Supplementary Materials

**Text S2.** Figure S3 shows the errors for a 10-day segment in June for two different model configurations. Overall, the configurations had slightly higher wind speed error magnitudes (maximum hourly average errors of 3 m s-1 within 1000m from the surface with the *SLUCM\_MYJ\_def* simulation) when compared to the surface stations errors (Section 3.3.2; Fig. 11a) and ACARS (Section 3.3.3; Fig. 12). Part of this discrepancy might be explained by the fact that these observations were taken for only 10 days and at a single point (lidar in Figure 1a), unlike the surface station and ACARS analysis which consisted of data points that were recorded at various points throughout the city and over a longer period of time.

 All of the model configurations had similar error structures throughout the boundary layer. There is an inconsistency present between the model and observations in the early morning hours (roughly 12 UTC to 16 UTC) and these errors are likely due to mistiming of the deepening of the PBL and incorrect rate of PBL development. The morning to early afternoon errors within the boundary layer decrease from 3 m s-1 with *SLUCM\_MYJ\_real* to 2 m s-1 with *BEP\_MYJ\_real*. The simulated wind speeds during the evening and early morning hours of the BEP UCM simulations outperformed all other model configurations. The errors seem to start at the surface and then extend upwards (up to about 1000m above mean sea level) as the day advances, which suggest a link between PBL growth misrepresentation and the magnitude of the simulated wind speed error.