Data was collected from four residential houses, situated in Gambelas, Faro, in the south of Portugal. All four are detached houses, with two floors and garden, where families live. Two of the houses have triphasic meters, while the others are monophasic. The former will be denoted as TH1 and TH2, while the latter are coined MH1 and MH2. TH1 has a PV system and a energy storage, MH1 has a PV system, and the others do not have any renewable energy source.

TH1 was used in NILMforIHEM project, that started in 2019. For this reason, and because it was used for objectives a) to e) above, has much more data for a much larger period of time. This house and the three additional houses were employed for project HEMS2IEA, which started in 2021. Only electric consumption data was recorded for these three houses. Recorded data for the four houses spans from November 2021 until July 2022. After this date, as one of the houses had major works, data was reduced to three houses.

MH1 has a PV installation, composed of 20 Sharp NU-AK panels, arranged in two strings, each panel with a maximum power of 300W. The inverter is a Kostal Plenticore Plus converter (KI), which also controls a BYD Battery Box HV H11.5 (with a storage capacity of 11.5 kWh). Consumption data is measured by a Carlo Gavazzi (EM340) energy meter, and variables related with the Inverter/battery are measured by a Kostal smart meter. The house electric panel is a Schneider panel consisting of 16 monophasic circuit breakers, plus a triphasic one. The house also has available four TP-Link HS100 Wi-Fi Smart Plugs (SP), one Intelligent Weather.

In the three additional houses, only electric consumption is measured. For this reason, in TH2, a Carlo Gavazzi EM340 meter was installed. In MH1 and MH2, Carlo Gavazzi EM112 (one-phase) meters were installed, providing a subset of variables acquired by the EM340.

The datasets are divided in months, starting in January 2020, and ending in February 2023, spanning therefore more than three years. They are Matlab data files, with the format 'v7', which can be loaded using the usual '*load*' Matlab command. Notice that the use of this format enables the data to be read directly by other languages, such as python, using the function *loadmat* in *scipy.io*.

The sensing devices are categorized in 8 categories, and within each category, there might be different appliances.

Table 1. Categories and Devices classifications

Categories	**Device numbers**	**Comments**
1 - Wibees	1-19	See Tables 2 and 3.
2 – EM340	20,31	See Table 10
3 - Inverter	21	See Tables 4-8
4 – Smart Plugs	22-24,34	See Table 11
5 – Weather Station	25	See Table 12
6 - SPWS	26 - 29	See Tables 13-16
7 – Air Conditioner	30	See Table 17
8 – EM 112	32-33	See Table 9

The variables measured by the Wibeees are shown in Table 2.

Table 2. Variables measured by Wibeees. All variables are matrices (except *ndtvec,* which is a vector)

Variable	**Variable Name**	**Units**
Time basis	dtvec	datetime
Number of Samples	ndtvec	
Voltage	Vvec	Volts
Current	Ivec	Ampere
Frequency	Fvec	Hertz
Active Power	APvec	kW
Reactive Power	RPvec	kVAR
Apparent Power	ApPvec	kVA
Power Factor	PFvec	
Active Energy	AEvec	kWh
Inductive Reactive Energy	IREvec	kVARh
Capacitive Reactive Energy	CREvec	kVAh

There are 16 monophasic WBs and 1 triphasic. The monophasic WBs range from 1 to 15, and 19. The triphasic one ranges from 16 to 18, corresponding to each one of three phases. Data iss ampled at 1 second. The most important electric appliances in TH1 are shown in Table 3.

Table 3. Wibees and major appliances

Wibeee	**Appliances**
1	Alarm
2	Swimming-pool Pump and lights, Garden Appliances
3	Illumination 1st floor
4	Illumination and Plugs Ground Floor (Hall, Garage, Bedroom, Bathroom)
5	Air conditioner Office 1st floor
6	Garage and outside gates
7	Air conditioners B12 and B13
8	Pantry plugs (Thermo-Accumulator, Washing and Drying machines)
9	Kitchen plugs (Dish Washing Machine, Microwave, etc)
10	Air conditioner A14
11	Kitchen plugs (2 fridges, Coffee Machine, etc
12	Living Room plugs (TV, Sound System, etc)
13	Plugs 1st Floor (Computing Equipment, towel heater, ceramic heater)
14	Air conditioner Living Room
15	Illumination Ground Floor (Living room, Dining Room, Kitchen)
16	Burner Stove 1
17	Burner Stove 2
18	Oven
19	Weather Station, Data Acquisition System

The data acquisition of the wibeees is asynchronous. This means that there is a time basis for each device. The different time basis are stored in the matrix *dtvec*. The number of samples for each device is stored in the vector *ndt*.

There are several variables associated with the inverter/battery. These variables are sampled at a 1 minute rate. They are detailed in Tables 4-8.

Table 4. Variables related with the inverter/battery in TH1. All variables are vectors (except *ndtINVvec*, which is a scalar). This table represents grid connection values

Variable	**Variable Name**	**Units**
Time Basis	dtINVvec	datetime
Number of Samples	ndtINVvec	
Active Power (L1)	INVPMAP_L1vec	kW
Active Power (L2)	INVPMAP_L2vec	kW
Active Power (L3)	INVPMAP_L3vec	kW
Total Active Power	INVPMAP_sysvec	kW
Reactive Power (L1)	INVPMRP_L1vec	kVAR
Reactive Power (L2)	INVPMRP_L2vec	kVAR
Reactive Power (L3)	INVPMRP_L3vec	kVAR
Total Reactive Power	INVPMRP_sysvec	kVAR
Apparent Power (L1)	INVPMApP_L1vec	kVA
Apparent Power (L2)	INVPMApP_L2vec	kVA
Apparent Power (L3)	INVPMApP_L3vec	kVA
Total Apparent Power	INVPMApP_sysvec	kVA
Current (L1)	INVPMI_L1vec	Ampere
Current (L2)	INVPMI_L2vec	Ampere
Current (L3)	INVPMI_L3vec	Ampere
Voltage (L1)	INVPMV_L1vec	Volt
Voltage (L2)	INVPMV_L2vec	Volt
Voltage (L3)	INVPMV_L3vec	Volt

Table 5. Variables related with the inverter/battery in TH1. This table represents DC values

Variable	**Variable Name**	**Units**
DC Power (string 1)	INVDCP_L1vec	kW
DC Power (string 2)	INVDCP_L2vec	kW
DC Power (string 3)	INVDCP_L3vec	kW
DC Power (All)	INVDCP_sysvec	kW
DC Current (string 1)	INVDCI_L1vec	Ampere
DC Current (string 2)	INVDCI_L2vec	Ampere
DC Current (string 3)	INVDCI_L3vec	Ampere
DC Voltage (string 1)	INVDCV_L1vec	Volt
DC Voltage (string 2)	INVDCV_L2vec	Volt
DC Voltage (string 3)	INVDCV_L3vec	Volt

Table 6. Variables related with the inverter/battery in TH1.This table represents Battery values

Variable	**Variable Name**	**Units**
State of Charge	INVBASCvec	%
Charge Current	INVBCCvec	Ampere
Discharge Current	INVBDCvec	Ampere
Charge (-)/Discharge(+) Power	INVBCDPvec	kW
Gross Capacity	INVBGCvec	kW
Number of cycles	INVBNCvec	
State of Charge	INVBSCvec	%
Temperature	INVBTvec	ºC
Voltage	INVBVvec	Volt

Table 7. Variables related with the inverter/battery in TH1. This table presents total values.

Variable	Variable Name	Units
Total DC power (sum of all PV inputs)	INVTotDCpowervec	W
Total AC Charge (AC-side to battery)	INVTotACchargevec	Wh
Total AC discharge energy (battery to grid)	INVTotACdischargevec	Wh
Total DC energy from PV1	INVTotDCPV1vec	Wh
Total DC energy from PV2	INVTotDCPV2vec	Wh
Total DC energy from PV3	INVTotDCPV3vec	Wh
Total DC PV energy (sum of all PV inputs)	INVTotDCPVvec	Wh
Total energy (AC-side to grid)	INVTotACenergyvec	Wh
Total DC charge energy (DC-side to battery)	INVTotDCchargevec	Wh
Total DC discharge energy (DC-side from battery)	INVTotDCdischargevec	Wh

Table 8. Variables related with the inverter/battery in TH1. This table shows home and other variables

Variable	Variable Name	Units
Active_Power (L1)	INVAP_L1vec	kW
Active_Power (L2)	INVAP_L2vec	kW
Active_Power (L3)	INVAP_L3vec	kW
Total Active_Power	INVAP_sysvec	kW
Current (L1)	INVI_L1vec	Ampére
Current (L2)	INVI_L2vec	Ampére
Current (L3)	INVI_L3vec	Ampére
Voltage (L1)	INVV_L1vec	Volt
Voltage (L2)	INVV_L2vec	Volt
Voltage (L3)	INVV_L3vec	Volt
Total Apparent Power	INVApP_sysvec	kVA
Total Reactive Power	INVApP_sysvec	kVAR
Total Home Consumption Rate	INVCRvec	kW
Power Consumption (from Battery)	INVPCBvec	kW
Power Consumption (from Grid)	INVPCGvec	kW
Power Consumption (from PV)	INVPCPVvec	kW
Energy Consumption (from Battery)	INVECBvec	kWh
Energy Consumption (from Grid)	INVECGvec	kWh
Energy Consumption (from PV)	INVECPVvec	kWh
Total Energy Consumption	INVECPVvec	kWh
Invertor Generation Energy	INVGEvec	kWh
Invertor Generation Power	INVGPvec	kW
Inverter State	INVISvec	
Manager State	INVMSvec	
Power Factor	INVPFvec	%
Power Limit	INVPFvec	
Work Time	INVWTvec	
Daily Yeld	INVYDvec	
Monthly Yeld	INVYMvec	
Yearly Yeld	INVYYvec	
Total Yeld	INVYTvec	

There are several variables associated with the EM112 and EM340 meters. These variables are sampled at a 1 second rate. They are detailed in Table 9 for the monophasic meters, and in Table 10, for the triphasic ones. The variables might be vectors (if only one house is measured in the corresponding period) or matrices (if there are measurements available for the two houses).

Table 9. Variables related with home consumption, from EM112 meters.

Variable	**Variable Name**	**Units**
Time Basis	dtEM1vec	datetime
Number of Samples	ndtEM1vec	
Voltage	EMVvec	Volt
Current	EMIvec	Ampere
Active Power	EMAPvec	kW
Reactive Power	EMRPvec	kVAR
Apparent Power	EMApPvec	kVA
Power Factor	EMPFvec	%
Frequency	EMF1Pvec	Hertz
Total Active Energy	EMAEP1Pvec	kWh
Partial Active Energy	EMRE1Pvec	kWh
Total Reactive Energy	EMRE1Pvec	kVARh
Partial Reactive Energy	EMREP1Pvec	kVARh
Power Demand	EMDP1Pvec	kW
Peak Power Demand	EMDPP1Pvec	kW

Table 10. Variables related with home consumption, from EM340 meters.

Variable	**Variable Name**	**Units**
Time Basis	dtEMvec	datetime
Number of Samples	ndtEMvec	
Voltage (L1-L2)	EMVL1_L2vec	Volt
Voltage (L1-N)	EMVL1_Nvec	Volt
Voltage (L3-L1)	EMVL3_L1vec	Volt
Voltage (L3-N)	EMVL3_Nvec	Volt
Voltage (L2-L3)	EMVL2_L3vec	Volt
Voltage (L3-N)	EMVL2_Nvec	Volt
Voltage (L-L)	EMVL_L_sysvec	Volt
Voltage (L-N)	EMVL_N_sysvec	Volt
Current (L1)	EMI_L1vec	Ampere
Current (L2)	EMI_L2vec	Ampere
Current (L3)	EMI_L3vec	Ampere
Active Power (L1)	EMAP_L1vec	kW
Active Power (L2)	EMAP_L2vec	kW
Active Power (L3)	EMAP_L3vec	kW
Total Active Power	EMAP_sysvec	kW
Reactive Power (L1)	EMRP_L1vec	kVAR
Reactive Power (L2)	EMRP_L2vec	kVAR
Reactive Power (L3)	EMRP_L3vec	kVAR
Total Reactive Power	EMRP_sysvec	kVAR
Apparent Power (L1)	EMApP_L1vec	kVA
Apparent Power (L2)	EMApP_L2vec	kVA
Apparent Power (L3)	EMApP_L3vec	kVA
Total Apparent Power	EMApP_sysvec	kVA
Power Factor (L1)	EMPF_L1vec	%
Power Factor (L2)	EMPF_L2vec	%
Power Factor (L3)	EMPF_L1vec	%

Total Power Factor	EMPF_sysvec	%
Frequency	EMFvec	Hertz
Active Energy (L1)	EMAE_L1vec	kWh
Active Energy (L2)	EMAE_L2vec	kWh
Active Energy (L3)	EMAE_L3vec	kWh
Total Active Energy	EMAETvec	kWh
Partial Active Energy	EMAEPvec	kWh
Total Reactive Energy	EMRETPvec	kVARh
Partial Reactive Energy	EMARPvec	kVARh
Power Demand	EMDPvec	kW
Peak Power Demand	EMDPPvec	kW

The maximum number of Smart Plugs existent in TH1 was 4. Data was sampled at 1 second. The measured variables are represented in Table 11.

Table 11. Variables measured by Smart Plugs. All variables are matrices (except *ndtSPvec*, which is a vector)

Variable	**Variable Name**	**Units**
Time Basis	dtSPvec	datetime
Number of Samples	ndtSPvec	
Voltage	VSPvec	Volts
Current	ISPvec	Ampere
Active Power	APSPvec	kW
Active Energy	AESPvec	kWh
Signal Power	SPRssivec	dbs
On/off	SPOnvec	0/1

The Intelligent Weather Station measures data minute by minute. The variables are shown in Table 12.

Table 12. Variables measured by the IWS. All variables are vectors

Variable	**Variable Name**	**Units**
Time Basis	dtWSvec	datetime
Air Temperature	WS_ATvec	ºC
Relative Humidity	WS_RHvec	%
Solar Radiation	WS_RADvec	W/m2

The Self-Powered Wireless Sensors measure variables in 4 compartments of TH1: in the first floor, the Hall, Bedrooms 1_2 and 1_4, and in the ground floor, the Lounge. Data is sampled at 1 minute intervals. The measured data is shown in Tables 13 to 16.

Table 13. Variables measured at Hall in the first floor (appliance 26). All variables are vectors

Variable	**Variable Name**	**Units**
Time Basis	dtH1vec	datetime
Air Temperature	H1_ATvec	ºC
Relative Humidity	H1_RHvec	%
Light	H1_Lvec	Lumens

Table 14. Variables measured at Bedroom 1_2 in the first floor (appliance 27). All variables are vectors

Variable	**Variable Name**	**Units**
Time Basis	dtB12vec	datetime
Air Temperature	B12_ATvec	ºC
Relative Humidity	B12_RHvec	%
Light	B12_Lvec	Lumens

Table 15. Variables measured at Bedroom 1_4 in the first floor (appliance 28). All variables are vectors

Variable	**Variable Name**	**Units**
Time Basis	dtB14vec	datetime
Air Temperature	B14_ATvec	ºC
Wall Temperature	B14_WTvec	ºC
Movement	B14_M	%

Table 16. Variables measured at Lounge in the ground floor (appliance 29). All variables are vectors

Variable	**Variable Name**	**Units**
Time Basis	dtLvec	datetime
Air Temperature	L_ATvec	ºC
Relative Humidity	L_RHvec	%
Movement	L_M	%

Finally, Table 17 illustrates the variables measured by the Air Conditioner at bedroom 1_4. Data is measured at 1 minute intervals.

Table 167 Variables measured by the AC in bedroom B_14, first floor. All variables are vectors

Variable	**Variable Name**	**Units**
Time Basis	dtACvec	datetime
Indoor Temperature	AC_ITvec	ºC
Reference Temperature	AC_RTvec	ºC
Outdoor Temperature	AC_OTvec	ºC
Power State (On/Off)	AC_PSvec	0/1
Swing Mode (On/off)	AC_SMvec	0/1
Eco Mode (On/off)	AC_EMvec	0/1
Turbo Mode (On/off)	AC_TMvec	0/1
Operational Mode	AC_SMvec	
Fan Speed	AC_FSvec	%

The*** *vec* variable are a raw version of the variables ***, with possibly interpolated data (please see below). The time basis for each one of the 34 devices is, as already specified, different, and expressed in each dt*vec variables.

As for processing a single time basis is needed, all variables have been downsampled to a 5 minutes sample time, where the values for each sample are the mean values of the corresponding variable, during the corresponding 5 minutes interval. Energy variables have been donwsampled to a 1 hour interval.

With real-time measured data, there is always the possibility of having missing or invalid data. All measured data is pre-processed, to check for possible gaps. If the number of consecutive missing values is less than 7, the values are interpolated with a moving median scheme; if not they are left as 0 and the period with no data is marked.

Data are also validated. At present only the ranges of temperature, humidity and solar radiation are verified. Valid ranges are:

- Smart Plugs: Current [0 inf]
- WS: AT [-10 50]; RH [0 120]; RAD [0 1500]
- SPWS Hall: AT [-10 50]; RH [0 120]
- SPWS Bed 1_2: AT [-10 50]; RH [0 120]
- SPWS Bed 1_4: AT [-10 50]; RH [0 120]; M [0 100]
- SPWS L: AT [-10 50]; RH [0 120]; M [0 100]
- AC: AC_RT [-10 50]; AC_IT [-10 50]; AC_OT [-10 50]

The information about interpolated data, gaps and faults can be found at the data file with the extension *_stat.mat*. This information can be seen in the following matrices (notice that the categories and device numbers in Table 1 are used here):

- *STEM*, *ENDEM* – matrices with the number of rows equal to the number of appliances, recording the start and the end of periods without data
- *STON*, *ENDON* - start and end samples of the periods with data
- *nEM*/*nON* - number of periods without data/with data
- *inicio*/*fim* - beginning/end of the data acquisition for each appliance
- *ttotal* - total number of seconds of the specified period of analysis

Each gap can be inspected with:

- *gaps* - array of records with all the gaps. Their structure is:
  - *devices* (category of the appliance)
  - *num* (appliance number)
  - *k* – sample index for the start of the gap

- *tbeg*/*tend* - time of the start/end of the gap

- *tgap* - total duration (in secs) of the gaps for each appliance

Faults can be inspected with:

- *tfault* - information about the total duration of the faults: array with 7 records for each device group. Each record has the following fields:
  - *num*: number of variables checked for the category
  - *dev*: array of records with the number of appliances in the group which are checked for validity: each record has:
    - *nvars* (number of variables checked)
    - *var* (variable names)
    - *t* (total faulty time for the specified variable).
  - *faulttot* – array with records for each fault. It has the following fields:

    devices – category of the appliance

    num – appliance number

    var – variable inspected

    kbeg/*kend* – sample numbers where the fault started/ended

    tbeg/*tend* - time the fault started/ended

- *nsamplesint*/*nsamples* - number of interpolated samples/total number of samples per appliance