Boundary-layer height and surface stability at SMEAR II, Hyytiälä, Finland in ERA5 and observations

Victoria A. Sinclair[1], Jenna Ritvanen[1,2], Gabin Urbancic[2], Irina Statnaia[1,2], Yurii Batrak[3], Dmitri Moisseev[1,2], and Mona Kurppa[1]

[1]Institute for Atmospheric and Earth System Research / Physics, Faculty of Science, University of Helsinki, PO BOX 64, FI-00014
[2]Finnish Meteorological Institute, Helsinki, Finland
[3]Development Centre for Weather Forecasting, Norwegian Meteorological Institute, Oslo, Norway

Correspondence: Victoria A. Sinclair (victoria.sinclair@helsinki.fi)

Abstract. The planetary boundary layer (BL) height and stratification are key parameters in determining the exchange of heat, momentum, moisture and trace gases between the surface and the free troposphere. Numerous different methods have been used to quantify the BL height and these methods have been applied to a wide variety of observational data sets obtained from different instruments and to numerical model output. We investigate the BL height at the Hyytiälä SMEAR II station in southern

5 Finland diagnosed from radiosonde observations, a microwave radiometer (MWR) and ERA5 reanalysis. Four different algorithms are used to diagnose the BL height from the radiosondes. The diagnosed BL height is sensitive to the method used and the level of agreement, and the sign of systematic bias, between the 4 different methods depends on the surface-layer stability. For example, for very unstable situations, the median BL height diagnosed from the radiosondes varies from 600 m to 1500 m depending on which method is applied. Good agreement between the BL height in ERA5 and diagnosed from the radiosondes

10 using Richardson number-based methods is found for almost all stability classes, suggesting that ERA5 has adequate vertical resolution near the surface to resolve the BL structure. However, ERA5 overestimates the BL height in very stable conditions highlighting the on-going challenge for numerical models to correctly resolve the stable BL. Furthermore, ERA5 BL height differs most from the radiosondes at 18 UTC suggesting ERA5 does not resolve the evening transition correctly. This study has also shown that BL height estimates from the MWR are reliable in unstable situations but often are inaccurate under stable

15 conditions when, in comparison to ERA5 BL heights, they are much deeper. The errors in the MWR BL height estimates originate from the limitations of the manufacturers algorithm for stable conditions and also the mis-identification of the type of BL. A climatology of the annual and diurnal cycle of BL height and observed surface layer stability was created. The shallowest (353 m) monthly median BL height occurs in February and the deepest (576 m) in June. In winter there is no diurnal cycle in BL height, unstable BLs are rare yet so are very stable BLs. The shallowest BLs occur at night in spring and summer and

20 very stable conditions are most common at night in the warm season. Finally, using ERA5 gridded data we determined that the BL height observed at Hyytiälä is representative of most land areas in southern and central Finland. However, the spatial variability of the BL height is largest during daytime in summer reducing the area over which BL height observations from Hyytiälä would be representative of.

1 Introduction

The boundary layer (BL) is the lowest part of the troposphere and is in direct contact with the Earth's surface. The BL is a critical part of the Earth system since all exchange processes between the land surface and the free troposphere take place through the BL, for example, the exchange of moisture, heat, momentum and trace gases such as CO_2. Two key characteristics of the BL are its height and stability. These characteristics are of critical importance for air quality and pollution dispersion,

30 with higher concentrations of pollutants occurring when the BL is shallow and stably stratified. The boundary-layer height and stability are also crucial variables to include in the analysis of surface-based aerosol and trace gas concentration observations as often measurements performed at ground-based stations are generalised to represent conditions throughout the BL and across larger horizontal scales. The accuracy of such an assumptions depends on spatial and temporal variations in the BL height.

Unfortunately, given its critical importance, the boundary-layer height cannot be measured directly but must instead be

35 estimated from observations made above the surface. Therefore, estimates of the boundary-layer height are challenging as vertical profiles of temperature and wind speed are required at high spatial resolution. These vertical profiles can be obtained from radiosonde soundings, tall towers or masts, or ground-based remote sensing instruments such as ceilometers, Doppler Lidars, wind profilers and microwave radiometers. Radiosondes provide direct observations of temperature, humidity and wind speed at high vertical resolution but such soundings typically only occur twice, or at most, four times a day and globally there

40 are very few sounding stations resulting in limited spatial coverage. However, some radiosonde stations have long temporal records often extending back for at least 20 years. Observations from tall towers or masts provide high temporal resolution vertical profiles but these measurements typically extend only a few hundred meters above the surface and thus are only capable of resolving shallow BLs. Ground-based remote sensing observations provide vertical profiles of attenuated backscatter or turbulence at high temporal resolution which can be used to derive the boundary-layer height. However, a general limitation of

45 remote-sensing based instruments is the limited spatial extent due to a small, albeit constantly growing, number of instruments worldwide and the absence of very long time series of observations.

Climatologies of the BL height have been created using observations from radiosondes, remote-sensing instruments and gridded reanalysis data sets. Seidel et al. (2010) analysed radiosondes from 505 stations worldwide over 10 years and used 7 different methods to estimate the BL height. They found substantial sensitivity of BL heights to the estimation method,

50 with parcel based methods producing lower estimates and methods that find the maximum in potential temperature gradient or minimum in relative humidity gradient producing the largest values of BL height. Seidel et al. (2010) also note that radiosonde soundings, with 2-4 profiles per day, are insufficient to capture the diurnal cycle. Baars et al. (2008) and Granados-Muñoz et al. (2012) both used remote-sensing instruments to create one-year climatologies of BL height for specific European stations. Similarly, Collaud Coen et al. (2014) developed an operational BL height detection method using remote-sensing instruments

55 and different algorithms to determine the BL height and compared the results to radiosonde based estimates and values from a high resolution numerical weather prediction model. For the two Swiss stations they considered, the microwave radiometer

provided convective boundary-layer heights in good agreement with the radiosonde sounding while the model in general overestimated the depth of the BL.

In contrast to observations, gridded reanalysis data sets have uniform spatial coverage across the globe, long and consistent
60   time series with no gaps, and the most recent reanalysis datasets have 3 or even 1 hour temporal resolution allowing the diurnal cycle to be resolved and therefore are well suited to created climatologies from. For example, von Engeln and Teixeira (2013) created a 20-year, world-wide climatology based on ERA-Interim reanalysis and in a comparison to both sea- and land-based radiosondes found ERA-Interim values of BL height to be in good agreement with the observations. Seidel et al. (2012) also analysed ERA-Interim reanalysis and radiosondes for a 24 year period over Europe and the continental United States. They
65   conclude that a bulk Richardson number based method was optimal to diagnose BL height from radiosonde, reanalysis, and climate model data sets and that BL height estimates from ERA-Interim were deeper than those from the radiosondes.

Despite these numerous climatological studies on BL height, there are few studies which consider climatologies of boundary-layer or surface-layer stability. Liu and Liang (2010) analysed radiosondes from 14 major field campaigns around the world and, in addition to estimating the BL height, classified the BL type as either stable, neutral, or unstable, finding that the
70   occurrence frequencies follow a narrow, intermediate, and wide Gamma distribution, respectively, over both land and oceans. Zhang et al. (2018) considered similar stability classes over China during summertime and found that 70% of the radiosondes they considered sampled convective BLs, 26 % were neutrally stratified and only 4% were stably stratified.

In this study we primarily focus on the BL height and stability at the SMEAR (Station for Measuring Ecosystem–Atmosphere Relations Hari et al., 2013), station Hyytiälä (also referred to as SMEAR II) which is located in southern Finland (the station
75   is described in detail in section 2). SMEAR II has extensive, long-term surface based measurements of atmospheric, forest and soil variables, and in particular, there are world-renown, long-term measurements of aerosol and trace gases starting from 1996 (Hari and Kulmala, 2005). Some previous studies have investigated the boundary-layer height at SMEAR II, however many studies were case studies or short-term studies (e.g. Lauros et al., 2007; Eerdekens et al., 2009; Ouwersloot et al., 2012). A recent more in-depth study of boundary-layer turbulence was presented by Manninen et al. (2018) who analysed one year of
80   Doppler lidar data from Hyytiälä to identify the main source of turbulent mixing. They show that there is almost no diurnal cycle in the occurrence or source of turbulence in winter whereas in summer surface based turbulence has a strong diurnal cycle. Notably they also find a considerable amount of nocturnal mixing near the surface, especially during the winter.

The aim of the current study is to provide an in-depth, long-term analysis of boundary-layer height and stability at SMEAR II by combining a range of observations made over the 25-year history of the station from a wide range of different instruments
85   along with the most modern global reanalysis data set from ECMWF, ERA5. The measurement station and data used is described in section 2. While we focus on one specific station, it should be noted that the methods described here are applicable, in principle, to many stations worldwide, for example all European Aerosols, Clouds, and Trace gases Research InfraStructure (ACTRIS) stations and Atmospheric Radiation Measurement (ARM) program locations. Furthermore, this study is timely as many scientists analysing long-term time series of surface based measurements of aerosol or trace gases are now starting
90   to incorporate meteorology and in particular BL dynamics, and primarily BL height, into their analysis. However, very few

stations with long-term surface measurements of aerosol and trace gases have equivalently long time series of vertical profiles from which boundary-layer height can be determined.

Given the different observation types and the difference in data between observations and reanalysis, different methods to estimate boundary-layer height need to be applied. Therefore, we begin this investigation with an overview of various methods to calculate the BL height using different data sets (section 3). In section 4, we show some examples of how these different methods, applied to observational and reanalysis data sets, estimate the boundary-layer height in both stable and unstable conditions. A systematic, statistical comparison between the different methods of diagnosing the BL height is given in Section 5. A long-term climatology of BL height and surface-layer stability is presented in section 6 and an estimate of the spatial representativity of the measurements of BL height made at SMEAR II is made based on ERA5 reanalysis data in section 7. Lastly, in section 8 the results are discussed and conclusions presented.

2 Measurement station and Data

2.1 SMEAR II

The SMEAR II station (Hari and Kulmala, 2005; Petäjä et al., 2016) is located in southern Finland (61°51'N, 24°17'E, 181 m a.s.l.), 220 km northwest of Helsinki (Fig. 1). The station is situated in a Boreal Scots pine forest stand which is 59 year old (in 2021) and has a mean canopy height of around 18 m (Bäck et al., 2012). A small lake, Kuivajärvi, (surface area of 8.4 km^2) is located directly beside the station (Fig. 1). SMEAR II is part of the European Aerosols, Clouds, and Trace gases Research InfraStructure (ACTRIS) and is a certified Integrated Carbon Observation System (ICOS) atmospheric station. Measurements began in 1996 and consist of extensive aerosol, trace gas and standard meteorological measurements which are made at the surface and on instrumented towers. The bulk of the in-situ measurements are made either at ground level or on an instrumented tower at the main station location indicated by "M" in Figure 1 Ground-based remote sensing instruments have gradually been added to the site since 2014. However, due to the very clean atmosphere, it is challenging to use ceilometers and Doppler lidars to derive BL height at this station due to the small signal to noise ratio (Manninen et al., 2018).

2.2 Radiosondes

Regular, continuous radiosondes are not released from SMEAR II however radiosondes have been released during intensive campaigns. An intensive measurement campaign called The Biogenic Aerosols - Effects on Clouds and Climate (BAECC) took place in Hyytiälä in from 1 February - 13 September 2014 (Petäjä et al., 2016, and also see the Data availability statement at the end of this paper). During the campaign, alongside with many other measurements, radiosonde soundings were released four times per day at 00, 06, 12 and 18 UTC (local time is UTC+2 hr in winter, UTC+3 hr in summer). The soundings were released from a small field approximately 200 m southeast from the main SMEAR II station (location marked by R in Fig. 1). Vertical profiles of dry and wet bulb temperature, relative humidity, pressure, altitude above sea level, and horizontal wind speed and direction are available with measurements every 10 - 15 m and extending from the surface to usually at least 20

km above the surface. There are several methods of estimating BL height from the radiosonde profile and these are described in section 3. Although the radiosonde data from the BAECC campaign was transmitted into the Global Telecommunication System (GTS), by inspecting the ERA5 analysis increment files, we conclude that these soundings have not been assimilated into ERA5.

2.3 Microwave Radiometer

In this study a humidity and temperature profiling microwave radiometer (MWR; Rose et al., 2005; RPG Radiometer Physics GmbH, 2014a) was used to provide an additional estimate of BL height. The radiometer has been a part of the ACTRIS cloud profiling station since June 2018. The MWR is located 200 m southeast from the main SMEAR II station (R in Fig. 1). The MWR measures emissions in two frequency bands, 22-31 GHz and 51-58 GHz, which correspond to water vapour and oxygen absorption lines. To further improve the retrieval of temperature profiles within the BL, measurements at 7 elevation angles, starting at 5° from horizontal, were performed. The profiles are available approximately every 10 minutes, with vertical resolution varying from 30 m below 1200 m to 200 m above that. The temperature profiles are retrieved from the measurements with an artificial neural network algorithm (Solheim et al., 1998) that was trained for the climate of Southern Finland using several years of radiosonde observations from the Finnish Meteorological Institute (FMI) station in Jokioinen. These measurements were used to simulate MWR observations by applying radiative transfer computations. The simulated MWR and radiosonde observations were used to train the neural network. A previous study on a similar MWR found random error between the MWR temperature profiles and radiosonde measurements increasing from 0.5 K in the lower BL to 1.7 K at 4 km altitude (Löhnert and Maier, 2012), while the accuracy (as root mean square error) of the temperature profiles given by the manufacturer ranges from ± 0.7 K in the BL to ± 1.0 K above 2000 m. Retrieval of the BL height from the temperature profiles is described in Section 3.2.

Since the MWR measures rather weak emission signals, such things as a wet antenna radome due to rain can severely affect the quality of the measurements. To address this potential issue, the MWR assembly includes a heating element and air blower that continuously dries the radome and the actual radome has a hydrophobic coating. However, this coating will wear off gradually over time and we would expect the quality of the measurements to decrease as the coating decays. To detect rain events and remove observations that were potentially affected by a wet radome, the MWR setup includes a Vaisala WXT 530 weather station, which uses an impact sensor for the precipitation detection. Because this sensor may miss light precipitation, OTT Parsivel[2] (Tokay et al., 2014) and OTT Pluvio[2] measurements were also used to flag the data that could be potentially affected by rain. OTT Parsivel[2] is a present weather sensor capable of measuring precipitation intensity and raindrop size distribution, and OTT Pluvio[2] is a weighing precipitation gauge. Both instruments are located within 30 m of the MWR. To avoid using MWR measurements affected by rain, measurements at times when rain intensity exceeded $0.1 \, \mathrm{mm\,h^{-1}}$, corresponding to accumulation of 0.008 mm, during the previous 5 minutes, detected by either the OTT Parsivel[2] disdrometer or OTT Pluvio[2], were excluded from the analysis.

2.4 Eddy Covariance and surface stability

155 Eddy covariance (EC) is the method applied to compute vertical turbulent fluxes within the atmospheric BL. The vertical turbulent fluxes of scalar quantities are obtained by calculating the covariance of the scalar and the vertical component of the wind speed. At SMEAR II it is possible to use the EC method due to the deployment of two Gill Solent 1012R ultrasonic anemometers at 23 m and 46 m. These instruments measure the three wind speed components and sonic temperature at 10 Hz. The turbulent fluxes analysed in this study were computed using the EC package EddyUH, a public software package
160 developed at the University of Helsinki for post-field processing of eddy covariance data (Mammarella et al., 2016).

We retrieved the processed data of the sensible heat flux (H), the friction velocity (u_*), air temperature and pressure as 30-minute averages covering the time period 1 January 1997 to 31 December 2019 from the SmartSMEAR public online database (https://smear.avaa.csc.fi). This data set contains missing data which is due to the assumptions made during the computation of the fluxes (e.g. stationarity) and technical instrumentation limitations (e.g. icing in winter). This topic, and the general quality
165 control of EC data, is discussed in detail by Mammarella et al. (2016) and Vickers and Mahrt (1997).

Using this data we computed the Obukhov parameter, L using

$$L = -\frac{T_0 u_*^3}{\kappa g \overline{w'T'}}.$$ (1)

where κ is the von Kármán constant, g is the gravitational constant, T_0 is a reference temperature where we used the 30-minute mean temperature at 16.8 m, $\overline{w'T'}$ is the vertical velocity and temperature covariance (also called the kinematic heat flux) given
170 by

$$\overline{w'T'} = \frac{H}{\rho c_p}$$ (2)

where ρ is air density and c_p is the heat capacity of air at constant pressure. Using the Obukhov length we can compute the stability parameter, ζ, defined as,

$$\zeta = \frac{z}{L}$$ (3)

175 where z is the height above the surface. Negative values of the stability parameter imply unstable stratification whereas positive values indicate stable conditions. When attempting to classify the stability near the surface, ζ has the advantage compared to the Richardson number in that only one level of observations is required. In this study, we define 6 stability classes to characterise the state of the surface layer. The stability classes are defined and subsequently referred to as: very unstable ($\zeta < -1$), weakly unstable ($-1 < \zeta < -0.01$), near-neutral unstable ($-0.01 < \zeta < 0$), near-neutral stable ($0 < \zeta < 0.01$), weakly stable ($0.01 <$
180 $\zeta < 1$), very stable ($\zeta > 1$).

2.5 ERA5 reanalysis

To complement the observations, we also use ERA5 reanalysis data (Hersbach et al., 2020). ERA5 is the fifth generation of ECMWF atmospheric reanalysis, and as is the case for all reanalysis, was produced by combining a vast number of observations

and a numerical weather prediction model. Due to the involvement of a modelling system, reanalysis cannot be considered as truth in the same manner as an observation can but it should be considered to be a best-guess of the state of the atmosphere. The numerical model forming the basis of ERA5 is the Integrated Forecast System (IFS) Cycle 41r2. Reanalysis data sets differ notably from operational archives produced by numerical weather prediction models. Although operational models have much higher resolution (typically from 1 to 4 km in contemporary regional numerical weather prediction systems) than global reanalysis data sets, due to frequent model updates, they can not provide consistent long term data sets, required for climatological studies.

ERA5 data covers the time period from 1950 to present with a time resolution of 1 hour and has the nominal spatial resolution of 31 km (or 0.28125 degree longitude and latitude near the equator). Spectral fields in ERA5 are stored with T639 truncation and grid-space fields are stored on N320 reduced Gaussian grid. Here we used data from 1 January 1979 to 31 December 2019, re-gridded to a regular 0.25 degree grid. BL height is provided as a diagnostic (see section 3.3) and we extract the BL height covering most of Finland and Scandinavia. Most of the analysis is based on BL height from the grid cell nearest to the Hyytiälä SMEAR II station which, in ERA5, has an altitude of 139 m a.s.l and has a land fraction of 0.91. In addition, we also extract a limited number of vertical profiles of potential temperature, relative humidity and the zonal (u) and meridional (v) wind speed components from which the scalar wind speed is calculated. These profiles are extracted on the native model levels, of which there are 137, with vertical grid spacing ranging from 20 to 100 m in the lowest 1 km.

3 Methods to estimate the boundary-layer height

A well-known qualitative definition of the boundary layer is that it is "the lowest part of the troposphere that is directly influenced by the presence of the Earth's surface, and responds to surface forcing within a timescale of about an hour or less" (Stull, 1988). However, many other definitions exist (e.g. Seibert et al., 2000) and objective definitions are required when developing automated methods to quantitatively calculate values of BL height. The methods applied can depend on what variables are available and on the type (i.e. convective, neutral or stable) of BL. In general, it is simpler to identify the height of the convective BL compared to the stable BL as typically a capping inversion is present, the sensible heat flux changes sign with height, and a clear decrease in turbulence is present. Stable BLs are not characterised by inversions and Seibert et al. (2000) note that for the stable BL "no universal relationship seems to exist between the profiles of temperature, humidity or wind and turbulence parameters (heat or momentum fluxes, turbulent kinetic energy TKE)." In this section, we describe the methods applied to the radiosonde, microwave radiometer and reanalysis data sets to quantify the BL height.

3.1 BL height from radiosonde soundings

To estimate the BL height from the radiosonde soundings, we use the planetary boundary layer height (Sivaraman et al., 2013) value added product (VAP) developed by the Atmospheric Radiation Measurement (ARM) user facility. This VAP was applied to the radiosonde profiles of pressure, dry bulb temperature, relative humidity, wind speed, and wind direction by ARM. As there is no "truth" to evaluate the accuracy of the BL height estimates, and since the BL height is somewhat subjective,

four different methods to estimate the BL height from radiosondes are included in this VAP. The difference between the four methods can be considered as a partial estimate of the uncertainty in the diagnosed BL heights. The four methods are those proposed by Heffter (1980) (referred to as H80), Liu and Liang (2010) (referred to as LL10), and a bulk Richardson number method (Seibert et al. (2000)) using two different critical Richardson number values (0.25 and 0.5) as thresholds to define the
220   top of the BL (referred to as $R_{0.25}$ and $R_{0.50}$). Before the BL height is calculated, the radiosonde data is sub-sampled onto a uniform 5 hPa vertical grid.

The H80 method determines the BL height using vertical gradients in potential temperature. First the radiosonde input data is further smoothed using a three-point moving average to a vertical resolution of 15 hPa to reduce localised noise and then the vertical gradient (lapse rate) of potential temperature is calculated. Inversion layers (up to 5) are then identified as contiguous
225   heights where the potential temperature lapse rate is greater than 0.005 K m^{-1}. The heights of the top and bottom of these inversion layers, and the change in potential temperature across each layer is saved. The BL height is then identified as the lowest inversion layer in which the potential temperature difference between the base and top of the inversion is greater than 2 K. If no such layer is found below 4 km, the BL height is set to the height of the inversion layer with the largest maximum potential temperature gradient.

230   The LL10 method requires the type of the BL to be determined before the BL height is estimated. The BL type (convective, stable or neutral residual layers) is identified using the potential temperature difference between the 5th and 2nd measurement above the surface (i.e. over a 15-hPa deep near-surface layer). This potential temperature difference is compared to a stability threshold, δ_s which has a value of 1 K. If $\theta_5 - \theta_2 < -\delta_s$, a convective BL is present, if $\theta_5 - \theta_2 > \delta_s$, a stable BL is present, and if $-\delta_d \leq \theta_5 - \theta_2 \leq \delta_s$, a neutral residual layer is declared. For convective BLs and neutral residual layers, the BL height is
235   defined to be the height at which an air parcel rising adiabatically becomes neutrally buoyant. To determine the level of neutral buoyancy, for each vertical level, the potential temperature difference between this level and the level closest to 150 m above the surface is calculated. An upward scan is performed until the first level when this difference exceeds a instability threshold ($\delta_u = 0.5$) is identified. Starting from this level, a second upward scan and second threshold is then applied to account for over-shooting parcels. The final BL height is therefore the first level above the level of neutral buoyancy where the vertical
240   potential temperature gradient exceeds 0.004 K m^{-1}.

If the BL is stable, first the level at which the vertical potential temperature gradient reaches a minimum is identified. If this level is a local peak, defined as the difference in the potential temperature gradient between this level and the level below is less than -40 K km^{-1}, or if there is an inversion layer in the two layers above, then this level is defined to be the BL height. In stable conditions, the Liu and Liang (2010) method also checks for the presence of low-level jets as shear-driven turbulence associated
245   with these features can influence the BL height. A low-level jet is identified if the wind speed has a localised maximum which is at least 2 ms^{-1} stronger than the wind speed in the levels immediately above and below. If a low-level jet is identified, the BL height is defined as the height of the wind maximum. If both the criteria based on potential temperature gradients and wind-shear are met, then the BL height is taken as the lower of the two heights. If neither the condition based on stability nor on wind shear are met, a missing value of -9999 is specified. This occurs in less than 2% of the soundings from the BAECC
250   campaign.

The Richardson number based methods involve calculating the bulk Richardson number:

$$Ri_b = \left(\frac{gz}{\theta_{v0}} \right) \left(\frac{\theta_{vz} - \theta_{v0}}{u_z^2 + v_z^2} \right) \tag{4}$$

where g is the acceleration due to gravity, θ_{v0} and θ_{vz} are the virtual potential temperature at the surface and height z, respectively and u_z and v_z are the horizontal wind speed components at height z. From the radiosonde observations, θ_{v0} is calculated

255   using the lowest measurement which is usually at a height of 2 m. The Richardson number is the ratio of the buoyant and mechanical production terms of the TKE budget equation. At a critical value of the Richardson number, the mechanical production of turbulence is balanced by buoyant consumption and at Richardson numbers equal or larger to the critical value the flow can not sustain turbulence (Stull, 1988). Thus, the BL height is diagnosed to be the height at which Ri_b reaches a critical value. Two critical values, 0.25 and 0.5 are considered in the present study.

3.2   BL height from microwave radiometer measurements

In this study we use the BL height estimates provided directly from the MWR manufacturer's software (RPG Radiometer Physics GmbH, 2014b). The MWR provides an estimate of the BL height based solely on the vertical profile of potential temperature as no vertical profiles of wind speed are available from the MWR. The potential temperature (θ) is calculated from the retrieved temperature profile at each altitude, z, using

265   $$\theta(z) = T(z) \left(\frac{p_0}{p(z)} \right)^{R/c_p} \tag{5}$$

where $T(z)$ is the temperature at altitude z, $p(z)$ is the pressure at altitude z estimated from the barometric formula for the standard atmosphere, p_0 is a reference pressure, R is the ideal gas constant, and c_p is the specific heat capacity at constant pressure. The potential temperature profile is then used to determine whether the BL is unstable or stable. The BL is classified as unstable if $\theta(z) < \theta(z_0)$ for any $z > z_0$ i.e. if potential temperature decreases with height. In this case, the BL height z_{BLH}

270   is defined through the parcel method (Holzworth, 1964) as the height at which the potential temperature is equal to the surface potential temperature, i.e. $\theta(z_{\mathrm{BLH}}) = \theta(z_0)$. Thus, for unstable cases the MWR defines the BL height as the level of neutral buoyancy and, unlike the H80 and LL10 methods, does not attempt to account for over-shooting thermals. Otherwise, the BL is classified as stable, and the BL height is defined to be the height at which the vertical derivative of the potential temperature, θ', has a localised minimum. If $\theta'(z)$ has no minimum, then $z_{\mathrm{BLH}} = 0$. Very similar methods have been applied to MWR

275   temperature profiles by Collaud Coen et al. (2014) and Moreira et al. (2020) to diagnose the BL height.

3.3   BL height from ERA5

The BL height diagnostic in ERA5 is required to be valid for all types and depths of boundary layer given the global nature of ERA5. Within ERA5, the BL height is diagnosed online using an algorithm originally proposed by Vogelezang and Holtslag (1996). This is a Richardson number based method and the BL height is defined as the lowest model level at which the bulk-

280   Richardson number reaches the critical value of $R_B = 0.25$. Additional details are in the IFS documentation (ECMWF, 2015).

4 Case studies

In this section we show examples of the radiosonde and MWR data and how the diagnosed BL height relates to the observed vertical profiles of meteorological variables. In addition, for each presented case study, we compare the observations to ERA5 data. A systematic, statistical comparison is presented in section 5.

285 ### 4.1 BL height diagnosed from Radiosondes

An example of vertical profiles of potential temperature, relative humidity and wind speed measured by radiosondes is illustrated in Fig. 2. In addition, the BL height diagnosed by the 4 methods described in section 3 are also indicated in Fig. 2. This specific case study, covering the 7-8 May 2014, includes 5 radiosonde ascents and highlights the typical diurnal evolution of the BL height and the sensitivity of the diagnosed BL height to the method.

290   At 00 UTC on 7 May, the potential temperature at the surface is below 0°C and increases rapidly with height within the lowest few hundred meters indicating a stably stratified BL. The surface-layer stability class is very stable at this time ($\zeta > 1$). Above this stable layer, the potential temperature is almost constant with height up until 1.2 km a.g.l where a weak inversion is evident above which another neutral layer is present that is capped at $\sim$2 km with an inversion and a rapid decrease in relative humidity. This strongly suggest that a residual layer is present. The BL height estimates from the 4 methods applied to the

295 radiosonde profiles range from 184 m ($R_{0.25}$ method) to 307.1 m (LL10 method).

At 06 UTC, the surface has warmed, the strong surface-based stable layer has started to mix out and the stability class is now weakly unstable ($-1 < \zeta < -0.01$). Notable differences in the diagnosed BL height are evident at this time; the H80 method identifies the BL top to be co-located with the inversion at the residual layer and thus indicates a BL height of 2468 m whereas the remaining three methods have heights from 248 m to 503 m. The large difference between the H80 and the other methods

300 occurs for two reasons. Firstly, the H80 method requires a strong inversion — a temperature difference of at least 2 K — and therefore does not identify the weak inversion at 500 m as the BL top. In contrast, the LL10 method only requires a 0.5 K temperature difference. Secondly, unlike the Richardson number approach H80 does not consider the wind speed profile, which at 06 UTC shows a localised maximum.

By 12 UTC the BL is well mixed with potential temperature almost constant with height until 2.6 km where a temperature

305 inversion is present, co-located with a strong vertical gradient in relative humidity. This is consistent with the very unstable stability class ($\zeta < -1$) present at this time. The diagnosed BL heights again show a large variation ranging from 1073 m ($R_{0.25}$) to 2793 m (H80) and based on visual analysis it appears that both Richardson number-based methods underestimate the height of the convective, well-mixed layer.

At 18 UTC the surface has started to cool and the surface layer was defined as weakly unstable ($-1 < \zeta < -0.01$) and then

310 near-neutral unstable ($-0.01 < \zeta < 0$) during the 1 hr period from 18 - 19 UTC. In addition, a low-level jet is beginning to develop. This complex situation results in divergent estimates of the BL height. The H80 and LL10 methods maintain a deep BL whereas, due to the inclusion of the wind profile, both Richardson number methods diagnose shallower BLs, the top of which is just above the low-level jet. In this case, the LL10 method identifies the BL type to be neutral as although the potential

temperature does increase with height near the surface, the increase is less than 1 K (the required minimum inversion strength

315   above the convective BL top or below the stable top (Liu and Liang, 2010)). Therefore, the LL10 scheme does not consider the wind speed profile and does not detect the low-level jet hence over-estimating the BL height. By 00 UTC on 8 May, the surface has cooled resulting in a strong stable surface layer (the stability class is very stable at this time and also the LL10 method determines the BL to be stable now) and a strong low-level jet is present. All methods diagnose the BL height to be shallow with heights ranging from 235 m (H80) to 339 m ($R_{0.5}$).

320   Overall, the vertical profiles from the radiosonde observations and ERA5 correspond qualitatively well to each other, but the radiosonde observations capture more details and sharper vertical gradients in the profiles due to the higher vertical resolution of the radiosondes compared to ERA5. For example, ERA5 does not produce the local maxima in the wind speed at $z \approx 1250\,\mathrm{m}$ and $z \approx 2400\,\mathrm{m}$ at 12:00 UTC. ERA5 also diagnoses the BL height to have the same classical diurnal cycle as identified from the radiosonde soundings. In general, ERA5 agrees reasonably well with the radiosondes and at 00, 06 and 12 UTC on 7 May

325   is within the range of values estimated from the radiosondes. At 12 UTC, the ERA5 BL height is much higher than the $R_{0.25}$ method. Given that ERA5 BL height is estimated using a Richardson number approach also with a critical values of 0.25, this is at first surprising. This discrepancy is likely explained by the local maximum of wind speed at $z \approx 1250\,\mathrm{m}$ observed by the radiosonde but not visible in ERA5, which leads to differences in the mechanical production term in $R_{0.25}$ compared to ERA5.

4.2   BL height diagnosed from Microwave radiometer measurements

330   Next, we show an example of the temperature profiles and BL height derived from the microwave radiometer measurements and compare these to ERA5. Figure 3 shows the MWR and ERA5 temperature profiles along with a cloud profiling product generated from cloud radar and lidar measurements at the SMEAR II site (CLU, 2019), and the stability class timeseries on 16 June 2019. This day had no precipitation measured by the ground instruments that would affect the quality of the MWR measurements.

335   The temperature profile and diurnal evolution in ERA5 is in good agreement with the temperature measured from the MWR (Fig 3a-b). Some differences are present, for example, ERA5 shows a temperature inversion between 1.5 - 2 km a.g.l. at 12 - 16 UTC that is not present in the MWR temperature profile.

  The BL heights diagnosed by the MWR and ERA5 show a strong diurnal cycle. A shallow BL is identified from 00 - 04 UTC (3am - 7am local time), after which the BL grows quickly in depth to 1 - 1.5 km a.g.l. A similar evolution of the BL

340   height can be seen in the cloud classification product (Fig. 3c), where the growth of the BL can be seen by the increasing height of the "Aerosol & insects" and "Insects" categories. After 10 UTC (1pm local time), the MWR BL height mostly follows the 10 °C isotherm at an altitude of 1 km a.g.l, while the BL height from ERA5 is higher, typically at 1.5 km a.g.l.. The BL height in the cloud classification product is between 0.8 - 1.2 km a.g.l., indicated approximately by the top of the "Insects" category (Chandra et al., 2010; Franck et al., 2021), and resembles the MWR BL height evolution even though the MWR estimates

345   the BL height slightly higher. This suggests that between 10 - 17 UTC, ERA5 overestimates the BL height. At 17 UTC (8pm local time) the BL height determined from both the MWR and ERA5 show the rapid collapse of the convective BL into a stable BL, and after 17 UTC the BL height estimated in ERA5 is generally in good agreement with the measurements from the

MWR. However, after 22 UTC, the BL identified by the MWR is more than 1 km deeper than the BL diagnosed by ERA5 and furthermore, the MWR estimates of BL height vary significantly during the night.

350   The MWR also identifies whether the BL is stable or unstable (indicated by filled/unfilled markers in Fig 3a-b). During daytime the BL is identified as unstable and during night mostly stable, even though some points are identified as unstable, and the MWR stability identification follows closely the stability class identified from the EC observations (Fig. 3d). The MWR BL height values identified as unstable agree well with the ERA5 BL height values. However, for the stable values the agreement is poor, suggesting that the MWR does not provide accurate BL height estimates in stable conditions. This point is further

355 investigated in Section 5.3.

5   Systematic comparison of BL height diagnosed from different methods

In this section, a systematic, statistical comparison between the different estimates of BL height is presented. First, the four BL height estimates from the radiosonde data are compared to each other and then to ERA5 BL height estimates. Second, the diagnosed BL height from the MWR is compared to ERA5. In particular, we aim to determine how well the BL height in ERA5

360 represents observations and identify whether the level of agreement between observations and ERA5 depends on the observed surface-layer stability. To quantify the level of agreement between the observations and ERA5, we first calculate the Pearson's correlation coefficients (r, Chang and Hanna (2004)). In addition, to quantify any systematic bias between the two estimates of BL height the fractional bias (FB) is calculated using

$$FB = \frac{\overline{h_{obs}} - \overline{h_{era5}}}{0.5\left(\overline{h_{obs}} + \overline{h_{era5}}\right)} \tag{6}$$

365 where h_{obs} is the BL height estimate from observations (either the MWR or radiosondes) and h_{era5} is the BL height from ERA5. FB varies between -2 and +2 and has a perfect value of zero. If FB is positive, the observations have deeper BLs than ERA5 and vice-versa, if FB is negative observations have shallower BLs than ERA5. As a reference, $|FB| = 0.67$ and $|FB| = 0.4$ indicates that $\overline{h_{obs}}$ and $\overline{h_{era5}}$ differ by a factor of two and a factor of 1.5, respectively.

    The normalised root mean square error (nRMSE) is also calculated following Equation 3 in Chang and Hanna (2004):

370    $nRMSE = \sqrt{\dfrac{\overline{(h_{obs} - h_{era5})^2}}{\overline{h_{obs}} \cdot \overline{h_{era5}}}}.$ (7)

nRMSE is always positive and a value of zero would indicate a perfect fit between the observations and ERA5. In contrast to FB, the nRMSE quantifies all sources of error (random and systematic) and the normalisation makes the values non-dimensional and thus this is a measure of relative error, not absolute error.

5.1   Comparison of the 4 Radiosondes methods

375 Figure 4 compares the diagnosed BL height from the 4 radiosonde methods for all times in the BAECC campaign period (1st February 2014 - 13th September 2014) to each other. The corresponding statistical values (r, FB and nRMSE) are provided

in Table 1. The Pearson's correlation coefficient for each comparison is positive and statistically significant at the 95% level (Table 1) which was to be expected given all methods use the same data to diagnose the same variable. However, all panels in Figure 4 show considerable scatter indicating that the diagnosed BL height can depend strongly on which method is applied.

380   Most notable in Figure 4 is that the BL heights diagnosed by the H80 method differ the most from the other three methods. This is quantitatively demonstrated by the r values which are in the range 0.542 to 0.624 and are lower than the correlation coefficients between the LL10 method and the two Richardson number methods (0.70 and 0.72). The FB values calculated between the H80 method and all other methods are positive and exceed 0.5. This means that the BL heights diagnosed by the H80 method are consistently larger than those from the other three methods which is consistent with the case study shown

385   in Figure 2. Closer analysis reveals that in the case of deep (> 1km) BLs, the H80 method almost always over-estimates the height of the BL compared to the LL10 method and both Richardson number methods. This is very likely because of the strong inversion (2 K) required by the H80 method and because the H80 method was designed to incorporate turbulent mixing caused by buoyant thermals overshooting their level of neutral buoyancy.

When very shallow (< 100 m) BLs are considered, the H80 method estimates shallower BLs than the LL10 method (Fig.

390   4a) but deeper BLs than the $R_{0.25}$ method (Fig. 4c). Furthermore, when the H80 method diagnosed BL heights between 500 - 1000m, there is good agreement with the $R_{0.25}$ method. This shows that although the bias of the H80 method is on average positive, how well the diagnosed BL heights compare to those from other methods does depend on the height and therefore, potentially, the stability of the BL.

The LL10 method agrees relatively well with both Richardson number approaches (Fig. 4b, d) and in both cases r exceeds

395   0.7. In comparison to the $R_{0.25}$ method, the LL10 method diagnoses slightly deeper BLs on average (FB=0.186) and this positive bias is particularly pronounced for deeper BLs. In contrast, when the LL10 scheme diagnoses BL heights in the range 200 - 1000 m, the BL height estimate from the $R_{0.25}$ method is deeper. The FB between the LL10 and $R_{0.5}$ method is small and negative indicating that the LL10 method diagnoses slightly shallower BLs than the $R_{0.5}$ method. The colours in Figures 4a, b and d show the BL type identified by the LL10 scheme. Neutral and convective BLs (red) are systematically deeper than

400   the stable BLs (blue). In this study, only 2 out of 805 soundings were defined to be convective, whereas 597 (74%) where defined as neutral and 206 (26%) as stable by the LL10 scheme. This differs from the surface stability class estimated from eddy covariance data which, if only February to October is included to ensure a fair comparison with the radiosondes, shows that 38% of times are classed as very unstable or weakly unstable, 19% of times are near-neutral stable or near-neutral unstable and 43% are weakly stable or strongly stable.

405   Of all the comparisons, in terms of the FB and nRMSE the best agreement is between the LL10 and $R_{0.5}$ methods whereas the best correlation (r=0.96) is between the $R_{0.25}$ method and the $R_{0.5}$ method (Fig. 4f). The $R_{0.25}$ method gives systematically lower estimates of the BL height than the $R_{0.5}$ method (FB of -0.261) which is to be expected given that the Richardson number generally increases in height as the amount of turbulence decreases.

To further understand the differences between the 4 different radiosonde methods we now consider the distributions of

410   diagnosed BL height for each synoptic observation time (Fig. 5) and for different stability classes (Fig. 6). The shallowest BLs in all four methods are diagnosed at 00 UTC and the deepest at 12 UTC. At 00 UTC, the median value diagnosed by the H80

method is very similar to the $R_{0.25}$ method and smaller than the median values diagnosed by the LL10 and the $R_{0.5}$ methods. This is not the case at 06, 12 and 18 UTC where, similar to shown in Fig. 4, the H80 method has consistently deeper BLs. Figure 5 also indicates that the LL10 method has a narrower distribution of diagnosed BL heights at 00 UTC and especially at 415 06 UTC compared to the other 3 methods.

Analysing the relation between the BL heights diagnosed from the 4 different methods based on observed surface-layer stability allows a more physically-based understanding than by analysing this in terms of time of day. For very unstable conditions ($\zeta \leq$ -0.1, Fig. 6a), the median BL height values vary considerably, from ~600 m to over 1500 m, between the 4 different methods. The shallowest BLs, in terms of the mean, median and 3rd quartile are diagnosed by the $R_{0.25}$ method 420 closely followed by the $R_{0.5}$ method which diagnoses slightly deeper BLs. The LL10 scheme and especially the H80 method produce the deepest BLs when the surface layer is very unstable. In general, similar behaviour between the 4 schemes is observed for weakly unstable and near-neutral unstable conditions. The biggest difference is that the LL10 scheme is in better agreement with the Richardson number methods as the surface layer approaches neutral stratification.

For very stable conditions ($\zeta \geq$ 0.1, Fig. 6b), all methods diagnose very shallow BLs and the absolute differences are very 425 small. The H80 method has the largest number of outliers, and given this method occasionally diagnosed BL heights greater than 1 km when the eddy covariance observations show the surface layer is very stable, it suggests that what is identified as the BL top is likely the top of a residual layer. When weakly stable and near-neutral stable conditions are considered, the LL10 scheme has the shallowest BLs and the narrowest distributions. The H80 method and the $R_{0.5}$ method have the deepest BLs in terms of the median value whereas the H80 method has the largest variation in the diagnosed BL height. Thus, in conclusion the 430 H80 method produced the deepest BLs for all stability classes except for the very stable case. How the LL10 scheme compares to the two Richardson number methods depends strongly on stability class with the LL10 scheme diagnosing deeper BLs for unstable cases and shallower BLs for stable conditions.

5.2 Comparison of the radiosondes methods to ERA5

Figure 5 also compares each of the 4 radiosonde BL height estimates to ERA5 for 00, 06, 12 and 18 UTC and a quantitative 435 comparison is presented in Table 2. At 00 UTC the best agreement is obtained between ERA5 and the $R_{0.5}$ method. This is confirmed by a high correlation coefficient of 0.834 and very small nRMSE and FB values (Table 2). However, the distribution of BL heights in broader in the $R_{0.5}$ method compared to ERA5 and compared to the other three radiosonde methods. Good agreement (r=0.801) is found between the $R_{0.25}$ method and ERA5 at 00 UTC although the FB is negative demonstrating that on average ERA5 diagnoses systematically deeper BLs than the $R_{0.25}$ method at night. Poor agreement exists between 440 both the LL10 and H80 methods and ERA5 at 00 UTC. A positive FB is found between ERA5 and the H80 method at 00 UTC indicating that, on average, the H80 method diagnoses deeper BLs than ERA5 at night. In contrast, a negative FB occurs between the LL10 method and ERA5 at 00 UTC.

At 06 UTC, very similar behaviour is observed as at 00 UTC: the best agreement occurs between ERA5 and the $R_{0.5}$ method, good agreement is present between ERA5 and the $R_{0.25}$ method, and poor agreement between H80 and LL10 and 445 ERA5. Figure 5 shows that while the LL10 and the two Richardson number methods have BL height distributions which are

largely similar (in terms of their median and inter quartile range (IQR) values) the H80 method has a much broader distribution (large IQR) and larger mean BL height than ERA5.

At 12 UTC, the ERA5 BL height distribution agrees well with the two Richardson number methods and the LL10 method (Fig. 5). Statistically the best agreement at 12 UTC is between the LL10 method and ERA5 whereas the worst agreement is 450 between the H80 method and ERA5 (Table 2). At 12 UTC all methods except the $R_{0.25}$ method have a positive FB meaning that these three method diagnose deeper BLs than ERA5 (see also Fig. 5). At 18 UTC, the BL heights diagnosed from both the H80 and LL10 scheme agree very poorly with ERA5 estimates (correlation coefficients less than 0.01) and the FB values are very large and positive; the H80 method has a FB of 1.087 which corresponds to an over-estimation of BL height by more than a factor of 3 compared to ERA5. In contrast, much better agreement occurs between ERA5 and both Richardson number 455 methods with the best agreement between ERA5 and the $R_{0.25}$ method. The case study shown in Fig. 2 at 18 UTC shows very similar behaviour.

Given that ERA5 uses a Richardson number approach to diagnose BL height with a critical Richardson number of 0.25, we would expect the best agreement between the $R_{0.25}$ method applied to radiosonde data and ERA5. However, this is only the case at 18 UTC. Furthermore, systematic bias is evident between the $R_{0.25}$ method applied to radiosonde data and ERA5 that 460 changes sign during the day. At 00 and 06 UTC, when BLs are typically shallow, ERA5 slightly overestimates the BL height in comparison to the $R_{0.25}$ radiosonde method. This may indicate that ERA5 struggles to resolve shallow BLs due to the limited vertical resolution. However, ERA5 also overestimates (but typically only by ∼10%) the BL height at 12 UTC when BLs are deep. In contrast, at 18 UTC, ERA5 has shallower BL heights than the $R_{0.25}$ applied to radiosonde data.

Figure 6 compares the different radiosonde methods to ERA5 but for different stability classes, identified using the EC 465 observations (section 2.4). For the most unstable BLs, very good agreement exists between the $R_{0.5}$ method and ERA5. ERA5 estimates deeper BLs than the $R_{0.25}$ method (negative FB, Table 3), shallower BLs than the LL10 method (by approximately 23%) and much shallower (almost by a factor of 2) BLs than the H80 method. The IQR of the ERA5 BL distribution is very similar to that from the $R_{0.25}$ method whereas the other three radiosonde methods, and in particular the H80 method, have larger IQRs and thus more variability in the diagnosed depth of the BL in very unstable conditions. For weakly unstable and 470 near-neutral unstable BLs, (Fig. 6a) the best, and very good, agreement exists between the $R_{0.25}$ method and ERA5 (Table 3). ERA5 predicts shallower BLs than all 4 radiosonde methods (the FB is positive for all methods - Table 3): the bias is very small for the $R_{0.25}$ method, moderate for the $R_{0.5}$ and LL10 methods and very large for H80.

For near-neutral stable and weakly stable BLs (Fig. 6b), again ERA5 agrees best with the $R_{0.25}$ method. The correlation coefficient is high (0.793) and the FB is very close to zero. ERA5 diagnoses shallower BLs than the $R_{0.5}$ method, and much 475 shallower BLs than the H80 method, for near-neutral stable and weakly stable BLs – similar to what was found for near-neutral unstable and weakly unstable BLs. In contrast, ERA5 predicts deeper BLs (negative FB) compared to those from the LL10 scheme for near-neutral stable and weakly stable BLs which is opposite of what was found for the three unstable classes. Furthermore, the LL10 scheme and ERA5 have very small correlation coefficients that are not statistically significant which suggests the LL10 scheme and ERA5 are not in good agreement for near-neutral stable and weakly stable BLs.

480   For the most stable BLs, all schemes and ERA5 diagnose the shallowest BLs of all stability classes. ERA5 has a larger IQR and thus more variability than the 4 radiosonde methods (Fig. 6b), however the number of outliers for all radiosonde methods is greater than in all other stability classes. For these very stable cases, there is a large negative FB between ERA5 and both Richardson number methods indicating that ERA5 significantly over-estimates the BL height. In particular, the FB between the $R_{0.25}$ method and ERA5 is -0.971 meaning that, on average, BLs in ERA5 are more than twice as deep as detected from the

485 $R_{0.25}$ method applied to the radiosonde observations. However, despite the systematic bias and the known challenges models have representing stable BLs correctly, there is still a reasonable agreement between ERA5 and the two Richardson number methods — both have correlation coefficients of ~0.6. The mean BL height predicted by ERA5 is smaller than that from the H80 method, and the FB also indicates that ERA5 has shallower BLs than the H80 method for the very stable class but this result is strongly influenced by the number of outliers. A small, although still statistically significant, correlation coefficient

490 exists between H80 and ERA5 indicating a poor level of agreement overall. Similarly, a very small correlation coefficient exists between the LL10 scheme and ERA5 (as was the case for the other stable classes) indicating that in very stable conditions the LL10 scheme diagnoses the BL height very differently to ERA5.

    In summary, when ERA5 is compared to the $R_{0.25}$ method we conclude that ERA5 accurately predicts the BL height in the majority of situations. The most notable exceptions are that ERA5 overestimates the height of stable BLs and for weakly

495 unstable BLs and at 18 UTC ERA5 has notable shallower BLs than the $R_{0.25}$ method. This suggests that ERA5 lacks the vertical resolution to fully resolve very shallow stable BLs and that ERA5 likely does not capture the evening transition well. However, as long as these two limitations are considered, ERA5 can be used as a basis for a long-term climatology of BL height at Hyytiälä.

5.3   Microwave Radiometer and ERA5

500 To compare the MWR BL heights, reported approximately every 10 minutes, with the ERA5 values output as hourly values, a 1-hour median of the MWR BL height values was used. Additionally, as was the case for the radiosonde comparison to ERA5, the surface stability classes derived from the eddy covariance observations are used to bin the MWR data and to determine the effect of the surface-layer stability on the agreement between MWR and ERA5 BL height values.

    Figure 7a shows the distributions of both the MWR and ERA5 BL height values, categorised according to the surface

505 stability class. For the unstable surface stability classes, the distributions agree well. Interestingly, the correlation coefficient in the very unstable class is lower than for the weakly unstable and near-neutral unstable surface stability classes (Table 4). The highest correlation (0.65) and smallest nRMSE (0.016) was achieved in the weakly unstable class (Table 4). The FB is small and positive for the most unstable class, almost zero for the weakly unstable class, and small and negative for the near-neutral unstable class (Table 4). This means that for very unstable BLs the MWR diagnoses deeper BLs than ERA5 but for less

510 unstable BLs the MWR has shallower BLs than ERA5.

    However, when moving towards stable conditions, the level of agreement between the MWR BL heights and ERA5 decreases and the number of outliers increases, especially in the MWR distribution. For the near-neutral stable case, the correlation coefficient is only 0.19 and negative correlations exist for the weakly stable and very stable bins (Table 4). The FB is large and

negative for the near-neutral stable case indicating that the MWR identified shallower BLs than ERA5 in these situations. In the very stable stability class, the MWR BL height values are notably larger than the ERA5 BL height values; the third quartile value of the BL heights in ERA5 is less than the first quartile in the MWR BL heights (Fig. 7a) and the FB is large and positive (Table 4). Furthermore, the MWR distribution is very broad with a large IQR and is much broader than the corresponding distribution from ERA5 (Fig. 7a). These statistical results from the very stable case support our finding from the case study (Section 4.2) that the height of the BL diagnosed by the MWR under stable conditions can be significantly overestimated. Similarly, Jiang et al. (2021) found that the MWR stable BL algorithm overestimated the BL height significantly compared to BL heights determined from ceilometer data in cases where the temperature differences between the ground and 1 km altitude was very small, indicating a strongly stable BL. Jiang et al. (2021) concluded that the overestimation was due to limitations in the MWR algorithm used to determine the stable BLH.

To further investigate the limitations in the MWR BL height algorithm, we filter the data so that all BL height values estimated when the MWR algorithm assesses the BL to be stable are removed (Fig. 7b). This filtering removes outliers from the distributions in the stable surface stability classes and improves the level of agreement between ERA5 and the MWR in almost all surface stability classes (Table 4). Notably, now the distribution of the MWR BLH values in the very stable surface stability class is narrow and the distributions of MWR and ERA5 BLH values span approximately the same altitude interval. However, the correlation coefficient in this class, while positive and larger than for the unfiltered dataset, is no longer statistically significant at the 95 % confidence level due to small sample size. The correlation increases also in the unstable surface stability classes, with the highest increase (0.164) occurring in the very unstable class. This suggests that the MWR has misclassified the stability of some measurements. Additionally, after removing the stable MWR BLH values, the fractional bias is negative in all surface stability classes. That is, the unstable BL height from the MWR observations is consistently lower than in ERA5. However, the high correlation and the low FB and nRMSE values for the unstable surface stability classes suggest that although the Richardson number method applied to ERA5 data, and the parcel method applied to MWR data differ, the diagnosed BL heights are consistent.

Figure 8 shows the timeseries of the MWR BL height hourly median values (calculated only from the unstable MWR BL height values) along with the ERA5 BL height values. The surface stability class (determined from EC observations) is indicated by the colour of the MWR circle marker. For clarity, the surface stability classes have been divided only into unstable and stable compound classes, with very, weakly and near-neutral classes combined. The seasonal cycle of the BL height is seen clearly in Figure 8. For both datasets, the BL heights from April until September show high peaks and clear diurnal cycles. Note that the stable BL height values from the MWR have been filtered out, so most of the MWR BL heights during summer nights are not seen in Figure 8. In the summer (June to August), ERA5 tends to estimate visibly deeper BLs during daytime than the MWR. This is confirmed if we calculate the FB for each month individually (not shown); a negative FB is found in June, July and August between ERA5 and the MWR BL heights. Furthermore, the case study of June 16 (Section 4.2) also showed that ERA5 estimated deeper daytime BLs than the MWR. In winter, the BL height identified by both ERA5 and the MWR is shallow and the variation during the day is much smaller than in summer. Good agreement, especially in December–January, is seen between the MWR and ERA5 BL heights in the winter months

The MWR and ERA5 BL heights display similar diurnal cycles throughout the timeseries, except for during October 2018

550   where noticeable differences exist both in BL height as well in the diurnal cycle. This poor agreement can mostly be accounted for by two factors, both of which cause the MWR BL height estimation to fail. Firstly, the temperature profile retrieval can fail. Reasons for this include a wet radome or light drizzle that is not detected by the precipitation sensors used to filter rain cases. Additionally, the retrieval can be affected by inconsistent cloud base temperature retrievals from the infrared radiometer due to scattered cloud cover. This effect can be seen on 1 October 2018 (Fig. S6) and can be identified by the inconsistent (noisy)

555   temperature profile. Note that in this case, even though the precipitation sensors used to remove rain cases did not capture the drizzle, the weather station included in the MWR was able to flag some of the temperature profiles during the case as affected by rain (not shown), but the flag did not appear in the BL height values that were consequently accepted. In this case, the ERA5 BL height follows the insect layer and boundary layer cloud development seen in the cloud classification product closely and thus performs better.

560   Secondly, the BL height retrieval algorithm can fail to identify the BL height even though the temperature profiles are visually sound. An example of this is seen on 28 October (Fig. S7), where the MWR BL height has a clear diurnal cycle peaking during the afternoon, but the values are distinctly too high. Even though the surface stability indicates an unstable surface layer during the day, the MWR algorithm is not capable of determining the BL height correctly in this case. Again here, the ERA5 BL height follows the maximum altitude of aerosol echo in the cloud classification product and provides a

565   much more realistic estimate of the BL height.

Notably, the surface stability class does not seem to determine whether the MWR succeeds or fails in identifying the BL height during the autumn and winter months (Fig. 8). In October 2018 several of the days where MWR BL overestimates the BL height are actually classified as unstable by the EC measurements and as such, as we have shown previously, the MWR should provide good BL height estimates. And to the contrary, several days with mostly stable surface stability, for example

570   22-23 November 2018, show good agreement and similar diurnal cycles between the ERA5 and MWR BL height values. This indicates that in addition to the stability of the boundary layer, local meteorological conditions should be considered carefully and the MWR BL height values should be compared to other data sources when deciding whether they are trustworthy or not.

6   Boundary-layer height and surface-layer stability climatology at Hyytiälä

Unfortunately, no long-term measurements of BL height from either radiosondes nor remote sensing instruments are available

575   from Hyytiälä, as is the case for many long-term surface stations. However, from the results presented in section 5, we conclude that ERA5 represents the depth of the BL at Hyytiälä very well except under very stable conditions and thus can be used as the basis for a long-term climatology. To better understand the meteorological causes of the BL height annual and diurnal climatology presented in this section, we also present a climatology of surface-layer stability class as determined from the eddy covariance observations.

580   The annual cycle of the BL height in ERA5 is shown in Figure 9a. The shallowest boundary layers occur in winter (December, January, February - DJF), with the lowest monthly median value of 353 m occurring in February, which is the statistically

the coldest month monthly in southern Finland; the mean February temperature at Hyytiälä averaged over 1979 to 2019 is -7.4°C. The full distribution of BL height at the 4 standard synoptic times (00, 06, 12 and 18 UTC, Figure S1) shows that the distribution of BL heights is positively skewed and has a similar shape at all times. In DJF, it is rare that the BL height exceeds

585 1 km and very shallow boundary layers (i.e. in the lowest bin 0 - 125 m in the histogram in Figure S1), are less common than moderately shallow boundary layers (i.e those in the 2nd to 4th bins, 125 m - 500 m).

During spring (March, April, May - MAM), there is a rapid increase in the BL height and the deepest BLs of all months (largest extremes) occur in May (Fig. 9a). This is despite May not being the warmest month of the year; the monthly mean temperature in May at SMEAR II is 9.1°C in May compared to 16.1°C in July. In addition, the variability in the BL height

590 in May is very large. During April and May, the mean BL height is larger than the median indicating that the BL height distribution is less Gaussian and more positively skewed. This is confirmed by Figure S1 which shows that, particularly at 00 and 18 UTC, the distribution of BL height in MAM is strongly positively skewed and also that the most common BL height is very shallow - less than 125 m.

The highest median value of the BL height occurs in June (576 m) and similarly to May, there is a high degree of variability

595 in the BL height (Fig. 9a). The variability is primarily related to the diurnal cycle — all hours of the day are included in Figure 9a. There is also a notable decrease in the median BL height across the three summer months (June-July-August, JJA) which is potentially caused by the pronounced decrease in the length of day and thus the incoming solar radiation between June and August: there is 19 hr 30 min of daylight on 21 June and 15 hr 19 minutes on 21 August. The full distributions of BL height in JJA (Fig. S1) also indicate that the distribution at 12 UTC in JJA is less skewed than in MAM and the peak is shifted to higher

600 values.

In autumn (September, October and November, SON), the median BL height is similar across all of these three months (Fig. 9a), however, the 75th percentile, the extremes (whiskers) and the variability are all smaller in November than in September. Particularly, in October and November, the mean and median BL heights are similar suggesting that the BL height distributions are approximately Gaussian. Figure S1 shows that the BL height distributions in the autumn months are more Gaussian than

605 in spring or summer but that even in autumn the BL height distribution has a small positive skew. Figure S1 also indicates that very shallow BLs are rare in autumn even at night. This, and other aspects of the diurnal cycle, are considered shortly.

The annual cycle in the observed surface-layer stability class (Fig. 9b) can also be considered and related to the annual cycle in BL height. To a first order approximation, we would expected deeper BLs to be more unstable than shallow BLs. The annual climatology shows that the unstable classes are most common in the warm season and the stable classes most common

610 in the cold season (Fig. 9b). The occurrence of all three unstable classes (all red colours) is almost constant from April until July whereas August, usually regarded as a summer month, has fewer times that are classified as unstable. The increase in unstable classes starting in April is consistent with the ERA5 BL height shown in Figure 9a which shows April is the month where deep BLs start to become evident. Figure 9b also shows that the very unstable class has a strong annual cycle peaking (18.6%) in July and rarely (<2.1%) occurring between November and February. Similarly, the weakly unstable class also has

615 a pronounced annual cycle but is most common in May (26.6%) which is the same month that the deepest BLs develop. The near-neutral unstable class occurs rarely in all months (the largest occurrence of 10.1% occurs in February) and it has a weak

annual cycle. The three different stable stability classes all have different annual cycles. The near-neutral stable class occurs most frequently in the cold season and least frequently in summer and is more common than the near-neutral unstable class, particularly in November and December when 20.6% and 18.2% of all times are classified as near-neutral stable. The weakly

620   stable class has an annual cycle with two peaks, the largest peak in October (27.8%) and a smaller secondary peak in March (21.1%) and a minimum in May (14.2%). Interestingly, the very stable stability class is most common in the warm season and least common in the cold season. A caveat to the annual cycle in stability class is that the number of missing data points also has a strong annual cycle with many more missing observations during winter than summer, mostly likely related to icing of the sonic anemometer.

625   The mean diurnal cycle of the BL height for each month is shown in Figure 10a. In all months the maximum mean BL height occurs at 12 or 13 UTC (2 or 3pm local time depending on season). This shows that the radiosondes released at 12 UTC most likely do measure the deepest BLs which is not the case for all locations worldwide when the standard synoptic times of 00 and 12 UTC often do not coincide with the deepest BLs. The largest diurnal cycle in mean BL height occurs in May, closely followed by June, and is caused by both the large maximum values during the day and the small minimum values at night. The

630   variation in the diurnal cycle of BL height in May is shown in Figure 11a. During daytime, there is a large variation in potential BL heights with maximum values almost reaching 3 km however 50% of BLs at 12 UTC in May have a height between 1100 and 1900 m. The median values are similar to the mean values during day time as the distribution is broad and symmetric (Fig. S1) but the mean values are greater than the median BL height at night as the distribution is highly non-Gaussian and instead resembles a Gamma distribution.

635   November, December and January have very small diurnal cycles and in December there is no clear diurnal cycle at all which is directly related to the very small diurnal cycle in incoming solar radiation in December. Figure 11b further emphasises the lack of any diurnal cycle as even the variability in the BL height is almost constant with time of day.

Figure 10a shows that the shallowest night time BLs do not occur in winter and Figure S1 demonstrates that very shallow, night-time BLs occur more often in spring and summer than in winter. This may contradict expectations where it is often

640   assumed that the shallowest BLs develop during the coldest part of the year. However, shallow BLs in winter and in summer are caused by different processes. In spring and summer, clear calm nights are needed and spring likely has more shallow BLs than summer as summer nights are short (maximum 19 h 40 min of daylight in Hyytiälä), and hence there is not much time for a shallow BL to develop. In winter, cold conditions are required for shallow BLs to develop. Winter likely has more shallow BLs than Autumn due to the fact that ground is snow covered in winter and there is less solar radiation in winter which gives

645   more time for inversions to develop.

Figure 10b shows the diurnal cycle of the stability class for all months together. As expected the very unstable and weakly unstable classes have a pronounced diurnal cycle peaking between 11 - 13 UTC and rarely occurring at night. The very stable and weakly stable classes have the opposite diurnal cycle and it is rare that weakly or strongly stable conditions occur during the day. Noticeably, there is no diurnal cycle in the amount of missing data which may suggest that missing data is not

650   strongly influenced by stability which does have a strong diurnal cycle. By considering all months at once, Figure 10b may hide interesting details and therefore we also consider the diurnal cycle of stability class for May (Fig. 11c) and December

separately (Fig. 11d). In May, almost all times with valid measurements between 6 and 16 UTC (9am and 7pm local time) are classed as unstable with the majority falling in the weakly unstable class. Between 21 UTC and 4 UTC, the majority of times are classed as near-neutral stable. In December, there is a more pronounced diurnal cycle in the stability class than in the BL

655   height. The near-neutral stable class has the largest diurnal cycle peaking at night and is rare during the day whereas the very unstable class is exceptionally rare in December. Of note is that at night shallower BLs occur in May than in December despite that the occurrence of the very stable class is slightly larger at night in December in comparison to May.

However overall, the diurnal cycle of ERA5 diagnosed BL heights is largely consistent with the diurnal cycle in the observed stability (i.e deeper BLs occur more often when the surface layer is unstable and shallower BLS are more prevalent when

660   observations indicate a stable surface layer). This, in addition to the statistical comparison of ERA5 against the radiosonde and MWR measurements presented in section 5 gives further confidence that ERA5 can resolve the BL height at Hyytiälä well.

7   Spatial variability of BL height in ERA5

So far we have only considered the BL height from one single grid point from ERA5. However, the BL height is likely to vary spatially, for example, due to changes in surface land use. The aim of this section is to determine how much the BL height

665   varies in the region surrounding Hyytiälä depending on both month and time of day. Firstly, this enables us to determine over how large an area the now ongoing BL height measurements made with the MWR at Hyytiälä are representative of. Secondly, this also enables us to determine if the conclusions drawn in sections 5 and 6 are also valid over the surrounding areas.

To assess the spatial variability of the BL height we use a subset of the global ERA5 reanalysis data set. Spatially, we consider a considerable part of Northern Europe (see Fig. 12) and analyse the time period from January 1979 to December

670   2019 using ERA5 data with 1-hour temporal resolution. This data set thus consists of time series of BL height from many grid points. For each calendar month, we calculate the Pearson correlation coefficients (averaged over all 42 years) between the time series of BL height extracted at the Hyytiälä location (and extensively analysed in this study) and the time series of the BL height from each of the individual grid cells within the selected area. Here we assume that larger areas of high spatial correlations indicate lower spatial variability, and conversely smaller areas of high correlations indicate high variability within

675   the BL height field. As a limit of 'high' correlation, we set the correlation coefficient to be greater than 0.75.

When all hours of the day are considered (shading in Fig. 12) the highest correlations are closest to Hyytiälä and correlations exceeding 0.75 cover most land areas in southern and central Finland. There is a moderate annual cycle with the largest areas of high correlation occurring between April and July and the smallest areas in the winter months. May has the largest area of high correlations which, in addition to southern and central Finland, also extends to some small regions of Sweden and large parts

680   of Estonia. This means that, when all hours of the day are considered, the spatial variability of the BL height is slightly higher in the cold season compared to spring and early summer. It also indicates that in May, when many observation campaigns take place in Hyytiälä (e.g. Laakso et al., 2007; Lampilahti et al., 2021), measurements of BL height are representative of a relatively large area. The difference in the BL height over sea and lakes compared to over land is also evident in Fig. 12 as there are no sea areas where the correlation exceeds 0.75.

685   The spatial variability of the BL height also depends on the time of day (Figs. S2 - S5. Figure 12 also shows the extent of the "high" correlation but only when the 12 UTC values of BL height are considered. This daytime variability of the BL height has a strong annual cycle with highest variability (smallest area of high correlation) during the period from May to September and the lowest variability in the cold season. This corresponds to a reduced spatial representativeness of daytime observations of BL height at Hyytiälä in the summer compared to in the winter. The high degree of spatial variability at 12 UTC in summer

690   could be attributed to deep convective BLs which can be highly variable and are much more strongly influenced by the surface heat flux and thus the surface type and amount of incoming radiation (and hence cloud cover) than stable BLs which are often also influenced by shear driven turbulence. When only BL height values at 00 UTC are considered (Fig. S2), the annual cycle is much smaller than at 12 UTC. Furthermore, between November and March the area with high correlations is about the same size at all 4 standard synoptic times (Figs. S2 - S5) which is consistent with the lack of diurnal cycle in ERA5 BL heights at

695   Hyytiälä (Figs. 10a, 11a,b)

       Although using ERA5 estimates of the BL height indicates the spatial variability of the BL height to some extent, these estimates should be taken with utmost care as they generally underestimate the true variability. This is due to the spatial resolution of ERA5, which is about 31 km, and thus does not allow for accurate representation of small-scale land surface features. These irregularities, such as small lakes or patches of different vegetation affect the evolution of the BL and increase its

700   spatial variability. A more accurate estimate of spatial variability would require a data set with much higher spatial resolution, however, this is not within the scope of this current study.

8   Discussion and Conclusions

In this study, we examined the BL height and surface stability at the SMEAR II station, Hyytiälä in southern Finland. A systematic, statistical comparison between four different methods of diagnosing the BL height from radiosonde data and ERA5

705   data was presented and BL height estimates from a microwave radiometer (MWR) were presented and also compared to ERA5. Due to no time overlap between the radiosondes and the MWR data, these two observation types could not be compared. A unique aspect of our comparison is that we quantified the effect of surface-layer stability on how well different methods agree with each other. When the 4 different methods applied to radiosonde data were compared, there were positive statistically significant correlations between all methods however considerable scatter was present meaning that the diagnosed BL height

710   can depend strongly on the method used. On average, the BL diagnosed from the H80 method was consistently deeper than the BL diagnosed from the other three schemes and this was especially true when deep BLs were considered. This is in agreement with Seidel et al. (2010) who found that methods which use the vertical gradient of potential temperature to identify the BL height (as H80 does), produce the larger values of BL height than other methods.

       The level of agreement between the 4 methods applied to radiosonde data depends strongly on the stratification of the surface

715   layer. For the very unstable classification, the median value of the BL height varied from ∼600 m to over 1500 m again with the H80 method producing the deepest BLs. While the LL10 method diagnoses deeper BLs than both Richardson number methods for unstable cases, the opposite is true for stable cases. The deeper BLs in almost all situations diagnosed by the H80

method very likely arise due to the very strong (2 K) inversion required by this method and that the BL top is defined to be at the level where the potential temperature is 2 K warmer than the base of the inversion. Physically, this definition means that the H80 method considers the impact of over-shooting thermals. This means that the H80 method is appropriate for diagnosing the maximum potential depth that mixing can occur over. However, the H80 method is not physically robust for stable BLs and is best suited to be applied only to convective situations.

A good degree of correlation ($r > 0.6$) is found between ERA5 and the two Richardson number methods at all synoptic times and for almost all stability classes. This good agreement between ERA5 and, in particular, the $R_{0.25}$ method suggests that the high vertical resolution of ERA5 (24 model levels below 1.5 km and grid spacing of 25 - 120 m in the BL) appears to be sufficient to capture the BL structure in most situations. This is a key advance over previous reanalysis such as ERA-Interim which only had 12 levels below 1.5 km. However, for the very stable class ERA5 estimates deeper BLs than the $R_{0.25}$ method and the correlation is poor. Thus, ERA5 still cannot capture the depth of very stable BLs accurately, which is likely due to deficiencies in the BL parameterization or lack of resolution.

In contrast, poor correlation ($r < 0.6$) exists between ERA5 and H80 at all synoptic times and all stability classes. ERA5 and LL10 only agree well at 12 UTC and for the very unstable and weakly unstable case. This is likely due to the method that LL10 uses to determine the type of BL and in particular the value of the stability threshold, δ_s, which is 1 K. When Liu and Liang (2010) developed this method and tested it using radiosondes from 14 field campaigns, they found that typically 60% of soundings were classed as neutral. The number of neutral cases found in this study is larger (74%), and much larger than the percentage of neutral cases suggested by the observed surface stability classes (19%). This suggest that the LL10 neutral category includes convective BLs and also potentially stable BLs as well. Misclassifying the BL as neutral when it is convective has no impact on the diagnosed BL height whereas misdiagnosing a stable BL as neutral does as the presence of low-level jets is not considered which leads to the over-estimation of the BL height. We suggest that in the future a systematic analysis of the impact of the inversion strength threshold on the diagnosed BL height should be conducted and that this threshold may vary from location to location.

It is also notable that for the H80, LL10 and the $R_{0.50}$ method the worst comparison with ERA5 in terms of r, FB and nRMSE occurs at 18 UTC. This indicates that defining the BL height during the collapse of the convective BL and transition to more stable conditions is challenging and potentially also that ERA5 struggles to capture this process accurately. Users of ERA5 BL height should be cautious using BL height estimates during the evening transition period. A caveat to the comparison between ERA5 and the radiosonde based estimates of BL height is that although the BAECC radiosondes were not assimilated into ERA5, soundings from the nearby operational stations of Jokioinen ($\sim$ 125 km SSW of Hyytiälä) and Jyväskylä ($\sim$ 100 km NW of Hyytiälä) were.

This study has also shown that BL height estimates from the MWR agree reasonably well with ERA5, and thus very likely represent the "true" BL height well, but only in certain conditions. For unstable situations when the BL is well-mixed, the MWR derived BL height agrees well with ERA5 and thus we deem it a reliable product. ERA5 does overestimate the BL height compared to MWR during June-August 2018, however, given that ERA5 is a model based data set, it is not clear based on our data whether the MWR also overestimates the "true" BL height. A key outcome of our analysis is that the

MWR does not reliably estimate the BL height under stable conditions, which at Hyytiälä, occur commonly at night between April and September. We hypothesise that this is due to the algorithm used by the MWR software under stable conditions and that potentially users could improve on this. Furthermore, some of the errors in the MWR diagnosed BL height arise as the classification of the stability type (stable vs unstable) fails; checking the accuracy of this classification using alternative observations is recommended. Finally, we identified that the MWR algorithm can also fail to identify the BL height under unstable conditions, as identified from the EC measurements, if there is fog/drizzle or scattered cloud cover. This is particularly true in autumn and winter.

As ERA5 agrees well with observations of BL height, a climatology of the annual and diurnal cycles in BL height determined from ERA5 was computed and presented alongside a climatology of the observed surface-layer stability. The shallowest monthly median BL height occurred in February and it is rare that the BL height exceeds 1 km in December-February. In winter there is almost no diurnal variation in the BL height; days are short, cloudy conditions are common and given the very high zenith angle incoming solar radiation is weak. Manninen et al. (2018) also note that there is almost no diurnal cycle in the occurrence of turbulence during winter at Hyytiälä. The surface layer is rarely unstable in winter, due to limited incoming solar radiation and thus weak surface heat fluxes. Also in winter, the very stable class rarely occurs and is less common than the weakly or near-neutral stable cases. The lack of strongly stable layers in winter is likely due to the prevalence of cloud which would limit surface nocturnal cooling. This hypothesis is supported by the results of Manninen et al. (2018) who showed that most of the nighttime turbulence during winter in Hyytiälä is associated with cloud.

During spring, the height of the diagnosed BL rapidly increases with deeper BLs occurring in May than in either July or August which are warmer months. Also of note is that the shallowest BLs of anytime occur at night in spring and consequently, the BL height has a very large diurnal cycle in April and May. The very shallow BLs occur at night in spring because often high pressure and clear skies prevail allowing strong radiative cooling to occur and, in comparison to summer, the nights are long enough to enable stable conditions and shallow BLs to become established.

The BL height had a maximum median value in June and then decreased during the remains of the summer due to the decrease in the length of the day. Very shallow BLs were less common at night in summer than in spring, possibly because nights are shorter. In autumn, the diurnal cycle of BL height is weaker than spring or summer. One potential reason is that synoptic-scale weather patterns and thermal advection may have a stronger influence on the BL height in Autumn compared to other times of year as strong extra-tropical cyclones are more common in autumn than spring or summer (Laurila et al., 2021). Strong cold-air advection over a warmer surface can lead to strong surface heat fluxes and deep and well mixed BLs (Sinclair et al., 2010).

Finally, an estimate of the spatial representativity of the measurements of BL height made at SMEAR II is made based on ERA5 reanalysis data. The analysis shows that, when all hours of the day are considered, the BL height at Hyytiälä is representatitive of most land areas in southern and central Finland. However, the spatial variability of the BL height depends on the time of day especially between May and September. The area which the Hyytiälä BL height values are representative of is much smaller at 12 UTC than 00 UTC and also smaller in summer than in winter.

The results presented here highlight the difficulty in accurately measuring the BL height and demonstrate that to have reliable and accurate estimates in all conditions, a range of measurements is needed. Thus, if BL height estimates are used to better understand surface based measurements of trace gases or aerosol particles, an appreciation of these challenges and knowledge of likely sources of error, and under what conditions they primarily occur, is necessary. Furthermore, this study has shown the large annual, seasonal and spatial variability in the BL height in a high latitude, yet fairly spatially homogeneous and flat location. Lastly, it is encouraging to report that modern reanalysis products with high temporal and spatial resolution can capture the BL height and its evolution well in most situations and, in the absence of observations, can be used with confidence.

Data availability. The measurements from the SMEAR II station are available at the SmartSMEAR portal (https://avaa.tdata.fi/web/smart/smear). The observations from the FMI automatic weather station used in this study are available at the FMI open data portal (https://en.ilmatieteenlaitos.fi/open-data). The data from the BAECC campaign are available at the Atmospheric Radiation Measurement (ARM) User Facility portal (https://www.arm.gov/research/campaigns/amf2014baecc). The ERA5 reanalysis data are available at the Copernicus Climate Data Store portal (https://cds.climate.copernicus.eu/). The ECOCLIMAP-SG land cover map is available from https://opensource.umr-cnrm.fr/projects/ecoclimap-sg. The ground-based remote-sensing data (the cloud profiling products shown in Figures 3c, S6c and S7c) used in this article are generated by the European Research Infrastructure for the observation of Aerosol, Clouds and Trace Gases (ACTRIS) and are available from the ACTRIS Data Centre using the following link: https://hdl.handle.net/21.12132/1.d28c281574434f92.

Author contributions. VAS jointly conceived the study, wrote the majority of the manuscript, performed the climatological analysis presented in section 6 and contributed to the interpretation of all results. JR performed the analysis and interpretation of the microwave radiometer data and contributed to the writing of sections 3.2, 4.2 and 5.3. GB analysed the eddy covariance data, computed the stability classes and contributed to the interpretation of these results. IS compared the radiosonde BL height estimates to those from ERA5 data. YB designed, performed and interpreted the spatial representativeness analysis presented in section 7 and contributed to writing section 7. DM jointly conceived the study, provided expert knowledge about the microwave radiometer data, and provided guidance on the interpretation of all results. MK selected and analysed the radiosonde case study and contributed to the interpretation of all results. All authors provided comments on drafts of the manuscript

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. This work was partially funded by Academy of Finland Flagship funding (grant no. 337549). We acknowledge the open data policy of the National Land Survey of Finland as the present study uses the following data from the National Land Survey of Finland: Elevation model 10 m, NLS orthophotos, Cadastral index map. We acknowledge ACTRIS for providing the CLU (2019) dataset in this study, which was produced by the Finnish Meteorological Institute, and is available for download from https://cloudnet.fmi.fi/.

References

Baars, H., Ansmann, A., Engelmann, R., and Althausen, D.: Continuous monitoring of the boundary-layer top with lidar, Atmos. Chem. Phys., 8, 7281–7296, https://doi.org/10.5194/acp-8-7281-2008, 2008.

Bäck, J., Aalto, J., Henriksson, M., Hakola, H., He, Q., and Boy, M.: Chemodiversity of a Scots pine stand and implications for terpene air concentrations, Biogeosciences, 9, 689–702, https://doi.org/10.5194/bg-9-689-2012, 2012.

Chandra, A. S., Kollias, P., Giangrande, S. E., and Klein, S. A.: Long-Term Observations of the Convective Boundary Layer Using Insect Radar Returns at the SGP ARM Climate Research Facility, Journal of Climate, 23, 5699–5714, https://doi.org/10/c8smk3, 2010.

Chang, J. C. and Hanna, S. R.: Air quality model performance evaluation, Meteorology and Atmospheric Physics, 87, 167–196, https://doi.org/https://doi.org/10.1007/s00703-003-0070-7, 2004.

CLU: Cloud profiling product Classification; 2019-06-16, https://hdl.handle.net/21.12132/1.d28c281574434f92, generated by the cloud profiling unit of the ACTRIS Data Centre, 2019.

Collaud Coen, M., Praz, C., Haefele, A., Ruffieux, D., Kaufmann, P., and Calpini, B.: Determination and climatology of the planetary boundary layer height above the Swiss plateau by in situ and remote sensing measurements as well as by the COSMO-2 model, Atmos. Chem. Phys., 14, 13 205–13 221, https://doi.org/10.5194/acp-14-13205-2014, 2014.

ECMWF: Part IV: Physical Processes, chap. 4, pp. 1–190, IFS Documentation, ECMWF, operational implementation 12 May 2015, 2015.

Eerdekens, G., Yassaa, N., Sinha, V., Aalto, P. P., Aufmhoff, H., Arnold, F., Fiedler, V., Kulmala, M., and Williams, J.: VOC measurements within a boreal forest during spring 2005: on the occurrence of elevated monoterpene concentrations during night time intense particle concentration events, Atmos. Chem. Phys., 9, 8331–8350, https://doi.org/10.5194/acp-9-8331-2009, 2009.

Franck, A., Moisseev, D., Vakkari, V., Leskinen, M., Lampilahti, J., Kerminen, V.-M., and O'Connor, E.: Evaluation of convective boundary layer height estimates using radars operating at different frequency bands, Atmospheric Measurement Techniques Discussions, pp. 1–27, https://doi.org/10.5194/amt-2021-142, 2021.

Granados-Muñoz, M., Navas-Guzmán, F., Bravo-Aranda, J., Guerrero-Rascado, J., Lyamani, H., Fernández-Gálvez, J., and Alados-Arboledas, L.: Automatic determination of the planetary boundary layer height using lidar: One-year analysis over southeastern Spain, J. Geophys. Res., 117, https://doi.org/10.1029/2012JD017524, 2012.

Hari, P. and Kulmala, M.: Station for Measuring Ecosystem-Atmosphere Relations (SMEAR II), Boreal Env. Res., 10, 315–322, 2005.

Hari, P., Nikinmaa, E., Pohja, T., Siivola, E., Bäck, J., Vesala, T., and Kulmala, M.: Station for measuring ecosystem-atmosphere relations: SMEAR, in: Physical and physiological forest ecology, pp. 471–487, Springer, 2013.

Heffter, J. L.: Air Resources Laboratories atmospheric transport and dispersion model (ARL-ATAD), https://www.osti.gov/biblio/5015173, 1980.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, https://doi.org/10.1002/qj.3803, https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.3803, 2020.

Holzworth, G. C.: Estimates of Mean Maximum Mixing Depths in the Contiguous United States, Mon. Wea. Rev., 92, 235–242, https://doi.org/10/bdt589, 1964.

Jiang, Y., Xin, J., Zhao, D., Jia, D., Tang, G., Quan, J., Wang, M., and Dai, L.: Analysis of Differences between Thermodynamic and Material Boundary Layer Structure: Comparison of Detection by Ceilometer and Microwave Radiometer, Atmos. Res., 248, 105 179, https://doi.org/10/gjpp32, 2021.

855 Laakso, L., Grönholm, T., Kulmala, L., Haapanala, S., Hirsikko, A., Lovejoy, E. R., Kazil, J., Kurten, T., Boy, M., Nilsson, E. D., et al.: Hot-air balloon as a platform for boundary layer profile measurements during particle formation, Boreal Env. Res., 2007.

Lampilahti, J., Leino, K., Manninen, A., Poutanen, P., Franck, A., Peltola, M., Hietala, P., Beck, L., Dada, L., Quéléver, L., et al.: Aerosol particle formation in the upper residual layer, Atmos. Chem. Phys., 21, 7901–7915, 2021.

Laurila, T. K., Gregow, H., Cornér, J., and Sinclair, V. A.: Characteristics of extratropical cyclones and precursors to windstorms in Northern
860 Europe, Weather Clim. Dynam. Discuss, https://doi.org/https://doi.org/10.5194/wcd-2021-44, in review, 2021.

Lauros, J., Nilsson, E. D., Dal Maso, M., and Kulmala, M.: Contribution of mixing in the ABL to new particle formation based on observations, Atmos. Chem. Phys., 7, 4781–4792, https://doi.org/10.5194/acp-7-4781-2007, 2007.

Liu, S. and Liang, X.-Z.: Observed diurnal cycle climatology of planetary boundary layer height, J. Climate, 23, 5790–5809, 2010.

Löhnert, U. and Maier, O.: Operational Profiling of Temperature Using Ground-Based Microwave Radiometry at Payerne: Prospects and
865 Challenges, Atmospheric Measurement Techniques, 5, 1121–1134, https://doi.org/10/gb8r4b, 2012.

Mammarella, I., Peltola, O., Nordbo, A., Järvi, L., and Rannik, Ü.: Quantifying the uncertainty of eddy covariance fluxes due to the use of different software packages and combinations of processing steps in two contrasting ecosystems, Atmospheric Measurement Techniques, 9, 4915–4933, 2016.

Manninen, A. J., Marke, T., Tuononen, M., and O'Connor, E. J.: Atmospheric boundary layer classification with Doppler lidar, J. Geophys.
870 Res., 123, 8172–8189, 2018.

Moreira, G. de A., Guerrero-Rascado, J. L., Bravo-Aranda, J. A., Foyo-Moreno, I., Cazorla, A., Alados, I., Lyamani, H., Landulfo, E., and Alados-Arboledas, L.: Study of the Planetary Boundary Layer Height in an Urban Environment Using a Combination of Microwave Radiometer and Ceilometer, Atmospheric Research, 240, 104 932, https://doi.org/10/gh6f3d, 2020.

Ouwersloot, H., Vil à-Guerau de Arellano, J., Nölscher, A., Krol, M., Ganzeveld, L., Breitenberger, C., Mammarella, I., Williams, J., and
875 Lelieveld, J.: Characterization of a boreal convective boundary layer and its impact on atmospheric chemistry during HUMPPA-COPEC-2010., Atmos. Chem. Phys., 12, 9335–9353, 2012.

Petäjä, T., O'Connor, E. J., Moisseev, D., Sinclair, V. A., Manninen, A. J., Väänänen, R., von Lerber, A., Thornton, J. A., Nicoll, K., Petersen, W., Chandrasekar, V., Smith, J. N., Winkler, P. M., Krüger, O., Hakola, H., Timonen, H., Brus, D., Laurila, T., Asmi, E., Riekkola, M.-L., Mona, L., Massoli, P., Engelmann, R., Komppula, M., Wang, J., Kuang, C., Bäck, J., Virtanen, A., Levula, J., Ritsche, M., and
880 Hickmon, N.: BAECC: A Field Campaign to Elucidate the Impact of Biogenic Aerosols on Clouds and Climate, Bulletin of the American Meteorological Society, 97, 1909–1928, https://doi.org/10.1175/BAMS-D-14-00199.1, 2016.

Rose, T., Crewell, S., Löhnert, U., and Simmer, C.: A Network Suitable Microwave Radiometer for Operational Monitoring of the Cloudy Atmosphere, Atmospheric Research, 75, 183–200, https://doi.org/10/fbfkh7, 2005.

RPG Radiometer Physics GmbH: RPG-Radiometer Physics GmbH, https://www.radiometer-physics.de/, 2014a.
885 RPG Radiometer Physics GmbH: Instrument Operation and Software Guide: Operation Principles and Software Description for RPG standard single polarization radiometers (G5 series), https://www.radiometer-physics.de/download/PDF/Radiometers/HATPRO/RPG_MWR_STD_Software_Manual%20G5.pdf, 2014b.

Seibert, P., Beyrich, F., Gryning, S.-E., Joffre, S., Rasmussen, A., and Tercier, P.: Review and intercomparison of operational methods for the determination of the mixing height, Atmospheric Environment, 34, 1001–1027, 2000.

890   Seidel, D. J., Ao, C. O., and Li, K.: Estimating climatological planetary boundary layer heights from radiosonde observations: Comparison of methods and uncertainty analysis, J. Geophys. Res., 115, 2010.

Seidel, D. J., Zhang, Y., Beljaars, A., Golaz, J.-C., Jacobson, A. R., and Medeiros, B.: Climatology of the planetary boundary layer over the continental United States and Europe, J. Geophys. Res., 117, 2012.

Sinclair, V. A., Belcher, S. E., and Gray, S. L.: Synoptic controls on boundary-layer characteristics, Boundary-layer meteorology, 134, 387–

895   409, 2010.

Sivaraman, C., McFarlane, S., Chapman, E., Jensen, M., Toto, T., Liu, S., and Fischer, M.: Planetary Boundary Layer (PBL) Height Value Added Product (VAP): Radiosonde Retrievals, Tech. rep., Department of Energy Office of Science Atmospheric Radiation Measurement (ARM) Program, DOE/SC-ARM/TR-132, 2013.

Solheim, F., Godwin, J. R., Westwater, E. R., Han, Y., Keihm, S. J., Marsh, K., and Ware, R.: Radiometric profil-

900   ing of temperature, water vapor and cloud liquid water using various inversion methods, Radio Science, 33, 393–404, https://doi.org/https://doi.org/10.1029/97RS03656, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/97RS03656, 1998.

Stull, R. B.: An Introduction to Boundary Layer Meteorology, vol. 13, Springer Science & Business Media, 1988.

Tokay, A., Wolff, D. B., and Petersen, W. A.: Evaluation of the New Version of the Laser-Optical Disdrometer, OTT Parsivel2, Journal of Atmospheric and Oceanic Technology, 31, 1276 – 1288, https://doi.org/10.1175/JTECH-D-13-00174.1, https://journals.ametsoc.org/view/

905   journals/atot/31/6/jtech-d-13-00174_1.xml, 2014.

Vickers, D. and Mahrt, L.: Quality control and flux sampling problems for tower and aircraft data, Journal of atmospheric and oceanic technology, 14, 512–526, 1997.

Vogelezang, D. H. P. and Holtslag, A. A. M.: Evaluation and model impacts of alternative boundary-layer height formulations, Boundary-Layer Meteorology, 81, 245–269, 1996.

910   von Engeln, A. and Teixeira, J.: A planetary boundary layer height climatology derived from ECMWF reanalysis data, J. Climate, 26, 6575–6590, 2013.

Zhang, W., Guo, J., Miao, Y., Liu, H., Song, Y., Fang, Z., He, J., Lou, M., Yan, Y., Li, Y., et al.: On the summertime planetary boundary layer with different thermodynamic stability in China: A radiosonde perspective, J. Climate, 31, 1451–1465, 2018.

Table 1. Statistical measures between BL height estimates. r is the Pearson's correlation coefficient and nRMSE is the Normalised Root Mean Square Error. All correlation coefficient are statistically significant at the 95% level. The best value in each column is highlighted in bold. The variables listed as the x-variables are taken as the observations in Equation 6 and those listed as the y-variables take the place of the model in Equation 6. Thus, if the x-variable has larger BLH than the y-variable, then the FB > 0. If the x-variable has smaller BLH than the y-variable, then the FB < 0.

x-variable	y-variable	Pearson's r	fractional Bias (FB)	nRMSE
Heffter	Liu Liang	0.624	0.585	0.612
Heffter	$R_{0.25}$	0.542	0.769	0.833
Heffter	$R_{0.50}$	0.578	0.530	0.550
Liu Liang	$R_{0.25}$	0.701	0.186	0.187
Liu Liang	$R_{0.50}$	0.721	**-0.076**	**0.076**
$R_{0.25}$	$R_{0.50}$	**0.960**	-0.261	0.263

Table 2. Statistical measures between BL height estimates from radiosonde measurements and ERA5. r is the Pearson's correlation coefficient, FB is fractional bias and nRMSE is the Normalised Mean Square Error. Values were calculated for different time of the day. Radiosonde data is taken as the observations in Equation 6. Correlation coefficients which are statistically non-significant at the 95% level are marked in italics. For each measure, the best value between the four different datasets is marked in bold.

Time of day	Heffter vs ERA5			Liu Liang vs ERA5			$R_{0.25}$ vs ERA5			$R_{0.50}$ vs ERA5		
	r	FB	nRMSE	r	FB	nRMSE	r	FB	nRMSE	r	FB	nRMSE
00 UTC	0.372	0.170	0.170	0.253	-0.277	0.280	0.801	-0.318	0.322	**0.834**	**0.089**	**0.089**
06 UTC	0.193	0.581	0.607	0.417	-0.365	0.371	0.679	-0.303	0.307	**0.711**	**-0.007**	**0.007**
12 UTC	0.539	0.499	0.515	**0.794**	**0.071**	**0.071**	0.645	-0.107	0.107	0.696	0.083	0.083
18 UTC	*0.092*	1.087	1.295	*0.001*	0.588	0.615	**0.644**	**0.193**	**0.194**	0.622	0.493	0.509
All	0.478	0.637	0.671	0.605	**0.079**	**0.079**	0.754	-0.111	0.111	**0.770**	0.153	0.153
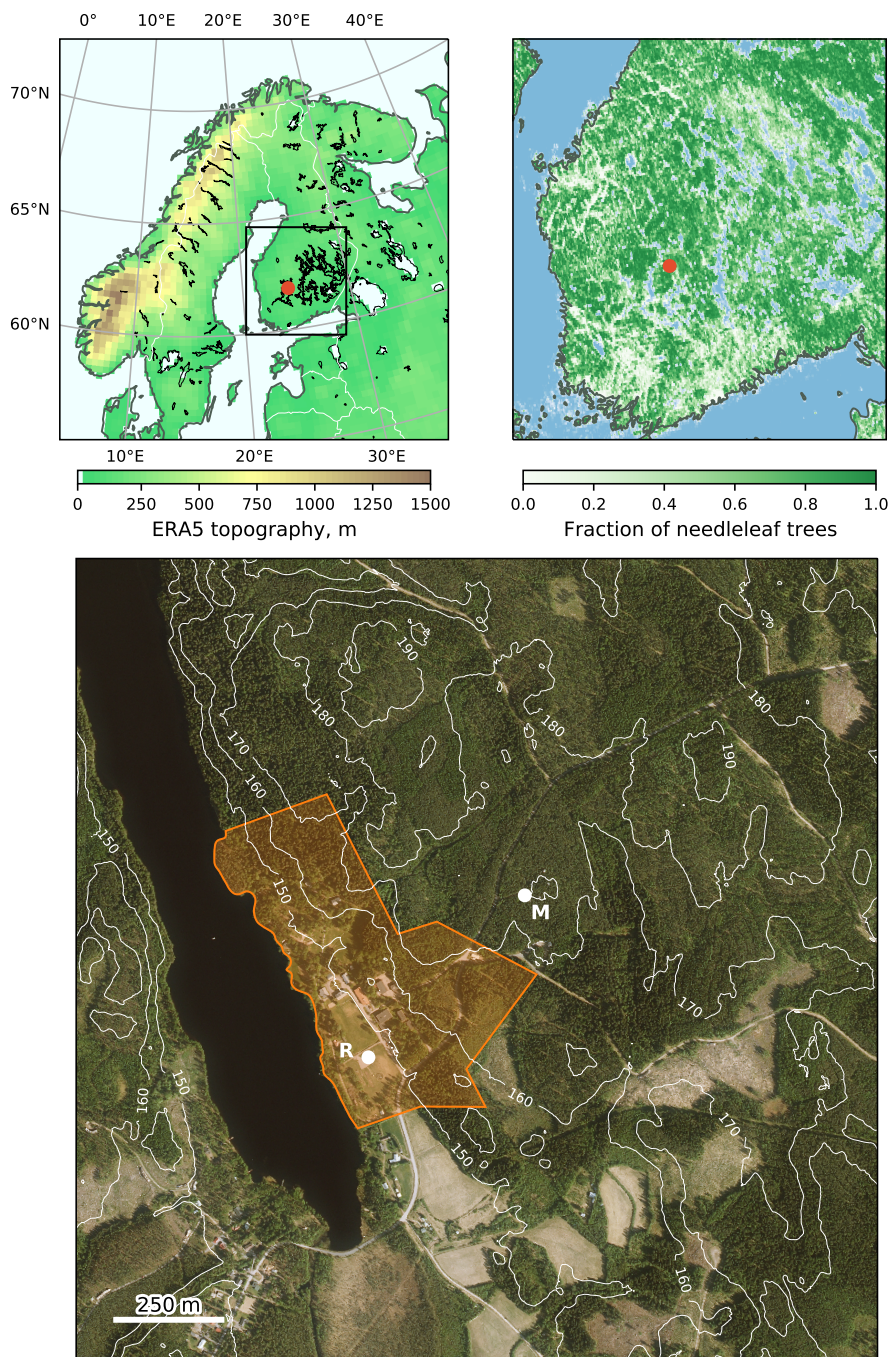
Figure 1. Location of Hyytiälä as shown by a red dot in the top two panels which show the topography (m) from ERA5 (left) and the fraction of needle leaf trees obtained from ECOCLIMAP-SG land cover map (right). The bottom panel shows the local area around Hyytiälä. The local topography (obtained from the National Land Survey of Finland) is shown in white contours (contour interval 10 m) and the area of the station is marked in orange. The location of the MWR is marked by an R and the mast where the eddy covariance measurements are made is marked by an M.
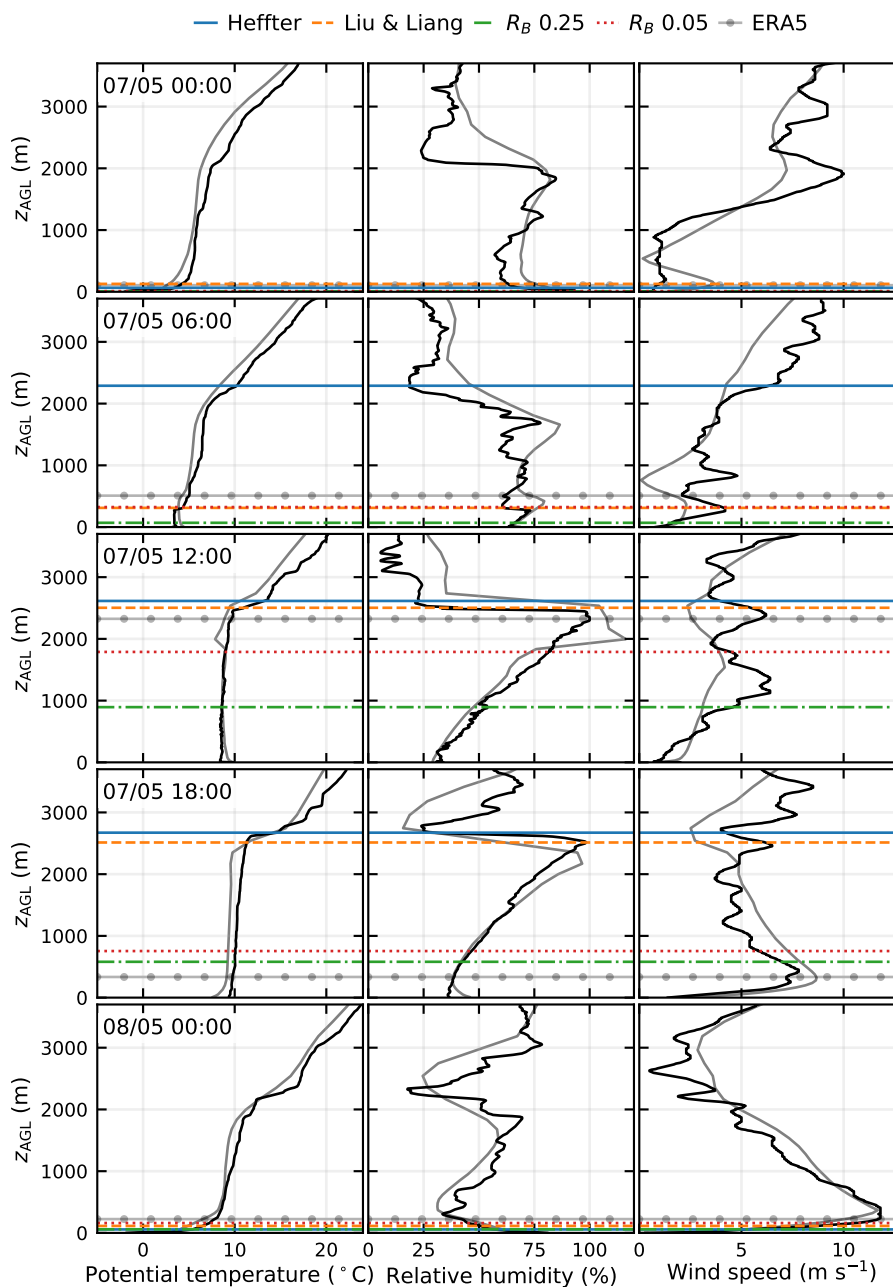
Figure 2. Vertical profiles of air temperature (left column), relative humidity (middle column) and horizontal wind speed (right column) from radiosonde soundings (black solid line) and ERA5 (grey solid line) and the estimates of BL height using different methods at SMEAR III on 7-8 May 2014. The BL height is estimated using the methods by Heffter (1980) and Liu and Liang (2010) as well as using a threshold value for the bulk-Richardson number (R_B) of 0.05 and 0.25. z_{AGL} stands for height above ground level. Times are given in UTC.
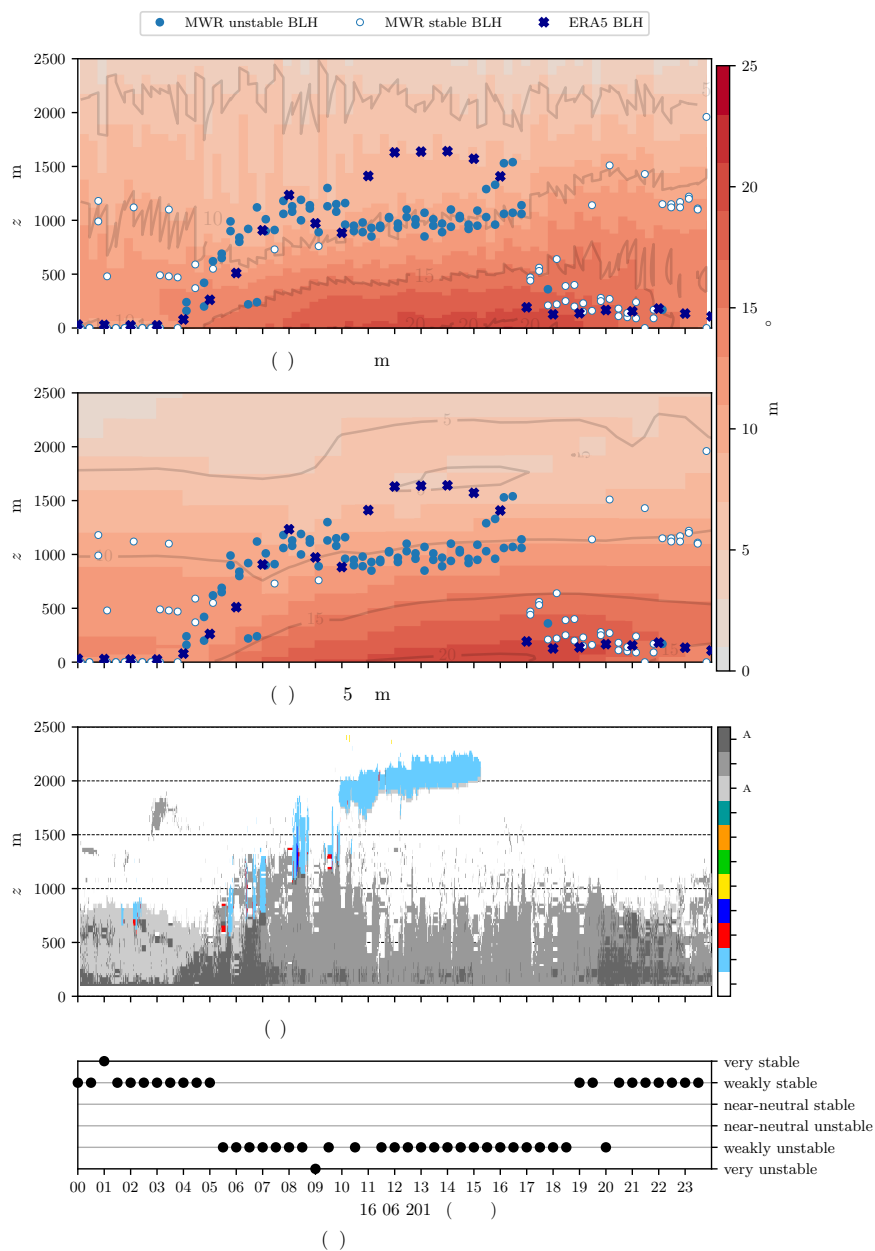
Figure 3. (a) Temperature profiles from MWR measurements and (b) ERA5, (c) a cloud profiling product (CLU, 2019) and (d) the stability regimes for 16 June 2019. Panels (a-b) have the MWR unstable BL height (filled circle), MWR stable BL height (empty circle), and ERA5 BL height (cross) plotted. The grey lines in panels (a-b) denote isotherms in 5 °C intervals. z_{AGL} stands for height above ground level.
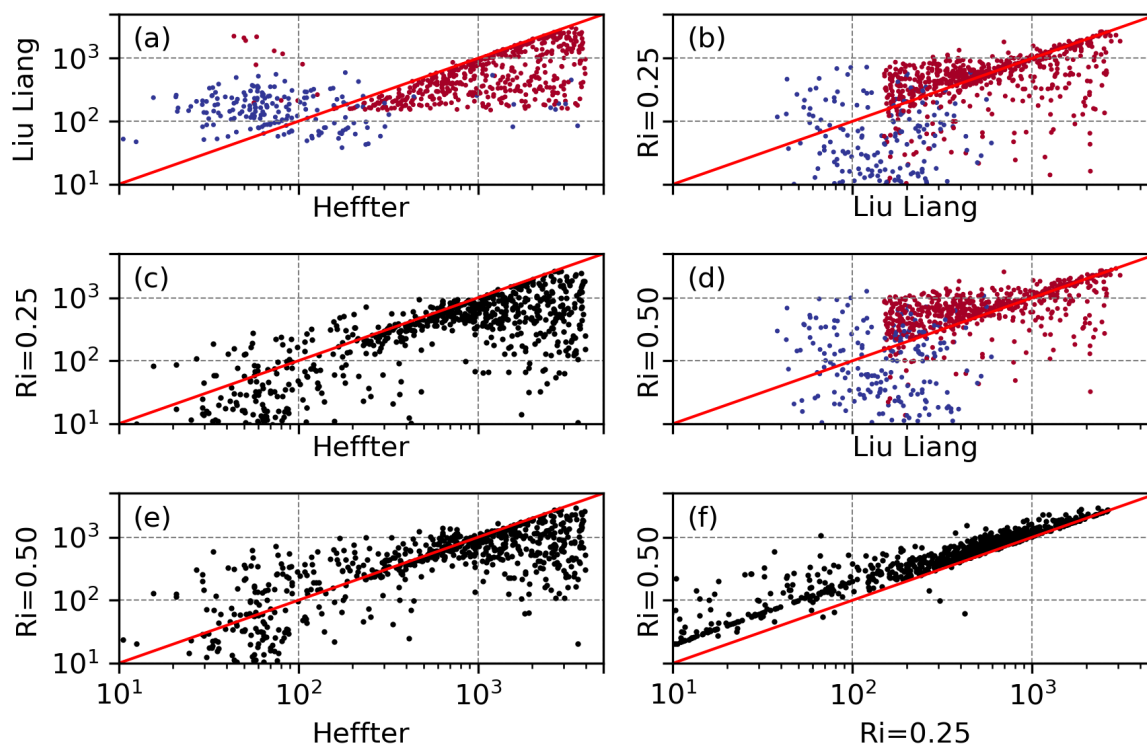
Figure 4. Scatter plots showing the relation between the BL height diagnosed from radiosondes taken during BAECC using 4 different methods. Data is from 1st February 2014 - 13th September 2014. In panels a, b and d, red points are when the LL10 scheme has diagnosed either a neutral or unstable BL and blue points are stable BLs. Red solid lines shows the 1-to-1 line. Note the logarithmic scale
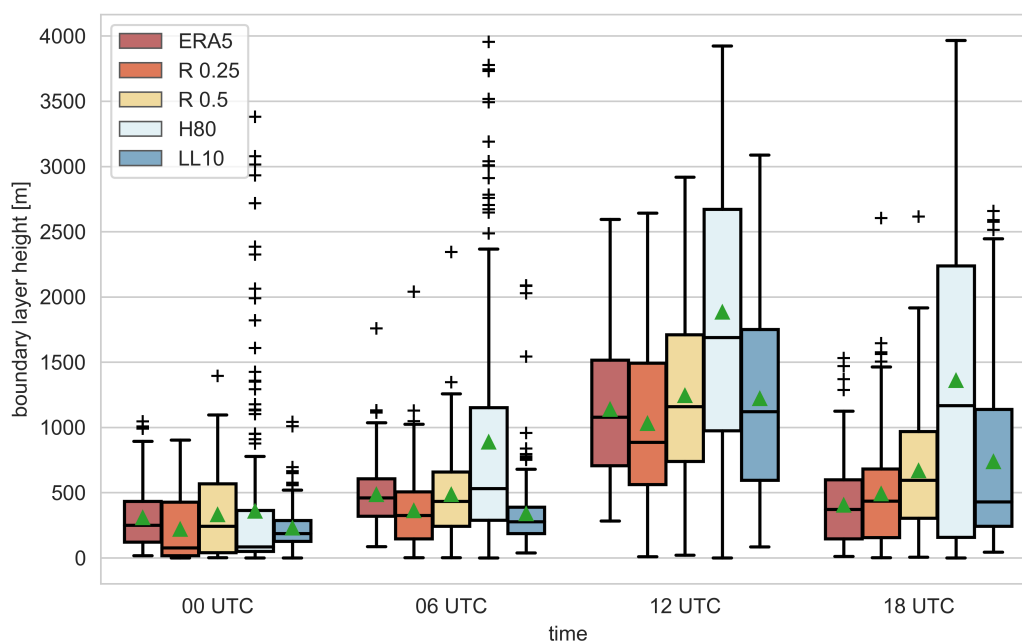
Figure 5. Boundary layer height distributions from ERA5 and from the 4 different methods applied to the radiosondes divided by time of the day. The boxes in each time bin are colour coded according to the legend. The shaded boxes extends from the first quartile (Q1) to the third quartile (Q3) values of the data i.e the interquartile range (IQR). The black solid line shows the median and the green triangles the mean values. The whiskers extend to Q1 - 1.5 IQR and Q3 + 1.5 IQR. Black points beyond the whiskers show the outliers.
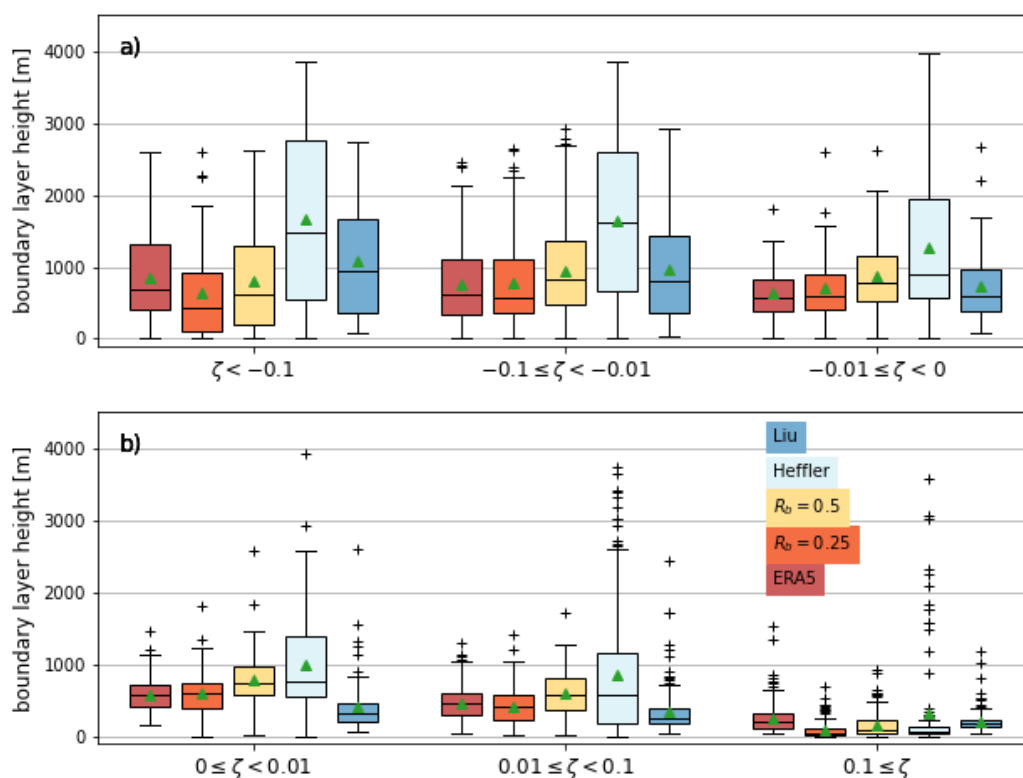
Figure 6. Boundary-layer height distributions from ERA5 (red) and from the 4 different methods applied to the radiosondes divided by stability class computed from the EC observations. Unstable classes are shown in (a) and stable classes in (b). The boxes in each stability bin are color coded according to the legend and the boxes and whiskers are defined as in Fig. 5
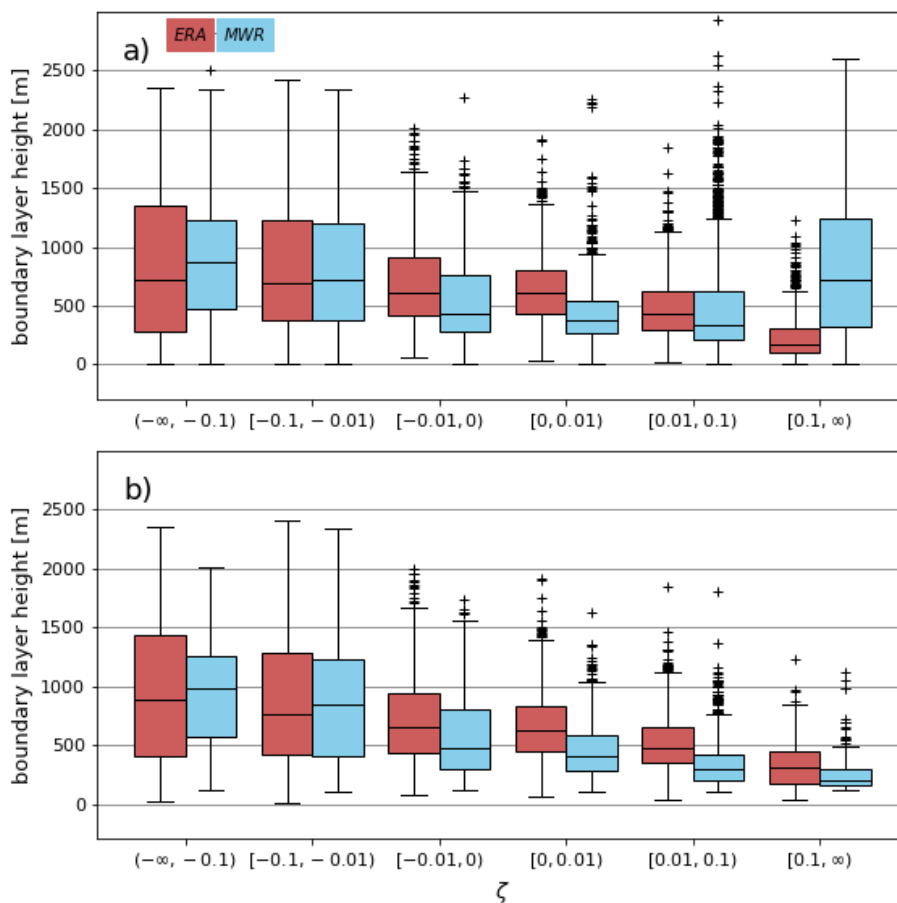
Figure 7. Boundary layer height distributions from ERA5 (red) and from the MWR (blue) divided by the stability class computed from the EC observations. The boxes and whiskers are defined as in Fig. 5. Panel (a) includes all MWR data. Panel (b) includes only times when the MWR has determined the BL to be unstable.
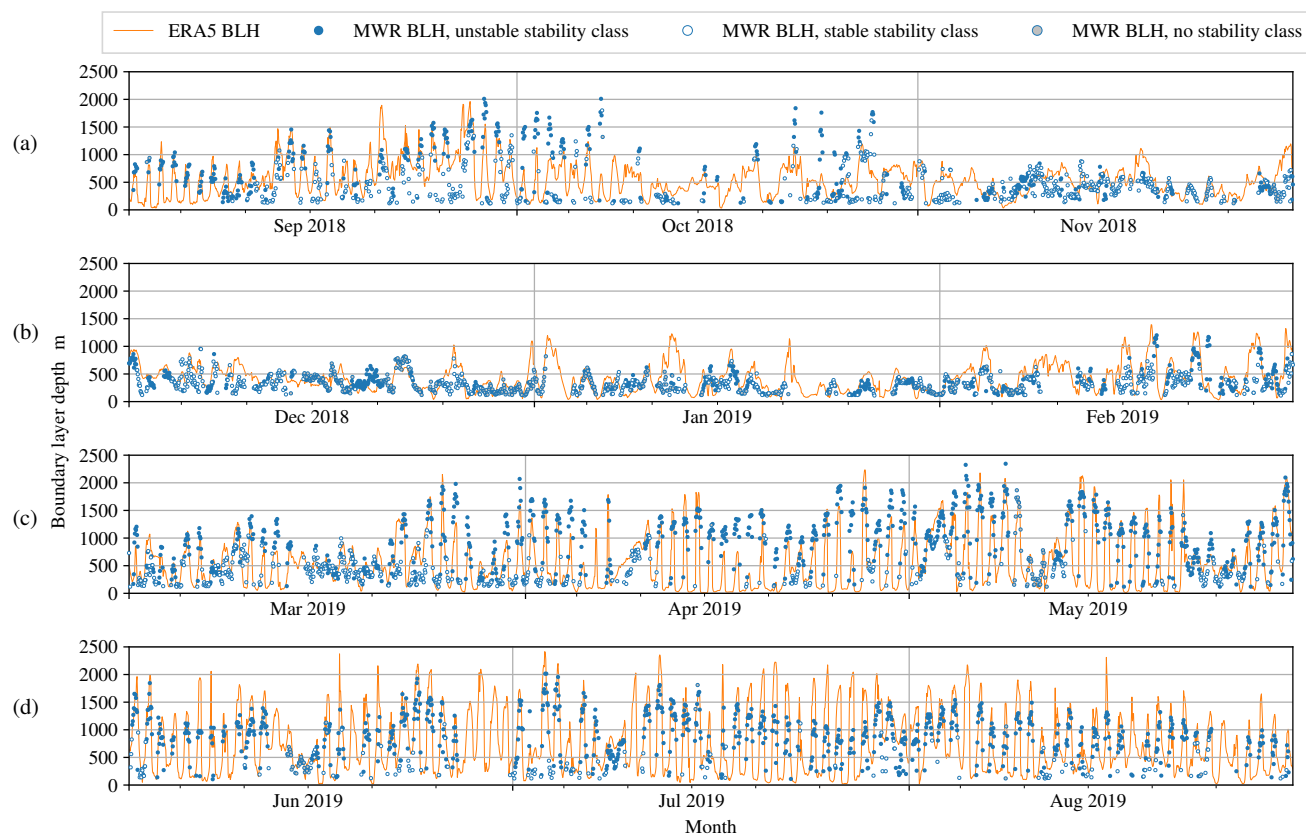
Figure 8. Timeseries of ERA5 and MWR boundary layer heights for (a) September 2018 – November 2018, (b) December 2018 – February 2019, (c) March – May 2019, and (d) June – August 2019. The orange line indicates the ERA5 BLH. The circles indicate the hourly median values of MWR unstable BLH; the times stability class indicates unstable BL (blue filled circles), the times the stability class indicated a stable BL (empty circles), and the times when stability class was missing (grey filled circles). Vertical lines denote the change of month, and ticks denote the 5th, 10th, 15th, 20th, and 25th days of the month.
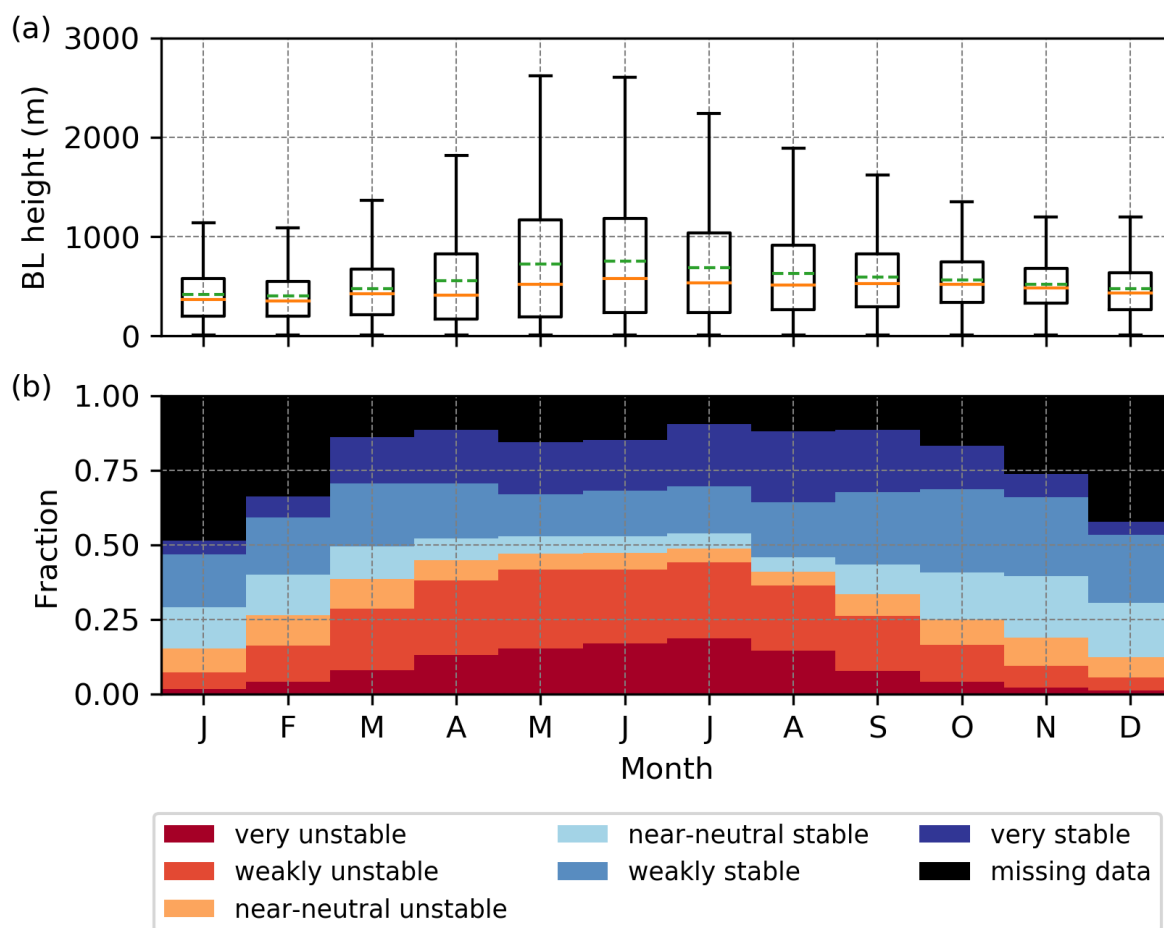
Figure 9. Annual mean cycles of (a) the ERA5 diagnosed BL height and (b) stability class. In (a) ERA5 data at the grid point closest to Hyytiälä is plotted. Orange lines show the median values, green lines the mean value. Boxes represent the 25th to 75th percentile (interquartile range, IQR) and whiskers extend to the 25th percentile minus 1.5 IQR and the 75th percentile plus 1.5 IQR. Outliers are not plotted. In (b) the fraction of each time that each stability class is observed is plotted for each month.
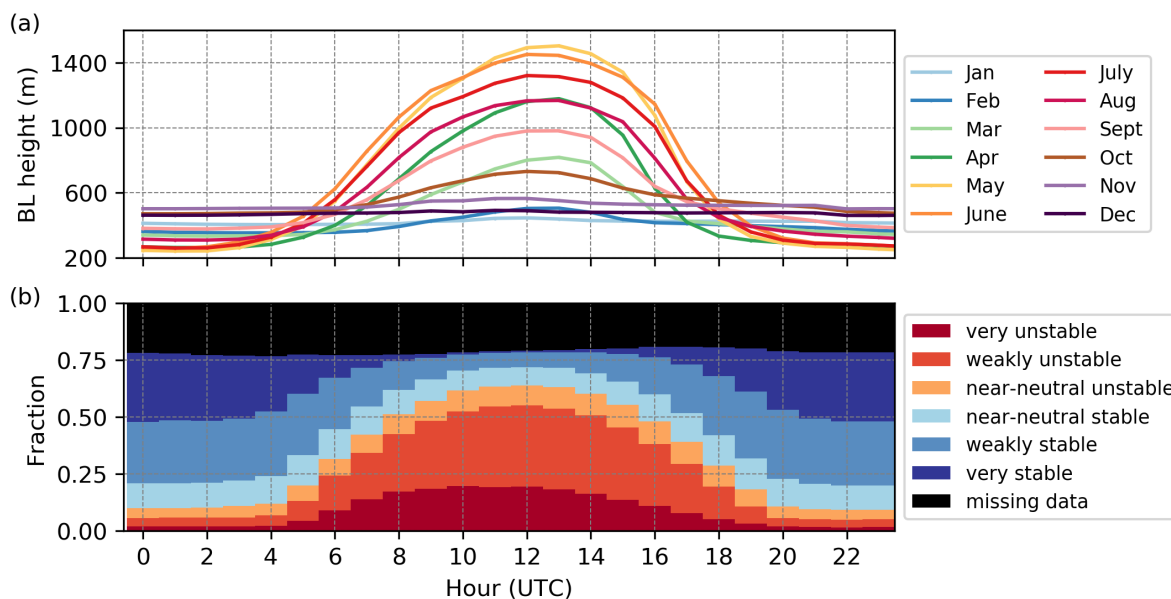
Figure 10. (a) Mean diurnal cycle of the BL height from ERA5 for each month averaged over 1979 - 2019. Data from every hour of the day is included. (b) Mean diurnal cycle of stability class for all months averaged over 1997 - 2019. The fraction of all possible observations times that each stability class is observed is shown.

Table 3. Statistical measures between BL height estimates from radiosonde measurements and ERA5. r is the Pearson's correlation coefficient and NMSE is the Normalised Mean Square Error. Values were calculated for different stability classes. Radiosonde data is taken as the observations in Equation 6. Correlation coefficients which are statistically non-significant at the 95% level are marked in italics. For each measure, the best value between the four different datasets is marked in bold.

stability	Heffter vs ERA5			Liu Liang vs ERA5			$R_{0.25}$ vs ERA5			$R_{0.50}$ vs ERA5		
	r	FB	NRMSE	r	FB	NRMSE	r	FB	NRMSE	r	FB	NRMSE
$\zeta < -0.1$	0.553	0.644	0.680	0.650	0.214	0.215	0.749	-0.328	0.332	**0.799**	**-0.0983**	**0.0985**
$-0.1 \leq \zeta < -0.01$	0.312	0.766	0.829	0.644	0.269	0.272	0.814	**0.0415**	**0.0415**	**0.820**	0.246	0.248
$-0.01 \leq \zeta < 0$	*0.0882*	0.690	0.736	0.247	0.158	0.159	0.630	**0.104**	**0.104**	**0.671**	0.323	0.327
$0 \leq \zeta < 0.01$	0.567	0.589	0.617	*0.170*	-0.329	0.333	**0.793**	0.0306	0.0306	0.768	0.291	0.294
$0.01 \leq \zeta < 0.1$	0.243	0.589	0.617	*0.118*	-0.301	0.304	0.746	**-0.103**	**0.104**	**0.754**	0.257	0.259
$0.1 \leq \zeta$	0.301	0.201	0.202	*0.103*	**-0.190**	**0.191**	0.593	-0.971	1.11	**0.604**	-0.457	-0.469
All	0.502	0.636	0.671	0.646	**0.0773**	**0.0773**	0.789	-0.112	0.113	**0.803**	0.151	0.151
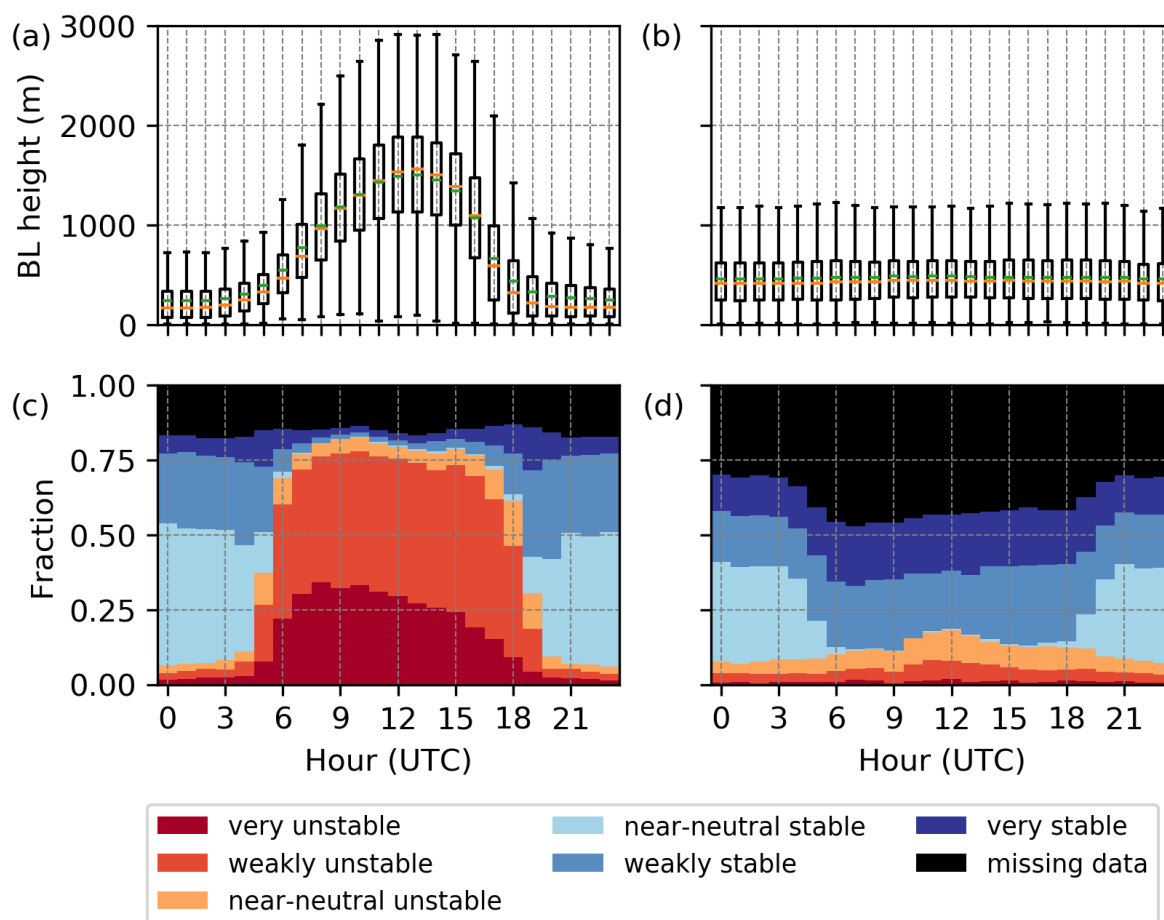
Figure 11. Variation in the (a,b) BL height and occurrence of different stability classes (c,d) as a function of time of day in (a,c) May and (b,d) December. In (a) and (b) boxes represent the 25th to 75th percentile (interquartile range, IQR) and whiskers extend to the 25th percentile minus 1.5 IQR and the 75th percentile plus 1.5 IQR. Outliers are not plotted.
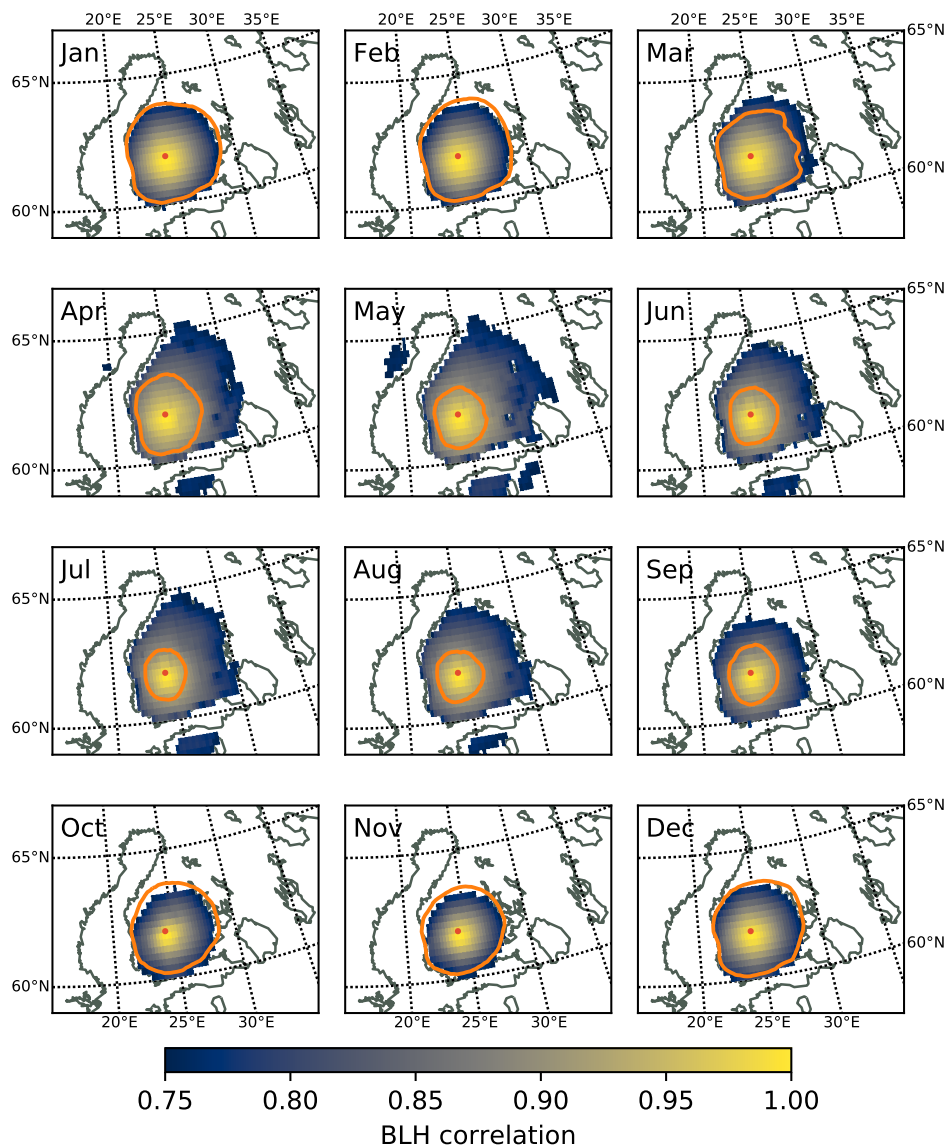
Figure 12. Monthly correlations between ERA5 BL height series from the grid cell closest to the location of the Hyytiälä station and the neighbouring ERA5 grid cells calculated over the time period from January 1979 to December 2019. Values of the correlation coefficient less than 0.75 are masked out. Position of the Hyytiälä station is marked with a red dot. Also on the panels, areas of high correlation ($r > 0.75$) computed for 12UTC time is outlined.

Table 4. Statistical measures between BL height estimates from microwave radiometer and ERA5. r is the Pearson's correlation coefficient and nRMSE is the normalised root mean square error. Values were calculated for different stability classes using both all MWR BLH values, and only MWR BLH values determined with the algorithm for unstable BL. Correlation coefficients which are statistically non-significant at the 95% level are marked in italics.

Stability	All MWR BLH values			Only unstable MWR BLH values		
	r	FB	nRMSE	r	FB	nRMSE
$\zeta < -0.1$	0.376	0.049	0.049	0.540	−0.019	0.019
$-0.1 \leq \zeta < -0.01$	0.650	−0.016	0.016	0.690	−0.016	0.016
$-0.01 \leq \zeta < 0$	0.624	−0.234	0.236	0.745	−0.230	0.231
$0 \leq \zeta < 0.01$	0.182	−0.366	0.373	0.562	−0.373	0.379
$0.01 \leq \zeta < 0.1$	−0.075	0.051	0.051	0.391	−0.377	0.384
$0.1 \leq \zeta$	−0.286	1.152	1.409	*0.152*	−0.219	0.220
All	0.259	0.107	0.107	0.681	−0.141	0.142