The new \textit{pbl\_met}: an open-source library for building meteorological processors and advanced data processing tools

Patrizia Favaron

Received: 29 April 2021 / Accepted: 16 September 2021 / Published online: 12 January 2022
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2022

Abstract
This paper presents the \textit{pbl\_met}, an open-source library for processing meteorological data collected in the planetary boundary layer (PBL) using conventional and modern sensors (examples of the latter include three-axial ultrasonic anemometers, SODAR/RASS, and wind-sensing LIDARs). The routines present in the \textit{pbl\_met} encompass the estimation of meteorological quantities (including global and net radiation and turbulence), the processing of ultrasonic anemometer raw data streams, the approximation of mixing height, and various facets of meteorological data management, summarization, and cleaning. The existence in \textit{pbl\_met} of estimation routines of diverse types allows to construct estimation paths of different depths, permitting the crafting of meteorological processors tailored to the data available. The overall quality of results obtained using estimation paths of differing depths has been investigated in a test case. Important differences have been observed, especially in mixing height, illustrating in the same time the wide range of processing options permitted by the \textit{pbl\_met} and the importance to select the best ones according to the data actually available; as some of the options allow the direct use of data from modern sensors, the \textit{pbl\_met} can be used as a platform for building next-generation meteorological processors for dispersion modeling applications.

Keywords Software library · Meteorological processor building · Modern meteorological processors · Planetary boundary layer

1 Introduction

All local-scale atmospheric pollutant dispersion models in current use require as input a description of the state of the planetary boundary layer (PBL). The actual data composing this description vary from model to model but always include two broad categories: wind (possibly three-dimensional) and turbulence information.
The nature and design of each dispersion model dictates the details of the meteorological data to be used as input. All these model-bound data sets differ in form and contents from the information commonly released by meteorological stations: hence the need of meteorological processors estimating the met model input given commonly available data. Most existing meteorological processors, e.g., AERMET (for use with AERMOD), CALMET (for CALPUFF and CALGRID), or MPP-FMI (for CAR-FMI and UDM-FMI), perform their estimations by the scheme described in van Ulden and Holtslag (1985) or similar ones. This choice reflects the fact that until recently, turbulence was not routinely measured at meteorological stations and had to be consequently derived from commonly available measurements.

In addition to the classic model-bound meteorological processors mentioned above, specific codes have been developed to address special problems. One example is SIMETPRO (Abiye et al., 2019), whose development stem from the necessity of taking into account environmental conditions very different from the ones typically found in the temperate Northern Hemisphere climates more characteristic of the places where famous meteorological processors have been developed.

The need of developing local processors based on van Ulden-Holtslag and similar schemes is not new and is at the origin of the first PBL_MET, a FORTRAN 77 library collecting a sample of relationships found in literature (Fraternali and Sozzi 1993). The main purpose of the PBL_MET was to help modeling practitioners to develop their own special-purpose meteorological processors using commonly available electromechanical sensors.

Meantime the quick evolution of technology fostered the construction of meteorological networks of a new type, able to get direct information on PBL state, using, e.g., ultrasonic anemometers (Schotland 1955). Examples of these networks, currently in operational use, are the SHAKEUP net and the Lazio micro-meteorological network (owned and run by the environmental protection agencies of Lombardy and Lazio, respectively). Other sensors like SODARs, SODAR/RASS, and LIDAR wind profiler have also gained popularity, and some installations are now releasing data continuously.

The availability of all these new types of measurement has created the need of a type of meteorological processor, not yet available, able to use the new data types whenever available, and to revert to conventional data would the new ones be accidentally missing.

The focus of the present work is to present the new pbl_met, an evolution of the old PBL_MET designed to help building meteorological processors of this new type; this approach is considered more useful and productive than describing one “modern” meteorological processor, as it enables in principle anyone to construct their own, tailored to their specific needs.

In order to fulfill its assigned mission, the new pbl_met must provide services which did not exist in the old PBL_MET. In particular, two entire sections have been added for processing raw ultrasonic anemometer data using a simplified form of eddy covariance (Aubinet et al. 2012; McMillen 1988) and for vertical wind and temperature profile management. Statistical and data management modules have been also added, as the library focus shifted from estimating turbulence to supporting its direct measurement. And last, significant parts of the old PBL_MET have been updated, in order to make the library easier to use and to improve the accuracy of some selected estimation functions.

Additions and rewrites have been made trying at best to maintain the user programming experience as cohesive as possible: this would have not been as easy to accomplish, would the same functionalities be spliced among different specialized libraries.

Overall, the changes made to the old PBL_MET justify a name change to pbl_met, instead of just increasing the version number.
To date various specialized libraries and tools are available, especially for eddy covariance. Examples are EddyPro and TK3 (Papale et al. 2006; Fratini and Mauder 2014). Libraries like the computational kernel of EddyPro and TK3 are however strongly focused towards flux measurement, and their use in the context of dispersion modeling could be in many cases cumbersome, compared to the dedicated pbl_met. Besides, their use would leave open the problem of estimating the mixing height $z_i$, a very important parameter in dispersion modeling.

It is worth mentioning the pbl_met is a work in progress and will continue to be in the foreseeable future as new measurement techniques emerge, old ones refine, and new needs arise.

2 The pbl_met library conceptual structure

2.1 Logical anatomy of conventional and modern meteorological processors: requirements for pbl_met

As pbl_met purpose is to help practitioners to develop meteorological processors for use in dispersion modeling, its general contents should match the anatomy of existing and future met processors.

The scheme still most commonly used for conventional meteorological processors is given in van Ulden and Holtslag (1985). It starts from the acknowledgement that the dispersion of pollutants in the atmosphere is directly affected by the vertical profiles of wind (causing distance transport), turbulence (affecting diffusion and dilution), and temperature (determining the rate of rise of a buoyant plume); these vertical profiles should be determined wherever dispersion occurs, that is, at least until the PBL thickness $z_i$. As these data are not normally measured, they must be estimated from existing information, namely wind, temperature, and radiation data available from common surface stations.

According to van Ulden and Holtslag (1985), the state of the PBL can be described using few scaling parameters, namely the PBL thickness $z_i$, the Obukhov length $L$, the aerodynamic roughness length $z_0$, the friction velocity $u_*$, the Deardoff velocity $w_*$, and the temperature scale $\theta_*$. These parameters constitute in particular the backbone of the Monin-Obukhov “surface layer” similarity theory, which connects the PBL parameters with wind and temperature statistics taken in the lower part of the PBL (Stull 1988). These parameters also connect with other important quantities, the sensible heat flux $H_0$ among them, which contribute to the operational description of the PBL and, together with some of the other parameters, are input to dispersion models. In the same paper (van Ulden and Holtslag 1985), specific estimation paths are indicated for turbulent parameters using either by the vertical profiles of wind speed and virtual temperature or by the surface energy budget method. As most readily available meteorological measurements occur at surface stations, the latter approach will be outlined.

Input data to the surface energy budget method are values of net radiation $R_n$ and soil heat flux $G$. The balance equation $R_n - G = H_0 + H_e$, where $H_0$ and $H_e$, the sensible and latent surface heat fluxes (Oke 1987), can be used to express other turbulent quantities. As $R_n$ and $G$ are seldom measured in conventional stations, provision must be taken to estimate them using more primitive meteorological data.
The most common kind of surface meteorological stations is composed by a horizontal anemometer, a thermo-hygrometer, a rain gauge, and in some cases a global radiometer. As a consequence, $R_n$ and $G$ must be themselves estimated starting from extant parameters.

In addition to turbulence and wind parameters, dispersion models need knowledge of the mixing height $\approx z_i$, that is, the PBL thickness. In van Ulden and Holtslag (1985), some indications are given for estimating it under stable, neutral, and convective conditions.

The estimation of $R_n$, $G$, $H_0$, $H_e$, $u_*$, $\theta_*$, and $z_i$ starting from existing measurement is then the task of any “conventional” meteorological processor, with important details which differ from site to site and with the specific data needs by the dispersion model addressed.

By the way, the scheme outlined in (van Ulden and Holtslag 1985) is not the only existing for conventional meteorological processors. Another, more recent, example is given in Abiye et al. (2019).

All conventional meteorological processors are based, however, on a common requirement: using “readily available meteorological data,” which meant those still produces using common electromechanical sensors like cup-vane anemometers, resistive thermometers, capacitive hygrometers, pyranometers, and the like.

In the last years, a new generation of meteorological sensors became available and easily deployable, for example, ultrasonic anemometers and wind and temperature profilers. To date, various specialized networks came to existence, like SHAKELUP net by ARPA Lombardia and the micro-meteorological net by ARPA Lazio, both composed by surface stations integrating conventional sensors with an ultrasonic anemometer. SODAR, SODAR/RASS, and LIDAR are also diffusing.

As a consequence, some of the data required by dispersion models (like, e.g., $u_*$, $H_0$, and wind/temperature vertical profiles) are available by direct measurement and no longer need to be estimated. Use of directly measured data instead of estimates promises to get more realistic results, which is desirable in view of both regulatory and research modeling applications.

Here lies a problem: to date, the meteorological processors in common use by practitioners (like AERMET and CALMET) ignore the newly available “advanced” meteorological data and insist in replacing them with estimates. This, in turn, damps modeling results quality while making the real usefulness of advanced meteorological networks questionable.

By the way, knowledge and use of directly measured turbulence may help addressing today’s problems, like understanding odor dispersion and air-bound spread of particulate-like matter like SARS-CoV-2. These delicate, fine-scale modeling tasks can only be accomplished by dispersion models from a detailed micro-meteorological understanding of both urban and rural environments. Focus in 1985, when the original PBL_MET was conceived, was very different, the problem then being quantifying the overall effects of massive pollutant releases protracted over long times.

It is the author’s strong feeling that for this to be possible, a new, “modern” generation of meteorological processors is due. These processors will inevitably need to be flexible, that is, able to provide estimates whenever direct data are not available because of failures or lack of sensors, but also to yield high quality meteorological fields when advanced data are available.

One possible approach to solve this problem could have been to write a “new” meteorological processor and let it compete with all others existing. A different, and in the author’s mind more productive, approach was followed: to collect a set of common processing blocks allowing anyone to construct their own modern processors. An example of this collection is the pbl_met.
Time has passed since the seminal paper (van Ulden and Holtslag 1985), and this gave an occasion of both modernizing the “classical” estimation paths and allowing use of direct turbulence and profile measurements. The latter point, that is direct turbulence and profile data, meant enlarging the scope from collecting as much estimation routines as possible, as it was in the old PBL_MET, to include wider data processing tools.

2.2 Classical estimation of PBL data using pbl_met

The classical side of pbl_met roughly follows the energy budget path in van Ulden and Holtslag (1985), under the conservative assumption that the surface stations in use are “minimal,” that is, equipped with temperature, relative humidity, wind speed, and wind direction sensors. Would any other sensor be present, its measurements may be used directly, so that the corresponding pbl_met estimation routine can be ignored. Notice no measurement of solar radiation or ground heat flux is required: as these quantities are key to the surface energy balance, routines exist allowing to estimate them.

Estimation of solar global radiation, \( R_g \), is supported in pbl_met by various methods having different complexity. The recommended \( R_g \) estimations in pbl_met (routines “Clear-SkyRg_Accurate” and “GlobalRadiation”) follows the path presented in Duffie and Beckman (2020), which expresses clear-sky radiation by the formula

\[
R_{cs} = \left( K_B + K_D \right) R_a, \]

where

\[
K_B = 0.98 \exp \left[ -\frac{0.00146 P_a}{K_t \sin \beta} - 0.075 \left( \frac{W}{\sin \beta} \right)^{0.4} \right]
\]

is the clearness index for direct radiation (dimensionless) (Allen 1996), \( K_t \) is the turbidity coefficient (dimensionless, equals 1 for perfectly clean air, and 0.5 or less for extremely dusty or polluted area), \( P_a \) is the atmospheric pressure (kPa), \( \beta \) is the angle of Sun above the horizon, \( W \) is the precipitable water (mm), \( K_B = \begin{cases} 0.35 - 0.36 K_B & \text{if } K_B \geq 0.15 \\ 0.18 + 0.82 K_B & \text{if } K_B < 0.15 \end{cases} \) (Allen 1996) is the transmissivity index for diffuse radiation, and \( R_a \) the solar radiation above the atmosphere (in W/m\(^2\)). \( R_g \) is in turn a function of time and position defined using astronomical relations (Duffie and Beckman 2020). Finally, global solar radiation is expressed as \( \hat{R}_g = R_{cs} \cdot \left( 1 + b_1 N^{b_2} \right) \), where \( N \) is the cloud cover (0 to 1), \( b_1 = -0.75 \), and \( b_2 = 3.4 \) (Kasten and Czeplak 1980; De Bruin and Holtslag 1982; van Ulden and Holtslag 1985). Estimating the global radiation then requires knowledge of the cloud cover, a value not typically sampled by automatic weather stations: if not measured, the value of \( N \) may be gathered at WMO or airport nodes nearby or assumed constant (e.g., 0.2 in case of no precipitation and 1.0 with precipitation. Alternatively, if a measurement of \( R_g \) is available, then a value of \( N \) can be found by solving the equation \( \hat{R}_g(N) = R_g \) (an approach used in routine “Cloud_Cover”).

Net radiation is also seldom available at weather stations, and an estimate is then necessary. The recommended estimation path in pbl_met (routine “NetRadiation”) proceeds through the separate estimation of shortwave and longwave components, through the partitioning \( R_n = R_{ns} - R_{nl} \) (Brutsaert 1982), and given \( R_{ns} = R_g \cdot (1 - \alpha) \), where \( \alpha \) is the albedo coefficient, and \( R_{nl} = \sigma \cdot f_{cd} \cdot \left( 0.34 - 0.14 \sqrt{e_a} \right) \cdot T_a^4 \).

Here \( \sigma = 5.6722 \times 10^{-8} \) Wm\(^{-2}\)K\(^{-4}\) indicates the Stefan-Boltzmann constant, \( f_{cd} = 1.35 \frac{\rho_v}{R_g} - 0.35 \) is the dimensionless cloudiness factor, \( e_a \) is the water vapor pressure (kPa), and \( T_a \) the air temperature (K) (Brunt 1932).

The ground heat flux \( G \), another of the parameters allowing estimation of PBL state according to van Ulden and Holtslag (1985), is estimated by the method shown in Choudhury et al (1987) using the formula \( G = K_G \cdot \exp(-0.5 \cdot LAI) \cdot R_n \), where \( K_G \) is a dimensionless...
constant equal to 0.4 whenever \( R_n > 0 \) and to 2 otherwise, and \( LAI \) is the leaf area index (unitless). The recommended estimation routine in \( pbl\_met \) is “GroundHeatFlux.”

Overall, the recommended estimation path from position to net radiation and ground heat flux allows to specify site- and situation-dependent parameters, namely the turbidity coefficient \( K_r \), the albedo \( \alpha \), and the leaf area index \( LAI \). This in turn permits meteorological processor builders to site-customize their products, with no special a priori assumption about stations location and change with time.

The net radiation, along with temperature, wind speed, and cloud cover, is key to the estimation of PBL parameters \( u_s, \theta_s, H_0 \), and \( z_0/L \). The approach used (routine “PBL_Parameters”) follows closely (van Ulden and Holtslag 1985) and is not detailed here.

### 2.3 Direct calculation of PBL data using eddy covariance

Fast wind and temperature sensors like the ultrasonic anemometer (Schotland 1955) can yield “sonic quadruples” (the three wind vector components and sonic temperature) at a high sampling rate, in the order of 10 samples per second or more, and at a resolution of about 1 cm/s for speed, 1 hundredth of °C for temperature.

The resulting data stream contains basically the same information estimated in van Ulden and Holtslag (1985), like, for example, the friction velocity \( u_s \), turbulent sensible heat flux \( H_0 \), and Obukhov length \( L \). This information needs being extracted rather than estimated. The extraction may be done by various procedures, among which the combination of eddy covariance method (Aubinet et al. 2012) with Monin-Obukhov surface layer similarity theory (Monin and Obukhov 1954; Foken 2006).

It is worth noting the development of eddy covariance method mostly occurred within the context of “scalar” fluxes, that is, the vertical turbulent transport of CO\(_2\), H\(_2\)O, and other gases in crops and forest ecosystems (Valentini 2003; Moene and van Dam 2014). The quantities of most interest to air-quality applications, and then to the \( pbl\_met \), are a relatively small subset of all this huge work, limited to wind components, sonic temperature, and, if present in the data set, water and carbon dioxide.

In any implementation of eddy covariance, one of the first decisions to take is which averaging scheme to use between a first-order autoregressive filter (McMillen 1988) and block averaging (Kaimal and Finnigan 1994). In \( pbl\_met \) the block averaging scheme is used on data, after optional steps of trend and spike removal (Aubinet et al. 2012).

In the simplified form of eddy covariance used in \( pbl\_met \), and limiting attention to the case most interesting to dispersion modeling (i.e., regardless of the presence of H\(_2\)O and CO\(_2\) data in ultrasonic anemometer data stream), Reynolds decomposition is applied to wind components \( u, v, w \) and temperature \( T \). Means \( \overline{u}, \overline{v}, \overline{w}, \overline{T} \); fluctuations \( u', v', w', T' \); and (co)variances \( u'u', v'v', w'w', T'T', u'v', u'w', v'w', u'T', v'T', w'T' \) are then computed and transformed from the acquisition to the streamline reference by applying the rotations described in Aubinet et al. (2012). The resulting rotated means, fluctuations, and (co)variance are then translated into the \( \overline{u} \) (van Ulden and Holtslag 1985) PBL status quantities. In particular, \( u_s = \left( \overline{uw}^2 + \overline{vw}^2 \right)^{1/4} \), \( H_0 = \rho C_p \overline{w'T'} \), and \( L = -\frac{\overline{u'T'}}{k \overline{v'T'}} \) where \( \rho \) is air density, \( C_p \) is air constant pressure thermal capacity, \( \kappa \approx 0.4 \) is von Karman constant, and \( g = 9.807 \text{m/s}^2 \) is the gravity constant.

In addition to estimating PBL parameters from conventional meteorological data and calculating them from raw ultrasonic anemometer readings, the \( pbl\_met \) allows their import from externally prepared data sets, as produced using existing eddy covariance systems found in existing micro-meteorological stations.
In this case, the processing is minimal, mostly reducing to importing externally computed turbulence parameters and validating/gap-filling them.

To this purpose, \textit{pbl\_met} includes routines for exploring statistically, summarizing and analyzing data sets and time series. The calculations made are gap-aware versions of the basic statistics described, for example, in (Bulmer 1967) and (Chatfield 2004).

### 2.4 Estimation of mixing height and profiles

The mixing height \(z_i\) is a very important parameter in dispersion modeling (DeVisscher 2014), and most models in widest use (e.g. CALPUFF and AERMOD) demand its value.

To date, \(z_i\) can be measured by various direct methods, like, for example, LIDARs and modified ceilometers, and indirectly as well by Radon concentration in constant emission areas (Vecchi et al. 2019). None of these methods is still widely adopted, and the easiest way to get it is by estimation.

In \textit{pbl\_met}, this task is accomplished (routine “EstimateZi”) from temperature \(T\), friction velocity \(u_\ast\), turbulent sensible heat flux \(H_0\), and (optionally) the Brunt-Väisälä frequency \(N\). The algorithm proceeds on daily basis, operating differently over night and day, used respectively as proxies of stable and possibly convective situations.

On night-time the non-advective form of Zilitinkevich estimate is used (Zilitinkevich and Baklanov 2002), \(z_i = \sqrt{\frac{\alpha_0+b^2-b}{2a}}\), with \(a = \left(\frac{\gamma L}{u_\ast}\right)^2\), and \(b = \frac{1}{L} + \frac{N}{26 u_\ast} + \sqrt{\frac{1.7}{1.4}}\frac{T_a}{1.4u_\ast}\), \(f\) the Coriolis parameter, and \(T_a = T + 273.15\) is air temperature (\(K\)).

The non-advective Zilitinkevich estimate is complemented with the mechanical estimate \(z_i \approx 1330\ u_\ast\), and the larger of the two is retained as the true \(z_i\).

On daytime, the \textit{Gryning-Batchvarova} model

\[
\left(\frac{dz}{dt} - w_h\right) \cdot \left\{\frac{z_i}{(1+2\Lam\_z-2b)L} + \frac{c_nT}{\gamma (1+\Lam\_z-2b)L}\right\} = \frac{w^{\ast T}/\gamma}{\mu}
\]

is used (Batchvarova and Gryning 1991; Batchvarova and Gryning 1994) assuming the temperature lapse rate above the PBL is \(\gamma = \frac{a}{z_{1+1}} - b + c \cdot z\) where \(a = 3\), \(b = 1.98 \cdot 10^{-3}\), and \(c = 2.27 \cdot 10^{-6}\) are site-dependent parameters, and \(A = 0.2\), \(B = 2.5\), and \(C = 8.0\) are empirically derived “universal” parameters. \(w_\mu\), the subsidence velocity at PBL top, is assumed to be 0.

For the Gryning-Batchvarova model to be entirely specified, an initial condition must be selected. In \textit{pbl\_met} this condition is defined as \(z_i(t_0) = 0\), where \(t_0\) is beginning of daytime.

The non-linear differential equation of Gryning-Batchvarova model is non-stiff, and in \textit{pbl\_met}, it is solved numerically using the fourth-order explicit Runge-Kutta method (Stoer and Bulirsch 1993).

It is worth noting the Gryning-Batchvarova is still mathematically valid also in case \(L > 0\) (stable conditions). In \textit{pbl\_met} this behavior is employed on daytime when some of the data indicate stability, which could well happen. In these stable times, the estimate of \(z_i\) decreases, instead of increasing when \(L < 0\).

The new \textit{pbl\_met} contains various functions for estimating vertical profiles, wind velocity among them.

The wind profile estimate made available by \textit{pbl\_met} (Sozzi et al. 2002) is based on Rossby similarity theory, and its form is \(u(z) = \frac{u_\ast}{\gamma} \log \left(\frac{z}{z_0}\right) - A_\mu(z)\), \(v(z) = -\frac{u_\ast}{\gamma} B_\mu(z)\) with the similarity functions \(A_\mu(z)\) and \(B_\mu(z)\) defined by \(A_\mu(z) = a_\mu \cdot \frac{z_0 - z}{z_i} + a_\mu \cdot \left(\frac{z_0}{z_i}\right)^2\) and \(B_\mu(z) = b_\mu \cdot \frac{z_0 - z}{z_i} + b_\mu \cdot \left(\frac{z_0}{z_i}\right)^2\).
The formulae used yield a smooth profile, which in many cases is just an approximation of reality as observed using SODARs or other remote sensing equipment. In case these measured profiles are available, the formulae given above may be used for gap filling or extrapolation of measured profiles above the maximum measurement height.

2.5 Other quantities and procedures

The routines presented in Sections 2.1 to 2.4 are the most significant in the context of meteorological data processing for dispersion modeling.

The pbl_met library contains many more procedures in addition to the ones mentioned in Sections 2.1 to 2.4. A full description is beyond the scope of this work. Their general scope may however be gained from the technical Appendix, where a short description of the library modules is given.

A complete list of pbl_met routines and data types, excessively long to be included in this work, can be found in the Supplementary Information (file “pbl_met_routines.txt”).

Some of the procedures not covered in this work may be used to build alternative estimation paths to the ones described here for global and net radiation. In most cases, these alternative routines are inherited from the past and have been retained for compatibility.

Other routines not described here may be useful in fields other than the processing of meteorological data for dispersion modeling. Provisions for evapotranspiration are, for example, given. Data clipping, validation, and gap filling are also provided, with support for univariate and multivariate time series.

2.6 The pbl_met test procedure

In order to test the pbl_met routines, it is necessary to assemble them into a procedure which reads available conventional meteorological and PBL data and process them allowing appropriate comparisons. In a sense, this procedure is similar to a meteorological processor, with additional provisions for saving intermediate results.

As the collected data are affected by sensor and data processing accuracy and resolution, which can affect the results of comparisons. To date, a “higher order” station which can be used for the systematic validation of pbl_met routines is not yet available (although a specific validation campaign can be designed). It is however still possible to evaluate the effect of choosing different estimation ways and to compare estimates with measurements whenever possible.

The procedure adopted operates by first collecting the available data and then processing them using various methods. The collected data are assumed to include wind speed $U$, conventional temperature $T$, relative humidity $R_h$, global and net radiation $R_g$ and $R_n$, friction velocity $u_*$, sensible heat flux $H_0$, and stability parameter $z/L$. We may assume $U$, $T$, $R_h$, $R_g$, and $R_n$ are collected using conventional sensors of the type found in commonly available meteorological stations; $u_*$, $H_0$, and $z/L$ are collected by an eddy covariance system fed by an ultrasonic anemometer. It is important that all these data are collected at a single site and in time sync.

In the description of the test procedure, the convention is adopted to denote with $\hat{x}$ the estimate of a measured quantity $x$.

The test procedure contains three estimation pathways. The first is a “deep estimate” $\left(\hat{R}_g^-;\hat{R}_g^+\right) \rightarrow \left(\hat{R}_n^-;\hat{R}_n^+\right) \rightarrow \left(\hat{u}_*^-;\hat{H}_0^-;\hat{u}_*^+;\hat{H}_0^+\right) \rightarrow \left(\zeta_i^-;\zeta_i^+\right)$, where apices – and + designate,
respectively, values obtained assuming a high or low atmospheric turbidity ($K_t = 1$ and 0.5, respectively) for global radiation and then carrying on the computations step by step “preserving apex”, that is, for example, $\hat{R}_g^{-} \rightarrow \hat{R}_g^{1}$. The second estimation path starts with measured global radiation, $R_g \rightarrow \hat{R}_g^{[2]} \rightarrow \hat{R}_g^{[3]}$, and the third with measured net radiation, $R_n \rightarrow \hat{R}_n^{[3]} \rightarrow H_0^{[3]} \rightarrow \hat{z}_i^{[3]}$. The fourth, and final, path is a “shallow” estimate starting from measured PBL parameters, $\hat{u}_s, H_0 \rightarrow \hat{z}_i^{[4]}$. Moving from the deep to the shallow estimate corresponds to changing from the classical to the most aggressive modern meteorological processor, passing through intermediate cases. Decreasing the estimation depth also reduces progressively the number of configuration options available: this may diminish the freedom of modelers on one side but also decrease opportunities for processor users’ error.

In the deep path, global radiation is estimated as clear sky radiation given the station position and time and measured values of temperature and relative humidity and pressure (routine “ClearSkyRg_Accurate”). The estimate is repeated assuming $K_t=0.5$ for $\hat{R}_g^{-}$ and $K_t=1$ for $\hat{R}_g^{+}$. The two estimates $\hat{R}_g^{-}$ and $\hat{R}_g^{+}$ are then compared with the measured $R_g$. The comparison is made both visually, using a comparison plot (a scatter plot of data in which the axes have identical size, and with the bisector of first and third quadrant, corresponding to a perfect match), and statistically. In this latter case, the normalized absolute difference

$$NAD = \frac{|x_i - y_i|}{|x_i + y_i|}$$

is employed as a relative indicator of dissimilarity (Hanna and Chang 2012).

A regression line is also calculated, and its data are presented in tabular form. Global radiation estimates $\hat{R}_g^{-}$ and $\hat{R}_g^{+}$ are then used as input for estimating the respective net radiations $\hat{R}_n^{-}$ and $\hat{R}_n^{+}$, given the water vapor pressure $e_a$, absolute temperature $T_a$, aerodynamic surface roughness $z_0$, and cloudiness factor $C$ (routine “NetRadiation”). The absolute temperature $T_0$ is obtained from measured temperature $T$ using the simple conversion $T_0 = T + 273.15$. The water vapor pressure $e_a$ is estimated from $T_a$, atmospheric pressure $P_a$, and the wet-bulb temperature $T_w$, itself estimated in terms of relative humidity $R_h$, pressure $P_a$, and temperature $T_a$ by the approximation described in Stull (2011). The cloudiness factor $C$ is assumed to be 1 for estimates $\hat{R}_n^{-}$ and $\hat{R}_n^{+}$, for which no global radiation is available and calculated as the daily average of $1.35 \cdot \left( R_g / \hat{R}_g^{-} \right) - 0.35$ when $R_g > 0$ for $\hat{R}_n^{[2]}$.

Once the net radiation has been estimated, it is used along the land use code $l$, aerodynamic roughness length $z_0$, temperature $T$, wind speed $U$, and cloud cover $N$; they all are used to estimate the PBL parameters. In the test procedure, $l=4$ (scattered urban), $z_0=0.023$m, and $N=0$. All other values are measured.

Finally, the mixing height $z_i$ is estimated from PBL parameters, temperature, and position.

The same calculations are made using the same assumptions in all other estimation chains, except the chains’ first step which always correspond to measured data.

### 2.7 The SHAKEUP test data set

Testing of the whole `pbl_met` is under way and will continue in future as new routines are added or changed.

Here test results of the part of `pbl_met` described in this work are presented, using data from the SHAKEUP station of Cinisello Balsamo Parco Nord (9.206016 E, 45.542443 N), from 1 March 2021 to 30 April 2021 (courtesy ARPA Lombardia). The station location can be considered of rural type, on grassland with about 6m trees nearby.

The SHAKEUP micro-meteorological network consists of 6 surface measurement nodes equipped with a three-axial USA-1 ultrasonic anemometer (METEK GmbH), a MeteoFlux...
Core V2 eddy-covariance system (Servizi Territorio srl), a set of conventional sensors, and a data-logger (CAE SpA).

The conventional sensors include a temperature and relative humidity probe, a rain gauge, a pyranometer, and a net radiometer. The ultrasonic anemometer and conventional sensors allow to get all the data required by the \textit{pbl\_met} test procedure described in the previous section.

3 Results and discussion

3.1 Global (clear-sky) solar radiation

The scatter plots in Figure 1 compare the estimates $\hat{R}_g^-$ and $\hat{R}_g^+$ with the measured global radiation $R_g$.

Figure 2 shows the time evolution of measured and estimated data on a specific day.

Figure 1 indicates a quite large proportion of global radiation estimates stay close to the first and third quadrant bisector. This is confirmed by the relatively low value of the NAD indicator, which is 0.1493 for high turbidity and 0.1434 for low turbidity data. The adjusted $r^2$ is 0.8628 for $\hat{R}_g^-$ and 0.8649 for $\hat{R}_g^+$.

The presence of outliers in the upper-left side of the diagram, that is points with high $\hat{R}_g^-$ or $\hat{R}_g^+$ and significantly low $R_g$, depend on high cloud cover situations: estimates $\hat{R}_g^-$ and $\hat{R}_g^+$ are in fact obtained in clear sky. Figures 2 and 3 show estimate and measure over a “good” and a “cloudy” day, respectively.

The spread of global radiation data points is caused by the presence of atmospheric phenomena which may affect the ground energy balance or by calibration issues able to change pyranometer readings. Reflective obstructions and refractive effects may in fact decrease or increase the measured data, without affecting estimates. The possible presence and position of clouds in the sky may also have an important effect: if no cloud is visible (ideal case), the pyranometer collects only direct and diffuse radiation. If clouds are present and screening the Sun radiation is attenuated and the pyranometer reading is lower than the cloudless

Fig. 1 Comparison plots for global radiation (dashed line, 1st and 3rd quadrant bisector; solid line, through-zero regression line)
condition. When clouds are present but not in the direct path between the Sun and the pyranometer in addition to direct and diffuse radiation, reflected radiation is also present, and the pyranometer reading may be higher if no clouds were present. It is worth noting that the attenuation and reflection may occur under the same cloud cover value, contributing to make $R_g$ estimation a difficult task. The deep estimate of PBL parameters and $z_i$, which depend on the $R_g$ estimate, is also uncertain.

### 3.2 Net radiation

The difficulties in estimating $R_g$ outlined in Section 3.1 may be overcome by measuring $R_g$ directly. In the typical estimation path used by the test procedure, knowledge of $R_g$ precedes immediately the estimation of the net radiation $R_n$, so that the degree of uncertainty
affecting the first directly acts on the other. Figures 4 and 5 present the comparison of the various net radiation estimates with measured values.

The outliers in Figure 4 have the same explain as in Figure 1: net radiation values, estimated from estimated global radiation, still carry the lack of information about cloud cover. This scattering is no longer visible in Figure 5, in which the estimate has been performed using the cloud cover aware measurement of $R_g$. The better overall agreement of $R_n^{(2)}$ with $R_n$ is also confirmed by the NAD indicator, as $NAD(R_n^-, R_n) = 0.3096$, $NAD(R_n^+, R_n) = 0.2542$ and $NAD(R_n^{(2)}, R_n) = 0.2037$. Similarly, $r^2(R_n^-, R_n) = 0.8408$, $r^2(R_n^+, R_n) = 0.8422$, and $r^2(R_n^{(2)}, R_n) = 0.9831$. 

---

**Fig. 4** Comparison plots for net radiations $\hat{R}_n^-$ and $\hat{R}_n^+$ estimated from $\hat{R}_g^-$ and $\hat{R}_g^+$ (dashed line, 1st and 3rd quadrant bisector; solid line, through-zero regression line)

**Fig. 5** Comparison plot for net radiation $R_n^{(2)}$ estimated from measured net radiation $R_n$ (dashed line, 1st and 3rd quadrant bisector; solid line, through-zero regression line)
3.3 PBL parameters

Temperature, wind speed, and net radiation (along with the land use code I, aerodynamic roughness length $z_0$, and anemometer height $z_r$) are used for estimating the PBL parameters $u_*$ and $H_0$. Three different paths are used in the test procedure: from $\tilde{R}_n^-$ to $\tilde{R}_n^+$ (deep estimation path), from $R_n^{(2)}$, and from the measured $R_n$, obtaining, respectively, $\tilde{u}_*-\tilde{u}_*^{(2)}, u_*^{(3)}$, and $u_*$ and $\tilde{H}_0^-, \tilde{H}_0^+, H_0^{(2)},$ and $H_0^{(3)}$, which may be compared to their measured counterparts $u_*$ and $H_0$. In order to reduce clutter, plots are shown for sensible heat flux $H_0$ and its estimates only: friction velocity is only described numerically. It can also be noted that $u_*$ estimates do not depend on $R_n$, so $\tilde{u}_*-\tilde{u}_*^{(2)}$. The same can be said for all other estimates, whose values depend only on wind speed $U$ and temperature $T$, which are always measured. Attention may then be limited to one estimate, $\tilde{u}_*^{(3)}$, assumed as representative of all others by identity. It is also worth noting the measured $u_*$ is only available with one decimal digit in the original ARPA database, which may limit the usability of comparisons.

Figures 6, 7, and 8 show a pattern like the one observed for net radiation estimates: use of the cloud cover blind deep estimates $\tilde{R}_n^-$ and $\tilde{R}_n^+$ induces a large scattering in data points, which is somewhat reduced if cloud cover is taken into account as in $R_n^{(2)}$ and $R_n$. Additionally, Figure 7 shows that use of estimated $R_n^{(2)}$ tends to induce a systematic overestimation which is somewhat mitigated if $H_0$ estimation is made from measured $R_n$, as seen in Figure 8.

Indeed, $\text{NAD}(\tilde{H}_0^-, H_0) = 0.5237$, $\text{NAD}(\tilde{H}_0^+, H_0) = 0.4679$, $\text{NAD}(\tilde{H}_0^{(2)}, H_0) = 0.4162$, and $\text{NAD}(\tilde{H}_0^{(3)}, H_0) = 0.2653$, confirming the estimates get “less worser” as the estimation path gets shallower: the less estimation steps, the better. It is also $r^2(\tilde{H}_0^-, H_0) = 0.6710$, $r^2(\tilde{H}_0^+, H_0) = 0.6732$, $r^2(\tilde{H}_0^{(2)}, H_0) = 0.8526$, and $r^2(\tilde{H}_0^{(3)}, H_0) = 0.8545$, not contradicting the claim than the shallowest estimate $\tilde{H}_0^{(3)}$ is “better.”

Finally, $\text{NAD}(u_*^{(3)}, u_*) = 0.3907$ and $r^2(u_*^{(3)}, u_*) = 0.9205$, indicating a decent linear correlation between estimate and measurement but quite a poor NAD score (almost twice the value observed for $H_0^{(3)}$).

![Fig. 6](image-url) Comparison plots for sensible heat flux $H_0^-$ and $H_0^+$ estimated from $\tilde{R}_n^-$ and $\tilde{R}_n^+$ (dashed line, 1st and 3rd quadrant bisector; solid line, through-zero regression line)
No $z_i$ measurement is available at Cinisello Parco Nord station, and then direct comparisons between estimate(s) and measurement cannot be performed.

It is nevertheless possible to see in a single graph the consequence of the choice of the estimation path.

In Figure 9 it can be seen the two shallower estimates $\hat{z}_i^{(3)}$ and $\hat{z}_i^{(4)}$ yield visibly smaller values than $\hat{z}_i^{-}$ and $\hat{z}_i^{+}$. The overestimate is quite mild under good insolation days, as happened on 8 March 2021 but is close to one order of magnitude on the cloudy 29 April
The “intermediate” estimate $z_i^{(2)}$ compares well to $z_i^-$ on day 8 March 2021 and to $z_i^{(3)}$, $z_i^{(4)}$ on 29 April 2021.

## 4 Conclusions and future work

As it was seen in the result section, the routines contained in the `pbl_met` library are various enough to allow their users to construct various kinds of meteorological processors and can be used to build different estimation paths for quantities relevant to dispersion modeling.

This is visible in the case study presented in the results section, where different estimation paths of different depth for turbulence parameters and the mixing height have been compared. The results obtained by the various estimation paths change significantly with the estimation depth. Results from the deeper path, in which turbulence indicators are reconstructed only from commonly available meteorological data (wind, temperature, and relative humidity), show a large scattering respect to PBL parameters and mixing height obtained from the shallowest path, where eddy covariance has been used to measure the PBL parameters; in the specific case of the mixing height, the deeper estimation paths tend to overestimate the value obtained in the shallowest estimation path. This effect is not unexpected, as the deeper estimation paths are blind to important phenomena which are difficult to guess, like, for example, the cloud cover.

The important point here is that the `pbl_met` diverse set of functions allows to tailor the estimation chain to the data available, including the measured PBL parameters which are not normally employed in conventional meteorological processors. That is, the `pbl_met` can be a starting point for building “modern” meteorological processors for dispersion modeling, able to take direct advantage from the rich set of turbulence data available from meteorological stations like the ultrasonic anemometer based ones in the SHAKEUP network; the presence in `pbl_met` of routines allowing to build traditional meteorological processors offers the possibility to revert to classical estimations when the measured PBL parameters are not available for any reasons (e.g., a temporary failure of the ultrasonic anemometer).

The important differences observed in mixing height between a deep “classical” and a shallow “modern” estimation paths can be reasonably expected to have important effects.
on the results of dispersion models: this has not been investigated here and may constitute the subject of a future work.

Appendix Technical details

The *pbl_met* library can be downloaded from its GitHub project repository, 
[https://github.com/serv-terr/pbl_met](https://github.com/serv-terr/pbl_met)

The repository is organized in subdirectories, among which the main source space, “/core.”

Other important directories are “/examples,” where source example code can be found, and “/doc,” for manual sources. Other directories exist as well.

The *pbl_met* test procedure mentioned in this paper can be found in the “/test/physical_tests/” and is composed by the main program “zi.f90,” the data access module “shakeup.f90,” an example of the station list file “shakeup_stations.csv,” and the data used in paper, in directory “/test/physical_tests/data/.”

The *pbl_met* is written in Fortran 2003 with minor use of Fortran 2008 extensions and is designed to be included in user projects as text source. Users may well package the library as a static or dynamic object library.

To reference the *pbl_met* from user code, it is sufficient to add the USE *pbl_met* statement.

In principle, any Fortran 2008 compiler may be used to compile the *pbl_met* sources, with possibly minor modifications due mainly to the use of “quiet NaN” (IEEE 1985) to encode invalid and missing values. The *pbl_met* sources compile without modification, with compile options as in the *makefiles* mentioned, using the open-source gfortran compiler (current version).

The *pbl_met* is composed by a main module named *pbl_met*, which includes all operational modules. Modules existing to date are listed in Table 1. The list of functions, data types, and data structures is extensive.

| Module       | Contents                                                                 |
|--------------|--------------------------------------------------------------------------|
| pbl_base.f90 | Repository for useful constants, valid/invalid entries coding, and utility procedures; also contains data types for simple INI file parsing, and spline functions |
| pbl_depth.f90| Estimation of $z_i$ by Gryning-Batchvarova model (day-time) and non-advective Zilitink-evich method (night-time) |
| pbl_evtrn.f90| Estimation of evapotranspiration by ASCE method, based on Penman-Monteith equation (Monteith 1981; Penman 1948; Penman 1956) |
| pbl_simil.f90| Surface layer similarity functions, both point-wise and profile-wise |
| pbl_stat.f90 | Functions supporting exploratory statistics of meteorological data, including auto- and cross-covariances and correlations, and a univariate time series data type |
| pbl_thermo.f90| Functions pertaining thermodynamics, psychrometry, radiometry |
| pbl_time.f90 | Support for time stamps and their management/conversion. Two forms are specified, a low-resolution one fitting into a 32 bit integer, and a high-resolution encompassing 64 bits (“double precision floating point”) |
| pbl_wind.f90 | Functions and data types supporting anemology, both general and ultrasonic anemometer specific, including eddy covariance |
| pbl_met.f90  | Module collecting all of the above, for convenience |
types and data type member procedures can be found in file “pbl_met_routines.txt” in the
Supplementary Information.

For historical reference, and to benefit users of FORTRAN 77, the old PBL_MET
(Sozzi and Fraternali 1994) is still distributed as part of the new, in PBL_MET_old directory
within the main pbl_met repository. As mentioned previously, the scope of old PBL_MET
is limited to classical PBL parameter estimation from conventional meteorological
stations.

It can be used at the user’s risk.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s42865-021-00042-9.

Acknowledgements I sincerely thank the meteorological team of ARPA Lombardia, who provided me the
testing data, my former professor Roberto Sozzi, who introduced me to the fascinating subject of boundary
layer meteorology, and my recently passed away colleague and friend Daniele Fraternali, who together with
Roberto Sozzi developed the former, and still useful, PBL_MET.

Declarations

Conflict of interest The author declares no competing interests.

References

Allen RG (1996) Assessing integrity of weather data for use in reference evaporation equation. J Irrig Drain
Eng 122(2):97–106
Aubinet M, Vesala T, Papale D (eds) (2012), Eddy Covariance – A Practical Guide to Measurement and
Data Analysis, Springer
Batchvarova E, Gryning SE (1991) Applied model for the growth of the daytime mixed layer. Bound-Layer
Meteorol 56:261–274
Batchvarova E, Gryning SE (1994) An applied model for the height of the daytime mixed layer and the
entainment zone. Bound-Layer Meteorol 71:311–323
Brunt D (1932) Notes on radiation in the atmosphere: I. Q J R Meteorol Soc 58:389–420
Bulmer MG (1967) Principles of Statistics. Oliver and Boyd, Second edition
Brutsaert W (1982) Evaporation into the Atmosphere: Theory, history and applications. Kluwer Academic
Publishers
Chatfield C (2004) The Analysis of Time Series: an Introduction. Chapman & Hall, Sixth edition
De Bruin HAR, Holtslag AAM (1982) A simple parameterization of the surface fluxes of sensible and latent
heat flux during daytime compared with the Penman-Monteith concept. J Appl Meteorol 21:1610–1621
DeVisscher A (2014) Air Dispersion Modeling: Foundations and Applications. Wiley
Duffie JA, Beckman WA (2020) Solar Engineering of Thermal Processes, Photovoltaics and Wind. Wiley,
Fifth edition
Falge E, Baldocchi D, Olson R (2001) Gap filling strategies for defensible annual sums of net ecosystem change. Agricolt Forest Meteorol 107(1):43–69
Foken T (2006) 50 Years of the Monin-Obukhov Similarity Theory. Bound-Layer Meteorol 119:431–447
Fraternali D, Sozzi R (1993) PBL_MET: a software library for advanced meteorological and air quality process ing, Air pollution 93 Monterey (Mexico) 16-18
Fratini G, Mauder M (2014) Towards a consistent eddy-covariance processing: an intercomparison of Eddy-
Pro and TK3. Atmos Meas Tech 7:2273–2281
Hanna S, Chang J (2012) Acceptance criteria for urban dispersion model evaluation. Meteorog Atmos Phys
116:133–146
Holtslag AAM, van Ulden AP (1983) A simple scheme for daytime estimates of the surface fluxes from
routine weather data. J Clim Appl Meteorol 22:517–529
IEEE (1985) IEEE Standard for Binary Floating-Point Arithmetic. ANSI/IEEE Std:754–1985
Kaimal JC, Finnigan JJ (1994) Atmospheric Boundary Layer Flows: Their Structure and Measurement. Oxford University Press
Kasten F, Czeplak G (1980) Solar and terrestrial radiation dependent on the amount of and type of clouds. Sol Energy 24:177–189
McMillen RT (1988) An eddy correlation technique with extended applicability to non-simple terrain. Bound-Layer Meteorol 43:231–245
Moene AF, van Dam JC (2014) Transport in the Atmosphere- Vegetation- Soil Continuum. Cambridge University Press
Monin AS, Obukhov AM (1954) Osnovnye zakonomernosti turbulentnogo peremeshivanija v prizemnom sloe atmosfery (Basic laws of turbulent mixing in the atmosphere near the ground). Trudy geofiz Inst AN SSSR 24(151):163–187
Monteith JL (1981) Evaporation and surface temperature. Q J R Meteorol Soc 107:1–27
Oke TR (1987) Boundary Layer Climates. Routledge, Second edition
Papale D, Reichstein M, Canfora E, Aubinet M, Bernhofer C, Bernard L, Kutsch WL, Rambal S, Valentini R, Vesala T, Yakir D (2006) Towards a more harmonized eddy covariance CO2 fluxes: algorithms and uncertainty estimation. Biogeosci Discuss 3(4)
Penman HL (1948) Natural evaporation from open water, bare soil and grass. Proc R Soc Lond A193:120–146
Penman HL (1956) Estimating evaporation. Trans Am Geophys Union 37:43–50
Reynolds O (1895) On the dynamical theory of incompressible viscous fluids and the determination of the criterion. Phylosophical Transact Royal Soc London 186:123–164
Schotland RM (1955) The measurement of wind velocity by sonic waves. J Meteorol 12:386–390
Sozzi R, Fraternali D (1994) PBL_MET library for advanced meteorological and air quality data processing. Proceedings of the 2nd Workshop on Intercomparison of Advanced Practical Short-Range Atmospheric Dispersion Models
Sozzi R, Georgiadis T, Valentini M (2002) Introduzione alla turbolenza atmosferica – Concetti, stime, misure, Pitagora
Stoer J, Bulirsch R (1993) Introduction to Numerical Analysis. Springer, Second edition
Stull RB (1988) An Introduction to Boundary Layer Meteorology. Kluwer Academic Publishers
Stull RB (2011) Wet-bulb temperature from relative humidity and air temperature. J Appl Meteorol Climatol 50:2267–2269
van Ulden AP, Holtslag AAM (1985) Estimation of atmospheric boundary layer parameters for diffusion applications. J Clim Appl Meteorol 24:1196–1207
Valentini R (ed.), (2003) Fluxes of Carbon, Water and Energy of European Forests, Springer
Vecchi R, Piziali FA, Valli G, Favaron P, Bernardoni V (2019) Radon-based estimates of equivalent mixing layer heights: a long-term assessment. Atmos Environ 197:150–158
Zilitinkevich S, Bakanov A (2002) Calculation of the height of the stable boundary layer in practical applications. Bound-Layer Meteorol 105:389–409