Politecnico di Torino

ScuDo
Scuola di Dottorato ~ Doctoral School
WHAT YOU ARE, TAKES YOU FAR

Doctoral Dissertation
Doctoral Program in Metrology (38thcycle)

Measurement quality in data products for meteorology and climate

By

Alberto Bottacin

******

Supervisor(s):
Dr. Chiara Musacchio, Supervisor
Dr. Graziano Coppa, Co-Supervisor

Doctoral Examination Committee:
Prof. Aleksandra Ewa Kowal, Referee, Institute of Low Temperature and Structure Research of the Polish Academy of Sciences
Prof. Fabio Madonna, Referee, Università degli Studi di Salerno
Prof. Sabrina Grassini, President, Politecnico di Torino

Politecnico di Torino
2025

Declaration

I hereby declare that, the contents and organization of this dissertation constitute my own original work and does not compromise in any way the rights of third parties, including those relating to the security of personal data.

Alberto Bottacin
2025

* This dissertation is presented in partial fulfillment of the requirements for **Ph.D. degree** in the Graduate School of Politecnico di Torino (ScuDo).

To Gaia, the reference of my life

Acknowledgements

Every PhD is a journey that can feel stressful and never-ending without good traveling companions. At INRiM, I found mentors, colleagues, and friends to whom I am infinitely grateful for the support received over these three years. I would particularly like to thank all the members of the Applied Thermodynamics sector (AE05), who contribute to create a vibrant and peaceful environment in which to work.

Thank you, Andrea, for seeing potential in me and welcoming me into the great family of metrologists dedicated to improving weather and climate monitoring. Thank you, Chiara, for the infinite patience you have shown: I often showed up at your door for advice without notice, and you never turned me away. Thank you, Graziano, for being my point of reference, especially for data analysis: your help and experience allowed me to learn years' worth of knowledge better and faster. Thank you, Enrico, for the immeasurable technical support you gave me, both in the laboratory and in the field: part of this work would not have been possible without your help.

A special thanks goes to my young colleagues. Thank you, Natalì, for listening to me when I felt stuck and needed advice. Thank you, Laura, for showing immediate interest in my work: discussing my experiments with you helped me improve several aspects of Chapter 4.

My sincere thanks go to my chemist colleagues (Michela, Fra, and Fr) who helped me and shared much of their expertise regarding gas analyzers. Without you, I would not have been able to carry out the work in Chapter 5.

An immense thank you goes to Tutor for being, on many occasions, an additional supervisor for my thesis. I am infinitely grateful for the time you dedicated to me and for the excellent advice you always managed to give.

Thanks to all the colleagues at “Casa Masse”, especially Stefano, who made this journey feel lighter. A proper thank you goes to Marina, who hosted me in her office for a year during the laboratory renovations.

Finally, thank you, Nicola: you introduced me to the measurement of turbulent fluxes, guiding me every step of the way. Without you, Chapter 5 would not exist.

I undertook this journey in the middle of uncertainty, spending a year at home before actually starting the PhD. I thank my parents for supporting me and for never doubting the choices I made.

This journey would never have happened without the person by my side. If she had not encouraged me to apply for this doctorate (completely different from my undergraduate degree), I would regret it now. I would have missed out on a world that I love living in and to which I wish to contribute as a researcher. Thank you, Gaia, for believing in me before anyone else and for helping me make my dream come true.

Abstract

Climate change is one of the defining challenges of the 21st century. The ability to detect, attribute, and project changes in the Earth's climate system depends critically on the availability of long-term, high-quality observational records of essential climate variables. Many climate datasets are derived from networks originally designed for operational weather forecasting rather than for the precise detection of long-term trends. Consequently, they often lack the traceability, redundancy, and systematic uncertainty evaluation that are hallmarks of metrology.

The application of metrological principles to atmospheric measurements is not an easy task: a complete evaluation of their uncertainty budget represents a challenge, especially due to the complex environmental thermodynamic interactions involved at the Earth's surface. This is particularly true for air temperature, which is the key quantity in evaluating global warming compared to preindustrial times. Past EURAMET-funded projects, MeteoMet and MeteoMet 2, addressed some key aspects of air temperature measurements by performing experiments aimed at improving our knowledge of the corresponding uncertainty budget. Despite the effort, fundamental difficulties are still met in evaluating the many thermodynamic contributions to the total uncertainty, from the laboratory calibration to the environmental effects in the field.

In addition to air temperature, other quantities used to study the state of the climate need metrological standardization. Turbulent fluxes have long been studied to understand the interaction between the atmosphere and the ground, but also to determine the health of ecological systems (e.g. forests, croplands, grasslands) and their response to climate change. The eddy covariance technique has emerged in the last four decades as the most popular method to measure turbulent fluxes, mainly thanks to the technological progress that has permitted its implementation. Despite the establishment of continental-scale networks and infrastructure, implementing the

eddy covariance methodology following standardized procedures and instruments, a rigorous metrological assessment of the technique is still missing, as proven by the absence of a stated measurement uncertainty along with the flux estimates.

Recent works highlighted that another variable, not related to atmospheric physics, started to be affected by the air temperature anomaly: the air/rock temperature in caves. Correctly measuring this variable is crucial for monitoring the underground climate, known to be stable and little influenced by the external conditions. The main metrological challenge concerns the harsh conditions in which a measurement system must operate, requiring periodic verification to ensure the SI traceability chain is not broken by sensor drifts or failures of the acquisition device.

This dissertation seeks to demonstrate how a rigorous application of metrology can enhance the quality and robustness of weather and climate observations, but also to propose new methodologies for quantifying currently unknown uncertainty components. The research specifically focused on improving the quality of three key variables: air temperature, CO_2 turbulent flux, and temperature in subterranean environments (caves), through a series of dedicated laboratory and on-field experiments. The main objectives of this project were:

- To define climate reference data and prototype a Climatological Reference Station.
- To improve the measurement uncertainty budget of air temperature, focusing on three uncertainty sources: datalogger contribution, influence of rain and different exposure due to installation's height;
- To develop a comprehensive framework for evaluating the measurement uncertainty of CO_2 turbulent fluxes by propagating the uncertainties associated with the models' input quantities;
- To demonstrate the feasibility of on-site thermometer verification in caves and promote a cost-effective, easily transportable experimental setup for this purpose.

Collectively, this research demonstrates that the application of metrology is not confined to laboratory settings but can be systematically applied across the full spectrum of climate monitoring to transform observational data into defensible,

SI-traceable datasets, which are essential for informing high-stakes decision-making in global climate mitigation and adaptation strategies.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background and Motivation

Climate change is one of the defining challenges of the 21st century. The ability to detect, attribute, and project changes in the Earth's climate system depends critically on the availability of long-term, high-quality observational records of Essential Climate Variables (ECVs). These records are the foundation for both fundamental research and policy-relevant assessments, such as those of the Intergovernmental Panel on Climate Change (IPCC). Yet the production of such records is not trivial. Many climate datasets are derived from networks originally designed for operational weather forecasting rather than for the precise detection of long-term trends. Consequently, they often lack the traceability, redundancy, and systematic uncertainty evaluation that are hallmarks of metrology. The importance that metrology has for the World Meteorological Organization (WMO) is established by its signing of the metre convention in 2010 and, later, by the creation of expert teams on measurement uncertainty (ET-MU) and on quality, traceability and calibration (ET-QTC). Furthermore, the recent WMO Decision INFCOM-2/Doc. 7.4(2) (Feb. 2023) on "Uncertainty Assessments" requests to further promote, organize and coordinate field experiments and studies, necessary to refine and improve the uncertainty evaluation and traceability of measurements in different observing networks.

The application of metrological principles to the atmospheric measurements is not an easy task: metrologists usually work with perfectly controlled experimental conditions, whereas atmospheric variables are intrinsically non-stationary. A

complete evaluation of the uncertainty budget for these measurements represents a challenge for metrology, especially due to the complex environmental thermodynamic interactions involved at the Earth's surface. This is particularly true for air temperature, which is the key quantity in evaluating global warming compared to preindustrial time [1]. Despite such importance, fundamental difficulties are still met in evaluating the many thermodynamic contributions to the total uncertainty, from the laboratory calibration to the environmental effects in the field. To bridge this gap, the Consultative Committee for Thermometry (CCT) included in its Strategy Planning 2021 – 2030 actions for the establishment of a practical definition of air temperature and for the development of methods on how to evaluate the uncertainty in air temperature measurements.

This dissertation is situated at the interface between these two domains, seeking to demonstrate how a rigorous application of metrology can enhance the quality and robustness of weather and climate observations, but also to propose new methodologies for quantifying currently unknown uncertainty components. In particular, laboratory and on-field experiments, along with data analyses, were conducted to improve the uncertainty budget of some quantities of climatological interest, such as air temperature, CO_2 turbulent flux and cave temperature. Aside from the straightforward relevance of air temperature, turbulent fluxes are addressed because they represent the main approach for understanding the ecosystem responses to climate change [2]. Recent works highlighted instead that the temperature in some caves started to be affected by the atmospheric temperature anomaly [3], demonstrating the importance of monitoring this type of environment that was less influenced by the atmospheric conditions, up to now. The enhancement of the measurement quality of these quantities allows the creation of datasets that are not only internally consistent but also defensible when used for high-stakes decision-making in mitigation and adaptation.

The specific needs and objectives of this thesis are reported in the next section, along with a brief description of the activities reported in each Chapter.

1.2 Needs and Objectives

The need for reference-quality observations has been recognized by the Global Climate Observing System (GCOS), which distinguishes between baseline and

reference networks. Reference networks such as the GCOS Reference Upper Air Network (GRUAN) aim to provide traceable, uncertainty-quantified measurements that anchor the broader observing system. However, for surface-based climate variables, such a reference framework is only beginning to emerge.

Chapter 2 of this thesis contributes to this effort by defining and prototyping a Climatological Reference Station (CRS). By embedding redundancy, traceability, and uncertainty evaluation in its design, the CRS represents a practical step toward operationalizing the concept of climate reference data. This development directly addresses the limitations of current surface networks, where inhomogeneities caused by instrument changes, relocations, and siting differences often necessitate complex homogenization procedures. A preventive approach, rooted in metrology, offers a more robust path forward.

Chapters 3 and 4 contribute to a better understanding of the uncertainty components affecting air temperature measurements. Two broad categories of sources of uncertainty can be distinguished. The first relates to the measurement system itself, which is investigated in Chapter 3. Historically, much attention has focused on radiative errors, leading to the design of shields and ventilated screens. However, other system-related influences are equally relevant. In particular, experiments are performed to analyze the systematic deviations introduced by the thermal sensitivity of datalogger acquisition electronics, which are often considered as black-boxes, trusting the not-so-transparent manufacturer specifications. Another error never investigated before concerns the effects of rainfall on shielded thermometers, which can perturb measurements through wetting and evaporative cooling. The deviations are estimated by reproducing a controlled precipitation and comparing the measurements with a reference system. Such effects, though often transient or small in magnitude, may accumulate into systematic biases in long-term records if not identified and corrected.

The second category of uncertainty sources arises from instrument siting and installation practices. The WMO recommends that air temperature sensors shall be installed between 1.25 and 2.00 m above ground level, a convention likely rooted in historical practice rather than empirical optimization. Yet within this accepted range, systematic differences in records can occur, as stratification and near-surface processes determine the vertical temperature profile. An intercomparison of thermometers at different heights is conducted, using modern artificially ventilated

screens, to quantify such deviations with metrological rigor and understand their dependence on environmental factors, as it is described in Chapter 4.

By considering both system-related and siting-related uncertainties, the thesis highlights that the reliability of air temperature records cannot be ensured by calibration alone. Instead, a comprehensive metrological approach is required—one that integrates laboratory characterization, field intercomparisons, and correction models grounded in physical understanding. Together, the analyses in Chapters 3 and 4 exemplify how subtle but systematic influences can be quantified and mitigated, thereby improving the traceability and comparability of one of the most important climate variables.

While much of climate monitoring focuses on state variables such as temperature and humidity, fluxes of energy, water, and carbon are equally critical for understanding the dynamics of the Earth system and the response of ecosystems to climate change. The 2022 GCOS Implementation Plan [4] states that "Understanding and estimating surface fluxes is essential for improving projections of climate change and planning adaptation and response measures". The Eddy Covariance (EC) method, based on measurements of a set of quantities, is the most widely used technique for measuring surface fluxes, underpinning international networks such as ICOS and AmeriFlux. Aware of the importance of having homogeneous measurement procedures, ICOS produces instructions with practical and technical information, in terms of the instrument setup, measurements and its management for the flux measurements by the ICOS members. However, these guides include very limited information about the optimal traceability chain, calibration procedures and uncertainty evaluations. Yet uncertainty quantification in EC remains uneven and often incomplete, almost nullifying the very same effort of measuring: measurements without a statement of uncertainty are incomplete and meaningless. That is why Chapter 5 aims to extend the metrological approach to EC by constructing complete uncertainty budgets and deriving effective sensitivity coefficients. The methodology is specifically applied for the estimation of CO_2 fluxes, but can be adapted for other fluxes, such as those of sensible or latent heat, thanks to the nature of the EC technique. A case study illustrates how uncertainties in flux measurements propagate into estimates of annual carbon budgets. This work demonstrates the applicability of the Guide to the Expression of Uncertainty in Measurement (GUM) [5] to flux observations, thereby strengthening the methodological rigor of network-level reporting.

Not all climate monitoring takes place in standard meteorological settings. Subterranean environments such as caves offer valuable natural laboratories for studying long-term climate variability, but they also present unique challenges for measurement verification, inevitable for measurement systems under constant saturation of water vapor in air. Chapter 6 presents the Bossea cave case study, where a dense thermometer network is used to monitor underground climate. By applying in situ verification protocols, systematic offsets due to datalogger faults were identified and corrected, ensuring comparability across the network. This work illustrates how metrological principles can be adapted to unconventional contexts, extending the reach of climate reference measurements beyond traditional meteorological stations.

Through the three thematic studies (air temperature, gas fluxes and cave temperature), the thesis demonstrates that measurement science is not an ancillary concern but a core component of climatology. Ensuring that climate records are traceable, comparable and uncertainty-quantified is a prerequisite for robust detection of climate change and for the credibility of the evidence base that informs global action. Therefore, the introduction of metrology into climate monitoring is both timely and necessary. As climate science increasingly underpins policy and societal decisions, the credibility of observational data cannot be taken for granted. By systematically addressing uncertainties, this dissertation contributes to building a foundation of reference-quality observations, aligning with the mandates of WMO, GCOS, and ICOS.

1.3 Structure of the Thesis

The dissertation is primarily divided into three thematic studies, as outlined in the objectives. The first is about air temperature (Chapters 2-4), the second concerns eddy covariance measurements (Chapter 5), whereas the third is related to cave temperature (Chapter 6). Each technical chapter, belonging to one of the themes, starts with an extensive introduction specific to the described content, followed by the methodology, results and/or discussion sections. The last part of each chapter is a summary of the main achievements which briefly reports the main results and conclusions. Finally, in Chapter 7 the overall conclusions are drawn, focusing on the impact of the research.

Chapter 2

Climatological Reference Station

Ground-based stations play a vital role in generating data essential for assessing local and global climate trends and variations. These networks comprise numerous observing sites, each equipped with diverse instruments and managed and maintained differently. Currently, the WMO Technical Regulations stipulate that every Member shall establish and maintain at least one Climatological Reference Station (CRS) (https://community.wmo.int/). These stations are defined as climatological observation points that collect data and associated metadata specifically to identify climate trends. Such stations are expected to provide long periods (typically no less than 30 years) of homogeneous records, ideally in environments where human-induced changes have been, and are anticipated to remain, minimal. Ideally, these records should be extensive enough to enable the identification of long-term climate changes, as outlined in the Manual on WIGOS [6]. However, it is important to note that a universally agreed-upon definition for the core instrumental and technical specifications of a CRS, along with standardized reference measurement procedures, is currently lacking. This absence leads to a variety of approaches among different nations, National Meteorological and Hydrological Services (NMHSs), and research institutes, thereby diminishing the comparability of results across geographical areas and time. Furthermore, the scarcity of CRSs significantly contributes to the substantial effort required for data harmonization and bias detection at local, regional, and global levels, as their absence necessitates reliance on other stations more prone to such biases.

It is neither expected nor realistic for every station in an observing network to meet the rigorous CRS standard. Nevertheless, the monitoring of ECVs is anticipated to increase in both temporal and spatial density. New stations and networks are likely to be established specifically to measure atmospheric and surface parameters for detecting climate change, including those related to extreme events. Within these networks, a limited number of stations are required to serve as references, supporting a tiered observational approach: these are the stations designated as CRSs.

The tiered-network strategy is founded on the understanding that achieving reference-quality measurements everywhere is neither necessary, economically viable, nor technically practical. Instead, these high-quality measurements are needed in sufficient locations to build confidence in the remaining observations. Beyond the recent initiative by the Global Climate Observing System (GCOS) to define features and promote the establishment of a GCOS Surface Reference Network (GSRN) [7], valuable initiatives like the United States Climate Reference Network (USCRN) [8] now employ a common design and consistent measurement principles for climatological reference stations. Other climatological reference networks exist globally, but comparing them remains challenging due to inconsistencies in instrumentation and site locations, which often fail to provide the stable environmental conditions necessary to avoid undetectable biases. Conversely, despite significant efforts in infrastructural networking, a common approach for designing climatological reference stations is missing and not yet planned in many regions, including Europe.

The purpose of this chapter is to outline the characteristics required for climatological reference data quality, highlighting the challenges climatologists are facing using historical time series. The definitions, specifications and technical characteristics of climate reference data and stations for ground-based networks presented in the following were developed within the project EMPIR 19SIP03 - CRS, whose outcomes are published in [9]. To support the deployment of such definitions and requirements, a prototype of a CRS has been developed and installed by INRiM, whose technical specifications are reported here alongside with first considerations from field tests.

2.1 Challenges with historical data series

Homogeneous long-term observational datasets are extremely rare on a global scale, making homogenization a critical component of climate data processing. These efforts are indispensable for correcting both abrupt and gradual change-points, as well as other potential biases, which may be artificially introduced into the data. Such biases often result from alterations in station location or local site environment, changes in instrument shelter type or instrumentation, or evolving observing and reporting methodologies [10]. Determining the precise timing of these global-scale changes is challenging, as they are infrequently documented alongside the station measurements.

Homogenization algorithms [11, 12] are developed to discern these change-point signals from background noise and subsequently apply approximate adjustments. The majority of currently available algorithms utilize comparisons with neighboring stations for both detection and adjustment. While substantial changes are relatively straightforward to identify, the detection of minor changes poses a greater challenge, despite their fundamental importance for understanding climate evolution. Furthermore, certain changes may be isolated to a single station, whereas others could be network-wide (e.g., a synchronized change in instrument shelter or observing system across an entire network), which complicates their detection [13, 14]. At daily and sub-daily temporal scales, homogenization practice becomes a particularly formidable challenge.

Such intensive mathematical and computational efforts are predominantly necessitated by the scarcity of data series originating from stations specifically designed to generate data immune from undocumented changes, continuously traceable to recognized standards, and accompanied by quantified measurement uncertainties.

Numerous long-term and centennial stations are situated within urban areas, where historical meteorological observatories were originally established. As cities expand and industrial activities intensify, coupled with increased vehicular traffic and the heating and cooling demands of buildings, these urban effects introduce significant impacts on the data quality of such stations. This often results in diminished representativeness of the surrounding climate, increased uncertainties, and reduced stability. Despite a few exceptions [15, 14], these urban influences are frequently

unstable over time, often exhibiting seasonal and daily cycles, and/or weather-type dependencies in the magnitude of the biases.

Relocating centennial stations could mitigate the introduction of biases attributable to the urban environment; however, the potential disruption to the continuity of the time series might outweigh the benefits of moving to a more uncontaminated site. Among the extant long-term stations, only a limited number are located in open fields or other stable environments, characterized by high quality of the surrounding site. These stations hold paramount relevance within climate networks, even when the site is affected by minor obstacles. Another issue, even for these representative historical stations, is the frequent lack of precise instrument model information, including details related to instrument type, exposure/orientation, and instrument shelter. Maintenance and calibration are also crucial. Well-documented maintenance procedures are essential for discerning potential breaks in time-series homogeneity and their associated biases. However, the systematic application of calibration practices was historically uncommon.

2.2 Key differences between weather and climate data

Many ground-based stations now incorporated into climate-relevant networks initially served as meteorological stations for weather monitoring and forecasting. Consequently, in most countries, the majority of climate-relevant observing stations continue to be managed by National Meteorological and Hydrological Services (NMHSs). This persists even in nations where a distinct organization is responsible for climate data management and analysis, such as in the US (where the National Weather Service manages the network, while the National Centers for Environmental Information handle data management) and New Zealand (where MetService and NIWA manage networks, with NIWA also curating climate data).

The fundamental goal of climate-relevant observations is to characterize the long-term attributes of the climate system, whereas weather observations primarily focus on describing the current atmospheric state and its short-term variations. This divergence in purpose necessitates distinct requirements for observing weather versus climate. To accurately evaluate and comprehend the climate, particularly climate variability and change, it is crucial to obtain homogeneous, continuous, and high-quality observations for specific climate elements. Such data are fundamen-

tal for deriving satisfactory climatological references. Furthermore, each climate observation must be accompanied by a comprehensive set of metadata, providing users with essential information regarding the recording process (how, where, when, by whom) and proper interpretation and usage. The GCOS climate monitoring principles [10], endorsed by the former Commission for Climatology (CCI), were developed specifically to minimize the impacts of inhomogeneities. The temporal resolution of data also represents a key difference between weather and climatological data. In the contemporary era, weather forecasting applications have limited utility for data with only daily time resolution. However, data comprising just one or two observations per day, provided they are consistent and reliable, can still be highly valuable for climatological purposes.

Considering these aforementioned requirements, not all weather-related data are suitable for climatological applications for the following reasons:

- The length of the record is (or remains) too short, though in certain circumstances, a well-equipped station can nonetheless contribute to validating other stations within a tiered approach or be incorporated into gridded analyses.
- The station is equipped with low-quality instruments or is situated at a site characterized by the presence of nearby obstacles that compromise the representativeness and/or stability of the measurement results.
- Metadata are inadequately documented.
- Insufficient maintenance leads to the introduction of artificial biases into the data series (e.g., due to unrecognized sensor drift not periodically verified or calibrated, or instrument changes implemented without specific comparison procedures between old and new systems).
- Data are biased by temporal changes in observing procedures and schedules.
- Observations are influenced by changes in the local environment and station surroundings, thereby exhibiting biases that hinder their reliable support for climate assessments, products, or services.
- Methodological aspects pose challenges (e.g., current synoptic observations typically require measurements of maximum temperatures only between 6 a.m. and 6 p.m. and minimum temperatures only between 6 p.m. and 6 a.m., thus

omitting 12 hours of extreme observations within a 24-hour period. Such a practice is insufficient for comprehensive climate and extreme event analyses).

The inherent differences between climate-relevant and weather observations underscore the necessity, as explicitly stated by the WMO, GCOS, and the broader climate community, for observing stations and networks to be specifically designed to generate data suitable for evaluating and understanding the climate. Nevertheless, high-level meteorological observations can still contribute data suitable for the same purpose.

2.3 Definitions and requirements

An objective of the project was to discuss and propose a clear definition of climate reference data and station with the aim of its adoption into relevant vocabularies and regulatory documents. The proposed definition is reported in Table 2.1 and is structured into two main components: a concise and unambiguous definition, followed by a series of explanatory notes designed to clarify key aspects.

The definition of climate reference data relies on a clear definition of the "measurand", which is the intended quantity to be measured/estimated, a key concept in metrology. Sometimes the measurand is ill-defined, leading to underrated issue and difficulties in implementing measurement procedures, including instrumental capabilities. In climatology, normally the measurand is not the single measured record, but the variability in time of the interested quantity and estimated from the measurements. Such a relative definition of the measurand helps reduce the impact of the uncertainty components on every single record, transferring the overall uncertainty in the measurand (the trend) to statistical methods, under the condition that the measurand is stable and representative of the climate signal under study.

On the other hand, the definition of climate reference station is just "operative", which means it is identified as any calibrated instrument which is able to produce climatological reference data for a chosen ECV. Compliance with standards promoted in WMO Technical Regulations and guidelines, including the Guide to Instruments and Methods of Observations [16], is preconditioned.

Table 2.1 Definitions of climate reference data and station.

Climate reference data	A series of traceable measurement results able to quantify the variability and change of climate-relevant variables. *Notes:* *i.* The result of the measurement is a single record of the observed parameter, while the overall measurand is the variability and change of the variable. *ii.* To be traceable, a measurement result requires that each instrument involved in the measurement process is related to a reference standard of the System of Units (SI) or other standards through a documented unbroken chain of calibrations. *iii.* The absolute requirement of a measurement is that it be made in such a way that after accounting for all sources of uncertainty it can be concluded that the true value of the measurand lies within the reported uncertainty interval with specified confidence. The result of a reference-grade measurement is such that it can be used to improve the quality of other (lower-tier) measurements.
Climate reference station	A climatological reference station is an instrumental installation able to generate climatological reference data. *Notes:* *i.* Measured data must be continuous and representative of the local environment. *ii.* The station must be stable in its location and siting characteristics for decades and equipped with top quality instrumentation. *iii.* The instrumentation needs to be well maintained with regular maintenance and calibration to constantly keep documented data traceability. *iv.* Changes in instrumentation must be limited, motivated and documented. Parallel observation periods must be planned prior to any instrumental change.

A further aspect in the definition of the measurand, measuring techniques can also introduce problems in interpreting the measurement results. A key example is measuring air temperature, the key variable observed in climate studies for compar-

ing extreme values and anomalies. While it might seem obvious that a thermometer measures the temperature of the air, this is not physically what happens. Indeed, a contact thermometer gives an indication of its heat equilibrium at that time, under those specific conditions of heat exchange. This implies that the temperature value is obtained under a fundamental non equilibrium between the radiative, convective, contact heat transfers and condensation, icing and evaporation phenomena. This is a fundamental thermodynamic principle that turns into large contributions to the uncertainties and even corrections in air temperature measurements. Adopted techniques evolved in time, to reduce errors and the magnitude of the quantities of influence, solar radiation at first. Hence, the result of the measurement is a compromise: a temperature value as representative as possible of the air temperature at that time in those conditions. When reference observations are needed in climatology to compare, harmonize other values, or to represent trends of an area, then a definition of what a "reference" temperature value is required. This requires a definition of the measurement conditions to be considered as reference. For example: the air temperature value is taken at zero radiation or extrapolated at zero total radiation; under steady conditions, or at air speed minimum of, or even at extrapolated zero wind speed. In the end, it's the definition itself of the measurand, the atmospheric air temperature, that is even still missing [17].

2.4 Technical features of CRSs

A climate reference station needs to be equipped with high-quality instrumentation, procedures and technology that will generate the best achievable estimates of values of the observed quantities. The station shall be located at a site that is representative of its regional climate and adheres to strict siting requirements. Locations shall be chosen where no significant changes in the surrounding areas are expected for the next century. Site and equipment maintenance need to be a priority to avoid degradation and uncontrolled instrumental drifts over time due to environmental conditions that may affect the measurements. Instruments shall be calibrated at regular intervals, to correct for drift and maintain traceability. Changes in instrumentations, procedures and technology shall be limited, motivated and documented. Parallel observation periods shall be planned prior to any change, and the time series of both new and old stations retained indefinitely. Uncertainty on measurement results shall be evaluated

including calibration uncertainty, instruments characteristics and field environmental influences.

2.4.1 Reference-grade measurements and quantities of influence

A climate reference station must be equipped with a set of instruments and procedures capable of reliably and consistently measuring surface-based atmospheric ECVs. These include: air temperature, water vapour (typically measured via relative humidity), surface radiation budget, wind speed and direction, air pressure, and both liquid and solid precipitation. Historically, climatology focused primarily on analyzing near-surface air temperature and precipitation trends derived from meteorological observations. Consequently, air temperature and precipitation are considered mandatory variables for CRS installations, except in climatic zones where precipitation levels are extremely low.

Over time, the scope of climate monitoring has expanded to include additional variables that contribute to understanding climate evolution. The following measurements are strongly recommended at CRS sites: relative humidity, solar radiation, wind speed and direction (at a height of 10 m), air pressure, and soil temperature.

Further optional variables are also encouraged where relevant, particularly for their relation to terrestrial ECVs or their potential influence on instrument performance. These include: snow cover, land surface temperature, soil temperature and soil moisture, surface albedo, river discharge, and groundwater (when present nearby). In permafrost regions, measurements such as the permafrost temperature profile, active layer thickness, and other cryospheric indicators are also recommended. Additionally, air quality observations may be included in the reference-grade measurements conducted at a CRS.

CRS stations are also required to measure Associated Quantities of Influence (AQI), as outlined in Table 2.2. These are supporting measurements needed to ensure reference-quality data for other variables. For example, producing a reference air temperature measurement requires supporting observations of solar radiation, relative humidity, precipitation, and wind. While AQIs are essential, they do not need to meet full reference-grade quality standards; this means they may involve lower calibration and maintenance requirements, and uncertainty budgets are not strictly defined. However, a Quality Check (QC) process must be applied continuously to

AQI instruments. This QC should follow the minimum standards for field verification, as defined by the WMO [16].

If an AQI also happens to be one of the station's reference variables, the same measurements can be used for both purposes. In the air temperature example, precipitation is required at reference quality, but the other AQIs (solar radiation, relative humidity, and wind) do not necessarily need to be measured at reference level.

Table 2.2 Schematic table for measured parameters and associated quantities of influence (From GCOS-226 [7]). Primary measurements are air temperature (T_{air}), relative humidity (RH), solar radiation (SR), wind speed and direction (WS/WD), precipitation accumulated and intensity (PA/PI), soil temperature (T_{soil}) and soil moisture (SM). Reference measurements require auxiliary measurements to detect the influencing factor with respect to an associated quantity of influence.

Primary Measurements	**Associated Quantities of Influence**								
	T_{air}	RH	R	WS/WD	P_{air}	PA/PI	T_{rain}	T_{soil}	SM
T_{air}		X	X	X		X		X	
RH	X		X	X	X	X			X
SR	X	X				X			
WS/WD	X					X			
P_{air}	X			X					
PA/PI	X			X					
T_{soil}						X			
SM						X			

2.4.2 Traceability to establish comparability in time and space

Data comparability in time and space is the main deliverable of an installation providing measurement results for climatology. The fundamental prerequisite is that recorded data are the result of measurements made by means of traceable instruments. Establishing documented traceability to a measurement process turns the observations into a robust set of values, comparable in time and space and among other different (but traceable) measurements. When climate reference stations are required to validate other systems, such as remote sensing, then the traceability is

essential, to guarantee that the response of a system under validation is correctly checked. This also applies when reference stations are used to check other stations in a network. Instrument traceability shall be established at installation, through calibration prior to deployment, and preserved in time, through procedures involving periodic tests, checks and re-calibration. Regular maintenance and calibration are key aspects of metrological best practice for reference networks and stations. Frequencies of on-site maintenance for a CRS are reported in Table 2.3.

As a general rule:

- Regular field inspection shall be made every 6 months and/or at need, following for example extreme events or evidence of malfunctioning. The inspection can lead to repair / substitution of instruments. In the case of manual observers, refresher training should be provided, especially if there is evidence of systematic errors in the data.
- Field verifications against travelling equipment shall be performed every year, to check instruments' correct working conditions. The verification requires a threshold limit for a pass/fail evaluation. Verification failures shall be followed by an immediate recalibration.
- Calibration should be repeated every 24 months. Longer time intervals shall only be considered if warranted by the instruments' quality, their exposure, the environmental conditions of the site, their ageing and the prescriptions from the manufacturers.

Standardized calibration procedures shall be adopted by all network stations to document instrumental traceability. To avoid interruption in the data series during the calibration process, replacement of sensors shall be adopted: a) in a circular way (calibrated sensors are used at every recalibration to replace sensors to be calibrated) or b) by temporarily substituting the sensor under calibration with another calibrated one. Field calibration can also be performed in exceptional cases, when difficulties are met in removing the sensors (for example in the case of very old historical systems or in special conditions). Field calibration shall be organized in order to cover the whole range of variability of the sensor, and the likely range of climate conditions, and to establish traceability as for the laboratory calibration, although the uncertainties can be larger.

2.4.3 Redundancy

Measurement redundancy, i.e. the use of multiple measuring sensors for the same variable, is recommended at reference stations. Initially this may be only for some of the observed quantities such as air temperature but can be extended to other instruments. Redundancy represents one way to assess aspects of both traceability and comparability. By using multiple, co-located traceable instruments to measure the same parameter, both the single values and the resultant data series can be compared. Disagreement between the data series can highlight measurement problems or sensor drift, which would be undetectable with a single sensor. Data management practices within the NMHS must enable storing of each sensor's data. This includes cases where the co-located sensors are of different types, such as a manual rain gauge and a tipping bucket rain gauge.

Agreement results in a lower contribution of the statistical uncertainty components in the overall measurement uncertainty. Redundancy also reduces the amount of missing data in the event of an outage which affects a single sensor, which is particularly valuable for additive elements such as precipitation where an outage prevents monthly or annual totals from being calculated.

As each reference station will ensure that all measurements are traceable to standards and have a quantified uncertainty budget, having identical equipment at each site is not needed. At the same time, networks made of identical stations can offer the opportunity to define unique calibration and maintenance procedures, same recording and transmission principles and data loggers, and thus a more uniform data stream and a better opportunity to replace instruments having similar spare items for the whole network. At a national level identical equipment should be encouraged , and where required compromise should be adopted in order to reach an optimal set of identical and specific installations. An example of this would be in the transition from manual to automated measurements, or where new technologies become available for improved measurements.

2.4.4 Managing instruments change

Regardless of whether by choice or necessity, the challenge of an instrument change is to manage all transitions in such a manner that the effects upon long-term series

continuity are minimized and the associated uncertainty well understood. A fundamental metrological principle stipulates that replacing one instrument with another at the same location should pose no problem when the old and the new instrument are both fully traceable to standards. In reality this idealized concept is not fully met since different instruments or sensors may react differently to the same external environmental factors. Under the conditions that both instruments are calibrated, the change does not affect the accuracy of data, but only the uncertainties associated with the quantities of influence (for example a different influence of radiation on a changed temperature sensor). For compensating such effects, parallel observations of sufficient length and encompassing a suitable range of different climate types are required, meeting the whole (or the majority of the) range of variability of each quantity. Such prescription is easily applicable to any station or network but is at present rarely documented in information or metadata.

Table 2.3 Typologies and frequencies of on-site maintenance for a CRS.

Activity	**Time interval**	**Provides**	**Results in**
Inspection	6 months/At occurrence	Repair/Substitution	System OK
Check/Verification	12 months	Pass/Fail	Tolerance limit
Calibration	24 months	Establish traceability	Uncertainty

2.4.5 Uncertainty evaluation

"When reporting the result of a measurement of a physical quantity, it is obligatory that some quantitative indication of the quality of the result be given so that those who use it can assess its reliability. Without such an indication, measurement results cannot be compared, either among themselves or with reference values given in a specification or standard. It is therefore necessary that there be a readily implemented, easily understood, and generally accepted procedure for characterizing the quality of a result of a measurement, that is, for evaluating and expressing its uncertainty." [5]

In accordance with the International Vocabulary of Metrology [18] and the Guide to the Expression of Uncertainty in Measurement [5], the evaluation of measurement uncertainty must incorporate all quantifiable contributions. Uncertainty is defined as the range within which the true value of the measurand is plausibly expected to lie.

It should be reported as twice the standard deviation, providing a 95 % confidence interval. The GUM distinguishes between two fundamental methods for uncertainty evaluation: Type A, based on statistical analysis of observations, and Type B, derived from instrument specifications and other information sources.

For meteorological applications in particular, four primary sources of measurement uncertainty can be identified:

1. *Instrumental components*: These include sensor-specific properties such as response time, drift, and sensitivity, as well as calibration accuracy and interactions with auxiliary components like radiation shields or data loggers. These uncertainties are typically assessed in controlled laboratory conditions or documented by manufacturers.

2. *Environmental effects on instrument response*: These encompass factors like precipitation (e.g., sensor overcooling), wind, solar and terrestrial radiation, condensation, icing, and exposure to extreme climates (e.g., polar or desert environments). Such influences must be evaluated under site-specific field conditions or through intercomparison campaigns.

3. *Measurand definition*: This refers to how well the quantity being measured is defined and how representative it is of the parameter of interest. This includes siting classification and related metadata.

4. *Statistical processing*: Uncertainties can also arise from data processing methods—for instance, when calculating mean values over defined intervals or detecting specific events (e.g., maximum values or anomalies).

A summary of key uncertainty sources contributing to each category is provided in Appendix A.1. Each source contributes to the overall uncertainty budget, calculated using Gaussian error propagation. It is important to recognize that calibration uncertainty represents just one part of this total, the simplest to be determined, which is in many cases the smallest contributor. Environmental factors are more difficult to study, as will be described in Chapter 3, leading likely to the largest uncertainty components.

Controlling Environmental Influences and Calibration Conditions

Instrument manufacturers are constantly working to develop designs aimed at reducing sensitivity to environmental conditions, thereby minimizing their impact on sensor response. Likewise, testing and calibration laboratories are expected to maintain controlled environmental conditions (e.g., temperature and humidity) to ensure that the influence of such factors remains negligible during calibration. Instruments are thus calibrated in stable settings, enhancing the reliability of their link to reference standards.

Among the many sources of uncertainty encountered in field measurements, environmental influences on sensor behavior represent some of the largest contributions. Variability in field conditions, including precipitation, wind, solar and terrestrial radiation, condensation, and extreme temperatures, can affect sensor performance in complex and often non-linear ways. It is virtually impossible to make field observations fully immune to these effects. As a result, environmental influences must be explicitly included in the measurement process and accounted for in the uncertainty evaluation.

Two main scenarios can occur:

a) If the effect can be numerically evaluated, then a correction can be applied to the readings, and a contribution to the uncertainty shall be included, in terms of uncertainty of the correction.

b) If the effect of an influencing factor cannot be corrected, this directly becomes a component of uncertainty which in principle is larger than the uncertainty of the correction.

Even for identical types of instrumentation, these effects can vary significantly depending on local environmental conditions. Therefore, site-specific experiments are necessary to evaluate the true impact of environmental variability on measurement accuracy. Although standardized prescriptions are not yet established, it is recommended that qualified personnel or manufacturers develop tailored methods to assess environmental influences. In most cases such experiments require a couple of identical systems, one exposed to the effect under test and the other protected at best. For example, the overcooling effect of the rain on thermometers reading can be evaluated by deploying two identical instruments (thermometer and shield) in

the same site under the rain and protecting one of the two. A detailed protocol to perform such test is published in [19] about the effect of the reflected radiation from a snow-covered surface on the accuracy of thermometers.

2.5 CRS Protoype

With the purpose of studying, comparing and confirming the prescribed definitions and requirements, a prototype for a CRS has been developed and installed in May 2023 by INRiM. As extensively described, the station shall be located at a site that is representative of its regional climate and adhering to strict siting requirements, according to the GSRN specifications [7]. Locations should be chosen where no significant changes in the surrounding areas are expected, but also should be as uniform as possible, without obstacles in the vicinity of the instrumentation that could induce biases in the measurements. Other characteristics have also been considered, such as security, ease of access for verification and maintenance routine. An example of data product, for the GCOS defined ECV "Air temperature near surface" is also provided, aligned to the GSRN requirements.

WMO identifies some of the sources of error due to the site features and proposes a classification scheme [16]. Each site will need to be large enough to house all instrumentation without adjacent instrumentation interfering with one another, with no shading or wind-blocking vegetation or localized topography, and at least 100 m from any artificial heat sources. The 100 m distance is based on a precautionary evaluation to avoid the effect of obstacles, although recent studies [20] demonstrated that at 50 m from the measuring point most of the influences are reduced to negligible effects, within the instrumental uncertainties. This allows including among the climate reference stations of a network also those installations placed in sites with the presence of minor issues also at less than 100 m (for example paths or road that are not asphalt or concrete, isolated trees, gentle slope etc.). Based on the considerations above, a site has been selected, taking into account specific requirements such as: the vicinity to INRiM (ease of access to the site, reduce carbon impact and person time) and the possibility to sign a formal agreement for the use of the land in a public area to avoid possible changes in the use of the terrain with new buildings or similar.

The identified site, shown in Figures 2.1 and 2.2, is located in the "Park of Stupinigi", a public area 3 km from INRiM, in the municipality of Nichelino. The

area is flat, covered by natural grass, and no obstacles are present within 100 m radius, apart from a small dirt path. Thanks to its characteristics and according to the WMO siting classification, the site is in class 1. In the Köppen classification, the climate zone corresponding to the site is Cfa (continental, no dry season, hot summer).

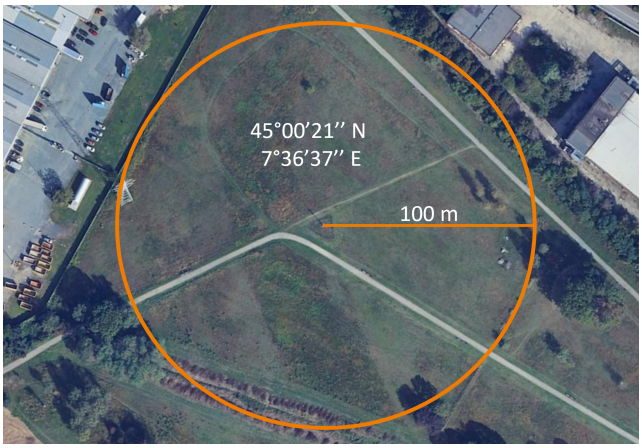


Fig. 2.1 Google Earth picture of the site, with the 100 m radius free of obstacles around the station, positioned in the centre of the red circle. The coordinates of the station are reported in the picture. It is worth mentioning that all trees have been cut in the area internal to the red circle, according to the WMO prescriptions, in order to reduce uncertainties in the data representativeness and measurement errors.

Following the requirements in terms of data quality for primary measurements (at reference level) and need to measure the so-called AQI, the configuration of the INRiM Climate Reference Station, shown in Figure 2.2. The station is based on the instruments, reported in Table 2.4, and an associated datalogger with transmitter. The required electric power is generated by solar panels and a battery, in order to test a configuration that can be installed anywhere, including remote locations (polar sites, high mountains, deserts). Measurements are recorded continuously with no missing data and all instruments proved their quality and stability.

In the second half of 2023, the INRiM Climate Reference Station was activated and data started to be recorded and transmitted. The datalogger records values from each sensor at 10 s intervals and transmits data buffers at 1 h frequency. The logger reads and transmits raw data only: for example, temperature values are recorded as resistance values, since the station uses Pt100 thermometers. The same applies for the other quantities. At a post processing level, raw data is transformed in processed values, reported in the associated units. For example, temperature is translated into

(a)

(b)

(c)

Fig. 2.2 (a) Picture of the CRS, (b) Detail of the redundant system for measuring air temperature composed by an aspirated screen (on the left) and two naturally ventilated screens (in the middle and on the right), (c) Picture of the anemometer and pyranometer used for measuring the AQI at the thermometers level.

degrees Celsius by applying the resistance-to-temperature calibration curve's coefficients, while radiation is changed from millivolts to watts applying the pyranometers' sensitivity. All data is also associated to the instrument uncertainty, given by the calibration uncertainty and all other parameters characterizing the response of each instrument. For each observed quantity, all other associated quantities of influence are recorded and presented at each time.

2.6 Summary of main achievements

Ground-based stations are essential for assessing local and global climate trends and variations. Climatologists use historical time series coming from weather stations designed for validation of weather forecasting and not for climate monitoring, hence

Table 2.4 Calibration uncertainty of instruments installed on the CRS prototype.

Variable	**Instrument**	**Calibration Uncertainty**
Temperature (reference)	PRT 100 (4-wire connection)	0.012 °C
Temperature - relative humidity (AQI)	Vaisala HMP155	0.05 °C 3% RH
Precipitation	SIAP TP200	Total: 2%–10% Intensity: 2%
Air pressure	Paroscientific DIGIQUARTZ	10 Pa
Wind speed & direction (reference)	GILL HS50 3-axis ultrasonic anemometer	Speed: < 1 m/s Direction: < 1°
Wind speed & direction (AQI)	Gill Windsonic	Speed: 2% @ 12 m/s Direction: 2° @ 12 m/s
Solar radiation (direct and reflected)	2 Hukseflux LP02 Second class pyranometers	< 1.8%

a huge effort is required for data homogenization, which becomes particularly difficult at the sub-daily scale and for instruments not regularly maintained and calibrated. Although the relevance of climate monitoring, a standardized definition of climate reference data and station, along with standardized reference measurement procedures, is currently lacking. The work presented in this chapter addresses this problem by proposing practical definitions, along with a series of technical notes, reported in Table 2.1:

- **Climate reference data**: a series of traceable measurement results able to quantify the variability and change of climate-relevant variables.
- **Climate reference station**: a climatological reference station is an instrumental installation able to generate climatological reference data.

A key point of the first definition is that the measurand is not just the quantity measured by an instrument, but is its variability over time. On the other hand, although the definition of CRS is operative, the notes specify some technical requirements such as assured continuity of measurements, representative site identification and

regular intervals for calibration, verification and maintenance (as reported in Table 2.3).

Additional details on technical features of CRSs are given in Section 2.4. For generating climate reference data, accurate measurements of ECVs require two fundamental characteristics: (a) to be traceable, thus recorded by calibrated and constantly maintained instruments and (b) to be accompanied by the values of all identified associated quantities of influence (the other ECVs influencing the response of the instrument or the measurand itself). The data product must then be a complete set of values, including the uncertainties related to both the primary measurement and the effect of the AQIs on its value and uncertainty. For example, the effect of rain as overcooling bias of thermometers and associated recovery time to reach again accurate temperature records, the uncertainty in the correction due to reflected radiation in case of snow under the temperature sensors, the wind induced reduction of biases on different solar shields etc. Nevertheless, the complete evaluation of the uncertainties and the corrections (and associated components of uncertainty), due to the effects of the AQIs, at present is almost impossible for most of the main ECVs of interest in climatology. Reference data products should in any case include, together with the records of the main ECVs, all the values of the AQIs (see Table 2.2), recorded at the same time.

To study and confirm the prescribed definitions and requirements of this work, a prototype for a CRS was developed and installed in May 2023 by INRiM (Figure 2.2) aligned with the WMO top class requirements. The installed instruments and processing of data are briefly described in Section 2.5. The creation of an “ad hoc” climate reference station, installed in a well-maintained free of obstables site, allows an active role in managing the implementation and data production. Furthermore, because INRiM became the WMO *Measurement Lead Centre on Traceability and Field Metrology*, the CRS acts also as a special site for research and development on instrument field validation and intercomparison, testing of new and evolving systems, evaluation of drifts and maintenance intervals and pros and cons in specific and challenging uses.

Chapter 3

Measurement system uncertainty of air temperature measurements

Modern contact thermometers are being continuously improved, together with the calibration procedures, to reduce the uncertainty in their use, now easily limited to the order of the millikelvin in the range of interest. Calibration procedures are well defined and cover a multitude of different sensors and methods, from the primary standards to the working instruments.

Despite such technological advances, measuring atmospheric air temperature still requires deep understanding and research [21]. Indeed, an agreed procedure for the calibration of thermometers in air is missing. The evaluation of complete uncertainty budgets for practical field measurements, taking into account the rapidly changing conditions and the numerous quantities of influence, is also a challenging task. The measurand itself, *air temperature* is also not yet fully defined [17] as already mentioned in the previous chapter. Although it is clear that the temperature of air corresponds to the temperature of the gas called “air", what is missing is a definition of air, which is a mixture of gases with non-constant concentrations, with a special concern about water vapour and CO_2. Evaporation and condensation of water are the main factors affecting the concentration of water vapour in air, whereas CO_2 concentration is increasing due to anthropogenic emissions. The variability of air composition corresponds to different thermophysical properties, leading in the end to a different interaction between the contact thermometer and the air components. Therefore, a clear definition of air is needed to establish standardized procedures

for calibration of thermometers in air, although it is not yet clear how it should be defined: dry air? With what amount of water content? Water content measured within which target uncertainty? Air stream should be null or stationary?

While a precise definition of air temperature is unclear, physicists have long investigated the potential errors a thermometer can make when measuring the temperature of a gas [22], specifically air. The radiative error affecting resistance thermometers exposed to radiation is likely the most important one, which led to the invention and introduction of wood Stevenson screens at the end of the nineteenth century for meteorological measurements. Since then, many studies have been conducted for optimizing the material of the screen [23] and its geometry for enhancing air ventilation around the thermometer, ensuring the best thermal contact between air and the thermometer sheath. For understanding the physics behind the radiative error, analytical models have been developed [24–27], which showed key aspects such as the dependence of the radiative error on the square root of sensor diameter. However, these models are based on assumptions, such as stationarity, and often do not take into account the coupling with air due to the radiation shield, as done for instance with a semi-empirical model in [28] for estimating the temperature error inside naturally ventilated shelters. To overcome this limitation and better understand ventilation inside screens, Computational Fluid Dynamics (CFD) simulations coupled with optical models, like ray tracing, have been performed over specific shelters/shields [29–32]: the aim is to estimate the radiative error, depending on wind speeds, input solar irradiance, sun elevation or other parameters. These simulations are crucial for a reliable estimation of the radiative error, however their application is still limited since in most cases only the measurement system is modelled, without considering the environment close the measurement site which can introduce spurious heat sources (for instance radiative emission from buildings). Another approach to tackle the radiative error is the development of new measurement sensors/systems which are able to directly estimate the air temperature with no correction needed. Noteworthy examples are the radiation compensating thermocouple (RCTC) [33, 34] and the energy-balance thermocouple (EBTC) [35, 36]. Both examples can be described as “ensemble” sensors constituted by three thermocouples having different characteristics. The RCTC is composed by thermocouples with different diameters that are electrically connected in a such a way that the respective radiative errors compensate. On the other hand the EBTC is constituted by three thermocouples inside brass globes having coatings with different emissivities: solving the energy

balance equations it is possible to estimate at the same time the air temperature, the short-wave and long-wave radiative loads. The lower sensitivity to radiation of these sensors was validated using resistance thermometers inside aspirated shields, which, however, integrate the air temperature below the screen, a characteristic that is not fully coherent with a *true* air thermometer. The need of having a good reference induced the development of non-contact acoustic thermometers for meteorological applications [37], which are truly insensitive to radiation [38], but rely on the knowledge of the air thermophysical properties to relate the speed of sound measurements to the temperature of the medium (the air). In this case the model uncertainty is still too large, leading to measurement uncertainty greater than 0.1 °C, which requires further improvements.

Although of crucial importance, solar radiation is only one of the factors affecting the uncertainty of air temperature measurements. The Measurement Quality Classification Scheme shown in Figure 3.1 is an instrument developed by the WMO [16] for visualizing the different groups of uncertainties affecting the measurements, in particular:

- uncertainties associated to calibration and the measurement system itself;
- uncertainties due to coupling of the sensor with the medium. In the case of air thermometers, these uncertainties are introduced by the use of solar screens;
- uncertainties resulting from maintenance and verification of the measurement system;
- uncertainties due to the associated quantities of influence;
- uncertainties which result from instrument siting and exposure, not dependent on the employed measurement system.

Documented traceability of measurement results requires full understanding and evaluation of all these uncertainty components, meaning being able to write down a complete uncertainty budget for air temperature measurements. A number of experiments have been already addressed for solving specific aspects of atmospheric air temperature measurements, quantities of influence and uncertainty components. The MeteoMet projects [39, 40] led by INRiM contributed to the investigation of some of these effects, mainly in terms of single components in the overall measurement uncertainty budget. The projects generated a positive impact, encouraging

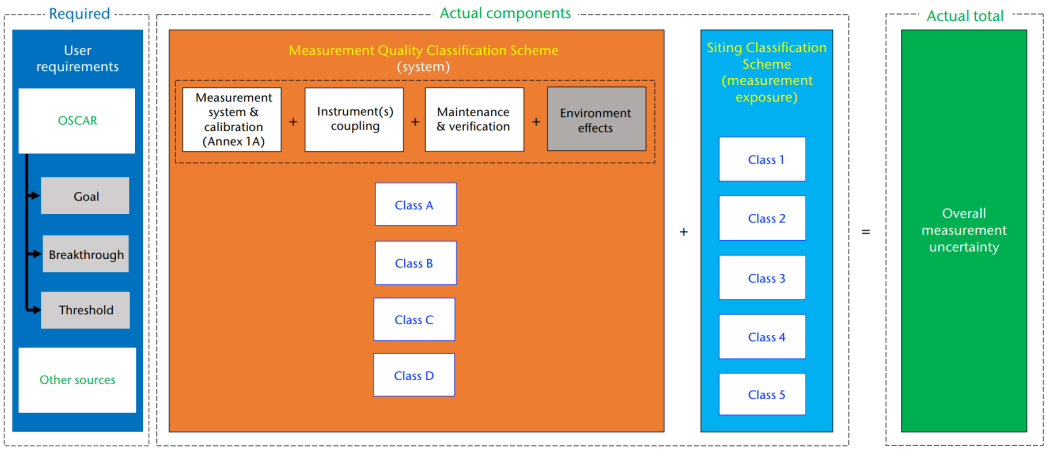


Fig. 3.1 Measurement Quality Classification Scheme [16].

collaboratiosn between metrologists and meteorologists. Now several aspects that can introduce uncertainties have been (partially) addressed, such as: on site calibration of instruments [41–44]; ageing of solar shields [45]; time response of sensors [46–48]; time drift of resistance thermometers [49]; self-heating errors both due to sensor probe and accompanying electronics [50, 51]; albedo effect due to snow cover [19, 52] and siting effects due to the presence of a road or a building [20, 53].

While this remarkable effort, many other uncertainty components have to be estimated. Within the activities of the PhD project, two further sources of system uncertainties have been analysed: the acquisition electronics of meteorological dataloggers and the effect of rain on air temperature measurement. Indeed, even if meteorological dataloggers are widely spread, they are rarely analysed as they are: Data AcQuisition boards (DAQ) developed for making, in most of the case, resistance measurements, corresponding to temperatures after applying the calibration curve of thermometers. For this reason, the results of the characterization of a meteorological datalogger are presented in Section 3.1. Similarly, the effect of rain on temperature measurements performed using different types of solar shields has never been a field of research. Within MeteoMet 2 [40], an experimental setup was developed to investigate this effect. The measurements of the conducted experiments have been analysed as part of the PhD project and published in [54]. A description of the experimental activities, the analysis of measurements and results are reported in Section 3.2.

3.1 Datalogger characterization

When performing temperature measurements, two main elements are required:

- a temperature sensor (thermometer), characterized by a certain quantity, such as electrical resistance, that varies with temperature;
- an acquisition device, which is designed to measure the temperature-sensitive quantity characteristic of the thermometer.

In calibration labs, reference thermometers are usually thermo-resistances (for instance Pt25, Pt100, thermistors) which are measured by bridges (AC or DC) or lab-grade temperature scanners. The acquisition devices must be calibrated, for instance against standard reference resistors, to ensure SI traceability. Then the resulting calibration uncertainty is added to the uncertainty budget of thermometers that are under calibration using the same bridge/scanner.

Similar reasoning applies to air temperature measurements for meteorological and climatological applications. The majority of studies cited in the previous section focus on the sensor itself, which is generally read during experiments with lab-grade equipment, like DC bridges. On the other hand, when comparisons between different thermometers are involved, as in the experiment for evaluating the albedo effect [19], it is not so crucial to have a high-quality instrument. Indeed, if the thermometers are read with the same instrument, uncertainty components due to the acquisition device cancel out when working with relative differences.

The dataloggers for meteorological applications are multi-purpose data acquisition boards, designed with Analog-to-Digital converters (ADC) and voltage/current references to interface with analog sensors. At the same time, they are also able to communicate with digital sensors that have their own acquisition unit, whose measurements are transmitted to the datalogger by an integrated processing unit through a chosen communication protocol.

Despite their relevance for guaranteeing accurate air temperature measurements, the datalogger contribution to the overall uncertainty budget is rarely evaluated. An explanation may be that this product is primarly used by people in national meteorological and hydrological services who do not have the knowledge and competencies to assess its performance, being more prone to trust the manufacturer's specifications.

Furthermore, there are examples in literature of newly designed microcontroller-based dataloggers, whose performances are tested against commercial models [55] and not against standards traceable to the SI, proving that there is a lack of metrological knowledge and/or resources for performing such tests. Fortunately, there are also some virtuous examples: for instance in [56] Whiteman, Hubbe and Shaw try to evaluate the performance of an inexpensive datalogger by comparing the corresponding temperature measurements against a reference PRT in a stirred liquid bath. On the other hand, Wu et al. developed an "active resistance network" for generating reference values of resistance used to test the datalogger of an automatic weather station [57]. However, what is missing in both cases is a complete evaluation of the uncertainties involved in the performance assessment.

Even though the scientific literature lacks works analyzing in detail datalogger performances, the main part of the acquisition is played by the ADC, which has been extensively studied over the years [58–60]. A IEEE standard describes how to determine the ADC static transfer function and test its dynamic behavior [61]. Although it is crucial to understand the uncertainties resulting from the ADC voltage measurements [62], they still remain a portion of the overall uncertainty budget for resistance measurements. In fact, dataloggers measure resistances mainly through the volt-ammeter method, thus by generating a reference current and measuring the voltage across the temperature sensor. The main issue for the characterization of this technique is that little information is available about the internal circuitry, complicating the characterization of the single components involved in the measurements (source current and measured voltage).

To characterize the datalogger, it is more convenient to consider the instrument as an ohmmeter, i.e. an indicator of resistance values, without paying too much attention to its internal circuits. Under this assumption, techniques used for calibrating resistance bridges could be potentially used, in particular those based on resistance networks composed of standard resistors [63]. Although this approach allows the smallest calibration/verification uncertainty, it requires a number of resistors that are not always available, even in a calibration laboratory.

A methodology for the characterization of a meteorological datalogger, identical to the one used in the CRS prototype described in Chapter 2, is here presented. The tests have been performed implementing a procedure similar to the one described in [64], based on a stirred liquid bath, to verify the datalogger's performance when

exposed to different environmental conditions. First a description of the methodology and experimental setup is given, then the results are discussed, mainly focusing on the uncertainty evaluation of the errors performed by the datalogger.

3.1.1 Methodology and experimental setup

The datalogger under test is a SIAP DA18K, as that installed at the CRS prototype. This model has a 24-bit ADC on board (no manufacturer or model specified). The specifications for voltage and temperature measurements are reported in Table 3.1, as written in the corresponding datasheet, with no information about resistance measurements or the value of the source current, which is described in the manual [65] as "impulsive" (to limit the thermoresistance self-heating) without further information about stability or electrical noise.

Table 3.1 Specification of the datalogger DA18K as reported from the datasheet.

Quantity	**Resolution**	**Accuracy**	**Uncertainty** (3σ)
Voltage (differential)	$0.3\,\mu V$	$\pm(10\,\mu V + 0.1\,\%\,\text{measure})$	$3\,\mu V$
Temperature	$0.0003\,^{\circ}C$	$0.02\,^{\circ}C$	$0.005\,^{\circ}C$

The ADC specifications seem compatible to what is currently available on the market and allow, in principle, a resolution of at least $0.001\,^{\circ}C$ as required for a CRS. On the other hand, it is not clear what "accuracy" means and how it is evaluated. It could probably refer to the difference with respect to a reference instrument owned by the manufacturer, whereas uncertainty may be related to the electrical noise associated to the sampled signals and not to the overall measurement uncertainty. A similar argument can be adapted to temperature specifications, even though it is quite erroneous to directly report them with almost no information about the probe used and the calibration performed. The only available note is that the temperature specifications are computed considering a Pt100 thermo-resistance.

Based on the information about resolution and assuming a nominal sensitivity for a Pt100 of $0.4\,\Omega/^{\circ}C$, it is possible to estimate the source current as:

$$I \approx 0.3\,\mu V/(0.4\,\Omega/^{\circ}C \cdot 0.0003\,^{\circ}C) = 2.5\,mA,$$

which is reasonable for resistance measurements, especially considering its impulsive nature to limit the self-heating. The information about the current is crucial for temperature measurements since thermometers can be calibrated in different ways. Sometimes accredited laboratories specify the extrapolated resistance at a null current value [50], therefore for zero self-heating. Instead others labs calculate the calibration curve with resistances measured using a declared source current, as done at the INRiM laboratory. Not only it is necessary to know the current, but also how it is supplied to the thermo-resistance: the same value of current supplied in impulsive mode (no current flowing from one measurement to the other) results in a smaller self-heating with respect to continuously maintaining the resistance under voltage. However, the self-heating for the on-off case is more difficult to evaluate, in particular when, as for the datalogger under test, no info is available about how long the current is supplied before taking the voltage measurements across the thermo-resistance.

Fig. 3.2 Picture of the experimental setup where is visible the datalogger inside the climatic chamber, the circulating bath (bottom-left corner) and the resistance bridge (on the table).

Despite the missing information, the best method to verify the resistance measurements performed by the datalogger is to use reference resistance values, generally obtained through resistance networks [63] composed by standard resistors with calibration uncertainty as low as about 5 ppm. The lack of this high-quality resistors

in our laboratory led to the development of a different methodology to test the datalogger.

Inspired by Gersak et al. [64], the reference resistance values are obtained using reference Pt100 thermometers inside a circulating bath, which allows a stabilization in the mK range, that is suitable for testing the declared temperature uncertainty. The datalogger has been tested by comparing two thermometers inside the bath, one read by a resistance bridge and the other by the datalogger. Furthermore, the datalogger was put in a climatic chamber for varying its operating conditions and verify if its performance is temperature-dependent.

The instruments used in the tests, visible in Figure 3.2, are listed here below:

- three reference Pt100, serial numbers 1, 2 and 3, with calibration certificates released by the INRiM temperature calibration laboratory, stating an expanded calibration uncertainty ($k = 2$) of 0.01 °C above 0 °C and 0.02 °C below 0 °C using a source current, constantly supplied, of 1 mA.

- a DC resistance bridge, the Fluke SuperThermometer (ST) 1595A, capable of measuring resistances with a declared 1-year 95 % confidence interval of 5 ppm or 0.000024 Ω for resistances between 0 and 120 Ω, using a 1 mA current. The bridge is used to read the reference thermometers under the same conditions they were calibrated, thus supplying continuously 1 mA of current.

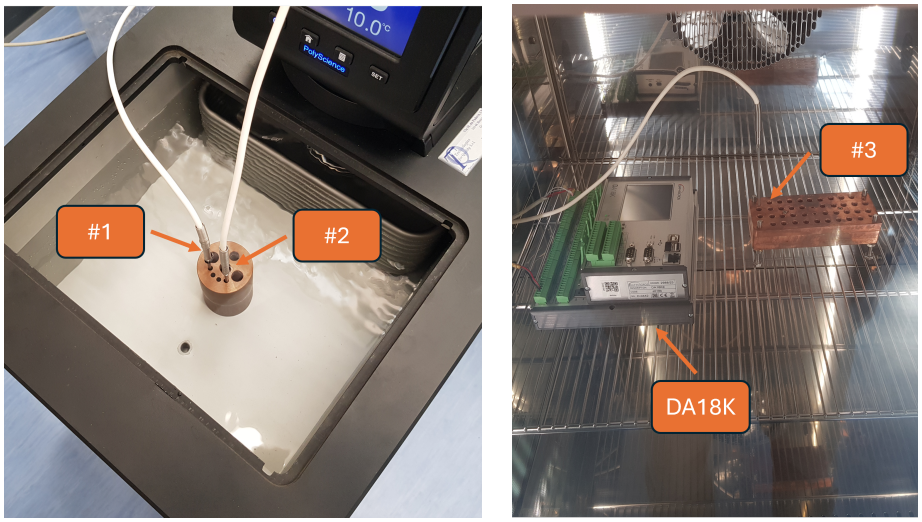
- a PolyScience bath used with pure ethanol for temperatures below 0 °C and deionized water in the range above 0 °C;

- a copper comparator block used inside the circulating bath to enhance temperature stability and reduce temperature gradients when comparing thermometers;

- a climatic chamber from Kambic, model KK-190 CHLT, having temperature stability in the order of 0.1 °C. However, in this case the specifications are not so relevant since the environmental temperature is a quantity of influence. Then, the chamber is used for maintaining the datalogger at a certain temperature, surely within ±1 °C.

The tested nominal selected bath temperatures are $[-20, -10, 10, 20, 30, 40, 50]$ °C, whereas $[-10, 20, 40]$ °C are the nominal values of the selected temperature of the

chamber. The relative humidity is set to 50 %rh and is not varied during the experiment. Each combination of the nominal bath and chamber temperatures were tested, even those that are improbable in a real case, like $T_{\text{bath}} = -20\,^{\circ}\text{C}$ and $T_{\text{chamber}} = 40\,^{\circ}\text{C}$, since the scope was to test the specifications for any input resistance/temperature under the operating conditions stated by the manufacturer ($-40 \div 80\,^{\circ}\text{C}$).



(a) (b)

Fig. 3.3 (a) Picture of the reference thermometers 1 and 2 inside the comparator block in the circulating bath. (b) Picture of the datalogger inside the climatic chamber. The reference thermometer 3 inserted in another comparator block is used to check the stability condition of the chamber.

The test procedure, for each combination of bath and chamber temperature, is divided in two phases:

1. **Reference comparison**. Reference thermometers 1 and 2 are inserted in the bath inside the copper comparator block as shown in Figure 3.3a. The sensors are measured with the resistance bridge, sampling one sensors at a time every 4 s. After the temperature has been stabilized within $\pm 0.01\,^{\circ}\text{C}$, measurements are acquired for 30 minutes. Then, the best stabilization interval of at least 5 minutes is selected for computing the average temperatures of the sensors, identified by $\overline{T}_{1\text{-ST}}$ and $\overline{T}_{2\text{-ST}}$, whose subscripts recall the sensor number and how it is measured. Since the thermometers are identical and measured with the same procedure employed during calibration, the difference

between them is an estimate of the non-uniformity of the temperature field due to the considered medium (copper comparator inside a bath). Therefore, the difference $\Delta T_{\text{REF}} = \overline{T}_{\text{2-ST}} - \overline{T}_{\text{1-ST}}$ is evaluated and will be used in the second phase to remove the average inhomogeneity and make a fair comparison between the datalogger and the reference system.

2. **Datalogger comparison**. After the first phase, the reference 2 is connected to the datalogger, which is inside the chamber at the set chosen temperature, while both references are kept in the same position within the comparator inside the bath. To guarantee that the datalogger case reaches a thermal equilibrium with the environment, the reference 3 is put inside the chamber in a copper block, as shown in Figure 3.3b. When the readings of the reference stabilize within $\pm 1\,^{\circ}\text{C}$, it is assumed that also the temperature of the datalogger case is reasonably stable for the test purpose. The datalogger has been configured to make one resistance measurement, converted in temperature in post-processing, every 10 s. Despite the different sampling frequency, the same procedure was adopted for the identification of stable intervals, which in this case can have durations between 3 and 8 minutes. Nevertheless, the different number of measurements in the intervals is taken into account for the uncertainty evaluation by the stability and repeatability terms, as described in the following subsection. After the stable interval definition, the temperature difference is calculated: $\Delta T = \overline{T}_{\text{2-DAQ}} - \overline{T}_{\text{1-ST'}}$, where the apostrophe for reference 1 is used to not confuse it with the temperature estimated during the first phase. As mentioned, to remove the effect of the medium inhomogeneity and make a better comparison, the ΔT_{REF} estimated in the first phase is subtracted to obtain the final quantity of interest: $\Delta T_{\text{DAQ}} = \Delta T - \Delta T_{\text{REF}}$.

Each quantity aforementioned is subjected to a measurement uncertainty, which is evaluated as described in the following section.

3.1.2 Results and Discussion

The two phases were executed one after the other, in both cases waiting for stabilization. First, the temperature differences estimated from the comparison of reference thermometers and the associated uncertainty analysis are presented. Then, the re-

sults concerning the comparison of the datalogger against the reference system are reported.

Figure 3.4 shows, as an example, the 30-min time series of reference thermometers, measured by the resistance bridge, for a bath temperature of 30 °C. As is clear from the plot, the curves are smooth, not affected by electrical noise, and slowly varying, following temperature variations of the liquid. The difference between the thermometers is due to the residual temperature gradient in the liquid and, consequently, in the comparator. After having found the best stability intervals, the temperature differences were calculated as described in Subsection 3.1.1, which are reported in Table 3.2. The estimates of temperature inhomogeneity of the comparator

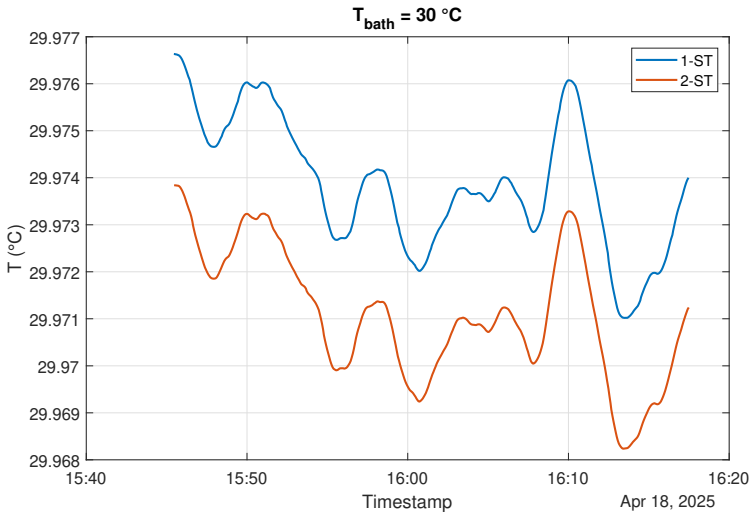


Fig. 3.4 Example of a 30-min time series of the reference thermometers for a $T_{\text{bath}} = 30\,^{\circ}\text{C}$ during the first phase of the test.

are, in magnitude, between 0.7 mK and 2.8 mK. These values are comparable to prior knowledge available on the use of a copper comparator inside the Polyscience bath. No clear trend is visible with the variation of the bath temperature, with positive differences below 0 °C, likely due to the use of a different liquid (ethanol instead of deionized water), and negative differences above 0 °C. In the table, close to the estimates, the standard measurement uncertainty is reported in round brackets, evaluated as follows:

$$u(\Delta T_{\text{REF}}) = \sqrt{u^2_{\text{1-stab}} + u^2_{\text{2-stab}} + u^2_{\text{1-cal}} + u^2_{\text{2-cal}} - 2\,\text{cov}_{\text{B}}(T_{\text{1-ST}}, T_{\text{2-ST}})}, \qquad (3.1)$$

Table 3.2 Estimated differences, and corresponding uncertainties, between reference thermometers compared at different set temperatures of the circulating bath.

$T_{\text{bath}}/^{\circ}\text{C}$	$\Delta T_{\text{REF}}/\text{mK}$
-20	1.8(0.2)
-10	0.7(0.4)
10	-1.6(0.2)
20	-2.4(0.3)
30	-2.8(0.3)
40	-2.8(0.2)
50	- 2.5(0.04)

where u_{stab} is the uncertainty contribution due to the instability of the signal, evaluated by the standard deviation of a uniform probability distribution, whereas u_{cal} is the uncertainty accounting for calibration. Since the two thermometers were calibrated at the same time using the same experimental apparatus, the type-B covariance between them can be estimated by their calibration uncertainty, thus compensating in Equation 3.2 the uncertainty for the calibration of the two sensors. This can be done since the covariance has a negative sign, which is the reason why it is recommended to work in "relative terms" using temperature differences. The uncertainty contributions for the resistance measurements also compensate for a similar reason. Despite the uncertainty for $T_{\text{bath}} = -10\,^{\circ}\text{C}$, which is about 50 % of the corresponding $u(\Delta T_{\text{REF}})$ (the smallest in absolute value), all the others are smaller than 15 %, which is more than reasonable for the intended use.

Moving to the results of the second phase, Figure 3.5 shows an example of a 30-min time series of the reference thermometers for a $T_{\text{bath}} = -40\,^{\circ}\text{C}$ and $T_{\text{chamber}} = -10\,^{\circ}\text{C}$, acquired during this part of the test. In this case, the reference 2 measured by the datalogger (solid orange line) shows a clear error and noisier behavior characterized by different harmonics, likely explainable by electrical noise or instability of the current source between consecutive measurements. To better investigate the associated noise, the histograms of $T_{\text{2-DAQ}}$ were plotted: Figure 3.6 is an example for $T_{\text{bath}} = 40\,^{\circ}\text{C}$ and $T_{\text{chamber}} = -10\,^{\circ}\text{C}$, while Figure 3.7 reports all the histograms for each combination of bath and chamber temperatures. It is quite evident, especially for the reported example, that the noise may be modeled by a bimodal distribution. Such noise shall be considered as an additional source of un-

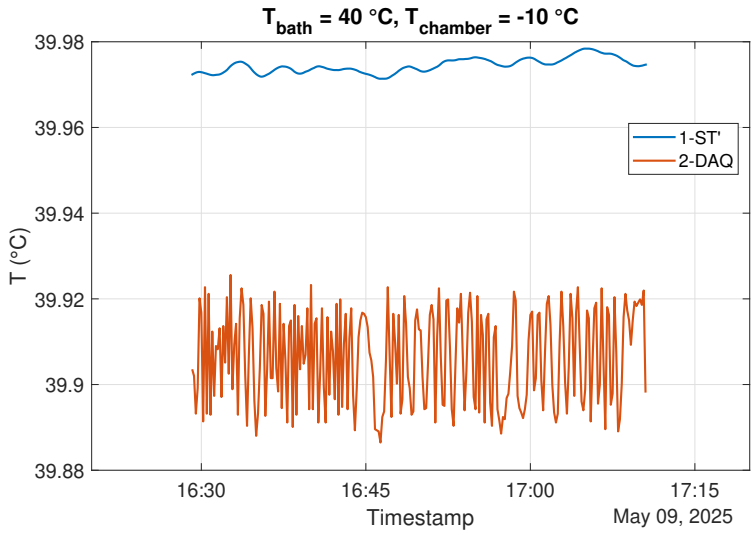


Fig. 3.5 Example of a 30-min time series of the reference thermometers for a $T_{\text{bath}} = 40\,^{\circ}\text{C}$ and $T_{\text{chamber}} = -10\,^{\circ}\text{C}$, during the second phase phase of the test. Reference 1 (solid blue line) was measured by the resistance bridge, whereas reference 2 (solid orange line) was measured by the datalogger.

certainty when evaluating the uncertainty budget for each temperature measurement performed by the datalogger, which essentially randomly samples from the noise distribution. Although a higher sampling frequency is mandatory for a better noise characterization, it is insightful to compute the empirical standard deviation of the measurements to have a coarse estimate of the dispersion across the mean value, which is highlighted by the dashed vertical red line in Figure 3.7. The estimated dispersions are reported in Table 3.3 and are comparable to the calibration uncertainty of the thermometers, i.e. in the range of 10 mK, hence it is not negligible. The estimated standard deviations are not dependent on the operating conditions, whereas they show a small trend with the bath temperature. Even in the case of better sampling, this type of estimate could be useful only for the evaluation of the standard uncertainty and not of confidence intervals. The best way to compute them is to model the distribution with some probability density functions (for instance the sum of two Gaussians) and use them for the determination of quantiles corresponding to the chosen coverage probability. Such a procedure requires a better noise characterization, which was not possible with the available instrumentation.

The calculated mean values in Figure 3.7 consider the entire 30-min interval. Following the procedure as the first phase, the temperature difference ΔT is evaluated by taking the best stabilization interval. Then the estimate is corrected for the thermal

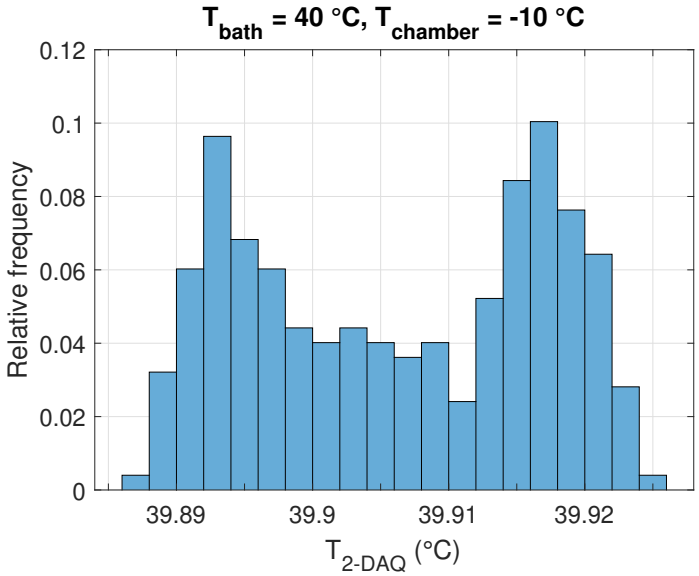


Fig. 3.6 Example of histogram of the temperature deviations between reference 1 and 2 during the datalogger comparison for $T_{\text{bath}} = 40\,^{\circ}\text{C}$ and $T_{\text{chamber}} = -10\,^{\circ}\text{C}$. The histogram is scaled such that the value for each bin corresponds to the associated relative frequency. The dashed vertical line refers to the mean value of the distribution, whose value is shown in the attached label.

gradient by subtracting ΔT_{REF}, obtaining in the end ΔT_{DAQ}. The measured values are reported in Table 3.4, with the corresponding uncertainties evaluated as follows:

$$u(\Delta T_{\text{DAQ}}) = \sqrt{u^2_{\text{1-stab}} + u^2_{\text{2-repeat}} + u^2_{\text{2-sh}} + u^2(\Delta T_{\text{REF}})}. \tag{3.2}$$

The term $u_{\text{1-stab}}$ is evaluated as above by taking a uniform distribution of the temperature signal for reference 1. On the other hand, reference 2 shows a more random behavior, thus an uncertainty evaluated through a repeatability contribution $u_{\text{2-repeat}}$ is considered, i.e. $\sigma(T_{\text{2-DAQ}})/\sqrt{N}$ where N is the number of measurements in the chosen interval. This type of evaluation is suitable even when the values are considered sampled from a bimodal distribution, as described above. The only assumption is that each performed measurement is independent from the others, that is reasonable by looking at the time series like in Figure 3.5. Then, the term $u_{\text{2-sh}}$ represents an uncertainty term due to the different self-heating of the sensor compared with its calibration condition. Since the datalogger supplies an impulsive current (estimated to be 2.5 mA) only during the acquisition interval, the self-heating of the sensor is smaller with respect to a condition of a constantly supplied current of 1 mA (as expressed by the calibration certificate). To take into account this fact, which is also

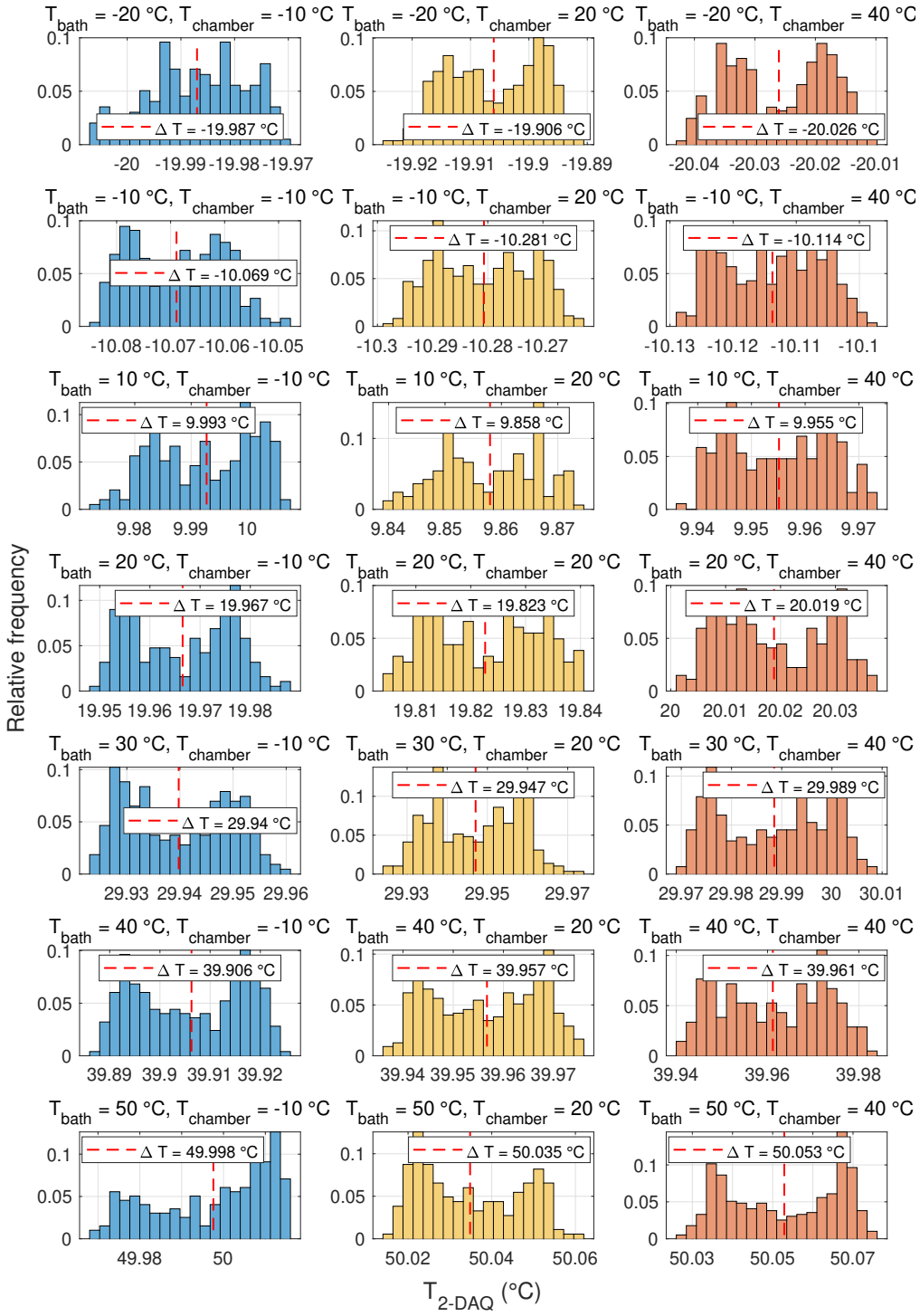


Fig. 3.7 Histograms of the temperature deviations between reference 1 and 2 during the datalogger comparison for each combination of T_{bath} and T_{chamber}. The histograms are scaled such that the value for each bin corresponds to the associated relative frequency.

Table 3.3 Empirical standard deviation of T_{DAQ} for each combination of T_{bath} and T_{chamber}.

	$\sigma(T_{\mathrm{DAQ}})/\mathrm{mK}$		
$T_{\mathrm{bath}}/^{\circ}\mathrm{C}$	$T_{\mathrm{chamber}} = -10\,^{\circ}\mathrm{C}$	$T_{\mathrm{chamber}} = 20\,^{\circ}\mathrm{C}$	$T_{\mathrm{chamber}} = 40\,^{\circ}\mathrm{C}$
−20	9.0	7.9	8.4
−10	8.5	8.6	7.7
10	8.8	8.8	9.4
20	10	10	9.6
30	9.7	11	11
40	11	11	11
50	14	12	14

the reason for a potential bias of the datalogger measurements, the self-heating due to 1 mA of current was estimated by using the “ZERO-POWER" function of the Fluke 1595A, which simply extrapolates the temperature value for null current and computes the difference with respect to a given current. This estimate is used as $u_{\text{2-sh}}$, which is an overestimation given that the self-heating is not totally eliminated in datalogger measurements. Although this approach is conservative, it allows to consider the self-heating in the uncertainty budget, which was estimated 4.7 mK for every T_{bath}, a value comparable to or greater than other uncertainty contributions. The last term, $u(\Delta T_{\mathrm{REF}})$, is the uncertainty of the applied correction reported in Table 3.2 as already commented.

The estimated ΔT_{DAQ} and the corresponding standard uncertainty, shown in Table 3.4 and plotted in Figure 3.8, prove that the datalogger performance is highly sensitive to the environmental temperature. The measurements performed for $T_{\mathrm{chamber}} = -10\,^{\circ}\mathrm{C}$ are affected by an error between 0.06 and 0.07 °C, which decreases for greater temperature of the chamber and becomes comparable with the calibration uncertainty for $T_{\mathrm{chamber}} = 40\,^{\circ}\mathrm{C}$. In an IP68 box, as it is used for protecting the datalogger in an outdoor environment, the internal environment is not fully thermilized with the outside air because of the constant supply of heat from the stored electronics and due to the use of insulating material placed on the internal walls of the box. Nevertheless, the internal temperature still varies to a certain degree following the external conditions, hence a correction of the measurements shall be considered. Note that, even in ideal conditions of 20 °C, the error is between 0.025 and 0.035 °C, which is not negligible. The measured temperature differences

Table 3.4 Estimated differences, and corresponding uncertainties, between reference thermometer 1, measured with the resistance bridge, and reference 2 measured with the datalogger, for each combination of the temperature of the bath and that of the climatic chamber.

	$\Delta T_{\mathrm{DAQ}}/\mathrm{mK}$		
$T_{\mathrm{bath}}/^{\circ}\mathrm{C}$	$T_{\mathrm{chamber}} = -10\,^{\circ}\mathrm{C}$	$T_{\mathrm{chamber}} = 20\,^{\circ}\mathrm{C}$	$T_{\mathrm{chamber}} = 40\,^{\circ}\mathrm{C}$
−20	-63(4.9)	-36(4.8)	-17(4.8)
−10	-66(4.9)	-30(4.8)	-16(4.8)
10	-69(4.9)	-27(5.0)	-11(5.0)
20	-69(5.0)	-25(5.1)	-13(4.9)
30	-66(5.0)	-23(5.3)	-11(5.0)
40	-66(5.2)	-24(5.0)	-14(4.9)
50	-61(5.1)	-25(4.8)	-5.9(5.2)

have an associated uncertainty which is about 5 mK for each combination of bath and chamber temperature, hence the relative uncertainty increases when the error decreases.

Since the values of ΔT_{DAQ}, associated to a set temperature of the chamber, are visibly comparable within two standard uncertainties, it is not possible to identify a trend with the variation of the bath temperature. In this case, assuming that the bath temperature has no influence and that each temperature difference refers to the same measurand, it can be convenient to estimate a consensus value. This is generally done using a weighted average, considering the inverse of variances as weights. The equations for estimating this quantity and evaluating the associated standard uncertainty are here adapted from [66] :

$$\overline{\Delta T_{\mathrm{DAQ}}} = \frac{\sum_{i=1}^{N_{\mathrm{bath}}} \frac{\Delta T_{DAQ,i}}{u^2(\Delta T_{DAQ,i})}}{\sum_{i=1}^{N_{\mathrm{bath}}} \frac{1}{u^2(\Delta T_{DAQ,i})}} \tag{3.3}$$

$$u(\overline{\Delta T_{\mathrm{DAQ}}}) = \sqrt{\frac{1}{\sum_{i=1}^{N_{\mathrm{bath}}} \frac{1}{u^2(\Delta T_{DAQ,i})}}} \tag{3.4}$$

Using these equations, it is possible to estimate the weighted average for each T_{chamber}, which is highlighted in Figure 3.8 by the black dashed lines, whereas the dotted lines delimit the interval $\pm u(\overline{\Delta T_{\mathrm{DAQ}}})$ covered by the corresponding standard

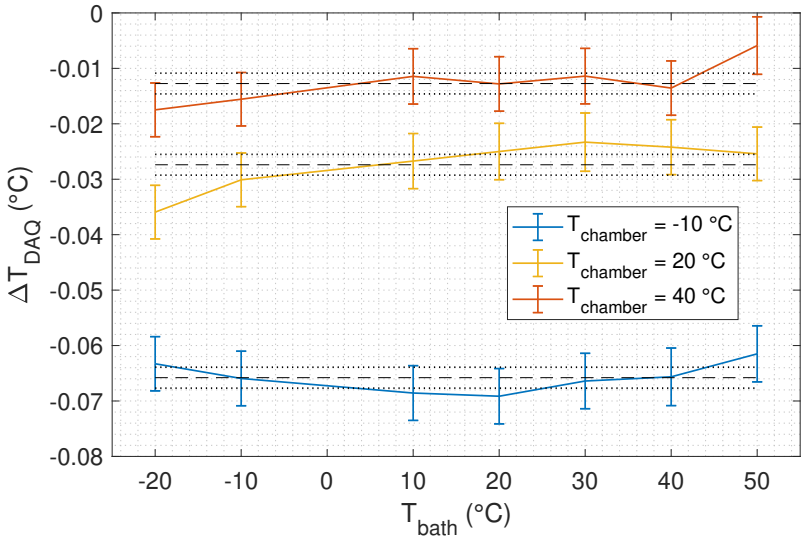


Fig. 3.8 Temperature difference ΔT_{DAQ} plotted as a function of bath temperature, for each value of the set temperature of the climatic chamber. The uncertainty bars correspond to the standard uncertainty reported in Table 3.4. The dashed black lines represent the consensus value $\overline{\Delta T_{\mathrm{DAQ}}}$ estimated with the variance-weighted average of all the values for a certain temperature of the chamber, whereas the dotted lines delimit the interval $\pm u(\overline{\Delta T_{\mathrm{DAQ}}})$ covered by the corresponding standard uncertainty.

uncertainty. The numerical values of the average, the absolute and relative uncertainty are shown in Table 3.5. Note that the uncertainty decreases since more estimates of the same measurand, as assumed in this case, add more information and allow to reduce the uncertainty over the best estimate. Indeed, even for the smallest correction, the relative uncertainty is below 15 %.

$T_{\mathrm{chamber}}/^{\circ}\mathrm{C}$	$\overline{\Delta T_{\mathrm{DAQ}}}/\mathrm{mK}$	$u(\overline{\Delta T_{\mathrm{DAQ}}})/\mathrm{mK}$	$u_r(\overline{\Delta T_{\mathrm{DAQ}}})/\%$
-10	-66	1.9	3.0
20	-27	1.9	7.4
40	-13	1.9	15

Table 3.5 Weighted average $\overline{\Delta T_{\mathrm{DAQ}}}$ with corresponding absolute, $u(\overline{\Delta T_{\mathrm{DAQ}}})$, and relative, $u_r(\overline{\Delta T_{\mathrm{DAQ}}})$, uncertainty.

The reported consensus errors should be used to correct the temperature measurements in the following way:

$$T_{\text{2-DAQ}_c} = T_{\text{2-DAQ}} - \overline{\Delta T_{\mathrm{DAQ}}}. \tag{3.5}$$

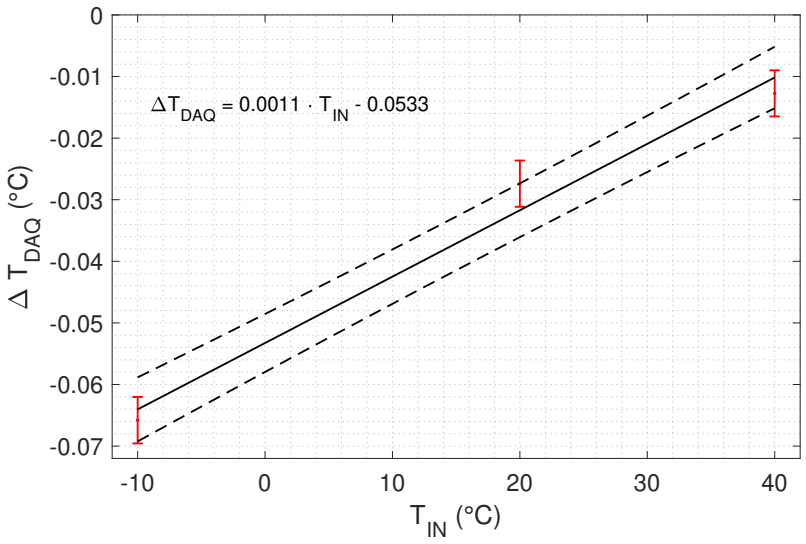


Fig. 3.9 Weighted linear least-squares fit, solid black line, used to determine the temperature correction of the measurements, performed with the datalogger, as a function of the device case temperature T_{DAQ}. The values highlighted in red correspond to the weighted mean of the estimated averages plotted with a coverage factor $k = 2$, as also for the prediction interval of the fit highlighted by the dotted black lines.

This correction is valid only for specific values of the environmental temperature, whereas to apply it in operating conditions it is necessary to estimate a temperature-depending correction $\overline{\Delta T_{\mathrm{DAQ}}}(T_{\mathrm{IN}})$, where T_{IN} is the temperature inside the box where the datalogger is stored. To evaluate this function, an uncertainty-weighted least-squares fit is made. More tested chamber temperatures should be considered for better function identification of the error correction, but for the sake of presenting the methodology, a linear function is assumed over the conducted measurements. The internal temperature is approximated with the temperature of the chamber for the tested points. The obtained fit and the corresponding coefficients are shown in the formula in Figure 3.9. The uncertainty of the coefficients resulting from the fit is propagated through the formula and added in quadrature with the uncertainty of the consensus values for evaluating the uncertainty of the predicted correction. Eventually, the standard uncertainty associated with the predicted correction is 2.5 mK, independently of the datalogger temperature.

After applying the correction, the type-B uncertainty associated with a single temperature measurement done by the datalogger is obtained by combining the

uncertainty contributions evaluated for the characterization:

$$u(T_{\text{2-DAQ}_c}) = \sqrt{u^2_{\text{2-cal}} + u^2_{\text{2-sh}} + u^2_{\text{noise}} + u^2(\Delta T_{\text{DAQ}}(T_{\text{IN}}))}. \tag{3.6}$$

Note that the u_{noise} can be estimated by taking the standard deviation of the temperature distribution shown in Figure 3.7, whereas the term $u_{\text{2-sh}}$ is highly dependent on the medium with which the thermometer is in contact, whose estimation is a tough task especially for air temperature measurements. Under the assumptions made for the characterization and neglecting the self-heating term, the overall standard measurement uncertainty is evaluated to be between 0.013 °C and 0.015 °C which is adequate for meteorological applications. However, if no correction is applied, it is necessary to add the error as an additional uncertainty term to take it into account. In this case, the type-B standard measurement uncertainty varies between 0.014 °C and 0.076 °C. Probably the accuracy specification reported by the manufacturer in Table 3.1 refers to the error compared with a reference thermometer, which, however, is not true as demonstrated by the laboratory test. Also the uncertainty specification likely corresponds to the noise affecting the measurement system, but it is highly underestimated by the manufacturer.

Despite the specific results shown here, the key point of this characterization was to present an effective experimental methodology and analysis of measurements for testing meteorological dataloggers. The uncertainty due to the use of dataloggers is just one uncertainty component of air temperature measurements that is usually neglected. In the following subsection, another component is analysed, which is the influence of rain on air temperature measurements, related to the group of uncertainties due to environmental effects.

3.2 The effect of rain on air temperature measurements

Atmospheric factors, such as wind, rainfall, solar radiation and relative humidity influence the air temperature measurement, thus compromising the accuracy of meteorological thermometers. Solar screens are used to protect the sensor and reduce the radiative error affecting the measurements, but they are not perfect and expose the thermometer to a microclimate condition which may not be representative of the ambient conditions [35]. For these reasons, field intercomparisons of artificially/naturally ventilated screens have been conducted to analyze the error affecting the air temperature measurement inside the screen with respect to a chosen reference system [67, 68]. A major effect causing biases and uncertainty in air temperature measurements is the influence of rain on the temperature sensor and its protecting screen. Although the psychrometric effect is well known [69], another error on the temperature readings due to rainfall have been recognized in the standard ISO 17714 (2007) [69], which recommends test methods for characterizing air thermometers with their radiation screen. In annex A of this documentary standard, it is written: "Another effect is that the temperature of the precipitation is generally lower than the temperature of the air [...] This can suddenly cool the screen at a different rate than the air (up to 5 K in 5 min)". Even though the high impact that this effect can have on measurements, there is no further information that quantifies it and no studies are available in literature. For this reason, the work presented here aims to highlight the influence of the rain in the readings of temperature sensors in automatic weather stations (AWS), with radiation screens of different types, such as naturally and artificially ventilated. The study is conducted by introducing an artificial rainfall generator and an adequate system to measure air temperature. The first is necessary to simulate controlled rainfall with varying water temperature and intensity, whereas the latter is needed since "There is no recognized reference system for measuring the true air temperature"[69]. Then the measurements of the sensors under rainfall conditions are then compared to the reference system and the results are analyzed and discussed. Finally, additional recommendations on a test method and further work are advised.

3.2.1 Methods

This section describes how the test setup has been designed and constructed, including a detailed description of the artificial rainfall simulator and of the reference system for air temperature measurement. For practical reasons, the test setup, shown in Figure 3.10, is built in a tarpaulin covered outdoor shaft, next to the laboratory building. The tests are performed by comparing sensors, with screens of different manufacturers and operating under rainfall conditions, with respect to the reference system, which is used to accurately measure air temperature in dry condition. The operating principle of the test set-up is described here following. Tap water is used for the rainfall generator and its temperature is controlled for testing at different rain temperatures. Cooling of water is performed by means of a plate heat-exchanger connected to a brine cooling system. The resulting water temperature, T_W, is monitored using a PT500 sensor. Experiments are performed with water temperatures between 6 °C and 14 °C.

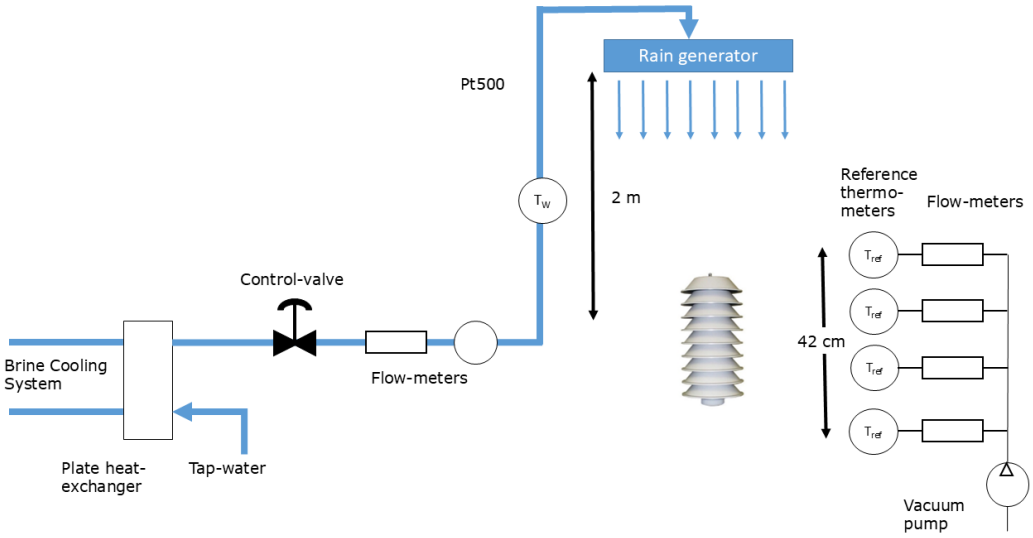


Fig. 3.10 Schematics of the set-up for testing the performance of meteorological air-thermometers under influence of rain.

The flow rate is controlled with a manual control valve regulating the water supply pressure and checked with a flowmeter. The accumulated volume is also measured with a mechanical flow meter. The air temperature is not controlled. Meteorological thermometers of different design are placed under the rainfall simulator and

measurements are made relative to the reference system. The latter consists of four aspirated thermometers, designed and constructed at Danish Technological Institute (DTI). The aspiration is obtained through a vacuum pump, whereas the flow rate is controlled with a valve and measured with a flowmeter. The reference thermometers are positioned beside the device under test (DUT) for measuring the air temperature but protected from being directly exposed to the rain. All the equipment used for the measurements is placed in a waterproof box. For data acquisition a LabVIEW® program is developed, whereas data analysis is conducted using MATLAB®. Design details of the rainfall generator and of the reference system are here reported.

Rain generator

The design of an artificial rain generator must ensure realistic amount of rain, which at the same time must be evenly distributed above the test area. Different approaches have been tested but it was found that using a spiral shaped perforated water hose was the best since it generates a rainfall intensity in an acceptable range and area distribution. The water hose is suspended approximately two meters above the reference thermometer setup and the DUT (see Figure 3.11a). Rainfall intensities from approximately 0-100 mm/h can be reached using this setup by regulating the water supply pressure. The rainfall intensity corresponding to the manually set flow rate is reported in Table 3.6. Moreover, the spatial variation of the rainfall intensity is measured. Several identical laboratory measuring cups are positioned in a matrix below the rainfall and, after a given period, the generator is shut off and the water level in each of the glasses is measured. The estimated spatial distribution of the rainfall intensity is then plotted as shown in Figure 3.11b. It is observed that the maximum intensity is not centered over the experimental area. Nevertheless, the DUT is placed in the center since a suitable intensity is achieved. At the boundary of the area, the intensity is below 5 mm/h, thus being a suitable spot to place the reference system. Tests were made where the position of the DUT in the rain area is changed to see if an uneven distribution of water has any effect and no significant change was found. The droplet size was not controlled or measured during the tests. As the drops were created passively, their diameter is expected to be in between 2 and 4 mm, as smaller droplets would stick to the generator, and larger droplets are unstable and will break into multiple smaller. droplets

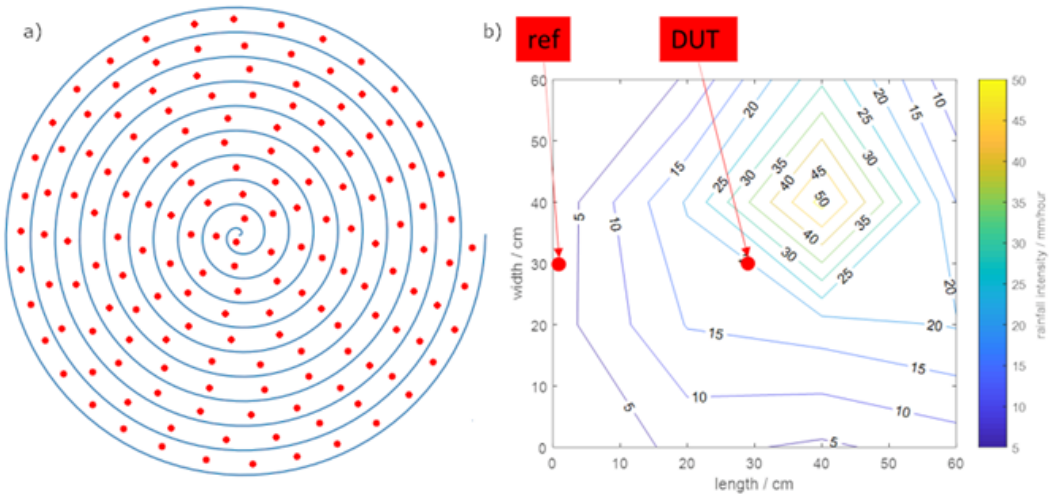


Fig. 3.11 a) Schematics of the rain generator. A standard water hose is perforated with equal distances between each hole and curled up in a spiral. The rain generator is placed approximately 2 meters above the reference thermometer setup and the DUT. b) The spacial rain intensity over the rain area is measured. The red boxes "ref" and "DUT" indicates where the thermometer reference setup and the DUT instrument are placed, respectively.

Reference system for air temperature measurement

When constructing a contact thermometer for air temperature measurement the thermodynamics raises several general issues that need consideration to reduce measuring errors (e.g. [70, 71]). The convection heat flux from air to the sensor element must be increased and the errors by conduction and radiation must be decreased. When constructing the reference measurement standard the following issues were taken into account:

- The time constant of the reference measurements system needs to be faster than the DUT and, because the air temperature is not actively controlled, fast enough to avoid lag errors become an issue.
- The sensor need to be sufficiently shielded to reduce radiation error.
- Using resistance thermometers, self-heating of the reference sensors is an issue.

To achieve a fast time constant and optimization of the heat flux from air to sensor the following strategy was realized: a sensor with small dimensions was chosen

and the heat transfer from the air to the sensor was optimized by increasing the airflow around the sensor. Self-heating effects were reduced by using low measuring currents for the resistance thermometers and thereby limiting the power dissipated in the sensor. Concerning dimensions, thermistors are a good choice due to their small size and fast reaction times. Glass encapsulated thermistors have proven to be reliable in practical air temperature applications [72–74] and a guideline of the Consultative Committee for Thermometry (CCT) on the calibration, linearization, and uncertainty evaluation is available [75]. Because of these qualities, Negative Temperature Coefficient (NTC) glass encapsulated thermistors are chosen, with a nominal resistance of 10 kΩ at 25 °C was chosen. A Fluke 1586A Super-DAQ Precision Temperature Scanner® is chosen as indicator since it uses a sufficiently low and constant measurement current of 10 µA, limiting self-heating effects. Overall, the measurement accuracy for this indicating device is specified to be ±3 mK. To increase the energy-transfer from the surrounding media to the sensor, an aspiration thermometer is constructed as suggested by Michalski et al. [71]. To increase the redundancy, thus the reliability of the reference temperature data, four such thermometers are assembled and labelled as: Meteo1, Meteo2, Meteo3 and Meteo4. Each thermistor is mounted strain-free to four varnished copper wires. The wires are then twisted together pairwise to minimize electromagnetic interference. This also results in a more rigid structure of the final sensor. The sensor is then mounted "straight through" a ¼" T-tube fitting, where the tube diameter has been widened. The end of the T from which the wire extends is sealed. Air can now be aspirated past the sensor element by connecting a pump to the remaining connector of the T fitting. Errors due to radiation from external sources may be significant [76], therefore a cylindrical radiation screen is constructed. This construction (dimensions and materials) follows the guidelines laid out in ISO 7726 [77] and Baker et al. [70]. The cylindrical screen is made from 0.4 mm thick sheets of reflective aluminum (low emissivity). The inner diameter of the screen is 10 mm. The inside of the cylinder is painted black by Nextel Velvet Coating 811-21® with an emissivity of approximately 0.98.

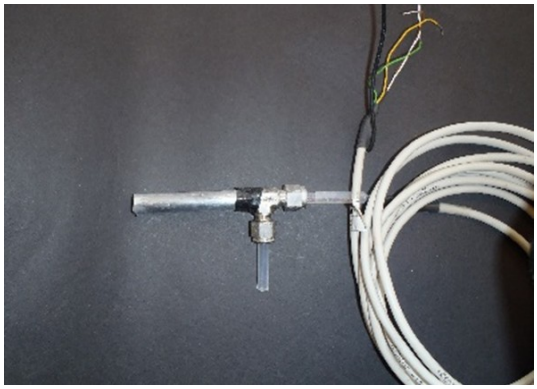


Fig. 3.12 Picture of the reference aspirated thermometer developed by DTI.

When air is aspirated, the heat transfer to the sensor is increased and in combination with the shielding of radiation sources, the sensor temperature should approximate the air temperature. Before using the reference system to evaluate the effect of the rain, the four reference aspirated thermometers were calibrated by comparison to a Standard Platinum Resistance Thermometer (SPRT) in a liquid bath, to ensure the traceability to the national standard. The thermistors were placed during the calibration in glass tubes in still standing air to evaluate the self-heating effect. They were also mutually validated by comparison in a climatic chamber with the set-up that is routinely used in DTI's calibration laboratory for calibrating air temperature and relative humidity sensors [78, 79]. The final expanded measurement uncertainty is below 10 mK for all the constructed thermometers and takes into account the calibration process (setup, least-squares fit, self-heating) and the uncertainty related to the indicator. Note that this uncertainty is strictly related to the instrument and it does not consider any effect due to quantities of influence (as it could be rain or solar radiation). The optimal flow rate was found experimentally by comparing the developed aspirated thermometers to four calibrated PT100 and a Hart Scientific Model 1560® Thermometer Readout in a climatic chamber. This set-up has been validated in the EURAMET P1061 intercomparison [78]. The optimal flow rate is found to be 2.5 l/min, corresponding to an air speed of about 1.3 m/s, which is used in all the experiments reported in this paper. Thanks to the larger immunity to solar radiation due to the small size of the employed thermistors, the tuned air speed is enough to guarantee a predominant heat exchange by convection, reducing possible temperature measurement interference due to radiative transfer of heat.

3.2.2 Tests and results

The developed experimental setup is used to compare two different thermometers, with associated screens produced by different manufacturers, having comparable calibration uncertainty. The devices under test, shown in Figure 3.13, are labelled DUT1 and DUT2. DUT1 is a sensor, of not specified type, shielded by a naturally ventilated screen and DUT2 is an artificially ventilated PT100 sensor. The sampling time for both sensors was set to 5 s. To investigate the influence of rain temperature and rain intensity on the performance of the devices, they were placed, one by one, under the rainfall generator and compared to the reference system in still or very slow moving air. During tests, the difference between water and air temperature is ideally kept constant (unless it is changed on purpose). It is reasonable that, as the distance between ground and the rain generator is only a few meters, the air volume exposed to rain reaches a near-thermal equilibrium when the generator is switched on. In most real-life situations, at first a fast drop in air temperature is expected, followed by an equalization between rain and air temperature, as described in Byers et al. (1949) [80]. This should be considered when interpreting the experimental results. In the following, details on the tests performed on both devices are given with the discussion on the comparison of the performances.

(a) (b)

Fig. 3.13 (a) DUT1: naturally ventilated thermometer (not specified type) (b) DUT2: artificially ventilated PT100.

Starting from the DUT1 test, the rainfall intensity, set during the experiment using the control valve, is shown in Figure 3.14. At the beginning, the generator is turned on setting a flow rate of $0.6\,\mathrm{l/min}$ ($\approx 20\,\mathrm{mm/h}$), then the rate is decreased to $0.4\,\mathrm{l/min}$ ($\approx 15\,\mathrm{mm/h}$) and finally to $0.2\,\mathrm{l/min}$ ($\approx 10\,\mathrm{mm/h}$). The difference between water and air temperature is approximately $-13\,\mathrm{K}$ in the first part of the test and then decreased to about $-3\,\mathrm{K}$ after having changed the flow rate to $0.2\,\mathrm{l/min}$.

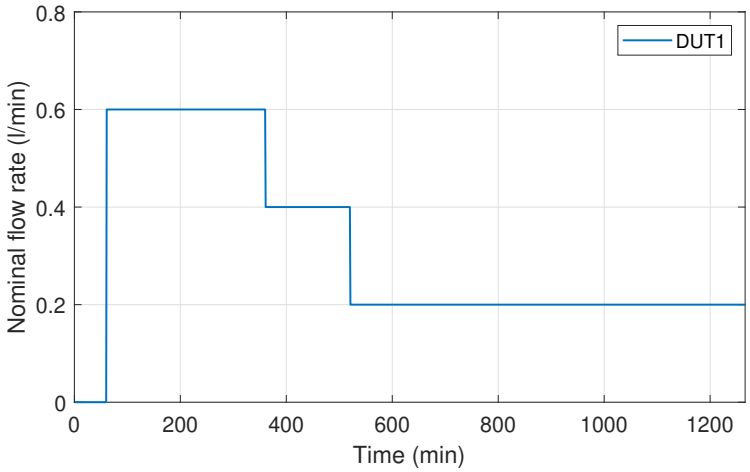


Fig. 3.14 Nominal flow rate set during the exposure of DUT1 sensor and screen under rainfall conditions.

A 5-minute average is used to analyze the data (for both DUTs) in order to filter out not-significant variations of the sensors' behavior due to fast transients of the air temperature or due to not perfect stability of the generator. The 5-min averages of the air temperature by the reference system (T_{REF}), the water temperature (T_W) and the air temperature measured by the DUT1 (T_{DUT1}) are shown in Figure 3.15. A difference is apparent between DUT1 and the reference system, which decreases in the last part of the curve when the flow rate and the difference $T_W - T_{REF}$ is decreased.

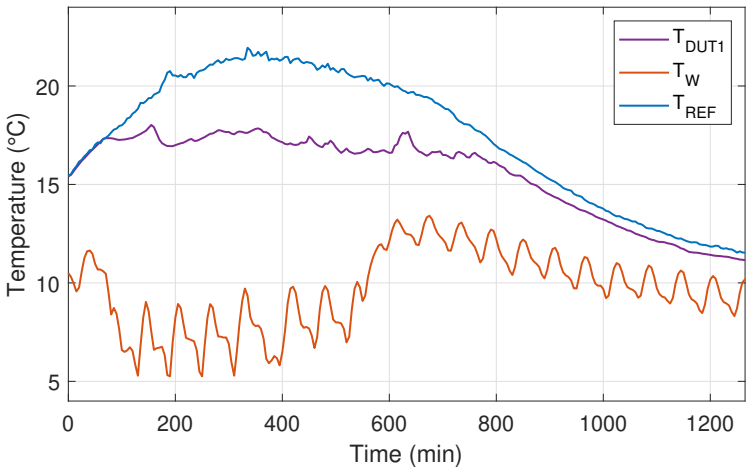


Fig. 3.15 5-minute average measurements related to DUT1 sensor and screen under rainfall conditions: T_{REF} is the air temperature measured by the reference system (blue line), T_W is the temperature of water measured by the PT500 (orange line) and T_{DUT1} is the temperature read by the DUT1 (purple line).

To make an estimate of the error $T_{\mathrm{DUT1}} - T_{\mathrm{REF}}$ for different rain intensities, the same data are plotted, now considering differences in Figure 3.16. There are two vertical scales, one for the DUT1 error and the other for the difference between water and air temperature. The vertical dashed black lines refer to the instants when the flow rate changes (nominal values are those shown in Figure 3.14). At the beginning the error is approximately zero since the DUT1 is outside the experimental area to verify if it is aligned with the reference system (as seen in Figure 3.15). The initial decrease, after having turned on the generator, is probably caused by dew deposited on the sensor. Likely, sprays generated by the impact of water with the ground produced dew deposition on the screen and on the sensor. The evaporation of water droplets from the screen plates decreases the temperature of the microclimate inside the screen, while the evaporation from the sensor could directly decrease its temperature. This situation could be similar to what happens in nature in the first instances of rainfall or when the rain intensity is very low. This phenomenon does not affect the reference system thanks to the continuous flow of air.

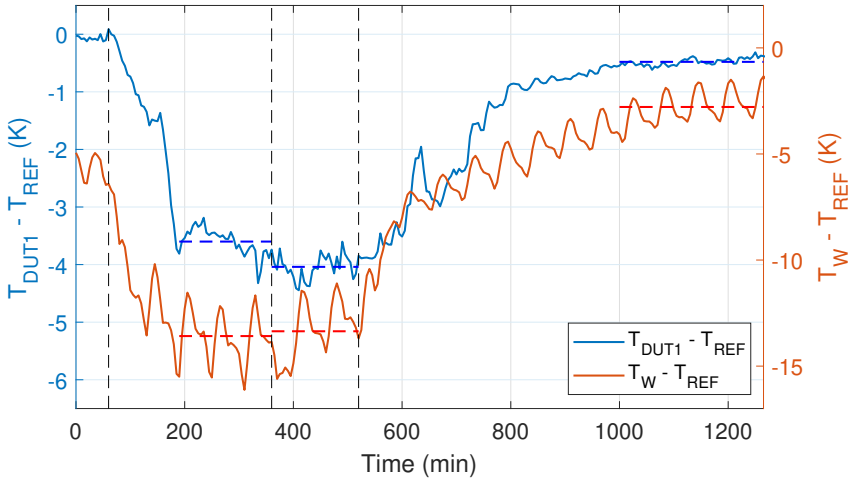


Fig. 3.16 5-minute average temperature differences. The blue curve is the DUT1 error, $T_{\mathrm{DUT1}} - T_{\mathrm{REF}}$, whereas the orange curve is $T_{\mathrm{W}} - T_{\mathrm{REF}}$, that is the difference between water and air temperature measured by the reference system. The estimated measurement error and rain-air temperature difference are displayed by horizontal dashed lines (in blue and orange respectively), whereas the vertical dashed black lines refer to the instants when the flow rate changes (nominal values are those shown in Figure 3.14).

On the other hand, air can stay inside naturally ventilated shields for longer times when wind is low, allowing heat transfer between air and shield, with a consequent decrease of air temperature. The device is then placed in the rain, at about minute

150, which produces a significant cooling of DUT1. When the water-air temperature difference was around -13 K, the error became approximately constant at about -4 K, even for a smaller flow rate equal to 0.4 l/min. In the last part of the test, the flow rate is decreased to 0.2 l/min and the water-air temperature difference is set to approximately -3 K (as mentioned above), which is reached around 7 hours later. This slow behavior is explained by the fact that the intensity of the heat transfer from the air to the sensor and shield is very low in the nearly still standing air and that the air temperature (not controlled) is decreasing during that part of the experiment, thus requiring more time to reach an almost constant difference. During this time interval the rain intensity remained constant, therefore the (absolute) decrease observed by the measurement error from -4 K to about -0.5 K can be brought back to the decrease (in magnitude) of the difference $T_W - T_{REF}$. Table 3.6 reports the estimated values for $T_{DUT1} - T_{REF}$ and $T_W - T_{REF}$ for different part of the experiment (corresponding to different set flow rate). The uncertainty for the average values is calculated considering the instrumental uncertainty of the thermometers and the dispersion of the signal in the time interval of interest.

Table 3.6 Estimations of the measurement error $T_{DUT} - T_{REF}$ (with associated standard uncertainty) for DUT1 and DUT2, at different rain-air temperature differences and flow rates (FR).

	FR = 0.2 l min^{-1}		FR = 0.4 l min^{-1}		FR = 0.6 l min^{-1}	
DUT	$T_{DUT} - T_{REF}$	$T_W - T_{REF}$	$T_{DUT} - T_{REF}$	$T_W - T_{REF}$	$T_{DUT} - T_{REF}$	$T_W - T_{REF}$
1	-0.5 (0.1) K	-2.8 (0.8) K	-4.0 (0.2) K	-13.4 (1.4) K	-3.6 (0.3) K	-13.6 (1.2) K
2	-0.3 (0.1) K	-3.4 (0.7) K	-0.9 (0.3) K	-12.1 (1.1) K	-1.1 (0.2) K	-12.1 (1.1) K

The same experiment is reproduced for DUT2. Similarly to DUT1, Figure 3.17 reports the flow rate set during the execution of the experiment, Figure 3.18 shows the 5-min averages of T_{REF}, T_W and T_{DUT}, whereas Figure 3.19 displays the 5-min average differences. Even though the water-air temperature difference is set as for DUT1 experiment, the estimated averages $T_W - T_{REF}$, are not exactly the same since the (outdoor) air temperature is not controlled and due to the instability of the cooling system. However, the performances of the two devices under test can be considered approximately comparable since the $T_W - T_{REF}$ differences are comparable within their standard uncertainty. Indeed, the overall behavior of DUT2 is similar to DUT1, but the measurement error is lower thanks to the artificial ventilation. Indeed for

0.6 l/min as rainfall intensity the average DUT2 error is $-1.1(0.2)$ K instead of $-3.6(0.3)$ K for DUT1. A decrease of the flowrate to 0.4 l/min does not produce a significant variation in the error. Finally, in the last part of the experiment, for 0.2 l/min and decreasing the water-air temperature difference (in magnitude), the DUT2 error decreases to a value comparable to that of DUT1. Estimations of the measurement error with associated standard uncertainty are again reported in Table 3.6.

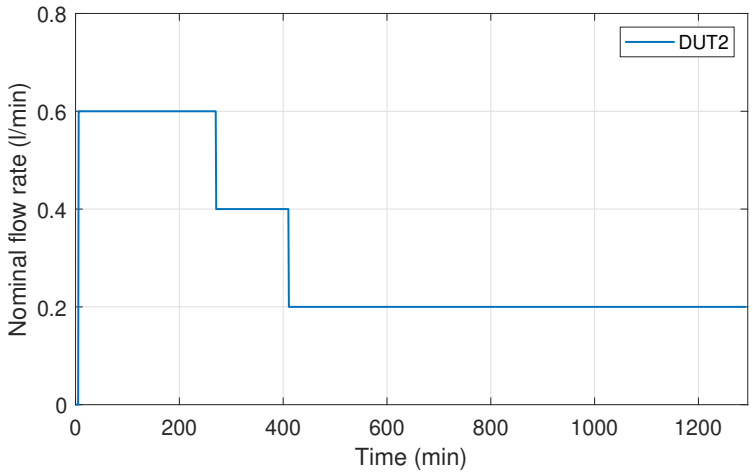


Fig. 3.17 Nominal flow rate set during the exposure of DUT2 sensor and screen under rainfall conditions.

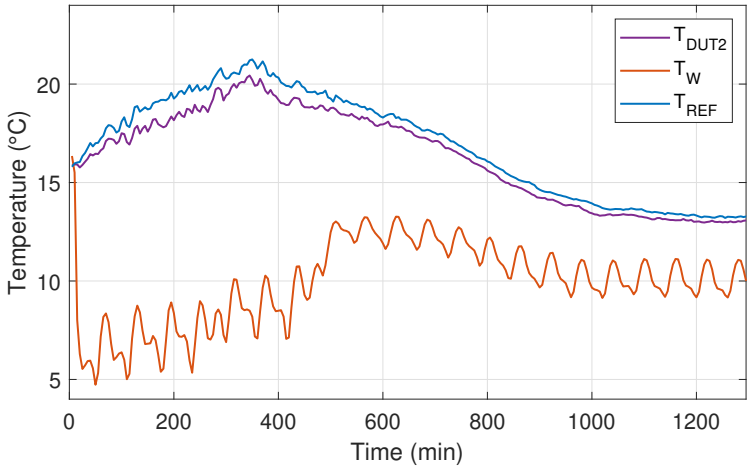


Fig. 3.18 5-minute average measurements related to DUT2 sensor and screen under rainfall conditions: T_{REF} is the air temperature measured by the reference system (blue line), T_{W} is the temperature of water measured by the PT500 (orange line) and T_{DUT2} is the temperature read by the DUT1 (purple line).

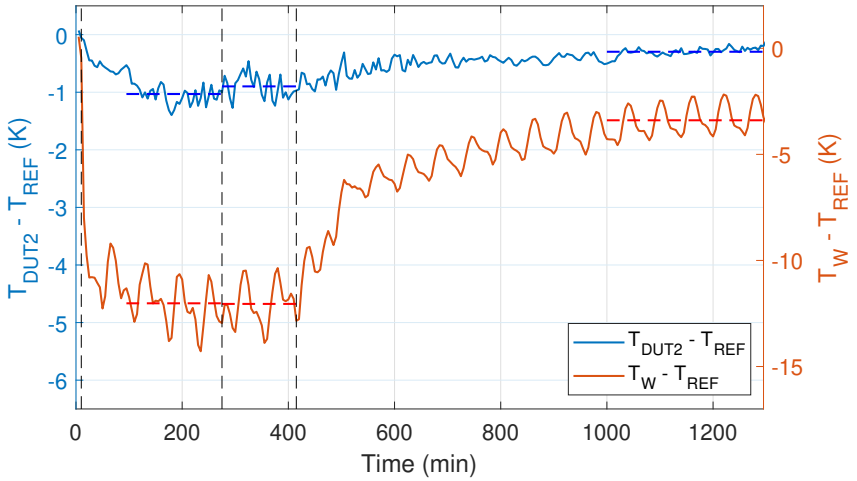


Fig. 3.19 5-minute average temperature differences. The blue curve is the DUT2 error, $T_{\mathrm{DUT2}} - T_{\mathrm{REF}}$, whereas the orange curve is $T_{\mathrm{W}} - T_{\mathrm{REF}}$, that is the difference between water and air temperature measured by the reference system. The estimated measurement error and rain-air temperature difference are displayed by horizontal dashed lines (in blue and orange respectively), whereas the vertical dashed black lines refer to the instants when the flow rate changes (nominal values are those shown in Figure 3.17).

3.2.3 Discussion

The performed tests show that the thermometer in a naturally ventilated screen is highly affected by rainfall. The maximum quantified error is equal to -4.0 K with an expanded uncertainty of 0.4 K ($k = 2$), corresponding to a rainfall intensity of 15 mm/h (0.41/min) and an average difference $T_{\mathrm{W}} - T_{\mathrm{REF}}$ equal to -13.4 K with expanded uncertainty of 2.8 K ($k = 2$). Naturally ventilated screens should allow a good coupling between the thermometer and the air through paths with the external environment (multiple parallel paths for multi-plates screens or spiral path as for DUT1). In slowly mowing or still air, as is the case in this experiment, the convective heat-flux from the outside air to the sensor is low. The walls of the screen are cooled by the rain, which reduces the air temperature inside the screen. The measurement chamber and the thermometer may also be wettened through splashes and dew deposition adding to the effect. This result shall be considered as a worst-case scenario since in most real-life situations the difference between rain and air temperature can be smaller than what is simulated here. The presence of wind would result in improved heat-transfer to the sensor reducing the measurement error but enhances evaporative cooling of the screen making the error larger. However,

the wind will also shorten the recovery time needed to measure again the correct temperature after rainfall. Overall, the artificially ventilated thermometer performs better than the naturally ventilated one under rainfall conditions as simulated in the experiments. The maximum quantified error is equal to -1.1 K with an expanded uncertainty of 0.4 K ($k = 2$), corresponding to a rain rate of 20 mm/h (0.6 l/min) and an average difference $T_W - T_{REF}$ equal to -12.1 K with expandend uncertainty of 2.2 K ($k = 2$). The difference in performance of DUT1 and DUT2 are as expected as the heat-transfer from the air to the sensor is improved as discussed above for DUT2. As for DUT1, the walls of the screen are cooled by the rain. The measurement chamber and the thermometer may also be wettened, but the continuous air flow inside the artificially ventilated screen would speed up evaporation of deposited water. Furthermore, the thermometer inside DUT2 is continuously in contact with fresh air, which stays in the screen for a very short time and is not cooled significantly by heat-transfer to the screen walls. An error remains, even though small, when the difference $T_W - T_{REF}$ and the rainfall intensity is decreased, which is comparable to the results obtained with naturally ventilated DUT1. This implies that also in mild rainfall conditions the thermometers could measure something not representative of the air temperature. As for DUT1, these results could be seen as a worst-case scenario, but they allow to say that the tested artificially ventilated thermometer is much less affected by rainfall than the naturally ventilated one. Since the error in real use is both dependent on the temperature difference between air and water, the rain intensity (in a lesser degree), and the wind speed, it is very difficult to apply a practical correction based on performed laboratory experiments. However, these tests give a first quantitative indication of the behavior of thermometers under rainfall conditions, highlighting the importance of the difference between rain and air temperature and how carefully air temperature measurements, under this condition, should be analyzed. To get more accurate data for quantifying the effect on different types of screens, more experiments need to be done under realistic conditions.

3.3 Summary of main achievements

This chapter addressed two critical but often overlooked sources of uncertainty in air temperature measurements: the contribution of meteorological dataloggers and the influence of rainfall on temperature sensors within different radiation screens. Both

investigations were conducted within the broader context of developing complete uncertainty budgets for air temperature measurements, essential for establishing reliable climate reference stations and ensuring measurement traceability.

Datalogger characterization: key findings

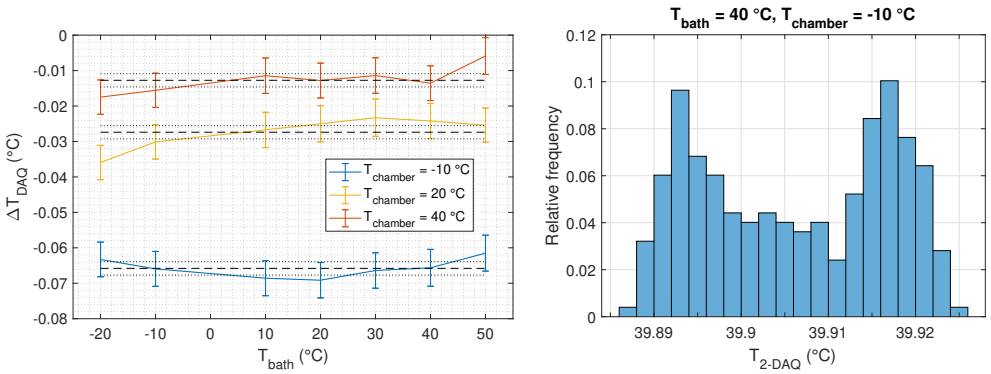
The characterization of the SIAP DA18K meteorological datalogger revealed significant systematic errors that are typically neglected in meteorological applications. The experimental methodology is based on the generation of reference resistance values using a Pt100 thermometer in a circulating bath, which first are read by a DC resistance bridge and then by the datalogger. The datalogger was placed inside a climatic chamber to test its performance by varying the operating temperature (T_{chamber}). The comparison was performed for different combinations of the bath and chamber temperatures. To compensate the fact that the instruments read the reference values in different time intervals, the comparison was performed using the temperature difference of two reference thermometers inside a comparator block in the bath (as shown in panel (a) of Figure 3.3), which is less sensitive to the instability of the bath. First both thermometers (1 and 2) are read by the resistance bridge, then, during the second phase, reference 2 is read by the meteorological datalogger. A picture of the complete experimental setup is shown in Figure 3.2.

Principle findings of the characterization include:

- **Temperature-dependent systematic error**: The datalogger exhibited substantial temperature-dependent biases ranging from $-66\,\text{mK}$ at $T_{\text{chamber}} = -10\,^{\circ}\text{C}$ to $-13\,\text{mK}$ at $T_{\text{chamber}} = 40\,^{\circ}\text{C}$, which far exceed the manufacturer's specifications. Figure 3.20a shows the temperature differences with the evaluated standard uncertainty calculated using Equations 3.1-3.2, that is about 5 mK for all the combinations. Because the dependence on the bath temperature is not significant, a consensus value is calculated, by a weighted average, for each tested chamber temperature, plotted with a black dashed line.

- **Electrical noise characteristics**: Temperature measurements done by the datalogger showed bimodal distributions with standard deviations between 7.7 and 14 mK, indicating complex noise patterns that cannot be adequately characterized without additional measurements with a greater sampling frequency. An example of the distribution of measurements is shown in Figure 3.20b.

- **Temperature correction**: Using the consensus values for biases, a correction depending on the external temperature is evaluated by a linear weighted least-squares fit. The standard uncertainty of the corrected temperature read by the datalogger is evaluated by Equation 3.6, adding to the calibration uncertainty the terms for the electrical noise and the applied correction. If no correction is applied, the error estimate should be included in the uncertainty budget, ranging in this case between 0.067 °C and 0.017 °C of standard uncertainty. Instead, the uncertainty of the corrected temperatures is smaller, about 0.011 °C, demonstrating the value of proper characterization and correction procedures.



(a) (b)

Fig. 3.20 (a) Temperature difference ΔT_{DAQ} plotted as a function of bath temperature, for each value of the set temperature of the climatic chamber. The bars correspond to the standard uncertainties associated with the estimated biases. (b) Example of histogram of the temperature deviations between reference 1 and 2 during the datalogger comparison for $T_{\mathrm{bath}} = 40\,^{\circ}\mathrm{C}$ and $T_{\mathrm{chamber}} = -10\,^{\circ}\mathrm{C}$.

The developed correction method provides a practical approach to field measurements, though requiring knowledge of the internal temperature of the box in which the datalogger is stored. Some dataloggers have an internal temperature sensor whose readings could be used to effectively determine the correction function. With the performed measurements a temperature-dependent systematic bias was proved, demonstrating also that the manufacturer specifications have low metrological rigor. Although the reported results are specific to the tested datalogger, likely the estimated error is typical of the model, which means that different samples from the same manufacturer would exhibit similar errors within a certain tolerance because of the manufacturing processes. On the other hand, dataloggers from different manufacturers may show different error patterns, hence confirming the importance of their

characterization for improving the quality of air temperature measurements. To further investigate this aspect, the characterization will be performed in the future to different models, extending the results also to direct measurements of the impulsive current and to a better noise characterization.

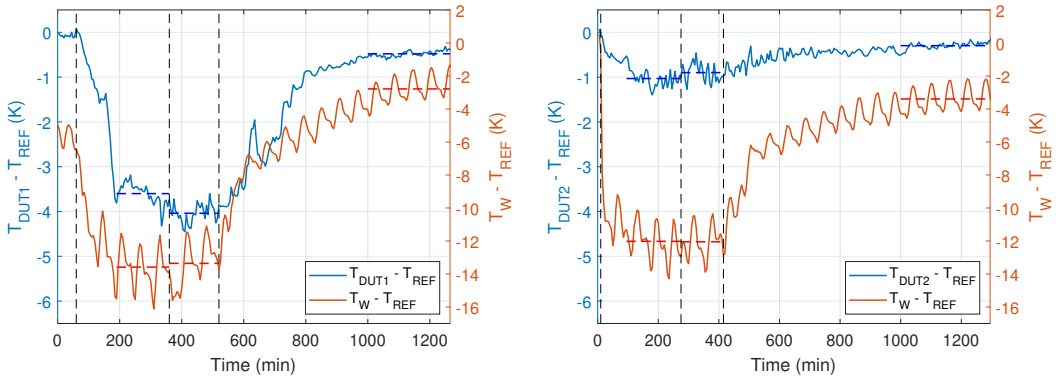
Rain Effect on Air Temperature Measurements: Key Findings

A rainfall generator and a reference system not influenced by rain, used to measure air temperature, have been designed and constructed by the Danish Technological Institute to investigate the effect of rain temperature and rain intensity on shielded air thermometers. Several tests were performed, whose measurements were analyzed in the context of the PhD project to establish the effect of rain on two different types of radiation screens. The reference system is constituted by NTC thermistors in aspirated radiation shields with optimized airflow (2.5 l/min, ≈1.3 m/s), which achieved measurement uncertainties below 10 mK and provided crucial baseline measurements for quantifying rain effects. The controlled quantification of the rain effect was performed by comparing, for different rainfall intensities, the readings of calibrated thermometers inside the screens (exposed to the generated rainfall) with those of the reference system. Critical findings include:

- **Naturally ventilated screens**: The maximum quantified error reached -4.0 K with an expanded uncertainty of 0.4 K ($k = 2$) for a rainfall intensity of 15 mm/h with a water-air temperature difference of -13.4 K. This represents a worst-case scenario but demonstrates the potential magnitude of rain-induced errors. An example of measurements considered for the estimation of the temperature difference is reported in Figure 3.21a.
- **Artificially ventilated screens**: These systems performed significantly better, with maximum error of -1.1 K with an expanded uncertainty of 0.2 K ($k = 2$) for a rainfall intensity of 20 mm/h. Figure 3.21b shows the measurements used to quantify the measurement error.
- **Physical mechanisms**: The accuracy of thermometers inside ventilated screens is influenced by how heat and moisture are exchanged between the air, the screen walls, and the sensor. In naturally ventilated screens, weak air movement leads to low convective heat transfer, so when rain cools the screen walls

(and sometimes wets the sensor), the internal air cools as well, biasing the temperature measurement. Wind can improve convective coupling with the air (reducing error), but it also increases evaporative cooling of wet surfaces (increasing error). In artificially ventilated screens, continuous airflow brings in fresh air that is less affected by cooled screen walls and speeds up evaporation of deposited water, which reduces measurement errors compared to natural ventilation. Nonetheless, some error persists during rainfall, especially when the rain–air temperature difference increases.

- **Intensity dependence**: While water-air temperature difference emerged as the primary factor, rainfall intensity showed secondary effects, particularly in the transition from dry to wet conditions where initial dew formation contributed to cooling.



(a) Naturally ventilated screen

(b) Artificially ventilated screen

Fig. 3.21 Comparisons between the screens and the reference system. The vertical dashed black lines refer to the instants when the flow rate changes (nominal values are those shown in Figures 3.14 and 3.17).

Although the developed methodology highlighted the superior performance of an artificially ventilated system under adverse conditions, the experiments were conducted in still or low-wind conditions, which represent worst-case scenarios for naturally ventilated screens. Furthermore, only two screen types were tested using an artificial rain generator that, while innovative, may not fully replicate natural rainfall characteristics. These limitations shall be overcome in future studies by refining the rainfall generator and performing a broader intercomparison to enhance the generalizability of the results.

The influence of wind speed was not investigated in this study, besides it plays a crucial role as described above. Applying a correction for the error, based on the reported experiments, is therefore not practical. However, this preliminary study gives a quantitative indication of the behavior of thermometers under rainfall conditions, highlighting the importance to further investigate this problem. The reported methodology can help to develop a standardized test protocol to qualify the device with regards to the measurement error as a function of the rain and to the speed it recovers after rainfall. To develop a standardized test, the following minimum requirements should be specified:

- Requirements of a suitable rain generator: uniformity of rainfall and surface area covered.
- Use of a suitable reference measurement system for measuring air temperature.
- Distance between rain generator and device under test.
- Rain intensity given as a realistic range.
- Difference between air and rain temperature and stability.
- Time from the switch on to the switch off of the generator.
- Maximum wind speed allowed.

As climate science demands increasingly accurate and reliable temperature records, the systematic characterization of measurement system uncertainties becomes not just scientifically valuable but essential for maintaining confidence in climate observations and the decisions based upon them. Future climate reference stations will benefit from incorporating these findings, both in terms of specific corrections and general approaches to uncertainty evaluation. The chapter contributes to the broader goal of establishing measurement systems capable of detecting and accurately quantifying climate change signals while maintaining full metrological traceability.

Chapter 4

Influence of installation height on air temperature measurements

An important aspect of atmospheric measurements considered in climatological studies is the metadata about the instruments' siting. This information is crucial for detecting biases in the temperature time series caused by changes in the environment close to the installation, for instance due to construction of buiildings or the advancement of a forest. Therefore, these variations must be homogenized before using the measurements for climatological studies.

As presented in Chapter 2, an ideal scenario for climatologists is to use reference measurements that are representative of the environment around the station. Since the majority of surface stations do not comply with this requirement due to biases introduced by heat sources or sinks in their proximity, the WMO provides a siting classification scheme in its Guide No. 8 [16]: the stations are divided in 5 classes, as shown in Figure 3.1, with a decreasing representativeness of measurements, for a certain quantity, from 1 to 5. An additional uncertainty component, related to the imperfect definition of the measurand because of siting, is suggested for each class and shall be added to the uncertainty budget of the measurements. Even though not explicitly reported, the guide specifies that the suggested uncertainties are mainly estimated through comparative studies. To mention a few recent works of this kind, in [20] Coppa et al. estimated the air temperature bias introduced by the presence of a road close to a weather station, whereas in [53] Izquierdo et al. gave an estimation of the temperature bias due to buildings close to the station.

When speaking about siting and representativeness, it is straightforward to think about obstacles or erroneous exposure of instruments in the horizontal direction. However, the siting of an instrument is defined in [81] by the WMO rapporteur Ehinger as "the act of finding a geographical location for it, defined both horizontally and *vertically*". For temperature measurements the installation height determines the exposure of the thermometers to layers of air that interact differently with the ground surface. As clearly stated in the WMO guide "the height above ground level is specified because large vertical temperature gradients may exist in the lowest layers of the atmosphere that can influence the temperature measurement" [16]. Hence, a standardized height interval for the installation of thermometers is suggested: "air temperature should be representative of the free air conditions surrounding the station over as large an area as possible, at a height of between 1.25 m and 2 m above ground level". For the siting classification it is also specified that "the height should never be less than 1.25 m" and that "the respect of the higher limit is less stringent, as the temperature gradient versus height is decreasing with height". An example is then reported, stating that the temperature difference for sensors located between 1.5 and 2 m is less than 0.2 °C. No explanations or comparison works are reported for justifying this estimate.

Despite the recommendation of clearly stating the height of the installed thermometer, it is not transparent how the suggested installation interval was defined. As can be understood from the guide, the lower limit was likely chosen to avoid too large an influence from the ground, even though it remains unclear how the value was estimated. The vertical temperature gradient close to the surface was studied in detail since the beginning of the twentieth century, as demonstrated by the measurements reported by Geiger in his fundamental book on the climate near the ground [82]. Furthermore, micrometeorologists extensively studied temperature profiles close to the surface by applying the Monin-Obukhov (MO) similarity theory [83–90]. However, the performed measurements are mainly concentrated during specific days and are not designed for the comparison of thermometers inside screens at different heights for continuous meteorological or climate monitoring, hence they cannot justify the chosen interval.

After an investigation across historical documents, it was found that the standardized installation height interval dates back to the Conference of Directors of the International Meteorological Organization (IMO) held in Washington during September and October 1947 [91]. During the conference, a series of technical

regulations, proposed by the just-born Commission for Instruments and Methods of Observations (CIMO), were voted to standardize atmospheric measurements among national weather services [92]. One of these regulations, the number 136, was about "observations de la temperature de l'air au sol" and states that "at stations on land the air temperature SHOULD be measured at a height between 1.25 and 2 metres above the ground. For stations with a considerable snow cover greater heights are permissible". The very same recommendation is also reported in the first CIMO provisional guide (1950) to instruments and methods of observation [93]. It remains fuzzy how the interval was chosen: before the conference in Washington, in August 1947, IMO commissions (including the CIMO) met in Toronto to discuss technical aspects and to define the regulations that were voted in the following conference of directors. Unfortunately, these meetings were mainly informal, with no available reports on the discussed subjects.

A possible explanation for the suggested lower bound can be formulated by considering the unit conversion from feet to meters of the Stevenson screen's height, which was the main screen used in UK after World War II. Indeed, the Stevenson screen's legs were reported to be 4 ft long [94], which is about 1.22 m. Considering the thickness of the internal wooden board, it is reasonable that the lower bound of the sampling volume was estimated to be about 1.25 m and chosen for the regulation of temperature measurements. On the other hand, the upper limit at 2 m likely was chosen for practical reasons, also considering that the air temperature at a greater height is less representative of human-life conditions.

Although the ambiguous origin of the recommendation, the potential bias introduced by different installation heights was subject of investigation by the CIMO in 1972. The WMO rapporteur W. R. Sparks in the first part of its report on "The effect of thermometer screen design on the observed temperature" [95] reviewed three works (the oldest dating back to 1922), that actually compared thermometers inside screens at different heights. Sparks concluded that the permitted height differences can produce differences in the monthly mean maximum and minimum temperature of about 0.4 °C. His final recommendation was that "a stricter definition of air temperature for routine meteorological use should be agreed. This definition should include a standard height and an averaging time for the observation".

Time have passed since Sparks' review: for some reasons his recommendations were not followed, leaving the same installation height requirement from 1947.

Nowadays modern thermometers and solar screens are employed across national weather services which can be installed on poles at any height in the suggested range, with no limitations imposed by the old wooden screens. However, experiments evaluating potential temperature differences conducted with modern instruments are missing. To bridge this lack of knowledge, an experiment was conducted to evaluate with metrological rigor the influence of the installation height on temperature measurements performed with Pt100 sensors inside identical forced ventilated screens. In the following, Section 4.1 describes the design of the experiment, the experimental site and the applied methodology for the analysis of measurements. Section 4.2 shows the results of the analysis, highlighting the impact of the installation height on climatological statistics such as the monthly average of daily maxima and minima. A discussion accompanies the results, focusing in particular on the uncertainty evaluation of the estimated temperature differences. Finally, Section 4.3 summarizes the main achievements of the experiment.

4.1 Materials and Methods

The conceived experiment for evaluating the influence of installation height, within the standardized range, is an intercomparison between Pt100 thermometers inside identical artificially ventilated solar screens. The employed installation is the very same used in [20], consisting of:

- 8 Pt100 temperature sensors connected to a customized ADC, each one mounted in a Young 43502 mechanically ventilated radiation shield. The thermometers were calibrated before the start of the experiment at INRiM laboratory, as will be explained in Section 4.1.1.
- A pyranometer, model Hukseflux LP02, measuring solar radiation. Manufacturer calibration is assumed, with a final uncertainty of less than 0.9 %.
- A sonic anemometer, model Gill WindSonic, mounted at about 2.5 m above ground. Its uncertainty has been estimated by the manufacturer as ± 1 % at 12 m/s.

- 2 hygrometers, model ETM-30 by Lombard&Marozzini, mounted inside multi-plate solar screens. Manufacturer calibration is assumed with a stated uncertainty of 0.5 %.

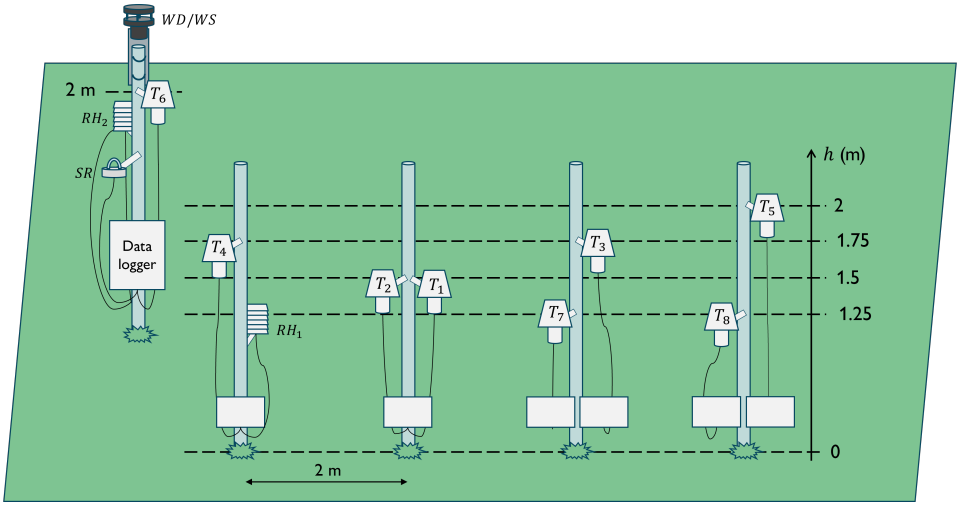


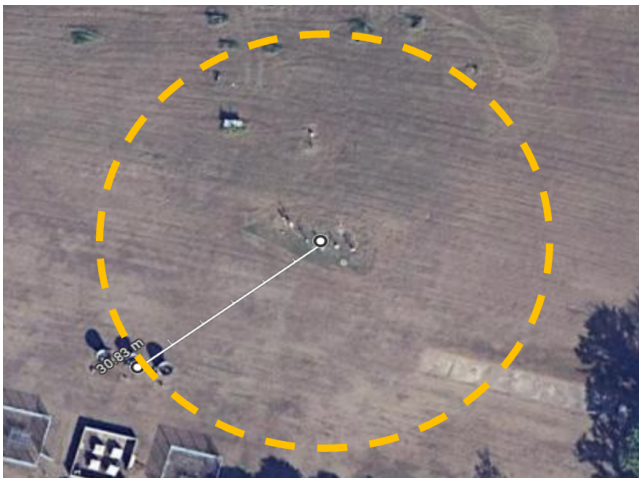
Fig. 4.1 Scheme of the experimental setup. The reported labels for each instrument/quantity are used for the analysis of measurements.

The instruments are mounted on 5 different poles, distant about 2 m from each other, and are labelled as shown in the sketch of Figure 4.1. To make the system redundant and less prone to failures, two thermometers are installed for each height level covered by the experiment, that are 1.25 m, 1.5 m, 1.75 m and 2 m. To minimize as much as possible the reciprocal influence of the screens, a maximum of two thermometers per pole were considered, orienting the screens towards south with an angle of $\approx 120^\circ$ between them. The main datalogger, reading several of the auxilliary sensors measuring the associated quantities of influence, is mounted in a dedicated pole to minimize the obstruction of wind flow due to the larger footprint of the electronic housing. All measurements are sampled simultaneously every 10 s, while once every 30 s the average of the measurements is calculated and stored into the database. Note that wind measurements are vectorially averaged.

The experimental setup was installed at INRiM campus over a grass field, which was regularly maintained, as shown in Figure 4.2a. The closest heat source (some antennas out of service) is approximately 30 m distant from the experiment (see Figure 4.2b), hence allowing to associate a WMO Class 2 to the site. Eventual errors induced by heat sources are here considered negligible since they shall affect the thermometers in the same way.

The intercomparison lasted from September 2023 to September 2025, a sufficiently long period to ensure that instruments are exposed to the greatest possible meteorological variability during the arc of the project. Regular maintenance, approximately once a month, was performed to check the correct operation of the system and to clean the screens (especially the fans) from the accumulated dirt.

(a)



(b)

Fig. 4.2 (a) Photo of the experiment. (b) Area free of obstacles and heat sources around the experiment, having a radius of 30 m.

4.1.1 Calibration

Before the start of the experiment, the thermometers were calibrated at INRiM laboratory in May 2023. The sensors were calibrated in a climatic chamber (Kambič model) at the nominal temperatures $\{-15,-5,0,5,15,25,35,40\}$ °C. The resistance of the devices under calibration (DUC) was read by a 3-wires technique using the customized datalogger employed later on field. This procedure helps to avoid the introduction of additional errors and uncertainties due to the acquisition electronics and the length of the wires. The reference temperature is given by two Pt100 thermometers, s/n 1 and 3, traceable to the ITS-90 and having an expanded calibration uncertainty ($k = 2$, 95 % coverage probabiity) of 0.01 °C for $T \geq 0\,°C$ and 0.02 °C for $T < 0\,°C$. The reference thermometers were read with the Fluke 1594A DC resistance bridge that supplies the same current, 1 mA, used during the accredited calibration of the sensors.

Fig. 4.3 Photo of the experimental setup during the calibration phase.

To minimize relative differences among thermometers during the field operation, all the temperature sensors were calibrated at the same time. The uniformity of the temperature field among the sensors was increased by using a copper comparator block, the same shown in Figure 3.3b. The reference were placed at the extremes of one of the main diagonal to estimate temperature homogeneity of the block and the corresponding average value, which was used as reference for the calculation of the calibration curve.

Due to the unavailability of the ice machine at the moment of calibration, it was not possible to make the ice-point required for calculating the Callendar-Van Dusen equation normally used for platinum resistance thermometers [96]. As alternative, a

second order polynomial equation was considered, having the form:

$$R(T) = aT^2 + bT + c, \tag{4.1}$$

which is easy to invert to obtain the estimated temperature from the measured resistance. The coefficients of this equation for each DUC are computed by a weighted least-squares regression, considering the inverse of resistance uncertainty as a weight. Table 4.1 reports the results of the calculation for each sensor.

Table 4.1 Calibration coefficients for the DUCs obtained through a weighted least-squares regression of Equation 4.1.

s/n	$a/(\Omega/^\circ\mathrm{C}^2)$	$b/(\Omega/^\circ\mathrm{C})$	c/Ω
DUC1	-5.845×10^{-5}	0.391	99.684
DUC2	-5.727×10^{-5}	0.392	99.924
DUC3	-6.039×10^{-5}	0.392	99.955
DUC4	-6.312×10^{-5}	0.392	99.939
DUC5	-7.459×10^{-5}	0.392	99.915
DUC6	-6.208×10^{-5}	0.392	99.929
DUC7	-6.255×10^{-5}	0.392	99.936
DUC8	-6.672×10^{-5}	0.392	99.906

The uncertainty budget, for each DUC and calibration point, was then evaluated by considering the following uncertainty components:

- u_{cal}: calibration uncertainty of the reference.
- u_{stab}: uncertainty related to the time stability of the reference temperature, evaluated in a conservative way as the biggest difference max-min among the two reference thermometers, weighted by a uniform distribution.
- u_{spatial}: uncertainty related to the spatial homogeneity of the temperature across the comparator block, evaluated as the maximum temperature difference between the two reference Pt100s weighted by a uniform distribution.
- u_{res}: uncertainty associated with the resolution of the data acquisition system used to measure the resistances of the DUCs. The value of $0.01\,\Omega$ is propagated using the calibration curve.

- u_{disp}: uncertainty associated with the repeated resistance measurements of the DUC and evaluated as $\sigma(R_{\text{DUC}})/\sqrt{N}$, where N is the number of measurements for the specific calibration point. The value is then propagated through the calibration curve.

- u_{interp}: uncertainty associated to the residual of the fitting procedure, which are then propagated through the calibration curve.

The expanded uncertainty budget $U_{\text{DUC}} = k \cdot u_{\text{DUC}}$ ranges between 0.4 and 0.5 °C ($k = 2$) depending on the sensor and calibration point, with only one case of 0.6 °C for DUC1 at 25 °C of nominal temperature. The following table reports an example of a detailed uncertainty budget for DUC5, which is pretty similar to those associated with the other devices. The dominant component is the calibration uncertainty of the reference thermometer, followed by the resolution of the resistance measurement system. The components relative to stability (especially for $T < 0\,°\text{C}$), spatial uniformity and interpolation have similar weights on the final budget. The evaluated uncertainty

Table 4.2 Detailed calibration uncertainty budget for DUC5. All the numerical entries of the table are expressed °C.

T_{ref}	u_{cal}	u_{stab}	u_{spatial}	u_{res}	u_{disp}	u_{interp}	u_{DUC}	U_{DUC} ($k = 2$)
−15.329	0.010	0.0044	0.003	0.007	0.0017	0.005	0.024	0.05
−5.228	0.010	0.0035	0.002	0.007	0.0018	0.001	0.026	0.05
−0.258	0.010	0.0032	0.002	0.007	0.0014	0.007	0.018	0.04
4.804	0.005	0.0011	0.002	0.007	0.0016	0.010	0.022	0.04
14.977	0.005	0.0005	0.001	0.007	0.0005	0.004	0.021	0.04
24.916	0.005	0.0004	0.001	0.007	0.0018	0.004	0.026	0.05
34.901	0.005	0.0002	0.001	0.007	0.0012	0.018	0.020	0.04
39.902	0.005	0.0002	0.001	0.007	0.0009	0.017	0.022	0.04

in laboratory conditions is generally sufficient for meteorological monitoring since bigger errors and uncertainties are introduced by the environmental factors (see Section 3.2 for an example about the influence of rain).

4.1.2 Data analysis procedure

The analysis of the performed measurements is conducted over 10 minutes. The reason is that in atmospheric science analyses the average reduces the noise affecting

air temperature measurements due to turbulence in close proximity to the ground. Furthermore, the use of artificially ventilated screens introduces additional noise which shall be smoothed out to extrapolate significant temperature differences in the vertical direction. Indeed, an additional moving average with a sliding window of 30 minutes is applied to smooth the remaining noise and identify only significant temperature variations. The temperature values calculated from this operation are the starting point for any processing step described in the following. Note that the measurements of the associated quantities of influence are also averaged, paying particular attention to the vectorial nature of the wind speed.

The intercomparison is performed by evaluating the temperature differences with respect to a reference height level, chosen to be 2 m. The use of temperature differences allows to neglect common uncertainty components of the total budget. Considering that the thermometers are read by the same measurement system, the very same used during calibration, the only components that must be kept for the system uncertainty are those related to dispersion and interpolation because they are specific for each sensor. Other components not considered during calibration are the sensor drift and the shield ageing: the former was proved in [20] to be no more than $0.01\,^{\circ}\mathrm{C/y}$, which is also compensated by relative measurements, whereas the latter was reduced by regular maintenance and, in principle, any potential error is shared by all the screens.

The uncertainties due to environmental factors are more difficult to determine: it is reasonable to assume that all the thermometer+screen systems are influenced in the same way by the external environmental conditions. Nevertheless, the different installation height changes the exposure to such factors, which will result in a different temperature measurement with respect to the reference.

The following temperature differences are evaluated for each timestamp, using the 10-minute averages, one for each installation height under study:

$$\Delta T_{1.25} = T_8 - T_6, \tag{4.2}$$

$$\Delta T_{1.50} = T_2 - T_6, \tag{4.3}$$

$$\Delta T_{1.75} = T_3 - T_6, \tag{4.4}$$

where T_8, T_2, T_3, T_6 are the temperature labels, shown in Figure 4.1, corresponding respectively to an installation height of 1.25 m, 1.5 m, 1.75 m and 2 m. The notation

ΔT_h indicates a temperature difference involving the temperature at height $h \in \{1.25, 1.50, 1.75\}$ m. When $\Delta T_h > 0\,^{\circ}$C means that the thermometer at level h is warmer than the reference at 2 m. On the other hand, if $\Delta T_h < 0\,^{\circ}$C the thermometer at h is colder compared to the reference.

The choice of the thermometers involved in Equations 4.2-4.4 is driven by the fact that during the experiment several power failures occurred, leading to missing data and irregular anomalies. For instance, after a few months, the time series of T_1 showed significant deviations from the others, hence it was identified as a fault thermometer and not considered for the analysis. Fortunately, the use of a redundant system allowed the identification of the most reliable thermometers, which are used for the data analysis. Anomalous measurements of the other sensors are removed from the time series by visual inspection, keeping only good-quality measurements.

The horizontal temperature difference at different heights across the experimental area is estimated to verify the presence of significant gradients. Spikes and outliers in the time series of the horizontal temperature differences are identified and removed because in many cases they are the result of fast transients, which are not significant for such kind of analysis. The adopted algorithm is a well-established moving window method based on the evaluation of the median and the median absolute deviation. The measurements that verify the following relation are removed from the dataset and not used for the rest of the analysis:

$$|\Delta T - \mathrm{med}_w(\Delta T)| > k \cdot 1.4826 \cdot \mathrm{mad}_w(\Delta T). \tag{4.5}$$

The notation med_w and mad_w refers to the moving median and moving median absolute deviation considering a width w for the sliding window. The factor 1.4826 is necessary to comply with the case of a gaussian distribution of data ($1.4826 \cdot \mathrm{mad}(X) \approx \sigma(X)$), whereas the coverage factor k is chosen equal to 3 to filter measurements that are outside the 99 % of the data distribution inside the window. For the horizontal gradient the width of the window is chosen equal to $24 \cdot 6 = 144$, which corresponds to one day of measurements. This choice is legitimated by the random behaviour of the horizontal gradient that does not show a strong time dependence. The good-quality estimates of the horizontal gradient are used to evaluate an additional uncertainty component that takes into account the spatial representativeness of the air temperature measurements.

A similar filtering procedure is also performed for the time series of ΔT_h, the quantity of interest, but using a different width of the moving window. Indeed the vertical gradient strongly depends on the hour of the day, implying that the window must represent a fration of the day in order to have a better estimate of the median value to use as reference. The chosen window in this case is $6 \cdot 8 = 48$, corresponding to one third of the day.

The last preprocessing step before analyzing the ΔT_h is the classification of measurements based on the stability conditions of the surface layer, that is the portion of the boundary layer (BL) in contact with the surface. The reason for such a classification is that the vertical gradient varies with the diurnal cycle, being positive for stable conditions (stratification of air layers during nighttime) and negative during unstable conditions (ground warmer than air during daytime). For making such a division, measurements can be grouped by using the solar radiation: daytime data are for $SR > 0\,\mathrm{W/m^2}$, whereas nighttime values are for null radiation. Even though the correspondence between the time of the day and stability conditions is not always true, for weather and climate monitoring it is more useful to adopt this classification since it is generally not possible to determine the stability conditions at the weather station sites: in absence of more thermometers at various heights (rarely installed by national weather services), the determination of atmospheric stability requires the use of sophisticated instruments such as the tri-axial sonic anemometers, which is seldom available. Therefore, by following the described subdivision, the results of the analysis will be much easier to apply to other weather stations.

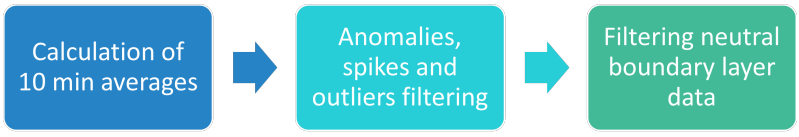


Fig. 4.4 Flow chart of the preprocessing steps before analyzing the results.

The surface layer can be observed in an additional condition, the neutral state, typically appearing when clouds cover a large part of the sky (overcast). Under this state, the vertical temperature profile is approximately constant, winds are moderate to strong and there is little heating or cooling from the surface [97]. Due to its nature, the state can appear both in daytime and nighttime, hence contaminating both the states described above. Since the neutral condition is the least significant for the investigation of the vertical temperature differences, an algorithm is developed to

identify and remove the corresponding measurements. The filtering is based on the computation of a running standard deviation over the temperature differences, which is smaller than a certain threshold δt for the neutral condition:

$$\sigma_w(\Delta T_h) < \delta t. \tag{4.6}$$

The symbol σ_w indicates the moving standard deviation computed for a certain window w defined by the number of data points inside it. The reason behind the chosen filter is that the temperature differences in neutral conditions are less dispersed within a certain time interval. The parameters of the filters are tuned on the data, finding optimal values of $w = 15$ (2 and a half hours) and $\delta t = 0.15\,^{\circ}\mathrm{C}$. The application of a filter of this kind is mandatory when treating large datasets, as in this case, since a manual classification would be unfeasible. Nevertheless, the approach is empirical and not perfect, allowing possible misclassifications. As it will be described in detail in Section 4.2, the risk is compensated by working with probability distributions for days and nights, obtained by a number of measurements that is larger compared to that of data classified in the neutral state.

Eventually, after the preprocessing steps summarized in Figure 4.4, the ΔT_h are analyzed by calculating the main moments, mean and variance, and representative values, such as the median and the mode, of the measurement distributions for day and nighttime. The behaviors of daily minima and maxima at different heights are also compared since they are crucial quantities for climatological studies. The uncertainty of the best estimate $\overline{\Delta T_h}$, for each season and stability condition, is evaluated by applying the following model (the use of relative measurements compensates some components as described above):

$$u^2(\overline{\Delta T_h}) = u^2_{\mathrm{interp}} + u^2_{\mathrm{disp}} + u^2_{\mathrm{site}} + u^2_{\mathrm{distr}}(\Delta T_h), \tag{4.7}$$

where:

- u_{interp} is the calibration uncertainty contribution due to interpolation, evaluated as the quadrature sum of the averages of residuals (among the calibration points) for each thermometer involved in ΔT_h.
- u_{disp} is the component corresponding to the quadrature sum of the average uncertainties (among the calibration points) due to repeated measurement for each thermometer involved in ΔT_h.

- u_{site} is the uncertainty associated with the representativeness of measurement (measurand uncertainty) within the experimental area. It is evaluated by taking the standard deviation of the temperature difference distribution in the horizontal direction.

- $u_{\text{distr}}(\Delta T_h)$ is the uncertainty component that considers, for each height level, the dispersion of the ΔT_h values around the best estimate, which is calculated from their distribution. This variability depends on the different exposure of the instrument, but also on the atmospheric conditions in different daily cycles.

The quantitative results of the preprocessing steps, the analysis of the intercomparison and the associated uncertainty analysis are reported in details in the following section.

4.2 Results and Discussion

The dataset of 10-minute averages consists of 88541 valid instances, interspersed with missing data intervals, clearly visible in Figure 4.5. Several power interruptions, related to failure events of the local electrical grid, led to many missing intervals (May 2024, July 2024, January 2025, July 2025) shorter than or equal to one week. Instead, the two main interruptions (about one month) are in August 2024 and April 2025: the first caused by an extreme meteorological event above INRiM campus, which interrupted the power supply during the vacation period, whereas the second was due to the cutting of some cables, during the mowing of the grass on the field by an external company, whose replacement took some time. The coverage of two years ensures that at least 30 days of measurements are available for each month, hence compensating for the missing periods.

The temperature time series of T_3 in Figure 4.5 has a minimum value of $-5.46\,^{\circ}\text{C}$ and a maximum one of $36.21\,^{\circ}\text{C}$, which are within the calibration interval, hence no values must be discarded. Concerning the associated quantities of influence, no anomalous behavior is observed. The differences between the measurements of relative humidity are because RH_1 is sampled at 1.25 m, hence being more influenced by the ground. This difference is also highlighted by the distribution of measurements shown in Figure 4.6 in the upper-left panel: the relative frequency of relative humidity increases with the magnitude, with RH_1 having a peak around 100 %rh, whereas RH_2 has a double peak at about 78 and 90 %rh. The other panels show the distribution of

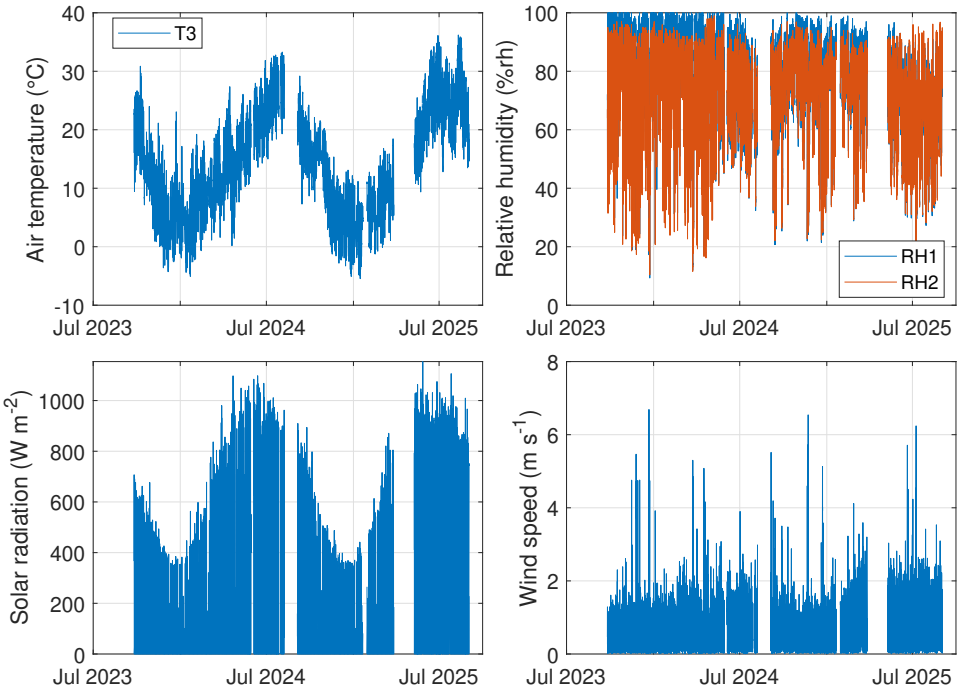


Fig. 4.5 Time series of the quantities measured by the experimental setup.

solar radiation and wind speed. The most populous bins for solar radiation are close to zero, which reflects the solar diurnal cycle. The distribution of wind speed and orientation is shown in the bottom panels. The mode of the wind speed distribution is 0.3 m/s, whereas the mean and the median are 0.69 m/s and 0.58 m/s respectively. The experimental site is in general characterized by low wind speeds: 78.7 % of measurements are between 0 and 1 m/s, about 19 % between 1 and 2 m/s and only the 2.3 % is above 2 m/s, corresponding to very few occasional events. To better understand the distribution of direction angles, it is useful to plot the measurements in a wind rose plot as in Figure 4.7. Some "blind angles" with a smaller relative frequency are visible between 80 ° and 90 ° and between 170 ° and 180 °, likely due to the presence of buildings obstructing the wind flowing from those directions. The wind blew more frequently from 90 ° to 110 °, with a wind speed smaller than 2 m/s. Only between 260 ° and 270 ° the wind speed reached values greater than 2 m/s, corresponding to fhön winds blowing from Susa Valley towards the city of Turin.

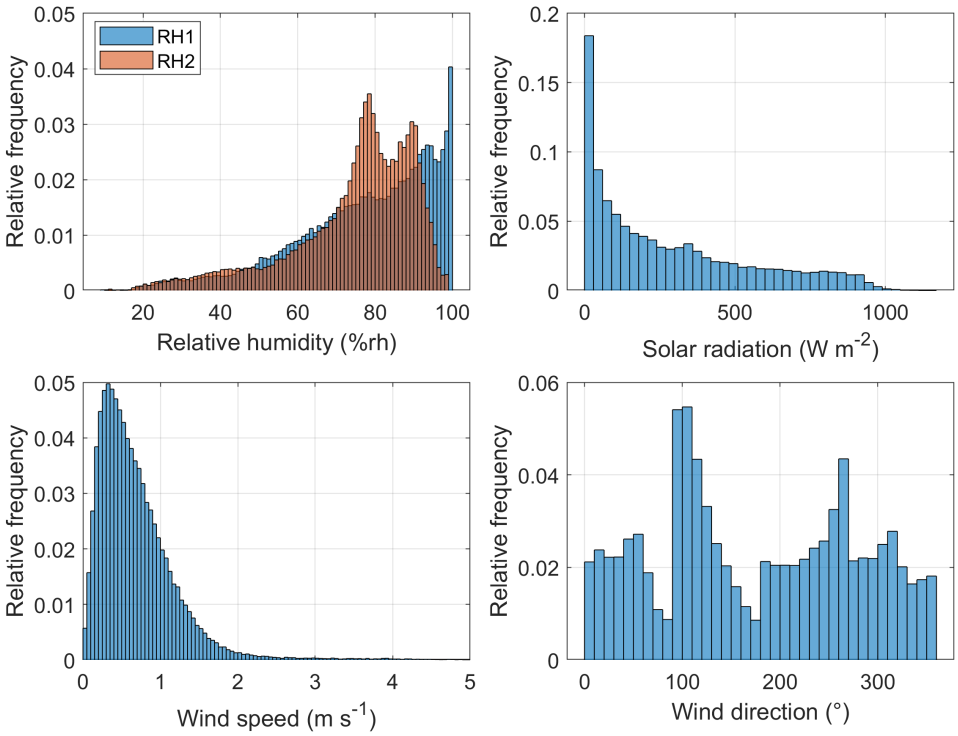


Fig. 4.6 Histograms of the associated quantities of influences measured by the experimental setup.

4.2.1 Preprocessing

After the visual inspection, the horizontal gradient across the experimental site is investigated. Ideally, since the representativeness of measurements varies with the installation height, a comparison between the thermometers installed at 1.75 m is considered for the analysis. The temperature difference $\Delta T_{\text{site}} = T_4 - T_3$ is calculated and used for the spike and outliers filtering phase as described in Section 4.1. The filtered data ($\sim$ 2303 instances), plotted as red dots on the time series in Figure 4.8a, are removed from the dataset and not considered in the following. Using the remaining measurements, the distribution of ΔT_{site} is evaluated and plotted in panel b of Figure 4.8. The empirical mean, median and standard deviation are equal to $-0.021\,^{\circ}\text{C}$, $-0.022\,^{\circ}\text{C}$ and $0.041\,^{\circ}\text{C}$ respectively. The estimated mean value is smaller than the calibration uncertainty, proving the good representativeness within the experimental area. Furthermore, the ΔT_{site} are evenly distributed across the mean, as proved by its (almost) coincidence with the median. For these reasons, it is decided not to correct the measurements for eventual horizontal gradient, but to

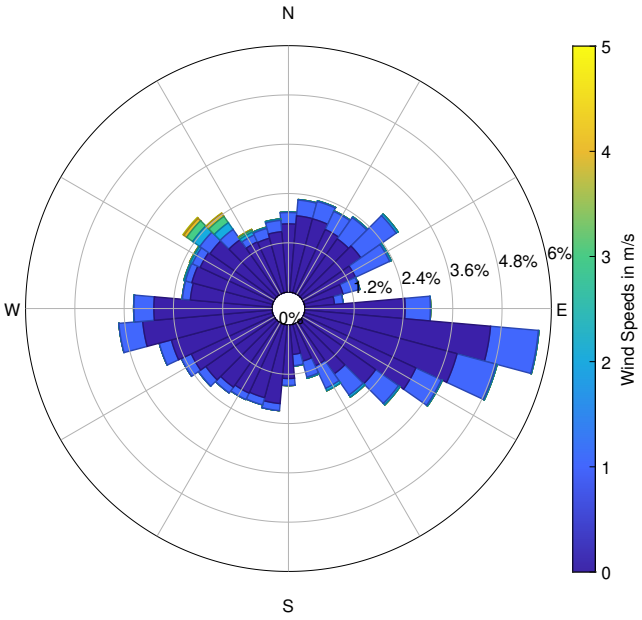


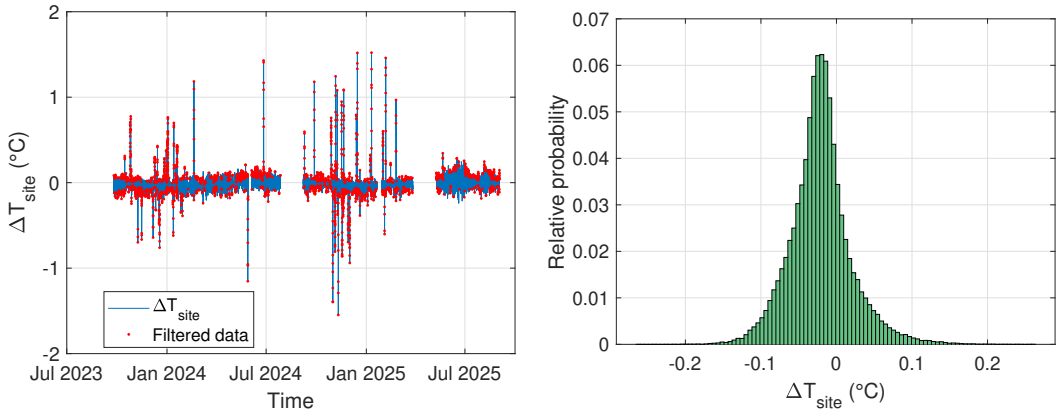
Fig. 4.7 Wind rose plot of the measurements sampled at the experimental site.

consider the dispersion as an additional uncertainty $u_{\text{site}} = \sigma(\Delta T_{\text{site}})$, used later to evaluate Equation 4.7.

A similar filtering procedure is then applied to the ΔT_h for removing spikes and outliers. Eventually, the size of the dataset for ΔT_h is 75430 ($h = 1.25\,\text{m}$), 72324 ($h = 1.50\,\text{m}$) and 75749 ($h = 1.75\,\text{m}$). The size slightly changes with h due to the empirical nature of the filter, which is not capable of removing exactly the same measurements at each height. The thermometers also presented some independent anomalous behaviors lasting a couple of hours (likely induced by a reduction of fan speed) which were removed from the time series. The results of the filtering process are reported in Figure 4.9 by taking as an example the time series of $\Delta T_{1.75}$.

The distributions of ΔT_h obtained from the good-quality measurements are visualized in Figure 4.10. All the histograms have a similar shape but at different scales, with a main peak close to zero broadened towards negative values. This shape is the result of the daily cycle affecting the vertical gradient, with a broadening towards negative values mainly due to measurements associated with nighttime.

The ΔT_h corresponding to neutral conditions also partially contributes to the peak, not allowing the estimation of a non-biased average value corresponding to only clear sky or partially cloudy days. This aspect drove the implementation of an empirical classification of the boundary layer based on Equation 4.6. Figure 4.11



(a) (b)

Fig. 4.8 (a) Filtering of ΔT_{site} from spikes and outliers as described in Section 4.1. (b) Distribution of ΔT_{site} evaluated with non-filtered measurements.

reports an example of the application of the criterion on the $\Delta T_{1.75}$ time series. The black dots represent the measurements associated with the neutral boundary layer, whereas the red and blue dots correspond to days ($SR > 0\,\text{W/m}^2$) and nights. The criterion works relatively well for classifying the measurements between 09/02/2024 and 11/02/2024, during which the vertical gradient was approximately null (rainfall event). Similar results can be found in the rest of the time series, even if some misclassifications may happen. In the worst case, measurements associated with the neutral BL remain in the daytime and nighttime classes, but they have a smaller weight by working with probability distributions. An example is reported in Figure 4.12, which shows the distributions for $\Delta T_{1.75}$ within each class. The smooth peak of the histogram for nighttime temperature differences is related to the broadening mentioned above, whereas the peak of daytime measurements contributes to the main peak visible before the classification (see Figure 4.10). The measurements classified in the neutral BL are also localized around zero, with a very small dispersion of about 0.03 °C, which proves that they truly belong to a BL state with approximately null vertical gradient. Similar considerations can be made for the distributions of $\Delta T_{1.25}$ and $\Delta T_{1.50}$, which are even more visibly affected by nighttime measurements. In the end, after the preprocessing steps, the number of measurements in each class for the considered ΔT_h is reported in Table 4.3. Despite some differences related to spikes and anomalies filtering, the number of data for the ΔT_h in each class is similar, allowing a fair comparison between them.

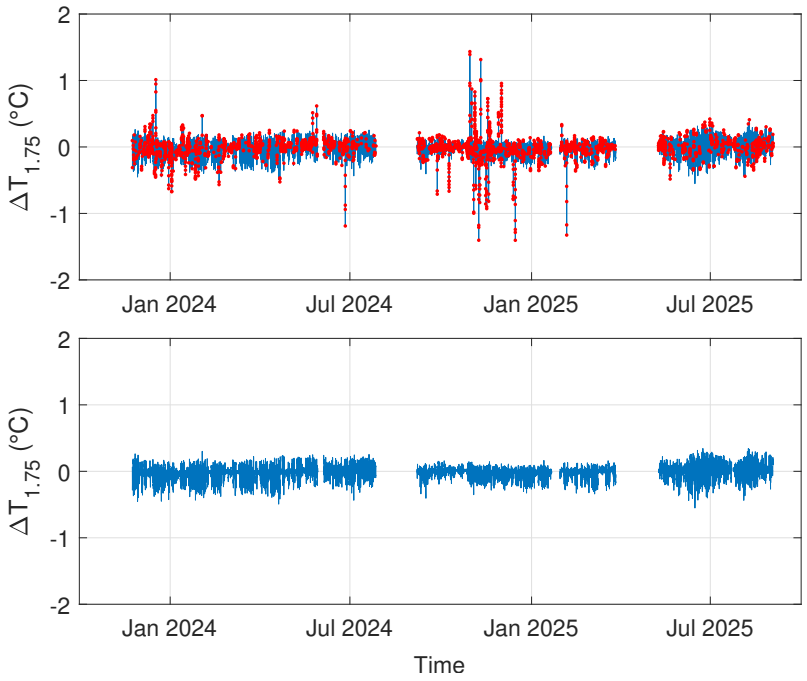


Fig. 4.9 Example of the filtering process applied to the vertical temperature difference $\Delta T_{1.75}$. The complete time series is shown in the upper panel, with the red dots corresponding to the filtered data. The bottom panel shows instead the time series after the removal of spikes and outliers.

4.2.2 Analysis of daytime and nighttime measurements

After the classification, the estimated ΔT_h are analyzed with respect to their own class (nighttime, daytime and neutral BL). To make a better visual comparison between the distributions of the ΔT_h within the same class, a kernel density estimation is performed. The method allows for the estimation of a Probability Density Function (PDF) with moments and quantiles that are almost coincident with those empirically estimated (differences in the order of mK may appear, which are irrelevant for the present study). A comparison of the estimated PDF is then done, using all the available measurements, and reported in Figure 4.13, whereas Table 4.4 lists the quantitative estimation of the main statistical quantities.

At nighttime, the distribution of ΔT_h appears to have a longer left tail and shifted towards smaller values for decreasing h, which is physically reasonable. Indeed, temperature measurements in forced-ventilated screens are the result of the (partial) integration of the temperatures characterizing the air column below them: at the lowest height the screen sucks air that is colder because closer to the ground. Further-

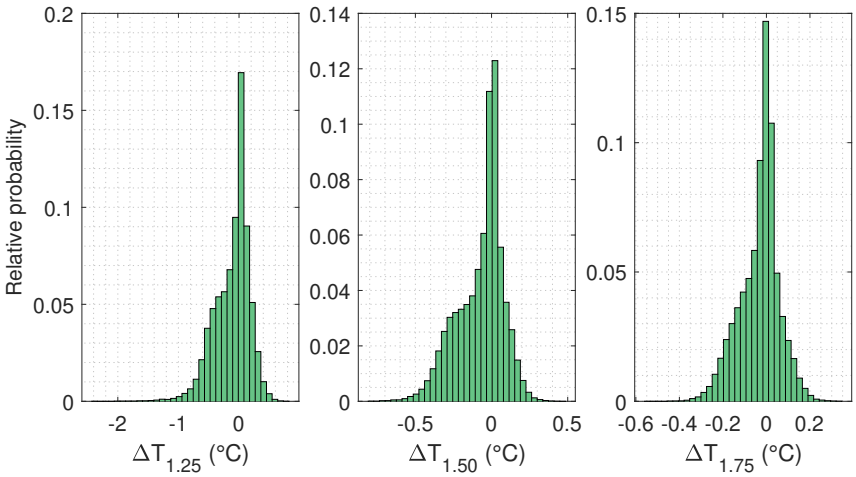


Fig. 4.10 Distribution of ΔT_h after the filtering process.

Table 4.3 Size of the dataset for each class.

	N_{nights}	N_{days}	N_{neutral}
$\Delta T_{1.25}$	28169	29446	17815
$\Delta T_{1.50}$	27128	27599	17597
$\Delta T_{1.75}$	28120	29478	18061

more, water may condense on surfaces during nights and be aspirated into the screen, cooling the thermometer and eventually introducing a bias in the measurements which can reach $-1\,^{\circ}\text{C}$. This phenomenon may happen at every height, but it is more probable closer to the ground due to colder air layers. In quantitative terms, the average value $\overline{\Delta T_h}$ changes from $-0.36\,^{\circ}\text{C}$ at 1.25 m to $-0.11\,^{\circ}\text{C}$ at 1.75 m, whereas the standard deviation $\sigma(\Delta T_h)$ decreases from $0.25\,^{\circ}\text{C}$ to $0.08\,^{\circ}\text{C}$. The behavior of the dispersion can be explained by the fact that the ground radiative cooling and condensation phenomena vary for different nights, which is reflected in a different bias. Thermometers closer to the ground are more sensitive to this variability, hence the dispersion of measurements is larger.

A different behavior is visible for the daytime PDF, appearing more symmetric around the mean for $h \in [1.50, 1.75]$ m, whereas for $h = 1.25$ m a longer left tail is still present. In all cases, the values range approximately from $-0.5\,^{\circ}\text{C}$ to $0.5\,^{\circ}\text{C}$. Negative values of vertical temperature difference, normally associated with stable conditions, are present since the choice of dividing the measurements based on solar radiation. At sunrise the solar radiation starts to be different from zero, but the

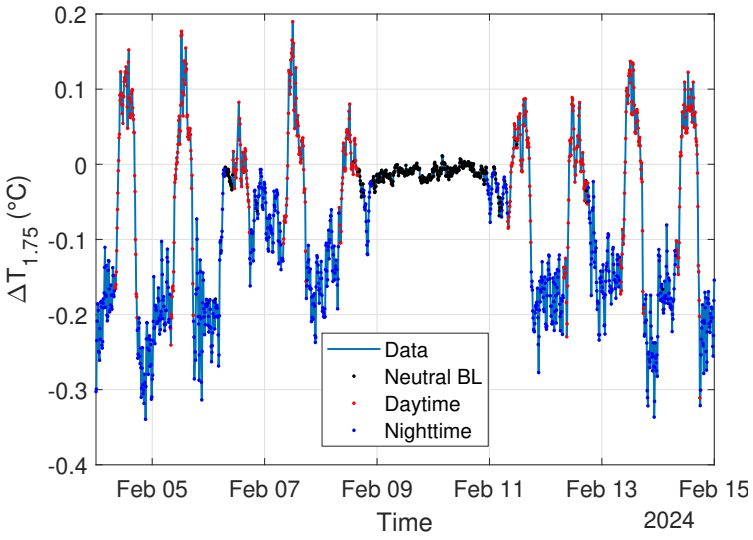


Fig. 4.11 Example of classification of the boundary layer state for a portion of the $\Delta T_{1.75}$ time series. Red dots correspond to temperature differences during daytime, blue dots during nighttime, whereas the black dots are associated with measurements characterized by a neutral boundary layer.

temperature difference is still negative due to the stable surface layer. After some time, the ground starts to warm, generating turbulence that disrupts the atmospheric stability and inverts the sign of the temperature profile. As for nighttime, the dispersion of measurements is bigger for smaller installation heights due to greater influence from the ground and increased inhomogeneity. The mean of the distribution also shifts towards larger values for decreasing h since at daytime the air closer to the ground is warmer. Quantitatively, the main peak (the most frequently measured value) moves from 0.13 °C at $h = 1.25$ m, to 0.04 °C at $h = 1.50$ m and then to 0.02 °C at $h = 1.75$ m. The average values are slightly different from the modes, but still are about one order of magnitude smaller than nighttime estimates. Accidentally, the calculated dispersion during the day is comparable with that at nighttime for each height level.

Even if less relevant since by definition it is the best condition to compare the thermometers, the measurements belonging to the neutral BL class are also analyzed. Similar considerations can be done about the dispersion of values, even if the variation is much less pronounced. The mean values range from 0.008 °C at 1.25 m to −0.005 °C at 1.50 m, with a difference that is included in the uncertainty of the site, thus not being significant. These results prove again that the employed empirical classification works, with a certain dispersion around the averages, as

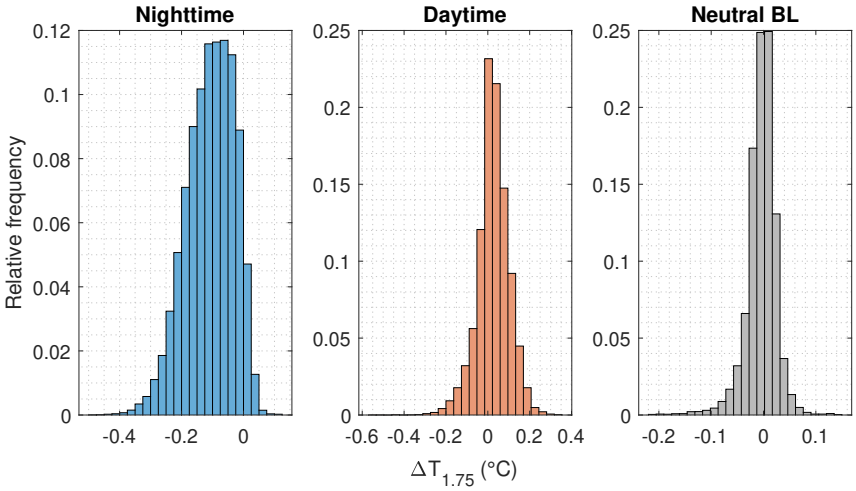


Fig. 4.12 Histograms of $\Delta T_{1.75}$ for each class.

shown by the PDF, which is due to the natural variability of the system and to the possible misclassifications.

Table 4.4 Estimation for each class of the main statistical quantities: mean, median, mode and standard deviation.

Class	h/m	$\overline{\Delta T_h}$/°C	med(ΔT_h)/°C	mode(ΔT_h)/°C	$\sigma(\Delta T_h)$/°C
Nighttime	1.25	−0.363	−0.341	−0.336	0.252
	1.50	−0.204	−0.198	−0.251	0.130
	1.75	−0.107	−0.101	−0.078	0.078
Daytime	1.25	0.058	0.092	0.130	0.231
	1.50	0.021	0.033	0.036	0.119
	1.75	0.026	0.026	0.016	0.076
Neutral BL	1.25	0.008	0.023	0.031	0.091
	1.50	−0.005	0.003	0.006	0.049
	1.75	−0.004	−0.002	0.009	0.029

A complete uncertainty analysis is performed to evaluate the standard uncertainty associated with the mean values $\overline{\Delta T_h}$, which are representative of the measured temperature difference over the entire arc of the experiment. The uncertainty components of the budget, described above and entering in Equation 4.7, are reported in Table 4.5. The combined effect of the instrumental uncertainties, u_{interp} and u_{disp}, with the measurand uncertainty u_{site}, is about 0.04 °C, which is one order of magnitude smaller than the uncertainty resulting from the weather variability u_{dist}. As one would

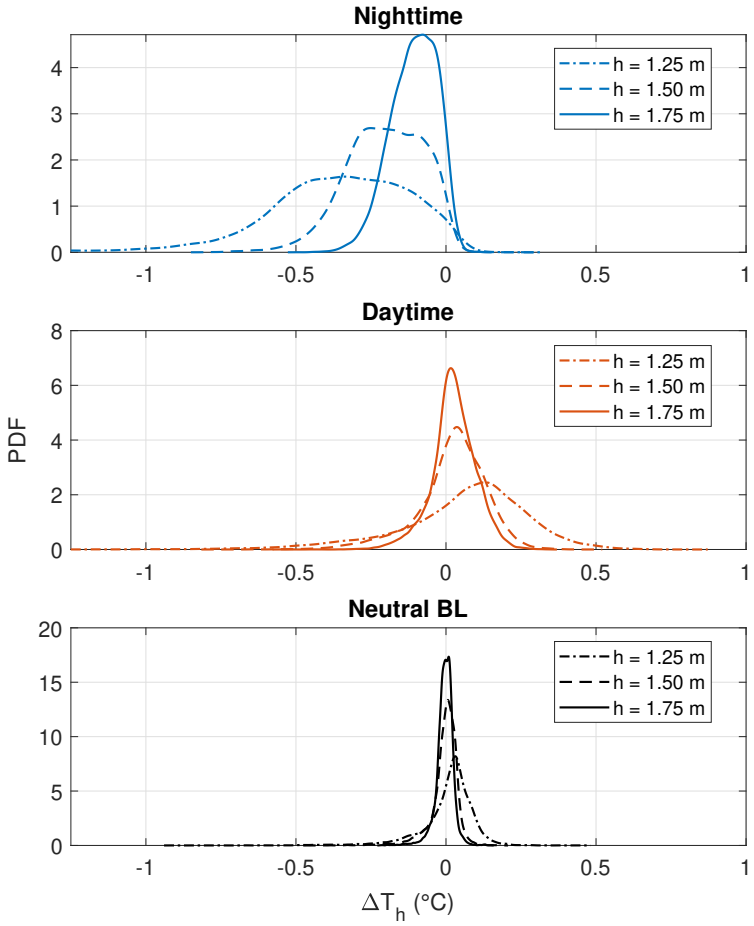
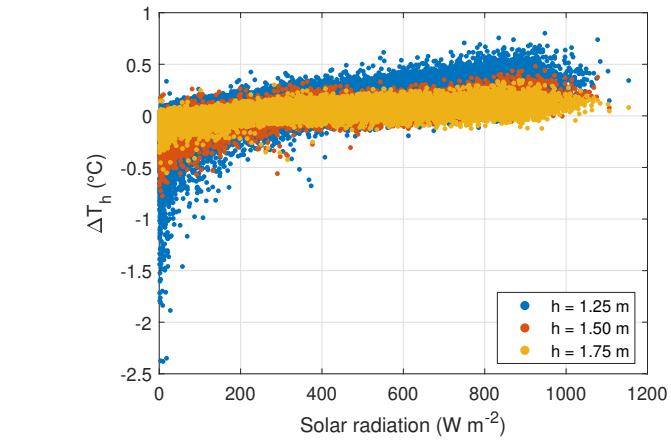


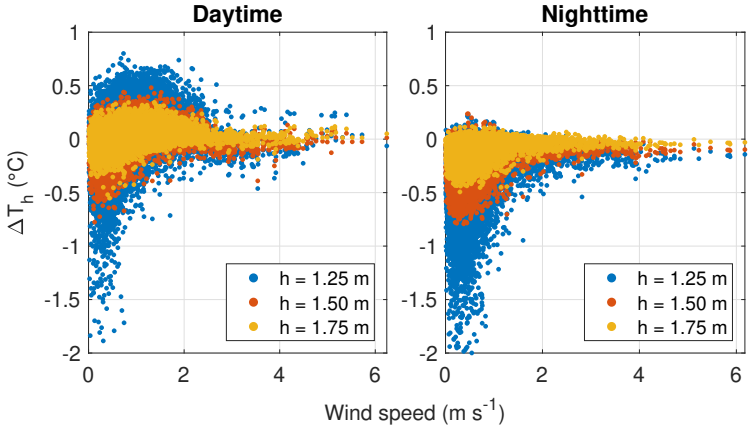
Fig. 4.13 Comparison of the Probability Density Functions (PDFs) of $\overline{\Delta T_h}$ for each height level and analyzed class (daytime, nighttime, neutral BL).

expect, the overall uncertainty is almost coincident with u_{dist}, that is evaluated as the standard deviation of the PDFs commented above. The uncertainty at nighttime and daytime are comparable for each height level, with values at 1.50 m and 1.75 m that differ by 0.04 °C. The uncertainties for $h = 1.25$ m are the biggest because of the greater dispersion of measurements.

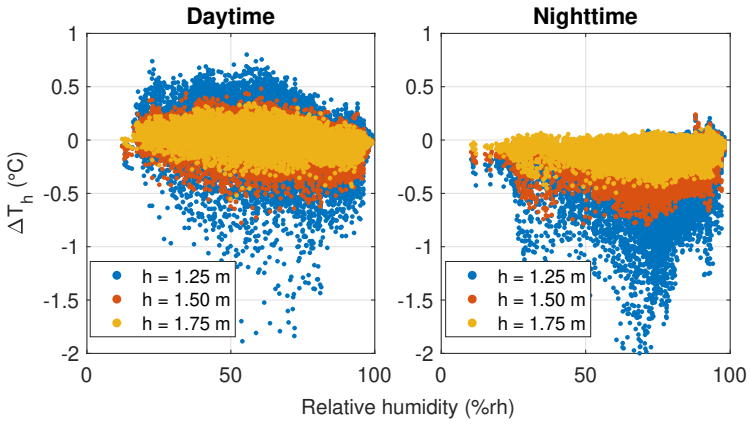
The relation with the associated quantities of influence is also analyzed to understand their influence on the measurements. The scatter plots in Figure 4.14 are a good tool to investigate this aspect. Starting from the incoming solar radiation, it is visible a non-linear trend between ΔT_h and SR, which varies with the installation height. The possible radiative error affecting the measurements is strongly reduced by the forced ventilation. Even if present, the error would similarly affect all the thermometers. The dependence on SR shown by the plot is an "indirect" relation:



(a)



(b)



(c)

Fig. 4.14 Scatter plots of the temperature differences ΔT_h with respect to solar radiation (a), wind speed (b) and relative humidity (c).

Table 4.5 Uncertainty budget associated to $\overline{\Delta T_h}$ for each analyzed class.

Class	h/m	$\overline{\Delta T_h}/^\circ\mathrm{C}$	$u_{\mathrm{interp}}/^\circ\mathrm{C}$	$u_{\mathrm{disp}}/^\circ\mathrm{C}$	$u_{\mathrm{site}}/^\circ\mathrm{C}$	$u_{\mathrm{dist}}/^\circ\mathrm{C}$	$u(\overline{\Delta T_h})/^\circ\mathrm{C}$
Nighttime	1.25	−0.363	0.006	0.002	0.041	0.252	0.25
	1.50	−0.204	0.005	0.002	0.041	0.130	0.14
	1.75	−0.107	0.005	0.002	0.041	0.078	0.09
Daytime	1.25	0.058	0.006	0.002	0.041	0.231	0.23
	1.50	0.021	0.005	0.002	0.041	0.119	0.13
	1.75	0.026	0.005	0.002	0.041	0.076	0.09
Neutral BL	1.25	0.008	0.006	0.002	0.041	0.091	0.10
	1.50	−0.005	0.005	0.002	0.041	0.049	0.06
	1.75	−0.004	0.005	0.002	0.041	0.029	0.05

larger irradiating power corresponds to a warmer temperature of the ground, hence also a warmer air close to it. This is the reason why, on average, an increasing solar radiation corresponds to a stronger vertical temperature gradient. In the following, it will be shown how this relation can be exploited to estimate a temperature correction for daytime measurements.

The wind speed has the opposite influence on ΔT_h: an increase in the wind speed induces the creation and enhancement of eddy vortices by shear stress. The greater turbulence permits a better mixture of air close to the ground, hence flattening the vertical temperature profile. This phenomenon is visible in Figure 4.14b: for wind speed greater than $2\,\mathrm{m/s}$ the temperature differences become closer to zero, with a reduced dispersion of the distribution. This behavior does not depend on the time of day, as it is clearly visible for both daytime and nighttime measurements.

The relation with relative humidity, plotted in panel (c) using RH_2, is instead less straightforward. The response of ΔT_h to a changing relative humidity appears to be flat. The dispersion of daytime values does not depend on the magnitude of relative humidity, whereas for nighttime, especially at $h = 1.25\,\mathrm{m}$, the dispersion slightly increases, likely due to the condensation of water on surfaces as described above.

The results here reported clearly show that, on average, a non-negligible difference affects the measurements of forced ventilated thermometers installed at different heights within the range recommended by the WMO. Especially significant are the average errors during nighttime, reaching a value of $-0.36\,^\circ\mathrm{C}$ at $1.25\,\mathrm{m}$, whereas those related to daytime are smaller, lower than $0.1\,^\circ\mathrm{C}$. Since the measurements are

affected by several sources of dispersion, especially due to weather variability, the standard uncertainty associated with the average estimates is evaluated. The interval $\overline{\Delta T_h} \pm u(\overline{\Delta T_h})$ covers $\approx$70 % of nighttime measurements and $\approx$75 % of daytime measurements. For a precise definition of the coverage factor for a given probability it is necessary to use the empirical PDF since closed-form distributions are not so easy to apply to the data.

4.2.3 Seasonal analysis

It is relevant to verify if there is a seasonal variation in the vertical temperature difference. The measurements are therefore divided by season and analyzed as in the previous section. Two other quantities are estimated, which are the seasonal average of the daily maximum and minimum temperature difference, indicated by $\max(\Delta T_h)$ and $\min(\Delta T_h)$. These values give an indication of an average "worst case scenario" during the day, which is useful to understand the maximum and minimum errors that can be made. An example of this additional evaluation is reported in Figure 4.15, considering all the good-quality measurements together with no seasonal division. As expected, the $\max(\Delta T_h)$ and $\min(\Delta T_h)$ follow the same trend of $\overline{\Delta T_h}$ for days and nights, but with a different rate of variation. The average maximum error ranges, for decreasing values of height, from 0.12 °C to 0.29 °C, whereas the average minimum error is in the range -0.20 °C to -0.62 °C. The uncertainties corresponding to these estimates (evaluated as above using Equation 4.7), as well as the numerical values for each class and season, are reported in Table 4.6.

The average temperature profiles for each season are plotted in Figure 4.16, whereas the seasonal variation for each quantity is visualized in Figure 4.17. Concerning daytime measurements, the temperature difference slightly varies between winter and summer, with an increase of 0.03 °C for 1.50 and 1.75 m. The temperature difference at 1.25 m presents a greater increase, almost reaching 0.1 °C during spring and summer. Its variation is much more visible for the average daily maximum, which becomes 0.4 °C in summer. Similar considerations can be done for nighttime data: the greatest variation is seen again for $h = 1.25$ m, with an average daily minimum changing from -0.51 °C in autumn to -0.77 °C in summer. A possible explanation for these variations could be attributed to a greater growth rate of grass on the field during spring and summer. Despite the effort of periodically cutting the grass, it is difficult to maintain its height always the same, especially during the

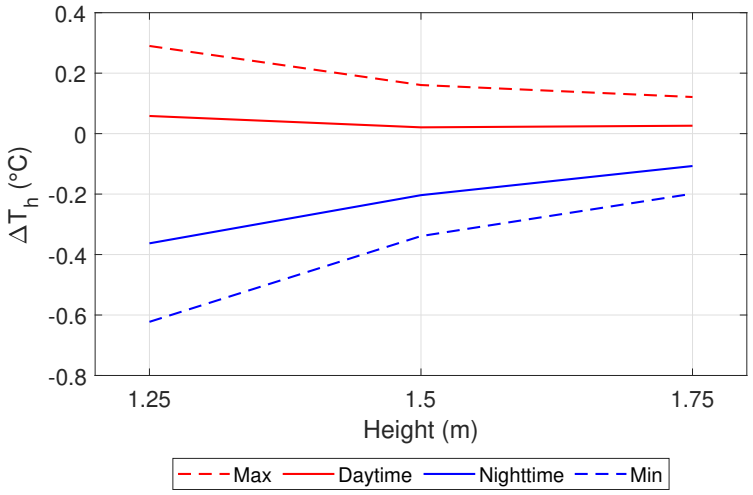


Fig. 4.15 Plot of the mean temperature differences for daytime and nighttime measurements (solid red and blue lines) as a function of installation heights. The average daily maximum and minimum of ΔT_h are also reported as a reference for a worst-case scenario (dashed red and blue lines).

hotter months. The variability in the grass height may have changed the effective installation height, thus enhancing the influence from the ground, especially for the thermometer at 1.25 m. Although this aspect has a major influence on maxima and minima, the average value is less influenced and gives a quantitative and robust estimation for the installation height effect, which seems not to be strongly dependent on the season.

4.2.4 Climatological analysis

The observed temperature differences may affect some quantities generally estimated by climatologists, such as the monthly mean maximum and minimum temperatures, as also appointed by Sparks in [95]. Due to the missing sampling intervals, it is possible to estimate monthly average quantities with a complete time series only between December 2023 and February 2025, without considering August 2024 because of the problems after the extreme localized meteorological event that hit the INRiM campus. In mathematical terms, the following quantities are estimated:

$$\Delta T_{\max}^{(h)} = \overline{T_{\max}^{(h)}} - \overline{T_{\max}^{(\mathrm{REF})}}, \tag{4.8}$$

$$\Delta T_{\min}^{(h)} = \overline{T_{\min}^{(h)}} - \overline{T_{\min}^{(\mathrm{REF})}}. \tag{4.9}$$

Table 4.6 Results of the seasonal analysis with corresponding standard uncertainties reported in round brackets. The superscript "d" stands for daytime, whereas "n" is related to nighttime measurements.

Season	h/m	$\overline{\Delta T_h^{(\mathrm{n})}}$/°C	$\overline{\min(\Delta T_h)}$/°C	$\overline{\Delta T_h^{(\mathrm{d})}}$/°C	$\overline{\max(\Delta T_h)}$/°C
All	1.25	−0.36(0.25)	−0.62(0.37)	0.06(0.23)	0.29(0.16)
	1.50	−0.20(0.14)	−0.34(0.18)	0.02(0.13)	0.16(0.10)
	1.75	−0.11(0.09)	−0.20(0.11)	0.03(0.09)	0.12(0.09)
SON	1.25	−0.31(0.20)	−0.51(0.28)	0.02(0.18)	0.21(0.10)
	1.50	−0.17(0.12)	−0.28(0.17)	0.01(0.11)	0.14(0.07)
	1.75	−0.08(0.08)	−0.16(0.11)	0.01(0.07)	0.09(0.07)
DJF	1.25	−0.34(0.21)	−0.57(0.25)	0.03(0.19)	0.21(0.11)
	1.50	−0.22(0.13)	−0.36(0.16)	0.01(0.13)	0.12(0.09)
	1.75	−0.12(0.09)	−0.21(0.11)	0.01(0.08)	0.08(0.07)
MAM	1.25	−0.39(0.25)	−0.61(0.39)	0.08(0.23)	0.32(0.16)
	1.50	−0.22(0.14)	−0.33(0.21)	0.02(0.13)	0.16(0.10)
	1.75	−0.12(0.09)	−0.20(0.13)	0.03(0.09)	0.12(0.08)
JJA	1.25	−0.42(0.34)	−0.77(0.49)	0.07(0.27)	0.41(0.16)
	1.50	−0.18(0.15)	−0.35(0.20)	0.04(0.13)	0.23(0.10)
	1.75	−0.09(0.09)	−0.20(0.11)	0.04(0.09)	0.19(0.08)

The overline stands here for the monthly average of daily maximum and minimum temperatures, which is applied to the thermometers at the investigated height levels and to the reference sensor at 2 m. Since the monthly mean is applied to each term and assuming that the dispersion of the daily $T_{\max}$ around the mean value is comparable for each height, the uncertainty associated with such quantities is mainly due to the measurement system and the inhomogeneities of the site, hence equal to 0.04 °C. The quantitative evaluation of $\Delta T_{\max}^{(h)}$ and $\Delta T_{\min}^{(h)}$ for each month is shown in Figure 4.18. A bias is detected in both maximum and minimum temperatures. During warmer months the bias increases for maxima, whereas in colder ones increases (in magnitude) for the minima. The values for February 2025 does not follow the yearly-trend and may be reasonably considered an outlier for this analysis. The differences for 1.50 and 1.75 m are smaller than 0.1 °C for maxima and smaller (in magnitude) than −0.26 °C for minima. The bias at 1.25 m is enhanced, reaching the largest value of 0.35 °C in July 2024 for maxima and −0.5 °C for minima in April 2024. These results are coherent with what was found by the few experiments

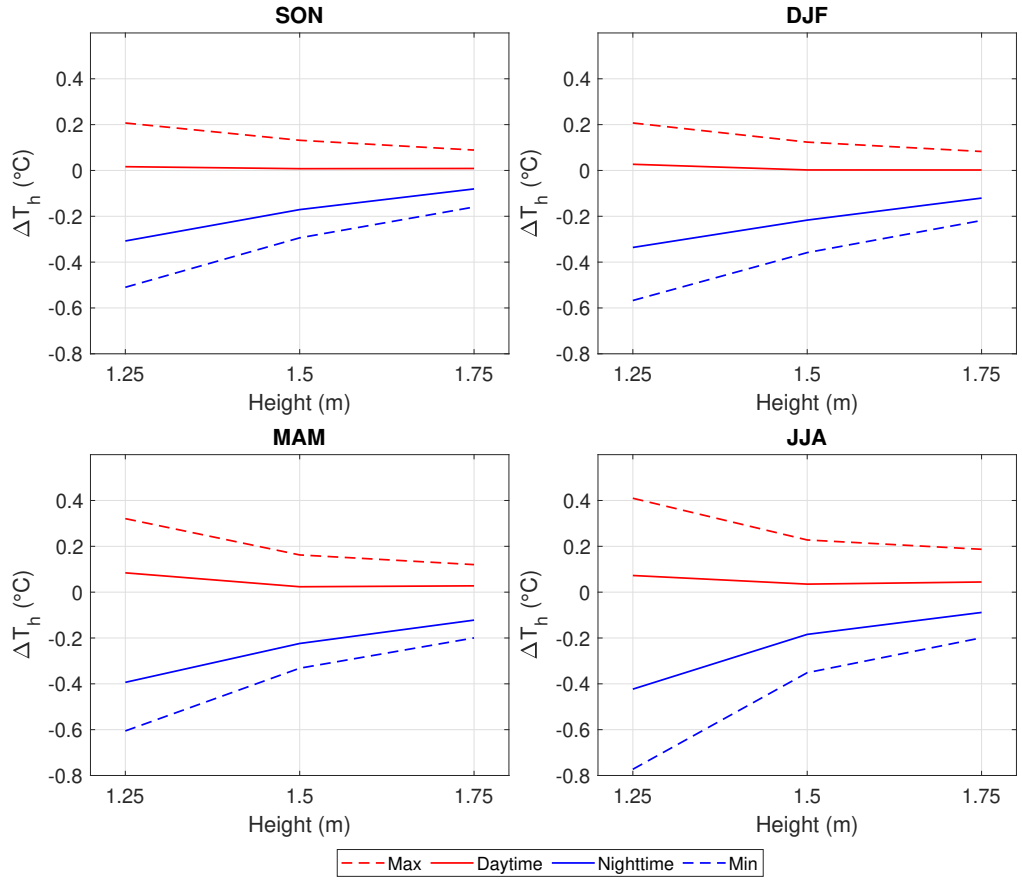


Fig. 4.16 Plots by season of the mean temperature differences for daytime and nighttime measurements (solid red and blue lines) as a function of installation height. The average daily maximum and minimum of ΔT_h are also reported as a reference for a worst-case scenario (dashed red and blue lines).

performed at the beginning of the last century [95], even if different instrumentation and shield technology were used. If not properly tackled, these differences can enter into the processing of data and affect the accuracy of climatological products.

A possible correction for the vertical gradient could be useful for climatologists to better compare thermometers from different meteorological networks that are installed at different heights. Such correction shall depend on the type of screen, but also on the type of climate of a certain region. Although the measurements performed for this experiment are not suited for the estimation of such a general correction, a methodology is here proposed to correct the daytime measurements based on the fit of the estimated temperature differences with respect to solar radiation, as shown in Figure 4.14a. An empirical equation is proposed, fitted and applied to correct

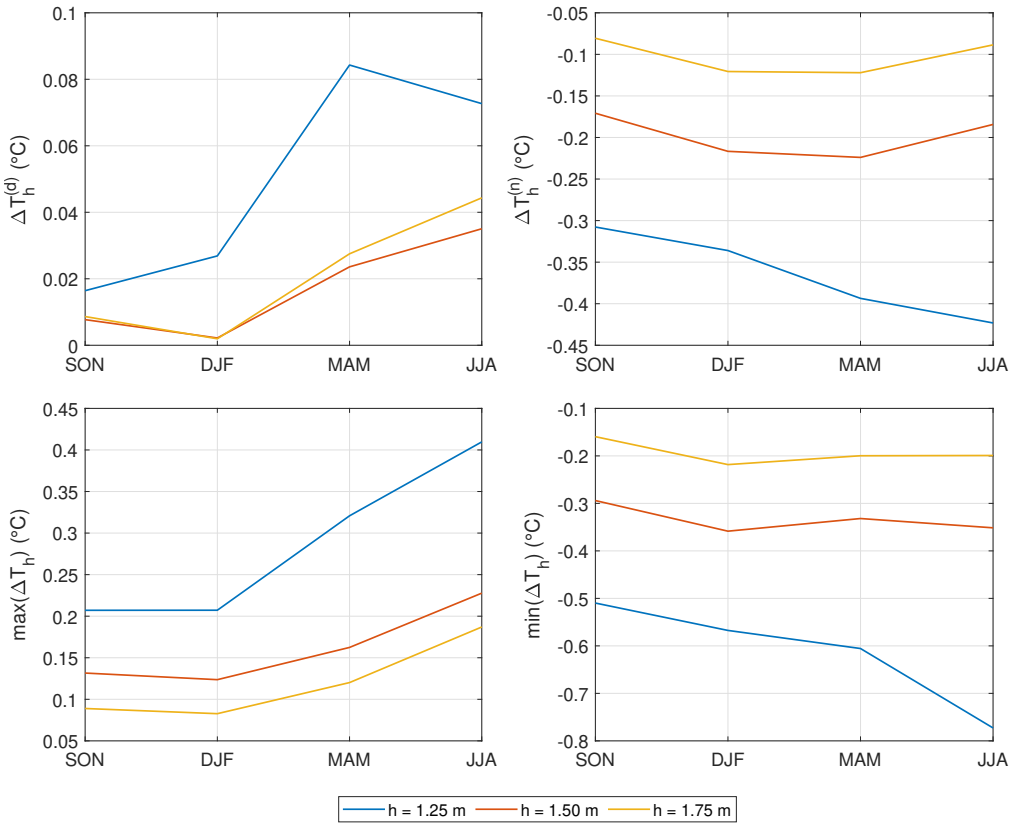


Fig. 4.17 Analysis of the seasonal variation for each installation height. The superscript "d" stands for daytime, whereas "n" is related to nighttime measurements.

the data, which are then used to evaluate the monthly maximum and minimum temperatures to verify the possible reduction of the estimated bias.

A rational polynomial function, reported in Equation 4.10, is considered for estimating the correction. The choice was driven by the need to follow the behavior of ΔT_h for values of solar radiation close to zero. Since the dispersion of data is not constant with SR, it is necessary to fit the equation through a weighted least-squares regression.

$$\Delta T_h = \frac{a}{\sqrt{SR}} + b\sqrt{SR} + c\,SR + d \tag{4.10}$$

The inverse of the dispersion for a given solar radiation is used as a weight in the fit. In particular, the dispersion is estimated by taking the running standard deviation, with a window of 50 data, and smoothing the result with a moving average of 1000 measurements. The estimated dispersion is added to the uncertainty budget of

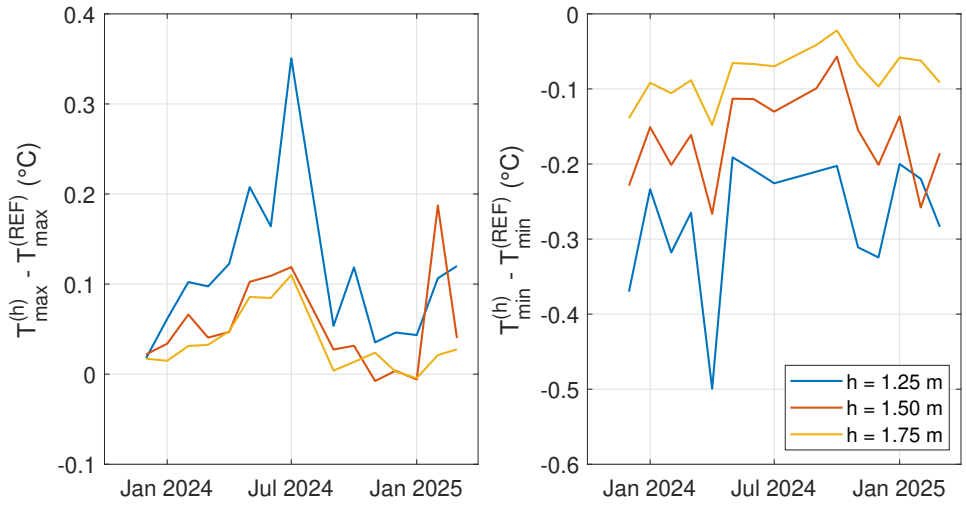


Fig. 4.18 Difference in the monthly maximum and minimum temperature, for each installation height, compared to the reference at 2 m.

the predicted correction to ensure that the weather variability observed during the experiment is properly accounted for.

The calculated regression coefficients, along with their standard uncertainty, are reported in Table 4.7, whereas the fit of $\Delta T_h(SR)$ is plotted for each height in Figure 4.19. The solid lines are the average corrections, whereas the dashed lines identify prediction intervals corresponding to $\pm 2\,u_{\text{fit}}$. The standard uncertainty of the predictions is obtained by propagating the uncertainty of the fitted coefficients and adding in quadrature the estimated dispersion. The uncertainty values are larger at 1.25 m, varying from 0.08 °C to 0.21 °C, whereas at 1.50 m and 1.75 m they are smaller than 0.1 °C.

Table 4.7 Coefficients of Equation 4.10 obtained through weighted least-squares regression. The number in parentheses is the standard uncertainty of the corresponding coefficient referred to the last digits of the quoted result.

h	$a/(°\text{C}\sqrt{W/m^2})$	$b/(°\text{C}/\sqrt{W/m^2})$	$c/(°\text{C}/(W/m^2))$	$d/°\text{C}$
1.25	−0.175(21)	0.027 03(75)	−0.000 230(18)	−0.2874(85)
1.50	−0.079(10)	0.013 71(42)	−0.000 120(10)	−0.1527(46)
1.75	−0.0005(58)	0.009 70(27)	−0.000 095(7)	−0.1011(27)

The estimated corrections are then applied to the temperature measurements. To check the effectiveness of such correction, the temperature differences are again evaluated and plotted against the solar radiation as in Figure 4.20. The ΔT_h are now

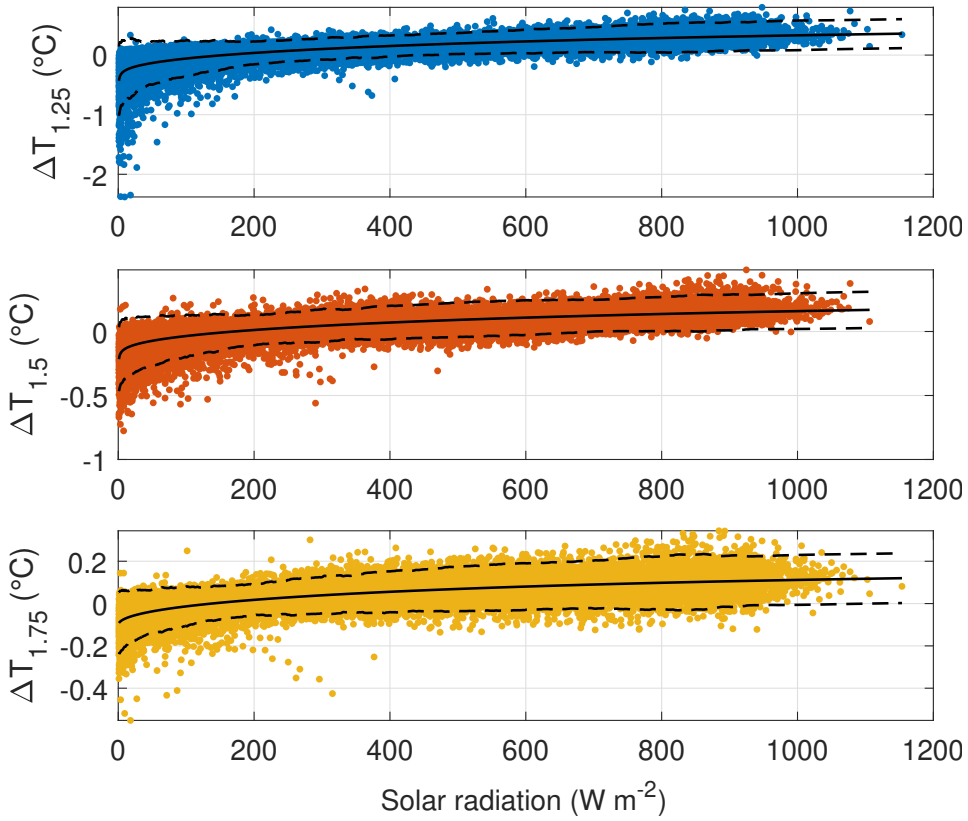


Fig. 4.19 Fit of Equation 4.10 on the daytime ΔT_h (solid black line). The dashed lines correspond to an expanded uncertainty of the predictions considering a coverage factor $k = 2$.

centered around zero with the same dispersion as before, but with no trend resulting from the diurnal cycle.

Afterward, the monthly mean maximum and minimum temperatures are again evaluated to verify eventual improvements. The plots in Figure 4.21 are compared to those in the left panel of Figure 4.18: excluding February 2025, all the differences are between $-0.07\,^{\circ}\mathrm{C}$ and $0.14\,^{\circ}\mathrm{C}$. The improvement is particularly visible for the bias at 1.25 m, with a maximum difference changing from $0.35\,^{\circ}\mathrm{C}$ to $0.14\,^{\circ}\mathrm{C}$. Furthermore, the differences are now more random around zero, hence being also negative. This random behavior may play a role in annual and decade averages because biases with opposite signs may compensate with each other.

The climatological analysis and the proposed correction for daytime measurements proved that experiments of this kind, with a complete evaluation of uncertainties, are crucial for reducing biases, but also for considering all the variability that may affect the measurements at a certain height with respect to a reference.

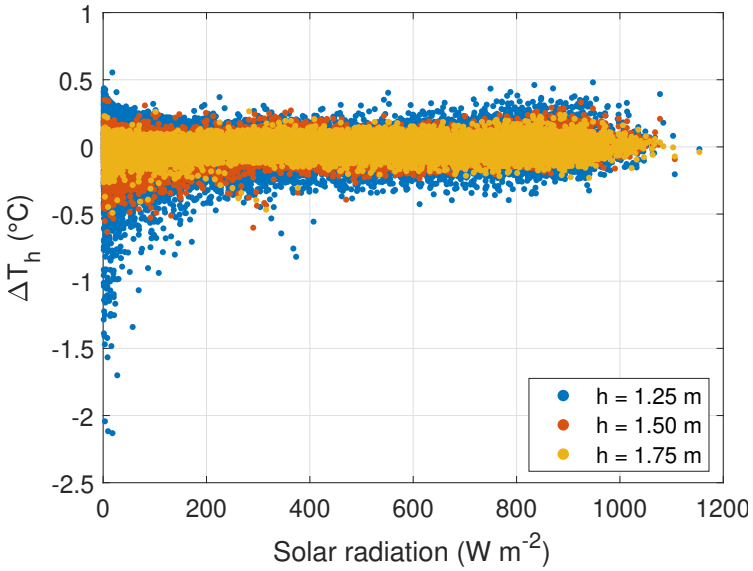


Fig. 4.20 Scatter plot of the corrected daytime ΔT_h with respect to the solar radiation.

The conducted metrological evaluation is a key point of the analysis, which was missing in the early experiments reported by Sparks to the WMO [95]. Finally, a precise standardized installation height, especially when using modern solar screens, should be determined and recommended through the main guide of atmospheric measurements [16]. At least, as an alternative, corrections like the one proposed in this work should be evaluated for the homogenization of data before performing climatological analyses.

4.3 Summary of main achievements

Chapter 4 examines how the installation height of thermometers within the standard World Meteorological Organization (WMO) interval of 1.25–2 m affects air temperature measurements and, consequently, climatological data products. Although the WMO has prescribed this interval since the first provisional guide (1950) to instruments and methods of observations [93], this choice was probably conventional rather than the outcome of systematic scientific evaluation. The chapter addresses this gap by investigating whether small variations in sensor height introduce systematic biases in climatological records, especially in the estimation of temperature extremes.

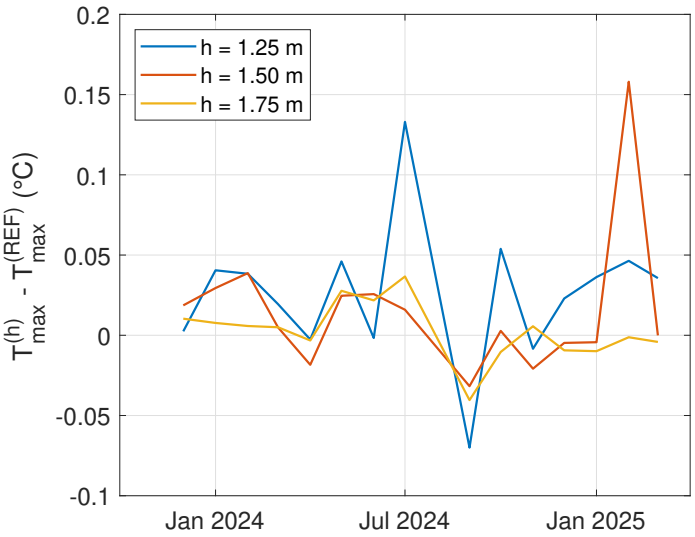


Fig. 4.21 Difference in the monthly maximum temperature, for each installation height, compared to the reference at 2 m using the daytime corrected measurements.

To explore this question, an intercomparison experiment (shown in Figure 4.2a) was carried out, with metrological rigor, using eight platinum resistance thermometers, each housed in mechanically ventilated radiation shields (model Young 43502) to minimize radiative biases. The sensors were mounted at four distinct heights: 1.25, 1.50, 1.75, and 2.00 m, with the 2 m level serving as reference. Additional instruments recorded auxiliary variables such as solar radiation, wind speed, and relative humidity, enabling the identification of environmental influences on vertical temperature gradients. The experimental site was an open grass field, selected to avoid obstacles and artificial heat sources, although it was characterized by relatively low wind speeds, an important factor for vertical mixing.

First, all sensors were carefully calibrated by comparison with INRiM reference thermometers, traceable to the SI, using a climatic chamber (see Section 4.1.1). The achieved expanded uncertainty ($k = 2$) was about 0.05 °C for all the sensors. During the experiment, data were collected continuously, averaged over 10-minute intervals, and further smoothed with a 30-minute moving window to suppress turbulence-driven noise. Faulty records and outliers were filtered out using quality-control procedures empirically tuned for the acquired dataset. The analysis focused on temperature differences (ΔT_h) between each installation height and the 2 m reference. The use of relative measurements compensates for common components of the

instrumental uncertainty, hence permitting a better determination of biases affecting the temperature records.

The results revealed systematic vertical gradients in air temperature that varied between day and night (Section 4.2.2), across seasons (Section 4.2.3), and with changing environmental conditions. During the daytime, sensors closer to the ground generally recorded warmer values: the solar radiation warms the ground, which in turn transfers heat to near-surface air layers thanks to the turbulent motion of air. At 1.25 m, the daytime bias relative to 2 m reached about 0.13 °C as the most probable value. At night, the opposite effect dominates: radiative cooling of the ground and stable boundary-layer stratification led to cooler temperatures at lower levels. The nighttime bias was most pronounced at 1.25 m, with an average offset of −0.36 °C. The dispersion affecting the measurements is comparable at each height for both conditions, being one of the uncertainty components associated with the best estimates. Figure 4.22 reports a summary of the results, showing the mean temperature bias, with the evaluated standard uncertainty, observed during the experiment for daytime and nighttime measurements.

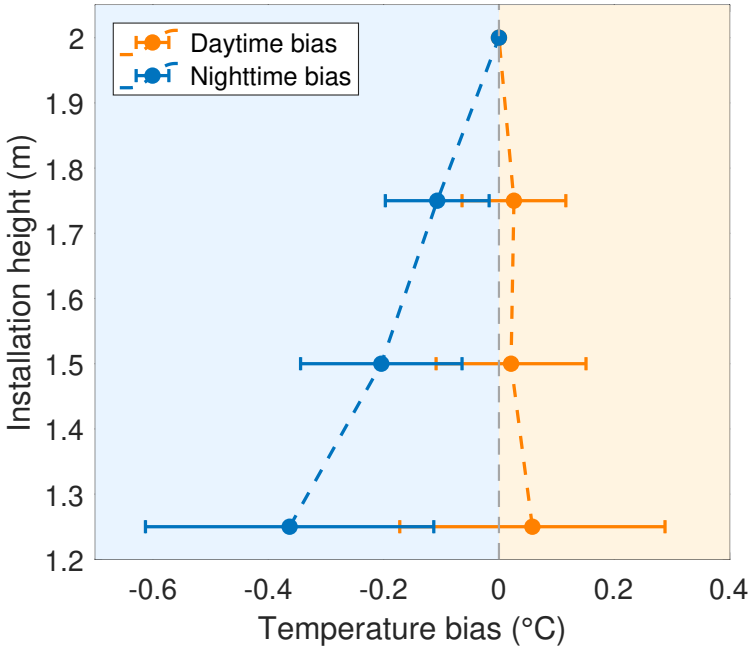


Fig. 4.22 Mean temperature bias and corresponding standard uncertainties evaluated at different heights for daytime and nighttime measurements.

Seasonal analysis highlighted that these effects were not constant throughout the year. In summer, when solar radiation is stronger and vegetation is more active, the

maximum daily difference at 1.25 m was up to 0.41 °C. In contrast, the nighttime minimum temperature difference was as low as −0.77 °C. These amplified differences likely reflect the combined influence of soil–atmosphere heat exchange, grass growth, and reduced nocturnal mixing. In winter, biases were generally smaller but still detectable.

The chapter also emphasizes that such differences are climatologically significant (Section 4.2.4). Monthly means of daily maxima and minima showed biases of up to 0.35 °C (maxima) and −0.5 °C (minima) at 1.25 m. In contrast, heights closer to 1.5–1.75 m exhibited much smaller deviations: typically less than 0.1 °C for maxima and −0.26 °C for minima. While modest in magnitude, these systematic shifts can distort long-term temperature records, particularly when evaluating climate extremes or constructing homogenized datasets.

The study also explored how meteorological factors modulate the height-dependent biases (see Figure 4.14). Solar radiation strongly amplifies near-surface warming during the day, while higher wind speeds reduce vertical gradients by enhancing turbulent mixing. Relative humidity plays a secondary role, influencing cooling through condensation processes at night.

Recognizing the need to mitigate these biases, a model is developed based on the relation with the solar radiation to correct the daytime data. When applied to the dataset, this correction substantially reduces systematic errors. For example, the maximum bias at 1.25 m decreases from 0.35 °C to 0.14 °C, improving comparability across installation heights (see Figure 4.18 compared to Figure 4.21). This approach demonstrates that post-processing can partially compensate for variations in siting, though prevention through stricter installation guidelines remains preferable.

To conclude, the installation height within the WMO's permitted interval can significantly influence measured air temperatures, particularly extremes, through mechanisms linked to surface–atmosphere interactions and boundary-layer stratification. These effects are strongest at 1.25 m and progressively diminish toward 2 m. While seemingly minor, the biases may accumulate over decades, compromising the homogeneity of climate records. Therefore, the study recommends either narrowing the definition of installation height (e.g., closer to 2 m) or implementing standardized correction methods to ensure comparability across networks. This evidence-based refinement of siting protocols could strengthen the reliability of reference climato-

logical datasets and reduce the need for extensive homogenization efforts in climate trend analysis.

Chapter 5

Metrology for fluxes: Eddy Covariance measurement uncertainty

The Eddy Covariance (EC) methodology in the last four decades has emerged, consolidated and proven to be an essential tool in the scientific research about turbulent exchange of fluids [97–100]. In particular, it has been fruitfully applied as a reference micrometeorological tool in fields like ecological climatology, agrometeorology or hydrology to directly quantify the dynamics of greenhouse gases (GHGs), mass and energy fluxes between the atmosphere and ecologically relevant surfaces, e.g. forests, croplands, grasslands or water bodies [99, 101–103]. Its relevance and worldwide popularity is mostly due to the robustness of the methodology and to its ability to measure continuously the surface turbulent exchanges and their budgets on time scales ranging from hours to decades and with minimal disturbances to the environment. The technological development of the measurement and computational tools that also occurred in the last decades, made this methodology a fit for purpose tool not limited for scientific research, but also for constant monitoring activities: an example is the establishment of continental scale networks and infrastructure like ICOS (Integrated Carbon Observation System) in Europe, NEON (National Ecological Observation Network) in Noth America or TERN (Terrestrial Ecosystem Research Network) in Australia. Moreover, new applications of the EC methodology are promoting the installation of new sites, e.g. for research and monitoring of climate, meteorology and pollution in urban environments or in agricultural production contexts, where the scientific investigation is proceeding alongside practical applications. An interesting and recent example is the possibility of using the EC methodology to provide a direct

estimate of GHGs exchange in the context of Monitoring, Reporting and Verification (MRV) supporting the Emission Trading System (ETS) or to provide independent estimates of emission factors used in the country scale inventories of GHGs, e.g. by organizations like IPCC and FAO [104–106].

However, the EC has been so far mostly confined to the academic research and minor attention has been paid to the metrological features that a technique should possess in order to provide robust, reliable and comparable measurements. As a consequence, despite significant advances has been done in specific aspects like the different setups and processing options [107, 108] or the modelling tools for carbon flux interpretation or upscale [109, 103, 110, 111], a framework to express the measurement uncertainty according to internationally accepted guidelines, like the JCGM guides [5, 18, 112], has not been yet defined. In particular, an analytical expression of the measurement uncertainty, applicable to all types of instruments used for the EC estimation of turbulent flux and satisfying the accepted metrological guidelines, has not been implemented thoroughly.

In the past two decades, several highly detailed analyses of "errors" have been performed to discuss specific issues of the EC methodology (Chapter 7 in [99]). In the EC community, these errors were traditionally divided in "random errors', such as those arising from empirical sampling or instrumental noise [113–116], and the "systematic errors" mainly associated to data processing or measurement issues such as the concentration and the synchronization biases [117, 118]. However, almost none of them approached the problem with metrological rigor, especially concerning systematic effects. Every correction has an associated uncertainty which must be evaluated and added to the uncertainty budget of flux measurements. Although this operation is spread among metrologists, it is not in other fields such as micrometeorology, as demonstrated by the total absence of evaluated uncertainty to associate with corrections despite the great effort put into estimating systematic errors. The lack of metrological knowledge is also confirmed by an erroneous sentence in Chapter 7 of the book by Aubinet et al. [99], one of the most cited book in the EC field, which is here reported: "Some of the systematic errors in flux measurements are well characterized, and corrections [...] have been developed for these biases [...]. However, in many cases, the corrections for these errors are imperfect, and thus *some uncertainty remains even after the correction is applied*.". This statement is false and not metrologically sound: an uncertainty remains not because the correction is imperfect, but because, even in the ideal case of a correction with null uncertainty,

a measurement has a minimum uncertainty associated with the use of a calibrated instrument traceable to the SI. Therefore, *the measurement uncertainty is always present, even when no correction is applied.*

A complete evaluation of the uncertainty budget for EC flux measurements is missing, even in the absence of systematic effects. Indeed, measurements of different physical quantities are involved in the EC, having a corresponding measurement uncertainty that must be considered and propagated through the model used to define the measurand (flux), that is the intended quantity to be measured [119]. The uncertainty evaluation should allow the analysis of the sensitivity to all the measured quantities and should be coherent and flexible enough to be applicable to the widest possible range of measurement values and devices with a reasonable compromise between complexity, implementation efforts and representativeness [5]. Upon these considerations, a method to quantify the measurement uncertainty of EC carbon flux estimates based on analytical propagation is developed and presented in this chapter. Focus is put on the propagation of instrumental uncertainty, for each involved quantity, considering the flux equation for an enclosed-path analyzer. Furthermore, the main applied correction for the unlevelled anemometer is considered, showing how uncertainties should be propagated when a corrected model is used. Then, the developed method is applied to single measurements and modelled values to quantify the impact of measurement uncertainty on annual carbon budget in a typical case study, represented here by an ICOS station (labelled San Rossore 2) monitoring GHGs fluxes from a Mediterranean forest with up-to-date instruments and procedures.

The chapter is organized as follows: Section 5.1 briefly introduces the the theoretical aspects of the EC technique and gives an overview of the random and systematic effects studied at the moment. Then Section 5.2 describes how the uncertainty of the input quantities can be propagated through the measurement model, but also how the uncertainty of the annual carbon budget is evaluated starting from the uncertainty of CO_2 fluxes. The application of the developed methodology to the case study is presented in Section 5.3, which also reports the implications of accounting measurement uncertainty for the annual carbon budgets. Eventually, the conclusions with summarized results and final considerations are given in Section 5.4.

5.1 Background on Eddy Covariance

The Eddy Covariance method provides measurements of gas emission and consumption rates and allows measurements of momentum, sensible heat, and latent heat (e.g., evapotranspiration, evaporative water loss, etc.) fluxes integrated over areas of various sizes. Eddy Covariance measurements are typically made in the surface boundary layer, which is approximately $20-50$ m high in the case of unstable stratification and a few tens of meters in stable stratification [99]. Fluxes are approximately constant with height in the surface layer, hence measurements taken in this layer are representative of the fluxes from the underlying surfaces which are desired to be known. Here atmospheric turbulence is the dominant transport mechanism, justifying the use of the Eddy Covariance approach to measure the fluxes. A physical insight of the mechanism is given below by following the reasoning in [120], then a simplified mathematical derivation of the main equation is presented alongside the main assumptions of the techniques. At last, an overview of the instrumental errors and uncertainty contributions studied until now is given.

Physical insight

Turbulent eddies are mainly created by a combination of two mechanisms: shear and buoyancy. When an horizontal wind encounter physical obstacles like plant canopies, buildings, or uneven terrain, it creates mechanical shear which causes the formation of turbulent eddies. Larger eddies also create their own shear, which in turn generates smaller eddies. This transfer of energy from large to small scales is known as the turbulent energy cascade. On the other hand, buoyancy occurs when air rises or falls due to differences in density, typically caused by temperature, disrupting the horizontal wind flow.

Farther from the ground, larger eddies are more common and rotate slowly, transporting material at lower frequencies. Conversely, closer to the ground, smaller eddies are more likely to be the main transporters which rotate faster and shift the transport to higher frequencies, making fast, small-scale air movements dominant. In practice, air is always a mix of different eddy sizes, with transport occurring across a wide range of frequencies. This includes everything from large, slow movements that take hours, to very small and fast movements that happen in the order of 0.1 s.

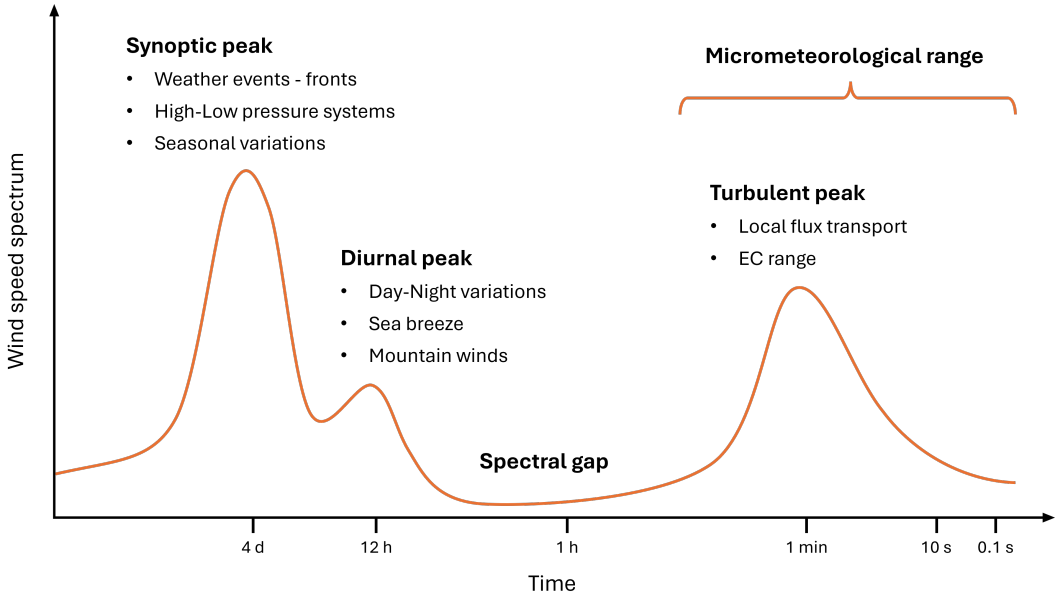


Fig. 5.1 Wind speed frequency spectrum sampled at Brookhaven National Laboratory at about 100 m, reproduced from Van der Hoven.

The micrometeorological flux transport at the lower frequency is limited to time periods of about 30 minutes to a few hours, as shown in Figure 5.1 reporting the wind power spectrum measured by Van der Hoven in [121]. Slower movements due to fronts and high-low pressure systems are considered at the synoptic scale, which is much larger in both space and time than the turbulence generated by the local ecosystem or area being studied. Between the turbulent peak and the diurnal peak, caused by daily phenomena (such as sea breeze) there is the so-called spectral gap in which no significant transport is present. The Eddy Covariance method addresses the separation of the localized turbulent transport from much larger synoptic scale variations by using Reynolds decomposition. This technique calculates fluxes over an averaging period of 30 minutes to a few hours, with each variable (X) being expressed as the sum of its mean value ($\overline{X}$) over that period plus an instantaneous deviation from the mean (X'), that is $X = \overline{X} + X'$. This process significantly reduces the influence of large-scale synoptic variations, ensuring that the measured flux is primarily from the local area and transported by locally generated turbulence.

To grasp how eddies favor the transport of a certain quantity, for instance the dry molar fraction S of a gas, it is convenient to look at a single point in space above the ecological surface of interest, for instance a forest, as shown in Figure 5.2. At time 1, eddy number 1 moves air parcel 1 downward with speed $V_{z,1}$. In

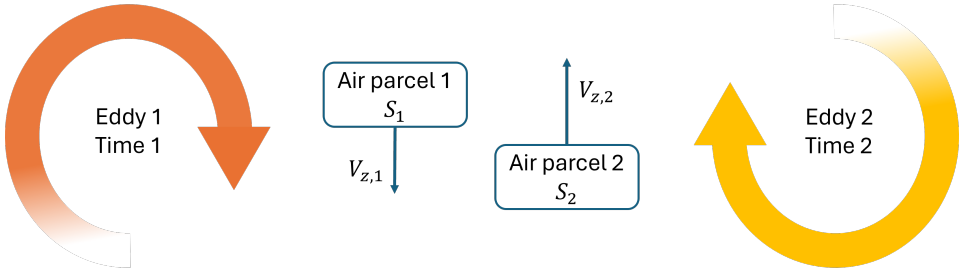


Fig. 5.2 Representation of turbulent transport associated with a single point in space. The depicted eddies refer to two different time instants in which the air parcels, characterized by a certain quantity S, are moved in the vertical direction.

the same point, at time 2, eddy number 2 moves the air parcel 2 upward with speed $V_{z,2}$. Each air parcel has its own characteristics, such as gas concentrations, temperature, humidity, etc. If these characteristics and the speed of the vertical air movement are measured, the vertical upward or downward fluxes of gas and water vapor concentrations, temperature, and humidity can be estimated. For example, if at one moment three molecules of CO_2 went up, and in the next moment only two molecules of CO_2 went down, then the net flux over this time was upward, and equal to one molecule of CO_2. Mathematically, such vertical flux can be represented as covariance between measurements of vertical velocity, the upward and downward movements, and the concentration of the entity of interest. Measuring this requires highly advanced instruments that can detect rapid, tiny changes in concentration, density, or temperature with both high speed and resolution. While the standard Eddy Covariance method relies on several assumptions and requires expensive, high-end equipment, it provides a nearly direct and continuous measurement of flux as long as those assumptions are met or properly accounted for.

Mathematical derivation

Starting from the conservation equation of a scalar quantity in the atmosphere, it is possible to derive the general expression for the vertical flux of a gas due to turbulent transport:

$$F = \overline{\rho V_z S}, \tag{5.1}$$

where ρ is the air density, V_z is the vertical wind speed component and S is the dry mole fraction of the gas of interest in the air. The overline indicates that the average over time of this product, evaluated in a 30-minute interval or a few hours of interest,

must be taken to estimate the mean flux. Note that in the micrometeorology the wind speed components are usually labelled as u,v and w. Instead, the notation V_x,V_y, V_z is used in this work to avoid confusion with the symbol for uncertainty.

As mentioned above, since eddies at the synoptic level are not relevant for local fluxes, Reynolds decomposition is used to break the terms into means and deviations:

$$F = \overline{(\overline{\rho} + \rho')(\overline{V_z} + V_z')(\overline{S} + S')}. \tag{5.2}$$

The detailed product is simplified by removing the terms depending on the average of the deviations, which is zero by construction. Then the remaining terms are the following:

$$F = \overline{\rho}\,\overline{V_z}\overline{S} + \overline{\rho}\,\overline{V_z'S'} + \overline{V_z}\,\overline{\rho'S'} + \overline{S}\,\overline{\rho'V_z} + \overline{\rho'V_z'S'}. \tag{5.3}$$

At this point, import assumptions are made in the conventional Eddy Covariance formulation:

- Air density fluctuations are assumed to be negligible, which is reasonable when the technique is applied over flat and vast spaces. This assumption allows to neglect three terms on the right-hand side of the above equation.
- The mean vertical flow is assumed to be negligible over horizontal homogeneous terrain, meaning that stationarity conditions must be ensured with no flow diversions or conversions. This hypothesis allows to remove an additional term depending on $\overline{V_z}$.

Under these assumptions, the classical equation for Eddy Covariance gas fluxes is derived:

$$F = \overline{\rho}\overline{V_z'S'}, \tag{5.4}$$

where the product $\overline{V_z'S'}$ expresses the empirical covariance between the vertical wind speed and the gas concentration, thus giving the name to the technique. Generally the flux is reported in $\mu\text{mol}\,\text{m}^{-2}\,\text{s}^{-1}$ and the dry molar fraction is in µmol/mol. Apart from the air density factor, the covariance represents the quantity intended to be measured (the measurand), which is unusual for a measurement model since in metrology the covariance is used as a metric to account for correlation among different quantities. This observation would be relevant below when the sampling uncertainty is introduced.

It is convenient to express the Eddy Covariance equation by making explicit the air molar density:

$$F = \frac{\overline{P}}{R\overline{T}}\overline{V_z' S'}. \tag{5.5}$$

The quantity P is the air pressure of air, the term R is the gas constant, whereas T is the air temperature. The estimation of the covariance requires the use of instruments able to detect very small changes at a very high frequency, whereas the sampling speed requirements for pressure and temperature are more relaxed under the assumption of negligible air density fluctuations. High-end sonic anemometers coupled with infrared gas analyzers (IRGAs) are used for measuring the covariance since have a sampling frequency of 10-20 Hz. These instruments are installed as shown in Figure 5.3. On the other hand, barometers and thermometers used for meteorological applications can be used for estimating the mean air density, provided that the sampling interval is set to 30 s, as suggested by Rebmann et al. in [107].

Fig. 5.3 Configuration of a gas analyzer coupled with a sonic anemometer used for EC measurements.

Although other assumptions concerning the representativeness of measurements must be met, they are less relevant for the purpose of this chapter, which is the analysis of the instrumental uncertainties. An overview of the state of the art related to the errors and uncertainties affecting the Eddy Covariance measurements is now presented below.

Errors and uncertainties in EC

Systematic errors in the flux measurements could arise mainly for two reasons: (i) assumptions not met and (ii) instrumental errors that affect the measurement system as a whole or just specific components, as the anemometer or the IRGA. The first type of error is critical to the application of the technique, whereas the second type is more related to technological and physical aspects of the instruments. As mentioned in the introduction to the chapter, the errors were and are still studied by researchers and manufacturers to find optimal corrections. Nevertheless, uncertainties associated to corrections are seldom evaluated, being an open field for metrological research.

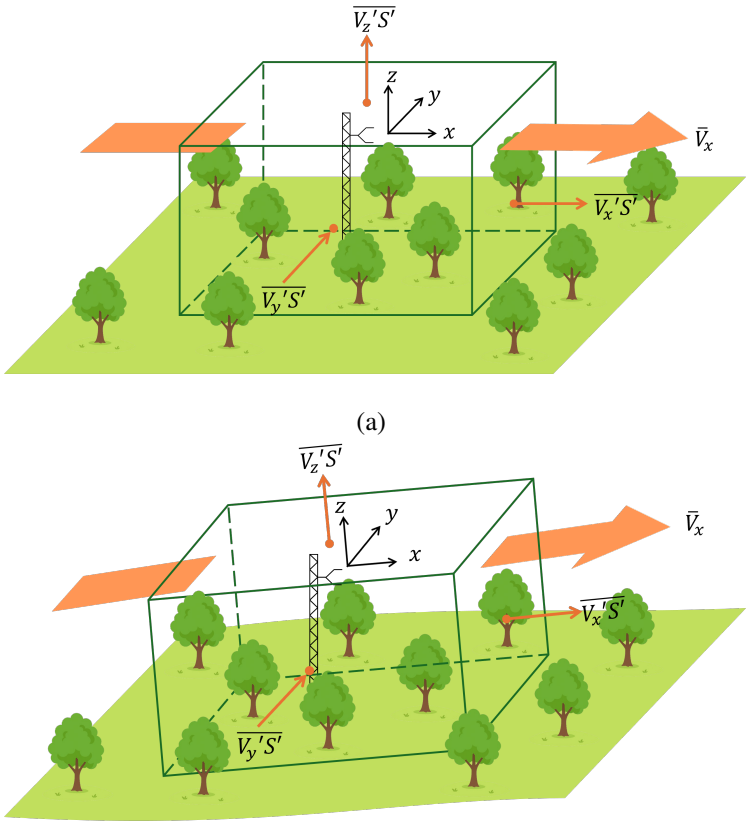


Fig. 5.4 (a) The coordinate axes at the sampling volume are aligned with the mean wind vector, which is parallel to the homogeneous surface. (b) The complex topography, an incorrect leveling or a short averaging period may introduce a tilt error, which must be compensated by a rotation of coordinates as described in the text.

As described above, one core assumption of the EC method is that, on average, there is no net vertical wind movement. To meet this assumption, the pole holding the sonic anemometer must be perfectly perpendicular to the average wind streamline, as shown in Figure 5.4a. However, this is not always the case due to factors like the mean wind streamline not being perfectly parallel to the ground, complex topography, slight leveling errors with the instrument, or large-scale, low-frequency air movements, as represented in Figure 5.4b. Therefore, a coordinate rotation is often performed on the data. This process mathematically adjusts the coordinate frame so that the measured vertical turbulent flux can be corrected to closely approximate the total flux, even when the initial assumption on V_z is not perfectly met. The planar fit and the double rotation methods are the most employed for correcting the coordinate frame [120], an operation that is always performed and mandatory for applying the EC technique. For this work the double rotation method is considered: the first rotation is used to impose $\overline{V_y} = 0$, whereas the second one sets $\overline{V_z} = 0$. In the end the flux is obtained by a linear combination of the sampled covariances in the three directions:

$$F = \overline{\rho}\,(A\,\overline{V_x' S'} + B\,\overline{V_y' S'} + C\,\overline{V_z' S'}), \tag{5.6}$$

where A, B and C are the coefficients derived from the application of the rotation matrices. Equation 5.6 represents the measurement model that should at least be considered when performing EC measurements, in particular when uncertainty is propagated, as will be shown in Section 5.2. For this purpose, it is particularly relevant knowing the expression of each rotation coefficient, which is a function $f(\overline{V}_x, \overline{V}_y, \overline{V}_z)$ where the time average values of the wind speed components are involved. The explicit expressions are the following:

$$A = -\frac{\overline{V_x}\,\overline{V_z}}{\sqrt{\overline{V_x}^2 + \overline{V_y}^2 + \overline{V_z}^2}\,\sqrt{\overline{V_x}^2 + \overline{V_y}^2}} \tag{5.7}$$

$$B = -\frac{\overline{V_y}\,\overline{V_z}}{\sqrt{\overline{V_x}^2 + \overline{V_y}^2 + \overline{V_z}^2}\,\sqrt{\overline{V_x}^2 + \overline{V_y}^2}} \tag{5.8}$$

$$C = \frac{\sqrt{\overline{V_x}^2 + \overline{V_y}^2}}{\sqrt{\overline{V_x}^2 + \overline{V_y}^2 + \overline{V_z}^2}} \tag{5.9}$$

Sometimes, the stationarity assumption is not ensured (even after rotating the measurement framework) because of significant trends in the time series which do not allow to estimate fluctuations with a simple block average. The main approach in this case is to detrend the measurements through a least-squares regression [122] before calculating the fluctuations. Even though researchers have doubts about the application of EC under such non-stationarity conditions, they never questioned the uncertainty of the parameters estimated from the fluctuating time series, which in principle should be considered and propagated through the model.

The last important error related to EC assumptions is due to the chosen time interval for the empirical estimation of the covariance. Too long intervals could include non-turbulent contributions to the covariance, whereas a too short interval acts as a high-pass filter, leading to missed input from larger eddies and underestimated fluxes [123, 120]. Time intervals of 30 minutes or 1 hour are suggested, but may not be the best everywhere since the optimal choice is strongly site-dependent. An empirical approach is often applied for the determination of the optimal interval: different reasonable averaging times (from 10 minutes to 4 hours) are attempted, then the one that maximizes the covariance is chosen.

As mentioned above, the other errors affecting the EC technique are related to instruments. Some of them concern the measurement system as a whole and are restricted to the time domain. Three well-known errors of this type are the time lag [99, 132], the time synchronization [118] and the frequency attenuation error [125]. The time lag error takes into account the fact that the anemometer and the IRGA are not located in the same place, thus they sample the same air parcel at different times. One way to minimize this error is to shift the time series of one of the two instruments until the covariance is maximized. On the other hand, the time synchronization error is strictly related to time differences in the clocks of the two instruments that may lead to erroneous measurements. This problem is merely technological and is solved by using advanced synchronization protocols such as the Precision Time Protocol (PTP) [120]. The frequency attenuation of the covariance signal is instead inevitable since depends on the physics of the measurement technique. The time response of the sensors, the tube attenuation, high/low-pass filters and other factors contribute to the frequency loss. Generally, a transfer function model taking into account each contribution is assembled and used to correct the cospectra, i.e. the fourier transform of the estimated covariance.

Table 5.1 List of systematic errors affecting the EC technique. For each error it is reported which part of the method is affected, the possible correction and what are the main scientific references.

Error	Affecting:	Correction	Reference
Unlevelled anemometer	Model assumption	Coordinate rotation	[124]
Non-stationarity	Model assumption	Detrending	[122]
Unsuited time integration	Model assumption	Emprirical	[123, 120]
Time lag	Anemometer + IRGA	Covariance maximization	[99, 120]
Time synchronization	Anemometer + IRGA	PTP protocol	[118]
Frequency attenuation	Anemometer + IRGA	Cospectra correction	[125]
Angle of attack	Anemometer	Angle-dependent calibration	[126–128]
Flow distorsion	Anemometer	-	[120, 129]
Density fluctuations	IRGA	WPL corrections or fast dry mole fraction	[130, 131]
Spectroscopic effect	IRGA (laser)	Instrument-specific	[132]
Band-broadening	IRGA (NDIR)	Instrument-specific	[133, 134]
Concentration bias	IRGA	Periodic verification and corrections	[117]

The anemometer and IRGA measurements may be subjected to errors themselves. For instance the anemometer has an imperfect response to winds coming from steep angles [126, 127], which can be corrected by an angle-dependent calibration [128]. Another problem is that the instrument itself influences the measurand through distorsion of the air flow [129]. This problem is still under evaluation (see pages 97-104 in [120] for an extensive literature review), but a generalized correction is unfeasible since it would be strongly depend on instrument design and geometry.

Concerning the IRGA, the most important error investigated for open-path analyzers (those without a sampling tube) is the one arising from density fluctuations induced by temperature (thermal expansion) and water vapour (dilution) variations in the air parcel crossing the instrument. This error can be compensated by taking into account this phenomenon in the measurement model, thus adding the so called Webb-

Pearman-Leuning (WPL) terms [130, 131]. Another solution is to use enclosed-path analyzers which are capable to output fast dry mole fraction by measuring in the sampling cell fast temperature, pressure, water vapor concentration and gas density of interest. In this way it is possible to use the model in Equation 5.5 without adding any terms. The temperature and water vapor fluctuations induces also spectroscopic effects on IRGA based on lasers, thus on the interaction between the air and single emitted lines. Corrections are instrument-specific and available for some models [132]. The most spread analyzers based on non-dispersive infrared radiation (NDIR) are also affected by errors, mainly because of band broadening due to the presence of other gases in the sampling volume. This effect is well-known and already corrected by manufacturers [133, 134]. Eventually, an important systematic error that is common to both type of IRGA is the concentration bias, which is associated with actual or apparent instrumental drifts or biases due to thermal expansion, dirt contamination or aging of components. A correction is proposed by Fratini et al. [117] based on knowledge of instrument calibration curves and on periodic on field verifications.

Although the described errors, listed in Table 5.1, result from the most studied systematic effects for the EC method, the uncertainty associated with each correction was rarely quantified. In the past, comparisons between different processing schemes were performed, whose results were used as input for uncertainty evaluation of fluxes [135, 108]. A more metrological approach requires an uncertainty evaluation for each correction, starting from the uncertainty of the single components and propagating it through the expression of the corrected model.

Despite the lack of metrological rigor, a type of uncertainty is always evaluated in the main processing schemes, which is the so-called “random uncertainty” [136, 99, 115, 116]. Although the classification of different types of uncertainty is improper according to the GUM principles [5], this uncertainty should take into account all the effects that are stochastic in nature. Two main effects that are generally evaluated are the instrumental noise and the imperfect sampling of the covariance. Following the strategy suggested by Mauder et al. [115], the instrumental noise for each instrument is estimated through the method proposed by Lenschow et al. [137], exploiting the autocorrelation function of the sampled time series. Using this technique, the estimated value embeds the electronic noise proper of the instrument, but also the noise associated with the sampled physical process. The evaluated noise variance is then propagated through the model, which normally is considered with no applied corrections. The overall contribution u_{noise} is then added in quadrature to

u_{sampling}, that is the uncertainty evaluated for taking into account turbulence sampling error. Indeed, Finkelstein and Sims in [113] note that errors can arise because large eddies, which are responsible for much of the total flux, cannot be adequately sampled during a 30 minutes integration period. They proposed an approach for the evaluation of the variance of the empirical covariance, which is widely recognized among the EC community. The overall random uncertainty is then evaluated by $u_{\text{random}} = \sqrt{u_{\text{noise}}^2 + u_{\text{sampling}}^2}$.

In addition to the described uncertainties, evaluated or not, an essential contribution is missing: the instrumental uncertainty resulting from the calibration and use of the instrument. This uncertainty *must* always be considered and propagated through the measurement model to ensure the traceability of measurements to the SI. This contribution may be small compared to other terms, but this judgement should be determined by numerical evaluation and not based on common thoughts. An attempt of instrumental uncertainty propagation is shown in the following section, considering also different assumptions underlying the measurement process, such as type-B independence or correlation among the measurements.

5.2 Uncertainty propagation in EC

The first step for the evaluation of the measurement uncertainty is the definition of the employed measurement model. The model for CO_2 flux considered in this work is expressed by Equation 5.6, which implies the double rotation of wind speed measurements as described in the previous section. This model is more suited when using enclosed-path analyzers, which do not require the WPL correction normally adopted for open-path analyzers. To simplify reading, the model is again reported below:

$$F = \overline{\rho}\,(A\,\overline{V_x'S'} + B\,\overline{V_y'S'} + C\,\overline{V_z'S'}).$$

The number of measurements used for estimating the covariances and the air density could be different. Indeed, air pressure and temperature are normally sampled every 30 s, whereas concentration and wind speed components are generally measured 10-20 times every second (10-20 Hz). For this reason, the variable N_X will indicate the number of individual measurements of the quantity X in a given averaging interval, in this case with a fixed duration of thirty minutes as in most of the EC applications. Usually $N_P = N_T = N_{\text{met}}$ where the subscript refers to measurements

of meteorological variables with a lower sampling frequency. Similarly, $N_S = N_{V_x} = N_{V_y} = N_{V_z} = N_{\text{turb}}$ where the subscript refers to measurements of turbulent variables.

Since the flux is estimated using an indirect measurement model [112], the uncertainty of each measurement must be propagated following the law of uncertainty propagation prescribed in the GUM [5]. This work focuses on the propagation of the instrumental uncertainty related to the use of an instrument. A representative value is chosen for each instrument by looking at its specifications or calibration certificates.

The general propagation of the uncertainty for a quantity $X \in \{P, T, S, V_x, V_y, V_z\}$ through the model can be easily written in matrix formulation in the following way:

$$u^2(F) = \begin{bmatrix} J_P \\ J_T \\ J_S \\ J_{V_x} \\ J_{V_y} \\ J_{V_z} \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} M_P & & & & & & \\ & M_T & & & & & \\ & & M_S & & & & \\ & & & M_{V_x} & M_{V_xV_y} & M_{V_xV_z} \\ & & & M_{V_yV_x} & M_{V_y} & M_{V_yV_z} \\ & & & M_{V_zV_x} & M_{V_zV_y} & M_{V_z} \end{bmatrix} \begin{bmatrix} J_P \\ J_T \\ J_S \\ J_{V_x} \\ J_{V_y} \\ J_{V_z} \end{bmatrix} = J^{\mathrm{T}} M J. \quad (5.10)$$

The J_X vectors correspond to the Jacobian storing the partial derivatives with respect to quantity X and evaluated for the corresponding number of measurements. Instead, the terms M_X are the variance-covariance matrices for the measurements of X, whereas M_{XY} are the cross-covariance matrices between measurements of different quantities, which are not zero only for the three wind speed components because taken with the same instrument (the tri-axial sonic anemometer). The larger number of measurements involved in the model requires to make some assumptions to simplify the propagation without compromising the relevance of the obtained results.

5.2.1 Assumptions

To make the propagation and the analysis of flux uncertainties simpler, the same uncertainty is associated to all the measurements of the same quantity (homoscedastic case). Even though this assumption could be stringent, the methodology proposed in this work can be generalized to the heteroscedastic case, which implies a better knowledge of the uncertainty budget for the calibration of each instrument, information not usually available.

To apply the law of uncertainty propagation, it is necessary to quantify possible correlations among the measurements for evaluating the M_X covariance matrices. The type-B correlation resulting from the use of the same device (or measurement system) is here addressed. This type of correlation takes into account that different measurements of the same quantity X are obtained using the same calibrated instrument. Other correlations, such as time autocorrelation, are here neglected supposing that the instrumental uncertainty is decoupled from the field effects. Again, this assumption is not so stringent since the uncertainties emerging from these aspects shall be added to the instrumental one and treated separately. In mathematical terms, this assumption can be formulated in the following way:

$$X(t) = x + \delta X_{\text{inst}} + \delta X_{\text{field}}(t) \tag{5.11}$$

The term x is a realization of the aleatory quantity X obtained through a measurement, whereas δX_{inst} and $\delta X_{\text{field}}(t)$ are the stochastic components that account for instrumental uncertainty and environmental effects, having zero mean and finite variance. Note that only $\delta X_{\text{field}}(t)$ depends on the sampling time, whereas the variance $u^2_{\text{inst}}(X) = \mathrm{V}[\delta X_{\text{inst}}]$ derives from the calibration of the instrument. The covariance of two measurements of X is then evaluated as:

$$\mathrm{cov}(X(t_i), X(t_j)) = \mathrm{cov}(\delta X_{\text{inst},i}, \delta X_{\text{inst},j}) + \mathrm{cov}(\delta X_{\text{field}}(t_i), \delta X_{\text{field}}(t_j)) \tag{5.12}$$

where $\delta X_{\text{inst},i}$ and $\delta X_{\text{inst},j}$ have identical probability distributions with the same variance. Since only time-independent effects are considered, the overall covariance of two measurements X_i, X_j is approximated as:

$$\mathrm{cov}(X_i, X_j) \approx u^2_{\text{inst}} \cdot r(X_i, X_j). \tag{5.13}$$

The evaluation of the instrumental correlation coefficient $r(X_i, X_j)$ between different measurements of the same device should be in principle based on the analysis of common components in the uncertainty budget for different calibration points. Since the information on calibration budgets is not accessible to every user, a minimum correlation is estimated by considering that the same standard has been used for the entire calibration range. Regardless the instrument, this correlation is the minimum one that should be considered for measurement results traceable to the SI. Under this assumption, the correlation coefficient for any two measurements of the same

instrument is estimated as [138]:

$$r_{\min}(X) = \frac{u^2_{\text{ref}}(X)}{u^2_{\text{inst}}(X)}, \tag{5.14}$$

where $u^2_{\text{ref}}(X)$ is the squared uncertainty of the reference measurement standard for quantity X, which is used to calibrate the instrument. As mentioned, the instrumental uncertainty, from now on indicated by $u(X)$, is chosen based on a priori metrological knowledge, calibration certificates (when available) or datasheet specifications.

5.2.2 Uncertainty budget and effective sensitivity coefficients

By manipulating Equation 5.10 it is possible to group in a convenient way the uncertainty of the measurements associated with the same quantity and express the flux uncertainty as:

$$u^2(F) = u^2_P(F) + u^2_T(F) + u^2_S(F) + u^2_{V_x}(F) + u^2_{V_y}(F) + u^2_{V_z}(F) + c_{V_zV_yV_z}(F). \tag{5.15}$$

The $u^2_X(F)$ are the uncertainty contributions to the overall flux uncertainty due to measurements of X, whereas $c_{V_xV_yV_z}(F)$ is the covariance contribution coming from the instrumental correlation between the three wind components, which are the only quantities measured with the same instrument. Indeed, the anemometer is usually calibrated against the same reference standard for all three directions, implying that the measurements for different wind speed components are instrumentally correlated. The covariance resulting from this correlation has been separated to understand its influence on the flux uncertainty evaluation because it can be negative and decrease the overall uncertainty budget. For $X \in \{P, T, S, V_x, V_y, V_z\}$ the uncertainty contributions result from the application of the law of uncertainty propagation, which leads to:

$$u^2_X(F) = \left[\sum_{i=1}^{N_X} \left(\frac{\partial F}{\partial X_i} \right)^2 + r(X) \sum_{i=1}^{N_X} \sum_{j \neq i}^{N_X} \frac{\partial F}{\partial X_i} \frac{\partial F}{\partial X_j} \right] u^2(X) = \overline{J_X}^2 u^2(X). \tag{5.16}$$

The scalar term $\overline{J_X}$ is here defined as an effective sensitivity coefficient. It is effective because sensitivity coefficients are generally computed as $\partial F / \partial X_i$ and quantify how sensitive the output values are with respect to variation of input measurements. In this case, with N_X measurements of X, it is convenient to quantify an overall

sensitivity coefficient ($\overline{J_X}$) which expresses the sensitivity with respect to all the measurements of the same quantity. In the case of identical uncertainty for the three axes of the anemometer, the covariance contribution $c_{V_xV_yV_z}(F)$ can be written defining a corresponding effective sensitivity coefficient, which is computed as:

$$c_{V_xV_yV_z}(F) = 2\, r_{\text{anem}}\Big(\sum_{i=1}^{N_X}\sum_{j=1}^{N_X}\frac{\partial F}{\partial V_{x,i}}\frac{\partial F}{\partial V_{y,j}} + \sum_{i=1}^{N_X}\sum_{j=1}^{N_X}\frac{\partial F}{\partial V_{x,i}}\frac{\partial F}{\partial V_{z,j}} + \sum_{i=1}^{N_X}\sum_{j=1}^{N_X}\frac{\partial F}{\partial V_{y,i}}\frac{\partial F}{\partial V_{z,j}}\Big)u^2_{\text{anem}} = \overline{H}_{V_xV_yV_z}u^2_{\text{anem}} \quad (5.17)$$

The term u_{anem} is the chosen uncertainty for the anemometer, which in general is a single value identical for the three axes ($u(V_x) = u(V_y) = u(V_z) = u_{\text{anem}}$). A similar consideration can be made for the correlation coefficients among different components which are assumed identical: $r(V_x) = r(V_y) = r(V_z) = r(V_x,V_y) = r(V_x,V_z) = r(V_y,V_z) = r_{\text{anem}}$. The exact analytical formulation of the partial derivatives, i.e. the sensitivity coefficients, and the derivation of the effective sensitivity coefficients are reported in Appendix A.2. Table 5.2 shows an example of the uncertainty budget for a single flux measurement, compiled by using the defined quantities. In the absence of an effective sensitivity coefficient, a sensitivity value shall be associated with each measurement for a given instrument, that is impractical. Note also that for the cross-covariance $c_{V_xV_yV_z}$ only the square root of the absolute value is reported, along with the same sign as $\overline{H}_{V_xV_yV_z}$.

5.2.3 Uncertainty of the annual carbon budget

Eddy Covariance measurements are mainly used to estimate the balance of a certain quantity emitted to or absorbed from the atmosphere. Specifically, CO_2 fluxes are used to estimate the annual budget of carbon exchanged between the atmosphere and a forest, a cropland or any other ecological interesting surface. Mathematically, the budget is simply obtained by the cumulative sum of the fluxes integrated over the sampling interval:

$$NEE = \Delta t \cdot m_C \cdot \sum_i F_i. \quad (5.18)$$

The budget is indicated with the acronym Net Ecosystem Exchange (NEE) and is expressed in $\mathrm{g/m^2}$; Δt corresponds to the sampling interval expressed in s, whereas

Table 5.2 Example of uncertainty budget for a single EC flux measurement. The reported values are specific for the considered 30 min sampling interval, corresponding to a CO_2 flux of $-1.77\,\mu\text{mol}\,\text{m}^{-2}\,\text{s}^{-1}$.

X	$u(X)$	$\overline{J_X}(\overline{H}_{V_xV_yV_z})$	$u_X(F)(\sqrt{\lvert c_{V_xV_yV_z}\rvert})$
P	30 Pa	$2.08\times10^{-4}\,\mu\text{mol}\,\text{m}^{-2}\,\text{s}^{-1}\,\text{Pa}^{-1}$	$0.0062\,\mu\text{mol}\,\text{m}^{-2}\,\text{s}^{-1}$
T	0.05 °C	$1.76\times10^{-4}\,\mu\text{mol}\,\text{m}^{-2}\,\text{s}^{-1}\,°\text{C}^{-1}$	$8.8\times10^{-6}\,\mu\text{mol}\,\text{m}^{-2}\,\text{s}^{-1}$
S	4 ppm	$0.05\,\text{mol}\,\text{m}^{-2}\,\text{s}^{-1}$	$0.20\,\mu\text{mol}\,\text{m}^{-2}\,\text{s}^{-1}$
V_x	0.1 m/s	$0.15\,\mu\text{mol}\,\text{m}^{-3}$	$0.15\,\mu\text{mol}\,\text{m}^{-2}\,\text{s}^{-1}$
V_y	0.1 m/s	$0.14\,\mu\text{mol}\,\text{m}^{-3}$	$0.014\,\mu\text{mol}\,\text{m}^{-2}\,\text{s}^{-1}$
V_y	0.1 m/s	$0.69\,\mu\text{mol}\,\text{m}^{-3}$	$0.069\,\mu\text{mol}\,\text{m}^{-2}\,\text{s}^{-1}$
$V_xV_yV_z$	0.1 m/s	$0.15\,\mu\text{mol}^2\,\text{m}^{-6}$	$0.019\,\mu\text{mol}\,\text{m}^{-2}\,\text{s}^{-1}$
F	-	-	$0.21\,\mu\text{mol}\,\text{m}^{-2}\,\text{s}^{-1}$

m_C is the molar mass of carbon expressed in g/µmol since carbon fluxes are generally expressed in $\mu\text{mol}\,\text{m}^{-2}\,\text{s}^{-1}$. Only good-quality data are used for the purpose, corresponding to half-hourly intervals for which conditions of stationarity and well-developed turbulence are satisfied. The discarded measurements shall be replaced with estimates obtained with a model to ensure non-biased quantification of the budget. Several techniques were developed for "gap-filling" the removed data using physical or data-driven models.

Beyond the measurement uncertainty described above, the one associated with the gap-filled values is also rarely taken into account. On average, approximately one-third of the measurements are removed from the flux time series [99] and replaced with modeled values, whose uncertainties may be significantly larger than those of the remaining measurements. Furthermore, the gap-filling models are trained with good-quality measurements that have their own uncertainty, which in turn should be propagated for evaluating the uncertainty of the model predictions. The impact of instrumental uncertainty on an annual carbon budget is therefore evaluated by performing a gap-filling procedure, for the considered case study, with a simple and widely accepted [139, 99] light-curve model based on input short-wave radiation SW_{in}:

$$F = \frac{\alpha\,\beta\,SW_{in}}{\alpha\,SW_{in}+\beta} + \gamma \tag{5.19}$$

The model parameters α, β and γ are calculated for 1-month time windows, using only half-hourly records that passed the turbulent quality check (QC) system. The non-linear weighted least-squares (NWLS) fitting procedure described in [140] is applied since allows to use the uncertainty of flux measurements as weights and to evaluate the uncertainty of the calculated parameters. The uncertainty of the modeled flux estimates is then obtained by further propagating the uncertainty of the coefficients and of the input quantity SW_{in}, taking also into account the dispersion of values around the fitted curve. In the end, knowing the uncertainty associated with each half-hourly flux estimate, it is possible to quantify the uncertainty related to the annual carbon budget.

Another important overlooked aspect is the correlation between different measured fluxes. Even if the fluxes are measured in disjoint integration intervals, thus associated with different turbulent conditions, the use of the same instruments leads to correlation between different half-hourly flux measurements which shall be considered in the uncertainty evaluation of the carbon budget. The covariance between two flux estimates F_1 and F_2 can be expressed as:

$$\begin{aligned}\operatorname{cov}(F_1,F_2) = &\sum_{X\in\{F,T,S\}}\sum_{ij}\frac{\partial F_1}{\partial X_{1,i}}\frac{\partial F_2}{\partial X_{2,j}}r(X)u^2(X)\\ &+\sum_{X\in\{V_x,V_y,V_y\}}\sum_{Y\in\{V_x,V_y,V_y\}}\sum_{ij}\frac{\partial F_1}{\partial X_{1,j}}\frac{\partial F_2}{\partial Y_{2,j}}r_{\text{anem}}u^2_{\text{anem}}(X)\end{aligned} \quad (5.20)$$

The above equation must be evaluated for each couple of fluxes for assembling the complete variance-covariance matrix of flux measurements. Note that for the anemometric variables it is necessary to also consider the flux cross-covariances related to different wind speed components sampled in different time intervals. Eventually, the uncertainty of the annual budget is calculated simply as:

$$u^2(NEE) = \sum_i u^2(F_i) + \sum_{ij}\operatorname{cov}(F_i,F_j) \quad (5.21)$$

In the case of instrumentally independent flux measurements, only the first sum is considered, thus covariances are neglected. Moreover, the covariance between the measured and modeled values is here assumed to be zero since complex to estimate due to the non-linearity of the model and the iterative process used to optimize the cost function. To show the importance of measurement uncertainty and instrumental

correlation in the gap-filling procedure, a comparison is performed between the annual carbon budgets estimated assuming independent, perfectly correlated and minimally correlated measurements, whose results are shown in the next section. Furthermore, it will be shown how the fit of Equation 5.19 can vary when flux measurement uncertainty is neglected compared to the case in which it is considered in the estimation of model parameters.

5.3 Case study: San Rossore 2

For testing the equations described above, the uncertainty framework has been applied to the 2022 measurements of an EC station monitoring GHGs fluxes from a Mediterranean evergreen forest, the San Rossore 2 ICOS Class 2 Ecosystem Station, located in Central Italy between the city of Pisa and the Tyrrhenian See (coordinates: $43°43'40.30''N$ $10°17'3.98''E'$ (https://meta.icos-cp.eu/resources/stations/ES_IT-SR2). The ICOS ecosystem stations network ensures that measurements are collected with a high level of standardization, comparability among sites and superior technical standards. Moreover, all the raw and elaborated data and metadata are highly structured and harmonized and freely accessible through the FAIR (Findable, Accessible, Interoperable and Reusable) principle. This allows every reanalysis, also those including the raw data recorded on field like the one presented here, to be easily and universally applicable to the entire network, facilitating scientific research and exchange. Specifically, the San Rossore 2 is a Class 2 ICOS labelled station in a forest, hence all the collected variables, measurement setups and processing schemes are fully described in the ICOS documentations and related publications [107, 108]. For the purpose of this analysis, the six variables which the flux model of Equation 5.6 is built upon, have been measured through the following instrumentations: LI-COR LI7200 infrared gas analyzer (LI-COR Inc., Lincoln, NE, USA) for S, a Gill HS-50 ultrasonic tri-axial anemometer (Gill Ltd, Lymington, Hampshire, UK) for V_x, V_y, V_z, a HMP155 thermohygrometer (Vaisala Oyj, Finland, EU) for T and a PTB110 barometer (Vaisala Oyj, Finland, EU) for P. The most important structural and ecological features of the forest, characterized by the dominance of domestic pines (Pinus pinea L., 1752) as dominant species with average height of 19 m at the time of measurements, are described in detail in [141]. It is of relevance here that the half-hourly estimates of CO_2 fluxes, provided by ICOS lie in the range

$-40\,\mu mol\,m^{-2}\,s^{-1}$ to $30\,\mu mol\,m^{-2}\,s^{-1}$, where by micrometeorological convention negative flux values correspond to carbon uptake driven by photosynthesis and positive values to release by respiratory and metabolic processes. The annual NEE measured since 2013 in the forest lie in the range $-537.5\,g/m^2$ to $-68.8\,g/m^2$, with a nominal value for 2022 estimated by ICOS as $-190.6\,g/m^2$ [142, 143]. The radiation variable used for the modelling and the gap-filling of the missing measurement records (see Equation 5.19), is the Short Wave Incoming Radiation (SW_{in}) and it is measured by a CNR-4 Pyrgeometer (Kipp&Zonen, Delft, The Netherlands, EU).

Reasonable values of instrumental uncertainties and minimum correlation coefficients considered for the application of the proposed method are listed in Table 5.3. These values results from checking the specifications of the instruments composing the Eddy Covariance measurement system. Just to report an example, the accuracy specification of the gas analyzer, 1 % of reading, was interpreted as a standard uncertainty. Since for simplicity only homoscedastic uncertainties are considered, the

Table 5.3 Instrumental uncertainty and minimum correlation coefficient considered for each instrument.

	P	T	S	V_x	V_y	V_z
$u(X)$	30 Pa	0.05 °C	4 ppm	0.1 m/s	0.1 m/s	0.1 m/s
$r_{\min}(X)$	0.0625	0.04	0.36	0.0625	0.0625	0.0625

percentage was applied to a reference value of 400 ppm to retrieve a reasonable value for the IRGA instrumental uncertainty, which is 4 ppm in this case study. A similar reasoning can be done for the barometer and the anomemeter which usually have a measurement uncertainty expressed in relative terms. Concerning the minimum correlation coefficients, not much information is available for their determination from the calibration process employed by the manufacturers. For the anemometer information coming from the air speed key comparison in [144] was used to determine the typical ratio between the reference instrument and the device under calibration. For the barometer was used the database of accredited laboratories in Italy (Accredia website) and at their calibration capabilities. For all calibration laboratories, the reference instrument had an uncertainty of about one fourth of the final uncertainty, thus giving 0.0625 as correlation coefficient. Then, a similar procedure was followed for the IRGA, finding useful information in the standard [145] for the verification of gas concentration in tanks used for calibration of analyzers, giving a correlation value of 0.36. In the end, only for thermometers' calibration more information was

available given the metrological experience in the sector: an expanded uncertainty ($k = 2$) of 0.01 °C covering the entire calibration range is reasonable for secondary standards used to calibrate meteorological thermometers. Depending on the employed calibration set-up the final calibration uncertainty can vary, however a value of 0.05 °C of expanded uncertainty ($k = 2$) is reasonable for accredited laboratories or manufacturers that calibrate a large volume of instruments per year. Then, the correlation coefficient is just $(0.01/0.05)^2 = 0.04$.

The developed framework for uncertainty propagation is valid for any type of correlation between instruments, hence a comparison between the implementations of the method for the following three cases has been made:

a) Independent measurements, obtained for $r(X) = 0, \forall X$.

b) Perfectly correlated measurements for $r(X) \approx 1, \forall X$. A value of $r(X) = 0.99$ is chosen to avoid numerical problems due to singular matrices, which does not compromise the results of this work.

c) Minimum correlation of measurements for values of $r_{\min}(X)$ reported in Table 5.3 and quantified as described above using Equation 5.14.

Before applying the methodology and analyzing the calculated uncertainties, it is necessary to test its validity. Indeed, the suggested method for uncertainty propagation is based on the linearization of the measurement model at a given point determined by measurements. The model has a non-linear nature, as highlighted by the rotation coefficients reported in Equations 5.7-5.9, hence shall be validated by a numerical method such as Monte Carlo (MC) [146]. Due to the large number of involved measurements (18000 records for the gas concentration and each wind speed component) and the computational complexity of Monte Carlo, it is unfeasible to verify the linear propagation for each flux measurement that will be considered in the case study, covering an entire year of measurements (17520 sampled fluxes). Therefore, a comparison between the Monte Carlo simulation and the developed analytical method has been performed only for some representative flux measurements that respect the assumptions of the techniques, hence taken under turbulent conditions and identified by using the Monin-Obukhov (MO) stability parameter available from the ICOS dataset. Both the independent and correlation assumptions have been tested, using the same number of MC samples, 10^5, corresponding to the

upper limit imposed by the hardware available at the time of the simulation. For the second assumption, only the minimum correlation has been tested, which is the more realistic and less degenerate case.

After validation, the analytical method was applied to calculate the flux uncertainties of the 2022 time series. The distribution of absolute and relative flux uncertainties has been analyzed to understand the typical values associated to different magnitudes of the flux. As for the gap-filling, only good-quality data that passed the QC, based on the nonstationarity ratio test [147] and instrumental diagnostics, were considered for the analysis. The calculated flux uncertainty has also been used to detect other lower quality data corresponding to outliers in the uncertainty time series detected by a simple gradient threshold algorithm. These apparent spikes could be due to both methodological aspects, normally related to low degrees of stationarity or underdeveloped turbulence not detected by the QC-method, or to other transient instrumental effects not detected by the diagnostic routine.

In the end, the gap-filling procedure is applied, as described in Subsection 5.2.3, for comparing the annual carbon budget and the associated uncertainty obtained for the different correlation cases. First the results are reported, then they are discussed putting the attention on the role of input instrumental uncertainty.

5.3.1 Results

The law of propagation of uncertainty (LPU) was applied to the time series of San Rossore 2 for each available flux measurement. For instance, Figure 5.5 is a zoom of the time series where the calculated uncertainties, under the independence assumption, are superimposed on the estimated fluxes considering a coverage factor $k = 2$.

The evaluated uncertainties were validated by comparison with MC simulations for flux measurements conforming to all the assumptions of the EC technique. Since the impossibility of performing the comparison for each measurement considered in the case study due to the extensive computation time, three MC simulations were performed for the records corresponding to the timestamps reported in Table 5.4. Starting from the first row, the MO stability parameter was evaluated to be respectively -2.58, -0.105 and -0.598. The values are compatible with unstable atmospheric conditions [97], hence the chosen measurements are suited for com-

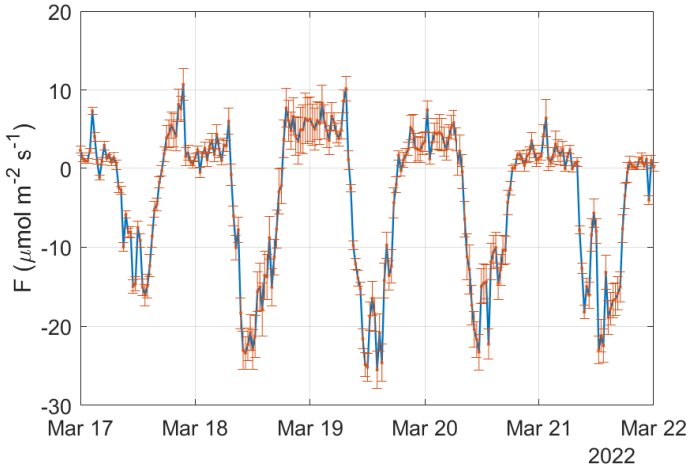


Fig. 5.5 Portion of the time series of CO_2 flux measurements. The bars correspond to the calculated expanded uncertainties, for a coverage factor $k = 2$, assuming independent measurements.

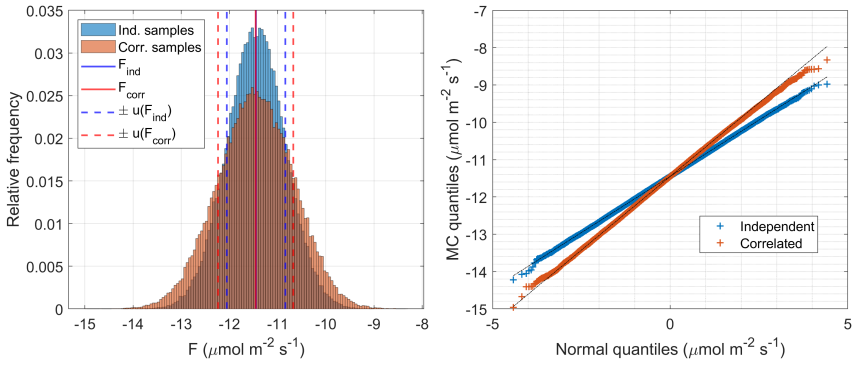
parison with MC. Focusing on the numerical results, the difference between the independence and correlation assumptions, investigated in detail below, is similar for LPU and MC, proving that the proposed method can catch their differences. The flux estimates between LPU and MC are consistent within three significant digits, whereas the evaluated uncertainties have approximately two common significant digits, with just a few exceptions (as for the third row under the independent assumption). These results, even if partial and related to only some flux measurements, suggest the validity of the proposed model under both the assumption of independence and correlation of measurements. However, the numerical results say little about the shape of the probability distribution associated with the measurement. The histograms of the MC samples are therefore plotted, as shown in Figure 5.6, along with the associated Q-Q plot. The distribution can reasonably be considered Gaussian in all cases as also proved by the empirical quantiles following a linear relation with the normal ones. Likely, this is due to the low values of input uncertainties compared to the degree of model nonlinearities, which allows to validate the LPU approach. Furthermore, the normality of the distribution also determines the coverage probability corresponding to $k = 2$, which is about 95 %. Nevertheless, only standard uncertainties will be considered for the analysis of the case study, following a conservative approach because of the few cases tested by MC comparison.

Table 5.4 Results of the comparison between the Law of Propagation of Uncertainty (LPU) applied to the Eddy Covariance technique and the corresponding Monte Carlo (MC) simulation. Three flux measurements are considered, associated with the reported timestamps, corresponding to atmospheric turbulent conditions.

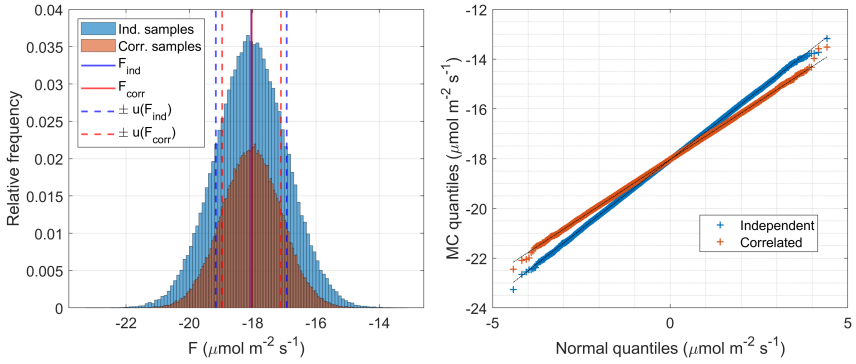
Timestamp	Assumption	$F/(\mu\text{mol}\,\text{m}^{-2}\,\text{s}^{-1})$		$u(F)/(\mu\text{mol}\,\text{m}^{-2}\,\text{s}^{-1})$	
		LPU	MC	LPU	MC
03/09/2022 11:00	Independence	-11.450	-11.449	0.594	0.606
	Correlation	-11.450	-11.457	0.784	0.780
07/11/2022 12:30	Independence	-18.039	-18.044	1.111	1.119
	Correlation	-18.039	-18.036	0.928	0.930
10/08/2022 11:30	Independence	-6.628	-6.627	0.633	0.643
	Correlation	-6.628	-6.622	0.507	0.516

The frequency distribution of absolute and relative uncertainties, computed by using the LPU approach considering independent or correlated measurements, is visualized in Figure 5.7. For all assumptions, the distribution is similar to a χ^2 distribution with a main peak (the mode) and a relatively long tail to the right, with values of the relative uncertainty well exceeding the threshold of 100 %. Since the distribution is not symmetrical, with values in the tail that heavily influence the mean, the median has been assumed as the best metric for estimating values representative of each uncertainty component that have similar distributions [148]. The median of absolute and relative uncertainties is equal to $0.51\,\mu\text{mol}\,\text{m}^{-2}\,\text{s}^{-1}$ and 6.2 % for independent measurements, respectively, while in the case of perfectly correlated measurements the median attains values of $0.67\,\mu\text{mol}\,\text{m}^{-2}\,\text{s}^{-1}$ and 8.8 %. On the other hand, the median of the distribution for minimally correlated measurements is closer to the independent case with $0.53\,\mu\text{mol}\,\text{m}^{-2}\,\text{s}^{-1}$ for absolute uncertainty and 7.7 % for the relative one.

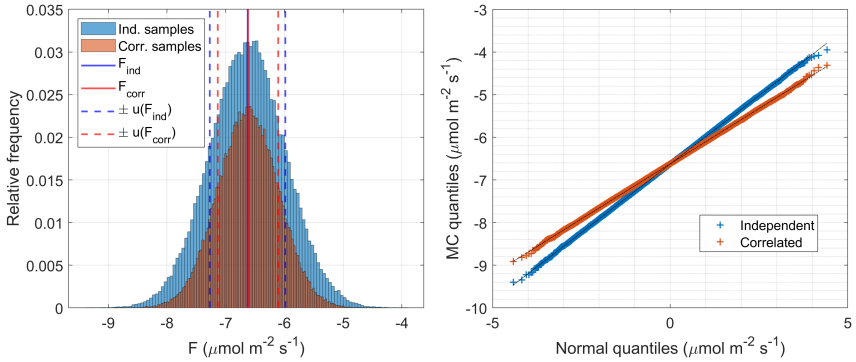
To analyze how the uncertainty terms vary with the magnitude of the measured flux, the measurements have been binned, with similar numerosity, with respect to flux values, and the median of the uncertainty components has been also calculated for each bin. These results are shown in Figure 5.8, whose plots on the left report medians of absolute uncertainties, whereas on the right those of the relative ones. Subplots 5.8a, 5.8b and 5.8c are related to independent measurements, perfect correlation or minimal correlation, respectively. In the case of independent measurements, the flux measurement uncertainty varies from a minimum absolute value of ap-



(a) 03/09/2022 11:00



(b) 07/11/2022 12:30



(c) 10/08/2022 11:30

Fig. 5.6 The left panels show the distribution of the MC samples under the independent (blue) and correlation assumption (orange) for each measurement involved in the MC validation. Solid lines correspond to the estimated fluxes, whereas the dashed lines are related to the evaluated standard uncertainties. The right panels show the Q-Q plots used to verify the normality of the distributions.

proximately $u(F) = 0.22\,\mu\text{mol}\,\text{m}^{-2}\,\text{s}^{-1}$ to a maximum of $u(F) = 0.99\,\mu\text{mol}\,\text{m}^{-2}\,\text{s}^{-1}$, proportional to the flux magnitude. The combined relative uncertainty has the opposite trend and it is inversely proportional to flux magnitudes: it attains values lower

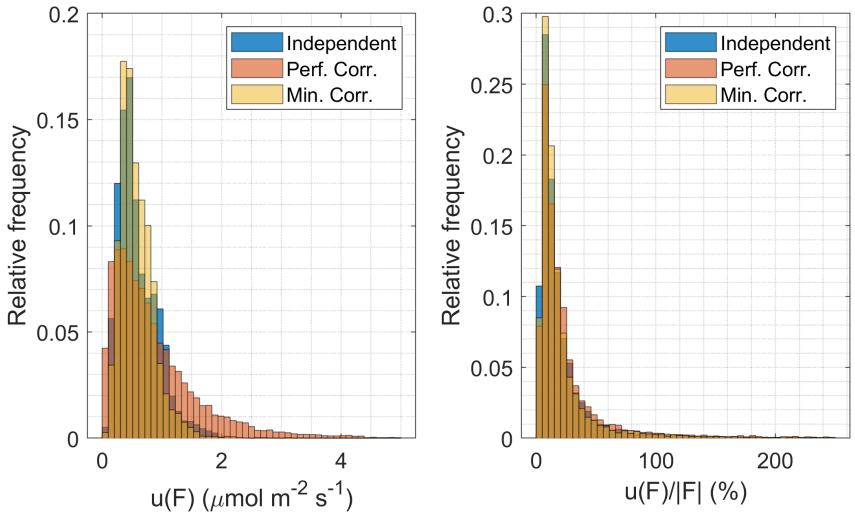


Fig. 5.7 Distribution of absolute (left) and relative (right) uncertainties evaluated under different assumptions: independent (blue), perfectly correlated (orange) and minimally correlated measurements (yellow).

than $u_r(F) = 10\,\%$ and decreases slowly with increasing flux magnitudes above a threshold flux of approximately $F = \pm 6\,\mu\text{mol}\,\text{m}^{-2}\,\text{s}^{-1}$. When the flux magnitude decreases toward smaller values and tends to zero, the relative measurement uncertainty sharply rises up to approximately 205 % for the smallest flux magnitudes. For almost the whole measurement range, the combined uncertainty is not distinguishable from the contribution of the concentration measurements alone. Only another term is not completely negligible, that is the contribution from the measurement of the vertical wind speed: $0.02\,\mu\text{mol}\,\text{m}^{-2}\,\text{s}^{-1} \leq u_{V_z}(F) \leq 0.09\,\mu\text{mol}\,\text{m}^{-2}\,\text{s}^{-1}$ The term $u_{V_z}(F)$ rises to values comparable to the contribution from concentration measurements, $u_S(F)$, only for $F > 10\,\mu\text{mol}\,\text{m}^{-2}\,\text{s}^{-1}$.

Figure 5.8b shows the case of perfect instrument correlation. Focusing first on absolute uncertainties, the minimum combined value decreases, compared to the independent case, to $0.18\,\mu\text{mol}\,\text{m}^{-2}\,\text{s}^{-1}$ whereas the maximum one increases to $2.4\,\mu\text{mol}\,\text{m}^{-2}\,\text{s}^{-1}$. The uncertainty shows again an increasing trend with flux values, with different weights associated with the uncertainty components: here the main contribution comes from the vertical wind speed, which is almost not distinguishable from the overall uncertainty. The plot also shows the covariance term (label '$V_xV_yV_z$') resulting from the correlation of different wind speed components. Since the covariance can be negative, to fairly compare with $u_X(F)$ the following quantity has been plotted: $\text{sgn}(c_{V_xV_yV_z})\sqrt{|c_{V_xV_yV_z}|}$. The term is mainly negative, with

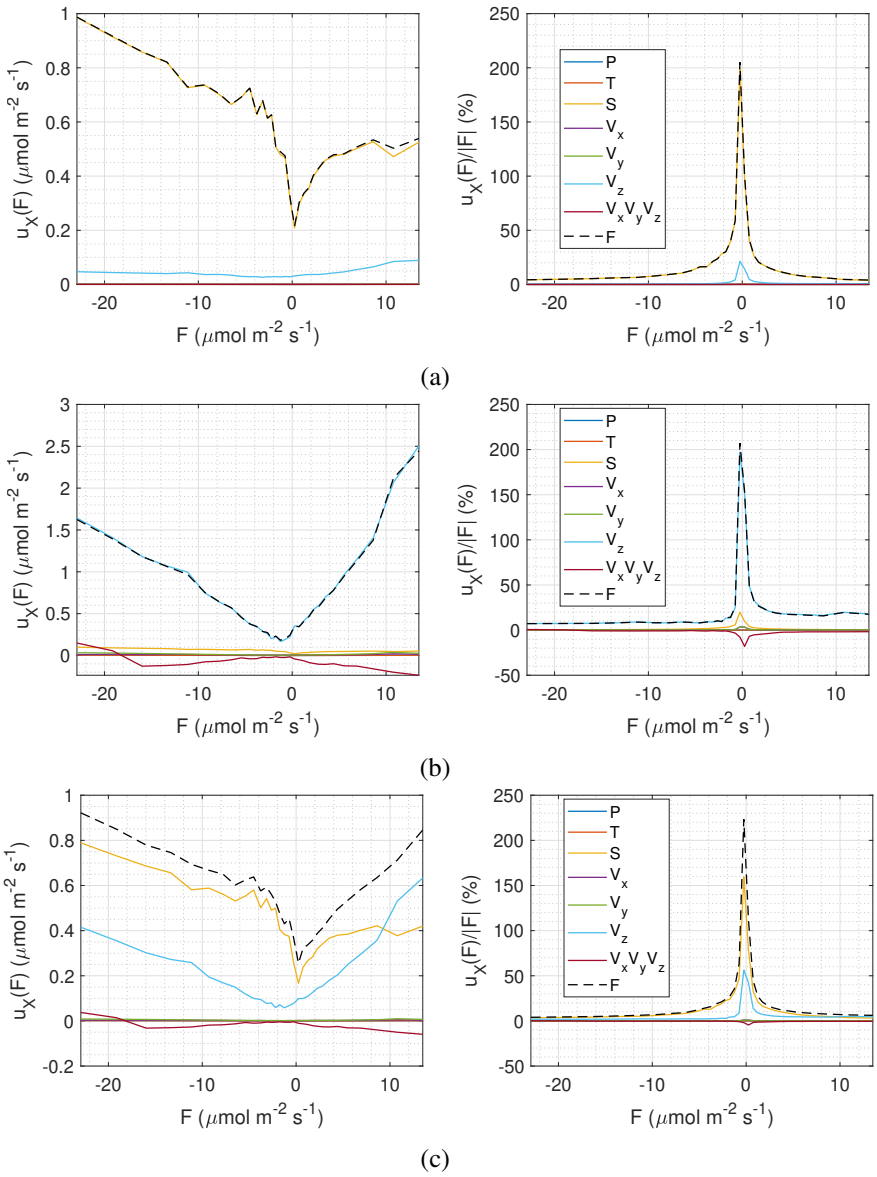


Fig. 5.8 Median of the absolute (left) and relative (right) uncertainties components, binned with respect to flux values. The shown curves are related to the distribution of $u_x(F)$, with X reported in the legend. The curve corresponding to $V_xV_yV_z$ refers to the distribution of the covariance ($c_{V_xV_yV_z}$) between the three anemometer components. Since it can be negative, the distribution is associated to the quantity $\mathrm{sgn}(c_{V_xV_yV_z})\sqrt{|c_{V_xV_yV_z}|}$ to make a fair comparison with $u_x(F)$. The different panels correspond to the different assumptions considered in this work: (a) independent measurements, (b) perfectly correlated measurements and (c) minimal correlated measurements.

positive values only for $F < -18\,\mu\mathrm{mol\,m^{-2}\,s^{-1}}$, and increases with flux magnitude. Even though its effect is to decrease the combined uncertainty, the values are an order

of magnitude smaller with respect to $u_{V_z}(F)$, thus no clear impact is visible. For what concerns the relative uncertainties, the combined uncertainty still decreases with flux values within $\pm 2\,\mu\text{mol}\,\text{m}^{-2}\,\text{s}^{-1}$, then it stays between 7 to 10 % for negative fluxes and between 15 to 25 % for positive fluxes. Also in this case, when the flux magnitude decreases toward smaller values and tends to zero, the relative measurement uncertainty increases sharply, rising up to approximately 207 % for the smallest flux magnitudes.

Results for the third assumption of partially correlated measurements are finally reported in Figure 5.8. The minimum and maximum of the combined absolute uncertainty are close to the independent case: $0.25\,\mu\text{mol}\,\text{m}^{-2}\,\text{s}^{-1}$ and $0.92\,\mu\text{mol}\,\text{m}^{-2}\,\text{s}^{-1}$, respectively. The combined relative uncertainty is visibly different only for the maximum, approximately equal to 223 %, slightly higher than the case of independent measurements. For this assumption the contributions of vertical wind speed and gas concentration are both not negligible. The uncertainty due to the gas analyzer is the largest for negative fluxes, whereas for positive ones the uncertainty component due to vertical wind speed increases and becomes the dominant one for $F > 9\,\mu\text{mol}\,\text{m}^{-2}\,\text{s}^{-1}$. The other components have a smaller order of magnitude, with the wind speed covariance being negative for $F > -19\,\mu\text{mol}\,\text{m}^{-2}\,\text{s}^{-1}$.

Table 5.5 reports the median of the effective sensitivities for each variable and for the three assumptions considered in this work. The largest sensitivity is always connected, by far, to the measurement of the vertical wind speed, regardless of the correlation level of the measurements. After the vertical wind speed, gas concentration has the largest sensitivity when the measurements are independent or minimally correlated, whereas the horizontal wind speed components become predominant in the perfectly correlated case. In the latter case, the sensitivity to the concentration measurement drops to the second last place in the sensitivity ranking due to the nature of relative measurements entering into the calculations of covariances (ideally whould be zero for correlation coefficient equal to 1). Eventually, the atmospheric pressure shows the smallest sensitivity coefficient regardless of the correlation level.

Figure 5.9 shows an example of fit for the model expressed in Equation 5.19 using the measurements performed in April 2022. The black curve is obtained by a Non-linear Ordinary Least-Squares (NOLS) procedure, which does not take into account the measurement uncertainty of individual flux values, whereas the red one is obtained using a Non-linear Weighted Least-Squares (NWLS) method, which does

Table 5.5 Median of the effective sensitivities for each assumption considered in the propagation of measurement uncertainty.

	Independent	Perf. Corr.	Min. Corr.
$\mathrm{Me}[\overline{J_P}]/(\mu\mathrm{mol\,m^{-2}\,s^{-1}\,Pa^{-1}})$	4.6×10^{-6}	3.6×10^{-5}	1.0×10^{-5}
$\mathrm{Me}[\overline{J_T}]/(\mu\mathrm{mol\,m^{-2}\,s^{-1}\,{}^{\circ}C^{-1}})$	0.0016	0.012	0.0030
$\mathrm{Me}[\overline{J_S}]/(\mathrm{mol\,m^{-2}\,s^{-1}})$	0.12	0.0093	0.093
$\mathrm{Me}[\overline{J_{V_x}}]/(\mu\mathrm{mol\,m^{-3}})$	0.0054	0.093	0.025
$\mathrm{Me}[\overline{J_{V_y}}]/(\mu\mathrm{mol\,m^{-3}})$	0.0068	0.11	0.030
$\mathrm{Me}[\overline{J_{V_z}}]/(\mu\mathrm{mol\,m^{-3}})$	0.43	7.2	1.9
$\mathrm{Me}[\overline{H}_{V_xV_yV_z}]/(\mu\mathrm{mol^2\,m^{-6}})$	0	-0.51	-0.032

account for uncertainty during parameters' calculation. Table 5.6 reports the values of the model coefficients estimated by the different models. The parameters estimates are significantly different: with respect to the NOLS model, the relative variation of α, β and γ is of 31 %, -6.8 % and -38 % respectively. Moreover, the relative uncertainty of the coefficients is strongly reduced in the NWLS case: it changes from 42 %, 5.9 % and 57 % to 5.3 %, 3.3 % and 8.4 % for α, β and γ, respectively, with relative variations of -87 %, -44 % and -85 %.

Table 5.6 Estimated coefficients and corresponding relative uncertainties for the light-curve model, reported in Equation 5.19, using two different procedures: Non-linear Ordinary Least-Squares (NOLS) and Non-linear Weighted Least-Squares (NWLS) method.

Fit parameter	NOLS	NWLS
$\alpha/(\mu\mathrm{mol\,W^{-1}})$	$-0.073(42\,\%)$	$-0.046(5.3\,\%)$
$\beta/(\mu\mathrm{mol\,m^{-2}\,s^{-1}})$	$-33.6(5.9\,\%)$	$-35.9(3.3\,\%)$
$\gamma/(\mu\mathrm{mol\,m^{-2}\,s^{-1}})$	$4.85(57\,\%)$	$3.03(8.4\,\%)$

Since the NWLS leads to smaller uncertainties for the model parameters, it has been considered in the final part of the measurement uncertainty implementation for filling the gaps in the time series. The cumulative annual NEE obtained by combining measured and NWLS-modeled fluxes is shown on the left of Figure 5.10, whereas the cumulative uncertainty of NEE, obtained by complementing the measurement uncertainty with the NWLS-propagated uncertainty for missing records, is shown on the right. A comparison of the different assumptions about the independence of measurements is done, especially for analyzing the different

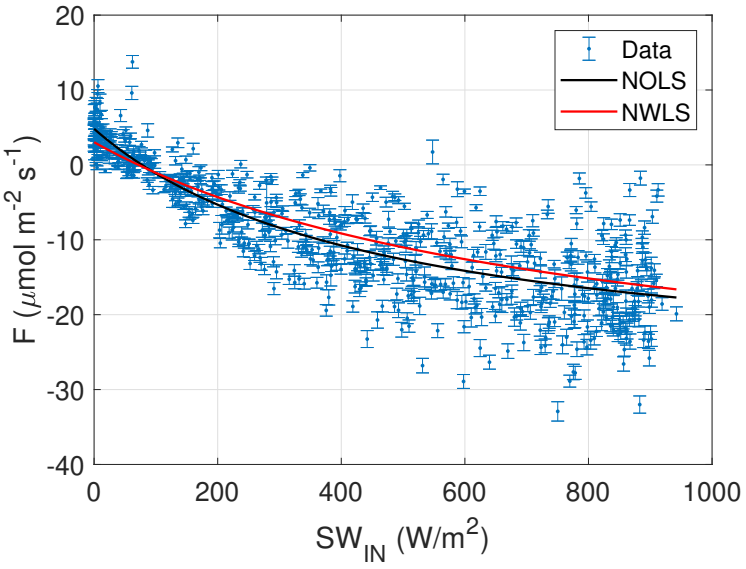


Fig. 5.9 Fit of the light curve model (Equation 5.19) using measurements, performed during April 2022, that passed the QC. The fit in black is obtained through a Non-linear Ordinary Least-squares (NOLS) procedure, whereas that in red is obtained considering a Non-linear Weighted Least-Squares (NWLS) procedure, with weights (shown in blue) corresponding to the inverse of the evaluated standard measurement uncertainties.

behavior of $u(NEE)$. The cumulative uncertainty obtained by applying the method described by Finkelstein and Sims (F&S) [113] is also reported as a reference value. The cumulative NEE estimates show the same time behavior, with a final budget varying from $-415.6\,\mathrm{g\,m^{-2}}$ for the independent case to $-447.1\,\mathrm{g\,m^{-2}}$ for the case of perfect correlation. The numerical values of the annual budget for each considered assumption are reported in Table 5.7: the perfect correlation leads to the largest uncertainty of $39.6\,\mathrm{g\,m^{-2}}$, which is a 8.9 % of the corresponding budget. The smallest uncertainty is for the independent case: $5.6\,\mathrm{g\,m^{-2}}$, 1.3 % of the estimated budget, which is comparable to the F&S estimate. Eventually, the assumption of minimum correlation leads to an uncertainty of $14.3\,\mathrm{g\,m^{-2}}$, 3.3 % of the budget calculated under the same assumption.

5.3.2 Discussion

The frequency distribution and statistical analysis of the combined measurement uncertainty reported in Figure 5.7 suggests the non-negligible presence of large uncertainties in the Eddy Covariance flux estimates, as evidenced by the positive

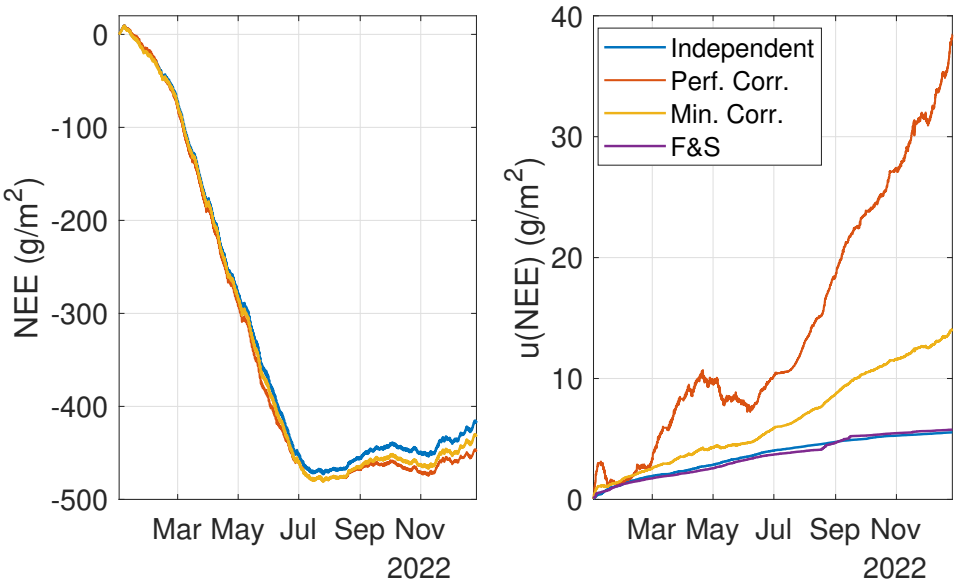


Fig. 5.10 Annual cumulative carbon budget with the associated cumulative uncertainty. The annual carbon budget is reported in the left panel for the three different options, that is blue for independent measurements, red for perfect correlation and yellow for minimum correlation. The uncertainty of NEE calculated by simply summing in quadrature the flux uncertainties is shown in the right panel with the same colors, with the violet line representing the F&S sampling error that is reported as a reference value.

skewness of the distribution. However, the characteristic median is always below the values of 1 µmol m^{-2} s^{-1} and 10 % for absolute and relative expression, respectively. The overall trends of absolute and relative combined measurement uncertainties can be considered similar, with a sharp increase (decrease) of the relative (absolute) uncertainty for small fluxes and vice versa for larger flux magnitudes. The most significant insights can be found in the different behavior of the contributions from the measurements of individual variables when independent or perfectly correlated measurements are considered. In the case of independent measurements (Figure 5.8a), that could be imagined as an ideal case when each variable is measured by a specific device, the largest contribution comes from the measurements of the gas concentration. When assuming perfectly correlated measurements (Figure 5.8b), the largest contribution comes indeed from the measurement of the vertical wind speed. The largest values of the combined measurement uncertainty are associated to perfectly correlated measurements, while the lowest are associated to totally independent measurements of the six variables used in the EC model. In practice, it is reasonable to assume that actual values lie between these two extreme configurations.

Table 5.7 *NEE* and the corresponding uncertainty evaluated considering independence, perfect and minimum correlation for the flux measurements. The values are compared to the uncertainty evaluated with the method reported by Finkelstein and Sim [113].

Model	$NEE/(\mathrm{g\,m^{-2}})$	$u(NEE)/(\mathrm{g\,m^{-2}})$	$u_\mathrm{r}(NEE)/\%$
Independent	-415.6	5.6	1.3
Perfect Corr.	-447.1	39.6	8.9
Minimal Corr.	-430.0	14.3	3.3
F&S	-	5.8	-

The explanation for this significant and somehow surprising difference is the following: the concentration measurements are involved in the calculation of covariances only through relative differences which have, in the case of perfect instrumental correlation, approximately null uncertainty and thus the contribution to the combined uncertainty is very small. Another striking difference between these two cases, is the largest absolute values of the uncertainty associated with perfectly correlated measurements, especially for positive fluxes, that is upward fluxes. It is also noteworthy the presence of the term $c_{V_x V_y V_z}$, that is contributing only in case of perfectly and minimally correlated measurements and is basically connected with the use of a single device to measure more than one outputs, that is the triaxial anemometer. The different uncertainty composition between independent and correlated measurements could have profound consequences also on the practical side, e.g. by suggesting the evaluation of a "full independent measurement system" instead of those in which the three wind speed components are measured by a single device. In practical terms, this would mean to replace a sonic triaxial anemometer with three anemometric devices, eventually with much lower accuracy and costs or complexity, to produce an output with comparable or even lower measurement uncertainty than the traditional one based on dependent wind speed measurements.

When it comes to the application of the measurement uncertainty to data analysis, the results are tangible and the consequences relevant. Firstly, the estimates of the parameters used for the *NEE* model and their uncertainties change significantly between the NOLS and NWLS fits. When the measurement uncertainty is taken into account weighing each individual measurements by its own uncertainty, not only the numerical estimates of the parameters α, β and γ are different, but their relative uncertainties also do. Specifically, there is a substantial reduction in the relative uncertainty of all the three parameters. In other words, when using the

NOLS or NWLS, both the modelled NEE and its uncertainty change and this would have consequences on all the applications of the EC datasets: this means that all the functional relationships, that is light-curves or Penman-Monteith (just to name a few classical applications of EC) and, consequently, carbon, evapotranspiration or energy budgets would change in a not-negligible manner. There are few doubts about the fact that using a NWLS fit represents an improvement in the data analysis compared to NOLS and this is probably the first main practical consequence of adopting the proposed method to quantify the measurement uncertainty of individual EC measurements, because the NWLS was simply not applicable before. Secondly, the more subtle but not less important distinction between independent and correlated measurements leads also to differences in the final estimates and relative uncertainties that are of great relevance for both the experimental and the modelling community. In the study case, that is obviously site-specific but representative of a state-of-the art EC station, the difference in the EC annual carbon budget of a forest between the extreme cases of totally independent or totally correlated measurements is about 7.0 %, that is $-415.6\,\mathrm{g\,m^{-2}}$ and $-447.1\,\mathrm{g\,m^{-2}}$, respectively.

The differences between the uncertainties are even more important and peculiar than the differences of the annual budget estimates. The $u(NEE)$ shows dramatic variations for the different cases of independence, perfect or minimum correlation between the measurements. Since the independence assumption does not consider the covariances between different flux measurements (see Section 5.2, Equation 5.20), the cumulative uncertainty is monotonic and cannot decrease over time. Surprisingly, the cumulative uncertainty retrieved by the F&S method is almost coincident with the one for independent measurements. In principle, the F&S method only evaluates the uncertainty due to empirical sampling of the covariance between gas concentration and vertical wind speed, without making any considerations about measurement uncertainty. On the other hand, when instrumental correlation is considered, the $u(NEE)$ can decrease over time, as clearly visible in the right panel of Figure 5.10, which is particularly evident for perfectly correlated measurements. Moreover, the uncertainty can also increase at a rate which is much different with respect to the independent case, due to positive covariances between flux measurements. The curve for minimum correlation lies between the independent and the perfect correlation case, allowing the decrease of uncertainty due to negative covariances.

It is reasonable to assume that results of the same order of magnitude could be obtained for any other EC station and the methodological considerations can be

considered valid everywhere in the flux community. Until now, the literature has never used any SI traceable quantification method of the measurement uncertainty for the EC data interpretation. Together, these two final considerations highlight the importance of carefully quantifying the flux measurement uncertainty and the relevance of the proposed method, which implements, to the authors knowledge for the first time, the instrumental uncertainty in a way that is coherent with metrological recommendations from international organizations like BIPM and ISO.

5.4 Summary of main achievements

The Eddy Covariance (EC) technique has emerged in the last four decades as a crucial tool for measuring greenhouse gases and energy fluxes. However, the large complexity of the measurement model, constituted by many quantities measured by different instruments, may lead to numerous errors (briefly described in Section 5.1) that have been addressed by manufacturers and researchers. The proposed corrections are applied to each flux measurement to ensure that, at the time of recording, all the assumptions behind the model are respected. As stated in the GUM [5], each correction has its own uncertainty that shall be added to the budget of the flux measurement: this recommendation is ignored by the end-users of the EC system, which nowadays lacks of a complete evaluation of the measurement uncertainty. The same calibration uncertainty of the instruments is not a source of concern, since it was considered irrelevant compared to other involved uncertainties, such as that arising from the covariance sampling error [113].

This study proposes a quantification of the measurement uncertainty in the Eddy Covariance-based estimates of turbulent fluxes coherent with the GUM. One of the mentioned corrections is considered in the measurement model, which is the coordinate rotation of the anemometer's axes, needed to ensure its alignment with the wind flow. Each rotation coefficient depends on the time averages of wind speed components, adding further complexity to the calculation of the sensitivity coefficients.

Due to the lack of information about the calibration budget of the instruments, generally used trusting manufacturer's specifications, homoscedastic assumption is considered, which also allowed to retrieve effective sensitivity coefficients, one for each involved quantity (reported in Appendix A.2), instead of evaluating sensitivity

coefficients for each measurement of a certain quantity in the usual 30 minutes interval (i.e. 18000 values for each wind speed component). The cases of independent and correlated measurements are both considered, focusing on the instrumental correlation arising from the use of an instrument performing multiple measurements involved in the estimation of the same flux.

The developed framework for uncertainty propagation was tested over a state-of-the-art micrometeorological station for estimating carbon fluxes and annual budgets of a forest. The method and its basic principles are valid for any other EC system, except for minor modifications needed to account for substances different than CO_2. Instrumental input uncertainties and correlation coefficients are chosen by analyzing the manufacturers' specifications and investigating the typical uncertainties of the reference instruments used during calibration by accredited laboratories or other national metrology institutes. After applying the suggested method to the flux time series, attention is paid to the propagation of the evaluated uncertainties through the model for gap-filling (a simple light curve), a mandatory step for estimating reliable annual carbon budgets. For the correlated case, the instrumental correlation between different flux measurements is also estimated and taken into account for evaluating the uncertainty of the annual carbon budget.

In the end, the main achievements are as follows:

1. The Law of Propagation of Uncertainty (LPU) has been fruitfully applied to an EC dataset to quantify the combined uncertainty of each half-hourly flux and to break it down into contributions stemming from the measurements of individual variables. The LPU approach has been validated through Monte Carlo simulations (Figure 5.6), which confirmed its validity since the corresponding uncertainties are comparable within the first two significant digits (Table 5.4).

2. The distribution of uncertainties for the 2022 time series shows positive skewness (Figure 5.7), with a characteristic median below the threshold of $1\,\mu mol\,m^{-2}\,s^{-1}$ and 10 % for absolute and relative uncertainty, respectively. The largest relative uncertainty is obtained for the smaller flux magnitudes. The largest uncertainty term varies depending on the considered assumption, with the vertical speed components having a greater weight for the correlated case (Figure 5.8).

3. The quantification of the measurement uncertainty for each half-hourly record allows the implementation of a non-linear weighted fitting procedure that can be proposed as an advanced modelling strategy, compared to ordinary fitting procedures. The information coming from uncertainties changed the numerical estimates of all parameters in the light curve model (Figure 5.9), which showed lower uncertainties compared to the ordinary fitting procedure (Table 5.6). The more precise estimate of the model coefficient would benefit the whole community of flux data users.

4. The assumptions of independence, partial or perfect correlation between measurements play a major role and determine large differences both in the estimation of annual carbon budget and in the quantification of the corresponding uncertainty (Figure 5.10 and Table 5.7). The uncertainty for the independent case is almost the same as for the sampling error [113], whereas for the correlated case it is more than double, proving that the instrumental uncertainty component is not negligible in the overall budget.

Ideally, the suggested approach will increase the reliability of turbulent flux EC estimates, e.g. allowing inter-comparison of flux measurements obtained with different set-up or comparing different estimates over time performed with the same instruments. Additionally, an analytical expression of the measurement uncertainty would also be useful for determining technical specifications of instruments and associated calibration uncertainties required for a given application to meet the target uncertainty on the defined final products.

Chapter 6

Verification of thermometers in caves: the Bossea case study

The temperature within natural caves is a crucial parameter for understanding subterranean environments and their interaction with the external climate [149, 150]. Historically, cave temperatures were often considered to be remarkably constant, particularly in deep, isolated sections, typically aligning with the mean annual surface temperature of the overlying area [151–153]. This perceived constancy led to assumptions that the subterranean environment offered a stable habitat, with minimal daily or seasonal cues for cave-dwelling organisms [152]. However, advancements in measurement technologies and dedicated research have revealed that cave temperatures, while significantly damped compared to surface variations, exhibit a more dynamic and complex pattern influenced by a multitude of factors [150].

The primary mechanism governing heat transfer from the surface to caves is thermal conduction through the bedrock [151]. The surrounding rock acts as a low-pass filter, smoothing out diurnal and seasonal temperature fluctuations, causing cave temperatures to react to long-term temperature drifts with a notable delay [154, 150, 151]. This delay is influenced by the mountain's thermal inertia and the underground thermal properties [155]. While conduction is a dominant factor, convection, driven by air and water fluxes, also plays a significant role in temperature dynamics [155, 149, 156, 157]. Inflowing water, with its large thermal capacity, is often a primary determinant of cave temperature. Airflows, especially near cave entrances, can introduce external thermal signals and cause thermal stratification or

create cold/hot air traps [155, 150, 156]. Atmospheric pressure variations can further induce subtle but permanent temperature variations, known as Pressure-Induced Temperature (PIT) variations, even at great depths [158–160]. The interplay of these mechanisms creates varied thermal patterns across different cave sections and climates, ranging from highly correlated to the surface to showing extreme delays or no clear relationship [156].

Monitoring cave temperatures has become increasingly vital for understanding and tracking global climate change. Caves serve as valuable natural laboratories and archives for paleoclimate studies, as the stable isotopic composition of speleothems can act as a proxy for past climate conditions. More recently, direct, long-term temperature measurements within caves have provided compelling evidence of ongoing global warming [151]. For instance, studies in the European Alps have detected statistically significant warming trends in cave air temperatures, at about half the rate observed outside [3]. Similarly, historical temperature comparisons in Slovenian show caves indicate significant warming, often linked to rising outside temperatures [161, 162]. This thermal decoupling, where the cave interior warms with a delay of decades, offers a unique perspective on climate alteration.

The implications of climate warming for subterranean environments and their unique ecosystems are profound [152, 163]. Cave-dwelling species, having evolved in thermally stable conditions, are highly sensitive to even subtle temperature changes and may face challenges due to low adaptive potential and limited dispersal capabilities, potentially leading to local extinctions. While caves can act as thermal refugia for some surface species, they may simultaneously become “dead-end traps" for their obligate inhabitants [152]. Beyond direct biological impacts, warming can affect crucial cave processes such as ice formation and persistence, with some ice caves already experiencing enhanced melt and projected to lose perennial ice within decades [3]. Changes in temperature also influence cave ventilation patterns, humidity, and condensation processes, which are critical for speleothem growth and the overall cave microclimate.

Anthropogenic activities, particularly tourism, represent another significant factor influencing cave temperatures. Heat input from visitors, lighting, and electrical equipment can noticeably increase cave air temperatures, especially in frequently visited areas. This is evident in Postojnska Jama, where both the Pulpito and Sepolcro sites show a strong correlation between air temperature and visitor numbers [162].

Understanding these impacts is crucial for the effective management and conservation of geoheritage sites, including prehistoric caves with invaluable artwork. Examples include Chauvet, Esparros, and Pech Merle Caves in France, where temperature monitoring is critical for preserving prehistoric paintings [164, 159, 160]. Studies in these caves highlight the importance of understanding and mitigating the thermal impact of even limited human presence. Such impacts are not unique to caves, as underground urban infrastructure also experiences significant temperature changes due to human activities, leading to subsurface heat islands [165].

The development of automated digital dataloggers over the past two decades has revolutionized speleoclimatology, enabling long-term measurements with higher temporal resolution of cave air, water and rock temperatures. Designing effective cave monitoring system involves careful consideration of sensor placement, resolution, accuracy, sampling frequency and data accessibility, mainly depending on the measurand is intended to be measured [166]. Although modern dataloggers can offer high reliability in normal operating conditions, the extreme environment of the cave, with 100 %rh of relative humidity, can alter the performance of the device mostly because of condensing water. Low-cost dataloggers such as the HOBO MX2205 or the TinyTag Plus 2, often employed for cave monitoring, are housed in robust waterproof (IP68) cases designed for use in harsh and outdoor applications. Such instruments can measure just one or two probes, requiring the installation of many acquisition units for monitoring more sites of interest inside the cave. Another downside is that such devices are normally fixed on rocks very close to the monitoring point, often not easy to reach for maintenance, for instance to substitute batteries. A different approach is to use dataloggers of higher quality specifically design for environmental monitoring that are able to read more thermometers at the same time. These dataloggers can measure thermometers with a 4-wires technique, reducing the effect of parasitic resistances of the wires and thus the temperature error. The possibility of using long wires, even longer than 10 m, permits the monitoring of more points or quantities of interest, such as air/rock temperature gradients, located far from the acquisition unit. Compared to low-cost option, these dataloggers have better performance, especially in terms of resistance measurement uncertainty, a contribution in the overall uncertainty budget for temperature measurements. More details on how this type of datalogger can be tested are given in Section 3.1. Nevertheless, these devices are not designed to be rated IP68, hence some cautions should be taken when installing them in extreme environments such as in the caves.

Despite the adopted strategy, what is missing in literature is a rigorous application of metrology principles to the temperature measurements performed in caves. In many studies a comparison is involved, between the internal and the external environments or between different caves, requiring that SI traceability is established. However, the term "traceability" is never mentioned in literature and likely it is not guaranteed in the mentioned studies. Furthermore, long-term monitoring of temperature requires that the thermometers shall be periodically verified and/or calibrated, to remove potential instrumental drifts, something that is partially taken into account only in one recent study [3]. To show how a metrologically rigorous method can enhance the quality of temperature measurements in caves, a case study based on the monitoring of the Bossea cave is reported. In particular, the results of a verification campaign are presented to show how such methodology is crucial for correcting temperature errors, for instance due to power supply failures or drift of the measurement system. A description of the installed thermometers network and related issues is given in Section 6.1. The methodology considered for the verification campaign, along with the associated results, is reported in Sections 6.2 and 6.3. The campaign highlighted a systematic error affecting the datalogger, which was estimated and corrected as reported in Section 6.4.

6.1 Bossea's thermometers network

In 2021 the Politecnico of Torino, in collaboration with INRiM, installed a new network of thermometers inside the cave of Bossea, a show cave located in the province of Cuneo in Piedmont. The monitoring of physical (temperature, pressure, air and water flow) and chemical parameters (for instance CO_2 and Rn concentrations) inside the cave is active since 1969, making Bossea the most prominent and comprehensive underground karst laboratory in operation in Italy. Since the first installations, the equipment has undergone several renovations to keep up with technological advancements and improve the measurement quality [167]. The adopted strategy for the new network was to install a datalogger, in the five monitoring points shown in Figure 6.1, capable of acquiring more than one sensor having length of wires up to 60 m. The calibration of the Pt100 thermometers was carried out at INRiM using a thermostat bath and comparing them with a reference setup. Note that the ALMEMO dataloggers directly apply the standard Callendar-van Dusen equation, given in the

standard IEC 60751:2022 [168], to resistance measurements. A polynomial curve was used to correct the temperature readings having the form:

$$T_{\text{cal}} = a\,T_{\text{raw}}^2 + (1+b)\,T_{\text{raw}} + c, \tag{6.1}$$

where T_{cal} is the corrected temperature after calibration, T_{raw} is the raw temperature reading given by the datalogger, whereas a, b and c are the coefficients of the curve obtained through calibration. A key aspect is that the sensors under calibration were read using the same datalogger installed in the cave, the ALMEMO 5690, hence ensuring the traceability of the whole measurement chain: datalogger-wires-sensor. The expanded calibration uncertainty declared with a 95 % of confidence was 0.03 °C for all the sensors, suitable for long-term detection of temperature anomalies and comparison between different thermometers in this environment. Additional information about the calibration is available in [169].

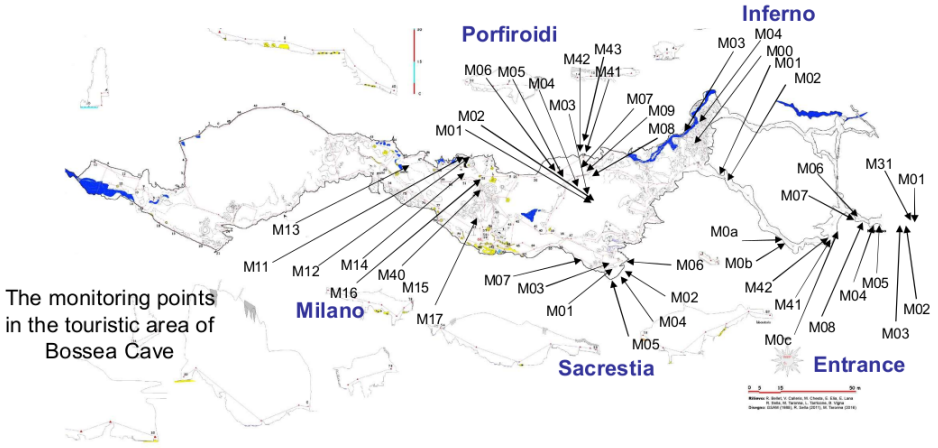


Fig. 6.1 Map of Bossea cave showing the monitoring points where thermometers are installed.

Although the chosen datalogger is designed specifically for environmental monitoring, the manufacturer recommends to use it in non-condensing conditions, a requirement that cannot be met in a cave. To avoid condensation of water on the case, an electrical heater was installed inside the plastic boxes housing the dataloggers. In this way, the temperature inside the box slightly increases, compared to the cave, which avoids condensation by reducing the relative humidity inside the box. An example of installation is shown in Figure 6.2 where the electrical heater is visible.

Much of the instrumentation in Bossea requires electricity for its operation. While it is easy to access, given that it is a show cave, there is a strong dependence

Fig. 6.2 Example of datalogger installation in a monitoring point of the Bossea cave. On the internal wall, on the right, it is visible the electrical heater used to reduce the relative humidity and avoid water condensation on the datalogger.

on it: power supply failures and voltage dips are in fact the main causes of data gaps. During 2023, a power failure lasted about a month and a half from April 25th to June 11th. The prolonged absence of power supply favored the condensation of water in the housings, which led to faults in the measurement system. Figure 6.3 shows the time series of three thermometers installed at *Porfiroidi*, one of the monitoring sites, which measure the temperature in meta-volcanite rocks at different depths. It is clear, especially for the rock temperature at 1 m depth, that an offset was introduced in the temperature signal because of an unknown measurement error.

In the period 05/03/2025-06/03/2025 an on-field campaign was conducted to verify the calibration of the PRTs and better investigate the faults of the measuring system. The verification is planned to be a routine activity to guarantee the required measurement accuracy and comparability among the several sensors. The thermometers under test were read using their same datalogger, guaranteeing the verification on the entire measurement chain: datalogger-wires-sensor.

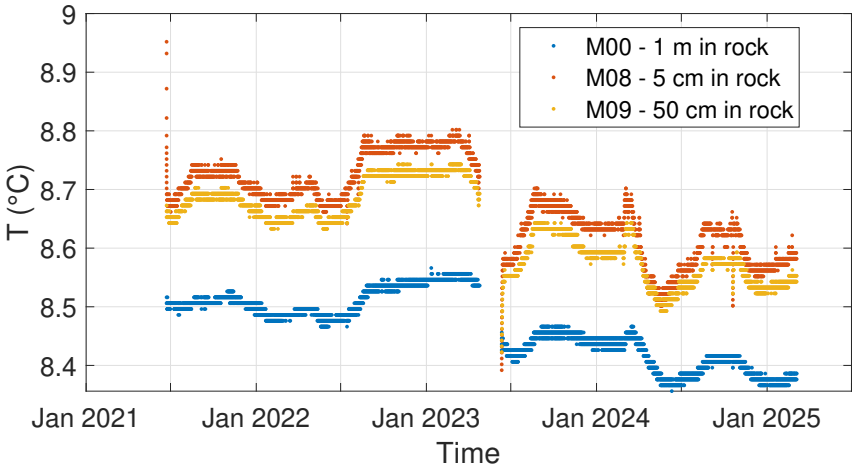


Fig. 6.3 Time series of three thermometers installed at Porfiroidi, one of the monitoring sites, which measure the temperature in metavolcanites at different depths.

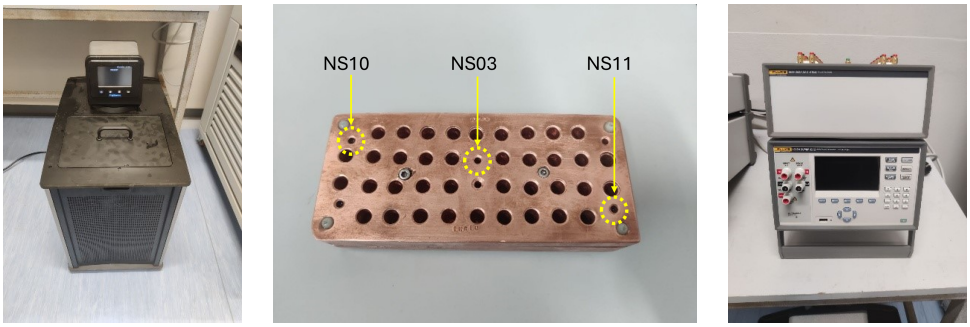
The cave monitoring sites involved in the campaign are those called *Sacrestia* and *Porfiroidi* (see Figure 6.1). Considering the limited time available, the harsh environmental conditions and the difficulties of removing the sensors from the installed positions, only a subset of sensors were verified in a liquid bath, transported by INRiM, for each site. Furthermore, sensors were tested only at one temperature of interest, which is fit-for-purpose considering the small variability of the temperature in the cave between 8 and 10 °C.

Another goal of the campaign was to test a second procedure based on the direct use of a comparator block in air and prove whether the natural thermal stability of the cave is enough for verification purposes. This method would provide the users with an easy to use system for self-validation in the future, based on the involvement only of a block, a high quality multimeter/resistance bridge and travelling reference thermometers provided by INRiM, thus excluding the use of the liquid-bath thermostat, which is heavy to transport and handle in the difficult route and challenging conditions inside the caves.

6.2 Methodology of the verification

The thermometers were verified by comparison with three reference Pt100 (internal s/n NS03, NS10 and NS11) traceable to the ITS-90. A PolyScience thermostatic bath (Figure 6.4a) filled with demineralized water, suitable for the measurement

range of interest, was used to generate a controlled temperature environment. A copper comparison block was crafted specifically for this purpose to increase the homogeneity and stability between the references and the thermometers under test, which was first used in liquid and then moved into air to test the second verification procedure. Figure 6.4b shows the comparator and the positions of the reference thermometers in the block, whereas Figure 6.4c shows the Fluke Super-DAQ 1586A temperature scanner used to read the reference Pt100s in 'scan' mode setting a 'medium' sampling rate (1 s for each channel), corresponding to one record every 3 s.



(a) (b) (c)

Fig. 6.4 Instruments used for the in situ verification: (a) The PolyScience thermostatic bath. (b) A copper comparator block where reference sensors are positioned as highlighted in the picture. (c) The Fluke Super-DAQ 1586A temperature scanner.

Because of the unique conditions inside the cave, a dedicated validation protocol was written, which includes analysis of measurement procedures and of the uncertainty components. For this reason, before transporting the instruments to the cave, characterization measurements were performed in the lab to check:

- Thermometers self-heating using the temperature scanner instead of a full resistance bridge (used for their calibration).
- Stability and homogeneity of the comparison block.
- Performance validity of the Super-DAQ for environmental conditions found in the cave ($T_{\text{env}} \approx 10\,^{\circ}\text{C}$, $RH \approx 100\,\%\text{rh}$).

Self-heating

The temperature scanner is a data acquisition system designed to read many thermometers at the same time. Given its scanning nature, the system supplies current

(1 mA, the same used for calibration) only during the sampling interval (1 s for each channel in our configuration). This type of acquisition differs from the employed calibration procedure for which a resistance bridge, the ASL/WIKA F18, was used. Indeed, the AC resistance bridge continuously supplies current to the measured resistances, resulting in different self-heating with respect to the on/off behavior of the scanner. Since the calibration curves refer to stabilized conditions using the AC bridge with 1 mA of current, we estimated the self-heating of the thermometers considering the operating conditions during the validation procedure, which will enter in the uncertainty budget of the validation accounting the different behaviors between our measuring system and the one used for calibration. The three Pt100s were inserted in the block, in the corresponding positions, inside the bath set at 8 °C and read using the Fluke 1595A resistance bridge, which has a function to estimate the self-heating by varying the input current. The self-heating errors for each sensor are those in Table 6.1. In principle, a similar procedure should be also performed in

Table 6.1 Self-heat estimated for the reference Pt100s.

Serial number	$\Delta T_{\text{self-heat}}/\text{mK}$
NS03	4
NS10	4
NS11	3

air to evaluate the self-heating for the second validation procedure that was tested. Nevertheless, best practices for self-heating evaluation in air are not yet defined since it strongly depends on air flow intensity and on the exposure of system to the flow, which were not possible to define for each cave site. For this reason, the total estimates of the self-heating were not applied as corrections, instead they have been used as components in the uncertainty budget for the procedure based on the comparison block in air.

Stability and homogeneity of the comparator

Before employing the comparison block for the validation, it was characterized in terms of homogeneity and stability inside the bath at 8 °C using the reference Pt100s. The stability is estimated by the difference between the maximum and minimum of the temperature recorded by one of the reference sensors in a 15-min

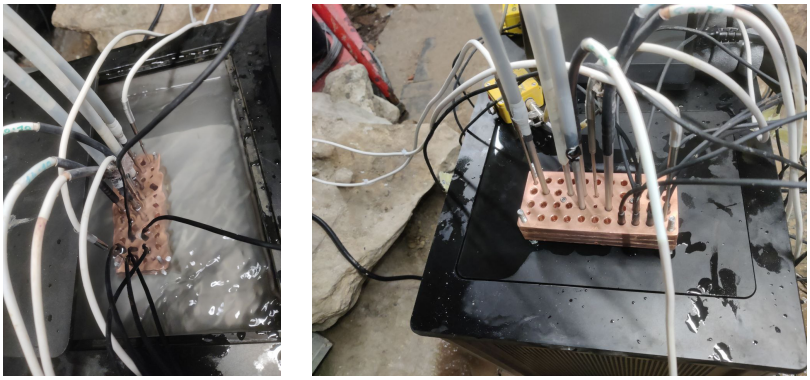
window once the oscillations are within ± 5 mK. Thus, the stability of the block was estimated as the maximum stability among reference sensors, which was equal to 1 mK. Then the homogeneity was estimated as the maximum difference between the thermometers positioned in the block during the same time frame, which resulted equal to 4 mK. The characterization of the block in air was not performed since the stability and air flow of the climatic chamber is not representative of the cave environment. However, stability and homogeneity were evaluated on site for each monitoring site and validation procedure.

Testing cave conditions

The performance of the Super-DAQ was then verified in harsh environmental conditions such as those in the cave. The measuring system, reading a stable reference in the block inside the bath set at 9 °C, was placed in a climatic chamber (Kambic KK-190 CHLT) first at $T_{\text{chamber}} = 20\,^{\circ}\text{C}$, $RH = 50\,\%\text{rh}$ and then at $T_{\text{chamber}} = 10\,^{\circ}\text{C}$, $RH = 100\,\%\text{rh}$ to verify if there was a significant change in the measurements. The difference between the two ambient conditions was estimated to be equal to 1.3 mK, with an associated standard uncertainty of 4.8 mK, proving that the obtained measurements are comparable and within the specification of the scanner, hence no further corrections were considered.

6.3 In situ verification

On 05/03/2025 the instruments were transferred to the cave for the measurement campaign. The verification process started at the Sacrestia site where six thermometers were validated by comparison (see Figure 6.5a) using the bath set at 9.5 °C, which was the closest to the measurement conditions of the sensors under test. The dataloggers (ALMEMO 5690) used to measure the thermometers under verification were configured to sample each channel every 5 s. After the comparison in liquid, the block has been exposed to air (see Figure 6.5b) and let it stabilize at the ambient conditions. A preliminary analysis made on site showed that the stabilization of the block in air was comparable to the bath and enough for the verification campaign (more details here below). The main difference is that the block and the thermometers stabilize at a non-controlled temperature, which is however representative of the



(a) (b)

Fig. 6.5 (a) Comparator block inside the bath. (b) Comparator block exposed to free air.

ambient temperature. Due to the measured performance, for the other monitoring site (Porfiroidi) only the block in air was used (Figure 6.6) to simplify the logistics and the instruments' transport.

Fig. 6.6 Comparator block exposed to free air during verification at Porfiroidi.

For each verification site a stabilization interval, defined as a 15-min time window in which the oscillations are smaller than ± 5 mK, is identified from the values of the reference Pt100s. The reference temperature $\overline{T}_{\mathrm{ref}}$, representative of the average block temperature, is estimated by taking the average of all the reference Pt100s during the stabilization interval. Using this estimate, for each sensor under verification is evaluated the temperature difference $\Delta T_{\mathrm{V},i} = \overline{T}_{\mathrm{ref}} - \overline{T}_{\mathrm{cal},,i}$, where $\overline{T}_{\mathrm{cal},i}$ is the mean temperature value obtained from the measurements of the i-th sensor under test applying the corresponding calibration curve. The uncertainty $u(\Delta T_{\mathrm{V},i})$ is calculated using the following components which were evaluated from the measurements for each sensor:

- $u_{\text{ref-cal}}$: calibration uncertainty of the reference.

- $u_{\text{ref-sd}}$: uncertainty associated with the repeated measurements of the reference during stability and evaluated as $\sigma(T_{\text{ref}})/\sqrt{N_{\text{ref}}}$, where N_{ref} is the number of measurements.

- $u_{\text{ref-res}}$: uncertainty associated with the resolution of the reference data acquisition system.

- u_{stab}: uncertainty related to the time stability of the reference temperature, evaluated as the max-min of the reference time series weighted by a uniform distribution.

- u_{spatial}: uncertainty related to the spatial homogeneity of the temperature across the comparator block, evaluated as the maximum temperature difference between the reference Pt100s weighted by a uniform distribution.

- $u_{\text{self-heat}}$: uncertainty associated with the self-heating of the reference Pt100s and evaluated by weighted with a uniform distribution the maximum $\Delta T_{\text{self-heat}}$.

- $u_{\text{test-cal},i}$: uncertainty associated with the calibration of the sensor under test.

- $u_{\text{test-sd},i}$: uncertainty associated with the repeated measurements of the sensor under verification and evaluated as $\sigma(T_{\text{cal},i})/\sqrt{N_i}$, where N_i is the number of measurements.

Eventually the total standard uncertainty is obtained by summing in quadrature all the above components:

$$u(\Delta T_{\text{V},i}) = \sqrt{u^2_{\text{ref-cal}} + u^2_{\text{ref-sd}} + u^2_{\text{ref-res}} + u^2_{\text{stab}} + u^2_{\text{spatial}} + u^2_{\text{self-heat}} + u^2_{\text{test-cal},i} + u^2_{\text{test-sd},i}}. \tag{6.2}$$

After having evaluated the uncertainty associated with ΔT_i, the comparison index E_i for the verification is calculated as:

$$E_i = \frac{|\Delta T_{\text{V},i}|}{2u(\Delta T_{\text{V},i})}. \tag{6.3}$$

The criterion for determining validated sensors is the following: thermometers with $|E_i| < 1$ are considered to have a valid calibration, otherwise they should be recalibrated. However, considering the small variability of the temperature measured

by the thermometers and assuming a constant drift across the calibration interval, the sensors readings can be corrected provided that a further uncertainty term is added to the calibration uncertainty as suggested by the GUM [5]:

$$T_{\mathrm{cor},i} = T_{\mathrm{cal},i} + \Delta T_{\mathrm{V},i}, \tag{6.4}$$

$$u(T_{\mathrm{cor},i}) = \sqrt{u^2_{\mathrm{cal},i} + u^2(\Delta T_{\mathrm{V},i})}, \tag{6.5}$$

where: $T_{\mathrm{cor},i}$ is the corrected temperature corresponding to the reading $T_{\mathrm{cal},i}$, $u(T_{\mathrm{cor},i})$ is the uncertainty of the corrected temperature and $u_{\mathrm{cal},i}$ is the calibration uncertainty of the thermometer.

The temperature differences with respect to the reference temperature, the corresponding uncertainty and the comparison index are reported for each sensor in the following tables related to a specific monitoring site. For the Sacrestia verification, the results for the two different verification procedures are reported.

Sacrestia

The results of the comparison in liquid for each sensor are shown in Table 6.2, whereas Table 6.3 reports the results for the validation procedure using the block into air. For this site, the estimated reference temperature using the bath is $\overline{T}_{\mathrm{ref}} = 9.419\,^{\circ}\mathrm{C}$, whereas for the comparison in air is $\overline{T}_{\mathrm{ref}} = 9.790\,^{\circ}\mathrm{C}$.

Table 6.2 PRT verification in liquid at Sacrestia site.

S/n	$\overline{T}_{\mathrm{cal}}/^{\circ}\mathrm{C}$	$\Delta T/^{\circ}\mathrm{C}$	$u(\Delta T)/^{\circ}\mathrm{C}$	$U(\Delta T)/^{\circ}\mathrm{C}\,(k=2)$	E
M01	9.33	0.092	0.016	0.033	2.83
M02	9.32	0.102	0.016	0.033	3.12
M03	9.33	0.088	0.012	0.024	3.70
M04	9.34	0.078	0.016	0.033	2.41
M05	9.32	0.100	0.016	0.033	3.07
M06	9.34	0.083	0.016	0.033	2.55

As is it clear from the comparison index, none of the sensors passed the verification, thus they should be corrected as described above. The uncertainty $u(\Delta T_{\mathrm{V}})$ is identical for the two validation procedures, proving that using the comparison block in air ensures a sufficiently stable and uniform reference system to be employed for on-site testing. Indeed, the estimated stability and homogeneity in air were 4 mK

Table 6.3 PRT verification in air at Sacrestia site.

S/n	$\overline{T}_{\mathrm{cal}}/°\mathrm{C}$	$\Delta T_{\mathrm{V}}/°\mathrm{C}$	$u(\Delta T_{\mathrm{V}})/°\mathrm{C}$	$U(\Delta T_{\mathrm{V}})/°\mathrm{C}\,(k=2)$	E
M01	9.70	0.090	0.016	0.033	2.76
M02	9.68	0.108	0.016	0.033	3.32
M03	9.71	0.083	0.012	0.024	3.52
M04	9.75	0.043	0.016	0.033	1.30
M05	9.68	0.110	0.016	0.033	3.36
M06	9.72	0.073	0.016	0.033	2.25

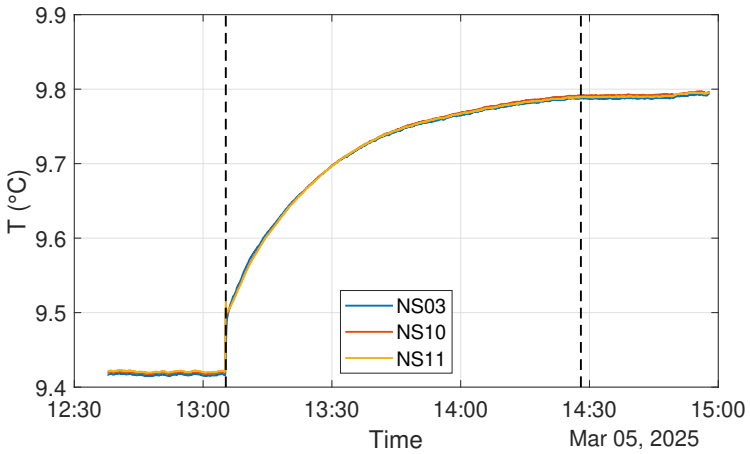


Fig. 6.7 Measurements of the reference thermometers showing the time required to stabilize the comparator block exposed to free air. The time instants highlighted by the black dashed lines correspond to the removal of the block from the bath and to the reached stabilization of the block.

and 6 mK, which are small compared to other uncertainty terms and about the same estimated on site for the bath. An estimate of the time required for stabilization can be made by looking at the temperature time series of the references in Figure 6.7. The time required for stabilization is about an hour and a half, with a thermalization rate of approximately 0.2 °C/h. This value strongly depends on the air flow around the block and can vary in different locations of the cave. Nevertheless, with the purpose of future periodic verifications, one or more comparators could be left in the cave close to the monitoring sites, thus ensuring that they are already thermalized and ready for taking measurements.

Porfiroidi

The estimated reference temperature using the block in air is $\overline{T}_{\text{ref}} = 8.990\,^{\circ}\text{C}$. Also for this site the thermometers do not conform with their calibration, with an absolute error that is much more pronounced, more than $0.5\,^{\circ}\text{C}$, compared to the sensors in Sacrestia. Assuming a constant error after the power failure during spring 2023, the

Table 6.4 PRT verification in liquid at Porfiroidi site.

S/n	$\overline{T}_{\text{cal}}/^{\circ}\text{C}$	$\Delta T_{\text{V}}/^{\circ}\text{C}$	$u(\Delta T_{\text{V}})/^{\circ}\text{C}$	$U(\Delta T_{\text{V}})/^{\circ}\text{C}\,(k=2)$	E
M00	8.42	0.574	0.021	0.042	18.7
M03	8.39	0.602	0.016	0.033	18.5
M04	8.40	0.591	0.016	0.033	18.2
M08	8.39	0.598	0.016	0.033	18.4
M09	8.41	0.577	0.012	0.024	24.5
M41	8.41	0.579	0.016	0.033	17.8
M42	8.39	0.598	0.016	0.033	18.4

measurements are corrected and plotted in Figure 6.8. It is evident that applying the estimated correction enhances the temperature offset that was shown in Figure 6.3, even changing its sign. The corrected measurements, at least those close to the validation event, are considered trustworthy since relying on estimates done by the reference measurement system, which was tested in laboratory for the cave environment. On the other hand, after the first calibration of the monitoring setup, an on-site verification was not performed, trusting the corrected records of the dataloggers. The shown discrepancy visible in the time series suggests that also the measurements before the power failure are affected by some sort of error due to the datalogger, which was not detected before. A further investigation of this problem and a possible solution are presented in Section 6.4.

6.4 Datalogger faults and error correction

An additional campaign of measurements was conducted on 06/06/2025 to check the correct operation of the dataloggers. An ALMEMO 710, which operates as the 5960 model but with the advantage of being portable, was recently purchased by INRiM and was used for the purpose. At each monitoring site, a comparison between the

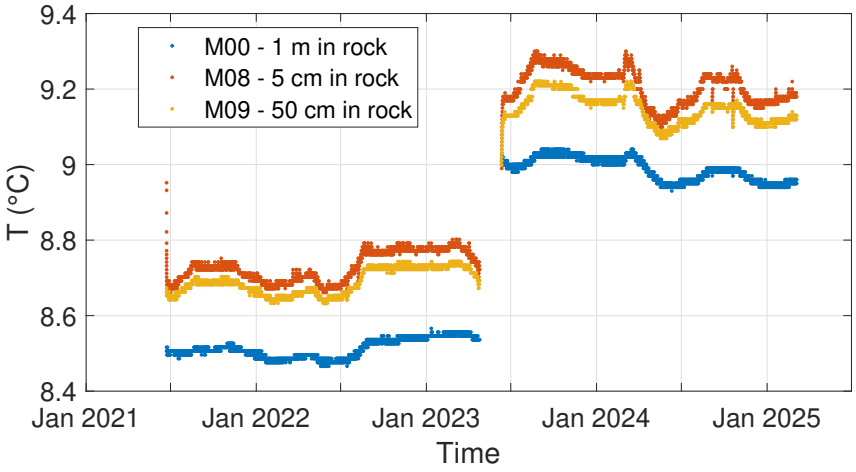


Fig. 6.8 Time series of three thermometers installed at Porfiroidi. The measurements after the power failure are corrected using the verification results. The correction enhances the temperature offset between the two portions of the time series, proving that an additional error affects the measurements before the failure event.

dataloggers was done by reading each sensor first with the 5960 and then with the 710. The goal was to test the raw acquisitions, without applying any calibration curve. This method is consistent because the dataloggers were used in sequence to read thermometers that are subjected to a fairly stable environment. The difference $\Delta T_{\mathrm{D},i} = T_{\text{raw-ref},i} - T_{\text{raw-DAQ},i}$ is evaluated for each sensor to verify channel-depending effects: $T_{\text{raw-ref},i}$ is the raw reading of the i-th sensor using the reference datalogger, whereas $T_{\text{raw-DAQ},i}$ is the raw reading performed with the installed datalogger. Since the acquisition circuit is shared among the sensors, the average of the observed differences, ΔT_{D}, is taken as the best estimate. A standard type-A measurement uncertainty is evaluated by considering independent measurements between the different channels, which is reasonable since the sensors are installed in different points of the cave.

Table 6.5 Temperature differences evaluated on 06/06/2025 by comparing the readings of the installed sensors obtained with the installed datalogger, the ALMEMO 5960, and with a reference portable acquisition device, the ALMEMO 710.

Site	$\Delta T_{\mathrm{D}}/^{\circ}\mathrm{C}$	$u(\Delta T_{\mathrm{D}})/^{\circ}\mathrm{C}$
Sacrestia	0.231	0.003
Porfiroidi	0.725	0.001

The results of these evaluations, reported in Table 6.5, prove that a non-negligible difference is observed, which leads to a temperature error that can be estimated by evaluating the following equations:

$$T_{\text{cal'}} = a\,(T_{\text{raw}} + \Delta T_{\text{D}})^2 + (1+b)\,(T_{\text{raw}} + \Delta T_{\text{D}}) + c, \tag{6.6}$$

$$\varepsilon_{\text{D}} = T_{\text{cal}} - T_{\text{cal'}} = -a\Delta\overline{T}_{\text{D}}^2 - 2aT_{\text{raw}}\Delta T_{\text{D}} - (1+b)\Delta T_{\text{D}} \approx -\Delta T_{\text{D}}. \tag{6.7}$$

The term T_{cal} represents in this case the measurements obtained by applying the calibration curve to the raw readings which are affected by a systematic error. On the other hand, the term $T_{\text{cal'}}$ corresponds to the measurement obtained by correcting the raw readings adding the temperature difference ΔT_{D} before applying the calibration function. The error ε_D is estimated by the difference between the two terms, which is approximately equal to $-\Delta T_{\text{D}}$ because the predominant term c is sorted out, leaving the a and b coefficients which are one order of magnitude smaller than ΔT_{D}.

The temperature correction ΔT_{D} is coherent in sign with ΔT_{V}, but can only partially explain it. To align the quantity $T_{\text{cal'}}$ to the corrected temperature T_{cor} seen in the previous section, it is necessary to evaluate another correction applied to the calibration-corrected values:

$$\Delta T_{\text{S}} = \Delta T_{\text{V}} + \varepsilon_{\text{D}} \approx \Delta T_{\text{V}} - \Delta T_{\text{D}}, \tag{6.8}$$

$$T_{\text{cor}} = a\,(T_{\text{raw}} + \Delta T_{\text{D}})^2 + (1+b)\,(T_{\text{raw}} + \Delta T_{\text{D}}) + c + \Delta T_{\text{S}} \tag{6.9}$$

Since $\Delta T_{\text{D}} > \Delta T_{\text{V}}$, the additional correction to be applied is negative, meaning that the resistance read during validation was larger with respect to the reference, and not the opposite. This behavior is compatible with what known for PRTs, since possible oxidation or induced mechanical strain on the sensing element generates a greater resistance. For this reason, the subscript "S", which stands for "Sensor", is used to distinguish such a correction. Taking as an example the thermometers M00, M08 and M09 at Porfiroidi, the calculated ΔT_{S} is $-0.16\,^{\circ}\text{C}$,$-0.12\,^{\circ}\text{C}$ and $-0.14\,^{\circ}\text{C}$, which converted in linear instrumental drift correspond to about $0.03\,^{\circ}\text{C/y}$. Platinum thermometers usually have a time drift per year in the mK range, but works studying the drift in constant high-saturation conditions at a reasonably stable temperature are missing. Probably, the long power interruption also affected the sensors in a not determinable, introducing a steep increase of their resistance, which is difficult to isolate from the standard time drift of a Pt100.

The uncertainty of the temperature estimated in this way is evaluated by:

$$u(T_{\text{cor}}) = \sqrt{u^2_{\text{cal}} + u^2(\Delta T_{\text{D}}) + u^2(\Delta T_{\text{S}}) - 2\text{cov}(\Delta T_{\text{D}}, \Delta T_{\text{S}})} = \sqrt{u^2_{\text{cal}} + u^2(\Delta T_{\text{V}})}. \tag{6.10}$$

In the end, considering the covariance between ΔT_{D} and ΔT_{S}, the uncertainty is identical to that of Equation 6.5, thus it is the same evaluated with the verification procedure. Although both the estimates and the uncertainties do not changes, the main advantage of reformulating the temperature correction in the form expressed by Equations 6.8-6.9 is that the datalogger error is decoupled from the error of the sensor. The comparison using the reference datalogger was much easier to perform with respect to the verification process because it was necessary to transport only a small datalogger in its own transportable case. This procedure should be repeated every 3/4 months to check the stability of the datalogger error. In case of changes in time of ΔT_{D}, the above equations can be evaluated using the updated value to align the measurements to the (last) verification event. Instead, a complete verification, as described in Section 6.3, should be performed every two years to reduce the exposure of laboratory-grade instruments to the cave environment, but also because of the logistical issues one has to deal with for transporting such equipment.

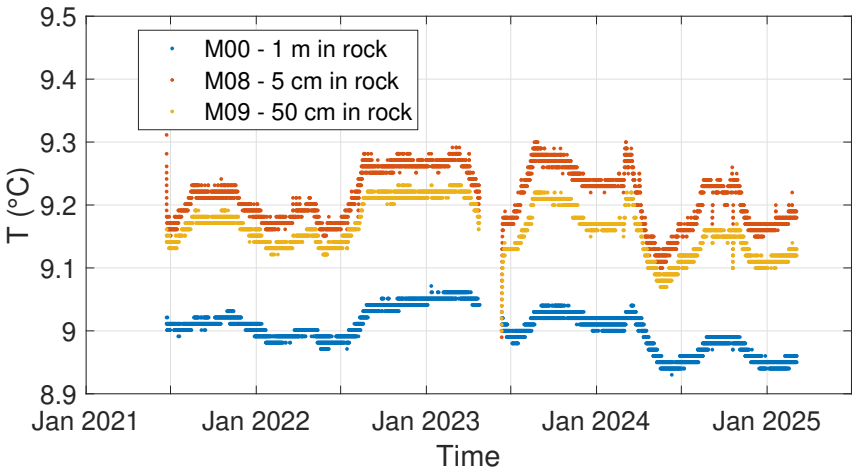


Fig. 6.9 Harmonized time series of three thermometers installed at Porfiroidi. The measurements before and after the power failure are corrected using the error model described in Section 6.4.

The developed error model allows to tackle the problem of time series harmonization at Porfiroidi site, reducing the temperature offset introduced by the data gap, as previously shown in Figure 6.3. As already proved, correcting only the measure-

ments after the power failure enhances the discontinuity between the two parts of the time series (see Figure 6.8). Assuming that the corrected measurements after the failure event are reliable, it is also necessary to correct the measurements before the data gap to harmonize the series. Indeed, the error affecting the datalogger readings could have been present also before, but with a different intensity. A correction of the readings ΔT_{D_0} is then visually estimated by trial and error and applied to the measurements. Feasible identified values are in the interval $[0.49, 0.52]\,^{\circ}\mathrm{C}$: the average is taken as best estimate, whereas the associated uncertainty is evaluated by considering a uniform distribution over the interval. Therefore, the applied correction is $\Delta T_{\mathrm{D}_0} = 0.505\,^{\circ}\mathrm{C}$ with a standard uncertainty $u(\Delta T_{\mathrm{D}_0}) = 0.01\,^{\circ}\mathrm{C}$, which should be added in quadrature to the calibration uncertainty for evaluating the overall uncertainty of the corrected measurements. Figure 6.9 shows the result of the harmonization, with the corrected measurements having an expanded uncertainty, with a 95 % of confidence, of $0.04\,^{\circ}\mathrm{C}$ for both portions of the time series.

6.5 Summary of main achievements

This chapter focuses on the results of an on-field verification campaign, conducted to verify the calibration of the Pt100s network, installed in the Bossea cave, after a power line failure event that lasted more than a month. The main goal was to ensure traceability to the national standards and comparability between different thermometers. A dedicated verification procedure, based on the use of a thermostatic bath on site, was conceived and performed by the INRiM staff, as also the analysis of all the uncertainty components based on laboratory measurements and on-site evaluations. A second product of the campaign was the performance assessment of the verification based on a copper comparator block (Figures 6.5 and 6.6) exposed to free air, which resulted in equivalent results compared to the use of a thermostatic bath. In particular, the temperature stability and uniformity were estimated to be 4 and 6 mK, which are suitable for the intended use. The main drawback of using the comparator block in air is the longer stabilization interval, which is in the order of a couple of hours but can vary depending on the ventilation conditions in the cave. Independently from the employed method, the verification proved that none of the sensors conformed with their calibration, with a deviation of about $0.6\,^{\circ}\mathrm{C}$ for some thermometers. The differences against the reference system can be used to correct

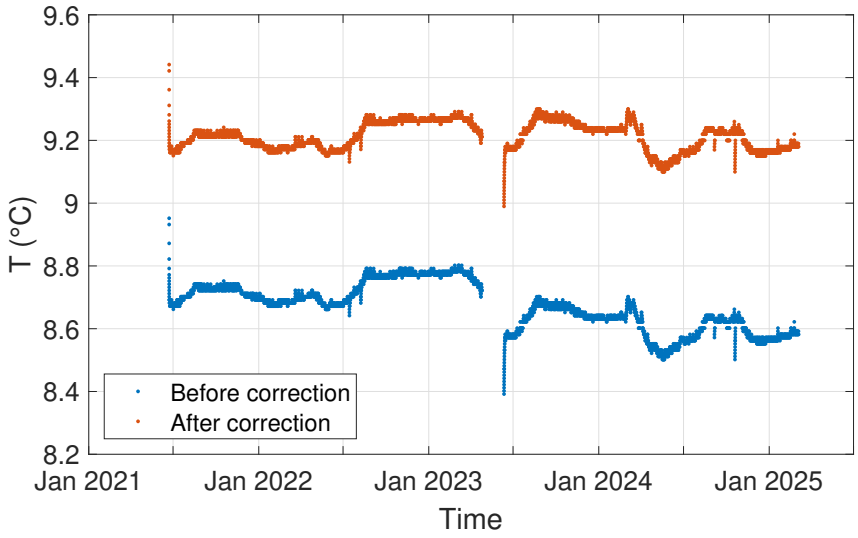


Fig. 6.10 Comparison between the time series of sensor M08, installed at Porfiroidi site in rock at 5 cm depth, before and after applying the correction of measurements.

the measurements thanks to the small variability of the temperature signal, which is in the order of 0.1 °C. This approach avoids to carry the sensors to the laboratory for a complete calibration, which is more expensive and requires more time. An additional uncertainty associated with the correction has been also evaluated from the verification process which is about 0.03 °C with $k = 2$ corresponding to a coverage probability of 95 %.

The campaign also revealed an error associated to the datalogger readings, which was further investigated by a comparison with a reference datalogger. An error model (Equations 6.6-6.10) was developed to compensate such behavior and align the temperature correction with the one obtained by the verification procedure. The use of this error model also allowed to harmonize the measurement time series, removing the temperature offset that was introduced by the power supply failure as shown in Figure 6.10. After correcting the values, the final type-B measurement uncertainty associated with the measurement is about 0.04 °C with $k = 2$ and a coverage probability of 95 %.

The reported results prove that a rigorous metrological approach is mandatory when dealing with measurements, especially to make sure that the measurement data products are traceable to the national standards. Without a guaranteed traceability of the instruments, the comparison between different temperature sensors is not reliable, even meaningless. Furthermore, a periodic verification allows to detect an

eventual instrumental drift, that could be confused for a climatological temperature anomaly. It is undeniable that the application of metrology principles requires more effort, in terms of competences and available instrumentation, particularly in extreme conditions such as those in a cave. This is the reason why the proved performance of the validation procedure in air will drive in the future the realization of a self-validation system, which will facilitate periodic test of the sensors without the need of an expensive instrument as the liquid-bath thermostat, which is also heavy to transport and handle in the challenging conditions inside the caves. The self-validation system will be used not only for the Bossea cave, but also in the context of a citizen science project called "Underground Climate Change" (UCC) that has the goal of establish a climate monitoring network of italian caves, at the moment composed by 18 caves. Even if standardized instruments were chosen and an installation protocol was defined, only low-cost dataloggers are installed, such as the HOBO MX2205, due to the limited budget of the project. To ensure comparability between different caves are therefore essential periodic verifications of the probes against a reference system, demonstrating the crucial role of the methodology reported in this chapter.

Chapter 7

Conclusions

This dissertation has examined how the rigorous principles of metrology can be systematically applied to climate monitoring. Although surface and flux measurements are routinely collected worldwide, the quality of these observations is often compromised by subtle but systematic sources of error. These range from instrumental limitations of acquisition systems to siting practices that bias readings, from environmental interactions with sensors to the absence of routine verification. If left unaddressed, such issues propagate into climatological datasets, undermining their reliability for detecting variability, attributing change, and supporting international assessments.

The central ambition of this work has been to demonstrate that climate monitoring must move beyond routine data collection and adopt a reference paradigm, where observations are traceable to the SI, redundantly measured, and accompanied by uncertainty budgets. Across the case studies developed in this thesis, this ambition has been addressed in practical terms: from the design of a Climatological Reference Station (Chapter 2), to detailed analyses of air temperature uncertainties (Chapters 3–4), to the propagation of uncertainties in ecosystem fluxes (Chapter 5), and finally to the verification of unconventional measurement sites (Chapter 6).

7.1 Main Scientific Contributions

The first major contribution is the design and deployment of a Climatological Reference Station (CRS). This prototype, realized at INRiM, integrates redundancy,

calibration traceable to the SI, and comprehensive uncertainty analysis. By embedding metrological principles into the architecture of the station, the CRS addresses the shortcomings of conventional automatic weather stations, which are prone to inhomogeneities arising from untracked changes in sensors or siting. The CRS represents a practical step toward the vision of a GCOS Surface Reference Network (GSRN). Its configuration showed how redundancy between instruments can provide internal consistency checks, while detailed calibration ensures traceability.

A second contribution is the quantification of measurement system uncertainties in air temperature observations. Two critical sources of uncertainty were investigated: (i) the sensitivity of datalogger electronics to case temperature, and (ii) the effect of rainfall on shielded thermometers. Laboratory characterization of the SIAP DA18K datalogger revealed systematic biases linked to thermal sensitivity. These deviations are larger for decreasing ambient temperature, being about 0.07 °C at −10 °C, which far exceeds the manufacturer's declared specifications of 0.02 °C. Additionally, the rainfall experiments demonstrated that wetting of naturally ventilated shields can cause transient cooling of up to 4 °C (Figure 3.16), with recovery times that are shorter for forced-ventilated screens. These findings highlight that acquisition electronics and environmental interactions are integral parts of the measurement process and must be explicitly included in uncertainty budgets. Collectively, this chapter contributes to the Measurement Quality Classification Scheme promoted by the WMO, providing empirical evidence for categories of uncertainty that were previously under-characterized.

A third contribution lies in the quantification of the influence of installation height on air temperature records, a long-standing but insufficiently quantified source of variability. By simultaneously operating eight calibrated Pt100 sensors at 1.25, 1.50, 1.75, and 2.00 m, the study demonstrated systematic biases between heights, with maxima up to 0.35 °C and minima down to −0.65 °C at 1.25 m relative to the 2 m reference. These biases were especially pronounced under stable nighttime conditions and during summer, when vegetation growth enhanced near-surface gradients. A correction model based on solar radiation successfully reduced daytime biases, decreasing the maximum offset at 1.25 m from 0.35 °C to 0.14 °C. These results show that height-dependent effects are climatologically significant, particularly for extremes, and call for either a stricter standardization of installation height or systematic correction procedures in climate data homogenization.

The fourth contribution is the extension of metrological approaches to eddy covariance CO_2 flux measurements, producing one of the first GUM-compliant uncertainty budgets for turbulent fluxes. By constructing a full uncertainty budget and deriving effective sensitivity coefficients, the work quantified how uncertainty components propagate into estimates of net ecosystem exchange and annual carbon budgets. The application of the methodology to the measurements taken at San Rossore's site revealed the decisive role of correlations between flux measurements in shaping annual carbon budgets. Moreover, the use of a weighted regression approach for gap-filling incomplete flux time series showed how uncertainty information can influence parameter estimation of flux–meteorology relationships and the corresponding uncertainty. These methodological advances have direct relevance for networks such as ICOS and AmeriFlux, where harmonized uncertainty reporting is still developing.

Finally, the dissertation contributed by adapting verification protocols to an unconventional environment, such as that of a cave. The dense network of Pt100 thermometers in the Bossea cave was subject to an in situ verification campaign following a power failure. Using thermostatic baths and comparator blocks, deviations of up to 0.6 °C were identified between calibration and operational values. Correction models were then applied to harmonize the time series. This study demonstrated that verification is essential even in seemingly stable subterranean environments and that metrological practices can be successfully transferred beyond conventional meteorological stations.

Together, these contributions illustrate that the metrological approach is not limited to laboratories but can be successfully applied across the full spectrum of climate monitoring contexts.

7.2 Broader implications and limitations

The dissertation demonstrates that uncertainties in climate monitoring are multidimensional, encompassing instrumentation, acquisition systems, siting, and environmental influences. By systematically quantifying these factors, the thesis contributes to ongoing international initiatives such as the GCOS Surface Reference Network (GSRN) and the WMO-led Global Basic Observing Network (GBON).

The developed CRS prototype can serve as a testbed for future standardization of reference-grade sites, whereas the datalogger characterization is crucial to have a complete knowledge of the data acquisition chain for reference measurements. Furthermore, the developed protocol will be the foundation for laboratory testing of dataloggers which will be used in the intercomparison of climate monitoring stations joining the pilot phase of the GSRN. The tests and intercomparison phase will be conducted by INRiM in the following two years exploiting the same facilities and experimental area reported in Chapters 3-4.

The corrections for rainfall and installation height provide actionable improvements to WMO No. 8, especially for siting and exposure classification of thermometers. The propagation of uncertainty in EC fluxes directly supports the methodological development of ICOS and similar flux networks, becoming the first step to fill the identified metrological gap. Finally, the verification protocols for cave thermometers highlight how reference principles can be extended to unconventional monitoring contexts, which starts to draw the attention of climatologists.

Conceptually, the work reinforces the principle that prevention is preferable to correction. While statistical homogenization will remain necessary for historical records, the future of climate monitoring should be built around systems that minimize inhomogeneities at their source. This shift in philosophy has the potential to improve the reliability of climate records while reducing the effort required for post-processing.

At the same time, several limitations must be acknowledged. The CRS prototype represents a single-site implementation, and its scalability to diverse climates and operational conditions remains to be demonstrated. The rainfall and installation height experiments were conducted under specific conditions, and replication across different climates, vegetation covers, and wind regimes would strengthen the conclusions. The eddy covariance uncertainty budget relied on assumptions about correlation structures that need validation using detailed calibration budgets at multiple sites. Finally, the thermometer verification in the cave using the copper comparator in air needs to be refined and further tested to ensure its efficacy in different caves, with stabilization times compatible with on-field activities. These limitations point not to shortcomings but to natural boundaries of a project conducted with limited resources in a defined time window.

7.3 Future Perspectives

Looking forward, several avenues of research and development emerge from this work. A clear priority is the expansion of instruments installed on the CRS for measuring other ECVs, such as soil moisture. The recently concluded SoMMeT project investigated the metrological traceability of Cosmic Rays Neutron Sensing (CRNS) probes, which are capable of estimating soil moisture based on the counting of secondary epithermal neutrons interacting with water molecules in the soil. Inspired by the issues encountered by the members of the project consortium, INRiM started an activity for the characterization of this kind of probe using the research nuclear reactor at the LENA facility in Pavia. The aim is to generate neutron fluxes comparable to those in the outdoor environment, which would allow to test the detector electronics without reaching saturation. The research is underway, following the metrological principles that characterize this dissertation.

Further controlled experiments are needed to refine the understanding of environmental influences on air temperature measurements. The effects of precipitation, solar radiation, and wind should be further systematically studied across a variety of shield designs and site conditions. A solar simulator has been recently installed at INRiM facilities, specifically designed for use in combination with the temperature-controlled wind tunnel developed within the MeteoMet projects. After the necessary characterization of the combined systems, the tunnel will be used to produce environmental conditions controlled in temperature, wind speed, pressure and solar radiation, enabling the implementation of new tests on solar screens. At the same time, activities should be conducted to reduce the model uncertainty affecting acoustic thermometry in air. The reason is that a practical metrological principle states that if you do not manage to reduce an uncertainty source, it is best to remove it from the beginning. Non-contact methods are ideally immune to the radiative error, hence they will likely complement the contact air temperature measurements in the future. Nevertheless, attention should be paid on the sensitivity of instruments used for non-contact techniques, such as sonic anemometers, which may show thermal sensitivity as shown for the datalogger in Chapter 3.

In the field of flux measurements, the developed framework for uncertainty propagation will be adapted to fluxes of latent and sensible heat or other scalar quantities such as methane. Additional corrections will be considered in the measurement model, especially those accounting for non-stationarity and frequency attenuation, to

evaluate an uncertainty budget as complete as possible. Once a common methodology for uncertainty propagation is defined, coordinated efforts are required to characterize the instrumental correlation of eddy covariance measurements, which is specific to the adopted instrumentation. Eventually, particular attention shall be paid to the gap-filling procedure, nowadays performed using Machine Learning or Neural Networks models, whose reliability is yet to be studied, especially concerning the propagation of input uncertainties and the uncertainty evaluation of the predictions. In the end, establishing standardized GUM-compliant uncertainty reporting within ICOS and similar networks would enhance comparability and improve confidence in carbon budget assessments.

Calibration and verification protocols should also be standardized for the monitoring of temperature in caves. This unconventional environment is usually studied by researchers or amateurs who lack metrological knowledge. In addition, low-cost instruments are often used without knowing or having tested their possible time drift, the main source of error and uncertainty when detecting temperature anomalies. Proactive and scheduled campaigns would ensure continuity and traceability of long-term records, reducing the risk of undetected biases and increasing the comparability between different caves.

In summary, this dissertation has demonstrated that the systematic integration of metrology into climate monitoring is both feasible and necessary. Through experimental campaigns, methodological developments, and case studies, it has provided new insights, tools, and prototypes that strengthen the foundations of climate science. The work contributes to the shift toward reference-quality observations, reinforcing the idea that climate data must be not only long and continuous but also traceable, comparable, and uncertainty-quantified. At a time when global decisions depend on robust climate evidence, the importance of such an approach is both scientific and societal.

References

[1] IPCC. Summary for policymakers - climate change 2021: The physical science basis. contribution of working group i to the sixth assessment report of the intergovernmental panel on climate change. Technical report, IPCC, 2021.

[2] Baldocchi D. D. How eddy covariance flux measurements have contributed to our understanding of global change biology. *Global Change Biology*, 26, 2020.

[3] Obleitner F., Spötl C., Trüssel M., and Niederberger B. Climate warming detected in caves of the European Alps. *Scientific Reports*, 2024.

[4] GCOS. Gcos 244–2022. the 2022 gcos implementation plan, 2022.

[5] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML. Evaluation of measurement data — Guide to the expression of uncertainty in measurement. Joint Committee for Guides in Metrology, JCGM 100:2008.

[6] WIGOS. *WMO-No. 1160: Manual on the WMO integrated global observing system*, 2019.

[7] GCOS. Gcos 226–2019. surface reference network (gsrn): Justification, requirements, siting and instrumentation options, 2019.

[8] USCRN. United states climate reference network, 2020. Accessed November 2020.

[9] A. Merlone, Beges G., Bottacin A., et al. Climatological reference stations: Definitions and requirements. *International Journal of Climatology*, 2024.

[10] WMO. *WMO-No. 1238: Manual on the high-quality global data management framework for climate*, 2019.

[11] Caussinus H. and Mestre O. Detection and correction of artificial shifts in climate series. *Applied Statistics*, 53, 2004.

[12] Domonkos P. Adaptive caussinus-mestre algorithm for networks of temperature series (acmant). *International Journal of Climatology*, 2011.

[13] Begert M. and Frei C. Long-term area-mean temperature series for switzerland—combining homogenized station data and high resolution grid data. *International Journal of Climatology*, 38, 2018.

[14] Trewin B., Braganza K., Fawcett R., et al. An updated long-term homogenized temperature data set for australia. *Geoscience Data Journal*, 7, 2020.

[15] Jones P. D. and Lister D. H. The urban heat island in central london and urban-related warming trends in central london since 1900. *Weather*, 64, 2009.

[16] WMO. *WMO-No. 8: Guide to Instruments and Methods of Observation*, 2024.

[17] Merlone A. Thermal metrology for climate: a review of projects, activities and open issues. *Measurement Science and Technology*, 32, 2021.

[18] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML. International Vocabulary of Metrology — Basic and general concepts and associated terms (VIM). Joint Committee for Guides in Metrology, JCGM 200:2012. (3rd edition).

[19] Musacchio C., Coppa G., and Merlone A. An experimental method for evaluation of the snow albedo effect on near-surface air temperature measurements. *Meteorological Applications*, 26, 2019.

[20] Coppa G., Quarello A., Steeneveld G.-J., Jandrić N., and Merlone A. Metrological evaluation of the effect of the presence of a road on near-surface air temperatures. *International Journal of Climatology*, 41, 2021.

[21] de Podesta M., Underwood R., Bevilacqua L., and Bell S. Air temperature measurement challenges in precision metrology. *Journal of Physics: Conference Series*, 1065, 2018.

[22] Robinson C. S. Radiation effects in thermometry. *The Journal of Industrial and Engineering Chemistry*, 13(9), 1921.

[23] Fuchs M. and Tanner C. B. Radiation shields for air temperature thermometers. *Journal of Applied Meteorology*, 4, 1965.

[24] Biltoft C. A. Calibration and quality control for new meteorological instrumentation, part ii. Technical Report DPG-FR-91-701, U.S. Army Dugway Proving Ground, 1990. Final Report.

[25] Nakamura R. and Mahrt L. Air temperature measurement errors in naturally ventilated radiation shields. *Journal of Atmospheric and Oceanic Technology*, 22, 2005.

[26] Erell E., Leal V., and Maldonado E. Measurement of air temperature in the presence of a large radiant flux: An assessment of passively ventilated thermometer screens. *Boundary-Layer Meteorology*, 114, 2005.

[27] de Podesta M., Bell S., and Underwood R. Air temperature sensors: dependence of radiative errors on sensor diameter in precision metrology and meteorology. *Metrologia*, 55, 2018.

[28] Bernard J., Kéravec P., Morille B., et al. Outdoor air temperature measurement: A semi-empirical model to characterize shelter performance. *Climate*, 7(26), 2019.

[29] Richardson S. J., F. V. Brock, Semmer S. R., and Jirak C. Minimizing errors associated with multiplate radiation shields. *Journal of Atmospheric and Oceanic Technology*, 16, 1999.

[30] Yang J., Liu Q., Dai W., and Ding R. A temperature error correction method for a naturally ventilated radiation shield. *Journal of Atmospheric and Solar–Terrestrial Physics*, 149, 2016.

[31] Dobre M., Sestan D., and Merlone A. Air t measurement uncertainty associated to a mounting configuration temperature sensor-radiation shield, 2018.

[32] Pearce J. V., Bevilacqua L., Underwood R., and Bell S. A validated multiphysics model of a meteorological instrument shelter. In *AIP Conf. Proc.*, volume 3230, 2024.

[33] Daniels G. E. Measurement of gas temperature and the radiation compensating thermocouple. *Journal of Applied Meteorology*, 7, 1968.

[34] Palmer T. Y. Comparison of aspirated and radiation-compensating thermocouples. *Fire Technology*, 1970.

[35] Lin X., Hubbard K. G., Walter-Shea E. A., Brandle J. R., and Meyer G. E. Some perspectives on recent in situ air temperature observations: modeling the microclimate inside the radiation shields. *J. Atmos. Ocean Tech.*, 18, 2001.

[36] Maruyama A., Matsumoto Y., and Nakagawa H. Multiple-globe thermometer for measuring the air temperature without an aspirated radiation shield. *Agricultural and Forest Meteorology*, 292–293, 2020.

[37] Underwood R., Gardiner T., Finlayson A., et al. A combined non-contact acoustic thermometer and infrared hygrometer for atmospheric measurements. *Meteorological Applications*, 22, 2015.

[38] Pisani M., Astrua M., and Merlone A. Non-contact thermometer for improved air temperature measurements. *Sensors*, 23(4), 2023.

[39] Merlone A., Lopardo G., Sanna F., et al. The meteomet project – metrology for meteorology: challenges and results. *Meteorological Applications*, 22, 2015.

[40] Merlone A., Sanna F., Beges G., et al. The meteomet2 project – highlights and results. *Measurement Science and Technology*, 29, 2018.

[41] Bertiglia F., Lopardo G., Merlone A., et al. Traceability of ground-based air-temperature measurements: A case study on the meteorological observatory of moncalieri (italy). *Int J Thermophys*, 36, 2015.

[42] Lopardo G., Bellagarda S., Bertiglia F., et al. A calibration facility for automatic weather stations. *Meteorol. Appl.*, 22, 2015.

[43] Aranda N. G. and Merlone A. On-site calibration procedure and uncertainty contributions on air temperature sensors. *International Journal of Thermophysics*, 45, 2024.

[44] Coppa G., Musacchio C., Becherini F., et al. On-site calibration of instruments in the arctic: assessment of temperature records at climate change tower in ny-Ålesund, svalbard. *Measurement Science and Technology*, 2024.

[45] Lopardo G., Bertiglia F., Curci S., Roggero G., and Merlone A. Comparative analysis of the influence of solar radiation screen ageing on temperature measurements by means of weather stations. *International Journal of Climatology*, 2013.

[46] Harrison R. G. Lag–time effects on a naturally ventilated large thermometer screen. *Quarterly Journal of the Royal Meteorological Society*, 137, 2011.

[47] Burt S. and de Podesta M. Response times of meteorological air temperature sensors. *Quarterly Journal of the Royal Meteorological Society*, 146, 2020.

[48] Burt S. and Baker D. V. Improved time constant of a newly released air temperature sensor and its implications. *Quarterly Journal of the Royal Meteorological Society*, 2025.

[49] Kowal A., Merlone A., and Sawiński T. Long-term stability of meteorological temperature sensors. *Meteorol Appl*, 27, 2020.

[50] García Izquierdo C., Hernández S., González A., et al. Evaluation of the self-heating effect in a group of thermometers used in meteorological and climate applications. *Meteorol. Appl.*, 26, 2019.

[51] Pavlasek P., Merlone A., Sanna F., et al. Determination of automatic weather station self-heating originating from accompanying electronics. *Meteorological Applications*, 27, 2020.

[52] Musacchio C., Coppa G., Begeš G., Hofstätter-Mohler C., Massano L., Nigrelli G., Sanna F., and Merlone A. Effect of snow-covered ground albedo on the accuracy of air temperature measurements. *Atmospheric Measurement Techniques*, 14, 2021.

[53] Garcia Izquierdo C., Coppa G., Hernández S., and Merlone A. Metrological evaluation of the building influence on air temperature measurements. *Atmosphere*, 15, 2024.

[54] Nielsen J., Bottacin A., Østergaard P., et al. Evaluation of the influence of rain on air surface temperature measurements. *Measurement Science and Technology*, 36, 2025.

[55] Rosiek S. and Batlles F. J. A microcontroller-based data-acquisition system for meteorological station monitoring. *Energy Conversion and Management*, 49, 2008.

[56] Whiteman C. D., Hubbe J. M., and Shaw W. J. Evaluation of an inexpensive temperature datalogger for meteorological applications. *Journal of Atmospheric and Oceanic Technology*, 17, 2000.

[57] Xiangjuan W., Wenhua L., Hongyan X., and Yingquan Z. Research on temperature channel calibration of data-acquisition unit of automatic weather station. In *The Tenth International Conference on Electronic Measurement & Instruments (ICEMI'2011)*, 2011.

[58] Carullo A., Parvis M., and Vallan A. A traveling standard for the calibration of data acquisition boards. *IEEE Transactions on Instrumentation and Measurement*, 2004.

[59] Rauth D. A. and Randal V. T. Analog-to-digital conversion. *IEEE Instrumentation & Measurement Magazine*, 8(4), 2005.

[60] Rapuano S., Daponte P., Balestrieri E., et al. Adc parameters and characteristics. *IEEE Instrumentation & Measurement Magazine*, 8(5), 2005.

[61] IEEE-SA. Ieee standard for terminology and test methods for analog-to-digital converters. Technical report, IEEE, 2023.

[62] Braudaway D. W. Uncertainty specification for data acquisition (daq) devices. *IEEE Transactions on Instrumentation and Measurement*, 55(1), 2006.

[63] White D. R., Jones K., Williams J. M., and Ramsey I. E. A simple resistance network for calibrating resistance bridges. *IEEE Transactions on Instrumentation and Measurement*, 46(5), 1997.

[64] Geršak G., Bojkovski J., Batagelj V., et al. On-site environmental conditions–do they really affect the measuring accuracy? *Journal of Testing and Evaluation*, 36(6), 2008.

[65] Siap+Micros S.p.A. *Data acquisition systems User Manual*, 2022.

[66] Fan Zhang N. The uncertainty associated with the weighted mean of measurement data. *Metrologia*, 43(3), 2006.

[67] van der Meulen J. P. and Brandsma T. Thermometer screen intercomparison in de bilt (the netherlands), part i: understanding the weather-dependent temperature differences. *Int. J. Climatol.*, 28, 2008.

[68] Brandsma T. and van der Meulen J. P. Thermometer screen intercomparison in de bilt (the netherlands), part ii: description and modeling of mean temperature differences and extremes. *Int. J. Climatol.*, 28, 2008.

[69] ISO 17714. Meteorology—air temperature measurements—test methods for comparing the performance of thermometer shields/screens and defining important characteristics. Standard, International Organization for Standardization, 2007.

[70] Baker H. D., Ryder E. A., and Baker N. H. *Temperature Measurement in Engineering*, volume II. Wiley, 1961.

[71] Michalski L., Eckersdorf K., Kucharski J., and McGhee J. *Temperature Measurement*. Wiley, 2001.

[72] Nielsen J. and Barendregt C. The use of thermistors for establishing the temperature conditions in a climatic chamber. In *Proc. of TEMPMEKO2004*, 2004.

[73] Bosma R. and Peruzzi A. climatic chamber for temperatures up to 180 °c and pressures up to 0.5 mpa. *Int. J. Thermophys.*, 35, 2014.

[74] van Geel J. L. W. A., Bosma R., van Wensveen J., and Peruzzi A. Thermistors used in a climatic chamber at high temperature and humidity. *Int. J. Thermophys.*, 36, 2015.

[75] Consultative Committee on Thermometry (CCT). *Guide on secondary thermometry, Thermistor Thermometry*. BIPM, 2014.

[76] Tegeler E. Radiation effects and its consequences on measurements in climatic chambers. In *Proc. of TEMPMEKO2004*, 2004.

[77] ISO 7726. Ergonomics of the thermal environment—Instruments for measuring physical quantities. Standard, International Organization for Standardization, 2007.

[78] Heinonen M. et al. Euramet p1061 comparison of air temperature calibrations, final report. Technical report, EURAMET, 2013.

[79] MeteoMet project (EURAMET JRP-contract number ENV07). metrology for pressure, temperature, humidity and airspeed in the atmosphere, final publishable jrp report. Technical report, EURAMET, 2015.

[80] Byers H. R., Moses H., and Harney P. J. Measurement of rain temperature. *J. Meteor.*, 6, 1949.

[81] Ehinger J. Wmo-no. 589: Siting and exposure of meteorological instruments. Technical Report WMO-No. 589, WMO, 1993.

[82] Geiger R. *The Climate Near the Ground*. Harvard University Press, 1950.

[83] Best A. C. Transfer of heat and momentum in the lowest layers of the atmosphere. *Geophysical Memoirs*, 65, 1935.

[84] Baum W. A. On the vertical distribution of mean temperature within the microclimatic layer. *Bulletin of the American Meteorological Society*, 29, 1948.

[85] Yokoi T. Measurements of the air temperature profile near the ground by two laser beams. *Journal of Atmospheric and Oceanic Technology*, 3, 1986.

[86] Högström U. Non-dimensional wind and temperature profiles in the atmospheric surface layer: A re-evaluation. *Boundary-Layer Meteorol*, 42, 1987.

[87] Högström U. Review of some basic characteristics of the atmospheric surface layer. *Boundary-Layer Meteorol*, 78, 1996.

[88] Arya S. P. Micrometeorology and atmospheric boundary layer. *Pure and Applied Geophysics*, 162, 2005.

[89] Foken T. 50 years of the monin–obukhov similarity theory. *Boundary-Layer Meteorol*, 119, 2006.

[90] Boekee J. et al. Rethinking the roughness height: An improved description of temperature profiles over short vegetation. *Boundary-Layer Meteorol*, 190, 2024.

[91] IMO. Conference of directions (washington 1947) final report. Technical report, International Meteorological Organization, 1947.

[92] Tannehill I. R. Meetings of the international meteorological organization (i. m. 0.) in toronto and washington, 1947. *Bulletin of the American Meteorological Society*, 1947.

[93] IMO. Provisional guide to international meteorological instrument and observing practice. Technical report, International Meteorological Organization, 1950.

[94] Scott R.H. *Instructions in the use of meteorological instruments*. 1875.

[95] Sparks W. R. Wmo-no. 315: The effect of thermometer screen design on the observed temperature. Technical Report WMO-No. 315, WMO, 1972.

[96] Pearce J. V., Rusby R. L., Yamazawa K., Rudtsch S., Iacomini L., Lopardo G., White D. R., and Tew W. L. *Guide to Secondary Thermometry: Industrial Platinum Resistance Thermometers*. Consultative Committee for Thermometry under the auspices of the International Committee for Weights and Measures, Bureau International des Poids et Mesures, 2021. Last updated 22 November 2021.

[97] Stull R.B. *An Introduction to Boundary Layer Meteorology*. Kluwer Academic Publishers, 1988.

[98] Kaimal J. C. and Finnigan J. J. *Atmospheric Boundary Layer Flows: Their Structure and Measurement*. Oxford University Press, 1994.

[99] Aubinet M., Vesala T., and Papale D. *Eddy Covariance: A Practical Guide to Measurement and Data Analysis*. Springer, 2012.

[100] Foken T. *Micrometeorology*. Springer, 2017.

[101] Bonan G. B. *Ecological climatology: concepts and applications*. Cambridge university press, 2003.

[102] Monson R. K. and Baldocchi D. D. *Ecosystem-atmosphere interactions*. Cambridge University Press, 2014.

[103] Baldocchi D. D., Chu H., and Reichstein M. On the processing of net ecosystem exchange data for the ameriflux network. *Bulletin of the American Meteorological Society*, 99, 2018.

[104] Njoroge S., Smith P., and Caddeo A. Monitoring, reporting and verification of soil organic carbon change: A review. *Soil Security*, 6, 2022.

[105] Moore D., Novick K., and Papale D. Estimating IPCC emission factors for Agricultural and Forest Ecosystems using a multi-decade standardized eddy covariance data record (FLUXNET), 2025. EGU General Assembly 2025.

[106] Burba G. Harvesting carbon: Lakes, ponds & wetlands a guide to utilizing direct flux measurements for assessment and verification of carbon sequestration and GHG emission rates over areas, 2025. ESS Open Archive, preprint.

[107] Rebmann C. et al. Icos eddy covariance flux-station site setup: a review. *International Agrophysics*, 32, 2018.

[108] Sabbatini ans others S. Eddy covariance raw data processing for co2 and energy fluxes calculation at icos ecosystem stations. *International Agrophysics*, 32, 2018.

[109] Papale D. et al. Towards a standardized processing of net ecosystem exchange measured with eddy covariance technique: algorithms and uncertainty estimation. *Biogeosciences*, 3, 2006.

[110] Baldocchi D. D. and Peñuelas J. How forests write the poetry of the anthropocene. *New Phytologist*, 222, 2019.

[111] Jung M. et al. Scaling carbon fluxes from eddy covariance sites to globe: a model-data integration approach. *Reviews of Geophysics*, 58, 2020.

[112] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML. Guide to the expression of uncertainty in measurement — Part 6: Developing and using measurement models. Joint Committee for Guides in Metrology, JCGM GUM-6:2020.

[113] Finkelstein P. L. and Sims P. F. Sampling error in eddy correlation flux measurements. *Journal of Geophysical Research: Atmospheres*, 106, 2001.

[114] Hollinger D. Y. and Richardson A. D. The uncertainty in eddy covariance measurements and its application to data filtering. *Global Change Biology*, 11, 2005.

[115] Mauder M. et al. A strategy for quality and uncertainty assessment of long-term eddy-covariance measurements. *Agricultural and Forest Meteorology*, 169, 2013.

[116] Rannik Ü., Peltola O., and Mammarella I. Random uncertainties of flux measurements by the eddy covariance technique. *Atmospheric Measurement Techniques*, 9, 2016.

[117] Fratini G., McDermitt D. K., and Papale D. Eddy-covariance flux errors due to biases in gas concentration measurements: origins, quantification and correction. *Biogeosciences*, 11, 2014.

[118] Fratini G. et al. On the impact of synchronization errors on eddy covariance fluxes. *Atmospheric Measurement Techniques*, 11, 2018.

[119] Taylor J. R. *An introduction to error analysis: the study of uncertainties in physical measurements*. University Science Books, 1997.

[120] Burba G. *Eddy Covariance Method for Scientific, Regulatory, and Commercial applications*. LI-COR Biosciences, 2022.

[121] Van der Hoven I. Power spectrum of horizontal wind speed in the frequency range from 0.0007 to 900 cycles per hour. *Journal of the Atmospheric Sciences*, 1957.

[122] Gash J. H. C. and Culf A. D. Applying a linear detrend to eddy correlation data in real time. *Boundary-Layer Meteorol.*, 79, 1996.

[123] Lenschow D. H., Mann J., and Kristensen L. How long is long enough when measuring fluxes and other turbulence statistics? *Journal of Atmospheric and Oceanic Technology*, 11, 1994.

[124] Wilczak J. M., Oncley S. P., and Stage S. A. Sonic anemometer tilt correction algorithms. *Boundary-Layer Meteorol.*, 99, 2001.

[125] Moore C. J. Frequency response corrections for eddy correlation systems. *Boundary-Layer Meteorol.*, 37, 1986.

[126] Nakai T. and Shimoyama K. Correction of sonic anemometer angle of attack errors. *Agric. Forest Meteorol.*, 136, 2006.

[127] Nakai T. and Shimoyama K. Ultrasonic anemometer angle of attack errors under turbulent conditions. *Agric. Forest Meteorol.*, 162–163, 2012.

[128] van der Molen M. K., Moors E. J., and Elbers J. A. Angle of attack dependent calibration of sonic anemometers. *Agric. Forest Meteorol.*, 122, 2004.

[129] Kochendorfer J. et al. How Well Can We Measure the Vertical Wind Speed? Implications for Fluxes of Energy and Mass. *Boundary-Layer Meteorology*, 145, 2012.

[130] Webb E. K., Pearman G. I., and Leuning R. Correction of flux measurements for density effects due to heat and water vapour transfer. *Quart. J. R. Met. Soc.*, 106, 1980.

[131] Lee X. and Massman W. J. A perspective on thirty years of the webb, pearman and leuning density corrections. *Boundary-Layer Meteorol.*, 139, 2011.

[132] Burba G., Anderson T., and Komissarov A. Accounting for spectroscopic effects in laser-based open-path eddy covariance flux measurements. *Glob. Change Biol.*, 25, 2019.

[133] McDermitt D. K., Welles J. M., and Eckles R. D. Effects of Temperature, Pressure and Water Vapor on Gas Phase Infrared Absorption by CO2. Application note, LI-COR, inc., 1993.

[134] LI-COR Biosciences. The Importance of Water Vapor Measurements and Corrections. Application Note AppNote129 / IRG4-110, LI-COR Biosciences, 2023.

[135] Mammarella I. et al. Quantifying the uncertainty of eddy covariance fluxes due to the use of different software packages and combinations of processing steps in two contrasting ecosystems. *Atmos. Meas. Tech.*, 2016.

[136] Billesbach D. P. Estimating uncertainties in individual eddy covariance flux measurements: comparison of methods and a proposed new method. *Agric. Forest Meteorol.*, 151, 2011.

[137] Lenschow D. H., Wulfmeyer V., and Senff C. Measuring second- through fourth-order moments in noisy data. *Journal of Atmospheric and Oceanic Technology*, 17, 2000.

[138] van der Veen A. M. H. Uncertainty propagation in calibration hierarchies. *Metrologia*, 57, 2020.

[139] Lasslop G., Reichstein M., Papale D., et al. On the choice of the driving temperature for eddy-covariance carbon dioxide flux partitioning. *Agricultural and Forest Meteorology*, 150, 2010.

[140] Malengo A. and Pennecchi F. R. A generalized least-squares method for the analysis of interlaboratory comparisons with internal consistency. *Metrologia*, 50, 2013.

[141] Arriga N. et al. Mediterranean pine forests: Comparison of fluxes and tree rings of pinus pinaster aiton and pinus pinea l. *Agricultural and Forest Meteorology*, 373, 2025.

[142] Arriga N., Goded I., and Manca G., 2022. ICOS Ecosystem thematic centre, 2022. Warm winter 2020 ecosystem eddy covariance flux product from San Rossore 2.

[143] Arriga N. et al., 2025. ETC L2 ARCHIVE from San Rossore 2, 2019–2024, ICOS RI.

[144] Terao Y. et al. Final report on the apmp air speed key comparison (apmp.m.ff-k3). *Metrologia*, 47, 2010.

[145] UNI 6143. Analisi del gas - Metodo comparativo per la determinazione e la verifica della composizione delle miscele di gas per calibrazione. Standard, UNI - Ente Italiano di Normazione, 2007.

[146] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML. Evaluation of measurement data — Supplement 1 to the "Guide to the expression of uncertainty in measurement" — Propagation of distributions using a Monte Carlo method. Joint Committee for Guides in Metrology, JCGM 101:2008.

[147] Mahrt L. Flux sampling errors for aircraft and towers. *Journal of Atmospheric and Ocean Tehcnology*, 1998.

[148] Wilks D. S. *Statistical methods in the atmospheric sciences*, volume 100. Academic press, 2006.

[149] Badino G. Cave temperatures and global climatic change. *Int. J. Speleol.*, 33, 2004.

[150] Badino G. Underground Meteorology - "What's the weather underground". *Acta Carsologica*, 39, 2010.

[151] Domínguez-Villar D. et al. Is global warming affecting cave temperatures? Experimental and model data from a paradigmatic case study. *Climate Dynamics*, 45, 2015.

[152] Mammola S. et al. Climate change going deep: The effects of global climatic alterations on cave ecosystems. *The Anthropocene Review*, 6, 2019.

[153] Mejía-Ortíz L., Christman M. C., Pipan T., and Culver D. C. What's the temperature in tropical caves? *PLoS ONE*, 15, 2020.

[154] Beltrami M., Ferguson G., and Harris R. N. Long-term tracking of climate change by underground temperatures. *Geophys. Res. Lett.*, 32, 2005.

[155] M. M., Luetscher and Jeannin P.-Y. Temperature distribution in karst systems: the role of air and water fluxes. *Terra Nova*, 16, 2004.

[156] Ravbar N. and Kosutnik J. Variations of karst underground air temperature induced by various factors (Cave of Županova jama, Central Slovenia). *Theor Appl Climatol*, 116, 2014.

[157] Gabrovšek F. How do caves breathe: The airflow patterns in karst underground. *PLoS ONE*, 18, 2023.

[158] Perrier F., Le Mouël J.-L., and Richon P. Spatial and Temporal Dependence of Temperature Variations Induced by Atmospheric Pressure Variations in Shallow Underground Cavities. *Pure Appl. Geophys.*, 167, 2010.

[159] Perrier F. et al. Temperature variations in caves induced by atmospheric pressure variations—Part 1: Transfer functions and their interpretation. *Geosystems and Geoenvironment*, 2, 2023.

[160] Perrier F. et al. Temperature variations in caves induced by atmospheric pressure variations—Part 2: Unveiling hidden thermal signals. *Geosystems and Geoenvironment*, 2, 2023.

[161] Šebela S., Baker G., and Luke B. Cave Temperature and Management Implications in Lehman Caves, Great Basin National Park, USA. *Geoheritage*, 11, 2019.

[162] Šebela S. and Turk J. Comparison of historical and current temperatures in show caves (Slovenia). *SN Applied Sciences*, 4, 2022.

[163] Medina M. J. et al. Temperature variation in caves and its significance for subterranean ecosystems. *Sci. Rep.*, 13, 2023.

[164] Bourges F. et al. Conservation of prehistoric caves and stability of their inner climate: Lessons from Chauvet and other French caves. *Science of the Total Environment*, 493, 2014.

[165] Rotta Loria A. F. The silent impact of underground climate change on civil infrastructure. *Communications Engineering*, 2023.

[166] Peyraube N. et al. Designing a cave air monitoring system: Guide and feedback from 15 years of monitoring the Cussac Cave (France). *International Journal of Speleology*, 54, 2025.

[167] Peano G. Il monitoraggio ambientale nella grotta di Bossea: problemi tecnici e soluzioni adottate. *Le Grotte d'Italia*, 3, 2002.

[168] IEC. Industrial platinum resistance thermometers and platinum temperature sensors. Standard, International Electrotechnical Commission, 2022.

[169] Messina G. Temperature, pressure and carbon dioxide trends in karst environment. Master's thesis, 2023.

Appendix A

A.1 key sources of uncertainty in climate reference measurements

Table A.1 summarizes key uncertainty sources resulting from instrumental components, environmental effects, measurand definition and statistical processing.

A.2 Effective sensitivity coefficients for EC

The analytical expressions of the effective sensitivity coefficients described in Section 5.2 are here reported. The key components are the partial derivatives $\frac{\partial F}{\partial X_i}$ for each variable:

$$\frac{\partial F}{\partial P_i} = \frac{1}{N_{\text{met}}} \frac{F}{R\overline{T}} \tag{A.1}$$

$$\frac{\partial F}{\partial T_i} = -\frac{1}{N_{\text{met}}} \frac{F\overline{P}}{R\overline{T}^2} \tag{A.2}$$

$$\frac{\partial F}{\partial S_i} = \frac{\overline{\rho}}{N_{\text{turb}}} (AV'_{x,i} + BV'_{y,i} + CV'_{z,i}) \tag{A.3}$$

$$\frac{\partial F}{\partial V_{x,i}} = \frac{\overline{\rho}}{N_{\text{turb}}} (D_x + AS'_i) \tag{A.4}$$

$$\frac{\partial F}{\partial V_{y,i}} = \frac{\overline{\rho}}{N_{\text{turb}}} (D_y + AS'_i) \tag{A.5}$$

$$\frac{\partial F}{\partial V_{z,i}} = \frac{\overline{\rho}}{N_{\text{turb}}} (D_z + AS'_i) \tag{A.6}$$

The coefficients $D_{x/y/z}$ represent the partial derivatives of the covariances sum with respect to the average wind speed values:

$$D_x = \frac{\partial A}{\partial \overline{V_x}}\overline{V_x'S'} + \frac{\partial B}{\partial \overline{V_x}}\overline{V_y'S'} + \frac{\partial C}{\partial \overline{V_x}}\overline{V_z'S'} \tag{A.7}$$

$$D_y = \frac{\partial A}{\partial \overline{V_y}}\overline{V_x'S'} + \frac{\partial B}{\partial \overline{V_y}}\overline{V_y'S'} + \frac{\partial C}{\partial \overline{V_y}}\overline{V_z'S'} \tag{A.8}$$

$$D_z = \frac{\partial A}{\partial \overline{V_z}}\overline{V_x'S'} + \frac{\partial B}{\partial \overline{V_z}}\overline{V_y'S'} + \frac{\partial C}{\partial \overline{V_z}}\overline{V_z'S'} \tag{A.9}$$

To evaluate the above expressions, it is necessary to calculate the partial derivatives of the rotation coefficients (reported in Equations 5.7-5.9) with respect to the three average wind speed components, for a total of 9 expressions:

$$\frac{\partial A}{\partial \overline{V_x}} = \frac{\overline{V_z}(\overline{V_x}^4 - \overline{V_y}^4 - \overline{V_y}^2\overline{V_z}^2)}{(\overline{V_x}^2 + \overline{V_y}^2)^{\frac{3}{2}}(\overline{V_x}^2 + \overline{V_y}^2 + \overline{V_z}^2)^{\frac{3}{2}}} \tag{A.10}$$

$$\frac{\partial A}{\partial \overline{V_y}} = \frac{\overline{V_x}\,\overline{V_y}\,\overline{V_z}(2\overline{V_x}^2 + 2\overline{V_y}^2 + \overline{V_z}^2)}{(\overline{V_x}^2 + \overline{V_y}^2)^{\frac{3}{2}}(\overline{V_x}^2 + \overline{V_y}^2 + \overline{V_z}^2)^{\frac{3}{2}}} \tag{A.11}$$

$$\frac{\partial A}{\partial \overline{V_z}} = -\frac{\overline{V_x}(\overline{V_x}^2 + \overline{V_y}^2)^{\frac{1}{2}}}{(\overline{V_x}^2 + \overline{V_y}^2 + \overline{V_z}^2)^{\frac{3}{2}}} \tag{A.12}$$

$$\frac{\partial B}{\partial \overline{V_x}} = \frac{\overline{V_x}\,\overline{V_y}\,\overline{V_z}(2\overline{V_x}^2 + 2\overline{V_y}^2 + \overline{V_z}^2)}{(\overline{V_x}^2 + \overline{V_y}^2)^{\frac{3}{2}}(\overline{V_x}^2 + \overline{V_y}^2 + \overline{V_z}^2)^{\frac{3}{2}}} \tag{A.13}$$

$$\frac{\partial B}{\partial \overline{V_y}} = \frac{\overline{V_z}(\overline{V_y}^4 - \overline{V_x}^4 - \overline{V_x}^2\overline{V_z}^2)}{(\overline{V_x}^2 + \overline{V_y}^2)^{\frac{3}{2}}(\overline{V_x}^2 + \overline{V_y}^2 + \overline{V_z}^2)^{\frac{3}{2}}} \tag{A.14}$$

$$\frac{\partial B}{\partial \overline{V_z}} = -\frac{\overline{V_y}(\overline{V_x}^2 + \overline{V_y}^2)^{\frac{1}{2}}}{(\overline{V_x}^2 + \overline{V_y}^2 + \overline{V_z}^2)^{\frac{3}{2}}} \tag{A.15}$$

$$\frac{\partial C}{\partial \overline{V_x}} = \frac{\overline{V_x}\,\overline{V_z}^2}{(\overline{V_x}^2 + \overline{V_y}^2)^{\frac{1}{2}}(\overline{V_x}^2 + \overline{V_y}^2 + \overline{V_z}^2)^{\frac{3}{2}}} \tag{A.16}$$

$$\frac{\partial C}{\partial \overline{V_y}} = \frac{\overline{V_y}\,\overline{V_z}^2}{(\overline{V_x}^2 + \overline{V_y}^2)^{\frac{1}{2}}(\overline{V_x}^2 + \overline{V_y}^2 + \overline{V_z}^2)^{\frac{3}{2}}} \tag{A.17}$$

$$\frac{\partial C}{\partial \overline{V_z}} = -\frac{\overline{V_z}(\overline{V_x}^2 + \overline{V_y}^2)^{\frac{1}{2}}}{(\overline{V_x}^2 + \overline{V_y}^2 + \overline{V_z}^2)^{\frac{3}{2}}} \tag{A.18}$$

The effective sensitivity coefficient of the variable X is obtained by combining the above expressions, as shown in Equations 5.16 and 5.17. The analytical expressions

for general instrumental correlation coefficients, considering the homoscedastic assumption, are the following:

$$\overline{J_P} = \sqrt{\frac{1+r(P)(N_{\text{met}}-1)}{N_{\text{met}}}}\frac{|F|}{R\overline{T}} \tag{A.19}$$

$$\overline{J_T} = \sqrt{\frac{1+r(T)(N_{\text{met}}-1)}{N_{\text{met}}}}\frac{|F|\,\overline{P}}{R\overline{T}^2} \tag{A.20}$$

$$\overline{J_S} = \sqrt{\frac{1-r(S)}{N_{\text{turb}}}}\overline{\rho}\sqrt{\begin{bmatrix}A & B & C\end{bmatrix}\begin{bmatrix}\sigma_t^2(V_x') & \overline{V_x'V_y'} & \overline{V_x'V_z'}\\ \overline{V_y'V_x'} & \sigma_t^2(V_y') & \overline{V_y'V_z'}\\ \overline{V_z'V_x'} & \overline{V_z'V_y'} & \sigma_t^2(V_z')\end{bmatrix}\begin{bmatrix}A\\B\\C\end{bmatrix}} \tag{A.21}$$

$$\overline{J_{V_x}} = \frac{\overline{\rho}}{\sqrt{N_{\text{turb}}}}\sqrt{(1+r(V_x)(N_{\text{turb}}-1))D_x^2+(1-r(V_x))A^2\sigma_t^2(S)} \tag{A.22}$$

$$\overline{J_{V_y}} = \frac{\overline{\rho}}{\sqrt{N_{\text{turb}}}}\sqrt{(1+r(V_y)(N_{\text{turb}}-1))D_y^2+(1-r(V_y))B^2\sigma_t^2(S)} \tag{A.23}$$

$$\overline{J_{V_z}} = \frac{\overline{\rho}}{\sqrt{N_{\text{turb}}}}\sqrt{(1+r(V_z)(N_{\text{turb}}-1))D_z^2+(1-r(V_z))C^2\sigma_t^2(S)} \tag{A.24}$$

$$\overline{H}_{V_xV_yV_z} = 2\overline{\rho}\,r(V_x,V_y,V_z)\,(D_xD_y+D_xD_z+D_yD_z) \tag{A.25}$$

Table A.1 List of some key sources of uncertainty contributing to the overall budget. The first and second columns report the general sources of uncertainty and useful references. The corresponding components are listed in more detail in the third column.

Sources of uncertainty	**Components**	**Details**
Instrument *References:* • MQCS • Calibration guidelines • GUM	Measurement System and Calibration	• Construction quality • Resolution • Instrument and logger calibration • Linearity • Hysteresis • Time constant • Drift with temperature • Sampling method • Sampling frequency • Processing algorithm • Digitization and rounding • Response time
	Instrument Coupling	• Radiation screen • Static pressure head • Rain gauge fence screen
	Maintenance and Verification	• Frequency of maintenance • Instrument and system drift with time • Instrument and system ageing • Instrument and system faults (that affect data but do not cause failure) • Cleanliness of instrument and site • Sensor mechanical stress during transport and operation
Environment Effects *References:* • MQCS • Scientific literature • GUM	Effects on instruments, not detectable in laboratory	• Evaporation of precipitation on screen (overcooling) • Wind effects on measurement • Condensation on temperature instrument • Solar radiation effects on measurement, including reflected radiation • Icing • Exposure to extreme limits of use
Site *References:* • Siting classification • Scientific literature	Effects of obstacles at less than 100 m from the measuring points	• Roads • Trees • Building • Water sources • Slopes
Statistical components *References:* • GUM	Type A uncertainties	• Datalogger sampling procedures, mean, standard deviation • Statistics on big data