Appendix S1: Supplementary methods

Contents

1 Accuracy of GTI cropland data layer **1**

1.1 Accuracy assessment method 1

1.2 Statistics 2

2 Mapping Rules **3**

1 Accuracy of GTI cropland data layer

The following describes an accuracy assessment of the GeoTerraImage (GTI) 2011 cropland data layer, which provided the inspiration for developing `DIYlandcover`, and the motivation for developing and testing the software over South Africa, under the aegis of the Mapping Africa project.

1.1 Accuracy assessment method

We used the same 609 1 km^2 quality control (Q) sites described in the main text to conduct a modified accuracy assessment of the GTI data. The mapping rules for GTI were different than those used for Mapping Africa, and included more categories of cropland. We therefore followed their inclusion criteria (e.g. they mapped orchards, fallow lands, and other categories that we excluded) in developing a reference for assessing the accuracy of GTI maps. We calculated the accuracy as follows:

- We divided each 1 km^2 grid cell into 25 smaller cells of 200 X 200 m (4 ha)
- We visually assessed the extent to which fields (as defined by GTI's criteria) visible in underlying Google Maps imagery (accessed using the OpenLayers plug-in in QGIS) overlapped each 4 ha subcell, to the nearest 5% of coverage. This was achieved by overlaying a second grid of cells on each 4 ha sub-cell, and recording whether this cell was >50% (field present) or <50% occupied by a field (field absent).
- We also assessed, to the nearest 5%, how much of each sub-cell was covered by GTI's cropland polygons, and recorded where and how much the coverage agreed or disagreed with the reference.

- From this, we calculated, first within each and then across all Q cells, the counts of true positives and negatives (TP and TN), and false positives and negatives (FP and FN):

$$TP = \sum_{Q=1}^{609} \sum_{i=1}^{25} fT = 1 \wedge fR = 1$$

$$FP = \sum_{Q=1}^{609} \sum_{i=1}^{25} fT = 1 \wedge fR = 0$$

$$FN = \sum_{Q=1}^{609} \sum_{i=1}^{25} fT = 0 \wedge fR = 1$$

$$TN = \sum_{Q=1}^{609} \sum_{i=1}^{25} fT = 0 \wedge fR = 0$$

Where f is the fraction (to the nearest 0.05) of each sub-cell i in each quality control cell Q, T is the test layer (the GTI cropland layer), R is reference (the visually identified field), and 1 equals presence and 0 absence.

- From this, we were then able to calculate across all 609 test cells (15,225 subcells) sites the following accuracy measures:

$$Sensitivity = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{FP + TN}$$

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

$$TSS = Sensitivity + Specificity - 1$$

With TSS being the True Skill Statistic (Allouche et al., 2006)

Back to top

1.2 Statistics

```
library(lmisc)
library(readxl)
library(data.table)
proot <- "DIYlandcover-devel"
setwd(full_path(proj_root(proot), proot))
dat <- data.table(read_excel("external/ext_data/gti_accuracy_assessmentf.xlsx"))
dat[tn == 25, flds := 0]  # count number without fields
dat[tp == 0 & fp > 0, flds := 0]
dat[tp > 0 | fn < 0, flds := 1]  # count number with fields

# All sites
sens <- dat[, sum(tp) / sum(tp, fn)]
spec <- dat[, sum(tn) / sum(fp, tn)]
acc <- dat[, sum(tp, tn) / sum(tp, fp, fn, tn)]
tss <- sens + spec - 1

# summary of tp, tn, fp, fn
tptn <- dat[, lapply(.(tp, fp, fn, tn), sum)]
setnames(tptn, toupper(colnames(dat)[2:5]))


# With fields
sensf <- dat[flds == 1, sum(tp) / sum(tp, fn)]
specf <- dat[flds == 1, sum(tn) / sum(fp, tn)]
accf <- dat[flds == 1, sum(tp, tn) / sum(tp, fp, fn, tn)]
tssf <- sensf + specf - 1

# Without fields
```

```
sensnf <- 1
specnf <- dat[flds == 0, sum(tn) / sum(fp, tn)]
accnf <- dat[flds == 0, sum(tp, tn) / sum(tp, fp, fn, tn)]
tssnf <- sensnf + specnf - 1

acctab <- rbind.data.frame("Sensitivity" = c(sens, sensf, sensnf),
                           "Specificity" = c(spec, specf, specnf),
                           "Accuracy" = c(acc, accf, accnf),
                           "TSS" = c(tss, tssf, tssnf))
colnames(acctab) <- c("All", "Fields", "NoFields")
acctab <- round(acctab, 2)
```

TP, FP, FN, TN

```
tptn
```

```
##        TP     FP     FN      TN
## 1: 2342.4 142.45 295.85 12444.3
```

Accuracy statistics

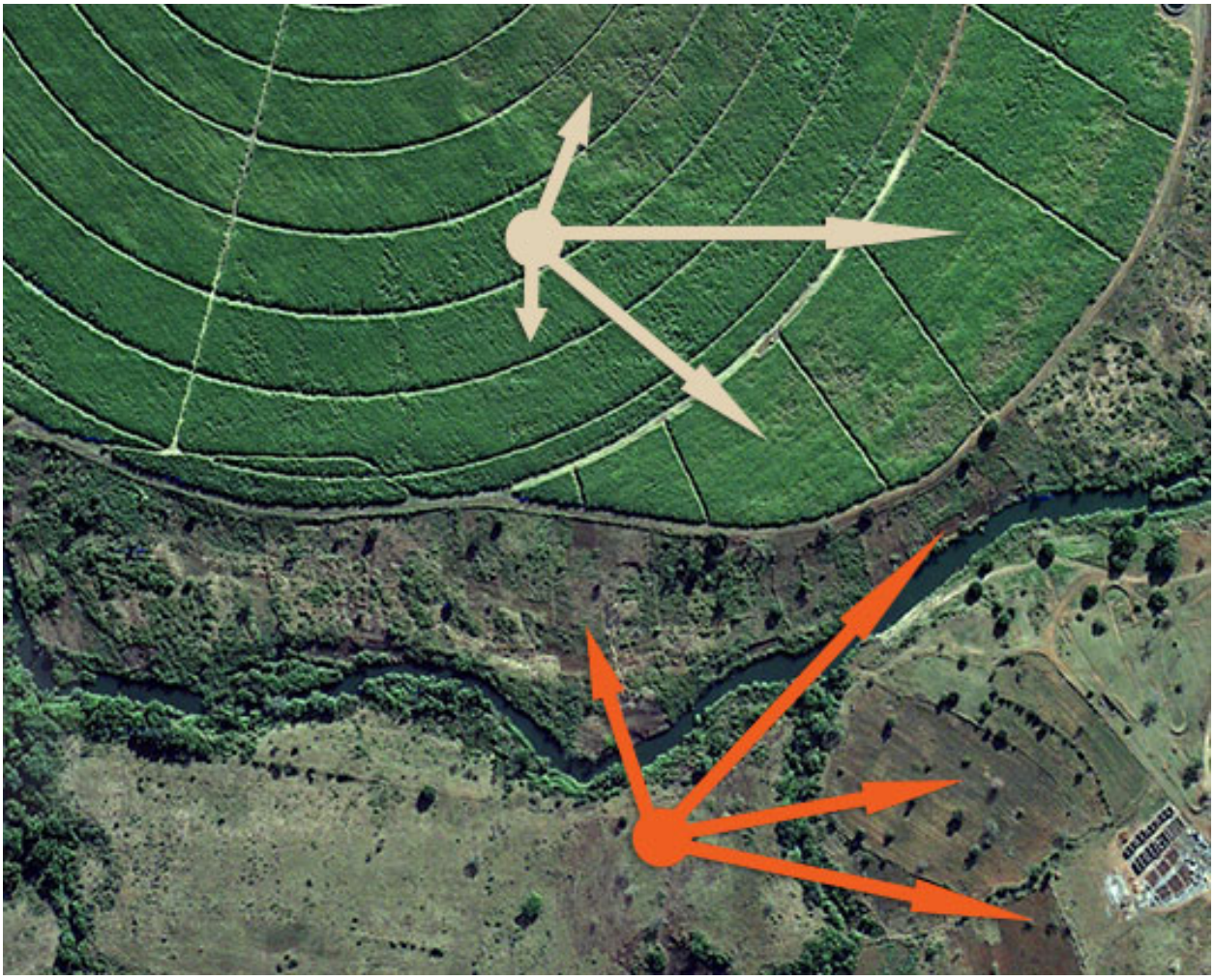
```
acctab
```

```
##              All Fields NoFields
## Sensitivity 0.89   0.93        1
## Specificity 0.99   0.97        1
## Accuracy    0.97   0.95        1
## TSS         0.88   0.89        1
```

Back to top

2 Mapping Rules

The digitizing rules for the Mapping Africa were followed in the trial run. These are illustrated in detail on the project website. The illustration for Rule 4 (below) captures the variability in the arable field class that workers were asked to map.



In the figure above, the red arrows indicate smallholder fields directly on the other side of the river from a commercial sugarcane field (tan arrows). In addition to being much smaller, the smallholder fields are less distinct at their edges, and contain remnant fruit trees, making them harder to distinguish from the surrounding vegetation.

Back to top