11. Guidelines for data analysis from dynamic experimental campaigns. Physical aspects

11.1 Summary

This chapter presents guidelines for using time series analysis methods and tools for estimating the thermal performance of buildings and building components. The specific target is to obtain key performance metrics such as heat loss coefficients, time constants, solar aperture, effective thermal capacity etc.

The document is integrated in a more comprehensive work. This chapter is the first part is mainly dealing with physical aspects and specific complexity and problems that may occur due to the experimental conditions. It may be considered as a question: what quality and what information does the data contain for analysis? Minimum steps to carry out data analysis are reported and different alternative analysis approaches are outlined.

The focus is mainly on the most critical aspects particularly regarding energy performance assessment of buildings and building components. More general techniques also required to carry out data analysis are briefly presented in this document including references for more comprehensive information.

Common exercises described in previous chapters facilitated to identify frequent mistakes leading to unjustified high spread in the results and inaccurate parameter estimates. This chapter focuses on criteria that must be considered to avoid these mistakes. Graphs are used to illustrate the considered aspects.

It is assumed that the reader is familiar with basic principles of heat transfer. There are many text books dealing with this topic such as Refs. 17, 27, etc. It is also assumed that the reader has some background on measurement techniques that are well described in the literature and standards (Refs. 28 to 31, etc.).

A case study is considered to facilitate the understanding of some of the recommendations given in this report. The presented case study consists of a round robin test box built in the framework of IEA EBC Annex 58. References are given for additional case studies that help to understand the different aspects discussed and are included within this document.

A second part (Report of Subtask 3 – Part 2, Ref. 19) focuses on statistical aspects and may help to make decisions in choosing a correct model and analysis of residuals. Both documents must be considered as complementary and on some occasion they may overlap, in particular for processing data for input to the modelling work.

11.2 Introduction

Analysis and modelling of data obtained from experiments under real climate conditions require special attention to the treatment of the data during all steps of the elaboration process. The interest for these techniques and their application has grown in recent years by industry. This interest has pushed standardisation activities such as CEN/TC 89/WG13 and research initiatives such as IEA EBC Annex 58. In general it concerns numerous observations by measurements at regular interval of physical processes. For the case of experimental work and analysis for the energy performance assessment of buildings the physical processes are importantly thermal transfer between a controlled indoor environment and a variable outdoor environment. In principle all these thermal transfer processes are well known physical ones, e.g. conduction, convection and radiation. On many occasions data is produced by people carrying out the technical work of setting up an experiment and controlling the process of data acquisition. Raw data is made available for analysis with the purpose of producing one or a few output results (see Table 2 for clarification).

Often mathematicians do not have the profound knowledge about the experiment and receive the data with limited information. The guidelines will address therefore also a brief introduction on the most common issues and will address several issues that deal with examining the data before any data treatment takes place.

As an introduction some basic information on temperature measurements is given as it is considered as important for a proper analysis of the measured data. The measurement of temperatures and thermal flows is performed by sensors based on the applied physical properties of the sensitive part of it: resistances (PT100), thermo-couples (like Cu-Co) and electronic devices. A correct measurement of the target temperature is required and a closer look will be given within the context of thermal performance of a building corresponding to the transfer of heat through the building envelope.

In

Figure *94* a schematic view is given of a building for which it is important to recognise that an indoor- and outdoor environment exists, separated by the building envelope. The heat transfer is importantly defined by the air temperatures.

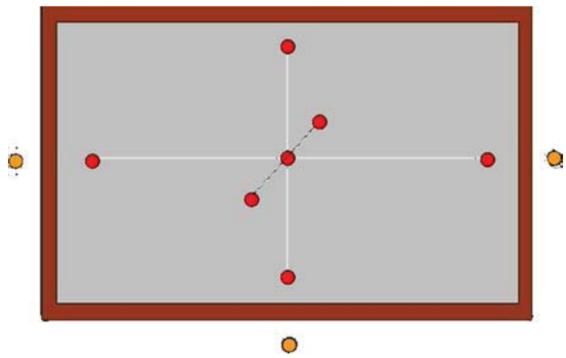


Figure 94: Schematic view of a building.

To guarantee an accurate temperature measurement, the temperature of the sensing element must be identical to the temperature of the measurand. Different strategies can be followed to fit this objective, depending on the kind of temperature that is measured. For example:

- Air temperature measurement: Convection is the main mechanism helping to equalise air and sensor temperatures so it must be enhanced, and the conductive and radiative heat transfer should be avoided. This can be achieved by using a fan and shielding of the sensor.
- Surface temperature measurement: a surface is the separation between two distinctive entities, usually air or water and solid. Installation of the sensor must guarantee, that its response to the different heat transfer phenomena (conduction, convection, and radiation), must be as similar as possible to the response of the measured surface. This is achieved using small sensing elements and integrating them as much as possible with the corresponding surface. This may be done by different techniques depending on the surface, for example: painting the sensors in the same paint as the surface, or covering them with tapes having similar properties

as the surfaces, etc. If sensor integration is not good enough high uncertainties can be found mainly associated to the different response of sensor and measured surface to the radiative component.
- Comfort temperature measurement: the perceived temperature from radiation and convection is usually measured with a black bulb. The conductive thermal component is avoided.

Radiation is the main source of wrong temperature measurements that give false signals from the sensor and hence false information to the mathematical models about the physical processes. The disturbing radiation may arrive from solar radiation, heat sources such as badly shielded electric heaters and incandescent light bulbs. Shielding of air temperature sensors is therefore necessary in particular for those that could be hit by solar radiation near to window openings and those sensors that are placed in a space where electric heaters or light bulbs are used. Ambient air temperature is usually measured with a special device that is double shielded and ventilated by natural or forced ventilation (with a small fan that draws air past the sensor). See Figure 96.

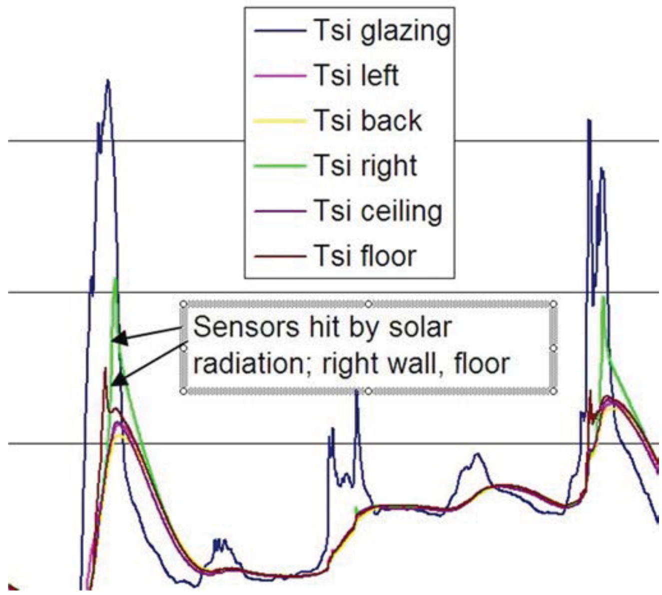


Figure 95: Surface temperature sensors hit by solar radiation as function of time.

Figure 96: Air temperature measurement.

In Figure 95 an example is given of the effect of solar radiation on the sensor. The temperature given by the sensors not hit by solar radiation is lower than the temperature given by sensors hit by solar radiation. Compare the surface temperature of the ceiling with that of the floor or right wall. One important question here is: does this sensor installation guarantee that the temperatures of the sensing elements are the same as the measured surfaces? Good documentation of experiment set up including detailed description of sensor installation is an indispensable complement for high quality data sets, and would be very helpful to answer this kind of questions.

As schematically illustrated in Figure 94, air temperatures are measured at different places. The reason is that for indoor environments without forced ventilation and sometimes where there is mechanical ventilation (for example with displacement ventilation systems), stratification occurs that can result in temperature difference in the order of several degrees. When the model is working with one signal representing the indoor temperature one has to know the uncertainty of that signal. The experiment can reduce that uncertainty by using a ventilator guaranteeing a certain limit on the stratification which provides a better quality input signal for the mathematical model. In Figure 97, the differences to average and extremes (minimum and maximum of 7 sensors as positioned in Figure 94) are given for a space of 37m^3 that can serve as input for analysis work. Note that the increase is during a 24 hour heating period.

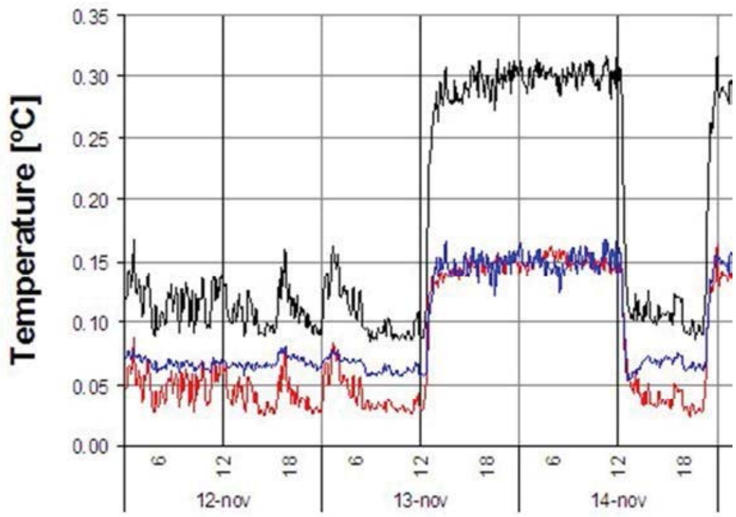


Figure 97: Indoor air temperatures; difference to average

The main conclusion from this introductory section is that in general one has to understand what a measured signal represents. What information is available from the single sensor signal or from a group of signals (such as the average from 7 indoor air temperatures that are supposed to represent one indoor air temperature)?

To get more knowledge about measurements and what information is contained in the observations a pragmatic approach for checking is proposed:

- Graphical plot of signals; it may indicate outliers, sudden changes as function of expected time constants.
- Statistical methods; average and variance of group of sensors; check of expected limits.

The measured data leads to ‘raw data’ with a certain accuracy. It should be stored and documented for later analysis or modelling work.

It leads to the following overview of the interaction between experimental work and analysis of the obtained measured data. It can be applied to heat transfer through the building envelope, either a wall or the whole building.

Table 1: Summary of interaction between experimental work and data analysis

Physical processes and Experiments	**Mathematics and Statistics**
Physical Object, Processes, Parameters	
Feedback	General mathematical description of methodology for object, processes, etc.
Experimental set-up Data information required Data collection	
Data pre-processing	Mathematical method, model and parameters Analysis techniques
Parameter conversion	Parameter identification and Mathematical assessment
Physical parameters and performance expression	
	Reporting of performance value and estimated uncertainty

One may recognise 8 steps:

1. the measurement of several phenomena of the physical process under investigation
2. the production and storage of raw data including quality aspects
3. pre-processing of data at measurement level (controlled by the acquisition system)
4. processing of data for model input (depends on analysis method)
5. representation of the physical system by a mathematical model
6. model identification from measured data
7. post-processing of the results and applied model
8. conversion and reporting of the final result, including uncertainty

Part 1 of the guidelines is mainly concerned with items 4, 5 and 7. Some overlap may occur where items 5 and 7 are concerned. Part 2 of the guidelines are dealing mainly with items 6 and 7.

As may be clear from the statements above, the measurements and the analysis of the observed data are dealing with variable conditions, sometimes induced, sometimes unexpected. The way to deal with it is the use of appropriate dynamic mathematical and statistical techniques complemented with general knowledge of the physical processes.

The next paragraph introduces the application of dynamic methods for the assessment of limited number of physical parameters to be identified from a huge amount of available data.

Dynamic tests allow modelling buildings and building components from experimental test campaigns carried out under dynamic test conditions. One of the main strength of these methods is that they permit the extraction of intrinsic characteristic parameters from time varying measurements. These features are very useful to carry out energy performance assessment of “as built” buildings, under outdoors weather and in occupied conditions where these conditions are dynamic.

These analysis approaches must be able to deal with features of the particular experimental conditions of these test campaigns. Although some characteristic parameters can be obtained from dynamic tests applying steady state approaches under certain test conditions (section 11.6.3), it is evident that, in general, time varying measurements call for the application of system identification techniques and time series analysis tools. However as built, in outdoors weather and in occupancy conditions requires the capability to take into account any other phenomena brought by these test conditions. In practice, this means that these test conditions require modelling additional terms to complete energy balance equations that may be not necessary in other different test conditions (such as well controlled tests in laboratories and steady state).

These real test conditions are usually linked to many complex physical phenomena, which case modelling may become very difficult. In this case simplifications criteria discussed in section 11.6.2 play an important role to obtain accurate results.

GENERAL CONCEPT

The objective is the identification of a mathematical modelling with application to energy performance assessment in the built environment. The problem is stated as: from many measured data to a few estimated values.

Table 2: Process to obtain a few estimated values (characteristic parameters) from many measured data. Main elements of the methodology.

INPUT	METHODOLOGY	OUTPUT
Many observations from time and space Physical processes Literature General knowledge	Description of physical processes into mathematical equations. Method should fulfil the aim taking into account the searched output	Limited value(s) Period Performance Efficiency Data
	Pre-processing Model choice Iteration process Post-processing Statistical tests Model validation External tests	

How to get from many observations as input for the calculation process to one or a few limited output values for reporting? In that process the accuracy of input data, the propagation of the errors in the calculation process and the required accuracy of the reported value are of high importance.

Once data has been produced (raw data), from a dedicated experiment, it is assumed that these data contain all information describing the physical processes that a mathematical model is supposed to analyse. Treatment of the raw data is therefore crucial and should be performed by someone who has knowledge about the physical processes as well as the experimental set-up. Pre-processing of data for the purpose of mathematical modeling is therefore important and should be carried out with caution.

Reduction of observations and signals on the input side implies the examination of the uncertainty of the input data to the calculation model. More about this aspect can be found below and illustrated in the figures.

It is very frequent confusing and misinterpreting dynamic test conditions with time dependent parameters. Test conditions can be steady state or dynamic. Parameters can be constant (intrinsic) or time varying. Dynamic analysis must be robust giving stable estimates for constant parameters and allowing the identification of clear dependencies for non-constant indicators. Dynamic conditions don’t change classical definitions of physical parameters. Equations used to extract these parameters from experimental data must take into account all the relevant effects which are present in given test conditions. Consequently dynamic conditions should not change parameters but call for equations adapted to the test conditions.

It must be taken into account that some performance characteristics, which are not constant by definition, can be can be handled as constant in practice provided that their variation is under the range of uncertainty of the parameter estimates. Discerning when these approximations are valid requires some knowledge of building physics. For example U-value is not a constant parameter. However it is used as constant in many practical applications. This is a reasonable assumption under some conditions, but could it be incorrect in some cases such as poorly insulated walls and windows.

This document describes guidelines for dynamic analysis for estimating the thermal performance of buildings and building components focusing on physical aspects. Minimum steps to carry out data analysis are reported and different alternative analysis approaches are outlined. This document is mainly focused on the most critical aspects particularly related energy performance assessment of buildings and building components. More general techniques which are also required to carry out data analysis are briefly presented in this document including references to more comprehensive information. Case study considered in chapters 3, 4, 5, 6 and 7, is used to facilitate the understanding of some of the recommendations given in this chapter. References for additional case studies that help to understand the different aspects discussed are included in the document.

This is considered as a multidisciplinary problem that requires application of knowledge in different areas such as physics and statistics. This first part is mainly dealing with physical aspects. A second part focused in statistical aspects has been also elaborated. Both documents must be considered as complementary.

11.3Physical parameters

Definition given by International standards are used in this document - particularly the following included in ISO 7345:1987 (Thermal Insulation – Physical Quantities and Definitions):

- **Thermal resistance, R:** Temperature difference divided by the density of heat flow rate in the steady state condition. Units: m^2K/W.
- **Thermal conductance, Λ:** Reciprocal of thermal resistance from surface to surface under conditions of uniform density of heat flow rate. Units: $W/(m^2K)$.
- **Thermal transmittance, U:** Heat flow rate in the steady state divided by area and by the temperature difference between the surroundings on each side of a system. Units: $W/(m^2K)$.

The following are defined by the ISO 13790:2008(E) (Energy performance of buildings - Calculation of energy use for space heating and cooling).

- **Heat transfer coefficient:** Heat flow rate divided by the temperature difference between two environments; specifically used for heat transfer coefficient by transmission or ventilation. Units: W/K.
- **Transmission heat transfer coefficient:** Heat flow rate due to thermal transmission through the fabric of a building, divided by the difference between the environment temperatures on either side of the construction. Units: W/K.
- **Ventilation heat transfer coefficient:** Heat flow rate due to air entering an enclosed space, either by infiltration or ventilation, divided by the difference between the internal air temperature and the supply air temperature. Units: W/K.

Characterization by system identification techniques requires a lumped representation of building fabric. The following parameters are usually considered in such lumped representation of a given building envelope:

If you want to list something, proceed as follows:

- UA: Overall thermal transmittance coefficient: the heat flow rate in the steady state divided by the temperature difference between the surroundings on each side of the system or component, in W/K. For the 1-D case the U-value, in W/m^{2}K.
- gA: Total solar energy transmittance or solar aperture: the heat flow rate leaving the component at the inside surface, under steady state conditions, caused by solar radiation incident at the outside surface, divided by the intensity of incident solar radiation on the component, in m^2. For the 1-D case the g-value [-].

  gA and g can show some variability, taking into account the theoretical expressions to calculate the g value of opaque components (equation (6) in chapter **¡Error! No se encuentra el origen de la referencia.**) and glassing (equation (9) in chapter 5). This variability can be assumed negligible for opaque components, in most practical applications. Considering glassing, these parameters depend also on the transmittance of the glassing that depends on the incidence angle, and consequently on the time of the year. Smaller variability than the uncertainty on the parameter estimates is expected, but it should be analysed and discussed in each particular case study. The variability of these parameters for the opaque wall and round robin box is discussed in sections 2.2.1 and 0 respectively.

Interpretation of effective heat capacities in lumped representations: As mentioned in the introduction, dynamic tests allow modelling buildings and building components from experimental test campaigns carried out under dynamic test conditions. The starting point of this analysis is considering energy balance equations that include the measured variables and characteristic parameters that must be identified. Dynamic energy balance equations of any given system must include terms representing energy accumulated in the system. This accumulation is governed by effective heat capacities that can be estimated as the other defined parameters. However these heat capacities can be considered as auxiliary parameters that regulate the energy that is accumulated in the system and depend on the characteristics of the system and its boundaries. Then these parameters are considered effective heat capacities and must not be compared to the theoretical heat capacities obtained from the characteristics of the construction materials of the building envelope.

11.4Experimental aspects related to identification

This section discusses the aspects that must be taken into account and the features that must be implemented to optimise the test regarding data analysis for identification.

<u>Analysis objectives</u>

First, experiment design must take into account the final objectives of the identification analysis. The generic objectives listed below will be considered for the following discussion:

- Characterize the heat losses of the envelope of one given zone through the heat transfer coefficient to its boundary zone. The boundary could be the outdoor or any other zone.
- Characterize the solar gains of the main zone through its overall gA-value.
- Characterise the dynamic performance of the main zone through an effective heat capacity.

<u>Main zone and boundaries</u>

The main zone and boundaries must be well defined and clear for analysis and consequently they must be taken into account for experimental design.

Requisites related to the main inputs and driving variables

The experiment design must guarantee that the phenomena that must be characterised are happening and strong enough for their analysis. Phenomena are considered strong enough in this context, when the amplitude of the corresponding driving variable is significantly higher than the uncertainty in its measurement. Otherwise signal to noise is poor.

The same criteria must be applied to the amplitude of any other variable required to complete the energy balance equation used in the analysis.

Applying these conditions to the considered generic objectives, the following conditions are necessary:

- To identify **heat transfer coefficient to the outdoors**, the experiment set up must ensure strong enough heat loss through the corresponding components. This is achieved maximising the temperature difference between the air in the main zone and the outdoor air, which is the driving variable in this case.

  This maximisation must be implemented taking into account the limitations due to the safety recommendations of the construction materials. Limitations due to the variability of characteristic parameters with temperature must be also considered. Too high indoor temperature that could lead to damage construction materials as well as variations of parameters that are not seen in actual operational conditions, must be avoided.

- Analogous considerations must be done to identify **heat transfer coefficients to the boundaries**. In this case the main driving variable, is the temperature difference between the air in the main zone and the air in the adjacent zones. This difference must be maximised taking into account the limitations mentioned in the previous paragraph.
- To identify the **overall gA-value**, solar gains must be strong enough during the experiment. This is achieved when the experiment contains sunny days, when solar radiation is high, which is the driving variable in this case.
- To identify the **effective heat capacity** the system must be excited by dynamic input signals in a wide range of frequencies covering the characteristic time constants of the system. This is achieved e.g. by applying a ROLBS power sequence. See statistical guidelines for further information of the ROLBS sequence (ref. 19). Some examples can be seen in section 4.2.5 chapter 4, section Figure 98 in this chapter, and ref. 23
- Notice that **heating in the main zone** during the experiment is necessary to maximise temperature differences between indoor air and ambient and adjacent zones. Free running tests may lead to poor signal to noise ratios in the temperature difference measurements and problems with identifiability. Additionally, the heating power is an important variable to complete the energy balance equation used in the analysis, so it must be strong enough in the experiment.

The experiment should be designed aiming to maximise the degree of accomplishment of all these conditions, for the given weather in the test site and the constraints resulting from the heating systems available in each case.

Homogeneity of the indoor air temperature

The different sources of heat such as heating devices and solar radiation can lead to some degree of inhomogeneity of the indoor air temperature contributing to the uncertainty budget of the parameter estimates.

A fan could help to achieve better homogeneity. However it must be taken into account that this strategy could introduce perturbations in the interior convection coefficients, so its usefulness is limited to cases when these perturbations could be considered as negligible.

When unavoidable air stratification is present in realistic experimental campaigns, the uncertainty in parameter estimates due to temperature stratification in indoor air must be taken into account and evaluated. Measurements giving information on air stratification are very useful. This information on air temperature distribution is useful to investigate the following issues:

- Uncertainty in parameter estimates due to temperature stratification in indoor air.
- Different options to achieve an optimum representation of indoor air temperature, for example: considering spatial averages, weighted averages, a reduced selection of measurement points, or any other.

11.5 Minimum steps for data analysis

Data analysis must consider at least the following steps:

1 Pre-processing: Any pre-processing carried out must be reported. Participants in the analysis exercises were encouraged to report data overview based on plots, discussion about quality of data and their suitability to fit objectives, etc.

2 Modelling approach: The methods and models used must be described. The hypotheses and approximation about the physics behind the considered candidate models must be justified. Schematic representations of heat flows in the building are recommended to support explanations. The process of model selection and the decisions made in this process must be explained. The software tools used to identify the parameters must be mentioned.

3 Validation: The validity of the results must be demonstrated. Statistical criteria are very useful in this process. Results must not contradict physical consistency. The process followed to demonstrate the validity of the results must be explained.

4 Results using different data must be compared. Since the data comes from the same physical system the best model should give similar results for two (or more) data series.

5 Results: A value estimated for each parameter and its corresponding uncertainty must be clearly marked as the final result.

6 A list of the hypotheses and approximation of the physics behind the model finally selected to give the final results, must be given together with the final result.

7 Conclusions: Any relevant finding resulting from the analysis, about the results, about the experiment set up and measurement campaign, etc., must be summarised.

8 Reporting: Reports must include at least a section devoted to each of points 1 to 5.

Feedback among the different points should be made in every phase of the process. Is the model accepted? It is advisable to apply more than one method to get a better understanding of the whole problem. Common sense should always be used and all the available physical and statistical knowledge should be used whenever possible.

11.6Data analysis

This section summarises the aspects that must be considered. Sub-sections 11.6.1, 11.6.2 and 11.6.4 are common to all the analysis approaches considered in this document. Section 11.6.3 focuses on specific aspects of some different approaches that can be chosen.

11.6.1 Preprocessing

Data analysis starts with a qualitative analysis of the quality of the data based on data overview. The objective of this data overview is to detect any abnormal behaviour in the tendencies of the measured variables, sensor failures, etc.

This qualitative analysis is based on the prior physical knowledge of the thermal system under study. Knowledge of the measurement principles of the sensors, transducers and data acquisition systems, can help to interpret unexpected behaviour of the recorded data.

The result of this analysis can be the rejection of some measurements, the decision of correcting some problem on the experiment set up and repetition of the experiment, or the acceptance of the data for the analysis.

Averaging and filtering

Averaging is frequently used as filtering and also as resampling technique. Other filtering and resampling techniques can be applied. Discussing filtering and resampling techniques is out of the scope of this document. However some relevant issues regarding their application are discussed in the following.

In general any averaging and filtering carried out must be justified, explaining its interpretation, which are the beneficial performances that are expected applying it, why improvements are expected, etc.

The application of filtering techniques is useful when there is certainty that their effect is removing information in the data that doesn’t correspond to the phenomena that we are studying in the building or building component. However filtering and averaging could have harmful effects if it removes relevant information to the process under study.

As an example the tests of the round robin box in Almería is considered, where tests have been optimised regarding analysis. Particular focus was put in these tests in optimising the accuracy of the measurement of heating power, and regarding amplitude and time resolution (see section 5.5.1 of this document). Amplitude of the heating power, set point of indoors air temperature and dead band have been set such that switching frequencies in the heating power are lower enough than the sampling frequency to guarantee good time resolution. Wrong resampling can make ineffective this experimental set up optimization, and lead to conclude erroneously that signal to noise ratio is poor. See Figure 98 and Figure 99.

In addition it is known that the amplitude of the heating power in these tests only can have two values depending on the status of the switch (on/off), with very small fluctuations around the “on value” (due to the stability of the power source). However averaging can bring some values very far from the actual values of the heating power, and consequently introduces an unnecessary source of uncertainty. See Figure 100.

Consequently in this case this resampling gives poorer signals than the originals so it makes no sense.

The effects of different averaging and resampling periods in other measured variables which are relevant in this case study are shown in Figure 101 (indoor air temperature), Figure 102 (vertical south global solar radiation in clear sky days), Figure 103 (vertical south global solar radiation in cloudy and clear sky days) and Figure 104 (outdoor air temperature).

Averages are suitable in steady state methods which are integral approaches where averages are used to represent integrals. The situation is radically different in differential approaches where the interpretation of time averaging as integration makes no sense.

11.6.2 Construction of candidate models based on hypotheses derived from prior physical knowledge

The starting point of the analysis is considering energy balance equations that include the measured variables and characteristic parameters that must be identified. The characteristics of the studied component and given test conditions are taken into account to build all the candidate models. This step is common to all the analysis approaches.

Then candidate models must be written trying to answer the following questions:

- What is the system to which the energy balance equation will be referred to?. Is it a volume?, is it a flat surface? What is the considered system and their boundaries?

  Some case studies where the energy balance considered for modelling is referred to a surface are reported in Refs. 34 and 35. A case study where the energy balance considered for modelling is referred to a volume is reported in Ref. 36.

- What are the phenomena that theoretically participate in the considered energy balance equation?
- Which of these phenomena are relevant in practice to the considered case study under the given test conditions?
- What is the most efficient way of modelling each phenomena considered as relevant in each case study?. Notice that here efficiency is referred to model accuracy, cost of used measurement devices, and model simplicity. Modelling should include in its objectives the maximisation of accuracy and the minimisation of costs of measurement devices as well as minimising model complexity.

  It must be taken into account that sometimes expressions that could be considered more accurate in principle, could in practice give bad performance. This behaviour could be explained since these expressions bring to the models information and also uncertainty, and in some cases the weight of the introduced uncertainty could be higher that the weight of the introduced information. Any of the following issues could explain this behaviour:

  - Sometimes these expressions are approximation themselves. In this case including effects that have low influence on the global energy balance could bring more uncertainty than information.
  - The contributions associated to each of the variables incorporated in such expressions (the more variables the more uncertainty). As a consequence, including effects that have low influence and are depending on many measured variables could lead to models with bad performance.

    Statistical tests can give very useful support to study this issue.

  - Being far from the scope of applicability of these expressions.

    Consequently it is possible that using more sophisticated expressions could lead to more inaccurate models at the end. Conversely we can find very simple expressions that however capture the main essences of the studied process leading to accurate models.

- Which are the main driving variables of each of the phenomena recognised as relevant for the considered case study?
- Which variables must be considered inputs and outputs according to causality?

If it is not possible to answer some of these questions a priori, several candidate models according to the different possibilities can be considered and evaluated.

If too many options are identified constructing candidate models in this way, it is useful to stablish some prioritised order first studying independently each of the relevant effects identified and then combining those that evidence improvement regarding models not including them. This strategy to construct candidate models was followed in the work reported in ref. 48.

Specific recommendations to construct candidate models for a case study: round robin text box.

This section further discuses criteria to construct candidate models focusing on the round robin test box considered in chapters 3, 4, 5, 6 and 7.

The first step to construct candidate models to fit this objective is to identify the system to which the energy balance equation will be referred to. In this case this energy balance will be referred to the volume of air confined by the building envelope.

To propose candidate models we need to formulate hypotheses according to the following questions 1 to 3. Candidate models must be based on approximations derived from these hypotheses.

1 Are the solar gains through the opaque walls (outside to inside) relevant regarding the other terms in the energy balance equations?.

- Candidate models considering only solar gains through the window make sense for this case study for data recorded in Belgium during winter (chapter 4), taking into account: 1. The low levels of solar radiation (mainly diffuse), 2. Due to the sun position in winter, the strongest global solar radiation is on the south vertical external surface, 3. Probably high insulation materials in the opaque walls, 4. White external surface that contribute to reduce solar gains.
- This issue can be investigated considering mono-dimensional analysis, based on energy balances of energy flows through the internal surfaces of the opaque walls. Heat flux measurement devices, on the internal face of each wall, measure the net heat flux density through the internal surface of each wall. These devices are used to carry out this analysis:
    - Studying if solar radiation influences this net heat flux through these surfaces, is useful to discern if solar radiation through the opaque walls is relevant or negligible.
    - The identification of U and g values of the walls using this measurement and the energy balance through the corresponding surface is useful to discern if g-value of the opaque walls is relevant or negligible.

  The mono-dimensional analysis of the roof in Spain (chapter 5) in summer can be studied as an unfavourable case with high levels of incident solar radiation, to evaluate if the g value is or isn’t negligible on the opaque walls. In this surface the global solar radiation is high compared to other orientations and also compared to tests in Belgium. If it is concluded that solar radiation is not relevant in this roof we could extrapolate that it is not relevant in any other opaque wall of the round robin box.

  Carrying out analogous mono-dimensional analysis to study floor, and all the others opaque walls is strongly recommended because it would be very useful to confirm and validate the results and conclusions obtained from the analysis of the roof.

- Would it be sufficient to compare plots of the heat flux measurements through the different walls, to conclude that solar gains are relevant when we observe differences in the heat fluxes and these differences are correlated to solar radiation?.

  No, this is not enough to do this affirmation. If we see that heat flux through the roof is lower than heat flux through the floor it may be: either because we have solar gains that reduce the net heat losses through the roof, or because the solar radiation that enters through the south window reaches the internal surface of the floor and increases the net heat flux through the floor. Similar considerations can be done for east and west walls. See Figure 105.

  It must be taken into account that differences smaller than the uncertainty in the measurement of the heat flux sensor cannot be interpreted as actual differences. It must be remembered that the minimum uncertainty in this measurement is 5% (Uncertainty of the heat flux sensor).

2 Is it possible to assume a constant UA value?, Has the external surface heat transfer coefficient a relevant dependence on the wind speed?, Have the surface coefficients a relevant dependence on the temperature of the corresponding surfaces?, Have the thermophysical properties materials a relevant dependence on temperatures?.

  Mono-dimensional analyses are useful and recommended also to study these issues.

3 Is air leakage relevant to the energy balance equation when the whole volume is considered to calculate the UA value?. If it is relevant we need to model this term. It makes sense to consider this contribution negligible in this case study, but there are alternatives that can be considered in case of being not negligible as justified below.

- Preliminary tests for air-leakage were done and reported by BBRI (see Chapter 4 in this document). These tests demonstrated that air-leakage is negligible.
- Even if air leakage is not negligible, the corresponding contribution to the energy balance equation could be considered negligible if the outdoor air flow is already heated up to the indoor air temperature level before being delivered to the room, then there is no heat loss involved. (Ref. 38, page 45).
- If the air enters the test room via a shortcut, then a 100% correction for the effect of air flow on heat balance is required. In practice this phenomena can have an effect on the HLC between zero (fully pre-heated) and maximum (no preheating at all). (Ref. 38, page 45).
- When air leakage between indoors and outdoors is not negligible, a coefficient depending on the pressure difference between both environments could be included. In practice pressure difference is dependent on temperature difference and also on wind speed and direction. Several options for this effect considering first order approximations of these two variables are suggested by Ref. 38 (section 5.1.).
- Improvements could be investigated considering other alternatives reported for this effect in the literature (ref. 40).

11.6.3 Modelling

This section briefly introduce some of the most usual modelling approaches and make emphasis on the physical aspects in each particular case. The main focus on these sections is on the key steps that require application of physical criteria, which are highlighted here. These aspects are crucial to obtain accurate results.

Relevant statistical aspects of these modelling approaches are described in detail in the statistical guidelines (Ref. 19).

Steady state approaches

In steady state all the physical quantities are time independent, according to the definition of steady state given by ISO 9251: 1987. Consequently steady state equations applied to raw data are not valid in dynamic test campaigns.

However integrated dynamic equations become analogous to steady state equations, using time averaging to represent integrals when the integration period is long enough for the accumulation terms to become negligible compared to the other energy flows in the equation. In this case steady state equations can be applied to dynamic test campaigns.

The applicability of these methods to dynamic data has significant constraints as discussed by Ref. 42. This reference reports on the study of the errors in the U-value estimates for different walls, using the steady-state equation. This study is done for instantaneous measurements as well as considering averages representing integrals. It concludes that instantaneous measurements can't provide accurate estimates of the U-values but its accuracy is significantly improved if time integrated variables are used. It is also shown that the error in the U-value estimation is minimised by using a multiple of 24 hours as the integration period. It affirms that the minimum valid integration period depends on the characteristics of the wall and weather conditions. It also demonstrates that longer integration periods are required when temperature fluctuations are higher, temperature difference between indoors and outdoors are lower and walls are heavier.

Average method and (multi-) linear regression methods based on averages belong to this family of models, which base their validity in using averages having their origin as integrals of dynamic formulations.

These models give very bad results when:

- Premises for applicability are not met. More details about these requisites for applicability for average methods are reported in ISO 9869 (Ref. 43).
- Inappropriate energy balance equations are applied in (multi-) linear regression methods. It must be remembered that criteria given in section 3.2 are valid also in this approach to construct candidate models.

It must be remarked that when a product or other operation with different variables is considered in a model, all the products and other operations must be done before the average. This is easily understood keeping in mind that time integral of dynamic energy balance equations are behind time averaging in the origin of these approaches.

These approaches applied to dynamic test campaigns are not efficient, but they can be useful in certain applications. Their main drawback is that sometimes extremely long test periods are required.

These methods were considered at the beginning of the PASSYS project (ref. 44), but were superseded by dynamic approaches that are more accurate, efficient and suitable for dynamic experimental conditions (Refs. 45 to 47).

This analysis is recommended as a first exploratory approach to each new problem. Such preliminary analysis is useful for indicating the order of magnitude of thermal parameters, to investigate which are the most relevant effects in the energy balance equations, what approximations are more suitable for each relevant effect, etc. (Refs. 7 and 48).

Linear models in transfer function form

These models are described in Refs. 49 and 50. Ref. 47 presents the application of this type of models to estimate the thermal properties of building components from outdoor dynamic testing, imposing appropriate physical constraints.

The following is focused on ARX models as an example without losing generality. In these models the output, y(t), is expressed as linear function, using constant coefficients, a_i and b_i, of a number, s, of past readings from the inputs, u(t), and also from a number, r, of past readings of output itself as follows:

$$y(t)+a_1 y(t-\Delta t)+\ldots+a_r y(t-r\Delta t)=b_0 u(t-b\Delta t)+\ldots+b_s u(t-(b+s)\Delta t)+e(t) \qquad (25)$$

Where Δt is the sampling interval, e(t) represents the model error.

Considering a dynamic ARX model including the same variables as the steady state energy balance equation of the given system, the required physical parameters are found from the coefficients of the dynamic ARX model, by imposing as physical constraint that the following equations must be coincident:

1. The steady-state energy balance equation of the considered system.
2. The ARX model, when all its inputs and outputs are constant.

This implies that the coefficient of each variable must be the same in both equations. In the case of steady steady-state energy balance equation these coefficients are the target physical parameters. In the case of the ARX model, with constant inputs and outputs, the coefficients become algebraic combination of the a_i and b_i, parameters obtained for the dynamic ARX model.

Then the steady-state physical parameters can be found by comparing these two equations (ref. 47). This comparison is possible provided that the ARX model contains the same variables as the steady state energy balance equation of the considered system. Consequently the first step in this analysis approach is to deduce and write the appropriate steady state energy balance equation that must be based on previous physical knowledge.

For example, if the steady-state energy balance equation for a given system is $\Phi=U(T_i-T_e)-gG_v$, then the dynamic ARX model used to obtain the U and g values must contain the following variables: Φ, T_i, T_e and G_v.

Different candidate models can be taken into account as discussed in section 11.6.2, taking into account different approximations to write the steady-state energy balance equation.

Once the variables that must be included in the model have been identified, it is necessary to decide which of them are considered inputs and which of them are considered as outputs. This assignment must be based on causality of these variables. Usually, either indoor temperature or heat flux is selected as the model output. Multi-output models can be also considered and have shown very good performance in some case studies where test were carried out under unfavourable test conditions (Ref. 51).

Notice that although this is a linear approach, some non-linear effects can be considered by a change of variable. An example is reported in ref. 58. In that work, long wave effects depending on surface temperature raised to the fourth power T^4, were included in the models through a new variable $y=T^4$.

<u>Models in continuous time state space form based on SDE</u>

The continuous–discrete stochastic state space model is a model that consists of a set of nonlinear discrete, partially observed stochastic differential equations (SDEs) with measurement noise, i.e.

$$dx_t = f(x_t, u_t, t, \theta)dt + \sigma(u_t, t, \theta)d\omega_t \qquad (26)$$

$$y_k = h(x_k, u_k, t_k, \theta) + e_k \qquad (27)$$

where $\theta \in \Theta\ R^p$ is parameter vector; $f(\cdot)\in R^n$, $\sigma \in \Theta\ R^{nxn}$ and $h(\cdot)\in R^l$ are nonlinear functions; $\{\omega_t\}$ is an n-dimensional standard Wiener process and $\{e_k\}$ is an l-dimensional white noise process with $e_k \in N(0, S(u_k, t_k, \theta))$.

This class of models is further described in Ref. 52 and treated by the tool called Continuous Time Stochastic Modelling (CTSM), Ref. 53.

The state space representation is very useful and flexible to represent physical systems governed by differential equations, which offers a very high potential to model a wide variety of physical systems.

Diffusion terms and modelling errors ($\sigma(u_t, t, \theta)d\omega_t$ and e_k in the previous equations) allow achieving very accurate parameter estimates.

It must be highlighted that the system equations can include measured as well as non measured states which is a very useful feature in modelling physical systems.

RC models can be considered here. However these family of models are a reduced subset of the state space models that can be used and don't make use of these capabilities to their full extent.

Case studies applying this approach are reported in Refs. 35, 54 and 55. Ref. 54 reports the application of RC models to a full size building. Ref. 35 reports the modeling of a building integrated and ventilated photovoltaic system. It demonstrated that a description of the nonlinear heat transfer is essential for modelling of surface effects due to long wave radiation and convection in this particular system. It showed also the usefulness of certain approximations to take into account unmeasured effects. Ref. 55 reports the application of the same model to evaluate the effect of different set ups in the heat transfer coefficient for the same PV integrated modules.

11.6.4 Model validation

Validation is a key issue of the problem. Statistical criteria are very useful guiding and optimising the process of model selection. Results must not contradict physical consistency.

The following criteria for model selection and validation suggested by Norlén (ref. 56) are recommended:

1. Fit to the data. The model residuals should be 'small' and 'white noise'. A necessary condition for 'white noise' is that there should be neither autocorrelation in the residuals nor correlation between the residuals and the input variables.
2. Internal validity. The model should agree with data other than that used for parameter estimation (cross validation).
3. External validity. The model results should not (without strong reason) conflict with previous experience or other known conditions.
4. Dynamic stability. In steady state, the model should provide an output after a temporary change in an input variable that gradually fades out (if the model is intended to describe dynamic characteristics).
5. Identifiability. It should be possible to determine the parameters of the model uniquely from the data.
6. Simplicity. The model should be as small as possible.

If the residuals contain periodicities, a study of the correlation in the frequency domain can be more useful to reveal such periodicities than the autocorrelation function. The periodogram shows how the variation of the residuals is distributed on frequencies. For white noise this variation is equally distributed, i.e. the cumulative periodogram is a straight line from (0,0) to (0.5,1). See Ref. 57 for further details.

Particular care must be taken in the interpretation of residuals with a frequency of 24 hours. In a first approach a non-negligible correlation between the model residuals and solar radiation could lead to the suggestion for a more detailed description of the solar radiation in the model for further improvements. However it must not be forgotten that, many variables can show a relevant correlation with solar radiation, so any other effect not properly modelled can show problems in the residuals in the same frequency as the solar radiation. This behaviour is highlighted in sunny locations where levels of solar radiation are high. The effects in the following list can be mentioned as examples of phenomena that could produce this behaviour.

If you want to list something, proceed as follows:

- Air leakage that can depend on wind speed and/or outdoor air temperature, both depending on solar radiation.
- Long wave effects resulting from high surface temperatures due to solar radiation. One example of this issue is reported in 58.
- U depending on thermal conductivities which depend themselves on the temperature of the materials that finally depend on solar radiation.
- Wrong resampling disregarding the sampling theorem (See section 0 and Figure 99).
- Etc.

If validation criteria don't fit, models must be rejected and new models must be reformulated. Conclusions from the previous analysis and validation process are very useful to construct new models, reconsidering hypothesis and rewriting candidate models according reviewed hypothesis. Redesigning experiment set up and carrying out new experiments might be necessary.

Practical recommendations

This section reports some practical recommendations reported in 38. These recommendations are based on experience from different European projects related to energy assessment of building components, such as IQ-TEST, etc. (ref. 39).

A. Preliminary check on residuals:

Plot of average and root mean square of the residuals are very useful in a preliminary analysis to detect problems related to data and modelling. In particular the following checks are recommended:

- If the root mean square (RMS) value of residuals is 'unexpectedly' large: check data file or model for errors.
- The mean value of residuals should be significantly smaller than the RMS value; if not: systematically biased output and check data file or model for errors.
- If at certain point the residual is (or starts to be) higher than roughly three times the RMS value: check for erroneous input or output data at that point.

B. Evaluation on the basis of confidence intervals:

If a parameter has a very high large confidence interval: Look at parameter significance and correlation between parameters then adapt the model to remove insignificant and highly correlated parameters:

- If not highly correlated with one or more other parameters, freeze the parameter at some value indicated by the identification result or at a value which is physically meaningful.

- If highly correlated with one or more other parameters, then freeze one of this group of parameters: choose the one with the highest deviation from what seems physically feasible.

Freezing parameters and establishing constraints among for different parameters in the model according to physical knowledge is usually helpful to solve these problems.

C. Evaluation of the characteristic parameters and individual parameters with physical meaning:

It must be taken into account that some individual parameters have only mathematical and no physical meaning; these cannot be checked for their physical meaning; for instance individual conductances and capacitances inside a construction without internal measurements.

The following checks are useful regarding parameters with physical meaning:

- Check if values deviate from values expected based on prior knowledge......
- Check if the confidence interval of a function of parameters is unexpectedly high; if so the test sequence is not appropriate for this function or the model is not appropriate (non-linearities?,...etc.).

D. Residual statistical analysis:

Autocorrelated residuals indicate that the model is not capable of following the dynamics of the system. This may be due to any of the following issues:

If you want to list something, proceed as follows:

- A missing input or error in the input data. This can be solved trying to find the missing input variable
- A higher order influence of one or more already known input variables. In this case a higher order model may lead to improved results.

Combining parameter values such as string of identical resistances and capacitances can be an useful strategy to avoid that the model becomes overparametrised.

A statistical post processing program may help to detect which information is still contained in the residuals.

E. Cross-validation:

Compare results with results on different data set. A convenient way is to repeat the identification on the data set with twice the original length and check for consistency. Since the data comes from the same physical system the best model should give similar results for both (or more) data series.

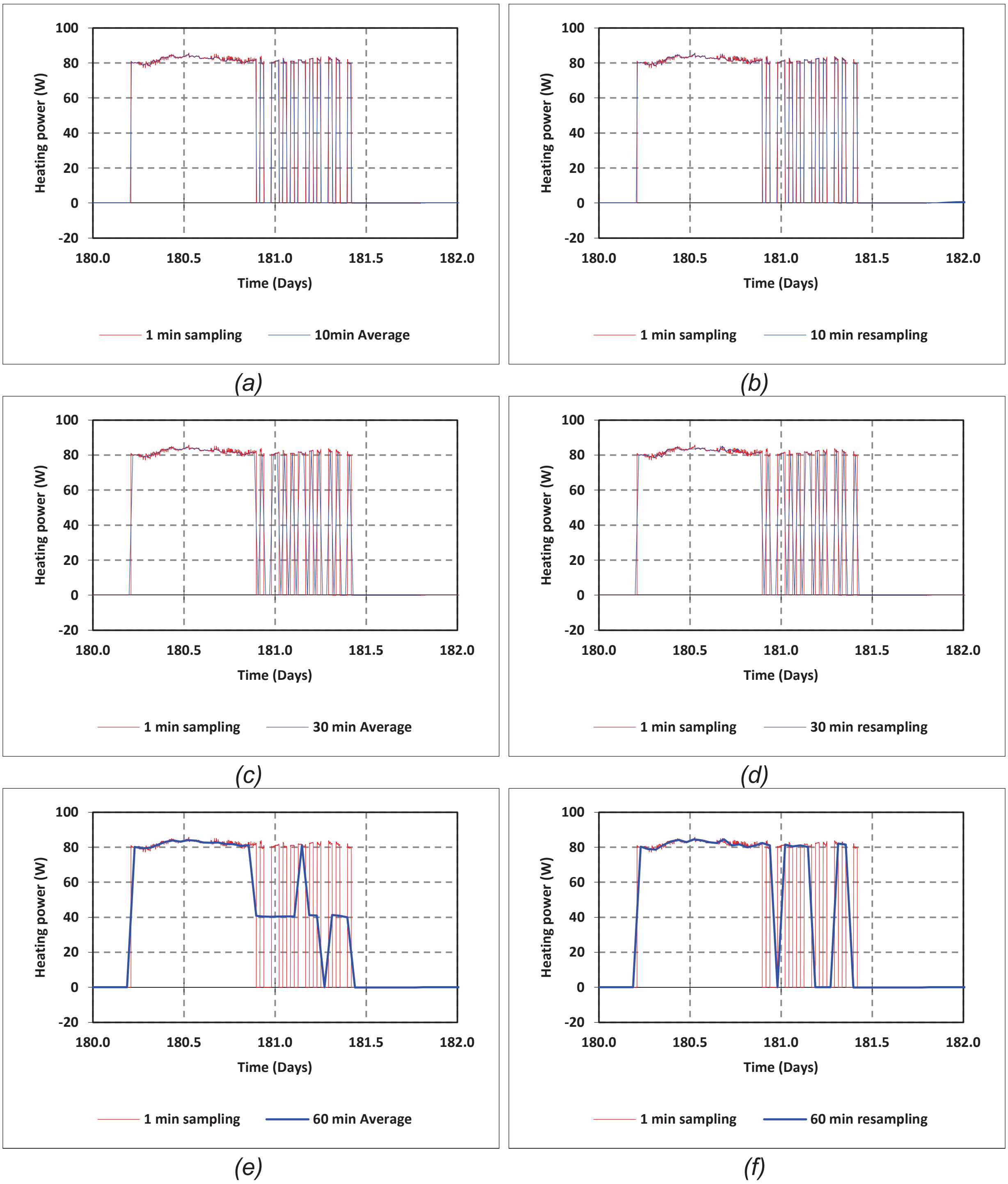


Figure 98: Effects of different resampling periods and techniques on power measurement in a ROLBS sequence. (Series 5 CE4. Test in Almería). Different behaviour is observed for the different averaging and resampling intervals. Faithful representation of the actual variable for 10 minutes resampling and averaging is seen in graphs (a) and (b). However relevant information on the measured variable is lost in some periods when longer intervals are considered for resampling and averaging. This is particularly evident in graphs (e) and (f), with 60 minutes averaging and resampling, around the time interval 181.0 to 181.5.

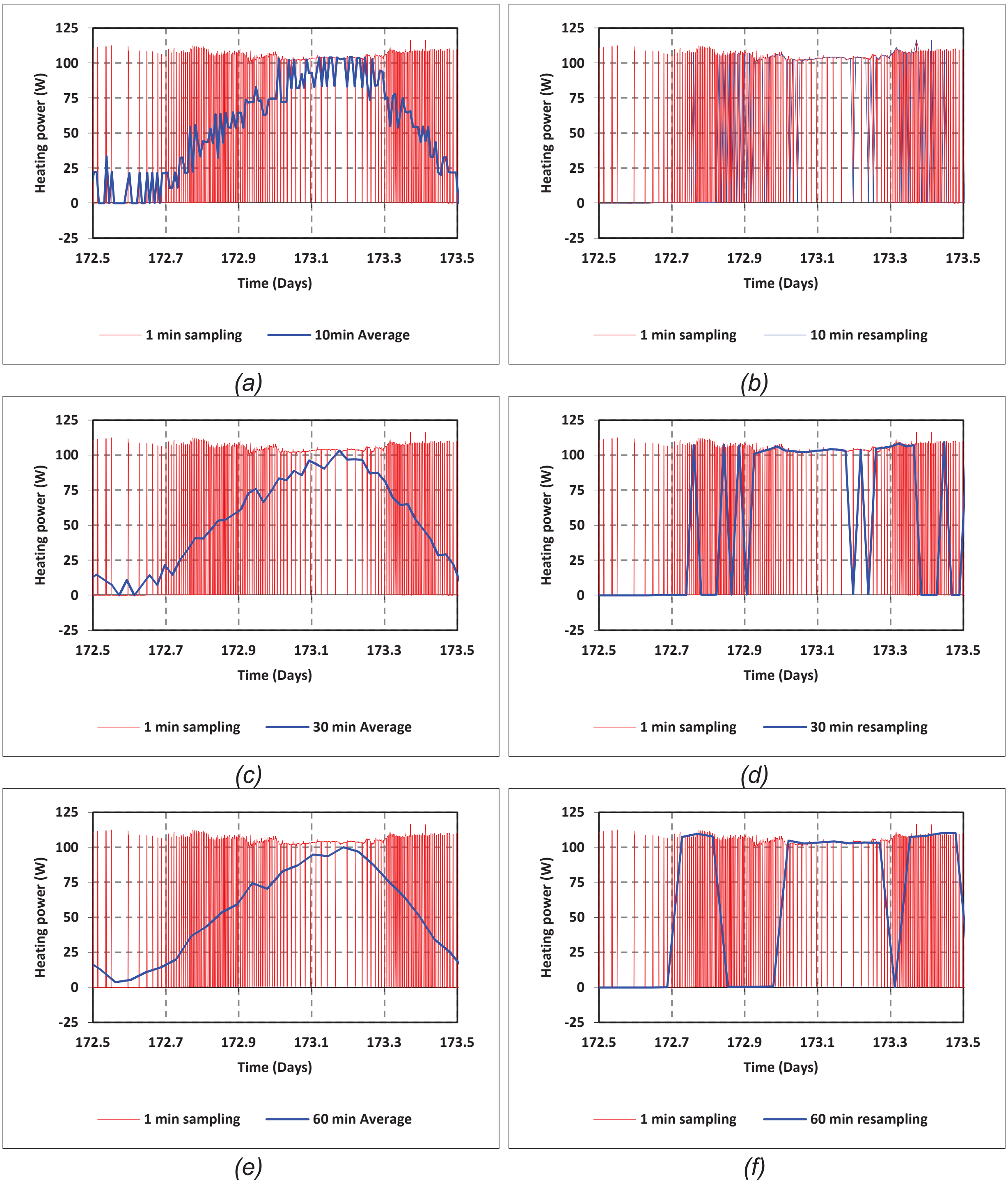
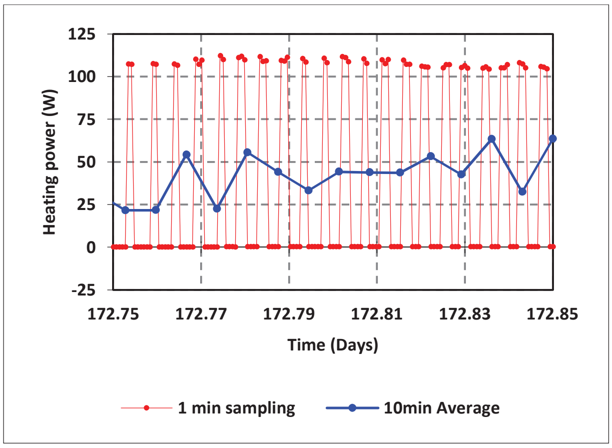
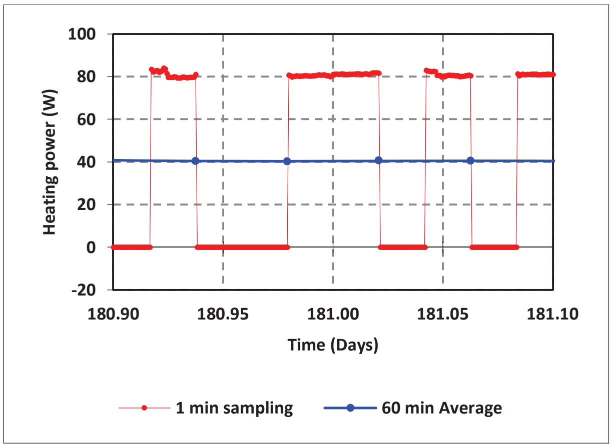


Figure 99: Effects of different resampling periods and techniques on measurement of heating power. (Series 4 CE4. Test in Almería). Relevant information on the measured variable is lost when resampling and averaging are applied in all cases.

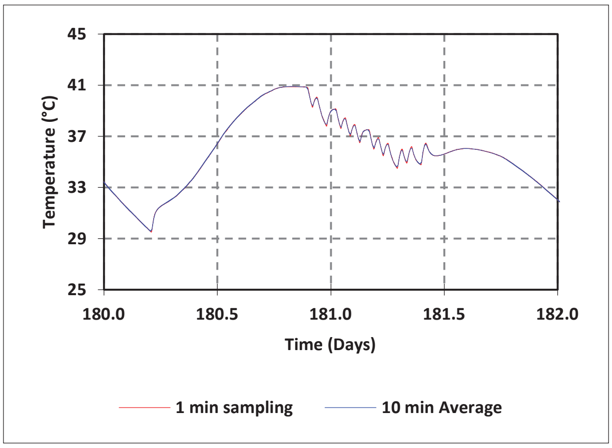




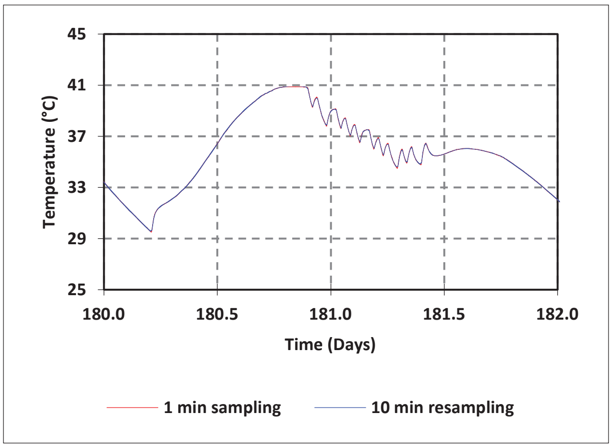
(a). Power measurement in a Co-heating test. (Series 4 CE4. Test in Almería).

(b) Power measurement in a ROLBS sequence. (Series 5 CE4. Test in Almería).

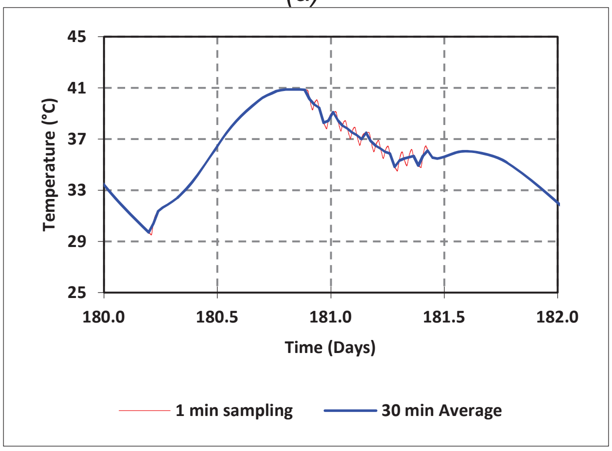
Figure 100: Some examples of resampling eliminating relevant information.



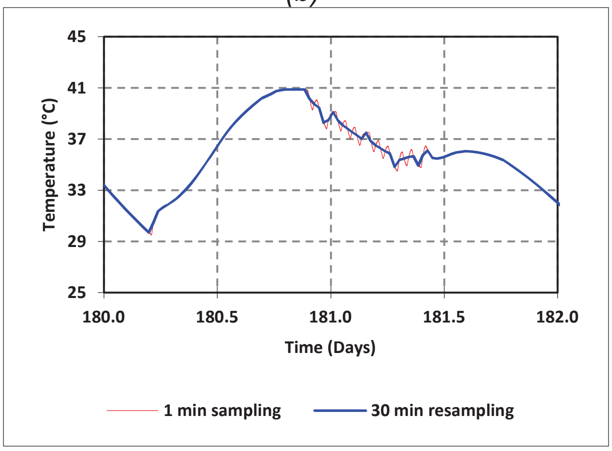
(a)



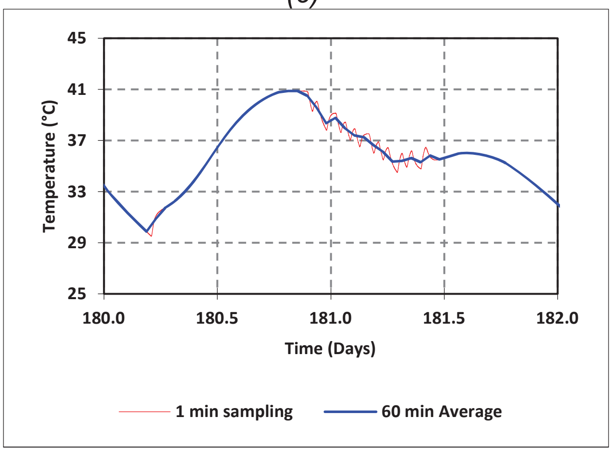
(b)



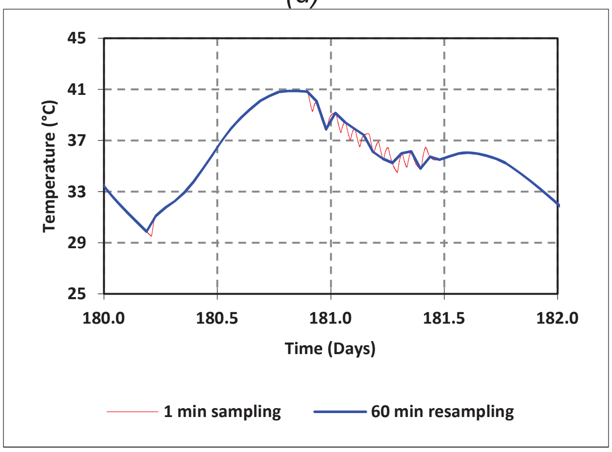
(c)



(d)



(e)



(f)

Figure 101: Effects of different resampling periods and techniques on measurement of indoor air temperature. (Series 5 CE4. Test in Almería). Faithful representation of the actual variable for 10 minutes resampling and averaging is seen in graphs (a) and (b). However relevant information on the measured variable is lost in some periods when longer intervals are considered for resampling and averaging. This is particularly evident in graphs (e) and (f), with 60 minutes averaging and resampling, for the time interval 181.0 to 181.5.

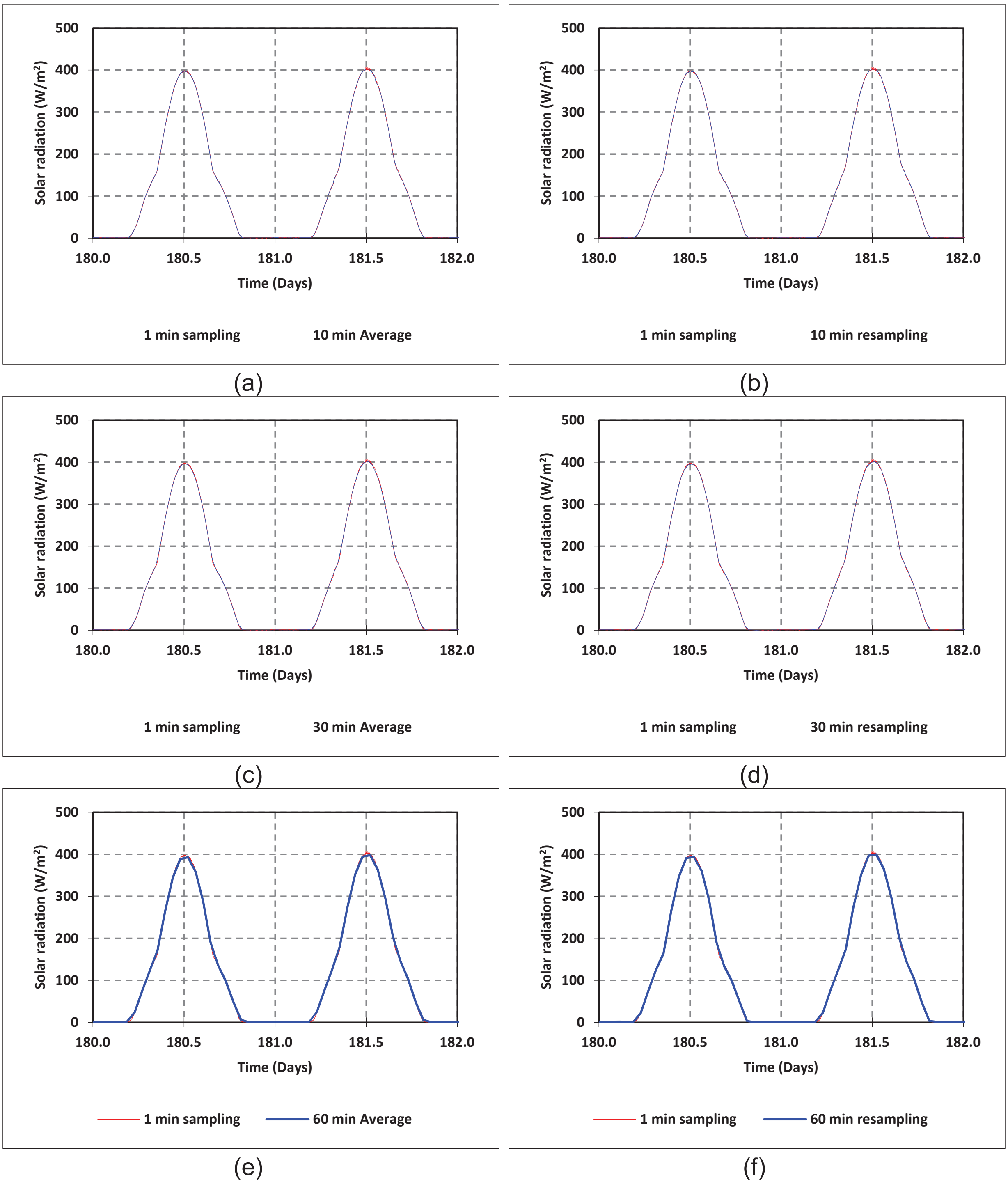
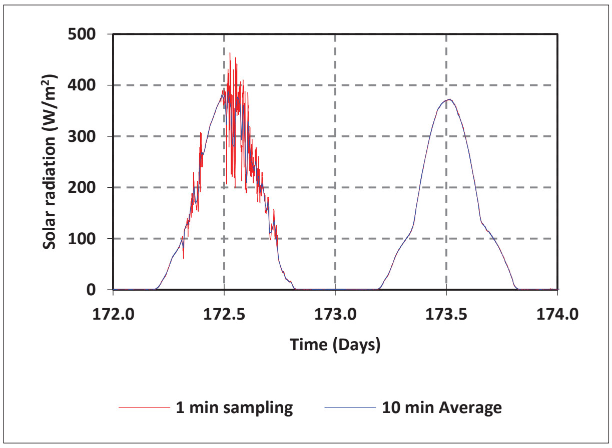
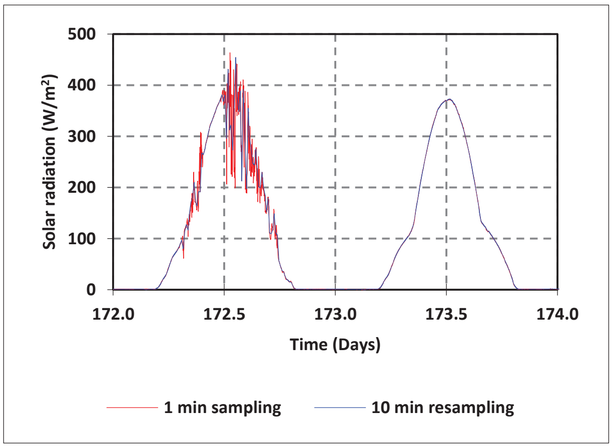


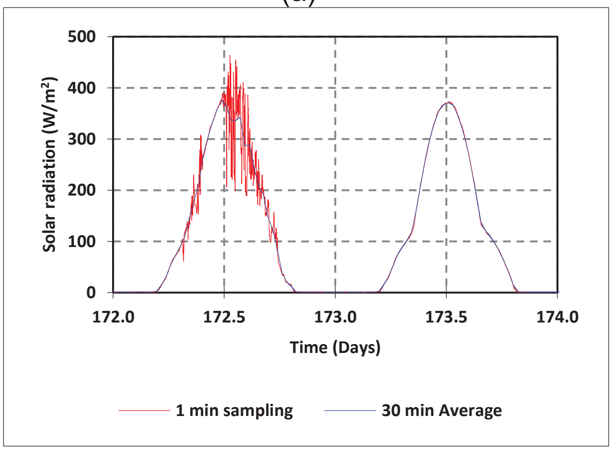
Figure 102: Effects of different resampling periods and techniques on measurement of vertical south global solar radiation. (Series 5 CE4. Test in Almería). The different averaging and resampling frequencies give faithful representation of the actual variable on sunny days. Different performance is observed for cloudy days (see Figure 103).



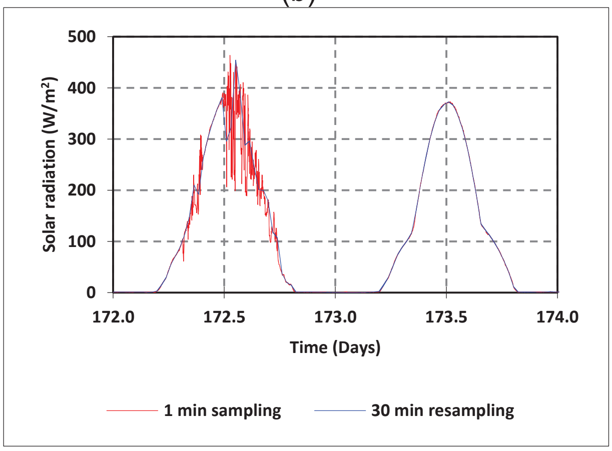
(a)



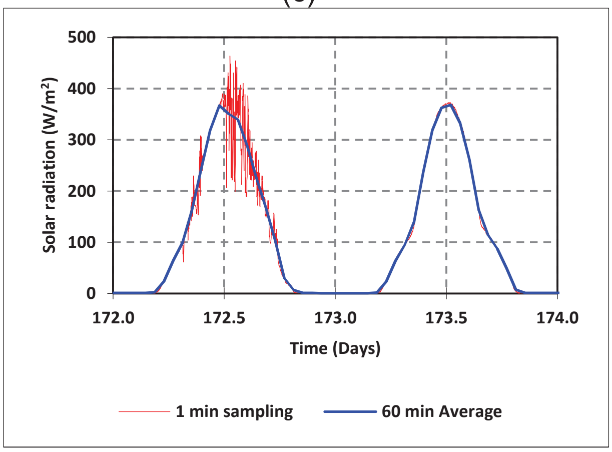
(b)



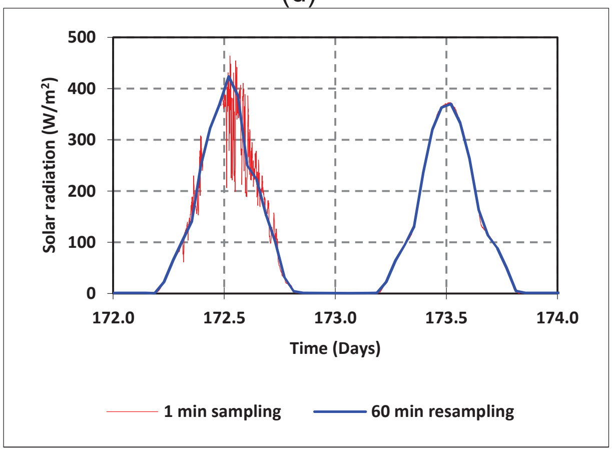
(c)



(d)



(e)



(f)

Figure 103: Effects of different resampling periods and techniques on measurement of vertical south global solar radiation for sunny and partly cloudy days. (Series 5 CE4. Test in Almería). Different behaviour is observed for clear sky and partly cloudy days. In the case of clear sky, the different averaging and resampling frequencies give faithful representation of the actual variable. However when days are partly cloudy the original measurement present relevant variations that are lost when resampling and averaging are applied.

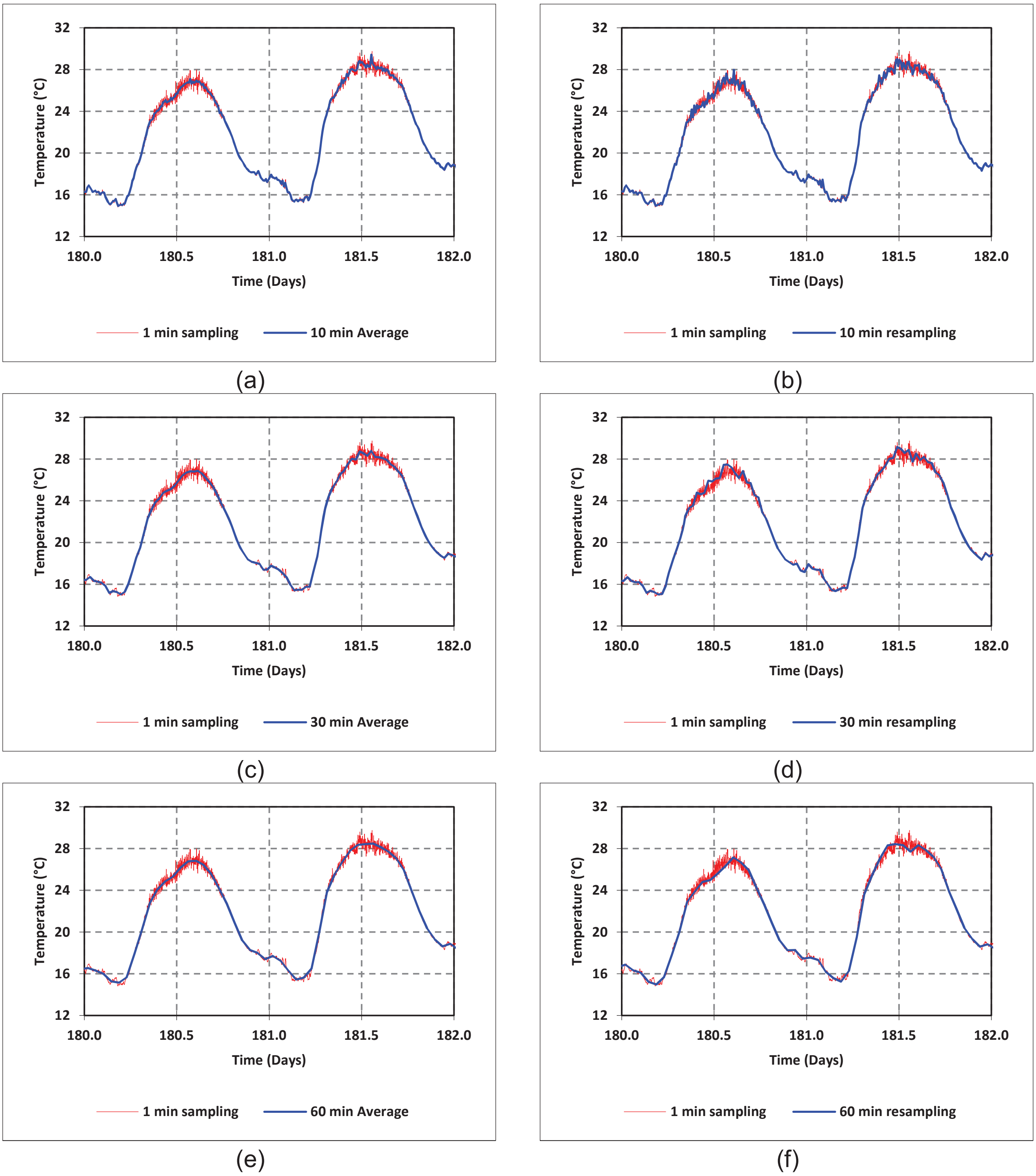


Figure 104: Effects of different resampling periods and techniques on measurement of outdoor air temperature. (Series 5 CE4. Test in Almería). Some information is lost for the different frequencies of averaging and resampling applied. Equivalent behaviour is observed for the different averaging frequencies, which eliminate small oscillations in high frequencies mainly around midday. In principle this filtering is not considered harmful because these oscillations seem to follow a white noise pattern in the range of the uncertainty in the measurement of this variable.

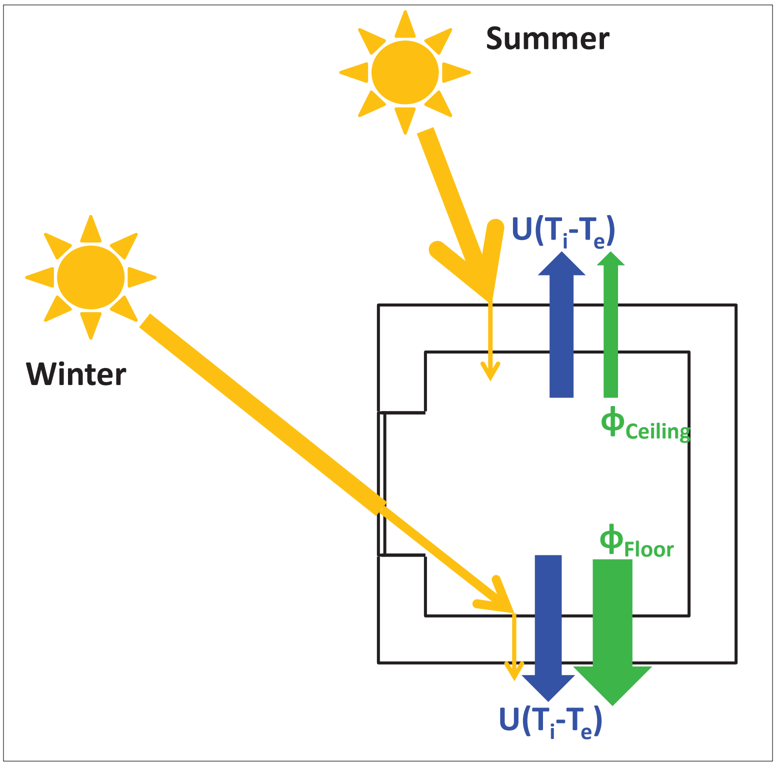


Figure 105: Different possibilities of heat flows due to solar radiation: 1.- Negligible. Then the net heat flow (green) is due to heat lost (blue). 2.- Non negligible from outdoor to indoor. Then the corresponding contribution is subtracted from the heat lost to give the net heat flow (as shown in this example through ceiling). 3.- Non negligible from indoor to outdoor. Then the corresponding contribution is added to the heat lost to give the net heat flow (as shown in this example through the floor).