

## ASSESSMENT OF NUMERICAL WEATHER FORECASTS AGAINST NEAR-SURFACE OBSERVATIONS OVER AN ANNUAL CYCLE

John M Edwards \*, J R McGregor, M R Bush, J Bornemann  
Met Office, U.K.

### 1. INTRODUCTION

The parametrization of the stable boundary layer in numerical weather prediction models is problematic: it is well-known that operational parameterisations imply more turbulent mixing than can be justified on theoretical grounds (Beare 2004, Cuxart 2006). However, attempts to use more physically based parameterisations lead to poorer model performance, with the surface and the surface layer becoming too cold: since screen-level temperatures are such an important forecasting product this is unacceptable.

Much has been learnt about the stable boundary layer in models from intensive field studies (*e. g.* Poulos 2002), where it is possible to focus on physical processes, but it the particular features of an individual case may obscure generic model behaviour. The aim of the current work is to complement this approach by comparing the statistical behaviour of numerical weather forecasts with detailed field observations for a single site across a complete annual cycle. The period from October 2006 to September 2007 was selected for this study. Two principal goals for the study were to characterize the surface flux budget and to understand the relationship of the surface and screen temperatures.

### 2. OBSERVATIONAL DATA

The observational data considered here were obtained at a field site at Cardington, Bedfordshire, UK. These data comprise upward and downward SW and LW fluxes, a radiometric surface temperature, temperature and humidity at the screen level (1.2 m here) and temperature, humidity and wind data at heights of 10, 25 and 50 m on a mast. The high-frequency data from the mast are used to form covariances, from which turbulent fluxes are inferred, as well as yielding mean data.

---

\*Corresponding author address: John M Edwards, Met Office, FitzRoy Road, Exeter, Devon, EX1 3PB, U.K.; email:john.m.edwards@metoffice.com

### 3. NUMERICAL WEATHER FORECASTS

All forecasts considered here were made with the Met Office Unified Model. Three forecasting configurations were studied here: the global configuration (GBL), with a horizontal resolution of 40 km over the UK, the North Atlantic European configuration (NAE), with a resolution of 12 km, covering the whole of Europe and the north Atlantic east of Newfoundland, and the 4-km high-resolution forecasting configuration (UK4) with a domain covering the UK. The vertical resolution of all configurations at the time was the same, with 38 vertical levels, 13 of which lie below 3 km; the boundary layer scheme of (Lock 2000) is allowed to operate over these levels. All three configurations represent turbulence in the stable boundary layer using a Richardson number scheme, although there are differences of formulation between the configurations, with the NAE and UK4 configurations having sharper tails than the global model. Each configuration is run 4 times a day at 6-hourly intervals and a number of diagnostics are archived from each run at hourly intervals, providing a database of forecasts from which the point values closest to the field site have been extracted for this study. In the following only results from the NAE and UK4 configurations are considered.

### 4. CONSISTENCY OF TEMPERATURE MEASUREMENTS

Since turbulent fluxes in the SBL are fairly small and high accuracy is desired in temperature forecasts, it is important to consider the consistency and accuracy of the observations: this is particularly pertinent to the relationship of the surface and screen temperatures.

Air temperatures are measured using platinum resistance thermometers (PRTs), while surface temperatures are measured using an IR thermometer. To assess the consistency of these measurements, Fig. 1 shows the ratio of differences in potential temperature between the mast (10 or 50 m) and the

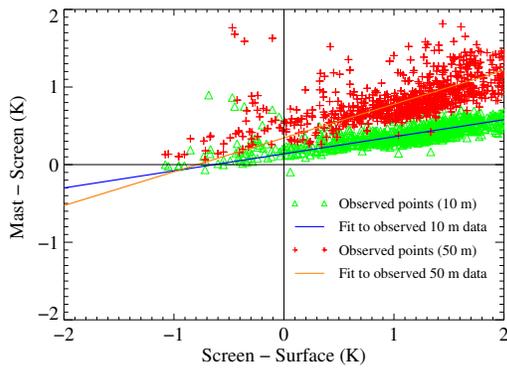


Figure 1: Scatter plot of the difference in potential temperature between the mast and screen against that between the screen and the surface for nearly neutral conditions in January 2007. Mast data for 10 m and 50 m are shown, together with best fits to these data.

screen (1.2 m) levels against the difference between the screen and surface temperatures for nearly neutral conditions with low LW surface cooling during January 2007. For a given height of observation on the mast, the data points would be expected to lie on a line through the origin with a slope of  $\log(z_{\text{mast}}/z_{\text{screen}})/\log(z_{\text{screen}}/z_{0h})$ , where  $z_{0h}$  is the thermal roughness length. The plot shows fits to the raw data which do not pass through the origin. Since the PRTs are believed to be more accurate, this may suggest that the IRT has a warm bias of 0.5–1 K for these data. Data for other months suggest that the warm bias rises as the surface temperature increases. This interpretation is broadly consistent with the analysis of Lapworth (pers. comm.), who found that the IR surface temperature showed a warm bias relative to *in situ* measurements of the grass temperature. The possibility of a warm bias must be borne in mind when interpreting measurements of the surface temperature.

A contrary bias may exist under clear skies. Since the IR temperature has been inferred from a radiative measurement, taking the surface emissivity to be 1.0, whereas the actual emissivity of dry grass is slightly lower, there may be an apparent cold bias, superimposed on an underlying warm bias in conditions of significant surface cooling, which is why the fits above were made using only data with low surface cooling.

The figure shows one other feature: the slope of fitted lines suggest a thermal roughness length of the order of  $1 \mu\text{m}$ , which is considerably smaller than

that used operationally (0.015 m).

## 5. FORECASTS OF SCREEN TEMPERATURE

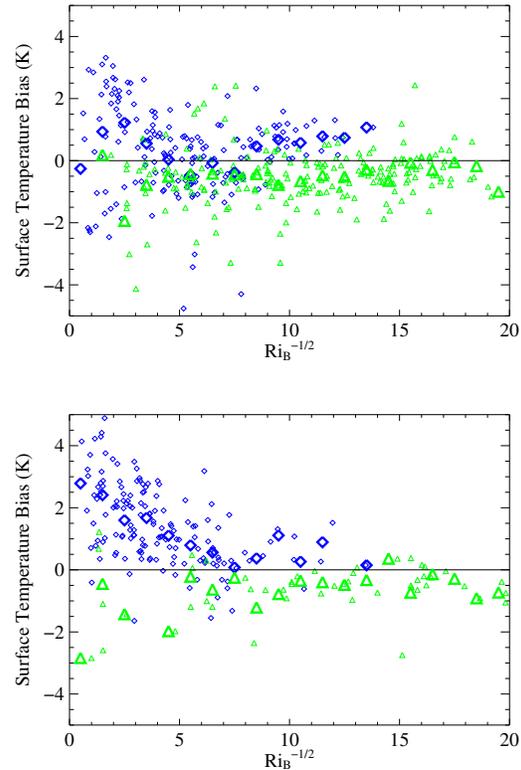


Figure 2: Scatter plots of the differences in surface temperature between the NAE forecast and the observations against the inverse square root of the bulk Richardson number for the surface layer for conditions of weak surface cooling (net LW cooling  $< 20 \text{ Wm}^{-2}$  in both the model and the observations), shown in green, and for conditions of strong surface cooling (net LW cooling  $> 80 \text{ Wm}^{-2}$ ) in blue both for the model and for the observations. Small symbols show the actual data points and large symbols bin averages for increments of 1 in  $Ri_B^{-1/2}$ . The upper panel shows results for December–February and the lower panel for March–May.

The forecasting of screen temperature is particularly difficult in very stable conditions where standard surface similarity theory may not apply (Sorbian 2006). To forecast the screen temperature accurately, it is necessary to predict the surface skin temperature and to represent the relationship be-

tween this temperature and the screen temperature correctly.

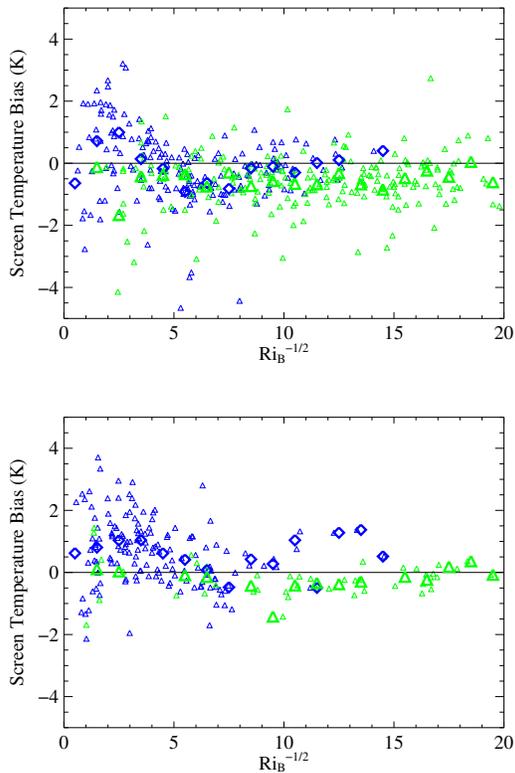


Figure 3: As above, but for the screen temperatures.

Figure 2 shows the errors in the forecast surface temperature for nocturnal conditions from the NAE configuration for December–February (upper panel) and March–May (lower panel). The bulk Richardson number for the lowest 10 m of the atmosphere is used as a measure of stability, with the specific form  $Ri_B^{-1/2}$  being employed, since this is proportional to the wind speed. Data have been selected for conditions of strong surface cooling (net upward LW  $> 80 \text{ Wm}^{-2}$ ), shown in blue and weak surface cooling (net upward LW  $< 20 \text{ Wm}^{-2}$ ), shown in green. The raw data show considerable scatter, principally due to errors in forecasts of cloud cover, so bin averages for each increment of 1 in  $Ri_B^{-1/2}$  are constructed. Under low surface cooling in winter (green points in the upper panel) the forecast surface temperature shows an apparent cold bias of around 0.5 K across all stability classes; but this bias may be partly apparent if there is a warm bias in the raw observations. Under strong surface cooling, there is an apparent warm error in the model, which is slightly more difficult to interpret because of the need to allow for

the greater impact of the surface emissivity in conditions of strong cooling. This warm bias in conditions of strong surface cooling is more apparent in spring (lower panel), where the apparent bias reaches almost 3 K in the most stable conditions. This is too large to explain in terms of the surface emissivity, especially if there is an underlying warm bias in the observed IR surface temperature, as surmised above. We cautiously interpret these results as showing that the model's surface temperature has a warm bias in conditions of strong surface cooling and high stability.

Figure 3 shows the corresponding errors in the screen temperature for winter upper panel and spring (lower panel). Again, these show a cold bias in conditions of weak surface cooling and a warm bias in very stable conditions of high surface cooling, although the warm bias in conditions of strong cooling in spring is somewhat smaller than that found for the surface temperature. This suggests that the forecasting model may underestimate the difference in potential temperature between the screen and the surface in very stable conditions, and we now consider the implications of this for the forecasting of screen temperatures.

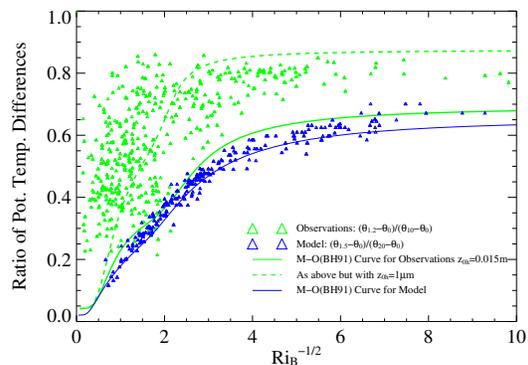


Figure 4: Scatter plot of the ratio of potential differences between the screen and the surface to those between the reference height (mast or lowest model level) and the surface for nocturnal conditions in April 2007 for the NAE (blue) and the observations (green). Solid curves show the theoretical curves which the data should follow according to Monin-Obukhov theory with the thermal roughness length used operationally. Additionally, for the observations, the theoretical line obtained with a thermal roughness length of  $1 \mu\text{m}$  is shown as the dashed curve.

Considering the data for stable conditions during

April 2007 (a very calm and settled month), it was found that differences between the forecast screen and surface temperatures barely exceeded 2 K, while observed differences in the range 4–6 K were not uncommon. Screen temperatures are forecast from the surface temperature and that on the lowest atmospheric level, using an interpolation that depends on surface similarity theory. It is therefore useful to consider the ratio of differences in potential temperature,  $(\theta_{\text{screen}} - \theta_{\text{surface}})/(\theta_{\text{lowest-level}} - \theta_{\text{surface}})$  against our measure of stability,  $Ri_B^{-1/2}$ . The blue symbols in figure 4 show this ratio for the NAE configuration (for nocturnal conditions during April 2007), together with the line along which they should fall, predicted using the similarity functions of (Beljaars 1991) and the roughness lengths used operationally for this grid-point. The forecast data essentially follow this curve<sup>1</sup>. Since the model's definition of the screen height (1.5 m) differs from that at which the observations were taken (1.2 m) and the model's lowest temperature level is at 20 m, while the observed temperature on the mast is at 10 m, it is not possible to show exactly the same ratio of temperatures from the observations; however, adjusting to the heights of the observations, and using the model's roughness lengths, the observational data should follow the solid green curve. In practical terms this is close to the blue curve; but the observational data do not follow it at all closely, particularly at high stabilities, where the ratio of potential differences is much higher than that predicted theoretically. Reducing the thermal roughness length to  $1 \mu\text{m}$ , as suggested by the data in Fig. 1, produces the dashed line, which fits the observed points much better, but still underestimates the ratio of potential differences at high stabilities. This may indicate the influence of other processes such as decoupling, radiative effects and gravity currents not represented within standard surface similarity theory.

### 5.1 Diurnal Composites of Surface and Screen temperatures

The upper panel of figure 5 shows a diurnal composite of the surface and screen temperatures for the months December–February and March–May from the observations, while the lower panel shows the error in the NAE and UK4 forecasts relative to the observed values. In winter there is an apparent cold bias in the surface temperature of the NAE (which may partly represent a warm bias in the observations,

<sup>1</sup>Some scatter is expected since the forecast data are stored to a limited precision to economise on storage.

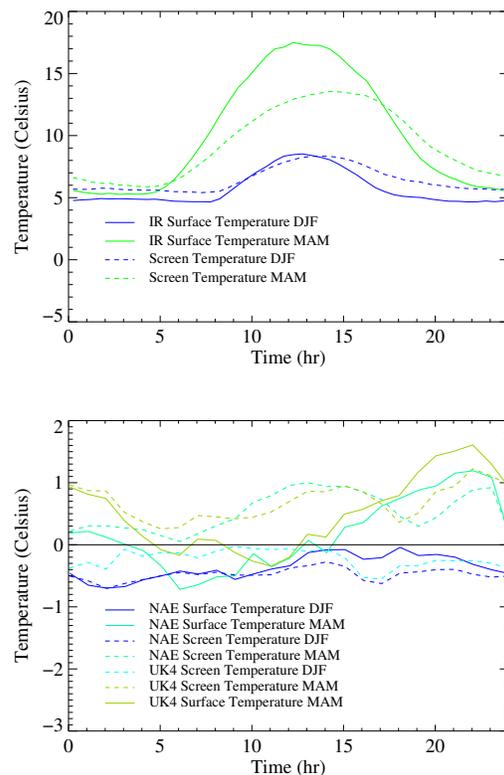


Figure 5: Diurnal composites of the observed surface (solid) and screen temperatures (dashed) for December–February (blue) and March–May (green) in the upper panel. The lower panel shows the differences in the diurnal composites between the forecast and the observations for the same variables and months. Dark blue and bluish green curves refer to the NAE configuration and light blue and olive green to the UK4 configuration: surface temperatures were not archived for the UK4 model at this period.

and in fact, the warmth of the observed surface temperature relative to the screen temperature around noon is difficult to reconcile with the very weak sensible heat fluxes around noon shown later), but the diurnal signal is one of relative warming during the day with more gradual cooling at night. Later, it will be shown that the model's shortwave warming during the day is excessive. The diurnal cycle of the screen temperature shows a cold bias of a similar magnitude and a weaker relative warming during the day, together with a period of cold errors around the evening transition. Other evidence, not shown here suggests that this may be due to a failure the assumptions of standard similarity theory in very stable conditions.

In spring, both models show a significant warm bias in the surface temperature during the earlier part of the night. To the extent that the observed surface temperature may have a warm bias, the actual bias in the model may be larger, though it is also necessary to bear in mind that clear-sky conditions were more prevalent in this season, so that allowing for the emissivity of grass becomes more significant in interpreting the results. The screen temperature shows a more consistent warm bias, though with a prominent relative cold signal during the transition.

## 6. WINDS AT 10 m

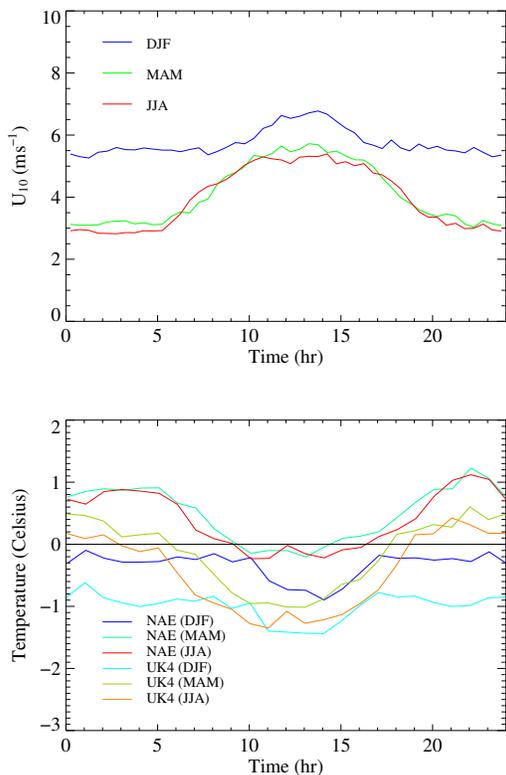


Figure 6: Diurnal composites of the observed wind speed at 10 m for winter (blue), spring (green) and summer (red) in the upper panel. The lower panel shows the differences between the forecast and observed values for the NAE and UK4 configurations.

Figure 6 shows diurnal composites of the observed wind speed at 10 m for winter, spring and summer, together with the relative errors in the NAE and UK4 forecasts. Whilst the models show different biases, there is a clear persistent diurnal signal of winds being relatively too fast at night and relatively too slow

by day.

## 7. THE OBSERVED SURFACE FLUX BUDGET

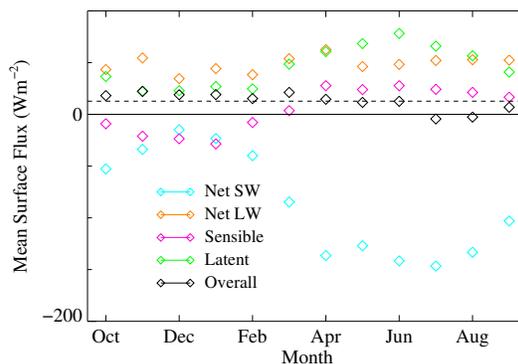


Figure 7: The observed annual cycle of the atmospheric surface flux budget, showing monthly mean values for the radiative, turbulent and overall fluxes.

In order to understand errors in the parametrization of the stable boundary layer, it is important to characterize the surface flux budget. Over the annual cycle the net atmospheric flux at the surface should be close to 0. How closely the flux budget is balanced provides a check on the consistency of the observations as a useful prelude to comparison with the model.

Figure 7 shows the annual cycle of the surface flux budget from the observations, with the convention that upward fluxes are positive. There is a strong annual cycle in the SW flux, with the anomalously sunny conditions of April 2007 being apparent. The net LW flux is more constant through the year, at around  $50 \text{ Wm}^{-2}$ . The sensible heat flux is negative from October through to February, indicating the importance of modelling the stable boundary layer during the winter months. The overall net flux is shown in black and this shows an overall net cooling bias of  $13 \text{ Wm}^{-2}$ . Superimposed on this is an oscillation with an amplitude of a similar size. We conclude that the annual cycle of the surface flux budget is generally represented realistically by the observations, but that there appears to be a bias, which must be borne in mind when interpreting model data.

Further work is required to isolate the cause of the net cooling imbalance. One possibility is that downward latent heat fluxes at night are underestimated, but there is also some evidence, involving comparison between the upward LW and the surface and near

surface temperatures, to suggest that LW cooling of the surface may be overestimated.

## 8. DIURNAL COMPOSITES OF SURFACE FLUXES

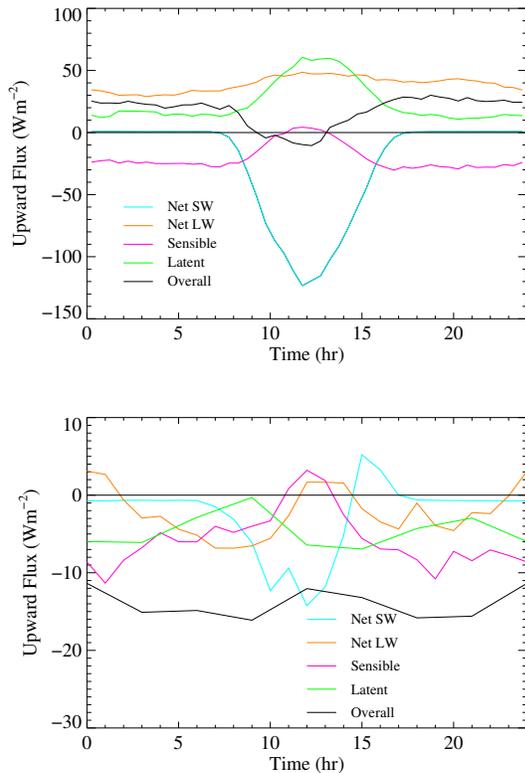


Figure 8: The diurnal composite of the observed surface fluxes for December–February (upper panel) and the differences between the NAE forecast and the observed values (lower panel).

Figure 8 shows diurnal composites of the surface fluxes from observations for winter, together with the relative errors in these fluxes from the NAE model. SW warming during the day is overestimated, which is consistent with the relative warming of the surface during the day discussed above. It is also consistent with other studies of the model which show that cloud cover is underpredicted. The model's sensible heat flux is too negative at night, consistent with the excessive turbulent mixing expected from the operational parametrization. The latent heat flux shows a smaller negative error, although this may partly be due to a bias in the observations. Relative to the raw measurements, the LW flux agrees well during the night and only shows excessive relative cooling

close to noon. The slight overall lower LW cooling in the model is not immediately easy to reconcile with a lack of cloud cover; but if the observed net LW cooling is actually overestimated, as possibly suggested by the bias in the annual mean surface flux budget, the interpretation would then be that the model does indeed show excessive LW cooling during the night.

## 9. CONCLUSION

A detailed comparison of forecasts over one year against data from one particular field site has been made. The use of full year's observational data has enabled useful checks on the consistency of the observational data to be made, which have highlighted potential issues in the interpretation of the data. It is apparent that high accuracy is required in observations to be used in characterizing model performance.

The comparison has strongly suggested that the model's screen temperature is tied too closely to the surface temperature in strongly stable conditions.

The closure of the annual mean surface flux budget to  $13 \text{ Wm}^{-2}$  with a realistic annual cycle is very encouraging, but work is required to explain the bias. The errors in individual components of the surface flux budget in a composite diurnal cycle in winter are below  $15 \text{ Wm}^{-2}$ . Taking account of possible biases in the observed fluxes, it has been *tentatively* suggested that the model may overestimate LW cooling of the surface because of a lack of cloud cover, and that this may compensate for an overestimate of turbulent warming. Although more work is required, it also seems possible that the warm bias in the surface temperature noted on calm clear evenings in spring, when the calculation of the radiative flux is less subject to uncertainties in cloud modelling, may be due to excessive turbulent heating of the surface. Thus, in these conditions, reducing the degree of turbulent mixing might actually be beneficial.

Overall, we conclude that a detailed comparisons of the surface flux budget between the model and observations have the potential to improve our understanding of the behaviour of the stable boundary layers in numerical forecasting models.

## 10. REFERENCES

- Beare, R. J., J. Cuxart, A. A. M. Holtslag, I. Esau, J.-C. Golaz, M. A. Jimenez, M. Khairoutdinov, B. Kosovic, D. Lewellen, T. S. Lund and J. K. Lundquist, A. McCabe, M. K. MacVean, A. F. Moene, Y. Noh, S. Raasch and P. Sullivan,

2006: An intercomparison of large-eddy simulations of the stable boundary layer. *Boundary-Layer Meteorol.*, **118**, 247–272

Beljaars, A. C. M. and A. A. M. Holtslag, 1991: Flux parametrization over land surfaces for atmospheric models. *J. Appl. Met.*, **30**, 327–341

Cuxart, J., A. A. M. Holtslag, R. J. Beare, E. Bazile, A. Beljaars, A. Cheng, L. Conangla, M. Ek, F. Freedman, R. Hamdi, A. Kerstein, H. Kitagawa, G. Lenderink, D. Lewellen, J. Mailhot, T. Mauritsen, V. Perov, G. Schayes, G. J. Steeneveld, G. Svensson, P. Taylor, W. Weng, S. Wunsch and K. M. Xu, 2006: Single-column model intercomparison for a stably stratified atmospheric boundary layer. *Boundary-Layer Meteorol.*, **118**, 273–303

A. P. Lock, A. R. Brown, M. R. Bush, G. M. Martin and R. N. B. Smith, 2000: A new boundary layer mixing scheme. Part I: Scheme Description and Single-Column Model Tests. *Mon. Wea. Rev.*, **128**, 3187–3199

Poulos, G. S., W. Blumen, D. C. Fritts, J. K. Lundquist, J. Sun, S. P. Burns, C. Nappo, R. Banta, R. Newsom, J. Cuxart, E. Terradellas, B. Balsey and M. Jensen, 2002: CASES-99: A Comprehensive Investigation of the nocturnal boundary layer. *Bull. Amer. Meteor. Soc.*, **83**, 555–581.

Sorbjan, Z., 2006: Local structure of turbulence in stably stratified boundary layers. *J. Atm Sci.*, **63**, 1526–1537