilities are likely to have significant grouping whereas residential areas without schools or sizeable transport facilities are likely to have less grouping.

The percentage of people crossing on red without placing a demand is even more variable and affected by a whole range of factors. This includes:

- How heavily trafficked the road is.
- What the weather is like (people are more likely to cross on red if they are cold/wet).
- How many lanes they need to cross.
- How rushed they are.
- How long pedestrians have to wait before receiving a green man.

It may be possible to account for the factors mentioned above in one of two ways. A direct approach can be applied which attempts to predict the impact of the two factors separately. Or an indirect, combined approach could be applied which simply scales the predicted appearance frequency by a set value to improve the prediction.

In order to apply the first approach, and account for these two factors separately a significantly larger study would be required to calculate the percentage reduction in pedestrian count required due to each factor at various sites. An attempt could then be made to determine a set of scaling values to apply to sites based on site characteristics and location. This scaling value could then be used to adjust the pedestrian count data. However, producing such a value which still leads to conservative results in the majority of cases will prove very difficult.

The second approach has been attempted on the sites considered in this paper. A scaling value has been calculated by comparing the predicted probability with the actual probability. The table below shows the average scaling value required at each site, as well as the range of scaling values.

Table 3: Summary of scaling values.

Reference Number	Percentage Reduction in Predicted Probability Required		
	Average	**Minimum**	**Maximum**
0079	55%	19%	82%
0383	75%	55%	93%
0628	49%	25%	73%
1121	68%	51%	90%
1157	60%	42%	71%

The table shows that there is a large degree of variance in the results, and also that there are discrepancies between the different sites, indicating that any attempt at computing such a scaling value would also require a larger study.

Having pointed out that this approach can lead to significant overestimation of the appearance frequency, it is worth noting that this approach is still a beneficial technique for two main reasons. Firstly, it is based upon collected data, and it outputs an appearance frequency which can be justified, instead of simply modelling it as every cycle and/or every other cycle. Secondly, in many cases it will provide benefits when compared to simply modelling it in every cycle, and as such it has the potential to aid pedestrian schemes in gaining approval.

Four of the five junctions included in this paper has been modelled in LinSig. Traffic surveys were undertaken at the same time as the pedestrian count surveys used in this paper. The LinSig models cover both the busiest AM and PM peak as well as an interpeak from 13:00-14:00. The results presented in the table below show the PRC values output by the LinSig model for each of the junctions when the pedestrian stage is modelled as appearing in every cycle, every other cycle and using the approach outlined in this paper (note that a model of site 0628 has not been included in this paper as it is closely associated with several other junctions and hence adds complexity which is unnecessary for this paper). For more details on how to apply the approach outlined in this paper to LinSig models refer to the next section.

Table 4: PRC results from LinSig models.

Reference Number	Peak	Predicted Pedestrian Stage Appearance Frequency	Practical Reserve Capacity		
			Every other cycle	Using predicted value	Every cycle
0079	08:00-09:00	83%	38.7	23.8	18.3
	13:00-14:00	37%	59.4	63.1	29.7
	17:15-18:15	52%	-4.5	-7.1	-25.3
0383	07:45-08:45	82%	40.8	30.2	26.6
	13:00-14:00	51%	114.8	110.2	93.1
	17:00-18:00	75%	51.3	42.2	35.9
1121	13:00-14:00	55%	89.6	87.6	63.2
	16:45-17:45	80%	7.3	-1.3	-7.6
1157	17:30-18:30	70%	29.3	19.6	3.4

The table demonstrates that, there is a significant difference in efficiency between modelling the pedestrian stage as appearing in every cycle or every other cycle. Without any data to determine which is the more appropriate it is hard to justify the every other cycle approach and as such the every cycle approach often dominates the results.

The table also shows that by using the approach recommended in this paper it is possible to achieve significant efficiency improvement when compared to the every cycle approach. This improvement can be justified as the approach in this paper has been shown to be conservative. The efficiency improvements offered by this paper can therefore be relied upon when undertaking scheme approval processes.

It is however worth noting that the approach presented in this paper does not allow for any additional demand drawn to the proposed scheme upon completion of the scheme. If additional demand needs to be considered, then the pedestrian survey data should be factored accordingly and from there on the approach is identical.

Applying the Theory to Proposed Schemes

In order to apply this approach to the modelling for proposed schemes the first step is to use Equation 5. Once a predicted appearance frequency has been calculated this can be applied to a LinSig model using bonus greens. This paper will not go into detail of how to use bonus greens, but a summary is provided here.

First, it is necessary to calculate the amount of lost time to traffic in a normal cycle where the pedestrian stage appears. It is then necessary to calculate the amount of lost time to traffic in a cycle in which the pedestrian stage does not appear. By subtracting this from the first value the amount of additional lost time in a cycle due to the appearance of the pedestrian stage can be calculated. This additional lost time can then be multiplied by the appearance frequency of the pedestrian stage to give an adjusted lost time which is representative of the appearance frequency. Bonus greens can be calculated by subtracting this adjusted lost time from the total additional lost time. The pedestrian stage should then be modelled as appearing in every cycle but with bonus greens applied to the relevant traffic lanes.

The approach above will lead to efficiency improvements at the junction when compared to simply modelling the pedestrian stage appearing in every cycle without the use of bonus greens.

Conclusions

This paper has demonstrated a conservative, probabilistic approach that can be applied to pedestrian survey data to enable the appearance frequency of a demand dependent pedestrian stage to be predicted. This allows for pedestrian stages to be modelled based on surveyed data instead of simply assuming an appearance frequency. This approach demonstrates improved efficiency when compared to simply assuming the pedestrian stage appears in every cycle. The aim is that this will make it easier to obtain approval for proposed pedestrian improvement schemes.

It has been noted that there are discrepancies between the predicted and actual appearance frequencies due to pedestrians arriving in groups and pedestrians crossing without placing a demand. These discrepancies always result in a conservative output which avoids the risk of overestimating junction efficiency; however, it can lead to underestimating efficiency.

It has also been noted that no allowance for additional demand drawn by schemes has been made in this paper.

Next Steps

- Quantify the number of pedestrians arriving in groups in order to determine average group size figures which can be used to adjust the pedestrian count data.
- Quantify the number of pedestrians crossing on red without placing a demand in order to determine a percentage crossing on red value which can be used to adjust the pedestrian count data.
- Quantify additional demand drawn due to schemes in order to apply an uplift to recorded pedestrian count data.

Appendix A

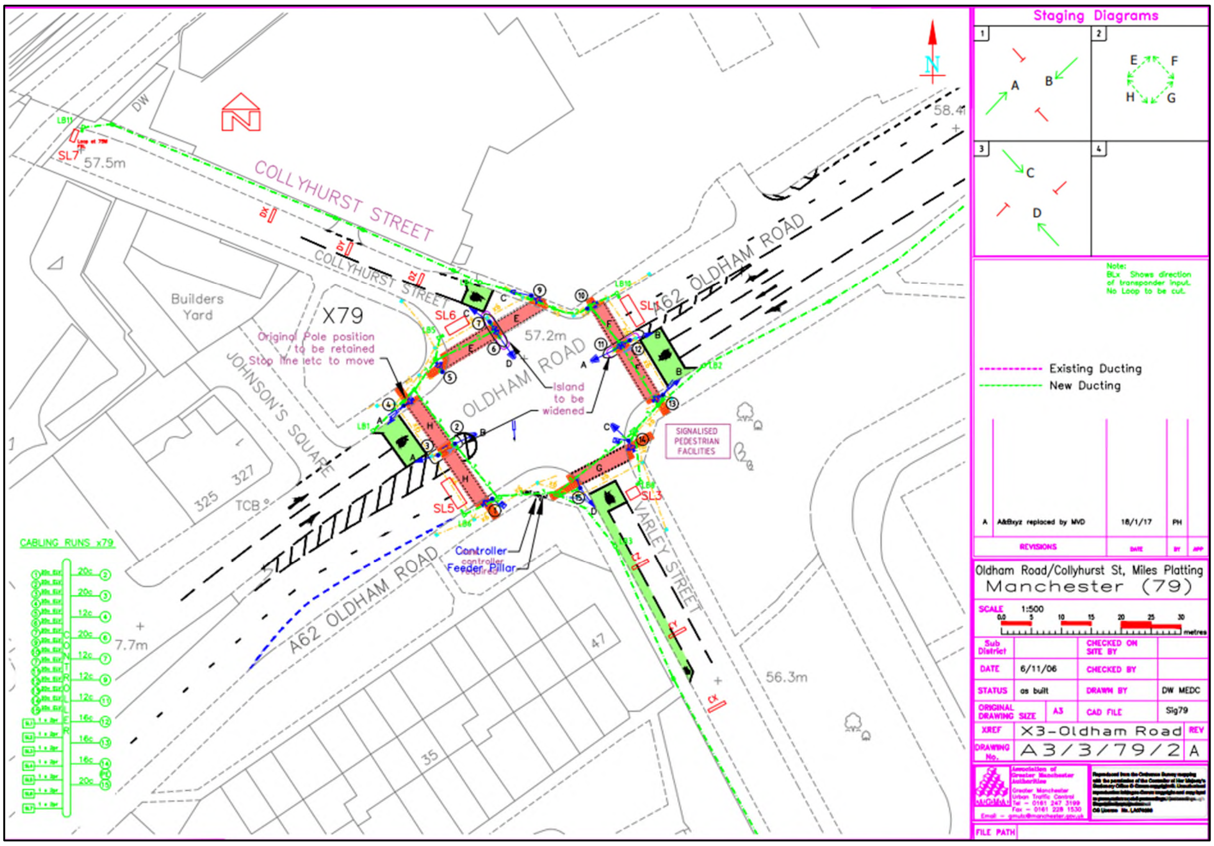

Figure 6: Site drawing for 0079.

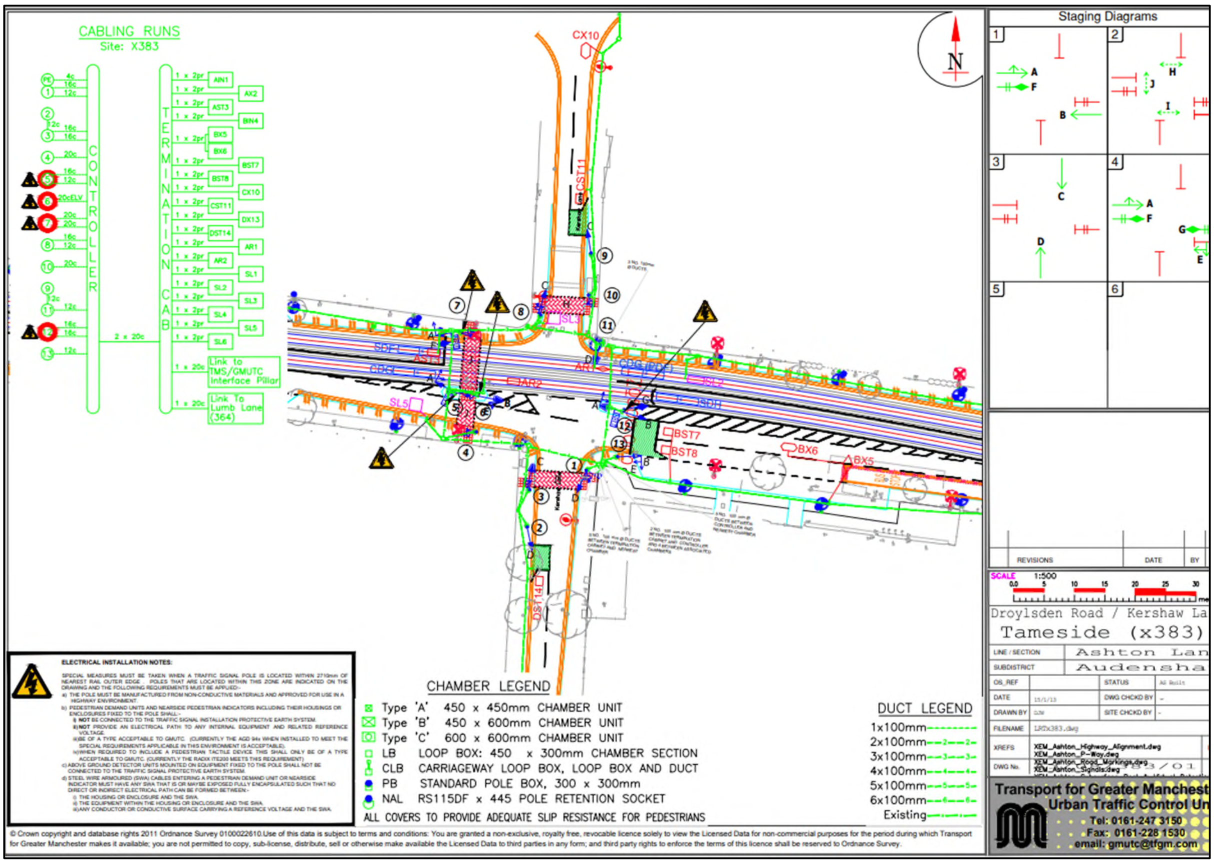

Figure 7: Site drawing for 0383.

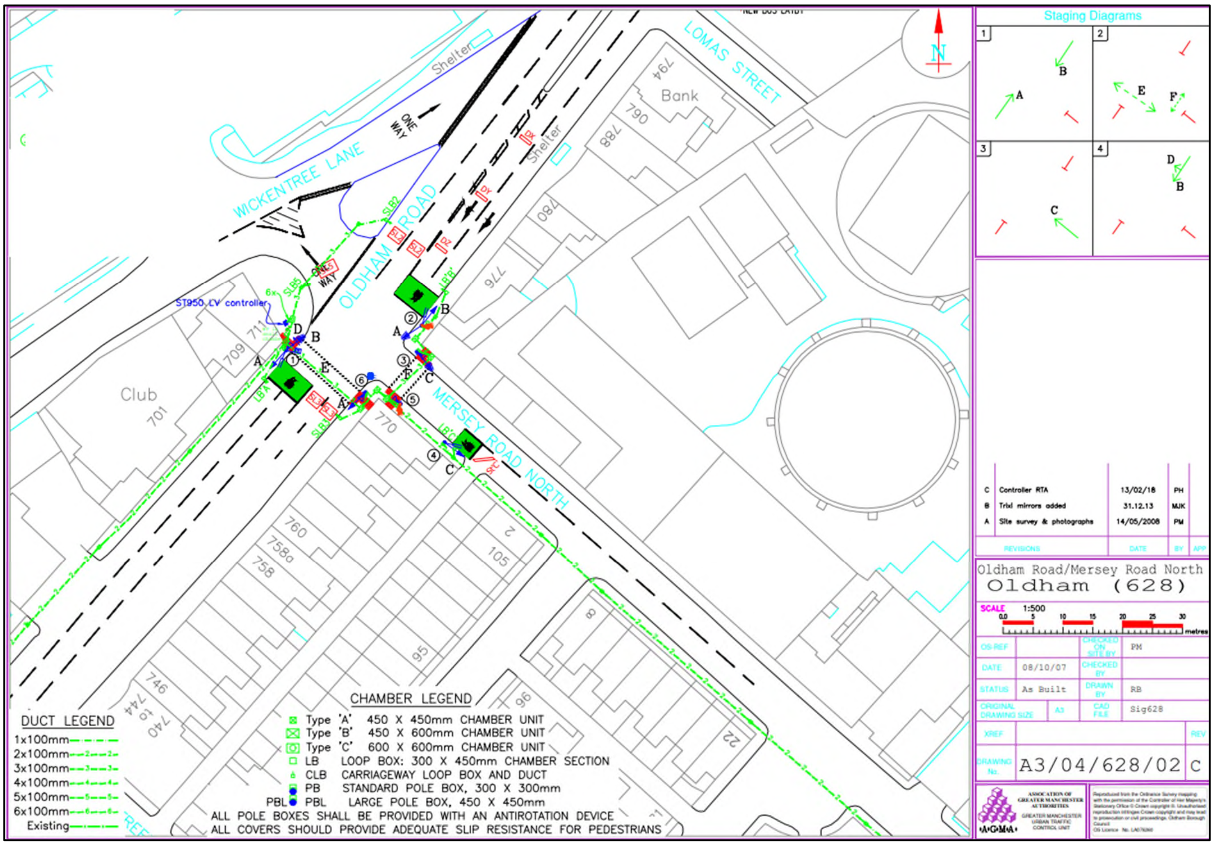

Figure 8: Site drawing for 0628.

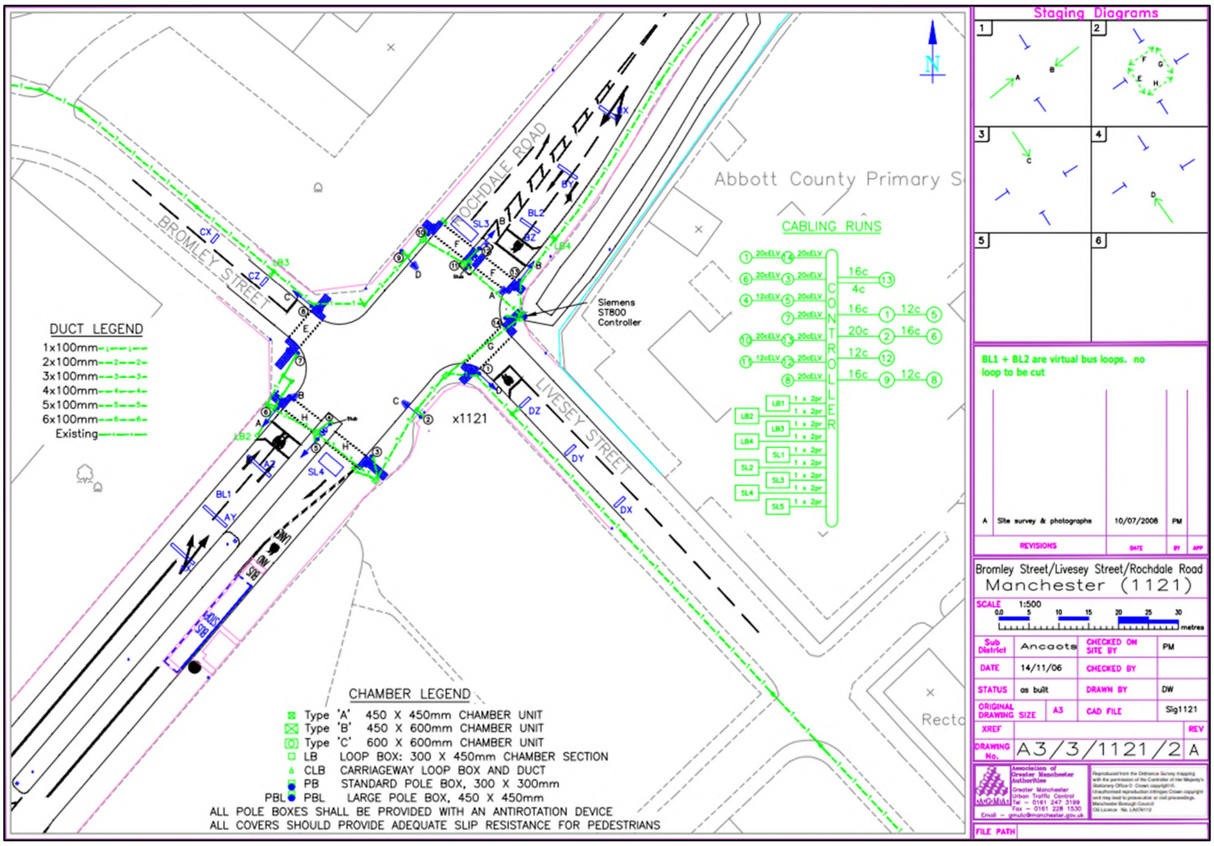

Figure 9: Site drawing for 1121.

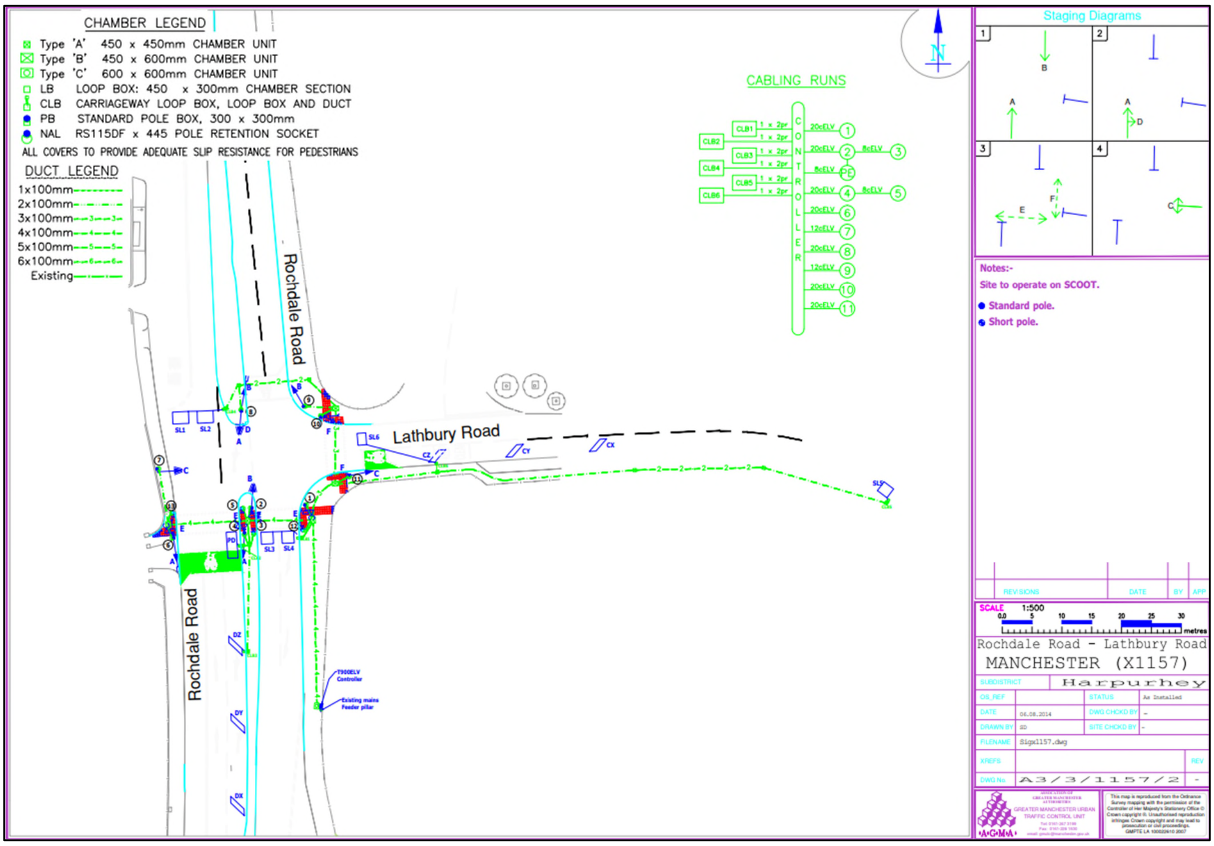

Figure 10: Site drawing for site 1157.