COMP6200 Portfolio 2

The goal of the second portfolio is to reproduce some work on predicting the energy usage of a house based on Internet of Things (IoT) measurements of temperature and humidity and weather observations.

Introduction: the electricity consumption in domestic buildings is explained by two main factors: the type and number of electrical appliances and the use of the appliances by the occupants. Naturally, both factors are interrelated. The domestic appliances use by the occupants would leave traceable signals in the indoor environment near the vicinity of the appliance, for example: the temperature, humidity, vibrations, light and noise. The occupancy level of the building in different locations could also help to determine the use of the appliances. In this work, the prediction was carried out using different data sources and environmental parameters (indoor and outdoor conditions). Specifically, data from a nearby airport weather station, temperature and humidity in different rooms in the house from a wireless sensor network and one sub-metered electrical energy consumption (lights) have been used to predict the energy use by appliances.

This work explores several questions. Is the weather data obtained from a nearby weather station representative enough to improve the appliances energy consumption prediction? Can the temperature and humidity measurements from a wireless network help in the energy prediction? From all the data used in prediction models, which parameters are the most important in energy prediction?

Table 1
List of appliances in each room or house zone.

Room	Equipment
Laundry	Small Fridge, Upright freezer, Wine Cellar for 160 bottles, Washing machine, Dryer, Internet router, internet hub, Network Attached Storage
Garage*	Rain water pump, electric garage door
Kitchen	Fridge, Induction cooktop, Kitchen hood, Microwave, Oven, Dishwasher, Coffee machine
Dining	WIFI booster, ZigBee coordinator, electrical blinds
Living	TV 138 cm, Hard drive enclosure, DVD player, cable box, laptop, Ink-jet printer, electric blinds
Office	2 desktop computers, 3 computer screens, 1 router, 1 laptop, 1 copier-printer, electric blinds
Ironing	Alarm clock, radio, Iron, electric blind
Room 1	Alarm clock, radio, electric blind, 2 lamps
Room 2	Desktop computer, monitor, alarm clock, electric blind
Room 3	Laptop, alarm clock
Game	93 cm TV, Internet router, DVD player, PlayStation
Bathroom 1	2 electric toothbrushes, hair dryer
Bathroom 2	2 electric toothbrushes
Attic*	Computer, Musical Instruments, Amplifier

Note: The listed equipment in the rooms marked with an * are outside the measurement range of the wireless sensor network.

Dataset and Exploratory Analysis: The combined data set is split in training and test validation using CARET'S create data partition function. 75% of the data is used for the training of the models and the rest is used for testing.

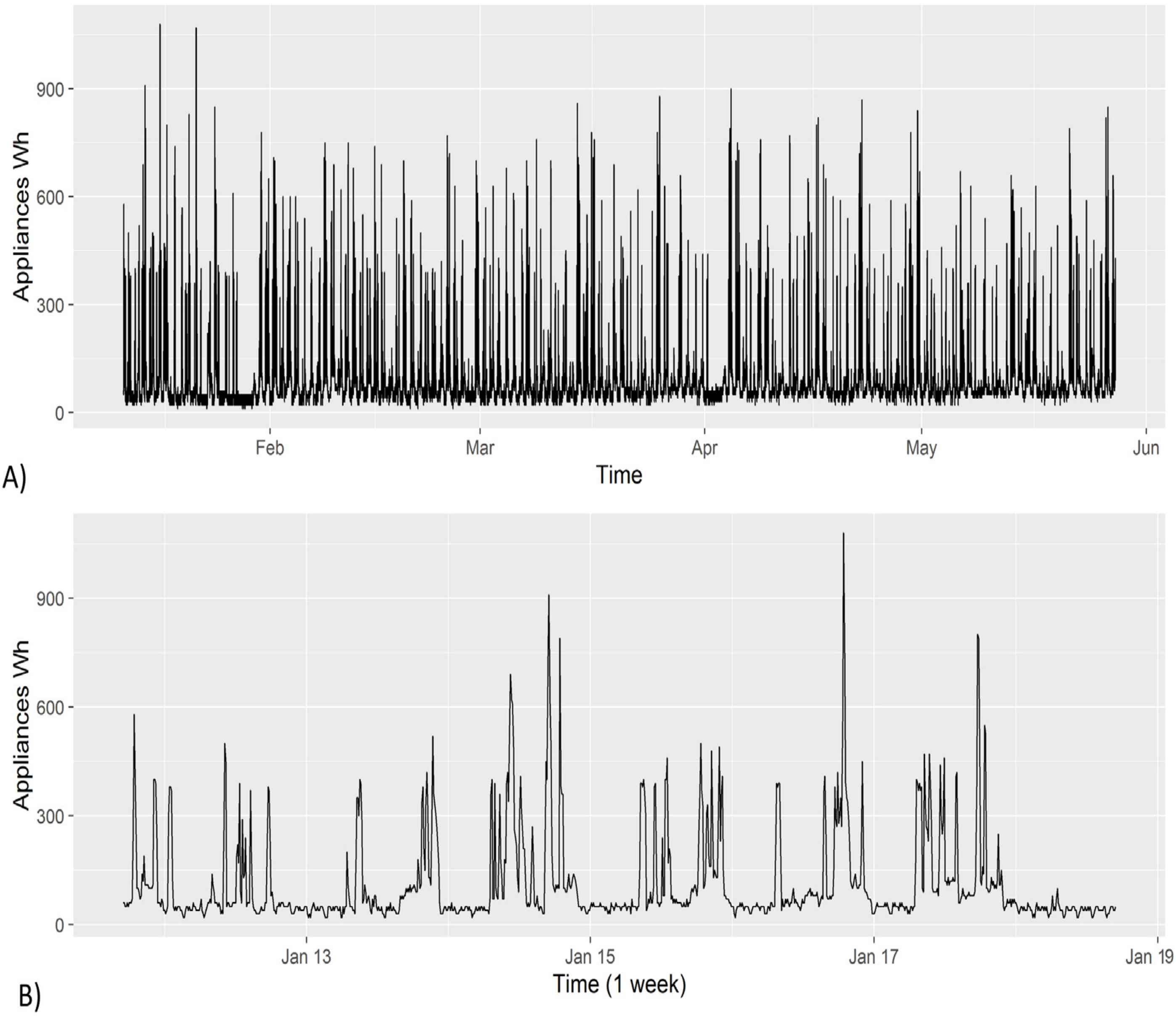

Fig 1. (A) Appliances energy consumption measurement for the whole period, (B) A closer look at the first week of data.

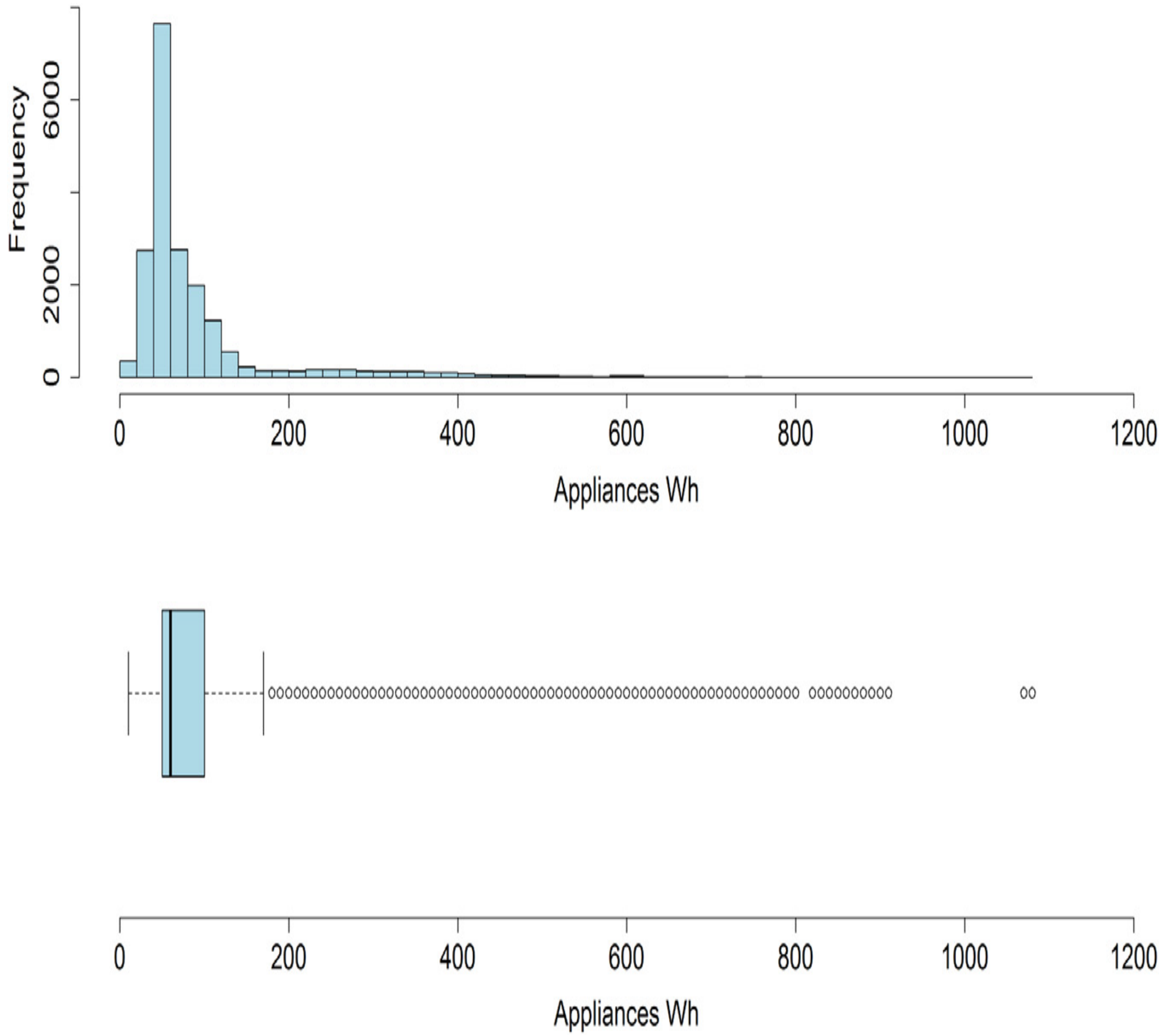

Fig. 2. Appliances energy consumption distribution. Top: histogram, bottom: boxplot. The histogram shows the frequency of energy consumption in the interval (bar width), and the boxplot shows the location of the median with the black line.

Fig. 3 shows that there is a positive correlation between the energy consumption of appliances and lights (0.19). The second largest correlation is between appliances and T2. For the indoor temperatures, the correlations are high as expected, since the ventilation is driven by the HRV unit and minimizes air temperature differences between rooms. For example, a positive correlation is found with T1 and T3.

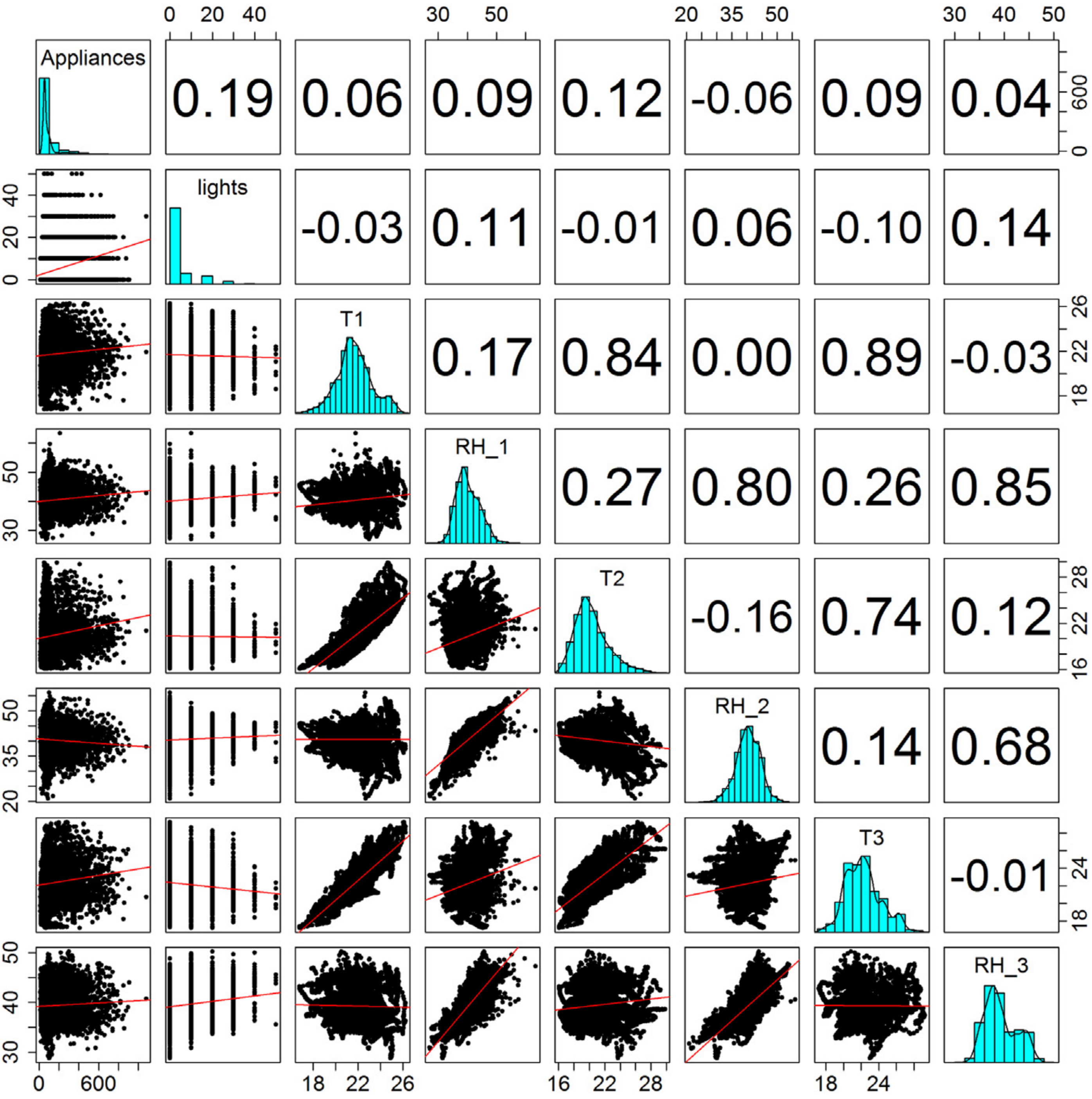

Fig. 3. Pairs plot. Relationship between the energy consumption of appliances with: lights, T1, RH1, T2, RH2, T3, RH3. T1 and RH1 correspond to the kitchen conditions; T2and RH2 correspond to the living room conditions.

An hourly heat map was created for four consecutive weeks of data to identify any time trends (See Figure 4). As can be clearly seen, there is a strong time component in the energy consumption pattern. The energy consumption starts to rise around 6 in the morning. Then around noon, there are energy load surges. The energy demand also increases around 6 pm. There is no clear pattern regarding the day of the week.

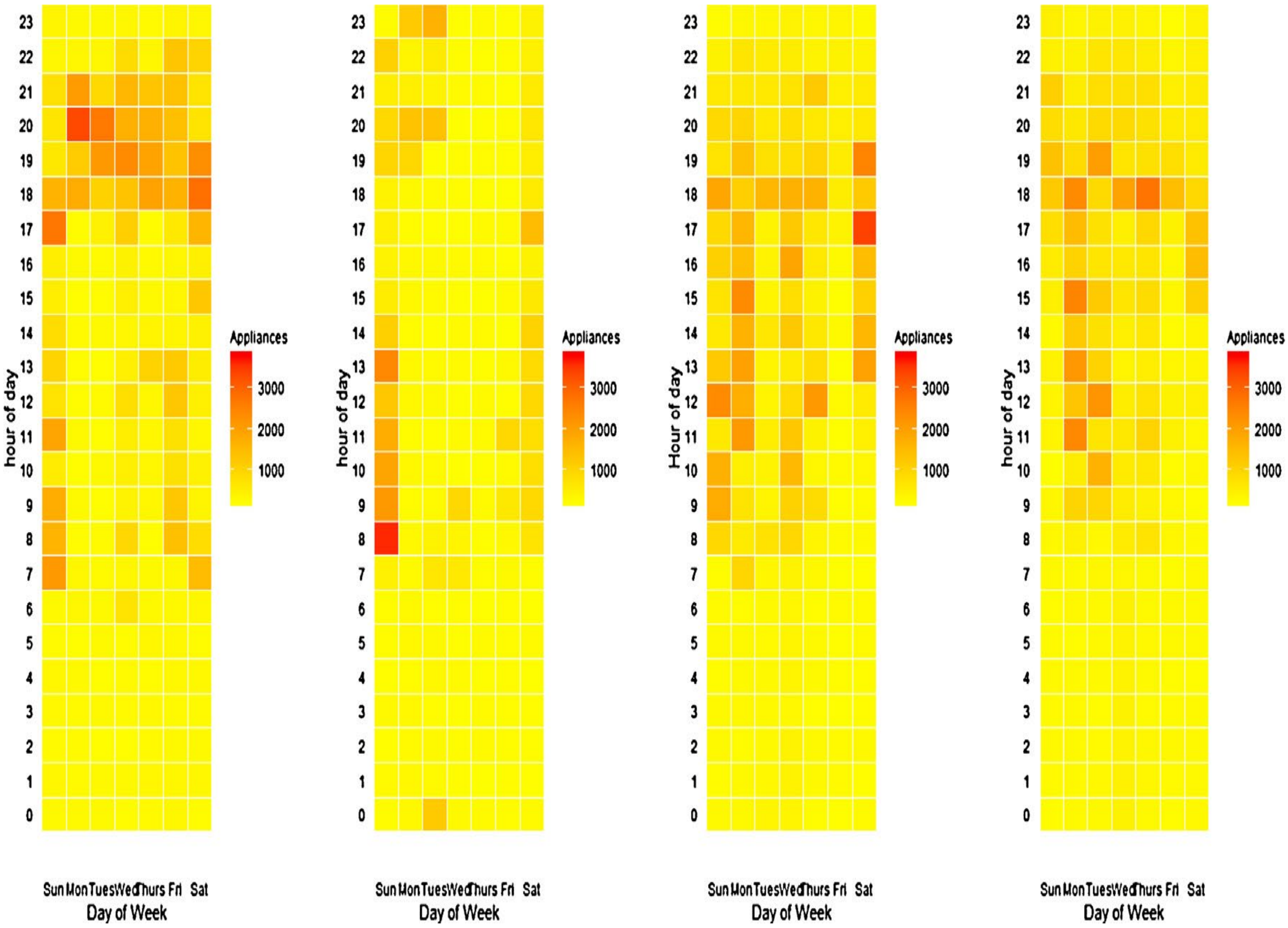

Fig. 4. Hourly energy consumption of appliances heat map for four consecutive weeks.

Table 2
Data variables and description.

Data variables	Units	Number of features
Appliances energy consumption	Wh	1
Light energy consumption	Wh	2
T1, Temperature in kitchen area	°C	3
RH1, Humidity in kitchen area	%	4
T2, Temperature in living room area	°C	5
RH2, Humidity in living room area	%	6
T3, Temperature in laundry room area	°C	7
RH3, Humidity in laundry room area	%	8
T4, Temperature in office room	°C	9
RH4, Humidity in office room	%	10
T5, Temperature in bathroom	°C	11
RH5, Humidity in bathroom	%	12
T6, Temperature outside the building (north side)	°C	13
RH6, Humidity outside the building (north side)	%	14
T7, Temperature in ironing room	°C	15
RH7, Humidity in ironing room	%	16
T8, Temperature in teenager room 2	°C	17
RH8, Humidity in teenager room 2	%	18
T9, Temperature in parents room	°C	19
RH9, Humidity in parents room	%	20
To, Temperature outside (from Chièvres weather station)	°C	21
Pressure (from Chièvres weather station)	mm Hg	22
RHo, Humidity outside (from Chièvres weather station)	%	23
Windspeed (from Chièvres weather station)	m/s	24
Visibility (from Chièvres weather station)	km	25
Tdewpoint (from Chièvres weather station)	°C	26
Random Variable 1 (RV_1)	Non dimensional	27
Random Variable 2 (RV_2)	Non dimensional	28
Number of seconds from midnight (NSM)	s	29
Week status (weekend (0) or a weekday (1))	Factor/categorical	30
Day of week (Monday, Tuesday... Sunday)	Factor/categorical	31
Date time stamp	year-month-day hour:min:s	–

Performance of Regression Models:

Table 3
Training and testing data set.

Data set	Number of observations
Training	14,803 and 32 variables
Testing	4932 and 32 variables

Table 5
Models performance.

Model	Parameters/features	Training				Testing			
		RMSE	R^2	MAE	MAPE %	RMSE	R^2	MAE	MAPE %
LM	Light, T1,RH1,T2,RH2,T3, RH3,T4, RH4,T5,RH5,T6, RH6, T7,RH7,T8,TH8,T9,RH9, To,Pressure,Rho,WindSpd, Tdewpoint, NSM, WeekStatus, Day of Week	93.21	0.18	53.13	61.32	93.18	0.16	51.97	59.93
SVM Radial	Light,T1,RH1,T2,RH2,T3,RH3, T4,RH4,T5,RH5,T6,RH6,T7,RH7,T8,TH8,T9,RH9,To, Pressure,Rho,WindSpeed, Tdewpoint,NSM, WeekStatus, Day of Week	39.35	0.85	15.08	15.60	70.74	0.52	31.36	29.76
GBM	Light,T1,RH1,T2,RH2,T3,RH3, T4,RH4,T5,RH5,T6,RH6, T7,RH7,T8,TH8,T9,RH9,To, Pressure,Rho,WindSpeed, Tdewpoint,NSM, WeekStatus, Day of Week	17.56	0.97	11.97	16.27	66.65	0.57	35.22	38.29
RF	Light,T1,RH1,T2,RH2,T3,RH3, T4,RH4,T5,RH5,T6,RH6,T7, RH7,T8,TH8,T9,RH9,To, Pressure,Rho,WindSpeed, Tdewpoint,NSM, WeekStatus, Day of Week	29.61	0.92	13.75	13.43	68.48	0.54	31.85	31.39

Feature Importance:

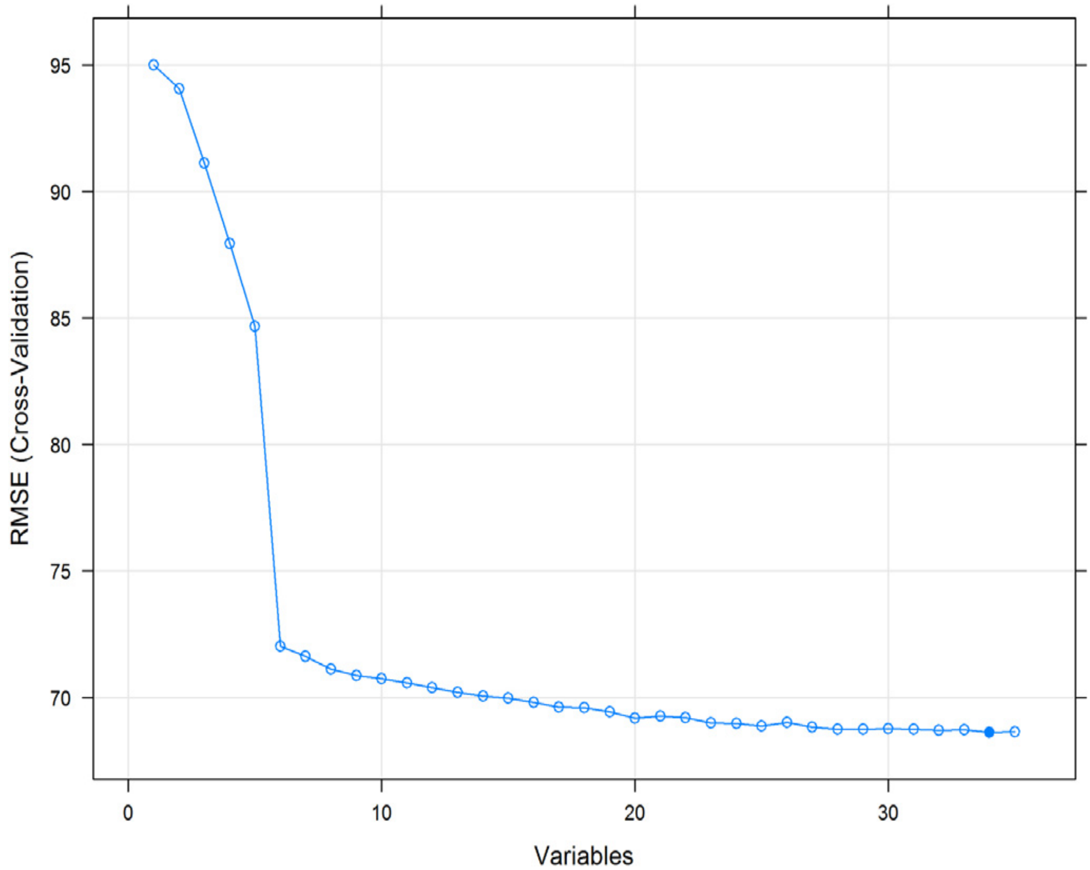

Fig. RMSE using the RFE algorithm. The optimal number of predictors (34) is shown with the filled dot.