Integration of GIS Data in Health Care Utilization

Georg Romstorfer[1,*], Stephanie Parragh[1], Günter Schneckenreither[1,2],
Michael Landsiedl[2],Patrick Einzinger[2], Martin Scheuringer[3]

[1]Vienna University of Technology, Institute for Analysis and Scientific Computing, Wiedner Hauptstraße 8-10,
1040 Vienna, Austria; *georg.romstorfer@tuwien.ac.at
[2]DWH Simulation Services, Vienna, Austria; [3]Main Association of Social Security Institutions, Vienna, Austria

Abstract The process of choosing medical providers for consuming health care services, highly depends on local properties like medical provider (e.g. physicians and hospitals) and population density or distances between single agents of these groups. As a consequence data containing geographical information has to be used. Obtaining, pre-processing and integrating this data into a dynamic model turns out to be a challenging task.

We present an agent-based dynamic model of Health Care Utilization that uses geographical data from two sources: Map data from OpenStreetMap for constructing the GIS-environment and raster data on population density from Statistics Austria. The presented model is held relatively simple as the main focus lies rather on the analysis of GIS data integration with agent based models of health care than on a valid mapping of health care utilization.

GIS data comes in various formats and one has to decide on a general format that can be used all over the model. We integrated all our data into a PostGIS database, where it can be easily maintained as all data can be found at a single location (in contrast to solutions with several data files). AnyLogicTM was used for simulation, but does not provide any means of displaying GIS information from a PostGIS database. As a consequence a class for displaying this information on a map had to be developed. It uses GeoTools for accessing the GIS enabled database.

As a result we show differences when modelling with and without GIS Data and give an outlook on future possibilities of simulation with geographically enhanced data.

Introduction

Modelling health care utilisation is a necessary task if one wants to raise the efficiency of hospitals or get information on the regional health care supply. Incorporating GISData into these models makes them more accurate but also adds more complexity.

Is it really worth the effort? There is no general answer to this question but the decision has to be made for each individual case. This paper will give a detailed view on which tasks have to be done to include GIS-Data into an agent-based model and show the differences compared to a model without geographical data. The used model is just a test model of patients utilising hospitals with simplified rules. These patients want to increase their life quality (e.g. get healthy) while hospitals want to maximise their performance (e.g. income, utilisation, patients' satisfaction).paragraph of a section or of a subsection, or for a paragraph with paragraph heading ('subsubsection'), in special cases in paragraphs directly after listings or figures, if it improves readability.

1 Model Description

The model is agent-based with two types of agents: patients and hospitals.

1.1 Patients

The health of the patients can vary from one to the other of the following states: healthy, sick1 or sick2. There are two different treatments which can cure these diseases. Treatment A cures disease 1 and treatment B cures disease 2 and if the patient is sick of disease 1 it doesn't cure him but it makes him feel better. Each patient's life quality is measured by 1 if he is healthy, which decreases if he gets sick or does not get any treatment and increases if he gets an appropriate treatment. If a patient gets sick he puts himself on the appointment list of the hospital and waits for his treatment.

The patient checks every week, if he gets some treatment and if not his satisfaction with the hospital decreases. The satisfaction decreases too, if the patient gets wrong treatments and therefore has to stay longer in hospital.
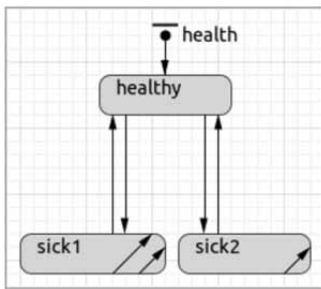
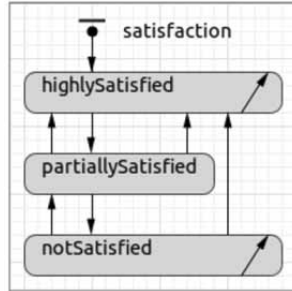**Figure 1**. Patient's health states.

Figure2. Patient's satisfaction states.

The data of the life quality of each patient is continually saved. Periodically each patient's mean life quality of the last period is compared to the life quality threshold. If the mean life quality falls under this threshold the patient becomes less satisfied and if the mean life quality gets much higher than the threshold he becomes more satisfied.

If the patient stays dissatisfied for a longer time he chooses another hospital. It is also possible, that the patient recommends a hospital to others if he his highly satisfied for a longer period.

Choosing hospitals. If a patient chooses a new hospital because he is not satisfied with his old one, he cancels his appointment and deletes his connection to the hospital.For each hospital the patient calculates a weight according to the following formula:

$$weight = \frac{(recommandations + 1) \cdot \ln(unitsPerWeek + 1 - treatments)}{(distance + 1)^2 \cdot \ln(unitsPerWeek + 1)}$$

Variable	Meaning
recommendations	number of patients that recommend that hospital
unitsPerWeek	number of patients thatcan be trated per week
distance	Distance to the hospital
treatments	Number of patients that get treatment in the hospital

If the distance is greater than a maximum distance the weight is divided by 1000 as a penalty. For the simulation with GIS-data the maximum distance was set to 60km and for the simulation without GIS-Data it was set to 100px.

A conclusion of Urach et al. [1] is that the willingness of patients to participate in a rehabilitation programme is inverse proportional to the *distance* to the rehabilitation centre. In our case there is a similar situation just with hospitals instead of rehabilitation centres. The result of testing out some reasonable inverse proportional functions was that inverse quadratic dependence of distance seems to be a good choice.

The ratio *unitsPerWeek* to treatments represents the utilisation of the hospital. If the hospital is running at normal operating capacity the ratio barely influences the weight but if the hospital is almost full (e.g. *treatments* equal *unitsPerWeek*) the weight decreases rapidly to zero. This dependence is best described by a logarithmic function of these variables.

Each *recommendation* increases the weight of the hospital. More recommendations lead to a higher probability of the hospital. After calculating these weights they are divided by their sum to get probabilities which sum up to 1. Then a hospital is randomly chosen with the calculated probabilities.

1.2 Hospitals

Each hospital maintains a list of appointments which holds patients waiting for treatment and a list *prolongedStays* which holds patients that have already gotten some treatment, but are not yet cured and therefore have to stay longer in hospital.

Hospitals are initialized with a certain number of beds from which the variable *unitsPerWeek* is calculated. This variable holds the number of patients that can be treated per week. The event *treat* is called *unitsPerWeek* times per week. Patients in the list *prolongedStays* are preferred over patients in the list *appointments*.

When the hospital treats a patient it has at first to decide on the type of treatment. Since the diseases have very similar symptoms, the hospital only knows a probability for the diseases. This probability is retrieved from a triangluar distribution dependenton the actual patient's health state (Figure 3). Zero represents the actual patient's disease whereas 1 represents the other disease.

Then the hospital compares this probability to a threshold specific for each hospital. If the probability is greater than the threshold, the patient gets the wrong treatment else the right one. The health of the patient changes according to the health-statechart afterwards.
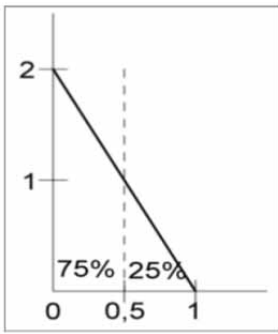
Figure 3. Triangular probability distribution for diseases with threshold at 0.5

The hospital adds the costs of the treatment to its income, increases its number of treatments and also increases the number of successful treatments, if the patient was cured.

Periodically the hospital adjusts the threshold for choosing treatments. The goal for the hospital is to find a good balance between successful treatments of the patients, high income and high utilisation. For accomplishing this the hospital calculates the following values for the adjustment period:

$$treatmentPerformance = \frac{successfull\ Treatments}{total\ Treatments}$$

$$utilizationPerformance =$$
$$= \frac{totalTreatments}{unitsperweek \bullet adjustementIntervall}$$

$$incomePerformance =$$
$$= \frac{income}{total\ Treatments \bullet \max(treatmentCosts)}$$

Each of these values lies between 0 (bad performance) and 1 (good performance). The average of these values represents the *performance* of the hospital.

By comparing this performance to the previous one, the hospital decides how to change the threshold for choosing the treatment type. Initially the hospital increases the threshold by a specified amount until the performance of the last period is greater than the actual performance. Then it starts decreasing the threshold until the last performance is again greater than the actual one etc.

2 Obtaining GIS-Data

From the model we now know that distances between patients and hospitals have to be calculated when patients choose a hospital. For this calculation their positions have to be known. The 'Main Association of Austrian Social Security Institutions' [2] provided the addresses of public hospitals.

Positioning patients is the second problem. First of all it's not possible to get the address of every Austrian inhabitant, because of data protection. Secondly the model would take too much time to process about eight million patients. The solution was to use the MGI-Lambert raster [3] with associated population data [4] in combination with the permanent habitable area [5] to position one agent for every 1000 persons in the intersection area of a raster cell with the permanent habitable area. The map [6] where hospitals and patients are displayed was obtained from Geofabrik GmbH [7] which offer dumps of the OpenStreetMap [8] database.

3 Preprocessing GIS-Data

The obtained data came in very different formats listed in the following table:

Data	Format
Hospitals	Excel - Sheet
MGI Lambert raster	Shapefile
Population data	CSV - File
Permanent habitable area	Shapefile
Map	OSM – Dump (XML)

The decision was made to put all this data into a PostGIS [9] database. PostGIS is an extension to PostgreSQL [10] that adds support for geographical objects. So all that data had to be converted into that format.

The geographic coordinates for each of 126 hospitals was retrieved from the webservice Nominatim [11] and stored together with the address, number of beds and the name of the hospital in the database. When dealing with geographical data, not only the data format matters but also the coordinate reference system (CRS) in which the data is stored.

These coordinate reference systems describe how coordinates are assigned to points on the earth surface, which earth model is used (e.g. sphere, ellipsoid, …), which units are used (e.g. meters, degrees, miles, …) and how these coordinates get projected onto a map.

All the data was converted into the WGS84 [12] CRS that stores positions as pairs of latitude and longitude. For the display of our data the WGS 84 / Pseudo-Mercator [13] CRS was used, because it is widely used by web mapping and visualisation applications. The shapefiles were loaded into our PostGIS database and converted to WGS84 with opensource tools (Figure 4 and Figure 5).

The population data was imported into the database by SQL-commands which created a new table that contains the geographic raster cells associated with the population data, so that each row in this table holds the geographic boundary of the raster cell and the population that lives in that cell. Figure 4 shows each raster cell colored by its population size. With the map we processed as follows: At first an area of the European map was cut out large enough to contain the whole Austrian boundary. Afterwards the created OSM-file was imported into our PostGIS database. Reducing the data to the administrative boundaries was the last step and resulted in the actual data of interest (Figure 6).
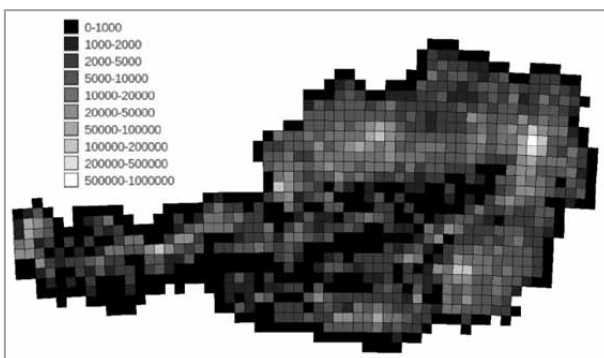

Figure 4. Raster with population density.

4 Integrating GIS-Data into the Model

The model was implemented in AnyLogic™. So this part will show how to integrate GIS-Data in that simulation environment. AnyLogic™ has some basic support for GIS-environments which means that agents can be positioned by latitude and longitude, distances can be calculated and shapefiles can be displayed as a map.

Unfortunately there is no support for PostGIS databases. The GIS-environment of AnyLogic™ works satisfactorally, so it was reused. The major problem was the map, that can just display shapefiles. Fortunately the map class *ShapeGISMap* derives from the abstract class *AbstractShapeGISMap*.
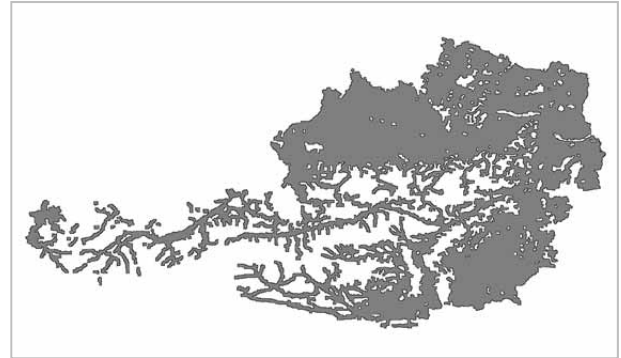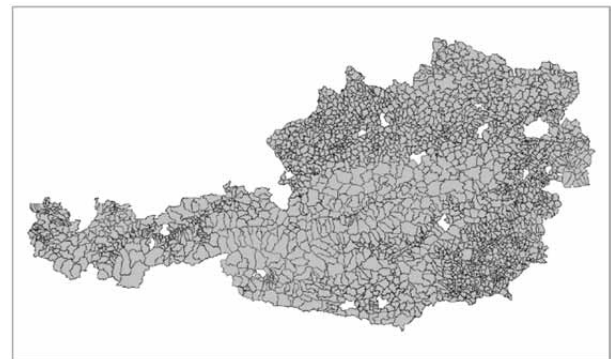

Figure 5. Permanent habitable area.

Figure 6. Austrian municipalities from OpenStreetMap.

So we derived our own map class *ShapeGeoTools-Map* that uses GeoTools [14] to access the GISData in the database. We also implemented the functions for calculating the distance between two points, zooming and panning and added some mouse event handling routines. The result is a fully integrated map that can be navigated by mouse like anybody is used to from well known web mapping applications. We added layers of the raster, the permanent habitable area and the Austrian municipalities. During simulation one can switch between these maps. The hospitals were initialized with their positions and the number of available beds, which we simply read from the database using the AnyLogic™ database connector. For each raster cell we uniformly positioned one agent per 1000 people in the intersection of that cell with the permanent habitable area. That gives us 7971 agents.

5 Calculating Distances

When calculating distances one has to take care which distance is meant: cartesian, geodesic or the length of the way from one point to the other. For our problem when calculating distances from a patient to hospitals

the way-distance would be most precise. Since calculation effort of way-distance is much higher than that for geodesic distance we did a comparison of these two distances based on our data to see how high the error would be, if we use geodesic distance instead of way-distance. The following table shows the comparison for 7971 patients and 126 hospitals. Distances were calculated from each patient to each hospital. Calculation of geodesic distances took less than 1 second whereas way-distances took 4 days. Way-distances were calculated by Gosmore[15] and the OSM-File containing the bounding box of Austria.

Calculation was done using an Intel® Core™ i7-2620M processor which features 4 2.7GHz CPUs and 4GB RAM. Since Gosmore took 2GB RAM at peak times only two instances could be run at the same time. The following table shows the results:

Average absolute distance differences	68205.95 m
Average relative distance difference	32.57%
Maximum absolute distance differenc	283631.17 m
Maximum relative distance differenc	179.35%
Equal nearest hospital	11.59%
Equal nearest hospitals (ordered)	3.15 hospitals
Equal hospitals in nearest 5	4.36 hospitals
Equal hospitals in nearest 10	8.83 hospitals
Equal hospitals in nearest 20	18.50 hospitals

We calculated the absolute distance difference between geodesic distance and waydistance for the distance between each patient and each hospital. The relative distance difference is the distance difference divided by the mean value of both distances. From these value we calculated the average and maximum.

The value of *equal nearest hospital* gives the percentage of patients that have the same nearest hospital at both calculation methods. The next value ist similar: Not only the nearest hospital is taken into account but also the following hospitals when ordered by distance. It gives the average number of nearest hospitals that match both calculation methods. The three last values give the average number of hospitals that are the same in the set of the first 5, 10 or 20 hospitals.

From these figures one can see that the difference between these two distances is high but the order of the hospitals when ordered by distance to a certain patient is preserved. And since the order of the hospitals counts when a patient chooses one of them, the absolute error does not matter that much.

6 Simulation Results

We built one model with GIS-Data and another one that also uses a spatial environment but positions patients and hospitals randomly to see where the differences are. The number of hospitals and patients and the number of beds in each hospital were the same in both models to make them comparable.
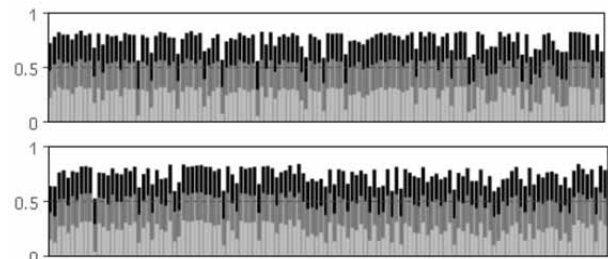

Figure 7. Average utilisation, income and treatment performance per hospital (without/with GIS-Data).

The comparisen of the hospitals' perfomance charts (Figure 7) shows that there are hospitals with low perfomance in both cases. These lie on the border of the environment in case of simulation without GIS-Data and in areas with less population in case of simulation with GIS-Data. These hospitals can be clearly identified when using GIS-Data. Further analysis showed that the eleventh hospital in the right chart is too big (too many beds) for the area's population.

In the following figures we see the hospitals and the connecting lines to patients that are associated with each hospital. Here we can clearly see that GIS-Data has great influence on the distribution of hospitals as they are not uniformly distributed on the area. The next figures show the distribution of patients. Black patients have a distance greater than 60km/100px to their hospital while white patients live almost next to a hospital. Hence black areas could be a hint on missing hospitals.

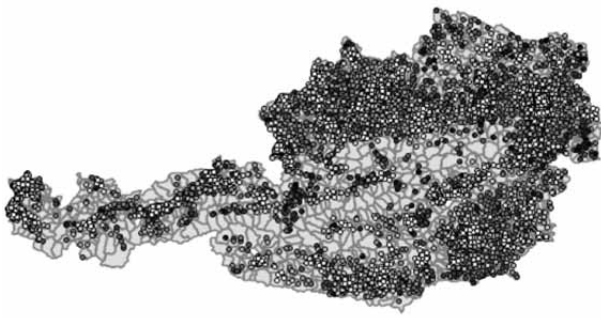**Figure 8**. Hospitals and connections to patients.

Figure 9. Patients coloured by distance to associated hospital

7 Conclusion

The integration of GIS-Data has a big impact on the simulation. Quite some effort is necessary for incorporating GIS-Data into the model. During simulation it mainly influences the distribution of patients and hospitals. Visualisations of the simulation take great advantage of the additional data. Further possibilities are regional statistics of any kind (e.g. health care supply or average distances per municipality). One could also assign different costs to treatments depending on geographic location or add other regional data.

References

[1] C. Urach, F. Breitenecker, F. Miksch, N. Popper, I. Wilbacher, G. Zauner: *Distance Dependence on the Willingness of Patients to Participate in Ambulatory Heart Rehabilitation Programs*. Internal Report.

[2] Hauptverband der österreichischen Sozialversicherungsträger, 18 July 2011, http://www.hauptverband.at

[3] STATISTICS AUSTRIA – Geodata, 12 July 2011, http://www.statistik.at/web_de/downloads/raster/r10000m.zip

[4] STATISTICS AUSTRIA – Statistical Grids MGI-Lambert, 12 July 2011, http://www.statistik.at/web_de/downloads/raster/r10000m_mit_daten_txt.zip

[5] STATISTICS AUSTRIA – Geodata, 12 July 2011,http://www.statistik.at/web_de/downloads/geodaten/statistik_austria_04_dsr_sehrstarkgen.zip

[6] GEOFABRIK Downloadbereich, 12 July 2011,http://download.geofabrik.de/osm/europe.osm.bz2

[7] GEOFABRIK, 12 July 2011, http://www.geofabrik.de

[8] OpenStreetMap, 12 July 2011, http://www.openstreetmap.org

[9] PostGIS, 12 July 2011, http://postgis.refractions.net

[10] PostgreSQL, 12 July 2011, http://www.postgresql.org

[11] Nominatim, 18 July 2011, http://wiki.openstreetmap.org/wiki/Nominatim

[12] EPSG::4326, 12 July 2011, http://www.epsg-registry.org/report.htm?type= selection&entity= urn:ogc:def:crs:EPSG::4326&reportDetail=short&style= urn:uuid:report-style:default-with-code&style_name=OGP%20Default%20With%20Code &title=EPSG::4326

[13] EPSG::3857, 12 July 2011, http://www.epsg-registry.org/report.htm?type =selection&entity=urn:ogc:def:crs:EPSG::3857&reportDetail=short&style=urn:uuid:report-style:default-with-code&style_name=OGP%20Default%20With%20Code &title=EPSG::3857

[14] GeoTools *The Open Source Java GIS Toolkit*, 13 July 2011, http://geotools.org

[15] Gosmore, July 2011, http://wiki.openstreetmap.org/wiki/