VII. Urban Modelling

Abhandlungen - Anthropogeographie
Institut für Geographische Wissenschaften, FU Berlin
Band 52, Berlin 1994

DUVA - A CONCEPT FOR A METADATA DRIVEN STATISTICAL PRODUCTION AND INFORMATION SYSTEM

Norbert Kopp
Statistical Office of Berlin,Germany

The DUVA-concept is an integrated base, which contains the whole process control for statistics production from the survey about plausibility, storing, confidentiality, editing and dissemination.

Every step in the process of production will be standardized independently of a specific statistic. With that it is possible to realize in the production independent modules for every step. About definition of the different interfaces, this concept can be developed step by step and in a division of labour.

For storing data there are two models. Our input interface to the system are standardized basic files, which can contain micro data or aggregated data. The second model are macro files with precompressed aggregated data. Because both data models are standardized it is possible to use a macro data generator.

All needed descriptions for basic and macro files like structure of records, variables, categories of attributes are stored in a metadata information system. All available information in that system will be used in an automatic way for further steps in the process of production.

Transformations in the statistical production process are realized on the basic of rules. Therefore the metadata information system knows different types of rules (reference tables, algorithms, lists of conditions, hierarchies) which can be used in an isolated or combined way.

For classifications the DUVA-concept has an own key concept. It guarantees, that it will be possible to combine horizontal or vertical different hierarchical levels from one or more classifications. About a system of reference tables you can combine them always in the same way.

User interface for the DUVA-system is a thesaurus. About that thesaurus the user can communicate with the system in his natural language. In the thesaurus there is no redundancy. The problem of homonyms, which normally appears in thesauruses, are solved, since the words are self-explanatory thanks to the contextual approach to the references. All words are found in that context they were used in the system to describe a given situation.

In DUVA there is an approach to solve the statistical confidentiality on the level of basic files. With our program SAFE (standardized anonymization functions for individual data), which is realized as a prototype, all known problems like primary and secondary confidentiality and also the problem of dominance can be solved.

DUVA contains many export interfaces to other systems, for example to SPSS, SAS, PC-AXIS, MS-Excel and so on. DUVA is conceptualized as an open system so that it is possible to integrate more standardized products in this system.

The Aims of a Statistical Information System

1. The superior aim of a statistical information system (of official statistics) should be to satisfy the statistical needs of the economy and the society in the best way possible. For that, it would be necessary to adjust the statistical information system consistently to user needs.

However, such user orientation can be made operational only indirectly, since the needs for statistical information are not constant, but undergo permanent changes, depending on the processes of social and economic changes. In addition, the users of statistics represent a broad range of sometimes quite different needs, so that it is practically impossible to adjust the planning involved in the conception of a statistical information system accordingly.

2. In order to be able to further pursue the first aim mentioned, however, without depending on changing needs of the various users, it is necessary to pursue a second aim, namely to make the entire data stock of official statistics the basis on which the different wishes of users are satisfied.

However, taking into account the manifold possibilities of evaluation, such data-oriented approach is possible only on the basis of micro data. This inevitably raises problems of data protection and of ensuring the statistical secret.

3. Thus, a third aim has to be mentioned, i.e., while implementing this data-oriented approach, it is necessary to strictly observe the statistical secret. The assurance given to respondents, that their data will be used for statistical purposes only, is not only prescribed by law, but also a necessary condition for them to accept the collection of statistics. If the respondents obliged to present information are not ready to supply truthful data, official statistics would hardly be in a position to provide the public with reliable information. It would be counterproductive if the data-oriented approach endangered the observance of the principle of scientific neutrality, objectiveness and neutrality regarding the collection and presentation of statistical results.

However, to keep the statistical secret and at the same time to try to satisfy all the wishes of users regarding the extent, the form and the required quality of data, is too costly and too time-consuming. So far, because anonymization efforts have been taken as a rule with regard to concrete evaluations and because tailor-made solutions are more expensive than general statistical publications, this often collides with the capabilities of statistical offices in terms of capacities available.

4. Thus, with a view to cost-benefit aspects it is necessary to look for possibilities to satisfy individual user requirements in an economical way. An economical means to achieve this aim is the multiple use of the data once collected, based on consistent standardisation of the whole stock of data and of the instruments used, since the value of statistical information is not lost in the course of utilisation, but, as a rule, even increases, when this information is used repeatedly. Consistent standardisation of the whole data stock and of the instruments to be used would make it possible to achieve this aim.

However, almost all statistical techniques have developed over the time and are characterised by a broad range of incompatible applications and organisational forms. Thus, in view of the abundance and complexity of tasks to be fulfilled by a statistical office, together with the capacity problems already existing and constraints by concrete deadlines to be observed, it seems to be almost hopeless to achieve standardisation.

5. That leads to a fifth aim: it is necessary for the statistical information system also to support statistical offices in coping with the traditional tasks of evaluation by providing means for their rationalisation in order to ease the burden on current statistical production and to set free the resources required for standardisation purposes. This should not be mixed up with the automation of traditional work processes, since solutions, once they have been developed, can be transferred to other applications only after appropriate standardisation. Such independence from data and programs is the only means to achieve that different work fields can profit from each other.

However, it will be seen quickly that this cannot be done without complete and standardised descriptive information. But so far such metadata have neither been standardised, nor been elaborated in a form generally suited for further processing.

6. Thus, the advantages of repeated uses, which are achieved thanks to such independence from data and programs, will not become effective, unless we strive for a standardised approach to the storage of metadata existing in the most different forms in data files, registers or even just in the minds of the employed staff. In view of the capacities available, this can only be done successfully, if metadata are recorded without redundancy and, if possible, at the very moment, when they are created, and if they can automatically be used later on in the statistical production process.

However, this requires looking at the entire statistical production process in its totality. But in view of the extent and the complexity of the range of tasks a statistical office is confronted with, there is a lack of transparency preventing even insiders from seeing all the structures and interrelationships of the various elements in the entire statistical production process.

7. Thus, in order to reduce the complexity of the overall system, an attempt should be made to elaborate roughly outlined structures of a statistical production process and to define the interfaces between its various components. The different components can only be refined as isolated elements, after this has been done.

Such project naturally raises scepticism, since, on the one hand, roughly outlined structures are too abstract for many critics to accept them as a strategic approach to possible improvements. On the other hand, concrete steps which might rapidly lead to visible success risk to produce failures in development, such as the deficits of presently existing statistical information systems.

8. For that reason the total concept to be developed should be set up so that a frame is created which makes it possible to start implementation in sub fields immediately. Thus, on the principle of trial and error it is then possible to develop prototypes suited on the one hand to collect practical experience as quickly as possible, and, on the other hand, to better assess and simultaneously to concretize the total concept and to protect it from possible future deficiencies.

As this work has to be done in parallel with current statistical production, which, however, in view of the problems of time and capacity, which exist anyway, must not be burdened additionally, such research-induced approach will never be free of friction in a statistical authority. On the other hand, one has to recognise that in view of the many interdependencies and the lack of transparency such work, if it is only done outside the statistical offices, without their active participation, can hardly have good chances of success.

9. For that reason it is particularly important that in statistical offices it should be possible to pursue an open systems approach, which allows to approach these aims gradually, perhaps restricted to several sub fields only, but at the same time ensures co-operation, excludes duplication of work and guarantees that partial steps, which are at first implemented in isolation from each other, are compatible and that elements which are unknown at a given moment or have to be newly developed can be integrated into the system later on without difficulty.

An orientation towards hardware and software standards is not helpful here in view of the methodological problems involved. In order to solve, in particular, the semantic problems, it is necessary to standardise the contents of the subject matter, which, in our opinion, can be achieved if a thesaurus is implemented as a user interface providing access to all metadata.

As no experience has so far been collected in official statistics with the use of thesauruses and as, basically, only linguists and librarians have traditionally dealt with this field of work, the said efforts might increase the risk that in addition to already existing communication problems in the field of information experts, data processing specialists, statisticians, lawyers, miscellaneous users etc. with their different technical terms, new difficulties for mutual understanding will directly be programmed.

10. In spite of these problems it is absolutely necessary to establish co-operation, because without interaction between offices and institutions and, above all, without interdisciplinary interaction it will not be possible for us to cope with the tasks involved in the implementation of a metainformation system. This can only be done on the road of dialogue, of learning jointly, of reaching agreement on an absolutely voluntary basis.

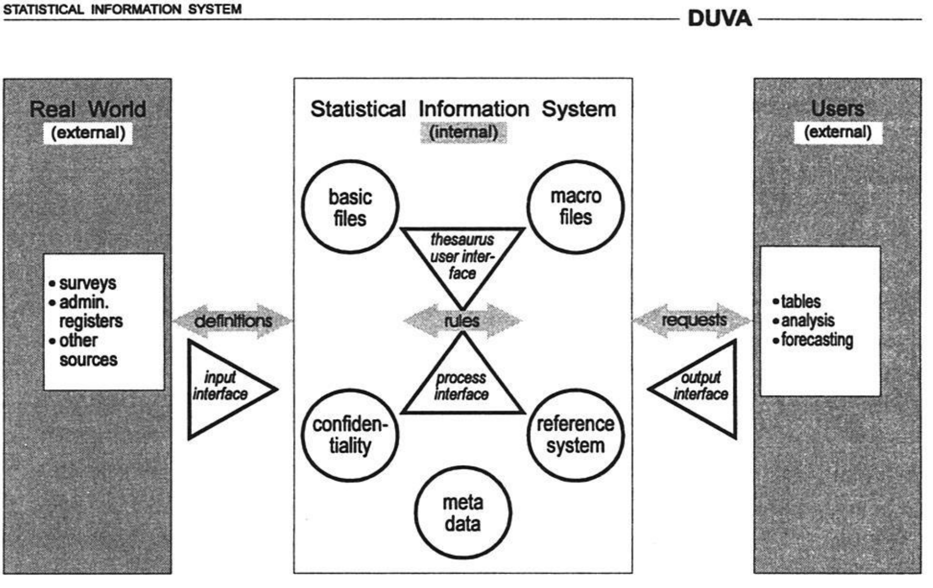
One should not be too optimistic in assessing the readiness of many to cooperate in such project. Mostly, there will be no co-operation, because the various interests are too different. It is indeed easier to stick to the currently applied methods of work, despite obvious deficits in the structure and operation of the present statistical information systems. In addition, most systems were developed in respect of concrete hardware and software solutions so that a new approach will be rejected as long as the new techniques available will not abolish all the deficits immediately and completely.

However, in view of the positive experience we gathered in the co-operation project of large cities in Germany, we are optimistic and believe that co-operation is possible and can be successful.

Elements and Interfaces of a Metadata Driven Statistical Information System

On this background and taking into account the profile of requirements described above (Figure 1), which should be met by a statistical information system, we should like to present below the major components which we found out as well as the possibilities of subject-related standardisation.

Figure 1 Outline of a Metadata Driven Statistical Information System
Definition of Standards and Interfaces



1. All plausibilised micro data from statistical surveys, automated administrative registers and statistically relevant results from other sources are potential data inputs for the statistical information system.

Theoretically, data input might begin with the survey itself, even with the planning for a survey. However, since for many years statistical offices have dealt with a broad range of very different statistics, partly involving highly complicated techniques, such approach would without necessity complicate the setting-up of a metadata-driven statistical information system or even render it impossible.

2. The system's input interface is the data model for those basic files, during the formation of which data collected and already plausibilised are stored in a standardised form and described in the metadata system, with regard to populations, objects, attributes and variables, categories, units of measurement and additional information.

3. All access operations in the system are made via a thesaurus (user interface), so that it is possible to proceed ahead step by step, to avoid redundancies, to recognise interrelations immediately and to integrate new words at any time without difficulty.

4. Relationships of any kind between objects, attributes and categories in the various basic files are established at this input interface by separately maintained classifications, references and reference tables, rather than by the basic files themselves. The semantic and delimitation problems, which are sometimes very difficult, have already been solved by others in advance and, thus, need not be regarded as an additional burden on the input process.

5. In order to ensure statistical confidentiality, all basic files are duplicated and anonymized in a standardised form. Thus, a data base, which is accessible by standard instruments, is available for evaluation purposes and, nevertheless, the occurrence of confidentiality cases is excluded. The implementation of this module is of utmost importance for efficient performance of the total system.

6. Evaluations as such, including the derivation of new attributes, are done in a standardised way, using the rules of the metainformation system. Data model No. 2 includes macro files which can be further processed flexibly in interchangeable formats.

7. All metadata regularly generated in the production process are stored automatically. Cumulatively they lead to permanent extension of the metainformation system and extend (and improve), with each further evaluation, the knowledge base for future uses.

8. Transparency for all elements of the statistical information system is provided by a registration or reference system, in which all components of the statistical information system are documented, which helps to avoid duplication of work and supports economical fulfilment of orders.

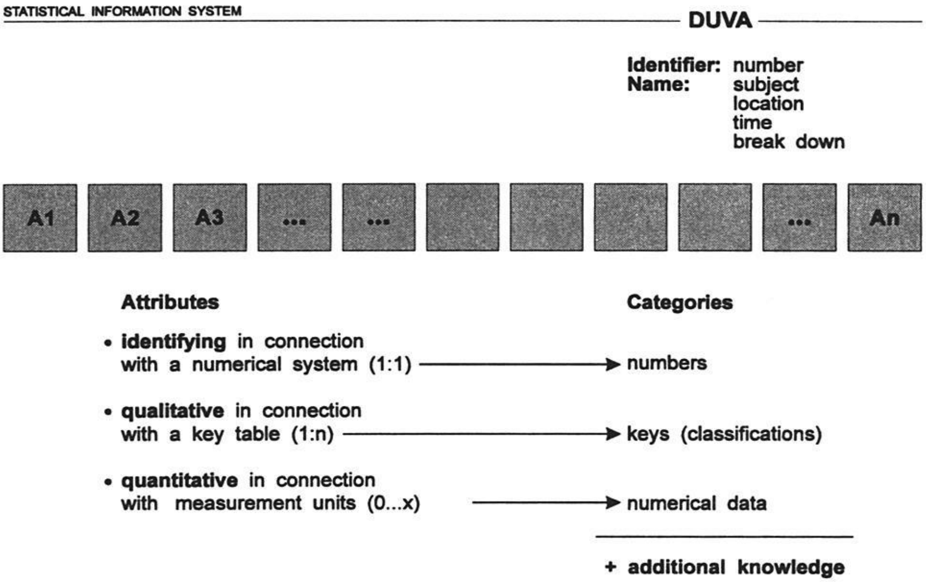
9. Thanks to the thesaurus, all queries for data can be drafted in natural language and be used for the collection of additional use-related metadata with the help of an electronic order processing system (output interface).

10. The electronic order processing system makes it possible to further develop and optimise the statistical information system permanently, in accordance with user needs. On the one hand, the user needs not grope his way through the specifications of the system, on the other hand, thanks to the integration of requirement-induced metadata the system is gradually and automatically adjusted by the users themselves according to their specific data needs.

How this can concretely be done shall be described in more detail below, by specifying the modules - basic files, statistical confidentiality, macro files, reference system and metainformation system - and the aspects of standardisation will be examined more closely by describing again 10 topics for each individual complex.

The data model for data input is a flat file used for the storage of plausible micro data (Figure 2). We believe that, if the following standards are observed, it will be possible to store all initial data of the statistical information system in a standardised form.

Figure 2 Data Model I: Basic Files



The standards for generating basic files are:

1. All data of a basic file must originate from one single survey or data source.
2. All data of a basic file have to belong to a population of objects unambiguously defined in terms of subject, space and time.
3. All attributes of a basic file have to be stored per each object in a data record of fixed length, arranged in an identical order.
4. All data of a basic file per each object have to be complete and free of redundancy.
5. All data of a basic file per each object have to have the same rank of plausibility.
6. All records of a basic file have to contain an unambiguous identification mark (record identifier) in the first field.
7. All qualitative attributes of a basic file have to be coded identically for identical categories.
8. All quantitative variables in a basic file have to use identical units for their value data.
9. All information required for the generation and utilisation of the basic file has to be described in the metainformation system.
10. All additional information, which is relevant for later interpretation of the basic file, must be recorded and stored separately as a non-formalised, free text.

Since the various statistical surveys in statistical offices cannot be seen in isolation, but should be regarded as components of a total system of mutually harmonised and interlaced surveys (full surveys, sector statistics, samples and the data from administrative registers), it is very important for the integrity of the statistical information system that it is possible to link all those different statistics to each other, which have to be stored in separate basic files.

This requirement is met, for example, by key tables. Reference tables are drawn up in the metainformation system, which are used, for example, to assign items within classifications, to define relationships of superiority and inferiority or to combine different classifications.

Figure 3 shows schematically how these reference tables permit that in the case of qualitative attributes the corresponding data in a basic file can be substituted without difficulty by corresponding other keys, so that it is possible to implement any sort of combination or grouping on a one-to-one basis.

There is no doubt that the generation of basic files is not an easy task. But it should be seen that the work required is done by those who are professionally engaged in statistics and so are suited best to solve the substantial problems that may arise.

Statistical Confidentiality

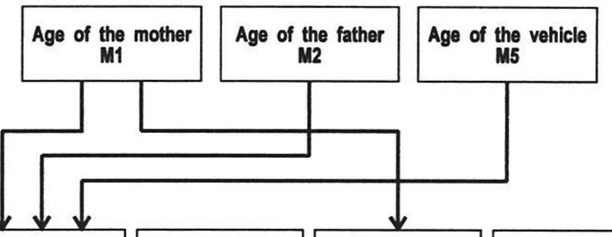
To anonymize data for confidentiality reasons, traditionally, means for all statistical offices to satisfy conflicting aims, i.e. to find a sound compromise between confidentiality, on the one hand, and adequate data quality, on the other hand. We have developed ten criteria, which would have to be observed, if this conflict in aims, with which official statistics are confronted as a consequence of the confidentiality issue, is to be solved with standard instruments.

Due to the fact that confidentiality measures are taken at the end of the statistical production process and always involve individual expenditure, it was necessary, in order to establish standardisation, to look for solutions which would make it possible to deal with confidentiality in an early phase of the statistical production process, similarly as in the case of plausibility checks.

Reference tables are suited to ensure a standardised approach to the confidentiality issue, because it is possible to replace keys which, in a certain combination, raise confidentiality problems by other corresponding keys.

Figure 3 Key Concept

Categories



Keys

Age in Years S1	Age in steps of 5 Years S2	Age in steps of 10 Years S3	Age in 4 Groups S4
0 0 < 1 Year	1 0 < 5 Years	1 0 < 10 Years	1 0 < 6 Years
1 1 < 2 Years	2 5 < 10 Years	2 10 < 20 Years	2 6 < 14 Years
2 2 < 3 Years	3 10 < 15 Years	3 20 < 30 Years	3 14 < 18 Years
3 3 < 4 Years	4 15 < 20 Years	4 30 < 40 Years	4 18 < 65 Years
4 4 < 5 Years	5 20 < 25 Years	5 40 < 50 Years	5 65 and older
5 5 < 6 Years	6 25 < 30 Years	6 50 < 60 Years	
6 6 < 7 Years	7 30 < 35 Years	7 60 < 70 Years	
7 7 < 8 Years	8 35 < 40 Years	8 70 < 80 Years	
8 8 < 9 Years	9 40 < 45 Years	9 80 < 90 Years	
9 9 < 10 Years	10 45 < 50 Years	10 90 < 100 Years	
10 10 < 11 Years	. .. .	11 100 and older	
. .. .	. .. .		
. .. .	. .. .		
. .. .	21 100 and older		
100 100 and older			

Reference Table

S1	→ S2
0	1
1	1
2	1
3	1
4	1
5	2
6	2
7	2
8	2
9	2
10	3
.	.
.	.
100	21

S1	→ S3
0	1
1	1
2	1
3	1
4	1
5	1
6	1
7	1
8	1
9	1
10	2
.	.
.	.
100	11

S1	→ S4
0	1
1	1
2	1
3	1
4	1
5	1
6	2
7	2
8	2
9	2
10	2
.	.
.	.
100	5

S2	→ S3
1	1
2	1
3	2
4	2
5	3
6	3
7	4
8	4
9	5
10	5
.	.
.	.
21	11

So far, standardised basic files offer a good basis, on which it is possible to organise this work so that all records with their concrete combinations of categories occur at least three times in the basic file. At first, this seemed to be unrealistic, in our opinion, but, in fact, it is principally feasible, because, though individual data are actually required to provide a data base which can be used flexibly to satisfy the most different evaluation needs, the statistical results themselves, however, always relate to statements about sets and subsets and to their structures and trends. But this is something for which just a few segments of the data base are utilised.

With our program system SAFE (standardised anonymization functions for individual data), which is now under development, we pursue an approach to confidentiality which uses copies of the basic files to exchange various keys of categories so that, on the one hand, the data stock includes only those records which occur at least three times, but, on the other hand, the frequencies of the various categories and also those of combinations of categories remain largely intact.

The perturbation technique used for the said purpose cannot be applied in the case of current economic statistics. Here it is better to solve confidentiality problems by using grouping techniques. And in order to avoid cases of predominance, it has turned out to be advisable to use the technique of suppression, because it does not prevent the results from being further processed by external users.

Doubtless, our SAFE confidentiality approach involves considerable manipulations with the initial statistical data, but the advantages connected with such standardised approach are substantial. For that reason, to arrive at an assessment of possible constraints and data quality, it is planned to also provide regular information about the extent to which evaluations made on the basis of anonymized data deviate from those based on original data.

But data quality also means, in our opinion, that the provision of data should not be delayed without necessity or be made more expensive by anonymization measures that may be required, which, however, is actually unavoidable with the present ways of production. In this sense data quality is worst in such cases where for considerations of data protection the supply of data, which would otherwise be possible, is either delayed or even rejected entirely.

Macro Files

With regard to statistical evaluations it is not advisable to operate directly on the level of basic files, even though they may be anonymized. According to experience, only some attributes are required for evaluations, and not the entire set of data, so that it is possible to raise the efficiency of work considerably, if the data are pre-compressed. This relates even more to standardised cases. How this can be done shall be demonstrated on a simple evaluation, for example, a file containing data on 100,000 individuals.

Population:	all inhabitants (100 000 persons) of the region Berlin-Mitte at a defined date.		**Subject** **Place** **Time**

Objekt:	individuals		
Serial No.	**Attribute**	**Category of attributes**	**Key**
M0	Identifier	serial numbers	1, ..., n
M1	Economic activity	economically active economically inactive	1 2
M2	Sex	male female	1 2
M3	Nationality	German non-German	1 2
M4	Marital status	single married widowed divorced	1 2 3 4

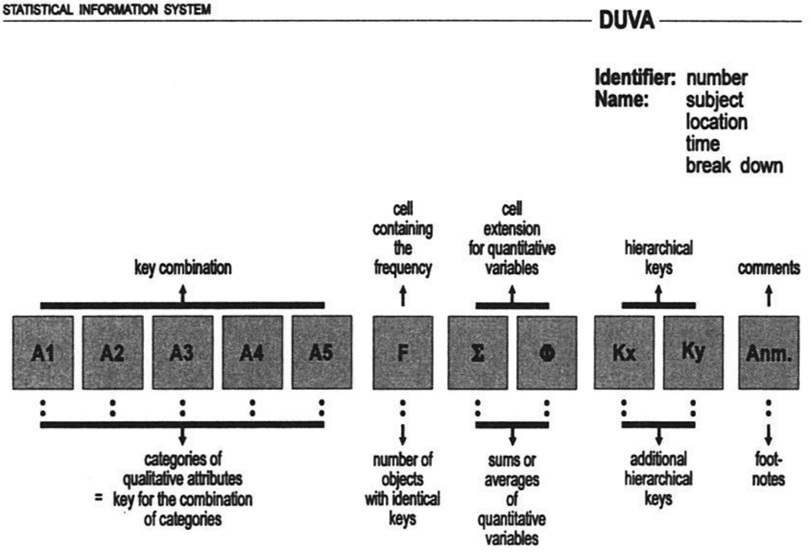
If all records in this file are sorted by attribute and category, it is comparatively easy to total the successively arranged identical records into one single record which is identified by the keys of the categories contained in a given combination so that it is necessary to just add the case number of the records totalled.

A large number of records were totalled in this way without any loss of statistical information. The theoretically imaginable number of data records in such a macro file is the result from the multiplication of the number of categories for all attributes, where the number of data records in the macro file can never be larger than the number of records in the intermediate file. In our example a maximum of 32 (2 x 2 x 2 x 4) data records may occur.

With this data model a standardised data base is available, which permits evaluations in any combination for all attributes included in the basic file, without making it necessary to use individual records. Its structure can easily be described completely, since it contains the same names as the basic file.

The structure of this macro file is determined entirely by qualitative attributes (Figure 4). It is, of course, also possible to transfer quantitative variables into macro files. However, they have no influence on the structure of the macro file, because they have always only one dimension, namely the unit of measurement. That is why quantitative values can be added to the frequency field as a sort of cell extension and also be stored in this macro file in the form of value sums or averages. In the same way it is also possible to store various hierarchical keys in this cell extension and even to include, if necessary, footnote references or also additional relevant information, which may be important for the interpretation of the cell contents.

Figure 4 Data Model II: Macro Files



Since basic files must not include data that can be determined in the statistical production process, it is necessary to ensure that all operations, which are characteristic of statistical evaluations, can be made during macro file generation by using the metainformation system. On the level of files, objects, attributes and categories, this includes operations of selection, combination, computation, etc.

The macro file data model can be evaluated at will, together with the corresponding metainformation. An isolated evaluation of the macro files may be done, for example, by using PC axis developed in Sweden. But the normal case, probably, is that macro files are generally used as a standardised intermediate product for further processing in the most different forms both inside and outside the statistical offices.

Reference System

Traditionally, the classical output of a statistical office is the transfer of tables, statistical reports, charts and comments. In addition, there are internal supply tables as well as on-line storage capabilities, the transfer of diskettes, CD-ROM, geographic and other specialised data banks in various fields.

A statistical information system should take account of this variety of uses and of the specific requirement profiles of the various users, without compelling the user to read through thick catalogues or to acquire specific knowledge in order to receive the statistics he needs.

It is important to grant all users equally easy access to the information potential and, at the same time, to use techniques which, as a feedback, analyse experience collected during the processing of user requirements and, in particular, their modifications in order to further develop the components of the system.

This is the purpose that our concept of a reference system, for which the following ten standardisation requirements were drawn up, is to serve:

1. Everything available in the statistical information system should be transparent to the user. The metainformation system with its descriptive information, to which each user should have access, serves this purpose.

2. All users should have access to the statistical information system at any time and in any way suited for them. There must not be barriers or unreasonable obstacles resulting from different states of knowledge or the use of different technologies. This easy access is realised by a thesaurus, since the use of natural language enables everyone to utilise the statistical information system. This approach makes it even possible to solve the problems raised by the use of different languages.

3. Data requirements should be met as inexpensively as possible. For that purpose the registration system provides information about the availability of final or semi-finished products (for example publications or macro files) or about the possibility of compiling results out of basic files. This makes it possible to find the financially most favourable solution.

4. This approach should be supported by electronic order processing to document and trace electronically all queries to the system irrespectively of whether they come from outside or inside the statistical office. By that, it is possible at the same time to implement a dynamic, use-related extension of the metadata, because new queries lead to new metadata and implementations in a form, different from that practised up to now, should simultaneously be seen as a service rendered in advance for future evaluations.

5. Electronic order processing can also be used systematically for dynamic system adaptation, by replacing parts which have proved to be less satisfactory in practice by better solutions so that the system is optimised continuously. For that purpose, the weak points are analysed periodically and appropriate improvements are made. In our opinion, this approach to optimization is a promising field for the use of expert systems.

6. Further possibilities to increase the efficiency of the system are provided by the thesaurus, which makes it possible to integrate into the metainformation system additional requirements which cannot be satisfied on the basis of a data-oriented approach, since the system can evaluate only those data which already exist in the system. Thus, unsuccessful queries can be evaluated systematically, in order to find out, for example, if missing data can either be replaced approximately by alternative

data or if an attempt should be made to open up new input-output sources. This would lead to a demand-driven extension of the data stock available.

7. As it is always extremely difficult for a statistical office to make outsiders understand the usefulness of its work, in particular that of a statistical information system, it is necessary to document systematically - in the sense of controlling - how the system is utilised. This can be achieved by interlacing the electronic processing system with an electronic accounting system, which combines the queries entering with the expenses for replying to them and automatically draws up invoices for those data supplies which are not free of charge.

8. Statistical offices have not only to take account of presently existing information needs, but they are also obliged to deal with quantitative historiography so that future clients, for example historians, can later receive statistics from the statistical information system. The appropriate means to take account of such future information needs is to register all activities of the system consistently in archives.

9. The service of a statistical information system should also provide for the possibility to establish links and to grant access to external data. For that purpose, it is necessary not only to supply descriptive information about the data available and the corresponding possibilities for processing them, but also to keep in the system a directory of sources and literature, which should be as complete as possible and continuously updated, so that access to text comments and external data sources is ensured at any time.

10. Last but not least, apart from the man-machine interface the system should also offer an individual form of service. For that it is continuously necessary to have organisation charts and to define responsible employees so that it is possible quickly to contact, if necessary, competent experts of the statistical offices who are able to give individual advice at any time.

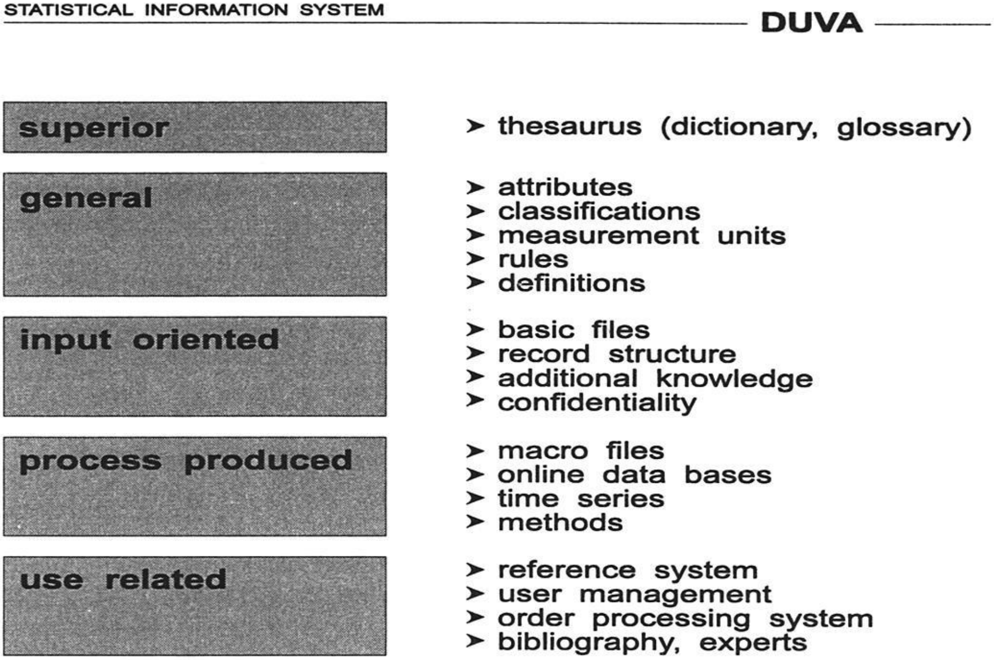
If these standards are taken into account, it will also be possible, in our opinion, to fulfil the expectations linked with a statistical information system, namely rendering optimal service to the clients of a statistical office, better than up to now, because, in addition to the supply of data, users are granted a possibility to exert influence in the individual components of the system by their concrete demands, and likewise it will be possible gradually to build up the metainformation system required for this purpose. How this can concretely be done will be described in the section that follows.

Metainformation System

What was said above was a description of the components needed for an operational metainformation system and of the possibilities to build it up gradually without running the risk that newly emerging components and demands cannot be integrated later into the existing system.

In this context, a superior element of such a system is the thesaurus, which at first is restricted to the inclusion of individual words used during the generation of individual basic files and the storage of the respective classifications. Then, links between these words and to the various elements will be established gradually (Figure 5).

Figure 5 Data Model III: Meta Information System



Attributes, classifications, units of measurement, rules and definitions have to be described as general components of this metainformation system. Input-oriented metadata include descriptions of basic files, of record structures and of the relevant extra knowledge required. Here the issue of confidentiality protection is dealt with in our project, too.

All metadata which are generated in the production process are stored with the process in which they are generated. This is done during metadata-driven production of macro files, during further processing operations, for example, in an on-line storage or for the production of time-series, and during the application of methods of mathematical statistics, which are used in the system.

Use-related metadata originate with the uses of the registration system, in particular, with electronic order management, supplemented by a bibliography and a list of organisational units and responsible staff members.

Final Remarks

The abundance of requirements with regard to subject-related standards of a metainformation system shows clearly that a large package of tasks has to be mastered here. It is certainly too much for an individual institution to cope with. However, in our opinion, promising fields of co-operation present themselves, in particular, with science. In addition, it is possible to approach the problem step by step.

As has been said, it is advisable to start by developing prototypes in selected sub fields, which can be used for new findings and later improvements.

In conclusion, it should be stressed that, in our opinion, without systematic elaboration of a theoretically founded concept for the solution of the metainformation problem, which is of relevance to the entire statistical information system, it will not be possible in future, either, to arrive at generally satisfactory statistical information systems, which means that, from this point of view, efforts in this field are justified in any case.

A NEW ERA FOR URBAN MODELLING?

Denise Pumain
Université Paris I, France

In the last two decades, several interesting innovations have appeared in the field of urban research. New paradigms such as the dynamics of open systems, self-organisation, synergetics, chaos, evolution, were recognised as conveying fruitful analogies for urban theory. New types of modelling were investigated, as sets of non-linear differential equations for spatial systems, cellular automata, multi-agents models, fractal growth, neural networks, evolutionary models... However, most of the new ideas have added very little to the pre-existing urban theory, and most of the models which have been built neither have been applied nor compared to other models.

Large urban models of the sixties failed mainly because of a lack of technical means for managing the huge amounts of spatially-disaggregated data that were required. Now that those problems may be solved thanks to the development of Geographical Information Systems, there is still a risk for urban modelling of an inability to meet the social demand because of an increasing trend of disconnecting "pure" "theory" and mathematical "modelling" from empirical research. It is suggested that the innovative power of the moment rests in the conjunction of new tools together with old ideas, large amounts of facts and relevant questions, and this would be a further incentive to develop a closer cooperation between the members of both Commissions.

Key Words: Urban Models, Urban Theory, Mathematical Models, Simulation Models

Introduction

Is there a real need for organizing, under the auspices of the IGU, an exchange of ideas between urban geographers and mathematical modelers? Urban geography is certainly a field where models are more frequently used than any other branch of the discipline. Most mathematical models which are developed in geography deal with urban questions or are conceived in an urban framework.

However, there is still quite a lot of serious and interesting urban geographical literature or teaching courses which ignore the issues of urban modeling. Not only qualitative studies of urban perception and collective representations or of urban hermeneutics are usually made without any help of mathematical tools, but new facts and prospective debates about urbanism or social problems are described and discussed, cross-cultural comparisons of urban systems are made without referring to the possible relevant existing models. On the other hand, how many mathematical models are built using urban objects as a pretext only and are developed only for the sake of mathematical consistency rather than any real geographical, practical or theoretical objective? How many models have appeared in one publication only, without any comparison of their properties with those of previous models, neither receiving the slightest beginning of an application to any kind of real data?

Such gaps and delays in communication are frequently observed in science, but they may be detrimental to the discipline. It seems to be a good moment for developing more interactions between the interests of urban geographers and mathematical modelers, since a series of more or less recent and promising innovations have appeared in the field of urban modeling. It is not our purpose to provide an impossibly exhaustive review of those new urban models. Several specialised reviews have already been published, about operational intra-urban models (WEGENER, 1994), or computer-oriented urban modeling (BATTY, 1992), or about models of systems of cities (MULLIGAN, 1984, PUMAIN, 1991), or urban models in general (BERTUGLIA, LA BELLA, 1991, BOURNE, 1993). After recalling the conditions under which those models could be useful and really used, we shall make some remarks about new theoretical models and modeling tools which could be of interest for urban geographers.

1. About the utility of models

Maximising utility has long been an objective, as a mathematical constraint, hypothesized by urban modelers for the behaviour of urban actors... What about the utility of urban models themselves for urban geographers? From recent reviews of urban modeling, two main broad fields of utilization may be distinguished: models as a means for the formalization and testing of urban theories, and models as pedagogical tools for teaching or for assistance to urban decision-makers.

Mathematical modeling is often referred to as an activity in theoretical research. However, perhaps under the influence of a similar trend in economics, the significance of "theory" has shifted from geography to mathematics. Is it necessary to recall that the quality of mathematical urban modeling is not only made of mathematical consistency but also its relevance and significance for the urban field? Even if theory remains a pure intellectual construction, its chances of success are linked to the number of facts that it is able to encompass and to enlighten.

A second utility of mathematical models is in their use as didactic tools, particularly when they are included in computer software packages. Mathematical models or simulation models on computers have the advantage of putting together several elements of urban theory and producing quantified tables of data or cartographic images, which are the results of a variety of sometimes rather complex interactions. The consequences of some theoretical hypotheses of spatial urban form and of its evolution may be studied and tested. The model user is also invited to do experiments by changing various parameter values and the rules of the model. Two main categories of users may be interested in such tools: first, students and young researchers may improve their knowledge of urban theory; second, urban managers and decision-makers can try to evaluate the possible consequences of some change they would like to bring into an urban situation.

Such practical urban models have met with a reasonable success, especially when they are designed as games. A software package like SIMCITY is known and used even by people who are not specialists of the urban field. There is here a challenge for urban geographers. If they want to avoid the severe regression which has been observed for instance in the average quality of the production of thematic maps with the increasing use of computer-assisted cartography and geographical information systems, specialists of the urban field should intervene personally in the process of designing those software dealing with urban problems.

2. New theories for urban modelling

Since the end of the seventies, urban models have contributed to a better integration of sometimes old ideas or empirical research findings into a more formalized urban theory. To my sense, two main improvements may be quoted in this field: bifurcation theory and self-organizing systems on the one hand (ALLEN and SANGLIER, 1979, ALLEN et al., 1981, WILSON, 1981), and individual spatial behaviour and random choice theory on the other hand (DOMENCICH and MACFADDEN, 1985).

Self-organisation theory has been useful for describing the evolution of urban systems at the level of a whole urban area (ALLEN et al., 1981, WILSON, 1981), or of a system of cities (ALLEN, SANGLIER, 1979, SANDERS, 1992). Models were transferred from physics to geography, using nonlinear differential equations for the simulation of urban change. Their novelty is not in the urban theory that they are referring to: agglomeration economies, economic base theory, distance-decay interactions and principles of central place theory are the main theoretical building blocks of those models. But for the first time those theoretical elements are integrated together in the same models; the models are both dynamic and spatial; they allow for modelling qualitative changes in urban structure from quantitative variations in the parameters. The models also try to derive in an explicit way the behaviour of a system at a higher level from the shape of the interactions between its elements observed at a lower level.

For all of those reasons, such dynamic urban models represent progress because they conciliate old ideas about urban systems and empirical observations about urban change with new mathematical and computing tools for experimenting and testing them (PUMAIN, 1989).

Major progress also has been made in the field of urban theory by developing much better insight, at a micro-level, into the behaviour of urban actors. Most operational models of today, even if referring to equilibrium theory, are far away from the concept of a perfectly informed and utility-optimizing homo economicus (MACMILLAN, 1993). Several surveys and empirical research have brought new and useful insights about the effects of a lack of information and uncertainties on spatial decisions, and to what extent differences in individual trajectories and tastes could induce a large dispersion of behaviours. Decisive advances have been made in the fields of residential mobility and housing strategies, job search behaviour and about the choice of transportation modes (CLARK, 1985, TIMMERMANS, BORGERS, 1985). The location processes of various kinds of urban industries, public facilities and retail services were investigated in the same way (BIRKIN and WILSON, 1986, WRIGLEY, 1988). Random utility or discrete choice theory and classical techniques of multinomial logit models are used for integrating such knowledge into predictive models of urban development (WEGENER, 1994).

Fractal geometry is another appealing field for urban modeling, for several reasons. First, it provides a mean for describing spatial structures where an internal order is compatible with a large part of randomness in the location and size of subsystems. Fractal structures also are revealing nested hierarchical principles and multifractals are helpful for describing systems where several levels of organisation are recognisable. Applications to urban systems, until now, have been of several kinds: the fractal structure of built-up areas has been investigated by comparing the length of the perimeter and the surface of urban areas, as well as by measuring the fractal dimension of the urban borders, or of the built-up space (FRANKHAUSER, 1994, BATTY, LONGLEY, 1994). Several attempts have been made at simulating the spatial extension of urban built-up areas with models of fractal growth (BATTY, 1991, FRANKHAUSER, 1994). However, we still lack of the means for validating the similarity between the computed and observed images. On the other hand, there is not always a clear identification of the social processes which are actually formalised by the parameters in such models. It is then interesting that for instance FRANKHAUSER (1994) identifies three kinds of urban processes which may lead to fractal structures of built-up areas, including the duality between urban extension and transportation networks ramification, polycentric growth, and the trend to preserve empty places inside urban spaces.

Fractal formalization also may lead to an improvement in the spatial theory of urban systems (ARLINGHAUS, 1985 and 1989). Such an approach invites us to integrate into models of the development of systems of cities the idea that, instead of using an Euclidean space as a backcloth, it could be more relevant to consider a fractal space, linking various speeds of movements in geographical space according to the scale under consideration.

Several other possibilities of modeling urban spatial dynamics by developing analogies with physical processes or mathematical computing have been explored, but without giving rise to enough empirical testing for validating their relevance for urban studies. This may be the case, however promising the models may appear, for instance, for equations of spatial diffusion (ZHANG, 1990), or for systems of neural networks (FISCHER, 1993, DIAPPI , OTTANA, 1994, WHITE, 1989), or for the theory of evolution applied to urban modeling (ALLEN, MACGLADE, 1987, RABINO, 1993).

Progress in theoretical urban geography may come not only from transferring and adapting ideas from other disciplines. New observations may lead to the reformulating of urban theories, leading to more general description of urban systems: for instance, the careful studies of several kinds of interactions taking place between cities help to develop a broader concept of city networks than those included in central place theory (DEMATTEIS, 1990). Progress also can emerge from the possibility of testing "old" hypotheses or models, with newly available large sets of properly prepared data. For instance, a large part of the literature on the rank-size rule and city primacy is invalidated because it uses data sets that are too limited or non comparable urban definitions (PARR, 1985, GUERIN-PACE, 1993). By using a large data base, including strictly comparable figures (MORICONI-EBRARD, 1993), deeper insights have recently been brought into questions like the domain of validity of the Paretian model for the distribution of city sizes, the generality of a stochastic model for the repartition of urban growth within an urban system, or to the justification of Jefferson's "law of primacy".

3. New tools for urban analysis and simulation

At least as significant as the new theoretical paradigms for urban research are the new modelling tools now available. M. WEGENER (1994) shows how those merely technical improvements succeeded in solving almost all theoretical difficulties which justified LEES famous "requiem for large scale models" in 1973. M. WEGENER identifies at least twenty research centres all over the world where operational integrated urban models are used and experimented. He distinguishes between "unified" models, which are tightly integrated behind a single modeling objective, and "composite" models, where models of urban subsystems (such as, for instance, population, housing, transportation, employment...) are only loosely linked within a broader framework. Whatever the chosen orientation, the existence and operationality of those models is linked to the development of large and reliable data bases, of appropriate software such as Geographical Information Systems, as well as to the above mentioned theoretical progress.

The rather good performance of henceforth "classical" modeling tools of urban geography when included in Geographical Information Systems is also evidenced for instance by the success of a programme like GMAP as developed since 1991 at the School of Geography at the University of Leeds (CLARKE, 1990, CLARKE and WILSON, 1987, BIRKIN et al., 1990). Micro-simulation procedures as well as spatial interaction models or even neural network approaches are integrated in a library of available programs working on urban or regional data bases, and used for solving problems of location or of marketing for private firms or governmental agencies. The question may be raised of the possible feedback of such applied studies on fundamental research: it may help to elaborate more operational modeling tools, but carrying out innovative research may become more difficult.

The introduction of systems for managing geographical information by computers may also change our way of constructing models. Among the first attempts at modeling the spatial evolution of urban systems are the Monte Carlo simulations by MORRILL, following a method first initiated for the simulation of the spatial diffusion of innovations by T. HÄGERSTRAND. MORRILL made simulation models at the scale of an urban area, for describing the extension of the Negro ghetto in Seattle with simple rules of social spatial segregation (MORRILL, 1965a). He also simulated the development of a system of central places according to migration flows between places (MORRILL, 1965b). In this direction of spatial simulation, progress in computer science and artificial intelligence opens new avenues for urban modeling.

A further set of models use cellular automata. A cellular automaton consists of an array of cells which may be in any one of several states. At each iteration, the state of each cell may remain the same or change to another state according to the states of neighbouring cells. Some models, like the famous "game of life", are very simple, but more complex cellular automata are useful in geography if they allow for several possible states and for sophisticated definitions of neighbourhood. W. TOBLER (1979) first mentioned cellular automata as one of the "geographical type" of models. H. COUCLELIS draw attention to their use for modeling micro-macro relationships in spatial dynamic models (1985) and for deriving complex dynamics from simple rules (1988). R. WHITE (1991) applied this formalism to simulate the evolution of land use patterns within urban areas. WHITE and ENGELEN (1993) checked the fractality of the simulated urban spatial structures.

Other simulation methods like "multi-agent systems" may allow more flexibility in spatial modeling of urban systems by generating the evolution of a global system from rules about local interactions of various types and ranges between "agents" of different kinds (BURA et al. 1993). Like all simulation models, cellular automata or multi-agent systems raise delicate problems of calibration and validation, which are not solved at the moment. However, they may be helpful tools for communicating the intuitive complexity of spatial urban dynamics to decision-makers.

The progressive inclusion of spatial analysis tools and of dynamic models within GIS may also change our way of doing research and our conception of modeling in the future. The power and speed of computing, the quasi immediacy of response between the formulation of an hypothesis and its testing, together with the instantaneous visual display of results may invite the development of a kind of "soft" modeling (MIKULA et al. 1994). The drawback of such flexibility is the development of too many "ad hoc" models with a great loss in simplicity and generality, which remain however two decisive properties of good models.

4. Urban theory without models

It should not be forgotten that a large share of interesting urban research is conducted and results are obtained without any reference to nor the help of mathematical modeling. Many of the fields where progress has recently been registered belong to that category. Among them, one could quote a few themes for which special sessions were organised at the Berlin meeting of the IGU Commission on Urban Development and Urban Life:

- counter-urbanization and metropolisation (which future for the contained urban areas and the hierarchised settlement systems?);
- social integration, spatial segregation and exclusion ("polarisation of urban space");
- city image and city marketing (connected to the questions of urban identity and of hermeneutics);
- urban poverty and the role of informal economic sectors in the urban development.

All those questions are socially important. New ideas and theories are emerging about them. They have been enlightened by significant results from empirical research and by connecting ideas coming from other fields of social sciences. Despite their social and scientific interest, they heave nor received major contributions from urban modelers. Not yet? Is the urban modeling to be conceived mainly as a formalisation process, which would come after the social, intuitive and empirical aspects of an urban question have been explored by more traditional means of investigation? Are the mathematical or the "purely theoretical" formulations real barriers to a dialogue with the empirical research and with the experimentation?

Such interrogations bring back to the question asked in the first part of this paper, about the utility of urban models.

Conclusion

So many more authorized papers already have been written on the status of modeling in urban geography and so many other deeper studies should be made about this topic that we can only consider the few remarks above as a short contribution for bringing together the work of our two Commissions and as an incentive to develop linkages through further meetings.

BIBLIOGRAPHY

ALLEN, P., SANGLIER, M. 1979. "A dynamic model of growth in a central place system", Geographical Analysis, 11: 256-272.

ALLEN, P., BOON, F., DENEUBOURG, J.L., DE PALMA, A., SANGLIER, M. 1981. Models of Urban Settlement and Structure as Dynamic Self-organizing Systems. Washington, DC: US Department of Transportation.

ANAS, A. 1984. "Discrete choice theory and the general equilibrium of employment, housing, and travel networks in a Lowry-type model of urban economy", Environment and Planning A, 16: 1489-1502.

BATTY, M. 1976. Urban Modelling. Cambridge: Cambridge University Press.

BATTY, M. 1992. "Urban Modelling in Computer-Graphic and Geographic Information System Environments", Environment and Planning B, 19: 663-688.

BATTY, M., LONGLEY, P. 1994. Fractal cities, a geometry of form and function. London and San Diego: Academic Press.

BERTUGLIA, C.S., LA BELLA, A. eds. 1991. I sistemi urbani. Milano: Franco Angeli, 2 vol.

BIRKIN, M., CLARKE, G.P., CLARKE, M., WILSON, A.G. 1990. "Elements of a model-based geographic information system for the evaluation of urban policy", in L. Worrall, ed. Geographical Information Systems: developments and applications. London: Belhaven Press, pp. 133-162.
BOURNE, L.S. 1993. "The demise of gentrification: a commentary and prospective view", Urban Geography, 14, 1: 95-107.
BURA, S., GUÉRIN-PACE, F., MATHIAN, H., PUMAIN, D., SANDERS, L. 1993. "Multi-agents systems and the dynamics of a settlement system", Budapest, 8th Colloquium on theoretical and Quantitative Geography.
CAMAGNI, R., DIAPPI, L., LEONARDI, G. 1986. "Urban Growth and Decline in Hierarchical systems: a supply-oriented dynamic approach", Regional Science and Urban Economics, 16: 145-160.
CLARK, W.A.V. 1985. Residential segregation in American cities. University of California at Los Angeles.
CLARKE, M. 1990. "Regional Science in industry and commerce: from consultancy to technology transfer", Environment and Planning B, 17: 257-268.
CLARKE, M. and WILSON, A.G. 1987. "Towards an applicable human geography", Environment and Planning A, 19: 1525-1541.
COUCLELIS H. 1985. "Cellular worlds: a framework for modelling micro-macro dynamics", Environment and Planning A, 17: 585-596.
COUCLELIS, H. 1988. "Of mice and men: what rodent population can tell us about complex spatial dynamics", Environment and Planning A, 20: 99-109.
DEMATTEIS, G. 1990. "Modelli urbani a rete", in F. Curti, L. Diappi, a cura di Gerarchie e reti di città: tendenze e politiche. Milano: Franco Angeli.
DIAPPI, L., OTTANA, M. 1994. City network system: the neuronal approach. Milano: Politecnico.
DENDRINOS, D.S., MULLALY, H. 1985. Urban Evolution: Studies in the Mathematical Ecology of Cities. Oxford: Oxford University Press.
FIK, T.J., MULLIGAN, G.F. 1990. "Spatial Flows and Competing Central Places: Towards a General Theory of Hierarchical Interaction", Environment and Planning A, 22: 527-549.
FISCHER, M. 1993. "New Developments in Mathematical modeling", Budapest, 8th Colloquium on Theoretical and Quantitative Geography.
FRANKHAUSER, P. 1994. La fractalité des structures urbaines. Paris: Anthropos.
HAAG, G. 1994. "The rank-size Distribution of Settlements as a Dynamic Multifractal Phenomenon", Chaos, Solitons and Fractals, 4, 4: 519-534.
HAAG, G., MUNZ, M., PUMAIN, D., SANDERS, L., SAINT-JULIEN, T. 1992. "Interurban Migration and the Dynamics of a System of Cities", Environment and Planning A, 24: 181-198.
LEPETIT, B., PUMAIN, D. eds. 1993. Temporalités urbaines. Paris: Anthropos.
LOMBARDO, S., PUMAIN, D., RABINO, G., SAINT-JULIEN, T., SANDERS, L. 1987. "Comparing Urban Dynamic Models: the unexpected differences in two similar models", Sistemi Urbani, 2.
MACMILLAN, W.D. 1993. "Urban and Regional modelling: getting it done and doing it right", Environment and Planning A, Anniversary Issue, 56-68.
MIKULA, B., MATHIAN, H., PUMAIN, D., SANDERS, L. 1994. "Dynamic modelling and Geographical Information System: for an integration", Paris, unpublished paper.
MORICONI-EBRARD 1993. L'urbanisation du monde depuis 1950. Paris: Anthropos.
MORRILL, R.L. 1965a. "Migration and the spread and growth of urban settlement", Lund Studies in Geography, Serie B, 26.
MORRILL, R.L., 1965b. "The Negro Ghetto: problems and alternatives", Geographical Review, 339-361.
MULLIGAN, G.F. 1984. "Agglomeration and central place theory: A review of the literature", International Regional Science Review, 9: 1-42.
PUMAIN, D. 1989. "Spatial dynamics and urban models", in J. Hauer, H. Timmermans, N. Wrigley, eds. Urban dynamics and spatial choice behaviour, Dordrecht: Kluwer Academic Publisher, pp. 155-173.
PUMAIN, D. 1991. "City Size Dynamics in Urban Systems", Cambridge, Department of Archaeology, Symposium on Dynamic modeling and human systems.
PUMAIN, D., SANDERS, L., SAINT-JULIEN, T. 1989. Villes et auto-organisation. Paris: Economica.
RABINO, G. 1993. "The Evolution Theory and Urban Modeling", Budapest, 8th Colloquium in Theoretical and Quantitative Geography.
RACINE, J.B. 1993. La ville entre Dieu et les hommes. Genève: Presses bibliques et universitaires et Paris: Anthropos.
ROSSER, J.B. 1991. From Catastrophe to Chaos: a General Theory of Economic Discontinuities. Boston: Kluwer.
SANDERS, L. 1992. Système de villes et synergétique. Paris: Anthropos.

TIMMERMANS, H., BORGERS, A. 1985. Spatial Choice Models: Fundamentals, Trends and Prospects. Eindhoven: University of Technology.

TOBLER, W. 1979. "Cellular Geography", in S. Gale, G. Olsson, eds. Philosophy of Geography. Dordrecht: Reidel.

WEGENER, M. 1994. "Operational Urban Models, State of the Art", Journal of the American Planning Association, 60, 1, Winter.

WEIDLICH, W., HAAG, G. 1982. Concepts and Models of a Quantitative Sociology. Heidelberg: Springer Series in Synergetics.

WHITE, R. 1989. "The Artificial Intelligence of Urban Dynamics: Neural Network Modelling of Urban Structure", Papers of the Regional Science Association, 67: 43-53.

WHITE, R. 1991. "A Cellular Automata Approach to the Evolution of Urban Land Use Patterns", Stockholm, 7th colloquium of Theoretical and Quantitative Geography.

WHITE, R., ENGELEN, G. 1993. "Fractal urban land use patterns. A cellular automata approach to the evolution of urban land use patterns", Environment and Planning A, 25.

WILSON, A.G. 1981. Catastrophe Theory and bifurcation: applications to Urban and Regional Systems. London: Croom Helm.

ZHANG, W.B. 1990. "Stability versus instability in urban pattern formation", Socio-spatial dynamics, 1: 41-56.

Abhandlungen - Anthropogeographie
Institut für Geographische Wissenschaften, FU Berlin
Band 52, Berlin 1994

COMPETING ORDER PARAMETERS IN A SELF-ORGANIZING CITY

Juval Portugali and Itshak Benenson
Tel-Aviv University, Israel

In this paper we consider the city as a complex, open, and thus self-organized system, and describe it by means of City-1: a cell space model specifically designed to examine the impact of international migration on the urban dynamics. Our specific case study concerns the recent inflow of new immigrants from ex-USSR countries to Israel's towns and cities. We perceive cities from the perspective of HAKEN's synergetics approach to self-organization in which ordered steady-state in the system is reached as a consequence of a competition among, and enslavement of, some order parameters. In this paper we focus on the interplay among two order parameter which are central to the urban dynamic: the cultural order parameter and the economic order parameter.

Introduction

Self-organization is a property of open, complex and consequently far from equilibrium, systems. 'Open' in the sense that they exchange matter, energy or information with their environment and 'complex' in the sense that the number and properties of their constitutive parts are undefinable (whether quantitatively or qualitatively is a matter of philosophical dispute).

Though the theory originated within the disciplines of physics and chemistry (e.g. HAKEN 1983a, 1983b; NICOLIS and PRIGOGINE 1977), the 'city', from the start, was referred to as an example for a self-organizing system. First, as a metaphor to convey the notion of self-organization (NICOLIS and PRIGOGINE 1977), and later as a subject-matter in the general attempt to expand the notion of self-organization to the socio-human domain (ALLEN 1981; ALLEN and SANGLIER 1981; ALLEN et al. 1985; WEIDLICH 1987; DENDRINOS and SONIS 1990).

HAKEN's synergetics approach to self-organization suggests that since the complexity of such systems constrains our ability to treat them causally and mechanistically, it will be more useful to focus on their morphological behavior. Such an examination revealed that the evolution of self-organizing systems follows a distinct routinized path: long periods of steady dynamics, interfered by relatively short periods of strong fluctuations and chaos. According to HAKEN (1985) while in steady state, the system is governed by a certain order parameter; while in fluctuations and chaos, several order-states compete, until one wins, enslaves the system, brings it to a new steady state and thus becomes the new order parameter.

In two previous studies (PORTUGALI, BENENSON and OMER 1994; PORTUGALI and BENENSON) we have suggested seeing cities and metropolises as open, complex, far from equilibrium, and thus self-organizing, systems. Perceiving cities as such we have developed two models: a cellular automata (CA) model we termed *City* on which various heuristic games could be played as means to study phenomena of sociospatial segregation of national, ethnic, and other groups in the city, and a cell-space model, termed *City-1*, on which the relations between international migration and the internal structure of cities, could be played, with special reference to the recent migration waves to Israel from ex-USSR countries. The aim of the present paper is to use *City-1* as means to study cultural spatial segregation in light of the interrelation between two order parameters which play a central role in shaping the socio-spatial structure of cities: the cultural order parameter (COP) and the economic order parameter (EOP).

The model

The formal description of our model is given in the Appendix. Generally speaking, *City-1* is a *cell-space* (CS) model (ALBIN 1975; COUCLELIS 1988; TOBLER 1979) whose territory is a 2D lattice of cells H_{ij} {0 < i < n; 0 < j < m), each of which might be considered a single house. Individuals occupy or leave the houses and thus participate in generating the migration dynamics of the city and its socio-spatial structure. One individual only can occupy a house. Each individual is characterized by *status and tendency. Status* refers to the individual's economic ability, while *tendency* his/her potential to improve status. Individuals are of two *origins*: *Veterans* and *Olim* (Hebrew for Jewish new immigrants; singular *Ole*). Depending on *status* and *tendency,* each individual can either buy or rent a house, that is, a house can be in two *forms* of occupation: rented or privately owned.

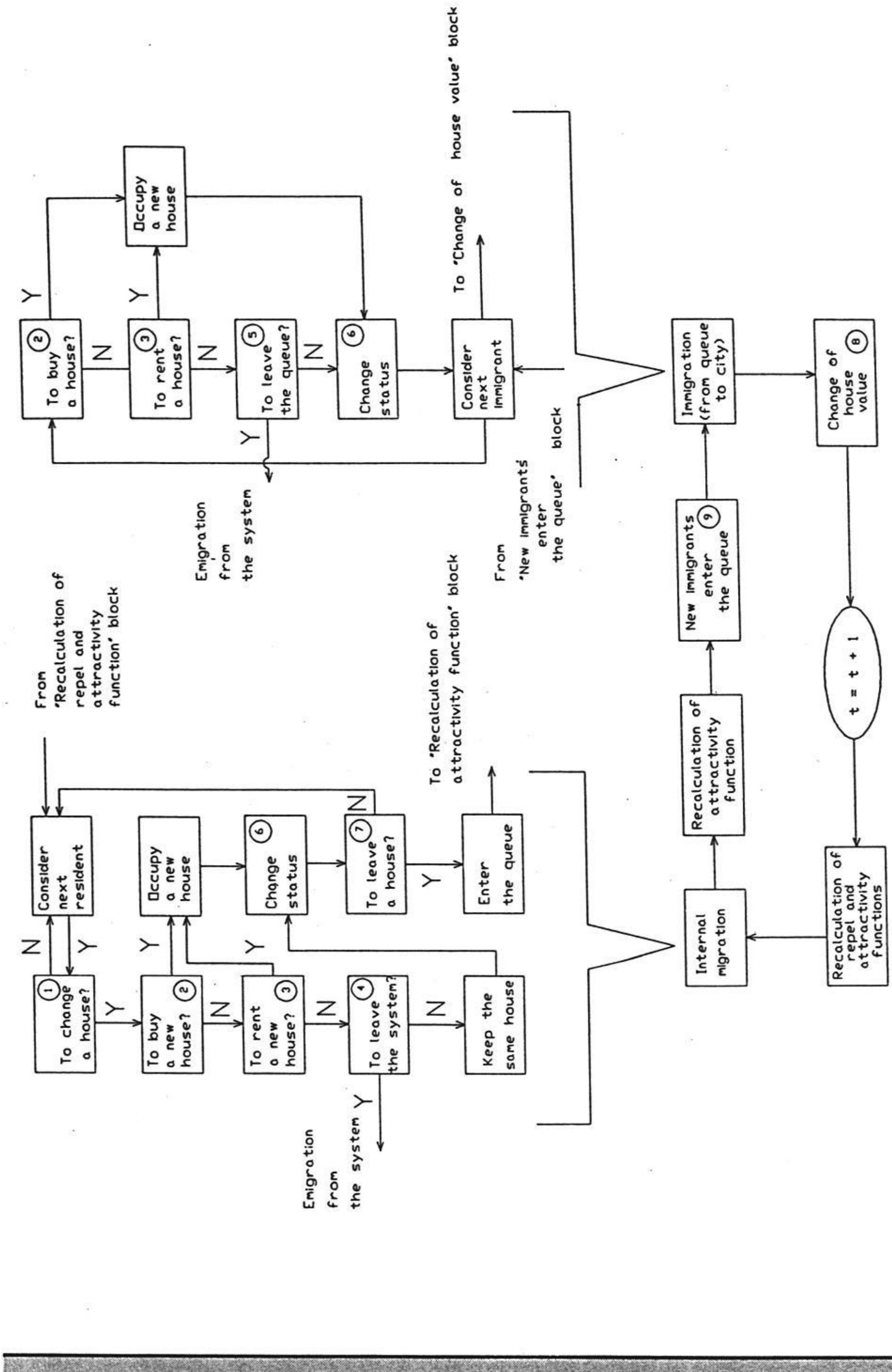
Each house has a *value.* The value of a vacant house is some function of the values of its neighbors. The value of an occupied house is, additionally a function of the *status* of the occupant.

An individual can either rent or buy a house. Both cases imply *payments* which are proportional to the *value* of the house and both also effect the *status* of the paying individual in a way that we define below.

Each sequential time-step every individual takes the following series of decisions (Figure 1): A resident of the city decides whether to stay in, or leave his present house, and in the latter situation, whether to rent or buy a new house in the city, to altogether leave the system, or only to leave his house in the city. In the latter case the resident becomes " homeless", i.e. he or she enters the queue for houses. A new immigrant (veteran or Ole) enters the city by joining the queue and then decides whether to buy/rent a house, further wait in the queue for housing, or altogether leave the queue (and thus the city). The various individuals take such decisions by cormparing their own properties (status, tendency) to the properties of their houses, the vacant houses, and their nearest neighbors.

The decisions have probabilistic and deterministic components. They are probabilistic in the sense that the larger the gap between the properties of the individual and his/her neighbors, the higher is the probability that he/she will decide

Figure 1 Structure of Decision Making



to leave the house, and visa versa with respect to occupying a place. They are deterministic in the sense that the individual can buy or rent the house when his/her status is sufficiently high compared to the value of a house, or when he/she cannot afford paying the rent, or the mortgage, of a house in the city and has to leave the city. The result of the latter is a *queue* of individuals who want to occupy a house within the city territory, but cannot afford doing so.

Competing order parameters in the city

Cultural (ethnic and/or national) and socio-economic spatial segregations are typical phenomena in cities in general and in cities which are subject to international migration processes, in particular. This came out also from our previous study (PORTUGALI and BENENSON) in which we have used City-1 as means to simulate and study the implications of the recent migration wave of Jews from ex-USSR countries, to Israeli cities. We have found that cultural relations and socio-economic relations, as formulated in our model, act as two order parameter which compete and interact in a complex way in the evolution of the city. While the aim of our previous paper was to study spatial segregation in the city, our present paper aims, as noted above, to study the complex interplay between the cultural order parameter (COP) and the economic order parameter (EOP).

For this purpose we have played on City-1 three scenarios which described the dynamics of the city and the interplay between the EOP and COP when (i) the EOP is dominant in a certain way, (ii) the COP is dominant in a certain way, and (iii) both the COP and EOP act together. In the following we will describe each scenario in a sequence and examine their evolution by means of three devises: segregation indices (Figure 3), three moments (T=20, T=40, T=120) of the evolving spatial cultural segregation (Figure 4), and the same three moments of the evolving land value surface (Figure 5). Note that Figure 5 can be seen as a surrogate to the evolving spatial economic segregation of individuals in the city. In all scenarios the initial mean value and STD of the status of veterans is four times higher than that of Olim; mean tendency of Olim is four times higher than that of veterans; STD of Olim tendency is four times lower than that of veterans.

(i) When the EOP is dominant in a certain way

The EOP is dominant when we assume two forms of economic antagonism: First, the probability for an individual to leave his house increases with the growth of the difference between the individual's and his neighbors' economic status. Second, the attractivity of a house increases with the increase of the economic improvement of its environment (Figure 2). As can be seen in Figure 3 and 4, in this case the city evolves with no cultural spatial segregation. The city's evolution is thus dominated by the EOP as some economic segregation can be observed (Figure 5). It is important

Figure 2 Repel/Attractivity Functions for Economic/Cultural Interactions

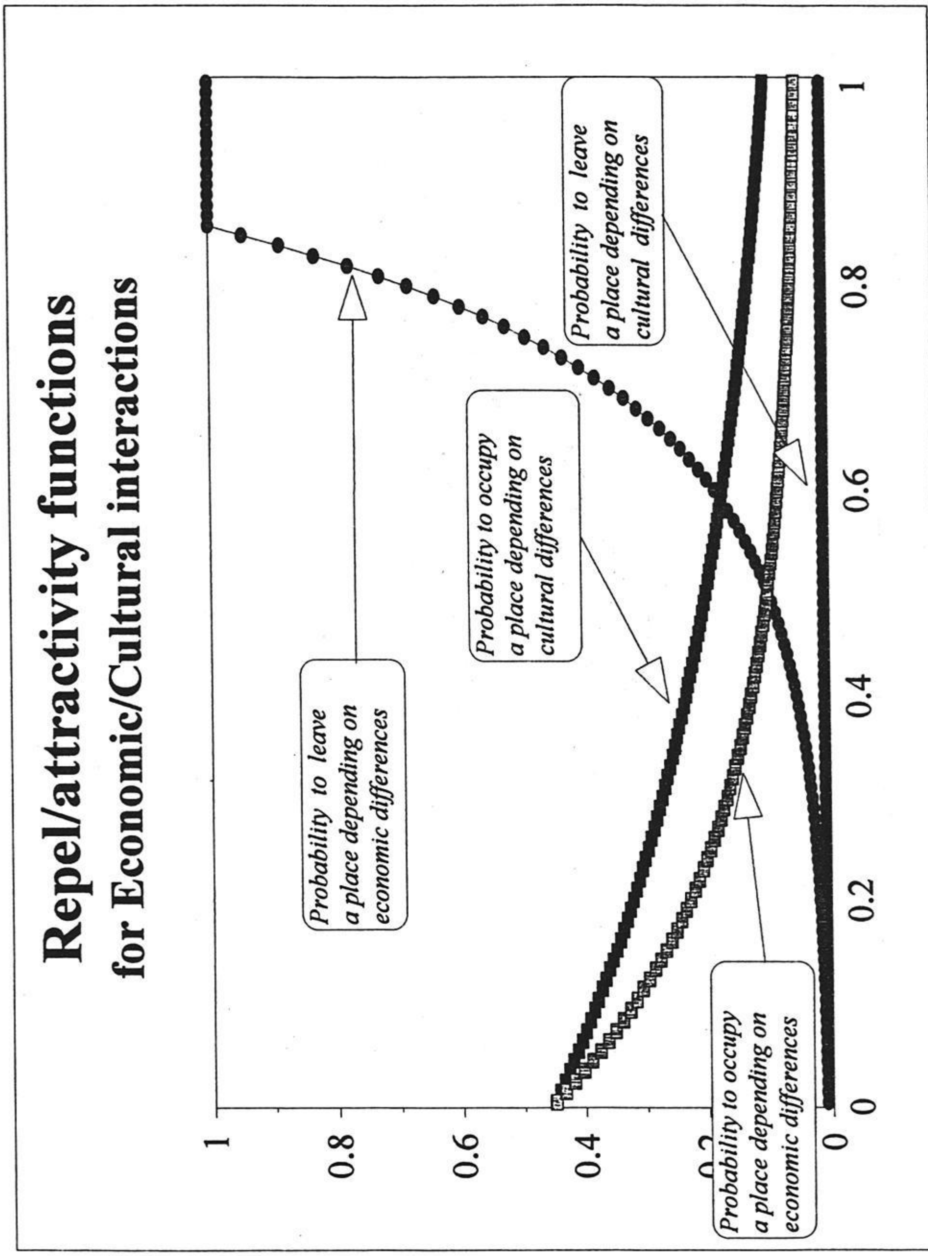
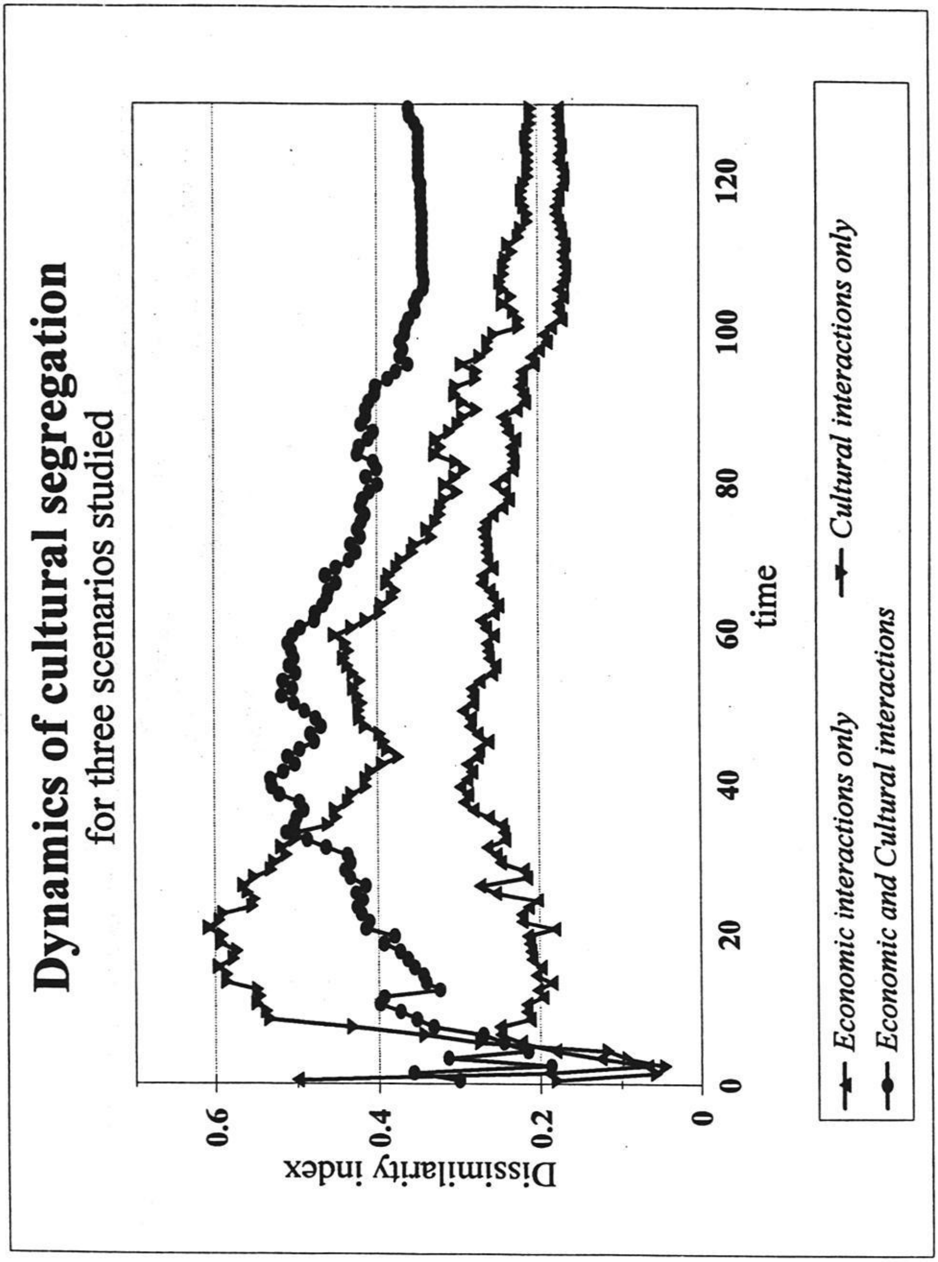


Figure 3 Dynamics of Cultural Segregation for Three Scenarios Studied



to note, however, that despite the fact that this scenario is governed by economic forces only, economic segregation is not very distinguished.

(ii) When the COP is dominant in a certain way

Here there is no economic antagonism and the COP is dominant in the sense that we assume a "mild" cultural antagonism between veterans and Olim. The latter shows up, first, in that the individual's decision to leave a house does not depend on the characteristics of his neighbors, and second, in that when choosing a new house the individual prefers neighbors of his/her own kind. "Prefers" implies that the attractivity of a fully friendly environment is 0.45 compared to 0.1 of a fully strange one (Figure 2). In this case the evolution of the city and the interplay between EOP and COP is more complex than in the previous scenario: At the beginning ($0< T<40$), the COP slaves the EOP and dominates the city's evolution: the city is segregated culturally (Figure 3 and 4) and economically (Figure 5). Then gradually the EOP and the COP neutralize each other, until in T=120 no cultural segregation (Figure 4), nor economic segregation (Figure 5) can be observed. Apparently this is a result of the higher economic tendency of Olim relative to the veterans.

(iii) When COP and EOP interact

Here we assume the existence of both economic and cultural antagonism as above, and the repel and attraction functions are thus the sum of both antagonisms as defined in the Appendix. This scenario is rather interesting. At first, the evolution of the city is governed fully by the COP - a strong trend towards cultural segregation can be observed in Figure 3. Then segregation decreases but do not disappear, and in fact it can be observed all the way to T=120. The interesting part is the role played here by the EOP. Indeed, as can be seen in Figure 3, it acts to moderate cultural spatial segregation, but at the same time it does not eliminates it, but actually acts as a catalyst and reinforces it: the COP enslaves the EOP and the city becomes culturally (Figure 4) and economically (Figure 5) segregated in line with the COP.

Figure 4 Dynamics of Olim/Veterans Spatial Distribution

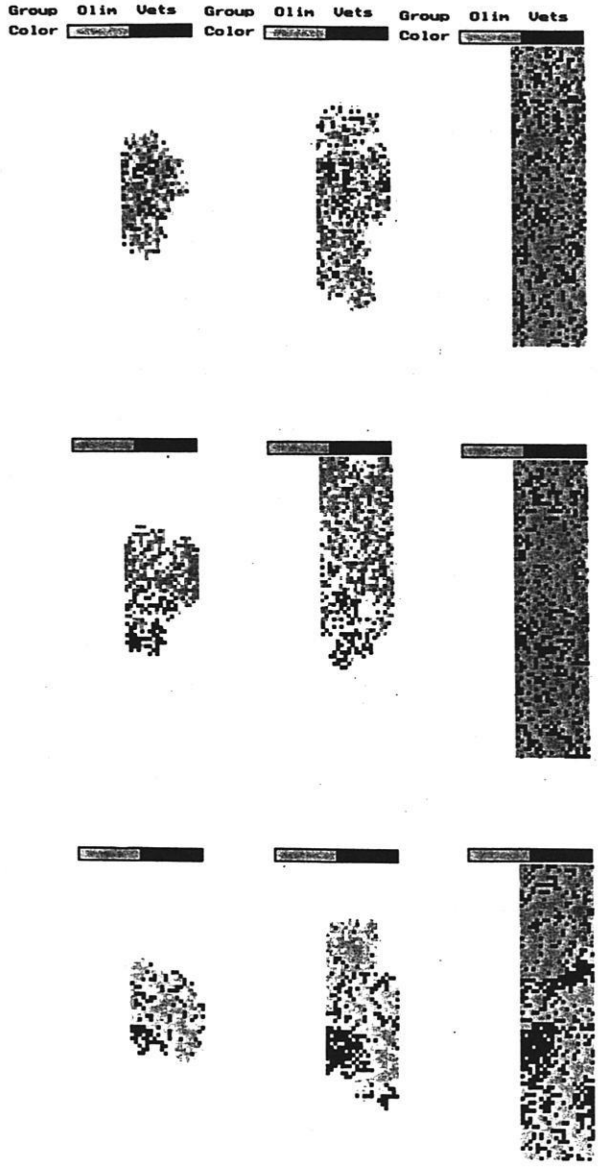
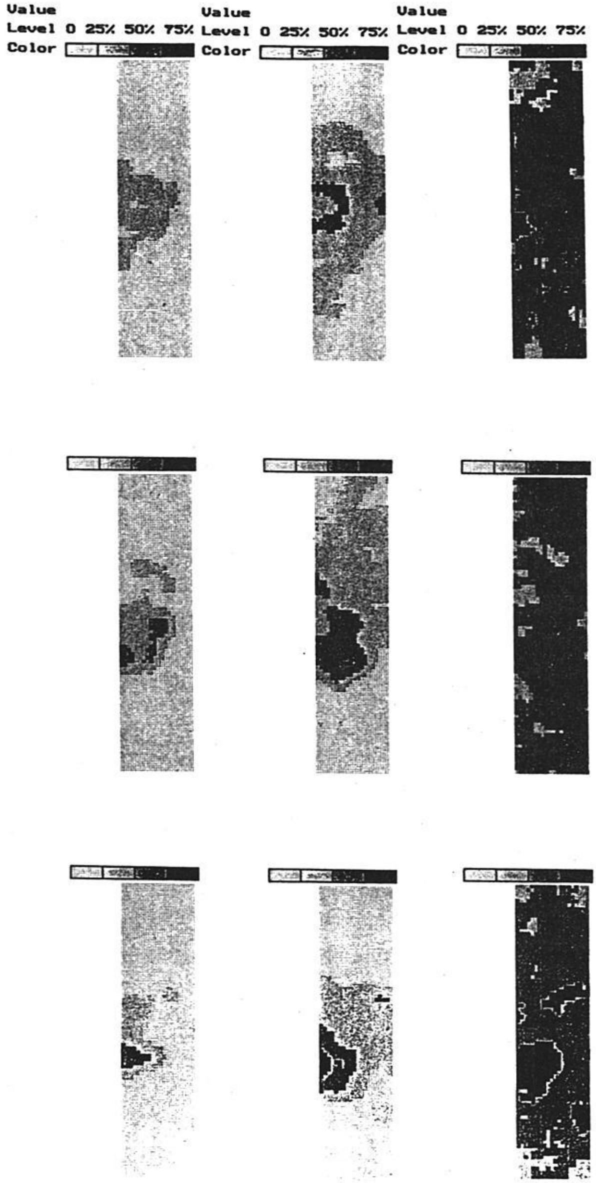


Figure 5 Dynamics of Land Value Maps



Economic interactions only

Cultural interactions only

Economic and Cultural interactions

t = 20 t = 40 t = 120

Appendix

CS model City-1 consists of *territory, transformation rules* , and *immigration-emigration* rules. Transformation rules define the properties and occupants of each house at the next time-step depending on properties and occupants of all the neighboring houses. Immigration-emigration rules define the flows of the individuals into- and out of the city.

Consider individual **P** with status S_p ($0 \leq S_p \leq 1$), tendency T_p ($0 \leq T_p \leq 1$), and origin O_p ($O_p \in$ {**Veteran, Ole**}). When **P** occupies a house/cell H_{ij} whose value is V_H ($0 \leq V_H \leq 1$), we denote an occupation form as F_p ($F_p \in$ {**Owner, Renter**}). The payment of an individual occupying house **H** is W_H ($0 \leq W_H \leq 1$). The status of an individual and the value of a house depend on time. We shall mark it explicitely, when necessary.

We consider as neighbors individuals and houses in the 5x5 square around a house, that is, individuals occupying houses from the set of

$$U(H_{ij}) = \{H_{kl} \mid \max(0, i-3) < k < \min(n, i+3), \max(0, j-3) < j < \min(m, j+3)\}$$

Below $S_{neighbors}$ is an average status of the individuals in the $U(H_{ij})$, $O_{neighbors}$ is a fraction of individuals in the $U(H_{ij})$, whose origin is the same as an origin of individual **P**, situated in the H_{ij}.

Each time step the sequence of decisions and parameters changes is calculated in the following way (equations' numbers correspond to the numbers of the boxes in the flow chart at the Figure 1):

Transformation rules

1. To change a house?

The probability for an individual **P** to leave house **H** depends on the occupant's origin, fraction of his/her neighbors whose origin differs from occupants' one and the absolute value of the difference between occupants' status and the status of his/her neighbors:

$$R_P = \alpha 1(O_P) + \beta 1(O_P) \exp(\gamma 1(O_p) \mathrm{Abs}(S_P - S_{neighbor})) + \alpha 2(O_P) + \beta 2(O_P) \exp(\gamma 2(O_P)(1 - O_{neighbor})) \quad (1.1.)$$

We suppose that R_P increases monotonously with the increase of **$\mathrm{Abs}(S_P - S_{neighbors})$ and $1 - O_{neighbors}$**

2. To buy a new house?

This deterministic condition is applied to queuing individuals who want to enter the city as well as to residents who want to change their houses. An individual **P** can buy house **H** of value V_H if his/her status and tendency are sufficiently high relative to the value of the house. That is

$$S_P > \delta_0 + \delta_1 V_H, \quad T_P > \tau_0 + \tau_1 V_H \quad (2.1)$$

where δ_0, δ_1, τ_0, τ_1 are constants.

An individual **P** chooses a house from the set of vacant houses satisfying conditions (2.1). The "attractivity" of a house **H** (**$Q_P(H)$**) is estimated as follows:

$$Q_P(H) = \alpha_3(O_P) + \beta_3(O_P)\exp(\gamma_3(O_P)S_{neighbors}) + \alpha_4(O_P) + \beta_4(O_P)\exp(\gamma 4(O_P)O_{neighbors}) \qquad (2.2)$$

We suppose that $Q_P(H)$ monotonously increases with an increase in the $S_{neighbors}$ and $O_{neighbors}$ and interpret "attractivity" as the probability to occupy (buy or rent) a vacant "house" H_{ij} when it is the only possible choice.

An individual occupies one of the houses satisfying (2.1) or fails to occupy any of the houses in the following way. Let us denote as Λ the set of the houses, satisfying (2.1), plus some special element H_O, corresponding to failure of the attempt to occupy. We define the probability p_{ij} to occupy house $H_{ij}\ \varepsilon\ \Lambda$ as

$$p_{ij} = s_{ij} / \sum_{H_{ij}\ \varepsilon\ \Lambda} s_{ij} \qquad (2.3)$$

where $s_{ij} = Q(H_{ij})s_0 / (1 - Q(H_{ij}))$, $s_0 = \prod_{H_{ij}\ \varepsilon\ \Lambda / \{H_0\}} (1-Q(H_{ij}))$

3. To rent a new house?

If an individual did not succeed to buy a house, he/she tries to rent one. In the same manner as above the individual chooses from the set of houses satisfying condition (3.1) which is similar to the second of conditions (2.1):

$$T_P > \tau_2 + \tau_3 V_H \qquad (3.1)$$

where τ_2, τ_3 are constants.

Each vacant house, satisfying condition (3.1), can be rented. The "attractivity" of the vacant houses is estimated according to (2.2). The probability to rent one of the houses **H** satisfying (3.1) is calculated according to (2.3).

4. Change of Status

Simultaneously with the above activities regarding spatial location, the individual's status changes every time-step. The status, at the next time-step, of individual **P** located at house **H,** depends on his/her current status, tendency and payment in the following manner

$$S_p(t+1) = \begin{cases} \max\{0, S_P(t) + T_P(t) - W_H(F_P(t))\} & \text{if } T_P(t) - W_H(F_P(t)) \leq 0 \\ \min\{1, S_P(t) + T_P(t) - W_H(F_P(t))\} & \text{if } T_P(t) - W_H(F_P(t)) > 0 \end{cases} \qquad (4.1)$$

where $S_P(t)$, $T_P(t)$ are defined as above, **t** denotes the number of the time-step.
We define payment $W_H(F_P(t))$ as the right parts of formulae (2.1) and (3.1):

$$W_H(F_P(t)) = \begin{cases} \tau_0 + \tau_1 V_H(t) & \text{for } F_P(t) = \text{owner} \\ \tau_2 + \tau_3 V_H(t) & \text{for } F_P(t) = \text{renter} \end{cases} \qquad (4.2)$$

Queuing individuals have no payment and consequently the changes in their status are determined as follows:

$$S_P(t+1) = \min\{1, S_P(t) + T_P(t)\} \qquad (4.3)$$

5. Change of Value

The value of a vacant house **H** at the next time-step is defined as the average of the status of individuals occupying houses or the value of the vacant houses in the neighborhood **U(H).** That is

$$V_H(t+1) = ((S_1(t) + S_2(t) + S_3(t) + ...) + (V_1(t) + V_2(t) + V_3(t) + ..))/25 \quad (5.1)$$

where **D < 1** is a decrement of value.

In the first pair of brackets we include status of individuals occupying houses in **U(H)**, and in the second pair the values of the vacant houses in **U(H)**.

Immigration-emigration rules

6. To leave the system?

A resident who could not buy a new house nor rent one, stays at his/her current house in the city or leaves the system. It is supposed that L_P, the probability to leave the system, is:

$$L_P = const \quad (6.1)$$

A new immigrant who failed to buy or rent a house either returns to the queue or else leaves the system.

7. To leave the queue?

An individual entering the queue is trying to occupy a house each time-step. If he/she did not succeed to occupy a house during some pre-determined time-steps ($T_{threshold}$), he/she leaves the system. Below

$$T_{threshold} = const \quad (7.1)$$

8. To leave the city?

This is a deterministic rule. An individual will leave the city's territory and enter the queue if

$$S_P(t) < S_{threshold} \quad (8.1)$$

where $S_{threshold}$ is a threshold status level.

9. Inflow of individuals

Each time-step a constant number of Veterans and Olim enter the city. Their initial status and tendency are assigned randomly and independently according to the given normal distributions with

$$mean = \mu(O_P), \quad STD = STD(O_P) \quad (9.1)$$

that is mean and average differ for Veterans and Olim.

BIBLIOGRAPHY

ALBIN, P.S. 1975. The Analysis of Complex Socioeconomic Systems. Lexington: Lexington Books.

ALLEN, P.A. 1981. "The evolutionary paradigm of dissipative structures", in E. JANTSCH ed. The Evolutionary Vision. Boulder: Westview Press.

ALLEN, P. and SANGLIER, M. 1981b. "Urban evolution, self-organization and decision making", Environment and Planning A, 13: 169-183.

ALLEN, P., SANGLIER, M., ENGEREN, G. and BOON, F. 1985. "Towards a new synthesis in the modeling of evolving complex systems", Environment and Planning B, 12: 65-84.

COUCLELIS, H. 1985. "Cellular worlds: a framework for modeling micro-macro dynamics", Environment and Planning A, 17: 585-596.

COUCLELIS, H. 1988. "Of mice and men: what rodent population can teach us about complex spatial dynamics",Environment and Planning A, 20: 99-109.

DENDRINOS, D.S. and SONIS, M. 1990. Chaos and Socio-Spatial Dynamics. New York, Berlin: Springer.

HAKEN, H. 1983a. Advances Synergetics. Berlin: Springer.

HAKEN, H. 1983b. Synergetics: An Introduction. Berlin: Springer.

HAKEN, H. ed. 1985a. Complex Systems - Operational Approaches. Berlin: Springer.

NICOLIS, G. and PRIGOGINE, I. 1977. Self-Organization in Nonequilibrium Systems: From Dissipative Structures to Order Through Fluctuations. New York: Wiley-Interscience.

PORTUGALI, J. and BENENSON, I. (submitted). Artificial planning experience by means of a cell-space model: simulating international migration in the urban process.

PORTUGALI, J., BENENSON, I. and OMER, I. (forthcoming). Socio-spatial residential dynamics, stability and instability within a self-organizing city, Geographical Analysis.

TOBLER, W.R. 1979. "Cellular geography", in S. GALE and G. OLSSON, eds. Philosophy in geography. Dordrecht: Reidel.

WEIDLICH, W. 1987. "Synergetics and social science", in R. GRAHAM and A. WUNDERLIN, eds. Lasers and Synergetics. Berlin, New York: Springer.