9.1 The WUDAPT Project: Status of Database and Portal Tools

G. Mills[1], B. Bechtel[2], M. Foley[1], Ching.J.[3], L See[4], J. Feddema[5]

[1]University College Dublin, Ireland, [2]University of Hamburg, Germany, [3]University of North Carolina, USA, [4]IIASA, Austria, [5]University of Victoria, Canada. Corresponding Author: Gerald Mills, School of Geography, Newman Building, UCD, Belfield Dublin 4, Ireland. Gerald.Mills@ucd.ie

1. Introduction

Urbanisation is associated with the intense transformation of the landscape through: the replacement of natural surface cover; the use of manufactured materials that have distinct thermal and radiative properties (and are mostly impermeable) and; the imposition of buildings close to one another. Together, this process creates an urban landscape with a complex and heterogeneous three-dimensional 'surface'. In addition they are places of focussed energy and material use and of waste generation. Consequently, cities are drivers of environmental changes at a hierarchy of scales and represent a spatial nexus where the aggregated actions of individuals exert significant control on global climate changes (e.g. Oke et al., 2017). However, despite their global significance, we know very little about the spatial composition of most cities and what we do know is often too coarse or inconsistent to undertake scientific inquiry or make meaningful comparisons.

This paper introduces the World Urban Database Access Portal Tool (WUDAPT) project, which is creating a global database of cities using: freely available software and data; the local knowledge of urban experts and; critically, a common framework for describing urban landscapes (Mills et al., 2015)

2. Literature Review

Although cities are extremely heterogeneous, this property is largely a function of scale. At the micro-scale, the variety of urban features (buildings, parks, roads, trees, etc.) and their spatial juxtaposition creates great surface diversity that is difficult to generalise. At a local-scale however much of this diversity is repeated such that cities can be characterised by distinct patches associated with typical assemblages of these features; these areas are often described using varying categories of land-cover and land-use. In many cities, these categories are correlated and are linked to land-use management policies. Although there are no universally accepted categories of land use and land cover, there are national and regional databases that employ common descriptors to map cities. The European Environment Agency's (EEA) **Urban Atlas**, for example, is a set of self-described reliable, inter-comparable, high-resolution land use maps for functional urban areas; currently there are 697 FUAs in the database[1].

At a global scale, the available information is much reduced and the available urban databases are based on delineating urban extents (or footprints), without characterising the internal structure. Estimates of extent based on population size or density rely on census counts based on administrative units that do not correspond to the urbanised landscape. Remotely sensed data can offer a consistent means of identifying the urban cover globally but the output depends on the source of information, specifically the spectral and spatial resolution of the sensor. Moreover, estimates of urban extent must grapple with issues of spatial contiguity and the interspersing of signals from natural cover and built cover within the urban area. Potere *et al.* (2009) examined eight global urban maps to determine which was the most accurate in assessing urban extent; the dataset derived from the MODIS satellite at a resolution of 500 m was the most accurate of those compared. Overall however, the existing data on cities internationally has limited value as it uses a binary classification (urban/natural) with no details on the physical structure or functional layout of cities.

Ideally, global urban data would inform us of the spatial variation of climate relevant variables related to form and function across cities. Form describes the physical structure of the city including the impervious (or vegetated) surface cover, the construction materials used (fabric) and the three-dimensional built geometry. Function refers to the intensity of human activities and the concomitant fluxes of materials, energy, water etc. that result in waste emissions into the overlying air, adjacent watercourses or nearby soil. In a modelling framework, these properties would be described by urban canopy parameters (UCPs) in that they describe the corrugated urban surface in terms of building heights, green area fraction, impervious surface cover, etc. (Ching, 2013).

If this information were available at sub-urban scales it would be an invaluable resource to support climate and environmental research, energy management, risk assessment and policy design and implementation (e.g. Jackson et al., 2010). The WUDAPT project seeks to provide these data for cities worldwide. Critically, these data are gathered using protocols to ensure consistency and quality. It takes a pragmatic approach to the acquisition of useful data which is classified according to level of detail. The lowest level (L0) provides a basic physical geography of cities based on the Local Climate Zone (LCZ) concept; higher levels (L1&2) seek to capture detailed information on urban elements, that is the buildings, roads, trees, etc. and their variation across the urban landscape (Mills et al., 2015).

---

[1] http://www.eea.europa.eu/data-and-maps/data/urban-atlas

In the following we present the status of the database and the portal tools before discussing the next steps.

3. WUDAPT data

A basic physical geography (L0) of each city consists of an LCZ map that is linked to typical values for key UCPs. L0 data will provide a sampling framework for gathering more detailed urban data (L1&2) in future developments of WUDAPT. Here the focus is on the workflow for generating L0 data, assessments of quality and results for a select city.

The LCZ scheme categorises landscapes into 17 basic types, 10 of which are urban (Stewart and Oke, 2012). The scheme was designed to describe the controls on the near-surface air temperature that give rise to the urban heat island (UHI) effect. These controls are represented by value ranges indicating mean building height, impervious surface fraction, anthropogenic heat flux, etc. (Table 1) for 'homogeneous' local scale (> 1km^2) areas known as **neighborhoods**. The LCZ scheme has been used successfully in a number of projects and cities; for example to design and analyse the urban thermal effect (e.g. Alexander and Mills, 2014), discriminate among energy exchanges in cities located in different climates (Alexander et al, 2016) and has been used to examine the climatic implications of urban development paths (Alexander et al, 2015).

Many of the LCZ variables are also core variables for atmospheric models and can provide guidance for selecting model parameters. Hence, a map of a city that identifies its neighbourhoods in terms of LCZ types is also a map of landscape variables suited to some urban climate studies.

LCZ type	(H/W)	H (m)	λ_b	λ_P
Compact high-rise	>2	>25	40-60%	<10%
Compact mid-rise	0.75-2	10-25	40-70%	<20%
Compact low-rise	0.75-1.5	3-10	40-70%	<30%
Open high-rise	0.75-1.25	>25	20-40%	30-40%
Open mid-rise	0.3-0.75	10-25	20-40%	20-40%
Open low-rise	0.3-0.75	3-10	20-40%	30-60%
Lightweight low-rise	1-2	2-5	60-90%	<30%
Large low-rise	0.1-0.3	3-4	30-50%	<20%
Sparsely built	0.1-0.25	3-4	10-20%	60-80%
Heavy industry	0.2-0.5	5-15	20-30%	40-50%

Table 1: Local Climate Zone (LCZ) types and typical value ranges for selected variables: height to width ratio (H/W), mean height of buildings (H), plan fraction of buildings (λb) and plan fraction pervious area (λP) . Source: Stewart and Oke, 2012.

3.1 Methodology

The workflow for generating LCZ maps consists of four steps:

1. A city of interest is identified and images from the Landsat8 satellite, which are freely available, are collated.
2. The researcher identifies parts of the city that typify the LCZ types; these areas are digitised in Google Earth and used as training and evaluation areas in image analysis.
3. GIS and Remote Sensing software (SAGA) uses training areas from step 2 to automatically classify the Landsat images from step 1 into LCZ types.
4. The urban expert reviews the automated classification and steps 2-3 are repeated until the expert is satisfied.

LCZ values are assigned using a supervised Random Forest (RF) classification algorithm implemented in SAGA software. RF is a non-parametric classifier that creates an ensemble of classification 'trees' based on a randomly drawn samples of surface characteristics extracted from the ROI from both within the training areas and outside. The RF classifier uses these probability trees to assign LCZ types to each cell based on likelihood. The accuracy and computing efficiency of this classifier are considered outstanding even for large data- and feature-sets. Further details of the classification process can be found elsewhere (Bechtel et al., 2015).

Fig. 1 shows the LCZ map generated for the region around Milan, Italy. The map depicts a densely built city core surrounded by low-rise open landscape. The development of the city along the main transport routes is clear as is the incorporation of the outlying satellite towns.

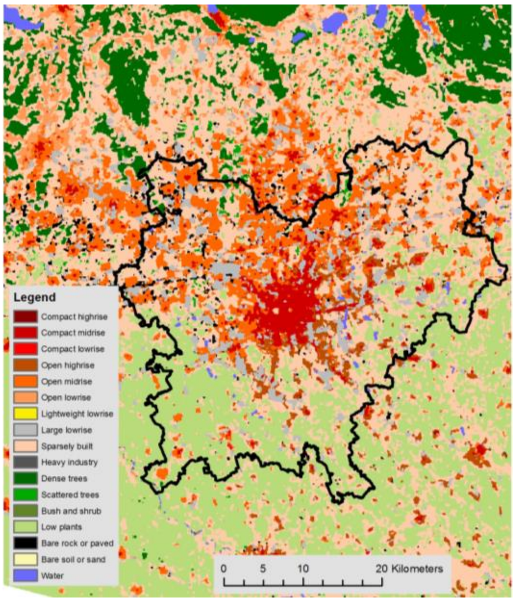

Fig 1. WUDAPT level 0 data for Milan, Italy

3.2 Quality control

Inevitably the LCZ protocol results in errors that must be managed through an evaluation process. The most critical aspect of error management is the quality and quantity of the training areas (TAs), which are used to guide the RF classifier and to evaluate the reliability of the result. These have to be acquired for each city, since the spectral properties of each LCZ will vary with building materials, the bio-physical background and the atmospheric and surface conditions at time of image acquisition. To minimise any errors, considerable effort is placed on the training of the experts and the independent evaluation of LCZ maps.

The statistics that support the decision on inclusion of an LCZ map are based mainly on measures of reliability, that is, the degree to which the final product is sensitive to the selection of TAs. The error dataset is created using a bootstrap method that divides the available TAs into two parts: a training set used by the RF classifier and; an evaluation set used to compare predicted LCZ values against those assigned by the urban expert. This division is performed 25 times by randomly assigning TAs into training and evaluation sets that are then summarised. Measures of consistency include:

1. *Overall accuracy* (OA), that is, the proportion of predicted cell values that match with the values assigned by urban experts. This statistic can be calculated for all LCZ types, the urban LCZ types (LCZ 1-10) and the combined urban (LCZ 1-10) and natural (LCZ A-D & F types).
2. *The Kappa (κ) coefficient* which measures both the successes and failures of the predicted LCZ values.

As a general rule, if these statistics exceed 0.6, we consider that the information in the map is sufficiently robust that it can enter the WUDAPT database; in other words, further work can improve the map but it meets our criteria for acceptance. In the case of Milan (Fig. 1), the OA score is 0.66 and κ=0.60. In a further assessment, the OA measure was weighted to account for the climatic impact of mis-classification (that is, the consequence of LCZ 1 being classified as LCZ A is worse than classified as LCZ 2, for example); here the Milan score is 0.89. Of course the map products are not static and improvements in the TAs will lead to improved versions of Level 0 data in the database.

Independent assessment of the quality of L0 data can be done where there is other comparable data on the urban surface. For Milan (and other European cities) such data are available in the Urban Atlas. Fig. 2 was generated by converting LCZ codes (Fig. 1) to classes of pervious cover as shown in Table 1.

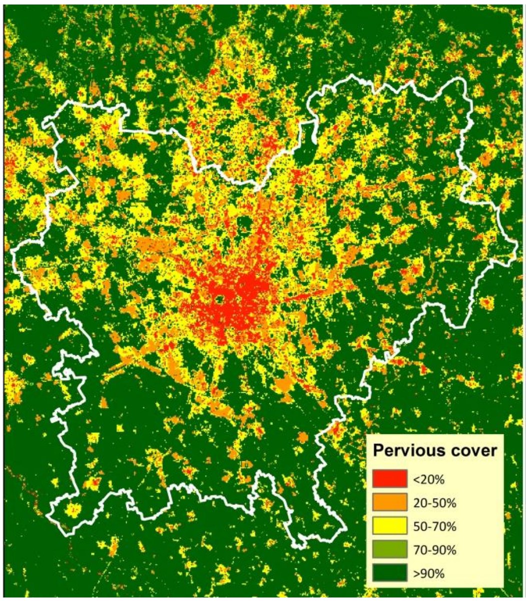

Fig 2. Pervious cover based on LCZ types (Fig. 1) and associated percent pervious surface (λp) from Table 1.

The replacement of natural cover by impervious manufactured materials to make roads, car parks, buildings etc. is a ubiquitous outcome of urbanisation that has profound implications for the surface energy and water budgets, with consequences for the local climate (e.g. urban heat island formation) and hydrology (e.g. urban flooding), respectively. As such, it is a key parameter for many models. A similar variable in the Urban Atlas is the sealing cover (%) based on the following categories:

- Dense Urban Fabric (50% - 80%)
- Medium Density Urban Fabric (30% - 50%
- Low Density Urban Fabric (10% - 30%)
- Very Low Density Urban Fabric (< 10%)

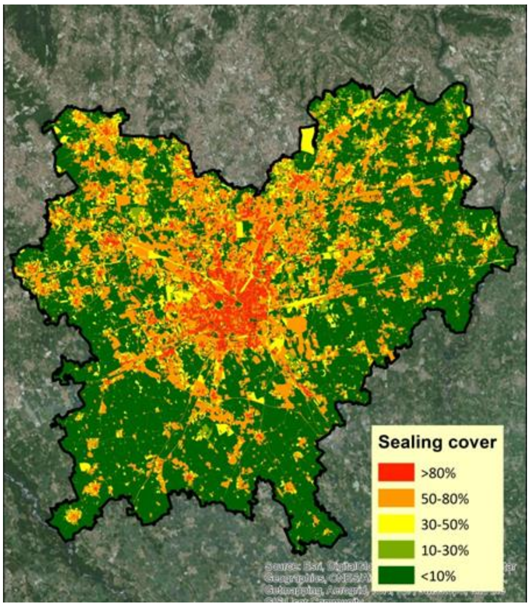

Fig 3. Sealing cover as a percent for Milan as captured by the European Urban Atlas.

Fig. 3 shows a map of sealing cover of Milan using the Urban Atlas data. Figs 2&3 are strongly correlated although the categories used by each differ. Milan itself appears at the centre of an urbanised landscape that extends outward along the road system. The intensity of urbanisation is apparent from the pattern that shows that the cover becomes progressively less impermeable with distance from the city centre. One should note that the boundary for Milan encloses only part of the contiguous urban area which includes smaller settlements that appear as patches of low pervious cover. Conversely much of Milan includes managed 'natural' cover.

3.3 Status

Currently there are over 100 cities being processed through the quality control process described above. Most of these cities are located in North America, Europe, South America and parts of Asia. So far, the database has grown 'organically' based on the interests of the participants but we have adopted a more strategic approach to gather more information on very large cities (such as the C40 group[2]) for which there is additional climate information and the cities of India and Africa. As data become available, it will be announced on the WUDAPT website.

4. WUDAPT portal tools

An integral component of WUDAPT is the development of tools that enable urban data to be extracted in a variety of formats. These portal tools will support a myriad of climate-related research issues including: assessing urban risks and evaluating resilience to current and projected climate changes; evaluating urban adaptation and mitigation policies; mapping air pollution and greenhouse gas emissions and; urban planning and design to mitigate urban heat island impacts. Currently, the portal allows users to submit and extract L0 data created using the WUDAPT protocol. Examples of two portal tools being developed are W2W and SCALER.

The W2W was designed to convert the spatial distribution of LCZ types into urban canopy parameter values for the Weather Research Forecasting (WRF) model. WRF has a few urbanisation schemes that account for the complex aerodynamic, radiative and energetic exchanges at the urban surfaces (Chen et al., 2011). The most complex of these schemes is BEP+BEM which provides a multilevel description of the urban canopy and includes a building energy model that links indoor and outdoor climates (e.g. Gutiérrez et al., 2015). These types of models represent the epitome of urban climate science but are limited in their application partly because of lack of urban data. Brousse et al. (2016) used L0 data to build an urban database for Madrid as part of an urban heat island study. Subsequently, Brousse developed W2W to create WRF-ready data for the region of interest at a suitable spatial scale; this tool converts the LCZ numerical descriptors (Table 1) into urban canopy parameters (UCPs). As such it provides a standardized methodology for bringing a city into this modeling framework. As more detailed urban data are acquired, the UCP data generated by W2W can be supplemented or made more precise.

Urban climate models require gridded UCPs at appropriate scales based on user needs and model physics. Typically this means that available datasets must be reformatted to remove extraneous details, which are not recoverable. SCALER uses the principle of the Multiple Resolution Analysis (MRA) to manage the multi-scale grid requirements of users (Neophytou et al., 2015). Its unique feature is its ability to retain sub-grid data on the input parameters as the selected model grid scale is increased. This allows the impact of sub-grid UCP variability on resulting model outputs to be examined and enables a clearer understanding of the role and impact of such parameters on the behavior of a complex urban system.

5. Discussion & Conclusions

WUDAPT is a community-based initiative and as such its development is largely led by the interests and needs of those that participate. Our immediate goals are to: increase the coverage of L0 data by including more cities, especially those for which there is very little data currently; develop protocols for acquiring data at Levels 1 & 2 (see Ching et al., 2017); build portal tools to support climate research.

There is still a considerable amount of work to be completed on the L0 data, especially in the realm of quality assurance but it is worth pointing out that even it its current form, these data can support a range of scientific projects:

- First, the LCZ maps provide a common basis to examine the physical layout and occupancy of cities.
- Second, the maps provide a context that can guide strategies for making observations, geo-locating environmental data and assisting interpretation. The LCZ maps can be used to as a framework to create observation networks that sample the urban landscape and its diversity efficiently. Moreover, the maps provide a means of geo-locating existing observations and assessing where there are gaps.
- Third, value ranges implied by LCZ types can be used to guide the selection of parameter values to describe the urban surface for hydrometerological and climate modelling.
- Finally, the protocol permits the generation of urban data that is useful for climate research in a timely fashion. Given the rate of planetary urbanisation and the dearth of consistent urban information, WUDAPT fills an important gap.

The WUDAPT project has great promise as a platform for knowledge sharing on cities and contributing to the creation of an urban climate science.

[2] http://www.c40.org/

6. References

Alexander, P.J. and Mills, G., 2014. Local climate classification and Dublin's urban heat island. Atmosphere, 5(4), pp.755-774.

Alexander, P.J., Mills, G. and Fealy, R., 2015. Using LCZ data to run an urban energy balance model. Urban Climate, 13, pp.14-37.

Alexander, P.J., Bechtel, B., Chow, W.T.L., Fealy, R. and Mills, G., 2016. Linking urban climate classification with an urban energy and water budget model: Multi-site and multi-seasonal evaluation. Urban Climate, 17, pp.196-215.

Bechtel, B., Alexander P.J., Böhner J., Ching J., Conrad O., Feddema J., Mills G., See L. and 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, no. 1 (2015): 199-219.

Brousse, O., Martilli, A., Foley, M., Mills, G. and Bechtel, B., 2016. WUDAPT, an efficient land use producing data tool for mesoscale models? Integration of urban LCZ in WRF over Madrid. Urban Climate, 17, pp.116-134.

Chen, F., Kusaka, H., Bornstein, R., Ching, J., Grimmond, C.S.B., Grossman-Clarke, S., Loridan, T., Manning, K.W., Martilli, A., Miao, S. and Sailor, D., 2011. The integrated WRF/urban modelling system: development, evaluation, and applications to urban environmental problems. International Journal of Climatology, 31(2), pp.273-288.

Ching, J. (2013). A perspective on urban canopy layer modeling for weather, climate and air quality applications. Urban Climate, 3, 13-39.

Ching, J., L. See, C. Ren, V. Masson, J. Hildalgo, X. Wang, G. Mills, J. Feddema, 2017: The WUDAPT Framework for Generating Urban Morphology, Material Composition and Activity Data for Modeling, Extended Abstract (9.4), 13th AMS Symposium of the Urban Environment, Seattle, WA..

Gutiérrez, E., González, J.E., Martilli, A., Bornstein, R. and Arend, M., 2015. Simulations of a heat-wave event in New York City using a multilayer urban parameterization. Journal of Applied Meteorology and Climatology, 54(2), pp.283-301.

Jackson, T.L., Feddema, J.J., Oleson, K.W., Bonan, G.B. and Bauer, J.T., 2010. Parameterization of urban characteristics for global climate modeling. Annals of the Association of American Geographers, 100(4), pp.848-865.

Mills G., Bechtel B., Ching J., See L., Feddema J., Foley M., Alexander P. and O'Connor M. (2015) An Introduction to the WUDAPT project. Proceeedings of ICUC9, Toulouse France. Available at http://www.wudapt.org/.

Neophytou, M., P. Mouzourides, A, Kyprianou, R. Choudhary and J. Ching, 2015: Sensitivity of mesoscale models to scale dependent UCP inputs, an example from energy demand, 9th International conference on Urban Climate (jointly with 12th Symposium on the Urban Environment), Toulouse, France, July, 2015

Oke T.R., Mills G., Christen A. and Voogt J.A. (2017) Urban Climates. Cambridge University Press. Forthcoming

Potere, D., Schneider, A., Angel, S. and Civco, D.L., 2009. Mapping urban areas on a global scale: which of the eight maps now available is more accurate?. International Journal of Remote Sensing, 30(24), pp.6531-6558.

Stewart, I. D., & Oke, T. R. (2012). Local climate zones for urban temperature studies. Bulletin of the American Meteorological Society, 93(12), 1879-1900.

.