

PERK Tutorial (Ver01)

Kishore Jagadeesan

08/09/2022

Table of contents

1	Introduction	5
2	Data sources:	6
2.1	Prescription Data For England:	6
2.2	WWTP Data:	6
2.3	API properties	7
3	Workflow	8
4	PERK Features	10
5	Upload Data	11
6	Predicted (PC)	13
6.1	Predicted: Prescription	13
6.2	Predicted: Predicted Concentration	15
7	Measured (MC)	17
8	Predicted vs Measured	20
8.1	Predicted vs Measured: Predicted vs Measured	20
8.2	Predicted vs Measured: Prediction Accuracy	22
9	Acknowledgements	24
9.1	Disclaimer	24
	References	25

List of Figures

3.1	PERK: Workflow.	9
4.1	PERK: Features	10
5.1	PERK: Upload Data	11
6.1	Predicted: Prescription - AV Panel.	13
6.2	PC: kg/month.	14
6.3	PC: PNDP.	14
6.4	Predicted: Predicted Concentrations - AV Panel.	15
6.5	PC: concentration/month.	16
6.6	PC: concentration/period.	16
7.1	Measured: Measured Concentrations - AV Panel.	17
7.2	MC: concentration/month.	19
7.3	MC: concentration/period.	19
8.1	Predicted vs Measured: Predicted vs Measured - AV Panel.	20
8.2	PCvsMC: PEC-I.	21
8.3	PCvsMC: PEC-II.	21
8.4	Predicted vs Measured: Prediction Accuracy - AV Panel.	22
8.5	PCvsMC: PA-I.	23

List of Tables

5.1	Upload Data: Data Input	11
6.1	Predicted: Prescription Sub-Panel	14
6.2	Predicted: Predicted Concentrations Sub-Panel	15
7.1	Measured Panel	17
8.1	Predicted vs Measured: Predicted vs Measured Sub-Panel	21
8.2	Predicted vs Measured: Prediction Accuracy Sub-Panel	22

1 Introduction

The aim of PERK is to predict and visualize concentrations of pharmaceuticals in the aqueous environment.

PERK acronym for **P**redicting **E**nvironmental concentration and **RisK**, is an R/Shiny application tool (Team et al. 2013), aims to facilitate automated modelling and reporting of predicted environmental concentrations of a comprehensive set of pharmaceuticals derived from a wide range of therapeutic classes with different mode of action.

The tool helps users,

- to input their measured concentration,
- to compare the predicted and measured concentrations of the APIs by means of the PEC/MEC ratio,
- to establish whether the predicted equations used tend to underestimate or overestimate measured values.
- It provides a consistent interactive user interface in a familiar dashboard layout, enabling users to visualise predicted values and compare with their measured values without any hassles.
- Users can download data and graphs generated using the tool in .csv or publication ready images (.pdf, .eps).

2 Data sources:

2.1 Prescription Data For England:

This tool uses the prescription data from PrAna(Jagadeesan et al. 2022), an R package to calculate and visualize England NHS prescribing data.

The data used in PrAna are as follows,

- Prescribing data and Practice information are from the monthly files published by the NHS Business Service Authority, used under the terms of the Open Government Licence.
- BNF codes and names are also from the NHS Business Service Authority's Information Portal, used under the terms of the Open Government Licence.
- dm+d weekly release data is also from the NHS Business Service Authority's Information Portal, used under the terms of the Open Government Licence.

2.2 WWTP Data:

The following dataset are provided from WWTP collaborators,

- Catchment map used to define the boundaries and capture the GP Practices inside the catchments for the prescription data calculations.
- Daily flow data used to calculate the load and population equivalent.
- Population Equivalent number of inhabitants per catchment zone.
- Site information required to predict information such as recovery percentage.
- Water quality parameters to predict population equivalent.

2.3 API properties

- Metabolites and Excretion factors collected from research articles and data repositories such as Drug bank.
- Recovery percentage collected from research articles, calculated from measured concentration from previous experiments, predicted using WWTP site information.
- Physio-chemical properties collected from research articles and data repositories.
- Site information required to predict information such as recovery percentage.
- Eco-toxicity data collected from research articles and data repositories.

3 Workflow

The workflow in this tutorial consists of the following steps, as in the Figure [3.1](#).

- Upload Data: Download template for the dataset and upload in the corresponding input holders.
- Analysis and Visualisation (AV) Panel: Click on the relevant analysis and visualisation panel. PERK features three AV panels (1) Predicted, (2) Measured, (3) Predicted vs Measured.
- Analysis and Visualisation settings (AVS): Click respective analysis and visualisation setting (AVS) tab, to select the option to analyse and visualise datatable/plot
- Plot settings: Click on the plot settings such as, color and line width for the better/suitable visualisation.
- Download data: Click on the download buttons to download generated plot/data in publication friendly .pdf/.eps or .csv files.

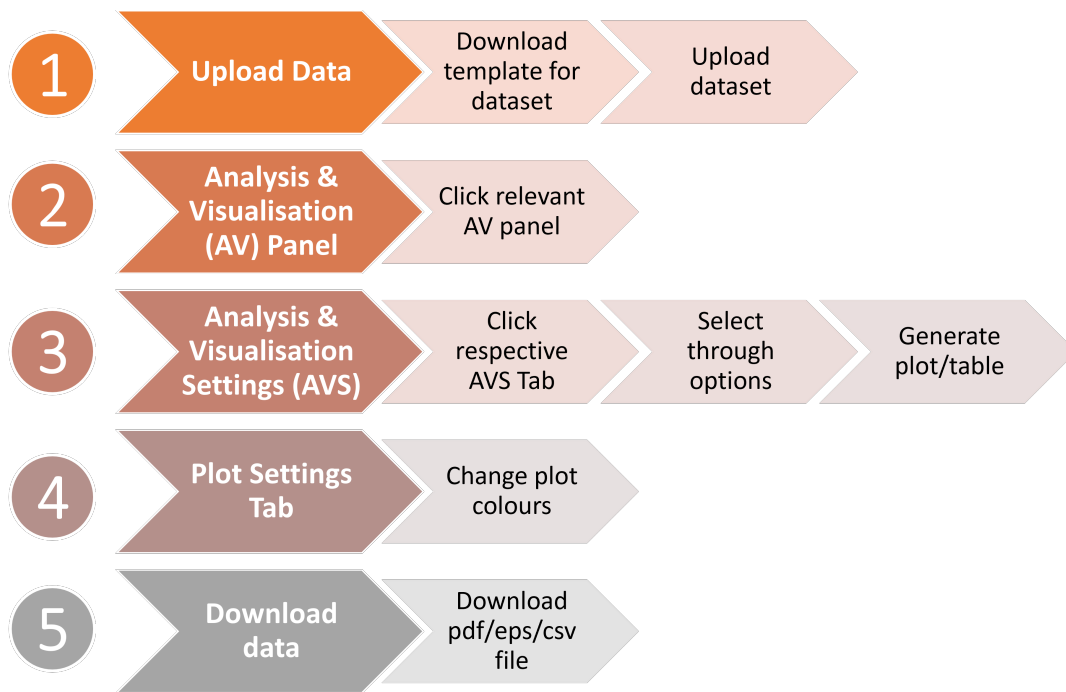


Figure 3.1: PERK: Workflow.

4 PERK Features

- PERK consist of several features, broadly categorized as three panels
 - (1) Upload Data (Section 5)
 - (2) Predicted (Section 6)
 - (3) Measured (Section 7)
 - (4) Predicted vs Measured (Section 8)
- Overview of the individual panels and their options can be found in Figure 4.1 and will be discussed in the following sections.

Predicted (PC)

- Prescription Trends
 - kg/month
 - mg/day/1000 inhabitants
- Prediction Trends (influent, effluent, river)
 - (ng/l) - Individual month, and Selected period

Measured (MC)

- Measured Trends (influent, effluent, river)
 - Concentration (ng/L)
 - Daily load (mg/day)
 - Population normalised daily load (mg/day/1000 inhabitants)
- Individual day, month, and Selected period.

Predicted vs Measured (PC vs MC)

- PC vs MC (influent, effluent, river)
- Prediction accuracy plot
- Individual month, and Selected Period

Figure 4.1: PERK: Features

5 Upload Data

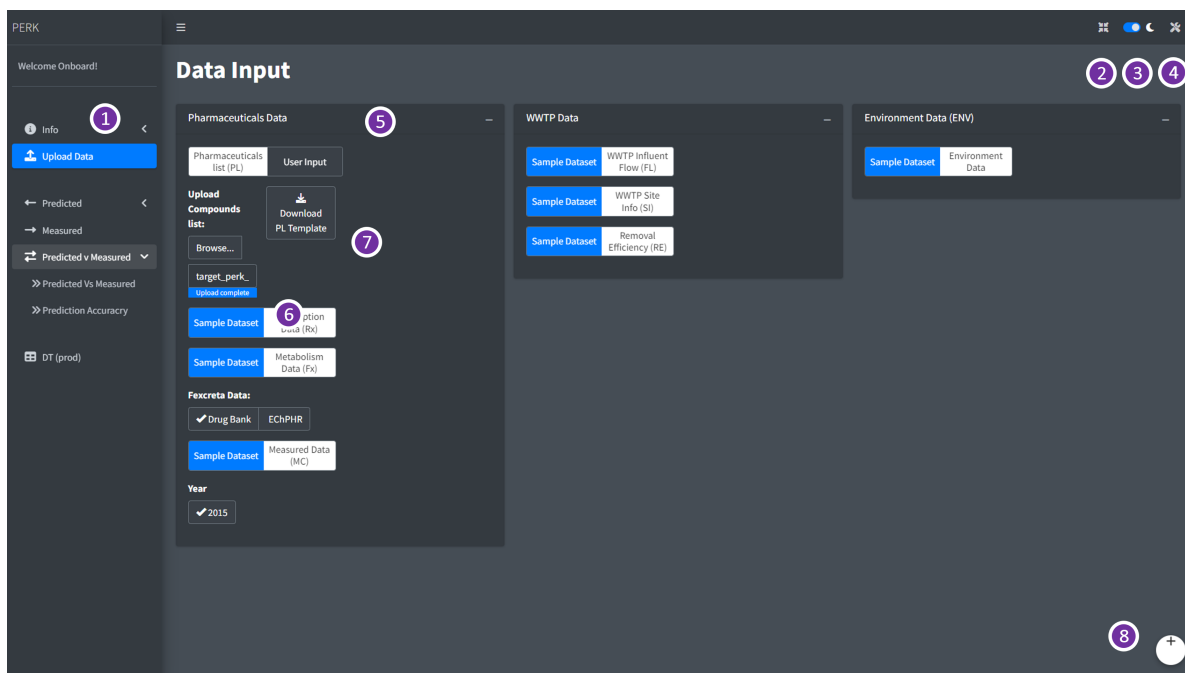


Figure 5.1: PERK: Upload Data

Table 5.1: Upload Data: Data Input

Part	Remarks
1	Analysis and Visualisation (AV) Panel
2	Full screen
3	Dark and Light mode
4	Plot settings
5	Data selection Area
6	Upload File Button
7	Download Template for the file
8	User Logout

- In this panel, user can Download template for the dataset and upload in the corre-

sponding input holders as in Figure 5.1.

- User can click on the **Download template** button to generate the comma separated value (.csv) file.
- Once the template is downloaded, user can add in or convert their dataset to corresponding template and upload it in the corresponding input holders to do the analysis and visualisation.

6 Predicted (PC)

- Predicted (PC) panel, has two sub-panels
 - (1) Prescription: to analyse and visualise the prescription trends and
 - (2) Predicted Concentrations: to analyse and visualise the prediction trends based on user inputs.

6.1 Predicted: Prescription

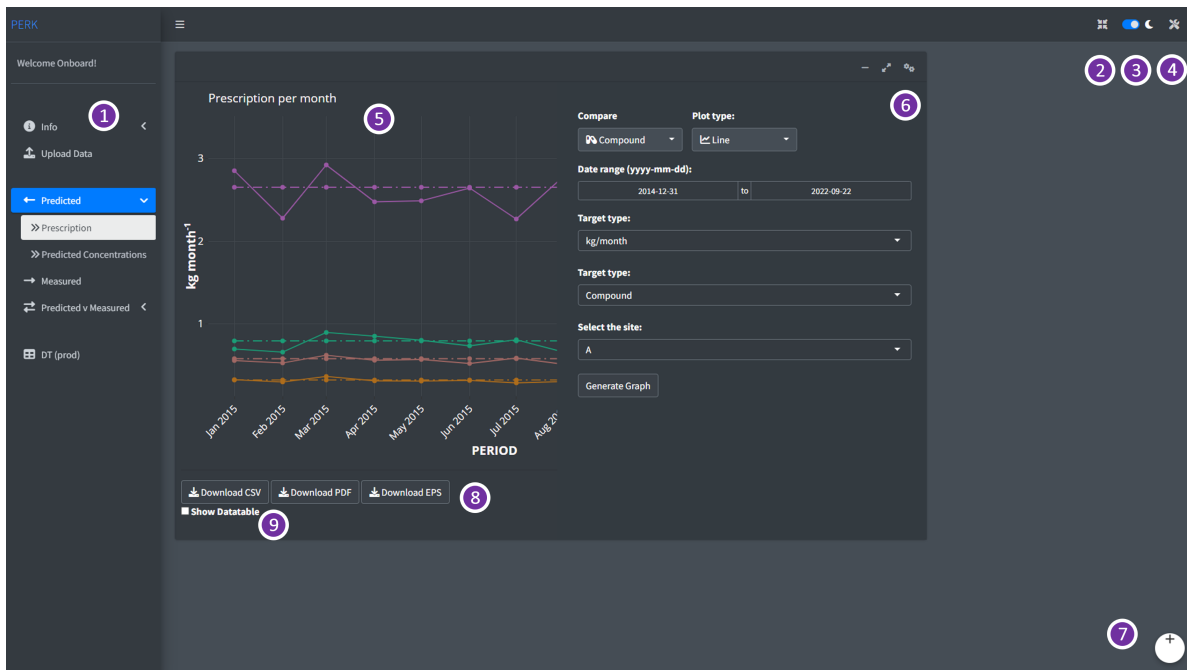


Figure 6.1: Predicted: Prescription - AV Panel.

- Different parts of the **Predicted: Prescription** sub-panel and PERK dashboard is highlighted in the Figure 6.1 and listed in the Table 6.1.
- In Prescription sub-panel, user can select the period of their interest using the **Data Range** option, and select prescription type (raw or population normalized) value using

Target type and the site using **Select the site** options in the analysis and visualisation settings (AVS) tab, as in Figure 6.1

Table 6.1: Predicted: Prescription Sub-Panel

Part	Remarks
1	Analysis and Visualisation (AV) Panel
2	Full screen
3	Dark and Light mode
4	Plot settings
5	Plot generated based on user selection
6	Analysis and Visualisation settings (AVS) panel
7	User log-out
8	Download buttons to download the generated plot as .pdf or .eps and data as .csv format
9	Show Datatable

- Prescription trends in the Predicted (PC) panel, can generate long-term month wise raw prescription trends (kg/month), as in Figure 6.2 and population normalized daily loads based on prescription (PNDP) (mg/day/1000 inhabitants) as in Figure 6.3 .
- User can download the generated plot as publication-friendly images in .pdf/.eps format, user can also download the images in .png format and data generated for the plot as .csv file using the download buttons present below the plot.
- User can view the data table by checking the **Show Datatable** check box present below the download buttons.

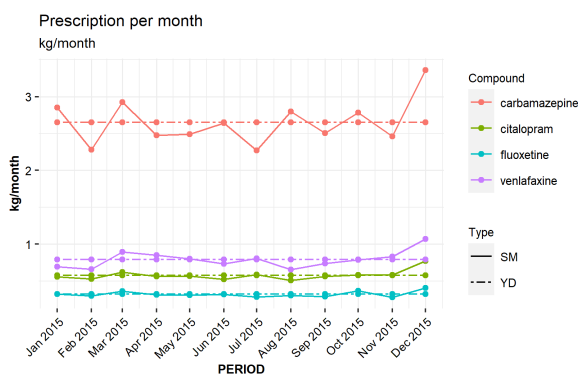


Figure 6.2: PC: kg/month.

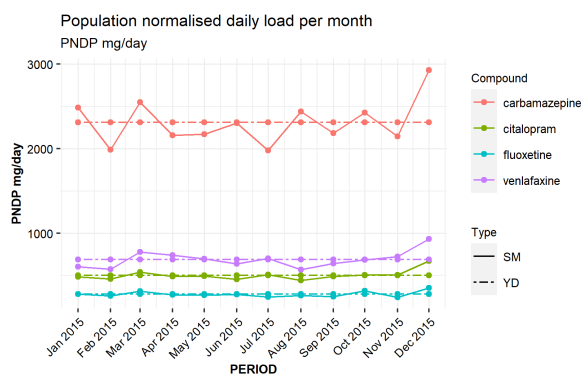


Figure 6.3: PC: PNDP.

6.2 Predicted: Predicted Concentration

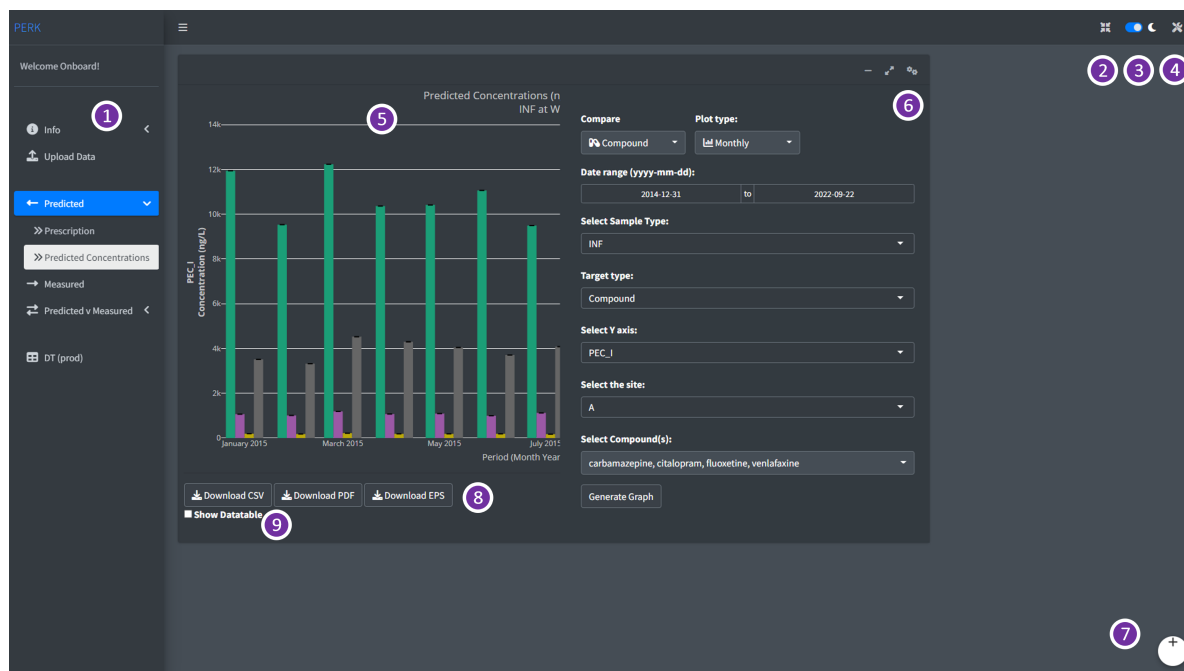


Figure 6.4: Predicted: Predicted Concentrations - AV Panel.

- Different parts of the **Predicted: Predicted Concentrations** sub-panel and PERK dashboard is highlighted in the Figure 6.4 and listed in the Table 6.2.

Table 6.2: Predicted: Predicted Concentrations Sub-Panel

Part	Remarks
1	Analysis and Visualisation (AV) Panel
2	Full screen
3	Dark and Light mode
4	Plot settings
5	Plot generated based on user selection
6	Analysis and Visualisation settings (AVS) panel
7	User log-out
8	Download buttons to download the generated plot as .pdf or .eps and data as .csv format
9	Show Datatable

- In the predicted concentrations sub-panel, user can select the period of their interest using the **Data Range** option, and select prediction sample type (wastewater influent

INF, wastewater effluent EFF and river RDOWN) using **Sample type** and the site using **Select the site** options in the analysis and visualisation settings (AVS) tab, as in Figure 6.4

- Two types of prediction values can be visualised in this panel based on the prescription data,
 - PEC_I: This prediction considers prescription based on individual month,
 - PEC_II: This prediction is based on the prescription per year.

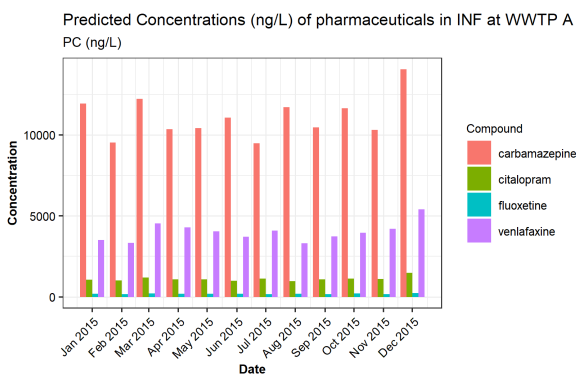


Figure 6.5: PC: concentration/month.

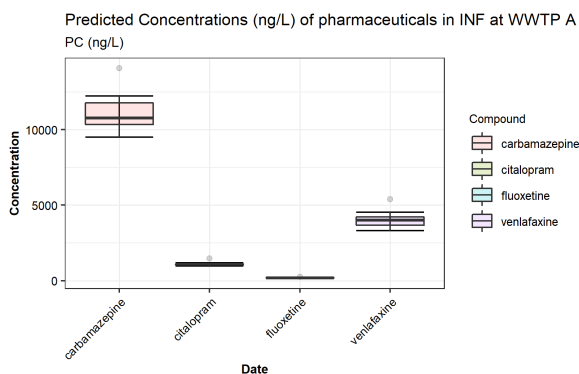


Figure 6.6: PC: concentration/period.

- This panel visualise prediction per month for the selected period (ng/L) as in the Figure 6.5, and total prediction per selected period (ng/L), as in the Figure 6.6
- In addition, this panel also enables to compare month wise and total predicted concentration of selected pharmaceuticals over different environmental matrices, such as, INF, EFF and RDOWN and compare over different WWTPs in the study.
- User can download the generated plot as publication-friendly images in .pdf/.eps format, user can also download the images in .png format and data generated for the plot as .csv file using the download buttons present below the plot.
- User can view the data table by checking the **Show Datatable** check box present below the download buttons.

7 Measured (MC)

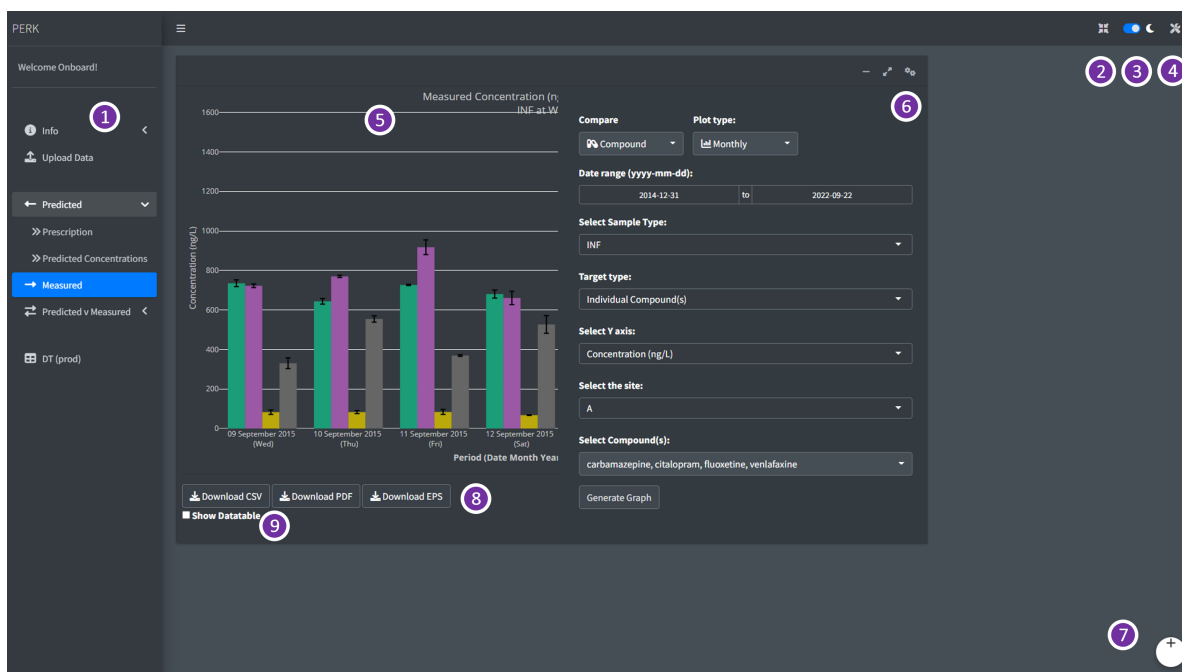


Figure 7.1: Measured: Measured Concentrations - AV Panel.

- Different parts of the **Measured** tab and PERK dashboard is highlighted in the Figure 7.1 and listed in the Table 7.1.

Table 7.1: Measured Panel

Part	Remarks
1	Analysis and Visualisation (AV) Panel
2	Full screen
3	Dark and Light mode
4	Plot settings
5	Plot generated based on user selection
6	Analysis and Visualisation settings (AVS) panel
7	User log-out

Part	Remarks
8	Download buttons to download the generated plot as .pdf or .eps and data as .csv format
9	Show Datatable

- In **Measured** panel as in the Figure 7.1, user can select the period of their interest using the **Data Range** option, and select sample matrix type (**INF** - wastewater influent, **EFF** - wastewater effluent, **RDOWN** - River Downstream, **RUP** - River upstream, **SPM** - Solids) using **Sample type**, based on the user input dataset.
- User can select the measurement type (Concentration, **DL** - Daily Load, **PNDL** - Population normalised daily load) using **Measurement Type**, and the site by **Select the site** options in the analysis and visualisation settings (AVS) tab, as in Figure 7.1
- User can download the generated plot as publication-friendly images in .pdf/.eps format, user can also download the images in .png format and data generated for the plot as .csv file using the download buttons present below the plot.
- User can view the data table by checking the **Show Datatable** check box present below the download buttons.
- Three types of measured values can be visualised in this panel based on the measurement data uploaded by the user,
 - **Concentration (ng/L)**: This is the raw concentration values based on individual measurements.
 - **DL (mg/day)**: This is the Daily Load (DL) values based on measurements normalised with the daily flow of wastewater for the **INF** and **EFF**, and river for the **RDOWN** and **RUP**.
 - **PNDL (mg/day/1000 inhabitants)**: This is the Population Normalised Daily Load values calculated based on the population in the WWTP catchment and the daily flow.
- This panel visualise measurement per month for the selected period (**ng/L**) as in the Figure 7.2, and total measurement per selected period (**ng/L**), as in the Figure 7.3

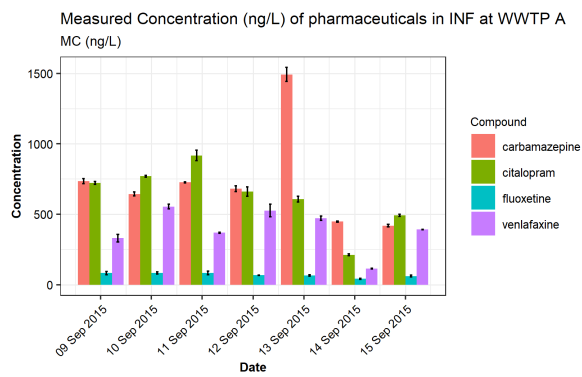


Figure 7.2: MC: concentration/month.

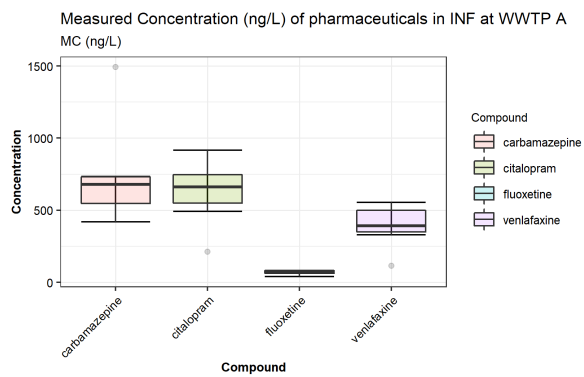


Figure 7.3: MC: concentration/period.

8 Predicted vs Measured

- Predicted vs Measured (PC vs MC) panel, has two sub-panels
 - (1) Predicted vs Measured: to analyse and visualise the predicted trends vs measured trends and
 - (2) Prediction Accuracy - PA: to analyse and visualise the prediction accuracy.

8.1 Predicted vs Measured: Predicted vs Measured

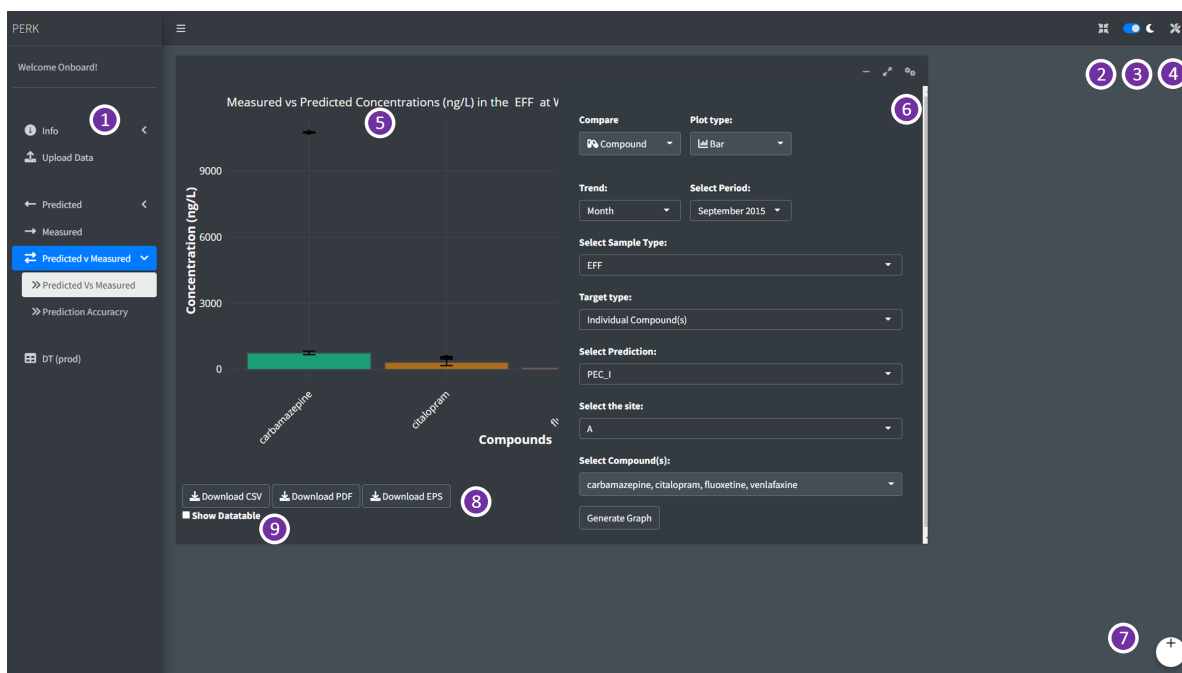


Figure 8.1: Predicted vs Measured: Predicted vs Measured - AV Panel.

- Different parts of the Predicted vs Measured: Predicted vs Measured sub-panel and PERK dashboard is highlighted in the Figure 8.1 and listed in the Table 8.1.

Table 8.1: Predicted vs Measured: Predicted vs Measured Sub-Panel

Part	Remarks
1	Analysis and Visualisation (AV) Panel
2	Full screen
3	Dark and Light mode
4	Plot settings
5	Plot generated based on user selection
6	Analysis and Visualisation settings (AVS) panel
7	User log-out
8	Download buttons to download the generated plot as .pdf or .eps and data as .csv format
9	Show Datable

- In **Predicted vs Measured** sub-panel, user can select the period of their interest using the **Select Period** option, and select sample type (wastewater influent INF, wastewater effluent EFF and river RDOWN) using **Select Sample type** and the site using **Select the site** options in the analysis and visualisation settings (AVS) tab, as in Figure 8.1
- Predicted vs Measured trends in the **Predicted vs Measured (PCvsMC)** panel, can generate measured concentration vs PEC_I and PEC_II, predictions based on monthly prescription as in Figure 8.2 and prediction based on the prescription per year Figure 8.3 respectively.

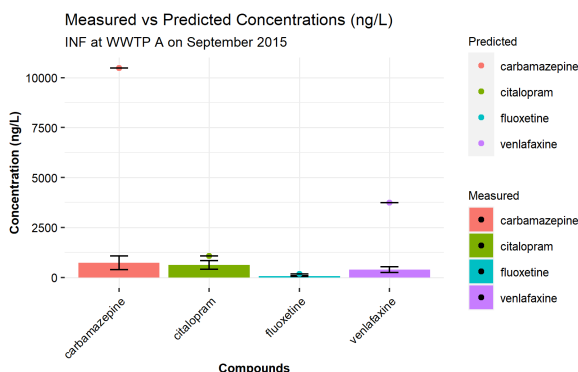


Figure 8.2: PCvsMC: PEC-I.

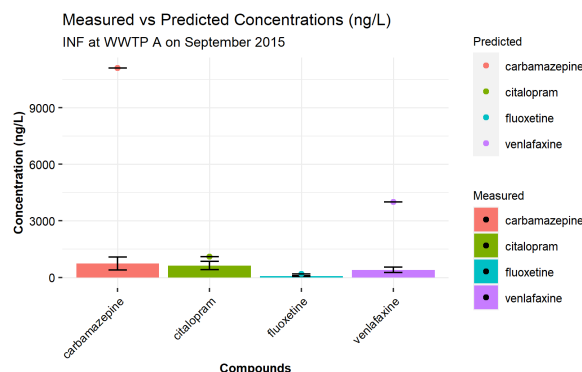


Figure 8.3: PCvsMC: PEC-II.

- User can download the generated plot as publication-friendly images in .pdf/.eps format, user can also download the images in .png format and data generated for the plot as .csv file using the download buttons present below the plot.
- User can view the data table by checking the **Show Datable** check box present below the download buttons.

8.2 Predicted vs Measured: Prediction Accuracy

- Different parts of the Predicted vs Measured: Prediction Accuracy sub-panel and PERK dashboard is highlighted in the Figure 8.4 and listed in the Table 8.2.

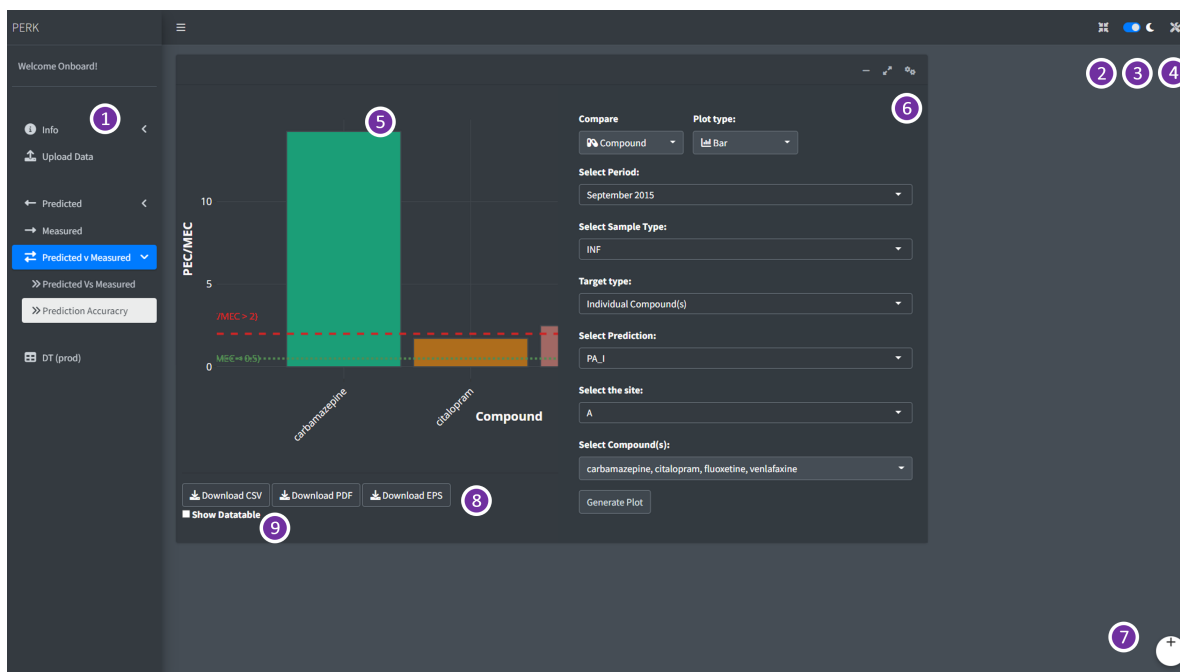


Figure 8.4: Predicted vs Measured: Prediction Accuracy - AV Panel.

Table 8.2: Predicted vs Measured: Prediction Accuracy Sub-Panel

Part	Remarks
1	Analysis and Visualisation (AV) Panel
2	Full screen
3	Dark and Light mode
4	Plot settings
5	Plot generated based on user selection
6	Analysis and Visualisation settings (AVS) panel
7	User log-out
8	Download buttons to download the generated plot as .pdf or .eps and data as .csv format
9	Show Datatable

- In Prediction Accuracy sub-panel, user can select the period of their interest using the Select Period option, and select sample type (wastewater influent INF, wastewater

effluent EFF and river RDOWN) using **Select Sample type** and the site using **Select the site** options in the analysis and visualisation settings (AVS) tab, as in Figure 8.4

- Prediction Accuracy trends in the **Prediction Accuracy (PA)** panel, can generate trends in PA_I and PA_II, predictions based on monthly prescription as in Figure 8.5 and prediction based on the prescription per year respectively.

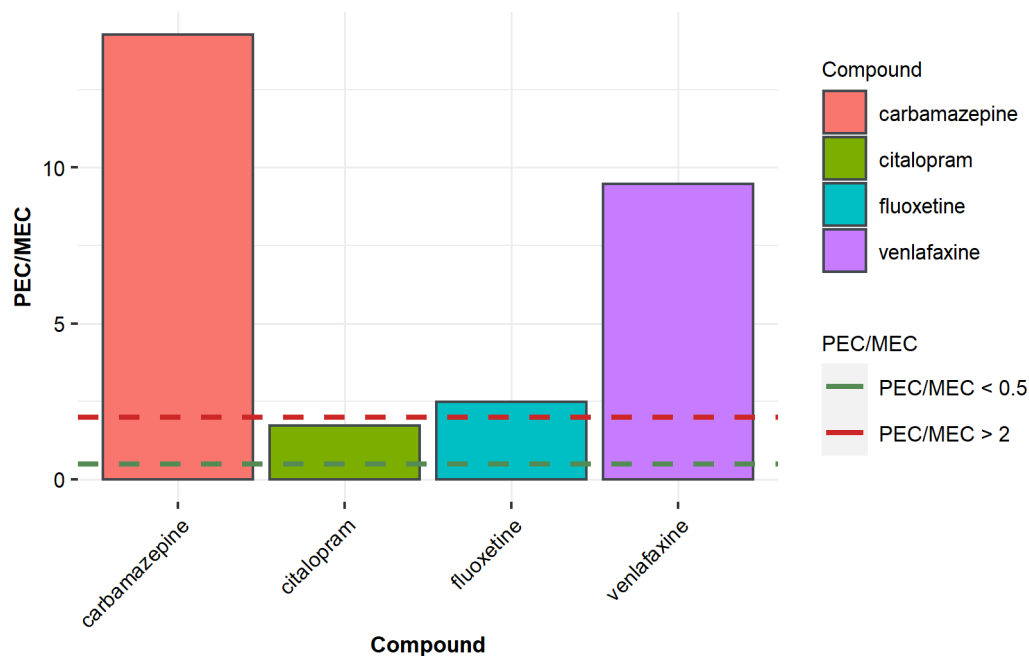


Figure 8.5: PCvsMC: PA-I.

- User can download the generated plot as publication-friendly images in .pdf/.eps format, user can also download the images in .png format and data generated for the plot as .csv file using the download buttons present below the plot.
- User can view the data table by checking the **Show Datatable** check box present below the download buttons.

9 Acknowledgements

This package was built as a part of the **Wastewater Fingerprinting for Public Health Assessment (ENTRUST)** and **Innovative Pathway Control (IPC)** project funded by Wessex Water and EPSRC IAA (grant no. EP/R51164X/1).

9.1 Disclaimer

We accept no liability for any errors in the data or its publication here: use this data at your own risk. You should not use this data to make individual prescribing decisions.

References

- Jagadeesan, Kishore Kumar, James Grant, Sue Griffin, Ruth Barden, and Barbara Kasprzyk-Hordern. 2022. “PrAna: An R Package to Calculate and Visualize England NHS Primary Care Prescribing Data.” *BMC Medical Informatics and Decision Making* 22 (1). <https://doi.org/10.1186/s12911-021-01727-z>.
- Team, R Core et al. 2013. “R: A Language and Environment for Statistical Computing.”