The MAESTRO turbulence dataset derived from the SAFIRE ATR42 aircraft

Louis Jaffeux[1], Marie Lothon[1], Fleur Couvreux[4], Dominique Bouniol[4], Grégoire Cayez[3], Lilian Joly[5], Jérémie Burgalat[5], Cyrille De Saint Leger[3], Hubert Bellec[3], Olivier Henry[3], Dyaa Chbib[3], Tetanya Jiang[3], Vergez Gilles[3], and Sandrine Bony[2]

[1]Laboratoire d'Aérologie, University of Toulouse, CNRS, UPS, Toulouse
[2] LMD/IPSL, CNRS, Sorbonne University, Paris, France
[3] SAFIRE, CNRS, Météo-France, Toulouse, France
[4] CNRM, Météo-France, Toulouse, France
[5] AEROLAB, CNRS, Reims Champagne Ardenne University, Reims, France

Correspondence: Louis Jaffeux (louis_jaffeux@gmx.fr) and Marie Lothon (marie.lothon@cnrs.fr)

Abstract. The MAESTRO airborne field campaign took place between August 10 and September 10 2024 over the North Atlantic tropical ocean near the Cabo Verde Islands. Its goal was to investigate the processes that control the mesoscale organization of clouds with a payload of probes and sensors, as well as vertically and horizontally-pointing radars and lidars. A particular attention was ~~paid~~ given to the role of coherent structures in the boundary layer and mesoscale cloud organization. This focus motivated the acquisition of high-resolution measurements of temperature and water vapor to capture turbulence dynamics in the subcloud layer. To achieve this, six hygrometers and four temperature sensors were deployed, including a new fast-rate hygrometer called FAST-WAVE. This article describes the turbulence dataset, prepared on the basis of these measurements. It consists in 25 Hz segmented time series of calibrated water vapor mixing ratio, temperature, and three-dimensional wind ~~,~~ taken with an air speed of approximately 100 m s^{-1}, relative to the aircraft, their corresponding fluctuations, as well as ~~turbulent moments,~~ statistical moments, covariance, and integral length scales. In total, 40 hours of stabilized legs data were gathered in a wide range of mesoscale contexts and local cloud conditions, with nearly 13 hours consisting of high-quality boundary-layer samples. This paper describes the methodological choices ~~made for all the computations,~~ required to obtain the dataset, including calibrations, and corrections ~~that were~~ applied to the original measurements. The collection of NetCDF files composing this dataset is publicly available on the AERIS website.

1 Introduction

The role of clouds in global climate dynamics remains associated with many mysteries and open questions. The simulation of clouds in numerical models also remains challenging, especially the representation of the diversity and behavior of low-level clouds over the oceans (Konsta et al., 2022). The tropical ocean is populated by many cloud types, ranging from small shallow cumulus to deep convective clouds. Low-level clouds are ubiquitous in this environment and their albedo makes them key players in the energy budget of the Earth (Bony and Dufresne, 2005). These clouds interact with their environment ~~by exchanging latent heat through evaporation and condensation~~ through condensation and evaporation, and by ~~absorbing solar~~

interacting with solar and terrestrial radiation. Using observations, several configurations have been described in the boundary layer with respect to convection, turbulence, and cloudiness. For example, shallow convection has been linked to organized structures in the subcloud layer, such as roll vortices creating cloud streets (LeMone and Pennell (1976)). In another case, the turbulence in the subcloud layer was found to be locally suppressed through cloud precipitation or evaporation while a cold pool formed, enhancing turbulent fluxes in the boundary layer at its front (de Szoeke et al., 2017). In another configuration the boundary layer and ~~the~~ clouds above were decoupled (Nowak et al., 2021). Understanding how such scenarios interplay, influence local circulation, contribute to broader cloud organization patterns, and, finally, affect the radiative impact of clouds, relies on the acquisition of representative observations.

~~Tropical marine clouds inherently exist at the intersection of mesoscale organized patterns and chaotic variability, where large-scale circulations and external drivers, such as the warm tropical ocean, supply the energy that generates turbulence.~~ Starting with the GATE experiment, scientists have formulated hypotheses about the organization of tropical clouds ~~for decades (Warner et al., 1979), and~~ (Warner et al., 1979). More recently systematic classifications of cloud organization using satellite images have been proposed over the tropical ocean (Stevens et al., 2020; Janssens et al., 2021). Links between the variety of cloud patterns and near-surface wind and stability of the low troposphere were pointed out (Bony et al., 2020) and confirmed by ground based instrumentation over Barbados (Schulz et al., 2021). Data from the ~~EUREC4A field campaign (Bony et al., 2017, 2022)~~ EURECA field campaign (Bony et al. (2017) and Bony et al. (2022)) were used to understand these different patterns ~~(Bony et al. in prep).~~ (Bony et al., 2025). Using LES simulation, Dauhut et al. (2023) reproduced and analyzed one case of meso-scale cloud organization observed during ~~EUREC4A~~ EURECA and highlighted the importance of cold pools and mesoscale circulation in its formation. Other LES studies used this dataset to evaluate whether current models are able to reproduce cloud organizations (Schulz and Stevens, 2023; Jansson et al., 2023). As is often the sign of success in scientific research, ~~EUREC4A~~ EURECA answers questions by raising new ones, that have already inspired the community to return to the field and collect new data.

From August to ~~October~~ September 2024, a network of international field campaigns, named ORCESTRA (Organized Convection and EarthCARE Studies over the Tropical Atlantic) (Stevens et al., 2025), took place over the tropical Atlantic to investigate the organization of convection, namely MAESTRO (Mesoscale organisation of tropical convection), PERCUSSION (Persistent EarthCare underflight studies of the ITCZ and organized convection) (Groß et al., 2026), BOW-TIE (Beobachtung von Ozean und Wolken – Das Trans ITCZ Experiment), PICCOLO (Process investigation of clouds and convective organization over the tropical ocean), CLARINET (CLoud and Aerosol Remote sensing for EarThcare) (Bell, 2024), CELLO (Cloud and EarthCARE caL/vaL Observations), SCORE (Sub-Cloud Observations of Rain Evaporation), and STRINQS (Soundings and TuRbulent eddy measurements in the ITCZ with a Network of QuadcopterS) projects. A common objective of these campaigns, is to collect relevant observations in the atmosphere in the vicinity of the Inter-Tropical Convergence Zone (ITCZ). These eight research projects therefore gathered around this focal point under the ORCESTRA umbrella to plan simultaneous field campaign deployments, first in the ~~eastern~~Eastern, then in the ~~western~~ Western North Tropical Atlantic Ocean. This collaboration fostered numerous instrumental synergies and joint missions, including coordination with spaceborne operations, as well as encouraging long-term research cooperation that will build on the data collected during the campaigns.

Within ORCESTRA, the MAESTRO field campaign used the instrumented ATR-42 research aircraft from SAFIRE to investigate the physical processes governing the mesoscale organization of shallow and deep convection. This campaign ~~follows the earlier mentioned 2020 EUREC4A field campaign and aims to pursue the~~ pursues the scientific questions that arose from ~~it~~the earlier mentioned EURECA project. The interplay between coherent structures in the sub-cloud layer and ~~the mesoscale organization of~~ clouds is one of the primary objectives of MAESTRO. A combined analysis of airborne measurements of cloud properties and spatial distributions, and in-situ ~~turbulent~~ observations of typical turbulence scales and fluxes in the ~~subcloud~~ sub-cloud layer, was envisioned to address this issue. State-of-the-art turbulence measurements are therefore a main requirement for the completion of the objectives of MAESTRO. Retrievals of proper turbulent fluctuation time series and moment computations require ~~expertise in~~ specific instrumentation, data filtering, and processing.

The goal of this article is to present the collected data, ~~to~~ describe the associated workflow ~~, and to share them with~~ that produced higher level turbulence statistics, and make them publicly available to the scientific community. This dataset includes high-frequency calibrated time series and turbulent fluctuation time series for temperature, 3D-wind, and water vapor mixing ratio, as well as corresponding ~~turbulent~~ statistical moments and associated errors. Section 2 describes the context of the campaign, the instrumentation used for in-situ turbulence measurements, and the sampling strategies. Section 3 presents the data processing, including improvements made to the original high-rate data, in particular for the wind corrections and the calibrations of high-rate temperature and humidity sensors, as well as the computation methodology for fluctuations, ~~turbulent~~ statistical moments, and integral length scales. Section 4 presents an overview of the final products and discusses the strengths and limitations of the dataset, as well as possible future improvements. Finally, section 5 concludes the article with recommendations for the users of the dataset and its future within the MAESTRO project.

2 The MAESTRO field campaign~~: acquisition strategy~~

This section provides a brief overview of the MAESTRO campaign, first outlining its objectives ~~,~~ and the encountered conditions, then describing the instrumentation used for turbulence measurements, and finally the adopted sampling strategies.

2.1 ~~Description of the campaign~~

~~The MAESTRO project aimed to 1) test hypotheses for mechanisms explaining the mesoscale organization of shallow and deep convection, 2) investigate whether and how this organization influences the Earth's radiation budget, 3) provide an extensive dataset to evaluate high resolution models in their prediction of mesoscale cloud-related organizations. To achieve these objectives, the field campaign targeted mesoscale conditions ranging from shallow to deep convection. Various cloud types and their organizations were systematically observed, with consistent sampling of the marine boundary layer, cloud base level, and mid-troposphere. The campaign documented several phenomena and cloud organizations ranging from cloud streets, convergence lines, cold pools, isolated short-lived convective cells up to large mesoscale convective systems (MCS). Operating from~~

2.1 Description

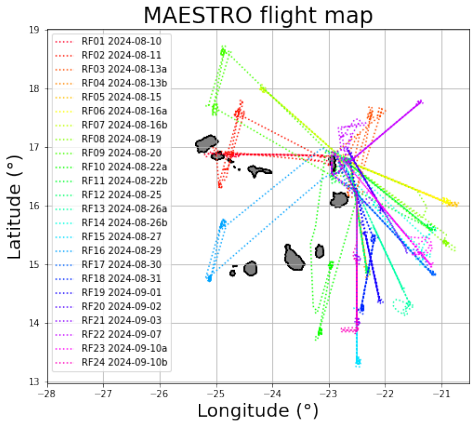


Figure 1. Trajectories of the 24 MAESTRO Research Flights (RF) of the ATR-42. Each flight is named RFXX (MAESTRO nomenclature) or 2024-MM-DDa or b (ORCESTRA nomenclature).

Operating out of Sal Island, Cabo Verde, ~~the MAESTROfield campaign consisted in 24 airborne missions performed between August 10 and September 10 2024. The ATR-42 research aircraft from SAFIRE was equipped with a payload of~~ MAESTRO's research flights leveraged the SAFIRE ATR-42 aircraft, equipped with advanced in-situ and remote sensing ~~instruments, including horizontally and vertically pointing radars and lidars. The~~ instrumentation. More than 40 hours of stabilized straight flight transects (covering roughly 15,000 km in total) provided high-rate in-situ measurements of temperature, ~~3-dimensional~~ three-dimensional wind, and water vapor~~, collected during the 24 MAESTRO research flights (later referred to as RF), amount to more than 40 flight hours of stabilized legs corresponding roughly to 15 000 km of air sampled~~. The projected flight trajectories are ~~shown~~ displayed in Figure 1. The ~~south-eastern part of the Cabo Verde area was preferably sampled as the air masses originated~~ campaign predominantly sampled the southeastern Cabo Verde region to capture air masses originating from the east, ~~thereby promoting air unaltered by land , and the southern region was closest~~ minimizing land influences, while the southern sector's proximity to the northern edge of the ITCZ ~~which is a region of key interest~~ held particular relevance for the ORCESTRA community. ~~In the vicinity of~~ Near Sal airport, ~~from~~ four to six ~~soundings were performed daily to characterize the whole depth of the troposphere~~vertical soundings were conducted daily (Winkler et al., 2026), characterizing the entire tropospheric depth and contributing essential meteorological context.

~~Trajectories of the 24 MAESTRO Research Flights (RF) of the ATR-42. Each flight is named RFXX (MAESTRO nomenclature) or 2024-MM-DD-a or -b (ORCESTRA nomenclature).~~

2.2 ~~Instrumentation for turbulence measurements~~

The ATR-42 from SAFIRE is an instrumented airplane equipped with 4 high-rate temperature probes, and 6 hygrometers. Horizontal and vertical wind components are derived from measurements taken by a five-hole radome positioned at the nose of The MAESTRO project primarily aims to test hypotheses regarding the mesoscale organization of shallow and deep convection. Central to this investigation is the role of coherent structures in the boundary layer, such as cold pools and convective plumes, which are believed to significantly influence cloud development and thus contribute to mesoscale convective organization. These elements can affect the statistical properties of the aircraft, Pitot air speed sensors, and an inertial navigation unit (INS), following the equations of Brown et al. (1983). Although some of this instrumentation acquires data with rates exceeding 100 Hz, the high-rate measurements were resampled to a common 25 Hz frequency consistent with EUREC4A (Brilouet et al., 2021), where the same aircraft and similar instrumentation were used. Given an aircraft ground speed of 100 m. s^{-1}, 25 Hz is equivalent to a 4 m spacing between each measurement. The hygrometers used included :

- WVSS-2 based on laser absorption spectroscopy with acquisition frequency of 0.5 Hz.
- the newly developed FAST-WAVE (Fast Aircraft Spectrometer for Tracking Water Vapor Emissions), based on laser absorption spectroscopy with acquisition frequency of 100 Hz.
- the Li7500 (Licor), based on near infrared absorption, with a 50 ms response time, and 50 Hz acquisition frequency.
- 1011 C, a chilled mirror measuring the dew point temperature with a response time of 1 °C/sec and acquisition frequency of 1 Hz.
- HUMAERO and UCAP, two calibrated capacitive sensors sensitive to relative humidity and associated with their own temperature sensors,with acquisition frequencies of 1 and 40 Hz,respectively.

The temperature measurements were performed using two platinum wire thermometers (deiced and non-deiced E102AL Rosemount) , two fine wire probes with wire diameters of 5 and flow (de Szoeke et al., 2017; Zilitinkevich et al., 2006), underscoring the importance of characterizing turbulence to identify their signatures. Table 1 summarizes the cloud types, mesoscale features, cloud cover, and maximum cloud top sampled during each flight. The 24 flights documented diverse phenomena: populations of shallow cumuli (RF01-07,09,10μm, measuring impact temperatures faster than ,16), convergence lines (RF13-15, 17-19), cold pools (RF04, 06), isolated convective cells (RF13,14), and, towards the end of the campaign, mesoscale convective systems (RF21,23,24). Mid-tropospheric clouds (StCu and AlCu) were consistently observed and directly sampled with the ATR in-situ instrumentation in five flights (RF5, 11, 12, 20, 22). Overall, cloud cover ranged from 28 to 100Hz (Ballantyne and Moss, 1977) . More information on the thermodynamic ATR intrumentation is available in part 3.1 of Bony et al. (2022)%, and cloud tops spanned from 1 to more than 14 km, providing the diversity needed to investigate organization processes across scales. A significant shift in synoptic and cloud conditions was documented throughout the campaign, from trade wind regimes with scattered shallow convection in August to deeper conditions in September, with the occurrence of mesoscale convective systems.

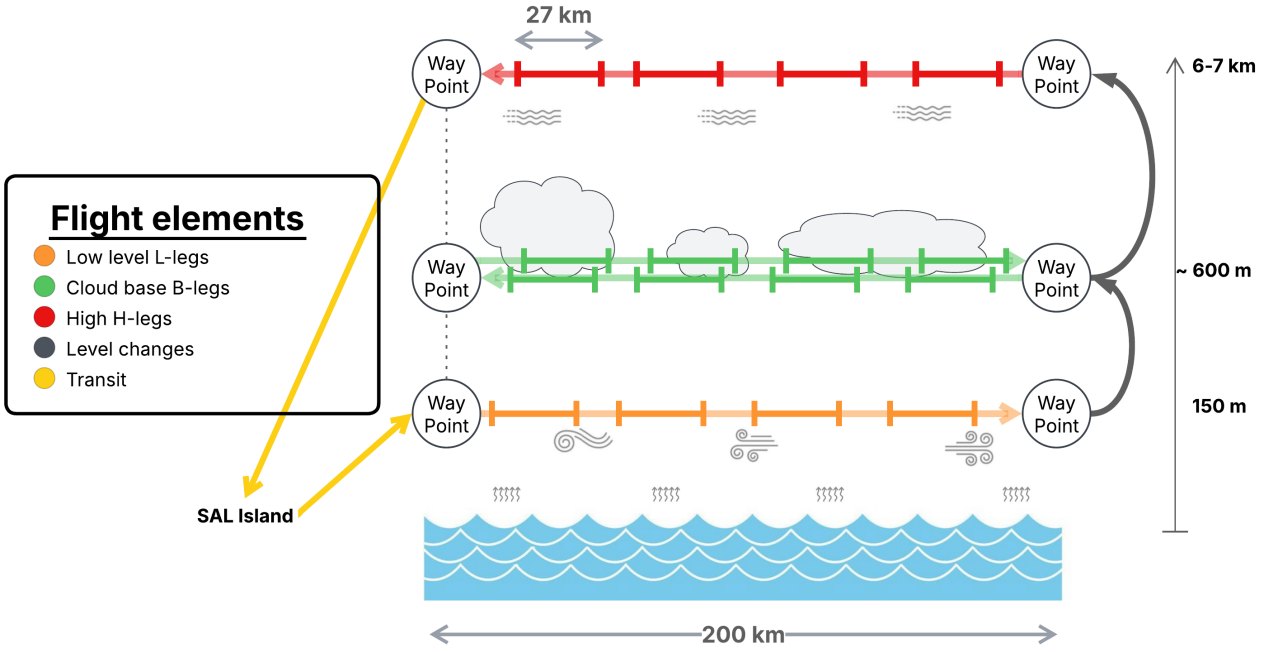


Figure 2. Schematic representation of a flight adopting the Standard strategy. Single arrows symbolize flight elements, such as stabilized legs, transits, or ascents. Line segments bounded by orthogonal end bars symbolize flight segments defined with the segmentation algorithm (see section 3.1.1).

2.2 Flight sampling ~~strategy~~strategies

The MAESTRO campaign was dedicated to the gathering of an extensive dataset, representative of various ~~convective clouds~~ cloud organizations and their surrounding environments~~ranging from shallow, low-level cumulus clouds to deeper convective systems~~, including in situ measurement of subcloud and cloud base levels, as well as cloud geometry with horizontal and vertical remote sensing. To achieve this, a main strategy, denoted in this article as "Standard", was designed as a series of stacked horizontal legs at predefined flight levels~~were planned and executed. A typical MAESTRO flight pattern consisted of a sequence of~~ . For nine of the 24 flights, the Standard strategy was implemented for systematic analysis (see Table 1). They consist of sequences of 200 km-long transects ~~, designed as follows:~~

- ~~A single H-type (High)transect at mid-tropospheric altitude (6000 meters above sea level), used for downward-looking remote sensing and in situ measurements of ice clouds and air motion.~~
- ~~Two B-type (Base)back-and-forth transects at cloud base, enabling both in situ and remote sensing observations of the cloud base region, with particular emphasis on cloud–subcloud thermal interactions.~~

– ~~One final L-type (Low) transect below the clouds dedicated to in situ measurements within the subcloud layer and upward-looking remote sensing.~~

denoted as high altitude (H-type), cloud base (B-type), and low altitude (L-type), and performed in this order. A schematic representation of a typical MAESTRO flight plan is given in Figure 2. The sequence order was sometimes modified, placing the high-altitude transect last. This change improved fuel efficiency by taking advantage of the lighter weight of the aircraft during ascent, enabling higher altitudes to be reached and additional legs to be performed within the spared flight time. ~~Finally,~~ S-type (Surface) legs ~~, were sometimes conducted depending on the conditions. They were flown exclusively during day time~~ were occasionally flown at about 60 meters above the sea surface to document turbulence ~~in the surface layer~~at the lowest altitude available to the ATR, and could only be flown during day time. However, potential alteration of other instruments due to exposure to sea salt was suspected, in particular for the horizontally pointing Lidar. For this reason, these legs were flown only on a few occasions during the campaign. ~~An example of a typical MAESTRO flight plan is given in Figure 2.~~

~~Schematic profile of a typical MAESTRO flight plan. Single arrows symbolize flight elements, such as stabilized legs, transit, or ascents. Line segments bounded by orthogonal end bars symbolize flight segments defined with the segmentation algorithm.~~ In the event of exceptional convective activity in the vicinity of Sal Island, flight plans were adapted to avoid dangerous areas and prioritize regions of highest scientific interest, this adaptive strategy is designated as "Case Study" (corresponding flights are listed in Table 1). The nomenclature used to name each transect remained similar, allowing the data from all MAESTRO flights to be merged in a single, coherent dataset. Additional segments noted as M- and T-types, were occasionally defined during Case Study flights. These legs enabled upward and downward pointing of remote sensing instrumentation to characterize cloud geometry, as well as to collect in-situ microphysical and thermodynamical measurements in specific regions. Finally, the latter portion of RF22 was dedicated to in-flight calibration maneuvers (described in more details in appendix section A).

~~Along this primary flight plan~~In addition to the Standard and Case Study strategies, the campaign also featured flights ~~coordinated~~ focused on coordinated measurements with other projects that were part of ORCESTRA. The strategy in all these flights are denoted as "Coordination" in tables 1 if the Standard was not strictly maintained. On seven flights (RF4, RF7, RF11, RF12, RF15, RF16, and RF21), the ~~german~~ German DLR HALO aircraft, operating as part of the PERCUSSION campaign, deployed drop-sondes in a circular pattern around the operational area of the ATR-42. This enabled estimates of mesoscale vertical velocity within the region of interest ~~, similarly to what was done during EUREC4A~~ (Bony and Stevens, 2019; George et al., 2021). During ~~RF18 and~~ RF16 ~~, coordination~~ and RF18, cooperation with the King Air aircraft from the CELLO project was performed with both airplanes flying in close formation, either side by side or one above the other. These coordinated operations not only completed the measurements performed on a given day, offering exceptional ~~case study~~ scientific opportunities, but also made it possible to compare simultaneous remote sensing and in-situ techniques used by the different platforms. Flights RF02, RF03, RF09, RF10, RF16 and RF18 were synchronized and aligned along the satellite track of the newly launched EarthCARE satellite (Wehr et al., 2023) for the calibration and validation of its onboard remote sensing instruments. The flights coordinated with EarthCARE varied only slightly from the Standard flight plan described in Figure 2, but kept its main elements as much as possible. In particular, to reach Earthcare overpasses, legs were often used to reach the position on time and were then not vertically stacked. During flights RF04, RF05, RF07, RF12, RF15, RF19, and RF20,

RCM and/or Sentinel1 satellites acquired Synthetic Aperture Radar ~~images on~~ (SAR) images collocated with the aircraft track. SAR is sensitive to sea surface roughness, especially centimeter-scale waves . For a complete review of the SAR capabilities see the review of Asiyabi et al. (2023). To take advantage of this, the flight plans were adjusted so ~~that low-level legs (~~S- ~~or~~ and L-type ~~)~~ legs coincided with the satellite overpasses, providing unique opportunities to investigate sea surface–atmosphere interactions. ~~Finally, in the event of exceptional convective activity in the vicinity of Sal Island, flight plans were adapted to avoid dangerous areas and prioritize regions of highest scientific interest.~~

~~The flight coordinated with EarthCARE varied slightly from the typical flight plan described in Figure 2, but kept its main elements as much as possible. A few other flights (e.g RF19 and RF20) used different flight plans adapted to specific cloud organizations. However, the nomenclature used to name each transect is similar, allowing the data from all MAESTRO flights to be merged in a single, coherent dataset. To characterize the additional leg types that were not part of the typical strategy, new categories were defined :~~ This general leg nomenclature is consistent across all MAESTRO datasets, while the specific segmentation described in Section 3.1.1 applies to the turbulence dataset presented here. All segment types are summarized as follows :

- H-type (High) transect at mid-tropospheric altitude (6000 meters above sea level), used for downward-looking remote sensing and in situ measurements of ice clouds and air motion.
- B-type (Base) back-and-forth transects at cloud base, enabling both in situ and horizontally pointing remote sensing observations of the cloud base level, with particular emphasis on cloud–subcloud interactions.
- L-type (Low) transect below the clouds dedicated to in situ measurements within the subcloud layer and upward-looking remote sensing.
- S-type (Surface) flight legs, flown at 60 meters above sea level, similar in purpose to L-type legs but at lower altitude and for comparisons with SAR radars.
- M-type (mid-level) legs, corresponding to intermediate altitudes ~~,~~between 1 and 4 km.
- T-type (top) legs, corresponding to low-cloud tops, between 1 and 1.5 km.

~~These legs enabled upward and downward pointing of remote sensing instrumentation to characterize cloud geometry, as well as the collection of in-situ microphysical and thermodynamical measurements in specific regions. This general leg nomenclature is common to all MAESTRO datasets.~~

~~For the turbulent dataset, the legs undergo an automatic segmentation algorithm, creating 4.5-minute segments ($\approx$ 27 km) by minimizing the heterogeneity of all 5 turbulent fluctuations simultaneously. This quantification allows the identification of sufficiently homogeneous segments for reliable turbulent moment estimates. To do this an heterogeneity score $\mathcal{H}$ has been defined as follows:~~

$$\mathcal{H} = \sum_{v \in [U,V,W,T,q]} \sqrt{\int_0^{\Delta T} \left(100 \cdot \frac{\int_0^t \left(F^{(v)}(x)\right)^2 dx}{\int_0^{\Delta T} \left(F^{(v)}(x)\right)^2 dx} - 100t \right)^2 dt}$$

where F^v refers to the fluctuation time series for the variable v and ΔT is the segment duration. $\mathcal{H}$ represents the average percentage deviation from an ideal uniform accumulation of variance over the time series summed over all five variables. In particular, the cumulative variance $(\int_0^t \left(F^{(v)}(x)\right)^2 dx)$ is scaled as a percentage of the total variance over the segment $(\int_0^{\Delta t} \left(F^{(v)}(x)\right)^2 dx)$ and compared to a linear rate of accumulation, calculating the root-mean square error between the two functions. Finally, the resulting averaged percentage error is summed for all five variables to obtain $\mathcal{H}$. The algorithm to automatically optimize segment placements is described in the appendix Figure **??**. This algorithm ensures that the maximum number of segments is obtained, and sequentially minimizes $\mathcal{H}$ (i.e. maximizes the homogeneity of corresponding time series) .

2.3 Instrumentation for turbulence

Initially, the segment length was defined as 5 minutes (≈ 30 km) following the recommendations of Lenschow et al. (1994) for reliable turbulent moment estimations. Since some legs were shorter than this initial threshold, this length was adjusted to maximize data usage while maintaining a consistent duration for each segment, easing their comparison. A quantitative summary of the obtained segments, categorized by flight strategies and leg types, is presented in Table **??**. **Strategy**

Figure 3 presents the histograms of $\mathcal{H}$ calculated for segments of each category throughout the entire MAESTRO campaign. As expected, low-level segments are generally more homogeneous than those collected at higher altitudes. Using an upper threshold of 30% to define homogeneity (green shaded area in Figure 3), most S-type The SAFIRE ATR-42 is a research aircraft equipped with a payload of in-situ and L-type segments qualify as homogeneous and are suitable for boundary-layer turbulent moment computation. This results in a total of 170 homogeneous boundary-layer segments (totalling all S-, L-, and B-type segments with $\mathcal{H}$<30%) included in the MAESTRO turbulence remote-sensing instruments. The wind vector is derived from a five-hole nose radome (gust probe), in combination with Pitot-static measurements and an inertial navigation system (INS), following the standard formulation for aircraft wind retrieval (Brown et al., 1983). While some instruments acquire data at rates up to 100 Hz, all variables distributed in the turbulence dataset were resampled to a common 25 Hz frequency, consistent with the EURECA processing choices (Brilouet et al., 2021). At a typical true airspeed of about 100 m s^{-1}, this corresponds to an along-track spacing of about 4 m. We stress that the stored sampling frequency in the original MAESTRO files (25 Hz) should not be confused with the intrinsic temporal response of each instrument: for some sensors (e.g., chilled-mirror hygrometers), the physical response time is much slower than 25 Hz even if the signal is interpolated at that rate. Table 2 summarizes the main characteristics of each in-situ sensor, distinguishing acquisition sampling, nominal response, and the role of each instrument in the final dataset.

Segments with $\mathcal{H}$ values between 30% and 60% (blue shaded area in Figure 3)can be considered of intermediate levels of heterogeneity. These likely contain local-scale structures, such as thermals and clouds, which explains their predominance among B-type segments. Such structures are important to characterize for the broader MAESTRO campaign. These segments may still be used for analysis, but the moment estimates are less reliableSix hygrometers were operated during the campaign to document water vapor variability across a wide range of conditions and to enable redundancy-based quality assessment. The instruments include: WVSS-2 (laser absorption spectrometer), FAST-WAVE (fast laser absorption spectrometer), an open-path

LI-7500 (LI-COR; near-infrared absorption), a 1011C chilled-mirror hygrometer measuring dew point temperature, and two capacitive relative-humidity sensors (HUMAERO and UCAP) associated with their own temperature measurements. Four temperature sensors were also operated: two Rosemount E102AL platinum resistance thermometers (deiced and non-deiced) and two fine-wire probes (5 and 10 μm) designed for fast-response temperature measurements (Ballantyne and Moss, 1977) . Additional information on the ATR-42 thermodynamic instrumentation and its general operating principles is provided in Bony et al. (2022).

MAESTRO also served as an opportunity to introduce new instrumentation, one that is particularly relevant to turbulence and cloud studies is the new fast-response hygrometer, FAST-WAVE. This instrument was developed by GSMA (Groupe de Spectrométrie Moléculaire et Atmosphérique) based on the AMULSE technology (Joly et al., 2016), using wavelength modulation spectroscopy. The instrument was housed in a certified antenna mounted on the aircraft fuselage. To validate the integration and acquisition chain, four dedicated test flights were performed in southwestern France in spring 2024. These flights were used to verify the instrument installation, acquisition protocols, and the main data treatment procedures prior to MAESTRO operations.

~~Segments with $\mathcal{H}$ values exceeding 60% (red shaded area in Figure 3) should not be used for turbulence statistics, as such calculations rely on the assumption of homogeneity. These high heterogeneity scores likely indicate the influence of more laminar flow, non turbulent, heterogeneous or intermittent turbulence large-scale dynamics, consistent with~~ During the campaign, quick-look time series and spectra were systematically examined after each flight to identify instrumental or processing issues. Three main points relevant for turbulence applications emerged :

1. A non-zero mean vertical wind bias (typically $\sim$0.6 m s^{-1} when averaged over complete stabilized legs) indicated a calibration issue in the wind retrieval; dedicated in-flight maneuvers enabled us to diagnose and correct constant biases in radome angles, and the correction procedure is described in Appendix A.

2. The LI-7500 exhibited a persistent narrowband spectral artifact peak (from 6 to 12.5 Hz) attributed to potential vibrations of the antenna, which strongly contaminates the high-frequency part of the ~~observation that they predominantly occur in H-type segments collected at high altitudes.~~

   ~~As discussed in Section **??**,~~ humidity spectrum; therefore, this instrument is not retained in the final turbulence dataset.

3. A dedicated intercomparison of the three potential reference hygrometers (1011C, WVSS-2, HUMAERO) is provided in Appendix B. While the 1011C chilled mirror provides an absolute measurement in principle, its limited thermal response hampers its ability to track airborne humidity variability in heterogeneous conditions encountered during MAESTRO. WVSS-2 showed the best overall robustness and agreement with the mirror under conditions where the mirror response is not limiting, whereas HUMAERO exhibited varying differences with respect to the 1011C and is therefore not considered as a potential reference.

4. The non-deiced Rosemount E102AL was selected as the temperature sensor for the MAESTRO turbulence dataset. Fine-wire probes, despite their faster intrinsic response time, are highly sensitive to cloud droplet impaction: the less

protective housing means that droplet collisions are more likely, rendering the measurement less reliable in cloudy conditions (e.g., Lenschow, 1986). With approximately 38% of all segments being B-type legs, where cloud droplet encounters are likely, this sensitivity represents a significant limitation for this dataset. The deiced Rosemount sensor avoids droplet impaction through resistive heating, but the heating element introduces a small systematic warm bias that degrades the accuracy of temperature fluctuation measurements in the ~~associated random errors increase significantly above the 30% heterogeneity threshold, resulting in high uncertainty in turbulence moments and integral length scales.~~ clear-air conditions that constitute the majority of MAESTRO legs. The non-deiced Rosemount combines a sufficient acquisition frequency and response time for 25 Hz turbulence measurements with robustness across the full range of MAESTRO conditions, and was therefore retained as the reference temperature sensor.

~~In the case of L-type and S-type legs, high heterogeneity scores typically result from encountering localized atmospheric events such as cold pools, rainfall, or particularly stable layers.~~

~~Histograms of $\mathcal{H}$ (Heterogeneity score) for each segment type. Shaded regions highlight qualitative thresholds: scores below 30% (green) indicate homogeneous segments, while those above 60% (red) relate to heterogeneous segments. Scores between 30% and 60% (orange) represent an intermediate heterogeneity range, where segments may be suitable for analysis but warrant cautious interpretation.~~

3 Data processing

~~This section presents the processing workflow and discusses how instrumental redundancy is leveraged to ensure the quality of the MAESTRO turbulent dataset. It describes the calibration and correction methodologies applied to the five variables used for the extraction of turbulent features, along with~~ Reliable turbulence statistics require measurements acquired under quasi-homogeneous flow conditions. The processing workflow described in the current section was designed to optimize this requirement. The wind retrievals received special attention during the MAESTRO campaign and are described in detail in Appendix A. These corrections were applied independently of the processing steps described here and the ~~rationale behind these adjustments.Finally, an evaluation based on amplitude and spectral behavior analysis is presented to determine the most suitable sensor for temperature and humidity measurements and to validate the overall data quality .~~ obtained wind serve as the starting point of the methodology. The MAESTRO turbulence dataset comprises two complementary data products: (1) 25 Hz time series of calibrated wind components, water vapor mixing ratio, and temperature, along with their derived fluctuations, provided both for individual segments and complete flight legs; and (2) segment-scale turbulence statistics including average values, second- and third-order statistical moments, covariances, integral length scales, spectral scaling exponents, and turbulent kinetic energy dissipation rate.

3.1 ~~Processing workflow~~

First, all details required to produce the fluctuation time series are described: the filtering of the original time series (Sect. 3.1), the segmentation of the data based on homogeneity of wind fluctuations (Sect. 3.1.1), and the calibration of the humidity

signal (Sect. 3.1.1). Second, the restitution of segment-scale statistical properties is described, including computational details for moments and integral length scales (Sect. 3.2), the TKE dissipation rate, and quality assessment metrics: random and systematic errors in statistical moment estimates, and identification of instrumental noise limitations affecting TKE dissipation rate estimates (Sect. 3.3).

3.0.1 ~~Computation of fluctuations~~

~~Turbulent fluctuations~~

3.1 Segmented fluctuation time series

Fluctuations $x'(t)$ are defined, for any given variable $x(t)$, as:

$$x(t) = \overline{x(t)} + x'(t) \tag{1}$$

where $\overline{x(t)}$ is the mean component of $x(t)$. There are multiple ways to ~~compute this mean component and thereby extract the corresponding~~ extract fluctuations, including removing segment average, detrending, and high-pass filtering. For the MAESTRO ~~Turbulence dataset, this is achieved by applying~~ turbulence dataset, a high-pass filter is applied with a cutoff frequency of 0.018 Hz, equivalent to filtering mesoscale features larger than approximately 5 km, following ~~step by step the methodology presented in Brilouet et al. (2021)~~the methodology of Brilouet et al. (2021). Because the use of ~~frequency cutting~~ frequency-cutting filters involves arbitrary choices, in particular the numerical method used to approximate the Fourier transform, alternative fluctuations ~~using~~ computed with a straightforward linear detrending method are also ~~computed and provided~~ used later in the dataset to evaluate the impact of the high-pass filtering on statistics.

Once segmentation and pre-processing have been performed ~~, the turbulent~~ (see Sects. 3.1.1 and 3.1.1), the fluctuations at 25 Hz are computed for each segment and for the ~~5 thermodynamical variables U_L, V_T~~five variables U_l, V_t, W, θ, and q, where ~~U_L~~ U_l is the horizontal ~~stream-wise~~ streamwise wind component, ~~V_T~~ V_t the horizontal transverse wind component ~~, computed~~ (obtained by projecting the geographical wind components onto the ~~mean wind at the segment scale~~direction of the segment-averaged wind), W is the vertical air speed, θ the potential temperature, and q the water vapor mixing ratio. ~~Figure ?? shows an example of such time series for a sub-cloud layer leg of RF06. In additionto this set of variables~~In addition, fluctuations of the horizontal wind components are ~~also~~ provided in the geographical frame of reference (U_{geo} and V_{geo}, the zonal and meridional ~~wind components~~components, respectively) and in the aircraft frame of reference ~~of the plane~~(U_x and V_y, the longitudinal and transverse components, respectively).

~~Fluctuation time series from segment L2-1 of RF06 for U_L, V_T, W, θ, and q, respectively.~~

3.1.1 ~~Computation~~ Segmentation of ~~turbulent moments and integral length scales~~the data

The ~~25 Hz fluctuations are used to compute turbulence moments. Second order moments provide information about the energy contained in the fluctuations. For their computation, the eddy correlation method is used, so that for any two variables x and y,~~

and for segments of duration ΔT :

$$\overline{x'y'} = \frac{1}{\Delta T} \int_0^{\Delta T} x'(t)y'(t)\,dt$$

aircraft legs undergo an automatic segmentation algorithm that creates 4.5-minute segments ($\approx$ 27 km) by minimizing the heterogeneity of fluctuations of the three wind components simultaneously. This quantification allows the identification of sufficiently homogeneous segments for reliable statistical moment estimates. The fluctuations of the corrected wind in geographical coordinates are used for the computation of the heterogeneity score.

Figure 3.2.2 shows a computed autocorrelation function, its integral, and the spectrum derived from a segment time series. Integral length scales are obtained as the integral of the autocorrelation function for For a segment with wind perturbations from the high-pass filtered fluctuation or cross correlation up to its first crossing with the x-axis (Lenschow et al., 1994) (Figures 3.2.2a and b). The integral length scale represents a measure of the size of the largest eddies in a turbulent flow and indicates the average length over which turbulent structures are sustained before they start to transfer energy to smaller scale eddies in the Kolmogorov energy cascade (Figure 3.2.2c).

Example of autocorrelation function, its integral, and obtained spectrum for the water vapor mixing ratio measured with FAST-WAVE for segment L1-1 of RF09.

The TKE dissipation rate ε is also computed. It measures the rate at which turbulent kinetic energy is converted into heat through viscous effects at small scales. Because this parameter governs the energy budget, it is an important result of the analysis of the turbulent flow. It can be estimated with different numerical methods. In this case, it is computed assuming Kolmogorov's scaling from power spectra such as the one shown in Figure 3.2.2c, using the formula of Lambert and Durand (1998) for the vertical wind w in the inertial subrange which is defined beyond the integral length scale and up to angular wave numbers corresponding to 25 Hz ($\approx$ 1.6 rad.m^{-1}, at 100 m.s^{-1}). Consequently, ε is defined as: three components in the geographical frame of reference $\{U_{\text{geo}}, V_{\text{geo}}, W\}$, the heterogeneity score $\mathcal{H}$ is defined as:

$$\varepsilon = \sigma_f^3 2\alpha (k_{IL}^{-2/3} - k_{max}^{-2/3})^{-3/2} \frac{1}{3 \cdot n} \sum_{p \in \{U_{\text{geo}}, V_{\text{geo}}, W\}} \sqrt{\sum_{i=1}^{n} \left(\frac{i}{n} - C_p(i) \right)^2} \quad (2)$$

$$\mathcal{H} = \frac{1}{3 \cdot n} \sum_{p \in \{U_{\text{geo}}, V_{\text{geo}}, W\}} \sqrt{\sum_{i=1}^{n} \left(\frac{i}{n} - C_p(i) \right)^2}$$

where k_{IL} and k_{max} are the wave numbers corresponding to the vertical wind integral length scale and the upper bound of $C_p(i)$ is the normalized cumulative variance at sample index i:

$$C_p(i) = \sum_{j=1}^{i} \frac{P_p(j)^2}{n\,\sigma_p^2} \quad (3)$$

with $P_p(j)$ the perturbation value at index j for component p, n the total number of data points in the segment, σ_p^2 the wavelength spectrum, respectively, σ_f is the variance of the integrated vertical wind spectrum between k_{IL} and k_{max}, and α is the Kolmogorov coefficient, assumed equal to 0.52. This formulation shows that ε is directly linked to the velocity gradients

and is higher in regions of strong velocity variations and shear. In atmospheric modeling, TKE dissipation rate determines the energy dissipation at the subgrid scale and is therefore required for closing the Navier-Stokes equation. The values obtained for the MAESTRO datasetare showed in Figure 10 and described in Section 4. p over the segment, and i/n representing an ideal uniform accumulation of variance. $\mathcal{H}$ is dimensionless and independent of segment length, so that values remain comparable across segments of different duration or resolutions. The design of this score is based on work done by Bernard-Trottolo (2001) . While the original definition only used the vertical wind component, we added both horizontal components to account for heterogeneity in these directions.

3.2 Pre-Processing

As most field campaigns, MAESTRO was an opportunity to develop and test new instrumentation, including a new high-rate hygrometer : FAST-WAVE. This instrument was developed after the Atmospheric Measurements by Ultra-Light SpEctrometer (AMULSE) technology by the GSMA (Groupe de Spectrométrie Moléculaire et Atmosphérique) (Joly et al., 2016), based on wavelength modulation spectroscopy. It was housed in a certified antenna, which was mounted on the airplane fuselage. It is the first time that such an hygrometer is used onboard a research aircraft. To validate the instrumental payload and specifically test this sensor prior to the field missions, 4 test flights were performed in the South West of France during the spring of 2024. These flights helped ensure the quality of instrument integration on the aircraft, data treatment procedures, and acquisition protocols. The subsequent analysis raised important concerns about the wind retrieval with unusual values for averaged angles of attack and sideslip, resulting in non negligible averaged vertical wind. For long enough period, the averaged vertical wind is expected to vanish and be very close to $\mathcal{H}$ quantifies how unevenly the variance of the wind fluctuations is distributed along a segment. In the rest of the article and in the dataset, $\mathcal{H}$ is expressed in %. A homogeneous segment in which variance accumulates steadily throughout will have $\mathcal{H} \approx 0$ ± 0.1 m.s^{-1}. This is usually used as a criterion in %, while a segment containing an abrupt regime change will exhibit a markedly nonlinear cumulative variance profile and therefore a large $\mathcal{H}$. A single dirac step of variance results in the calibration of airborne wind measurements (Mallaun et al., 2015; Khelif et al., 1999) . This hypothesis was generally valid during the straight legs of the MAESTRO campaign. However, significant correlation (25%) was noted outside of these legs. The current subsection details the train of thought that led to corrections of the airplane wind , and the calibration methodology for temperature and water vapor mixing ratiomaximum heterogeneity score $\mathcal{H} = 50$ %. The score is invariant to time reversal but sensitive to the shuffling of time steps, which is the desired behavior for detecting regime transitions rather than simply measuring total variance. $\mathcal{H}$ has two purposes : (i) to be used in conjunction with the segmentation algorithm to optimize the partitioning of flight legs into segments representing single turbulent regimes, (ii) as a quantitative description of the wind field homogeneity for each sample.

3.1.1 Wind corrections

The air velocity components are obtained by subtracting the ground velocity of the plane, measured by the INS, to the air velocity with respect to the aircraft, deduced from the differential pressures measured by the five-hole nose radome and a Pitot tube. Aircraft flight inevitably disturbs the air flow, and therefore all sensing probes mounted on any aircraft are not

directly measuring the meteorological undisturbed airflow. To mitigate this effect, gust probes are often mounted on the nose of Flight legs are divided into segments of 270 s (4.5 min), using an alternating bi-directional algorithm to minimize overlap and optimize coverage. For each flight leg, the aircraft and calibrated, which is the case for the 5-hole radome of the ATR-42. In addition, there are multiple potential sources of errors that include uncertainty about the positions and alignments of the INS and radome nose, and instrument performance. In the case of turbulence studies, high accuracy in these measurement is of primary importance. number of fitting segments (n_{seg}) is determined. The remaining duration R is distributed evenly across all segments to obtain an adjusted segment duration: $\Delta t = 270\,\mathrm{s} + R/n_{\mathrm{seg}}$.

Calibration maneuvers were performed during RF22, on 7th of September 2024, as described in Lenschow (1986) to confirm and resolve the apparent wind measurement issue. In particular, a reverse heading maneuver was performed by flying at constant height and true heading starting with the wind in the back, then taking a 90-degree heading change to the right, after that taking a 180-degree heading change to the right, and finally taking a 90-degree turn to the left . This maneuver was performed in 27 minutes at an altitude of 2.2 km, outside of the boundary layer. The initially computed wind during the reverse heading maneuver is shown in Figure A1 (panels a and b). The basic hypothesis behind this maneuver is that the horizontal and vertical winds should be independent of the attitude parameters of the plane. In particular with different heading mean values, during the first and last straight segments and during the second and third one, Two boundaries are initialized at the start and end points of a given leg: the left boundary $t_L = t^{\mathrm{leg}}_{\mathrm{start}}$ and the retrieved 3D-wind should be nearly identical. right boundary $t_R = t^{\mathrm{leg}}_{\mathrm{end}}$. Segments are then extracted alternately from the left and right sides of the remaining leg portion. For even iterations ($n = 0, 2, 4, \ldots$), a search window $[t_L, t_L + \Delta t]$ is defined; the starting position $t_{\min}$ that minimizes $\mathcal{H}$ over a 270-s segment is identified, the segment is set to $[t_{\min}, t_{\min} + 270\,\mathrm{s}]$, and the left boundary is updated to $t_L = t_{\min} + 270\,\mathrm{s}$. For odd iterations ($n = 1, 3, 5, \ldots$), a search window $[t_R - \Delta t, t_R]$ is defined; $t_{\min}$ is found analogously, the segment spans $[t_{\min}, t_{\min} + 270\,\mathrm{s}]$, and the right boundary is updated to $t_R = t_{\min}$. After each iteration, R and Δt are computed again. This process continues until all n_{seg} segments are extracted, working progressively from both ends toward the center. This heuristic balances positioning constraints across all segments while ensuring maximum coverage of each leg.

The initially obtained horizontal wind (FigureA1a) strongly varies in direction and strength when crossing the same air volume after a 90 Figure 3 presents the histograms of $\mathcal{H}$ obtained for each segment category throughout the entire MAESTRO campaign. As expected, low-level segments are generally more homogeneous than those collected at higher altitudes. Visual inspection of the time series guided the choice of indicative thresholds to distinguish homogeneous, quasi-homogeneous, and heterogeneous segments; these thresholds are not strict quality gates but serve as guidance for users. Using an upper threshold of 10 % to define homogeneity (green shaded area in Figure 3), 8 (89%) S-type and 180 degrees heading change. Moreover the vertical wind shows a strong positive change as the plane turns to the right, and a strong negative change as the plane turns to the left (see FigureA1b)(95%) L-type segments qualify as homogeneous and are suitable for boundary-layer statistical moment computation. In total, 355 boundary-layer segments (all S-, L-, and B-type segments with $\mathcal{H} < 10$ %) are included in the MAESTRO turbulence dataset as homogeneous. This indicates that the vertical wind is heavily correlated with the roll of the aircraft.

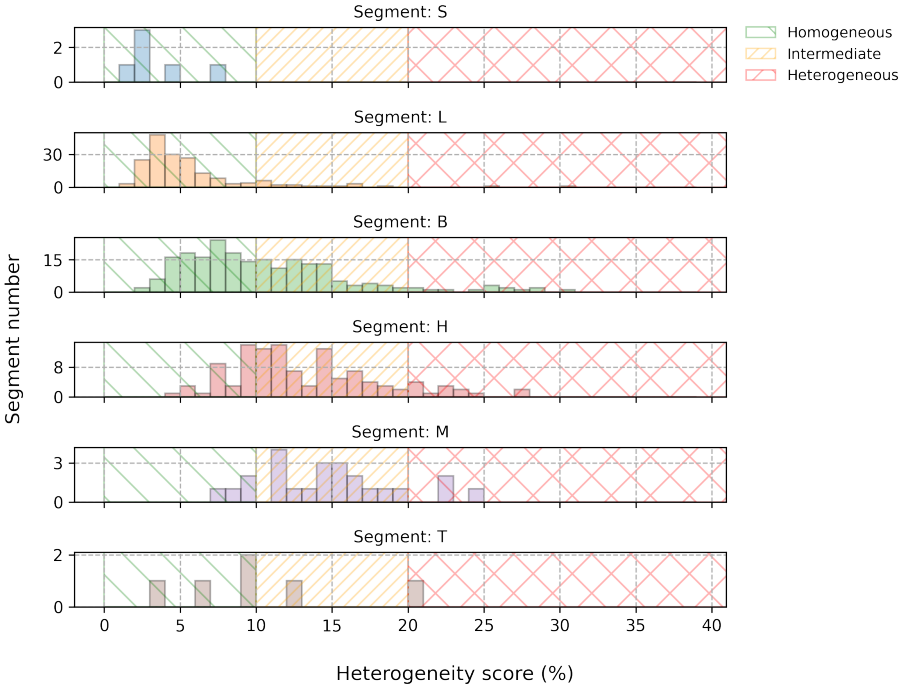


Figure 3. ~~Computed wind vectors during the reverse heading maneuver with and without wind corrections. Initially computed horizontal~~ Histograms of $\mathcal{H}$ (~~a~~heterogeneity score) ~~and vertical~~ for each segment type. Shaded regions highlight indicative thresholds: scores below 10 % (~~e~~green) ~~winds along the plane's trajectory. Corrected horizontal~~ indicate homogeneous segments; scores above 20 % (~~b~~red) ~~and vertical~~ indicate heterogeneous segments; intermediate scores (~~d~~orange) ~~winds along the plane's trajectory~~warrant cautious interpretation.~~This maneuvers took place during RF22 on September 7, 2024.~~

~~The adopted approach to compute biases in both radome angles using the data from stabilized legs consists in a parameter optimization (PO) that uses the wind computation formulas and optimizes two biases for the angles of attack and sideslip to minimize the vertical wind averaged for the whole campaign. This methodology is similar to the ones adopted by the German Aerospace Center DLR (Giez et al., 2021; Mallaun et al., 2015). However, rather than determining a set of biases for each flight, the method was used once, at the scale of the whole campaign assuming a non drifting sensor defect common to the entire dataset . The other hypothesis used in these studies, namely the minimization of the correlation between the vertical wind and the roll angle of the plane, was not required. Instead, simultaneous optimization of both angle biases reduced this correlation to negligible levels, from 25 to 1.4 %. The calculations produced biases of -0.015 and -0.545 degrees for the attack and sideslip angles, respectively.~~ Segments with $\mathcal{H}$ between 10 % and 20 % (orange) can be considered quasi-homogeneous. These likely contain intermittent cloud coverage or alternating air from the boundary layer and free troposphere (particularly for B-types). In the case of L- and S-types, quasi-homogeneous segments typically result from localized atmospheric events

such as cold pools, rainfall, or regime transitions. These segments may be used for analysis, but moment estimates are less reliable.

~~The wind has then been computed again, taking into account (subtracting) the obtained biases for both measured incidence angles. The horizontal wind corrected along the plane trajectory for the reverse heading maneuver (Figure A1c) shows a direction and strength close to identical as the airplane travels through the same air volumes. The vertical wind along the trajectory (Figure A1d) does not show any significant increase during the right turns or decrease during the left turn, demonstrating that roll and vertical wind are now uncorrelated. The main flaws that were initially identified in both the horizontal and vertical wind fields have thus been corrected by taking into account two fixed biases for the angle of attack and sideslip that were estimated using the full campaign dataset. The adopted correction will affect both the mean wind fields and the extracted turbulent moments and retrieved spectra. The latter is evaluated in the following subsection ??.~~ Segments with $\mathcal{H}$ exceeding 20 % (red) should not be used for usual turbulence statistics, as such calculations rely on the assumption of quasi-homogeneity. These high scores indicate laminar, non-turbulent, heterogeneous, or intermittent flow, and predominantly occur in high altitude segments classified as H-types. The time-series data can still be used to study gravity waves below the filtered length scales.

3.1.1 ~~Calibrations of temperature and humidity instruments~~

3.1.1 Calibration of Humidity Sensors

~~Fast humidity measurements have long been a challenge for airborne platforms, because fast sensors are usually calibrated in controlled environments at the ground and experience much harsher conditions during aircraft operations. The methodology applied here and taken from Brilouet et al. (2021), relies on the synergy between fast and slow reference sensors , allowing frequent calibrations of the fast sensor with a reference below a common low frequency, accounting for possible instrumental drifts. The calibration methodology is applied for each segment and for each fast sensor relative to each reference sensor mentioned in Section 2.3. Figure 4 illustrates this methodology with an example for~~ The accurate measurement of high-frequency humidity fluctuations from an aircraft platform is a longstanding technical challenge. Fast sensors often degrade over time when exposed to cloudy, dusty, or precipitation environments, or may be affected by noise due to imperfect mounting, while slow sensors are relatively stable but lack high-frequency fidelity.

We emphasize that this dataset is optimized for turbulence analysis. The primary objective is to provide accurate turbulence statistics. Consequently, the proper scaling of fluctuation amplitudes and minimal instrumental noise are critical, whereas absolute offsets in mean values are of secondary importance. For example, a 1 g kg^{-1} bias in mean mixing ratio does not affect the variance $\overline{q'^2}$, but a 10% error in the calibration slope increases the variance by 21%. Our approach therefore prioritizes the accurate reproduction of fluctuation amplitudes across scales relevant to turbulence (sub-kilometer to a few meters), rather than absolute mean accuracy. Users requiring accurate absolute values of water vapor mixing ratio, for instance to compute precise relative humidity values, should be aware that the mean value of each calibrated FAST-WAVE ~~during RF06. It consists in four steps: Low pass filtering of~~ signal is that of the WVSS-2 reference, whose absolute accuracy relative to the 1011C is characterized in Appendix B.

It is common practice to calibrate a fast instrument against a slower, more stable reference by maximizing the low-frequency agreement between the two signals, quantified by the coefficient of determination R^2 after low-pass filtering (Brilouet et al., 2021). This corrects the primary amplitude scaling and instrumental drift. The calibration is performed independently for each segment rather than once per flight or per campaign, for two reasons. First, fast sensors on airborne platforms can experience gradual or abrupt changes in their response characteristics during a flight, due to exposure to clouds, precipitation, salt, and wide ranges of temperature and humidity; a segment-scale calibration accounts for these slow drifts and environment-induced changes. And second, the MAESTRO flight legs sample distinct meteorological regimes at very different altitudes: temperature ranges from approximately +28 °C near the surface to −10 °C at 6000 m, and water vapor mixing ratio spans nearly an order of magnitude between the boundary layer and the mid-troposphere. Capacitive and spectroscopic sensors are known to exhibit environment-dependent sensitivities: the sensitivity of capacitive sensors to relative humidity is a nonlinear function of temperature, while laser absorption spectrometers can be affected by pressure and temperature variations. A segment-scale calibration ensures that the retrieved slope and offset are locally appropriate for the thermodynamic regime actually sampled, rather than imposing a single correction averaged over heterogeneous conditions.

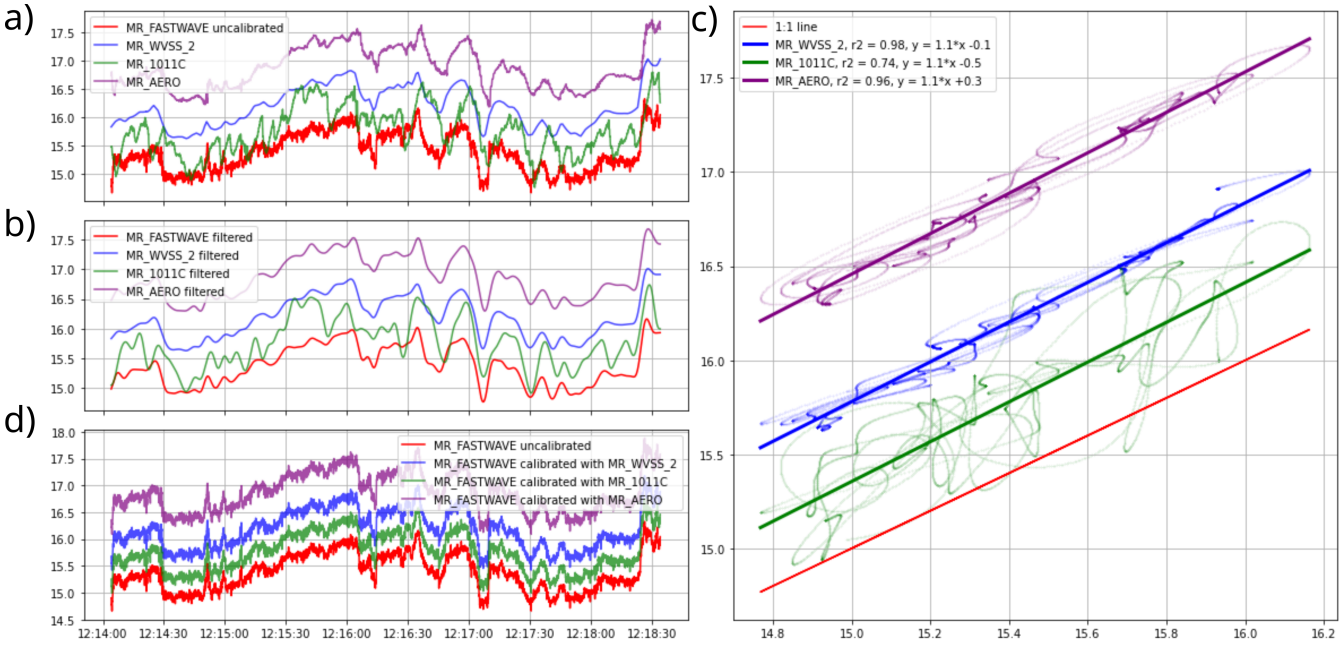


Figure 4. Example application of the calibration methodology for RF06: (a) uncalibrated time series for FAST-WAVE and the three slow sensors (g kg^{-1}); (b) corresponding low-pass filtered time series; (c) linear regression of WVSS-2 (y-axis) against FAST-WAVE; (d) calibrated FAST-WAVE time series.

Figure 4 illustrates the procedure for an example segment. For each segment and fast sensor, the fast and ~~slow sensor with a 1/6 Hz cutoff frequency (Figure~~reference signals (Figure 4a) ~~. Optimal lag search to maximize correlation between the two filtered signals (Figure~~are low-pass filtered with a cutoff at 1/6 Hz (Figure 4b). ~~Linear regression between the fast and~~

low-pass filtered signals (FigureThe optimal lag between sensors is determined by maximizing their cross-correlation, and a linear regression is fitted to align amplitude and offset (Figure 4c). Application of the retrieved slope and offset Finally, the amplitude and offsets are applied to the original fast signal (Figure sensor signal to obtain calibrated time-series (Figure 4d). Segment-sensor pairs with $R^2 > 0.9$ indicate satisfactory agreement on large-scale structure.

Example application of the calibration methodology taken from RF06 : a) uncalibrated time series for FAST-WAVE and the slow reference sensors (in g/kg) b) corresponding low-pass filtered time series, c) linear regression plots for each reference (yaxis) with respect to FAST-WAVE, d) calibrated FAST-WAVE time series for each reference.

For a given segment and fast instrument, R^2 values are compared to identify the optimal reference sensor for acquiring the highest quality signals. Following the initial methodology, correlation coefficients exceeding 0.90 indicate successful calibrations. This approach significantly helped monitoring potential instrument degradation throughout the campaign. Coupled with a spectral analysis of high-rate sensors, the best fast and reference pair of sensors are picked for high frequency temperature and humidity measurements to be used in We acknowledge that this approach has a limitation in segments where the thermodynamic field is nearly horizontally homogeneous,for instance in weakly turbulent or laminar layers. In corresponding segments, the low-frequency variability available for the regression is small, the whole dataset, based on the correlation coefficients and the characteristics of the power spectra. This analysis is presented in the following subsection **??**. fitted slope is uncertain, and the calibration is less constrained. For these cases, the R^2 criterion ($R^2 > 0.9$) indicates where the low-frequency variability is potentially insufficient to produce a reliable regression. This is consistent with the general observation that weakly turbulent H-type segments, where horizontal homogeneity is expected, also show the lowest R^2 values in the calibration results.

3.2 Post-Processing and Quality Analysis

3.1.1 Amplitude and Spectral Domain Quality Analysis : Choice of best sensors

On the one hand, the corrections of winds and the calibrations of temperature and water vapor mixing ratio modulate the amplitude of the fluctuations and thereby significantly influence the extracted turbulent moment computations. For example, a calibration slope of 1.1 will increase the variance by 21%, in comparison with a 1.0 slope . Therefore, ensuring the validity of the employed procedures is crucial. On the other hand, the spectral behavior of the parameters should be compatible with Kolmogorov's turbulence theory. The current subsection explores both aspects with respect to each variable measured during the campaign, in order to qualify the processed data : first for water vapor mixing ratio, then for temperature, and finally for the three wind components.

3.1.1 Water Vapor Mixing Ratio

Although technical limitations of the 1011C and chilled-mirror hygrometer and the HUMAERO capacitive sensor have been highlighted in Appendix section B and reported in the instrumentation section 2.3, we included them in the calibration procedure to further confirm our choice. The outcome of the rapid humidity calibrations calibrations of fast humidity sensors is presented in Figure 5, categorized by segment type and reference sensor each possible reference-fast sensor couplings.

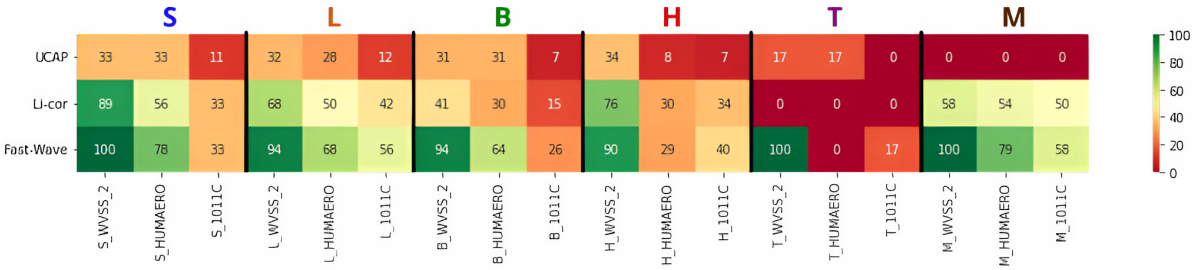


Figure 5. Results of the fast moisture sensor calibration. Percentage of segments with fast-slow sensor low frequency linear regression coefficient R^2 values above 0.9 for each segment category (heights) and each fast and reference humidity sensor.

FAST-WAVE demonstrates the best agreement across all segment types and for each reference sensors, with only ~~minor~~ rare exceptions in a few T and H segments. Among the reference sensors, ~~WVSS2~~ WVSS-2 consistently shows the best agreement across all segment types and fast sensors, except for the T-segments, where UCAP aligns better with HUMAERO than ~~WVSS2~~WVSS-2. Overall, the results demonstrate ~~that~~ poor compatibility between UCAP and any reference sensor. FAST-WAVE and ~~WVSS2~~ WVSS-2 are the most reliable combination for humidity calibration, achieving effective performance across 94% of the segments. This highlights their robust compatibility and suitability for diverse ~~segment types, with only very few exceptions. In addition to the calibrations, a quality analysis of the spectral behavior of each humidity sensor was performed on the datasetto determine which sensor best matched the expected theoretical turbulence spectrum. For each segment, the spectrum of the signal is computed using Fast Fourier Transform (FFT) . Then, the wavelengths greater than the integral length scale are resampled to be equally spaced on a logarithmic scale and~~ conditions. On the basis of these results and to provide a consistent dataset, FAST-WAVE fluctuations, calibrated against the WVSS-2 are selected as the base signal for humidity fluctuations and computations of turbulent statistics.

3.2 Statistical description of segments

In the dataset, ~~together with~~ only horizontal components in the wind-aligned frame of reference (U_l and V_t) are used for the statistical description of individual segments. However, in the files, the suffixes are not repeated to to maintain clarity. By projecting the wind onto the direction of the segment-averaged wind, the variance and covariance estimates represent fluctuations relative to the local flow, rather than with respect to fixed geographic or aircraft axes. This approach ensures that the statistics are physically interpretable and consistent across segments with differing sampling or wind directions.

3.2.1 Second- and third-order statistical moments

The 25 Hz fluctuations are used to compute segment-scale statistics. To compute these estimates, the only horizontal wind component used are those in the ~~corresponding energy values, they are used to perform a linear regression in log-log space over the such-defined inertial sub-range (as illustrated in Figure 3.2.2c)~~ wind-oriented frame of reference. Second-order moments (variance for every of the five variables: U_l, V_t, W, θ, q and covariance of each possible variable couple) provide

information about the energy contained in the fluctuations. In the moment dataset, variances are identified by the VAR_ prefix, and covariances by the COVAR_ prefix. For any two variables x and y and a segment of duration ΔT, the eddy-correlation method gives:

$$\overline{x'y'} = \frac{1}{\Delta T}\int_0^{\Delta T} x'(t)\,y'(t)\,\mathrm{d}t \tag{4}$$

where the overbar denotes a temporal average over the segment. The variance of a single variable x is the special case $\overline{x'^2}$.

~~Several important metrics can be obtained from this operation : 1) the slope which should be as close as possible to -5/3 for a turbulent sample,~~

Among covariance estimates, those of direct physical relevance include the vertical turbulent fluxes of momentum ($\overline{U'W'}$, $\overline{V'W'}$), heat ($\overline{W'\theta'}$), and moisture ($\overline{W'q'}$). The momentum fluxes quantify the vertical transport of horizontal momentum and drive the wind shear structure of the boundary layer. $\overline{W'\theta'}$ controls buoyancy-driven convection and is a key driver of boundary-layer growth. $\overline{W'q'}$ governs the supply of moisture to the cloud layer, important to boundary layer convection and associated cloud formation. $\overline{U'V'}$ quantifies horizontal shear production of TKE. And finally, the temperature-humidity covariance ($\overline{\theta'q'}$) is related to buoyancy effects in moist air and can be related to the partitioning between heat and moisture contributions to the buoyancy flux.

Turbulent kinetic energy per unit mass (TKE) is defined as:

$$TKE = \frac{1}{2}\left(\overline{U'^2} + \overline{V'^2} + \overline{W'^2}\right) \tag{5}$$

and is provided in the dataset under the variable name TKE. It gives a global measure of the intensity of turbulent motions and is a key input for turbulence parameterization schemes in numerical models.

Third-order moments and skewness are also computed for the five variables. The third-order moment $\overline{x'^3}$ and the normalized skewness $Sk(x') = \overline{x'^3}/(\overline{x'^2})^{3/2}$ characterize the asymmetry of the probability distribution of fluctuations. Non-zero skewness of vertical velocity ($Sk(W')$) can be indicative, for instance, of meteorological elements such as narrow updrafts and broad downdrafts, typical of convective boundary layers (Lenschow, 1986). In the dataset, third-order moments are identified by the M3_ prefix, and skewness by the SKEW_ prefix. All second- and ~~2)~~third-order moments, as well as skewness values, are provided both from high-pass filtered fluctuations and from detrended fluctuations (identified by the _DET suffix), to allow users to assess the sensitivity of the statistics to the choice of mean-field removal method.

3.2.2 Integral length scale and inertial-range spectral slope

The integral length scale L_x is defined as the distance over which fluctuations of a variable x remain significantly correlated. It is obtained as the integral of the autocorrelation function up to its first zero crossing (Lenschow et al., 1994) (Figures 3.2.2a,b):

$$L_x = U_a \int_0^{\tau_0} R_{xx}(\tau)\,\mathrm{d}\tau \tag{6}$$

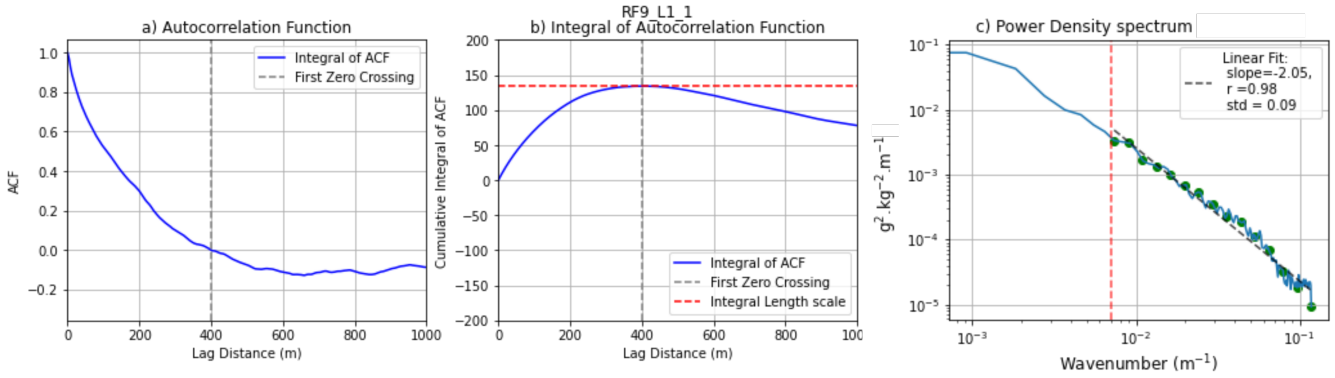
where $R_{xx}(\tau)$ is the normalized autocorrelation function, τ the time variable, and τ_0 its first zero crossing, and U_a is the true airspeed used to convert from lag time to length. The integral length scale gives a characteristic size of the ~~correlation coefficient which indicates the match with the turbulent cascade deduced from the Kolmogorov Similarity hypothesis. The inertial sub-range definition used here implies that significant noise levels would result in lower slope values and low correlation coefficient. For the three hygrometers, the two spectrum parameters are displayed in Figure ?? as~~largest energy-containing eddies and indicates the average spatial extent over which turbulent structures persist (Figure 3.2.2c). It is a key parameter for characterizing the dominant turbulent structures in the boundary layer, and for quantifying sampling adequacy through the random error estimates (Sect. 3.3.2).

~~The results of fast temperature calibrations are displayed in Figure ?? for each segment category and reference sensor, in the same way as for fast moisture sensors in Figure 5.The non deiced temperature sensor is more in line with the fast instruments for every segment type, except for the M and T categories. The two fine wires show similar performance, with an edge to the 5 μm one. Overall, the results demonstrate that 5 μm fine wire and non deiced temperature sensor are the most reliable combination for temperature calibration, achieving effective performance across 79% of the segments. This highlights their robust compatibility and suitability for diverse segment types, with only minor exceptions.~~ Integral length scales are computed for each of the five variables (L_U, L_V, L_W,L_θ, L_q) as well as for all ten cross-correlation pairs (e.g. $L_{W\theta}$,L_{Wq}, etc.), and are provided in the segment statistics files. In particular, L_W is used as input to the TKE dissipation rate computation described below. For each variable, the inertial-range spectral slope is also extracted. The power spectral density is computed using a Fast Fourier Transform on the fluctuation time series, and a linear regression in log-log space is performed over the inertial subrange, defined from the wavenumber corresponding to the integral length scale $k_{IL} = 2\pi/L_x$ up to the Nyquist wavenumber (as illustrated in panel (c) of Figure 3.2.2).

~~Similarly to what was done with the humidity sensors, a quality analysis of the spectral behavior of each temperature sensor was performed on the dataset to determine which sensor matched best the theoretical expected turbulence spectrum. For the analysis, the non deiced Rosemount was included as it is also a fast sensor, with acquisition frequency greater than the required 25 Hz. The estimated correlation coefficient and slopes are shown in Figure ??. Spectral analysis of humidity sensors : results of linear regressions by instrument and segment type. The light blue area covers correlation coefficients below 0.9, the red dotted lines indicate slope values of -2 and -4/3, the red dashed line represents the -5/3 slope expected by the theory of turbulence.~~

3.2.6 Turbulent kinetic energy dissipation rate

~~Following the two criteria, the non deiced Rosemount probe demonstrates the best behavior for any segment type despite the two fine wires being specifically designed for fast rate measurements with effectively faster response times and frequencies. This non-deiced probe also presents the advantage not to require further calibration, which adds uncertainty in the turbulent~~



3.2.3 ~~Temperature~~

~~Results~~ for segment L1-1 of ~~the temperature calibration~~RF09. ~~Percentage~~ In panel (c), 'r' is the determination coefficient, and 'std' the standard deviation of ~~segments with R^2 values above 0.9 for each category~~ equally spaced points of the spectrum and ~~each fast and reference temperature sensor~~the fitted curve in log space.

3.2.4 ~~Temperature~~

~~Results~~ for segment L1-1 of ~~the temperature calibration~~RF09. ~~Percentage~~ In panel (c), 'r' is the determination coefficient, and 'std' the standard deviation of ~~segments with R^2 values above 0.9 for each category~~ equally spaced points of the spectrum and ~~each fast and reference temperature sensor~~the fitted curve in log space.

Figure 6. Example autocorrelation function (a~~2D representation for each of the three statistically significant segment types. Comparatively,~~), its integral (b), and the ~~slopes obtained with FAST-WAVE are~~ derived power spectrum (c) for the ~~closest to the expected -5/3 value for all segment types. The correlation coefficient is also more often above the 0.9 threshold. Only for some type-L segments, UCAP provides better coefficients than~~ water vapor mixing ratio measured with FAST-WAVE ~~, demonstrating its potential as a fast-rate instrument,however with less reliability. Overall, FAST-WAVE demonstrates the best spectral behavior among all high-rate hygrometers. As a result of both the amplitude and spectral comparisons, we concluded that, for the MAESTRO dataset, FAST-WAVE, calibrated using the WVSS-2, is the best sensor choice to use for humidity measurements.Spectral analysis of humidity sensors : results of linear regressions by instrument and segment type. The light blue area covers correlation coefficients below 0.9, the red dotted lines indicate slope values of -2 and -4/3, the red dashed line represents the -5/3 slope expected from classical turbulence theory.~~

3.2.5 ~~Temperature~~

~~Results~~ for segment L1-1 of ~~the temperature calibration~~RF09. ~~Percentage~~ In panel (c), 'r' is the determination coefficient, and 'std' the standard deviation of ~~segments with R^2 values above 0.9 for each category~~ equally spaced points of the spectrum and ~~each fast and reference temperature sensor~~the fitted curve in log space.

moments estimates. We chose to consider this reference fast sensor asthe best sensor choice for temperature fluctuation measurement. The TKE dissipation rate ε measures the rate at which turbulent kinetic energy is converted into heat through viscous effects at small scales. In the case of aircraft measurements with a spatial resolution of 4 m, the dissipation range lies beyond the scope of direct observations.

3.2.7 3D Wind

Histograms of the ratio before and after correction for the variance of the three wind components and the turbulent kinetic energy for S, L, B, and H segments. To statistically assess the effect of wind corrections, Following standard practice for aircraft turbulence measurements (Lenschow, 1986), ε is computed from the vertical wind component W, which is less affected by mesoscale horizontal variability and more reliably captures isotropic fluctuations in the inertial subrange. Assuming locally homogeneous and isotropic flow (Kolmogorov, 1991), the one-dimensional power spectrum of W in the variance computed on each leg of the campaign was compared before and after correction. This comparison is summarized in the form of histograms in Figure ??. The variance of the horizontal components of the retrieved wind vector see significant change from the radome corrections. Variance variations above 20% are reached mostly in B and H segmentsinertial subrange takes the form:

$$S_W(k) = \tfrac{4}{3}\,\alpha\,\varepsilon^{2/3}\,k^{-5/3} \tag{7}$$

where $\alpha = 0.52$ is the Kolmogorov constant for the vertical-velocity spectrum, taking the values obtained by Sreenivasan (1995) and Wyngaard and Coté (1971). This expression assumes that the spectral slope in the inertial subrange is exactly $-5/3$. Using the formula of Lambert and Durand (1998), and defining the inertial subrange from the integral length scale wavenumber k_{IL} up to an upper bound $k_{\max}$ (discussed below), ε is computed as:

$$\varepsilon = \sigma_f^3 \left[2\alpha \left(k_{IL}^{-2/3} - k_{\max}^{-2/3}\right)\right]^{-3/2} \tag{8}$$

where $\sigma_f = \sqrt{\int_{k_{IL}}^{k_{\max}} S_W(k)\,\mathrm{d}k}$ is computed by integrating the W power spectral density over the inertial subrange. The distributions of the ratios of variance between streamwise and transverse components are symmetric for every segment types.The TKE and

In the absence of instrumental noise, $k_{\max}$ is set to the Nyquist wavenumber (12.5 Hz $\approx$ 0.8 rad m^{-1} at 100 m s^{-1}). However, the accuracy of ε estimates is fundamentally limited by instrumental noise at high wavenumbers, which becomes relevant in low-turbulence conditions where W fluctuation amplitudes are small. Following Muschinski et al. (2001), we assessed the spectral noise floor of the vertical wind measurements. Analysis of $S_W(k)$ across all segments revealed that instrumental noise begins to influence the spectrum when spectral density values fall below approximately 2×10^{-6} m^1 s^{-2}. Such contamination can be observed exclusively in some H-type segments at mid-tropospheric altitudes; for all boundary-layer segment types (S, L, B), no spectral contamination by instrumental noise was detected.

To prevent noise from biasing ε estimates, $k_{\max}$ is therefore defined as the minimum of the Nyquist wavenumber and the first wavenumber at which $S_W(k)$ falls below 1×10^{-5} m^1 s^{-2}, which is a conservative threshold approximately five times the

observed noise floor. In practice, this truncation affects primarily weak-variance H-type segments; boundary-layer segments retain their full inertial subrange.

Let us consider the uncertainty of ε associated with the assumption $S_W(k) \propto k^{-5/3}$ throughout the integration range in Eq. (7). When the observed spectral slope deviates from $-5/3$, the integrated vertical wind variance σ_f^2 changes, introducing a bias in the retrieved ε. To quantify this sensitivity, we numerically evaluated Eq. (8) for a spectrum of the form $S_W(k) = \frac{4}{3}\alpha\varepsilon^{2/3}k^s$, where the slope $s = -5/3 + \delta$ is varied. ε is kept constant, and the retrieval of ε uses Eq. (8), still assuming the $-5/3$ slope and meaning the error associated with a slope deviation is not only sensitive on δ, but also slightly on the integration range as σ_f is still evaluated on the s-scaled spectrum.

For parameters representative of L-type segments (integral length scale $L_W = 50$ m corresponding to $k_{IL} \approx 0.13$ rad m^{-1}, ~~however, are not affected by the radome corrections. Because TKE is half the sum of the 3 variances, it follows that the main effect of the corrections is a redistribution of the variance between the two horizontal components depending on the wind direction relative to the airplane heading.~~ $k_{\max}$ equal to the Nyquist wavenumber ≈ 0.8 rad m^{-1}), a slope deviation of $\delta = +0.1$ (i.e., an observed slope of -1.57 instead of $-5/3$) leads to an underestimation of ε by approximately -18 %, while a deviation of $\delta = -0.1$ (slope of -1.77) causes an overestimation of approximately +23 %. For parameters representative of H-type segments (integral length scale $L_W = 200$ m corresponding to $k_{IL} \approx 0.03$ rad m^{-1}, $k_{\max}$ equal to the Nyquist wavenumber ≈ 0.8 rad m^{-1}), a slope deviation of $\delta = +0.1$ (i.e., an observed slope of -1.57 instead of $-5/3$) leads to an underestimation of ε by approximately -30 %, while a deviation of $\delta = -0.1$ (slope of -1.77) causes an overestimation of approximately +43 %. Segments where the parameter corresponding to the extracted spectral slope of the vertical wind (Slope_W) deviates substantially from $-5/3$ should be treated with caution when interpreting absolute values of ε.

~~Spectral analysis of wind components by segment type. The light blue area covers correlation coefficients below 0.9, the red dotted lines indicate slope values of -2 and -4/3, the red dashed line represents the -5/3 slope expected by the theory of turbulence.~~

~~To enable a direct comparison with previously presented results for temperature and water vapor mixing ratio, the~~ Unlike fixed-scale approaches (e.g. Muschinski et al., 2001), which use a fixed wavelength range, our formulation dynamically adapts the spectral ~~characteristics of the wind field are also examined. Figure **??** presents the power spectral density (PSD) distributions for each wind component. The analysis is conducted in a manner consistent with the methodology applied to the scalar fields. Across all segments classified as surface-layer (S-type) and lower-boundary-layer (L-type) , the derived spectral slopes and corresponding correlation coefficients exhibit strong agreement with theoretical predictions, particularly with the well-established -5/3 power-law scaling in the inertial subrange. This consistency suggests that turbulence in these regions exhibits the expected quasi-isotropic cascade behavior. At higher altitudes, progressing into the upper boundary layer and free troposphere, the spectral slope and correlation coefficient values demonstrate increased variability for all three wind components. This broadening in parameter space reflects the transition to more laminar, or large-scale dynamical regimes, where turbulence becomes weaker, or increasingly intermittent .~~ range to the integral length scale of each segment. In weakly turbulent H-type segments, L_W typically reaches 200 m, so the integration extends over a substantial wavenumber range before encountering the noise floor,

yielding meaningful ε estimates even in relatively quiescent conditions. In highly turbulent boundary-layer segments, the integral length scale is tens of meters and the full inertial subrange up to the Nyquist limit is retained.

Similar methodological choices have been made in comparable aircraft studies (Brilouet et al., 2021; Shaw and Businger, 1985; Wacławc , enabling direct comparison with prior field campaigns. Users requiring precise absolute values of ε at low turbulence levels are advised to consult the corresponding spectral slopes ($Slope_W$) and integral length scale (L_W) included in the segment statistics files as additional quality indicators that can be used to test the validity of the assumptions used.

3.3 Quality assessment

It is important to note that the accuracy of turbulence statistics is inherently dependent on the intensity of turbulence itself. In weakly turbulent, intermittent, or near-laminar conditions, predominantly for some tropospheric segments (B-, H-, T-, and M-types), the small magnitude of fluctuations may approach instrumental detection limits, making statistical estimation more uncertain. These conditions are identifiable through multiple quality indicators: low variance values, relatively high random errors, spectral slopes near or above -1 signalling possible noise contamination, and high heterogeneity scores $\mathcal{H}$, which often indicate transitions between turbulent and laminar regimes or intermittent turbulence. Users should interpret moments from weakly turbulent segments with caution.

3.3.1 ~~Errors associated with turbulent moments~~Random error

3.3.2 ~~Random error~~

Beyond instrumental limitations, the reliability of ~~turbulent moment estimations depends on the sampling and filtering~~ statistical moment estimates depends on sampling conditions. Following ~~Lenschow et al. (1994), each of these effects can be accounted for by computing random and systematic errors, respectively. The random error (ϵ_r)~~ Lenschow et al. (1994), the random error ϵ_r is inherent to the ~~sampling conditions, in particular the finite~~ finite segment length relative to the integral length ~~of the sample used to compute the turbulent moments relatively to the corresponding integral length scales~~scale. For any two variables x and y~~, the random error associated with their covariance can be estimated as:~~ :

$$\epsilon_{\underline{r}r} = \sqrt{\frac{2L_{xy}}{L}}\sqrt{1+\frac{1}{r_{xy}^2}}\sqrt{1+\frac{1}{r_{xy}^2}} \tag{9}$$

where L_{xy} is the integral length scale of $x'y'$, L the ~~length of the used segment~~ segment length, and r_{xy} ~~is~~ the correlation coefficient between x and y. For the variance of a single variable x~~, the random error is defined as:~~ :

$$\epsilon_{\underline{r}r} = 2\times\sqrt{\frac{L_x}{L}}\sqrt{\frac{L_x}{L}} \tag{10}$$

In both cases, ~~obtaining reliable estimates (low values of ϵ_r) requires the sample length to be significantly larger that the integral length scale ($L >> L_{xy}$ or $L >> L_x$, respectively). Figure~~ reliable estimates require $L \gg L_{xy}$ (or $L \gg L_x$).



Figure 7. Histograms of random errors for the variances of all five variables, for the whole campaign and each segment type. Colors indicate the heterogeneity class of each segment.

Figure 7 shows the histograms of ~~random errors associated with the variances of all 5 variables~~ ϵ_r for the whole campaign and for ~~each type of segments. All 5 variances exhibit similar trend with respect to the 6 segment types. Variance estimates are relevant mostly for boundary-layer categories.~~ S-, L-, B-, and H-type segments. For the three boundary-layer categories (S, L, ~~and~~ B), most ~~of the~~ segments show variance random ~~error~~ errors below 20%~~and~~, with positively skewed distributions. ~~This positive skewness suggests that while the majority of random error values remain low—indicating sufficiently large sample sizes for reliable estimates—there is a minority of segments that suffered sampling limitations. Interestingly, the B category, which was expected to be highly heterogeneous (as illustrated in Figure 3) due to its diverse conditions—such as intermittent cloud encounters, in-cloud segments, and out-of-cloud segments—displays a single-mode distribution of random errors. This is surprising, given that such variability should result in significantly different integral length scales and therefore different level of random errors giving a distribution with several modes.~~ For the H category, ~~which are the only statistically significant population of segments~~ the only segment type with a statistically large sample size outside the boundary layer, the ~~random~~

~~error~~ distributions are negatively skewed and centered around ~~values ranging from~~ 20~~to~~ –30%~~. Most segments exhibit high random errors, with a few segments still achieving lower error values. Such a pattern is consistent with sampling conditions in the free atmosphere, where~~ , consistent with the larger integral length scales ~~are usually obtained, leading to relatively high random errors. The few samples with smaller integral length scales indicate that in a few cases, turbulence might still be at play~~expected in the free atmosphere. The heterogeneity score coloring confirms overall agreement between $\mathcal{H}$ and ϵ_r: homogeneous segments (green) exhibit low random errors, while the high ϵ_r values in H-type segments with elevated $\mathcal{H}$ (red, orange) reflect the mismatch between the fixed segment length and the large integral scales of laminar or intermittent flow.

3.3.3 Systematic error



Figure 8. Histograms of ~~random errors associated with~~ the ~~variances of~~ filtering-induced variance loss for all ~~5~~ five variables, for the whole campaign and ~~for~~ each segment type~~of segments. Colours indicate heterogeneity class of corresponding segments.~~

3.3.4 ~~Systematic error~~

~~As mentioned in Subsection ??, the~~

The definition of the mean field ~~and consequently, the method used for its computation and calibration,~~ directly influences the magnitude of the fluctuations~~. The systematic error stems from the loss of information due to the choices made for computing fluctuations, in this case~~ , and the choice of a 5-km high-pass ~~filtering. To estimate this loss,~~ cutoff removes variance at scales larger than 5 km. To quantify this filtering effect, which is called systematic error (ϵ_S) the covariance of the high-pass-filtered series (~~F_{fil}~~$F_{\rm fil}$) is compared to ~~the covariance~~ that of the detrended series (~~F_{det}~~): $F_{\rm det}$):

$$\epsilon_{sf} = \frac{F_{det} - F_{fil}}{F_{det}} \frac{F_{\rm det} - F_{\rm fil}}{F_{\rm det}} \tag{11}$$

~~These errors are~~ This quantity is computed for every ~~second order moment , including~~ second-order moment (variances and covariances~~, and are~~) and provided in the dataset~~in order to give an estimate of the removed variance due to scales larger than 5 km. Figure~~ . It is not a systematic error in the classical sense (cf. Lenschow et al., 1994, who define systematic error as the inherent bias fro . Instead, it quantifies the variance suppressed by the choice of cutoff scale.

Figure 8 shows the histograms of ~~systematic errors associated with the variances of all 5 variables for the whole campaign and for each type of segments~~ϵ_f for all five variables and each segment type. For the ~~three~~ boundary-layer ~~segment~~ types (S, L, ~~and~~ B), the ~~histograms are very different in the case of~~ distribution of ϵ_f for vertical wind variance ~~relatively to the case of horizontal components, with a much narrower distribution positioned~~ is much narrower and confined entirely below 25%~~. This is explained by the very little effect large scales have on~~ , reflecting the small contribution of large scales to the vertical wind in the ~~turbulent~~ mixed layer. For the other ~~4 variables, even in the boundary layer, much~~ four variables, substantially more variance is ~~usually removed with the 5 km scale~~ removed by the 5-km filter, revealing the relatively ~~higher influence of the meso-scale on these thermodynamic parameters.~~ stronger mesoscale influence on thermodynamic fields. The heterogeneity score coloring shows a strong link between $\mathcal{H}$ and ϵ_f: non-linear trends in the large-scale component, associated with regime transitions or mesoscale structures, produce both high heterogeneity scores and large filtering effects.

~~Histograms of systematic errors associated with the variances of all 5 variables for the whole campaign and for each type of segments.~~

4 Dataset summary

~~As an illustration of the turbulence measured during the MAESTRO campaign, Figure 9 presents histograms of the variances, computed from the high-pass filtered series for each segment category. The instruments used to compute these moments are the one selected for the dataset, after corrections and calibrations (see subsections ?? and ??). For the three wind components, the variance is usually higher within the boundary layer, in particular in L-segments. These high variance values indicate more turbulent wind behavior near the ground, which is characteristic of the atmospheric boundary layer. More precisely, it can be related to thermal convection or shear-driven turbulence, for example. For temperature and water vapor mixing ratio,~~

~~cloud bases (B-type segments) are associated with the highest variance values. The usually sharp gradients in temperature and humidity at cloud base contribute to enhanced fluctuations, driven by latent heat release and vertical mixing. Some of these higher variance values can therefore be attributed to the crossing of this inversion in the case of relatively high heterogeneity segments. However, they may also be indicative of processes related to the dynamics of clouds such as condensation, evaporation, and entrainment.~~

~~Histograms of computed variance for each segment category and the 5 high-pass filtered variables.~~

~~Figure 10 shows the TKE dissipation rate measured during MAESTRO. The general trend is an increase towards the sea surface (segments L and S) whose roughness contributes to turbulence. The obtained values for boundary-layer segments (L and S types) are in-line with the one that can be found in the literature (for example in Lambert and Durand (1998)) ranging between 10^{-4} and 10^{-3} $m^2.s^{-3}$. For higher altitudes, the values decrease substantially as flow regimes become more laminar and less energetic.~~

~~Histograms of TKE dissipation rates for the turbulence dataset.~~

~~The dataset is~~

The nomenclature for the dataset file names and their content is given in Table 3. The dataset is divided by flight, ~~and for each flight three kinds of NetCDF files are generated~~legs and segments as explained below:

1. ~~Segment-scale moments~~ SEGMENT_STATISTICS (one file per flight). The ~~turbulent~~ statistical moments and associated errors, along with segment-averaged values for each thermodynamic variable~~(referenced in Table ??)~~, are provided for each segment. Additionally, the starting, and ending coordinates, mean true heading, true airspeed, and ground speed for a given ~~flight~~ segment are included.

2. ~~Segment-scale time series~~ SEGMENT_TIMESERIES (one file per segment). For each segment, one file is generated that includes the 25 Hz geographical coordinates, the corrected horizontal wind components in the geographic, stream-wise, and plane's frame of reference, the corrected vertical wind component, the calibrated water vapor mixing ratio, and potential temperature. In addition, fluctuations of the horizontal stream-wise wind components, vertical wind component, water vapor mixing ratio, and potential temperature are provided for high-pass filtered data, as well as detrended series.

3. ~~Leg-scale time series and moments~~ LEG_TIMESERIES (one file per ~~leg). Files similar to the two previous categories are also generated at the scale of entire legs~~flight elements). For each leg, or specific flight element (D, A, and V, meaning decent, ascent, and Velocity Azimuth Display, respectively) one file is generated with a 25 Hz temporal resolution, without constraints on homogeneity. The variables within these files include 3D-Wind in the three coordinate systems (streamwise, aircraft frame of reference, and geographical), absolute humidity, and temperature. The purpose of this approach is to provide a suitable dataset for users wanting to study the variability along ~~the entire leg~~ entire legs from turbulence to mesoscale, develop specific analyses, revise segmentation, ~~or~~ identify specific meteorological patterns or events in the time series, ~~such as thermals or cool pools~~or use ascents and descents to obtain vertical atmospheric profiles.

~~The nomenclature~~

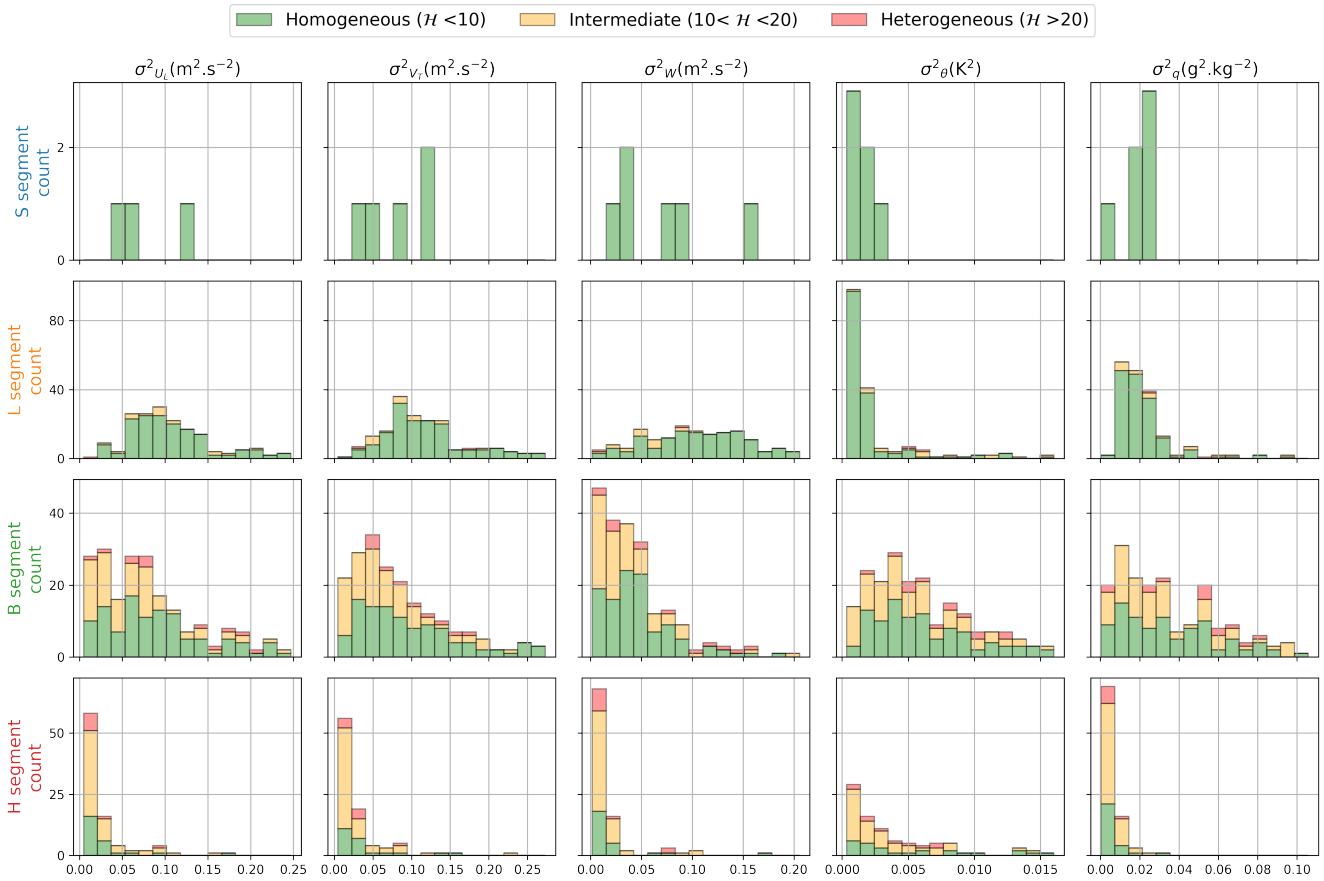


Figure 9. Histograms of computed variance for each segment category and the 5 high-pass filtered variables.

As a preview of the turbulence measured during the MAESTRO campaign, Figure 9 presents histograms of the variances, computed from the high-pass filtered series for each segment category. The instruments used to compute these moments are the one selected for the dataset~~file names is given in Table **??**.~~ , after corrections and calibrations, as described in section 3. For the three streamwise wind components, the variance is usually higher within the boundary layer, in particular in L-segments. These high variance values indicate more turbulent flow near the ocean surface, which is characteristic of the atmospheric boundary layer.

~~File nomenclature for the MAESTRO turbulence dataset. *YYYYMMDD* corresponds to the flight date (e.g., 20240815), *NN* to the campaign flight number (e.g., 09), *ZZ* to the SAFIRE flight number identifier, *X* to the leg identifier and *i* the segment identifier (e.g., L2_1, the first segment of L2), and *V* the file version. **Product Moments** MAESTRO_ORCESTRA_ATR_TURBULENCE_nc **Fluctuations** MAESTRO_ORCESTRA_ATR_TURBULENCE_FLUCTUATIONS_RF*NN*_as2400*ZZ*_L*X*_*i*_*YYYYMMDD*_v*V*.nc **Leg scale calibrated time series** MAESTRO_ORCESTRA_ATR_TURBULENCE_CALIBRATED_RF*NN*_as2400*ZZ*_L*X*_*YYYYMMDD*_v*V*.nc~~ For temperature and water vapor mixing ratio, cloud-base segments (B-type) are associated with the

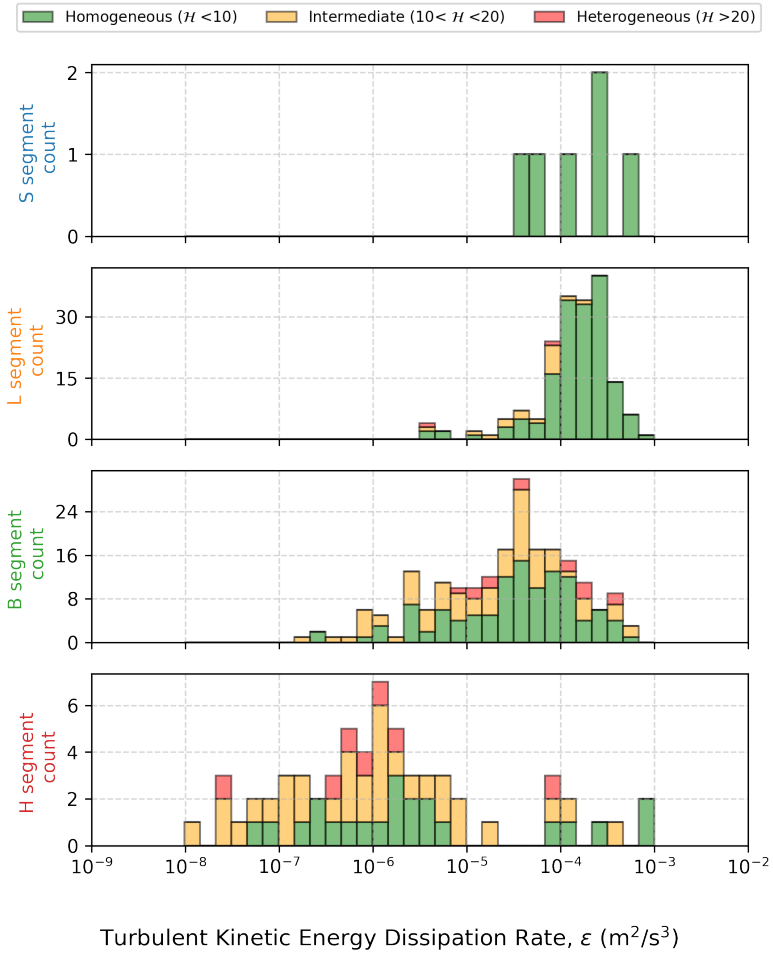


Figure 10. Histograms of TKE dissipation rates for the turbulence dataset.

highest variance values. B-type legs are horizontal transects flown at cloud-base level. Two possible contributing factors can be identified: (1) the aircraft may alternately encounter clear air and cloud interior during partially cloudy segments, generating along-track horizontal gradients in both T and q due to the contrasting thermodynamic properties of these two environments, and (2) even in segments that are fully within the clouds, the cloud interior itself is a region of strong thermodynamic heterogeneity, where active condensation, evaporation, and mixing processes produce large local fluctuations in temperature and humidity.

Figure 10 shows the TKE dissipation rate measured during MAESTRO. The general trend is an increase towards the sea surface (segments L and S) whose roughness contributes to turbulence. The obtained values for mixed layer segments (L and S types) are in-line with the one that can be found in the literature (for example in Lambert and Durand (1998)) ranging between

10^{-4} and 10^{-3} $m^2.s^{-3}$. For higher altitudes, the values decrease substantially as flow regimes become more laminar and less energetic.

5 Conclusions

The article presents the MAESTRO turbulence dataset, based on the in-situ wind, temperature, and moisture measurements from the instrumented ATR-42 from SAFIRE during the MAESTRO field campaign. It presented the sampling strategy and processing details. The quality of the variables of the dataset has been assessed, including the computation of systematic and random errors. On the one hand, the moment dataset can be used to perform process studies or model validation, as it summarizes the turbulence and mean data effectively in few key metrics that can be easily compared across segment types or flights. On the other hand, leg-scale time series dataset can be used to identify larger scales spatial variations that were cut off by the filtering using the calibrated time series, and ~~turbulent~~ fluctuations enable finer analysis to reprocess the data, or perform specific analyses with various methods such as wavelet analysis. Providing both approaches ensures the best and broadest utilization of the turbulence data collected during the MAESTRO field campaign.

With the data quality analysis, a few comments and recommendations for users of this dataset can be formulated. Nan values can be present in the moment and legs time series files. In the moment files, inertial slopes may contain NaN values if the inertial range is too small to be able to derive a reliable slope estimate or if the regression coefficient is below 0.9. The TKE dissipation rate ε can also take NaN values if the inertial range of the vertical wind component spectrum is too small. In the leg time series, NaN values are associated with INS malfunctions affecting altitude, latitude, longitude, static pressure, true heading, and all wind components. NaN values can be found in 12 of the 269 leg time series files. In the segment time series, NaN values are avoided through the use of the segmentation algorithm. For turbulence statistics, segments with a heterogeneity score below 30 can be considered homogeneous and should be prioritized. Additionally, the treatment of FAST-WAVE data will be improved to better resolve frequencies higher than 25 Hz. A dedicated 100 Hz dataset will be released on the AERIS website following this publication.

While being a valuable resource for the broad atmospheric research community, the present dataset was primarily collected to address the MAESTRO research questions. The data presented in this article will be used to examine the relationships between turbulence characteristics within the marine boundary layer and the organization of convection in the tropics. As one of several MAESTRO and ORCESTRA datasets, it will be integrated with other observations to conduct analyses targeted at these research questions. Other datasets include sounding near Sal airport, airborne remote sensing instruments, satellite observations, and in-situ microphysics. Among the satellite products that were gathered during the campaign, SAR images are particularly promising. They document the centimeter-scale waves at the surface of the ocean (Atlas, 1994), thus they are relevant to the study of turbulence in the marine boundary layer and sea-air-cloud interactions (Brilouet et al., 2023; Kudryavtsev et al., 2005). Seven flights of the campaign were planned with respect to these acquisitions so that the plane was concomitantly performing a full leg at 500 ft above the surface of the ocean. Exploring these data could lead to better interpretations of SAR images and plane turbulence data, and ultimately improve the current understanding of air-sea interactions within the tropical

convective environment. This represents one of the direct research avenues for this dataset and a potential building block for further analyses, hopefully advancing the more fundamental questions that motivated MAESTRO and ORCESTRA.

6 Code and data availability

Code is availble on request to the corresponding authors. The turbulence dataset is available at https://doi.org/10.25326/812 (last access:24/09/2025, Jaffeux and Lothon (2025)). The raw datasets used for this study are available on the AERIS website (last access:02/03/2026, Bony (2024a) and Bony (2024b)).

Appendix A: Wind Corrections

The air velocity components are obtained by subtracting the ground velocity of the aircraft, measured by the INS, from the air velocity with respect to the aircraft, deduced from differential pressures measured by the five-hole nose radome and a Pitot tube. While gust probes are calibrated to mitigate flow disturbance effects, multiple sources of error can affect wind retrieval, including uncertainties in instrument alignment and performance. The full wind component equations in geographical coordinate are the following:

$$U_{geo} = -U_a \cdot D^{-1}\Big[\sin\psi\cos\theta + \tan\beta\big(\cos\psi\cos\phi + \sin\psi\sin\theta\sin\phi\big) + \tan\alpha\big(\sin\psi\sin\theta\cos\phi - \cos\psi\sin\phi\big)\Big] + U_p - L\big(\dot\theta\sin\theta\sin\psi - \dot\psi\cos\psi\sin\theta\big) \tag{A1}$$

$$V_{geo} = -U_a \cdot D^{-1}\Big[\cos\psi\cos\theta - \tan\beta\big(\sin\psi\cos\phi - \cos\psi\sin\theta\sin\phi\big) + \tan\alpha\big(\cos\psi\sin\theta\cos\phi + \sin\psi\sin\phi\big)\Big] + V_p - L\big(\dot\psi\sin\psi\cos\theta + \dot\theta\cos\psi\sin\theta\big) \tag{A2}$$

$$W = -U_a \cdot D^{-1}\Big[\sin\theta - \tan\beta\cos\theta\sin\phi - \tan\alpha\cos\theta\cos\phi\Big] + W_p + L\dot\theta\cos\theta \tag{A3}$$

where: $D = \sqrt{1+\tan^2\alpha+\tan^2\beta}$ is the normalization factor of the rotation matrix, ψ is the aircraft heading angle, θ is the aircraft pitch angle, ϕ is the aircraft roll angle, α is the angle of attack, β is the angle of sideslip, U_a is the true airspeed, U_p, V_p, W_p are the airplane speeds relative to the ground in the east, north, and vertical directions, respectively, $\dot\theta = \frac{d\theta}{dt}$ and $\dot\psi = \frac{d\psi}{dt}$ are the time derivatives of pitch and heading angles, $L = 8.05$ m, is the distance between the INS and the radome, U_{geo}, V_{geo}, W are the wind velocity components in the east, north, and vertical directions, respectively.

Initial analysis of the MAESTRO dataset revealed two significant indicators of wind measurement issues: (1) non-negligible averaged W values during stabilized legs (0.6 $\pm$0.1 m.s^{-1}, whereas values close to 0 $\pm$0.1 m.s^{-1} are expected for sufficiently long averaging periods), and (2) substantial correlation (25%) between ϕ and W outside of straight legs. These findings motivated dedicated calibration maneuvers during RF22 to diagnose and correct the calibration of the radome attack angles.

Following the methodology of Lenschow (1986) and similar to Mallaun et al. (2015), we performed a parameter optimization using data from all stabilized legs to compute biases in both radome angles (α and β). The approach minimizes the campaign-averaged vertical wind by optimizing constant biases for α and β, assuming a non-drifting systematic offset common to the entire dataset. This simultaneous optimization of both angle biases reduced the roll-vertical wind correlation from 25% to 1.4%, and yielded biases of -0.015° and -0.545° for the attack and sideslip angles, respectively.

Calibration maneuvers were performed during RF22, on the 7th of September 2024, as described in Lenschow (1986) to confirm and resolve the apparent wind measurement issue. In particular, a reverse heading maneuver was performed by flying at constant height and true heading starting with the wind in the back, then taking a 90° heading change to the right, after that taking a 180° heading change to the right, and finally taking a 90° turn to the left. This maneuver was performed over 27 minutes at an altitude of 2.2 km, outside of the boundary layer. The initially obtained horizontal wind (Figure A1a) strongly varies in direction and strength when crossing the same air volume after a 90 and 180 degrees heading change. Moreover the vertical wind shows a strong positive change as the plane turns to the right, and a strong negative change as the plane turns to the left (see Figure A1c). This is an example of the vertical wind being heavily correlated with the roll angle of the aircraft. The wind field was recomputed incorporating these bias corrections for all MAESTRO flights. Validation using a reverse heading maneuver confirms that the corrected horizontal wind maintains consistent direction and magnitude when the aircraft traverses the same air volume at different headings (see Figure A1c), and that vertical wind is now independent of aircraft roll (see Figure A1d).

Appendix B: Intercomparison of humidity reference sensors

This analysis evaluates reference instruments against their known limitations and examines their mutual agreement under varying atmospheric conditions. In principle, the 1011C chilled-mirror hygrometer provides the most accurate humidity measurement available on the ATR-42, as it directly measures dew point temperature through condensation/evaporation rather than relying on calibrated spectroscopic or capacitive responses. This makes it the natural reference for evaluating the other instruments. However, its thermal response time (limited to 1 $K.s^{-1}$ warming rate) constrains its ability to track rapid humidity changes. In heterogeneous atmospheric conditions, this limitation can introduce systematic errors, as discussed below. The WVSS-2 is a calibrated laser absorption spectrometer with negligible response time, while the HUMAERO is a calibrated capacitive sensor sensitive to relative humidity, which is then converted to mixing ratio using an associated temperature measurement. Unlike the 1011C, both the WVSS-2 and HUMAERO acquire independent, near-instantaneous measurements at each sampling point.

Figure B1 compares segment-averaged mixing ratio values from the WVSS-2 and HUMAERO against those from the 1011C. The WVSS-2 agrees with the 1011C within ±5% for 509 segments (90.4% of the dataset), whereas the HUMAERO agrees for only 236 segments (41.9%). The HUMAERO systematically overestimates mixing ratio by more than 5% in 317 segments (56.3%). In contrast, the WVSS-2 shows excellent agreement across nearly all segment types, with deviations primarily confined to specific conditions (ice clouds and high-humidity segments) that can be physically explained (see below). Given

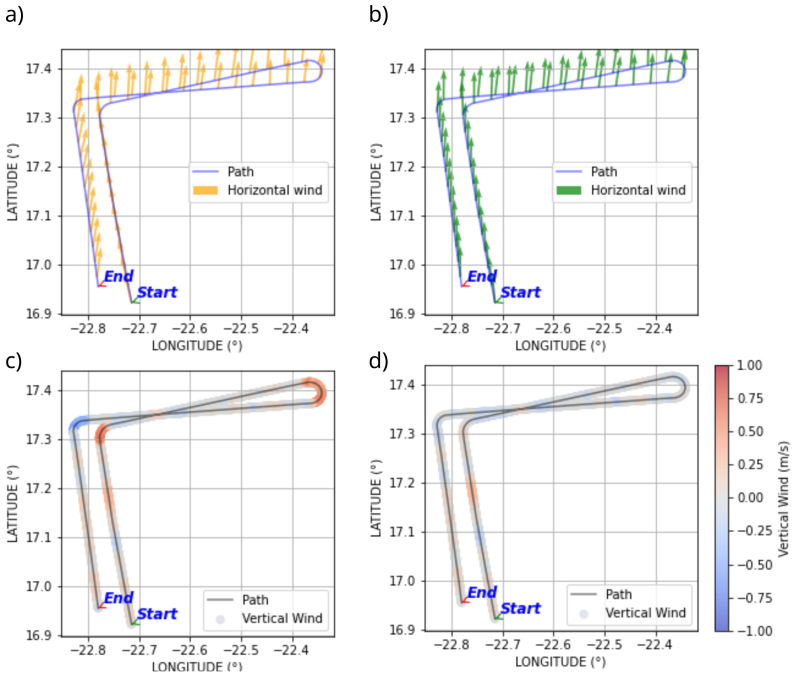


Figure A1. ~~Segmentation algorithm flowchart~~Computed wind vectors during the reverse heading maneuver with and without wind corrections. Initially computed horizontal (a) and vertical (c) winds along the plane's trajectory. Corrected horizontal (b) and vertical (d) winds along the plane's trajectory. This maneuvers took place during RF22 on September 7, 2024.

that the majority of segments were acquired in clear-sky conditions, where the 1011C's thermal inertia is less problematic and its measurements are expected to be accurate, the underestimation by the HUMAERO cannot be attributed to limitations of the reference instrument. The capacitive sensor's conversion from relative humidity to mixing ratio also introduces additional uncertainty through its dependence on accurate temperature measurement and calibration stability under varying ambient conditions. Based on this evidence, the HUMAERO is excluded from further consideration as a reliable reference sensor for the MAESTRO dataset.

Figure B2 shows example time series from one of the mixed layer segments and one of the cloud base segments for the two remaining reference sensors (WVSS-2 and 1011C). The 1011C exhibits clear ramping phases when adapting to higher humidity/dew point values, while it cools down rapidly to measure decreasing humidity gradients thanks to the strong airstream. This asymmetric response to positive and negative humidity gradients causes a systematic underestimation of the average mixing ratio during such segments where local heterogeneity of the moisture field exceeds the adaptation capacity of the mirror. Figure B3 shows average values for each segment, obtained from the two reference sensors and sorted by segment type. Most cloud base segments (in green) fall below the 1:1 line, reflecting the fact that the behavior described in Figure B2. Conversely, L-type segments generally lie slightly above the 1:1 line. These latter segments are very moist and approach the edge of the WVSS-2 calibration range. Similar issues may affect B-types as well.

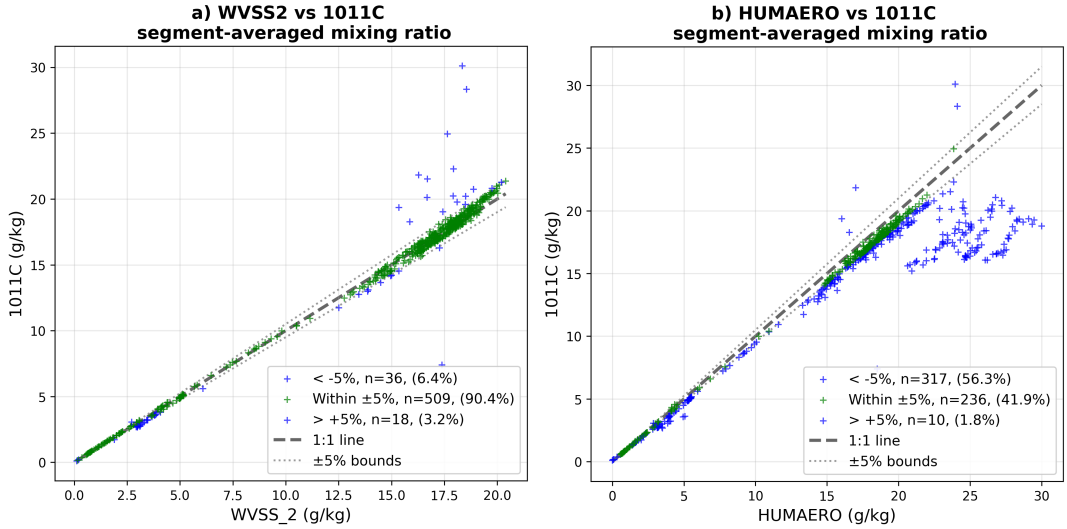


Figure B1. Comparison of mixing ratio values averaged over segments for the WVSS-2 (a) and HUMAERO (b) with respect to the 1011C instrument. In green, points where instruments agree within ±5%, in blue points where the 1011C measurements are lower by more than 5%, and in red higher than 5%.

Within the 36 segments (6.4% of the dataset) for which the WVSS-2 values exceed those of the 1011C by more than 5%, 26 belong to saturated H-type segments (traversing ice or mixed phase clouds), 6 to B-type segments, and 4 to M-type segments. When the deviation threshold is narrowed to 2%, 109 segments exhibit deviations below this threshold, including 59 B-type, 39 H-type, 3 T-type, and 8 M-type segments. These results support the sustained hypothesis, attributing the disagreement between the WVSS-2 and 1011C to limitations of the 1011C rather than overestimation by the WVSS-2. Of the 18 segments for which the 1011C measured significantly higher mixing ratios compared to the WVSS-2, 14 belong to the B-type, 1 each to the H- and S-types, and 2 to the L-type. These segments are characterized by high relative humidities or cloud fraction. Inspection of these few anomalous segments reveals large mixing ratio jumps exceeding 10 g/kg in the 1011C signal that strongly affect the mean values. Although we cannot provide a physical explanation for this behavior, an environmental origin is ruled out, as none of the other five instruments recorded similar patterns, and such variability is not expected in the atmosphere. No post-processing correction was applied to compensate for the 1011C's thermal inertia. Inversion techniques to correct for the limited warming rate (1 $K.s^{-1}$) would likely introduce additional uncertainty in particular when combined with the calibration of fast sensors. Instead, we account for this limitation by selecting WVSS-2 as the primary reference, which combines robustness and high accuracy under all MAESTRO conditions.

Author contributions. LJa wrote the article; LJa wrote the analysis code; ML and LJa analysed the data; ML, SB, FC, DB revised the manuscript. SAFIRE processing team (GC, CSL, HB, HO, DC, TJ) produced and curated the initial ATR 25Hz dataset which was used to

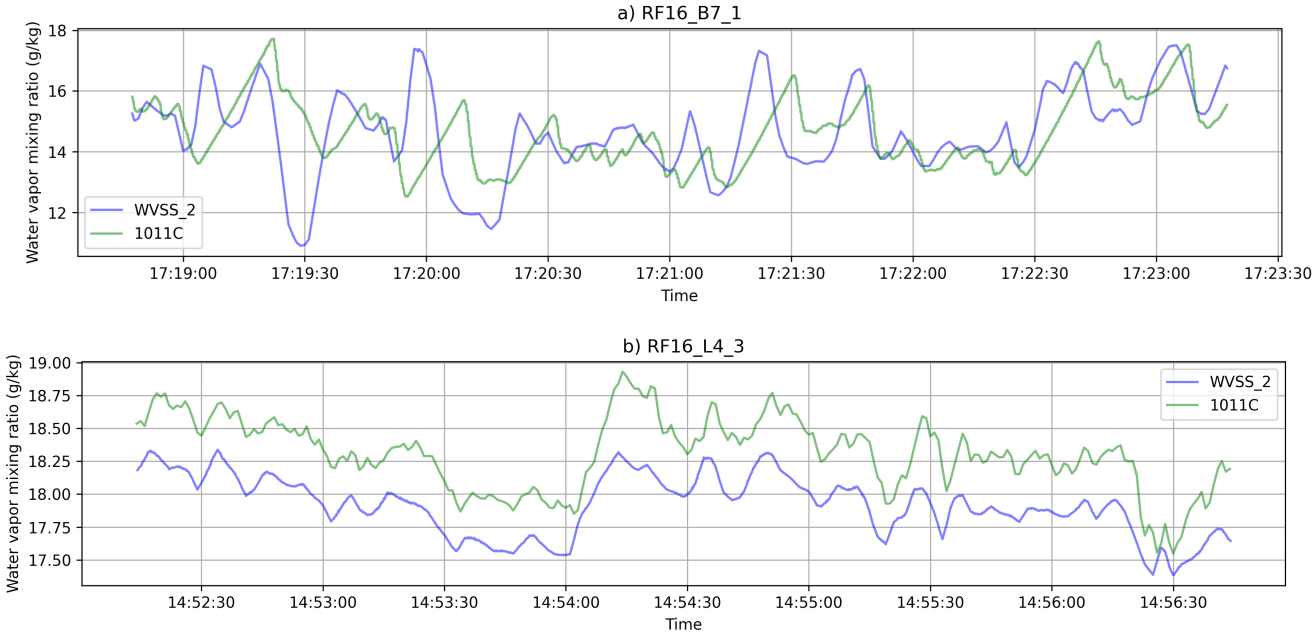


Figure B2. Example of mixing ratio time series from the 1011C and WVSS-2 instruments, showing ramping structures due to the relatively slow warming rate of the 1011C, compared with instantaneous recording of laser absorption from the WVSS-2 for a B-type segment(a) with cloud encounters and a clear sky L-type segment(b) from the same RF16 flight.

produce the final turbulence dataset. JB and LJo adapted the FAST-WAVE hygrometer to be embarked on the ATR and curated its data. SB is the lead scientist and coordinator of the MAESTRO campaign.

Competing interests. The authors declare no competing interests.

Acknowledgements. ATR42 Airborne data were obtained using the aircraft managed by Safire, the French facility for airborne research, an infrastructure of the French National Center for Scientific Research (CNRS), Météo-France, and the French National Center for Space Studies (CNES). The authors gratefully acknowledge all the SAFIRE staff, technicians, engineers, pilots and directors for their considerable efforts and involvement in the realization of the MAESTRO airborne operations. This research received funding from ERC Advanced Grant (MAESTRO grant no. 101098063) and ESA (contract no. 281042).

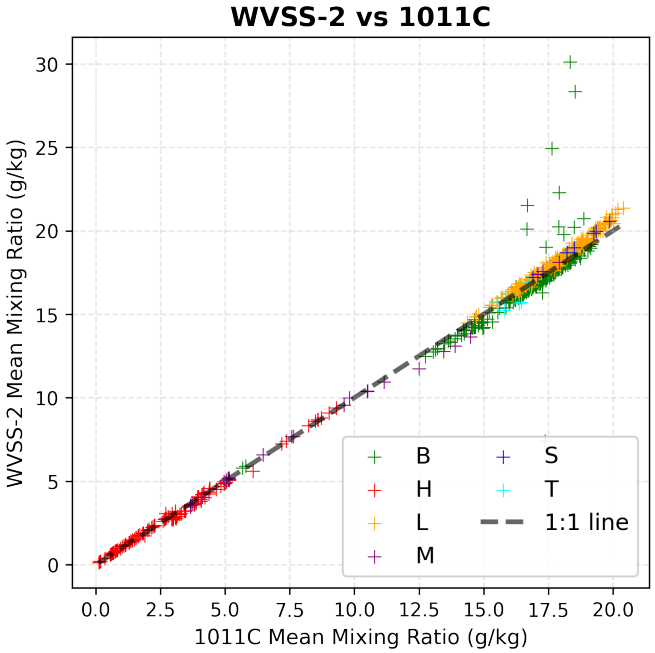


Figure B3. Comparison of mixing ratio values averaged over segments for the 1011C and WVSS-2, by segments types.

References

Asiyabi, R. M., Ghorbanian, A., Tameh, S. N., Amani, M., Jin, S., and Mohammadzadeh, A.: Synthetic aperture radar (SAR) for ocean: A review, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 16, 9106–9138, 2023.

Atlas, D.: Footprints of storms on the sea: A view from spaceborne synthetic aperture radar, Journal of Geophysical Research: Oceans, 99, 7961–7969, 1994.

Ballantyne, A. and Moss, J.: Fine wire thermocouple measurements of fluctuating temperature, Combustion Science and Technology, 17, 63–72, 1977.

Bell, M.: Collaborative Research: AGS-FIRP Track 2-Process Investigation of Clouds and Convective Organization over the atLantic Ocean (PICCOLO), NSF Award Number 2331202. Directorate for Geosciences, 23, 31 202, 2024.

Bernard-Trottolo, S.: Diagnostic de la présence de structures cohérentes au sein de la couche limite atmosphérique: campagnes TRAC98 et MAP99, Ph.D. thesis, Toulouse 3, 2001.

Bony, S.: MAESTRO-2024_SAFIRE-ATR42_SAFIRE_FAST Temperature and Humidity Fast Observation 25HZ, https://doi.org/10.25326/738, [dataset], 2024a.

Bony, S.: MAESTRO-2024$_S AFIRE-ATR42_S AFIRE_C ORE_T DYN Thermodynamic and dynamic data 1HZ$, https://doi.org/10.25326/744, [da

Bony, S. and Dufresne, J.-L.: Marine boundary layer clouds at the heart of tropical cloud feedback uncertainties in climate models, Geophysical Research Letters, 32, 2005.

Bony, S. and Stevens, B.: Measuring area-averaged vertical motions with dropsondes, Journal of the Atmospheric Sciences, 76, 767–783, 2019.

Bony, S., Stevens, B., Ament, F., Bigorre, S., Chazette, P., Crewell, S., Delanoë, J., Emanuel, K., Farrell, D., Flamant, C., et al.: EUREC 4 A: A field campaign to elucidate the couplings between clouds, convection and circulation, Surveys in Geophysics, 38, 1529–1568, 2017.

Bony, S., Schulz, H., Vial, J., and Stevens, B.: Sugar, Gravel, Fish, and Flowers: Dependence of Mesoscale Patterns of Trade-Wind Clouds on Environmental Conditions, Geophysical Research Letters, 47, e2019GL085 988, https://doi.org/https://doi.org/10.1029/2019GL085988, e2019GL085988 10.1029/2019GL085988, 2020.

Bony, S., Lothon, M., Delanoë, J., Coutris, P., Etienne, J.-C., Aemisegger, F., Albright, A. L., André, T., Bellec, H., Baron, A., et al.: EUREC 4 A observations from the SAFIRE ATR42 aircraft, Earth System Science Data Discussions, 2022, 1–61, 2022.

Bony, S., Poujol, B., Mckim, B., Rochetin, N., Lothon, M., Windmiller, J., Maury, N., Dufaux, C., Jaffeux, L., Chazette, P., et al.: Evidence for the role of thermal and cloud merging in mesoscale convective organization, Atmospheric Chemistry and Physics, 25, 17 331–17 362, 2025.

Brilouet, P.-E., Lothon, M., Etienne, J.-C., Richard, P., Bony, S., Lernoult, J., Bellec, H., Vergez, G., Perrin, T., Delanoë, J., et al.: The EUREC 4 A turbulence dataset derived from the SAFIRE ATR 42 aircraft, Earth System Science Data, 13, 3379–3398, 2021.

Brilouet, P.-E., Bouniol, D., Couvreux, F., Ayet, A., Granero-Belinchon, C., Lothon, M., and Mouche, A.: Trade wind boundary layer turbulence and shallow precipitating convection: New insights combining SAR images, satellite brightness temperature, and airborne in situ measurements, Geophysical Research Letters, 50, e2022GL102 180, 2023.

Brown, E. N., Friehe, C., and Lenschow, D.: The use of pressure fluctuations on the nose of an aircraft for measuring air motion, Journal of Applied Meteorology and Climatology, 22, 171–180, 1983.

Dauhut, T., Couvreux, F., Bouniol, D., Beucher, F., Volkmer, L., Pörtge, V., Schäfer, M., Ayet, A., Brilouet, P.-E., Jacob, M., et al.: Flower trade-wind clouds are shallow mesoscale convective systems, Quarterly Journal of the Royal Meteorological Society, 149, 325–347, 2023.

de Szoeke, S. P., Skyllingstad, E. D., Zuidema, P., and Chandra, A. S.: Cold pools and their influence on the tropical marine boundary layer, Journal of the Atmospheric Sciences, 74, 1149–1168, 2017.

Delanoë, J., Protat, A., Jourdan, O., Pelon, J., Papazzoni, M., Dupuy, R., Gayet, J.-F., and Jouan, C.: Comparison of airborne in situ, airborne radar–lidar, and spaceborne radar–lidar retrievals of polar ice cloud properties sampled during the POLARCAT campaign, Journal of Atmospheric and Oceanic Technology, 30, 57–73, 2013.

Derrien, M. and Le Gléau, H.: MSG/SEVIRI cloud mask and type from SAFNWC, International Journal of Remote Sensing, 26, 4707–4732, 2005.

George, G., Stevens, B., Bony, S., Pincus, R., Fairall, C., Schulz, H., Kölling, T., Kalen, Q. T., Klingebiel, M., Konow, H., et al.: JOANNE: Joint dropsonde Observations of the Atmosphere in tropical North atlaNtic meso-scale Environments, Earth System Science Data, 13, 5253–5272, 2021.

Giez, A., Mallaun, C., Nenakhov, V., and Zöger, M.: Calibration of a nose boom mounted airflow sensor on an atmospheric research aircraft by inflight maneuvers, 2021.

Groß, S., Ewald, F., Stevens, B., Wirth, M., Dekoutsidis, G., Ehrlich, A., Kouklaki, D., Krüger, K., Rosenburg, S., Volkmer, L., et al.: Persistent EarthCARE underflight studies of the ITCZ and organized convection (PERCUSION): Contribution to EarthCARE Validation, EGUsphere, 2026, 1–41, 2026.

Jaffeux, L. and Lothon, M.: MAESTRO_2024_Turbulence_Dataset, https://doi.org/10.25326/812, [dataset], 2025.

Janssens, M., Vilà-Guerau de Arellano, J., Scheffer, M., Antonissen, C., Siebesma, A. P., and Glassmeier, F.: Cloud patterns in the trades have four interpretable dimensions, Geophysical Research Letters, 48, e2020GL091 001, 2021.

Jansson, F., Janssens, M., Grönqvist, J. H., Siebesma, A. P., Glassmeier, F., Attema, J., Azizi, V., Satoh, M., Sato, Y., Schulz, H., et al.: Cloud botany: Shallow cumulus clouds in an ensemble of idealized large-domain large-eddy simulations of the trades, Journal of Advances in Modeling Earth Systems, 15, e2023MS003 796, 2023.

Jen-La Plante, I., Ma, Y., Nurowska, K., Gerber, H., Khelif, D., Karpinska, K., Kopec, M. K., Kumala, W., and Malinowski, S. P.: Physics of Stratocumulus Top (POST): turbulence characteristics, Atmospheric Chemistry and Physics, 16, 9711–9725, 2016.

Joly, L., Maamary, R., Decarpenterie, T., Cousin, J., Dumelié, N., Chauvin, N., Legain, D., Tzanos, D., and Durry, G.: Atmospheric Measurements by Ultra-Light SpEctrometer (AMULSE) Dedicated to Vertical Profile in Situ Measurements of Carbon Dioxide (CO2) Under Weather Balloons: Instrumental Development and Field Application, Sensors, 16, 1609, 2016.

Khelif, D., Burns, S., and Friehe, C.: Improved wind measurements on research aircraft, Journal of atmospheric and oceanic technology, 16, 860–875, 1999.

Kolmogorov, A. N.: The local structure of turbulence in incompressible viscous fluid for very large Reynolds numbers, Proceedings of the Royal Society of London. Series A: Mathematical and Physical Sciences, 434, 9–13, 1991.

Konsta, D., Dufresne, J.-L., Chepfer, H., Vial, J., Koshiro, T., Kawai, H., Bodas-Salcedo, A., Roehrig, R., Watanabe, M., and Ogura, T.: Low-level marine tropical clouds in six CMIP6 models are too few, too bright but also too compact and too homogeneous, Geophysical Research Letters, 49, e2021GL097 593, 2022.

Kudryavtsev, V., Akimov, D., Johannessen, J., and Chapron, B.: On radar imaging of current features: 1. Model and comparison with observations, Journal of Geophysical Research: Oceans, 110, 2005.

Lambert, D. and Durand, P.: Aircraft to aircraft intercomparison during SEMAPHORE, Journal of Geophysical Research: Oceans, 103, 25 109–25 123, 1998.

LeMone, M. A. and Pennell, W. T.: The relationship of trade wind cumulus distribution to subcloud layer fluxes and structure, Monthly Weather Review, 104, 524–539, 1976.

Lenschow, D., Mann, J., and Kristensen, L.: How long is long enough when measuring fluxes and other turbulence statistics?, Journal of Atmospheric and Oceanic Technology, 11, 661–673, 1994.

Lenschow, D. H.: Aircraft measurements in the boundary layer, in: Probing the atmospheric boundary layer, pp. 39–55, Springer, 1986.

Malinowski, S. P., Gerber, H., Jen-La Plante, I., Kopec, M., Kumala, W., Nurowska, K., Chuang, P., Khelif, D., and Haman, K.: Physics of Stratocumulus Top (POST): turbulent mixing across capping inversion, Atmospheric Chemistry and Physics, 13, 12 171–12 186, 2013.

Mallaun, C., Giez, A., and Baumann, R.: Calibration of 3-D wind measurements on a single-engine research aircraft, Atmospheric Measurement Techniques, 8, 3177–3196, 2015.

Muschinski, A., Frehich, R., Jensen, M., Hugo, R., Hoff, A., Eaton, F., and Balsley, B.: Fine-scale measurements of turbulence in the lower troposphere: An intercomparison between a kite-and balloon-borne, and a helicopter-borne measurement system, Boundary-layer meteorology, 98, 219–250, 2001.

Nowak, J. L., Siebert, H., Szodry, K.-E., and Malinowski, S. P.: Coupled and decoupled stratocumulus-topped boundary layers: turbulence properties, Atmospheric Chemistry and Physics, 21, 10 965–10 991, 2021.

Schulz, H. and Stevens, B.: Evaluating large-domain, hecto-meter, large-eddy simulations of trade-wind clouds using EUREC4A data, Journal of Advances in Modeling Earth Systems, 15, e2023MS003 648, 2023.

Schulz, H., Eastman, R., and Stevens, B.: Characterization and evolution of organized shallow convection in the downstream North Atlantic trades, Journal of Geophysical Research: Atmospheres, 126, e2021JD034 575, 2021.

Shaw, W. J. and Businger, J. A.: Intermittency and the organization of turbulence in the near-neutral marine atmospheric boundary layer, Journal of Atmospheric Sciences, 42, 2563–2584, 1985.

Sreenivasan, K. R.: On the universality of the Kolmogorov constant, Physics of Fluids, 7, 2778–2784, 1995.

Stevens, B., Bony, S., Brogniez, H., Hentgen, L., Hohenegger, C., Kiemle, C., L'Ecuyer, T. S., Naumann, A. K., Schulz, H., Siebesma, P. A., et al.: Sugar, gravel, fish and flowers: Mesoscale cloud patterns in the trade winds, Quarterly Journal of the Royal Meteorological Society, 146, 141–152, 2020.

Stevens, B., Bony, S., David, R., Delanoë, J., Gross, S., Klocke, D., Windmiller, J., Wing, A., and Wu, Y.: Orcestra: Organized convection and earthcare studies over the tropical Atlantic, Tellus, 2025.

Wacławczyk, M., Ma, Y.-F., Kopeć, J. M., and Malinowski, S. P.: Novel approaches to estimating the turbulent kinetic energy dissipation rate from low-and moderate-resolution velocity fluctuation time series, Atmospheric Measurement Techniques, 10, 4573–4585, 2017.

Wacławczyk, M., Gozingan, A. S., Nzotungishaka, J., Mohammadi, M., and P. Malinowski, S.: Comparison of different techniques to calculate properties of atmospheric turbulence from low-resolution data, Atmosphere, 11, 199, 2020.

Warner, C., Simpson, J., Martin, D., Suchman, D., Mosher, F., and Reinking, R.: Shallow convection on day 261 of GATE/mesoscale arcs, Monthly Weather Review, 107, 1617–1635, 1979.

Wehr, T., Kubota, T., Tzeremes, G., Wallace, K., Nakatsuka, H., Ohno, Y., Koopman, R., Rusli, S., Kikuchi, M., Eisinger, M., et al.: The EarthCARE mission–science and system overview, Atmospheric Measurement Techniques, 16, 3581–3608, 2023.

Winkler, M., Rixen, M., Beucher, F., Couvreux, F., Nam, C. C., Peyrillé, P., Schmidt, H., Segura, H., Wieners, K.-H., Alkilani-Brown, E., et al.: RAPSODI: radiosonde atmospheric profiles from ship and island platforms during ORCESTRA, collected to Decipher the ITCZ, Earth System Science Data, 18, 1833–1854, 2026.

Wyngaard, J. and Coté, O.: The budgets of turbulent kinetic energy and temperature variance in the atmospheric surface layer, Journal of Atmospheric Sciences, 28, 190–201, 1971.

Zilitinkevich, S. S., Hunt, J., Esau, I. N., Grachev, A., Lalas, D., Akylas, E., Tombrou, M., Fairall, C., Fernando, H., Baklanov, A., et al.: The influence of large convective eddies on the surface-layer turbulence, Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography, 132, 1426–1456, 2006.

MAESTRO/SAFIRE name/Date	TO-L times (UTC-1)	Sampled conditions	Cloud cover	Max CT	Sampling Strategy
RF01 / as240023 / 10-08-2024	11:05 - 13:54	ShCu	47%	1.0 km	Coordination
RF02 / as240024 / 11-08-2024	14:33 - 18:13	ShCu	80%	2.0 km	Coordination
RF03 / as240025 / 13-08-2024	14:20 - 17:37	ShCu	19%	1.5 km	Coordination
RF04 / as240026 / 13-08-2024	19:11 - 22:34	ShCu, CP	27%	2.0 km	Standard
RF05 / as240027 / 15-08-2024	16:22 - 20:09	ShCu, AlCu	93%	7.0 km	Standard
RF06 / as240028 / 16-08-2024	10:36 - 14:17	ShCu, Rain, CP	55%	2.5 km	Standard
RF07 / as240029 / 16-08-2024	16:06 - 19:38	Sh and StCu	74%	2.5 km	Standard
RF08 / as240030 / 19-08-2024	15:02 - 18:29	Anvil Cloud	50%	7.0 km	Case Study
RF09 / as240031 / 20-08-2024	14:03 - 17:33	ShCu	51%	1.5 km	Coordination
RF10 / as240032 / 22-08-2024	13:55 - 17:32	ShCu	41%	2.5 km	Coordination
RF11 / as240033 / 22-08-2024	19:33 - 22:42	AlSt	40%	7.5 km	Standard
RF12 / as240034 / 25-08-2024	19:21 - 22:41	AlCu	87%	7.5 km	Standard
RF13 / as240035 / 26-08-2024	12:38 - 15:58	Cg, CL	33%	14.0 km	Standard
RF14 / as240036 / 26-08-2024	17:52 - 20:52	Cg, CL	40%	12.0 km	Case Study
RF15 / as240037 / 27-08-2024	19:00 - 22:34	CL	28%	3.0 km	Standard
RF16 / as240038 / 29-08-2024	13:52 - 17:40	ShCu	44%	1.5 km	Coordination
RF17 / as240039 / 30-08-2024	19:53 - 23:15	CL	75%	2.5 km	Standard
RF18 / as240040 / 31-08-2024	13:57 - 17:30	CL	58%	2.0 km	Coordination
RF19 / as240041 / 01-09-2024	19:03 - 22:33	CL	57%	2.0 km	Case Study
RF20 / as240042 / 02-09-2024	19:31 - 22:55	StCu	85%	1.0 km	Case Study
RF21 / as240043 / 03-09-2024	16:33 - 19:51	MCS	92%	13.0 km	Case Study
RF22 / as240044 / 07-09-2024	10:03 - 13:38	StCu	56%	1.5 km	Case Study
RF23 / as240045 / 10-09-2024	12:36 - 15:54	MCS	78%	14.5 km	Case Study
RF24 / as240046 / 10-09-2024	17:36 - 20:45	MCS	97%	13.0 km	Case Study

Table 1. Summary of the 24 SAFIRE ATR-42 research flights during the MAESTRO field campaign. Each flight sampled specific cloud types or mesoscale phenomena, with the three sampling strategy types adopted (described in subsection 2.2). Times of take off and landing (TO-L) are given in local time (UTC-1). **Cloud type:** ShCu: shallow cumulus; StCu: stratocumulus; AlCu: altocumulus; AlSt: altostratus; Cg: congestus; **Phenomena:** CP: cold pools; MCS: mesoscale convective system; CL: convergence line. Cloud Cover was obtained from the Météosat Seconde Génération (MSG) satellite in the smallest lat-lon box containing the flight track with the image closest to the middle time of the flight as diagnosed from the algorithm from Derrien and Le Gléau (2005)b in the framework of Eumetsat SAF Nowcasting. **Max CT**: Maximum Cloud Top, measured with the onboard vertical airborne radar RASTA (Delanoë et al., 2013). Some flights classified as "Case Study" or "Standard" also included coordination elements (see subsection 2.2 for details).

Table 2. In-situ instrumentation used for the MAESTRO turbulence dataset. The dataset is distributed at 25 Hz for all variables. The table distinguishes the electronic sampling rate from the native instrument response time. NC : Not communicated

Quantity	~~Flight list (RF)~~ Model	~~Number of Flights~~ Manufacturer	~~S~~ Principle	~~L~~ Signal Output Frequency	~~B~~Internal Response Time	Precision	Role in dataset
Attitude angles & Ground Speed	Inertial Navigation System (AIRINS model 6005214)	iXblue	Fiber optic gyroscope, GPS Antenna	100 Hz	<3 ms	<0.1 $m.s^{-1}$	Wind Computations
True Air Speed	Pitot (Static Pressure) and Radome central hole (Total Pressure)	Rosemount	differential pressure sensors	200 Hz	15 ms	0.25 hPa	Wind Computations
Attack/Sideslip angles	Vertically/Horizontally aligned pressure measurements	Rosemount	Differential pressure	33 Hz	15 ms	0.2 hPa	Wind Computations
Temperature	Rosemount E102AL (non-deiced)	Collins	Platinum resistance	200 Hz	>0.1 s	0.3 K	~~H~~**Retained** as fast absolute sensor
Temperature	~~T~~ E102AL (deiced)	~~M~~ Collins	~~Total~~ Heated Platinum resistance	200 Hz	>0.1 s	0.3 K	Backup (icing)
~~Typical MAESTRO~~ Temperature	~~4~~Fine wire, 5~~, 6, 7, 8, 11, 12, 13, 15, 17, 23~~ μm	~~11~~ SAFIRE	~~3~~ Fine-wire thermometer	~~78~~ 200 Hz	~~131~~ >10 ms	~~68~~ NC	~~0~~
Temperature	~~0~~ Fine wire, 10 μm	~~280 Case study~~ SAFIRE	~~14, 19, 20, 21, 22, 24~~Fine-wire thermometer	~~6~~ 200 Hz	~~0~~ >10 ms	~~67~~ NC	~~26~~
Water vapor	~~22~~ FAST-WAVE	~~6~~ GSMA, Univ. Reims	~~24~~ Wavelength modulation spectroscopy	~~145~~ 100 Hz	10 ms	NC	**Retained**: fast q sensor; calibrated against WVSS-2
~~Coordination~~ Water vapor	~~1,~~ Water Vapor Sensing System 2~~,~~ (WVSS-2)	FLYHT Aerospace Solutions	Tunable diode laser absorption	0.5 Hz	3~~, 9, 10, 16, 18~~ s	~~7~~ 5 %	~~6~~ **Retained**: slow reference for calibration
Water vapor	~~44~~ LI-7500	~~60~~ LI-COR Environmental	~~28~~ Open-path NIR absorption	~~0~~ 50 Hz	~~0~~ >50 ms	~~138~~ 1 %	Excluded: vibration-induced spectral artifact
~~Total~~ Dew point	1011C	~~24~~ Buck Research Instruments	~~9~~ Chilled-mirror hygrometer	~~189~~ 1 Hz	~~217~~ 1 $K\,s^{-1}$ (warming rate)	~~118~~ 1 K	~~6~~ Reference in principle; limited by thermal inertia (App. B)
Humidity (RH)	~~24~~ HUMAERO	~~563~~ Enviscope	Capacitive RH + T sensor	1 Hz	TBD	1.7 %	Excluded: systematic bias vs. other

Table 3. ~~List of~~ MAESTRO turbulence dataset structure and variables~~calculated over each~~. File nomenclature: file names begin with MAESTRO_ORCESTRA_ATR_TURBULENCE_* *YYYYMMDD* = flight date (e.g., 20240815), *NN* = campaign flight number (e.g., 09), *ZZ* = SAFIRE flight identifier, *X* = leg identifier, *i* = segment identifier (e.g., ~~their standard notation~~L2_1), *V* = file version. Unless stated otherwise, statistics were computed using high-pass filtered fluctuations and ~~NetCDF~~ using the wind-oriented frame of reference (U_t and V_t). Quantities obtained form detrended series are annotated with the _DET suffix. Random and systematic errors are available for each covariance and variance variable~~names~~.

~~Category~~Variable type	~~Notation(s)~~Varible name in the NetCDF file ~~NetCDF variable name(s)~~
~~General characteristics~~	SEGMENT_STATISTICS
File name :	*SEGMENT_STATISTICS_RF*NN*_as2400*ZZ_YYYYMMDD*_v*V*.nc
~~Segment name~~ Segment metadata and aircraft position, true heading, ground speed, and true air speed:	~~name~~name, date, (START_*, END_*, MEAN_*) time/lat/lon, MEAN_alt, MEAN_THDG, MEAN_GS, MEAN_TAS
~~Segment start/end t_i, t_f, t time_start, time_end Segment start/end/mean position lat_i, lon_i, lat_f, lon_f, lat, lon lat_start, lon_start, lat_end, lon_end, lat, lon Mean altitude z alt Mean heading $THDG$ MEAN_THDG Mean ground speed GS MEAN_GS Mean true airspeed TAS MEAN_TAS~~ Mean ~~static pressure~~ quantities:	~~P_S MEAN_PS~~ MEAN_PS, MEAN_U_geo, MEAN_V_geo, MEAN_WS, MEAN_WD, MEAN_TS, MEAN_W, MEAN_T, MEAN_MR
~~Mean geographic wind~~ Second- and Third-order moments (from high-pass filtered and detrended):	~~U_{geo}, V_{geo} UWE_mean, VNS_mean~~ VAR_U, VAR_V, VAR_W, VAR_T, VAR_MR, COVAR_UV, COVAR_UW, COVAR_VW, COVAR_UT, COVAR_UMR, COVAR_VT, COVAR_VMR, COVAR_WT, COVAR_WMR, COVAR_TMR, M3_U, M3_V, M3_W, M3_T, M3_MR
~~Heterogeneity score~~ Skewness :	~~$\mathcal{H}$ H_score~~ SKEW_U, SKEW_V, SKEW_W, SKEW_T, SKEW_MR
~~Mixing ratio correlation coefficient~~ Integral scales :	~~R^2_{MR} R2_MR~~ L_U, L_V, L_W, L_T, L_MR, L_UV, L_UW, L_VW, L_UT, L_UMR, L_VT, L_VMR, L_WT, L_WMR, L_TMR
~~Horizontal wind speed and direction~~ Other derived quantities :	~~V_H, DD WS_MEAN, MEAN_WD~~ TKE, TKE_DR, Slope_U, Slope_V, Slope_W, Slope_T, Slope_MR
~~**Turbulent quantities**~~ Quality metrics :	Heterogeneity_score, R2_MR, ERR_S*, ERR_R*
~~Second-order moments~~	SEGMENT_TIMESERIES