User's Guide for the Rapid Assessment of Vulnerability from Emissions Upstream (RAVEN STREAM) online tool

Corresponding Author: John D. Hader, PhD

Institutional email address: john.hader@empa.ch
Last updated: June 7th, 2024

Table of contents

1 Introduction 3
1.1 Background 3
1.2 Purpose, scope, and disclaimer 3
1.3 What users will need to run RAVEN STREAM: 3
1.4 Overview of analysis 4
2 Running the tool 4
2.1 Data collection and preparation 5
2.2 Data import, parameterization, and running 8
3 Interpreting output 10
3.1 Risk visualisations and prioritisation tables 10
3.2 Uncertainty visualizations 12
3.3 Downloading output and additional supplementary files 14
4 Troubleshooting common errors 17
5 Future modifications and other notes 17
6 References 18

1 Introduction

1.1 Background

The Rapid Assessment of Vulnerability from Emissions Upstream (RAVEN STREAM) tool is intended as an easy-to-use, open-access online tool to help sewage treatment plants (STPs) assess risk from potential spills of industrial chemicals used upstream. RAVEN STREAM was developed as part of the ECORISK2050 project (Welch et al., 2022), as a collaboration between Stockholm University and the Käppala STP in Stockholm, Sweden. A comprehensive presentation of the tool's development, along with a case study of its application to the Käppala STP, is detailed in a paper by (Hader et al., 2023).

1.2 Purpose, scope, and disclaimer

RAVEN STREAM is intended as a preventative screening tool to identify industrial chemical products that are used upstream of an STP that, in the event of a down-the-drain spill, could cause damage to the biologically-active regions of the receiving STP. RAVEN STREAM enables prioritization of chemicals that pose the highest risk, and thus where the STP's resources can be directed for communication with the higher-risk upstream industries for further information gathering and/or implementation of additional spill mitigation measures (if needed). RAVEN STREAM should not be used as an emergency response tool for decision support on whether or not to divert influent sewage away from an STP for release into the environment untreated to prevent possible contamination and damage to the plant. A decision on whether or not do divert influent that is contaminated with a spilled chemical should be based on a comprehensive cost/benefit analysis of possible damage to the plant, long-lasting effects, and the corresponding environmental damage that may result from diverting or not diverting contaminated influent. Such an analysis is beyond the scope of the analysis conducted by RAVEN STREAM. Furthermore, RAVEN STREAM is provided with no warranty or guarantee regarding its ability to identify upstream chemicals that may damage an STP in the event of a chemical spill. The authors assume no liability for any damage caused by decisions made based on the results of the tool, as the results of the tool are provided purely for exploratory purposes. Nonetheless, the authors hope the use of the tool can promote and facilitate discussion between STPs and their upstream industrial clients regarding the safe management of industrial chemicals with respect to possible damage to these crucial buffers between humans' wastewater and the environment.

1.3 What users will need to run RAVEN STREAM:

In order to run the RAVEN STREAM tool, users will need the following data and other pieces of information:

- Data on the storage patterns of industrial chemical products at industrial facilities connected to the STP, including:
    - The mass of chemical product held in stock at a given time (where available)
    - The mass of chemical product used annually
    - The constituent chemicals (i.e., ingredients) of the chemical products, identified by their CAS numbers.

- The volumes of the activated sludge region(s) of the plant, and/or the volume of the anaerobic digestor(s)
- The flow ratios of influent into the separate regions of the plant (e.g., if there are multiple, separated activated sludge regions with differential influent fractions treated)

1.4 Overview of analysis

(Hader et al., 2023) outline the details of how the RAVEN STREAM tool was generated, its input data, assumptions, and an example of its application to a STP in the Stockholm, Sweden metropolitan area. Briefly, the RAVEN STREAM tool calculates the concentration that the constituent chemicals of a bulk industrial chemical product would have in the receiving STP if all of the bulk industrial chemical product held at an upstream facility were involved in a down-the-drain chemical spill. The predicted constituent chemical 'exposure concentrations' in the STP are compared to information on how toxic the constituent chemicals are. This enables a calculation of the possible 'risk' the chemicals pose to the treatment plant. The equation used to calculate constituent chemical risk to the STP is as follows:

$$Constituent\ Chemical\ Risk = \frac{Chemical\ Concentration\ in\ STP}{Chemical\ Toxicity\ Towards\ STP\ Microorganisms}$$

The risk of the bulk industrial chemical product is calculated by summing all of the risks of the individual constituent chemicals in the bulk product. The concentration a constituent chemical would have in the STP is based on the bulk mass of the chemical product the industry holds in stock, the percent contribution of the constituent chemicals in the bulk product, and the volume of the receiving compartment of the STP. The chemical toxicity information is obtained from the webpages of the European Chemicals Agency (ECHA), which provide information on chemical toxicity towards sewage treatment plant microorganisms for thousands of chemicals that are registered for use in the European Union. The toxicity metric that is employed is the predicted no-effect concentration, or PNEC value, which indicates the concentration at and below which no negative effect is expected from the chemical in a STP. This is a threshold value, and provides no information on the expected magnitude of effect that could be expected if the concentration of the chemical in the STP exceeded the PNEC. Rather, if the concentration were to exceed the PNEC (i.e., a risk value of greater than 1), then this indicates an adverse outcome from the chemical on the STP could occur. Conversely, a risk value below 1 indicates an adverse outcome at the STP from the given chemical is unlikely.

In order to more fully understand the details and uncertainties of the chemical risk analysis applied within the RAVEN STREAM tool, users are strongly encouraged to read the (Hader et al., 2023) publication before using RAVEN STREAM, which is available open access through this link: https://pubs.rsc.org/en/content/articlelanding/2023/va/d3va00067b

2 Running the tool

The landing page of the RAVEN STREAM tool is shown in Figure 1, which will be used throughout Sections 2 and 3 to guide the reader through the steps of using the tool. Throughout the text, references to numbers denoted as "(X)" refer to the numbered points in

Figure 1 annotating key buttons or parts of the tool (e.g., "button (2)" refers to the "Load Sample Run" button of the tool).

A sample run has been prepared of the RAVEN STREAM tool to quickly demonstrate the functionality of the tool, not requiring preparing or loading in any new data or parameterisations. To load these default data and STP values, press the "Load Sample Run" button (2) of the tool. This will import a 'fake' dataset saved within the tool, and also populate the necessary parameterisations for a hypothetical STP. Then, click the "Run RAVEN STREAM" button (8), and proceed to Section 3 for information on interpreting results.

Alternatively, instructions on how to prepare and format real input data, how to import these data into the tool, and how to parameterise the tool for a real STP of interest are provided in Sections 2.1 and 2.2 below.

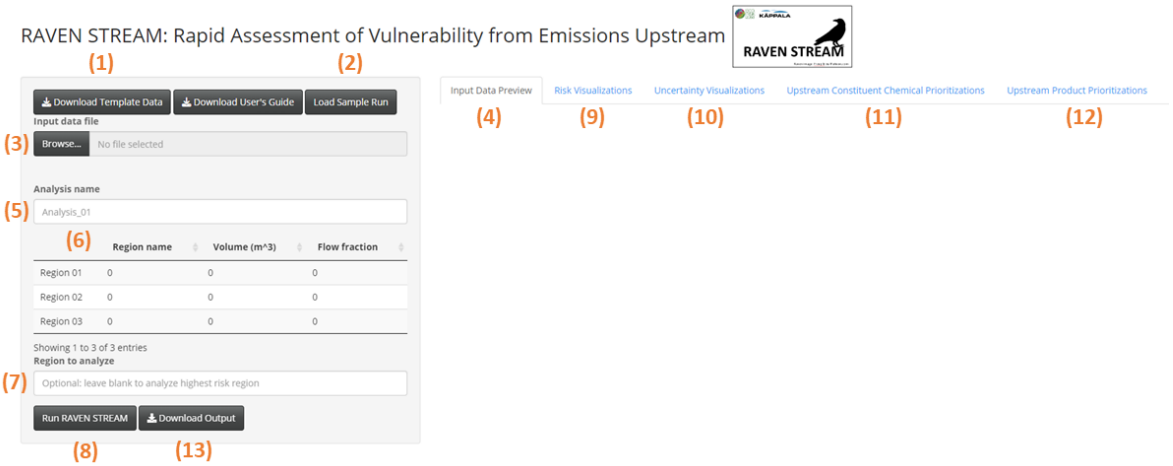


Figure 1. Base landing page of the RAVEN STREAM online tool. Numbers annotate key parts of the tool that are referenced throughout the User's Guide.

2.1 Data collection and preparation

Prior to using the tool, users must collect data from upstream industrial facilities that are connected to the STP on the masses of chemical products held in stock, the masses of chemical products used per year, and the constituent chemical ingredients of the products. The required data is described below in the "Description of data file" sub-section. These data can be input to the template dataset provided for the tool. A description of how to access the template dataset is provided below, followed by a description of the data that must be collected from the connected upstream industries.

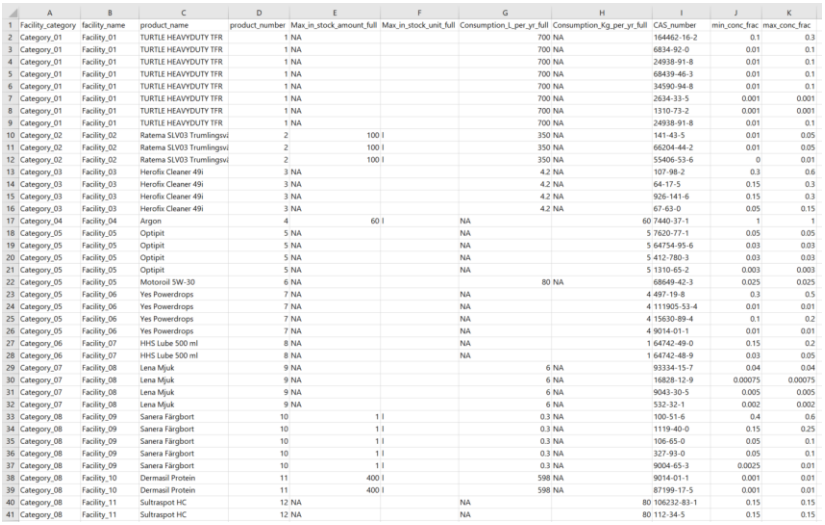
Opening template data while preserving CAS numbers

A template dataset is provided within the RAVEN STREAM tool to help users organise their industrial chemical product usage data. To download this template, press the 'Download Template Data' button (1) (see Figure 1). A dialogue box will open up, allowing the user to download the template data. This file has the '.csv' file extension, however the delimiter between entries in this template file is a semicolon (;). Data can be input by the user to the

.csv file via any method (e.g., using Microsoft Excel, Google Spreadsheets, R, Python, etc.) as long as the data file is formatted according to the tool requirements (see below). Guidance is provided below assuming the data will be organised using Microsoft Excel within the Windows operating system.

Chemical information is provided in the input data file on the basis of Chemical Abstracts Service Registry Numbers (CAS-RN, or simply "CAS numbers"; see this webpage: https://www.cas.org/support/documentation/chemical-substances/faqs). A common issue with using CAS numbers in Microsoft Excel is that the formatting, which involves numbers and dashes, can be misinterpreted by Excel to represent calendar dates. By opening a .csv file with CAS-numbers in Excel, the CAS-numbers can be automatically converted to dates, corrupting the data in the file. To avoid this, it is best to copy the data from the .csv file opened as a text file directly into a newly-opened Excel file and manually set the CAS number column to 'Text'. To do this, start by closing any open instances of Excel. Then, navigate to the Template Data file, right click on the file and select 'Open with' -> 'Notepad'. Within the file, right click and select 'Select all', and then press CTL + C to copy the data to the clipboard.

Then, open Excel, right click Cell A1, and under "Paste options" select the "Keep text only (T)" option to paste the information from the template data file into Excel. Then, in the top ribbon of Excel, navigate to Data -> Text to Columns. In the dialogue box that opens, leave the "Original data type" option as "Delimited", then click "Next". In the next page, under "Delimiters", uncheck "Tab" and instead check the "Semicolon" option, then click "Next". In the next page, under "Data preview", scroll the data preview window to the right to see the 'CAS_number' column (which will be the third to last column). Click this column to highlight it, and under the 'Column data format' section at the top of the dialogue box, select 'Text'. Click 'Finish', and the data will be populated into the columns of the Excel file (as in Figure 2). This procedure forced Excel to recognize the CAS_number column as being in text format, and avoided the automatic conversion of the CAS numbers to Date format. All other data columns should have been automatically recognised correctly by Excel.



	A	B	C	D	E	F	G	H	I	J	K
1	Facility_category	facility_name	product_name	product_number	Max_in_stock_amount_full	Max_in_stock_unit_full	Consumption_L_per_yr_full	Consumption_Kg_per_yr_full	CAS_number	min_conc_frac	max_conc_frac
2	Category_01	Facility_01	TURTLE HEAVYDUTY TFR	1	NA		700	NA	164462-16-2	0.1	0.3
3	Category_01	Facility_01	TURTLE HEAVYDUTY TFR	1	NA		700	NA	6834-92-0	0.01	0.1
4	Category_01	Facility_01	TURTLE HEAVYDUTY TFR	1	NA		700	NA	24938-91-8	0.01	0.1
5	Category_01	Facility_01	TURTLE HEAVYDUTY TFR	1	NA		700	NA	68439-46-3	0.01	0.1
6	Category_01	Facility_01	TURTLE HEAVYDUTY TFR	1	NA		700	NA	34590-94-8	0.01	0.1
7	Category_01	Facility_01	TURTLE HEAVYDUTY TFR	1	NA		700	NA	2634-33-5	0.001	0.001
8	Category_01	Facility_01	TURTLE HEAVYDUTY TFR	1	NA		700	NA	1310-73-2	0.001	0.001
9	Category_01	Facility_01	TURTLE HEAVYDUTY TFR	1	NA		700	NA	24938-91-8	0.01	0.1
10	Category_02	Facility_02	Ratema SLV03 Trumlingsvi	2	100	l	350	NA	141-43-5	0.01	0.05
11	Category_02	Facility_02	Ratema SLV03 Trumlingsvi	2	100	l	350	NA	66204-44-2	0.01	0.05
12	Category_02	Facility_02	Ratema SLV03 Trumlingsvi	2	100	l	350	NA	55406-53-6	0	0.01
13	Category_03	Facility_03	Herofix Cleaner 49i	3	NA		4.2	NA	107-98-2	0.3	0.6
14	Category_03	Facility_03	Herofix Cleaner 49i	3	NA		4.2	NA	64-17-5	0.15	0.3
15	Category_03	Facility_03	Herofix Cleaner 49i	3	NA		4.2	NA	926-141-6	0.15	0.3
16	Category_03	Facility_03	Herofix Cleaner 49i	3	NA		4.2	NA	67-63-0	0.05	0.15
17	Category_04	Facility_04	Argon	4	60	l	NA		60 7440-37-1	1	1
18	Category_05	Facility_05	Optipit	5	NA		NA		5 7620-77-1	0.05	0.05
19	Category_05	Facility_05	Optipit	5	NA		NA		5 64754-85-6	0.03	0.03
20	Category_05	Facility_05	Optipit	5	NA		NA		5 412-780-3	0.03	0.03
21	Category_05	Facility_05	Optipit	5	NA		NA		5 1310-65-2	0.003	0.003
22	Category_05	Facility_05	Motoroil 5W-30	6	NA			80 NA	68649-42-3	0.025	0.025
23	Category_05	Facility_06	Yes Powerdrops	7	NA		NA		4 497-19-8	0.3	0.5
24	Category_05	Facility_06	Yes Powerdrops	7	NA		NA		4 111905-53-4	0.01	0.01
25	Category_05	Facility_06	Yes Powerdrops	7	NA		NA		4 15630-89-4	0.1	0.2
26	Category_05	Facility_06	Yes Powerdrops	7	NA		NA		4 9014-01-1	0.01	0.01
27	Category_06	Facility_07	HHS Lube 500 ml	8	NA		NA		1 64742-49-0	0.15	0.2
28	Category_06	Facility_07	HHS Lube 500 ml	8	NA		NA		1 64742-48-9	0.03	0.05
29	Category_07	Facility_08	Lena Mjuk	9	NA			6 NA	93334-15-7	0.04	0.04
30	Category_07	Facility_08	Lena Mjuk	9	NA			6 NA	16828-12-9	0.00075	0.00075
31	Category_07	Facility_08	Lena Mjuk	9	NA			6 NA	9043-30-5	0.005	0.005
32	Category_07	Facility_08	Lena Mjuk	9	NA			6 NA	532-32-1	0.002	0.002
33	Category_08	Facility_09	Sanera Färgbort	10	1	l		0.3 NA	100-51-6	0.4	0.6
34	Category_08	Facility_09	Sanera Färgbort	10	1	l		0.3 NA	1119-40-0	0.15	0.25
35	Category_08	Facility_09	Sanera Färgbort	10	1	l		0.3 NA	106-65-0	0.05	0.1
36	Category_08	Facility_09	Sanera Färgbort	10	1	l		0.3 NA	327-93-0	0.05	0.1
37	Category_08	Facility_09	Sanera Färgbort	10	1	l		0.3 NA	9004-65-3	0.0025	0.01
38	Category_08	Facility_10	Dermasil Protein	11	400	l		598 NA	9014-01-1	0.001	0.01
39	Category_08	Facility_10	Dermasil Protein	11	400	l		598 NA	87199-17-5	0.001	0.01
40	Category_08	Facility_11	Sultraspot HC	12	NA		NA		80 106232-83-1	0.15	0.15
41	Category_08	Facility_11	Sultraspot HC	12	NA		NA		80 112-34-5	0.15	0.15

Figure 2. Template data file loaded into Microsoft Excel, with CAS numbers correctly read in avoiding accidental conversion to Date format.

Description of data file:

A description of the data and required format in each of the columns is provided in Table 1. Data input to the RAVEN STREAM tool must be formatted so data from the various facilities, chemical products, and their constituent chemical ingredients are input down the rows. Each row of data must contain data for a unique constituent chemical ingredient in a given chemical product. Therefore, information on the facility category and name, product name and number, max-in-stock amount and unit, and consumption per year amount are repeated down the rows of data for each unique constituent chemical ingredient in a given chemical product (refer also to Figure 2).

Table 1. RAVEN STREAM input data file column formatting requirements.

Column name	Meaning	Required format	Note
Facility_category	Category of the industrial facility (e.g., hospital, airport, car wash)	Text	Should be provided by the reporting upstream industrial facility
facility_name	Name of the industrial facility	Text	
product_name	Name of the industrial chemical product	Text	
product_number	Number assigned to the industrial chemical product	Numeric	Should be assigned by the STP operator running RAVEN STREAM
Max_in_stock_amount_full	Maximum amount of the chemical product held in stock at the facility at any given time	Numeric	Missing values indicated with "NA"
Max_in_stock_unit_full	Unit of measure corresponding to the Max_in_stock_amount_full column	Text (can be 'l' or 'kg' for litres or kilograms, respectively).	Leave blank if missing value
Consumption_L_per_yr_full	Average used amount of the chemical product at the industrial facility in one year, in litres.	Numeric	Only provide data for either one or neither of these columns (leave one or both as "NA", respectively).
Consumption_Kg_per_yr_full	Average used amount of the chemical product at the industrial facility in one year, in kilograms	Numeric	
CAS_number	The CAS number of the constituent chemical ingredient in the bulk chemical product.	CAS number format. See this link to CAS.org	Chemicals with improperly formatted

			CAS numbers are not analysed in the tool.
min_conc_frac	The minimum fraction contribution of the constituent chemical ingredient to the bulk chemical product.	Numeric	The values in these two columns may be the same for a given constituent chemical.
max_conc_frac	The maximum fraction contribution of the constituent chemical ingredient to the bulk chemical product.	Numeric	

Regarding the data columns whose labels' start with "Max_in_stock" and "Consumption" (i.e., columns E through H), (Hader et al., 2023) provide additional information on the need for these two different pieces of information. In brief, industries may not have explicit information regarding the maximum amount of chemical held in stock at a given time, but rather only information on how much of the chemical product they use throughout the year. To estimate the max-in-stock masses of the chemical products that do not have max-in-stock data provided, an extrapolation algorithm is calculated within RAVEN STREAM based on chemical products that have both the max-in-stock and yearly usage data provided. This enables estimation of the max-in-stock masses of chemical products given only their yearly usage data. Thus, whenever possible, it is best to report both the max-in-stock as well as the yearly usage data of a chemical product from the upstream industries. If neither of these pieces of information are provided, risk to the receiving STP cannot be estimated for this chemical product. Furthermore, if a sufficient number of chemical products do not have both max-in-stock and yearly usage data provided, then the extrapolation algorithm within RAVEN STREAM will not be very robust for making max-in-stock predictions for chemical products with only yearly usage data. See Section 2.1.2 of (Hader et al., 2023) for more information, as well as Section 5 of this document for additional discussion of uncertainty/future work in this area of the tool.

Data from an upstream industrial chemical product use survey can be directly input to the Template data file using Microsoft Excel, or this data template file can simply be used as the structure for organising a survey data file in another software (e.g., R or Python). Once all of the upstream chemical usage data is input to the data file, save the file as a .csv file. Note that non-latin characters (e.g., ä, ö, ü) may cause data reading/display issues within RAVEN STREAM, and users should avoid using these characters in the input data as the tool may not run correctly if they do. The file should be saved as a .csv file, but can have comma or semicolon separators (note that Excel will by default place semicolons or commas between entries differently depending on the country).

2.2 Data import, parameterization, and running

Once the input data file has been prepared, in the RAVEN STREAM tool click the 'Browse' button (3). A dialogue box will appear, allowing the user to navigate to the input data file. Select the input data file and click 'Open'. The data will be imported into the tool, and a preview of the data will display to the right of the screen (see Figure 3; ensure the 'Input Data Preview' tab (4) is selected along the top of the tool). Briefly ensure that data within each column of the preview matches the expected data that was input to the tool.

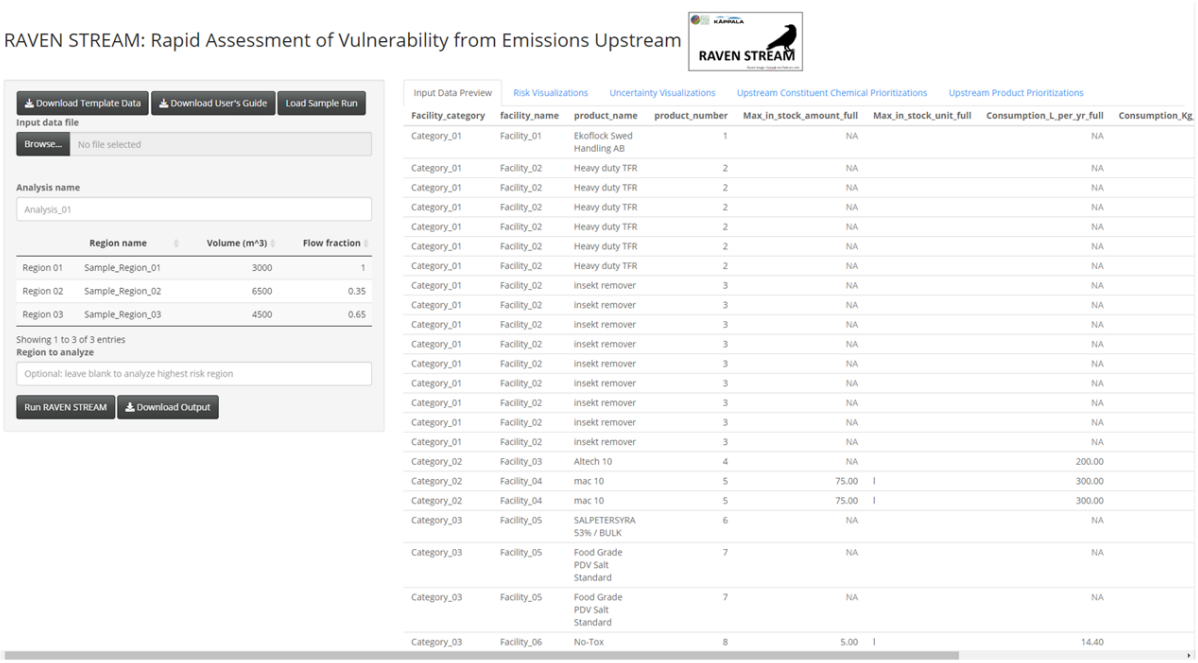


Figure 3. The RAVEN STREAM online tool showing a preview of the data that has been input to the tool.

Under the 'Analysis name' text box (5), enter a name for the analysis (e.g., "Test_01"). It is recommended that no spaces be used between characters or words in this name, rather the underscore or dash character should be used.

Below the 'Analysis name' text box is a table where information on the region or regions of interest in the sewage treatment plant are entered (6). The different regions that may be of interest in the STP could be, for example, the activated sludge region or anaerobic digestor(s) in the plant. For each region of interest, enter the name, volume (in cubic meters), and the "Flow fraction" of the region in the corresponding columns of the input table. The "Flow fraction" column is used to define how much of the influent wastewater (and therefore spilled chemical) would contaminate that region of the plant. For example, in the (Hader et al., 2023) study, the Käppala STP had three regions of the plant of interest: an activated sludge region that was physically divided into a smaller and larger region, and an anaerobic digestor region. 36% of the influent wastewater was treated in the smaller region of the activated sludge compartment, while 64% was treated in the larger region. Thus, the "Flow fractions" for these regions would be input as 0.36 and 0.64, respectively. For the anaerobic digestors, 100% of the sludge was diverted to this region, so the "Flow fraction" was set to 1 (which would simulate a scenario where all of the chemicals would be present in the sludge, rather than liquids). While only three regions can be defined at a time in the current version of RAVEN STREAM, unused rows can be left at their default values and these will be ignored in the rest

of the analysis (as long as at least one valid row is provided). See (Hader et al., 2023) for more details on flow fractions, and assumptions related to spilled chemical partitioning to liquids versus solids in the influent wastewater.

In the "Region to analyse" text box (7), leave this blank if you would like to have risks calculated and displayed for the region of the STP that has the highest risks (based on the corresponding volume and influent flow fraction). Alternatively, enter the name of the region that you would like to have the risk values generated for, regardless of whether it is the highest risk region.

Finally, press the 'Run RAVEN STREAM' button (8). A progress bar will display in the bottom-right of the tool, indicating the overall progress of the calculations. The tool should only take a few seconds to run if the input dataset contains between hundreds and thousands of chemical products. Once the tool has finished processing, the progress bar will briefly display "Processing complete; explore other tabs for results", and will then disappear.

3 Interpreting output

3.1 Risk visualisations and prioritisation tables

Navigate to the 'Risk Visualisations' tab (9). Two figures will be displayed, the first shows risks to the STP from the constituent chemicals in the input data, and the second shows 'additive' risks based on the summed risks of all constituent chemicals in a given bulk industrial chemical product (Figure 4). The first figure (panel A) shows the risks from the constituent chemicals, with the x-axis denoting the predicted concentration that the chemicals will have in the event of a down-the-drain spill of the bulk chemical product, based on their masses, the volume of the region of interest in the STP, and the influent flow fraction. The y-axis denotes the calculated risk that the constituent chemicals pose to the plant, based on their predicted concentrations and their toxicity (i.e., PNEC). Pink squares on the plot denote instances where the bulk chemical product (which is made up of the constituent chemicals) had its max-in-stock mass estimated based on its yearly usage data provided by the industry, while green circles indicate that this max-in-stock mass for the chemical product was directly reported by the industries (see Section 2.1 for details). The second figure (panel B of Figure 4) displays the additive risks of the constituent chemicals present in each of the bulk chemical products. In each of these risk visualization figures, a "Potential Risk = 1" line is denoted along the y-axis. A risk value of 1 or below indicates that the given constituent chemical or product is unlikely to have an adverse impact on the sewage treatment plant, based on the provided data. However, a risk value above 1 indicates that an adverse impact is possible from the constituent chemical or bulk product.

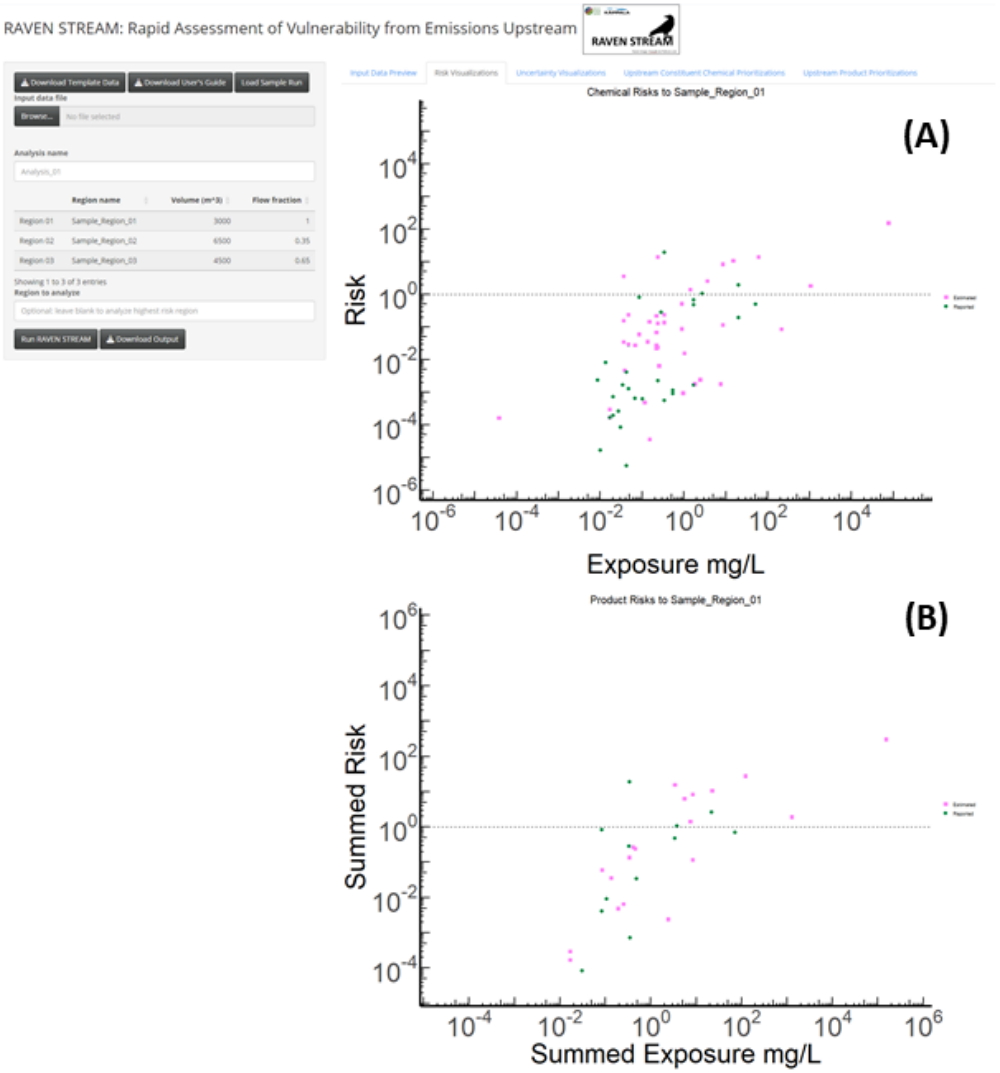


Figure 4. Figures displayed under the "Risk Visualisations" tab of the RAVEN STREAM online tool after processing of the Sample data set for the (panel A) constituent chemicals and (panel B) bulk chemical products.

The information in these two risk visualisation figures is also presented in the form of prioritization tables, which can be accessed under the "Upstream Constituent Chemical Prioritizations" and "Upstream Product Prioritizations" tabs (11 & 12). These tables provide information on the upstream industrial facility, the names of the bulk chemical products, the CAS numbers of the chemicals, and the risk values for the plotted points from the figures in the Risk Visualizations tab (see Figure 5). The number of chemicals or bulk chemical products for which information is displayed at one time can be modified by changing the "Show 10 entries" drop-down box above the table (see annotation on Figure 5). Each column can also be ranked in ascending or descending order, or alphabetical order, by clicking the up/down arrows to the right of each column header.

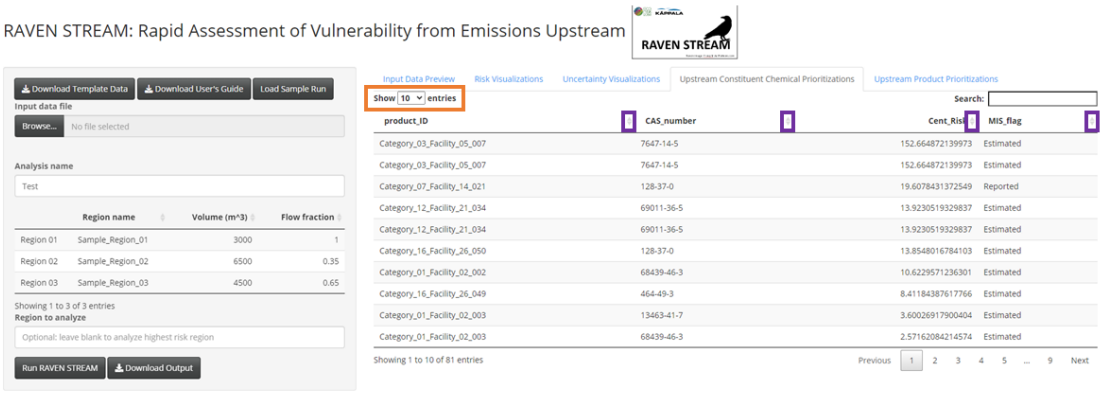


Figure 5. Constituent chemical prioritisation table, based on the analysis of the Sample dataset in the RAVEN STREAM online tool. Annotations show options to increase the number of constituent chemicals displayed on the page at one time (orange box) and the option to sort each column in ascending or descending order (purple boxes). A similar prioritisation table is also provided in the tool for the bulk industrial products (i.e., which are based on the per-product sum of the constituent chemical risks shown here), and can be found in the tab labelled "Upstream Product Prioritizations".

3.2 Uncertainty visualizations

Visualizations of uncertainty in the chemical risk estimates are provided under the "Uncertainty Visualizations" tab (10). The first figure in this tab (panel A, Figure 6) shows the uncertainty in the risk values of the constituent chemicals, considering uncertainty in the toxicity of the chemicals, the uncertainty in the percent contribution of the constituent chemicals to the bulk chemical product, and (if applicable) uncertainty in the amount of bulk chemical product held in stock upstream. The second figure (panel B) displays a semi-quantitative estimate of uncertainty for the risk associated with the bulk chemical products. Risk of a bulk product is calculated by assuming additive toxicity of the constituent chemicals. Such chemical risk though can only be calculated for chemicals that have corresponding toxicity information in the internal database of toxicity derived from the European Chemicals Agency (see section 2.2 of (Hader et al., 2023)). Since not all chemicals have toxicity data reported in this database, some fraction of constituent chemicals under analysis are likely to not be represented in the estimate of risk for the bulk chemical product. This figure displays the fraction of the constituent chemicals that do have toxicity values available.

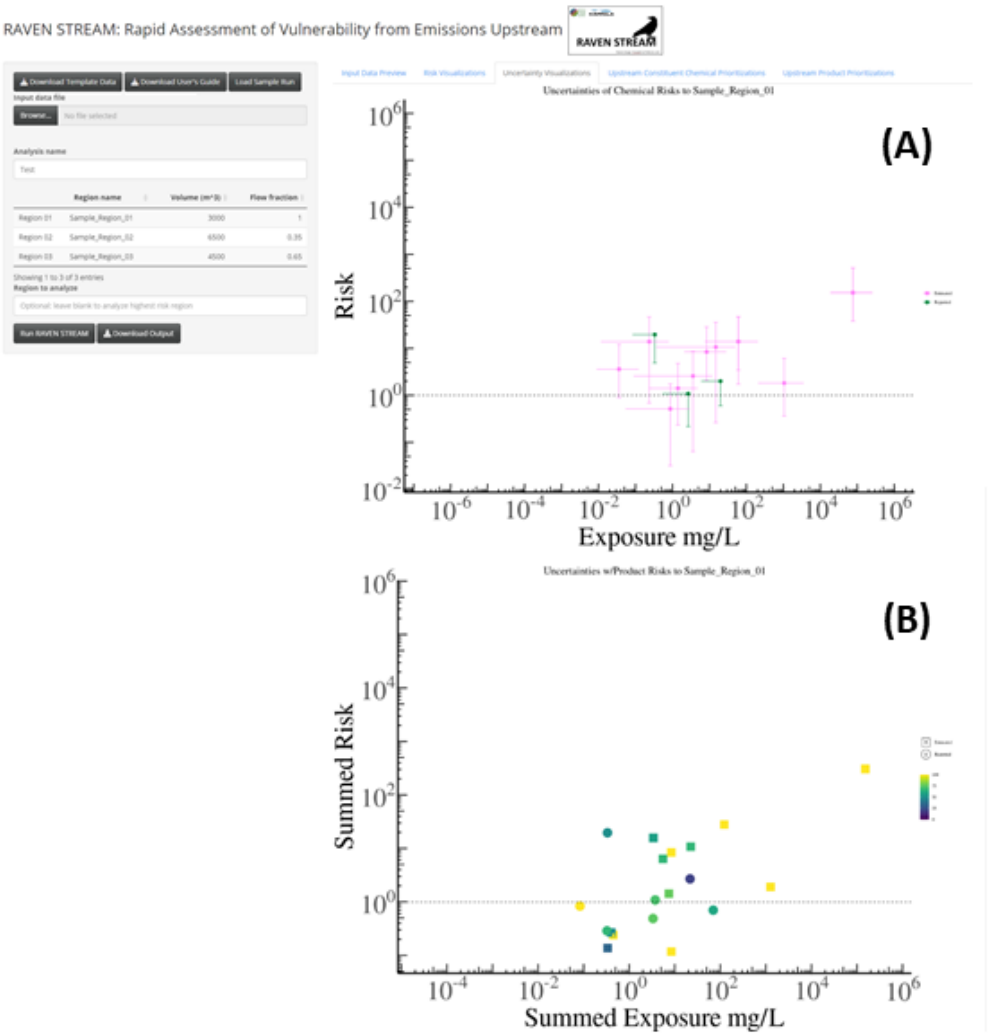


Figure 6. Figures shown under the "Uncertainty Visualisations" tab of the RAVEN STREAM online tool. Panel A shows the risks and uncertainties from the constituent chemicals (points correspond to the same values shown in Figure 4, panel A), while error bars denote uncertainty in these values. Panel B shows a semi-quantitative estimate of uncerntainty in the risk values of the bulk chemical products, whereby shading indicates the fraction of constituent chemicals for which toxicity data was available per chemical product (since risk could not be calculated for constituent chemicals without toxicity data). Note that the bounds on the X and Y axes in both panels are not the same as those in Figure 4, to focus in on and enable more detailed analysis of chemicals with risk values near or above 1.

The risk visualisation figures, the corresponding prioritisation tables, and the uncertainty figures can be used to identify the upstream facilities that the STP could delegate additional resources to for further information gathering, and potentially discussions regarding supplementary spill risk mitigation strategies that may need to be implemented to ensure the highest risk industrial chemical products are not involved in a down-the-drain spill. While the risk visualisation and prioritisation tables provide a deterministic estimate of risk to the plant, it is also important to consider, qualitatively and quantitatively, the associated uncertainties in the Uncertainty Visualization tab. For example, constituent chemicals with a deterministic

risk value below the threshold of 1, but with a large uncertainty range that far exceeds one on the upper end, should potentially be prioritized for additional information gathering over a chemical with a risk value just above one but a low uncertainty. Similarly, a bulk chemical product whose additive risk value is slightly below one but with a relatively small percentage (e.g., less than 75%) of the chemicals reflected in the toxicity estimate should potentially be prioritized for additional information gathering over a bulk chemical product with a risk value slightly over 1 but with 100% of its constituent chemicals having toxicity values available for risk calculation.

3.3 Downloading output and additional supplementary files

All of the figures and prioritization tables generated and displayed within the RAVEN STREAM tool can be downloaded for further use outside of the tool by pressing the "Download Output" button (13). The screen will then turn grey, and a dialogue box will appear at the top stating "Preparing RAVEN STREAM output for download". At this point, the tool is processing all of the images and tables into a zip file for download, and may take a minute or more to complete. Once processing has finished, the output zip file will either download automatically to the user's computer, or a dialogue box will appear allowing the user to download it to a specific directory. Once unzipped, the user can then access all of the figures and prioritization tables generated from the RAVEN STREAM analysis, as well as several other figures and .csv files with supplementary information from the analysis. Table 2 contains the names of all of these files, along with additional information needed to interpret them:

Table 2. Names and descriptions of the contents of the files contained in the .zip file downloaded from the RAVEN STREAM online tool after an analysis. Note that in the File name column, the "[Region]" will have the name of the STP region for which the risk analysis was conducted.

File name	File type	Contents	Notes
Analysis summary statistics [Region]	.csv	Summary counts of various metrics from the overall analysis.	Metric definitions provided in Table 3
Constituent Chemicals Risk Plot [Region]	.png	Figure showing the estimated exposure and risk that the constituent chemicals pose to the region of the STP under analysis.	From "Risk Visualizations" tab of tool
Constituent Chemicals Uncertainty Plot [Region]	.png	Figure showing the estimated exposure and risk, along with uncertainty bounds, of the constituent chemicals (focused on chemicals above and around the Risk = 1 line).	From "Uncertainty Visualizations" tab of tool
Counts by Categories [Region]	.csv	Summary counts of various metrics (mainly related to the constituent chemicals and products), on the basis of industrial facility category type.	Metric definitions provided in Table 4
Max-in-stock adjustment CDF [Region]	.png	Cumulative distribution of the number of months of chemical product held in stock across industries who provided data on their maximum amount held in stock and yearly usage of chemical products.	The 25th, 50th, and 75th percentiles of this distribution are used as a conversion factor for the low, central, and high estimates of max-in-stock amounts when industries only report their yearly

			usage (see (Hader et al., 2023) for more details).
Max-in-stock adjustment histogram [Region]	.png	Histogram of the same information presented in the "Max-in-stock adjustment CDF" figure.	Note that all values above 30 were put into this highest bin, to enable easier viewing of the figure.
Max-in-stock adjustment percentiles [Region]	.csv	Selected percentiles from the distribution of the conversion factor used to estimate the maximum amount of chemical product held in stock at an industrial facility, based on its yearly usage data.	See (Hader et al., 2023) for more details
Prioritized Constituent Chemicals Risks [Region]	.csv	Prioritization table of constituent chemicals. "MIS_flag" column reports the "Max in stock" flag, i.e., whether the mass of the product was "Estimated" or "Reported". The "Low_Exp", "Cent_Exp", "High_Exp" and "Low_Risk", "Cent_Risk", "High_Risk" columns denote the low, central tendency, and high-end estimates of exposure and risk (respectively), considering uncertainties (see (Hader et al., 2023) for more details).	From "Upstream Constituent Chemical Prioritizations" tab of tool
Prioritized Product Risks [Region]	.csv	Prioritization table of chemical products. Column definitions similar to table for prioritisation of constituent chemicals.	From "Upstream Product Prioritizations" tab of tool
Product Fractions Plot [Region]	.png	Figure showing the percent of constituent chemicals per product for which toxicity data were available (and thus could have risk estimates calculated for). This is a semi-quantitative estimate of the uncertainty associated with the risk of the chemical products (focused on chemicals above and around the Risk = 1 line).	From "Uncertainty Visualizations" tab of tool
Products Risk Plot [Region]	.png	Figure showing the estimated exposure and risk that the bulk chemical products pose to the region of the STP under analysis.	From "Risk Visualizations" tab of tool

Table 3. Definitions of the metrics in the rows of the "Analysis summary statistics" file output from the RAVEN STREAM online tool.

Metric	Definition
Num_unique_facilities	Number of upstream industrial facilities represented in the analysed data.
Num_unique_prods	Number of unique chemical products that are represented in the analysed data (based on product name, repeats counted as 1).
Num_unique_chems	Number of unique constituent chemicals represented across all chemical products (based on CAS number, repeats counted as 1).
Num_valid_PNEC	Number of unique chemicals (based on CAS number) in the chemical toxicity database for which valid toxicity information is available.
Num_toxic_PNEC	Number of unique chemicals (based on CAS number) in the chemical toxicity database that indicate possible toxicity towards sewage treatment plants (i.e., had a PNEC value).
Num_non_toxic_PNEC	Number of unique chemicals (based on CAS number) in the chemical toxicity database for which hazard towards sewage treatment plants was unlikely or not identified.

Num_range_PNEC	Number of unique chemicals (based on CAS number) in the chemical toxicity database for which a range of PNEC values was available (as opposed to a single value).
Num_chems_risk_1_or_over	Number of constituent chemicals in the analysed upstream data for which the risk value was greater than or equal to 1 (based on the central value, i.e., not the low or high risk estimate).
Num_prods_risk_1_or_over	Number of chemical products in the analysed upstream data for which the risk value was greater than or equal to 1 (based on the central value, i.e., not the low or high risk estimate).
Num_uncert_chems_high_risk_1_or_over	Number of constituent chemicals in the analysed upstream data for which the high-end risk value was greater than or equal to 1 (i.e., to provide a measure of uncertainty for the "Num_chems_risk_1_or_over" value, which is just based on the central value).
Num_uncert_prods	Number of chemical products in the upstream chemical data for which risk values were calculated
Num_uncert_prods_under_75	Number of chemical products in the upstream chemical data for which fewer than 75% of the constituent chemicals in the product had PNEC toxicity values* provided.
Num_uncert_prods_75_and_over	Number of chemical products in the upstream chemical data for which 75% or more of the constituent chemicals in the product had PNEC toxicity values* provided.
Num_uncert_prods_0.1_to_1.0	Number of chemical products in the upstream chemical data for which the central risk value is greater than or equal to 0.1, but less than 1 (i.e., their risk value is relatively close to the threshold of concern).
Num_uncert_prods_0.1_to_1.0_under_100	Number of chemical products in the upstream chemical data for which the central risk value is greater than or equal to 0.1, but less than 1, and fewer than 100% of the constituent chemicals have valid PNEC toxicity values* provided.

* "valid PNEC toxicity values" corresponds to both valid numeric PNEC values for the constituent chemicals, as well as instances where constituent chemicals were documented to have hazards unlikely or not identified. Instances where no or obviously erroneous information on chemical toxicity was provided (i.e., a PNEC value of 0) are considered as missing PNEC toxicity information.

Table 4. Definitions of the columns in the "Counts by Categories" file output from the RAVEN STREAM online tool.

Column label	Definition
Facility category	Industrial facility category.
Prod_num	The number of chemical products used by each facility category type (if the same chemical product is used at two facilities in a given category, this would count as two, but if the same product has multiple entries for one facility, this would only be counted as one).
Num_prods_max	*Number of chemical products (i.e., Prod_num) in the facility category for which only a max-in-stock amount of the product was reported.
Num_prods_yrl	*Number of chemical products (i.e., Prod_num) in the facility category for which only a yearly-usage amount of the product was reported.
Num_prods_both	*Number of chemical products (i.e., Prod_num) in the facility category for which both a yearly-usage amount and a max-in-stock amount of the product was reported.
Chem_num	The total number of constituent chemicals in the analysed dataset per facility category (repeated chemical CAS numbers across and within products are all counted uniquely).
Toxic_PNEC_num	Number of the chemicals in the category (i.e., Chem_num) which had documented information indicating they were toxic to sewage treatment plants (i.e., not a missing PNEC value and not a documented indication of hazard not found/unlikely).

Non_toxic_PNEC_num	Number of the chemicals in the category (i.e., Chem_num) which had documented information indicating a hazard was not found or was unlikely towards sewage treatment plants.
Missing_PNEC_num	Number of the chemicals in the category (i.e., Chem_num) which had missing PNEC toxicity information towards sewage treatment plants.

* Note that since the number of products is based on unique product names per facility (i.e., a product reported twice at the same facility with the same name will be counted once), only the first instance of these products would be reflected in these counts.

4 Troubleshooting common errors

Several common errors that users make while conducting an analysis with the RAVEN STREAM online tool have been identified. These errors, and how to fix them, are provided here to help users avoid some of these issues:

- CAS numbers converting to Date format in Excel:
  - As discussed in Section 2.1, opening an Excel (.xlsx or .xls) file that contains CAS numbers may, depending on how the file has been formatted, automatically convert some CAS numbers to a Date format within Excel. To avoid this, instructions are provided in Section 2.1 of this User's Guide.
- Non-latin characters in input data file:
  - Users should exclude non-latin characters (e.g., ä, ö, ü) from their input data files. Otherwise, the tool may have issues displaying the input data in the 'Data preview' tab due to the way that characters are encoded between the input file and the tool within the R programming language, which could cause the tool to crash.

Users who identify any errors in the tool itself or encounter any other issues are strongly encouraged to report these to the corresponding author (see title page for contact information).

5 Future modifications and other notes

As of the public release of the RAVEN STREAM online tool, there have been fewer than a dozen users of the tool. While particular attention has been put to making the tool as straight-forward to use as possible, there will undoubtedly be areas where the tool can be improved. Furthermore, due to time constraints around publishing the tool, there are several aspects of the workflow of the tool known to the developers that not as efficient as possible, and are key targets for improvement in a future version of the tool. Some of these planned future improvements are:

- Reformatting of input data: Namely, the units required for the Max-in-stock and yearly usage data of chemical products are not intuitive. A future update will streamline data formatting requirements to only allow kilograms, and will utilise only one column per metric.
- Allow setting of a user-defined number of compartments to be analysed.
- Revise/harmonise some of the constituent chemical and chemical product counting: These are primarily intended to be used as summary statistics/data checks for the user's data. Some of the counting algorithms could be made more consistent (e.g.,

currently products with the same name reported twice at a given facility are not counted twice; see Table 4).

- Develop a backup method for yearly usage to max-in-stock extrapolation: As indicated in Section 2.1, it is best for the user to report both the max-in-stock as well as the yearly usage data of a chemical product from the upstream industries. These data are used to develop an extrapolation algorithm for chemical products that do not have max-in-stock values. If the user provides only a few instances of chemical products with both max-in-stock and yearly usage data (i.e., the dataset is quite small), then the extrapolation algorithm that will be applied to chemical products with missing max-in-stock data will not be very robust (and if no such data are provided, the tool will likely experience severe problems). A future version of RAVEN STREAM could provide a "backup" extrapolation algorithm to use for chemical products that only have yearly usage data provided, if the user input data contains little or no data on chemical products with both max-in-stock and yearly usage data.

It should also be noted that while the intended use of RAVEN STREAM was for analysing real collected survey data on chemical product storage at industries upstream from an STP, analysis of 'hypothetical' chemical storage and spill scenarios are also possible with the tool. This enables estimation of what storage patterns upstream would be required to present a risk to the receiving treatment plant. For example, if several chemicals of concern are known or thought to be employed at industries upstream from the STP, a test dataset of these chemicals could be developed with a range of hypothetical upstream storage masses. These data could then be run through the RAVEN STREAM tool to determine which of these chemicals and corresponding storage masses could pose a potential risk to the STP in the event of a spill. This hypothetical risk prioritisation information could be used to query the upstream industries about specific chemicals to determine if the storage masses meet or exceed the levels of concern for posing an unacceptable risk to the STP (i.e., as opposed to performing a comprehensive survey of all chemicals used at the industries upstream).

The RAVEN STREAM online tool, and this accompanying User's Guide, are intended to be starting points for an upstream chemical risk framework system. Hopefully, this framework can be improved upon and grow as the user base expands and feedback is provided from the various users with different use cases, priorities, concerns, and logistical situations surrounding their sewage treatment plants. Users are encouraged to report problems, suggest areas of improvement, or provide other comments on the RAVEN STREAM tool and this User's Guide to the corresponding author (see title page of this document for contact information). Such feedback is greatly appreciated.

6 References

Hader, J.D., Frenzel, M., Scullin, J., Plaza, E., MacLeod, M., 2023. Prioritizing toxic shock threats to sewage treatment plants from down-the-drain industrial chemical spills: the RAVEN STREAM online tool. Environ. Sci.: Adv. 10.1039.D3VA00067B. https://doi.org/10.1039/D3VA00067B

Welch, S.A., Lane, T., Desrousseaux, A.O.S., van Dijk, J., Mangold-Döring, A., Gajraj, R., Hader, J.D., Hermann, M., Kutteyeri, A.P.A., Mentzel, S., Nagesh, P., Polazzo, F., Roth, S.K., Boxall, A.B.A., Chefetz, B., Dekker, S.C., Eitzinger, J., Grung, M., MacLeod, M., Moe, S.J., Rico, A., Sobek, A., van Wezel, A.P., 2022. ECORISK2050: An Innovative Training Network

for predicting the effects of global change on the emission, fate, effects, and risks of chemicals in aquatic ecosystems [version 2; peer review: 2 approved].