Urban surface cover determined with airborne lidar at 2 m resolution – implications for surface energy balance modelling

Annika Nordbo[1], Petteri Karsisto[1,3], Leena Matikainen[2], Curtis R. Wood[3], Leena Järvi[1*]

[1]Department of Physics, University of Helsinki, Finland
[2]Finnish Geodetic Institute, Kirkkonummi, Finland
[3]Finnish Meteorological Institute, Helsinki, Finland

* corresponding author: Leena Järvi
Postal address: Erik Palménin aukio 1, PO Box 48, 00014 University of Helsinki, Finland
email: leena.jarvi@helsinki.fi

Abstract

Urban surface cover largely determines surface–atmosphere interaction via turbulent fluxes and its description is vital for several applications. Land-cover classification using lidar has been done for small urban areas (<10 km^2) whereas surface-cover maps in atmospheric modelling often have resolutions >10m. We classified land cover of the urban/suburban area (54 km^2) of Helsinki into six classes based on airborne lidar data, and an algorithm for machine-learning classification trees. Individual lidar returns were classified (accuracy 91%) and further converted to 2-m-resolution grid (95% accuracy). Useful lidar data included: return height and intensity, returns-per-pulse and height difference between first and last returns.

The sensitivity of urban surface-energy-balance model, SUEWS, to simulate turbulent sensible and latent heat fluxes was examined. Model results were compared with eddy-covariance flux measurements in central Helsinki. An aggregation of the surface-cover map from 2 to 100 m reduced the fraction of vegetation by two thirds resulting in 16% increase in simulated sensible heat and 56% reduction in latent heat flux. Street trees became indistinguishable already at 10 m resolution causing 19% reduction in modelled latent heat flux. We thus recommend having surface-cover data with 2 m resolution over cities with street trees, or other patchy vegetation.

Keywords

Urban surface cover classification, street trees, lidar scanning, energy budget model, eddy-covariance, surface energy balance model (SUEWS)

Highlights

- Lidar returns from a 54 km^2 urban area were classified into 6 surface cover classes.
- Decision tree classification of lidar data alone led to 91% accuracy.
- The commonly used 10-metre map resolution is still not fine enough to resolve street trees.
- Maps of 10-m resolution already cause a 19% error in modelled latent heat flux.

1. Introduction

Urban surface cover is strongly interrelated with several surface properties and processes in the surface–atmosphere interaction of an urban area. The replacement of pervious surfaces and vegetation by impervious materials alters the surface energy balance: most notably, the available energy is used for the vertical turbulent sensible heat flux rather than evaporation (e.g. Grimmond and Oke, 1991; Nordbo et al., 2012a; Loridan and Grimmond, 2012) which leads to the intensification of the inadvertent urban heat island effect, observed in many cities (Arnfield, 2003). The reduced evaporation and percolation is accompanied by a larger runoff, and together they affect the water balance of a city and cause a higher flood risk (Xiao et al., 2007); though this could give rise to opportunities for improving the urban environment (Coutts et al., 2013). The lack of vegetation also minimizes carbon dioxide (CO_2) uptake via photosynthesis and reduces the buffering effect of air pollutants (Demuzere et al., 2014). Indeed, the fraction of natural area has been shown to be a very strong proxy for predicting CO_2 emissions from cities (84% explanation power, Nordbo et al., 2012b).

The exact determination of the fraction of vegetation has received a lot of attention due to links to urban heat island mitigation. The simulated urban heat island in Tokyo was shown to be up to 1.5°C weaker when patchy vegetation was taken into account in a simulation (Hirano et al., 2004). By patchy vegetation, we mean small and isolated vegetation surrounded by impervious surfaces (one example is street trees).Vegetation has been concluded to have a cooling effect in an urban environment due to shadowing effects and evaporative cooling, especially in summer (Simpson, 2002; Bowler et al., 2010; Lindberg and Grimmond, 2011; Coutts et al., 2013). Thus, there is a clear need for land cover classification that adequately resolves patchy vegetation.

Urban surface cover is commonly available only with a coarse resolution (10–50 m) and without building and tree height information. At such a resolution, patchy vegetation is almost entirely disregarded. Airborne Light Detection And Ranging (LiDAR, hereafter lidar) measurements are based on active remote sensing technology where visible or near-infrared (NIR) light is sent and received in high frequency pulses (Beraldin et al., 2010). Lidar technology provides a solution to the resolution problem: it allows detailed scanning of a 3D complex urban structure with a high horizontal and vertical accuracy with a relatively low cost. Furthermore, the intensity of returned lidar pulses is an indicator for the colour of the target hit: which improves the ability for land cover determination. Generally, the determination of urban morphology is comparatively simpler since it relies only on the height information of the returned lidar pulse; whereas intensity information is additionally needed for surface-cover classification.

The use of the lidar scanning technique is increasing its popularity in the determination of urban land cover (Table 1). Classification has been mainly done using object-based approaches (Blaschke, 2010), where pixels are first divided into segments which are then classified into land-cover classes using different methods (Brennan and Webster, 2006; Im et al., 2008; Matikainen and Karila, 2011; Buján et al., 2012; Zhou, 2013;-Dunne et al., 2013 in Table 1). Classification trees, a machine learning algorithm presented in Breiman et al. (1984), or other related approaches, have been used in many studies, and overall accuracies over 90% have been reached (e.g., Im et al., 2008; Mancini et al., 2009; Matikainen and Karila, 2011 in Table 1). The number of surface-cover classes ranges from 2 to 10 among the research done within a decade, and the most commonly used variables used for classification are: height above ground and intensity or/and a colour channels from an orthoimage (geometrically corrected aerial photograph, Table 1). Segmentation and classification have typically been carried out using rasterized data. This is the most straightforward approach when an orthoimage is used together with lidar data. Classification of individual lidar returns has mainly been used in a pre-processing stage to group lidar points into ground points and non-ground points. Matikainen and Karila (2011) made a comparative surface-cover classification study at a suburban area using five techniques and concluded that a combination of lidar scanning and orthoimage data gives the best result. Lidar scanning can also be used for updating maps through an automatic detection of changes in buildings (Matikainen et al., 2010), for calculating solar irradiance in urban canopy (Tooke et al., 2012) as well as for predicting building energy demand (Tooke et al., 2014).

This study has two main aims:(i) to determine the land cover of a large area within Helsinki at a very fine

resolution (2 m) for use in modelling of the surface-atmosphere exchange, in addition to creating an elevation map as a part of the land cover classification, and (ii) to quantify the sensitivity of urban energy balance modelling on land cover resolution at an urbanized micrometeorological measurement site, thus addressing errors caused by too coarse maps commonly used to describe the surface in urban energy balance models.

Surface cover is determined by airborne lidar scanning data, together with a classification tree algorithm. The classification is conducted initially for each individual lidar return, rather than for pixels or segments as has been done in several previous studies. This is done due to the large and very heterogeneous surface and the need for a generic method that could be applied at other micrometeorological stations. The study area of concern is the second largest study worldwide among urban areas for which classification has been done using lidar data (Table 1).The resulting surface cover data are then used for simulating surface energy fluxes using the Surface Urban Energy and Water Balance Scheme (SUEWS version 2014b, Järvi et al., 2011; Järvi et al., 2014). The simulation is done for an eddy-covariance flux measurement site, located in the lidar scanning domain (Nordbo et al., 2013). Recommendations for an adequate surface cover database resolution are finally given.

Table 1. Review of work on urban surface cover classification using lidar data. OA is overall accuracy.

Study area (surface area in brackets)	Classes (number in brackets)	Classification method (software in brackets□)	Data used in classification¤	Reference data (number of data points in brackets)	OA (%)	Reference
mixed forest and estuary to an urban area (12 km²)[†]	saturated and non-saturated intertidal sediments, saturated or stressed and lush ground cover vegetation, low and tall deciduous and coniferous trees, roads and bare soil, bright-roofed structures, dark-roofed structures, water (10)	segmentation into homogeneous regions (eCognition) and classification with a manually decided hierarchy	z, z_g, I, RPP	polygons from orthoimages (53217)	94	Brennan and Webster, 2006
semi-urban (< 0.5 km²)[†]	building, tree, grass, road/parking lot, other artificial object (5)	segmentation into homogeneous regions (Definiens*) and machine-learning classification trees (C5.0)	z, z_n, z_g, I	points from visual inspection of orthoimages (200)	>90	Im et al., 2008
urban (0.16 km²)[†]	building, grass, land, tree (4)	machine-learning algorithm Real AdaBoost with classification trees (N/A)	z, z_n, Δz; R, G, B, NIR and NDVI from orthoimage	polygons from orthoimages (<13126)	96%	Mancini et al., 2009
urban to suburban (13 km²)	building, vegetation (2)	structures in RPP image (N/A)	z, z_n, RPP	orthoimage	83%[Δ]	Goodwin et al., 2009
suburban (training 0.7 km², validation 2.7 km²)	building, tree, vegetated ground, non-vegetated ground (4)	segmentation into homogeneous regions (eCognition) and machine-learning classification trees (Matlab)	z_n, NDVI from orthoimage^	points from visual inspection of orthoimages and other sources (269)	97	Matikainen and Karila, 2011
urban to suburban (19 sub-domains of 0.16 km²)	building, trees (2)	buildings separated using building footprint data; most of the z > 2.5 m features were trees	z	-	-	Lindberg and Grimmond, 2011
suburban (0.4 km²)[†]	building, high vegetation, low vegetation, pavement/road, bare earth (5)	segmentation into homogeneous regions (Definiens*) and classification with a manually decided classification tree based on boxplots	z, I, Δz, RVI from orthoimage	pixels from visual inspection of orthoimages (75)	96	Buján et al., 2012
suburban (6 km²)[†]	building, pavement, trees and shrubs, grass (4)	segmentation into homogeneous regions (eCognition) and classification with a manually decided hierarchy	z, I, ΔI	points from visual inspection of orthoimages (300)	91	Zhou, 2013
urban (0.3 km²)[†] residential (<0.1 km²)[†]	building, bare ground, grass, high vegetation, miscellany (5)	partially supervised hierarchical classification	z, lidar-NDVI¤ from orthoimage	1122 N/A	90 92	Guan et al., 2013
urban to suburban (369 km²)	building, bare soil, trees, grass, road, other paved, water (7)	iterative rule-based classification (eCognition Developer)	z, z_d, I	points from visual inspection of orthoimages and lidar data (3000)	94	O'Neil-Dunne et al., 2013
urban to suburban (54 km²)	building, impervious, grass, low vegetation, high vegetation, water° (5)	classification trees (Matlab)	z, I, RPP, Δz	points from visual inspection of orthoimages (51 944)	91	This study

□ eCognition: Trimble GeoSpatial, C 5.0: RuleQuest Research, Release 2.0, Matlab: Mathworks Inc

¤ z – height of first return above ground, z_n – height of last return, z_g – ground height, I – intensity, ΔI – intensity difference between the first and last return, RPP – returns-per-pulse, Δz – height difference between the first and last return, z_d = elevation deviation in each grid cell, NDVI – normalized difference vegetation index, RVI – ratio vegetation index from the red and near infra-red bands, lidar-NDVI – NDVI where lidar intensity used

* former and later eCognition

[†] Area not given in the reference, evaluated value   ^Several lidar and orthoimage attributes were tested but only NDVI selected by the classification tree algorithm   [Δ] 83% of building objects were discriminated from surrounding features

° Note that water was defined from a mask or using a flood-fill algorithm

2. Materials and methods of surface cover classification

2.1 Study area

The data included in this study cover an area of 6 km by 9 km (54 km^2), consisting of six tiles from the national lidar database and covering central Helsinki (Figure 1). Helsinki is the capital of Finland located on the coast of the Baltic Sea with a population within Greater Helsinki of about one million. The city is characterized by a very heterogeneous surface with a dense city centre (although without skyscrapers) and a large central park which is mostly a mixed forest. The vicinity of the sea causes a marine climate which can change to a continental climate, depending on the prevailing airmass. Climatological (1981−2010) monthly mean temperatures vary between −4.7°C in February and 17.8°C in July (Pirinen et al., 2012). The large variation is caused by the range in daylight hours from 6 to 19. The average annual precipitation is 655 mm.
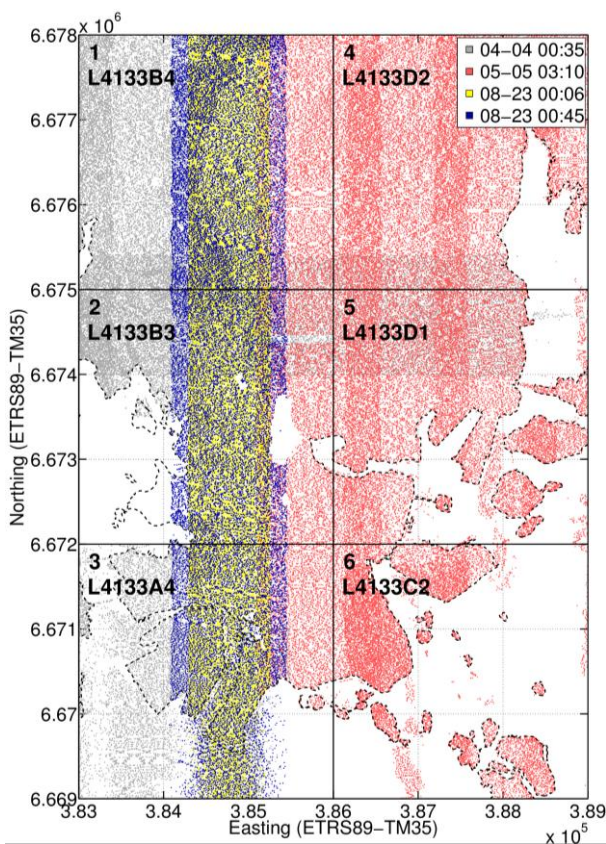

Figure 1. Available lidar scanning data from different days in 2008 (see colours and times in legend, "month day hour minute" as mm-dd hh:MM). Only 0.5% of data are shown. Data tiles (3 km x 3 km) are marked with a solid black line and numbered 1−6 in addition to the database code names (starting L4133). The coastline is marked with a dashed black line.

2.2 Lidar data and supporting information

The lidar data were collected during four flights on three days in 2008: April 4th, May 5th and August 23rd (Figure 1, Figure 2; National Land Survey of Finland, 2012). The data from August were divided into two flights due to a change in recorded intensities although the flight paths were almost the same. The data are part of an open database that aims at providing a nationwide lidar scanning coverage for the determination of a new digital elevation model (DEM) with a 2 m resolution. The flight altitude was 1900 m (except for 1300 m on

23rd of Aug) and measurements were done with two Optech ATLM GEMINI lidars with a pulse frequency of 50 kHz. The maximum off-nadir view angle is 20° and the pulse footprint is about 0.5 m (beam divergence 0.25 mrad). The accuracy of the airborne laser scanner is mainly determined from the GNSS (global navigation satellite system) accuracy (typically within 2–5 cm), from the IMU (inertial measurement unit) quality and range to the target. Thus, typically the height accuracy of ALTM Gemini in Finnish conditions is 0.15 m and the planimetric accuracy is better than 0.6 m (range 2 km). The dataset includes the coordinates of each return (ETRS-TM35FIN), height (m a.s.l.), pulse number (i.e. a sequential ID number within each flight), return number per pulse (1–4) and return intensity. Up to four returns per sent pulse were recorded: the first three returns and the last return. Altogether there are over 40 million data points in the six-tile dataset. The point density was on average 0.76 # m^{-2} for the six tiles, including water area with almost no returns and areas where flight tracks overlap. Given the point density and footprint, the surface cover classification resolution was chosen to be 2 m.

In addition to the lidar data, geographical information on the coastline and building edges were used (HSY, 2008). A mask with a 2 m resolution was created for both using ArcMAP (ArcGIS 10.0, Esri). Shapefile lines were first converted to polygons and polygon geometry was repaired to exclude self-intersection, unclosed rings and duplicate vertex. The repaired polygons were then converted to a raster image. The sea mask is especially important because lidar data were mostly not available for the sea, since pulses are poorly returned from water surfaces.

2.3 Pre-processing of raw lidar data

The processing is summarised with a flowchart (Figure A.1), and as follows:

The lidar data post-processing tool FUSION (version 3.3, McGaughey, 2009) was used to create a Digital Surface Model (DSM, Figure 2 a) and filter outliers from the data with individual lidar returns. First, obvious outliers were removed using the *FilterData* algorithm (those data-points more than 4 × st.dev. within a 10 x 10 window). Second, the *GroundFilter* algorithm was used to extract data-points that probably represent the 'ground points' (3 x 3 median filter). Third, since many building roof data-points were included in the 'ground points', the *PolyClipData* algorithm was used to exclude building data-points based on an independent dataset of building footprint shapefiles (HSY, 2008). Fourth, the *Gridsurfacecreate* algorithm was used to make a Digital Terrain Model (DTM, Figure 2 b). Fifth, outliers were further omitted based on the distance to the DTM: the lidar return height should be from −1 m to 100 m. Sixth, a set of ground points were defined and a new DTM (Figure 2 b) and a new DSM (Figure 2 a) were created. Seventh, a normalized DSM (nDSM = DSM − DTM, Figure 2 c) was created to show heights of trees and buildings a.g.l. (above ground level), and the lidar point data (individual returns) were saved in ASCII format with height data in metres a.g.l.

The backscatter intensity values were corrected for spherical loss (i.e. varying flight height) following standard procedures (e.g. Ahokas et al., 2006). The reference height used here was 1900 m (compare with Table 2). The incidence angle also affects the intensity (e.g. Kaasalainen et al., 2011) but it was not corrected for, since it had a minor effect: if the maximum scan angle of 20° is used as an approximation, the maximum correction is 8%. The inclusion of the incidence angle correction did not to improve the surface cover classification and thus it was omitted for simplicity. An atmospheric correction was not done, since data were divided according to flights and classification trees were used for each flight separately, and thus the correction became redundant.
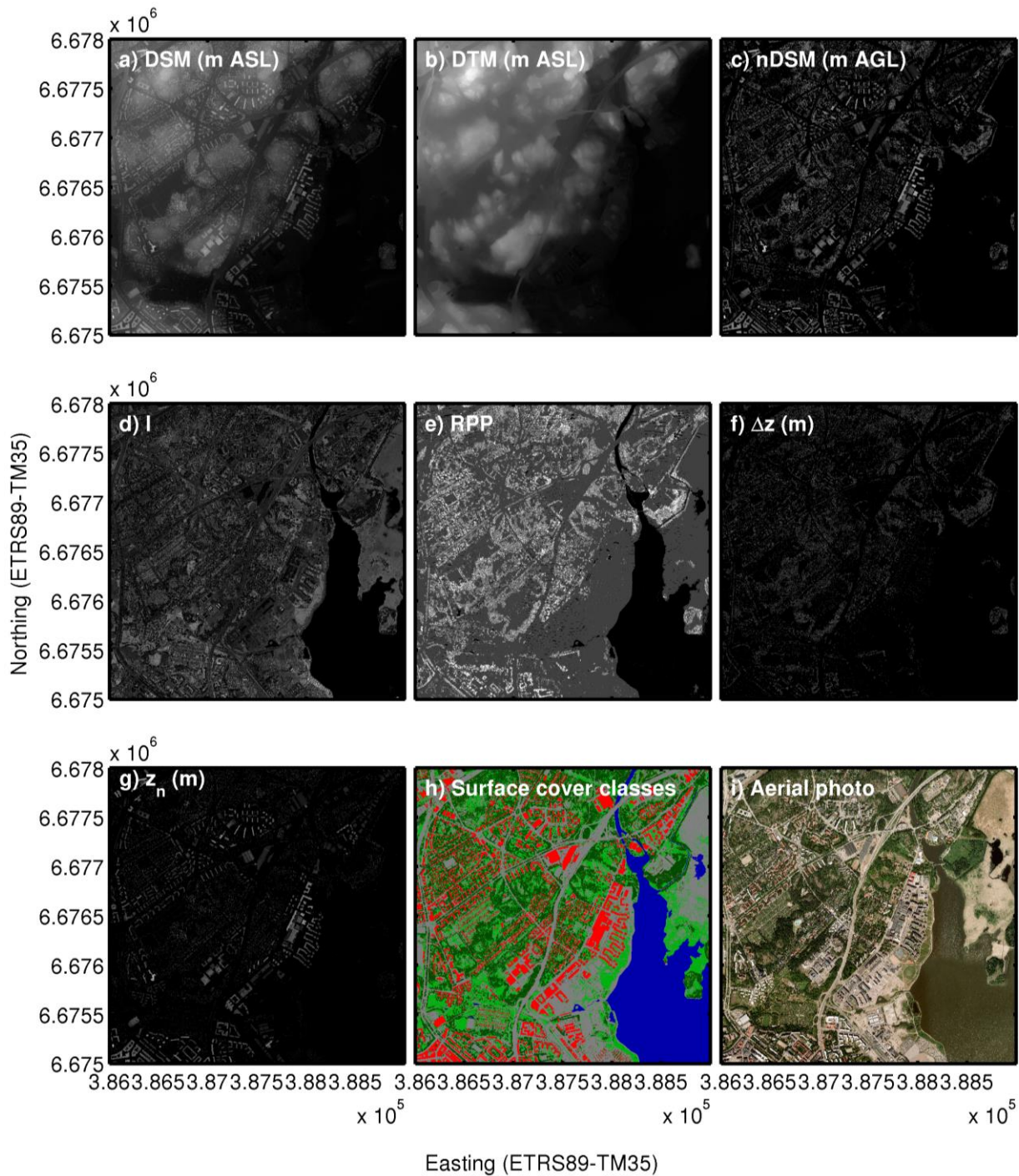
Figure 2. *Maps of the 3km x 3km area for which algorithms were developed (tile 4): a) digital surface model (DSM), b) digital terrain model (DTM), c) normalized DSM (nDSM = DSM − DTM), d) return intensity (I), e) returns-per-pulse (RPP), f) height difference between first and last return (Δz), g) height of last return (z_n), h) surface cover classes and i) aerial ortho photograph. The resolution of all maps is 2 m and maps a−d), f), g) show medians and the map in e) the maximum of lidar points in each grid cell. The surface cover classes in h) are the same as in Figure 4, i.e.: red−buildings, grey−impervious, light green−grass, medium brown−low vegetation, dark green−high vegetation, blue−water.*

Table 2. Lidar scanning flight details: *date and time (UTC+3), flight altitude, instrument, number of flight lines within the six analyzed tiles, number of lidar data points and number of data points in the training dataset.*

Time (UTC+3)	Flight altitude (m)	Instrument	Number of flight lines	Number of data points	Number of training data points
Apr 4 2008 00:35–02:00	1900	04SEN161	2	9676484	11900
May 5 2008 03:10–04:49	1900	05SEN180	4	15190786	13175
Aug 23 2008 00:06–00:45	1300	04SEN161	1	9585022	13221
Aug 23 2008 00:45–01:22	1300	04SEN161	1	6468180	13648

2.4 Surface cover classification

Data from tile 4 (Figure 1 and 2) were used for developing the algorithms introduced in the following. The step-by-step data processing is explained with the aid of Figure 2 (see also Appendix A.1.). First, individual lidar returns are classified using classification trees and the data are converted to a grid. The gridded data are then improved by a set of post-processing steps at pixel level. Orthoimage data are used for creating training and validation data, but a normalized difference vegetation index from the images is not used in classification due to colour differences in such a large area.

(1) Classification of individual lidar returns using classification trees

The land cover classification was done using Matlab (R1012a, MathWorks MATLAB®) into six surface cover classes: built, impervious (including paved, rock, gravel), grass, low vegetation, high vegetation and water. Data from first lidar returns were mainly used, constituting 83% of all available data, in addition to information assigned based on additional returns that correspond to the first return. The classification was done using a classification tree algorithm that had six input variables: height above ground (z, Figure 2 c), return intensity (I, Figure 2 d), returns-per-pulse (RPP, Figure 2 e), height difference between the first and last return (Δz, Figure 2 f), height of the last return (z_n, Figure 2 g), intensity difference between the first and the last return (ΔI; not shown, since later is proven not to be relevant). Returns that belong to the same pulse are in succession in the data and thus the returns corresponding to the first return could be identified based on the return number per pulse (ranging from 1 to 4).

Different types of objects and surface types have different reflectances of NIR light which results in different return intensities (Figure 2 d). Generally, the return intensity is lowest for a dark surface such as a paved road or a dark roof and water appears the darkest since it absorbs the NIR laser pulse. Wet surfaces also appear darker than dry (Kaasalainen et al., 2010) but this is not relevant here since the flights were preceded by at least three non-rainy days. Buildings can have any intensity value since a large range of roof materials is possible. The intensity of trees depends on the species (Kim et al., 2009) and whether or not leaves are on trees, but the intensity range is more constrained than that of buildings. Since the intensity and height range of buildings and high vegetation overlaps, information on e.g. returns-per-pulse is additionally needed to distinguish them. The height of grass and impervious areas (paved, rock, gravel) can be the same and thus they are distinguished by the intensity.

For each flight, a training dataset was created by randomly selecting lidar returns for which the surface cover type is known based on an orthoimage (altogether 51 944 data points, Table 2): For each flight and

class, points that could belong to the class were overlaid on an orthoimage and points were selected in groups of tens of points according to visual inspection. Buildings were defined to have a height above 2 m and low vegetation between 0.5 m and 2 m in the selection of points that could belong to the class in concern. Half of each dataset was randomly chosen to be left for validation and the other half is used for training the classification tree. First, an optimal minimal leaf area (i.e. minimum amount of data points in each ending point of a tree) and an optimal prune level (i.e. amount of splits in the tree) were calculated for each flight. The Gini's diversity index was used as a splitting criterion. Second, a prediction for the validation dataset was done using the optimized tree, and the prediction capability was assessed. The importance of each of the four input variables was calculated based on the risk associated to splits on every predictor, using the Predictor Importance algorithm in Matlab. Third, the classification tree was used for predicting the land cover class for the whole dataset. Finally, data points with an assigned surface cover class were converted to a grid using the mode of points falling in each pixel. If several values were the mode, the mode was taken randomly, and a missing value was not allowed to be the mode.

Individual lidar returns were classified, instead of pixels or segments, for two reasons. (i) The classification was not done at pixel level due to the complex urban land cover where several land cover types can collocate even in a 2 metre resolution pixel. Thus, the average or median height or intensity within a pixel would not describe the land cover correctly. (ii) Individual lidar returns were used instead of segmentation since the analysed area was large and included a variety of surface types which thus could be more generically classified. The approach also does not rely on specific special segmentation software and can be used for other areas too.

(2) Post-processing at pixel level
Separation of buildings and trees. To improve the separation of buildings and trees (we can currently safely ignore so-called green roofs in Helsinki due to their small amount), a multiple return image was created. Areas covered with trees typically return multiple echoes whereas building roofs only return one. An exception to this are building edges that often give two returns, one from the roof and one from the ground. The multiple return image gives the maximum number of returns-per-pulse within a pixel (Figure 2 e). Only pulses for which z < 30 m and for which Δz < 30 m were included in order to minimize the amount of building points falsely included: lidar returns from tall building edges often have Δz > 30 m. Even with this constraint, wall structures were still included in the RPP image. Thus, an altered form of the filtering method proposed by Goodwin et al. (2009) was used to remove linear structures. A 9 x 9 pixel moving kernel was run through all pixels. If the centre point has more than one return, a line with 36 different angles is run through the centre point. Two options for setting the RPP value to one were used: (i) if ≥ 4 of the 9 pixels on the line and only 4 outside the line have multiple returns or (ii) if all 9 line pixels have a maximum of two returns and there are only 16 outside the line have multiple returns. Once the linear structures have been removed, all pixels with RPP ≥ 2 were classified as low or high vegetation. Dense trees can also only have one return and thus individual building points inside a forest are omitted with a 3 x 3 moving kernel that is repeated seven times: if ≥ 5 pixels are low or high vegetation and ≤ 3 building pixels, the pixel is replaced by high vegetation. Finally a building mask (section 2.2) is still applied. The effort for separating buildings and trees at pixel level was nevertheless done because the building mask does not include all of the smallest buildings and structures. Thus, the net effect of the mask is that it ensures that for instance buildings with complex glass roofs are not classified as trees.

Determination of water. Since water mostly absorbs the lidar pulse, data availability over water is low and an alternative method for locating water has to be used. The location of sea was determined using a

mask based on coastlines (section 2.2). Inland waters were flood-filled using manually pre-determined flood-fill seeds. Rivers were also flood-filled, but the seeds were automatically detected by finding linear structures of missing data (like for the RPP image). The kernel size was 31 pixels, and ≥80% of data on the line and ≤30% of the whole kernel had to be missing data in order for the line points to be considered as flood-fill seeds.

Removal of bridge edges. Bridge edges appear as buildings since they are often high above the ground level determined by the DEM. A 21 x 21 pixel and a 11 x 11 pixel kernel are used to convert those points to impervious: a line is rotated around a building pixel (like for the RPP image) and if < 20% of the kernel is building and at least all but two of the pixels on the line are building, the pixel is determined as the seed of a flood-fill algorithm. Two different sized kernels are used in order to find both large vehicle bridges and smaller pedestrian bridges.

Removal of grass points inside impervious land. Since impervious and grass land are almost solely distinguished by the intensity value, the classification is more prone to errors than other classes. A 3 x 3 moving kernel was used for replacing grass pixels surrounded by impervious land if ≥ 8 pixels in the kernel were impervious.

Final smoothing at pixel level. The final surface cover map (Figure 2 g) is retained with a five time repetition of a 3 x 3 pixel randomly moving kernel to fill in missing data: if ≤ 3 of the kernel pixels are missing, the pixel is replaced by the mode.

Aggregation. The smoothed surface cover map was aggregated to seven resolutions (4, 6, 10, 20, 30, 50 and 100 m) in addition to the original 2 m resolution. This was done with a mode filter: the class with most occurrences was chosen, or if there were several classes with the same occurrence, the aggregated class was randomly picked.

3. Materials and methods of flux measurements and modelling

3.1. Eddy-covariance site and measurements

The eddy-covariance measurements, used to evaluate the energy balance model performance, were done at the Hotel Torni site located in the centre of Helsinki (Figure 3). The site is one of the measurement sites of the SMEAR III (Station for Measuring Ecosystem-Atmosphere Interactions) and the site is also part of the intensive Urban Boundary-layer Atmosphere Network in Helsinki (Wood et al., 2013). The site is characterized by dense urban land-use with a high fraction of built-up area with buildings and paved streets in all directions. Street trees and small parks are typical for the patchy greenness of Helsinki city centre. The average building height is 24 m and the built fraction is 55% (Nordbo et al., 2013), and the area is a source of CO_2 (4.8 kg C m^{-2}, Nordbo et al., 2012b).

Measurements of fluxes of sensible heat (*H*) and latent heat (*LE*) using the eddy-covariance measurement technique have been on-going since September 2010 (Nordbo et al. 2013). The flux measurements are done on a mast mounted on the north-west corner of the hotel's tower and thus flow distortion hinders data retrieval when the wind originates from 50–185°. The equipment consists of a three-dimensional sonic anemometer (USA-1, Metek Ltd, Germany) and an infrared gas analyser (LI-7200, LI-COR Biosciences, Lincoln, NE, USA). Flux calculations and quality screening are performed according to widely accepted procedures (Aubinet et al., 2012; Nordbo et al., 2012a; Nordbo et al., 2013). The majority (75%) of the flux has been estimated to originate from within a kilometre of the tower. For more information on the site characteristics, measurement setup and source area analysis, see Nordbo et al. (2013).

The flux measurements are accompanied by auxiliary meteorological measurements needed for model

forcing and these measurements were conducted at two locations in Helsinki. The Elisa mast is located 550 m south-east of the tower (Figure 3) and the SMEAR III – Kumpula site 4100 m north-east of the tower (Figure 4). Air temperature data from the Elisa mast were used whereas other data needed to force the model (solar radiation, air pressure, wind speed, relative humidity and precipitation) were measured in Kumpula. Details on instrumentation on the Elisa mast can be found in Wood et al. (2013), whereas details on the Kumpula site can be found in Järvi et al. (2009). Meteorological data from Kumpula were gapfilled as described in Järvi et al. (2012). If air temperature data were missing from the Elisa mast, a seasonal linear fit with data from Kumpula was used to replace the missing values.
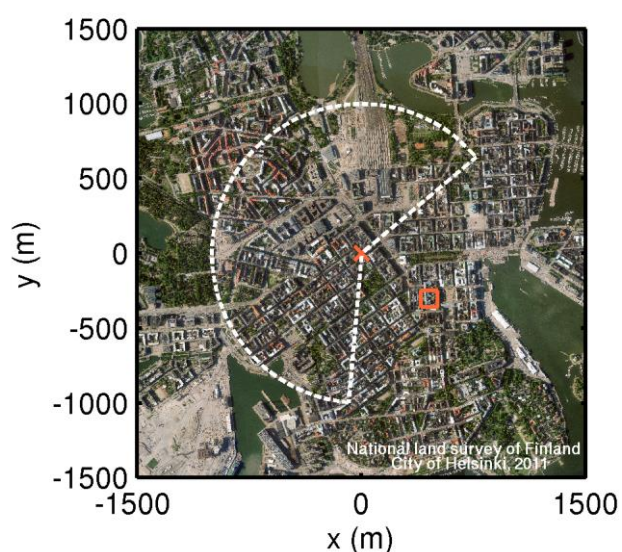

Figure 3. Orthoimage of the area around the micrometeorological measurement station Hotel Torni (red cross); x is positive eastwards and y is positive northwards. An area with a 1000 m radius and without instrument flow distortion is also shown (white dashed line) in addition to the Elisa mast (red square).

3.2. SUEWS model and model runs

The model used in the study is the Surface Urban Energy and Water balance Scheme (SUEWS, Järvi et al., 2011; Järvi et al., 2014). It simulates the energy and water balances in a local or neighbourhood scale using meteorological forcing data and information about the surface cover. Urban surface is divided into seven surface cover types (buildings, paved, evergreen/deciduous trees and shrubs, grass, irrigated grass and water) and rates of evaporation-interception for each single-layer surface type are calculated. The model takes also into account underground soil, and water movements both at the surface layer and soil layer are allowed. The urban energy balance components are separately resolved: the latent heat flux is calculated from the modified Penman–Monteith equation and the sensible heat flux as a residual from the available energy minus the latent heat flux. The model is run with an hourly time step, but for the water balance calculations it adopts a 5-minute time-step due to correct response to precipitation and snowmelt events.

SUEWS was run for one grid (or model) areain the centre of Helsinki for 18 months from July 2011 to December 2012. The model area was chosen to be a 1 km radius circular area centred at the Hotel Torni site, since this area is most representative for the flux measurements (Nordbo et al. 2013) against which the model performance is evaluated. The same area was simulated using eight different surface-cover fractions

(resolution 2, 4, 6, 10, 20, 30, 50 and 100 m) aggregated for the 1 km radius circle. The different resolution surface-cover fractions, conversely, were aggregated from the 2-m-resolution lidar data. Only data from summer months June to August 2012 were used in this study since we wanted to concentrate on the full leaf period in the sensitivity analysis. Table 3 shows some of the site-specific model parameter values of SUEWS. The average building and tree heights for the whole area (1000 m radius) were calculated from the original 2-m-resolution dataset and were used for all runs. More information about the default parameters that are not given in Table 3 can be found in Järvi et al. (2014).

Table 3: Site specific model parameters used in this study.

Parameter	Value
Latitude (°)	60.1678
Longitude (°)	24.9387
Model area (ha)	314.2
Model height (m)	60
Building height (m)	17.9
Tree height (m)	8.3
Roughness length (m)	1.4
Displacement height (m)	14.9
Population density (pop ha^{-1})	80.9
Starting day for irrigation (day of year)	152
Ending day for irrigation (day of year)	243
Fraction of automatic irrigation	0.30
Initial conditions	
Days since precipitation	4
Previous day air temperature (°C)	21
Surface water states (mm)	0
Initial soil storage, buildings (mm)	10
Initial soil storage, vegetation and paved surface (mm)	54

The model performance was evaluated using statistics commonly used in studying urban energy balance models (Grimmond et al. 2010). The used statistics include the root mean square error (RMSE), systematic and unsystematic RMSE, mean bias error (BIAS) and index of agreement (d). In addition, we calculated the centralized root-mean-square deviation

$$CRMSD = \sqrt{\frac{\sum_{i=1}^{n}[(y_i - \langle y \rangle) - (x_i - \langle x \rangle)]^2}{n}}, \text{ (W m}^{-2}) \tag{1}$$

where *y* denotes the simulation result and *x* is the measured value. The angle brackets <> denote the mean value. The CRMSD value deviates from the traditional RMSD value in that the bias has been removed by using subtracting the mean values.

4. Results and discussion

4.1. Evaluation of surface cover classification

The large amount of data (over 40 million data points) including four flights were separately classified using the classification tree machine learning technique. The minimal leaf area ranged from 14 to 66, and the number of nodes (i.e. splits in classification tree) from 15 to 65 (Table 4). Height above ground was always the

most important classifier, followed by RPP and intensity. The 3rd flight is an exception since the order of importance of RPP and I was reversed and the classification tree algorithm did not use any other variables. For the other flights, Δz was the fourth important classifier, z_n was used for the first flight and both ΔI and z_n were used for the second flight. The height of the second and third return was not used in any decision trees and thus they were omitted from all analysis in this paper.

***Table 4. Characteristics of the surface cover classification using classification trees for each flight**: minimal leaf area, number of nodes, accuracy per class (*user's accuracy*) and for all classes and fraction of data for which the prediction probability is <0.6.*

Flight time (UTC+3)	Mini-mal leaf area	Num-ber of node	Accuracy per surface cover class						Prediction probability <0.6 (% of all data)
			built	imper-vious	grass	low veg.	high veg.	All	
Apr 4 2008 0:35–02:00	27	31	88.9	86.6	83.9	99.8	92.6	90.4	7.0
May 5 2008 03:10–04:49	14	65	93.1	87.5	86.6	99.9	90.0	91.6	2.9
Aug 23 2008 00:06–00:45	64	15	93.3	83.6	93.6	99.9	84.6	89.6	0.1
Aug 23 2008 00:45–01:22	66	25	89.5	87.8	92.3	99.9	90.0	90.8	1.8
All	-	-	91.3	86.4	89.3	99.9	89.2	90.6	-

Intensity was found more important than height elsewhere (Im et al., 2008), but the surface types are not comparable. The intensity value, that was the main distinguisher between grass and impervious surfaces, was 35, 21, 21, 18 for the four flights, respectively. This result emphasizes the importance of separate classification schemes for each flight since intensities are not easily cross-calibrated. Differences can be caused for instance by the lack of leaves and grass in Helsinki in April compared with the other months, or the wetness of the surface on the flight day, or aerosol optical thickness (no atmospheric correction). The difference between the 3rd and 4th flight remains small, as predicted, since they were from the same day.

RPP was a strong distinguisher between tall vegetation and buildings, as has been shown elsewhere (Lindberg and Grimmond, 2011), although intensity was here a common distinguisher, too. Δz was used for separating impervious surfaces and grass with a limit of about $\Delta z > 1$ m for grass. The Δz limit is high considering that the height distinguisher for grass was 0.5 m (section 2.4), but it must be noted that the height above ground is based on the gridded DTM and thus $\Delta z > 1$ m can occur although $z < 0.5$ m. The Δz criterion was used only for 0–10% of the grass data of the four flights. It was expected that Δz would be good for detecting building edges and trees, but the classification tree algorithm used it only for the data from the 1st flight and for classifying 2% of the data. These results lead us to expect that RPP was a good enough distinguisher and thus Δz was commonly not needed anymore. Furthermore, z_n distinguished 1% of building/tree data of the 2nd flight and 4% of impervious/grass data of the 1st flight. ΔI was used in the classification tree of the 2nd flight to distinguish <1% of building/tree data and it can be deemed unnecessary.

For individual lidar returns in each flight, the accuracy for high vegetation and buildings was generally ≥ 90%, whereas for grass and impervious the value was mainly below 90% (Table 4). This is because impervious and grass surface cannot be clearly distinguished by RPP or the height, since the classification is mainly based on the intensity value. The errors in buildings and high vegetation were due to confusion between them and the same applies for grass and impervious surfaces. The fraction of data points for which

the prediction probability is <0.6 (i.e. the likelihood that the point belongs to a particular class) ranged from 0.1% to 7%. The overall accuracy for all flights is 91% before the individual lidar returns are made into a grid. The range in prediction accuracy between the four flights is very small (0.03) which verifies that the use of separate classification schemes was a suitable way to overcome the difficult problem of cross calibration between intensity values from different flights.

For the gridded data, the amount of lidar pixels that coincide with building pixels from the mask ranges between 88% and 94% for the six tiles (Table 5), which is roughly in accordance with the accuracy for individual lidar returns. In order to increase the accuracy of the final gridded surface cover dataset, the building mask was used for introducing 100% completeness for buildings. After this and other pixel level post-processing steps (see section 2.4), 1–2% of data were missing from the six tiles (Table 5). The final accuracy is estimated to be >95% after all data processing steps.

The overall accuracy of the surface-cover classification is consistent with reported accuracies among similar studies (Table 1) although comparison between studies is difficult due to different locations of concern, horizontal resolution and the number of classes. Studies that rely only on lidar data and use object-oriented segmentation and classification trees have reached overall accuracies of 94% for a large mixed land cover area (Brennan and Webster, 2006), >90% for a small semi-urban area (Im et al., 2008) and 91% for a medium-sized suburban area (Zhou, 2013). Studies which have used complementary orthoimage data and varying methods report accuracies of 96% for a small urban area (Mancini et al., 2009), 97% for a medium-sized suburban area (Matikainen and Karila, 2011), 96% for a small suburban area (Buján et al., 2012), 90% for a small urban and 92% for a small residential area (Guan et al., 2013). Furthermore, Im et al. (2008) conclude that metrics based on shape or texture did not improve the classification significantly, i.e. information on mean height, standard deviation of height and intensity were enough for classification.

Table 5. Statistics of data for each data tile.

Tile	Number of data points	Point density[¤] (# m^{-2})	Missing data before final smoothing (%)	Lidar building point in mask building point (%)[◊]
1	11 114 592	1.24	2	88
2	9 490 478	1.05	1	94
3	6 823 906	0.76	1	86
4*	5 851 619	0.65	2	93
5	5 039 354	0.56	1	92
6	2 600 523	0.29	2	94

* Tile 4 is the area for which algorithms were developed (Figure 2)
[¤] Includes also water area where data are usually not available
[◊] Percentage of building grid points based the classification trees of lidar points that were also building grid points based on the building mask.

4.2 Surface cover and height in central Helsinki

The surface cover classification of the whole domain is presented in Figure 4. Water covers 33% of the surface area and 1% of the area had no data coverage. Of the remaining land area, 27% is vegetation (16% high vegetation, 2% low vegetation, 9% grass), 27% impervious (including bare rock or sand) and 13% buildings. The fraction of impervious surfaces is large due to the presence of harbour areas, and fraction of high vegetation increased by the central park (north-south-directed strip in the north-west corner of Figure 4).

The height data (a.g.l.) are available both for each lidar return and as gridded height data (nDSM, Figure 2c). The distribution of tree height has a mode at 4 m and is positively skewed. The highest trees are found in the central park (Figure 5), and 95% of the trees are shorter than 20 m. The distribution of building height is bimodal: the first peak is at 7.5 m which corresponds to a two-storey detached house (north and north-west in the domain) and the second is at 22 m which corresponds to the homogenous building height in Helsinki city centre (around Hotel Torni in Figure 5).
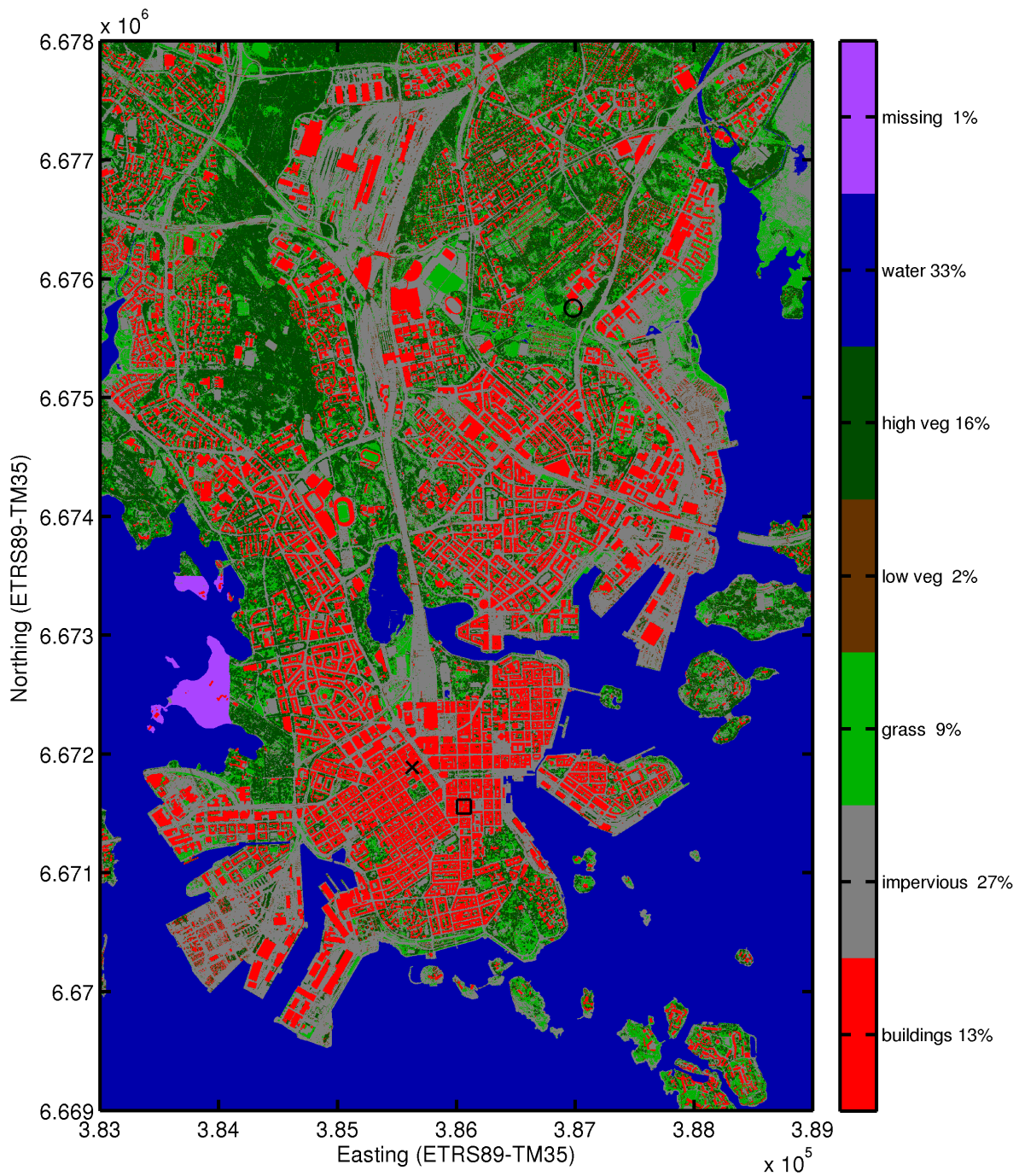
Figure 4. *Surface cover classes in the whole 9 km by 6 km study domain. Colours per class and the fraction of the class are given in the colourbar. Measurement sites are also indicated: Hotel Torni (black cross), Elisa Tower (black square) and SMEAR III – Kumpula (black circle).*
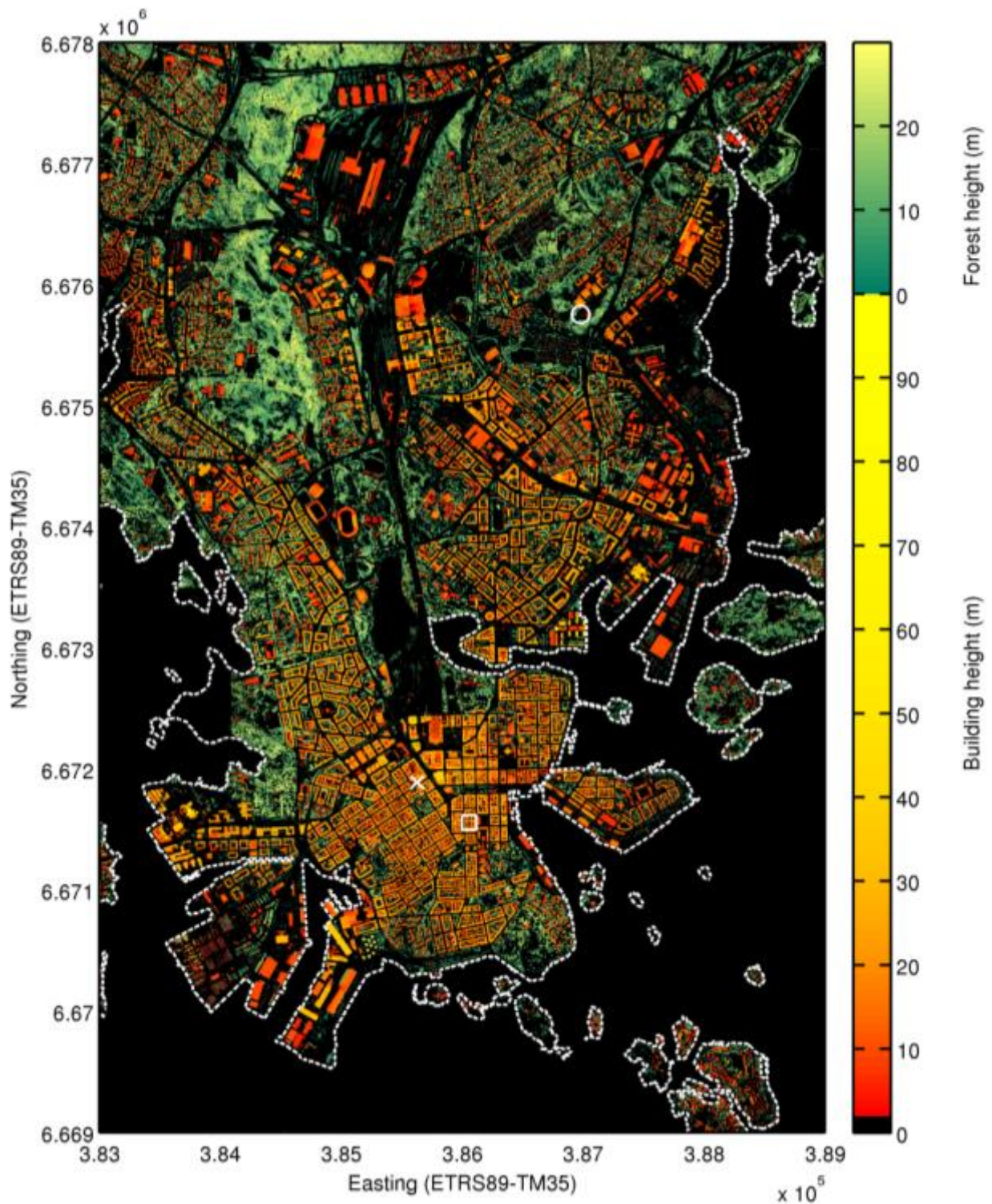
Figure 5. *Building and forest heights (m, see colourbar) in the whole 9 km by 6 km study domain. Forest includes here low and high vegetation. Measurement sites are also indicated: Hotel Torni (white cross), Elisa Tower (white square) and SMEAR III – Kumpula (white circle). The coastline is shown with a white dashed line.*

4.3 Effect of aggregation on surface cover fractions and modelled energy balance

The area around the Hotel Torni measurement station is mainly buildings (37%, Figure 6) and impervious surfaces (41%), followed by grass (7%), low vegetation (3%) and high vegetation (12%). These values are for a 2 m resolution and an area with a radius of 1 km, excluding directions with flow distortion (thus direct comparison with Nordbo et al., 2013 is not possible). The high-resolution vegetation fraction of 22%, which includes all three vegetation classes, drops to half when the resolution is aggregated to 30 m and further to a third of its original value when the resolution is dropped to 100 m (Figure 6). Street trees disappear almost completely already at a 10 m resolution (not shown). The fraction of impervious surfaces remains nearly constant whereas the fraction of buildings increases up to 50% when the resolution is 100 m. To conclude, aggregation favours the most dominant surface types on the expense of the less common types.
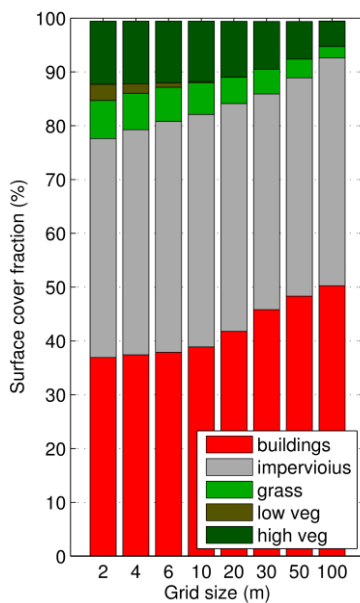

Figure 6. *Surface cover fraction (%) within 1 km from the Hotel Torni measurement site as a function of grid size (m). The surface cover map was aggregated with a mode filter for the shown grid sizes. Only wind directions which are not affected with mast flow distortion are included (see **Figure 3**).*

Eight model runs were made for the different surface cover resolutions in Figure 6, and the simulated turbulent heat fluxes were compared to measured ones. The average diurnal course of measured H ranged from 10 W m^{-2} in the night, to 230 W m^{-2} afternoon (not shown). LE varied from a nocturnal flux of 17 W m^{-2} to an afternoon value of 92 W m^{-2}. The SUEWS model overestimated H on average only by 2 W m^{-2}, when using a 2 m resolution surface-cover map, but there was a night-time underestimation of 15 W m^{-2} and a day-time overestimation of 8 W m^{-2}. The latent heat flux was underestimated by the model throughout the day by 23 W m^{-2}, on average. This led to a modelled median day-time Bowen ratio ($\beta=H/LE$) of 4.0 whereas the measured value was 1.8. The LE underestimation is commonly observed in urban energy balance modelling studies (e.g. Grimmond et al., 2010). In a dense city centre, the anthropogenic water emissions from people and traffic are detected by measurements (Sailor et al. 2011), but they are not included in the model, which can cause some discrepancy between the results.

Table 6. Error statistics for the different grid sizes. CRMSE = centralized root mean square error, RMSE = root mean square error, RMSES = systematic root mean square error, RMSEU = unsystematic root mean square error, MBE = mean bias error and d = index of agreement (Grimmond et al., 2010; Willmott et al., 1985).

	Grid size (m)	BIAS (W m^{-2})	CRMSE (W m^{-2})	RMSE (W m^{-2})	RMSES (W m^{-2})	RMSEU (W m^{-2})	MBE (W m^{-2})	*d*
	2	2.1	66.1	66.1	2.4	66.1	2.1	0.90
	4	2.5	66.1	66.2	2.8	66.1	2.5	0.90
	6	4.8	67.4	67.6	4.9	67.4	4.8	0.90
H	10	6.5	68.3	68.6	6.9	68.3	6.5	0.90
	20	8.9	70.4	71.0	10.5	70.2	8.9	0.89
	30	14.6	75.0	76.4	18.8	74.1	14.6	0.89
	50	18.2	80.0	82.0	25.7	77.9	18.2	0.88
	100	19.8	79.0	81.4	26.0	77.2	19.8	0.88
	2	−23.2	43.4	49.2	42.2	25.4	−23.2	0.59
	4	−24.1	43.4	49.7	43.2	24.6	−24.1	0.58
	6	−26.4	43.6	51.0	45.5	23.0	−26.4	0.56
LE	10	−28.4	44.0	52.3	47.6	21.7	−28.4	0.54
	20	−29.5	44.3	53.2	48.8	21.2	−29.5	0.53
	30	−33.7	45.6	56.7	53.3	19.6	−33.7	0.50
	50	−33.5	45.7	56.7	53.1	19.9	−33.5	0.50
	100	−38.4	47.7	61.2	58.4	18.4	−38.4	0.47

When the resolution of the surface cover map is decreased (4 m – 100 m), the correspondence between the measured and simulated fluxes worsens systematically with all calculated statistics (**Figure 7**, Table 6). The bias of *H* increases from 2 W m^{-2} up to 20 W m^{-2} for 100 m resolution (Figure 7a), which corresponds to a 16% change in flux. This is due to more energy being transferred via sensible rather than latent heat flux when the fraction of vegetation decreases, as is expected. In the same manner, the underestimation of *LE* grows from 23 W m^{-2} to 38 W m^{-2} (**Figure 7**a) which corresponds to a high 56% change in flux since *LE* was originally smaller

than *H*. Note that already an aggregation from 2 m to 10 m increases *H* by 2% and decreases *LE* by 19%. Due to the high sensitivity, the median *β* increases from 4.0 to an extreme value of 23.6 when dropping the resolution to 100 m. Furthermore, the CRMSD for *H* increases from 66 W m^{-2} to 80 W m^{-2} and that of *LE* from 44 W m^{-2} to 48 W m^{-2} (**Figure 7**b) which shows that the scatter in the fluxes increases in addition to the systematic bias. Similarly increase is also seen in other RMSE's (Table 6). For H, RMSEU is throughout larger than the RMSES, whereas for LE, the situation is vice versa. This indicates that in the case of LE the model performance could likely be increased e.g. by changing parameter values or accounting anthropogenic water emissions, but for H the difference is of random origin.

These results apply only to the growing season and outside this the importance of patchy vegetation to the energy partitioning is likely to be smaller. At such times, the surface fraction of buildings due to anthropogenic activities is likely to be more important.
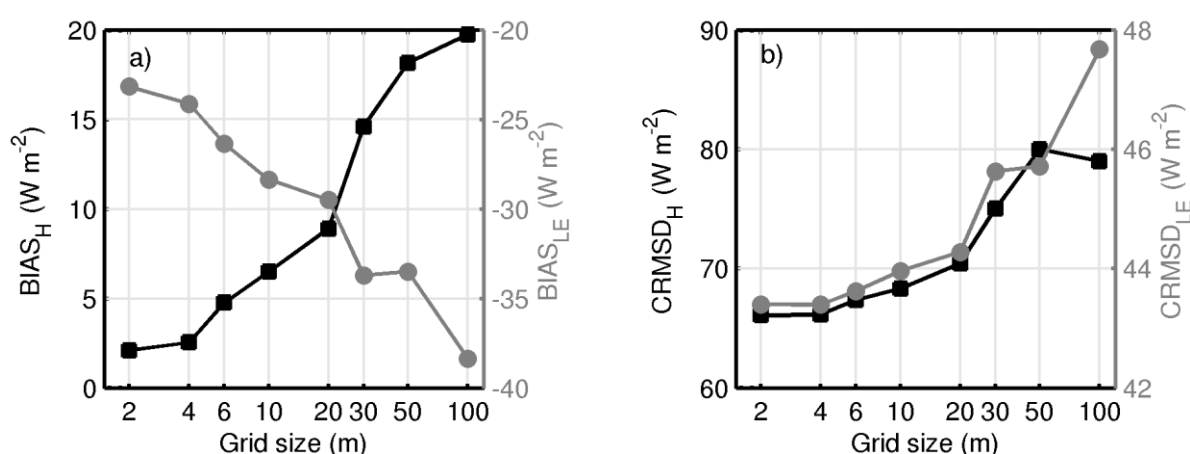

Figure 7. *a) Bias between the simulated and measured flux and b) centralized root-mean-square deviation as a function of land cover grid size. Values are given for sensible heat flux (left y-axis, black) and for latent heat (right y-axis, grey) in both subplots.*

5. Summary and conclusions

The study showed one of the largest published studies of an urban to suburban area (54 km^2) classified based on airborne lidar data. Individual lidar returns were classified instead of pixels or segments using solely classification trees (overall accuracy 91%). The accuracy was worst when distinguishing impervious and grass, because the separation relies almost solely on return intensity. Number of returns-per-pulse distinguished buildings from trees but buildings with special roofs, such as glass structures, were the most challenging to classify. The classified lidar returns were further converted into pixels, filtered at a pixel level, and combined with building and water masks, leading to a very-fine 2m spatial resolution and a final accuracy of 95%.

We recommend, when classifying individual lidar returns, using data on return height above ground and intensity, number of returns-per-pulse and the height difference between the first and last return (last being least essential). We also recommend using separate classification trees for individual flights to avoid the problem of cross-calibration of intensity data. The use of a flood-fill algorithm was found useful for defining rivers where there are no lidar returns.

The final surface cover map was aggregated from 2 m to 100 m and a surface energy balance model was run with the different surface cover resolutions and the simulation results were compared with measured data.

The fraction of vegetation dropped to a third of its original value when the resolution is reduced to 100 m. This led to a 16% increase in simulated sensible heat flux and a 56% reduction in latent heat flux since available energy was directed to sensible heat rather than evaporation due to the lack of vegetation. All statistics showed systematic worsening of the model performance with increasing horizontal surface resolution. We recommend using a high-resolution surface cover map if the city in concern has street trees or other type of patchy vegetation. The error in *LE* caused by a 10 m resolution map is already large (19%) at our study site.
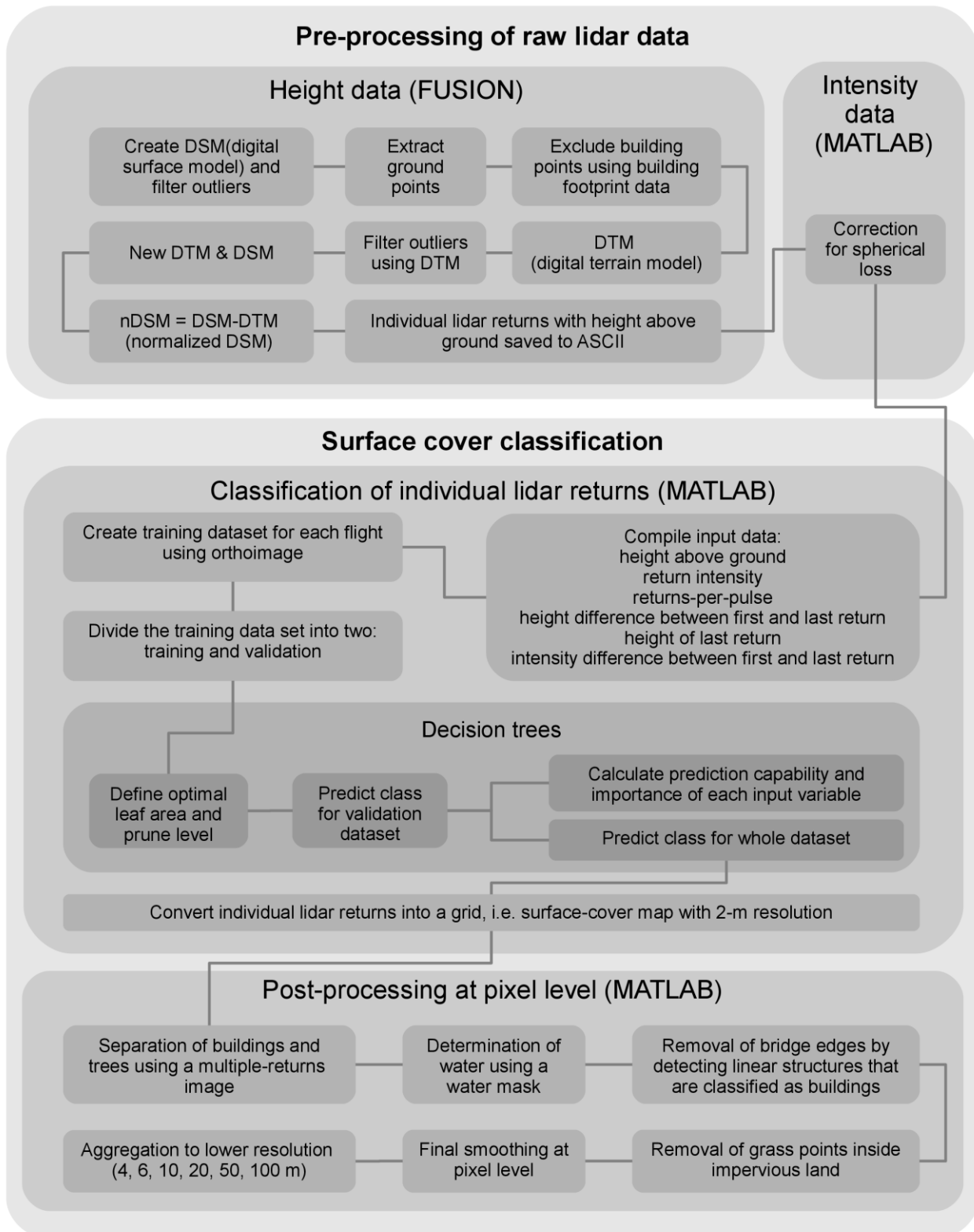
Acknowledgements

Appendix

Figure A.1 Summary of surface cover classification methods.

References

Ahokas, E., Kaasalainen, S., Hyyppä, J. and Suomalainen, J. 2006. Calibration of the Optech ALTM 3100 laser scanner intensity data using brightness targets. ISPRS Comission I Symposium.

Arnfield, A.J. 2003. Two decades of urban climate research: A review of turbulence, exchanges of energy and water, and the urban heat island. *Int. J. Climatol.,* 23, 1-26.

Aubinet, M., Vesala, T. and Papale, D. 2012. Eddy Covariance - A Practical Guide to Measurement and Data Analysis. *Springer,*

Beraldin, J. -., Blais, F. and Lohr, U. 2010. Laser scanning technology. In: *Airborne and Terrestrial Laser Scanning* (G. Vosselman and H. -. Maas). Whittles Publishing, Dunbeath, Scotland, UK, 1-42.

Blaschke, T. 2010. Object based image analysis for remote sensing. *Isprs Journal of Photogrammetry and Remote Sensing,* 65, 2-16.

Breiman, L., Friedman, J. H., Olshen, R. A. and Stone, C. J. 1984. *Classification and Regression Trees* (Anonymous). Wadsworth, Inc., Belmont, CA, USA,

Brennan, R. and Webster, T.L. 2006. Object-oriented land cover classification of lidar-derived surfaces. *Can. J. Remote Sens.,* 32, 162-172.

Buján, S., Gonzalez-Ferreiro, E., Reyes-Bueno, F., Barreiro-Fernandez, L., Crecente, R. and co-authors. 2012. Land Use Classification from Lidar Data and Ortho-Images in a Rural Area. *Photogramm. Rec.,* 27, 401-422.

Coutts, A.M., Tapper, N.J., Beringer, J., Loughnan, M., Demuzere, M. 2013. Watering our Cities: the capacity for Water Sensitive Urban Design to support urban cooling and improve human thermal comfort in the Australian context.' *Progress in Physical Geography*, 37, 2–28.

Goodwin, N.R., Coops, N.C., Tooke, T.R., Christen, A. and Voogt, J.A. 2009. Characterizing urban surface cover and structure with airborne lidar technology. *Canadian Journal of Remote Sensing,* 35, 297-309.

Grimmond, C.S.B. and Oke, T.R. 1991. An Evapotranspiration-Interception Model for Urban Areas. *Water Resour. Res.,* 27, 1739-1755.

Grimmond, C.S.B., Blackett, M., Best, M.J., Barlow, J., Baik, J.-. and co-authors. 2010. The International Urban Energy Balance Models Comparison Project: First Results from Phase 1. *J. Appl. Meteorol. Climatol.,* 49, 1268-1292.

Guan, H., Ji, Z., Zhong, L., Li, J. and Ren, Q. 2013. Partially supervised hierarchical classification for urban features from lidar data with aerial imagery. *Int. J. Remote Sens.,* 34, 190-210.

Hirano, Y., Yasuoka, Y. and Ichinose, T. 2004. Urban climate simulation by incorporating satellite-derived vegetation cover distribution into a mesoscale meteorological model. *Theoretical and Applied Climatology,* 79, 175-184.

HSY. 2008. Seutu CD - a dataset by the Helsinki Region Environmental Services Authority.

Im, J., Jensen, J.R. and Hodgson, M.E. 2008. Object-based land cover classification using high-posting-density LiDAR data. *Giscience & Remote Sensing,* 45, 209-228.

Järvi, L., Grimmond, C.S.B. and Christen, A. 2011. The Surface Urban Energy and Water Balance Scheme (SUEWS): Evaluation in Los Angeles and Vancouver. *Journal of Hydrology,* 411, 219-237.

Järvi, L., Grimmond, C.S.B., Taka, M., Nordbo, A., Setälä, H. and co-authors. 2014. Developments of the Surface Urban Energy and Water balance Scheme (SUEWS) for cold climate cities. *In press in Geoscientific Model Developments,*

Järvi, L., Nordbo, A., Junninen, H., Riikonen, A., Moilanen, J. and co-authors. 2012. Seasonal and annual variation of carbon dioxide surface fluxes in Helsinki, Finland, in 2006-2010. *Atmos. Chem. Phys. Discuss.,* 12, 8355-8396.

Järvi, L., Hannuniemi, H., Hussein, T., Junninen, H., Aalto, P.P. and co-authors. 2009. The urban measurement station SMEAR III: Continuous monitoring of air pollution and surface-atmosphere interactions in Helsinki, Finland. *Boreal Environ. Res.,* 14, 86-109.

Kaasalainen, S., Pyysalo, U., Krooks, A., Vain, A., Kukko, A. and co-authors. 2011. Absolute Radiometric Calibration of ALS Intensity Data: Effects on Accuracy and Target Classification. *Sensors,* 11, 10586-10602.

Kaasalainen, S., Niittymaki, H., Krooks, A., Koch, K., Kaartinen, H. and co-authors. 2010. Effect of Target Moisture on Laser Scanner Intensity. *IEEE Trans. Geosci. Remote Sens.,* 48, 2128-2136.

Kim, S., McGaughey, R.J., Andersen, H. and Schreuder, G. 2009. Tree species differentiation using intensity data derived from leaf-on and leaf-off airborne laser scanner data. *Remote Sens. Environ.,* 113, 1575-1586.

Lindberg, F. and Grimmond, C.S.B. 2011. Nature of vegetation and building morphology characteristics across a city: Influence on shadow patterns and mean radiant temperatures in London. *Urban Ecosyst.,* 14, 617-634.

Loridan, T. and Grimmond, C.S.B. 2012. Characterization of Energy Flux Partitioning in Urban Environments: Links with Surface Seasonal Properties. *Journal of Applied Meteorology and Climatology,* 51, 219-241.

Mancini, A., Frontoni, E. and Zingaretti, P. 2009. Automatic extraction of urban objects from multi-source aerial data. Proceedings of CMRT09.

Matikainen, L. and Karila, K. 2011. Segment-Based Land Cover Mapping of a Suburban Area-Comparison of High-Resolution Remotely Sensed Datasets Using Classification Trees and Test Field Points. *Remote Sensing,* 3, 1777-1804.

Matikainen, L., Hyyppa, J., Ahokas, E., Markelin, L. and Kaartinen, H. 2010. Automatic Detection of Buildings and Changes in Buildings for Updating of Maps. *Remote Sensing,* 2, 1217-1248.

McGaughey, R.J. 2009. FUSION/LDV: Software for LiDAR data analysis and visualization. United States. *Department of Agriculture, Seattle,*

National Land Survey of Finland. 2012. PaITuli - Spatial data for research and teaching.

Nordbo, A., Järvi, L. and Vesala, T. 2012a. Revised eddy covariance flux calculation methodologies - effect on urban energy balance. *Tellus Ser. B-Chem. Phys. Meteorol.,* 64, 18184.

Nordbo, A., Järvi, L., Haapanala, S., Moilanen, J. and Vesala, T. 2013. Intra-City Variation in Urban Morphology and Turbulence Structure in Helsinki, Finland. *Bound. -Layer Meteorol.,* 146, 469-496.

Nordbo, A., Järvi, L., Haapanala, S., Wood, C.R. and Vesala, T. 2012b. Fraction of natural area as main predictor of net CO2 emissions from cities. *Geophys. Res. Lett.,* 39, L20802.

O'Neil-Dunne, J.P.M., MacFaden, S.W., Royar, A.R. and Pelletier, K.C. 2013. An object-based system for LiDAR data fusion and feature extraction. *Geocarto Int.,* 28, 227-242.

Pirinen, P., Simola, H., Aalto, J., Kaukoranta, J.-., Karlsson, P. and co-authors. 2012. Tilastoja Suomen Ilmastosta 1981-2010 (Climatological Statistics of Finland 1971-2000). *Finnish Meteorological Institute Reports,* 1, 1-96.

Sailor, D. J. 2011. A review of methods for estimating anthropogenic heat and moisture emissions in the urban environment. *Int. J. Climatol*., 31, 189–199.

Simpson, J.R. 2002. Improved estimates of tree-shade effects on residential energy use. *Energy Build.,* 34, 1067-1076.

Tooke, T.R., van der Laan, M. and Coops, N.C. 2014. Mapping demand for residential building thermal energy services using airborne LiDAR. *Appl. Energy,* 127, 125-134.

Tooke, T.R., Coops, N.C., Christen, A., Gurtuna, O. and Prevot, A. 2012. Integrated irradiance modelling in the urban environment based on remotely sensed data. *Solar Energy,* 86, 2923-2934.

Willmott, C.J. et al., 1985. STATISTICS FOR THE EVALUATION AND COMPARISON OF MODELS. *Journal of Geophysical Research-Oceans*, 90(NC5): 8995–9005.

Wood, C., Järvi, L., Kouznetsov, R., Nordbo, A., Joffre, S. and co-authors. 2013. An overview on the Urban Boundary-layer Atmosphere Network in Helsinki. *BAMS,* 94, 1675-1690.

Xiao, Q., McPherson, E.G., Simpson, J.R. and Ustin, S.L. 2007. Hydrologic processes at the urban residential scale. *Hydrol. Process.,* 21, 2174-2188.

Zhou, W. 2013. An Object-Based Approach for Urban Land Cover Classification: Integrating LiDAR Height and Intensity Data. *Ieee Geoscience and Remote Sensing Letters,* 10, 928-931.