MODELLING POPULATION DENSITY IN THE BALTIC STATES USING THE DIGITAL CHART OF THE WORLD AND OTHER SMALL SCALE DATA SETS

J. Sweitzer[1] and S. Langaas[2]

[1] The Beijer International Institute of Ecological Economics, Box 50005, S-104 05 Stockholm Sweden
[2] GRID-Arendal, c/o Dept. of Systems Ecology, S-106 91 Stockholm, Sweden

Abstract. *Using a Geographic Information Systems approach to analyse the Baltic Sea drainage basin allows us to rationally merge economic and ecological data and thus aid in developing efficient coastal management policies. The objective of this study is to develop a model which can estimate population density within the Baltic Sea drainage basin area. The three Baltic States, Estonia, Latvia, and Lithuania, were used as a test area to develop the initial population density model. The model was developed using sub-national population statistics together with digital data on land use and infrastructure. Digital data were compiled from various sources including ESRI's Digital Chart of the World and the European Space Agency's Remote Sensing Forest Map of Europe. We use a single objective, multi-criteria model which assumes that settlement is influenced in varying degrees by land cover, road and railroad infrastructure, and the location of urban areas. Output from the model is in the form of a population density map. We conclude that the population density model is an improvement over other methods used to analyse population density in this region and that it is particularly useful for estimating the population of a drainage basin area.*

INTRODUCTION AND OBJECTIVE

Human activity in the drainage basin strongly impacts the Baltic Sea (Ambio, 1990; Helsinki Commission, 1992). Despite this, no coherent source of information exists which provides the population distribution within the Baltic Sea drainage basin. Population statistics at the county or municipality level are available for all countries within this region. These data however, provide no information on the population within the drainage basin in the cases where the drainage basin only partially covers the administrative region.

The objective of this study is to estimate population density in the Baltic Sea drainage basin. Using a Geographic Information System (GIS) approach, we have developed a model which can be used to estimate population density (persons/km^2). The three Baltic States, Estonia, Latvia, and Lithuania, were used as a test area to develop the initial model. We use a single objective, multi-criteria model which assumes that settlement is influenced in varying degrees by land cover, road and railroad infrastructure, and the location of major urban areas. Output from the model is in the form of a digital population density map which can be overlaid with a digital drainage basin map in order to assess population density within the nine meso-scale sub-drainage basins which cover the three Baltic States. This paper describes the data and methodology used to develop the population density model and presents the results of the model.

PRIMARY DATABASES

The population density model was developed using IDRISI (version 4.1), a raster-based GIS software package. The model uses raster data at a one square kilometre resolution. Model inputs include population statistics and six raster maps: administrative regions, proximity to roads, proximity to railroads, proximity to major cities, constraint areas, and land cover. The land cover map consists of the following categories: forest, open lands, urban areas, lakes, land-lake interface, and glaciers. The input maps were created by compiling data from several sources which are briefly described in this section.

Digital Chart of the World (DCW)

Much of the data for this study were extracted from the Digital Chart of the World (DCW) database, a 1:1,000,000-scale vector basemap of the world produced by Environmental Systems Research Institute, Inc. (ESRI). The DCW consists of 17 data layers related to the following themes: national and ocean boundaries, populated areas, railroads, roads, utilities, drainage, hypsography, land cover, ocean features, physiography, aeronautical features, cultural landmarks, transportation structure, and vegetation. The primary data source for the DCW was the Operational Navigation Chart (ONC) series, a group of 1:1,000,000 scale paper maps created by the United States Defence Mapping Agency. These paper maps were scanned, vectorised, and edited by ESRI and are available in ARC/INFO format. For a detailed description of the production process, see ESRI (1992).

Sweitzer, J. and Langaas, S. 1995. Modelling population density in the Baltic Sea States using the Digital Chart of the World and other small scale data sets. In Gudelis, V. Povilanskas, R. and Roepstorff, A. (eds.) 1995. Coastal Conservation and Management in the Baltic Region. Proceedings ot the EUCC -WWF Conference, 2-8 May 1994, Riga - Klaipeda - Kaliningrad., pages 257-267.

The ONC charts were designed to assist pilots flying at medium or low altitudes and are also used for U.S. military operational planning, intelligence briefings, and preparation of visual cockpit displays (DMA, 1981). The DCW coverages show some bias due to the navigational and military slant of its source data. In general, ground features with distinctive and consistent visual patterns such as roads, railroads, utility lines, lakes, rivers, and the perimeter of urban areas are highly detailed and well represented in the DCW. Features with less distinct or inconsistent borders such as wetlands or cultivated areas, have little significance to a military pilot and thus are not well represented in the DCW. In addition, tertiary roads as well as urban roads and railroads are not included in the DCW. At the scale of this study, we felt the military-navigational biases of the DCW data did not compromise the quality of the roads, railroads, populated areas, and lakes coverages. All the other classes we had expected to extract from the DCW, such as wetlands, cultivated areas, and vegetation, were inconsistent or inaccurate and therefore rejected for use in this study.

ESA Remote Sensing Forest Map of Europe

The European Space Agency (ESA) Remote Sensing Forest Map of Europe was a contribution to the World Forest Watch project of the International Space Year 1992 (ISY 92). The main objective of the work was to develop a map depicting the forest areas of Europe. The primary data source for the forest map was 1km^2 satellite data from the National Oceanic and Atmospheric Administration (NOAA-11) Advanced Very High Resolution Radiometer (AVHRR) satellite sensor system. A total of 72 different AVHRR images were used, the majority being from 1990-92. Only images from the summer months of June, July and August were applied. Images were classified into forest /non-forest using a supervised classification approach (Bayesian maximum likelihood) and various ancillary map information. An accuracy assessment was carried out by comparing the classified AVHRR images to classified Landsat MSS images for 32 randomly selected 100x100km areas within Europe. The summed overall accuracy (OA)parameter for all 32 test areas were 82.5%. Further details about the procedure used and accuracy assessment are given in ESA (1992) and GAF (1992). The final digital forest map, available as public domain data, is a vector ARC/INFO coverage to which major lakes were added.

Administrative Regions

A digital map of administrative units at the same level as the population statistics is required for the model. In the three Baltic states, the most detailed population statistics available are at the district level. District maps for these countries were obtained from three different sources. Digital maps were provided by the Estonian Environment Information Centre and the Latvian Ministry of Regional Development and Environment Protection. The Lithuanian district borders were digitized from a 1:600,000 scale paper map (Valstybinis Zemetvarkos Institutas, 1990). The three digital maps were merged to form a single district-level administrative region map covering the Baltic States.

Population statistics

Population statistics for Estonia, Latvia, and Lithuania were obtained from national statistical yearbooks (Statistical Office of Estonia, 1992; State Committee for Statistics of the Republic of Latvia, 1992; Valstybinis Zemetvarkos Institutas, 1990). District-level urban and rural population statistics from 1989 were used for this study.

DATA PRE-PROCESSING

All the data layers were edited and rasterized in ARC/INFO, then converted into IDRISI format. Further processing was required in IDRISI to generate the land cover map. The land cover map was created by overlaying three coverages: lakes, urban areas, and the rasterized ESA forest map. The final land cover map has five classes: urban areas, water bodies, land-water interface, forest, and open lands (representing both agriculture and pasture lands). Land-water interface is an artefact of overlaying two different data sources. These pixels are found along the lake shores and coastlines, where the DCW data shows more detail than the ESA forest data. We consider these pixels to represent non-urban land since that is how they are classified in the DCW data and we take DCW as the more correct source due to its greater detail. We have no indication whether these pixels represent open land or forest however, since in the ESA forest data they are classified as water.

METHODOLOGY: SPATIAL MODELS TO DERIVE POPULATION DENSITY

Without having the co-ordinates of each individual, the only feasible way to determine spatial population distribution is through modelling. Most population density maps represent the average population density for a statistical unit such as a country, a district, or a town. These maps can be considered the simplest "models" of the distribution of population density. When the intention is to determine the population within a region which does not coincide with the statistical unit boundaries however, this method is insufficient. We have attempted to develop a model which can be used in such cases. Our approach uses ancillary spatial information to more realistically dissaggregate the population of a statistical unit within its boundaries. The model works independently with urban and rural populations. The resulting urban and rural population density maps are then combined to produce a complete population density map.

Modelling of Urban Population Density

To model the urban population, we assume equal population density at the district level. Urban area is calculated by summing the urban grid cells ($1 km^2$) within each district boundary. Urban population is divided by urban area and the resulting value, population per km^2, is assigned to each urban grid cell of the appropriate district. The result is a map of urban population density.

Although assuming equal urban population density may seem simplistic, a number of studies have found linear relationships between urban areas and population (reviewed in Stern 1985). To assess whether a linear relationship was appropriate to use in the Baltic case we examined the relationship for a number of sample towns in the three Baltic states. The population statistics for these towns were taken from national statistical yearbooks, and their areas were derived from the land cover map. The relationship between population and area for these towns is shown in Figure 1. As can be seen, although not a perfect relationship, the linear model holds quite well.

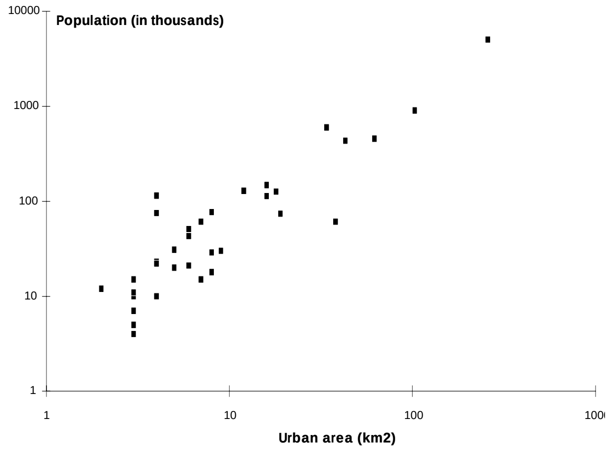


Figure 1. Relationship between urban population and urban area in the Baltic States

Modelling of Rural Population Density

To model rural population density we use the Single Objective, Multi-Criteria Evaluation (MCE) approach (Eastman 1993, Eastman et al. 1993). The MCE approach determines the *suitability* of a tract of land for a specific purpose given some criteria. Criteria are divided into *constraints* and *factors*. Constraints are areas which have no suitability. Factors are defined as spatial features which influence the suitability of a tract of land for the given objective. Each factor is assigned a *weight* according to its relative importance. A prerequisite of the MCE analysis is that constraint maps are binary (0/1) and factor maps are standardised to a common range of values. The relationship can be expressed mathematically by Equation 1:

$$S = \sum w_i x_i \times \prod c_j \qquad (1)$$

where

S = suitability

w_i = weight of factor i

x_i = value of factor i

c_j = value of constraint j

In the derivation of the weights for the entire study area it is a precondition that $\Sigma w_i = 1$. We use the MCE approach to determine the "suitability" for rural settlement of each 1 km^2 pixel of a Baltic States map. This map is then used together with population statistics to generate a rural population density map. This study is meant to be descriptive, not prescriptive. We use the term "suitability" here to mean the *likelihood* of finding rural people.

The constraints The constraints were defined from the land cover map to be (i) urban areas and (ii) lakes. If data resources were available, we also would like to exclude (iii) wetlands and (iv) restricted-settlement areas (National parks, etc.).

The factors In the MCE approach, factors are input as raster maps standardised to a common range of values. There are several steps involved in the choice and preparation of the factor inputs: (1) selecting factors, (2) preparing factor maps (determining the range of influence and the standardisation procedure), and (3) assigning relative importance (weights) to the factors. Each step requires subjective decision making. In this case study we completed the first two steps ourselves. For the last step we relied upon the opinions of six researchers from the Baltic States (two from each country). The six persons had previously been exposed to the MCE analytical approach and were considered to have above-average national knowledge of the relationship between various factors and the rural population density.

Selection of factors Given our data resources we decided that the following features were related to the distribution of the rural population: (1) proximity to roads, (2) proximity to railroads, (3) land cover, and (4) proximity to major cities (with population >200,000). The selection of factors (1) through (3) are based upon intuitive geographic knowledge. Proximity to major towns have elsewhere been found to be a significant determinant of distribution of population (Eklund and Deichman 1991). Others factors considered, but rejected by us for the Baltic States were (i) proximity to major lakes, (ii) proximity to major rivers, and (iii) proximity to the coastline.

Preparation of factor maps Proximity to Roads and Proximity to Railroads: We assumed that a likely "zone of influence" of both roads and railroads was a five kilometre range. If the roads map had been perfect (all roads included) it would have been more appropriate to assign a shorter range of influence to roads. The influence roads and railroads have on rural settlement was assumed to decrease with distance. We determined that being close to infrastructure should count more than being farther away. Thus, for the 0-5 km range of influence around each road or railroad, we applied a power function. We decided that the influence of the infrastructure would decrease by a factor of $\mathbf{d}^2$, where $\mathbf{d}$ = distance away from infrastructure. Beyond five kilometres from any given road or railroad, the influence on rural settlement is zero. For the factor maps, we chose to use a nominal standardisation value range of 100-0. We converted the actual physical distances in the proximity maps into a continuous range of values with the highest suitability pixels given the value 100 and the lowest the value zero.

Proximity to big cities: By examining a map of average rural density/km^2 for all the districts in the Baltic States, we found 0-100 km to be a suitable range of influence of major cities. Within this range, we assumed that influence decreases linearly with distance from the city. We then standardise the proximity values so that grid cells less than one kilometre away from the major cities are given the value 100 and cells 100 km or beyond the cities are given a value of zero.

Land cover: All the non-constraint classes of the land cover map (forest, open lands, and land-lake interface) are considered to have varying degrees of influence on rural settlement. The standardisation procedure in this case involves transforming these classes into a "continuous" range of values between 100-0. The standardisation value given to each land class represents the relative degree of influence that class is assumed to have on rural settlement. We decided to give the highest value (standardisation value of 100) to the open lands class which includes both agricultural and grazing land. The urban and lake classes, which are defined as constraints, are given a standardisation value of zero. The forest class is given the lowest value (standardisation value of 10). This represents our assumption that the probability of having rural population in the Forests is 1/10th the probability of having rural population in the open lands. The land-lake interface class (unclassified rural land) is believed to have an equal probability of belonging to either the forest or the open land class. We therefore give it a standardisation value of 55 $[(SV_F + SV_O)/2]$.

Weighting of factors In this step the six researchers from the three Baltic states were asked to weigh the relative importance of the four factors. All six stated that settlement patterns in the three countries were similar and did not feel that separate weights should be determined for each country. The weighting process is subjective. It was carried out through pairwise comparisons between the four factors, which means that six pairwise comparisons were made by each researcher. The weighting scale used consists of nine qualitative terms that are associated with nine quantitative values (see below).

EXTREMELY	VERY STRONGLY	STRONGLY	MODERATELY	EQUALLY	MODERATELY	STRONGLY	VERY STRONGLY	EXTREMELY
LESS IMPORTANT					MORE IMPORTANT			
1/9	1/7	1/5	1/3	1	3	5	7	9

The quantitative values were then entered in to a Pairwise Comparison Matrix (PCM) which facilitates calculation of each factors' relative weight (Eigen vector) and a Consistency Ratio which tells if the researcher's opinions were internally consistent (see Eastman *et al.* 1993 for details). In our case, the opinions of two of the Baltic researchers were not internally consistent and thus rejected. We therefore averaged the factor weights given by the four remaining researchers and entered these into the PCM. A Consistency Ratio of 0.08 was obtained which signifies a high degree of consistency. The relative weights for the various factors were calculated to be (in decreasing order of importance):

proximity to major cities = 0.3737
proximity to roads = 0.2916
land cover type = 0.2492
proximity to railroads = 0.0855

These weights were then entered into the Multi Criteria Evaluation (MCE) module. The output of the MCE is a rural population suitability map. In this map the highest suitability value, 100, is assigned to grid cells that are on open lands lying less than 1km from the roads, railroads and major towns. The lowest suitability value is given to grid cells on forest land lying further than 5 km away from roads and railroads, and more than 100 km away from major cities. Cells covered by constraint areas have a suitability of zero.

Using the suitability map, we can determine the proportion of total rural population likely to occur in each 1km^2 grid cell. This proportion is represented by the ratio: $\mathbf{s_i/\sum s_i}$ where s_i is the suitability of grid cell$_i$. Rural population density per square kilometre can then be estimated using Equation 2:

$$\mathbf{RPD_i = s_i/\sum s_i * P_r = s_i * P_r/\sum s_i} \qquad (2)$$

where RPD_i = rural population in grid cell i (= rural population /km^2)
s_i = suitability of grid cell i
P_r = total rural population

The sum of suitability values ($\sum s_i$) for all rural grid cells within each district was calculated by overlaying the administrative map with the suitability map and summing the suitability values within each district boundary. Rural population of each district was divided by the appropriate summed suitability value ($P_r/\sum s_i$). Each cell in the suitability map (s_i) was then multiplied by this district-specific ratio to generate a rural population density map. Finally, the urban and rural population density maps were merged.

RESULTS AND DISCUSSION

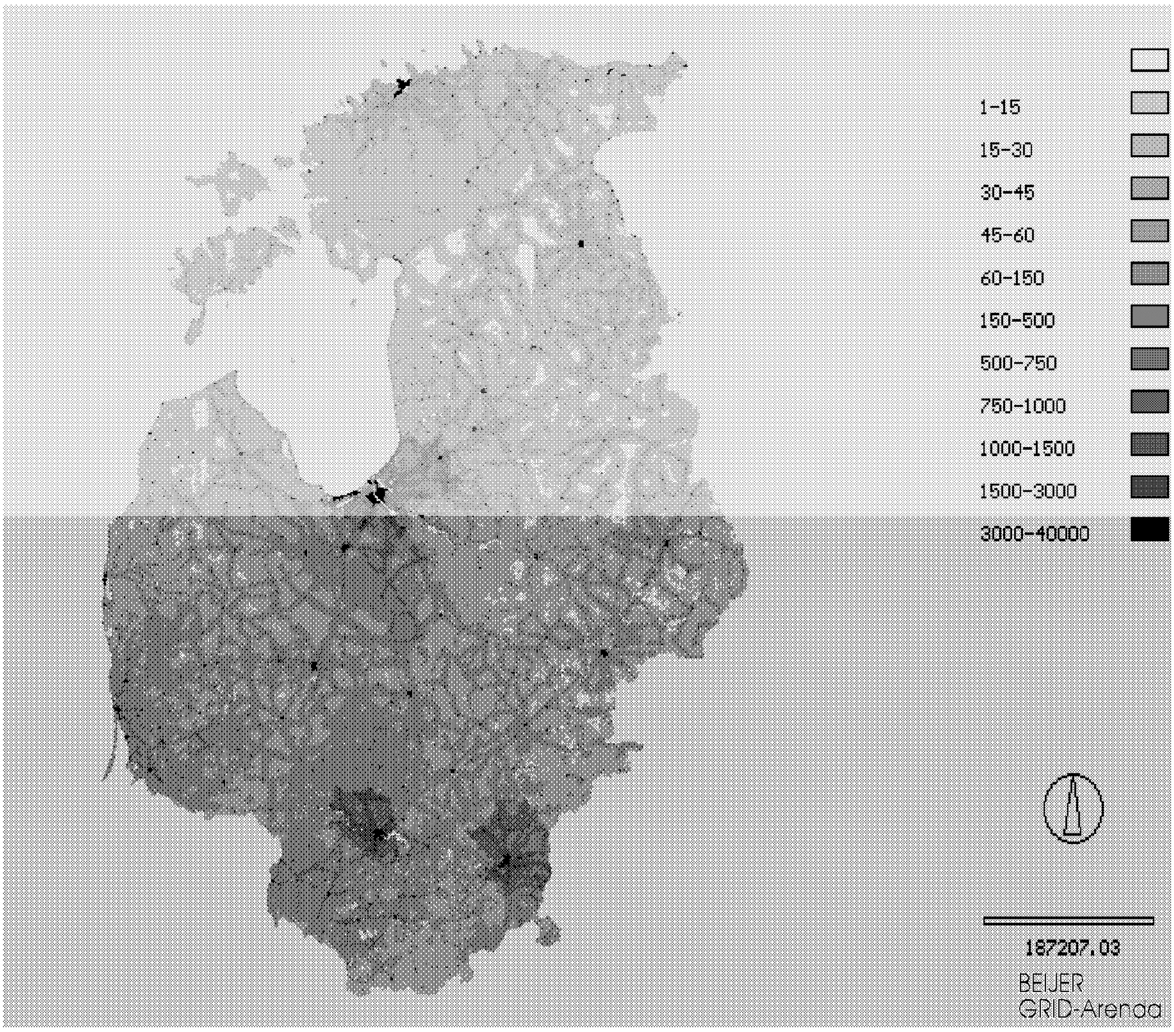
The population density map derived from the model is shown in Figure 2. A visual assessment shows that the model produces reasonable results. Quantitatively validating the results of the model is difficult. To assess the accuracy of our results, we need access to higher quality population density data for this region. To the best of our knowledge, such data is not currently available in the Baltic States. We compared our results to a 100 km^2 resolution population density map of Latvia and found similar trends in the data. The methodology used to create the Latvian population density map however, did not lead us to believe that it's results were more reliable then our own (Baranovs, pers. com.). A population density map of Estonia which uses some data obtained by geo-coding persons is currently under production (Roosaare, pers. com.). When this map is available, it will be used to assess the results of our model. We are continuing to search for data sources which can be used to validate the model.

The population density model has several weaknesses which limit its predictive capabilities. The model relies heavily on subjective decisions, the method used to assess urban population is simplistic and the model results are impaired by error sources in the data. There is no way to judge the reliability of the population statistics, for example. Additionally, we know that some of the GIS data inputs have flaws. The roads coverage is not complete and the land cover map is highly generalised. Nevertheless, it provides an improvement over previous attempts to determine population density in this area.

CONCLUSIONS

Despite its weaknesses, the population density model offers several important contributions. It is an improvement over the traditional method of analysing population density which is to assume even population distribution over the area for which statistics are available. This is of particular significance in drainage basin areas which don't necessarily coincide with administrative district boundaries. Furthermore, because the methodology is consistent, the model can be easily improved as more detailed input data becomes available. This is an advantage over traditional cartographic methods of mapping population distribution. Finally, the model is flexible. It can easily be applied to heterogeneous areas, such as the coastal zone of the Baltic Sea, where the factors influencing population settlement differ by region. As more detailed GIS data becomes available, we will continue to improve this model by addressing the weaknesses and providing more rigorous verification studies. Future research plans involve modelling the population density for each country within the Baltic Sea drainage basin and using the model output to estimate population density within 87 of the meso-scale sub-drainage basins of the Baltic Sea.

Figure 2. Population Density of the Baltic States (persons/km^2)



ACKNOWLEDGEMENTS

We gratefully acknowledge the contributions of Harijs Baranovs, Jaak Jaagus, Karlis Kalviskis, Linas Kliucininkas, Arturas Razinkovas, and Jüri Roosaare. This research was carried out within the framework of the European Union DG XII-D (Contract EV5V-CT92-0183) and funded by the Swedish Environmental Protection Board, the Swedish Council for Planning and Cooridnation of Research (FRN), and the Norwegian Research Council. Partial funding came from The Gulf of Riga Project. The GIS dataset developed for this study will be available as public domain data through GRID-Arendal (Contact S. Langaas, address noted above).

REFERENCES

Ambio. 1990. "Current Status of the Baltic Sea" Special Report Number 7. September, 1990.

Baranovs, Harijs. 1994. Ministry of Regional Development and Environment Protection, Jurmala, Latvia. Personal Communication.

Defense Mapping Agency (DMA). 1981 "Product Specifications for Operational Navigation Charts Scale 1:1,000,000". (First Edition January, 1981 with updates through 1987) PS/1AB/120. DMA Aerospace Center, St. Louis, MO.

Eastman, J.R. 1993. Decision theory and GIS. In: Proceedings AfricaGIS'93

Eastman, J.R., P.A.K. Kyem, J. Toledano and W. Jin. 1993. GIS and Decision Making. Vol. 4 in Explorations in Geographic Information Systems Technology, UNITAR, Geneva.

Eklundh and Deichman, 1991. "Estimated Population Densities for Africa" Part 6 in *Global Digital Datasets for Land Degradation Studies: A GIS Approach.* GRID Case Study Series No. 4, UNEP/GRID, Nairobi.

Environmental Systems Research Institute, Inc. (ESRI). 1992 "The Digital Chart of the World for use with ARC/INFO Data Dictionary" Redlands, CA

European Space Agency (ESA). 1992. Remote Sensing Forest Map of Europe. ESA/ESTEC, Nordwijk, The Netherlands. 18 pages + map.

Gesellschaft für Angewandte Fernerkundung (GAF). 1992. Feasibility study of Regional Forest Inventory Maps of the ISY - "World Forest Watch" Project. Final Report prepared for ESA/ESTEC. 80 pages + annex.

Helsinki Commission. 1992. "Background Document for the Baltic Sea Environmental Declaration, 1992." Diplomatic Conference on the Protection of the Marine Environment of the Baltic Sea Area. Conference Document No. 5/2.

Roosaare, Jüri. 1994. Department of Geography, Tartu University, Tartu, Estonia. Personal Communication.

State Committee for Statistics of the Republic of Latvia. 1992. "Statistical Yearbook of Latvia" Riga, Latvia.

Statistical Office of Estonia. 1992. "Estonia: A Statistical Profile" Tallinn, Estonia.

Stern, M. 1985. Census from heaven? Population estimates with remote sensing techniques. Meddelanden från Lunds Universitets Geografiska Institution, Avhandlingar XCIX.

Valstybinis Zemetvarkos Institutas. 1990 "Lietuvos Respublikos Administracinis Zemelapis" 1:600,000. Respublikinis Zemetvarkos Projektavimo Institutas. Vilnius, Lithuania.