

1 Performance of three machine learning algorithms for predicting soil  
 2 organic carbon in German agricultural soil

3 Ali Sakhaee<sup>1</sup>, Anika Gebauer<sup>2</sup>, Mareike Ließ<sup>2</sup>, Axel Don<sup>1</sup>

4 <sup>1</sup>Thünen Institute of Climate Smart Agriculture, Braunschweig, Germany

5 <sup>2</sup>Department Soil System Science, Helmholtz Centre for Environmental Research – UFZ, Halle (Saale), Germany

6  
 7 Correspondence to: Ali Sakhaee (a.sakhaee@thuenen.de)

8 **Supplements**

9 **Table S1: The range of parameters for tuning in full dataset (AP1 and AP1L) and mineral and organic soil subsets**  
 10 **(AP2 and AP2L).**

|     | Parameter         | Full dataset | Mineral soil | Organic soil |
|-----|-------------------|--------------|--------------|--------------|
| SVR | C                 | 1-100        | 1-50         | 1-200        |
|     | epsilon           | 0-5          | 0-1          | 0-5          |
|     | gamma             | 0.001-1      | 0.001-1      | 0.001-1      |
| RF  | mtry              | 3-13         | 3-13         | 3-13         |
|     | number of trees   | 1000         | 1000         | 1000         |
| BRT | number of trees   | 100-3000     | 100-3000     | 100-3000     |
|     | shrinkage         | 0.001-0.1    | 0.001-0.1    | 0.001-0.1    |
|     | interaction depth | 1-5          | 1-5          | 1-5          |
|     | bag fraction      | 0.5-0.9      | 0.5-0.9      | 0.5-0.9      |

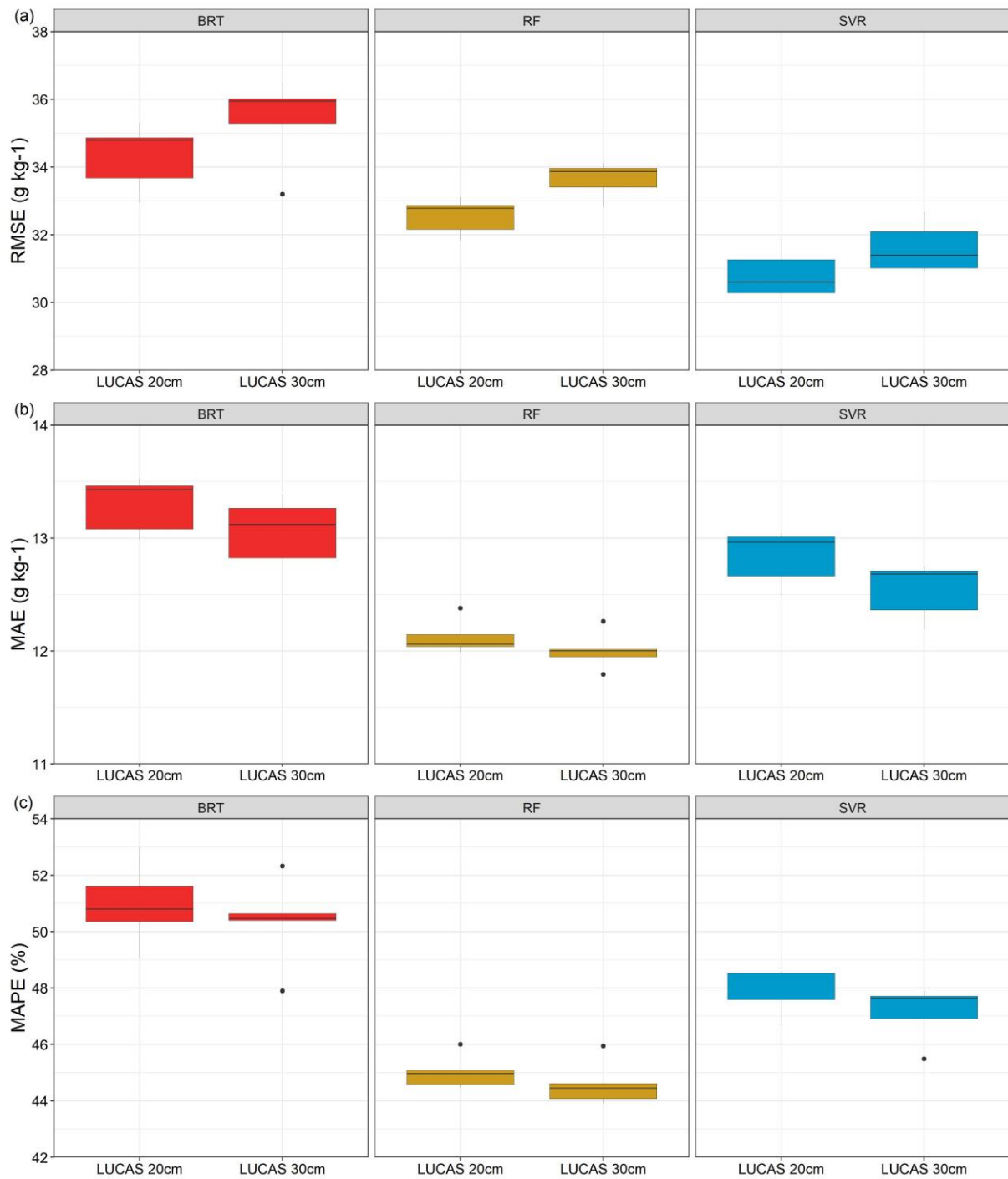
11 **Table S2: Error metrics of the algorithms: A) built on the German Agricultural Soil Inventory, B) including LUCAS**  
 12 **data in the training set.**

|   | Algorithm | RMSE  | MAE  | MAPE | Approach |
|---|-----------|-------|------|------|----------|
| A | BRT       | 32.9  | 12.4 | 50.9 | AP1      |
|   | RF        | 33.2  | 12.3 | 48.6 | AP1      |
|   | SVR       | 31.6  | 12.3 | 47.4 | AP1      |
|   | BRT       | 9.5   | 6.2  | 35.9 | Mineral  |
|   | RF        | 9.1   | 5.9  | 34   | Mineral  |
|   | SVR       | 9.2   | 5.8  | 31.8 | Mineral  |
|   | BRT       | 107   | 90.4 | 48.5 | Organic  |
|   | RF        | 106.1 | 89.3 | 48.2 | Organic  |
|   | SVR       | 101.7 | 86.9 | 45.6 | Organic  |
|   | BRT       | 22    | 9.1  | 36.3 | AP2      |
|   | RF        | 21.7  | 8.8  | 34.5 | AP2      |
|   | SVR       | 21    | 8.6  | 32.3 | AP2      |
|   | Algorithm | RMSE  | MAE  | MAPE | Approach |
| B | BRT       | 31.3  | 11.8 | 47.4 | AP1L     |
|   | RF        | 32.5  | 12.1 | 46.8 | AP1L     |

|  |     |       |      |      |         |
|--|-----|-------|------|------|---------|
|  | SVR | 32.6  | 12.3 | 46.4 | AP1L    |
|  | BRT | 9.4   | 6.2  | 35.6 | Mineral |
|  | RF  | 9.1   | 6    | 34.6 | Mineral |
|  | SVR | 9.1   | 5.8  | 31.7 | Mineral |
|  | BRT | 105.4 | 88.4 | 45   | Organic |
|  | RF  | 104.1 | 86.2 | 43.5 | Organic |
|  | SVR | 100.2 | 81.7 | 40.2 | Organic |
|  | BRT | 21.7  | 9    | 36   | AP2L    |
|  | RF  | 21.4  | 8.7  | 34.9 | AP2L    |
|  | SVR | 20.7  | 8.4  | 31.9 | AP2L    |

13 Table S3: List of covariates, their abbreviations and reference.

| SCORPAN ID | Covariates   | Abbreviation |
|------------|--|--------------|
| S          | Net erosion  | Net-Ero      |
|            | Available water capacity   | AWC          |
|            | Total nitrogen   | TN           |
|            | pH   | pH           |
|            | Soil organic map   | Peat         |
|            | Clay content   | Clay         |
| C          | Multi-annual grid of annual sunshine duration over Germany                             | Sun-Dur      |
|            | Multi-annual grids of number of summer days over Germany                               | Summ-D       |
|            | Multi-annual grids of monthly averaged daily minimum air temperature (2m) over Germany | Min-temp     |
|            | Multi-annual grids of precipitation height over Germany                                | Precip       |
| O          | Landuse  | DLM          |
| R          | Digital elevation model  | EU-DEM       |
|            | Slope  | Slope        |
|            | Aspect north south direction   | Aspect-NS    |
|            | Aspect east west direction   | Aspect-EW    |
|            | Plan Curvature   | Plan-Curv    |
|            | Profile curvature  | Prof-Curv    |
|            | Topographic wetness index  | TWI          |
|            | Geomorphographic map   | GMK          |
| P          | Large-scale landscape unit map (Bodengrosslandschaft)                                  | BGL          |
|            | Large-scale soil climate region map (Bodenklima)                                       | Bod-klim     |
|            | Hydrological unit  | HUK-HE       |
| N          | X coordination   | x            |
|            | Y coordination   | y            |



14

15 **Figure S1: Boxplot comparing algorithm performance. LUCAS at the original sampling depth (20 cm) versus LUCAS**  
 16 **with depth extrapolated (30 cm): A) RMSE (g kg<sup>-1</sup>), B) MAE (g kg<sup>-1</sup>) and C) MAPE (%).**

17

18