Predicting Normal and Anomalous Urban Traffic with Vectorial Genetic Programming and Transfer Learning*

John Rego Hamilton[1], Anikó Ekárt[1], and Alina Patelli[1]

Aston University, Birmingham, UK
Department of Computer Science
{j.connorregohamilton1, a.ekart, a.patelli2}@aston.ac.uk

Abstract. The robust and reliable prediction of urban traffic provides a pathway to reducing pollution, increasing road safety and minimising infrastructure costs. The data driven modelling of vehicle flow through major cities is an inherently complex task, given the intricate topology of real life road networks, the dynamic nature of urban traffic, often disrupted by construction work and large-scale social events, and the various failures of sensing equipment, leading to discontinuous and noisy readings. It thus becomes necessary to look beyond traditional optimisation approaches and consider evolutionary methods, such as Genetic Programming (GP). We investigate the quality of GP traffic models, under both normal and anomalous conditions (such as major sporting events), at two levels: *spatial*, where we enhance standard GP with Transfer Learning (TL) and diversity control in order to learn traffic patterns from areas neighbouring the one where a prediction is needed, and *temporal*. In the latter case, we propose two implementations of GP with TL: one that employs a lag operator to skip over a configurable number of anomalous traffic readings during training and one that leverages Vectorial GP, particularly its linear algebra operators, to smooth out the effect of anomalous data samples on model prediction quality. A thorough experimental investigation conducted on central Birmingham traffic readings collected before and during the 2022 Commonwealth Games demonstrates our models' usefulness in a variety of real-life scenarios.

Keywords: Nature-inspired computing for sustainability · Resilient urban development · AI-driven decision support systems · Intelligent and safe transportation · Urban traffic prediction.

1 Introduction

Designing, building and maintaining an accessible, safe and cost-effective urban traffic infrastructure are key milestones on the road to meeting the UN's Sustainable Development goals[1]. The complex decision making involved, particularly at

* This work is supported by the Engineering and Physical Sciences Research Council (Grant Number EP/R512989/1).

[1] Available at https://sdgs.un.org/goals, in particular, goal 11, action 11.2.

the level of local administration, can be significantly streamlined when robust and reliable predictions of future traffic through key areas of the urban road network are made available by computational means rather than by exclusively relying on human expertise.

To that end, data-driven, Artificial Intelligence methods are widely recognised ways of producing computationally efficient models of dynamic, large-scale transport networks, such as modern day cities [3]. Within that category, evolutionary optimisation techniques, Genetic Programming (GP) in particular, offer an attractive solution to the urban traffic and modelling prediction problem. When combined with Transfer Learning (TL), traditional GP has been shown to predict vehicle flow with competitive accuracy, through urban junctions where traffic readings are either unavailable or unreliable [7, 12]. Transferring traffic patterns learnt on areas that are topologically similar to, yet geographically distinct from, the region where a traffic model is needed but difficult or impossible to obtain practically is particularly important when vehicle flow is affected by atypical events (accidents, major sporting events, construction work, etc.). Besides accurately capturing vehicle flow under typical and atypical conditions by *spatial* means, as afforded by TL, traffic anomalies can also be mitigated on a *temporal* level. One way of achieving that is by equipping TL-enhanced prediction algorithms with a lag operator, making it possible to train traffic models on a dynamically adjusted number of past readings, counting backwards from an arbitrarily chosen sample in the training data set. We refer to this approach as *skipping*. Alternatively, the effect of anomalous data on the traffic models' prediction accuracy can be attenuated by applying adequate linear algebra operators that aggregate training samples. We refer to this approach as *smoothing*.

Our goal is to investigate the impact of spatial (TL) and temporal (skipping and smoothing) anomaly handling on the prediction accuracy of GP traffic models. We achieve this by proposing a rigorously validated, robust and competitively accurate approach to the urban traffic modelling and prediction problem. This approach rests on two original contributions:

1. A novel traffic prediction algorithm, $\overline{\text{GENTLER}}$, equipped with TL and smoothing, that produces models transferable to areas different to those where initial training occurs, without compromising prediction accuracy, and is tolerant to anomalous training samples.
2. A rigorous experimental investigation of $\overline{\text{GENTLER}}$'s prediction capability under both normal and anomalous traffic conditions. To enable that, $\overline{\text{GENTLER}}$ is deployed on several areas of the Birmingham city centre and trained on traffic data collected both before and during the 2022 Commonwealth Games. GENTLER [12], a precursor algorithm featuring TL and skipping, is used as a benchmark. The full factorial experiment presented in section 5 demonstrates the competitiveness of the two algorithms in a variety of traffic conditions.

A summary of relevant work conducted in the traffic modelling and prediction domain is given in section 2. Section 3 presents a simple yet illuminating example that highlights the value adding potential of incorporating vectorial elements into our existing algorithm. The paper's two original contributions are the topics of sections 4 and 5. Section 6 outlines our conclusions.

2 Background

Traffic prediction by means of traditional time series modelling and machine learning has been receiving significant attention [4, 11]. Traditional time series modelling, based on the auto-regressive integrated moving average (ARIMA) method and its variants, is well-suited to cases where short-term predictions (a few hours into the future) are sufficient. Conversely, machine learning approaches such as supervised learning underpinned by deep neural networks [4, 11, 1] and hybrid models [10], have proven successful for short, medium and longer term predictions (up to one week). Yet, they typically require extensive training [1]. This downside can be mitigated by turning to evolutionary algorithms: versions specifically tailored to tackle complex problems related to urban transport have been successfully applied to automate traffic signal management [5, 13], design bus route networks with a reduced environmental impact [8], locate electric vehicle charging stations [6], etc.

Another promising approach to traffic prediction is related to transfer learning, a technique that enables transferring models trained on data collected from one area (source) to a neighbouring one (target), where traffic readings may not be available. Li et al. report competitive results when applying transfer learning to highway traffic [9]; their findings are predicated on source and target traffic exhibiting similar patterns, which is less likely when it comes to the complex road networks within major cities. Modelling traffic through those networks requires a more sophisticated version of transfer learning: one example is GENetic programming with Transfer LEarning (GENTLE) [7], an algorithm that produces robust models of urban traffic, transferable from source to target junctions with no need for additional training on the latter. A lag operator ensures that historical traffic flow values can be taken into account in a computationally efficient way, without inflating the terminal set with lagged input terms. GENTLER [12] is an extension of GENTLE whereby, in the case of transfer from multiple sources, the models copied over from one source junction to the next are supplemented with a configurable amount of random trees, in an attempt to achieve a better exploration-exploitation trade-off. When its parameters are optimally configured, GENTLER yields models of significantly better accuracy than GENTLE.

GENTLE and GENTLER use traditional scalar-based symbolic regression. By contrast, Vectorial Genetic Programming (VE-GP), a recent approach specifically designed to model time series [2], utilises vector terminals and vectorial functions, both aggregate and cumulative. VE-GP is applied on a real-life physiological time series prediction problem. When the specifics of the healthcare problem domain are abstracted out, sufficient similarities to traffic prediction remain to justify incorporating certain VE-GP components into GENTLER.

3 Motivating Scenario

The example analysed in this section is a simplified version of a real-world traffic modelling problem. It uses a basic road layout to explain how anomaly handling mechanisms operate across space (TL) and time (lag operator enabling skipping and linear algebra operators facilitating smoothing) in order to produce accurate traffic predictions that hold under both typical and anomalous conditions.

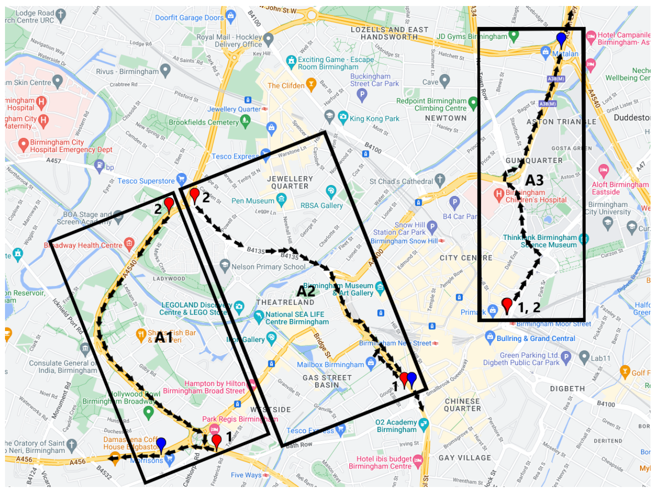

Fig. 1: Three central Birmingham areas monitored before and during the 2022 Commonwealth Games: sensors 1 and 2 measure inflow traffic; the unlabeled sensor captures outflow traffic.

3.1 Traffic Anomaly Handling Over Time: Skipping and Smoothing

Let us consider a topology with two inflow lanes, x_0 and x_1, and one outflow lane, y. This could be assimilated to any of the three areas shown in Fig. 1, where sensors 1 and 2 monitor inflow, whilst the remaining sensor records outflow, every Δt minutes, over a time interval of length L. Assuming $\Delta t = 15$ and $L = 60$, the traffic readings are stored in matrix R.

$$R = \begin{bmatrix} x_0(t_0) & x_0(t_1) & x_0(t_2) & x_0(t_3) \\ x_1(t_0) & x_1(t_1) & x_1(t_2) & x_1(t_3) \\ y(t_0) & y(t_1) & y(t_2) & y(t_3) \end{bmatrix} \quad (1)$$

The number of rows in matrix R is given by the number of monitored lanes, whereas the number of columns is equal to the number of traffic readings collected throughout the monitoring interval L. Symbol $x_0(t_0)$ represents the number of vehicles passing through input lane x_0 at the beginning of the monitoring interval, $x_1(t_1)$ stands for the number of vehicles recorded Δt minutes later through input lane x_1, whilst $y(t_3)$ marks the number of vehicles exiting through outflow lane y at the end of the monitoring interval. The other symbols in matrix R are to be interpreted in a similar fashion.

Model Representation GENTLER builds models of the traffic flowing through lane y by combining scalar terminal nodes from set T and arithmetic operator nodes from set O, both given in equation (2). Operator / stands for protected division, which returns 1 whenever the right hand side operand is 0. Set O also includes a unary *lag* operator that returns the value of its input delayed by one sampling interval, Δt.

$$T = \{x_0, x_1, y\} \quad O = \{+, -, \times, /, lag\} \tag{2}$$

In the case of $\overline{\text{GENTLER}}$, traffic models are represented as combinations of the vector terminals and linear algebra operators given in equation (3). The elements of set $\overline{T}$ correspond to the rows in matrix R, in the order given by their indices. The first four elements of set $\overline{O}$ are unary aggregate operators that output, respectively, the mean, sum, maximum and minimum of their input vectors. The following four elements perform the same operations cumulatively, whereas the final four represent the traditional vector addition, subtraction, element-wise product and element-wise protected division. The complete definitions (with examples) of all operators in set $\overline{O}$ are available in [2].

$$\begin{aligned} \overline{T} = \{\overline{x_0}, \overline{x_1}, \overline{y}\} \\ \overline{O} = \{\mathbf{V\_mean}, \mathbf{V\_sum}, \mathbf{V\_max}, \mathbf{V\_min}, \\ \mathbf{C\_mean}, \mathbf{C\_sum}, \mathbf{C\_max}, \mathbf{C\_min}, \\ \mathbf{VsumW}, \mathbf{V\_W}, \mathbf{VprW}, \mathbf{VdivW}\} \end{aligned} \tag{3}$$

Let us consider the tree-like traffic models M and $\overline{M}$ in Fig 2. The output of the GENTLER tree M represents the predicted vehicle flow, $\hat{y}(t)$, through the output lane.

$$\hat{y}(t) = (x_0(t) + x_1(t-2))/c, t \in [t_0, \ldots, t_3], \quad c = const \tag{4}$$

The $\overline{\text{GENTLER}}$ model $\overline{M}$ outputs a vector, the elements of which are predicted outflow values at each time instant in the monitoring interval.

$$\begin{aligned} [\hat{y}(t_0), \hat{y}(t_1), \hat{y}(t_2), \hat{y}(t_3)] = 1/c \times [x_0(t_0), x_0(t_1), x_0(t_2), x_0(t_3)] \\ +1/4c \times (4x_1(t_0) + 3x_1(t_1) + 2x_1(t_2) + x_1(t_3)), \quad c = const \end{aligned} \tag{5}$$

Model Evaluation In order to determine the prediction accuracy (i.e., the fitness) of the traffic models, their outputs are compared against the readings in the third row of matrix R. Assuming that only the first three columns in

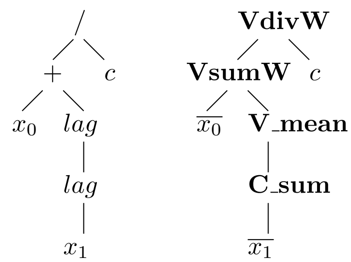

Fig. 2: Model representation in GENTLER (tree M, on the left) and in $\overline{\text{GENTLER}}$ (tree $\overline{M}$, on the right).

matrix R are used for training, the fitness of model M is $\sqrt{\frac{1}{3}\sum_{i=0}^{2}(\hat{y}(t_i) - y(t_i))^2}$, a scalar. The values $\hat{y}(t), t \in [t_0, \dots t_2]$ featured in equation (6) are those given in equation (4).

$$\begin{aligned} 3 \times (fitness(M))^2 = \sum_{i=0}^{2}(\hat{y}(t_i) - y(t_i))^2 = & \\ (x_0(t_0)/c - y(t_0))^2 & \quad : t_0 \\ +(x_0(t_1)/c - y(t_1))^2 & \quad : t_1 \\ +((x_0(t_2) + x_1(t_0))/c - y(t_2))^2 & \quad : t_2 \end{aligned} \tag{6}$$

By contrast, the fitness of $\overline{M}$ is a vector, as shown in equation (7) where values $\hat{y}(t), t \in [t_0, \dots t_2]$ are given in equation (5). However, $\overline{\text{GENTLER}}$ calculates the root mean squared error (RMSE) of candidate models by squaring and averaging the elements of the fitness vector, leading to a scalar value much like in the case of GENTLER.

$$fitness(\overline{M}) = [\hat{y}(t_0) - y(t_0), \hat{y}(t_1) - y(t_1), \hat{y}(t_2) - y(t_2)] \tag{7}$$

The fitness of models M and $\overline{M}$ reveals the different ways in which GENTLER and $\overline{\text{GENTLER}}$ are equipped to handle anomalies in the training data. In cases where traffic is affected by road construction, accidents, popular sporting events, etc., GENTLER manages the disruption by skipping over (potentially) anomalous readings: as shown in equation (6), the prediction generated by model M is unaffected by samples $x_1(t_2)$ and $x_1(t_1)$. $\overline{\text{GENTLER}}$ employs a different mechanism: the unary operators in set $\overline{O}$ have a smoothing effect in that data samples are aggregated and cumulated (see second line of equation (5)) which reduces the impact of anomalous readings on the prediction accuracy.

3.2 Traffic Anomaly Handling Across Space: Transfer Learning

In a real-world setting, it is often the case that some areas of a city's road network, say A1 and A2 in Fig. 1, are reliably monitored, whilst others are not (the equipment sensing traffic through A3 is faulty, has been incorrectly installed or is missing altogether). Regardless, an accurate model of A3 traffic is needed all the same (e.g., to help decision makers determine the layout of new roads to be built in the area in order to streamline vehicle flow, thus reducing delays, accident rates and pollution). In the absence of native data, A3 traffic models need to be trained on neighbouring junctions A1 and A2 then transferred over to A3 (see [7] for a detailed explanation of the transfer learning process). For the transfer to be successful, the exogenous model (trained on A1 and A2) would need to predict traffic as accurately as an indigenous one (trained on A3, assuming data were available to make that possible). The experimental investigation documented in section 5 demonstrates that this is indeed the case: exogenous models produced by both GENTLER and $\overline{\text{GENTLER}}$, on various combinations of typical and anomalous training data sets predict traffic with an accuracy that is comparable to that of indigenous models.

4 $\overline{\textbf{GENTLER}}$ Explained

We propose $\overline{\text{GENTLER}}$, vectorial GENetic Programming with Transfer LEarning and Randomisation, that features the following components:

1. classic Symbolic Regression enhanced with transfer learning and bespoke exploration-exploitation tuning (also found in GENTLER [7]);
2. vectorial representation (i.e., vector terminals and linear algebra operators enabling smoothing); and
3. vectorial fitness, customised with a penalty mechanism designed to punish models with scalar outputs (i.e., vectors filled with copies of the same element), thus increasing the selection pressure in favour of trees with (true) vectorial outputs, which are more likely to be accurate traffic predictors.

Algorithm 1 illustrates how GENTLER and $\overline{\text{GENTLER}}$ evolve traffic models. Each row in matrix *models* (line 1) represents the population at a given generation: for GENTLER, the candidate models will be similar to M in Fig. 2 (note the *skipping*-enabling *lag* operator), whereas for $\overline{\text{GENTLER}}$, the trees will look like $\overline{M}$. Line 2 initialises the *estimator*, i.e., the object wrapper for *gp learn*'s $SymbolicTransformer^2$, with all expected evolutionary parameters (see Table 2 caption). On line 3 the *estimator* evolves the population of candidate models, for G_1 generations on training data collected from the first source (src1, which can represent any area in Fig. 1).

The fit method in the *gp learn* library implements the classic GP loop; one of the steps involved is fitness calculation. This is presented in algorithm 2, which runs for every element in $models[G_i]$, where G_i is the current generation. Lines 2 through 5 illustrate the classic RMSE calculation, over all *trn* samples in the training data set (containing readings collected from src1 if $G_i \leq G_1$ or from the second source, src2, if $G_i > G_1$). This is where *smoothing* occurs, as an effect

[2] https://gplearn.readthedocs.io/en/stable/reference.html#symbolic-transformer

Algorithm 1 GENTLER and $\overline{\text{GENTLER}}$ essential logic

$models \leftarrow [\,][\,]$
$estimator.init()$
$models[G_1] \leftarrow estimator.fit(getTrainData(\text{src1}))$

################ **Transfer Learning with Randomisation**
if $mix == 0$ **then**
  $models[G_1 + 1] \leftarrow makeRndModels(N)$
else
  if $mix == 1$ **then**
    $models[G_1 + 1] \leftarrow models[G_1]$
  else
    $models[G_1 + 1] \leftarrow getBest(models[G_1], hof, cmp, mix \times N)$
      $\cup\ makeRndModels((1 - mix) \times N)$
  end if
end if####################################

$models[G_2] \leftarrow estimator.fit(getTrainData(src2), models[G_1 + 1])$
return $getBest(models[G_2])$

of the cumulative and aggregation operator featured in $\hat{y}(t)$. The result, $rmse$, represents the fitness of the model. Fitness calculation in $\overline{\text{GENTLER}}$ implies an additional step (lines 6 through 9): should the output of the current model $\overline{M}$, i.e., the vector containing predicted values $\hat{y}$ for each of the time instants in the training set, contain identical elements (indicating a scalar output), chances are that model will not yield a competitive prediction accuracy, therefore its fitness is increased by a factor of 100 (*pen* on line 8).

Once the population at generation G_1 becomes available, algorithm 1 continues to the transfer learning stage. This occurs when traffic is to be predicted through a (target or destination) area where historical vehicle flow data are not available (because sensing equipment is absent, faulty or incorrectly installed, or because the target has not yet been built). In these cases, training data have to be collected from adjacent locations (e.g., neighbouring junctions where reliable traffic readings are available), called sources. For illustration, suppose that any one of the areas in Fig. 1 is the destination. Any combination of the remaining two areas (sources) can be used to train the traffic models aimed to predict traffic through the destination; GENTLER assumes two such sources, src1 and src2. In that context, transfer learning consists in training the candidate models on data collected from src1 for G_1 generations (from 1 to G_1) and then on data collected from src2 for $G_2 - G_1$ generations (from $G_1 + 1$ to G_2). Thus, the traffic patterns learnt from src1 data and further refined on src2 data are transferred over to the destination to produce predictions (since all possible destinations shown in Fig. 1 already exist and are monitored, we use those data to validate the transferred model). The selection process of the trees to be transferred between sources (from generation G_1 to $G_1 + 1$) is controlled by three parameters:

Algorithm 2 GENTLER and $\overline{\text{GENTLER}}$ fitness calculation

$rmse = 0$
for t **in** $[t_0, \ldots, t_{trn}]$ **do**
  $rmse \leftarrow rmse + (\hat{y}(t) - y(t))^2$
end for
$rmse \leftarrow \sqrt{rmse/(trn+1)}$

################### **penalty ($\overline{\textbf{GENTLER}}$ only)**
if $\hat{y}(t)$ *identical*, $\forall t$ **in** $[t_0, \ldots, t_{trn}]$ **then**
  $rmse \leftarrow rmse \times pen$
end if################################

return $rmse$

- hof the number of trees (in order of fitness, starting with the most accurate) to include in the **h**all **o**f **f**ame and consider for transfer;
- cmp the number of least correlated **comp**onents (trees) within the hall of fame to be transferred to src2; and
- mix a real number between 0 and 1 representing the proportion of trees transferred from src1 relative to random trees.

Lines 5 through 13 show the three possible ways of putting together the N models in the population at generation $G_1 + 1$, depending on the mix value:

- All models are random (lines 5, 6): all traffic patterns learnt on the first source are lost; training starts afresh on src2 (equivalent to pure exploration).
- No models are random (lines 8, 9): all traffic patterns learnt on the first source are transferred over to src2 where training resumes (equivalent to pure exploitation).
- Some models are random (lines 11, 12): $mix \times N$ of the most accurate, least correlated cmp trees trained on src1 are transferred over to src2, and the remaining positions are filled with randomly generated models (balance between exploitation and exploration).

The configurable amount of random trees injected at generation $G_1 + 1$ dynamically adjusts the exploration-exploitation ratio: if optimally chosen, the mix value will make it possible to leverage the traffic patterns learnt from the first source as well as maintain a healthy amount of genetic diversity.

Once the transfer from the first to the second source is complete, the fit function is called again, on line 14, to evolve the models at generation $G_1 + 1$ for $G_2 - G_1$ additional generations, on src2 training data. Method $getBest$ (line 15) returns the most accurate model at generation G_2; its output can then be used to predict traffic through the destination (and be validated, if destination traffic readings are available).

5 Experimental Investigation

5.1 Sensor Selection and Traffic Readings Analysis

Transport for West Midland (TfWM) has 178 sensors installed across Birmingham[3]. Out of the ones covering the city centre, we eliminated those with null outputs (most likely due to faults) and those that recorded significantly fewer than 96 samples a day (i.e., captured data less frequently than every 15 min). The remaining sensors monitor vehicle flow through areas A1, A2 and A3 in Fig. 1. A stream of arrows indicates the direction of traffic: inbound vehicles are counted by sensors 1 and 2, whereas outbound ones are captured by the third sensor (unlabeled) in each of the considered areas.

The nine sensors collected data before and during the 2022 Commonwealth Games that took place in Birmingham: readings captured between the 25th of April and the 29th of May are taken to represent traffic under normal conditions, whereas values recorded between the 25th of July and the 28th of August are indicative of traffic under anomalous conditions (i.e., disrupted by restrictions put in place to accommodate for the various sporting events). We conducted full factorial experiments with GENTLER and $\overline{\text{GENTLER}}$, considering all possible combinations of source and destination areas, with relevant models trained and validated (60-40 data split) under normal and anomalous conditions; the results are presented in Table 2.

The TfWM traffic monitoring exercise is young: measurements are of poor quality compared to other major EU municipalities[4]. The two principal challenges are unevenness (standard deviations reported in Table 1 are very high compared to averages, particularly in A2) and scarceness (most acutely in A3, where there are no sensors monitoring traffic through the main road, which is intensely used by motorists going round the Clean Air Zone to avoid fees). Although these data quality related problems are bound to negatively impact the accuracy of the traffic models, we chose not to eliminate outliers during preprocessing, as there is no way to determine whether they are indicative of sensor malfunctions or of spikes in real traffic; their deletion would either deplete the data pool or overly-sanitise it, making it difficult for the ensuing experiments to authentically highlight the strengths of GENTLER and $\overline{\text{GENTLER}}$. Instead, we chose to use the native models (trained and validated on data from the same area) as a baseline, in order to ascertain the relative quality of non-native ones (trained and validated on data collected from different areas), thus showcasing the effectiveness of transfer learning with smoothing and skipping under realistic (in this case, sub-optimal) conditions. We argue that this is a valid experimental approach, as transfer learning is meant to provide models of traffic through areas where native data are not available (roads that are not monitored or yet to be built). For comprehensive evidence that the two algorithms work efficiently on high quality data, the interested reader is referred to [7, 12].

[3] https://data-tfwm.opendata.arcgis.com
[4] Such as Darmstadt: https://www.ui.city/en/.

Table 1: Output sensor data used to validate all traffic models: relevant stats.

	Normal			Anomalous		
	A1	**A2**	**A3**	**A1**	**A2**	**A3**
Count	1344	1344	1344	1344	1344	1344
Mean	77.73	9.73	160.58	55.60	26.45	150.46
SD	43.00	16.08	83.43	46.28	18.92	80.16
Min	0	0	6	0	0	9
Max	194	100	383	209	173	337

Table 2: GENTLER and $\overline{\text{GENTLER}}$ model accuracy: hof, cmp, mix = (717, 68, 0.28); for all other evolutionary parameters see [12]; RMSE is lowest over 30 runs; % RMSE is the ration of RMSE to the mean of the corresponding validation dataset in Table 1.

		NN		NA		AN		AA	
	Combination	**RMSE**	**% RMSE**	**RMSE**	**% RMSE**	**RMSE**	**% RMSE**	**RMSE**	**% RMSE**
GENTLER	A1 -> A1	20.19	25.97	15.59	28.04	20.41	26.26	15.51	27.89
	A2 -> A1	20.58	26.48	16.85	30.31	20.73	26.67	16.92	30.42
	A3 -> A1	415.62	534.67	372.11	669.20	523.20	673.06	240.99	433.40
	A2, A3 -> A1	42.04	54.08	38.35	68.97	21.12	27.17	17.33	31.17
	A3, A2 -> A1	29.78	38.31	24.09	43.33	30.33	39.02	24.57	44.20
	A2 -> A2	5.60	57.54	10.65	40.28	5.56	57.15	10.67	40.35
	A3 -> A2	216.92	2228.22	324.29	1225.66	326.51	3354.00	338.71	1280.16
	A1 -> A2	5.68	58.42	10.75	40.63	5.69	58.46	10.81	40.86
	A3, A1 -> A2	9.72	99.91	16.59	62.70	8.69	89.29	12.67	47.88
	A1, A3 -> A2	6.68	68.68	14.20	53.69	6.68	68.68	14.20	53.69
	A3 -> A3	48.33	30.10	57.46	38.19	50.54	31.47	66.09	43.92
	A2 -> A3	151.90	94.59	140.44	93.34	137.97	85.91	139.49	92.71
	A1 -> A3	150.89	93.96	138.20	91.85	127.33	79.29	126.50	84.07
	A2, A1 -> A3	117.73	73.31	129.76	86.24	124.36	77.44	127.36	84.64
	A1, A2 -> A3	115.82	72.12	127.11	84.48	121.37	75.58	128.44	85.36
$\overline{\text{GENTLER}}$	A1 -> A1	19.28	24.80	15.72	28.27	19.13	24.61	15.34	27.58
	A2 -> A1	21.01	27.03	16.53	29.73	20.69	26.62	16.37	29.45
	A3 -> A1	42.45	54.62	38.68	69.57	47.13	60.63	51.86	93.27
	A2, A3 -> A1	32.02	41.19	24.13	43.39	42.34	54.47	38.62	69.45
	A3, A2 -> A1	30.62	39.39	27.48	49.42	33.28	42.82	24.61	44.26
	A2 -> A2	5.65	58.11	10.40	39.33	5.67	58.31	10.23	38.68
	A3 -> A2	12.77	131.20	20.31	76.78	47.25	485.35	48.88	184.75
	A1 -> A2	5.73	58.87	10.98	41.52	5.65	58.10	10.57	39.95
	A3, A1 -> A2	7.13	73.30	14.34	54.23	12.12	124.59	15.27	57.74
	A1, A3 -> A2	9.85	101.19	13.61	51.43	13.14	135.00	20.26	76.58
	A3 -> A3	48.16	29.99	52.82	35.11	44.80	27.90	53.03	35.24
	A2 -> A3	94.47	58.83	91.18	60.60	94.47	58.83	91.18	60.60
	A1 -> A3	86.28	53.73	97.83	65.01	83.27	51.85	85.76	57.00
	A2, A1 -> A3	110.19	68.62	99.88	66.38	105.45	65.66	96.59	64.19
	A1, A2 -> A3	89.48	55.72	83.12	55.24	100.27	62.44	100.14	66.55

5.2 Results Discussion

We ran GENTLER and $\overline{\text{GENTLER}}$ on all source and destination combinations achievable considering the three areas in Fig. 1. The Combination column in Table 2 lists the source area(s), where training is performed, to the left of the arrow and the destination area, where validation takes place, to the right of the arrow. The experiments we conducted are of two types: homogeneous, where training and validation take place on data collected under matching conditions, and heterogeneous, where training is performed on normal readings and validation on abnormal ones and vice-versa. Experimental results in the former category are presented in the columns headed NN and AA, whereas findings in the latter category are reported in columns NA and AN (the first letter refers to training and the second to validation).

Modelling Traffic under Normal Conditions This segment of the experimental analysis is aimed at investigating the impact of transfer learning on model accuracy. It relies on data presented in the two columns of Table 2 under the NN heading.

With an average of 9.7 outgoing vehicles (see Table 1), the flow of traffic through the A2 area under normal conditions is severely limited. Output traffic in the A1 area is significantly higher (77.7 vehicles passing by the Morrisons sensor, on average), with A3 being by far the busiest of the three (160.6 average output flow through the Matalan monitoring point, more than double the volume of traffic out of A1). However vehicle flow through A3 is also the least even, with acute variations in the number of recorded outgoing vehicles, as indicated by a standard deviation of 83.4, twice as high as in the case of A1. These statistics suggest that models trained and validated on data collected from the A1 area are likely to be the most accurate of the three. This is confirmed by the experimental results: out of the three native models produced by GENTLER, the one trained and validated on A1 data has the highest accuracy (relative RMSE of 26%) and transfers over efficiently to A2: the model trained on A1 data and validated on A2 data is of comparable accuracy to the A2 native model (relative RMSE of 58.4% in the case of the former, compared to 57.5% in the case of the latter). The same applies to $\overline{\text{GENTLER}}$ models: 58.9% relative error when training on A1 and transferring to A2, compared to 58.11% achieved by the A2 native model.

The extreme variations in the A3 data make it very difficult to benefit from vehicle flow patterns learnt on the much better behaved A1 traffic: the GENTLER model trained on A1 data and validated on A3 data is more than three times less accurate than the native A3 model, a result that doesn't change when the source junction is A2. When it comes to $\overline{\text{GENTLER}}$, single source transfer learning works better: training models on A1 and, respectively, A2 data results in an accuracy loss (relative to the native A3 model) that is over one order of magnitude smaller than in the case of GENTLER.

The situation improves further when training occurs on A1 and A2 data, subsequently as opposed to separately, before the transfer to A3: the GENTLER models' relative RMSE drops from the 93% - 94% range (which is the case for single source transfer learning), to the 72% - 73% one (achieved for multiple source transfer learning). $\overline{\text{GENTLER}}$ models, single source and multiple source, with A3 as the destination, are all in the 53.7% - 58.8% band, with the exception of the one trained on A2 and then on A1, which is approximately 10% less accurate. As expected, A3 models transfer over very poorly in all single and multiple source combinations. Whilst GENTLER completely fails to apply patterns learnt during A3 training in order to model traffic through A1, $\overline{\text{GENTLER}}$ manages to produce a single source transfer model that is half as accurate as the A1 baseline (the former has a prediction error of 54.6% as opposed to 24.8% in the case of the latter). $\overline{\text{GENTLER}}$ continues to generate single source transfer learning models of superior accuracy to that of GENTLER ones when A3 is the source and A2 the destination. However, both algorithms yield multiple source transfer learning models that significantly outperform single source ones, for all three destinations considered in this study.

This part of the experimental analysis indicates that, under normal conditions (i.e., when vehicle flow through both source and destination junctions is relatively even, as is the case for A1 and A2 but not for A3), transfer learning models and native ones are comparably accurate regardless of overall traffic volume (which is much higher through A1 than A2). Multiple source transfer learning models (that benefit from two bouts of training on different source areas before being validated on the destination one) are consistently superior to single source models. In situations where native models are not available, this experimental conclusion supports our claim that transfer learning models can be be confidently used as competitive substitutes.

Modelling Traffic under Anomalous Conditions This segment of the experimental analysis is aimed at comparing transfer learning with skipping (GENTLER) against transfer learning with smoothing ($\overline{\text{GENTLER}}$), in terms of their efficiency at producing traffic models that are competitively accurate, even though they were trained and validated on data affected by anomalies (i.e., vehicle flow disruptions caused by restrictions put in place during the Commonwealth Games). The relevant data are included in the two columns of Table 2 under the AA heading.

Out of the three areas, A2 continues to have the lowest outflow (see columns under the Anomalous heading in Table 1). However, compared to pre-Games readings, traffic exiting A2 through the roundabout at the entrance to the Chinese Quarter is three times as busy. Taken in conjunction with the decreased average flow out of A2 and A3, this suggests that, during the Games, a significant part of Five Ways and Moor St traffic was diverted via the roads around The Mailbox and New St Station. This appears to have had very little effect on the vehicle flow evenness: standard deviation levels under anomalous conditions are comparable to pre-Games ones.

Overall, both skipping and smoothing are efficient at modelling the changes in traffic patterns during the Games:

– A1 predictions stay in the same accuracy band. The A1 native model produced by GENTLER (skipping) has an average RMSE of 27.9% compared to 26% before the Games, whereas the $\overline{\text{GENTLER}}$ model (smoothing) performs at 27.6% compared to 24.8% pre-Games.
– A2 predictions become better: skipping takes the average RMSE down from 57.5% to 40.3%, whereas smoothing achieves a model accuracy increase of almost 20 percentage points. Since A2 is the recipient of traffic diverted from the other two areas during the Games, the fact that both temporal anomaly handling mechanisms we propose are effective at accurately capturing that dynamic is evidence in support of our contributions' value.
– A3 predictions become marginally worse: skipping leads to an increase in the average prediction error from 30.1% pre-Games to 43.9%, whilst smoothing performs better, causing a precision loss of only 5 percentage points.

Transfer learning interacts with skipping in the expected way: GENTLER fails to transfer traffic patterns learnt whilst training on A3 (the area where the output validation data set has the highest standard deviation) to either A1 or A2 destinations. However, when training takes place on A3 and a second source, skipping brings down the average error recorded in the destination area by as much as 52 percentage points compared to pre-Games levels. The transfer learning, smoothing combination leads to single source transfer models comparable to those produced by GENTLER. They are outperformed by $\overline{\text{GENTLER}}$ models trained on multiple sources, most notably when A2 is the destination: there is a drop in prediction error of 15 to 24 percentage points compared to pre-Games levels. This adds to the above mentioned evidence attesting to the efficiency of our temporal anomaly handling mechanisms, in that they are now shown to also be effective in combination with spatial anomaly handling (transfer learning).

Modelling Traffic under Mixed Conditions This segment of the experimental analysis is aimed at investigating the efficiency of transfer learning with skipping (GENTLER) and transfer learning with smoothing ($\overline{\text{GENTLER}}$) at predicting anomalous traffic based on patterns learned from normal one and vice-versa. The investigation is based on the data in the four columns at the centre of Table 2 (headed NA and AN).

The performance of the native models produced by GENTLER indicates that skipping is just as effective under heterogeneous conditions (i.e., models are trained on normal traffic and validated on anomalous one or vice-versa) as it is under homogeneous ones (i.e., models are both trained and validated on either normal traffic or abnormal one): the two heterogeneous A1 native models are in the 26.3% - 28% error range, where the homogeneous ones also lie, with a similar conclusion to be drawn for A2 and A3. It is particularly relevant to note that regardless of whether their training occurs under normal or anomalous circumstances, native models predict either type of traffic comparably well. This is most obvious in the case of A2: the AN and NN models are practically equally accurate (approximately 57% average prediction error), whilst the same can be said about the NA and AA models (circa 40% average prediction error). Analysing the performance of $\overline{\text{GENTLER}}$ native models yields similar experimental findings. This implies that, whenever it is necessary to model future traffic anomalies (e.g., predict vehicle flow during the 2026 edition of the Games) for which anomalous training data is yet to become available, skipping and smoothing make it possible for training to be performed on data collected under normal circumstances, without compromising the prediction accuracy.

Discounting combinations that include source A3, where vehicle flow is too uneven to allow for the efficient transfer of learnt patterns across areas, all single and multiple source transfer learning models perform within an accuracy band of approximately 10 percentage points. When combining knowledge learnt by training on A1 and A2 (in either order), skipping produces models capable of predicting normal traffic within a 4% error margin, regardless of whether training took place under normal or anomalous circumstances. When it comes to predicting anomalous traffic, the same source combination, when skipping is applied, brings the margin down to 2.4%. Also note that all four multiple source transfer learning models produced by GENTLER, where A3 is the destination, outperform native models by at least 8% and as much as 21%. Smoothing is less effective than skipping at producing multiple source transfer models that rival the accuracy of native ones (when A3 is the destination, the prediction quality of $\overline{\text{GENTLER}}$ models worsens by as much as twofold). Yet, the observation regarding the relative competitiveness of heterogeneous and homogeneous models continues to be valid in the case of smoothing: $\overline{\text{GENTLER}}$ models trained on A1 and A2 (in either order) are within the 2.1% - 10.3% accuracy band.

6 Conclusions

Traffic modeling and prediction are central to efficient intelligent transportation which, in turn, is a key component of the smart cities initiative and the UN's Sustainable Development strategy. It is thus essential that efficient algorithms be developed to produce traffic predictions with competitive accuracy in a variety of practical settings: the area where traffic is being predicted has not been fitted with sensing equipment, the traffic predictions will inform city planners' decision making concerning the layout of new roads, the traffic readings used for training were collected during sporting events or other kinds of short and medium term disruption, etc.

To cater to such needs, we propose $\overline{\text{GENTLER}}$, a traffic modelling and prediction algorithm that leverages Genetic Programming enhanced with Transfer Learning and randomisation, on the one hand, and presents increased tolerance to training data anomalies, on the other hand. The former quality enables $\overline{\text{GENTLER}}$ to predict vehicle flow through areas where traffic data are not available, by learning from readings collected on neighbouring areas of the road network. The latter feature is afforded by linear algebra functions that mitigate the effect of training outliers via aggregation and cumulation.

$\overline{\text{GENTLER}}$ produces competitive models regardless of whether the training and validation data were collected during typical or anomalous traffic conditions.

We support that claim with a comprehensive set of experimental results obtained by running $\overline{\text{GENTLER}}$ on traffic readings obtained before and during the 2022 Birmingham Commonwealth Games. Those results indicate that the prediction accuracy of $\overline{\text{GENTLER}}$ models does not deviate from the GENTLER reference in a statistically significant way. In other words, when native models are not available, or heterogeneous predictions are required, running transfer learning equipped with smoothing and, respectively, skipping, selecting the most accurate of the resulting models and using it to predict traffic through the destination area will yield a level of accuracy that is comparable to baseline.

References

1. Almeida, A., Brás, S., Oliveira, I., Sargento, S.: Vehicular traffic flow prediction using deployed traffic counters in a city. Future Generation Computer Systems **128**, 429–442 (2022)
2. Azzali, I., Vanneschi, L., Bakurov, I., Silva, S., Ivaldi, M., Giacobini, M.: Towards the use of vector based gp to predict physiological time series. Applied Soft Computing **89**, 106097 (2020)
3. Bliemer, M.C., Ban, X.J., Leclercq, L., Qian, S., Unnikrishnan, A., Yang, X.S.: Special issue on dynamic transportation network modelling, emerging technologies, data analytics and methodology innovations. Transportation Research Part C: Emerging Technologies **142**, 103778 (2022)
4. Boukerche, A., Tao, Y., Sun, P.: Artificial intelligence-based vehicular traffic flow prediction methods for supporting intelligent transportation systems. Computer Networks **182**, 107484 (2020)
5. Cacco, A., Iacca, G.: Simulation-driven multi-objective evolution for traffic light optimization. In: Castillo, P.A., Jiménez Laredo, J.L., Fernández de Vega, F. (eds.) App. of Evol. Comp. pp. 100–116. Springer (2020)
6. Cintrano, C., Toutouh, J.: Multiobjective electric vehicle charging station locations in a city scale area: Malaga study case. In: Jiménez Laredo, J.L., Hidalgo, J.I., Babaagba, K.O. (eds.) App. of Evol. Comp. pp. 584–600. Springer (2022)
7. Ekárt, A., Patelli, A., Lush, V., Ilie-Zudor, E.: GP with transfer learning for urban traffic modelling and prediction. In: 2020 IEEE CEC. pp. 1–8 (2020)
8. Kajihara, S., Sato, H., Takadama, K.: Generating duplex routes for robust bus transport network by improved multi-objective evolutionary algorithm based on decomposition. In: Castillo, P.A., Jiménez Laredo, J.L. (eds.) Applications of Evolutionary Computation. pp. 65–80. Springer (2021)
9. Li, J., Guo, F., Sivakumar, A., Dong, Y., Krishnan, R.: Transferability improvement in short-term traffic prediction using stacked LSTM network. Transportation Research Part C: Emerging Technologies **124**, 102977 (2021)
10. de Medrano, R., Aznarte, J.L.: A spatio-temporal attention-based spot-forecasting framework for urban traffic prediction. Applied Soft Computing **96**, 106615 (2020)
11. Nagy, A.M., Simon, V.: Survey on traffic prediction in smart cities. Pervasive and Mobile Computing **50**, 148–163 (2018)
12. Patelli, A., Hamilton, J.R., Lush, V., Ekárt, A.: A gentler approach to urban traffic modelling and prediction. In: 2022 IEEE CEC. pp. 1–8 (2022)
13. Wittpohl, M., Plötz, P.A., Urquhart, N.: Real time optimisation of traffic signals to prioritise public transport. In: Castillo, P.A., Jiménez Laredo, J.L. (eds.) Applications of Evolutionary Computation. pp. 162–177. Springer (2021)