

Read Me: Replication Directory Overview

30 July, 2021

README OVERVIEW

Author	Logan Stundal
Date	July 26, 2021
Purpose	<p>This document explains script execution order to replicate all results in the paper / appendix as well as produce all figures and tables. Data imports and data exports are noted for each script.</p> <p>Estimation results are exported to <code>/Results/Replication-Estimates/</code> Figures are exported to <code>/Results/Replication-Figures</code> Published models were compressed and exported to <code>/Results/Published-Models/</code></p>

These replication files are provided at:

- [GitHub repository](#)
- [Harvard Dataverse](#)

Software

The primary analysis in this paper was conducted using R-INLA version 21.02.23 compiled on Feb 22, 2021. The tar.gz installation file for this version of INLA is provided here:

- <https://inla.r-inla-download.org/R/stable/src/contrib/>
- https://inla.r-inla-download.org/R/stable/src/contrib/INLA_21.02.23.tar.gz

These models were estimated with R version 4.0.4 (2021-02-15) running on Windows 10 x64 (build 19043).

Data

There are two data files stored in the `Replication/Data` directory which are used to reproduce all models or figures:

- `Replication/Data/farc_events.Rdata` contains all data necessary to replicate the analysis. This file contains an r-object called `dat` which is a named list comprising of 4 data frames: `2002-2004` , `2005-2007` , `2008-2009` , and `2002-2009` each corresponding to a temporal aggregation of the data implied by the name. These data frames each contain the following variables:

Variable Name	Description
department	Colombia Administrative level-1, Department name
municipality	Colombia Administrative level-2, Municipality name
yr_grp	Years for cross-section
cinep	FARC activity, count (CINEP source)
cinep_bin	FARC activity, binary-indicator (CINEP source)
icews	FARC activity, count (ICEWS source)
icews_bin	FARC activity, binary-indicator (ICEWS source)
icews_cinep_under	FARC activity, ICEWS underreporting relative to CINEP, binary-indicator
ged	FARC activity, count (GED source)
ged_bin	FARC activity, binary-indicator (GED source)
ged_cinep_under	FARC activity, GED underreporting relative to CINEP, binary-indicator
area_km2	Municipality area

Variable Name	Description
area_km2_ln	Municipality area, logged
distance_bogota_km_ln	Distance (km) between municipality centroid and Bogota, logged
distance_bogota_km	Distance (km) between municipality centroid and Bogota
pop_sum	Population count at municipality level
pop_sum_ln	Population count at municipality level, logged
terrain_ri_mean_m	Terrain roughness indicator, municipality average (meters)
centroid_mun_lat	Municipality centroid, latitude
centroid_mun_long	Municipality centroid, longitude

- **Replication/Data/colombia.Rdata** - an unprojected simple feature of Colombia's international border used to provide a sharply contrasting boundary in maps containing posterior mean, probability, and standard deviation estimates of the Gaussian random field.
- **Replication/Data/colombia2.Rdata** - an unprojected simple feature object of level-2 administrative boundaries (Municipalities) used for mapping purposes.

Scripts

The following sections detail how the R scripts provided in `Replication/Scripts/` reproduce the published results. These replication scripts export results as follows: -

Results/Replication-Estimates ~ contains Rdata files of estimates used to reproduce all tables and figures. - **Results/Replication-Figures** ~ contains png files for all figure exports - **Results/Published-Models** ~ contain a compressed Rdata file with pre-estimated models from executing code in `replication-models.R`

R-INLA models with random fields have large file sizes (~130Mb * 5 models * 4 time cuts = ~2.5Gb). To reduce file sizes, after estimating models in the primary `replicaton-models.R` script, quantities of interest necessary for figures or tables are extracted or computed and exported as vectors rather than complete model objects. Compressed full model results are stored as a compressed Rdata object in `Results/Published-Models`

Script Name	1-descriptive.R
Purpose	Reproduces the descriptive data overview and comparison presented in paper Section 3.2 including Figures 1 & 2 and Tables 1 & 2. This script also produces Appendix Table 1.
Imports	Data/farc_events.Rdata Data/colombia.Rdata Data/colombia2.Rdata
Exports	figure_main_1.png - Figure 1 figure_main_2.png - Figure 2
STATUS	DONE

Script Name	2-models-spde.R
Purpose	Estimates all main models reported graphically in the main draft as well as in table format in the Appendix.
Imports	Data/farc_events.Rdata
Exports	Replication-Estimates/parameter-data.Rdata - All model parameter and HPD estimates Replication-Estimates/field-data.Rdata - Projected estimates of Gaussian field mean and SD Replication-Estimates/range-data.Rdata - Spatial field decay and range estimates Replication-Estimates/pred-data.Rdata - Model predicted outcomes on probability scale Published-Models/published-models-spde.Rdata - Compressed INLA model estimates presented in the published article
STATUS	DONE

Script Name	3-models-spem.R
--------------------	------------------------

Script Name	3-models-spem.R
Purpose	Estimates discrete spatial probit error models (spem) presented in Appendix tables A3 and A4 as well as Figure A3 (spem ROC comparison) and Figure A4 (spem predicted probabilities). This script and associated export also provides standard vanilla probit estimates.
Imports	Data/farc_events.Rdata
Exports	Published-Models/published-models-spem.Rdata figure_appendix_4.png
STATUS	DONE

Script Name	4-figures_coefplots.R
Purpose	Produces figures 3 and 4 which present INLA model parameter estimates and credibility intervals for included regressors and GRMF parameters.
Imports	parameter-data.Rdata
Exports	figure_main_3.png figure_main_4.png
STATUS	DONE

Script Name	5-figures-ROCs.R
Purpose	Produces figure 5 which presents ROCs and AUC estimates comparing ICEWS and GED model performance against CINEP ground truth observations.
Imports	Data/farc_events.Rdata Replication-Estimates/pred-data.Rdata Published-Models/published-models-spem.Rdata

Script Name	5-figures-ROCs.R
Exports	figure_main_5.png figure_appendix_3.png figure_appendix_5.png figure_appendix_6.png figure_appendix_7.png
STATUS	DONE

Script Name	6-tables.R
Purpose	Produces Appendix tables A5-A8 which present INLA GRMF model estimates in table format with median and 95% HPD estimates.
Imports	parameter-data.Rdata
Exports	None. custom_table() function will return LaTeX code when evaluated in an RMarkdown document compiled to pdf.
STATUS	DONE

Script Name	7-figures-field_range-estimates.R
Purpose	Produces figures related to posterior Gaussian Markov Random Field. Figure 8 - posterior field maps and Figure 9 - Posterior spatial error range and decay.
Imports	Replication-Estimates/field-data.Rdata Replication-Estimates/range-data.Rdata Data/colombia.Rdata
Exports	figure_appendix_8.png - Figure 8 figure_appendix_9.png - Figure 9
STATUS	DONE