Knowledge Discovery and Causality in Urban City Traffic: A study using Planet Scale Vehicular Imagery Data

Damien Fay
Computer Science Department
University College Cork, Ireland
d.fay@4c.ucc.ie

Gautam S. Thakur
Oak Ridge National Laboratory
Oak Ridge, USA
thakurg@ornl.gov

Pan Hui
Dept of Computer Science and Engineering
HKUST, Hong Kong
panhui@ust.hk

Ahmed Helmy
CISE
University of Florida, Gainesville, USA
helmy@cise.ufl.edu

ABSTRACT

The increase in number of vehicles has created problems in many cities across the globe. Building comprehensive knowledge base about global city dynamics and traffic distribution is a key step to provide fundamental solution to the problems. In this paper, we examine a readily available data source; the existing infrastructure of traffic cameras around the world. We have collected real time traffic data from 2,700 public online traffic camera distributed across 10 cities in four continents for a duration of six months. Our platform allows us to automatically search public cameras, collect and process imagery data, remove outliers, and extract traffic density from those images in a highly scalable way. A time series model employing a co-integrated vector autoregression model is presented in which traffic forecasts may be produced and regions of the city not well observed may be suggested. In addition, a topological comparison of six of these networks is presented.

Keywords

Urban Infrastructure, Vehicular Traffic, Causality, GIS

Categories and Subject Descriptors

H.2.8 [**[Database Management]**]: Database Applications-Data mining, Image databases, Spatial databases and GIS.

General Terms

Experimentation, Human Factors, Measurement.

1. INTRODUCTION

The increase in number of vehicles has created a problem of traffic congestion in many world cities. In attempting to solve this problem, isolated approaches like improving the design of road and intersections, and changing the usage patterns have been considered. Although it has rewarded to some extent, the root cause for such an aggregation still persist. We believe these approaches should be augmented with a comprehensive picture of cities' structural dynamics and the traffic distribution across its key intersection in a collective manner, and more global cities should be sampled to build a rich knowledge base for such studies.

In this paper, we utilize for the first time a readily available source; the existing global infrastructure of thousands of video cameras, providing a continuous stream of street images from dozens of cities around the world. We introduce a novel monitoring, analysis and prediction framework, which consists of a network of planet-scale public webcams, fast and scalable traffic density estimation algorithm, and statistical models for traffic forecasting. Our dataset consists of 125 million images from over 2,700 traffic web cameras in 10 cities/states for six months, with a overall size of 7.5 terabytes. These regions are spread across North America, Europe, Asia, and Australia. In this paper, we have selected six cities with similar time granularly for a fair comparison of this study.

Our algorithm to estimate traffic density employes, scalable, and effective background subtraction technique to process millions of traffic camera images, and build an extensive library of spatio-temporal vehicular density data. Based on the mixture of gaussians, this algorithm is robust to outliers (camera errors) and sensitive to frequently changing lighting conditions. A comparison with ground truth (of number of cars) shows a near linear correlation to allow analysis at a network level.

Moreover, we have constructed a time series model for the traffic data using a co-integrated vector autoregression model, to reveal the extent to which the traffic observed at a point in the network is explained by that observed at another location. We employ a *Granger network*; a network in which causal links are identified; to provide a sparse representation and also to reveal the major pathways in network. A major contribution here is the actual model itself, which is the building block for understanding city dynamics.

To this end, our contributions are:

- We provide a novel framework to study the cause and effect relationship of traffic causality in urban streets using thousands of on-line web cameras. In future, we also plan to release the dataset to the research community,

ACM SIGSPATIAL IWCTS'13, Nov 05-08 2013, Orlando, FL, USA

- we establish that causality on motorways (inter-states) are far more perceivable than city's local traffic. This is a distinguishing factor that can be used for profiling the cities, and
- we construct an empirical time series model based on real data that considers the interaction between the traffic observed at different points in a city. The models produce forecasts with a typical error (PMSE) of 3-9%. It also reveal that traffic camera networks tend to be disassortative; many smaller junctions feeding into larger ones. It is envisaged that models constructed here will form the basis of future studies into vehicular traffic dynamics in cities based on real data.

2. MEASUREMENT AND PRE-PROCESSING

Table 1 details the regions used in this study and the extent of the data and time span of the sample. The snapshots taken at every camera (intervals ranging from 20-60 secs.) first pass a background estimation and subtraction phase. These are then used to estimate the *traffic density* arriving per unit time as opposed to a car count.

While a car count might seem preferable to a traffic density measure, there are several practical challenges. A car count requires a far greater computational cost due to the effort required to isolate each object. Traffic congestion further complicates matters when cars occlude each other, making it difficult to segregate cars based on edge structures. In addition, vehicles at the far end of the road are small in the image and cannot be detected by these algorithms. These web cameras are installed at strategic intersections and highway segments, which are critical in assessing issues such as traffic jams and evacuation mitigation plans. Deploying such cameras everywhere is not feasible, so we believe by covering these key locations, the results from this study will provide substantive understanding on the causes of causality and traffic prediction.

2.1 Background Subtraction

Background subtraction is a standard method for object localization in image sequences with fixed cameras, where the frame rate is lower than the velocity of the objects to be tracked (i.e. cars move out of the scene typically at a rate exceeding 1 minute). The basis for models of background are based on the observation that *background* does not change significantly (in comparison to foreground/objects) across time. Any part of an image that does fit with that model is deemed as *foreground/object*. These foreground regions are then further processed for the detection of desired objects.

The background model used here assumes that the distribution of background pixel values may be modeled as a weighted sum of Gaussian distributions. Our approach follows closely that proposed by [3, 9, 10] because of their reliability and robustness to sensitive changes in the lighting conditions. In our approach, the observed pixel value is modeled by a weighted sum of Gaussian kernels. Let x_t represent a pixel value in the t^{th} frame, then the probability of observing this value is assumed to be:

$$p(x_t) = \sum_{i=1}^{K} w_i^t * \mathcal{N}(\mu_{i,t}, \Sigma_{i,t}) \tag{1}$$

where $\mathcal{N}(\mu_{i,t}, \Sigma_{i,t})$ is the i^{th} kernel with mean $\mu_{i,t}$ and covariance matrix $\Sigma_{i,t}$, and w_i^t is the weight applied to that kernel such that $\sum_i w_i^t = 1$. We assume that RGB channels are uncorrelated thus the covariance matrix for each kernel is diagonal. When a new frame arrives, the pixel values are compared to the kernels to determine if it is likely that this value was drawn from a distribution with $\mathcal{N}(\mu_{i,t}, \Sigma_{i,t})$ (using for example a 95% confidence interval). If so, $\mu_{i,t}$, $\Sigma_{i,t}$ and w_i are updated using exponential filters; if not a new kernel is created and the existing kernel with the lowest w_i is eliminated (see [10] for specifics). Short lived kernels and their associated pixels are deemed to be possibly foreground producing a binary map. Morphological operations are then applied to this map to remove noise and any blobs with area smaller than a certain threshold.

The view of most cameras used in this study is along the direction of the road and this perspective skews the size of objects on an image [5]. To counter this effect, we weigh each foreground pixel with the exponent of it's distance from the bottom of the image. Thus a pixel in the bottom of the image will be weighted less (object appear larger at the bottom than on the top) than a pixel at the top. While this weighting is not exact and does produce some warping as we shall see in the next section it is not excessive but it is simple and does not require manual tuning to each camera.

To test the performance of the car density capture, six cameras were selected at random were examined by hand and labeled to produce a *ground truth* count for the number of cars. This ground truth was then regressed against the measured car density to check that the relationship is linear. Overall, the analysis shows that while there are some errors, the relationship between the actual and measured number of cars is sufficiently clear to allow analysis at a network level.

3. GRANGER NETWORKS

At this point the data has been distilled so that each camera node has an associated time series. Specifically, define $y_i(t)$ as the time series associated with the i^{th} node at time t. Note, $y_i(t)$ is a time series of *traffic density* at each traffic cam and is linearly related to the number of vehicles at that traffic camera. However, from $y_i(t)$ alone, it is not possible to tell the number of cars that have actually passed through the junction. A composite view of the network can, however, be formed by looking at all the junctions together by examining the *empirical* behavior[1] of the time series.

There is a strong correlation between all the time series in this study. Indeed, there is even a strong correlation between the time series of different cities. However, these are *spurious correlations* and are caused by factors such as the traffic increasing in the lead up to rush hour; regardless of location. Thus standard correlation analysis is misleading in this case. What is of interest, is the movement of traffic around the city; the flow of one time series into another. Note that this flow is directional and requires identifying the causations in the network; the type of network thus described is known as a *Granger Network* (GN) [6, 2].

A Granger network is a weighted directed graph as $G = (V, E)$ where V is the set of vertices (nodes) and E is the set of edges (links). The adjacency matrix of G, $A(G)$, has an entry if $v \to u$, and zero otherwise

$$A(G)(u,v) = \begin{cases} w_{u,v}, & \text{if } v \to u|q \\ 0, & \text{if } v \nrightarrow u|q \end{cases} \tag{2}$$

where $w_{u,v}$ is the strength of the connection between u and v as now explained. Given that v Granger causes u does not mean that the strength of the causality is strong; it merely means that it is consistent. The strength of the connection is instead measured by the coefficient of variance explained [7] by inclusion of a (Granger) causal variable:

$$R^2_{v\to u} = \frac{\hat{\sigma}^2_{\epsilon_{u|v,q}} - \hat{\sigma}^2_{\epsilon_{u|q}}}{\hat{\sigma}^2_u} \tag{3}$$

[1] i.e. no underlying model of physical transport is assumed.

Table 1: Global Webcam Datasets

City	# of Cameras	Duration	Interval	Records	Database Size
Beaufort	70	30/Nov/10 - 01/Mar/11	30 sec.	24.2 million	1150 GB
Connecticut	120	21/Nov/10- 20/Jan/11	20 sec.	7.2 million	435 GB
London	182	11/Oct/10 - 22/Nov/10	60 sec.	1 million	201 GB
Sydney	67	11/Oct/10 - 05/Dec/10	30 sec.	2.0 million	350 GB
Toronto	89	21/Nov/10 - 20/Jan/11	30 sec.	1.8 million	325 GB
Washington	240	30/Nov/10 - 01/Mar/11	60 sec.	5 million	400 GB
Total	**768**	-	-	**41.2 million**	**2861 GB**

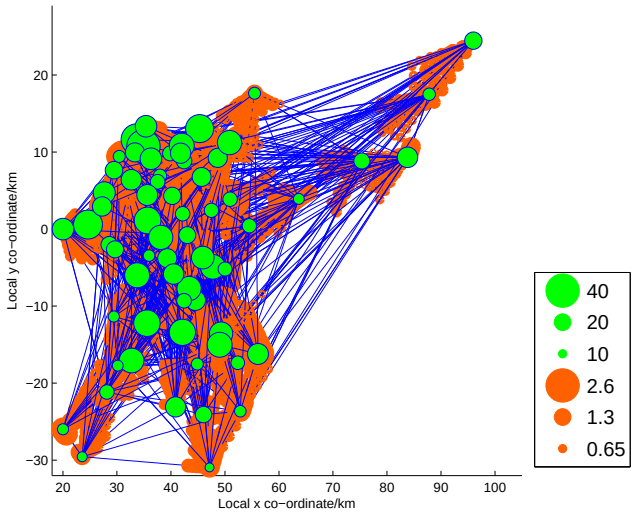

Figure 1: Granger network for Sydney.

where $\hat{\sigma}_u^2$ is the variance of u. In addition, the total variance explained at a node, denoted R_u^2, is the increase in variance explained in u by inclusion of *all* the information from its neighbors.

In summary, a Vector Auto-regressive model is constructed and the Hypothesis is tested. If Hypothesis is rejected then a casual link is said to exist between $y_i(t)$ and $y_j(t)$ whose strength, $w_{i,j}$ is the increase in $R^2_{v\to u}$ of $z_i(t)$ by inclusion of $z_j(t)$.

4. RESULTS

4.1 Network Analysis

Figure 1 shows the Granger network constructed using all the data from the Sydney traffic system. This figure shows quite a bit of information and is now explained. The X and Y axis are local $X-Y$ co-ordinates constructed from GPS co-ordinates using the Carlson and Clay model [4]. The green nodes are located at the actual local $X-Y$ co-ordinates for each camera. The size of the green node is proportional to the total variance explained at that node (given all the incoming nodes), i.e. R_u^2 as shown in the legend. The only edges shown are those for which the alternate hypothesis is accepted. The weight on an edge from u to v, i.e. $R^2_{v\to u}$ (Equation 3) is represented by orange circles proportional to the weight (shown in the legend) and are located at the tail incident on v; that is, they represent the variance explained flowing into a node from the source. As can be seen from the figure the network has a sparse structure with links existing mainly between physically proximate nodes or in the case of exterior nodes, pointing inward towards the city centre. Specifically, of a total of $N^2 = 4,356$ possible links, only 119 are found to be significant.

Figure 2 shows the co-integrating factor for the Sydney network

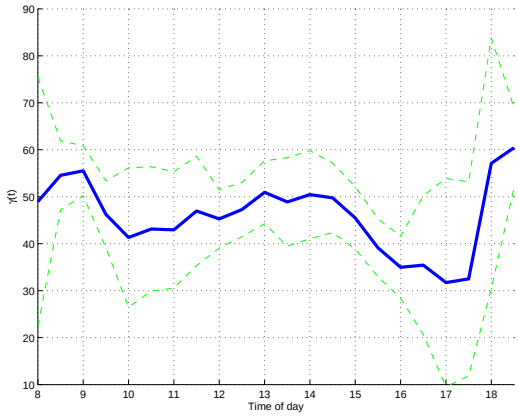

Figure 2: The co-integrating factor for Sydney. (Blue smoothed, ± 95% conf. intervals in green)

averaged by each 15 minute interval. This may be thought of as the average state of the network and shows clearly the morning and evening rush hours. It is interesting to note that just before the evening rush hour there is a lull in the traffic and in addition quite a large traffic volume during lunch. In section 4.2 this fact will be used to disaggregate the data by hour of the day into three regions.

Returning to Figure 1, it can be noted that adjacent nodes tend to have similar total R_u^2 values. This is not surprising; a well observed location (in the Granger causal sense) is likely to have many neighbors which re-enforce each other. However, it is interesting to ask if there are regions in which the R_u^2 values are lower than expected and at which an extra camera might be beneficial. The analysis carried out here employs a *Gaussian Process* (GP) to interpolate the R_u^2 values. A GP is the appropriate interpolation to use here as it meets required constraints on the problem; the function should tend to zero in the absence of cameras and the degree to which one camera effects another is unknown and needs to be estimated from the data. This is based on the simplifying assumption that the total variances' explained at the nodes are samples from a random process which is spatially correlated with isotropic correlation[2]:

$$C_{x_1,x_2} = E[[R^2_{x_1} - \mu_{R_u^2}][R^2_{x_2} - \mu_{R_u^2}]] = f\{\|x_1, x_2\|\} \quad (4)$$

where C_{x_1,x_2} is the covariance between two locations $x_1 \in \Re^2$ and $x_2 \in \Re^2$, $E[\cdot]$ denotes the expectation operator, $\mu_{R_u^2}$ is the average value of the random process (to be estimated), f denotes a function and $\|x_1, x_2\|$ is the physical (Euclidean) distance between x_1 and x_2. In addition, it is assumed that the process tends to zero; i.e. that far from a camera no information is known. A Matérn Kernel[3]

[2]That is, the correlation between the observed R_u^2 values at two points depends on distance alone and not the direction.

[3]The Matérn Kernel is a commonly used kernel for interpolation and is similar to a Gaussian kernel in that it dies away monotonically with a decay rate that can be adjusted using two parameters;

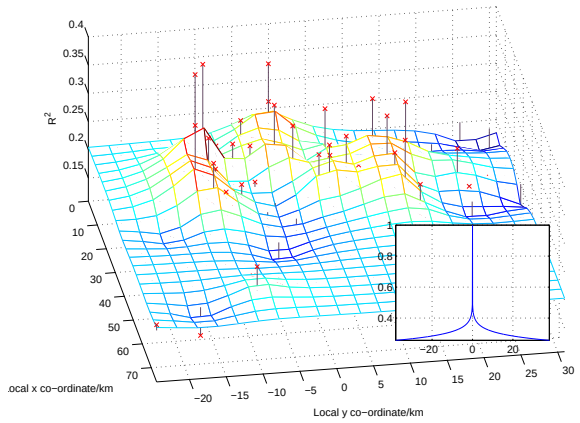

Figure 3: A GP model for distribution of R_u^2. (Sydney all hours; the Matérn Kernel is shown in the inset, $\{\nu = 0.02, \theta = 21\}$)

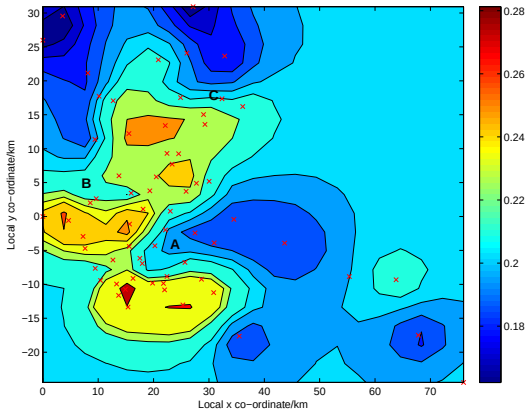

Figure 4: A contour plot of the GP distribution of R_u^2 with locations of possible interest labeled A,B and C.(Sydney all hours)

is used to represent $f\{\|x_1, x_2\|\}$ and consists of two parameters, θ and ν (to be estimated from the data), with θ controlling the scale and ν the shape of the kernel (see [8] for a description of the effect of varying these parameters). A full description of a GP is beyond the scope of this paper; what is important in the current context is the width and shape of the optimized kernel and of course the interpolation itself.

Figure 3 shows the interpolation of R_u^2 across the sample domain. Overall, the interpolated values are quite high in the city centre and drift towards the process mean ($\hat{\mu}_{R_u^2} = 0.2$)) quite quickly. In addition the inset in Figure 3 shows the Matérn Kernel which is estimated from the data ($\{\nu = 0.02, \theta = 21\}$). This kernel is quite 'peaked' such that the effect that a sample (i.e. an R_u^2 value) has on another location is mostly local. However, the kernel is quite wide at the bottom showing that a weighting of approximately 0.2 - 0.4 is still felt up to approximately ten kilometers away. Figure 4 shows the same information as a contour plot; this is more suited to identifying interesting regions.

In Figure 4 there are three regions labeled A, B and C, in which the interpolation appears to be lower than expected (these were selected subjectively). An inspection of the street map around C shows its proximity to a major Highway (M7). We suspect that region C contributes to the traffic observed on the M7. Only three cameras are deployed throughout the M7 and none of them are in C's vicinity. It is possible that inclusion of few cameras around this region would increase the information available to the network. B is located in a park area covering 32 km^2 called Georges national park (which explains the low coverage) while A is located in central Sydney on a major highway, again the M7 that has low coverage. A similar case is observed in all the networks.

4.2 Disaggregation and forecast analysis

The aim of this section is to demonstrate a set of forecasts from the eVAR(p) models. The data is first disaggregated into three regions; the morning (7am to 10am), noon (10:15am to 3pm) and the afternoon (3:15 to 7 pm). In addition, within each dataset 2/3 and 1/3 is kept back as an *out of sample* test set. Results are reported as a prediction mean squared error (PMSE) at various forecast horizons from 1 to 6 steps ahead (1 step=5 minutes). The data is normalized to between 0 and 1, where 1 is the maximum traffic density seen at a junction. This is to facilitate the reader to easily gauge the amplitude of the forecast errors.

Figure 5 presents a panel in which a single 6 step-ahead forecast

an excellent description may be found in [8].

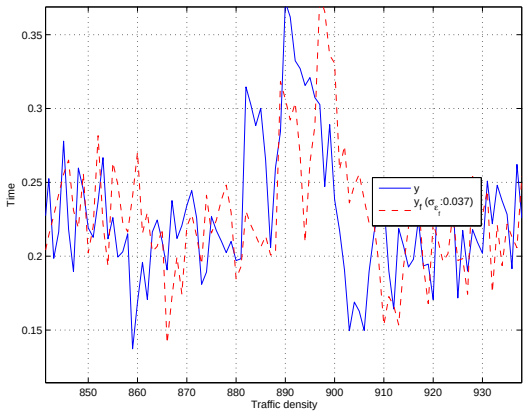

Figure 6: A sample forecast versus the actual for the traffic density at camera 16. (Sydney morning dataset, 6-step ahead forecast or 30 minutes).

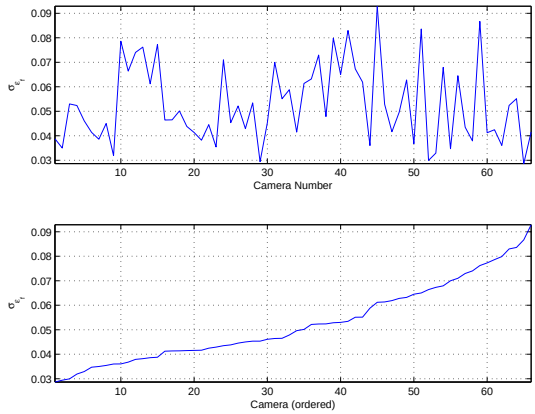

Figure 7: The prediction MSE for Sydney morning dataset (6-steps ahead or 30 minutes).

is shown for the network as whole. The software shows how the congestion in the future can be identified using the current state of the network. The regression in the bottom left hand panel however shows that there is still a significant forecast error in some cases.

Figure 6 shows a sample forecast of the traffic at camera 16 (a city center camera) for Sydney over the entire duration of the dataset. The forecast horizon here is 6 steps ahead. As can be

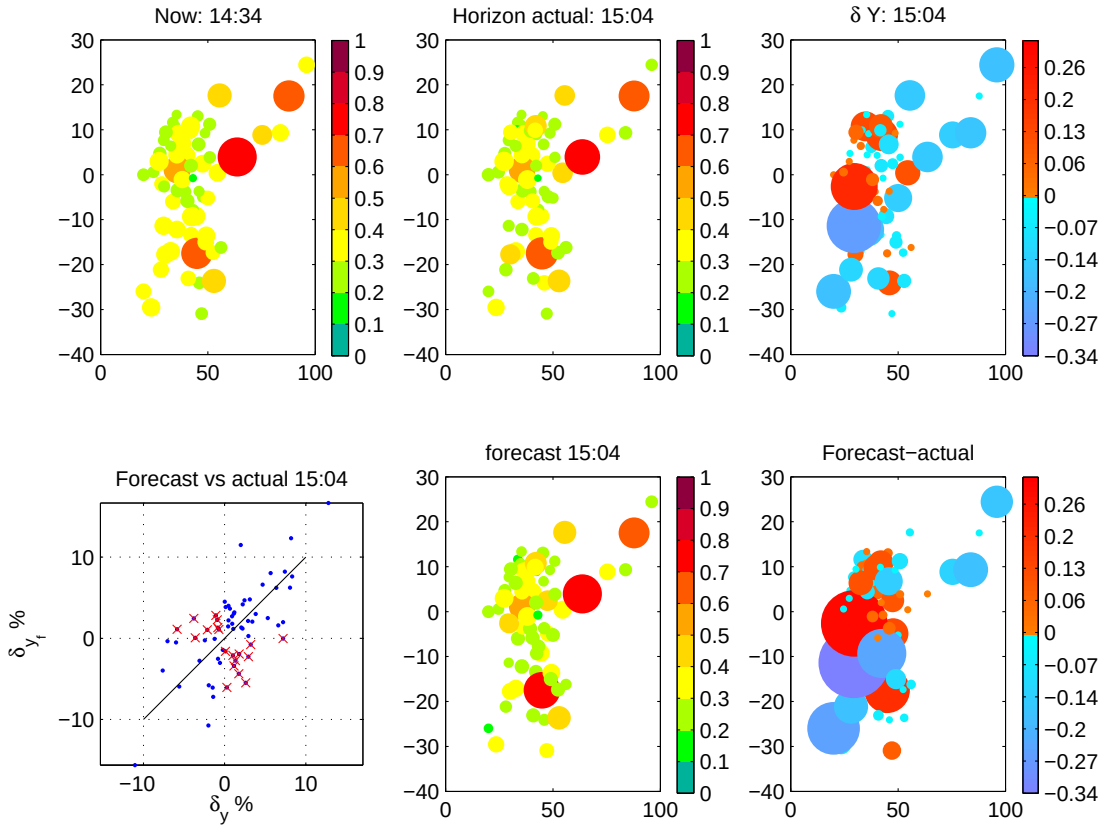

Figure 5: The panels shows the state of the network at 14:34 and 30 minutes later at 15:04. The top row shows actual states (a) and (b) while (c) shows the change between 14:34 and 15:04. The size and color of the circles reflects the traffic density as a % of the maximum and are located at the local X-Y coordinates of the junctions. The bottom panels show the forecasts; first the errors against the actual (d), then forecast itself (e) and then forecast change (f). (Sydney afternoon dataset, 6-step ahead forecast or 30 minutes).

seen the forecasts look reasonable but no more can be said without looking at the residual statistics.

Figure 7 shows the square root of the PMSE for the morning data set in Sydney. There are 66 camera's and so 66 values to report which are sorted lexically (i.e. by camera number) and also in ascending order (bottom panel). These vary from 0.03 to 0.09 meaning that the expected value of an error is about 3-9%.[4]

As the forecast horizon increases the forecast error is expected to increase as is shown in Figure 8. There is no discernible difference between the three times periods.

4.3 Comparison of global city networks

This section compares the topological characteristics of Granger networks for the six cities involved in this work. As networks are weighted directed graphs, weighted metrics are required for the comparison (see [1] for definitions of the metrics used below).

The average weight of a link for each city, $\bar{w}$, shows a marked difference in each network. While Beaufort, and to a lesser degree WDC, and London tend to have links that convey more information The networks for Sydney, Connecticut, and London exhibit links do not convey much information suggesting either these networks can benefit from more cameras or that the information arriving at a junction is incoming from many directions (a typical case in a dense city environment). The average weighted clustering coefficient, $\bar{\gamma_k}$, which measures the local cohesiveness, confirms this hypothesis

[4]The standard error and the expected error are not identical.

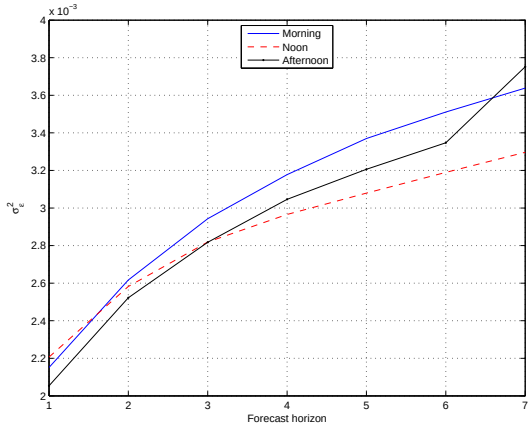

Figure 8: The increase in forecast error with forecast horizon for Sydney (disaggregated by morning, noon and afternoon; averaged across all camera's).

with Beaufort and WDC again scoring highest.

For the average nearest neighbors degree we find that $\bar{k}^w_{nn} < \bar{k}_{nn}$, indicating that edges with larger weights tend not to point to neighbors with larger degree; the implication being that on average the information flowing into a junction need not come from high degree neighbors. In addition, as $\bar{k}^w_{nn}$ and $\bar{k}_{nn}$ are both positive the networks are found to be dis-assortative; high degree nodes tend to

Table 2: Summary of topological characteristics for the GN's.

City	$\bar{w}$	$\bar{\gamma_k}$	k_{nn}	k^w_{nn}
Sydney	0.10	0.0015	23.19	22.18
Beaufort	0.77	0.0132	48.60	42.10
Connecticut	0.02	0.0019	1.37	1.18
GTA	0.08	0.0015	3.46	2.81
London	0.16	0.0013	34.34	28.36
WDC	0.24	0.0033	29.75	16.67

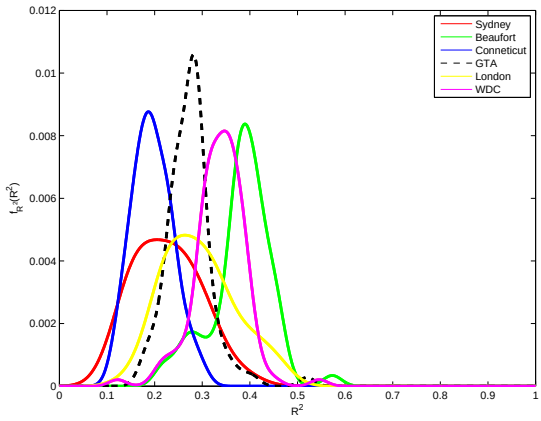

Figure 9: A comparison of the distributions of R^2_u values for the 6 cities.

be connected to low degree nodes indicating that important junctions tend to be meeting points for many smaller junctions. It is notable that in the case of GTA and Connecticut this is not significant as these networks are composed of mainly motorways in which each junction is mainly a link in a long chain along the network. Figure 9 compares the distribution of R^2_u values at each node/junction. Again, it is seen that for WDC and Beaufort that this distribution is higher with Sydney and Connecticut have a distribution centered around the lowest means. A detailed technical report of this work is available here [11].

5. CONCLUSION

Vehicular traffic congestion is becoming a critical problem in major cities throughout the world. In order to address that problem, we have proposed a framework for the systematic monitoring, measurement, analysis and prediction of urban traffic density at a global scale. The proposed framework is highly scalable and can be used for any city that utilizes online traffic web cameras for traffic analysis. So far, we have analyzed five cities and a state to examine the cause and effect of traffic causality that is contributing to widespread congestion. The time series analysis in this paper threw up some compelling results. Firstly, the location of traffic cameras may not be optimal and there are regions in all the networks we surveyed in which there a lack of camera coverage, which effects the information in the network as a whole. One possibility is the introduction of new cameras in those areas or possibly the integration of information from alternative sources such as smart phone data (or even to alert smart phones to the use of their data at a particular location). While only three cases were focused on in the analysis in Section 4.1, further investigation is envisaged. The disaggregated forecasts produced show that reasonable forecasts of the city traffic may be created up to 30 minutes ahead. The PMSE of these forecasts varied widely (from 0.03 to 0.09) and this in itself allows the city operators to see which cameras are not providing adequate information. We envisage that these forecasts can be improved in future analysis with the aid of a Bayesian framework especially in taking into account invalid camera readings that degrade the results.

The Granger networks themselves showed a high degree of information exchange between the nodes, while still presenting a sparse representation of the network as a whole. This sparse representation is important, as it reveals the main information pathways in the network, which apart from being statistically important also provides a succinct summary of the network to traffic managers. While comparing these networks across the different cities, two types of cities emerged; those based on mainly motorway based road networks, which exhibit a higher degree of cohesion, compared to those based on a city street type topology. The differing clustering coefficients in each network present an interesting dilemma; given finite resources it is best to pool cameras such that they reinforce each other's information (as in the case of Beaufort) or spread the cameras evenly around the city (as in the case of London).

6. REFERENCES

[1] A. Barrat, M. Barthélemy, R. Pastor-Satorras, and A. Vespignani. The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences of the United States of America*, 101:3747–3752, 2004.

[2] Adam B. Barrett, Lionel Barnett, and Anil K. Seth. Multivariate Granger causality and generalized variance. *Physical Review E*, 81(4):041907+, apr 2010.

[3] Y. Benezeth, P.M. Jodoin, B. Emile, H. Laurent, and C. Rosenberger. Review and evaluation of commonly-implemented background subtraction algorithms. In *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, pages 1 –4, dec. 2008.

[4] C.G. Carlson and D.E. Clay. The earth model âĂŞ calculating ïňĄeld size and distances between points using gps coordinates. *Site SpeciïňĄc Management Guidelines series-11, Potash and Phosphate Institute (PPI)*, 1999.

[5] David A. Forsyth and Jean Ponce. *Computer Vision: A Modern Approach*. Prentice Hall Professional Technical Reference, 2002.

[6] C. W. J. Granger. Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica*, 37(3):424–438, 1969.

[7] R. Johnson and D. Wichem. *Applied multivariate statistical analysis*. 1982.

[8] T. J. Santner, B. Williams, and W. Notz. *The Design and Analysis of Computer Experiments*. Springer-Verlag, 2003.

[9] Y. Sheikh and M. Shah. Bayesian modeling of dynamic scenes for object detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27:1778–1792, 2005.

[10] C. Stauffer and W.E.L. Grimson. Adaptive background mixture models for real-time tracking. In *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, volume 2, 1999.

[11] Gautam Thakur and Pan Hui. Planet Sensing: Modeling vehicular mobility on a planet-scale. `https://sites.google.com/site/planetsensing/`, 2013.