PBL Netherlands Environmental
Assessment Agency

PLANETARY SECURITY: IN SEARCH OF DRIVERS OF VIOLENCE AND CONFLICT

Part II: Inferences through Machine Learning

Background Report

Hans Visser, Sophie de Bruin, Joost Knoop, Paul Vethman and Willem Ligtvoet

01 December 2019

PBL

Planetary security: in search of drivers of violence and conflict.
Part II: Inferences through Machine Learning

The Hague, 2019
PBL publication number: 3405

Corresponding author
Hans.Visser@pbl.nl

Authors
Hans Visser, Sophie de Bruin, Paul Vethman, Joost Knoop, Willem Ligtvoet

Ultimate responsibility
PBL Netherlands Environmental Assessment Agency

Production coordination
PBL Publishers

PBL Netherlands Environmental Assessment Agency is the national institute for strategic policy analysis in the fields of the environment, nature and spatial planning. We contribute to improving the quality of political and administrative decision-making by conducting outlook studies, analyses and evaluations in which an integrated approach is considered paramount. Policy relevance is the prime concern in all of our studies. We conduct solicited and unsolicited research that is both independent and scientifically sound.

Contents

Summary and findings

What are the major origins and drivers of different types of conflict? Sorting out the main causes of conflict and war is difficult and often shaped by ideological believes. Even today, historians and political scientists have discussions on the primary causes of the First World War. There are several types of conflict, ranging from international and civil wars to local conflicts, riots and revolution. And there are many theories that explain these different types of conflict, which mostly focus on economic conditions and a range of factors that can foster grievances and greed, creating incentives to initiate or join a conflict.

The first report in this study was directed to the available data and databases on national scales, with special attention to the reliability of these databases. In the present report, we focus on **identifying conditions** that are associated with high risk levels of conflict and violence, on national scales, and we assess to what extend socio-economic and environmental indicators — with a special focus on water — play a role. Our assessment starts with identifying conditions affecting conflict risk that are deliberated in academic and popular literature, yielding a theoretical framework for the statistical analyses.

The study design connects to the development cooperation policy of the Dutch central government. This policy seeks to achieve the following goals in developing countries (BuZa, 2018):

- *Prevent conflicts and instability.*
- *Reduce poverty and social inequality.*
- *Promote sustainable growth and climate action worldwide.*

Statistical approach

Our analysis of international databases is aimed at finding a hierarchy of conflict conditions, also denoted as a hierarchy of conflict drivers. The rationale of this approach is illustrated in **Figure A**. In traditional statistical analyses it is assumed, or at least suggested, that a dependent variable, here a risk indicator for conflict and violence, can be related to drivers which cause spatial or temporal patterns in conflict (the left panel in Figure A). However, this approach is not realistic since all these indicators are mutually correlated, yielding arrows with heads at both sides and an extra arrow between driver 1 and 2 (middle panel). In fact, conflict can also be seen as a driver in this scheme and one-directional **causality** becomes less obvious.

The approach chosen here follows a variation on the middle panel where we identify a hierarchy in drivers, based on the strength of associations (explanatory power) of various drivers and expert judgment (right panel). We avoid the term causality. We further note that the processes leading to an outbreak of conflict are complex, multifaceted and highly context specific. This makes **predicting** conflict hard and controversial. Therefore, we do not try to predict such outbreaks in this study. We rather try to develop a diagnostic system based on the conditions identified, given sufficient strengths of the models we will find.

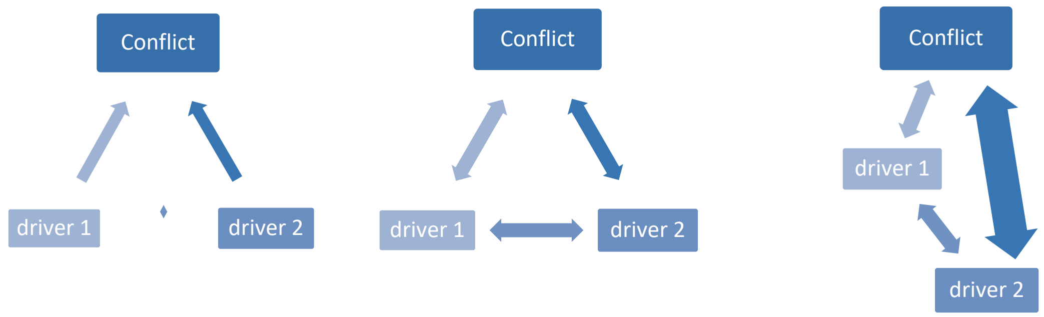


Figure A Three visualisations for the relationship between a conflict-related indicator and two drivers. The left panel shows one-directional linkages which are often interpreted as being causal. The middle panel presents the more realistic situation: all three indicators are interrelated. The right panel shows the situation followed in this study: based on explanatory power and expert judgement drivers can be re-arranged in a hierarchical manner, some indicators are interpreted as more structural whereas other indicators, affected by the structural drivers, are more proximate. We avoid the term 'causal'; drivers are seen as conditions that relate to conflict. The thickness of arrows corresponds to the explanatory power of specific drivers.

To find a hierarchy in conflict drivers, we selected a Machine Learning technique which deals with both linear and non-linear relationships and which can cope with potential drivers which are mutually correlated (the statistical problem of multicollinearity, the arrow with double heads between drivers 1 and 2 in figure A). The method is called Random Forest and is based on a recurrent estimation of so-called Regression Trees. The method characterizes non-linear relationships by identifying driver thresholds rather than assuming linear relationships
a-priori.

Random Forest models consist of an ensemble of Regression-Tree models where the number of models will lie between 100 and 1000 in practice. Each of these models is estimated on 2/3 of the available country data and predictions are made for the remaining 1/3. The procedure is illustrated in **Figure B**. As a final step the Random Forest method combines the results from the individual regression trees to find a hierarchy in drivers, the so-called importance function. This importance function can be interpreted as the levels of association between individual drivers and the conflict indicator.

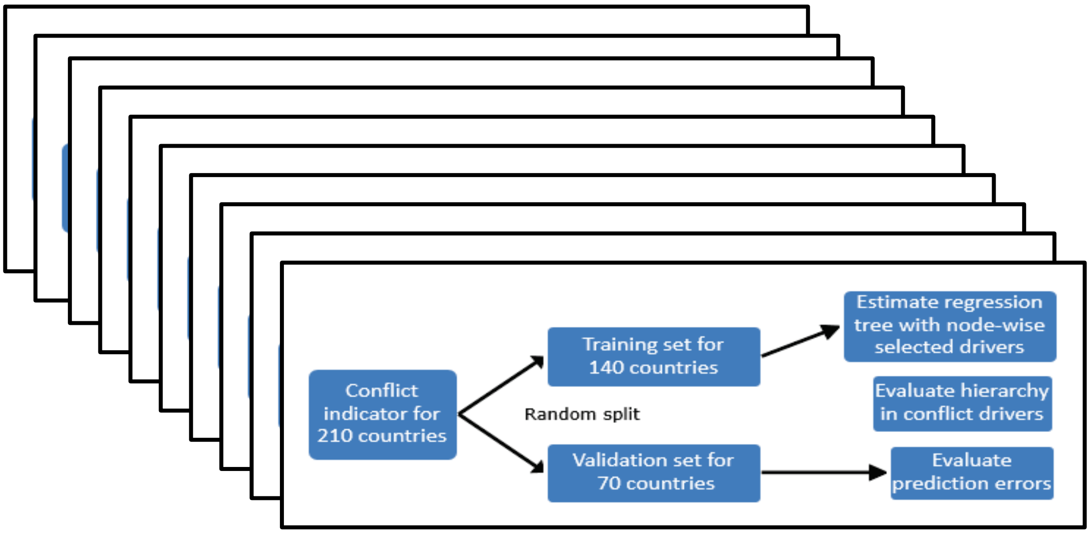


Figure B Scheme showing the estimation procedure of a Random Forest. This Machine Learning approach consists of the estimation of an ensemble of, say, 1000 regression trees which form 'the forest'. Each ensemble member has the same structure as shown in the individual rectangles. However, ensemble members differ as for one aspect: the original data set (here conflict data for 210 countries) is *randomly* split into a training set for which a regression tree is estimated (140 countries), and a validation set (70 countries). The regression tree shows which drivers have the strongest relationship to conflict. Next to that, the tree is used to predict the conflict data for independent country data (the validation set). As a final step the importance of potential drivers is derived from the total set of 1000 regression trees.

Data

Since there is no unique (composite) indicator which fully covers the field of conflict and violence, we introduce three such indicators which highlight different aspects of conflict risks; we perform our modelling approach for each of these indicators separately.

The first indicator is the Global Peace Index which is composed from 23 underlying peace-related indicators (such as the number of violent demonstrations, weapons imports, nuclear and heavy weapons capabilities and military expenditure). The second indicator is based on death counting related to violence (homicide rates) and conflicts (state-based conflicts, non-state conflicts and one-sided violence against civilians). The third indicator concerns the number of uprooted people and is composed from migration flows forced by conflicts (taken as the sum of internally displaced people and refugees). All three indicators have a continuous scale, rather than stating that a country is in a state of war or not, which would introduce a binary indicator having values '0' or '1'.

A potential set of drivers — denoted as 'regressors' in statistical terms or 'predictors' in Machine Learning terms — is derived from studying the leading conflict literature. We propose a theoretical framework consisting of eight driver groups: (1) economic inequality and poverty, (2) grievances and discrimination, (3) governance and corruption, (4) demographics and education, (5) availability of resources such as ores, oil and fertile land, (6) conflicts in neighbouring countries, (7) infant mortality and malnutrition, and

(8) water and climate-related indicators. The last group covers indicators such as the availability of clean water, improved sanitation, and the humanitarian impacts of floods and droughts.

Results
After a pre-selection of indicators we selected 16 potential drivers for conflict and violence, listed in **Figure C**. The graph shows the main results from this study. We found that Governance, GDP per capita and Inequality are the dominating drivers/conditions, although differences across the three conflict indicators — shown in the legend — are considerable.

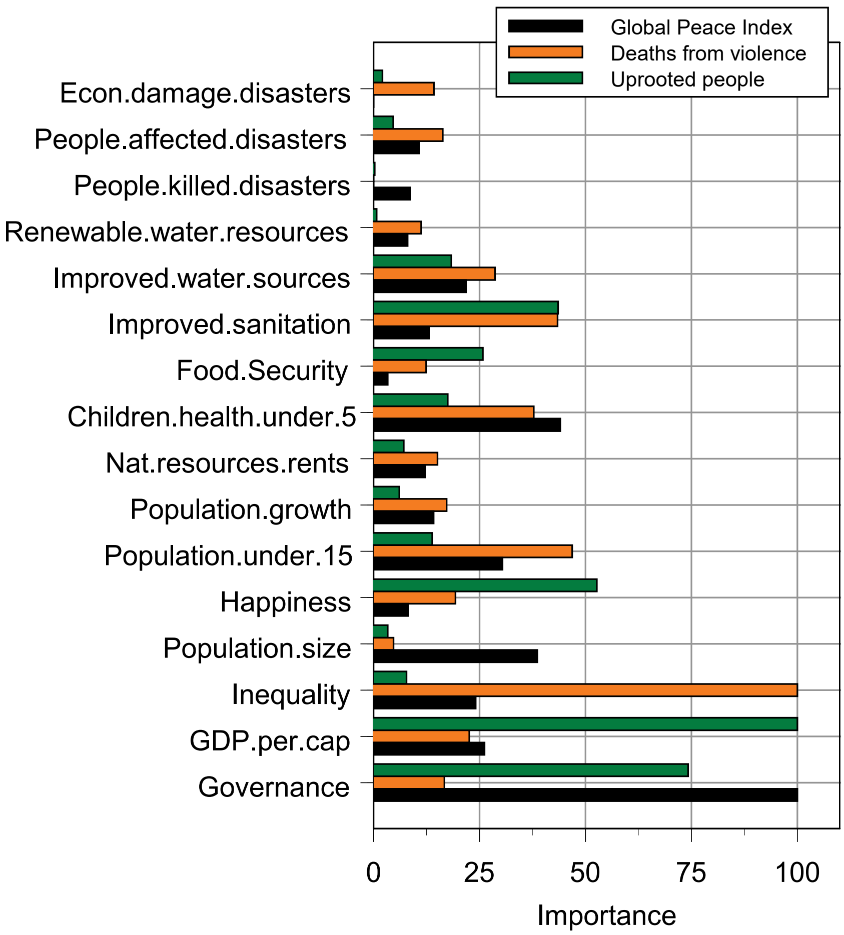


Figure C Importance functions for three conflict indicators: Global Peace Index (bars in black), Deaths from violence and conflict (bars in orange), and People uprooted by conflicts (bars in green). The latter group combines internally displaced people as well as refugees which leave the country. Importance functions, based on 16 selected drivers, are estimated using the Random Forest approach.

At the other end of the scale we found several factors that hardly affect conflict indicators. These mainly include environmental and resources-related factors, such as economic damage and people affected by water-related disasters (floods, droughts, tsunamis) and natural resources rents. The latter indicator stands for the total natural resources rents as percentage of each country's GDP (where 'total' stands for the sum of rents from oil, natural gas, coal, minerals and forest). Therefore, these statistical analyses do not support the general idea that environmental indicators and water-related indicators in particular affect conflict risk on a national scale.

We argue that governance and corruption, GDP per capita and inequality can be considered as more structural (ultimate) conditions, affecting conflict directly but also indirectly via other indicators. These structural factors show the highest correspondence to conflict and are regarded as fundamental drivers in the literature (e.g., Acemoglu, 2012; Mach et al.,2019), and in five out of the six conflict studies analyzed for this case study. Other factors shown in Figure C – such as access to improved sanitation, youth bulge and health of children under the age of 5 - are considered to be proximate, more direct factors.

Conclusions

This study adds to the extensive literature on conflicts and violence and yields the following more general conclusions.

First, it matters which (composite) indicator one chooses for 'conflict and violence' in a statistical analysis. It appears that findings such as presented here, highly depend on the indicator chosen. Many studies only choose only one indicator such as deaths from state-based and non-state-based violence, and one-sided violence against civilians which than is coded as a binary variable ('conflict' or 'no conflict'). The indicators chosen here, have a continuous scale and reflect risk levels of conflict and violence seen from different angles.

Second, although (statistical) conflict analysis can be done by including a wide range of drivers, governance and socio-economic development (modelled as 'GDP per capita' and 'Inequality') are the dominating factors, both for conflict and violence in itself and for the influence of proximate variables. This conclusion is consistent with findings in the conflict literature, notably the recent study of Mach et al. (2019) who assessed the linkages between conflict, climate and other drivers by structured judgments of 11 experts in the field.

Third, environmental and water-related indicators do not relate very well to conflict and violence, at least if based on national-scale analyses. This conclusion is consistent with the majority of findings in the conflict literature (e.g., Buhaug, 2015; Gleick and Iceland, 2018; Schmeier et al., 2019; Mach et al. 2019) and Table 2.1 in this report. However, some researchers do find such relationships and thus show contrasting conclusions (Hsiang et al., 2013; Abel et al., 2019).

There are a number of explanations for these seemingly opposing results. We name three of them here. Environmental conditions might play important roles on *local* scales which level off on *national* scales (e.g., De Bruin et al., 2018). Next to that, water-related disasters and climate extremes do not 'automatically' lead to more grievances, and thus to higher levels of conflict and violence. It can bind people too, leading to cooperative management (e.g., Ostrom et al., 1999; Schmeier et al., 2019). Finally, the analyses given here have an global extent. However, if these analyses would be performed for certain regions, such as for African countries or countries within the EU, quite different results might emerge (stratification, in statistical terms).

In addition, it should be noted in this context that water-related disasters lead to the highest number of people either affected or killed, compared to earth quakes and violent conflict (Figure 7.1). Thus, although water-related disasters are not a major driver for conflict — at least according to national census data and surveys — the social and economic disruption they cause, is enormous.

Fourth, data quality is a reason of concern for any quantitative study on the relationship between conflict and potential drivers, including the analysis presented in this report (Jerven, 2013; Visser et al., 2018; Arnold, 2019; Espey, 2019). We have found that data quality is limiting statistical analyses in two ways. If information on driver X is missing for country Y, this country will be omitted from the analysis simply because statistical methods cannot cope with missing data in **any** variable. And if data are available, they may not be reliable due to a low level of statistical capacity or definitional uncertainties. Both situations will be especially true for the least developed countries with comparatively low levels of governance (weak institutions).

To check the robustness of results presented in Figure C we performed a number of sensitivity analyses: (i) the application of a set of indicators with a much wider scope than shown in Figure C (cf. Appendix A); (ii) imputation of missing country data before estimating Random Forest models; (iii) re-analysis of data using Regression Trees; and (iv) the visual analysis of scatterplots along with LOESS trends (Appendix C, and Figures 5.2B, 5.3B and 5.4B). Results are consistent with those shown in Figure C (apart from minor differences).

Fifth, we found that the explanatory power of Random Forest models is moderate, namely 46% on average. Especially, countries with weak institutions (low levels of governance) show only moderate prediction accuracies. One explanation could be that data in poor countries are less reliable (the conclusion above). Another explanation could lie in the fact that we did not include all relevant indicators in our analysis. Next to that, not all events or influences that may play a role, can be translated to global data sets. These aspects are dynamic or unique, such as the end of the cold war, which included secret negotiations not possible to catch in a number. Factors not explicitly included are the role of international inferences (proxy wars) and the role of ethnicity and religion.

Sixth, we found that statistical results shown here underpin the main objectives of the Dutch government, set out in the policy document *Investing in Global Prospects* by the Minister of Foreign Affairs, Sigrid Kaag (BuZa, 2018). The document emphasises the importance of SDG 16 – peace, justice and strong institutions, corresponding to one of the main findings shown in Figure C where 'justice and strong institutions' can be seen as synonymous with 'governance'. Next to that, striving to reduce poverty (SDGs 1 and 8), and social inequality (SDGs 5 and 10) is consistent with the importance shown in Figure C (GDP per capita and Inequality; the last is a composite for poverty in combination with economic and gender inequality). The relationship between SDGs and the hierarchy shown in Figure C is visualised in **Figure D**.

Finally, it is important to note that our analysis, and that of many other authors, is based on historical data. It may be questioned to what extent the results of these analyses will be representative **for the future**, given the high population growth in vulnerable regions, the increasing impact of climate change and increasing weather extremes (e.g. IPCC, 2018 and 2019). Today many examples are found of increasing violence on local scale between farmers and cattle farmers related to increasing water stress and increasing tensions in some transboundary river basins and a further increase may be expected (e.g. Ligtvoet et al., 2018; De Bruin et al., 2018).

Figure D Seventeen Sustainable Development goals (SDGs). SDGs which arise in this report as dominant in relation to conflict and violence, are enlarged. These are SDG #16 (related to peace, governance and corruption), SDGs # 5 and 10 (related to economic and gender inequality) and SDGs #1 and 8 (GDP per capita, poverty).

As noted by Mach et al. (2019), developments in population growth, climate change, sea level rise, and the uncertainties about societal responses, governance capacities and adaptation limits add to the complexity and uncertainties in a diagnostic system. And ranges within conflicts will not be resolved autonomously, or narrow down, especially in vulnerable areas, putting extra pressure on governance capacities. For example, Gleick and Iceland (2018) show that future risks of conflict may increase significantly if the governance capacities do not improve.

Thus, as for the future, the situation sketched in Figure C is expected to change due to increasing pressures of climate change with risks on water shortages, land erosion and food insecurity, and due to limits of adaption. See stylized scheme in **Figure E**.

Outlook

In this study we have developed a broad knowledge base regarding the drivers of conflict on a global scale, based on national data. In future research, much can be done to improve these results. We name the following topics:

- It would be interesting to break down this analysis per region or continent (stratification in statistical terms). Are governance, inequality and GDP of similar importance in Asia compared to Africa? Or show high risk countries different patterns compared to “the average”?

- We did not explore the role of foreign powers in conflict, that of international interventions. The same holds for the role of ethnicity and religion, as well as the role of land degradation. These topics could be explored in future research.

- The lack of data or the presence of poor data in less-developed countries hampers statistical modelling. Therefore, it is important to keep an eye on the development of data in these countries, hopefully with improved reliability. This is of importance since these countries face higher conflict risks.

- The analysis of indicators defined on national scales only, might be a limiting factor. In the study *The Geography of Future Water Challenges*, we analysed data on water province levels where water provinces fall within the borders of individual countries. One way forward would be to gather indicator data for these spatial scales. Random Forest models could be estimated for these water-province based data.

- Linkages between conflict and 'water' — or more generally 'environmental and climate change' — are not fixed and **might change in the near future**, especially since climate change is expected to exacerbate conflict–climate connections (Figure E). Therefore, it is important to perform statistical analyses such as done here, with updated and, hopefully, improved data in the coming years.

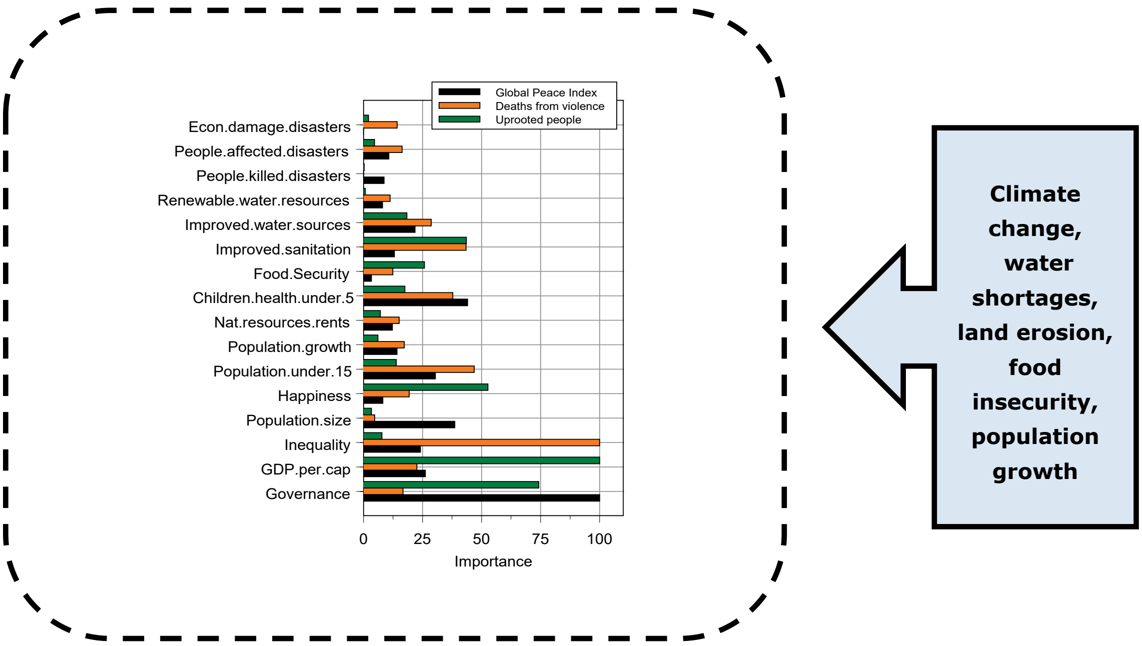


Figure E Future relationships are expected to change due to increasing pressures from climate change, population dynamics and land erosion. Detailed discussions on adaptation strategies are given by IPCC (2018, 2019).

1 Introduction

The potential links between environmental change, including climate change and changing weather patterns, migration and conflict have received growing attention from scientists, media and global institutions over the last decade. But despite the increasing role that environmental change has played in global security analyses and conventions, research on these topics has not fully matured or reached consensus on the existence of causal relationships. As for drivers of conflicts, violence and migration multiple explanations are found in the literature, varying from poverty and inequality, availability of resources (fertile land, ores, oil, water), grievances and greed, ineffective governance and corruption.

The role that water-related impacts may play in the eruption of conflict has received more attention in recent years (Gleick, 2014; Gunasekara et al., 2014), although relationships found are often to be context-specific (Niasse, 2005; Von Uexkull et al., 2016) or found to be negligible compared to other factors (Theisen et al., 2012).

To deepen scientific insights in these complex processes and to strengthen the knowledge–policy interface, PBL Netherlands Environmental Assessment Agency participates in the Planetary Security Initiative. Recent studies highlighted the possible role of water — too little, too much, too dirty — and the links between water security and conflicts. See the following reports: 'The geography of future water challenges' (Ligtvoet et al., 2018) and 'Linking water security threats to conflict' (De Bruin et al., 2018).

In the present study, we have chosen a wider scope to make our analysis as complete as possible. We explore and analyse a broad range of global databases containing human security indicators on national scales, varying from socio-economic indicators, climatic/weather indicators, indicators for food production to political indicators (corruption, governance, conflicts and violence). In this way, we can analyse the possible influence of water-related factors compared to other environmental factors and the broader social, economic and political factors.

The results of the study are published in two parts. The Part -I-report was published in 2018, and was directed to the available data and databases on national scales, with special attention to the reliability of these databases. These databases have multiple applications as we have shown:

- monitor human security issues such as formulated in the Sustainable Development Goals (SDGs) ,
- support research in the field of disaster risk reduction as coordinated by the UN Office for Disaster Risk Reduction and the Sendai Framework,
- support climate change adaptation research (e.g., IPCC, 2018 and 2019),
- identify hotspots of conflict and violence, this to prioritise humanitarian aid programs, such as the Central Response Fund (CERF), initiated by the UN Office for the Coordination of Humanitarian Affairs,

- feed statistical analyses and integrated assessment models that aim to analyse and predict impacts of climate change in relation to poverty, water-related tensions, migration flows and conflicts (e.g., Ligtvoet et al., 2018; Abel et al., 2019),
- feed early-warning systems for famine, such as the FEWS network, and political violence and risk assessments such as the ViEWS system introduced by Hegre et al. (2019) for Africa.

The part-II-report focuses on identifying conditions that are associated with high levels of conflict and violence, on a national scale, and assesses to what extend socio-economic and environmental indicators play a role, with a special focus on water.

The study design connects to the development cooperation policy of the Dutch Government which seeks to achieve the following goals in developing countries (BuZa, 2018):

- ***Prevent conflicts and instability.***
  Especially in fragile and conflict-affected countries, development is lagging behind. These regions are breeding grounds for radicalisation and migration.
- ***Reduce poverty and social inequality.***
  Despite the decline in global poverty, extreme poverty still exists. The government seeks to take targeted measures to reduce poverty even more. Despite the progress that has been made, inequality has increased due to social exclusion, discrimination and violence. The Netherlands is investing in giving everyone a fair chance by supporting organisations that defend human rights, women's rights and the environment.
- ***Promote sustainable growth and climate action worldwide.***
  Starting in 2018, the government is spending an additional amount of up to €80 million a year for measures in developing countries to fight climate change.

Our assessment starts with identifying conditions affecting conflict risks that are deliberated in academic and popular literature, yielding a theoretical framework for the statistical analyses. This conceptual framework consists of eight driver groups (Chapter 2; upper left panel Figure 1.1). Statistical methods, based on Machine Learning techniques, are introduced in Chapter 3, along with some considerations on causality and endogeneity.

Then, we introduce in Section 4.1 three indicators which stand for various aspects of conflict and violence. These indicators are analyzed with respect to a wide set of potential drivers introduced in Section 4.2 (upper right panel Figure 1.1). Before estimating relationships by the Random Forest approach we perform a pre-selection of potential drivers by use of scatterplot matrices and corresponding correlations (Section 5.1 and Appendix A; lower left panel Figure 1.1). This selection leads to a set of 16 drivers which are analyzed by Random Forest analyses as presented in Section 5.2 (lower right panel Figure 1.1).

In the discussion we address a number of issues, namely (i) interpretation of findings in relation to findings in the literature, (ii) stability of results over time and (iii) the linkage between our findings and the Sustainable Development Goals (SDGs). Conclusions and a future outlook are given in Chapter 7.

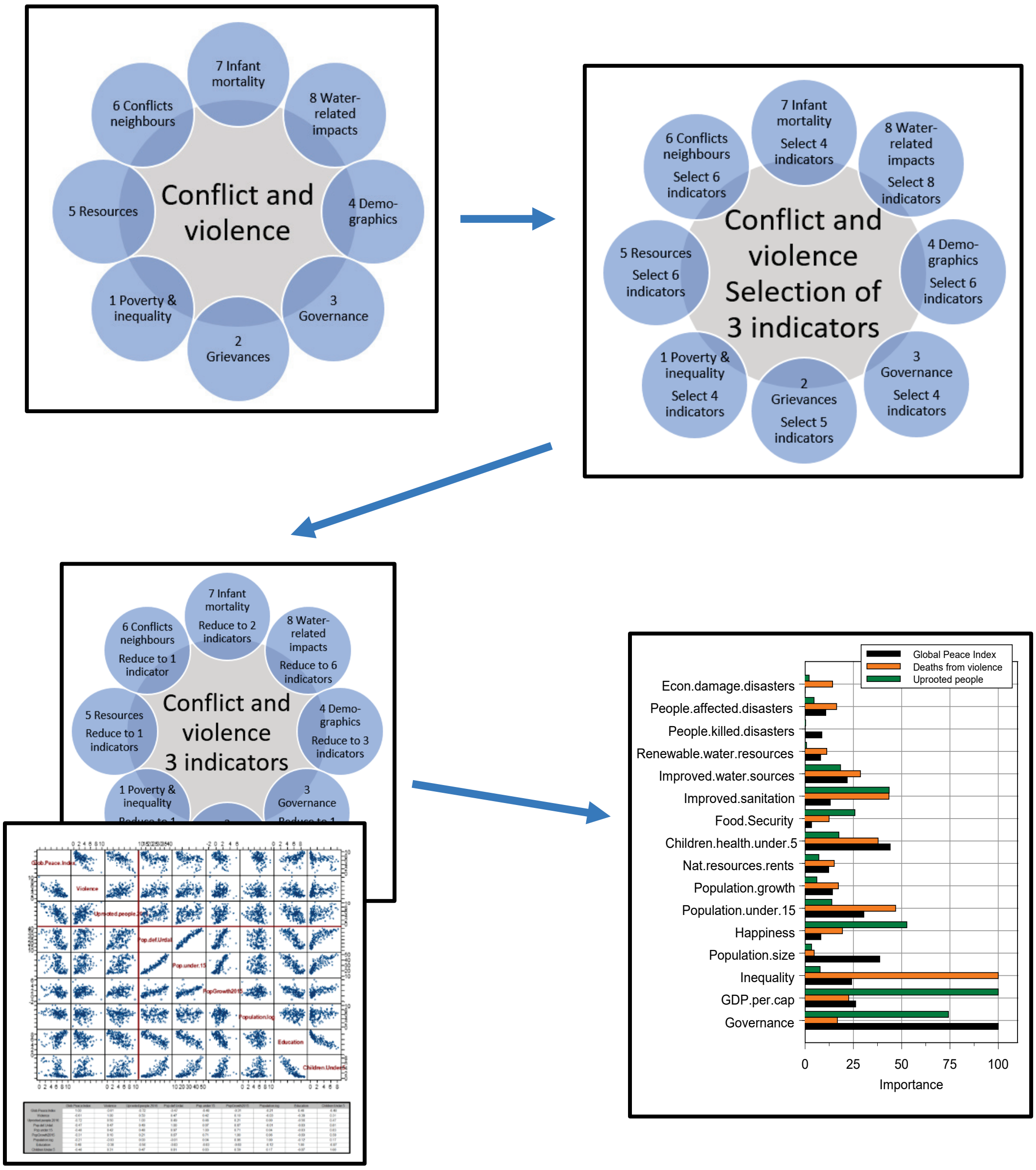


Figure 1.1 Schematic modelling approach in four stages. First, a theoretical framework is derived from the conflict literature, leading to eight driver groups (upper left panel). Second, specific indicators are chosen which resemble the essence of each of these groups in combination with three indicators which stand for various aspects of conflict and violence (upper right panel). Third, a pre-selection of drivers is performed by scatterplots and correlation matrices (lower left panel), leading to a selection of 16 potential drivers which are analysed with respect to each of the three conflict indicators as a final stage. This leads to the hierarchy of drivers shown in the lower right panel.

2 Literature on conflicts

What are the major origins and drivers of different types of conflict? Sorting out the main causes of conflict and war is difficult and often shaped by ideological believes. Even today, historians and political scientists have discussions on the primary causes of the First World War (Collier, 2007).

There are several types of conflict, ranging from state failures, international and civil wars to local conflicts, riots and revolution (Stohl et al., 2017). There are numerous theories that explain different types of conflict, which mostly focus on economic conditions and a range of factors that can foster grievances and greed, creating incentives to initiate or join a conflict. People and the organizations they are part of need reasons to start a conflict, whether these motives are legitimate or not. Limited perspectives for development, poverty, high economic and social inequality, such as discrimination, grievances due to former conflicts — also denoted as “frozen conflicts” — and unequal distributions of resources can all be motives to justify rebellion against authorities (Collier, 2007; Bara, 2014).

These perceived reasons to start a conflict may only be materialized when there are opportunities to start a conflict. A united and competent regime can handle potential insurgents, but also shocks like natural disasters, while weakened and paralyzed regimes cannot handle insurgencies, possibly leading to civil war or oppression (Goldstone et al., 2010; Besley and Persson, 2011).

In this chapter, several conditions affecting conflict risk in a broad sense are introduced that are deliberated in academic and popular literature. It must be noted though, that the outbreak of conflict is multifaceted and the effects and complex interactions of conflict variables are context specific. This makes **predicting** conflict hard and controversial (Cederman and Weidmann, 2017; Bowlsby et al., 2019). Therefore, we do not try to predict such outbreaks in this study, as done by Hegre et al. (2013, 2016, 2019), Ward et al. (2013), Halkia et al. (2017) and Witmer et al. (2017). This report rather focuses on identifying conditions that are associated with high levels of conflict on a country scale and assesses the extent to which environmental indicators — with a focus on water — may play a role.

In recent years, the world has become slightly more violent due to a small number of highly violent conflicts, increasing the absolute number of refugees (UNHRC, 2015). The absolute number of state-based and non-state-based conflicts has also increased since around 2011, whereas one-sided conflict, when one party uses violence against a non-violent group, has declined. See Figure 2.1 upper panel, where data are shown over the period 1975-2017.

This report does not go into the historical details of specific conflicts. For such details we refer to the interactive UCDP website where one can zoom in for individual countries and their history of conflicts. An example is given for Rwanda in the lower panel of Figure 2.1. Other references are Andrews (2017) and Freedman (2017).

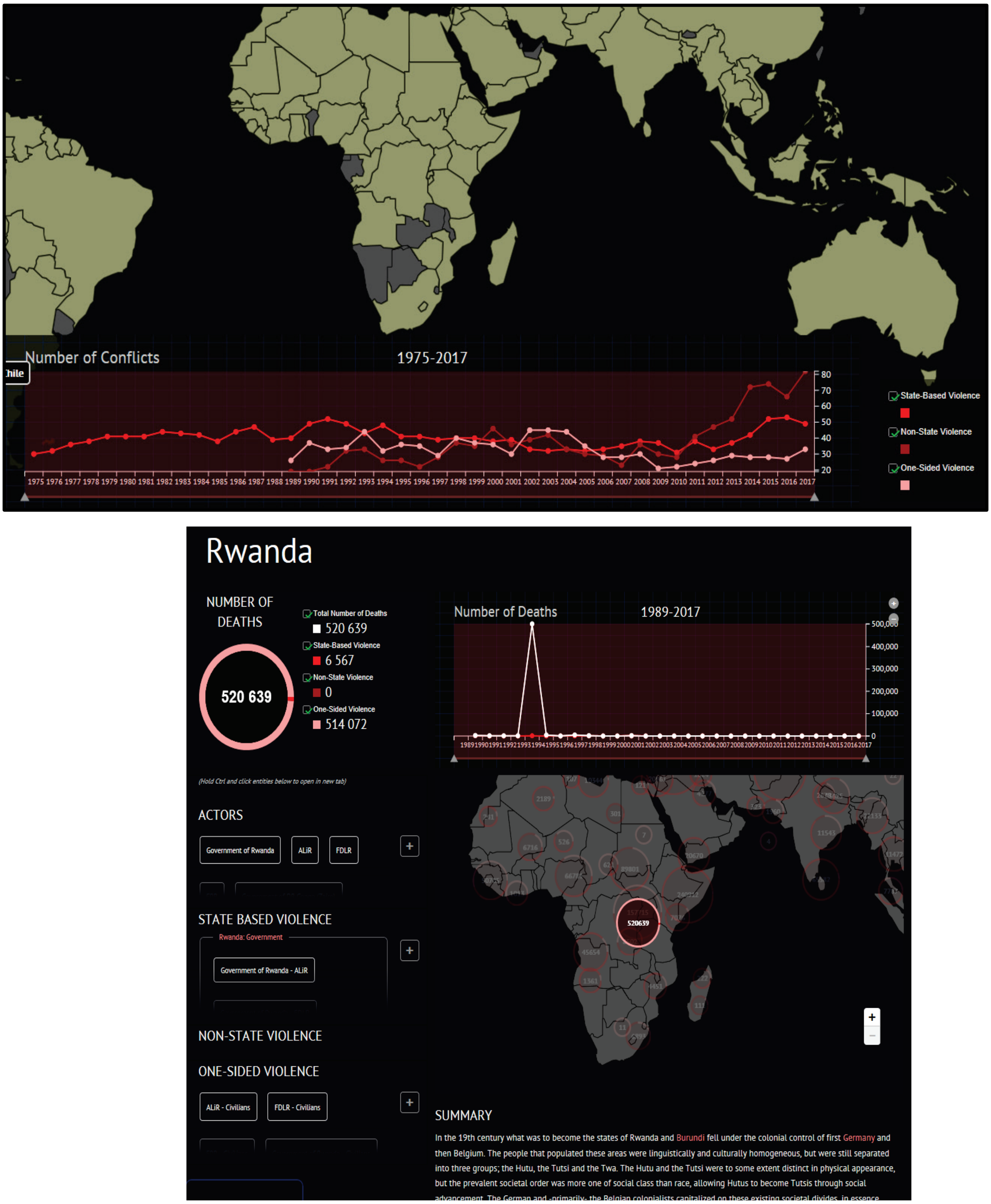


Figure 2.1 Number of conflicts world-wide since 1975 (upper panel). Data are from the Uppsala Conflict Data Program (UCDP) and screen shot is taken from their interactive website: http://www.ucdp.uu.se/. Three types of violence are discerned: state-based (light red), non-state (dark red) and one-sided (rose). For definitions please see Section 2.2.7 in the Part-I-report. By zooming in statistics for individual countries are displayed, here for Rwanda (lower panel). By zooming in further detailed descriptions are given for individual conflicts, such as for the extreme one-sided violence in 1994 where around 500.000 people were killed.

In this chapter we describe the first stage of the modeling approach followed in this report, illustrated in Figure 1.1 and reprinted below:

2.1 Indicators for incentives and opportunities

Six often cited but diverse studies assessing conflict variables are discussed to identify major indicators. These studies, summarized in Table 2.1, do not define conflict or war in the same way, neither do they make use of the same model characteristics. Therefore, we compare their conclusions in a qualitative rather than a quantitative way.

Next to that, variables used in the models and studies are defined differently. Two out of six studies are forecasting models giving future conflict projections based on historic data, while the other four studies analyse the main driving (explanatory) variables. It is still valuable to analyse the conclusions of these studies, to see whether, and if so how, water-related environmental changes pose a risk to conflict.

In general, the studies emphasize the role of *governmental institutions, economic inequality, poverty* and *demographics*, although the studies do not agree on the major variable(s). The studies do agree on the observation that civil wars are disproportionately concentrated in *poor* parts of the world where inequality is high (Besley and Persson, 2011). The article of Hegre et al. (2013) is the only study predicting armed conflict on the long term, towards 2050. This study does not take variables into account that cannot be adequately forecasted, such as political institutions that are central in the other studies.

The dominant factor for conflict and insurgency according to both the leading studies of Collier and Hoeffler (2004), and Fearon and Laitin (2003) are primarily the lacking of economic opportunities for deprived populations favouring insurgency, like joining rebellion organizations. Collier and Hoeffler focus on an economic calculus of costs and opportunities for the control over commodities, with an additional effect from fear of domination by ethnic majorities or *grievances* resulting from *former conflicts*.

Not only former conflicts play a role; Bara (2014) and Goldstone (2010) also take the tensions in *neighbouring countries* into account. Hegre at al. (2013) and Collier and Hoeffler (2004) see ethnic cleavages as an increased risk for conflict, however Fearon and Laitin

Table 2.1 Major causes for conflict according to six leading conflict studies.

Authors	**Summary**	**Main (explaining) variables** (factor number as in Figure 8.2)
1 Goldstone et al. (2010): *A global model for forecasting political instability*	This model distinguishes countries that experienced **intrastate** instability from countries that did not, built on onsets of p political instability based on events from 1955 – 2003. The model uses few variables, of which political institutions is regarded as the most dominant one by far.	- Instable political institutions (3) - High infant mortality (7) - Conflict in neighbouring states (6) - Political/economic discrimination (2)
2 Hegre et al. (2013): *Predicting armed conflict 2010 – 2050*	This model predicts **global and regional** armed conflicts for the 2010 – 2050 period based on data from 1970 to 2009. Predictions are made for no conflict, minor conflict and major conflict.	- Population size (4) - Infant mortality rate (7) - Demographic composition (4) - Education levels (4) - Oil dependence (5) - Ethnic cleavages (11) - Neighbouring characteristics (6)
3 Fearon and Laitin (2003): *Ethnicity, insurgency, and civil war*	This study searches for the causes of **intrastate** conflict by using data from 1945 to 1999. The authors reject a focus on ethnic or religious characteristics as a root cause for conflict. Factors that favour insurgence explain increased risk on conflict. This study includes colonial wars where others do not.	- Poverty, slow economic growth (1) - Political instability (3) - Rough terrain (10) - Large populations (4)
4 Besley and Persson (2011): *The logic of political violence*	This study analyses whether **intrastate** political violence emerges in the form of repression or civil war and which economic and political factors drive one-sided (repression) or two-sided (civil war) violence.	- Political institutions – policies (3) - Shocks affecting individual incomes and aid, and the timing of shocks (-)
5 Collier and Hoeffler (2004): *Greed and grievance in civil war*	Analyses of causes of civil war in the period 1960-1999. Grievances and opportunities are being approached as main incentives for war, although proxies for these factors are hard to find.	Grievances: - High inequality (1) - A lack of political rights (3) - Ethnic & religious division (11) Opportunity: - Capture of resources (2,5) - Gaining power (-)
6 Bara (2014): *Incentives and opportunities: A complexity-oriented explanation of violent ethnic conflict*	This study uses the method of qualitative comparative analyses from 1990-2009. The study shows that the discussion concerning whether conflict is opportunity driven or incentive driven is a false one. Both incentives and opportunities must be present to drive a conflict.	- Conflict trap (2, 13) - Bad neighbourhood (6) - Ousted rulers (1, 2) - Resource curse (2, 5)

(2003) conclude that more ethnically diverse countries are not more likely to experience civil war. Both Bara and Collier and Hoeffler focus on *resources*, in terms of abundance and scarcity, as a driver of conflict.

Goldstone et al. (2010) and Besley and Persson (2010) emphasize that the role of *state institutions* is more important for the development of conflict than economic incentives. A united and competent regime can handle potential insurgents or shocks like natural disasters, while weakened and paralyzed regimes cannot handle insurgencies, possibly leading to civil war or oppression. Within these studies, the access to resources is indirectly part of the analyses. The study of Besley and Persson study how natural disasters cause negative shocks on wage rates and how these shocks are related to the occurrence of civil conflict.

Within the model of Goldstone et al. (2010), child mortality is used as a proxy for the availability of *sufficient food and water*, health care and sanitation. However, the availability of, and access to *resources* and the impacts of shocks on societies is considered to be at least partly dependent on policies and the *capabilities of institutions*.

For a general overview of conflict drivers (or 'pillars of conflict') we refer to Stohl et al. (2017). The recent study of Mach et al. (2019; their Figure 3a and Supplementary Table 1) presents a hierarchy of conflict drivers based on expert elicitation. We will discuss their results in Section 6.2.

Box 2.1 Governance

Governance has been defined in many different ways. In this report the governance indicator of the INFORM database is used for the analyses. This composite indicator is a combination of two independent indicator:

Governance Effectiveness

'Reflects perceptions of the quality of public services, the quality of the civil service and the degree of its independence from political pressures, the quality of policy formulation and implementation, and the credibility of the government's commitment to such policies.' (Kaufman and Kraay, 2015).

Corruption

The CPI scores and ranks countries based on how corrupt a country's public sector is perceived to be. It is a composite index, a combination of surveys and assessments of corruption, collected by a variety of reputable institutions (Transparancy International, 2017).

2.2 Conceptual model

Based upon the analyses of the authors summarized in Table 2.1, related literature and — additionally — the role of environmental change, we propose a theoretical framework, denoted here as 'conceptual model', covering a wide range of drivers which relate to conflict risks. The model is shown in Figure 2.2. Here, the (potential) drivers of conflict are grouped into eight factors for which we give a short explanation, along with references to the literature. Factors 9 through 13, shown at the right of the graph, are reviewed separately in Section 2.3

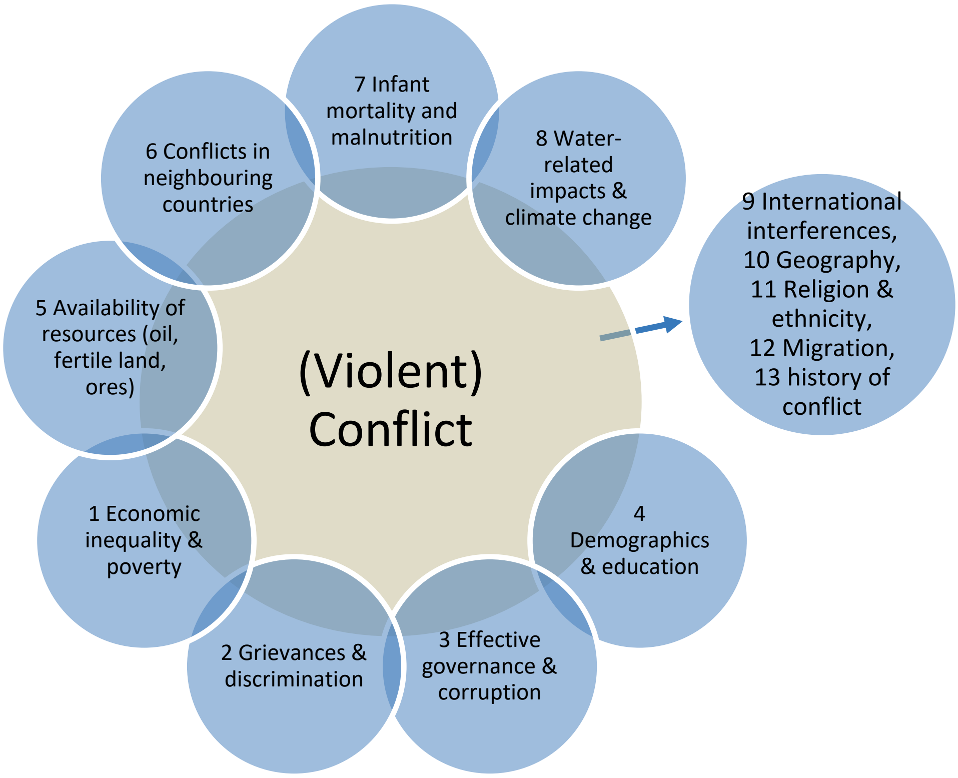


Figure 2.2 Conceptual model showing eight drivers of violence and conflict, all taken from the conflict literature, including related literature on climate change, migration and conflict.

1. Economic inequality and proverty. A major risk factor for conflict is poverty. Especially in countries were economic inequality is high and were the poor have limited economic perspective (Ward et al., 2013; Bara 2014; Håvard et al., 2016). As stated in Section 8.1, the absence of economic opportunities for deprived populations may favour insurgency, such as joining rebellion organizations. Numerous studies agree that civil wars are disproportionately concentrated in poor parts of the world where material inequality is high (Besley and Persson, 2011).

Paul Collier, a leading scientist in the field of conflict studies, found that halving the starting income in low-income countries doubles the risk of civil war (Collier, 2007). On the individual level, the relationship between poverty and conflict can be attributed to the limited perspectives people have and that there may be less to lose in a material sense. Conflict may be a tool to improve living conditions and change the status quo. At the state level, poverty can lower resilience to rebellion and unrest since government effectiveness is low, stripping capacity for public goods provision, and limiting the projection of power and authority, whether soft or coercive (Collier et al., 2003).

2. Grievances and discrimination. Political or economic discrimination of religious, ethnic or political (minority) groups can be a reason for people to be dissatisfied with the government and develop grievances. Countries with high levels of state-led discrimination face around triple the relative chances on civil war compared to countries without discrimination, according to Goldstone (2010). Bara indicates that 'ousted rulers' – groups ousted from a position of power – increase the risk on conflict in a situation of instability. Not all studies agree upon this finding though. For example, Collier (2007) argues that oppressed and discriminated people are often not in the position to rebel, since these groups have little access to resources and hold little power. An overview of current research has been given by Stohl et al. (2017, p. 17-31).

An example is the discrimination and suppression of the Rohingya minority in Myanmar. The decade long suppression of this group resulted in conflict between the army and militant Rohingya. However, most Rohingya people do not have the power or resources to fight the military and therefore their only option is to flee to a saver region, mostly in Bangladesh and Thailand.

3. Effective governance and corruption. A number of studies reject the idea that either poverty and economic marginalization or discrimination and suppression are the decisive causes of conflict within a state. Discrimination, leading to grievances and a lack of economic perspective can create tense societies, but if political institutions are cohesive and powerful, no (violent) conflict can develop according to a number of prominent studies (Fearon and Laitin, 2003; Collier and Hoeffler, 2004; Besley and Persson, 2011). A strong regime may defeat any form of insurgency within a country and, therefore, decrease the opportunity that insurgents are successful (Goldstone et al., 2010). The central observation of this idea is that a united and competent regime is able to handle potential insurgents or shocks such as natural disasters, while weakened and paralyzed regimes cannot handle insurgencies, possibly leading to civil war or oppression. Thereby, local populations may not trust weak states and choose their own paths (Collier et al., 2003). An example may be North Korea, where poverty is widespread and grievances may exist, but conflict is absent, as far as we know of.

The development of motives to start a conflict and the presence of opportunities to start a conflict may be intertwined with each other. Bara (2014) shows in her study that the combination of incentives *and* opportunities is required for conflict because a group/country should be both willing and able to rebel or resist. Thereby, it is hard to distinguish incentives and opportunities, according to this study. A minority group can have reasons to rebel because of political exclusion, but the optimal moment is when political instability in the power center arises.

The role of effective/uneffective governance is highlighted in the work of the economist Acemoglu, be it with emphasis on economic development, rather than conflicts. See Acemoglu et al. (2005) and Acemoglu and Robinson (2012).

4. Demographics and education. An important characteristic of population is its age structure and it has been identified by several studies as a risk factor for conflict. Countries that have large youth populations relative to the older generations, so-called 'youth bulges', are found to face higher conflict risks, especially in conditions of economic stagnation (Goldstone (2002); Urdal, 2004; LaGraffe, 2012; Hegre et al., 2013).

Aspects other than rioting and revolution in societies with little economic perspectives for the youth have been linked to youth bulges. A study of Nordas and Davenport (2013) found that large youth cohorts increase the repression of state authorities, since the younger cohorts of the population are more likely to challenge authority. The Arab Spring has been linked by a number of studies to the large youth cohorts in countries such as Egypt, Yemen, Syria and Tunisia (LaGraffe, 2012; Malik and Awadallah, 2013). Young, often educated people, were mostly the ones who took the streets and started the revolution. Simply sad because of their limited economic perspectives and dissatisfaction with the authoritarian regimes, although there were more causes and grievances involved in the escalation of some of the revolutions (Hoffman and Jamal, 2012).

5. **Availability of resources**. The presence or absence of resources can increase conflict risks in different ways. Scarcity or inaccessibility of resources can cause livelihood insecurity (poverty and limited economic perspectives have been taken into account in factor 1). Rising food prices have also been linked to unrest in society, especially in societies that are already politically instable (Smith, 2014; Bellemare, 2015; Natalini et al., 2015). Resource scarcity does not per se lead to conflict: fighting during dry, water scarce years for example is suicidal in some pastoral societies (Eaton, 2008). Therefore, droughts may lead to *increasing levels of cooperation* and *reconciliation* instead of rising conflict risks (Theisen, 2012).

The abundance of resources, especially resources valuable in an international context (oil, ores) - the so-called ‘resource curse’ - can increase conflict risks in a number of ways. The extraction of natural resources has been linked to corruption, suppression, economic decline and civil war in numerous case studies (Basedau and Lay, 2009). Extractive governments or industries can increase feelings of inequality and grievances, what might decrease state legitimacy (Acemoglu and Robinson, 2012).

Resource abundance makes it more interesting for domestic groups to engage in quasi-criminal activities (Humphreys, 2005). Thereby, the availability of funds to finance rebellion can increase opportunities for success (Collier and Hoeffler, 2004). It is not always the government or local groups in a certain resource-rich region that increase conflict risk: the presence of resources may be an incentive for outsiders to engage or foster conflict (Humphreys, 2005). During times of conflict, resource-rich places such as oil fields often are primary targets (Collier, 2007).

Recently, several studies pointed to the role of 'water' as an important natural resource and directly or indirectly connected to violence and conflict. We name Raleigh and Vik Bakken (2017), De Bruin et al. (2018) and Gleick and Iceland (2018).

6. **Conflicts in neighbouring countries**. The role of conflicts and violence in neighbouring countries has been proposed by Goldstone et al. (2010) and Hegre (2013). Other studies that investigate these geographically-oriented relationships, are the PhD theses of Buhaug (2005) and Höhne-Sparborth (2018). The last study treats a number of case studies such as for Zambia, next to examples from Malawi, Belize, Jordan and Thailand. These case studies demonstrate that the mechanisms identified have relevance in varied conflict situations, but that the net effect of individual channels of spill-over are dependent on local risk factors and policies.

In the present study we do not aim to predict conflict and violence, be it in the short- or the long-term. However, we will test the role of neighbouring countries by adding this variable to the set of explanatory variables in Chapter 11.

7. Infant mortality rate. Rodwan Abouharb and Kimball (2007) show the importance of infant mortality measures in political studies, economics and demography. Infant mortality rates are sensitive to distributional issues, and especially, how well governments provide for their citizen's economic and social welfare. As such, infant mortality rates will interact with conflict and violence on national scales (Goldstone et al., 2010; Hegre et al., 2013), although this indicator serves as a proxy for low levels of development.

8. Water-related impacts and climate change. The potential relationship between water and climate-related impacts has been the starting point of the work of PBL for the Planetary Security Initiative, resulting in studies by Ligtvoet et al. (2018) and De Bruin et al. (2018). Please see the following website. As stated in the Introduction of the Part-I report of this study, water, climate and conflict relationships are contested and dependent on contextual factors. Studies towards the causes of conflict hardly align their results with water or climate-related impacts, and only if related to economic shocks in terms of wages and aid (Besley and Persson, 2011).

However, it can be argued that the main drivers of conflict can be linked to water in different ways and on different scales. Since water-related issues are not expected to become sole causes of conflict, it is important to analyse indirect, conjunctural effects via adverse economic and livelihood impacts (Buhaug, 2016). The most direct one is the intensification of poverty and economic inequality because of water-related events. Especially when events occur in cascade, people that are already poor, are likely to fall (back) into poverty (Hallegatte et al., 2016; Olsson et al., 2014). The World Bank has estimated that the effects of climate change could push 100 million people more into poverty by 2030, unequally divided across the world, leaving socio-economic drivers of water stress out of the estimation (Hallegatte et al., 2016).

Next to that, extreme water or climate-related events increase stresses on resources important for livelihoods, eroding the legitimacy of states (Femia and Werrell, 2017). Increasing food prices may also trigger social unrest where already deprived people face livelihood uncertainties (Bellemare, 2015; Smith, 2014). The stability of institutions will be mainly threatened by the ability and political will to handle projected changes, rather than just the effects of climate change. Decreasing legitimacy can be linked to decreased governance effectiveness, and, again, peace.

In most parts of the world though, climate and water conditions do not constitute a *direct* threat for peace on the national scale. Regions already prone to conflict, face higher risks of conflict related to water security threats, especially on the local level.

Box 2.1 Did climate change contribute to the Syrian war?

Some studies stated that there is evidence that the lingering drought in the Fertile Crescent contributed to political unrest via agricultural failures, livestock mortality and large-scale migration (Kelley et al., 2015; King, 2015). Some scholars opposed this firm conclusion by showing that there is no reliable evidence that antropogenic climate change impacted the droughts in Syria, let alone that these droughts caused pre-conflict migration or that migration levels affected conflict risk (Selby et al., 2017). Fact is that some parts of Syria are short on water and that this caused problems in agriculture, resulting in higher food prices. This is not only a result of the weather patterns, but also largely of resource mismanagement causing humanitarian problems (De Châtel, 2014). All together, it is not possible to measure the relative contribution of increasing water stress prior to the Syrian War. But it can be shown that decreasing harvests — due to resource-mismanagement or droughts — added to rising food prices, economic marginalisation of farmers and (temporal) migration can be proved (Gleick, 2014).

2.3 Factors not included in this report

Not all drivers proposed in the literature are contained in Figure 2.2. Next to these eight drivers, five others are proposed, shown at the right-hand side of the scheme. Here, we shortly describe these factors (9) up to (13), and give arguments why they are left out in our analyses.

9. Proxy wars / international interferences / international politics. In this report we do not account for the complicated relationships between states that can be expressed in other countries, so called proxy wars or internationalized conflicts. Different international coalitions in different wars, have been fighting each other in **another** country. Most recently, international coalitions influenced the wars in Syria and Yemen. Proxy wars are not a new phenomenon: the Cold War was an era of proxy wars in which the United States fought with the USSR, for example in Vietnam (Mamdani, 2005). A long list of proxy wars is given here.

Countries can also choose to empower specific domestic rebel groups that can undermine the power of the enemy state (Salehyan, 2010), influencing opportunities of rebel groups. For example, the United States funded insurgencies in Nicaragua and Afghanistan in the 1980’s to destabilize the central governments (Salehyan, 2010).

Strand et al. (2019) show that the number of countries involved in internationalized conflicts has increased dramatically over the past decades (their Figure 4). They state that *research indicates that internationalized conflicts are more persistent and less likely to find a political solution. This durability can be due to aspects of the conflict itself, but may also be driven by the increasing number of parties involved in the conflict, which means more actors who can potentially block a deal.*

Other forms of international interference are humanitarian aid funds, such as the UN Central Emergency Response Fund (CERF, see Section 3.2.2 of the Part-I-report). There are indicators for aid dependency of countries such as the World Bank indicator 'Net ODA received as percentage of GNI', where ODA stands of Official Development Assistance and GNI for Gross National Income.

We did not incorporate indicators for international (proxy) interferences since it is unclear if such interferences lead to more (or less) tensions or violence in countries, or the other way around, that countries with weak institutions and high levels of violence attract foreign interferences (be it military support for certain groups or humanitarian aid). And, subsequently, these last interferences themselves can weaken or strengthen countries (and so on, and so on). Next to that, it is not easy to find a proper indicator for proxy interferences.

10. Geographical distribution of conflicts. According to some studies, certain geographic landscapes (mountainous areas, distance to the centre of power, jungle areas) are important factors when rebellious groups want to be successful (Fearon and Laitin, 2003; Buhaug, 2005). This is because 'rough terrain' poses opportunities for groups to be relatively uncontrolled. Next to that, communities in these areas often receive less state support than people living in the capital. They are therefore more supportive for rebellious organizations. The existence of these landscapes is not taken into account in this study since the role of these landscapes is (i) contested and (ii) may play a role in local insurgency only (Tollefsen and Buhaug 2015).

Conflict risk is not only linked to certain socio-economic factors, but risks are also intertwined with geographic location. The risk of civil conflict is seven to ten times higher in drylands and tropical zones than in cooler, continental climate zones (Buhaug and Rudolfsen, 2015). The explanation for this distinction is not well understood. Poverty, however, is more frequent in drylands and tropical zones, via agricultural activity, and to the prevalence of diseases and the restrictions to develop infrastructures for economic development (Sachs et al., 2001). See Figure 2.3.

Climatic conditions have influenced human developments over a history of thousands of years, heavily influencing and reinforcing the 'unequal' distribution of wealth and power

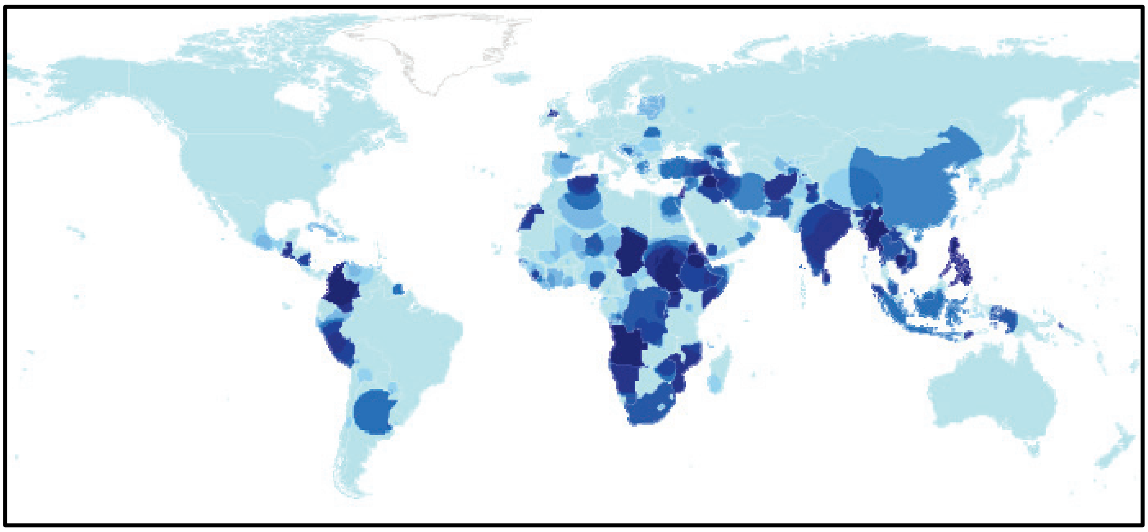


Figure 2.3 Global distribution of armed civil conflict in the period 1946-2014. Darker shades indicate more persistent zones of conflict (Buhaug and Rudolfsen, 2015).

today (Diamond, 1998; Fenske, 2013). These long-term perspectives on the interlinkages between climate, poverty and geographical location suggest that intensifying harsh climatic conditions in already vulnerable places may increase future conflict risk. But these suggestions cannot be tested in this report since only present-day annual data are included, not long-term changes.

11. Religion and ethnicity. Religion plays a role in several contemporary conflicts. As examples we name (i) the declaration of an Islamic State in Iraq and Syria, (ii) the Buddhist majority that is undermining the rights of the Muslim Rohingya minority in Myanmar, and (iii) the conflict between Muslim Sudan and the new former country of South Sudan, with a predominantly Christian population.

Religion can provide legitimacy and identities for certain actions based on rules and behaviour, providing identities and religious institutions for powerful and resource-rich organizations (Fox, 2017). History is full of religious conflicts, such as the crusader wars in the Middle East. However, religion is not taken into account in this report for two reasons. First, religion is not used as a variable by the studies used for this report (Table 2.1). Second, religion is only of importance in conflicts where economic, social or political inequalities are explained by religion (Fearon and Laitin, 2003).

Ethnicity is another variable that has is not taken into account. Although a popular narrative in media, there is no consensus concerning the risks diverse ethnic or religious compositions of a country bring along (Besley and Persson 2011; Ward et al., 2013). Some violent conflicts have been fought along ethnic lines, such as the ethnic violent conflict in Rwanda; other countries that had a bloody civil war, were ethnically ‘pure’, such as Somalia (Collier, 2007). Some studies find some risks for societies in which one ethnic group is in power – ethnic dominance – but this relationship has only been observed in economically less developed countries (Goldstone et al., 2010; Besley and Persson, 2011).

12. Migration*.* Forced migration, also defined as displacement, is caused by violent conflict, but some scholars see an increased conflict risk if large flows of migrant come into regions. Brzoska and Fröhlich (2016) identified three types of receiving areas that are conflict-prone as a result of in-migration:

- *Regions with extreme resource scarcity*: if receiving areas already face absolute resource scarcity (food, water) for different reasons, incoming migrants may be seen as competitors, possibly increasing tension or even conflict.
- *Regions with high level of conflict*: in regions where tension over identities or interests are high, the potential of migrants to become a conflict driver or trigger of conflict is relatively high (compared with peaceful areas), especially when migrants influence identity conflicts.
- *Regions with exclusive identities*: when migrants arrive into communities unwilling to accept others, tension or even conflict can arise. This can occur due to political discourses emphasising economic scarcity due to migrants, or within communities fearing the erosion of traditions, when migrants are perceived as different and threatening.

We did not include migration as factor in this study since it is (partly) contained in the violence- and conflict-related variables introduced in Section 4.1. The term 'partly' is added since migration can be initiated by violence or conflicts - which is meant here - or by economic deprivation.

13. History of conflict. Some studies highlight the role of historical conflicts as a predictor for the present (and future) conflict situation. We name the study of Hegre at al. (2016) who use an indicator for historical conflicts and violence next to indicators for GDP per capita and levels of education. Furthermore, the study of Mach et al. (2019), based on expert elicitation, names a history of conflicts as an important driver (cf. Figure 6.4 — reprinted from their study).

Since the present study is not directed to prediction and the fact that historic conflicts are strongly reflected in present-day indicator values (strong persistence), we will not use an indicator for 'historical conflicts' in the next Chapters.

The temporal persistence of indicators has been illustrated in Section 4.1 of the Part-I-report for a number of indicators. As for the Global Peace Index we show its temporal persistence in Figure 2.4. Here, we compare 2008 values for 163 countries with their 2016 values. The graph shows a strong clustering around the 1-to-1 line, apart from a small number of countries which came into war after 2008. An example is Syria where a civil (proxy) war started in 2011. The position of Syria in the graph is far above the 1-to-1 line.

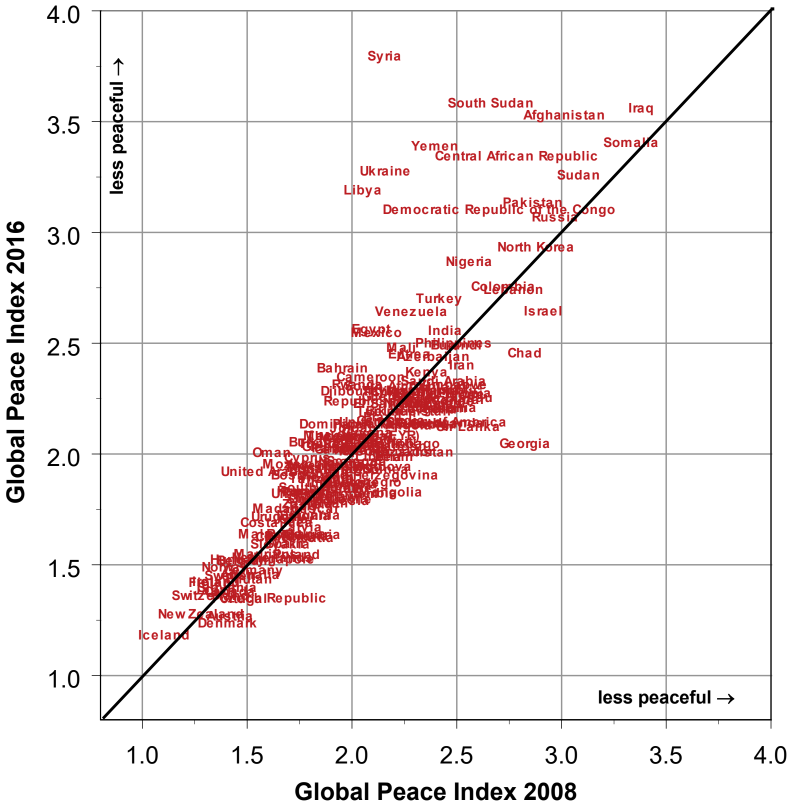


Figure 2.4 Scatterplot for the Global Peace Index across 163 countries. The 2008 values are plotted on the x-axis and 2016 values on the y-axis. Countries at the left of the 1-to-1 line show changes to less peaceful conditions from 2008 to 2016 and countries lying at the right vice versa.

3 From conceptual model to statistical inferences

3.1 Preliminary comments on causality issues

All factors given in Table 2.1 and Figure 2.2 will show some relationship or association to violence and conflict. However, it is difficult to say in what sense or to what degree these factors lead to global patterns of violence and conflict. Here, we assume that all factors named in the literature are in some way related to conflict, and may thus be seen as potential threat multipliers. Since we selected the drivers from the main-stream conflict literature, false correlations/relationships are unlikely. However, factors will differ in strength and explanatory power, introducing a hierarchy in explanatory strength.

The relative importance of drivers follows from the techniques we will describe in Sections 3.2 and 3.3. However, to make the hierarchy in conflict drivers better understandable, we will adopt the terminology used in the field of biology and economics, that of structural (ultimate) factors and proximate factors. Here, a *proximate* or direct factor is seen as a factor which is closest to, or immediately responsible for some effect. This is in contrast to a higher level or upper level effect, denoted as *structural effect*, respectively. This latter factor is more fundamental in nature. For example, Hofman (2001), Acemoglu et al. (2005) and Acemoglu and Robinson (2012) analyse long-term economic developments to find the structural (ultimate) causes of economic growth. In Section 6.1 we will analyse how the factors given Table 2.1 and Figure 2.2 might fit into a similar hierarchical scheme.

We note that showing causality between two variables is a complex issue. One problem is that of **endogeneity** of two variables under consideration. Fearon (2010, p. 44) states:

The second major problem one faces when trying to use governance indicators to assess the causes of economic growth or civil war onset is endogeneity. If an indicator is well correlated with contemporaneous growth or civil war onset, we cannot infer causality, because it could be that the observation of growth is leading the experts to think that governance is good, or that the observation of civil war leads them to infer that governance or institutions are bad.

Another example of endogeneity can be found in the work of Angus Deaton (Deaton, 2013; Weil, 2015) who clarified the interrelationships between health, income and institutional quality, both across countries and over time. More formally, if we model a dependent variable Y_i by adding a regressor X_i, how can we know that X_i causes Y_i, and not the other way around?

In this study we take a simple position in this discussion. We only select those regressors that are identified as conflict drivers in the main stream literature. These regressors are interpreted as **conditions** that relate to violence and conflict, and we avoid the term 'causality'. See Figure 3.1 for a schematic sketch of the approach followed here.

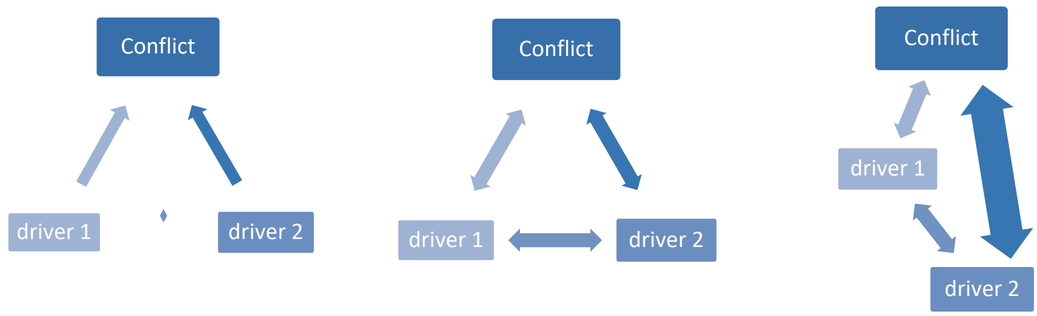


Figure 3.1 Three visualisations for the relationship between a conflict-related indicator and two drivers. The left panel shows one-directional relationships which are often interpreted as being causal. The middle panel presents the more realistic situation: all three indicators are interrelated. The right panel shows the situation adapted in this study: based on explanatory power and expert judgement drivers can be re-arranged in a hierarchical manner. We avoid the term 'causal'; drivers are seen as conditions that relate to conflict.

3.2 Machine Learning and Regression Trees

We will make a number of statistical inferences based on the data sets described in Chapters 2 and 3 of the Part I report. To do so, we first explore simple univariate relationships by showing a number of scatterplots and scatterplot matrices. The advantage of scatterplots is the visual presentation of data. All types of characteristics can be seen, such as the presence of linear or parabolic (quadratic) relationships, exponential relationships, threshold phenomena, and the presence of outliers (Tukey, 1977).

However, the fact that some variable X shows no clear association to a variable Y in a scatterplot, does not prove that there is no relationship at all. It could be that this relationship is masked by that between Y and some other variables X_2, X_3, .. Therefore, we perform a series of *multi-driver analyses* where a number of explanatory variables are applied simultaneously. To do so, we follow a non-linear approach as a second step.

In most studies indicators of violence and conflict are correlated with explanatory variables using Multiple Regression models. Here, a variable Y_i is assumed to have a linear relationship with variables $X_{1,i}$, $X_{2,i}$, $X_{3,i}$, … , with i being an index for country number 1, .. , 191:

$$Y_i = \alpha_o + \alpha_1 X_{1,i} + \alpha_2 X_{2,i} + \ldots + \alpha_N X_{N,i} + \varepsilon_i \quad (1)$$

where parameters α_o , … , α_N are constants and ε_i a random noise process (e.g. James et al. 2013 — Chapter 3). We note that Model (1) can be estimated for all regressors available, or combined with a selection procedure to select the relevant regressors (forward and backward procedures).

However, since the relationships we are looking for, may show non-linear behaviour, such as quadratic or threshold-like patterns, we apply a method based on *regression trees* which is part of the wider group of Classification And Regression Tree (CART) models. These models were introduced by Breiman et al. (1984) and gained popularity within the field of ***machine learning***. We refer to James et al. (2013, Ch. 4) and Kuhn and Johnson (2016, Chapter 8). Applications in the field of air pollution research are given by Visser and Noordijk (2002), and in the field of biology by Visser and Wortelboer (2013).

The Regression Tree model is non-linear, with the general form:

$$Y_i = f\,(X_{1,i}\ , X_{2,i}\ \ldots , X_{N,i}\ ) + \varepsilon_i \tag{2}$$

We note that model (1) and model (2) have two aspects in common. First, both models allow for *conjunctural causation*. This refers to the possibility that a certain condition has no effect on the outcome on its own, but only in combination with other conditions. Second, both models share *equifinality*. Equifinality means that the same phenomenon can be explained by different, mutually non-exclusive sets of conditions.

However, both models differ as for a third aspect, called *causal asymmetry*. Causal asymmetry implies that (a combination of) conditions causing a certain outcome (here: conflict and violence) are not necessarily a mirror image of those conditions causing the absence of this outcome. This aspect is important since the presence or absence of violent conflicts cannot be treated as simple binary oppositions (Ide, 2015, Section 3.1).

Regression Tree analysis is a statistical technique that divides the set of measurements Y_i into two subsets on the basis of one of the proximate variables $X_{1,i}$ .... $X_{N,i}$. The criterion for the division of the subsets is the minimization of the variance of the two subsets. Suppose that variable $X_{j,i}$ is selected. Now the Y_i values fall into two subsets: one group for countries where [$X_{j,i}$ < threshold c], and one complementary group of countries where [$X_{j,i}$ ≥ threshold c]. After this first step in the analysis, one of the subsets is itself again divided into two new subsets, with again the criterion of minimization of variance. Eventually, this leads to a 'tree' of classes describing the influence of the proximate variables shown in Eq. (2).

Once a tree is generated we want to check if all the nodes in the tree are needed or if we should *prune* the tree. The rationale for pruning the tree is that we want to have a model for our data that is as *parsimonious* as possible, while keeping certain desirable characteristics intact (such as the predictive power of the tree). The final nodes of the tree are called 'leaves'. By averaging all Y_i values that correspond to that specific leaf we get an RT prediction mean(Y_i) for countries that fall in the particular node. The 'predictive power' of the full tree is found by calculating (i) the sum of all squared errors $[Y_i - \text{mean}(Y_i)]^2$ for each leave (also denoted by the term 'deviance'), (ii) add these squared errors for all leaves, and (iii) compare this sum to the initial sum of squares if no explanatory variables were available.

3.3 Random Forests

A variation on Regression Trees which will be applied here, namely **Random Forests**, a method also initiated by Leo Breiman (Breiman, 2001). This method estimates an ensemble of, say, 1000 Regression Trees by splitting the data set at hand randomly in a training set, being two third of the data, and a test set, being one third of the data set. From this test set a bootstrap sample is made by random selection to make a full simulated set of records (thus, if the original set contains M countries, the bootstrapped set contains M countries too). This procedure is also denoted as 'bagging' which stands for Bootstrap aggregating. One advantage of bagging procedures is the stability of model predictions (Kuhn and Johnson, 2016 – their Section 8.4). In other words, the risk of 'over-fitting' is much smaller than that if only one regression tree is estimated (as described in Section 3.2).

Each of the 1000 regression trees is used to predict the values of the test set (being roughly one third of the full data set) to evaluate the predictive strength of the trees. These predictions are denoted in jargon as 'Out Of Bag (OOB) data' and can be used to evaluate a Root Mean Squared Error (RMSE), based on all these OOB data. It is important to note that these predictions are for records **not** contained in the trees estimated and does not come from fitting to the model to the data at hand.

One important aspect of the trees estimated within the Random Forest approach is that of splitting at a specific node in the tree. In a traditional regression tree the best regressor for a certain splitting is chosen from the full set of M regressors. However, within the Random Forest approach the model choses amongst a subset of M/3 regressors, which are randomly chosen from the full set of M regressors. This approach has the important advantage that subsequent regression trees will be quite different in structure.

The procedure is summarized in Figure 3.2 for a situation with a conflict data set for 210 countries (all data for a fixed year). Since the structure of these 1000 regression trees cannot be averaged/aggregated in the normal way, two methods have been developed to show **the relative importance** of the independent variables $X_{1,i}$ .... $X_{N,i}$. These methods are denoted as "Incremental MSE" and 'Incremental node purity'. For details of these importance measures, please refer to the publication of Liaw and Wiener (2002) and Liaw (2018). For a general description please refer to James et al. (2013, Chapter 8), and Kuhn and Johnson (2016, their Chapter 8 and more specific Section 8.5).

One aspect of the Random Forest approach is important to mention here: the potentially distorting role of correlated regressors (a problem also denoted as multicollinearity). Kuhn and Johnson (2016, p. 202) argue that such between-predictor correlations lead to a (small) dilution of importance function values. That is, importance values of two drivers which are mutually correlated will have both (slightly) lower values if compared to the situation where one of these two drivers would be omitted. Thus, multicollinearity has, unavoidable an influence on the importance functions estimated, but not in the sense that one variable will dominate the either.

Another aspect worth mentioning, is that of the imputation of missing driver data. Within the framework of Random Forest data can be interpolated using the so-called proximity matrix. The (i,j) element of the proximity matrix produced by the Random Forest software is the fraction of trees in which countries i and j fall in the same terminal node. The intuition is that 'similar' countries should be in the same terminal nodes more often than dissimilar ones (Liaw and Wiener, 2002; Liaw, 2018).

In summary, we choose for the Random Forest approach for a number of reasons. First, the predictive strength is evaluated on independent data (the validation set). Second, the principle of 'bagging' assures more stable predictions compared to a single regression tree. Third, we can evaluate importance measures which can be interpreted as levels of association. Third, the approach is able to estimate both linear and non-linear relationships between an independent variable and regressors. Finally, the approach is attractive due to the three aspects set out in Section 3.2, that of conjunctural causation, equifinality and causal asymmetric.

A disadvantage of the Random Forest approach is that we do not have a single model which shows how interactions between regressors $X_{1,i}$ .... $X_{N,i}$ and independent variable Y_i are estimated. In other words, we do not have a concrete regression tree. Instead we have 300 trees or alike, which are all equally likely and all with different splits. For the study at hand this disadvantage is not relevant since we mainly seek the *importance* of potential drivers which is given by the importance functions contained in the random forest approach. For the conflict cases at hand, we also analysed regression trees in Appendix E.

A simulation example illustrating the Regression Tree approach and Random Forests, is given in Appendix B.

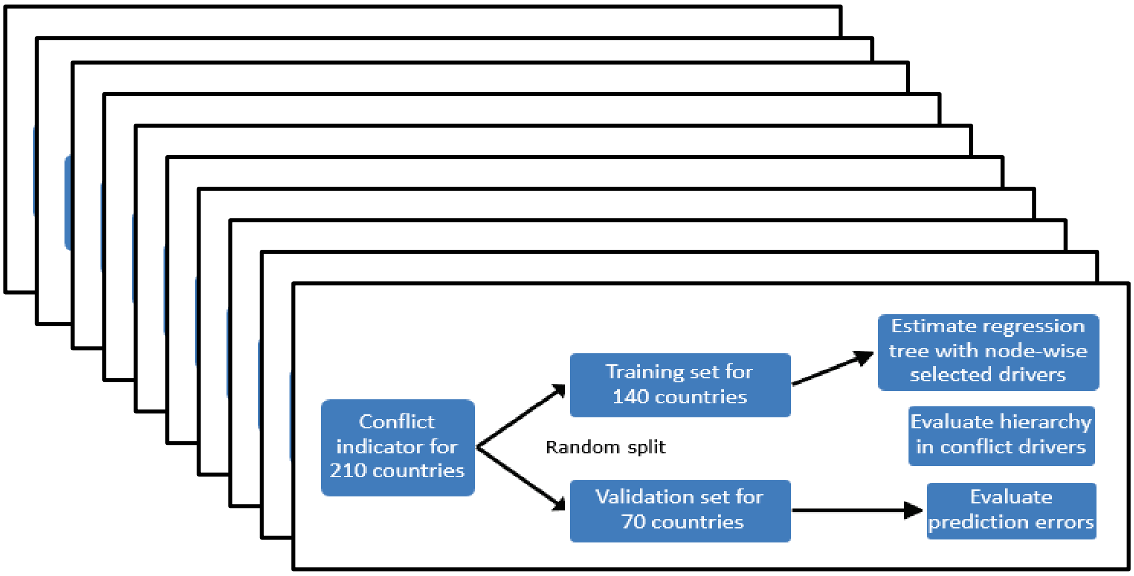
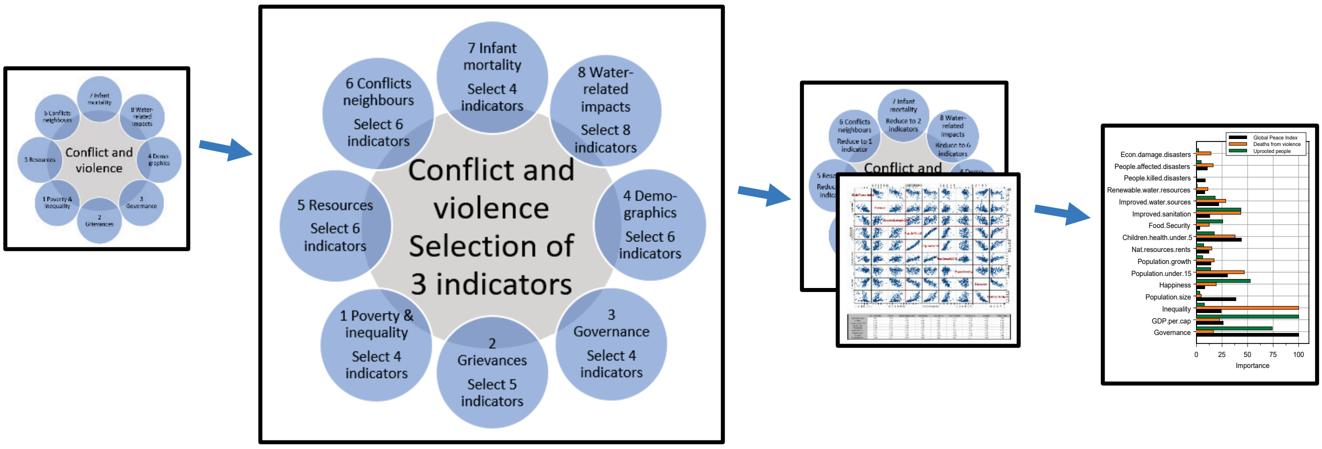


Figure 3.2 Scheme showing the estimation procedure of Random Forests. This Machine Learning approach consists of the estimation of an ensemble of 1000 regression trees which form 'the forest'. Ensemble members differ since the original data set (here conflict data for 210 countries) is *randomly* split into a training set for which a regression tree is estimated (140 countries), and a validation set (70 countries). The regression tree shows which drivers show the strongest relationship to conflict. Here, relationships may be of linear or non-linear nature. The method is able to cope with interrelationships such as shown in the left two panels of Figure A. Next to that the tree is used to predict the conflict data for independent country data (the validation set). As a final step the importance of potential drivers is derived from the total set of 1000 regression trees.

4 Choosing conflict-related indicators

To estimate relationships between conflict and violence indicators on the one hand and potential drivers on the other hand we have to choose dependent and independent variables, as explained in Section 3. A description of three conflict and violence indicators is given in Section 4.1 and a description of regressors in Section 4.2. This is the second stage in the scheme shown in Figure 1.1 and reprinted below.



4.1 Indicators of violence and conflict-dependent variables

The centre of Figure 2.2 is 'Violent conflict'. It is not straightforward which indicator or composite indicator represents this term best and a number of choices can be made. Next to that, indicators for violence and/or violent conflicts will be blurred by a number of uncertainties such as summed up in Chapter 5. We name (i) missing or uncertain data for countries with poor statistical capacity and (ii) the role of propaganda and politics in the supply of this type of data.

Therefore, we propose three indicator choices here, each with advantages and disadvantages. By using these indicators **in parallel** — as dependent variables in Regression Tree models — we make our inferences more robust.

Note 1: as stated in Section 2.2.7 we do not discern between three types of conflict (state-based conflicts, non-state conflicts and one-sided violence).

Note 2: in formulating violence and conflict indicators we will choose indicators such that all countries will have data other than zero. For example, we do not choose 'the number of battle field deaths' as an indicator since many countries are not in conflict and would have zeros. This choice eases statistical inferences.

The Global Peace Index
The first indicator we choose is the Global Peace Index (GPI**)**, derived for the year 2016 by the Institute for Economics and Peace (IEP). The GPI is a composite indicator based on 23 underlying violence-related data and given in Table 2.3 in Section 2.2.6 of the Part-I-report. Its strong side is the wide range of conflict- and violence-related aspects taken into account, varying from perception of criminality, homicide rates, political terror, access to small arms and deaths from internal fighting (measures for 'internal peace') on the one hand, and the number of refugees and IDPs, weapon exports, armed services personnel rate and deaths from external conflicts (measures for 'external peace') on the other hand.

However, the strength of the GPI is a weakness as well. As for all composite indicators, sub-indicators have to be scaled and given weights to compose the final composite, as explained in Section 3.1 of the Part-I-report. The way this is done, is based on expert judgement and contains some form of subjectivity. Another assumption underlying the GPI is that all sub-indicators have an increasing (or decreasing) relation to the level of peacefulness of countries. However, it is not completely clear that such relationship is linear for all 23 sub-indicators. For example, it could be that an increasing number of (nuclear) weapons *increases* peacefulness to a certain level and *decreases* peacefulness if the level of armament crosses a certain threshold, leading to a more U-shaped relationship.

Deaths from violence and conflict
A second indicator we choose, is the death rate by violence and conflict, denoted hereafter as 'Violence'. To compose this indicator, we combined the three UCDP databases on conflict-related deaths and the number of intentional homicides of UNOCD. For a description of these four databases we refer to Section 2.2.7, and Figures 2.6A and 2.6B of the Part-I-report. Conflict-related deaths were calculated as the mean over the years 2011-2016, while homicides are calculated as the mean over the years 2011-2015 (at present, 2016 data are not available). To find death *rates* rather than *absolute numbers* we added the country-averaged number of deaths and divided these data by the population size of each country.

As a final step we to took the logarithm of these data — this to reduce the influence of extreme death rates in statistical analyses — and scaled these figures to a minimum value '0' and a maximum value of '10'.

The strength of this second indicator is its simplicity: it is not a composite indicator such as GPI and its interpretation is straightforward. It is directly based on violence-related data by counting deaths. Next to that, countries with no conflict-related deaths still show differences in homicide rates which is a measure for peacefulness in this group of countries. In this way all countries can be ranked.

Again, the strength of this indicator might be seen as its weakness as well. It's strength is its easy interpretation: deaths from violence and conflicts are a direct outcome of tensions in countries. However, factors such as human rights, presence of weapons, the number of refugees or the relationship to neighbouring countries are not included (as in the GPI). Next to that, Section 2.2.7 of the Part-I-report shows that homicide data are provided by individual countries which use (slightly) different definitions. Furthermore, not all countries provide these data on an annual basis, thus some interpolation is needed to fill in years with no data.

We note that the indicator 'Deaths from violence and conflict' is contained in the Global Peace Index as can be seen from Table 2.3 of the Part-I-report, but only for a small part (13% of the weights can be attributed to conflict-related deaths).

Uprooted people

The third indicator we choose is a measure for the rate of displacement of people due to violence and conflict. This displacement follows from the sum of people who flee the country (refugees) and the number of people who flee but stay within their own country (internally displaced people due to conflicts, or IDPs in short). We will denote this indicator by the term 'Uprooted people'. Data for refugees come from UNHCR (Section 2.2.9) and data on internally displacements come from IDMC (Section 2.2.9 and Figure 2.7). All data are for *new* refugees and *new* IDPs and are for the year 2016. As for the Violence indicator we calculated rates, applied a log-transformation and scaled from '0' to '10'.

Strength and weaknesses of this indicator are similar to the indicator for violence and conflict.

We note that the INFORM database of JRC also contains an indicator for 'uprooted people'. However, their indicator is different from the one defined here. Furthermore, we note that the ‘Uprooted people’ indicator is contained in the Global Peace Index for only 6% as can be seen from Table 2.3 of the Part-I-report.

Indicators compared

A comparison between the GPI, the Violence indicator and the indicator for Uprooted people is given in Figure 4.1. The upper panel shows the scatterplot between GPI and Uprooted people, the lower panel the scatterplot for GPI and Violence. The green line is the one-to-one line if the information in both variables would be identical. The correlation coefficient R for the upper panel accounts for -0.72 and for the lower panel -0.61. Clearly, the information contained in these indicators is partly different. This is easy to understand if we look at the 23 sub-indicators and their weights shown in Table 2.3.

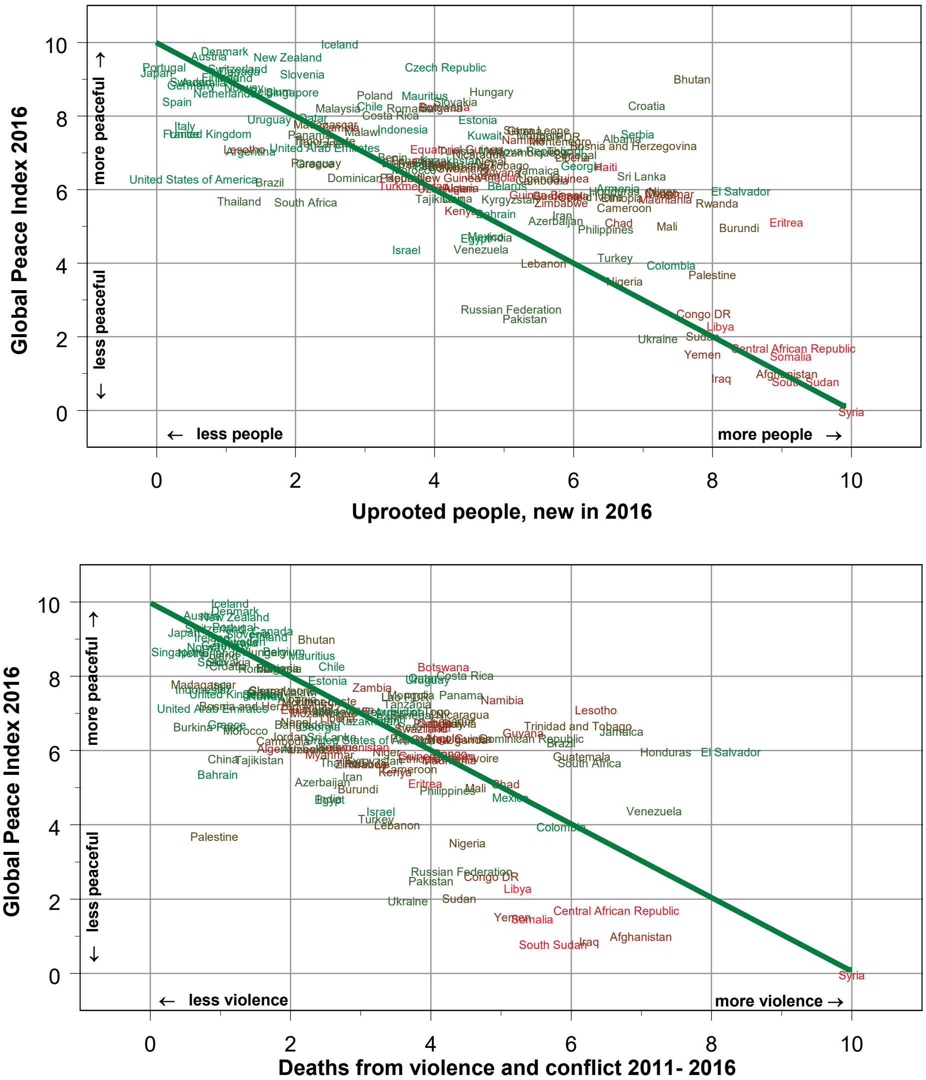


Figure 4.1 Scatterplots showing the relationship between Uprooted people and the Global Peace Index (upper panel), and the relationship between Deaths from violence and conflicts and the Global Peace Index (lower panel). Colours from green to red correspond to the World Bank statistical capacity indicator (red means low capacity). The green lines are the one-to-one lines in case the information contained in the x- and y-variable would be identical.

4.2 Drivers of conflict - independent variables

Given the indicators formulated in Section 10.1 and the "driver carrousel" shown in Figure 2.2, we have chosen a number of indicators which are representative for these drivers. Indicators are summarized in Table 4.1 and refer to databases described in Chapters 2 and 3 of the Part-I-report.

Table 4.1 Overview of independent variables (regressors), according to the categorization shown in Figure 2.2. Section numbers refer to the Part-I-report.

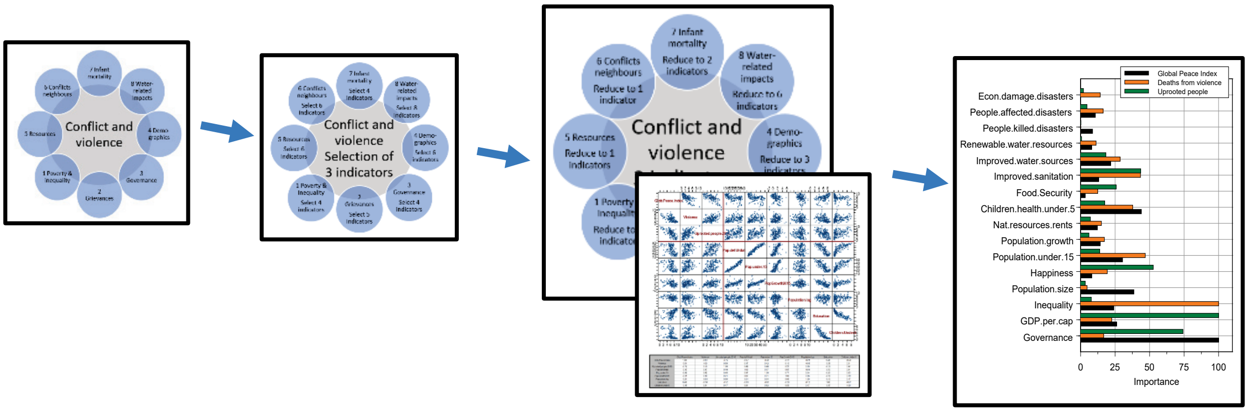
Driver group	**Variable 1**	**Variable 2**	**Variable 3**	**Variable 4**	**Variable 5**	**Variable 6**
1 Economic inequality and poverty	GDP PPP per capita from the JRC INFORM database 2016 (Section 2.2.10)	Percentage of people living on 1.9 dollar a day at 2011 prices (see Section 2.2.3 and Appendix A #12)	Poverty according to the Multiple Poverty Index (Section 2.2.5)	Economic and gender inequality, a composite indicator from the JRC INFORM database 2016 (Section 2.2.10)		
2 Grievances and discrimination	Happiness, a perception indicator for the year 2016 (Section 2.2.8.	Unemployment (World Bank indicator, 2016, see Section 2.2.3 and Appendix A #17)	Group grievances (Fragile States, Section 3.2.3)	Human rights and rule of law (Fragile States, Section 3.2.3)	Factionalized elites (Fragile States, Section 3.2.3)	
3 Effective governance and corruption	Corruption Perceptions Index 2016 from Transparency International (Section 2.2.4)	Effective governance, taken from the World Bank set of governance indicators (Section 2.2.4)	Governance for the year 2016 (INFORM database, Section 2.2.10). This indicator is a composite of the first two indicators	The Fragile States Indicator (Section 3.2.3)		
4 Demographics and education	Youth bulge indicator computed as the number of people between 15 and 24 years of age, relative to the number of people being 15 years of age and older	Youth bulge indicator computed as the number of people under 15 years of age, relative to the total population (UN Population Division, Section 2.2.5)	Population growth as annual percentage for the year 2016 (World Bank indicator, Section 2.2.3 and Appendix A #11).	Population size 2016 (World Bank indicator).	Level of education, taken from the UN-DP Human development report 2016 (Section 2.2.5).	

	(UN Population Division, Section 2.2.5)					
5 Availability of resources	Energy import/export (Word Bank, Section 2.2.3, Appendix A #3)	Ores and metal exports (Word Bank, Section 2.2.3, Appendix A #10)	Total natural resources rents (Word Bank, Section 2.2.3, Appendix A #16)	Forest as percentage of the total area of a country (Word Bank, Section 2.2.3, Appendix A #4)	Agricultural land as percentage of the total area of a country (Word Bank, Section 2.2.3, Appendix A #1)	Renewable water per capita (Word Bank, Section 2.2.3, Appendix A #15)
6 Conflicts in neighbouring countries	Mean GPI of surrounding countries (Section 9.1)	Minimum GPI in one of the surrounding countries (Section 9.1)	Mean Violence in surrounding countries (Section 9.1)	Max. Violence in one of the surrounding countries (Section 9.1)	Mean number of Uprooted people in surrounding countries (Section 9.1)	Max. number of Uprooted people in one of the surrounding countries (Section 9.1)
7 Infant mortality and malnutrition	Mortality in children under 5 years of age (JRC INFORM database 2016, Section 2.2.10)	Malnutrition in children under 5 years of age (JRC INFORM database 2016, Section 2.2.10)	Health of children under 5 years of age (JRC INFORM database 2016, Section 2.2.10)	Food safety		
8 Water-related impacts and climate change	Improved water sources (Word Bank, Section 2.2.3, Appendix A #5)	Improved sanitation facilities (Word Bank, Section 2.2.3, Appendix A #6)	Rate of people affected by water-related disasters (CRED database, Section 2.2.1)	Rate of people killed by water-related disasters (CRED database, Section 2.2.1)	Economic damage due to water-related disasters, relative to GDP PPP (CRED database, Section 2.2.1)	Renewable water per capita (Word Bank, Section 2.2.3, Appendix A #15)
	Aridity Index (PBL IMAGE database)	Agricultural rents as perc. of GDP (World Bank database, Section 2.2.3)				

5 Drivers of conflict: results

We perform the statistical analysis in two stages. First, we analyse the data described in Table 4.1 by visual presentation of the data along with the calculation of correlation matrices (Section 5.1 and Appendix A). Second, we apply the Random Forest approach to a selection of drivers (Section 5.2). This pre-selection of regressors prior to modelling is needed (i) to reduce the complexity of the models, and (ii) to reduce the problem of highly correlated regressors leading to multicollinearity. See Kuhn and Johnson (2016, Section 3.4) for more details. A third argument comes from the fact that several indicators summarized in Table 4.1, suffer from a large number of missing data. Since the Random Forest approach only works for those countries which have data for *all* regressors in the model, a (great) number of countries would be excluded. Thus, a reduction in drivers might help here.

Both stages are on the right of the scheme below.

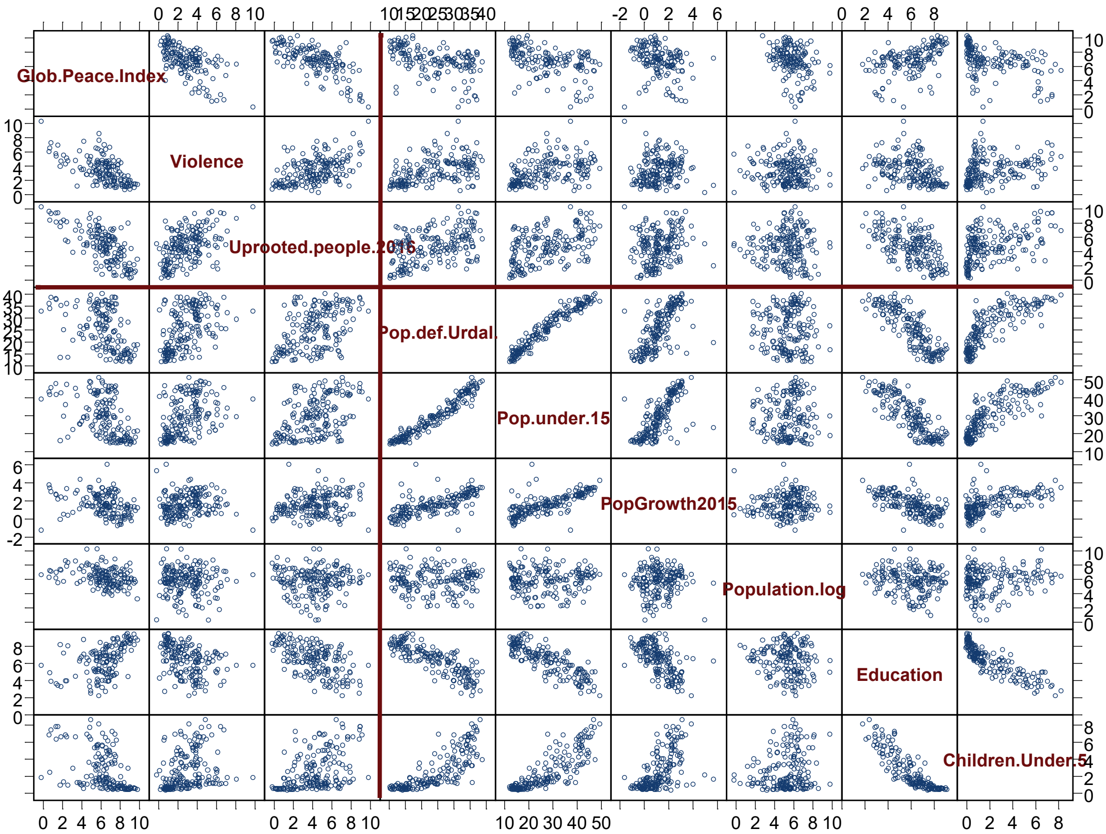


5.1 Pre-selection of data

As a first stage in the modelling approach we analyse relationships per factor by a visual presentation of data, along with correlation matrices. Relevant indicators are given in Table 4.1 which refer to the drivers shown in Figure 2.2. Based on the results and inferences, we choose a smaller number of indicators, if necessary.

Results from this first stage are summarized in Appendix A. Here, we give an example for factor 4 'Demographics and Education'. To find associations between GPI, Violence and Uprooted people on the one hand and demographic indicators on the other, we selected the following (composite) indicators: (1) a youth-bulge indicator computed as the number of people between 15 and 24 years of age, relative to the number of people of 15 years of age and older, as proposed by Urdal (2011), (2) a youth-bulge indicator computed as the number of people under 15 years of age, relative to the total population, (3) population growth as annual percentage for the year 2016, (4) population size, (5) an indicator for the level of education, and (6) an indicator for malnutrition and mortality of children under 5 years of age. See Table 4.1, factor 4. We note that the sixth indicator on child health is also part of factor 7.

The scatterplot matrix and corresponding correlation matrix are given in Figure 5.1A. The correlation matrix shows intermediate correlations to the three youth-bulge related indicators, with the highest value between GPI/Uprooted people and the percentage of people under 15 years of age. Correlations lie in the range of $0.42 \leq |R| \leq 0.49$. The relationships with population growth or population size is lower, in the range of $0.0 \leq |R| \leq 0.31$. The Education index shows intermediate correlations, in the range of $0.38 \leq |R| \leq 0.56$. Finally, the index for malnutrition and mortality among children under 5 years of age shows intermediate correlations: $0.31 \leq |R| \leq 0.48$.



	Glob.Peace.Index	Violence	Uprooted.people.2016	Pop.def.Urdal.	Pop.under.15	PopGrowth2015	Population.log	Education	Children.Under.5
Glob.Peace.Index	1.00	-0.61	-0.72	-0.47	-0.48	-0.31	-0.21	0.48	-0.48
Violence	-0.61	1.00	0.50	0.47	0.42	0.10	-0.03	-0.38	0.31
Uprooted.people.2016	-0.72	0.50	1.00	0.49	0.48	0.21	0.00	-0.56	0.47
Pop.def.Urdal.	-0.47	0.47	0.49	1.00	0.97	0.67	-0.01	-0.83	0.81
Pop.under.15	-0.48	0.42	0.48	0.97	1.00	0.71	0.04	-0.83	0.83
PopGrowth2015	-0.31	0.10	0.21	0.67	0.71	1.00	0.06	-0.60	0.59
Population.log	-0.21	-0.03	0.00	-0.01	0.04	0.06	1.00	-0.12	0.17
Education	0.48	-0.38	-0.56	-0.83	-0.83	-0.60	-0.12	1.00	-0.87
Children.Under.5	-0.48	0.31	0.47	0.81	0.83	0.59	0.17	-0.87	1.00

Figure 5.1A Scatterplot matrix for three violence/conflict indicators, five indicators with a demographical character, and one indicator for education. The lower panel shows the correlation matrix for these variables.

Next to these inferences, the scatterplot matrix shows important relationships between indicators 'the percentage of people under 15 years of age', 'annual population growth', 'level of education' and 'malnutrition and mortality of children under the age of 5'. All these factors are mutually highly correlated with values in the range of $0.59 \leq |R| \leq 0.87$, where $R = -0.87$ is for education and malnutrition/mortality of children under the age of 5.

These high correlations correspond to what is denoted in the literature as the *poverty trap*: in countries with high levels of poverty and infectious diseases, mortality under children will be high and the birth-rate of women will stay high. As a consequence, poverty will increase more and possibilities for education will shrink. As a result, child mortality and birth rates might increase even more (e.g. Sachs et al., 2001).

The strong relationship between the percentage of people under the age of 15 on the one hand, and population growth on the other hand, follows from Figure 5.1B.

Based on these results we selected all variables shown in Figure 5.1A as input for the Random Forest analyses, except the youth bulge indicator following the approach of Urdal.

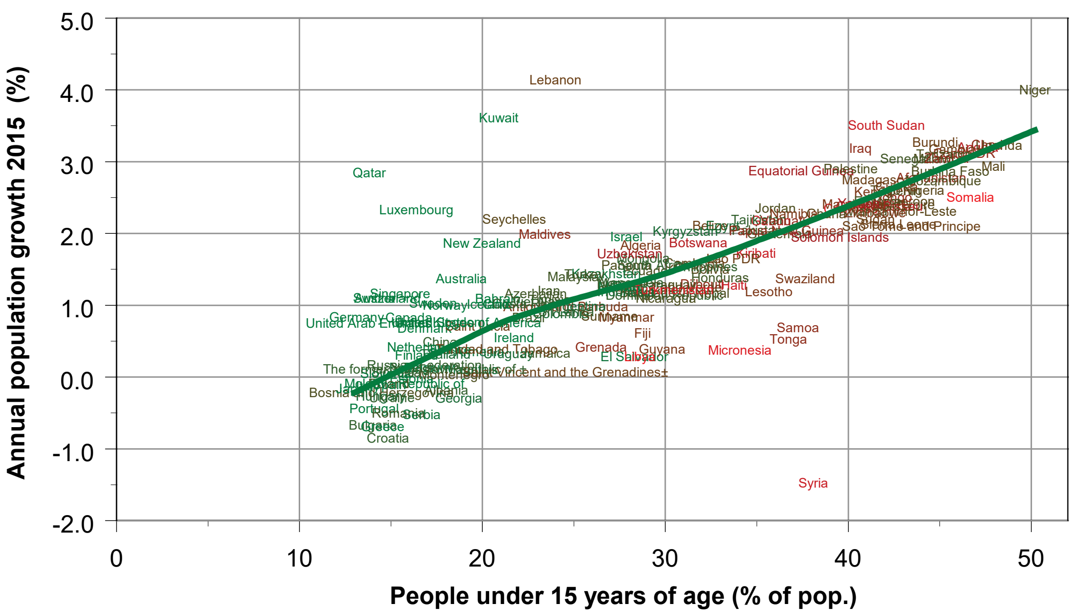


Figure 5.1B The relationship between annual population growth for the year 2015 and the percentage of people under the age of 15. The trend is estimated by the LOESS routine. Colours from green to red correspond to the World Bank statistical capacity indicator (red means low capacity).

5.2 Random Forest analyses

In the preliminary phase we have studied relationships in a single-driver (univariate) setting. However, relationships may be conjunctural, that is: an independent variable X_1 may show a weak relationship to a dependent variable Y, while its importance may change in combination with another variable X_2. Therefore, we apply Random Forest (RF) analysis to find an importance hierarchy in variables X_1, ... , X_N, as explained in Section 3.3.

We selected the following set of independent variables (regressors) which showed intermediate to high correlations in the first-stage analysis (Section 5.1 and Appendix A). These variables are a subset of those given in Table 4.1:

- GDP per capita PPP, percentage of people living on less than 1.9 dollar a day (factor 1),
- Happiness, unemployment, economic and gender inequality (factor 2)
- Governance as a combination of effective governance and corruption (factor 3),
- Youth bulge taken as the percentage of people under 15 years of age, populations size and population growth as percentage per year, level of education (factor 4),
- Agricultural land as percentage of total area, idem forestry, energy imports, total natural resources rents, renewable water per capita (factor 5),
- Health of children under 5 years of age, and food safety (factor 7),
- Improved access to water sources, improved sanitation, rate of people affected by water-related disasters, idem people killed, idem economic damage as percentage of GDP, aridity index (factor 8).

The role of conflicts and violence in neighbouring countries (factor 6) will be treated separately.

Since our aim is to analyse a common set of drivers in relation to each of the conflict indicators, we further reduced the set of drivers **to a common set of 16 drivers** which are relevant to each aspect of conflict and violence. Therefore, a number of indicators which showed lower correlations are omitted in the following analysis (e.g., education, aridity, agricultural land as percentage of total area). We note that these indicators need not to be irrelevant and we tested their role in Random Forest models not described in this report.

The Global Peace Index

We start to analyse the Global Peace Index as dependent variable Y_i and the variables $X_{1,i}$, … , $X_{23,i}$ mentioned above. The method we follow has been explained in Section 3.3 and by a simulation example shown in Appendix B. The importance functions estimated by the Random Forest algorithm, are shown in Figure 5.2A.

We find the following ranking of drivers: (1) Governance, (2) Health of children under the age of 5, (3) Population size, (4) Population under the age of 15 (youth bulge indicator), and (5) Income and gender inequality. Next to these top rankings it is interesting to note that several water-related indicators have only a minor importance. We name the three impact indicators from water-related disasters and the indicator for Renewable internal fresh water resources per capita (upper four drivers in Figure 5.2A).

The relationship between the Global Peace Index and Governance is shown in more detail in Figure 5.2B. The graph shows that the relationship between both variables is non-linear, with a tipping point around Governance values of 2.5.

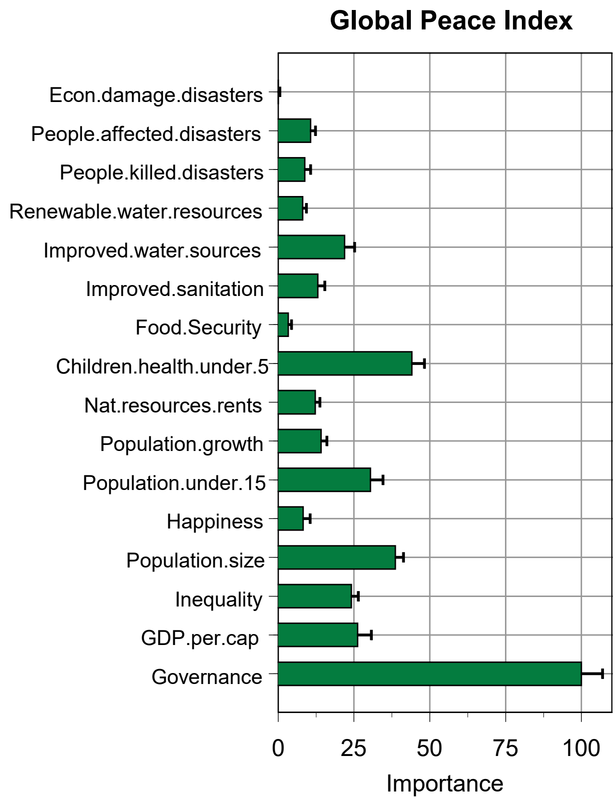


Figure 5.2A Importance of independent indicators for the Global Peace Index as dependent variable, based on 300 regression trees. Explained variance: **49%**. Analysis is for 129 countries which have data for all 16 indicators considered.

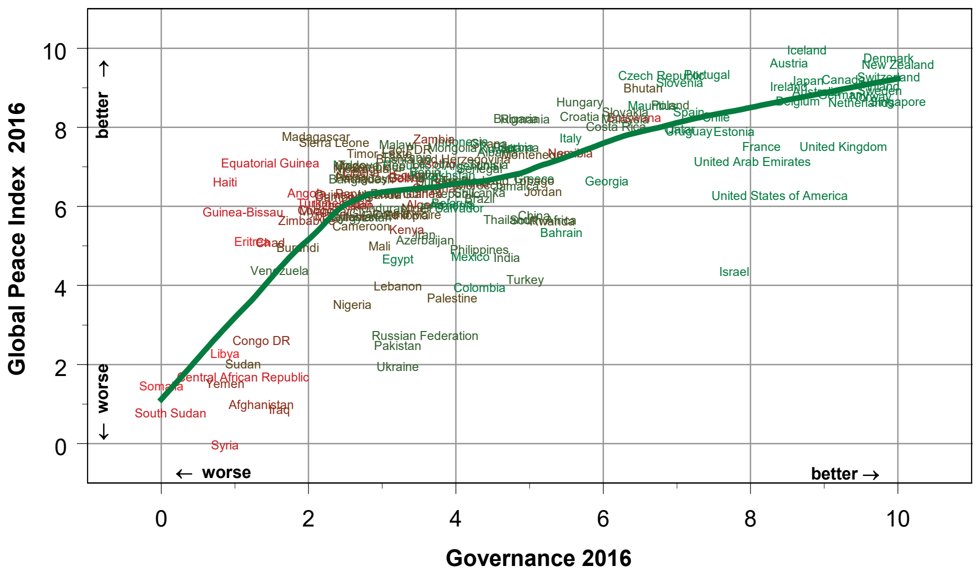


Figure 5.2B The relationship between the Global Peace Index (2016) and Governance for 157 countries. The trend is estimated by the LOESS routine. Colours from green to red correspond to the World Bank statistical capacity indicator (red means low capacity). Note that the number of countries in Figure 5.2A is lower due to missing values in the full set of regressors.

Due to the set-up of the Random Forest procedure we estimated for each country around 100 true predictions (originating from 300 regression trees which were split into a training set of 2/3 of the countries available). We have plotted these predictions in Appendix D (Figure D.1A). This plot is important since we can see that several poor countries are not predicted very well. These are Syria, Afghanistan, Iraq, Pakistan and Ukraine. Note that not all (poor) countries are shown in this graph due to missing values in the regressors (such as Somalia or South Sudan).

We repeated the analysis above where missing data in driver indicators were imputed using the so-called proximity matrix, contained in the Random Forest approach (Section 3.3). The number of countries now increases from 129 to 157. Figure D.1B shows that the prediction performance is slightly better (explained variance increases from 49% to 56%). The importance function is identical to that shown in Figure 5.2A.

Deaths from violence and conflict

The importance functions estimated with the Random Forest algorithm, are shown in Figure 5.3A. If we combine both functions, we find the following ranking: (1) Inequality, (2) Population under the age of 15 (also denoted as 'youth bulge'), (3) Percentage people with

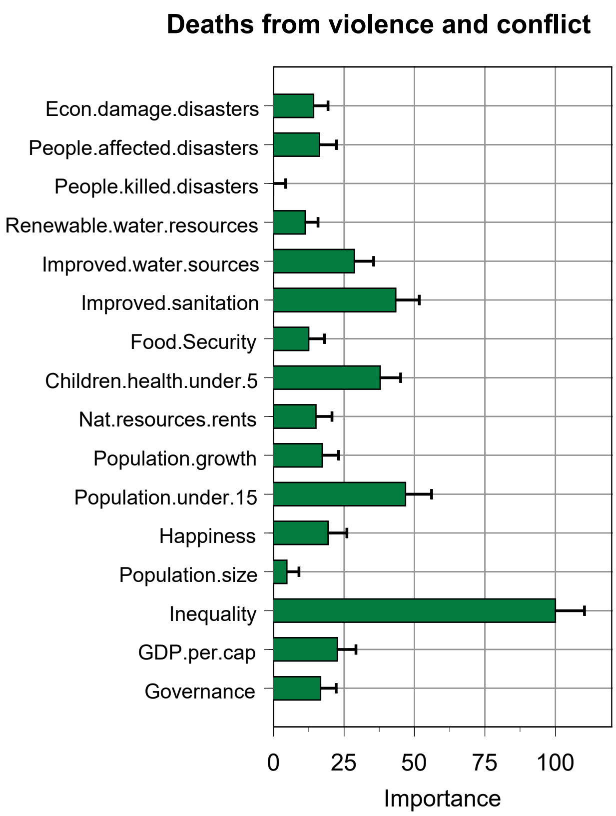


Figure 5.3A Importance of independent indicators for Violence as dependent variable, based on 300 regression trees. Left method is 'Incremental MSE' and left method is 'Incremental node purity'. Explained variance: **42%**. Analysis is for 134 countries which have data for all indicators considered.

improved sanitation and (4) Health of children under the age of 5. Next to these top rankings it is interesting to note that a few water-related indicators have only a minor importance. We name the three impact indicators from water-related disasters and the indicator for Renewable internal fresh water resources per capita. Also, the importance of Population size is small.

The relationship between deaths by violence and conflict on the one hand and Inequality on the other hand, is shown in more detail in Figure 5.3B. The graph shows that the relationship between both variables is linear. The importance function is identical to that shown in Figure 5.3A.

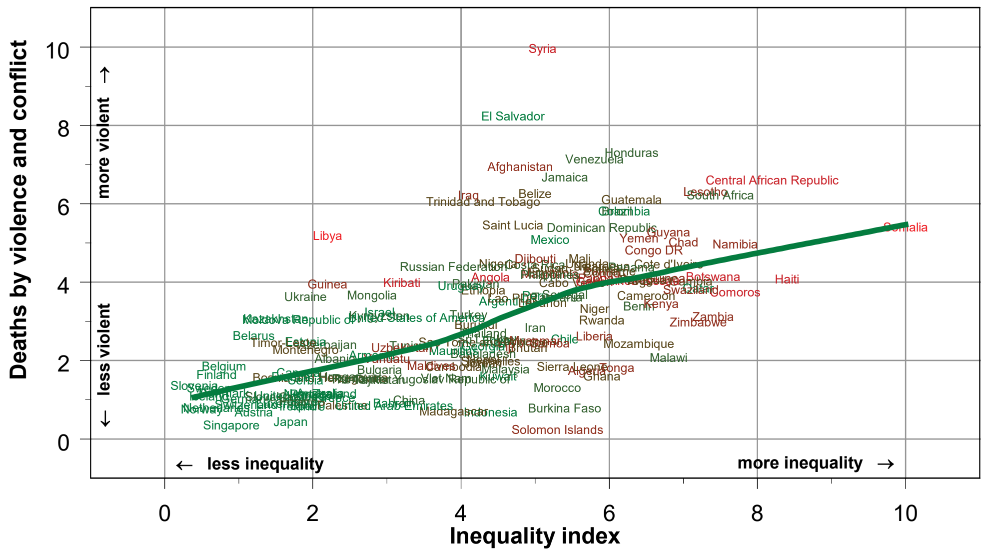


Figure 5.3B The relationship between the Inequality (2016) and the indicator for deaths by violence and conflict (period 2011-2016), for 173 countries. The trend is estimated by the LOESS routine. Colours from green to red correspond to the World Bank statistical capacity indicator (red means low capacity). Correlation coefficient R accounts for 0.60. Note that the number of countries in Figure 5.3A is much lower due to missing values in the full set of regressors.

Due to the set-up of the Random Forest procedure we estimated for each country around 100 true predictions (originating from 300 regression trees which were split into a training set of 2/3 of the countries available). We have plotted these predictions in Appendix D (Figure D.2A).

Again, this plot is important since we can see that several poor countries are not predicted very well. Examples are El Salvador, Afghanistan, Iraq and Trinidad and Tobago. Note that not all (poor) countries are shown in this graph due to missing values in the regressors (such as Somalia, South Sudan or Syria).

We repeated the analysis above where missing data in driver indicators were imputed using the proximity matrix (Section 3.3). The number of countries now increases from 134 to 188. Figure D.2B shows that the prediction performance equals that of that shown in Figure D.2A. The importance function is identical to that shown in Figure 5.3A.

Uprooted people

The importance functions estimated with the Random Forest algorithm, are shown in Figure 5.4A. If we combine both functions, we find the following hierarchy: (1) GDP per capita, (2) Governance, (3) Happiness and (4) Percentage of people with improved sanitation. Next to these top rankings it is interesting to note that a few water-related indicators have only a minor importance. We name the three impact indicators from water-related disasters and the indicator for Renewable internal fresh water resources per capita. Also, the importance of Inequality and Population size is small.

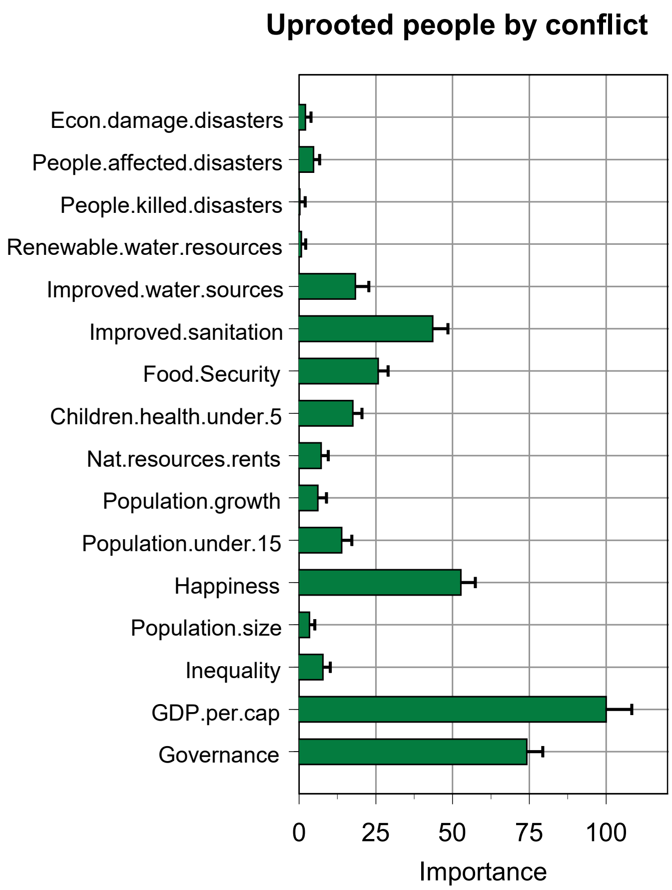


Figure 5.4A Importance of independent indicators for Uprooted people as dependent variable, based on 300 regression trees. Left method is 'Incremental MSE' and left method is 'Incremental node purity. Explained variance: **46%**. Analysis is for 134 countries which have data for all indicators considered.

The relationship between Uprooted people and GDP per capita PPP is shown in more detail in Figure 5.4B. Here, the relationship between both variables is more or less linear.

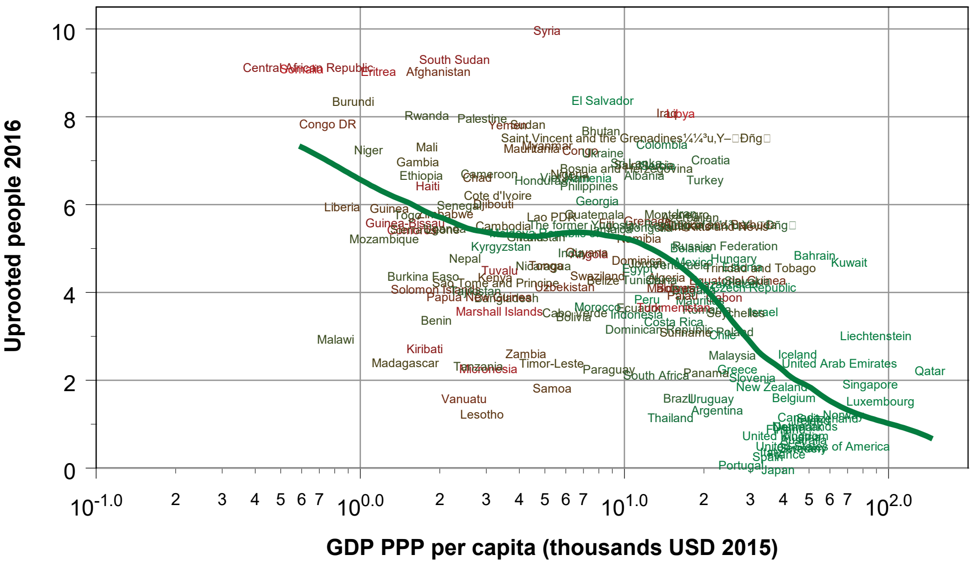


Figure 5.4B The relationship between the GDP per capita and the indicator for Uprooted people, for 191 countries. The trend is estimated by the LOESS routine. Colours from green to red correspond to the World Bank statistical capacity indicator (red means low capacity). Note that the number of countries in Figure 5.4A is much lower due to missing values in the full set of regressors.

Due to the set-up of the Random Forest procedure we have for each country around 100 true predictions (originating from 300 regression trees which were split into a training set of 2/3 of the countries available). We have plotted these predictions in Appendix D (Figure D.3A).

We repeated the analysis above where missing data in driver indicators were imputed using the so-called proximity matrix (Section 3.3). The number of countries now increases from 134 to 191. Figure D.3B shows that the prediction performance is slightly better. The importance function is identical to that shown in Figure 5.4A.

Role of neighbouring countries
To check the role of tensions and conflicts in neighbouring countries, we re-estimated the trees shown in Figures 5.2A, 5.3A and 5.4A with one extra variable: the mean conflict level in neighbouring countries (factor 6 in Figure 2.2). As for the Global Peace Index this extra variable has minor importance. Thus, tensions in neighbouring countries are not found to be a leading factor here. However, for the other two violence indicators — Deaths by violence and Uprooted people — this extra variable has a profound influence and becomes the variable with dominant importance. See Figure 5.5 for Uprooted people.

We note that scatterplot between violence indicators and violence indicators in neighbouring countries are shown in Section A.6 of Appendix A.

As for the last two violence indicators these findings correspond to the prediction models presented by Hegre et al. (2013, 2016) and Goldstone et al. (2010), given in Table 2.1.

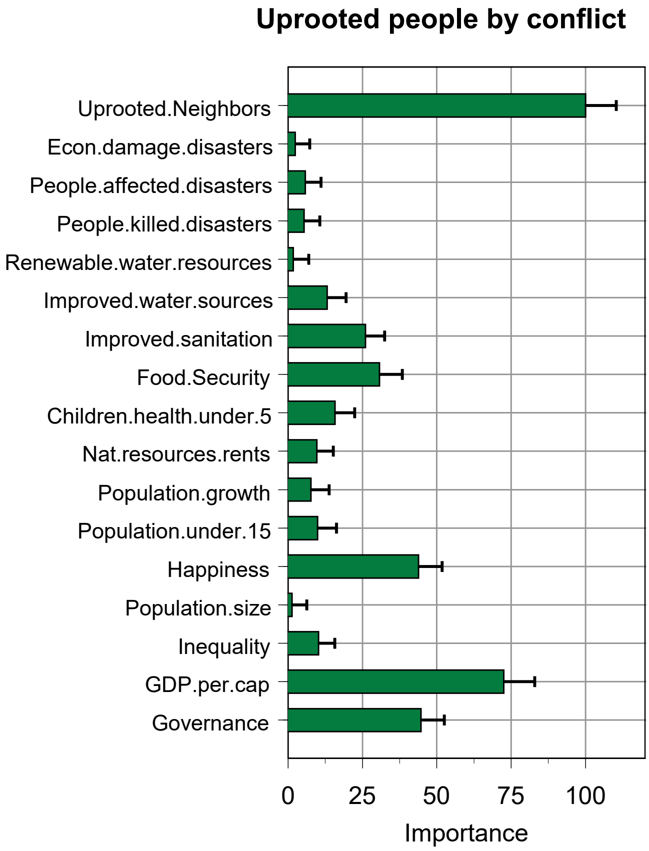


Figure 5.5 Importance of independent indicators for Uprooted people as dependent variable, based on 300 regression trees. Left method is 'Incremental MSE' and left method is 'Incremental node purity. Explained variance: **52%**. Analysis is for 121 countries which have data for all indicators considered.

6 Discussion

6.1 Ranking conflict drivers and interpretation

If we summarise the importance of explanatory indicators across all three random forest models, we come to the hierarchy summarized in the left panel of Figure 6.1 (cf. Figures 5.2A, 5.3A and 5.4A). Clearly, Governance, GDP per capita and Inequality are the main factors found in these three analyses, although differences per conflict indicator are large.

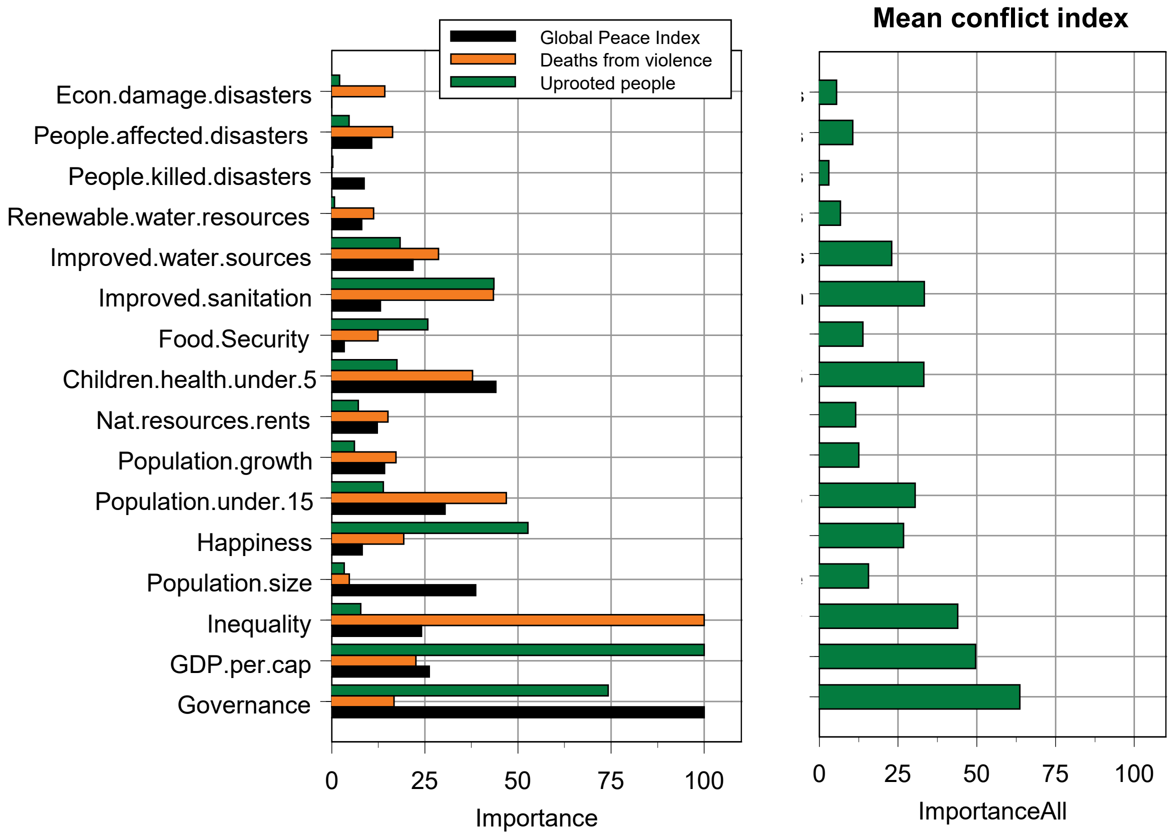


Figure 6.1 Importance functions for three conflict indicators combined. Left panel: combination of graphs 5.2A, 5.3A and 5.4A. Right panel: importance function averaged over conflict indicators shown in the left panel.

To arrive at a concluding hierarchy in order to properly interpreted the results, we have averaged the scores of all three conflict indicators, as shown in the right panel of Figure 6.1. The most influential indicators are (1) Governance, (2) GDP per capita PPP and (3) gender and economic Inequality. Population under the age of 15 (‘youth bulge’), Health of children under the age of 5, People with improved sanitation and (un)Happiness are tied equal at the fourth position.

The hierarchy shown in Figure 6.1 can be interpreted and labelled in terms of structural and proximate drivers. To do so we have rearranged the findings shown Figure 6.1. As structural factors we propose governance and corruption, GDP per capita and inequality. These factors show the strongest though differing relationships to conflict indicators. Proximate factors are influenced by the structural factors.

For example, bad sanitation is caused by a lack of good governance, income and gender inequality, not the other way around. Following the same reasoning, the health of children under the age of 5 is influenced by the structural factors, and not the other way around. Drivers with weaker associations — such as Food Security, Natural resources rents, aridity and levels if education — could be added to the group of proximate drivers.

The proximate factors, in combination with the structural factors, eventually determine how people feel, to what extent they are (un)happy, feel grievances and discrimination (Stohl et al., 2017 – Chapter 2). The hierarchy in terms of structural and proximate conditions is shown in Figure 6.2.

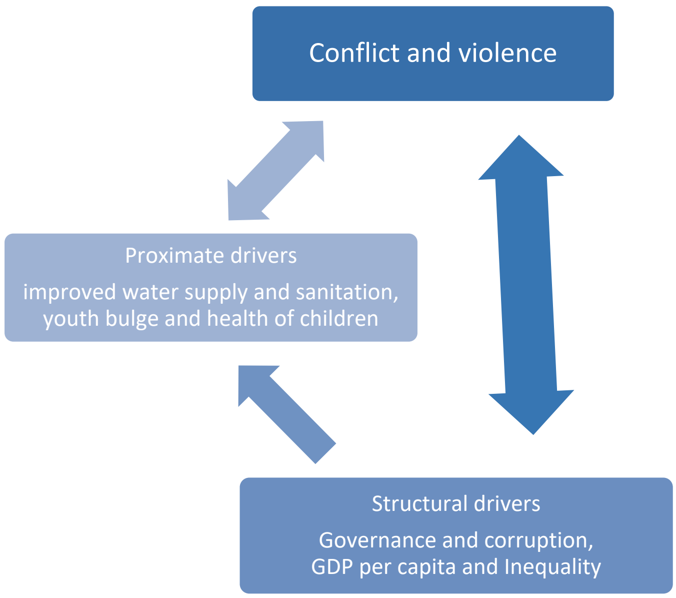


Figure 6.2 Driver hierarchy interpreted in terms of structural and proximate factors/conditions. Indicators correspond to those named in Figure 6.1.

Although we conceptualize governance, GDP per capita and inequality here as equal worthy structural factors, it has to be noted that governance may be seen as the most important structural factor (cf. the right panel of Figure 6.1). Five out of the six studies used for this analysis (Goldstone, Fearon and Latin, Besley and Persson, Collier and Hoeflier, Bara) mention governance and the importance of institutions as a stirring factor in conflict. Additionally, we modelled Governance, as dependent variable, in relation to the drivers summed up in Table 4.1. It shows that Governance is tightly related to these drivers (explained variance of the Random Forest model accounts for 79%). See Figures D.4 and D.5 for details.

6.2 Relationship to findings in the literature

Results compared to the literature summarized in Table 2.1

If we compare the results shown in Figures 6.1A and B with those presented in Table 2.1, we find that this analysis mirrors the conflict literature well. **Governance** — our number one explanatory factor — is named by Goldstone et al. (2010), Fearon and Laitin (2003), Besley and Persson and by Collier and Hoeffler (2004). As for the last, we interpret a lack of political rights as being close to low levels of governance. Thus, our statistical result coincides very well with the factor mentioned most often in Table 2.1. **GDP per capita** is only named by Fearon and Laitin, be it in terms of ‘poverty and slow economic growth’. **Inequality** is named by Collier and Hoeffler (2004) as an important driver.

Furthermore, **the health of children under the age of 5** (which indicator is a combination of child mortality and malnutrition, taken from the INFORM database) is named by Goldstone et al. (2010) and Hegre et al. (2013), be it in terms of ‘infant mortality’. Political and economic discrimination as named by Goldstone et al. lies close to the interpretation of **happiness/unhappiness**. The same holds, to some extent, to the role of 'bad neighbourhoods' as mentioned by Bara (2014).

It is also worth mentioning that, vice versa, several drivers mentioned in Table 2.1, are ranked low in the Random Forest analyses — as summarized in Figure 6.1A — or were omitted in the preliminary analyses (Section 5.1). We mention the dependence on oil and ores (Hegre et al. 2013; Collier and Hoeffler, 2004; Bara, 2014) which is close to our indicator **natural resources rents** (taken from the Word Bank indicator database). Thus, here our results deviate from three of the authors mentioned in Table 2.1.

However, it is important to note that this variable can act as threat multiplier or may play an important role on a local scale which might level off on a national scale. For example, one study finds a positive impact of mining on conflict *on local levels* in Africa (Berman et al., 2017). These authors combined georeferenced data on mining of 14 minerals with information on conflict events at a spatial resolution of 0.5° x 0.5° for Africa between 1997 and 2010. Their results suggest that the historical rise in mineral prices might explain in part the level of violence across countries over the period.

In terms of ultimate and proximate drivers, mining activities will lie at the proximate side (cf. Section 3.1). We come to this point in more detail in the next section.

Furthermore, we find that **all three indicators on water-related disasters** (floods, droughts, tsunamis) and **renewable water sources** show the lowest rankings in Figure 6.1A and 6.1B. Apparently, these indicators are not of importance to explain conflict risks, at least for data on national scales. This finding corresponds to the absence of water-related drivers in Table 2.1.

Finally, we found that **conflict and violence in neighbouring countries** influences two out of three conflict indicators. This corresponds to the findings of Goldstone et al. (2010) and Hegre et al. (2013), and this also reflects a study by Buhaug and Gleditsch (2008). Although we did not analyse why a spill-over effect can be observed, Buhaug and Gleditsch find that conflict is more likely in places where ethnic groups are tied over boundaries. They also find that contagion is primarily a feature of separatist conflict.

Results compared to the Positive Peace framework of IEP

It is interesting to compare our findings with the Positive Peace framework developed by the Institute for Economics and Peace (IEP, 2018). The Institute for Economics and Peace defines *positive peace* as the attitude, institutions, and structures that create and sustain peaceful societies. All drivers found in the Random Forest analyses are (more or less) present in their framework consisting of eight pillars, reprinted in Figure 6.3.

The left four pillars — from 'Acceptance of the Rights of Others' to 'Well Functioning Government' — relates to Governances, 'Equitable Distribution of Resources' relates to Inequality, 'High Levels of Human Capital' relates to GDP per capita, and 'Good Relations with Neighbours' relates to Conflict and violence in neighbouring countries (cf. Figure 5.5). Only 'Free Flow of Information' is not easy to couple to our results, although it could be argued that it can be coupled to 'equality', since the absence of inaccessibility of reliable information diminishes social inclusion.

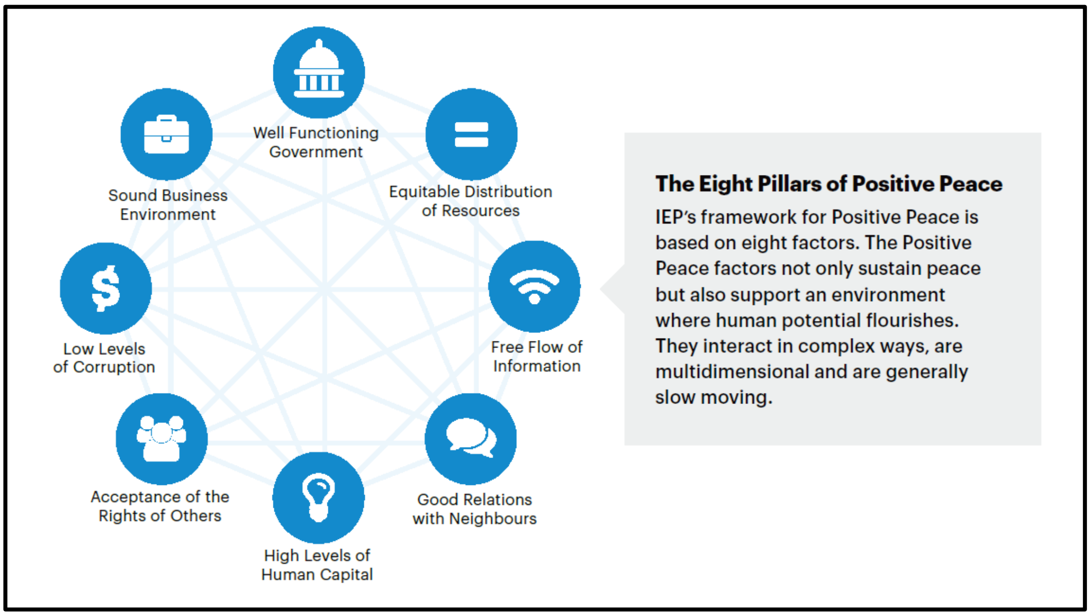


Figure 6.3 Positive Peace framework of the Institute for Economics and Peace. Source: IEP, 2018.

Results compared to the expert elicitation presented by Mach et al. (2019)

Mach et al. (2019) present research findings on the contested relationship between climate and conflict, based on the structured judgments of 11 experts from diverse disciplines. The group of experts is a sample of the most experienced and highly cited scholars on the topic.

The experts agree that climate has affected organised armed conflict within countries. However, they conclude that other drivers are judged to be substantially more influential. In the first step 16 potential conflict drivers were selected and in a second step these selected drivers were ranked by the expert team as for the strength of its influence and the uncertainty in this judgement. These judgements are summarized in Figure 6.4 (reprinted from the Nature publication). The uncertainties around their influence estimates are also given.

Do the most and least influential drivers found here correspond to those found in their study? Figure 6.4 shows that the most influential drivers are 'Low socioeconomic development', 'Low state capabilities' and 'Intergroup inequality'. Drivers that are judged as the least influential, are 'climate variability and/or change', 'physical geography' and 'vertical inequality'. It shows that the uncertainty estimates for 'climate' are highest among these 16 drivers.

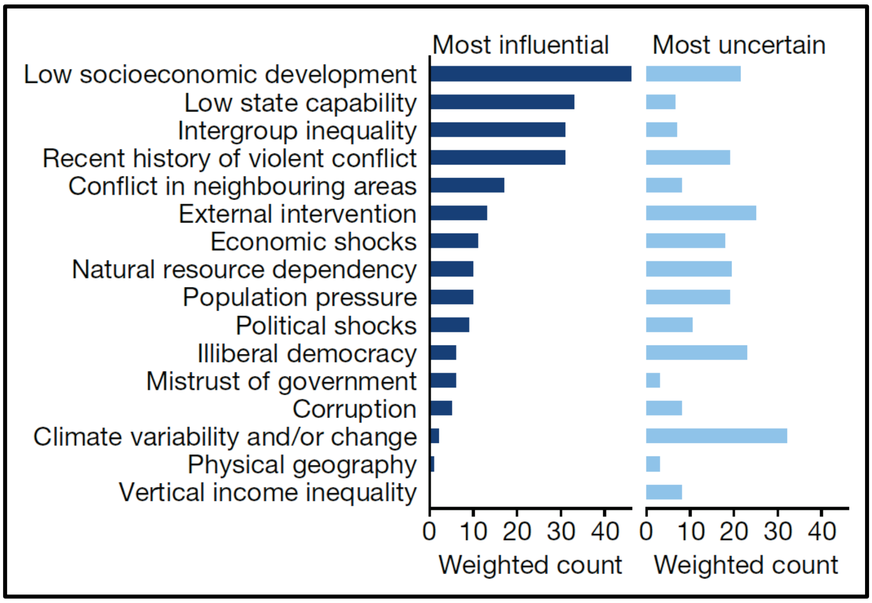


Figure 6.4 Drivers of conflict risk as identified by a group of 11 experts. Source: Mach et al. (2019, their Figure 3a).

How do these results, gained by expert judgements, relate to the findings based on Machine Learning in combination with the driver framework shown in Figure 2.2? Since the short descriptions of drivers in Figure 6.4 do not correspond 1-to-1 to descriptions given in this report, we relate findings using the driver descriptions given by Mach et al. in their Supplementary Table 1 (partly reprinted in Appendix E).

It shows from the first column of Table E.1 that 'socioeconomic development' may be interpreted as 'GDP per capita' (among other measures). GDP per capita is the second important indicator in our Random Forest estimates, summarized in Figure 6.1. The second variable 'low state capabilities' stands for (i) nations with ineffective or absent formal institutions or (ii) weak states, and can reasonably be interpreted as nations with low levels of governance. This is the most important driver in our Random Forest estimates.

'Intergroup inequality' deals with inequality among different groups and can originate from political, economic, ethnic and social dimensions according the definition given in Table E.1. This driver corresponds only partly with the inequality indicator chosen in this study (a combination of economic and gender inequality). However, both indictors rank at the third position as for their influence/importance.

The role of climate, be it as variability or change, is not explicitly considered in our study. We focused on water-related indicators which partly coincide with 'climate' (impacts of weather-related disasters). Both approaches find a limited and/or highly uncertain role for this indicator.

Figure 6.4 also shows the importance of a recent history of violent conflicts, and conflicts in neighbouring countries. The last factor is also identified in Section 5.2 and Figure 5.5. The role of recent history of violence is treated in Section 2.3, item #13 and Figure 2.4. The role of 'external intervention' is also treated in Section 2.3, item #9.

6.3 How well do models describe the data?

Although it is not our intention to predict conflict outbreaks, Random Forest models allow for an evaluation of true conflict-risk predictions for all countries in the model. That is due to the principle of Bootstrap Aggregation explained in Section 3.3. For each country we have around 100 true predictions for the conflict indicator at hand (we estimated 300 Regression Trees where one third of the countries is not used for model fitting). By looking at these predictions we can evaluate how well the model describes the data at hand.

The prediction performance of the Random Forest models is shown in Appendix D. A scatterplot for each of the three conflict indicators — Global Peace Index, Deaths by conflict and violence and People uprooted by conflict and violence — is shown in the graphs D.1, D.2 and D.3, respectively.

All graphs show the same pattern: predictions deteriorate if the conflict Index values show values lower than the threshold of 4.0 (Figures D.1A and B), higher than 5.0 (Figures D.2A and B) and higher than 6.0 (Figures D.3A and B). In all three cases predictions **underestimate** the true national conflict value.

A second observation is that countries with large prediction errors are mainly, but not exclusively, those with low statistical capacities (the countries in red colours in Figures D.1, D.2 and D.3). Examples are Syria, Afghanistan, Iraq, Congo DR, Burundi, Trinidad and Tobago, Palestine and Lesotho. Exceptions are El Salvador, Israel, South Africa and Thailand (countries in green colours).

We note that prediction errors improve in two cases if we interpolate missing data in drivers before the estimation of the final Random Forest model (Figures D.1B and D.3B).

As for the first group of countries (large prediction errors and low statistical capacity) a range of uncertainties — as summed up in the Part-I-report (Chapter 5) — could well explain the pattern. Next to that, factors not considered, such as the proxy character of many present-day conflicts, may play a role (factor #9 in Section 2.3).

A third observation is that **variance explained** by the Random Forest models lies around 46%, which we judge as not particular high. Here, the term 'variance explained' points to the variance of model predictions – coming from the validation data sets - in relation to the true values of the Global Peace Index, the index for Deaths from conflict and violence, and the index for Uprooted people. A visual presentation of indicators which is consistent with this observation, has been given in Figures 5.2B, 5.3B and 5.4B.

Our overall conclusion on the model performance is that nations in a poor state of peacefulness or in a high state of conflict and violence — expressed in terms of deaths and uprooted people — are not very well predicted by the models estimated. We suggest two explanations which need further research: (i) the quality of data used in this study, and (ii) a simplification of conflict–drivers relationships as shown in Figure 2.2, where drivers such as international interferences in conflicts or the role of ethnicity has not been taken into account.

Our findings are consistent with a number of studies which challenge the role of statistical models in **conflict prediction** (Cederman and Weidmann, 2017; Detges, 2017; Bowlsby et al., 2019).

6.4 Stability of results over time, sensitivity analyses

The analyses given here are based on spatial variability, positioned around the year 2016. Temporal variability has not been considered. Could that weaken the inferences made here? To answer this question, we first checked the temporal variability of dependent and independent variables. Second, we added a water-related shock indicator, taken from the INFORM database, to the Random Forest models shown in Section 5.2.

As stated in Section 4.1 of the Part-I-report, many drivers of conflict only show slow changes over time. For example, we compare the corruption perception of the World Bank for the years 2000 and 2016 in graph 4.1A. The graphs show a strong linear 1-to-1 relationship. Many other indicators show slow increasing patterns over time, with only a few exceptions for those countries which experience the onset of (local) conflicts.

An example is given for the Human Development Index (HDI, Section 2.2.5 of the Part-I-report) in Figure 6.5A for 189 countries. HDI is a composite indicator widely applied to cover wellbeing in countries. It stands for a mixture of the Gross national income per capita, level of education and life expectancy at birth, where each indicator is given equal weight. Most countries show parallel improvements over time, with only a few exceptions.

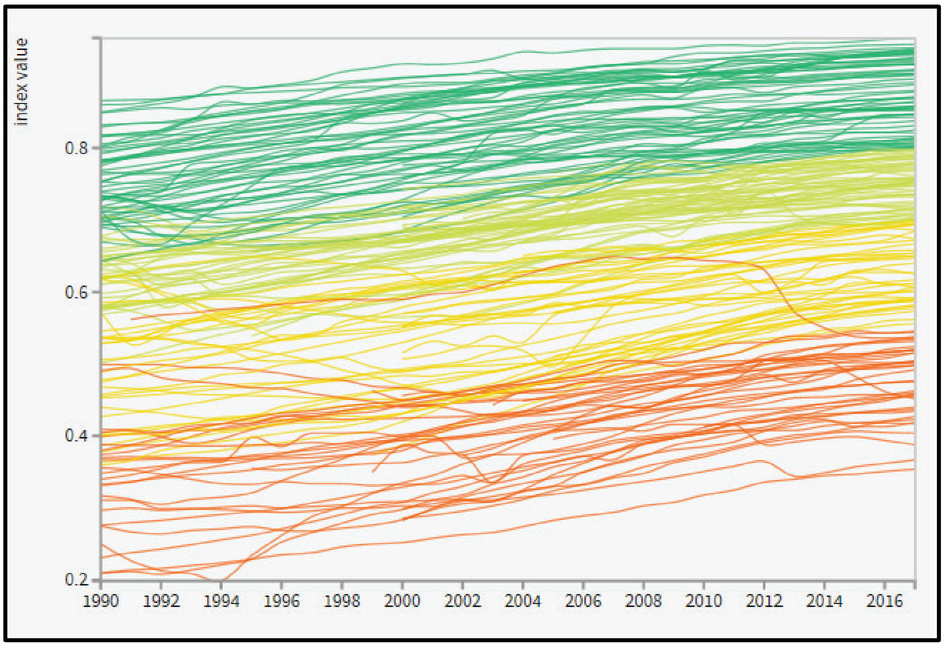


Figure 6.5A Human Development Index for 189 countries, covering the period 1990-2017. This graph and related graphs can be downloaded from: http://hdr.undp.org/en/data. Lowest curve is for Niger, highest curve is for Norway.

Figure 6.5B shows the relationship between the Global Peace Index for the years 2008 and 2018 (the graph is an update from the graph shown in Figure 2.4). Again, the graph shows a strong linear 1-to-1 relationship with a few exceptions. The level of peacefulness in Ukraine, Libya, Yemen and Syria deteriorates due the onset of conflicts. Only Georgia shows a clear improvement. From these observations we conclude that temporal changes in importance functions as shown in Section 5.2 are unlikely, except for a small number of countries with conflict outbreaks.

Next to slow changing conflict drivers some indicators might show shocks which influence conflicts with certain time lags. To check that hypothesis, we analysed an indicator for natural disasters as presented in the INFORM database (indicator #36 shown in Table 2.4B of the Part-I-report). This indicator stands for the relative number of people among the population affected by natural disasters in the three years preceding our base year 2016. Here, natural disasters are earthquakes, tsunami's, droughts, heat waves, cold waves, storms and flooding.

We have added this indicator to the Random Forest analyses given in Section 5.2. For all three conflict indicators we found that this shock indicator showed **lowest** importance relative to the importance functions shown in Figures 5.2A, 5.3A and 5.4A. Although this exercise does not prove that shocks from natural disasters would not influence conflict tensions in countries, we conclude that **direct** effects are unlikely. However, we note that such shocks can act as threat multiplier in countries with weak institutions.

Next to water/weather related shocks political and/or economic shocks may play a role. Political shocks are diverse such as the death of a dictator, falling out of former allies, cancelled or contested elections or external intervention. Economic shocks could be price shocks or economic crises. See Mach et al. (2009, Supplementary Table 1) for details. Here, we have assumed that these shocks are reflected in the wide set of regressors summarized in Table 4.1.

Another topic is that of the **robustness** of results presented in Figures 5.2A, 5.3A and 5.4A. We have performed various of such analyses. First, we checked the dependence of the Random Forest results by filling in the data gaps in driver data. To do so, we used the interpolation technique based on Random Forest and the so-called proximity matrix. Details of this approach are given in Section 3.3 and Liaw (2018).

Due to interpolation the number of countries which can be includes, rises from 129 to 157 in case of the Global Peace Index, from 134 to 188 in case of Deaths by conflict and violence, and from 139 to 191 in case of Uprooted people. Prediction scatterplots are shown in Figures D.1B, D.2B and D.3B. It appears that the importance functions stay unchanged and that a slightly better prediction performance is found.

Second, we re-estimated the Random Forest results by the related method of Regression Trees (explained in Section 3.2 and illustrated in Appendix B). Results are consistent with those found by Random Forest in that the main drivers – governance, GDP per capita and inequality – are dominant here too (Figures B.1, B.2 and B.3). Small differences are found for indicators with lower importance.

Third, the fact that we (i) formulated three indicators for conflict and violence, instead of one, and (ii) took a wide initial set of potential drivers (much wider than the 16 shown in Figure 6.1), makes our results more robust compared to studies with only one dependent variable and a limited set of potential drivers.

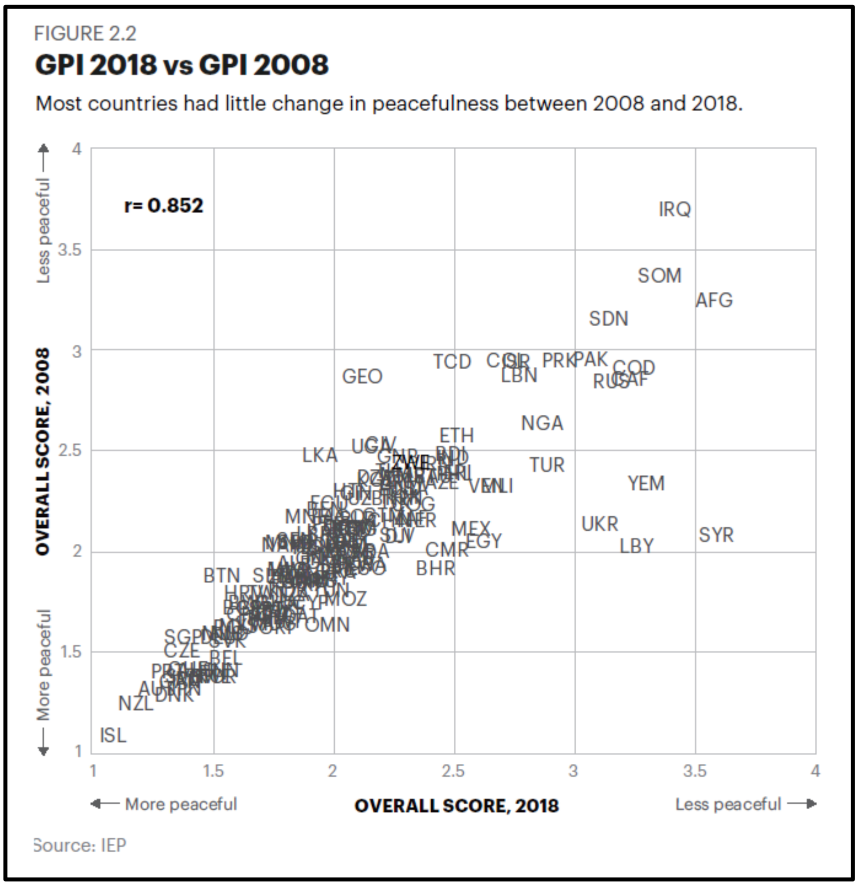


Figure 6.5B Global Peace Index (= overall score on x-axis) for 190 countries and positioned in two years: 2018 (x-axis) and 2008 (y-axis). Three letter codes stand for ISO3 country codes. Source: IEP (2018).

6.5 The Sustainable Development Goals and Dutch foreign policy

In 2015 all United Nation member states adopted the 17 Sustainable Developments Goals (SDGs) for 2030. These SDGs are:

1. End poverty in all its forms everywhere
2. End hunger, achieve food security and promote sustainable agriculture
3. Healthy lives for all
4. Inclusive and equitable quality education for all
5. Gender equality and empowerment of all women and girls
6. Water and sanitation for all
7. Access to affordable and sustainable energy for all
8. Inclusive economic growth, employment and decent work for all
9. Infrastructure for sustainable industrialisation
10. Reduce inequality within and among countries
11. Make cities inclusive, safe, resilient and sustainable
12. Sustainable consumption and production
13. Urgent action to combat climate change
14. Conserve and sustainably use the oceans and seas
15. Protect ecosystems, forests and biodiversity
16. Promote peaceful societies, effective institutions and access to justice for all
17. Revitalise the global partnership for sustainable development

The SDGs promote peace and stability in fragile and unstable regions. They can eliminate the breeding grounds for conflicts and radicalisation and restore trust between citizens and the state, as stated by the UN.

SDGs are the guiding principles for Dutch development cooperation policy, set out in the policy document *Investing in Global Prospects*. As noted in Chapter 1, there is special attention paid to the following goals in developing countries (BuZa, 2018):

1. *Prevent conflicts and instability*
2. *Reduce poverty and social inequality*
3. *Promote sustainable growth and climate action worldwide*

Results from this study have relevance for the Dutch efforts to work on the SDGs on a global scale since the majority of indicators shown in Figure 6.1 are related to SDGs:

- Water-related disasters <-> SDG #13,
- Renewable water resources <-> SDGs #6 & #13,
- Improved water supply and sanitation <-> SDG #6,
- Food security <-> SDG #2,
- Health of children under the age of 5 <-> SDGs #2 & #3,
- Happiness <-> SDG #3,
- Inequality <-> SDGs #5 & #10,
- GDP per capita <-> SDG #8,
- Governance <-> SDG #16.

Furthermore, Education, treated in Appendix A.4, corresponds to SDG #4, and indicators for poverty, treated in Appendix A.1, correspond to SDG #1.

All SDGs are connected and influence each other (as shown in the middle and left panel of Figure 3.1), and although SDG #16 includes both conflict and governance — in terms of institutions, rule of law, accountability and corruption — we have found that working on SDGs #5, #8, #10 and #16 play a central role. We have visualised these central roles in Figure 6.5 by enlarging the more fundamental SDGs based on the findings in this report.

The Dutch policy document emphasises the importance of SDG 16: peace, justice and strong institutions. Based on our analysis investing in good governance will not only lead to lower conflict risks, but it will also support the feasibility of the other 16 goals. Thus, our results give support to the first guiding principle from the policy note: prevent conflicts and instability.

Next to that, the second guiding principle – that of reducing poverty and social inequality – corresponds to the second and third driver identified in this study: GDP per capita and inequality, where the latter is defined as a composite of economic and gender inequality.

The third guiding principle – that of promoting sustainable growth and climate action worldwide – is less clearly reflected in our results, although the indicators 'improving water supply and sanitation', 'improving child health' and 'food security' can be linked to this principle.

We conclude that the main results found here, underpin the guiding principles of the Dutch development cooperation policy.

Figure 6.5 Seventeen Sustainable Development goals (SDGs). SDGs which arise in this report as dominant in relation to conflict and violence are enlarged. These are SDG #16 (related to peace, governance and corruption), SDGs # 5 and 10 (related to economic and gender inequality) and SDGs #1 and 8 (GDP per capita, poverty).

7. Conclusions and future research

Conclusions

By applying Machine Learning techniques, we have explored a wide set of conflict drivers which are proposed in the conflict literature. Since the essence of conflict and violence — defined here on national scales — is multifaceted, and thus not easy to grasp in one (composite) indicator, we have introduced three such indicators, each highlighting various aspects of conflict risks. All models were estimated for these violence indicators in parallel. The modelling approach is summarized in Figure 1.1 as a four-step procedure.

Results have been summarized in Figure 6.1 and compared to similar results in the literature, most notably the recent study of Mach et al. (2019) which is based on expert elicitation (Figure 6.4).

We draw the following conclusions:

- Although statistical conflict analysis can be done by including a wide range of regressors, **governance and socio-economic development** (the latter modelled as 'GDP per capita' and 'Inequality') play a fundamental role. This conclusion is consistent with findings in the conflict literature (e.g. Mach et al., 2019).

- Environmental and water-related indicators have large economic and humanitarian impacts (Figure 7.1), but do not relate very well to conflict and violence, at least if based on national-scale analyses. This conclusion is both consistent and inconsistent with findings in the conflict literature which show contrasting results. Buhaug (2015), Gleick and Iceland (2018) and Mach et al. (2019) show results consistent with those presented here, while Hsiang et al. (2013) and Abel et al. (2019) do find statistical relationships between conflict, climate and environmental indicators.

- We have given a number of explanations for these seemingly contrasting results. First, is important to note that the environmental and climate-related indicators can act as threat multipliers, lying in the proximate group of drivers, shown in Figure 6.2. Thus, they should be seen as contributing factors to instability as concluded by Gleick and Iceland (2018). Second, environmental and climate-related indicators might play important roles on a local scale which level off on a national scale (e.g. De Bruin et al., 2018). Third, water-related disasters and climate extremes do not 'automatically' lead to more grievances, and thus to higher risk levels of conflict and violence. It can bind people too (e.g., Ostrom et al., 1999). Finally, the analyses given here have a global extent. If these analyses would be performed for certain regions, such as for African countries or countries within the EU, quite different results might emerge (stratification, in statistical terms).

  Next to these arguments, it should be noted that water-related disasters lead to the highest number of people affected and people killed if compared to similar numbers by earth quakes and violent conflict (Figure 7.1). Thus, although weather-related disasters are not a major driver for conflict – at least as seen in national census data and surveys - their impact on social and economic disruption is enormous.

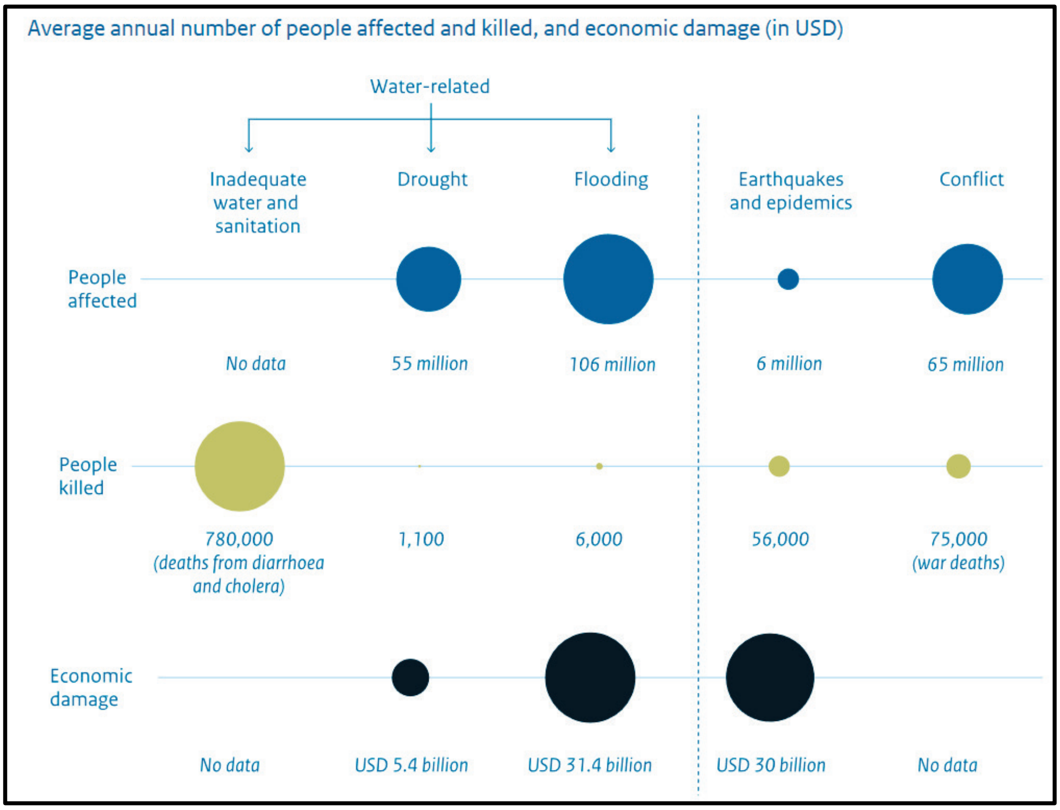


Figure 7.1 Comparison of average annual impacts from disasters, diseases and conflict. Data sources: CRED (droughts, floods and earthquakes, 1996-2015), UCDP (people killed, 1989-2014), UNHCR (people affected by conflicts, 2015) and WHO. Graph taken from Ligtvoet et al. (2018).

- Data quality is a reason of concern for any quantitative study on the relationship between conflict and potential drivers, including the analysis presented in this report (Jerven, 2013; Arnold, 2019; Espey, 2019; Section 6.3 this report). We have found that data quality is limiting statistical analyses in two ways. If information on driver X is missing for country Y, this country will be omitted from the analysis simply because statistical methods cannot cope with missing data in **any** variable. And if data are available, they can be unreliable due to a low level of statistical capacity or definitional uncertainties (cf. Section 5 in the Part-I-report). Both situations are especially true for poor countries with low levels of governance. And unfortunately, these are the countries which one wants to add to one's analysis the most.

- We found that the explanatory power of Random Forest models is moderate. Especially, countries with weak institutions (low levels of governance) show low prediction accuracies. One explanation could be that data in poor countries are less reliable (the conclusion above). Another explanation could lie in the fact that we did not include all relevant indicators in our analysis. Next to that, not all events or influences that may play a role, can be translated to global data sets. These aspects are dynamic or unique, such as the end of the cold war, which included secret negotiations not possible to catch in a number. Factors not explicitly included, are the role of international inferences (proxy wars) and the role of ethnicity and religion.

Limitations

This study has a number of limitations which we will describe briefly. The first limitation we address, concerns our choice for data on a national scale. One drawback has been illustrated in Figure 2.1 of the Part-I-report: each country is counted as one, irrespective of its areal extent or population size. For example, Africa contributes with 55 countries while North and South America contribute with 35 countries. And data from Chile or the Netherlands are given equal importance as data from India or China.

Another drawback is that many factors or drivers play a role on local rather than national scales. This is particular true for water-related indicators. This study finds that, on a national scale, water-related factors do not, or hardly, affect conflict risk. However, conflicts that rise over water are mostly *local* conflicts, as noted in the third conclusion. The report by De Bruin et al. (2018) discusses 10 pathways in which water-related events are coupled to conflict, although not all case studies support the observed pathways. Seven out of these ten pathways describe local conflicts which all erupt in rural areas, except for the pathway discussing the risk of food price spikes.

The third limitation comes from the fact that our study is time-invariant. All indicators are chosen to be representative for the year 2016. It is assumed that if all indicators would be available for the year 2000 or alike, we would find importance functions comparable to those presented in Figure 6.1. This is a reasonable assumption as long as indicators vary slowly over time (as illustrated in Figures 6.5A and B). However, (1) economic shocks due to the onset of conflicts or rebellion, (2) political shocks due to regime changes or cancelled/contested elections or (3) climate shocks due to land-falling cyclones or large-scale flooding just preceding the year 2016 may have influenced values for the year 2016. These potential effects are not considered in this study.

A fourth limitation of the study is the lack of data for those countries that score high on conflict, such as South-Sudan, Yemen or Somalia. It holds for both Regression Tree and Random Forest analyses that a country is omitted if only one (or more) regressors appears to be missing. To increase the number of countries, Figures 5.2B, 5.3B and 5.4B have been included to show these important relationships with a higher number of countries but at the expense of only one regressor. Another approach is imputing missing data, as in Figures D.1B, D.2B and D.3B. This option should be deepened in our view.

Future research

In this study we have developed a broad knowledge base regarding the drivers of conflict on a global scale, based on national data. In future research, much can be done to improve and extend our results. It would be interesting to break this analysis down per continent (stratification in statistical terms). Are governance, inequality and GDP of similar importance in Asia compared to Africa? Or show high risk countries different patterns compared to 'the average'?

A second aspect we did not explore further is the role of foreign powers in conflict: proxy wars / international interventions. The role of foreign intervention is discussed in the study of Mach et al. (2019, their Figure 3a). However, there is no indicator available to assess this aspect, although the amount of foreign war funding could be a way to take this shadowy influence into account. This topic could be deepened further. More generally, we could extend and improve the theoretical basis of this study as summarized in Figure 2.2.

A third aspect concerns the lack of data or the presence of poor data in less-developed countries. Data is sometimes so poor that it becomes unreliable (Jerven, 2013, 2016; Espey, 2019). In future research, it is important to keep an eye on the development of data in these countries and its reliability, more so because these countries face higher conflict risks. To quote Espey: *Unless governments establish competent monitoring systems, the world will not reach the UN Sustainable Development Goals.*

A fourth aspect is the limiting influence of selecting data on national scales only. In the study *The Geography of Future Water Challenges*, we analysed data on water province levels where water provinces fall within the borders of individual countries. One way forward would be to gather indicator data for these spatial scales.

Finally, linkages between conflict and 'water', or more generally 'climate change', are not fixed and **might change in the near future**, especially since climate change is expected to exacerbate conflict–climate connections. Thus, as for the future, the situation sketched in Figure 6.2 will change due to increasing pressure of climate change, water shortages, land degradation and food insecurity, and due to limits of adaption (Figure 7.2; Van der Esch et al., 2017; IPCC, 2018 and 2019). For an in-depth discussion we refer to Mach et al. (2019, Supplementary Tables 2,3 and 4).

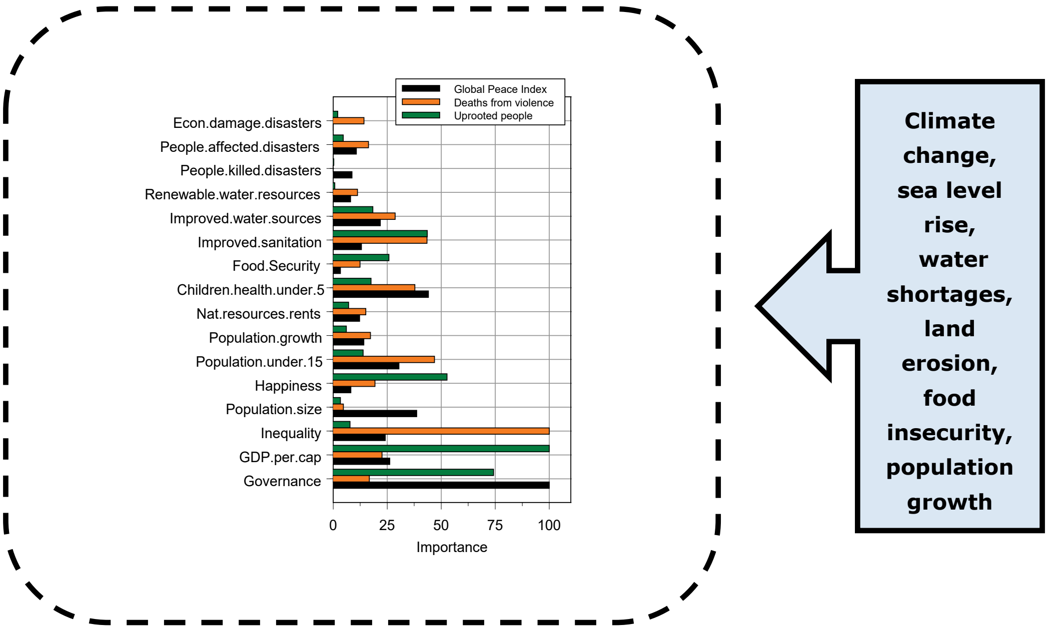


Figure 7.2 Future relationships might change due to increasing pressures from climate change and population dynamics and land erosion. Detailed discussions on adaptation strategies are given by IPCC (2018, 2019).

In conclusion, it is important to perform statistical analyses such as done here, with updated and, hopefully, improved data in the coming years.

Acknowledgements

We wish to thank our colleagues Ezra Berkhout, Guus den Hollander and Ton Dassen for valuable comments on early drafts of this report.

References

Abel, G.J., Brottrager, M., Cuaresma, J.C., Muttarak, R., 2019. Climate, conflict and forced migration. Global Environmental Change 54, 239-249.

Acemoglu, D, S. Johson and J.A. Robinson (2005). Institutions as a fundamental cause of long-run growth. Chapter 6, in Handbook of Economic Growth, Vol. 1A (Eds. P. Aghion and S.N. Durlauf), Elsevier.

Acemoglu, D. and J. A. Robinson (2012). Why Nations Fail. London, Profile Books LTD.

Adams, C., T. Ide, J. Barnett and A. Detges (2018). Sampling bias in climate-conflict research. Nature Climate Change, 8, 200-203.

Andrews, J. (2017). The world in conflict. Understanding the world's troublespots. The Economics Books, London.

Arnold, C. (2019). Death counters. Nature, vol. 566, 7 Feb., 22-025.

Bara, C. (2014). Incentives and opportunities - A complexity-oriented explanation of violent ethnic conflict. Journal of Peace Research, 51(6), 696-710.

Basedau, M. and J. Lay (2009). Resource curse or rentier peace? The ambiguous effects of oil wealth and oil dependence on violent conflict. Journal of Peace Research, 46(6), 757-776.

Bellemare, M. F. (2015). Rising food prices, food price volatility, and social unrest. American Journal of Agricultural Economics, 97(1), 1-21.

Berman, N., M. Couttenier, D. Rohner and M. Thoenig (2017). This Mine is Mine! How minerals fuel conflicts in Africa. American Economic Review 107(6), 1564-1610.

Besley, T. and T. Persson (2011). The Logic of Political Violence. The Quarterly Journal of Economics.

Bongaarts, J. and B.C. O'Neill (2018). Global warming policy: is population left out in the cold? Science, 361, 650-652.

Bowlsby, D., E. Chenoweth, C. Hendrix and J.D. Moyer (2019). The future is a moving target: predicting political instability. British Journal of Political Science, doi:10.1017/S0007123418000443, 1-13.

Breiman, L., J.H. Friedman, R.A. Olshen and C.J. Stone (1984). Classification and regression trees. Wadsworth, Belmont, California.

Breiman, L. (2001). Random Forests. Machine Learning, 45, 5-32.

Bruin, S. de, J. Knoop, W. Ligtvoet and H. Visser (2018). Linking climate change and water security to conflict; an exploration of pathways. PBL background report. The Hague, Netherlands Environmental Assessment Agency.

Brzoska, M. and C. Fröhlich (2016). Climate change, migration and violent conflict: vulnerabilities, pathways and adaptation strategies. Migration and Development, 5(2), 190-210.

Buhaug, H. (2005). The geography of armed civil conflict. PhD Thesis Norwegian University of Science and Technology, Trondheim.

Buhaug, H. and Gleditsch (2008). Contagion or confusion? Why conflicts cluster in space. International Studies Quarterly 5292), 215-233.

Buhaug, H. and I. Rudolfsen (2015). A Climate of Conflicts? Conflict Trends.

Buhaug, H., J. Nordkvelle, T. Bernauer, T. Böhmelt, M. Brzoska, J. W. Busby, A. Ciccone, H. Fjelde, E. Gartzke, N. P. Gleditsch, J. A. Goldstone, H. Hegre, H. Holtermann, V. Koubi, J. S. A. Link, P. M. Link, P. Lujala, J. O'Loughlin, C. Raleigh, J. Scheffran, J. Schilling, T. G. Smith, O. M. Theisen, R. S. J. Tol, H. Urdal and N. von Uexkull (2014). One effect to rule them all? A comment on climate and conflict. Climatic Change, 127(3), 391-397.

Buhaug, H. (2016). Climate Change and Conflict: Taking Stock. Peace Economics, Peace Science and Public Policy, 22(4), 331-338.

BuZa (2018). Investing in Global Prospects. Report Ministry of Foreign Affaires, May 18, 2018. Download from here .

Cederman, L.-E. and N. B. Weidmann (2017). Predicting armed conflict: time to adjust our expectations? Science, 355(6324), 474-476.

Collier, P., V. L. Elliot, H. Hegre, A. Hoeffler, M. Reynal-Querol and N. Sambanis (2003). Breaking the Conflict Trap. Civil War and Development Policy. A World Bank Policy Research Report. Washington, The World Bank.

Collier, P. and A. Hoeffler (2004). Greed and grievance in civil war. Oxford economic papers, 56(4), 563-595.

Collier, P. (2007). The Conflict Trap. Bottom billion. Why the Poorest Countries Are Failing and What Can Be Done About It, Oxford University Press.

De Châtel, F. (2014). The role of drought and climate change in the Syrian uprising: Untangling the triggers of the revolution. Middle Eastern Studies, 50(4): 521-535.

Deaton, A. (2013). The great escape: health, wealth and the origins of inequality. Princeton University Press.

Detges, A. (2017). Climate and conflict: reviewing the statistical evidence. A summary for policy-makers. Climate Diplomacy Berlin, Adelphi.

Diamond, J. (1998). Guns, germs and steel. The fates of human societies. Vintage Publishing.

Eaton, D. (2008). The business of peace: raiding and peace work along the Kenya–Uganda border (Part I). African Affairs, 107(426), 89-110.

Espey, J. (2019). The missing ingredient for a better world: data. Nature 571, 299-299.

Fearon, J. D. and D. D. Laitin (2003). Ethnicity, insurgency, and civil war. American Political Science Review, 97(01), 75-90.

Fearon, J.D. (2010). Governance and civil war onset. World Development report 2011.

Femia, F. and C.E. Werrell (2017). Climate Change, the Erosion of State Sovereignty, and World Order. In C. E. Werrell & F. Femia (Eds.), *Epicenters of climate and security: The new geostrategic landscape of the anthropocene* (pp. 11-19). The Center for Climate and Security.

Fenske, J. (2013). Does land abundance explain African institutions? Economic Journal,123(573), 1363-1390.

Fox, J. (2017). Religion: Influences on Domestic and International Relations. States and Peoples in Conflict M. Stohl, M. I. Lichbach and P. N. Grabosky. New York, Routledge 50-68.

Freedman, L. (2017). The future of war. A history. Penguin books, UK.

Gleick, P. H. (2014). Water, Drought, Climate Change, and Conflict in Syria. Weather, Climate, and Society 6(3), 331-340.

Gleick, P.H. and C. Iceland (2018). Water, security, and conflict. Issue brief WRI and Pacific Institute.

Goldstone, J. A., R. H. Bates, D. L. Epstein, T. R. Gurr, M. B. Lustik, M. G. Marshall, J. Ulfelder and M. Woodward (2010). A Global Model for Forecasting Political Instability. American Journal of Political Science, 54(1), 190-208.

Gunasekara, N., Kazama, S., Yamazaki, D., & Oki, T. (2014). Water conflict risk due to water resource availability and unequal distribution. *Water resources management, 28*(1), 169-184.

Halkia, S., S. Ferri, I. Joubert-Boitat, F. Saporiti and M. Kauffmann (2017). The Global Conflict Risk Index (GCRI) regression model: data ingestion, processing, and output methods. JRC technical report JRC108767.

Hallegatte, S., M. Bangalore, L. Bonzanigo, M. Fay, T. Kane, U. Narloch and A. Vogt-Schilb (2016). Shock Waves : Managing the Impacts of Climate Change on Poverty. World Bank report.

Håvard, H., B. Halvard, V. C. Katherine, N. Jonas, T. W. Stephanie and G. Elisabeth (2016). Forecasting civil conflict along the shared socioeconomic pathways. Environmental Research Letters 11(5), 054002.

Hegre, H., J. Karlsen, H. M. Nygård, H. Strand and H. Urdal (2013). Predicting Armed Conflict, 2010–2050. International Studies Quarterly, 57(2), 250-270.

Hegre, H., H. Buhaug, K.V. Calvin, J. Nordkvelle, S.T. Waldhoff and E. Gilmore (2016). Forecasting civil conflict along the shared socioeconomic pathways. Env. Res. Letters, 11(5), 054002.

Hegre, H., M. Allansson, M. Basedau, et al., 2019. ViEWS: a political violence early-warning system. Peace Research 56(2), 155-174.

Hoffman, M. and A. Jamal (2012). The youth and the Arab spring: cohort differences and similarities. Middle East Law and Governance, 4(1), 168-188.

Hsiang, S.M., M. Burke and E. Miguel (2013). Quantifying the influence of climate on human conflict. Science 341, Sept. 13.

Hsiang, S. M. and M. Burke (2014). Climate, conflict, and social stability: what does the evidence say? Climatic Change, 123(1), 39-55.

Huang, J., H. Yu, X. Guan, G. Wang and R. Guo (2015). Accelerating dryland expansion under climate change. Nature Climate Change, NCLIMATE28-37, 1-7.

Humphreys, M. (2005). Natural resources, conflict, and conflict resolution uncovering the mechanisms. Journal of Conflict Resolution, 49(4), 508-537.

Ide, T. (2015). Why do conflicts over scarce renewable resource turn violent? A qualitative comparative analysis. Global Environmental Change, 33, 61-70.

IEP (2018). Global Peace Index 2018. Annual report on measuring peace by the Institute for Economics and Peace.

IPCC (2018). Global warming of 1.5 ºC. Special report on the impacts of global warming of 1.5 °C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty. See: http://ipcc.ch/report/sr15/.

IPCC (2019). Climate change and land. Special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems. See: https://www.ipcc.ch/report/srccl/.

Jerven, M. (2013). Poor numbers. How we are misled by African development statistics and what to do about it. Cornell University Press, London.

Jerven, M. (2016). Africa by numbers: reviewing the database approach to studying african economies. Research note. African Affairs 115/459, 342-358.

JRC (2017a). INFORM Index for Risk Management. Results 2017.

JRC (2017b). INFORM Index for Risk Management. Concept and methodology 2017 (M. Marin-Ferrer, L. Vernacci and K. Poljansek). Report JRC106949.

Kaufman, D. and A. Kraay (2015). Worldwide Governance Indicators, World Bank.

Kelley, C. P., S. Mohtadi, M. A. Cane, R. Seager and Y. Kushnir (2015). Climate change in the Fertile Crescent and implications of the recent Syrian drought. Proceedings of the National Academy of Sciences, 112(11), 3241-3246.

King, M. D. (2015). The weaponization of water in Syria and Iraq. The Washington Quarterly, 38(4), 153-169.

Kuhn, M. and K. Johnson (2016). Applied predictive modeling. Springer, New York.

LaGraffe, D. (2012). The youth bulge in Egypt: An intersection of demographics, security, and the Arab Spring. Journal of Strategic Security, 5(2), 65.

Liaw, A and M. Wiener (2002). Classification and regression by randomForest. R News, 2/3, 18-22.

Liaw, A (2018). Package 'randomForest'. CRAN publication to be downloaded from here: https://www.stat.berkeley.edu/~breiman/RandomForests/

Ligtvoet, W., J. Knoop, S. de Bruin, D. van Vuuren, H. Visser, K. Meijer, R. Dahm and L. van Schaik (2017). Water, climate and conflict: security risks on the increase? Briefing note Planetary Security Initiative (PBL, Deltares and Clingendael).

Ligtvoet, W., A. Bouwman, J. Knoop, S. de Bruin, K. Nabielek, H. Huitzing, J. Janse, J. van Minne, D. Gernaat, P. van Puijenbroek, J. de Ruiter and H. Visser (2018). The geography of future water challenges. PBL infographics report.

Mach, K.J., C. M. Kraan, W. N. Adger, H. Buhaug, M. Burke, J.D. Fearon, C.B. Field, C.S. Hendrix, J.-F. Maystadt, J. O'Loughlin, P. Roessler, J. Scheffran,K.A. Schultz and N. von Uexkull (2019). Climate as a risk factor for armed conflict. Nature, doi:10.1038/s41586-019-1300-6. Supplementary Information (notably Tables 1 and 2) available at MachSI2019.pdf .

Malik, A. and B. Awadallah (2013). The economics of the Arab Spring. World Development, 45, 296-313.

Mamdani, M. (2005). Good Muslim, bad Muslim: America, the Cold War, and the roots of terror, Harmony.

Natalini, D., A. W. Jones and G. Bravo (2015). Quantitative assessment of political fragility indices and food prices as indicators of food riots in countries. Sustainability, 7(4), 4360-4385.

Niasse, M. (2005). Climate-induced water conflict risks in West Africa: Recognizing and coping with increasing climate impacts on shared watercourses. *Human Security Climate Change*, 21-23.

Nordas, R. and C. Davenport (2013). Fight the youth: youth bulges and state repression. American Journal of Political Science, 57(4), 926-940.

Olsson, L., M. Opondo, P. Tschakert, A. Agrawal and S.E. Eriksen (2014). Livelihoods and poverty. *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press.

Ostrom, E., J. Burger, C.B. Field, R.B. Norgaard and D. Policansky (1999). Revisiting the Commons: local lessons, global challenges. Science 284, 278-282.

Raleigh, C. and H. Urdal (2007). Climate change, environmental degradation and armed conflict. Political Geography, 26(6), 674-694.

Raleigh, C. and I. Vik Bakken (2017). Water stress and conflicts in Africa. PRIO policy brief, 11.

Ross, M.L. (2004). What do we know about natural resources and civil war? Journal of Peace Research, 41(3), 337-356.

Sachs, J. D., A. D. Mellinger and J. L. Gallup (2001). The geography of poverty and wealth. Scientific American, 284(3), 70-75.

Salehyan, I. (2010). The delegation of war to rebel organizations. Journal of Conflict Resolution, 54(3), 493-515.

Schmeier, S., C. Baker, J. Blauw, C. Iceland, K. Meijer and R. Sasse (2019). Are water and conflict linked and what actually links them? Water, Peace and Security opinion peace.

Selby, J., O. S. Dahi, C. Fröhlich and M. Hulme (2017). Climate change and the Syrian civil war revisited. Political Geography, 60, 232-244.

Smith, T. G. (2014). Feeding unrest. Journal of Peace Research, 51(6), 679-695. doi:doi:10.1177/0022343314543722.

Stohl, M, M.I. Lichbach and P.N. Grabosky, eds. (2017). States and peoples in conflict. Routledge, New York.

Strand, H, S.A. Rustad, H. Urdal and H.M. Nygard (2019). Trends in armed conflict, 1946-2018. PRIO Policy brief, Conflict trends 03-2019.

Theisen, O. M., Holtermann, H., & Buhaug, H. (2012). Climate wars? Assessing the claim that drought breeds conflict. *International Security, 36*(3), 79-106.

Transparancy International (2017). What is corruption? Retrieved December 2017, from https://www.transparency.org/what-is-corruption#define.

Tukey, J.W. (1977). Exploratory data analysis. Addison-Wesley Publishing Company.

UN-DP (2016). UN report on human development.

UNHRC (2015). UNHCR, The Environment & Climate Change. V. Türk, S. Corliss, J. Riera et al. Geneva, Switzerland, The UN High Commissioner for Refugees.

Urdal, H. (2011). A clash of generations? Youth bulges and political violence. Paper UN-DP meeting 21-22 July 2011.

Van der Esch, S., B. ten Brink, E. Stehfest, M. Bakkenes, A. Sewell, A. Bouwman, J. Meijer, H. Westhoek and M. van den Berg. Exploring future changes in land use and land condition and the impacts on food, water, climate change and biodiversity. Scenarios for the Global Land Outlook. PBL report 2076.

Visser, H. and H. Noordijk (2002). Correcting air pollution time series for meteorological variability. With an application to regional P_{10} concentrations. RIVM report 722601 007.

Visser, H. and R. Wortelboer (2013). The ecological quality of surface waters in the Netherlands: an analysis with Regression trees. PBL publication number 1075 (in Dutch).

Visser, H., S. de Bruin, J. Knoop, W. Ligtvoet, and A. Bouwman (2018). Planetary security: in search of drivers of violence and conflict. Part I: national-scale databases and reliability issues. PBL report 3404.

Von Uexkull, N., Croicu, M., Fjelde, H., & Buhaug, H. (2016). Civil conflict sensitivity to growing-season drought. *Proceedings of the National Academy of Sciences, 113*(44), 12391-12396.

Ward, M. D., N. W. Metternich, C. L. Dorff, M. Gallop, F. M. Hollenbach, A. Schultz and S. Weschle (2013). Learning from the Past and Stepping into the Future: Toward a New Generation of Conflict Prediction. International Studies Review, 15(4), 473-490.

Weil, D.N. (2015). A review of Angus Deaton’s The Great Escape: Health, Wealth, and the Origins of Inequality. J. opf Economic Literature 53(1), 201-114.

Witmer, F.D.W., A.M. Linke, J. O'Loughlin, et al. (2017). Subnational violent conflict forecasts for sub-Saharan Africa, 2015-2065, using climate-sensitive models. Journal of Peace Research, 54(2), 175-192.

World Bank Group (2016). High and dry: climate change, water, and the economy. Washington DC, The World Bank.

World Bank (2018). Learning to realize education's promise. World Development Report.

Appendix A Pre-selection of drivers

In this Appendix we follow the conflict-driver carrousel shown in Figure 2.2. Each of the driver groups (1) up to (8) is treated in a separate section in the same order. Variables are taken from the data sets described in Part I of this study (Chapter 2). The approach is visual and univariate, that is, indicators for conflict and violence are compared to potential explanatory variables *one-by-one*. Potential drivers are summarised in Table 4.1.

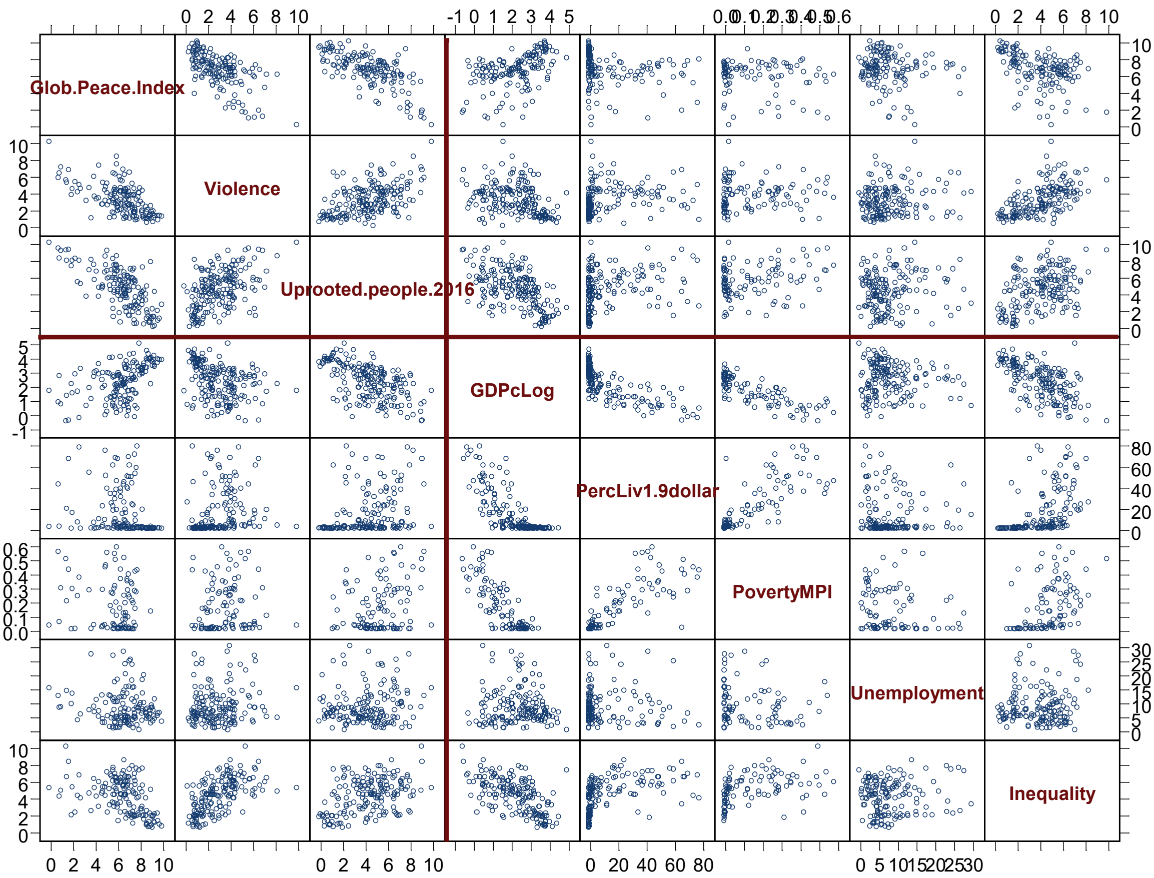
This approach allows us to find and select relevant indicators from a wide set of potential relevant indicators, as summarized in Table 2.5 through 2.6 of the Part-I-report. In each of the Sections A.1 through A.8 the following topics are described (where relevant):

- which driver indicators show the strongest/weakest relationship with one of the indicators GPI, Violence and Uprooted people?
- are there other remarkable linkages?
- are results consistent or inconsistent with findings reported in the literature?

A.1 Economic inequality and poverty

To find associations/relationships between GPI, the indicator for Violence and the indicator for Uprooted people on the one hand, and indicators for inequality and poverty on the other, we selected the following (composite) indicators: (1) GDP PPP per capita, (2) percentage of people living on 1.9 dollar a day, (3) poverty according to the Multiple Poverty Index, (4) unemployment and (5) economic and gender inequality (details are given in Table 4.1).

The scatterplot matrix and corresponding correlation matrix are given in Figure A.1A.



	Glob.Peace.Index	Violence	Uprooted.people.2016	GDPcLog	PercLiv1.9dollar	PovertyMPI	Unemployment	Inequality
Glob.Peace.Index	1.00	-0.61	-0.72	0.49	-0.23	-0.17	-0.12	-0.44
Violence	-0.61	1.00	0.50	-0.35	0.22	0.07	0.14	0.60
Uprooted.people.2016	-0.72	0.50	1.00	-0.57	0.30	0.23	0.11	0.42
GDPcLog	0.49	-0.35	-0.57	1.00	-0.72	-0.64	-0.07	-0.61
PercLiv1.9dollar	-0.23	0.22	0.30	-0.72	1.00	0.88	-0.03	0.52
PovertyMPI	-0.17	0.07	0.23	-0.64	0.88	1.00	-0.19	0.38
Unemployment	-0.12	0.14	0.11	-0.07	-0.03	-0.19	1.00	0.06
Inequality	-0.44	0.60	0.42	-0.61	0.52	0.38	0.06	1.00

Figure A.1A Scatterplot matrix for three violence/conflict indicators and five indicators for economical poverty and inequality. It holds for each individual scatterplot in the matrix that the maximum number of bullets (= countries) is 191, but this number could be much less due to missing data. The lower panel shows the correlation matrix for these variables.

Remarkable is the intermediate role of GDP per capita PPP (the GDPcLog variable) in relation to the three violence indicators. The correlation coefficients account for $0.35 \leq |R| \leq 0.57$. The relationship between GDPc and the Global Peace Index is shown in more detail in Figure A.1B. The green to red shades correspond to the World Bank quality indicator for statistical capacity (Section 5.1). The LOESS trend shows only a minor increase with rising income per capita, from GPI values around 6 to values around 8.

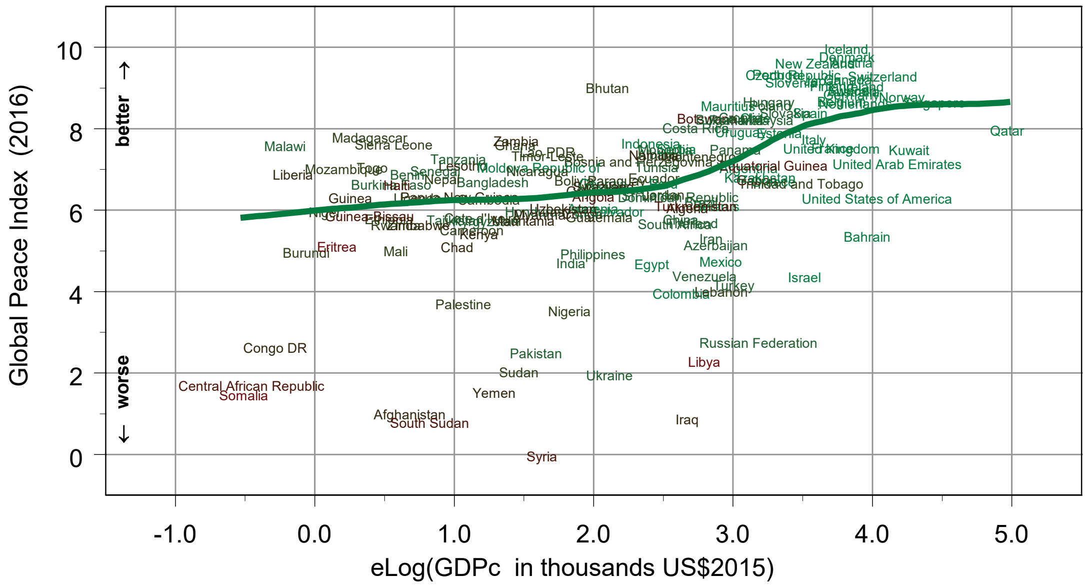


Figure A.1B The relationship between GDP per capita (2015) and the Global Peace Index (2016). The trend is estimated by the LOESS routine. Colors from green to red correspond to the World Bank statistical capacity indicator (red means low capacity).

Both poverty indicators — Percentage of people living on less than 1.9 dollar a day and Poverty MPI — are only weakly correlated to each of the indicators GPI, Violence and Uprooted people (R values in the range of $0.07 \leq |R| \leq 0.30$). Correlations for the variable Unemployment are even lower, in the range of $0.11 \leq |R| \leq 0.14$.

The indicator for Inequality shows the highest correlations (R= 0.60 for Violence). The relationship between Inequality and Violence is shown in more detail in Figure A.1C. The trend in the graph shows a linear relationship between Inequality on the one hand and deaths by violence and conflict on the other hand.

The result for Inequality corresponds to findings of Collier and Hoeffler (2004). A leading role of income per capita and poverty in relation to violence and conflict, as suggested by Fearon and Laitin (2003) and Collier (2007), is not substantiated by our findings (cf. Figure A.1B). The low correlation for unemployment is consistent with the literature summarised in Table 2.1: none of these authors names this factor as leading in their conflict-driver analyses.

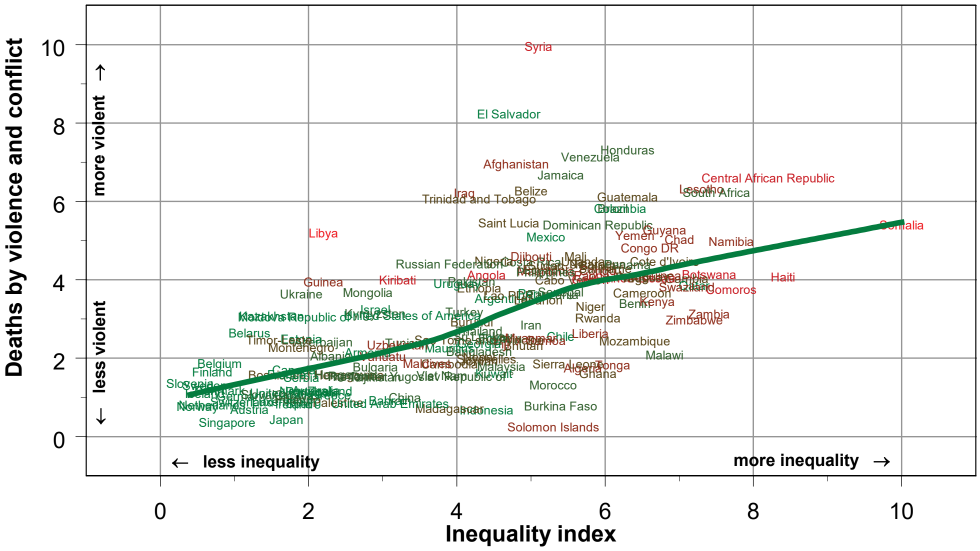


Figure A.1C The relationship between the Inequality and the indicator for deaths by violence and conflict (period 2011-2016). The trend is estimated by the LOESS routine. Colors from green to red correspond to the World Bank statistical capacity indicator (red means low capacity). Correlation coefficient R accounts for 0.60.

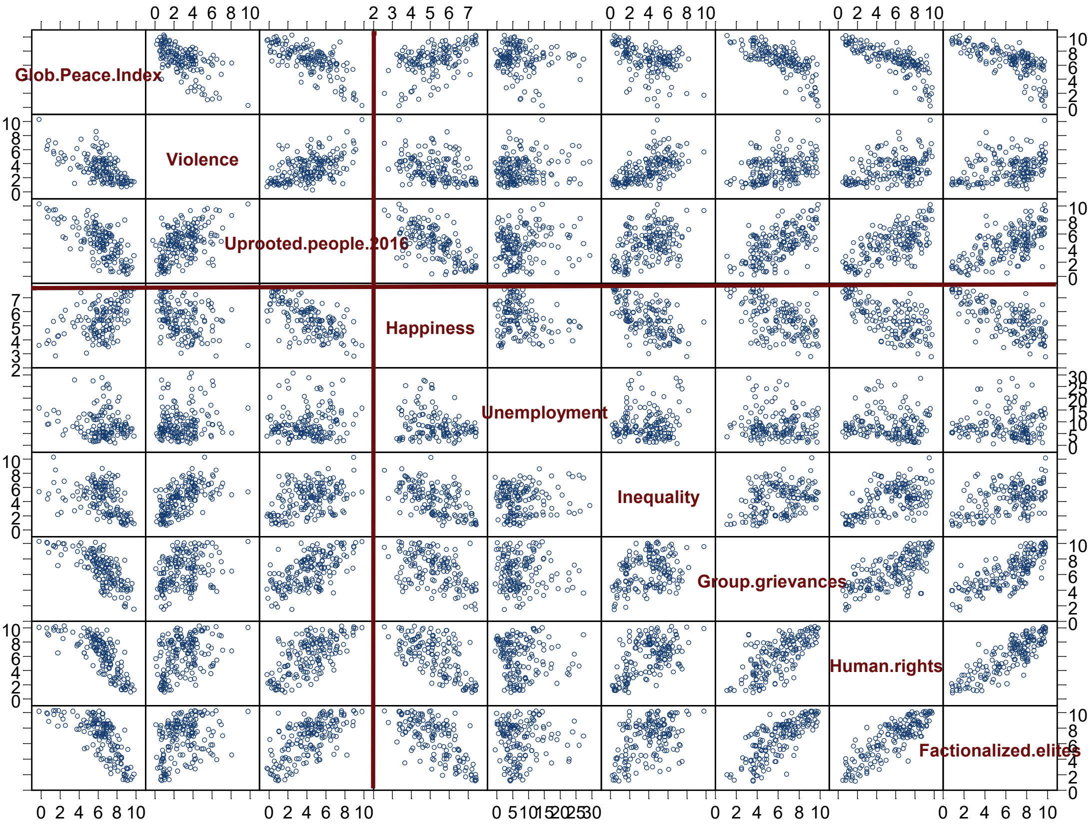
A.2 Grievances and discrimination

To find associations/relationships between GPI, Violence and Uprooted people on the one hand and indicators for grievances on the other, we selected the following (composite) indicators: (1) (un)happiness, a perception indicator for the year 2016, (2) unemployment, (3) economic and gender inequality, (4) group grievances, (5) human rights and (6) factionalised elites.

The scatterplot matrix and corresponding correlation matrix are given in Figure A.2.

The indicator for (un)happiness shows intermediate correlations with GPI, Violence and Uprooted people, with a maximum value of R= -0.69 in relation to Uprooted people. However, the variable Unemployment shows only marginal correlations, $0.11 \leq |R| \leq 0.14$. The indicator for Inequality, here framed in the context of grievances, has a reasonable correlation with Violence, as noted in Section A.1: $0.42 \leq |R| \leq 0.60$.

The three grievance factors from the Fragile States Index show high correlation values, especially for the GPI. Highest value is found for the Human rights index and the Global Peace Index (R= -0.76). However, this result might be explained from the fact that these indicators (partly) share the same underlying information with GPI. Thus, high correlations may not come as a surprise here. Furthermore, the correlations with Violence are low ($0.32 \leq R \leq 0.39$).



	Glob.Peace.Index	Violence	Uprooted.people.2016	Happiness	Unemployment	Inequality	Group.grievances	Human.rights	Factionalized.elites
Glob.Peace.Index	1.00	-0.61	-0.72	0.51	-0.12	-0.44	-0.73	-0.76	-0.72
Violence	-0.61	1.00	0.50	-0.36	0.14	0.60	0.32	0.39	0.33
Uprooted.people.2016	-0.72	0.50	1.00	-0.69	0.11	0.42	0.63	0.68	0.67
Happiness	0.51	-0.36	-0.69	1.00	-0.21	-0.57	-0.50	-0.65	-0.72
Unemployment	-0.12	0.14	0.11	-0.21	1.00	0.06	-0.05	-0.04	0.02
Inequality	-0.44	0.60	0.42	-0.57	0.06	1.00	0.30	0.55	0.48
Group.grievances	-0.73	0.32	0.63	-0.50	-0.05	0.30	1.00	0.70	0.76
Human.rights	-0.76	0.39	0.68	-0.65	-0.04	0.55	0.70	1.00	0.85
Factionalized.elites	-0.72	0.33	0.67	-0.72	0.02	0.48	0.76	0.85	1.00

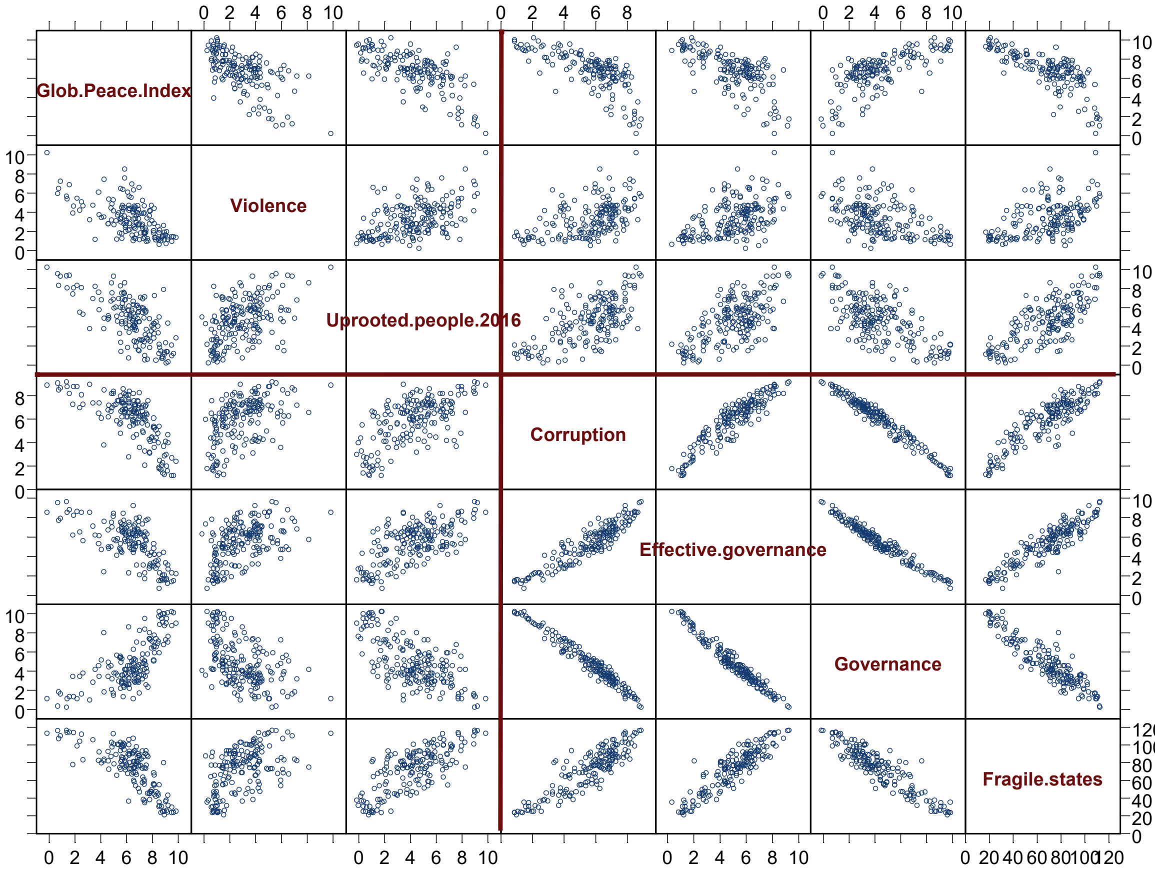
Figure A.2 Scatterplot matrix for three violence/conflict indicators and six indicators for grievances, (un)happiness or discrimination. The lower panel shows the correlation matrix for these variables.

The role of Inequality and the high correlation with the Human rights indicator correspond to the findings of Collier and Hoeffler (2004, 'a lack of political rights'). Furthermore, both factors correspond to the factor 'political and economic discrimination' suggested by Goldstone et al. (2010).

A.3 Effective governance and corruption

To find associations between GPI, Violence and Uprooted people on the one hand and indicators for grievances on the other, we selected the following (composite) indicators: (1) Corruption Perceptions Index 2016, (2) Effective governance, (3) Governance for the year 2016, taken as a composite of indicators (1) and (2), and the Fragile States Indicator. Indicators (1) and (2) have been shown in Figure 2.4 of the Part-I-report.

The scatterplot matrix and corresponding correlation matrix are given in Figure A.3A.



	Glob.Peace.Index	Violence	Uprooted.people.2016	Corruption	Effective.governance	Governance	Fragile.states
Glob.Peace.Index	1.00	-0.61	-0.72	-0.73	-0.71	0.74	-0.78
Violence	-0.61	1.00	0.50	0.50	0.50	-0.50	0.52
Uprooted.people.2016	-0.72	0.50	1.00	0.66	0.66	-0.66	0.77
Corruption	-0.73	0.50	0.66	1.00	0.94	-1.00	0.90
Effective.governance	-0.71	0.50	0.66	0.94	1.00	-0.98	0.92
Governance	0.74	-0.50	-0.66	-1.00	-0.98	1.00	-0.93
Fragile.states	-0.78	0.52	0.77	0.90	0.92	-0.93	1.00

Figure A.3A Scatterplot matrix for three violence/conflict indicators and four indicators for governance and corruption. The lower panel shows the correlation matrix for these variables.

The first observation from the scatterplot matrix is the high correlations between the four governance indicators (4 x 4 panel lower right). For example, the correlation between the INFORM indicator Governance and the Fragile States indicator is **-0.93**. And the correlation between Corruption and Effective Governance is **0.94**.

It shows that the INFORM Governance indicator is strongly related to the GPI, with R= 0.74. The correlations to Violence and Uprooted people are somewhat lower: -0.55 and -0.66, resp. The relationship to the Fragile States Indicator is high, but we note that many characteristics of violence and conflict are incorporated in this composite indicator. Thus, high correlations are not unexpected here (cf. Section 3.2.2).

The relationship between Governance and the GPI is shown in more detail in Figure A.3B. The graph shows a parabolic rather than a linear relationship between both indicators.

The role of governance, also denoted as 'institutions' in the literature, has been highlighted in the work of Goldstone et al. (2010), Fearon and Laitin (2003) and Besley and Persson (2011), as shown in Table 2.1. The role of governance in relation to conflicts and water stress is highlighted in a recent PRIO study by Raleigh and Vik Bakken (2017) and a WRI/Pacific Institute Issue Brief by Gleick and Iceland (2018).

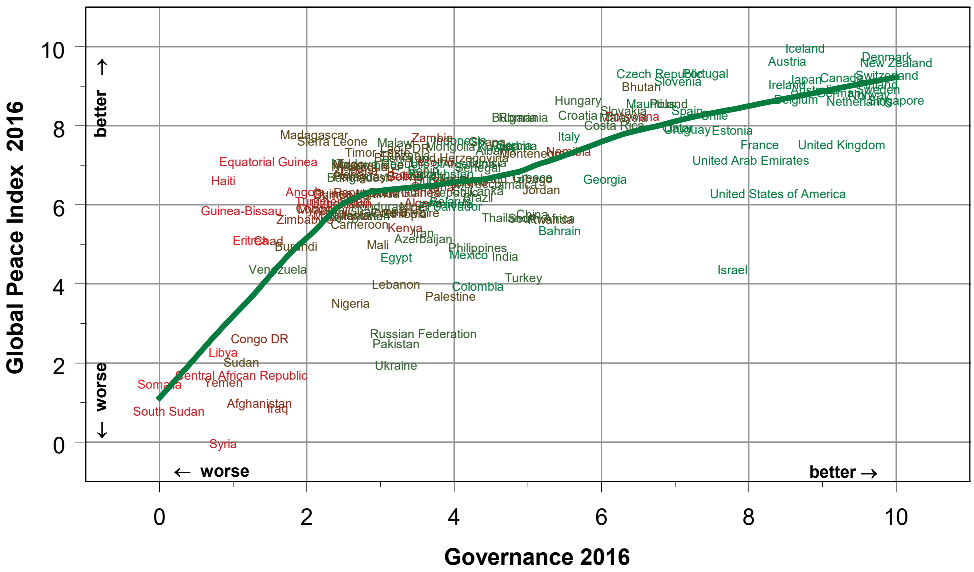


Figure A.3B The relationship between the Global Peace Index and Governance. The trend is estimated by the LOESS routine. Colors from green to red correspond to the World Bank statistical capacity indicator (red means low capacity).

A.4 Demographics and education

To find associations between GPI, Violence and Uprooted people on the one hand and demographic indicators on the other, we selected the following (composite) indicators: (1) a youth bulge indicator computed as the number of people between 15 and 24 years of age, relative to the number of people being 15 years of age and older, as proposed by Urdal (2011), (2) a youth bulge indicator computed as the number of people under 15 years of age, relative to the total population, (3) population growth as annual percentage for the year 2016, (4) population size, (5) an indicator for the level of education, and (6) an indicator for malnutrition and mortality of children under 5 years of age.

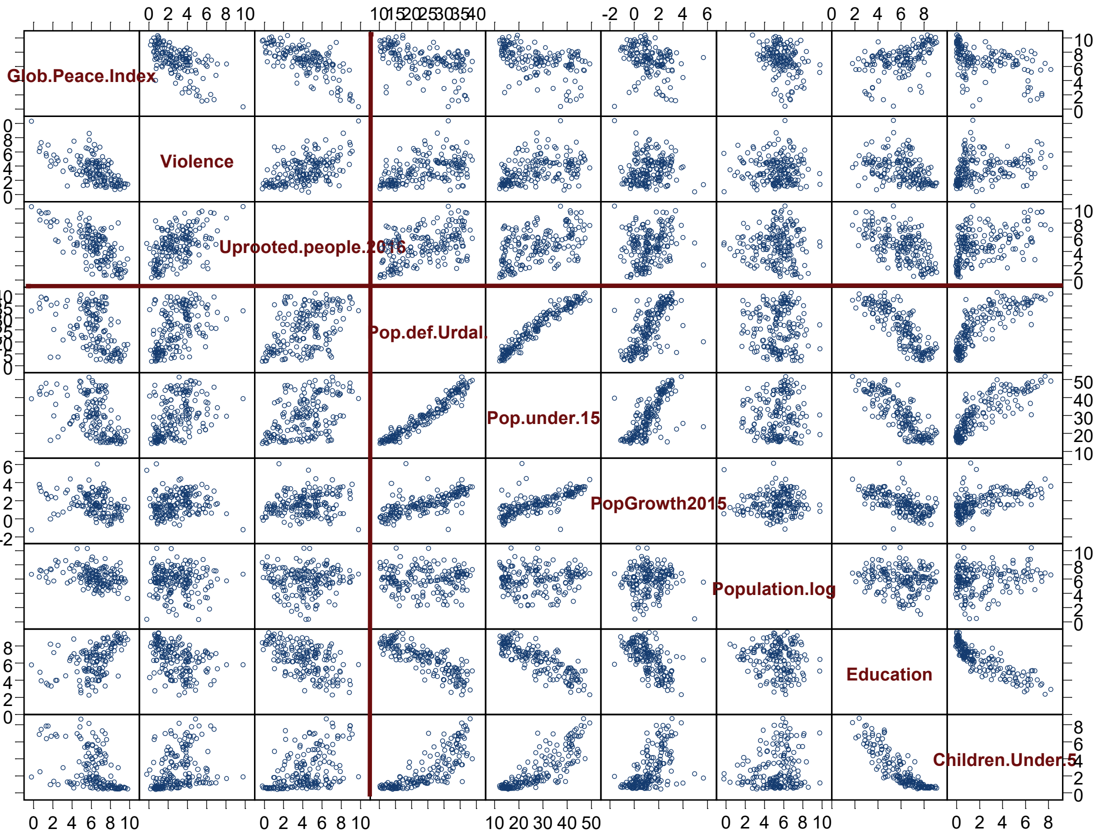
The scatterplot matrix and corresponding correlation matrix are given in Figure A.4A.

The correlation matrix shows intermediate correlations to the three youth-bulge related indicators, with the highest value between GPI/Uprooted people and the percentage of people under 15 years of age. Correlations lie in the range $0.42 \leq |R| \leq 0.49$. The relationships to population growth or population size is lower, in the range of $0.0 \leq |R| \leq 0.31$.

The Education index shows intermediate correlations, in the range of $0.38 \leq |R| \leq 0.56$. The relationship between Education and Uprooted people is shown in more detail in Figure A.4B. The graph show a linear relationship between both indicators in the sense that more education relates to fewer uprooted people.

Finally, the index for malnutrition and mortality among children under 5 years of age shows intermediate correlations: $0.31 \leq |R| \leq 0.48$.

Next to these inferences, the scatterplot matrix shows important relationships between indicators 'the percentage of people under 15 years of age', 'annual population growth', 'level of education' and 'malnutrition and mortality of children under 5'. All these factors are mutually highly correlated with values in the range of $0.59 \leq |R| \leq 0.87$, where R= -0.87 for education and malnutrition/mortality of children under the age of 5.



	Glob.Peace.Index	Violence	Uprooted.people.2016	Pop.def.Urdal.	Pop.under.15	PopGrowth2015	Population.log	Education	Children.Under.5
Glob.Peace.Index	1.00	-0.61	-0.72	-0.47	-0.48	-0.31	-0.21	0.48	-0.48
Violence	-0.61	1.00	0.50	0.47	0.42	0.10	-0.03	-0.38	0.31
Uprooted.people.2016	-0.72	0.50	1.00	0.49	0.48	0.21	0.00	-0.56	0.47
Pop.def.Urdal.	-0.47	0.47	0.49	1.00	0.97	0.67	-0.01	-0.83	0.81
Pop.under.15	-0.48	0.42	0.48	0.97	1.00	0.71	0.04	-0.83	0.83
PopGrowth2015	-0.31	0.10	0.21	0.67	0.71	1.00	0.06	-0.60	0.59
Population.log	-0.21	-0.03	0.00	-0.01	0.04	0.06	1.00	-0.12	0.17
Education	0.48	-0.38	-0.56	-0.83	-0.83	-0.60	-0.12	1.00	-0.87
Children.Under.5	-0.48	0.31	0.47	0.81	0.83	0.59	0.17	-0.87	1.00

Figure A.4A Scatterplot matrix for three violence/conflict indicators, five indicators with a demographical character, and one indicator for education. The lower panel shows the correlation matrix for these variables.

These high correlations correspond to what is denoted in the literature as the *poverty trap*: in countries with high levels of poverty and infectious diseases, mortality under children will be high and the birth rates will increase. As a consequence, poverty will increase more and possibilities for education will shrink. As a result child mortality and birth rates will increase even more (e.g. Sachs et al., 2001).

The strong relationship between the percentage of people under the age of 15 on the one hand, and the level of education on the other hand, follows from Figure A.4C.

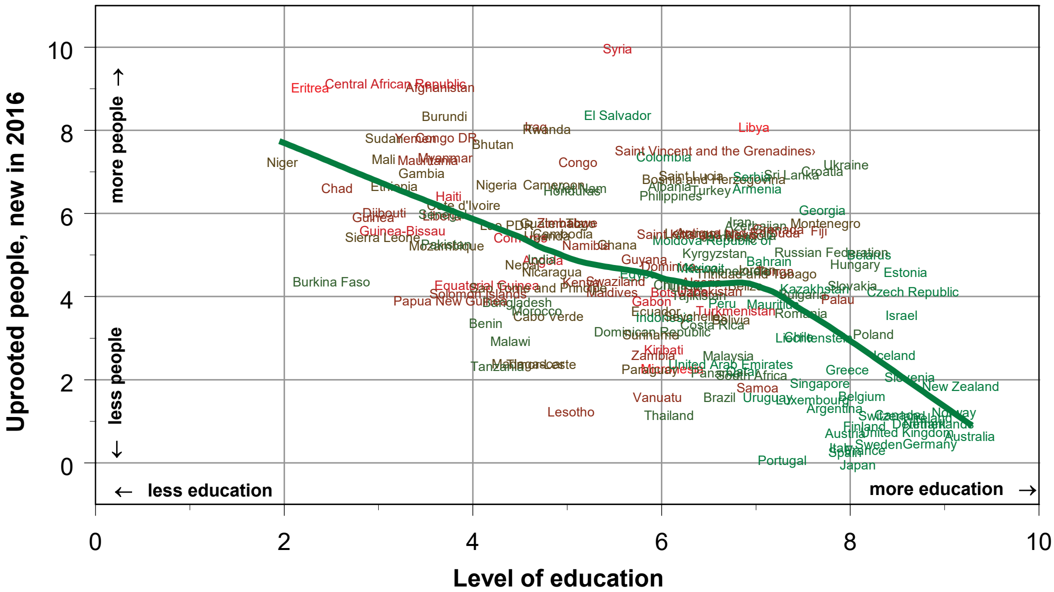


Figure A.4B The relationship between Uprooted people, new in 2016, and the level of education. The trend is estimated by the LOESS routine. Colors from green to red correspond to the World Bank statistical capacity indicator (red meaning low capacity).

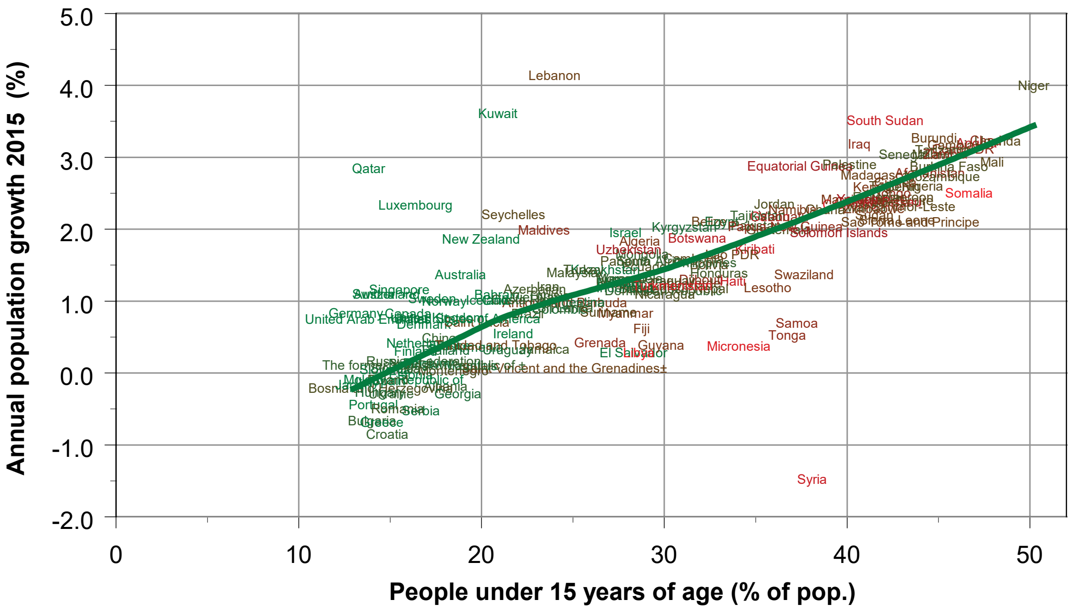


Figure A.4C The relationship between annual population growth for the year 2015 and the percentage of people under the age of 15. The trend is estimated by the LOESS routine. Colours from green to red correspond to the World Bank statistical capacity indicator (red meaning low capacity).

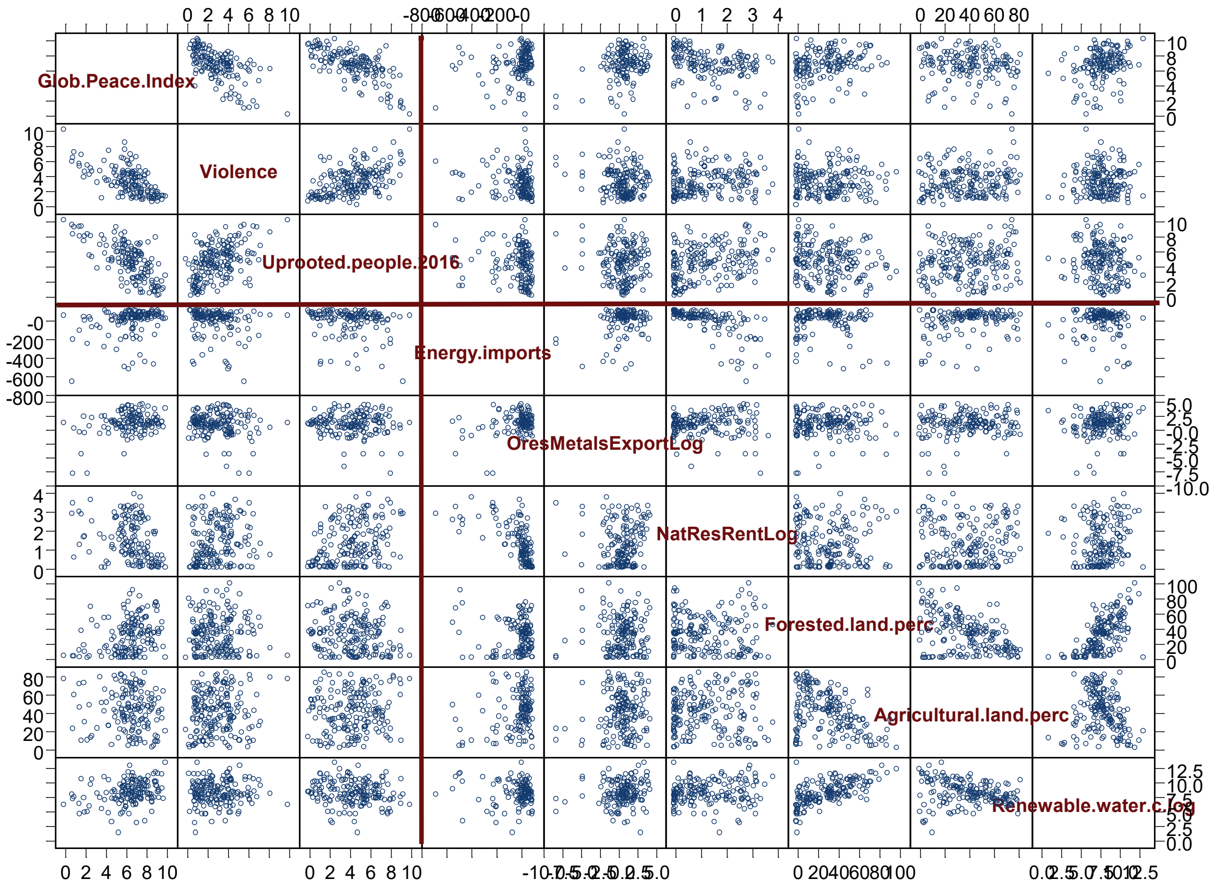
How do these findings relate to the conflict drivers identified by various authors named in Table 8.1? Our findings correspond mostly with those of Hegre et al. (2013) who highlight the role of demographic composition and levels of education. High infant mortality is named by Goldstone et al. (2010). As for the role of education please refer to a recent extensive report of the World Bank (2018): 'Learning to realize education's promise'.

The role of population size is indicated by both Hegre et al. (2013) and Fearon and Laitin (2003), but this role is not substantiated by the results found here ($0 \leq |R| \leq 0.21$). We note that Table 2.1 does not name the role of population growth as potential driver in Table 2.1. This aspect is highlighted in a number of other studies, especially in the context of climate change (Bongaarts and O'Neill, 2018 — and references therein).

A.5 Resources: abundance or scarcity?

To find associations between GPI, Violence and Refugees on the one hand and indicators for resources on the other hand, we selected the following (composite) indicators from the World Bank database (Section 2.2.3): (1) energy imports/exports (as percentage of total energy use, Appendix A #3), (2) ores and metal exports (as percentage of total merchandise exports, Appendix A #10), (3) total natural resources rents (as percentage of GDP, Appendix A #16), (4) forest as percentage of the total area of a country (Appendix A #3), (5) agricultural land as percentage of the total area of a country (Appendix A #1) and (6) renewable fresh water resources per capita (Appendix A #15).

The scatterplot matrix and corresponding correlation matrix is given in Figure A.5A.



	Glob.Peace.Index	Violence	Uprooted.people.2016	Energy.imports	OresMetalsExportLog	NatResRentLog	Forested.land.perc	Agricultural.land.perc	Renewable.water.c.log
Glob.Peace.Index	1.00	-0.61	-0.72	0.24	0.28	-0.31	0.26	-0.09	0.30
Violence	-0.61	1.00	0.50	-0.15	-0.15	0.18	-0.01	0.10	0.04
Uprooted.people.2016	-0.72	0.50	1.00	-0.15	-0.10	0.32	-0.17	0.13	-0.15
Energy.imports	0.24	-0.15	-0.15	1.00	0.25	-0.58	-0.04	0.22	-0.06
OresMetalsExportLog	0.28	-0.15	-0.10	0.25	1.00	0.03	0.01	0.12	0.14
NatResRentLog	-0.31	0.18	0.32	-0.58	0.03	1.00	-0.04	0.02	0.11
Forested.land.perc	0.26	-0.01	-0.17	-0.04	0.01	-0.04	1.00	-0.43	0.63
Agricultural.land.perc	-0.09	0.10	0.13	0.22	0.12	0.02	-0.43	1.00	-0.20
Renewable.water.c.log	0.30	0.04	-0.15	-0.06	0.14	0.11	0.63	-0.20	1.00

Figure A.5A Scatterplot matrix for three violence/conflict indicators and six indicators related to resources of countries. The lower panel shows the correlation matrix for these variables.

The results show only low correlations for all six indicators, with correlations in the range of 0.01 ≤ |R| ≤ 0.32. These low correlations also correspond to the visual inspection of 3*6 scatterplots in the upper right rectangle of Figure A.5A. The best result is found for the relationship between 'Natural resources rents as percentage of GDP' and Uprooted people: R= 0.32.

We re-plotted their scatterplot in Figure A.5B where 'Natural resources rents' are plotted in their original scale as percentage of GDP. The graph illustrates the complex relationship between these two variables, especially for natural resource rents in the range of [0%, 20%].

An explanation for these perhaps unexpected low correlations is not easy to give. The role of resources as a driver for conflict and violence has been highlighted by Collier and Hoeffler (2004), Hegre et al. (2013) and Bara (2014). Clearly, the weak relationships found here do not substantiate a *direct* relation to conflict and violence.

The complexity of conflict conditions has been reviewed by Ross (2004): 'What do we know about natural resources and civil war'. A more recent study of Ide (2015) tried to find relationships in a conjunctural causation study, analysing 20 conflicts of which 7 turned into violence.

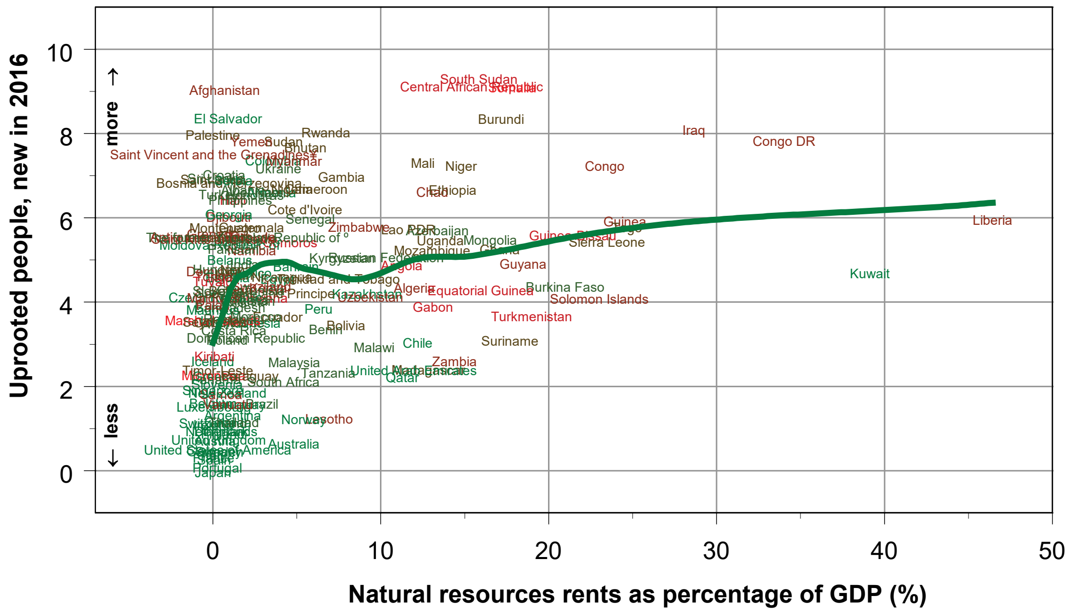
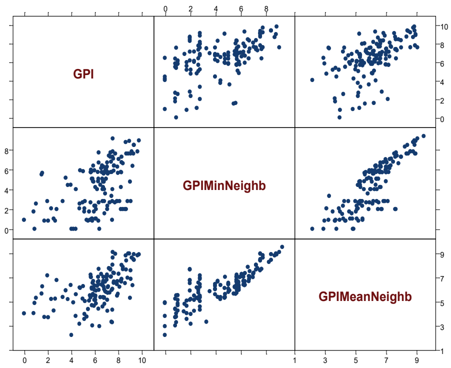


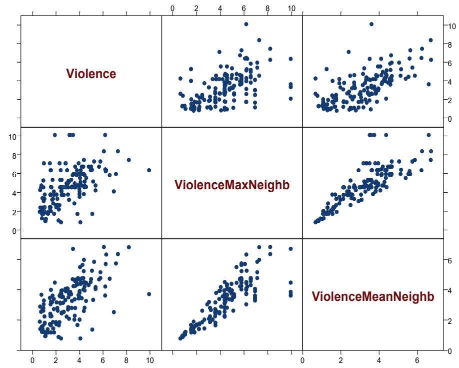
Figure A.5B Scatterplot between Natural resources rents and Uprooted people. Natural resources include oil, natural gas, coal next to minerals and forestry.

A.6 Conflicts in neighbouring countries: spill-over effects

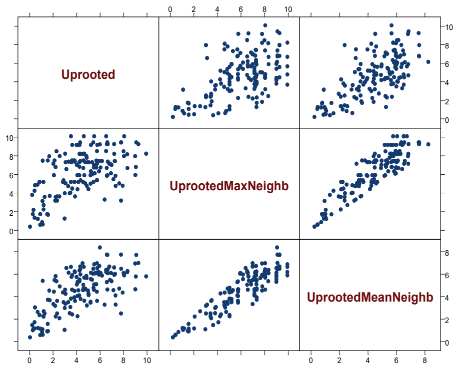
This study does not aim to predict conflict and violence, be it on the short-term or the long-term. However, we tested the role of neighbouring countries by adding this variable to the set of explanatory variables listed in Section 5.2. The scatterplot matrices shown in Figure A.6A show that conflict situations and tensions in neighbouring countries play an intermediate role. As for the Global Peace Index we find a correlation coefficient of R = 0.55, for Violence R= 0.66, and for Uprooted people R= 0.72. The last scatterplot is shown in Figure A.6B in more detail. The relation is not linear but shows a more parabolic shape. Countries lying under the green trend line, such as Yemen, show a much higher number of uprooted people than in the surrounding countries.



	GPI	GPIMinNeighb	GPIMeanNeighb
GPI	1.00	0.52	0.55
GPIMinNeighb	0.52	1.00	0.83
GPIMeanNeighb	0.55	0.83	1.00



	Violence	ViolenceMaxNeighb	ViolenceMeanNeighb
Violence	1.00	0.52	0.66
ViolenceMaxNeighb	0.52	1.00	0.83
ViolenceMeanNeighb	0.66	0.83	1.00



	Uprooted	UprootedMaxNeighb	UprootedMeanNeighb
Uprooted	1.00	0.59	0.72
UprootedMaxNeighb	0.59	1.00	0.89
UprootedMeanNeighb	0.72	0.89	1.00

Figure A.6A Scatterplot matrices for the global Peace Index (upper left), Violence (upper right) and Uprooted people (lower middle). Method explained in Section 4.2 of the Part-I-report.

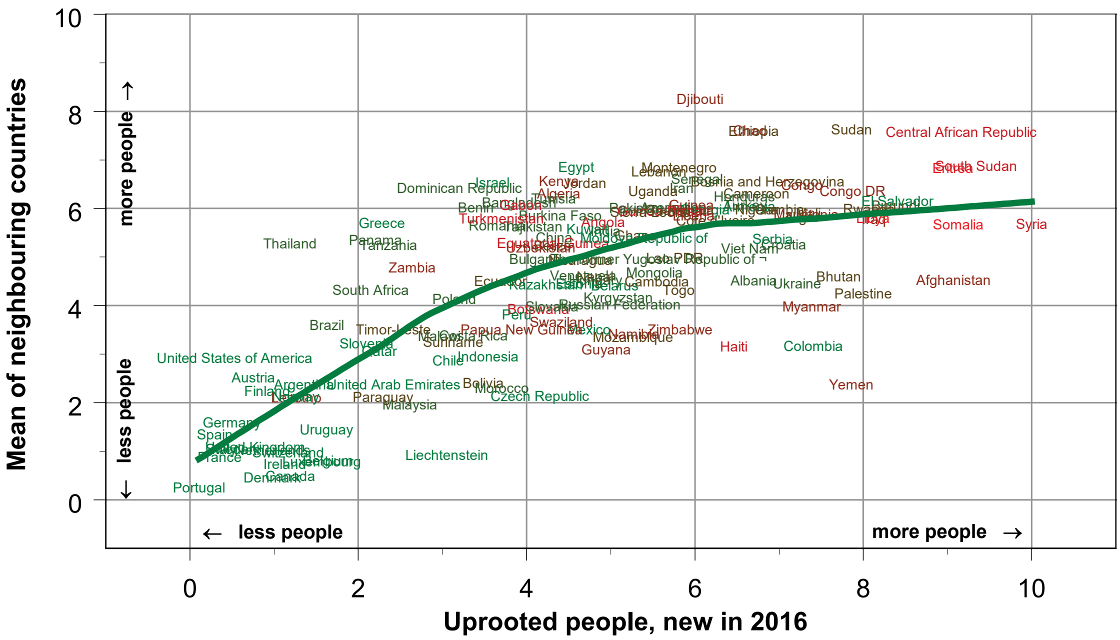
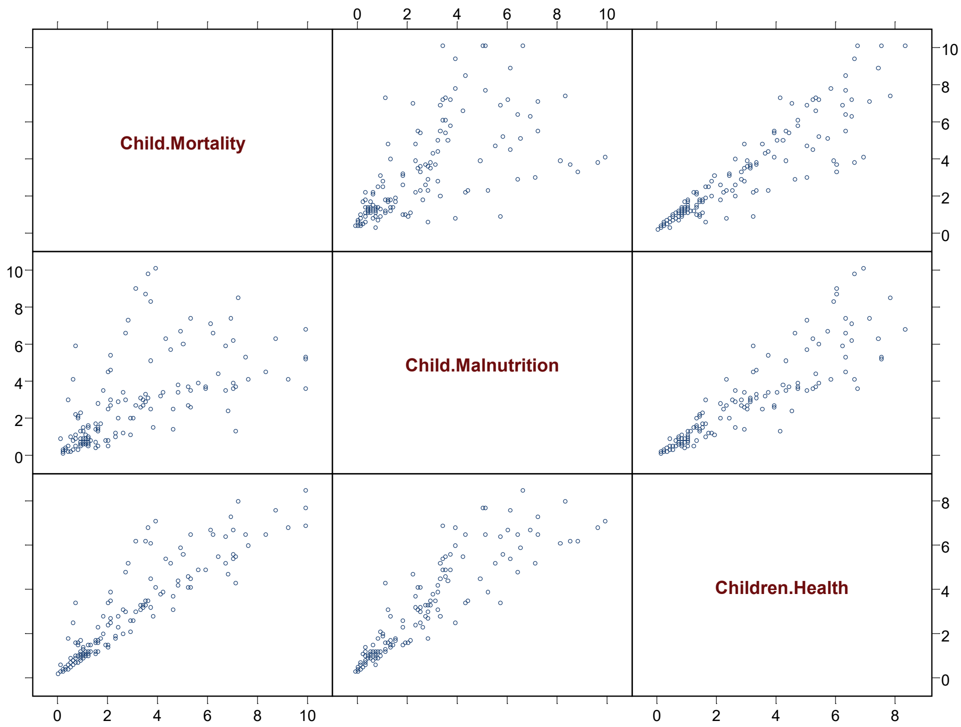


Figure A.6B Scatterplot between Uprooted people in a country and the mean value of Uprooted people in all of the surrounding countries.

A.7 Infant mortality and malnutrition

Here, the composite indicator 'Children health' is used, taken from the JRC INFORM database. This composite is a combination of two underlying indicators: child mortality and child malnutrition for children of 5 years and younger. The scatterplot in Figure A.7 shows that the composite is highly correlate two both underlying indicators (R = 0.90 and 0.93). From this result we conclude that the composite 'Child health' is a reasonable choice for child health care .



	Child.Mortality	Child.Malnutrition	Children.Health
Child.Mortality	1.00	0.64	0.93
Child.Malnutrition	0.64	1.00	0.90
Children.Health	0.93	0.90	1.00

Figure A.7 Scatterplot matrix for child mortality, child malnutrition and a combination of these two indicators, denoted as child health. These indicators were taken from the INFORM database.

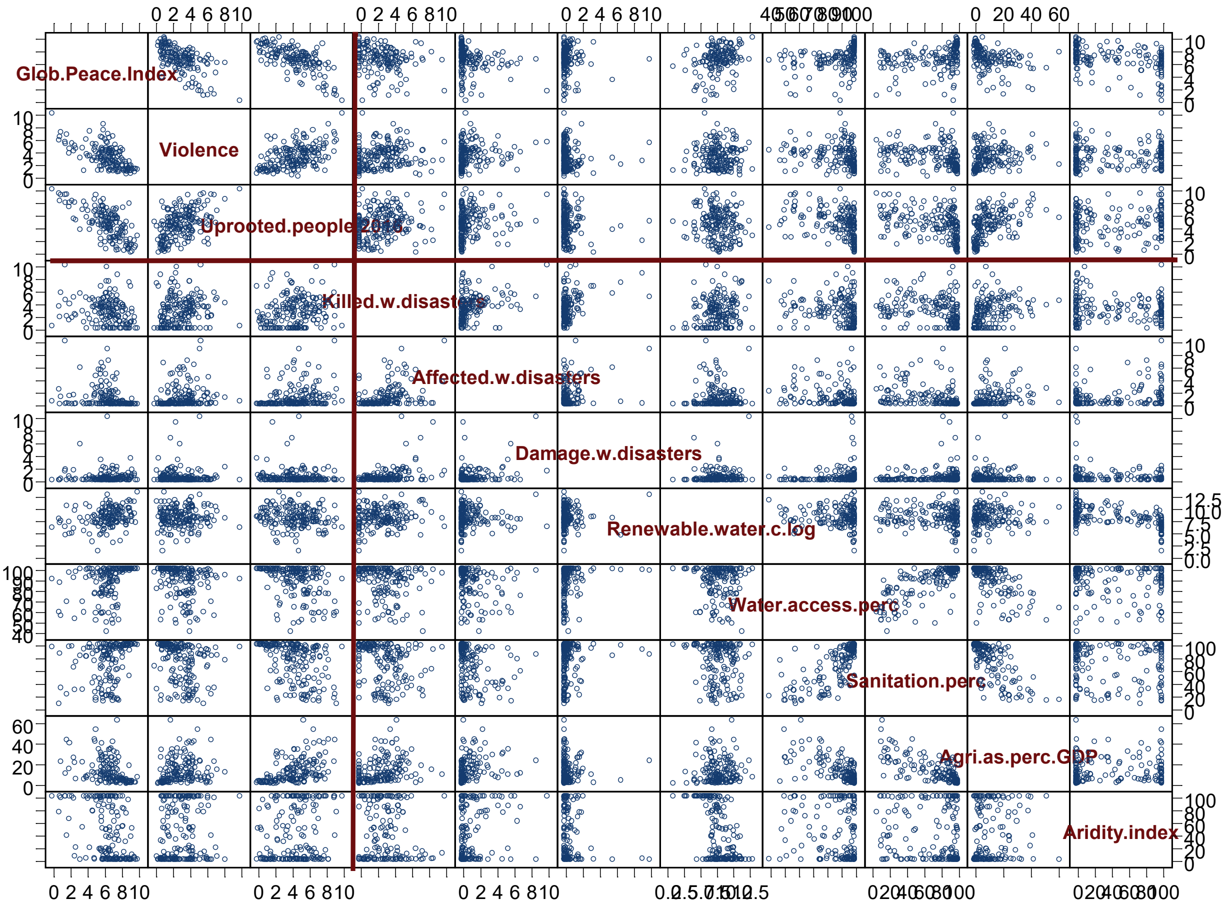
A.8 Water-related impacts and climate change

To find associations between GPI, Violence and Uprooted people, on the one hand, and indicators for water-related impacts, on the other, we selected the following three indicators for impacts of natural disasters: (1) the number of people killed as percentage of the total population, (2) the number of people affected as percentage of the total population, and (3) the direct economic damage relative the GDP of each country. Next to these disaster-related indicators we added (4) renewable internal freshwater resources per capita (indicator #15 in Appendix A), (5) improved water sources as % of population with access (indicator #5 in Appendix A), (6) improved sanitation facilities as % of population with access (indicator #6 in Appendix A), (7) agricultural rents as % of GDP, and (8) the percentage of the country area that can be categorized as dryland (from 0 to 100%). For definitions on dryland in relation to aridity — as a function of precipitation and evaporation — please refer to Huang et al. (2015).

The scatterplot matrix and corresponding correlation matrix is given in Figure A.8A. The correlation matrix shows a large number of low correlation values. As for the relation between GPI, Violence and Uprooted people, on the one hand, and water-related disasters on the other, correlations lie in the range of $0.01 \leq |R| \leq 0.25$. Values for improved water access and improved sanitation are somewhat higher: $0.25 \leq R| \leq 0.39$. Values for agricultural rents as percentage of GDP lie in the range of $0.21 \leq |R| \leq 0.48$. Finally, the aridity index — expressed here as the percentage of dryland per country — shows $0.08 \leq |R| \leq 0.40$.

As an illustration we re-plotted the scatterplot matrices for the aridity index and agricultural rents in Figure A.8B. The upper panel shows the complex interactions between the GPI and the percentage of drylands in a country. The scatter in GPI for countries which are 100% drylands is enormous: from an GPI values of zero (Syria) to 10 (Iceland). This shows that there is no simple, direct relation between these two indicators. A stronger relation is shown in the lower panel, be it not a linear relation. The graph shows that countries with agricultural rents above 30% of its GDP have higher values for the Uprooted people indicator. However, the scatter for countries with percentages ≤ 30% is enormous.

The rather low correlations found here illustrate the conclusions from a number of recent studies in that (1) water-related indicators play an important role in conflict situations, but not as a main drivers at present times. Interactions are indirect (Raleigh and Vik Bakken, 2017; Ligtvoet et al., 2018; De Bruin et al., 2018; Gleick and Iceland, 2018).



	Glob.Peace.Index	Violence	prooted.people.201	Killed.w.disasters	Affected.w.disasters	Damage.w.disasters	enewable.water.c.lo	Water.access.perc	Sanitation.perc	Agri.as.perc.GDP	Aridity.index
Glob.Peace.Index	1.00	-0.61	-0.72	-0.19	-0.25	-0.01	NA	0.34	0.27	-0.36	-0.40
Violence	-0.61	1.00	0.50	0.15	0.20	-0.05	NA	-0.25	-0.28	0.21	0.08
Uprooted.people.2016	-0.72	0.50	1.00	0.20	0.21	-0.08	NA	-0.39	-0.39	0.48	0.19
Killed.w.disasters	-0.19	0.15	0.20	1.00	0.45	0.43	NA	-0.24	-0.29	0.31	-0.03
Affected.w.disasters	-0.25	0.20	0.21	0.45	1.00	0.34	NA	-0.27	-0.33	0.28	0.20
Damage.w.disasters	-0.01	-0.05	-0.08	0.43	0.34	1.00	NA	0.15	0.13	0.01	-0.13
Renewable.water.c.log	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Water.access.perc	0.34	-0.25	-0.39	-0.24	-0.27	0.15	NA	1.00	0.77	-0.60	-0.17
Sanitation.perc	0.27	-0.28	-0.39	-0.29	-0.33	0.13	NA	0.77	1.00	-0.71	-0.08
Agri.as.perc.GDP	-0.36	0.21	0.48	0.31	0.28	0.01	NA	-0.60	-0.71	1.00	0.10
Aridity.index	-0.40	0.08	0.19	-0.03	0.20	-0.13	NA	-0.17	-0.08	0.10	1.00

Figure A.8A Scatterplot matrix for three violence/conflict indicators and eight indicators concerning water-related impacts. The lower panel shows the correlation matrix for these variables.

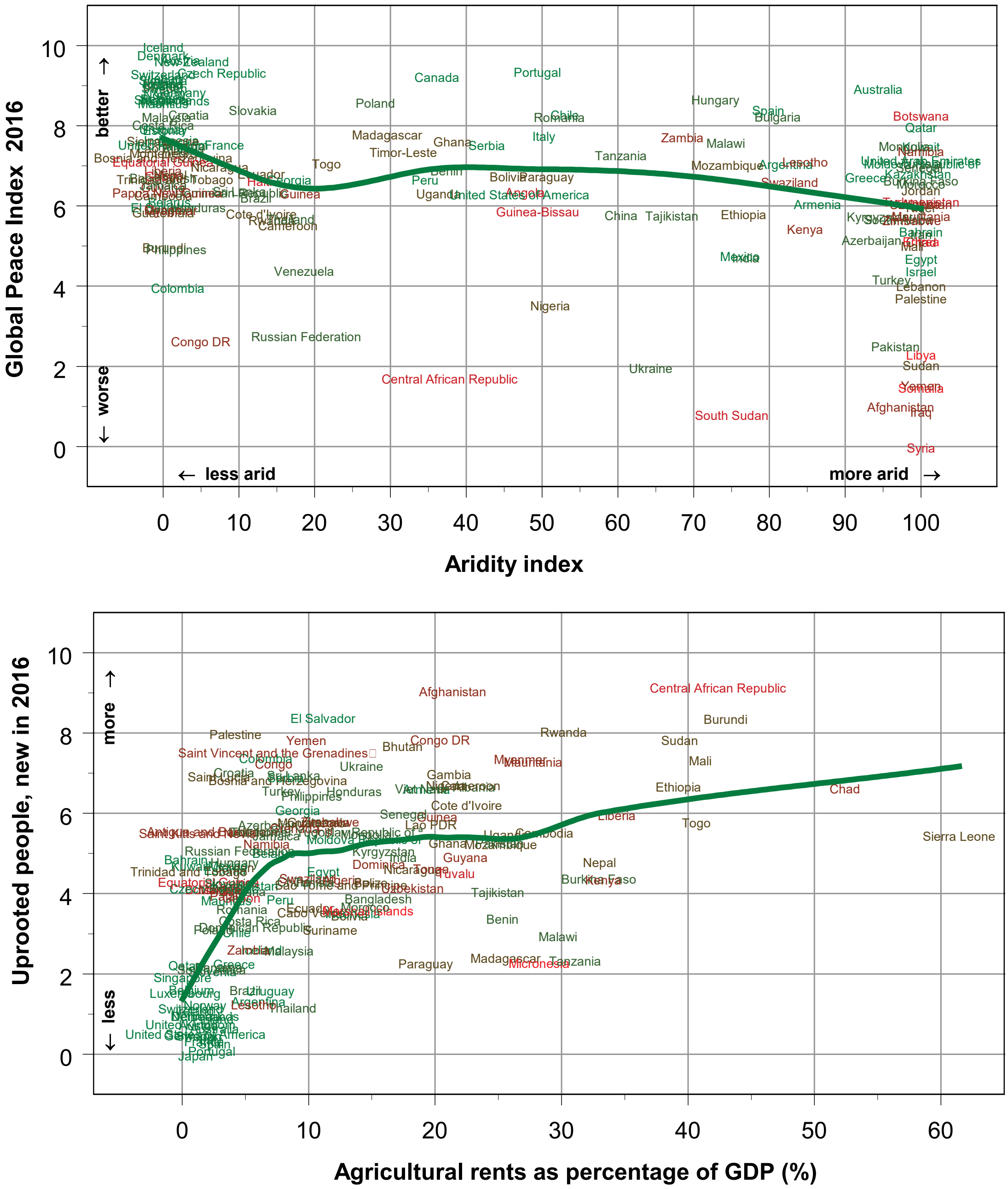


Figure A.8B The relation between GPI and aridity (upper panel), and Uprooted people and agricultural rents (lower panel). The trends are estimated by the LOESS routine. Colours from green to red correspond to the World Bank statistical capacity indicator (red meaning low capacity).

Appendix B Regression Trees and Random Forests by simulation example

An example of a regression tree is given in Figure B.1 where we model a **simulated** violence indicator Y_i in 151 countries from a series of a series of 13 potential explanatory variables $X_{1,i}$, ….. $X_{13,i}$:

"Governance","GDP.per.cap","Inequality","Population.size","Population.under.15", "Nat.resources.rents","Children.health.under.5","Food.Security","Improved.sanitation","Improved.water.sources","Renewable.water.resources", "People.killed.disasters","Education".

The simulated violence indicator Y_i is **linearly** composed from the indicators 'Inequality' and 'People killed from water-related disasters' as follows:

RandomForestViolence2$ViolenceTest <- ((RandomForestViolence2$People.killed.disasters + 2*RandomForestViolence2$Inequality)/3.0) + rnorm(191,0.5)

Thus, we have given the indicator Inequality double weight. Note 1: both indicators have values between 0 and 10, with '10' denoting the worst situation for both variables. Note 2: the number of countries in the databases is 191 but since we need countries with non-missing values for all variables, the number of countries reduces to 151. Note 3: all drivers selected above come from the databases described in the Part-I-report.

The average value of Y_i over all 151 countries is given in the upper node of the tree: 4.50 with a deviance of 518. The first split in the tree is for the variable *Inequality*. If this indicator is under the threshold value of 3.35, mean Y_i values decrease to 2.52 (left ellipse). If this indicator is above the threshold, mean Y_i values decrease to 5.60 (right ellipse).

The predictive power of this split can be calculated from the deviances shown under the ellipses: the initial deviance of 518 lowers to a deviance of 52+137 = 189. Thus, the reduction in deviance by this single spit is 100 * (518 - 189) / 518 = 64%. The full predictive power of this tree is 100 * (518 - 74) / 518 = 86%, with the number 74 equalling the sum of deviances in all six leaves of the tree.

The Splus tree software also produces a pruning graph shown in Figure B.2. Here, we have chosen to prune to tree to six leaves (= end nodes). Software output is given in Box B.1.

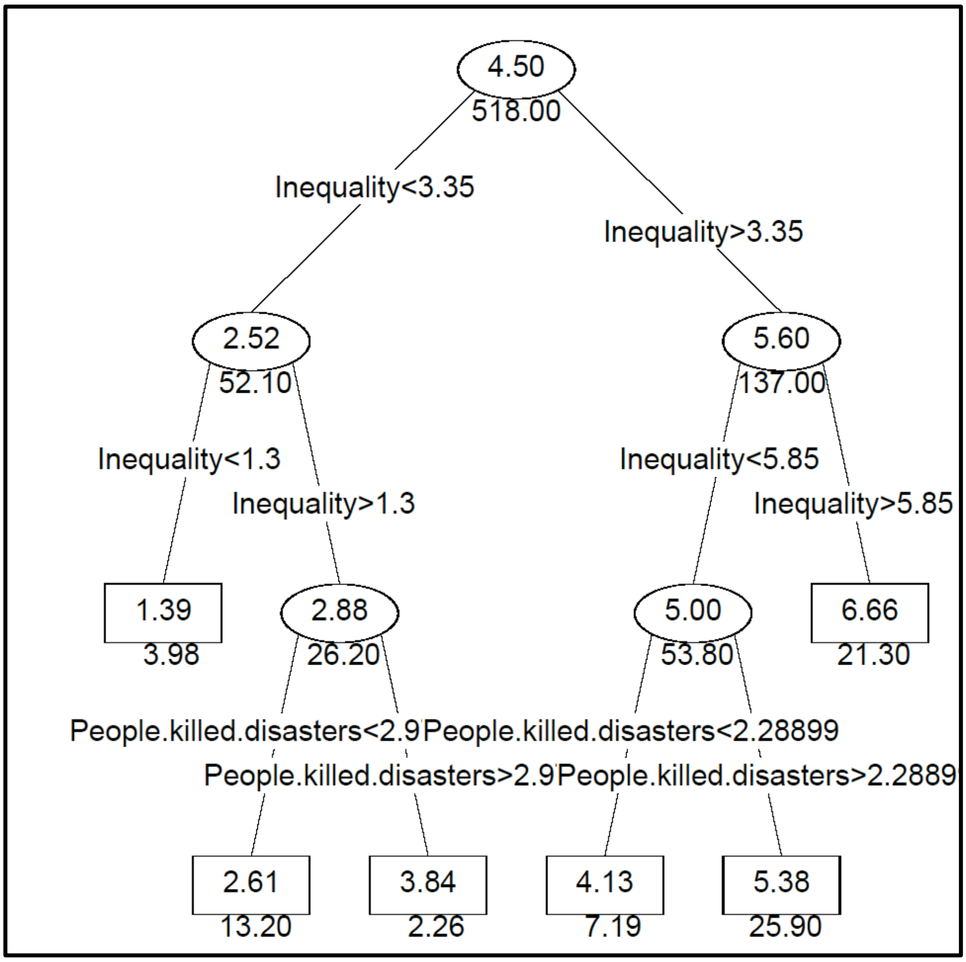


Figure B.1 Regression tree estimate for Simulated violence as dependent variable and estimated for 151 countries. The software output is given in Box B.1. Inequality is a measure for income and gender inequality and lies in the range of [0.0, 10.0], with higher values pointing to more inequality. The independent variable People.killed.disasters has values in the range of [0.0, 10.0] with higher values pointing worse conditions. Variance reduction is 86%

Box B.1 Output Regression Tree model shown in Figure B.1 after pruning to 6 leaves (decision based on the graph shown in Figure B.2).

```
Variables actually used in tree construction:
[1] "Inequality"              "People.killed.disasters"
Number of terminal nodes:  6
Residual mean deviance:  0.5091 = 73.82 / 145
Distribution of residuals:
        Min.     1st Qu.      Median        Mean     3rd Qu.
   Max.
 -1.709e+000 -5.517e-001 -1.508e-002 -5.544e-016  4.634e-001
   2.876e+000

[[2]]:
node), split, n, deviance, yval
      * denotes terminal node

 1) root 151 517.700 4.497
   2) Inequality<3.35 54  52.050 2.519
     4) Inequality<1.3 13   3.981 1.387 *
     5) Inequality>1.3 41  26.160 2.877
      10) People.killed.disasters<2.97161 32  13.200 2.606 *
      11) People.killed.disasters>2.97161 9   2.260 3.841 *
   3) Inequality>3.35 97 136.600 5.598
     6) Inequality<5.85 62  53.780 5.000
      12) People.killed.disasters<2.28899 19   7.193 4.131 *
      13) People.killed.disasters>2.28899 43  25.910 5.384 *
```

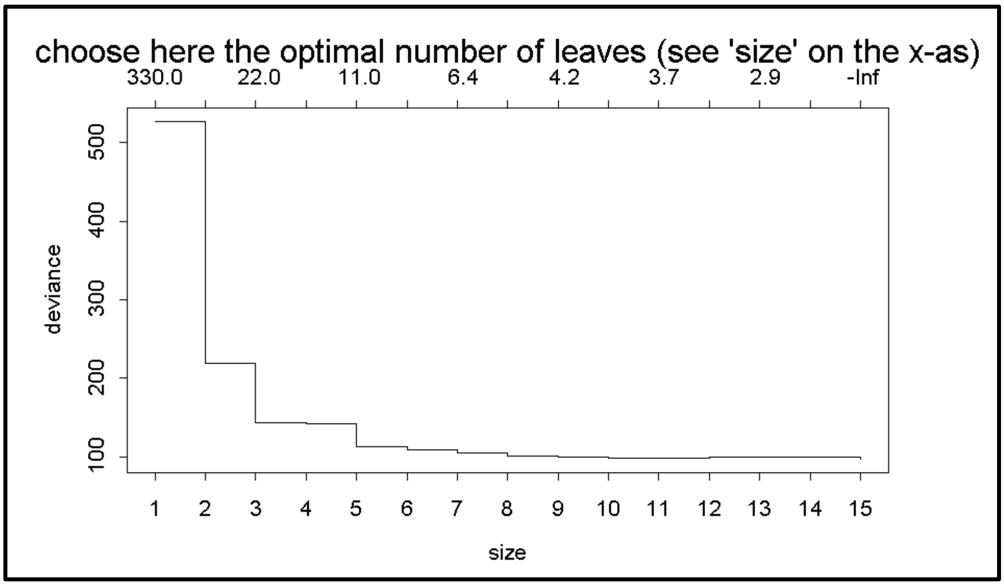


Figure B.2 Deviance reduction as a function of tree size (the final number of leaves in the tree). Here we have chosen to prune the tree to 6 leaves (shown in Figure B.1).

Next to the Regression Tree approach we have estimated a Random Forest model to the same data. Two importance functions and their combination to one function is shown in Figure B.3. The functions show that the correct regressors are found with Inequality having double weight.

Due the principle of bootstrap aggregation we find around 100 predictions for each country (we estimate 300 trees, where the software splits data into 2/3 for training and 1/3 for validation). These predictions are shown in Figure B.4. Here, country predictions are averaged over 100 predictions.

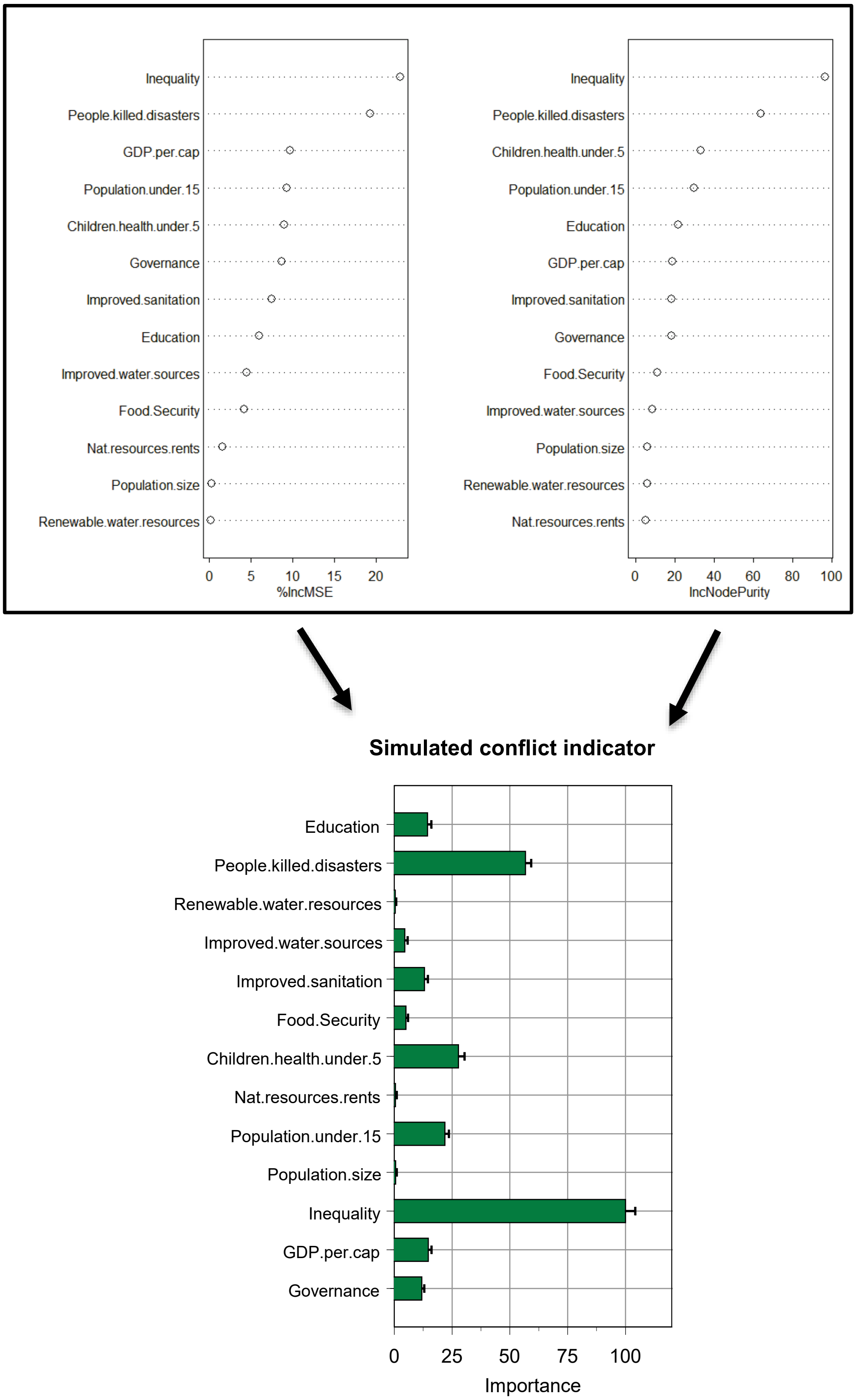


Figure B.3 Two basic importance functions and the way these two functions are combined to one function (based on scaling x-values from 0 to 100 and subsequent averaging).

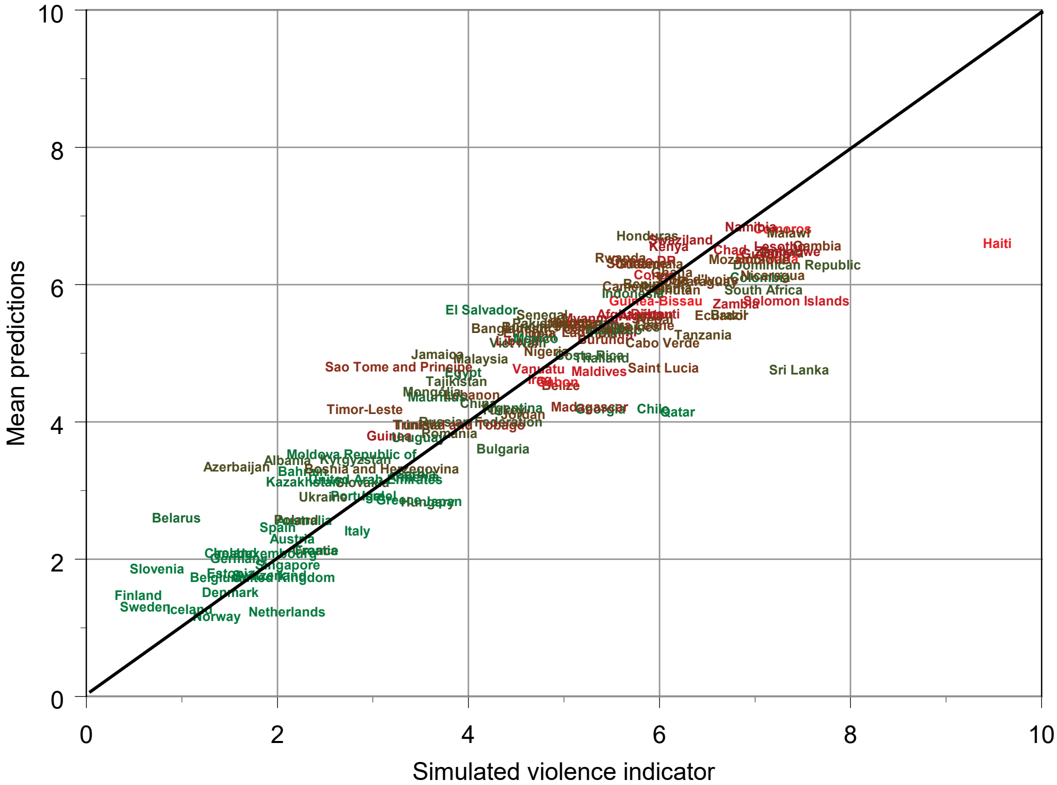


Figure B.4 Prediction scatterplot for 151 countries. Since we estimated 300 regression trees where each tree is estimated on a bootstrap sample (consisting of two third of the countries and then extended again to 151 countries by random sampling) we have generated 100 true predictions for each country. These predictions are averaged per country and these mean predictions are shown in the graph. The mean squared error (MSE) over all predictions is 0.61. The squared correlation between predictions and true value for each of the 300 trees lies around 0.82. Colors from green to red correspond to the World Bank statistical capacity indicator (red means low capacity).

Appendix C Regression Trees for three conflict indicators as dependent variable

Here, we repeat the analysis as described in Section 5.2 by analysing the three conflict indicators by Regression Trees. Thus, we show alternatives for Figures 5.2A (Global Peace Index), 5.3A (Deaths from violence and conflict) and 5.4A (People uprooted people by conflict and violence).

The first Regression Tree is shown in Figure C.1. It is important to note that countries with missing data for one or more variables are omitted in the analysis (identical to Random Forest). Therefore, the number of countries reduces from 191 to 133. The 58 countries which are left out, are mostly nations with less than 500,000 inhabitants — such as Grenada, Dominica, Malta and Samoa — or in some cases countries in conflict, such as Eritrea, Central African Republic and Somalia.

The tree shows that only three variables are of importance, namely governance (factor 3), population size (factor 4) and inequality (factor 2). Furthermore, the order of importance of these three variables is directly seen since this importance decreases if we go down from the upper node of the tree to the lower leaves (the rectangular boxes). We find a variance reduction of **68%** by including these three regressors.

We note that the reduction of each split in the tree can be calculated from the deviances shown under each node and leave. Here, the first split of Governance > 5.5 versus ≤ 5.5 gives an variance reduction of 100 * (487 - 251 - 47) / 487 = 39%.

In summary, the hierarchy found is (1) governance, (2) population size and (3) inequality.

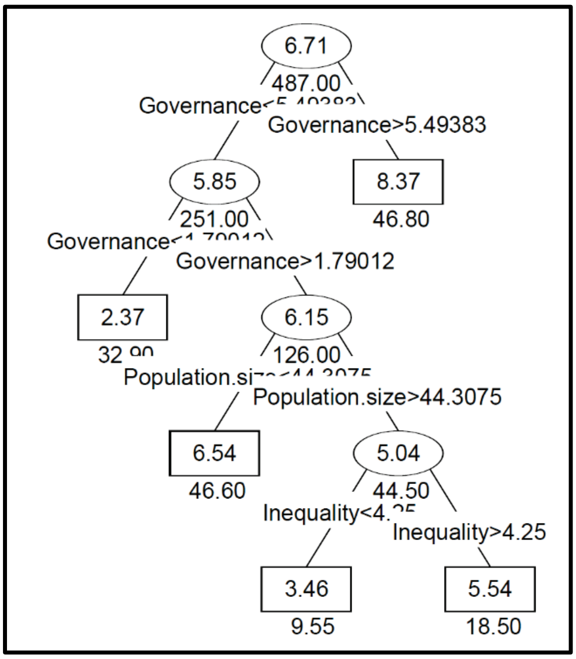


Figure C.1 Regression tree with the **Global Peace Index** (GPI) as dependent variable. The GPI lies in the range of [0.0, 10.0] with high values pointing to countries with better peace conditions. The independent variable Governance also lies in the range of [0.0, 10.0] with higher numbers pointing to better governance. Population.size is expressed in millions. Inequality lies in the range of [0.0, 10.0] with higher values pointing to more inequality. Explained variance: **68%**. Analysis is for 133 countries which have data for all indicators considered.

Number of terminal nodes: 5

Residual mean deviance: 1.206 = **154**.3 / 128

node), split, n, deviance, yval
    * denotes terminal node

1) root 133 **487**.100 6.706
  2) Governance<5.49383 88 250.900 5.852
    4) Governance<1.79012 7 32.910 2.373 *
    5) Governance>1.79012 81 125.900 6.153
      10) Population.size<44.3075 60 46.570 6.541 *
      11) Population.size>44.3075 21 44.550 5.044
        22) Inequality<4.25 5 9.549 3.459 *
        23) Inequality>4.25 16 18.500 5.540 *
  3) Governance>5.49383 45 46.780 8.375 *

The second regression tree for Deaths from violence and conflict is shown in Figure C.2. Here, only two regressors dominate, namely Inequality and Governance where inequality is the most important factor. The total reduction in variance is **46%**. Thus, the explanatory power of the tree model is lower than for the Global Peace Index (which is 68%).

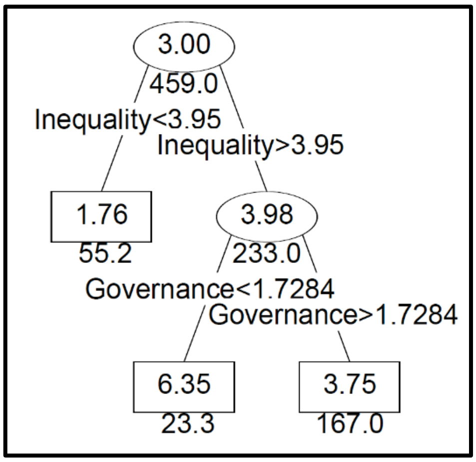


Figure C.2 Regression tree with **Deaths by violence and conflict** as dependent variable. The Violence indicator lies in the range of [0.0, 10.0] with high values pointing to countries with more deaths by violence and conflict. The independent variable Governance also lies in the range of [0.0, 10.0] with higher numbers pointing to better governance. Inequality lies in the range of [0.0, 10.0] with higher values pointing to more inequality. Explained variance: **46%**. Analysis is for 140 countries which have data for all indicators considered.

Variables actually used in tree construction: "Inequality", "Governance"

Number of terminal nodes: 3

Residual mean deviance: 1.792 = **245.5** / 137

node), split, n, deviance, yval
* denotes terminal node

1) root 140 **458.70** 3.000
2) Inequality<3.95 62 55.21 1.764 *
3) Inequality>3.95 78 233.50 3.982
6) Governance<1.7284 7 23.31 6.353 *
7) Governance>1.7284 71 167.00 3.748 *

The third Regression Tree is shown in Figure C.3, for People uprooted people by conflicts and violence. Here, two new variables come in: GDP per capita and (un)happiness. The variables governance and inequality are selected here as well (as in the other two regression trees). The order of importance of these variables is (1) GDP per capita, (2) governance, (3) (un)happiness and (4) inequality. The variance reduction from all splits in the tree accounts for **64%**. This is comparable to the variance reduction we find for the Global Peace Index, shown in Figure C.1.

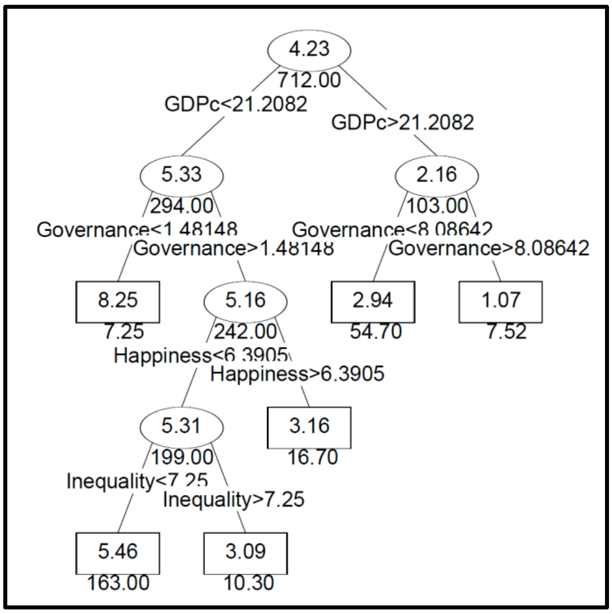


Figure C.3 Regression tree with **Uprooted people** as dependent variable. This indicator lies in the range of [0.0, 10.0] with high values pointing to countries with more uprooted people. The independent variable Governance also lies in the range of [0.0, 10.0] with higher numbers pointing to better governance. GDPc - i.e. GDP per cap PPP - is expressed in thousands USD2015. Inequality lies in the range of [0.0, 10.0] with higher values pointing to more inequality. Explained variance: **64%.** Analysis is for 138 countries which have data for all indicators considered.

Number of terminal nodes: 6

Residual mean deviance: 1.965 = **259.4** / 132

node), split, n, deviance, yval
    * denotes terminal node

1) root 138 **712** 4.226
  2) GDPc<21.2082 90 294.400 5.328
    4) Governance<1.48148 5 7.246 8.250 *
    5) Governance>1.48148 85 242.000 5.156
      10) Happiness<6.3905 79 199.400 5.308
        20) Inequality<7.25 74 162.900 5.458 *
        21) Inequality>7.25 5 10.280 3.092 *
      11) Happiness>6.3905 6 16.720 3.155 *
  3) GDPc>21.2082 48 102.900 2.159
    6) Governance<8.08642 28 54.720 2.937 *
    7) Governance>8.08642 20 7.517 1.069 *

Appendix D Prediction performance three conflict indicators and governance

Here, we show the prediction scatterplots corresponding to the Random Forest analyses given in Section 5.2 for each of the three conflict and violence indicators (graphs D.1, D.2 and D.3).

The prediction scatterplot where Governance is the dependent variable (Section 6.3 and Figure 6.2A) is given in Figure D.4. The correspondending importance function is shown in Figure D.5.

To find the influence of missing data in potential drivers of conflict and violence we re-estimated the models which underly those used for Figures D.1, D.2 and D.3. The difference lies in the interpolation of missing data in the drivers by use of the Random Forest proximity matrix beforehand. The advantage is that more countries are included in the analysis. An example for the Global Peace Index is given in Figure D.6. The number of countries increases from 129 in Figure D.1 to 157 in Figure D.6. Also the variance explained shows an improvement: from 49% in Figure D.1 to 56% in Figure D.6.

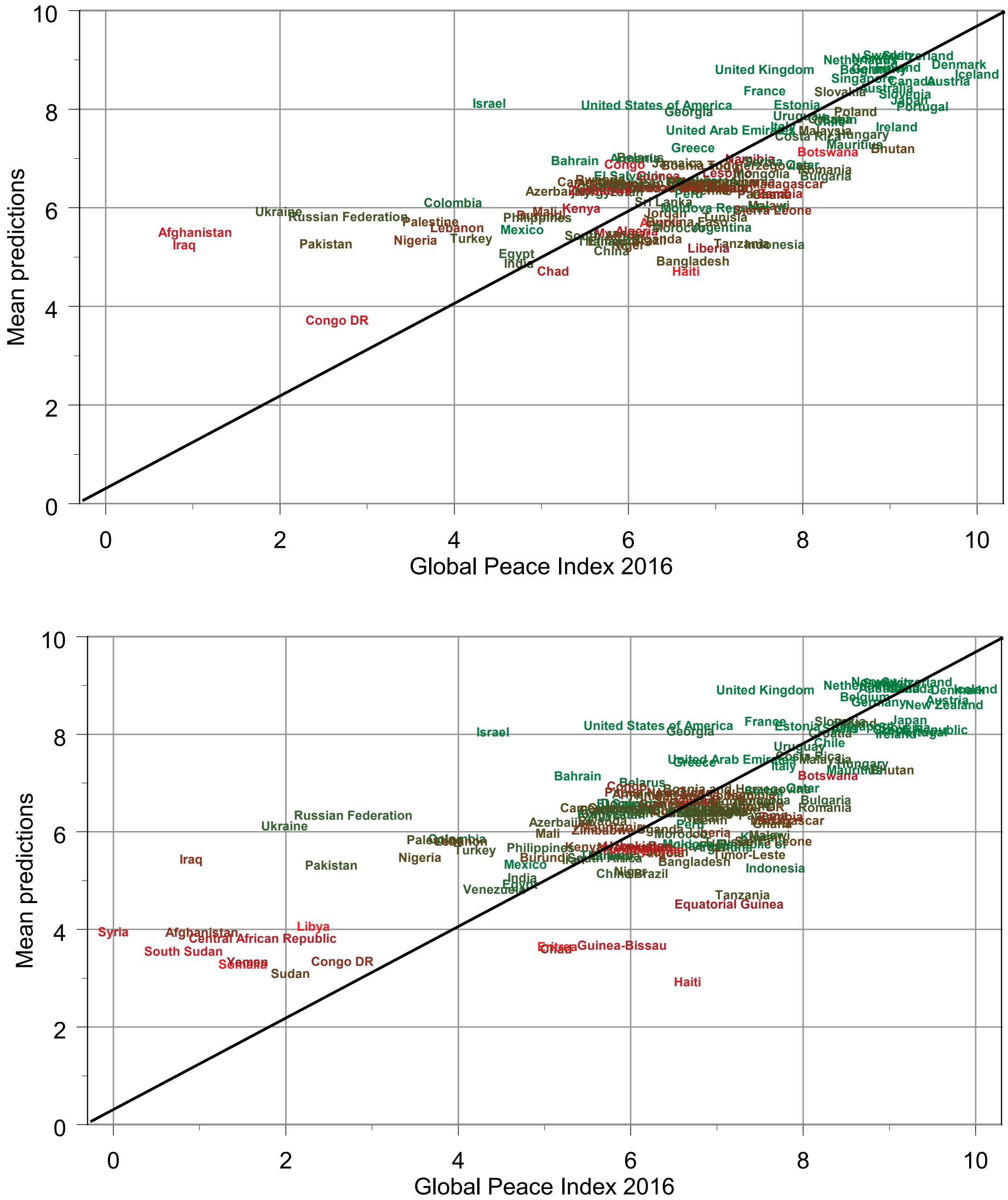
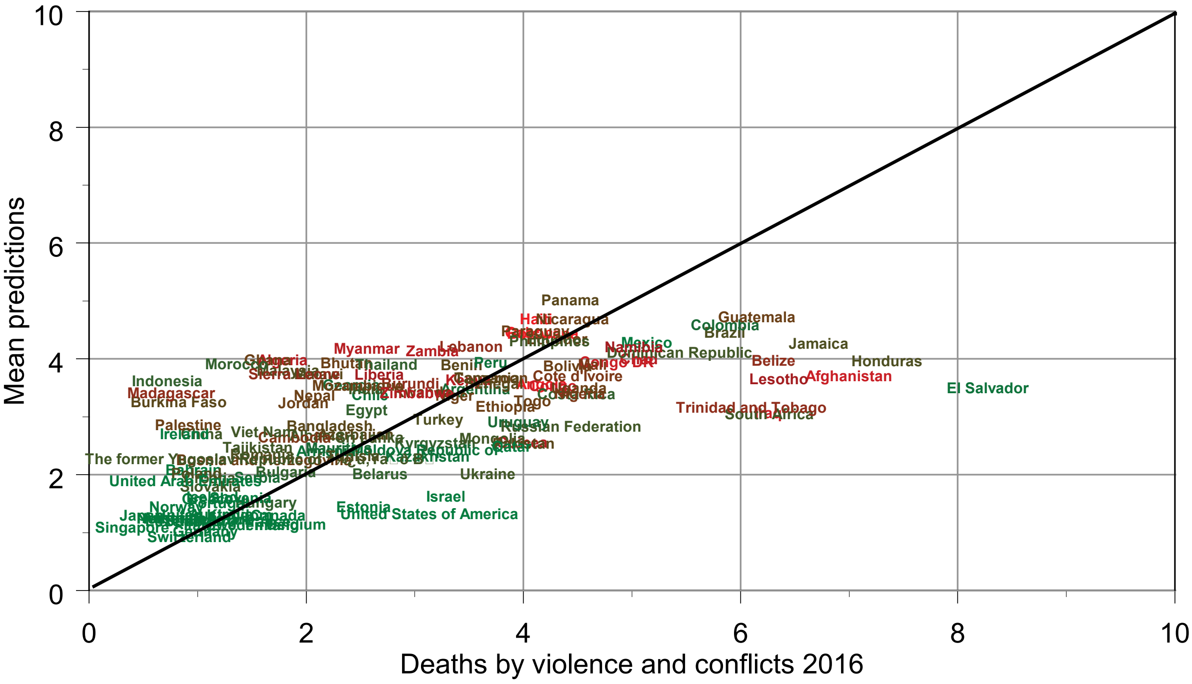


Figure D.1 A: prediction scatterplot for the Global Peace Index as dependent variable (129 countries). Explained variance is 49%. B: idem where missing driver data were imputed (1.6% of all driver data were missing, 157 countries). Explained variance now is 56%. Colors from green to red correspond to the World Bank statistical capacity indicator (red meaning low capacity).



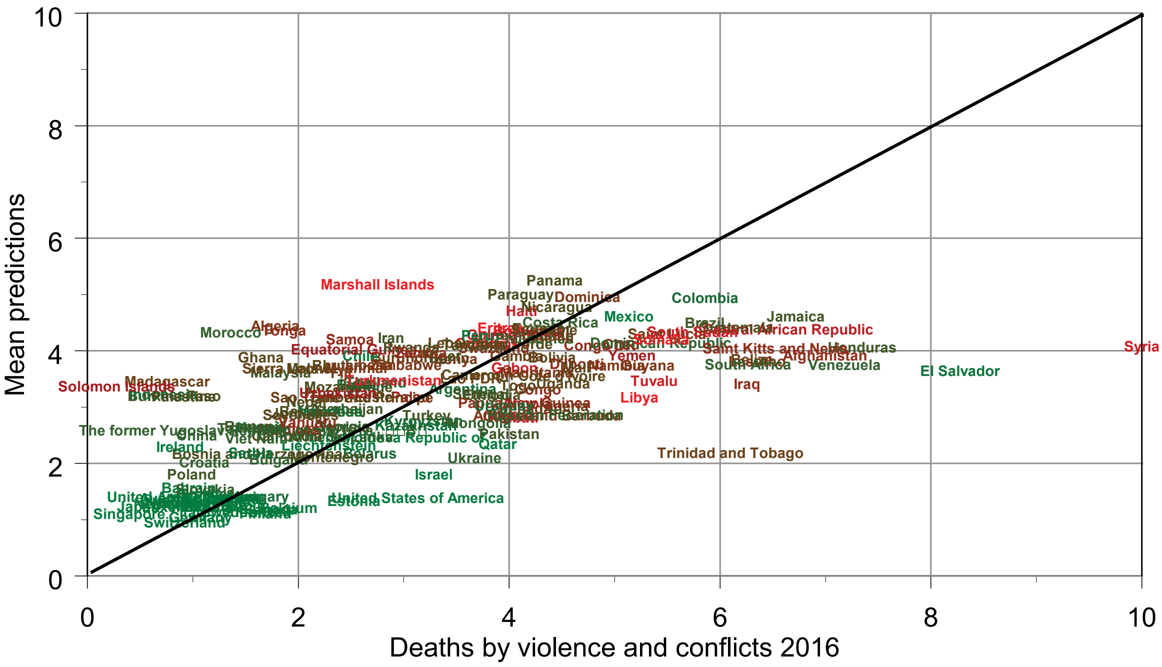
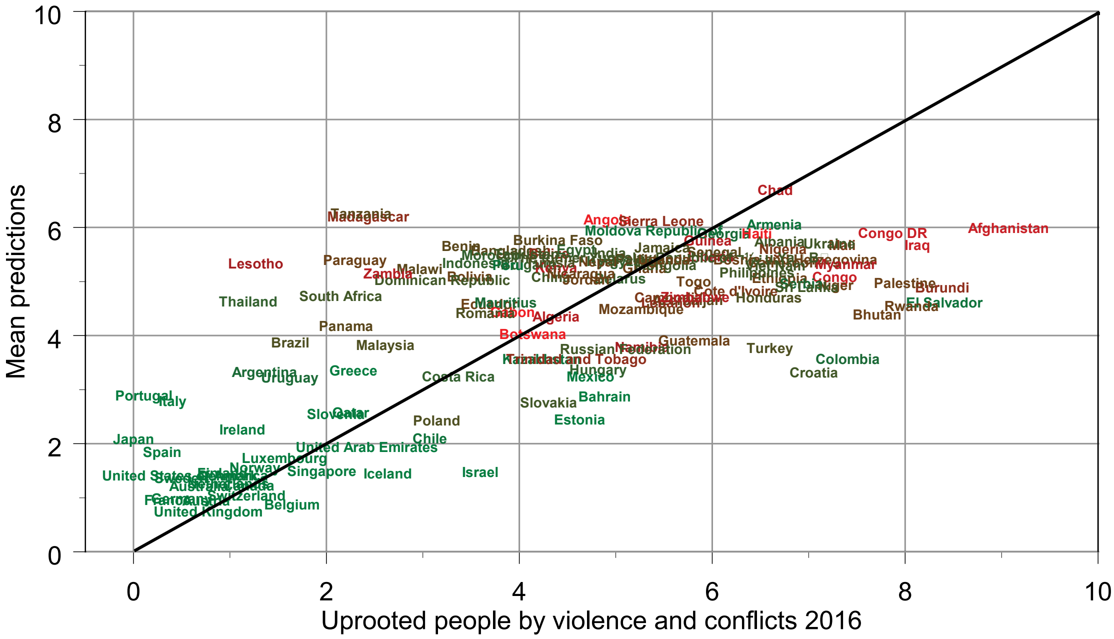


Figure D.2 A: prediction scatterplot for Deaths by conflict and violence as dependent variable (134 countries). Explained variance is 41%. B: idem where missing driver data were imputed (3.6% of all driver data were missing, 188 countries). Explained variance now is 38%. Colors from green to red correspond to the World Bank statistical capacity indicator (red meaning low capacity).



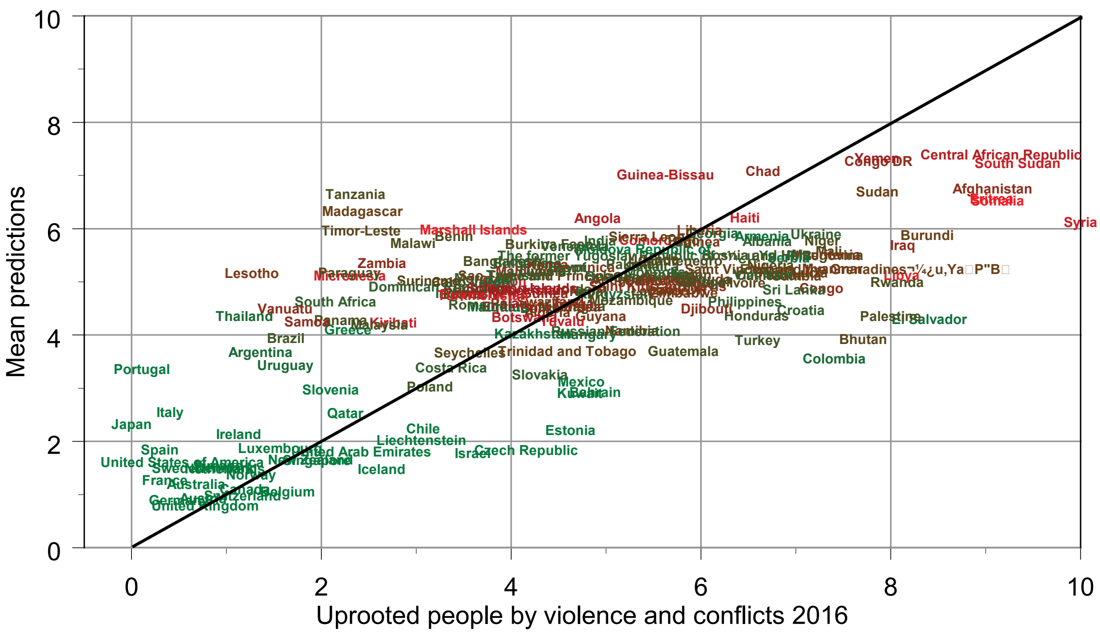


Figure D.3 A: prediction scatterplot for Uprooted people by violence and conflict as dependent variable (134 countries). Explained variance is 46%. B: idem where missing driver data were imputed (3.7% of all driver data were missing, 191 countries). Explained variance now is 47%. Colors from green to red correspond to the World Bank statistical capacity indicator (red meaning low capacity).

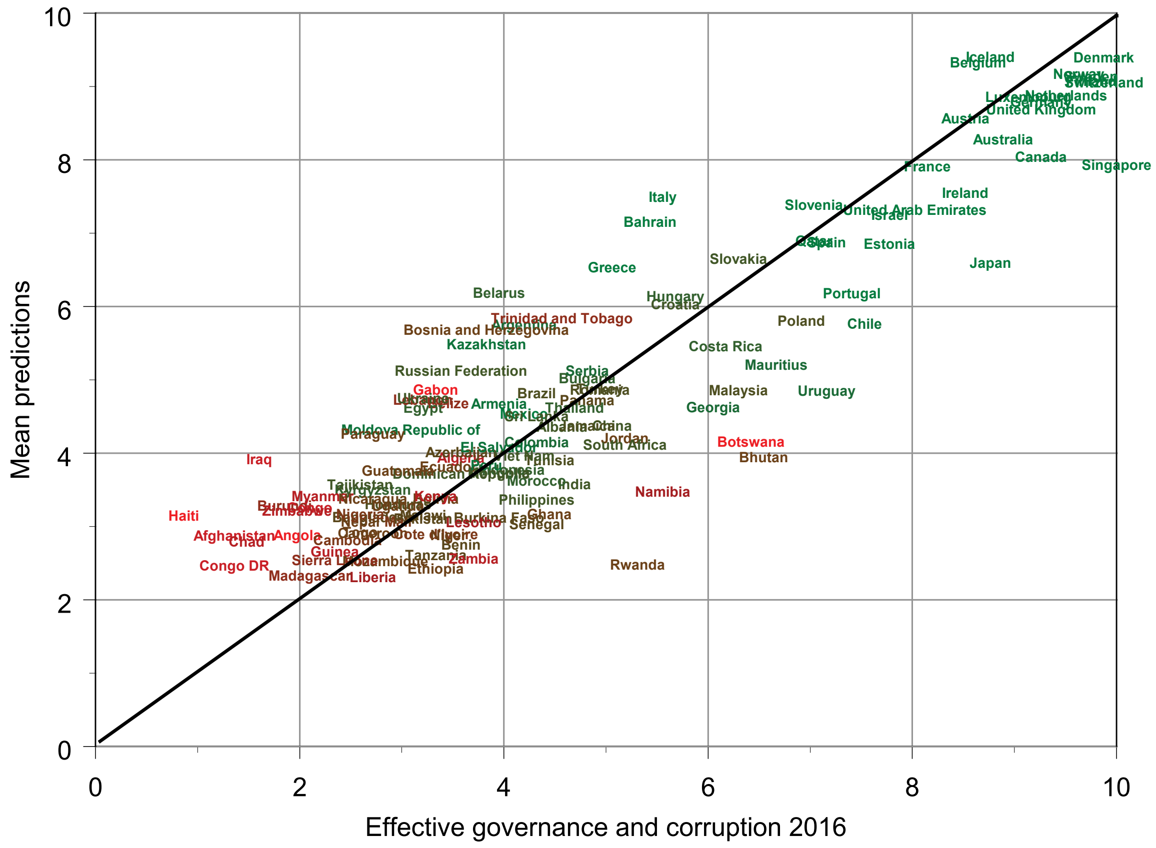


Figure D.4 Prediction scatterplot for Governance as the dependent variable (131 countries). Explained variance is 79%. Colours from green to red correspond to the World Bank statistical capacity indicator (red meaning low capacity).

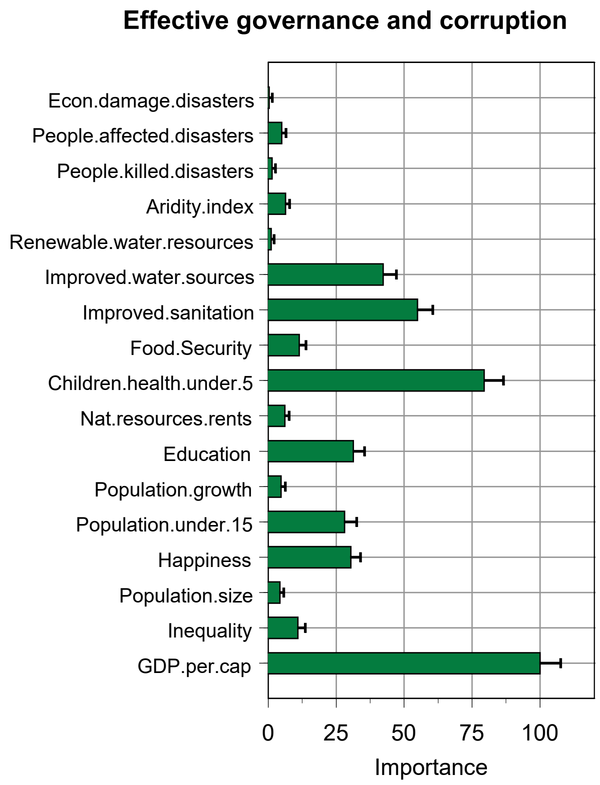


Figure D.5 Random Forest model with **Governance** as the dependent variable. Explained variance: **79%.** Analysis is for 131 countries that have data for all the indicators considered.

Appendix E Description of the drivers as taken from Mach et al. (2019)

Table E.1 Three main conflict drivers identified by Mach et al. (2019). These are the upper three drivers shown in Figure 6.3.

Conflict drivers	**Importance for conflicts to date and uncertainties**
Low socioeconomic development Definition**:** the well-being of people and the opportunities they have to be economically productive, often proxied by GDP per capita among other measures	• Low socioeconomic development, as GDP per capita, is one of the best predictors of intrastate conflict onset and incidence. It is the single most robust covariate in cross-section and time series. [Evidence: *robust*; Agreement: *high*] • However, there is uncertainty about why, including correlation versus causation—the degree to which socioeconomic development as per capita GDP is directly related to conflict risk as compared to proxying for other mechanisms (e.g., economic shocks, grievances, low state capability, recent history of violent conflict). Isolating the role of socioeconomic development from other correlated factors is empirically challenging. [Evidence: *medium*; Agreement: *low*] • Low socioeconomic development, especially in combination with inequalities, could increase grievances and motivations for violence. Such effects are especially relevant to people who are not the poorest (e.g., extremely poor populations can't afford military equipment). [Evidence: *medium*; Agreement: *low*] • The opportunity cost theory of rebel mobilisation posits that rebel leaders are more able to recruit and fund soldiers in a society where alternative means of income are scarce. [Evidence: *medium*; Agreement:*medium*] • Low socioeconomic development interacts with and can contribute to low state capability (e.g., via low tax revenue), another underlying driver of conflict. Low state capability reduces efficient provision of services and goods thereby increasing grievances, it limits socioeconomic development (e.g., via the absence of effective, impartial political and legal institutions and underpinning bureaucracy), and it also reduces the projection of authority, including policing and monitoring capabilities and accommodation of claims. [Evidence: *medium*; Agreement: *medium*] • Conflict traps undermine both economic development and state capability. [Evidence: *robust*; Agreement: *high*]
Low state capability Definition: low coercive capability, limited ability to regulate and distribute power among claimants, and low bureaucratic capability	• Intrastate conflict is concentrated in weak states. [Evidence: *robust*; Agreement: *high*] • Low state capability together with political shocks explains much historical civil war. New states forming from former colonies have low state capability (e.g., given little history of taxation and local authority) and new political competition (e.g., given new authority such as a UN seat or profits from cash-crop marketing). There is associated vulnerability of the center following political shocks that favor rebel groups or reduce capability of the center. [Evidence: *medium*; Agreement: *medium*]

to provide services and goods. Low state capability can be linked to low socioeconomic development and the political context.	• State ability to project power across its territory, with coercive capability to defeat mobilizing rebel groups, is important because small rebel groups can do enormous, long-lasting damage. [Evidence: *medium*; Agreement: *high*] • Ineffective or absent formal institutions for power sharing and transfer increase reliance on violence (e.g., informal bargains backed by threat of violence). [Evidence: *medium*; Agreement: *medium*] • Political stakes are high in weak states because access to services, resources, and security requires access to political power, with winners and losers. (Expert input: 1, 6) [Evidence: *medium*; Agreement: *medium*] • In weak states, inefficient provision of services, including their role in buffering shocks, leads to grievances for which the state doesn't have administrative capacity or willingness to learn about, address, or suppress. (Expert input: 4, 6, 8, 9) [Evidence: *medium*; Agreement: *medium*] • There is uncertainty because it is difficult to isolate the role of state capability from other correlated factors (e.g., GDP per capita is a poor proxy for state capability, even though it has been used as such). [Evidence: *robust*; Agreement: *medium*]
Intergroup inequality Definition: horizontal, systematic inequality among groups	• Intergroup inequality across multiple dimensions (e.g., political, economic, ethnic, social) is robustly linked to conflict. [Evidence: *robust*; Agreement: *high*] • Inequalities among groups can drive formation of conflict parties and enable mobilization, through identifiable differences and collective identities motivating self-sacrifice. For example, it can be easier to mobilize groups along ethnic lines, even if the grievance itself is political or economic, unrelated to ethnicity. [Evidence: *medium*; Agreement: *high*] • Different societal cleavages can serve as the foundation for identity groups challenging the state. Across such potential cleavages, ethnicity has been prominent in contemporary conflicts, perhaps because it is detectable, permanent, and often geographically clustered (e.g., far from the capital with less access to power and less state monitoring and suppression). Such conclusions have emerged despite uncertainty stemming from poor measures and conceptualization of ethnic diversity. [Evidence: *medium*; Agreement: *medium*] • Political inequality, including exclusion from power, can provide a basis for conflicts. For example, less access to services and resources, decreased economic prospects, or greater vulnerability to state predations can lead to mistrust or mobilization to increase share of power. [Evidence: *medium*; Agreement: *high*] • Inequality in the size, strength, and capability of groups contributes to conflict in determining groups' threat capacity and associated access to power. Lack of threat of force to hold government accountable increases conflict likelihood for excluded groups. [Evidence: *limited*; Agreement: *medium*] • Economic inequality among groups can also increase conflict risk, sometimes interconnected with political inequality (e.g., via systematic exclusion from employment opportunities) or exacerbated grievances, noting that the poorest groups may not have sufficient

	capacity to mobilize. [Evidence: *medium*; Agreement: *high*] • There is uncertainty, especially around correlation versus causation, because intergroup inequality is slow changing, difficult to measure, and difficult to isolate from other conflict drivers (e.g., ethnic differences versus economic conditions). [Evidence: *medium*; Agreement: *low*]