THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Fuel choice, fuel switching and improved cook stoves in Vietnamese households: Analysis, models and proposals for new solutions

NIKLAS VAHLNE

Department of Energy and Environment
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2015

Fuel choice, fuel switching and improved cook stoves in Vietnamese households: Analysis, models and proposals for new solutions

NIKLAS VAHLNE
ISBN: 978-91-7597-215-2

Doktorsavhandlingar vid Chalmers teknisks högskola
NR: 3896
ISSN 0346-718X

Energy Technology
Department of Energy and Environment
Chalmers University of Technology
SE-412 96 Gothenburg
Sweden
Telephone: +46 (0) 31 772 10 00

Printed by Chalmers Reproservice
Gothenburg, Sweden 2015

Abstract

A majority of rural households in the developing world use solid biomass fuels for cooking. This use has severe negative health effects, is often either expensive or time consuming, and contributes to global warming. Options for policy interventions include the promotion of improved cook stoves (ICS) and enabling households to switch to more modern fuels, like liquefied petroleum gas.

The main aims of this thesis is 1) to explore whether rural households' fuel use can be modeled in new ways that focus on prediction, 2) to investigate whether area level differences in fuel use may have impacts for ICS programs, and 3) to address new solutions for ICS programs in areas where the current fuel use is mainly collected biomass.

Methods used to model fuel use are ordinary linear regression and a machine learning algorithm called Random Forest. Further models are developed in order to evaluate possible implications and proposed solutions for ICS programs based on variations in current fuel use. The papers use data from two different surveys. The first data set is from a survey, carried out in the Vĩnh Phúc province of northern Vietnam in 2010. The second survey is representative of most of rural Vietnam and was collected in 2002, 2005, and 2008 as part of the Vietnam Rural Electrification program.

The results from the regression and Random Forest analysis include new ways to model fuel use, enabling easy and accurate prediction. The results also provide possible alternative explanations for some previous modeling results. The modeling of stove interventions reveals large potential differences between communes, as well as a possible non-linear relationship between stove efficiencies and benefits, but also large uncertainties in estimations depending assumptions of fuel choice. Lastly, a model for a new type of ICS program that offers possibilities to overcome some of the barriers to adoption and sustained use reported by previous studies is evaluated.

Combining the conclusions from the respective papers, a possibility of modeling variations in possible outcomes for stove programs, and the effects of such programs based on area descriptions, is found. However, further research is needed in order to make more robust estimations.

Keywords: fuel ladder, fuel choice, rural households, developing, Vietnam, cooking, improved stoves

List of papers

The thesis is based on the following appended papers:

I. **Rurality as a policy relevant measure for commune level differences in fuels used for cooking in rural Vietnam**
Niklas Vahlne, Erik O. Ahlgren. Submitted for publication

II. **Policy implications for improved cook stove programs – a case study of the importance of village fuel use variations**
Niklas Vahlne, Erik O. Ahlgren. *Energy Policy* 66 (2014): 484-495.

III. **On LPG use in rural Vietnamese households**
Niklas Vahlne. Submitted for publication

IV. **Energy efficiency at the Base of the Pyramid: a system–based market model for improved cooking stove adoption**
Niklas Vahlne. Erik. O Ahlgren *Sustainability* 6.12 (2014): 8679-8699.

Contents

1 Introduction

1.1 Background

In the rural areas of many developing countries, a major part of the used energy comes from biomass such as wood and agricultural residues (Foell et al., 2011; IEA, 2010). However, inefficient combustion, which is the norm, entails several problems. One of the most serious is indoor pollution, which is a major cause of illness and death in many developing countries (Bruce et al., 2000; Torres-Duque et al., 2008) and which mostly affects women, since they are usually responsible for the cooking duties, and the children in their care. Other often mentioned problems include the time-consuming firewood collection (Rehfuess et al., 2006). Indoor air pollution is listed as one of the most serious health problems and is estimated by the World Health Organization to cause around 4.3 million premature deaths each year (WHO, 2015).

Perhaps counterintuitively, since biomass is generally considered to be a renewable fuel, inefficient combustion of biomass, even if harvested sustainably, is a major contributor to global warming, mainly due to the formation of black carbon. Other compounds, such as methane, are also formed during inefficient combustion (Bond and Sun, 2005; Bond et al., 2011; Ramanathan and Carmichael, 2008). Increasing combustion efficiencies or enabling households to use other fuels may provide a cost-effective, quick reduction of the emissions of climate warming compounds (Bice K, 2008; Kopp and Mauzerall, 2010). In areas where the collection of firewood contributes to deforestation, the combustion also contributes to an increase in atmospheric CO_2, the most common and widely acknowledged greenhouse agent.

Within the Kyoto Protocol, the international agreement for greenhouse gas (GHG) emission reductions, the Clean Development Mechanism (CDM) is an option for participating countries to support projects expected to reduce emissions in the developing world (UNFCCC, 1997, 2001) instead of reducing domestic emissions. The aims of CDM also include the funding of projects aimed at increasing local sustainability. Projects aimed at improving the cooking situation in the developing world would thus be suitable since, in addition to reducing GHG emissions, a substantial increase in welfare for poor households is also possible. However, one of the main climate warming agents emitted from inefficient combustion, i.e., black carbon, is currently not included in the Kyoto Protocol.

Options for interventions aiming to improve the energy situation for households include providing improved cook stoves (ICSs) and subsidies for alternative fuels, such as liquefied petroleum gas (LPG), as well as increasing the rate of electrification. Many ICS programs have not achieved their anticipated level of success (Gifford, 2010; Mobarak et al., 2012). Oft-cited success stories include the Chinese national stove program, which managed to reach a majority of the rural Chinese households (Sinton et al., 2004; Smith et al., 1993), and the Kenyan Jiko stove (Bailis et al., 2009). However, the Chinese program primarily achieved success in more well-off areas where there were also biomass shortages (Smith et al., 1993). The Jiko stove was designed for charcoal, a commercial fuel, mainly for use in urban areas (Bailis et al., 2009). The question of how to design stoves and programs that efficiently reach the majority of rural households in the developing world remains unresolved.

Household’s fuel choices have been studied extensively. The predominating model in many early studies was the so called fuel ladder model in which household’s where assumed to abandon dirty and, what the researchers believed to be, less convenient fuels, such as firewood, when their economic situation allowed them to do so, e.g. (Leach, 1992). Masera et al. (2000) found that many households do not abandon fuels but rather add more fuel’s to their fuel mix, partly because, for some tasks and energy services, the traditional fuels and ways of cooking can be more suitable, a behavior labeled fuel stacking, in contrast to the fuel ladder model.

The study of household fuel choices is of interest, not only for possible intervention, for example through fuel subsidies, but also in order to estimate the possible benefits of any such policies. Furthermore, there seems to be a correlation between fuel choice and adoption of improved cook stoves, and some of the circumstances that increase the likelihood of households using modern fuels also increase the propensity for using improved cook stoves (see, e.g., Barnes et al., 1994; Shrimali et al., 2011). Numerous studies have explored the topic of fuel choice (Heltberg, 2004; Hosier and Dowd, 1987a; Leach, 1992; Masera et al., 2000). A common conclusion is that, with higher income, households are more likely to start purchasing cleaner, more efficient fuels (Hosier and Dowd, 1987a; Leach, 1992). However, many more factors have been found to correlate with the usage of more modern fuels (Lewis and Pattanayak, 2012; van der Kroon et al., 2013). Several papers have also studied household firewood collection, in particular when connected to degraded forests (Amacher et al., 1996; Cooke, 1998; Heltberg et al., 2000; Van't Veld et al., 2006). These previous studies of fuel choice and fuel wood collection form an important basis for the interpretation of the modeling of fuel choice and the evaluation of the policy suggestions presented in this thesis.

Models of rural fuel use have previously mainly been concerned with finding and measuring causes to household's fuel choices based on other household characteristics. Many of these models have used different specifications, and the results are therefore either conflicting or together describe a very complex interplay between different factors (Lewis and Pattanayak, 2012; Takama et al., 2012; van der Kroon et al., 2013). One alternative interpretation of this complexity is that these factors are correlated with and partly describe area specific settings. Previous studies have, for example, found that electrification, education and household size are important (Lewis and Pattanayak, 2012). These factors can all be expected to be systematically different in urban and rural context. Factors found that explicitly point to area level influences, such as the distance to firewood, or population density have also been found to correlate with fuel choice in various studies (van der Kroon et al., 2013). Area level settings may also influence relative fuel prices and the household's opportunity costs for time.

Although the paths to different fuel choices may be many and complex, finding simple and common denominators can provide opportunities both for prediction and further understanding. The ability to predict current fuel use can be important for baseline calculations and useful for the design of future studies and projects, and, in combination with other studies, also used as a basis for policy formulation.

The main aim of previous models has often neither been to predict current fuel use, nor has the included variables been chosen as to facilitate such usage of the models. Thus, in this thesis the possibility of constructing models for prediction purposes is explored. Further, for interventions aiming at increased use of ICS or modern fuels initial adoption and sustained use is vital (Ruiz-Mercado et al., 2011). This has been found to be partly dependent on the fuel use before ICS interventions, and models describing differences in fuel use may therefore be useful. It has been found that ICS programs are less successful especially in areas where collected biomass is currently the dominating fuel, wherefore the situation in this type of areas is in need of special attention, and there is a demand for new concepts and ideas.

1.2 Aim and scope

The main aims of this thesis is 1) to explore whether rural households' fuel use can be modeled in new ways that focus on prediction, 2) to investigate whether area level differences in fuel use may have impacts for ICS programs, and 3) to address new solutions for ICS programs in areas where the current fuel use is mainly collected biomass. The first aim is met through two studies of household fuel use, Papers 1 and 3. The aim of Paper 1 is to create a model of current fuel use based on area descriptions, rather than individual household characteristics. Paper 3 aims to find a minimal set of characteristics, from both the household and area levels, that achieves maximal prediction of both current and future LPG use in households. The second aim is met in Paper 2, which explores whether local area differences in fuel use may have implications for ICS programs. The third aim is met through Paper 4, which attempts to evaluate an ICS program that could increase households' incentives and abilities to obtain state-of-the-art cook stoves based on a CDM-like funding mechanism even in areas where the combustion of collected biomass can be considered to be CO_2-neutral.

Papers 1 and 3 rely heavily on Vietnamese data and the conclusions are therefore currently limited to household's cooking in Vietnam. Paper 2, while using Vietnamese data, is applying a new method based on area dependent differences in household's fuel mixes which may be applicable beyond the region of study. Paper 4 is a conceptual model, not based on any particular data set.

1.3 Outline

This thesis is organized as follows: Chapter 2 presents a more in-depth introduction to solid fuel use in developing countries and possible policy interventions aimed at improving the rural households' cooking situation. It also presents the contribution of black carbon to global warming, the history of CDM projects, and prospects for future stove projects. Chapter 3 is a brief summary of the four articles, including the aims, methods, and results. Chapter 4 contains a discussion of the methods used in the various papers. Chapter 5 discusses generally the outcomes of this research. Finally, Chapter 6 ends the thesis with some suggestions for further research. The four papers are appended to this thesis.

2 Solid fuel use in developing countries

2.1 The problems related to the use of biomass fuel

The use of solid biomass as a cooking fuel has long been regarded as a problem by both scientists and policymakers. Yet, the reason it has been considered a problem has changed somewhat over the years. In the 1950s, India attempted to introduce chimneys to poor households, aiming to decrease indoor air pollution (Kshirsagar and Kalamkar, 2014). However, the issue really got attention at the global level a couple decades later when the World Watch Institute released a report titled *The Other Energy Crisis: Firewood* (Cooke et al., 2008; Eckholm, 1975). This report anticipated a global crisis, as a growing population with increasing demand for fuelwood would lead to wood being consumed faster than the forests could regrow. Extrapolating those contemporary trends in 1975 led to the conclusion that not only would major deforestation occur, but millions of households would be unable to prepare their meals. The goals of programs aiming to improve household's cooking were then mainly to reduce deforestation, and to delay "the other energy crisis".

These conclusions were later disputed because the World Watch Institute's study did not recognize that much of the collected firewood is not taken from standing trees, but rather from scrubs, bushes, and prunable trees, which can regenerate. Also unrecognized were the households responding to fuelwood shortages by reducing their fuelwood use or supplementing their fuel by producing their own sources, including planting trees, and switching to agricultural residues (Dewees, 1989).

A number of studies have investigated household's collection of firewood. Briscoe (1979) describes a complex pattern of energy use between different social classes in Bangladesh. Although all the energy used consists of various types of agricultural residues, firewood, and dung, differences in the availability and perceived attractiveness of the fuels between classes could be coupled with different usage patterns. Reddy et al. (1982) noted differences between villages based on the accessibility of wood. Households in villages without accessible wood nearby relied more on purchased wood, but also on wood grown on their own lands.

In more recent papers households have been modeled as units optimizing their utility to infer regression models for firewood collection (Amacher et al., 1996; Cooke, 1998; Heltberg et al., 2000; Van't Veld et al., 2006). The households collect firewood until the opportunity cost of doing so is higher than the price of firewood on the market (if such a market exists), or the opportunity cost becomes higher than the opportunity cost of using any other collectable fuel such as agricultural residues. We also expect them to stop collecting firewood if the opportunity cost becomes higher than the utility obtained from using the collected wood. Labor-constrained households are more likely to react to firewood scarcity with more costly measures (Cooke et al., 2008), i.e., they would have to pay a higher opportunity cost by diverting more labor into additional firewood collection or purchase a commercial fuel. According to Dewees (1989), if households collect less, it does not necessarily mean that less is available. Rather, it could signify a constraint on the labor supply within a household.

Studies of firewood collection in Nepal have reached contradictory conclusions concerning whether households increase or decrease their time collecting when forests become degraded (Amacher et al., 1996; Cooke, 1998; Kumar and Hotchkiss, 1988). However, these differences may correspond to variations in the areas and conditions of the studies. Although all three studies use Nepalese household data, the study by Amacher et al (1996) is in the Terai (marshy grasslands) and mid-hills whereas the studies by Kumar and Hotchkiss (1988) and (Cooke, 1998) focus on the hill areas. If firewood collection takes more and more time, with decreasing success, households will turn to collecting agricultural residues or buying fuel instead.

If, however, no other source becomes competitive, time spent collecting firewood is likely to increase. It is therefore possible to explain the reduced collecting time found in (Amacher et al., 1996) with the existence of more alternatives, including more agricultural residues, but also greater access to commercial fuels. It is also possible that the opportunity cost of time differs in the described regions making time spent on collection more expensive; the Terai is the most industrial region of Nepal, implying more opportunities for other time consuming activities. As of today the link between the collection of firewood and deforestation remains unclear (Arnold et al., 2006) due to the difficulty of distinguishing this activity from others, such as clearing land for agricultural purposes. However, recent estimates are that 27–34% of the firewood used globally can be considered to be unsustainable, albeit with large variations between regions (Bailis et al., 2015).

After the concerns of ”the other energy crisis” faded, attention has again been turned to the health aspect of cooking with solid biomass. Indoor air pollution resulting from incomplete combustion of biomass fuels eventually attracted renewed attention. With an estimated annual mortality of 4.3 million (WHO, 2015), it is increasingly understood as one of the most serious global health problems. Reducing indoor air pollution became the new focus, by introducing either chimneys (emitting smoke outside) or increased combustion efficiency (in order to decrease emitted particles) – or both. Pope et al. (2009) presented a compilation of several dose-response studies of $PM_{2.5}$ (particulate matter with a diameter less than 2.5 micrometers) and found a log-linear relationship between the inhalation dose of $PM_{2.5}$ and the relative risk of mortality from ischemic heart disease, cardiopulmonary disease, and cardiovascular disease, implying greater benefits of reducing a certain amount of exposure at the lower end of the scale than at the higher end. Smith and Peel (2010) used this mapping to show the difference in benefits for various policies. Grieshop et al. (2011) used the results in a comparative analysis of different cook stove alternatives.

There is however a gap in the research between the daily inhalation dose of 1 mg/day and 10 mg/day of $PM_{2,5}$, as pointed out by Smith and Peel (2010). This is the interval of the estimated exposure from traditional cooking with solid biomass. One must therefore assume this relationship holds true for the interval of interest when estimating health benefits from various stove alternatives. However, there is direct evidence of improved health following ICS adoption and usage (Ezzati and Kammen, 2001). Recently, the term indoor air pollution has been substituted for household air pollution, due the acknowledgement that, especially in densely populated areas, pollution from cooking with solid biomass also affects the outdoor environment. This realization also implies that adding a chimney as the sole solution for increasing health may be insufficient.

The incomplete combustion that emits unhealthy particles also generates compounds that contribute to global warming. Most severe among these are methane and so-called black carbon. Black carbon will be discussed further in Section 2.3. Furthermore, in areas where the collection of biomass is causing deforestation, the activity also leads to an increased atmospheric concentration of CO_2. The provision of ICSs has been argued to be one of the most cost-effective ways to both improve health and combat global warming (Smith, 2008).

Use of solid fuels for cooking implies other problems for households as well. Many households collect their needed fuel, usually firewood, agricultural residues, and dung. Other households instead purchase fuels, such as firewood, charcoal, and coal. Households that collect fuels often spend much of the day to do so, at the expense of tending crops and animals, looking after children, studying, or various income generating activities. A large welfare increase may therefore result from households spending less time collecting firewood. Other households purchase fuels, either because they can earn more money than the cost of gathering firewood, or collecting firewood simply is not possible in their area. Many of these households have to spend a substantial portion of their income on firewood. If these households could use purchased fuels more efficiently, it could increase welfare through direct monetary savings or the ability to cook more food.

Efforts to improve the cooking situation of poor households may lead to co-benefits in many dimensions. Interest in making these changes should include many different actors, from the households to local authorities, national governments, and the international community.

2.2 Policy interventions

Two types of intervention are generally considered as viable policy options. The first is to help households achieve higher combustion efficiency, using the same fuels they currently use. This can be done by promoting ICSs. The second option is to stimulate households to abandon solid fuel in favor of cleaner options such as LPG and electricity. This process, which also occurs spontaneously as households increase their socio-economic status, has been studied extensively to understand how to stimulate it by manipulating different factors and for making projections.

2.2.1 Fuel switching

Fuel switching studies have generally concluded that as household income and socio-economic status rise, they start using modern fuels (Barnes et al., 2004; Campbell et al., 2003; Hosier and Dowd, 1987a; Leach, 1992). Partial fuel switching (Heltberg, 2004) and fuel stacking (Masera et al., 2000) describe a common situation in the developing world, where households do not completely switch to modern fuels. Instead, the same household might use several different fuels.

Partial fuel switching has sometimes been interpreted as households using different fuels for the same activity, such as using LPG for much of the year except when harvesting, when the household might have access to large amount of agricultural residues. The term fuel stacking on the other hand emphasize that for some tasks and energy services, the traditional fuels and ways of cooking can be more suitable. Recently, measurement have shown that the fuel stacking behavior can have substantial consequences for the effectiveness of ICSs, since cooking is only one energy service obtained by the open fire, and furthermore the ICS may not be suitable for all types of cooking performed with the traditional stoves or over open fire (Ruiz-Mercado and Masera, 2015).

Both Masera et al. (2000) and Heltberg (2005) analyze the combination of LPG and wood, in particular in relation to why households continue to use purchased fuel even after investing in an LPG stove. In some cases, cooking with purchased firewood is more expensive than using LPG, if the cost of the initial investment in the stove and canisters is disregarded (Heltberg, 2005). Masera et al. (2000) find that households continue to use wood because LPG stoves are not very suitable for making tortillas, while Heltberg (2005) finds that households that rely solely on LPG tend to live in bigger cities than households using both LPG and wood, and interprets this as these households can more easily buy the types of food that are not suitable for cooking on an LPG stove instead of preparing them themselves. Because of the different types of cooking in East Asia compared with South and Central America, these findings are not directly transferable. However, it has been noted that rice straw, a type of agricultural residue, is especially good for some types of frying because of its flame properties (Wen and En-Jian, 1983). Therefore, agricultural residues may occasionally be preferred by some households.

There are possible area level effects on household fuel choice, for example an area’s level of rurality which can be linked to the availability of biomass, availability of modern alternatives and the opportunity cost of time. The availability of collectable biomass is not always included in the models of rural fuel choice. However, several variables, correlated with rurality, which in turn may be correlated with the availability of biomass, often are included in the energy ladder models. Evidence for the latter correlations is given in Heltberg (2000), who uses population density as a proxy for availability of firewood in a model for firewood collection. Greater availability of biomass lowers the opportunity cost of collection, thus increasing the competiveness of these fuels. This may affect fuel choices in two ways: First, a household will use collected fuels as long as the opportunity cost of collecting and using those fuels (sometimes called the inconvenience cost) is lower than the cost of purchasing and using any other fuel from the market. Second, the extent of a household's cooking that is performed with a commercial fuel will influence the relative cost of using that fuel on a per-meal basis if an initial investment in appliances is needed in order to use these fuels. For example, in order to use LPG, investment in both a stove and canisters is needed, which means that the utilization time very much determines the difference in cost between purchased firewood and LPG.

Studies of fuel switching do not always include the availability of collectable fuels (Campbell et al., 2003; Heltberg, 2004; Leach, 1992). An extensive study of fuel switching using household data from eight developing countries was conducted by Heltberg (2004). In both rural and urban areas of many of the studied countries, several household factors, including household expenditure, education, and access to indoor water and electricity, correlated significantly with fuel switching. Peng et al. (2010) examine factors determining whether households use biomass and, if so, the amount of biomass used. These studies do not explicitly account for availability of biomass. Heltberg (2004) divides the samples between urban and rural households, while Peng (2010) uses dummy variables for different landscapes: plains, hills, and mountains. Other studies, such as Heltberg (2005), include measurements of biomass availability, distance to firewood collection, but find little influence on fuel choice from distance to firewood collection. Gundimeda and Köhlin (2008) on the other hand do find an effect of forest cover on fuel choice.

Fuel switching is generally occurring faster in urban areas than in rural areas (Gundimeda and Köhlin, 2008; Heltberg, 2004; Heltberg, 2005). Possible explanations for these different fuel switching rates include a lack of infrastructure for modern fuels, lower (or non-monetary) and more sporadic income, a more traditional lifestyle and lower opportunity cost of time, together with a higher availability of collectable fuels. The availability of biomass has been found to also have a strong influence on the process of urban fuel switching (Barnes et al., 2004).

Several factors that have been used to model fuel switching may be correlated with a higher degree of urbanization. For example; electrification, indoor tap water (Heltberg, 2004), and number of years since electrification (Peng et al., 2010), education and occupation. Also income, the most characteristic variable of the energy ladder model, may also be correlated with population density. Heltberg (2004) found a correlation between LPG use for cooking and access to electricity in both urban and rural areas. This was also found in urban samples by Barnes et al. (2004), who offer two possible explanations: (i) “access [to electricity] proxies for market development. In that case, fewer barriers would constrain other modern fuels in cities where electricity is available,” and (ii) “availability of lighting and other appliances spurs people to a greater acceptance of modernity and modern fuels.” Heltberg (2005) adds two further possible explanations: (iii) “areas that are in some sense more ‘modern’ (for example large as opposed to small towns, and places with better infrastructure) get connected first to the grid” and (iv) “assume that energy needs are organized in a hierarchy where electricity is the most desired innovation and modern fuel is further down the list of priorities.” It is possible to enumerate further: (v) an open fire provides services other than solely heat for cooking; these may include lighting, conservation of food, and repelling of insects. Access to electricity entails more convenient ways of achieving these extra energy services, and it may therefore be considered that electricity reduces the open fire to simply being a means for cooking and, as a result, the open fire loses in competiveness. (vi) this point is closely related to point (iii), namely that areas that get connected first are generally more densely populated and can thus on average provide a lower availability of collectable fuels (compare Barnes et al., 2004). (vii) Electrification may lead to an increase in income, both in agricultural and non-agricultural areas, which in turn may drive fuel switching. An increase in income from agriculture after electrification due to better irrigation possibilities has been found (Khandker et al., 2009). Finally, (viii) a more densely populated area is also likely to provide more opportunities for wage work, which in turn could affect the opportunity cost of firewood collection. Items (ii), (iv), and (v) can thus be seen as explanations for why electricity may cause fuel switching, while (i), (iii), (vi), and (vii) are explanations for why electrification correlates with fuel switching in regression models.

The question of why there is a correlation between LPG use and household access to electricity and whether there is any causation involved is still unresolved (Köhlin et al., 2011). However, several of the explanations offered are related to different degrees of urbanization. The finding that electricity also affects the LPG adoption in rural areas may thus be due to the definition of these areas as being rural areas with and without electricity, while they perhaps could have been defined as more or less rural in the sense of household density.

Across many studies, several factors have been found to correlate with fuel switching, and likewise, in many cases, also been argued for to be causes for fuel switching. However, even if these are actual causes, they cannot all be affected through policy. Although development in other areas, such as electrification, has been shown in some studies to correlate with LPG adoption, as noted this connection is still uncertain and debated. Increasing the rate of urbanization can hardly be done for the sake of increased LPG usage. Reducing access to firewood is not possible or desirable, since it would decrease welfare for households unable to make the transition. Education may show benefits but implies a long-term treatment. Increasing employment would enable households to purchase more expensive fuels, but would also increase their opportunity cost of collecting firewood. However, one can assume that the respective governments are already trying to increase the number of work opportunities for other reasons.

More direct actions include subsidizing the fuels deemed desirable, like LPG. However, the experience from such interventions is that most targeted households still cannot afford the transition, and instead the beneficiaries tend to be those who already use LPG, industries, and transportation companies. This has generally not been shown to be a cost-effective approach, at least on its own (Arze del Granado et al., 2012). Poor households may not start to use LPG even when subsidized, as the fuels they are currently collecting remain the most economic option or because the cost of the LPG stove is too high. Attempted subsidies of the stoves instead of the fuel (or a combination of fuel and stove subsidies) may however be more promising (Arze del Granado et al., 2012; Edwards and Langpap, 2005). In many places the cost of LPG is already on par with firewood, on a per-meal basis, after the investment in the LPG stove. Thus, subsidizing LPG stoves may be a viable option in areas where most households are currently purchasing firewood, but may not significantly affect households currently collecting their fuel.

2.2.2 ICS types and characteristics

Besides trying to accelerate fuel switching, policies may aim to promote household adoption of improved cook stoves with higher efficiencies and lower emission than traditional stoves.

Traditional stoves are often either of three-stone or tripod type, i.e., basically an open fire with an arrangement of stones or metal so that pots can be placed over the fire. Another traditional type of stove is a mud stove enclosing the fire on three sides, which can be more fuel efficient by isolating the fire and directing the heat. However, mud stoves may increase indoor air pollution compared with the open fire, by limiting the air supply (Kshirsagar and Kalamkar, 2014). There are many types of ICSs, from locally manufactured mud stoves to new generation ICSs, sometimes labeled ABS (advanced biomass stoves) (Kshirsagar and Kalamkar, 2014). The simplest ICSs are better designed mud or ceramic stoves with added features such as better enclosure of the fire, better heat transfer, fuel grates, and sometimes chimneys to direct flue gases away. Often aims of such designs include affordability and local manufacture, by local artisans using locally available materials.

At the other end, ABSs include gasifier stoves and rocket stoves. Gasifier stoves utilize the principle of gasification: the fuel is first combusted in an oxygen deficit environment, creating syngas (which mainly consists of CO and H_2). In a secondary stage the syngas is combusted, creating the flame that heats the pots. Gasifier stoves have currently reached the best levels of emissions of all ICS types, but are not yet at the level of LPG stoves (Grieshop et al., 2011). The other type of ABS is the rocket stove, a metal cylinder utilizing direct combustion. Rocket stoves are generally on par with gasifier stoves in fuel efficiency, but generally do not reach as low emissions. Both the rocket stoves and the gasifier stoves need electric fans to achieve maximal combustion efficiencies (although several designs without are also available), which are either battery operated or driven by small thermo-electric generators, running on the heat from the stove itself. This latter option is particularly interesting for households currently lacking access to electricity and when combined with products that let this electricity serve other purposes, like charging mobile phones. Disadvantages of the so-called ABSs are mainly that they are expensive and too advanced for local manufacturing. A further downside is that they often require finely chopped fuels, which may be an inconvenience for households.

Some types of cooking may be more or less suited for an ICS, e.g., baking of different kinds of bread. Like the fuel stacking behavior (see Chapter 2.2.1 and Masera et al., 2000; Ruiz-Mercado et al., 2011), households may only use the ICS for some types of cooking. Thus the most efficient ICS model may not always achieve the highest overall efficiency for a household.

The recently found log-linear dose-response curve from indoor air pollution on health (Pope et al., 2009; Smith and Peel, 2010; and Chapter 2.1) may justify extra efforts to distribute stoves that significantly decrease emissions, because the positive effect on health accelerates with greater reduction. Still, at present only LPG stoves achieve sufficiently clean combustion to reach WHO recommendations for indoor air pollution levels. Furthermore, if the ICS only partly displace the open fire these low levels will neither be achieved without additional efforts.

As touched upon in Chapter 2.1, the aims of improved stove programs have shifted over the years. For example, if the only objective is to improve indoor air pollution, it may be enough to add a chimney. However, this does little for improving efficiency, which may be an aim if deforestation or household economy is a policy objective. Moreover, in densely populated areas it may increase outdoor pollution. The aim of policy makers to promote ICS may influence not only the stove design, but naturally also what stove and usage parameters are researched and recorded (Ruiz-Mercado et al., 2011), from exposure to PICs to stove efficiencies. However, Ruiz-Mercado et al. (2011) argue that some of the most important stove parameters, on which all others depend, are often overlooked and not sufficiently measured, namely initial adoption and sustained usage. New methods using real time measurements connected with mobile IT solutions now enable scholars to measure usage over time to more properly estimate actual benefits of stoves when adopted (Pillarisetti et al., 2014).

2.2.3 ICS programs and dissemination

A number of stove types were presented in the previous section. The number of alternatives continues to grow, but the promotion of these innovations has proven difficult.

ICSs have shown to be able to reduce emissions harmful to both human health and the climate (Grieshop et al., 2011). Furthermore, they can significantly reduce the needed input of both firewood and time (Jetter and Kariher, 2009). Hence, the stoves would enable households to gain a cleaner indoor environment and more efficient energy utilization, which in turn would save them time (and money if currently purchasing firewood). The respective countries would have a healthier and more productive population, and might be able to save natural resources. For the international community, distributing ICS could be a cost-effective way to reduce global warming. Thus, there are significant co-benefits spanning a wide range of stakeholders. So why doesn’t a large-scale distribution take place?

Several barriers have been identified over the years. First, although the stoves span a large range of alternatives, those that substantially increase efficiency and reduce emissions are expensive, at least for an average household in the developing world. If the household is currently purchasing firewood, the payback time may range from half a year to several years. Few households can afford to take such a risk. If a household relies solely on self-collected firewood, which is normal in rural areas, the payback can only be achieved if the time saved is used for income-generating activities. However, possibilities for earning money are often rare, and many households may not have sufficient funds for making the investment in the first place. Currently, the stove options that have demonstrated substantial improvements are expensive (50-100 USD), and cheaper options may not reduce emissions or the demand for firewood enough to be attractive at all, as was the case in a recent trial in India (Hanna et al., 2012).

There is also a gender dimension to this liquidity problem. In many places women are responsible for cooking and the collection of firewood, although the latter activity is sometimes shared, while the husband is usually the head of the household and thus in charge of household spending. Even if the person responsible for firewood collection and cooking may appreciate time savings and a cleaner indoor environment, this may not be perceived as an important issue for the household head (Miller and Mobarak, 2011).

Most of the poor households in the developing world will not be able to acquire ICS without external help. External help might mean subsidizing the stoves, partly or fully. Another option is to provide a beneficial payment plan, so that households do not have to make a large upfront payment for a new technology with unclear benefits. This latter option relies somewhat on anticipated potential monetary savings or at least on the household having some sort of monetary income. In such circumstances, payment plans have in some cases been shown to drastically increase households' willingness to buy an ICS. At least this has been shown for ICSs aimed for the commercial fuel charcoal (Beltramo et al., 2014). However, the difference between charcoal and firewood is not only that charcoal is always commercial, but the design of an improved charcoal stove might be easier than the design of a stove meant to replace the open fire.

Several programs (and studies aimed to measure uptake) have provided ICS (for firewood) for free or almost free, and many have yielded discouraging results. Several papers have concluded that households do not value what is given for free (Barnes et al., 1994; Martin et al., 2011). This attitude is often found among field workers (Bensch and Peters, 2012), not only with respect to ICSs but to other technologies, such as mosquito nets. There may be a selection bias underlying this conclusion, however. Households that have bought an ICS are households that already expect advantages from using an ICS. Furthermore, a household purchasing an ICS has a say in the decision, and can thus pick a model suitable for its needs rather than being given a generic, often low quality model. Lastly, a household with money to purchase a stove is also more likely to purchase fuel and thus may have direct monetary incentives to use the stove. These explanations are not the same as saying that households that receive an ICS for free do not value the ICS just because it is free.

Another possibility that seems equally or perhaps more plausible is that stoves that have been handed out for free have not been of sufficient quality and have not been suited for all cooking needs. Further, the open fire may be valued for uses besides cooking, and these extra services may be more relevant to the poor. Some of these uses may actually be in conflict with the intention of the ICS technology. For example, some level of indoor smoke may be desired as an insect repellent and as a way to preserve food by hanging the food high in the kitchen and in this way smoke it slightly. Further uses of an open fire may include heat and light, services that are also decreased when ICS technology are introduced. Lastly, the fire may be of social value and serve as a family gathering point, or it may be associated with religious practices (Shankar et al., 2014). Satisfying these uses with the ICS design would be challenging.

However, if there is an anticipation that stoves will be continued to handed out for free, incentives for stove maintenance may decrease. This interpretation of "not valuing what is free" might still be true.

Even without attempting to meet these alternative values of open fires, the design issue is still challenging and failure to meet the needs of the households may result in substantial losses in efficiency improvements. For example, a household may be used to cooking two meals at the same time, while the ICS can only cook one, thus forcing the household to use both the ICS and an open fire, and thus the efficiency improvements are lower. The opposite may also be true, i.e., an ICS designed for cooking two meals at the same time may be suboptimal when cooking only one. Furthermore, some types of cooking may not be suitable at all with certain ICS designs; baking bread is a typical example. The overall efficiency gains would, again, not satisfy the ICS designers' expectation. Some stoves require finely chopped wood, which could require more work. All these examples suggest that stoves may need to be designed not only for different regions but to give individual households more choice in the type of stove they are to use. These observations may help explain the failure of many stove programs employing a top-down approach, in which efficiencies were optimized in laboratory settings and then stoves were distributed to households.

There is a further trade-off to consider between stove cost and stove performance and quality. There are many reasons for choosing a low cost stove; however this may also lower the increase in household welfare. Many programs have chosen low cost stoves, either because of budget limitations, still wanting to reach a large number of households, or because there are hopes of establishing local markets. The establishment of local markets may be sought for several reasons. For ICS use to continue, the limited lifetime of stoves implies either that stoves must continually be subsidized, or that after initial trials, households need to see the benefit and continue to renew or repair their stoves. Stoves may have to be in the lower end of the cost range for this to happen, however. Furthermore, stove programs often anticipate co-benefits like the creation of local work opportunities, including stove manufacturing, installing, and repairing. This ambition also requires that the stove be built from locally available materials and of a relatively simple design.

On the other hand there are also stove programs advocating expensive stoves, stoves that are so much more efficient and clean that households experience sufficient improvements when using them. However, these programs may need to be continually funded, and the stoves are not manufactured locally and thus do not serve efforts to raise the number of local jobs.

2.3 Black carbon and global warming

There is a broad consensus regarding the existence of anthropogenic global warming. This means that, due to human activities, an increase of various compounds in the atmosphere is causing temperature increases. The main mechanism behind global warming is the greenhouse effect, which means that various molecules in the earth's atmosphere absorb more light in the wavelengths dominating in the earth's radiation spectrum compared to that of the sun's. Many human activities emit gases with this characteristic, the best known and most important being the combustion of fossil fuels. There are, however, still significant uncertainties regarding feedback mechanisms, i.e. how an increase in global temperature *per se* will cause even more global warming, for example through increased cloud formation, methane leakage from the tundras and changed albedo from decreasing ice caps (Allison et al., 2009; IPCC, 2007).

The major contributor to climate change from the burning of biomass is black carbon (Bond et al., 2011; Grieshop et al., 2011), albeit not mainly through a greenhouse effect. The estimations of the black carbon impact on global warming contain some uncertainties, partly because black carbon acts on the climate through several different mechanisms, e.g. by lowering the albedo of ice, polar ice caps and glaciers, altering the albedo of clouds and directly absorbing radiation (Ramanathan and Carmichael, 2008). Further difficulties in estimating the effect stem from the very short atmospheric lifespan of black carbon compared with greenhouse gases. Because of the short lifespan, the exact effect of black carbon emissions is determined by both the location and time of the emission. Different locations means different probabilities for the black carbon to land on ice cover or interfere with clouds. Furthermore, since incoming radiation is higher closer to the equator, black carbon emitted there will have a larger impact due to direct radiation absorption than black carbon emitted closer to the poles (Bond et al., 2011). The ice caps have melted faster than predicted. An explanation for this could be the exclusion of lingering black carbon on icy surfaces from climate models (Ramanathan and Carmichael, 2008), although several other mechanisms could also cause this discrepancy (Arzel et al., 2006; Pedersen et al., 2009).

The story is further complicated by the fact that many sources, especially biomass sources of black carbon, also emit so-called organic carbon, which is thought to have a cooling effect on the climate. Not only is this effect also uncertain, but the relative emissions from different sources are sensitive to type of biomass and burning characteristics (Grieshop et al., 2011).

Although uncertainties remain, the evidence of the contribution of black carbon to global warming is steadily increasing (Bond et al., 2011; Haywood and Ramaswamy, 1998; Jacobson, 2001; Ramanathan and Carmichael, 2008). And although the short lifespan poses some regional variations in the effect of black carbon, this characteristic may also enable fast climate action, which could have immediate benefits for the climate (Ramanathan and Carmichael, 2008). This could prove to be vital in light of more and more evidence that climate goals may be difficult to reach in time (Matthews and Caldeira, 2008). Targeting black carbon might also seem particularly attractive considering the inbuilt inertia in the energy-climate system (Myhrvold and Caldeira, 2012).

Global warning potential (GWP) values for black carbon have been calculated in several papers, though several authors have pointed out the difficulty of comparing black carbon with greenhouse gases (Bond et al., 2011) due to the short atmospheric life of black carbon (i.e. from days to weeks). Hansen et al. present GWP values of 2000, 500 and 200 for 20, 100 and 500 year horizons, respectively (Hansen et al., 2007). The values from Hansen also include the cooling effect of organic carbon. The most recent study (Bond et al., 2011) finds the GWP values for black carbon to be 2900±1500 and 830±440 for the 20 and 100 year horizons, respectively; for organic carbon the corresponding values are -160 (-60,-320) and -46 (-18,-90). However it has been argued that the GWP metric is unsuitable to handle short-lived climate agents like methane and black carbon in particular (Shine et al., 2007). This is because the GWP value remains the same, not accounting for time of emission or climate goals. One can argue that a short-lived agent such as black carbon should be counted as more valuable when one risks overshooting targets, but as less important when emitted today. Metrics weighing such considerations have been proposed (Shine et al., 2007). However, possible feedback effects, such as a warmer climate leading to increased methane emissions from the tundra and decreased albedo from glaciers and polar caps, are not properly reflected here either.

2.3.1 Black carbon and discounting

Discounting of future emission is routinely applied in energy systems and economic and climate projections. The discount rate is generally considered to consist of two components: one reflecting the pure time preference of having something sooner rather than later and the other reflecting the notion that society is expected to increase its wealth and hence a future expense can be considered to be relatively lower. Without jumping into the controversy regarding the appropriate value to be used as discount rate (Nordhaus, 2007; Sterner and Persson, 2008), it is interesting to contemplate the discrepancies that arise due to the fact that the GWP metric is adjusted to reflect the impact of a climate warming agent over a certain amount of time, be it 20, 100 or 500 years. In energy-economic models projecting future scenarios, discounting is routinely based on the date of emission.

Now let us compare CO_2 and black carbon in the framework of discounting. CO_2 is a long-term climate gas, and the effects of emitting CO_2 into the atmosphere will last for centuries. Black carbon on the other hand is short-lived in the atmosphere, on the order of weeks. The GWP metric is calculated as the total effect of an emission on the climate during a certain period as, the GWP_{100} and GWP_{20}, are the total effects of a certain gas within a 100 and a 20 year period respectively. If we were discounting the cost of emission reduction per CO_2-equivalent, our energy systems model would tell us that the cost of emitting a CO_2-equivalent in year one is the same whether it is black carbon or CO_2.

Now assume we were to instead discount the climate benefits of the emissions. Further assume that the effect of CO_2 is exponentially declining so that in 100 years the climate effect is about one-third of the initial effect and that black carbon has its entire climate effect in the first year (for simplicity; in reality the effect lasts for only a few weeks, as noted above). Discounting this climate effect to the current year, with a discount rate of 1.05, yields a value of the reduction of black carbon almost four times higher than the corresponding value for CO_2 for the same amount of emitted CO_2-equivalents in year one.

This example is not meant to be physically correct; rather, it is just an example to explain the problem of discounting in energy systems models where climate warming agents have different life spans. The climate effects also depend on when in time emissions occur, as the environmental damage of climate change is not linear with emissions. Furthermore, in the example, the climate cost is assumed to be proportional to the climate forcing and occurring simultaneously. This is also a simplification.

2.4 The Clean Development Mechanism

The Clean Development Mechanism (CDM) is one of several flexibility mechanisms created under the Kyoto Protocol. It allows for the developed nations (Annex 1 nations) to perform part of their CO_2 mitigation in the developing world through the purchase of Certified Emission Reductions (CERs). CERs are awarded to projects that have reduced GHG emissions. The CDM has three goals: to promote sustainable development, to assist the Annex 1 nations in meeting their emission targets by finding lower cost options for mitigation, and to assist in meeting the overall goals of the UNFCCC (UNFCCC, 1997; Wara, 2008). Another reason for implementing the CDM was to encourage greater participation in the Kyoto Protocol: the CDM gave developing nations the opportunity to receive foreign investment, and developed nations will have a cheaper option in case domestic mitigations prove too expensive. Indeed one of the main proponents of flexibility mechanisms in the Kyoto Protocol was the U.S. (Bodansky, 2001). There is also hope that the CDM will induce technology transfer, which can promote sustainable development beyond the CDM projects themselves (Paulsson, 2009).

From an intuitive economic perspective, flexibility mechanisms make sense, by allowing the market to find the least cost-options. Extending the flexibility mechanisms beyond the capped nations should allow for finding the cheapest options globally. Furthermore, if it is possible to find cheaper options for emissions reductions and at the same time improve local sustainability, this seems like a win-win policy.

There are many reasons why options to decrease emissions are cheaper in the developing world than in developed countries. Many basic efforts have already been made within certain areas in developed nations, like increasing the efficiency of power plants, albeit for reasons other than GHG reduction. Further improvements might prove more difficult and expensive than raising the performance of low-efficiency technologies in the developing world. In some cases, in transition economies where the energy system is rapidly expanding, it might be cheaper to influence this development in a more sustainable direction than altering mature energy systems in the industrialized world.

2.4.1 Criticism of the current CDM mechanism

Because the CDM is a mix of a market solutions and subsidy schemes on the receiving end, it is subject to criticism generally directed toward subsidy policies. Receivers are given money to reduce their GHG emissions. It is therefore possible to draw some conclusions from past experiences of subsidy programs. The inefficiency of subsidy programs can partly be blamed on so-called “free riders,” i.e. actors who would have performed the change even in the absence of the subsidy. This dilemma is often not considered in subsidy programs. Empirical evidence suggests a high level of free riding in subsidy programs. Should free riding also happen within the CDM, it is even more alarming than just being an inefficient policy. Since money generated from the non-mitigation of GHGs in one place is given to a free rider whose behavior was unchanged by it, the result is an increase in GHG emissions compared with if the policy had not been in place (Sterk and Wittneben, 2006). This problem is acknowledged through the criteria for additionality, i.e. that a project must be proven not to have occurred in the absence of CDM funding. This leads to conflicting goals, however: one cannot seek out the cheapest option, but only the cheapest option among those that are not considered profitable. This is known as Grubb’s paradox. Already in 1998, Grubb aired concerns about the structure of what would become CDM: “The difficulty of ensuring that emissions savings are additional is compounded by the paradox that the most ‘cost-effective’ projects will be the least ‘additional’” (Grubb et al., 1999).

The CDM functioning so that the market seeks the most cost-effective mitigation options in the third world. However on top of this, is the objective of additionality; but these two objectives are in conflict. When estimating a project’s cost-effectiveness, to be sure that a project is truly additional it should be far from cost-effective, but the aim of having a market-based mechanism is to find cost-effective solutions. It is also worth considering that the entities responsible for the projects and for verification and monitoring, and also the buyers of CERs, all have incentives to claim that the most cost-effective projects are also additional. If the entity behind the project can prove that an already cost-effective project is also additional, they can make more money than if they choose to undertake a project that is only cost-effective when CDM money is put into the project. For the buyers of CERs, funding cost-effective projects means meeting their obligations at lower costs (depending on what sets the price of CERs).

The goal of sustainability is not monetized and is therefore generally overlooked. The benefits of the project depend on the amount of CERs it generates. Deciding whether a project is helping to achieve sustainability goals is up to the host country, and no aid is offered from official methodologies (Sterk and Wittneben, 2006). Many projects with questionable contributions to local sustainability has previously been accepted (Pearson, 2007; Wara, 2008).

2.4.2 CDM for black carbon

Black carbon is currently not included in any international climate agreements, partly due to the many uncertainties surrounding its effects on the climate. However, this knowledge has improved over the last decade so that the negative effects have become increasingly clear. Despite increased knowledge, it remains unresolved how to evaluate black carbon against other climate warming agents. However, even if assigning a value at the lower limit of the confidence interval of the possible GWP estimates of black carbon, valuable funding for ICS programs would still be possible.

Including black carbon in CDM may make it possible to fund ICS projects. There are already a few stove projects funded through carbon credits within the framework of CDM (Simon et al., 2012). In these projects, the carbon credits are given on the assumption that parts of the firewood is harvested unsustainably and that reduced deforestation is achieved due to a more efficient use of firewood.

Considering the criticism of many CDM projects, stove projects may be an interesting option aimed at increasing local sustainability, both in terms of improving health and increasing local employment opportunities. Furthermore, including black carbon in international climate agreements may stimulate increased participation by developing countries in future agreements

(Bond and Sun, 2005). However, formulating a policy framework that includes both developed and developing countries may prove complicated. Furthermore, the issue of how to value black carbon against the greenhouse gases remains open.

3 Summary of the papers

Papers 1 and 3 are mainly descriptive in nature, intended to increase the knowledge of rural fuel choices. Despite being mainly descriptive, policy conclusions can be drawn from Papers 1 and 3 when combining the results from these papers with those of previous (or future) research. Paper 2 is concerned with policy implications for stove programs based on differences in fuel use patterns between areas. Paper 4 is an evaluation of a new type of stove program aiming to enable poor households to adopt ICS technology.

3.1 Paper 1: Rurality as a policy relevant measure for commune level differences in fuels used for cooking in rural Vietnam

Paper 1 is primarily a descriptive article utilizing two different rural surveys for modeling fuel use on commune levels in Vietnam. The results may be useful for example in calculations of emissions and of which biomass resources could be freed up if households were able to switch to more efficient cooking practices, either self-achieved or through various aids. The model can also assist in both understanding and quantifying the mechanisms behind fuel choices and thus aid in the formulation of policies. The model is not, however, based on economic theory and the parameter estimates should not be understood as causal effects.

Based on previously found differences in fuel use between urban and rural settings (Barnes et al., 2004; Heltberg, 2004), a model is formulated that uses some easily obtainable measures of rurality in order to predict fuel use at commune level in Vietnam. Instead of the rather abstract concept rurality, commune mean income, distance to nearest town and the average amount of land holdings per household in each commune (village average land; VAL) is used. VAL is closely related to household density, but is constructed only from the, by the households reported, land holdings, i.e. commons and land owned by other entities is not included. The models constructed with these variables are very accurate, in terms of a high level of prediction, and can thus be used to describe fuel patterns in different communes.

Although not the main aim of the models, the results in the paper may have implications for some previous models of rural fuel choices that aim to infer causality in their parameter estimations. Several studies of rural fuel choice have been performed in the past, including (Heltberg, 2004; Hosier and Dowd, 1987b; Leach, 1992). The underlying hypothesis is that households are switching to more modern, clean and convenient fuels as their income levels and socio-economic status increase.

Other papers deal with the gathering of collectable fuels, such as firewood and agricultural residues (Amacher et al., 1996; Cooke, 1998; Van't Veld et al., 2006). Several of the papers concerned with fuel switching lack a comprehensive description of the fuel gathering process; see, e.g., (van der Kroon et al., 2013) for a review.

An interesting finding from the Vĩnh Phúc area is that use of agricultural residues is more common in less rural environments. In more household-dense, i.e. in one sense less rural, areas, the households use agricultural residues and purchase fuels to a greater degree. The behavior of turning to privately produced fuels, such as agricultural residues, is described in the literature as the response by households to degraded forests, i.e., when gathering firewood becomes more difficult or when the shadow cost of collecting firewood increases by other mechanisms (Van't Veld et al., 2006).

It is argued in Paper 1 that some of the results generated in previous studies may instead be explained by their association with rurality. Household density and distances to nearest towns are used as explanatory variables in Paper 1 partly because, as a predictive tool, reducing reliance on household data may be beneficial for future studies and policy interventions, and partly because it is argued in the paper that some measurement of rurality, such as household density, should be considered as a confounder of many other variables. Consequently, if the causal effects of other parameters are sought in the future, it is reasonable to check whether the inclusion of rurality in one’s regression in any way alters the results. Even if rurality is in itself only an umbrella for many variables, e.g., the opportunity cost of gathering firewood and access to alternative fuels, including it in models may lead to alternative interpretations. While, strictly speaking, not proving or disproving one or the other, the knowledge that one’s model can be equally well described by a rurality measurement may provide alternative interpretations and lead to new routes of research.

An example of such findings is the link between electrification and LPG usage found in previous studies (Heltberg, 2004; Peng et al., 2010). Whether this is merely a correlation or whether true causation exists is still a matter for debate (Köhlin et al., 2011). Electrification tends to be more common in densely populated areas and closer to towns, which are the same areas where the availability of collected firewood is lower and outlets selling modern fuels are more common. Availability of various fuels is one example of a possible mechanism behind fuel choice in more or less rural settings.

The usefulness of the models is mainly in future estimations of variation of policy outcomes in different areas, predicted from only a few, easily obtainable area level variables. However, the model presented in Paper 1 does not in itself provide such predictions of effects; instead it is only able to predict the current situation, and as it is based on 2008 data, it might benefit from an update. The model thus needs to be complemented with experiments to be truly useful. Such experiments have been carried out during recent years and include randomized trial of ICS adoption in order to find average treatment effects(Bensch and Peters, 2012; Hanna et al., 2012; Marron et al., 2008) as well as discrete choice models exploring household preferences for fuels and stoves (Takama et al., 2012; van der Kroon et al., 2014). Considering previous understanding of ICS adoption (Barnes et al., 1994), it is possible that outcomes of such experiment could vary with rurality, and the model could therefore also be used in the experimental design. The model in combination with such experiments could then be used to predict outcomes in different areas.

3.2 Paper 2: Policy implications for improved cook stove programs — A case study of the importance of village fuel use variations

Paper 2 deals with possible implications of local variations in fuel mixes for ICS programs. A large-scale distribution of improved cook stoves (ICS) could lead to a reduction in emissions harmful to people’s health and the climate, as well as a reduction in the time, or money, spent obtaining firewood. However, most ICS programs have displayed a low rate of success (Bailis et al., 2009; Gifford, 2010). Success has mostly been restricted to areas where firewood or charcoal is already purchased (predominantly urban areas), thus making ICS an attractive option to save money (Barnes et al., 1994), as well as areas where there is a firewood shortage. This may be due to that households don’t consider the benefits, saved time spent on firewood collection and improved indoor environment, to outweigh the inconveniences from having to change their cooking practice. However, even if the design of the stoves attracted households to actually use them, without effectively saving money, household may not be induced to maintain stoves, thus limiting the program to a short-term intervention.

Many rural households utilize several types of fuel. This may be due to limited access to collectable fuels, limited time available for collection, economic restrictions, or the fact that different fuels are more or less suitable for various purposes (Masera et al., 2000). Households may collect fuel from several sources and may also purchase several different types of commercial fuels. These fuel use patterns may to a large degree be area-dependent; see Section 3.1. These differences may have implications for stove and program design. Furthermore, depending on these fuel combinations, ICS programs may face varying chances of success in different areas.

The aim of this paper is to estimate variations between different geographical areas in terms of possibilities and incentives to sustain a market for ICSs and whether the potential reductions in CO_2 equivalents may vary substantially between areas. Furthermore, the paper aims to explore the possible implications of multiple fuel use on an ICS distribution program and to examine whether this has any implications for stove design.

A rural energy survey was carried out in the Vĩnh Phúc province in northern Vietnam. Using household data from six communes, each household is treated as a separate model, for which emission reductions and monetary savings are calculated. The results are then presented as means and medians per village for various combinations of stove assumptions and behavioral assumption. Emissions per kilogram of biomass used are calculated for two different stoves: one representing a cheap option based on the former generation of stoves distributed in China, and one being a top-of-the-line fan-powered gasifier stove. The reduced need for firewood per cooked meal is allowed to vary from 10% to 80%. The calculations are also performed for two different assumptions regarding the response of the household to obtaining an ICS. The first assumption is that households keep using their existing fuel mix even after obtaining the ICS. The second assumption is that households alter their behavior towards a higher degree of biomass use if this becomes more economical. The latter is assumed to occur in two instances: first, when collected biomass is assumed to last longer, due to higher stove efficiency, and thus is able to displace commercial fuels; and second, when commercial firewood becomes more economical on a per-meal basis, households are assumed to replace their use of LPG and/or coal in favor of commercial firewood.

When it comes to money saved as a function of stove efficiency, both linear relationships and relationships with both decreasing and increasing derivatives are found in the studied villages. The linear relationship is found for villages already dominated by commercial firewood, since no displacement effects are present. In contrast, the relationship with decreasing derivatives occurs in areas where the fuels used are primarily a mix of collected fuel and commercial biomass, meaning that the positive effect on monetary savings reaches a saturation point when the commercial fuels have been displaced, after which further improvements do not result in further monetary savings. The opposite occurs in areas dominated by fossil fuels, with higher stove efficiencies allowing the small amount of collected fuels to displace more and more commercial fuels, and more households abandoning coal and LPG in favor of firewood as a consequence.

The emission calculation shows that, if only the gases currently included in the Kyoto Protocol are included in the calculation, with high efficiency stoves, the reduction of GHG emissions in a fossil fuel-dominated village is on par with or even surpasses the other villages. However, when black carbon is also included, the total CO2eq reductions are larger in the villages that are more reliant on biomass. It should be noted, however, that it is difficult to assign a GWP value to black carbon (Bond and Sun, 2005), and consequently a GWP value at the lower end of the current estimations is used. For several conditions there is strong dependence on the (highly uncertain) assumptions regarding how households will alter their fuel mix, leading to a need for further research in this area. It should be emphasized that the paper is limited to the possible implications of variations in fuel mixes on ICS programs. There are many other aspects that designers of ICS programs may want to consider. It is also limited to the activity of cooking, while the open fire may provide other energy services not provided by an ICS (Ruiz-Mercado and Masera, 2015).

3.3 Paper 3: On LPG use in rural Vietnamese households

Models for fuel switching in previous studies have used a large number of different variables (Lewis and Pattanayak, 2012; van der Kroon et al., 2013). However, when the aim is not to predict a dependent variable based on a number of independent variables, cf. Paper 1, but instead the causal effect of the individual independent variables when altered, complete knowledge (or assumption) of the data generating process is necessary in order to obtain unbiased estimates (Pearl, 2009). Without this information, a regression model can only be used for predicting y from a given x, and not what would happen to y if x were changed. In econometric applications of regression modeling, the data generating processes are assumed to be known. Yet if the process of fuel switching is known, why are there differences in model specifications between the studies? If instead assuming that the process is not known, there are more efficient tools that are less relying on the modelers assumptions than regression. The interpretation will however be different, and the results should either be used for developing further non-causal modeling or for suggesting further research.

The method used in Paper 3 is a rather new machine learning algorithm that has gained much attention within the medical sciences, especially in genetics-related research, called Random Forests (Breiman, 2001; Strobl et al., 2009b). Normally, papers dealing with fuel switching have often utilized either logit or probit regression models. The econometric approach would be to formulate a model by reasoning, based on theory and previous research, while the statistical approach would be to use some sort of data driven variable selection approach. The reasons for choosing a data mining approach over the econometric approach is that there is no absolute certainty regarding which variables to include in such a model. This will be further discussed in Sections 4.4 and 4.5.

Assuming a statistical approach, there are several reasons for choosing the Random Forest rather than regression. First, regression can be very unstable in a variable selection approach. It is also sensitive to the choice of functional form. Both of these problems are magnified due to multicollinearity if there are many correlated variables considered, and especially if all possible interactions are to be included. And finally, any ranking between variables is not well defined without a given model specification. With Random Forests, the functional form need not be specified in advance, and all possible interactions are automatically considered. The Random Forest algorithm enables estimating the relative importance of the included variables for predicting the outcome which also can be used for a more stable variable selection.

It should be noted that the results of this paper can only be used as guidance for further studies, not as found causal effects or directly as basis for policy. If the results are interpreted in such a way they are still relying on assumptions. One can think of the method in Paper 3 as a way of moving some of the assumptions to after the modeling, i.e., the model results are less affected by the researcher's preconceptions, compared with an econometric model. Assumptions are then needed when interpreting the results, possibly with the aid of previous research. Nevertheless, it presents a possibility for a more unified approach of research when the underlying model is unknown.

Data collected during the rural electrification program in Vietnam (Khandker et al., 2009) is used for Paper 3. Although this data set is less detailed than the data set used in Paper 2 when it comes to amount of fuels used, it does enable the construction of a model whereby the households' main fuel is the dependent variable. In Paper 3, the Random Forest algorithm is used to classify households based on whether they are LPG users or not. The data electrification program data is useful because of its larger sample size, representativeness of much of rural Vietnam, and that the data was collected from three different years.

The aim of this paper is primarily to perform an exploratory data analysis, through the use of Random Forest type algorithm, to find a set of factors that are associated with rural households that are using, or will start to use, LPG. The outcomes of the algorithm represent a ranking of the most important variables for classifying households as LPG or non-LPG users as well as classifying households into future LPG users (i.e. predicting who will have started to use LPG in 2008 based on what is known in 2005).

Apart from current income, it appears that the history of household income and wealth (as indicated by number of appliances, as well as certain appliances) are important predictors for fuel choice. Considering the results for current fuel use, it is possible that an earlier high level of income may have enabled households to make this transition in the past and the households in this case continues to use an already purchased stove. However, past income was shown to also be a useful predictor of future fuel choice, possibly indicating that households need to have a stable income for a long period of time before making the transition. The high upfront cost involved when switching to LPG has been pointed out as a prohibitive factor. Although in many cases the household would actually save money in the long run by making this switch, the payback time is not short enough or there are no liquid funds in the first place to make the purchase (Edwards and Langpap, 2005; Heltberg, 2005). The algorithm also chose the number of appliances as well as refrigerator and rice cooker; this can be interpreted as solely signifying household wealth. However, the importance of refrigerator and rice cooker may also signify that these appliances complement the LPG stove in accomplishing the services previously performed by the open fire.

Furthermore, household density and distance to the nearest town, two variables found in Paper 1 to predict not only fuel choice but also the amount of collectable fuels used, were found also in this paper to be useful predictors of LPG usage on household level and were chosen by the algorithm despite the plethora of other variables available to choose from. However, relatively richer households are less sensitive to these two parameters.

Another important outcome is that many of the variables used in several of the previous studies did not increase the predictive ability of the model when included. Although, this fact does not prove that they wouldn’t be included given a larger data set nor that there is no causal link between them, it is still interesting that fuel choices can be modeled, in this data set, to an equally or even higher degree of predictive accuracy without these variables, using only a history of income and measurements of rurality.

The results in Paper 3 can be used to guide future regression modeling, of the predictive type, such as the model in Paper 1. Furthermore, the results can be used as a basis for further studies, aiming to test whether any causality is involved between the chosen predictors and LPG usage. It can also be used as a check for the econometric type of models; if parameters not previously included are found to be of high importance in the Random Forest modeling, there may be a need to rethink the econometric model. In a causal model, both the inclusion of variables that should not be there and the exclusion of important variables may cause results to be biased, depending on what effect that is to be measured and the causal structure. These issues will be discussed further in Sections 4.4 and 4.5.

3.4 Paper 4: Energy efficiency at the base of the pyramid: A system-based market model for improved cooking stove adoption

As described in Section 2.2.3, programs trying to distribute ICS to households have a long history. Most of them have had limited success, with the Chinese program in the 1980s as an important exception (Smith et al., 1993). At the same time as households are using a vast amount of biomass inefficiently, damaging their health and the environment, many nations are aiming to increase the share of biomass in their electricity production. Reasons for this can be environmental concerns, but energy security concerns and economic objectives may also be involved. The wood used residentially by households constitutes a large part of the total energy used in Asia (as defined by International Energy Agency, and excluding China). The estimated annual amount used is 270 million tons oil equivalent (MTOE), while the amount of coal used for electricity production in the same area is 260 MTOE (IEA, 2012). Paper 4 is an attempt at designing a program that uses a growing demand for biomass, together with an updated CDM in order to help poor households in the developing world to obtain better cooking technology and at the same time increase their income and stimulate national development of biomass technologies.

This paper proposed a model in which households willing to install an ICS are allowed to sell excess collected biomass to an agent who in turn sells the biomass to, e.g., electricity-production companies. The amount households are allowed to sell can either be regulated by the price or by setting a fixed quota each household are allowed to sell. Households are thus given incentives to both purchase stoves and use them efficiently. In this way the model mimics both the Chinese program but also a more urban setting where fuels are often commercial; see Section 2.2.3 and (Barnes et al., 1994; Smith et al., 1993). Note also that the extra income coupled with a payment plan, this model may overcome both the income and liquidity problems often pointed out as principal barriers to commercial stove distribution. Furthermore, since the model couples efficient use of firewood with extra income for the households, it may create a market feedback mechanism where households and stove manufacturers interact to continuously improve stove performance and suitability for households.

Two methods are used in this paper. First, a cost and emission calculation of a simple hypothetical program, based on the model principles, is performed in order to establish whether it is possible to finance the program based on selling carbon credits based on the anticipated reductions in emissions, both with and without black carbon. Second, the policy model and its possible consequences are qualitatively evaluated based on previous literature on fuel choices, firewood collection, ICS programs, and deforestation.

Under some assumptions, the paper finds that the model can be financed solely based on reduced carbon dioxide emissions from displaced coal. The cost calculation shows that at a price of CERs within historical values, the logistics of the program is covered together with compensation on par with current minimum wages in parts of the developing world, if firewood collection were a full time job.

Households are generally sensitive to the shadow cost of work; increasing the price paid for firewood will likely increase gathering, but also increases the shadow price of saved wood. If a functioning market already exists, the price increase will provide a greater incentive to use the wood more efficiently. If no prior market exists, each additional mass unit of wood the household collects will have a higher opportunity cost than the unit before. Allowing the households to sell some wood would increase the incentives for reduced use of firewood.

Based on current knowledge, the program appears fully functional and furthermore can be financed via carbon credits from reduced emissions. However, further studies are needed in order to fully understand the possible consequences of the policy. For example, possible long-term consequences include risks for deforestation or the alteration of tree species, and hence may have implications for local ecosystems. Furthermore, there may be negative consequences for non-collecting households that are currently purchasing firewood, through an increased price on firewood. Furthermore, how to implement and monitor the program, and regulate a quota remains an unresolved issue.

3.4.1 A note on implementation and further discussion

The paper is conceptual in nature and cannot address all issues related to the proposed model. Questions regarding practical implementation of the model are avoided, to prevent the proposed model from being permanently coupled to particular details of implementation. Moreover, a reader of the paper may worry that the proposed model is much too complex, and the streams of money and an artificially raised price on firewood are all frail mechanisms that invite corruption. In this section, a number of these questions will be addressed by invoking examples of possible implementations. Note however that these are not the only possible implementations of the proposed model. An issue that has been raised by reviewers of the paper is how to prevent raising prices on wood sold to households that have purchased an ICS, and how a quota can be enforced.

In order to enforce a quota, one must be able to easily identify the households. One possibility is to register the fingerprints of household members, which can then be used to identify the households through a mobile device (such as a modern smartphone). If a quota is enforced, every time a household sells wood they must identify themselves, and the same household cannot sell more wood until the following week. With a price set higher than the current local market price, the household with an ICS may choose, instead of collecting, to purchase wood from the market and then sell it on. However, with the quota in place, the total extra outtake remains the same while the households face better incentives than before to use the wood efficiently.

A related objection, under circumstances where no quota is thought to be needed, is that it would be possible for a village to designate one household to purchase an ICS while the other households in the village only sell through this household. With no quota set per household, this could be a strategy (and may be one further reason for enforcing a quota). Note however that the underlying assumption of the model is that there are ICSs that are suitable for the households and enable them to save wood, and therefore installing an ICS is the economic alternative because firewood collection is limited by the amount of time one is able or willing to spend on firewood collection.

By using mobile devices identifying households, the whole chain can be operated without any exchange of paper money. When a household is identified by fingerprints, the appropriate sum can automatically be transferred to the household’s mobile account. Mobile phones are already used extensively for mobile banking in the developing world, through various solutions using phone numbers as identification and without requiring smart phone technology. It should also be mentioned here that the cost of a low-end mobile phone today is only a fraction of the costs of an advanced ICS and can be included in the payment plant if the household does not have one already. Furthermore, by using mobile technology, the amount the drivers collect can automatically be monitored and compared with what is delivered to the final destination, e.g., a power plant. Mobile technology has already proven an important method enabling poor and rural households in the developing world to use bank services (Dermish et al., 2011) and indications that it also could be used as a mean to decrease corruption has also been reported (Economist, 2013).

4 Reflections on methodology

This chapter discusses the methods used in the appended papers, which are the main work of this thesis. Section 4.1 presents some background on the choices made for the case study and survey, which Paper 2 and parts of Paper 1 are based on. In sections 4.2 and 4.3 the methods and modeling used in Paper 2 and Paper 4 are discussed. Section 4.4 discusses the choices of statistical models used in Paper 1 and 3, and argues for how these models can be a valuable complement to causal modeling. Section 4.5 describes the Random Forest and bagging algorithm in greater detail and is accompanied by simulations that may clarify the interpretation of the results in Paper 4.

4.1 Case studies, selection and survey

Papers 1 (partly) and 2 (fully) are based on data from a survey conducted in six communes in the Vĩnh Phúc province in northern Vietnam. The Vĩnh Phúc province in the northwestern part of the Red River delta was chosen for the study because it encompasses three distinctive geographical areas, the delta lowland, hilly midlands, and the mountainous areas that characterize much of northern Vietnam (although other provinces also have this characteristic). Two communes were chosen from each of these geographical areas. The purpose of this was to investigate whether differences in fuel patterns between the communes appeared to be related to geographical area or if differences on commune level was such that other descriptions on commune level is needed to capture any differences.

The survey was designed in cooperation with the Institute of Energy in Hanoi and was carried out by that Institute and the Vietnam Women's Union. The questionnaire was given to 40 households in each commune, for a total of 240 households. The questionnaire covered the households' basic socio-economic status in aspects such as income, education and household size, and agricultural practices along with current energy use for cooking, space heating, water-heating, and electrical appliances.

The data collection, and what conclusions can be drawn from studying it, can be criticized at both commune and household level. At the commune level because of the small number of communes that were sampled, and also because no clear sampling strategy was employed except choosing two representatives from each topographical area. To overcome the imperfections at village level, recasting groups and omitting some groups in order to check whether this still leads to the same conclusions has been implicitly performed in the papers. This is standard conduct in case studies (Yin, 2009).

The randomness in selection of the households might also be questioned, and thus also how well statistical methods can be used to describe the communes based on the sampled households. This flaw is due to the absence of a list of households from which the survey participants can be randomly selected prior to data collection. In the absence of such a list, households have been sampled according to certain rules, such as "ask each 20th house one passes." This approach may still miss certain categories of households and thus be biased and, furthermore, only certain parts of the area intended to be represented may actually have been sampled. However, one can argue that even if the actual communes are not properly reflected, the conclusions remain valid for the samples, and thus for areas with such characteristics. The main conclusions for Paper 1 were also validated by using a second data set.

Because of the small number of villages sampled, the phrase "a case study" is appended to the title of Paper 2. This terminology is meant to signify the limitations and uncertainties of generalizing in a case study. It does not mean an in-depth analysis of a specific case, which is the most common use of "case study." The study would have benefited from data collection from more communes. It is called a case study on account of limitations in the sample size.

4.2 Modeling in Paper 2

Although ICS programs have been most successful in areas with constraints on firewood collection (Barnes et al., 1994) (Bailis et al., 2009), cost-benefit analyses of improved cook stoves generally assume that households use a single biomass fuel; see, e.g., (Smith and Haigler, 2008). This assumption of a single source is true for many areas, but where firewood collection is restricted households are more prone to both adopt ICS and to use more varied fuels.

In a resource-constrained environment, households often use different fuels for similar purposes. In China (Kaul and Liu, 1992) and Vietnam (Tuan and Lefevre, 1996; and Paper 2), a combination of collected and purchased wood, agricultural residues, and coal are often used, on a yearly basis, to provide the total household energy needed for cooking. The reasons for this can for example be seasonal variations in availability for different fuels (Kaul and Liu, 1992) or the fact different fuels may be more or less suitable for different purposes (Masera et al., 2000).

There are many uncertainties regarding the benefits of an improved stove program. For example: Do improved cook stoves help households reduce their need of firewood as described by Bensch and Peters (2012), or is there no decrease in firewood use at all (Hanna et al., 2012)? How do emissions and efficiencies measured in laboratories relate to actual emissions in households (Roden et al., 2009)? How will stove usage change over time (Hanna et al., 2012)? And how are health issues related to reductions in emission levels (Smith and Peel, 2010)? What GWP value should be assigned to black and organic carbon, respectively, and what is their mass ratio in emissions from various cooking technologies in actual use by households (Bond et al., 2013)? Furthermore, there are different methods for calculating reduced need for firewood and the share of this wood that is to be assumed sustainable (Lee et al., 2013). In Paper 2, it is explored how the introduction of an ICS may change household fuel choice and how this may influence the effects of ICS adoption. This adds more uncertainties regarding the benefits of ICS dissemination.

The modeling in Paper 2 is not a system model, but an aggregate of several separate household models in which the effects of ICS adoption are calculated for various assumptions. The model calculates the emissions and money saved for different assumptions on stove performance and at different efficiency increases. The assumptions include how households will alter their fuel mix as a result of an increased efficiency of biomass combustion. The purpose of the model is not to make accurate predictions but rather to identify the uncertainties in benefits that arise, both from actual differences between areas and due to the lack of knowledge about how an improved stove may alter household fuel choices. However, the article also suggests possibly exploiting these differences for initial deployment. A widespread adoption of ICS also implies possible system effects, which are not accounted for in the aggregated household models. As an example, firewood prices and the efficiency of collecting firewood may be affected by whether and/or how other households use an ICS, thereby reducing the total collective need for firewood.

It is also a simplification on household basis since it simply assumes a proportional displacement of other fuels as cooking with biomass becomes more efficient. However, households might use the ICS for just some of their cooking tasks and as a complement to the open fire. Recent measurements show that households often use ICSs, together with the open fire as well as with more modern fuels (Ruiz-Mercado and Masera, 2015).

4.3 Modeling in Paper 4

There are two models in Paper 4, one concerning the logistics cost of the proposed model and one concerning the anticipated household response to the program. The paper asks “would this scheme be possible?” and answers from these two perspectives. The logistics model is a simplistic calculation repeated for a number of different input assumptions. The model of household participation and response in the program is based on previous research, but formulated mathematically.

The modeling choices are intended to not be specific in terms of locations and assumptions. Furthermore, many more questions remain before the proposed model’s success in a real life setting can be tested. For example, how to design the organizational body of the proposed program and how to ensure proper accounting at both household and institutional level would need careful attention. Rigorous procedures for similar projects are needed, particularly in light of the criticisms of previous and current CDM projects as discussed in section 2.4.

Including more site-specific information and technical options might find lower, but site-specific cost estimates in the logistics model. As long as the question is whether the proposed model is feasible, however, a positive answer based on a suboptimal solution is sufficient. But Paper 4 should be understood as a first step in proposing and evaluating a new type of program, and certainly not as a policy recommendation, at this stage.

A model that balances the investment costs of learning and changes in cooking practices, and maintains a sustainable removal of forest products without obstruction of local agriculture, can be investigated through quantitative modeling. However, this model would require sophisticated inputs including fuel collectors’ opportunity cost of time, values for the cost of collecting wood in man-hours, the households’ perceived costs of learning and adopting new cooking practices, the perceived benefits from stoves, the loss of other values derived from open fires, information on local biomass regrowth and status, the assumption of not placing quotas on purchase of biomass fuel, and knowledge of stove performance and maintenance costs. This data would be difficult to collect.

4.4 Causal modeling and data mining

Statistical models based on observational (non-experimental) data can be of two different types, causal or associative. Causal modeling attempts to answer the question: if one changes X, what happens to Y? Which is different from associative modeling, which asks the questions: if one finds X, what is Y?

The models used in this thesis are both associative models. Paper 1 uses traditional regression techniques, while Paper 3 employs Random Forest, a novel machine learning, or data mining, method invented in 2001 by Leo Breiman (Breiman, 2001). (Although for the most part “bagging,” a subset of Random Forest is used.)

Whether meaningful causal statements can be made after studying non-experimental data was heavily debated during the last century, with mainly mathematical statisticians on one side arguing the negative, while academics from other disciplines argued the affirmative. For the purpose of the subsequent discussion, causal modeling is regarded as meaningful, in accord with the conventions of the field of research of Papers 1 and 3. It will be argued here, however, why also the non-causal models are meaningful both intrinsically, and for questioning and complementing the results of previous causal models, and possibly for guiding future analyses.

In this thesis two approaches of associative modeling are employed. The first is a straightforward linear regression model without causal interpretation, but which may actually be useful for prediction purposes. Furthermore, this model can be seen as an orientation of fuel choices in Vietnam. The second technique is a method for data mining called Random Forest. It is argued below why such an approach may complement the causal modeling approaches.

If one wishes to construct a causal model, it may be useful to know that these fuel choices vary with rurality. Conclusions drawn about the causality of other variables that also vary with rurality, without a full description of the other aspects of rurality that may influence fuel choices, may be faulty.

Paper 1, for example, investigates how rurality correlates with fuel choice. Although not the main aim, one possible use of this relation can be to question some of the causal statements that are made regarding fuel choices, both in terms of existence of any causal link but also the size of such effects. The degree of rurality of an area could be a confounding variable for many of the variables found to be influential by various studies, such as income, availability of collectable and modern fuels, occupation, education, and electrification. Should as researcher find a correlation between a variable and fuel choice and believes this to have causal effect it may be useful to also perform the regression with measurements of rurality included. Although strictly, this procedure can neither prove nor disprove the existence of a causal link, it can be informative for the reader to know whether the found effect is heavily altered or vanishes in the presence of a measurement of rurality, just to point towards the possibility of alternative interpretations.

4.4.1 Causal modeling

A causal model is supposed to be derived from theory and previous knowledge. Econometric causal models should be derived from economic theory (for example Gundimeda and Köhlin, 2008). Causal modeling of observational data measures the effect of one explanatory variable on an outcome of interest. This practice is not criticized here, despite its hazards, because in many cases this is the only possible approach. However, careful consideration is needed not only to which factors influence the outcome, but also to how these factors influence each other. Furthermore, one needs to specify, assuming not all explanatory factors are independent, which of the variables one is trying to measure, i.e., the same regression model may not be suitable for measuring the effect of all the included variables at the same time. If uncertainties persist about which other explanatory variables are needed to reduce bias, one can explore how much a variable affects the result.

To measure causal effects with a regression model using observational data, one assumes that there is a causal process, that the form of this causal structure is known, with regards to which variables affect each other and the function that maps this effect from certain variables onto others, that all the needed variables are available (not necessarily all the variables that are involved in the process) and that the variables are measured accurately, or that any measurement error in the explanatory variables are low (if the size of this error is known this can be adjusted for). Under these conditions it is possible to measure the effect of various explanatory variables on the dependent variables.

In some cases results can falsify the assumed causal structure, but the results cannot prove that the assumed causal model is correct. This point is consequential for the interpretation of many studies, not least of which are review papers that summarize many regression studies measuring the significance of explanatory variables; this will be discussed later. In summary, causal modeling can be used when one wants to measure the causal effect of one variable on another, a causal relationship is assumed to be known, and the needed variables are available. And although observational data can be used to statistically prove that the causal model is wrong, this is not always done.

However, causal (econometric) modeling should not be used for discovering which variables affect the outcome, but to measure the causal effects that are assumed. (There are algorithms that can in some circumstances detect causal structures in data, although they also rely on assumptions. However these algorithms are not what is discussed here.) Many studies employ a partly econometric and partly exploratory approach in which a number of variables are assumed to be able to affect the outcome. All the variables are then included in the same regression model, and causality is inferred for the variables found to be significant. However, the inclusion or exclusions of variables can cause other variables to both loose and gain significance. Furthermore, as noted, lack of adherence to causal structure between the included explanatory variables makes it unclear what effects one is actually measuring. The inclusion of several correlated variables in the regression also increases the variation of the estimates, and can cause parameter estimations to change both in sign and size (this last statement is true also without accepting a causal interpretation of the regressions). Because of these issues, interpreting the results of such models can be difficult. Since these models are often neither tested nor designed for prediction, and the variables are not chosen to be suitable for prediction, the utility of such models in this regard, can be limited.

It is argued here that in some applications, previous modeling can benefit from being complemented with more data driven exploratory approaches. The applied economic theory, which by necessity simplifies real circumstances, may not encompass all the possible choices and circumstances faced by individuals and households. This might lead to both the omission of variables and the wrong functional forms. If omitted variables are confounders, their omission may bias estimates of the variables of interest. Furthermore, the data used for estimation may not properly reflect the theoretical variables, which can also produce bias. Data collections often contain more variables than are, or could be, included in econometric approaches. Using approaches that allow the inclusion of these in an analysis may lead to clues for further investigations.

4.4.2 Data mining as a complement

Data mining is often used as a derogatory term in econometric contexts, but the econometric and machine learning fields may be using different definitions of the term. The derogatory use of "data mining" seems to describe the dubious method of running random regressions and making assertions of causality for any significant relationship that appears.

One of the most effective methods for associative modeling is Random Forest. Random Forest is an ensemble method, meaning that a large number of weak learners (a weak learner being a classifier that is only weakly correlated with the variable of interest), classification or regression trees are grown and the prediction of the forest is the mean prediction (or majority vote) of all the individual trees. In order to achieve variation between the trees, a random subsample of the observations is used for each tree, and a random subset of the variables is considered for partitioning in each node.

This section is concerned with the reasons for using Random Forest as a complement to more traditional regression and econometric techniques; please refer to section 4.5 for a more technical introduction.

The literature on household's fuel choice illustrates why Random Forest is a useful complement to traditional regression techniques. During the last few decades, dozens of papers have been written about fuel switching. Though most of them agree on income as an important parameter, there are substantial differences in the other parameters included (Lewis and Pattanayak, 2012; van der Kroon et al., 2013). Consequently, which variables end up being significant in a single regression depends to a great deal on what variables are included in the first place. Different papers thus end up with different sets of significant variables, sometimes leading to different policy suggestions. Not only is it important to include the right variables, but the functional form, must be specified before the measuring the causal effect.

In Random Forest it is not necessary to specify functional form, including interactions, before the analysis. Considering the various claims of relationships made in different papers, it would be useful to achieve some ranking of how important the various variables are to the determination of fuel choices, in order to guide further studies. In parametric regression, the term *importance* is not clearly defined (although, given a certain model, then such a ranking could also be performed in parametric settings). How importance is calculated in the Random Forest framework will be explained in section 4.6.

Random Forest may serve as a benchmark for the predictive accuracy possible to achieve given the data for classification or regression problem at hand (Strobl et al., 2009b). If substantial improvement in predictive ability can be obtained through Random Forest over a parametric model, there may be reason to re-evaluate the original model. If prediction when using Random Forest is highly dependent on variables not initially included in the parametric model, or if their functional form is different, there might be a need to re-evaluate the original parametric model. However, a better prediction in a new model does not necessarily imply that previous parameter estimates are wrong; this is depends on the causal structure. Furthermore, the inclusion of other variables may also be a sign of measurement errors. This use of Random Forest can then be used to check the models that estimate the causal effect of a selected variable where the causal structure and functional forms was assumed to be known. It is however important to note that the interpretation of the Random Forest results does not necessarily imply causal relationships but are only a measurement of the predictive capabilities.

There are, of course, also downsides to using Random Forest. The most obvious is that there are no regression coefficients with corresponding error margins. Furthermore, although one does not need to specify interactions, obtaining those relations after the forest has been grown is not entirely straightforward. It is, however, possible to graph the relationship between a variable and a response, called the partial dependence plot. This is done by creating new data sets for each value of the variable of interest, holding this variable constant in each data set, while all the other variables take on their original values. Subsequently running each new data set through Random Forest generates mean predictions, which can be plotted for each value, and a curve describing the relationship is obtained. The partial dependence can be plotted with different values of other variables, and differences seen in the partial dependence plots can then indicate interaction effects.

4.5 More on Random Forests and bagging

Random Forest and bagging are ensemble methods based on Classification And Regression Trees (CART). The aim of these methods is to achieve as high prediction capabilities as possible, given the data and the classification or regression problem at hand.

In CART, a search algorithm finds cutting point in one of the predictor variables for dividing the data set into two subsets for optimal sorting of the outcome variable. The two data sets are then further subdivided until a stopping criterion is reached (Breiman et al., 1984; Strobl et al., 2009b). The result is a tree structure where the data is divided into purer and purer groups in each node. A single tree is suitable for interpretation and can reveal interactions. However, the CART algorithm is very sensitive to small changes in the data, i.e., repeating the experiment by collecting new data or bootstrapping the original data often lead to completely different trees.

Bagging and Random Forest grows a large number of trees and uses averages for regression or a majority vote for classification. This technique is generally referred to as bagging (Breiman, 1996). Subsamples are made by drawing almost two-thirds (0.632) of the data, and a new tree is constructed for each subsample.

A split in a single tree is only optimal in its node and does not account for further subdivisions. To find the variables and splits that would be optimal for the whole tree would be a computation-heavy operation. Introducing an aspect of randomness in the variable selection may ensure that more splits are included and subsequently enhance prediction (Strobl et al., 2009b). On top of the bagging procedure, in Random Forests, only a random subset of predictor variables in each node is considered in order to increase independence between trees. This procedure has been proven to reduce prediction errors (Breiman, 2001).

However, constructing trees based on node purity may favor continuous variables, or factor variables with many levels, over those with fewer levels, simply because the existence of many possible split points increases the likelihood of finding an optimal split purely by chance. In order to achieve unbiased importance in the presence of mixed types of predictor variables, conditional trees have been introduced (Hothorn et al., 2006b) and subsequently Random Forests based on such trees (Strobl et al., 2009a). The main difference is that the variable chosen for the split is taken as the variable with the strongest association (in term of lowest p-value) in a statistical test and then the optimal split is sought only in this variable.

4.5.1 Out-of-Bag Error – an internal cross validation procedure

Random Forest provides an internal cross validation procedure. In each tree, only a randomly selected subset of the data is used for constructing the tree. Hence, after the whole forest has been grown, for each data point roughly one-third of the forest has been constructed without this data point. Each data point is then classified using only the third of the forest where that data point did not aid in construction. The misclassification rate for this procedure is referred to as the out-of-bag (OOB) error.

Breiman (1996) argues that the OOB error is as accurate as using testing and training data. However, if the OOB error is also used for variable selection and tuning, information leaks from the “out-of-bag” to the “bag”, hence a proper test data set should be used in these cases. To get unbiased OOB error estimates, it is necessary to grow as many trees as needed for convergence (Breiman, 2001), i.e., until the OOB error stabilizes. The OOB is more comparable to a cross-validation error than measurements based on R^2 that are based on the same data points as were used for construction of the model.

4.5.2 The output from the algorithms

As mentioned the algorithm does not provide any regression coefficients with associated errors. Instead the outcome is black box classifiers. In order for interpretation of the result partial dependence and importance of the variables are used.

4.5.2.1 Partial Dependence

Although the functional form need not be assumed and specified beforehand, the functional form is not a straightforward result from the algorithm either. Rather the outcomes, besides a black box classifier, include a ranking of the importance of the variables considered and the so-called partial dependence for each variable. The partial dependence describes the influence of a variable on a certain outcome.

The partial dependence is defined as (Berk et al., 2009):

$$\tilde{f}(x) = \frac{1}{n}\sum_{i=1}^{n} f(x, x_{iC}),$$

where x is the variable for which partial dependence is of interest and x_{iC} includes the other variables. For classification purposes, the function f is:

$$f(x) = \log p_k(x) - \frac{1}{K}\sum_{j=1}^{K} \log p_j(x)$$

where K represents the number of classes, k is the class for which the partial dependence of variable x is sought and p_j the proportion of votes for class j. A new data set is created for each value of the variable of interest, in which this variable is constant in each data set whereas all other values take on their original values. Each data set is then used to predict a single value. These values from are then used to construct a graph predicting the response at the different values of the variable of which the partial dependence is sought. No adherence to possible correlations is included and generated data points used for creating the partial dependence may include data combinations never observed in the original data. The partial dependence cannot be interpreted as a causal relationship and is not based on a statistical model (Berk et al., 2009). However, in an exploratory study, the ability to detect possible nonlinearities and interactions can be very useful for further analysis.

Although interactions and functional forms need not be specified but are rather generated, their reporting is not yet fully developed (Touw et al., 2013). However, it is possible to create bivariate partial dependence plots in three dimensions to examine the interaction of two variables (Cutler et al., 2007). The partial dependence can also be plotted, conditional on different values of other variables; differences in the shape of the relationship can then indicate interaction effects, which is the approach used in Paper 3.

4.5.2.2 Importance

The Random Forest algorithm has a built-in measurement of variable importance. Several different approaches to measuring the importance of outcomes are available, from simply calculating the number of times the variable is chosen in the forest, to the mean effect the variable has on node purity across the splits, and lastly (and the most commonly used approach), the reduction in prediction after the variable is permutated (Breiman and Cutler, 2013; Strobl et al., 2009b). By performing a permutation of a variable, the link between the predictor and the response is broken; hence, the decrease in prediction accuracy after the permutation can be taken as a measure of the importance of the variable.

However, in Random Forests, variables that are not needed for prediction may still receive high importance if they are correlated with variables truly useful for prediction, because only a subset of the variables are considered for splitting in each node. (Although bagging in many cases reduces this effect, correlated variables may be given too much importance; this will be explored further in 4.5.4).

In the ordinary Random Forest algorithm, the importance values may also be biased toward variables that are either continuous or multilevel. A good split is more likely to be found if many possible splits are available. This property is inherited from the CART algorithm. A solution is to instead use forests based on conditional trees (Hothorn et al., 2006a), where association is estimated in a separate procedure. An algorithm based on these trees has been developed, called Conditional Forest (Strobl et al., 2008).

Furthermore, in Conditional Forest another solution to the exaggerated importance of spurious correlation has been proposed, called conditional importance (Strobl et al., 2009a), where permutations are only performed within the cells defined by splits (the splits obtained when the forest was grown) in correlated variables. However, another possibility is to compare the Random Forest importance with the importance values achieved through bagging (Grömping, 2009). In this case, it seems important to check that the predictive abilities do not decrease when using bagging compared with Random Forest, since this would indicate that some splits usable for prediction are no longer found, which would mean a decreased importance for the variables involved in these splits.

4.5.3 Simulations using bagging

In this chapter we will explore how the Random Forest can detect associations in the data in controlled settings. Although a number of examples will be shown where the bagging algorithm is able to find the correct causal structure, such an interpretation of the results will still be argued against in Paper 3, backed up by further examples. However, the results in this section can still be useful for understanding bagging and the results of Paper 3.

In this section we illustrate both the possibility of recovering a causal structure in a data set and some possible pitfalls when doings so. These examples take the causal model depicted in Figure 1 as the starting point. In this example X1 and X2 are roots. X1 causes bX11 and bX12, while X2 causes bX21 and bX22. Together X1 and X2 cause Y. Variables Z1 to Z4 are unrelated to each other and all the other variables and are added as opportunity for the algorithm to make erroneous choices. The data is generated by drawing X1 and X2 from a normal distribution. The variables that are further down (children), including Y, are generated by adding the parents plus a normal distribution. For simplicity, all normal distributions are with the same mean and dispersion and all linear models have all coefficients equal to one. The variables bX11–bX22 are correlated with Y through their common parents; however, they are not themselves causes for Y. This is an example of a spurious correlation in causal analysis. Conditioning on the true causes should cause the effect of the spuriously correlated variables to disappear. In order to do this with Random Forest, the special case bagging must be employed, so that in every split the algorithm has access to all possible variables (or use conditional importance (Strobl et al., 2009a). In the bagging framework we then simply make sure that X1 and X2 are available for the algorithm and hope that the importance of bX11–bX22 will be approximately zero.

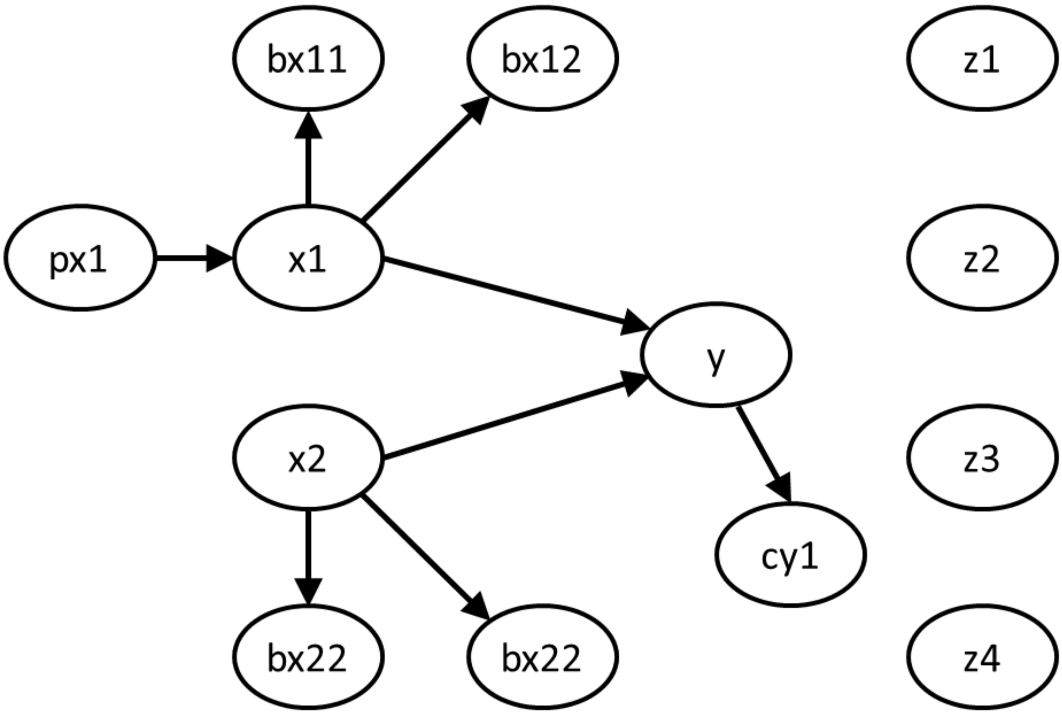

Figure 1: Causal model for simulations.

The results of 100 such forests, based on a structure as is shown in Figure 1, minus px1 and cy1, are shown in Figure 2. These results show that in this example the algorithm has no problem separating the spuriously correlated variables from the true causes.

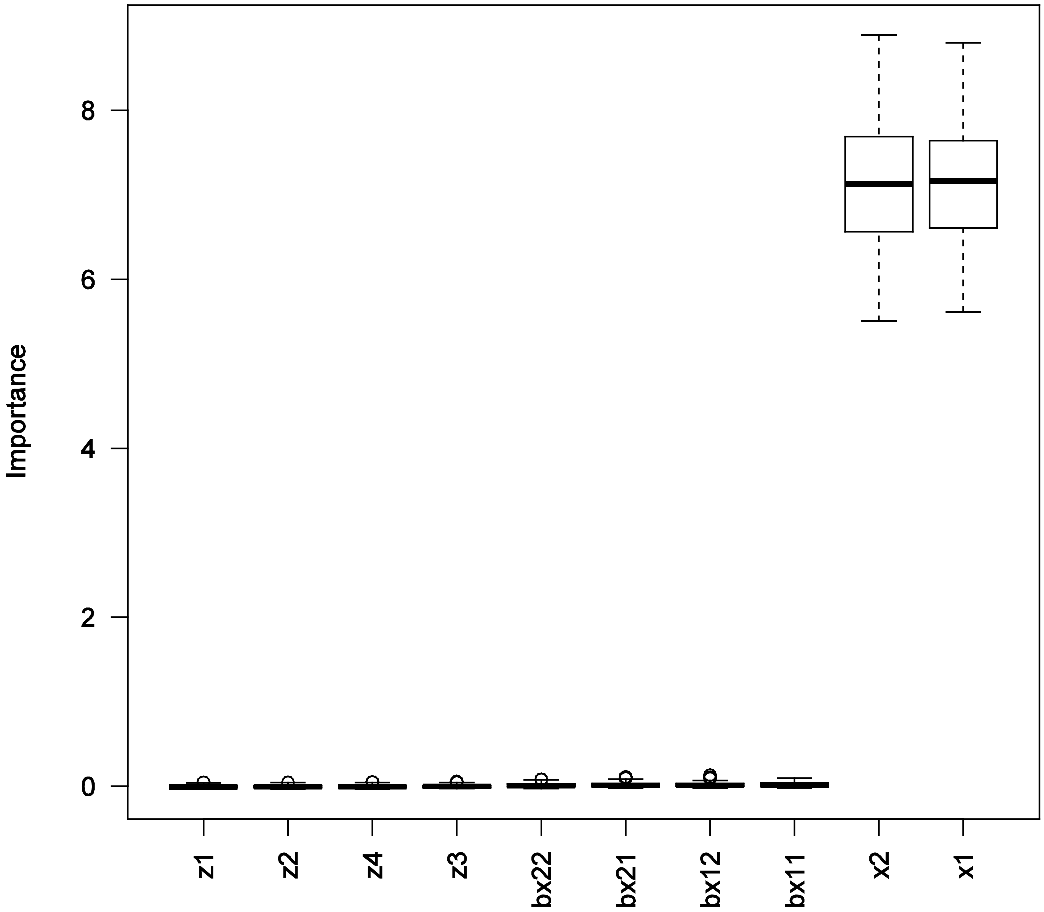

Figure 2: Spurious correlated variables (bx11–bx22) are not selected when the algorithm has access to the true causes.

If a parent to one of the true causes is included, i.e., a true cause for X1 (labeled px1), in this case too, the bagging algorithm has no problem separating immediate causes from causes “higher up the chain;” see Figure 3. These examples show that it might actually be possible to discern causal structures in the data, by using bagging. This can be achieved by starting from a variable for which dependencies are sought, and finding all immediate parents. New runs can in turn find the parents’ parents, and so on. However, using and interpreting the results in such a way assumes that all relevant variables are in the data (and measured accurately). Since this is likely an approach one would use in exploratory settings, this cannot be guaranteed other than in simulations.

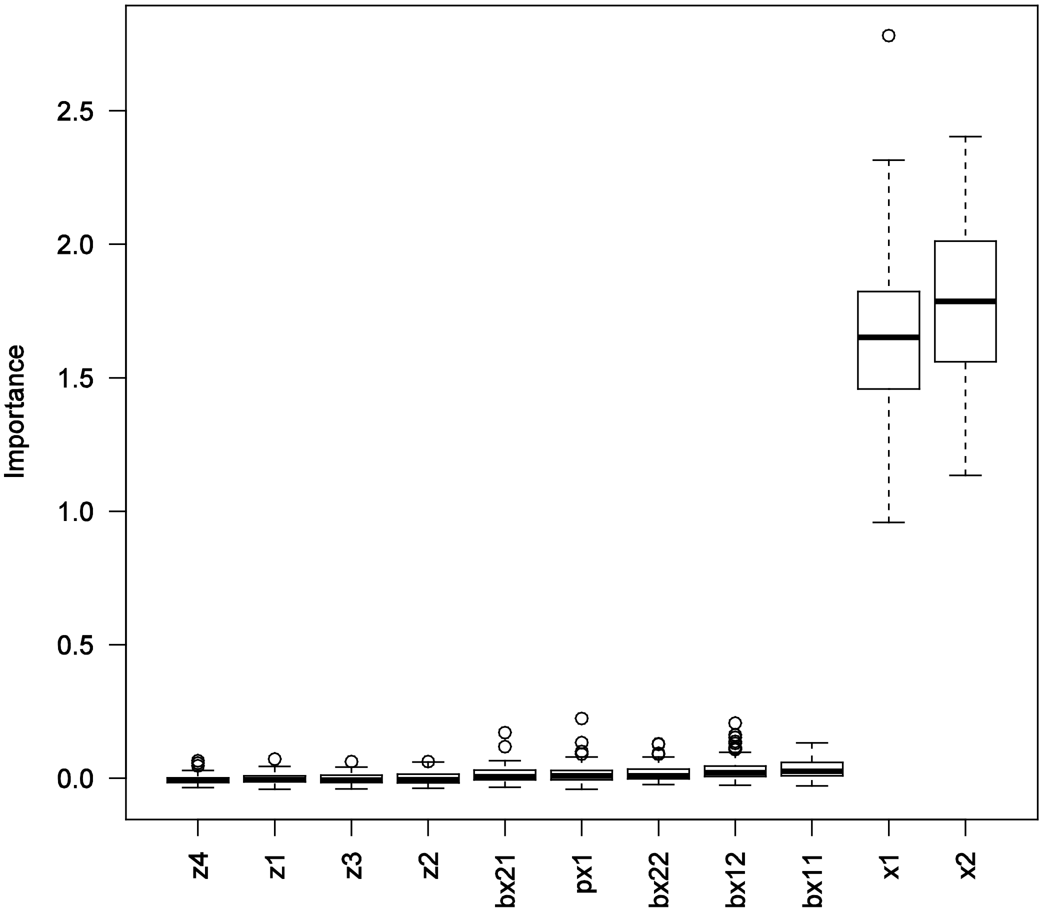

Figure 3: A parent of X1 (px1) is included in the simulation

As for any other approach, missing variables have consequences for the results of bagging as well. It the true cause is not included in the data, the algorithm will choose other variables correlated with the true cause. Figure 4 shows how the algorithm in such a case would use the spuriously correlated variables bX11 and bX12 to predict Y when X1 has been removed.

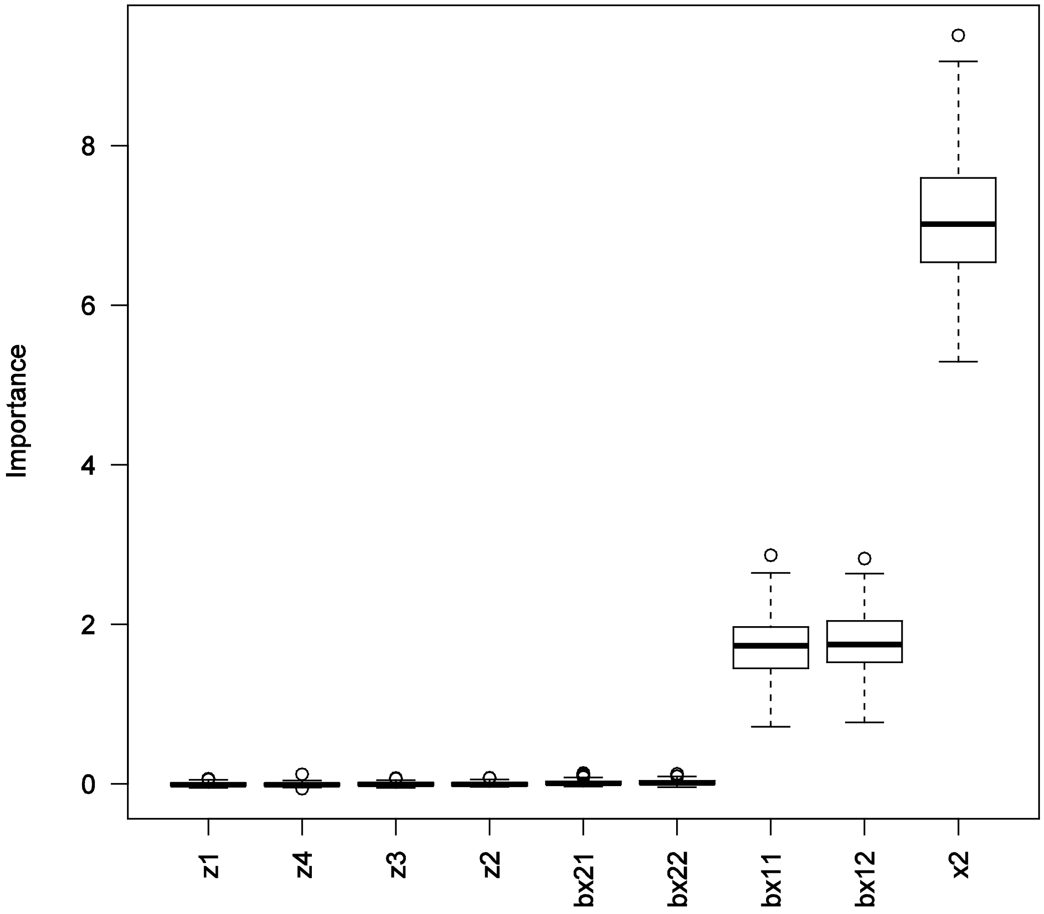

Figure 4: The omitted variable: Here X1 has been removed from the calculation, so the algorithm instead uses the spuriously correlated variables bX11 and bX12 to predict Y.

Data in many real world applications are seldom perfectly measured. Adding a measurement error to X1 has much the same effect on bX11 and bX12 as when X1 is completely removed, as seen in Figure 5. Now the algorithm uses all three variables (X1, bX11, bX12) for predicting Y. In Figure 5 a normally distributed measurement error has been added to X1 after the causal effect on bX11, bX12 and Y has been calculated. Since the measurement error is also drawn from a standard normal distribution, X1 with measurement error is essentially another bX11 or bX12. Note however that this measurement error is rather large compared to the original variable (the measurement error is drawn from a normal distribution with the same parameters as the one used for the original variable).

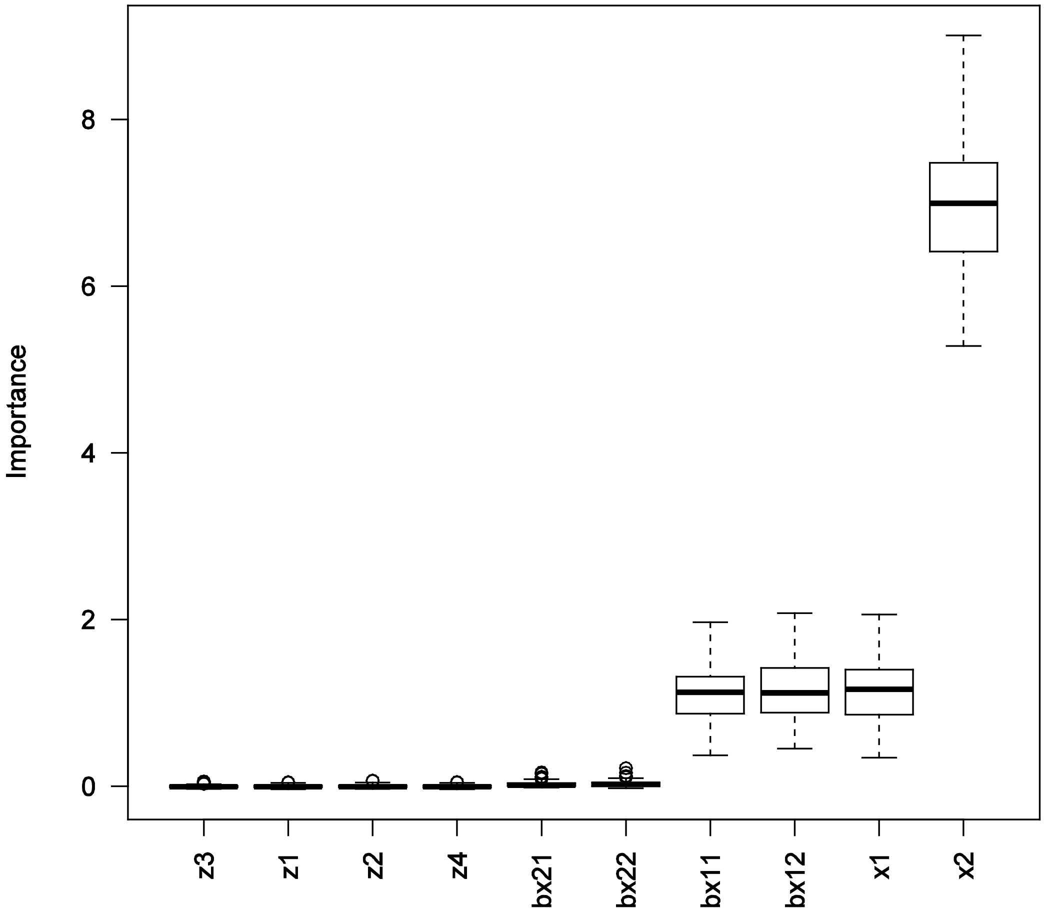

Figure 5: Measurement error is introduced in x1

A causal interpretation of the bagging results would also require knowledge of causal directions, i.e., which variable is the cause and which is the effect. Figure 6 is an example of the predicted variable itself causing an effect in another variable. The algorithm finds this variable and uses it for prediction. To use the information for meaningful models, the direction of causes must be established.

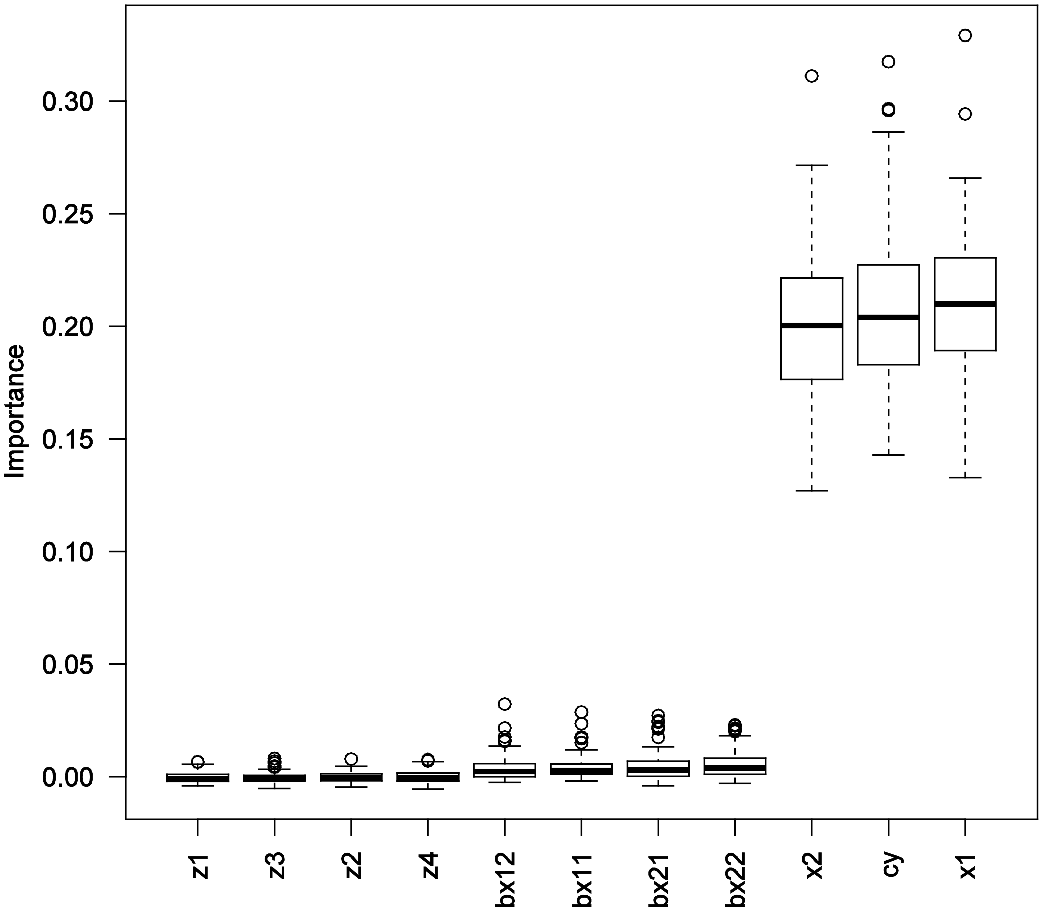

Figure 6: The data includes a child of the dependent variable (cy)

There are actually circumstances where directions of causality can be established through adhering to some rules. A collider, which is a variable caused by two variables, can be separated from a root, which is causing two variables, by noting that the variables leading into the collider or away from the root, are conditionally correlated if we have found a collider, i.e., the variables become correlated when the collider is conditioned on. One can check this using linear regression, yet the same effect is also visible using bagging.

Note, however, that in combination with measurement errors and possible omitted variables, as in Figure 4 and Figure 5, establishing causes and directions quickly becomes difficult (and data may be generated in more complex ways than those illustrated as an example in Figure 1). The results in Paper 3 are most likely also affected by measurement errors and omitted variables. Hence these results should mainly be understood as guidance for further studies and not as found causal effects. However, the results in this section seem to suggest that if a causal process is sufficiently indicated by the data, bagging can find this process in the data. It is also possible to construct examples where a spuriously correlated variable (i.e., not causing) receives high importance in the Random Forest and bagging framework. For example, this can be done by including a variable that is caused by several of the variables that are also causing the outcome. The appliance variables in Paper 3 might be such variables. The conditional importance (Strobl et al., 2008) tools specifically treat these situations. This functionality has been shown to work well in simulated data with known structures and no measurement error.

Finally, although results like those reported in Paper 3 should not be interpreted in a causal way, the results in this section show the value of studying examples where a model (or different models) claims to capture causal explanation. In these circumstances, checking that the results from bagging do not contradict this model can provide useful insights. Results suggesting a different model than that assumed in the causal modelling could indicate omitted variables, different functional forms (including interactions), and measurement errors.

5 Main findings and concluding remarks

Paper 1 offers two kinds of contributions. First, it offers new interpretations of previous studies of rural fuel choices. Second, it clarifies the urban vs. rural division by including the inverse of household density and distance to nearest town as predictive variables in regressions. Because these variables show great explanatory power, fuel use may be better predicted with few variables, enabling improved mapping of the fuel usage patterns of different communes.

In Paper 3 it is found that households that have sustained a high level of income over a longer period of time are more likely to use or start to use LPG. There is also an interaction effect with rurality, as described by VAL and distance to nearest town, in which LPG usage decline with increased rurality. However the wealthier households are less sensitive to the degree of rurality. Moreover, it is found that households are more efficiently classified into groups of LPG users and non-LPG users without the inclusion of many previously used variables when the measurements of rurality, history of income and certain appliances are available for the algorithm, possibly suggesting alternative interpretations.

One interesting outcome from Papers 1 and 3 is the prediction of policy effects based on household density and distances from nearby towns, instead of more detailed household surveys. This predictability could be beneficial in certain circumstances, suggesting the use of overall area characteristics to calculate the benefits ICS dissemination might bring to various areas.

Paper 2 contributes a new method for the calculation of benefits from improved cook stoves in communes based on existing fuel use patterns. This method evaluated possible variations in benefits of ICSs in different areas. One outcome of Paper 2 is the demonstration of potentially meaningful differences in benefit estimation based on whether fuel patterns, before and after ICS adoption, is taken into account. However, Paper 2 is neither a comprehensive cost-benefit analysis nor a basis for policy on its own.

Together, the papers hint at the possibility of combining statistical models (Papers 1 and 3) with more technical modeling (Paper 2) in order to achieve useful predictions of the benefits from stove dissemination programs. Using models to predict rural fuel use in combination with other types of modeling may provide insights into the potential for these technologies. However, as already mentioned, ideally these papers should be complemented with actual experiments and in-field measurements, and one should be reluctant to base policies directly on the descriptive models presented in this thesis.

Although no hard comparisons between modelling approaches has been presented in this work, it is argued here that complementing econometric models with other types of statistical models can be beneficial for interpretation and further research. An example here is that many variables that previously have been argued to cause fuel switching also increase with rurality and are thus likely to partly describe this transition besides their own possible effect on fuel choice if not a complete description of the underlying mechanisms are also included. Another example is the connection between many of these variables and a stable economy. Important to note here is that the models presented in this thesis can never prove that the findings from previous studies are wrong but merely suggest alternative explanations.

The highly unequal income levels between the developed and the developing worlds combined with the projected costs for climate change mitigation are such that it would be possible to pay households enough money for them to be able to purchase improved stoves and for using them efficiently. In Paper 4, a model where saved fuel wood due to ICS adoption is used by households as currency in order to pay for improved stoves has been presented. The proposed model would, under certain assumptions, also enable funding based on carbon credits, even if black carbon is not included in any future climate agreement and no net CO_2-emissions from the combustion of biomass is assumed.

6 Suggestions for future research

The papers presented in this thesis should all be interpreted as possibilities for further research rather than as policy indicative in themselves. Paper 1 needs to be complemented with research and experiments for the different type of areas as described by the model. Paper 1 can also be used in the design of such experiments.

Paper 2 reveals great uncertainty in benefits-calculations based on how households may alter their fuel use after obtaining an improved cook stove. A natural continuation would therefore be to follow a stove program in an area where varied fuel use is common. Extending randomized experimental approach to both stove types and different area conditions would clarify what level of success that can be anticipated and compared to the results in Paper 2. Paper 1 can be used as a basis for describing areas with different conditions in an experimental design and thus possibly, after experiments, different anticipated outcomes.

The results in Paper 3 might be used as a basis for further studies of rural fuel choices, or for further development of statistical models. Paper 3 is exploratory in nature and the factors identified as important in classification of households requires further research before any causal links are proven. However, it may still be fruitful to examine how to inform an effective policy building on the possibly identified link between household's that are using LPG and their history of income, for example through an increased use of payment plans for the purpose of increased LPG adoption. Furthermore, the association between certain appliances and LPG use might possibly point towards links that goes beyond the explanation that these appliances are only signifying wealth, and could be studied further.

Paper 4 should be interpreted as a conceptual piece. If its ideas gain traction after scrutiny and deliberation, they can motivate further studies of the organizational structure, energy system modeling and finally experimental pilot studies in careful chosen sites with the cooperation and agreement of the anticipated participatory households. Further studies of how to incorporate such a model under a future CDM-like policy would also be needed.

Acknowledgements

First of all, I would like to thank my supervisor, Erik Ahlgren, for all the help you have given me.

I would also like to thank the Vietnam Women’s Union, the Institute of Energy in Hanoi and Mr. Nguyen M. Bao for cooperation regarding the survey used as a data source for two of the papers in this thesis.

I am also grateful to all my fantastic colleagues and especially the EST group for so many fun and insightful discussions. Especially big thanks to the colleagues who have given me tremendous moral support: Érika Mata, Eoin Ó Broin, Patrick Moldenhaur, Olivia Sanchez, Lina Reichenberg, Johan Rootzén, Mikael Odenberger, German Maldonado and Viktor Andersson.

I also would like to thank Marie, Inger and Katarina for all the help with everything.

I have had interesting discussions, which have been invaluable for this work, with Rebecka Jörnsten, Gunnar Köhlin, Daniel Johansson, Madelene Ostwald, Vera Liskovskaja, Olle Nerman, and Marc Jaccard. Thank you for all your help!

Invaluable help has also been given from my great friend Stefan Öberg who has not only provided essential discussion and help throughout this work, but also was vital in the finalization of this work.

Last but not least, I thank my mother, brother, family and friends for everything! I love you all!

This study was made possible through a SIDA-SAREC research grant and through funding contributions from the European Commission to the POEM (Policy options to engage Asian emerging economies in a post-Kyoto regime) project, grant 226282, and from the Foundation in memory of J Gust Richert.

Gothenburg, May 2015

References

Allison, I., Bindoff, N., Bindschadler, R., Cox, P., De Noblet, N., England, M., Francis, J., Gruber, N., Haywood, A., Karoly, D., 2009. The Copenhagen Diagnosis, 2009: updating the world on the latest climate science. The University of New South Wales Climate Change Research Centre (CCRC), Sydney, Australia.

Amacher, G.S., Hyde, W.F., Kanel, K.R., 1996. Household fuelwood demand and supply in Nepal's tarai and mid-hills: Choice between cash outlays and labor opportunity. World Development 24, 1725-1736.

Arnold, J., Köhlin, G., Persson, R., 2006. Woodfuels, livelihoods, and policy interventions: changing perspectives. World Development 34, 596-611.

Arze del Granado, F.J., Coady, D., Gillingham, R., 2012. The Unequal Benefits of Fuel Subsidies: A Review of Evidence for Developing Countries. World Development 40, 2234-2248.

Arzel, O., Fichefet, T., Goosse, H., 2006. Sea ice evolution over the 20th and 21st centuries as simulated by current AOGCMs. Ocean Modelling 12, 401-415.

Bailis, R., Cowan, A., Berrueta, V., Masera, O., 2009. Arresting the Killer in the Kitchen: The Promises and Pitfalls of Commercializing Improved Cookstoves. World Development 37, 1694-1705.

Bailis, R., Drigo, R., Ghilardi, A., Masera, O., 2015. The carbon footprint of traditional woodfuels. Nature Climate Change 5, 266-272.

Barnes, D.F., Krutilla, K., Hyde, W., 2004. The Urban Household Energy Transition. Energy, poverty, and the environment in the developing world. World Bank. Washington DC https://www.esmap.org/sites/esmap.org/files/Rpt_UrbanEnergyTransition.pdf.

Barnes, D.F., Openshaw, K., Smith, K.R., Plas, R.v.d., 1994. What Makes People Cook with Improved Biomass Stoves? A Comparative International Review of Stove Programs. World Bank Technical Paper No. 242 http://www-wds.worldbank.org/external/default/WDSContentServer/IW3P/IB/1999/08/15/000009265_3970311122727/Rendered/PDF/multi_page.pdf.

Beltramo, T., Levine, D.I., Blalock, G., 2014. The Effect of Marketing Messages, Liquidity Constraints, and Household Bargaining on Willingness to Pay for a Nontraditional Cookstove. CEGA Working Papers https://escholarship.org/uc/item/4cq705v3.

Bensch, G., Peters, J., 2012. A Recipe for Success? Randomized Free Distribution of Improved Cooking Stoves in Senegal (March 1, 2012) SSRN: http://ssrn.com/abstract=2030746.

Berk, R., Sherman, L., Barnes, G., Kurtz, E., Ahlman, L., 2009. Forecasting murder within a population of probationers and parolees: a high stakes application of statistical learning. Journal of the Royal Statistical Society: Series A (Statistics in Society) 172, 191-211.

Bice K, e.a., 2008. Black Carbon: A Review and Policy Recommendations. (Woodrow Wilson School of Public and International Affairs, Princeton University, Princeton, NJ) http://tinyurl.com/wws2008-bc.

Bodansky, D., 2001. Bonn voyage: Kyoto's uncertain revival. The National Interest 65, 45-55.

Bond, T., Doherty, S., Fahey, D., Forster, P., Berntsen, T., DeAngelo, B., Flanner, M., Ghan, S., Kärcher, B., Koch, D., 2013. Bounding the role of black carbon in the climate system: A scientific assessment. Journal of Geophysical Research: Atmospheres 118(11), 5380-5552.

Bond, T.C., Sun, H., 2005. Can Reducing Black Carbon Emissions Counteract Global Warming? Environ. Sci. Technol. 2005, 39, 5921-5926.

Bond, T.C., Zarzycki, C., Flanner, M.G., Koch, D.M., 2011. Quantifying immediate radiative forcing by black carbon and organic matter with the Specific Forcing Pulse. Atmospheric Chemistry and Physics 11, 1505-1525.

Breiman, L., 1996. Bagging predictors. Machine learning 24, 123-140.

Breiman, L., 2001. Random forests. Machine learning 45, 5-32.

Breiman, L., Cutler, A., 2013. Random Forests. https://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm Accessed: 20130805.

Breiman, L., Friedman, J., Olshen, R., Stone, C., 1984. Classification and regression trees (CART) Wadsworth International Group. Belmont, CA, USA.

Briscoe, J., 1979. Energy use and social structure in a Bangladesh village. Population and Development Review, 615-641.

Bruce, N., Perez-Padilla, R., Albalak, R., 2000. Indoor air pollution in developing countries: a major environmental and public health challenge. Bulletin of the World Health Organization 78, 1078-1092.

Campbell, B.M., Vermeulen, S.J., Mangono, J.J., Mabugu, R., 2003. The energy transition in action: urban domestic fuel choices in a changing Zimbabwe. Energy Policy 31, 553-562.

Cooke, P., Köhlin, G., Hyde, W.F., 2008. Fuelwood, forests and community management-Evidence from household studies. Environment and Development Economics 13, 103-135.

Cooke, P.A., 1998. Intrahousehold labor allocation responses to environmental good scarcity: a case study from the hills of Nepal. Economic Development and Cultural Change 46, 807-830.

Cutler, D.R., Edwards Jr, T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., Lawler, J.J., 2007. Random forests for classification in ecology. Ecology 88, 2783-2792.

Dermish, A., Kneiding, C., Leishman, P., Mas, I., 2011. Branchless and mobile banking solutions for the poor: a survey of the literature. innovations 6, 81-98.

Eckholm, E., 1975. Other energy crisis: firewood. Worldwatch paper series.

Economist, 2013. Zapping mosquitoes, and corruption. The Economist Newspaper Limited Q2 2013.

Edwards, J.H.Y., Langpap, C., 2005. Startup costs and the decision to switch from firewood to gas fuel. Land Economics 81, 570-586.

Ezzati, M., Kammen, D.M., 2001. Indoor air pollution from biomass combustion and acute respiratory infections in Kenya: an exposure-response study. The Lancet 358, 619-624.

Foell, W., Pachauri, S., Spreng, D., Zerriffi, H., 2011. Household cooking fuels and technologies in developing economies. Energy Policy 39, 7487-7496.

Gifford, M.L., 2010. A Global Review of Cookstove Programs. MS. Thesis Energy and Resources Group UC Berkeley, CA http://www.eecs.berkeley.edu/~sburden/misc/mlgifford_ms_thesis.pdf.

Grieshop, A.P., Marshall, J.D., Kandlikar, M., 2011. Health and climate benefits of cookstove replacement options. Energy Policy 39, 7530-7542.

Grubb, M., Vrolijk, C., Brack, D., 1999. The Kyoto Protocol: A Guide and Assessment. Earthscan.

Grömping, U., 2009. Variable Importance Assessment in Regression: Linear Regression versus Random Forest. The American Statistician 63, 308-319.

Gundimeda, H., Köhlin, G., 2008. Fuel demand elasticities for energy and environmental policies: Indian sample survey evidence. Energy Economics 30, 517-546.

Hanna, R., Duflo, E., Greenstone, M., 2012. Up in Smoke: The Influence of Household Behavior on the Long-Run Impact of Improved Cooking Stoves. NBER Working Paper No. 18033 http://www.nber.org/papers/w18033.

Hansen, J., Sato, M., Kharecha, P., Russell, G., Lea, D.W., Siddall, M., 2007. Climate change and trace gases. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 365, 1925-1954.

Haywood, J.M., Ramaswamy, V., 1998. Global sensitivity studies of the direct radiative forcing due to anthropogenic sulfate and black carbon aerosols. J. Geophys. Res. 103, 6043-6058.

Heltberg, R., 2004. Fuel switching: evidence from eight developing countries. Energy Economics 26, 869-887.

Heltberg, R., 2005. Factors determining household fuel choice in Guatemala. Environment and Development Economics 10, 337-361.

Heltberg, R., Arndt, T.C., Sekhar, N.U., 2000. Fuelwood Consumption and Forest Degradation: A Household Model for Domestic Energy Substitution in Rural India. Land Economics 76, 213-232.

Hosier, R.H., Dowd, J., 1987a. Household fuel choice in Zimbabwe : An empirical test of the energy ladder hypothesis. Resources and Energy 9, 347-361.

Hosier, R.H., Dowd, J., 1987b. Household fuel choice in Zimbabwe: An empirical test of the energy ladder hypothesis. Resources and Energy 9, 347-361.

Hothorn, T., Hornik, K., Zeileis, A., 2006a. Unbiased recursive partitioning: A conditional inference framework. Journal of Computational and Graphical Statistics 15.

Hothorn, T., Hornik, K., Zeileis, A., 2006b. Unbiased recursive partitioning: A conditional inference framework. Journal of Computational and Graphical Statistics 15, 651-674.

IEA, 2010. International Energy Agency, 2010. World Energy Outlook 2010, Paris.

IEA, 2012. Statistics & Balances (2009).

IPCC, 2007. Climate Change 2007: Synthesis Report. Contribution of Working Groups I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. IPCC.

Jacobson, M.Z., 2001. Strong radiative heating due to the mixing state of black carbon in atmospheric aerosols. Nature 409, 695-697.

Jetter, J.J., Kariher, P., 2009. Solid-fuel household cook stoves: Characterization of performance and emissions. Biomass and Bioenergy 33, 294-305.

Kaul, S., Liu, Q., 1992. Rural household energy use in China. Energy 17, 405-411.

Khandker, S., Barnes, D., Samad, H.A., Minh, N.H., 2009. Welfare impacts of rural electrification: evidence from Vietnam. World Bank Policy Research Working Paper No. 5057 http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1476699.

Kopp, R.E., Mauzerall, D.L., 2010. Assessing the climatic benefits of black carbon mitigation. Proceedings of the National Academy of Sciences 107, 11703-11708.

Kshirsagar, M.P., Kalamkar, V.R., 2014. A comprehensive review on biomass cookstoves and a systematic approach for modern cookstove design. Renewable and Sustainable Energy Reviews 30, 580-603.

Kumar, S.K., Hotchkiss, D., 1988. Consequences of deforestation for women's time allocation, agricultural production, and nutrition in hill areas of Nepal. Research Report-International Food Policy Research Institute.

Köhlin, G., Sills, E.O., Pattanayak, S.K., Wilfong, C., 2011. Energy, gender and development: what are the linkages? where is the evidence? World Bank Policy Research Working Paper No. 5800 http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1931364.

Leach, G., 1992. The energy transition. Energy Policy 20, 116-123.

Lee, C.M., Chandler, C., Lazarus, M., Johnson, F.X., 2013. Assessing the climate impacts of cookstove projects: issues in emissions accounting. Challenges in Sustainability 1, 53-71.

Lewis, J.J., Pattanayak, S.K., 2012. Who adopts improved fuels and cookstoves? A systematic review. Environmental health perspectives 120, 637.

Marron, T., Riojas-Rodriguez, H., Schilmann, A., Romieu, I., Masera, O., 2008. Factors that modify the adoption of an improved chimney stove (Patsari) reducing indoor air pollution in rural Mexico. Epidemiology 19, S262.

Martin, W.J., Glass, R.I., Balbus, J.M., Collins, F.S., 2011. A major environmental cause of death. Science 334, 180-181.

Masera, O.R., Saatkamp, B.D., Kammen, D.M., 2000. From Linear Fuel Switching to Multiple Cooking Strategies: A Critique and Alternative to the Energy Ladder Model. World Development 28, 2083-2103.

Matthews, H.D., Caldeira, K., 2008. Stabilizing climate requires near-zero emissions. Geophysical research letters 35, L04705.

Miller, G., Mobarak, A.M., 2011. Intra-Household Externalities and Low Demand for a New Technology: Experimental Evidence on Improved Cookstoves. Unpublished manuscript http://casi.sas.upenn.edu/sites/casi.sas.upenn.edu/files/iit/Miller%20and%20Mobarak.pdf.

Mobarak, A.M., Dwivedi, P., Bailis, R., Hildemann, L., Miller, G., 2012. Low demand for nontraditional cookstove technologies. Proceedings of the National Academy of Sciences 109, 10815-10820.

Myhrvold, N.P., Caldeira, K., 2012. Greenhouse gases, climate change and the transition from coal to low-carbon electricity. Environmental Research Letters 7, 014019.

Nordhaus, W.D., 2007. A Review of the" Stern Review on the Economics of Climate Change". Journal of Economic Literature, 686-702.

Paulsson, E., 2009. A review of the CDM literature: from fine-tuning to critical scrutiny? International Environmental Agreements: Politics, Law and Economics 9, 63-80.

Pearl, J., 2009. Causal inference in statistics: An overview. Statistics Surveys 3, 96-146.

Pearson, B., 2007. Market failure: why the Clean Development Mechanism won't promote clean development. Journal of Cleaner Production 15, 247-252.

Pedersen, C.A., Roeckner, E., Lüthje, M., Winther, J.G., 2009. A new sea ice albedo scheme including melt ponds for ECHAM5 general circulation model. Journal of Geophysical Research 114, D08101.

Peng, W., Hisham, Z., Pan, J., 2010. Household level fuel switching in rural Hubei. Energy for Sustainable Development 14, 238-244.

Pillarisetti, A., Vaswani, M., Jack, D., Balakrishnan, K., Bates, M.N., Arora, N.K., Smith, K.R., 2014. Patterns of stove usage after introduction of an advanced cookstove: the long-term application of household sensors. Environmental science & technology 48, 14525-14533.

Pope, C.A., Burnett, R.T., Krewski, D., Jerrett, M., Shi, Y., Calle, E.E., Thun, M.J., 2009. Cardiovascular Mortality and Exposure to Airborne Fine Particulate Matter and Cigarette Smoke. Circulation 120, 941-948.

Ramanathan, V., Carmichael, G., 2008. Global and regional climate changes due to black carbon. Nature Geosci 1, 221-227.

Reddy, A.K.N., 1982. Rural energy consumption patterns--A field study. Biomass 2, 255-280.

Rehfuess, E., Sumi, M., Prüss-Üstün, A., 2006. Assessing Household Solid Fuel Use: Multiple Implications for the Millennium Development Goals. Environ Health Perspect 114, 373-378.

Roden, C.A., Bond, T.C., Conway, S., Osorto Pinel, A.B., MacCarty, N., Still, D., 2009. Laboratory and field investigations of particulate and carbon monoxide emissions from traditional and improved cookstoves. Atmospheric Environment 43, 1170-1181.

Ruiz-Mercado, I., Masera, O., 2015. Patterns of Stove Use in the Context of Fuel–Device Stacking: Rationale and Implications. EcoHealth, 1-15.

Ruiz-Mercado, I., Masera, O., Zamora, H., Smith, K.R., 2011. Adoption and sustained use of improved cookstoves. Energy Policy 39, 7557-7566.

Shankar, A., Johnson, M., Kay, E., Pannu, R., Beltramo, T., Derby, E., Harrell, S., Davis, C., Petach, H., 2014. Maximizing the benefits of improved cookstoves: moving from acquisition to correct and consistent use. Global Health: Science and Practice 2, 268-274.

Shine, K.P., Berntsen, T.K., Fuglestvedt, J.S., Skeie, R.B., Stuber, N., 2007. Comparing the climate effect of emissions of short- and long-lived climate agents. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 365, 1903-1914.

Shrimali, G., Slaski, X., Thurber, M.C., Zerriffi, H., 2011. Improved stoves in India: A study of sustainable business models. Energy Policy 39, 7543-7556.

Simon, G.L., Bumpus, A.G., Mann, P., 2012. Win-win scenarios at the climate–development interface: Challenges and opportunities for stove replacement programs through carbon finance. Global Environmental Change 22, 275-287.

Sinton, J.E., Smith, K.R., Peabody, J.W., Yaping, L., Xiliang, Z., Edwards, R., Quan, G., 2004. An assessment of programs to promote improved household stoves in China. Energy for Sustainable Development 8, 33-52.

Smith, K.R., Haigler, E., 2008. Co-benefits of climate mitigation and health protection in energy systems: scoping methods. Annu. Rev. Public Health 29, 11-25.

Smith, K.R., Haigler, Evan, 2008. Co-Benefits of Climate Mitigation and Health Protection in Energy Systems: Scoping Methods. Annu. Rev. Public Health 2008. 29:11–25.

Smith, K.R., Peel, J.L., 2010. Mind the Gap. Environmental Health Perspectives 118, 1643-1645.

Smith, K.R., Shuhua, G., Kun, H., Daxiong, Q., 1993. One hundred million improved cookstoves in China: How was it done? World Development 21, 941-961.

Sterk, W., Wittneben, B., 2006. Enhancing the clean development mechanism through sectoral approaches: definitions, applications and ways forward. International Environmental Agreements: Politics, Law and Economics 6, 271-287.

Sterner, T., Persson, U.M., 2008. An even sterner review: Introducing relative prices into the discounting debate. Review of Environmental Economics and Policy 2, 61-76.

Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., Zeileis, A., 2008. Conditional variable importance for random forests. BMC bioinformatics 9, 307.

Strobl, C., Hothorn, T., Zeileis, A., 2009a. Party on! The R Journal 1/2, 14-17.

Strobl, C., Malley, J., Tutz, G., 2009b. An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. Psychological methods 14, 323.

Takama, T., Tsephel, S., Johnson, F.X., 2012. Evaluating the relative strength of product-specific factors in fuel switching and stove choice decisions in Ethiopia. A discrete choice model of household preferences for clean cooking alternatives. Energy Economics 34, 1763-1773.

Torres-Duque, C., Maldonado, D., Perez-Padilla, R., Ezzati, M., Viegi, G., on behalf of the Forum of International Respiratory Societies Task Force on Health Effects of Biomass Exposure, 2008. Biomass Fuels and Respiratory Diseases: A Review of the Evidence. Proc Am Thorac Soc 5, 577-590.

Touw, W.G., Bayjanov, J.R., Overmars, L., Backus, L., Boekhorst, J., Wels, M., van Hijum, S.A., 2013. Data mining in the Life Sciences with Random Forest: a walk in the park or lost in the jungle? Briefings in bioinformatics 14, 315-326.

Tuan, N.A., Lefevre, T., 1996. Analysis of household energy demand in Vietnam. Energy Policy 24, 1089-1099.

UNFCCC, 1997. Kyoto protocol to the United Nations framework convention on climate change. United Nations Framework Convention on Climate Change. FCCC/CP/L.7/Add1, Kyoto.

UNFCCC, 2001. The Marrakesh Accords and The Marrakesh declaration. http://www.unfcccc.int.

Van't Veld, K., Narain, U., Gupta, S., Chopra, N., Singh, S., 2006. India's firewood crisis re-examined. RFF Discussion Paper No. 06-25 http://papers.ssrn.com/sol3/papers.cfm?abstract_id=901625.

van der Kroon, B., Brouwer, R., van Beukering, P.J., 2014. The impact of the household decision environment on fuel choice behavior. Energy Economics 44, 236-247.

van der Kroon, B., Brouwer, R., van Beukering, P.J.H., 2013. The energy ladder: Theoretical myth or empirical truth? Results from a meta-analysis. Renewable and Sustainable Energy Reviews 20, 504-513.

Wara, M., 2008. Measuring the Clean Development Mechanism’s Performance and Potential. 55 UCLA L. REV., 1759, 1770.

Wen, W., En-Jian, C., 1983. Our views on the resolution of China's rural energy requirements. Biomass 3, 287-312.

WHO, 2015. Household (Indoor) Air Pollution. http://www.who.int/indoorair/en/ Accessed: 2015-03-15.

Yin, R.K., 2009. Case study research: Design and methods. Sage.