Contents lists available at ScienceDirect

# Environmental Development

journal homepage: www.elsevier.com/locate/envdev

# Assessing the application of gender perspectives in land restoration studies in Ethiopia using text mining

Wuletawu Abera [a,*], Lulseged Tamene [a], Dawit Mekonnen [b],
Natalia Estrada Carmona [c], Marlène Elias [d], Deepa Joshi [e], Ermias Aynekulu [f]

[a] *Alliance of Bioversity International and International Center for Tropical Agriculture (CIAT), Addis Ababa, Ethiopia*
[b] *International Food Policy Research Institute (IFPRI), Addis Ababa, Ethiopia*
[c] *Alliance of Bioversity International and CIAT, France*
[d] *Alliance of Bioversity International and CIAT, Italy*
[e] *International Water Management Institute (IWMI), Colombo, Sri Lanka*
[f] *Centre for International Forestry Research – World Agroforestry (CIFOR–ICRAF), UN Avenue, P.O. Box 30677-00100, Nairobi, Kenya*

A R T I C L E   I N F O

A B S T R A C T

Restoration of degraded land is key to enhancing land productivity and farmers' wellbeing in sub-Saharan Africa. Evidence shows that the benefits of land restoration are tremendous, ranging from biophysical benefits in soil health, agricultural productivity, ecosystem services, to socio-economic dimensions such as improving farmers' income and livelihoods. Yet one issue that is rarely considered is how the outcomes of restoration initiatives affect different social groups, specifically women, men, and the marginalized. Our study aims to understand the extent of gender and social inclusion in land restoration studies in Ethiopia. Through analyzing 314 peer-reviewed scientific articles on land restoration initiatives and studies for Ethiopia published between 1985 and 2020, we aimed to: i) develop a novel approach that would allow for the semi-automatization of analyzing gender aspects in a large set of scientific documents using machine learning techniques, ii) assessed the current evidence on gender integration in peer-reviewed studies related to RDL, and iii) identified what gender issues are discussed in RDL literature. We employed text mining techniques to analyze the literature for keywords and to classify articles into three main classes: *gender blind, gender sensitive,* and *gender transformative*, based on the extent to which gender and social inclusion keywords appeared in the studies. Our analysis showed that gender and social inclusion are rarely considered in landscape restoration studies in Ethiopia, as most studies and interventions are heavily focused on the biophysical and environmental aspects of restoration. Despite an exponential increase in the use of gender and social inclusion keywords after 2005, most of the research (58 percent of articles) remains *gender blind*. The most frequently used related terms remained generic, e.g. community, access, participatory, and tenure. Only 38% of the studies could be classified as *gender sensitive* to a certain degree, and a strikingly low number of articles (4 percent) – which consider the systematic challenges women face – could be defined as *gender transformative*. This implies that scientific evidence about how land restoration has benefited men and women, and to what extent it could transform gender and power dynamics, is lacking. Our findings demonstrate the urgent need to consider gender and social inclusion to ensure associated benefits of land restoration practices.

---

* Corresponding author.
*E-mail address:* Wuletawu.Abera@cgiar.org (W. Abera).

## 1. Introduction

Land degradation and climate change are two major processes that impact biogeochemical cycles (e.g. water, carbon, and nitrogen), affecting the capacity of land to sustainably provide the ecosystem services — including agricultural productivity and biodiversity conservation — needed for human and other species' survival and wellbeing (Mooney et al., 2009; Montoya and Raffaelli, 2010; Sannigrahi et al., 2020; Orwin et al., 2015; Lamarque et al., 2014; Sutton et al., 2016; Bai et al., 2013). Land degradation also increases agroecological systems' exposure to climate change and reduces the effectiveness of adaptation options (Webb et al., 2017). At least 40 percent of the world's population are negatively impacted by land degradation and escalating climate change and its impacts, such as biodiversity loss, resulting in increased conflict over available land and resources, massive human migration (IPBES, 2018), and reduced food availability due to lower crop productivity (Wheeler and Von Braun, 2013).

Land restoration is considered to be the main approach to solving land degradation and is a sustainable strategy to recover lost ecosystem services and reverse soil deterioration to enhance agricultural productivity and benefit livelihoods in Ethiopia (Abera et al., 2020; Mekuria et al., 2011; Balehegn et al., 2019; Araya et al., 2015; Araya et al., 2016; Desta et al., 2021; Temmerman et al., 2013; Laughlin, 2014).

On the other hand, relatively little evidence can be found on the social and economic benefits of restoring degraded lands (RDL), particularly related to the long-term impacts on gender roles and norms, and the equitable sharing of benefits from restored lands. There is, however, evidence that shows the linkages between land and water management interventions, degraded ecosystems, and gender inequalities (Broeckhoven and Cliquet 2015; Collantes et al., 2018). Yet despite potential synergies between restoration initiatives and gender equality outcomes, the gender dimension in land restoration research remains poorly addressed (Sijapati Basnett et al., 2017; Elmhirst et al., 2017).

The gender dimension in restoration is experienced in differentiated costs, risks, and benefits for different groups of women and men, shaped by various contextually-specific, socio-economic, and cultural factors. In addition, social norms and gender roles can reduce empowerment and limit the role women play in promoting land restoration activities (Catacutan and Villamor, 2016; Agarwal, 2018). For example, access to resources such as land, credit, technology, and knowledge about new practices is more limited for women in developing countries than for men (Catacutan and Villamor, 2016; Agarwal, 2018). When women in rural communities do not hold decision-making positions and their voice are not heard they may lose interest or lack the capacity to engage in land restoration activities (Forsythe et al., 2015). Furthermore, enhancing the value of land through restoration can affect the desirability of the land, and may thus pose the risk of land dispossession for those with insecure land rights (Elias et al., 2021; Sijapati Basnett et al., 2017). This implies that land, water, and ecosystem degradation can not only disproportionately affect marginalized communities and women, but that restoration activities could also exacerbate existing inequalities if not properly taken into account during the design and implementation of restoration interventions. Because asset-poor groups and individuals are highly dependent on common-pool resources, such as forests, water bodies, and ecosystems, restoration initiatives that affect these areas will significantly impact their livelihoods and wellbeing (Adams et al., 2016). Additionally, there are gendered norms around women's and men's access to, control of, and responsibilities related to land, water, and other environmental resources. For example, initiatives that are labor intensive can exacerbate women's pre-existing heavy labor burdens, without ensuring their control of any associated benefits. Thus, restoration initiatives have gendered implications for women's and men's time, labor, benefits, and exclusions. Understanding and addressing these issues is critical to ensuring that restoration initiatives do not further exacerbate marginalization and inequalities and, at the very least, that they serve to maintain the existing gender and social equality in landscape restoration (Elias et al., 2021; Sijapati Basnett et al., 2017).

Conversely, studies show that the inclusion of marginalized communities and women in tenure security have a high likelihood of engaging women in restoration (McLain et al., 2021; Elias et al., 2021). For instance, a study by Collantes et al. (2018) showed that the gender-responsive land certification and registration process undertaken in Ethiopia in the early 2000s increased landowners' likelihood of investing in soil and water conservation measures by 20–30 percent. Gender-inclusive restoration initiatives and gender-transformative approaches that can disrupt the underlying causes of the interconnected inequalities that women face in relation to water (irrigation and domestic), land (agriculture), and ecosystem services (access to forests and related natural resources) can facilitate women's equal and meaningful participation, and leadership in natural resource governance (Collantes et al., 2018; Sijapati Basnett et al., 2017). Contributions of women's participation, decision, and leadership role in land restoration - are limited across Africa (de Siqueira et al., 2021; Mbile et al., 2019; Coleman and Mwangi, 2013). One notable exception is the Green Belt Movement, led by the Nobel Prize winner Wangari Maathai in Kenya (1940–2011). This is a unique example of a social justice initiative on restoration and conservation based on the needs of marginalized rural women (Leigh 2005).

The focus on Ethiopia is a conscious choice, because of the unique situation of land ownership and rights in Ethiopia. Given the country's communist political history, more particularly, land reform and nationalization in March 1975, the State became the only legitimate owner of land in the country. Subsequent reforms and changes in the political system in Ethiopia have resulted in significant changes in the State's ownership and control of land, nonetheless there are distinct ambiguities and variations in relation to land ownership across different regions of the country. As part of an ongoing effort to formalize land holdings in recent decades, the Ethiopian government introduced a joint titling scheme, the Revised Family Code (2000), which would improve tenure security, for men as well as for women (Namubiru-Mwaura, 2014). Likely as a result of these joint titling requirements, the country has a relatively small gender gap in landholdings as compared to global averages (Doss et al., 2015). However, for women who receive only usufruct rights to land, without ownership, there has been little change in practice for individual land use title holders, particularly since they are prohibited from selling or mortgaging land (Deininger et al., 2008). There is no rural land sale market, and the land rental market is

restricted except for Amhara (Deininger et al., 2008, Deininger et al., 2011).

As customary inheritance practices are largely patrilineal, this can serve to restrict women's land rights (Girma and Giovarelli, 2019). The primary way to gain access to land (to use) is by inheritance from parents, with older sons given preferential treatment (Bezu and Holden, 2014; Kosec et al., 2018). Women access land by marrying men with land (Bezu and Holden, 2014) or living with an adult son who inherited the land rights" (Komatsu et al., 2022; 3). Furthermore, joint titling does not necessarily lead to joint management and decision-making, nor does it equate to equality and parity in size and quality of landholding (Kumar and Quisumbing, 2010).

A social norm exacerbated and reinforced by patrilocal customs - that the husband is in charge of land management, and that only men may cultivate land with oxen means that women landholders commonly resort to renting out much of their land. Women (and especially women living in male-headed households) are least likely to adopt restoration practices and agricultural innovations to address degradation, due to a lack of access to resources including land, water and labor; capital and credit; and extension services (Crossland et al., 2021). Women also have a higher labor burden associated with unpaid care work, which prevents them from spending as much time on restoration practices as men (Kumasi and Asenso-Okyere 2011). In sum, as a result of their relatively marginal landholdings and cultural restrictions concerning land management, women are often the most vulnerable to degradation and simultaneously the least able to participate in restoration initiatives.

This paper examines the gender perspectives in the existing scientific knowledge base on restoration of degraded lands, benefits and impacts of restoring degraded landscapes (RDL) with the aim of making RDL efforts and studies more gender transformative using Ethiopia as a case study. Specific objectives are to: i) develop a novel approach that would allow semi-automatization of analyzing gender aspects in a large set of scientific documents (i.e. through machine learning), ii) assess the current evidence on gender integration in peer-reviewed studies related to RDL; and iii) identify what gender issues are discussed in RDL literature. Using a text mining approach, we generate robust scientific evidence on the extent to which gender and other social inequalities are integrated into RDL initiatives and studies in Ethiopia to better understand and address the key challenges and capitalize on opportunities for inclusive and sustainable development. We hope this approach can help researchers to spot/identify timely gender and social inclusion patterns, topics and gaps that can be observed and improved for future studies.

## 2. Methods

### 2.1. Document sourcing and searching

Our study considered literature focused on Ethiopia, where a large number of land restoration initiatives and studies have been undertaken/implemented (Abera et al., 2020). Variations and combinations of the following search terms were used to find articles specific to this study: land restoration, forest landscape restoration, sustainable land management, soil water conservation, land rehabilitation, Ethiopia, impact, benefit(s), livelihood(s), gender, community, female headed, social transformation, woman, female, inclusion, equality, and equity. The search for literature, i.e. peer-reviewed articles, was conducted in July 2020. Specifically, we used the following search codes ("Ethiopia" AND "land restoration" OR "forest landscape restoration" OR "sustainable land management" OR "soil water conservation" OR "land rehabilitation" AND/OR "impact, benefit" OR "livelihood" OR "gender" OR "community" OR "female headed" OR "social transformation" OR "woman" OR "female" OR "inclusion" OR "equality", OR "equity"). Databases searched included the CGIAR publication database, Wiley Online Library, JSTOR, Elsevier, Science Direct, Sage Journals, Taylor and Francis Online, ProQuest, Springer Science and Business, ResearchGate, United States Department of Agriculture: National Agricultural Library, Semantic Scholar, and CORE. Snowballing enabled us to gather literature beyond the above-cited venues. We additionally used Google Scholar to locate open access articles. Based on the title and abstract review, articles that did not fall into the specified themes or geographic area of Ethiopia are excluded manually.
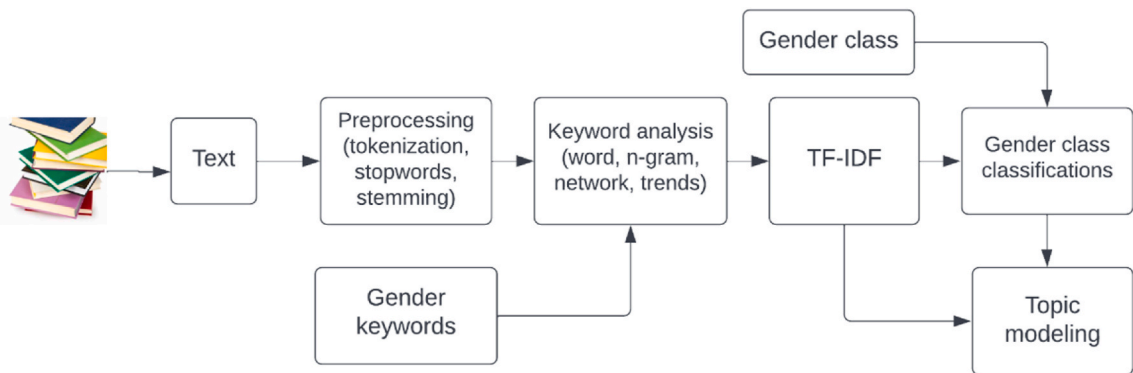


**Fig. 1.** Methodological workflow used to assess the gender responsiveness of scientific literature on land restoration in Ethiopia (Key: TF-IDF: Term frequency - inverse document frequency).

### 2.2. Text mining and modeling

Text mining is an efficient and automated process to derive high-quality information from a large set of documents (Cohen and Hunter, 2008; Rzhetsky et al., 2009). In this study, we text mined 312 peer-reviewed articles. This text-mining approach followed the following steps: (i) preprocessing text for a clean, comparable, and useable text format; (ii) identifying and extracting gender and social inclusion keywords mentioned in the land restoration literature; (iii) analyzing keyword (in general) and keyword (inclusion-related) frequencies and associations; (iv) evaluating gender-inclusion levels in each peer-review article; and (v) developing topic modeling to disentangle the most important topics discussed in gender-sensitive or -transformative RDL literature in Ethiopia (Fig. 1).

#### 2.2.1. Preprocessing text

Extraction of information from a large set of documents using text mining requires a set of preprocessing steps. The first operation is the conversion of pdf to text format using the pdf tools R package (Ooms, 2020). Subsequently, the raw text is converted to smaller text units (a process called tokenization) including character, word, sentences, lines, paragraphs, and n-grams). In this study, we have segmented the raw data into words and n-grams (particularly 2-g) using *tidytext* (Silge and Robinson, 2016) package. N-grams display the words that co-occurred the most (a 2-ngram is, for example, natural + resources).

In natural language processing (NLP), words that ensure sentence structure and grammar but contain little information are called stopwords (e.g., is, of, was, the). We used the snowball package (Bouchet-Valat and Bouchet-Valat, 2020) to remove stopwords, which reduces computational demand. The third and important preprocessing step is stemming, a process of reducing each word to its basic forms. For example, crop, crops, cropping, cropped will all be reduced to crop* (Bouchet-Valat and Bouchet-Valat, 2020). We used the snowball R package for stemming words.

#### 2.2.2. Identifying and extracting gender and social inclusion keywords

We identified a set of gender and social inclusion keywords through an iterative process, relying on the gender and social inclusion literature, co-authors, and gender experts to identify a set of keywords relevant to refer/infer aspects of gender and social inclusion (Appendix 1). For example, we screened articles for gender and social inclusion keywords such as: affirmative, benefit, co-benefits, community, disaggregate, elder, engagement, equal, equality, equitable, equity, female, gender, gender-responsive, income, inclusiveness, livelihood, mainstream, ownership, participation, participatory, power, sex, smallholder, tenure, women and youth. We used the stemmed version of the keywords and identified matching keywords with the R package fuzzyjoin (Robinson and Silge, 2020).

The keyword screening process provides insights on the level of overall attention paid to gender and social inclusiveness in the land restoration literature. We quantified the frequency of each keyword in every article to explore and characterize the nature of the gender representations in the literature. We then supplemented frequency analysis with network analysis to further identify connection of keywords. In network analysis, words are depicted as nodes, while their co-occurrence is represented as edges or links. The co-occurrence analysis is conducted by aggregating all articles that contain co-occurring keywords. For an article to be included in this analysis, we set a minimum of three co-occurrences of the specified keywords. By comprehending the network's structure, including central nodes, clusters, and branches, this approach provides valuable insights into the text's underlying themes and relationships.

#### 2.2.3. Analyzing word and keyword frequencies and associations

We used term frequency (tf) to identify the most frequently occurring words in each article and through time. Additionally, we weighted article word and keyword frequencies by the inverse of their frequency across all articles, a metric called term frequency-inverse document frequency (tf-idf) statistics (Tokunaga and Makoto, 1994; Havrlant and Kreinovich, 2017). The tf-idf is a statistical measure that helps to identify relevant words in an article and a group of articles, and is calculated as follows (Ramos, 2003):

$$tf - idf(t, d, D) = tf(t, d) * idf(t, D)$$

where *t* is for a specific term or word, d represents a specific article, *tf (t,d)* is the frequency of *t* in *d*, *D* represents all articles, *idf(t,D)* is the inverse article frequency, which can be viewed as a weight. The *idf (t, D)* is estimated as:

$$idf(t, D) = log \frac{N}{n_t}$$

where *N* is the total number of articles, and $n_t$ is the number of articles with the term *t* in it. We used the R *tidytext* package to generate the tf-idf statistics for each word and document (Robinson and Silge, 2020). We explored word and keyword associations through a network analysis, which helps to understand what words or keywords are central to the network and keep it connected.

#### 2.2.4. Identifying the level of gender sensitivity and inclusion

Adapting work done by Pederson et al. (2014), which provided gender classifications for projects and programs based on their level of engagement for gender considerations, we classified peer-reviewed articles into three main groups: *gender blind*, *gender sensitive*, and *gender transformative*. Articles that did not include any account of gender — either through disaggregation of data by sex, or some kind of inference or discussion of gender and social inclusion — were categorized as *gender blind*. *Gender-sensitive* (or gender-aware) studies are those that consider gender roles, differences, and/or inequalities (e.g., in access to resources), but do not consider their underlying causes. If an article presented data or analysis by gender in any section, it was categorized as *gender sensitive*. Finally, *gender*

*transformative* articles were research that aims to understand the systemic bottlenecks and barriers to gender equality and/or how to challenge these barriers in order to advance gender equality. With regard to restoration, examples of this research include how gender norms limit women's land rights and restoration opportunities, their rights to restoration benefits, and/or how restoration initiatives can challenge discriminatory norms that limit women's resource rights and incomes.

With our gender experts, we manually classified the articles into three gender classes based on a full text reading. Then we used 75% of the articles iteratively to train the machine learning model. The model analyzed those documents and extracted features (e.g., rare, unique and common words in each document and across classes) which allowed the model to "understand" why the article is assigned to a certain category (i.e. gender blind, gender sensitive or gender transformative). We used multinomial regression, based on *glmnet* R package, because it has shown better performance in other studies (Friedman et al., 2021). The trained model is tested with 25%of the articles. To obtain the best combination of *glmnet lesso* model parameters, we used 10-fold cross validation for the training dataset. In addition to the overall model performance, we used a confusion matrix between manually-assigned categories by experts and automatically assigned categories by the predictive model to further examine the predictive capacity of the algorithm for the three gender classes.

### 2.2.5. Topic modelling

We conducted topic modelling to examine the types of gender issues and topics raised in relation to land restoration in Ethiopia. To this end, topic modeling (Nikolenko et al., 2017) was done on the aggregate of gender inclusive (gender sensitive and transformative) articles because there were too few articles to analyze each category. Topic modeling is an unsupervised machine learning technique that does not require a predefined list of tags and training data prepared by humans. We used *topicmodels* R package (Grün and Hornik, 2011), which uses Latent Dirichlet allocation (LDA) algorithm (Blei et al., 2003) to map each article into a set of topics and a set of words for a given topic. Topic modeling tools provide a fast and feasible way of extracting latent semantic structure in documents and establishing links between the latent topics, which would otherwise be impossible or time consuming for a large dataset. Topic coherence and prevalence are the two indexes used to estimate the number of topics and also as diagnosis of quality of defined topics. Coherence is defined as the average or median of pairwise word similarities formed by top words of a given topic. It is a measure of how often the top tokens in each topic appear together across all articles. Prevalence of a topic is the probability of topics distribution across all articles (Röder et al., 2015; Maier et al., 2018). As abovementioned, the two metrics are estimated for articles classified as gender sensitive and gender transformative classes.

## 3. Results and discussions

### 3.1. Overall gender-related n-grams (keywords) analysis

Our literature sourcing resulted in 312 peer-reviewed articles covering a wide range of approaches in land restoration, from testing specific methodologies to impact assessments of land restoration on ecosystem services. Results show that the common elements discussed in these articles are disproportionately centered on biophysical aspects and specifically on key environmental resources such as land, soil, water, and forests, and their management and conservation (Fig. 2a). In land restoration articles, the top-four 2-g are: land degradation, water conservation, soil erosion, and land management (Fig. 2b). Northern Ethiopia is also frequently mentioned, indicating that this area has been the center of the land degradation studies in the country. This is in line with the meta-analysis by Abera et al. (2020) that shows that most studies have been conducted in the northern part of Ethiopia due to the high prevalence of land degradation in the region compared to the rest of Ethiopia.

The most-frequently mentioned gender and social inclusion keywords across articles include: community, participation, benefits, tenure, ownership, and smallholder, each mentioned in above 100 articles (32 percent) (Fig. 2c). Words such as female, gender, and sex were mentioned in less than 10 percent of the articles (Fig. 2c). Keywords such as equality, equity, equal, and norms, suggesting a higher level of inclusion, are mentioned in only 5–10 percent of the articles. Keywords pertinent to gender-sensitive and gender-transformative concepts — such as negotiation, mainstreaming, leadership, ethnicity, and equality — were present only in 3–5 articles (Fig. 2c). Gender and social inclusion keywords that were absent from the land restoration literature for Ethiopia include: ancestry, disability, diversity, identity, intersectionality, minority, multicultural, and many other important terms that relate to social diversity (Fig. 2c). This suggests that the scientific knowledge and work on land restoration poorly addresses gender and social inclusion in relation to land restoration in Ethiopia. However, the 2-g analysis on the gender and social inclusion keywords suggests that aspects such as community base, smallholder farmers, tenure security, and income generation are frequently discussed (Fig. 2d).

Our network analysis of the keywords (Fig. 3a and b) reinforces the focus on biophysical and environmental aspects in RDL research but shows the dominant role of keywords such as soil, land, farm, agriculture, and develop (Fig. 3a). Words such as fertility, moisture, nutrient, improvement, control techniques (bunds and terrace), frequently orbit around the word "soil". Whereas restoration, rehabilitation, resources, certification and tenure are the most frequently orbiting words around "land". Practices, products, sustainability, and development orbit around "agriculture", and size, system, and income orbit around "farm" (Fig. 3a). When focusing on gender and social inclusion keywords, we found that livelihood, gender, labor, control, and women are keywords that connect the whole constellation of inclusive-related words (Fig. 3b).

### 3.2. Gender-related keyword trend analysis

We found a positive trend over time in the use of social inclusion related terms within the restoration literature (Fig. 4). Before
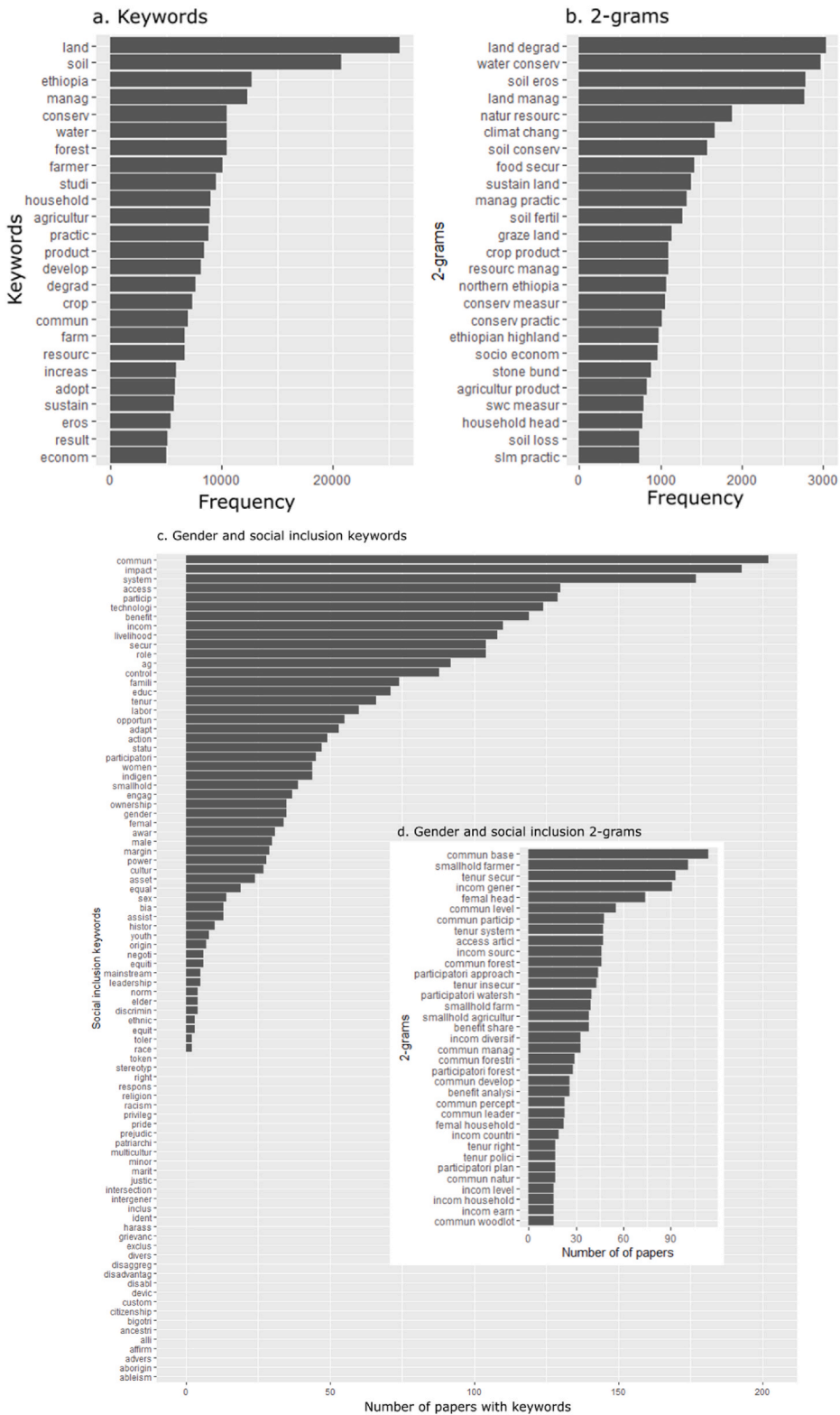
**Fig. 2.** Overall frequency of use of (a) keywords, and (b) 2-g; and gender and social inclusion (c) keywords, and (d) 2-g in RDL literature in Ethiopia.

2005, very few land restoration research articles use social inclusion terms such as asset, communities, culture, security, role, system, equity, female, rights, and sex. However, after 2005, the use of these words grows exponentially, particularly for words such as access, community, income and participatory. The use of other keywords that are more related to gender and social inclusion, such as equity,



**Fig. 3.** Network analysis of (a) keywords (number of articles >20) and (b) gender and social inclusion-related words (number of articles> =5) based on RDL literature relating to Ethiopia.

inclusiveness, and negotiation, however, have been growing at a slower rate (Fig. 4). An increasing number of articles on land restoration in relation to social inclusion is a good sign in general as it aims to address the call for land restoration that provides livelihood benefits to people and community, not just biophysical benefits.

### 3.3. Gender inclusion grade using document classification

The gender experts classified the 312 publications into three gender classes: gender blind (58 percent of the articles), gender sensitive (38 percent) and gender transformative (4 percent). Relatively large percentage (38%) of gender sensitive studies can be misleading as we consider articles that have gender disaggregated data, discussed gender and social inclusion terms/concepts versus land restoration, or provided interpretation or inference of it in general. In a strict sense, while gender sensitive studies seem to be large in number, gender transformative research in which the advancements of restoration research in relation to gender inclusion needs to achieve is rare. Our efforts to predict an article's gender classification resulted in an overall model predictive power of 69 percent. This indicates that combining text mining and multinomial regression models is a promising strategy for enabling the automatization of documents or article classification into gender grades, although further research is needed to improve the predictive power of this combined method. Interestingly, the model was able to accurately classify articles using or applying a gender-transformative approach (accuracy = 82 percent) (Fig. 5). However, gender-blind and gender-sensitive articles were properly classified approximately half of the time (accuracy = 53 percent and 58 percent, respectively). In general, this workflow can be extended to automate: 1) reviewing land restoration and gender insights at regional and global scale, 2) monitoring gender and social inclusion in relation to land restoration related studies, and 3) developing framework that can facilitate the inclusion of gender and social inclusion perspective
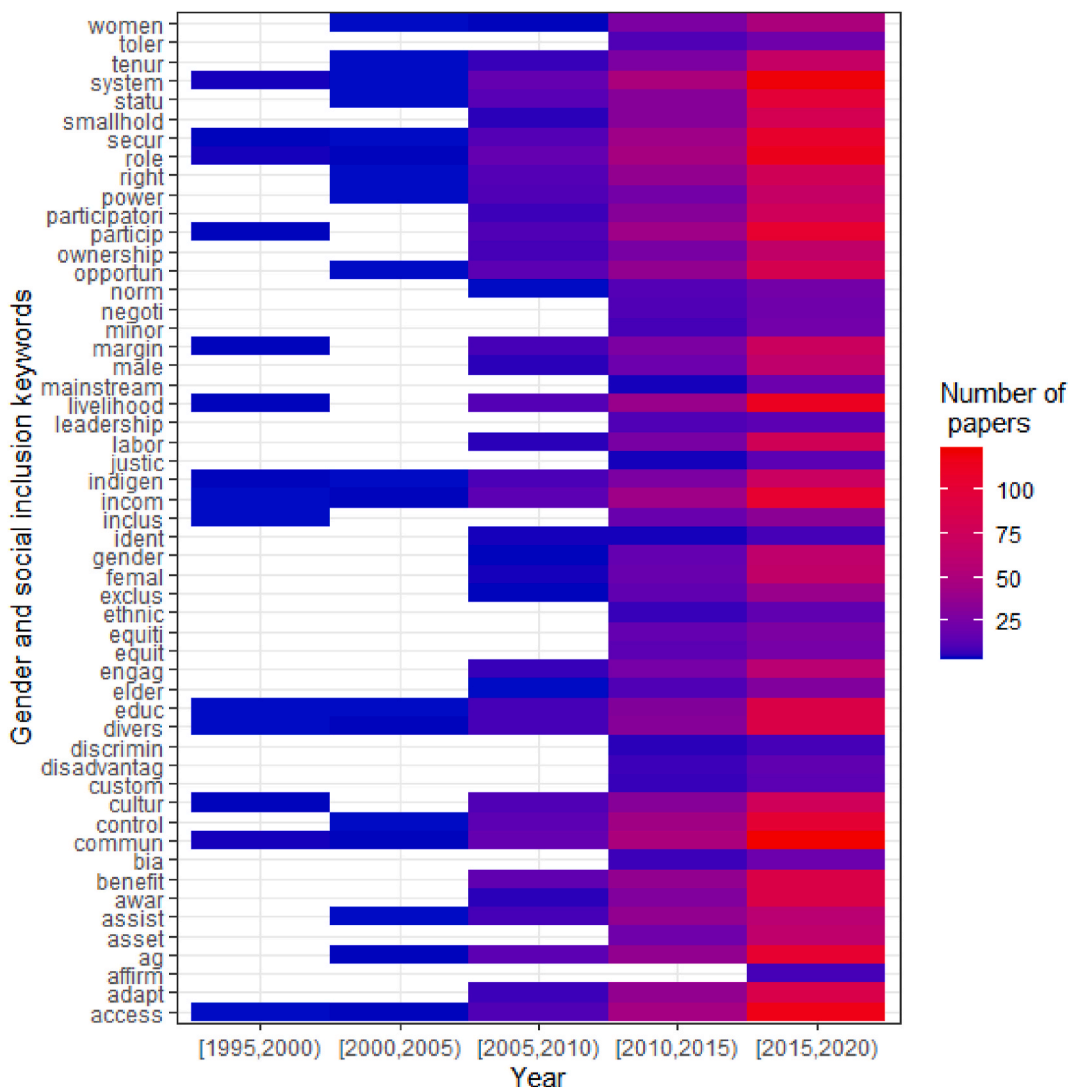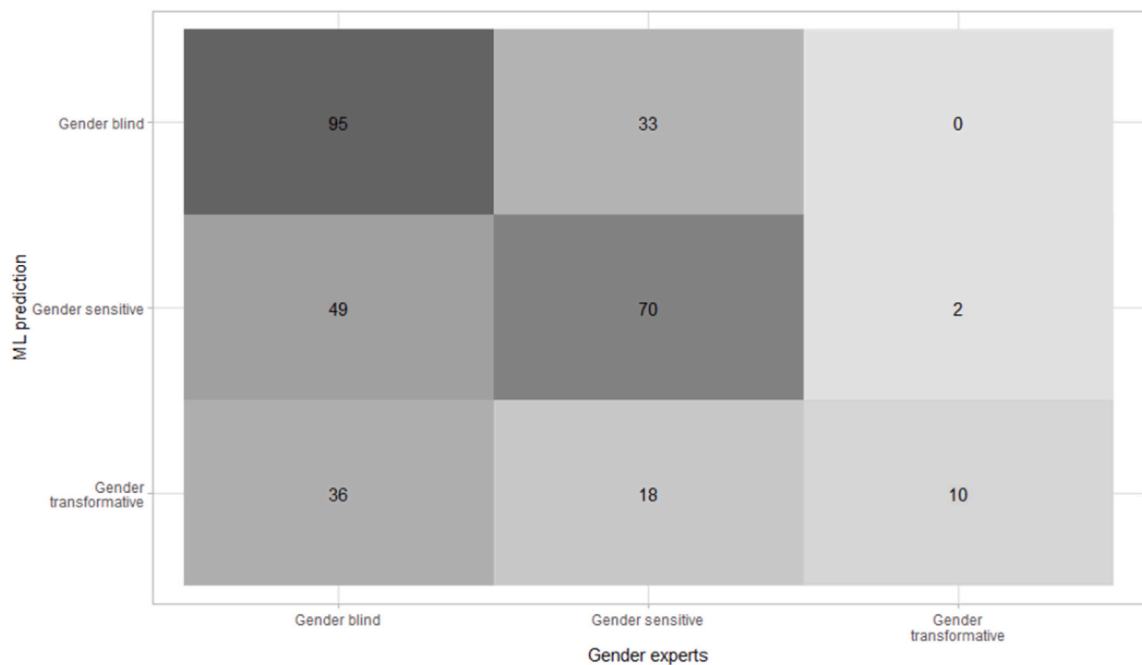


**Fig. 4.** Temporal trend (1995–2020) in the use of gender and social inclusion keywords in land restoration research in Ethiopia.

**Fig. 5.** Confusion matrix showing the number of articles classified into three gender classes defined by experts in x-axis and predicted gender class by ML (machine learning) text mining algorithm in the y-axis. Gender blind n: 181 articles; gender sensitive n: 121 articles; and gender transformative n: 12 articles.

when doing land restoration studies in future.

*3.4. Gender topics discussed using topic modeling*

The topic modeling analysis (Table 1) identified seven broad thematic areas — 1) Land management; 2) Forest/church restoration; 3) Agricultural systems; 4) Household technology adoption; 5) Community-based interventions; 6) Soil conservation; and 7) Agricultural products – often discussed in the articles classified as gender sensitive or transformative. Under each topic, a different set of words, and gender and social inclusion keywords are frequently discussed (Table 1). As expected, the most-discussed topic (e.g., highest prevalence) is around land management, whereas soil conservation and topics related to agriculture were the least mentioned. The highest probability is obtained for land management thematic area followed by church forest. Thematic area 1 is related to land management, tenure and its role and impact on the livelihood of smallholder farmers. This can be expected as large body of studies are conducted on land restoration and management in Ethiopia. Thematic area 2 is related to the contribution of church and forest conservation and afforestation. Many studies provided the contribution of the church to forest conservation in Ethiopia (Sahle et al., 2021; Wassie et al., 2009; Aerts et al., 2016). Forests, in the vicinity of orthodox church, are part of the religious spaces and considered as sacred. Thousands of churches in Ethiopia have been widely acknowledged for their ecological value. Wassie et al. (2009) estimated availability of 35,000 orthodox church communities whereas Aerts et al. (2016), using remote sensing, estimated as many as 19,400 church forests in the Ethiopian highlands, with a total area of 39,000–57,000ha. However, the latter two thematic areas seem to discuss key aspects for inclusion, for example, "agricultural products" was the only topic that listed labor in the top-ten gender and social-inclusion keywords, whereas "soil conservation" seemed to focus on the household level. Across the topics, "household technology adoption" seemed the most intersectional topic, as it discussed a wide range of ideas including age; household head; education; female; participation; family; and gender. Overall, the topic coherence matrices shows that low topics' quality and limited interpretability. This could be due to the articles included in this analysis are those contains both the biophysical and social and gender issues together, which could affect topic quality. The highest coherence matric value is obtained for soil conservation topic, with only "household" social inclusion keywords (Table 1). When the topic has more gender and social inclusion keywords, the coherence matrices decline likely due to the limited probability of co-occurrence of biophysical and gender inclusion keywords.

**4. Conclusions and implications**

This study has aimed to understand the extent of gender and social inclusion in land restoration studies in Ethiopia, through a review of 314 peer-review journal articles published between 1985 and 2020. The approach we applied in this research facilitated our exploration and analysis of the content of restoration-related research in Ethiopia through a gender and social inclusion lens. There is a large body of evidence attesting the importance of gender and social inclusion dimensions in supporting positive intervention

**Table 1**

There are seven main topics mentioned or discussed in the land restoration studies literature in Ethiopia, which have been classified as gender sensitive or gender transformative. Topics are organized by the probability of occurring (prevalence) and the quality assessed as the level or related words (coherence). A traffic color scheme is used to classify prevalence and coherence values as high (green), medium (yellow) and low (red) based on the percentile distribution.

| Topic | Top 10 Keywords | Gender and social inclusion keywords | Prevalence (probability) | coherence (quality) |
|---|---|---|---|---|
| 1. Land management | land; management; agriculture; resource; secure; sustain; food; develop; tenure; research | tenure; livelihood; smallholder; rights; policy; rural; impact; role | 22 | 0.06 |
| 2. Forest/church restoration | forest; land; tree; manage; exclosure/enclosure; livelihood; plant; church; product; access | livelihood; local; villages; participation; benefit | 17 | 0.06 |
| 3. Agricultural systems | Agriculture; system; climate; change; livelihood; develop; food; work; conserve | promote; social; sustain; traditional; community; credit | 16 | 0.07 |
| 4. Household technology adoption | household; adopt; farm; variable; access; size; factor; education; practices; extension | age; household head; education; female; participation; family; gender; | 16 | 0.08 |
| 5. Community-based interventions | community; farmer; participation; implement; practice; govern; member; technological; management | Community; women; benefit; location; decision; role; impact | 14 | 0.05 |
| 6. Soil conservation | soil; conservation; fertility; water; farmer; household; practice; adopt; erosion; bund | household; | 9 | 0.14 |
| 7. Agricultural products | land; income; crop; farm; irrigation; cash; product; market; invest; diversification | access; livelihood; labor; | 7 | 0.08 |

engagement and outcomes (Leisher et al., 2016). Yet, our findings show that the majority of restoration research in Ethiopia remains blind to gender and social inclusion. Despite the exponential growth of gender and social inclusion keywords after 2005, research that examines the underlying causes of inequality (gender transformative) in restoration actions remains extremely rare.

Costs and benefits for different groups of women and men related to land restoration is shaped by considering multiple, contextually-specific, socio-economic, and cultural factors, which can, to a considerable extent, determine the success of restoration initiatives. Thus, full and equal participation of marginalized groups – particularly minorities or often-ignored agents of change, such as women – is needed in the planning and design of restoration initiatives to ensure better outcomes and shared benefits.

Ethiopia is one of the countries globally with the largest, if not the most extensive, experience and efforts in restoring ecosystems, as reflected in the large body of peer-reviewed publications on this very topic. Hence, our findings can serve future global, regional, and national efforts in more adequately shaping land restoration. Making these efforts inclusive and intersectional will require the collective action of all the stakeholders involved in restoration – from the donors specifying in their calls for funding that project design includes, international organizations, governments, research organizations, and communities, down to individual smallholders – to ensure that intervention designs have strong gender/social inclusion foundations. In particular, closer monitoring and measurement of the impact of restoration efforts is needed across all stakeholders and rightsholder groups, as well as extending these efforts beyond the biophysical aspects into social relationships. It is unclear to date how and to what extent restoration efforts affect the most vulnerable or least-represented stakeholders in terms of poverty, power imbalances, and the benefits deriving from restoration. These are important gaps to be filled within restoration initiatives that will also help to achieve multiple Sustainable Development Goals by 2030.

Scientists and researchers working on land management need to integrate inclusion aspects and gender analysis into their studies in order to inform and guide land management interventions, programs, and policies. Policies and interventions that consider and address the existing gender gaps in power, ownership, access to, and benefits from land resources are needed to improve adoption and scaling out of restoring degraded landscapes in Ethiopia and beyond. Our findings highlight the importance of developing enabling factors and policies to guide gender-transformative projects in land restoration and encourage studies highlighting gender inclusiveness in land restoration practices and impact assessments of such practices.

The limitation of this study is that the sample size is relatively small for machine learning algorithms. In addition, the keywords that are used as gender and social inclusion can be improved by developing standardized conceptual inferences such as gender and social inclusion dictionary. Dictionary-based text mining analysis is based on the community (e.g. gender community) priori definition of each keyword and its relations with large conceptualization. These are common in sentiment analysis and well-advanced dictionaries in the medical field. With similar efforts to this study, gender and social inclusion dictionaries can be developed and used to facilitate automatic text mining analytics on land and land management in relation to gender and social inclusion. The low performance of topic modelling can be improved by complimenting the machine learning process with human experts in more efficient way of interacting and go through each important keywords in each topic and article.

**Declaration of competing interest**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

No data was used for the research described in the article.

## Acknowledgements

## Appendix 1

*Gender and social inclusion keywords used for analysis*

"ableism", "aboriginal", "accessibility", "accessible", "adaptive technologies", "adverse impact", "affirmative action", "age", "ally", "racism", "assistive", "bias", "bigotry", "sex", "male", "female", "disability", "culture", "custom", "diversity", "race", "ethnicity", "religion", "education", "discrimination", "diverse", "elder", "opportunity", "equity", "exclusion", "privilege", "gender", "identity", "harassment", "historical disadvantage", "inclusion", "indigenous", "intergenerational", "power", "prejudice", "pride", "minorities", "ancestry", "citizenship", "ethnic origin", "family status", "marital status", "stereotype", "systemic", "access", "intersectionality", "marginalization", "multiculturalism", "patriarchy", "justice", "participation", "tokenism", "tolerance", "leadership", "control", "benefits", "role", "livelihood", "women", "community", "Youth", "Equal", "tenure", "smallholder", "participatory", "ownership", "asset", "right", "awareness", "norms", "negotiation", "disaggregate", "gender-responsive", "affirmative","mainstream", "engagement", "income", "grievance", "security", "equitable", "inclusive", "labor".

## References

Abera, W., Tamene, L., Tibebe, D., Adimassu, Z., Kassa, H., Hailu, H., Mekonnen, K., Desta, G., Sommer, R., Verchot, L., 2020. Characterizing and evaluating the impacts of national land restoration initiatives on ecosystem services in Ethiopia. Land Degrad. Dev. 31 (1), 37–52.

Adams, C., Rodrigues, S.T., Calmon, M., Kumar, C., 2016. Impacts of large-scale forest restoration on socioeconomic status and local livelihoods: what we know and do not know. Biotropica 48 (6), 731–744.

Aerts, R., Van Overtveld, K., November, E., Wassie, A., Abiyu, A., Demissew, S., Daye, D.D., Giday, K., Haile, M., TewoldeBerhan, S., Teketay, D., 2016. Conservation of the Ethiopian church forests: threats, opportunities and implications for their management. Sci. Total Environ. 551, 404–414.

Agarwal, B., 2018. Gender equality, food security and the sustainable development goals. Curr. Opin. Environ. Sustain. 34, 26–32.

Araya, T., Nyssen, J., Govaerts, B., Deckers, J., Cornelis, W.M., 2015. Impacts of conservation agriculture-based farming systems on optimizing seasonal rainfall partitioning and productivity on vertisols in the Ethiopian drylands. Soil Tillage Res. 148, 1–13.

Araya, T., Nyssen, J., Govaerts, B., Deckers, J., Sommer, R., Bauer, H., Gebrehiwot, K., Cornelis, W.M., 2016. Seven years of resource-conserving agriculture effect on soil quality and crop productivity in the Ethiopian drylands. Soil Tillage Res. 163, 99–109.

Bai, Z., Dent, D., Wu, Y., de Jong, R., 2013. Land Degradation and Ecosystem Services. Ecosystem Services and Carbon Sequestration in the Biosphere. Springer, Dordrecht, p. 357–381.

Balehegn, M., Haile, M., Fu, C., Liang, W., 2019. Ecosystem-based adaptation in Tigray, Northern Ethiopia: a systematic review of interventions, impacts, and challenges. In: Handbook of Climate Change Resilience. Springer, Cham, p. 1–45.

Bezu, S., Holden, S., 2014. Demand for second-stage land certification in Ethiopia: evidence from household panel data. Land Use Pol. 41, 193–205.

Blei, D.M., Ng, A.Y., Jordan, M.I., 2003. Latent dirichlet allocation. J. Mach. Learn. Res. 3, 993–1022.

Bouchet-Valat, M., Bouchet-Valat, M.M., 2020. Package 'SnowballC'. R Package version 0.7. 0.

Broeckhoven, N., Cliquet, A., 2015. Gender and ecological restoration: time to connect the dots. Restor. Ecol. 23 (6), 729–736.

Catacutan, D.C., Villamor, G.B., 2016. Gender roles and land use preferences—implications to landscape restoration in Southeast Asia. In: Land Restoration. Academic Press, pp. 431–440.

Cohen, K.B., Hunter, L., 2008. Getting started in text mining. PLoS Comput. Biol. 4 (1), e20.

Coleman, E.A., Mwangi, E., 2013. Women's participation in forest management: A cross-country analysis. Global Environ. Change 23 (1), 193–205.

Collantes, V., Kloos, K., Henry, P., Mboya, A., Mor, T., Metternicht, G., 2018. Moving towards a twin-agenda: gender equality and land degradation neutrality. Environ. Sci. Pol. 89, 247–253.

Crossland, M., Valencia, A.M.P., Pagella, T., Magaju, C., Kiura, E., Winowiecki, L., Sinclair, F., 2021. Onto the farm, into the home: how intrahousehold gender dynamics shape land restoration in Eastern Kenya. Ecol. Restor. 39 (1–2), 90–107.

de Siqueira, L.P., Tedesco, A.M., Meli, P., Diederichsen, A., Brancalion, P.H., 2021. Gender inclusion in ecological restoration. Restor. Ecol. 29 (7), e13497.

Deininger, K., Ali, D.A., Alemu, T., 2008. Impacts of land certification on tenure security, investment, and land markets: evidence from Ethiopia. In: World Bank Policy Research Working Paper, p. 4764.

Deininger, K., Ali, D.A., Alemu, T., 2011. Impacts of land certification on tenure security, investment, and land market participation: evidence from Ethiopia. Land Econ. 87 (2), 312–334.

Desta, G., Tamene, L., Abera, W., Amede, T., Whitbread, A., 2021. Effects of land management practices and land cover types on soil loss and crop productivity in Ethiopia: A review. Int. Soil Water Conserv. Res. 9 (4), 544–554.

Doss, C., Kovarik, C., Peterman, A., Quisumbing, A., Van Den Bold, M., 2015. Gender inequalities in ownership and control of land in Africa: myth and reality. Agric. Econ. 46 (3), 403–434.

Elias, M., Joshi, D., Meinzen-Dick, R., 2021. Restoration for whom, by whom? A feminist political ecology of restoration. Ecol. Restor. 39 (1–2), 3–15.

Elmhirst, R., Siscawati, M., Basnett, B.S., Ekowati, D., 2017. Gender and generation in engagements with oil palm in East Kalimantan, Indonesia: insights from feminist political ecology. J. Peasant Stud. 44 (6), 1135–1157.

Forsythe, L., Morton, J., Nelson, V., Quan, J., Martin, A., Hartog, M., 2015. Strengthening Dryland Women's Land Rights: Local Contexts, Global Change.

Friedman, J., Hastie, T., Tibshirani, R., Narasimhan, B., 2021. Package 'glmnet'. CRAN R Repository.

Girma, H., Giovarelli, R., 2019. The gender implications of joint land titling in Ethiopia. Gates Open Res. 3 (917), 917.

Grün, B., Hornik, K., 2011. Topicmodels: an R package for fitting topic models. J. Stat. Software 40 (13), 1–30. https://doi.org/10.18637/jss.v040.i13.

Havrlant, L., Kreinovich, V., 2017. A simple probabilistic explanation of term frequency-inverse document frequency (tf-idf) heuristic (and variations motivated by this explanation). Int. J. Gen. Syst. 46 (1), 27–36.

IPBES, 2018. The IPBES assessment report on land degradation and restoration. In: Montanarella, L., Scholes, R., Brainich, A. (Eds.), Secretariat of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. Bonn, Germany.

Ooms, Jeroen, 2020. Pdftools: Text Extraction, Rendering and Converting of PDF Documents. R package version 2.3.1.

Komatsu, H., Ambel, A.A., Koolwal, G., Yonis, M.B., 2022. Gender norms, landholdership, and rural land use fee and agricultural income tax in Ethiopia. Land Use Pol. 121, 106305.

Kosec, K., Ghebru, H., Holtemeyer, B., Mueller, V., Schmidt, E., 2018. The effect of land access on youth employment and migration decisions: evidence from rural Ethiopia. Am. J. Agric. Econ. 100 (3), 931–954.

Kumar, N., Quisumbing, A.R., 2010. Policy Reform towards Gender Equality in Ethiopia: Little by Little the Egg Begins to Walk. International Food Policy Research Institute, Wasgington DC.

Kumasi, T.C., Asenso-Okyere, K., 2011. Responding to land degradation in the highlands of Tigray, Northern Ethiopia. Int. Food Pol. Res. Inst. 1142, 44.

Lamarque, P., Lavorel, S., Mouchet, M., Quétier, F., 2014. Plant trait-based models identify direct and indirect effects of climate change on bundles of grassland ecosystem services. Proc. Natl. Acad. Sci. USA 111 (38), 13751–13756.

Laughlin, D.C., 2014. Applying trait-based models to achieve functional targets for theory-driven ecological restoration. Ecol. Lett. 17 (7), 771–784.

Leisher, C., Temsah, G., Booker, F., Day, M., Samberg, L., Prosnitz, D., Agarwal, B., Matthews, E., Roe, D., Russell, D., Sunderland, T., 2016. Does the gender composition of forest and fishery management groups affect resource governance and conservation outcomes? A systematic map. Environ. Evid. 5 (1), 1–10.

Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., et al., 2018. Applying LDA topic modeling in communication research: toward a valid and reliable methodology. Commun. Methods Meas. 12 (2–3), 93–118.

Mbile, P.N., Atangana, A., Mbenda, R., 2019. Women and landscape restoration: a preliminary assessment of women-led restoration activities in Cameroon. Environ. Dev. Sustain. 21, 2891–2911.

McLain, R., Lawry, S., Guariguata, M.R., Reed, J., 2021. Toward a tenure-responsive approach to forest landscape restoration: a proposed tenure diagnostic for assessing restoration opportunities. Land Use Pol. 104, 103748.

Mekuria, W., Veldkamp, E., Tilahun, M., Olschewski, R., 2011. Economic valuation of land restoration: the case of exclosures established on communal grazing lands in Tigray, Ethiopia. Land Degrad. Dev. 22 (3), 334–344.

Montoya, J.M., Raffaelli, D., 2010. Climate Change, Biotic Interactions and Ecosystem Services.

Mooney, H., Larigauderie, A., Cesario, M., Elmquist, T., Hoegh-Guldberg, O., Lavorel, S., et al., 2009. Biodiversity, climate change, and ecosystem services. Curr. Opin. Environ. Sustain. 1 (1), 46–54.

Namubiru-Mwaura, E., 2014. Land Tenure and Gender: Approaches and Challenges for Strengthening Rural Women's Land Rights. © World Bank, Washington, DC. https://openknowledge.worldbank.org/handle/10986/21033. License: CC BY 3.0IGO.

Nikolenko, S.I., Koltcov, S., Koltsova, O., 2017. Topic modelling for qualitative studies. J. Inf. Sci. 43 (1), 88–102.

Orwin, K.H., Stevenson, B.A., Smaill, S.J., Kirschbaum, M.U., Dickie, I.A., Clothier, B.E., et al., 2015. Effects of climate change on the delivery of soil-mediated ecosystem services within the primary sector in temperate ecosystems: a review and New Zealand case study. Global Change Biol. 21 (8), 2844–2860.

Pederson, A., Greaves, L., Poole, N., 2014. Gender-transformative health promotion for women: a framework for action. Health Promot. Int. 30 (1), 140–150.

Ramos, J., 2003. Using tf-idf to determine word relevance in document queries. In: Proceedings of the First Instructional Conference on Machine Learning, Vol. 242, pp. 29–48. No. 1.

Robinson, D., Silge, J., 2020. Tidytext: text mining using dplyr, ggplot2, and other tidy tools. R package version 0 2, 4.

Röder, M., Both, A., Hinneburg, A., 2015. Exploring the space of topic coherence measures. In: Proceedings of the Eighth ACM International Conference on Web Search and Data Mining, pp. 399–408.

Rzhetsky, A., Seringhaus, M., Gerstein, M.B., 2009. Getting started in text mining: part two. PLoS Comput. Biol. 5 (7), e1000411.

Sahle, M., Saito, O., Reynolds, T.W., 2021. Nature's contributions to people from church forests in a fragmented tropical landscape in southern Ethiopia. Global Ecol. Conserv. 28, e01671.

Sannigrahi, S., Zhang, Q., Joshi, P.K., Sutton, P.C., Keesstra, S., Roy, P.S., et al., 2020. Examining effects of climate change and land use dynamic on biophysical and economic values of ecosystem services of a natural reserve region. J. Clean. Prod. 257, 120424.

Sijapati Basnett, B., Elias, M., Ihalainen, M., Paez Valencia, A.M., 2017. Gender matters in forest landscape restoration: a framework for design and evaluation. CIFOR.

Silge, J., Robinson, D., 2016. tidytext: text mining and analysis using tidy data principles in R. J. Open Source Softw. 1 (3), 37.

Sutton, P.C., Anderson, S.J., Costanza, R., Kubiszewski, I., 2016. The ecological economics of land degradation: impacts on ecosystem service values. Ecol. Econ. 129, 182–192.

Temmerman, S., Meire, P., Bouma, T.J., Herman, P.M., Ysebaert, T., De Vriend, H.J., 2013. Ecosystem-based coastal defence in the face of global change. Nature 504 (7478), 79–83.

Tokunaga, T., Makoto, I., 1994. Text categorization based on weighted inverse document frequency. In: Special Interest Groups and Information Process Society of Japan. SIG-IPSJ.

Wassie, A., Sterck, F.J., Teketay, D., Bongers, F., 2009. Effects of livestock exclusion on tree regeneration in church forests of Ethiopia. Forest Ecol. Manag. 257 (3), 765–772.

Webb, N.P., Marshall, N.A., Stringer, L.C., Reed, M.S., Chappell, A., Herrick, J.E., 2017. Land degradation and climate change: building climate resilience in agriculture. Front. Ecol. Environ. 15 (8), 450–459.

Wheeler, T., Von Braun, J., 2013. Climate change impacts on global food security. Science 341 (6145), 508–513.