Visualizing Transportation Emissions

Corentin Godeau, James Grissom, Brian Gabriel Hill, Prathik R Kaundinya,
Benjamin Mouscadet, Oksana Riou

Introduction - Motivation

The current scientific consensus on climate change or global warming and its consequences is best summed up by the 2015 IPCC (Intergovernmental Panel on Climate Change), in particular the 3rd chapter[6] which emphasizes its future consequences on our planet and civilization, from rising sea levels to heatwaves and hurricanes. Human activity has warmed the world by about 1°C since pre-industrial times, and this warming is expected to increase to at least 1.5 or 2°C at minimum. This warming is caused by the emissions of greenhouse gas such as carbon dioxide[1] : over 500 billion metric tons of carbon (1835t of CO_2) have been emitted since the industrial revolution, and the emissions are likely to reach more that 1 trillion ton by 2050.

With this motivation in mind, our group is focusing on the transportation aspect of carbon emissions, and our goal is to help improve the way we visualize and think about these emissions. This paper will explain gap in current analyses that we will solve by first detailing an extensive literature survey on transportation emissions and current tools and methods for visualization. Next, we will explain our proposed method for solving the gap we have identified. We then create an experiment to test the performance of our tool, and finally we will conclude with the results of that experiment and some conclusions about our tool and the lessons we have learned.

Problem Definition

We started by looking at what tools were already created, and we found some interesting tools with a few key shortcomings. Most transportation companies offer tools on their websites to evaluate the environmental impact of a given journey, but such information remains specific to the chosen transportation and company and does not allow users to compare the alternatives. For example an airline website will simply state the emissions for a single flight, which is helpful but doesn't cause the consumer to think critically about how to reduce that number. There are a few tools that enable such comparisons with a user-friendly interface, but also have shortcomings on the scope of the analysis. For example, the mapping tool "Map My Emissions" does offer a good comparison of some transportation types but does not include planes' emissions nor show how these statistics change from year to year.

In order to give consumers a better idea of the magnitude and results of transportation in regard to carbon emissions, we have created a tool that allows a user to visually analyze the results of multiple types of transportation over a selected route of travel. Given that currently technology is not yet available to measure individual road vehicles emissions in a cost-effective manner, we will predict over generalized classes of vehicles [3]. In doing this, the user should understand the current impact of transportation as well as the future impact on the environment. Our tool should be interactive and allow users to select different origin and destination cities, modes of transport, and years for prediction.

Survey

Our literature survey consisted of both general background on how the transportation industry contributes to climate change as well as finding specific tools that would be inspiration for our project.

The transportation industry is a major emitter of greenhouse gases, some of which being even more potent than CO_2[8, 15] such as nitrous oxide which has 298 times the potency of CO_2. Here potency (or radiative forcing) is defined as the change in the flux of heat that is radiated away by earth (in $W.m^{-2}$). Along the carbon dioxide, such gases are directly responsible for global warming, to which aviation is a net contributor. Around 3.5% of the total human-induced warming can be attributed to aviation[12, 18](which is about as much as Canada), and this figure is expected to grow by 2050, possibly up to 15%. Indeed, transportation is and will likely remain the fastest growing sector, consequently leading to ever-increasing emissions. Literature agree on a likely increase of around 130% of international aviation CO_2 emissions between 2005 and 2025[11, 16]. However calculating the impact of aviation emissions on climate change is tricky ; the consumed fuel depends on many flight factors and does not scale linearly with the distance travelled[9]. Complex models do exist but have limited accuracy.

The second chapter of 2015 IPCC report[16] emphasizes the critical need to reduce transportation growth in order to achieve low warming goals. Electric cars and trains are not carbon free either. Depending on the country, the electricity can be produced with coal or oil power plants. In 2003, the global electricity supply sector accounted for 37.5% of global carbon dioxide emissions[17] and will likely surpass 4 Gt of carbon by 2020 (compared to 2.1 Gt of carbon in 2003). Countries relying on nuclear energy benefit from a mostly carbon-free electricity, and therefore from carbon-free trains and electric cars. On the other hand coal and oil based electricity makes electric transportation contribute to climate change as much as fossil fuels based transportation.

Consequently and as emphasized again by the IPCC report[16], it is a critical matter that every single person take some action to limit their personal transportation emissions. Some people have argued in favor of carbon rationing[2, 14] and carbon taxes which would incentivize environmentally friendly behaviors. However these policies put a lot of responsibility on the people who do not necessarily have the means to evaluate their personal carbon contribution. This burden is made even heavier by the "psychological distance" to climate change[10, 13]. For one to feel concerned, it is important and helpful to have a regular reminder of the consequences of one's actions. Because climate change occurs over such a long period of time, small personal habits look innocuous.

This is why it is important to develop so-called eco-feedback technologies, in other words tools that provide feedback on individual or group behaviors with a goal of reducing environmental impact[5]. These technologies are based on the working hypothesis that people lack awareness and understanding of their personal impact, which is supported by literature[10] and that this is one of the main reasons for their lack of actions. A lot of these are focused on energy saving and resource consumption at home[7] and use the power of visualization to induce behavior change. We want to focus more on the transportation part, like Ubigreen[4] which uses mobile phones to monitor and provide feedback about transportation and commutation activities. The reported effectiveness of this tool is inspiring, but it lacks the ability to plan one's trip beforehand, and only provide a posterior feedback. Moreover, the feedback itself is not really quantified (it is either positive or negative), whereas we aim at providing the numbers themselves in order to emphasize the differences between the impacts of all transportation means.

Proposed Method - Intuition

Why should it be better than the state of the art?

Our goal is to produce an all-in-one tool evaluating the environmental impact of any trip with diverse transportation modes regardless of the company. Adding to the classical individual-oriented mapping tools our website will output general information and trends reflective of the population. Our project seeks to create a tool that will allow for comparison between both types of transportation methods as well as showing trends on the popularity of methods.

With our tool, individuals concerned about their environmental impacts and traveling often enough will be able to monitor and try to decrease their contribution to climate change. It is hoped that explaining personal carbon emission output based on transport mode chosen will allow users to choose an efficient mode of transport between cities. Moreover, informing users of emission levels on a broader level (e.g. forecasting total emissions on direct flights between a pair of cities) serves to show that the pollution caused by air travel will only be exacerbated by the growth in air travel between cities.

To measure the usefulness of our website we will enable users to inform us about the helpfulness of the provided information thanks to a feedback feature.

Proposed Method - Approach

After the midterm evaluation of our project and paper, the team is using the following schedule:

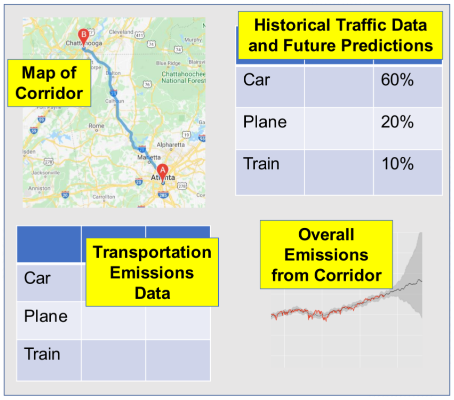
Week 9-10: Database search for all (campus closing and classes moving online disrupted activities these two weeks)

Week 11-12: Website and basic map set-up for Corentin and Brian / Data clean-up for Prathik, Benjamin, and Oksana. Midterm progress report for James and Brian

Week 13-14: Developing frontend for Corentin, Oksana and Benjamin / Developing backend for James, Brian and Prathik.

Week 15-16: Connecting frontend and backend for Corentin, Oksana and Prathik and / Final report and software documentation for James, Benjamin and Brian.

Our project proposal presentation contained the following concept for our tool:



We have obtained flight traffic data from the Bureau of Transport Statistics (bts.gov) datasets site for the years 2010-2019; this would allow for a larger training set to provide more accurate forecasting capabilities for our model.

Our model building consisted of these steps:

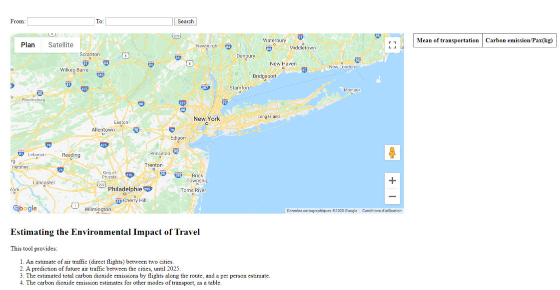
- Writing a script to download and parse transportation data from the U.S. Bureau of Transportation Statistics (www.bts.gov)
- Design the front-end of the site - the inputs would be the origin city and destination city, and the model would display the distance, historic air traffic rates, projections of traffic and carbon emissions due to aviation. It would also display estimated emissions for other vehicles (different car classes and train).
- Develop the backend to accept arguments, generate the AR model and pass values to the front-end as necessary with a Flask framework.

Data Preprocessing:
For the data parsing, air traffic data from 10 past years (2010-2019) were downloaded from the US DOT travel statistics database. The data consisted of monthly breakdowns for each direct active route between airport pairs in the US. While not lacking data in any time period, there is extraneous data that is irrelevant to our tool's objective. As a first, the flight routes with zero passengers (freight routes) were filtered out. Then, small planes (filtered by seating capacity<20 to eliminate private jet flights) were filtered out. For our annual forecasting problem, data from all months were aggregated into another table containing only cumulative yearly data.

Another problem we had to solve was how to aggregate information for a city with multiple airports (such as New York City) and map a city to an airport code. This was done with the help of an auxiliary dataset that contained coordinates of ~27,000 airports in the U.S. along with their coordinates. We decided to select all airports within a range of ± 0.4°of latitude and longitude (translates to a circle with radius of 28 miles from the input city coordinates). The flights to/from within this radius for both origin and destination were aggregated to measure total intercity passenger traffic.

Basic Website:
The first step was to learn how to use the Google Maps API. Our basic goal was to show a map with directions between two cities. Using the very clear tutorials available on the Google Developers website, we managed to display a map on our frontend. Then, using the Google Routes API, we queried for the routes between two given locations. Improving the work we already had, we added the possibility for the user to enter an origin and a destination, and a query was sent to the Google Routes API. Cities could be qualified by a state name to enhance accuracy. If a route was found, we displayed it on the map. When querying such routes, an interesting point is that the API sends back additional metadata that can be useful for our end goal such as the distance between the two locations. Once we had a semi-finalized version of the backend, Flask was set up to enable communication between the backend and the frontend. Flask is a micro web framework written in Python that does not require particular tools or libraries. It has no database abstraction layer, form validation, or any other components where pre-existing third-party libraries provide common functions. We have a flask application running on a dedicated server. This method was chosen for ease of use as it is a standard library for backend communication.

Here is an example of the front-end interface of the project, prior to a city pair query being made:



Predictive Model:
Given a specific origin and a specific destination we have been able to extract from the collected data the total number of people who traveled by plane between this origin and this destination for different years. The goal is to predict how this number is going to change for coming years. The models that we have implemented to compute such predictions are auto-regressive (AR) models which are often used to depict time-varying mechanisms in nature or society. An AR model is characterized by an order p and uses past values to compute a new value as follows: $Y_t = \sum_{i=1}^{p} \varphi_i Y_{t-i}$ where $\varphi_i, ..., \varphi_p$ are the p parameters of the model. We thus used past statistics to compute the parameters of AR models and used these parameters to predict future statistics. For instance, here are past statistics about the number of people traveling by plane from New-York to Boston:

2015	2016	2017	2018	2019
558, 854	544, 618	464, 539	415, 641	371, 359

And here are statistics for coming years predicted using an AR model of order 3:

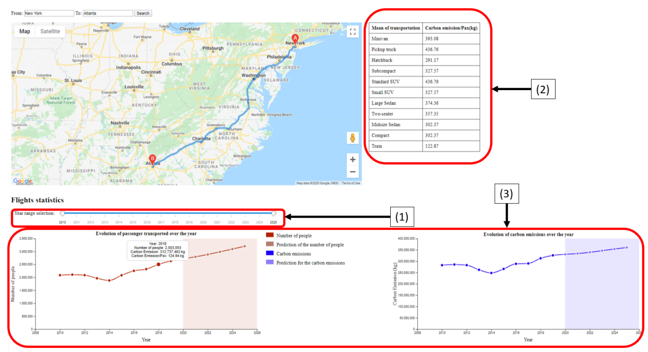
2020	2021	2022	2023	2024
328, 153	292, 368	259, 675	230, 677	205, 052

Computation of carbon emissions:
To compute the carbon emissions produced by a passenger traveling by plane between a given origin and a given destination during a year we used the following method:

- Obtain the seating capacities of planes used to operate on the query segment.
- For each type of aircraft divide the number of seats corresponding to monthly statistics by the seating capacity of the aircraft.
- Use that information to obtain the exact number and types of aircrafts which operated on the query segment during the year.
- Approximate the fuel consumption model of each aircraft by using data mapping aircraft type to fuel consumption values for different distances.
- Determine the fuel consumption in kg of each aircraft knowing the air distance between cities.
- Convert this fuel consumption to amount of CO_2 emission with a standard factor (3.15 kg of CO_2 for 1 kg of fuel).

An important note is that this would yield a worst-case upper bound of emissions - this is because the load factor (number of passengers/number of seats) on these flights may vary drastically, but are almost always <1. As such, the CO_2 emission per person would approach the best-case scenario on high-demand routes (e.g. NYC - Chicago).

Display of Data:
After computing the required information, the data is displayed on the front-end with simple layout for ease of interpretability. The key elements of the output are shown in the figure below:



In the above figure, (1) corresponds to an interactive slider that allows the user to adjust the range of years as necessary to understand the trend in aviation emissions over years. (2) is the main computation result plots: The shaded portions correspond to future emissions made by the AR model. There is also a mouseover tooltip that displays further relevant information in each year such as the number of people traveling, the cumulative CO_2 emission for that year and a per-person estimate made by the described method. (3) is the emission estimate for other modes of transport (different classes of cars and train). An important note is that the car estimates are based on a single passenger assumption, and fuel efficiency could be improved drastically by increasing the number of passengers transported. Also, the train estimate is based on an average national load factor in 2019 for Amtrak trains of 54.6%.

Experiments - Testbed

The major experimental goal of this project is to provide users with insight into the impact of travel on carbon emissions. The three major questions we would help the user answer are:

- What is my personal footprint if I choose a certain mode of transport between cities?
- If I choose public transport, which is a more economical option? (plane/train).
- In the big picture, how much will emissions vary over the next few years if we don't implement

energy efficiency plans? (assuming status quo in terms of regulations, technology, etc.)

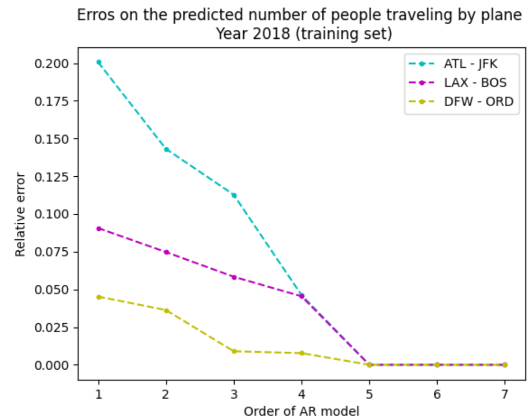
For us, as developers, we seek to answer the following questions:

- What are the shortcomings of the AR model we have chosen, and what are the pitfalls in this method of estimation?
- How do we quantify model performance, and choose which order of AR model to use for predictions?
- What are the sources of error in our prediction? Can we estimate their effect on the output?

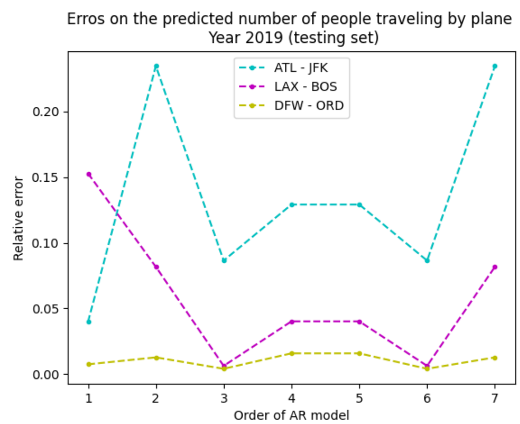
Experiments - Observations

So far, we observed that the data sets are relatively clean, but contains extraneous information. Filtering it out based on our protocol and aggregating by origin-destination pairs yields the required information. One of the challenges that we faced was the fact that there were non-standard airplane identifier codes between data sets. In order to solve this problem we had to either discover a mapping between these codes through research or observations, or somehow aggregate or average these values, which would slightly reduce the overall accuracy of our model. We chose to do the former, as it is likely that these codes were changed by the publishers of the respective databases purposely, and the aggregation of these values would still provide valuable information for our project.

Another critical aspect of the model was deciding the order of the AR model we would generate. We expected this to depend on the amount of training data available, and how far ahead we would want to make predictions in our tool. To ascertain this, we performed a train-test validation on the years 2019 and 2018 for 3 random sample routes (Atlanta to NYC, Los Angeles to Boston and Dallas to Chicago). The data from years [2010,2018] comprised of the training set, while data from 2019 was holdout for the test. We were successfully able to quantify the relative error of prediction in the training and test sets. The training set performance (how well the AR model fits the training data) is summarized below:



The test performance to measure predictive capability(2019) is shown below:



It is observed that AR models of order 3 consistently provided the best result - we believe this is because it is the optimal result from the bias-variance tradeoff. The highest order models fit the training data perfectly, but perform poorly with prediction; the lower order models consistently under-fit the data leading to poor predictions. This observation motivated us to display results from the third order AR model on the site. As such, we believe that using data from even before would necessitate an increase in AR model order because of greater data range - however, this may also be detrimental to model performance because of lack of relevance over such a long time span.

The model also predicts a slight decrease in air traffic between some segments, which is a trend that has been observed on some routes over the past few years. However, the majority of random routes tested showed an increase in passenger air travel leading up to 2020 and predicts a further rise as well. We also seek to identify some sources of error in our calculations and prediction. An inherent weakness on modeling future air traffic is the inability to forecast fluctuations based on some large-scale global events such as the current COVID-19 pandemic that would greatly reduce worldwide air traffic in 2020. As such, the predictive method smoothly extrapolates the data to future years. Nevertheless, it produces a reasonable estimate barring such drastic events that alter the pattern of air travel. Another potential minor source of error is the different seating configurations operated by different airlines on different planes. For instance, Delta Airlines operates the 777-200 LR plane with 273 passengers in a 3-class configuration, but American airlines uses 290 seats: a ~5% difference in seating capacity. As mentioned in our project proposal, we also conducted user feedback survey to understand potential methods to improve our tool based on what users would like to see.

Peer Survey Results

In order to gain knowledge about the strengths and weaknesses of our tool, we published a survey that was sent to 12 different participating members. The survey includes two questions with a 1-5 rating of user interface and data usefulness where a rating of 1 is not intuitive and not helpful respectively, and a 5 is very intuitive and helpful data.

The survey also includes four mandatory short answer questions: "What part of our tool did you like the most?", "What other data would you like displayed?", "How can we improve the user interface?" and "What additional functionality would you like to see?". There was also a non-mandatory short answer question to ask for general thoughts and improvements that could be made.

On user interface, we received an average rating of 4 out of 5. On the usefulness of the data, we received an average rating of 4.25 out of 5. Given the feedback from the short answer questions, most users seemed to have thought that the visualization was a strong suit of our tool. We received positive comments on the map API as well as the ability to compare the different modes of transportation. Many of the users wanted to see more on the significance of the data. Some of the given ideas were display how many number of trees could offset these carbon emissions, using predictions for different types of vehicles and not just planes, and a handful of users wanted to see where this data was coming from.

For the data vizualization, there were a few small errors on different browsers with overlapping text and tables, but most users expressed interest in knowing what the tool did before entering cities, as there was not a lot of explanation given unless the user was told beforehand. Another small addition that would be nice is a loading bar as the backend takes some time to compute.

Overall most users reported a positive experience, most requesting small changes that are easy to implement. After receiving feedback we added a slider that allowed the users to select the years to predict for plane emissions. This change was received positively.

Conclusions

We have achieved our target of assessing and predicting the carbon emission impact of aviation, along with other modes of transport for comparison. We are looking to compare the forecasts of several models to approximate these values, and hope that our tool can be used to highlight the environmental impact of aviation, and promote environmental awareness in society.

Our tool allows the user to evaluate their transportation choices for a variety of modes of transportation, and interactively select future predictions to gain knowledge about the severity of carbon emissions within different routes of travel.

In the future, it would be beneficial for this tool to be more robust to different modes of transportation in order to provide the user with a deeper level of understanding of their personal carbon footprint as well as trends across the population.

At a high level, the user should be able to interactively select scenarios such as changes in air traffic, road patterns, additions in public transportation, etc. in order to visualize the potential changes in carbon emissions. Additionally, the user should be able to understand the impact of the amount of carbon that our tool displays for different modes of transport in an intuitive fashion.

The peer survey results have provided us with valuable information on how the user interface can be improved in the future in order to express data in a cleaner fashion and make the tool more intuitive. We also received validation that the data we are providing is interesting to our peers, and that the user interface is at a satisfiable level.

Distribution of Team Member Effort

All team members have made equally substantial contributions to this project.

References

[1] Myles R Allen, David J Frame, Chris Huntingford, Chris D Jones, Jason A Lowe, Malte Meinshausen, and Nicolai Meinshausen. 2009. Warming caused by cumulative carbon emissions towards the trillionth tonne. *Nature* 458, 7242 (2009), 1163–1166.

[2] Tina Fawcett. 2004. Carbon rationing and personal energy use. *Energy & Environment* 15, 6 (2004), 1067–1083.

[3] Ye Feng, Don Fullerton, and Li Gan. 2005. *Vehicle choices, miles driven, and pollution policies.* Technical Report. National Bureau of Economic Research.

[4] Jon Froehlich, Tawanna Dillahunt, Predrag Klasnja, Jennifer Mankoff, Sunny Consolvo, Beverly Harrison, and James A Landay. 2009. UbiGreen: investigating a mobile tool for tracking and supporting green transportation habits. In *Proceedings of the sigchi conference on human factors in computing systems.* 1043–1052.

[5] Jon Froehlich, Leah Findlater, and James Landay. 2010. The design of eco-feedback technology. In *Proceedings of the SIGCHI conference on human factors in computing systems.* 1999–2008.

[6] O Hoegh-Guldberg, D Jacob, M Taylor, M Bindi, S Brown, I Camilloni, and G Zhou. 2018. „Impacts of 1.5 C Global Warming on Natural and Human Systems "in Global Warming of 1.5 C: An IPCC Special Report on the impacts of global warming of 1.5 C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty. *V. MassonDelmotte, P. Zhai, HO Pörtner, D. Roberts, J. Skea, PR Shukla, A. Pirani, W. Moufouma-Okia, C. Péan, R. Pidcock, S. Connors, JBR Matthews, Y. Chen, X. Zhou, MI Gomis, E. Lonnoy, T. Maycock, M. Tignor, T. Waterfield (eds.)* (2018), 175–311.

[7] David Holstius, John Kembel, Amy Hurst, Peng-Hui Wan, and Jodi Forlizzi. 2004. Infotropism: living and robotic plants as interactive displays. In *Proceedings of the 5th conference on Designing interactive systems: processes, practices, methods, and techniques.* 215–221.

[8] Atul K Jain, Bruce P Briegleb, K Minschwaner, and Donald J Wuebbles. 2000. Radiative forcings and global warming potentials of 39 greenhouse gases. *Journal of Geophysical Research: Atmospheres* 105, D16 (2000), 20773–20790.

[9] Christian N Jardine. 2005. Calculating the environmental impact of aviation emissions. *Environmental Change Institute, Oxford University Centre for Environment* (2005).

[10] Anja Kollmuss and Julian Agyeman. 2002. Mind the gap: why do people act environmentally and what are the barriers to pro-environmental behavior? *Environmental education research* 8, 3 (2002), 239–260.

[11] Andrew Macintosh. 2008. Overcoming the barriers to international aviation greenhouse gas emissions abatement. *Air and Space Law* 33, 6 (2008), 403–429.

[12] Andrew Macintosh and Lailey Wallace. 2009. International aviation emissions to 2025: Can emissions be stabilised without restricting demand? *Energy Policy* 37, 1 (2009), 264–273.

[13] Rachel I McDonald, Hui Yi Chai, and Ben R Newell. 2015. Personal experience and the 'psychological distance'of climate change: An integrative review. *Journal of Environmental Psychology* 44 (2015), 109–118.

[14] Gillbert E Metcalf and David Weisbach. 2009. The design of a carbon tax. *Harv. Envtl. L. Rev.* 33 (2009), 499.

[15] Stephen A Montzka, Edward J Dlugokencky, and James H Butler. 2011. Non-CO 2 greenhouse gases and climate change. *Nature* 476, 7358 (2011), 43–50.

[16] Joeri Rogelj, Drew Shindell, Kejun Jiang, Solomone Fifita, P Forster, Veronika Ginzburg, Collins Handa, Haroon Kheshgi, Shigeki Kobayashi, Elmar Kriegler, et al. 2018. Mitigation pathways compatible with 1.5 C in the context of sustainable development. (2018).

[17] Ralph EH Sims, Hans-Holger Rogner, and Ken Gregory. 2003. Carbon emission and mitigation cost comparisons between fossil fuel, nuclear and renewable energy resources for electricity generation. *Energy policy* 31, 13 (2003), 1315–1326.

[18] John Whitelegg and Nick Williams. 2000. *The plane truth: aviation and the environment.* Ashden Trust.