# Workflows

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **WORKFLOW** | **NAME** | **IN-&OUT** | | **R Scripts** |
| WF01 | Data Subsets | Input | “buist2012\_bloodairpartition\_data.xlsx” (B1/B2)  “paixao2013\_tissuebloodpartition\_invivo\_data.xlsx” (P1)  “paixao2013\_tissuebloodpartition\_invivo\_data.xlsx” (P2)  “HFC\_PCs\_2014.xlsx” (Sheet: Reported values) | S01a\_FolderFilePaths |
| S02\_B1B2transformation |
| S03\_PCvesselcal |
| Output | “Data\_subsetlist.xlsx” (sheets: All subsets, With HFCs, without HFCs)  “NameCorrectionslist.xlsx” | S04\_P1namecorrection |
| S05\_P2namecorrection |
| Comment |  | S06\_Untransformed |
| S07\_Logtransformed |
| S08\_HFCdataSuffix |
| WF02 | Training data format | Input | “buist2012\_bloodairpartition\_data.xlsx” (B1/B2)  “paixao2013\_tissuebloodpartition\_invivo\_data.xlsx” (P1)  “paixao2013\_tissuebloodpartition\_invitro\_data.xlsx” (P2) | S01b\_FolderFilePaths |
| S02\_B1B2transformation |
| S03\_PCvesselcal |
| Output | “TrainingData.xlsx” (including smiles) | S04\_P1namecorrection |
| Comment |  | S05\_P2namecorrection |
| WF03 | Data Statistics | Input | “Data\_subsetlist.xlsx” (sheets: All subsets) | S01c\_FolderFilePaths |
| Output | Density plots (placed in DensityPlots)  “DescriptiveStats.xlsx” (Sheet: All data) | S09\_Subsetnamecolumn |
| S10\_Descriptivestats |
| Comment |  | S11\_StatsReordering |
| PS1\_Densityplot |
| WF04 | Experimental data | Input | “HFCexperimental.xlsx” | S01d\_FolderFilePaths |
| Output | “HFCAverage.xlsx” (sheet: All, sheet: Scenarios) | S12\_ExperimentalInfo |
| Comment |  | S13\_averagingTestsubjects |
| S14\_averagingExdata |
| WF05 | LROP1 | Setting | Model Generation | S00.2\_Modelgenerations |
| Input | “Data\_subsetlist.xlsx” (sheets: All subsets) | S01e\_FolderFilePaths |
| Output | “OptimizationStats.xlsx) (sheet: LROP1) | S15\_LROP1thresholds |
| Comment |  | S16\_LROP1Indicator |
| WF06 | LROP2 | Setting | Model Generation | S00.2\_Modelgenerations |
| Input | “Data\_subsetlist.xlsx” (sheets: All subsets)  “OptimizationStats.xlsx) (sheet: LROP1) | S01f\_FolderFilePaths |
| S17\_BFEmodelname |
| Output | BFEmodels (zip)  “OptimizationDescriptor.xlsx” (sheet: LROP2) | S18\_BFELimit |
| S19\_Descriptorlist |
| Comment |  | S20\_BFEinfo |
| WF07 | LR1 | Setting | Model Generation | S00.2\_Modelgenerations |
| Input | “Data\_subsetlist.xlsx” (sheets: All subsets)  “OptimizationStats.xlsx) (sheet: LROP1)  “OptimizationDescriptor.xlsx” (sheet: LROP2)  “HFC\_PCs\_2014.xlsx” | S01g\_FolderFilePaths |
| S21\_LRTransposing |
| S22\_LRmodelname |
| S23\_LRTrainingOrganizing |
| Output | “ModelStats.xlsx” (Sheet: LR1)  “ModelOutputdata.xlsx” (Sheet: LR1)  Observed-predicted plot  Models (pmml) | S24\_LRhfcOrganizing |
| S25\_LRModelstats |
| S26\_ModelPredictionInfo |
| PS02\_Modelstatplot |
| Comment |  | S31\_G1Melting |
| PS03\_G1comparison |
| WF08 | RF1 | Setting | Model Generation  List of nodesizes  List of number of Trees | S00.2\_Modelgenerations |
| S00.3\_RFsettings |
| S01h\_FolderFilePaths |
| Input | “Data\_subsetlist.xlsx” (sheets: All subsets)  “HFC\_PCs\_2014.xlsx” | S26\_ModelPredictionInfo |
| S27\_RFvariables |
| Output | “ModelStats.xlsx” (Sheet: RF1)  “ModelOutputdata.xlsx” (Sheet: RF1)  Observed-predicted plot  Models (zip)  Graphic comparison of all RF1 models | S28\_RFTrainingOrganizing |
| S29\_RFhfcOrganizing |
| S30\_RFModelstats |
| PS02\_Modelstatplot |
| S31\_G1Melting |
| Comment |  | PS03\_G1comparison |
| WF09 | RFOP1 | Setting | Model Generation | S00.2\_Modelgenerations |
| Input | “Data\_subsetlist.xlsx” (sheets: All subsets)  “ModelStats.xlsx” (Sheet: RF1) | S01i\_FolderFilePaths |
| S28\_RFTrainingOrganizing |
| S32\_RFseedselection |
| Output | “OptimizationOutputdata.xlsx” (sheet: RFOP1)  “OptimizationStats.xlsx” (sheet: RFOP1)  RFOP1 plot (change in RMSE and R2) | S33\_Nodesizelist |
| S34\_RFOP1Modelstats\_nodesize |
| S35\_Treelist |
| Comment | Included only log-transformed data and the use of molecular descriptors | S36\_RFOP1\_Modelstats\_treelist |
| S37\_RFOP1Optimumsettings |
| S38\_RFOP1Melting |
| PS04\_RFOP1plot |
| WF10 | RF2 | Setting | Model Generation | S00.2\_Modelgenerations |
| Input | “Data\_subsetlist.xlsx” (sheets: All subsets)  “HFC\_PCs\_2014.xlsx”  “OptimizationStats.xlsx” (sheet: RFOP1) | S01j\_FolderFilePaths |
| S26\_ModelPredictionInfo |
| S28\_RFTrainingOrganizing |
| Output | “ModelStats.xlsx” (Sheet: RF2)  “ModelOutputdata.xlsx” (Sheet: RF2)  Observed-predicted plot  Models (zip) | S29\_RFhfcOrganizing |
| S30\_RFModelstats |
| S39\_RFOP1variables |
| PS02\_Modelstatplot |
| Comment | Included only log-transformed data and the use of molecular descriptors |  |
| WF11 | RFOP2 | Setting | Model Generation | S00.2\_Modelgenerations |
| Input | “Data\_subsetlist.xlsx” (sheets: All subsets)  “OptimizationStats.xlsx” (sheet: RFOP1) | S01k\_FolderFilePaths |
| S28\_RFTrainingOrganizing |
| S40\_RFOP2variables |
| Output | “OptimizationOutputdata.xlsx” (sheet: RFOP2)  “OptimizationDescriptors.xlsx” (sheet: RFOP2)  Plot of percental change in R2 and RMSE | S41\_RFOP2Modelstats |
| S42\_RFOP2Descriptorlist |
| S43\_RFOP2Melting |
| Comment | Included only log-transformed data and the use of molecular descriptors | PS05\_RFOP2descriptors |
|  |
| WF12 | RF3 | Setting | Model Generation | S00.2\_Modelgenerations |
| Input | “Data\_subsetlist.xlsx” (sheets: All subsets)  “HFC\_PCs\_2014.xlsx”  “OptimizationDescriptors.xlsx” (sheet: RFOP2)  “ModelStats.xlsx” (Sheet: RF2) | S01l\_FolderFilePaths |
| S44\_RF3Descriptorlist |
| S39\_RFOP1variables |
| S26\_ModelPredictionInfo |
| Output | “ModelStats.xlsx” (Sheet: RF3)  “ModelOutputdata.xlsx” (Sheet: RF3)  Observed-predicted plot  Models (zip) | S28\_RFTrainingOrganizing |
| S29\_RFhfcOrganizing |
| S30\_RFModelstats |
| PS02\_Modelstatplot |
| Comment |  |  |
| WF13 | Prediction and APD | Setting | - | S01m\_FolderFilePaths |
| Input | “Data\_subsetlist.xlsx” (sheets: All subsets)  “HFC\_PCs\_2014.xlsx”  “OptimizationDescriptors.xlsx”  “ModelStats.xlsx” | S45\_Modelfilelist |
| S44\_RF3Descriptorlist |
| S46\_PCAtrainingdata |
| S47\_PCinfo |
| Output | “PredictionADP.xlsx” (PC prediction for each HFC + reliability of each model)  PCA0-PCA1 plot with HFCs | S48\_PredictionADPformatting |
| PS06\_PCAplot |
|  |
| Comment |  |  |
| WF14 | Model Selection | Setting | Row filters selection “MD” and “Log” models (in metanode) | S01n\_FolderFilePaths |
| Input | “PredictionADP.xlsx” (sheet: All) | S49\_ReliabilityFilter |
| Output | “ModelSelection” (sheets: “ReliableModels”, “TopModels”, “SelectionHFC”, “SelectionGeneral”) | S50\_ModelSelection |
| S51\_DomainDifference |
| Comment | Final selection was done manually, judging statistics and overall reliablity | S52\_SelectionInfo |
| WF15 | Sensitivity analysis | Setting |  | S01o\_FolderFilePaths |
| Input | HFCaveragedata.xlsx (sheet: scenario)  HFC\_PCs\_2014.xlsx (sheet: reported values) | S53\_SAPCvaluelist |
| S54\_BaselineParameters |
| Output | SensitivityEndpoints.xlsx (sheets: “Baseline” & “SA”)  Line plot for each exposure scenario (change in PC vs change in endpoint)  Column plot for each exposure scenario (slope of sensitivity)  Column plot of all slopes of sensitivity | PBTK\_2hExposure |
| S55\_EndpointCal |
| S56\_AddScenario |
| S57\_SAparameters |
| S58\_AddScenarioSA |
| S59\_EndpointChangeCal |
| Comment |  | S60\_SAdatamelting |
| S61\_SASlopeCal |
| S62\_SASlopeDatamelting |
| PS07\_SALinePlot |
| PS08\_SAColumnPlot |
| PS09\_SAOverallPlot |
| WF16 | PBTK vs Experimental | Setting |  | S01p\_FolderFilePaths |
| Input | “ModelSelection.xlsx” (sheet = SelectionHFC)  “HFCaveragedata.xlsx” (sheet = All)  “HFC\_PCs\_2014.xlsx” (sheet = Reported values) | S63\_ScenarioList |
| S64\_PCtransformation |
| S65\_SourceColumn |
| S66\_MCcombos |
| Output | “PBTKexOutputdata.xlsx” (sheets = Scenario)  “PBTKexEndpoints.xlsx” (sheet = All) | S54\_BaselineParameters |
| PBTK\_2hExposure |
| S67\_MCminmax |
| Comment |  | S68\_AddSource |
| S69\_TimepointSelection |
| S70\_DataComparison |
| PS10\_TKvsEXplot |

# TK prediction workflows

|  |  |  |  |
| --- | --- | --- | --- |
| **WORKFLOW** | **NAME** | **IN-&OUT** | **R Scripts** |
| Inhalation TK prediction | Setting | Molecule format (SMILES, IUPAC, MarvinSketch or SDF)  Molecule input  Parameters:   * Body weight (kg) * Body height (m) * Ventilation rate (at work) (L/min) * Ventilation rate (at rest) (L/min) * Clearance (L/min) * Work hours per week * Work days per week * Length of simulation (in days) * Exposure level (in ppm) | tkS01a\_FolderFilepaths |
| S64\_PCtransformation |
| S44\_RF3Descriptorlist |
| tkS02\_PCtransposition |
| tkS03\_PCAtrainingdata |
| S47\_PCinfo |
| tk09\_ADPreportInfo |
| tkS04\_Indicator |
| tkS05\_PBTKparameters |
| tkS06\_CmaxAUC |
| tkPS01\_TKprofile |
| tkPS02\_TKextremevalues |
| Input | “ModelSelection.xlsx” (sheet = SelectionGeneral)  “Data\_subsetlist.xlsx” (sheet = All subsets)  “OptimizationDescriptors.xlsx” (sheet = previoiusOP) | tkPS03\_PCAplot |
| Output | Full applicability report  PCA plot for all PCs  Graphic presentation of PBTK prediction | PBTK\_WorkWeek |
| Comment |  |  |
| Inhalation TK prediction (multiple) | Setting | Molecule format (As typed table or excel table)  Parameters (typed in table)   * Body weight (kg) * Body height (m) * Ventilation rate (at work) (L/min) * Ventilation rate (at rest) (L/min) * Clearance (L/min) * Work hours per week * Work days per week * Length of simulation (in days) * Exposure level (in ppm) | tkS01a\_FolderFilepaths |
| tkS07\_UniqueScenarios |
| S64\_PCtransformation |
| S44\_RF3Descriptorlist |
| tkS08\_PCtranspositionMultiple |
| tkS03\_PCAtrainingdata |
| S47\_PCinfo |
| tk09\_ADPreportInfo |
| tkS04\_Indicator |
| tkS05\_PBTKparameters |
| tkS06\_CmaxAUC |
| tkPS01\_TKprofile |
| Input | “ModelSelection.xlsx” (sheet = SelectionGeneral)  “Data\_subsetlist.xlsx” (sheet = All subsets)  “OptimizationDescriptors.xlsx” (sheet = previoiusOP) | tkPS02\_TKextremevalues |
| Output | Full applicability report  PCA plot for all PCs  Graphic presentation of PBTK prediction | tkPS03\_PCAplot |
| Comment |  | PBTK\_WorkWeek |
| Inhalation TK prediction (HFCs) | Setting | Exposure scenario defined by HFC, gender and exposure level in ppm.  Parameters (Value selection):   * Work hours per week * Work days per week * Length of simulation (in days) | tkS01b\_FolderFilepaths |
| S63\_ScenarioList |
| S64\_PCtransformation |
| tkS11\_PCtranspositionHFC |
| S44\_RF3Descriptorlist |
| tkS10\_PCAtrainingdataHFC |
| Input | “ModelSelection.xlsx” (sheet = SelectionHFCs)  “Data\_subsetlist.xlsx” (sheet = All subsets)  “HFCaveragedata.xlsx” (sheet = All”)  “OptimizationDescriptors.xlsx” (sheet = previoiusOP) | tkS13\_HFCinfo |
| S47\_PCinfo |
| tk09\_ADPreportInfo |
| tkS12\_PBTKparametersHFC |
| Output | Full applicability report  PCA plot for all PCs  Graphic presentation of PBTK prediction compared to *in vivo* data from exposure study  Numeric comparison of PBTK prediction and *in vivo* data | PBTK\_2hExposure |
| S66\_MCcombo |
| S54\_BaselineParameters |
| S67\_MCminmax |
| S69\_TimepointSelection |
| S70\_DataCompariso |
| Comment |  | tkPS03\_PCAplot |
| PS10\_TKvsEXplot |

# Evaluation work flows

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **WORKFLOW** | **NAME** | **IN-&OUT** | | **R Scripts** |
| eWF01 | RFgenerations | Setting | - | eS01.1\_FolderFilePaths |
| Input | “ModelStats.xlsx” (All RF sheets) | S31\_G1Melting |
| Output | Graphic comparison of R2 and RMSE for all three generations of RF | ePS01\_RFgenerationsPlot |
| Comment |  |  |

## WORKFLOWS for TK prediction

|  |  |  |  |
| --- | --- | --- | --- |
| Name | Function | Settings | Output |
| Inhalation TK prediction (HFCs) | Predicts TK profile for the selected HFC-exposure scenario (HFC + exposure level + gender) for the average test person of that exposure scenario, using the QSPR-predicted partition coefficients.  The predicted TK profile is visually and numerically compared to  *in vivo* data from exposure studies. The numerically evaluation includes AUCtlast, t1/2 (assuming 1. Order elimination) and an RMSE-based index. A low index indicate good agreement between prediction and *in vivo* data.  The visual comparison includes errorbars making minimum and maximum values of the experimental data as well as the minimum and maximum range of the PBTK prediction based on a pseudo Monte Carlo analysis (considering the group of test subjects as a population)  NB: Pseudo-MonteCarlo analysis can take some time to complete. | 1. Exposure scenario (defined by HFC, level and gender) 2. Number of days for the work week simulation. Default: 7 days | 1. Applicability domain report for each partition coefficient (APD-threshold, distance, prediction reliability) 2. PCA plot for each partition coefficient (visual indication of applicability domain) 3. Work-week TK profile (plot) 4. Experimental setup TK-profile compared with *in vivo* data 5. Numerical comparison of PBTK predicted profile and *in vivo* data |
|  |  |  |  |