# Creating value from energy data : the disaggregation challenge

## Introduction

All stakeholders in the energy industry, governments, utilities, distribution network operators (DNOs) and consumers themselves acknowledge a growing need for higher consumers' involvement, in order to make them more aware and responsive while bringing them more serenity with respect to their bill, their comfort, and their ecological imprint.

Indeed, the residential sector totalizes 40% of the total energy consumption in the world and over 35% of the electricity consumption. The residential consumers are by definition small, scattered and individually unpredictable as compared to industrial or commercial players. It is therefore difficult to reach them and even more problematic to manage their demand. Thus, the residential sector is the most challenging segment of the future smart grids.

Governments target long term behavior changes towards more frugality because they know that, on the long run, whatever the energy mix, "the best energy is still the energy you don't use".

Utilities on the other hand are for obvious reasons not altogether enthusiastic about energy savings. Yet, they target customers' loyalty, and to this end, they have to comply with their aspiration, namely, in short, to know how much they spend and for what, and how to save money and facilitate their lives. While interacting with their consumers about those legitimate concerns, utilities have the opportunity to gather valuable information about detailed consumption typologies.

Moreover, distribution network operators, in charge of network security, have to anticipate more and more local production and consumption peaks. Indeed, in the near future, local production peaks will originate from the development of renewable energy systems such as residential photovoltaic generation and local wind mills, whereas new consumption peaks may stem from new devices and habits, such as electrical vehicles drawing power in the evening when HVAC are still on. Home energy management, including residential storage capacity management and DR schemes may become a key component of future global smartgrid management.

Consumers themselves, with various motivations, ranging from the economic to the ecological, want to be reassured that their energy consumption and consequently their energy bill are not excessive, that their devices are not dysfunctional, and that they can take advantage of economic opportunities related to local production and demand response.

After all, electricity, along with gas and water, are the only goods people have to buy "on the fly" without really knowing how much they will pay in the end.

## Our vision

WattGo ran several studies from 2013 to 2014 on French households' motivations and interests regarding their electricity and energy bill. The key findings of these surveys[1] were :

- 61% of households are asking for notifications and alerts about unusual electrical consumption.

- 53% of households want to know the cost of each electrical appliance usage.

- 50% of households struggling to pay their electric bill, are asking for a bill forecast service.

- 73% of households would accept to receive alerts, via e.g. SMS, at peak load times, asking them to limit electrical appliances use.

In line with those figures, we believe that utilities will have to move up another gear to answer their customers' concerns and ensure their loyalty, by delivering a package of services such as:

- orienting each consumer towards the right offer, along with the availability of a home energy management dashboard (displaying bill forecasts, advices on dynamic pricing, pie charts of energy end use) in order to help them manage their energy,

- providing each household with the right message at the right time, about devices and usages, through specific alerts and contextual advices, in order to generate peace of mind,

- giving consumers relevant benchmark data together with real time feedback.

## Load curve disaggregation

A number of those services are enabled only by load curve disaggregation. This is why it has been coined as representing the "holy grail" for utilities or energy services companies [4]

Load curve disaggregation or Non-Intrusive Load Monitoring (NILM) refer to algorithms which aim at extracting individual appliance or end-use level data from a home load curve at a single point (the electricity meter). No plug level additional sensors are needed and yet (major) appliances (or end uses) can be detected: the disaggregation algorithm can determine when and how much they draw power as if they were monitored…

Many consumers, even those who are highly motivated by neighborhood comparisons, competitions and social platforms, found that basic energy monitors that simply display their energy consumption had virtually no impact in helping them save energy.

On the contrary, knowing which individual appliance or device uses energy, when and how much, may help them focus on the largest uses and identify opportunities for demand response.

Disaggregation leads to different types of services, depending on its being operated on line or off line. Real Time or Near Real Time is required for equipment specific alerts (for example a water heater left "on" during an absence or a sudden change in the load pattern of a cooling appliance), whereas off line calculation is perfectly fit to energy pie charts and end-use bill forecasts. The three kinds of services mentioned above use both on line and off line calculation.

## Our approach

Energy services companies in the residential sector were more or less originally divided into two groups of players: those coming from the NILM community, focusing on electrical load curve measurement and analysis (e.g., Bidgely) and those relying on sociological knowledge to address consumers' behavior change through a wide range of informative actions and incentive schemes (e.g., Tendril, Opower)

While all players tend now to extend their original skills to provide a more comprehensive range of services, WattGo has from the start chosen to run on two legs:

- a panel of households, representative of French electricity consumption, monitoring in real time their consumption at the meter, and describing their devices and usages by a comprehensive questionnaire: PowerMetrix

- a novel set of self-learning NRT disaggregation tools: HomePulse.

Indeed, we believe that prior knowledge and understanding of the devices and usages "embedded" in a home load curve is key to complete any disaggregation task under operational conditions, however advanced the algorithms may be. This is why we took up the challenge to learn about devices and usages directly from aggregated load curves, building our knowledge base from an initial succinct set of

records up to a substantial archive as we go along with our panel load curves disaggregation task.
The approach can easily be extended to any beta-tester customers' community in any country, as soon as the initial customer base encompasses the most country-specific devices and usages.

## Powermetrix, a representative panel of load curves

Building a representative panel (as from 2012) led us to recruit, through various recruitment channels, some 1300 voluntary households, half of them being equipped with various sensors, including at least a medium resolution (5" to 10") sensor acquiring data on their energy meter (mainly electricity meter, some gas meters have been instrumented as well). For each panelist, monitoring is conducted continuously and the resulting data are securely transferred to our servers to be fed into our database and processed.
These energy consumption data are enhanced by hundreds of quantitative and qualitative variables that we have collected via online questionnaires about devices and usages, providing unique insight into households' electricity consumption.
We developed our own "big data friendly" IT framework, running Spark real time framework and Cassandra database, in order to achieve massive NRT implementation and full scalability.
Thanks to our panel, we already deliver statistical analysis and home centric market research studies to the main market players, such as :

- households' segmentation according to electricity consumption,

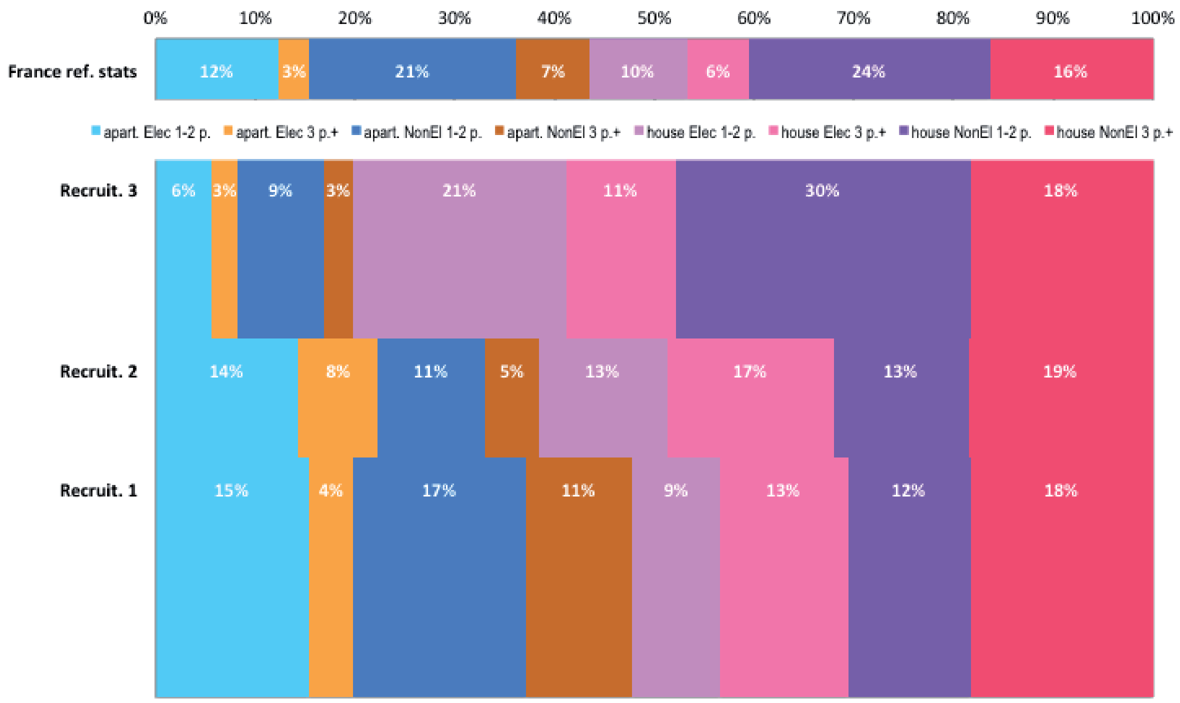
- impact of specific endogenous or exogenous data,

- "Usage and Attitudes" surveys over specific household electrical appliances,

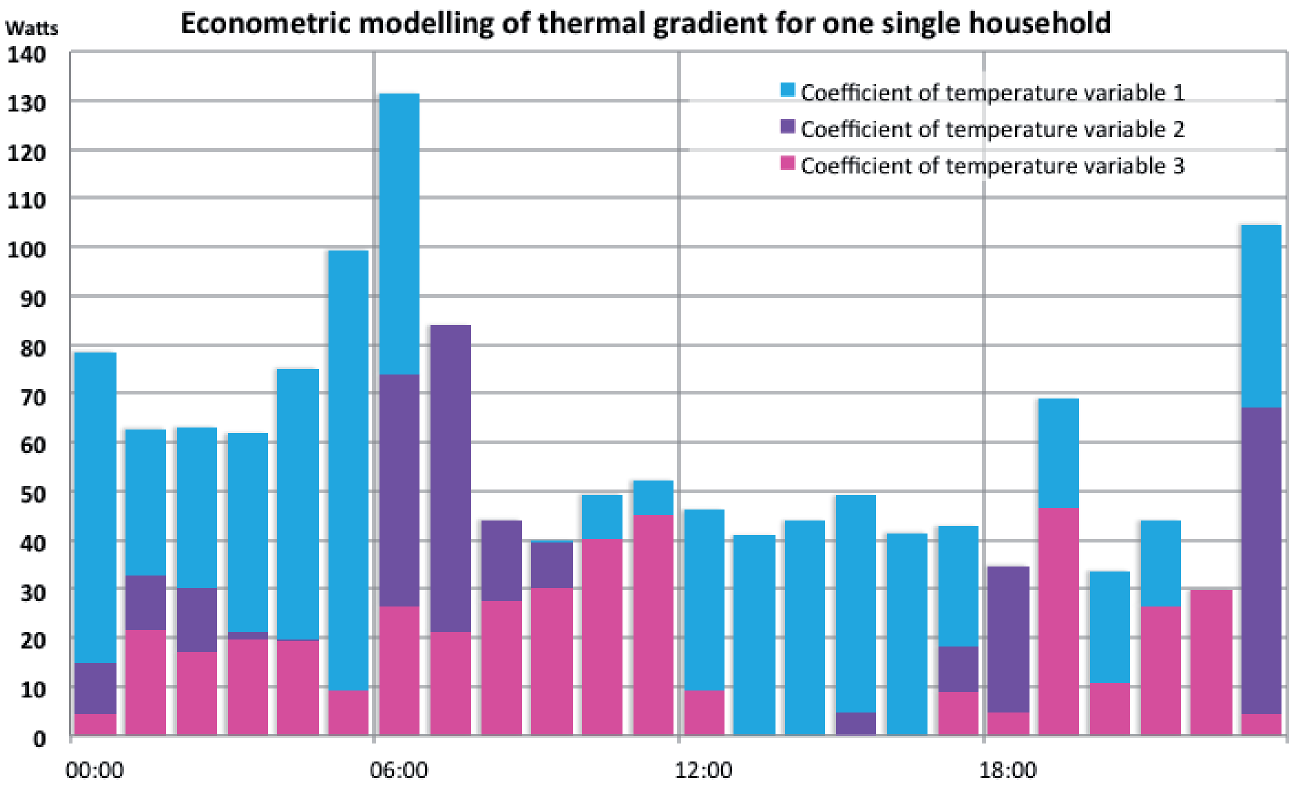
- « before/after», «with/without » or prospective analyses to identify and test new concepts of services related to energy management and smart home (attractiveness, usages…),

- econometric study of temperature-related electricity consumption feeding long term projections or shorter-term forecasting,

- analyses of society issues, such as fuel poverty.




How alternative recruitment channels (here 1, 2, 3) providing contrasted profiles, help in building a representative panel of households.




Econometric modeling of temperature-related electricity consumption at individual home level brings powerful forecasting and demand response capabilities

## Home Pulse, a set of cutting-edge disaggregation algorithms

Requirements

Our goal was to find a way to disaggregate the load curve in order to deliver the services listed above under realistic and sensible operational conditions.

This general requirement breaks down into the more specific following ones: 1. disaggregation on the fly (or Near Real Time) 2. of the main appliances or end uses (heating, water heating, air conditioning, fridges and freezers, cooking, baseline load, stand by load), 3. without requiring previous learning of individual appliance consumption sequences, 4. without thorough description of the appliances brought in the individual home, 5. without specific hardware development on the meter, 6. performing on both purely resistive and inductive appliances.

We nevertheless accept a learning period lasting a few weeks, during which the collection of aggregated data helps in training the algorithm without delivering results to the customer; at the end of this learning period, we may ask the customer targeted questions in order to confirm or infirm key assumptions about devices, and thus improve accuracy.

State of the art of academic research

The disaggregation problem was introduced first by Hart in the late 1980s [1]. Looking more closely at the disaggregation problem, approaches known or thought of so far (see methods overview in [2, 3 and 4, table A1]) fall into one of the two following schemes:

- either they extract first, from the aggregated load data, whatever signatures, features or characteristics of individual unknown appliances embedded in the load curve they can find, and then they try next to “match” them to some appliance in a data base, in order to label them,
- or they learn beforehand about the appliances brought in the individual home, as precisely as possible (each particular appliance’s load profile through time, which captures the usage as well), and this knowledge feeds a breakdown algorithm.

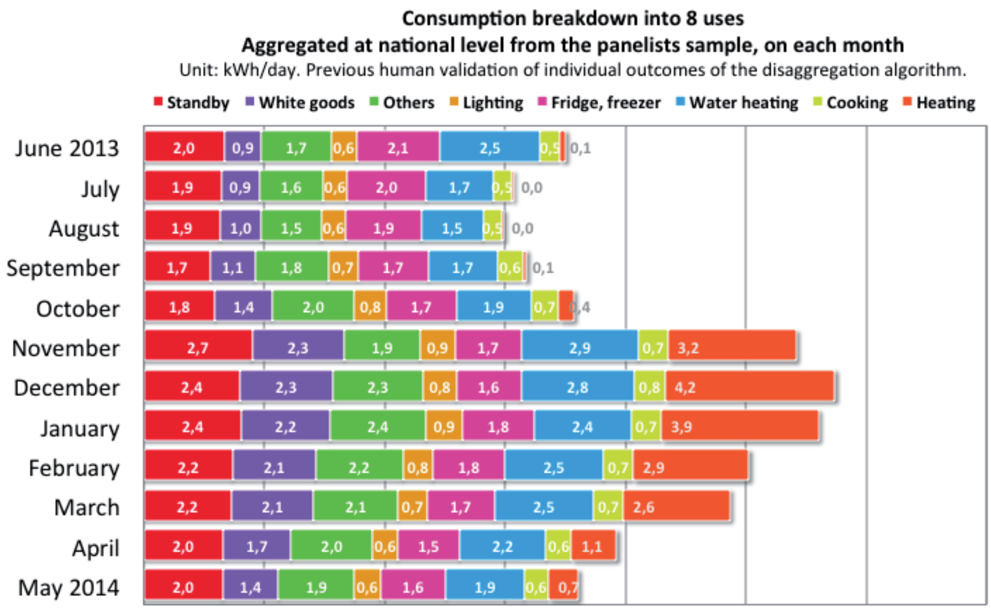
Methods in this second group need training on home individual appliance data (they are called “supervised” methods); this runs in direct contrast with our requirement 3 above, which only outlines the aim of NILM, which is not to require the "intrusion" of plug-level sensors.

To our knowledge, methods in this second group systematically rely on Hidden Markov Models.

Mathematically, the load curve disaggregation problem clearly lies in the application field of Factorial Hidden Markov Models (FHMM). These are classic, albeit non trivial and computationally heavy methods, based on hidden states Markov chains (indeed intractable in an NRT context). Each major appliance in the home is represented by such a chain, whose model parameters have theoretically to be learned beforehand from the individual appliance data (supervised methods). Results of such disaggregation methods based on FHMM methods are presented in [5,6]. If trained on other appliances of the same category, typically in other homes, standard FHMM methods perform very poorly. The Reference Disaggregation Energy Data Set (REDD) release paper [7] clearly shows the drop in performance when training data are different from test data.

However, these standard FHMM methods may not be representative of the current state of the art. Indeed, some extensions of the standard FHMM methods have been proposed recently such as in [8,9]. E.g. some recent thesis work [10] attempted to tune FHMM methods in order to learn HMM model parameters for some devices (mainly refrigerators) directly from home aggregated data. To that purpose, general appliance HMM models representing appliance types are proposed. They consist in prior distribution over the HMM parameters, the parameters of these prior distribution being learned from a few appliance instances of the type (using an appliance data base such as Trace Base [11]). The author proposes a method to tune the general appliance model to the specific household appliance instance using only aggregate data. The crucial point is then to extract the appliance signature from the aggregate load for a sufficient time period. The latter applies to fridges and freezers, whose patterns appear quite clearly at night if neither air conditioning nor heating are on.

We had the opportunity to test extensively by ourselves a FHMM disaggregation method. It was implemented so as to train the major appliances models on more or less individual appliances data extracted empirically from the aggregated load curve (according to customer's answers to a short questionnaire about their main appliances and some heuristics). The method required several weeks of training to produce any reliable results, proved extremely sensitive to incorrect devices' lists as well as to unexpected alterations in the load curve profile.




In case of electrical heating and/or high level of noise, most usages including standby, water boiler and white appliances were overestimated by 120 to 150% at national level

Whatever the future improvements in the HMM field, we consider it likely that these methods:

- will remain unfit to NRT implementation (in conflict with requirement 1),
- will end up in requiring previous training on individual appliances data (in conflict with requirements 3 and/or 4).

In fact, most of published disaggregation methods fall into the first category. They nevertheless differ greatly from each other, according to:

- the kind of "electrical signature" that is extracted from the load curve, which in turn depends crucially on acquisition frequency range and nature of data,
- the type of "matching algorithm" used to recognize and label the appliances which could have been detected, which of course depends on the mathematical description of this signature (parametric or not, probabilistic description allowing Bayesian inference or not…)

## Our choices

Having discarded HMM methods as unable, for the time being, to fulfill requirements 1, 3 and 4, we turned to the first scheme "extract first, then match".

According to requirement 5, we discarded any method requiring custom hardware to acquire and preprocess the electrical signal (detection of high range harmonics and /or on/off transient signatures, i.e. above 10 kHz acquisition frequency).

According to requirements 2 and 6, we discarded any method relying directly on detection (on a millisecond basis from a CT clamp for example) of waveform perturbations due to inductive receptors (such as HVAC, fridges, pool pumps, heat pumps etc.), which is bound to perform poorly for purely resistive appliances such as electricity heating, water heater and ovens.

Anticipating the roll out of smart meters with 1s to 10s data HAN delivery capacities, we chose to monitor current meters with easily self-installed optical sensors delivering active power data on a few seconds basis.

## Our algorithms (patents pending)

We developed accordingly a set of original algorithms to:

- Extract appliances' signature from the load curve,

- Describe those signatures with a few (10 to 20) relevant parameters,

- Perform Bayesian inference to label the signatures,

- Enhance continuously an appliances database, and model parameters coded in this database by way of probability densities,

- Optimize signature extraction for highest recognition at the labelling step.

## Achievements so far

A first implementation of the technology is performed for Budget Telecom and its service Qinergy.com, which offers residential customers a detailed analysis of their electricity consumption (real time dashboard, history, consumption targets) including the overall consumption break down into several main end uses: baseline load and standby, water heating, fridge and freezer, heating. The service delivers weekly end-use bills, trends and benchmarks as well as annual end-use bill forecasts (notifications and specific alerts are coming soon).

## References


1. G.W. Hart. Nonintrusive appliance load monitoring. Proceedings of the IEEE, 80 (12): 1870–1891, Dec 1992. ISSN 0018-9219. doi: 10.1109/5.192069.

2. M. Zeifman and K. Roth. Nonintrusive appliance load monitoring: Review and outlook. Consumer Electronics, IEEE Transactions on, 57 (1): 76–84, February 2011. ISSN 0098-3063. doi: 10.1109/TCE.2011.5735484.

3. A. Zoha, A. Gluhak, M. A. Imran, and S. Rajasegarar. Non-intrusive load monitoring approaches for disaggregated energy sensing: A survey. Sensors, 12 (12): 16838–16866, 2012. ISSN 1424-8220. doi: 10.3390/s121216838. URL http://www.mdpi.com/1424-8220/-12/12/16838.

4. Carrie Armel, K., et al., Is disaggregation the holy grail of energy efficency? The case of electricity, Energy policy (2012), http://dx.doi.org/10.1016/j.enpol.2012.08.062

5. Marc Bons. Modèles a source markovienne cachée appliqués a l'analyse non-intrusive des principaux usages domestiques de l'électricité. PhD thesis, 1996.

6. J. Zico Kolter and Tommi Jaakkola. Approximate inference in additive factorial HMMs with application to energy disaggregation. In Neil D. Lawrence and Mark Girolami, editors, AISTATS, International Conference on Artificial Intelligence and Statistics, volume 22 of JMLR Proceedings, pages 1472–1482. JMLR.org, 2012.

7. Zico Kolter and Matthew J. Johnson, Reference Energy Disaggregation data Set (REDD), Precourt Energy Efficiency Center, Standford University.

8. Hyungsul Kim, Manish Marwah, Martin F. Arlitt, Geoff Lyon, and Jiawei Han. Unsupervised disaggregation of low frequency power measurements. In SDM'11, pages 747–758, 2011. URL http://www.cs.uiuc.edu/~hanj/pdf/sdm11_hkim.pdf.

9. Johnson, M. J., and Willsky, A. S. (2013). Bayesian nonparametric hidden semi-Markov models. The Journal of Machine Learning Research, 14(1), 673-701.

10. Oliver Parson, Unsupervised training methods for non-intrusive appliance load monitoring from smart meter data, University of Southampton, PhD Thesis.

11. Andreas Reinhardt, Paul Baumann, Daniel Burgstahler, Matthias Hollick, Hristo Chonov, Marc Werner, Ralf Steinmetz: On the Accuracy of Appliance Identification Based on Distributed Load Metering Data. Proceedings of the 2nd IFIP Conference on Sustainable Internet and ICT for Sustainability (SustainIT), pp. 1-9, 2012.

Authors:

Jean-François Hochedez, PhD, Research Director Wattgo; Romain Benassi, PhD, Researcher Wattgo; Florence Richard, Research Manager Wattgo; Frédéric Lefevbre-Naré, Chief Scientific Director Wattgo.

Dr Jean-François Hochedez graduated from Ecole Centrale des Arts et Manufactures in 1987 and received his PhD from the Paris XI Orsay University in 1990. As a staff scientist at the Royal Observatory of Belgium, he has been Principal Investigator and co-Investigator of several space instruments and led its Solar Physics department. His preferred approaches in Research rely on the application of mathematics and instrumentation to the preparation and exploitation of observational data. This is as well the line that he brings to the presented Wattgo developments.

Romain Benassi graduated from Télécom Bretagne, with a specialization in signal processing, and received a Ph.D. from Supélec on optimization and statistics issues. He works now, for more than a year, as a research engineer in Wattgo, a start-up in the field of energy, mainly on topics such as load curve disaggregation and statistical studies about residential consumption of electricity.

Florence Richard graduated from France's Ecole Polytechnique school of engineering and from ENSAE ParisTech (Paris Graduate School of Economics, Statistics and Finance). She holds a MS in Probability and Stochastic Models and spent 13 years as Research Scientist at ONERA (the French Aerospace Lab) where she specialized in Guidance, Navigation and Control. She was head of missiles'performance team. She joined WattGo's team in June 2014 as Senior Scientific Advisor.
Frédéric Lefebvre-Naré graduated from Ecole Polytechnique in 1987, École Nationale des Ponts et Chaussées in 1991, holds a post-graduate degree in Econometrics from ENSAÉ Paris, 1994, and attended the 9th session of IHEE (Institute for Advanced Business Studies), 1998-99. He devoted most of his career to applying maths and statistic modeling to social sciences, marketing research, economics and big data at large. He served as CSO and CTO of media audience tracking institute Médiamétrie, as co-CEO of Evalua, a public policies evaluation consultancy, as advisor to Presidential candidate François Bayrou, and is WattGo's CSO since 2013s.

For more information: info@wattgo.com
www.homepulse.fr
January 2015