Development Progress Valerie

Valerie de Rijk

2024-04-25

Development process V de Rijk - logbook

This document will provide details on bug fixes and other adjustments made to SimpleBox in R and Excel during my internship at the RIVM-DMG department.

Updating of documentation

My first task was to update documentation in user vignettes and Readme.md files. This is an ongoing process, but one that is important for the publication of the repositories on Github. An example of adjustments made is for the initial Readme.md file in SbooScripts.

Initial version

```
##Vignettes and their use

Getting started and participate: start.md

The workflow to use and add data and restructuring the csv file before commit: CSVdata.md

Making a variable in the SB context: FirstVars.md

Creating and checking flows AirFlow.md

The workflow to use when adding a k process relative to a compartment, species or scale:
processFlow.md

An example for adding and testing a new variable: testRainDropRadius.md
```

New updated version

```
# SimpleBox v4.0

This repository and its supporting repository [SBOO](https://github.com/rivm-syso/SBoo)

are an implementation of the [SimpleBox model](https://www.rivm.nl/publicaties/simplebox-

40-a-multimedia-mass-balance-model-for-evaluating-fate-of-chemical-substances), that was

previously in Excel. SimpleBox (SB) is a multimedia mass balance model of the so-called

'Mackay type'. It simulates environmental fate of chemicals as fluxes (mass flows)

between a series of well-mixed boxes of air, water, sediment and soil on regional,

continental and global spatial scales. The model is capable of handling chemicals,

nanomaterials and plastics.

## Installation

SB is split over two projects. In order to use SBOO you need two projects with the
```

parent folder in common. This repository and the SBOO repository thus need to be stored in the same folder. The working directory of the project should always be set to this folder. More info on the R-servers at RIVM can be found [here] (http://wiki.rivm.nl/inwiki/bin/view/StatNMod/RStudio%2BServer).

Dependencies

Packages needed for both repositories are defined in [requirements](requirements.txt)

Set-Up

This repository includes several vignettes detailing both the set-up of the [project] (/vignettes/Development/Readme.md) and theoretical implementation of the [SB Model] (/vignettes/Readme.md).

- The Development folder details the technical construction of the project and is most suitable for developers. The vignette [start](/vignettes/Development/start.md) explains the basics of object-oriented modelling and how this is implemented in this repository. Multiple other vignettes are useful if aiming to develop, such as [CSV-reordering] (/vignettes/Development/CSVdata.Rmd), [creating variables] (vignettes/Development/FirstVars.Rmd), [creating and checking flows] (/vignettes/Development/AirFlow.Rmd), [creating processes] (vignettes/Development/processFlow.Rmd) and [testing new variables] (/vignettes/Development/testRainDropRadius.Rmd).
- The vignette folder outlines the theoretical implementation of the SimpleBox model, following the structure of Chapter 3 in [Schoorl et al. (2015)] (https://www.rivm.nl/bibliotheek/rapporten/2015-0161.html)

License

*TODO add license

Testing of SimpleBox4Nano - R vs Excel with nAg_10nm

The initial test presented large absolute and relative differences.

Diffusivity

The largest differences were present between flows in air from Solid to Particulate species (e.g., aAs_AaP) and Solid to Attached species (e.g., aAs_aAA). These relative differences were present on all scales. As such, debugging was needed. The problems seemed to lie with the implementation of the HeteroAgglomeration.air flux. This is a complicated function which is dependent on other functions as well. This led to two observations:

1. The Excel Version missed multiplication with the Cunningham factor for aerosols and other particles that were not the analyzed nanoparticle.
2. The R-version had a bug in f_diffusivity function leading to the function returning solely the first part of the function.

differences.png

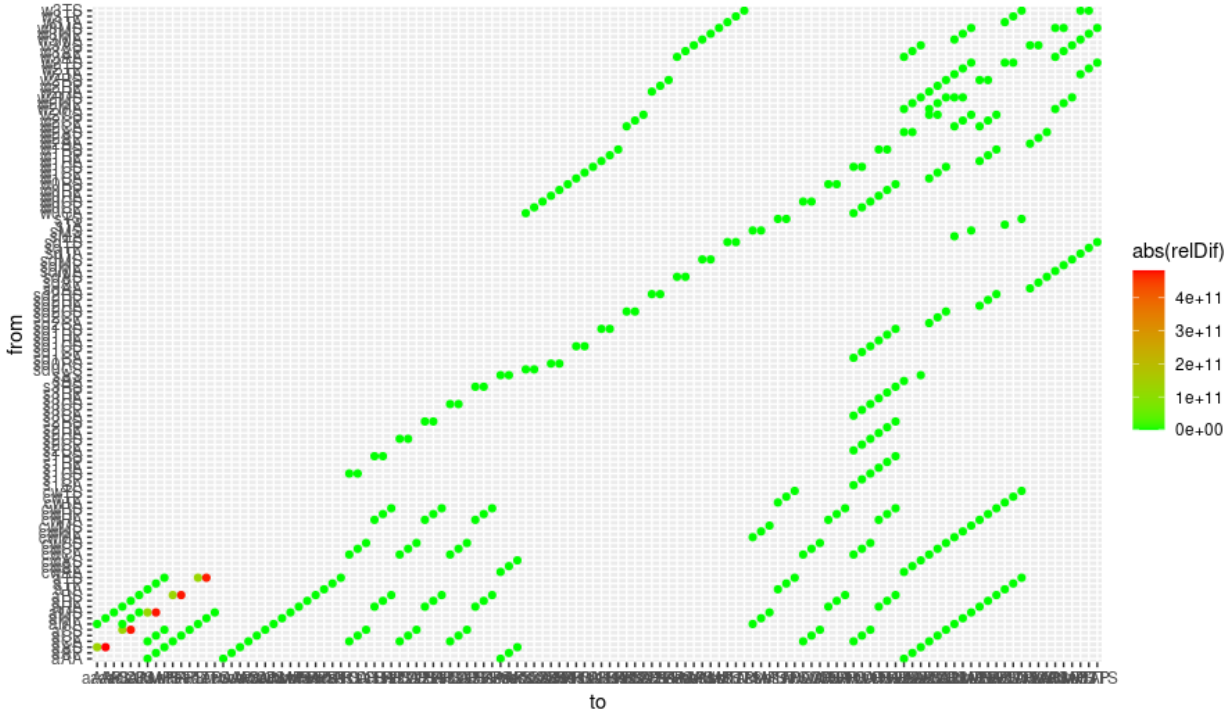


Figure 1: Relative differences between flows showing up to differences in 4e+11.

R-Version

The following code was changed from:

```
f_Diffusivity <- function(Matrix, Temp, DynVisc, rad_species, Cunningham = NULL) {
  if(!is.numeric(Temp)){
    warning("Temp missing in f_Diffusivity")
    return(NA)
  }
  if(!is.numeric(DynVisc)){
    warning("DynVisc missing in f_Diffusivity")
    return(NA)
  }
  if(!is.numeric(rad_species)){
    warning("rad_species missing in f_Diffusivity")
    return(NA)
  }
  kboltz <- constants::syms$k
  if (Matrix == "air") {
    if (is.null(Cunningham))
      Cunningham <- f_Cunningham(rad_species)
    return (kboltz*Temp*Cunningham)/(6*pi*DynVisc*rad_species)
  } else {
    (kboltz*Temp)/(6*pi*DynVisc*rad_species)
  }

}
```

to

```
f_Diffusivity <- function(Matrix, Temp, DynVisc, rad_species, Cunningham = NULL) {
  if(!is.numeric(Temp)){
    warning("Temp missing in f_Diffusivity")
    return(NA)
  }
  if(!is.numeric(DynVisc)){
    warning("DynVisc missing in f_Diffusivity")
    return(NA)
  }
  if(!is.numeric(rad_species)){
    warning("rad_species missing in f_Diffusivity")
    return(NA)
  }
  kboltz <- constants::syms$k
  if (Matrix == "air") {
    if (is.null(Cunningham))
      Cunningham <- f_Cunningham(rad_species)
    return ((kboltz*Temp*Cunningham)/(6*pi*DynVisc*rad_species))
  } else {
    (kboltz*Temp)/(6*pi*DynVisc*rad_species)
  }

}
```

This changed the output from:

```
 3.631107e-21
```

to

```
7.095256e-11
```

and the relative differences (for one flow comparison from)

Flow	R-old	Excel-Old	Abs difference	Relative Difference	R-new	Abs difference	Relative difference
aAS_aAP	2.97e-19	1.43e-7	1.43e-7	4.8e+11	1.43e-07	-8.02e-10	5.62e-03

Excel Version

I changed the Excel version to include the Cunningham factor for aerosols in all scales by adjusting the formulas. The adjustments led to the following results within the relative differences graph, effectively reducing the relative differences even further.

differences update.png

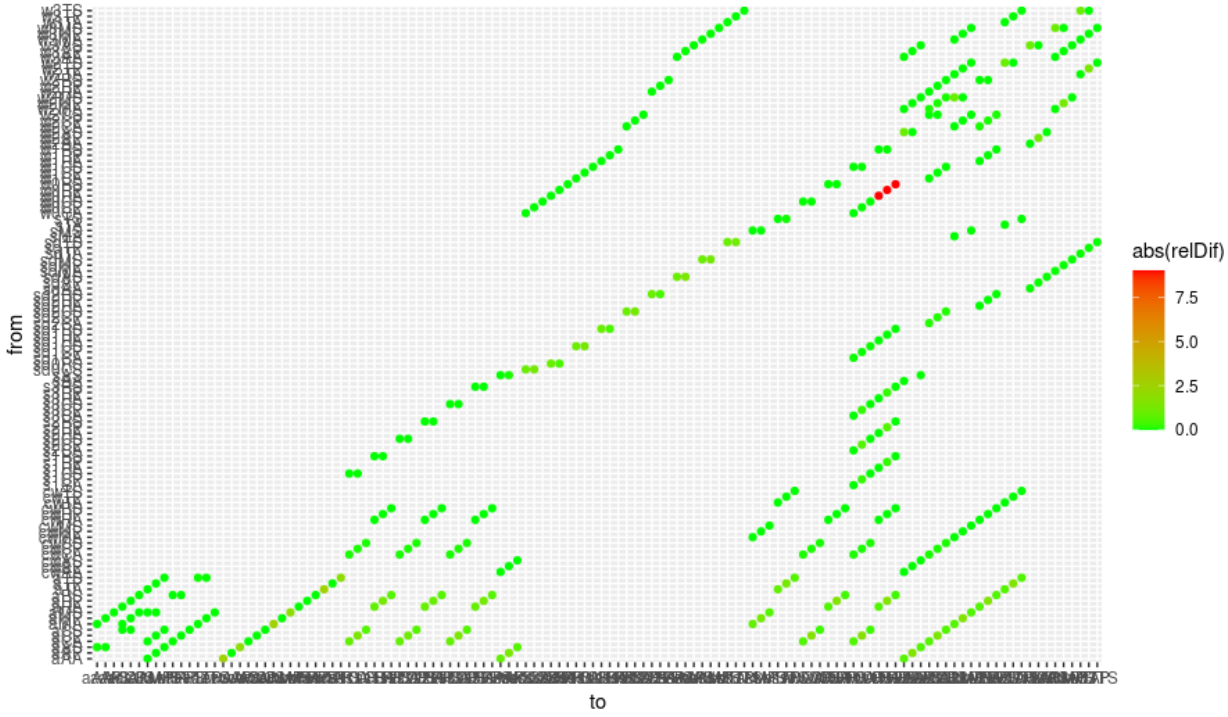
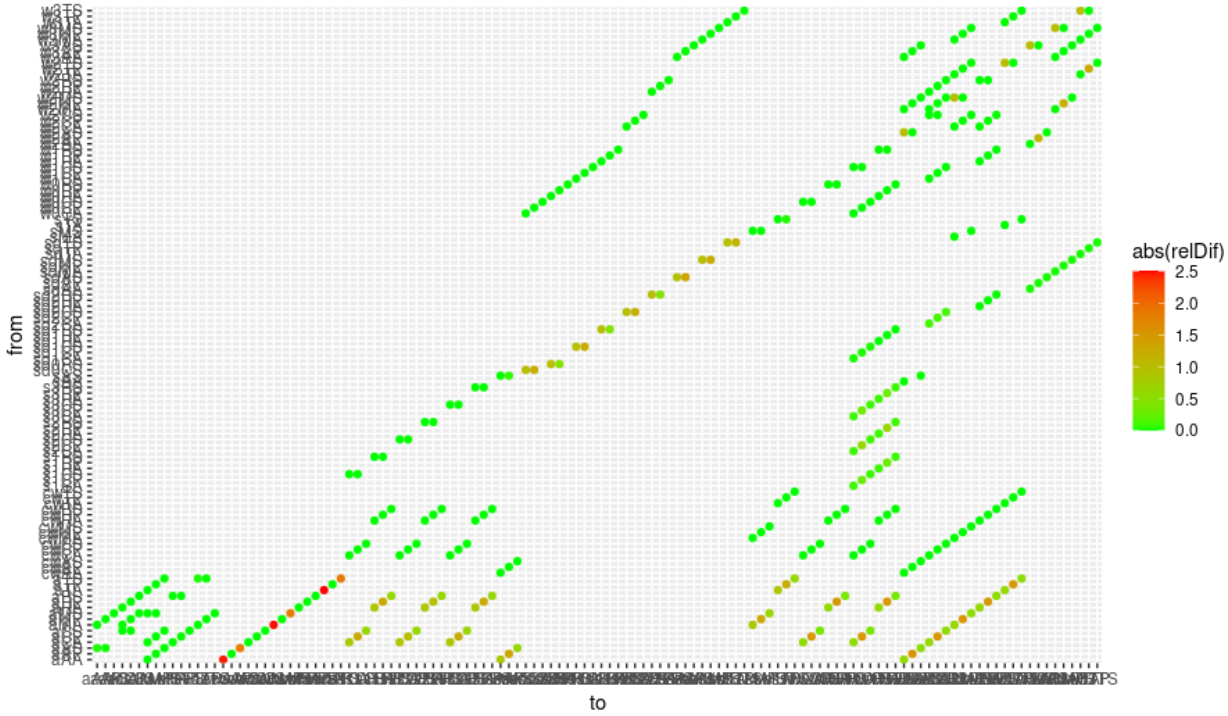


Figure 2: Relative differences after update



Relative differences and absolute differences were still very large.

Cloudwater scavenging

The next largest differences were found in the flows considering aggregated particles in air to aggregated particles in cloudwater (e.g., aTA_cwTA)

Flow	R-old	Excel-Old	Abs difference	Relative Difference	Excel-New	Abs difference	Relative difference
aTA_cwTA	2.32e-7	8.14e-7	5.81e-7	2.50	2.32e-7	5.6e-12	2.41e-5

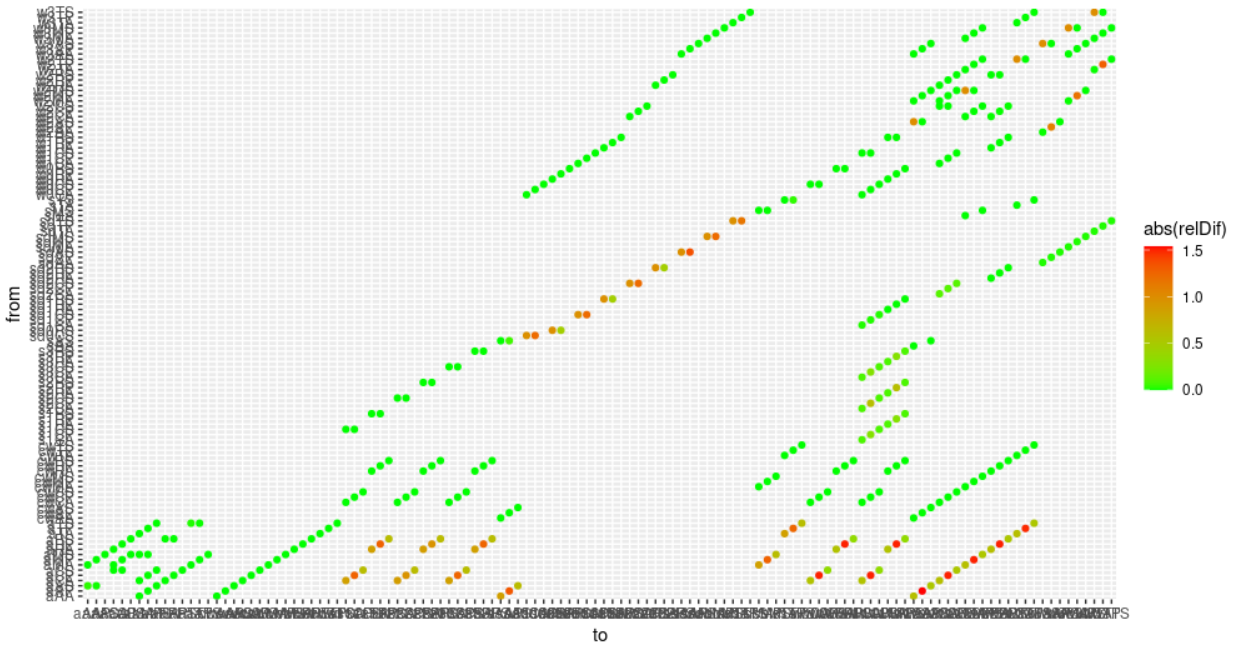


Figure 3: Relative differences after cloudwater scavenging

Dry Deposition

The SimpleBox Excel version has a different way of implementing Dry Deposition (through constants) and thus differs from the R version per definition.

Created a function of constant DRYDEPvelocity. This fixes the discrepancy between the flows for aAP and other Attached species in air. Leads to exact same flow, so model works properly.

Dry Deposition for aggregates

Settling velocity of particles is implemented different in Excel and R (Factor 18 instead of 9).

Heteroagglomerations water and soil (sediment)

The total collection coefficient (a sum of Brownian collection, interception collection and gravitational collection) was erroneous in Excel. The Brownian collection is computed correctly as described in Meesters et al. (2014) as presented in the next code chunk.

However, in Excel, the first term in BrownSFP (in most scales) replaced ASPDF by PecletNumberFP, causing a faulty value. When this was fixed, the discrepancy was also fixed. In some other cells, ASPDF was replaced by porosity, this also fixed the discrepancy.

Sedimentation

An issue considering the flow from sea to deepocean is caused by the fact that it includes both sedimentation and advection (as one flow). The R version only has a working advection flow. It should be split in two flows (advection and sedimentation). This caused the following discrepancy.

Flow	R-old	Excel-Old	Abs difference	Relative Difference	Excel-New	Abs difference	Relative difference
w2TP_w3TP	2.27e-7	5.21e-7	2.94e-7	1.29			

Heteroagglomeration

The flux from particles to aggregated particles were not consistent between Excel and R was inconsistent, see table below.

Flow	R-Old	Excel-Old	Abs difference	Relative difference	R-New	Abs difference	Relative difference
sd1CS_sd1CA	1.13e-1	2.25e-3	1.1e-1	9.8e-1	2.25e-3	2.92e-10	1.29e-7

The problem was in the fact that colloid concentration differed 1e2 and the fact that the colloid concentration in sd is multiplied by 2 in the Excel script, but not always consistently; Issue #85. Changed this in the R version by adapting the data to multiply (and change the factor 100).

*Heteroagglomeration on Arctic Scale)

In Excel, the calculation of the Peclet number differed in the heteroagglomeration flux because it referred to the wrong scale (diffusivity in Water on continental scale instead of respective global scales). Fixed this.

Correction factor for erosion

In the R version, a correction factor is applied to consider the fact that erosion mostly happens in the top soil. In Excel, this is also implemented but only for flows where both erosion and runoff are concerned. In cases where particles are attached, and so only erosion happens, this is not implemented. This is adjusted in Excel since it should be there as well.

Flow	R-Old	Excel-Old	Abs difference	Relative difference	Excel-New	Abs difference	Relative difference
s2CP_w1CP	1.09e-11	4.36e-12	6.36e-12	6.03e-1	1.01e-11	9.134260e-13	8.3e-2

Resuspension from sediment to continental sea

There was a discrepancy between the two flows. This was caused by the fact that the NET sedimentation rate for continental seas was put to 0. When this was fixed, the following flow was also fixed, but there is probably another mistake

Flow	R-Old	Excel-Old	Abs difference	Relative difference	R-New	Abs difference	Relative difference
sd2CA_w2CA	9.8e-9	8.73e-9	1.06e-9	1.09e-1	8.88e-9	1.55e-10	1.74e-2

Found other mistake. Settling velocity of Suspended Particles in Excel is calculated by =2.5/(24*3600) instead of Stokes, leading to a discrepancy.

Function correction factor water percentage of runoff to.subcompartment

In Excel, an update was created that corrected for the compartment runoff is flowing towards to in the following way: (Area_runoff_to)/ (Total water area). This is only valid for the scales Regional and Continental. This was not yet implemented in R. To this extent I created the following variable:

```
FracROWatComp <- function(all.landFRAC, all.Matrix, Matrix, SubCompartName, ScaleName) {
  compFrac <- all.landFRAC$landFRAC[all.landFRAC$SubCompart == SubCompartName & all.landFRAC$Scale ==  S
  all.landFrac <- as.tibble(all.landFRAC)
  all.Matrix <- as.tibble(all.Matrix)
  mergeddata <- left_join(
    x = all.landFRAC,
    y = all.Matrix,
    by = join_by(SubCompart))

  if ((Matrix == "water") & (ScaleName %in% c("Regional", "Continental"))) {
    # total landfrac of (fresh) water compartments
    waterFrac <- mergeddata |>
      filter(Matrix == "water" & Scale == ScaleName) |>
      summarise(waterFrac = sum(landFRAC, na.rm = TRUE)) |>
      pull(waterFrac)
    return(compFrac / waterFrac)
  } else {
    return(1)
  }
}
```

Testing of SimpleBox4Plastic