

Data Dictionary

ISA Literature Screening Dataset

(Last revised: May 25, 2022)

[semantic_map_<year>.csv](#)

[reference_metadata_<year>.csv](#)

[citation_context_<year>.csv](#)

semantic_map_<year>.csv

This table represents a semantic annotation of the outline (table of contents) of an Ozone ISA. Each row corresponds to a **“section”** (broadly defined at any level, which includes paragraphs between that section heading and the next section heading) listed in the outline. Sections in the first few chapters of an ISA, including “Preamble”, “Preface”, “Legislative and Historical Background”, “Executive Summary”, “Integrative Summary”, and “Integrated Synthesis”, are not included because these chapters do not constitute the actual systematic literature survey as those later chapters on specific subject matters.

Table 1. Semantic map schema.

Column Name	Description	Example
isa_year	Year in which this ISA was released.	2013
level_all_num	Complete section number. Information on this line of the table only covers the body text directly under this section number (between this section heading and the next section heading). For example, “3.4.1” only covers text directly following Section 3.4.1, not those texts under Section 3.4.1.1 or Section 3.4.1.2.	3.4.1
level1_num	Level-1 section number. It is the 1st number in level_all_num.	3
level2_num	Level-2 section number.	4

	It is the 2nd number in level_all_num.	
level3_num	Level-3 section number. It is the 3rd number in level_all_num.	1
level4_num	Level-4 section number. It is the 4th number in level_all_num.	.
level5_num	Level-5 section number. It is the 5th number in level_all_num.	.
level6_num	Level-6 section number. It is the 6th number in level_all_num.	.
level1_char	Level-1 section title in characters. It corresponds to the chapter title.	Atmospheric Chemistry and Ambient Concentrations
level2_char	Level-2 section title in characters. It corresponds to the section title, if any.	Background O3 Concentrations
level3_char	Level-3 section title in characters. It corresponds to the subsection title, if any.	Contributions from Natural Sources
level4_char	Level-4 section title in characters. It corresponds to the subsubsection title, if any.	(empty)
level5_char	Level-5 section title in characters. It corresponds to the subsubsubsection title, if any.	(empty)
level6_char	Level-6 section title in characters. It corresponds to the subsubsubsubsection title, if any.	(empty)
chapter_category	Chapter level category label.	Atmospheric
topic	Section-level topic label. If topic = "Supplementary", it means this section either does not contain subject matter or only summarizes other sections that do contain subject matter.	Background ozone
discipline	Section-level discipline level.	Atmospheric

reference_metadata_<year>.csv

This table is the complete set of metadata for all references, including those found in search results and those cited in the topical (non-supplementary) sections of the Ozone ISA.

Table 2. Reference metadata schema.

Column Name	Description	Example
REFERENCE_ID	HERO ID. The numerical identifier of a reference in EPA's Health & Environmental Research Online (HERO) database. It can be used to retrieve reference metadata in https://hero.epa.gov/hero/index.cfm/search/index . This field is the primary key of the table. It cannot be empty.	1065937
PMID	PubMed ID. The numerical identifier of a reference in PubMed. It can be used to retrieve reference metadata in https://pubmed.ncbi.nlm.nih.gov/ . This field can be empty if the reference does not have a PMID.	15803198
WOSID	Web of Science ID. The numerical identifier of a reference in Web of Science. This field can be empty if the reference does not have a WOSID.	WOS:000228093000014
DOI	Digital Object Identifier of a reference. This field can be empty if the reference does not have a DOI.	10.1038/nrg1578
TITLE	Title text of a reference.	Gene-environment interactions in human diseases
AUTHORS	List of authors of a reference, separated by semicolon.	Hunter, DJ
YEAR	Year of publication of a reference.	2005

<p>ABSTRACT</p>	<p>Abstract text of a reference.</p>	<p>Tropospheric ozone (O₃) negatively impacts human health and ecosystems, and is a greenhouse gas. Wildfires are a source of tropospheric O₃, and studies show that wildfires are increasing in North America. In this study, we present a critical review of O₃ production from wildfires focusing on three key topics: the influence of wildfire emissions on O₃ production; the influence of photochemistry on wildfire O₃ production; and regulatory issues associated with wildfire O₃ production in the United States. Observations of O₃/CO range from approximately 0.1 to 0.9, and are caused by the interplay of numerous factors including fire emissions, efficiency of combustion, chemical and photochemical reactions, aerosol effects on chemistry and radiation, and local and downwind meteorological patterns. Using average O₃/CO ratios for major biomes, we estimate global wildfires produce approximately 170 Tg of O₃ per year, which is 3.5% of all global tropospheric O₃ production. Areas of uncertainty in wildfire O₃ production include the net effect of aerosols on chemical and photochemical reactions within a fire plume, the impact of oxygenated volatile organic compounds and nitrous acid on O₃ production, and the interplay of variables that lead to extreme O₃/CO values. Because wildfire frequencies are likely increasing and have been shown to contribute to elevated O₃ at air quality monitoring sites, it is important to better understand the emissions,</p>
-----------------	--------------------------------------	--

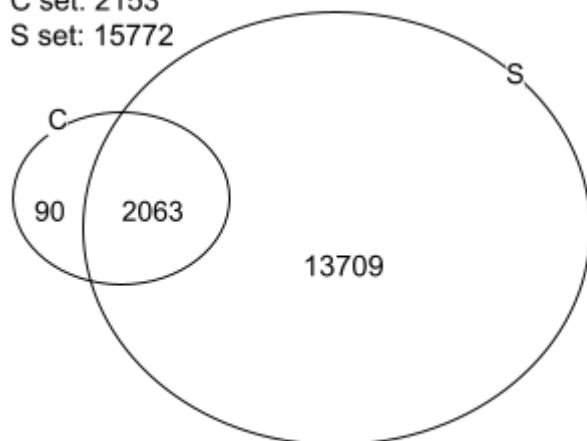
		photochemistry and impacts of these fires.
CITED	Whether the reference is cited in any non-Supplementary section (see the description of Column "topic" in the semantic map schema above).	Y: yes. N: no.
IN_SEARCH	Whether the reference was in the systematic search result set. Some references are cited in non-supplementary sections not through a systematic search, but by known-item search (known by experts who authored that section).	Y: yes. N: no.

Venn diagrams:

2013:

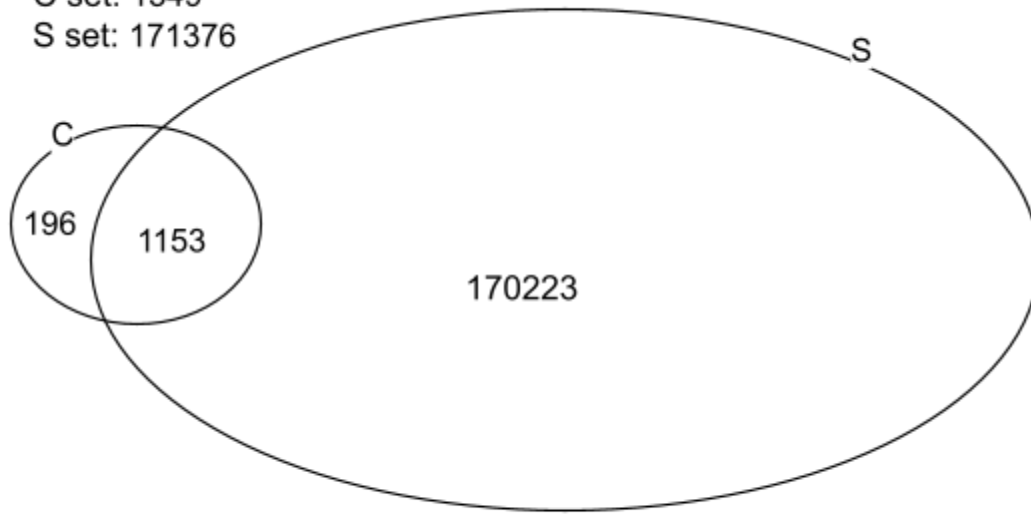
C set: 2153

S set: 15772



Total: $90+2063+13709 = 15862$

2020:
C set: 1349
S set: 171376



Total: $196+1153+170223 = 171572$

citation_context_<year>.csv

This table contains the information for each citation instance in the non-supplementary sections (see the description of Column “topic” in the [semantic map schema](#) above). Each row corresponds to a different instance of a citation. The same reference can be cited multiple times across the ISA. In fact, the same reference can be cited multiple times (i.e., have multiple instances of citations) within one section or even within one paragraph.

Table 3. Citation context schema.

Column Name	Description	Example
INSTANCE_ID	Citation instance ID. This field is the primary key of the table. It cannot be empty.	230
REFERENCE_ID	HERO ID. The numerical identifier of a reference in EPA's Health & Environmental Research Online (HERO) database. It can be used to retrieve reference metadata in https://hero.epa.gov/hero/index.cfm/search/index . One HERO ID (i.e., one reference) can be cited in multiple instances throughout the ISA.	47756
level_all_num	Complete section number. Information on this line of the table only covers the body text directly under this section number. For example, “3.4.1” only covers text directly following Section 3.4.1, not those texts under Section 3.4.1.1 or Section 3.4.1.2.	3.4.1.1
level1_num	Level-1 section number. It is the 1st number in level_all_num.	3

level2_num	Level-2 section number. It is the 2nd number in level_all_num.	4
level3_num	Level-3 section number. It is the 3rd number in level_all_num.	1
level4_num	Level-4 section number. It is the 4th number in level_all_num.	1
level5_num	Level-5 section number. It is the 5th number in level_all_num.	.
level6_num	Level-6 section number. It is the 6th number in level_all_num.	.
level1_char	Level-1 section title in characters. It corresponds to the chapter title.	Atmospheric Chemistry and Ambient Concentrations
level2_char	Level-2 section title in characters. It corresponds to the section title, if any.	Background O3 Concentrations
level3_char	Level-3 section title in characters. It corresponds to the subsection title, if any.	Contributions from Natural Sources
level4_char	Level-4 section title in characters. It corresponds to the subsubsection title, if any.	Contributions from the Stratosphere
level5_char	Level-5 section title in characters. It corresponds to the subsubsubsection title, if any.	(empty)
level6_char	Level-6 section title in characters. It corresponds to the subsubsubsubsection	(empty)

	title, if any.	
ANCHOR_TEXT	The citation text snippet that makes the reference.	Shapiro, 1980
CONTEXT_PARAGRAPH	The text of the paragraph in which this citation takes place. In other words, the context of this citation instance.	<p>The basic atmospheric dynamics and thermodynamics of STE were outlined in the 2006 O3 AQCD (U.S. EPA, 2006b); as noted there, stratospheric air rich in O3 is transported into the troposphere. Ozone is produced naturally by photochemical reactions in the stratosphere as shown in Figure 3-1. Some of this O3 is transported downward into the troposphere throughout the year, with maximum contributions at mid-latitudes during late winter and early spring mainly coming from a process known as tropopause folding. These folds occur behind most cold fronts, bringing stratospheric air with them. The tropopause should not be interpreted as a material surface through which there is no exchange. Rather these folds should be thought of as regions in which mixing of tropospheric and stratospheric air is occurring (Shapiro, 1980). This imported stratospheric air contributes to the natural background of O3 in the troposphere, especially in the free troposphere during winter and spring. Significant intrusions of stratospheric air occur in "ribbons" ~200 to 1000 km in length, 100 to 300 km wide and about 1 to 4 km thick (Wimmers et al., 2003; Hoskins, 1972). Thus, these intrusions are large scale three-dimensional events and should not be thought of as one-dimensional. STE also occurs during other seasons including summer.</p>