# **Text S4: Cleanup and merging operations for land boundary datasets.**

The first cleanup operation involved intersecting CAR polygons with themselves. The second operation involved deleting repeated polygons. After deleting the repeated polygons, the intersected polygons were used to erase the original dataset without the repeated polygons. After that, the intersected file was merged with the erased one, completing the empty spaces and leaving no overlaps among them.

The second operation involved removing overlaps between the datasets and merging them for improved accuracy. For this purpose, a hierarchy among datasets was established. CAR was judged to be the first dataset to be maintained as it was assumed that landowners who have CAR are more prone to comply with environmental legislation, as CAR is the primary monitoring instrument of this legislation. The other datasets are only complements contributing to identify rural properties. Many property boundaries identified on other datasets were also on CAR. Those properties that are in INCRA or Terra Legal, for example, but not in CAR, actually will be noncompliant with the Forest Code if they do not register with CAR by May 2016. See supplemental tables S1 and S2 for changes in land areas after the cleanup operations.

As a result, this combined dataset with CAR, LAU, INCRA and Terra Legal sub-datasets generated an output with 48,909 polygons and 45,605,704 hectares without overlaps (Table S3). When comparing these figures with census data for agriculture from the Brazilian Institute of Geography and Statistics (IBGE, 2006), Mato Grosso has a total of 112,987 rural establishments (including properties, settlements and land occupations) with a total area of 48,688,711 hectares. Thus this research identified 43.2% of rural establishments in number and 93.7% in area from IBGE (2006) census data, which is not spatial. This difference clearly demonstrates what is already known: this research did not include settlements and all land occupations, only private properties and land occupations geo-referenced by the Terra Legal program, which account for only 5,496 land occupations with a total area of 674,022 hectares (SERFAL, 2014).

A subsequent operation eliminated properties located in the Pantanal biome because no significant soy production was found there, which represented 1,527 polygons with a total area of 3,527,138 hectares. As a result, this analysis operated with a total of 47,382 polygons and a total area of 42,078,566 hectares in the Amazon and Cerrado biomes of Mato Grosso (Table S3).