

Dear Mark,

Following your e-mail on November 7<sup>th</sup> 2022, we have now made the revisions and would like to submit the revised version of the article “*A Framework for Producing Small Area Estimates Based on Area-Level Models in R*” (ID 2022 - 159) for review to the **R Journal**.

We further submit a separate file that includes the responses to the raised comments and a detailed account of the changes we made to the package and paper.

The manuscript has been read and approved by all authors and has not been sent elsewhere for publication. We did our best to comply with the journal guidelines.

The R package emdi facilitates the estimation of regionally disaggregated indicators using small area estimation (sae) methods and provides tools for model building, diagnostics, presenting, and exporting the results.

The package version 1.1.7 includes unit-level small area models that assume access to auxiliary information from population micro-data. However, in developed countries like Germany and the UK, population micro-data are not publicly available and access to such data is even challenging within gatekeeper organizations. Instead, population-level auxiliary data are often only available at some aggregate level. Therefore, area-level models and various extensions have been added to the package emdi since the release of version 2.0.0. Area-level models are less demanding with respect to (a) data requirements, as only aggregates are needed for estimating the indicator of interest, and (b) computational resources.

The area-level extensions include amongst others (a) transformed area-level models with back-transformations, (b) spatial and robust estimators, and (c) area-level models that account for measurement errors. Corresponding mean squared error estimators are implemented for assessing the uncertainty of the point estimates. User-friendly tools like a stepwise variable selection function, model diagnostics, benchmarking options, high quality maps and export options of the results enable the user a complete analysis procedure – from start to finish. The functionality of the package is demonstrated by illustrative examples based on synthetic data for Austrian districts.

The latest version of package emdi 2.1.3 combines a wide range of SAE models with several tools that enable an entire data analysis and therefore expands the existing packages for the following reasons. 1. None of the existing packages contains such a variety of different area-level models. 2. emdi includes also area-level models that are not available in existing packages like adjusted variance estimation methods and transformation options for the standard FH model. 3. Package emdi offers user-friendly tools that go beyond model estimation for the new and existing methods. To the best of our knowledge package emdi is the most comprehensive and user-friendly package in the field of sae.

We are also constantly working to expand emdi to meet the current demands in the field of sae and associated with this among others official statistics. Therefore, we welcome suggestions and are happy to exchange ideas with our users via Github (see also our contributinal guidelines on Github).

Our paper “A Framework for Producing Small Area Estimates Based on Area-Level Models in R” does not only describe the extension of package emdi to area-level models but also offers an opportunity for potentially less experienced sae users to start working with area-level sae models. In particular, the reader is led systematically through the estimation process from an applied perspective including a) model building with the help of a step function based on information criteria, b) model estimation, c) model assessment and diagnostics, d) presentation of the results, and e) exporting the results to Excel or OpenDocument Spreadsheet. Tables and flow diagrams

including different model extensions and estimation possibilities provided by emdi support the prospective sae user/reader of the paper.

Furthermore, we provide a replication script of the used R code and describe the computational setup (including R version, package versions etc.) in order to guarantee reproducible results.

Yours Sincerely,

Sylvia Harmening, Ann-Kristin Kreutzmann, Sören Schmidt, Nicola Salvati, Timo Schmid