Year: 2021

Recycled text and risk communication in natural gas pipeline environmental impact assessments

Hileman, Jacob D ; Angst, Mario ; Scott, Tyler A ; Sundström, Emma

Abstract: Under the U.S. National Environmental Policy Act (NEPA), energy infrastructure projects that are permitted by federal agencies require preparation and publication of an environmental impact assessment. However, fifty years after the passage of NEPA, agencies' compliance behaviors, and how these behaviors might shape the risks associated with energy infrastructure, remain largely unexplored. Here, we consider how assessment documents from forty-six of the largest U.S. natural gas pipeline mega-projects address landslide risks. Using a series of text mining and content analysis methods, we evaluate the prevalence of recycled text across assessments. We find that text similarity does not correspond closely to reported risk levels – in many cases, common verbiage is used and only project-specific details (e.g., locations, numeric figures) are substituted. While such approaches likely expedite preparation of assessments and facilitate knowledge transfer between projects, we argue that common text potentially hinders clear communication of differential risks to decision-makers and the public, who may lack the technical expertise to contextualize the magnitude and severity of reported figures. In light of ongoing policy efforts to streamline lengthy and costly energy infrastructure permitting processes under NEPA, it is vital that such efforts do not undermine the risk communication requirements of the review process.

DOI: https://doi.org/10.1016/j.enpol.2021.112379

Originally published at:
Hileman, Jacob D; Angst, Mario; Scott, Tyler A; Sundström, Emma (2021). Recycled text and risk communication in natural gas pipeline environmental impact assessments. Energy Policy, 156:112379.
DOI: https://doi.org/10.1016/j.enpol.2021.112379

Energy Policy 156 (2021) 112379

Contents lists available at ScienceDirect

Energy Policy

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

Research Article

Recycled text and risk communication in natural gas pipeline environmental impact assessments

Jacob D. Hileman [a,b,*], Mario Angst [c,d], Tyler A. Scott [e], Emma Sundström [a]

[a] Stockholm Resilience Centre, Stockholm University, Kräftriket 2B, 10691, Stockholm, Sweden
[b] Centre of Natural Hazards and Disaster Science, Uppsala University, P.O. Box 256, 75105, Uppsala, Sweden
[c] Swiss Federal Institute of Aquatic Science and Technology, Überlandstrasse 133, 8600, Dübendorf, Switzerland
[d] Digital Society Initiative, University of Zurich, Rämistrasse 69, 8001, Zürich, Switzerland
[e] Department of Environmental Science and Policy, University of California-Davis, One Shields Avenue, Davis, CA, 95616, USA

ARTICLE INFO

Keywords:
Environmental impact assessment
Federal Energy Regulatory Commission
Landslide risk
Natural gas pipeline
Risk communication
Text similarity

ABSTRACT

Under the U.S. National Environmental Policy Act (NEPA), energy infrastructure projects that are permitted by federal agencies require preparation and publication of an environmental impact assessment. However, fifty years after the passage of NEPA, agencies' compliance behaviors, and how these behaviors might shape the risks associated with energy infrastructure, remain largely unexplored. Here, we consider how assessment documents from forty-six of the largest U.S. natural gas pipeline mega-projects address landslide risks. Using a series of text mining and content analysis methods, we evaluate the prevalence of recycled text across assessments. We find that text similarity does not correspond closely to reported risk levels – in many cases, common verbiage is used and only project-specific details (e.g., locations, numeric figures) are substituted. While such approaches likely expedite preparation of assessments and facilitate knowledge transfer between projects, we argue that common text potentially hinders clear communication of differential risks to decision-makers and the public, who may lack the technical expertise to contextualize the magnitude and severity of reported figures. In light of ongoing policy efforts to streamline lengthy and costly energy infrastructure permitting processes under NEPA, it is vital that such efforts do not undermine the risk communication requirements of the review process.

1. Introduction

Encouraged by favorable energy policies and advances in extraction technology, the U.S. natural gas industry has grown rapidly over the past two decades (Wang et al., 2014). One of the most visible, and controversial, signs of this expansion is the construction of interstate gas transmission pipelines. These large pipelines require certification from the Federal Energy Regulatory Commission (FERC), which has the broad authority to determine the existence of public need, siting of the pipeline and related infrastructure, and approve or reject proposed projects (FERC, 1999). One condition for certifying a pipeline is that FERC must undertake a comprehensive environmental impact assessment (EIA) process, as mandated by the U.S. National Environmental Policy Act (NEPA) (EPA, 2013a). This process is intended to weigh the need for new pipelines against the social and environmental risks associated with each project. While pipelines ultimately require a variety of different state and federal permits, the EIA is the centerpiece of the massive, drawn-out regulatory process for reviewing and certifying pipeline projects.

EIA rules such as NEPA exist in many countries worldwide (Glasson and Therivel, 2013; Larsen et al., 2018), but the effectiveness of these procedural regulations remains uncertain (Emerson and Baldwin, 2019). One particular tension for pipelines, and large linear infrastructure projects more generally, is how well these decision support tools can account for complex interdependencies and uncertainties associated with system design, siting, and operations across the extensive geographic scope of the projects (Gregory et al., 2020). The social and environmental risks associated with major gas pipelines are well-documented (Bergquist et al., 2020; Entrekin et al., 2011; Jackson et al., 2014; Vengosh et al., 2014; Vidic et al., 2013; Witter et al., 2013). However, the lengthy, comprehensive environmental review documents within which risks are currently presented to decision-makers and the public are a challenging medium for presenting the dynamic and highly

---

* Corresponding author. Department of Government, Uppsala University, PO Box 514, 751 20, Uppsala, Sweden.
E-mail addresses: jacob.hileman@statsvet.uu.se (J.D. Hileman), mario.angst@eawag.ch (M. Angst), tascott@ucdavis.edu (T.A. Scott), emma.sundstrom@su.se (E. Sundström).

https://doi.org/10.1016/j.enpol.2021.112379
Received 14 July 2020; Received in revised form 10 May 2021; Accepted 14 May 2021

Abbreviations

Federal Energy Regulatory Commission (FERC)
National Environmental Policy Act (NEPA)
Environmental Impact Assessment (EIA)
Environmental Impact Statement (EIS)
Environmental Assessment (EA).

varied risks associated with different pipeline projects. Given time demands (Doyle, 2017), resource limitations, and isomorphic pressures – e.g., to streamline (Lyles, 2017), standardize (CEQ, 2019), and avert litigation (Ruple and Race, 2020) – agencies are not strongly incentivized to deviate from existing procedures, compare results to other projects, or devote extra effort to EIA production. Thus, even when EIAs published by FERC clearly present best available risk metrics, the manner in which these risks are presented – largely within recycled boilerplate text – hinders risk communication.

Drawing on forty-six onshore natural gas pipeline “mega-projects” (Flyvbjerg, 2014) certified by FERC from 1997 to 2019, we address how differences in the level of landslide risk among pipelines are reflected, or not reflected, in the form of text similarity across their respective EIAs. Landslides are a growing concern among regulators and the public; a number of recent high-profile pipeline explosions have been attributed to landslide events (PHMSA, 2020; Soraghan, 2019). In this study, we focus on a particular section of the assessments – the Geologic Hazards section – where landslides and other geophysical risks that pipeline projects may encounter are characterized and quantified, and risk minimization and mitigation strategies are presented. Clear articulation of these risks within NEPA documentation is key for informing decision-makers and the public about project costs and benefits, and the relative merits of different projects. However, pressures associated with the EIA process potentially hinder the ability of these decision-making tools to convey risks adequately, and to distinguish between higher- and lower-risk projects.

Using a series of text mining and content analysis methods, we evaluate the prevalence of recycled text across natural gas pipeline EIAs. We find that text similarity does not correspond closely to reported risks, and in many cases common verbiage is used and only project-specific details (e.g., locations, numeric figures) are substituted. While such approaches likely expedite preparation of assessments and facilitate knowledge transfer between projects, we argue that common text potentially hinders clear communication of differential risks to decision-makers and the public, who may lack the technical expertise to contextualize the magnitude and severity of reported figures. In light of ongoing policy efforts to streamline lengthy and costly energy infrastructure permitting processes under NEPA, it is vital that such efforts do not undermine the risk communication requirements of the review process.

The following sections provide background on natural gas pipeline infrastructure and the NEPA process. We then explain how we frame risk as it relates to landslide hazards and pipeline projects, and articulate the relationship between this specific instantiation of risk and text similarity in project EIAs. Next, we describe the methods we used to identify the sample of U.S. natural gas pipeline mega-projects, quantify landslide risks, and analyze text similarity among the projects. Following the presentation of the results, we discuss the main findings and illustrate their significance using passages of text from the EIAs. Lastly, we highlight the policy implications of the findings as they relate to FERC’s role in permitting and certifying gas pipelines and other energy infrastructure, such as hydropower facilities and bulk transmission lines, and suggest several promising avenues for future research on text similarity in NEPA documentation for U.S. energy projects.

2. Environmental impact assessments and risk communication in natural gas pipeline projects

2.1. Natural gas pipelines and landslide risks

In 2018, the U.S. natural gas pipeline system included over 3 million miles of mainline pipes, and transported around 75 trillion cubic feet of gas per year (EIA, 2020a). Gas transmission pipelines are a core component of this overall system, and are used to transport large volumes of methane over long distances – often hundreds of miles – from production well sites and processing plants to regional and global distribution hubs (EIA, 2020a). These pipelines consist of a large-diameter mainline pipe, along with smaller lateral pipes that move gas to and from the mainline. While we focus solely on the pipeline itself in this study, pipeline projects may include a range of related infrastructure, such as compressor stations, access roads, staging areas, and liquefied natural gas facilities.

Gas pipelines must contend with a number of risks during construction and operation: they are not confined within a secure industrial site (Jo and Crowl, 2008), they transport hazardous and explosive material at high pressure (Jo and Ahn, 2002), and they are vulnerable to seismic events and other natural hazards (Lanzano et al., 2013). Landslides, in particular, are a growing pipeline safety concern (PHMSA, 2020). The overall level of landslide risk associated with U.S. natural gas pipelines has increased considerably since the late 1990s, in large part due to the expansion of pipeline infrastructure in the landslide-prone Appalachian Mountains (EIA, 2017; USGS, 2016). This is not an idle concern – multiple recent gas pipeline explosions have been linked to landslide events (PHMSA, 2018, 2019; Soraghan, 2019), including the Leach XPress Pipeline in this dataset. Within six months of coming online in 2018, a section of the 36-inch diameter Leach XPress mainline in West Virginia ruptured, and subsequently exploded, due to a landslide-related event (PHMSA, 2019). In response to this wave of high-profile pipeline explosions, the Pipeline and Hazardous Materials Safety Administration issued an advisory bulletin to owners and operators of pipelines titled, “Pipeline Safety: Potential for Damage to Pipeline Facilities Caused by Earth Movement and Other Geological Hazards” (PHMSA, 2019).

Still under construction at the time of this study, the Mountain Valley Pipeline – which crosses more miles of high landslide risk terrain than any other project in this dataset – has already had to contend with numerous landslide events. In 2018, a slope failed along the construction right-of-way during a heavy rain and resulted in the temporary closure of a public road in Virginia (DEQ, 2018). In a letter filed with FERC in 2019, the pipeline company requested emergency authorization to address a landslide in West Virginia, stating: “The progression of the slide caused additional area outside the limits of disturbance to destabilize, uprooted numerous large trees, has the potential to impact an aquatic resource, and has progressed to the point where a residence directly downslope is unsafe to be occupied.” (MVP, 2019) More recently, environmental inspectors in West Virginia identified a site where earth “slips” resulted in a situation where “the installed pipe shifted due to the movement of the slips in at least three locations.” (FERC, 2020a).

2.2. Natural gas pipeline assessment and permitting

Natural gas pipelines in the U.S. are built and operated by private companies, while the siting of interstate gas pipelines and projects involving liquefied natural gas terminals are regulated by FERC (FERC, 2018). Under NEPA, all actions taken by federal agencies are subject to environmental assessment (EPA, 2013a). Thus, while FERC itself does not build or own pipelines, in order to certify pipeline siting and operations FERC must assess a project's environmental impacts. For the most part, there are three possible levels of assessment: (1) Categorical Exclusions, where an agency or legislative act has designated a class of actions as not individually or cumulatively having a significant environmental effect, and exempted from further review (CEQ, 2020); (2) Environmental Assessments (EAs), which study whether or not a proposed action will have significant environmental effects (EPA, 2013b); and (3) Environmental Impact Statements (EISs), wherein an agency is required to take a "hard look" at a proposed action if it is determined it will have significant environmental effects (EPA, 2013a). EAs and EISs are similar in substance; both comprehensively assess projects' environmental impacts and consider different decision alternatives. However, EISs are much more extensive and detailed, take longer and cost more to complete (DOE, 2017), and are required by law to undergo a public review and comment process (EPA, 2013a).

Major gas transmission pipelines like the mega-projects studied here typically carry the potential for significant environmental impacts and most frequently fall under the EIS standard, although eight of the forty-six pipelines we examine do use EAs.[1] The prevalence of EISs is due to the fact that the elongated footprint of the pipeline corridor impacts large swaths of terrestrial and aquatic habitats (Southerland, 2004; Xiao et al., 2017) – potentially affecting threatened and endangered species – and pipelines may cross any number of public and private roads, streams and wetlands, forests and farmland, and historically and culturally significant areas. The U.S. Natural Gas Act also grants companies building pipelines the right to condemn private property through eminent domain (CRS, 2019). As such, the preparation of an EIS typically involves multiple rounds of consultation with relevant state and federal agencies, knowledge experts, affected landowners, and the general public.

A pipeline EIS is a significant bureaucratic undertaking, and while NEPA completion statistics vary depending on time windows and reporting methods used, in general an EIS process takes around two to five years (Johnson, 2016). This period involves a series of scoping, review, draft, and comment procedures, after which the lead agency – in this case, FERC – publishes a final EIS and record of decision (EPA, 2013b). Notably, EISs and EAs are a purely procedural requirement; FERC and other federal agencies involved in the consultation process are not required to reduce environmental impacts in their final decisions, but rather are required to take a comprehensive "hard look" at all prospective impacts (Austin et al., 2004) and provide an informed science-based rationale for their decision to approve or reject a proposed pipeline project (CEQ, 2018).

2.3. Text recycling and risk characterization in EIAs

Given the comprehensive scope and extended timelines of the EIA process for pipeline projects, there are good reasons to expect that FERC – and the private consulting firms hired to aid in the preparation of EIAs – will utilize reusable language and templates as a way to expedite and economize the process. Indeed, from an organizational perspective, standardized pre-written text (i.e., boilerplate) has many advantages; EIAs are typically authored by interdisciplinary teams, and pre-formatted structures provide a way to organize disparate work and information (Hays, 1983). Additionally, EIAs are often the subject of litigation (Ruple and Race, 2020), and well-vetted (and possibly already legally-reviewed) language potentially reduces erroneous wording and legal exposure (Ben-Shahar, 2007).

At the heart of this study is whether the pressures associated with the EIA process, which favors standardized language and consistent reporting of project attributes, potentially hinder the ability of these decision-making tools to convey varying levels of landslide risk across pipeline projects. To the extent that verbiage is consistent and only numeric risk measures are substituted across different documents, it is unlikely that EIAs are able to clearly and adequately characterize risks for decision-makers and the public, which is a key NEPA requirement. Reliance on recycled text may downplay heterogeneous local circumstances and make objective determination of risk more difficult, and arguably prioritizes the completion of the EIA process over the substance of the documented impacts.

While there are many different risk considerations associated with major gas pipelines, in this analysis we consider risk from an infrastructural perspective: how the geophysical setting exposes pipelines to varying levels of landslide risk (Gregory et al., 2020). EIAs for pipeline projects typically contain a Geologic Hazards section where landslides and other geophysical risks are characterized and quantified, and risk minimization and mitigation strategies are discussed. We specifically examine whether the degree of risk from landslides, measured as the number of miles of a pipeline that cross identified high landslide risk areas, corresponds to unique discussion and characterization of risk within the Geologic Hazards section of pipeline EIAs. In other words, do EIAs for pipelines with greater empirically measured landslide risks discuss these risks differently as compared to lower-risk pipelines? Or conversely, do assessments generally tend to employ common language and substitute relevant project-specific metrics regardless of the level of landslide risk?

3. Data and methods

3.1. Identifying natural gas pipeline mega-projects

The first step in the data collection process was to develop a set of criteria for identifying the sample of natural gas pipelines included in this study. We specifically focus on the largest gas pipeline "mega-projects," and therefore define the threshold for inclusion based on pipeline diameter and length. The pipelines contained in this dataset are all at least 100 miles long and have mainline pipe diameters of 24 inches or greater. While gas transmission pipelines can vary widely in diameter – on the order of 6–48 inches, depending on where they are located within the larger networked pipeline system (PST, 2015) – around 93% of the gas pipelines under FERC's jurisdiction that are over 100 miles long are also 24 inches or greater in diameter (EIA, 2020b). We further refined the 100-mile length criterion to specify that these miles of pipe must be located onshore, be part of a contiguous project, and consist of new pipeline (rather than replacements or upgrades) (Table 1). While we only consider gas pipelines within FERC's jurisdiction, over 80% of gas pipeline mega-projects fall under FERC permitting requirements (EIA, 2020b).[2]

While gas transmission pipelines vary in basic features such as length

---

[1] EAs were drafted in cases where FERC's initial assessment concluded these projects would result in "no significant impact" and therefore did not require full EISs. Determining why these eight pipeline projects did not meet the EIS standard is not central to the aims of this study. The designation may be due to any number of contextual factors, such as these projects using more existing rights-of-way or transiting across areas deemed to be of lower environmental risk. Regardless, both EAs and EISs are expected to address potential risks such as geologic hazards, so comparing across documents for both types of projects is worthwhile.

[2] These figures take into account all new pipelines, laterals, and expansion projects for which both pipeline length and diameter data were available in the Energy Information Administration's natural gas pipelines dataset.

Table 1
Criteria used for identifying U.S. natural gas pipeline mega-projects.

Pipeline Criteria		Description	Application of Criteria
Diameter		Natural gas pipelines vary in size, and it is common for the same project to utilize a range of different pipe diameters. The mainline is the largest component, while laterals often consist of smaller pipe.	To be included in this study, the mainline pipe had to be at least 24 inches in diameter.
Length[a]	*Onshore*	Natural gas pipelines are built onshore and offshore, and the same project may contain both onshore and offshore components.	To be included in this study, at least 100 miles of the pipeline had to be located onshore. The project may also have offshore components, but we do not consider these components here.
	Contiguous	Natural gas pipeline projects are not always composed of a singular connected piece of infrastructure, and may include multiple interconnected components in geographically distant areas.	To be included in this study, the project needed to include at least 100 miles of interconnected – but not necessarily consecutive – pipeline.
	New Pipeline	Natural gas pipeline projects may involve installation of new pipe, pipe replacement, and/or pipeline "loop." Loop refers to lengths of pipe that are installed adjacent to an existing pipeline segment to increase the volume of transported gas.	To be included in this study, at least 100 miles of the pipeline had to consist of new mainline and laterals. We do not consider loop or replacement pipe here.

[a] We include two pipelines in the dataset that were just under 100 miles long (see Table 2). This is in order to avoid accidentally excluding relevant pipelines, as distances given in the EIAs were often rounded and therefore a small margin of error is present in the total pipeline length estimates.

and diameter, they do not vary much in terms of structural and material characteristics (e.g., right-of-way requirements, pipe composition, etc.). Pipeline construction and design standards are spelled out by state and federal regulations, and pipeline technology is well established. Instead, the primary sources of variance in pipeline landslide risks relate to the geophysical differences between different routes and potential social and environmental consequences. In effect, this means we are able to analyze how FERC assesses a largely homogenous set of projects located under varying risk environments.

We utilized FERC's list of "approved major [natural gas] pipeline projects" (FERC, 2020b) to identify the natural gas pipelines that meet the diameter and length criteria for this study. At the time of data collection, the FERC list included pipeline projects from 1997 up through 2019. We reviewed each project on the list – nearly five hundred projects in all – and noted each project that was over 100 miles long. Next, we reviewed the EIAs for each of these projects in order to determine whether the pipelines satisfied the diameter and more refined length criteria. Through this process we identified forty-six natural gas pipeline mega-projects certified by FERC from 1997 to 2019 (Fig. 1). We identified an additional nine pipeline projects over 100 miles long for which EIAs were either not locatable or were designated "critical energy infrastructure projects" and not publicly accessible, and hence could not be included in this study.

3.2. *Measuring landslide risk*

In assessing landslide risks to pipelines, we consider risk in its most general form, where greater exposure to potential landslide hazards corresponds to higher risk (Kaplan and Garrick, 1981). We measure risk as the total miles of high landslide risk areas crossed by each project. We use this measure as the risk from potential landslide hazards is a function of the length, not the relative fraction, of a pipeline that is routed across high-risk areas. While longer pipelines have the potential to cross more high landslide risk areas on account of their greater lengths, the empirical data presented here demonstrate this is not the case in practice. This is due to the fact that landslide hazards are not randomly distributed across the United States, and further reflects decisions made by the project developers to route – or conversely, to avoid routing – pipelines across high-risk areas.

To empirically measure landslide risks across pipeline projects, we rely on information reported in the EIAs for each of the forty-six gas pipeline mega-projects. Pipeline EISs and EAs frequently employ the Landslide Overview Map of the Conterminous United States for identifying low, moderate, and high landslide risk areas (USGS, 2016). Landslide risk is typically reported in terms of incidence (i.e., where landslides have previously occurred) and susceptibility (i.e., where future landslides are more likely to occur). Following this classification system, high landslide incidence/susceptibility is defined as regions where greater than 15% of the map area has experienced landslides/is susceptible to future landslides. Given idiosyncrasies in the way landslide risks were reported across all forty-six EIAs, here we consider high landslide risk to be those areas classified as high incidence and/or susceptibility.[3] We report the total miles of each pipeline that cross high landslide risk areas (Table 2), and do so specifically for the reach of pipeline identified using the methodology outlined in the previous section (i.e., onshore, contiguous, new pipeline). We only consider landslide risks associated with the pipeline itself, not for other project infrastructure (e.g., compressor stations, access roads, etc.). We did not independently review the landslide information provided in the EISs and EAs, but we note each study was deemed acceptable by FERC, as each of the forty-six pipeline projects was certified by the Commission.

We assess the landslide risk associated with each pipeline project in the analysis, as well as examine three qualitative, aggregated risk groups. The three levels of risk we examine are "no high-risk areas", "some high-risk areas", and "considerable high-risk areas". Pipelines that do not traverse any high landslide risk areas belong to the first category ($n = 23$). We made the distinction between the latter two categories based on a natural split observed in the empirical data; pipelines crossing 36 or fewer miles of high landslide risk areas ($n = 14$), and pipelines crossing 54.5 or more miles of high-risk areas ($n = 9$) (with no observations falling in between 36 and 54.5 total high-risk miles). In this manner, we are able to examine the levels of recycled text and landslide risk on a project-by-project basis, as well as interpret general trends among the different groups of pipelines.

3.3. *Measuring text similarity*

In order to prepare the forty-six pipeline EISs and EAs for text similarity analysis, we converted the Geologic Hazards section of each PDF document into plain text files using the *pdftools* package (Ooms, 2018) in the R programming language (R Development Core Team, 2016). All but two of the documents were provided as machine-readable PDFs on the FERC Online eLibrary website, with the two exceptions being image files of scanned hard copies. We used optical character resolution with the *tesseract* R package (Ooms, 2019) to extract text from these image files. We manually reviewed each of the text files for accuracy against the original documents, and corrected any errors prior to analysis. We consider only the main body of text in the Geologic Hazards section, and

---

[3] This is of course a simplified approach. We do not attempt to parse out the various triggers of landslides (e.g., earthquakes, heavy precipitation, etc.), nor do we distinguish between risks associated with different types of landslides (e.g., debris flows, earth slumps). We further recognize that the classification system used in the United States Geological Survey map is a regional, not site-specific, metric.

Table 2
Summary of U.S. natural gas pipeline mega-projects from 1997 to 2019.

Pipeline Name[a]	Pipeline Location[a] (U.S. States)	EIA Docket Number[b]	EIA Issuance (year)	Mainline Diameter (inches)	Pipeline Length[c] (miles)	High Landslide Risk (miles)
Northern Border Pipeline	IA, IL	CP95-194	1997	36	243.1	0.2
Medicine Bow Lateral Project[d]	CO, WY	CP99-102	1999	24	149	0
FGT Phase IV Expansion Project	FL	CP99-94	2000	36	113	0
Millennium Pipeline	PA, NY	CP98-150	2001	36	384.4	33.2
Gulfstream Pipeline	AL, MS, FL	CP00-6	2001	36	306.3	0
Cheyenne Plains Pipeline	CO, KS	CP03-302	2004	36	387.2	0
Piceance Basin Expansion Project	CO, WY	CP05-54	2005	24	141.8	0.2
Entrega Pipeline	CO, WY	CP04-413	2005	42	328.1	0
Carthage-Perryville Pipeline	LA, TX	CP06-85	2006	42	172.1	0
Cypress Pipeline	FL, GA	CP05-388	2006	24	167	0
Cheniere Trail Pipeline	LA	CP05-357	2006	42	116.8	0
Guardian Expansion Project	WI	CP07-8	2007	30	119.2	0
Southeast Supply Header Pipeline	LA, MS, AL	CP07-44	2007	42	270.7	0
Southeast Expansion Project	AL, MS	CP07-32	2007	42	110.8	35.3
Kanda Lateral Project[d]	UT, WY	CP07-14	2007	24	123.7	12
Elba Express Pipeline	GA, SC	CP06-470	2007	42	187.9	0
Phoenix Expansion Project	AZ	CP06-459	2007	42	260.7	0
Louisiana Pipeline	LA	CP06-449	2007	42	135.7	0
East Texas-Mississippi Expansion Project	LA, MS	CP06-446	2007	42	240.3	36
Rockies Express West Project	CO, KS, MO, NE, WY	CP06-354	2007	42	795.3	0
Midcontinent Express Pipeline	OK, AL, LA, MS, TX	CP08-6	2008	42	510.3	84.9
Fayetteville-Greenville Expansion Project	MS, AR	CP07-417	2008	36	263.8	31.2
Gulf Crossing Pipeline	TX, OK, LA	CP07-398	2008	42	356.3	28.2
Rockies Express East Pipeline	MO, IL, IN, OH	CP07-208	2008	42	639.1	140
High Plains Expansion Project	CO	CP07-207	2008	30	163.7	0
Fayetteville Express Pipeline[d]	AR, MS	CP09-433	2009	42	185	18
Bison Pipeline	MT, ND, WY	CP09-161	2009	30	301.2	67.4
Ruby Pipeline	NV, OR, UT, WY	CP09-54	2010	42	675.2	25.7
ETC Tiger Pipeline[d]	LA, TX	CP09-460	2010	42	175.4	0
Appalachian Gateway Project[d]	PA, WV	CP10-448	2011	30	102.5	7.8
Virginia Southside Expansion Project[d]	VA	CP13-30	2013	24	98	0
Constitution Pipeline[e]	NY, PA	CP13-499	2014	30	124.4	0
Rover Pipeline	MI, OH, PA, WV	CP15-93	2016	42	510.3	224.1
Southeast Market Pipelines	AL, FL, GA	CP15-17	2016	36	642.5	0
Dalton Expansion Project[d]	GA	CP15-117	2016	30	114.9	18.3
Mountaineer XPress Pipeline	WV	CP16-357	2017	36	170.5	170.5
NEXUS Gas Transmission Project	MI, OH	CP16-22	2017	36	256.6	9
Mountain Valley Pipeline	VA, WV	CP16-10	2017	42	303.5	225.6
Atlantic Coast Pipeline[e]	NC, VA, WV	CP15-554	2017	42	604.5	187.2
Leach XPress Pipeline	OH, PA, WV	CP15-514	2017	36	133.6	118
Atlantic Sunrise Pipeline	PA	CP15-138	2017	42	185.9	54.5
Northern Access Pipeline[d]	NY, PA	CP15-115	2017	24	99	0
Midcontinent Supply Header Pipeline	OK	CP17-458	2018	36	234.1	0
PennEast Pipeline	NJ, PA	CP15-558	2018	36	120.2	23.5
Louisiana-Texas Connector Project	LA, TX	CP17-21	2019	42	170.2	0
Rio Bravo Pipeline	TX	CP16-454	2019	42	137.9	0

[a] It is not uncommon for pipeline EIAs to contain multiple different projects. The name and location given here refer to the primary project included in the analysis.

[b] Pipeline projects often have multiple associated FERC docket numbers, however, the number provided here represents the primary EIS/EA document. The combination of the docket number and the year the EIS/EA was issued will identify the relevant assessments utilized in the analysis.

[c] The length reported here does not include offshore pipeline segments, pipeline loop, pipe replacement, or unrelated projects also contained within the same EIA.

[d] These pipeline projects use EAs, as opposed to full EISs.

[e] The Constitution and Atlantic Coast pipelines were cancelled by their respective developers in 2020, but are included here as FERC previously approved both projects.

do not include the content of any footnotes, tables and figures, or table and figure captions in the analysis.

The Geologic Hazards section is a passage of text within the larger Environmental Analysis section of the EIAs. We focus on this section given our specific interest in the characterization of landslide risks. While landslides may also be discussed in other sections of the EISs and EAs, the general purpose of the Geologic Hazards section is to characterize and quantify geophysical risks, describe risk minimization and mitigation strategies, and summarize these risks for decision-makers and the public. More detailed information, including raw data, is typically provided in appendices and supplemental filings. Since the Geologic Hazards section also discusses other hazards (e.g., earthquakes, karst terrain, flooding, etc.), in analyzing the text we subset only those sentences containing the following set of key landslide-related terms: landslide, slip, debris flow, slope failure, ground failure, slope movement, slope instability, and steep slope. Since pipeline EIAs can also contain multiple different projects, prior to analysis we removed the landslide-related sentences for any unrelated projects. We note that two of the forty-six pipeline projects – the Cheyenne Plains and Elba Express pipelines – do not contain any discussion of landslide risks in their

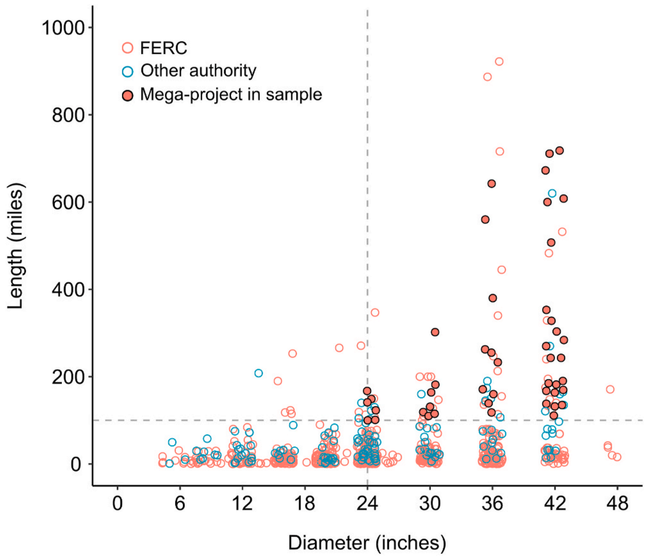

Fig. 1. Scatterplot (with jittered points) displaying the length and diameter of U.S. natural gas pipeline projects (EIA, 2020b), and the sample of mega-projects included in this study. The grey dashed lines denote the length and diameter thresholds.

Geologic Hazards sections, and therefore they are not included in the text similarity analysis.

The fundamental approach we use for assessing text similarity is the co-occurrence of words and phrases in each text. We draw on term frequency-inverse document frequency (*tf-idf*) weights to assess pairwise similarity between the project texts (Schütze et al., 2008). We operationalize term frequency (*tf*) as the number of times each individual, unique word appears in a document. Inverse document frequency (*idf*) is the natural log of the total number of documents in the corpus divided by the number of documents containing the same term – in this case, the same word – and reflects the prevalence of a term *t* across the entire corpus of documents *d*. The product of tf_d and idf_t gives *tf-idf* weight, which reflects the relative importance of a given term as a distinguishing feature in the document. The *tf-idf* weight for a given term in a given document goes up as term count within the document increases, and down as the term occurs in more documents. Thus, weights are highest for terms that occur frequently in a small number of documents (Schütze et al., 2008).

The vectors of *tf-idf* weights for each pairwise combination of EIA documents are then compared using cosine similarity – i.e., the cosine of the angle between the respective vectors of *tf-idf* weights for all observed terms in each document (Han et al., 2011). This provides a score between 0 and 1, where 1 indicates perfect similarity between the texts, and 0 represents perfect dissimilarity. An advantage of cosine similarity is that it is length-normalized to facilitate comparison of different length texts (Schütze et al., 2008). Although *tf-idf* weights are a "bag of words" approach – the order of words and phrases in a text is not factored in – this approach is still well suited for identifying areas of highly similar text. In theory, two texts could have a high *tf-idf* weight cosine similarity and very little similar prose, but in practice word order and word occurrence are not random. Thus, texts with high *tf-idf* weight cosine similarity scores correspond to cases with highly similar, if not almost identical, prose.

We additionally note that prior to analyzing the texts we performed a suite of standard text preprocessing steps, including employing a common dictionary of stop words (e.g., an, of, the) and removing proper nouns. Removing stop words that are not of substantive interest allows text similarity measures to be based strictly on words that define the meaning of the various texts. Proper nouns, such as project names and state names, are removed because they also hinder cross-text comparison. These words and phrases inherently differ between EIA documents, and thus typically have high *tf-idf* weights (since they are frequently used in one particular EIA and sparingly, if ever, used in others), but these weights do not reflect interesting differences in project risk assessment. We provide all the data files and code for the text analysis on a publicly accessible repository (Hileman et al., 2021).

3.4. Analyzing correspondence between text similarity and landslide risks

The procedure described in the preceding section yields a text similarity score $textsim(p_i,p_j)$ for every pairwise combination of pipeline projects p_i and p_j in the dataset. We computed an analogous risk similarity score $risksim(p_i,p_j)$ for every pair of pipelines based on the Euclidean distance between absolute lengths of identified high landslide risk areas. We take the log of the Euclidean distance in order to distinguish between pairs of pipelines with similar overall risks, but where, for example, one pair of pipelines belongs to the "some risk" group and the other pair to the "considerable risk" group. For pipelines p_i and p_j crossing high landslide risk areas of length *l*, this amounts to:

$$risksim(p_i,p_j) = 1 - \sqrt{\left(log_{10}(l_{p_i}) - log_{10}\left(l_{p_j}\right)\right)^2} \tag{1}$$

To assess the text and landslide risk similarity for each pair of pipelines, we apply a straightforward measure of differences between text similarity and risk similarity for every pair of pipelines. First, we apply a scale-invariant transformation to $risksim(p_i,p_j)$ so that both measures are scaled on the same 0 to 1 interval. We then directly compute the difference in risks: $diff(p_i,p_j) = textsim(p_i,p_j) - risksim(p_i,p_j)$. This results in a joint text-risk similarity score that ranges from −1 to 1, and which has a direct qualitative interpretation in the context of the study. Values close to 1 signify pipeline projects with very similar landslide-related texts and very different landslide risk levels (e.g., $textsim(p_i,p_j) = 0.9$ and $risksim(p_i,p_j) = 0.1$). Values close to −1 signify projects with very different landslide-related texts and very similar levels of landslide risk (e.g., $textsim(p_i,p_j) = 0.15$ and $risksim(p_i,p_j) = 0.95$). Returning to the aims of this study, pairs of pipelines with negative similarity scores are what we might expect to see: projects that employ more unique, project-specific text to describe landslide risks. Pairs of pipeline projects with large positive scores indicate the use of more generic, recycled text across projects with varying levels of landslide risk.

We recognize that some amount of recycled text is expected in EIAs, given the extended timeline and scope of the EIA process favors strategies that expedite and economize the process. Therefore, we assume the values for $textsim(p_i,p_j)$ are likely to vary in more or less predictable ways, given the specific features of a pair of pipeline projects. To assess this phenomenon, we model $textsim(p_i,p_j)$ for all 946 pairs of pipeline projects using a generalized linear model within a Bayesian framework. The model is premised on the intuition that similar locations and levels of landslide risk for any two projects should increase their text similarity scores. The model also reduces measurement-dependent variation in text similarity that is introduced on account of the varying number of landslide-related sentences each project contains. In other words, the model controls for the general tendency of text similarity scores to increase as the number of landslide-related sentences increases, which reflects the fact that a higher volume of text presents more opportunities for text similarity to occur.

We use posterior parameter distributions from the model to compute distributions of expected text similarity scores conditional on individual project features – i.e., the level of landslide risk, the geographic location of the project, and the number of landslide-related sentences. We then assess whether pairs of pipeline projects have considerably higher $textsim(p_i,p_j)$ scores than expected, which we define as being above the upper bound of the 86% high posterior density interval. It is important to emphasize that this is not a causal model, and we do not interpret the parameter estimates directly. Instead, our goal is to estimate a set of model parameters that predict text similarity reasonably well, while providing an uncertainty estimate in order to identify observations that are well outside that estimate. We include a detailed mathematical description of the full model as supplementary material (Appendix A).

4. Results

The forty-six natural gas pipeline mega-projects in this dataset vary in length from nearly 100 miles long to just under 800 miles long, and are routed across anywhere from 0 to 225 miles of high landslide risk areas (Fig. 2). Half of the forty-six pipelines do not cross any identified high-risk areas. The four pipelines with the highest levels of landslide risk – respectively, the Mountain Valley Pipeline, Rover Pipeline, Atlantic Coast Pipeline, and Mountaineer XPress Pipeline – cross more miles of high-risk areas than all the other pipelines in the dataset combined. Each of these four pipelines was certified by FERC during a two-year period over 2016–2017. This time period corresponds to the buildout of pipeline infrastructure in the landslide-prone Appalachian Mountains, where six of the ten mega-projects approved for construction in 2016 and 2017 are located.

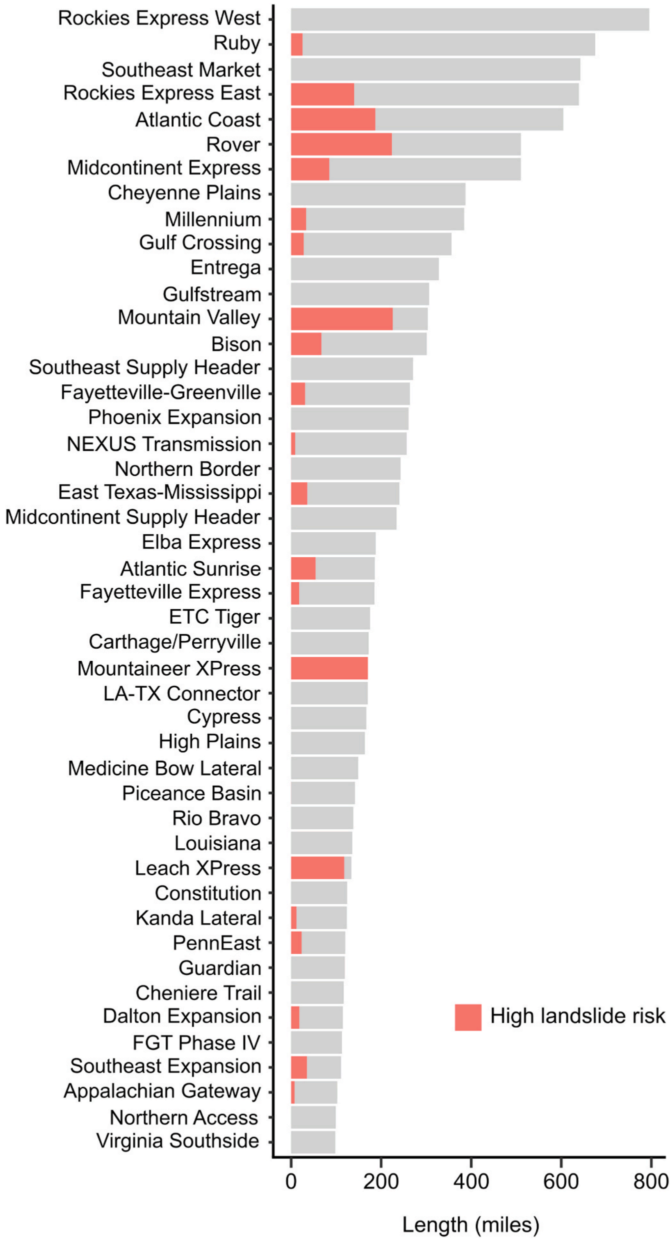

Fig. 2. Bar plot displaying the length of each pipeline project in the dataset, and the miles of high landslide risk areas crossed by each, with projects ordered by pipeline length.

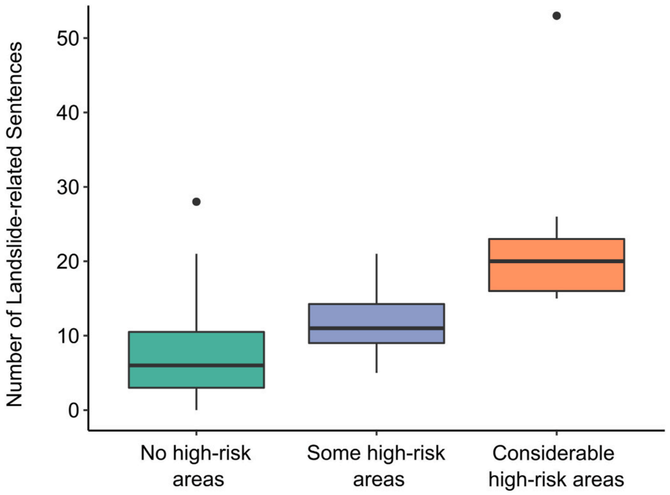

Fig. 3. Number of sentences containing landslide-related terms for pipelines that cross no high landslide risk areas low (n = 23), some high-risk areas (n = 14), and considerable high-risk areas (n = 9).

The variation in levels of landslide risk is largely reflected in the number of landslide-related sentences contained in the Geologic Hazards section of the EIAs for each pipeline project (Fig. 3). Overall, pipelines that do not cross any high-risk areas tend to contain fewer sentences describing landslide risks, while projects that cross considerable high-risk areas contain the most sentences discussing landslides. These results, however, do not provide any indication of how landslide risks are actually being characterized and discussed within the various project EIAs.

In terms of text similarity, we find that the three landslide risk categories contain projects employing both generic recycled text and project-specific descriptions (Fig. 4). Overall, text similarity across pipeline projects within each risk group tends to increase with risk. The median text similarity score for pipelines in the “no high-risk areas” group is 0.19, which increases to 0.25 in the “some high-risk areas” group, and increases further to 0.36 in the “considerable high-risk areas” group. However, there is substantial spread in the scores within each group, and the “no risk” and “some risk” groups contain several pairs of

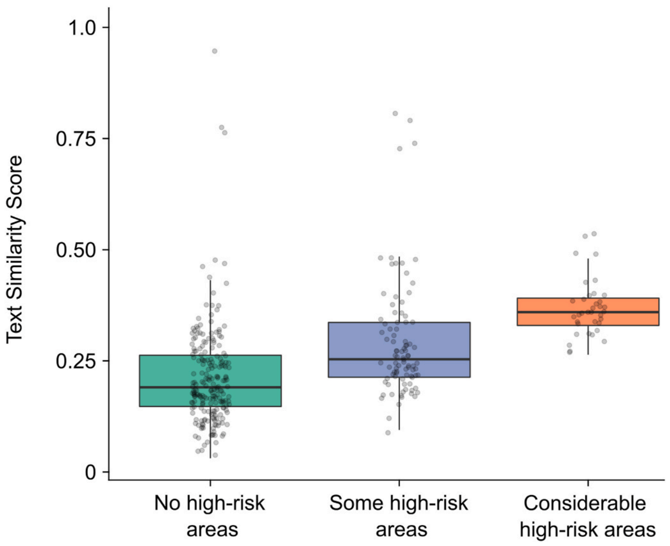

Fig. 4. Boxplots displaying the distributions of text similarity scores within the no high-risk, some high-risk, and considerable high-risk groups. Each point represents one pair of pipeline projects.

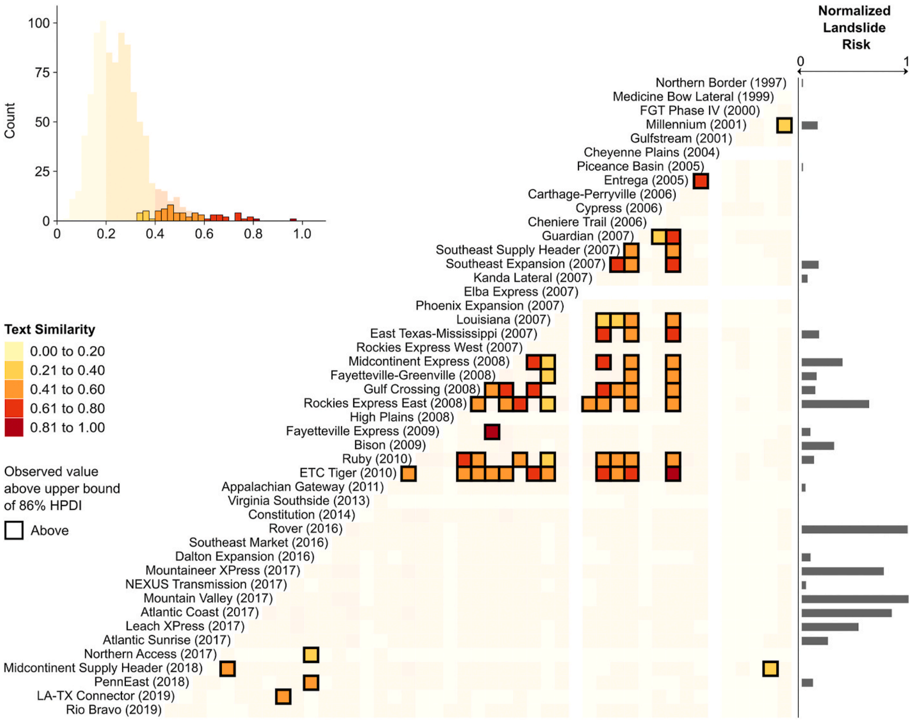

Fig. 5. Matrix highlighting the specific pairs of pipeline projects with higher-than-expected text similarity scores, along with an inset histogram displaying the distribution of similarity scores for each pair of projects, and a normalized measure of landslide risk for each project. Scores above the upper bound of the 86% high posterior density interval (HPDI) are denoted by the outlined bars and cells. For ease of reference, the projects are presented in the same order as they appear in Table 2.

pipeline projects with text similarity scores greater than 0.75. These results illustrate that, while there is variation in the levels of text similarity across all sets of pipeline projects, most projects share at least some common text with other projects in the dataset. This baseline level of similarity is expected – the Geologic Hazards section of the EIAs describes the same basic set of geophysical issues, for the same reasons, so some degree of text similarity is inherent due to common subject matter and relevant terms.

Given that there is a relatively high baseline level of text similarity expected since these EIA sections all address the same set of issues (as opposed to comparing, for example, a Geologic Hazards section and a Cultural Resources section), our focus here is primarily on pairs of projects with higher-than-expected similarity. These cases display commonality in the text going beyond shared subject matter and purpose. Fig. 5 displays results for all pairs of pipeline projects in the dataset, and highlights those projects where the observed text similarity scores are outside the expected range of values, based on our statistical model accounting for the projects' locations and risk profiles. While only 7% of all pairs of projects in the dataset share texts that are more similar than expected – a finding that holds for a range of low and high text similarity scores – the presence of groups of projects with very similar texts suggests the use of recycled text may be more closely related to features of the EIA process (e.g., time period, authorship teams, responsible sub-organization), rather than substantive similarities or differences in landslide risk across projects. The loose cluster of around a dozen projects in the middle of the matrix represents pipelines that were certified by FERC between 2006 and 2010. Most of these projects, although not all, are located in states along the Gulf of Mexico in the southeast and south-central U.S. These projects vary considerably in overall length, as well as landslide risk profiles. Multiple pipeline projects in this cluster share highly similar texts with seven to ten other pipeline projects in the cluster.

Fig. 6 displays the joint text-risk similarity score for each pair of pipeline projects in the dataset. Negative scores (blue cells) refer to cases where landslide risks are relatively similar, but the EIA texts are relatively dissimilar. Positive scores (red cells) refer to the inverse – similar text, but dissimilar landslide risks. What is of focal interest are the outlying small clusters of projects with high positive text-risk similarity scores. In these cases, a common document format and means of characterizing landslide risks was used despite considerable differences in the level of landslide risk. There are thirteen sets of projects that score between 0.41 and 0.7. These pairs of projects tend to be located in the same regions and occur at similar points in time. For example, the Rover, Mountaineer XPress, and Atlantic Coast pipelines were approved over 2016–2017, and all are routed across portions of the Appalachian Mountains in West Virginia. Similarly, the Southeast Expansion, East Texas-Mississippi, and Midcontinent Express pipelines were approved over 2007–2008, and all are routed across Mississippi and neighboring states in the south-central U.S. Overall, the high rate of low text-risk similarity sets of pipeline projects (many observed scores closer to −1) relative to high text-risk similarity sets (no observed scores above 0.7) is expected. After all, the text from a given EIA cannot be highly similar to two other EIAs which are themselves quite different. Thus, when viewing all pairwise comparisons at once, the overall density of highly similar sets should decrease as the number of projects increases.

5. Discussion

5.1. Unpacking text similarity and landslide risks in pipeline EIAs

In thinking about the implications of text recycling in the assessment of energy infrastructure projects, it is important to emphasize that standardized, common text is not inherently problematic. Starting from scratch on every EIA would be inefficient and make it harder to draw upon an agency's existing basis of knowledge, and streamlined procedures are expected to the extent that they help federal agencies, in this case FERC, accomplish their aims. The point of this analysis is not to argue that recurring text in EIAs is evidence that FERC is shirking its

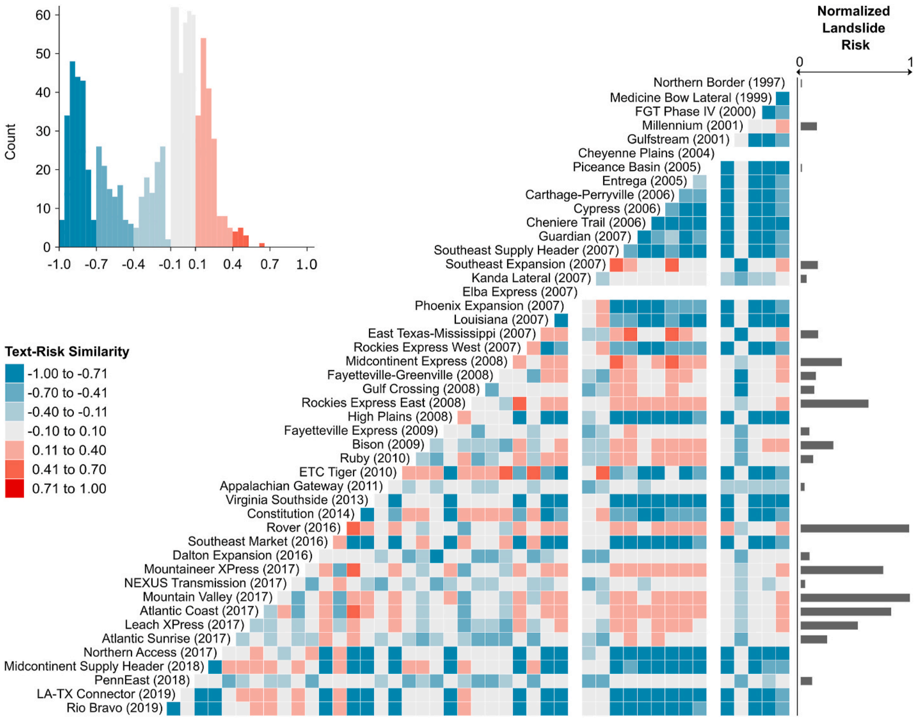

Fig. 6. Matrix displaying the joint text-risk similarity score for each pair of pipeline projects, along with an inset histogram displaying the distribution of similarity scores, and a normalized measure of landslide risk for each project. For ease of reference, the projects are presented in the same order as they appear in Table 2.

responsibility, as recycled text alone is not sufficient evidence for such an argument. Rather, we seek to consider the implications of this practice – that existing bureaucratic procedures and the administrative demands of the EIA draft, notice, and comment process can be ill suited for clearly portraying and conveying risks associated with major energy infrastructure projects.

To understand how and why this might be the case, we review several examples from the EIAs of the pipeline mega-projects studied here. Of particular interest in this study is the presence of projects with high text similarity scores and low landslide risk similarity (i.e., pairs of pipeline projects with similarity scores closer to 1 in Figs. 5 and 6), indicating a high level of recycled text in spite of considerable differences in landslide risk. For example, the Gulf Crossing and Midcontinent Express pipelines use highly similar text to discuss general sources of landslide risk. These two projects were both certified by FERC in 2008 and are both located in the south-central U.S., yet they respectively cross 28.2 and 84.9 miles of high landslide risk areas. Given the landslide-related texts of their EIAs are significantly more similar than expected, factors other than overall landslide risk, such as project location, may be generally driving the utilization of recycled text. Or, it may signal that FERC believes these differences in landslide risk are not of any practical consequence. Regardless, in spite of the observed difference in the levels of landslide risk, the following paragraph characterizing landslide hazards exists in the final EISs for both pipeline projects – and indeed, it is present in multiple other project EISs located in the same region, and approved around the same time – with minor variations in word-for-word text noted in brackets:

> "Several factors contribute to slope failures and subsequent landslides including the degree of slope or tilt of geologic materials, the composition of the materials, the amount of man-made disturbance of the materials, proximity to seismic activity, and the amount of rainfall exposure. Generally flat areas were selected for the location of the proposed [compressor and meter station/aboveground facility] sites; therefore, slope failure is not expected at [these above ground facility sites/those facility locations]. [However], slope failures and landslides represent a potential hazard along portions of the proposed Project route that would traverse areas of side slopes and rolling terrain, [or areas identified as potentially prone to landslide events]. Factors that would increase the potential for slope failures along [steep] slopes and rolling terrain include cutting along slopes, the weight of construction equipment, and unusually high precipitation." (FERC, 2008a, 2008b).

We also find examples in the EIAs where verbiage is highly similar, but numeric risk measures are substituted across projects with very different risk profiles. For example, the Mountain Valley Pipeline crosses 225 miles of high landslide risk areas – the most out of all forty-six pipelines in this dataset – while the recently cancelled Constitution Pipeline was not slated to cross any high landslide risk areas. Respectively, the two pipelines were approved by FERC in 2017 and 2014, and are located in Virginia/West Virginia and New York/Pennsylvania. The following passages from their respective final EISs showcase different numeric figures being presented in very similar passages of text:

> "Several locations were identified as having the susceptibility for landslides within the vicinity of the proposed pipeline. Approximately 25 miles of the pipeline route in Pennsylvania is considered to have a moderate to low susceptibility to landslides. In New York, approximately 15 miles of the proposed alignment has a moderate to low susceptibility to landslides, and the remaining 83 miles has a low susceptibility." (FERC, 2014).

> "Several locations were identified as having a high incidence of and high susceptibility for landslides within the vicinity of the MVP. About 152 miles (77 percent) of the MVP pipeline route in West Virginia is considered to have a high incidence of and high susceptibility to landslides. In Virginia, about 51 miles (48 percent) of the proposed alignment has a high incidence of and high susceptibility to landslides." (FERC, 2017).

When viewed side-by-side it becomes apparent that, while the two excerpts are highly similar, one project faces much greater landslide risk than the other. The Constitution Pipeline crosses no areas of high landslide susceptibility/incidence, while the Mountain Valley Pipeline is routed across more than 200 miles of such high-risk areas. This risk has been borne out in practice. As noted previously, Mountain Valley Pipeline has had to contend with numerous landslide events along its 303-mile long route through Virginia and West Virginia, even while construction remained largely stalled due to regulatory challenges and federal permit suspensions (FERC, 2019). However, in the EIA documents from which these excerpts are drawn, no comparisons are made – the individual studies do not compare the reported risks of the focal project to other similar projects, or otherwise contextualize an appropriate level of concern. Instead, one must read multiple EISs to understand that the risks reported for the Mountain Valley Pipeline are quite considerable relative to its peers. Thus, the use of a standardized template and reporting language deter effective, clear risk communication.

A similar issue arises with respect to reported mitigation strategies. For example, the ETC Tiger Pipeline does not cross any high landslide risk areas, while the Midcontinent Express Pipeline is routed across more than 80 miles of high-risk areas. Both projects are located in the south-central U.S., and were respectively approved by FERC in 2010 and 2008. The respective EA and final EIS for these projects each describe how *"construction of the pipeline [and restoration of disturbed areas] would be accomplished in accordance with [ETC Tiger's/MEP's] Plan, which includes measures to control runoff and erosion that would minimize the potential for slope failures."* (FERC, 2008b, 2010) Furthermore, *"[ETC Tiger/MEP] would [also] implement specialized two-tone construction techniques to provide for safe working conditions in areas [of side slopes] potentially susceptible to slope failures."* (FERC, 2008b, 2010) A number of other pipeline projects approved between 2006 and 2010, and located in various regions of the U.S., also contain this exact same passage of text. Without searching across multiple EIA documents, however, interested parties would be unaware that this is just generic language rather than a project-specific mitigation strategy.

5.2. *Future research considerations*

The objective of this study was chiefly to develop and apply a method for measuring recycled text in EIAs as a function of empirically observed landslide risk. The examples of recycled text we provide illustrate potential problems with the widespread use of this practice to expedite the EIA process for gas pipeline projects. More generally, however, there are many other useful research applications of the text analysis methods used in this study. The energy sector is rife with policies and procedures that result in large document corpuses, including FERC hydropower facility licensing processes (Ulibarri, 2015), electric utility resource plans (Wilkerson et al., 2014), state and local regulations governing facility siting and utility operations (Ottinger et al., 2014; Schumacher and Yang, 2018), and – as explored in this study – environmental review processes for infrastructure projects and management programs (Scott et al., 2020). Tools for measuring recycled or similar text can be applied to these corpuses to trace the diffusion of policy ideas (Wilkerson et al., 2015), map patterns of coordination in energy policy networks (Hsu and Rauber, 2021), and measure policy similarity (Linder et al., 2020).

Furthermore, while this study measures similarity in EIA texts, it does not measure topical content. Future work using automated content analysis methods, including unsupervised machine learning techniques, such as topic modeling, and supervised machine learning techniques such as word embeddings and part-of-speech tagging, can shed further light on this issue by identifying the different content foci of EIAs and how risk is discussed. This can be used to develop a more holistic picture of what concepts and ideas are more or less likely to be recycled. In addition to landslide hazards, future studies should consider investigating text similarity in relation to other topics. For example, environmental justice and the siting of energy infrastructure (Finley-Brook et al., 2018; Emanuel, 2017), and contributions of U.S. energy projects to global climate change (Howarth et al., 2011), were recently both named as priority areas for FERC moving forward (Skibell, 2021).

6. Conclusion and policy implications

Given EIAs are a procedural requirement without fixed decision standards under NEPA, it is not clear whether there is presently any objective standard for weighing the social and environmental risks of pipeline projects that would lead FERC or other agencies to deem a project too risky to approve. In this respect, it is noteworthy that FERC certified all forty-six pipelines in our sample, regardless of the reported differences in risk. Environmental protection is not FERC's primary mandate, and NEPA places only procedural, not distributive, requirements on agency decisions. Thus, it is not surprising that energy supply and security might take precedence in these decisions. However, it is striking that since 1999, FERC has only rejected two natural gas pipeline applications – roughly 0.4% of all the applications it received – and neither was rejected due to social or environmental risks (Mattei and Sanzillo, 2020; Tierney, 2019). One possible explanation for this is that unsuitable projects are rejected before even undergoing formal EIA review. Given the time and costs involved in EIA preparation, it stands to reason that both applicants and the agency have an incentive to identify – and avoid – particularly problematic projects prior to assessment. However, if the EIA process is meant to thoroughly vet projects and garner new information through public notice and comment, the fact that not one assessment resulted in a pipeline being rejected points to the possibility that EIA review of pipelines is largely *pro-forma*. This comports with our finding that landslide hazards are presented similarly, in spite of significant differences in underlying risks.

The homogeneity we observe with respect to risk presentation in pipeline EIAs indicates that the current approach does not do an adequate job of communicating risks to decision-makers and the public. While observed specifically in the case of natural gas pipeline mega-projects, there is little reason to expect that this problem is limited to large pipelines, or pipeline infrastructure alone. For smaller projects, the magnitude of the risk may be lessened by reduced pipeline volumes, but agencies face the same pressures to streamline reporting and communication of risk during the permitting process. If anything, the diminished public salience of smaller pipelines and more limited scope of EA reviews (and if applicable, Categorical Exclusion documents) likely increases formulaic risk communication behaviors. More generally, the case of pipeline EIAs echoes existing literature describing how the technocratic, rational model of the EIA process (Bartlett, 1997; Bartlett and Kurian, 1999) can be a hindrance to effective use of this critically important decision tool (Bradbury, 1994; Ortolano and Shepherd, 1995). For instance, in cases where text is highly similar and numerical risk figures are substituted, readers – including FERC staff – would have to study multiple EIAs for different projects to get a full sense of what the relative risk is for any one given project. Newer models of risk assessment, presentation, and communication offer considerable potential to improve how risks are reported within EIA processes (Gregory et al., 2020; Lundgren and McMakin, 2018; van der Vegt, 2018).

While EIA reform efforts often focus on macro-level factors such as length and time to completion (Doyle, 2017; Thomas et al., 2019), an important consideration for future research and policy discussion is how NEPA and related EIA laws can maintain procedural consistency and adhere to common regulatory standards, while providing a more robust and effective assessment of landslides and other geophysical project risks. One could envision, for instance, developing a standardized approach for presenting project risks in comparison to other similar projects. This would allow agencies to retain some of the potential benefits of standardized language (e.g., speed, clarity, and legal precedent), while enhancing risk communication effectiveness. Energy policy scholars have documented the role that risk perception plays in shaping public acceptance and stakeholder support for many kinds of energy infrastructure (Dowd et al., 2011; Mah et al., 2014; Songsore and Buzzelli, 2014; Wadley et al., 2019). Developing clearer ways to present and discuss project risks within the EIA process will certainly not alleviate all project conflicts, but nonetheless will provide better information to inform public risk perception and more closely align the stated aims of the EIA process with practice. Such efforts are already underway – FERC recently released a Notice of Inquiry seeking public input on the natural gas pipeline certification process (FERC, 2021), which presents new opportunities for key stakeholders and the general public to advocate for changes in how landslides and other high-consequence events are evaluated in EIAs.

In summary, this study explored the question of whether the text of EIAs prepared for natural gas pipelines varies across projects with different geophysical risk profiles. We find evidence that pipeline projects facing different levels of landslide risk use similar, and at times nearly identical, passages of recycled text to characterize and summarize landslide hazards, and discuss risk minimization and mitigation strategies. Policy analysis and risk communication scholars have long noted the challenges associated with summarizing rare high-consequence events like landslides within the framework of decision tools such as EIAs (Dooley, 1985; Fairley, 1981; Stewart and Leschine, 1986). However, decision-makers and the public typically lack the time and technical expertise to analyze the underlying data or review past studies, so focused high-level summaries in EIAs are critical for informed decision-making. Given the pressures that preparers of EIAs face to be simultaneously timely, comprehensive, and brief (Doyle, 2017), and to avoid litigation (Ruple and Race, 2020), it is understandable that EIAs might take on a boilerplate quality with substitution of verbiage across projects, and the insertion of project-specific metrics around generic recycled text. Rather than imply that every project EIA should contain entirely novel text, the key takeaway is that the model of risk analysis and communication used in EIAs merits reconsideration. Present practices risk transforming an important decision tool into a more rote, formulaic procedural task, and with potentially profound consequences.

CRediT authorship contribution statement

Jacob D. Hileman: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. **Mario Angst:** Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. **Tyler A. Scott:** Investigation, Writing – original draft, Writing – review & editing. **Emma Sundström:** Formal analysis, Methodology, Writing – original draft.

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.

Acknowledgments

The authors wish to thank Daniel Nohrstedt, Chris Vrettos, and Federico Holm for their helpful comments and suggestions on earlier versions of this paper, along with the three anonymous reviewers for the time and effort they put into providing valuable feedback during the review process.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.enpol.2021.112379.

Funding and declaration of interests

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The authors declare no conflicts of interest.

References

Austin, J., Carter, J.M., Klein, B.D., Schang, S.E., 2004. Judging NEPA: A“ Hard Look” at Judicial Decision Making under the National Environmental Policy Act. Environmental Law Institute. https://www.eli.org/sites/default/files/eli-pubs/d14-12.pdf.

Bartlett, R.V., 1997. The rationality and logic of NEPA revisited. Environmental policy and NEPA: past, present, and future, 51–60.

Bartlett, R.V., Kurian, P.A., 1999. The theory of environmental impact assessment: implicit models of policy making. Pol. Polit. 27 (4), 415–433. https://doi.org/10.1332/030557399782218371.

Ben-Shahar, O., 2007. Boilerplate: the Foundation of Market Contracts. Cambridge University Press.

Bergquist, P., Ansolabehere, S., Carley, S., Konisky, D., 2020. Backyard voices: how sense of place shapes views of large-scale energy transmission infrastructure. Energy Research & Social Science 63, 101396. https://doi.org/10.1016/j.erss.2019.101396.

Bradbury, J.A., 1994. Risk communication in environmental restoration programs. Risk Anal.: An Official Publication of the Society for Risk Analysis 14 (3), 357–363. https://doi.org/10.1111/j.1539-6924.1994.tb00252.x.

CEQ, 2018. CEQ Guidance. Council on Environmental Quality. https://ceq.doe.gov/guidance/guidance.html.

CEQ, 2019. Update to the Regulations Implementing the Procedural Provisions of the National Environmental Policy Act (No. CEQ–2019–0003). Council on Environmental Quality. https://www.whitehouse.gov/wp-content/uploads/2020/01/NEPA-NPRM-01092020_Pre-publication-version.pdf.

CEQ, 2020. Categorical Exclusions. NEPA.gov: National Environmental Policy Act. Council on Environmental Quality. https://ceq.doe.gov/nepa-practice/categorical-exclusions.html.

CRS, 2019. This Land Is Your Land? Eminent Domain under the Natural Gas Act and State Sovereign Immunity. Congressional Research Service. https://crsreports.congress.gov/product/pdf/LSB/LSB10359.

DEQ, 2018. MVP Site Inspection Reports: Cahas Mountain Road. Virginia Department of Environmental Quality. https://www.deq.virginia.gov/Portals/0/DEQ/Water/Pipelines/MVP_CIR_Cahas%205-21-18%20.pdf?ver=2018-07-27-111529-823.

DOE, 2017. Lessons Learned Quarterly Report. U.S. Department of Energy. Energy.gov. https://www.energy.gov/nepa/listings/lessons-learned-quarterly-report.

Dooley, J.E., 1985. Risk Theory and the Environmental Assessment Process. Environ. Impact Assess. Technol. Assess. Risk Anal. 15–47. https://doi.org/10.1007/978-3-642-70634-9_2.

Dowd, A.-M., Boughen, N., Ashworth, P., Carr-Cornish, S., 2011. Geothermal technology in Australia: investigating social acceptance. Energy Pol. 39 (10), 6301–6307. https://doi.org/10.1016/j.enpol.2011.07.029.

Doyle, M., 2017. Interior: Order Limits Most NEPA Studies to a Year. E&E News, p. 150. https://www.eenews.net/stories/1060059865.

EIA, 2017. Appalachia Region Drives Growth in U.S. Natural Gas Production since 2012. U.S. Energy Information Administration. https://www.eia.gov/todayinenergy/detail.php?id=33972.

EIA, 2020a. Natural Gas Pipelines. U.S. Energy Information Administration. https://www.eia.gov/energyexplained/natural-gas/natural-gas-pipelines.php.

EIA, 2020b. U.S. Natural Gas Pipeline Projects [Data set]. https://www.eia.gov/naturalgas/pipelines/EIA-NaturalGasPipelineProjects.xlsx.

Emanuel, R.E., 2017. Flawed environmental justice analyses. Science 357 (6348). https://doi.org/10.1126/science.aao2684, 260-260.

Emerson, K., Baldwin, E., 2019. Effectiveness in NEPA decision making: in search of evidence and theory. J. Environ. Pol. Plann. 21 (4), 427–443. https://doi.org/10.1080/1523908X.2019.1615421.

Entrekin, S., Evans-White, M., Johnson, B., Hagenbuch, E., 2011. Rapid expansion of natural gas development poses a threat to surface waters. Front. Ecol. Environ. 9 (9), 503–511. https://doi.org/10.1890/110053.

EPA, 2013a. National Environmental Policy Act. Environmental Protection Agency. https://www.epa.gov/nepa.

EPA, 2013, July 31. National Environmental Policy Act Review Process. Environmental Protection Agency. In: https://www.epa.gov/nepa/national-environmental-policy-act-review-process.

Fairley, W.B., 1981. Assessment for catastrophic risks. Risk Anal.: Off. Publ. Soc. Risk Anal. 1 (3), 197–204. https://doi.org/10.1111/j.1539-6924.1981.tb01416.x.

FERC, 1999. Certification of New Interstate Natural Gas Pipeline Facilities (Policy Statement). Federal Energy Regulatory Commission. https://www.ferc.gov/legal/maj-ord-reg/PL99-3-000.pdf.

FERC, 2008a. Gulf Crossing Pipeline: Final Environmental Impact Statement. Federal Energy Regulatory Commission.

FERC, 2008b. Midcontinent Express Pipeline Project: Final Environmental Impact Statement. Federal Energy Regulatory Commission.

FERC, 2010. ETC Tiger Pipeline Project: Environmental Assessment. Federal Energy Regulatory Commission.

FERC, 2014. Constitution Pipeline and Wright Interconnect Projects: Final Environmental Impact Statement. Federal Energy Regulatory Commission.

FERC, 2017. Mountain Valley Project and Equitrans Expansion Project: Final Environmental Impact Statement. Federal Energy Regulatory Commission.

FERC, 2018, August 14. About FERC - What FERC Does. https://www.ferc.gov/about/ferc-does.asp.

FERC, 2019. Cessation of Certain Activities: FERC Correspondence with Applicant (Oct. 11, 2019). Federal Energy Regulatory Commission.

FERC, 2020a. Mountain Valley Pipeline Environmental Compliance Monitoring Program: Weekly Summary Report (April 5-11, 2020). Federal Energy Regulatory Commission.

FERC, 2020, January. Approved Major Pipeline Projects (1997-Present). Federal Energy Regulatory Commission. https://ferc.gov/industries/gas/indus-act/pipelines/approved-projects.asp.

FERC, 2021. Notice of Inquiry: Certification of New Interstate Natural Gas Facilities (174 FERC ¶ 61,125). Federal Energy Regulatory Commission. https://ferc.gov/sites/default/files/2021-02/C-1-PL18-1-000.pdf.

Finley-Brook, M., Williams, T.L., Caron-Sheppard, J.A., Jaromin, M.K., 2018. Critical energy justice in US natural gas infrastructuring. Energy Research & Social Science 41, 176–190. https://doi.org/10.1016/j.erss.2018.04.019.

Flyvbjerg, B., 2014. What you should know about megaprojects and why: An overview. Proj. Manag. J. 45 (2), 6–19. https://doi.org/10.1002/pmj.21409.

Glasson, J., Therivel, R., 2013. Introduction To Environmental Impact Assessment. Routledge.

Gregory, R., Satterfield, T., Boyd, D.R., 2020. People, pipelines, and probabilities: clarifying significance and uncertainty in environmental impact assessments. Risk Anal.: Off. Publ. Soc. Risk Anal. 40 (2), 218–226. https://doi.org/10.1111/risa.13409.

Han, J., Pei, J., Kamber, M., 2011. Data Mining: Concepts and Techniques, 3rd Edition. Elsevier.

Hays, R., 1983. Prescriptions for using boilerplate. IEEE Trans. Prof. Commun. 26 (2), 60–62. https://doi.org/10.1109/TPC.1983.6448684.

Hileman, J., Angst, M., Scott, T., Sundström, E., 2021. Replication files for “Recycled text and risk communication in natural gas pipeline environmental impact assessments” [Data set]. https://zenodo.org/record/4742964.

Howarth, R.W., Santoro, R., Ingraffea, A., 2011. Methane and the greenhouse-gas footprint of natural gas from shale formations. Climatic Change 106 (4), 679. https://doi.org/10.1007/s10584-011-0061-5.

Hsu, A., Rauber, R., 2021. Diverse climate actors show limited coordination in a large-scale text analysis of strategy documents. Communications Earth & Environment 2 (1), 30. https://doi.org/10.1038/s43247-021-00098-7.

Jackson, R.B., Vengosh, A., Carey, J.W., Davies, R.J., Darrah, T.H., O’Sullivan, F., Pétron, G., 2014. The environmental costs and benefits of fracking. Annu. Rev. Environ. Resour. 39 (1), 327–362. https://doi.org/10.1146/annurev-environ-031113-144051.

Johnson, K., 2016. Annual National Environmental Policy Act (NEPA) Report 2015. https://ceq.doe.gov/docs/get-involved/NAEP_2015_NEPA_Annual_Report.pdf.

Jo, Y.-D., Ahn, B.J., 2002. Analysis of hazard areas associated with high-pressure natural gas pipelines. J. Loss Prev. Process. Ind. 15 (3), 179–188. https://doi.org/10.1016/S0950-4230(02)00007-4.

Jo, Y.-D., Crowl, D.A., 2008. Individual risk analysis of high-pressure natural gas pipelines. J. Loss Prev. Process. Ind. 21 (6), 589–595. https://doi.org/10.1016/j.jlp.2008.04.006.

Kaplan, S., Garrick, B.J., 1981. On the quantitative definition of risk. Risk Anal.: Off. Publ. Soc. Risk Anal. 1 (1), 11–27. https://doi.org/10.1111/j.1539-6924.1981.tb01350.x.

Lanzano, G., Salzano, E., de Magistris, F.S., Fabbrocino, G., 2013. Seismic vulnerability of natural gas pipelines. Reliab. Eng. Syst. Saf. 117, 73–80. https://doi.org/10.1016/j.ress.2013.03.019.

Larsen, S.V., Hansen, A.M., Nielsen, H.N., 2018. The role of EIA and weak assessments of social impacts in conflicts over implementation of renewable energy policies. Energy Pol. 115, 43–53. https://doi.org/10.1016/j.enpol.2018.01.002.

Linder, F., Desmarais, B., Burgess, M., Giraudy, E., 2020. Text as policy: measuring policy similarity through bill text reuse. Pol. Stud. J.: The Journal of the Policy Studies Organization 48 (2), 546–574. https://doi.org/10.1111/psj.12257.

Lundgren, R.E., McMakin, A.H., 2018. Risk Communication: A Handbook for Communicating Environmental, Safety, and Health Risks. John Wiley & Sons.

Lyles, K.C., 2017. Expediting the NEPA process via a document management system and virtual GIS-based NEPA platform. Environ. Pract. 19 (3), 139–147. https://doi.org/10.1080/14660466.2017.1359012.

Mah, D.N.-Y., Hills, P., Tao, J., 2014. Risk perception, trust and public engagement in nuclear decision-making in Hong Kong. Energy Pol. 73, 368–390. https://doi.org/10.1016/j.enpol.2014.05.019.

Mattei, S., Sanzillo, T., 2020. FERC’s Failure to Analyze Energy Market Forces: Risks to Ratepayers, Landowners and the Overall Economy. Institute for Energy Economics and Financial Analysis. http://ieefa.org/wp-content/uploads/2020/12/FERCs-Failure-to-Analyze-Energy-Market-Forces_December-2020.pdf.

MVP, 2019. Supplement to Variance Request No. A-78. FERC Online eLibrary (Accession Number 20190808-5134) [Letter to Federal Energy Regulatory Commission].

Ooms, J., 2018. pdftools: Text Extraction, Rendering and Converting of PDF Documents. https://CRAN.R-project.org/package=pdftools.

Ooms, J., 2019. tesseract: Open Source OCR Engine. https://CRAN.R-project.org/package=tesseract.

Ortolano, L., Shepherd, A., 1995. Environmental impact assessment: challenges and opportunities. Impact Assessment 13 (1), 3–30. https://doi.org/10.1080/07349165.1995.9726076.

Ottinger, G., Hargrave, T.J., Hopson, E., 2014. Procedural justice in wind facility siting: recommendations for state-led siting processes. Energy Pol. 65, 662–669. https://doi.org/10.1016/j.enpol.2013.09.066.

PHMSA, 2018. Notice of Proposed Safety Order [Letter to Columbia Gas Transmission Llc]. https://www.phmsa.dot.gov/sites/phmsa.dot.gov/files/docs/news/58376/columbia-gas-npso-july-2018.pdf.

PHMSA, 2019. Pipeline Safety: Potential for Damage to Pipeline Facilities Caused by Earth Movement and Other Geological Hazards. Pipeline and Hazardous Materials Safety Administration. https://www.phmsa.dot.gov/regulations-fr/notices/2019-08984.

PHMSA, 2020. Safety Awareness Overview. Pipeline and Hazardous Materials Safety Administration. https://www.phmsa.dot.gov/safety-awareness/pipeline/safety-awareness-overview.

PST, 2015. Pipeline Briefing Paper #2: Pipeline Basics & Specifics About Natural Gas Pipelines. Pipeline Safety Trust. http://pstrust.org/wp-content/uploads/2015/09/2015-PST-Briefing-Paper-02-NatGasBasics.pdf.

R Development Core Team, 2016. R: A Language and Environment for Statistical Computing. CRAN.

Ruple, J.C., Race, K.M., 2020. Measuring the NEPA litigation burden. Environ. Law 50 (2), 479–522. https://www.jstor.org/stable/26939867.

Schumacher, K., Yang, Z., 2018. The determinants of wind energy growth in the United States: drivers and barriers to state-level development. Renew. Sustain. Energy Rev. 97, 1–13. https://doi.org/10.1016/j.rser.2018.08.017.

Schütze, H., Raghavan, P., Manning, C.D., 2008. Introduction to Information Retrieval, Vol. 1. Cambridge University Press, Cambridge, United Kingdom.

Scott, T.A., Ulibarri, N., Perez-Figueroa, O., 2020. NEPA and national trends in federal infrastructure siting in the United States. Rev. Pol. Res. 37 (5), 605–633. https://doi.org/10.1111/ropr.12399.

Skibell, A., 2021. Glick Unveils Environmental Justice, Climate Plans. E&E News. https://www.eenews.net/stories/1063725039.

Songsore, E., Buzzelli, M., 2014. Social responses to wind energy development in Ontario: the influence of health risk perceptions and associated concerns. Energy Pol. 69, 285–296. https://doi.org/10.1016/j.enpol.2014.01.048.

Soraghan, M., 2019. Landslides, Explosions Spark Fear in Pipeline Country. E&E News. https://www.eenews.net/stories/1060472727.

Southerland, M.T., 2004. Environmental impacts of dispersed development from federal infrastructure projects. Environ. Monit. Assess. 94 (1–3), 163–178. https://doi.org/10.1023/b:emas.0000016886.16085.39.

Stewart, T.R., Leschine, T.M., 1986. Judgment and analysis in oil spill risk assessment. Risk Anal.: Off. Publ. Soc. Risk Anal. 6 (3), 305–315. https://doi.org/10.1111/j.1539-6924.1986.tb00223.x.

Thomas, C., Garrett, C., Maynard, B., Houser, K., 2019. NEPA streamlining yet again: Will the diet work this time? Nat. Resour. Environ. 33 (3), 34–37.

Tierney, S.F., 2019. FERC’s Certification of New Interstate Natural Gas Facilities: Revising the 1999 Policy Statement for 21st Century Conditions. Analysis Group.

Ulibarri, N., 2015. Collaboration in federal hydropower licensing: impacts on process, outputs, and outcomes. Publ. Perform. Manag. Rev. 38 (4), 578–606. https://doi.org/10.1080/15309576.2015.1031004.

USGS, 2016. Landslide Overview Map of the Conterminous United States. https://pubs.usgs.gov/pp/p1183/plate1.html.

van der Vegt, R.G., 2018. Risk assessment and risk governance of liquefied natural gas development in Gladstone, Australia. Risk Anal.: Off. Publ. Soc. Risk Anal. 38 (9), 1830–1846. https://doi.org/10.1111/risa.12977.

Vengosh, A., Jackson, R.B., Warner, N., Darrah, T.H., Kondash, A., 2014. A critical review of the risks to water resources from unconventional shale gas development and hydraulic fracturing in the United States. Environ. Sci. Technol. 48 (15), 8334–8348. https://doi.org/10.1021/es405118y.

Vidic, R.D., Brantley, S.L., Vandenbossche, J.M., Yoxtheimer, D., Abad, J.D., 2013. Impact of shale gas development on regional water quality. Science 340 (6134), 1235009. https://doi.org/10.1126/science.1235009.

Wadley, D.A., Han, J.H., Elliott, P.G., 2019. Risk hidden in plain sight: explaining homeowner perceptions of electricity transmission infrastructure. Energy Pol. 132, 744–753. https://doi.org/10.1016/j.enpol.2019.06.022.

Wang, Q., Chen, X., Jha, A.N., Rogers, H., 2014. Natural gas from shale formation – the evolution, evidences and challenges of shale gas revolution in United States. Renew. Sustain. Energy Rev. 30, 1–28. https://doi.org/10.1016/j.rser.2013.08.065.

Wilkerson, J., Larsen, P., Barbose, G., 2014. Survey of Western U.S. electric utility resource plans. Energy Pol. 66, 90–103. https://doi.org/10.1016/j.enpol.2013.11.029.

Wilkerson, J., Smith, D., Stramp, N., 2015. Tracing the flow of policy ideas in legislatures: a text reuse approach. Am. J. Polit. Sci. 59 (4), 943–956. https://doi.org/10.1111/ajps.12175.

Witter, R.Z., McKenzie, L., Stinson, K.E., Scott, K., Newman, L.S., Adgate, J., 2013. The use of health impact assessment for a community undergoing natural gas development. Am. J. Publ. Health 103 (6), 1002–1010. https://doi.org/10.2105/AJPH.2012.301017.

Xiao, J., Shi, P., Wang, Y.-F., Yu, Y., Yang, L., 2017. A framework for quantifying the extent of impact to plants from linear construction. Sci. Rep. 7 (1), 2488. https://doi.org/10.1038/s41598-017-02443-3.