1 How to enhance Automatic Number Plate Recognition surveys when deriving travel patterns – the value of simulation

Ben Mackley and Tom van Vuren, Mott MacDonald

2 Introduction

The recording and matching of number plates, licence plates or registration plates has long been part of the toolkit for developing trip matrices for use in transport planning and modelling. (Watling, 1994) refers to (Makowski and Sinha, 1976) as having reported the first statistical approach to the matching of partial number plates. Software packages were developed, of which the best-known ones in the UK were MVMACH (MVA Systematica, 1987), NOPCOP (Lucas C. F., 1986) and Micromatch (Colin Buchanan and Partners, 1986).

Partial number plates recorded by hand using pen and paper or by using dictaphones were labour intensive and subject to human error, whilst the sampling approach inevitable led to further problems in the determination of OD matrices, mainly false positives and false negatives, skewing the resulting trip distribution.

As a result, the use of number plate matching became limited to fewer projects, and mainly localised ones, such as junction movement studies. Cameras became the preferred means of recording the number plates and although the capabilities and reliability of these cameras increased over time, the analysis of the resulting videos remained a mainly manual task, depending on human intervention. It has only been recently that the quality of the cameras and the supporting software has enabled the development and use of Automatic Number Plate Recognition (ANPR) techniques; and these services are now offered by most commercial survey organisations.

ANPR surveys, capturing in principle almost all passing vehicles, unbiased and at a known location on the network) are a welcome addition to other ways of estimating travel patterns, that each have their own shortcomings:

- Roadside interviews (RSIs) are becoming increasingly difficult to carry out due to Health and Safety concerns and impact on traffic conditions (see e.g. (Daily Post, 2017));
- Techniques for using data obtained from mobile network operators, though attractive from a point of view of bias and sample size, are still being developed and verified (Tolouei et al, 2015)
- GPS-based data sources (such as obtained from TrafficMaster) suffer from low (approx. 1-2%) and potentially biased sample rates (van Vuren and Carey, 2011)

The attraction of ANPR data is that in principle it is possible to observe 100% of the traffic passing cameras. In this paper we address one of the problems unique to ANPR surveys – caused by capture rates not being quite 100%. The resulting errors are similar to those encountered using partial number plate techniques in the past – false positives and false negatives resulting in a skewed trip distribution. We present a simulation-based approach to handle low capture rates and illustrate how such an approach can help reduce errors and strengthen the resulting trip matrices for practical use.

3 Other work in this area

There are existing papers that consider the use of ANPR as a tool in route flow estimation such as (Castillo, Menéndez and Jiménez, 2008) and (Castillo et al., 2010), which also contain consideration of ‘error recovery’. However, the topic is considered in terms of designing ANPR data collection to prevent and reduce errors, with discussion around using duplicate cameras on key links. Our paper will address the issue from a different angle, assuming the ANPR data collection to be external and existing, and hence looks at post-survey methods of correction. It also only considers extraction of OD pairs from the data, a simplification in comparison to route flow estimation, but the typically desired output for the transport modelling industry. Overall, this paper is intended to demonstrate practical application and validation rather than a theoretical approach.

There are also many existing papers regarding ANPR error reduction in the process of matching number plates, one example is (Oliveira-Neto, Han and Jeong, 2012), and dealing with misread number plates as opposed to missed number plates. Again, this paper assumes the matching of number plates to have been carried out externally, and for appropriate error-reduction methods to have already been applied.

4 ANPR Overview

4.1 Automatic Number Plate Recognition Technology and Applications

In simplest terms, Automatic Number Plate Recognition (ANPR) technology captures and records passing vehicles by identification of their vehicle registration number (VRN). A common application of this technology is to track movements of VRNs via a network of cameras, forming chains of each capture instance, with timestamps for each capture instance. The outputs can be processed into many formats, but in this paper we will be working with *trip chains*; ordered sequences of cameras corresponding to the capture instances of a VRN. In this paper, we will display trip chains in the following format:

- For a trip through camera C_1, then camera C_2 and then camera C_3 we write the trip chain as $C_1 > C_2 > C_3$;
- For a trip through camera C_1 only, we simply write the trip chain as C_1.

Trip chains and their associated timestamps can be used for a number of uses:

1 Identification of each VRN by vehicle and emission classification (using databases such as those maintained by the DVLA);
2 Extraction of origin and destination camera pairs, with disaggregation by time and vehicle class;
3 Average journey times between cameras;
4 Analysis of specific camera-to-camera movements, with potential for analysis of routing patterns or turning movements.

These are valuable tools in a wide-range of transport modelling applications with **1** having obvious strength for environmental modelling. We will limit our scope to exploring applications of **2**; analysing extracted origin and destination data and examining its integrity. In doing this, vehicle classifications have been disregarded and arbitrary time periods have been considered, but expanding the scope with respect to these is a straightforward reworking.

4.2 Extracting Origin-Destination Information from ANPR

The extraction of OD pairs from ANPR outputs is not a demanding task thanks to the format of trip chains and their 1-1 correspondence with trips. OD pairs are obtained by extracting the first and last camera from each trip chain and the resulting OD matrix can be converted into zone-to-zone trips by use of a camera-to-zone correspondence. For simplicity, we will only consider camera-to-camera trips.

The value of ANPR OD matrices depends on how the trips are extracted from the ANPR captures; if no duration constraint is applied then tours are likely to be represented as a single trip with the same origin and destination. To maximise the number of genuine trips captured, we have used data that applies the following constraints when building trip chains:

1 No trip duration longer than two hours; when building a trip chain, if the next capture instance pushes the trip duration over two hours then the trip chain is not extended further. A new trip chain is started from that capture instance instead.
2 Depending on their location with respect to the network, cameras have been deemed as either internal or external (inbound & outbound) cameras.
  a. An inbound camera will automatically start the formation of a new trip chain;
  b. An outbound camera will complete a trip chain being built.

External cameras either mark the extents of the ANPR network or the entrance to a specific purpose site, and hence we consider traffic passing these points to be beginning or ending their trips with respect to the survey.

The value of ANPR OD matrices is also dependent on camera locations, with considerations for individual cameras and the camera network as a whole. Camera networks can be set-up to focus on areas of interest and appropriate use of external cameras can help to ensure individual trips are captured appropriately. Depending on the purpose and focus points of an ANPR survey, either partial OD movements or full OD movements can be obtained.

4.3 Measuring ANPR Camera Performance

Although ANPR cameras have improved significantly since first being developed, they still have their limitations in recognising 100% of passing VRNs. In modern day surveys, cameras can capture between 85% and 95% of the passing vehicles; according to one supplier, a 92.5% average sample rate on 1,500,000 plates was collected on the M25 in daylight and night conditions (Intelligent Data, 2015). Compared to traditional methods of data collection, this is a very high sample rate.

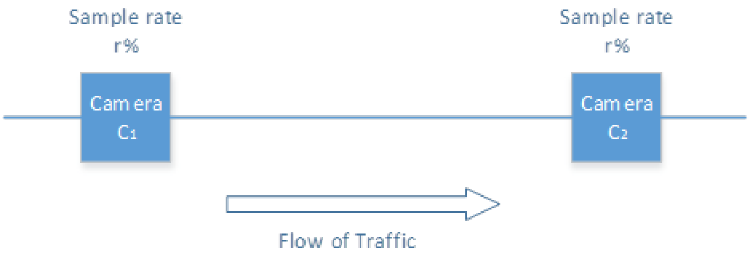
These benchmark rates can be made more difficult to achieve by external factors such as lighting, weather and camera alignment/set-up (Gurney et al., 2013). Temporary ANPR installations, as most transport surveys require, increase the chances that camera set-up and positioning is sub-optimal. Hence questions remain on how large an impact poor capture rates have on extracted OD movements and how is best to correct for these impacts. This forms the motivation for this paper.

To consider ANPR camera performance, we first need a measure to evaluate against. This is achieved using independent video cameras at ANPR camera sites, to capture footage that can be processed into volumetric counts. Against these volumetric counts, the number of detected VRNs at each location can be compared to form a sample rate for each camera. The temporal resolution depends on how the volumetric counts have been carried out; for our purposes, we have calculated a sample rate for each camera during each 15-minute period of the day.

4.3.1 A Simple Network Application

To form a base understanding of how sample rates affect observed data, we first consider a simple example shown in **Figure 1**. From this example, we hope to identify a number of sampling consequences which can help us to understand analysis of more complex networks.

Figure 1: Simple Network Example



Within this network, we define the following traffic:

- The cameras are both on the same one-way link (direction from C_1 to C_2);
- We define 100 vehicles with trip chain $C_1 > C_2$;
- We also define that all vehicles passing C_1 must pass C_2 and vice versa. Hence there are 0 vehicles with trip chain C_1 and 0 vehicles with trip chain C_2,

- We define that both cameras have the same sample rate r, although these are independent of each other.

Under these conditions we can calculate that $C_1 > C_2$ will be observed as having $100r^2$ trips, C_1 and C_2 will each have $100r(1-r)$ trips and $100(1-r)^2$ trips will not have been observed at all. This is illustrated for explicit samples rates in **Table *1***.

Table 1: Effect of sample rates on a two camera scenario

Trip Chain	Origin	Destination	100%	95%	90%	85%	80%
$C_1 > C_2$	C_1	C_2	100	90.25	81	72.25	64
C_1	C_1	C_1	0	4.75	9	12.75	16
C_2	C_2	C_2	0	4.75	9	12.75	16
Undetected	n/a	n/a	0	0.25	1	2.25	4
Total Detected	-	-	100	99.75	99	97.75	96

An immediate observation that can be made from **Table *1*** is that sampling does not lead to a reduction in trip volume for all trip chains/OD pairs; $C_1 > C_2$ has a reduction in trip volume, whereas C_1 and C_2 each have an increase in trip volume. To explore why this occurs, we consider the three effects that sampling can introduce on the vehicle's origin and destination:

1 *Sampling can have no effect on the origin and destination*; this occurs when vehicles are at least observed at the first (origin) camera and last (destination) camera on their route.

2 *Sampling can alter the observed origin and destination*; this occurs when a vehicle is observed by at least one camera, but not observed at either the origin or destination camera or both.

3 *Sampling can mean the vehicle is not observed*; this occurs when a vehicle is not observed by any of the cameras that it passes.

Since our output of interest are OD values, we consider the impact of **1-3** in this respect. For any given OD pair:

- **1** has no effect on the OD trip volume;
- **3** leads to a decrease in trip volume with no effect on the trip volumes of any other OD pairs;
- **2** leads to a decrease in trip volume but will lead to an increase in trip volume for the new 'altered' OD pair.

Hence in **Table *1***, the reduction in trip volume for $C_1 > C_2$ is caused by both **2** and **3**, and the knock-on effect of **2** causes an increase in trip volume for C_1 and C_2. Applying these observations to the general case, we consider that the sampled trip volume for an OD pair is structured as follows:

$$sampledOD = originalOD - k_1 + k_2$$

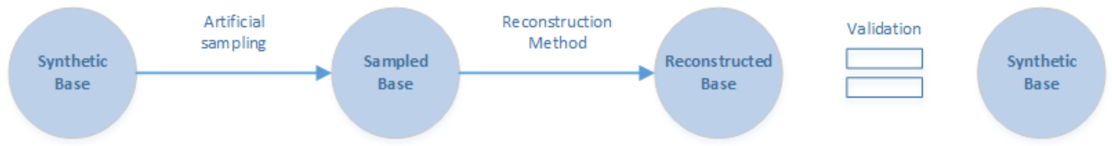
Here we call k_1 the **reduction factor** which represents the decrease caused by **2** and **3**, and we call k_2 the **redistribution factor** which represents the increase caused by the knock-on effect of **2**. To develop reconstruction methods, we will need to determine for each OD pair what the magnitude of the reduction and redistribution factors are. Determining the magnitude of the reduction factor will be significantly easier as it depends only on the trip volume of the original OD pair and on the sample rate of the origin and destination camera. Being able to visualise the impacts of sample rates in the quantitative way that **Table *1*** does will be a powerful tool for shaping how these corrective factors should be applied.

4.3.2 Application to Surveyed Data

In transport modelling, it is commonplace to apply methods to correct or augment incomplete input data. Direct validation of these methods is difficult, as the input data has most likely been chosen as the best representation of reality and so there is no better representation to validate against. Validation is often obtained later once the data has been applied to a network or model but this is limited in how well it can correct the input data, particularly if there is a risk of structural error or bias. This is particularly relevant here given the observations made in 4.3.1 regarding redistribution and the clear structural error it introduces.

In 4.3.1, we have simulated sampling scenarios on a synthetic base dataset and made observations from the results. Given a reconstruction or correction method, it would not be difficult to apply the method to the sampled scenarios and then compare the results to the base dataset by means such as linear regression (of OD or trip chain volumes). This is in effect a form of direct validation as is demonstrated in **Figure 2**. Hence our intention is to develop a generic 'simulation approach' that can be applied to correction methods on more complex ANPR networks and outputs. Below we address some further requirements for this and address how to make this applicable in a practical context.

Figure 2: Simulation Approach



5 Simulation Approach

As is the case in 4.3.1, the base situation needs to be defined and known, to allows us to carry out the validation step as described in **Figure 2**. Given a set of ANPR surveyed trip chains, it is suggested that a synthetic base should be chosen that is as similar to reality as possible. It seems feasible that problems could arise when simulating sampling on a set of trip chains with an average of 10 cameras per trip chain, which might not arise in a set of trip chains with an average of 2 cameras per trip chain.

Analysis of this has not been carried out by the author and is outside the scope of this paper. This paper will aim to demonstrate the effect that sample rates can have and that a reconstruction method exists and is feasible, using a single synthetic base. This synthetic base will be derived from a set of trip chains received from an ANPR survey containing around 50 camera sites. Further work in this area is desirable and could take two suggested forms:

1. Finding suitable candidates to be used as a synthetic base given a set of ANPR outputs;
2. Demonstrate a given reconstruction method can be shown to be effective against a wide range of varying synthetic bases, removing the need to have a synthetic base similar to reality.

5.1.1 Artificial Sampling Method

An assumption that we make when simulating the sampling is that, at a given camera, a vehicle is equally as likely to be missed as any other vehicle. As is stated in (Castillo, Menéndez and Jiménez, 2008), this is an idealised assumption but given the information available this is the best assumption that can be made.

Using this assumption, we want to replicate cameras being removed from trip chains with a probability derived from the sample rate. In our synthetic base, we have trip chains containing sequences of cameras and the timestamp of each camera occurrence. We also have the sample rate of each camera per 15-minute period in the day. Hence for each camera C_i in an arbitrary trip chain T, we have the sample rate r for the camera at the time it was passed. Using a pseudo-random number generator, we can directly imitate the occurrence of vehicles being random sampled by the ANPR cameras. To do this, we generate a number $x \in [0,1]$ and apply the following algorithm;

1. If $x \leq r$ then do nothing;
2. If $x > r$ then remove C_i from T.

5.1.2 Impact of sampling scenarios on Total Trip Volume

Now that we have a synthetic base and a method to impose sampling rates on this base, we can visualise the impact some sampling scenarios have. The sampling scenarios we have tested involve imposing the same sampling rate to all cameras, with the scenarios ranging between 55% and 100%.

Scenario 100 – Sample rate for all cameras $r = 100\%$ (Synthetic Base)
Scenario 95 – Sample rate for all cameras $r = 95\%$
Scenario 90 – Sample rate for all cameras $r = 90\%$
Scenario 85 – Sample rate for all cameras $r = 85\%$
Scenario 80 – Sample rate for all cameras $r = 80\%$
Scenario 70 – Sample rate for all cameras $r = 70\%$
Scenario 60 – Sample rate for all cameras $r = 60\%$
Scenario 50 – Sample rate for all cameras $r = 50\%$

For each sampling scenario, we can compare the OD values obtained from the synthetic base, with the OD values obtained after applying the sampling scenario to the synthetic base. Each scatter diagram below displays a scatter point per OD pair, where the x co-ordinate is the OD value in the synthetic base, and the y co-ordinate is the OD value in the corresponding sampling scenario. By comparing the points to the line $y = x$ we can establish how an OD pair has been affected by the sampling scenario. The linear regression statistics displayed in the plot area should also give a quantitative idea of how the overall trip volume level and the overall OD distribution have been affected. As shown in **Figure 3**, where $r = 100\%$ (no sampling), we are aiming for a gradient of 1 and R^2 value of 1.

Figure 3: Scenario 100 vs. Synthetic Base

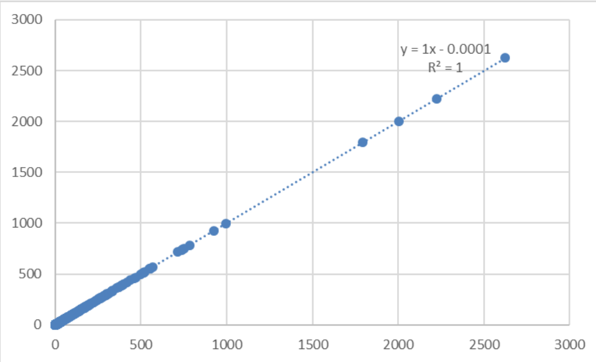


Figure 4: Scenario 95 vs. Synthetic Base

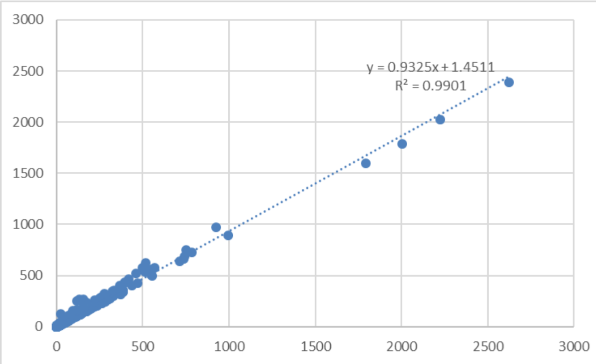


Figure 5: Scenario 90 vs. Synthetic Base

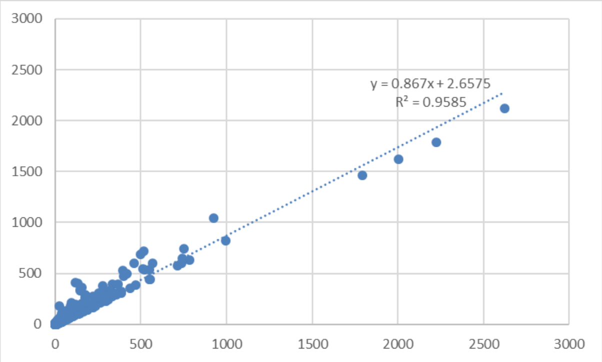


Figure 6: Scenario 85 vs. Synthetic Base

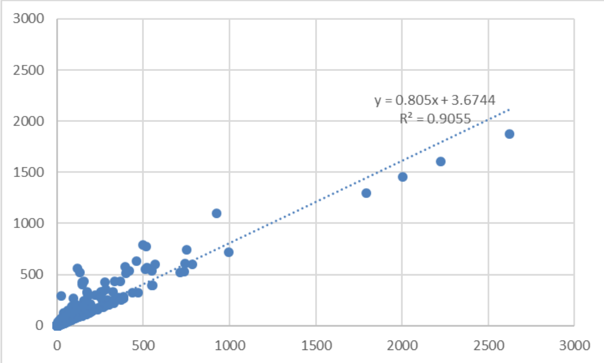


Figure 7: Scenario 80 vs. Synthetic Base

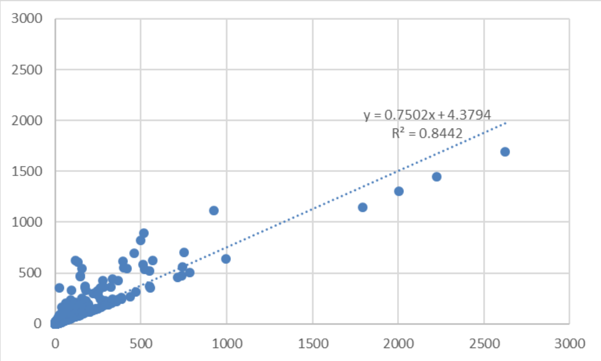


Figure 8: Scenario 70 vs. Synthetic Base

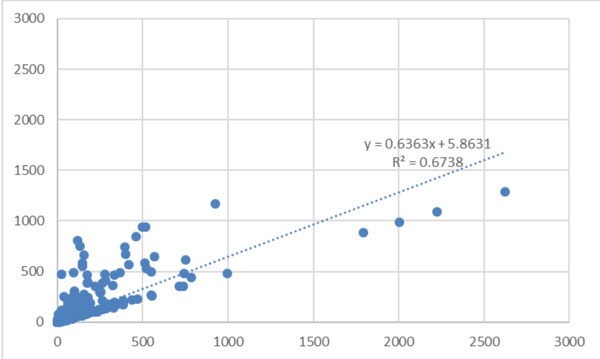


Figure 9: Scenario 60 vs. Synthetic Base

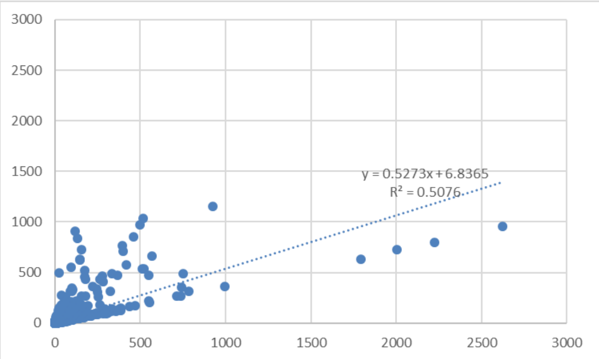
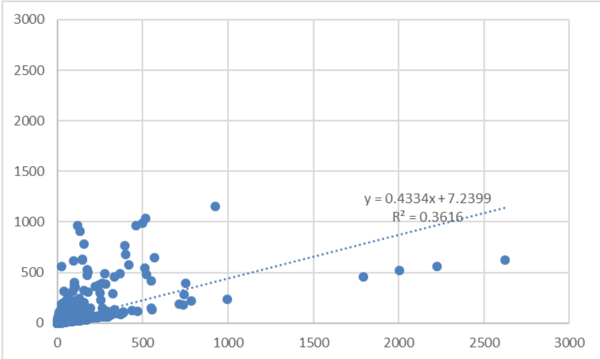


Figure 10: Scenario 50 vs. Synthetic Base



If we consider **Figure 4** and **Figure 5** then we can see that the data shows a strong fit for the line $y = x$ with R^2 values around 0.99 and 0.95 respectively. However, there are a group of high flow OD pairs which are positioned noticeably below the line $y = x$ and a cluster of points with lower flow that are sitting above the line. These further demonstrate the idea of a reduction factor and a redistribution factor. The points sitting below the line $y = x$ correspond to those where the reduction factor is greater than the redistribution factor, and the points sitting above correspond to those where the redistribution factor is greater than the reduction factor. These effects are only exacerbated as we consider more aggressive sampling scenarios and in **Figure 9** and **Figure 10** there is very little correlation, demonstrated by the very poor values of R^2.

6 Correction Methods

Now that we have created a synthetic base, applied sampling scenarios and established a method of comparing resulting datasets, we now look at developing a reconstruction method to return a sampled dataset to the synthetic base (or as close to). As detailed in **Figure 2**, we can compare the synthetic base with the reconstructed dataset to examine the extent with which the reconstruction method has been successful. Furthermore, we can compare regression statistics between the base and sampled datasets, with regression statistics between the base and reconstructed datasets to see if at least an improvement has been achieved.

6.1 Naïve Method

The first method we will test as a corrective method will only take into account the sample rates of the origin and destination camera to factor up each trip chain, and therefore OD pairs, to their original value. In a more formal sense, for each trip chain T (from the sampled dataset) with origin camera C_O and destination camera C_D, we calculate the sample rate of C_O, r_O, and the sample rate of C_D, r_D, over the time period we are building the OD matrix for. If T currently has a trip volume v then we replace v by

$\frac{v}{c_O c_D}$. We apply this to Scenario 80 (sample rate $r = 80\%$) below. This method is naïve as it only serves to increase trip volumes and hence can only be tackling the reduction factor, without consideration of the effect of redistribution. However, there is value in assessing how effective it is as the simplicity of the calculations involved are desirable.

Figure 11: Scenario 80 vs. Synthetic Base

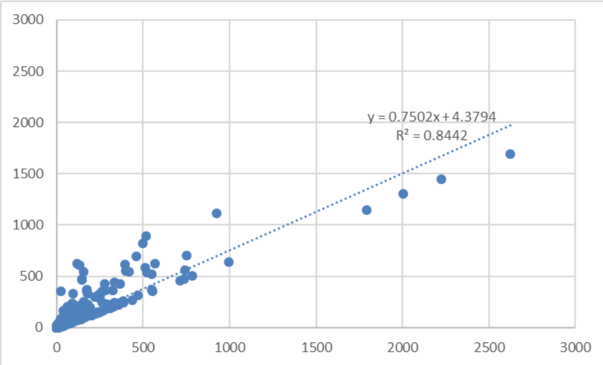


Figure 12: Scenario 80 Naïve Reconstruction vs. Synthetic Base

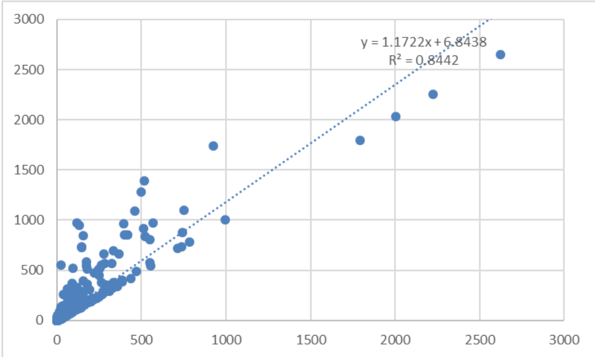


Table 2: Effect of sample rates on synthetic base

	Base	Sampled	Reconstructed
Total Volume of Trips	69298	65253	101960
% of Original Trips	100%	94.2%	147.1%

By comparing **Figure 11** with **Figure 12**, we can see from the R^2 value that this method does not improve the correlation with y=x. Additionally, ***Table 2*** shows that the total volume of trips is now badly overestimated; the naïve reconstruction only exacerbates the OD pairs that already sit above $y = x$ in **Figure 11**. This confirms the expectation that the naïve reconstruction method would not take the redistribution factor into account.

Another observation is that in **Figure 12**, there are a number of points which now lie on or in a very close neighbourhood of the line $y = x$, indicating that the naïve reconstruction method is effective at correcting the trip volumes for certain OD pairs. These points are almost certainly points where there is no contribution from the redistribution factor, indicating that the naïve reconstruction method is proficient in dealing with the reduction factor alone.

6.2 A More Considered Method

The naïve reconstruction method highlighted the importance of developing a method that tackles both the reduction and redistribution factor. In this section we will outline a more considered approach which is intended to do this. This method has been developed in a heuristic manner and by following a loose set of requirements, derived from **1** and **2**.

1 When adding to any trip volume, we must identify if it is adding to the overall matrix volume or if it has been redistributed from elsewhere (redistribution factor);

2 Where added trip volume is being redistributed from elsewhere, we must be able to identify where it has come from and remove it, to ensure that the overall matrix volume is maintained.

To explain the algorithm that we have developed we use the following terminology and notation;

- Let $set(T)$ be the collection of unique trip chains

- Let T be a trip chain with volume v. We define v^* as the reconstructed volume of T calculated through the algorithm
- Let C_O be the first/origin camera in T. C_O has sample rate r_O
- Let C_D be the last/destination camera in T. C_D has sample rate r_D
- Let T_O be the trip chain T with C_O removed. T_O has volume v_O and reconstructed volume v_O^*
- Let T_D be the trip chain T with C_D removed. T_D has volume v_D and reconstructed volume v_D^*

Algorithm A	**Comment**
1 For each T in $set(T)$, define $v^* = v$	v^* will be calculated from v by adding and subtracting factors
2 For each multi-camera T in $set(T)$	Cycle through each T with more than one camera in the chain (order is not significant)
a. Calculate $a_O = \frac{v}{r_O} - v$	Calculates the expected trip volume missed due to sampling by the origin camera
b. Set $v^* = v^* + a_O$	This satisfies **1**
c. If T_O is in $set(T)$, then set $v_O^* = abs(v_O^* - a_O)$	A vehicle travelling the path of T that is missed by the origin camera will have been recorded under T_O. To satisfy **2**, we must subtract this trip from v_O
d. Calculate $a_D = \frac{v}{r_D} - v$	Calculates the expected trip volume missed due to sampling by the destination camera
e. Set $v^* = v^* + a_D$	This satisfies **1**
f. If T_D is in $set(T)$, then set $v_D^* = abs(v_D^* - a_D)$	A vehicle travelling the path of T that is missed by the destination camera will have been recorded under T_D. To satisfy **2**, we must subtract this trip from v_D
3 For each T in $set(T)$, define $v = v^*$	Update the volumes to the newly calculated volume

Algorithm B	**Comment**
1 For each T in $set(T)$, define $v^* = v$	v^* will be calculated from v by adding and subtracting factors
2 For each single-camera T in $set(T)$	Cycle through each T with only one camera in the chain(order is not significant)
a. Calculate $a_O = \frac{v}{r_O} - v$	Calculates the expected trip volume missed due to sampling by the origin (only) camera
b. Set $v^* = v^* + a_O$	This satisfies **1**
3 For each T in $set(T)$, define $v = v^*$	Update the volumes to the newly calculated volume

By running **Algorithm A** followed by **Algorithm B**, we now have a set of trip chains with a reconstructed trip volume, from which a reconstructed OD matrix can be derived.

In step **2** in both **Algorithm A** and **Algorithm B**, note that order is not significant. This is because all addition and subtraction factors, are calculated from the initial v . If an order were applied and v were constantly updated to be v^*, then we risk a particular trip volume v decreasing to 0 (due to steps **5** and **8**) before it has received any uplift from steps **4** and **6**. In this instance the uplift would then be 0 because $v = 0$, which clearly in some cases could be an issue. In the absence of any logical way to order, it is assumed best to use the method outlined above.

If we compare the results of applying this algorithm we see the following results;

Figure 13: Scenario 95 vs. Synthetic Base

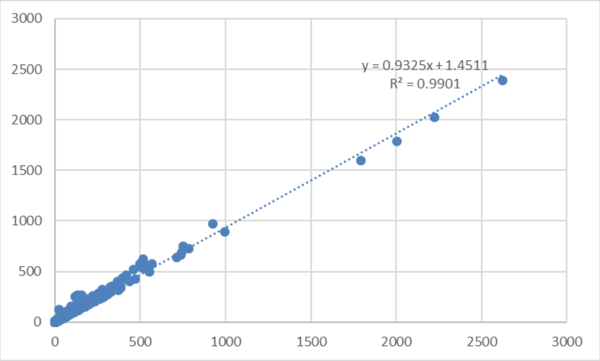


Figure 14: Scenario 95 Reconstruction vs. Synthetic Base

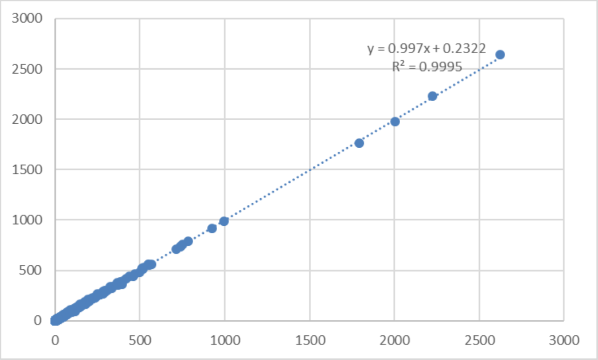


Figure 17: Scenario 90 vs. Synthetic Base

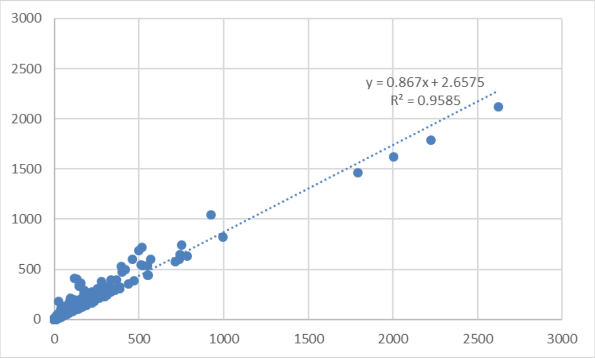


Figure 18: Scenario 90 Reconstructed vs. Synthetic Base

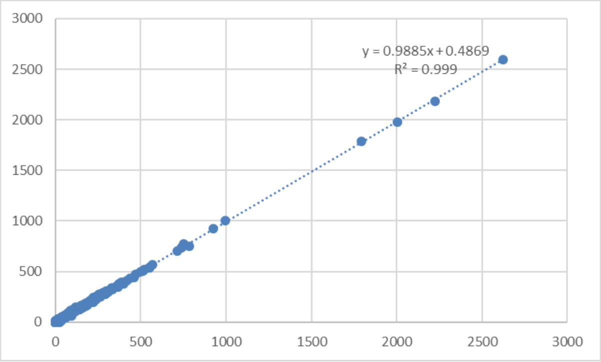


Figure 15: Scenario 80 vs. Synthetic Base

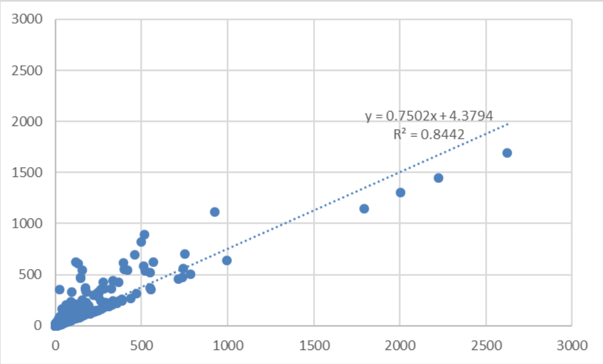


Figure 16: Scenario 80 Reconstructed vs. Synthetic Base

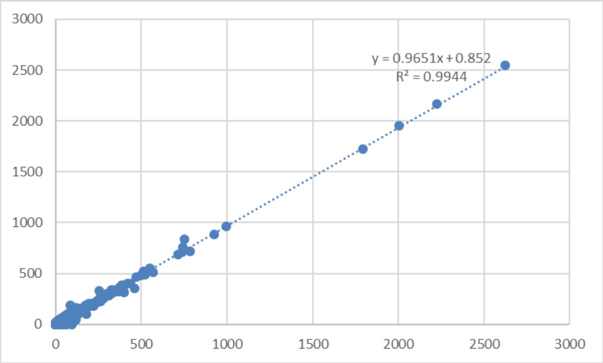


Figure 19: Scenario 70 vs. Synthetic Base

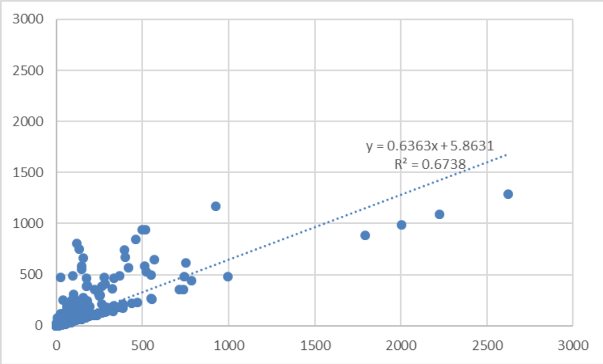


Figure 20: Scenario 70 Reconstructed vs. Synthetic Base

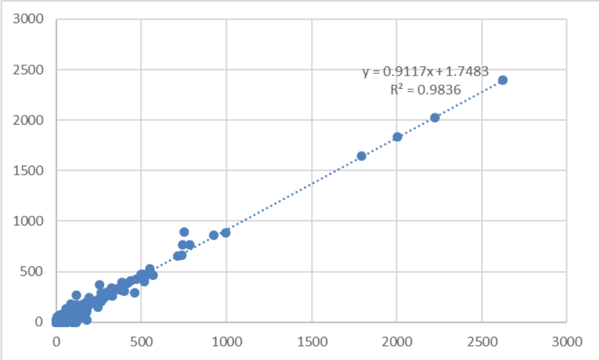


Figure 21: Scenario 50 vs. Synthetic Base

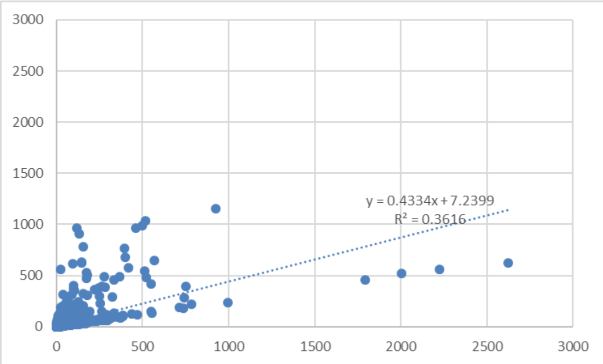
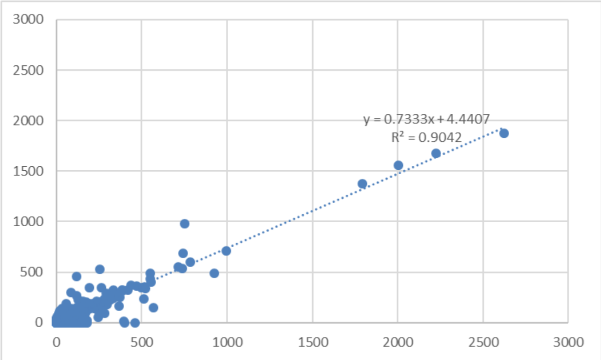


Figure 22: Scenario 50 Reconstructed vs. Synthetic Base



In the above figures, we can compare before and after reconstruction by comparing the gradients and R^2 values in the left and right scatter graphs. As can be seen, these demonstrate a significant improvement in correlation and gradient compared to the sampled scenarios we are reconstructing from, and compared to the results demonstrated by the naïve method.

To explain why the algorithm is in two parts; if we include the single-camera trips in Algorithm A (and not running Algorithm B) then the single camera trips are noticeably higher than the multi-camera trips. This is evident in **Figure 23**, where the red points (single-camera trips) seem to have a correlation independent of the rest of the data. As you can see in **Figure 24**, splitting the algorithm visibly improves the correlation and gives an improved value of R^2. There are intuitive explanations for this although no in depth analysis has been carried out; the decision to split the algorithm has been made purely based on observations during simulation.

Figure 23: With Single Trips Included in Algorithm A (and no Algorithm B)

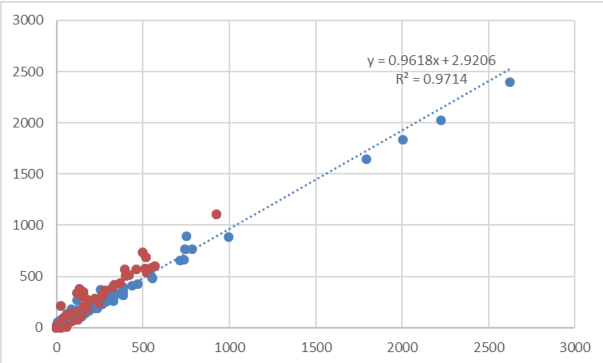
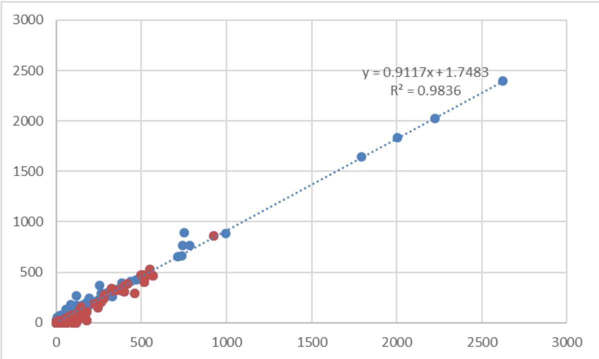


Figure 24: With Single Trips Included in Algorithm B



7 Conclusions

Poor ANPR camera performance has an unusual effect on OD matrices extracted from ANPR datasets, affecting not only their volume but also their distribution. We have shown that, using simple simulation methods and assumptions, these effects can be mapped and quantified. The methods that we derived reduce these impacts giving a strong correlation between base and reconstructed results, with the level of success depending on the extent of camera performance that is being corrected for. Importantly, methods like this can help to give further confidence in ANPR surveys as a mass collection method for obtaining origin-destination information. Particularly, our simulation approach can be a cost-effective way to recover lost information if external factors heavily affect the surveys, such as weather, impacting on the camera performance.

The purpose of this paper is as an introduction; there is further work that can be carried out to investigate the effects of sampling on different synthetic bases and whether reconstruction methods can be effective regardless of the structure of the ANPR data it is applied to. Further opportunities could be to explore the impact of sampling scenarios that may be encountered following ANPR surveys, such as a single camera or small group of cameras performing particularly poorly, and to what extent reconstruction methods can deal with error recovery in this scenario.

To the transport planning industry as a whole, a more general application of this paper comes from the basis of the simulation approach. If similar simulation approaches can be applied to error correction in other data sources and data manipulation methods, then this could provide valuable insight into the methods that we apply on a daily basis and guide enhancements to more accurate and reliable modelling.

8 References

Castillo, E., Menéndez, J. and Jiménez, P. (2008). Trip matrix and path flow reconstruction and estimation based on plate scanning and link observations. *Transportation Research Part B: Methodological*, 42(5), pp.455-481.

Castillo, E., Gallego, I., Menendez, J. and Rivas, A. (2010). Optimal Use of Plate-Scanning Resources for Route Flow Estimation in Traffic Networks. *IEEE Transactions on Intelligent Transportation Systems*, 11(2), pp.380-391.

Daily Post. (2017). *A55 traffic queues for miles as 'ridiculous' Welsh Government carries out census*. [online] northwales. Available at: http://www.dailypost.co.uk/news/north-wales-news/a55-traffic-queues-miles-ridiculous-12806781 [Accessed 1 Apr. 2017].

Gurney, R., Rhead, M., Lyons, V. and Ramalingam, S. (2013). *The effect of ANPR Camera Settings on System Performance*. 1st ed. [ebook] Available at: http://uhra.herts.ac.uk/bitstream/handle/2299/12527/The_effect_of_ANPR_Camera_Settings_on_System_Performance_RG_MR.pdf?sequence=2 [Accessed 5 Apr. 2017].

Intelligent Data. (2015). *Intelligent Data - ANPR Capability*. [online] Slideshare.net. Available at: https://www.slideshare.net/PaulONeill7/intelligent-data-anpr-capability [Accessed 1 Apr. 2017].

Makowski, G. and Sinha, K. (1976). A statistical procedure to analyze partial license plate numbers. *Transportation Research*, 10(2), pp.131-132.

Oliveira-Neto, F., Han, L. and Jeong, M. (2012). Online license plate matching procedures using license-plate recognition machines and new weighted edit distance. *Transportation Research Part C: Emerging Technologies*, 21(1), pp.306-320.

Tolouei, R., Alvarez-Indave, P.,Duduta, N. (2015). Developing and Verifying Origin-Destination Matrices Using Mobile Phone Data. The LLITM Case. *Proceedings of the 2015 European Transport Conference, Frankfurt.*

van Vuren, T and Carey, C (2011) Building Practical Origin-Destination (OD/Trip) Matrices from Automatically Collected GPS data. *Proceedings of the 2011 European Transport Conference, Glasgow.*

Watling, D. (1994). Maximum likelihood estimation of an origin-destination matrix from a partial registration plate survey. *Transportation Research Part B: Methodological*, 28(4), pp.289-314.