

Evaluation of forage grass biomass estimation models using multispectral drone imaging across multiple sites

Raquel A. Oliveira ^{1*}, Roope Näsi ¹, Jonas Edström ^{1,3}, Joel Pitkänen ^{1,4}, Panu Korhonen ², Oiva Niemeläinen ², Niko Koivumäki ¹, Jere Kaivosoja ², Eija Honkavaara ¹

¹ Dept. of Remote Sensing and Photogrammetry, Finnish Geospatial Research Institute (FGI) in National Land Survey of Finland, Vuorimiehentie, 02150, Espoo, Finland - (raquel.alvesdeoliveira, roope.nasi, niko.koivumaki, eija.honkavaara)@nls.fi

² Natural Resources Institute Finland (Luke), Finland – (panu.korhonen, jere.kaivosoja)@luke.fi, oivaniemelainen@gmail.com

³UPM Energy, Helsinki, Finland, jonas.edstrom@upm.com

⁴Aiforia Technologies Plc, Helsinki, Finland, joel.pitkanen@aiforia.com

Keywords: Multispectral, Unmanned Aerial System (UAS), Grass, Biomass, Machine Learning.

Abstract

Sustainable grassland management practices enhance key ecological functions, including carbon sequestration, biodiversity conservation, and the maintenance of soil fertility essential for climate change mitigation. Accurate and reliable estimation of grass biomass is essential for decision making on harvesting time and rate of fertilizer application. Remote sensing and data analysis technologies offer unprecedented opportunities for monitoring grassland dynamics, yet methodological challenges persist in generalizing remote sensing-based models for different growths and different areas. This study investigates the estimation of grass biomass of different growth stages during two years using multispectral UAS-based remote sensing. A leave-one-out cross-validation was conducted using five harvest datasets to train and test random forest (RF) and partial least squares regression (PLSR) models, assessing estimation accuracy within individual sites. This was followed by a cross-site evaluation, where models trained using data from other locations were tested on each harvest date to evaluate model generalizability. The estimation models within the Maaninka site yielded at the best NRMSEs 14.7%, but exceeded 55% on two cutting dates. Incorporating data from multiple sites improved generalization or maintained similar accuracy across test dates. The findings indicated that using data from various locations can improve model stability, especially in cases where local data does not provide strong predictive information.

1. Introduction

Grasslands are critical ecosystems that contribute to key ecological functions, including carbon sequestration, biodiversity conservation, and the maintenance of soil fertility. Implementing sustainable grassland management practices enhances these functions, thereby supporting climate change mitigation and promoting long-term ecosystem stability. Agricultural grasslands—particularly those used for harvested forages—provide a significant portion of the feed for livestock in Northern Europe.

Remote sensing technologies have the potential to provide tools for estimating forage yield. Taking advantage of 3D information acquired from cameras (e.g. Patel et al. 2024) or LiDAR (Hütt et al. 2024) utilizing Uncrewed Aircraft System (UASs) as a platform is an option to estimate biomass and yield in grasslands. Furthermore, multiple studies have utilized spectral information including multispectral (e.g. Hernandez et al. 2024) and hyperspectral (Franceschini et al. 2022, Oliveira et al. 2020, Oliveira et al. 2024, Wengert et al. 2022) datasets along with diverse modelling techniques to estimate forage biomass, the resulting models are often site-specific. Since the growing conditions varies, it is still challenge to obtain a model which generalizes across different locations or time periods with varying environmental conditions (Bazzo et al., 2023). Grass type and harvest timing influence biomass prediction (Viljanen et al., 2018). Lussem et al. (2020) analysed a multi-year dataset to predict forage yield based on plant height derived from low-cost UAV imagery. Wengert et al. (2022) has investigated the biomass estimation using eight multiple grass sites in natural grasslands in Germany using data from hyperspectral imaging data collected for different harvest dates of same year. They

verified that the camera had potential in the application, but more studies are needed considering spectral range and more growth periods.

This study investigates the estimation of biomass for different growth stages/cuts and sites, utilizing UAS multispectral remote sensing data features—particularly vegetation indices—to identify those that are most optimal and stable across different harvest times. Data from tree experimental sites of grass in two years (2021 and 2022) were utilized. To evaluate within-site estimation accuracy of partial least square regression (PLSR) and random forest (RF), a leave-one-out cross-validation was performed where, in each iteration, models were trained on four harvest datasets of the same area and tested on the remaining one. Additionally, a cross-site evaluation was performed by training models on all data from other locations leaving the test date from the target site out, enabling assessment of model transferability and comparison between within-site and cross-site prediction performance.

2. Materials and Methods

2.1 Study area and Datasets

Three experimental grass trial fields located Finland were used (Figure 1). **Isokyrö** trial consisted of 240 test plots (Figure 1a), with two different species of fescue. Most of the test plots (228) consisted of tall fescue, while there were only 12 test plots of meadow fescue. The test plots were organized in four columns with 60 test plots in each column. The test plots with meadow fescue were randomly located within the field. **Ylistaro** comprised 54 test plots (Figure 1c), out of which 33 were fescue

plots and 21 were timothy plots. The sites Isokyrö and Ylistaro were separated by approximately 10 km, whereas Maaninka site was located about 250 km from both. **Maaninka** field included 60 plots with dimensions of 1.5 x 8 m and each plot separated by a similar sized cover plot (Figure 1b). The plant material was pure timothy (*Phleum pratense* L. cultivar ‘Nuutti’) cultivated as typically for silage grass. Different nitrogen treatments were applied to each plot to generate greater variation between the plots.

In-situ data biomass (fresh yield - FY) measurements were performed across June–August of 2021 and 2022 (Table 1). These dates corresponded to different growth periods in Finland. Table 1 shows the mean and standard deviation of the plots fresh yield measurements from three different study areas—Maaninka, Isokyrö, and Ylistaro—collected across multiple grass cuts and years.

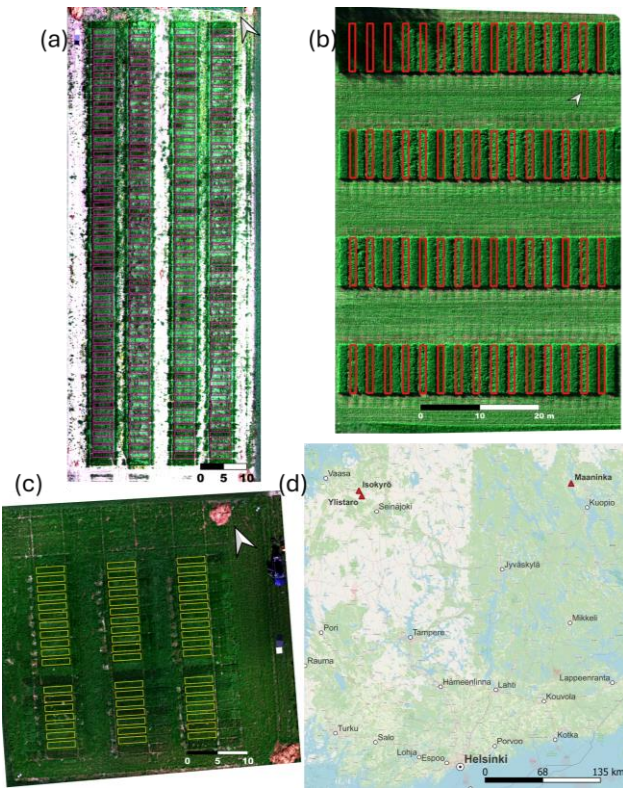


Figure 1: Trail sites (a) Isokyrö, (b) Maaninka and (c) Ylistaro. (d) Location of the study areas, data by OpenStreetMap.

Isokyrö has two datasets from 2021, showing a substantial increase in yield from the second (18,115 kg/ha) to the third cut (28,621 kg/ha). Ylistaro, with three harvest records, presented its highest yield in the first cut of 2022 (31,588 kg/ha), the highest single value across all sites, while the lowest was in the second cut of 2022 (13,961 kg/ha). Maaninka site has the most datasets, with first and third cut of 2021, first, second cut and third cut of 2022, thus, it was selected to be the site used for evaluating different setups of the regression models.

Maaninka FY ranged from a high of 21,947 kg/ha during the first cut of 2022 to a low of 8,071 kg/ha during the second cut of 2022. During these cuts, many plots exhibited high biomass

values, while a smaller number showed lower biomass due to the absence of nitrogen fertilization (Figure 2). The standard deviation of FY in the first cuts were greater than in the second and third cuts, primarily due to the increased variability introduced by plots with reduced or no nitrogen application. In contrast, FY in the second and third cuts was more uniformly distributed and generally lower in magnitude (Figure 2).

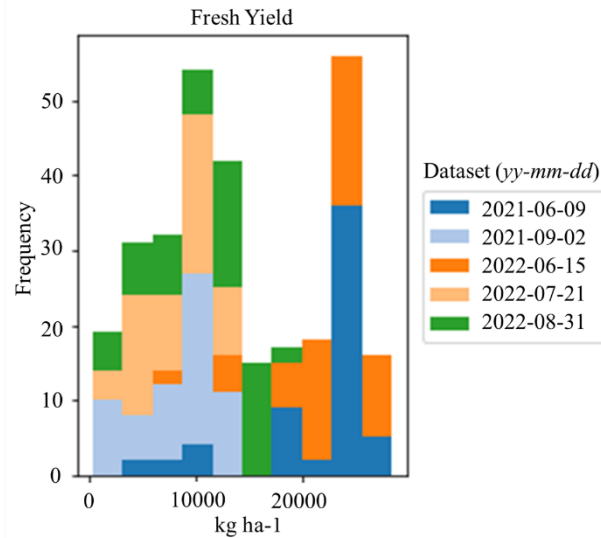


Figure 2: Histogram of fresh yield in-situ data measurement for each date.

The UAS campaigns for high-resolution remote sensing data acquisition were carried out using a DJI Matrice 300 RTK (M300), carrying a multispectral Micasense Altum camera, with 2064 × 1544 pixels images in five VNIR bands: blue (475 nm), green (560 nm), red (661 nm), red-edge (717 nm), and NIR (842 nm), and an additional thermal band at 160 × 120 pixels for a wavelength of 11,000 nm, which was not utilized. The flights were conducted from approximately 100 m above ground level. Flight dates for aerial monitoring were closely preceded or coincided with the harvest dates.

	Growth	Date harvest	FY (Kg/ha)		Flight date
			Mean	Std	
Maaninka	First	2021-06-10	21554	5849	2021-06-09
	Third	2021-09-02	8424	3825	2021-08-26
	First	2022-06-15	21947	4749	2022-06-14
	Second	2022-07-21	8071	3327	2022-07-20
	Third	2022-08-31	11169	4971	2022-08-30
Isokyrö	Second	2021-07-21	18115	2500	2021-07-21
	Third	2021-09-16	28621	2461	2021-09-16
Ylistaro	Third	2021-09-16	18129	4215	2021-09-16
	First	2022-06-16	31588	4815	2022-06-16
	Second	2022-09-07	13961	2538	2022-09-07

Table 1: Harvest dates, mean and standard deviation of fresh yield (FY kg /ha) and flight dates for each study site.

2.2 Data process and analysis

The geometric and radiometric processing of datasets were executed in Agisoft Metashape 1.7 software, resulting on approximate 3–4 cm spatial resolution multispectral reflectance orthomosaics. Ground control points measured with RTK GNSS receiver of approximated 2.5 cm accuracy were utilized to ensure geometric accuracy among the multitemporal datasets. The radiometric calibration and reflectance transformation were performed considering the Micasense calibration parameters, and its reflectance panel and irradiance sensor.

Several vegetation indices (VIs) (see Oliveira et al., 2024) were computed from the multispectral orthomosaics. The Pearson's Correlation Coefficients (PCC) between the biomass and VIs were analysed. To enhance model robustness under these temporally variable conditions, only the most predictive VIs—identified through the correlation analysis—were selected as input features for the estimation models.

To evaluate the potential for developing models that generalize across varying temporal conditions and sites, Random Forest (RF) and partial least squares regression (PLSR) were applied to estimate FY using the datasets collected from different harvests and across two growing seasons. First, all features were standardized to have zero mean and unit variance. The number of trees in the RF was 500 and the ratio of features chosen at each node was 0.3. The number of components in the PLS regression was 3.

First tests were performed using only data from the Maaninka study site. From the five grass harvest datasets—two from the 2021 season and three from 2022, a leave-one-out cross-validation approach was used. In each iteration, four of the five harvest datasets from Maaninka were used to train the PLSR and RF regressors, while the remaining dataset (corresponding to a single Maaninka cut date) was held out for testing. This procedure was repeated until each date had been used once as the test set. The estimation accuracy was quantified coefficient of determination (R^2) and Root Mean Square Error (RMSE) and Normalized Root Mean Square Error (NRMSE) (Table 2).

To complement this within-site validation, a second evaluation setup was implemented to assess model transferability across locations. In this cross-site experiment, models were trained on all available data from all sites (Isokyrö and Ylistaro and Maaninka) except one test date from Maaninka. The trained models were evaluated on each Maaninka harvest date, allowing for a comparison of within-site and cross-site model performance.

3. Results

Figure 2 presents the absolute PCC values between the VIs and FY for the Maaninka datasets, and Figure 3 presents the average PCCs for each site. Several vegetation indices have a high correlation with FY ($PCC > 0.8$) in Maaninka. There were, however, some vegetation indices that only correlated well with FY for specific growth stage. The best performing vegetation indices based on the PCC from Maaninka field trial for each cut were BNDVI, DATT1, GBNDVI, GNDVI1, NDRE, OSAVI, RDVI1 and SAVI (Figure 2). The PCCs were lower for Isokyrö and Ylistaro compared to Maaninka, the highest PCCs for Ylistaro were ARVI, DATT1, NDRE, OSAVI, SAVI, and TCI. The 8 VIs with the highest PCCs from Maaninka area were utilized to build the regression models.

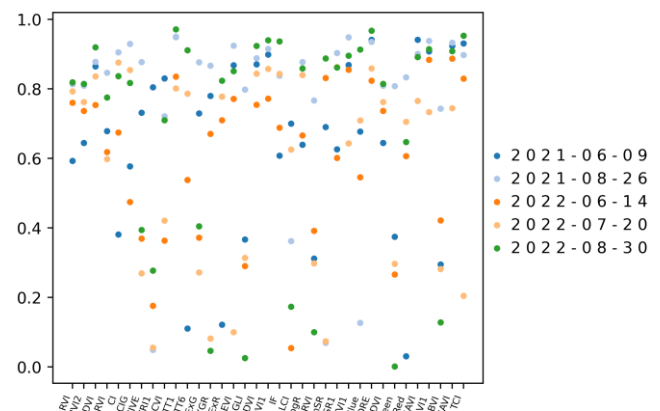


Figure 3. Absolute correlation coefficient between each vegetation index and fresh yield (FY) in Maaninka field trial in different times.

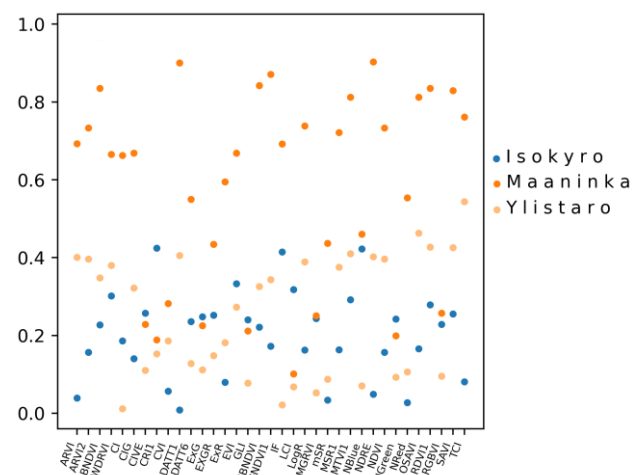


Figure 4. Average correlation coefficient between each vegetation index and fresh yield (FY) in cut datasets per field trial.

The predicted FY accuracies for both RF and PLSR in the within-site test are presented in the Table 2 and the across-sites performance is presented in Table 3. Figure 3 presents scatterplots of predicted versus in-situ measured FY values for each iteration of the cross-validation performed using RF. The plots are arranged chronologically based on the harvest date used as the test set, starting with the first harvest from 2021. Predictions generated using within-site are displayed in the left column, while those from the across-sites are shown in the right column.

The best results within Maaninka site were achieved for the first harvest in 2021, with NRMSE 14.71% using RF, the RF model showed a good fit for the lower FY values (< 12000 kg/ha), while predictions for yields above 17500 kg/ha showed both over- and underestimations (Figure 3). In contrast, the second cut of 2022 (2022-07-20) model had the worst results (RF NRMSE = 92.41%, PLSR NRMSE = 154.96%). For the second and third cuts of 2022, both RF and PLSR achieved similar results, although with high NRMSE for the third cut (NRMSE over 55%).

	Trained on 4/5 Maaninka cut					
	RF			PLSR		
Test Data (Maaninka)	RMSE kg /ha	NRMSE %	R ²	RMSE kg /ha	NRMSE %	R ²
2021-06-09	3180	14.71	0.71	3860	17.85	0.57
2021-08-26	2069	24.56	0.7	3197	37.95	0.29
2022-06-14	6175	28.13	-0.72	6447	29.37	-0.87
2022-07-20	7459	92.41	-4.11	12507	154.96	-13.37
2022-08-30	6230	55.78	-0.6	6169	55.23	-0.57

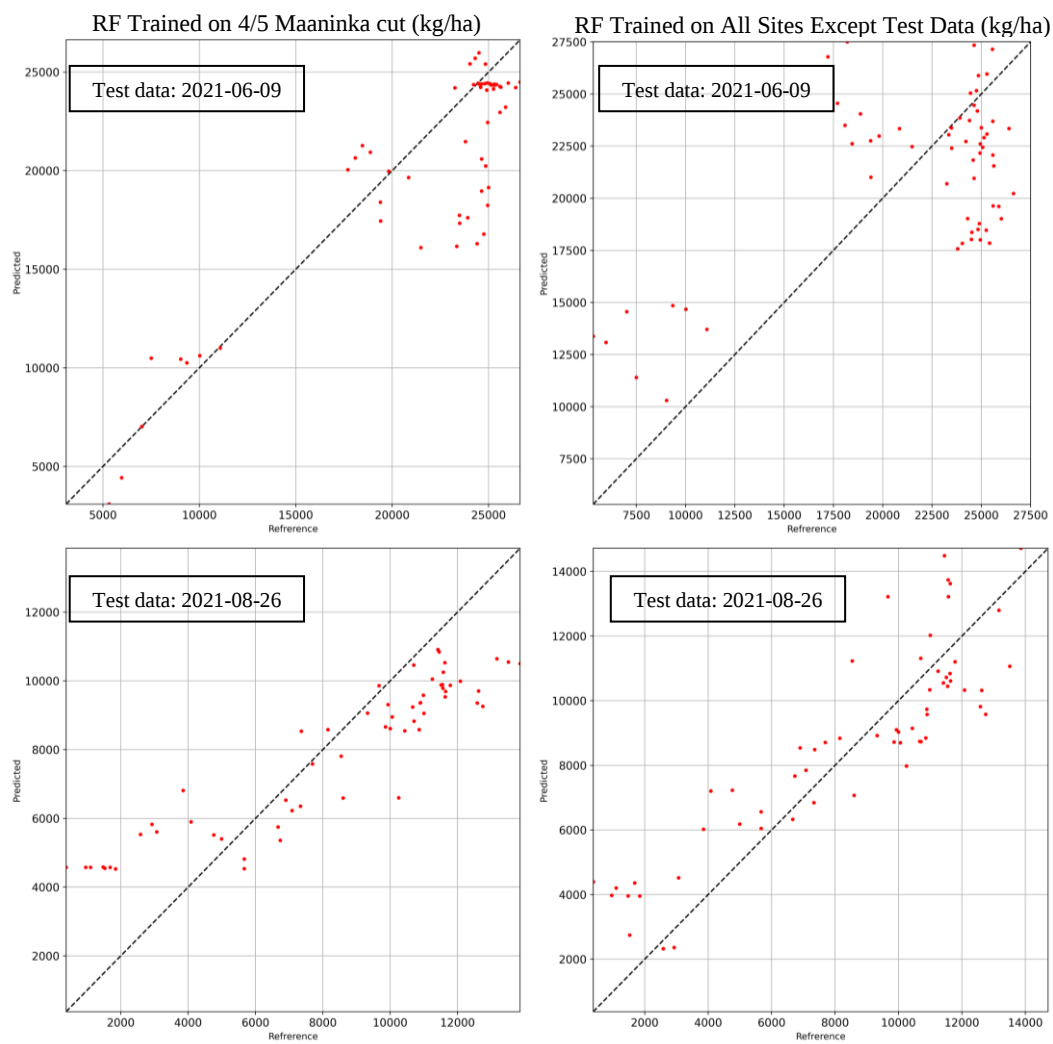
Table 2: Performance of random forest (RF) and partial least squares regression (PLSR) models for the leave-one-out cross-validation (LOOCV) within Maaninka datasets – where the model was trained on 4 out of 5 dates from Maaninka and tested on the left-out date. R²: coefficient of determination and NRMSE: Normalized Root Mean Square Error.

In the cross-site setup (Table 3), RF models performed better for the 2021 cuts (NRMSE 21.4%) while PLSR models provided better NRMSE for 2022 cuts with NRMSEs 20.68%, 33.93% and, 60.7%, respectively for first, second and third cuts of 2022. Overall, the cross-site models (Table 3) performed similar or better than the models using only same site (Table 2). Notably, for the 2022-07-20 test date, the cross-site model significantly outperformed the model using only the same site

(NRMSE of 39.77% vs. NRMSE of 92.41%), suggesting that including external site data can in some cases improve robustness, especially when local data is less informative.

	Trained on All Sites Except Test Data					
	RF			PLSR		
Test Data (Maaninka)	RMSE kg /ha	NRMSE %	R ²	RMSE kg /ha	NRMSE %	R ²
2021-06-09	4610	21.39	0.37	5181	24.03	0.2
2021-08-26	1806	21.44	0.77	2653	31.49	0.51
2022-06-14	6187	28.19	-0.73	4540	20.68	0.07
2022-07-20	3210	39.77	0.05	2739	33.93	0.31
2022-08-30	7462	66.81	-1.29	6780	60.69	-0.89

Table 3: Performance of random forest (RF) and partial least squares regression (PLSR) models for the leave-one-out cross-validation (LOOCV) within Maaninka datasets *Cross-Site Transfer* – where the model was trained on all available data from the other sites plus all Maaninka dates except the test Maaninka date. R²: coefficient of determination and NRMSE: Normalized Root Mean Square Error.



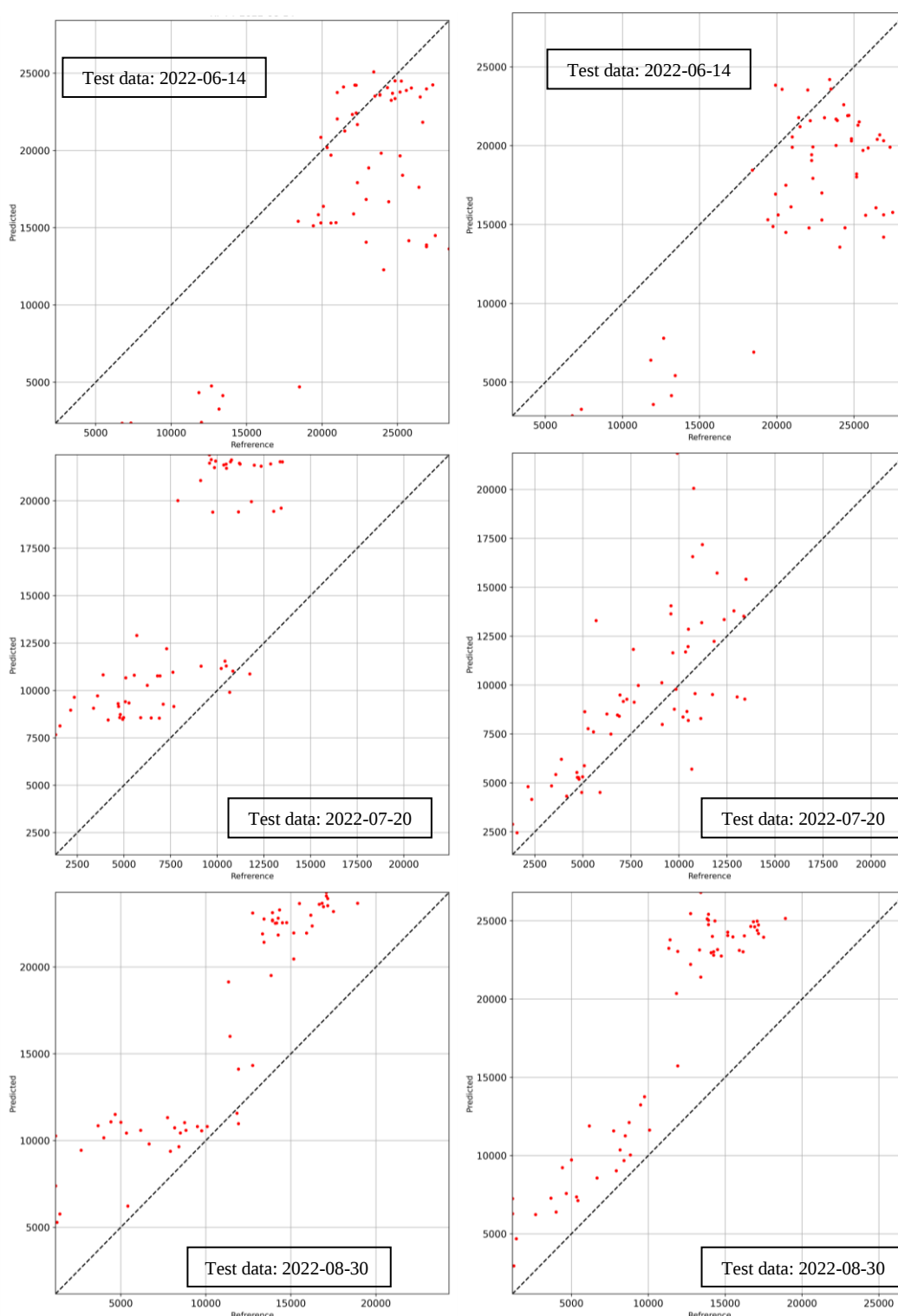


Figure 5. Scatter plot of the predicted fresh yield within Maaninka datasets using random forest (RF). The **left column** contains the scatter plots where the model was trained on 4 out of 5 dates from Maaninka and tested on the left-out date, while the **right column** where the model was trained on all available data from the other sites plus all Maaninka dates except the test Maaninka date. The cuts are ordered from top to bottom in chronological order.

4. Conclusions

The objective of this study was to evaluate estimating biomass at a given harvest using data exclusively from other harvests and, where applicable, from different locations. Although the

study area remained consistent, variations in environmental conditions across growth periods and years introduced challenges for model generalization. The results demonstrated that incorporating data from multiple locations can enhance model robustness, particularly when local data are limited or less informative. Random Forest regression outperformed PLSR

within Maaninka site; however, PLSR exhibited slightly better performance and generalizability in tests involving multi-location datasets.

Future research should incorporate longer multi-year time series and multi-location datasets, and potentially include additional data sources beyond remote sensing, to enable a more robust evaluation of model performance. Furthermore, studies on generalization potential of deep learning models will be of great interest.

Acknowledgements

This research was funded by CyberGrass 2.0 -project (ID NPA0800223) in the European Union Interreg NPA programme, by the European Agricultural Fund for Rural Development: Europe investing in rural areas (North Savo ELY Centre) - "Remote sensing methods to support feed production REHUDROONI" (Grant no. 145346), by the European Regional Development Fund for "CyberGrass I - Introduction to remote sensing and artificial intelligence assisted silage production" project (ID 20302863) in the European Union Interreg Botnia-Atlantica programme, by "Säilörehun korjuuajan tarkentaminen" project (ID 413577) and by the Ministry of Agriculture and Forestry of Finland project NC Grass - Dairy and beef industries in Finland: Progressing pathways to carbon-neutrality by 2035 (VN/28562/2020-MMM-2). The field trial in Maaninka was funded mainly by the European Agricultural Fund for Rural Development ("Production Resilience from Grass - VarmaNurmi" project).

References

- Bazzo, C. O. G., Kamali, B., Hütt, C., Bareth, G., & Gaiser, T. (2023). A review of estimation methods for aboveground biomass in grasslands using UAV. *Remote Sensing*, 15(3), 639.
- Franceschini, M. H., Becker, R., Wichern, F., & Kooistra, L. (2022). Quantification of grassland biomass and nitrogen content through UAV hyperspectral imagery—active sample selection for model transfer. *Drones*, 6(3), 73.
- Hernandez, A., Jensen, K., Larson, S., Larsen, R., Rigby, C., Johnson, B., ... & Sinton, S. (2024). Using Unmanned Aerial Vehicles and Multispectral Sensors to Model Forage Yield for Grasses of Semiarid Landscapes. *Grasses*, 3(2), 84-109.
- Hütt, C., Isselstein, J., Komainda, M., Schöttker, O., & Sturm, A. (2024). UAV LiDAR-based grassland biomass estimation for precision livestock management. *Journal of Applied Remote Sensing*, 18(1), 017502-017502.
- Lussem, U., Schellberg, J., Bareth, G. Monitoring Forage Mass with Low-Cost UAV Data: Case Study at the Rengen Grassland Experiment. *PFG—Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, 88, 407-422.
- Patel, M. K., Ryu, D., Western, A. W., Fitzgerald, G. J., Perry, E. M., Suter, H., & Young, I. M. (2024). A new multispectral index for canopy nitrogen concentration applicable across growth stages in ryegrass and barley. *Precision Agriculture*, 25(1), 486-519.
- Oliveira, R. A., Näsi, R., Niemeläinen, O., Nyholm, L., Alhonoja, K., Kaivosoja, J., ... & Honkavaara, E. (2020). Machine learning estimators for the quantity and quality of grass swards used for silage production using drone-based imaging spectrometry and photogrammetry. *Remote Sensing of Environment*, 2426, 111830.
- Oliveira, R. A., Näsi, R., Korhonen, P., Mustonen, A., Niemeläinen, O., Koivumäki, N., ... & Honkavaara, E. (2024). High-precision estimation of grass quality and quantity using UAS-based VNIR and SWIR hyperspectral cameras and machine learning. *Precision Agriculture*, 25(1), 186-220.
- Viljanen, N., Honkavaara, E., Näsi, R., Hakala, T., Niemeläinen, O., & Kaivosoja, J. (2018). A novel machine learning method for estimating biomass of grass swards using a photogrammetric canopy height model, images and vegetation indices captured by a drone. *Agriculture*, 8, 70. <https://doi.org/10.3390/agriculture8050070>
- Wengert, M., Wijesingha, J., Schulze-Brüninghoff, D., Wachendorf, M., & Astor, T. (2022). Multisite and Multitemporal Grassland Yield Estimation Using UAV-Borne Hyperspectral Data. *Remote Sensing*, 14(9), 2068. <https://doi.org/10.3390/rs14092068>