One Earth

Article

Social media data shed light on air-conditioning interest of heat-vulnerable regions and sociodemographic groups

Graphical abstract

Highlights

- Social media data can be used for estimating the drivers of air-conditioning adoption
- Sociodemographic factors distinguish online air-conditioning interest globally
- Heat-vulnerable regions and groups show high online interest in air conditioning

Authors

Sibel Eker, Alessio Mastrucci, Shonali Pachauri, Bas van Ruijven

Correspondence

sibel.eker@ru.nl

In brief

As temperatures increase, air conditioners are becoming increasingly vital for climate adaptation, but widespread air conditioning can result in growing greenhouse gas emissions and impose challenges on the resilience of power grids. It is thus vital to understand future demand for cooling globally, but the extent and drivers of air-conditioning adoption remain unclear because the available data are limited. Here, we use social media advertising data across 136 countries to fill the knowledge gap and show that age, relationship status, and parenthood are key drivers of air-conditioning adoption worldwide.

Eker et al., 2023, One Earth *6*, 428–440
April 21, 2023
https://doi.org/10.1016/j.oneear.2023.03.011

CellPress
OPEN ACCESS

One Earth

Article

Social media data shed light on air-conditioning interest of heat-vulnerable regions and sociodemographic groups

Sibel Eker,[1,2,3,*] Alessio Mastrucci,[2] Shonali Pachauri,[2] and Bas van Ruijven[2]
[1]Nijmegen School of Management, Radboud University, Nijmegen, the Netherlands
[2]International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria
[3]Lead contact
*Correspondence: sibel.eker@ru.nl
https://doi.org/10.1016/j.oneear.2023.03.011

SCIENCE FOR SOCIETY Climate change exacerbates life-threatening heat waves, making air conditioning (AC) units, which enable comfortable indoor temperatures, a vital adaptation approach. However, AC can generate hydrofluorocarbons—a much stronger greenhouse gas than CO_2. AC use is also electricity hungry, challenging power grid stability and potentially increasing CO_2 emissions in regions where electricity is generated from fossil energy. To enable AC units to offer reliable and climate-friendly cooling services, especially in those heat-vulnerable regions in the Global South, it is essential to know the extent and drivers of future AC demand, but the available data are very limited. In an effort to better understand AC demand and motivations, we leverage rich and accessible social media advertising data. Our analysis confirms that such data can act as a reasonable proxy of willingness to buy AC units. We show that age, relationship status, and parenthood are key factors influencing AC adoption, particularly in regions that are increasingly vulnerable to heat waves.

SUMMARY

Cooling homes with air conditioners is a vital adaptation approach, but the wider adoption of air conditioners can increase hydrofluorocarbon emissions that have high global warming potential and carbon emissions as a result of more fossil energy consumption. The scale and scope of future cooling demand worldwide are, however, uncertain because the extent and drivers of air-conditioning adoption remain unclear. Here, using 2021 and 2022 Facebook and Instagram data from 113 countries, we investigate the usability of social media advertising data to address these data gaps in relation to the drivers of air-conditioning adoption. We find that social media data might represent air-conditioning purchasing trends. Globally, parents of small children and middle-aged, highly educated married or cohabiting males tend to express greater interest in air-conditioning adoption. In regions with high heat vulnerability yet little empirical data on cooling demand (e.g., the Middle East and North Africa), these sociodemographic factors play a more prominent role. These findings can strengthen our understanding of future cooling demand for more sustainable cooling management.

INTRODUCTION

Space cooling is a primary end use of energy in the buildings sector and the fastest growing.[1] It accounted for nearly 16% of the final electricity consumption in buildings in 2020.[2] Energy consumption for space cooling has more than tripled since 1990 and is expected to triple again by 2050 if no energy efficiency measures are implemented.[1] The share of space cooling in the global final energy demand for buildings is projected to rise from 4% in 2010 to 11%–37% in 2100.[3] One of the major drivers of this increase is the need for more cooling to adapt to increasing temperatures and heatwaves caused by climate change.[4] Compared with baseline scenarios in which energy demand is driven by population and income growth alone, a pervasive need for cooling in multiple sectors is estimated to increase the global energy demand by 11%–27% in 2050 even in a moderate warming scenario.[5] Being the fastest-growing end use of energy, space cooling stresses the trade-off between climate change adaptation and mitigation, besides being tighly linked to all 17 UN Sustainable Development Goals.[6]

Alongside climate change, macro-socioeconomic trends such as rising population and income have been considered the main drivers of cooling demand.[1,6] Recent research, however, shows that demographics and household characteristics also play a major role in adoption of air conditioning (AC) in residential buildings. Urbanization, age structure, family size, and the presence of young children in the family strongly affect the AC adoption decision of households in eight Organisation for Economic Cooperation and Development (OECD) countries.[7] The role of such demographic factors is similar in developing countries. Analyses of household surveys from Brazil, India, Mexico, and Indonesia suggest that the decision to purchase AC is influenced by housing conditions, education, employment, gender, and the household head's age, besides income as a key factor.[8] Income and education influence AC adoption more significantly in developing countries than in developed countries.

The influence of demographic, socioeconomic, and geographic heterogeneity on the adoption of AC appliances underlines the importance of taking this heterogeneity into account in modeling studies that estimate future cooling demand and mitigation options. Currently, integrated assessment models (IAMs) highly underestimate the additional cooling needs of the building sector because they assume energy demand to be driven only by income, population, and unchanging climatic conditions.[9] Considering the long-term, socially heterogeneous demand responses driven by the penetration of AC appliances would not only enhance the realism and accuracy of energy demand models but also facilitate the creation of feasible mitigation and adaptation pathways.[10,11] This consideration of heterogeneity would also make the models more relevant to understand the ties between cooling consumption and well-being.[12]

Modeling the demographic, socioeconomic, and geographic heterogeneity of cooling demand requires heterogeneous data on a global scale. Data availability for AC ownership, however, is limited to household surveys from a few countries only.[9] Household characteristics and demographic factors that can play a role in AC adoption are not consistently measured across these surveys. As a result of such limitations, modeling studies are often based on historical data drawn from a few countries, for instance, applying the relationship between AC ownership and climatic conditions in the United States (US) to other regions,[13] undermining the effect of geographic and socioeconomic differences on global cooling demand.

Major social media platforms, such as Facebook (including Instagram) and Twitter, report audience sizes of various demographic groups and cultural interests aggregated from online activities, such as likes, posts, events, and browsing history. This online data footprint of individuals can help elucidate consumption behavior, lifestyle-change tendencies, and their drivers.[14] Despite the innovative use of social media data in analyzing several social phenomena, such as obesity prevalence,[15] disaster response and risk perception,[16] urban sustainability,[17] and gender inequality,[18] its use in analyzing consumption behavior has been limited to a few studies. For instance, the geotags of tweets are used for estimating daily travel demand in specific locations,[19,20] and Facebook audience-size data are used for analyzing the drivers of low-carbon dietary choices on a global scale.[21]

Social media data can address the limitations of available data and complement conventional data sources in terms of geographic, temporal, and contextual scope as a low-cost data source. Covering billions of users and collected by the same algorithms, social media data are consistent across countries and provide large-scale information that cannot be obtained from surveys with inevitably limited sample sizes. Retrieving these data from the application program interfaces (APIs) of the social media platforms bears a much lower cost than conducting large-scale surveys in multiple countries. Additionally, social media data are based on users' posts, likes, purchases, and other online activities, and these activities can represent observed behavior as opposed to self-reported answers to survey questions, which are often biased by response styles, such as socially desirable responding.[22,23] In the case of AC adoption, for instance, social media data can represent users' online research for appliances before or after the purchase. However, as a result of the aggregation of various online activities and the black-box nature of the algorithms used in this aggregation, social media data might not represent the actual behavior, which can be directly measured by purposeful surveys. Therefore, it becomes necessary to identify the similarities and differences between social media data and purposeful surveys in order to be able to derive useful insights from the social media data.

Here, we address the lack of global data on the extent and drivers of AC adoption by using social media advertising data collected from Facebook and Instagram across 113 countries. We first compare the social media advertising data (audience size of Facebook and Instagram users interested in AC) with household surveys from four countries to identify the similarities and differences between the two datasets in terms of the extent and heterogeneity of AC interest and adoption. We find that social media advertising data correlate best with the trends in AC ownership, that is, recent purchases. We then scale the analysis of social media advertising data up to a global level and find that heat-vulnerable regions, such as the Middle East, North Africa, Pacific Asia, and Balkan countries, have the highest online interest in AC. Globally, social heterogeneity of AC adoption becomes visible in online interest, too, given that middle-aged and highly educated groups, married or cohabiting individuals, and parents of small children tend to express a higher online interest in AC. We demonstrate that social media advertising data can be useful in evaluating future demand for cooling and that age, relationship status, and parenthood are the key factors of willingness to purchase AC units. Such could be especially useful for sustainable cooling management, especially in regions that are vulnerable to heat stress but have little available empirical data.

RESULTS

Social media data as a proxy for AC adoption

To analyze the relation between social media data and actual consumption patterns, we compare the fraction of the social media audience interested in AC (F_i) averaged over two seasons, with the fraction of survey-respondent households that recently purchased air conditioners (S_i). We expect this comparison to demonstrate the similarities and differences between the two data sources, yet we do not expect the social media data to predict the survey data given the inherent differences between the two sources. The surveys cover four countries with different socioeconomic and climatic conditions and measure the number of households that own or have recently purchased an AC unit.

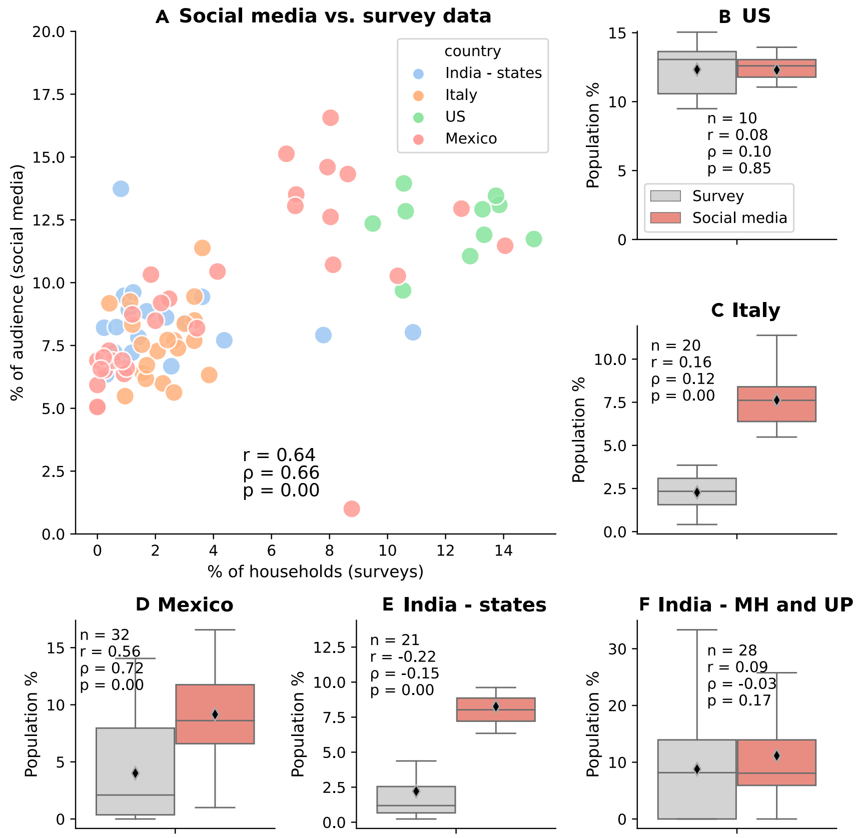

Figure 1. Comparison of social media data and household surveys

(A) Point-by-point comparison of percentage of social media audience interested in AC (y axis) and the percentage of households that recently purchased AC (x axis) in the states or regions of four countries. Each data point refers to a region or state. The social media data are averaged over two data-collection points. r and ρ refer to Pearson and Spearman correlation coefficients, respectively. p shows the p value of a two-sided Wilcoxon signed-rank test on the null hypothesis that the survey and social media data follow the same distribution. With $p < 0.05$, this null hypothesis is rejected.

(B–E) Boxplots showing the range of percentages of social media audience interested in AC (red) and the percentages of households that recently purchased AC (gray) in the US, Italy, Mexico, and India, respectively. n is the number of states or regions. The boxes show the quartile range, where black diamonds mark the mean, and whiskers (error bars) extend to ±50% of the interquartile range.

(F) Boxplots showing the range of percentages of social media audience interested in AC (red) and the percentages of households that “intend” to purchase AC (gray) in the semi-urban areas of two states of India, Maharashtra (MH) and Uttar Pradesh (UP).

Among these two metrics, we found that the social media audience fraction interested in AC better correlates with the fraction of households with recent purchases (experimental procedures; Figure S1), which can be explained by pre- or post-purchase online research activity.

We find that AC interest of the social media audience (F_i) and households’ recent purchase rates (S_i) are of a comparable order of magnitude between 0% and 17.5% at the regional level (Figure 1A). The two data sources are positively correlated, yet the low value of correlation coefficient (r) or rank correlation (ρ) indicates discordances in actual values. In the US and the semi-urban areas of India in two states (Figures 1B and 1F), the mean and median values of the social media audience interest and household purchases are similar, and the similarity of their distribution is statistically significant ($p > 0.05$) based on a Wilcoxon signed-rank test. A relatively strong correlation is observed in Mexico. In Italy and India (Figures 1C and 1E), there is a gap between the ranges of the two data sources, such that the social media audience fraction interested in AC is steadily exceeding the fraction of households that recently purchased AC. In semi-urban areas of India (Figure 1F), we observed a higher correlation between the social media audience fraction and the fraction of households that “intend” to purchase an air conditioner rather than those with actual recent purchases (see experimental procedures).

From this comparison, we conclude that social media data reflect similar orders of magnitude for the trends of AC adoption, that is, recent purchases that indicate the rate of change of the appliance ownership stock, despite noticeable and statistically significant discrepancies from the survey data. These discrepancies can be explained by the different data-collection years, the penetration rate of social media platforms, and the fact that not all online activity on a subject such as AC results in an actual purchase. Therefore, social media data might not be an accurate predictor of AC adoption, yet they can complement conventional data sources as an approximate indicator of purchasing trends.

Social media data explain social heterogeneity

Age, gender, and education are considered the main demographic drivers of cooling demand,[7,8] and they are commonly recorded heterogeneity attributes in both data sources. To analyze how the two data sources represent the social heterogeneity of AC demand, we compare the effect of age, gender, and education on audience interest and the effect of these attributes on purchasing rates in the five household surveys. Figure 2 shows a visual comparison of social media and survey data across consistent categories of these attributes.

Among genders, males show a higher online interest in AC than females in all countries, indicated by a statistically significant difference between the males and females in the social media dataset ($p_{soc} < 0.05$ for a Wilcoxon signed-rank test with an alternative hypothesis that the interest of males is higher than that of females). Selected semi-urban areas of India are an exception to this gender difference. According to the survey data, this gender difference is statistically significant only in Italy and the selected semi-urban areas of India. Surveys from the US, Indian states, and Mexico show that a lower or similar fraction of households where the main income earner is male have recently purchased an air conditioner in comparison with households

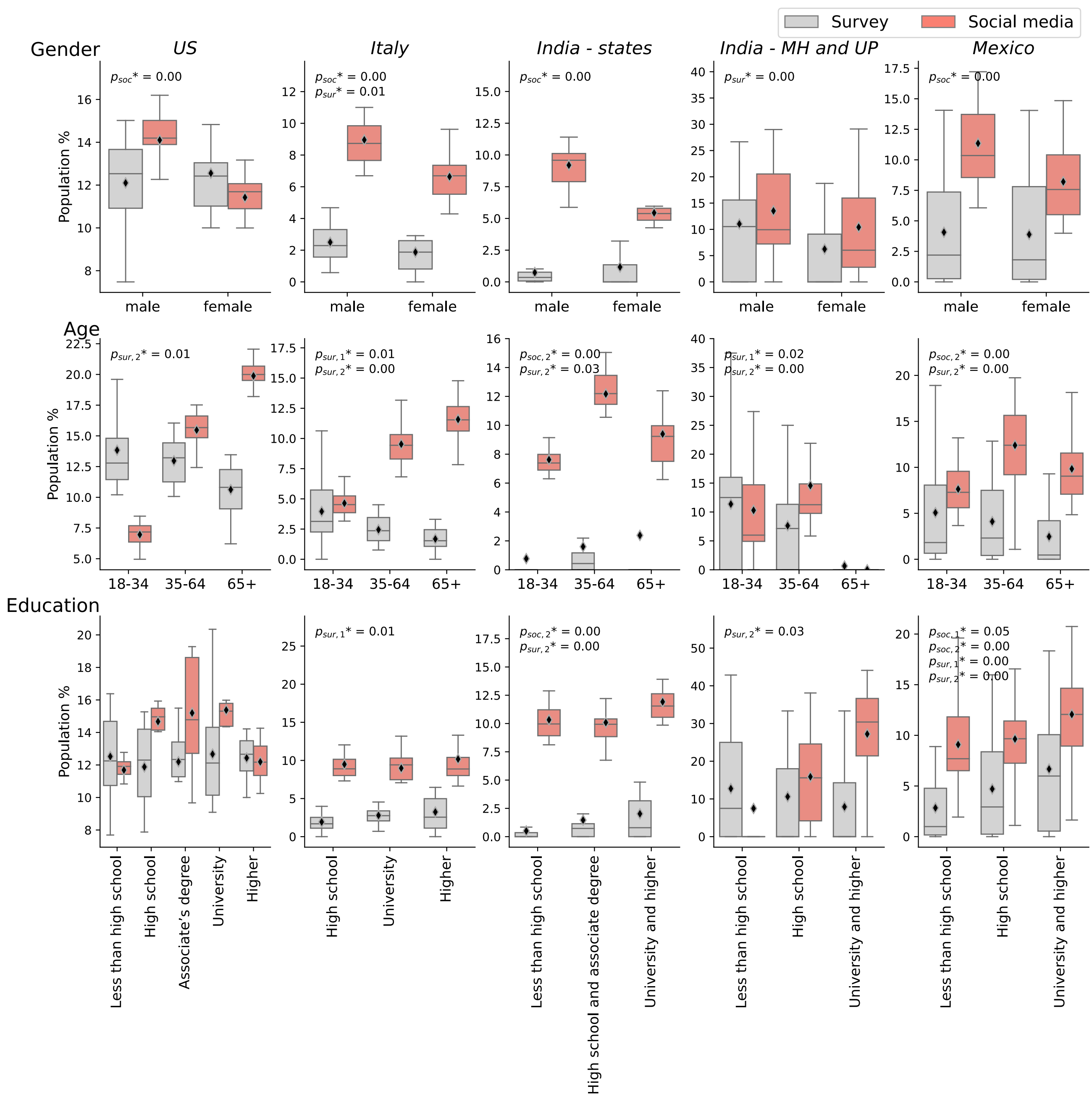

Figure 2. Comparison of social media data and household surveys for gender, age, and education

Boxplots showing the range of percentages of social media audience interested in AC (averaged over two data-collection points) and the percentages of households that recently purchased AC (or intend to do so) in the US, Italy, Mexico, and India. The boxes show the quartile range across states or regions within countries, where black diamonds mark the mean, and whiskers (error bars) extend to ±50% of the interquartile range. In the IRES survey across India's states, there are very few states with respondents who recently purchased AC and where the household head is young (18–34) or old (65+) or with less than a high school education. Therefore, there is no variation in the data that results in visible boxplots. p values of the one-sided Wilcoxon signed-rank test are reported in the cases where $p \leq 0.05$, indicating a statistically significant support for alternative hypotheses that social media interest (purchase rate) is higher in males than females, in younger age cohorts than older age cohorts, and in higher education levels than lower ones. See the experimental procedures for the description of statistical tests.

with a female main income earner. This finding is similar to previous studies that report mixed findings about the effect of gender across different countries.[8] Therefore, we conclude that a higher fraction of males demonstrate online interest in AC as captured by the social media data, yet the purchase rates do not show a similar uniform gender difference between the male-headed or female-headed households.

For age, the results are variant across countries, too. According to all surveys except the state-level survey in India, as the age of the household head increases, the mean value of recent AC purchases declines. The difference in distributions is statistically significant especially for the difference between the 35–64 and 65+ age cohorts ($p_{sur,2} < 0.05$). This finding is consistent with the previous findings that older people in the OECD countries use less AC.[7] The social media data show a different pattern in AC interest across age cohorts. Social media users aged 35–64 are more interested in AC than those aged 18–34, but the audience above 65 years old has a lower interest than the middle-aged cohort, except in the US and Italy. There are multiple possible explanations for this discrepancy, even though a detailed analysis of these is beyond the scope of this research. The 35- to 64-year-old people might be more interested in AC than the 18-to 34-year-old people because AC adoption correlates positively with home ownership,[7,8] and the average age of a first-time home buyer has increased even in affluent countries, such as the UK and US.[24,25] The higher interest of the 65+ age cohort than of the middle-aged population in the US and Italy can be due to the social media bias toward high-income, highly educated people in the older-age cohorts, who are more likely to adopt AC than lower-income groups.[26] The increasing interest of older people in AC adoption in recent years might offer another explanation for this discrepancy given that the surveys are not as recent as the social media data.

For education, the findings from the survey data resonate with the previous findings in the literature that show a positive relationship between education level and AC adoption and that education plays a more important role in developing countries.[8] The higher AC purchase rates among the highly educated are more visible and statistically significant in Italy, India (state level), and Mexico. (In India and Italy, only the difference between high school and university graduates is statistically significant.) In the US, there are no statistically significant differences between the education levels, as the boxplots of the distribution across ten divisions show. The social media data show a positive relationship between the education and AC interest in Mexico and India, whereas education does not seem to influence online AC interest in high-income countries (Italy and the US), similar to the survey data. Therefore, the effect of education on AC adoption differs across countries, and these differences are represented also by the social media data.

Global interest in AC and its drivers

We extend the analysis of social media data to a global level to understand the worldwide extent and drivers of online interest in AC. Figure 3A shows the extent of AC interest of the Facebook and Instagram audiences in 113 countries where social media data are available and sufficiently representative of the total population. The highest interest is observed in the Middle East and North Africa (MEA), Pacific Asia (PAS), and Eastern Europe and Balkans (EEU), which are reported to be increasingly vulnerable to heatwaves.[27–29] Given that the penetration rates are also high in these countries (Figure S2), social media data can be considered useful indicators of the increasing AC interest and the social heterogeneity of it in these vulnerable regions.

On a global scale, gender affects the online AC interest as observed in the US, Italy, Mexico, and India (Figure 3B). The mean percentage of the male audience interested in AC (14.14%) is 19% higher than that of females (11.88%). The distribution of AC interest across countries is flatter in the Global North than in the Global South, where the AC interest is skewed toward lower values, but a long tail indicates a few countries with high interest, such as those in MEA, PAS, and the Balkans. This implies a relative homogeneity in the Global North but a strong heterogeneity across the Global South countries.

The effect of age (Figure 3C) and education (Figure 3D) is also similar to our previous findings. The middle-aged cohort has a higher mean percentage interest in AC (16.6%) than younger (10.4%) and older (14.4%) cohorts. Global North countries are relatively homogeneous in terms of AC interest among the 35- to 64-year-old cohorts, indicated by a flat distribution, yet among the older population, only a few countries such as Cyprus, Hungary, and Greece have a high fraction of interest. This finding implies that the elderly population, who is more vulnerable to heat stress but known to be reluctant to adopt AC, might be changing their behavior toward AC adoption in these countries where heat stress is increasing. Education plays a relatively minor role in differentiating the AC interest on a global scale. The higher the education level, the higher the interest in AC in the Global South, except for the audience with less than a high school education. In the Global North, the difference between the mean value of each education level is marginal. Education plays a more important role in the Global South than in the Global North such that there is a wider difference between the average interest rate of high school graduates and post-graduates in the Global South (14.2%–18.9%) than in the Global North (14.2%–14.6%).

Relationship (Figure 3E) and family status (Figure 3F) are two other household characteristics that affect AC interest. The mean fraction of the social media audience who is married or cohabiting and interested in AC is 26%, and this is 53.8% higher than the AC-interested audience fraction with any relationship status. On average, a slightly higher fraction of the married or cohabiting audience in the Global North (26.4%) than in the Global South (25.9%) is interested in AC. Relationship status distinguishes AC interest more in the Global South such that the average AC interest of married or cohabiting audience is 60% more than that of the general audience, whereas this difference is 40% in the Global North. Having young children in the household affects AC interest significantly, too. The social media data show that 46.2% of parents of 0- to 8-year-old children are interested in AC on average, whereas this fraction is 42.4% for parents of 9- to 18-year-olds and 18.2% for those with any family status. Having young children affects AC interest more in the Global South than in the Global North.

The two other important factors that are known to affect AC ownership are income and climatic conditions. Figure 3G shows the relationship between AC interest, which is an indicator of purchase rates rather than ownership, and gross domestic product (GDP) per capita across countries in the Global North and South separately and across states of the US, Brazil, India, and Mexico. In the Global North, a negative relationship exists, yet the statistical significance of this relationship disappears when the outliers are removed (Figure S4). This negative relationship can be attributed to the already high AC ownership and hence relatively low social media interest, as observed in the wealthier states of the US, such as the District of Columbia, New York, and Massachusetts, or to the lower need for AC due to climatic conditions, as is the case in high-income northern European countries, such as Norway and Luxembourg. In the Global South, GDP per capita

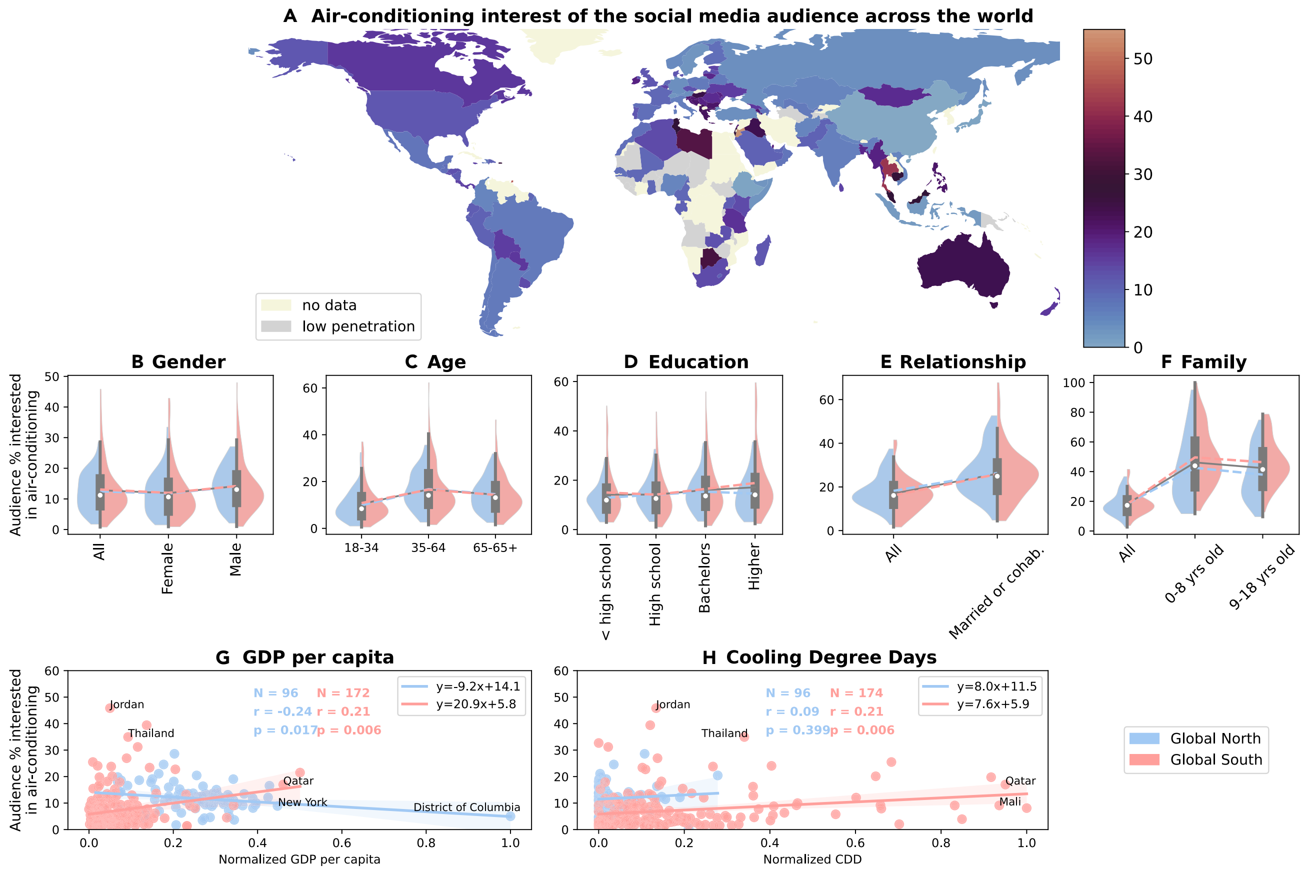

Figure 3. AC interest of the global social media audience and its relation to main demographic, economic, and climatic factors

(A) Audience percentage interested in AC (F_i) across 113 countries, where the countries with no data and low social media penetration rates are colored in gray and beige, respectively.

(B–F) Violin plots showing the distribution of audience percentage interested in AC across 113 countries, separately for Global North (GN) and Global South (GS), for each category of gender (total, female, male), age cohorts, education level, relationship status (total audience and the audience who is married or cohabiting), and family status (total audience, parents of 0- to 8-year-old children, parents of 9- to 18-year-old children). Gray bar plots in the middle show the distribution for all countries, not separated as GN and GS, marked by the median (white dot), interquartile range (box), and ±50% of the interquartile range (whiskers). Solid gray lines and dashed blue and red lines connect the mean values in each category, for all countries, GN, and GS, respectively.

(G and H) Linear regression plots of the relationship between air-conditioning interest and (normalized) GDP per capita and cooling degree days (CDDs) in 135 countries and the states of US, Mexico, Brazil, and India, resulting in 281 data points, of which 12–14 are removed because of the lack of GDP and CDD data. (A different form of regression function does not provide a better fit, as the scattered residuals in Figure S3 show.) Normalization of the x axes was done with respect to the minimum and maximum values: \$411 and \$203,173 for GDP per capita and 0 and 1,528 for CDD, respectively. For each region, the number of data points (N), Pearson correlation coefficient (r), and the p value of the regression coefficients are reported. Table S1 lists the GN and GS countries, and Table S2 shows the descriptive statistics.

correlates positively with the AC interest even though the highest AC interest is in countries with relatively low GDP per capita, such as Jordan and Thailand. Alongside income, climatic conditions indicated by cooling degree days (CDDs) correlate positively with the audience fraction interested in AC in countries of the Global South and states of India, Mexico, and Brazil (Figure 3H). In the Global North, there is no statistically significant relation between CDDs and AC interest. This implies that climate is not a significant factor affecting AC interest at the national level, at least in the range that CDDs are currently distributed across the Global North countries.

AC interest in vulnerable regions

PAS, MEA, EEU, and Africa (AFR) are regions that are highly vulnerable to heat stress,[27,28] but there are few data available to capture social heterogeneity of AC adoption in most of the countries in these regions. These regions, except AFR, are those where the social media audience is most interested in AC, as Figures 3A and S2 show. The social media audience also represents a relatively high fraction of the population in these regions, that is, 55.1%, 65.8%, 57.3%, and 15.4% in PAS, MEA, EEU, and AFR, respectively. Therefore, we focus on these four regions to address the data gap on the drivers of AC adoption and explain how AC interest on social media relates to gender, age, education, and family characteristics (Figure 4).

Similar to the global level, males tend to have a higher online interest in AC than females in these regions. The relationship of AC interest with age and education is also similar to the global results. The middle-aged cohort (35–64) is the most interested

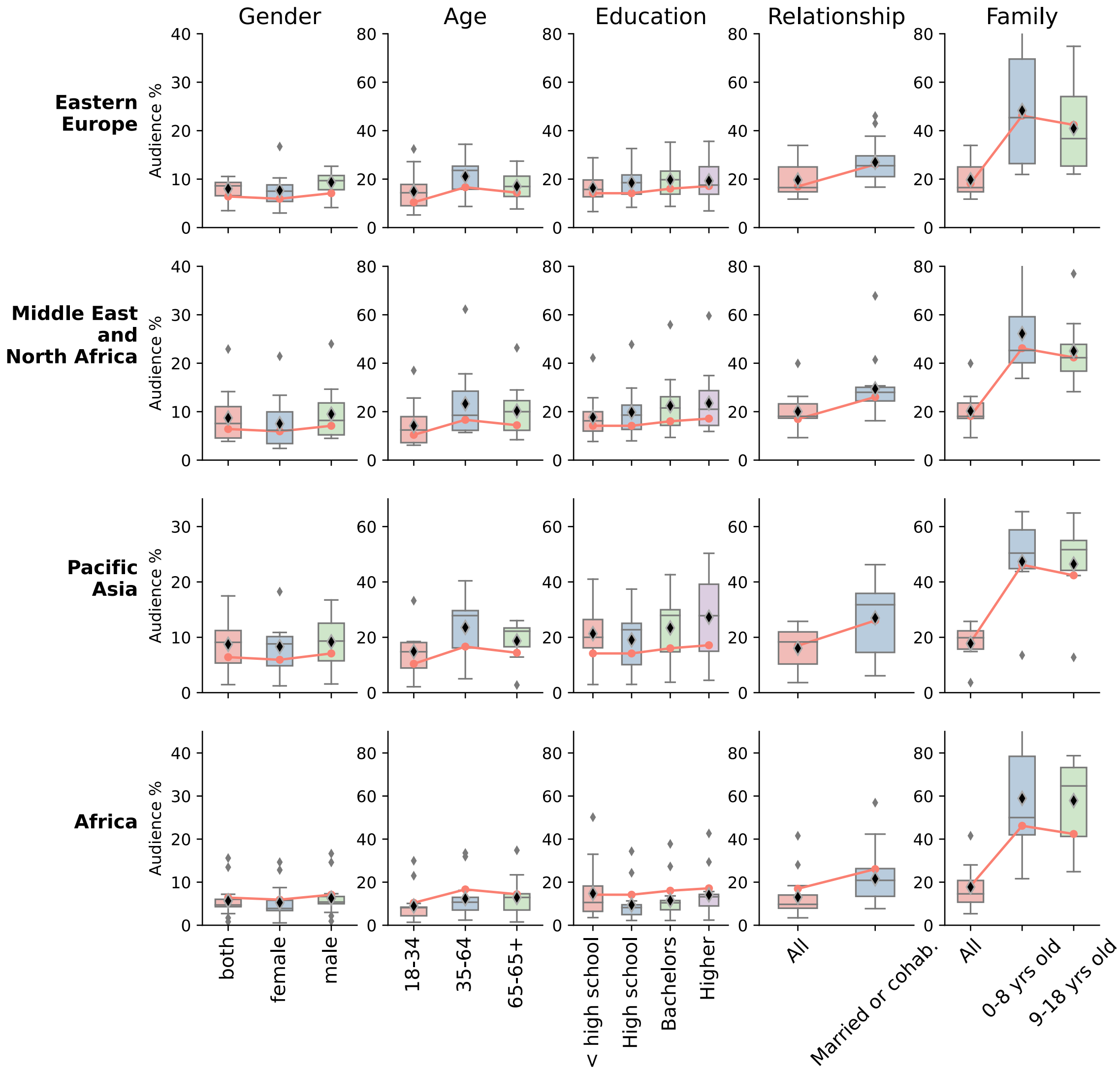

Figure 4. AC interest of the social media audience in Eastern Europe and Balkans (EEU), Middle East and North Africa (MEA), Pacific Asia (PAS), and Africa (AFR) and its relation to main demographic factors

Each row shows the distribution of audience fractions across 14, 14, 7, and 17 countries of the EEU, MEA, PAS, and AFR, respectively. Table S3 lists these countries. The boxes show the quartile range, where black diamonds mark the mean, whiskers extend to ±50% of the interquartile range, and outliers are marked by small gray dots. The red lines connect the global mean values in each category. Table S4 shows the main statistics.

in AC adoption in these regions, except in AFR, where older populations are slightly more interested in AC than the middle aged, on average. The mean AC-interested fractions of the middle-aged audience are 58.3%, 64.2%, and 41.7% higher than that of the younger audience in PAS, MEA, and EEU, respectively, whereas this difference is 39% in AFR.

The higher the education level, the higher the interest in AC. These effects are more striking in PAS and MEA. The AC interest of the highest education level is 18%, 28%, and 33% higher than that of the lowest education level in EEU, PAS, and MEA, respectively. In AFR, even though the audience interest is low, the AC interest of post-bachelor graduates is around 50% higher than that of high school graduates, which might also reflect their higher earning capacities. In all regions, AC interest is higher among married or cohabiting audiences and those who have small children than among the general audience. In PAS and AFR, parenting young children more significantly affects AC interest than it does in other regions, where the AC-interested fraction of parents of young children is around three times higher than the AC-interested fraction of the general population.

DISCUSSION

Space cooling is a requirement for adaptation to increasing temperatures caused by climate change yet is also a mitigation challenge due to the associated energy consumption. Therefore, although the adoption of AC appliances is expected to cover the cooling gap in regions vulnerable to heat stress, it can put additional pressure on energy demand. Future estimates of global AC adoption and the subsequent energy demand are often based on limited data, which does not adequately capture the geographical and social heterogeneity of AC adoption. In this study, we address this data gap, exploring demographic and social drivers of global AC adoption using social media data.

Research and policy implications

A key finding of this study was that the fraction of the social media audience interested in AC correlates better with the actual AC purchase rates than ownership rates given that online activity often results from an interest in buying. The metric used in existing studies to analyze the extent and drivers of AC adoption is the ownership rate, that is, the fraction of households that own an AC unit. The ownership rate represents a stock that accumulates as households acquire new AC units or discard the old ones. For estimating the future dynamics of this ownership stock, purchase rates are important as the stock's inflow. Furthermore, because new AC units are likely to be more energy efficient than the older ones, purchase rates indicate how rapidly the AC stock can be transformed into an energy-efficient one. Therefore, social media data as an indicator of purchase rates can complement other data sources in estimating the future of the AC stock.

Another key finding was that demographic factors and household characteristics play an important role in AC adoption on a global scale: male, middle-aged, highly educated social media audiences in a cohabiting relationship or parenting 0- to 8-year-old children tend to be more interested in AC adoption across the world. These global findings from social media data are, with exceptions, in line with those from household surveys we examined from the US, Mexico, Italy, and India and with other studies that reached similar conclusions based on surveys from the OECD countries, Brazil, and Indonesia.[7,8]

The social heterogeneity of AC interest we observed in the social media data might affect future AC ownership in several ways. First, population growth is a main driver of global AC stock not only through the presence of more people but also through the higher AC demand of households with small children than of households with older or no children. Second, middle-aged and young adults are more interested in AC adoption than the elderly. Even though this finding implies an adaptation challenge for the vulnerable elderly population, it also means a mitigation opportunity. If young and middle-aged population cohorts acquire energy-efficient appliances today, this can curb energy consumption in the decades to come. Gender differences both in online interest and actual purchase rates are striking. Men are known to make purchases of durable goods in most households,[30] and this might explain the online gender difference in AC interest. However, female-headed households have lower income levels, especially in developing countries, and the fraction of female-headed households has been increasing.[31] Therefore, gender inequality could pose another challenge to increasing AC penetration across the world.

Our global analysis has highlighted geographic and socioeconomic differences in AC adoption. We found that the Global North countries are relatively homogeneous in terms of interest of their social media audiences in AC. Even though the effect of gender, age, and education is similar both in the Global North and South, the countries in the former are more uniformly distributed in terms of the interest in AC adoption, whereas the latter show less interest on average, except in a few wealthy countries, and a skewed distribution. An exception to the homogeneity in the Global North is observed for the oldest age cohort. The elderly social media audience of a few heat-vulnerable Eastern European (EEU) countries, such as Greece, Hungary, and Cyprus, express a higher interest in AC adoption than the same age cohort of other Global North countries, indicating that heat stress and relative prosperity might be motivating a population group that is otherwise known to be reluctant to acquire AC.

PAS, the Middle East, and Eastern Europe are regions where social media audiences have the highest interest rate in AC adoption. These are regions where heat stress has been increasing and is expected to increase further,[27] hence the high interest rates indicate a high adaptation potential. Across the demographic and household characteristic groups, these regions have a stronger AC interest than the global average interest. Considering that population, education, and income levels are still growing rapidly in these regions, meeting increasing AC demand with efficient and low-carbon cooling systems and equipping the vulnerable population with such systems is of utmost importance for mitigation and adaptation. In AFR, however, where the increase in heat vulnerability is expected to be highest later in the century as a result of climate change and population increase,[28] a very low fraction of the social media audience shows an interest in AC. Even though social media penetration is low in AFR, it is expected to be biased toward urban, high-income, and highly educated audiences, which coincides with groups that are interested in AC in the rest of the world. Therefore, considerably lower online interest in AC in AFR implies that adaptation could still be a big challenge in this region.

This study has three key findings regarding the use of social media data to investigate AC adoption. First, social media data do not shed light on AC ownership stocks but can be a proxy for the trends of ownership indicated by purchase rates. Second, social media data correspond to findings from household surveys on a similar order of magnitude and social heterogeneity, thus indicating their usability to augment the survey data for AC adoption rates. The predictive representativeness of social media data for the adoption rates cannot be evaluated on the basis of the existing surveys, though, given the inherent differences between the two data sources, such as the variables they measure or data-collection years. Third, social media data are helpful when used alongside conventional surveys to understand a social phenomenon. Surveys provide purposeful measurements of consumption, especially for definite questions such as AC ownership, whereas the social media data provide empirical support to scaling the local or national surveys up to larger samples of population and to the global level.

Usability of social media data for such purposes is not without shortcomings, though. Low social media penetration through the total population in low-income and conflict-prone countries or at high spatial resolution levels, such as for cities and towns instead of countries and regions, limits the usability of social media data. This low penetration should be considered when social media data are incorporated into any analysis, for instance, through the exclusion of countries with low penetration rates, as we did in this study.

Limitations and future work

One of the main contributions of this study is in investigating the social heterogeneity of AC adoption on a global scale based on social media data alongside household surveys from four countries. In this analysis of heterogeneity, we compared the distribution of AC adoption indicators for different categories of population, distinguishing these by age, education level, or family status as the independent variables. This categorical nature of the independent variables limited the statistical methods available to compare across datasets (social media and surveys) and to identify the importance in the order of these independent variables. Future work can employ, for instance, machine-learning methods on multivariate social media audience-size data to determine the rank order of demographic and socioeconomic factors on a global scale and in different regions. Another limitation of this comparison between datasets is the unit of investigation, that is, individuals in the social media data and households in the survey data. This can lead to an overestimation of purchase rates based on social media data given that multiple individuals from the same household might engage in online activity related to AC, especially when characteristics such as age and education level are considered.

Modeling studies can benefit from this analysis of social media data in quantifying global cooling demand projections. Social media data can provide empirical support to calibrate not only statistical extrapolation models but also descriptive simulation models based on population segmentation. Combined with demographic macrotrends and projections, such data can enable accounting for geographical and social heterogeneity in plausible scenarios of future global AC adoption and energy demand for cooling.

EXPERIMENTAL PROCEDURES

Resource availability

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by Sibel Eker (sibel.eker@ru.nl).

Materials availability

All new materials generated in this study are provided via links in the data and code availability section.

Data and code availability

The full code, results, and datasets used and generated are available on GitHub: https://github.com/sibeleker/Facebook_Airconditioning.

Excel versions of Tables S2, S4, and S6 are available at Zenodo: https://doi.org/10.5281/zenodo.7770189.

Data collection

Social media data

We use the audience-size data of Facebook and Instagram, which are publicly available to any registered advertiser through the Facebook Marketing API. These data are in an aggregate form, hence no personal information is disclosed on the Marketing API or used in this study. The audience-size data are available for specific population groups defined by demographic factors such as geographic location, age, gender, and education level and by interest categories that refer to social, economic, and cultural interests. The demographic factors are mostly defined by users upon profile creation, and interests are either user defined or inferred by algorithms depending on users' posts, apps, ad clicks, page likes, and other activities.[32] In addition to the information collected from its own platforms (Facebook, Instagram, and WhatsApp), Facebook collects data from more than 30% of the most popular websites through cookie-enabled online tracking[33] and from mobile devices through location tracking and app downloads.[34]

The interest category we use in this study is "air conditioning" (with ID 6003711009918), determined on the basis of a keyword search on the Marketing API for available interest categories about space cooling. The metric we use to refer to the AC interest is the social media audience fraction interested in AC (F_i) for a specific location i, such as a country, state, or city. The Marketing API reports two metrics for audience size: daily active users (DAUs) and the lower and upper bounds of monthly active users (MAUs). We use DAUs as a measure of audience size and define F_i as denoted in Equation 1, where DAU_i^* is the audience size with AC interest at location i and DAU_i is the total audience size in that location.

$$F_i = \frac{DAU_i^*}{DAU_i} \quad \text{(Equation 1)}$$

Social media data may not represent the entire population because they are biased toward internet users. This bias was found to be nonsignificant in some studies,[35] yet the audience segmentation data are often corrected with respect to the penetration rate, i.e., the fraction of the total population that use social media.[36,37] We do not correct the audience fraction with respect to the penetration rate given that the objective of this study is not prediction. Still, to avoid overconfidence, we exclude countries with a low penetration rate and a low audience size. Figure S5 shows the distribution of penetration rate and audience size across 136 countries. The countries where the penetration rate is below the 25th percentile (19.8%) and the total audience size is below the 25th percentile (916,652) are excluded from our global analysis. This choice leads to 21 countries being excluded and 113 being included.

Penetration rates are high in the four countries for which household surveys are compared with the social media data, meaning that the social media data represent a large fraction of subnational population in these four countries. As shown in Figure S6, the average penetration rate is 53% across the 11 regions of the US, lower than the national penetration rate as a result of mobility and change in location settings. The average penetration rate is 73.9% across the 32 states of Mexico, 68.8% across 20 regions of Italy, and 36.4% across 21 states of India.

We collected the audience-size data from the Facebook Marketing API by using a Python interface called pySocialWatcher[38] in May–September 2021 and November 2022. Given that AC interest might depend on the season, this data collection in two different seasons was to avoid the seasonality bias. Figure S8 shows the difference between the audience fraction interested in AC in May–September 2021 and November 2022. The AC interest has slightly and moderately increased in the US and Italy, which have spring/summer and autumn seasons at these two data-collection points and have higher AC ownership in general, attributed to better socioeconomic conditions. In India, the AC interest is lower in almost all states in November 2022, which is considered an autumn season. In Mexico, where November temperatures are still relatively high, AC interest shows a higher increase than spring 2021. Throughout our analysis (in Figures 1, 2, 3, and 4), we use the arithmetic mean of the audience fractions obtained from the two data-collection points (Equation 2).

$$F_i = \frac{F_{i,\text{spring}} + F_{i,\text{autumn}}}{2} \quad \text{(Equation 2)}$$

The data and analysis scripts are available online (see data and code availability).

Household surveys

To test representativeness of the social media data, we compare the AC interest among the social media audience with the AC ownership and purchases reported in five household surveys from four countries: the US, Mexico, Italy, and India. This selection of surveys was motivated by the data availability on AC specifically and the representation of a broad range of contextual conditions, such as different climatic and socioeconomic conditions and household characteristics. The surveys were conducted between 2018 and 2020, as summarized in Table 1, which leads to a gap between the surveys and the social media data-collection year (2021 and 2022).

The surveys report the number of households that own at least one air conditioner ($H_{\text{own},i}$) and those that have purchased an air conditioner in the last 2 years ($H_{\text{pur},i}$). Additionally, the Prayas survey conducted in the semi-urban areas of two Indian states reports the number of households that intend to purchase an AC unit ($H_{\text{int},i}$). The metric we use in comparisons with the social

Table 1. Summary of the five household surveys in four countries and their characteristics

	US	Italy	Mexico	India (states)	India (towns)
Survey name	Residential Energy Consumption Survey (RECS)	Household Budget Survey	National Survey on Energy Consumption in Private Homes (ENCEVI)	India Residential Energy Survey (IRES)	Prayas Energy Consumption Patterns in Households
Reference	Energy Information Administration (EIA)[39]	Instituto Nazionale di Statistica (ISTAT)[40]	National Institute of Statistics and Geography (INEGI)[41]	Council on Energy, Environment and Water (CEEW)[42]	Prayas Energy Group[43]
Year	2020	2019	2018	2020	2019
Sample size (households)	18,496	18,718	28,953	14,850	4,200
Location unit	ten divisions (nine Census Bureau divisions with the Mountain division divided into two)	20 regions (NUTS level 2)	32 states (31 states at NUTS level 1 and the capital)	21 most populous states	27 semi-urban areas in Maharashtra and Uttar Pradesh

media data is the fraction of households who recently purchased an air conditioner in a specific location i (S_i), as defined in Equation 3, where $H_{\mathrm{pur},i}$ is the number of households with a recent purchase of an air conditioner and H_i is the total number of households in location i.

$$S_i = \frac{H_{\mathrm{pur},i}}{H_i} \tag{Equation 3}$$

An alternative metric that could be used in comparison is the fraction of households that own an air conditioner, that is, $H_{\mathrm{own},i}/H_i$. Ownership stock is commonly used in several studies as the main indicator of AC adoption.[44,45] The main inflow of AC ownership stock that determines the trends over time is the purchase rate. We find that the social media data better correlate with recent purchases than with ownership (Figure S1) given that consumers engage in online activity related to AC when they are in the process of acquiring it or just after. Previous studies also found that online activity reflects the trends, that is, the rate of change in a social phenomenon.[21] Therefore, we consider social media data as a proxy for AC purchases, not ownership. For the semi-urban areas of Maharashtra and Uttar Pradesh in India, where the population sizes are smaller and income levels are lower, social media data are better indicators of intentions to purchase an air conditioner rather than actual recent purchases. Therefore, for this survey, we report the comparison of the social media data with the household fraction that "intends" to purchase an air conditioner.

The surveys report the location of respondent households at different administrative levels, such as divisions in the US or states in Mexico. The Prayas survey reports rural and semi-urban areas of two states of India. For this survey, we excluded the rural areas and focused only on semi-urban ones given the low social media penetration rates in the rural areas. Therefore, the location i is specified according to the surveys, as listed in Table 1, and the social media data are retrieved for the same location to calculate comparable metrics S_i and F_i.

Other data sources

To calculate the social media penetration rates, we used the population data collected from the following sources: global population per country from the World Bank statistics[46] and population per state for the US,[47] Italy,[48] Mexico,[49] and India[50] from the national statistics of each country. Given that the census data of India are 10 years old, we estimated the 2021 population of each state and town with a correction multiplier (1.114), which represents the estimate for the national population increase between 2011 and 2021.[51]

To investigate the correlation of AC interest with social media and income levels (Figure 3G), we obtained the GDP per capita data per country from the World Bank statistics,[52] per state of the US from the Bureau of Economic Analysis,[53] and per state of India, Brazil, and Mexico from the OECD Regional Statistics database.[54]

We analyzed the relation of AC interest and climatic conditions (Figure 3H) on the basis of the mean CDDs for each country and the states of large countries with multiple climatic zones, such as the US, India, Mexico, and Brazil. CDDs are calculated as the sum over a year of positive daily differences between the outdoor temperature and a fixed temperature (balance temperature), assumed at 26°C based on prior studies.[13] We used a 30 year time series for outdoor temperatures, based on the observed historical weather dataset EWEMBI[55] (EartH2Observe, WFDEI, and ERA-Interim data Merged and Bias-corrected for ISIMIP), combining global climate data variables from a number of sources, consistently downscaled and bias corrected. CDD calculations were run over a spatial grid at 0.5° resolution to account for infra-regional climate differences, and results subsequently aggregated at state and country levels (E. Byers, personal communication).

Data analysis

For confidence building, we compare the social media data with the household surveys by visualizing the distribution of both metrics (F_i and S_i) and calculating the correlation coefficients between them. We calculate the Pearson correlation coefficient (r) on the basis of the absolute values and the Spearman (ρ) rank-correlation coefficient given that F_i and S_i are not directly comparable metrics, and the rank order of states or towns in the two datasets can be a better indicator. These correlation coefficients are calculated with the Python package *scipy.stats* v.1.6.2.[56]

To compare the distribution of survey and social media data (Figure 1), we use two-sided Wilcoxon signed-rank test.[57] The null hypothesis (H_0) of this test is that the distribution of the two samples is equal, and in particular, it tests whether the distribution of differences ($D_i = F_i - S_i$) is symmetric around zero based on the signed ranks of these differences. Equation 4 denotes this test, where $\tilde{\mu}_D$ is the median of differences. We assumed the level of significance to be 0.05 ($\alpha = 0.05$). A p value smaller than 0.05 means that the null hypothesis is rejected, and the differences between the two samples are not symmetric around zero. Therefore, in our study, p values higher than 0.05 indicate a statistically significant similarity between the distribution of the survey and social media data. The test statistic and p values are calculated with the Python package *scipy.stats* v.1.6.2,[56] and they can be seen in Table S5.

$$\begin{aligned} H_0 &: \tilde{\mu}_D = 0 \\ H_a &: \tilde{\mu}_D \neq 0 \end{aligned} \tag{Equation 4}$$

To compare the online AC interest (F_i) and purchase rates (S_i) between different demographic groups (Figure 2), we use Wilcoxon signed-rank test again with an alternative hypothesis indicating the direction of asymmetry around 0. The results of these tests can be seen in Table S6. For gender (Equation 5), the difference for each data source (s: soc, sur) and each region (i) is the difference between the data points (d) for males and females. The alternative hypothesis (H_a) is that males tend to show a higher online interest in AC than females or that the households with a male main income earner tend to have purchased AC more than households with a female main income earner, hence the median of the differences is greater than zero. A p value (p_s) smaller than

0.05 means that the null hypothesis is rejected, and males tend to show a higher interest or purchase rate than females.

$$D_{s,i} = d_{s,i,\text{male}} - d_{s,i,\text{female}} \quad H_0 : \tilde{\mu}_{D,s} = 0 \quad H_a : \tilde{\mu}_{D,s} > 0 \tag{Equation 5}$$

For age, we compare the 18- to 34-age cohort with the 35–64 cohort (Equation 6) and compare the 35–64 cohort with the 65+ cohort (Equation 7). The alternative hypothesis for each case is that the younger age cohort has a higher interest or purchase rate than the older cohort. A p value for the first case ($p_{s,1}$) smaller than 0.05 means that the 18–34 cohort tends to show a higher interest or purchase rate than the 35–64 cohort. A statistically significant p value ($p_{s,2}$) indicates the same relationship between the 35–64 and 65+ cohorts.

$$D_{s,i,1} = d_{s,i,18-34} - d_{s,i,35-64} \quad H_0 : \tilde{\mu}_{D,s,1} = 0 \quad H_a : \tilde{\mu}_{D,s,1} > 0 \tag{Equation 6}$$

$$D_{s,i,2} = d_{s,i,35-64} - d_{s,i,65+} \quad H_0 : \tilde{\mu}_{D,s,2} = 0 \quad H_a : \tilde{\mu}_{D,s,2} > 0 \tag{Equation 7}$$

For education, we test the alternative hypothesis that the AC interest or purchase rate is higher at higher education levels. Specifically, we compare the high school graduates with those with less than high school education (case 1; Equation 8) and compare those with a university degree or higher with high school graduates (case 2; Equation 9). Because the education group categorization is different across the countries, in case 1, we compare the university graduates with the high school graduates in the US and Italy and compare the higher (masters and PhD) education graduates with university graduates in case 2. A p value ($p_{s,1}$ and $p_{s,2}$) smaller than 0.05 indicates a statistically significant positive relationship between the education level and AC adoption indicators.

$$D_{s,i,1} = d_{s,i,\text{high school}} - d_{s,i,\text{less than high school}} \quad H_0 : \tilde{\mu}_{D,s,1} = 0 \quad H_a : \tilde{\mu}_{D,s,1} > 0 \tag{Equation 8}$$

$$D_{s,i,2} = d_{s,i,\text{university}} - d_{s,i,\text{high school}} \quad H_0 : \tilde{\mu}_{D,s,2} = 0 \quad H_a : \tilde{\mu}_{D,s,2} > 0 \tag{Equation 9}$$

Because the social media data are categorical, that is, the audience size is provided for a discrete set of age cohorts, education groups, or family characteristics, we analyze the global data by visualizing its distribution in each category and report the descriptive statistics. We discuss the main findings based on the comparison of mean values across the countries or states in each category. To analyze the relationship between the audience fraction interested in AC (F_i) and GDP per capita or CDDs, we report linear regression results, alongside the p values that test the null hypothesis of regression coefficients being zero, and the Pearson correlation coefficients. Linear regression does not result in an accurate fit between the two datasets, as shown by the relatively low correlation values and high p values of regression. However, the residues of linear regression show (Figure S3) that a nonlinear regression model does not represent the distribution of data better. Therefore, we use linear regression to derive the direction and scale of association rather than form a statistical model.

All box and violin plots were created with *seaborn* v.0.11.1,[58] and the maps were plotted with *geopandas* v.0.6.1.[59] All statistics, including correlation coefficients and linear regression coefficients, were calculated with *scipy.stats*. Further specifications of analysis can be found in the figure captions.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.oneear.2023.03.011.

ACKNOWLEDGMENTS

We thank Dr. Edward Byers from the International Institute for Applied Systems Analysis for providing the data for standard CDDs by country and state. This research was supported by European Union's Horizon 2020 and Horizon Europe research and innovation programs under grant agreement nos. 821124 (NAVIGATE) and 101081661 (WorldTrans).

AUTHOR CONTRIBUTIONS

All authors contributed to the design of the study. S.E. collected the social media data and conducted the analysis with contributions from A.M. in processing and analyzing the survey data. S.E. drafted the manuscript. All authors contributed to writing the final version of the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

One or more of the authors of this paper self-identifies as a gender minority in their field of research. One or more of the authors of this paper self-identifies as an underrepresented ethnic minority in their field of research or within their geographical location.

Received: July 21, 2022
Revised: August 30, 2022
Accepted: March 27, 2023
Published: April 21, 2023

REFERENCES

1. International Energy Agency (2018). The future of cooling. https://www.iea.org/reports/the-future-of-cooling.
2. International Energy Agency (2021). Cooling. https://www.iea.org/reports/cooling.
3. Levesque, A., Pietzcker, R.C., Baumstark, L., De Stercke, S., Grübler, A., and Luderer, G. (2018). How much energy will buildings consume in 2100? A global perspective within a scenario framework. Energy *148*, 514–527. https://doi.org/10.1016/j.energy.2018.01.139.
4. Arias, P.A., Bellouin, N., Coppola, E., Jones, R.G., Krinner, G., Marotzke, J., Naik, V., Palmer, M.D., Plattner, G.-J., and Rogelj, J. (2021). Technical summary. In Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, V. Masson-Delmotte, et al., eds. (Cambridge University Press), pp. 33–144. https://doi.org/10.1017/9781009157896.002.
5. van Ruijven, B.J., De Cian, E., and Sue Wing, I. (2019). Amplification of future energy demand growth due to climate change. Nat. Commun. *10*, 2762. https://doi.org/10.1038/s41467-019-10399-3.
6. Khosla, R., Miranda, N.D., Trotter, P.A., Mazzone, A., Renaldi, R., McElroy, C., Cohen, F., Jani, A., Perera-Salazar, R., and McCulloch, M. (2020). Cooling for sustainable development. Nat. Sustain. *4*, 201–208. https://doi.org/10.1038/s41893-020-00627-w.
7. De Cian, E., Pavanello, F., Randazzo, T., Mistry, M.N., and Davide, M. (2019). Households' adaptation in a warming climate. Air conditioning and thermal insulation choices. Environ. Sci. Pol. *100*, 136–157. https://doi.org/10.1016/j.envsci.2019.06.015.
8. Pavanello, F., De Cian, E., Davide, M., Mistry, M., Cruz, T., Bezerra, P., Jagu, D., Renner, S., Schaeffer, R., and Lucena, A.F.P. (2021). Air-conditioning and the adaptation cooling deficit in emerging economies. Nat. Commun. *12*, 6460. https://doi.org/10.1038/s41467-021-26592-2.
9. Colelli, F.P., and Cian, E.D. (2020). Cooling demand in integrated assessment models: a methodological review. Environ. Res. Lett. *15*, 113005. https://doi.org/10.1088/1748-9326/abb90a.
10. Brutschin, E., Pianta, S., Tavoni, M., Riahi, K., Bosetti, V., Marangoni, G., and van Ruijven, B.J. (2021). A multidimensional feasibility evaluation of low-carbon scenarios. Environ. Res. Lett. *16*, 064069. https://doi.org/10.1088/1748-9326/abf0ce.

11. Trutnevyte, E., Hirt, L.F., Bauer, N., Cherp, A., Hawkes, A., Edelenbosch, O.Y., Pedde, S., and van Vuuren, D.P. (2019). Societal transformations in models for energy and climate policy: the ambitious next step. One Earth *1*, 423–433. https://doi.org/10.1016/j.oneear.2019.12.002.

12. Rao, N.D., and Wilson, C. (2021). Advancing energy and well-being research. Nat. Sustain. *5*, 98–103. https://doi.org/10.1038/s41893-021-00775-7.

13. Mastrucci, A., Byers, E., Pachauri, S., and Rao, N.D. (2019). Improving the SDG energy poverty targets: residential cooling needs in the Global South. Energy Build. *186*, 405–415. https://doi.org/10.1016/j.enbuild.2019.01.015.

14. Rolnick, D., Donti, P.L., Kaack, L.H., Kochanski, K., Lacoste, A., Sankaran, K., Ross, A.S., Milojevic-Dupont, N., Jaques, N., Waldman-Brown, A., et al. (2022). Tackling climate change with machine learning. ACM Comput. Surv. *55*, 42. https://doi.org/10.1145/3485128.

15. Chunara, R., Bouton, L., Ayers, J.W., and Brownstein, J.S. (2013). Assessing the online social environment for surveillance of obesity prevalence. PLoS One *8*, e61373. https://doi.org/10.1371/journal.pone.0061373.

16. Kryvasheyeu, Y., Chen, H., Obradovich, N., Moro, E., Van Hentenryck, P., Fowler, J., and Cebrian, M. (2016). Rapid assessment of disaster damage using social media activity. Sci. Adv. *2*, e1500779. https://doi.org/10.1126/sciadv.1500779.

17. Ilieva, R.T., and McPhearson, T. (2018). Social-media data for urban sustainability. Nat. Sustain. *1*, 553–565. https://doi.org/10.1038/s41893-018-0153-6.

18. Garcia, D., Mitike Kassa, Y., Cuevas, A., Cebrian, M., Moro, E., Rahwan, I., and Cuevas, R. (2018). Analyzing gender inequality through large-scale Facebook advertising data. Proc. Natl. Acad. Sci. USA *115*, 6958–6963. https://doi.org/10.1073/pnas.1717781115.

19. Liao, Y., Yeh, S., and Jeuken, G.S. (2019). From individual to collective behaviours: exploring population heterogeneity of human mobility based on social media data. EPJ Data Sci. *8*, 34. https://doi.org/10.1140/epjds/s13688-019-0212-x.

20. Liao, Y., Yeh, S., and Gil, J. (2021). Feasibility of estimating travel demand using geolocations of social media data. Transportation *49*, 137–161. https://doi.org/10.1007/s11116-021-10171-x.

21. Eker, S., Garcia, D., Valin, H., and van Ruijven, B. (2021). Using social media audience data to analyse the drivers of low-carbon diets. Environ. Res. Lett. *16*, 074001. https://doi.org/10.1088/1748-9326/abf770.

22. Paulhus, D.L., and Vazire, S. (2007). The self-report method. In Handbook of research methods in personality psychology (The Guilford Press), pp. 224–239.

23. Chan, D. (2009). So why ask me? Are self-report data really that bad? In Statistical and Methodological Myths and Urban Legends: Doctrine, Verity and Fable in the Organizational and Social Sciences (Routledge/Taylor & Francis Group), pp. 309–336.

24. Department for Levelling Up Housing and Communities (2022). English Housing Survey data on owner occupiers, recent first time buyers and second homes. https://www.gov.uk/government/statistical-data-sets/owner-occupiers-recent-first-time-buyers-and-second-homes.

25. National Association of Realtors (2021). Profile of home buyers and sellers. https://cdn.nar.realtor/sites/default/files/documents/2021-highlights-from-the-profile-of-home-buyers-and-sellers-11-11-2021.pdf.

26. Anderson, M., and Perrin, A. (2017). Technology use among seniors. https://www.pewresearch.org/internet/2017/05/17/technology-use-among-seniors/.

27. Byers, E., Gidden, M., Leclère, D., Balkovic, J., Burek, P., Ebi, K., Greve, P., Grey, D., Havlik, P., Hillers, A., et al. (2018). Global exposure and vulnerability to multi-sector development and climate change hotspots. Environ. Res. Lett. *13*, 055012. https://doi.org/10.1088/1748-9326/aabf45.

28. Liu, Z., Anderson, B., Yan, K., Dong, W., Liao, H., and Shi, P. (2017). Global and regional changes in exposure to extreme heat and the relative contributions of climate and population change. Sci. Rep. *7*, 43909. https://doi.org/10.1038/srep43909.

29. García-León, D., Casanueva, A., Standardi, G., Burgstall, A., Flouris, A.D., and Nybo, L. (2021). Current and projected regional economic impacts of heatwaves in Europe. Nat. Commun. *12*, 5807. https://doi.org/10.1038/s41467-021-26050-z.

30. Organisation for Economic Cooperation and Development (2021). Gender and the environment: Building evidence and policies to achieve the SDGs. https://doi.org/10.1787/3d32ca39-en.

31. Milazzo, A., and van de Walle, D. (2015). Women left behind? (World Bank).

32. Facebook (2022). About detailed targeting. https://www.facebook.com/business/help/182371508761821?id=176276233019487.

33. Englehardt, S., and Narayanan, A. (2016). Online tracking: a 1-million-site measurement and analysis. Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security (Association for Computing Machinery).

34. Cabañas, J.G., Cuevas, Á., and Cuevas, R. (2018). Unveiling and quantifying Facebook exploitation of sensitive personal data for advertising purposes. 27th USENIX Security Symposium*18* (USENIX Security).

35. Rife, S.C., Cate, K.L., Kosinski, M., and Stillwell, D. (2016). Participant recruitment and data collection through Facebook: the role of personality factors. Int. J. Soc. Res. Methodol. *19*, 69–83. https://doi.org/10.1080/13645579.2014.957069.

36. Ribeiro, F.N., Benevenuto, F., and Zagheni, E. (2020). How biased is the population of Facebook users? Comparing the demographics of Facebook users with census data to generate correction factors. 12th ACM Conference on Web Science.

37. Kashyap, R., Fatehkia, M., Al Tamime, R., and Weber, I. (2020). Monitoring global digital gender inequality using the online populations of Facebook and Google. Demogr. Res. *43*, 779–816. https://doi.org/10.4054/DemRes.2020.43.27.

38. Araujo, M., Mejova, Y., Weber, I., and Benevenuto, F. (2017). Using Facebook ads audiences for global lifestyle disease surveillance: promises and limitations. Proceedings of the 2017 ACM on Web Science Conference.

39. US Energy Information Administration (2022). Residential Energy Consumption Survey (RECS). https://www.eia.gov/consumption/residential/data/2020/.

40. Instituto Nazionale di Statistica (ISTAT) (2019). Household Budget Survey. https://www.istat.it/en/archivio/180353.

41. National Institute of Statistics and Geography (INEGI) (2018). National survey on energy consumption in private housing units (2018). http://en.www.inegi.org.mx/programas/encevi/2018/.

42. Council on Energy, Environment and Water (2020). The India Residential Energy Survey (IRES). https://www.ceew.in/india-residential-energy-survey-ires.

43. Prayas Energy Group (2019). Energy consumption patterns in households. https://energy.prayaspune.org/our-work/article-and-blog/energy-consumption-of-households-echo-a-blog-series.

44. Mastrucci, A., Byers, E., Pachauri, S., Rao, N., and van Ruijven, B. (2022). Cooling access and energy requirements for adaptation to heat stress in megacities. Mitig. Adapt. Strateg. Glob. Chang. *27*, 59. https://doi.org/10.1007/s11027-022-10032-7.

45. Davis, L., Gertler, P., Jarvis, S., and Wolfram, C. (2021). Air conditioning and global inequality. Global Environ. Change *69*, 102299. https://doi.org/10.1016/j.gloenvcha.2021.102299.

46. The World Bank (2020). Population, total. https://data.worldbank.org/indicator/SP.POP.TOTL.

47. US Census Bureau (2019). State population totals and components of change: 2010-2019. https://www.census.gov/data/tables/time-series/demo/popest/2010s-state-total.html.

48. Instituto Nazionale di Statistica (ISTAT) (2021). Resident population on 1st January. http://dati.istat.it/?lang=en.

49. National Institute of Statistics and Geography (INEGI) (2020). Population by federal state. http://en.www.inegi.org.mx/programas/ccpv/2020/.

50. Census India (2011). Population enumeration. https://censusindia.gov.in/2011census/population_enumeration.html.
51. United Nations (2019). World population Prospects 2019 (UN Department of Economic and Social Affairs, Population Division). population.un.org.
52. The World Bank (2020). GDP per capita (current US$). https://data.worldbank.org/indicator/NY.GDP.PCAP.CD?name_desc=false.
53. US Bureau of Economic Analysis (2021). GDP per state in 2021, current US$. https://www.bea.gov/data/gdp/gdp-state.
54. Organisation for Economic Cooperation and Development (2020). Gross domestic product, large regions TL2. https://stats.oecd.org/Index.aspx?DataSetCode=RGION_ECONOM.
55. Lange, S. (2019). EartH2Observe, WFDEI and ERA-interim data merged and bias-corrected for ISIMIP (EWEMBI) V.1.1. https://doi.org/10.5880/pik.2019.004.
56. Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods *17*, 261–272. https://doi.org/10.1038/s41592-019-0686-2.
57. Haghighi, A.M., and Wickramasinghe, I. (2020). Probability, Statistics, and Stochastic Processes for Engineers and Scientists (CRC Press).
58. Waskom, M. (2021). Seaborn: statistical data visualization. J. Open Source Softw. *6*, 3021. https://doi.org/10.21105/joss.03021.
59. Jordahl, K., Van den Bossche, J., Fleischmann, M., Wasserman, J., McBride, J., Gerard, J., Tratner, J., Perry, M., Garcia Badaracco, A., and Farmer, C. (2020). geopandas/geopandas: v0.6.1. Version v0.6.1. https://doi.org/10.5281/zenodo.3946761.