Yugoslav Journal of Operations Research
7 (1997), Number 2, 257-266

ENTROPY MODEL TO SIMULATE ORIGIN-DESTINATION TRIP GENERATION

Savo M. JOVANOVIĆ

Faculty of Electrical Engineering, University of Belgrade,
Bulevar Revolucije 73, 1000 Belgrade,
Yugoslavia

Abstract: In this paper the general concept of entropy developed in physical is applied to forecast the transportation process parameters in a city or a broader region. The expression for entropy is obtained by generalization of the classical Hamiltonžs principle of the smallest action as the most general formulation of the movement of mechanical systems. The practical example of the application of this example of this concept to the real system, a medium sized city, shows that the approach presented in this paper gives quite good and useful results, and can be used for solution of various planning problems in transportation networks.

Keywords: Transportation, planning, forecasting.

1. INTRODUCTION

Traffic is the result of man's activities. Technological development is changing the scope and structure of traffic whose development has a feedback effect on technological development. This socio-economic cause-and-effect connection should be noted in traffic planning in order to furnish diagnose which is the only possible way to reach valid prognoses.

One of the most frequent methods in traffic forecasting is the gravitation model with various modifications [5], [3] adapted to the characteristics of the given environment and level of technological development. However, it would be erroneous to assume that the mechanisms of these manifestations are also changed by technological development. Natural subsystems change the response of their manifestations, reflecting changes in sets of influential parameters, but nature does not change its mechanisms.

The gravitation model is phenomenologically inconsistent regardless of the well-prepared forms of its responses, which can be understood as interpolation formalisms. Every force is proportional to the mass acceleration, though the exponents of physical variables in these relations cannot be arbitrary. On the other hand, the preparation of a diagnostic model requires a complete system and this is why a survey is very expensive, which is not of utmost importance in this case. What is important is that a valid long-term forecast cannot be expected.

Owing to the fact that traffic systems are stochastic, entropic systems tending towards disorganization and a continuous rise in entropy, during control of these systems the state of entropy can be considered to be the generalized thermodynamic function of the state of a nonthermodynamic system [1].

Following the suggestions of classical mechanics and using Hamilton's principle of the smallest action as the most general formulation of the movement of mechanical systems, with the generalization of adequate variables and concepts it is possible to formulate differential equations of phenomena in the most general form. In the same vein, we can write the basic relations of potential and kinetic energy, and using the law on linearity and Onsager's theorems which comprise the basis of thermodynamics of irreversible slowly changing states, we can obtain an expression for the onset of entropy. All these expressions have meaning in the deductive approach as well, when it is not possible to formulate a system of differential equations and when it is not possible to see a system of material points. Using the consistency of natural subsystems as a point of departure, with the assumption of generalized phenomenological relations it is possible to deductively identify the system by calculating phenomenological coefficients [4].

It is interesting to note here the similarity of the mentioned expressions, for the potential and kinetic energy of mechanical systems and for the onset of entropy of thermodynamic systems, as quadratic forms of position, speed and generalized force, respectively. Wiener's message is in this spirit: "The goal of cybernetics is to develop a common language and corresponding technique so that problems of control and communication may be encompassed in the general scope and a suitable repertoire of ideas and techniques may be found allowing the individual characteristics of manifestations to be classified under certain common concepts".

Taking the nature of traffic systems as a point of departure, we will present a model of the onset of entropy as the generator of a simulated origin-destination traffic, and using the inductive-deductive method, with a review of the mathematics of phenomenology, we will undertake the identification of the system and the numerical expression of phenomenological coefficients.

2. MATHEMATICAL MODEL

If the value of parameter P_k, of subsystem S_i is disturbed by increment Δp_k, then there will be a negative change in entropy which in the first approximation can be expressed by the quadratic form:

$$\Delta S = -1/2 \sum_{k,l} L_{kl} \Delta p_k \Delta p_l ,$$

so the onset of entropy or the change of entropy per time unit is:

$$\sigma = \frac{\partial(\Delta S)}{\partial \tau} - \sum_{k,l} L_{kl} \Delta p_k \Delta p_l$$

$$\sigma = \sum_{l} \Phi_l X_l$$

where $X_1 = \frac{\partial(\Delta S)}{\partial(\Delta p_l)} = -\sum_k L_{kl}\Delta p_k$, is the generalized thermodynamic driving force and $\Phi_1 = \frac{\partial(\Delta P_l)}{\partial\tau} = -\sum_k L_{lk}X_k$, describes the irreversible process and L_{kl} are the phenomenological coefficients.

Following Onsager's mutuality relation $L_{kl} = L_{lk}$.

Therefore, the onset of entropy of subsystem $S_i \subset S$, $(i = \{1,\dots,N\})$, by disturbing parameter P_k, P_k, $(k \in \{1,\dots,n\})$ can be expressed by the quadratic form:

$$\sigma = \sum_{k=1}^{n} \sum_{l=1}^{n} L_{kl} X_k X_l$$

or complementarily as the interaction of subsystems S_i and S_j:

$$\sigma^{ij} = \sum_{k=1}^{n} \sum_{l=1}^{n} L_{kl} X_k^{ij} X_l^{ij}$$

Since traffic appears as the result of sociological and economic differences, and traffic system is treated as static and entropic, origin-destination trips of the i-th and j-th traffic units will be equated with the function of the onset of entropy in the interaction of these units:

$$F_{ij} = \sum_{k=1}^{n} \sum_{l=1}^{n} L_{kl} X_k^{ij} X_l^{ij}$$

where

X_{kl}^{ij} - driving forces based on socio-economic parameters, and
L_{kl} - the phenomenological coefficients.

Based on the Onsager relation $L_{kl} = L_{lk}$ we can write:

$$F_{ij} = \sum_{k=1}^{n} \sum_{l=1}^{n} (2-\delta_{kl}) L_{kl} X_k^{ij} X_l^{ij}, \qquad \delta_{kl} = \begin{cases} 0, k \neq 1, \\ 1, k = 1 \end{cases}$$

Our problem is now reduced to the assumption of the form of the driving forces as a function of socio-economic parameters and the characteristics of the road network. Then by counting and conducting surveys a representative sample should be established in order to numerically determine the parameters of system L_{kl}.

2.1. Driving forces and relevant parameters

Every inhabited place that is characterized by a set of parameters as a measure of activity may be considered an elementary traffic unit. These parameters may be:

- spatial,
- demographic,
- economic,
- socio-economic,
- sociological, etc.

Two systems of traffic units are used in practice:

- the grid system and
- the zone system.

The grid system is based on uniform units of space. It is suitable when considering an urban environment.

The zone system is based on unequal units of space that represent parts of or entire administrative or statistical region with changeable boundaries. This system is suitable for studying regional road networks. It requires special criteria to establish and define zones. Thus they must:

- have a clear center of concentration that should be large enough so that together with its zone of influence it generates sufficiently large traffic compared to the region under observation,
- represent a homogenous whole with respect to the specific features of most characteristics.
- comprise areas which have a set of parameters as a measure of activity (usually defined statistically or administratively),
- be connected with basic traffic directions.

One zone may hold several traffic units. They are usually given a code number in the zone. The set of activity parameters may be arbitrary with respect to the system, but identical with respect to zones. It can contain parameters such as:

1. position
2. population size
3. population density
4. number of jobs
5. size of actively working population
6. size of actively working population in primary activities
7. size of actively working population in secondary activities
8. size of actively working population in tertiary activities
9. number of motor vehicles
10. degree of motorization
11. per capita income, etc.

The choice of k relevant parameters from the set of n statistical parameters is made by choosing combinations of the k-th class of n elements $\binom{n}{k}$ for which the correlation factor is maximum. The numerical value of the elements of every parametric vector $\{p_k\}$ is normalized in relation to the maximum element of that vector.

Now let us assume that driving forces X_k^{ij}, of activity p_k, between traffic units i and j are given by the expression:

$$X_k^{ij} = A_k^i p_k^i p_k^j (-\lambda_k^i d_{ij})$$

where

$d_{ij}, (i, j \in N)$ - elements of the trip matrix; distance between units i and j, travel time or travel cost $(N = \{1,\ldots,N\})$;

λ_k^i - factor of exponential change in the influence of activity k of the i-th unit;

It follows that:

$$d_k^i \sum_{j=1}^{N} p_j^k = \sum_{j=1}^{N} p_k^i d_{ij}$$

$$\lambda_k^i = \frac{1}{d_k^i} = \frac{\sum_{j=1}^{N} p_j^k}{\sum_{j=1}^{N} p_k^i d_{ij}}$$

If local, city traffic is under consideration, then the exponent is proportional to distance dij, and we will take reciprocal value d_{ij} and use the same expression. In both cases we will normalize values d'_{ij} and d_{ij} in relation to the maximum value.

Normalized parameter A_k^i can be obtained from the following:

$$a_k = \max(p_k^i p_k^j)$$

$$A_k^i a_k \max_i \sum_{j=1}^{N} \exp(-\lambda_k^i d_{ij}) = \sum_{j=1}^{N} \exp(-\lambda_k^i d_{ij})$$

Starting with individual activity parameter ($k = 1$), with an increase in k, there will be an increase in the number of linear equations for the determination of coefficients L_{kj}, from 1 to $\frac{k(k+1)}{2}$. This number requires a larger number of experimental data and this leads to numerical saturation and weak constraints of the linear system. We can conclude that enlarging the set of activity parameters increases the analytical determination, but will weaken the numerical determination beginning with some value of k. Thus we arrive at the optimal value of k that defines the system's indetermination.

It has been experimentally established that it is sufficient to take 7 activity parameters, $k = 7$. For larger values of k, there is a rise in the weak constraints to the linear system, owing both to the numerical dependence of the parameter values and the possibility of their mutual dependence.

3. EXPERIMENT

Based on the above expressions, a complete program system was developed to diagnose and forecast origin-destination trips with corresponding numerical processing. During this development, several real systems were analyzed, both regional and city, and the model diagnosis had a very high correlation factor of around 90%. We will restrict ourselves here to an analysis of the urban area of an old city with around 120,000 inhabitants.
This is a case when the forces of attraction rise proportionally to the distance d_{ij}.

3.1. Input data

The urban area is divided into two traffic zones: zone 01 and zone 02, with 6 and 4 traffic units, respectively. These units correspond administratively to local communities or city sections. We will treat them simply as active nodes in the network, and their centroid will be the center of the zone.

A set of seven relevant parameters was chosen by experimental means, whose size and structure corresponds to the maximum correlation factor.

This set is:

- position
- population
- number of jobs
- size of actively working population in secondary activities
- size of actively working population in tertiary activities
- degree of motorization
- per capita income

These data for every traffic unit are given in Table 1.

Table 1. Relevant activity parameters

node	1.	2.	3.	4.	5.	6.	7.
0100	1.0	0.790	0.208	1.410	0.940	1.053	1.300
0101	1.0	0.680	0.900	1.010	1.010	1.538	1.800
0102	1.0	0.600	1.340	0.537	1.253	1.370	1.900
0103	1.0	0.420	0.680	1.125	1.125	1.266	1.600
0104	1.0	0.060	6.982	0.180	0.100	1.000	1.300
0105	1.0	0.133	3.965	0.400	0.100	1.111	1.330
0106	1.0	0.900	0.600	1.072	1.608	1.538	1.800
0200	1.0	0.825	1.170	0.245	2.205	1.695	1.900
0201	1.0	0.899	0.280	1.602	1.068	1.492	1.900
0202	1.0	0.710	0.580	0.633	1.477	1.515	2.100
0203	1.0	0.420	0.140	0.214	0.856	1.923	2.300
0204	1.0	0.997	1.580	1.485	1.485	1.538	1.600

The second group of input data consists of the travel matrix $\{d_{ij}\};\ i,j \in N$ whose elements are the distances between nodes, travel time or, as already mentioned, travel cost that is calculated on the regional road network. This matrix is symmetrical and is given in the form of the triangular matrix in Table 2. This example is calculated with the reciprocal normalized matrix elements.

At the end follows a group of experimental data obtained by counting and surveying. The choice of data in this group is made by choosing the interactions of representative nodes, e.g. of small, medium and large nodes, and then no more than 10 pairs of unknown phenomenological coefficients are taken into the diagnosis process.

In addition to these input data, the program also provides checking of the experimental data, checking of the set of experimental parameters and differentiation between the regional and road network.

Pairs of experimental data and model's output results (upper values) in the diagnosis of phenomenological coefficients for the traffic network of this city are given in the Table 6.

Table 2. Distance matrix

node	0100	0101	0102	0103	0104	0105	0106	0200	0201	0202	0203	0204
0100	0	172	237	352	818	347	256	353	547	627	415	450
0101		0	178	293	804	330	234	308	378	380	350	278
0102			0	229	731	450	387	167	477	488	286	283
0103				0	626	619	485	273	646	470	240	427
0104					0	720	976	777	1167	840	643	1009
0105						0	371	589	650	816	654	540
0106							0	401	430	589	473	210
0200								0	438	300	192	409
0201									0	270	507	283
0202										0	266	507
0203											0	403
0204												0

3.2. Output results and parameters of the model

Based on experimental data and network parameters, the least squares method is used to form a system of linear equations of the n-th order where n varies from 1 to $\frac{k(k+1)}{2}$, i.e. fo$k = 7$ and $n = 28$. The system is weakly constrained with a symmetrical matrix. We used the Gauss-Seidel procedure to solve it in the diagnosis process, however, in order to more precisely solve the phenomenological coefficients for $k = 7$ and $n = 28$, we used an interactive procedure to solve the symmetrical weakly-constrained system through its eigen values.

In the diagnosis process, the combinations of given parameters start with $k = 1$, so that we receive its correlation factor R (Table 3).

Table 3: Correlation factors

	σ_{exp}	σ_{par}	n	R
1.	95.1481	88.5680	1	0.9138
2.	95.1481	54.7301	1	-0.1280
3.	95.1481	96.6779	1	0.6864
4.	95.1481	52.6642	1	-0.0391
5.	95.1481	53.5484	1	-0.1382
6.	95.1481	66.6079	1	0.5770
7.	95.1481	66.4710	1	0.6165

However, the maximum correlation factor is obtained with the participation of all seven parameters and is $R = 0.968124$. This information is received through the logic vector whose possible value states are 0 and 1, by the state:

$$(1,1,1,1,1,1,1).$$

Should the positive effects, e.g. parameters 2, 4 and 5, be excluded, then the values of this vector would be:

$$(1,0,1,0,0,1,1)$$

and the phenomenological coefficients would be obtained as the solution to the system of linear equations of the 10th order.

In this numerical process we also receive normalized model parameters λ_k^i and A_k^i, (Table 5).

Finally, by solving the weakly constrained system of linear equations we obtain the phenomenological coefficients L_{kl}; $k,l \in N(cardN = 28)$ (Table 4).

Table 4: Phenomenological coefficients $L_{kl} = \frac{1}{2} L_{kl},\quad k \neq l$

L_{kl}	1.	2.	3.	4.	5.	6.	7.
1.	235.732	21.579	43.617	45.349	9.118	191.862	199.835
2.		-2.653	2.435	3.065	-3.569	8.460	8.681
3.			2.538	3.874	1.142	17.398	17.850
4.				4.342	1.300	19.343	19.594
5.					-0.991	3.458	3.670
6.						41.362	86.169
7.							45.152

Now we have a complete model to generate origin-destination trips with numerical parameter values for the given traffic region:

Table 5: Normative system parameters

λA	1.		2.		3.		4.		5.		6.		7.	
	λ	A	λ	A	λ	A	λ	A	λ	A	λ	A	λ	A
0100	2.262	1.000	2.227	0.968	2.533	0.021	2.342	0.353	2.143	0.194	2.182	0.270	2.197	0.188
0101	1.873	0.980	1.750	0.987	2.362	0.018	1.760	0.372	1.748	0.192	1.897	0.256	1.888	0.179
0102	1.914	0.999	1.783	1.006	2.519	0.018	1.823	0.375	1.730	0.200	1.856	0.269	1.909	0.184
0103	2.416	0.982	2.299	0.975	2.858	0.019	2.694	0.327	2.254	0.193	2.311	0.267	2.320	0.186
0104	5.235	0.938	5.100	0.917	7.523	0.015	5.202	0.342	4.945	0.183	5.103	0.252	5.110	0.176
0105	3.315	0.963	2.968	1.003	4.396	0.017	2.967	0.380	3.035	0.192	3.277	0.256	3.294	0.178
0106	2.400	0.990	2.246	0.994	2.768	0.020	2.141	0.390	2.379	0.186	2.400	0.261	2.441	0.180
0200	2.084	0.998	2.103	0.949	2.869	0.017	2.068	0.364	2.122	0.184	2.038	0.267	1.998	0.189
0201	2.953	0.980	2.720	0.997	3.583	0.019	2.975	0.353	2.625	0.199	2.860	0.265	2.884	0.183
0202	2.842	0.987	2.688	0.984	3.720	0.018	2.773	0.364	2.681	0.192	2.697	0.270	2.740	0.186
0203	2.278	0.986	2.109	0.998	2.656	0.019	2.239	0.362	1.943	0.206	2.278	0.260	2.270	0.182
0204	2.417	0.983	2.279	0.983	3.393	0.017	2.340	0.365	2.268	0.192	2.360	0.264	2.338	0.185

$$F_{ij} = \sum_{k=1}^{n} \sum_{l=1}^{n} L_{kl} X_k^{ij} X_l^{ij} \quad \text{where}$$

$$X_k^{ij} = A_k^i p_k^i p_k^j \exp(-\lambda_k^i d_{ij}).$$

Normative parameters λ_k^{ij} and A_k^i are given in Table 5, for each $i \in N$ in pairs, and phenomenological coefficients L_{kl} are given in Table 4.

Table 6: Comparison of experimental and calculated data of the entropy model
Correlation factor $R = 96.81\%$

	0100	0101	0102	0103	0104	0105	0106	0200	0201	0202	0203	0204
0100		15 9	29	61 78	205	63	40	46 87	199	226	130	142
0101	18		24	108 90	341 277	161 95	60	115	174	172	143 164	90 91
0102	47	13 24		40 48	278 263	159 159	170	20	237	248	79 110	91
0103	65	51	37 23		182 157	163	153	55 43	239	149	18 33	110 120
0104	54	65	51	32		42	95	63	128	74	44	98
0105	21	22	52	95	100		32	99	117	168	133	80
0106	32	33 28	83 107	115 152	379 271	75		115 124	149	115 238	247 185	20
0200	103	98 100	15	49 66	276	195 218	171		98 201	114 101	31	130 171
0201	127	50 70	116 112	182 181	347 268	149 148	98	132 101		26	182 152	30
0202	165	81	127 133	79 112	143 115	220	195	44	31		32 34	143 147
0203	121	114	70 68	28 37	252 217	234	203	21	230	42 62		70 147
0204	124	50	54 49	126 121	450 275	215 146	20	128	36 57	180 188	72 136	

The drawback of quadratic forms is seen in the additive features of its members owing to the possibility of the induction of numerical errors of cutting. However, the probability of entropy decreasing in individual activities is small and thus the dominant effect of the positive values of the phenomenological coefficients is expected, which is confirmed by the results received.

Generated origin-destination traffic simulated using this model is given in Table 6 along with experimental data (lower values). The correlation factor of these data is 96.81%.

In addition to this example, several other real systems were concerned, the most important being the regional plan of Serbia. This regional system consists of 105 traffic units in 9 zones. The correlation factor of this model was 87.14%.

4. CONCLUSION

In this paper the general concept of entropy developed in physics is applied to forecast behavior in a transport system in a city or broader area. The practical example of the application of this concept on a real system, a medium-sized city, indicates that the method presented in the paper gives quite good and usable results. These results can be used directly to solve various planning problems and can also be used as input data for the program system that is used to determine the distribution of traffic in the given traffic network.

REFERENCES

[1] Jovanović, S.M., "Information models of regional road networks", Report IBK, 1345, Vinča, 1975.

[2] Ortuzar, J. de D., Willunsen, L.G., *Modelling Transport*, Willey, 1990.

[3] Petrović, M., *Elements of Mathematical Phenomenology*, Belgrade, 1991.

[4] Prigogine, I., *Introduction to the Thermodynamics of Irreversible Processes*, Charles and Thomas III., Springfield, 1955.

[5] Wilson, A.G., *Entropy in Urban and Regional Modelling*, Point-England, London, 1970.