Generating WUDAPT's Specific Scale-dependent Urban Modeling and Activity Parameters: Collection of Level 1 and Level 2 Data

Linda See[1], Jason Ching[2], Valéry Masson[3], Johannes Feddema[4], Gerald Mills[5], Marina Neophytou[6], Michael Foley[5], Martin O'Connor[5], Grega Milčinski[7], Marko Repse[7], Christoph Perger[1], Martina Duerauer[1], Steffen Fritz[1] and Benjamin Bechtel[8]

1 *International Institute for Applied Systems Analysis, Laxenburg, Austria, see@iiasa.ac.at*
2 *Institute for the Environment, University of North Carolina, Chapel Hill NC, USA, jksching@gmail.com*
3 *Centre National de Recherches Météorologiques, Météo France, Toulouse, France, valery.masson@meteo.fr*
4 *Department of Geography, University of Kansas, Lawrence KS, USA, feddema@ku.edu*
5 *School of Geography, Planning and Environmental Policy, University College Dublin, Dublin, Ireland, gerald.mills@ucd.ie*
6 *School of Engineering, University of Cyprus, Nicosia, Cyprus, neophytou@ucy.ac.cy*
7 *Sinergise, Ljubljana, Slovenia, grega.milcinski@sinergise.com*
8 *Institute of Geography, University of Hamburg, Hamburg Germany, benjamin.bechtel@uni-hamburg.de*

dated : 22 June 2015

1. Introduction

The Local Climate Zone (LCZ) framework of Stewart and Oke (2012) has been proposed as a generic, standardized classification of urban landscapes into different atmospheric regimes in support of the World Urban Database and Access Portal Tools (WUDAPT) initiative (Ching 2013, Ching *et al.* 2014, Mills *et al.* 2015). A workflow for the development of an LCZ map, which can be applied to any city worldwide, has been outlined previously in Bechtel et al. (2015). We refer to the LCZ characterization of a city as 'level 0' data collection. Since each LCZ is characterized by a range of values for a number of different parameters that are required by climate and weather models, it is possible to use the 'level 0' data as an initial starting point for determining these model parameters. However, this means that only one value is applied to each LCZ and this will either be the mean value of the range or a refinement based on local knowledge.

In order to determine the actual value of the parameter from the LCZ parameter range, additional data are required. LCZs can be used as a sampling frame to collect a representative sample of a variable that would allow more refined parameter values to be estimated. The sample data can be collected through the use of mobile and Web 2.0 technology in combination with crowdsourcing (Howe 2006) as outlined below. In this context, we refer to crowdsourcing as the use of urban experts and their networks of people, e.g. students, to collect the data, which will become part of the WUDAPT open database.

The underlying distribution of the sample data collected by the crowd or urban experts could then be used to distribute parameter values across the gridded area of a LCZ for use in a climate or weather model. We refer to this sampling approach as 'level 1' data collection since it provides more refined estimates of LCZ parameters but it is still an approximation. The most comprehensive or wall-to-wall estimate of LCZ parameters is referred to as 'level 2' data collection. Data at 'level 2' can be collected systematically if there are sufficient resources; ancillary databases (e.g. building footprint files) can also be used to extract some 'level 2' variables automatically.

This paper outlines two proposed data collection methods for 'level 1' variables and discusses possible sources of data for 'level 2'.

2. Variables for collection

The LCZ parameters can be divided into four main types, i.e. surface cover, geometry, materials and function (or urban metabolism). These can be further broken down into specific variables of interest as shown in Table 1, which can then be used to derive values for the LCZ parameters.

Table. 1 Variables of urban form and function for climate and weather models

Feature	**Variables of interest**
Cover	Land cover, vegetation type, vegetation organization
Geometry	Building height, width of streets, contiguous or isolated buildings, roof geometry
Material	Wall type, roof type, window type, road materials, window fraction on the wall, colour/albedo
Function	Building use, irrigation, road type, temperature setting, occupancy, air conditioning, shutters or shading, window opening, building age, building renovation post 1990

3. Level 1 data collection

3.1 Cover

One key input to climate and weather models is the areal fraction of different land cover types. For the pure land cover classes of LCZs 11 to 16, e.g. grassland, water, bare soil, etc., the information is available directly from the LCZ classification itself. However, for the urban LCZ classes 1 to 10, the areal fraction requires more information. One source of this information is very high resolution satellite imagery such as that found on Google Earth. Since areal fractions are dependent upon the underlying grid, where different climate and weather models will operate on different grid sizes and grid placements, a tool is needed that allows for collection of point data on a regularly spaced grid. Users can then calculate the areal fraction based on their own model set up, which is an envisaged function of the WUDAPT portal (Ching *et al.* 2015).

The Geo-Wiki tool for crowdsourcing and validation of land cover was modified for this purpose as shown in Figure 1. A regularly spaced grid of points was laid on top of the city of Dublin, with a point every 30 m. This density was chosen to create a comprehensive coverage from which a more optimum spacing can be determined for data collection in other cities in the future. The points that fell into LCZs 1 to 10 were then input to Geo-Wiki and users were asked to determine the land cover in the center of the blue ellipse from the choices on the right (Figure 1).

Fig. 1 The Geo-Wiki tool for collection of land cover types in Dublin on a point sample.

In total there were 297,149 points classified in Dublin from which fractional land cover can be calculated. Initial processing of the results indicated that the fractional land cover for LCZs 1 to 10 generally falls within the ranges originally suggested by Stewart and Oke (2012). Work is ongoing to examine the variation in fractional land cover across LCZs and to determine the spacing and number of samples for which land cover data should be collected for other cities. Once this work is completed, Geo-Wiki will be set up to collect land cover data for other cities for which 'level 0' data already exist. A mobile version for tablets will be developed in the future, which may speed up the classification process.

3.2 Geometry, materials and function

Some of the geometry, materials and function data can be collected from photographs of buildings. Photographs are available from numerous sources including Google StreetView, Panoramio, Flickr, Instagram etc. Google StreetView has the advantage of comprehensive coverage in many cities globally while other photograph repositories may be useful for supplementing locations where no streets are present.

A data collection tool for WUDAPT was developed in Geopedia, which is an online GIS (Geographic Information System) developed by Sinergise. The sample generated for land cover data collection via Geo-Wiki was laid over a typical building footprint file from OpenStreetMap (OSM - downloaded from GeoFabik.de). Although building footprints in OSM are often not complete, they are generally sufficiently detailed for sampling across LCZs. An example is shown in Figure 2 where the red buildings are those from the sample. A user then clicks on a building and a window appears on the left hand side. Once editing is enabled, the user can enter information (Figure 3a) about the Google StreetView photograph (Figure 3b), if a photograph is available. The information collected includes:

- Height of the building (indicated by number of stories);
- Dominant building material (brick, wood, concrete, stone, glass, mud, other);
- Pitch of the roof (flat, low, high);
- Type of building (residential, commercial, public, warehouse, other);
- Whether the building is detached or not;
- The proportion of the front of the building that is windows;

- The age of the building;
- A rough estimate of the albedo.

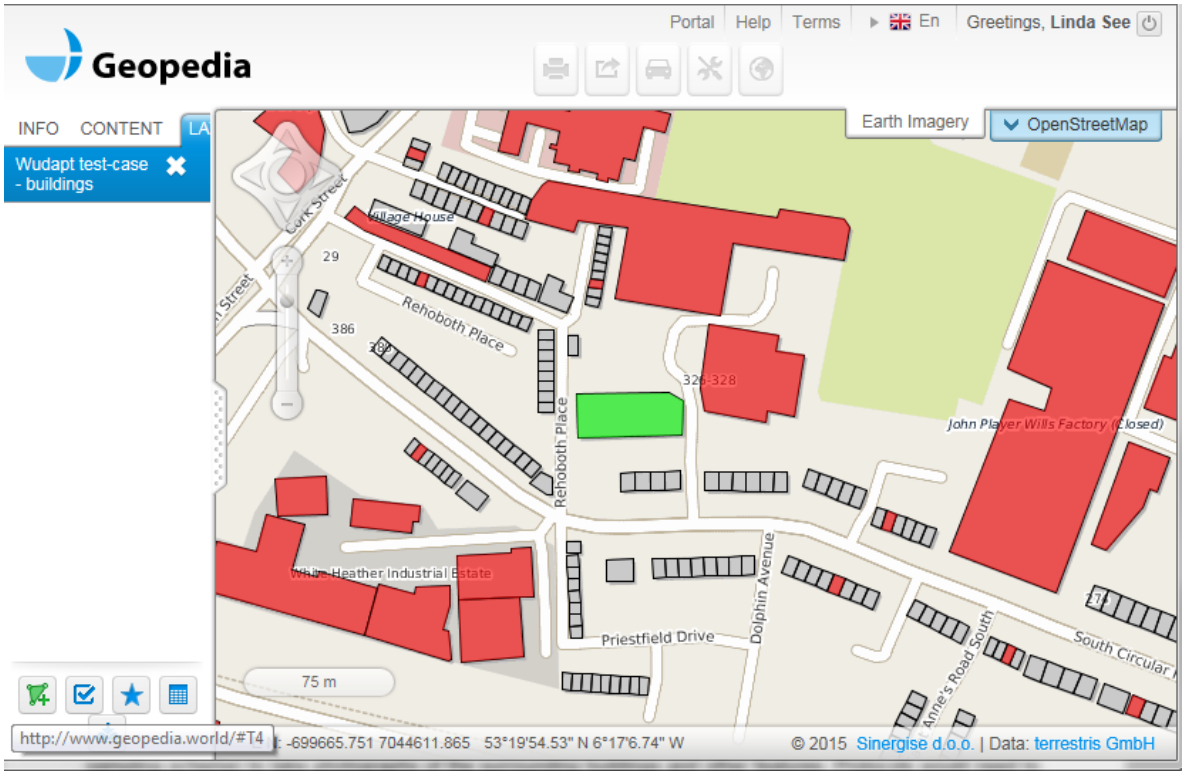

Fig. 2 Buildings from OpenStreetMap displayed in Geopedia. Buildings in red are in the sample for data collection, green are those for which data have already been collected and grey are outside the sample.

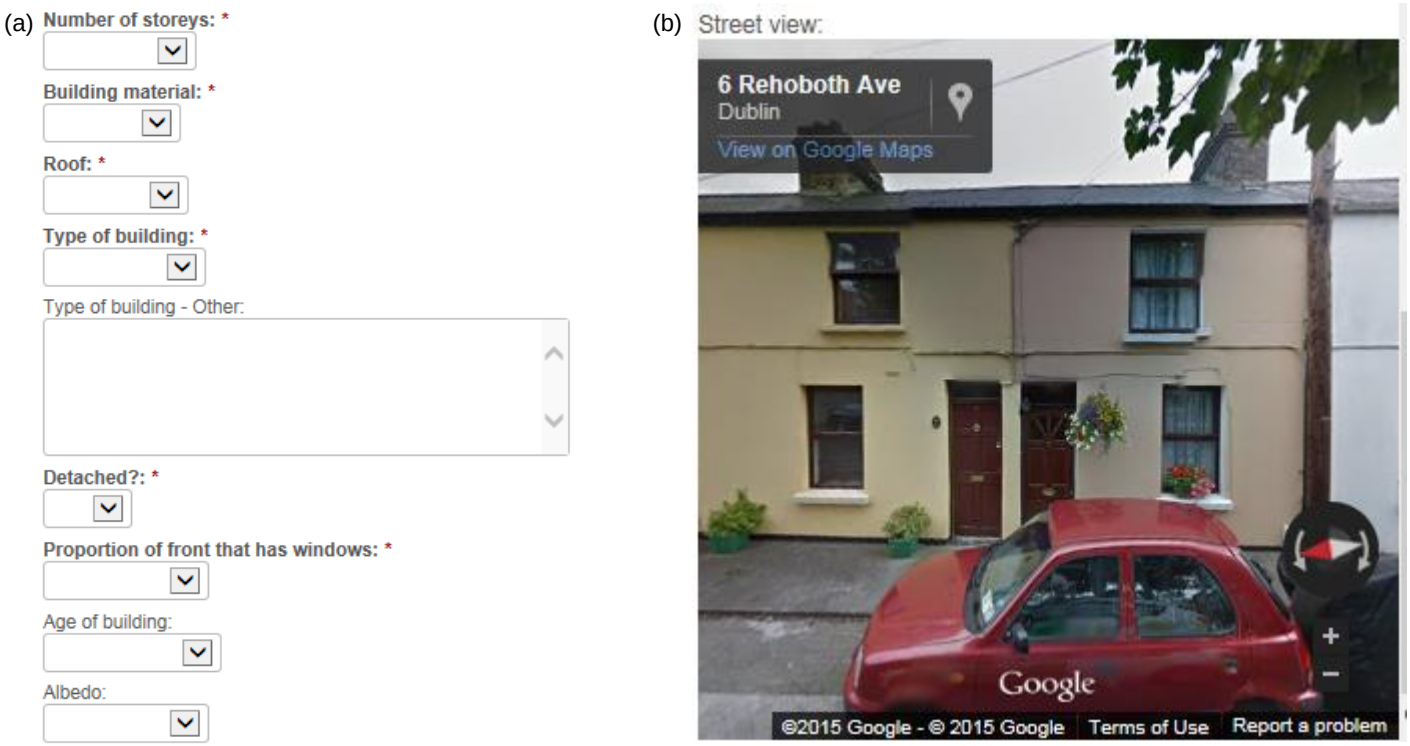

Fig. 3 (a) Data collection interface for the building shown in the (b) Google StreetView image

A second type of application could be built whereby individuals are sent to specific locations (based on a sampling scheme) to take photographs of the surrounding buildings and other features. Protocols would need to be developed, e.g. how the photographs should be taken to optimally capture the variables of interest. The photographs could be linked to a simple form for collecting the variables from Table 1, including ones that might be difficult to extract from viewing just the photograph alone.

4. Level 2 data collection

4.1 Cover

Area fraction of land cover could be calculated automatically in the WUDAPT portal for any grid size if certain data sets were available, e.g. a building footprint file, a road network, a map of urban greenspace, an impervious surface layer and other land cover types, e.g. from the European Environment Agency's Urban Atlas product for European cities. Using a combination of these layers, the calculation of area fraction of land cover could be undertaken using common geoprocessing functions like those found in a GIS.

4.2 Geometry, materials and function

For some cities, building height information might be obtained from planning departments while a building footprint file in combination with a street network would allow for the automatic calculation of canyon widths and the density of buildings. Albedo can be obtained from Landsat imagery (and other satellite sensors) but will only capture albedo of the roof and ground, and not the vertical dimension, e.g. fronts of buildings.

Level 2 data are not available for function (Table 1) as there are no remotely sensed products that can measure the variables characterizing human metabolism. These variables would need to be collected through sampling (level 1) or estimated via an empirical relationship. Some cities may have databases on variables such as building age but these will be in the minority and, most likely, they will not be openly available for use by the WUDAPT community.

5. Conclusions and next steps

This paper has addressed tools and methods for collection of 'level 1' and 'level 2' data. The main focus in 'level 1' has been on the development of an application for collecting land cover data based on Geo-Wiki and the collection of different variables that capture the geometry, materials and function of the urban landscape using Geopedia. The next steps will involve the use of the Geo-Wiki tool to collect the land cover of urban LCZs for around 20 cities where LCZ maps have already been created. The second tool will first be tested on Dublin and then rolled out to the other cities after the quality of information collected from the photographs is assessed. A mobile app will also be developed for collecting data on the ground, which will complement the online collection of information.

For more information and continuing developments in 'level 1' and 'level 2 data' collection, see the WUDAPT website (http://www.wudapt.org).

Acknowledgment

This work has been partly supported by the EU FP7 ERC grant CrowdLand (No. 617754).

References

Bechtel, B., Alexander, P.J., Böhner, J., Ching, J., Conrad, O., Feddema, J., Mills, G., See, L., Stewart, I., 2015: Mapping Local Climate Zones for a worldwide database of the form and function of cities. *ISPRS International Journal of Geo-Information*, **4**(1), 199–219.

Ching, J.K.S., 2013: A perspective on urban canopy layer modeling for weather, climate and air quality applications. *Urban Climate*, **3**, 13–39.

Ching, J., Mills, G., See, L., Bechtel, B., Feddema, J., Hanna, A., Milčinski, G., Masson, V., Neophytou, M., Mitra, C., O'Connor, M., Pereira, N., Steeneveld, G.-J., Stewart, I., Wang, X., Alexander, P., Foley, M., Gal, T., 2015: The portal component, strategic perspectives and review of tactical plans for full implementation of WUDAPT. *In*: *Proceedings of the ICUC9*. Presented at the ICUC9, Toulouse, France: Meteo France.

Ching, J., Mills, G., See, L., Bechtel, B., Feddema, J., Oleson, K., Stewart, I., Neophytou, M., Chen, F., Wang, X., Hanna, A., 2014: WUDAPT: Facilitating advanced urban canopy modeling for weather, climate and air quality applications. *In*: *Proceedings of the 94th American Meteorological Society Annual Meeting*. Presented at the A, Atlanta, Georgia: AMS.

Howe, J., 2006: The rise of crowdsourcing. *Wired Magazine*, **14**(6), 1–4.

Mills, G., Ching, J., See, L., Bechtel, B., Feddema, J., Masson, V., Stewart, I., Neophytou, M., O'Connor, M., Chen, F., Martilli, A., Grimmond, S., Alexander, P., Foley, M., Gal, T., Wang, X., Mitra, C., Pereira, N., Steeneveld, G.-J., 2015: Introduction to the WUDAPT project. *In*: *Proceedings of the ICUC9*. Presented at the ICUC9, Toulouse, France: Meteo France.

Stewart, I.D., Oke, T.R., 2012: Local Climate Zones for Urban Temperature Studies. *Bulletin of the American Meteorological Society*, **93**(12), 1879–1900.