Student thesis series INES nr 234

Modelling near ground wind speed in urban environments using high-resolution digital surface models and statistical methods

Lars Johansson

2012
Department of
Physical Geography and Ecosystems Science
Lund University
Sölvegatan 12
S-223 62 Lund
Sweden

Lars Johansson (2012). Modelling near ground wind speed in urban environments using high-resolution digital surface models and statistical methods
Master degree thesis, 30 credits in Geomatics
Department of Physical Geography and Ecosystems Science, Lund University

Modelling near ground wind speed in urban environments using high-resolution digital surface models and statistical methods

Lars Johansson

Master thesis, 30 credits, in Geomatics

Supervisors:

Jonathan Seaquist
Department of Physical Geography and Ecosystems Science
Lund University

Fredrik Lindberg
Department of Earth Sciences
University of Gothenburg

Acknowledgements

I gratefully appreciate all the efforts of my supervisors; Fredrik for suggesting the subject and for all help along the way, and Jonathan for his remarkable ability to ask the right questions and for the interesting discussions that follow such questions.
I thank Shiho Onumura at the Department of Earth Sciences, University of Gothenburg for her efforts with the CFD-modelling. Sofia Thorsson, Björn Holmer and Shiho read the thesis and gave very helpful comments, which I very much appreciate. Without the endless support and encouragement from Marja, this thesis would never have been written. Thank you so much.

Abstract

Wind is a complex phenomenon and a critical factor in assessing climatic conditions and pedestrian comfort within cities. This master's thesis attempts to quantify and model the relationship between near ground wind speed and urban geometry using two-dimensional raster data and variable selection methods based on Multiple Linear Regression. Such model can be implemented in a Geographic Information System (GIS) to assess spatial distribution of wind speed at the street scale in complex urban environments.

Wind speed data two meters above ground is obtained from simulations by computer fluid mechanics modelling (CFD) and used as a response variable. Utilizing a shadow-casting raster algorithm, four measures of urban geometry are derived from high-resolution surface models (DSM): Sky View Factor (SVF), Fetch, Frontal Area Index (FAI) and Angular Frontal Area Index (αFAI). To compute Fetch and FAI, the shadow-casting algorithm needs a search angle and search distance as input parameters. In order to evaluate the effect of these parameters, a number of settings are tested resulting in a total of 53 different predictors. Four DSMs and three wind directions are combined, resulting in twelve unique datasets from which observations are sampled. A sequential variable selection algorithm followed by all-possible subset regression was used to select candidate models for further evaluation.

The results show that models including SVF and Fetch explain general spatial wind speed pattern characteristics, but the prediction errors are large, especially so in areas with high wind speeds. However, all selected models did explain 90 % of the wind speed variability ($R^2 \approx 0.90$). The RMSE of normalized predicted wind speed (amplification factor) for city models ranges from between 0.01 (dense building geometry) to 0.22 (wide street aligned in the wind direction with air flow channelling). The differences between the selected candidate models are less than 0.02, and no model consistently performs "best" over all city models. Models that include FAI and αFAI did not improve on these results. Predictors adding information on width and height ratio and alignment of street canyons with respect to wind direction are possible developments to improve model performance. To assess the applicability of any derived model, the results of the CFD model should be thoroughly evaluated against field measurements.

Keywords: Geography, Physical Geography, Geomatics, Urban climate, Urban geometry, Wind speed, Digital surface model, Spatial variability, Geographic Information System

Sammanfattning

Vind är ett komplext fenomen och en viktig faktor att ta hänsyn till i studier relaterade till stadsklimat och klimatkomfort i städer. Föreliggande uppsats är ett försök att kvantifiera sambandet mellan marknära vind och stadens geometri utifrån tvådimensionella rasterdata och linjär multipelregression. En sådan regressionsmodell kan användas i ett geografiskt informationssystem (GIS) för att beräkna rumslig fördelning av vindhastighet i komplexa urbana miljöer.

Genom tredimensionell modellering med en Computer Fluid Dynamics-modell (CFD) erhölls vindhastighets-data för ett horisontellt skikt två meter över marken som här används som responsvariabel. Vidare användes en skuggkastningsalgoritm som utifrån en högupplöst höjdmodell, digital surface model (DSM), beräknar fyra olika mått på urban geometri: Sky View Factor (SVF), Fetch, Frontal Area Index (FAI) och Angular Frontal Area Index (αFAI). Algoritmen använder två olika parametrar för beräkningen av dessa derivat av höjdmodellen, sökavstånd och sökvinkel, och olika värden användes för att undersöka effekten av parametrarna. Detta gav totalt 53 olika prediktorvariabler beskrivande den urbana geometrin. Fyra olika höjdmodeller och tre olika vindriktningar genererade tolv unika dataset, stadsmodeller, från vilka cirka 2800 observationer av samtliga variabler samplades. Variablerna med största förklaringsgraden utgallrades och de bästa kombinationerna, modellerna, av dessa identifierades med hjälp av regression på alla möjliga kombinationer av de valda variablerna.

Resultaten visar att modeller med endast SVF och Fetch förklarar generella vindhastighetsmönster, men felen är relativt stora, särskilt i områden med höga vindhastigheter. Alla valda modeller förklarar dock cirka 90 % av vindhastighetens variation ($R^2 \approx 0.90$). RMSE för förstärkningsfaktor varierar mellan olika stadsmodeller; 0,01 (tät bebyggelse) till 0,22 (bred gatukanjon med kanalisering av vindflödet). Skillnaden mellan de valda modellerna är mindre än 0,02 och ingen av dem är konsekvent bättre än de andra över alla tolv stadsmodeller. FAI och αFAI ökade inte förklaringsgraden. Variabler som beskriver gatukanjonernas riktning i förhållande till vindriktningen och kanjonernas höjd och breddförhållande är intressanta utvecklingsmöjligheter. För att bedöma en modells användbarhet, måste resultatet av CFD-modelleringen utvärderas med hjälp av mätningar i fält.

Nyckelord: Geografi, Naturgeografi, Geomatik, Stadsklimat, Bebyggelsegeometri, Vindhastighet, Digital höjdmodell, Rumslig variation, Geografiska informationssystem.

Table of Contents

List of Figures

List of Tables

1 Introduction

The study presented in this master's thesis in geomatics is a first attempt to quantify and model the relationship between near ground wind speed and urban geometry using continuous two-dimensional raster data within a GIS. Wind speed is a very complex phenomenon and a critical factor in assessing climatic conditions and pedestrian comfort within cities. A simple computational approach may therefore be a useful GIS-tool in both research and management applications.

1.1 Background and motivation

The urban environment is receiving more attention in many research fields, including both physical and human geography disciplines. A city may be viewed as a spatial aggregation of buildings of different size, shape and distribution, which defines the urban canopy. This can have a large impact on the physical environment on both local and regional scales, for example by affecting the natural hydrological cycle or the radiation balance at the Earth's surface. The city also offers activities and opportunities that attract people, leading to increased global urbanisation and generating a need for research on critical environmental issues, many of them related to the urban climate.

The urban climate results from human activities and from the modification of surface conditions. New surface materials, associated with buildings and roads along with changes to the surface morphology contribute to this new set of conditions (Grimmond 2007). Human activities are associated with the emission of heat, moisture and pollutants. Wind speed, wind direction, air pollution, driving rain, radiation and daylight are examples of physical quantities that constitute the outdoor climate and change due to the shape, size and orientation of buildings and other obstacles such as trees etc. Both increased and decreased wind speed may potentially be regarded as unfavourable changes leading to uncomfortable conditions for pedestrians and insufficient removal and accumulation of air pollutants. Increased wind speed at pedestrian level is one of the problems that are considered most important, and studies of the pedestrian wind environment for large construction projects are required by many urban authorities (Blocken & Carmeliet 2004).

Wind in the urban environment is a highly complex phenomenon mainly dependent on the local geometry within the city structure. Traditionally wind flow is investigated using wind tunnel experiments or full-scale field measurements (S. Kim & Boysan

1999). An increasingly popular way to describe airflow in the urban environment is by using three-dimensional computational fluid dynamic (CFD) models where numerical methods are utilized to simulate the interaction between the atmosphere and the urban surface (Blocken & Carmeliet 2004; Murena et al. 2009). However, CFD-models are computational intense and allows for only a limited spatial extent to be analyzed (Lindberg et al. 2008).

Assuming wind speed at the pedestrian level is dependent on the local urban geometry, it could be predicted if the relationship between wind speed and the geometry were quantified; i.e. if there were an equation, or model, defining this relationship. Within a GIS, measures of the urban geometry may be represented by two-dimensional raster data computed from detailed urban elevation models, thus allowing for a computationally efficient implementation of a model utilizing local raster operations.

Sky View Factor (SVF) and *Frontal Area Index* (FAI) are two measures that are used in urban climatic research in order to quantify the spatial variation of urban geometry. SVF is a parameter that can be expressed as the ratio between radiation received (or emitted) by a planar surface and that received from (or emitted to) the entire hemisphere, which has been used in a variety of climate related studies such as energy exchange, spatial variations of urban air and surface temperature patterns and outdoor thermal comfort, as well as in the modeling of the urban climate and in urban planning (Lindberg & Grimmond 2010). FAI has been used in efforts to parameterize aerodynamic properties of the urban surface such as zero-plane displacement and roughness length (Grimmond & Oke 1999) and in mapping of urban roughness, ventilation paths, and wind dynamics (Gal & Unger 2009; Wong et al. 2010; Chen & Ng 2011). A novel measure related to FAI is introduced in this study, *Angular Frontal Area Index* (αFAI). Together with *Fetch*, which measures the distance to a change in aerodynamic properties of a surface (Oke 1988), these three measures are derived to represent urban geometry in this study.

1.2 Objectives

The overall objective with this study is to build a model for the prediction of pedestrian wind speed in complex urban settings using measures of urban geometry as explanatory variables (predictors). To achieve this, the influence on wind speed of four different measures is analyzed using statistical methods. These four measures, the predictors, are: SVF, Fetch, FAI, and αFAI.

The specific objectives of this study are:

- Investigate the relationships between wind speed and the predictors.

- Identify those measures, or combinations of measures of urban geometry that explain most of the wind speed variability.
- Evaluate possible candidate models by mapping predicted values and residuals.

The strategy is to use raster datasets representing simulated wind speed data and measures of urban geometry as response variable and predictor variables respectively. The predictor variables are derivatives of high-resolution digital surface data (DSM) computed by utilizing raster algorithms in MATLAB. The DSMs represent the spatial variation of the geometry in a complex real world urban environment. Wind speed data is simulated using a CFD-model, ENVI-met. Based on observations (samples) from these datasets, Multiple Linear Regression techniques are used to select important variables and to build models.

Hereafter, a model is referred to as a regression equation predicting mean wind speed (y) including one or more measures of urban geometry as explanatory variables ($x_1 \ldots x_i$). Such model can be implemented in a raster-based GIS application utilizing local operations on raster data.

It is important to note that any suggested model in this work is a model predicting wind speed simulated by a CFD-model, not wind speed as measured in the real world. To evaluate the results of the CFD-model using field measurements is not within the scope of this work.

1.3 Thesis structure

This thesis is divided into seven chapters. In Chapter 2, focus is on wind flow within the urban canopy. The impact of buildings and building arrangements on wind conditions are discussed and some research efforts to model wind speed are presented. In Chapter 3, some important aspects of regression analysis are examined with focus on variable selection and model building. The methodology of the study is outlined in Chapter 4 including data preparation, computation of DSM-derivatives and application of statistical methods. The results are presented in Chapter 5 followed by a discussion and the conclusions, Chapter 6 and 7.

2 Wind and the urban environment

The urban surface is one of the roughest on the Earth, which is why cities have major implications for the air flow at all scales (Grimmond & Oke 1999). The impact of the geometry of the urban environment is the single most important factor controlling wind at the local scale (streets), but it also contributes to the control of wind flow patterns in the surrounding rural areas, for example by trapping radiation, giving rise to the urban heat island and its associated wind flows. Thus, cities has a major influence on the *planetary boundary layer*, which is the part of the atmosphere where the conditions are characterized by the interaction between the atmosphere and the Earth's surface. Thus, compared to the surrounding areas, a city usually provides a rougher surface, but also a warmer and maybe drier set of surface conditions. As air flows from the countryside to the city, it adjusts to this new set of conditions and the *urban boundary layer* (UBL), which is a local to meso-scale phenomenon within the planetary boundary layer develops. It extends above cities, where the conditions are governed by the characteristics of the urban surface (Oke 1988). A closer look at the city reveals the *urban canopy layer* (UCL), which is the layer below roof level. This layer is generated by micro-scale processes between the buildings (Oke 1988; Arnfield 2003). The climate, and thus wind speed at any location within the UCL is controlled by the surface characteristics and the geometry in the immediate surroundings (Hunter et al. 1991), which is the topic of this chapter.

2.1 Wind flow in the urban canopy layer

The basic geometric unit of the urban environment is the street canyon, defined by a road and its adjacent buildings. The microclimate characteristics of a street canyon depends on its orientation with respect to incoming solar radiation and angle-of-attack of the wind, construction materials and canyon geometry (Oke 1988), that is the height, shape and distribution of the buildings. The wind flow characteristics of the street canyon has been the subject to numerous of studies; in the field e.g. (Nakamura & Oke 1988; Eliasson et al. 2006), by wind tunnel and water channel experiments, e.g. (Johnson & Hunter 1999; Princevac et al. 2010) and by the use of numerical models e.g. (Hunter et al. 1991; Baik & Kim 1999). To summarize, the wind flow in a street canyon is driven and determined by the interaction of the flow field above buildings and local effects such as topography, building geometry, traffic and other local features (Georgakis & Santamouris 2008).

When air encounters a building, it is deflected over the top, down the front (the windward wall) and around the sides. At the edges of the windward wall, the air

accelerates and the flow may separate from the building surface, initiating a reverse flow in low-pressure zones on the sides, top and leeward side of the building. On the leeward side, this recirculation (the wake zone) extends approximately two times the building height in the downwind direction. Above the lee eddy, converging air deflected over the roof produces a jet of high velocity. In front of the building, the air streaming down the windward wall produces a vortex at ground level with a recirculation in upwind direction. At some distance in front of the building, this counter flow meets incoming air and here a stagnation zone develops (Oke 1988; Wieringa et al. 2001; Blocken & Carmeliet 2004).

In general, wind speeds in the urban canopy layer are reduced compared to surrounding rural areas. Local high wind speeds at pedestrian level are associated with corners streams and the windward vortex. These effects are enhanced by tall buildings, which deflect faster moving upper air, thus increasing wind speed at the pedestrian level. Air that streams downward the windward side of a tall building may also enhance the lee eddy of any upwind building producing a strong vortex which may give wind speeds higher than those in the open (Oke 1988; Blocken & Carmeliet 2004).

Applied to a street canyon, this implies that different spacing between the buildings, i.e. canyon widths, with respect to building height may cause different flow patterns due to the interaction between upwind and downwind effects of buildings. Oke (1988) used the height to width ratio (H/W) of street canyons to categorize wind flow into three flow regimes: (1) isolated roughness flow, (2) wake interference flow and (3) skimming flow, Figure 1.

In the case of isolated roughness flow, the buildings are relatively widely spaced ($H/W < 0.3$) and other buildings do not affect the flow around them as if they were isolated obstacles (Figure 1a). As distance increases downwind from a building, the sheltering effect decreases and eventually the jet merges with the flow and readjusts to its undisturbed form (Oke 1988). Thus, in isolated roughness flow, the wind speed is expected to increase in the downwind direction behind a building fuelled by the merging jet, before encountering the next building downwind where a stagnation zone develops resulting in a decrease in wind speed.

At closer spacing ($0.3 < H/W < 0.7$) the wake zone of the upwind building interferes with the upwind stagnation zone of the next building downwind. This is the wake interference flow (Figure 1b). At more dense arrangements ($H/W < 0.7$), the main flow starts to skim above the roofs, driving a lee vortex in the cavity between the buildings beneath, i.e. in the street canyons (Figure 1c). In this situation, the bulk of the flow does not enter inside the street canyon. In a study using numerical methods,

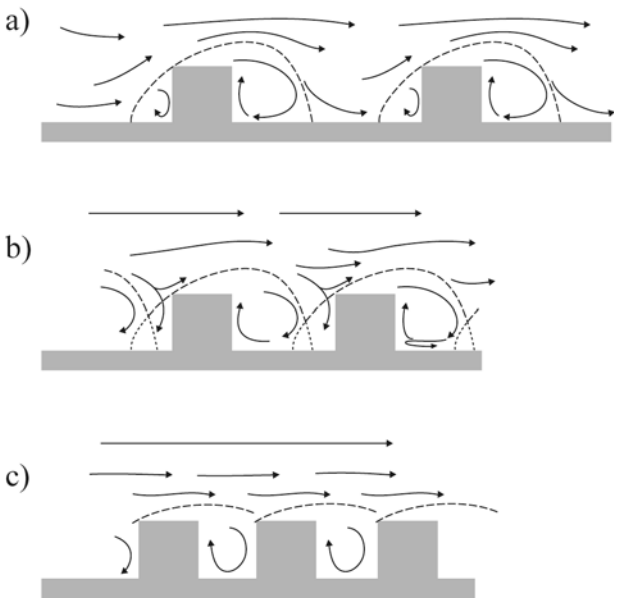

Figure 1 Flow regimes identified by Oke (1988): a) isolated roughness flow, b) wake interference flow, and c) skimming flow. (Adapted from Oke 1988).

Hunter et al. (1991) found evidence suggesting that the canyon length affects the threshold H/W ratios for the transition from one flow regime to another.

When the flow is perpendicular to the canyon axis, the recirculation on the canyon floor is directly opposite (G. T. Johnson & Hunter 1999) and for a H/W-ratio near unity and roof-top winds exceeding 1.5–2 ms^{-1}, the recirculation wind speed at the canyon bottom is of the order 1/3–1/2 of the roof-top wind (Britter & Hanna 2003). If the wind is oriented at some oblique angle to the canyon direction, the canyon vortex will still develop but having an along canyon component giving it a “cork-screw” motion (Britter & Hanna 2003; Oke 1988). The along-street channelling depends linearly on the component of the above-roof wind parallel to the street and the across-street recirculation vortex depends linearly to the component perpendicular to the street (Dobre et al. 2005). If the above roof-level flow is parallel to the canyon, channelling may cause wind speeds greater than in the open (Oke 1988) and the flow direction at the canyon floor is the same as above roof level (G. T. Johnson & Hunter 1999). Close to the canyon walls friction retards the wind flow. Hence, wind speed is expected to be higher in the centre of the canyon (Ahmad et al. 2005).

Irregular canyon geometry introduces complications to the flow patterns described above by controlling the production of turbulence both inside and outside the canyon. For example, Kastner-Klein (2004) showed the influence of roof shapes on the wind flow pattern in street canyons. But in general the resulting flow is qualitatively consistent with the general expected flow described above (Carpentieri et al. 2009; Britter & Hanna 2003).

Obviously, the geometry of the urban canopy layer is more complex than the one street canyon. The flow in street intersections is highly complex and strongly three-dimensional compared to the flow in street canyons (Carpentieri et al. 2009). For relatively regular street networks, field and wind tunnel investigations shows that the flow pattern may become more complicated at or near an intersection due to the interaction of winds between crossing street canyons (Wang 2007). When the wind flow parallel to one street reaches an orthogonal intersection, a large recirculation region develops at the entries to the crossing streets (Soulhac et al. 2009). Thus, the effect of street intersection extends into adjacent canyons up to a distance of about one street width (Tiwary et al. 2011). Princevac et al. (2010) experimented on airflow in between arrays of regular building blocks using water channel methods. The experiments showed that the flow along canyons (the main flow) parallel to the incoming flow was partially deflected sideways after the first row of buildings. Moreover, there was a counter flow back into the channel after the second row. The effect was enhanced by the presence of a tall building rising above surrounding buildings. Thus, at and near street intersections it is expected that wind speeds may be affected by either deflection or merging of airflows.

Without exploring this topic any further, it can be concluded that by just increasing the complexity of the urban geometry stepping up from a single regular street canyon to a regular network of intersecting orthogonal streets, the flow patterns become intricately more complex. Moreover, the flow at intersections is very sensitive to minor changes in the geometry (Robins et al. 2002), suggesting that any established relationship between geometry and wind flow at one location may not be generalized to other locations.

3 Statistical concepts

In research where data are observational, meaning that the variance in the response has not been manipulated by experimental design, difficulties in determining which casual factor or factors best explain observed responses may arise. In these situations, Multiple Linear Regression (MLR) is one of the most used techniques for describing such systems and identifying models consistent with the observed phenomenon (Whittingham et al. 2006). It is usually the case that models with fewer predictors have greater generality which is why scientific work usually is directed towards identifying relatively parsimonious models in agreement with observed data (Whittingham et al. 2006). Finding an appropriate subset of predictors to include in the model is often called the variable selection problem (Montgomery et al. 2001).

In the present study, variable and model building techniques based on MLR are used to identify those variables that explain most of the variability in wind speed, the response variable. In this chapter the assumptions and diagnostic methods of linear regression and some aspects of variable selection and model building are discussed.

3.1 Regression analysis and diagnostics

Linear regression analysis begins, as the name implies, with the assumption that there is a linear relationship between the response variable (y) and the explanatory variable (x), the regressor. By fitting a straight line to the data, the outcome is an equation (or model) that can be used to predict values of y from new values of x (Rogerson 2010). Apart from linearity, there are some other assumptions, or requirements, that should be addressed and if necessary corrected for. In short, the assumptions of MLR can by summarized as follows (Montgomery et al. 2001; Rogerson 2010):

1. Linear or approximately linear relationship between the response (y) and the regressors (x).
2. The error term has zero mean.
3. The error term has constant variance.
4. The errors are uncorrelated.
5. The errors are normally distributed.
6. There is no multicollinearity in the data.

In addition to these assumptions, there are some other potential problems that should be diagnosed:

7. Outliers and influential cases.

A linear relationship between any two variables x and y is represented by a straight line and the simple linear regression fit to the data is

$$y = \beta_0 + \beta_1 x_1 + \varepsilon$$

where the intercept β_0 and the slope β_1 are unknown constants (the regression coefficients) and ε is a random error component, the residuals. In the case of MLR, there is more than one regressor in the equation, thus

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k \beta_1 + \varepsilon$$

is the multiple linear regression extension of the simple linear regression equation with k regressors. The regression coefficients are estimated by least-squares estimation (Rogerson 2010; Montgomery et al. 2001).

The residuals may be viewed as the variability in y not explained by the regression model, why problems such as non-linear relationships or non-constant variance in the response becomes a property of the residuals (Montgomery et al. 2001) possible to detect by some residual diagnostic.

In order to use MLR when there is no linear relationship between the original variables, the variables should, if possible, be transformed. Common methods include logarithmic, quadratic or square root transformations. By transformation of the response and/or the regressor variables model inadequacies such as non-normal errors or non-constant variance may be eliminated, or at least reduced (Montgomery et al. 2001). If no subject-matter knowledge on the form of the relationships (i.e. logarithmic, quadratic etc.) is available, the appropriate transformations have to be chosen empirically. Scatter diagrams (response against regressor) can be used to investigate possible linear relationship between two variables but in the case of MLR this may not be adequate. This is particularly the case when several important regressors are interrelated. Instead, residual plots (residuals against predicted values or predicted values against the response) or partial residual plots are preferred (Montgomery et al. 2001). The residuals should be normally distributed and uncorrelated, that is, there should be no evidence of any trends or patters when plotting the residuals against any of the regressors, the response or the predicted values.

When two or more regressors are highly correlated to each other, multicollinearity is said to be present in the data. Effects of multicollinearity are large variances and large absolute values for the least-squares estimation of the regression coefficients. Also the regression coefficients may end up having different signs than expected and are very sensitive to changes in the data. The removal or addition of a few observations may have huge impact on the estimation of the regression parameters. In such cases,

inference based on the regression model may be misleading or erroneous. However, the linear combination of the parameters (the fitted model) may be a good predictor, even though the individual parameters are estimated poorly (Montgomery et al. 2001). The effect of multicollinearity can seriously undermine variable selection algorithms that use significance tests on parameter estimates as selection criteria, e.g. stepwise regression (Graham 2003; Whittingham et al. 2006).

Multicollinearity can be detected by calculating the variance inflation factor, VIF

$$VIF = \left(1 - R_j^2\right)^{-1}$$

where R_j is the coefficient of determination when x_j is regressed on the remaining regressors. A high value of R_j gives high VIFs, indicating a multicollinearity problem. Several critical values have been suggested: VIFs > 10 indicate severe multi-collinearity but values as low as two may have significant impact (Graham 2003). If the aim of the regression analysis is prediction only, and there is no interest in the particular relationships between the response and the regressors, then the problems with multicollinearity can be ignored (Graham 2003). If not, some method to deal with the problem has to be applied.

3.2 Variable selection and model building

To select a few important variables from a large set of candidate regressors is a compromise between two conflicting objectives. First, a model with as many regressors as possible is wanted so that the information content in these factors can influence the prediction of the response. Second, a model with as few regressors as possible is preferred due to the increase in predicted value variance with increased number of regressors. Also, many regressors means greater costs for data collection and model maintenance (Montgomery et al. 2001; J. B. Johnson & Omland 2004; Whittingham et al. 2006). In this study, the number of regressors equals the number of derivatives that have to be computed. Thus cost is equal to the processing load on the application used to compute DSM derivatives.

To solve the variable selection problem, a suite of algorithms has been developed of which stepwise methods are most commonly used and included in most software packages. In stepwise multiple regression, successive addition (forward selection) or removal (backward elimination) of significant or non-significant terms, or a combination of both (stepwise regression), is performed. However, the significance testing approach adopted by the stepwise methods has limitations and the method has been criticised. Whittingham (2006) mention three main shortcomings of stepwise multiple regression methods: (1) bias in parameter estimations (2) inconsistencies among model selection algorithms, and (3) an inappropriate focus or reliance on a

single best model, where data are often inadequate to justify such confidence. Moreover, there is often no explicit criterion of what is the best model (Burnham & Anderson 2004).

If there are not too many candidate variables to choose from, all-possible regression algorithms offer an alternative to stepwise methods. All-possible regression algorithms fit all possible equations involving one regressor, two regressors etc. (Montgomery et al. 2001). By using an appropriate criterion that involve both the fit and complexity, multiple candidate equations can be evaluated, hence avoiding the shortcomings of stepwise methods mentioned above (Whittingham et al. 2006; Johnson & Omland 2004). Competing models can also be evaluated using cross-validation subsets or a separate test set (Luo 2008). In principle, all possible subsets should be tried, but as the number of possible subsets increases exponentially with increasing number of candidate regressors ($2^p - 1$) more than 20–40 regressors is not feasible to include (Luo 2008; Montgomery et al. 2001).

By contrast to traditional hypothesis testing (significance testing) model selection offers a way to draw inferences from a set of competing hypothesis or models (Johnson & Omland 2004). To choose among models, to compare or weight models, several selection criterion are available, for example: R^2, Adjusted R^2, Mallow's C_p, Akaike's information criterion (*AIC*) and Schwart's Bayesian Information Criterion (*BIC*) (Luo 2008). R^2 is not a very good criterion as it always increases when a new term is added to the model. Because of the lack of an optimum value for R^2 it is not straightforward to use for selection (Montgomery et al. 2001). Adjusted R^2 adjusts for the number of included terms and increases only if the improvement of the model by adding a term is better that what would be expected by chance (Luo 2008).

AIC and *BIC* are defined as (Luo 2008):

$$AIC = 2h + n\ln\frac{RSS}{n}$$

$$BIC = h\ln(n) + n\ln\frac{RSS}{n}$$

where *RSS* is the sum of residual squares, h is the number of regressors in the model and n is the number of observations. The equations above have two components: one that measure the model fit to the observed data and one bias correction factor which increases as a function of the number of regressors in the model, discouraging over-fitting (Luo 2008; J. B. Johnson & Omland 2004). This means that both criteria account for both fit and complexity and the model with the lowest *AIC* or *BIC* value has the best relative fit given the number of parameters in the model (J. B. Johnson & Omland 2004; Whittingham et al. 2006).

Mallow's C_p is a criterion related to the mean square error of a fitted value and is calculated as (Mallows 1973):

$$C_p = \frac{RSS_p}{\hat{\sigma}^2} - n + 2p$$

RSS_p is the residual sum of squares for the subset regression with p regressors, $\hat{\sigma}^2$ is an unbiased estimate of σ^2, and n is the number of observations. C_p is frequently calculated by using the mean square error for the full model as an estimate of σ^2. In the presence of redundant variables this may be problematic due to overestimation of σ^2; consequently the values of C_p will be small. If the bias in the model is negligible, it can be shown that $C_p = p$. Thus, a model that is unbiased should have a C_p value close to p for that model and generally small values of are preferred (Montgomery et al. 2001).

The methods described above are objective techniques to identify (possibly) important variables and to screen out those that are (likely to be) redundant. In practise, it is also important to consider subject-matter knowledge on the phenomena under study in order to judge which predictors that may be important. In the present study, the choice of predictors are based on the assumption that the urban geometry is the most important factor controlling wind speed near ground in between buildings. It is also assumed that SVF, Fetch, FAI and αFAI are appropriate measures of the geometry. However, there is no a priori knowledge about the relative effect of these compared to each other. Hence, the methodology applied in the present study once the set of predictors have been computed is data driven and based on objective variable selection techniques.

4 Data and methodology

The purpose of the selected methods is to derive models (equations) predicting wind speed from a set of candidate explanatory variables describing the urban geometry. In this study, a model is equal to a regression equation estimated by MLR. The input data in the regression analysis, wind speed (response variable) and geometry measures (predictor variables) was sampled observations (grid cells) from derivatives of digital surface models (DSM). Four different DSMs, hereafter denoted *City Models*, were constructed as described in section 4.1. The simulation of wind speed was performed using the CFD-model ENVI-met (section 4.2) and the computation of DSM-derivatives by a shadow-casting algorithm implemented in MATLAB is described in section 4.3. The regression analysis and variable selection methodology are outlined in section 4.4.

4.1 City models

As test cases, four areas in Göteborg city centre were used to construct four different City Models reflecting different types of urban morphologies; dense building structure with narrow canyons, open areas and canyons with varying width and orientation, (Figure 2 and Table 1). For each City Model a DSM was obtained by vector to raster conversion of a detailed governmental database (Lindberg 2007). The resolution of the resulting grids was 2 meters and the geographic extents were 242×242 meters, i.e. 121×121 pixels. To facilitate the needs of the ENVI-met software, a buffer zone with zero elevation surrounding the DSM was added to the city model grid. Flat areas were chosen because the existence of topography would interfere with the effect of building geometry making any derived model highly biased by location. To eliminate any topographic influence, elevation between buildings was zeroed. The DSMs did not include vegetation.

Again, it is important to note that this work will not try to model wind speeds within the selected areas of Göteborg, but the wind speeds simulated by a CFD-model (ENVI-met) using the four constructed city models as input data.

4.2 Wind speed simulation

As the response variable, wind speed patterns for the four city models were simulated in ENVI-met 3.1 (Bruse 1999; Bruse n.d.). ENVI-met is a three-dimensional computer model designed to analyse small-scale interactions between urban design (geometry), and microclimate. The model is able to simulate the interactions between different urban surfaces, vegetation and the atmosphere at typical scales of 0.5 to 10 meters

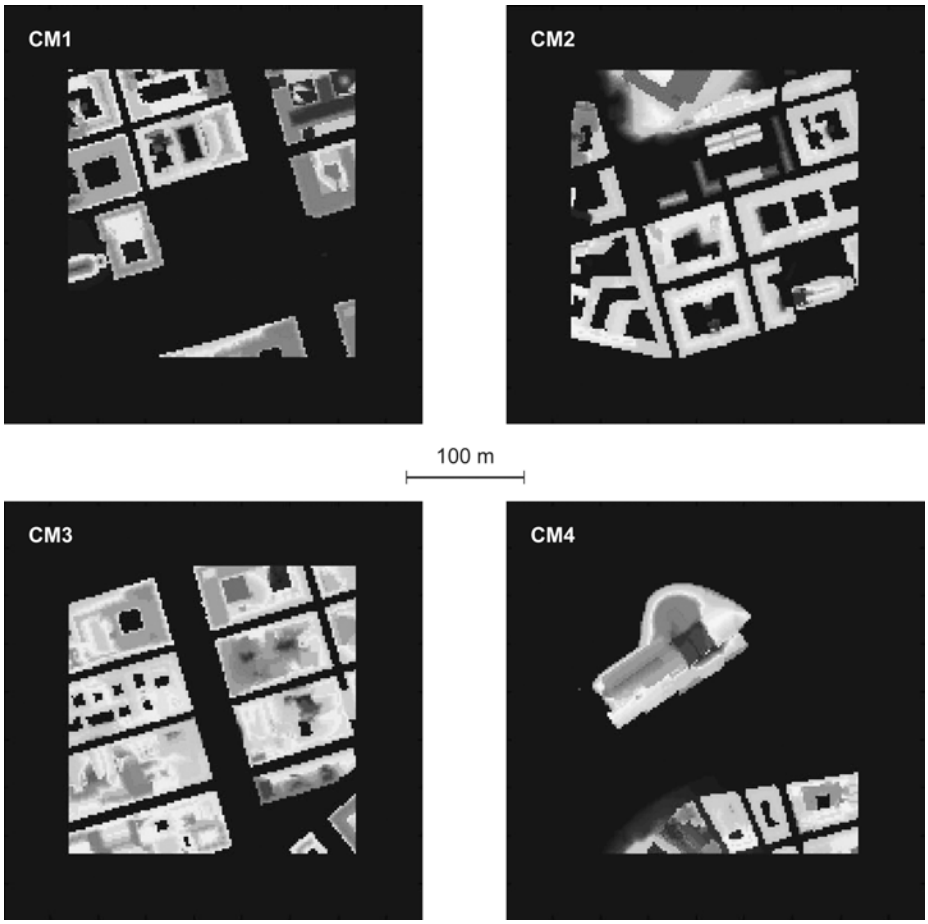

Figure 2 City Modes 1–4 (CM1–CM4).

Table 1 Description of the four city models.

City model	Description
CM1	An open area (square) in the centre and quite dense building structure with narrow canyons running approximately in west to east direction in the north. A wide street canyon runs through the model in approximately south to north direction.
CM2	Dense building structure with a small open area in the northwest. Street canyons are narrow and intersect at right angles.
CM3	A very dense building structure with narrow street canyons, divided by a wide street running from south to north.
CM4	Mainly open areas in between dense buildings in the south and a large isolated building in the north.

and a maximum time step of 10 seconds (Ozkeresteci et al. 2003). The software uses input values for buildings, vegetation, ground surfaces, climatic conditions, soils (Rosheidat et al. 2008), and then simulates the impact on various climatic parameters,

such as wind speed and direction, air temperature and humidity, turbulence, radiative fluxes, bioclimatology and gas and particle dispersion (Ozkeresteci et al. 2003).

ENVI-met allows for a maximum grid of 250×250×30 cells and takes a number of configuration parameters to set up the simulation environment. For this study the size of each model was 175×175 pixels, which include a zone of zero elevation around the actual DSM. This zone is needed for the software to establish a vertical logarithmic wind profile representing external undisturbed wind conditions. The roughness length parameter was set to 0.1 meters and input wind speed (undisturbed wind 10 meters above ground) to $6\,\mathrm{m\,s^{-1}}$. To study the effect of different wind directions with respect to the orientation of the geometry, three different wind directions, S (180°), SW (225°), and W (270°), were simulated for each of the four city models giving a total of 12 simulations. Simulation time was one hour, between six and seven in the morning, thus the result is mean wind speed ($\bar{u}$) during this time. It was assumed that an input wind speed of $6\,\mathrm{m\,s^{-1}}$ in the morning makes any thermal effects negligible.

In practice the combination of one city model and one wind direction may be regarded as one unique model, giving a total of 12 different datasets. Hence, hereafter City Model refers to a combination of one DSM (1, 2, 3 or 4) and one wind direction (S, SW or W), e.g. denoted $CM1_S$. From the three-dimensional simulation result, 2-dimensional (horizontal) grids were extracted at 2 meters above ground level. Finally, the wind speed grids were divided by input wind speed. Thus, the actual data used in the following regression analysis is the fraction of input wind speed, often called the *amplification factor* (Blocken & Carmeliet 2004), in this case:

$$\gamma_{x,y} = \frac{\bar{u}_{x,y}}{U_{ref}}$$

where $\gamma_{x,y}$ is the amplification factor in grid cell *x,y*, $\bar{u}_{x,y}$ is the simulated mean wind speed in grid cell *x,y* and U_{ref} is the external reference wind speed, here given by the ENVI-met configuration (input wind speed).

4.3 Computation of DSM derivatives

As stated above, the fundamental assumption of this study is that the geometry of the urban environment is the major factor influencing wind speed within the urban canopy and at street or pedestrian level (two meters above ground). Building height, size, shape and the spatial distribution of buildings constitute the geometry. Therefore, a DSM is a useful representation of the urban geometry. However, DSMs do not represent the geometry at locations in between buildings, and since these are the locations of interest in this study (obviously there will be no street level winds at

the same locations as any building), such measures must be derived. This motivates the computation of SVF, Fetch, FAI and αFAI, since these derivatives are measures of the geometry at locations in between buildings.

To compute the derivatives, a shadow-casting algorithm operating on raster data (Richens 1997; Ratti 2004; Lindberg & Grimmond 2010) was utilized in MATLAB. In the following sections the procedures are explained in more detail.

4.3.1 Sky View Factor

SVF is a measure of the degree to which the sky is obscured for a given point (Grimmond & Oke 1999). It is a dimensionless measure between 0 and 1; a value of zero represents total obstruction and a value of one free space.

Figure 3 illustrates the influence on SVF of the geometry surrounding a given point. It is evident from this illustration that both the distance to, and the height of the buildings affects the value of SVF. In Figure 3 the building B_1 obscures a larger part of the sky for point p_1 to which it is close than for point p_2, and it has no effect on the point p_3, since building B_3 at this location obscures it. Further, due to its height, B_1 obscures more of the sky than B_2 at location p_1. The segments indicated by the dotted lines represents the part of the sky that is not obscured for the three points. Compared to the corresponding segments for p_1 and p_3, $SVF_{p3} < SVF_{p1} < SVF_{p2}$.

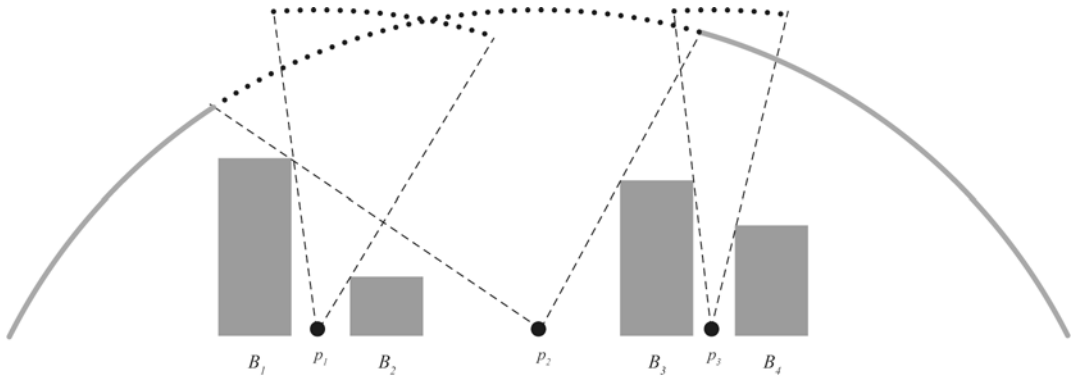

Figure 3 Influence of geometry on Sky View Factor (SVF) illustrated by a vertical cross-section. The SVF for point p_2 is indicated by the dotted line and is ≈ 1/3 of the "sky". The effect of building height and distance to building is clearly seen at point p_1 and p_3, which have lower SVF values. Note that in reality the visible proportion of the whole hemisphere defines SVF.

Using the shadow-casting algorithm, Ratti and Richens (1999) developed a method for calculation SVF from high resolution DSMs which have been evaluated and proved to be very accurate (Lindberg & Grimmond 2010). To calculate SVF, the shadow-casting algorithm is repeated for a large number of solar positions, distributed over the sky according to the concept of annulus weighting proposed by Steyn (1980). A predefined number of solar positions, distributed over a number of annulus levels,

i.e. levels of the same solar elevation angle, are used. For each level, azimuths are offset to give an optimal distribution of all solar positions on the hemisphere. The algorithm then computes the sum of weighted shadow maps obtained from each of the annulus levels (Lindberg & Grimmond 2010).

The algorithm computes five different measures of SVF: (1) the global SVF taking all directions into account and (2–5) four direction specific SVF, one for each of the cardinal axis north (SVF_N), east (SVF_E), south (SVF_S) and west (SVF_W) using a 180° field of view. Since the interest here is in the geometry's orientation with respect to wind direction, the four directional SVFs were recomputed accordingly. This can be done by computing the weighted sum of two of the four directional SVFs using local raster operations. For each model, SVF for the directions *upwind*, *downwind*, and the two directions perpendicular to the wind direction *left* and *right* were computed as illustrated in Figure 4. For example, if the wind direction is from SW (225°) the upwind SVF, SVF_{uw}, is the weighted sum of SVF_W and SVF_S, computed as follows:

$$SVF_{uw} = SVF_W \times w_W + SVF_S \times w_S$$

$$w_W = \frac{225 - 180}{270 - 180}$$

$$w_S = 1 - w_W$$

where w_W and w_S are the weights for west and south respectively. Note that the upwind direction is one of the wind directions used in the simulation of wind speed, i.e. 180°, 225° or 270° (section 4.2).

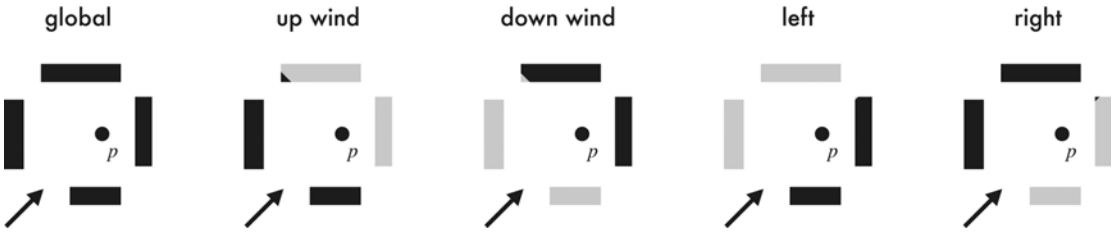

Figure 4 Five different measures of SVF. The rectangles represent buildings, and black indicates the part of the buildings surrounding point *p* that affects respective value of SVF at that point. The arrows indicate wind direction.

4.3.2 Fetch

Fetch is the distance, measured in the upwind direction, from a given point to a *leading edge*, that is, a change to a new and climatically different surface. Downwind from the leading-edge, the air must adjust to a new set of boundary conditions (Oke 1988). In this study, Fetch is conceptualised as the distance from a point *p* to the

nearest obstacle, e.g. a building, in the upwind direction. Compared to SVF, Fetch is not affected by building height.

For each cell, Fetch is computed for a set of directions specified by a *search angle, α*, which forms a circle sector with radius *r*, defined by the *search distance*, as illustrated in Figure 5. If no building is located within the search distance, the value of Fetch is equal to the search distance. The resulting Fetch-value for point *p* in Figure 5 is the average value of the distances d_1 to d_n represented by solid lines (between point *p* and the *x*-marks) for l_1 to l_n, where $n = \alpha$, normalized by the search distance, d_s:

$$Fetch_p = \frac{1}{d_s}\frac{\sum_{i=1}^{n} d_i}{n}, \quad n = \alpha, \quad 0 < Fetch_p < 1$$

Thus, Fetch is a value between 0 and 1. In Figure 5, two measurements (l_4 and l_5) have distances equal to the search distance, because no building is in "the line of sight".

The search distance and search angle are parameter settings in the algorithm and can in theory have any values. For example, a search angle of 360° would give a Fetch value representing the average distance to all buildings surrounding point *p*. The possible values in this study are restricted to values that keep the search sector within the boundaries of the built up part (the DSMs) of the city models. Figure 6 shows the maximum possible values, given that only a subset of the model is later sampled (section 4.3.4).

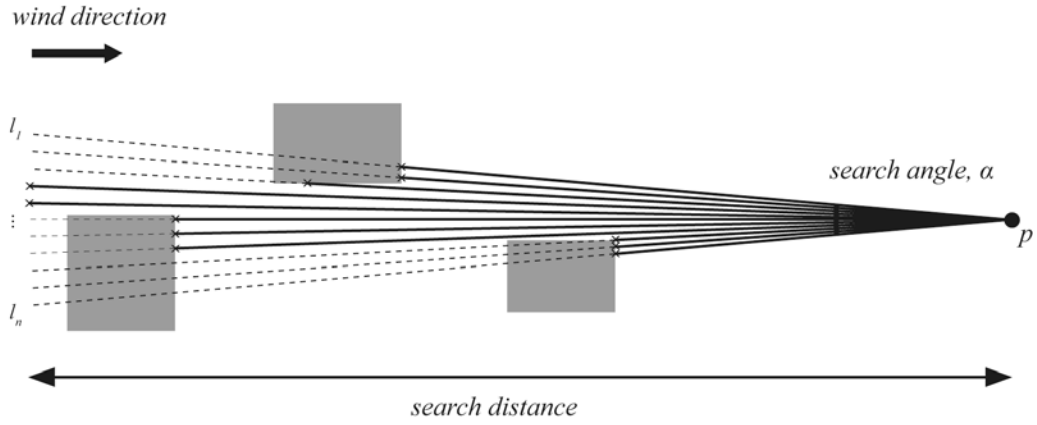

Figure 5 Given a search angel, α, and a search distance d_s the Fetch value for point *p* is the average distance to any building up wind from point *p*. In the figure, α = 11 and the distances are represented by solid lines ended by an *x*-mark.

The maximum possible values of search angle and search distance were found to be 61° and 144 meters respectively. To test the influence of search angle and search distance on the correlation to wind speed, all combinations of four different settings for each parameter were used:

Search angle: 11°, 21°, 41°, 61°

Search distance: 10 m, 50 m, 100 m, 144 m

Thus, a total of 16 different Fetch derivatives were computed. Table 2 summarises all derivatives.

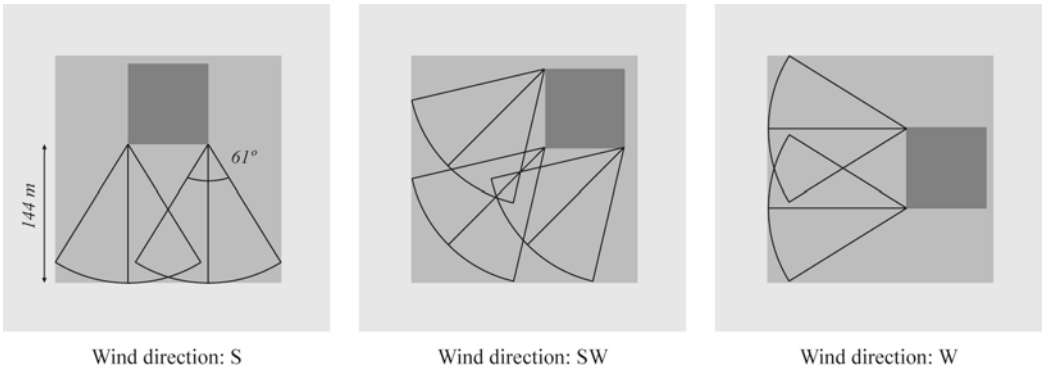

Figure 6 Illustration showing the maximum possible search angle and search distance for calculation of Fetch, FAI and αFAI. The medium grey area is the DSM part of the city model and the dark grey area is the areal subset used for sampling. To maximize search distance and search angle, the sampling area is different for the three different wind directions.

4.3.3 Frontal area index and angular frontal area index

FAI is a measure of building walls facing the wind flow in a particular direction and expressed as frontal area per unit horizontal area Figure 7a. It has a strong relationship with surface roughness and has been suggested as a good indicator for roughness in meso-scale meteorological models (Wong et al. 2010). However, in this study, FAI is conceptualized not as frontal area per unit horizontal area, but building height *h* facing the wind per unit distance *d*, as illustrated in Figure 7b. The value of FAI for point *p* is the sum of the heights h_1 and h_2 divided by the distance *d*. Compared to SVF and Fetch, FAI is a measure that is less "local" in that it accounts for all roughness elements (buildings) within a given distance.

In this study, a novel measure of urban geometry is introduced. Similar to FAI, αFAI is a measure of the angle between the ground surface and a vector pointing towards the top of the wall facing the wind per unit distance. The angle is expressed as the ratio between the building height and the distance from a given point to where the height is measured. For the example in Figure 7b, αFAI for point *p* is equal to $h_1 / d_1 + h_2 / d_2$ divided by the distance *d*.

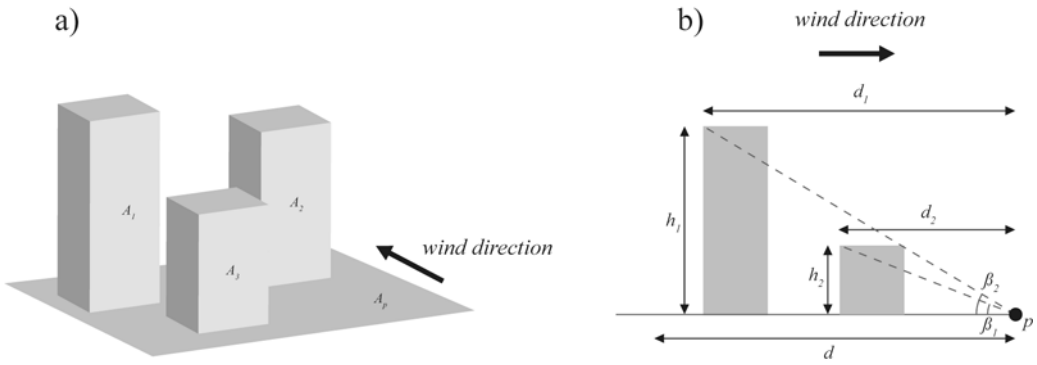

Figure 7 a) FAI is the sum of building walls facing the wind per unit horizontal area. b) Illustration of FAI and αFAI as conceptualized in this study.

The difference between these two measures is that in the case of αFAI, the positions of the buildings in relation to point *p* have influence on the value. Shifting the positions of the two buildings in Figure 7b will increase the value of αFAI (β_2 will be very large) but the value of FAI will be the same.

Computation of FAI and αFAI follows the same conceptual method as for the computation of Fetch described above. The computed values are average values of several measurements, given a search angle and a search distance, as illustrated in Figure 8.

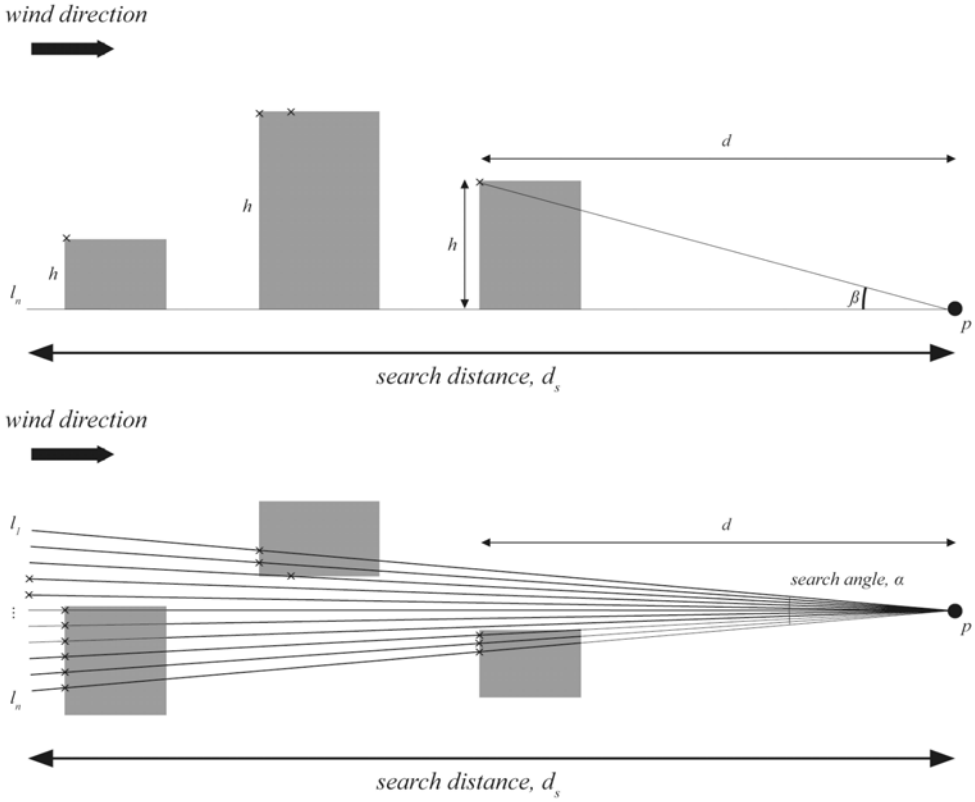

Figure 8 Given a search angel, α, and a search distance d_s FAI and αFAI for point *p* is the average value of FAI and αFAI along each line l_1 to l_n.

For simplicity, the same search distances and angles as for Fetch were used. Consider point *p* in Figure 8. The value of FAI is the average value of the building height *h* in each point marked by *x* along each line $l_1$to l_n, divided by the search distance, d_s :

$$FAI_p = \frac{1}{n}\sum_{i=1}^{n}\frac{\sum h}{d_s}, \quad n = \alpha$$

αFAI is computed in a similar way, but instead of the height, the ratio between the height *h* and the distance *d* between *p* and the points marked *x* along the lines $l_1$to l_n is used:

$$\alpha FAI_p = \frac{1}{n}\sum_{i=1}^{n}\frac{\sum \frac{h}{d}}{d_s}, \quad n = \alpha$$

4.3.4 Sampling

In total, 53 different DSM derivatives (Table 2) were computed for all four city models and three different wind directions, giving 648 raster datasets, including wind speed. This means that for each city model and wind direction, each location (pixel) represents an observation of wind speed plus 53 derivatives. To obtain a dataset to use in the statistical analysis, a sample of observations from each model was extracted and compiled into one data set. The sample was taken from a subset area in each model chosen so to avoid border effects from the simulation of wind speed and to allow for as long search distance as possible when computing Fetch, FAI and αFAI. Thus the location of the sample area was different for the three wind directions, previously illustrated in Figure 6. The sample areas were 43×43 pixels. Pixels that coincided with buildings were excluded from the sample that in total included little more than 2800 observations. Because values of variables at one location tend to be strongly associated with values at nearby locations (i.e. spatial autocorrelation), a random sampling design can provide redundant information when sample locations are close to one another and may not be preferred (Rogerson 2010). Instead, a cyclic sampling design optimises the distances between sampling points, therefore the number of samples needed to represent the spatial characteristics of the data is also optimized (Burrows et al. 2002). Here, a 3/7 cyclic sampling design was used. Applied on raster data this design takes samples at position (pixel) 0, 1 and 3 within a seven pixel long sequence. This pattern is repeated along both columns (y) and rows (x) for the whole raster, or as in this case, for the whole sample area described above.

Table 2 Summary of all computed derivatives. For Fetch, FAI and αFAI search angle and search distance is indicated by the first and second subscripted digit respectively, i.e. Fetch-three-one ($Fetch_{31}$) is Fetch computed using a search angle of 41° and a search distance of 10 meters.

Sky view factor				
SVF_g	SVF_{uw}	SVF_{dw}	SVF_{lw}	SVF_{rw}
	Fetch			
	10 m (1)	50 m (2)	100 m (3)	144 m (4)
11° (1)	$Fetch_{11}$	$Fetch_{12}$	$Fetch_{13}$	$Fetch_{14}$
21° (2)	$Fetch_{21}$	$Fetch_{22}$	$Fetch_{23}$	$Fetch_{24}$
41° (3)	$Fetch_{31}$	$Fetch_{32}$	$Fetch_{33}$	$Fetch_{34}$
61° (4)	$Fetch_{41}$	$Fetch_{42}$	$Fetch_{43}$	$Fetch_{44}$
Frontal area index				
	10 m (1)	50 m (2)	100 m (3)	144 m (4)
11° (1)	FAI_{11}	FAI_{12}	FAI_{13}	FAI_{14}
21° (2)	FAI_{21}	FAI_{22}	FAI_{23}	FAI_{24}
41° (3)	FAI_{31}	FAI_{32}	FAI_{33}	FAI_{34}
61° (4)	FAI_{41}	FAI_{42}	FAI_{43}	FAI_{44}
Angular frontal area index				
	10 m (1)	50 m (2)	100 m (3)	144 m (4)
11° (1)	αFAI_{11}	αFAI_{12}	αFAI_{13}	αFAI_{14}
21° (2)	αFAI_{21}	αFAI_{22}	αFAI_{23}	αFAI_{24}
41° (3)	αFAI_{31}	αFAI_{32}	αFAI_{33}	αFAI_{34}
61° (4)	αFAI_{41}	αFAI_{42}	αFAI_{43}	αFAI_{44}

4.4 Statistical methods

To assess the adequacy of the use of regression analysis to a given problem or dataset, possible violations of the regression assumptions should be diagnosed. Such diagnostic methods are primarily based on the study of the error terms, that is, the regression residuals (Montgomery et al. 2001). Since the regression residuals are part of the diagnostic results from a regression analysis, the methodological approach has to be iterative, where the initially specified model is edited and regressed again until the model is satisfactory.

Here, the regressors are 53 different raster dataset representing derivatives of urban geometry and the main goal is to identify those regressors, or the combination of regressors, that best explain the variability in wind speed. It is important to recognise the nature of the data set. As described above, several varieties of the same derivative (e.g. 16 different Fetch) are included which is likely to result in many redundant variables and high collinearity.

To achieve this, the following steps were performed:

1. Fitting the full model using MLR, i.e. the model containing all 53 regressors.
2. Examination of regression residuals and other diagnostics.
3. Transformation of response and regressors if appropriate. Check for improvement.
4. Selection of a set of candidate regressors using a sequential selection approach
5. All possible subset regression on the set of selected regressors.
6. Examine the results and select one (or more) model.
7. Predict wind speed using the suggested models and evaluate model performance.

4.4.1 Sequential variable selection algorithm

A first screening of the relatively important regressors was performed using a sequential selection algorithm implemented in MATLAB (step 4 above), in which regressors were added to the model in a sequential order, minimising the residual sum of squares *RSS* in each step. To avoid adding more regressors than necessary to improve the model, the process continued until a model selection criterion was satisfied (minimized AIC or BIC). Here, BIC were used in favour of AIC because it penalizes many regressors harder. The algorithm was applied as follows:

1. The dataset was partitioned into *L*-fold randomised cross-validation subsets using the `cvpartition`-function in MATLAB ($L = 100$).
2. For each subset *l*, a sequential selection scheme using MLR under BIC was applied. For each step *i*, the regressor minimising *RSS* was selected as the *i*th regressor (x_i) in the *l*th model and a binary value V_{li} indicating the appearance of the regressor in the model was stored. The score indicating the order of selection (i.e. the value of *i*) and *RSS*, R^2 and *BIC* was also stored. Regressors were added to the model as long as the value of *BIC* improves, i.e. decreases. This procedure results in 100 separate set of selected predictors, or models.
3. The frequency of appearance $F_i = \sum_{l=1}^{L} V_{li}$ and average score were computed for each regressor. A high value of F_i indicates that a regressor was included in many of the models and a low score indicates that it was selected early in the process (a low *i*). Based on the results, a subset of candidate regressors was chosen for further analysis.

Due to the fact that many regressors are varieties of the same derivative, it is reasonable to suspect a high degree of multicollinearity in the dataset. To account for this, the possibility of using Partial Least Square Regression (PLSR) instead of MLR

was tried. PLSR is a regression technique that is useful when there are many regressors compared to the number of cases in the data (high dimensional datasets), or when there is a high degree of multicollinearity. However, *RSS*, and thus *BIC*, computed on the full model by PLSR (`plsregress` in MATLAB) did not differ from values computed by MLR, which is why MLR was chosen as the regression method in the sequential selection algorithm. In MATLAB the function `regstats` were used in step two above.

4.4.2 All-possible subset regression

Based on the results from the sequential variable selection algorithm described above, a selected set of candidate regressors was subjected to the `regsubsets` function in the R-Package Leaps. `regsubsets` is an all-possible regression algorithm, fitting models to all possible combinations of regressors given as input. The output from the algorithm is a set of models including an intercept and *p* regressors, along with diagnostic statistics.

4.4.3 Evaluation of models

Four models were selected based on the results from the all-possible subset regression models. Predicted wind speed, or amplification factor, was computed for all city models. RMSE was calculated for the predictions within the sampling areas. For visual interpretation and evaluation of the model performances, predicted values and residuals were mapped.

5 Results

In this chapter the most important results are presented and illustrated by maps and diagrams, and summarized by tables. Because of the many city models, wind directions and derivatives, the possible number of maps and diagrams are very large. Here only some of them are presented as examples.

The chapter starts with a brief look at the simulated wind speed patterns and their relationships to the derivatives, section 5.1. After the specification and fitting of the full model, the influence of search angle and search distance are examined in section 5.2. The results of the variable and model selection are presented in Section 5.3. Finally, in section 5.4, the results from predicting wind speed using the selected models are presented.

5.1 Wind speed and derivatives

5.1.1 Wind speed simulation

The results of the wind speed simulation by ENVI-met are shown in Figure 9. To the left, wind is from the south (S) in the middle from southwest (SW) and to the right from the west (W). Wind speed is expressed as amplification factor (γ), i.e. the ratio between observed and external undisturbed wind speed. The red rectangles are the areas where samples were taken.

Looking at the top row (CM1) in Figure 9, the highest wind speeds ($\gamma > 0.6$) are found in the centre of the wide street (1) running almost parallel to the external wind direction in $CM1_S$ and in corner streams (2). Moderately high wind speeds ($0.4 < \gamma < 0.6$) are found in the open square (3). Here, a distinct difference in $CM1_S$ is apparent: the two corner streams (2) means that there are lower wind speeds in the middle of the square than at the flanks. Generally the opposite pattern is found in open areas. Also, in $CM1_{SW}$, wind speeds are higher in the down wind end of the square (4) leading into the wide street canyon.

Low wind speeds are found in sheltered areas (5), at the windward side of buildings (6), and in the narrow street canyons. There is no sign of the vortex at the windward side of buildings produced by the air deflected down along the windward wall.

The effect of angle-of-attack is visible in the wide street (1) where wind speed decrease as the wind shifts toward west and becomes more oblique to the street direction. In the open area, the wind pattern, i.e. the distribution of low wind speed close to the buildings, changes accordingly. This effect is not apparent in the narrow

canyons. Also the effect of surface drag by the canyon wall is visible, particularly in the wide street in the S and SW.

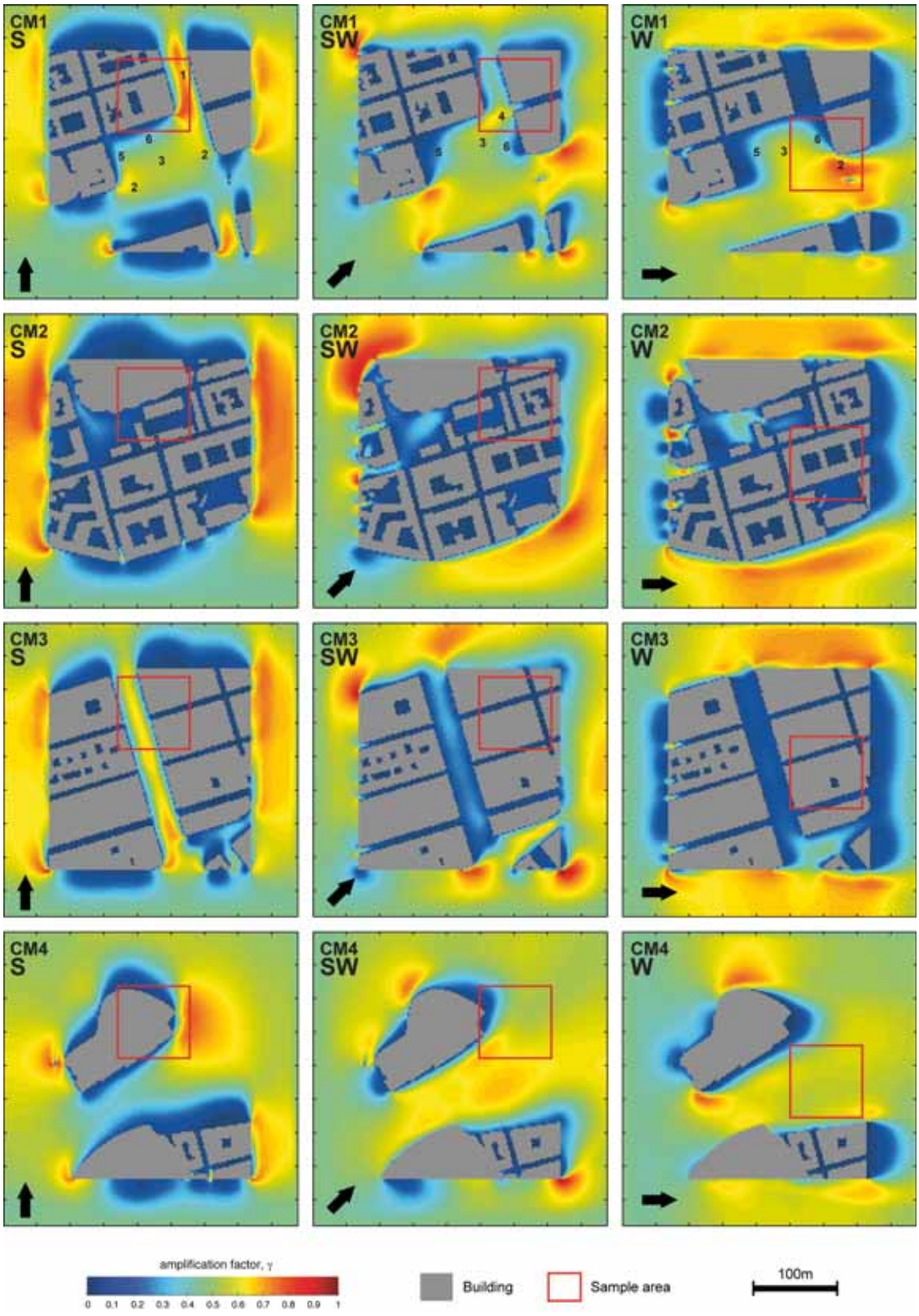

Figure 9 Simulated wind speed for all city models and all wind directions expressed as amplification factor, (γ).

5.1.2 DSM-derivatives

Figure 10 shows examples of derivatives for $CM1_{SW}$. Both SVF and Fetch show low values close to buildings and high values in open areas. As expected, the opposite is true for FAI and αFAI. The effects of search distance and search angle on the spatial patterns are clearly visible for Fetch, FAI and αFAI, which suggest that these settings are important. The main qualitative difference between all derivatives and wind speed is the lack of recognition of contextual wind pattern such as corner streams and channelling effects.

Sky View Factor

The upper row in Figure 11 shows scatter diagrams of wind speed against the three of the SVF regressors, not accounting for the effects of any other regressors. The diagrams show positive non-linear relationships between these variables. By transforming both wind speed and each of the SVFs by the natural log, i.e. $\ln(y) \propto \ln(x)$, gives the scatter diagrams in the bottom row in Figure 11. These diagrams show approximately linear relationships, but with large spread of observations at low values of amplification factor.

Fetch

The scatter diagrams in Figure 12 show very complex relationships between Fetch and wind speed. However, the same transformation (natural log) was found to be the best linear fit given this data set (bottom row in the figure). The spread of observations are highest at low values of both wind speed and Fetch. Also, at shorter search angles (exemplified with $Fetch_{22}$ in the figure), the correlation seems to be weaker.

Frontal area index

The scatterplots for FAI in Figure 13 indicate weak negative relationships with wind speed. In general lower wind speeds are found at higher FAI-values, that is, in more dense geometry, which is to be expected. As in the case of SVF and Fetch, the variance of the predictor is high at low wind speeds. There is no clear indication of the effect of search distance and search angle. Transformation of FAI, given that amplification factor was transformed by the natural log, did not increase the linearity.

Angular frontal area index

As illustrated in

Figure 14 αFAI shows an exceptionally negative exponential-like relationship to wind speed, which becomes approximately linear after the transformation of wind speed. At low wind speeds, the variability in αFAI is very high, and the opposite is true for low values of αFAI. Low values of αFAI are expected in open areas, not close to upwind

buildings and this may explain the variability in wind speed at low αFAI-values; both pixels in the open as well as close to the windward wall (down wind side of the open area) are included.

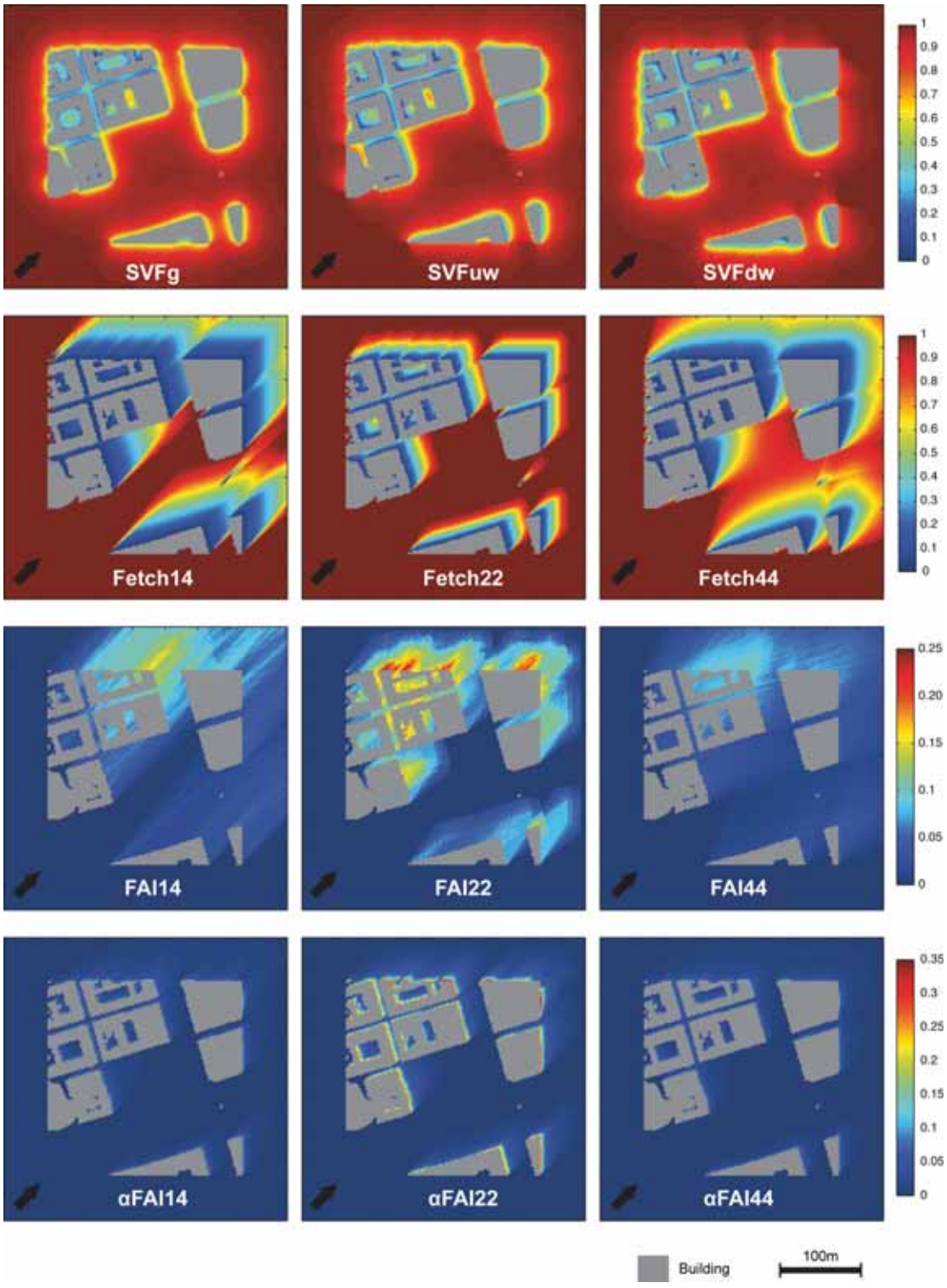

Figure 10 Examples of maps showing the four measures of urban geometry, $CM1_{SW}$.

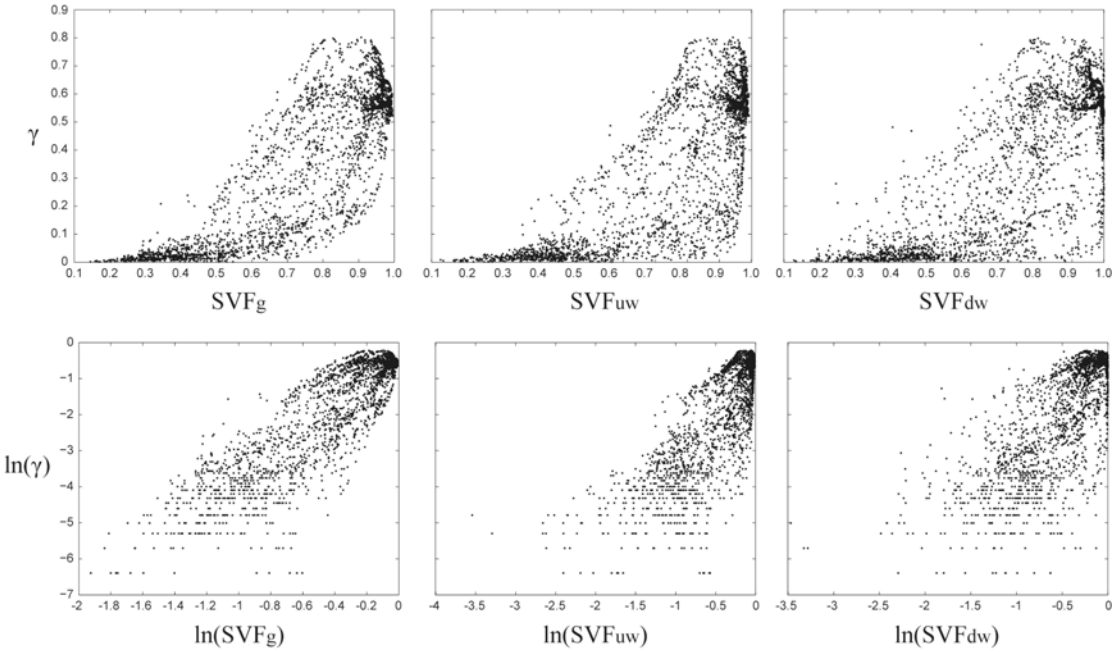

Figure 11 Scatterplots showing the relationship between amplification factor and three different SVF-derivatives (top row). Transformation of both amplification factor and SVF increases the linearity of the relationships (bottom row).

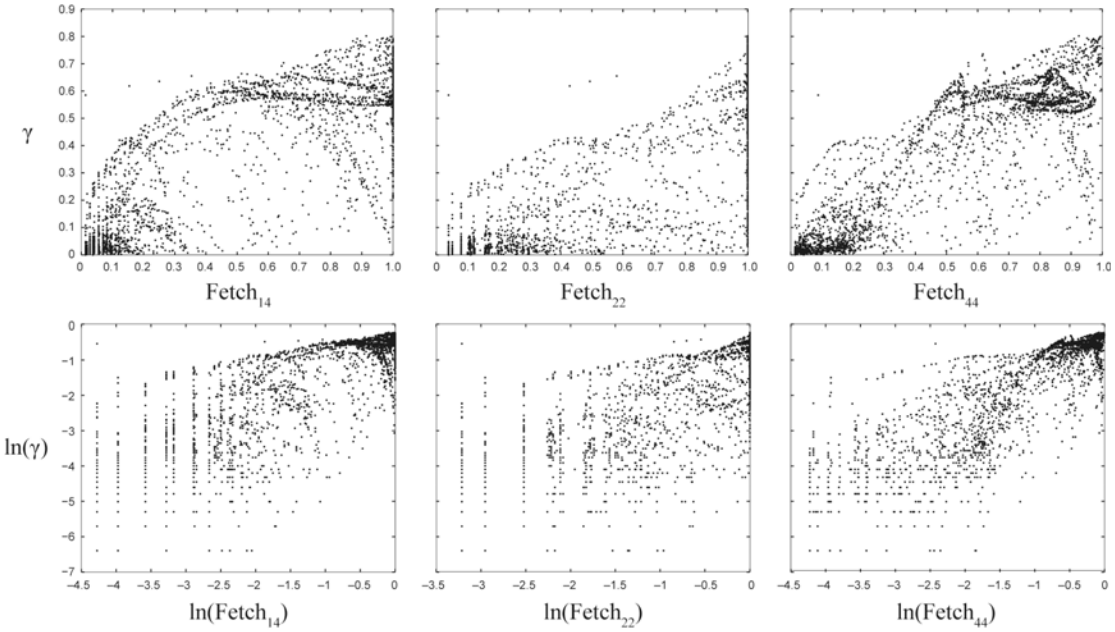

Figure 12 Scatterplots showing the relationship between amplification factor and three different Fetch-derivatives (top row) and the transformed variables (bottom row).

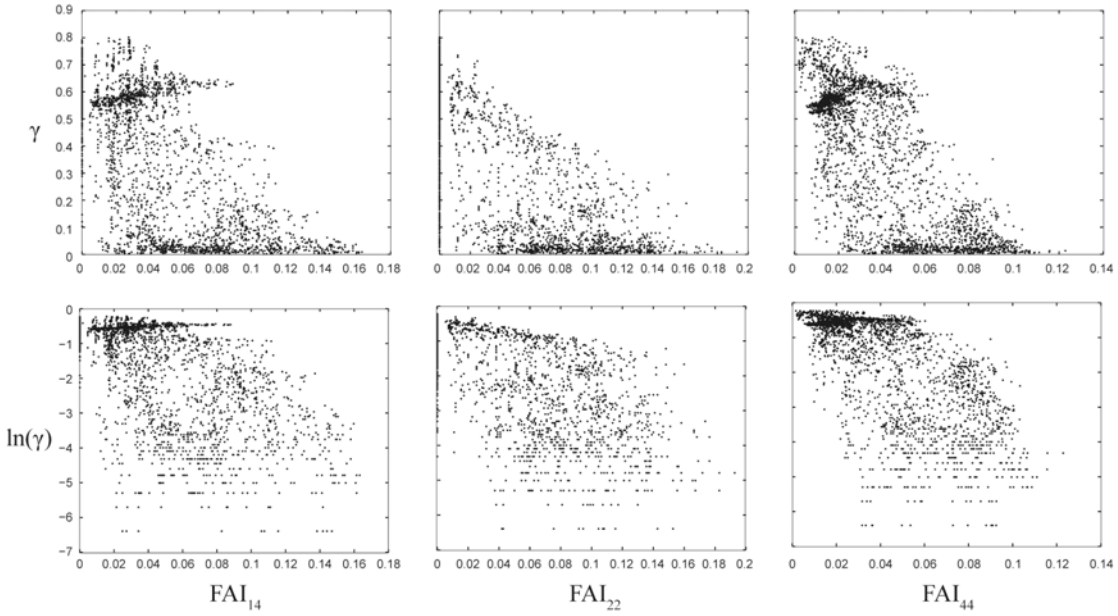

Figure 13 Scatterplots of amplification factor (γ) and FAI show a negative correlation (top row), which approximates a linear behaviour when γ is transformed (bottom row).

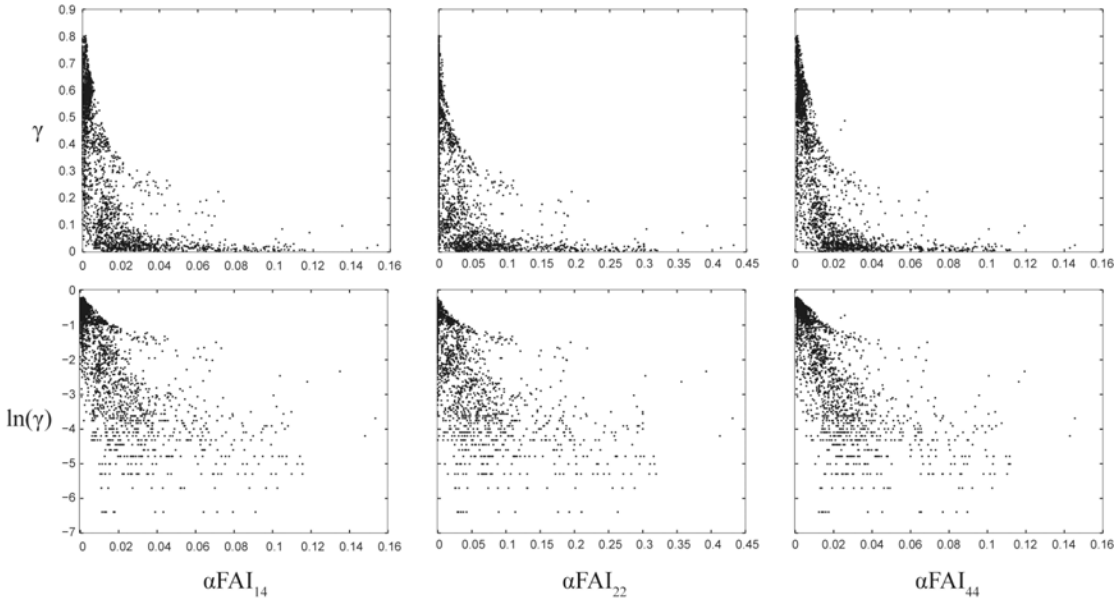

Figure 14 Scatterplots of amplification factor (γ) and αFAI show a negative exponential like correlation (top row), which approximates a linear behaviour when γ is transformed (bottom row). At low amplification factor values, the point cloud is very dispersed.

5.1.3 Fitting the full model

As a first investigation, scatterplots of wind speed against the regressors revealed possible nonlinear relationships. This was confirmed by fitting the full model. Figure 15 shows the residual diagnostics from the full model fit. The residuals are

approximately normal distributed (Figure 15a) but they indicate both non-linearity and non-constant variance (Figure 15b–c).

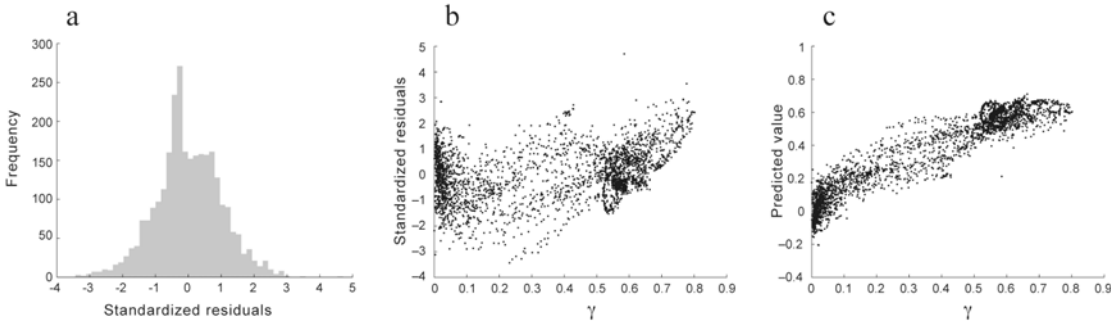

Figure 15 Regression diagnostics for fitting the full 53-regressor model: a) histogram of standardised residuals, b) standardised residuals against observed response and c) standardised predicted values against observed response.

Based on these findings, the response, the SVF and Fetch regressors were transformed by natural log, and the full model was fitted again with the transformed variables replacing the original. The results are shown in Figure 16.

The residuals are still reasonably normally distributed (Figure 16a) but the residual plots (Figure 16b–c) have improved considerably. The curved shape of the point clouds has almost disappeared indicating that the model specification is roughly correct with respect to the linearity assumption. Still, there is evidence of non-constant variance, mainly at low wind speed values, approximately at $\ln \gamma < -4$, where the spread of data points increase with a trend towards negative residuals (Figure 16c). Examination of multicollinearity showed extremely high VIF-values for the transformed set of regressors, with a mean of 10,971, reaching maximum at 223,169 for αFAI_{43}. SVF_{dw} have the lowest VIF-value, 24.2. In general, Fetch and αFAI were the most inter-correlated derivatives.

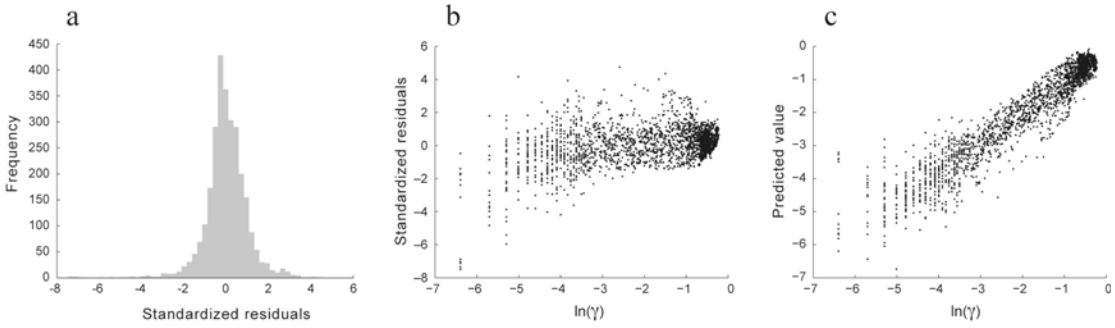

Figure 16 Regression diagnostics for fitting the full transformed 53-regressor model: a) Histogram of standardised residuals, b) standardised residuals against observed response and c) standardised predicted values against observed response.

5.2 Influence of search angle and search distance

One of the objectives of the present study is to investigate the influence of search angle and search distance on the correlation between wind speed and Fetch, FAI, and αFAI. As described in 4.3.2, four different search angles (α_s) and four different search distances (d_s) were used to calculate 16 different derivatives of Fetch, FAI and αFAI respectively. Figure 17 plots Pearson's correlation coefficient *r* for the linear correlation between wind speed and all derivatives of Fetch, FAI and αFAI as a function of α_s and d_s . Because linearity is assumed, the transformed values of wind speed and Fetch were used in the calculations. Pearson's coefficient is a bivariate measure of correlation assuming that the correlation is affected by no other variable than the dependent and one independent variable. Hence, it may not be appropriate to draw any conclusions on the importance of individual variables as part of a more complex model from these data. However, the correlation coefficient should provide a useful comparison between variables.

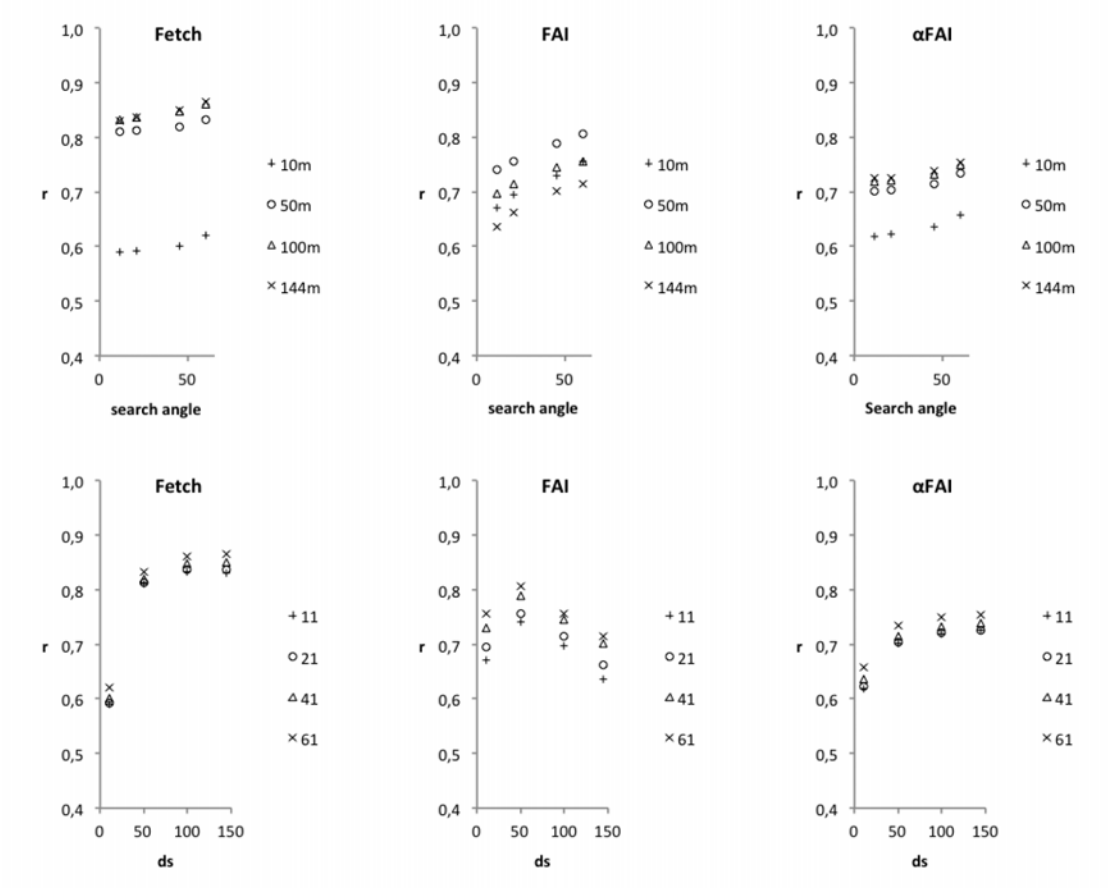

Figure 17 Pearson's correlation coefficients for wind speed (*u*) and Fetch (to the left), FAI (in the middle) and αFAI (to the right) against search distance (top row) and search angle (bottom row).

Generally, the plots in Figure 17 indicate high correlations and most values of *r* are above 0.6, peaking at ≈ 0.85. The effect of search angle is consistent over all

derivatives; increasing α_s has positive effect on *r*, which is most apparent in the case of FAI. The situation is somewhat more complex when it comes to search distance. For both Fetch and αFAI increasing search distance has a clear positive influence on the correlation coefficient and towards the maximum search distance, the curve flattens out. This is not the case for FAI. Here, a maximum correlation is obtained for all search angles at a search distance of 50 meters.

5.3 Model building

5.3.1 Variable selection

As a first screening, the set of transformed regressors was submitted to a variable selection algorithm utilising MLR in a sequential added variable approach under *BIC*, as previously described in section 4.4.1. In Figure 18 the regressors are ranked (from left to right) based on the frequency of appearance, F_i (bars) of all selected regressors (Fi ≤ 1), leaving out those never selected in any model ($F_i = 0$).

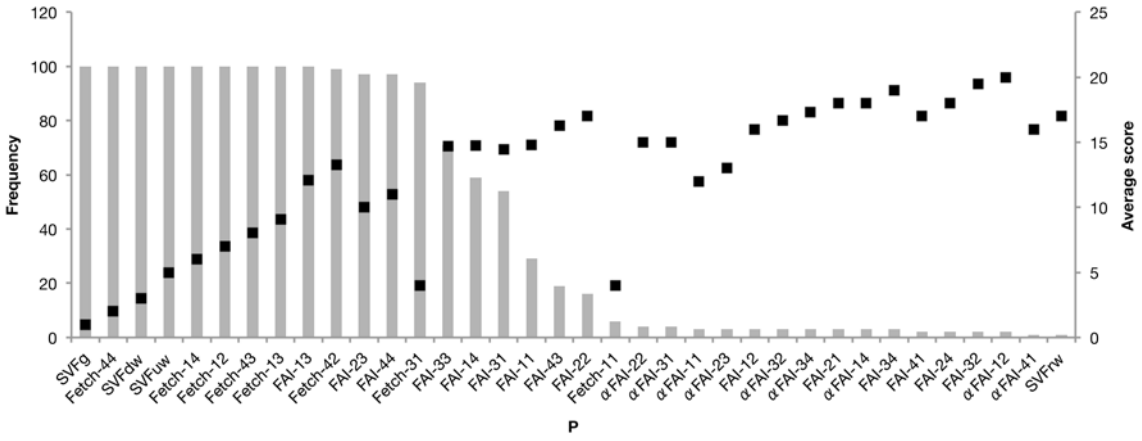

Figure 18 Frequency of appearance in 100 cross-validation subsets (bars) and average score (black dots). All regressors with $F_i \geq 1$ are included in the diagram. The order of selection is from left to right.

Thirteen regressors stand out having $F_i > 90$, of which nine appears in all models ($F_i = 100$). Among these top ranked regressors, Fetch and SVF derivatives dominate, but three FAI-derivatives are also included. αFAI is not favoured by the selection algorithm and is present in only a few models and having high average scores. The diagram also shows the average score (dots). The first regressor to be selected by the algorithm receives the score one, the second two, and so forth. The three best-ranked regressors have average scores of 1.00, 2.00, 3.00, respectively, indicating that they are selected in this order in all 100 models. Looking at the following regressors, the average score starts to fluctuate, meaning that the order of selection for these is not constant over all subsets. The general trend is that a low average score is associated with a high frequency and thus a high rank, but this trend flattens out at lower

frequencies (around rank 14) and there are large deviations. For example, consider $Fetch_{11}$, which has a very low average score (4.00), but a frequency of only six and is therefore ranked in place 20.

Adding more regressors to a model results in a decrease in *RSS* and an increase in R^2, which means that the more regressors the better fit. But with respect to parsimony, such criteria are not very useful and may result in overfitted models and unstable parameter estimations. Since the algorithm was run under *BIC*, a restriction was set on the number of regressors included in the models, this number ranging from 6 to 24. As seen in Figure 19, the decrease in *RSS* flattens out at around $p = 4$ and the same holds for the increase in R^2, why the *BIC* criterion seems to be too relaxed and allows too many regressors to enter the model.

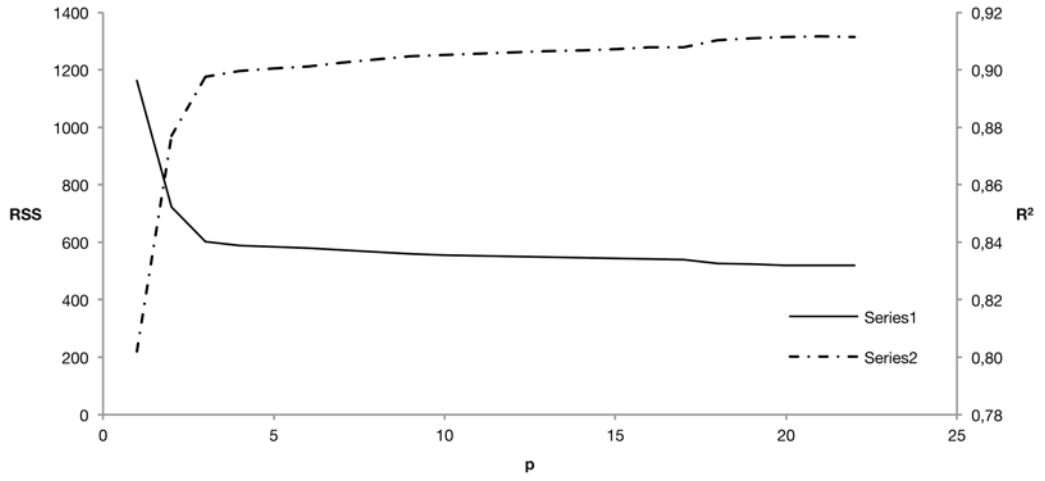

Figure 19 Average RSS and R^2 over all cross-validation subsets for each step (p) in the variable selection procedure.

Considering the diagram in Figure 19, it may be suggested that selecting more than the four best-ranked regressors may result in an overfitted model. However, the algorithm selects the i*th* regressor given that there already are regressors in the model. This means that should the order be different, for example by forcing some regressor into the first slot or by running an all-possible subset regression, the order of selection, and hence the ranking might be different.

Based on the results, the thirteen best-ranked regressors with respect to frequency of appearance ($F_i > 90$) were selected for further analysis in an all-possible subsets algorithm. This excludes all αFAI-derivatives from further analysis.

5.3.2 Model selection

The results of the all-possible subset regression on the selected regressors are presented in Table 3. For each number of regressors, *p*, the *two best of all possible*

models is shown. Not surprisingly, the best one-regressor model include SVF_g, and the second best include $Fetch_{44}$, which also were the two first selected regressors in all cross-validation subsets in the variable selection procedure previously described. With $p = 2$, however, SVF_g is left out and the combination of $Fetch_{44}$ and SVF_{dw} is proposed as the best model. Here $Fetch_{43}$ replaces $Fetch_{44}$ in the second best model.

Table 3 Summary statistics for all possible subsets regression.

Model No.	*p*	SVF_g	$Fetch_{44}$	SVF_{dw}	SVF_{uw}	$Fetch_{14}$	$Fetch_{12}$	$Fetch_{43}$	$Fetch_{13}$	FAI_{13}	$Fetch_{42}$	FAI_{23}	FAI_{44}	$Fetch_{31}$	R^2	$AdjR^2$	*RSS*	MS_{res}*	C_p	C_p**	*BIC*	*Mean VIF*
1	1	*													0,802	0,802	1174	0,418	3108,7	2800,0	–4528	
2	1		*												0,748	0,748	1491	0,531	4704,6	4312,5	–3857	
3	2		*	*											0,895	0,895	623	0,222	338,0	174,1	–6297	1,42
4	2			*				*							0,893	0,892	637	0,227	408,2	240,7	–6235	1,40
5	3	*	*	*											0,897	0,897	608	0,217	261,7	101,9	–6360	6,43
6	3		*	*	*										0,897	0,897	608	0,217	266,0	106,0	–6356	3,07
7	4		*	*	*									*	0,901	0,900	589	0,210	171,3	16,3	–6438	3,59
8	4	*	*	*										*	0,900	0,899	595	0,212	200,4	43,9	–6410	6,27
9	5		*	*	*	*								*	0,901	0,901	586	0,209	158,9	4,6	–6444	7,09
10	5		*	*	*						*			*	0,901	0,901	588	0,210	168,8		–6434	18,57
11	6		*	*	*	*	*							*	0,903	0,902	578	0,206	118,0		–6477	14,73
12	6		*	*	*	*			*					*	0,902	0,902	579	0,207	122,7		–6472	55,00
13	7		*	*	*	*		*	*					*	0,905	0,904	566	0,202	58,4		–6529	251,24
14	7		*	*	*	*	*	*						*	0,904	0,903	572	0,204	89,7		–6498	116,34
15	8		*	*	*	*		*	*		*			*	0,905	0,905	563	0,201	48,1		–6533	284,76
16	8		*	*	*	*		*	*			*		*	0,905	0,905	564	0,201	49,8		–6531	226,79
17	9		*	*	*	*		*	*		*	*		*	0,905	0,905	560	0,200	34,9		–6540	256,18
18	9		*	*	*	*		*	*			*	*	*	0,905	0,905	561	0,201	39,3		–6536	236,22
19	10		*	*	*	*		*	*		*	*	*	*	0,906	0,906	557	0,199	22,3		–6547	
20	10		*	*	*	*		*	*	*	*	*		*	0,906	0,906	558	0,199	24,1		–6545	
21	11		*	*	*	*		*	*	*	*	*	*	*	0,906	0,906	554	0,198	10,3		–6553	
22	11		*	*	*	*	*	*	*	*		*	*	*	0,906	0,906	557	0,199	22,4		–6541	
23	12	*	*	*	*	*		*	*	*	*	*	*	*	0,906	0,906	554	0,198	12,1		–6545	
24	12		*	*	*	*	*	*	*	*	*	*	*	*	0,906	0,906	554	0,198	12,2		–6545	
25	13	*	*	*	*	*	*	*	*	*	*	*	*	*	0,906	0,906	554	0,199	14,0		–6537	

* *n=2805* ** C_p *best subset regression on the regressors included in the top nine models only.*

The first appearance of FAI in any model is in model 16 (FAI_{23}). This is an indication of the importance of selection order, since this is not the FAI-derivative with the highest rank in the variable selection procedure.

Looking at the statistics for each model in Table 3 as illustrated in Figure 20, *RSS* and R^2 indicate a rapid increase in model performance as p increases. For $p = 1$, there is a relatively big difference between both models, but this difference becomes small at higher values of p. Above $p = 3$ there is very little difference.

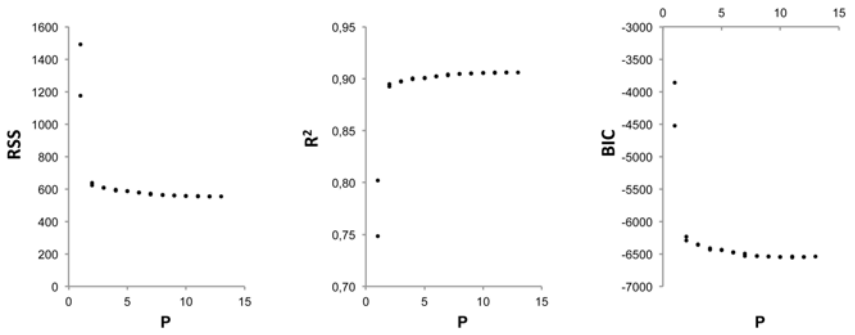

Figure 20 R^2, *RSS* and *BIC* as functions of the number of regressors in the model (based on Table 3).

Given the evidence in the diagrams, it seems appropriate to select a model with three or four regressors since adding more regressors do not improve *RSS* or R^2 very much. However, *RSS* and R^2 are not necessarily good criteria for model selection. Instead the algorithm uses Mallow's C_p to rank models. The diagrams in Figure 20 indicate that there are redundant regressors present in the dataset. Thus, a second set of Mallow's C_p was calculated using only the regressors included in models with a mean VIF-value less than ten as the full model.

A plot of C_p for all proposed models having $C_p < 100$ (Table 3) is shown in Figure 21. Filled dots are C_p-values when all thirteen regressors are in the full model and circles represent the situation with a subset of regressors in the full model. In the first case four models, number 21, 23, 24, and 25 are close to the line $C_p = p$ and should therefore be considered as candidate models. In the second case, model nine is a first choice having a C_p value less than p and only five regressors.

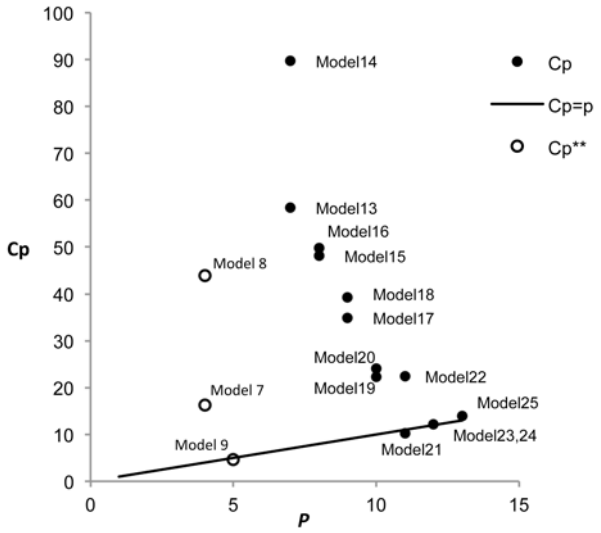

Figure 21 Mallow's C_p for models having $C_p < 100$ as a function of the number of regressors in the model. Filled dots represent the situation where all selected regressors are included in the full model and circles when only the highest ranked regressors are in the full model.

The presence of redundant regressors may be detected by looking for multicollinearity in the data. In Table 3 the mean VIF values for some models are listed and in Table 4 the details behind those figures i.e. the contribution by each regressor, are revealed. Models with $p = 1$ are not included because VIF is not applicable on one-regressor models. It is clear from this table that SVF_g is a major contributor to the total VIF in models having few regressors due to correlation to both the other SVF-regressors but also the Fetch-regressors. By examining the table it becomes evident that in general, adding two regressors of the same type e.g. two Fetches, increases the VIF values. For example, consider model 10 where both $Fetch_{44}$ and $Fetch_{42}$ are included and have

very high VIFs. However, VIF for $Fetch_{31}$ are still reasonably low as are the SVF-regressors in the model. Being the only derivative that accounts for the geometry in the downwind direction, SVF_{dw} adds less redundant information and has low VIF values through all models, except when SVF_g is also in the model.

Because of inter-correlation between regressors, the more regressors that are in a model, the higher the mean VIF, especially in this study where many regressors are very alike, e.g. different Fetches. The effect on FAI appears to be small and SVF_{dw} is less correlated to other regressors than SVF_{uw}.

Table 4 VIF-values for each regressor included in some the models suggested by the all-possible subset regression.

Model	SVF_g	$Fetch_{44}$	SVF_{dw}	SVF_{uw}	$Fetch_{14}$	$Fetch_{12}$	$Fetch_{43}$	$Fetch_{13}$	FAI_{13}	$Fetch_{42}$	FAI_{23}	FAI_{44}	$Fetch_{31}$	Mean VIF
3		1,42	1,42											1,42
4			1,40				1,40							1,40
5	10,25	3,18	5,87											6,43
6		3,52	1,63	4,05										3,07
7		4,12	1,89	5,13									3,22	3,59
8	11,35	4,16	6,75										2,82	6,27
9		13,25	1,89	5,29	11,80								3,23	7,09
10		30,96	1,93	5,62						48,85			5,47	18,57
11		13,27	1,89	5,43	30,72	31,98							5,06	14,73
12		13,26	1,89	5,45	147,73			158,06					3,62	55,00
13		498,52	1,92	5,64	307,32		582,56	358,99					3,75	251,24
14		336,41	1,91	5,64	37,54	47,42	380,42						5,07	116,34
15		566,18	1,94	5,76	312,26		895,10	364,66		126,13			6,06	284,76
16		524,93	1,92	5,67	310,78		603,66	359,03			4,07		4,25	226,79
17		580,67	1,94	5,82	314,53		897,77	364,66		129,82	4,19		6,26	256,18
18		661,78	2,00	5,76	328,62		732,92	372,55			9,42	8,62	4,27	236,22

Based on the results from the all-possible subset regression, four models were chosen for further analysis: model 3, model 6, model 7, and model 9. Hereafter they are denoted model A, B, C and D respectively.

Model A ($Fetch_{44}$ and SVF_{dw}) has only two regressors and is therefore the best candidate from a parsimony point of view. It is also the model of the four selected with the lowest degree of multicollinearity, indicated by the VIFs. Further, the MS_{Res} is almost as low as for competitors having more regressors but the C_p value is high, suggesting substantial bias in the predicted values. Model B is the second "best" model having three regressors, but the lower VIF-value than model 5 makes it a preferred candidate. Model C and model D adds more Fetch-derivatives to the model but still have reasonable low VIFs. Model D also has the most favourable C_p-value. At higher numbers of p, the slight increase in R^2 and decrease in *RSS* may not be sufficient considering the very high VIFs.

The coefficient statistics for models A–D obtained from the Linear Regression module in SPSS 19 are shown in Table 5. The figures emphasise previous findings:

$Fetch_{44}$ and SVF_{dw} contribute most to the explanation of wind speed variability, and between the two, $Fetch_{44}$ is superior. Standardised coefficients[1] (*Beta* in Table 5) give insight into the relative importance of each predictor in the model (Field 2005).

Table 5. Regression coefficients and statistics for the four selected candidate models, Model A–D.

MODEL A	Unstandardized Coefficients		Standardized Coefficients			95.0% Confidence Interval for B		Correlations			Collinearity Statistics
	B	Std.	Beta	t	Sig.	Lower	Upper	Zero-order	Partial	Part	VIF
Constant	–.216	.013		–16.945	.000	–.241	–.191				
$Fetch_{44}$	.792	.009	.619	84.915	.000	.774	.810	.865	.849	.520	1.415
SVF_{dw}	1.416	.023	.455	62.468	.000	1.372	1.461	.790	.763	.383	1.415

MODEL B	Unstandardized Coefficients		Standardized Coefficients			95.0% Confidence Interval for B		Correlations			Collinearity Statistics
	B	Std.	Beta	t	Sig.	Lower	Upper	Zero-order	Partial	Part	VIF
Constant	–.231	.013		–18.169	.000	–.256	–.206				
$Fetch_{44}$	.699	.015	.546	48.141	.000	.671	.728	.865	.673	.291	3.517
SVF_{dw}	1.344	.024	.432	55.830	.000	1.297	1.391	.790	.726	.338	1.633
SVF_{uw}	.313	.038	.100	8.224	.000	.238	.388	.831	.154	.050	4.054

MODEL C	Unstandardized Coefficients		Standardized Coefficients			95.0% Confidence Interval for B		Correlations			Collinearity Statistics
	B	Std.	Beta	t	Sig.	Lower	Upper	Zero-order	Partial	Part	VIF
Constant	–.194	.013		–14.830	.000	–.220	–.169				
$Fetch_{44}$	.756	.015	.591	48.837	.000	.726	.787	.865	.678	.291	4.124
SVF_{dw}	1.255	.025	.403	49.267	.000	1.205	1.305	.790	.681	.294	1.887
SVF_{uw}	.497	.042	.159	11.802	.000	.415	.580	.831	.218	.070	5.127
$Fetch_{31}$	–.415	.043	–.102	–9.557	.000	–.500	–.330	.601	–.178	–.057	3.224

MODEL D	Unstandardized Coefficients		Standardized Coefficients			95.0% Confidence Interval for B		Correlations			Collinearity Statistics
	B	Std.	Beta	t	Sig.	Lower	Upper	Zero-order	Partial	Part	VIF
Constant	–.197	.013		–15.031	.000	–.223	–.171				
$Fetch_{44}$	.841	.028	.657	30.374	.000	.787	.896	.865	.498	.181	13.252
SVF_{dw}	1.252	.025	.402	49.270	.000	1.202	1.302	.790	.681	.293	1.888
SVF_{uw}	.525	.043	.168	12.293	.000	.441	.608	.831	.226	.073	5.288
$Fetch_{14}$	–.085	.023	–.076	–3.700	.000	–.130	–.040	.830	–.070	–.022	11.797
$Fetch_{31}$	–.422	.043	–.104	–9.734	.000	–.507	–.337	.601	–.181	–.058	3.231

$Fetch_{44}$ and SVF_{dw} have values > 0.40 in all models, 0.402 being the lowest value for SVF_{dw} in model D. Measured this way, the other regressors may be regarded as less influential. The effect of adding regressors to a model can also be assessed by the correlation statistics[2]. For example, the zero-order correlation for SVF_{uw} in model B is –0.831 indicating a strong positive relationship to γ in the same magnitude as $Fetch_{44}$ and SVF_{dw}. However, accounting for the effects of $Fetch_{44}$ and SVF_{dw}, the partial and part correlation for SVF_{uw} drops to 0.15 and the part correlation to 0.05 respectively,

[1] Standardised coefficients are regression coefficients measured in standard deviation units and so are directly comparable (Field 2009).

[2] In Table 5 (SPSS output table), zero-order correlation is equal to Pearson's correlation between the outcome and the regressor. Partial correlation is the correlation between the regressor and the outcome controlling for the effects of all other regressors. Part correlation is the unique correlation between the regressor and the part of outcome not explained by the other regressors in the model (Field 2009).

while $Fetch_{44}$ and SVF_{dw} remain at high numbers. In the correlation statistics, the phenomenon of changing signs of a correlation when adding more regressors to a model is also illustrated, for example in the case of $Fetch_{31}$ in Model C.

5.4 Model performance: predicting wind speed

The four models A–D, where γ is the amplification factor

$$\ln\gamma_A = -0.216 + 0.792 \times \ln Fetch_{44} + 1.416 \times \ln SVF_{dw}$$

$$\ln\gamma_B = -0.231 + 0.699 \times \ln Fetch_{44} + 1.344 \times \ln SVF_{dw} + 0.313 \times \ln SVF_{uw}$$

$$\ln\gamma_C = -0.194 + 0.756 \times \ln Fetch_{44} + 1.255 \times \ln SVF_{dw} + 0.497 \times \ln SVF_{uw} - 0.415 \times \ln Fetch_{13}$$

$$\ln\gamma_D = -0.197 + 0.841 \times \ln Fetch_{44} + 1.252 \times \ln SVF_{dw} + 0.525 \times \ln SVF_{uw} - 0.085 \times \ln Fetch_{14} - 0.422 \times \ln Fetch_{31}$$

were used to predict wind speed for the four city models 1–4 and all wind directions (S, SW and W). The resulting wind speed maps for CM1 are shown in Figure 22.

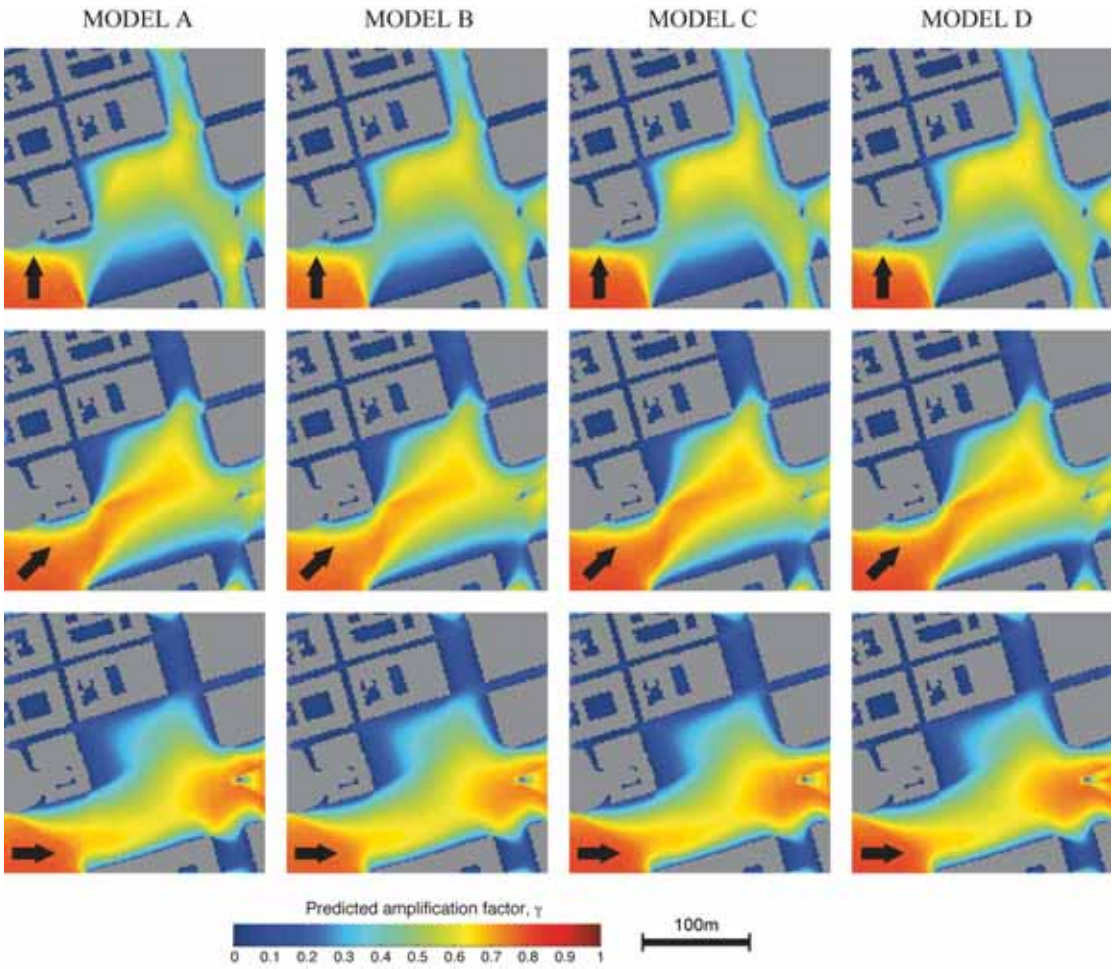

Figure 22 Predicted amplification factor for CM1, Models A–D. Arrows indicate external wind direction: S (top), SW (middle), and W (bottom).

The four models predict very similar wind speed pattern with almost no visible differences. In general the models predict the observed wind speed pattern fairly well if the spatial distribution of high and low winds are considered (compare to Figure 9) However, there are major deviations in the predicted values, which become apparent when mapping the residuals as in Figure 24.

The prediction power of the four models (A–D) was assessed by computing RMSE for the prediction of wind speed within the sampling areas for all twelve City Models. This includes pixels that were not sampled and consequently not used in the model building process. The result is displayed graphically in Figure 23. Two things are evident: (1) there are large differences between City Models (or at least sampling areas) and (2) there is no obvious “best” model.

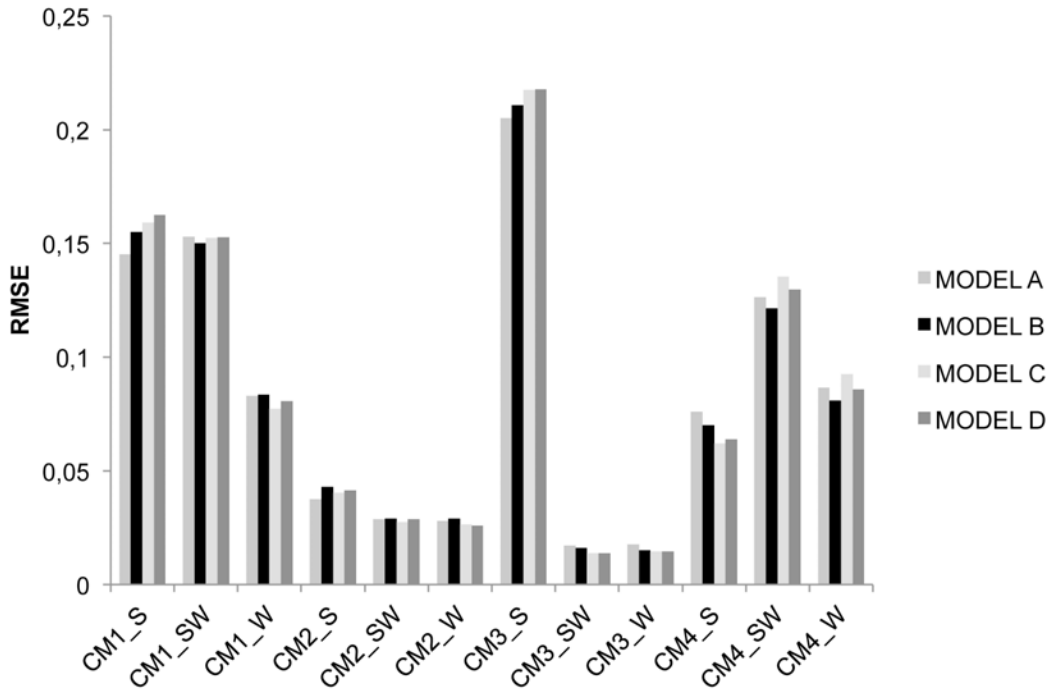

Figure 23 RMSE for all city models, A–D for all wind directions.

The mean of RMSE over all City Models for the four models are 0.084 (Model A), 0.084 (Model B), 0.085 (Model C) and 0.085 (Model D) confirming the small differences. There is a weak indication that Model A performs better in both $CM1_S$ and $CM3_S$ which suggests that Model A predicts wind in the wide canyon parallel to external wind better than the other models.

The spatial distribution of the prediction errors, the residuals, is shown in Figure 24. Negative residuals indicate under estimation and vice versa. It is evident that most contribution to the total error is due to underestimation at locations with high wind

speeds, such as corner streams. The effect of angle-of-attack is also clearly visible, comparing the three wind directions as the wind turns to west. Winds in the wide street canyon go from very high (S) to very low (W) and simultaneously the residuals go from high negative (under estimation) to slightly positive (over estimation).

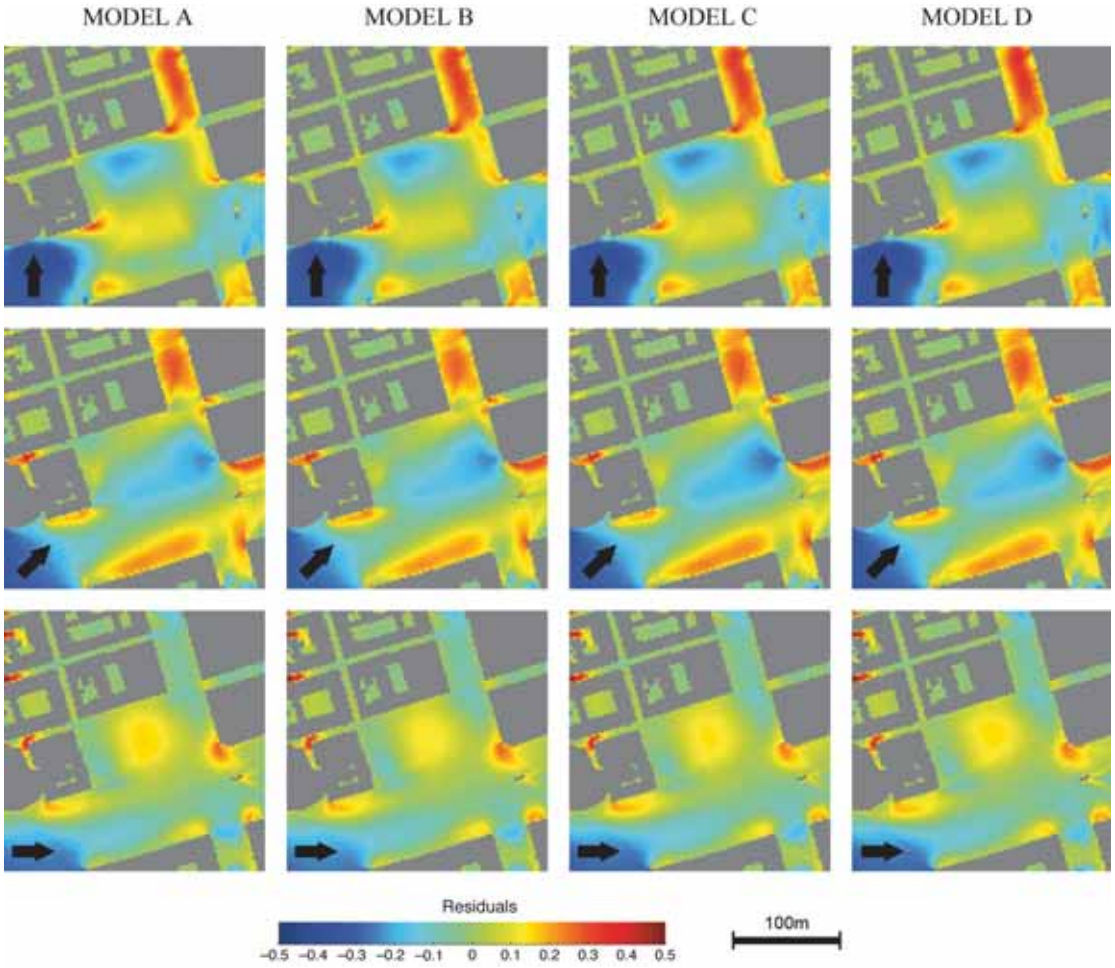

Figure 24 Maps showing the residuals of the predicted amplification factor for CM1. Arrows indicate external wind direction: S (top), SW (middle), and W (bottom).

6 Discussion

Based on the recognition of the influence on the microclimate by the building geometry within the urban canopy, possible relationships between near ground wind speed and the geometry have been investigated. The methodological strategy was to use simulated wind speed as response variable and measures of the geometry as predictor variables in multiple linear regression in order to estimate equations (models) for the prediction of wind speed. The results presented above show that many of the derived measures of the geometry show high correlation with observed wind speed. Furthermore, the estimated models are able to predict general wind speed distribution patterns fairly well. However, the results also point to some limitations of the applied methods. Primarily, they are due to (1) the modelling domain of the ENVI-met model, and (2) the spatial characteristics of the investigated phenomena. In this chapter the results are discussed within this context.

6.1 Spatial distribution of variables

6.1.1 Observed wind speed

The simulated wind speed patterns produced by ENVI-met has been examined by visual inspection of maps. In general, the maps shown in Figure 9 agree with the expected characteristics of wind flow distribution within the urban canopy for complex urban environments as described in Chapter 2. High wind speeds are found in corner streams and in the centre of wide canyons approximately parallel to the external wind direction. The highest wind speeds develop where these two co-occur (e.g. location (1), $CM1_S$ in Figure 9). Medium wind speeds are found in open areas while low wind speeds are located in close proximity to buildings; in narrow canyons, in the wake zones, and at the windward side of buildings facing open areas (e.g. location (6) in Figure 9). However, due to the windward vortex produced by the downward deflection of air down along windward walls, a zone with relatively high wind speeds is expected close to the wall and a stagnation zone is expected at some distance in the upwind direction. In the maps in Figure 9 there are no signs of the windward vortex. This suggests limitations of the capabilities of ENVI-met to simulate wind speed. Another possibility is that the conditions (geometry and configuration settings) are so that the vortex does not develop. The reason behind this is not investigated here, but if the simulated wind speed is erroneous, it has implications for the applicability for the use of any model estimated by this methodology in a real world scenario.

6.1.2 Measures of urban geometry

Of the four measures of urban geometry (i.e. SVF, Fetch, FAI and αFAI), the results show that SVF and Fetch contributes most to the explanation of wind speed variability. As a single predictor SVF_g comes out as ahead, followed by $Fetch_{44}$. By mapping both amplification factor and DSM derivatives, the spatial distribution of the quantities may be examined and compared. The results showed that some of the expected features of the distribution of wind speed such as sheltering by buildings has its approximate counterparts in many of the derivatives, e.g. in SVF_g, SVF_{uw} and in the Fetches.

Sky view factor

The maps shown in the top row in Figure 10 shows the distribution of three of the SVF derivatives. As expected, low values are found close to buildings and high values in open areas, which in a very general way correspond to the observed wind speed pattern. This suggests that the proximity to buildings and building height are important components explaining the spatial distribution of wind speed. SVF_{uw} and SVF_{dw} emphasise different locations, the leeward and windward sides of buildings respectively. Of all derivatives, only SVF-derivatives correspond to the decrease in wind speed at the downwind side of open areas, i.e. location (6) in Figure 9 and of these SVF_{dw} does “the best job”. The rate of change is high close to buildings which is why most areas have a SVF value higher than 0.7. This explains the appearance of the scatter diagrams in Figure 11 where there is high wind speed variability in this part of the diagrams. The SVF maps are also characterized by naturally looking smooth surfaces, similar to the wind speed maps in Figure 9. SVF_{rw} and SVF_{lw} do not seem to be of any importance given the used dataset.

Fetch

$Fetch_{44}$ is the one of the Fetches computed by maximum search distance and maximum search angle. As the second “best” predictor, it shows similarities with wind speed distribution at the leeward side of open areas and in narrow street canyons. Fetch is a measure of the distance to the nearest obstacle in the upwind direction, adding a directional component to the notion made above on the importance of proximity to buildings. A comparison to the other examples of Fetches in Figure 10 uncovers an important difference, resulting from the use of different search angle and search distance. Of these three maps, $Fetch_{22}$ is the one with the shortest search distance and smallest search angle. As is clearly seen in this map (and in Figure 25), a very large area (many pixels) has a value of one. This “saturation effect” means that, in this area, there is no variability in the predictor variable that can explain the variability in wind speed. This is also evident in the scatter diagrams in Figure 12,

where there are a large number of observations at e.g. $Fetch_{22}$ = 1. This effect decreases with increasing search angle and search distance, resulting in potential better predictors (e.g. $Fetch_{44}$). In the sampled dataset, the number of observations where Fetch is equal to one is for $Fetch_{11}$, $Fetch_{14}$, and $Fetch_{44}$ 2,334, 460 and 1 respectively (out of a total number of 2806). One conclusion of this is that the search distance should match the scale of open areas in the data.

The distribution of wind speed and Fetch along a profile running over an open area or across a wide street in downwind direction shows a contradictory pattern. This situation is illustrated in Figure 25. At the upwind end of the profile (A), both wind speed and Fetch increase. At some distance downwind, the increase in wind speed flattens out having a maximum just over half way. Approaching the down wind end of the profile (B), wind speed decreases. Fetch, on the other hand, continues to increase along the whole profile meaning that there is a non-linear relationship between these two variables along the profile. Hence, the correlation between two variables varies in space.

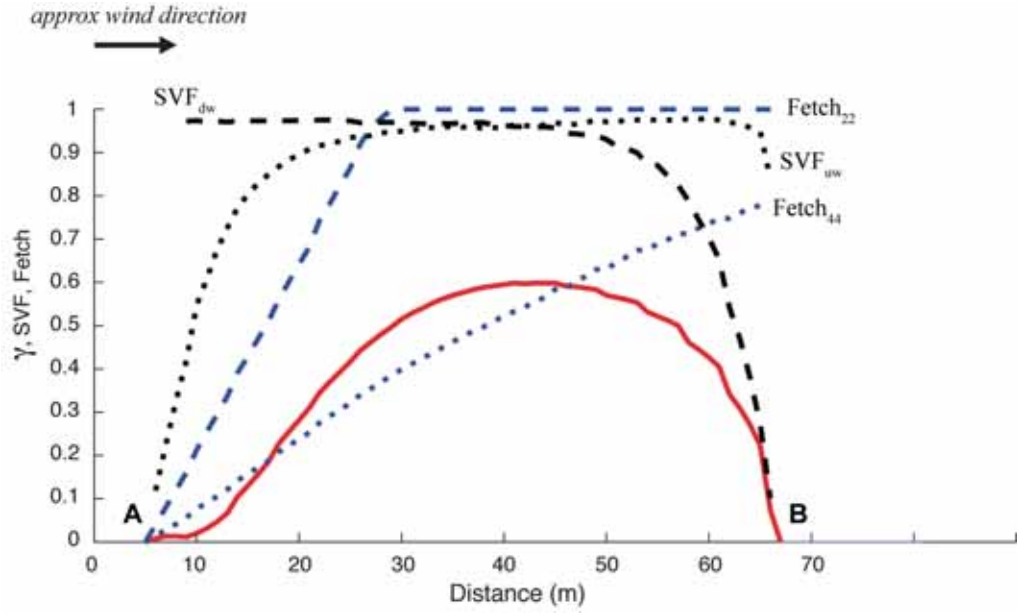

Figure 25 Variation of wind speed (red line) and four predictors along a profile running in approximately WSW direction across the open square in CM1. The variable profiles are for wind direction W.

Of the predictors included in Figure 25 only SVF_{dw} follows the decrease in wind speed approaching B. The combination of these two curves may serve as an illustration of the two variable model favoured by the best subset regression algorithm (Section 5.3.2)

Frontal Area Index and Angular Frontal Area Index

Three FAI-derivatives survived the variable selection algorithm but none of these were favoured by the all-subset regression. No αFAI-derivatives were selected, having F_i values < 10. This indicates that the highly local spatial variations in wind speed at pedestrian height do not correspond to the variations in FAI and αFAI. One suggestion might be that wind speed and FAI and αFAI correlate on an aggregated scale.

6.1.3 Transformed variables

Examination of scatterplots (section 5.1) suggested non-linear relationships between wind speed and all of the predictors, but due to the complex spatial patterns discussed above, the scatter diagrams show very dispersed patterns and weak trends with embedded structures and "subtrends" within the point clouds. Thus, any applied transformation is not likely to improve on all observations. However, the natural log used here considerably improved the full model fit. Figure 26 shows the transformed variables along the same profile as in Figure 25. The same discussion as untransformed variables above is relevant also for this figure, but a comparison of the two strengthen the evidence for the applied transformations; the apparent shape of the predictor curves matches that of wind speed better.

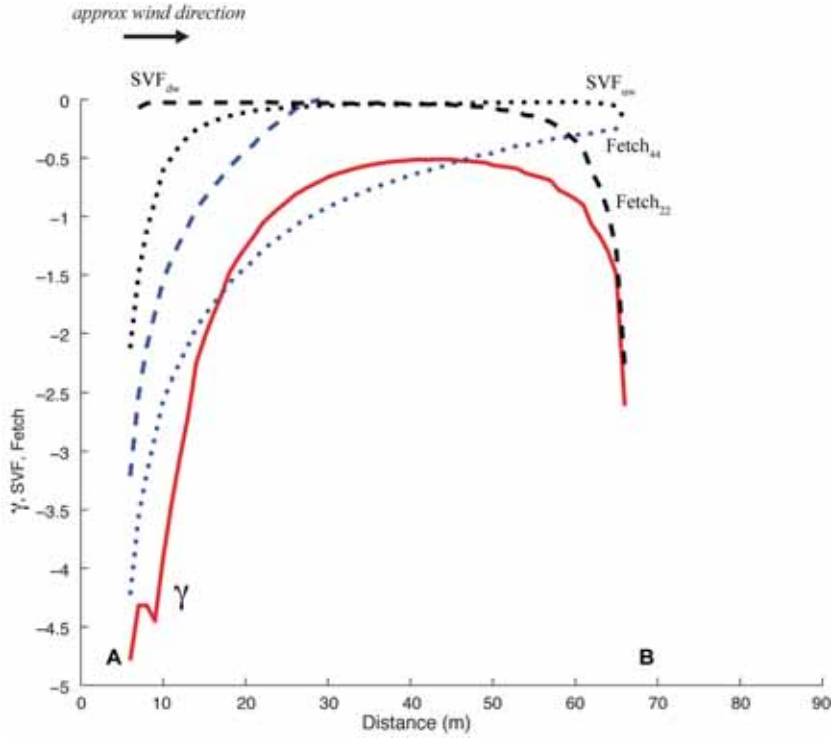

Figure 26 Variation of natural log-transformed wind speed (red line) and four predictors along a profile running in approximately WSW direction across the open square in CM1. The variable profiles are for wind direction W.

It is important not to draw hasty conclusions on this evidence alone; it is only one profile in one city model for one wind direction. However, it is an illustration of the effect of variable transformations in a spatial context, which is not possible to read from a scatter diagram.

6.1.4 Predicted wind speed

Looking at the results, there is no clear evidence in favour of any of the selected models A–D; no one consistently predicts wind speed much better over the twelve evaluation datasets than the others. There is also no evidence for the existence of any model *not* selected that should perform (much) better than any of the selected models, given the set of candidate regressors included in this study.

The small differences between the models implied by Figure 22 and Figure 23 may be explained by the fact that they do include the same predictors, $Fetch_{44}$ and SVF_{dw}, which are the predictors selected as number two and three by the sequential selection algorithm. Since these two predictors explain most of the variability in observed wind speed, the other predictors SVF_{uw}, $Fetch_{41}$ and $Fetch_{31}$ do not add information unique enough to make a great impact on the prediction. To illustrate this, the difference between model A and models B, C and D are mapped in Figure 27.

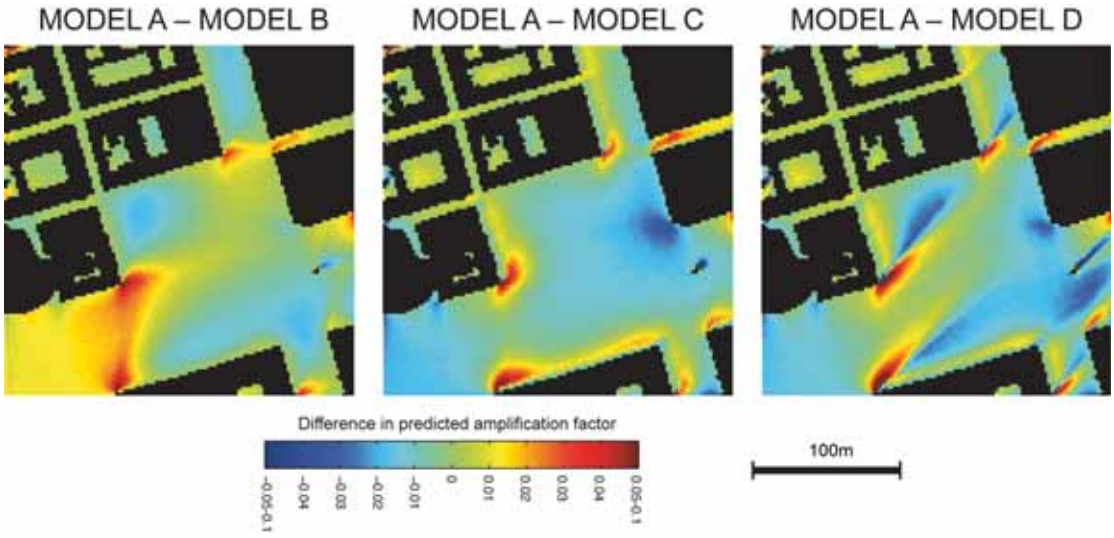

Figure 27 Differences in predicted amplification factor between model A and models B, C and D, $CM1_{SW}$. The colour scale is saturated at both ends to emphasise the spatial pattern in the differences.

The main differences between predicted and observed wind speed, as well as between models, are in areas where wind speed is high due to channelling or deflection by buildings (corner streams). In such locations all models underestimate wind speed both in magnitude and in areal extension. Conversely, all models are in good agreement with the observed low wind speeds in narrow canyons and in canyons perpendicular to the wind direction. Here, the differences between models are also

almost zero, especially between model A and B. This is for example demonstrated by the low RMSE for CM2, $CM3_{SW}$ and $CM3_{W}$ where narrow canyons and thus low wind speeds dominate the sampling areas (Figure 9). This suggests that the models are biased towards low wind speeds.

6.2 Comments on the methodology

6.2.1 The data

The small spatial domain of the CFD model limits the computation of DSM-derivatives (i.e. possible search distances and search angles) and the effective area that can be used for sampling. This in turn makes the choice of study areas very important, if they are to provide a dataset representative of different urban environments. For example, the sampling areas in $CM1_{S}$ and $CM1_{W}$ covers corner streams and not much more. Given the results, such areas are likely to influence a model in a negative way, i.e. increase the prediction errors. The small spatial domain may also be one reason for the exclusion of FAI and αFAI as they work on a different scale than the more local measures Fetch and SVF.

Even though the city models are constructed using real world urban geometry (DSM), they are simplifications due to the removal of topography. Also vegetation is not included in the models. This means that any model estimated with this method apply to idealized urban environments. Also the quality of the original GIS-data used to create the city models are of importance and may influence the results.

6.2.2 Model building

The scientific idea of parsimony is central to the concept of variable selection and model building. A model with a few predictors has both practical and statistical advantages; e.g. data acquisition is less costly, computations may be more effective, and the statistical properties of the estimated coefficients are more robust.

To select variables, an appropriate selection criterion in combination with expert knowledge about the investigated phenomenon is very important. In this study, being an exploratory effort on the edge of a data mining adventure, several statistics that selection could be based on were computed. The set of candidate predictors were chosen based on the assumption that they are appropriate spatial measures of urban geometry, which in turn is the major influence on wind speed in the urban canopy layer.

In order to identify those measures, or combinations of measures (models), of urban geometry that explain most of the wind speed variability, a sequential selection algorithm utilized RSS as selection criterion, was applied. The number of predictors

in the model was controlled by BIC; predictors were entered to the model as long as BIC decreased. Given the result of the final evaluation of model A–D, it can be concluded that this approach is too generous, adding too many regressors and overfitting the models. However, the computation of the frequency of appearance and average scores (Figure 18) serves well as tools to cut down the number of candidate predictors. The predictors with high *F* and low average scores are also the ones favoured by the all-possible subset regression. The use of such criteria is subjective and the threshold (e.g. $F_i > 90$) was set based on the results (Figure 18).

Not unexpected, the full set of predictors showed very high collinearity; all predictors are measures of the same geometry and to some extent they all measures distances to buildings. SVF_{dw} show consistently the lowest VIF-values. This is because it is a measure considering the geometry in the opposite direction (downwind) compared to all other predictors except SVF_g, which also include this direction. The reason for including many predictors of the same type (e.g. Fetch) was to examine the effect of search angle and search distance. The results show that adding more than one Fetch predictor to the model, does not gain much in predictive power.

High multicollinearity may alter the sign of the estimated coefficients so that they conflict with the assumed correlation to wind speed, i.e. positive or negative correlation. This happens in model C and D when additional Fetches enter the model. If only prediction is the purpose, this may be disregarded. However, it is an indication of an over estimated model including redundant data, which may be less predictive should it be applied on a new set of data. In this study, no thorough inspection of outliers and influence points were made. Such observations can have great impact on the results and should be evaluated and eventually removed from the dataset.

The all-possible subset regression algorithm ranks models based on computed statistics and thus is an objective method. However, the final selection of models (A–D) was a subjective choice guided by those statistics. A possible alternative to this is to use several models by model weighting, which can be done by AIC.

7 Conclusions and further developments

There is no thing as a true model; there is truth and there are models, and at most a model may be a useful approximation of that truth. In this thesis a first attempt to build a useful statistical model for the prediction of near ground wind speed from a set of measures of urban geometry has been described. Such model can be utilized in research and applications concerned with outdoor thermal comfort, urban climate modelling and climate planning. The specific objectives were to examine the relationship between the response and predictor variables and to identify those predictors that explain most of the wind speed spatial variability. From the set of selected candidate predictors, models were fitted and evaluated.

In general, the relationships between wind speed and the predictors are very complex and there is evidence indicating that correlations vary in space. Multiple Fetch, FAI and αFAI derivatives were computed using different search angle and search distance, and it can be concluded that these settings are important. For Fetch and αFAI, the correlation to wind speed improved with both increasing search angle and search distance. FAI had best correlation at a search distance of 50 meters. None of the predictors showed clear linear relationships to wind speed, which is why transformations were applied. Transforming wind speed, SVF and Fetch by the natural logarithm, leaving FAI and αFAI untransformed improved the fit of the full model considerably.

The variable selection and model building procedure favoured models including Fetch and SVF-predictors, of which $Fetch_{44}$ and SVF_{dw} explained most of the variability. The model with these two predictors explained 90 % of the wind speed (R^2 = 0.895). Adding more regressors did not improve predicted wind maps. Models having more predictors of the same type (e.g. multiple Fetches) suffered from severe multicollinearity problems.

It can be concluded that, despite absence of a good over all fit, much of the variability in wind speed may be explained by SVF and Fetch. However, wind speed patters due to deflection of the airflow (e.g. corner streams) or channelling does not allow for prediction using this method. To improve prediction ability, future work should be directed towards measures depicting the spatial context of the geometry, e.g. alignment of street canyons and height and width ratios of urban spaces, and to develop algorithms for the computation of these as continuous raster data. Also, the measures used in this study may be elaborated and optimized to improved model performance. This should be possible to achieve within a raster based GIS utilizing

the shadow casting algorithm and digital surface data. Another focus would be to try different regression techniques and variable section methods to improve model specification and estimation of regression coefficients.

Finally, the lack of the windward vortex in the simulated wind speed data does not agree with previous research on urban wind, and indicates possible shortcomings of the three-dimensional modelling procedure. Hence, the simulated wind speed obtained from ENVI-met should be thoroughly examined and validated in order to assess the applicability of the methods and any statistical model in estimating wind speed.

8 References

Ahmad, K., Khare, M. & Chaudhry, K.K., 2005. Wind tunnel simulation studies on dispersion at urban street canyons and intersections—a review. *Journal of Wind Engineering and Industrial Aerodynamics*, 93, pp.697–717.

Arnfield, A John, 2003. Two decades of urban climate research: a review of turbulence, exchanges of energy and water, and the urban heat island. *International Journal of Climatology*, 23(1), pp.1–26.

Baik, J.-J. & Kim, J.-J., 1999. A numerical study of flow and pollutant dispersion characteristics in urban street canyons. *Journal of Applied Meteorology*, 38, pp.1576–1589.

Blocken, B. & Carmeliet, J., 2004. Pedestrian Wind Environment around Buildings: Literature Review and Practical Examples. *Journal of Building Physics*, 28(2), pp.107–159.

Britter, R.E. & Hanna, S.R., 2003. Flow and dispersion in urban areas. *Annual Review of Fluid Mechanics*, 35, pp.469–496.

Bruse, M., 1999. *Die Auswirkungen kleinskaliger Umweltgestaltung auf das Mikroklima. Entwicklung des prognostischen numerischen Modells ENVI-met zur Simulation der Wind-, Temperatur-, und Feuchtverteilung in städtischen Strukturen*, PhD Thesis, Univ. Bochum, Germany.

Bruse, M. ed. *ENVI-met website* M. Bruse, ed. Available at: http://www.envi-met.com [Accessed December 14, 2011].

Burnham, K.P. & Anderson, D.R., 2004. Multimodel inference: Understanding AIC and BIC in Model Selection. *Sociological methods & research*, 33, pp.261–304.

Burrows, S. et al., 2002. Application of geostatistics to characterize leaf area index (LAI) from flux tower to landscape scales using a cyclic sampling design. *Ecosystems*, 5(7), pp.667–679.

Carpentieri, M., Robins, A.G. & Baldi, S., 2009. Three-dimensional mapping of air flow at an urban canyon intersection. *Boundary-Layer Meteorology*, 133, pp.277–296.

Chen, L. & Ng, E., 2011. Quantitative urban climate mapping based on a geographical database: A simulation approach using Hong Kong as a case study. *International Journal of Applied Earth Observations and Geoinformation*, 13(4), pp.586–594.

Dobre, A. et al., 2005. Flow field measurements in the proximity of an urban intersection in London, UK. *Atmospheric Environment*, 39(26), pp.4647–4657.

Eliasson, I. et al., 2006. Wind fields and turbulence statistics in an urban street canyon. *Atmospheric Environment*, 40(1), pp.1–16.

Field, A., 2005. *Discovering statistics using SPSS: (and sex and drugs and rock“n”roll)* 2nd ed. London: Sage Publications.

Gal, T. & Unger, J., 2009. Detection of ventilation paths using high-resolution roughness parameter mapping in a large urban area. *Building and Environment*, 44(1), pp.198–206.

Georgakis, C. & Santamouris, M., 2008. On the estimation of wind speed in urban canyons for ventilation purposes—Part 1: Coupling between the undisturbed wind speed and the canyon wind. *Building and Environment*, 43, pp.1404–1410.

Graham, M.H., 2003. Confronting Multicollinearity in Ecological Multiple Regression. *Ecology*, 84(11), pp.2809–2815.

Grimmond, C.S.B., 2007. Urbanization and global environmental change: local effects of urban warming. *The Geographical Journal*, 173(1), pp.83–88.

Grimmond, C.S.B. & Oke, T.R., 1999. Aerodynamic properties of urban areas derived from analysis of surface form. *Journal of Applied Meteorology*, 38(9), pp.1262–1292.

Hunter, L.J., Watson, I.D. & Johnson, G.T., 1991. Modelling air flow regimes in urban canyons. *Energy and Buildings*, 15(3-4), pp.315–324.

Johnson, G.T. & Hunter, L.J., 1999. Some insights into typical urban canyon airflows. *Atmospheric Environment*, 33(24-25), pp.3991–3999.

Johnson, J.B. & Omland, K.S., 2004. Model selection in ecology and evolution. *Trends in Ecology & Evolution*, 19(2), pp.101–108.

Kastner-Klein, P., Berkowicz, R. & Britter, R.E., 2004. The influence of street architecture on flow and dispersion in street canyons. *Meteorology and Atmospheric Physics*, 87(1-3).

Kim, S. & Boysan, F., 1999. Application of CFD to environmental flows. *Journal of Wind Engineering and Industrial Aerodynamics*, 81(1-3), pp.145–158.

Lindberg, F., 2007. Modelling the urban climate using a local governmental geo-database. *Meteorological Applications*, 14(3), pp.263–273.

Lindberg, F. & Grimmond, C.S.B., 2010. Continuous sky view factor maps from high resolution urban digital elevation models. *Climate Research*, 42, pp.177–183.

Luo, W., 2008. Spatial/Temporal Modelling of Crop Disease Data using High-Dimensional Regression.

Mallows, C.L., 1973. Some comments on Cp. *Technometrics*, 15(4), pp.661–675.

Montgomery, D.C., Peck, E.A. & Vining, G.G., 2001. *Introduction to linear regression analysis* 4th ed. Hoboken, New Jersey: John Wiley & Sons, Inc.

Murena, F. et al., 2009. Modelling dispersion of traffic pollution in a deep street canyon: Application of CFD and operational models. *Atmospheric Environment*, 43(14), pp.2303–2311.

Nakamura, Y. & Oke, T.R., 1988. Wind, temperature and stability conditions in an east-west oriented urban canyon. *Atmospheric Environment*, 22(12), pp.2691–2700.

Oke, T.R., 1988. *Boundary Layer Climates* 2nd ed. Routledge.

Ozkeresteci, I. et al., 2003. Use and Evaluation of the Envi-met Model for Environmental Design and Planning: An Experiment on Linear Parks. *Proceedings of the 21st International Cartographic Conference (ICC)*.

Princevac, M. et al., 2010. Lateral channeling within rectangular arrays of cubical obstacles. *Journal of Wind Engineering and Industrial Aerodynamics*, 98(8-9), pp.377–385.

Ratti, C., 2004. Raster analysis of urban form. *Environment and Planning B*.

Richens, P., 1997. Image processing for urban scale environmental modelling. *Proc 5th Intemational IBPSA Conference: Building Simulation 97, Prague.*

Robins, A., Savory, E. & Scaperdas, A., 2002. Spatial variability and source-receptor relations at a street intersection. *International Journal of Water, Air and Soil Pollution: Focus 2*, pp.381–393.

Rogerson, P.A., 2010. *Statistical Methods for Geography: A Student's Guide* Third Edition. Sage Publications Ltd.

Rosheidat, A., Hoffman, D. & Harvey, B., 2008. Visualizing pedestrian comfort using ENVI-met. *Third National Conference of IBPSA-USA*.

Soulhac, L. et al., 2009. Flow and dispersion in street intersections. *Atmospheric Environment*, 43(18), pp.2981–2996.

Steyn, D.G., 1980. The calculation of view factors from fisheye‐lens photographs: Research note. *Atmosphere-Ocean*, 18, pp.254–258.

Tiwary, A. et al., 2011. Air flow and concentration fields at urban road intersections for improved understanding of personal exposure. *Environment International*, 37(5), pp.1005–1018.

Wang, X., 2007. Effects of street orientation on dispersion at or near urban street intersections. *Journal of Wind Engineering and Industrial Aerodynamics*, 95, pp.1526–1540.

Whittingham, M.J. et al., 2006. Why do we still use stepwise modelling in ecology and behaviour? *Journal of Animal Ecology*, 75(5), pp.1182–1189.

Wieringa, J. et al., 2001. New revision of Davenport roughness classification. *Proc 3rd European & African Conference on Wind Engineering, Eidhoven, Netherlands*.

Wong, M.S., Nichol, J.E. & To, P.H., 2010. A simple method for designation of urban ventilation corridors and its application to urban heat island analysis. *Building and Environment*, 45(8), pp.1880–1889.

Department of Physical Geography and Ecosystems Science Lund University

The student thesis reports are available at the Geo-Library, Department of Physical Geography, University of Lund, Sölvegatan 12, S-223 62 Lund, Sweden.

Report series started 1985. The complete list and electronic versions are also electronic available at the LUP student papers (www.nateko.lu.se/masterthesis) and through the Geo-library (www.geobib.lu.se)

175 Hongxiao Jin (2010): Drivers of Global Wildfires — Statistical analyses

176 Emma Cederlund (2010): Dalby Söderskog – Den historiska utvecklingen

177 Lina Glad (2010): En förändringsstudie av Ivösjöns strandlinje

178 Erika Filppa (2010): Utsläpp till luft från ballastproduktionen år 2008

179 Karolina Jacobsson (2010): Havsisens avsmältning i Arktis och dess effekter

180 Mattias Spångmyr (2010): Global of effects of albedo change due to urbanization

181 Emmelie Johansson & Towe Andersson (2010): Ekologiskt jordbruk - ett sätt att minska övergödningen och bevara den biologiska mångfalden?

182 Åsa Cornander (2010): Stigande havsnivåer och dess effect på känsligt belägna bosättningar

183 Linda Adamsson (2010): Landskapsekologisk undersökning av ädellövskogen i Östra Vätterbranterna

184 Ylva Persson (2010): Markfuktighetens påverkan på granens tillväxt i Guvarp

185 Boel Hedgren (2010): Den arktiska permafrostens degradering och metangasutsläpp

186 Joakim Lindblad & Johan Lindenbaum (2010): GIS-baserad kartläggning av sambandet mellan pesticidförekomster i grundvatten och markegenskaper

187 Oscar Dagerskog (2010): Baösberggrottan – Historiska tillbakablickar och en lokalklimatologisk undersökning

188 Mikael Månsson (2010): Webbaserad GIS-klient för hantering av geologisk information

189 Lina Eklund (2010): Accessibility to health services in the West Bank, occupied Palestinian Territory.

190 Edvin Eriksson (2010): Kvalitet och osäkerhet i geografisk analys - En studie

om kvalitetsaspekter med fokus på osäkerhetsanalys av rumslig prognosmodell för trafikolyckor

191 Elsa Tessaire (2010): Impacts of stressful weather events on forest ecosystems in south Sweden.

192 Xuejing Lei (2010): Assessment of Climate Change Impacts on Cork Oak in Western Mediterranean Regions: A Comparative Analysis of Extreme Indices

193 Radoslaw Guzinski (2010): Comparison of vegetation indices to determine their accuracy in predicting spring phenology of Swedish ecosystems

194 Yasar Arfat (2010): Land Use / Land Cover Change Detection and Quantification — A Case study in Eastern Sudan

195 Ling Bai (2010): Comparison and Validation of Five Global Land Cover Products Over African Continent

196 Raunaq Jahan (2010): Vegetation indices, FAPAR and spatial seasonality analysis of crops in southern Sweden

197 Masoumeh Ghadiri (2010): Potential of Hyperion imagery for simulation of MODIS NDVI and AVHRR-consistent NDVI time series in a semi-arid region

198 Maoela A. Malebajoa (2010): Climate change impacts on crop yields and adaptive measures for agricultural sector in the lowlands of Lesotho

199 Herbert Mbufong Njuabe (2011): Subarctic Peatlands in a Changing Climate: Greenhouse gas response to experimentally increased snow cover

200 Naemi Gunlycke & Anja Tuomaala (2011): Detecting forest degradation in Marakwet district, Kenya, using remote sensing and GIS

201 Nzung Seraphine Ebang (2011): How was the carbon balance of Europe affected by the summer 2003 heat wave? A study based on the use of a Dynamic Global Vegetation Model; LPJ-GUESS

202 Per-Ola Olsson (2011): Cartography in Internet-based view services – methods to improve cartography when geographic data from several sources are combined

203 Kristoffer Mattisson (2011): Modelling noise exposure from roads – a case study in Burlövs municipality

204 Erik Ahlberg (2011): BVOC emissions from a subarctic Mountain birch: Analysis of short-term chamber measurements.

205 Wilbert Timiza (2011): Climate variability and satellite – observed vegetation

responses in Tanzania.

206 Louise Svensson (2011): The ethanol industry - impact on land use and biodiversity. A case study of São Paulo State in Brazil.

207 Fredrik Fredén (2011): Impacts of dams on lowland agriculture in the Mekong river catchment.

208 Johanna Hjärpe (2011): Kartläggning av kväve i vatten i LKAB:s verksamhet i Malmberget år 2011 och kvävets betydelse i akvatiska ekosystem ur ett lokalt och ett globalt perspektiv

209 Oskar Löfgren (2011): Increase of tree abundance between 1960 and 2009 in the treeline of Luongastunturi in the northern Swedish Scandes

210 Izabella Rosengren (2011): Land degradation in the Ovitoto region of Namibia: what are the local causes and consequences and how do we avoid them?

211 Irina Popova (2011): Agroforestry och dess påverkan på den biofysiska miljön i Afrika.

212 Emilie Walsund (2011): Food Security and Food Sufficiency in Ethiopia and Eastern Africa.

213 Martin Bernhardson (2011): Jökulhlaups: Their Associated Landforms and Landscape Impacts.

214 Michel Tholin (2011): Weather induced variations in raptor migration; A study of raptor migration during one autumn season in Kazbegi, Georgia, 2010

215 Amelie Lindgren (2011) The Effect of Natural Disturbances on the Carbon Balance of Boreal Forests.

216 Klara Århem (2011): Environmental consequences of the palm oil industry in Malaysia.

217 Ana Maria Yáñez Serrano (2011) Within-Canopy Sesquiterpene Ozonolysis in Amazonia

218 Edward Kashava Kuliwoye (2011) Flood Hazard Assessment by means of Remote Sensing and Spatial analyses in the Cuvelai Basin Case Study Ohangwena Region –Northern Namibia

219 Julia Olsson (2011) GIS-baserad metod för etablering av centraliserade biogasanläggningar baserad på husdjursgödsel.

220 Florian Sallaba (2011) The potential of support vector machine classification

of land use and land cover using seasonality from MODIS satellite data

221 Salem Beyene Ghezahai (2011) Assessing vegetation changes for parts of the Sudan and Chad during 2000-2010 using time series analysis of MODIS-NDVI

222 Bahzad Khaled (2011) Spatial heterogeneity of soil CO2 efflux at ADVEX site Norunda in Sweden

223 Emmy Axelsson (2011) Spatiotemporal variation of carbon stocks and fluxes at a clear-cut area in central Sweden

224 Eduard Mikayelyan (2011) Developing Android Mobile Map Application with Standard Navigation Tools for Pedestrians

225 Johanna Engström (2011) The effect of Northern Hemisphere teleconnections on the hydropower production in southern Sweden

226 Kosemani Bosede Adenike (2011) Deforestation and carbon stocks in Africa

227 Ouattara Adama (2011) Mauritania and Senegal coastal area urbanization, ground water flood risk in Nouakchott and land use/land cover change in Mbour area

228 Andrea Johansson (2011) Fire in Boreal forests

229 Arna Björk Þorsteinsdóttir (2011) Mapping *Lupinus nootkatensis* in Iceland using SPOT 5 images

230 Cléber Domingos Arruda (2011) Developing a Pedestrian Route Network Service (PRNS)

231 Nitin Chaudhary (2011) Evaluation of RCA & RCA GUESS and estimation of vegetation-climate feedbacks over India for present climate

232 Bjarne Munk Lyshede (2012) Diurnal variations in methane flux in a low-arctic fen in Southwest Greenland