Using European Time Use Data for Modelling Individual Exposure to Air Pollution

Lydia E. Gerharz[1*], Miranda Loh[2], Aileen Yang[3], Alexandra Kuhn[4]

[1]Institute for Geoinformatics, University of Münster, Germany

[2]National Institute for Health and Welfare, Kuopio, Finland

[3]Norwegian Institute for Air Research, Kjeller, Norway

[4] Institute for Energy Economics and the Rational Use of Energy, University of Stuttgart, Germany

Abstract. In the assessment of exposure to air pollution, individual behavioural patterns have a significant influence on daily exposure. The average exposure of an individual person over a time period is estimated by the sum of the ambient air pollutant concentration c_i multiplied by the time fraction t_i an individual spends at each location or microenvironment (ME) *i*. Time use statistics are essential for estimating the time fraction per ME.

In HEIMTSA, a European project on integrated health assessment, individuals' exposure to outdoor air pollution is modelled for Europe. For the model approach, the Multinational Time Use Survey (MTUS) data provides a harmonised set of national European time activity data. These activities are classified into the time individuals spend in five MEs: Home, Work, Outdoor, Transportation, and Other. The air pollutants' concentration in these microenvironments is calculated by environment specific infiltration factors and the modelled outdoor concentration.

The time activity approach allows the comparison of policy scenarios influencing behaviour and infiltration factors. Demographical and activity-based grouping, generated with clustering techniques, were compared as predictor variables for exposure. The results imply that demographical factors solely are not good predictors for exposure and that differences between countries exceed differences between demographic groups depending on the choice of infiltration factors. Although activity-based clusters are difficult to describe by socio-economic parameters, exposure and policy impacts are differentiated better. This approach could be useful to compare policy impacts and highlight extremely exposed subgroups in European countries.

Key words. Exposure modelling, harmonised time use, cluster analysis, population subgroups

INTRODUCTION

The assessment of health burdens caused by ambient air pollution has previously been done by using ambient air quality, as measured at fixed site monitors, as an exposure proxy. Relationships between the concentration of, for instance particulate matter (PM), and the health responses of the population were demonstrated by large scale studies like the American Cancer Society Study (Pope *et al.*, 2002) and Harvard Six Cities Study (Dockery *et al.*, 1993). In fact, the true *exposure* of individuals can vary considerably over a population, depending on individuals' behaviour and whereabouts. Exposure differs from ambient air quality because it describes the pollutant concentration in the environment that an individual comes into direct contact with. A person's exposure consists of multiple microenvironmental exposures. A microenvironment (ME) refers to a location or surroundings that a person spends time in and where the air pollutant concentrations are treated as homogenous (*e.g.* IPCS 2004). If we know the pollutant concentration c_i and the fraction of time t_i per day spent by an individual in each ME *i*, the average exposure *E* can be expressed as:

$$E = \sum_{i=1}^{n} c_i t_i \quad (1)$$

*contact author: Institute for Geoinformatics, Weseler Str. 253, 48151 Münster, Germany

+49 251 8339760, Lydia.Gerharz@uni-muenster.de

Modelling exposures in this way is considered an indirect exposure assessment approach (Klepeis 1999), as it does not directly measure a person's exposure. As one can see, variability in exposures depends on the variability in microenvironment concentrations and individuals' time activity patterns. The concentration c_i can be estimated by models or derived from measurement studies and combined with the time fraction t_i extracted from time-use diaries. The concentration in MEs varies over space and time as the emissions and distributions of air pollutants are not constant. Variability in time activity patterns also occurs in two dimensions: cross-sectionally between individuals or groups of individuals, and longitudinally, across time. One example of exposure models is presented by Zidek *et al.* (2005). The stochastic model pCNEM includes measurements of outdoor concentrations and parameter distributions for modelling the indoor concentration as well as US activity data from the National Human Activity Patterns Survey (NHAPS, Klepeis *et al.* 2001) to estimate the exposure probability distribution for a random individual in the population. Another example is HAPEM (Özkaynak *et al.*, 2008) which is provided by the US EPA and calculates estimates of exposure to ambient air pollutants for stratified population groups, using an air pollution model, and proximity and penetration factors to estimate ME concentrations.

In HEIMTSA, a European 6th framework project on integrated health impact assessment, methods and tools for scenario assessment are developed to quantify and compare the magnitude of current and future health impacts from different European environmental policies. One important part of the HEIMTSA framework is the quantification of exposure to outdoor air pollution. This approach includes modelling individual exposure to ambient air pollution on a European scale using pollutant modelling techniques and time activity data. The aim is to provide a full chain model from emissions to exposure, including policy changes that could affect emissions and thereby concentrations, outdoor-to-indoor infiltration factors and human behaviour. To provide a model methodology for the whole of Europe, harmonised European time use data sets are combined with regional infiltration factor estimates. The change in exposure by the latter two can be quantified by using an *Exposure Scaling Factor* (ESF).

In this paper, we will present the exposure modelling methodology developed for HEIMTSA on the example of exposure to PM2.5 using the Multinational Time Use Survey (MTUS, http://www.timeuse.org/mtus/) diary data. To investigate the influence of human behaviour on the exposure results, two different grouping techniques for the time use data were tested and the resulting exposure distributions compared. As a conclusion, the potential use and limitations of the approach will be discussed as well as potential future work.

METHODOLOGY

Exposure modelling

To demonstrate the model approach, individual exposure to PM2.5 (fine particles, diameters <2.5 µm) will be considered. If the ME *i* in eq. 1 is an indoor environment, a mass balance model can be used to estimate the concentration indoors (Hänninen *et al.*, 2004). In this model, the indoor concentration is composed of two sources: pollutant infiltration from outdoors and pollution caused by indoor sources. As we are focussing on the exposure caused by ambient air pollution, the indoor concentration is expressed as a function of infiltrated outdoor concentration, ignoring the indoor sources:

$$C_{in|out} = C_{out} \frac{pa}{a+k} = C_{out} F_i \qquad (2)$$

where, a = the air exchange rate, p = the particle infiltration factor and k = the particle deposition rate. This infiltration factor F_i for the outdoor concentration C_{out} represents the fraction of particles from the outdoors that pass into the microenvironment, and therefore lies between 0 and 1. Infiltration factors have been estimated for buildings in European cities in several studies, e.g. EXPOLIS (Hänninen *et al*., 2004) and RUPIOH (Hoek *et al*., 2008). The value of F_i depends on climate regions, housing types, ventilation systems and seasons. For travel MEs, a traffic enrichment factor is used which is usually larger than 1, as it reflects the relative concentration experienced in-traffic compared to the ambient background. Estimates of infiltration and traffic enrichment factors for the HEIMTSA exposure modelling are presented in the *Modelling framework & implementation* section.

Combining the infiltration factors with the time spent in each ME we derive the *Exposure Scaling Factor*:

$$ESF(s,c,p) = \sum_{i=1}^{n} F_i(c,p)t_i(c,s) \quad + E(c,p)t_{travel}(c,s) \tag{3}$$

which is calculated per population sub-group s, country c and pollutant p across n microenvironments and with t_i = time fraction of the day spent in ME i and E = traffic enrichment factor. The ESF is essentially a dimensionless scaling factor of the ambient concentration for the population living and working in each cell. ESFs are estimated using Monte Carlo modelling to generate distributions across the population.

An exposure module is being developed for use with an existing air quality decision support system. For the estimation of the spatial distribution of hazardous air pollutants over Europe, the system EcoSense is applied. Detailed discussion of the model can be found in Krewitt *et al.* (1999). EcoSense calculates the dispersion and chemical transformation of the respective pollutant and has the advantage of applying emission reduction scenarios to estimate impacts of policies on the concentration. The applied model outputs are annual average concentrations on a 50 x 50 km^2 cell size raster (EMEP-grid). To estimate exposures to the modelled ambient concentrations, we apply the ESF to the ambient concentration per grid cell. We will focus in this work on the influence of time use on the ESF estimates. Further steps to exposures were presented in Kuhn *et al.* (2009) and will be omitted here.

Time use data

To estimate the time fraction spent in each ME, usually diaries are used. As HEIMTSA aims to assess health impacts for the whole EU-30, a harmonised data set for Europe is necessary to ensure comparability between countries. We decided to use the datasets of the Multinational Time Use Survey (MTUS) which provides a set of national time use surveys in Europe. The available MTUS countries with data from the 1990s onwards are listed in Table 1, classified into regions based on sets of regional infiltration factors determined in Hänninen *et al.* (2009). MTUS includes 41 classified activities. As it is impossible to model the concentration for all 41 activities separately, we classified them into five MEs, namely work, home, travel, outdoor, and other. Unfortunately, the MTUS activity catalogue contains little information about the locations of the activities or travel modes which required the investigators to use subjective judgement as to which ME each activity would fit into. The preliminary categorisation is shown in Table 2.

Table 1:Countries and number of diaries included in exposure model.

Country	Region	Years of survey	No. of diaries
Austria	Central Europe	1992	25’162
France	Central Europe	1998/99	15’318
Germany	Central Europe	1991/92, 2001/02	61’625
Netherlands	Central Europe	1990, 1995, 2000	8’034
Slovenia	Central Europe	2000/01	12’273
Norway	Northern Europe	1990/91, 2000/01	13’804
Sweden	Northern Europe	2000/01	7’747
Italy	Mediterrean Europe	2002/03	52’206
Spain	Mediterrean Europe	2002/03	46’774
United Kingdom	Northwestern Europe	2000/01	20’980

For exposure modelling, usually time activity diaries are classified according to demographic factors like age and gender (Özkaynak *et al*., 2008). This grouping can lead to the problem of large variance in the time use behaviour within the groups due to the heterogeneity of group compositions. McCurdy & Graham (2003) recommended distinguishing additionally between seasons, temperature, precipitation and day-type as these factors have been found to be influential in determining exposures. On the other hand, increasing the number of variables by which to classify the population leads to a larger number of groups with decreasing numbers of individuals. This reduces the number of diaries available for simulation over a year-long period. As annual averages of air pollutant concentrations were applied, the time use pattern for a whole year was estimated by sampling diaries within each group to approximate the annual average time spent per subgroup in each ME and preserve intra-individual variability. For consistency across the MTUS results, we chose to use only single day diaries, although some national studies provided diaries of several days. To avoid having small sample sizes per group, we decided to distinguish between gender (male/female), age (<15, 15-64, >65) and employment status (work/nonwork) for the middle age group, yielding eight cohorts. These are not only hypothesized to have differing time activity patterns, but are also similar to the types of stratification typically used in air pollution epidemiology studies. Thus, they may have more relevance for health impact assessments.

When assessing differences in exposure due to behavioural changes induced by policies, one expects that the policy impacts would be much greater between population groups that are defined according to homogeneous behaviours. Following from this hypothesis, if the time use behaviour is too heterogeneously distributed within a single group, the policy effect in the population becomes less clear. To test this hypothesis, we investigated the impact of a second grouping technique, based on time spent in microenvironments, compared to the demographic grouping in exposure assessment. In order to divide the population into cohorts with similar behaviour, a clustering algorithm was applied. Clusters are defined as point clouds with members having a minimum distance to the cluster centre. A simple method to define these cluster centres is the iterative *k-means* algorithm (MacQueen 1967, Hartigan *et al.* 1979). For a given number *k* of cluster centres, this algorithm allocates the cluster centres randomly for the initial step and assigns each point to the nearest cluster centre using a defined distance measure. In the next iterative step, the cluster centres are moved to the new central point of the respective cluster and the allocation of the nearest member points follows again. The iteration stops if the centres do not change any more. Hierarchical clustering methods, namely agglomerative and divisive techniques, were also tested by us but found less effective and traceable.

When using the k-means algorithm, it is possible that the position of the final clusters could depend on the initial cluster allocation. Thus, the algorithm should be repeated several times to ensure that not only a local optimum is found. Another problem is that the number of clusters has to be defined *a priori*. Running the algorithm with different *k* values and comparing the amount of explained variance for each step could help to decide on a meaningful *k*. Clustering was tested for several countries and detailed results are presented for the MTUS data from Germany.

Table 2: Grouping of activities into MEs.

MTUS Activity codes	Variable Label	ME	MTUS Activity codes	Variable Label	ME
AV1	Formal work	Work	AV18	Excursions, trips	Outdoor
AV3	Second job		AV21	Walks	
AV4	School/classes		AV9	Gardening	
AV8	Odd jobs		AV10	Shopping	Other
AV5	Travel to/from work	Travel	AV11	Child care	
AV12	Domestic travel		AV14	Receive personal services	
AV17	Leisure travel		AV19	Playing sport	
AV2	Paid work at home	Home	AV20	Watching sport	
AV6	Cooking/washing up		AV23	Civic organizations	
AV7	Housework		AV38	Entertaining friends	
AV16	Sleep/naps		AV40	Pastimes/hobbies	
AV13	Dressing/toilet		AV41	Unknown activity	
AV30	Listening to radio		AV15	Meals/snacks	
AV31	Watching T.V.		AV22	At church	
AV32	Listening to music, etc.		AV24	Cinema/theatre	
AV33	Study		AV25	Dance/party, etc.	
AV34	Reading books		AV26	Social clubs	
AV35	Reading pa-pers/magazines		AV27	Pubs	
AV36	Relaxing		AV28	Restaurants	
AV39	Knitting/sewing		AV29	Visiting friends	
			AV37	Conversation	

Modelling framework & implementation

The ME infiltration estimates are based on the infiltration factors measured in EXPOLIS and classified into the regional representative values according to Hänninen *et al.* (2009) as shown in Table 1. Traffic enrichment factors were derived from literature review (Gulliver and Briggs, 2004 and Kaur *et al*., 2005).

As the EXPOLIS measurements were performed in homes, the work and other infiltration factors had to be estimated via comparison with other studies and expert judgement. By a comparison with the ISS (Instituto Superiore di Sanità) and the EXPOLIS study (Hänninen *et al.*, 2009), F_{work} was assumed to be 15 % lower than F_{home}. As the other indoor ME contains activities with an unspecified location which could be either indoors or outdoors, the infiltration factor is chosen between F_{home} and 1 which is representative for the outdoor ME. The standard deviation was transferred to the other F_i.

For the calculation of the ESF, time use distributions per ME of the respective cohort must be combined with the infiltration factors (eq. 3). We applied a Monte Carlo sampling of diaries from the time use data and from the infiltration factor distributions. A lognormal distribution was assumed for the infiltration factors (Hänninen *et al*., 2004). For the time use data, a non-parametric method was used where weekdays and weekend diaries were sampled with the ratio 5:2 for 365 days. The same applied for seasonal sampling with a ratio of 1:1 although we did not use seasonal infiltration factors for this analysis. For the ESF results, an analysis of variance (ANOVA) was performed to calculate the effectiveness of the chosen groups. All analyses and calculations presented here were performed in R 2.8.1.

RESULTS

Comparison of time use distributions

The time use distribution of the clusters and demographic groups based on the German MTUS data are shown in Figure 1 and Figure 2. For the number of clusters *k*, the value of 5 seemed to reduce the intra-group variability noticeably as can be seen by the size of the boxes of the clusters in Figure 1. Compared to the clusters, the demographic groups have wider distributions for the time spent in each ME. Nevertheless, performing an ANOVA on the groupings and time spent per ME was significant for the demographic groups as well as for the clusters. For the classification of working/not working individuals, the employment status provided by the MTUS dataset was used. It is clear from Figure 2, that this classification is not strict, as people with an unemployed status also spent time working. This might be due to the fact that the work classification also includes school. Also the distribution of time spent working per day is very large which also comes from the fact that part-time and full-time jobs are not treated separately and because weekday and weekend days are represented together in the plots.

A comparison of demographic groups and clusters between all MTUS countries showed a stable time use pattern for clusters. Even the percentage of diaries per cluster is comparable for similar clusters between countries.

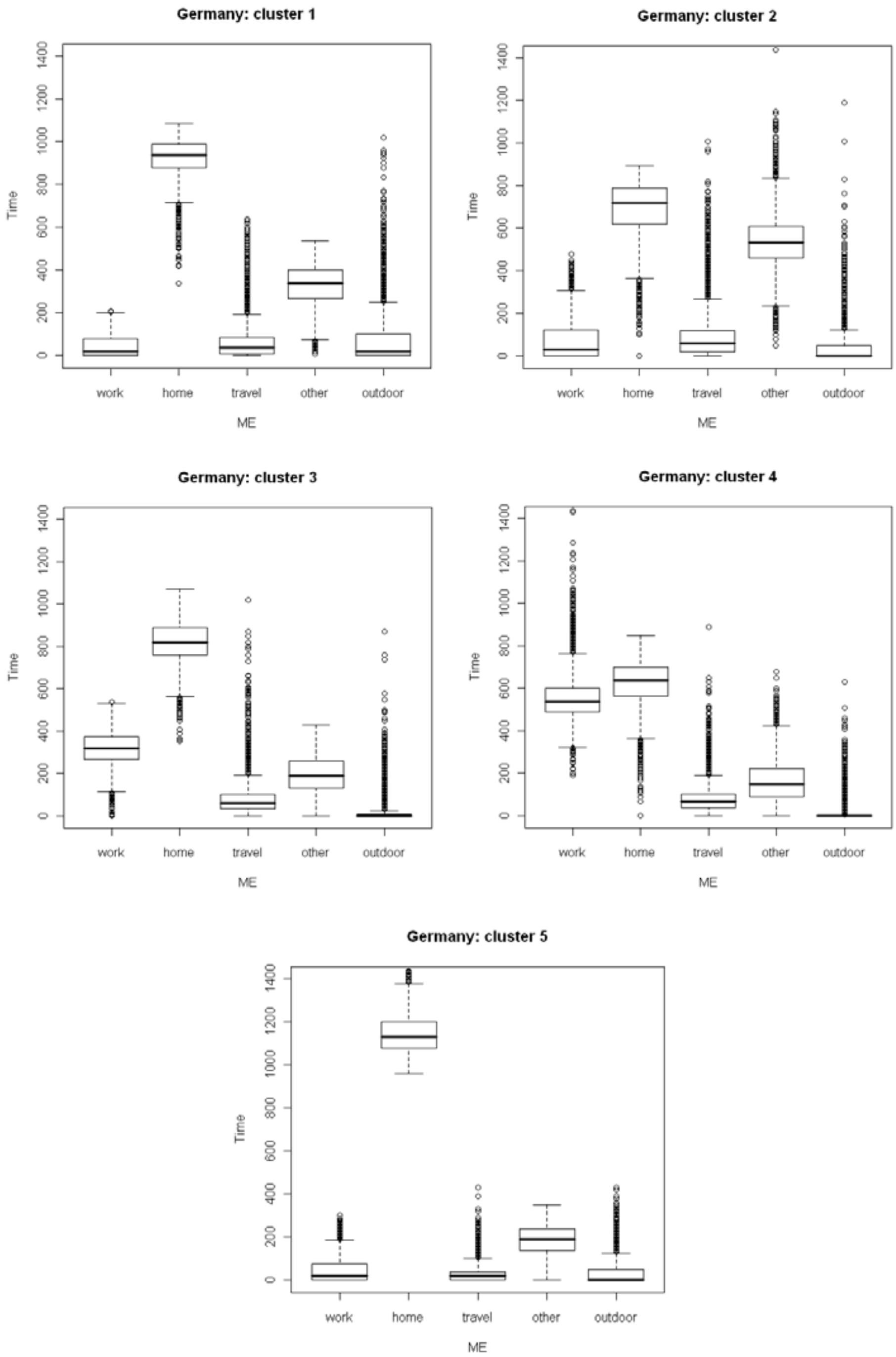

Figure 1. Boxplot of the time spent per ME for the five clusters.

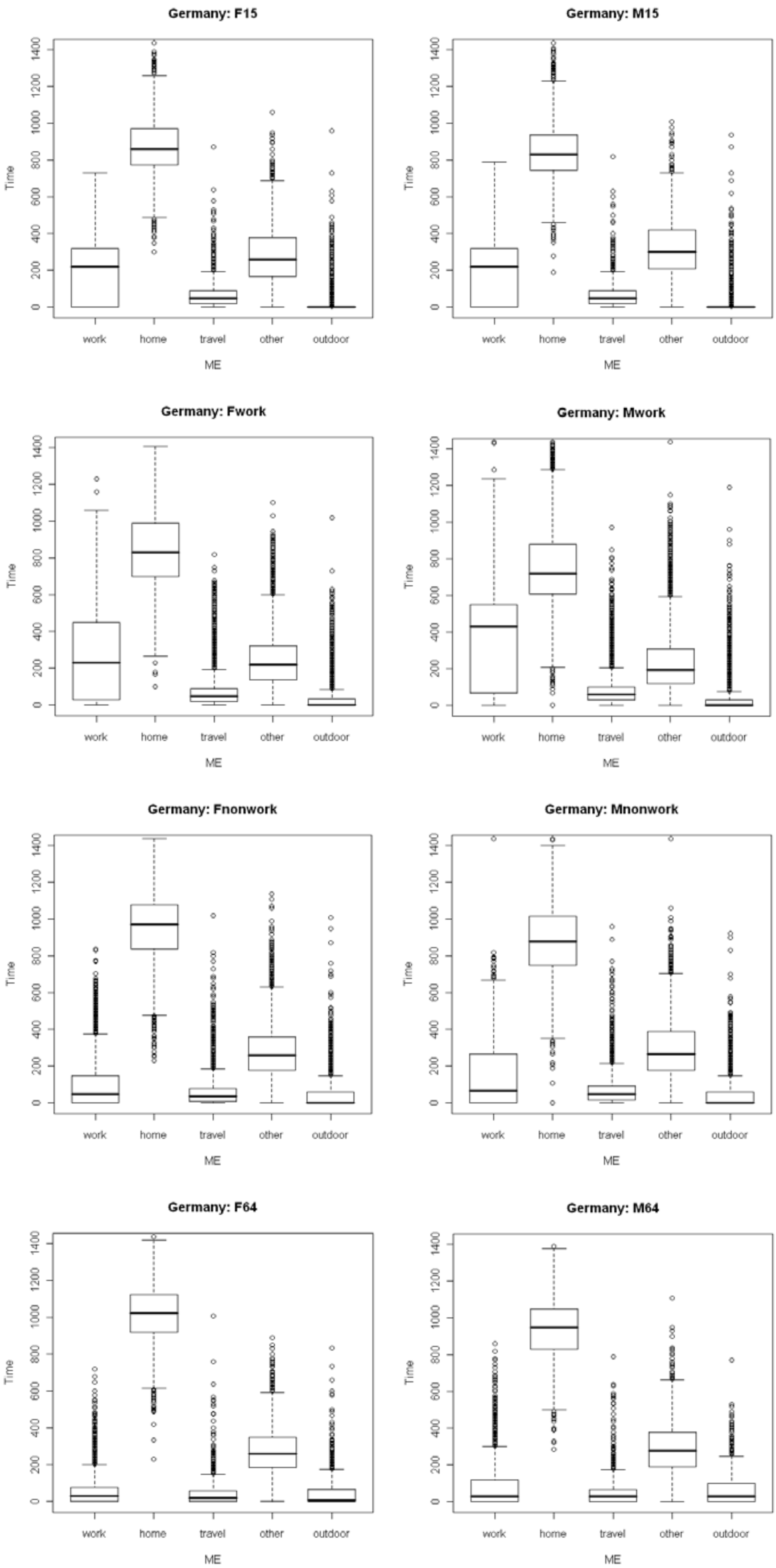

Figure 2: Boxplots of the time spent per ME for the eight demographic groups.

Table 3: Demographic factors for clusters.

Cluster	% women	% men	Average age
1	59 %	41 %	43.8
2	49 %	51 %	37.6
3	57 %	43 %	33.2
4	33 %	67 %	38.8
5	67 %	33 %	44.3

The average age and percentage of women and men per cluster are presented in Table 3. Except for cluster 4 and 5, male and female persons are nearly equally distributed within the clusters. The age averages show obvious and significant differences within a range of 11 years. Nevertheless, all ages can be found in all clusters, making clear differentiation difficult.

Exposure scaling factors

The estimated ESF distributions per population subgroup for Germany are presented as cumulative probability curves in Figure 3 with averages given in Table 44. ANOVA results are presented in Table 55. To assess the uncertainty due to the Monte Carlo sampling, a set of 100 ESF distributions was computed using the same methodology. The comparison of the ESF averages for each group in the 100 simulations shows an average standard deviation of 0.002 and maximal deviation of 0.012 (1.7 %) between single runs.

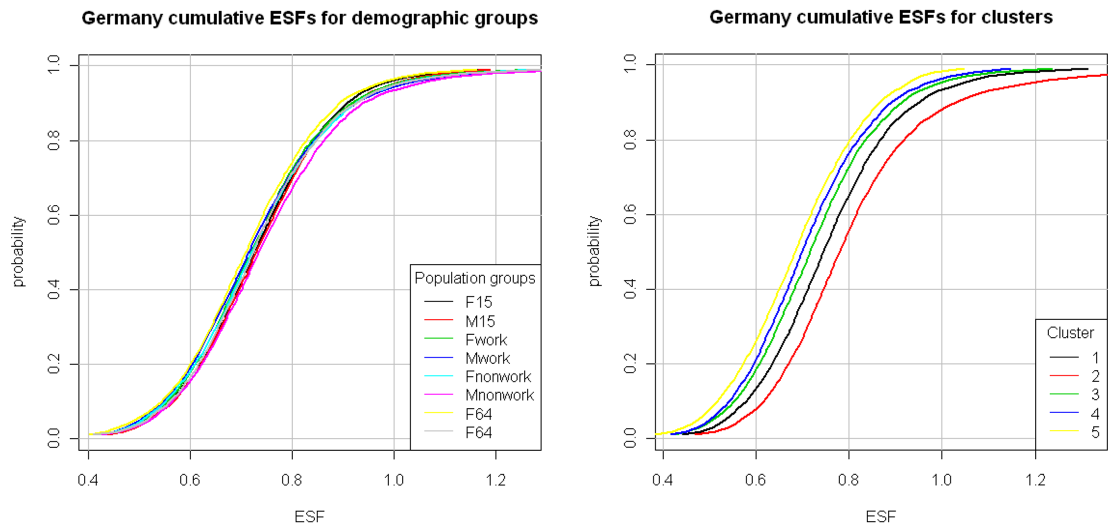

Figure 3: Cumulative probability plots of the Exposure Scaling Factor distributions for the different groups in Germany.

For the ESF averages (see Table 4), the range between the demographic groups is approximately 0.04 with females older than 64 (F64) having the lowest and non-working males between 15 and 64 (Mnonwork) having the highest ESF values. Both groups represent less than 10 % of the population questioned for the MTUS diaries. The largest population groups, working males and females between 15 and 64 years have a similar average ESF of nearly 0.73, which is very close to the population average. Generally, for the population groups less than 15 years and working people between 15 and 64 years, the differences between males and females are very small.

Table 4: ESF averages according to different groups and clusters for Germany.

Groups	Diary %	ESF
F15	5%	0.735
M15	5 %	0.741
Fwork	27 %	0.734
Mwork	30 %	0.733
Fnonwork	15 %	0.737
Mnonwork	8 %	0.754
F64	5 %	0.719
M64	4 %	0.740
Cluster 1	24 %	0.761
Cluster 2	11 %	0.816
Cluster 3	21 %	0.731
Cluster 4	24 %	0.713
Cluster 5	20 %	0.690

As shown in Table 4, the range between the ESF distributions per group is larger for clusters than for demographic groups, which is also shown in Figure 3. The range of ESF averages is approximately 0.13 with cluster 2, which is the smallest, having the highest value. This means that 11 % of the MTUS population have an ESF 11 % higher than the average of the whole population. The average ESF of this cluster is even 18 % higher than for the lowest 20 % of the population (cluster 5).

The variability within the ESF distributions is comparable for clusters and demographic groups. This means that the lower variability of time use per ME for the clusters is not reflected in the ESF results. Compared to the time use distribution, the variability in the infiltration factors seems to be larger and thus influences the variance in the ESF distributions more heavily.

Table 5: ANOVA results for ESF. Results significant at the 0.95 level are marked bold.

Independent variable	Number of groups	p-value
Gender	2	0.12
Employment status	2	**« 0.01**
Age group	3	**0.02**
Demographic groups	8	**« 0.01**
Cluster	5	**« 0.01**

The ANOVA result for gender and ESF is not significant. Demographic groups as well as clusters are significant predictors for ESFs (Table 5). From the single demographic factors (age, gender, employment), employment status is most significant.

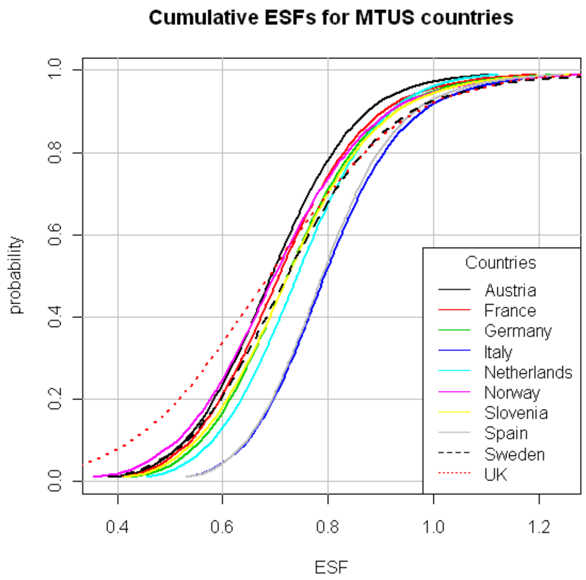

Figure 4: Cumulative probability plot for the ESF distributions in all MTUS countries.

In Figure 4 the results for the ESF distributions for all available MTUS countries is presented. Please note, that for France and Sweden no diaries for individuals younger than 15 years were available so the groups M15 and F15 are missing in the ESF estimation. It is clearly visible that the southern countries Italy and Spain have the highest values, both with ESF means > 0.8. The lowest ESF values were calculated for the UK with an average of 0.70. Germany has a mean of 0.734 which is near the European average.

For all of the countries, non-employed males reached one of the highest values. The largest difference between demographic subgroup ESFs was found in Italy ranging from 0.77 for F64 to 0.85 for Mnonwork. The variances of the ESF distributions are similar for most of the countries as seen in Figure 4. Only the UK, the variance is considerably larger than for the rest of the European countries which is a result of the larger variance in the infiltration factors for Northwestern Europe measured in the EXPOLIS study. Compared to the demographic grouping results presented in Figure 3, the differences between the countries are larger than the differences of population groups within the countries.

DISCUSSION

The presented approach provides a European wide set of Exposure factors for outdoor air using the harmonised European time activity dataset MTUS and a set of infiltration factors derived from EXPOLIS. Including time use data in European scale exposure assessments to evaluate impact of policies is certainly an advancement compared to approaches based on ambient concentration solely. Nevertheless, the influence of behavioral policies on subgroups would probably be rather small as the infiltration factor seems to have a larger impact on the

ESF than the time use. The infiltration factors are a large source of uncertainties due to the geographical extrapolation and adjustments to work and other indoor infiltration factors. Comparisons of different studies showed large differences between the infiltration factor results which have a greater impact on the exposure results than the time use differences between countries and population subgroups. The strong influence of the infiltration factor variance on the exposure results is clearly visible when comparing the country estimates. Also, the variability of the ESF within the clusters and the demographic groups are similar although the time use distributions have lower variability in the clusters which results form the dominating infiltration factor variances. The choice of the different geographically dependent infiltration factors explains most of the differences between the national ESF distributions. It is likely that the differences between the countries, classified into the same geographical regions, is clearly underestimated by applying the same infiltration factors.

Within a country, the differences in ESFs were only affected by the time use statistics. Due to the model composition, each Exposure Scaling Factor is a composition of individual diaries to represent weekday and seasonal variations over a year. Thus, the choice of "diary pool" for each ESF is important to preserve variability and extremes of the population. The results highlighted the difficulties of determining sensible groups, as demographic grouping showed fewer differences between the ESF estimates than the smaller set of clusters. Thus, the extremes for the ESF in the population are diluted in the demographic group results, showing only marginal deviations from the population average.

Hence, regarding the grouping of the population for the ESFs, demographic groups seem not to be the most effective predictors, although they are still significant. It has to be acknowledged here that the ANOVA results have to be treated with care, as significance is easily reached with a large number of samples. Clusters on the other hand seem more meaningful in terms of depicting differences in exposure and policy changes, but are hard to describe with available demographic variables. To complete the exposure analysis, outdoor air quality on a grid is combined with the population distribution and the subgroup ESFs on a grid. If we cannot describe clusters with the population parameters available for these grids, *i.e.*, age, gender, employment, clusters are of virtually no use for the exposure modelling chain. Clearly, the finer the granularity of distinction is, the narrower the time distribution and thus the better the ESF estimates are. In parallel, the size of the groups gets smaller with increasing distinction. The balance between a representative group size and a maximal reduction of variance led us to choose the current grouping constellation. Further investigations are necessary to identify variables that are more homogeneously distributed within the clusters and are thus useful for determination of clusters and ESF changes. Also the activity set of the MTUS data revealed difficulties in classification for exposure studies, as location information were missing. Adding the location to the time activity data could help better differentiate exposure and more easily evaluate potential policy changes.

CONCLUSION

In the presented exposure modelling approach, we combined harmonised time use data and infiltration factors to estimate Exposure Scaling Factors for population subgroups and assess changes due to policies. Therefore, common demographic grouping techniques were compared with activity –based grouping to assess differences in exposure for susceptible groups. As anticipated the time activity-based grouping shows clearer results in the ESF and thus allow a better identification of policy changes and extreme exposure impacts. We demonstrated that for behavioural and infiltration policies, individual activity data is inevitable to assess the impacts. Nevertheless, there is a need for more realistic assumptions regarding behavioral policies to enhance the full chain modelling.

For this time use driven exposure model, variables to describe time activity characteristics in terms of exposure relevance need to be identified, as these are inevitable for exposure characterisation. The demographic factors used in this study are not the best choice as their performance compared to the time use driven approach was rather bad. On the other hand, the activity clustering is not necessarily a better alternative as clusters cannot be easily transferred form national time use statistics are available to smaller, *e.g.* urban or rural, regions where the population composition can differ. The future work demands identification of better factors from time use research that can also be applied for exposure modelling.

ACKNOWLEDGEMENTS

This study was realised in the HEIMTSA project funded by the European Union Sixth Framework programme contract no. 036913. We thank Dr. Kimberly Fisher for providing us access to the Multinational Time Use Study datasets.

REFERENCES

Dockery DW, Pope CA, Xu X, Spengler JD, Ware JH, Fay ME, Ferris BG and Speizer FE. (1993) *An association between air pollution and mortality in six US cities*. N Engl J Med, 329 (24):1753-1759

Gulliver J. and Briggs D.J. (2004). *Personal exposure to particulate air pollution in transport microenvironments.* Atmos Environ, 38(1):1:8.

Hänninen OO, Lebret E, Ilacqua V, Katsouyanni K, Künzli N, Srám RJ and Jantunen M (2004) *Infiltration of ambient PM25 and levels of indoor generated non-ETS PM25 in residences of four European cities*. Atmospheric Environment; 38(37):6411-6423

Hänninen OO, Hoek G, Mallone S, Chellini E, Katsouyanni K, Kuenzli N, Gariazzo C, Cattani G, Marconi A and Jantunen M (2009) *Seasonal Patterns in Ventilation and PM Infiltration in European Cities*: *Comparison of available studies* 7th International Conference on Air Quality – Science and Application (Air Quality 2009), Istanbul, pp.24-27

Hartigan JA and Wong MA (1979) *A K-means clustering algorithm.* Applied Statistics 28: 100–108

Hoek G, Kos G, Harrison R, de Hartog J, Meliefste K, ten Brink H, Katsouyanni K, Karakatsani A, Lianou M, Kotronarou A, Kavouras I, Pekkanen J, Vallius M, Kulmala, M, Puustinen A, Thomas S, Meddings C, Ayres J, van Wijnen J and Hameri K (2008) *Indoor-outdoor relationships of particle number and mass in four European cities.* Atmospheric Environment; 42(1):156-169

IPCS Risk Assessment Terminology Part 2: IPCS Glossary of Key Exposure Assessment Terminology IPCS Harmonization Project Geneva, World Health Organization

Kaur S., Nieuwenhuijsen MJ and Colvile RN (2005). *Pedestrian exposure to air pollution along a major road in Central London, UK.* Atmos Environ 39(38):7303:7320.

Klepeis NE (1999) *An introduction to the indirect exposure assessment approach: modeling human exposure using microenvironmental measurements and the recent National Human Activity Pattern Survey*. Environ Health Perspect, 102(2):365–374

Klepeis, NE, Nelson W, Ott W, Robinson J, Tsang A, Switzer P, Behar J, Hern S and Engelmann W (2001) *The National Human Activity Pattern Survey (NHAPS): a resource for assessing exposure to environmental pollutants.* J Exposure Analysis and Environmental Epidemiology, 11:231–252

Koutrakis P, Briggs S and Leaderer B (1992) *Source Apportionment of Indoor Aerosols in Suffolk and Onondaga Counties, New York.* Environ Sci Technol, 26(3):521–527

Krewitt W; Heck T; Droste-Franke B; Trukenmüller A; Friedrich R (1999) *Implementation of the EcoSense model in Brazil/Latin America*, IAEA Contractual Service, Final Report, Agreement No 98CL9254

Kuhn A; Loh M; Gerharz L; Torras S; Yang A; Klotz V; Fantke P; Bartonova A; Friedrich R (2009) *Individual Exposure Modelling in an Integrated Decision Support System for air quality management*. 21st conference of the International Society for Environmental Epidemiology (ISEE). August 25 – 29, 2009. Dublin, Ireland

MacQueen, J (1967) *Some methods for classification and analysis of multivariate observations In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability.* eds L M Le Cam & J Neyman, **1**, Berkeley, CA: University of California Press, pp 281–297

McCurdy T and Graham SE (2003) *Using human activity data in exposure models: Analysis of discriminating factors.* J Exposure Anal Environ Epidemiol, 13:294-317

MTUS (2009) Multinational Time Use Study, Versions World 553, 580 and 60 (released 26 March 2009) Created by Jonathan Gershuny and Kimberly Fisher, with Evrim Altintas, Alyssa Borkosky, Anita Bortnik, Donna Dosman, Cara Fedick, Tyler Frederick, Anne H Gauthier, Sally Jones, Jiweon Jun, Aaron Lai, Qianhan Lin, Tingting Lu, Fiona Lui, Leslie MacRae, Berenice Monna, José Ignacio Giménez Nadal, Monica Pauls, Cori Pawlak, Andrew Shipley, Cecilia Tinonin, Nuno Torres, Charlemaigne Victorino, and Oiching Yeung Centre for Time Use Research, University of Oxford, United Kingdom http://wwwtimeuseorg/mtus/

Özkaynak H, Palma T, Touma JS, Thurman J (2008) *Modeling population exposures to outdoor sources of hazardous air pollutants*. J Exp Anal Environ Epidemiol 18:45–58

Pope CA, Burnett R, Thun M, Calle E, Krewski D, Ito K and Thurston G (2002) *Lung Cancer, Cardiopulmonary Mortality, and Long-term Exposure to Fine Particulate Air Pollution.* Journal of the American Medical Association, 287(9):1132–1141

Zidek J, Shaddick G,White R, Meloche J, Chatfield C (2005) *Using a probabilistic model (pCNEM) to estimate personal exposure to air pollution*. Environmetrics, 16:481–93