Deutsches Institut für
Wirtschaftsforschung

DIW BERLIN

www.diw.de

SOEP The German Socio-Economic Panel Study

SOEPpapers
on Multidisciplinary Panel Data Research

283

Jan Goebel • Michael Wurm • Gert G. Wagner

Exploring the Linkage of Spatial Indicators from Remote Sensing Data with Survey Data—the Case of the Socio-Economic Panel (SOEP) and 3D City Models

Berlin, March 2010

SOEPpapers on Multidisciplinary Panel Data Research
at DIW Berlin

This series presents research findings based either directly on data from the German Socio-Economic Panel Study (SOEP) or using SOEP data as part of an internationally comparable data set (e.g. CNEF, ECHP, LIS, LWS, CHER/PACO). SOEP is a truly multidisciplinary household panel study covering a wide range of social and behavioral sciences: economics, sociology, psychology, survey methodology, econometrics and applied statistics, educational science, political science, public health, behavioral genetics, demography, geography, and sport science.

The decision to publish a submission in SOEPpapers is made by a board of editors chosen by the DIW Berlin to represent the wide range of disciplines covered by SOEP. There is no external referee process and papers are either accepted or rejected without revision. Papers appear in this series as works in progress and may also appear elsewhere. They often represent preliminary studies and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be requested from the author directly.

Any opinions expressed in this series are those of the author(s) and not those of DIW Berlin. Research disseminated by DIW Berlin may include views on public policy issues, but the institute itself takes no institutional policy positions.

The SOEPpapers are available at
http://www.diw.de/soeppapers

ISSN: 1864-6689 (online)

German Socio-Economic Panel Study (SOEP)
DIW Berlin
Mohrenstrasse 58
10117 Berlin, Germany

Contact: Uta Rahmann | urahmann@diw.de

Exploring the Linkage of Spatial Indicators from Remote Sensing Data with Survey Data – The Case of the Socio-Economic Panel (SOEP) and 3D City Models*

Jan Goebel[1], Michael Wurm,[2,3] and Gert G. Wagner[1, 4]

[1] SOEP at DIW Berlin
[2] German Aerospace Center (DLR), German Remote Sensing Data Center (DFD), Oberpfaffenhofen
[3] Julius Maximilians University of Wuerzburg, Remote Sensing Unit
[4] Berlin University of Technology (TUB) and MPI for Human Development, Berlin

* A first and longer version of this paper under the title "Integration raumrelevanter Indikatoren in sozial- und verhaltenswissenschaftliche Analysen," has been published in German in H. Taubenböck and S. Dech. (eds.), Fernerkundung im urbanen Raum – Vom Satellitenbild zur Planungspraxis. Darmstadt 2010: Wissenschaftliche Buchgesellschaft. This contribution is part of the Feasibility and Planning Study on Fundamentally Improving and Expanding the Socio-Economic Panel (SOEP) Longitudinal Infrastructure Study, funded by the Federal Ministry of Education and Research (Project No. 01UW0706).

Abstract

This paper demonstrates the spatial evaluation of survey data from the German Socio-Economic Panel (SOEP) study using geo-coordinates and spatially relevant indicators from remote sensing data. By geocoding the addresses of survey households with block-level geographic precision (while preventing their identification by name and guaranteeing their complete anonymity), data on SOEP respondents can now be analyzed in a specific spatial context. In the past, regional analyses of SOEP based on official regional indicators (e.g., the unemployment rate) always had only very imprecise spatial information to work with. This limitation has now been overcome with the geocoded respondents' information. Within a protected unit of the fieldwork organization responsible for SOEP (TNS Infratest, Munich), the addresses of survey households can now be used to generate a variable describing the location of the household with block-level precision. At DIW Berlin, this additional variable is fed into a special computer infrastructure with multiple security layers that makes the socio-economic analysis possible. This paper demonstrates the use of this geographical location and remote sensing data to check respondents' subjective assessments of the location of their residence, and discusses the analytical potential of linking remote sensing data and survey data.

Zusammenfassung

Der Beitrag zeigt räumliche Auswertungsmethoden von Daten der Längsschnittstudie Sozio-oekonomisches Panel (SOEP) auf Basis von Geokoordinaten und Einbindung von raumrelevanten Indikatoren aus Fernerkundungsdaten. Durch die Verortung von Privathaushalten (ohne Darstellung der Identität und unter Gewährleistung vollständiger Anonymität) mit einer Genauigkeit auf Straßenabschnittsebene ist es erstmals möglich, Personen im Zeitraum 2000-2008, die mit der Stichprobe des SOEP befragt wurden, im kleinräumigen Kontext zu analysieren. Bisherige regionale Analysen des SOEP waren immer an administrative Grenze gebunden, diese Restriktion ist mit der Verortung der Haushalte aufgehoben. Aus den Koordinaten wird eine Variable *Lage* erzeugt, welche über eine spezielle, technisch mehrfach gesicherte Rechnerinfrastruktur den anonymisierten Daten der Befragungshaushalte zugewiesen wird und somit die Einschätzung über die Lage der Wohnhauses des Befragten mit einer objektiven Variable überprüft werden kann. Es wird anhand einiger Beispiele das Analysepotential aufgezeigt und diskutiert.

Keywords: remote sensing data, social sciences, behavioral sciences, multi-disciplinarity, SOEP

JEL Classification: C81, C83, R14

1 Introduction

This paper describes new approaches and presents the initial results of an interdisciplinary analysis of the linkage of data from the German Socio-Economic Panel (SOEP) to spatially relevant indicators derived from *remote sensing data*.

In the following, survey data of the SOEP on households and persons living in Germany are linked to detailed "3D city models". For these models digital image processing methods on the basis of high-resolution satellite images and altitude information from digital surface models (DSMs) serve as a basis for generating spatial parameters such as the *density* of built urban areas, or environmentally relevant indicators such as the designation of *green space* (or the percentage of green space). This information can be used as additional new variables in the SOEP data structure using state-of-the-art statistical methods. The anonymity of the households and persons surveyed is safeguarded throughout the entire analysis as well as in the results.

Due to a dearth of good data, there have been surprisingly few analyses in the social sciences based on spatial data (Bradburn, 2004) for a number of years, with the exception of those carried out by geographers. Turner (1998) explains the skepticism of social scientists toward remote sensing by the fact that they place more weight on the causes than on the actual physical locations of certain behaviors, and that, to put it simply, the variables they are interested in cannot be measured from the air (e.g. government policy, land tenure, distribution of power and wealth, market mechanisms or social standards).

Rindfuss and Stern (1998) describe a lack of involvement of social scientists in the development of new remote sensing sensors and technologies, as well as a low level of willingness of scientific disciplines to break with their traditions. They also find that typical socio-scientific data do not provide geographical coordinates, which makes it difficult to identify spatial locations and hence to carry out an integrated analysis with remote sensing data. The advantage of an interdisciplinary analysis is clear: remote sensing data and methods make it possible to monitor the physical surroundings of people's homes and to describe and assess them using objective processes (image classification). Only through the integration of contextual information describing people's neighborhoods can correlations between human beings and their physical environments be made possible.

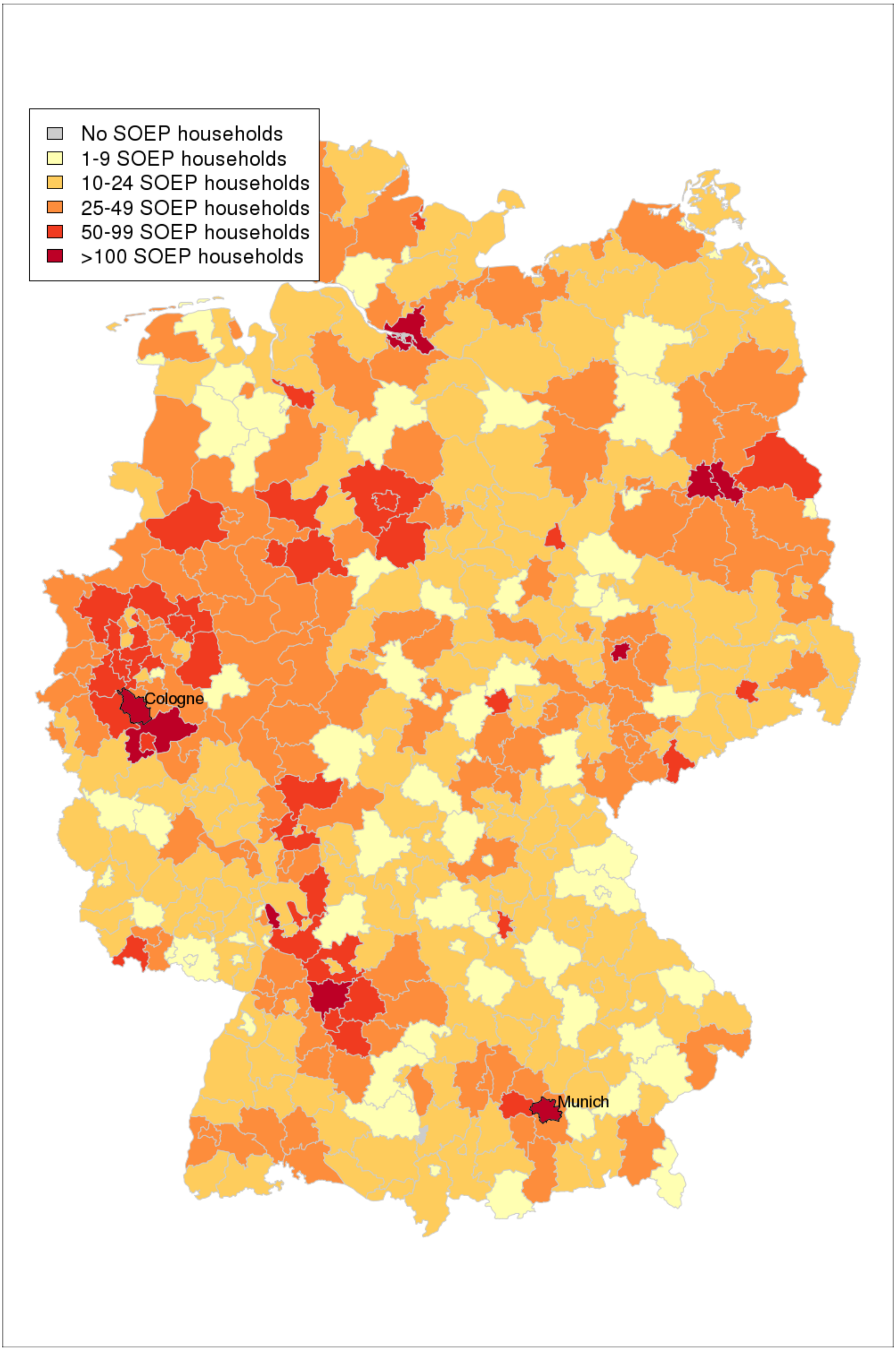
2 The SOEP Database

The German Socio-Economic Panel (SOEP) is a longitudinal survey of households and individuals that has been conducted annually in the western part of Germany since 1984 and was extended to eastern Germany in 1990 (see Wagner et al. 2007 and Frick et al. 2010). The structure of the SOEP is that of a panel survey, meaning that the same people are surveyed again and again, some of whom have already participated in 25 waves (approximately 2,500 individuals). Unlike cross-sectional studies, which provide only a one-off “snapshot” of respondents,,showing, for example, the percentage who are married, divorced, or living alone at the time of the survey, the panel design of the SOEP study makes it possible to conduct a repeated survey of the same individuals. This allows an analysis of processes of change at both the individual and household levels. Thus, if the proportion of respondents who are married, divorced, or living alone changes from one survey year to the next, it is possible to check whether or not it was the same people who changed status. Furthermore, with a sample size of over 20,000 people living in over 10,000 households, the SOEP is an unusually large sample for a scientific survey.

This large case number also makes it possible to draw conclusions about how specific geographical characteristics affect psychological, social, and economic behavior. Although the sample of 10,000 households from throughout the country is not sufficient for a detailed spatial description of the entire Federal Republic of Germany (see Fig. 1), it is possible to check their correlations with psychosocial, social, and economic characteristics for specific geographically differentiated types of living environments. For example, Gerstorf et al. (2010) show that life satisfaction in the last few years before a person dies depends on the strength of the economy and the unemployment rate in the district where a respondent lived during the last few years of his or her life.

While traditional SOEP analyses have always been restricted to administrative units, selecting the area using geo-coordinates makes it possible to define any location independently of any territorial changes in the administrative units.

Fig. 1: Distribution of SOEP households in Germany (Survey year 2008)

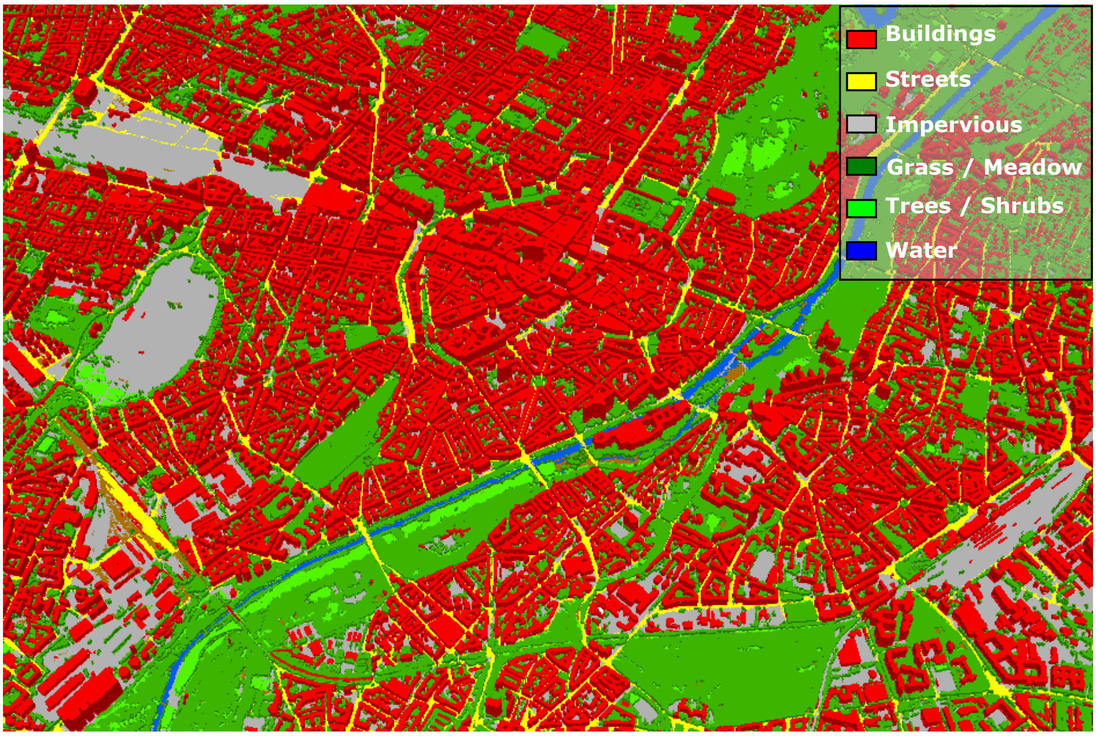


Source: SOEP 2009, authors' own calculations, rural and urban districts (counties) as from 31.12.2000

3 Methods

The objective of the present study is to show the correlations between objective spatial geodata and subjective indicators from socio-scientific survey data by conducting an interdisciplinary analysis of contextual information from remote sensing data and comparable variables from the SOEP data set, as described at the outset. In the context of a pilot study, we carried out analyses for the cities of Cologne and Munich. In the sample period, observations were available for 491 households in these two regions. For both regions studied, 3D city models were derived from spatially high-resolution (1m) remote sensing data (see Wurm et al., 2009a and 2009b).

Fig. 2: Data derived from spatially high-resolution remote sensing data



3D-city model derived by means of a digital surface model and very high resolution optical satellite imagery for a subset of the city center of Munich; viewed from south-east.

Source: DLR 2009, authors' own calculations

To this end, a transferable object-based approach was developed in order to first demarcate individual buildings from digital surface models (DSMs) and optical satellite image data and to deduce the number of stories for each building using the absolute altitude information. In the next step, information on land cover from the optical satellite image data (IKONOS) was produced with hierarchical segmentation (Esch et al., 2008) and classification (Taubenböck & Roth, 2007). By taking into account comparable basic data, this method guarantees a comparable objective database of both regions studied for the synergetic analysis that builds on this using survey data (see Fig. 2).

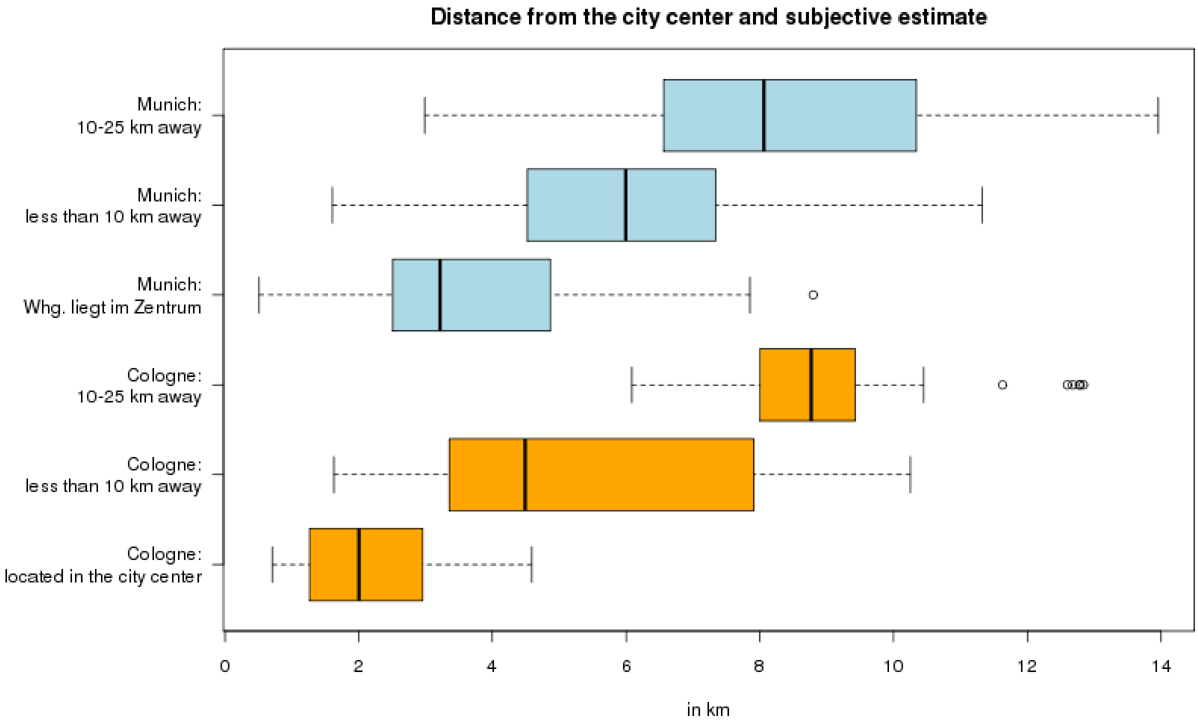
The three-dimensional city model shows various classes of land cover for the urban area ("road," "meadows/grassland," "woods/shrubbery," "water," "paved," and "undeveloped land") and individual buildings according to size and form. Thus, extensive information is available on location and size of buildings and of green space in the urban area. In the present study, initial correlations between the subjective survey information and the objective information derived from remote sensing data will be investigated. This requires using comparable variables in SOEP.

4 Results

A good means of checking whether it is feasible to link SOEP survey data with geo-referenced data is to use the distance between the surveyed household and the nearest city center. This variable is, on the one hand, estimated by the respondents during the interview and can also be measured using the coordinates. In order to do this, distances between the place of residence and city center are calculated using the above-mentioned coordinates of the households at block level and the coordinates of the respective city center in Cologne and Munich. This distance is assigned to the individual survey household as a new variable, taking into consideration the above-mentioned data protection provisions.

The last survey of households asking for self-assessment of the distance from the nearest city center took place in 2004, with respondents choosing from seven possible answers ("No response," "Home is located in the city center," "Less than 10 km," "10 to less than 25 km," "25 to less than 40 km," "40 to less than 60 km," "60 km and over"). A comparison between the estimates from the (subjective) responses given and the (objective) distances measured using the coordinates is presented in Fig. 3. The boxplots show that for both regions studied, greater measured distances are also generally reflected in greater distances stated in the responses.

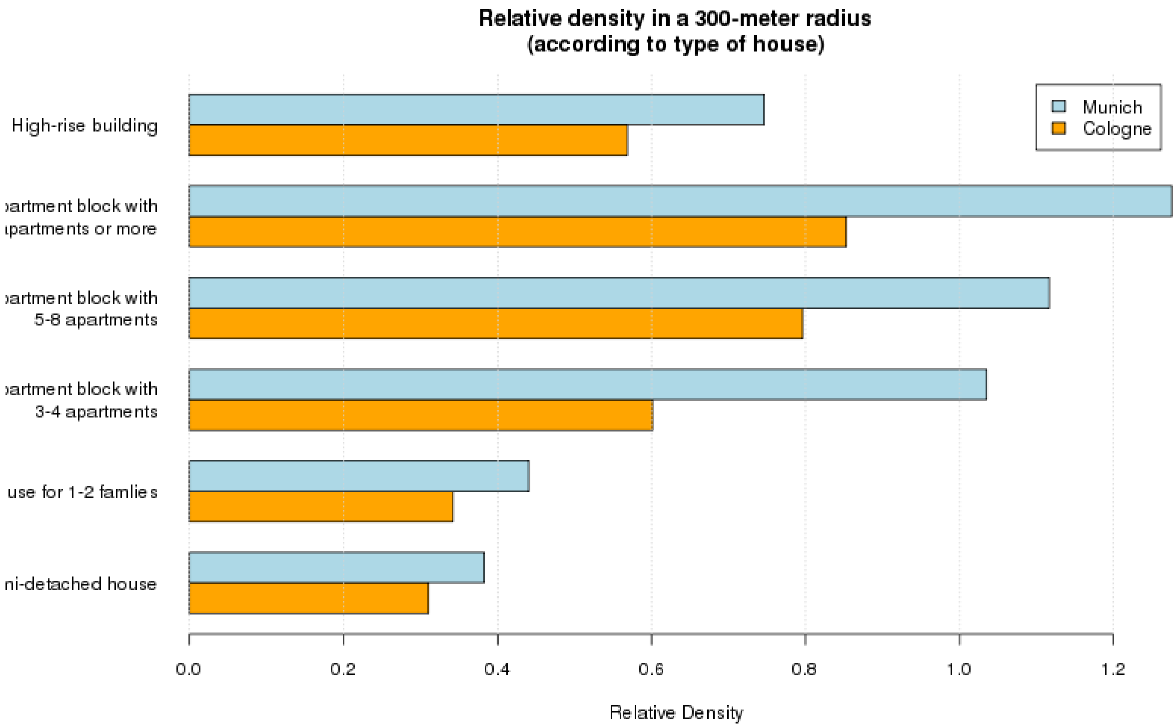
Fig. 3: Comparison between estimated distance and measured distance from the city center



Source: SOEP + DLR 2009, authors' own calculations

Differences in the regions studied are shown in detail: For the class "10 to less than 25 km," the median in the two cities is between 8.1 km (Munich) and 8.8 km (Cologne). The measured distance refers to distance between two points as the crow flies, while estimated distances refer mainly to the distance traveled. In Cologne those who live at a median distance of 2 km away from the "Alter Markt" (mean value = 2.4 km) consider themselves to be residents of the city center. In Munich, on the other hand, the median value is at 3.2 km (mean = 3.8 km) from the "Marienplatz". One possible cause of the difference in the estimation may lie in the city layout. Cologne city center is surrounded by parks that have developed from the cordon of forts around the city. This ring is at a distance of between 2.2 km in the west and approx. 2.5 km in the north and south from the center, and is both a distinctive and a dividing element in the cityscape. Further analyses regarding the subjective estimation on the location in the urban area are needed to shed more light on possible correlations.

Fig 4: Comparison between the average relative density and the subjective estimation of the residential area.



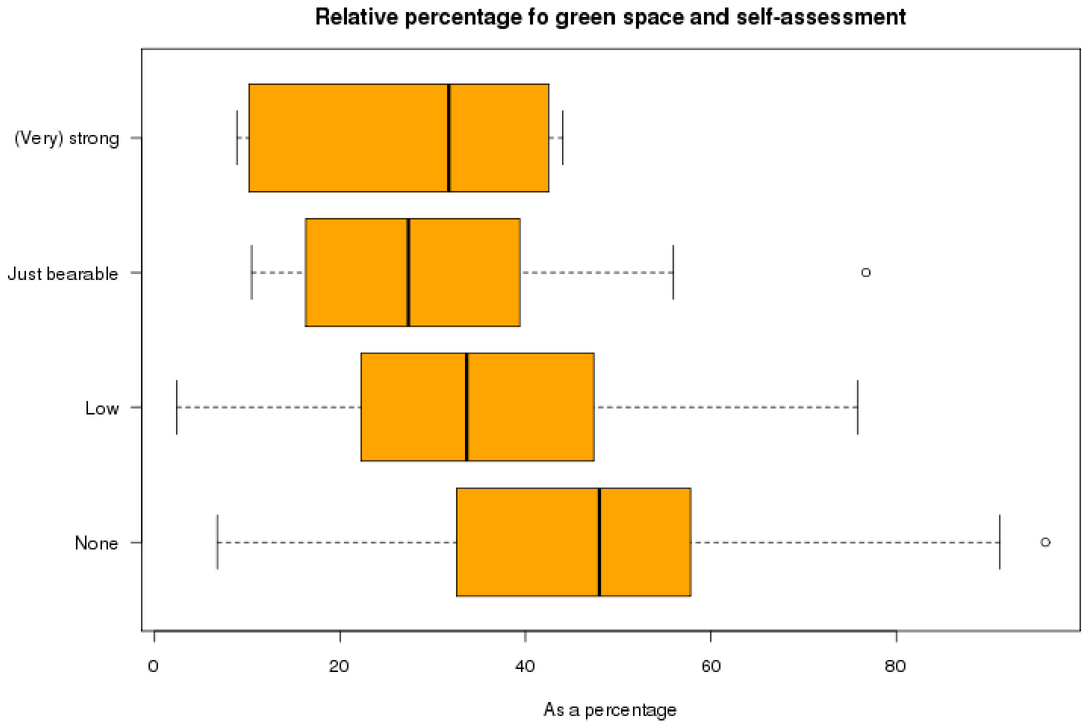
Source: SOEP + DLR 2009, authors' own calculations

The possible applications for linking spatial information with SOEP data are not limited to the comparison of SOEP *location* variables in the urban area with household location using geo-coordinates. One objective variable that can only be created using additional geodata is *density*. This indicator depicts the built environment and is the ratio of built area compared to the reference area. The built area is the area of all the full stories of a building and may be obtained from the three-dimensional urban model.

Hoffmeyer-Zlotnik (2000) ascertains the density of the respondents using the development structure in the immediate vicinity of their own home. This type of specification of the density of the neighborhood can be produced for all SOEP households using the 3D city model. The floor area of all buildings within a 300-meter radius of each household surveyed is calculated and divided by the overall area. This *relative density* can in turn be compared with subjective variables of SOEP respondents. Fig. 4 shows, for example, a comparison of the relative density with the interviewees' assessments of their residential area. For the two cities examined here, the mean values show that with a shift towards an increasing number of apartments, from "detached/semi-detached" to "apartment block with 9 or more apartments (but a maximum of 8 stories, that is, not a high-rise)", the relative density is also increasing in both cities.

Another variable that can be ascertained directly from the remote sensing data is the relative proportion of *green space* in the vicinity of the households surveyed. The green space derived from optical satellite image data, consisting of the classes "meadow/grassland" and "woods/shrubbery," can also be compared with the survey data. As with the calculation of the relative density, the proportion of green space is calculated for each household surveyed and its immediate environment (r=300m) and compared with SOEP indicators. Fig. 5 shows the correlation of the variables "To what extent do you consider yourself to be affected by the following environmental influences in your area: by a lack of accessible green spaces?" with the relative proportion of green space from the three-dimensional city model.

Fig. 5: Proportion of green space from satellite image data and adverse effect of a lack of green space.



Source: SOEP + DLR 2009, authors' own calculations

It has not been feasible to conduct a more in-depth and rigorous analysis to date due to the low case numbers in the two cities studied. It will not be possible to carry out really deep and stable analyses on the basis of SOEP until similar remote sensing data are available for more regions besides Munich and Cologne.

In order to at least illustrate the far-reaching analysis potential, however, we have also carried out a multivariate analysis of gross rents in Cologne and Munich. In SOEP, the head of household provides detailed information about the condition, facilities, and furnishings of the apartment, and also about the gross rent to be paid per square meter. In order to illustrate the additional explanatory power of the variables from remote sensing, the following procedure was used: in a first model, only the main effects of the variables from the SOEP survey were considered. The following variables were used as explanatory variables: house type, residential area, year of construction, need for renovation, central heating, balcony or terrace, cellar, and a dummy variable for Munich as place of residence.

In a second model, the three additional variables of *distance* from the center, *proportion of green space* and *urban density* in a 300-meter radius were added (from remote sensing data). Both models were then optimized automatically using a *stepwise* procedure (forward und backward selection) (including all possible interaction effects). Both models are shown in Table 1.

Owing to the small sample size and the resulting limitation in the number of variables to be estimated, the proportion of explained variance is relatively low. However, by including the variables from remote sensing, this figure increases (despite the small sample) by 3 percentage points to a good 14%. Information about the residential area from SOEP was still included in the first model, although it was not significant. The two variables about residential area and type of dwelling were no longer included in Model 2, and apparently were better explained by the additional variables of proportion of green space, urban density, and distance to the city center (see also Fig. 4 and Fig. 5).

After central heating, the density of development has the greatest positive effect on the rent per square meter of the households observed. The proportion of green space in the immediate vicinity of the home also has the effect of raising rents, with each additional percentage point of green space equating to an increase in the price per square meter of around 30 cents on average. Conversely, rents decrease with every kilometer further away from the center.

Table 1: Regression models to illustrate gross rent prices (per qm²)

	Model 1		Model 2	
	Coefficient	**Pr(>|t|)**	**Coefficient**	**Pr(>|t|)**
Residential area (SOEP) *Reference: Purely residential area. Old building*				
Purely residential area. New building.	-0.04	94.99%	-	-
Mixed area	-0.16	78.65%	-	-
Type of dwelling (SOEP) *Reference:Detached or semi-detached house (1-2 families)*				
Apartment block (3-8 apartments)	1.57	0.28%	-	-
Apartment block (with 9 apartments or more)	1.13	2.98%	-	-
Remote sensing variables				
Proportion of green space (%)	-	-	0.03	9.87%
Density of development	-	-	3.33	0.29%
Distance from the center; *Reference: in the city center*				
Distance <= 7km	-	-	-1.23	7.80%
Distance > 7km	-	-	-1.69	0.94%
N = 333 adj. R²:	0.11		0.14	

Additional control variables included: Year of construction, need for renovation, no central heating, no cellar, no garden and in Munich.
Source: SOEP + DLR 2009, authors' own calculations

5 Summary and Outlook

This paper demonstrates the spatial evaluation of data from the German Socio-Economic Panel (SOEP) study using geo-coordinates and linkage of spatially relevant indicators from remote sensing data. This data linkage also represents a completely new connection between two scientific disciplines. By geocoding the addresses of private households with block-level geographic precision (while preventing their identification by name and guaranteeing their complete anonymity), it is now possible to analyze data on respondents to the SOEP in the period 2000-2008 in a specific spatial context. In the past, regional analyses of SOEP data using official regional indicators (e.g., the unemployment rate) always had only very imprecise spatial information to work with. This limitation has now been overcome with the geocoded respondents' information. Any space can be defined using the geo-coordinates in combination with external geocoded information (see Lakes 2009, Hintze and Lakes 2009).

In the present study, the coordinates of SOEP's survey households were used to create a variable *location* which was assigned to the anonymized data of the household surveyed by means of a special computing infrastructure that is technically secured in several ways. This allows the respondent's subjective evaluation of the location of their place of residence to be checked against an objective variable. In this pilot study, the variables *density* and *proportion of green space* were derived from high-resolution remote sensing data and then compared with SOEP data on the two cities of Munich and Cologne.

Several examples were provided to demonstrate and discuss the analysis potential of geocoded survey data. All the variables studied confirm that objective and subjective surveys tend to produce similar results. With respondents' estimates of the distance of their residence from the city center, however, a greater variance between the objective variable and subjective evaluation is seen as the objective distance increases. The creation of comparable spatial indicators from remote sensing data should make it possible to create and jointly evaluate new variables for the whole of Germany in the near future.

References

Bradburn, N. M. (2004): Foreword to: Spatially integrated Social Science. In: Goodchild, M. F., Janelle, D. G. (Hrsg.) (2004): Spatially Integrated Social Science. Oxford University Press. ISBN 978-0-19-515270-8, 480 pages.

Donnay, J.-P, Barnsley, M. J. und Longley, P. A. (2001): Remote Sensing and Urban Analysis. In Donnay, J.-P, Barnsley, M. J. und Longley, P. A. (eds.): Remote Sensing and Urban Analysis, Taylor and Francis, London & New York, 3-18.

Esch, T., Thiel, M., Bock, M. und Dech, S. (2008): Improvement of image segmentation accuracy based on multi-scale optimization procedure. IEEE Geoscience and Remote Sensing Letters, vol. 5, no. 3.

Frick, J.R., Goebel, J., Engelmann, M., Rahmann, U. (2010, in press): The Research Data Center (RDC) of the German Socio-Economic Panel (SOEP), Schmollers Jahrbuch, European Data Watch.

Gerstorf, D., Ram, N., Goebel, J., Schupp, J., Lindenberger, U., & Wagner, G. G. (2010): Where people live and die makes a difference: Individual and geographic disparities in well-being progression at the end of life. SOEPpapers No. 287. Berlin. 49 pages.

Hintze, P. und Lakes, T. (2009): Geographically Referenced Data for Social Science. Working Paper No. 125. German Council for Social and Economic Data (RatSWD). 50 pages.

Hoffmeyer-Zlotnik, J.H.P. (2000): Regionalisierung sozialwissenschaftlicher Umfragedaten. Siedlungsstruktur und Wohnquartier. Westdeutscher Verlag, 249 pages.

Lakes, T. (2009): Geodata. Working Paper No. 89. German Council for Social and Economic Data (RatSWD). 14 pages.

Maktav, D., Erbek, F. S. und Jürgens, C. (2005): Remote Sensing of urban areas, International Journal of Remote Sensing, vol. 26, no. 4, pp. 665-659.

Owen, D., Green, A. und Elias, P. (2009): Review of geospatial resource needs. Review. Institute for Employment Research, University of Warwick. 49 pages.

Rindfuss R. R. und Stern, P. C. (1998): Linking Remote Sensing and Social Science: The Need and the Challenges. In: Liverman, D., Moran, E.F., Rindfuss, R.R. und Stern, P.C. (eds.): People and Pixels. Linking Remote Sensing and Social Science. National Academy Press, Washington D.C., 256 pages.

Taubenböck, H. und Roth, A. (2007): A transferable and stable object oriented classification approach in various urban areas and various high resolution sensors. Urban Remote Sensing Joint Event 2007.

Turner, II B.L. (1998): Frontiers of Exploration: Remote Sensing and Social Science Research. In Proceedings of Pecora 13, Symposium on Human Interactions with the Environment. Bethesda, Md.: American Society for Photogrammetry and Remote Sensing.

Wagner, G.G., Frick, J. R. und Schupp, J. (2007): The German Socio-Economic Panel Study (SOEP) – Scope, Evolution and Enhancements. Schmollers Jahrbuch 127 no. 1, pp. 139-169.

Wurm, M., Taubenböck, H., Roth, A. und Dech, S. (2009a): Urban structuring using multisensoral remote sensing data – By the example of the German cities Cologne and Dresden. In: Proceedings of 2009 Urban Remote Sensing Joint Event, Shanghai, China.

Wurm, M., Taubenböck, H., Krings, S., Birkmann, J., Roth, A. und Dech, S. (2009b): Derivation of population distribution for vulnerability assessment in flood-prone German cities using multisensoral remote sensing data. In: Proceedings of the SPIE Europe Remote Sensing, 11 pages.