Recommended Updates to the Soil-to-Groundwater Pathway in the U.S. EPA's 1996 Soil Screening Guidance

Seth Whiteaker[a,b], and Alex Valdez[a,c]

[a] *Virtual Student Federal Service Internship Program, Environmental Protection Agency, Washington, DC*

[b] *Ecological Management and Restoration Bachelor's Degree Program, University of California, Davis, CA*

[c] *Environmental Management Master's Degree Program, University of Maryland Global Campus, Adelphi, MD*

Abstract: This document provides an overview of the history, current use, and recommended improvements of the dilution attenuation factor (DAF) used when calculating soil screening levels (SSL), along with recommendations to address the assumption of an infinite contaminant source when calculating SSLs. This report focuses on the United States Environmental Protection Agency's (U.S. EPA) 1996 Soil Screening Guidance (SSG), which provides guidance for calculating SSLs as implemented in the Regional Screening Level (RSL) and Preliminary Remediation Goal (PRG) calculators. The U.S. EPA's SSG currently utilizes a single default nationwide DAF during the calculation of SSLs when a risk of soil-to-groundwater contaminant migration is present, which does not reflect the significant hydrologic differences across the U.S. Additionally, the SSG's calculation assumes an infinite source of contaminant is present at a contaminated site, which may not be an accurate reflection of the conditions at sites. This report examines the history and initial development of the DAF, reviews potential issues with the current use of the DAF, and provides recommendations for improvement to the DAF calculations, including hydrologic region-specific DAFs, and a calculator to correct the infinite source assumption inherent in SSL calculations. The report concludes with the limitations of this analysis and recommendations for future work to improve the DAF and SSG.

Keywords: Dilution attenuation factor; Soil screening guidance; Risk assessment; Groundwater contamination; Correction factor;

1. Introduction

1.1 Defining the problem

Groundwater pollution via soil contamination is a major concern in the United States (U.S with 50% of the U.S. population depending on a groundwater source for numerous domestic uses, for example drinking and bathing (United States Environmental Protection Agency [U.S. EPA], 2023a). Drinking contaminated groundwater may result in significant acute, chronic, and/or carcinogenic adverse health effects (U.S. EPA, 2015). Groundwater contamination via soil comes from various sources of pollution including, but not limited to municipal solid waste landfills, illegal waste disposal, leaking storage tanks, runoff, and naturally occurring sources (U.S. Geological Survey, 2018). The U.S. EPA designates many of these heavily contaminated sites on the National Priorities List (NPL) under the authority of the Comprehensive Environmental Response, Compensation, and Liability Act of 1980 (CERCLA). Identifying what chemicals at a site present an unacceptable risk to human or environmental health is key to protecting vulnerable areas or populations, along with determining which sites require remediation. Effective risk characterization screens chemicals whose concentration in a media does not present a significant excess risk; likewise, chemicals whose concentrations potentially pose an excess risk are identified for further assessment and, if needed, remediation. Various technical documents provide guidance on how to determine appropriate screening levels in different media.

1.2 Risk based soil screening levels

The U.S. EPA's 1996 Soil Screening Guidance (SSG)

is the current framework used to determine what contaminant concentrations in soil require action/remediation. In 1996, the U.S. EPA published the SSG to accelerate the evaluation and cleanup of contaminated soils (U.S. EPA, 1996b). The SSG provides generic and site-specific guidance on screening chemicals of potential concern (COPCs) during the risk assessment process for contaminants present in soil (U.S. EPA, 1996a). Risk based soil screening levels (SSL) of pollutant concentrations in soil are calculated from standardized equations that combine exposure likelihood and pollutant toxicity to identify a specific pollutant concentration in soil that may require remediation (U.S. EPA, 1996c). For contaminated soil, exposure pathways include ingestion, inhalation (via particulate volatilization), and contaminant leaching into water sources (U.S. EPA, 1996a). Assuming groundwater is present, contaminants in soil may migrate to groundwater aquifers, presenting new exposure pathways via water ingestion or dermal contact. Developing sufficiently conservative (but not overly restrictive) contaminant screening levels aids in the identification of COPCs at a contaminated site, helping focus cleanup efforts to those contaminants posing a threat to groundwater.

1.3 Basics of soil to groundwater contamination

Soil to groundwater contamination occurs when a contaminant moves through the vadose zone into the saturated zone (Šimunek & van Genuchten, 2006). After leaching into soil, various physical properties of a contaminant (e.g., vapor pressure and soil adsorption) and the hydrogeologic characteristics of a soil (e.g., porosity, moisture levels, organic content, etc.) affect the contaminant transport (*Soil Pollution*, 2018). For example, highly soluble chemicals (e.g., salts) easily move through porous surface soils; less soluble chemicals may reside in soil for longer periods (National Pesticide Information Center, 2021). Eventually, the contaminant will reach the water table (if groundwater is present), potentially posing a further hazard to environmental and human health. Due to the typically slow rate of groundwater flow, contamination may linger for decades (Geophysics Research Forum, 1984). It is imperative to protect groundwater resources by identifying and remediating contaminated soil that poses a risk of contaminant migration to groundwater. Standardized equations provide a uniform method to identify soil remediation screening levels when contaminant migration to groundwater is possible.

Due to differences in physical properties between inorganic and organic compounds/contaminants, the equations that determine the soil screening level-migration to groundwater (SSL-MGW) differ slightly. Equation 1 defines the SSL-MGW for inorganic contaminants, while Equation 2 defines the SSL-MGW for organic contaminants.

Equation 1- SSL-MGW for inorganic compounds (U.S. EPA, 1996b)

$$Soil\ screening\ level\left(\frac{mg}{kg}\right) = C_w \cdot \left\{K_d + \frac{\theta_w + (\theta_a \cdot H')}{\rho_b}\right\}$$

Where:

C_w = target soil leachate concentration (mg/L)
K_d = Soil-water partition coefficient (L/kg)
θ_w = Water-filled soil porosity (L_{water}/L_{soil})
θ_a = Air-filled soil porosity (L_{air}/L_{soil})
H' = Henry's law constant (dimensionless)
ρ_b = Dry soil bulk density (kg/L)

Equation 2- SSL-MGW for organic compounds (U.S. EPA, 1996b)

$$SSL\left(\frac{mg}{kg}\right) = C_w \cdot \left\{(K_{oc} \cdot f_{oc}) + \frac{\theta_w + (\theta_a \cdot H')}{\rho_b}\right\}$$

Where:

C_w = target soil leachate concentration (mg/L)
K_{oc} = Soil organic carbon-water partition coefficient (L/kg)
f_{oc} = Organic carbon content of soil (kg/kg)
θ_w = Water-filled soil porosity (L_{water}/L_{soil})
θ_a = Air-filled soil porosity (L_{air}/L_{soil})
H' = Henry's law constant (dimensionless)
ρ_b = Dry soil bulk density (kg/L)

The SSL is the final calculated output of the equation, dictating the concentration in the unsaturated soil below which the given contaminant does not present a health concern from subsequent contaminant leaching into groundwater (U.S. EPA, 1996c).

K_d, K_{oc}, and H' are chemical specific constants. θ_w, θ_a, ρ_b, and f_{oc} may be determined for a specific site, but for the soil screening guidance, the U.S. EPA uses default values defined in the SSG User's Guide (U.S. EPA, 1996b).

C_w is a chemical specific value/contaminant concentration goal, based on a 1 x 10⁻⁶ individual excess cancer risk for carcinogenic contaminants and a hazard quotient of 1 for non-carcinogenic risks, calculated in Equation 3.

Equation 3- Determination of the target soil leachate concentration (C_w) (U.S. EPA, 1996c)

$$C_w = (MCLG,\ MCL,\ or\ HBL) \cdot DAF$$

Where:

MCLG = Maximum contaminant level goal (chemical specific concentration)
MCL = Maximum contaminant level (chemical specific concentration)
HBL = Health based limit (chemical specific concentration)
DAF= Dilution attenuation factor (dimensionless)

The MCLG, MCL, and HBL are chemical specific values derived from the maximum acceptable exposure to a chemical based on the human dose-response to that chemical (Gilbert, 2012).

1.4 Basics of the Dilution Attenuation Factor

In the *Soil Screening Guidance: User's Guide*, the U.S. EPA (1996b, p. 30) defines/describes the dilution attenuation factor below:

"As soil leachate moves through soil and ground water, contaminant concentrations are attenuated by adsorption and degradation. In the aquifer, dilution by clean ground water further reduces concentrations before contaminants reach receptor points (i.e., drinking water wells). This reduction in concentration can be expressed by a dilution attenuation factor (DAF), defined as the ratio of soil leachate concentration to receptor point concentration. The lowest possible DAF is 1, corresponding to the situation where there is no dilution or attenuation of a contaminant (i.e., when the concentration in the receptor well is equal to the soil leachate concentration). On the other hand, high DAF values correspond to a large reduction in contaminant concentration from the contaminated soil to the receptor well."

Equation 4 details how to calculate a DAF value, while Equation 5 calculates the aquifer mixing zone depth, d (a factor used in equation 4).

Equation 4- Calculation of the Dilution Attenuation Factor (U.S. EPA, 1996b)

$$DAF = 1 + \left(\frac{Kid}{IL}\right)$$

Where:

K= aquifer hydraulic conductivity (m/yr)
i= hydraulic gradient (m/m)
d=aquifer mixing zone depth (m)

I= infiltration rate (m/yr)
L= length of area of concern parallel to ground water flow (m)

Equation 5- Calculation of aquifer mixing zone depth (U.S. EPA, 1996b)

$$d = \sqrt{(0.0112L^2)} + d_a\left(1 - e^{\frac{(-LI)}{(Kid_a)}}\right)$$

Where:

d_a= aquifer thickness (m)
All other factors are the same as Equation 4

Heath (1984, p. viii) defines hydraulic conductivity (K) as, "The capacity of a rock to transmit water; expressed as the volume of water that will move in unit time under a unit hydraulic gradient through a unit area measured at right angles to the direction of flow". The hydraulic gradient (i) refers to the slope of a water table (Newell, Hopkins, & Bedient, 1990). The New Jersey Department of Environmental Protection (NJ DEP) (2021, p. 16) defines the aquifer mixing zone depth (d) as, "the depth to which the contaminant is diluted in ground water". The aquifer thickness (d_a) refers to the depth from the top of the water table to the bottom of the saturated zone (Newell, Hopkins, & Bedient, 1990). The NJ DEP (2021, p. 15) defines infiltration rate (I) as, "the rate of recharge of precipitation to the ground water". Finally, the length of area of concern parallel to ground water flow (L) refers to the length of the contaminant source in soil adjacent and parallel to the groundwater flow; in the context of soil-to-groundwater (SGW) migration, it is the length of contaminated soil directly leaching into the aquifer (U.S. EPA, 1996c).

The U.S. EPA's (1996c) default DAF is 20. Essentially, for every one part of contaminant concentration in soil, the corresponding ground water concentration of that contaminant is expected to be 1/20 of that value. The DAF in the SSG only accounts for leachate dilution, not any other form of attenuation like adsorption or degradation. The hydrogeologic factors in equations 4 and 5 can vary greatly from site to site. An investigation of the history and potential update of the default DAF is the focus of the remainder of this document.

2. DAF History

2.1 DAF development by the U.S. EPA

The U.S. EPA's SSG relied on two studies to determine a nationwide DAF standard (NJ DEP, 2021). In one study, via a Monte Carlo simulation (15,000 iterations with various input parameters found nationwide), the SSG applied the U.S. EPA Composite Model for Leachate Migration with Transformation Products (EPA CMTP) to produce the standard/default DAF (U.S. EPA, 1996c). The second modeling effort utilized data from two large hydrogeologic surveys. The American Petroleum Institute (API) conducted one survey and compiled the useable data (known as the hydrogeologic database [HGDB]), while the other survey contained U.S. EPA's own data collected from nationwide dense non-aqueous phase liquid (DNAPL) contaminated Superfund sites.

Raw data from the HGDB is in Appendix F of the U.S. EPA's (1996c) SSG. API conducted a nationwide survey of approximately 8700 members of the Association of Ground Water Scientists and Engineers (Newell, Hopkins, & Bedient, 1990). From this survey, API collected 400 useable responses containing data on the location, hydrologic characteristics of the aquifer, geologic characteristics, and other relevant data. The data included quantitative values of the aquifer hydraulic conductivity, the aquifer mixing zone depth, infiltration rate, hydraulic gradient, and the depth to the top of the aquifer. The HGDB contains data from 48 of the 50 states, with Arkansas and North Dakota the only states not to have any suitable responses. Figure 1 provides approximate locations of surveys providing useable data. Thus, the SSG considered the data set to be reasonably representative of hydrogeologic conditions across the United States. Therefore, the SSG used the HGDB and its own data from DNAPL contaminated Superfund sites in a Monte Carlo simulation to determine a nationwide/default DAF of 20.

Figure 1: Approximate locations of survey sites in the

HGDB (Source, Newell, Hopkins, & Bedient, 1990)

2.2 Potential issues with a nationwide default DAF

While this default value has been used across the United States, the default DAF does not account for the large hydrogeologic variations across the country or even within relatively small regions. Heath (1984) identified 15 hydrogeologic regions (composed of various numbers of sub-regions) across the United States. The HGDB contains data from 13 of these 15 regions (Puerto Rico/U.S. Virgin Islands were excluded, and Alluvial Valleys were incorporated into the other 13 regions) (U.S. EPA, 1996c). Hydrogeologic characteristics related to the calculation of a DAF vary substantially from different regions. For example, a glaciated mountain valley will typically have vastly different hydrogeologic characteristics than a coastal beach. Therefore, a nationwide default DAF may not be appropriate, even when only using it as part of a screening equation.

2.3 State level DAFs

To account for these variations within New Jersey, the NJDEP (2021) developed their own default DAF and detailed the methods behind their calculations. The NJDEP guidance also acknowledges that a site-specific DAF may be used. During research into the history of the DAF, the NJDEP's report presents the most up to date and detailed reasoning behind their decision to calculate a state specific DAF. The NJ DEP used some of the same data as the original 1996 U.S. EPA SSG, but only for comparable hydrogeologic settings in state. Coincidentally, even though the NJDEP's chosen hydrogeologic parameters used to calculate a DAF differed from the U.S. EPA's, the NJDEP calculated a state specific DAF of 20, no different than the U.S. EPA's SSG. Still, based on the wide variation of hydrogeologic characteristics across different regions of the United States, state or hydrogeologic region specific DAFs are expected to vary considerably from the U.S. EPA's SSG default.

Based on the conducted research, all states either clearly state using a default DAF of 1 or 20, or no default DAF is mentioned (therefore assumed the state uses the U.S. EPA's default DAF value), or only use site specific and calculated DAFs. For example, the Nevada Division of Environmental Protection uses a default DAF of 1 or 20 for their equivalent of a screening guidance (called a leaching-based basic comparison level) (Nevada Division of Environmental Protection, n.d.). Similarly, as stated in Appendix A of Minnesota's Pollution Control Agency's (2013) "Soil leaching values", Minnesota chooses to use the U.S. EPA's default DAF of 20 for their SSL-MGW calculations;

however, when calculating a DAF using the default K, i, d, I, and L factors also provided in Appendix A, the calculated DAF is 9.4. The Minnesota Pollution Control Agency unequivocally states that the U.S. EPA's default DAF of 20 is sufficiently protective for most sites with a 0.5-acre source area or less. Modeling results for larger source areas are included in the HGDB of the SSG.

2.4 Current Use of the DAF

In conjunction with the rest of the U.S. EPA (1996c) SSG, the DAF is used during the risk assessment process when soil-to-groundwater contamination is a possibility. When calculating generic screening levels, the default of 20 is used. If calculating a site-specific SSL and a site-specific DAF is required, the factors listed in Equation 4 must be measured on site or at least inferred from a geologically similar site. A site-specific DAF will provide a more accurate screening level, allowing a site practitioner to make a better-informed decision on whether a site needs remediation for a certain chemical(s). However, due to the specialized equipment, survey techniques, time, and cost required to accurately measure the factors listed in Equation 4, a site-specific DAF may not be feasible during the early stages of a site investigation. With the DAF history, assumptions, and current use accounted for, one can suggest updates to the DAF.

3. Hydrogeologic region specific DAF

3.1 Reasoning for a hydrogeologic region specific DAF

As previously stated, hydrogeologic conditions across the United States vary greatly. A hydrogeologic specific (but still regionally "generic") DAF would better predict contaminant movement from soil-to-groundwater than a nationwide default DAF. A map of the HGDB defined hydrogeologic regions is presented in Figure 2. While hydrogeologic characteristics do not neatly follow political boundaries (i.e., hydrogeologic regions cross state and county lines), the delineated regions can easily be shared with users as a file compatible with geographic information systems (GIS) or through web maps that do not require the use of specialized software. This project uncovered little literature investigating region-specific DAFs. One
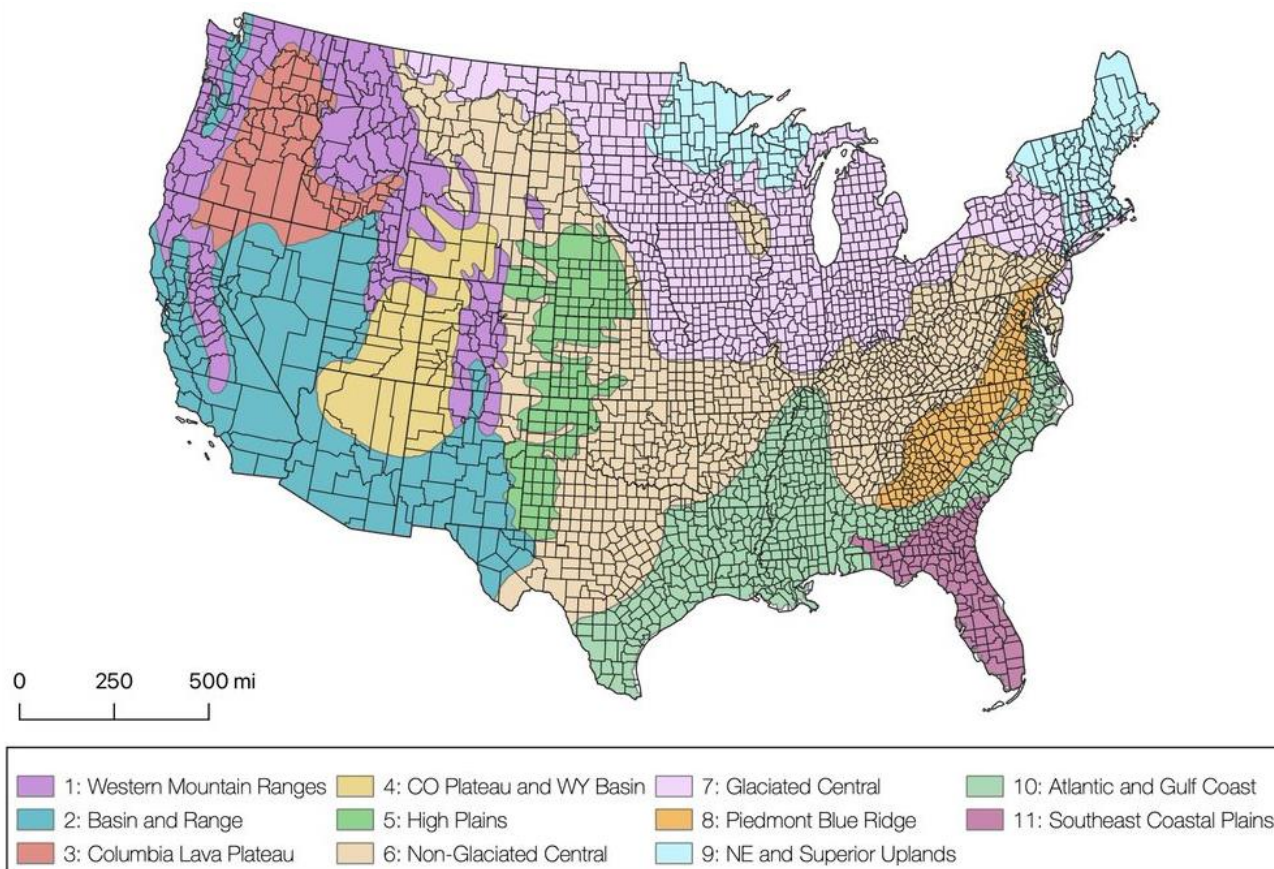

Figure 2: Hydrogeologic regions of the contiguous United States (12: Hawai'i and 13: Alaska not pictured). Data from Clawges and Price (1999).

exception was the Newell, Hopkins, & Bedient (1990) study conducted years before the U.S. EPA published the SSG, though this study did not specifically recommend hydrogeologic region-specific DAFs.

3.2 Summary of findings from Newell, Hopkins, & Bedient (1990)

Utilizing data from the (at the time) newly formed HGDB, Newell, Hopkins, & Bedient (1990) calculated average hydrogeologic parameters (aquifer hydraulic conductivity [K], hydraulic gradient [i], aquifer thickness [d_a], and infiltration rate [I]) for one HGDB defined hydrogeologic region and eight HGDB defined sublevel hydrogeologic settings. Of note, Appendix F of the HGBD refers to "regions" and "settings", while Newell, Hopkins, & Bedient refers to these same areas as hydrogeologic environments (U.S. EPA, 1996c; Newell, Hopkins, & Bedient, 1990). Additionally, Newell, Hopkins, & Bedient (1990) determined a national average of each hydrogeologic parameter. To determine a statistically significant result, Newell, Hopkins, & Bedient (1990) required a minimum of 20 data points for at least one of the hydrogeologic parameters. With Newell, Hopkins, & Bedient (1990) calculated hydrogeologic parameters, this project calculated a mean and median DAF of 1.06 and 1.11, respectively, for each of the hydrogeologic environments. As previously noted, the hydrogeologic environments in Newell, Hopkins, & Bedient did not match the HGDB defined hydrogeologic regions in the HGDB. Additionally, calculations in Newell, Hopkins, & Bedient omitted significant amounts of data in the HGDB. Therefore, this project strived to determine HGDB-defined hydrogeologic region specific DAFs.

3.3 Analyzing hydrogeologic region specific DAFs with the HGDB

Using data from the SSG's (1996c) Appendix F (HGDB raw data), this project compared existing DAFs for 13 hydrogeologic regions included in the HGDB. As seen in Figure 3, hydrogeologic-region-specific DAFs (especially the outliers) vary from each region, as do the number of samples in each region. Scrutinizing the listed half-acre source area DAFs in the SSG Appendix
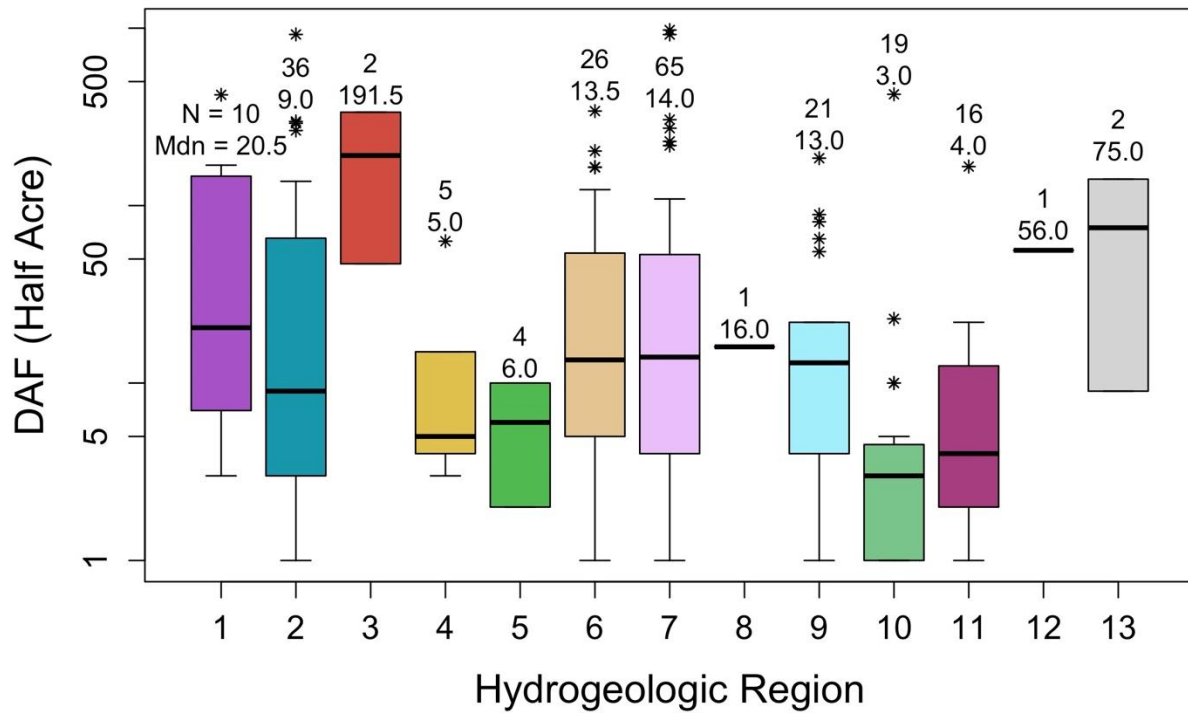

Figure 3: Region-specific DAFs, calculated using HGDB data (regions correspond to Figure 2). N is the number of samples in each region and Mdn is the median DAF. Regions 12 and 13 refer to Hawaii and Alaska, respectively.

F, DAFs range from 1 (the most conservative DAF) to the tens of thousands (and over 100,000 for one site) (U.S. EPA, 1996c). Similarly, more than half of the sites were drawn from regions 6 and 7, while more than half of the regions had 10 or less sample sites. The results displayed in Figures 3 are limited by the dataset present in the HGDB which, to the knowledge of the authors, has not been updated or expanded since 1989. While the current data in the HGDB is not expected to have significantly changed since 1989, expanding the data set would result in more statistically significant calculated DAFs.

4. Update to the infinite source assumption

4.1 The infinite source assumption in the U.S. EPA's 1996 SSG

Other assumptions made by the U.S. EPA's 1996 SSG may lead to the calculation of an overly conservative soil screening level. The development of the SSL-MGW assumed an infinite-source contaminant (U.S. EPA, 1996c); in essence, the SSL-MGW assumes the leachate source diluting into water remains constant over time, helping simplify subsequent mathematical calculations of SSLs in the 1996 guidance (Rixey, Garg, & He, 2000). This assumption is not accurate for most NPL sites; therefore, SSLs calculated using an infinite source assumption may not accurately represent contaminant migration in the environment.

4.2 Correcting the infinite source assumption

While technological and methodological limitations in 1996 necessitated the use of simplified factors in the SSG, various modern modeling tools (e.g., robust high-level programming languages like Python, R, and MATLAB) can simplify the process of determining a correction factor for the infinite source assumption. This project utilized the framework proposed by Rixey Garg, & He, (2000) to program a simple correction factor calculator in Python (Whiteaker, 2023) using both EPA-default and site-specific values (see Figure 4). Since the calculator is based in part on values from the existing retardation factor calculator (U.S. EPA, 2021), the supported chemicals are not comprehensive.

5. Limitations

5.1 Data Accuracy

As the DAF is a function of the factors previously

Figure 4: Finite-source correction factor calculator interface

defined in Equations 4 and 5, the accuracy of the calculated DAFs is directly derived from the accuracy of these hydrogeologic factors. Any discrepancy between a measured factor and the true value would lead to a subsequent error in the DAF. In an ideal survey, all measurements would be conducted by qualified experts with the same types of equipment and measuring techniques. However, the 1989 survey sent out by the API did not require any specific measuring equipment, method, or technique to provide hydrogeologic data. For example, survey responders used a combination of pump tests, slug tests, or laboratory permeability tests to provide hydraulic conductivity (K) values (Newell, Hopkins, & Bedient, 1990). In nearly 75% of responses, surveys determined the infiltration rate (I) via Darcy's Law, while the rest of the responses measured a contaminant plume's movement over a defined period to obtain a value. Other hydrogeologic factor's values were determined by the respondent's best judgement or are literature values.

Newell, Hopkins, & Bedient (1990) noted potential concerns for the respondent's data accuracy when reporting hydraulic conductivity values. The survey asked respondents to estimate the accuracy of their reported hydraulic conductivity values. Over half reported an accuracy of +- 50%, a large variance. Newell, Hopkins, & Bedient (1990) did not note the reason for such a large potential error. Additionally, almost half of those surveyed had less than five years of experience as a professional geologist, hydrologist, or hydrogeologist, suggesting inexperience could lead to potential error, especially as many values were determined from a responder's best judgement.

5.2 Need for GIS when determining hydrogeologic region

Figure 2 displays the different hydrogeologic regions

across the continental United States, with county level political boundaries outlined on the map. For many counties, it is obvious which hydrogeologic region applies. For example, any county within Louisiana would be classified as region 10- Atlantic and Gulf Coast. Alaska and Hawaii are also their own hydrogeologic region. For these locations, someone trying to determine which hydrogeologic specific DAF to use for the calculation of a soil screening level can simply refer to Figure 2.

However, for counties that share two or more hydrogeologic regions, someone trying to determine a generic DAF would need to refer to a GIS tool that can overlay the data in Figure 2 and accept an accurate coordinate. Most states have 2 or more hydrogeologic regions within their borders as well, posing a similar problem. While technology exists to perform this relatively simple check, it cannot be assumed that all local or state level environmental regulators would have access to these tools or the knowledge on how to use them. Additionally, the boundaries themselves are only accurate at the regional scale; even if the tools were available and accessible to all users, there would still be some degree of subjectivity when classifying NPL sites close to region boundaries, which are fairly common.

5.3 *Other factors varying by site location*

Some of the factors in Equation 1 and 2 (θ_w, θ_a, ρ_b, and f_{oc}) vary by site, but the U.S. EPA's (1996b) SSG only uses nationwide default values for these factors. However, this project only focused on the hydrogeologic factors in the DAF.

5.4 *Soil saturated with contaminant(s)*

Equations 1 and 2 (both taken from the 1996 SSG) both assume the soil is unsaturated. For heavily contaminated sites, this assumption may not hold. For organic contaminants, soil saturation implies that the absorption capacity of soil particles, the solubility of the water in soil, and the maximum saturation level of air in soil have all been reached and the soil contaminant will be in a free phase (NJ DEP, 2021). Calculating an SSL may not be appropriate for a site with contaminant saturated soil.

6. Future Work

6.1 *Expanding the dataset with relevant NPL sites*

This project proposes leveraging hydrogeologic data from relevant Superfund sites (those with soil-to-groundwater migration concerns and with the relevant data identified) to expand the data set for calculating DAFs. The U.S. EPA identified remedies for over 1500

Superfund sites across the United States, with 83% of these sites having groundwater remedies and 81% having soil remedies (U.S. EPA, 2023b). This suggests the hydrogeologic characteristic data needed to expand the HGDB already exists, albeit fractured across numerous remediation investigations/feasibility studies (RI/FS). Figure 5 indicates NPL locations in the contiguous United States, grouped by the hydrogeologic regions previously presented in Figure 2. Of note, not every site appearing in Figure 5 has a soil-to-groundwater contamination risk. Many of these sites were listed after the U.S. EPA published the SSG in 1996, meaning the U.S. EPA never utilized the measured hydrogeologic data from newer sites with soil to groundwater contamination concerns. With more usable data, the calculated region-specific DAFs will be more statistically significant than solely using data from the HGDB.

If a SGW contamination risk is present at a Superfund site, it is possible that the responding remediation team conducted a geological and/or hydrogeological study to determine the subsequent leaching to groundwater. This study may include measurements on the site's specific aquifer hydraulic conductivity (K), hydraulic gradient (i), aquifer thickness (d_a) (or aquifer mixing zone depth, [d]), and infiltration rate (I), along with potentially the length of the area of concern parallel to ground water flow (L). This data would be recorded in the RI/FS.

The U.S. EPA website archives RI/FS for Superfund sites, storing the RI/FS as a pdf file. It is feasible to search each of the completed RI/FS to identify the relevant hydrogeologic factors specified in Equation 4 and 5. To speed up the search for relevant data, a pdf reader can run a query to find and highlight specific words (e.g., hydraulic conductivity, hydraulic gradient, etc.) that may produce relevant data. A quality control check will need to filter irrelevant and/or repetitive findings. The usable data can then be matched to the hydrogeologic specific region and added to the HGDB, as described in section 6.1.

6.2 *Replicating or verifying original HGDB survey*

As reviewed in Section 5.1, the HGDB data quality is inconsistent. A new, standardized nationwide survey would provide hydrogeologic data with consistent collection methods. However, a survey of this magnitude would likely be expensive and take significant resources to accomplish. Additionally, the data retrieval explained in Sections 6.1 and 6.2 would likely not have used the same data collection methods as a hypothetical nationwide standardized survey. Alternatively, conducting a few randomized "spot checks" of HGDB or Superfund hydrogeologic data could provide a better understanding on the magnitude of data error in the HGDB.
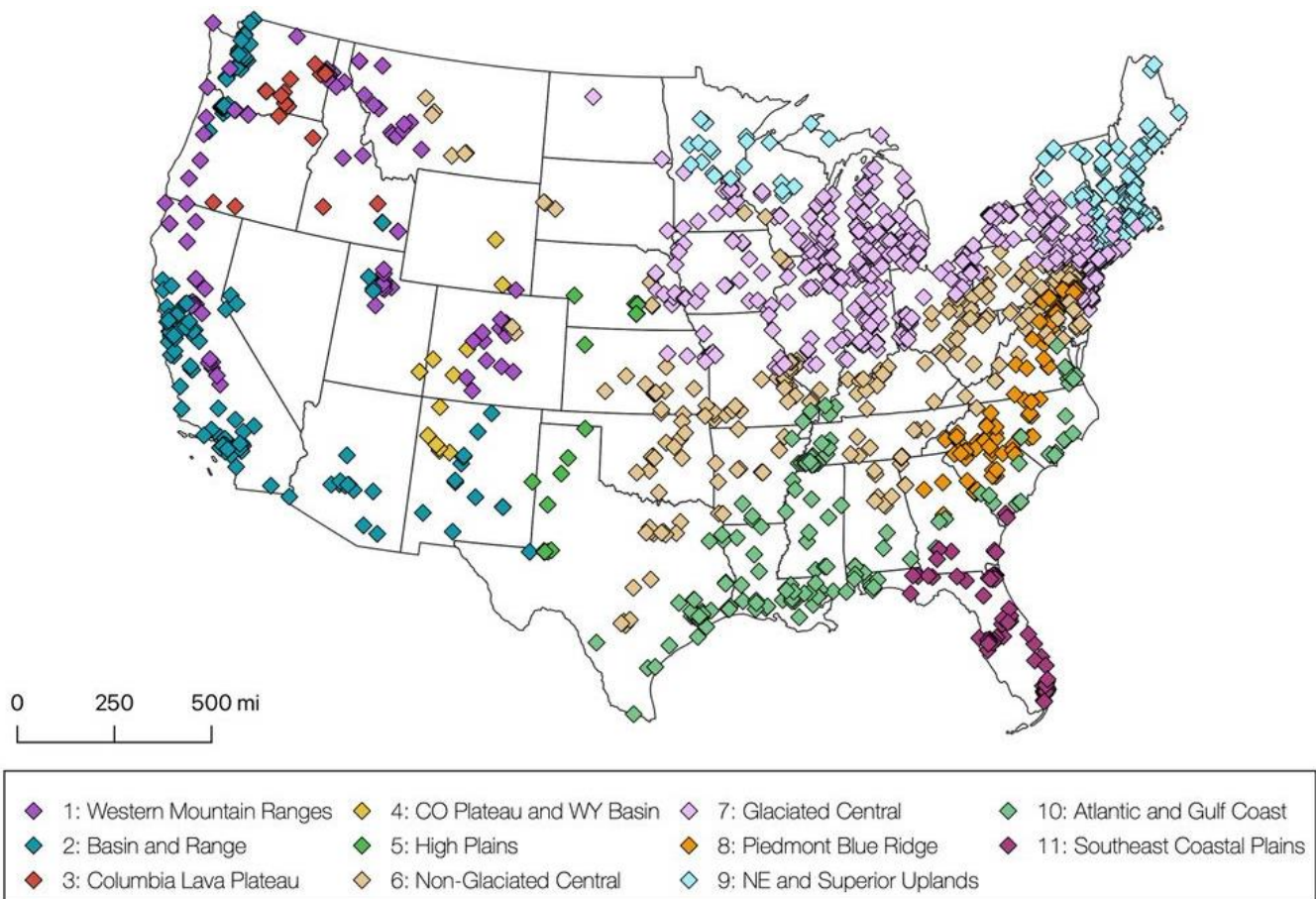
Figure 5: NPL sites by hydrogeologic region (12: Hawaii and 13: Alaska not pictured). Data from Clawges and Price (1999).

6.3 Identifying other sources of hydrogeologic data

While this report reviewed a significant amount of the publicly available literature and data sources relevant to the DAF, there may be other relevant hydrogeological databases that could further expand the HGDB dataset. Other government agencies, such as the U.S. Geological Survey and state environmental agencies, may have access to more routine groundwater monitoring data that could be used to calculate the DAF and other SSL parameters. One potential source of relevant data from industry could come from companies or entities involved in the shale/hydraulic fracturing (fracking) boom in the United States. Fracking involves drilling into the ground to retrieve oil and natural gas trapped in shale and hard rock formations (API, n.d.). Much of the shale boom started in the 2000s, well after the publishing of the U.S. EPA SSG (Yücel & Plante, 2019). It is possible that companies collected relevant hydrogeologic data when conducting fracking operations, which can be used to expand the HGDB. As described in Section 2.1, the API helped develop the survey that became the original HGDB, so there is precedent for working with private companies or organizations to gather hydrogeologic data. One limitation with using hydrogeologic data from hydraulic fracturing sites is that fracking in the continental United States is concentrated in a few basins, with most of the fracking occurring in Texas and North Dakota (Yücel & Plante, 2019). Further efforts could focus on identifying other sources of hydrogeologic data that could expand the HGDB.

6.4 Refining the hydrogeologic regions

While the hydrogeologic regions are certainly an improvement over the current national average, they are still somewhat arbitrary, relatively low resolution, and exclude states and territories outside of the contiguous United States. Future classification systems grounded more solidly in hydrogeology and geomorphology would be easier to generalize and may be more accurate predictors of the DAF and other site-specific data; for example, NPL sites on mountain slopes likely have similar hydraulic conductivities and

gradients, regardless of whether they occur in the Sierra Nevada, Rocky Mountains, or Appalachian Mountains.

In addition to direct inputs for the DAF calculation, efforts to extract data from NPL sites and other sources should thus consider aspects of the site's environmental setting such as topography, geologic history, and climate, among others. Once all this data is compiled into one standardized database, a principal component analysis or similar technique could be used to explore sources of variability between site-specific DAFs. If national maps of relevant hydrogeologic and environmental data were available, it would also be possible to use unsupervised learning methods such as k-means clustering to generate new hydrogeologic regions.

6.5 Publishing a web map for determining an applicable DAF

As described in Section 5.2, a user may have difficulty trying to determine which hydrogeologic region specific DAF to use when determining an SSL. If a hydrogeologic specific (but still "generic") DAF is adopted, the USGS or EPA could publish a map like Figure 2 where a user can input their location to determine the applicable DAF for a site in question. Alternatively, an interactive map where a user "clicks" on a location and the generic DAF is an output would be more helpful than a simple analog map. While this currently does not exist, developing and publishing a map with a GIS software or program (e.g., QGIS, ESRI's ArcGIS, or R using the shiny and terra packages) would not be difficult.

6.6 Updates to the SSG not related to the DAF or infinite source assumption

If NAPLs are present at a site with SGW concerns, additional criteria may be necessary to include in equations within the SSG. Criteria/factors for the SSG should be identified when soil is saturated with a contaminant. Finally, as mentioned in Section 5.4, other factors used in Equations 1 and 2 can typically have site specific values determined. However, EPA default values are usually utilized when calculating SSLs. Using a nationwide default value for a factor that can vary widely from site to site presents similar issues as a nationwide DAF. Further research into replacing single nationwide default values for θ_w, θ_a, ρ_b, and f_{oc} with more location specific (but still generic) values could lead to more accurate SSLs.

7. Conclusion

The U.S. EPA's 1996 Soil Screening Guidance provided a framework for determining soil screening levels that protect human and environmental health. As the guidance is over 25 years old, it is important to revisit the development of this guidance and suggest improvements. The SSG developed calculations for various exposure pathways, including for contaminant movement from SGW. Using hydrogeologic data available at the time, the U.S. EPA's SSG provided nationwide default values for various factors to be used when calculating SSLs for SGW contaminant migration, helping simplify the subsequent calculations. This review assessed aspects of the SGW calculations, specifically, how the current use of a default DAF and how the infinite source assumption affects these calculations. This review highlights the advantages of using hydrogeologic specific DAFs. Expanding on this suggestion, this paper provides hydrogeologic specific DAFs for 11 different regions, as defined in Figure 2. A calculator that can aid in determining a correction factor for the infinite source assumption is detailed as well. Limitations of this research and suggestions for future efforts related to improving the SSG conclude this paper. While significant effort is still required to turn these suggestions into official, published guidance, this paper outlines where the U.S. EPA and other interested agencies can start their efforts.

References

American Petroleum Institute. (n.d.). *Introduction-What is hydraulic fracturing?* Accessed on March 23, 2023, from https://www.api.org/oil-and-natural-gas/energy-primers/hydraulic-fracturing

Clawges, R., & Price, C. (1999). Digital data set describing ground-water regions in the conterminous US. 1.0. U.S. Geological Survey. https://water.usgs.gov/GIS/metadata/usgswrd/XML/ofr99-77_gwregion.xml

Geophysics Research Forum. (1984). *Groundwater contamination*. National Academies Press.

Gilbert, S. (2012). *A small dose of toxicology: The health effects of common chemicals*. (2nd ed.). Healthy World Press.

Heath, R.C. (1984). *Ground water regions of the United States*. U.S. Geological Survey. https://pubs.usgs.gov/wsp/wsp2242/pdf/wsp2242.pdf

The MathWorks Inc. (2023). MATLAB version: 9.13.0 (R2022b), Natick, Massachusetts: The MathWorks Inc. https://www.mathworks.com

Minnesota Pollution Control Agency. (2013, May). *Soil leaching values*. https://www.pca.state.mn.us/sites/default/files/c-r1-04.pdf

National Pesticide Information Center. (2021). *Water solubility*. http://npic.orst.edu/envir/watersol.html

Nevada Division of Environmental Protection. (n.d.). *Further evaluation of soil leaching to groundwater*. https://ndep.nv.gov/uploads/documents/eval-soil-leach-to-gw.pdf

Newell, C.J., Hopkins, L.P., & Bedient, P.B. (1990). A hydrogeologic database for ground-water modeling. *Groundwater*, 28: 703-714. https://doi.org/10.1111/j.1745-6584.1990.tb01986.x

New Jersey Department of Environmental Protection. (2021). S*oil and soil leachate remediation standards for the migration to ground water exposure pathway: Basis and background*. https://www.nj.gov/dep/srp/guidance/rs/bb_migration_gw.pdf

Rixey, W. G., Garg, S., & He, X. (2000). A methodology for accounting for the finite leaching characteristics of contaminated soils and oily wastes on groundwater dilution and attenuation. *Environmental Engineering Science*, 17(3), 117–127. https://doi.org/10.1089/ees.2000.17.117

Šimunek & van Genuchten. (2006). *Contaminant transport in the unsaturated zone theory and modeling*. United States Department of Agriculture. https://www.ars.usda.gov/arsuserfiles/20361500/pdf_pubs/P2169.pdf

Soil Pollution. (2018). Elsevier. https://doi.org/10.1016/C2016-0-02243-X - https://www.sciencedirect.com/topics/earth-and-planetary-sciences/contaminant-transport

United States Environmental Protection Agency. (1996a). *Quick reference fact sheet*. https://semspub.epa.gov/work/HQ/175229.pdf

United States Environmental Protection Agency. (1996b). *Soil screening guidance: User's guide*. https://semspub.epa.gov/work/HQ/175238.pdf

United States Environmental Protection Agency. (1996c). *Technical background document*. https://epa-sdcc.ornl.gov/pdfs/SSG_nonrad_technical.pdf

United States Environmental Protection Agency. (2015). Ground water contamination. https://www.epa.gov/sites/default/files/2015-08/documents/mgwc-gwc1.pdf

United States Environmental Protection Agency (2021). *Retardation Factor Calculator*. Web App. Available at https://www3.epa.gov/ceampubl/learn2model/part-two/onsite/retard.html.

United States Environmental Protection Agency. (2023a). Report on the environment: Ground water. https://www.epa.gov/report-environment/ground-water

United States Environmental Protection Agency. (2023b). Superfund remedy report: 17th edition. https://www.epa.gov/system/files/documents/2023-01/100003149.pdf

United States Geological Survey. (2018). Contamination of groundwater. https://www.usgs.gov/special-topics/water-science-school/science/contamination-groundwater

Yücel, M. & Plante, M. (2019). *GDP gain realized in shale boom's first 10 years*. Federal Reserve Bank of Dallas. https://www.dallasfed.org/research/economics/2019/0820

Whiteaker, S.J. (2023). *Finite Source Correction Factor Calculator*. (Version 1.0). Web App. Available at https://www.growingecology.com/finite-source.