The nutrient modeling R project readme

Gerald C. Nelson

July 20, 2016

# The nutrient modeling R project

The goal of this R project is to provide estimates of the nutrient consumption changes arising in scenarios that include socioeconomic and climate futures. It integrates data from

* FAO's Food Balance Sheets
* the Shared Socioeconomic Profiles used in the IPCC's Fifth Assessment Report and many other studies
* quantitative scenario modeling results from IFPRI's IMPACT model
* a lookup table that is used to convert food consumption by a representative consumer in a country (including optionally a measure of nutrient loss during food preparation) to it macro and micro constituents, and
* age- and gender-adjusted nutrient requirements.

The latest versions of the code and most of the data are at <https://github.com/GeraldCNelson/nutmod>

# Directory structure

The directory hierarchy of the project is as follows. Second level directories are in parentheses:

* data (IMPACTData) - contains .xlsx and .rds files generated by the dataPrep and dataManagement R scripts
* data-raw (FBSData, IMPACTData, NutrientData, SSPData) - data files from FAO, SSP, IMPACT, the nutrient lookup table
* R - R scripts
* results - results from the analysis, contains .xlsx and .rds files
* nutrientModeling (data, www) – code and data for the shiny app

# R code file naming conventions

All R code is in the R directory. File naming conventions are:

\* dataPrep.xxx.R - reads in the raw data from the data---raw directory and processes it into .rds (and sometimes .xlsx) files and writes these to the data or IMPACTData directory

\* dataManagement.xxx.R - reads in .rds data files from the data or IMPACTData directory, does more processing and writes writes these to the data or IMPACTData directory.

\* xxxFunctions.R - has generic functions used in the xxx R scripts.

A special functions script is nutrientModFunctions.R. In addition to R functions it holds the names of key variables such as file names and directory paths. All other scripts read this information in with functions from this script.

* nutrientCalcs.R - all final calculations are (currently) done in this script.

# Results files

All file names have a standard format - shortName, creation date, suffix. An example is metaData.2016-04-17.xlsx.

* suffix - .xlsx, .csv, or .rds (a compressed file format used in R)

## shortNames

* General - metaData, nut.requirements, food.group, staples
* requirements - req.EARxxx, req.RDA.vitsxxx, req.RDA.minrls, req.RDA.macroxxx, req.UL.vitsxxx,req.UL.minrlsxxx
* type of results - ratio of nutrient consumption by commodity/staple/food group to total consumption, ratio of nutrient consumption by commodity/staple/food group to requirement

# Variable naming conventions

* scenario - name of the IMPACT scenario, currently SSP2-GFDL, SSP2-MIROC, SSP2-NoCC
* region - the aggregation from individual countries to larger regions. Currently region\_code.IMPACT3
* year - 4 digit year with an X to start with so it is a character value. Example - X2020
* IMPACT\_code - a 4 or 5 digit code for each of the IMPACT commodities. Example - cwhea - the quantity of wheat consumed by a representative consumer
* food.group.code - the code for a food group. Example - cereals.
* staple.code - the code for a food group. Example - staple
* foodAvailpDay - the per capita quantity of an IMPACT commodity available for consumption by a representative consumer. Units - kg/day
* nutrient.code - a nutrient code combines the common name of the nutrient and the units it is in. Example - folate\_µg
* nutrient.Q - the quantity of a nutrient available from an IMPACT commodity per day
* nutrient.sum [.all, .staple, .foodGroup] - The sum of nutrient.Q for all commodities, for each staple group, for each food group.
* nutrient.ratio [.all, .staple, .foodGroup] - The ratio of a nutrient to the total consumed
* nutrient.req.ratio [.all, .staple, .foodGroup] - The ratio of a nutrient to the the daily requirement

# Processing steps

There are three levels of data processing. These can all be run automatically using the automate.R script

## Data preparation

Scripts that begin with dataPrep.xxx import data from original sources, do some manipulation convert to R formats (and usually excel files as well) and write to the data directory.

All files created in a data preparation script keep *just* the years in the keepYearList variable created in the nutrientModFunctions.R script, EXCEPT

* the FBS outputs which keep the years in keepYearList.FBS variable (to allow for 3 year averaging around the start year), and
* the population output which adds a year 0 to keepYearList for use in adding alcohol and fish to the IMPACT data.

### dataPrep.regions.R

This script contains functions to align regional aggregations of country data and writes a file with the alignments to df.regions.all.rds. The alignment is based on the ISO 3 digit codes as of November 2015. For a valid aggregation, all ISO codes must be assigned to one of the aggregations. New regional aggregations can be added in this script.

Should not need to be run unless a new regional aggregation is added.

### dataPrep.FBS.R

This script reads in the FAO Food Balance Sheet information from a zip file, keeps just the years in the variable keepYearList.FBS and deletes a lot of other data that are not needed, converts FAO country codes to ISO 3 digit codes, sums individual FBS items to the IMPACT code they are included in, and writes out results to an rds file, dt.FBSxxx. All FBS items are included.

Precursor script is dataPrep.regions.R

Should not need to be run unless a new FBS file is available.

### dataPrep.IMPACT.R

This script reads in the IMPACT data from a gdx file and writes out selected variables to a .rds file. The gdxrrw package for R is needed to run this script. It is available at this url, not from CRAN - <https://support.gams.com/gdxrrw:interfacing_gams_and_r>.

Outputs files for each of the variables in a category of variables. [varname].rds. The categories are land, region, commods, and world. CSEs are handled separately.

### dataPrep.nutrientData.R

This script reads in the nutrient lookup table (file name is held in nutrientLU) for IMPACT commodities (USDA GFS IMPACT Vx). It produces a data frame that has for 100 grams of each IMPACT commodity the amount of several nutrients adjusted for bone in to boneless and edible portion. The choice of using a cooking retention value is in nutrientCalcs.R.

### dataPrep.SSP.R

This script reads in the SSP data and renames the variables in the SSP file. It deletes all years except those in keepYearList (with year 0 added). It keeps population results only from the IIASA-WiC POP model. The output rdf file shortname is dt.SSPPopClean.

Only needs to be run when a new SSP data set becomes available.

### dataPrep.NutrientRequirements.R

This script calculates nutrient requirements for SSP age group categories. The requirements are imported in dataPrep.nutrientData.R and saved as df.nutrients. Precursor scripts are dataPrep.nutrientData.R and dataManagement.SSP.R.

and does manipulations to align the SSP population data with the nutrient requirements age and gender structure data. Outputs are nutrient requirements for a representative consumer in each country. These change by year and scenario.

## Data management

These scripts use the cleaned up data from the dataPrep scripts, does various manipulations and outputs new .rds files, and if not too big .xlsx files with the same content.

**dataManagement.SSPPop.R**

This script reads in the Shared Socioeconomic Profiles population data written by dataPrep.SSP.R. It deletes unneeded data and creates new ‘age’ group categories for pregnant and lactating women.

Currently aggregation of population data to the region is done here.

Precursor scripts are dataPrep.regions.R and dataPrep.SSP.R.

Only needs to be run when a new SSP data set becomes available.

**dataManagement.IMPACT.R**

This script reads in the IMPACT rds files prepared in dataPrep.IMPACT.R and creates the IMPACTfood data table for use in nutrientCalcs.R. It reads in dt.FoodAvail,

adds the fish and alcohol data, and pcGDPX0, PCX0 PWX0, and CSV and writes out dt.IMPACTfood.

Precursor scripts are dataPrep.regions.R, dataPrep.IMPACT.R, dataManagement.alcohol.R, dataManagement.fish.R

## Results preparation

**nutrientCalcs.R**

This script reads in the iMPACTfood data table and does some preliminary processing.

**nutrientCalcsProcessing.R**

NutrienCalcsProcessing.R writes out a large number of .rds, ,csv, and .xlsx (where file sizes are not too large) files with final results.

## Results display

**app.R**

The app.R script is in the nutrientModeling subdirectory. It follows the shiny software protocols and is used to display information for individual countries at https://nutrientmodeling.shinyapps.io/nutrientModeling/.