A Data-driven Approach for Supporting Policy Intervention in Sustainable Development

Pooja Bassin, Ph.D. Student, International Institute of Information Technology, Bangalore (corresponding author)
pooja.bassin@iiitb.ac.in
080 4140 7777
26/C, Hosur Road, Electronic City Phase 1,
Bangalore, Karnataka, 560100
Srinath Srinivasa, Professor, International Institute of Information Technology, Bangalore
Sridhar Mandyam, Professor, International Institute of Information Technology, Bangalore

Abstract

In recent years, there have been a number of research efforts worldwide, exploring the use of data-driven approaches to support sustainable development. This paper presents our work based on integrating disparate datasets to build explainable models of SDG indicators. Bayesian networks are employed to develop a potential causal structure of the modeled domain and attempt to quantify uncertainties associated within. For policy decisions, it is still a challenge to grasp the exact causal influences among the various factors. Therefore, the graphical representation of nodes as real-world objects and arcs as dependence relationships helps policy-makers to draw conclusions before and after policy interventions. We examine how intervening at one or at multiple variables manipulates the network which may shift the status of target indicators. Collection of Bayesian networks corresponding to each of the indicators of the sustainable development goals and inspecting intervention impacts that may flow down and across the models is one of the major areas of focus in our study.

Keywords: policy support, Bayesian networks, policy enunciation

1. Introduction

Sustainability has been a major impetus worldwide, with several countries exploring the use of Big Data, Web, Mobile, and IoT-based interventions for promoting sustainable development. In countries like India, much of the Internet adoption on the ground is driven by rural populations that have primarily used the Internet not so much for exchanging ideas, but as a core component of meeting basic needs, and livelihood generation[1]. This is also aided by a number of state-driven digital initiatives addressing basic needs like creating bank accounts for every citizen, providing Direct Benefit Transfer[2] to various kinds of marginalized and economically weaker sections, creating a Unified Payment Interface (UPI) for enabling payments through mobile phones and empowering small businesses, remote services like medical consultations based on mobile applications, and so on. Such initiatives are largely regarded as *sustainability* and *capability* initiatives. Abstract notions of sustainability and capability have become central conceptual instruments in areas like developmental economics, political science, and systems analysis and design since the late 1970s and early 1980s (Caradonna 2014; Dresner 2012; Lubin and Esty 2010; Sen 2005; Nussbaum 2003; Robeyns 2005). Sustainability is increasingly seen as a guiding principle for social justice, replacing "market forces" as the ultimate arbiter of fairness in free societies (McMichael, Butler, and Folke 2003; Loorbach 2007). United Nations 2030 agenda for Sustainable Development defines sustainability as having three pillars:

[1] Rural India propels digital revolution, registers 45\% internet growth; urban India matures at 11\% growth https://bestmediainfo.com/2020/05/rural-india-propels-digital-revolution-registers-45-internet-growth-urban-india-matures-at-11-growth/

[2] Direct Benefit Transfer https://dbtbharat.gov.in/

Economy, *Society*, and *Environment*. Sustainable development is defined as human development in a way that can continue to rely on the yields from the three pillars. Towards this end, sustainable development is divided into a set of 17 Sustainable Development Goals (SDGs), adopted by all UN member states[3]. Each SDG in turn is characterized by several sub-goals, leading to a total of 169 *targets* to be achieved by all participating member states. For instance, the first target under SDG 1 (No poverty), is as follows: "By 2030, eradicate extreme poverty for all people everywhere, currently measured as people living on less than US$1.25 a day."

Several member countries have initiated their own internal programs to promote SDGs as part of their agenda for governance. These include: ranking their states and provinces based on SDG scores, rewarding high-performing provinces, and either imposing a penalty or performing interventions on low-performing or "aspirational" provinces. *Niti Aayog*, an Indian government public policy think tank, has established a nodal agency for coordinating and monitoring Sustainable Development Goals in India known as SDG Vertical[4]. The objective of the agency is to monitor progress under the SDG framework and related initiatives at national and sub-national levels. Participating states under this program have embarked on data-driven approaches to address different indicators in order to improve the scores on one or more targets.

This paper presents our ongoing work in building a Policy Support System (PSS) that integrates datasets from different sources to build explainable Bayesian models of different indicators and perform analyses of policy interventions on them.

2. Related Literature

Several web-based initiatives called Policy Support Systems (PSS) have emerged to support policy formulation and intervention around SDGs (L 2017; Mulligan et al. 2020; Malhotra, Anand, and Singh 2018; Rıo Castro, Fern´andez, and Colsa 2020). A PSS is distinguished from Decision Support Systems (DSS) in that, PSS are technologies designed for policy formulation and the intervention itself, rather than just supporting specific decisions. Supporting policy formulation includes model building, simulation of complex inter-dependencies, and counter-factual or "what-if" analyses. Some example PSS initiatives around the world, include the following: Guppy (L 2017) proposes a Big Data-based policy support environment called SDG-PSS[5] to effectively deliver on SDG 6 (clean water and sanitation) targets[6]. Another web-based policy support system called Costing Nature helps in the development and implementation of conservation strategies focused on sustaining and improving ecosystem services[7]. The application of this tool is (Mulligan et al. 2020) to identify priority areas for sustainable management and to realize targets under SDG 6 at the country scale for Madagascar and the basin-scale for the Volta Basin. Within this SDG 6 priority areas footprint, Costing Nature also assesses synergies and trade-offs provided by this land for SDG 15 (life on land), SDG 13 (climate action), and SDG 2 (zero hunger).

The need for a PSS has been expressed in different ways by several other researchers as well. (Griggs et al. 2013) stress the need for a systemic approach to achieving SDGs and the need to

---

[3] UN SDGs https://sdgs.un.org/goals

[4] Niti Aayog Verticals Sustainable Development Goals https://www.niti.gov.in/sdg-vertical

[5] SDG 6 Policy Support System: https://sdgpss.net/en/

[6] UN SDG Goal 6: https://sdgs.un.org/goals/goal6

[7] CostingNature: http://www.policysupport.org/costingnature

go beyond decision-making in silos. They also propose an integrated framework (Griggs et al. 2014) combining six SDGs to collectively track indicators and targets and track the ways interventions towards one target affect others. (Thomas et al. 2021) emphasize the need for adaptive and dynamic models to support policymaking in the context of SDGs and how interventions are often rendered ineffective due to external factors. An integrated PSS model, MedAction PSS (Van Delden, Luja, and Engelen 2007) attempts to simulate bio-physical and socio-economic developments in Mediterranean watersheds by addressing the issue of policy-relevant outputs for policy-makers by designing a ternary diagram depicting the evolution of the system as described by Environmentally Sensitive Areas (ESA), Water Shortage and Farmer's Profit indicators. Despite the increasing interest in PSS for sustainable development, we still find a dearth of essential hermeneutics to reason about sustainability. As acknowledged by (Griggs et al. 2014), policy interventions do not happen in a vacuum. Policy interventions affect dynamical systems (societies and communities) that have their own self-interest, and interventions at a given place towards a given target often affect (either positively or adversely) the prospects for neighboring regions and other goals as well. Policy interventions may also result in unintended consequences and/or push-backs from the population – the so-called Cobra Effect[8]. In our ongoing work on designing a PSS for sustainable development, we address the problem of computational modeling of sustainability, and the effect of policy interventions.

3. Policy Support System

We propose a *Policy Support System* that models a policy's goals, targets, indicators, and sub-indicators to understand the impact of policy interventions using Big Data together with Bayesian networks to perform causal inference under uncertainty. *Policy Enunciator* as part of the PSS architecture acts as a language for policy-makers that highlights policy-related issues upon intervention. Bayesian belief networks are deployed to mimic real-world scenarios to model the domain of interest. Bayesian network is a probabilistic graphical model where each node represents a random variable and the arcs represent dependence relationships among different variables of the graph. To carry out the task of policy enunciation, we develop Bayesian networks for all the leaf nodes of the SDG tree. Fig.1 represents the "SDG-2 Tree" inspired by the state government report assessing agricultural productivity across various districts of the south Indian state of Karnataka[9]. The indicator is set as the target node of the network. It becomes the focus of our investigation where multiple factors with possible mutual causal relationships affect the likelihood of the occurrence of the indicator. The conditional inter-dependencies among the possible causal associations constitute the whole network.

To prepare a hand-crafted Bayesian network, a "dataframe'" is constructed. This is done by identifying various independent and dependent factors responsible for the key indicator of a model as shown in Fig.2. The construction helps in quickly identifying the factors to prepare a directed acyclic graph based on the availability of data.

A Bayesian network is developed for each SDG indicator selected from the SDG tree. They collectively form a "model ensemble". A model ensemble graphically represents a Sustainable Development Goal that constitutes the Policy Enunciator's model library. A model ensemble may have one or more nodes common across other models. It is evident that a model's objective to capture possible causal relationships for say '2.2.1: Percentage of children under 5 years who are stunted ' and '2.2.4: Percentage of children aged 0-4 years who are underweight

---

[8] The Cobra Effect https://fee.org/articles/the-cobra-effect-lessons-in-unintended-consequences/

[9] Application of Data Science to Achieve Sustainable Development Goal-2, Interim Report-1, Public Affairs Center, December, 2020

’ may certainly have one or many common aspects that affect the target variables[10].

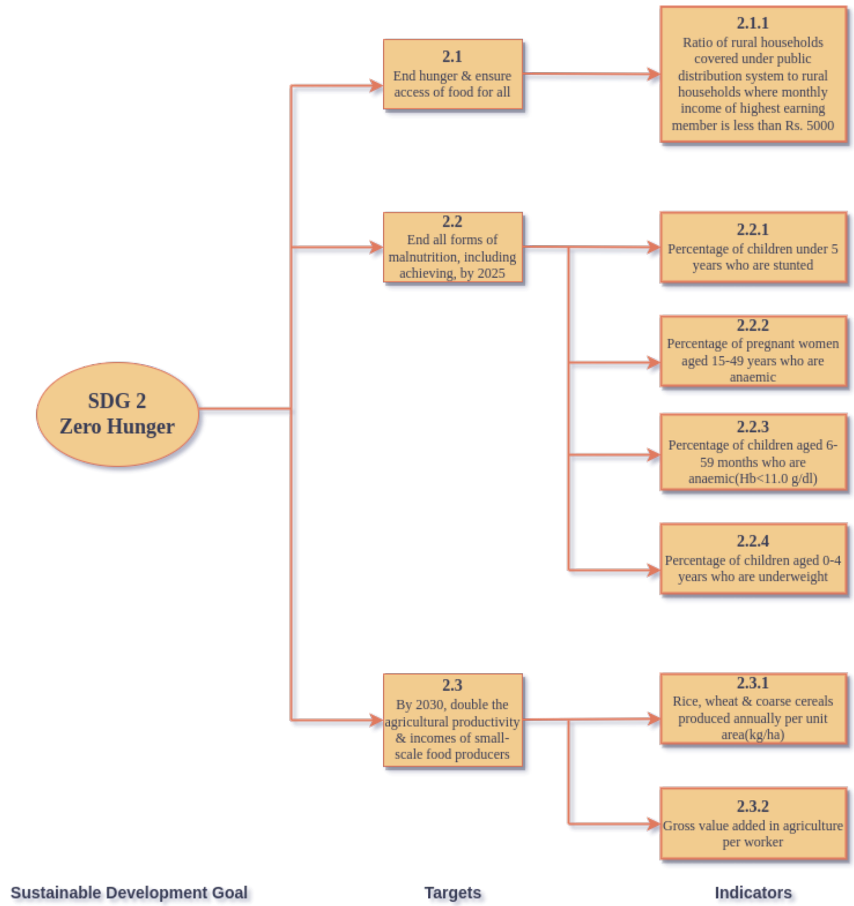

Figure 1: SDG-2 Tree

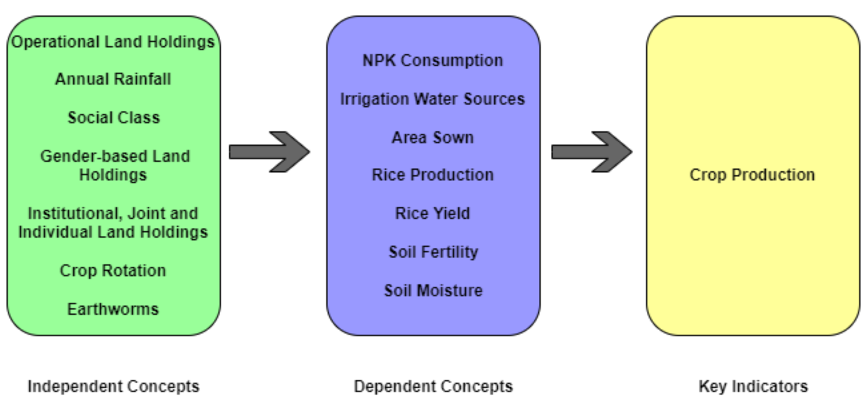

Fig. 2 Identification of independent, dependent concepts to form a “dataframe”

[10] WHO Malnutrition Factsheet https://www.who.int/news-room/fact-sheets/detail/malnutrition

3.1 Causal Inference in Policy Scenarios

The Policy Enunciator subsystem focuses on *intervention modeling*. Policy intervention modeling entails representing a policy decision in terms of changes in one or more variables. For instance, a policy decision mandating high consumption of 'fertilizers', would impact the overall 'rice productivity in the model shown in Fig. 3. A given policy specification may affect multiple such variables in multiple models. Also, there could be variables in common across multiple models.

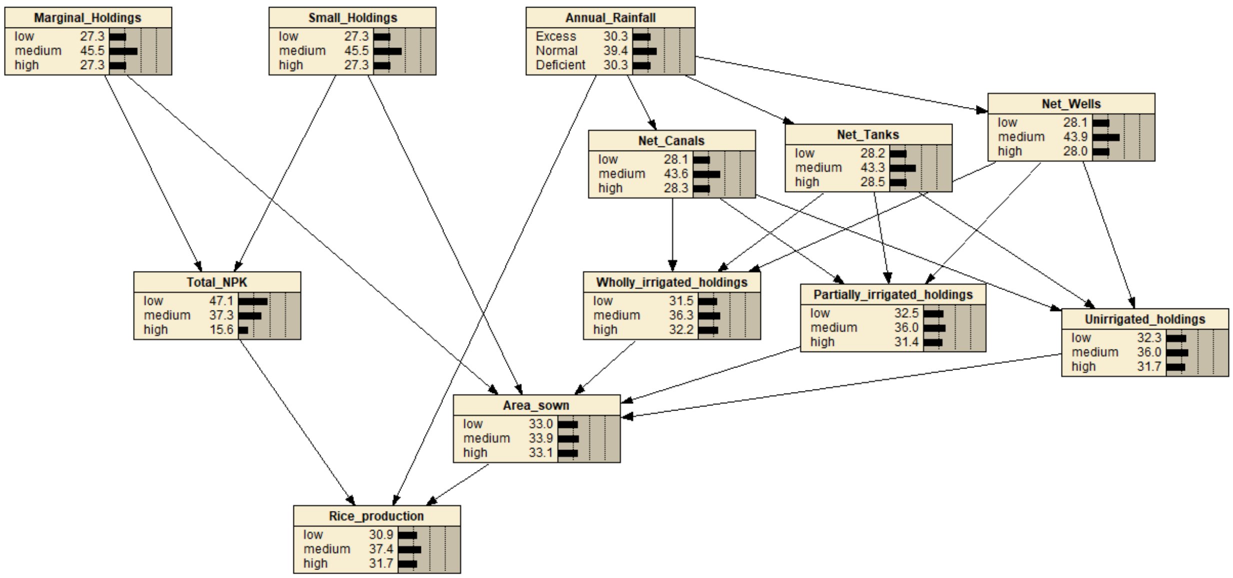

Fig. 3 Rice Production network in Karnataka

Hence, a policy intervention would mathematically translate into the *conditioning* of several variables across several models. Conditioning of these variables would propagate changes in variables downstream, leading to some change in the target variable of the respective Bayesian network. Such impacts are called *downstream impacts*. Policy interventions in one node also tend to have side effects on other nodes, and achieving the sustainability goals would require the formulation of policies that consider side effects and push-backs. We term these as *lateral impacts*. A use-case for 'Public Distribution System'[11] with direct and indirect effects is shown in Fig. 4. Identical variables (nodes) are represented with similar colors. Intervening on the nodes may show direct effects such as "stability in food prices", "reduction in hunger" or may have lateral impact of creating "bogus ration cards" for procuring food grains from fair price shops.

The primary motivation behind building the Policy Enunciator is to identify the downstream and lateral impacts within and across models when we intervene in one of the variables of the model.

3.2 Intervention Impacts

Intervening on a variable not only governs the target within its own model but also affects outcomes in other models that comprise matching nodes. Identifying lateral impacts is a challenge but downstream impacts can be graphically observed. Separate Bayesian networks are learned from data for different administrative subunits (in this case, districts). The probability distribution of the target variable "rice production" is reduced to a single weighted score.

[11] Public Distribution System https://dfpd.gov.in/pd-Introduction.htm

$$score = \sum_i w_i p_i$$

where w_i is the weight assigned to the level i of the node and p_i is the probability of i. We calculate an Average Treatment Effect (ATE) on the districts by deducting the mean score of the control group from the mean score of the treatment group.

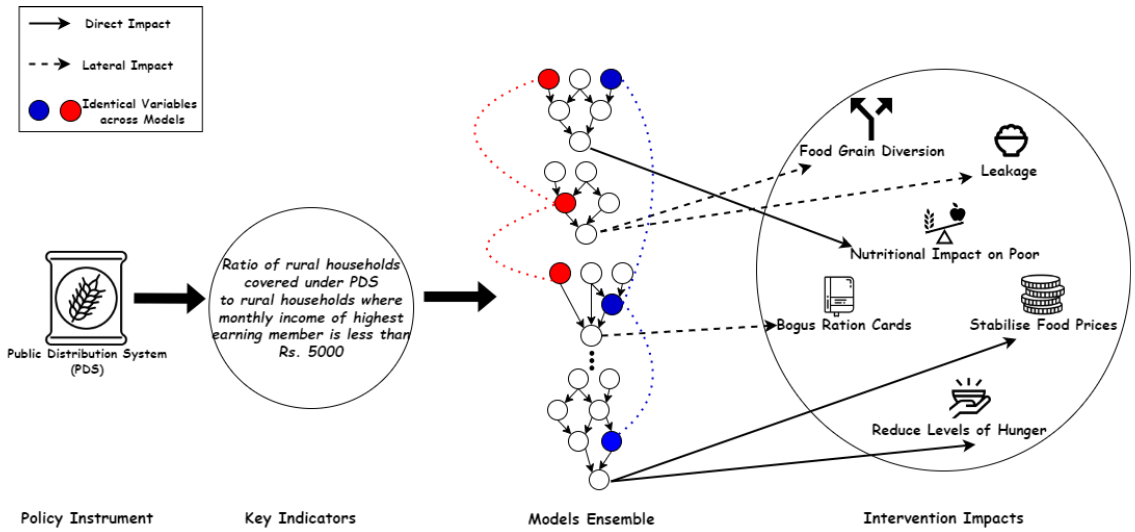

Fig. 4 Public Distribution System Use-case for Policy Enunciation

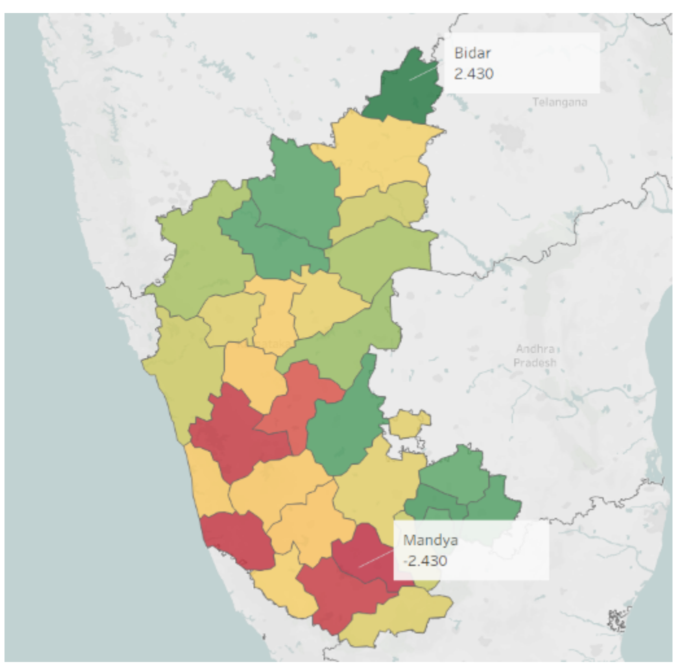

Fig. 5 Average Treatment Effect seen after ‘increased fertilizer consumption’ policy intervention in the districts of Karnataka

We perform a policy intervention ‘increase fertilizer consumption’, and visualize it over all the districts of Karnataka as seen in Fig.5. Green-colored regions have high ATEs and red-colored zones are negatively impacted with low values of ATE. We see a positive repercussion of the intervention in the district of ‘Bidar’ which is in the northeastern part of Karnataka and a negative impact of intervention in the ‘Mandya’ district in south Karnataka. Mandya district has fertile land and is known as the land of 5 rivers. It is also famously known as ‘Sugar City’ because sugarcane is a major crop grown here[12] whereas, Bidar falls in the dry belt of northern Karnataka[13]. We see conflicting impacts of fertilizer consumption in both districts. A fertile district like Mandya shows negative ATE whereas a dry district Bidar shows a positive treatment effect. The values convey to the policymaker that imposing the same policy intervention uniformly across all the regions of the state is not desirable. The same intervention may leave one region better and make many other worse off which is not a Pareto improvement.

3.3 Intervention Core

A Pareto improvement is an action that should benefit at least one individual but no one should be worse off (Pareto 1896). Given a set of administrative regions (districts) $r_1, r_2, ...r_n$ over which, an intervention p is proposed for improving the target variable t:

Let $t(r_i)$ be the score for variable t in region r_i and let $p_t r_i$ be the updated score for variable t in region r_i after policy intervention p. Policy intervention p for target variable t is said to result in a Pareto improvement iff:

$$\forall i,\ p_t(r_i) \geq t(r_i)\ \wedge\ \exists j,\ p_t(r_j) > t(r_j)$$

Let interventions $p_1, p_2, ..., p_n$ be a set of interventions that are shown to result in Pareto improvement for target variable t. For a given intervention p_k, let the term $v(p_k)$ indicate the set of all variables affected by the intervention. A given intervention p_k is said to be a “core” intervention iff:

$$\nexists p_m,\ v(p_m) \subset v(p_k)$$

4. Conclusion

This position paper presented our ongoing work on designing a PSS for sustainable development. The primary departure we make in this work is to look at sustainability from a holistic perspective modeling both downstream and lateral effects of policy interventions. This work is currently under active development with building Bayesian models for different targets taking the major chunk of the construction activity. We hope that the proposed PSS would be generic enough to be applicable to sustainability-related interventions anywhere in the world, given relevant data.

Acknowledgments

The authors thank the Department of Planning and Statistics, Government of Karnataka, India, and our collaborating agencies, Centre for Open Data Research(CODR), Public Affairs Centre, Jigani, Bengaluru, India.

---

[12] Mandya https://mandya.nic.in/en/

[13] Bidar https://bidar.nic.in/en/

References

Caradonna, Jeremy L. 2014. *Sustainability: A history.* Oxford University Press.

Dresner, Simon. 2012. *The principles of sustainability.* Routledge.

Griggs, David, Mark Stafford Smith, Johan Rockstr¨om, Marcus C Ohman, Owen Gaffney, Gisbert Glaser¸ ¨ Norichika Kanie, Ian Noble, Will Steffen, and Priya Shyamsundar. 2014. "An integrated framework for sustainable development goals." *Ecology and society* 19 (4).

Griggs, David, Mark Stafford-Smith, Owen Gaffney, Johan Rockstr¨om, Marcus C Ohman, Priya Shyamsundar¸ ¨ Will Steffen, Gisbert Glaser, Norichika Kanie, and Ian Noble. 2013. "Sustainable development goals for people and planet." *Nature* 495 (7441): 305–307.

L, Guppy. 2017. "Accelerating Water-Related SDG Success; 6 Steps and 6 Components for SDG 6." *UNU INWEH Policy Brief, Issue 4 United Nations University Institute for Water Environment and Health. Hamilton, Ontario, Canada.*

Loorbach, Derk. 2007. "Governance for sustainability." *Sustainability: Science, Practice and Policy* 3 (2): 1–4.

Lubin, David A, and Daniel C Esty. 2010. "The sustainability imperative." *Harvard business review* 88 (5): 42–50.

Malhotra, Charru, Rashmi Anand, and Shauryavir Singh. 2018. "Applying big data analytics in governance to achieve sustainable development goals (SDGs) in India." In *Data Science Landscape,* 273–291. Springer.

McMichael, Anthony J, Colin D Butler, and Carl Folke. 2003. "New visions for addressing sustainability." *Science* 302 (5652): 1919–1920.

Mulligan, Mark, Arnout van Soesbergen, David G Hole, Thomas M Brooks, Sophia Burke, and Jon Hutton. 2020. "Mapping nature's contribution to SDG 6 and implications for other SDGs at policy relevant scales." *Remote Sensing of Environment* 239:111671.

Nussbaum, Martha. 2003. "Capabilities as fundamental entitlements: Sen and social justice." Feminist economics 9 (2-3): 33–59.

Rıo Castro, Gema del, Marıa Camino Gonz´alez Fern´andez, and ´Angel Uruburu Colsa. 2020. "Unleashing the convergence amid digitalization and sustainability towards pursuing the Sustainable Development Goals (SDGs): A holistic review." Journal of Cleaner Production,122204.

Pareto, Vilfredo. 1896. *Cours d'´economie politique: profess´e `a l'Universit´e de Lausanne.* Vol. 1. F. Rouge.

Robeyns, Ingrid. 2005. "The capability approach: a theoretical survey."Journal of human development 6 (1): 93–117.

Sen, Amartya. 2005. "Human rights and capabilities." Journal of human development 6 (2): 151–166.

Thomas, A, BJ Cosby, R Gooday, H Lyons, G Siriwardena, E Kettel, R Matthews, K Beauchamp, M Petr, and BA Emmett. 2021. "Rapid adaptive modelling for policy support towards achieving Sustainable Development Goals: Brexit and the livestock sector in Wales." Environmental Science & Policy 125:21–31.

Van Delden, Hedwig, Patrick Luja, and Guy Engelen. 2007. "Integration of multi-scale dynamic spatial models of socio-economic and physical processes for river basin management." Environmental Modelling & Software 22 (2): 223–238.