



Bayesian Hierarchical Modeling of Historical Irrigation Patterns

A Master's Thesis

Presented to

Albrecht Daniel Thaer Institute for Agricultural and Horticultural Sciences
Faculty of Life Sciences
University of Humboldt

In Partial Fulfillment
of the Requirements for the Degree
Master of Science

Rachel Ledig
602220
ledigrac@hu-berlin.de
Born Nov 17, 1991 in California, USA

First Examiner

Dr. Tobias Krueger

Second Examiner

Dr. Christian Schleyer

Advisor

Dr. Fabian Stenzel

December 2021

I, Rachel Ledig, hereby declare that the present thesis has not been submitted as a part of any other examination procedure and has been independently written. All passages, including those from the internet, which were used directly or in modified form, especially those sources using text, graphs, charts or pictures, are indicated as such. I realize that an infringement of these principles which would amount to either an attempt of deception or deceit will lead to the institution of proceedings against myself. All errors are my own. *Berlin, December 2021*

Rachel Ledig

Approved for the Division
(Faculty of Life Sciences)

Dr. Tobias Krueger

Dr. Christian Schleyer

Table of Contents

| | |
|--|-----------|
| Chapter 1: Introduction | 1 |
| Chapter 2: Context | 3 |
| 2.1 Modeling Studies of Irrigation Expansion | 3 |
| 2.2 Establishing a Framework for Irrigation Expansion | 4 |
| 2.2.1 Drivers | 4 |
| Chapter 3: Methods | 11 |
| 3.1 Study Region | 11 |
| 3.2 Data | 12 |
| 3.2.1 External Data | 13 |
| 3.2.2 LPJmL Data | 16 |
| 3.3 Hypothesis | 21 |
| 3.4 Data Processing and Modeling | 22 |
| 3.4.1 Thinning | 22 |
| 3.4.2 Standardization | 22 |
| 3.4.3 Time Series Considerations | 22 |
| 3.4.4 Hierarchical Modeling | 22 |
| Chapter 4: Appendix | 25 |
| 4.1 Proof of Bayes' Theorem for Two Hypothesis from Hacking (2001) | 25 |
| 4.1.1 Bayes Theorem | 26 |
| 4.1.2 Model Comparison | 29 |
| 4.2 Digression: Covid Rapid Antigen Tests | 29 |
| 4.3 Predictor Selection | 30 |
| References | 31 |

List of Tables

| | |
|--|----|
| 4.1 Probabilities for incidence of Covid and rapid antigen test sensitivity. | 30 |
|--|----|

List of Figures

| | | |
|-----|--|----|
| 2.1 | Furrow, Sprinkler and Drip irrigation [@GoogleImages2021] | 5 |
| 3.1 | Study Area | 12 |
| 3.2 | Example calculations of Terrain Ruggedness Index (TRI) for differing geographies. | 13 |
| 3.3 | Terrain Ruggedness Index values for the study area. | 14 |
| 3.4 | Governance regimes of the study area in 2005. | 16 |
| 3.5 | Population density | 18 |
| 3.6 | Irrigation fraction per grid cell of the study area in 2005. | 20 |
| 3.7 | Histogram of the outcome variable, Irrigation Fraction (IF) for the entirety of the study region and period (1960 to 2005). | 21 |
| 3.8 | Directed Acyclic Graph for Beta | 23 |
| 3.9 | Directed Acyclic Graph for zero inflated parameter z_i | 24 |
| 4.1 | Effects of the prior distribution on the unnormalized posterior with a $\text{Normal}(0.5, 0.25)$ likelihood. First row: Uniform prior, Second row: $\text{Beta}(a = 0.75, b = 0.75)$ prior, Third Row: Exponential prior. Prior and likelihood are multiplied together to create the unnormalized posterior seen in the third column. | 27 |

Abbreviations

| | |
|-----|--------------------------|
| IF | Irrigation Fraction |
| DAG | Directed Acyclic Graph |
| TRI | Terrain Ruggedness Index |
| DEM | Digital Elevation Model |

Abstract

The preface pretty much says it all.

Second paragraph of abstract starts here.

To Mom, Dad, Sis, and Fi.

Chapter 1

Introduction

The practice of irrigation, and its subsequent impacts, have far reaching effects for the global economy, food and water supply, and environment. Irrigated land accounts for 20% of the global cultivated land area which derives 40% of the global food supply Portmann, Siebert, & Döll (2010), 70% of water withdrawals from ground water aquifers are used for irrigation (FAO, 2011), and

Ultimately, the agricultural system will continue to demand more water while competing with other sectors over decreasing water availability in the coming years. With this pressure, farmers will aim to be as efficient as possible by increasing their yields which, in certain areas and under certain conditions, can be done by irrigating their fields. Understanding where, when, and why farmers decide to begin irrigating is important to understand water demands now and in the future. Unfortunately, irrigation expansion is seldom studied, with few studies using statistical modeling techniques to investigate the role that biophysical, economic, and political factors play in the spread of irrigation infrastructure in the present. None do so historically. This leads to the first and overarching research question:

Which factors (drivers) have influenced the expansion of irrigation in both time and space over the past 100 years?

By understanding under which conditions irrigation infrastructure is built could be coupled with projections of future climate and economic scenarios, allowing us to get a better idea of scales of agricultural production, land use, and water needs in the coming years. In order to eventually extrapolate to the future, predictive accuracy needs to be addressed. This can be done by separating data into training and testing datasets, model tuning, and error estimation.

Several other questions smaller questions arise in regard to drivers of expansion which are interesting to investigate. Such as:

- Which has a larger influence on irrigation expansion patterns, biophysical or socio-economic factors?
- Do rates of irrigation expansion differ depending on crop type?

Both of these questions have interesting policy implications, should large trends be present in the model and analysis.

The time series component of this data set also gives rise to some interesting questions.

- Expansion rates of total irrigated area differ temporally, what could be the drivers for this difference?
- Is there a temporal delay in irrigation expansion based on the influence of certain drivers, and if so how much?

Chapter 2

Context

This chapter details the theoretical basis for this thesis. Initially, literature will be reviewed beginning with irrigation expansion drivers in Section 2.2 and followed by a detailed account of prior irrigation modeling studies in Section 2.1. Finally, an account of the basics of Bayesian inference will be described in Section ??.

2.1 Modeling Studies of Irrigation Expansion

Shifting gears, there are few studies that investigate the drivers or irrigation expansion. Those that do, only one takes in to account historical data or detail expansion rates based on regions or countries. Stefan Siebert et al. (2015), after collecting a global dataset of historical irrigation patterns, goes on to investigate trends and patterns in global irrigation. The authors note that rates of irrigation expansion can be divided into two main time periods, before and after 1950, documenting that from 1900 to 1950 global irrigation doubled and from 1950 to 2005 global irrigated area tripled. Different regions also experienced different growth rates, with regions such as Australia and Oceania, southeastern Asia, Middle and South Africa, Central America, and eastern Asia experiencing growth rates higher than that of the global average and the North of Africa and North America, experiencing lower than average growth rates. Europe and Western Asia also stand out here with a slow decrease of area equipped for irrigation seen toward the end of the study period, confirming what Angelakis et al. (2020) mentions regarding the disintegration of irrigation schemes after the dissolution of the USSR. In general Stefan Siebert et al. (2015) conclude that in 1900 irrigation took place predominately in arid locations, which expanded into more humid climes as the study period progresses, in fact, the percentage of global irrigation that occurred in arid regions remained relatively stable from 1900 to 2005, as where the percentage of global area equipped for irrigation went down in areas with rice cultivation and up in other wet areas. Ultimately, Stefan Siebert et al. (2015) connect changes in irrigation to several variables including population, aridity, and river discharge.

Neumann et al. (2011) attempt to model current irrigation patterns based on a variety of biophysical factors (slope, discharge, humidity, ET, evap, pop density, etc) and political factors (democracy, GDP, corruption, political stability) at two different

levels, grid cell (5 arc-minute grid cells) and at a country level. However, Neumann's study is only modeling the current irrigation pattern, with no historical perspective. The authors find that the addition of socio-economic information to the model does help improve the explanation the current irrigation pattern, but not only to a small extent. A delay existed, as current patterns are formed from decisions, policies, and economic conditions that occurred years before.

Puy, Lo Piano, & Saltelli (2020) discuss uncertainties of published projected irrigation expansion patterns by comparing these projections to a simple model that predicts irrigated area as a function of only population size, bounded by available water and land. A simple model was constructed for ease of interpretation, as the main purpose of this study is to discuss uncertainties present in this (seemingly) simple model and investigate how other published projections of irrigation expansion fall within Puy, Lo Piano, & Saltelli (2020)'s model's range of uncertainty. Puy, Lo Piano, & Saltelli (2020)'s justification for modeling irrigated area as a function of only population size stems from the previous studies showing that population is a variable that encapsulates much more complicated socio-ecological relationships.

2.2 Establishing a Framework for Irrigation Expansion

2.2.1 Drivers

Understanding the mechanisms behind irrigation and the features of irrigation systems is imperative to understanding why, when, and where irrigation is implemented. According to the FAO & Walker (1989), the choice by the farmer of irrigation technology is dependent on several intertwined socio-economic and biophysical concerns. Multitudes of literature exist detailing the considerations necessary when choosing an irrigation system, but little exists on why farmers choose to irrigate in the first place. Although the distinction is notable and directly equating the choice of a specific irrigation technology to the decision to begin irrigating is perhaps naive, it is important to understand that the decision to irrigate *is not possible* without the simultaneous choice of an irrigation technology. Therefore using the irrigation technology selection criteria outlined by FAO & Walker (1989) as a basis to discuss drivers of irrigation expansion is justified.

By using these areas as a guideline for what farmers are concerned about when establishing an irrigation system and why those concerns exist, not only can the causal relationship between external conditions and irrigation expansion be investigated, but appropriate predictors for the hierarchical model can also be selected. These eight components are detailed in the subsections below.

To begin, it is important to know a little bit about irrigation systems. Irrigation systems can be divided into two main systems: gravity fed which includes all types of surface irrigation, and mechanized pump systems, such as sprinkler or localized distributions (Bjorneberg, 2013). Surface irrigation, such as furrow, basin, or border irrigation, is the oldest, simplest, and cheapest method available and used on around

85% of the irrigated land in the world (Bjorneberg, 2013; Holzapfel & Mariño, 2008). Pumped irrigation systems are less common globally. Different types of irrigation can be seen in Figure 2.1.



Figure 2.1: Furrow, Sprinkler and Drip irrigation [@GoogleImages2021]

Water

Water supply is a core component of irrigation and can be divided into two main dimensions, water quantity, and water quality. As water quality matters more for the choice of a specify irrigation technology, it will not be discussed in detail below. Water quantity, however, is a large concern when deciding if a farmer can irrigate. In the natural world, water comes from three sources: rain, surface water (rivers, lakes, reservoirs, etc.), or groundwater (aquifers).

Precipitation, specifically rainfall, is not technically an irrigation water source as it is applied directly to crops without any irrigation system. As a result, rainfall, a free source of crop water, currently sustains roughly 80% of the global cultivated area which produces 60% of the total crop production (Program & UNESCO, 2009). Although direct rainfall is the predominant method for watering crops and can nullify the need for any irrigation technology, meeting crop water demand with rainfall can be difficult as rainfall can be extremely varied in both space and time. Spatially, different regions receive different annual amounts of rainfall, with tropical and temperate zones receiving more rainfall than semi-arid or arid regions (Forseth, 2010). However, mean annual rainfall does not show the whole picture, and when deciding whether the amount of precipitation in a region is sufficient to sustain crop growth, it is important to remember that rainfall can heavily vary in seasonality, as well. Wet and dry seasons dictate the boom and bust of vegetation in some tropical and sub-tropical regions [CITATION]. The coincidence of rainfall and crop growing season is integral for farmers as crop water demands must be met during the specific periods of the crop's growing cycle. One benefit of rain-fed agriculture is that rainwater is naturally of high quality, meaning that it is free of solids and harmful substances that other water sources may have (**douglascoxWaterQualityCrop2015?**). Although, it should be noted that patterns of rainfall are shifting with the effects of climate change. When long-standing climactic patterns change, differences in annual rainfall amounts, duration of the wet and dry seasons, and intensity of rainfall have consequences for both farmers and

crops, making planning for a productive growing season more difficult in some areas (Rockström et al., 2010). These changes have the potential to incentivize the shift from rainfed agriculture to irrigated.

Surface water is only applied to crops through an irrigation system, unlike rainwater. Surface water is water that is collected, from precipitation, runoff, and up-welling of groundwater resources, on the surface of the planet (in rivers, lakes, ponds, and dams). Surface water is the most common water source for irrigation as its supply is generally less variable than precipitation and easier to access than groundwater. As a result, 54% of the area available for irrigation is irrigated with surface water (S. Siebert et al., 2010; Thenkabail et al., 2009). Surface water, due to the fact that it is accumulative, potentially passes through many environments before it is extracted by the farmer and applied to the field, meaning that issues with water quality are more common. Pollution, sediment, and salts must be monitored to ensure that the surface water can be applied to crops and is able to pass through the specific irrigation distribution system (FAO & SAFR, 2001). Again, it should be pointed out that since the original source of most surface water is precipitation in some form, it is also susceptible to changes in precipitation patterns due to climate change.

Groundwater, water that is stored below ground in aquifers, is a collection of water that has infiltrated through the soil over many years. As a source of clean reliable irrigation water, groundwater is unparalleled (S. Siebert et al., 2010). The water stored in aquifers is of high quality and much more resistant to seasonal changes in precipitation than surface water. Groundwater's largest issue, with respect to its use as a source of water for irrigation, is access. Multiple methods for groundwater extraction exist but most are more costly than surface water withdrawals. However, in semi-arid or arid regions where surface water may also be seasonal or non-existent, extraction of groundwater for irrigation becomes more economically viable as its use can sustain large populations (E.g. regions in the middle east).

Ultimately, rainfall is always the default source for farmers to supply water to their crops, and only when rainfall is not sufficient during necessary parts of the crop growing season, do farmers consider using irrigation to meet the crop water demand.

Topography

Steep topography can prevent the implementation of the majority of most common and basic surface irrigation systems, such as furrow (using long channels to water crops) or basin (flooding of field sections) irrigation. Generally, for this type of irrigation to be used field slopes need to have less than a 5% incline, which limits the application of this simple type of irrigation to valleys and flatlands (FAO & SAFR, 2001).

Other types of irrigation, sprinkler or localized systems, can be utilized in steeper environments, but are often more costly than simple surface irrigation techniques. There are some ingenious examples of farmers working with topography to use simple irrigation methods and adapt to steeper slopes: rice terraces in South East Asia use basin irrigation to cultivate paddy rice on extremely steep slopes (Adachi, 2007). However, in general, steep slopes are rarely irrigated with surface irrigation methods.

Soil

Soil composition, in the context of irrigation, is important because different soils have different capacities to store and infiltrate water, meaning that different types of irrigation systems may be necessary to ensure that crops are able to get the water that they need (FAO & Walker, 1989). For example, sandy soils retain less water and drain faster, meaning that they must be irrigated more frequently but with less water. In this case, surface irrigation methods that apply large amounts of water to a field at one time may be less suited for these soils (FAO, Brouwer, & Heibloem, 1986). In addition, as the soil type dictates the infiltration rate of irrigation water and therefore the rate of application, the soil type combined with the specific irrigation method chosen may impose requirements on the consistency and volume of the water supply. For the context of this thesis, soil composition will be largely ignored as the oftentimes soil composition limits the type of irrigation used, not the absence or presence of it. However, it is important to note that as surface irrigation methods, which are more common worldwide, are less suitable for certain soil types (E.g. sandy soils), there may be circumstances in which given a specific soil type, irrigation would be inefficient.

Crops

Crop type also has the potential to influence the presence/absence of irrigation and the type used as different crops require different amounts of water over the growing period (FAO, Brouwer, Prins, Kay, & Heibloem, 1986). Staple crops that need to be irrigated are rarely irrigated with sprinkler or localized (drip) systems as these systems are much more costly than surface irrigation. Sprinkler and localized systems are almost always reserved for fruits and vegetables, so-called cash crops (FAO, Brouwer, Goffeau, & Heibloem, 1985).

Compatibility

Irrigation technologies must be a benefit not a burden to the farmer and the choice to irrigate must fit within the given growing and harvesting needs of a particular crop. Irrigation must not impede the usage of pruning or harvesting machinery (FAO & Walker, 1989). In addition, the farmers must be able to work with the chosen irrigation technology. Sprinkler and localized systems are more complicated than surface irrigation, however, surface irrigation requires more manual input from the farmer. Surface irrigation is also easier to maintain than other types of irrigation (FAO, Brouwer, & Heibloem, 1986). This is less of a concern for this thesis, as compatibility issues between crop type, irrigation system, farmer knowledge, and other farm machinery is not the main focus.

Economics

The economics of specific irrigation systems and the choice to irrigate is important for farmers, as even the simplest surface irrigation methods are not without costs. Water, energy, materials, labor, maintenance, and monitoring costs must all be considered

when deciding to begin irrigating and which irrigation system to choose (FAO & Walker, 1989). According to FAO & SAFR (2001), for small farms in Sub-Saharan Africa, up to 80% of the costs to develop an irrigation system stems from water resource procurement, such as construction of a dam. In addition, the authors note that the cost for cubic meter of water is increasing in this region as well, putting more economic pressure on the farmers.

Social Influences

However, there are ways to shift the costs of establishing irrigation from the individual farmer to the collective. Grass-roots, bottom-up irrigation cooperatives or schemes may be set up to share the construction, maintenance, and monitoring of shared irrigation and water transportation infrastructure (FAO & Walker, 1989). FAO & Walker (1989) note that there are often several levels involved in the organization of irrigation schemes. These include the individual farmers who participate at the field level; the local farmers' collectives who manage operation, maintenance, allocation, and conflicts; the governing body at the local or state level that deals with water distribution and funding. Each actor has different needs, power, and roles ensuring the management of local irrigation. Examples of irrigation schemes are common globally and historically.

External Influences

Sometimes external factors play a large role in stimulating irrigation expansion. Social trends, governmental programs, and conflicts among others can influence the rate of irrigation expansion at a sub-national, national, or regional level. This type of expansion could be labeled as top-down, as generally there are larger actors pushing or incentivising the construction of irrigation infrastructure, in comparison to the bottom-up dynamic of locally organized irrigation schemes.

Angelakis et al. (2020) details explicitly some of these trends that have occurred during the last century. In Spain, during the reign of Francisco Franco,¹ the government began a national water development plan to settle small farmers. This campaign created the foundation for Spain's expansive irrigated agricultural system today, the largest irrigated area in the European Union. Another instance that Angelakis et al. (2020) discusses is the sudden overall increase in irrigated area after the Second World War, which is attributed to two aspects: a sudden population increase and developments in technology. The population increase, due to increased birthrates and an extension of life span due to medical advances, created a demand for more food and advances in technology, specifically those that were necessary to optimize pumps needed for pressurized irrigation and water transportation systems. Both of these changes resulted in a rapid global expansion of irrigated area. Other examples exist such as the Green Revolution, or Third Agricultural Revolution,² also spurred trends

¹A point estimate is a single value, generally the average, given as the estimate value for a parameter.

²At the equator, this is roughly a 9.2km by 9.2km grid cell resolution.

of irrigation expansion, particularly in Asia (Angelakis et al., 2020). Angelakis et al. (2020) points out that large scale irrigation projects often succeed due to governmental involvement and support from changing market patterns.

Ultimately, the drivers of irrigation are complex and interconnected. When quantifying these forces behind an ever changing pattern of irrigation it is easy to generalize, but the reality of irrigation is that there are infinite combinations of conditions, contexts, times, and ambitions that determine the ultimate success or failure of an irrigation project. What worked in one place and time, is not guaranteed to result in a positive outcome years later and worlds apart. The above sections are not exhaustive in their description of the complexity that is involved in the planning, construction, and operation of the majority of irrigation projects, but they are a good start.

Chapter 3

Methods

3.1 Study Region

To be able to investigate the patterns of global historical irrigation expansion, the initial proposed study area encompassed the entirety of the globe's terrestrial land mass with recorded observations detailing every year for the 45-year study period. This constituted a data set of roughly three million observations. In an effort to decrease computation time, a logical reduction of the study area (and therefore data) was necessary.

Figure ?? illustrates the pattern of irrigation at the last time step (2005) of the study period. It is notable that irrigation does not exist at all in some places such as the higher latitudes and the Sahara. These areas also contain no cropland, no managed grasslands, and no pastureland. Upon the logic that irrigation cannot expand into areas that contain no agricultural land (whether it be cropland, grassland, or pasture land), all cells which contained no agricultural land over the course of the study period were removed, giving rise to the study area you see in figure 3.1.

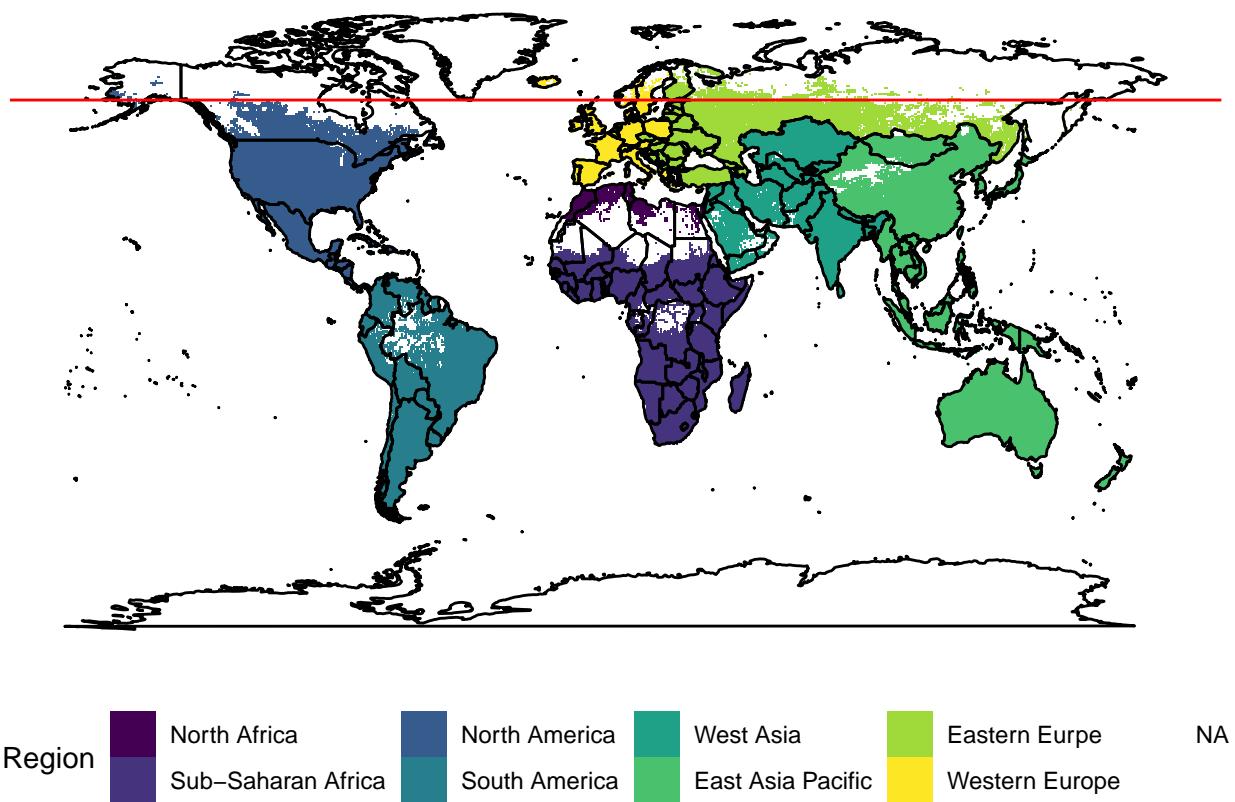


Figure 3.1: Final study area after the removal of cells with no agricultural land over the course of the 45 year study period. Colors represent regional groupings. Notice that excluded areas are located within the Amazon, the Sahara, Central Africa, the Middle East, Eastern China, and toward the Arctic Circle (represented at 60 degrees North with a red horizontal line).

3.2 Data

To appropriately choose data that encapsulates the concepts and relationships described in Section 2.2 the following datasets and variables were collected and selected. It should be noted that many of these datasets contain more than one variable that would conceptually represent the links between expanding irrigation and various drivers. The predictors presented below were selected in an iterative model building process for various reasons including data availability (both spatially and temporally), modeling constraints, and conceptual fidelity. For more information on all potential variables please see Appendix Section 4.3.

The primary variable of interest in this masters thesis, irrigation fraction, along with several other predictor variable were used as input for or were extracted from the Lund-Potsdam-Jena managed Land (LPJmL) model. Other datasets for various predictors were gathered from datasets external to those used in LPJmL. Detailed

descriptions of the acquisition of all data sets are in the sections to follow.

3.2.1 External Data

Topography: Ruggedness

In order to explain the relationship between topography and irrigation expansion, an index of a country's topographical heterogeneity was chosen. Other similar irrigation expansion studies similarly use slope (calculated from Digital Elevation Models (DEMs)) as a predictor for explaining irrigation expansion but do so using a 5 arcmin gridded slope dataset (Neumann et al., 2011). Unfortunately for this thesis, grid cells used for analysis are significantly larger¹, and an aggregation of slope to a larger grid cell was unfeasible for this thesis. Instead, a well known index of country topographical heterogeneity was chosen to represent the “ruggedness” of a country.

Terrain Ruggedness Index The Terrain Ruggedness Index (TRI) was developed by Riley, Degloria, & Elliot (1999) initially to be used to analyze the effects of topography on species habitats and behaviors. The authors developed a simple method to calculate the heterogeneity of a given area using DEMs. To do so, the sum change in elevation is calculated between a cell and the elevation of its eight nearest neighbors. These individual cell level TRIs can then be averaged for a given region, in this thesis's case, a country (Riley, Degloria, & Elliot, 1999). Adapted from Riley, Degloria, & Elliot (1999), a representation of different topography types and the respective TRI scores is displayed in 3.2. The exact dataset used for country level TRI is from Nunn & Puga (2012).

| (a) | (b) | (c) |
|-----------------|-----------------|-----------------|
| 100 125 100 | 200 175 200 | 200 210 220 |
| 125 200 125 | 175 100 175 | 210 225 225 |
| 100 125 100 | 200 175 200 | 205 210 220 |
| TRI = 700 | | |
| TRI = 700 | | |
| TRI = 100 | | |

Figure 3.2: TRI values a) peak geography, b) bowl geography, c) rolling hill geography. Values presented in the tables represent elevations. Notice that for both a) and b) the TRI is the same despite very different geography, but both values are much higher than c) which is a more homogenous landscape.

Description In Figure 3.3 the Terrain Ruggedness Index for the study region can be seen. Higher values indicate countries with more topographical heterogeneity, and

¹A point estimate is a single value, generally the average, given as the estimate value for a parameter.

lower values represent less.

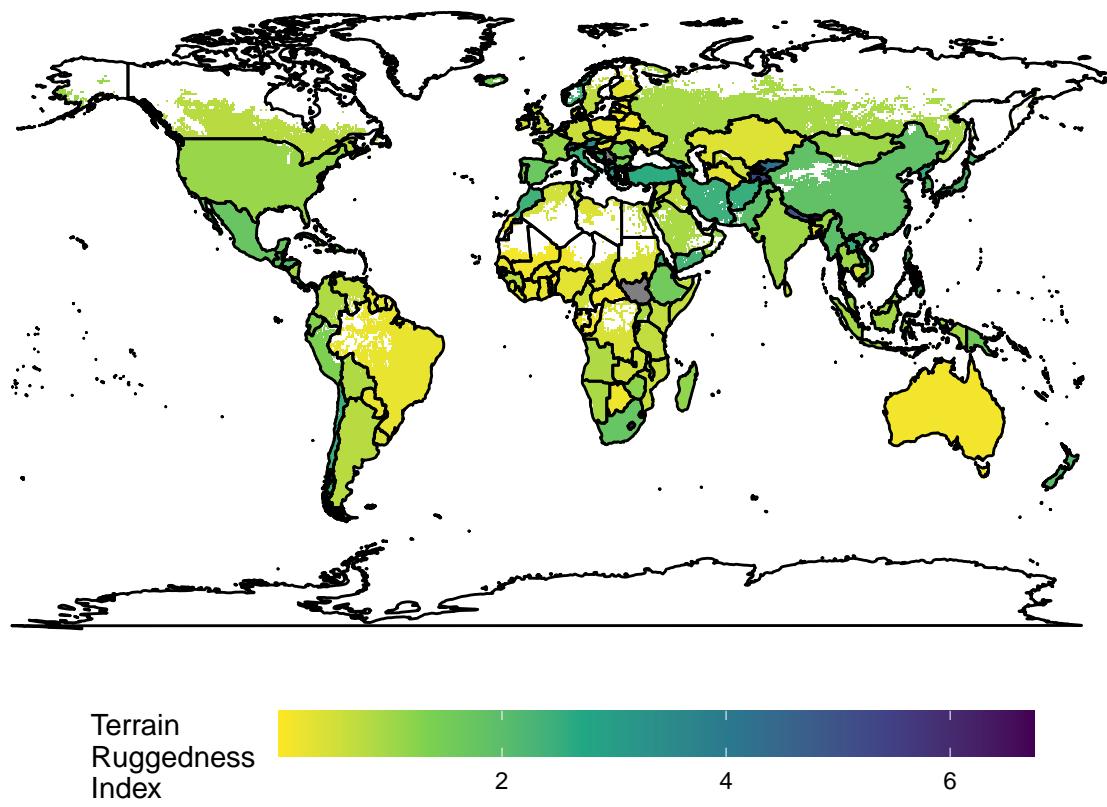


Figure 3.3: TRI values on a country level basis.

Economics: Income

Maddison Project Database The Maddison Project Database was developed by University of Groningen to document and track global historical economic trends to gain a better understanding of the drivers behind periods of global economic growth and stagnation (Bolt & van Zanden, 2020). The most recent update includes historical data for more than 160 countries. Using data collected from a variety of sources, the Maddison Project Database presents estimates of income, represented as Gross Domestic Product per capita, and the subsequent total population used to calculate income on a per person basis. Yearly estimates are provided on a per country basis, although due to data availability, estimates for some years (mainly prior to 1960) are not complete for all countries (Bolt, Inklaar, de Jong, & van Zanden, 2018). In order to standardize methods of estimations across time and space, the most current version of the Maddison database relies on the extrapolation method to determine yearly estimates. The total and per capita GDP were selected from benchmark year of 1990 and then estimates forwards and backwards are made based on growth rates from country national accounts (or other growth estimates). These estimates are then tested against independent benchmarks where they exist (Bolt & van Zanden, 2020).

Description

External Influences: Democracy

Bjørnskov-Rode Regime Data The Bjørnskov-Rode Regime Data developed by Bjørnskov & Rode (2018) presents a data set of political governance regimes and regime changes for 192 countries and 16 self governing territories between 1950 and 2016, including data regarding periods of colonial control, which other well known datasets such as Center for Systemic Peace (2020) used in Neumann et al. (2011) do not include. The Bjørnskov-Rode Regime Dataset classifies governmental regimes into democratic and non democratic categories, each with three classifications, giving rise to six different political regimes.

| Democracies | Non-Democracies |
|--------------------------------------|---|
| Parliamentary Democracies | Civilian Autocracies (Leader w/o military rank) |
| Mixed Democracies (w/weak president) | Military Dictatorships (Leader w/military rank) |
| Presidential Democracies | Royal Dictatorships |

In addition, a classification for “Colony” (i.e. most often “Former Colony”) was created to better understand understand the relationship between external governance and irrigation fraction.

Description A representation of governance regimes for the study area in the last period of the study period can be seen in Figure 3.4 below.

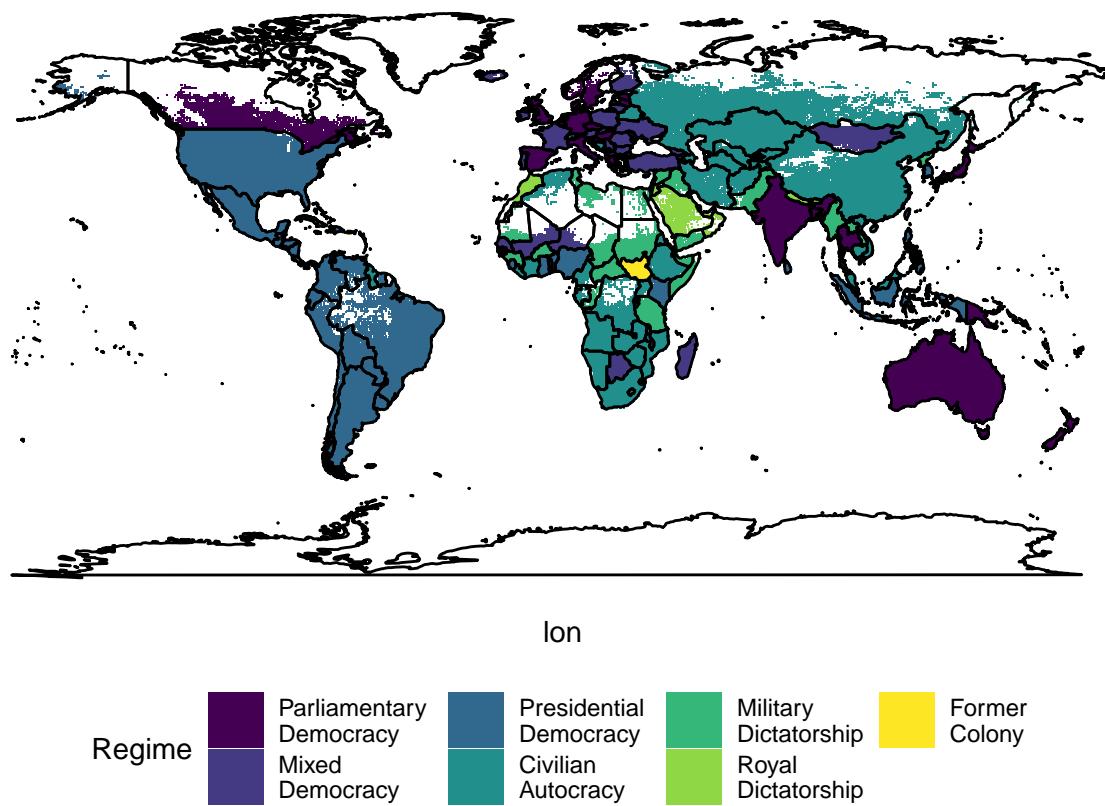


Figure 3.4: The political systems of the study area in the last year of the study period (2005).

3.2.2 LPJmL Data

The Lund-Potsdam-Jena managed Land (LPJmL) is a Dynamic Global Vegetation Model which simulates the global terrestrial carbon cycle and the corresponding response of vegetation patterns, both natural and managed, under a given set of climactic conditions (Schaphoff et al., 2018). Spatially and temporally explicit data regarding climate, landuse patterns, and anthropogenic activities among others is fed into the model and used to simulate, via the LPJmL's established biophysiological interconnections, a multitude of processes and outcomes that relate to global vegetation patterns and carbon cycles(PIK, n.d.). Four predictor variables were used as input for or were extracted from the LPJmL model. They are detailed in the section below.

Compatibility: Distance to the Next Irrigated Cell

LPJmL Derived Output

Description

Social Influences: Population Density

The History Database of the Global Environment (HYDE) The History Database of the Global Environment (HYDE) was first developed by C. Klein Goldewijk & Battjes (1997) to better understand and map historical land use patterns anthropogenic activities. Several updates to this dataset have taken place since the publication of the original dataset and currently HYDE Version 3.2 currently includes population estimates and land use patterns for the entirety of the Holocene, which spans the last 12,000 years (K. Klein Goldewijk, Beusen, Doelman, & Stehfest, 2017).

For the study period (1960-2005) of this masters thesis, estimates of total population were collected from the United Nations World Populations Prospects (Division, 2008). Then this data is harmonized with world wide spatial population patterns from Land Scan (Laboratory, 2014) which shows global population density at a 1km resolution and was used to allocate population totals to each available administrative unit (K. Klein Goldewijk, Beusen, Doelman, & Stehfest, 2017). Population data was then harmonized resulting in a final data product that yields population density, and other important land use data, on a $0.5^\circ \times 0.5^\circ$ latitude longitude global grid.

The HYDE population density estimates were then used as an input to LPJmL

Description A description of population density estimates from the year 2005 are represented in Figure 3.5.

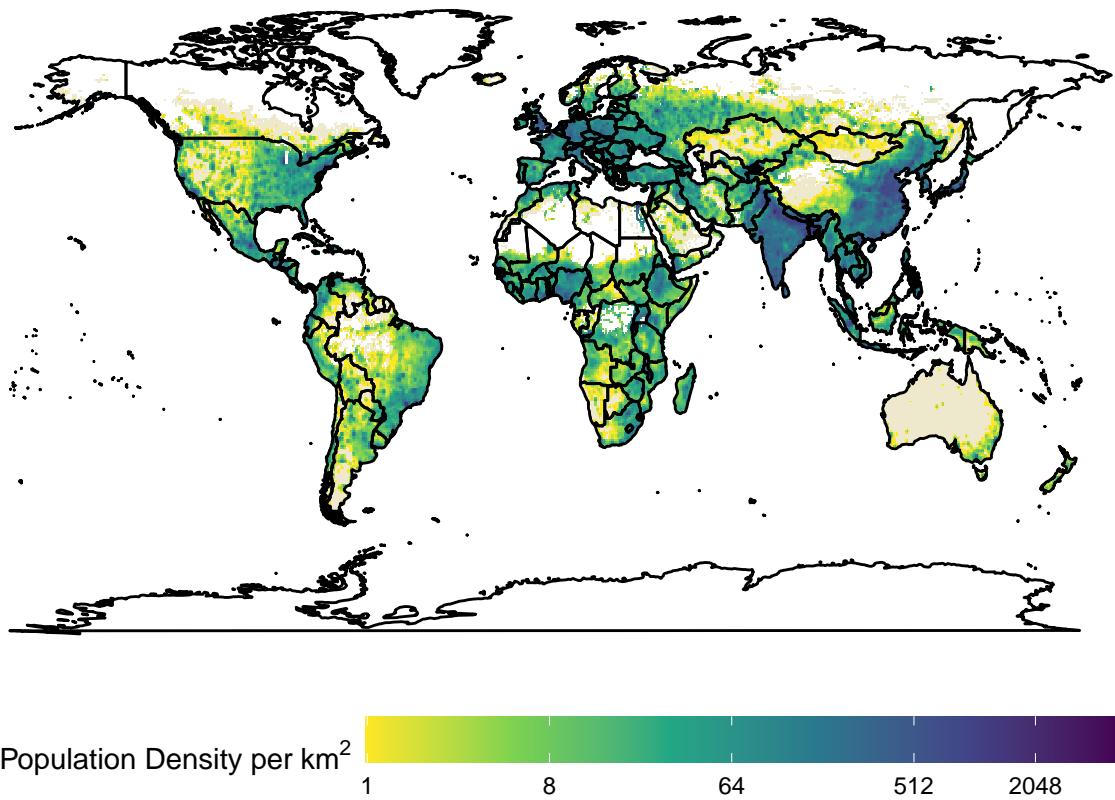


Figure 3.5: Population density

Water: Precipitation

To encapsulate the relationship between the presence/absence of irrigation or the amount of irrigation in a cell and water, precipitation was chosen to represent this link. Several other variables existed that would have also represented the connection between a water source and irrigation expansion, however to reduce multicollinearity, only one variable could be used to encapsulate all the effects of water. Through an iterative modeling process of checking the strength and stability of different predictor effects, precipitation was ultimately chosen. In addition, using precipitation as a predictor conveniently fits best within the conceptual framework better than others with the notion that farmers will begin to irrigate when there precipitation cannot meet crop demand.

Climate Research Unit Time Series Dataset Data for precipitation was collected from the Climate Research Unit of the University of East Anglia. This dataset is open-source and includes monthly precipitation data for a $0.5^\circ \times 0.5^\circ$ latitude longitude global grid cell resolution over a time period of 1901 to 2018. To collect this data, individual station observations were anomolized by using the each station's monthly mean from the years 1961 to 1990 to standardize observations as a percentage of the mean monthly value (value of -100 means 0 precipitation, value of 0 is equal to

monthly mean for a given station). These anomalies were then converted to a $0.5^\circ \times 0.5^\circ$ latitude longitude grid using angular distance weighting ultimately giving rise to a single monthly value for each grid cell (University of East Anglia Climatic Research Unit, Harris, Osborn, & Jones, 2021).

Description Below

Outcome Variable: Irrigation Fraction

Global Historical Irrigation Data Set Data for global historical irrigation patterns was collected from the Global Historical Irrigation Data Set (HID) (Stefan Siebert et al., 2015) which describes hectares of area equipped for irrigation (AEI) per grid cell at a 5 arcmin resolution² over a period of 105 years, from 1900 to 2005. By documenting global and historical irrigation patterns, Stefan Siebert et al. (2015) hoped to create a better understanding of the evolution of said patterns. It is worth noting that the dataset provided in Stefan Siebert et al. (2015) documents area equipped for irrigation, meaning area that is equipped with infrastructure to irrigate crops but not *necessarily* irrigated. In addition, rainwater harvesting³ is also not included in the summation of area equipped for irrigation. [CITATION]

To amass this data Stefan Siebert et al. (2015) used a variety of sources to collect national and subnational statistics including FAOSTAT (FAO, 2021b), EuroStat (Comission, 2021), and Aquastat (FAO, 2021a) along with other less collected sources like census data and statistical yearbooks. Data was recorded for 10 year timesteps until 1980 and five year timesteps until the termination of the study period in 2005. Data for the period prior to 1950 and for the year 2005 has higher levels of uncertainty in the measurements when compared to the data between 1950 and 2005, as irrigation data from international organizations (e.g. FAO) were unavailable prior and post. After collection the data was harmonized and downscaled to a 5 arcmin resolution. Special care was taken to ensure that high resolution data (at a 5 arcmin resolution) could be accurately summed to the subnational level, ensuring accuracy at different resolutions. In addition, the authors note that validation of this dataset was not possible due to the fact that all available data was used as input to create the HID (Stefan Siebert et al., 2015).

The HID, in combination with other datasets such the History Database of the Global Environment (C. Klein Goldewijk & Battjes (1997), see Section ??) and Climate Research Unit Time Series Data (University of East Anglia Climatic Research Unit, Harris, Osborn, & Jones (2021), see Section (ref?)(CRU)) were used was then fed into the LPJmL model. For the purposes of this masters thesis, a global irrigation pattern represented by percentage of irrigated area per cell at a $0.5^\circ \times 0.5^\circ$ latitude longitude global grid cell resolution with yearly observations for a period of 1960 to 2005 was extracted from the LPJmL model .

²At the equator, this is roughly a 9.2km by 9.2km grid cell resolution.

³The act of harvesting rainwater during rain and then later applying it to crops.

Description Levels of irrigation vary widely both geographically and temporally. In Figure 3.6 the global irrigation pattern for the last year of they study period (2005) is depicted.

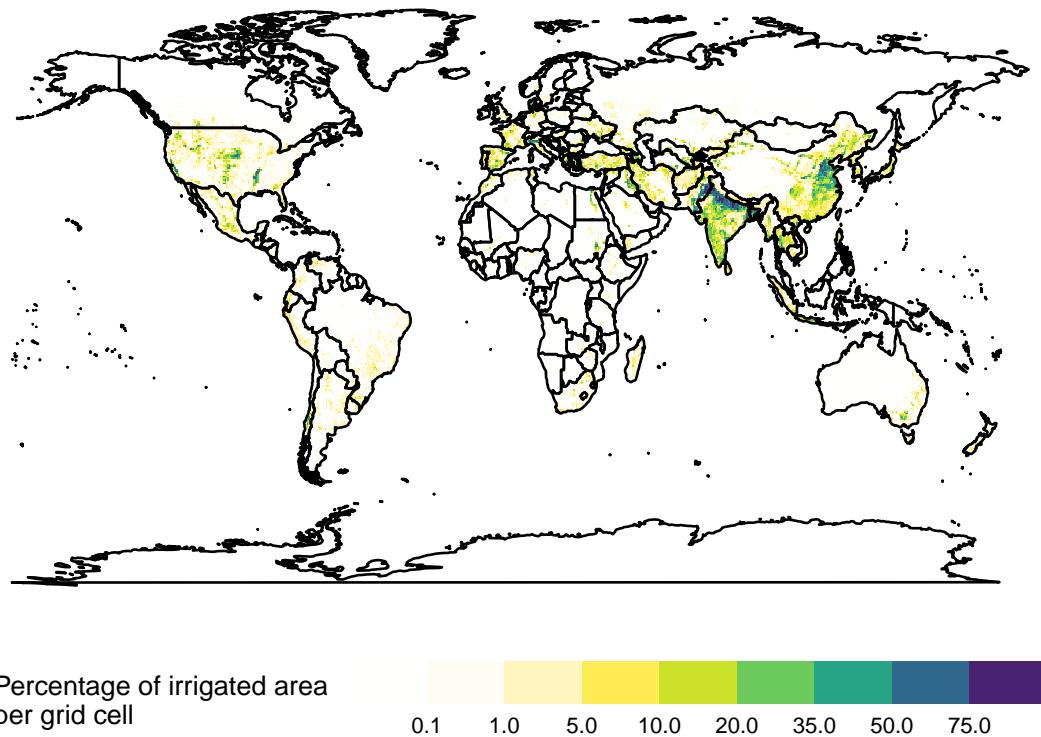


Figure 3.6: Percentage of a cell’s area covered by land equipped for irrigation. Represents the global irrigation pattern at a $0.5^\circ \times 0.5^\circ$ latitude longitude grid cell resolution for the last year of the study period (2005).

There are several “hotspots” of irrigation that occur, predominantly in the interior of the United States, north of India, Bangladesh, and Eastern China in which more than 30% of a cell is irrigated area. Otherwise, generally irrigation, where it exists, a small minority of cell area. No irrigation occurs in some places including towards arctic circle and in the Sahara, among other places.

When looking at the histogram of the data distribution of the target variable, irrigation fraction, it is clear to see that the data includes many zeros.

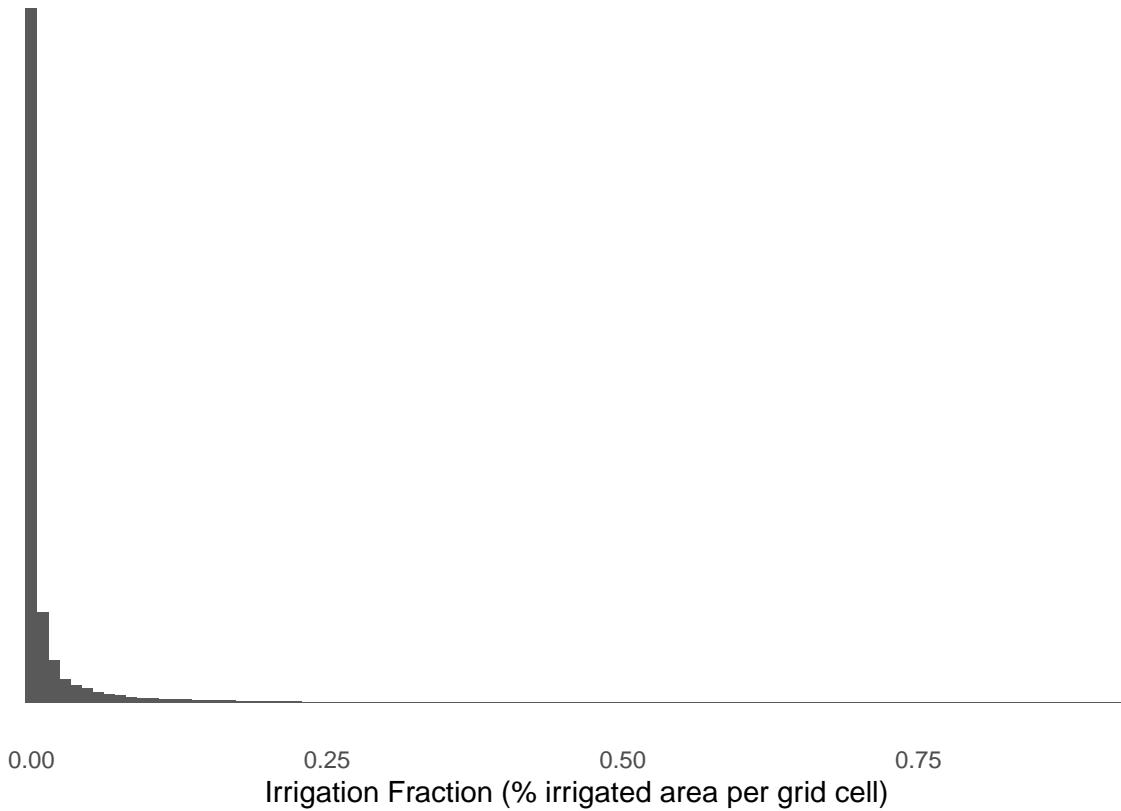


Figure 3.7: Histogram of the outcome variable, Irrigation Fraction (IF).

3.3 Hypothesis

Based on information discussed above several hypothesis can be formed to investigate the expected effects of different predictors. As discussed previously there are two outcome variables in this masters thesis: the presence/absence of irrigation and the amount of irrigation. Hypothesis are similar for both predictor variables. The rational is explained below in addition to the hypothesis.

For the amount of irrigation present in a cell the stated hypothesis for this thesis are:

1. A **decrease in precipitation** will result in an **increase in irrigation**.
2. A **decrease in population density** will result in an **increase in irrigation**.
3. An **increase in income (GDP per capita)** will result in an **increase in irrigation**.
4. A **decrease in topographical heterogeneity (ruggedness)** will result in an **increase in irrigation**.
5. An **increase in democratic values** will result in an **increase in irrigation**.

For the presence or absence of irrigation, the following hypothesis are proposed:

1. A decrease in distance to the next irrigated cell will result in an **increase of irrigated cells**.
2. An **increase in population density** will result in an **increase in irrigated cells**.
3. A decrease in precipitation will result in an **increase in irrigated cells**.
4. An **increase in income (GDP per capita)** will result in an **increase of irrigated cells**.

3.4 Data Processing and Modeling

3.4.1 Thinning

“The first law of geography: Everything is related to everything else, but near things are more related than distant things.” - Walter Tobler (Tobler, 1970)

For Spatially Autocorrelated Structures

For Temporally Autocorrelated Structures

3.4.2 Standardization

3.4.3 Time Series Considerations

3.4.4 Hierarchical Modeling

Bayesian hierarchical modeling has some unique features in comparison to other forms of statistical influence. in the context of this thesis, the two main advantages of Bayesian Inference are an improved propagation of uncertainty throughout the modeling process and an ability to include prior information into models, allowing the inclusion prior knowledge (Gelman, Hill, & Vehtari, 2020).

DAGs

A Directed Acyclic Graph (DAG)

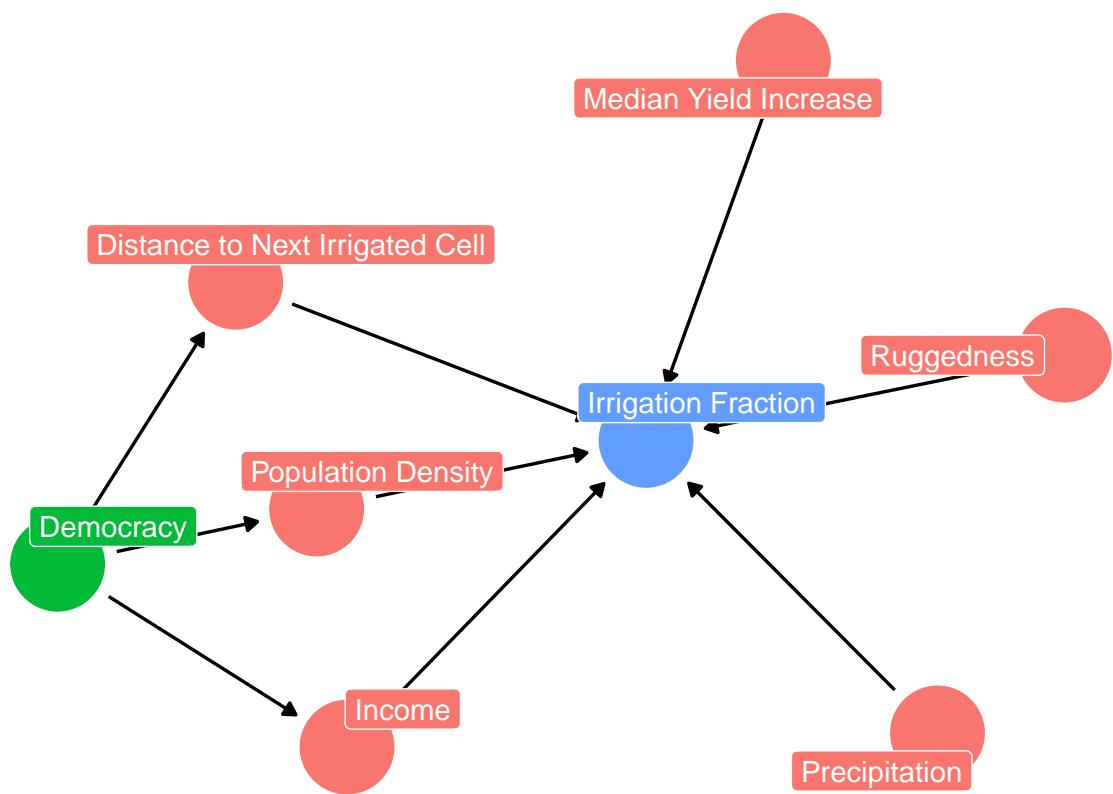


Figure 3.8: Directed Acyclic Graph for Beta

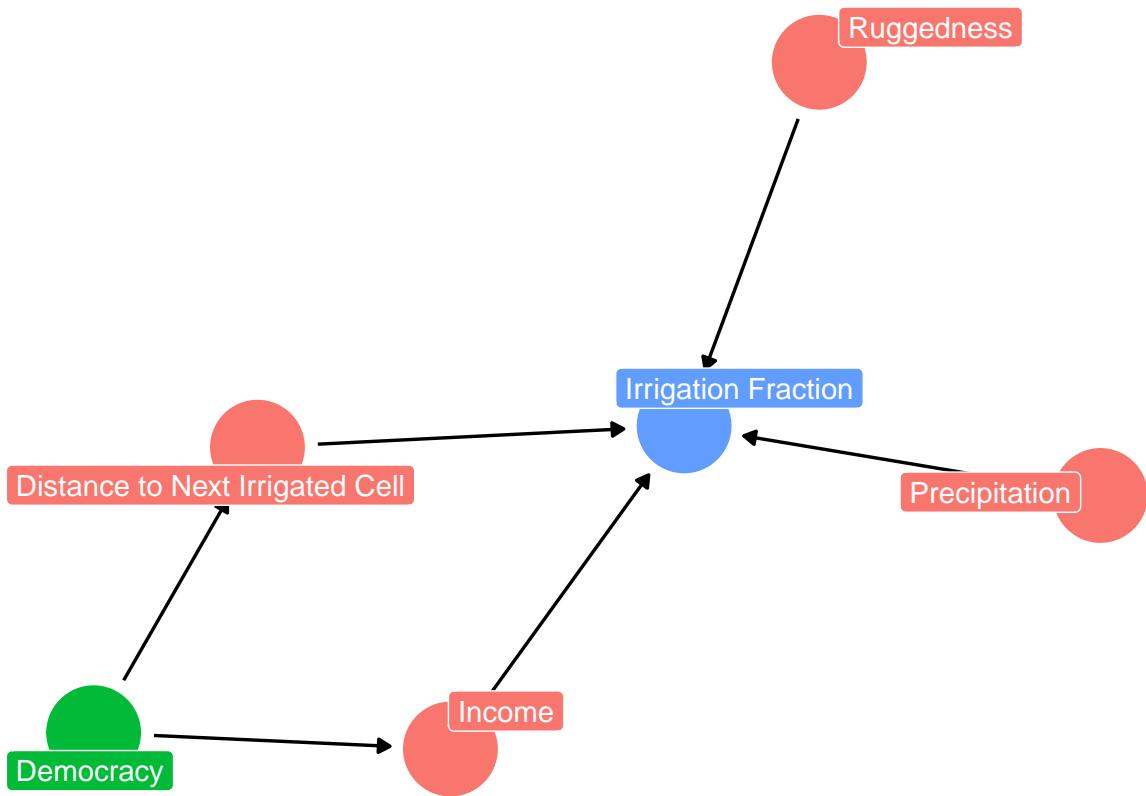


Figure 3.9: Directed Acyclic Graph for zero inflated parameter z_i

To test these hypothesis, a quick view of correlations was carried out.

```
fr if(knitr:::is_latex_output()) '\\appendix'
```

Chapter 4

Appendix

4.1 Proof of Bayes' Theorem for Two Hypothesis from Hacking (2001)

Definition of **Conditional Probability** taken from Hacking (2001) beginning on p.49:

$$\Pr(A|B) = \frac{\Pr(A,B)}{\Pr(B)} \quad (4.1)$$

Implying that for $\Pr(B) > 0$,

$$\Pr(A|B) * \Pr(B) = \Pr(A,B) \quad (4.2)$$

Knowing that,

$$\Pr(B,A) = \Pr(A,B) \quad (4.3)$$

Therefore,

$$\Pr(B|A) * \Pr(A) = \Pr(A|B) * \Pr(B) \quad (4.4)$$

Dividing by $\Pr(A)$, gives Bayes' Theorem (equivalent to Equation (4.8))

$$\Pr(B|A) = \frac{\Pr(A|B) * \Pr(B)}{\Pr(A)} \quad (4.5)$$

Assuming that $0 < \Pr(B) < 1$, the equation for **Total Probability** taken from Hacking (2001, p. 59) can be written:

$$\Pr(A) = \Pr(A|B) * \Pr(B) + \Pr(A|\text{Not } B) * \Pr(\text{Not } B) \quad (4.6)$$

Substituting Equation (4.6) into Equation (4.5), giving rise to the Equation (4.7), the base equation used for the basis of Equation (4.12),

$$\Pr(B|A) = \frac{\Pr(A|B) * \Pr(B)}{\Pr(A|B) * \Pr(B) + \Pr(A|\text{Not } B) * \Pr(\text{Not } B)} \quad (4.7)$$

4.1.1 Bayes Theorem

Bayes' theorem is the basis for Bayesian inference. Although Bayes' theorem can be manipulated into many forms, the most basic representation is Equation (4.8).

$$p(\theta|y) = \frac{p(y|\theta) * p(\theta)}{p(y)} \quad (4.8)$$

Where θ is an unknown parameter to be estimated and y is the observed data. To understand what is happening here, it is necessary to define some terms using accessible language from McElreath (2020, p. 37).

- The symbol $|$ can be read as “given” or “conditional on,” meaning that the probability of one event is dependent on the occurrence of another. This is also known as “Conditional Probability.”
- The outcome $p(\theta|y)$, is also known as the “posterior” and can be read as “the probability of the unobserved parameter given the observed data.”
- $p(y|\theta)$ is the probability of the data, or the likelihood.
- $p(\theta)$, represents previous knowledge about the unobserved parameter θ , and is called the “prior.”
- $p(y)$ is the “average probability of the data,” which is “the probability of the data, that has been averaged over $p(\theta)$,” as McElreath (2020, p. 37) mentions. The main function of $p(y)$ is to normalize the posterior (remember, $p(\theta|y)$).

All together Bayes' theorem looks like this (McElreath, 2020):

$$\text{Posterior} = \frac{\text{Probability of the Data} * \text{Prior}}{\text{Average Probability of the Data}} \quad (4.9)$$

It is important to remember that although Bayes' theorem can be used with point estimates¹, the advantage of Bayesian inference is that all terms in Equation (4.8) can be represented as probability distributions, communicating much more uncertainty about a parameter than point estimate statistics. From this point forward, assume that terms such as posterior, prior, and likelihood imply distributions not point estimates. For fun, a relevant example of a point estimate use of Bayes Theorem is listed in Section 4.2.

Even more so when using probability distributions, often times it is difficult to deal with the normalizing term, $p(y)$. However, as $p(y)$ does not depend on the unobserved parameter θ , it can be considered a constant and factored out giving rise to the more common used and less computationally difficult *unnormalized posterior density* (Gelman et al., 2013; McElreath, 2020).

$$p(\theta|y) \propto p(y|\theta) * p(\theta) \quad (4.10)$$

¹A point estimate is a single value, generally the average, given as the estimate value for a parameter.

Effectively, Equation (4.10) means that the posterior is proportional to the likelihood, $p(y|\theta)$, multiplied by the prior, $p(\theta)$. It's not necessary in most contexts to normalize using $p(y)$, proportionality works just fine.

The Prior

Specification of the prior distribution is not a trivial task, as the prior attempts to encapsulate the previous knowledge of a specific parameter and has impacts on the posterior distribution. The determination of the prior should be done before seeing the data and serves two main functions, communicating previous scientific or common knowledge about a parameters distribution and/or helping the model learn more efficiently (McElreath, 2020). Often times flat priors are used initially and improved upon iteratively. Effects of the prior on the posterior can be seen in Figure 4.1 which is adapted from McElreath (2020, p. 38) and Kurz (2019, sec. 2.4).

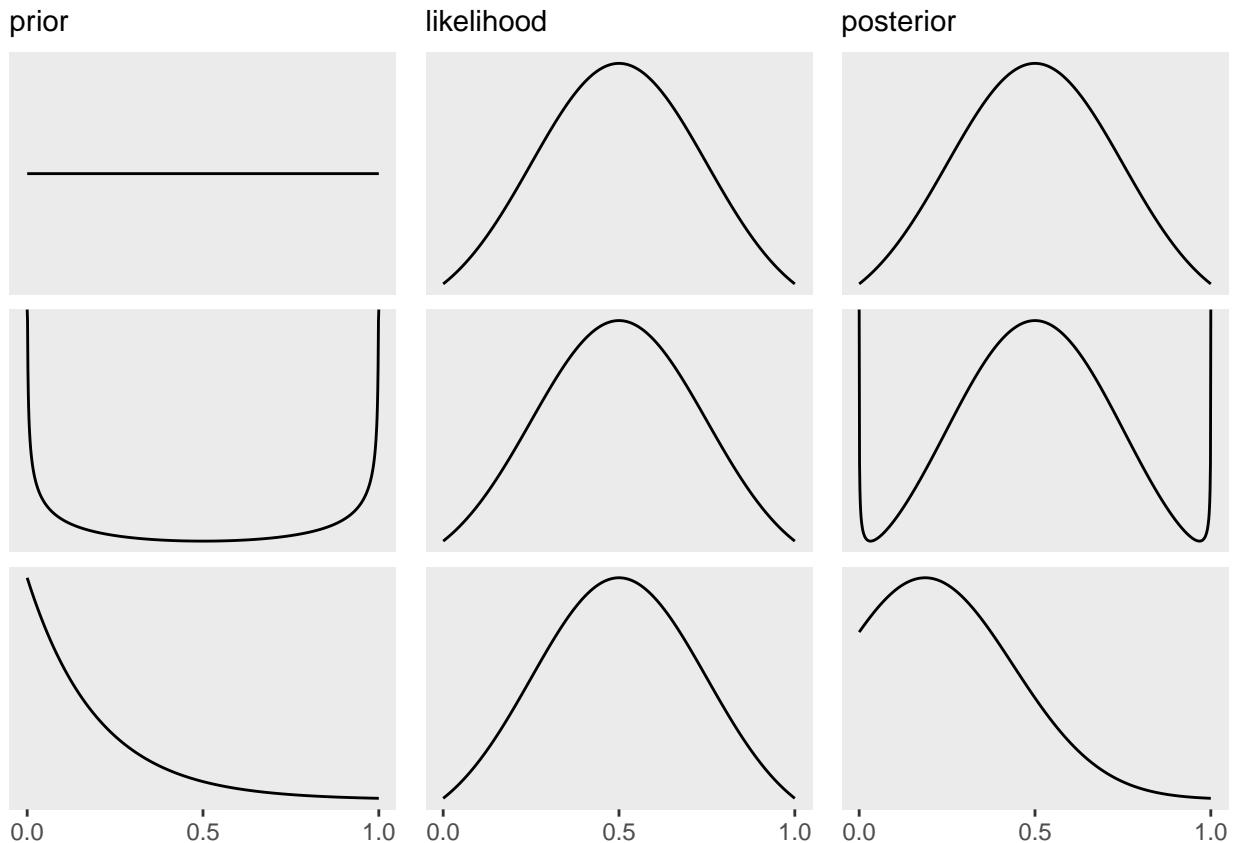


Figure 4.1: Effects of the prior distribution on the unnormalized posterior with a $\text{Normal}(0.5, 0.25)$ likelihood. First row: Uniform prior, Second row: $\text{Beta}(a = 0.75, b = 0.75)$ prior, Third Row: Exponential prior. Prior and likelihood are multiplied together to create the unnormalized posterior seen in the third column.

There are several different types of priors used when discussing Bayesian inference.

Some of the most common types of priors discussed are (Gelman et al., 2013):

- **Proper priors**, which integrate to 1 and **improper priors**, that don't. The fact that a prior (or any probability distribution for that matter) integrates to 1 is important not only computationally, as it makes it easier to manipulate and solve analytically, but an improper prior violates the assumption that probability distributions integrate to 1.
- **Informative priors** communicate information about a given parameter's distribution, **non-informative priors** do not. A classic example of a non informative prior is a flat or uniform prior (seen in the row 1 of Section 4.1) as it assumes that all probability values are equally likely which is almost never the case.
- To discuss the amount of information communicated via the prior distribution to the posterior, terms such as **Strongly Informative Priors** or **Weakly Informative Priors** can be used. Weakly informative priors are informative (i.e. they do contain information about the distribution of a given parameter) but serve to regularize the posterior only by setting bounds within which the posterior can vary its shape, meaning that they are not aiming to completely describe the distribution of a parameter. In addition, these priors are easily overwhelmed if the amount of data is large. Strongly informative priors, conversely, heavily influence the posterior distribution and are more likely to overwhelm the data.

The goal when specifying the prior is to have at least a weakly informative prior for each given parameter. This ensures that general scientific knowledge about the distribution of a parameter is communicated but if large amounts of data exist, the prior gives way to the overwhelming evidence of the likelihood (Gelman, 2020). Manipulation of the prior can be useful from can be used to solve or diagnose issues, given certain contexts and questions. However, it is important to note that “bad” (less suitable) priors can come with some unintended consequences such as allowing for more unstable predictions due sampling difficulties from bad geometry of the posterior (when using tools such as STAN (Stan Development Team, 2020)) or entertaining predictions that are not possible given the problem (ex. predicting negative values when logically only positive ones exist) (McElreath, 2020, sec. 5.1). Priors that are skeptical of extreme values are often times called *regularizing* priors (McElreath, 2020, sec. 7.5)

Gelman, Simpson, & Betancourt (2017) and Gelman (2020) discuss appropriate prior choice and emphasize that the identification of weak or strong priors is not possible without knowledge of the likelihood, which violates the idea that a prior should be constructed *a priori*, or before having any notion of the data or even the data generating process. As mentioned in Gelman, Simpson, & Betancourt (2017, sec. 1.2), this may be the most ideal way to construct a prior but in reality practical concerns such as model convergence, unknown parameters, knowledge of the modeler, etc. make this task impossible or very difficult. Hence, a division is created that separates Bayesian issues into two camps, those can/are willing to build a prior in a *subjective* way by incorporating expert/scientific knowledge into the prior *before* seeing the data or setting up the model, and those that “break” this rule by using some knowledge about the likelihood and data to construct their prior.

Ultimately, prior specification for Bayesian inference is necessary but is often times more of an art than a science. Iterative processes help with choosing the prior and seeing its effect on the posterior for a given question, dataset, and model. The overall aim is to balance practical and theoretical concerns when choosing most the suitable prior. Currently, a large area of research focuses solely on prior specification and usage within a Bayesian context aimed at demystifying prior specification and increasing standardization of prior specification processes and documentation.

4.1.2 Model Comparison

Multiple ways to compare models exist, some rely more heavily on a conceptual and causal understanding of the problem, others rely on predictive accuracy. Model comparison can occur *within-sample* or *out-of-sample*, meaning that comparisons calculated using data used to train the model (*within-sample*) or using data that has been left out when fitting the model (*out-of-sample*).

More informal methods of model comparison exist such as prior and posterior predictive checks, comparison of bayes' factor or bayesian R^2 among others. (Gelman, Hwang, & Vehtari, 2014). However, as these checks are done using within-sample data, and some can be prone to overfitting under certain circumstances (McElreath, 2020, Chapter 7)

4.2 Digression: Covid Rapid Antigen Tests

Imagine (you don't have to) that SARS-CoV-2 is circulating in Berlin. An individual takes a rapid antigen test and receives a positive test result. However, the media recently has been discussing that rapid antigen tests are less accurate than other types of tests to detect SARS-CoV-2, like the PCR. The individual thinks that perhaps the test was wrong, meaning that this person received a "false positive" and doesn't actually have Covid even though the test came back positive. Data can be gathered, a positive test result (the y), and a hypothesis can be constructed, the individual doesn't have Covid (the θ). Substituting this in to Bayes' theorem, the probability that this individual doesn't have Covid even though the test came back positive can be calculated (remember, $p(\theta|y)$) using Equation (4.11).

$$p(\text{No Covid}|\text{Pos}) = \frac{p(\text{Pos}|\text{No Covid}) * p(\text{No Covid})}{p(\text{Pos})} \quad (4.11)$$

Some extra information is necessary to fill in the missing pieces,

- According to Tagesspiegel (2021), 200 people are infected with SARS-CoV-2 for every 100,000 people in Berlin, also known as probability of having Covid ($P(\text{Covid}) = 0.002$). Since our hypothesis is that the individual *doesn't* have Covid, we can calculate $P(\text{No Covid}) = 0.998$. This represents our previous knowledge about how many people don't have Covid and is the prior ($p(\theta)$) in Bayes' theorem.

- My local Covid Test Center claims that their rapid antigen tests have a sensitivity of 96% and a specificity of 99%. Sensitivity is the ability to correctly identify those who have the disease and specificity is the ability to correctly identify those that do not have the disease.
 - So sensitivity represents the probability of testing positive when (or given) you have Covid, $P(\text{Positive}|\text{Covid}) = 0.96$. Conversely, the $P(\text{Negative}|\text{Covid}) = 0.04$
 - Specificity represents the probability of testing negative given you don't have Covid $P(\text{Negative}|\text{No Covid}) = 0.99$. Then, $P(\text{Positive}|\text{No Covid}) = 0.01$. This is the probability of the data, or the likelihood ($p(y|\theta)$).

From these statements, Table 4.1 can be created .

Table 4.1: Probabilities for incidence of Covid and rapid antigen test sensitivity.

| | Covid ($p = 0.002$) | No Covid ($p = 0.998$) |
|----------|-----------------------|--------------------------|
| Positive | 0.96 | 0.01 |
| Negative | 0.04 | 0.99 |

Last thing that needs to be discussed is the denominator, $p(\text{Positive})$ or $p(y)$. There is a manipulation of Bayes's rule using the law of Conditional Probability detailed in Hacking (2001) that can be used in cases where there are only two outcomes (two θ s). This can be applied here as is this the case, i.e. the person either has or doesn't have Covid. Equation (ref?)(eq:bayesspecial) details Hacking (2001) for this example. Proof for (ref?)(eq:bayesspecial) can be seen in 4.1.

$$P(\text{No Covid|Pos}) = \frac{P(\text{Pos|No Covid}) * P(\text{No Covid})}{P(\text{Pos|No Covid}) * P(\text{No Covid}) + P(\text{Pos|Covid}) * P(\text{Covid})} \quad (4.12)$$

From then using Table 4.1 and Equation (4.12), the probability of not having Covid given a positive test result can be calculated.

$$P(\text{No Covid|Pos}) = \frac{0.01 * 0.998}{0.01 * 0.998 + 0.96 * 0.002} \quad (4.13)$$

$$P(\text{No Covid|Pos}) = 0.84 \quad (4.14)$$

So, what does this mean? This $P(\text{No Covid|Pos}) = 0.84$ is the probability of not having Covid given a positive test result. This is quite high, so chances are even though the test came back positive, this individual does not have Covid. Perhaps we should take another rapid test or a PCR.

4.3 Predictor Selection

References

- Adachi, S. (2007). Agricultural Technologies of Terraced Rice Cultivation in the Ailao Mountains, Yunnan, China, *6*(2), 173–196. <http://doi.org/10.14956/asafas.6.173>
- Angelakis, A. N., Zaccaria, D., Krasilnikoff, J., Salgot, M., Bazza, M., Roccaro, P., ... Fereres, E. (2020). Irrigation of World Agricultural Lands: Evolution through the Millennia. *Water*, *12*(5), 1285. <http://doi.org/10.3390/w12051285>
- Bjorneberg, D. L. (2013). IRRIGATION | Methods. In *Reference Module in Earth Systems and Environmental Sciences*. Elsevier. <http://doi.org/10.1016/B978-0-12-409548-9.05195-2>
- Bjørnskov, C., & Rode, M. (2018). Regime Types and Regime Change: A New Dataset. *SSRN Electronic Journal*. <http://doi.org/10.2139/ssrn.3234263>
- Bolt, J., Inklaar, R., de Jong, H., & van Zanden, J. L. (2018). Rebasing 'Maddison': New Income Comparisons and the Shape of Long-run Economic Development, 69.
- Bolt, J., & van Zanden, J. L. (2020). Maddison style estimates of the evolution of the world economy. A new 2020 update, 44.
- Center for Systemic Peace. (2020, April). Polity Project: Polity 5 Dataset.
- Comission, E. (2021). Eurostat Database. <https://ec.europa.eu/eurostat/data/database>.
- Division, U. N. P. (2008). World Population Prospects, The 2007 Revision.
- FAO. (2011). *The state of the world's land and water resources for food and agriculture (SOLAW) Managing systems at risk*. Rome and London: FAO and Earthscan.
- FAO. (2021a). AQUASTAT. <http://www.fao.org/aquastat/en/databases/>.
- FAO. (2021b). FAOSTAT. <http://www.fao.org/faostat/en/#data>.
- FAO, Brouwer, C., Goffeau, A., & Heibloem, M. (1985). *Irrigation Water Management: Introduction to Irrigation*. Rome.
- FAO, Brouwer, C., & Heibloem, M. (1986). *Irrigation Water Management: Irrigation Water Needs*. Rome.

- FAO, Brouwer, C., Prins, K., Kay, M., & Heibloem, M. (1986). *Irrigation Water Management: Irrigation Methods* (No. Training Manual no. 5). Rome.
- FAO, & SAFR. (2001). *Irrigation manual: Planning, development, monitoring and evaluation of irrigated agriculture with farmer participation*. Harare: Food and Agriculture Organization of the United Nations (FAO), Sub-Regional Office for East and Southern Africa (SAFR).
- FAO, & Walker, W. R. (1989). *Guidelines for designing and evaluating surface irrigation systems*. Rome.
- Forseth, I. (2010). Terrestrial Biomes. *Nature Education Knowledge*. <https://www.nature.com/scitable/knowledge/library/terrestrial-biomes-13236757/>.
- Gelman, A. (2020, April). Prior Choice Recommendations. *Stan Dev GitHub*. <https://github.com/stan-dev/stan/wiki/Prior-Choice-Recommendations>.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian Data Analysis*. Chapman and Hall/CRC. <http://doi.org/10.1201/b16018>
- Gelman, A., Hill, J., & Vehtari, A. (2020). *Regression and Other Stories*. Cambridge University Press.
- Gelman, A., Hwang, J., & Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models. *Statistics and Computing*, 24(6), 997–1016. <http://doi.org/10.1007/s11222-013-9416-2>
- Gelman, A., Simpson, D., & Betancourt, M. (2017). The Prior Can Often Only Be Understood in the Context of the Likelihood. *Entropy*, 19(10), 555. <http://doi.org/10.3390/e19100555>
- Hacking, I. (2001). *An Introduction to Probability and Inductive Logic*. <http://doi.org/10.1017/CBO9780511801297>
- Holzapfel, E. A., & Mariño, M. A. (2008). Irrigation in Agriculture. In S. E. Jørgensen & B. D. Fath (Eds.), *Encyclopedia of Ecology* (pp. 2033–2039). Oxford: Academic Press. <http://doi.org/10.1016/B978-008045405-4.00628-5>
- Klein Goldewijk, C., & Battjes, J. (1997). A Hundred Year (1890 - 1990) Database for Integrated Environmental Assessments (HYDE, version 1.1). *PBL Netherlands Environmental Assessment Agency*.
- Klein Goldewijk, K., Beusen, A., Doelman, J., & Stehfest, E. (2017). Anthropogenic land use estimates for the Holocene HYDE 3.2. *Earth System Science Data*, 9(2), 927–953. <http://doi.org/10.5194/essd-9-927-2017>
- Kurz, A. S. (2019). *Statistical Rethinking with brms, Ggplot2, and the tidyverse*.

- Laboratory, O. R. N. (2014). LandScan: LandScan Global Population Database, The 2012 Revision. <https://landscan.ornl.gov/>.
- McElreath, R. (2020). *Statistical Rethinking: A Bayesian Course with Examples in R and STAN* (Second). Chapman and Hall/CRC.
- Neumann, K., Stehfest, E., Verburg, P. H., Siebert, S., Müller, C., & Veldkamp, T. (2011). Exploring global irrigation patterns: A multilevel modelling approach. *Agricultural Systems*, 104(9), 703–713.
- Nunn, N., & Puga, D. (2012). Ruggedness: The Blessing of Bad Geography in Africa. *Review of Economics and Statistics*, 17.
- PIK. (n.d.). LPJmL - Lund-Potsdam-Jena managed Land Potsdam Institute for Climate Impact Research. <https://www.pik-potsdam.de/en/institute/departments/activities/bios-water-modelling/lpjml>.
- Portmann, F. T., Siebert, S., & Döll, P. (2010). MIRCA2000Global monthly irrigated and rainfed crop areas around the year 2000: A new high-resolution data set for agricultural and hydrological modeling. *Global Biogeochemical Cycles*, 24(1). <http://doi.org/10.1029/2008GB003435>
- Program, W. W. A., & UNESCO. (2009). *The United Nations World Water Development Report 3: Water in a Changing World*. Paris.
- Puy, A., Lo Piano, S., & Saltelli, A. (2020). Current Models Underestimate Future Irrigated Areas. *Geophysical Research Letters*, 47(8), e2020GL087360. <http://doi.org/10.1029/2020GL087360>
- Riley, S., Degloria, S., & Elliot, S. D. (1999). A Terrain Ruggedness Index that Quantifies Topographic Heterogeneity. *International Journal of Science*, 5, 23–27.
- Rockström, J., Karlberg, L., Wani, S. P., Barron, J., Hatibu, N., Oweis, T., ... Qiang, Z. (2010). Managing water in rainfed agricultureThe need for a paradigm shift. *Agricultural Water Management*, 97(4), 543–550. <http://doi.org/10.1016/j.agwat.2009.09.009>
- Schapoff, S., Von Bloh, W., Rammig, A., Thonicke, K., Biemans, H., Forkel, M., ... Waha, K. (2018). LPJmL4 - A dynamic global vegetation model with managed land - Part 1: Model description. *Geoscientific Model Development*, 11(4), 1343–1375. <http://doi.org/10.5194/gmd-11-1343-2018>
- Siebert, S., Burke, J., Faures, J. M., Frenken, K., Hoogeveen, J., Döll, P., & Portmann, F. T. (2010). Groundwater use for irrigation a global inventory. *Hydrology and Earth System Sciences*, 14(10), 1863–1880. <http://doi.org/10.5194/hess-14-1863-2010>
- Siebert, Stefan, Kummu, M., Porkka, M., Döll, P., Ramankutty, N., & Scanlon, B. (2015). A global data set of the extent of irrigated land from 1900 to 2005. *Hydrology and Earth System Sciences*, 19, 1521–1545. <http://doi.org/10.5194/hess-19-1521-2015>

19-1521-2015

- Stan Development Team. (2020). RStan: The R interface to Stan.
- Tagesspiegel. (2021, May). Coronavirus-Karte: Deutschlandweite Fallzahlen in Echtzeit. *Tagesspiegel*. <https://interaktiv.tagesspiegel.de/lab/karte-sars-cov-2-in-deutschland-landkreise/>.
- Thenkabail, P. S., Biradar, C. M., Noojipady, P., Dheeravath, V., Li, Y., Velpuri, M., ... Dutta, R. (2009). Global irrigated area map (GIAM), derived from remote sensing, for the end of the last millennium. *International Journal of Remote Sensing*, 30(14), 3679–3733. <http://doi.org/10.1080/01431160802698919>
- Tobler, W. R. (1970). A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, 46, 234. <http://doi.org/10.2307/143141>
- University of East Anglia Climatic Research Unit, Harris, I., Osborn, T. J., & Jones, P. (2021). Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. *Scientific Data*, 7(1), 109. <http://doi.org/10.1038/s41597-020-0453-3>