

Research Article

Implementation of Proximal and Remote Soil Sensing, Data Fusion and Machine Learning to Improve Phosphorus Spatial Prediction for Farms in Ontario, Canada

A. Lachgar ^{1,2}, D. J. Mulla ^{1*} and V. Adamchuk ³

¹ University of Minnesota, Saint Paul, USA;

² Mohammed VI Polytechnic University, Ben Guerir, Morocco;

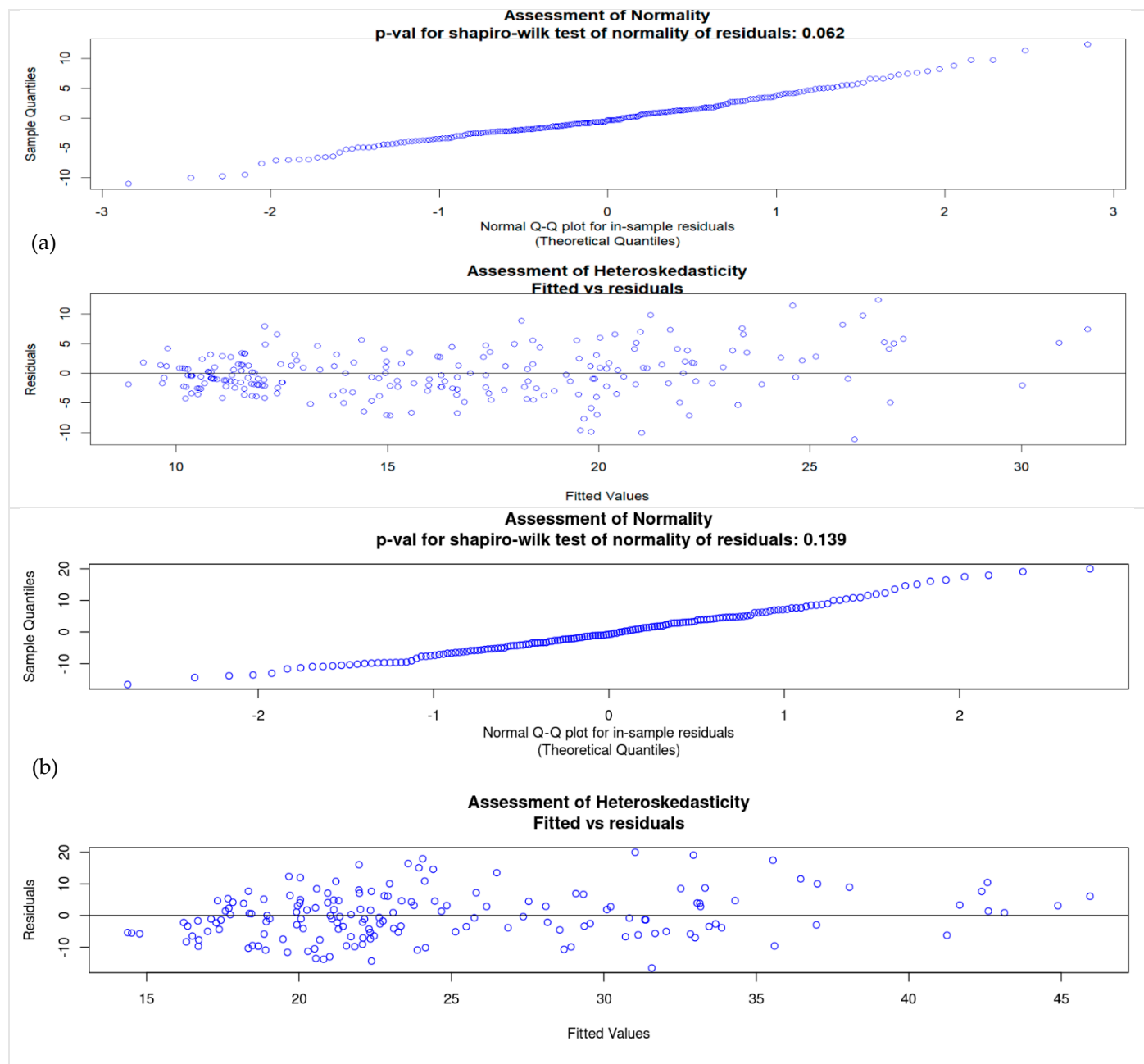
³ McGill University; Ste-Anne-de-Bellevue, Quebec, Canada

* Corresponding author: mulla003@umn.edu

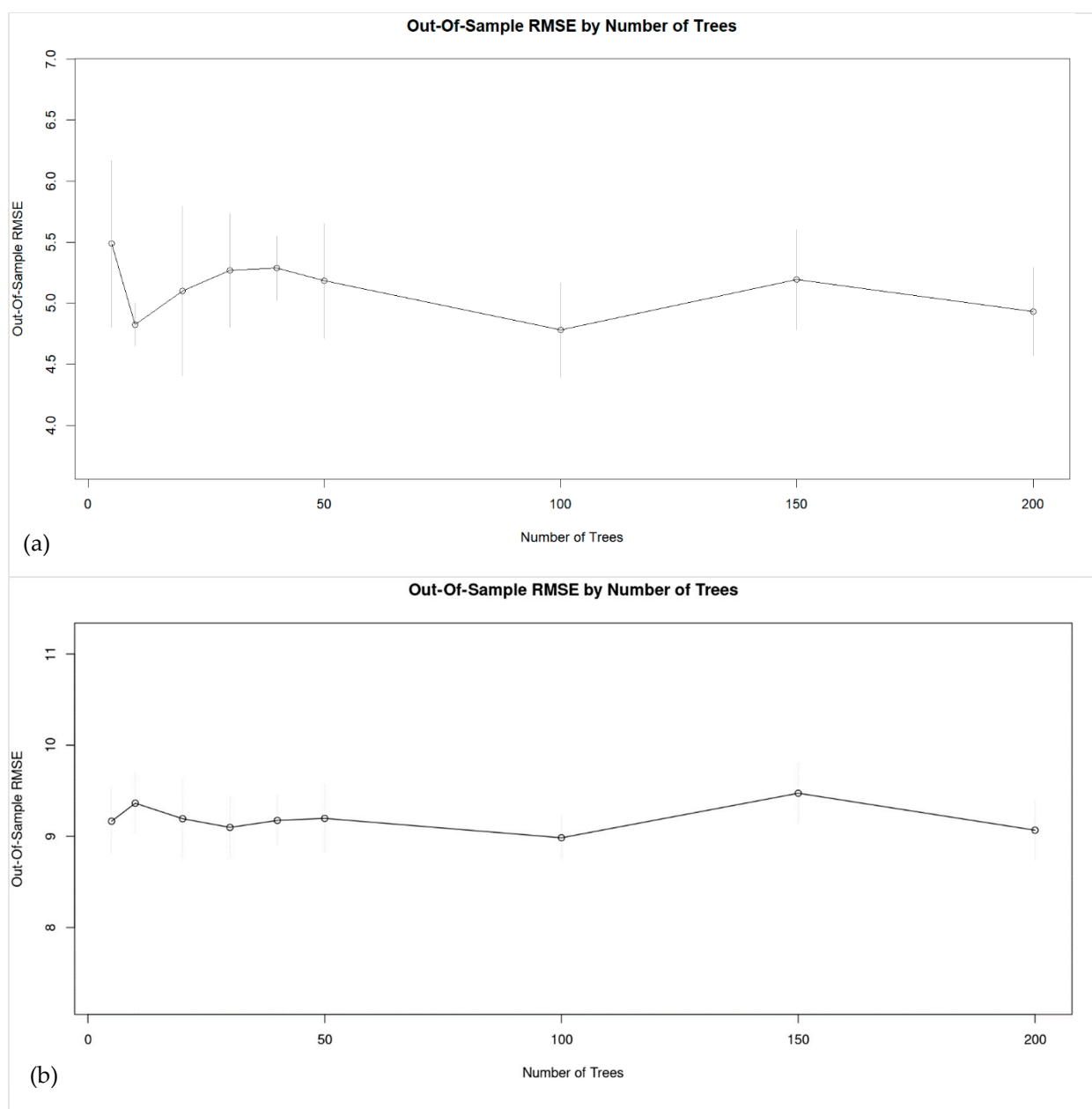
Supplementary Materials

Supplementary Table S1. Details of the fields considered in this research project, including size of the fields, number of soil samples, Proximal Soil Sampling (PSS) variables, number of PSS readings and acronyms for predictive models based on the geographic zone where fields are located. Dualem measured horizontal coplanar geometry-HCP and perpendicular geometry-PRP ECa, each at two depths (1 and 2).

| Ontario Zones | Field ID | Area(ha) | Soil samples | Soil samples/ha | PSS sensor | PSS variables | Number of PSS readings |
|---|----------|----------|------------------------------|-----------------|--------------------------|------------------------------|----------------------------------|
| UTM 18N zone (Greeley, eastern ON region) | NX | 44.47 | 74 (random-directed 2017) | 1.7 | DUALEM-21s-2017 | HCP1 HCP2 PRP1 PRP2 | 4850 4850 4850 4850 |
| | | | | | Trimble AgGPS 542-2017 | Topographic attributes | 4376 |
| | SH | 39.80 | 64 (random-directed 2016) | 1.6 | DUALEM-21s-2016 | HCP1 HCP2 PRP1 PRP2 | 9337 9337 9337 9337 |
| | | | | | Trimble AgGPS 542-2016 | Topographic attributes | 9429 |
| | VN | 20.43 | 51 (regular grid-2014) | 2.5 | DUALEM-21s (2016) | HCP1 HCP2 PRP1 PRP2 | 5073 5073 5073 5073 |
| | | | | | Trimble AgGPS 542 (2016) | Topographic attributes | 5181 |
| | KN | 48.14 | 119 (regular grid-2014) | 2.5 | DUALEM-21s (2016) | HCP1 HCP2 PRP1 PRP2 | 11427 11427 11427 11427 |
| | | | | | Trimble AgGPS 542 (2016) | Topographic attributes | 11662 |
| | HN | 39.37 | 99 (regular grid-2014) | 2.5 | DUALEM-21s (2016) | HCP1 HCP2 PRP1 PRP2 | 20129 20129 20129 20129 |
| | | | | | Trimble AgGPS 542 (2016) | Topographic attributes | 20090 |
| UTM17N zone (Guelph, western ON region) | RB | 30.32 | 72 (regular grid-2014) | 2.4 | DUALEM-21s (2015) | HCP1 HCP2 PRP1 PRP2 | 18524 18524 18524 18524 |
| | | | | | Trimble AgGPS 542 | Topographic attributes | 20813 |
| | LN | 21.14 | 62 (regular grid-2015) | 1.5 | DUALEM-21s (2015) | HCP1 HCP2 PRP1 PRP2 | 6931 6931 6931 6931 |
| | | | | | Trimble AgGPS 542 (2015) | Topographic attributes | 7110 |
| | Total | 243.67 | 479 | 1.9 | - | - | 704036 |



Supplementary Fig S1. Normality test of errors using Q-Q plot and the Shapiro-Wilk test, residual plot to assess heteroscedasticity within (a) eastern Ontario (UTM 18N) zone and (b) western Ontario (UTM 17N) zone. There is little evidence of the errors violating normality and homoscedasticity.



Supplementary Fig. S2. Optimization of the number of trees with the out-of-sample RMSE for (a) eastern Ontario (UTM 18N) zone and (b) western Ontario (UTM 17N) zone.

Supplementary Table S2. Data splitting for regional and global models

| | Total number of observation points | Train (75%) set for calibration | Test (25%) Set for validation |
|-------------------------------------|------------------------------------|---------------------------------|-------------------------------|
| Eastern region models | 296 | 224 | 72 |
| Western region models | 219 | 165 | 54 |
| Overall dataset for “global models” | 515 | 386 | 129 |

Supplementary Table S3. Data splitting for in-field models

| | Total number of observation points | Train (75%) set for calibration | Test (25%) Set for validation |
|-----------|------------------------------------|---------------------------------|-------------------------------|
| NX models | 70 | 54 | 16 |
| SH models | 61 | 47 | 14 |
| VN models | 47 | 37 | 10 |
| KN models | 117 | 89 | 28 |
| LN models | 51 | 39 | 12 |
| RB models | 69 | 54 | 15 |
| HN models | 99 | 76 | 23 |