Chapter VI

FORECASTING

Contents

Chapter VI

FORECASTING

Purpose of this chapter

The importance of forecasting in telecommunication network planning cannot be over-emphasised, reliable subscriber line forecasting in particular being crucial both for implementation planning and as a vital input for traffic forecasting in rural and metropolitan networks.

The aim of this chapter is therefore to describe the fundamental methods of forecasting which can be used for both subscriber line and traffic forecasts with an emphasis on defining the purpose of the forecast, identifying appropriate forecast indicators, identifying the zone in which the forecast is to be made and, above all, the segmentation into subscriber categories.

Outputs to be obtained

- Improved forecasting function in the planning group;
- Emphasis on forecasting in the planning processes of the organisation.

Inputs required

- Experienced personnel from other technical, commercial and operational departments to be transferred to the forecasting unit;
- Training in forecasting methods for relevant personnel.

Chapter VI

FORECASTING

6.1 Introduction

Forecasting is absolutely crucial for the successful planning of the telecommunication network. The best available planning, optimisation and dimensioning models would be completely useless without appropriate input data, so that the importance of forecasting which largely provides this input cannot be over-emphasised.

However, forecasting is at the same time a difficult procedure which demands a high degree of experience and skill and can also take considerable time and manpower to carry out. In addition, it is usually difficult to produce forecasts of adequate quality - by their very nature forecasts are almost invariably wrong and the question is usually how uncertain a particular forecast may be.

This chapter is arranged as follows:

Paragraph 6.2 (Forecasting principles) provides comprehensive background and guidelines on:

- the analysis of the actual forecasting situation;
- the selection models and methods;
- the design of inter-active forecasting schemes;
- improvement of forecast quality;
- the avoidance of common traps.

In addition attempts are made to encourage the use of logical and analytical methods, mathematical modelling; computerisation of data and methods; the use of all relevant data and models and the ignoring of all non-relevant information.

Forecasting methods for the specific cases of metropolitan, rural and national networks, together with the case of a network in a small country are given in §§ 6.3 to 6.7.

6.2 Forecasting principles

Generally, **Traffic** is generated by **Users** of the telecommunications services - traditionally called **Subscribers**. What is meant is rather main lines. We may still use the term, even if we nowadays can distinguish between several kinds of participants on the telecommunication market. This applies to all levels of the network: rural, urban, metropolitan, national, regional and also the world-wide international network level.

Since subscribers of different categories use telecommunications services very differently and also show different trends of development of the demand, it should in principle always be the best traffic forecast method to calculate the traffic intensities by carefully estimating the calling rates for each subscriber category and between all of the categories and then just multiply these values by the carefully forecast number of subscribers of each category in each location. The subscriber forecasts are needed in any case for planning, so from the traffic forecasting point of view they would be available without extra work. Furthermore, this principle would offer very good consistency between subscriber and traffic forecasts.

This is absolutely true concerning the lower levels: rural, urban and metropolitan networks. When it comes to the national level and upwards, however, we find that the subscriber forecasts are not needed in themselves for planning of the networks, since the switches on these levels as a rule are pure transit ones. However, subscriber forecasts are usually the most difficult ones to make, and we therefore generally forecast traffic directly for these higher levels. Furthermore, on these levels, traffic forecasting is a very special subject and almost always based on statistical demand analysis. Of course, one of the most important explaining variables still is the number of subscribers, so we can not ignore the subscribers completely.

Because of these circumstances, this subsection will apply in particular to subscriber forecasting, although the principles themselves are quite general. The other subsections comprise more detailed examples of forecasting methodologies for the rural, metropolitan, national and regional levels. For the first two cases, both subscriber and traffic forecasting are explained; for the last two, of course only traffic forecasting is considered.

The aim of subscriber forecasting is twofold. First of all, a considerable part of the planning of the telecommunication network must of course be based directly on the predicted subscriber distribution of the near or far future, but even the more traffic dependent network parts can not be properly planned and dimensioned without the supply of reliable subscriber forecasts, since these are important primary elements of the traffic forecasts.

Sometimes the subscriber prediction thus represents the final forecast product required; in other cases the subscriber forecast is just a necessary step in the traffic forecast process; but probably in most situations it has to play both these roles. It is therefore absolutely necessary that the purpose of the forecast is carefully considered before the process of problem analysis, method selection and forecast structure design starts.

The subscriber forecasting activity is indeed a quite heterogeneous process, characterised by a complicated interdependence between telecommunications development and private as well as professional life.

Great differences exist between subscriber classes, both concerning the development of demand for services and the way of using these services.

The economy, fluctuating between booms and depressions, the evolution of telecommunications, moving stepwise due to shifts between technical generations, and the rapid changes in social structures, are all factors which unfortunately imply trend shifts, thereby at least to some extent limiting the possibilities of using time series models alone.

Ideally, true econometric models, i.e. models of multi-equation simulation type should in principle be preferred, but the necessary degree of complexity for these models, lack of relevant data, and the need for forecasts based on lower geographical levels and for short time scales are circumstances that reduce the usefulness of such schemes.

Even the use of explicit causal and correlation methods is somewhat limited, which leaves us with the time series analysis, but mixed with judgmental or partly judgmental methods.

One way of improving the final results is to try and reconcile two or more completely separate forecasts, preferably based on different strategies. Some approaches are frequently used particularly in subscriber forecasting, e.g.

growth curves, and comparison methods. The variety of available information in a specific situation can often stimulate the introduction of new and original ways to forecast, in addition to the general standard methods.

All this emphasises the importance of starting up the forecasting activity by a careful selection of main strategies, and a choice of forecasting models and methods that agree with these strategies and with the structure of the forecasting problem and the available data.

It is of equal importance to work on the appropriate geographical and network level and with the correct time scale, as well as to treat different subscriber classes separately.

Fortunately, a first forecasting scheme may be very simple; with time and increased experience from follow-up studies, the ideal degree of sophistication may be reached.

6.2.1. Strategies and methods

The importance of reliable forecast results, especially for subscriber forecasts due to their crucial and double role in planning, necessitates a careful analysis of the STRUCTURE of the forecasting situation, the PROPERTIES of possible models, and the CAPACITY of the resources.

Suppose that we have studied a particular forecasting situation up to a certain point: We are absolutely clear about the purpose of the forecast, which should mean that we know exactly what is to be forecast, the required level of precision, and for what future point of time.

Further, we have an idea about the type, amount and quality of available numerical data or other kinds of information and also know at what point of time the forecast is needed as well as how much it may cost. Hopefully, we are aware of the amount and type of available resources, i.e. the personnel who can be used for the task, their skill, and the existing software tools - as well as those which could possibly be developed.

The initial procedure starts by finding a MAIN STRATEGY, which means that we try to find out reasonable qualitative relationships between the past, the present and the future, and between the forecast indicators and other variables.

Next we proceed a little closer to the real problem, by scrutinising the STRUCTURE of the forecast indicators and other variables in order to give preference to some main class of models.

Finally, we try to find out whether the structure of the forecasting problem is likely to remain unchanged in the future or not; this quality could be called STABILITY.

After these preparatory studies, we may start the exercise of method selection, provided that we know what kind of methods are available and how they relate to the actual STRATEGY, STRUCTURE and STABILITY, their other qualities, and how they could be applied and combined.

a) Forecasting as a complex activity

Almost all practical exercises in forecasting seem to create some new feature, as no application or situation is exactly similar. In addition, trend shifts of the subscriber demand development are frequent, which limits the usefulness of a number of methods. Therefore we can not expect that any kind of overall forecasting theory or method will be developed that will answer all the particular needs. On the contrary, forecasting is indeed an extremely complex activity and not at all a simple statistical problem.

Our forecasting scheme will seldom, if ever, be realised as one more or less complex model but rather as a complex of a number of quite simple methods or algorithms so that one very important element is then subjective judgement. Not only may some of the individual methods be of subjective judgement type, but the ultimate assessment of the future must be based on subjective judgement, and so must the whole forecasting framework.

Subjective judgement is constantly used in forecasting, for example to:

- analyse the requirements and the forecasting possibilities;
- design the structure of the forecast result;
- select methods;
- design an appropriate structure of the forecasting scheme;
- scrutinise given data;
- digitise non-numerical input data;
- set start values;
- adjust parameter values;
- assess and interpret intermediate and final results.

In fact, the entire forecast may be based on subjective judgement alone.

Subscriber forecasting in particular is probably both more complex and utilises also subjective judgement to a greater extent than other forecasting activities in the field of telecommunications.

The reason for an extensive use of subjective judgement in this field is of course not a belief that it would give more precise or reliable results than statistical methods; on the contrary, the performance is normally worse from that point of view. In many cases it simply has to be applied; it then becomes a fine art trying to benefit from its advantages and at the same time avoiding or diminishing the effects of the drawbacks.

b) Forecasting strategies

When we are at the point of analysing a particular forecasting application, designing the corresponding forecasting structure, and selecting the methods to be used in the forecasting scheme, we should first try to clarify the type of main philosophy we believe in and thus would try to apply. That is, we must choose a forecasting strategy in terms of assumed relationships between the present and the future.

This is particularly important in the case of subscriber forecasting since the field is then wide open for the application of all kinds of models and methods, even unorthodox ones, and for the use of both numerical data and non-numerical information.

There are different ways of defining forecasting strategies; one way being to distinguish between deterministic, symptomatic and systematic strategies.

The DETERMINISTIC strategy is based on the belief that there is a complete, or at least very close, causal relation between the present and the future. In other words, this strategy assumes that present conditions or actions determine the future outcome very closely. It would therefore pay to analyse the present very carefully and to try to calculate the future consequences. The actual model could be developed from or be tested against past data. We could also say the deterministic strategy in our case concentrates on the development of total demand by more or less sophisticated extrapolation into the future. Time series models are typical realisations of this strategy.

The SYMPTOMATIC strategy is based on the belief that there are signs in the present that indicate future development of the subject in question. The signs do not cause the development but are relevant early symptoms, especially for more or less dramatic future changes. This means that one should be aware of symptomatic, perhaps small changes that indicate turning points, often analysing the development of differences rather than that of total volumes. An example could be when a small, but rather fast developing increase of the demand for data modems may imply a potential large demand of data connections in the near future, which could be a motive for the implementation of a separate public data network.

One way to start such an analysis would be to assume certain different courses of development and try to imagine what kind of early symptoms would occur.

The SYSTEMATIC strategy is based on the belief that the courses of development follow rather strict laws or principles, such as socio-economic theories, only that the real world is so complicated, that until we have found out the underlying regularities, many changes seem to be purely accidental. There is always a way to discover these laws, namely to ignore or black out a part of the reality, which is the same as introducing a degree of abstraction in order to build a model of the reality.

An interesting point is that our strategy could be both deterministic and systematic at the same time, in so far as we could strongly believe in explicit causality according to the deterministic strategy, but believe that the explaining variables themselves develop according to complicated interdependencies of a systematic strategy.

Since there is no such thing as the perfect or even correct forecasting model, we should often try to follow two or more alternative forecasting courses and then compare and possibly reconcile the different results. When we do that, we may as well also try to apply different forecasting strategies.

c) Forecast models and methods

When we try to classify different forecasting methods we will find that the boundary between them is often rather vague. Even a “simple statistical method” is after all not so simple, when it comes to the actual use of it in a practical case, since quite a lot of subjective judgement is involved in that process, e.g. method selection, result interpretation, etc.

In fact, subjective judgement is involved in most of the available methods. We may, however, distinguish between ESSENTIALLY statistical methods and ESSENTIALLY non-statistical ones.

Each method may be characterised by three main elements: input form, processing procedure, and output form. Any method must always provide a numerical output. The essentially statistical methods process data employing statistical principles and therefore require a numerical input. The essentially non-statistical methods are based mainly on other principles, such as judgement, simple extrapolation, etc. They may still comprise ELEMENTS of statistical processing. The input is numerical, non-numerical or mixed. We may thus think of statistical methods as being quantitative and of non-statistical methods as being qualitative.

The choice of method depends to a great extent on the available data. A good set of relevant and reliable numerical data indicates generally that some kind of statistical methods should be applied, since even very simple statistical models normally outperform non-statistical approaches, especially pure subjective judgement, even if the judges happen to be very experienced in the field.

General weaknesses of the judgmental methods are:

- Large amounts of data increase the forecaster’s CONFIDENCE in the forecasts, but their QUALITY may even decrease;
- Probability and similarity are often confused - the size of a sample is neglected;
- There is usually too much belief in specific and circumstantial data;
- Man is conservative - future changes are often underestimated.

Even for a case where a good set of data is available, subjective judgement should however be utilised in a complementary way, so as to improve the quality of the forecast; things may be seen from different angles and important additional factors may be brought in, etc.

i) Non-statistical methods

Non-statistical methods are more dominant in subscriber forecasting than in traffic forecasting and can be sub-divided into two classes, viz. JUDGMENTAL METHODS, and COUNTING METHODS.

The JUDGMENTAL METHODS are those in which the forecasters or their stand-ins play the active part. Such methods are:

- NAIVE EXTRAPOLATION, where simple assumptions about the development during the coming time period, or simple, subjective extensions of the consequences of current events are applied: e.g. the assumption that 80 % of the households in new flats to be built in an area during a two-year period will demand a telephone line.

- NAIVE COMPOSITE, where detailed, local estimates are collected and aggregated, usually adjusted for biases and expected changes. The ordinary local short-term subscriber connection forecasts belong to a great extent to this category.
- JURY OF OPINION, or EXPERT METHODS, where a group of experienced people, not necessarily forecasters, but preferably related to the field, consent on the expected future outcome of a subject. One very typical example is discussion on possible, or perhaps desirable, subscriber demand saturation levels for different types of area.
- SCENARIO METHODS, from which the assumed future is revealed trough a series of snapshots. Often entirely qualitative, the method should still not be neglected as an overall element even in the field of subscriber forecasting, since new courses of development or the future use of new services may be indicated by its use.
- DELPHI TECHNIQUE, where the estimates of the future are made by an expert group as a successive series. Each member of the group makes his own, independent estimate, but at each step in the series he uses a summary of the whole group’s result from the preceding step to influence his own new estimate.
- COMPARISON METHODS, where the forecasts are based on past events or conditions which are analogous to the present or future situation; usually the analogy is taken from an area which is more developed than the one considered, the idea being that the latter will follow about the same course of development that has already taken place in the former one. The use of a comparison method alone will lead to poor forecasting. Nevertheless, comparison methods are frequently used for subscriber forecasting in developing areas.
- The COUNTING METHODS, for example MARKET RESEARCH METHODS, are mainly those where the customers are the primary source of information. The strength of these methods is of course that the customers’ opinions are directly focused, which in principle should reflect the future needs very well. A serious weakness is that the customers seen as a group are conservative and narrow-minded, usually overestimating the future need for traditional, well-known services and tending to neglect new ones. Methods of this type are:
    - MARKET TESTING, where the response from the representative samples of customers to for example new but available services is analysed and extrapolated to estimate the future need.
    - MARKET SURVEY, where a representative sample of potential customers are interviewed concerning their attitudes and intentions to utilise existent or future services. Again, the gathered data must of course be adjusted for biases.

ii) Statistical methods

The statistical methods are usually classified as belonging to two main classes: time-series methods, and casual methods which are forecasting methods based on explanatory variables other than time. The former class comprises methods such as moving averages, exponential smoothing, adaptive filtering, time series extrapolation, time series decomposition and ARIMA models; and the latter such methods as correlation methods, regression models and econometric models.

Referring to the forecasting strategies defined earlier, we may alternatively use another classification, namely the essentially DETERMINISTIC models, where the data are supposed to be describable by mathematical functions plus random variations; and the essentially STOCHASTIC models, where the stochastic relationships between neighbouring observations are the basis for defining the models.

The DETERMINISTIC model family comprises all the traditional trend models, seasonal models, growth models and regression models. One common property of these models is their relative simplicity but also the general structure of the analysis, where the forecaster starts by creating or selecting a model in most cases based on his subjective analysis of the actual forecasting situation. Thereafter, the available data is used for the estimation of the parameters of the selected model.

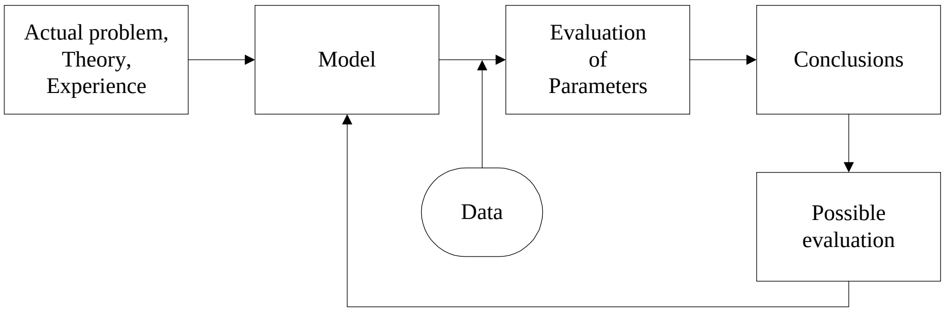


Figure 6.1: Deterministic analysis

The STOCHASTIC approach includes typically the ARIMA models, where the data are analysed as a first step. The result of the analysis is a proposed possible model and possible parameters. Then a decision must be made on whether the model is realistic or not.

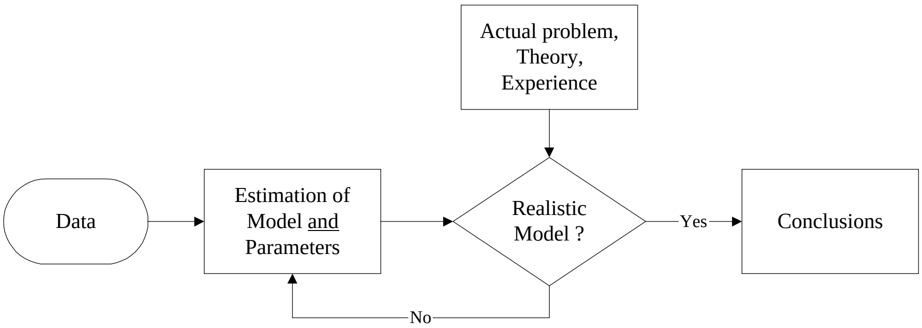


Figure 6.2: Stochastic analysis

The methods based on stochastic models are generally more sophisticated compared to deterministic methods. They often perform extremely well, but on the other hand they usually require extensive computer programmes, great skill, and fairly large quantities of stable data.

The choice of approach has sometimes more to do with forecasting resources and available data rather than with the performance of the different models.

The most common TIME SERIES METHODS are the following:

- MOVING AVERAGES, in which the future outcome is calculated as the average of a number of the most recent values.
- EXPONENTIAL SMOOTHING, where the forecast is based on a weighted combination of the forecast for the previous time and the most recent outcome.
- ADAPTIVE FILTERING, in which the forecast is based on a weighted combination of actual and estimated outcomes.
- TIME SERIES EXTRAPOLATION, where the forecast is calculated as a future extension of a least squares function fitted to a data series where time is the explaining variable.
- TIME SERIES DECOMPOSITION, where the outcomes from trend, seasonal cyclical and random components are estimated separately from a data series.
- ARIMA METHODS (Box-Jenkins analysis), which are complex, iterative procedures and produce autoregressive, integrated moving average models, make adjustments for seasonal and trend factors, estimate weighting parameters, and test the model.

The full advantages of the last two, time series decomposition and ARIMA methods, are generally apparent in the field of traffic forecasting, but all the methods are used also in subscriber forecasting, even if causality is supposed to be more important in the latter case.

A class of model which must not be forgotten is the GROWTH CURVE, where the development of demand is supposed to accelerate in the beginning, but slow down with time when much of the potential demand is satisfied. A hypothetical saturation level is then either calculated from historical data or simply imposed using more or less subjective judgement.

Growth curves are used particularly in subscriber forecasting for a number of reasons, one of them being that the use of a limited amount of historical data may be combined with subjective judgement, for example by adjustment of the saturation level both in time and magnitude.

The main CAUSAL METHODS are:

- CORRELATION METHODS, where the forecast is based on historic covariation patterns between variables.
- REGRESSION MODELS, where the predictive equation is based on minimisation of the residual variance of one or more explaining variables.
- ECONOMETRIC MODELS BASED ON SIMULTANEOUS EQUATIONS, where integrated systems of simultaneous equations of relationships between economic and other parameters are solved.

Sometimes, both regression models and econometric models based on simultaneous equations are simply called econometric models, the former comprising single-equation regression models; the latter may then be named multi-equation simulation models, or true econometric models.

The ECONOMETRIC models may be counted as a third class of models in addition to the DETERMINISTIC and STOCHASTIC classes, even if the boundary between statistical time-series models and economic theory has grown rather vague.

The causal methods are particularly attractive for subscriber forecasting. The choice of explaining factors is however different for different subscriber classes. The supply of such explaining factors is a general problem for the appropriate use of causal methods.

d) Use of models

In 6.2.1 b) we started on a philosophical level, identifying the possible main forecasting STRATEGIES: the deterministic, the symptomatic, and the systematic approach.

In 6.2.1 c) we classified the different MODELS of the forecasting problem, which lead us to choose from mainly two different model approaches: the DETERMINISTIC and the STOCHASTIC.

The ECONOMETRIC models were placed in a third class, but could perhaps be considered to be closely connected to the deterministic models.

In this subsection, the question is whether we can consider that the structure of the forecasting problem is likely to remain stable or not, over the forecasting period. In this respect, stability, we may distinguish between two approaches, the GLOBAL approach and the LOCAL approach.

i) Global versus local approach

The GLOBAL approach is based on the assumption that the corresponding model, the GLOBAL model, will describe the situation in a satisfactory way over the entire period of time, both the past and the future. The stochastic models and the econometric models all rely on this assumption. The general procedure consists of first of all, model identification and fitting using initial data, then repeated use of the model to obtain forecasts. The stability problem is covered by some form of quality control procedure applied to the forecast errors.

The LOCAL approach assumes that no situation or structure is likely to remain stable over any length of time. Therefore, the parameters of the fitted model must be permitted to change with time, when new data arrive. Even the model itself could be changed. The LOCAL approach is used in fitting deterministic models, the result being the ADAPTIVE methods in forecasting.

ii) Forecast processes

The total forecast process may be subdivided into the consecutive parts Identification, Fitting and Forecasting.

If we define subjective judgement and activities in which subjective judgement is a strong element as “Art” rather than “Science”, then the identification activity is a typical “Art”, even though there are methods for identification within classes of models, e.g. the Box-Jenkins approach, and among growth curves and regression models.

The FITTING activity consists much more of “Science”, even if there is a wide range of criteria of goodness of fit to choose from, not only the method of least squares, but e.g. the use of discounted methods; the forecaster may define his own criteria of fit using a cost of error function related to the actual application of the forecast.

The FORECASTING activity is from one aspect almost pure “Science”, namely to produce an unbiased result based on the use of the already chosen and fitted model, but the very last phase, being a matter of acceptance, rejection or modification is again rather an “Art”.

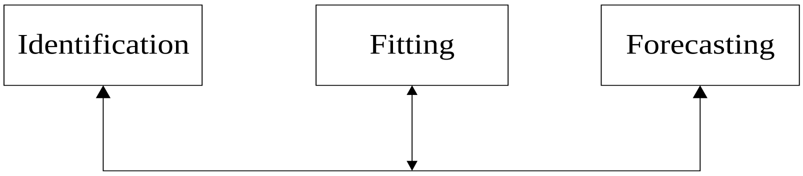


Figure 6.3: The forecast process

The fitting activity is quite general in statistics and more or less common to the different fields of application and will therefore not be treated here, but the identification activity and parts of the forecasting are specific for the different kinds of forecasting. The identification activity includes the method selection as well as the design of a complete forecasting scheme.

e) Improving the forecast

We always try to produce the best possible forecast, or at least a sufficiently good one. The available models have however been the same for several years, no dramatic developments in technique have taken place recently.

Thus, the efforts to improve the forecasts have concentrated on the problem of model selection. Such attempts have highlighted the properties of various methods: strengths, weaknesses and performance characteristics.

Using this kind of knowledge, we may thus select methods which fit in with the specific forecasting problem as well as possible, but we can improve the forecast quality even further by combining different forecasts in an effective way, by using subjective judgement as a complement to mathematical models, or by simulating a range of input assumptions.

i) Combination of forecasts

There is actually quite an extensive amount of research results available on the subject of combining forecasts to achieve improvements in which it is shown that the results of combined forecasts can be far better those produced by single techniques and expert analysis.

The best experts and the most sophisticated models can not outperform an effective scheme which combines results. Furthermore, the selection of a particular technique or the engagement of a certain expert to forecast the future is in itself a prediction - we predict that the model or the expert will produce a good forecast; therefore the combination of forecasts really offers a way of improving the forecast quality.

The process of combining forecast models may be a very simple one. If we scrutinise the different available models according to the circumstances of the specific forecasting problem, we can quite easily see what the strengths and weaknesses of the models are from different aspects, such as input data requirements, accuracy, ability to detect or reflect trend shifts, etc.

By matching two or more models with different properties, we may thus reduce the effects of the weaknesses of one model by the strengths of the other, but still retain the strengths of the first.

ii) Simulation of a range of input assumptions

By varying the input assumptions we can carry out a sensitivity analysis of the forecast. A whole range of probable outcomes are then established by varying the values of the input parameters of a particular model. The result of such a simulation is that the most critical parameters are highlighted, and that the range and the distribution of expected results are surveyed. One effect may be that we then can concentrate on the analysis, scrutinisation and possible adjustment of critical input data and simply accept non-critical parameter values.

Sensitivity analysis, based on this kind of simulation is very important both for long-term planning, in order that flexibility and robustness of future possible network strategies might be tested, and for short and medium-term planning as well, since the investment decisions should be taken in the light of the accuracy of the forecast.

iii) Selective application of subjective judgement

Quantitative forecasting often still includes elements of subjective judgement to some extent. In general, however it is recommended that reliance should be placed on the results from quantitative forecasts rather than on our own judgement, provided of course, that the method is reasonably relevant to the forecasting problem and that input data of sufficient amount and quality exist. Even very simple quantitative models outperform the assessments of top-rated experts. Judgmental adjustment of output values from a quantitative forecast reduces its accuracy, the reason being that the experts' ability to process information and control the relationships between variables is very limited, and that all intuitive forecasts are more or less biased.

However, in some situations subjective judgement should be used. Typical cases are when trend shifts can be expected, or when relationships between variables will change in the future. We should then at least try to use subjective judgement on the input side of the forecasting method rather than adjusting the output data. The reason for this is that on the input side data are more logically structured, which means that we can examine and adjust different data separately, but the output data are the result of a more or less complex process. Adjustment of output data therefore means that we interfere with the statistical processing in an uncontrolled way.

Other cases when subjective judgement should be brought in are when data are very limited or do not exist; for example in situations such as forecasting of future services to be introduced or when the forecast horizon is very far away.

The application of judgement should then be made on a structured basis, which in addition to the immediate higher quality probably obtained, also enables future updating in connection with follow-up studies to be carried out in a controlled and regular way.

The use of subjective judgement is much more frequent in subscriber forecasting than in traffic forecasting, and since the former is the basis for the latter, the design of the subscriber forecasting scheme is crucial for the whole planning process.

6.2.2 Choice of forecast approaches

The objective is the design of a forecasting scheme which functions well. In some situations, a single method will fulfil the needs but usually a much more complex pattern will be required. The main parts of such a scheme can be defined in different ways; one way is to consider the logically structured complex of selected methods and algorithms as being a central block, operating on different blocks of input data, being controlled by the user via a set of subjective judgement and decision points, and producing the final forecast.

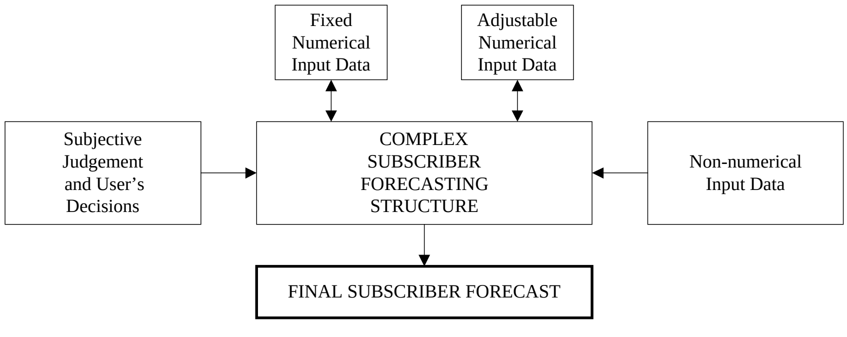


Figure 6.4: Outline of forecasting scheme

The users' block should then preferably contain pure subjective judgement and decisions only, so that all mathematically defined algorithms are included in the central block, even if they should happen to be parts of, for example, judgmental methods. The advantage of this type of structure is that the interface between the user and the central method block may be clearly defined, which gives a good basis for the design of an interactive computerised forecasting module. For each connection point between these blocks it should be stated whether the user's judgement is necessary, desirable or just possible; the importance of the decision and the possible sources of background information should be given as well.

The different types of input data may need some explanation. By fixed numerical indata is meant data which have to be taken at face value. An example may be official population forecasts for entire areas where the degree of uncertainty is very small. By adjustable numerical indata we may mean values that are proposed as being quite likely according to for example current trends, to limited measurements, or possibly to some kind of standard classification. One example may be a proposed, limited set of standardised future saturation levels of telephone penetration for populated areas, where the individual areas are classified as belonging to certain types.

The degree of adjustment in that case may be, for example, either that the areas may be reclassified, or that the proposed saturation levels are individually adjusted for all areas or for particular ones, before or after the preliminary calculation according to some forecast model.

By non-numerical indata is meant all kinds of verbal information such as, for example, general topological, social or economic descriptions of areas, or scenarios of future telecommunications. Such information could be used in different ways, either being digitised in an early phase to be used for parameter values in a regular calculating algorithm, or being used as it is, as a part of the judgement factors when a preliminary forecast result is analysed.

Bearing in mind the main structure of the forecasting scheme, we may however now focus on the delicate problem of selection of models to be used in the central method block.

a) Selection principles

Although the choice of a particular subscriber forecasting model to a great extent should be guided by its forecasting performance, other qualities and circumstances must be considered as well.

Important factors are, among others, the length of the forecast period, the functional form of the model and, in the case of causal methods, the accuracy of the historical and forecast values of the explanatory variables.

The length of the forecast period affects the decision to give preference to a particular model before others, along with the supply of historical data and the purpose of the forecasting. For instance, time series models may be appropriate for short to medium term forecasts when stability is no problem, when sufficient historical data are available, and when the causality is not of interest.

The development of the subscriber demand is usually considered as being an ideal subject for the application of causal models, but if the structure which generates the data is difficult to identify, there is usually no choice but to use a forecasting model which is based on historical data of the variable of interest.

The functional form of the model should also be considered. While it is true that a more complex model may reduce the model specification error, it is also true that it could considerably increase the effect of data errors.

Availability of forecasts for explanatory variables and their reliability record is another factor affecting the choice of a forecasting model. A superior model using explanatory variables which may not be estimated accurately can be inferior to an average model whose explanatory variables are accurately forecast.

When market stability is a problem, causal or association models which can handle structural changes should preferably be used.

When causality matters, simple models or ARIMA models may miss the target if they are used as single forecasting tools. Another case when the time series methods may fail is of course when historical data are insufficient.

Sometimes a particular method might be chosen simply because it is nicely presented in some article, and once chosen and understood by the forecaster, be invariably applied in all situations, irrespective of the different conditions and demands. In the long term, this will probably lead to poor forecasting; a better way is to reconsider the forecasting methodology continuously, i.e. to evaluate the different methods in the light of the changing situation.

At first glance, this matter seems to be quite simple; methods should be used which fulfil the specific requirements in the actual forecasting situations at the least possible cost. This means that it should be a problem of balancing the properties of the different models against the requirements of the situation. However, to work well, each method has to be considered separately for particular situations such as a certain minimum level of knowledge and skill of the user, particular computer and software possibilities, a certain amount and quality of input data, etc.

The selection can be made in many different ways, but in general a systematic approach should be used. In any case each possible method should be studied in detail so as to show clearly what the advantages and weaknesses are in general as well as for specific applications. Thus the problem can be considered as a matrix with all kinds of users' crucial questions on the input side, the range of available models on the output side, and the properties of the different models inside the matrix.

<table>
<tr><th colspan="2" rowspan="2">DIMENSIONS</th><th rowspan="2">QUESTIONS</th><th colspan="3">METHODS</th></tr>
<tr><th>- - - - -</th><th>Time Series Extrapolation</th><th>- - - - -</th></tr>
<tr><td>Time</td><td>Span
Urgency
...</td><td></td><td></td><td></td><td></td></tr>
<tr><td rowspan="3">Resource

Require-ments</td><td>Mathematical Sophisticatio n</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Computer</td><td></td><td></td><td></td><td></td></tr>
<tr><td>.....</td><td></td><td></td><td></td><td></td></tr>
<tr><td rowspan="4">Input</td><td>Antecedent</td><td>Are only limited past data available?</td><td></td><td>Past data are essential</td><td></td></tr>
<tr><td>.....</td><td></td><td></td><td></td><td></td></tr>
<tr><td>External Stability</td><td>Are significant shifts expected among variable relationships?</td><td></td><td>Can not validly reflect shifts</td><td></td></tr>
<tr><td>.....</td><td></td><td></td><td></td><td></td></tr>
<tr><td rowspan="3">Output</td><td>Detail</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Accuracy</td><td></td><td></td><td></td><td></td></tr>
<tr><td>.....</td><td></td><td></td><td></td><td></td></tr>
</table>

Figure 6.5 : Method selection matrix

With this structure in mind, we can do two things: we can select the single models by checking the model properties against the user's requirements and the user's resources against the models' requirements; and we can also combine two or more models so as to cover the users' requirements as well as possible.

For this, we must begin by specifying the requirements and the resources for the actual situation. We may distinguish between:

- GENERAL CONDITIONS, closely related to the forecasts methods, like time requirements, resources, input and output:
- SPECIFIC REQUIREMENTS, related to the aim of the forecast, the time horizon and the geographical levels;
- PARTICULAR CONDITIONS, e.g. extremely rapid growth, long waiting lists, limited funds, etc.

More specifically, we may work according to the following logical scheme in which the users' requirements and resources are designated as DIMENSIONS:

1) We consider possible relationships between the present and the future and between variables of interest and other factors, to find a sort of ideal STRATEGY.
2) We analyse the specific forecasting problem to see what type of STRUCTURE it has.
3) Now we judge whether this structure probably will remain the same or will be changed during the forecast period, i.e. we define the STABILITY.

4) Next, we use the specification of STRATEGY, STRUCTURE and STABILITY to make a preliminary classification the possible usefulness of the available methods. Example: VS = Very Suitable; MS = Moderately Suitable; LS= Less Suitable
5) Again we use these specifications, together with what we know about the GENERAL CONDITIONS, the SPECIFIC REQUIREMENTS, and THE PARTICULAR CONDITIONS, to find out which of the dimensions are crucial for the method selection. This means that each dimension must be connected to a QUESTION that can be answered with YES for cruciality and with NO otherwise. In this way we will obtain the set of crucial dimensions for the particular forecasting situation.
6) Now we can compare the crucial dimensions with the corresponding properties and requirements of the different available methods, starting with the VS-class. As an example, the degree of fit between dimensions and method properties could be classified as G = Good fit; A = Average fit; and B = Bad fit. The result will be a set of fit results for each method.
7) By comparison between the set of dimensions and the criterion of fit, we may now be able to give preference to one or some of the methods.

Hopefully the selected models agree with our main forecasting STRATEGY (deterministic, symptomatic or systematic), with the STRUCTURE of the situation (deterministic, stochastic or econometric approach), and with our views on the STABILITY(global or local approach). If they do not, we have to do our best to compensate by combining different forecasts, simulating input assumptions, or adding subjective judgement to the process.

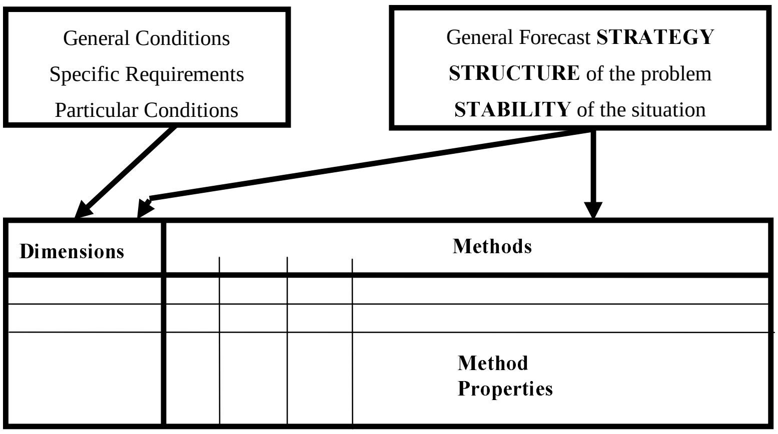


Figure 6.6: Identification of crucial dimensions, and preliminary classification of methods

Crucial Dimensions	Methods Class VS			Class AS	Class LS
			- - - - -		
	G	B			
	G	B			
	G	G			
	A	A			
	B	A			
	. . .	. . . .			

Figure 6.7: Method of property fit versus crucial dimensions.

With this structure, the method selection procedure may be computerised, since the properties of any number of models and methods, for any number of dimensions, can easily be stored in a data file.

b) Factors to consider

In any forecasting situation, there are some conditions or factors which are absolutely fundamental for the design of an effective and appropriate forecasting scheme and crucial for the arrival at a useful result -and consequently necessary to carefully consider before any other activity starts:

- The real PURPOSE of the forecast is the key to the situation;
- Suitable forecast INDICATORS are important not only in order to compose the desired forecasts results, but for the forecast process itself;
- The ZONING must be both realistic from a data collection point of view, and objective so as not to bias the network planning;
- TIME, LEVEL and SEGMENTATION may be applied dynamically in order to positively affect the forecast quality.

i) Purpose of the forecast

The first crucial question which should be asked is: what is the exact purpose of the forecast? Once this is answered, many of the remaining questions concerning the selection of the most appropriate forecasting method can also be answered. The definition of the purpose can of course vary considerably from case to case.

Let us take one example, where the purpose of the forecast is to produce the necessary input data for a long term planning of the telecommunications network of a particular, large city. The result of the long-term planning study is in its turn meant to be the necessary basic material for discussions and preliminary decisions on the choice of technology, network topology, routing strategy, including economic considerations, for a target network 15 years ahead, but also to show the necessary intermediate steps. The discussion will take place in five months time from now, but the network planning group will spend say, two months to prepare the alternative solutions. This leaves us a maximum of three months for the forecasting activities.

First of all, we understand that a subscriber distribution forecast as well as a traffic interest matrix forecast is needed. Consequently, we can not spend the three months entirely on the subscriber forecast. Subscriber forecasters must rather discuss the situation with the traffic forecasters immediately, in order to define what kind of subscriber demand parameters they need for the traffic forecasting, i.e. the type of indicators, the level of detail, the accuracy, and the form, e g if some kind of probabilistic intervals are necessary, and of course also the urgency.

After that, we must consider the part of the subscriber forecast that is to be used as a direct input to a network planning study. We know that the target network will be used so that this is a typical long-term forecast.

The urgency is high, and the frequency seems to be low since it is a one-time task so far; these facts could imply that the introduction of new methods or development of software may take too much time and be too costly; possibly a combination of different simple methods would be adequate. On the other hand, that would indicate a more complex forecasting structure, which nearly always leads to an increased amount of subjective judgement, and to the use of various types of information.

The question of the form of the forecast result is interesting; in our example the planning alternatives are going to be used as types of scenarios, reflecting the consequences of development trends that after all are quite uncertain; therefore we should probably try to produce at least an optimistic, a pessimistic, and an average forecast, but possibly we should also indicate some kind of confidence intervals.

However, it is our job to make the results as useful as possible, which also means that they must be quite simple and not contain ambiguous or too complicated information; if the forecaster cannot himself draw the conclusion from some unstructured mass of output data, it would of course be ridiculous to leave this task to the receiver of the forecast; he would not be able to use such data at all.

The purpose, together with the form, also affect the required quality; the forecast should of course meet some standards of validity, accuracy and credibility, but the precision is also important; should the precision be as high as possible, or could we work with a minimum precision level to be deducted from the definition of the purpose of the forecast?

We must always keep the purpose of the forecast in mind. Using the previous example, let us illustrate how we could otherwise very well produce a forecast of high precision that is both accurate and credible, but still NOT VALID. Say that we are calculating the future subscriber distribution in the area, and that we are doing it well, using the available data in the best possible way, applying all sorts of significance tests, and combining methods in order to improve the quality; but that we do not distinguish between subscriber classes, or that we specify such classes that do not reflect some future radical changes of the subscriber mix in a particular areas. Our output is maybe in itself correct, but if the purpose of the subscriber forecast is to create input data for the traffic interest forecast, then it is unfortunately invalid, since it most certainly will lead to erroneous traffic figures.

ii) Forecast indicators

The number of subscriber connections is measured by the number of main terminal lines i.e. the number of effective access possibilities to the general network and the use of well chosen forecast indicators will greatly facilitate a correct measurement of the number of these connections.

There are two main reasons why we must choose forecast indicators carefully. First of all, the purpose of the forecast tells us what the variables of interest are; secondly, the development of different indicators may follow very different courses and be dependent on explaining factors in quite different ways. The basic indicators are defined as follows:

TRANSFER

This is linked to the mobility of customers and mainly concerns households. Transfer takes place when a subscriber changes residence and wishes to continue to have the terminal in his new home. He requests the transfer of his line, which leads to two actions: the outgoing transfer which corresponds to giving up the original line (as in a cancellation), and the incoming transfer corresponding to a new request.

CANCELLATION

This occurs when an existing customer no longer wishes to subscribe to telecommunication service.

NEW DEMANDS

These are requests for connection which do not arise from transfers. The new request is therefore an indication of the “natural” increase in the number of subscriber connections.

DEMANDS IN HAND

These are requests for connection which have not yet been satisfied. They may be either transfers or new requests. The time spent waiting for a line connection may be “normal” (an unavoidable minimum of several days in some countries), or longer (several weeks or months, or even several years) where demand is very much greater than supply.

EXPRESSED DEMAND

While the number of connections is a measure of the actual size of the network, the factor of greatest interest for forecasting purposes is the number of lines requested by customers, i.e. the size of the network which would result if the supply corresponded permanently and exactly to the demand. It will be assumed that in the past this quantity was adequately approximated, at a given instant, by the sum (L + I) of connected lines and requests in hand. The connected lines (L)represent the satisfied expressed demand, and the requests (I) represent the unsatisfied expressed demand. The forecast of the demand may also be interpreted in terms of (L + I), the precise object of the planning exercise being to dimension the number of available lines so as to minimise the number of requests in hand.

The expressed demand depends in practice on the effectively available supply. If the latter evolves significantly, it may lead to a correlated evolution in the expressed demand without necessarily affecting the real potential demand. Thus, if the supply is far smaller than the real demand, potential customers will only be inclined to submit a request when connection becomes possible (e.g. when a new exchange is created). In this case, and if the operating Administration wishes to catch up with the backlog of requests, the forecast indicator (L + I) will have to be abandoned in favour of an indicator representing more accurately the actual demand and which is therefore capable of providing more realistic results.

iii) <u>Mechanism for developing the number of connections</u>

The diagram in Figure 6.8 shows the movements which affect the number of connections (L + I).

The total numbers of connected lines and demands in hand evolve under the influence of movements occurring between two points of time t and (t + t') according to the following formulas:

L = SDNL + SDIT - CAN - OT
I = DRNL - DWNL + DRIT - DWIT - SDNL - SDIT
L + I = DRNL - DWNL + DRIT - DWIT - CAN - OT

where

L = New connections
I = Total unsatisfied demands for connections
L + I = Net demand for connections
DRNL = Demand received for new lines
DRIT = Demand received for incoming transfers
DWNL = Demand withdrawn for new lines
DWIT = Demand withdrawn for incoming transfers
SDNL = Satisfied demand for new lines
SDIT = Satisfied demand for incoming transfers
CAN = Cancellations
OT = Outgoing transfers

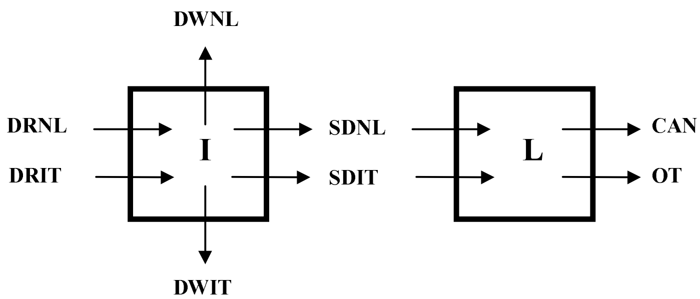


Figure 6.8: Relationship between the indicators

The number of connections (L + I) is therefore increased by new demands for connection and incoming transfers and reduced by cancellations by existing subscribers, outgoing transfers and demands withdrawn for incoming transfer and new lines.

According to the context and the actual number of connections, each of the above factors will have a greater or lesser relative importance. In periods of high growth rate, the increase in the number of connections (L + I) will be due mainly to new requests, with a residual volume of transfers and cancellations. In a more stable context, the volume of new requests will be relatively small, merely linked to the population growth; it will then become more important to follow migratory movements which will generate requests for transfers.

iv) Zoning

The planning of any telecommunications network implies some main basic activities, viz. calculation of suitable LOCATION of units such as terminal and transit exchanges, remote subscriber switches and multiplexers, cabinets and distribution points; BOUNDARIES between switching units; ROUTING principles and ROUTE dimensions. Since all these variables of interest of course will change during the planning period, it is necessary to up-date the future subscriber distributions and traffic interests at frequent intervals.

In the case of a metropolitan network, for example, the planning must be based on the division into so-called TRAFFIC AREAS which ideally should be homogenous from the point of view of the subscriber mix in each area, since then both the subscriber distribution and the traffic interest matrix may be easily calculated from this basic area division, for any division into switch unit areas. This should be kept in mind when an administrative zoning system is considered.

The need for demand forecasts relating to increasingly smaller geographical units involves problems of two types: information availability problems and zoning problems.

Information availability problems

The forecasting system requires the following type of data for its operation:

- Social and economic data (population, number of dwellings, number of employees in manufacturing industries and service industries, size of companies, etc.) which illustrate the network development context;

- Telecommunication data (equipment penetration rate by customer category) giving an evaluation of the potential market.

The need for social and economic data at different geographical levels makes certain demands on the national statistical system. It calls for an administrative zoning system dividing the national territory into units of an appropriate size; as well as an institution responsible for the collection of the statistical data at the various geographical levels implied by the zoning system. It is clear that the possibilities for demand forecasting at the various levels will depend essentially on the existence of such an information system, its proper functioning and its specification. However, while some sort of administrative zoning system almost always exists, there is more uncertainty about the existence and the performance of an associated information system. In many cases (e.g. in many developing countries), the absence of any information system or the paucity of the available data - the more pronounced the smaller the geographical unit - impose severe constraints on the choice of a forecasting system. It may be necessary for example, to use only global methods (e.g. overall telephone density) or methods which can only be applied at the national level. In every case, the smaller the geographical unit studied, the sparser will be the statistical data available. Thus the possibilities for analysis will be reduced as a geographical precision is gained. It appears that the only solution to the information availability problem is for Administrations to establish their own data collection systems.

Zoning problems

As discussed above, the demand forecasts necessary for planning are associated with existing or planned service areas (of switching centres, of cabinets, etc.) and are therefore applicable to a system of telecommunication zoning. It would however, be irrational to base a forecasting system directly on such a zoning system for several reasons, the most important of which is that the telecommunication zoning system is not static but evolves with the network structure (the creation of new units necessitates the redefinition of service areas).

The external information system supplying the social and economic statistics refers to an administrative division of the territory. In particular cases these administrative divisions may have affected the network structure; thus, in certain countries in which network structure has been modelled on the administrative organisation, there tends to be an excessive number of economically unjustified exchanges.

In most cases therefore, for the sake of consistency and to follow the structure of the external information system, demand forecasting studies will be based on the existing administrative zoning system, which has the advantage of being fixed. In every case, studies will be carried out in terms of geographical units which make possible the reconstitution of the service area being studied.

This has important implications for the Administration’s statistical system in that it must be compatible with the external information system. Telecommunication data will have to be available based on:

- the administrative zoning system to enable forecasts to be made;

- the telecommunications zoning system with, in each case, an associated matrix enabling forecasts for the service area to be reconstituted from the forecasts for an administrative division.

For example, when making forecasts for a switching area the service area of the switching unit is divided into distribution areas as far as the telecommunication zoning system is concerned whereas the territory is divided into localities and economic forecasts are made per locality as far as the Administrative zoning system is concerned. Figure 6.9 shows how the two zoning systems may be superimposed.

In order to make forecasts for the service areas of distribution cabinets on the basis of available forecasts at the locality level, it is necessary to draw up the matrix as shown in Figure 6.10. The coefficients of this matrix may change in the course of time as they will vary with the network structure.

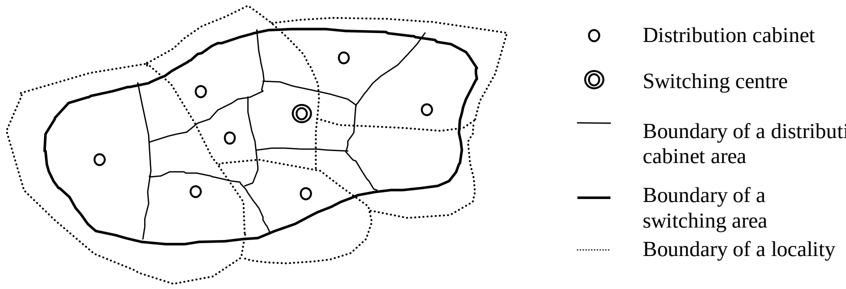


Figure 6.9: Administrative and telecommunications zoning systems

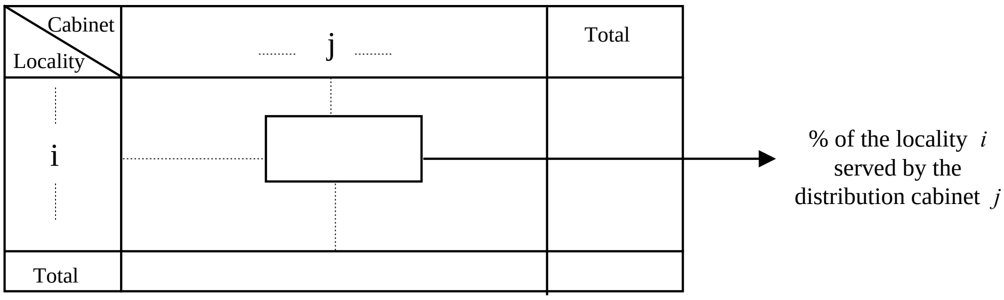


Figure 6.10: Matrix showing the correspondence between the two zoning systems

v) Time range, geographical level and subscriber segmentation

It may of course be possible to make a separate forecast for a case where all three of these factors are fixed. An example might be the forecast number of residential subscribers in a specific traffic area of a particular metropolitan network, exactly five years ahead.

In many cases, however, we actually move backwards and/or forwards, or possibly upwards and/or downwards, on the axes of time, level and subscriber categories.

Even when applying a single, very simple model for the development of the subscriber demand, such as for example the use of some version of growth curves, that movement still takes place, in that we move forward in time when we extrapolate from historical data into the future, but backward in that we let the imposed saturation level of the far future influence the more near future values.

Another, more obvious example of movement in time is when we make a long-term forecast as a basis for a target network and then let this long-term forecast play a limiting or smoothing role for the medium-term forecast.

Top Down - Bottom Up schemes are usually used in the context of geographical level, but we could as well use this approach also for time range and subscriber segmentation.

There are several reasons for the use of such schemes: we wish to obtain consistency between different forecasts; the best forecast model must operate on a level, time range or subscriber segmentation that is different from the final forecast variables of interest; we see the need to use several methods operating on different levels, whereby we can cross-check partial results by moving forward/backward and up/down.

Therefore the forecast procedure could be considered as a three dimensional system where we are free to move in any direction on all axes in order to arrive at a sufficiently good final forecast result. That means that we may apply the Top Down - Bottom Up Procedure to all three axes, whatever the requested level of the final forecast may be.

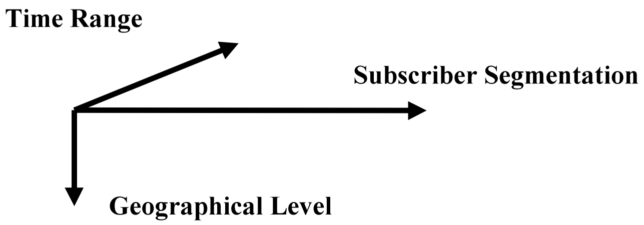


Figure 6.11: The three-dimensional forecasting system

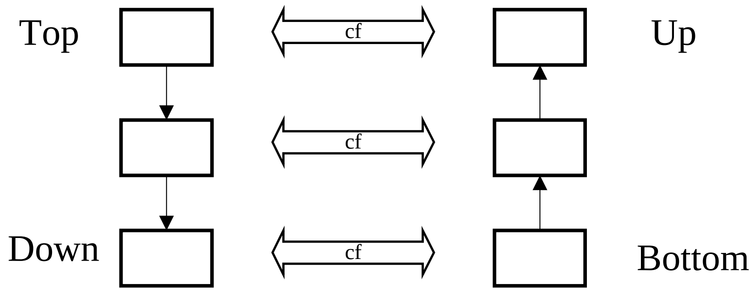


Figure 6.12: The top down-bottom up system

c) General conditions

Each method for consideration can be characterised by its specific qualities or properties, i.e. how accurate it is in general, whether it can detect direction changes, if it requires great skill and interactive use, if it depends on the supply of large amounts of input data, etc. All such qualities could be called dimensions, and the usefulness of a particular method should be judged in the light of the dimensions that are crucial for the specific forecasting problem.

If, for example, a high level of accuracy is not needed for a certain forecast task, then accuracy is not an important criterion for the choice of method in that particular case, and can probably be excluded from the set of dimensions to be used in the selection process.

Our goal is now to create a set of critical dimensions which is relevant for the forecasting problem in question and which can then be used in the model selection process.

Ideally, a specific question should be attached to each dimension, to be answered with YES or NO; YES would then indicate that the corresponding dimension is critical; NO would mean that the dimension can be neglected or at least seems to be of less importance.

We keep in mind that this very systematic approach to the method selection is particularly useful for subscriber forecasting, where we can often choose from a wide range of models and methods. However, as it is possible that none of the methods is really satisfactory, the resulting forecast method could be rather complex and contain some new approach as well as considerable subjective judgement. Traffic forecasting is from this point of view usually more straight-forward, but can of course present other problems.

i) Time requirements

DIMENSION: HORIZON

QUESTION: Is the forecast horizon short-term, medium-term or long-term?

This particular question can not be answered by a simple YES or NO. Furthermore the interpretation of short, medium and long term horizon will certainly be different for different applications.

DIMENSION: URGENCY

QUESTION: Is the forecast needed very quickly?

Again, this is a question where the answer depends on our standards. “Very quickly” may mean within an hour, the same day , in a weeks time, etc.

DIMENSION: FREQUENCY

QUESTION: Are updated forecasts needed more or less frequently?

Possibly the forecast must be updated, for example on a monthly or yearly basis. Then it is of course important that the applied model can be used in a systematic manner, so that comparative results from follow-up studies are preferably realised as modifications of parameter values for example.

ii) Resources

DIMENSION: SOPHISTICATION

QUESTION: Is mathematical skill limited?

Implementation, application and appropriate use of certain models and methods may require quite a high mathematical capability of the forecaster. If the personnel do not posses this capability, it is probably preferable to avoid the use of sophisticated models until the necessary skill is obtained.

DIMENSION: COMPUTERISATION

QUESTION: Are computer capabilities limited?

A modern computer is essential for the whole forecasting activity. It could be a microcomputer which works fast and can store large amounts of data, but even a personal computer would raise the level of the forecasting compared to manual calculations. Some methods, however, are in practice impossible to apply without a computer, the most sophisticated ones then requiring not only a standard software package but also extensive training of the personnel, e.g. the stochastic ARIMA models, and the true econometric methods, which are based on multi-equation simulation models.

DIMENSION: ECONOMY

QUESTION: Are only limited economical resources available?

Some methods are expensive to apply each time they are used, for example as market tests; other methods are expensive to implement, such as Box-Jenkins, but may be inexpensive to use once the programme package and the necessary training has been purchased.

iii) Input conditions

DIMENSION: PAST DATA

QUESTION: Are only limited past data available?

This is of course a rather critical dimension for most of the statistical methods, especially for stochastic models, but also for more sophisticated causal ones. In some cases, we may possess for example long data series of totals but very limited sets of segmented figures. This and similar situations call for combined methodologies.

DIMENSION: VARIABILITY

QUESTION: Are the input data series characterised by strong random or other fluctuations?

In general, statistical methods can absorb or isolate variations quite well.

DIMENSION: INTERNAL CONSISTENCY

QUESTION: Are policy changes expected from the telecommunications administration?

Policy changes can be radical, for example as decisions to invest in new services, introduce a new technique, or give priority to specific classes on the waiting list; or quantitative, such as changing the tariffs, etc. All such changes may affect the development of the subscriber demand considerably.

DIMENSION: EXTERNAL CONSISTENCY

QUESTION: Are external changes expected?

External changes concern the environment in which the Administration operates, such as strong social evolution in a country or in certain areas, migration, community policy decisions external to the telecommunication organisation’s domain, etc.

DIMENSION: EXTERNAL STABILITY

QUESTION: Are shifts in the relationships between significant variables expected?

These situations are very difficult whichever model is used, probably more so in the case of the causal methods than for the time series models, where early signs may be detected to some extent.

iv) Output requirements

DIMENSION: DETAIL

QUESTION: Are segmented forecasts required?

No particular problems are presented except possibly by some causal methods where accuracy may become a problem, and in particular by econometric models, where the considered variables may be useful only for an aggregate forecast.

DIMENSION: ACCURACY

QUESTION: Is a high level of accuracy required?

The possible accuracy of most of the methods is subject to the characteristics of the input.

DIMENSION: SHIFT DETECTION

QUESTION: Must shifts be detected early?

The ability to detect turning points is generally higher for some more sophisticated methods, such as econometric models, but some non-statistical methods may also offer this quality.

DIMENSION: SHIFT REFLECTION

QUESTION: Must shifts be reflected immediately?

The ability of reflecting turning points is excellent for causal methods in general, but also for the stochastic class of models, such as Box-Jenkins, once the shifts are identified.

DIMENSION: FORM

QUESTION: Is a probabilistic or interval forecast required?

Most of the statistical methods can provide a confidence interval.

d) Specific requirements

The different dimensions mentioned above are organised in four classes:

- Time Requirements
- Resources
- Input Conditions
- Output Requirements.

If we examine these dimensions, we see that the class Resources concerns mainly the size, structure, skill, level, budget and available computer tools of the forecasting organisation while the other classes of dimension are associated with the specific forecasting situation.

Also, these three other classes are of two types, insofar as the classes Time Requirements and Output Requirements are closely associated with the forecasting problem, i.e. the real purpose of the forecast, if in that is included also the geographical level, the time range, the subscriber segmentation, the zoning, etc.; while the class Input conditions concerns the available data and the degree of environmental stability, even if that stability must be examined in the light of some other possible variables, i.e. we must have some possible methods in mind.

This may seem somewhat contradictory to what is said about the selection principles, in 6.2.2 a). However, this is not the case; the selection cannot be carried out in a single step. We can not answer the questions on consistency and stability, associated with the class Input Conditions, before we have some idea about possibly significant variables, which indicates particular methods or classes of methods. On the other hand the answers to these questions are necessary as a basis for the final method selection. Thus, the selection must be carried out in two stages as shown in Figure 6.13.

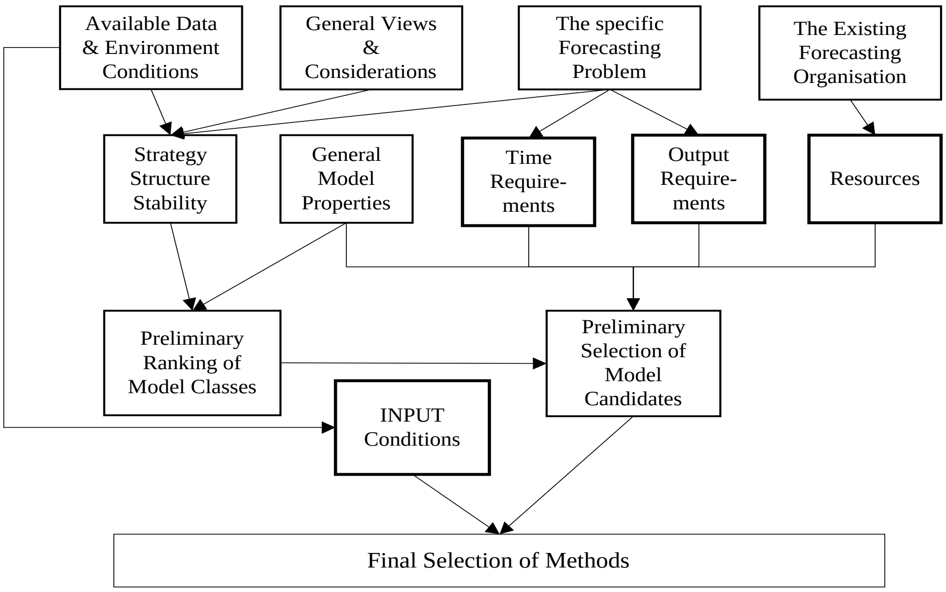


Figure 6.13: Process for selection of method

For the moment, however we concentrate on the problem of defining the set of dimensions, or in other words, of answering the question associated with all the different dimensions. Figure 6.14 shows what kind of information we must use.

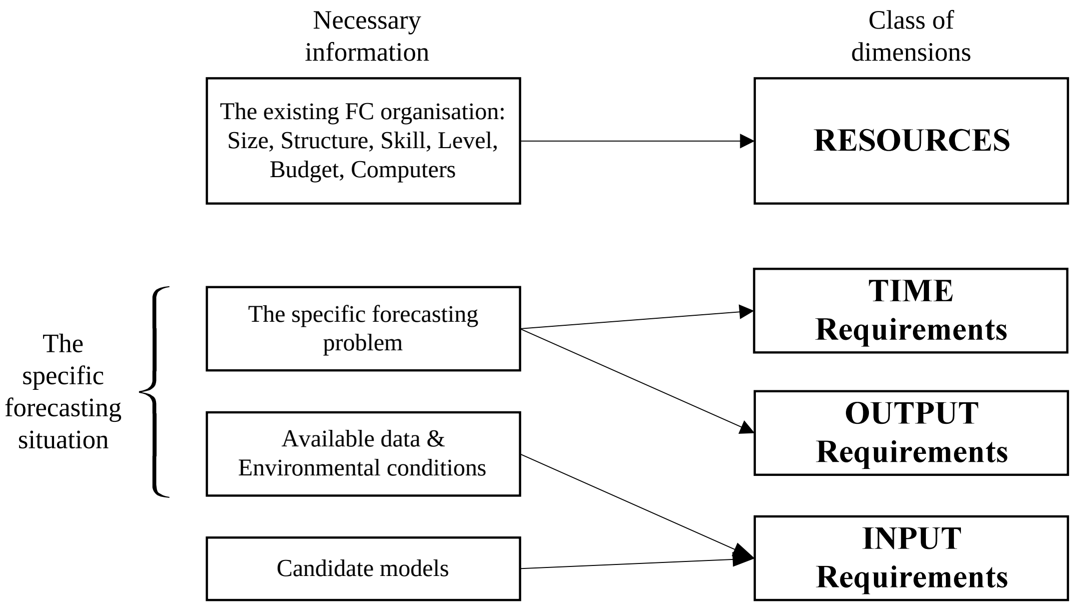


Figure 6.14 : Process of defining the vector of dimensions

To illustrate the definition of a set of "dimensions", let us again use the example from subsection 6.2.2 b), adding a few facts concerning the forecasting organisation and also about available data and environmental conditions.

Say that the forecast group consists of four people: one very experienced traffic engineer, specialised in forecasting during the last three years; one good mathematician and programmer; two telecommunication engineers; one of them experienced in network planning, the other one inexperienced, both of them interested in computer programming and numerical analysis.

The group has only personal computers. No forecasting standard package is implemented.

The available subscriber data consist of subscriber distribution per exchange area for the last 10 years, and per cabinet area for the last 2 years, but not segmented in subscriber classes. No strange variations are observed. A trial to segment the subscribers was made last year.

General social and economic data are available., but other information in short supply.

The economic and social structure of the metropolitan area has not been much changed over the years, but changes will occur in the future; a general economic rise due to industrialisation, some new outer zones will be populated, other zones will be restructured. The administration will give priority to waiting business subscribers from now on.

Our set of dimensions could thus be the following:

HORIZON:	L (The purpose is long term planning)
URGENCY:	Y (We have only a few weeks to go)
FREQUENCY:	N (In principle, a one-time job)
SOPHISTICATION:	N (The group is skilled enough)
COMPUTERISATION:	Y (No software)
ECONOMY:	N (No limitations are expressed)
PAST DATA:	Y (Inhomogenous, short series)
VARIABILITY:	N (No unusual fluctuations observed)
INTERNAL CONSISTENCY:	Y (New policy from administration)
EXTERNAL CONSISTENCY:	Y (Social structural changes)
EXTERNAL STABILITY:	N (Relationships between variables may remain stable)
DETAIL:	Y (Subscriber segmentation needed)
ACCURACY:	N (The purpose is strategic planning)
SHIFT DETECTION:	Y (Trend shifts must be identified)

SHIFT REFLECTION: Y (Trend shifts are crucial for long term development)
FORM: Y (An interval is needed for strategy scenarios)

It must be understood that the lack of space here of course makes it impossible to give a completely realistic example. A real case would comprise much more background material and require extensive analysis and discussion.

e) Properties of quantitative and qualitative methods

The quantity of method properties or characteristics is actually the content of the method selection matrix; in other words, the characteristics of the different possible methods, related to the various dimensions.

Having first of all created the set of crucial dimensions which is valid for the particular forecasting problem in question, and perhaps having established a preliminary ranking between model classes, we may then start selecting suitable methods by investigating to what extent they match the set of dimensions.

It must at all times be remembered however that the proposed method selection scheme is intended to be simply a useful tool in the very important and difficult task of designing a forecasting scheme. All situations are in fact somewhat unique: other dimensions than the ones proposed could be of importance; more relevant questions could be connected to the dimensions; the properties of different methods may not be of universal validity; other methods, versions of methods or improvements could be found. Furthermore, the description of method properties in this chapter is rather brief, so that the forecaster should in addition consult more general information, and he should also include his own experience from the use of the models in other cases.

Finally, having obtained the set of properties in the particular case, any particular method must not be rejected just because there is a B = a Bad Fit notation in the selected set. It is possible that the difficulty can be overcome or is not as important as was first thought, or the method could be of great value as complementary to other methods, etc.

In the following text, each of the dimensions is considered separately in terms of the associated forecasting methods:

HORIZON

We have mentioned earlier that two types of method play a more dominating role for subscriber forecasting than for traffic forecasting: the causal methods, due to obvious and easily understandable relationships between socio-economic factors and subscriber demand development; and the judgmental methods, mainly due to the need for subjective judgement when other methods are insufficient or too difficult to apply.

As far as the time horizon is concerned, the causal methods, also including association without expressed causality, work well for all time spans, the only crucial factor from that point of view being the supply of forecasts of the explaining variables. On a high geographical level such as the national level, this is generally no problem, while on the lowest levels such data may be unobtainable at least for the far future.

In the case of the judgmental methods a distinction is made between the more quantitative methods such as Naive Extrapolation, Naive Composite and Jury of Opinion, and the more qualitative ones such as Scenario methods, Delphi technique and Comparison methods; the more quantitative group being fit for short, possibly up to medium-term forecasts, due to the high degree of uncertainty of these methods when it comes to long term judgement; the other group, comprising the more qualitative methods, being designed especially for the long-term, may be used also for medium-term forecasts. This means, that for medium-term forecasts, no obvious judgmental method can be preferred, possibly different methods must be combined.

The group of Counting methods, for example market testing or market surveys may be a help, since these methods are actually designed for medium-term forecasts. Market surveys could also possibly cover long-term applications.

Finally, the time series methods are all excellent both for short-term and medium-term forecasts. Some of them, such as moving averages, time series extrapolation, the ARIMA methods, and growth curves, cover also long term applications very well, provided that good series of historical data are available.

URGENCY

This is a very interesting dimension insofar as it concerns both the status of the forecasting organisation and the individual methods. We must also remember that subscriber forecasting in general takes more time to execute than for example traffic forecasting.

However, we may here distinguish between development time and execution time. The development time must cover the activity with which we are now dealing, i.e. the design of the forecasting scheme, but must also cover the development and the implementation of methodologies and computer programmes as well as the necessary training. In the case of methods where external experts or the public are involved, such as some judgmental counting methods, the organisation of the activities and the production of the necessary forms must also be included.

The quantitative judgmental methods are usually excellent from the urgency point of view since they can be quickly arranged, executed and processed.

Quite different are the qualitative judgmental methods, especially scenario methods and Delphi technique, both the development times and the execution times being long. If these methods are used in urgent situations, the forecast quality may suffer considerably. The comparison methods may work moderately well provided that data are easily available from the beginning as data gathering is usually time consuming. The output from the calculations must probably also be further processed, which makes execution time moderately long.

The same applies for the market research methods, where the execution of the forecast could take a very long time, after a development time which at least is not short.

In the group of time series methods, moving averages, exponential smoothing and time series extrapolation, are all easy to develop, implement and learn, and are also very quickly executed. Adaptive filtering and time series decomposition offer the same quick execution, only that the development times are somewhat longer, and that time series decomposition may involve some difficulties in data collection.

As far as the ARIMA models are concerned the development time is definitely long; even if programme packages are purchased and implemented, extensive training is certainly required. Due to the strong element of inter-activity necessary for this approach, the execution time can only be moderately short. Taking into account that the ARIMA approach includes the possibilities and the requirements of the other time series methods we may judge it to offer even faster execution than these methods, for a certain defined level of forecast quality.

The growth curves are possibly a special case from the urgency point of view, since they are basically simple, may work with few or many data and may involve subjective judgement. They may on other hand be used in more sophisticated systems. This means that both development and execution can vary considerably but, used by experienced forecasters, they can lead to very quick results.

The last group, association or causal methods, comprises correlation methods, regression models, and (true) econometric models. The use of econometric models is normally very time consuming since the model building is so complicated; once the complete model is developed, the execution of forecasts is however rather quick. The regression models are in principle similar to the econometric models from the urgency point of view, only that they work with explicit expressions, so the development time may be only moderately long. The correlation methods may present some delays concerning data evaluation, otherwise they are moderately quick.

FREQUENCY

All the time series methods and the causal methods are easily updated, as are two of the judgmental methods, naive extrapolation and the expert methods such as jury of opinion. The other judgmental methods, as well as the market research methods are updated relatively easily, but the counting methods in particular require some planning of activities so that spontaneously raised updating requirements could cause difficulties.

Since subscriber forecasting schemes have a tendency to be more complex than other structures, updating could always be a problem, unless the scheme is very well planned for easy processing of new and additional data.

SOPHISTICATION

Some models require a very high level of understanding, the two most demanding being the ARIMA methods, and the econometric models.

A fundamental level of competence is required for the other causal methods, and for most of the time series methods. Two of the time series methods are relatively easy to use: moving averages, and exponential smoothing.

Market research methods require some skill to give useful results, and this also applies for the comparison methods.

This means that if the competence level is low, the forecasting group will have a tendency to use the most simple methods available, probably simple versions of growth curves, and quantitative judgmental methods. Thus, for the important task of subscriber forecasting causal methods are frequently not used at all, or ready-made expressions are perhaps picked up from examples in the literature.

COMPUTERISATION

A computer is essential for all causal methods, with the possible exception of the most simple correlation methods.

For the ARIMA methods and the adaptive filtering methods a computer is a necessity.

The data analysis for the market research methods requires a computer, and for the comparison methods it is very useful.

It is seen that, as for SOPHISTICATION, the lack of a suitable computer means a risk for the forecasting activity to take a wrong turn: possibly using simple time series methods even though trend shifts can be expected, or using simple judgmental methods too extensively, because more relevant, but at the same time also more sophisticated, models can not be applied due to the computation difficulties.

ECONOMY

Among the judgmental and the market research methods there is one extremely expensive method: market testing. The other extremes are the most primitive ones, naive extrapolation and naive composite, which are generally very cheap to use. The remaining methods can vary in cost, but since both preparations and personnel are involved, they can be only moderately economical.

Both the expert methods such as jury of opinion and the Delphi technique involve a number of experts; the scenario methods may involve considerable discussion, and the market surveys require good controls.

Among the statistical methods the ARIMA methods and the econometric models are very costly when it comes to development; the ARIMA programmes are expensive to obtain and will always require expensive training, often also modifications for their adaptation to the specific circumstances. The econometric models always require a substantial amount of development; however, the operating costs of both are low or moderate.

Especially economical among the time series methods are the moving averages, exponential smoothing, and time series extrapolation, provided that the necessary data are readily available.

Other methods, such as adaptive filtering, time series decomposition, and the less sophisticated of the causal methods, are only moderately expensive to develop, and rather inexpensive to operate.

Again, the growth curves can be used without much expense, but, with more sophistication, the cost will increase.

PAST DATA

This is a weak point for almost all statistical methods; possibly only exponential smoothing will work well with limited past data; the most sophisticated methods, such as the ARIMA approach, are completely dependent on long series of past data.

One exception is again the growth curves with their mixture of forward and backward movement on the time axis, and their use of both past data and subjective judgement for curve parameters. This is probably another reason for the popularity of the growth curves.

The judgmental market research methods can work well with only limited past data; one exception is comparison methods, where an extended history is essential.

VARIABILITY

Here the situation is more or less opposite compared to PAST DATA; the statistical methods can handle fluctuations effectively, possibly with some reservations for time series extrapolation, where the confidence interval of the forecast will widen because of the fluctuations in past data, and ordinary correlation methods, which may give poor results if covariation between variables is low.

All judgmental and counting methods handle fluctuations badly; there may however be the possibility of extending the view so as to decrease the influence of short-term influences, at least for more qualitative judgmental methods.

INTERNAL CONSISTENCY

Serious lack of internal consistency will more or less cancel the usefulness of all the time series methods.

Since such inconsistency is frequently due to policy changes in the Administration and the resulting changed economical conditions, the value of the time series methods for forecasting may sometimes be quite limited.

The causal methods will also suffer from such inconsistencies unless they can be appropriately related with, or reflected among, the explaining variables.

Of course, judgmental and market research methods may still work well in these situations, provided that the forecasters and the involved experts or customers are aware of likely inconsistencies.

EXTERNAL CONSISTENCY

This is a somewhat better situation than the foregoing; in particular the causal methods may work well, provided that the changes can be related to the predictor variables; the econometric models may be particularly good in this case.

As far as the time series methods are concerned, one necessary condition is of course that the changes have already started, so that there are at least some effects in past data. The most simple of these methods, moving averages and time series extrapolation, will still give poor results. Even the most sophisticated ones, such as the ARIMA approach, can only moderately well reflect such changes.

Among the judgmental and market research methods, the more qualitative judgmental ones, such as scenario methods and the Delphi technique, may have the best possibilities of reflecting external inconsistencies.

EXTERNAL STABILITY

This is quite another problem; if shifts in the relations between the variables can be expected, then all causal methods are more or less useless - they are based upon stable relationships.

The time series methods seem to be in a somewhat better situation; this may however be only an illusion, since the subscriber development of course in principle can always be explained by a number of factors; these factors as such are just blacked out and replaced by the time parameter. At least moving averages and ordinary time series extrapolation should be avoided in this situation.

The judgmental and the market research methods behave in the same way as for external inconsistencies.

DETAIL

All time series methods, and the most simple judgmental methods, such as naive extrapolation and naive composite should normally offer no difficulties in this respect. When well planned, other judgmental methods, and the counting methods, could also be used for detailed forecasts.

The causal methods, however, may be a problem, working well only for aggregated forecasts, at least if accuracy is important. One exception could be simple regression methods, where causality is not a necessary condition.

ACCURACY

This dimension is difficult to describe, since the various methods may be very accurate under certain favourable conditions but work poorly otherwise.

Naive extrapolation provides a rather limited level of accuracy, unfortunately difficult to evaluate.

Naive composite could be very accurate, provided that the local estimates are of that quality, and that biases and expected changes are considered, otherwise the accuracy may be low.

Jury of opinion is again difficult to evaluate; possibly dynamic conditions could favour this technique.

Scenario methods are of course also suitable for dynamic situations and especially for extended horizons. The absolute accuracy is of course never high with this method.

The Delphi technique is much of the same quality as the scenario methods. Comparison methods always present a great risk; the predictions are suspect; at the same time the forecaster may have a false feeling of control over the situation, since the basic material he uses is possibly good in itself, and so the weak point concerning the doubtful validity of the method is easily forgotten.

Market testing and market surveys can be quite accurate; this depends largely on the preparation, the execution and the data processing.

Moving averages are accurate provided that conditions are stable, as are time series extrapolation and time series decomposition.

The more advanced time series methods, exponential smoothing, adaptive filtering, and the ARIMA methods, normally provide excellent accuracy, at least for short-term forecasts. The ARIMA methods are perhaps preferable in that respect, working well also for medium-term considerations.

The causal methods normally provide good accuracy, much is due to the degree of explained variation and stability of relationships between variables.

SHIFT DETECTION

This is a dimension where the causal methods are useful. In case of non-causality, such as when ordinary correlation methods are used, a lagging relationship between the variables improves the shift detection ability.

The time series methods are weak in this respect; the ARIMA methods may sometimes be able to detect shifts; time series decomposition may do the same to some extent.

Among the judgmental and market research methods, naive extrapolation and naive composite have poor possibilities of shift detection; the other methods can be anything from poor to good.

SHIFT REFLECTION

Here the situation is somewhat different; the judgmental and counting methods have good possibilities to reflect shifts, once they are detected.

The causal methods work excellently, and so probably will the ARIMA methods; other time series methods are fair, if we exclude moving averages and the simple time series extrapolation, where possibilities are poor.

FORM

Interval or probabilistic forecasts require some kind of statistical processing; so that almost all statistical methods and the market research methods will provide such forecasts; exceptions may be exponential smoothing and adaptive filtering, where generally only point forecasts will be obtained.

Naive extrapolation, naive composite, and also comparison methods can provide approximate ranges only, on a purely subjective basis.

Jury of opinion, scenario methods and the Delphi technique offer somewhat better possibilities since the final forecasts are the results of statistical processing of subjective forecasts in themselves.

6.2.3 Socio-economic factors

To discuss socio-economic factors in the field of forecasting as being something particular or special, is irrelevant. Society as a whole is of course social and telecommunication is indeed a very typical social manifestation. Economy is a driving force of central importance to the whole social life, but is also a symbol for, and the result of, social position and activity. Thus, the relative demand for telecommunications is almost automatically a product of socio-economic factors, even if there is interdependence rather than explicit dependence between at least satisfied demand and these factors.

Since the telephone service is well established and has been extensively studied, we will take telephone development as an example in this subsection. Most of the principles and the reasoning will however also apply for other telecommunication services.

a) The systematic strategy for forecasting

The term socio-economic factors is indeed a very promising one - it suggests that telecommunication is a typical logical consequence of the complex interaction in a systematic world where social conditions and economy are the strong driving forces. Thus if we could only discover the mathematical laws regulating the relationship between significant variables we would have solved our forecasting problem, provided that sufficiently reliable, quantitative forecasts of the relevant explaining factors were available.

The keyword is causality, and the forecast steps would be:

- Find explaining variables which are considered truly causal for the variable of interest.
- Design a mathematical model expressing the causality.
- Apply the model to historical data to test and possibly revise the model, and to find the model constants.
- Find forecast values of the explaining variables.
- Calculate the future values of the variable of interest.

The question may be asked: Would it be appropriate to collect historical data and forecasts of many different variables of diversified types, to perform regression analysis for the variable of interest against the individual or combined variables, and, finally, to select a number of variables for use in the forecast model because they have shown the best correlation? The answer is definitely not - the correlation coefficient is a normalised measure of the strength of the linear relationship between variables. It is a purely mathematical expression and does not indicate any causal relationship between the variables involved. When two series are both rising with time it is possible to get a good correlation coefficient even though they are unrelated to each other. On the other hand, even if the two series have a low correlation coefficient, they can still be related to each other, especially for shorter periods.

Thus it is difficult to find true, causal explaining variables and the real meaning of causality has to be clarified. Is it correct that, for example, a future household income X, except for random variations, necessarily would lead to a demand Y in the real world, always and everywhere, if other variables were fixed? Of course it is not, but as a matter of fact, in the field of forecasting, good statistical performance makes us sometimes accept complete absurdities, the real reason for that being that we strongly WISH to find useful relationships and regularities from available data.

However, on the other hand, using the same example, we can probably agree with the opinion that groups of households with higher income also have a higher demand for modern telecommunication services RELATIVE TO THE GENERAL TREND. This is a key to our problem: we forecast for a certain level, for a certain category, and for a certain period of time. We must then remember that our explaining variables do not explain the main development of the demand globally in any of these respects; they can at best explain a part of the VARIATION between the development for the specific case and this main trend.

This again emphasises that our particular forecasting case is a point in the forecasting system with the three dimensions of time range, geographical level, and subscriber segmentation, and that the bottom up/top down principle preferably should be applied.

In many forecasting situations, however, we unfortunately have to abandon the bottom up/top down approach, or we may even have to apply correlation methods without causality. The main point is that we know what we are doing, trying to compensate by combining different possible methods, or at least trying to estimate the quantitative value of the forecast in the form of the degree of uncertainty, possibly applying some kind of sensitivity analysis.

b) Influence of the telecommunication services on demand

The demand for a particular telecommunication service among a particular subscriber category is affected not only by social conditions, economy and a global trend of development, but also by what we may call marketing activities, i.e. the telecommunication organisation’s efforts to advertise the services, tariff policy, etc.

One example of how we can structure services and subscriber categories from this point of view is shown in Table 3.1. The availability and category of the services as well as the market segmentation vary of course from country to country.

Market segments		General[a)] private subscriber	General[a)] business subscriber	Special[b)] business subscriber	Special[c)] institutional customer
Example of service offerings		Major utilisation			
Public services	Telephone	X	X	X	X
	Telex		X	X	X
	Telegraph	X	X		X
	Data services			X	X
	Facsimile		X	X	X
Special services	Wide area services		X	X	X
	Private line		X	X	X
	Private switched services			X	
	Digital data			X	X
	TV transit				X
	Radio transit				X
	Special customer applications			X	X

a) Largely passive marketing effort.
b) Active marketing effort.
c) Related institutions (broadcasting corporations, government, etc.) negotiative marketing.

Table 3.1 Sample of services

A second way of classifying various telecommunications services is to look at their age - old or new services - which to some extent determines their level of development in terms of technological maturity and market saturation. Each telecommunication service, like most other commercial products, has a life cycle spanning from introduction through growth and maturity phases to obsolescence, as depicted in Figure 6.15.

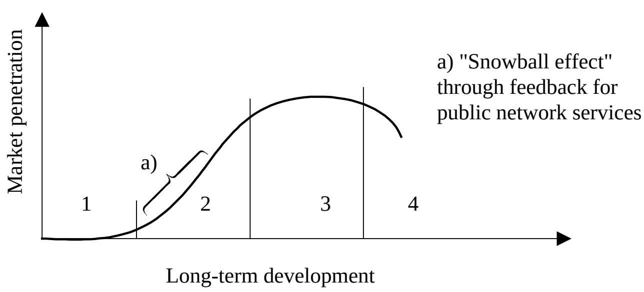


Figure 6.15: Service life-cycle curve

Depending on the initial time of introduction and the general economic climate, the various services listed in the table of market segments have reached a different level of maturity in each country. In some countries data services are now in phase 1, telephone main station growth and telex may have reached phases 2 or 3, and telegraphy may be in a phase 4. The decline in phase 4 depends on the price and effectiveness of available substitutes. In the area of telecommunications, this is normally based on technological advances. The life cycle curve may be at different phases for individual market segments of one product at the same time. In addition to this long-term basic trend, the rate of growth can be influenced during all phases of development by the marketing factors as well as by other factors, which we could call environmental. For public telecommunication services such as telephony, the following special factors of development must be taken into account:

- During the initial phase of growth the objective usefulness of the telephone depends on the quantity of other telephones in the network that can be reached. In addition, subjective factors influence potential subscribers, such as a private telephone as a prestige item (or telex for business customers), and psychological pressure on households without a telephone. The growth of the number of stations which can be called and subjective influences create a feedback effect which sustains an increasing rate of growth. Of course this growth may be controlled by supply constraints.

- At a more advanced phase of the life-cycle curve the factor of market saturation becomes important. For example, in the case of telephony, when practically all households are connected, the growth potential for private main stations in this segment is exhausted although in the case of main stations this would not lead to a decline, but to a levelling off of the curve. Other services offered, such as extension phones, or long distance calling become new marketing targets.

c) Environmental factors

Some environmental factors are clearly quantitative, such as many demographic factors - population, number of households etc. - and are thus relatively easy to use in statistical models, other factors of great importance are difficult or impossible to quantify, such as general technological development, political environment, etc. Some of the latter may be hidden in the parameter “general trend”, others are most often simply neglected.

The ambitious forecaster may, however, find ways to quantify these environmental factors, which is especially important if they are expected to change considerably in the future.

i) Technological environment

The technological environment in general is characterised by fast-growing research and development and an unprecedented swiftness of technological change. In the long run technological development is an important component when estimating the demand. This is due to the fact that this development allows for a better grade of service, new service features, and new service categories. A technically improved service means that a larger part of the population will demand the service in the future. New service categories apply advanced technology to areas of demand which could not be served earlier.

ii) Political environment

In many countries the political and legal environment can be a predominant factor in determining the supply of telecommunications services, based on budgetary allocations or on government regulation. If supply is restricted through such measures, the demand and the most effective exploitation of existing resources may still be influenced by marketing factors. Even if supply cannot be brought to a level where it meets basic demand, it is still important to identify segments and the extent of market demand as a basis for decision making on supply allocations and priorities.

iii) Culture

Culture can be defined as “that complex whole which includes knowledge, belief, arts, morals, law, customs, and any other capabilities and habits acquired by man as a member of society”. This is a classic definition of some of the major aspects of culture and emphasises that it is something which is learned.

Culture comprises the way in which we do things, use things and judge things, and it varies from society to society. Culture used to change slowly, for it expressed long-term responses to physical environment and other experiences. This change is now accelerating.

Based on cultural differences between countries, we must recognise that basic factors affecting telecommunication demands may be of varying importance from country to country. Evaluations of market factors and of demand must be adapted to the specific cultural background of an area. On the other hand, increased communication and travelling between countries nowadays tends to reduce cultural differences around the world.

iv) Demographic factors

Population (absolute number) and age structure of the population are factors which directly affect telecommunication demand. The size and the growth of these factors are strictly connected with birth and death figures and, to a smaller extent, with figures concerning migration.

The number of households (where households are understood as economically self-sufficient small communities): this factor represents a share of potential subscribers of telecommunication services. A study of the number of people living alone and of persons per household can prove useful for a better forecast of the number of households. In order to have greater sensibility in the analysis of the influence of this factor on telecommunication demand, the study of the distribution of the number of marriages of the various age groups may be meaningful for some countries.

Housing: a factor involving the number of houses available today and expected growth, in relation to the specific housing requirements of the various social categories, such as elderly people, single persons and families of various sizes.

Degree of urbanisation: understood to be the relative share of the population living in centres exceeding a specified number of inhabitants. An in-depth survey in this field is very useful to determine different habits and behaviours; varying needs with respect to telecommunications are to be expected at different degrees of urbanisation.

City structure: involving surveys on cities with reference to business and residential quarters, density of dwellings, etc.

Mobility: a study of the population comprising aspects such as the annual number of removals, daily commuter traffic (distance covered and number of people involved), seasonal mobility (e g due to vacation).

Second homes: the number of second homes is an increasingly important factor of additional demand for telecommunication services.

Demographic factors have a significant influence on market demand. For example, the size of households, which varies from country to country, determines the number of households within a given population, and thus the market quantity of households. The size of communities influences demand, since there is less demand for telecommunications within tightly knit rural communities. Furthermore, the size of communities and the distance between them affects the traffic demand between communities.

v) Number of consumers

The market consists of all the actual and potential consumers of a product or service. The size of the market can be measured by studying the overall size of the population. In many cases it is better to include the number of households as a variable rather than then size of the population; however, if this is done, it may be necessary to include household size as another variable: a constant number of households with an increase in average household size would be likely to lead to an increase in the consumption of many commodities. Sometimes the members of a population are weighted - for example, adults counting for more than children - so as to obtain the number of equivalent adults. Then the age and the sex-distribution of population will be of importance in any market which is segmented so that the product is, or can be, slanted towards particular groups.

The business sector can for instance be measured by studying the number of employees and the number of separate business entities.

vi) Needs and wants

The real needs of a country or region cannot always be assessed on the basis of the expressed demand, since this may have no significance. In the more developed countries with a more or less important telecommunication infrastructure, unsatisfied needs find their natural expression in the form of subscriber waiting lists and the difficulties encountered in handling traffic. Even here the clarity with which these needs emerge varies with traffic policy, the distribution of national income and the probability of obtaining satisfaction.

Sometimes, however, fairly large sectors of the population, which have remained at a rather low stage of development and whose standard of living is close to the minimum subsistence level, are quite resigned to living without means of telecommunication. No need is expressed, but this does not mean that a rapid means of communication is unnecessary.

vii) Buying habits

The following components are included:

- consumer personality;
- importance of purchase (availability of substitutes);
- social class;
- professional category;
- culture.

The influence of buying habits on telecommunication services is mainly on the use of equipment and then on the buying of equipment.

Shifting preference patterns may be one of the more substantial influences on a demand for a product, particularly over a period of some years ahead. Furthermore, it is unrealistic to assume that one consumer's preferences are independent of the preferences of other consumers. Changes in consumption therefore occur, independently of relative prices and income levels, because consumers learn from other consumers.

A decision to demand a given product is frequently a result either of previous personal experience of that product or of contact with other people who have used it

viii) Economic factors

The distribution of income, as well as the average level of income, may be among the determinants of the demand for a given service. A country with a few rich people and many poor is likely to have quite a different consumption pattern from that of a country with the same average level of income which is evenly distributed.

Similarly, looking at income distribution over time, a significant shift of the sort which could be caused by a radical reconstruction of tax structure might result in a substantial change in consumption patterns: middle class purchasing habits could replace those of the rich and the poor. Within any given country, however, changes in income distribution are normally fairly gradual and will not affect consumption of most customer products greatly except over the rather long term.

There are other possible shifts in income which are not necessarily discovered by examining average income levels and which may be of importance in determining consumer purchases. For example, some forecasters emphasise the idea of the threshold income level - that there is some level at which the household moves over the threshold of willingness and ability to buy some particular product of service. If the threshold income level could be identified and the number of households moving over it in any given year could be predicted, this would then obviously be of help in determining the potential market for the service in question. A market segmentation of the residential sector by income classes would then be very valuable.

Looking at the business subscriber sector in general, we know that increasing trade and tourism demand a well established telecommunication system. Increasing foreign trade causes an increasing demand for telecommunication services within the different sectors concerned.

The more developed sectors of a developing country are often limited to the capital and its surroundings, or to a port, an important agricultural area or an important raw materials deposit.

A special phenomenon in the developing countries is investments in so called growth-centres from which effects will be spread to the surrounding areas.

Because of this complex mixture, evaluation based on separate areas can give a higher degree of precision for the total country.

ix) Business and residential demand

With respect to both availability and usage, two principal segments are distinguished in some telecommunication services, particularly in the telephone service:

- the private market, i.e. households, or residential lines;
- the business market, i.e. companies, associations and institutions etc. together with the self-employed.

Often this latter market segment is subdivided into:

- business only;
- mixed business/private.

By “mixed business/private” it is meant that a subscriber’s line is used for both business and private traffic, for example in farm houses, small shop owners, family doctors with the practice in their houses, etc. This category unfortunately often overlaps the business market for the small users. Thus the following subscriber segmentation could be more appropriate:

- residential lines;
- single business lines (even enterprises with more than one line belonging to this category);
- multi-business lines (e g PBX-lines);
- public lines (pay phones).

Note that the pay phones are often used as the ONLY telephones in small villages for example.

There is a close relationship between the nature of the economic process in an industry and the need for communication. If within the business market a division is made according to economic activity in terms of primary, secondary and tertiary stages of industrial processing, there will be an increasing need for communication (in the same sequence). With regard to market analysis and demand determination, it is therefore important that both structural and temporary developments in the various branches of industry should be followed closely.

The size of the staff exercises an influence on the demand for telecommunication facilities. This applies in particular to private automatic branch exchanges, PABX (a great many extensions and mutual traffic) and mobile networks in some industries.

In addition, the composition of the total personnel establishment plays an important role. The proportion of “white-collar” workers is determinative of the telephone density of the relevant PBX; also, various provisions for the supply of information are more predominant among the white-collar workers than among the manual workers.

Specific patterns of interaction within an industry, between supply companies, manufacturers and distributors, their location in a country, and their requirements for speedy information flow also determine a demand pattern for telecommunication services.

d) Marketing factors

It is assumed that the environmental factors under which an Administration has to operate are given and cannot be changed or influenced.

There are, however, other factors affecting the market which are under the control of the Administration. These are termed marketing factors. They will determine how much demand arises, how much service will be given and how much work arises, given the environment. They can be used to overcome environmental factors, to accentuate them or to serve any other objective.

i) Tariff policy

It is a widely known and accepted fact of economic theory that normally the higher the price of a particular product or service the less will it be demanded. This statement is modified by the concept of demand elasticity; a product is said to be price inelastic when the relative change in demand is correspondingly smaller than the relative change in price. At the extreme, a product is highly price-elastic if revenue decreases with increasing prices; though the unit price increases, the number of units demanded falls sufficiently to yield less total revenue. It is unlikely that this would be a desirable change, though in some enterprises it might be, yielding the same profit with less effort and resources through a decrease in expenses. Usually, with price-elastic products, costs cannot be reduced in proportion and price increases lead to smaller profits, or larger losses per unit. This feature applies particularly to telephone services because of the large fixed costs existing in networks.

Elasticity of demand for new installations may be estimated taking into account the subjective price perception of potential customers. Price elasticity expresses the sensitivity of customers to the cost of the service. The elasticity parameter is calculated as the ratio of percentage change in demand (quantity sold per period) caused by a percentage change in price.

Besides the application of elasticity to new installations, an elasticity factor may be determined for the effect of a change in monthly fixed charges on subscribers' decisions to retain the service (telephone, telex, private line, etc.).

The factors can be calculated by analysing the effect of historic changes in tariffs for which related unit quantities can be identified.

In an inflationary environment it should be kept in mind that unchanged tariffs represent a relative decline in real price compared to a rising overall price index from year to year. Under a condition of price elasticity this will result in a stimulation of demand.

The degree of elasticity depends on several factors. Intuitively, demand elasticity will be low for business subscribers, and higher for private long-distance calls. The degree of availability of comparable substitute services (e.g. mail, telegram) also influences price elasticity. Finally, the degree of price elasticity depends on the level of income (business cost consciousness, or private disposable or discretionary income).

Elasticity is often lower with greater market penetration, and it may vary with the size of the price change. Elasticity may also be different for price increases and decreases.

Telecommunication services with installations on customer premises are normally billed on the basis of installation charge, monthly fixed charge and usage charges.

One investigation indicated an unbalanced perception of the different pricing elements by potential subscribers so that the installation charge was weighted more heavily than justified by its actual share of total longer term costs.

We may distinguish between satisfied demand, expressed demand and potential demand. The term potential demand indicates that there is a demand not expressed, possibly due to high tariffs, bad service, lack of advertising etc., i.e. different marketing factors.

The expressed demand should be closely related to the size of the waiting list. A substantial waiting list means normally, that possible investments are limited, i.e. non-optimal. In the long run it may, dependent on the expressed economic criteria, be optimal to remove the waiting list as quickly as possible and thereby reach the point where the investments can be optimal; the tariffs must then always be optimised. This can be properly done only if reliable forecasts of the satisfied, expressed and potential demand for connections and traffic are prepared.

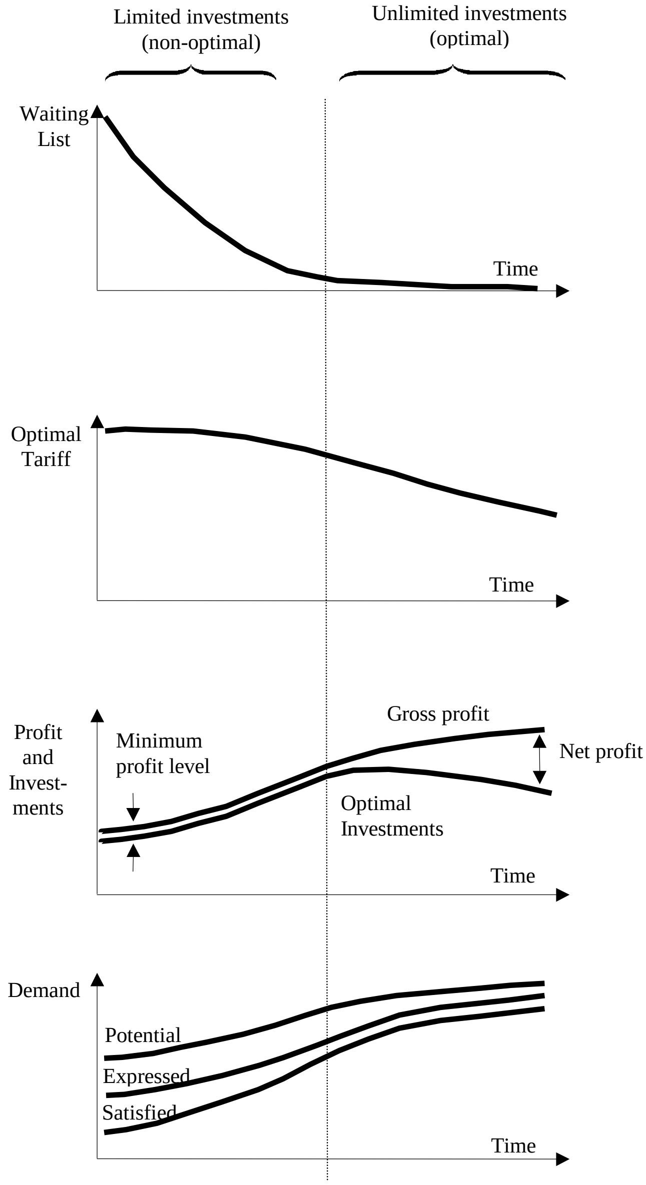


Figure 6.16: Example: Fast removal of waiting list

ii) Other marketing factors

Bad quality of service, limited range of available services, low product quality, poor installation service, frequent faults and long waiting times for repair, are as negative factors for the expressed demand as are shortages in supply. Such conditions make the forecasting process very difficult; the potential demand could be much greater than expressed, which diminishes the value of historical data; the development of demand will be highly dependent on how well the Administration can carry their updated policy into effect in the future.

e) The three phases of growth

The choice of forecasting methods, of short-, medium- and long-term techniques and of the structure of the general forecasting system to be used is conditioned by the development context of the network.

The network does not develop in a linear and uniform manner with time but passes through a number of phases. To facilitate reasonably simple forecasting schemes, this large variability is often approximated through a division into THREE distinct phases:

- a slow, often linear-like growth: the "starting" phase;
- a phase of accelerated growth: the "rapid growth" phase;
- a phase of decelerated growth: the "saturation" phase.

This model with three phases as shown in Figure 6.17 describes in many cases rather well the effect of the market for an item of equipment.

Each phase corresponds for example to different social and economic conditions. It is important to analyse the type of development through which the network is currently passing before determining the forecasting system.

Again, let us take telephony development as an important example.

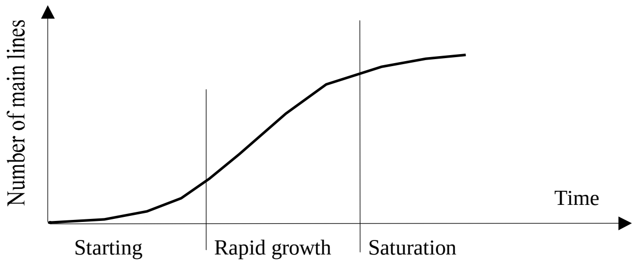


Figure 6.17: Example of the three phases of growth

i) Starting phase

The starting phase corresponds to a period of slow network development which is reflected by a low telephone density as indicated by the number of lines per 100 inhabitants.

The correlation between telephone density and the gross domestic product (GDP) per inhabitant shows that this phase corresponds to countries economically less developed.

The principal component of telephone demand in this phase is the “professional demand”, which arises mainly from the industrial and public administration sectors. It is therefore necessary to install the elementary telephone infrastructure which accompanies industrial development and without which such development would be hindered.

Studies carried out over a long period have shown that there is a correlation between industrial production and professional demand. In some cases, the growth of the network which occurs with industrial growth may be accelerated by increased use of telecommunications within industrial companies. The general form of the curve is often nearly linear.

Residential demand - telephones for private use - remains small during this phase, however, due to low incomes, insufficient consumer spending and the restricted usefulness of the telephone for social purposes, which reduces its attraction for potential users. The telephone is still considered a luxury, and not an everyday commodity, and priority is given instead to other customer goods or simply to the purchase of food in countries with the lowest per capita income.

The development of the number of connections may be restricted by the lack of funds for the heavy investment required for the equipment of a telecommunication network. With only limited financial resources, developing countries give priority to the more fundamental sectors considered to be of greater strategic importance, such as heavy industry, energy and transport. Forecasters must therefore take account of these considerations when formulating a methodology.

In the absence both of planning decisions which give a certain priority to the development of the network and of sufficiently reliable information as a basis for forecasting for the development period of residential demand (i.e. the start of the second phase), it would be advisable to programme short- and medium-term investments. Priority will therefore be given to this type of forecast, concentrating on the professional demand by analysing national and local economic development.

ii) Rapid growth phase

In the context of more advanced economic development, the importance of the telecommunication network increases in both the production sector, which demands a better quality of service and the infrastructure essential for economic development; and the private residential sector, where there is a change of attitude towards the telephone, in that it ceases to be a luxury and becomes an everyday item. This phenomenon is due to the rise in the standard of living, the acquisition of other consumer products and a more widespread acceptance of the telephone by society in general, which increases its usefulness. This latter aspect explains the phenomena of accelerated growth.

However, the potential demand for connections will not in itself increase the number of subscribers without a special investment effort, a long-term plan with specified marketing objectives and the financial means to achieve them.

This rapid growth phase, whose driving force is the residential sector, is often characterised by the demand being much greater than the supply and the supply largely determining the level of expressed demand: an excessive waiting time may discourage possible customers while a reduction in this time induces a new demand. Thus the paradox arrives in which the satisfaction of the demand has the effect of increasing the number of demands in hand. This continues until the network contains sufficient equipment.

This phase represents the stage of network definition. Forecasts will therefore be all the more important in that any serious forecasting errors may lead to costly planning errors.

Strategic plans for equipping the network will be established on the basis of long-term forecasts.

The global forecast to determine the overall cost of investments to be programmed will have more weight than local forecasts and the “descending” type of forecasting will replace the “ascending” type. Local forecasts will only be applied within their allocation of overall resources and will have to be consistent with the forecasts established for the whole network, the latter being the more reliable in this phase.

iii) Saturation phase

Network development then enters a new phase when the residential telephone density reaches a very high level. The network is then considered mature in that there is no longer a high level of potential demand.

The development of the network now follows a stable growth rate with supply adapting to demand.

The composition of the demand for connections undergoes a change. It is no longer made up mainly of new residential requests but becomes more diversified:

- Transfers become a significant part;
- The residential demand becomes a demand for extra equipment such as for weekend homes, second and third lines;
- The professional demand may be renewed by the development of service industries whose vocation is information processing and which require a higher rate of telephone equipment per employee than do manufacturing industries.

This new development context leads to a modification in forecasting methods:

- It becomes necessary to observe local population movements more closely and at finer levels: urbanisation, creation of industrial zones and migrations. The local forecasts become predominant and overall forecasts are the sum of local forecasts;
- New professional equipment philosophies need to be considered, mainly due to the introduction of new services such as teleconference calls or facsimile.

iv) Comparison of growth rates

The duration of the three phases may vary considerably from one area to another. Figure 6.18 shows different ways in which the number of subscribers may develop with time:

- Curve 1 depicts a well established industrialised area; the first two phases have been completed and the current phase is characterised by a gradual increase in the number of subscribers and by high density;
- Curve 2 depicts an area which is in the final stages of network development and will soon enter the saturation phase;
- Curve 3 depicts an area which is embarking upon a spectacular process of network development; the saturation phase need not be taken into account except for long-term planning purposes;
- Curve 4 depicts an area with inadequate domestic resources; telephone development will continue to be very slow.

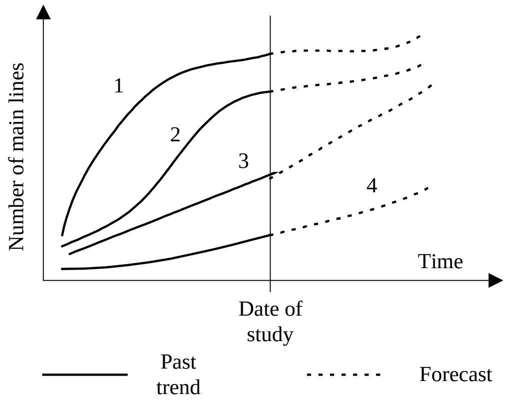


Figure 6.18: Progression of the number of lines

f) Linear econometric models

The systematic strategy is based on the belief that the development of telecommunications and other similar manifestations of human activity are controlled by strict socio-economic laws which in principle can be expressed mathematically, only that there is a complicated interdependence between an enormous amount of significant parameters. The appropriate approach would therefore be to black out less interesting parts of the system and run an extensive multi-equation simulation programme where input data should be the estimated and observed interrelationships between the remaining, more significant variables, i.e. a true econometric model.

i) The danger of the black box

Within the field of subscriber forecasting, true econometric models are rarely applied. Instead, explicit econometric models of a linear type, or linearised by transformation, are widely used.

The general linear model has the form:

$$Y(t) = a(1)\cdot x(1) + a(2)\cdot x(2) + \ldots + a(n)\cdot x(n) + b$$

where *Y(t)* is the dependent variable, *x(1), x(2)....* are the explaining factors, *a(1), a(2)....* are the equation constants, and b is the intercept of the function; and an example of a linearised model could be an expression of elasticity:

$$\ln Y(t) = a(1)\cdot \ln x(1) + a(2)\cdot \ln x(2) + \ldots + a(n)\cdot \ln x(n) + b$$

where *a(1), a(2) ...* are the elasticities of the dependent variable *Y* relative to the explaining factors *x(1), x(2) ...* and *b* is the intercept of the function.

Regression analysis on historical data is used to test the different explaining variables and to calculate *a(1), a(2) ...* and *b*.

The belief in causality could be illustrated as in Figure 6.19.

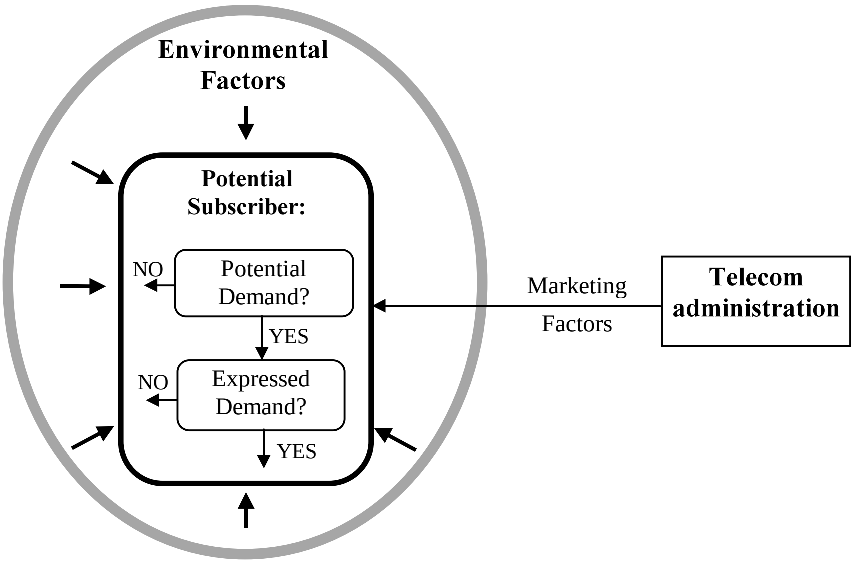


Figure 6.19: Principle of causality

In practice we may start with the following data :

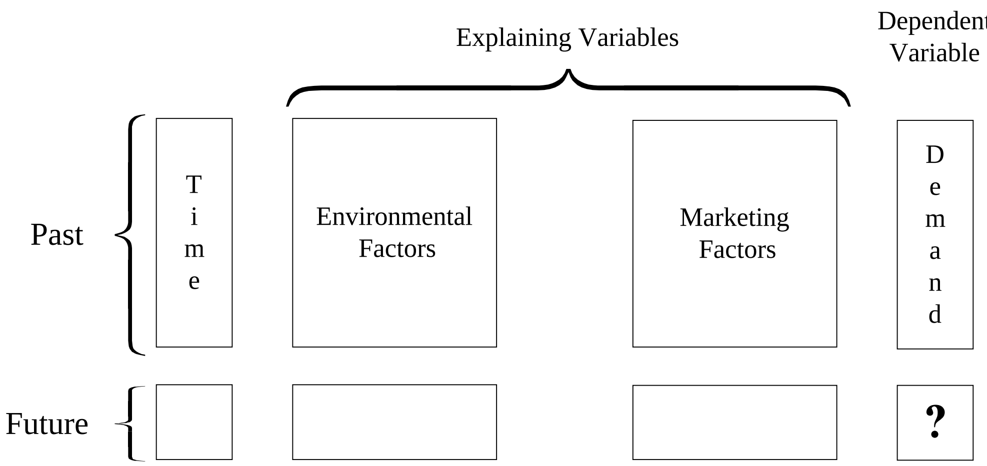


Figure 6.20: Start position for linear econometric analysis

After the statistical analysis we could have defined a function such as:

$$Y(t) = F(X, Y, Z, \mathrm{K}\ d)$$

where *Y(t)* is the dependent variable, *X, Y, Z*..... are the chosen explaining variables, and *d* is a trend development factor which takes into account all other influences.

Possibly in many cases no general trend development factor is used at all; a set of explaining factors is instead chosen so as to minimise the statistical correlation value. In these cases, although the original intention was to establish some form of causality, it is likely that no true causality exists and that, effectively, formal statistics have been used simply to create a “black box” which, when fed with input data, always produces an output, but which gives us no information at all about the real complex connections between its input and output variables. Usually, it is possible to find several similar model hypotheses which are equally good theoretically and statistically, but which would produce different forecasts.

ii) A linear econometric model example (Source: The Swedish Telecommunications Administration)

General models for different types of demand for telecommunication products are used both on central and local levels, as shown in Figure 6.21.

Table 3.2 shows the factors that, at the time(1977), were thought to affect the demand for various products. Changes in the number of main stations and stations are studied by dealing with the new orders and cessations separately. An example of an affecting factor for new orders, main stations, is “number of population (age groups)”. In this case “age group” means the number of eighteen-year-olds in the year in question.

The econometric models in use today have been chosen by means of many different analyses using different combinations of explaining factors, as shown in Tables 3.3 and 3.4. Different types of prices have been used (fixed prices and current prices). Various time shifts have also been added to the statistical analysis.

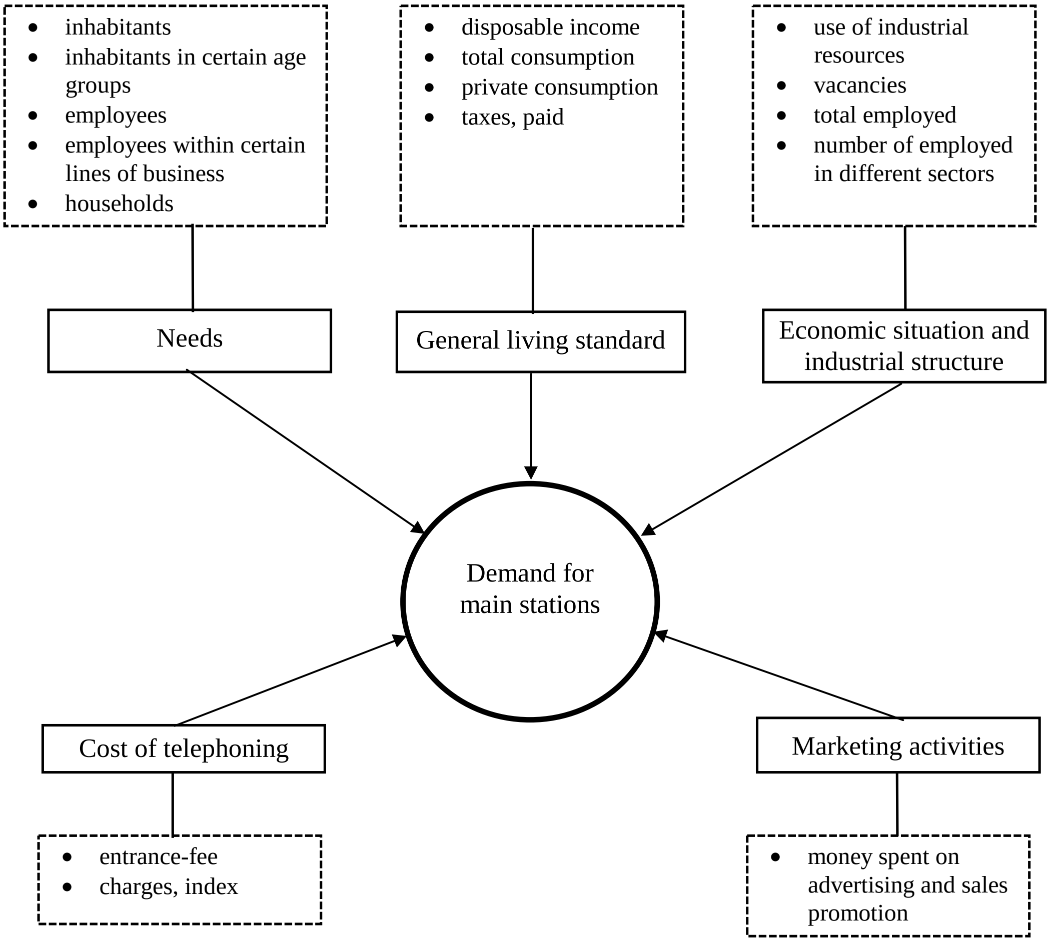


Figure 6.21: Demand model

PRODUCT			Number of population (age groups)	Private consumption	Economic index	Number of employees	Import and export	Money spent on advertising, sales promotion, etc.	Working telex main stations	Gross investment
MAIN STATIONS	**Residential sector**	**New Orders A**	•	•						
		B	•		•					
		Cessation	•							
	Business & Public Sector	**New Orders**			•	•				
	Second Dwellings			•						
STATIONS		**New Orders A**	•	•						
		B			•	•				
TELEX MAIN STATIONS		**New Orders**					•	•		
		Cessation			•				•	
		Removals			•					•

TABLE 3.2 : Explaining factors

RESIDENTIAL SECTOR	**BUSINESS & PUBLIC SECTOR**	**SECOND DWELLINGS**
New orders: • Entrance-fee • Economic index • Inhabitants (in different age groups) • Housing construction (new constructions) • Total employed • Consumption	New orders: • Economic index • Entrance fee • GNP • Consumption • Total employed	New orders: • Consumption • GNP • Entrance fee • Economic index • Number of secondary dwellings • Total employed
Cessation: • Deaths • Inhabitants (in different age groups) • Emigrants		

TABLE 3.3 Demand for telephone main stations: tested variables (Example)

TELEPHONE STATIONS	
New orders:	• Money spent on advertising, sales promotion, etc. • Charges • Inhabitants (in different age groups) • Economic index • One- and two-dwelling houses and terraced houses • Total emplyed • Consumption • Housing construction (new constructions)
Cessation:	• Deaths • Inhabitants (in different age groups) • Emigrants • Working main stations

TABLE 3.4 Demand for telephone stations: tested variables (Example)

g) Analytic models

In some cases, it is possible to adopt an analytic approach for demand forecasts, as for other complex problems.

This means dividing an inhomogeneous total market into more homogenous sub-markets and identifying by empirical means causal connections between the influence variables and the demand behaviour - in practical terms, on the basis of representative surveys and measurements. Demand models so obtained are causal models in the more narrow sense of the term.

i) Socio-economic user data

When data of a socio-economic character about telephone users are available, either directly from the Telecommunication Administration registers or, more likely, by making use of household sample surveys available from the Government's statistical authorities, the use of the telephone service can be investigated with respect to for example:

- income or expenditure per household;
- professional category of the head of the household;
- educational level of the head of the household;
- type of living quarters used by the household;
- price indices;
- etc.

Such an investigation can be further extended to examine the penetration of the telephone service by region of the country, for urban and/or rural areas, etc. according to the data that have been collected in the relevant sample surveys.

The most important socio-economic factors which seem to determine the use of the telephone service are:

- the income (or expenditure) per household;
- the professional category of the head of the household;
- the service access charges.

Sample surveys usually distinguish between the following, or similar categories:

- professional and technical workers;
- administrative and managerial workers;
- sales workers;
- clerical and related workers;
- service workers;
- production and related workers;
- labourers;
- agricultural workers, fishermen;
- non-active persons.

The above mentioned sample surveys cover mainly the so called residential user category. Other surveys on business enterprises etc. are also made, though less regularly. The estimate of the business/official telephone users must be based on particular investigations.

ii) Typical development trends by user category

Telephone services have usually first been provided to the business and official (Governmental) users of a country. These users are also known to make more use of the telephone services than the residential users and consequently, from a revenue point of view, are more attractive to the service provider. These categories are also quite insensitive to price changes.

The often high installation charges do not seem to affect the demand for services from the business and official telephone users. High installation charges will obviously have the effect of limiting the access to the service by lower income users. The most affected category then comprises the residential users.

The figures below show the telephone development trends for Japan since 1962 for the business and residential user categories. The same pattern is typical of any country, though the time frame will be specific to each country. This is dependent on national conditions, economic, political, etc.

This development scenario is typical of both developing and developed (industrialised) countries, though the effects are usually more obvious in the low telephone density countries. The fact that, when the penetration of services has reached a higher value, the marginal costs for providing services are lower allows lower charges to be applied.

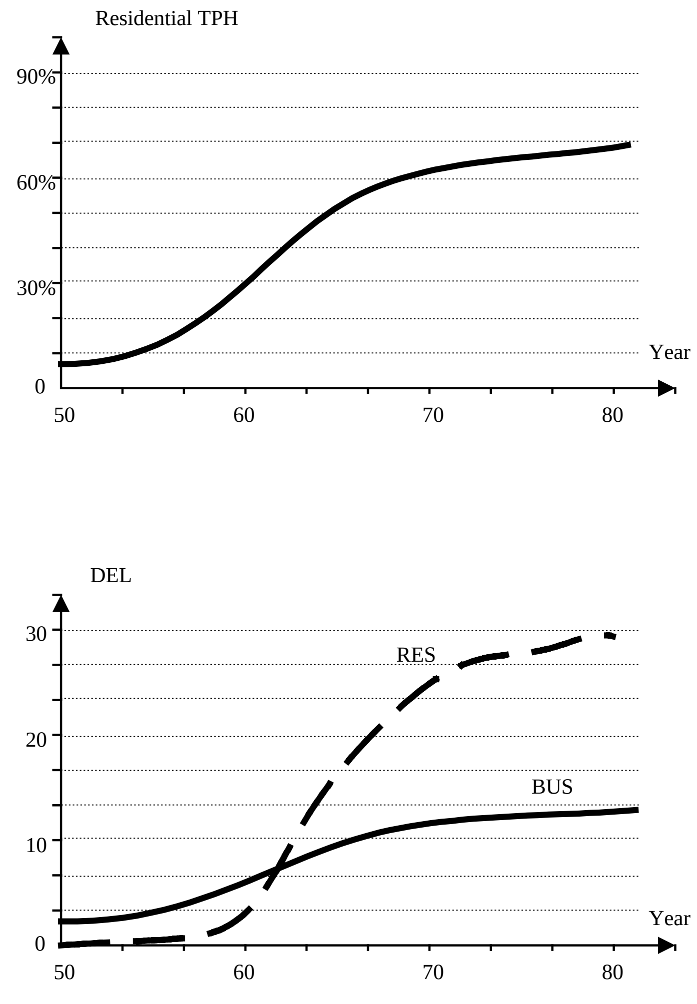


Figure 6.22: JAPAN: Development trends for business and residential users

6.2.4 Subscriber segmentation

It may be obvious but still needs stating that one seldom forecasts demand for telecommunications as such. Management clearly requires demand forecasts for each of the services separately. This is not only because the various telecommunication services may be at different points in their life cycle (from inception to saturation), but also because each may have a growth rate or trend (e.g.compare telephone and telegram demands today), which would make the forecaster's task of predicting a combined demand very difficult. More importantly, the management and operational tasks and problems (e.g. financing, equipping, marketing, staffing, charging, training) associated with the various demand functions will vary enormously, thus justifying the need for the separate demand forecasts.

In a similar way, within any particular telecommunication service, for example telephony, there will be justification from both the forecaster's and the administrator's viewpoints for separate demand forecasts for different aspects of the service or segments of the market. In the field of subscriber forecasting, this type of disaggregation is called subscriber segmentation, and traffic forecasting should also be carried out based on that segmentation.

It should be remembered that there are different aspects to consider before the final choice of a particular subscriber segmentation can be made:

1. output requirements derived from the defined purpose of the forecast;
2. analytic needs due to the heterogeneous market;
3. statistical requirements about aggregate sizes;
4. input constraints, i.e. limited supply of data and other significant resources.

Requirements 1 and 4 work in the same direction, while at least requirement 2 is contradictory to the last two items, for the following reasons:

- To produce a sufficiently accurate and reliable demand forecast using data from a very heterogeneous market, an analytic approach will in principle be the most safe and controllable model, i.e. different subscriber categories will develop their demands quite differently and due to different factors; therefore, each category should be studied individually. On the other hand, if the sizes of the aggregates are too small, the final forecasts may be of poor quality for statistical reasons. Another drawback of using small aggregates can be lack of data, since a disaggregation puts a higher demand on the (hopefully) existing data base.

- Individually taken, 1, 2 and 3 could consequently lead to quite contradictory decisions from this point of view.

- In principle, an optimal degree of segmentation could be found for each specific forecasting situation.

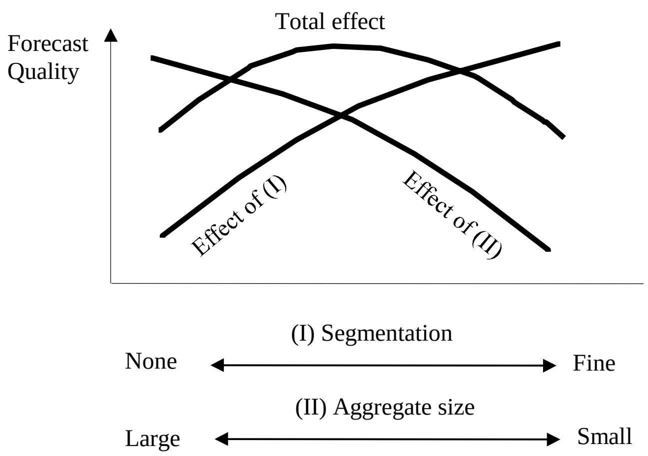


Figure 6.23: Effect of segmentation and aggregate size on forecast quality (example)

Thus the problem will be to match the output requirements against different possible complex forecasts models, the ranges of which are dependent on analytic needs, statistical requirements, and input constraints. In some respects, output requirements and analytic needs are, however, the main and positive items in the decision process, while statistic requirements and input constraints are negative factors, limiting the possibilities of optimal forecasting model design.

a) Systematic planning of subscriber segmentation

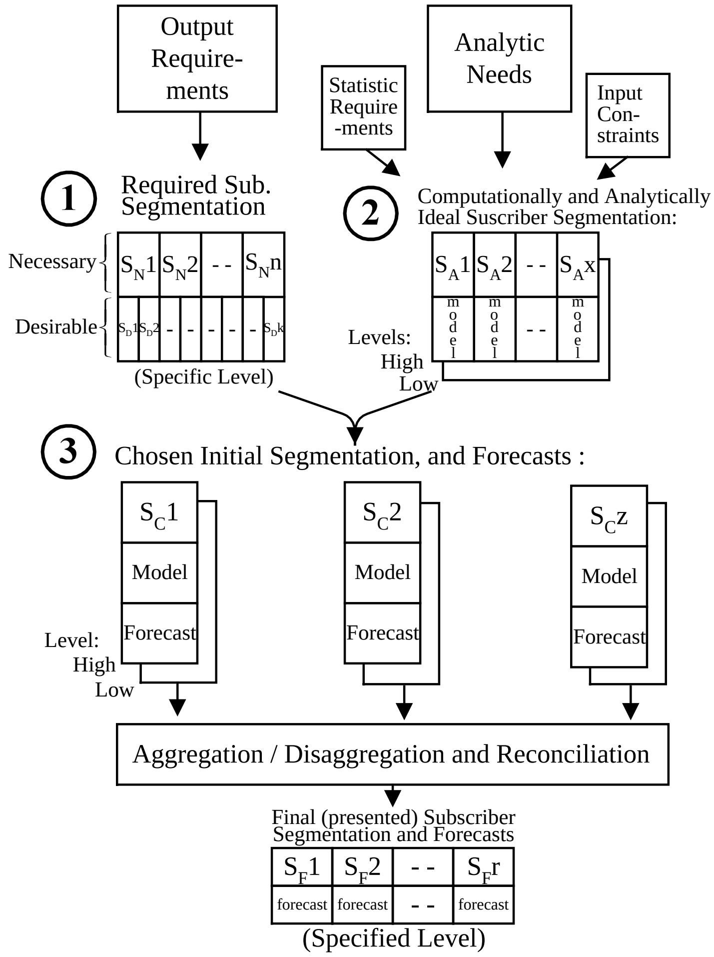


Figure 6.24: Choice of subscriber segmentation in the forecast process

Figure 6.24 illustrates a systematic approach to subscriber segmentation planning:

1. A careful analysis of the purpose of the forecasts reveals the output requirements, possibly divided into two levels, viz. the minimum acceptable segmentation from the user's point of view, and an optional, finer segmentation, that would improve the forecast and/or increase its usefulness.

2. The ideal subscribers segmentation from an analytical point of view is defined. Statistical requirements and input constraints are weighed in. The defined segmentation must refer to specific forecast models and to specific levels, certainly geographically, but possibly also in the time perspective. To avoid completely useless alternatives, the analysis must be carried out with a clear awareness of the segmentation derived from the output requirements.

3. An initial segmentation can now be defined, fulfilling the output requirements, at the same time sufficiently satisfying the analytical needs. Now not only appropriate models must be chosen, but also an algorithm for the aggregation and/or disaggregation of forecasts so obtained, in order to arrive at a final level for presentation.

Situations where the user's need can not be met completely with the help of statistical methods will frequently occur. In such cases, possibly non-statistical methods must be integrated in the forecasting system.

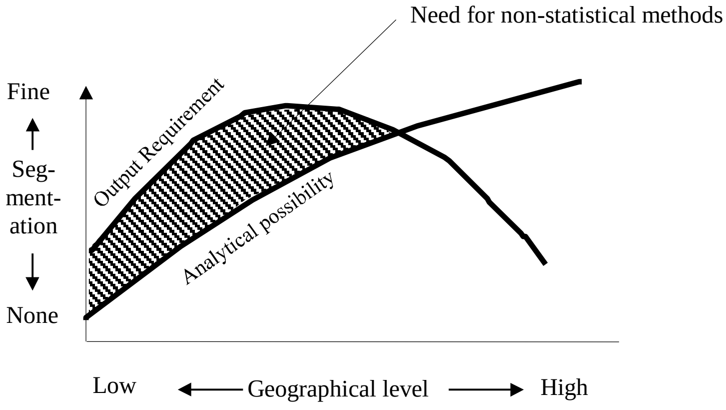


Figure 6.25: Need for non-statistical methods

b) Output requirements

There are two main purposes of subscriber forecasts, comprising one direct and one indirect objective. The first objective is the use of the subscriber forecast for marketing and planning of services and for the part of the planning of the network that is more or less traffic-independent; the other being the use of the subscriber forecasting as a basis for the traffic forecasting activity.

Long term planning of metropolitan networks, for example, requires a forecast traffic matrix, i.e. the traffic flows between a number of different areas, each individual area being geographically as homogenous as possible from the subscriber mix point of view, but the different areas having different subscriber mixes.

It is then very important to produce a subscriber forecast that is as well segmented as possible so as to permit an analytically adequate traffic forecast to be made.

Figure 6.26 illustrates the structure of a telephone forecasting system from the user's point of view, and Figure 6.27 shows a possible segmentation structure.

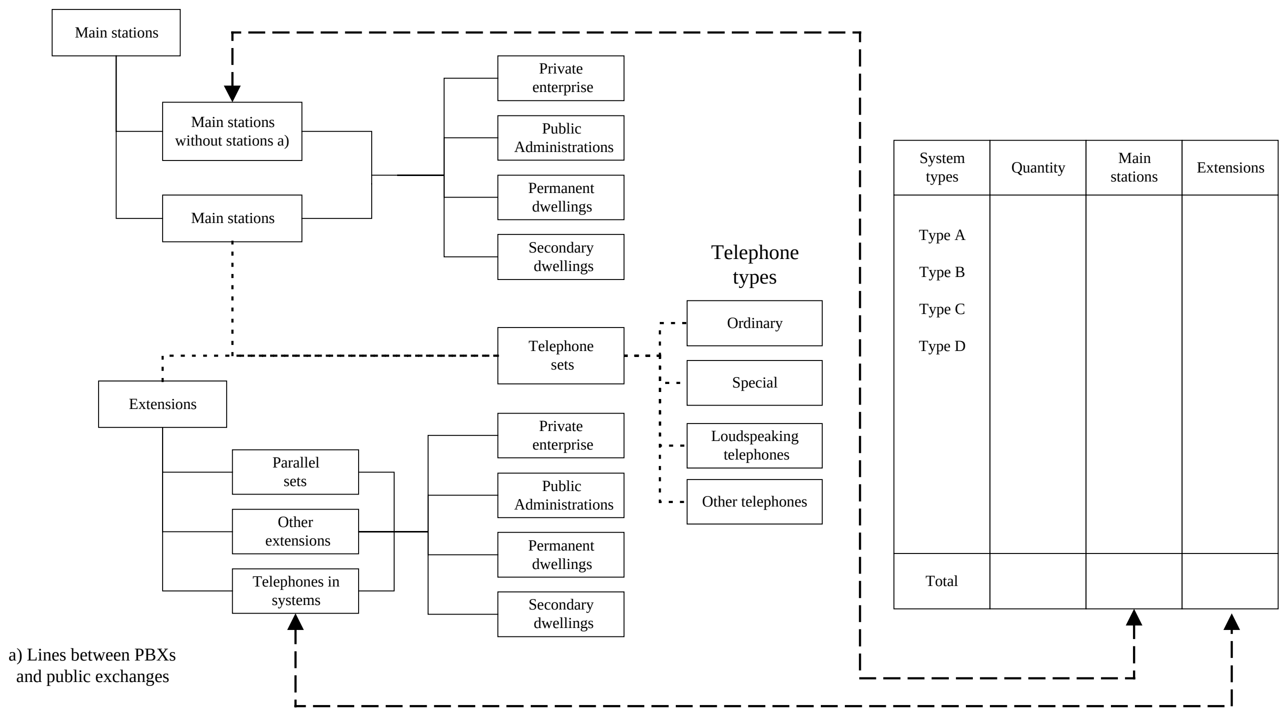


Figure 6.26: Components of a forecasting system (Example)

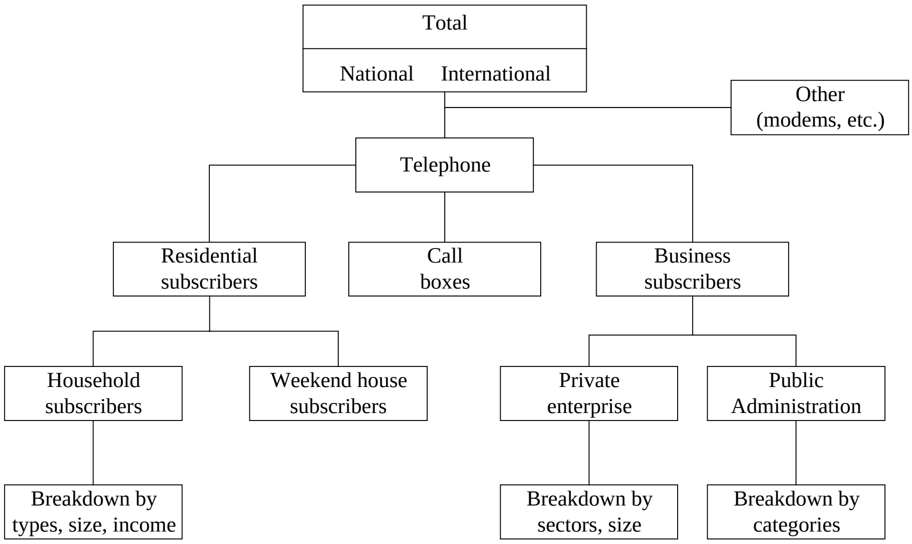


Figure 6.27: Segmentation structure (Example)

c) Analytical needs

As an example, consider only two subscriber categories; residential and business subscribers. How is the development of demand for these two classes explained?

Of course, in reality there are many of different influences, but let us concentrate on the main differences between the groups.

If we first take the residential group, we could try to define a sort of minimum level of demand, which we could call “the basic need”. We know that the definition of such a term will be floating and that both the absolute level of basic need as well as the proportion of the total demand that may be defined will change with time and also be completely different for different countries and areas.

In any case, the main driving forces behind the “basic need” are global social factors such as structure and services of the community, while the main driving forces behind the rest of the demand are more local social factors, i.e. such factors that create a WISH to posses the respective service. The individual household economy is not really a driving force behind this wish but is still a strong explaining factor, being the MEANS TO SATISFY THE WISH, so that the demand really is an expressed wish.

If we now take the business sector, we appreciate that the need for running the business must be the main reason behind the expressed demand, so that driving forces are business activities, trade, etc. Thus, we see that in the business sector, economy is the goal and telecommunication is one of the necessary means.

We could say that essentially there is an ACTIVE relationship between economy and demand, but there is a low price elasticity.

In contrast, in the residential group there is a PASSIVE relationship between economy and the demand, since telecommunications is the goal and economy is the necessary means. The price elasticity is usually also relatively high.

It follows from this reasoning, that the systematic strategy and the corresponding analytic approach requires a segmentation into at least these two categories.

6.2.5 Time scale and geographical level

When we discuss time scale, geographical level, subscriber segmentation, macro or micro forecasting etc., in many cases we then relate these concepts to the PURPOSE OF THE FORECAST, but often they rather have to do with possible or appropriate FORECAST MODELS, and sometimes with a more complex FORECASTING STRUCTURE. Let us now examine these terms more closely.

TIME SCALE is either short -term, for example each one of the first five years or long-term, for example more than five years ahead, often 10, 15 and 20. A bridge between the two is often required in form of a medium-term forecast.

Both short-term and long-term forecasts should regularly be made for all geographical levels; first of all because they are needed for planning purposes, but each one of them also being an important part in a dynamic reconciliation process, so that the long-term forecast is updated when short-term forecasts are revised. On the other hand, the long-term forecast can form an envelope for the shorter term variations.

However, the primary use of both short-term and long-term forecasts on the national level is, via network planning, a broad allocation of resources for land, buildings, vehicles, exchange equipment, local cable plant, junction and trunk circuits, terminals, PBX, personnel needed for construction, installation and maintenance, operators and other staff.

Broad allocation means planning from the financial, recruitment and organisational points of view.

On the local level the primary use of the forecasts is for the total planning of the network and, as a consequence of that, calculation of requirements for land, floor space and equipment. Timing of extensions of existing exchanges and network as well as commissioning of new exchanges is of course of great importance.

Thus time scale is mainly related to the PURPOSE of the forecast, but to some extent also to the required consistency and quality of the forecast.

GEOGRAPHICAL LEVEL could be the national level comprising the long distance network, the metropolitan network level, or the rural network level.

All these levels are service areas related to the PURPOSE of the forecast.

However, one problem is that important data, especially of the socio-economic type, basically are available for what we call administrative areas, whereas the telecommunication planning must be carried out for the service areas, which are often different from the administrative areas.

We must then work on a level of ELEMENTARY AREAS defined so that both the administrative areas and the service areas are aggregates of these elementary areas.

We could say that the logically necessary zoning system is the interface between the two types of areas.

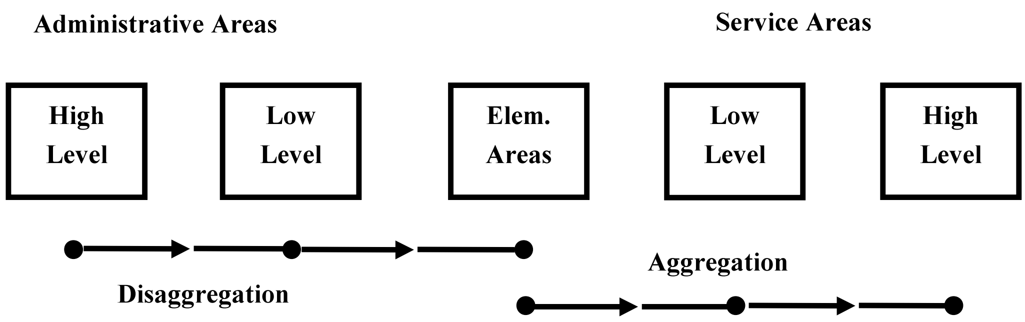


Figure 6.28: Illustration of a Zoning Concept

Both population forecasts and demand forecasts can of course in principle be made anywhere from the high level administrative area to the high level service area. Usually, however, population forecasts are made for high level administrative areas, thereafter being disaggregated down to low levels, the demand forecasts consequently being performed on these low levels with a following aggregation up to higher levels, (Bottom Up) possibly reconciled with direct high level demand forecasts (Top Down).

So we see that geographical level is related not only to the purpose of the forecast, but also to both input constraints and analytical needs.

SUBSCRIBER SEGMENTATION, as treated in section 6.2.4, is related to the analytical need and to the input constraints as well as to the purpose of the forecast.

When the purpose is to provide a basis for the traffic forecast, however, we could say that the corresponding degree of segmentation is again, indirectly, caused by analytical needs.

The concept of MACRO and MICRO FORECASTS is basically related to the forecast model; the macro forecast being based on summarising measures for the concerned area, such as employment, earnings, occupation and consumption of various commodities and services; the micro forecast rather being based upon details of the studied area; for example present and future land usage, city plans etc. Thus micro means detailed and macro means total or overall.

This may be a matter of choice of appropriate forecast model, but the purpose is also a strong factor, the micro forecast usually being the appropriate approach when for example exact distribution of the demand in an urban area is required.

The use of an ASCENDING or a DESCENDING approach is a pure matter of forecast model (due to analytic needs, etc.) and is used in the geographical context; the use of the ascending method simply means that the forecasts are made on a lower geographical level with a following aggregation up to the desired level, the descending method of course being the reverse procedure.

AGGREGATION and DISAGGREGATION are quite general terms related to both geographical level and subscriber segmentation. They are a matter of method rather than of purpose.

TOP DOWN/BOTTOM UP can be even more general, also including time, i.e. dynamic forecasts.

a) Time scale

Even if a changed course of short-term demand development can indicate the need for revision of the long - term forecast, the latter still provides the foundation for all other planning; it is very important that the quality of the long-term forecast is high.

Long-term forecasts are essential not only for determining specific land and building requirements and for rough estimates of required annual quantities of materials and of capital needs, etc., but in our rapidly changing world they are badly needed for strategic network planning, which is crucial for the whole telecommunications future of any country.

Short- and medium-term planning deals with the determination of how telecommunication plant should be built, in the light of the long-term plan.

A short-term forecast is usually made for day-to-day engineering needs and is normally derived for each year over for example a five-year period, while the time scale for medium-term forecasts depends on each plant’s lead time, i.e. the period between the identification of a need and its realisation.

It is necessary to update these forecast frequently, particularly short-term ones, in order to keep abreast of current technological developments, changes in the character of traffic, subscriber distributions, and residential and business developments. Furthermore, when any changes are made in the short- or medium-term forecasts, it is essential that the long-term forecasts be reviewed and updated as necessary.

i) Dynamic forecasting problems

In one sense, all intelligent forecasting starts with scenarios, even if no particular scenario methods have been applied, but the forecaster has used “common sense” to speculate about “reasonable” courses of development. Human beings are, however, usually optimists, at the same time almost always employing conservative thinking.

The optimistic view is reflected by the well known fact that local micro forecasters generally tend to overforecast the demand of their own area, while the conservative thinking makes us believe that everything will continue as before - which is however not at all the case in the real world.

As a matter of fact, this changing world will also cause almost all kinds of statistical forecast models to fail even if there were an unlimited supply of data, at least when used for long-term forecasting.

Time-series models will fail not only because of large statistical confidence intervals, but mainly because the trends WILL shift; causal models will all fail even if all kinds of explaining or correlated variables are forecast, since elasticities and interdependencies will change with time.

So we see that the tendency to use qualitative (non-statistical) methods for long-term forecasting is caused not only by lack of data but to a great extent also by trend shifts that can not be quantitatively predicted.

Beside qualitative methods, we may after all still try to take an analytical approach, using causal correlation models for medium and long-term forecasting. There are some very common mistakes which partly may have to do with time:

- CROSS-SECTIONAL DATA USED DYNAMICALLY

  The coefficients of the forecast equation may be derived from data about the variable of interest and about explaining variables taken from a set of different countries or areas.

  Such coefficients are NOT VALID for forecasting, since in reality there is a strong global development trend over time and this global trend would be more or less neglected with this kind of analysis.

  Occasionally, the situation is worse; that is if data from only one point of time but from many geographical areas are used; if the material were spread over time at least to some extent, it would probably be rather better.

- CORRELATION WITHOUT CAUSALITY

  The leaving out of true causal factors or the inclusion of correlated but non-causal factors could both lead to erroneous forecasts if predictable trend shifts in these factors occur. One way to discover causality or non-causality is to analyse trend shift effects in past data, preferably for the same area, possibly even for other areas.

- REGRESSION IN WRONG DIRECTION

  If we analyse a data series for two variables X and Y, we will find that the regression line $Y = f1(X)$ is different compared to the regression line $X = f2(Y)$. These regression lines are uni-directional. A third possibility is to use orthogonal regression, which will give a line between the other two, $f3(X,Y)$.

  In the case of true causality, regression in one direction, that is to say the appropriate one, is well motivated. Should however the direction by mistake be reversed, the forecast would be wrong.

  If there is correlation but not causality between two variables, then any direction of the regression will not be adequate; in that case orthogonal regression is better.

- LIMITED SAMPLING

  If the explaining variable is sampled within a given interval (which to some extent often is the case); this will generally lead to under-forecasting of the variable of interest, provided that the development is positive over time, or, in case of a causal explaining factor, that the elasticity is positive.

ii) Choice of models

Assuming a situation as shown in Figure 6.29, with a good series of historical data indicating both expressed and satisfied demand, it is seen that the expressed demand at first developed quite steadily, then there was a period of stagnation believed to have been caused by economic difficulties, but now the development seems to have recovered again.

If we define a set of target marketing factor values (not yet achieved) it is then theoretically possible to estimate the corresponding "potential" demand. This means that the expressed demand would have been equal to the estimated potential one if the real marketing factor values (e g tariff level, subscriber services etc.) had been the target ones.

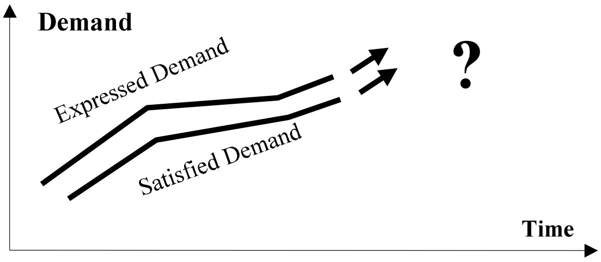


Figure 6.29: Forecasting situation (Example)

One of many possible approaches would then be the following:

- An attempt is made to estimate the potential demand in the past and on a long-term basis.
- The expressed demand is forecast on a short-term basis, using statistical methods, either deterministically (time-series) or analytically (causal model) or both. The stagnation period can cause trouble for the time series analysis but may not be entirely troublesome for the causal approach since a hypothesis of causality may be tested using such trend shift periods.
- It is assumed that the expressed demand and the potential demand will be the same in the far future. Therefore, an extrapolation between the short-term forecast and the long-term forecast of potential demand can be made.
- Possibly an estimate of the future satisfied demand according to different supply assumptions can be made.

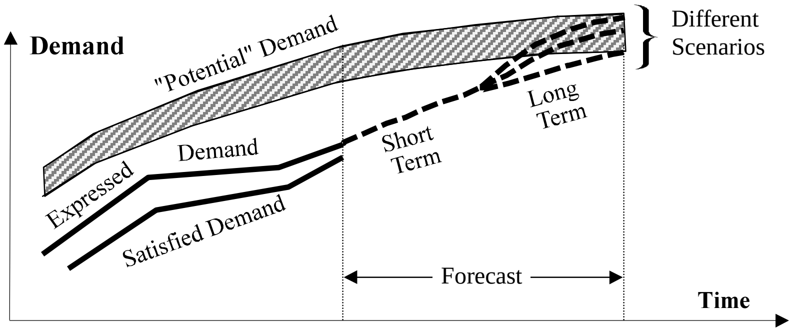


Figure 6.30: Forecast approach (Example)

The following questions now have to be answered:

- Should relative growth, absolute growth, or absolute level be forecast?
- Should the specific area be placed in relation to more global development trends, considering either other areas of the same type or higher levels, or should the outer world be ignored?
- Which models would be appropriate?

If we start with the first question, we could in general say that for short-term forecasts, absolute level may be appropriate in many situations, but in other cases, possibly for causal models, relative growth variations may be better explained by the causal factors. For long-term forecasts, absolute level is possibly less risky than relative growth.

As for the second question, global factors will generally have less influence in short time perspectives but for a more distant time horizon their importance will increase. Even small entities will, in time, often adapt to the development of the outer world.

In the case of the last question, a general scheme for model choice is described above.

The general effects of an increased time span are tendencies to apply qualitative instead of quantitative models, reduction of time-series analysis thereby increasing the need for causal and judgmental methods, and possibly an increased need for top-down approaches.

An interesting class of models is the Growth Curves, which could be used for the entire time span.

To use for example some type of logistic curve for estimation of the future saturation level from past data would however be very risky and uncertain, since very small changes in input data may lead to quite different calculated saturation levels. Instead of this, the saturation level should normally be predicted by other methods and then imposed on the logistic functions.

Figure 6.31 attempts to illustrate for what time span each model is best used, if it is to be used at all. The thickness of each black bar then corresponds to the usefulness of the model for different time spans. Note that the diagram does NOT demonstrate the RELATIVE importance of different methods.

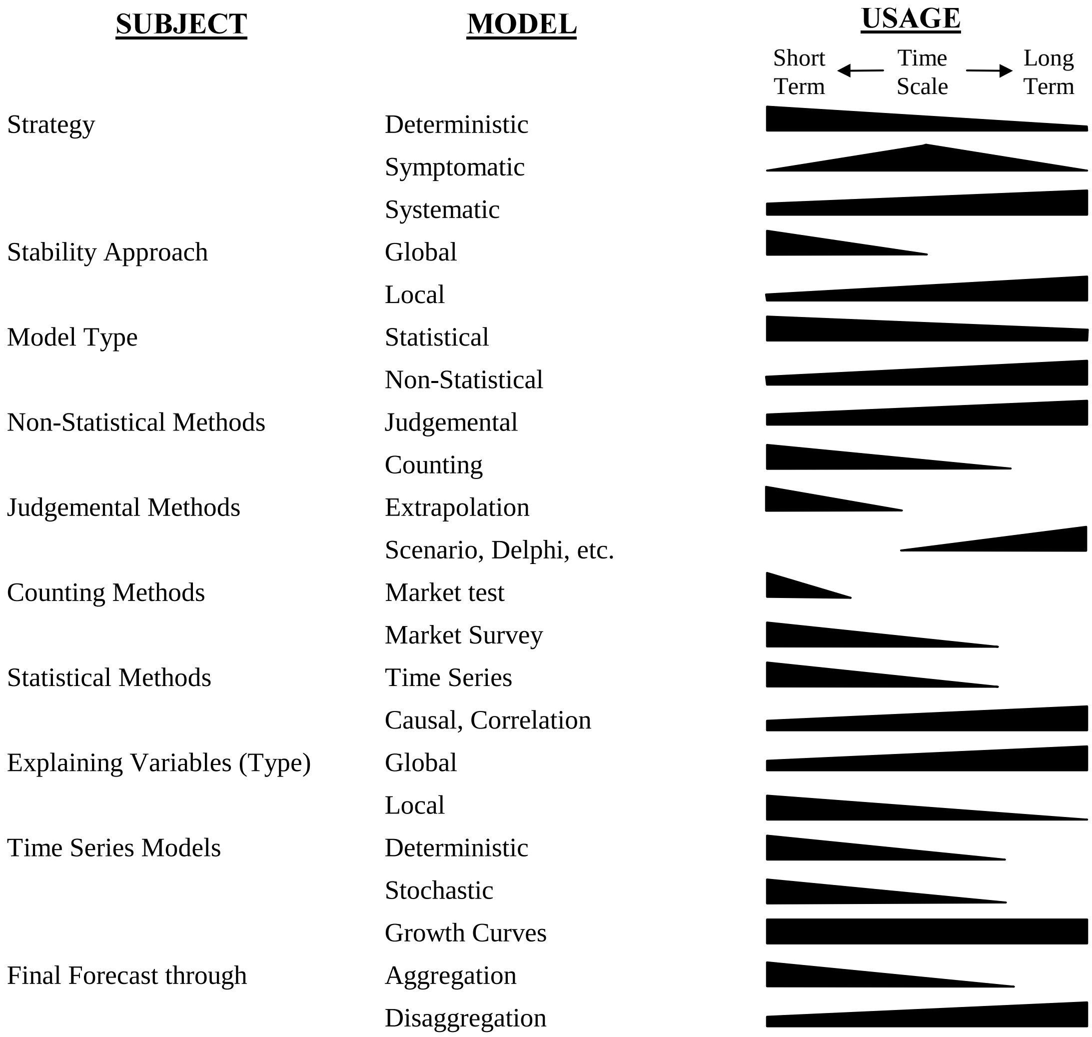


Figure 6.31: Individual model usage as a function of time scale (Example)

iii) Consistency over time

We have seen that no single forecasting method will be adequate for all time spans. Accordingly, different methods must be used for different periods, and this raises the problem of forecast reconciliation. When forecasts for overlapping and non-overlapping time periods are derived by different methods, discontinuities may occur in the growth curve.

No standard procedure can be laid down for resolving these problems, and often, reconciliation between forecasts derived by different methods is achieved by some essentially subjective method, such as the graphical interpolation of a smooth curve between the two forecasting regimes to remove a discontinuity or inconsistency. Such a procedure is illustrated schematically in Figure 6.32.

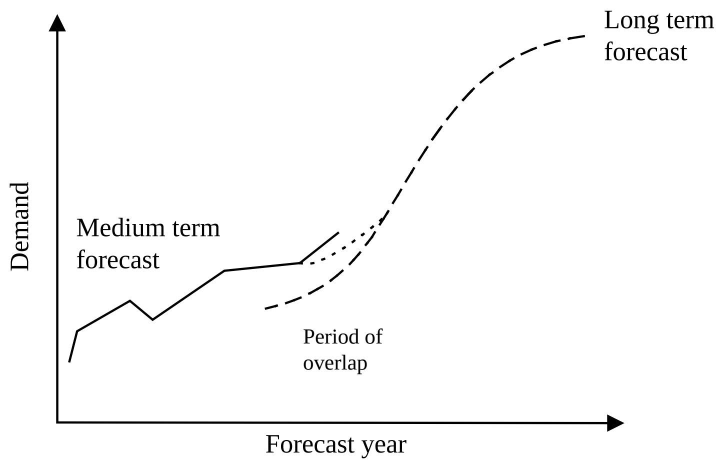


Figure 6.32: Interpolation between forecasts

Although subjective, this elementary approach to reconciliation need not be wholly arbitrary. From his knowledge of the properties of the models, and the degree of faith which he puts in each, the forecaster may give greater weight in the overlapping period to the projection from the model he considers more reliable at that point.

Although the methods of reconciliation between short and medium and long-term forecasts may sometimes be seen as excessively subjective, it is essential that discontinuities due to changes in forecasting technique be avoided. If not removed, they can lead to undesirable “lumpiness” in investment and unjustified policy changes.

b) Geographical level

We have seen that we may have to move both backwards and forwards on the time axis in our dynamic forecasting activity.

In a similar manner, but even more emphasised, we move upwards and downwards between geographical levels.

The purpose of the forecast defines as much as possible the level of the FINAL forecast, but the problem will be to decide upon appropriate intermediate working levels, i.e. the INITIAL forecast levels. If an initial level is different from the final one, then of course aggregation or disaggregation must take place after the corresponding initial forecast. Often we must work in both directions, thus applying the top down/bottom up principle.

The final forecast level is the same as the PLANNING level, and another important question is then to define the necessary input for planning, for example the degree of desired subscriber segmentation, subdivision of the area etc. The forecast output should of course be arranged in exactly that way.

i) Forecast requirements for network planning

For planning purposes, let us define three levels, viz. the national level, where the planning objectives are long distance networks and necessary transit exchanges; the metropolitan level, comprising multi-exchange networks, local and tandem exchanges, remote subscribers units and subscriber networks; and finally the rural level with rural exchanges and remote subscriber units.

Figure 6.33 intends to illustrate the main factors affecting the planning of the different networks and switching units, if by planning we mean the process of deciding types, sizes, functions, locations, logical and geographical boundaries etc.

PLANNING LEVEL	PLANNING OBJECTS	RELATIVE IMPORTANCE OF MAIN FACTORS	
		SUBSCRIBER DISTRIBUTION	TRAFFIC DISTRIBUTION
NATIONAL	The long distance network Transmission systems Transit exchanges		
RURAL	The rural network Transmission systems Rural exchanges Remote subscriber units		
METROPOLITAN	The Multi-Exchange network Remote subscriber units Transmission systems Local & tandem exchanges The subscriber network		

Figure 6.33: Main Factors for network planning relative geographical level

On the national level, we see that the future traffic distribution is the dominating planning factor.

If a subscriber forecast is required as an input to the traffic forecast, it will typically comprise the number of subscribers per district, segmented in residential and business subscribers, the latter possibly divided into single-line and multi-line. These subscriber forecasts will then often be used as some of the explaining factors in direct high-level traffic forecasts models where main trend and global economic variables are other important factors. The subscriber forecast is usually made directly (no aggregation).

At the metropolitan level, it is seen that both subscriber and traffic distribution are highly important. For the detailed exchange planning as well as for the exact planning of the subscriber networks, the exact location of each subscriber is necessary.

For the planning of exchange locations and boundaries and for inter-exchange planning, only the subscriber distribution per traffic area is required. This planning also requires, however, also the future traffic interest distribution between all traffic areas, and since such a traffic forecast requires fundamentally an analytical approach due to very different subscriber behaviour, the subscriber forecast for each traffic area must be segmented in different categories.

Finally, the rural level is characterised by a structure where the population is not so much evenly scattered over the country but rather concentrated to a range of small to large villages and towns.

The particular structure of the appropriate telecommunications networks implies that inter-village traffic volumes are usually not needed for planning, only total outgoing, incoming and internal traffic per village is required. This means that we need only forecast the number of subscribers per village. Segmentation of the subscribers is rarely needed and would often be difficult to obtain. On the other hand, villages can be of very different economic types, characterised by for example farming, tourist activities etc.

One solution may thus be to implement a village categorisation system which would have one, two or more dimensions, for example socio-economic category, administrative level (commune administration, schools etc.), size, general development trend (up, down,...) etc. Standardised data could be for example saturation densities, saturation calling rates etc.

A part of the forecasting activities would then be to continuously revise either for example saturation density levels etc., or village categories.

ii) Forecasting procedures for exchange planning

The bulk of capital expenditure is used for equipment to serve an exchange area and thus a close study is required of each exchange.

Patterns of growth vary considerably between differing types of exchange area according to the character of the locality and the nature of employment within and around the area. In addition to this local knowledge it is necessary to have a background of broad national assumptions about the future on which forecasts are based. These assumptions about an Administration's policies on prices etc., together with social and economic change mean that forecasts need to be issued regularly.

Forecasting from the bottom up: line plant

To meet the needs of forecasting and planning, line plant forecasts must be made in components which can readily be summarised. To achieve this objective the division of the exchange into forecasting units is determined by several limiting factors which emerge from a study of the make-up of an exchange area. An urban area generally consists of three different types of territory:

- The commercial and/or industrial centre.
- The surrounding residential area.
- The outer rural fringe.

In drawing boundaries within this area an attempt to separate the three types will be made, for each is likely to have different problems. Of equal importance are the natural boundaries of main roads, railways, rivers, etc. and these sometimes take preference in the division of the exchange into units.

Traditionally, this unit was the cabinet (or cross-connection point, CCP) i.e. the end point of the primary network. Nowadays, there could still be cabinets, but alternatively, remote subscriber switching units could be used, or possibly remote subscriber multiplexers. However, in this section irrespective of the equipment used the end point, which is the interface between the primary and the secondary network, will be designated as "cabinet".

Planners usually further subdivide the cabinet area into smaller units called distribution points (DPs) and the forecaster uses these DPs when making forecasts. One or several DPs, depending on the eventual telephone penetration, may form a forecasting "section" and the forecaster could divide the cabinet up into sections giving a separate forecast for each. In deciding whether a section shall consist of one or several DPs the forecaster will try to ensure that the DPs which make up the section have similar property characteristics.

The background work which enables forecasts to be produced for planners on request should proceed continuously. It consists of gathering information upon which individual section forecasts can be based. Such information has only relatively short-term value for little is usually known about residential or commercial development beyond the next five years, with the rare exception of new town planning. It is necessary that the closest possible liaison be established with those authorities which can provide assistance in this field. These are usually:

- The local authority, especially its planning department for new development, and the local public health office, who can advise on slum clearance.
- Local architects.
- Large building firms.

In addition it is very necessary to examine all the local newspapers and the journals of the construction industry as well as the planning minutes of local councils.

Most of the local information sources should be visited at regular intervals so that the progress of building development, planning proposals etc., may be regularly updated. The data gathered should be sorted into exchanges, stored on information files and plotted on graphs. This, together with information collected from within the telephone service on exchange growth enables the forecaster to build trend pictures for each exchange.

When a line plant request is received the forecaster should first fix a cut-off date for the revision, and then gather his information. This may consist of the exchange trend graphs and history file plus up-to-date line plant maps of the area under survey and engineering records of the existing subscribers. Both the maps and the records should then be sorted into "cabinets" for this unit is usually most suitable for the forecasting.

Having gathered as much information as possible a detailed survey may start by visiting each street and open space within the forecast area. Notes on the type of property, its age, condition, the number of open spaces, houses in large grounds and the surrounding countryside should be prepared. The likelihood of building occurring on the open spaces and big houses or grounds being replaced by higher density housing should all be noted and marked on the maps. The notes made on the visit should be compared against the known history of the exchange and reconciled to produce the forecast of both tenancies and connections for each cabinet. Since to accord with planning requirements sections never cross cabinet boundaries, it is a simple matter to combine all the sections forecasts contained within a cabinet to produce cabinet figures and in turn to aggregate these to provide exchange forecasts.

Note that when we come to network planning in urban areas, the same basic information about the subscriber distribution is used. This means that the appropriate boundary between traffic areas will probably not follow the streets but will rather pass straight through the building blocks.

Forecasting from the top down: exchange equipment

As has been described, line plant forecasts are the result of highly detailed area investigations somewhat tempered by social, national, economic and policy considerations. Of necessity they rely heavily on the opinions of field forecasters and it is proper that these opinions should be tested against the views of others. Just as sections can be aggregated into cabinets and cabinets into exchanges, so exchange forecasts can be combined to produce area forecasts and so on to national level. This may be reversed and in practice national forecasts usually are broken down to areas. The national forecast may be regarded as being authoritative and is then the envelope within which aggregated forecasts for exchange equipment should be contained.

This principle may lead to a need for reconciliation in two steps since the national forecast must be broken down to area forecasts, the areas at the same time being the aggregates of exchange forecasts, which are aggregates of the line plant forecasts.

One way is to agree on subscriber growth per area first, and then break down area forecasts into individual exchange forecasts.

At this point of reconciliation with the line plant, forecasts must begin. The process is not dissimilar to that used to make line plant forecasts except that is not necessary, or indeed possible, to undertake a detailed street by street exercise.

Forecasts could begin by referring to the exchange history file. From this may be noted all or some of the following for example:

- Past trend.
- Current exchange achievement.
- Rate of growth.
- Existing penetration.
- Changes or additions to housing patterns.
- Quality of housing.
- Patterns of commerce or industry and any changes which may be occurring.
- The policy on housing and industry which could be applied by the local authority.

Together with these items will be considered a mixture of local, regional and national factors to produce connection forecasts based on a carefully balanced blend of what has happened in the past allied to a reasoned review of the life style of the people who live in the area, the houses they occupy, the work they do, the rate at which the area is expanding or contracting, whether the area is attractive to industry and many other factors.

Finally, the exchange forecasts could be checked using for example:

- Graphs of achievement and past forecast.
- Details of all previous changes to forecasts and the reason for them.
- The history and the forecast of the exchange size as a percentage of the area size.
- Known plans of residential and business development.

The forecasts obtained by this approach should be subject to continuous scrutiny to maintain them in as up to date conditions as possible.

iii) Choice of models

We have seen that, in a way, there are great differences between forecasting procedures- detailed forecasting usually being based on detailed investigations and to a great extent on judgmental methods, while overall forecasting may be much more based on statistical, deterministic or systematic approaches.

But actually, these differences depend not so much on geographical level as on the planning objective, i.e.: line plant and exchange equipment planning, or network planning including topology (e.g. exchange locations and boundaries).

Considering the network planning requirements, i.e. defining low level as typically rural, medium level as typically metropolitan, and high level as being the national level, then it is found that the usefulness of a certain model is not so much affected by a difference in level as it is by different time horizons.

Figure 6.34 is intended to illustrate the usefulness of different models as a function of geographical level. The lowest level corresponds to detailed line plant and exchange equipment planning; above that, the diagram refers approximately to the rural, metropolitan and national levels. Note that the diagram does NOT show how important each model is as compared to others

SUBJECT	**MODEL**
Strategy	Deterministic
	Symptomatic
	Systematic
Stability Approach	Global
	Local
Model Type	Statistical
	Non-Statistical
Non-Statistical Methods	Judgemental
	Counting
Judgemental Methods	Extrapolation
	Scenario, Delphi, etc.
Counting Methods	Market test
	Market Survey
Statistical Methods	Time Series
	Causal, Correlation
Explaining Variables (Type)	Global
	Local
Time Series Models	Deterministic
	Stochastic
	Growth Curves
Final Forecast through	Aggregation
	Disaggregation

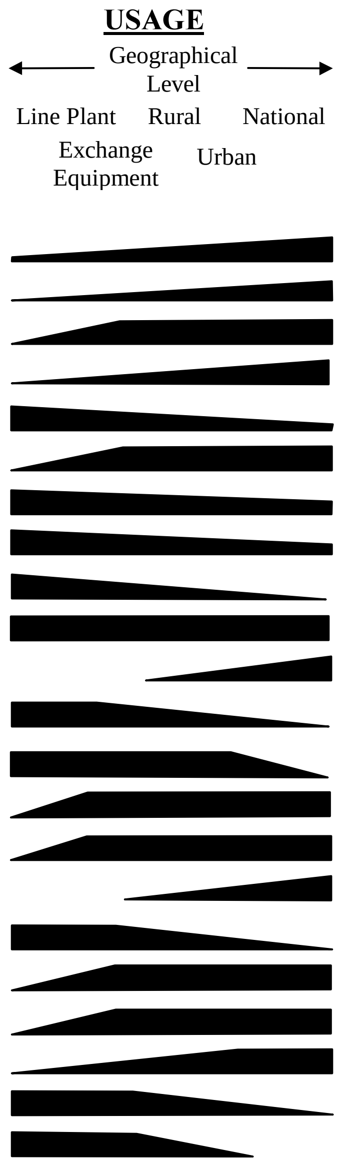


Figure 6.34: Individual model usage as a function of geographical level (Example)

6.2.6 Forecasting schemes

Throughout this section, we have tried to take an analytic view and to recommend systematic schemes rather than temporary, individualistic and isolated solutions; applying these principles on method selection, on the combined use of several methods, as well as on top-down/bottom-up reconciliation, and subscriber segmentation principles, etc.

At the same time, however, it has been argued that judgmental methods and subjectively based decisions are essential parts of the subscriber forecasting process, and that no two identical forecasting situations exist.

These claims are not contradictory; first of all, the forecasting scheme must be adapted to each specific situation, secondly, the forecasting scheme illustrated in Figure 6.4, § 6.2.2, is a framework within which the complex structure will vary from case to case.

The idea of systematic schemes and analytical thinking is that, without this approach, the forecaster would be left in an unknown and confusing situation when confronted with a reality that most certainly deviates from his predictions.

Causalities, elasticities and trend shifts would then be difficult or impossible to detect or at least to properly quantify; and the valuable possibility of analysis of the deviations may be lost. An analytical approach and a standardised, systematic work flow will facilitate a continuously improved forecasting methodology. The forecasting scheme should be designed in such a way that all models and algorithms are easily replaceable by alternative ones.

a) Initial data

An important and powerful source of information is the skill and experience of the forecaster, especially since he is to take a very active part in the forecasting process, always ready to take into account new combinations of intermediate results and to consider alternative, perhaps novel algorithms.

He has then to manipulate mainly two kinds of initial data, i.e. fixed and adjustable numerical input data, and non-numerical input data.

i) Numerical input data

A part of the fixed numerical input data are data which have already been scrutinised, revised and adjusted and are thus, ready for use. Typically, such data are prepared by other organisations and consist of population statistics and forecasts, economic data, etc.

Other collected numerical data are known as raw data, and should of course be checked from a reliability point of view. The checking should be active, i.e. suspicious and unreasonable values must be removed, and other appropriate adjustments should be made, e.g. in order to eliminate biases caused by particular circumstances, such as bad network performance, non-adequate marketing conditions, etc.

There is however another, frequently occurring obstacle which is encountered when the form of the collected and adjusted input data does not match the real need for segmentation. The possibly most serious problem of this type concerns the forecast of traffic interest matrices, where it actually is impossible to derive, unambiguously, the traffic interests from the subscriber distribution and the corresponding calling rates, even if both of these are specified as per traffic area and per subscriber category. Furthermore, if there are, say n categories, then n x n traffic interest matrices are needed for a perfect final forecast.

It is especially essential for the subscriber forecaster to be fully aware of these problems, since the subscriber forecast is one of the cornerstones in the total forecast process. In addition, the subscriber forecast part itself suffers from this kind of problem.

As an example, say that we are to forecast the number of subscribers in each zone, where the zones are the interfaces between administrative and service areas, and that we need to segment into residential, professional type 1,2,...etc., subscribers, but that data unfortunately are available only for single, 2 to 5, 6 to ...etc., lines. We would then need to break down the available data into single line residential, single line professional of type 1,...etc., subscribers, and then aggregate these subsets into the desired segments. We will however find that this decomposition can not be obtained from the available data, unless significant additional information is given.

This is now rarely the case, but observing that aggregation can always be done, and that very valuable sources of information nevertheless exist, namely experience, local knowledge, and comparative data from other areas, we find that an analytic approach based upon very primitive hypothesis testing can be taken.

The available data (which are aggregates of a form we do not like) can be split up into basic sets in a number of different ways, but the application of a reasonable hypothesis will direct us to one set of subsets.

This set, being a set of hypothetical basic data, may now be aggregated into the same form as the available data. Thus we have two sets of aggregate data, viz. the available aggregate data, and hypothetical aggregate data. These two should of course agree, provided that the available data are reliable, and that our hypothesis is correct.

Since collected and adjusted data never can be completely “true” and no hypothesis can perfectly describe the real world, we can of course not expect total agreement, but clear disagreement indicates a necessary revision of our hypothesis and/or a further scrutinisation of the available data.

When the hypothesis works satisfactorily, we may accept aggregates of the revised hypothetical basic data as suitable forecast data.

On a dynamic basis, of course, analyses of the quantitative disagreement between forecasts and real outcomes will produce additional information to be used for a continuous improvement of our hypotheses.

The formal procedure is as follows:

We define: SFD = Suitable Forecast Data

AAD = Available Aggregate Data

HAD = Hypothetical Aggregate Data

HBD = Hypothetical Basic Data

The relations between sets and subsets is illustrated in Figure 6.35 and the corresponding logical relations between the sets are shown in Figure 6.36.

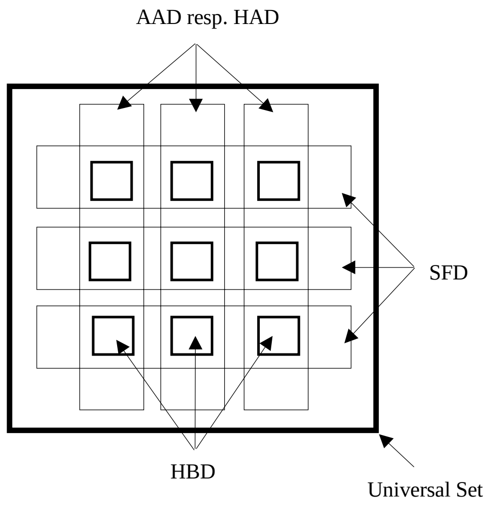


Figure 6.35: Relations between data sets and subsets

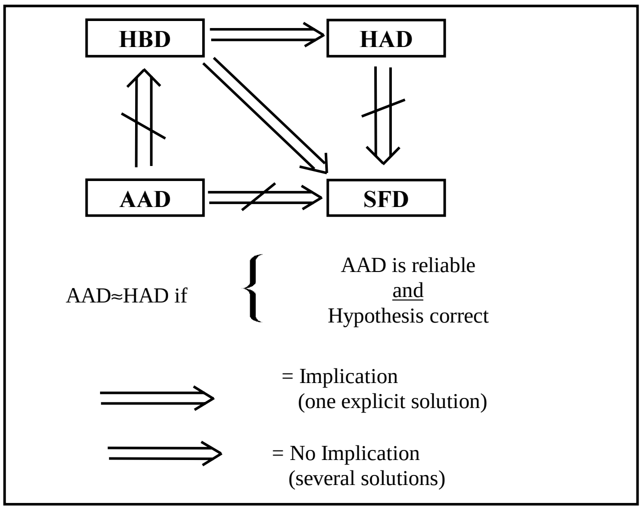


Figure 6.36 : Logical relations between sets

The total flow chart of the process of estimation of suitable forecast data from collected raw data is indicated in Figure 6.37.

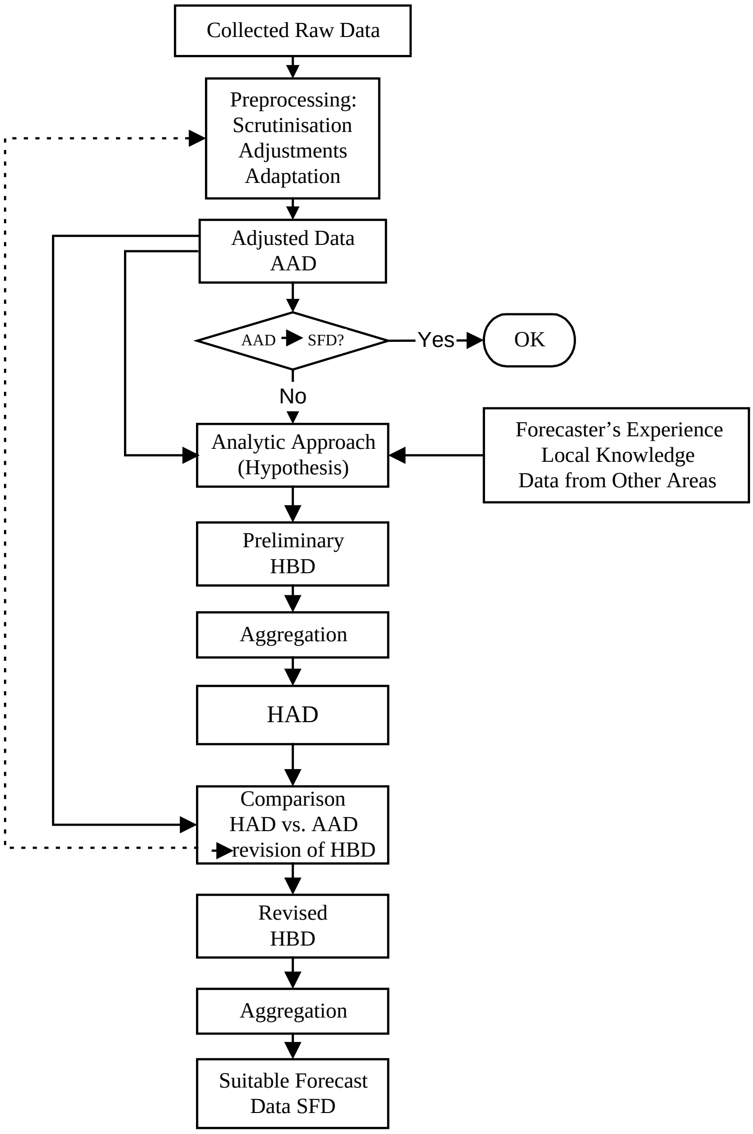


Figure 6.37: From collected raw data to suitable forecast data

ii) Adjustable numerical input data

The adjustability implies that the given data are more or less uncertain, based on recommendations, beliefs, limited, biased or non-targeting measurements, or estimations from various sources.

This may concern both past and present data as well as future trends or levels. If they are statistically based, confidence intervals may in principle be provided; in most situations, however, this is not the case.

In general, this uncertainty emphasises the need for an analytic approach including hypothesis testing by aggregation and check against given data, combined models, top-down/bottom-up procedures and continuous follow-up studies. Mathematical modelling of course facilitates this complex structure.

iii) Non-numerical input data

These data are in the form of verbal descriptions, for example concerning administrative or service areas or smaller entities.

The descriptions are usually given in terms of historical background and development, present levels, trends, and future levels. Indicators may be demographic, social and economic structures; the administrative, commercial and educational role in a more global hierarchy, e.g. no support, self-support, supporting a wider area, also its level, for example that the area includes a university, colleges etc.; and more specific items such as large enterprises.

This information can be used in at least two different ways. One way is simply that the forecaster should bear it in mind when he analyses the forecasting situation, scrutinises numerical indata, makes hypotheses, analyses intermediate results, etc. Another way, which initially requires more effort but may simplify the further work and also may be more fruitful in the longer perspective, is to digitise the verbal information, i.e. to characterise each area using a finite set of parameters.

This is quite reasonable since every description tries to express the qualities of the specific subject in terms relative to a reference frame which in its turn is a sort of average of all subjects in the more global set of subjects. This means that words such as None, Low, Medium, High, Very High etc., may be digitised into for example 0,1,2,3,4 etc. We may of course use more sophisticated scales as well. By proceeding in this way, all areas may be categorised in one, two, or several dimensions.

Some dimensions may be direct measures, such as saturation density levels, density growth rates etc.; other dimensions being indirect, for example administrative level = 3, economic trend = 2, etc.; in the latter case the parameter values are used in models of essentially causal type.

The advantage of such a digitisation and of the corresponding mathematical modelling is obvious: a continuos, analytically based updating of the methodology and/or revision of the parameter values is made possible, i.e. we will know what we are doing and we can keep the forecasting process under control, steadily improving it as time goes by.

b) Subjective judgement and user’s decisions

The properties of the main judgmental methods have been described, and we have also seen how subjective judgement is used in the process of method selection, preparation of suitable forecast data, subscriber segmentation, etc.

Let us now consider the interaction between programmed statistical methods and user’s decisions based on subjective judgement. Figure 6.38 illustrates the process.

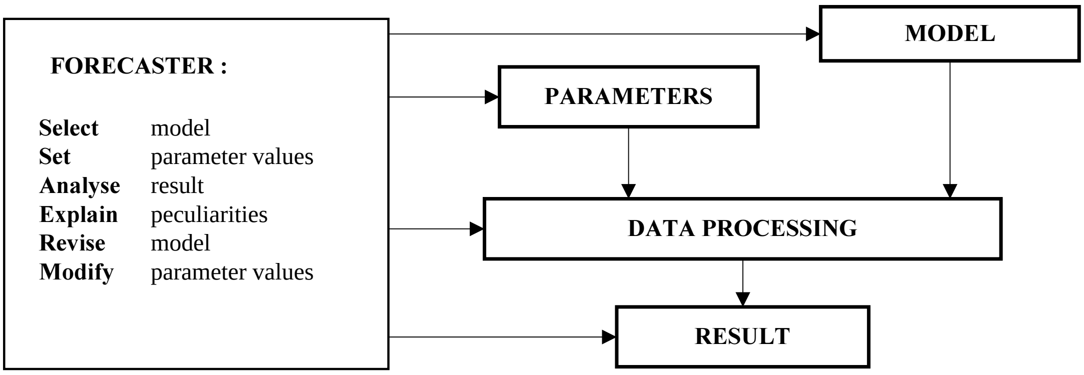


Figure 6.38: Interactive forecasting

A principle of fundamental importance is demonstrated here: the forecaster should not introduce direct changes of the result of the application of a chosen statistical model unless he has very good reasons to do so, since such changes correspond to undefined interference in the statistical analysis; therefore, the quality of the obtained result would also be completely undefined.

The greatest value of the forecaster's subjective judgement lies in the understanding of the mechanism behind the development of demand and in the ability to transfer diversified qualitative properties into digital data.

The human capacity of data processing, or just to consider even relatively small amounts of data, is however very limited, which means that forecasting based on such activities will generally lead to poor results.

Another poor methodology which is frequently used can be demonstrated as follows:

Suppose that the future population distribution in an area comprising a medium sized city and a fairly large rural fringe is to be forecast.

Our first goal may be to forecast the following aggregates:

- the total area (A)
- the metropolitan area (B)
- the rural fringe(C)

Suppose that (A) is forecast through disaggregation of the national forecast; (C) through subjectively judged growth rates; and (B) is estimated as the difference between (A) and (C).

Should one error happen to be positive and other one negative, then the effect on 1 may be very large.

A better way is to estimate all three quantities separately, either applying single methods, or by combined aggregation/disaggregation; and then to check whether (B) + (C) is approximately equal to (A).

Subjective judgement should then be used to design the process, to evaluate the result, and to revise the input parameters.

6.3 Forecasting for metropolitan networks

Introduction

Traditionally, subscriber forecasts are repeated at very long intervals. This is so because they are usually completely manual, i.e. expensive and time consuming and often produce results of poor quality. At the same time, sophisticated models have been developed for traffic forecasting. Unfortunately, these models are often fed with very unreliable data, so the final result is doubtful. The following points should be remembered when forecasting for this type of network:

- The subscriber forecast has a dominating influence on the estimation of locations and boundaries of exchanges and RSUs and the subscriber network.
- The traffic forecast dominates the planning of the junction network.
- The cost of the future junction network may represent only 5-15 % of the total network cost.
- What we can **measure** are the **present, carried** traffics in a possibly far from perfect network, i.e. where the traffic is disturbed.
- What we should **forecast** is **future** traffic **demand** in a hopefully much better network.
- We may have present traffics between exchanges, but we would need quantities between traffic areas.
- The present network carries inefficient traffic, but the future network not, we hope.
- Point-to-point traffic records may be missing.
- We would need to split inter-area traffics between subscriber categories.
- We may have traffics per route, but would need traffics per address.
- We would like to treat traffic values differently, depending on how reliable they are.
- Traffic profiles may be suppressed due to a poor network.
- Traffic records contain errors both due to sampling and to recording method.

From the above the following conclusions can be drawn:

- Efforts should be concentrated on s**ubscriber forecasting** to try to improve the methodology so that the result is more accurate and so that the forecasts can be easily updated on a regular basis.
- Traffic should not be underforecast. It would lead to bad network performance, while some overforecasting would cause only a minor cost increase.
- Traffic measurements alone are a very uncertain and poor basis for forecasting. Instead of trusting them, **logical** and **analytical** models should be used. Traffic records should of course be used, but only to check the parameter values of the model.

An important aspect is to clearly define the structure and detail of the forecast so that it is exactly what is needed for a proper planning of the future network.

First of all, the service area must be divided into basic zones and traffic areas. Typically for a medium to large city, 50 to 500 basic zones and 5 to 50 traffic areas are defined.

Basic zones and traffic areas are chosen in a way so as to facilitate proper subscriber and traffic forecasting.

The forecast is to be given in the form of the future number of subscribers of each subscriber category for each basic zone, and the traffic forecast is to be based on this detailed subscriber forecast. It is then aggregated into traffic interests between each traffic area.

We see that the final subscriber forecast is thus much more detailed compared to the final traffic forecast, but that the very detailed subscriber forecast is used to make the traffic forecast as good as possible.

To use existing exchange areas instead of traffic areas for forecasting would degrade the quality both of the final forecast and of the result of the planning of the network, as would also be the case if basic zones or subscriber categories were not defined.

This is explained in more detail in the section on “Estimation of traffic forecast parameters”.

The basic zones should furthermore be defined in such a way that each traffic area comprises a number of complete zones, at the same time each present exchange area should also comprise a number of complete zones. In that way, each basic zone belongs completely to one particular traffic area, but also to one particular exchange area.

Note that the **future** exchange area boundaries, which should be a result of the network planning, will often be quite different compared to the present ones.

6.3.1 Forecast of non-residential main lines

The following symbols are used in the text:

T	=	Total area (the whole city, or a relatively large sub-area)
MET	=	METropolitan area
TRA	=	TRAffic area
z	=	zone
t	=	point of time
NR	=	no. of Non-Residential lines
CB	=	no. of Coin Box lines
BUS	=	no. of BUSiness lines
P	=	no. of PBX lines
PS	=	no. of PBX lines, specified for identified, larger enterprises
PN	=	no. of PBX lines, Not specified
B	=	no. of single Business lines
BB	=	no. of single Business lines, registered as Business lines
BR	=	no. of single Business lines, registered as Residential lines
SS	=	no. of Special Service lines
PB	=	Proportion B of BUS (B/BUS)

’ (prime) means intermediate (temporary) value

Preferably, non-residential lines are split into several categories. We can distinguish at least the following types:

- Special Service lines
- Coin Box lines
- Single Business lines
- PBX lines

Each of these types requires a particular treatment in order to obtain the best possible subscriber forecast, eventually arriving at an appropriate traffic forecast.

a) Special service lines (SS)

It can be assumed that special service lines:

- attract only incoming traffic and originate none (or, at least, very little);
- are concentrated on one or, at most, only very few predetermined switching points of the network. Thus, in principle, we can treat the corresponding specification separately, in principle, as follows:

First, define different Special Service functions, e.g.:

a = Ambulance, b = Fire brigade, c = Information services of certain types, d = Other services.

Then determine those points in the network on which these service functions should be concentrated.

Alt.1: All services on the same and only point;

Alt.2: Several points but all services on each point;

Alt.3: Several points, each dedicated to a limited number of services in such a way that each service is concentrated on only a single point.

Alt.4: Several points, at least some of which offer only a limited number of all services; at least some services offered by more than one point.

Using the functions, a-d in the example, these alternatives can be illustrated as follows:

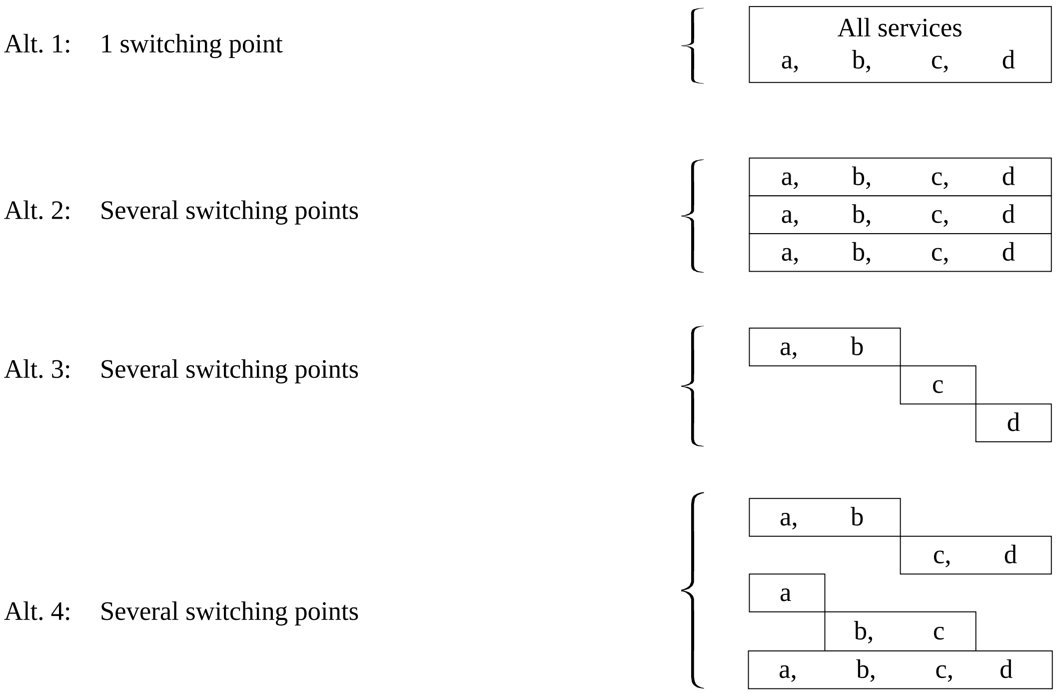


Figure 6.39: Alternative arrangements of special services

If we feel that it is necessary to calculate the corresponding traffic interests very accurately, we would then have to specify the following traffic cases for each traffic area:

Alt.1: Traffic $\rightarrow$ SS (only one traffic case)

Alt.2: Traffic $\rightarrow$ SS (only one traffic case)

Alt.3: Traffic $\rightarrow SS_{a+b}$, Traffic $\rightarrow SS_c$, Traffic $\rightarrow SS_d$ (according to the given example, 3 traffic cases)

Alt.4: Traffic $\rightarrow SS_a$, Traffic $\rightarrow SS_b$, Traffic $\rightarrow SS_c$, Traffic $\rightarrow SS_d$ (4 traffic cases)

For alternatives 2 and 4, the relation should be specified between each Traffic Area and the corresponding Special Service points. For alternative 4, this specification will be relatively complex. Alternatives 1-3 should, however, cover almost all real cases.

b) Coin box lines(CB)

The specification of coin box lines should raise no particular problems. Forecasts can usually quite easily be made for each zone of the city. After aggregation of zone forecasts into traffic areas, a top-down process could be combined with this bottom-up approach. One way to do this would be first to forecast the total number of coin box lines for the entire metropolitan area using a separate method, and then to use this total forecast as an envelope for the traffic area forecasts which obviously must be adjusted. The last step would then be to desegregate the traffic area forecasts down to zone forecasts, with the proportions between the original zone forecasts remaining unchanged.

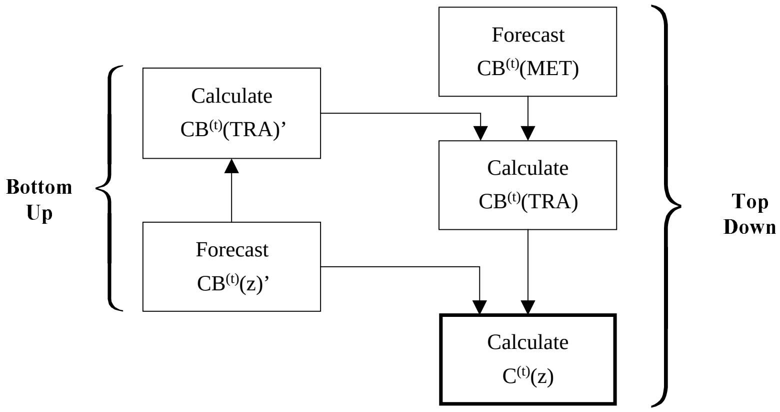


Figure 6.40 : Bottom up - top down forecast of coin box lines

c) Business lines (BUS)

The term "business lines" refers both to commercial lines and to non-commercial professional lines, as well as including official lines, e.g., governmental, hospital, and other similar lines.

Particularly from the traffic behaviour point of view, we can easily identify two main types of business lines, i.e., single business lines (B), and PBX lines (P). With respect to the single business lines, there is some uncertainty; normally lines that are used professionally are also registered as business lines (BB), but in quite a few cases, they may be registered as residential lines (BR), for example when the same premises are used both professionally and as a private residence. PBX lines do not present this kind of difficulty, but their specification may still be split into two different classes, depending on the available information:

One class includes those identified, larger enterprises for which the locations and the required number of lines for defined points of time are known (PS). The other class includes the demand that is not identified as dedicated to specific enterprises, but rather identified as the total demand for an area (PN).

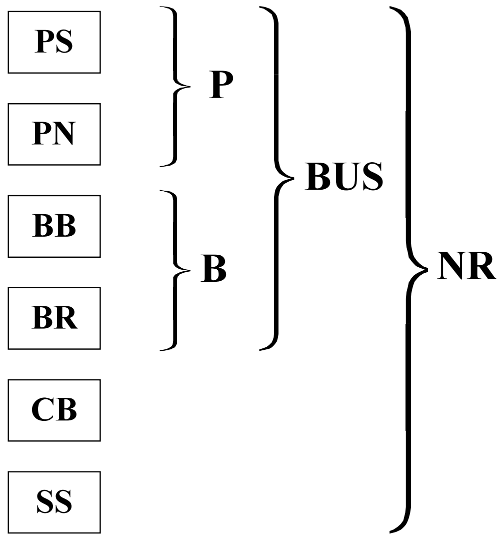


Figure 6.41 : Relation between different kinds of non-residential lines

d) Forecast schemes for single business lines (B) and PBX lines (P)

Depending on how detailed our basic forecasts are, different forecast schemes may be applied. Two alternatives are given below and illustrated in Figures 41 to 45:

Alt.1: Besides possible forecasts of PBX lines for identified, larger enterprises (PS), only the total number of business lines (BUS) can be forecast, separately for zones and for larger areas. A fair estimate of the proportion of single business lines (PB) is available, either for each zone or for larger areas.

Alt.2: More detailed basic forecasts are available, i.e., the total number of PBX lines (P) per zone, and the total number of single business lines (B), separately for zones and for larger areas. The single line business forecasts may be split into the two classes, BB and BR (see legend).

<u>Alt. 1 : Less detailed basic forecasts available</u>

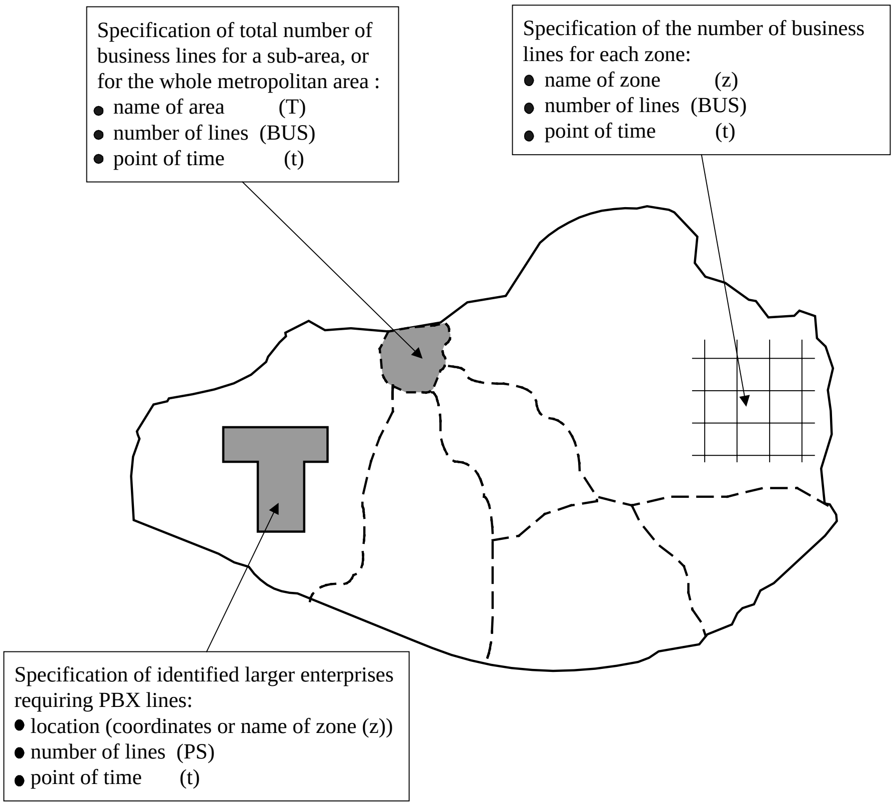


Figure 6.42 :Specification of business lines in a metropolitan area (Alt. 1)

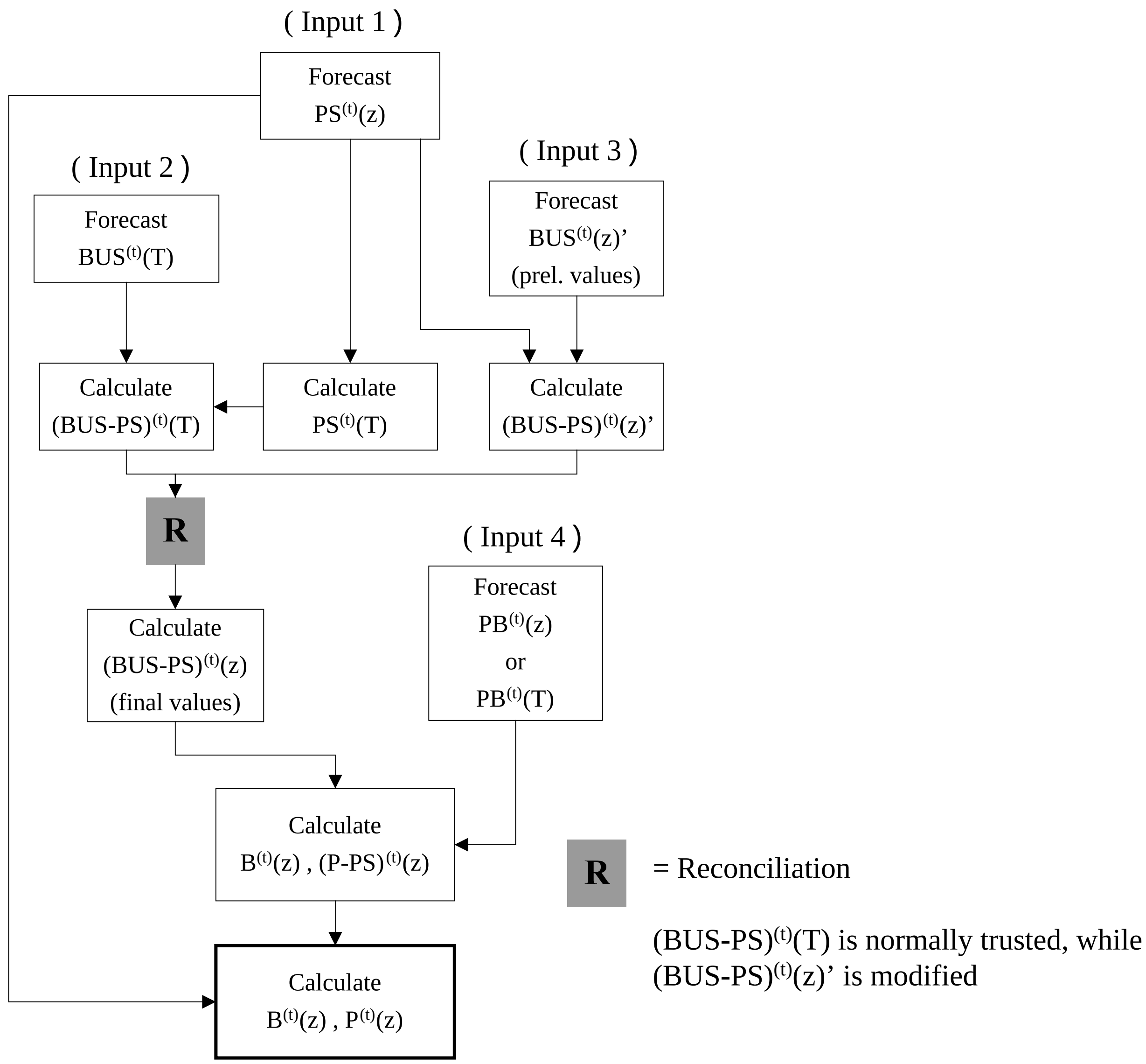


Figure 6.43 : Forecast scheme for PBX and single Business lines, Alt. 1

Alt. 2 : More detailed basic forecasts are available

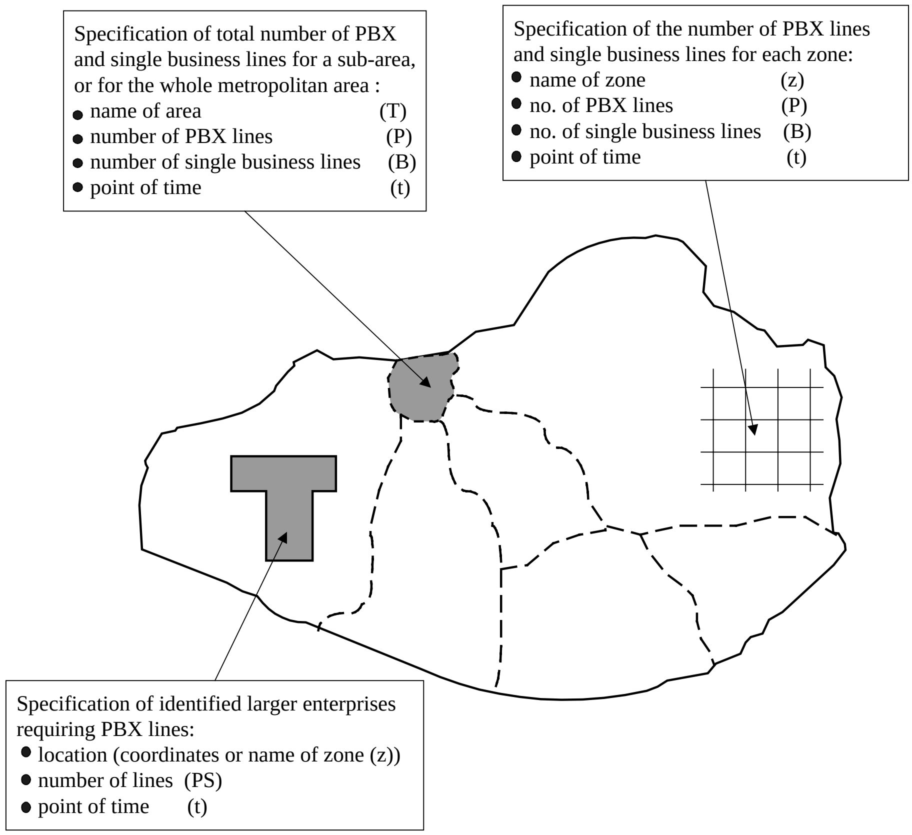


Figure 6.44 : Specification of business lines in a metropolitan area (Alt. 2)

In this case, PBX lines and single business lines can be forecast according to separate schemes.

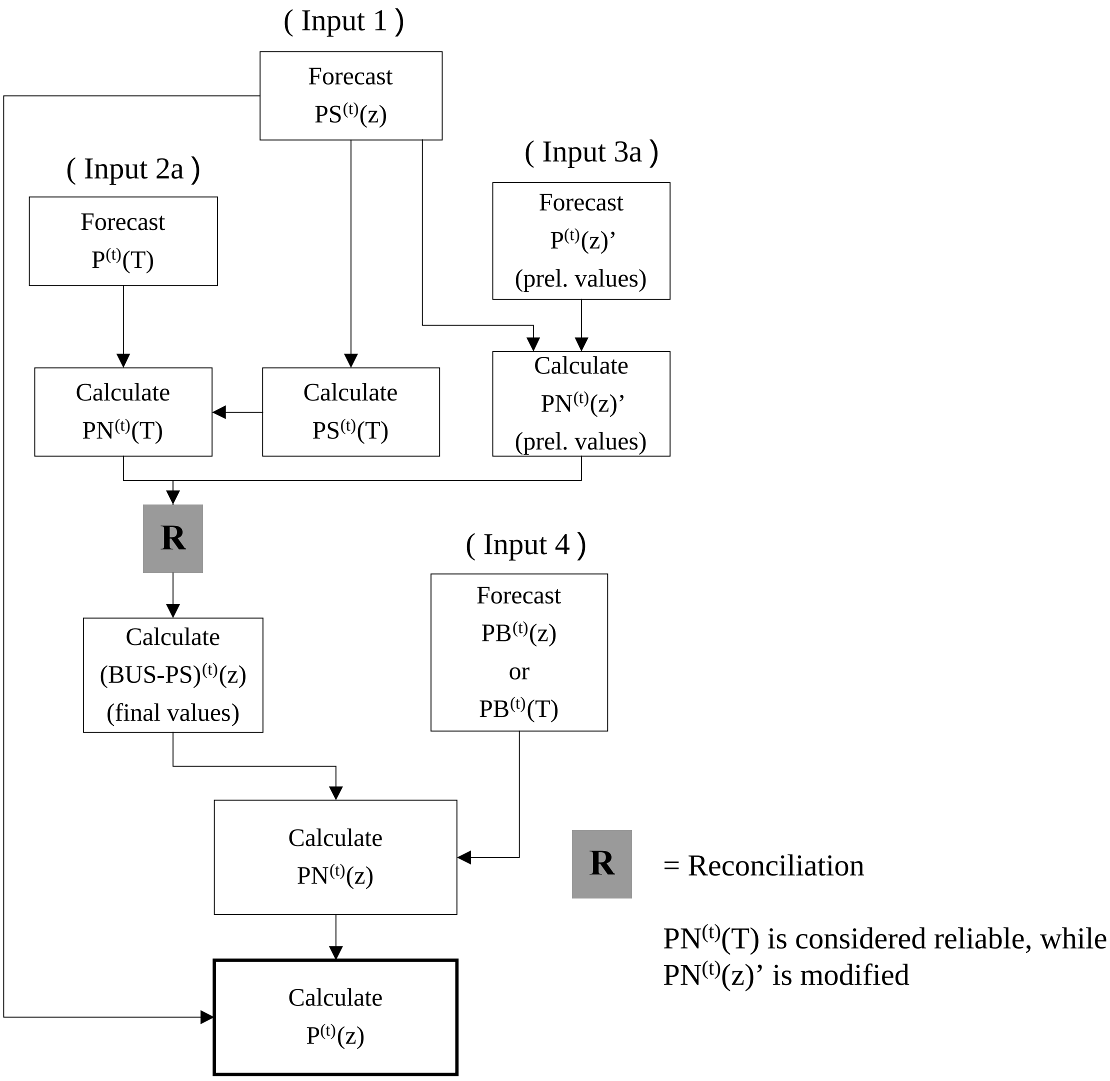


Figure 6.45 : Forecast scheme for PBX lines, alt. 2

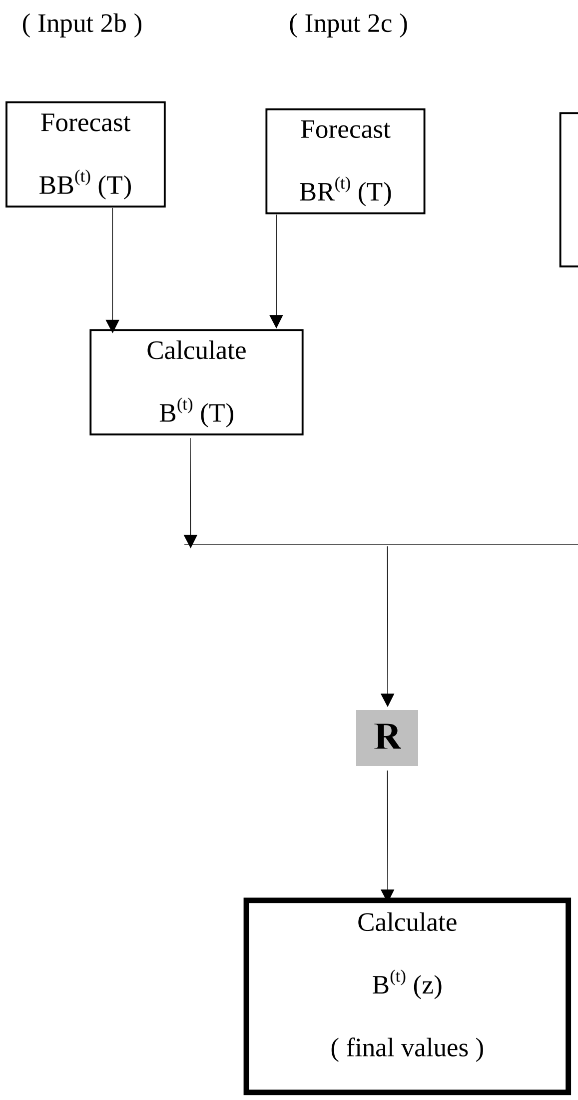


Figure 6.46 : Forecast scheme for single business lines, alt. 2

R = Reconciliation

$B^{(t)}(T)$ is considered reliable, while

$B^{(t)}(z)'$ is modified

e) Inputs to the business line forecasts

The inputs to the previously illustrated forecasting schemes are considered as follows:

Input 1: $PS^{(t)}(z)$

In some cases, it is possible to identify specific, large, future enterprises e.g., "A new hospital x will come into service at a point of time 6 years hence; the location will be in zone A2; 60 PBX lines will be required". Now, if the hospital is the only new large enterprise known in zone A2 until year 6, then $PS^{(6)}(A2)=60$.

Should another enterprise be planned for the same zone (A2), for example, "A large franchise of a sales company will be established in zone A2, in 5 years time, and 40 PBX lines will be required", then $PS^{(5)}(A2)=40$ and $PS^{(6)}(A2)=100$. (40+60=100!) The $PS^{(t)}(z)$ values are accepted (considered to be relatively reliable).

Input 2: $BUS^{(t)}(T)$

This is the total number of future business lines (P+B) in the entire metropolitan area or, if good segmented basic data are available, in some sub-areas.

Several different forecasting methods should be combined since this will be the envelope for the detailed forecasts. Ideally, regression methods based on professional, trade, and socio-economic development would be combined with trend analysis and non-statistical approaches like comparative methods and subjective judgement.

Input 2a: $P^{(t)}$ (T)

The total number of PBX lines (P) should be forecast using the same kind of scheme as used for total business lines (BUS), but with less emphasis on socio-economic factors and more emphasis on trade, business, and official activities.

Input 2b and 2c: $BB^{(t)}$(T) and $BR^{(t)}$(T)

The total number of single business lines (B) should be forecast again using the same scheme as used for total business lines (BUS), but unlike in Input 2a, **more** emphasis should be placed on socio-economic factors.

Depending on local conditions, and on the administration's policy, $BR^{(t)}$(T) can be neglected, i.e., set = 0; alternatively, $B^{(t)}$(T) can be forecast directly, including an estimate of $BR^{(t)}$(T).

Input 3: $BUS^{(t)}$(z)'

This is a micro-forecast by zone. Depending on the available information, many different kinds of direct methods can be used. Subjective judgement will be used to a large extent. Note that it is important to obtain a good **relative** distribution of business lines between the different zones since the values are only preliminary (absolute levels will be modified because **total** values for the entire city will be forecast separately).

Input 3a: $P^{(t)}$(z)'

The number of PBX lines in each zone will probably be forecast on the basis of city planning, information from medium or large sized companies, institutions and official organisations, and on detailed investigations of the development possibilities in the zones.

Again, the **relative** distribution of PBX lines between zones is more important than the absolute numbers.

Input 3b and 3c: $BB^{(t)}$(z)' and $BR^{(t)}$(z)'

The forecast of single business lines in each zone will rely on detailed zone development investigations even more than the PBX forecast does. Furthermore, in this case, the **relative** distribution between zones is of paramount importance.

As was the case with total forecasts, $BR^{(t)}$ (z)' may neglected, i.e., set =0; alternatively, $B^{(t)}$(z) may be forecast directly, including an estimate of $BR^{(t)}$(z)'.

Input 4: $PB^{(t)}$ (z), or $PB^{(t)}$ (T)

$PB^{(t)}$ (z) is preferable. Approximate values are acceptable. The basis for the forecast may be parameters which express the relative structure and character of the different zones.

6.3.2 Forecast of residential main lines

The following symbols are used in the text:

T = Total area (the whole city, or a relatively large sub-area)

z = Zone

t = point of time

RES = number of RESidential lines

HH = number of House Holds in the area in question

HP = Household Penetration in the area in question
Numerical value: from 0 to -or even above- 1.

HPG = Household Penetration - Global trend
Numerically, HPG ranges from 0 to 1.
Verbally, this corresponds to "Low" → "Average" → "High"

LER = Local Economy Relative to global economy.
Numerically, LER ranges from 0 to 1.
Verbally, this corresponds to "Bad" → "Average" → "Good"

HPCT= Household Penetration Curve Type
HPCT = VL means "Very Low" penetration
HPCT = L means "Low" penetration
HPCT = A means "Average" penetration
HPCT = H means "High" penetration
HPCT = VH means "Very High" penetration

HE = Household Economy in the area in question
HE = 1 means "High Income" families
HE = 2 means "Middle Income" families
HE = 3 means "Lower Income" families
HE = 4 means "Poor" families

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

A number of different methods can be used to make the residential lines forecast.

One widely used approach is to forecast each single future extension, aggregating the forecasts from houses to blocks, from blocks to zones, and from zones to exchange areas. One reason for choosing this approach is that the subscriber network must be planned in detail, the objective being short to medium-term planning.

The same forecast is then used for long-term network planning. This process is based on detailed investigations of each spot of the city, arrived at by visiting these spots, making notations on their present state and on the possible further development of different kinds of buildings in all locations, then comparing those notations with those collected on the previous visit, and finally manually estimating the likely number and kind of new extensions to be installed in the near future. The process is extremely time-consuming and is therefore not very frequently repeated.

The accuracy of the process is reasonably sufficient from the block perspective, whereas the accuracy of the process for aggregates is much lower.

For long-term planning purposes, we prefer good long-term forecasts for each zone covering the network structure from the top level down to, usually, only cabinet or remote subscriber unit level.

To achieve this, an alternative forecast scheme is proposed here.

a) From household penetration (HP) to number of residential lines (RES)

If we can forecast the number of households (HH) of a certain area, and if we are able to estimate the household penetration (HP), then we can calculate the demand (RES) using the following expression:

$$RES = HH \cdot HP$$

We can increase the accuracy of the forecast by sub-dividing the total number of households into the different economic levels (HE), provided that we can find good estimates of the penetration rates for those different levels. If we distinguish between four classes, HE=1,2,3,4, we obtain:

$$RES_{HE} = HH_{HE} \cdot HP_{HE} \qquad HE = 1,2,3,4$$

and

$$RES = \sum_{HE=1}^{4} RES_{HE} = \sum_{HE=1}^{4} HH_{HE} \cdot HP_{HE}$$

Now the following difficulty may arise:

For the **total** area (T), e.g., the city, it may be possible to forecast the number of households (HH) for each class HE=1,2,3,or 4, but for some or for all of the **zones**, perhaps only a kind of **weighted** value (HE) can be defined, e.g., for a particular zone (z), we estimate HE = 2.5, corresponding to an **average** penetration of the zone HP = 0.45 (example).

For the total area, we then obtain:

$$RES_{HE}(t) = HH_{HE}(t) \cdot HP_{HE} \qquad HE = 1,2,3,4$$

and

$$RES(T) = \sum_{HE=1}^{4} RES_{HE}(T)$$

whereas for a zone, we estimate:

$$RES(z)' = HH(z) \cdot HP(z)$$

If, however, HH per income level (HE) is known for a certain zone, then for that zone, of course,

$$RES_{HE}(z)' = HH_{HE}(z) \cdot HP_{HE} \qquad HE = 1,2,3,4$$

and

$$RES(z)' = \sum_{HE=1}^{4} RES_{HE}(z)'$$

We see that two elements are needed to make this kind of estimate, namely household penetration (HP), and number of households (HH).

In the formulae above, prime (') has been used to indicate a preliminary value since zone forecasts obtained in this way are normally used as an input to a somewhat more complex combination of forecasting methods where zone forecasts are matched against total values and are thus adjusted in the corresponding process as shown in Figure 6.47.

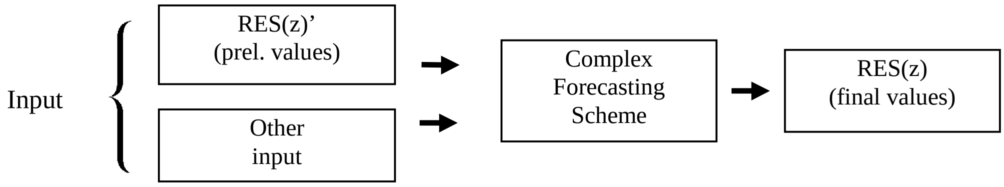
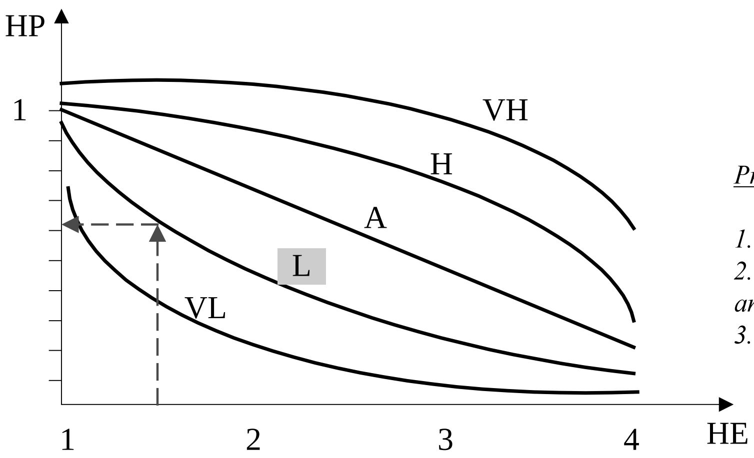


Figure 6.47

b) Household penetration (HP)

Household economy (HE) is only one of several variables which describe household penetration (HP); it is certainly a very important variable. Its relative significance is usually greater in countries where the **total** penetration rate is low compared to countries where telecommunications are already well developed. This results from a number of factors, such as the fact that even relatively poor families in rich countries are better off than their HE level counterparts in poorer countries. And, in addition, telecommunication tariffs are usually much lower in the wealthier countries.

It is possible to graphically represent these conditions such that the curves (HPCT) show household penetration (HP) relative to household economy (HE), each HPCT curve corresponding to the stage of relative economic development combined with the relative household penetration, in the particular area.

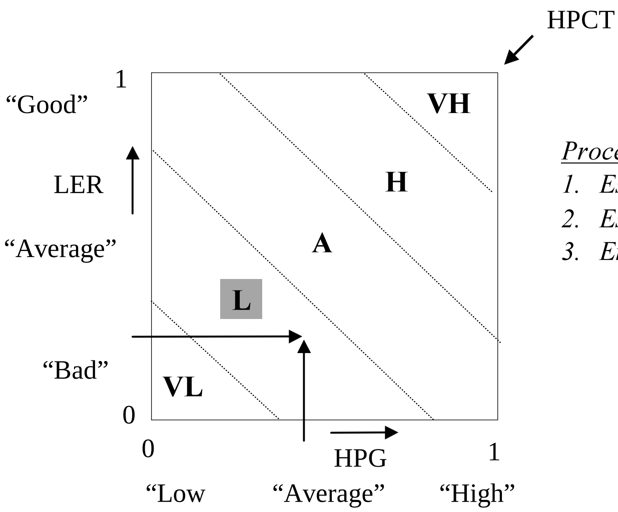


<u>Procedure</u> :

1. Choose curve HPCT (e.g. L=Low)
2. Enter HE (e.g. 1.5=between High Class and Middle Class households)
3. Read HP (e.g. 0.61 = 61% penetration)

Figure 6.48: Curves for determining penetration rates

A decision table, as shown below, can be used to choose the curve type (HPCT).



<u>Procedure :</u>
1. *Estimate HPG (eg ≈ 0.5 = “Average”)*
2. *Estimate LER (eg ≈ 0.25 = “Bad” to “Average”)*
3. *Entering HPG and LER, read HPCT (eg L = “Low”)*

Figure 6.49: Decision table for determining type of curve

Note that an area under study may correspond, for example, to HPCT=L at the **present** (point of) time, but may be expected to increase, e.g., to HPCT=A at a future (point of) time!

LER is defined as “Local Economy Relative to global economy”. The terms “Local” and “Global” need to be explained:

Definition of “Local”: If, from a socio-economic point of view, the character of the metropolitan area under study differs significantly from the other cities in the country, then the term “local” applies only to that area under study. If, however, the characters of the cities in any given country are fairly uniform, then “local” can mean all the cities in that country.

Definition of “Global”: that part of the world which most significantly influences both the economy and the social development of a particular country.

The method described above can be used for all levels, i.e., zones, traffic areas, and the total area. It is of the most value, however, when applied to investigations per zone as shown in Figure 6.49.

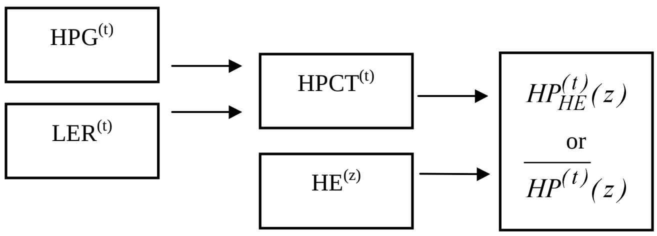


Figure 6.50

For higher levels (T), complex combinations of methods using several different, important forecasting methods and sources of data should be used as indicated in Figure 6.51.

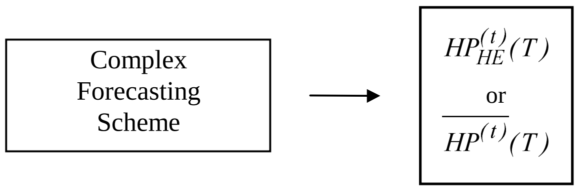


Figure 6. 51

c) Number of households

Two separate forecasts should be made; the first is a forecast per zone, based on detailed (micro) methods, where each zone is analysed in detail, and the second is total forecast, where a number of different methods and types of data are combined in a complex combination of methods, using, above all, socio-economic and regression models as indicated in Figure 6.52.

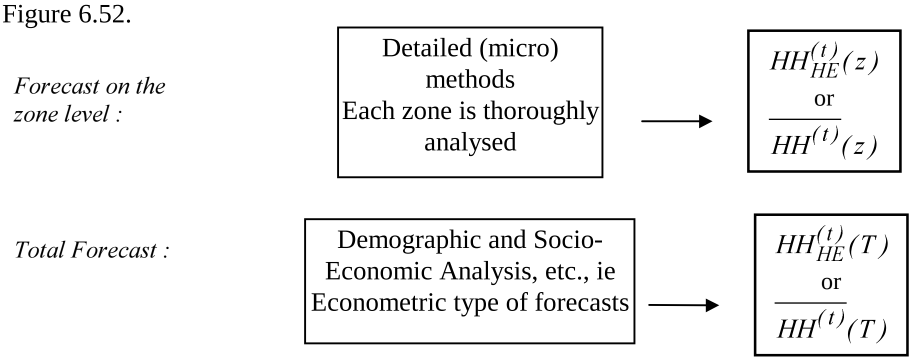


Figure 6.52:

d) Final forecast

In order to estimate the future number of residential lines per zone $(RES^{(T)}(z))$, two alternate paths can be used, depending on the supply of available data (see Figure 6.53).

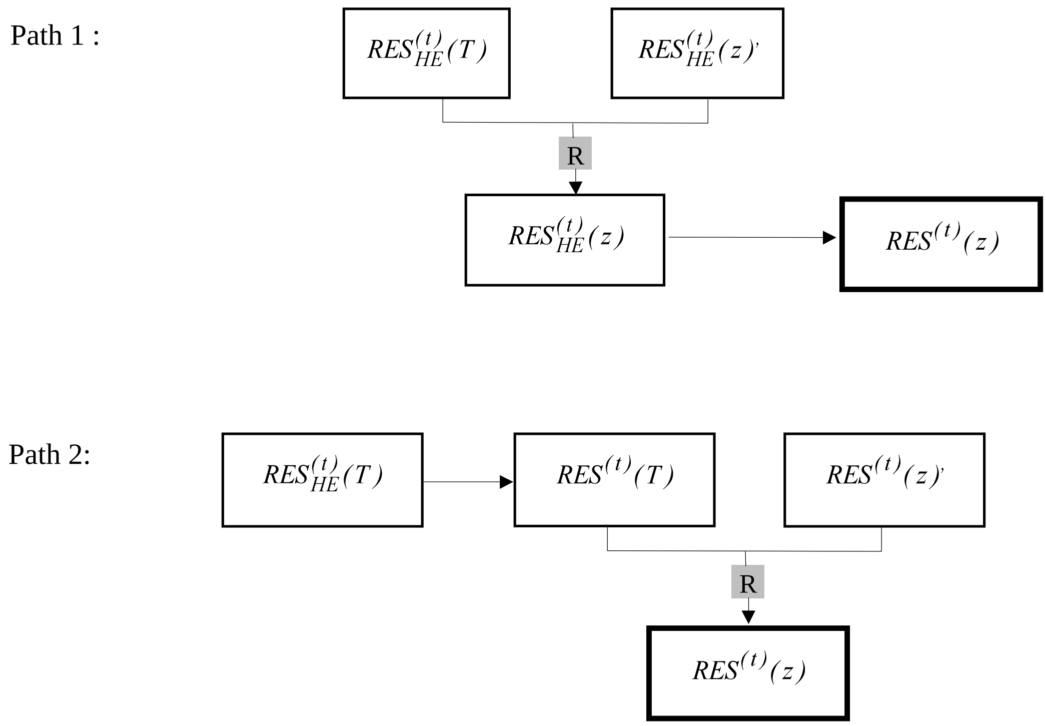


Figure 6.53

R means Reconciliation between total values and zone values. Normally, Totals should be accepted and zone values accordingly adjusted so that the sum of the adjusted values agrees with the Total value.

"Total" may refer to the entire metropolitan area, but if background information so permits, the entire area can be sub-divided into a number of sub-areas where a number of "Totals" must be defined.

In such a case, it may be possible to combine Path 1 and Path 2, e.g., if zone values $RES_{HE}^{(T)}(z)'$ (by household income level) are known for some sub-areas, but only $RES^{(T)}(z)'$ (zone totals) are known for other sub areas.

6.3.3 Estimation of traffic forecast parameters

The planning of a telecommunication network should be based upon a sound traffic forecast. A reliable traffic interest matrix is needed but is, however, difficult to achieve since recorded traffic data may be incomplete, of varying quality and perhaps not relevant in a future situation. The methodology presented here concentrates on the construction of the present traffic interest matrix, and it is hypothetical insofar as it constructs the matrix from assumed traffic characteristics using, as far as possible, the available recorded traffic data. It works step-by-step, correcting the assumed model parameter values between the steps, and it takes conceivable future changes of traffic characteristics into consideration. The scheme has a modular structure, i.e., the models are replaceable.

A forecast traffic interest matrix is needed in order to plan a telecommunication network for any future point of time T. An element of the matrix $A_{kl}^{(T)}$ should preferably denote the individual traffic interest from any traffic area k to any traffic area l. A commonly used forecasting procedure is based upon the assumptions about present traffic interests $A_{kl}^{(0)}$, the present subscriber distribution $n_k^{(0)}$, and a reliable forecast of the future subscriber distribution $n_k^{(T)}$. Furthermore, such a forecast should be made for each class of subscribers separately, the total forecast then being the aggregate of the separate ones.

More work has been carried out on the study of traffic growth models than on the study of the present traffic interests $A_k^{(\,)}$. In fact, the preparation of such a matrix presents great difficulties. An existing network usually contains a mixture of different types of analogue equipment, in many cases both crossbar and step-by-step systems. Network losses are often quite high, which would suggest high rates of repeated call attempts. Especially in step-by-step networks, such repeated call attempts cause abnormal holding times and considerable additional inefficient traffic load on the interconnecting routes. There are no, or only very limited, possibilities for traffic or call dispersion measurements; nor can the recorded route traffics be used for the calculation of traffic interests, since they carry not only inefficient traffic, but also an anonymous mixture of calls of different origins and destinations.

Even if we did have a procedure for deriving the present traffic dispersion from traffic records, such a matrix would not be a relevant basis for a sound forecast, since the future network is supposed to offer improved service and less inefficient traffic compared to the present one, to show a changed traffic profile, and perhaps be subject to changed subscriber behaviour due to changed tariff policy, etc.

Summarising these problems, we find that $A_k^{(\,)}$ is:

- generally impossible or difficult to obtain from traffic records;
- of varying quality; some values will be most uncertain, others will be missing.
- not relevant in a future situation.

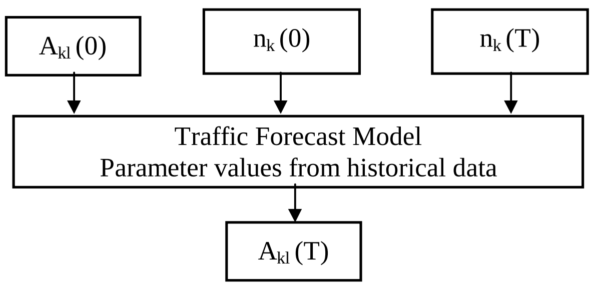


Figure 6.54: Traditional forecasting scheme

What we really need is a method which utilizes available recorded data as much as is reasonably possible but which is not absolutely dependent on a complete supply of such data. This implies a considerable amount of individual judgment and decision making, i.e., the model must be mixed.

The main idea of the procedure presented here is to define traffic parameter values which can be checked against traffic records in order to ensure, as far as possible, that they do not disagree with the present traffic situation. The parameters must be suitable as a basis for the future traffic interest forecast, which means that it must be possible to update the values according to expected changes in subscriber behaviour and network quality.

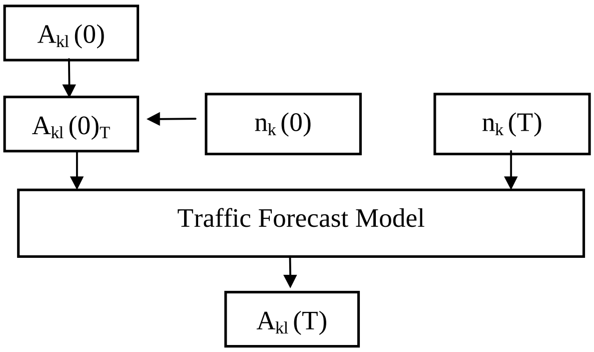


Figure 6.55 : Mixed model

a) Basic parametres

The following definitions apply:

A = Total traffic
a = Traffic per subscriber (main line)
y = Call intensity
h = Holding time
K = Dialing time per digit
B = Congestion level
R = Routing vector
d = Dispersion factor
W= Traffic interest weight
n = No. of subscriber lines

Subscripts:

b,c= Subscriber class no.
k,l= Traffic area no.
u,v= Exchange area no.
r = Route no.
o = Originating
t = Terminating
· = Total amount
0 = Present time
T = Future time
* = Recorded quantities
'(prime) means "intermediate (temporary) value"

i) Subscriber classes

A number of subscriber classes should be defined. A subscriber class should be reasonably homogeneous as regards traffic level and subscriber behaviour. It must, of course, also be possible to estimate the present and future distribution of the number of subscribers per class. Examples of subscriber classes are:

a) Residentials, high and middle class.
b) Residentials, lower class.
c) Single business lines of various kinds.
d) Lines to small PBXs.
e) Lines to larger PBXs.
f) Coin boxes.
g) Data users, switched lines.
h) Data users, leased lines.

ii) Traffic areas and exchange areas

An area where a telecommunication network exists is divided into a number of exchange areas. Traffic records are related to these exchange areas. In favourable cases, we may know some present traffic interests between exchange areas $A_{uv}^{(0)}$, as well as the number of subscribers per class b in each area, $n_{bu}^{(\ )}$. For planning purposes, however, we need to forecast the future traffic interests between traffic areas $A_{k}^{(T)}$ rather than $A_{uv}^{(T)}$. Furthermore, we want to make separate forecasts for different subscriber classes and then aggregate those into a total forecast.

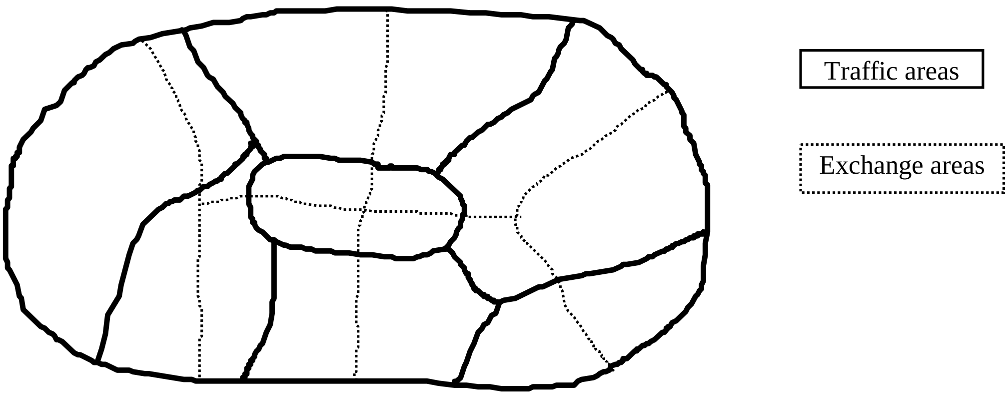


Figure 6.56: : Traffic areas and Exchange areas

This means that we should divide the entire area into traffic areas. Since we need to translate back and forth between exchange areas and traffic areas during the forecasting process both for number of subscribers per class and for traffic interests, each traffic area should be relatively homogeneous from the subscriber class point of view.

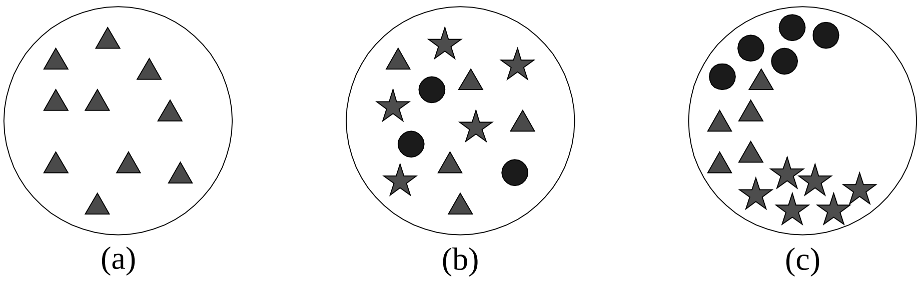


Figure 6.57: Traffic areas
(a) One subscriber class: suitable traffic area .
(b) Several classes, but well mixed: suitable traffic area.
(c) Unsuitable traffic area.

Under such circumstances, we can always calculate quite simply:

$$n_{bu} = \sum_k \frac{n_{ku} \cdot n_{bk}}{n_k} \qquad \text{and} \qquad n_{bk} = \sum_u \frac{n_{ku} \cdot n_{bu}}{n_u}$$

since

$$n_{bku} = \sum \frac{n_{bu} \cdot n_{bk}}{n_k} \qquad \text{and} \qquad n_{bu} = \sum_k n_{bku}$$

where n_{bu} = no. of subscribers of class b in exchange area u;
n_{bk} = no. of subscribers of class b in traffic area k; etc.

iii) Forecast parameters

Our aim is to forecast the future traffic interests between traffic areas, $A_{k}^{(T)}$. It is, of course, valuable from the planning point of view to have the possibility of separate forecasts for different kinds of traffic, e.g., data traffic on leased lines, business-to-business traffic, etc. But besides that, the final forecast is much more reliable if it is the aggregate of separate ones. Another point is that a forecast of total originating and terminating traffics, $A_{k.}^{(T)}$ and $A_{.}^{(T)}$ respectively, generally is more accurate than the point-to-point forecast, $A_{k}^{(T)}$. The ideal forecast is then the following:

i) Originating and terminating traffics per subscriber class and traffic area are forecast, $A_{bk.}^{(T)}$ and $A_{b.}^{(T)}$ respectively.

ii) These are aggregated, giving total originating and terminating traffics per traffic area, $A_{k.}^{(T)}$ and $A_{.}^{(T)}$ respectively.

iii) Independently of the total traffic forecasts, the point-to-point traffics between subscriber classes are forecast, $A_{bckl}^{(T)}$.

iv) These are aggregated, giving the point-to-point traffics for all subscribers, $A_{kl}^{(T)}$.

v) The originating and terminating traffic forecasts $A_{k.}^{(T)}$ and $A_{.l}^{(T)}$ respectively are accepted and thus distributed over the matrix, using the separate point-to-point forecast values $A_{kl}^{(T)}$ as distribution factors.

Consequently, we need traffic forecast parameters that can be checked against available traffic records, be adapted to future conditions, and can be used for the calculation of the desired traffic quantities, in combination with subscriber distribution data. Three such forecast parameters are central for the proposed procedure.

i) $a_{b.}$ = total originating traffic per subscriber line in subscriber class b. The property of this parameter is that it is relatively universal, i.e., there is little variation between different places of similar character and stage of development, and it is also fairly stable over time.

ii) d_{bc} = traffic dispersion factor. It shows how the originating traffic per subscriber of class b is spread over all classes. $\Sigma d_{bc} = 1$. The property of the parameter is a little 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.

iii) W_{bckl} = traffic interest weight. The parameter corresponds to the tendency of a subscriber in class b, in traffic area k to call a subscriber in class c because the latter is in area l. For example, a high class residential subscriber might show a clear tendency to call small shops provided that these shops are situated in the same area or in the city centre, but less tendency to call shops located further away or situated in a lower class residential district.This parameter is, of course, entirely of a 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.

b) Forecasting procedure

i) Calculations for the present point of time

The goal is to find realistic present values of the forecast parameters $a_{b.}^{(\ )}$, $d_{bc}^{(0)}$, and W_{bckl} for well defined traffic areas. The following procedure could be applied:

a) We collect the parts of the following data that are available:

$A_{0}^{*(0)}$ = Originating exchange traffic

$A_{t}^{*(0)}$ = Terminating exchange traffic

$A_{uv}^{*\ (0)}$ = Exchange-to-exchange traffic

b) We define subscriber classes and traffic areas, which implies that the following relation matrices should be prepared:

$n_{bk}^{(0)}$ = No. of class b subscribers in area k

$n_{ku}^{(0)}$ = No. of subscribers in traffic area k who are connected to exchange area u.

Because of the homogeneity principle applied to the choice of traffic areas, $n_{bu}^{(0)}$ can be derived from these relation matrices.

c) Since the point-to-point traffic in the matrix will be used as check values during the calculation of forecast parameter values, some kind of confidence intervals should be attached to them. The size of a confidence interval depends, of course, on how the particular exchange-to-exchange traffic value was derived. Cases where a great part or all of the traffic is routed via the local transit network may give rise to such uncertainty in the estimated point-to-point traffics that the value of such estimates is doubtful.

d) Now we determine the originating traffic per subscriber line in each subscriber class, $a_b^{(0)}$, in the following way: Solve the equation system:

$$\sum_b n_{bu}^{(0)} \cdot a_{b.}^{(0)} = A_{u.}^{*(0)} \qquad u = 1,2,....U$$

where U = no. of exchange areas.

If the total number of subscriber classes equals S, then we will get $\binom{U}{S}$ sets of solutions.

The assumption that the originating traffic per subscriber in a particular class is constant irrespective of the exchange area, cannot be absolutely true, of course, and since the "known" data $n_{bu}^{(0)}$ and $A_{u.}^{*(0)}$, furthermore, are more or less uncertain, some of the sets will look a little strange as they will also have extreme values, e.g., negative values and very high values. Fortunately, extremely low and extremely high values generally belong to the same sets. What we do is remove those sets from the total number. From the remaining acceptable sets, we calculate the most likely values of $a_{b.}^{(0)}$. There are several ways to do this, the simplest being to consider each class b separately and estimate $a_{b.}^{(0)}$ as the median of all accepted values. Another way is to apply the method of least squares to each class individually, or to consider all classes simultaneously. The flexibility of the so composed set may be increased by determining a confidence interval for each $a_{b.}^{(0)}$ value. Again, there are several possibilities. A statistically calculated 95% confidence interval could be used, <u>or</u> a fixed percentage around the chosen value, <u>or</u> ,perhaps, the whole range of values from the different sets.

e) We will need the terminating traffic per subscriber line in each class $a_{.c}^{(0)}$ as a check value when we determine the traffic dispersion factors, so we repeat the procedure as per d) above, but now solving the equation system

$$\sum_c n_{vc}^{(0)} \cdot a_{.c}^{(0)} = A_{.v}^{*(0)} \qquad v = 1,2,....U$$

Again, sets containing extreme values are rejected, and representative $a_{.c}^{(0)}$ values and confidence intervals are calculated from those remaining.

f) Now we come to the delicate problem of determining the traffic dispersion factors $d_{bc}^{(0)}$. The definition of d_{bc} is: the proportion of the originating traffic per subscriber line of class b that terminates among class c subscribers. Consequently, S d_{bc} = 1, and in the $d_{bc}^{(0)}$ matrix, we set the values row by row. To understand the idea, we imagine that Figure 6.57 is a picture shown on a visual data screen.

To guide us, our earlier determined $a_{b.}^{(0)}$ values are shown on the left of Figure 6.57. We should fill in the matrix on the top right, row by row, based on our experience and reasoning, and using local information. At the very bottom, the earlier determined $a_{.c}^{(0)}$ values with their confidence intervals are displayed.

When we have set all $d_{bc}^{(0)}$ values, our computer calculates the resulting values:

$$a'^{(0)}_{.c} = \frac{\sum_b n_b^{(0)} \cdot a_{b.}^{(0)} \cdot d_{bc}^{(0)}}{n_c^{(0)}}$$

appearing immediately below the $\mathrm{d}_{\mathrm{bc}}^{(0)}$ matrix.

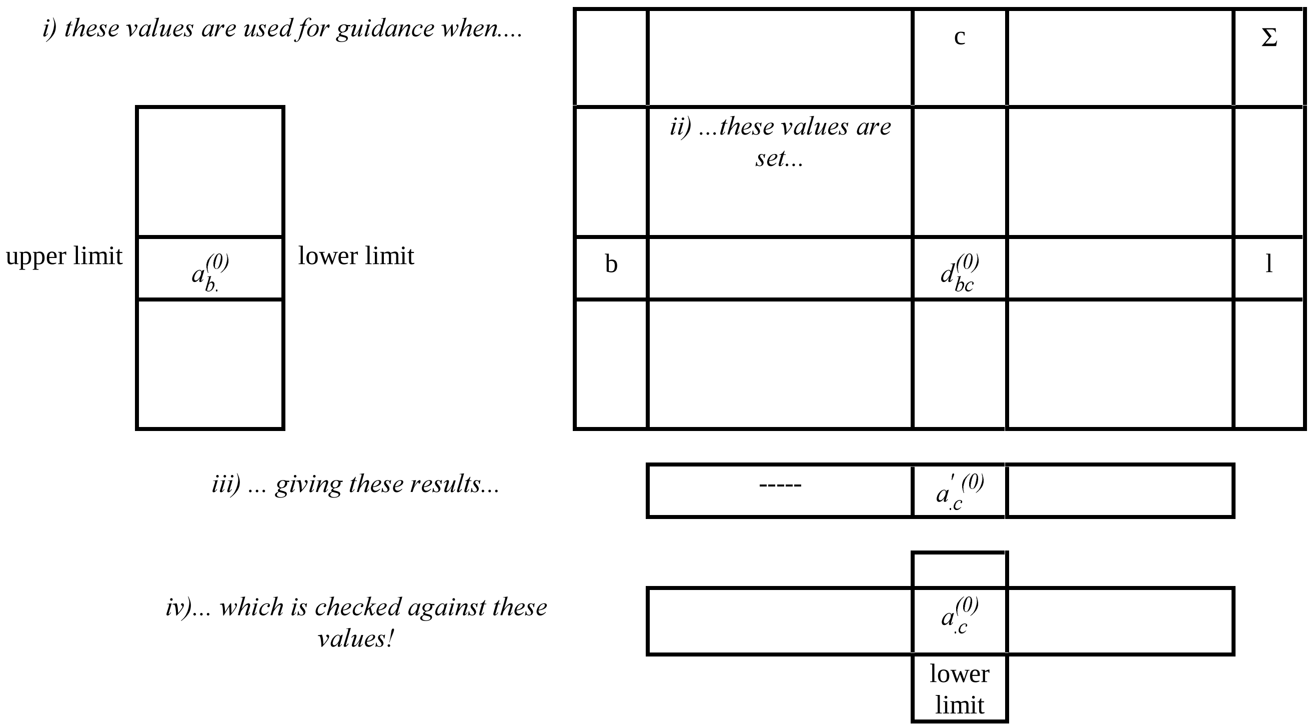


Figure 6.58 : Setting and checking of traffic distribution factors

The next step is to compare these resulting $a'^{(0)}_{.c}$ values with the check values $a_{.c}^{(0)}$ displayed further below, and then decide whether the observed differences can be accepted or not. If not, the $d_{bc}{}^{(0)}$ matrix is revised, which is quite simple because, for example, high $a'^{(0)}_{.c}$ values relate to high $d_{bc}^{(0)}$ factors, etc.

g) The traffic distribution weight W_{bc} is defined as a measure of the tendency of a class b subscriber in traffic area k to call a class c subscriber because that subscriber is in traffic area l. Therefore, each pair of b,c values can be treated separately in the process of setting the $W_{bckl}^{(0)}$ values. Furthermore, a very limited set of round values can be used, e.g., three values 1,2 or 3. In that case, 1="low", 2="normal", and 3="high". There may, of course, be reason to use a finer scale, viz., five values 1,2,3,4, or 5. In that case, 1="very low", 2="low", 3="normal", 4="high", and 5="very high".

Again, let us imagine that we are looking at the data display. If we set a pair of b,c values, a matrix filled with 3:s appears. The 3:s are default values, which will be used if we do not set other values.

h) All basic traffic parameters having now been determined, we can calculate:

$$A_{bckl}^{(0)} = \frac{a_{b.}^{(0)} \cdot d_{bc}^{(0)} \cdot n_{bk}^{(0)} \cdot n_{cl}^{(0)} \cdot W_{bclk}^{(0)}}{\sum_{l} n_{cl}^{(0)} \cdot W_{bclk}^{(0)}}$$

$$A_{kl}^{(0)} = \sum_{b} \sum_{c} A_{bckl}^{(0)}$$

and

$$A_{uv}^{(0)} = \sum_{k} \sum_{l} A_{kl}^{(0)} \cdot \frac{n_{ku}^{(0)} \cdot n_{lv}^{(0)}}{n_{k}^{(0)} \cdot n_{l}^{(0)}}$$

The values $A_{uv}^{(0)}$ can be checked against the known $A_{uv}^{*(0)}$ values (if any).

k \ l	1	2	3	4
1	3	3	3	3
2	3	3	3	3
3	3	3	3	3
4	3	3	3	3

k \ l	1	2	3	4
1	5	5	3	1
2	4	3	3	3
3	3	3	3	3
4	3	3	3	3

we change some values

Figure 6.59 : Setting of $W_{bckl}(0)$ - values

i) So far, we have estimated $a_{b.}^{(0)}$ as values common to all exchange areas. This assumption seems to be the only logical one since exchange areas are neither homogeneous from the socio-economic nor from the subscriber category point of view. What we actually need for planning, however, are forecast parameters related to <u>traffic</u> areas, and such areas will be chosen so as to be homogeneous. This means that it will usually be possible to identify the specific socio-economic character of each particular area.

This again raises the possibility of defining different $a_{b.}$ values for different areas, k, viz. $a_{b.}'$ which will greatly refine and improve our forecasts.

The estimation procedure will be as follows:

1. keep $W_{b}^{(0)}$ (no change);

2. keep $d_{bc}^{(0)}$ (no change);

3. set $a_{b.}'$ using the earlier obtained $a_{b}^{(0)}$ values for guidance;

4. Calculate hypothetical point-to-point traffics A' as:

$$A'^{(0)}_{uv} = \sum_k \sum_l A'^{(0)}_{kl} \cdot \frac{n^{(0)}_{ku} \cdot n^{(0)}_{lv}}{n^{(0)}_{k} \cdot n^{(0)}_{l}}$$

where

$$A'^{(0)}_{kl} = \sum_b \sum_c A'^{(0)}_{bckl}$$

and

$$A'^{(0)}_{bckl} = a^{(0)}_{b.} \cdot d^{(0)}_{bc} \cdot n^{(0)}_{bk} \cdot \frac{n^{(0)}_{cl.} \cdot W^{(0)}_{bckl}}{\sum_l n^{(0)}_{cl.} \cdot W^{(0)}_{bckl}}$$

5. Checking the hypothetical values $A'^{(0)}_{uv}$ against the already approved $A^{(0)}_{uv}$ values), adjust $A^{(0)}_{b.k.}$ until a reasonable agreement is obtained.

6. If the result is unsatisfactory, $W^{(0)}_{b\ k}$ and/or $d^{(0)}_{b}$ values can be revised.

ii) Calculations for the future point of time

The main object of this paper has been to present a procedure to find the present values of some traffic parameters that are important for a future traffic forecast. Therefore only brief comments will be made on how these parameters should be updated such that they continue to be relevant in a future situation.

The traffic profile for the hours of the day is often quite distorted in older networks. For example, the traffic ratio of the present busy hour to all hours might be, say, 1/12, while if the network were working at a good quality of service level, the value of the same ratio would be, say, 1/8. This indicates that the traffic parameters need to be revised. Depending on how much we know, or can reasonably guess, about the present conditions, this revision could be done in several ways. Two examples are given:

First example:

1. Adjust the $A^{(0)}_{uv}$ values individually for expected changes in the traffic profiles.
2. Calculate new $A^{(0)}_{u.}$ values.
3. Calculate new <u>sets</u> of $a^{(0)}_{b.}$, corresponding to the sets that were accepted previously.
4. Calculate the new $a^{(0)}_{b.}$ values but do not change $d^{(0)}_{b}$ or $W^{(0)}_{b\ k}$!

Second example:

Revise the $a^{(0)}_{b.}$ values directly. Then $a^{(0)}_{b.}$, $d^{(0)}_{b}$ and $W^{(0)}_{b\ k}$ should be updated to reflect the expected future conditions of development and subscriber characteristics, $a^{(T)}_{b.}$, $d^{(T)}_{b}$ and $W^{(T)}_{b\ k}$ respectively. Combining them with the subscriber forecast $n^{(T)}_{bk}$, results in the calculation of the future traffic interest matrices.

c) Conclusions

The traffic forecasting procedure presented here has the following properties:

1. Recorded data are used as far as reasonably possible for the calculation of forecast parameter values.
2. The forecaster's experience and local knowledge are used for setting hypothetical values of the remaining forecast parameters.
3. The hypothetical values are utilised to calculate quantities that can be checked against recorded data.
4. Where serious deviations occur, the forecast parameter values are revised by decisions made by the forecaster.
5. The calculations are based upon simple and replaceable algorithms, suitable for computer applications. Personal computers can very well be used.
6. It is a step-by-step procedure where judgement and decision-making are essential elements for each step. The procedure is thus best suited to interactive use.
7. The sensitivity of the forecast, based on variations of the basic parameter values, is therefore easily investigated.
8. The forecast is, likewise, easily updated when more traffic data has been collected.

6.4 Forecasting for rural networks

The population of a country is more or less concentrated in a great number of locations: large and medium sized cities, smaller towns and small to very small villages. At the same time, the country is administratively divided into areas, for example districts, each district being subdivided into smaller areas, for example communes. Thus the country may be described in terms of areas and locations.

The particular administrative division of any country has developed from historical, traditional, cultural and political factors as well as to some extent from current economical and social conditions.

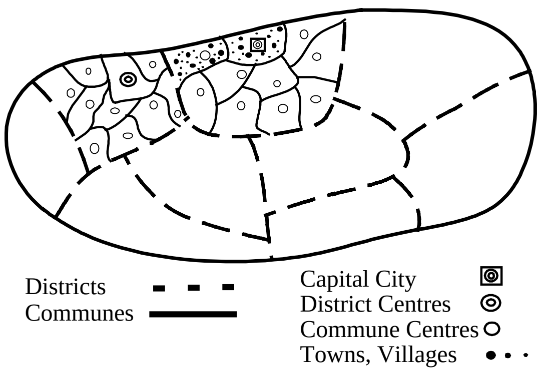
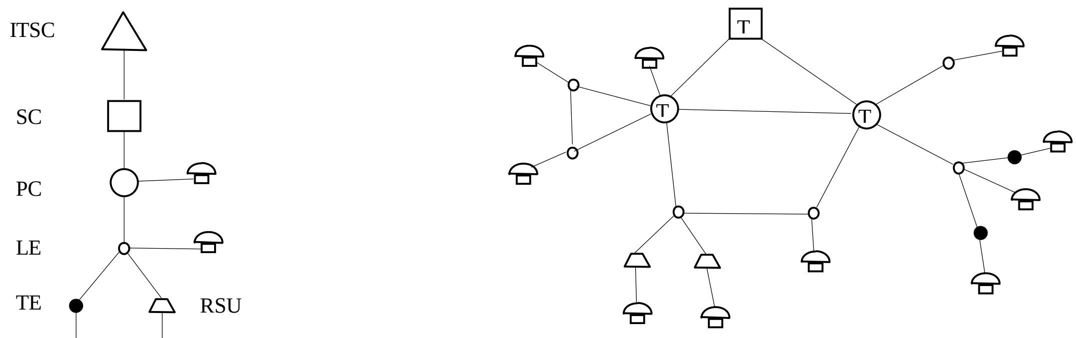


Figure 6.60: Typical administrative divisions

From the telecommunications point of view, the country is divided into service areas and organised in switching centres with an interconnecting network. This structure depends partly on the administrative division of the country and partly on technical and economical telecommunication considerations.

In principle, there is only one public network from the international switching centre(s) (ITSC) down to terminal exchanges (TE) or remote subscriber switching units (RSU), except that the local traffic in large and medium sized cities is usually served by multi-exchange networks.



National network hierarchy, local and LD-traffic

Multi-exchange network hierarchy, local traffic (Integrated or separated LD-traffic)

Figure 6.61: Network hierarchies

For the above reasons, switches on higher levels, for example secondary centres (SC) and primary centres (PC), are usually located in district centres and commune centres, local exchanges (LE) in towns and terminal exchanges (TE) and remote subscriber switches (RSU) in villages.

The future network(s) should be planned taking economical criteria much more into consideration, which means that **forecasting** should be carried out not so much considering the present switching and routing configuration, but rather concentrating on the real socio-economic and business structure.

From a practical planning point of view, the total network of a country is divided into somewhat more useable parts. First of all, the multi-exchange or metropolitan networks are planned separately. The rest of the total network is then divided into a higher level, constituting the national or long distance network as shown in Figure 6.62 and a lower level which comprises a number of rural networks. Such a rural network may for example correspond to a particular district, so it may serve a large number of populated places.

We see that each rural network may thus include for example a primary centre, a number of local exchanges and possibly a large number of remote subscriber units (RSU). The corresponding administrative structure may comprise a district centre, a number of commune centres and a large number of smaller towns and villages.

The forecast for each rural district should of course give the future number of subscribers in each town and village, possibly comprising hundreds of places.

The traffic forecast should in principle then consist of a matrix giving the traffic interests between all these locations. This is not practicable, and not even needed since there will clearly be no direct routes between most of these small places. On the higher level of the rural network, however, we may expect the need for alternative routing.

Thus, the traffic forecast should be organised in two parts:

On the higher level, involving the communes, a complete traffic interest matrix is needed;

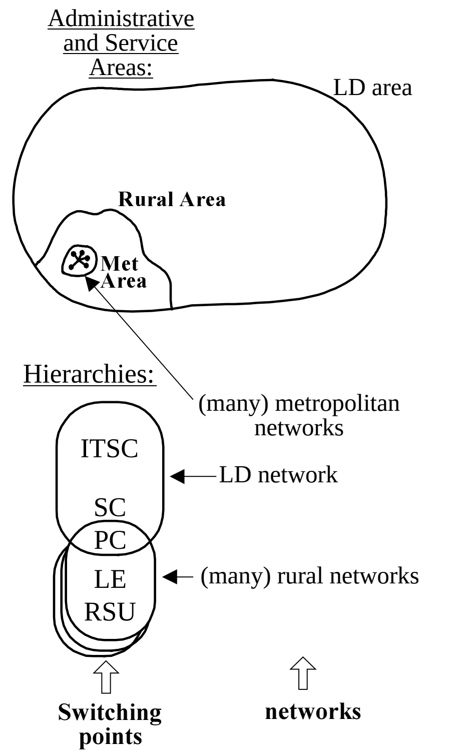


Figure 6.62 : Constituents of national network

On the lower level, comprising all the small towns and villages, a list, giving only the originating, the terminating and the internal traffic for each place will suffice.

			. . . c . . .	Σ
		A_{ij}		A_o
c			(I)	(II)
Σ		A_t	(II)	

Matrix : Traffic interests between communes

V	Sub	Calling rates			Traffics		
		TCR	PO	PI			
	(III)		(IV)			(V)	

List : Number of subscribers, Calling Rates and Traffics per village, town or city

Figure 6.63: Matrix and list for rural forecasting

We see that the **matrix** thus deals with the **areas** (**C** = Commune), while the **list** deals with the **places** (**v** = Village, town or city).

Six types of quantities are especially marked in Figure 6.63, viz.:

I = Point-to-point traffic between communes;

II = Total originating or terminating traffic per commune;

III = Number of subscribers per village, town or city;

IV = Calling rates TCR, Proportion of originating traffic PO, and Proportion of internal traffic PI per village, town or city;

V = Originating traffic $\mathbf{A_o}$, Terminating traffic $\mathbf{A_t}$ and Internal traffic $\mathbf{A_i}$ per village, town or city.

Prime (') means preliminary value.

The forecast can be carried out in two steps as shown in Figure 6.64:

Step 1 is to find calling rates TCR, PO and PI for the present time (t=0);

Step 2 is to forecast future calling rates, number of subscribers and preliminary point-to-point traffic between communes, and then to find the final traffics by reconciliation.

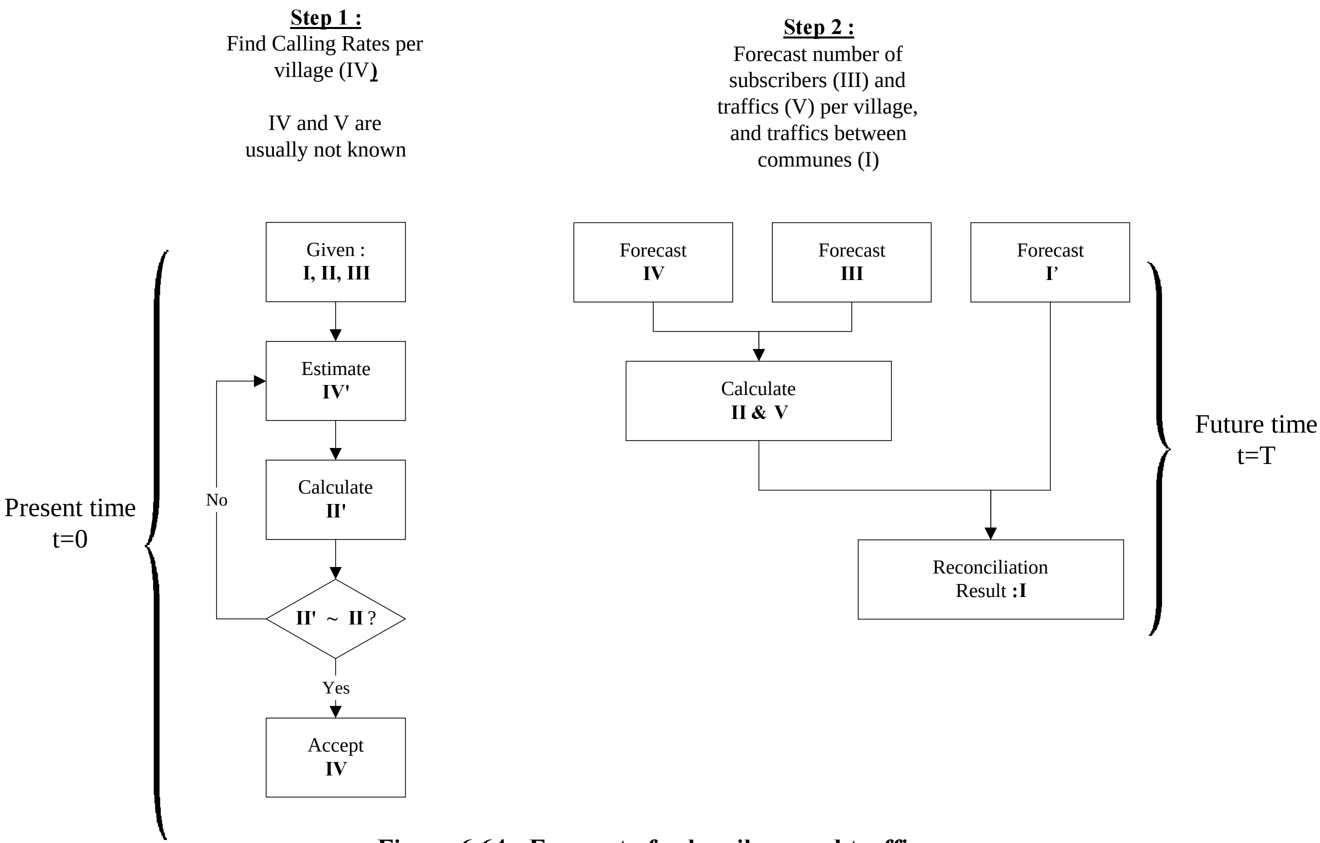


Figure 6.64 : Forecast of subscribers and traffics

The difficulty in this case is to forecast the future number of subscribers (III) and the calling rates (IV) for all these villages and towns. This subsection concentrates therefore on these problems in an attempt to make the forecast process both accurate and fast, once that it is well established, even if mathematical modelling and the necessary computer programming may need some initial efforts. The procedures involved are given in the following sections.

6.4.1 Forecasting procedures

The following symbols are used in this procedure:

SLEPT= Village category variable, composed of sub-variables S,L,E,P,T

S = Size S = 0 : -100 (population)
1 : 100-500
2 : 500-3000
3 : 3000-10000
4 : 10000-

L = Level L = 0 : (almost no functions)
1 : Basic, for example, police, fire, nurse, elementary school
2 : Many functions for own day-to-day needs
3 : Administrative and commercial centre for other towns (for example, for the whole community)
4 : Administrative and commercial centre for large area (for example, for the whole district)

E = Socio-Economic type E = 0 : Basic, for example, Agriculture, fishing ...
1 : Mixed, for example, Basic (i.e., as E=0 above)+ small industries and businesses
2 : Mixed, (i.e., as E=1 above) + large industries, businesses, administration

P = Private economic level P= 0 : Poor
1 : Average
2 : High

T = Population T = 0 : Rapidly decreasing development
1 : Slowly decreasing trend
2 : Constant
3 : Slowly increasing
4 : Rapidly increasing

POP = Population
D = Density, corresponding to total demand
DMAX= Density saturation limit for villages, corresponding to total demand
C = Future number of connected subscriber lines
PC = Proportion satisfied demand (connected lines / total demand)
TCR = Total Calling Rate (traffic per subscriber line)
PO = Proportion Originating traffic, out of total
PI = Proportion Internal traffic, out of total
v = Village
c = commune
d = district
n = country
t = point of time (0 = present year)

Note: “Density” should correspond to “Total Demand” whether satisfied or not. A properly maintained waiting list may thus be included. This goes for past, present and future points of time. For a certain village **v** this means:

Number of connected main lines $C_v^{(t)} = POP_v^{(t)} \cdot D_v^{(t)} \cdot PC_v^{(t)}$

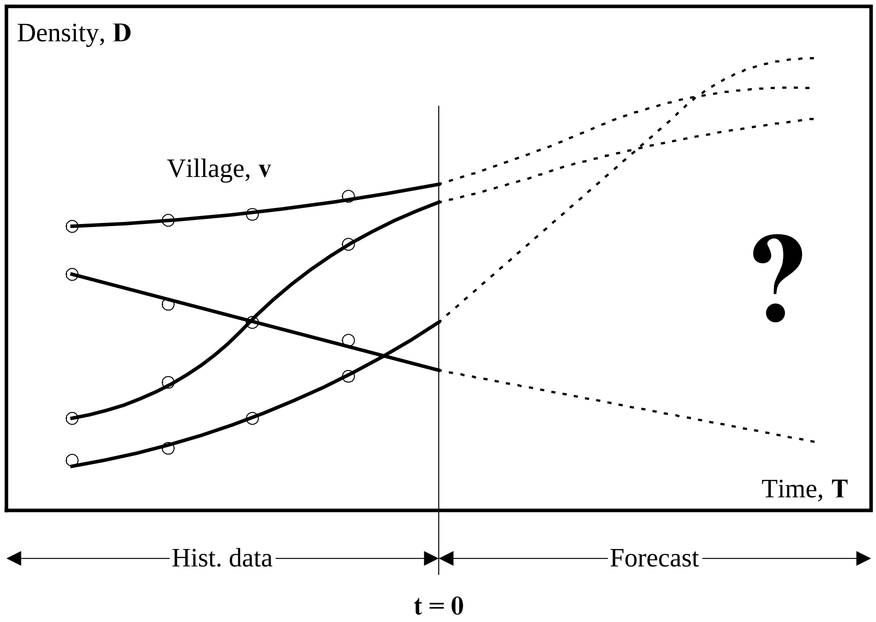
Total originating traffic $= C_v^{(t)} \cdot TCR_v^{(t)}$

For each rural area (for example, called District) which is to be planned, forecasts of the following are required:

(a) the traffic interests between all sub-areas (for example, called Communes);

(b) the long-distance traffic from and to each commune;

(c) the number of subscribers (main lines) and the originating, terminating and internal traffic for each village and town in the rural area. For the purposes of this subsection, which concentrates on item c), "villages" covers both types.



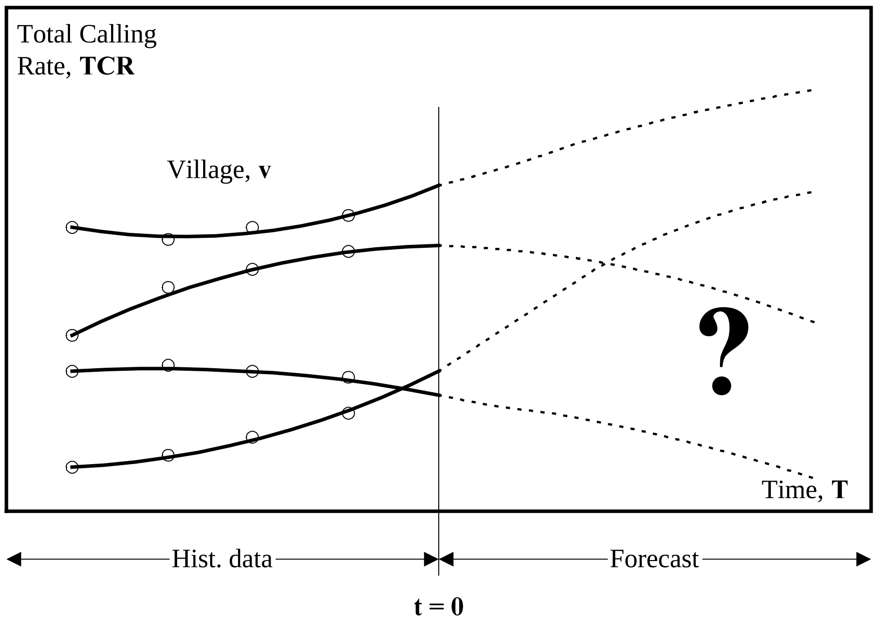


Figure 6.65*:* **Forecast problems (Forecast of PO and PI not shown***)*

Typically, there are about 10 to 20 communes in such a district, while there may be several hundred villages, and country-wide, there may be thousands of villages, but typically numbering about 10,000 to 50,000.

Furthermore, there are great differences between villages as regards population size, administrative and cultural level, socio-economic type, private economic level, and population development trend.

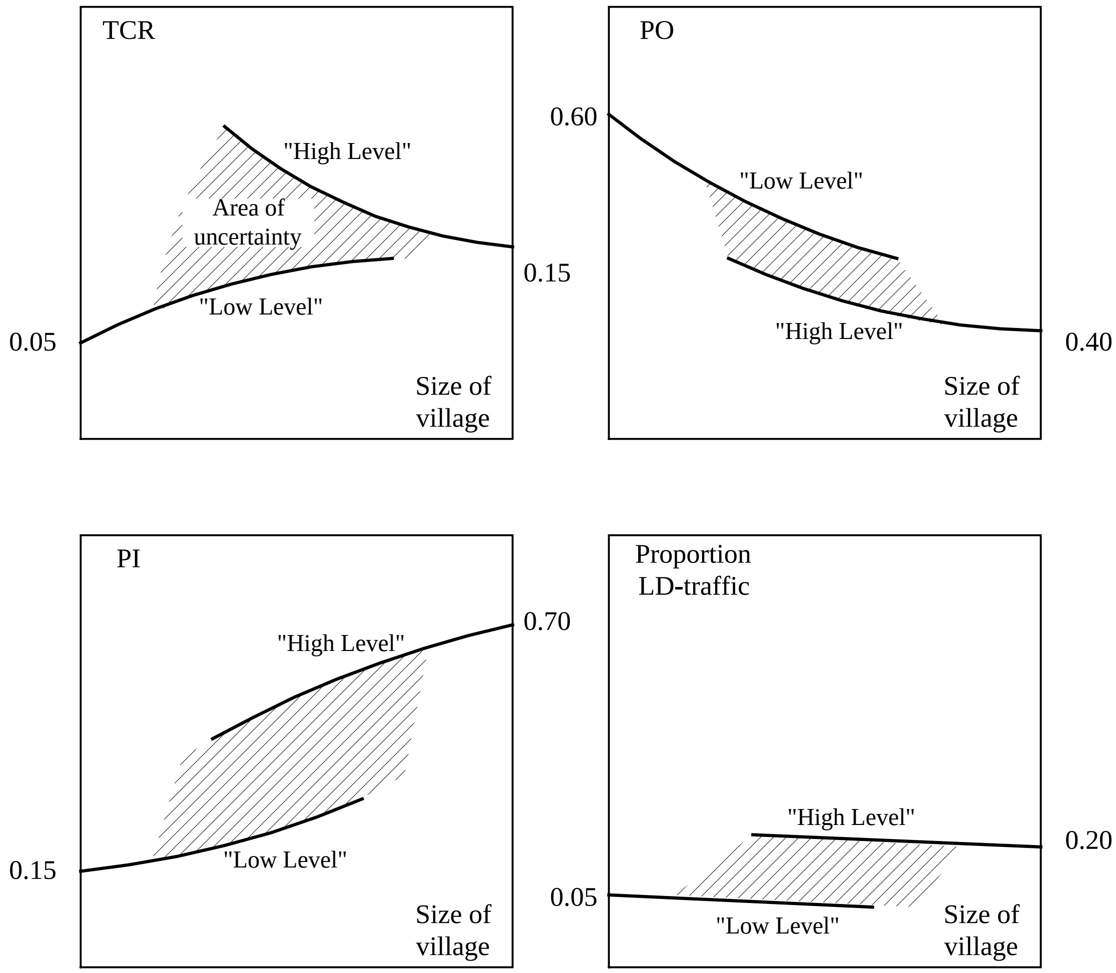


Figure 6.66 : Examples of how traffic characteristics for villages in a particular area could be described, using only two parameters - Size of a village, and, "Level" of a village. "Level" would then be the substitute for several parameters (compare with L, E, P and T). As a result of this rather rough way of describing the villages, areas of uncertainty would be quite large.

Consequently, there is a dilemma here; because of the differences between the villages, particular attention should be given to village forecasting, but, because of the large number of villages, the village forecast procedure has to be standardised and computerised as much as possible, applying some kind of mathematical model.

One class of model which can be used to forecast the development of the main line density for such a diversified mass of entities is the growth curve, for example, the exponential logistic model. Growth curve models have a number of useful properties. One such property is that they function correctly both if the supply of historical data is meagre or, on the contrary, if there are extensive data sets. Another welcome feature is that, when well used, growth curves can be used in a dynamic bottom-up/top-down process provided that a separately estimated saturation limit is superimposed on each curve, The basic growth curve function is in itself trend extrapolating forwards on the time axis (=bottom-up), while the superimposed saturation limit works in the reverse direction (=top-down).

The saturation limit should NOT, consequently, be a RESULT of the application of the growth curve model - that would change the process from this stable and reliable bottom-up/top-down method into a rather poor, uncontrollable, and unstable method.

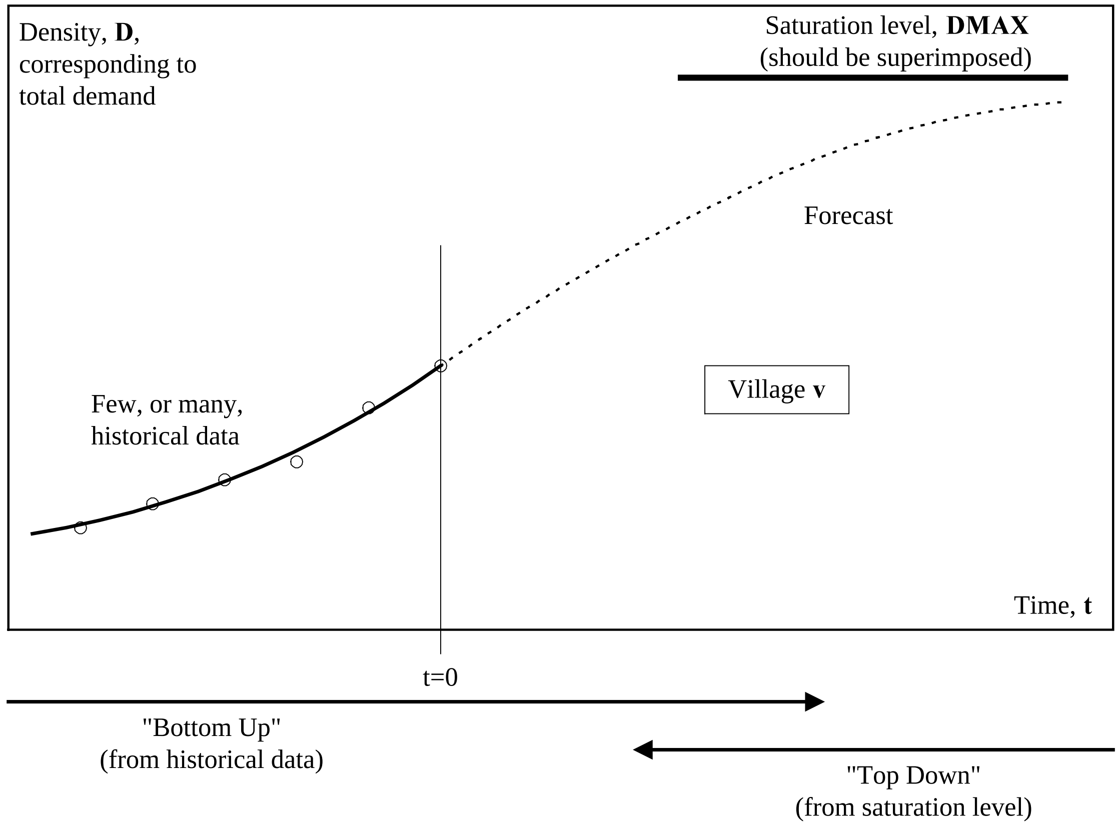


Figure 6.67: Use of growth curves

We can usually predict the development of the population or the number of households quite well. What we therefore need besides calling rate forecasts, is a forecast of density or penetration for each village.

Since we assume that there are at least some historical data available on densities per village, $D_v^{(t\geq 0)}$, we will need only the future density saturation limit for the same villages, $DMAX_v$, to estimate the future densities, $D_v^{>}$, which will merely be a matter of straight forward numerical calculation according to the growth curve model.

In the case of $DMAX$, since the saturation limit is usually (although not necessarily) meant as an asymptotic limit for growth curves, thus being defined for unlimited time, it should in principle remain constant for all points of time **t**, i.e. **t** would not be a parameter. $DMAX_v$ would thus be the appropriate notation. However, since any particular village may change character over time, leading to a changed development trend, we need to modify $DMAX$ also. Therefore, the variable is defined as $DMAX_v^{(t)}$, thus becoming an instrument not only to calculate one particular trend curve for each different village, but also to control trend changes.

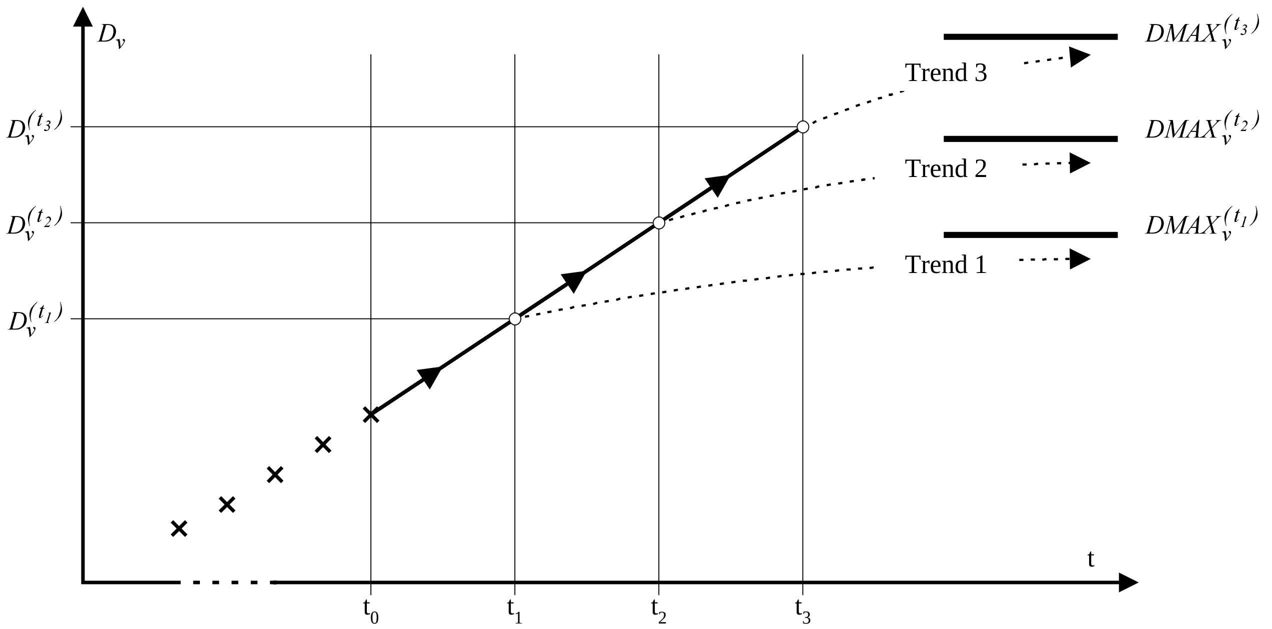


Figure 6.68: As an example, this figure aims to illustrate how the forecast for a particular village might be done, utilising three different trend lines, to be used for example in case of two trend shifts.

The way of modifying $DMAX_v^{(t)}$ for different points of time **t** is to change $SLEPT_v^{(t)}$ and all associated parameters.

In a situation where trend changes for the whole area **x** can be foreseen, the average saturation level $DMAX_x^{(t)}$ may be used as the instrument to modify all village forecasts simultaneously (and collectively).

These two instruments, $DMAX_v^{(t)}$ and $DMAX_x^{(t)}$, may be used in combination (producing a sort of development trend for the whole area but allowing deviations for a few individual villages).

The average future saturation limits $DMAX_x^{(t)}$ for a whole area **x** should be defined and used as an input to the algorithms which are intended to estimate the village saturation limits $DMAX_v^{(t)}$. Preferably, the variable should be defined for the whole rural area, i.e. the district (**x**=**d**). If defined for each commune (**x**=**c**) or for the whole country (**x**=**n**), the algorithms have to be modified.

As already mentioned, the final goal is to forecast the future number of connected lines and the originating, terminating, and internal traffic, for each village.

Once future population development is estimated, and since traffic quantities are calculated as the product of calling rates and connected main lines, we need only to forecast:

- future village density saturation limit $DMAX_v^{(t)}$;
- future calling rate $TCR_v^{(t)}$;
- future proportion of originating traffic $PO_v^{(t)}$;
- future proportion of internal traffic $PI_v^{(t)}$.

This can be done by using all the relevant historical data on densities and calling rates but, in addition, the known facts on the character status and future prospects of each village should be used.

Each village in a district can be well enough described and the descriptions can be digitalised by transferring them into the form of a numerical village category variable, $SLEPT_v^{(t)}$.

Computer algorithms to estimate the required forecast variables from given village information can then be constructed.

Input :

$D_v^{(t)}$; $TCR_v^{(t)}$; $PO_v^{(t)}$; $PI_v^{(t)}$
$(t \le 0)$

$DMAX^{(t)}$; $TCR_x^{(t)}$; $PO_x^{(t)}$; $PI_x^{(t)}$
$(t \ge 0$; $x = d, c$ or $n)$

$SLEPT_v^{(t)}$
$(t \ge 0)$

→ Computerised Algorithms →

Output :

$D_v^{(t)}$

$TCR_v^{(t)}$

$PO_v^{(t)}$

$PI_v^{(t)}$

Figure 6.69 : Estimation of forecast variables

Two immediate tasks have to be carried out:

- to analyse the logical relationships between various significant village characteristics and required output from the process illustrated above;

- to construct the algorithms which are to operate on $SLEPT_v^{(t)}$ and on $DMAX^{(t)}$, and, of course, to define all other variables that are to be used in these algorithms.

It is obvious that behind the demand for telecommunication services, there must be some driving forces. Depending on the means available, this demand might be satisfied by the provision of new main lines. Whether the new subscribers will use the newly provided services extensively or not, may also be a matter of the available means to satisfy the demand (one example of such means could be household economy). Thus the driving forces which could create a demand and the means that can help to satisfy this demand must be examined. This is a complicated procedure as not only are the factors to be examined interdependent rather than independent, but they are extremely numerous, and many are, in practice, next to impossible to quantify.

However, if some generalisations about a village, such as: the village is “relatively large”; its administrative, commercial and cultural level is “high”; the socio-economic level is “pretty advanced”; the people living there are, in general, at a “rather good” economic level, and that the village seems to have a “good chance” to grow larger “rapidly”, then a number of assumptions are available which can be used in a model -which will generate probable future scenarios upon which decisions on the telecommunication infrastructure planning can be based.

If the future telecommunication development in the village were to be based upon these verbal generalisations, it could be said that “fairly large” density growth leading to “a high” future density could be expected and that the usage of the provided services would also grow to “a relatively large figure”.

What is probably meant is that this village is quite a dynamic one, with many social and business activities, even to some extent dominating the surrounding area; in other words, it is believed that both the driving forces behind the telecommunication demand and the means to satisfy the demand are quite substantial.

It could now be appropriate to apply the concept of village categorisation by defining numerical values for all the parameters $S_v^{(t)}, L_v^{(t)}, E_v^{(t)}, P_v^{(t)}$ and $T_v^{(t)}$ for each village **v** and for each point of time $t \ge 0$. The settings for **t=0** will be used to check the model parameters, thus giving us a good basis for the forecasting. The overall density saturation limit $DMAX^{(t)}$ should also be defined, probably for the whole district, i.e. $DMAX_d^{(t)}$.

Neglecting for the moment the necessary checks and corresponding adjustments and iterations, the flowchart to determine $DMAX_v^{(t)}$, $TCR_v^{(t)}$, $PO_v^{(t)}$ and $PI_v^{(t)}$ may now examined in more detail and may be divided into three consecutive sub-flowcharts as shown in Figure 6.70.

These sub-flowcharts are used as follows:

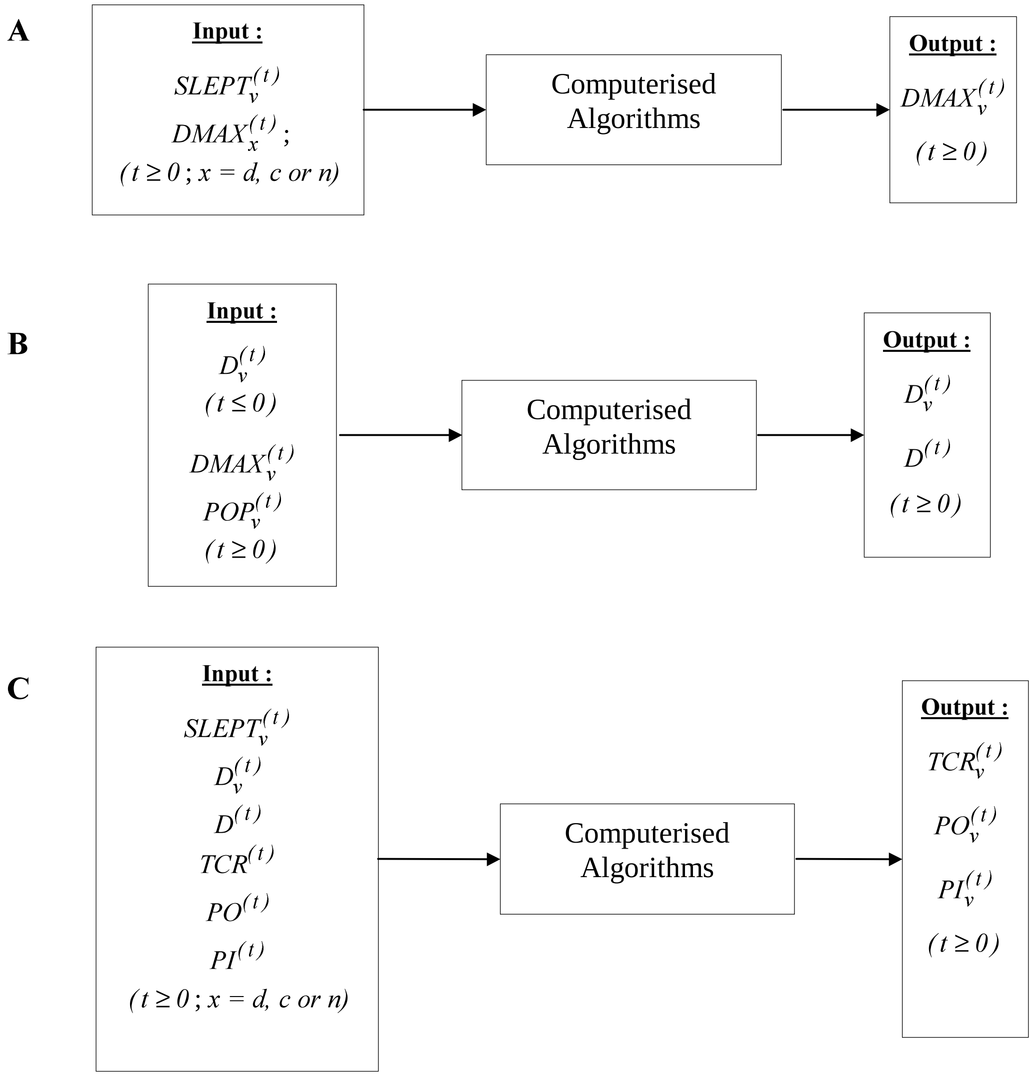


Figure 6.70 : Sub-Flowcharts:

<u>Sub-flowchart A :To calculate saturation levels per village</u>

Let $DMAX_x^{(t)}$ be the average saturation level in the area, for example the district (**x=d**). If a particular village **v** now is an absolutely "average" one from all significant aspects, then we should expect that the saturation level to be used for this village should be the same as the average level, i.e. that $DMAX_v^{(t)} = DMAX_x^{(t)}$.

If on the other hand the village differs significantly from the "average" village in a way reflected by the parameters $SLEPT_v^{(t)}$, the specific saturation level $DMAX_v^{(t)}$ using factors based on $SLEPT_v^{(t)}$ has to be calculated.

A simple expression for such an estimation could be in the form:

$$DMAX_v^{(t)} = DMAX_x^{(t)} \cdot FDS_v^{(t)} \cdot FDL_v^{(t)} \cdot FDE_v^{(t)} \cdot FDP_v^{(t)} \cdot FDT_v^{(t)}$$

where **FDS** = **F**actor for estimation of maximum density based on size of the village;

FDL = **F**actor for estimation of maximum density based on level of the village; etc.

In the case of an "average" village, all factors **FDS**, **FDL**,... should be = **1**, otherwise not.

Sub-flowchart B:

<u>To calculate future densities per village</u>

This is where the main growth curve algorithm is applied, e.g. the exponential logistic model. Two sets of data are required:

a) Past and present densities per village = Historical data;
b) Future saturation levels per village = Output of sub-flowchart **A**.

The calculation is a matter of fitting a curve to historical data under the constraint that the **imposed** saturation level must be the asymptote to the extrapolation of the curve.

From the resulting future densities per village $D_v^{(t)}$ the mean future densities per commune or for the whole district can be calculated.

Sub-flowchart C:

<u>To calculate future calling rates per village</u>

$TCR_x^{(t)}$, $PO_x^{(t)}$, and $PI_x^{(t)}$ are respectively the average calling rate, the average proportion of originating traffic, and the average proportion of internal traffic in the area, for example the district (**x**=**d**).

Again, if the particular village **v** is absolutely "average", it should be expected that these average values for the whole area would also apply for the village considered; if not, the specific calling rates , $PO_v^{(t)}$, and $PI_v^{(t)}$ using factors based on $SLEPT_v^{(t)}$ have to be calculated.

A set of simple expressions to determine the individual village parameters could be as follows:

$$TCR_v^{(t)} = TCR_x^{(t)} \cdot FTS_v^{(t)} \cdot FTL_v^{(t)} \cdot FTE_v^{(t)} \cdot FTP_v^{(t)} \cdot FTT_v^{(t)} \cdot FTD_v^{(t)};$$

$$PO_v^{(t)} = PO_x^{(t)} \cdot FOS_v^{(t)} \cdot FOL_v^{(t)} \cdot FOE_v^{(t)} \cdot FOP_v^{(t)} \cdot FOT_v^{(t)} \cdot FOD_v^{(t)};$$

$$PI_v^{(t)} = PI_x^{(t)} \cdot FIS_v^{(t)} \cdot FIL_v^{(t)} \cdot FIE_v^{(t)} \cdot FIP_v^{(t)} \cdot FIT_v^{(t)} \cdot FID_v^{(t)}.$$

where:

FTS = **F**actor for estimation of total calling rate based on size of the village;

FOS = **F**actor for estimation of proportion originating traffic based on size of the village;

FIS = **F**actor for estimation of proportion internal traffic based on size of the village;

FTD = **F**actor for estimation of total calling rate based on density in the village;

etc.

In the case of an "average" village, all factors **FTS**, **FTL** ... **FOS**, **FOL** ... **FIS**, **FIL** ... should be = **1**, otherwise not.

Note that $TCR_x^{(t)}$, $PO_x^{(t)}$ and $PI_x^{(t)}$ are input values to the algorithms. However, after having run the algorithms, thus obtaining $TCR_v^{(t)}$, $PO_v^{(t)}$ and $PI_v^{(t)}$ as an output, new values $TCR_x^{(t)}$, $PO_x^{(t)}$ and $PI_x^{(t)}$ can be calculated from this output. Thus, the whole calculation may be iterative.

From SLEPT to FDS

It is seen that the calculation of future densities and future calling rates involves the use of a number of "factors" $FDS_v^{(t)}$, $FDL_v^{(t)}$,...etc., which are based on $SLEPT_v^{(t)}$. Having first defined numerical values of $SLEPT_v^{(t)}$ $(t \geq 0)$, the problem will consequently be to translate these data into "factors".

It is first of all necessary to study how the required parameters **TCR**, **PO**, **PI** and **DMAX** are influenced by **S**, **L**, **E**, **P**, **T** and **D**. Considering **TCR**, for example; possibly this parameter is most strongly affected by the level of the village, i.e. parameter **L**. Perhaps the next strongest influence comes from the socio-**e**conomic type, i.e. parameter **E**, etc.

If this is represented by figures, so that "**1**" means the strongest influence, "**2**" means the next to strongest influence, etc., these influences can be shown in the form of a table as in Figure 6.71 below:

Influence on / by	TCR	PO	PI	DMAX
S	5	2	1	1
L	1	1	2	2
E	2	3	3	3
P	3	5	4	4
T	6	4	5	5
D	4	6	6	not applicable

Figure 6.71 : General influence on **TCR**, **PO**, **PI** and **DMAX** by **S**, **L**, **E**, **P**, **T** and **D**.
"1" represents strongest influence, "**2**" next-to strongest, etc. Note: Given figures are just <u>examples</u>

These influences can be illustrated graphically, but still in a qualitative manner, i.e. not introducing quantities (actual numerical values) as shown in the series of diagrams given in Figures 72 and 73.

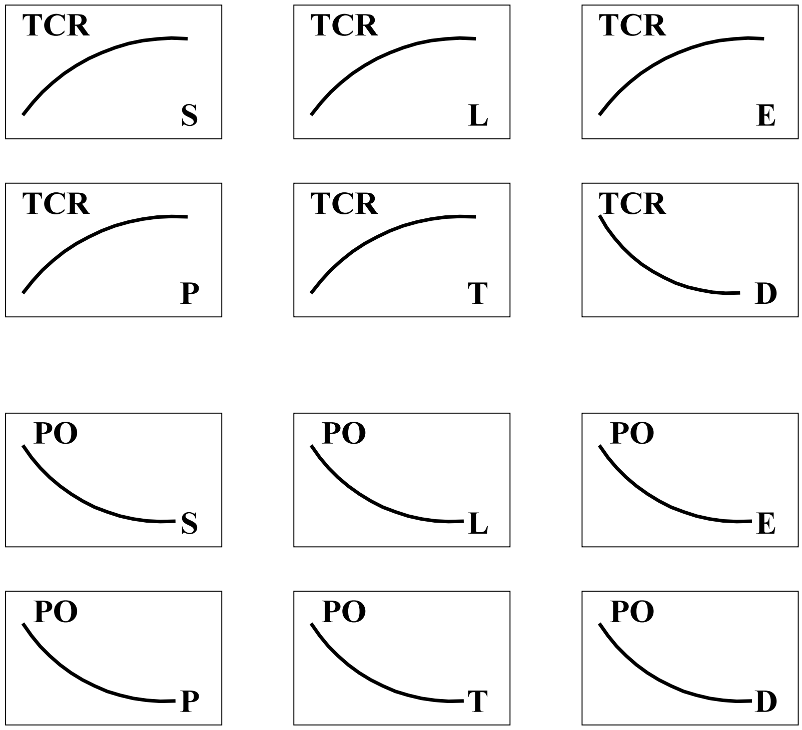


Figure 6.72: General influence on *TCR* and *PO* by *S,L,E,P,T* and *D*, expressed graphically.

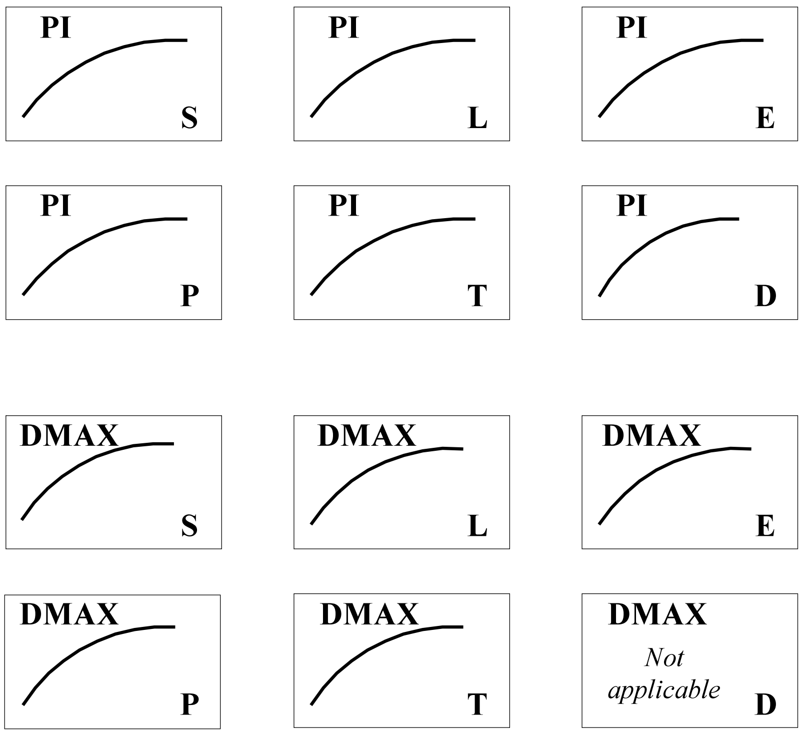


Figure 6.73: General influence on *PI* and *DMAX* by *S,L,E,P,T* and *D*, expressed graphically.

Inspecting all these curves, it is seen that they may generally be characterised by two parameters:

B = Inclination of the curve;
Z = Shape of the curve

Concerning the **inclination B** of a curve it is seen that a curve may be nearly horizontal (weak influence), positively steeper (larger, positive influence), or negatively steeper (larger, negative influence).

Thus **B** could be used to characterise the inclination as shown in Figure 6.74 below:

B = **Steepness** of curve, indicating degree of influence;

|B| = **1** represents largest inclination;

|B| = **2** represents next-to largest inclination; etc.

(**|B|** = **absolute value** of **B**)

\+ represents **positive** inclination;

- represents **negative** inclination.

{B}=

Influence on / by	TCR	PO	PI	DMAX
S	+3	-1	+1	+1
L	+1	-2	+2	+2
E	+1	-2	+3	+3
P	+2	-3	+3	+3
T	+3	-3	+3	+3
D	-2	-3	+3	not applicable

Figure 6.74: Influence of B (example only)

Examining the other property - the shape of a curve, **Z**. it is seen that all curves are more or less convex, others concave, and bent to different degrees, from slightly bent to quite strongly bent. Each particular curve has not a constant bend, but the bend decreases with increasing value of the influencing parameter.

It is also seen that all curves which have positive inclination are bent alike, for example in a concave manner, and that all curves having negative inclination are bent the other way - i.e. in a convex manner.

The factor **Z** can be used to characterise the shape of curves as shown in Figure 6.75 below:

Z = Shape of curve;
Z = **1** represents less bent curve;
Z = **2** represents medium bent curve;
Z = **3** represents more bent curve.

$\{Z\}=$

Influence on / by	TCR	PO	PI	DMAX
S	2	3	2	2
L	3	2	2	2
E	2	1	2	1
P	1	1	1	1
T	1	1	1	1
D	2	2	1	not applicable

Figure 6.75: Influence of Z (example only)

Any curve may then be calculated as

$$FYQ = \frac{|B| \cdot CB + \left(\sqrt[Z+CZ]{Q} - \sqrt[Z+CZ]{\overline{Q}} \right)}{|B| \cdot CB}$$

where

Y = letter **D**, **T**, **O** or **I** (representing **DMAX**, **TCR**, **PO** or **PI**);

Q = **S**, **L**, **E**, **P**, **T** or **D**;
(**Q** is interpreted as a letter in the name "**FYQ**", otherwise as a variable)

$\overline{Q}$ = average value of **S**, **L**, **E** ,**P**, **T** or **D**;

B = curve steepness parameter;

Z = curve shape parameter;

CB = scale constant for **B**;

CZ = scale constant for **Z**;

TB = **+1** for positive **B**-values, **1** for negative **B**-values;

|B| = absolute value of **B**.

Interpretation of $\overline{Q}$:

$\overline{S} = 2;\ \ \overline{L} = 2;\ \ \overline{E} = 1;\ \ \overline{P} = 1;\ \ \overline{T} = 1;$

$$\overline{D} = \frac{\sum_V POP_V \cdot D_V}{\sum_V POP_V}$$

<u>Interpretation of scale constants</u>:

CB and **CZ** may be used to manipulate a whole range of curves simultaneously.

Order of magnitude: $CB \approx 20;\ \ CZ \approx 0.5$

Note: A smaller value of $|B|\cdot CB$ indicates a steeper curve;
A smaller value of $Z{+}CZ$ indicates a less bent curve.
(**Z+CZ=1** will produce a straight line)

The effect of $|B|\cdot CB$ and $(Z + CZ)$ on FYQ is shown in Figure 6.76.

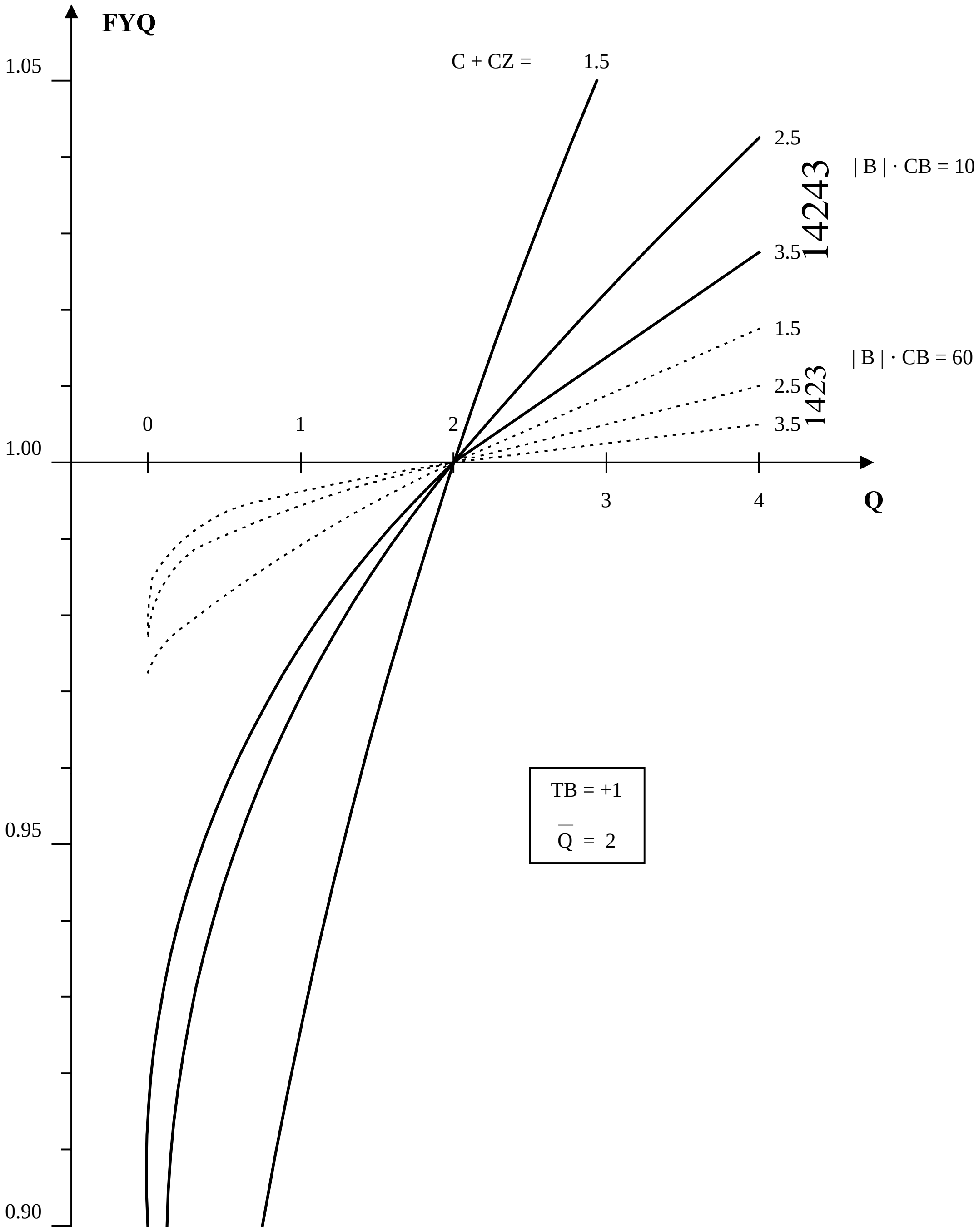


Figure 6.76: Demonstration of the effect on *FYQ* of the two quantities $|B|\cdot CB$ and *Z+CZ*, for different Q-values.

The algorithm can now be applied, a logical flow chart, as shown in Figure 6.77 for example, being followed.

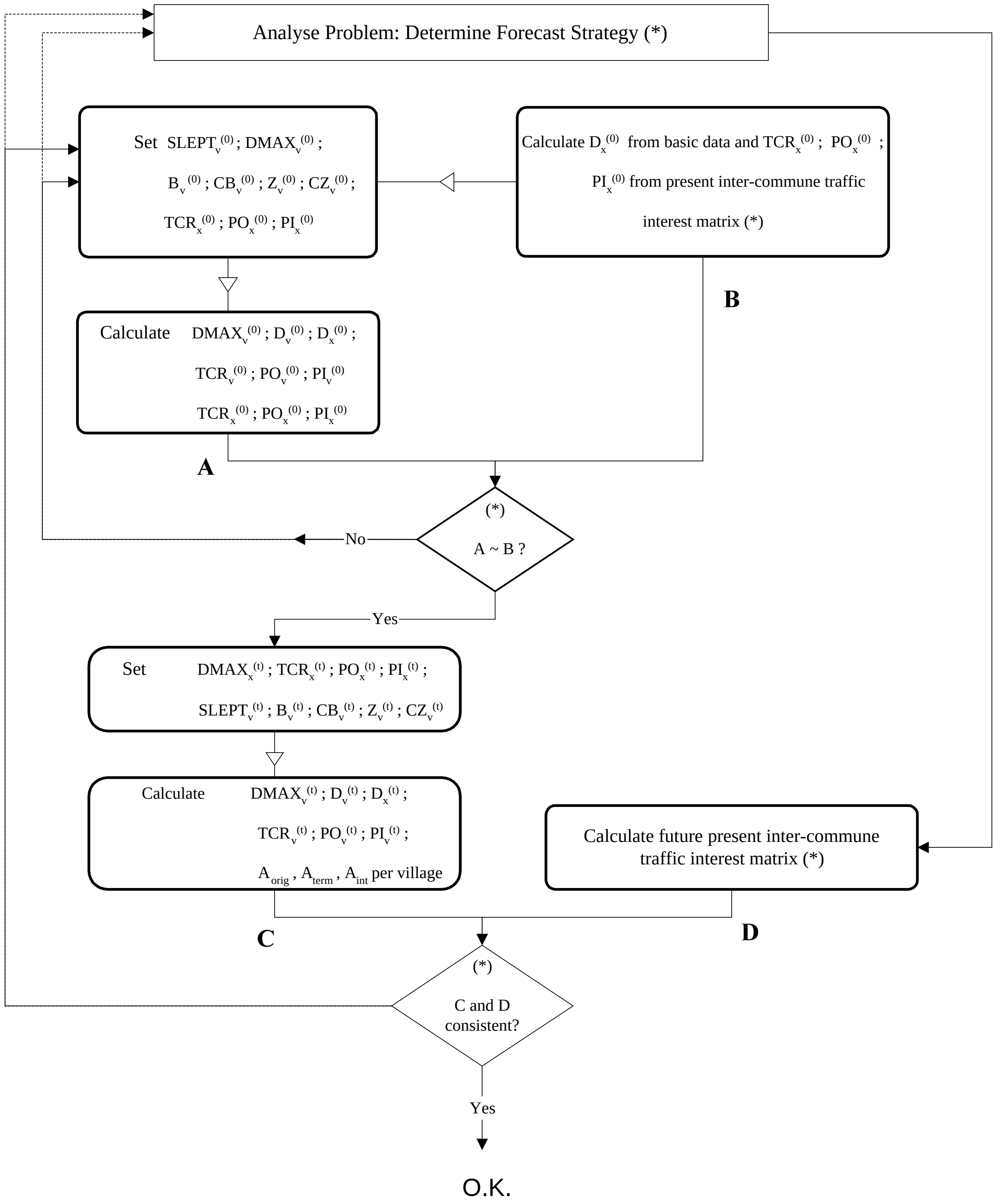


Figure 6.77: Forecast scheme for rural networks.

* The main problem analysis, calculations on the inter-commune traffic matrix, and decisions and comparisons between village and inter-commune traffic is not treated in this subsection.

When this flowchart is examined it is seen that some parameters appear both as input and output, and that sometimes a comparison or check is done between them. The explanation is as follows:

Basic data are generally aggregate data. It is not possible to disaggregate these data in a mathematically unambiguous way.

The calculations carried out are in the contrary quite detailed (disaggregated) and are based on the combined use of basic, aggregate data, and detailed, hypothetical parameter values.

If, after applying the algorithms, the results are aggregated, sets of data which are compatible with the basic data could be obtained, possibly with similar numerical values. If they are not, the situation has to be re-considered. It is possible that the differences could be acceptable and possibly not.

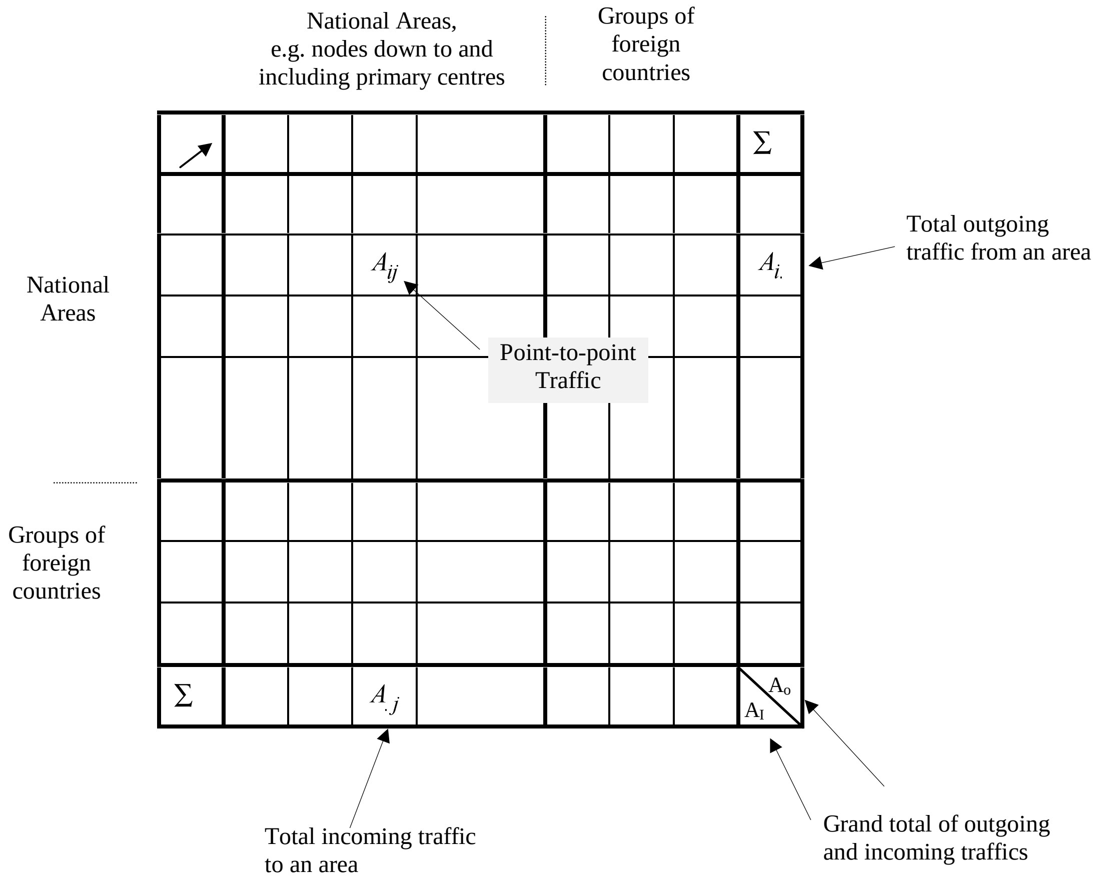
6.5 Traffic forecasting for national networks

The part of the network in a country which includes the nodes from the international switching centres (ITSC) down to and including the primary centres is usually called the National Network or the Long Distance (LD)Network.

The planning of this network involves to a great extent the choice of transmission systems and the optimisation and dimensioning of trunk routes. Another item is the optimisation of routing of all traffic including both national and international traffic, which becomes especially interesting if there is more than one ITSC in the country. The hierarchy of the network is also very important: should for example some of the primary centres (PC) be upgraded to secondary centres (SC) or perhaps the reverse? If so, the internal structure of such primary areas may be re-organised and the question is, how?

We see from this, that the traffic forecasts should go one step below the lowest level in the LD network, and that outgoing and incoming international traffic for each district in the country should be included. The forecast should be presented as a traffic matrix between all areas, and could thus become enormously large.

It is however not necessary to specify all individual foreign countries as areas. It will usually do to specify a few **groups** of countries. For example: say that the studied LD network is in a country in Central America. Possibly then group no. 1 would be the whole of North America, group no. 2 the other countries in Central America, group no. 3 South America, group no. 4 Spain, France and Great Britain, group no. 5 the rest of Europe, and group no. 6 the rest of the world. For another studied LD network the grouping of foreign countries would be different. It all depends on cultural, social and commercial communication between countries, and of course to some extent on where the foreign existing international switching centres are located.



It is furthermore not necessary to include the nodes on the lowest level in the matrix. For these nodes it will do to define originating, terminating and internal traffics in form of a long list as indicated below:

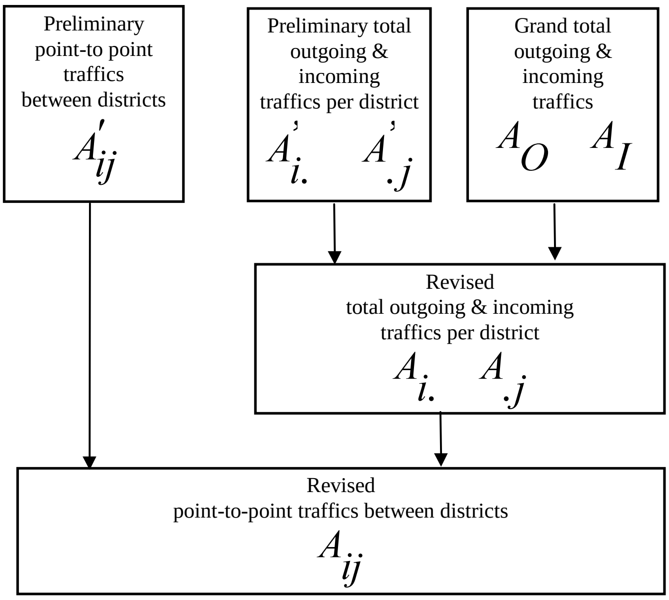
Areas	**Traffic**		
(All nodes)	**Originating**	**Terminating**	**Internal**

Of these two, the matrix and the list, the matrix of traffics is of dominating importance. This matrix comprises three types of quantities: The point-to-point traffics between areas, the total outgoing and incoming area traffics, and the grand total of outgoing and incoming traffics. Then there are two types of sub-quantities: national traffics and international traffics.

Now, two general principles should be followed:

1 Each one of these 3x2=6 types of quantities should be forecast **separately**, each one preferably by at least two different methods. That would give 6x2=12 forecasts. The use of two different methods should however be used together with a following reconciliation, and national and international traffics are written in different parts of the matrix, so after these operations we have the three different types to manipulate.

2 Now we should follow the top down principle, i.e. the grand total of outgoing and incoming traffics should be used as the envelope for the total outgoing and incoming area traffics, which will therefore be adjusted accordingly; and these adjusted area traffics should again be used as the envelope for the point-to-point traffics.



In principle, this process, which is a top down one, could be further improved by combining it with a bottom up approach as well. In practice, however, it might be difficult to produce so many separate forecasts due to possible lack of data.

In any case, there are many possible single methods which may be used for each one of these separate forecasts. No “best” algorithm for national traffic forecasting can be given, since conditions and availability of data vary so much between countries.

6.6 Aspects of forecasting for regional networks

A regional network may serve from very few up to quite a number of countries. Typical problems to be discussed between the countries are:

- How should the traffic interests between the countries be determined and how could 24 hour traffic profiles be defined and applied?
- How should responsibility for forecasting be shared and how should reconciliation between forecasts by different countries be done?
- How should traffic routing between the countries be determined, how should the countries share the transit traffic and who should provide transit traffic resources?
- What tariffs should be applied and how should revenues be shared?
- Could all countries agree to a common optimisation of the network in order to minimise the total network cost and thereby maximise the total common revenue?
- If so, could recommendations from the optimisation study be obeyed by all countries?

Let us first take a look at the particular region as a part of the world as shown in Figure 6.78.

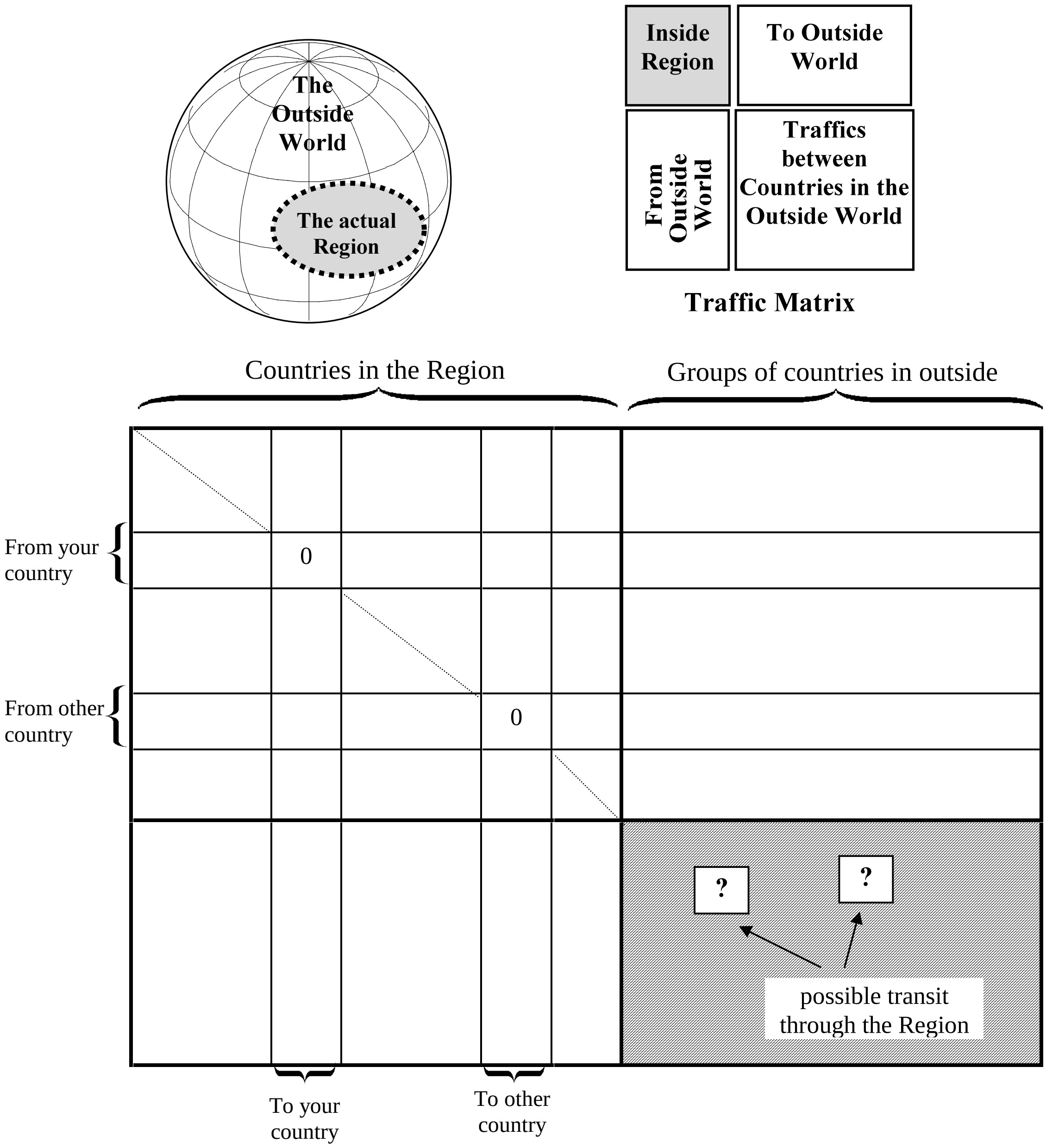


Figure 6.78

We see that the total traffic matrix contains four different main parts.

- One part represents traffics between the countries in the particular region.
- Next part shows traffics from each country to the outside world, i.e. countries outside the region.
- The third part are incoming traffics from the outside world to each country in the region. All these three quantities are important for the planning of the region.
- The fourth part, traffics between countries in the outside world, is hardly interesting, unless some part of it passes as transit through the studied region.

As for national forecasting, the countries in the outside world should be arranged in a few groups.

Figure 6.79 shows the traffic matrix again, but now parameter notations are added and partial and total sums of traffics are included.

↗		j		$\sum_j$		l		$\sum_l$	$\sum_j \sum_l$
i		A_{ij}		$\sum_j A_{ij}$		A_{il}		$\sum_l A_{il}$	$\sum_j \sum_l A_{ij,l}$
$\sum_i$		$\sum_i A_{ij}$		$\sum_i \sum_j$				$\sum_i \sum_l$	$\sum_i \sum_j \sum_l$
k		A_{kj}							
$\sum_k$		$\sum_k A_{kj}$		$\sum \sum_k$					
$\sum \sum_k$		$\sum_i \sum_k A_{i,kj}$		$\sum_i \sum_j \sum_k$					

i, j = country within REGION (or ITSC)
k, l = group of countries outside REGION

Figure 6.79

The best thing would of course be if all countries in the region could agree on a common optimisation and planning of the network. That would increase the available common revenue. See Figure 6.80, Alt. b): Overall Network Planning.

It is however very common for the countries to act otherwise, i.e. by a large number of bilateral discussions and agreements. This process is tedious, time consuming and expensive, and leads to higher network costs, i.e. to reduced common revenue. See Alt. a): The Sub-Planning Principle.

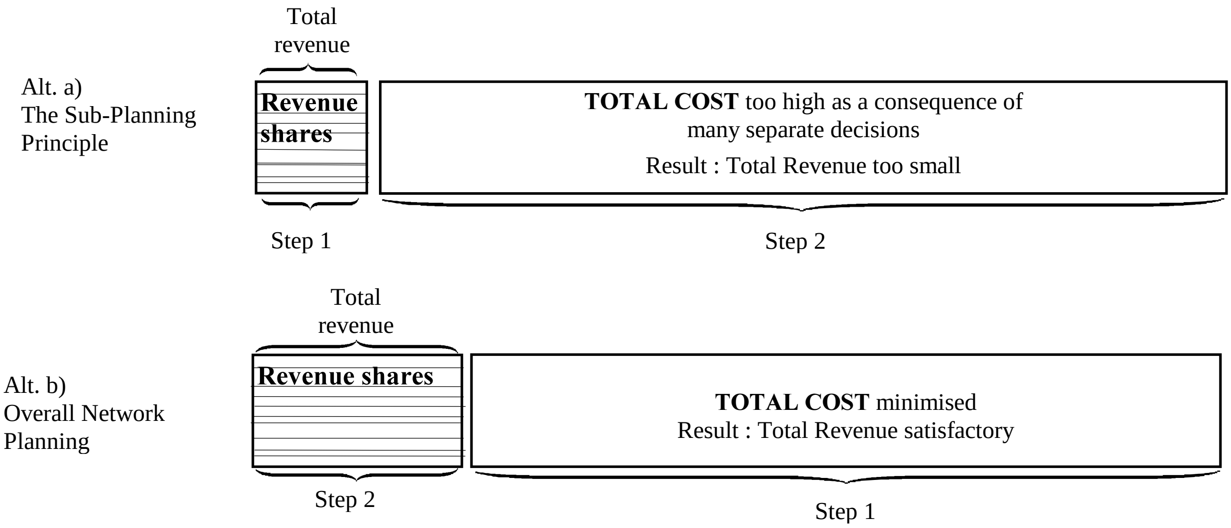


Figure 6.80

Provided that the countries could agree to common planning, different strategies for traffic data provisioning could be used:

Traffic Quantities (Country x):

Category I:

$$\sum_j \sum_l A_{ij,l} \qquad \sum_l A_{il} \qquad \sum_j A_{ij} \qquad \sum_k A_{kj} \qquad A_{ij} \qquad A_i \qquad A_{kj}$$

Category II:

$$\sum_i \sum_k A_{i,kj} \qquad \sum_i A_{ij} \qquad A_j$$

Strategies:

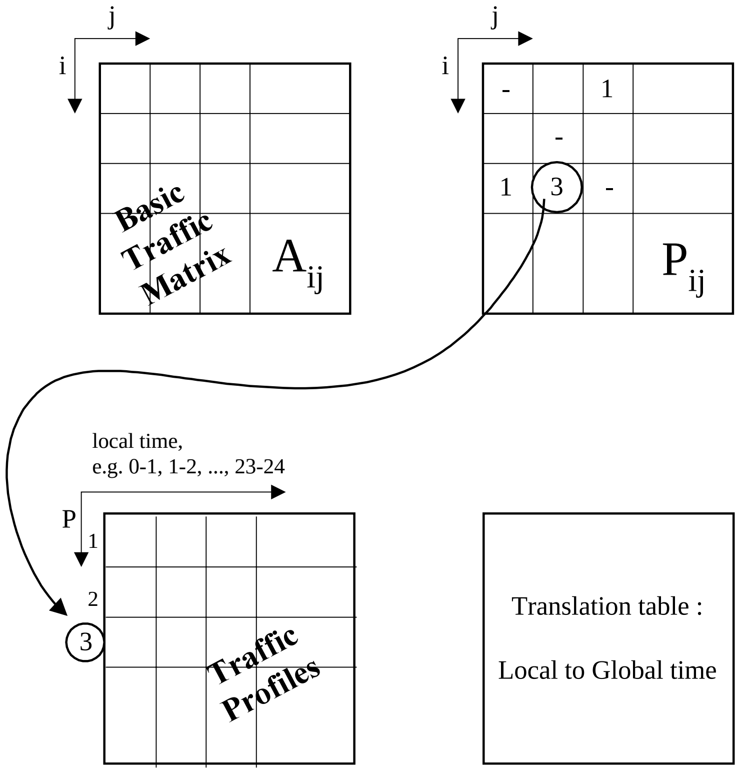
1 Each country (x) forecasts quantities from category **I** only
= no conflict, but country x has no control over terminating traffics from the region.

2 Each country (x) agrees with each other country (z) on just **two** traffic values, viz., A_{XZ} and A_{ZX}.

3 Each country forecasts **all** values above (both category **I** and **II**).
Reconciliation must take place since there will be **two** proposed values for each element. The reconciliation algorithm must be agreed upon from the beginning.

etc.

The problem of the application of traffic profiles could be solved as shown in Figure 6.81.



If A_{ij} = BH - values, then the profile =	0...	1.00	0...	-----	$\Sigma > 1$

If A_{ij} = 24-hour traffic volume, then the profile =	0...	0...	0...	-----	$\Sigma = 1$

Figure 6.81 : Application of traffic profiles

First of all a matrix of basic traffic quantities between all countries is defined (A_{ij}).

The next matrix gives a profile number for each pair of countries (P_{ij}). Several pairs may have the same profile number.

What we need then is a table defining each possible profile in the form of a vector with 24 values, one for each hour (local time). The vectors could be defined in two different ways dependent on how the A_{ij}-values are defined. If A_{ij} are busy hour values for example, then each profile vector contains one value = 1.0, corresponding to the busy hour traffic. All other values are of course ≤ 1 .

If, on the other hand, A_{ij} are 24-hour traffic volumes, then the **sum** of the profile vector is = 1.0, so that a conversion table between local and global time is all that remains to be provided.

6.7 A subscriber forecasting scheme for small countries

6.7.1 Introduction

Some countries may have quite a small population and consequently also a telecommunications network with few subscribers and a simple hierarchical structure. A few cities and towns usually hold the major part of the subscribers. Outside these towns, there are much fewer lines.

Usually, traffic forecasting has not been considered at all, and subscriber forecasting is done simply by visiting all studied areas, towns and villages in the country, walking the streets, inspecting the status and judging the likely future of each estate and house. A simple classification system may be used to indicate whether a house has a telephone line, is judged to get one “for certain” or is just “likely” to get one, for each of the nearest years. To increase the accuracy of the forecast, some interviewing of local administrative and other people may take place. Two main subscriber categories may be identified: residential users, and business and official users. Forecasting of coinbox lines (payphones) may be done separately.

These micro forecasts are then aggregated up to cabinet areas and exchange areas. A result of such studies is of course that the forecasting staff has a unique and very useful knowledge about all service areas in the country. A weakness of this purely bottom-up procedure is that forecasting errors are also aggregated so that the accuracy of the cabinet and exchange area forecasts depends entirely on how well the forecasters can judge on the micro level. Say that a forecaster has a positive general belief in the future. Then when choosing between zero and one line he will often choose **one**, instead of 5 he will maybe believe in 10 lines etc. On the micro level itself these absolute errors are small, but when aggregated they may cause wrong and very costly planning mistakes.

In the future, proper forecasting becomes more important: the network will grow and may need updating, new services will be added, etc. The problem to be solved is how to improve the forecasting for the future network. What is certain is that to retain a single forecasting method, trying to refine that method more and more, will hardly pay much in the form of improved accuracy.

6.7.2 Improving the forecasts

Forecasts can be improved by adopting the following techniques:

- Combine different methods;
- Work bottom up/top down;
- Combine subjective judgement with mathematical modelling;
- Use all relevant and reliable data;
- Work interactively in a forecasting system, i.e. the system may run various algorithms automatically, but the forecaster is to make decisions dependent on intermediate results and to manipulate the forecast parameters;
- Revise the forecast, parameters and even models continuously, based on follow-up studies.

A new forecasting scheme may start in a very simple manner, but may, if well run and maintained, improve steadily, since the forecasting activity will in itself produce new data in the form of observations of the forecasting performance. At the same time the forecasting staff will be more and more experienced and skilled.

The scheme proposed here is rather a simple one. It may in the beginning be run manually, but a computerisation would offer great advantages. Forecasts could be carried out and be updated very quickly, many scenarios could be tried, the sensitivity of the forecasts to various stimuli could be investigated, and the forecasters’ skill would improve dramatically. The amount of errors introduced by manual intervention in intermediate phases of the forecast procedure would also be diminished.

The scheme works on the following levels:

N= National level;

C&D= City and District level. If cities inside a district are treated separately (as Metropolitan Areas), then D=rest of the district;

E= Exchange area;

CAB= Cabinet area or RSU area;

DP= Distribution Point or the corresponding level.

The following subscriber categories are used:

RES= Residential main lines. To start with, the classifications according to the subscribers' applications are accepted. In future, an attempt to estimate the amount of lines reported as residential but used as business could be introduced;

BUS= PBX lines + single business lines (reported as such) + official lines. In future, it may be good to split into two categories, viz., PBX lines and single line business;

CB= Coin Box lines;

SS= Special Service lines.

The business lines may be split into two sub-classes to further increase the forecast quality:

BUSSPEC= Reliable estimations of business lines in specified locations, corresponding to well planned new enterprises or new branches of companies, etc.;

BUSOTH= All **other** estimated Business lines, i.e. BUS = BUSSPEC + BUSOTH

The forecasting procedure for business lines from the C&D level and down would therefore be:

1 BUSSPEC is estimated separately;

2 BUSOTH = BUS - BUSSPEC;

3 The Bottom Up/Top Down process is run on BUSOTH;

4 BUSSPEC IS ADDED BACK TO FINAL VALUES OF BUSOTH.

In the case of special service lines it may be preferable to exclude them from the general, integrated forecasting scheme, at least in the beginning. Some aspects of forecasting for coinbox lines are given in the subsection on Metropolitan forecasting of non-residential lines. In the beginning, it may be preferable to work out a separate forecasting procedure for CB, and later integrate it in the general scheme.

The total scheme for residential and business forecasts shown in Figure 6.82 is sub-divided into the following main parts or sub-processes:

I Line forecasting on the national level (N) with distribution into classes RES and BUS on N-level and on C&D-level;

II Estimation of forecast parameters POP (Population), HH (Households) and EMP (Employees per Economic Sector) on N-level with distribution on C&D-level;

(II) interacts with (I) both on N-level and on C&D-level!

III Bottom up/top down forecast procedure on levels C&D, D, E, CAB and DP:

The top down process distributes the RES and BUS forecasts on C&D-level from (I) + (II) over the E, CAB and DP levels according to Bottom Up aggregates.

There will be a tendency that top down values override the bottom up values to some extent, but not completely. It is rather a reconciliation that must take place on all levels. The process is interactive and thus moves up and down until the forecasters feel that strong arguments have had a reasonable influence, while weak arguments have had a less influence both on the distribution and on the possible adjustment of top down as well as on bottom up values.

IV Follow up process. This process is of equal importance compared to other sub-processes. It consists of comparing the real outcome with the previous forecast values. The comparison should be done on all levels and should affect the next forecast.

Process:

1) Observation + aggregation to all levels;

2) Comparison with forecasts;

3) Explaining differences;

4) Decision.

If a deviation from the forecast development curve is observed, there are three possible main decisions:

a) A jump + recovery (gradual return to curve);

b) A jump + return to trend (the whole curve is moved);

c) A trend shift (new curve).

Comments on the forecast process items (a) - (g). See Figure 6.82.

(a) Forecast of the total number of lines on N-level. Statistical demand analysis should be used. Explaining variables could be GDB per capita and POP. For the short term horizon (up to 5 years) a complementary model would simplify trend extrapolation.

(b) Forecast of POP, HH and EMP per economic sector, nation-wide. This is a forecast that should be done by the government. Probably, only figures for a certain year can be obtained, for example for the next 10-year period. If so, the forecasting group must try and complement these figures with other data, for example from international comparisons.

(c) Distribution of POP, HH and EMP per sector (result of (b)) over the C&D level. This is best done by applying a method that considers total and individual trends. See for example GAS 10/Ch. 6: The Pantograph Method.

(d) Forecast of RES and BUS on the national level. RES is calculated as population or HH multiplied by the density D, which is thus the variable to be forecasted. A simple and much used method is the use of the exponential logistic model. Input should be at least the two last historical values, and the estimated density saturation limit.

BUS is calculated individually for each economic sector by multiplying EMP in that sector by lines/EMP. The lines per EMP for a particular sector for a particular year is obtained from linear interpolation between two points - the present rate and the estimated far future rate.

International statistics show that these rates (lines per EMP per sector) are very stable.

(e) Same as (d), but on the C&D level the POP, HH and EMP data are the result of (c).

(f) Distribution on the N level of total lines to RES and BUS lines is done using the result of (d) as distribution factors.

(g) Distribution of RES and BUS lines from the N level to the C&D level is done using the result of (e) as distribution factors.

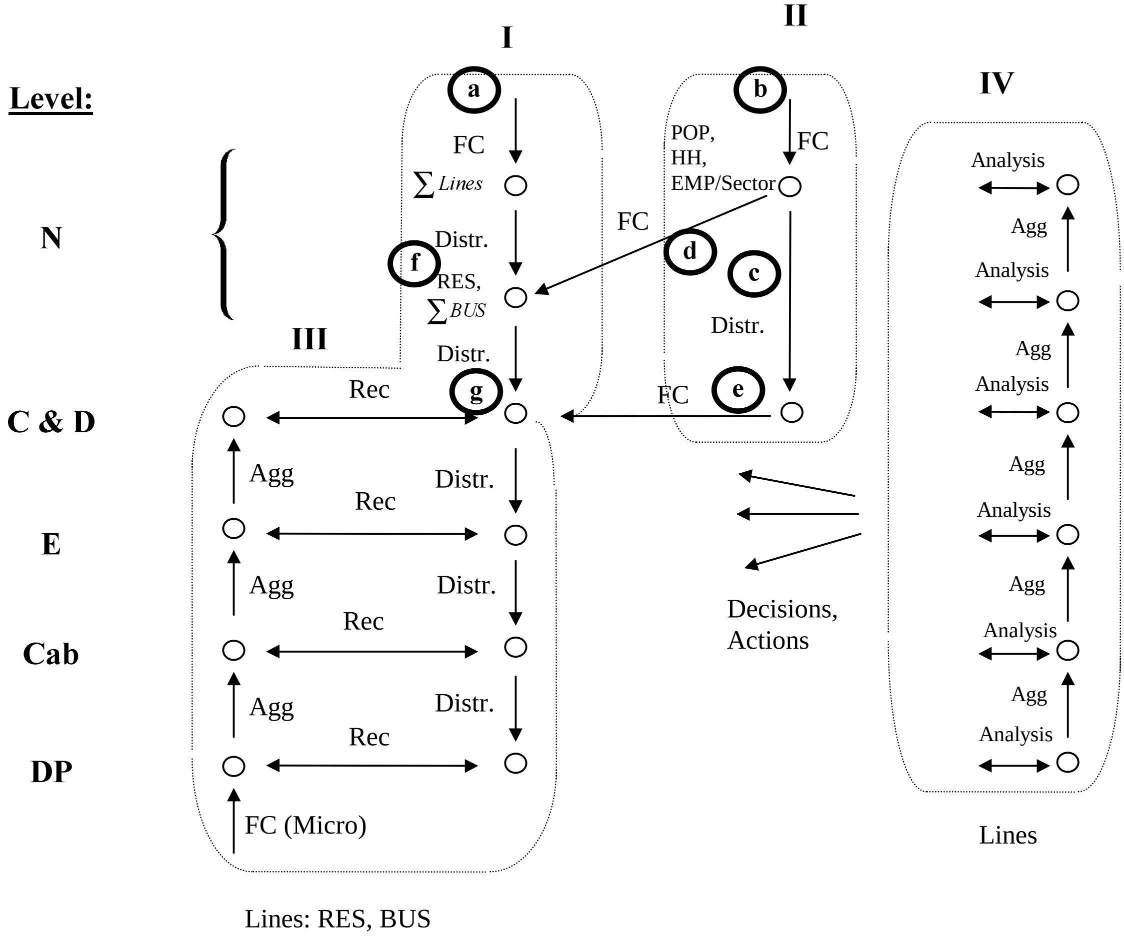


Figure 6.82 : Overall subscriber forecast process and sub-processes I, II, III and IV