Key conclusions of the first international urban land surface model comparison project

Article

Accepted Version

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

1 **Key conclusions of the first international urban land surface model comparison**

2 **project**

3

4 M.J. Best, C.S.B. Grimmond

5

6

7 AFFILIATIONS: Best – Met Office, Exeter, UK, and King's College London,

8 London, UK; Grimmond – University of Reading, Reading, UK

9

10 CORRESPONDING AUTHOR: Martin J. Best, Met Office, FitzRoy Road, Exeter,

11 EX1 3PB, UK.

12 E-mail: martin.best@metoffice.gov.uk

13

ABSTRACT: The first international urban land surface model comparison was

designed to identify three aspects of the urban surface-atmosphere interactions: (1) the

dominant physical processes, (2) the level of complexity required to model these, and

(3) the parameter requirements for such a model. Offline simulations from 32 land

surface schemes, with varying complexity, contributed to the comparison. Model

results were analysed within a framework of physical classifications and over four

stages. The results show that the following are important urban processes; (i) multiple

reflections of shortwave radiation within street canyons, (ii) reduction in the amount

of visible sky from within the canyon, which impacts on the net long-wave radiation,

(iii) the contrast in surface temperatures between building roofs and street canyons,

and (iv) evaporation from vegetation. Models that use an appropriate bulk albedo

based on multiple solar reflections, represent building roof surfaces separately from

street canyons and include a representation of vegetation demonstrate more skill, but

require parameter information on the albedo, height of the buildings relative to the

width of the streets (height to width ratio), the fraction of building roofs compared to

street canyons from a plan view (plan area fraction) and the fraction of the surface that

is vegetated. These results, whilst based on a single site and less than 18 months of

data, have implications for the future design of urban land surface models, the data

that need to be measured in urban observational campaigns, and what needs to be

included in initiatives for regional and global parameter databases.

Capsule Summary

The conclusions from the first international urban land surface model comparison

project have implications for future models, observations and parameter databases,

that extend beyond the urban modelling community

1. Introduction

Urban areas are often warmer than their surrounding rural environments, referred to as

the urban heat island (UHI). This urban warming has numerous effects, including the

initiation of convective storms (e.g., *Bornstein and Lin*, 2000), altering pollution

dispersion by adapting mixing through changes to atmospheric boundary layer

structure (e.g., *Sarrat et al.*, 2006, *Luhar et al.*, 2014), impacts on the production and

mixing of ozone (e.g., *Chaxel and Chollet*, 2009, *Ryu et al.*, 2013), enhanced energy

demand for summer-time cooling through air conditioning (e.g., *Radhi and Sharples*,

2013, *Li et al.*, 2014), impacts on urban ecology (e.g., *Pickett et al.*, 2008, *Francis and

Chadwick*, 2013) and increased mortality rates during heat waves (e.g., *Laaidi et al.*,

2012, *Herbst et al.*, 2014, *Saha et al.*, 2014). As such, it is important to be able to

accurately forecast urban warming and other meteorological variables for cities where

the majority of the World's population now lives.

Predictions of future climate suggest additional warming in urban environments

(*McCarthy et al.*, 2010, *Oleson et al.*, 2011). Indeed, the Inter-Governmental Panel on

Climate Change (IPCC) Working Group 1 Fifth Assessment Report (*IPCC*, 2013)

included at least one model that explicitly included an urban representation, and this

number is likely to increase in the future as the resolution of these climate models

increases to the extent that some urban areas are resolved. For future design of

60   buildings and planning of cities, it is important that the dominant processes that lead

61   to urban warming effects are considered. This requires the development of models

62   that can represent the most important features of the urban heat island be used for

63   reliable predictions.

64

65   The urban heat island results from differences in surface energy exchanges between

66   the urban environment and its surrounding rural area. Thus, understanding these

67   differences is needed to interpret the urban heat island. The differences in urban

68   surface energy exchanges arise through a number of processes. The geometry of a

69   street canyon will increase the incoming solar radiation and long-wave radiation that

70   are absorbed, due to multiple reflections and re-radiated from the 3-dimensional

71   structures. The orientation of street canyons and the elevation of the sun will impact

72   the reflected solar radiation, as a consequence of the depth to which the direct

73   sunshine can penetrate into the canyon. The reduced availability of water at the urban

74   surface, compared to natural vegetated or bare soil surfaces, means more of the

75   incoming solar radiation is transformed into heat rather than a flux of moisture into

76   the atmosphere. However, a larger proportion of this energy for heating is held within

77   the fabric of the buildings given the large thermal inertia of the materials, resulting in

78   changes in the diurnal cycle of urban temperatures. Moreover, an additional source of

79   heating within the urban areas comes from human activities such as transport, the

80   internal heating of the buildings and the metabolic rates of the people themselves

81   (e.g., Sailor and Lu, 2004).

82

83   All of these processes contribute to the differences in the energy balance between

84   urban and rural surfaces, but it is difficult to identify which are the dominant

85   processes just from observations as the processes cannot be separated because of the

86   complex nature of the environment. As such, the best way to study these processes

87   individually is by using urban land surface models (ULSMs) that have been

88   developed for weather and climate applications, i.e., exchange surface fluxes with an

89   atmospheric model. There are a number of such ULSMs that vary considerably in

90   their complexity (e.g., *Kusaka et al.*, 2001, *Fortuniak*, 2003, *Krayenhoff and Voogt*,

91   2007, *Hamdi and Masson*, 2008, *Lee and Park*, 2008, *Oleson et al.*, 2008a). Although

92   newer models often include more complex features than previous models, without

93   knowing the dominant processes and controls, it is difficult to quantify the impact of

94   each new feature.

95

96   The first urban land surface model comparison was designed to objectively assess and

97   compare the performance of a range of ULSMs for a single observational site. It

98   attempted to identify the dominant physical processes that need to be represented in

99   ULSMs by comparing models of varying complexity (Table 1). These models ranged

100   from simple bulk representations of the surface that have been applied to atmospheric

101   models for over a decade, representations of the facets of a street canyon (i.e., roofs,

102   walls and road) that have been used in weather and climate models, through to more

103   recently developed schemes that consider a complete energy balance at various levels

104   within the urban canyon that have been applied to stand alone single point studies.

105   Figure 1 shows a conceptual representation of the surface energy balance for these

106   models of varying complexity. Whilst the scale that these models typically represent

107   is larger than the size of the elements within a street canyon, a common feature is the

108   ability to predict the exchange of fluxes between the urban surface and the atmosphere

109   above it, i.e., the net all-wave radiation (Q^*), turbulent sensible (Q_H) and latent heat

110 (Q_E) fluxes, as measured from flux towers in numerous urban observational

111 campaigns.

112 The aim of the urban model comparison was to consider:

113 (1) What are the dominant physical processes in the urban environment?

114 (2) What is the level of complexity required for an ULSM to be fit for purpose?

115 (3) What are the parameter requirements for such a model?

116 Here we present an analysis of the model comparison results to address these

117 questions.

118

119 **2 Model Comparison design**

120 **2. 1 Observational data**

121 The criteria for selecting the evaluation dataset were; first it had not been used to

122 evaluate any ULSMs previously, and second it needed to cover an annual cycle to

123 allow assessment for different seasons. Model evaluation studies often result in the

124 development and optimisation of a model in order to obtain better representation of

125 the assessed metrics. Hence, using a dataset previously used by one or a sub-set of the

126 models to be evaluated would not enable a clean/independent objective assessment for

127 all of the models.

128

129 The dataset for a suburb of Melbourne (Preston) (*Coutts et al.*, 2007a, 2007b) that had

130 observations from 13 August 2003 to 13 November 2004 was selected. The

131 moderately developed, low-density housing area is classified by *Coutts et al.* (2007b)

132 as an Urban Climate Zone (UCZ) 5 (*Oke*, 2006), Local Climate Zone (LCZ) 6

133 (*Stewart and Oke*, 2012) or *Loridan and Grimmond* (2012) Urban Zone for Energy

134 exchange (UZE) medium density. The description of UCZ 5 is "medium

135 development, low density suburban with 1 or 2 storey houses, e.g., suburban housing"

136 (*Oke*, 2006), and as such the site is typical of suburban areas found in North America,

137 Europe and Australasia. The area has mean building height-to-width ratio of 0.42 and

138 mean wall-to-plan ratio of 0.4 (*Coutts et al.*, 2007b). The surface is dominated by

139 impervious cover (44.5% buildings, 4.5% concrete and 13% roads), with a pervious

140 cover of 38% (15% grass, 22.5% other vegetation and 0.5% bare ground or pools)

141 (*Coutts et al.*, 2007a).

142

143 The methods used to obtain the observed fluxes applied to our current analysis are

144 given in Table 2, with details (e.g., data processing) presented in the original

145 observation papers (*Coutts et al.*, 2007a, 2007b). In addition, the initial model

146 comparison results papers (*Grimmond et al.*, 2011, *Best and Grimmond*, 2013, 2014)

147 provide the site parameters. A continuous gap-filled atmospheric forcing dataset (474

148 days) to run the models was created for this study (see *Grimmond et al.*, 2011). To

149 evaluate the modelled fluxes (sensible heat flux, latent heat flux, net all-wave

150 radiative flux and net storage heat flux (ΔQ_S)) 30 min periods are used when no

151 observed fluxes are missing to allow consistent analysis between the fluxes (N=8865

152 or 38.9% of the full period).

153

154 **2.2 Data analysis**

155 To permit the research questions posed above to be considered, information about the

156 observational site was released to the modelling groups in stages. This enabled

157 analysis of the importance of the different types of information to model performance

158 through assessment of the change in model skill between the stages. The stages (Table

159 3), designed to correlate with ease of access to information for all cities globally,

160   involved release of (*Grimmond et al.*, 2011):

161   *Stage 1*: Atmospheric forcing data: (Table 3), typically provided by an atmospheric

162   model.

163   *Stage 2:* Vegetation and built fraction: two dimensional plan area characteristics of

164   the site. These can be determined from land cover datasets derived from satellite data.

165   *Stage 3*: Morphology: three dimensional characteristics of the site (Table 3.). These

166   can be interpreted from LiDAR (e.g., *Goodwin et al.*, 2009, *Lindberg and Grimmond,*

167   2011), aerial photographs (e.g., *Ellefsen*, 1990/1991), detailed satellite imagery (e.g.,

168   *Brunner et al.*, 2010), or simple empirical relations (e.g., *Bohnenstengel et al.*, 2011).

169   *Stage 4:* Building material parameters (Table 3): only obtainable from local

170   knowledge of the materials used in the construction of the buildings.

171   *Stage 5:* Observed fluxes: to allow parameter optimisation studies. Only a few groups

172   completed this stage, so these results are not presented here.

173

174   The results from 24 modelling groups are analysed, involving 21 independent models

175   (Table 1). Alternative versions of the same model were run by the same or

176   independent modelling groups, which resulted in 32 sets of model simulations being

177   submitted for all of the four stages (see full list in *Grimmond et al.*, 2011). Each group

178   completed a survey indicating the level of complexity used for various physical

179   processes within their models. From the latter, categories of physical processes were

180   established, with classes that cover the range of complexities (*Grimmond et al.*, 2010,

181   2011). These categories were chosen to investigate the importance of various physical

182   processes that could contribute to differences in the surface energy balance between

183   the urban and rural environments. Thus every model is assigned to a class in each

184   category based on the survey information. In this study, the complexity category

185 (Grimmond et al., 2011) is not considered as the focus is to separate the specific

186 physical processes. The categories, with the number of models in each class are

187 shown in Table 4.

188

189 Comparing the mean behaviour of the models in each of the classes as a reference

190 provides a method to determine the level of complexity that gives the best

191 performance for each category. These data are analysed to address the second research

192 question, where "fit for purpose" in this study is defined as being able to accurately

193 represent the energy exchange between the urban surface and the atmosphere (i.e., the

194 net all-wave radiation, turbulent sensible and latent heat fluxes).

195

196 Furthermore, by assessing the performance of the models across the categories for all

197 classes, it is possible to identify the physical processes that have the largest impact on

198 the performance of the models, hence identifying the dominant physical processes and

199 addressing the first research question.

200

201 **2.3 Methodology**

202 Initial results from the urban model comparison (*Grimmond et al.*, 2011) ranked the

203 models and assessed the performance of the various classes within the categories

204 using standard statistical measures. Here an alternative approach to assess the models'

205 performance is used, that considers the percentage of the models' data values that are

206 within observational error (E_{obs}). This gives a measure between zero (no values within

207 observational errors) and 100% (all values within observational errors, i.e., a 'perfect'

208 model). Although this type of analysis is not strictly benchmarking, as each model is

209 not being compared to an *a priori* metric, it could be considered as being closer to the

210   benchmarking ethos as having all data points within observational errors would be a

211   stringent metric.

212

213   The observational error estimates used in this analysis are for day-time fluxes based

214   on a percentage of the observed fluxes, as suggested by *Hollinger and Richardson*

215   (2005): net all-wave radiation flux 5%, turbulent sensible heat flux 10%, latent heat

216   flux 8%, and upward components of both shortwave and long-wave radiation fluxes

217   10%. As the net storage heat flux in the observational dataset is determined as the

218   residual of the surface energy balance, its observational error is assumed to be the sum

219   of the errors for the other terms (i.e., $Q*$, Q_H and Q_E), giving 23%. The night-time

220   error estimates are assumed to be double the day-time error estimates for each of the

221   fluxes. The absolute magnitude of fluxes during this period are typically small (order

222   of (10) W m^{-2}), hence changes in the percentage of the observed flux used as the error

223   estimates are likely to be within the reporting resolution (e.g. order of (1) W m^{-2}) of

224   the observations (especially the turbulent fluxes). Whilst these error estimates may be

225   indicative rather than the actual values, the results would not substantially change the

226   analysis presented.

227

228   The analysis was undertaken for each model (k) in each class (j) within each category

229   (i) (Table 4), for each flux, over each stage within the comparison, and separately for

230   day-time and night-time. From this the percentage of data within observational error

231   ($E_{obs\ i,j}$) was determined:

232    $$E_{obs,ij} = \frac{\sum\limits_{k=1}^{n_{ij}} M_k}{n_{ij}T} \times 100\%$$ (1)

233   where M is the number of points within observational error for model (k), n is the

234   number of models and T is the number of day-time or night-time points in the time

235   series as appropriate.

236

237   **3. Results**

238   Application of eqn. 1 to the sensible, latent and net storage heat fluxes, for each class

239   and category, at Stage 1 and Stage 4 (Table 3) are shown in Figure 2. The results

240   could range between 0% (i.e., no model data points within the observations errors) to

241   100% (i.e., all model data points within observational errors). The relative changes

242   between the stages are also shown if Figure 2, i.e., for stage (s) the change relative to

243   the previous stage (s-1) given by:

244    $$E_{obs,ij}^{s} / E_{obs,ij}^{s-1} \qquad\qquad (2)$$

245   Assessment of "between stages performance" allows an emphasis of the common

246   results across all of the classes and categories. It is scaled between 0% and 100%,

247   with 50% corresponding to no change between the stages (Figure 2).

248

249   Generally the results of the analysis, consistent with *Grimmond et al.* (2011), show

250   that the skill to model latent heat fluxes is improved between stages 1 and 2. Knowing

251   the plan area vegetation fraction (provided in Stage 2) is important for modelling the

252   latent heat flux. No other stages show a general increase in model performance across

253   the classes and categories for the fluxes shown in Figure 2. For the radiation fluxes

254   (Fig. 3), the largest changes evident between Stages 3 and 4 are for the reflected

255   shortwave radiation flux and are due to the specification of the bulk albedo at the site

256   (i.e., the ratio of the reflected outgoing shortwave radiation flux from the whole urban

257   surface to the incoming shortwave radiation flux, information released at Stage 4).

258   This is also consistent with the conclusions from *Grimmond et al.* (2011).

259

260 Model performance for the outgoing long-wave radiation flux has its largest changes

261 at night-time between Stages 3 and 4 (when the 3-d site morphological information

262 (Table 3) were made available, Fig. 3). This enhanced performance at night could be

263 related to improved estimates of the sky view factor which influences radiative

264 trapping, and/or from improved estimates of the difference in nocturnal surface

265 temperatures between building roofs and those of the roads and walls of the urban

266 canyons. Improved performance is not detected in the day-time outgoing long-wave

267 radiation flux (Fig. 3), probably because of the dominance of shortwave radiation at

268 this time. These results were not identified in *Grimmond et al.* (2011) as there was no

269 separate analysis for day-time and night-time.

270

271 It is evident from Figures 2 and 3 that the performance of the models for each of the

272 fluxes does not improve consistently for each stage, as might be expected. This

273 suggests that the models are not able to correctly make use of all of the information

274 that is provided at each of the stages and hence the design of the models, and the use

275 of their specific parameters, is not necessarily correct. This is discussed further in

276 Grimmond et al. (2011).

277

278 Each model is assigned to one class for every category (Table 4). This means that a

279 model with particularly good (or poor) performance will influence the results for its

280 class in each of the categories. The implications of this are that it is not possible to

281 ensure that the good performance from a particular class within one category is not

282 actually resulting from the results of a class from a different category. This potential

283 contamination of results by categories inhibits the analysis of the dominant physical

284   processes and the suitability of the models. Both the analysis presented in *Grimmond*

285   *et al.* (2011) and that in Figures 2 and 3 have this limitation, hence we will not

286   consider further any results in Figures 2 and 3 for any specific class or category.

287   Alternatively, to address this issue of cross-contamination, we repeat the complete

288   analysis using eqn. 1 separately for each category (c), but only considering the subset

289   of models from class (*a*). Hence for each class (*j*) in category (*i*) for the analysis of

290   eqn. 1, the models used are those that are in both class (*a*) of category (*c*) and class (*j*)

291   of category (*i*), of which there are $n_\alpha = n_{ca} \cap n_{ij}$, thus:

292    $$E_{obs,caij} = \frac{\sum_{k=1}^{n_\alpha} M_k}{n_\alpha T} \times 100\% \tag{3}$$

293   This gives the equivalent of 26 versions of Figures 2 and 3 (one for each class in each

294   category); although for a given subset of models it is inevitable that some classes will

295   not have any members and hence have no data. We then apply the following equation

296   for each of the stages to determine which of the original class of models has the best

297   performance:

298    $$P_{ca} = \frac{\sum N_m}{N_{tot} - \left(\sum N_{nd}\right) - 1} \times 100\% \tag{4}$$

299   where P_{ca} is the percentage of classes in the analysis that are improved from just the

300   subset of models (compared to the analysis with the full set of models),

301    $$N_m = \sum_{k=1}^{N_{tot}} \begin{cases} 1 \text{ if } E_{obs,caij} > E_{obs,ij} \\ 0 \text{ otherwise} \end{cases} \tag{5}$$

302   is the number of classes that are improved in the analysis, N_{tot} is the total number of

303   classes ($\sum ij = 26$) and

304    $$N_{nd} = \sum_{k=1}^{N_{tot}} \begin{cases} 1 \text{ if } n_{ca} \cap n_{ij} = 0 \\ 0 \text{ otherwise} \end{cases} \tag{6}$$

13

305   is the number of classes with no data.

306

307   Hence values of P_{ca} close to 100% relate to nearly all classes in all categories being

308   improved from the physical process represented in class (a) of category (c). This

309   indicates that this process and its representation are important to model performance.

310   Whereas values close to 0% relate to almost all classes in all categories being

311   degraded, suggesting that the representation of the physical process is detrimental to

312   model performance. Values around 50% have a similar number of classes that are

313   improved and degraded, suggesting that the representation of the physical process has

314   little impact on model performance. Hence the conclusions that can be drawn from

315   this analysis are more robust than those of Figures 2 and 3, and the previous study of

316   *Grimmond et al.* (2011).

317

318   For example, with models that have an infinite number of reflections (category R,

319   class i), the median of the results over the stages give a value of 88% for the night-

320   time net storage heat flux (Fig. 4). This results from 14 of the 16 possible classes

321   containing data being improved when considering only these models, demonstrating

322   that this is important for predicting this flux. However, models that have multiple

323   reflections (category R, class m) have a value of 12.5% for the night-time net storage

324   heat flux (Fig. 4). This results from only two of the possible 16 classes containing data

325   being improved, hence showing that this is detrimental to predicting the flux.

326

327   The results of Figure 4 show that for some classes (e.g., infinite reflections; category

328   R, class I, Table 4), there are some demonstrated improvements to a flux (e.g., LW_{up})

329   which is not obviously explained by the physics (e.g., how do infinite reflections of

330   shortwave radiation improve the outgoing long-wave radiation but not the reflected

331   shortwave?). Also, there are some classes that improve one particular flux, but not

332   other fluxes. For example, models that represent the net storage heat flux as the

333   residual of the surface energy balance (category S, class r, Table 4) demonstrate a

334   clear improvement for the day-time sensible heat flux, but not for the latent or the net

335   storage heat fluxes. This could be because with such models the sensible heat flux is

336   not constrained by the energy balance giving them the freedom to enable better

337   predictions of the sensible heat flux, whilst moisture availability is still the main

338   control for the latent heat flux.

339

340   There are many such conclusions that can be drawn from Figure 4. Here the focus is

341   on results that are consistent between the fluxes, or consistent for a particular flux

342   between the day-time and night-time.

343

344   Models with a bulk representation of the albedo and emissivity (category A_E, class 1,

345   Table 4), and a bulk representation of facets and orientation (category F_O, class 1; the

346   models in these two classes were identical), demonstrate an improvement in skill

347   during the day-time for nearly all fluxes, with the exceptions of the outgoing long-

348   wave radiation which shows little change in skill and net all-wave radiation fluxes

349   with only small improvements (Fig. 4). This class of models also shows an

350   improvement in the night-time sensible and latent heat fluxes, but degradation in the

351   radiative fluxes during the night. These improved results are most likely due to the

352   ability to utilize the observed bulk albedo directly. This class of models clearly

353   delivers the largest benefits across the fluxes and indicates the most significant

354   physical process to represent is the bulk albedo for the urban surface, because the net

355   shortwave radiation dominates the surface energy balance.

356

357   Improvements to the outgoing long-wave radiation flux and the net all-wave radiation

358   flux during both day-time and night-time are obtained from models that have a single

359   layer for each element of the urban environment (i.e., roofs and either urban canyons,

360   or walls and roads separately) in the morphology category (category L, class 2, Table

361   4; Fig. 4). Improvements to the night-time sensible heat flux and net storage heat flux

362   are also obtained from this class of models, but there is no improvement to these

363   fluxes during the day-time. This neutral day-time result in the sensible and net storage

364   heat fluxes may be explained by the negative impact on the outgoing shortwave

365   radiation flux, which dominates over the long-wave radiation flux during the day-

366   time. However, these results demonstrate the importance of presenting the difference

367   in radiative surface temperatures between the roofs and the urban canyon, due to the

368   non-linear relationship between the upward long-wave radiation and the radiative

369   temperature.

370

371   When considering the way in which the models represent vegetation (category V,

372   Table 4), we find that although including vegetation (classes s and i, Table 4) does

373   generally lead to an improvement for the fluxes, these improvements are not as

374   obvious as those from the bulk albedo or the single layer urban morphology. Hence

375   although these results confirm those presented in earlier studies on the comparison

376   (*Grimmond et al.*, 2011, *Best and Grimmond*, 2013, 2014), that representing

377   vegetation gives improved results, we demonstrate that the more robust analysis

378   presented here shows that this is not the most important physical process as was

379   concluded in these earlier studies. Getting the radiative fluxes correct from the

380 shortwave via the bulk albedo and the long-wave through the urban morphology are

381 required before the vegetation can influence the partitioning of energy between the

382 sensible and latent heat fluxes.

383

384 Previous studies on the urban comparison data have also concluded that models which

385 neglect the anthropogenic heat flux (Q_F) do at least as well as the models that include

386 this flux, although they were unable to explain this result (*Grimmond et al.*, 2011,

387 *Best and Grimmond*, 2013, 2014). However, the results in Figure 4 show that

388 although the class of models that neglect the anthropogenic heat flux (category A_N,

389 class n, Table 4) do improve some of the fluxes, the improvements are not consistent

390 over all of the fluxes. Moreover, this class of models within the anthropogenic heat

391 flux category is not always the one that delivers the best results. Hence we can

392 conclude that although the models that neglect the anthropogenic heat flux do show

393 some improved results, we cannot make any significant statements about the classes

394 within this category.

395

396 4. **Conclusions**

397 Prior conclusions from the ULSM comparison with daily (24 h) and seasonal analysis

398 include that: representation of vegetation is critical to model performance (*Grimmond*

399 *et al.,* 2011, *Best and Grimmond*, 2013), along with the associated initial soil moisture

400 (*Best and Grimmond*, 2014), and the bulk albedo is also important (Grimmond et al.,

401 2011). Notably, neglecting the distinctive urban anthropogenic heat flux was not

402 found to penalize performance (albeit in the suburban area the value is small) (*Best*

403 *and Grimmond*, 2013). However, this new analysis considering diurnal performance

404 (day, night) enables us to conclude that nocturnal radiative processes also benefit from

405   accounting for the enhanced long-wave trapping that occurs within urban areas.

406   Separating the radiative processes of the roof and the urban canyon is beneficial.

407

408   More critically, the more robust analysis presented here enables identification of a re-

409   prioritisation of the key physical processes: firstly, ensuring the use of the correct bulk

410   albedo for the urban surface; secondly, the outgoing long-wave radiative fluxes with

411   the representation of morphology separated into roofs and urban canyons; and thirdly,

412   the inclusion of vegetation. The implications of the bulk albedo is important for

413   observations as the temporal resolution of satellite estimates mean they will not

414   provide the variations by time of day that are observed (e.g., Christen and Voogt,

415   2004, Grimmond et al. 2004, Kotthaus and Grimmond 2014).

416

417   The current results for anthropogenic heat flux are consistent with the earlier studies:

418   that neglect of the relatively small magnitude flux at this site (study period mean =

419   ~17 W m^{-2}) is reasonable. This conclusion could well be different for urban

420   environments where this is a more significant term in the surface energy balance. The

421   flux is expected to be larger in other areas of Melbourne (e.g., as suggested from

422   analysis using the model of Lindberg et al. 2013) and for urban areas elsewhere. We

423   therefore recommend that future model comparisons ideally include areas of cities

424   with larger anthropogenic heat fluxes.

425

426   Thus to answer the three over-arching research questions of the urban model

427   comparison:

428   (i) The dominant physical processes in the urban environment that models need to be

429   able to simulate, in order, are; changes to the bulk albedo of the surface that result

18

430   from building materials and also shortwave trapping from the canyon geometry; the

431   reduction in outgoing long-wave radiation from the street canyon due to a reduced

432   sky view factor and the contrast between this and the roofs that see a full sky view;

433   and the evaporation from vegetation.

434   (ii)   For the current generation of ULSMs, the ability to utilize a bulk surface

435   albedo (category A_E, class 1, Table 4) and to be able to distinguish between the

436   roofs of buildings and the urban canyons (category L, class 2), and to have a

437   representation of vegetation (category V, classes s, i), results in the best

438   performance.

439   (iii)   The key parameters for ULSMs are the bulk surface albedo (information given

440   for Stage 4 influencing the upward shortwave radiation flux), the height to width

441   ratio of the urban canyons and the fraction of building roofs to the urban canyons

442   (information given for Stage 3 influencing the upward long-wave radiation flux),

443   and the vegetation fraction (information given for Stage 2 influencing the sensible

444   and latent heat fluxes).

445

446   The results, from this and the previous studies on the ULSM comparison, all suggest

447   that a simple representation for most of the physical categories is sufficient for this

448   type of application, i.e., determination of local scale fluxes (e.g. for use in the

449   coupling to an atmospheric model). The prior categorization of the models

450   (Grimmond et al., 2011, Best and Grimmond, 2013) into (simple, medium and

451   complex) complexity classes based upon the number of physical categories treated as

452   simple by a model demonstrated that the simple models performed best. This relative

453   success of simple models suggests that for simulating local scale fluxes, more

454   complex schemes deliver little additional benefit. Furthermore, the reduced parameter

455   requirements for simple schemes are advantageous for large scale applications, such

456   as global or regional scale modelling. However, it cannot be expected that this

457   conclusion would also hold for other applications, e.g., atmospheric dispersion within

458   street canyons of a specific city, as the simple models do not present some of the basic

459   physical requirements for such applications. Thus the requirement for the

460   development of more complex ULSMs does remain.

461

462   The implications of this study go beyond the urban environment. In general, we need

463   to balance the requirement for complexity within models against what is actually

464   required for a model to be fit for purpose. Hence new and more complex processes

465   should not be included in models unless it can be demonstrated that they are required.

466   In addition, consideration needs to be given to the availability of information to

467   specify parameters within complex models, and if such complexity can be justified

468   given the uncertainty range for the parameters. Also, the type of analysis used here

469   could be applied to any comparison study to ensure that the results are robust and not

470   contaminated by physical processes not being directly considered.

471

472   These key conclusions are based on the single site observational dataset of less than

473   18 months. This suburban site of low density housing, is typical of extensive areas in

474   North America, Europe and Australasia. Hence we might expect the results from this

475   study to be valid over a reasonable range of cities. However, most urban environments

476   have a range of zones (e.g. *Ellefsen*, 1991, *Grimmond and Souch*, 1994, *Stewart and*

477   *Oke*, 2012) with very different characteristics. So to test if the results presented here

478   are robust for other cities, similar "experiments" are required for additional sites with

479 differing climates and urban characteristics. Hence we recommend that further model

480 comparison projects are required for the urban community.

481

482 Despite these limitations, the results have implications for future development of

483 ULSMs and for the types of data that need to be collected in future urban

484 measurement campaigns (e.g., soil moisture, given its impact to limit transpiration and

485 the long timescales required for model spin-up, along with the conclusion that the

486 fraction of vegetation is important for urban areas) and/or the parameters that should

487 be collated systematically for cities around the world (e.g., *Ching et al.*, 2009, *Loridan*

488 *and Grimmond*, 2012, *Stewart and Oke*, 2012, *Ching*, 2013, *Faroux et al.*, 2013).

489

501

502

503

504 **References**

505

506 Best, M.J. (2005), Representing urban areas within operational numerical weather

507   prediction models. *Boundary-Layer Meteorol*, 114: 91–109.

508 Best M.J., C.S.B. Grimmond, M.G. Villani (2006), Evaluation of the urban tile in

509   MOSES using surface energy balance observations. *Boundary- Layer Meteorol,*

510   118: 503–525.

511 Best M.J., M. Pryor , D.B. Clark, G.G. Rooney, R.H.L. Essery, C.B. Ménard, J.M.

512   Edwards, M.A. Hendry, A. Porson, N. Gedney, L.M. Mercado, S. Sitch, E. Blyth,

513   O. Boucher, P.M. Cox, C.S.B. Grimmond, R.J. Harding (2011) The Joint UK Land

514   Environment Simulator (JULES), Model description – Part 1: Energy and water

515   fluxes. *Geosci Model Dev,* 4: 677-699

516 Best, M.J., C.S.B. Grimmond (2013), Analysis of the seasonal cycle within the first

517   international urban land surface model comparison, *Boundary-Layer Meteorol.,*

518   146, 421-446, doi: 10.1007/s10546-012-9769-7.

519 Best, M.J., C.S.B. Grimmond (2014), Importance of initial state and atmospheric

520   conditions for urban land surface models performance, *Urban Climate*. In press,

521   doi:10.1016/j.uclim.2013.10.006

522 Bohnenstengel, S.I., S. Evans, P.A. Clark, S.E. Belcher (2011), Simulations of the

523   London urban heat island, *Q. J. R. Meteorol. Soc.,* 137, 1625-1640, doi:

524   10.1002/qj.855.

525 Bornstein, R., Q. Lin (2000), Urban heat islands and summertime convection

526   thunderstorms in Atlanta: three case studies, *Atmos. Environ.*, 34, 507-516,

527   doi:10.1016/S1352-2310(99)00374-X.

528   Brunner, D., G. Lemoire, L. Bruzzone, H. Greidonus (2010), Building height retrieval

529   from VHR SAR imagery based on an iterative simulation and matching technique.

530   *IEEE Transactions on Geoscience and Remote Sensing*, 48, No.3,

531   doi:10.1109/TGRS.2009.2031910.

532   Chaxel, E., J.-P. Chollet (2009), Ozone production from Grenoble city during the

533   August 2003 heat wave, *Atmos. Environ.*, 43, 4784-4792,

534   doi:10.1016/j.atmosenv.2008.10.054.

535   Chen F., H. Kusaka, M. Tewari, J. Bao, H. Hirakuchi (2004), Utilizing the coupled

536   WRF/LSM/Urban modeling system with detailed urban classification to simulate

537   the urban heat island phenomena over the Greater Houston area. *Fifth Symposium*

538   *on the Urban Environment*, CD-ROM. 9.11. Amer. Meteor. Soc., Vancouver, BC,

539   Canada.

540   Ching, J.K.S. (2013), A perspective on urban canopy layer modelling for weather,

541   climate and air quality applications, *Urban Climate*, 3, 13-39.

542   Ching, J., M. Brown, S. Burian, F. Chen, R. Cionco, A. Hanna, T. Hultgren, T.

543   McPherson, D. Sailor, H. Taha, D. Williams (2009), National urban database and

544   access portal tool, *Bull. American Meteorol. Soc.,* 90, 1157-1168,

545   doi:10.1175/2009BAMS2675.1.

546   Christen, A., R. Voogt (2004), Energy and radiation balance of a central European

547   city, Int. J. Climatol., 24, 1395-1421, doi:10.1002/joc.1074.

548   Coutts, A.M., J. Beringer, N.J. Tapper (2007a), Characteristics influencing the

549   variability of urban CO_2 fluxes in Melbourne, Australia, *Atmos. Environ.,* 41, 51-

550   62.

551   Coutts, A.M., J. Beringer, N.J. Tapper (2007b) Impact of increasing urban density on

552   local climate: spatial and temporal variations in the surface energy balance in

553   Melbourne, Australia, *J. Appl. Meteorol.*, 47, 477-493.

554   Dandou A., M. Tombrou, E. Akylas, N. Soulakellis, E. Bossioli (2005), Development

555   and evaluation of an urban parameterization scheme in the Penn State/NCAR

556   Mesoscale model (MM5). *J Geophys Res,* 110: D10102.

557   doi:10.1029/2004JD005192.

558   Dupont S., P.G. Mestayer (2006), Parameterisation of the urban energy budget with

559   the submesoscale soil model. *J Appl Meteorol Climatol,* 45: 1744–1765.

560   Dupont S., P.G. Mestayer, E. Guilloteau, E. Berthier, H. Andrieu (2006),

561   Parameterisation of the urban water budget with the submesoscale soil model. *J*

562   *Appl Meteorol Climatol*, 45: 624–648.

563   Ellefsen, R. (1991), Mapping and measuring buildings in the canopy boundary layer

564   in ten U.S. cities. Energy and Buildings, 16, 1025-1049.

565   Essery R.L.H., M.J. Best, R.A. Betts, P.M. Cox, C.M. Taylor (2003), Explicit

566   representation of subgrid heterogeneity in a GCM land surface scheme. *J*

567   *Hydrometeorol*, 4: 530–543.

568   Faroux, S., A. T. Kaptue Tchuente, J.-L. Roujean, V. Masson, E. Martin, P. Le

569   Moigne (2013), ECOCLIMAP-II/Europe: a twofold database of ecosystems and

570   surface parameters at 1 km resolution based on satellite information for use in land

571   surface, meteorological and climate models, *Geosci. Model Dev.*, 6, 563-582,

572   doi:10.5194/gmd-6-563-2013.

573   Fortuniak, K. (2003), A slab surface energy balance model (SUEB) and its application

574   to the study on the role of roughness length in forming an urban heat island. *Acta*

575   *Universitatis Wratislaviensis,* 2542, 368-377.

576   Fortuniak K., B. Offerle, C.S.B. Grimmond (2004), Slab surface energy balance

577   scheme and its application to parameterisation of the energy fluxes on urban areas.

578   *NATO ASI, Kiev, Ukraine*; 82–83. Available from: www.met.rdg.ac.uk/urb

579   met/NATO ASI/talks.html (Last accessed 4–15 May 2010).

580   Fortuniak K., B. Offerle, C.S.B. Grimmond (2005), Application of a slab surface

581   energy balance model to determine surface parameters for urban areas. *Lund*

582   *Electronic Reports in Physical Geography,* 5: 90–91.

583   Francis, R.A., M.A. Chadwick. (2013), Urban Ecosystems: Understanding the Human

584   Environment. *Routledge*, 220pp.

585   Goodwin, N.R., N.C. Coops, T.R. Tooke, A. Christen, J.A. Voogt (2009),

586   Characterizing urban surface cover and structure with airborne lidar technology.

587   *Can. J. Remote Sens.*, 35, 297-309.

588   Grimmond, C.S.B., C. Souch (1994), Surface description for urban climate studies: a

589   GIS based methodology. *Geocarto. International*, 9, 47-59.

590   Grimmond C.S.B., T.R. Oke (2002), Turbulent heat fluxes in urban areas:

591   observations and local-scale urban meteorological parameterization scheme

592   (LUMPS). *J Appl Meteorol,* 41: 792–810.

593   Grimmond, C.S.B., J.A. Salmond, T.R. Oke, B. Offerle, A. Lemonsu (2004), Flux and

594   turbulence measurements at a densely built-up site in Marseille: Heat, mass (water

595   and carbon dioxide), and momentum. *J. Geophys. Res. Atmos.*, 109, D24101,

596   doi:10.1029/2004jd004936

597   Grimmond, C.S.B., M. Blackett, M.J. Best, J. Barlow, J.-J. Baik, S.E. Belcher, S.I.

598   Bohnenstengel, I. Calmet, F. Chen, A. Dandou, K. Fortuniak, M.L. Gouvea , R.

599   Hamdi, M. Hendry, T. Kawai, Y. Kawamoto, H. Kondo, E.S. Krayenhoff, S.-H.

600   Lee, T. Loridan, A. Martilli, V. Masson, S. Miao, K. Oleson, G. Pigeon, A.

601   Porson, Y.-H. Ryu, F. Salamanca, L. Shashua-Bar, G.-J. Steeneveld, M. Trombou,

602   J. Voogt, D. Young, N. Zhang (2010), The international urban energy balance

603   models comparison project: first results from phase 1, *J. Appl. Meteorol.*

604   *Climatol.*, 49, 1268-1292, doi: 10.1175/2010JAMC2354.1.

605 Grimmond, C.S.B., M. Blackett, M.J. Best, J.-J. Baik, S.E. Belcher, J. Beringer, S.I.

606   Bohnenstengel, I. Calmet, F. Chen, A. Coutts, A. Dandou, K. Fortuniak, M.L.

607   Gouvea , R. Hamdi, M. Hendry, M. Kanda, T. Kawai, Y. Kawamoto, H. Kondo,

608   E.S. Krayenhoff, S.-H. Lee, T. Loridan, A. Martilli, V. Masson, S. Miao, K.

609   Oleson, R. Ooka, G. Pigeon, A. Porson, Y.-H. Ryu, F. Salamanca, G.-J.

610   Steeneveld, M. Trombou, J. Voogt, D. Young, N. Zhang (2011), Initial results

611   from phase 2 of the international urban energy balance model comparison, *Int. J.*

612   *Climatol.*, 30, 244-272, doi:10.1002/joc.2227.

613 Hamdi, R., V. Masson (2008), Inclusion of a drag approach in the Town Energy

614   Balance (TEB) scheme: offline 1-D evaluation in a street canyon, *J. Appl.*

615   *Meteorol. Climatol.*, 47, 2627-2644.

616 Harman I.N., M.J. Best, S.E. Belcher (2004a), Radiative exchange in an urban street

617   canyon. *Boundary-Layer Meteorol,* 110: 301–316.

618 Harman I.N., J.F. Barlow, S.E. Belcher (2004b), Scalar fluxes from urban street

619   canyons. Part II: model. *Boundary-Layer Meteorol,* 113: 387–410.

620 Harman I.N., S.E. Belcher (2006), The surface energy balance and boundary layer

621   over urban street canyons. *Q J R Meteorol Soc,* 132: 2749–2768.

622 Herbst, J., K. Mason, R.W. Byard, J.D. Bilbert, C. Charlwood, K.J. Heath, C.

623   Winskog, N.E.I. Langlois (2014), Heat-related deaths in Adelaide, South Australia:

624   Review of the literature and case findings – An Australian perspective, *J. Forensic*

625   *and Legal Medicine*, 22, 73-78. doi:10.1016/j.jflm.2013.12.018

626   Hollinger, D.Y., A.D. Richardson (2005), Uncertainty in eddy covariance

627   measurements and its application to physiological models, *Tree Physiol.*, 25, 873-

628   885.

629   IPCC (2014), Working Group I Contribution to the IPCC Fifth Assessment Report

630   (AR5), Climate Change 2013: The Physical Science Basis. *Intergovernmental*

631   *Panel on Climate Change*, Geneva, Switzerland.

632   Kanda M., T. Kawai, M. Kanega, R. Moriwaki, K. Narita, A. Hagishima (2005a), A

633   simple energy balance model for regular building arrays. *Boundary-Layer*

634   *Meteorol*, 116: 423–443.

635   Kanda M., T. Kawai, K. Nakagawa (2005b), A simple theoretical radiation scheme

636   for regular building arrays. *Boundary-Layer Meteorol*, 114: 71–90.

637   Kawai T., M. Kanda, K. Narita, A. Hagishima (2007), Validation of a numerical

638   model for urban energy-exchange using outdoor scalemodel measurements. *Int J*

639   *Climatol,* 27: 1931–1942.

640   Kawai T., M.K. Ridwan, M. Kanda (2009), Evaluation of the simple urban energy

641   balance model using 1-yr flux observations at two cities. *J Appl Meteorol Climatol,*

642   48: 693–715.

643   Kawamoto Y., R. Ooka (2006), Analysis of the radiation field at pedestrian level

644   using a meso-scale meteorological model incorporating the urban canopy model. In

645   *ICUC-6, Göteborg, Sweden*, 12–16 June 2006.

646   Kawamoto Y., R. Ooka (2009a), Accuracy validation of urban climate analysis model

647   using MM5 incorporating a multi-layer urban canopy model. In *ICUC-7,*

648   *Yokohama, Japan*, 28 June–3 July 2009.

649   Kawamoto Y., R. Ooka (2009b) Development of urban climate analysis model using

650   MM5 Part 2 – incorporating an urban canopy model to represent the effect of

651   buildings. *J Environ Eng* (Transactions of AIJ) 74(642): 1009–1018 (in Japanese).

652   Kondo H., F.H. Liu (1998), A study on the urban thermal environment obtained

653   through a one-dimensional urban canopy model, *J Jpn Soc Atmos Environ*. 33,

654   179-192 (in Japanese)

655   Kondo H., Y. Genchi, Y. Kikegawa, Y. Ohashi, H. Yoshikado, H. Komiyama (2005),

656   Development of a multi-layer urban canopy model for the analysis of energy

657   consumption in a big city: structure of the urban canopy model and its basic

658   performance. *Boundary-Layer Meteorol,* 116: 395–421.

659   Kotthaus, S., C.S.B. Grimmond (2013), Energy exchange in a dense urban

660   environment - Part II: Impact of spatial heterogeneity of the surface, Urban

661   Climate, hppt://dx.doi/org/10.1016/j.uclim.2013.10.001.

662   Krayenhoff, E.S., J.A. Voogt (2007), A microscale three-dimensional urban energy

663   balance model for studying surface temperatures, *Boundary-Layer Meteorol.*, 123,

664   433-461.

665   Kusaka, H., H. Kondo, Y. Kikegawa, F. Kimura (2001), A simple singlelayer urban

666   canopy model for atmospheric models: comparison with multi-layer and slab

667   models, *Boundary-Layer Meteorol.*, 101, 329-358.

668   Laaidi, K., A. Zeghnoun, B. Dousset, P. Bretin, S. Vandentorren, E. Giraudet, P.

669   Beaudeau (2012), The impact of heat islands on mortality in Paris during the

670   August 2003 heat wave, *Environ. Health Perspectives*, 120, 254-259,

671   doi:10.1289/ehp.1103532.

672   Lee, S.-H., S.-U. Park (2008), A vegetated urban canopy model for meteorological

673   and environmental modelling, *Boundary-Layer Meteorol.*, 126, 73-102.

674   Lemonsu A., C.S.B. Grimmond, V. Masson (2004), Modelling the surface energy

675   balance of an old Mediterranean city core. *J Appl Meteorol,* 43: 312–327.

676   Li, C.B., J.J. Zhou, Y.J. Cao, J. Zhong, Y. Liu, C.Q. Kang, Y. Tan (2014), Interaction

677   between urban microclimate and electric air-conditioning energy consumption

678   during high temperature season, *Applied Energy*, 117, 149-156,

679   doi:10.1016/j.apenergy.2013.11.057.

680   Lindberg, F., C.S.B. Grimmond (2011), Nature of vegetation and building

681   morphology characteristics across a city: Influence on shadow patterns and mean

682   radiant temperatures in London. *Urban Ecosyst.*, 14, 617-623. doi:10.1007/s11252-

683   011-0184-5.

684   Lindberg, F., C.S.B. Grimmond, N. Yogeswaran, S. Kotthaus, L. Alen (2013),

685   Impacts of city changes and weather on anthropogenic heat flux in Europe 1995-

686   2015, *Urban Climate*, 4, 1-15. http://dx.doi.org/10.1016/j.uclim.2013.03.002.

687   Loridan T., C.S.B. Grimmond, S. Grossman-Clarke, F. Chen, M. Tewari, K.

688   Manning, A. Martilli, H Kusaka, M. Best (2010), Trade-offs and responsiveness of

689   the single-layer urban canopy parameterization in WRF: an offline evaluation

690   using the MOSCEM optimization algorithm and field observations. *Q J R Meteorol*

691   *Soc,* 136: 997–1019. doi:10.1002/qj.614.

692   Loridan T., C.S.B. Grimmond, B.D. Offerle, D.T. Young, T. Smith, L. Järvi, F.

693   Lindberg (2011), Local-scale urban meteorological parameterization scheme

694   (LUMPS): Longwave radiation parameterization and seasonality-related

695   developments. *J Appl Meteorol Climatol,* 50: 185-202.

696   doi:10.1175/2010JAMC2474.1

697   Loridan, T., C.S.B. Grimmond (2012), Multi-site evaluation of an urban land-surface

698   model: intra-urban heterogeneity, seasonality, and parameter complexity

699   requirements. *Q. J. R. Meteorol. Soc.*, 138, 1094-1113, doi:10.1002/qj.963.

700   Luhar, A.K., M. Thatcher, P.J. Hurley (2014) Evaluating a building-averaged urban

701   surface scheme in an operational mesoscale model for flow and dispersion, *Atmos.*

702   *Environ.*, 88, 47-58, doi:10.1016/j.atmosenv.2014.01.059.

703   Martilli A., A. Clappier, M.W. Rotach (2002), An urban surface exchange

704   parameterisation for mesoscale models. *Boundary-Layer Meteorol,* 104: 261–304.

705   Masson V. (2000) A physically-based scheme for the urban energy budget in

706   atmospheric models. *Boundary-Layer Meteorol,* 41: 1011–1026.

707   Masson V., C.S.B. Grimmond, T.R. Oke (2002), Evaluation of the Town Energy

708   Balance (TEB) scheme with direct measurements from dry districts in two cities. *J*

709   *Appl Meteorol,* 41: 1011–1026.

710   McCarthy, M.P., M.J. Best, R.A. Betts (2010), Climate change in cities due to global

711   warming and urban effects, *Geophys. Res. Letters*, 37, L09705,

712   doi:10.1029/2010GL042845.

713   Offerle B., C.S.B. Grimmond, T.R. Oke (2003), Parameterization of net all-wave

714   radiation for urban areas. *J Appl Meteorol,* 42: 1157–1173.

715   Oke TR (2006) Towards better scientific communication in urban climate. Theor.

716   Appl. Climatol. 84: 179-190. doi: 10.1007/s00704-005-0153-0

717   Oleson, K.W., G.B. Bonan, J. Feddema, M. Vertenstein, C.S.B. Grimmond (2008a),

718   An urban parameterization for a global climate model: 1. Formulation and

719   evaluation for two cities, *J. Appl. Meteorol. Climatol.*, 47, 1038-1060.

720   Oleson K.W., G.B. Bonan, J. Feddema, M. Vertenstein. 2008b. An urban

721   parameterization for a global climate model: 2. Sensitivity to input parameters and

722   the simulated heat island in offline simulations. *J Appl Meteorol Climatol*, 47:

723   1061–1076.

724   Oleson, K.W., G.B. Bonan, J. Feddema, T. Jackson (2011), An examination of urban

725   heat island characteristics in a global climate model, *Int. J. Climatol.*, 31, 1848-

726   1865. doi:10.1002/joc.2201.

727   Pickett, S.T.A., M.L. Cadenasso, J.M. Grove, P.M. Groffman, L.E. Band, C.G. Boone,

728   W.R. Burch, Jr., C.S.B. Grimmond, J. Hom, J.C. Jenkins, N.L. Law, C.H. Nilon,

729   R.V. Pouyat, K. Szlavecz, P.S. Warren, M.A. Wilson. 2008. Beyond urban

730   legends: an emerging framework of urban ecology, as illustrated by the Baltimore

731   Ecosystem Study. *BioScience*. 58(2):139-150

732   Pigeon G., M.A. Moscicki, J.A. Voogt, V. Masson (2008), Simulation of fall and

733   winter surface energy balance over a dense urban area using the TEB scheme.

734   *Meteorol Atmos Phys,* 102: 159–171.

735   Porson A., P.A. Clark, I.N. Harman, M.J. Best, S.E. Belcher (2010), Implementation

736   of a new urban energy budget scheme in the MetUM. Part II. Validation against

737   observations and model intercomparison. *Q J R Meteorol Soc,* 136: 1530-1542.

738   Radhi, H., S. Sharples (2013), Quantifying the domestic electricity consumption for

739   air-conditioning due to urban heat islands in hot arid regions, *Applied Energy*, 112,

740   371-380, doi:10.1016/j.apenergy.2013.06.013.

741   Ryu Y.-H., J.-J. Baik, S.-H. Lee (2011), A new single-layer urban canopy model for

742   use in mesoscale atmospheric models. *J Appl Meteorol Climatol,* 50: 1773-1794.

743   doi: 10.1175/2011JAMC2665.1

744   Ryu, Y.-H., J.J. Baik (2013), Effects of anthropogenic heat on ozone air quality in a

745   megacity. *Atmos. Environ.*, 80, 20-30, doi:10.1016/j.atmosenv.2013.07.053.

746   Saha, M.V., R.E. Davis, D.M. Hondula (2014), Mortality displacement as a function

747   of heat event strength in 7 US cities, *American J. Epidemiology*, 179, 467-474,

748   doi:10.1093/aje/kwt264.

749   Sailor, D.J., L. Lu (2004), A top-down methodology for developing diurnal and

750   seasonal anthropogenic heating profiles for urban areas. Atmos. Environ., 38,

751   2737-2748, doi:10.1016/j.atmosenv.2004.01.034.

752   Salamanca F., E.S. Krayenhoff, A. Martilli (2009), On the derivation of material

753   thermal properties representative of heterogeneous urban neighbourhoods. *J Appl*

754   *Meteorol Climatol,* 48: 1725–1732.

755   Salamanca F., A. Krpo, A. Martilli, A. Clappier (2010), A new building energy model

756   coupled with an urban canopy parameterization for urban climate simulations –

757   part I. Formulation, verification, and sensitivity analysis of the model. *Theor Appl*

758   *Climatol*, 99: 345-356. doi: 10.1007/s00704-009-0142-9.

759   Salamanca F., A. Martilli (2010), A new Building Energy Model coupled with an

760   Urban Canopy Parameterization for urban climate simulations – part II. Validation

761   with one dimension off-line simulations. *Theor Appl Climatol*, 99: 345–356.

762   Sarrat, C., A. Lemonsu, V. Masson, D. Guedalla (2006), Impact of urban heat island

763   on regional atmospheric pollution, *Atmos. Environ.*, 40, 1743-1758,

764   doi:10.1016/j.atmosenv.2005.11.037.

765   Stewart, I.D., T.R. Oke (2012), Local climate zones for urban temperature studies,

766   *Bull. American Meteorol. Soc.*, 93, 1879-1900, doi:10.1175/BAMS-D-11-00019.1.

767

Figure captions

769 Figure 1: Conceptual figure of how surface energy balance exchanges are included in

770 urban land surface models of different complexity. Note individual models have

771 simple and complex features (Grimmond et al., 2011).

772

773 Figure 2: For each flux and physical category class (Table 4), the percentage of

774 modelled data points within the specified observational errors (eqn. 1) for Stages 1

775 and 4 (grey) plus the change relative to the previous stage (eqn. 2; scaled between -

776 100% and 100%, shown by the horizontal dotted lines). Blue shading indicates an

777 improvement (> 0) and red degradation (< 0). Results are shown for day and night-

778 time (with day defined as incoming solar radiation flux greater than 0 W m^{-2}). Codes

779 definition for the physical categories and component classes (used in the x-axis) are

780 given in Table 4

781

782 Figure 3: As for Fig. 2, but for the radiative fluxes

783

784 Figure 4: The subset of models within a class of a category improved compared to all

785 models (P_{ca}, eqn. 4) ranked according to the median over the stages (for each flux, by

786 time of day (as for Fig. 2)). Shading shows the range of results over the stages, with

787 the individual results shown as horizontal lines within this. The colouring emphasises

788 the values of the median over the stages, with 100% corresponding to all classes

789 improved, 0% all classes degraded and 50% no change. Note X-axis code (Table 4)

790 order changes between subplots because of ranking (Colour text is to aid differences

791 to be noted).

792

793

Table 1: Urban land surface models (ULSMs) used to obtain results that are analysed

795   here. See Grimmond et al. (2010, 2011) for more details of the different model

796   versions and the number of groups that submitted simulations to the urban model

797   comparison.

798

Model name	References
Building effect parameterization (BEP)	Martilli et al. (2002) Salamanca et al. (2009, 2010) ; Salamanca and Martilli (2010)
Community Land Model – urban (CLM-urban)	Oleson et al. (2008a, 2008b)
Institute of Industrial Science urban canopy model	Kawamoto and Ooka (2006, 2009a, 2009b)
Joint UK land environment simulator (JULES)	Essery et al. (2003); Best (2005); Best et al. (2006); Best et al. (2011)
Local-scale urban meteorological parameterization scheme (LUMPS)	Grimmond and Oke (2002); Offerle et al. (2003); Loridan et al. (2011)
Met Office Reading urban surface exchange scheme (MORUSES)	Harman et al. (2004a, 2004b); Harman and Belcher (2006), Porson et al. (2010)
Multi-layer urban canopy model	Kondo and Liu (1998); Kondo et al. (2005)
National and Kapodistrian University of Athens model	Dandou et al. (2005)
Noah land surface model/single-layer urban canopy model	Kusaka et al. (2001); Chen et al. (2004); Loridan et al. (2010)
Seoul National University urban canopy model	Ryu et al. (2011)
Simple urban energy balance model for mesoscale simulation	Kanda et al. (2005a, 2005b); Kawai et al. (2007, 2009)
Slab urban energy balance model	Fortuniak (2003); Fortuniak et al. (2004, 2005)
Soil model for submesoscales (urbanized)	Duport and Mestayer (2006); Dupont et al. (2006)
Temperatures of urban facets (TUF)	Krayenhoff and Voogt (2007)
Town energy balance (TEB)	Masson (2000); Masson et al. (2002); Lemonsu et al. (2004); Pigeon et al. (2008), Hamdi and Masson (2008)
Vegetated urban canopy model	Lee and Park (2008)

799

800   Table 2: Methods used to obtain the observed fluxes used for comparison with the
801   ULSM. Sources: *Coutts et al.*, (2007a, 2007b). Height of observation for all fluxes: 40
802   m.
803

Flux	Instrument / Method	Sampling frequency (Hz.)	Averaging period (min)
SW_{up} LW_{up} Q^*	Kipp and Zonen CM7B and CG4 radiometers	1	30
Q_H	CSI CSAT 3D sonic anemometer	10	30
Q_E	CSI CSAT 3D sonic anemometer CSI Krypton hygrometer (Aug 2003 – Feb 2004), LiCOR LI7500 open-path infrared gas analyser (remaining period)	10	30
ΔQ_S	Residual of the surface energy balance	N/A	30
Q_F	Calculated (*Sailor and Lu* ,2004) : *Vehicles*: Numbers from survey (Nov. 2002 – Oct 2003 *Building* sector: 30 min electricity and daily natural gas statistics *Human metabolism*: Night, day and transition period metabolic rates, with population density statistics	N/A	Average monthly diurnal cycle at 30 min. resolution

804

805

806   Table 3. Information released at each stage of the comparison

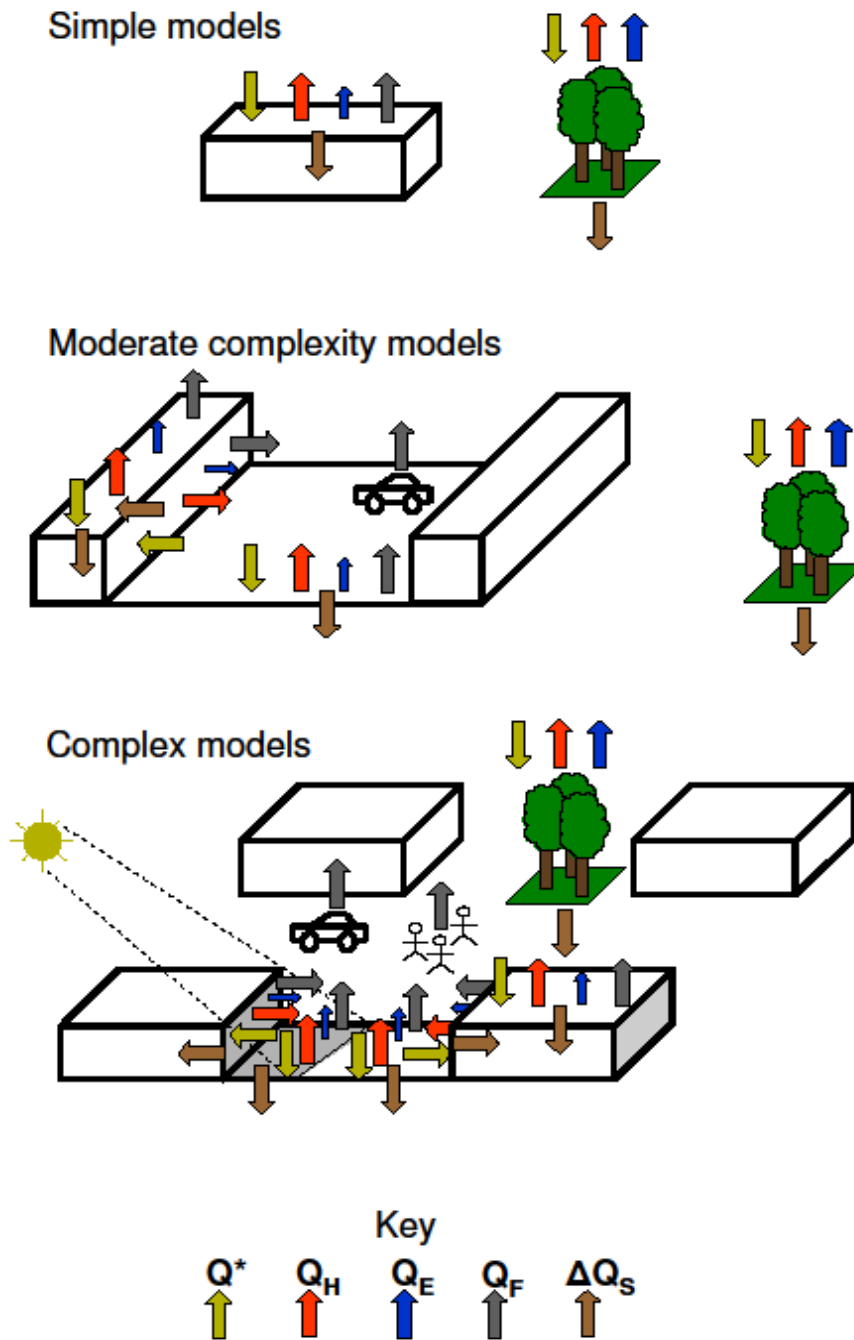| Stage	Information released
1	Atmospheric forcing data only
	(incoming shortwave radiation, incoming long-wave radiation, precipitation, atmospheric wind speed, temperature, specific humidity and surface pressure)
2	Vegetation and built fractions
3	Morphology
	(Building heights, height-to-width ratio, mean wall to plan area ratio, fraction of surface covered by buildings, concrete, road,)
4	Specific information on building materials
	(e.g., albedo and thermal properties of wall, road, roof)
5	Observed fluxes for parameter optimisation
	(Not considered in this study)

807

808 Table 4: Classes and physical categories used in the analysis of the urban comparison

809 results, including the number of models in each class (see *Grimmond et al.*, 2010,

810 2011 for more details). Colours are used on the plots to aid comparison.

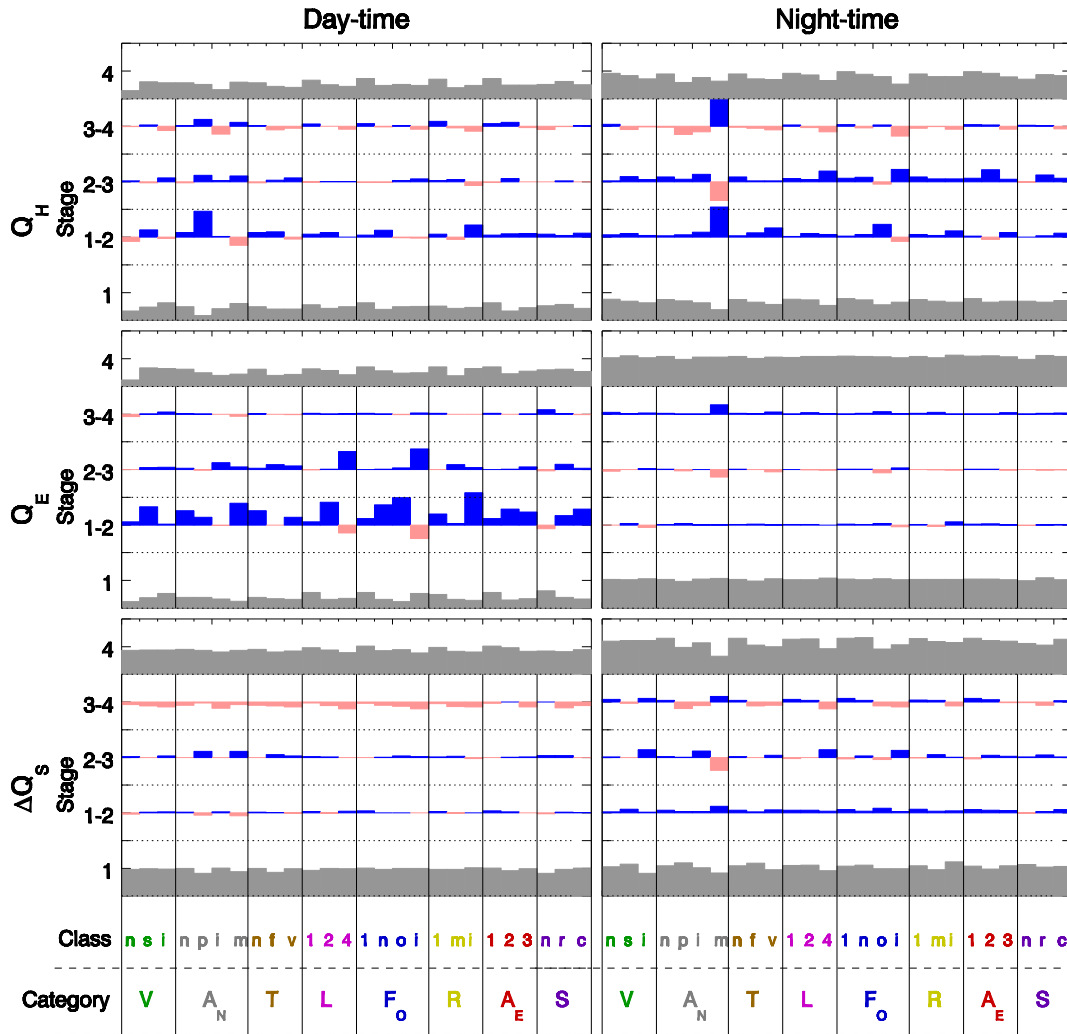| Category	Class			
Vegetation (V)	None (n)	Separate tile (s)	Integrated (i)	
No. of models	8	19	5	
Anthropogenic heat flux (A_N)	None (n)	Prescribed flux (p)	Internal building temperature (i)	Modelled (m)
No. of models	22	2	6	2
Temporal variation of the anthropogenic heat flux (T)	None (i.e., no flux) (n)	Fixed (i.e., time invariant flux) (f)	Variable (i.e., time varying flux) (v)	
No. of models	22	3	7	
Urban morphology (L)	Bulk (1)	Single layer (2)	Multiple layer (4)	
No. of models	6	20	6	
Facets & orientation (F_O)	Bulk (1)	Roof, walls, road without orientation (n)	Roof, walls, road with orientation, no intersections (o)	Roof, walls, road with orientation and intersections (i)
No. of models	5	17	6	4
Reflections (R)	Single (1)	Multiple (m)	Infinite (i)	
No. of models	11	13	8	
Albedo, emissivity (A_E)	Bulk (1)	Two facet (2)	Three facet (3)	
No. of models	5	4	23	
Net storage heat flux (S)	Net all wave radiation (n)	Surface energy balance residual (r)	Conduction equation (c)	
No. of models	3	6	23	

811

812

Figure 1: Conceptual figure of how surface energy balance exchanges are included in urban land surface models of different complexity. Note individual models have simple and complex features (Grimmond et al., 2011).

Figure 2: For each flux and physical category class (Table 4), the percentage of modelled data points within the specified observational errors (eqn. 1) for Stages 1 and 4 (grey) plus the change relative to the previous stage (eqn. 2; scaled between -100% and 100%, shown by the horizontal dotted lines). Blue shading indicates an improvement (> 0) and red degradation (< 0). Results are shown for day and night-time (with day defined as incoming solar radiation flux greater than 0 W m^{-2}). Codes definition for the physical categories and component classes (used in the x-axis) are given in Table 4.
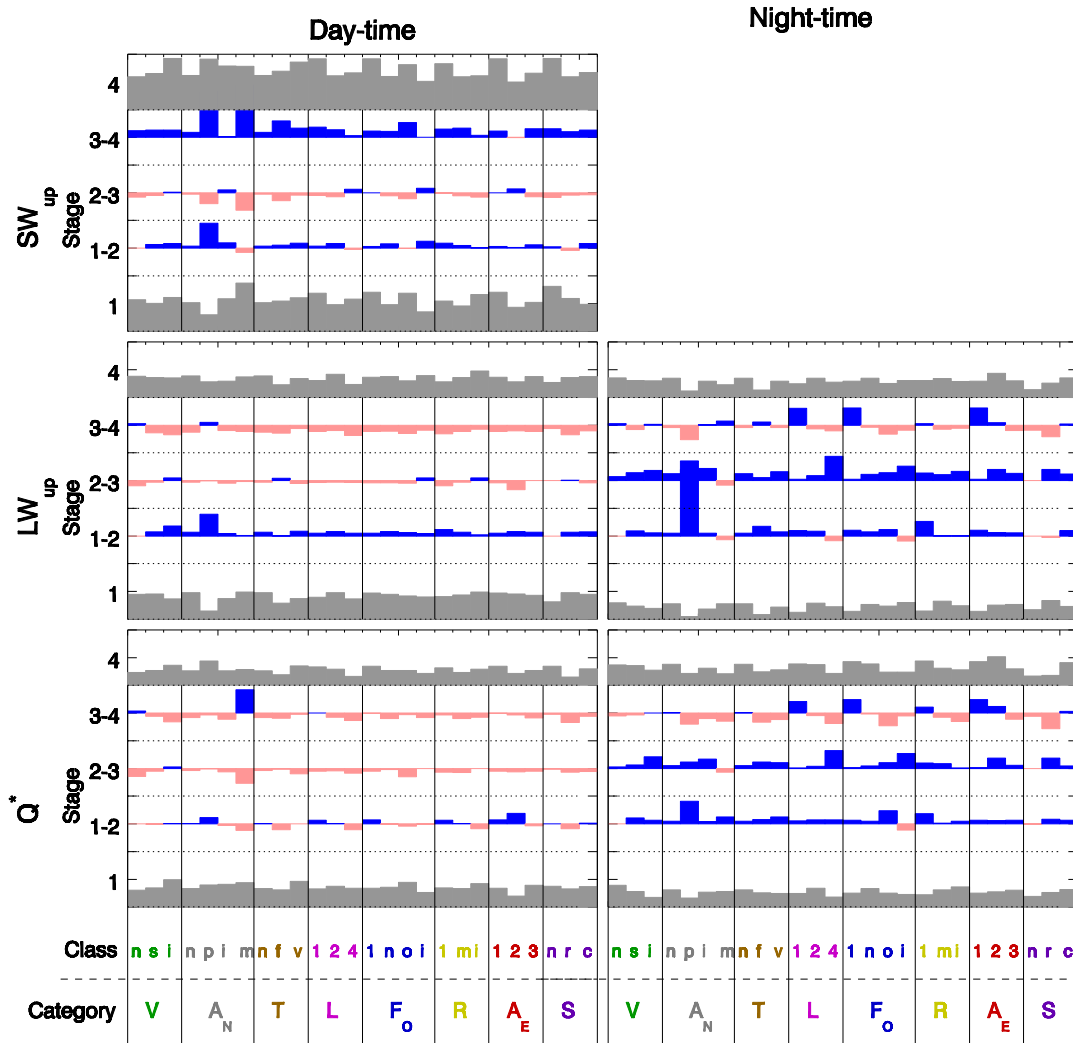
828

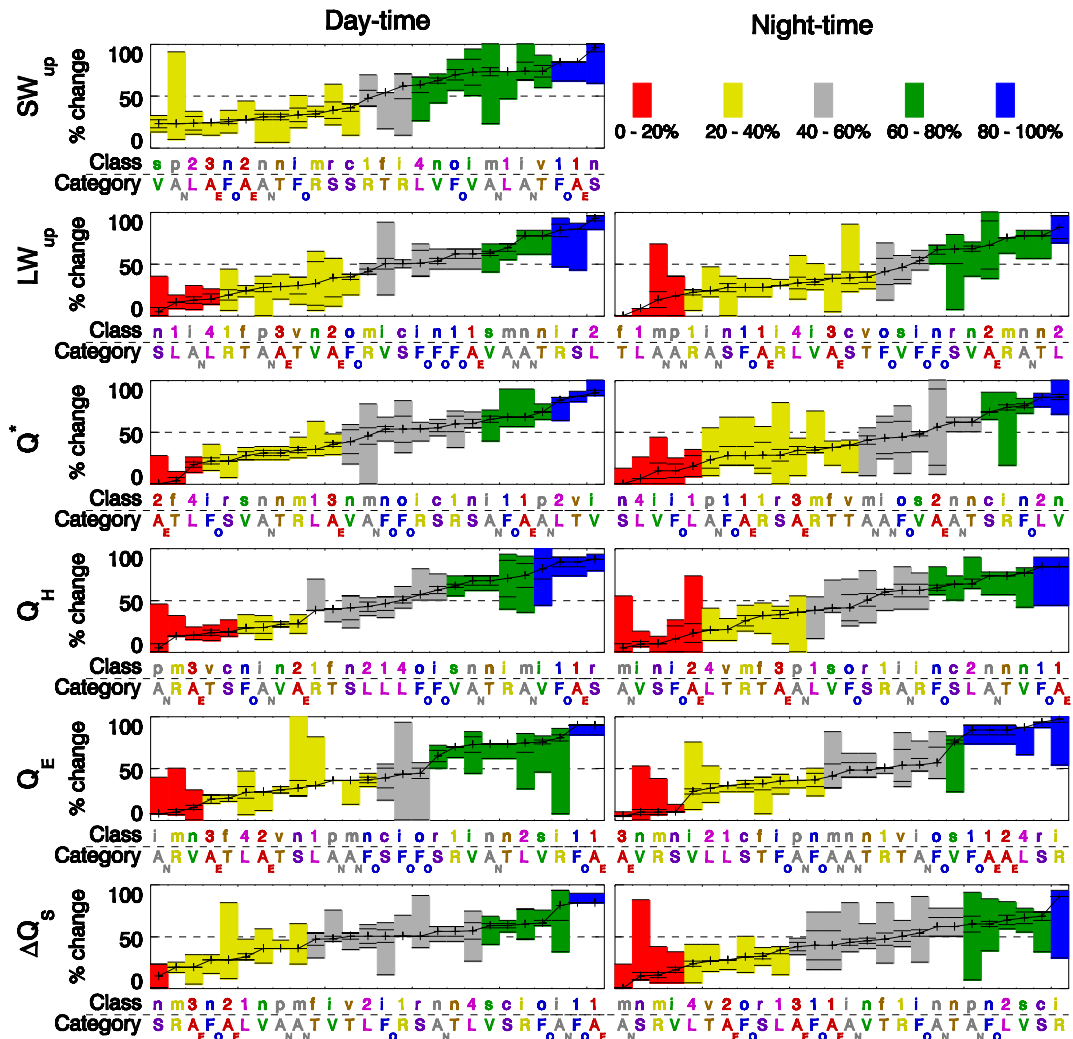829   Figure 3: As for Fig. 2, but for the radiative fluxes.

830

Figure 4: The subset of models within a class of a category improved compared to all models (P_{ca}, eqn. 4) ranked according to the median over the stages (for each flux, by time of day (as for Fig. 2)). Shading shows the range of results over the stages, with the individual results shown as horizontal lines within this. The colouring emphasises the values of the median over the stages, with 100% corresponding to all classes improved, 0% all classes degraded and 50% no change. Note X-axis code (Table 4) order changes between subplots because of ranking (Colour text is to aid differences to be noted).