A Metropolitan Case Study

Subscriber and Traffic Forecasting

Mr. H. Leijon, ITU

UNION INTERNATIONALE DES TELECOMMUNICATIONS
INTERNATIONAL TELECOMMUNICATION UNION
UNION INTERNACIONAL DE TELECOMUNICACIONES

A Metropolitan Case Study

Subscriber And Traffic Forecasting

CONTENTS

1. Notations Used In This Document

Traffics

A Total traffic
a Traffic per sub. (main line)

Zones

Z General notation
p,q Index

Traffic Areas (C=Centar, S=Susak, K= Kozala, Z=Zamet)

T General notation
k,l Index (1, 2, 3, 4)

Categories (R=RES, B=BUS, P=PBX, C=CB, N=B+P+C)

X General notation
b,c Index (R, B, P, C = 1, 2, 3, 4)

Exchange Areas

E General notation
u,v Index

Numbers

C Connected subs (main lines)
W Waiting list (size)
D Demand (expressed, i.e., D=C+W)
S Saturation value
P Population
H Households
E Employees

Other Parameters

F Correction factor

d Weight for the distribution of given sub. traffics (originating, excluding LD) to various categories

W Weight for the distribution of given category - to - category traffics per sub, between different traffic areas

. means "total"

pq (as index) means "from p to q"

t Time (-10,.........0,.........20)

’ (prime) means "intermediate (temporary) value"

2. Subscriber Forecast

We need to forecast $C_x^{(t)}(z)$ = number of connected main lines C for each category x and zone z for future point of time t = 5, 10, 15.

We have the following data:

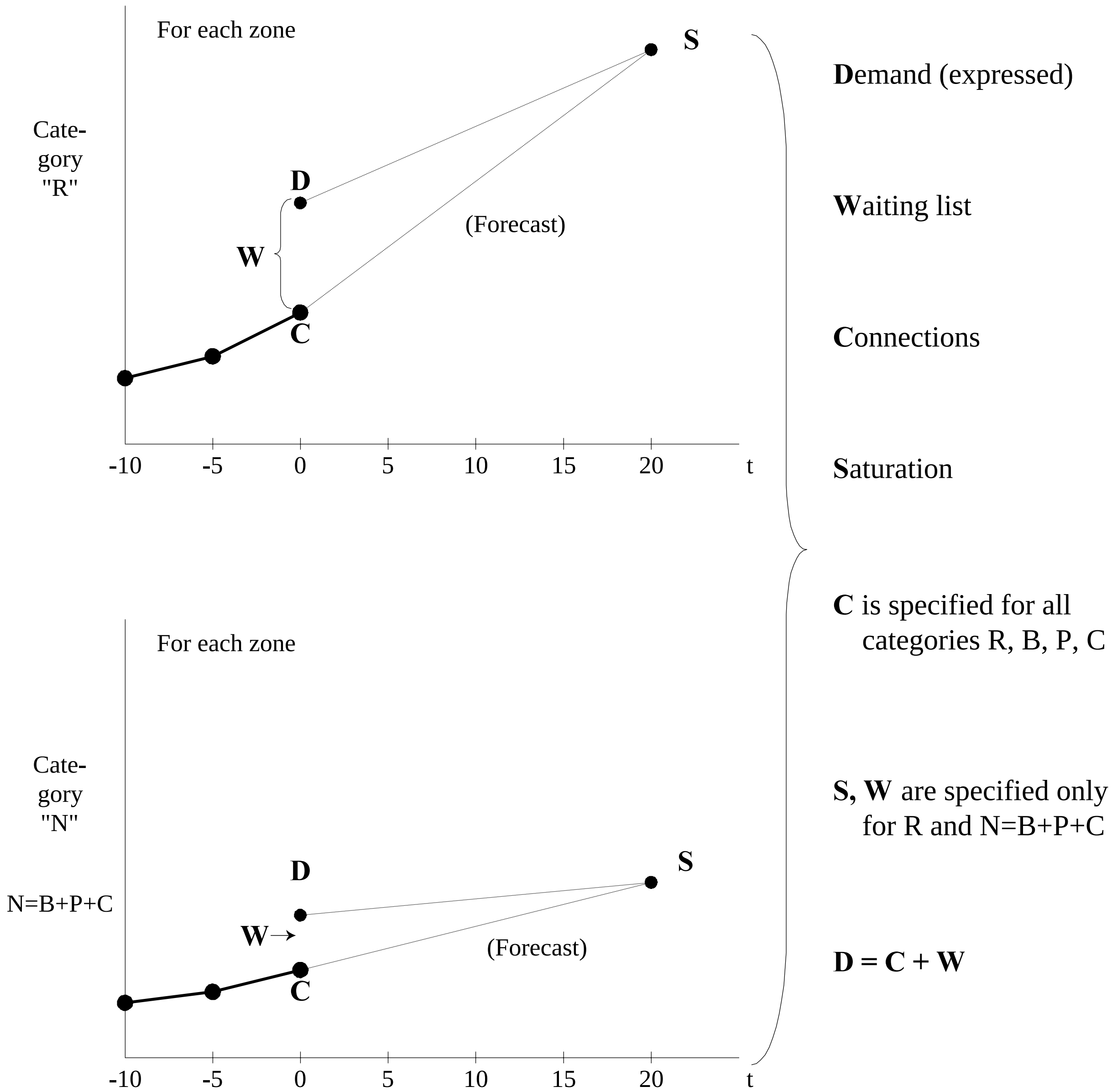
a) $P^{(t)}(T)$ = Population P per traffic area T for some years t from 1971 to 2005;

b) $E^{(t)}(T)$ = No. of employees E;

c) $H^{(t)}(T)$ = No. of households H;

d) Other data of various kinds;

e) See diagrams shown below:



We should accept as generally reliable the given historical data on:

the number of connected main lines	=	the satisfied demand C, and
the size of the waiting list	=	the expressed, unsatisfied demand W.

We observe, however, two important facts:

- the estimated values of saturation S, for most of the zones, are very high - in some cases extremely high - compared to the present number of main lines;

- these S-values are themselves forecasts, and thus may be very uncertain.

The obvious conclusions that can be drawn are that the saturation levels should be thoroughly analysed and in some cases adjusted, and that growth curves of the exponential type may not adequately describe development over the next 15 years. It would be preferable to use curves that are based on the development of other, explaining variables. Thus you have two main tasks:

- to adjust S;
- to find development curves.

Your decisions should be based on special studies using the material.

Examples of such studies:

- The development of employee distribution by sectors, for each traffic area.

- The development of individual RES, single line BUS and PBX penetration trends.

- The development of relative RES, BUS, and PBX penetration.

- The employee distribution by sectors, time and traffic areas.

- Business line penetration factors by sectors.

- Saturation levels for RES and BUS lines by sectors, time and traffic areas.

- Possible adjustments of the saturation levels that were proposed by the administration (Adjustments per traffic area or by individual zone?)

- The development of household size over time. Possible influence on the demand?

- Development of calling rates over time.
- Development of LD-traffic per line and per traffic area over time.

} These studies are more useful for the traffic forecasts.

- Will demand really be satisfied in the future (i.e. no waiting list)?

<u>The forecast procedure could be as follows:</u>

There are two mains reasons to adjust S-values:

- your investigations convince you that the S-values are not reasonable;

- you think that a substantial waiting list will still exist in the year 20 (S should not contain waiting).

The adjusted values should be fed into the data file.

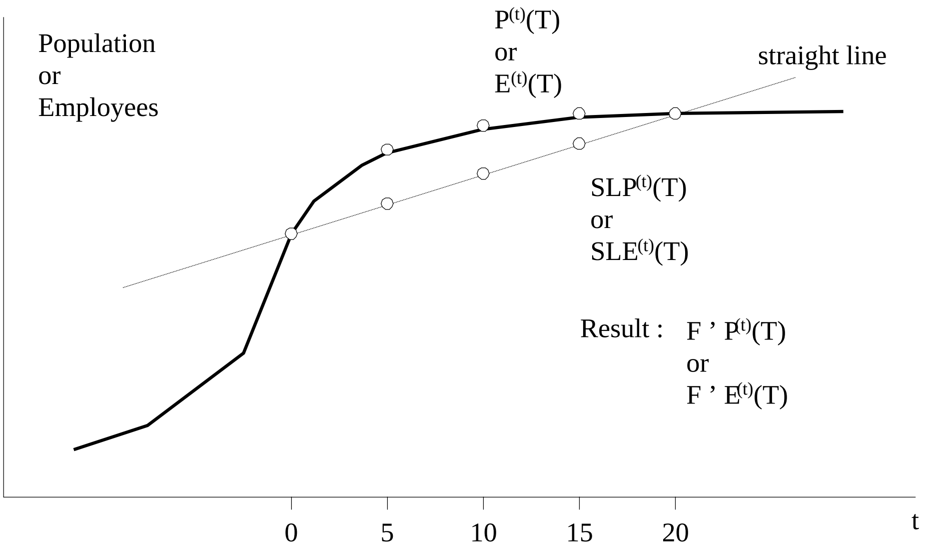
To calculate the development of C from year 0 to year 20, the following steps are followed:

A pre-stored computer algorithm will calculate the development as deviations from a straight line between $C^{(o)}$ and S. To use this algorithm, you need to read in some correction factors F for each point of time and for each traffic area. The algorithm will thus use the same factors for each zone in the same traffic area.

There are **many** different approaches used to estimate these correction factors. **One** very simple approach follows here:

For each traffic area T, study the past and the future development of P and E. The aim is to find sets of F-values which will be used to create development curves from $C^{(o)}$ to S, deviating from straight lines.

For each traffic area :



$$F'P^{(t)}(T) = \frac{P^{(t)}(T)}{SLP^{(t)}(T)}$$

$$SLP^{(t)}(T) = P^{(0)}(T) + \frac{t}{20} \cdot \left[P^{(20)}(T) - P^{(0)}(T)\right]$$

$$F'E^{(t)}(T) = \frac{E^{(t)}(T)}{SLE^{(t)}(T)}$$

$$SLE^{(t)}(T) = E^{(0)}(T) + \frac{t}{20} \cdot \left[E^{(20)}(T) - E^{(0)}(T)\right]$$

Adjust $F'P^{(t)}(T) \Rightarrow FP^{(t)}(T)$

and/or $F'E^{(t)}(T) \Rightarrow FE^{(t)}(T)$

as you wish.

Note: an alternative method could be used!

Feed $FP^{(t)}(T)$ and $FE^{(t)}(T)$ into the computer.

The pre-stored computer program will now calculate, for each traffic area:

$$C_R'^{(t)}(z) = FP^{(t)}(T)\cdot\left[C_R^{(0)}(z) + \frac{t}{20}\left(S_R(z) - C_R^{(0)}(z)\right)\right]$$

and

$$C_N'^{(t)}(z) = FE^{(t)}(T)\cdot\left[C_N^{(0)}(z) + \frac{t}{20}\left(S_N(z) - C_N^{(0)}(z)\right)\right]$$

for each zone in the respective areas.

$$C_N^{(0)} = C_B^{(0)} + C_P^{(0)} + C_C^{(0)}$$

$C_N'^{(t)}(z)$ is then split into $C_B'^{(t)}(z)$, $C_P'^{(t)}(z)$ and $C_C'^{(t)}(z)$:

$$C_B'^{(t)}(z) = C_N'^{(t)}(z)\cdot\frac{C_B^{(0)}(z)}{C_N^{(0)}(z)}$$

$$C_P'^{(t)}(z) = C_N'^{(t)}(z)\cdot\frac{C_P^{(0)}(z)}{C_N^{(0)}(z)}$$

$$C_C'^{(t)}(z) = C_N'^{(t)}(z)\cdot\frac{C_C^{(0)}(z)}{C_N^{(0)}(z)}$$

a) A table showing the result will be printed for your study and adjustment.

b) Specific cases may be plotted on request.
These two kinds of output are illustrated below.

You should now study the result and adjust

$C_{R,B,P,C}'^{(t)}(z) \Rightarrow C_{R,B,P,C}^{(t)}(z)$ = final forecast (maybe in just a few cases).

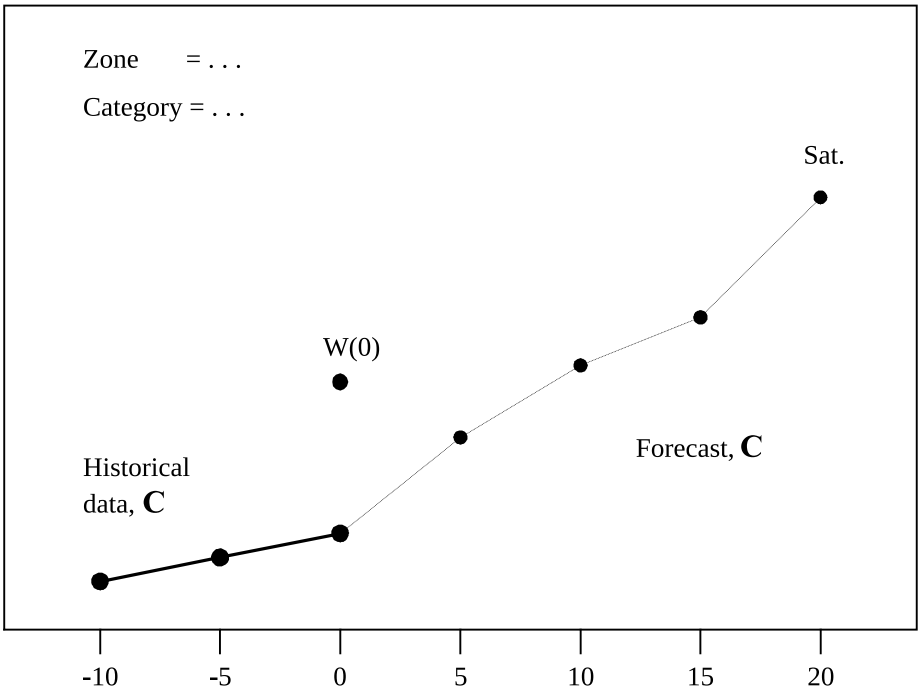
You should use your own judgement, based on the special studies you may make.

Feed the adjusted values into the computer.

a) Table : Connected main lines = $C^{(t)}(z)$

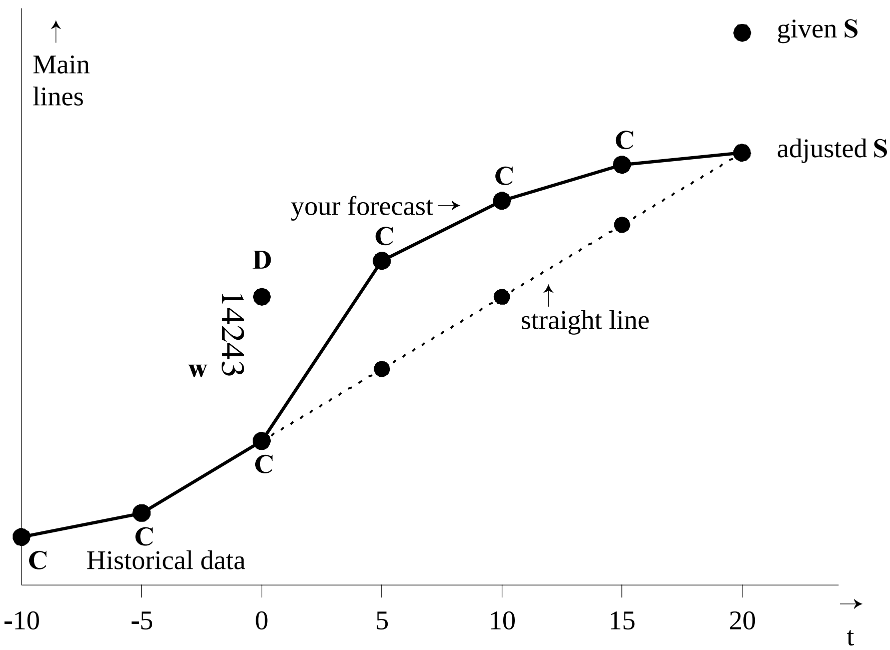
Tr.		Year 5					Year 10	Year 15
Area	Zone	RES	BUS	PBX	CB	Σ		
C	CA							
	CB							
	. . .							
	Σ							
K	KA							
	KB							
	. . .							
	Σ							
	. . .							

b) Plotted version (for particular zones, on request)



What you have obtained so far is illustrated below:

For each category
and
for each zone:



3. Traffic Forecast

We need to forecast traffic interests between all traffic areas, outgoing and incoming long-distance traffics for all areas, and special service traffic from each area, for three points in time: 5, 10, and 15 years in the future.

We have measured and estimated traffic matrices for exchange and traffic areas for three points of time: -10, -5, and 0 (present time). For the present time, we have also estimated calling rates for each traffic area and for each subscriber category. These figures are, of course, not the exact truth - it is sometimes difficult to distinguish between "residential" and "business" lines, and the figures are, in addition, probably based on samples.

The next thing to consider is that the traffic interests between different areas are composed of a large number of small subscriber-to-subscriber traffics, and different subscribers have very different telephone habits and needs.

The calling rate of PBX-lines and even of single line business subscribers is, for example, much higher than the average calling rate of residential subscribers.

Furthermore, we will find that residential subscribers probably will make more calls to other residentials than to business subscribers. The business subscribers they do call will probably often be the single line business ones, since these frequently represent shops, doctors, etc.

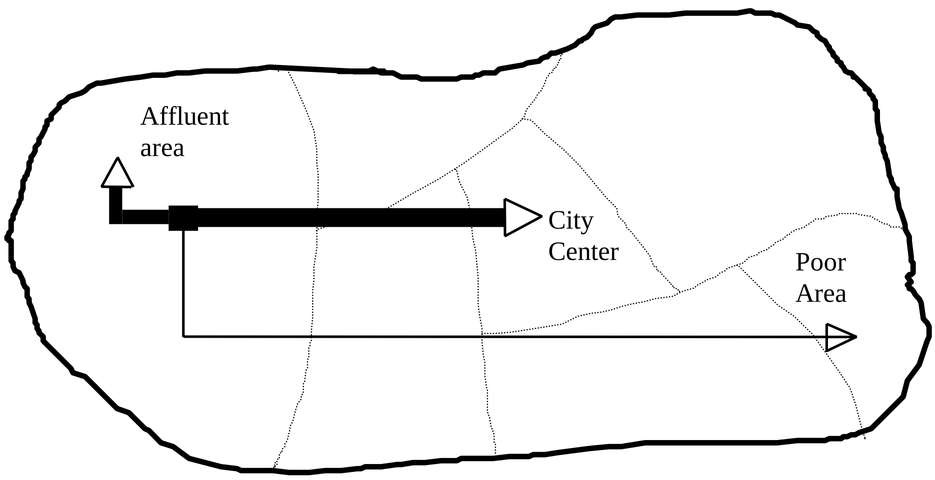
Business subscribers, on the other hand, are likely to call other business lines. This is not the only factor to consider, however. The distribution of calls does not depend solely on the subscriber category, but to a great exent on **where** the subscribers are located, i.e., to which traffic areas they belong.

For example, let us take the calls originated by residential subscribers with a fairly high standard of living.

Let us say that 30% of the local traffic originated by these affluent subscribers is directed toward single line business subscribers in the town, i.e., small shops, hairdressers, lawyers, doctors, etc. These calls will frequently go to businesses located in the same area or in the city center. It is unlikely that many of the calls originated by the affluent residential subscriber will be directed to poorer areas which, in addition, may also be more distant.

From the foregoing, we conclude that we would need to use a few parameters to describe and calculate the traffics between subscribers in different categories and in different areas.

If we could define some such parameters, we would have an excellent tool which would enable us to master the situation as the city develops and as different zones and areas change character and grow at different speeds.



"Residential to single line business traffic" case illustration

What we need is:

- one parameter which describes how much traffic is originated, on average, by a subscriber of a specific category and of a specific area;
- one parameter to describe the amount of that traffic that is directed to subscribers in each specific category;
- one parameter to describe how the traffic directed to a specific category of subscribers is distributed to different areas, depending on the significant characteristics of these areas.

In this document, the three parameters are called:

$a_{b.}(k)$ = Local (excluding LD) originating traffic from a subscriber of category **b** belonging to traffic area **k**

d_{bc} = Proportion of the traffic originated by a category **b** subscriber that is directed to category **c** subscribers

$W_{bc}(kl)$ = Weight that corresponds to the distribution between areas of that part of the trafficoriginated by category **b** subscribers in area **k** and which is going to category **c**.

Our general approach will then be to try to determine the present values of these three parameters first, then use the data we have, and finally try to estimate how the parameter values will change in future situations.

The general properties of the three parameters are:

$a_{b.}^{(t)}(k)$ The parameter is relatively universal, i.e., it does not vary too much between different places of similar character and stage of development, and it is also fairly stable over time.

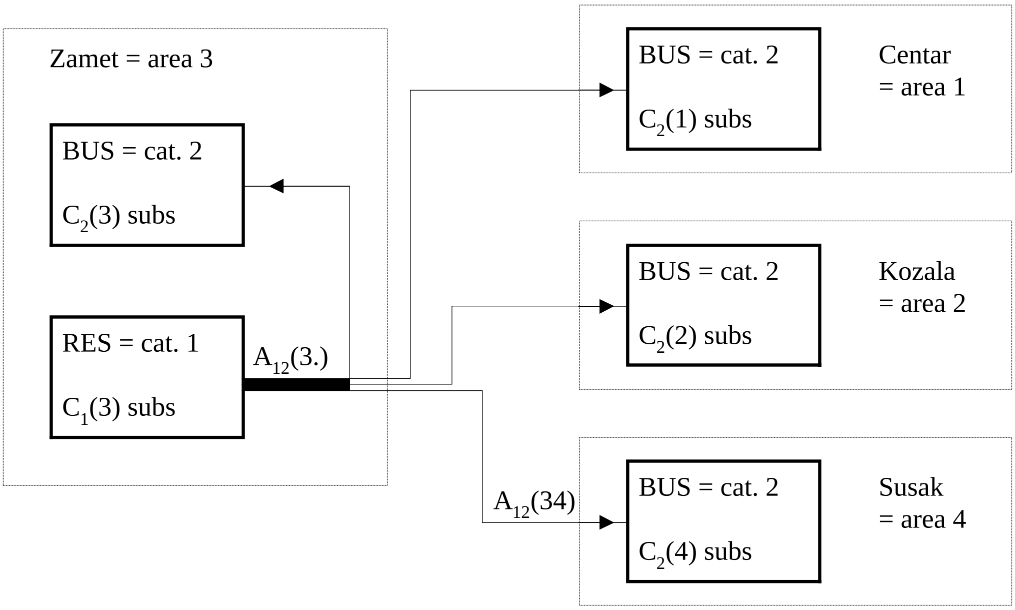
$d_{bc}^{(t)}$ The parameter is somewhat less universal than the first one, i.e., it is more locally influenced and its values also change more with the development of the area.

$W_{bc}^{(t)}(kl)$ The parameter is of an entirely local character, and its values may also change considerably with the development of the area. Fortunately, the individual weights can be taken as very round figures without causing serious errors in the aggregated traffic quantities.

For further details on this type of approach, please see the document "Metropolitan Networks : Estimation of Traffic Forecast Parameters".

The next pages demonstrate how the three parameters are used to calculate a partial traffic stream in the network.

Example : How **a**, **d** ,and **W** are used to calculate the traffic from residential subscribers in the **Zamet** area to single line business subscribers in the **Susak** area.



$$A_{12}(3.) = C_1(3) \cdot a_{1.}(3) \cdot d_{12} \text{ erlang}$$

$$A_{12}(34) = C_1(3) \cdot a_{1.}(3) \cdot d_{12} \cdot \frac{C_2(4) \cdot W_{12}(34)}{C_2(1) \cdot W_{12}(31) + C_2(2) \cdot W_{12}(32) + C_2(3) \cdot W_{12}(33) + C_2(4) \cdot W_{12}(34)} \text{ erlang}$$

The figure shows **all the a- d-** and **W-** values that affect the calculation of . $A_{12}(34)$ No other value will influence that calculation (except for C-values).

Cat. \ Area	1 Centar	2 Kozala	3 Zamet	4 Suzak
1 = RES			$a_{1.}(3)$	
2 = BUS				
3 = PBX				
4 = CB				

Cat.	1 = RES	2 = BUS	3 = PBX	Special Service	Σ
1 = RES			d_{12}		100%
2 = BUS					100%
3 = PBX					100%
4 = CB					100%

Category		1 = RES				2 = BUS				3 = PBX			
	Area	1 = C	2 = K	3 = Z	4 = S	1 = C	2 = K	3 = Z	4 = S	1 = C	2 = K	3 = Z	4 = S
1 = RES	1 = C												
	2 = K												
	3 = Z					$w_{12}^{(31)}$	$w_{12}^{(32)}$	$w_{12}^{(33)}$	$w_{12}^{(34)}$				
	4 = S												
2 = BUS	1 = C												
	2 = K												
	3 = Z												
	4 = S												

The formulae used in the computer program to calculate traffic quantities are:

$$\begin{cases} A_{bc}^{(t)}(kl) = \left[a_{b.}^{(t)}(k) \cdot d_{bc}^{(t)} \cdot C_b^{(t)}(k)\right] \cdot \dfrac{C_c^{(t)}(l) \cdot W_{bc}^{(t)}(kl)}{\sum_l C_c^{(t)}(l) \cdot W_{bc}^{(t)}(kl)} \\ A_{kl}^{(t)} = \sum_b \sum_c A_{bc}^{(t)}(kl) \end{cases}$$

k, l	=	1, 2, 3, 4
b	=	1, 2, 3, 4
c	=	1, 2, 3
t	=	0, 5, 10, 15

and

$$\begin{cases} A_{b \to Sp.S.}^{(t)}(k) = \left[C_b^{(t)}(k) \cdot a_{b.}^{(t)}(k) \cdot d_{b \to Sp.S.}^{(t)}\right] \\ A_{k \to Sp.S.}^{(t)} = \sum_b A_{b \to Sp.S.}^{(t)}(k) \end{cases}$$

k	=	1, 2, 3, 4
b	=	1, 2, 3, 4
t	=	0, 5, 10, 15

Start the forecast by studies and discussions on present values, especially, of course, $a_b^{(0)}(k)$, $d_{bc}^{(0)}$ and $W_{bc}^{(0)}(kl)$. Don't forget that in the first part of the process, LD-traffic is excluded!

After that, you can start working by computer. Whether you concentrate re-settings on the $d_{bc}^{(0)}$ table or on the $W_{bc}^{(0)}(kl)$ values is for you to decide.

A table to allow you to set the $W_{bc}^{(0)}(kl)$ weights now appears on the screen. The table is filled with default values (e.g., 3's). You should change the W values, as you see fit. Use, for example, values 1,2...5 where 1=very low; 2=low; 3=normal; 4=high; 5=very high. However, you could use any scale you like between 0 and 99, provided that each sub-row contains at least one value $\neq 0$.

Category		1 = RES				2 = BUS				3 = PBX			
	Area	1 = C	2 = K	3 = Z	4 = S	1 = C	2 = K	3 = Z	4 = S	1 = C	2 = K	3 = Z	4 = S
1 = RES	1 = C												
	2 = K												
	3 = Z												
	4 = S												
2 = BUS	1 = C												
	2 = K												
	3 = Z												
	4 = S												
3 = PBX	1 = C												
	2 = K												
	3 = Z												
	4 = S												
4 = CB	1 = C												
	2 = K												
	3 = Z												
	4 = S												

Now you should define present $d_{bc}^{(0)}$ values by using the following set of tables (appearing on your terminal):

		Traffic area				
Category		**C**	**K**	**Z**	**S**	**mean**
	R					
	B					
	P					
	C					

For your guidance

		Category			
Category		**R**	**B**	**P**	**S.S**
	R				
	B				
	P				
	C				

For your active work

Check that each rowsum = 100 (%)

(In the computer, d values are represented as fractions between 0 and 1)

$a_{b.}^{(0)}(T)$ = local originating traffic per subscriber include. special service, but excluding LD-traffic.
See table 3.7 in the main Data document.
Format: .xxxxx

$d_{bc}^{(0)}$ values are to be supplied by you.
The result appears below!

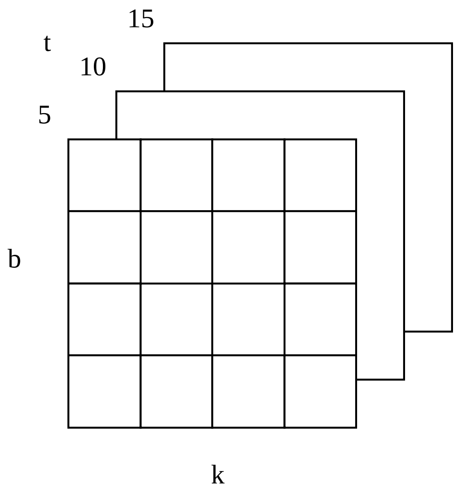
	Traffic area					
Traffic Area		**C**	**K**	**Z**	**S**	**mean**
	R					
	B					
	P					
	C					

Result : <u>Hypothetical</u> Present Traffic Matrix $A_{kl}^{(0)}$

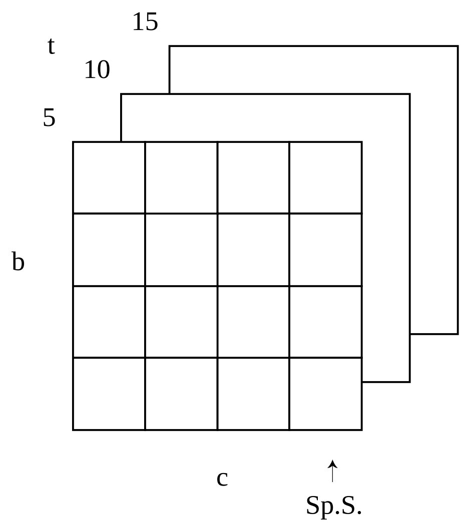
This table (a result of the given $a_{b.}^{(0)}(T)$ values and the set $W_{bc}^{(0)}(kl)$ and $d_{bc}^{(0)}$ values) is shown on the screen. You should now check the values against the "real" $A^{(0)}$ matrix (Table 3.6 in the main Data document). As a result, you may have to adjust the $W_{bc}^{(0)}(kl)$ values and/or the $d_{bc}^{(0)}$ values.

The process is consequently iterative!

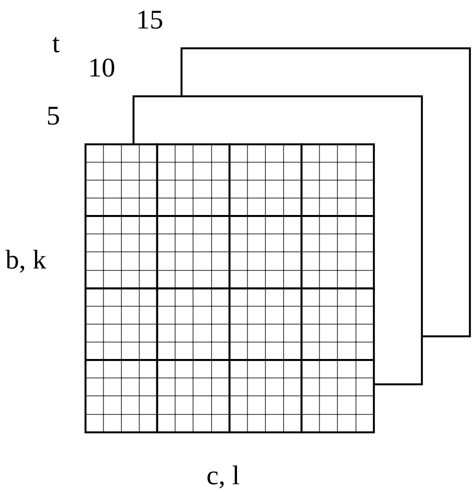
Now undertake the studies necessary to make decisions on future values, and then feed them into the computer.



$a_{b.}^{(t)}(k)$



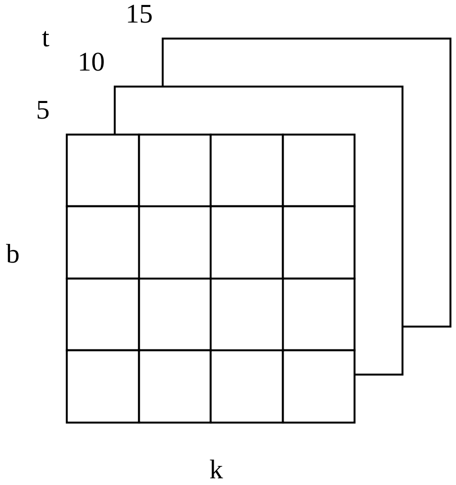
$d_{bc}^{(t)}$



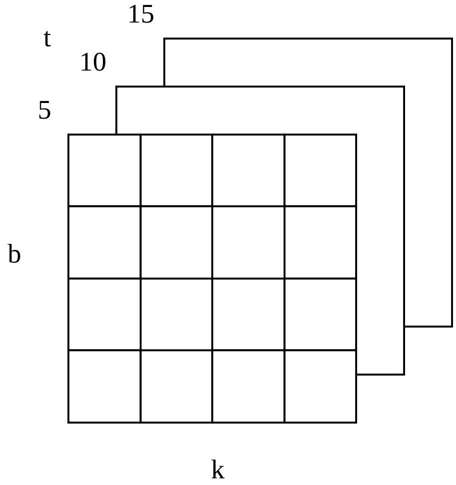
$W_{bc}^{(t)}(kl)$

One thing remains to be done - LD traffics!

Forecast future outgoing and incoming LD traffics per subscriber per category and traffic area, and feed these values into the computer.



Outgoing LD traffic/sub.
$a^{(t)}_{LDout}(b,k)$



Incoming LD traffic/sub.
$a^{(t)}_{LDinc}(b,k)$

Calculation formulae used in the program:

$$A^{(t)}_{LDout}(k) = \sum_b a^{(t)}_{LDout}(b,k) \cdot C^{(t)}_b(k)$$

$$A^{(t)}_{LDinc}(k) = \sum_b a^{(t)}_{LDinc}(b,k) \cdot C^{(t)}_b(k)$$

Result :

	Centar	Kozala	Zamet	Susak	Special Service	LD
Centar						
Kozala						
Zamet						
Susak						
LD					- - -	- - -

5
10
t 15

If you feel it is necessary to forecast total originating and terminating traffics per traffic area separately, you can do that and then use, for example, Kruithof's Double Factor Method for the reconciliation between those totals and the point-to-point values as per above. In that case, it is recommended that you exclude Special Service and LD traffics from the process. You have to do this manually, and the result should be fed into the data file. **GOOD LUCK!**

4. Survey Of The Forecast Process

SUBSCRIBER FORECAST :

* 1. Adjust S per zone;

* 2. Find $FP^{(t)}$ and $FE^{(t)}$ per traffic area;

3. **Output:** $C^{(t)}$ per zone;

* 4. Adjust individual $C^{(t)}$-values.

TRAFFIC FORECAST :

* 5. Set $W^{(o)}$ for combinations of categories and traffic areas;

* 6. Set $d^{(o)}$ for combinations of categories;

7. **Output:** Hypothetical $A^{(o)}$ matrix;

* 8. Compare hypothetical with given $A^{(o)}$ matrix;

* 9. If necessary, re-do $W^{(o)}$ or $d^{(o)}$, or both;

*10. Determine future values of $W^{(t)}$, $a^{(t)}$ and $d^{(t)}$;

*11. Forecast future LD traffics as calling rates per category and per traffic area;

*12. **Output:** Future traffic matrices.

---

* Requires special studies and/or discussions