



**C O M P L E X**

Knowledge Based Climate Mitigation Systems for a Low Carbon Economy



## Software package 1: integration of models

**Date**

**Monday, 20 June 2016**

**Report Number**

**D6.10**

VERSION NUMBER:

1.0

Main Author:

Anna Shchiptsova

### DIFFUSION LEVEL – PU

PU

**PUBLIC**

RIP

**RESTRICTED INTERNAL AND PARTNERS**

RI

**RESTRICTED INTERNAL**

CO

**CONFIDENTIAL**

**Coordinator:** [Nick Winder, nick.winder@ncl.ac.uk](mailto:nick.winder@ncl.ac.uk)

### INFORMATION ON THE DOCUMENT

<b>Title</b>	D6.10 Software package 1: integration of models
<b>Authors</b>	Anna Shchiptsova

### DEVELOPMENT OF THE DOCUMENT

<b>Date</b>	<b>Version</b>	<b>Prepared by</b>	<b>Institution</b>	<b>Approved by</b>	<b>Note</b>
20.06.2016	1.0	AS	IIASA		

CONTENTS

1 Overview.....1  
2 Description.....2-4  
3 Appendix: R CMD Check results for package modelIntegration.....5-6

## 1 Overview

### URL:

<http://www.iiasa.ac.at/web/home/research/researchPrograms/AdvancedSystemsAnalysis/modellIntegration-package.html>

Available data for download contains a Windows binary package, a bundled package (usable in all platforms), files with manuals and getting started information.

### Components:

Binary package: modellIntegration\_1.0.0.zip

Bundled package: modellIntegration\_1.0.0.tar.gz

Getting started: manual.pdf,  
                  README.md,  
                  introductionModellIntegration.html

### Installation:

R ( $\geq 3.2.5$ ) is required. If installation is done from the bundled package, then build tools (Rtools compatible with R version) are required. Installation is similar to any R package installation from local resources.

### Author and maintainer:

Anna Shchiptsova, International Institute for Applied Systems Analysis (IIASA)

### Repository:

[github.iiasa.ac.at/posteriorIntegration.git](https://github.com/iiasa/posteriorIntegration) (*not publicly available*)

### Development environment:

All components are developed using R v. 3.2.5. and Rtools v. 3.2.0.1948.

The package was passed 'R CMD check'. Results of automated checking is given in the appendix.

## 2 Description

The R package 'modelIntegration' implements aggregation of several probability distributions into a single integrated one. Suppose that, several independent methods are used to observe a deterministic element and each method represents the latter as a probability distribution. Thus, we deal with a family of probability distributions providing alternative descriptions to the same object. The problem is how to combine information from the prior estimates. This package implements the posterior integration method [Kryazhimskiy, 2013]. For comparison, an implementation of simple averaging of the input distributions is added.

### Methods

The posterior integration method [Kryazhimskiy, 2013; Kryazhimskiy, 2016] is based on the assumption that model outcomes are mutually compatible, i.e., we should observe identical outcomes after the use of model ensemble. Formally, the product probability distribution of the original estimates is

$$p(z) = \frac{p_1(z) \cdot p_2(z) \dots p_n(z)}{\sum_{z' \in Z} p_1(z') \cdot p_2(z') \dots p_n(z')}$$

where  $p_1, p_2, \dots, p_n$  are prior distributions on  $Z$  associated with the methods  $1, \dots, n$ .  $Z$  is a non-empty finite set, whose number of elements is bigger than one.

Alternatively, prior estimates can be combined using simple averaging. This approach represents the distribution of the outcomes of random tests, in each of which one of the priors is chosen at random with probability  $1/n$ , and then an outcome is picked up randomly according to the probability distribution based on the chosen method. Namely,

$$p(z) = \frac{p_1(z) + p_2(z) + \dots + p_n(z)}{n}$$

### Basic usage

The main method of the modelIntegration package is 'integrate'. It can work with several representations of probability distributions. The discrete distributions are supplied through 'pdfs' argument, which supports a 'table-based' format. A continuous distribution is discretized using the cdf, supplied in 'cdfs'. In this case, a bin center equals to a value of the corresponding outcome and a bin width is determined from the subsequent outcome values in the range. The identical range of the random variables (associated with each prior distribution) is set in the 'vals' argument.

```
#### R code ####
```

```
example1 <- integrate(  
  vals = forest_npp[, 1],  
  pdfs = as.list(forest_npp[c("LEA_Tundra", "DGVM_Tundra"])))
```

```
summary(example1)
```

```
----- output -----
```

```
##      Product Average
## mean 189.29034 213.6184
## std  42.78502 74.0616
-----
```

```
example2 <- integrate(
  vals = forest_npp90[, 1],
  pdfs = as.list(forest_npp90["LEA_Tundra"]),
  cdfs = list("DGVM_Tundra" = function(x)(pnorm(x, mean = 202, sd = 52))))
summary(example2)
```

```
----- output -----
##      Product Average
## mean 183.73562 212.92005
## std  43.87124 79.16872
-----
#####
```

### Aggregated distributions

The two integrated estimates can be accessed with '*product*' and '*average*' calls correspondingly. The package also supports a summary of descriptive statistics for the integrated distributions and the priors.

```
#### R code ####
```

```
example <- integrate(c(1, 2), list(c(0.75, 0.25), c(0.75, 0.25)))
product(example)
```

```
----- output -----
##  x prob
## 1 1 0.9
## 2 2 0.1
-----
```

```
average(example)
```

```
----- output -----
##  x prob
## 1 1 0.75
## 2 2 0.25
-----
```

```
statistics(example)
```

```
----- output -----
##      P1      P2 Product Average
## mean 1.2500000 1.2500000  1.1 1.2500000
## std  0.4330127 0.4330127  0.3 0.4330127
-----
#####
```

## References

- [1] Kryazhimskiy, A.V. (2013). Posterior integration of independent stochastic estimates. IIASA Interim Report. IR-13-006.
- [2] Kryazhimskiy, A.V. (2016). Posteriori integration of probabilities. Elementary theory. Theory of Probability and its Applications, 60(1): 62-87.
- [3] Kryazhimskiy, A., Rovenskaya, E., Shvidenko, A., Gusti, M. Shchepashchenko, D. & Veshchinskaya, V. (2015). Towards harmonizing competing models: Russian forests' net primary production case study. Technological Forecasting & Social Change, 98: 245-254.

**Appendix: R CMD Check results for package modelIntegration**

```
* using log directory '****/release 1.0.0/modelIntegration.Rcheck'  
* using R version 3.2.5 (2016-04-14)  
* using platform: x86_64-w64-mingw32 (64-bit)  
* using session charset: ISO8859-1  
* using option '--no-vignettes'  
* checking for file 'modelIntegration/DESCRIPTION' ... OK  
* checking extension type ... Package  
* this is package 'modelIntegration' version '1.0.0'  
* checking package namespace information ... OK  
* checking package dependencies ... OK  
* checking if this is a source package ... OK  
* checking if there is a namespace ... OK  
* checking for executable files ... OK  
* checking for hidden files and directories ... OK  
* checking for portable file names ... OK  
* checking whether package 'modelIntegration' can be installed ... OK  
* checking installed package size ... OK  
* checking package directory ... OK  
* checking DESCRIPTION meta-information ... OK  
* checking top-level files ... OK  
* checking for left-over files ... OK  
* checking index information ... OK  
* checking package subdirectories ... OK  
* checking R files for non-ASCII characters ... OK  
* checking R files for syntax errors ... OK  
* checking whether the package can be loaded ... OK  
* checking whether the package can be loaded with stated dependencies ... OK  
* checking whether the package can be unloaded cleanly ... OK  
* checking whether the namespace can be loaded with stated dependencies ... OK  
* checking whether the namespace can be unloaded cleanly ... OK  
* checking loading without being on the library search path ... OK  
* checking dependencies in R code ... OK  
* checking S3 generic/method consistency ... OK  
* checking replacement functions ... OK  
* checking foreign function calls ... OK  
* checking R code for possible problems ... OK  
* checking Rd files ... OK  
* checking Rd metadata ... OK  
* checking Rd cross-references ... OK  
* checking for missing documentation entries ... OK  
* checking for code/documentation mismatches ... OK  
* checking Rd \usage sections ... OK  
* checking Rd contents ... OK  
* checking for unstated dependencies in examples ... OK  
* checking contents of 'data' directory ... OK  
* checking data for non-ASCII characters ... OK  
* checking data for ASCII and uncompressed saves ... OK  
* checking installed files from 'inst/doc' ... OK
```



\* checking files in 'vignettes' ... OK  
\* checking examples ... OK  
\* checking for unstated dependencies in 'tests' ... OK  
\* checking tests ... OK  
Running 'testthat.R'  
\* checking for unstated dependencies in vignettes ... OK  
\* checking package vignettes in 'inst/doc' ... OK  
\* checking running R code from vignettes ... SKIPPED  
\* checking re-building of vignette outputs ... SKIPPED  
\* checking PDF version of manual ... OK  
\* DONE  
Status: OK