



HAL
open science

Human spatial dynamics for electricity demand forecasting: the case of France during the 2022 energy crisis

Nathan Doumèche, Yann Allioux, Yannig Goude, Stefania Rubrichi

► To cite this version:

Nathan Doumèche, Yann Allioux, Yannig Goude, Stefania Rubrichi. Human spatial dynamics for electricity demand forecasting: the case of France during the 2022 energy crisis. 2023. hal-04219577v2

HAL Id: hal-04219577

<https://hal.science/hal-04219577v2>

Preprint submitted on 14 Nov 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Human spatial dynamics for electricity demand forecasting: the case of France during the 2022 energy crisis

Nathan Doumèche^{1,2*}, Yann Allioux², Yannig Goude²,
Stefania Rubrichi³

¹*LPSM, Sorbonne University, 4 place Jussieu, Paris, 75005, France.

²OSIRIS, EDF R&D , Bd Gaspard Monge, Saclay, 91120, France.

³SENSE lab, Orange Innovation, 46 Av. de la République, Châtillon, 92320, France.

*Corresponding author(s). E-mail(s):

nathan.doumeche@sorbonne-universite.fr;

Contributing authors: yann.allioux@edf.fr; yannig.goude@edf.fr;

stefania.rubrichi@orange.com;

Abstract

Accurate electricity demand forecasting is crucial for energy security and efficiency, especially when relying on intermittent renewable energy sources. Recently, massive savings have been observed in Europe following an unprecedented global energy crisis. However, assessing the impact of such a crisis and of government incentives on electricity consumption behaviour is challenging. Moreover, standard statistical models based on meteorological and seasonal data have difficulty dealing with such sudden changes. Here, we show that mobility indices based on mobile network data significantly improve the performance of state-of-the-art models in electricity demand forecasting during France's government-pushed energy sobriety period. We start by documenting the drop in French electricity consumption during the winter of 2022-2023. We then show how our mobile network data captures work dynamics and how adding such mobility indices to models outperforms the state-of-the-art during that atypical period. In effect, our results characterise the effect of employment-related behaviour patterns on electricity demand.

Keywords: Electricity demand forecasting, machine learning, mobile phone data, Kalman filter, energy crisis, environmental and energy transitions

Energy is at the very core of modern economies and politics, powering industry, transport, residential use, and agriculture [1]. Over the past two years, Europe has experienced a major energy crisis with energy prices reaching levels not seen in decades [2]. Prices began to rise rapidly in the summer of 2021 as the global economy picked up following the easing of COVID-19 restrictions. Subsequently, the war in Ukraine led to a significant reduction in gas supplies, pushing gas prices even higher [3]. In this context, the European Union adopted the Council Regulation 2022/1854 in October 2022 [4], which established a series of emergency measures to mitigate the effects of this crisis, mainly by reducing electricity demand, with a binding reduction target of 5% during peak hours.

In France, where a significant proportion of nuclear plants were also offline [5], the government called for a voluntary effort to reduce energy consumption by 10% over two years and launched its own energy sobriety plan [6]. Various media documented a subsequent drop in France’s electricity demand in the winter of 2022-2023 [7–9]. Energy saving is also part of France’s long-term policy of ecological transition and energy sovereignty. Indeed, the energy sector’s impact on climate change is forcing changes in consumption patterns, which is fueling a growing interest in energy savings and the transition to sustainable energy sources [10–13]. In France, electricity is one of the most important components of the energy mix, accounting for 25% of its final energy consumption, and the French Ecological Transition Plan is based on massive electrification driven by decarbonised energy, coupled with energy savings [14, 15]. While modifying human behaviour (e.g., by encouraging remote working) has been identified as an important axis of the sobriety plan, a better understanding of how this relates to energy savings is crucial for energy planning.

Recently, artificial intelligence has been recognised as a powerful tool to support the mitigation of greenhouse gas emissions and tackle climate change [16]. In particular, machine learning techniques have been applied to electricity load forecasting to ensure the electricity grid remains balanced [17] and to reduce electricity wastage. As France’s electricity storage capacity is limited and expensive to run, electricity supply must match demand at all times. As a result, electricity load forecasting at different forecast horizons has attracted increasing interest over the last few years [18].

This article focuses on so-called short-term load forecasting, or 24-hour ahead load forecasting, which is particularly relevant for operational usage in industry and the electricity market [19, 20]. We address this problem both in terms of feature selection and model design. Most state-of-the-art models rely on historical electricity load data, seasonal data such as holidays or the position of the day in the week, and meteorological data such as temperature and humidity [19]. However, such data cannot accurately account for the complex human behaviours that affect the variability of energy demand, such as holidays and—increasingly—remote working. As a result, traditional models struggle to account for unexpected large-scale societal events such as the COVID-19 lockdowns or energy savings following economic, geopolitical, and environmental crises [21]. New data capturing consumption behaviours is therefore needed to better model electricity demand. Over recent decades, datasets generated from mobile networks, location-based services, and remote sensors in general, have been used to study human behaviour [22]. Indeed, geolocation from mobile phones

makes it possible to precisely characterise human flows [23–25]. For example, such data have been used to study disease propagation [26–29], traffic [30], the impact of human activities on biodiversity [31], and water consumption [32, 33]. In terms of day-ahead load forecasting, mobility data from SafeGraph, Google, and Apple mobility reports were strongly correlated with electricity load drops in the US during the COVID-19 outbreaks [34, 35], as well as in Ireland [36] and in France [37]. Although such data is quite informative about activity in urban areas, e.g., in retail stores and train stations, it does not precisely account for human presence and flows. Indeed, there is intrinsic bias in such data collection, corresponding for example to that of using (or not) a specific application. It is therefore necessary to take into account such biases when building models using this kind of data.

In this context, the originality of this paper relies on the use of high-quality human presence data provided by the mobile network operator Orange—representing 30–40% of the French market—to model electricity demand during France’s 2022–2023 sobriety period [38]. This dataset is based on adjusted mobile phone traffic volume measurements collected continuously and passively at the mobile network level, unlike most location-based services data where the user is required to opt-in, which may introduce biases. As a result, our mobile network-based signal can be considered representative of the underlying population’s data. In this article, we start by characterising electricity savings during the sobriety period in France. We then show that our mobility data from mobile networks are correlated with other well-known socio-economic indices that capture spatial dynamics of the population. Furthermore, we show that models using mobility data outperform the state-of-the-art in electricity demand forecasting by 10% with respect to usual metrics. Finally, we show that the *work* index we have defined (see Section 3.2) has a distinctive effect on electricity demand, and is able to explain observed drops in electricity demand during holidays. Other human spatial dynamics indices such as tourism at the national level did not prove to have a significant effect on national electricity demand.

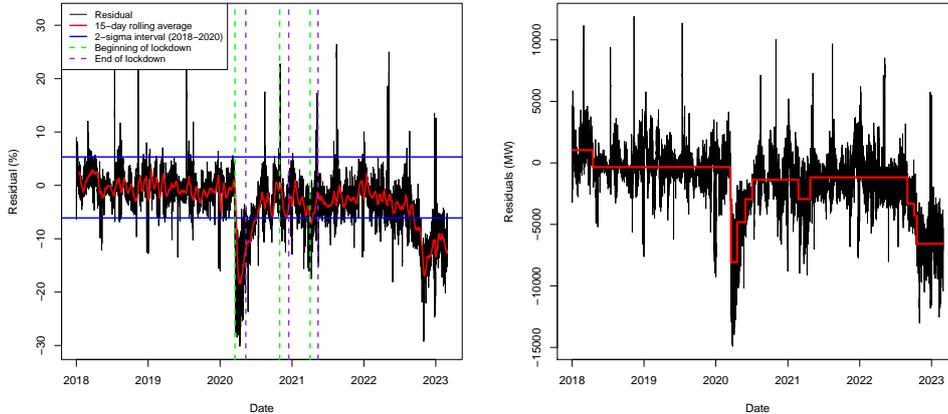
1 Results

1.1 Quantifying electricity savings

To quantify electricity savings, the effects of temperature and time seasonality must be removed from the French electricity demand data. The expected load given temperature and time, which we denote $\widehat{\text{Load}}$, is estimated using a generalised additive model (GAM). Figure 1 shows the residuals: $\text{res} = \text{Load} - \widehat{\text{Load}}$, where Load is the actual value of the electricity demand. This GAM was trained on the data from 01/01/2014 to 01/01/2018. The residuals were then evaluated from 01/01/2018 to 01/03/2023. Therefore, residuals measure the gap between the electricity demand at a given time and the expected demand with respect to its time and temperature dependency between 2014 and 2018. Negative residuals correspond to electricity savings.

In Figure 1 (left), the blue lines represent the 2σ variations over the period spanning 01/01/2018 to 01/01/2020, and correspond to typical variation in electricity demand around its expected value given the temperature and seasonal data. The holidays deviate strongly from the expected trend and correspond to the peaks in the residuals.

Fig. 1 Electricity demand corrected for the effects of temperature and annual seasonality.



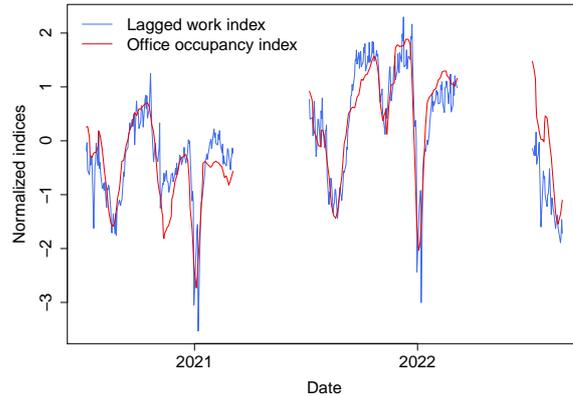
Left: Descriptive statistics of the residuals. **Right:** The ten most important change points are represented by a change in the red line. The red line is the mean of the residuals between the change points.

Note that the 15-day rolling average in red only exits this confidence interval during the lockdowns and the 2022–2023 winter’s sobriety period. This means that, during these events, the French electricity load is significantly lower than its expected values. As shown in Figure 1 (right), to detect these changes in the electricity demand, we ran a change point analysis [39, 40] using binary segmentation, which detects and orders the changes in the mean of the residuals. The two most important change points of the 2018-2023 period were at the beginning of the sobriety period (04/10/2022) and the beginning of the first COVID-19 lockdown (15/03/2020). During the sobriety period from 04/10/2022 to 01/03/2023, the residuals had a mean of -10.6%. This result was close to the assessment made by the French transmission system operator (RTE) whose estimate was of a 9% decrease in consumption during the winter of 2022-2023 [41]. Figure 1 (right) shows the ten most important change points in the residuals over the 2018-2023 period along with mean of the residuals between the change points. These results confirm that there was indeed a significant drop of around 11% in French electricity demand during the sobriety period of 04/10/2022 to 01/03/2023. As this drop is visible in the electricity load adjusted for temperature and time, this means that temperature and seasonal data are not sufficient to accurately explain these energy savings.

1.2 Mobile phone data and work dynamics

Here, we show how our dataset of aggregate mobile phone data efficiently captures the spatial dimension of social dynamics. In particular, we focus on its ability to quantify the work-related human presence (hereafter *work* index, see Section 3.2) over time. To investigate the ability of our dataset to characterise such behaviour, we used the

Fig. 2 Comparison of work indices.



Mobile network-based work index and the normalcy index’s office occupancy one. The work index in blue is lagged by 7 days. Weekends and bank holidays are excluded. Both indices have been standardised, i.e., the empirical mean has been subtracted and the result divided by the empirical standard deviation. The mobile network dataset only covers the period from July to March each year.

office occupancy index from The Economist’s normalcy index [42]. This index was developed during the COVID-19 pandemic to evaluate the impact of the latter and of government policies on human behaviour. It tracks eight variables (sports attendance, time at home, traffic congestion, retail footfall, office occupancy, flights, film box office, and public transport) at the national level; this open data can be found here: <https://github.com/TheEconomist/normalcy-index-data>. The *office occupancy* index is derived from Google’s COVID-19 community mobility reports, which are no longer being updated as of mid-October 2022 [43]. As the *office occupancy* index is only available from February 2020 to October 2022, and mobile-network dataset only covers the periods from July to February of each year, Figure 2 has missing values. As illustrated in Figure 2, the office occupancy variable is highly correlated (87%) with the 7-day lagged *work* index when excluding weekends and bank holidays. Moreover, our *work* index is more suitable than the *office occupancy* index for operational use, because it is seven days ahead of the *office occupancy* index. As detailed in Appendix A, it also contains more information, because it captures the reduction in office occupancy during weekends and holidays. This is very valuable because holidays are known to have a significant impact on electricity demand, while their effect is difficult to evaluate. This often leads to having to analyse regular days and holidays separately [44]. In the appendices, we demonstrate how tourism trends are related to another index from the same dataset.

1.3 Load forecasting with mobility data

The purpose of this paragraph is to measure the benefits of incorporating mobility data into state-of-the-art load forecasting techniques (see Appendix B for a more complete description of the models). In this field, the state-of-the-art is generally divided into three classes of forecasts [45, 46]: statistical models that approximate electricity demand by simple relationships between explanatory variables, data assimilation techniques that update a model using recent observations, and data-driven machine learning methods whose results may be more difficult to explain but are more expressive.

Here, we focus on the known state-of-the-art in French load forecasting. To evaluate the benefits of using mobility data to forecast France’s national electricity load, we ran a benchmark on the sobriety period, i.e., from 01/09/2022 to 28/02/2023. The training period spanned 08/01/2013 to 01/09/2022. Results are presented in Table 1 in terms of root mean square error (RMSE) and mean absolute percentage error (MAPE). For a full description of the models and metrics, please refer to the Methods section. Bold values highlight the best forecasts in each category. Overall, they show that incorporating mobility data improves the performance of the best forecast (aggregation of experts) by about 15% in RMSE and 10% in MAPE, and indeed, for all models the improvements are of a similar percentage. The NA values in the table are due to the fact that neither persistence nor SARIMA include exogenous data. These gains are significant, because they leave the confidence intervals obtained by bootstrapping (see Methods). Furthermore, the ranking of the models is consistent with recent studies [21, 47]. We remark that the time series bootstrap in the *random forests + bootstrap* improves the performance of the random forest algorithm without mobility data—confirming the results of [48]—but is not the case when adding mobility data. Finally,

Table 1 Benchmark with and without mobility data.

	Without mobility data		With mobility data	
	RMSE (GW)	MAPE (%)	RMSE (GW)	MAPE (%)
<i>Statistical model</i>				
Persistence (1 day)	4.0 ± 0.2	5.5 ± 0.3	N.A.	N.A.
SARIMA	2.4 ± 0.2	3.1 ± 0.2	N.A.	N.A.
GAM	2.3 ± 0.1	3.5 ± 0.2	2.17 ± 0.08	3.3 ± 0.1
<i>Data assimilation technique</i>				
Static Kalman filter	2.1 ± 0.1	3.1 ± 0.2	1.72 ± 0.08	2.5 ± 0.1
Dynamic Kalman filter	1.4 ± 0.1	1.9 ± 0.1	1.20 ± 0.08	1.7 ± 0.1
Viking	1.5 ± 0.1	1.8 ± 0.1	1.24 ± 0.07	1.7 ± 0.1
Aggregation of experts	1.4 ± 0.1	1.8 ± 0.1	1.16 ± 0.07	1.6 ± 0.1
<i>Machine learning</i>				
GAM boosting	2.6 ± 0.2	3.7 ± 0.2	2.4 ± 0.1	3.5 ± 0.2
Random forests	2.5 ± 0.2	3.5 ± 0.2	2.0 ± 0.1	2.7 ± 0.2
Random forests + bootstrap	2.2 ± 0.2	3.0 ± 0.2	2.0 ± 0.1	2.7 ± 0.2

holidays are known to behave differently from regular days [44]. We therefore ran the same benchmark (see appendix) when excluding holidays; the results suggest that

incorporating mobility data still significantly improves forecasting performance (see Table B1). These results all suggest that adding mobility data leads to RMSE and MAPE gains of around 10% when forecasting French electricity demand.

1.4 Explaining the impact of mobility

In this section we use variable selection to offer insight into the performance of forecasts that use mobility data. We also investigate the link between electricity demand and the work feature—which emerges as the second most explanatory variable in our variable ranking (see below).

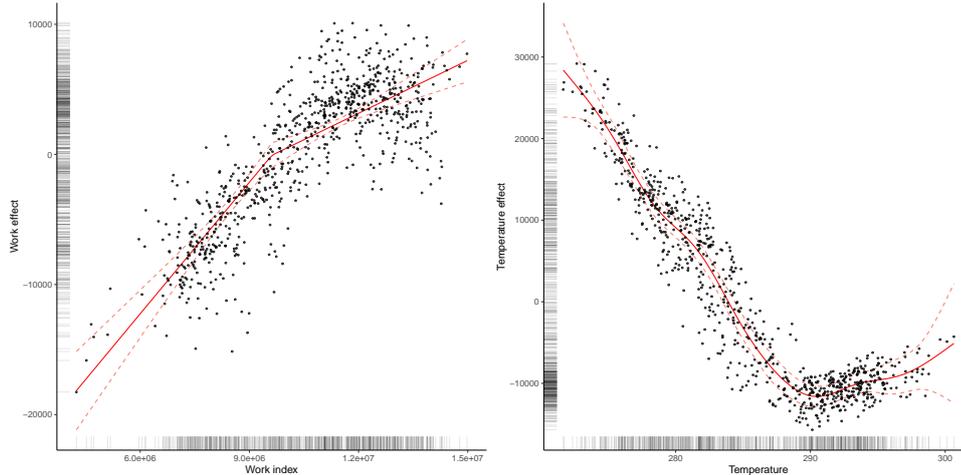
Variable selection

Combining the calendar, meteorological, electricity, and mobile network datasets resulted in 38 features. Some of these features are highly correlated, e.g., see Appendix A.1 for details on the correlation between the *temperature* and the *school holidays* features. Thus, to create highly explainable and robust forecasts, it is necessary to select a smaller number of highly explanatory features, in order to better understand how they relate to electricity demand. Nevertheless, typical variable selection methods based on cross-validation [49–51] are not directly applicable to time series, because the samples are not independent. To reduce the dimensionality of the problem, one solution is to rank the features by order of importance and then select the most important ones [52]. In this paper, we considered three such ranking methods: minimum redundancy maximum relevance (mRMR), Hoeffding D-statistics, and Shapley values. For multivariate time series, feature selection can be performed using the mRMR algorithm, which consists in selecting variables that maximise mutual information with the target [53, 54] (here the electricity load). We used the *mRMRe* package [55] in R for this analysis; the most important variables—in decreasing order of importance—were *temperature*, the *work* index, and the *time of year*. The Hoeffding D-statistic ranking and the Shapley value ranking results are detailed in Appendix D.1. All three rankings gave the same results, implying that the *work* index is more important than the calendar data. We note also that these analyses suggest that the *tourism* and *residents* indices do not appear to be of great importance with respect to France’s electricity demand (See Appendix D.1).

Impact of work dynamics on the electricity demand

Outperforming the state-of-the-art indicates that mobility data had explanatory power when it came to electricity demand during the sobriety period. However, this result does not provide any formal insights into the future performance of the index. We therefore ran a statistical analysis of the predictive ability of this mobility data. Moreover, state-of-the-art data assimilation techniques being difficult to analyse, we restrict ourselves to explainable models of the electricity demand. Since the effect of temperature is known to be nonlinear, we consider generalized additive models (GAMs) instead of standard linear regression. As temperature and then the *work* index were ranked as the two most important variables in our variable selection phase, we consider the electricity demand corrected for the effect of temperature. Figure 3 shows that the electricity demand increases with the work index, i.e., the higher the number of people

Fig. 3 Effects of features on electricity demand.



Left: Temperature-corrected electricity load as a function of the work index. **Right:** Work-index corrected electricity load as a function of temperature. Each black point is an observation at 10 a.m. The effect given by the GAM regression is shown in red on both plots. Dotted red lines correspond to the 95% confidence interval of the effects.

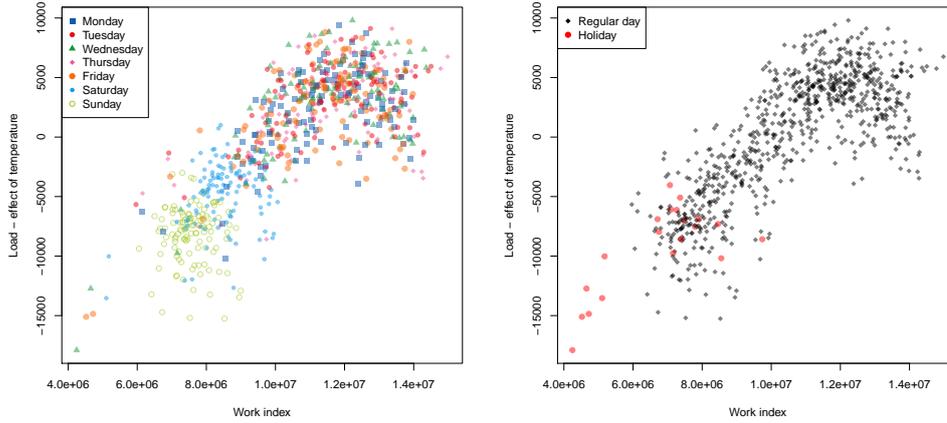
at work, the higher the electricity demand. Moreover, the *work* index accounts for several consumption behaviours. First, Figure 4 (left) shows how the index accounts for the effect of weekends, thus capturing weekly seasonality related to typical work behaviour. Indeed, we see that weekdays, Saturdays (in purple), and Sundays (in yellow) correspond to specific clusters of points with a lower *work* index. Figure 4 (right) shows that the *work* index is related to consumption differences during the holidays (in red) vs other days (in black). Note that they have the same relationship as on regular days. Therefore, the *work* index summarises in a single feature both the effects of the day of the week and the holiday (7 features).

Moreover, the analysis of the impact of our *work* index on electricity demand—when fixing the day of the week and excluding holidays, shows that lower *work* index correspond to lower electricity demand (see Section D.2 in the appendix). This shows that lower work dynamics are associated with energy savings. As expected, the effect is more pronounced during working hours. As a result, the work index is more informative than calendar information alone. In fact, models using the *work* index performed better on the atypical event of the sobriety period than models based on calendar data, which only capture seasonality in stationary signals (see Table D4 in the appendix). This suggests that the *work* index is explanatory of electricity demand.

2 Discussion

In this work, we have shown that the period spanning September 2022 to March 2023 was atypical in terms of France’s electricity demand. During this so-called sobriety period, we observed a decrease in electricity demand similar to what happened during

Fig. 4 Dynamics captured by the work index.



Each point is an observation of the electricity load corrected for temperature as a function of the work index at 10 a.m. between July 2019 and March 2022. **Left:** Dependence of the *work* index on the day of the week. **Right:** The holiday pattern.

the first COVID-19 lockdown. However, this period of significant electricity savings lasted for over six months, which is much longer than the 1-month COVID-related period. These observations are consistent with those of French media and France’s transmission system operators. These results suggest that additional phenomena to annual seasonality and temperature are responsible for the recent significant changes in the electricity consumption behaviour.

To better understand this collective energy-saving behaviour, we have introduced here mobility indicators from mobile network data. This is an original new way of tracking human mobility, which can help better characterise electricity demand. Indeed, the vibrancy of places varies over years, but also over the course of a day, which implies changes in the organisation and planning of such areas in terms of services, such as electricity demand, and more generally in terms of public policy. For instance, following the COVID-19 lockdowns, the tendency towards remote work was shown to have a significant impact on electricity demand in the 2019-2020 period [56, 57]. Indeed, the changes in work organisation catalysed by COVID-19 have the potential to reshape many aspects of modern societies and economies [58]. Understanding the spatial practices of human populations is therefore fundamental to operational decisions and research in many fields. However, individuals circulate through a number of places throughout the day—on average between 2.5 and 4 per person in French metropolitan areas [59]—month, and year, whether for housing, work, education, personal relationships, and leisure. Determining the appropriate index to measure and its level are both major difficulties in analysing mobility dynamics [60]. In France, for example, the collection of statistics on remote working has traditionally been limited to ad hoc flash surveys. Only from 2021 will these questions be included in the Continuous Employment Survey (the main reference statistical survey on the labour market) and from

2022 in the census, but with less frequent sampling compared to mobile phone data [60]. One response has been the processing of digital traces, in particular data from mobile networks. The quality of these data has evolved with both the development of network technology and the ability to account for intrinsic biases, such as selection (the individuals in a dataset may not be sampled at random within the population of interest) and overgeneralization (small datasets may not reflect well low-probability events). Our high quality dataset was designed to precisely quantify human presence over France at a very high frequency with respect to census or survey data, and has already been studied as such to account for residential behaviour [61]. Its advantage is that it allows one not only to quantify with a high degree of accuracy the population present at a given time and place, but also to characterise the way they inhabit that place (i.e., residing, working, or exploring). Here, we have shown that these indices are highly correlated with other public datasets, and that they also can account for human presence dynamics related to tourism and work (see also Appendix A.2).

As evidenced by our benchmark in Table B1, standard statistical models such as GAM struggled during the sobriety period. Indeed, for the same state-of-the-art GAM, the RMSE and the MAPE when excluding holidays were respectively 55% and 87% higher than in the same test period two years earlier (September 2019 to March 2020) [21]. Relying on a benchmark specific to France electricity load forecasting, we have shown that including mobile network mobility data in the analyses improves the state-of-the-art performance by around 10%. Although evaluating the cost of load forecasting error is a difficult task, it has been estimated that a 1% reduction in load forecasting error would save an energy provider up to several hundred thousand USD per GW peak [62]. In 2022, the average daily load peak in France was 58 GW. According to this estimation, the gain of 0.2% in MAPE in Table 1 resulting from exploiting mobility data would have amounted to tens of millions of USD per year at the national level.

In addition, we have shown that the *work* index accounts for several consumption behaviours, including the impact of weekends and holidays on the electricity demand. Remark that these dynamics are not specific to the sobriety period, which suggests that the benefits of using mobility data would generalise to the post-crisis period. Overall, the higher the *work* index, the higher the electricity demand.

Future lines of research include studying the work index at a 1-hour frequency, over longer periods, and at the finer geographical scale of French administrative regions. Indeed, as shown in Appendix A.1, mobile network data effectively capture human spatial dynamics other than those related to work, such as residence and tourism. Although in this paper we have not found a significant effect of such behaviours on national electricity demand, it might become visible when working at the regional level. Although we have shown that a reduction in the *work* index corresponds to a reduction in the electricity demand, further studies are needed to disentangle the effect of economic growth, employment rate, and remote working in this phenomenon. Moreover, we have focused in this work on mean forecast performance, i.e., on the ability of the forecast to predict the expected value of the electricity demand. Another interesting subject would be to evaluate the variance in the electricity demand given the *work* index, which would be helpful for practitioners when acting on the electricity

market. Finally, in practice, it currently requires several days to clean, aggregate, and adjust the indices. For operational use, further studies are therefore needed to quantify the impact of such a delay in the use of the *work* index on the performance of benchmark forecasts, or conversely, to study the predictive capabilities of the *work* index.

3 Methods

3.1 Open-source calendar, meteorological, and electricity datasets

The reference dataset runs from 08/01/2013 to 28/02/2023. It consists of calendar data (dates and holidays), meteorological data (temperature), and historical data (electricity power load at different time scales). All these data are public and distributed under the Etalab open source licence. The calendar data are extracted from the French open source database [63, 64]. This regroups holiday periods according to France’s three holiday timetables—in France holidays depend on the region you live in—as well as the French national holidays. This calendar dataset has no missing values. The meteorological data are extracted from the SYNOP Météo-France database [65]. Météo-France is the French public agency responsible for the national weather and climate service. The dataset consists of 3-hourly temperature measurements from 62 meteorological stations located throughout French territory. This dataset has many missing values, which we have imputed as follows. First, if a station has a missing value at time t and the station’s measurements are available 3 hours before and 3 hours after t , the missing value is imputed as the mean of these two measurements. If no such values are available, the missing temperature is imputed as the temperature of the nearest station. If however all stations in a region have missing values, the temperature of each station is imputed by taking the mean of the temperature at the same hour from the day before and the day after. Finally, the historical electricity load dataset is extracted from the RTE’s public releases [66]. RTE (Réseau de Transport d’Electricité) is France’s transmission system operator. It provides high quality data on regional electricity consumption in France with a frequency of 30 minutes. The national electricity load has no missing values, which is valuable since this is the final target throughout this article.

3.2 Mobility dataset

The reference dataset is complemented by mobility indices. These mobile phone data were provided by Orange’s Flux Vision business service [38], in the form of daily presence data reports. These include the number of visitors in the 101 geographical areas of mainland France, which correspond to the second level of national administrative divisions. For each location and each day, the data are stratified by the type of visitor (resident, usually present, tourist, excursionist, recurrent excursionist) and origin (foreign, local, non-local). The mobile phone data were anonymised in compliance with strict privacy requirements and audited by the French data protection authority (Commission Nationale de l’Informatique et des Libertés). Computation of

the presence data reports is based on the on-the-fly processing of signalling messages exchanged between mobile phones and the mobile network, usually collected by mobile network operators to monitor and optimise mobile network activity. Such messages contain information about the identifiers of the mobile subscriber and of the antenna handling the communication, the timestamp, and the type of event (e.g., voice call, SMS, handover, data connection, location update). Knowing the location of antennas makes it possible to reconstruct the approximate position of a communication device. All these data were then used to compute the total number of individuals in a given area, without saving any residual information that could be traced back to the individual users. More specifically, at any given day, each individual was characterised based on their pattern of movement and their origin as follows.

- Resident: person whose spends much of their time in the study area, and spent at least 22 nights (not necessarily consecutive) there over the past eight weeks.
- Usually present: person who is not a resident of the study area but has been seen in the study area repeatedly: more than four times in different weeks during the previous eight weeks.
- Tourist: person who spends the night in the study area who is neither resident nor usually present.
- Excursionist: person not staying overnight the night before and the night of the study day, and present less than 5 times during the day in the last 15 days.
- Recurrent excursionist: person who has not spent the night before and the current night in the study area and who has been present more than five times during the day in the previous 15 days.

The night corresponding to a given day is the period between 8 p.m. of that day and 8 a.m. of the following day. Moreover, origin is categorised as follows:

- Foreign: person with a foreign SIM card.
- Local: person with a billing address in the study area.
- Non-local: person with a billing address outside the study area.

This data was then corrected by Orange Flux Vision to account for spatial and temporal biases, and so to be representative of the general population. To this end, they use spatially-stratified market share data, socio-economic data from the national statistics institute (Insee), mobile phone ownership data also from Insee, and customer socio-demographic information provided upon subscription. From these data, we constructed three indices. The *work* index corresponds to the number of recurrent excursionists, the *tourism* index to the number of foreign plus non-local tourists, and the *resident* index to the number of residents plus “usually presents”. In this article, the mobility dataset covered the periods from 01/07/2019 to 01/03/2020, from 01/07/2020 to 01/03/2021, from 01/07/2021 to 01/03/2022, and from 01/07/2022 to 01/03/2023.

3.3 Benchmark models

Models were evaluated according to the following test errors. Let T_{test} be the test period, $(y_t)_{t \in T_{test}}$ the target, and $(\hat{y}_t)_{t \in T_{test}}$ an estimator of y . The root mean square error is defined by $\text{RMSE}(y, \hat{y}) = \left(\frac{1}{T_{test}} \sum_{t \in T_{test}} (y_t - \hat{y}_t)^2 \right)^{1/2}$ and the mean absolute

percentage error is defined by $\text{MAPE}(y, \hat{y}) = \frac{1}{T_{test}} \sum_{t \in T_{test}} \frac{|y_t - \hat{y}_t|}{|y_t|}$. Both errors are useful for operational uses. Since the sampling of time series are dependent, confidence intervals were obtained by time series bootstrapping [67] using the `tseries` package [68] forecasting; the GAM model was extracted from [21], the static and dynamic Kalman filters were inspired by [47], the Viking algorithm comes from [69], the GAM boosting parameters are from [70], and the random forest and random forest with bootstrap parameters were taken from [48]. A full description of these models can be found in the appendices.

4 Data and code availability

The source code detailing how to create the electricity dataset and all of the models used in this article is available at https://github.com/NathanDoumeche/Mobility_data_assimilation. The corresponding dataset is available on Zenodo at <https://zenodo.org/records/10041368>. Hence the change point results shown in Figure 1, as well as the dataset and the benchmarks without mobility data of Table 1, are directly reproducible for future research, and can easily be updated to work for new time periods of interest. Note however that the mobility indices are not publicly available.

Acknowledgments. This work was funded by Sorbonne University, EDF R&D, and Orange Innovation Research. The authors bear full responsibility for its findings and conclusions. We would like to thank Kevin Bleakley for his thorough proofreading of this article.

References

- [1] Shove, E. & Walker, G. What is energy for? *Social practice and energy demand. Theory, culture & society* **31**, 41–58 (2014).
- [2] Ferriani, F. & Gazzani, A. The impact of the war in Ukraine on energy prices: consequences for firms’ financial performance. *International Economics* **174**, 221–230 (2023).
- [3] Ruhnau, O., Stiewe, C., Muessel, J. & Hirth, L. Natural gas savings in Germany during the 2022 energy crisis. *Nature energy* (2023).
- [4] European Commission. Report from the Commission to the European Parliament and the Council on the review of emergency interventions to address high energy prices in accordance with Council Regulation (EU) 2022/1854 (05/06/2023). URL https://energy.ec.europa.eu/system/files/2023-06/COM_2023_302_1_EN_ACT_part1_v2.pdf.
- [5] RTE. Réactualisation des perspectives pour le système électrique pour l’hiver 2022-2023 (février 2023). Tech. Rep., Réseau de Transport d’Électricité (2023). URL <https://assets.rte-france.com/prod/public/2023-02/2023-02-24-reactualisation-hiver.pdf>.

- [6] French government. Plan de sobriété énergétique. Press release (10/06/2022). URL <https://www.ecologie.gouv.fr/sites/default/files/dp-plan-sobriete.pdf>.
- [7] Techniques de l'ingénieur. Consommation électrique : le plan de sobriété fournit ses premiers effets. Available: <https://www.techniques-ingenieur.fr/actualite/articles/consommation-electrique-le-plan-de-sobriete-fournit-ses-premiers-effets-117348/> (2022). [Accessed: May 11, 2023].
- [8] The New York Times. As Russia chokes Europe's gas, France enters era of energy sobriety. Available: <https://www.nytimes.com/2022/09/05/business/russia-gas-europe-france.html> (2022). [Accessed: May 11, 2023].
- [9] Le Monde. Électricité : la baisse inattendue de la consommation a permis d'éviter les coupures cet hiver. Available: https://www.lemonde.fr/planete/article/2023/03/16/electricite-la-baisse-inattendue-de-la-consommation-a-permis-d-eviter-les-coupures-cet-hiver_6165773_3244.html (2023). [Accessed: May 11, 2023].
- [10] Abdelaziz, E., Saidur, R. & Mekhilef, S. A review on energy saving strategies in industrial sector. Renewable and Sustainable Energy Reviews **15**, 150–168 (2011).
- [11] Rockström, J. et al. A roadmap for rapid decarbonization. Science **355**, 1269–1271 (2017).
- [12] Hoegh-Guldberg, O. et al. The human imperative of stabilizing global climate change at 1.5°C. Science **365**, eaaw6974 (2019).
- [13] de Maere d'Aertrycke, G., Smeers, Y., de Peufeilhoux, H. & Lucille, P.-L. The role of electrification in the decarbonization of central-western Europe. Energies **13** (2020).
- [14] Omar, H., Graetz, G. & Ho, M. Decarbonizing with Nuclear Power, Current Builds, and Future Trends, Ch. 6, 103–151 (John Wiley & Sons, Ltd, 2022).
- [15] RTE. Energy pathways to 2050. Tech. Rep., Réseau de Transport d'Électricité (2022). URL https://assets.rte-france.com/prod/public/2022-01/Energy%20pathways%202050_Key%20results.pdf.
- [16] Rolnick, D. et al. Tackling climate change with machine learning. ACM Computing Surveys (CSUR) **55**, 1–96 (2022).
- [17] Pinheiro, M., Madeira, S. & Francisco, A. Short-term electricity load forecasting—a systematic approach from system level to secondary substations. Applied Energy **332**, 120493 (2023).

- [18] Hong, T. et al. Energy forecasting: A review and outlook. IEEE Open Access Journal of Power and Energy **7**, 376–388 (2020).
- [19] Nti, I., Teimeh, M., Nyarko-Boateng, O. & Adekoya, A. Electricity load forecasting: a systematic review. Journal of Electrical Systems and Information Technology **7**, 2314–7172 (2020).
- [20] Hammad, M., Jereb, B., Rosi, B. & Dragan, D. Methods and models for electric load forecasting: a comprehensive review. Logistics, Supply Chain, Sustainability and Global Challenges **11**, 51–76 (2020).
- [21] Obst, D., de Vilmarrest, J. & Goude, Y. Adaptive methods for short-term electricity load forecasting during covid-19 lockdown in France. IEEE Transactions on Power Systems **36**, 4754–4763 (2021).
- [22] Blondel, V., Decuyper, A. & Krings, G. Understanding vehicular routing behavior with location-based service data. EPJ Data Science **4** (2015).
- [23] Deville, P. et al. Dynamic population mapping using mobile phone data. Proceedings of the National Academy of Sciences of the United States of America **111**, 15888–15893 (2014).
- [24] Blumenstock, J., Cadamuro, G. & On, R. Predicting poverty and wealth from mobile phone metadata. Science **350**, 1073–1076 (2015).
- [25] Lorenzo, G. et al. AllAboard: Visual exploration of cellphone mobility data to optimise public transport. IEEE Transactions on Visualization and Computer Graphics **22** (2016).
- [26] Bengtsson, L., Lu, X., Thorson, A., Garfield, R. & Schreeb, J. Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: a post-earthquake geospatial study in Haiti. PLoS Medicine **8**, e100–1083 (2011).
- [27] Blumenstock, J. Inferring patterns of internal migration from mobile phone call records: evidence from Rwanda. Information Technology for Development **18**, 107–125 (2012).
- [28] Rubrichi, S., Smoreda, Z. & Musolesi, M. A comparison of spatial-based targeted disease mitigation strategies using mobile phone data. EPJ Data Science **7** (2018).
- [29] Pullano, G., Valdano, E., Scarpa, N., Rubrichi, S. & Colizza, V. Evaluating the effect of demographic factors, socioeconomic factors, and risk aversion on mobility during the covid-19 epidemic in France under lockdown: a population-based study. Lancet Digit Health **2**, e638–e649 (2020).

- [30] Xu, Y., Clemente, R. & González, M. Understanding vehicular routing behavior with location-based service data. EPJ Data Science **10**, 1–17 (2021).
- [31] Filazzola, A. et al. Using smartphone-gps data to quantify human activity in green spaces. PLOS Computational Biology **18**, 1–20 (2022).
- [32] Terroso-Sáenz, F., Muñoz, A., Fernández-Pedaue, J. & Cecilia, J. Human mobility prediction with region-based flows and water consumption. IEEE Access **9**, 88651–88663 (2021).
- [33] Smolak, K. et al. Applying human mobility and water consumption data for short-term water demand forecasting using classical and machine learning models. Urban Water Journal **17**, 32–42 (2020).
- [34] Chen, Y., Yang, W. & Zhang, B. Using mobility for electrical load forecasting during the covid-19 pandemic. arXiv:2006.08826 (2020).
- [35] Ruan, G. et al. A cross-domain approach to analyzing the short-run impact of covid-19 on the U.S. electricity sector. Joule **4**, 2322–2337 (2020).
- [36] Zarbakhsh, N., Misaghian, M. & Mcardle, G. Human mobility-based features to analyse the impact of covid-19 on power system operation of ireland. IEEE Open Access Journal of Power and Energy **9**, 213–225 (2022).
- [37] Antoniadis, A., Gaucher, S. & Goude, Y. Hierarchical transfer learning with applications to electricity load forecasting. International Journal of Forecasting (2023).
- [38] Business, O. Flux vision. URL <https://www.orange-business.com/en/solutions/data-intelligence-iot/flux-vision>.
- [39] Killick, R. & Eckley, I. Changepoint: an R package for changepoint analysis. Journal of Statistical Software **58**, 1–19 (2014).
- [40] Aminikhanghahi, S. & Cook, D. A survey of methods for time series change point detection. Knowledge and Information Systems **51**, 339–367 (2017).
- [41] RTE. Winter 2022-2023. Tech. Rep., Réseau de Transport d'Électricité (2023). URL <https://www.rte-france.com/actualites/bilan-hiver-2022-2023-coupures-electricite-evitees-grace-baisse-consommation>.
- [42] The Economist. The global normalcy index. Available: https://www.economist.com/graphic-detail/tracking-the-return-to-normalcy-after-covid-19?utm_medium=pr&utm_source=inf-a (2023). [Accessed: July 11, 2023].
- [43] LLC, G. Google covid-19 community mobility reports. URL <https://www.google.com/covid19/mobility/>. [Accessed: Sept 21, 2023].

- [44] Krstonijević, S. Adaptive load forecasting methodology based on generalized additive model with automatic variable selection. Sensors (Basel) (2022).
- [45] Singh, A., Ibraheem, Khatoon, S., Muazzam, M. & Chaturvedi, D. Load forecasting techniques and methodologies: a review. 2012 2nd International Conference on Power, Control and Embedded Systems 1–10 (2012).
- [46] Wang, Z. et al. Benchmarks and custom package for electrical load forecasting. arXiv:2307.07191 (2023).
- [47] Vilmarest, J. & Goude, Y. State-space models for online post-covid electricity load forecasting competition. IEEE Open Access Journal of Power and Energy **9**, 192–201 (2022).
- [48] Goehry, B., Yan, H., Goude, Y., Massart, P. & Poggi, J.-M. Random forests for time series. REVSTAT-Statistical Journal **21**, 283–302 (2023).
- [49] Wasserman, L. & Roeder, K. High-dimensional variable selection. The Annals of Statistics **37**, 2178–2201 (2009).
- [50] Huang, J., Horowitz, J. & Wei, F. Variable selection in nonparametric additive models. Annals of statistics **38**, 2282–2313 (2010).
- [51] Marra, G. & Wood, S. Practical variable selection for generalized additive models. Computational Statistics & Data Analysis **55**, 2372–2387 (2011).
- [52] Genuer, R., Poggi, J.-M. & Tuleau-Malot, C. Variable selection using random forests. Pattern Recognition Letters **31**, 2225–2236 (2010).
- [53] Peng, H., Long, F. & Ding, C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. IEEE Transactions on Pattern Analysis and Machine Intelligence **27**, 1226–1238 (2005).
- [54] Han, M., Ren, W. & Liu, X. Joint mutual information-based input variable selection for multivariate time series modeling. Engineering Applications of Artificial Intelligence **37**, 250–257 (2015).
- [55] Jay, N. D. et al. mRMRe: an R package for parallelized mRMR ensemble feature selection. Bioinformatics **29**, 2365–2368 (2013).
- [56] Abu-Rayash, A. & Dincer, I. Analysis of the electricity demand trends amidst the covid-19 coronavirus pandemic. Energy Research & Social Science **68**, 101682 (2020).
- [57] Ku, A., Qiu, Y., Lou, J., Nock, D. & Xing, B. Changes in hourly electricity consumption under covid mandates: a glance to future hourly residential power

- consumption pattern with remote work in Arizona. *Applied Energy* **310**, 118539 (2022).
- [58] Vyas, L. New normal at work in a post-covid world: work–life balance and labor markets. *Policy and Society* **41**, 155–167 (2022).
- [59] MTECT. Enquête mobilité des personnes 2018-2019. Available: <https://www.statistiques.developpement-durable.gouv.fr/resultats-detailles-de-lenquete-mobilite-des-personnes-de-2019>. [Accessed: July 11, 2023].
- [60] Pora, P. Telework and productivity three years after the start of the pandemic. *Economie et Statistique / Economics and Statistics* **593** (2023).
- [61] Lévy, J., Piantoni, J. C. S. & François, J. Who lives where? Counting, locating, and observing France’s real inhabitants. *SocArXiv* (2023).
- [62] Hong, T. & Fan, S. Probabilistic electric load forecasting: a tutorial review. *International Journal of Forecasting* **32**, 914–938 (2016).
- [63] Etalab. Jours fériés en France. Available: <https://www.data.gouv.fr/fr/datasets/jours-feries-en-france/> (2023). [Accessed: May 11, 2023].
- [64] Etalab. Vacances scolaires par zones. Available: <https://www.data.gouv.fr/fr/datasets/vacances-scolaires-par-zones/> (2023). [Accessed: May 11, 2023].
- [65] Météo-France. Données SYNOP essentielles OMM. Available: <https://public.opendatasoft.com/explore/dataset/donnees-synop-essentielles-omm/> (2023). [Accessed: May 11, 2023].
- [66] RTE. éco2mix. Available: <https://www.rte-france.com/en/eco2mix/download-indicators> (2023). [Accessed: May 11, 2023].
- [67] Lahiri, S. *Resampling methods for dependent data* 1 edn. Springer Series in Statistics (Springer New York, NY, 2003).
- [68] Trapletti, A., Hornik, K. & LeBaron, B. tseries: Time series analysis and computational finance. URL: <https://cran.r-project.org/web/packages/tseries/index.html> **0.10-54** (2023).
- [69] Vilmarest, J., Browell, J., Fasiolo, M., Goude, Y. & Wintenberger, O. Adaptive probabilistic forecasting of electricity (net-)load. *arXiv:2301.10090* (2023).
- [70] Taieb, S. B. & Hyndman, R. A gradient boosting approach to the kaggle load forecasting competition. *International Journal of Forecasting* **30**, 382–394 (2014).
- [71] F. Michailesco, O. S. Migrations résidentielles post-covid : l’attractivité du périurbain légèrement renforcée. Tech. Rep., Insee (2023). [Accessed: Sept 21,

- 2023].
- [72] Bakhat, M. & Rosselló, J. Estimation of tourism-induced electricity consumption: The case study of balearics islands, spain. Energy Economics **33**, 437–444 (2011).
 - [73] Lai, T., To, W., Lo, W., Choy, Y. & Lam, K. The causal relationship between electricity consumption and economic growth in a gaming and tourism center: the case of Macao SAR, the People’s Republic of China. Energy **36**, 1134–1142 (2011).
 - [74] INSEE. Arrivées dans l’hôtellerie - Total - France métropolitaine. Available: <https://www.insee.fr/fr/statistiques/serie/010598571> (2023). ID 010598571, [Accessed: May 11, 2023].
 - [75] Emmanuel, T. et al. A survey on missing data in machine learning. Journal of Big Data **8** (2021).
 - [76] Shukla, S. & Marlin, B. A survey on principles, models and methods for learning from irregularly sampled time series (2021).
 - [77] Little, R. & Rubin, D. Statistical analysis with missing data Vol. 793 (John Wiley & Sons, 2019).
 - [78] Ma, J. et al. Transfer learning for long-interval consecutive missing values imputation without external features in air pollution time series. Advanced Engineering Informatics **44**, 101092 (2020).
 - [79] Zhang, Y. et al. Missing value imputation in multivariate time series with end-to-end generative adversarial networks. Information Sciences **551**, 67–82 (2021).
 - [80] Ayme, A., Boyer, C., Dieuleveut, A. & Scornet, E. Naive imputation implicitly regularizes high-dimensional linear models. arXiv:2301.13585 (2023).
 - [81] Box, G. & Jenkins, G. M. Time Series Analysis: Forecasting and Control (Holden-Day, 1976).
 - [82] Wood, S. N. Generalized additive models: an introduction with R (CRC press, 2017).
 - [83] Wood, S. & Wood, M. S. Package ‘mgcv’. R package version **1**, 29 (2015).
 - [84] Hand, D. J. Forecasting with exponential smoothing: The state space approach by rob j. hyndman, anne b. koehler, j. keith ord, ralph d. snyder. International Statistical Review **77**, 315–316 (2008).
 - [85] Kalman, R. E. A new approach to linear filtering and prediction problems. Journal of Basic Engineering **82**, 35–45 (1960).

- [86] Cesa-Bianchi, N. & Lugosi, G. Prediction, Learning, and Games (Cambridge University Press, New York, NY, USA, 2006).
- [87] Gaillard, P., Stoltz, G. & Erven, T. V. A second-order bound with excess losses. Conference on Learning Theory 176–196 (2014).
- [88] Gaillard, P. & Goude, Y. opera: Online prediction by expert aggregation. URL: <https://CRAN.R-project.org/package=opera> r package version 1 (2016).
- [89] Breiman, L. Random forests. Machine learning **45**, 5–32 (2001).
- [90] Breiman, L., Friedman, J., Olshen, R. & Stone, C. Classification and regression trees (Chapman & Hall/CRC, 1984).
- [91] Breiman, L. Arcing the edge. Tech. Rep., Citeseer (1997).
- [92] Friedman, J. H. Greedy function approximation: a gradient boosting machine. Annals of statistics 1189–1232 (2001).
- [93] Grinsztajn, L., Oyallon, E. & Varoquaux, G. Why do tree-based models still outperform deep learning on typical tabular data? Advances in Neural Information Processing Systems **35**, 507–520 (2022).
- [94] Makridakis, S., Spiliotis, E. & Assimakopoulos, V. M5 accuracy competition: Results, findings, and conclusions. International Journal of Forecasting **38**, 1346–1364 (2022).
- [95] Bühlmann, P. & Hothorn, T. Boosting Algorithms: Regularization, Prediction and Model Fitting. Statistical Science **22**, 477 – 505 (2007).
- [96] Hoeffding, W. A non-parametric test of independence. Annals of Statistics **19**, 293–325 (1948).
- [97] Bénéard, C., Biau, G., Veiga, S. D. & Scornet, E. Camps-Valls, G., Ruiz, F. J. R. & Valera, I. (eds) Shaff: Fast and consistent shapley effect estimates via random forests. (eds Camps-Valls, G., Ruiz, F. J. R. & Valera, I.) Proceedings of The 25th International Conference on Artificial Intelligence and Statistics, Vol. 151 of Proceedings of Machine Learning Research, 5563–5582 (PMLR, 2022).

Appendix A Datasets and features

In this appendix, we provide further insight into our exploratory analysis of the mobility dataset. We show how the indices also capture holiday dynamics at the regional level by comparing the mobile network-based *tourism* index with official tourism statistics from Insee, and by studying the temporal evolution of the *work* index.

A.1 Regional human presence indices

Although for the purpose of national forecasts we have only relied on national-level indices, mobile network data were also available at the regional level, which helped to better understand the data at hand. In order to obtain a preliminary understanding of the data, we computed Pearson’s product-moment correlation coefficient r between the human presence variables on the one hand, and the calendar and meteorological data on the other, for the regions of mainland France. This analysis confirmed that our indices matched several well-known human spatial dynamics. In large urban regions such as Île-de-France (IDF) we observed a negative correlation between the *residence* index and both of the calendar variables *school holidays* and *summer holidays* ($r = -0.65$ and $r = -0.84$, respectively), as well as *temperature* ($r = -0.70$). This captures how IDF residents leave their region during the holidays and then behave as tourists. Consistently, in regions that are traditionally popular holiday destinations—such as the coastal region of Provence-Alpes-Côte d’Azur (PACA), the *tourism index* variable was positively correlated with the calendar variables *school holidays* and *summer holidays* ($r = 0.58$ and $r = 0.86$, respectively) and the meteorological variable *temperature* ($r = 0.82$). The *work* index behaved similarly in both regions, with a negative correlation with the *weekly holiday* calendar variable ($r = -0.54$ in IDF and $r = -0.55$ in PACA). To better characterise the seasonal changes of the indices, we show the evolution of the daily *tourism*, *residence*, and *work* indices in IDF (Figure A1—top) and PACA (Figure A1—bottom). In line with the Pearson correlation, in a region with a high level of economic activity such as IDF, the *residence* and *work* indices tended to increase during off-peak periods and to decrease during holidays. We observed the opposite for the *tourism* index in PACA, which is a very tourist-oriented region. Moreover, unlike in IDF, the *work* index in PACA did not decrease significantly during the summer holidays. This might be explained by the different make-up of the respective labour markets, with a high proportion of tourism workers in PACA.

We can also clearly see the effects of the COVID-19 health crisis. In IDF, for example, the *tourism* and *work* indices significantly dropped during the crisis. These then gradually increased during the post-COVID period, but without reaching pre-COVID levels. This was especially pronounced for the *work* index, probably because of changes in work organisation triggered by the health crisis, and also as an effect of the energy crisis. In PACA, on the other hand, we observed a lower impact on tourism, partly due to tourist origin (there are more local tourists, i.e., who do not cross regional borders) than in IDF. Of note, the *residence* index seems to have gradually increased since the end of the COVID-19 crisis in PACA. This phenomenon of migration to certain regions of France has been documented by Insee in the report [71], but deserves a more in-depth analysis.

Fig. A1 Regional indices.



7-day rolling average of mobility indices for the Île-de-France (top) and Provence-Alpes-Côte d’Azur (bottom) regions. Indices have been standardised, i.e., the empirical means have been subtracted and the result divided by the empirical standard deviations. The mobile network dataset only covers the period from July to March in each 12-month period. Shaded areas correspond to regional school holidays, and horizontal grey lines mark the three main COVID-19 lockdowns in France.

A.2 The tourism index from mobile-phone data

The evaluated number of tourists and residents has been shown to be correlated with electricity demand in highly touristic areas [72, 73]. For this reason, we created and studied a *tourism* index at the national level. Traditionally, most similar assessments have been carried out on a monthly or annual basis. One strength of our mobile phone-based *tourism* index is that it can be calculated at finer temporal and geographical scales. To further assess the *tourism* index’s performance as a proxy for tourism activity, we compared its monthly average with the Insee tourism index [74], as shown in Figure A2. We obtained an 87% correlation between the two signals, showing that the tourism index efficiently captures tourism trends. However, our study found that tourism had no significant impact on French electricity demand (see Appendix D.1).

Fig. A2 Comparison of the Insee and the mobile network tourism indices.

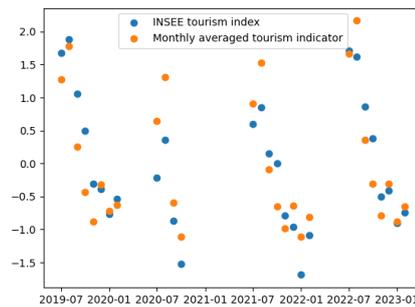
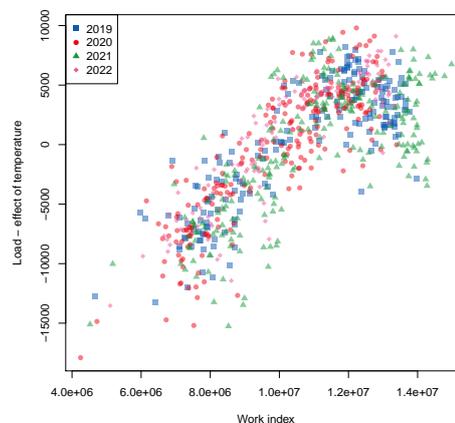


Fig. A3 Residuals as a function of the *work* index over the years 2019–2022.

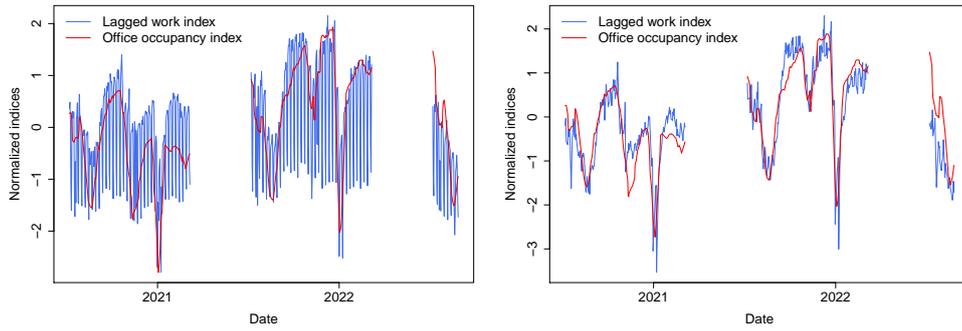


Each point is an observation between July 2019 and March 2022.

A.3 *Work* index and calendar features

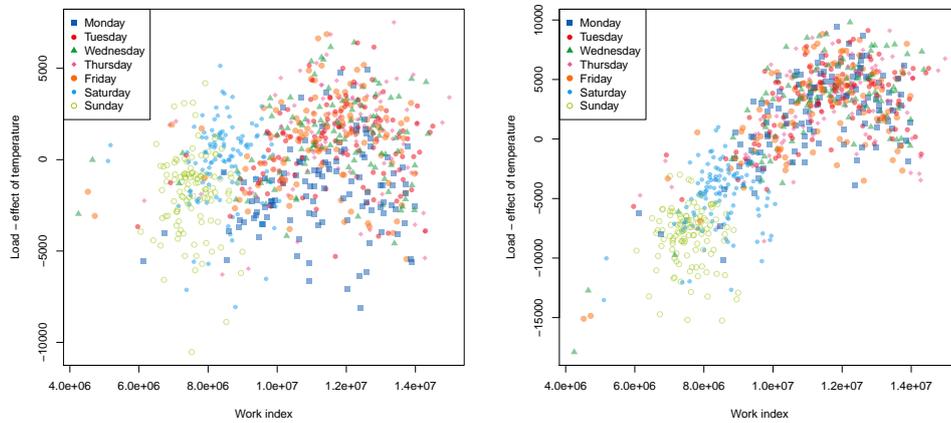
As explained in the introduction, several phenomena occurred between 2020 and 2023 that significantly changed human behaviour and affected French electricity demand. To better understand the impact of the *work* index on electricity demand, it was therefore important to see whether this dependence has changed over time. Figure A3 shows that the dependence of electricity demand in the *work* index has been stationary over the years 2019–2022. This shows that this relationship is robust to the aforementioned events, which is an argument to believe that the results of this article will generalise well to future periods of interest. In addition, as shown in Figure A4, unlike The Economist’s office occupancy index, our *work* index from mobile data captures the reduction in work activity due to weekends and holidays. Furthermore, as expected, Figure A5 shows that the *work* index is only useful for electricity demand forecasting during typical work hours. Indeed, the electricity demand corrected for the

Fig. A4 Comparison of work indices.



Comparison of the 7-day lagged mobile phone based index and the normalcy office occupancy index on all days (left), and when excluding weekends and holidays (right).

Fig. A5 Electricity demand corrected for temperature as a function of the *work* index for each day of the week.



Each point is an observation between July 2019 and March 2022. **Left:** 2 a.m. **Right:** 10 a.m.

temperature effect had a significant dependence in the index at 10 a.m., but not at 2 a.m. See Appendix D.2 for more details.

Appendix B Benchmark and models

In this appendix, we detail the framework and the models in Table 1.

B.1 Handling missing values in mobile network data

There are two types of missing data in our datasets. First, the datasets are regularly sampled time series but with different frequencies. Indeed, recall that the calendar and the electricity datasets have 30-minute frequencies, while the meteorological dataset has a 3-hour one, while the mobile phone dataset has a 1-day frequency. A common method to deal with differences in sampling frequency is to impute the missing value by interpolation [75]. The interpolation method for meteorological data is described in the Methods section of the main paper, while the Orange indices are considered constant throughout the day.

Second, the mobile network dataset only covers the periods from 01/07/2019 to 01/03/2020, 01/07/2020 to 01/03/2021, 01/07/2021 to 01/03/2022, and 01/07/2022 to 01/03/2023. Though various techniques have been developed to tackle sampling irregularities in time series [see, e.g., 76], dealing with large sets of consecutive missing values is still very challenging. The three main approaches when studying time series with consecutive missing values are deletion, imputation, and imputation with masking [75]. First, deletion consists of discarding any observation with at least one missing value. Though this is the simplest way to deal with missing values, it can introduce a bias if the missing data are not-at-random, i.e., if the missing data are actually informative with respect to the target [77]. Second, in regression, imputation techniques aim to “fill in” missing values. The state-of-the-art in time series imputation is wide-ranging and an active field of research [see, e.g., 78]. Note that imputation that maximises a regression model’s performance is not necessarily that which reconstructs missing values most accurately [79, 80]. This makes it more difficult to understand and explain the true effect of imputed features on a target variable. Third, imputation with masking consists of imputing the missing values and keeping track of which observations have been imputed by adding a new feature equal to 1 if the observation comes from an actual measurement, and 0 if it was imputed. In this paper, the pattern of missing data is regular, spanning each year from March to July, and does not depend on the explanatory variables (temperature, *work* index, etc.). Thus, to simplify the analysis, we chose the deletion framework and have not tried to impute the missing values of the mobile network indices.

Furthermore, Table 1 not only shows that mobile phone indices help to improve the performance of state-of-the-art forecasting algorithms, but also attests that this is still true even when comparing the complete open dataset with the incomplete mobile phone dataset. Indeed, on the one hand, models “without mobility data” were trained on the complete calendar, weather, and electricity datasets, spanning from 08/01/2013 to 01/09/2022. On the other hand, models “with mobility data” were created in a two-step process using the transfer learning framework presented in [37]. First, a model trained without mobility data from 08/01/2013 to 01/09/2022 provided an estimate \hat{Load} of the electricity demand $Load$. Then, another model was trained in the deletion framework to forecast the error $err = Load - \hat{Load}$, also known as the residual, using the mobile phone dataset. This second forecast is denoted by $e\hat{r}$. The final forecast was therefore the sum of the two forecasts $\hat{Load} + e\hat{r}$. Notice that this framework gives an advantage to the reference forecast “without mobility data”. In fact, the gains from using mobile phone data are much higher if the training periods of all models

are restricted to the period for which mobile phone data are available (although we have not included these results in the paper for the sake of simplicity). However, the framework we chose allowed us to assess the interest of using mobile phone data from an operational point of view. It ensured that the best models trained using the mobile phone dataset outperformed the best models trained on the full, open datasets. Therefore, the gains of 10% we obtained are likely to be much higher if we had access to a more complete mobile phone dataset. We chose this residuals method to account for the mobile phone data because it gave better results than directly training models “with mobility data” on all datasets restricted to the period for which the mobile phone data are available (once again, we have not included these results in the paper for simplicity).

B.2 Statistical models

B.2.1 Time series models

Persistence models are the simplest type of model for time series. They consist of estimating the target with its own lags and are common baselines in time series benchmarks because of their simplicity, ability to capture trends, explainability, and robustness to sudden changes in data distributions. In Table 1, the persistence estimator corresponds to a 24-hour lag in electricity demand.

Seasonal autoregressive integrated moving average (SARIMA) models [81] are also commonly used in time series analysis. Here, we trained one model for each of the 48 half-hours in a day to capture daily seasonality in the data. Each model was then fitted with weekly seasonality by running the `auto.arima` method of the `forecast` package in R.

B.2.2 Generalized additive models

Generalized additive models (GAMs) are a generalisation of linear regression. Instead of learning linear coefficients linking certain features $\mathbf{x} = (x_1, \dots, x_d)$ to a target y , a GAM learns the nodes and coefficients of the regression of the features onto the target with respect to a spline basis. More precisely, given a target time series $y = (y_t)_{t \in T}$ indexed by T , and some explanatory variables $\mathbf{x} = (x_{t,1}, \dots, x_{t,d})_{t \in T}$, the response variable y is written in the form:

$$y_t = \beta_0 + \sum_{j=1}^d f_j(x_{t,j}) + \varepsilon_t,$$

where $\varepsilon = (\varepsilon_t)_{t \in T}$ are independent and identically distributed (i.i.d.) random noise. Though the target y_t at time t is in fact a real number, each potentially explanatory time series $x_k = (x_{t,k})_{t \in T}$ has a dimension $d_k \geq 1$; i.e., at time t , $x_{t,k} \in \mathbb{R}^{d_k}$. Therefore, nonlinear effects of multiple variables are allowed, such as, for instance, $y_t = \beta_0 + f_1(x_{t,1}) + \varepsilon_t$ with $x_{t,1} \in \mathbb{R}^2$. The goal of GAM optimisation is to find the best nonlinear functions f_1, \dots, f_d to fit y . Thus, each nonlinear effect f_j is decomposed over a spline

basis $(B_{j,k})_{1 \leq j \leq d, k \in \mathbb{N}}$, with coefficients β_j , such that

$$f_j(x) = \sum_{k=1}^{m_j} \beta_{j,k} B_{j,k}(x),$$

where m_j corresponds to the dimension of the spline basis. The offset β_0 is chosen so that the functions f_j are centred. The coefficients $\beta_0, \beta_1, \dots, \beta_d$ are then obtained by penalised least squares. The penalty term involves the second derivatives of the functions f_j , forcing the effects to be smooth (see [82]).

The GAM model used in the experiments presented in Table B1 was taken from [21]. As is usual in load forecasting with GAMs, we considered one model per half-hour of the day, with the 48 half-hour time series treated independently. Therefore, 48 models were fitted, one for each half-hour of each day. Given a half-hour h , our model was therefore:

$$\begin{aligned} \text{Load}_{h,t} = & \sum_{i=1}^7 \sum_{j=0}^1 \alpha_{h,i,j} \mathbf{1}_{\text{DayType}_t=i} \mathbf{1}_{\text{DLS}_t=j} \\ & + \sum_{i=1}^7 \beta_{h,i} \text{Load1D}_t \mathbf{1}_{\text{DayType}_t=i} + \gamma \text{Load1W}_t \\ & + f_{h,1}(t) + f_{h,2}(\text{ToY}_t) + f_{h,3}(t, \text{Temp}_{h,t}) + f_{h,4}(\text{Temp95}_{h,t}) \\ & + f_{h,5}(\text{Temp99}_{h,t}) + f_{h,6}(\text{TempMin99}_{h,t}, \text{TempMax99}_{h,t}) + \varepsilon_{h,t}, \end{aligned} \quad (\text{B1})$$

where the timestamp t corresponds to the day, and:

- $\text{Load}_{h,t}$ is the electricity load on day t at time h .
- DayType_t is a categorical variable indicating the type of day of the week.
- DLS_t is a binary variable indicating whether t is daylight saving time or standard time.
- Load1D and Load1W are the loads of the previous day and previous week, respectively.
- ToY_t is the time of year, growing linearly from 0 at midnight when January 1 begins, to 1 on December 31 at 23:30 pm.
- $\text{Temp}_{h,t}$ is the national average temperature at time h on day t .
- $\text{Temp95}_{h,t}$ and $\text{Temp99}_{h,t}$ are exponentially smoothed temperatures with respective factors $\alpha = 0.95$ and 0.99 . For example, $\alpha = 0.95$ corresponds to

$$\text{Temp95}_{h,t} = \alpha \text{Temp95}_{h-1,t} + (1 - \alpha) \text{Temp}_{h,t}.$$

- $\text{TempMin99}_{h,t}$ and $\text{TempMax99}_{h,t}$ are respectively the minimal and maximal values of Temp99 on day t over all time instants i such that $i \leq h$.

We ran these models in R using the `mgcv` library [83]. We used the default thin-plate spline basis to represent the f_j 's, except for the time of year effect f_2 for which we chose cyclic cubic splines (see [82] for a full description of the spline basis). To replicate

the GAM of [47], the dimensions of the bases were taken equal to 5, except for f_2 which had a basis of dimension 20.

B.3 Data assimilation techniques

B.3.1 State space models

State space models are efficient in capturing time-varying structures (as opposed to seasonality) in time series [84]. In particular, the Kalman filter is a powerful mathematical and algorithmic tool introduced by [85] for state space model estimation. In electricity load forecasting, Kalman filters have been used to update the output of a GAM using recent observations of electricity demand [47].

Following the notation in Equation (B1), let $f(\mathbf{x}_t) = (1, f_1(x_{t,1}), \dots, f_d(x_{t,d}))^\top$. Our goal is to estimate a time-varying vector $\boldsymbol{\theta}_t \in \mathbb{R}^{d+1}$ such that $\mathbb{E}[y_t | \mathbf{x}_t] = \boldsymbol{\theta}_t^\top f(\mathbf{x}_t)$. This corresponds to adjusting the relative importance of each nonlinear effect, while preserving their shapes. This is achieved by considering the state space model

$$\begin{aligned}\theta_t - \theta_{t-1} &\sim \mathcal{N}(0, Q_t), \\ y_t - \theta_t^\top x_t &\sim \mathcal{N}(0, \sigma_t^2),\end{aligned}$$

where $\mathcal{N}(\mu, \sigma^2)$ is the multidimensional normal distribution with mean μ and variance matrix σ^2 , θ_t is the latent state, Q_t the process noise covariance matrix, and σ_t^2 the observation variance. Applying the recursive Kalman filter equations as described in section A of [47] provides us with both θ_t and the conditional expectation $\mathbb{E}[y_t | \mathbf{x}_t]$, which is known to be the best forecast, i.e., minimizes the mean square error conditional on past observations and exogenous covariates x_t . As in [47], we ran the three variants *static*, *dynamic*, and *Viking* of the Kalman filter. The *static* version is a degenerate case where Q_t is null, which leads to low adaptation. The *dynamic* variant supposes that $Q_t = Q$ and $\sigma_t = \sigma$ are constants obtained by grid search optimisation on past observations. Finally, the *Viking* version assumes that Q_t and σ_t are updated online (see [47] for more details). In Table 1, the GAM model used in the state space models is that from [21], while the *static* Kalman filter, *dynamic* Kalman filter, and *Viking* method are from [69].

B.3.2 Online aggregation of experts

Online robust aggregation of experts [86] is a model agnostic technique for time series forecasting. This approach combines various forecasts (called experts) based on their past performance, in a streaming manner. This method allows for adaptation to changes in distributions by tracking the best experts. Sequential expert aggregation assumes that the data are observed sequentially. The target variable Y (here electricity demand) is supposed a bounded sequence, i.e., $Y_1, \dots, Y_T \in [0, B]$, where $B > 0$. Our goal is to forecast this variable step by step for each given time t . At each time t , N experts offer forecasts of Y_t , denoted by $(\hat{Y}_t^1, \dots, \hat{Y}_t^N) \in [0, B]^N$. These experts can be the result of any process, such as a statistical model, a physical model, or human-based expertise. Then, the aggregation algorithm generates a forecast of Y_t by

the weighted average of the N forecasts:

$$\hat{Y}_t = \sum_{j=1}^N \hat{p}_{j,t} \hat{Y}_t^j,$$

where the weight $\hat{p}_{j,t} \in \mathbb{R}$ depends on the performance of \hat{Y}_t^j over the period $\{1, \dots, t-1\}$. Then, Y_t is observed and the next instance starts.

In our study, we ran the ML-Poly algorithm, first proposed by [87] and subsequently implemented in R in the `opera` package [88]. This algorithm identifies the best expert aggregation by giving more weight to experts producing the lowest forecasting error, making it noteworthy due to the absence of parameter tuning required. In Table 1, all of the estimators related to data assimilation techniques are combined, i.e., the GAM, the static Kalman filter, the dynamic Kalman filter, and the Viking estimator.

B.4 Machine learning

B.4.1 Random forests

Among the most robust machine learning techniques are random forests [89]. They consist of averaging a given number of decision trees generated by applying classification and regression trees [90] to different subsets of the data obtained by bagging and random sampling of covariates. Each decision tree estimates the target by a series of logical comparisons on the feature variables. An example of a decision tree of depth 3 is “if *temperature* > 30°C, if it is 10 a.m., and if it is a Wednesday, then *electricity demand* = 6 GW”. Random forests require very little prior knowledge about a problem, which makes them good for benchmarking in applied machine learning problems. In Table 1, the random forests all had 1000 trees of depth 6 (the square root of the number of features). Random forests are usually trained on random subsets of the training sample. To take advantage of the dependence of samples in time series, the random subsets can be drawn from a given number of consecutive measures. This is what occurs in the *random forest + bootstrap* architecture [48].

B.4.2 Gradient boosting

Gradient boosting [91, 92] consists of successively fitting the errors of simple models—called weak learners—and then aggregating them. This is an ensemble technique, like random forests. Gradient boosting usually outperforms random forests [93], at the cost of more parameters to calibrate. It has previously shown excellent performance on regression problems [93] and in forecasting challenges [94]. In tree-based gradient boosting algorithms, weak learners are decision trees, whereas in GAM boosting algorithms [95], weak learners are spline regression models.

B.5 Models with mobile phone data

As explained in Appendix B.1, the forecasts trained on the dataset “with mobility data” actually consisted of two models. The first model was trained on the entire

dataset “without mobility data”. The second model estimated the error of the first model on the dataset “with mobility data”. The GAM “with mobility data” is the sum of the GAM “without mobility data” and of the following GAM:

$$\begin{aligned} err_{h,t} = & \sum_{i=1}^7 \sum_{j=0}^1 \tilde{\alpha}_{h,i,j} \mathbf{1}_{\text{DayType}_t=i} \\ & + f_{h,7}(\text{ToY}_t) + f_{h,8}(\text{Work}_t) + f_{h,9}(\text{Residence}_t) + \varepsilon_{h,t}, \end{aligned}$$

with these abbreviations defined in Appendix B.2.2. The static Kalman filter, dynamic Kalman filter, and Viking estimators “with mobility data” were then computed by summing the effects of the two GAMs. The GAM boosting “with mobility data” was the sum of the boosted GAM “without mobility data” and of a boosted GAM with all variables (calendar, meteorological, electricity, and mobile phone). The random forest “with mobility data” was the sum of the *random forest + bootstrap* model “without mobility data” and of a random forest with all variables. The *random forest + bootstrap* “with mobility data” was the sum of the *random forest + bootstrap* model “without mobility data” and of a *random forest + bootstrap* with all variables.

B.6 Excluding holidays

As mentioned in Section 1.3 of the main paper, holidays are known to behave differently from regular days [44]. Therefore, we ran the same benchmark here, but excluding holidays, as well as the days directly before and after holidays, from both training and testing. Table B1 shows that, when excluding holidays, incorporating mobility data improved the best performance (aggregation of experts) by 8% in RMSE and 6% in MAPE. Once again, the global performance improvement across all models was around 10%. Note that these gains are significant, because they leave the confidence interval obtained by bootstrapping (see Methods).

Table B1 Benchmark excluding holidays.

	Without mobility data		With mobility data	
	RMSE (GW)	MAPE (%)	RMSE (GW)	MAPE (%)
<i>Model</i>				
Persistence (1 day)	4.0 ± 0.2	5.0 ± 0.3	N.A.	N.A.
SARIMA	2.0 ± 0.2	2.6 ± 0.2	N.A.	N.A.
GAM	1.70 ± 0.06	2.6 ± 0.1	1.55 ± 0.05	2.43 ± 0.08
<i>Data assimilation technique</i>				
Static Kalman filter	1.43 ± 0.05	2.20 ± 0.08	1.07 ± 0.04	1.63 ± 0.06
Dynamic Kalman filter	1.10 ± 0.04	1.58 ± 0.05	0.96 ± 0.03	1.39 ± 0.04
Viking	0.98 ± 0.04	1.33 ± 0.04	0.98 ± 0.03	1.41 ± 0.05
Aggregation of experts	0.96 ± 0.04	1.36 ± 0.04	0.88 ± 0.03	1.28 ± 0.04
<i>Machine learning</i>				
GAM boosting	2.3 ± 0.1	3.3 ± 0.2	2.2 ± 0.1	3.1 ± 0.2
Random forests	2.1 ± 0.1	3.0 ± 0.1	1.8 ± 0.1	2.4 ± 0.1
Random forests + bootstrap	1.9 ± 0.1	2.6 ± 0.1	1.8 ± 0.1	2.4 ± 0.1

Appendix C Change point detection

In this appendix, we detail and justify the use of the model used in Section 1.1 to assess energy savings, as well as the change point detection algorithm subsequently applied to it.

C.1 The seasonality model

The model used in Section 1.1 to capture the dependence of calendar and meteorological data on electricity demand is the following direct adaptation of the GAM from [21]:

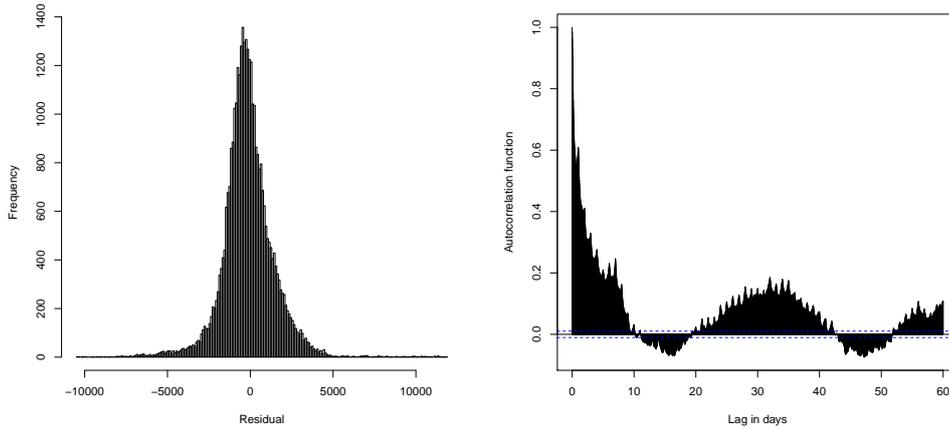
$$\begin{aligned} \text{Load}_{h,t} = & \sum_{i=1}^7 \sum_{j=0}^1 \alpha_{h,i,j} \mathbf{1}_{\text{DayType}_t=i} \mathbf{1}_{\text{DLS}_t=j} \\ & + f_{h,1}(\text{ToY}_t) + f_{h,2}(\text{Temp95}_{h,t}) + f_{h,3}(\text{Temp99}_{h,t}) \\ & + f_{h,4}(\text{TempMin99}_{h,t}, \text{TempMax99}_{h,t}) + \varepsilon_{h,t}. \end{aligned}$$

We note that this corresponds to removing the dependence on the timestamp t and on the lags Load1D and Load1W from equation (B1). On the one hand, these features were removed because they only captured the trend of the signal without explaining the phenomena at stake, which interfered with the interpretability of the model. On the other hand, the remaining features account for well-known repeated phenomena such as the effects of weekends (in DayType), holidays (in ToY), and heating and cooling (in the smoothed temperatures), thus helping to explain seasonality in the signal. This GAM was trained on data from 01/01/2014 to 01/01/2018. The residuals $\text{res} = \text{Load} - \hat{\text{Load}}$ were then evaluated from 01/01/2018 to 01/03/2023. Between 01/01/2018 and 01/01/2020, this GAM had an average MAPE of 2.1% and an average RMSE of 1.6 GW. This is comparable to the performance of the GAM in [21], which had an average MAPE of 1.6% and an average RMSE of 1.2 GW. At the cost of slightly lower performance, our GAM is more interpretable because it only takes seasonal phenomena into account. We therefore consider it to be a good model for forecasting what the electricity demand should be over a multi-year time horizon, assuming that electricity consumption behaviour remains unchanged.

C.2 Descriptive analysis of residuals

In this paragraph, we focus on the period spanning 01/01/2018 to 01/01/2020. As shown in Figure C6 (left), the residuals histogram is bell-shaped. Since we had $2 \times 365 \times 48 = 35040$ observations, we chose the number of breaks in the histogram to be $\lfloor \sqrt{35040} \rfloor = 187$, where $\lfloor \cdot \rfloor$ is the floor function. Student's t -test showed that the expectation of the residuals was significantly lower than zero ($p < 2.2 \times 10^{-16}$) and was contained in the interval $[-0.16 \text{ GW}, -0.12 \text{ GW}]$ with a probability of 95%. The empirical mean was -0.14 GW , while the empirical standard deviation was 1.6 GW. An Anderson-Darling normality test suggested that the residuals did not follow a normal distribution ($p < 2.2 \times 10^{-16}$). Moreover, as shown in Figure C6 (right), the

Fig. C6 Descriptive statistics of the residuals.



Left: Histogram of the residuals between 01/01/2018 and 01/01/2020. **Right:** Autocorrelation function of the residuals between 01/01/2018 and 01/01/2020. The dotted lines correspond to a confidence interval pertaining to the precision of the auto-correlation estimators.

autocorrelations of the residuals decreased slowly and were significantly greater than zero, suggesting that the residuals were not stationary. Further evidence for this came from a Box-Ljung test with a 1-day window ($p < 2.2 \times 10^{-16}$). Both the fact that the expectation of the residuals was significantly less than zero and the residuals were not stationary indicated that other phenomena than calendar seasonality and temperature are involved, though their impact appears to be moderate since the estimator performs well. This suggests that, even in this period without major events or decisions such as COVID-19 or sobriety, other features can be useful for better understanding electricity demand.

C.3 Ranking changes in the data distribution

The descriptive analysis shows that the residuals are not stationary. Therefore, from a statistical point of view, it is pointless to look for the change points observed in Figure 1 in absolute terms. In fact, the more precise the technique for detecting change points becomes, the more change points will be detected everywhere. This is why we need quantitative information about the importance of the change points in order to rank them and determine which are the most significant ones. A number of metrics have been developed to measure the importance of change points [40]. To assess the significance of the number of change points here, we sequentially compared the standard deviation of the residuals with the amplitude of the change points. This resulted in 10 change points being considered in the following analysis. The principle behind offline change-in-mean techniques is to segment the signal in such a way that approximating the signal by its mean in each segment results in the lowest possible variance. However, finding such an optimum would be computationally expensive for our time

series of around 70 000 observations. Therefore, we rely on faster algorithms which have been developed to find approximations to the optimal change points, such as binary segmentation, as was used in Figure 1.

Appendix D Statistical analysis

In this appendix, we provide further analyses related to the variable selection detailed in the Results section. to further justify the study of the *work* index in the statistical analysis of Section 1.4.

D.1 Variable selection: Hoeffding D-stastics and Shapley values

This paragraph complements the mRMR variable ranking performed in the Results section. To examine the variable selection process more closely, we computed the Hoeffding D-statistic, as shown in Table D2. This is a distribution-free measure of the dependence between variables [96]; the closer it is to 1, the greater the dependence. We then computed the Shapley values of the same variables using the SHAFF algorithm [97]; results are shown in Table D3. We see that with the three ranking methods, the three most important variables, in order of importance, were the *temperature*, the *work* index, and the *time of year*. Of note, the effect of the *work* index only became clear after filtering out the effect of the temperature on electricity demand. We see that the importance of *tourism* and *time of year* decreases when correcting electricity demand for temperature, due to their high correlation with temperature. As a result of this analysis, the *tourism* and *residents* indices did not seem to have a significant impact on French electricity demand.

Table D2 Hoeffding D-statistic.

	Temp95	Work	Residence	Tourism	Toy	Dow	Holidays
Load	0.29	0.035	0.093	0.19	0.069	0.010	0.000071
Load \ Temp	0.018	0.18	0.010	0.025	0.017	0.091	0.00066
Load \ (Temp,Work)	0.0057	0.0067	0.014	0.011	0.038	0.0057	-0.00001

The statistic was computed using all available days from July 2019 to March 2022. *Load \ (Features)* stands for the *Load* corrected for the effect of the *Features*.

D.2 *work* index and calendar features

The variable selection analysis in Appendix D.1 showed that the *work* indicator has a very strong effect on the electricity demand, being the second most explanatory variable. To better understand this effect, we compared—in Table D4—the performance of GAMs where we progressively added the features in the order of importance suggested by the variable selection analysis. The *Temp* GAM corresponds to the model

$$\text{Load}_{h,t} = f_h(\text{Temp95}_{h,t}) + \varepsilon_{h,t}.$$

Table D3 Shapley values.

	Temp95	Work	Residence	Tourism	Toy	Dow	Holidays
Load	0.31 ± 0.05	0.05 ± 0.02	0.06 ± 0.03	0.14 ± 0.01	0.21 ± 0.01	0.03 ± 0.02	0.002 ± 0.003
Load \ Temp	0.041 ± 0.007	0.26 ± 0.01	0.035 ± 0.005	0.072 ± 0.005	0.11 ± 0.02	0.19 ± 0.02	0.04 ± 0.01
Load \ (Temp, Work)	0.11 ± 0.01	0.10 ± 0.01	0.06 ± 0.01	0.07 ± 0.01	0.28 ± 0.01	0.04 ± 0.01	0.003 ± 0.008

Shapley values were computed on all available days from July 2019 to March 2022. $Load \setminus (Features)$ stands for the $Load$ corrected for the effect of the $Features$.

The $Temp + Work$ GAM corresponds to the model

$$Load_{h,t} = f_{h,1}(Temp95_{h,t}) + f_{h,2}(Work_{h,t}) + \varepsilon_{h,t}.$$

The $Temp + Time$ GAM corresponds to the model

$$Load_{h,t} = \sum_{i=1}^7 \sum_{j=0}^1 \alpha_{h,i,j} \mathbf{1}_{DayType_t=i} \mathbf{1}_{DLS_t=j} + \beta \mathbf{1}_{Holidays_t} + f_{h,1}(ToY_t) + f_{h,2}(Temp95_{h,t}) + \varepsilon_{h,t}.$$

The $Temp + Time + Work$ GAM corresponds to the model

$$Load_{h,t} = \sum_{i=1}^7 \sum_{j=0}^1 \alpha_{h,i,j} \mathbf{1}_{DayType_t=i} \mathbf{1}_{DLS_t=j} + \beta \mathbf{1}_{Holidays_t} + f_{h,1}(ToY_t) + f_{h,2}(Temp95_{h,t}) + f_{h,3}(Work_{h,t}) + \varepsilon_{h,t}.$$

The $Temp + Work + Lags$ GAM corresponds to the model

$$Load_{h,t} = f_{h,1}(Temp95_{h,t}) + f_{h,2}(Work_{h,t}) + \sum_{i=1}^7 \alpha_{h,i,j} \mathbf{1}_{DayType_t=i} Load1D_{h,t} + \beta Load1W_{h,t} + \varepsilon_{h,t}.$$

The $Temp + Time + Lags$ GAM corresponds to the model

$$Load_{h,t} = \sum_{i=1}^7 \sum_{j=0}^1 \alpha_{h,i,j} \mathbf{1}_{DayType_t=i} \mathbf{1}_{DLS_t=j} + \beta \mathbf{1}_{Holidays_t} + f_{h,1}(ToY_t) + f_{h,2}(Temp95_{h,t}) + \sum_{i=1}^7 \gamma_{h,i,j} \mathbf{1}_{DayType_t=i} Load1D_{h,t}$$

$$+ \lambda \text{Load1W}_{h,t} + \varepsilon_{h,t}.$$

The *All variables* GAM corresponds to the model

$$\begin{aligned} \text{Load}_{h,t} = & \sum_{i=1}^7 \sum_{j=0}^1 \alpha_{h,i,j} \mathbf{1}_{\text{DayType}_t=i} \mathbf{1}_{\text{DLS}_t=j} + \beta \mathbf{1}_{\text{Holidays}_t} + f_{h,1}(\text{ToY}_t) \\ & + f_{h,2}(\text{Temp95}_{h,t}) + \sum_{i=1}^7 \gamma_{h,i,j} \mathbf{1}_{\text{DayType}_t=i} \text{Load1D}_{h,t} + \lambda \text{Load1W}_{h,t} \\ & + f_{h,3}(\text{Work}_{h,t}) + \varepsilon_{h,t}. \end{aligned}$$

The p-values of the Fisher tests assessing the significance of the GAM effects were below 5% for all GAMs. We see in Table D4 that replacing calendar data by the *work* index was beneficial during atypical events whose behaviour differed from the past, i.e., the *sobriety* period here. Indeed, the time variables were only relevant during the *normal period* spanning from July 2023 to September 2023, during which they still benefitted from the *work* index. During the *sobriety* period, the time variables—which only reconstruct past behaviour—were less explanatory than the *work* index, which did not benefit from being coupled with them.

Table D4 Integration of mobility data in GAMs.

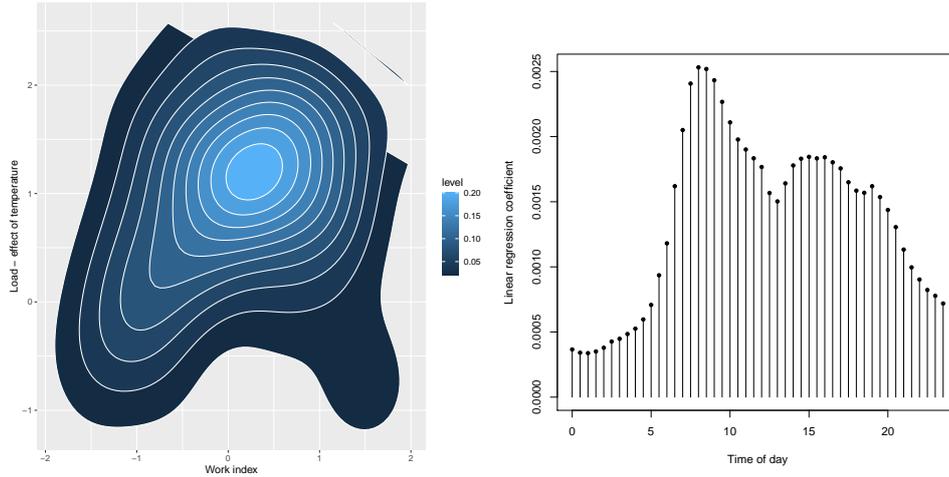
	Normal period		Sobriety	
	RMSE (GW)	MAPE (%)	RMSE (GW)	MAPE (%)
<i>Baseline</i>				
Persistence (1 day)	3.49	5.36	4.03	5.26
<i>GAM</i>				
Temp	3.53	6.78	6.13	9.60
Temp + Work	1.62	3.10	4.98	8.11
Temp + Time	1.33	2.46	5.60	9.62
Temp + Time + Work	1.09	2.02	5.24	9.00
Temp + Work + Lags	1.11	1.92	2.11	3.13
Temp + Time + Lags	0.89	1.52	2.61	4.29
All variables	0.80	1.38	2.60	4.32

This benchmark covers all days, including holidays.

D.3 Work dynamics

In Section 1.4, we explained how the *work* index captures the effects of both the *day of week* and *holidays* features. However, in both Section 1.3 and Appendix D.2, we showed that the *work* index improved the performance of the forecast, beyond the effect of the calendar features. Let us briefly take a closer look at this effect. To remove the effect of *time of day* and *holidays*, we worked on a specific day (here Wednesday) and removed holidays. Figure D7 (left) shows how the electricity demand on Wednesdays

Fig. D7 Effect of the *work* index on a given day at a given hour.



Left: 2d density plots of residuals as function of the *work* index at 10 a.m. on the Wednesdays between July 2019 and March 2022. **Right:** Regression coefficient of the *work* index on electricity demand corrected for the effect of temperature on the training set spanning July 2019 to March 2022.

was still positively influenced by the *work* index. Furthermore, as expected, Figure D7 (right) shows that the effect of the *work* index was more important during working hours (from 6 a.m. to 8 p.m.). These results confirm that on Wednesdays a high *work* index corresponded to high electricity demand. This effect could be due to economic growth (higher economic activity corresponding to both more people working which raises the *work* index, and to higher electricity demand) and to energy saving due to remote working (a lower office occupancy corresponding both to a lower *work* index and lower electricity demand).