Noname manuscript No.
(will be inserted by the editor)

Assessment of Crop Yield Prediction Capabilities of CNN using Multisource Data

**Petteri Nevavuori · Nathaniel Narra ·
Petri Linna · Tarmo Lipping**

Received: date / Accepted: date

Abstract The growing abundance of digitally available spatial, geological and climatological data opens up new opportunities for agricultural data based input-output modeling. In our study, we took a Convolutional Neural Network model previously developed on Unmanned Aerial Vehicle (UAV) image data only and set out to see whether additional inputs from multiple sources would improve the performance of the model. Using the model developed in a preceding study, we fed field-specific data from the following sources: near-infrared data from UAV overflights, Sentinel-2 multispectral data, weather data from locally installed Vantage Pro weather stations, topographical maps from National Land Survey of Finland, soil samplings and soil conductivity data gathered with a Veris MSP3 soil conducitivity probe. Either directly added or encoded as additional layers to the input data, we concluded that additional data helps the spatial point-in-time model learn better features, producing better fit models in the task of yield prediction. With data of four fields, the most significant performance improvements came from using all input data sources. We point out, however, that combining data of various spatial or temporal resolution (i.e., weather data, soil data and weekly acquired images, for example) might cause data leakage between the training and testing data sets when training the CNNs and, therefore, the improvement rate of adding additional data layers should be interpreted with caution.

Keywords crop yield prediction · CNN · multisource input · remote sensing · intra-field

P. Nevavuori
Mtech Digital Solutions Oy,
E-mail: petteri.nevavuori@mtech.fi

P. Linna · N. Narra · T. Lipping
Tampere University,
E-mail: {nathaniel.narra, petri.linna, tarmo.lipping}@tuni.fi

1 Introduction

The application of novel and performant deep learning techniques has seen an increasing trend in the last few years in the domain of Smart Farming and Precision Agriculture [6]. Multiple factors are at play: the abundance of open access satellite system spatial data, availability of commercial unmanned aerial vehicles (UAVs) mountable with external sensors, developments in the soil sensor and camera sensor technologies and the constant need to optimize the production of farms.

Convolutional Neural Networks (CNN), being a subset of deep learning, have been utilized in recent studies on crop yield prediction [6]. The spatial model architecture has been used in predicting cotton yield from RGB data taken at close proximity [15], cereal crop yield prediction from mid-altitude UAV RGB data [11], rice grain yield estimation [18] and crop yield prediction using multisource inputs on patch-scale [4]. In [11] we compared intra-field crop yield estimation performance with NDVI and RGB data from the earlier and later part of the growing season with a variety of CNN configurations. The focus of that study was to assess the generalization capability of a yield prediction model with UAV RGB data.

1.1 Objectives

In this study, we examine the effect of additional field-related spatial or spatial-like data on the intra-field crop yield prediction capabilities using data gathered from the earlier half of the growing season of 2018 (weeks 21 to 26). The objective of this study is to assess crop yield prediction capabilities with the best CNN model composition from [11] by varying the input data configuration. The focus of this study is to see whether additional data, such as weather data, soil and ground information and open-access Sentinel-S2 data would improve the point-in-time prediction performance compared to just using UAV-based RGB data. To limit the scope of the study, architectural and hyperparameter tuning of the CNN model is not addressed here to better isolate performance changes to data and the tuned out architectural and optimizer related hyperparameters were thus taken from [11].

2 Material and Methods

2.1 Data Acquisition

For this study, four crop fields were selected for data acquisition in the vicinity of Pori, Finland (61°29'6.5"N, 21°47'50.7"E) for the growing season of 2018. The field information is provided in Table 1. Following the conclusions of [11], only data from the earlier half of the growing season was considered for UAV and Sentinel-S2 data.

Table 1 The fields selected for the study in the proximity of Pori, Finland. The thermal time is calculated as the cumulative sum of temperature between the sowing and harvest dates. Mean yield has been calculated from processed yield sensor data for each field.

Field #	**Size (ha)**	**Mean yield (kg/ha)**	**Crop (*Variety*)**	**Thermal time (°C)**	**Sowing date**
1	7.59	5157.6	Wheat (*Mistral*)	1316.8	14 May
2	11.77	5534.3	Barley (*Zebra*)	1179.9	12 May
3	7.88	4166.9	Barley (*RGT Planet*)	1127.6	16 May
4	7.24	6166.0	Oats (*Ringsaker*)	1216.4	18 May

Table 2 General information of data sources and their original formats.

Source	**Type**	**Resolution/Step**	**Multitemporal**
UAV	Raster	0.3125 m/px	Yes
Sentinel-S2	Raster	[10,20,60] m/px	Yes
Soil samples	Vector	50 m	No
Veris MSP3	Vector	20 m	No
Topography	Vector	2 m	No
Weather	Tabular	-	Yes
Yield	Vector	Varying	No

The multisource input data for the fields consists of UAV-based RGB images, location data, multispectral Sentinel-2 [1] satellite data, sparsely collected and analyzed soil samplings, machine-collected soil information, topography information and local weather station data. General information about the original data sources are given in Table 2. Some of the data were collected during the growing season of 2018 either manually or automatically, while other data were acquired within one year time difference from the aforementioned season. A total of 39 layers constitute the input data sets, while a single layer, the crop yield, is used as the ground truth. These data are described next and the data layers are numbered for further reference.

2.1.1 UAV

It has already been demonstrated that UAV-based RGB data from the first half of the growing season works better than the data from the second half of the growing season and better than NIR only in crop yield prediction [11]. The UAV data of this study has also been used in [10]. The images were taken at average height of 150 meters with a minimum of three ground control points for geometric calibration. Color correction was performed pre-flight and illumination sensors were used for radiometric calibration. We selected UAV-based RGB data acquired for the first weeks after sowing (weeks 21 to 26 of 2018). Thus, every imaged field has five distinct UAV RGB rasters in the collected data set. The data were acquired with overfligths using a SEQUIOA (Parrot Drone SAS, Paris, France) multispectral camera mounted on a Airinov

Solo 3DR (Parrot Drone SAS, Paris, France) UAV. Field-wise orthomosaics were constructed with Pix4D (Pix4D S.A., Prilly, Switzerland) software. UAV data contains the following layers:

1. Red
2. Green
3. Blue

2.1.2 Sentinel-S2

The Sentinel-S2 satellite data for the fields was acquired from the Copernicus Open Access Hub (European Space Agency, Paris, France). The data were date-matched to UAV images during acquisition, prioritizing images where the algorithmically determined cloud probability was lowest. Thus, five Sentinel-S2 rasters with temporal spacing similar to the UAV data were selected for the data set. With the abbreviated names of product layers in brackets, the Level-2A Sentinel-S2 consists of the following layers:

4. Wavelength 0.443 μm (B01)
5. Wavelength 0.490 μm (B02)
6. Wavelength 0.560 μm (B03)
7. Wavelength 0.665 μm (B04)
8. Wavelength 0.705 μm (B05)
9. Wavelength 0.740 μm (B06)
10. Wavelength 0.783 μm (B07)
11. Wavelength 0.842 μm (B08)
12. Wavelength 0.865 μm (B8A)
13. Wavelength 0.945 μm (B09)
14. Wavelength 1.610 μm (B11)
15. Wavelength 2.190 μm (B12)
16. Aerosol optical thickness at 550 nm (AOT)
17. Scene classification layer (SCL)
18. Water vapour map (WVP)
19. Cloud probability (CLDPRB)
20. True color, red (TCIR)
21. True color, green (TCIG)
22. True color, blue (TCIB)

2.1.3 Soil samples

Soil samples were manually collected from the fields by ProAgria, an agronomic counseling instution, and sent to a Eurofins (Eurofins Viljavuuspalvelu, Mikkeli, Finland) laboratory for further analysis. Soil samples were collected with 50 m steps so that a single sample represented an area of 50 × 50 m. The samples were collected manually once during November 2018. Being point vectors, the data were rasterized with the `gdal_warp` program of the GDAL utility [17]. Soil sample data contains the following layers:

23. Calcium
24. Copper
25. Potassium
26. Magensium
27. Manganese
28. Phosphorus
29. Sulfur
30. Zink

2.1.4 Veris MSP3

To get a finer map of soil chracteristics, a MSP3 soil scanner (Veris Technologies, Salina, Kansas, USA) was used to map the fields at depths of 0-30 cm and 30-90 cm. The measurements were performed during April and May of 2019. The MSP3 measures the soil's electrical conductivity (EC), which is an indicator of soil compactness, wetness and soil type proportions. Additionally, the instrument measures the pH of the soil. Being irregularly spaced point data initially, data had to be rasterized from point vectors. The rasterization was done with the `gdal_warp` program of the GDAL utility [17]. Each field was measured once. Veris MSP3 data contains the following layers:

31. Shallow EC
32. Deeper EC
33. Ratio, (EC SH / EC DP)
34. Infra-red reflectance
35. Red reflectance
36. Soil pH

2.1.5 Topography

The National Land Survey of Finland conducts light detection and ranging (LiDAR) based elevation mappings on a regular basis in Finland. This data is openly available for anyone to download [2] and contains laser scanned point-cloud data with approximately one point per 2 m^2 [9]. The LiDAR data set was acquired for each of the four fields. The LiDAR data were converted from point-cloud data to spatial rasters using the ArcGIS (Esri, Redlands, California, USA) software. During the conversion, the data were interpolated to match UAV data data in terms of resolution. The topography data contains only the following layer:

37. Elevation information

2.1.6 Weather data

Weather data were collected with two separately located Vantage Pro2 (Davis Instruments, Hayward, California, USA) weather stations. As the fields constitute two distinct clusters, a weather station was placed in the immediate vicinity of each field cluster. While the stations log multiple variables with a time resolution of just minutes, we utilized accumulated daily statistics and matched data to UAV acquisition dates. Thus, five weather data maps were constructed for each field spacing matching the dates of the UAV data. The weather data contains the following layers:

38. Cumulative temperature sum
39. Cumulative rain sum

2.1.7 Yield data

As the task of regression is that of supervised prediction, the training of the CNN model requires information about the ground truth, the target values. These were acquired during the harvest of 2018 via yield mapping sensor devices attached to the harvesters, either with a CFX 750 (Trimble Navigation, Sunnyvale, California, USA) or Greenstar 1 (John Deere, Molinde, Illinois, USA). CFX 750 utilizes optical sensors to measure yield throughput and moisture. Greenstar 1 utilizes a kinetic mass flow sensor to measure yield throughput and a separate moisture sensor. The yield maps generated by the mapping equipment were initially in the form of vector point-clouds. The irregularly spaced points were filtered prior rasterization to contain only points where the yield was between 1500 and 15000 kg/ha and the harvester speed between 2 and 7 km/h, following the yield pre-processing methodology of [11]. Rasterization was then done by interpolating the yield data to form a raster image.

2.2 Data Preprocessing

2.2.1 Interpolation

The first step after the acquisition of data was to harmonize the spatial resolution across multiple different sources. The UAV data were initially downsampled to 0.3125 m/px, or 32 pixels per 10 meters. This is to match the method of data processing in [11]. Main reasons are to limit the inputs to reasonable size and to have the input dimensions conform to a power of 2 for GPU-based computations. The coarser data, namely Sentinel-S2, soil samples, Veris MSP3, elevation and yield data, required upsampling via interpolation to match this resolution. The interpolation was done using the GDAL utility's `gdal_grid` program with `invdist:power=3:smoothing=20` interpolation algorithm. As with the input data, also the target crop yield data were interpolated to UAV matching resolution. Example results of interpolation are depicted in Figure 1.

2.2.2 Input Feature normalization

After interpolation, the next step was to normalize the data. While absolute values could also be directly used, scaling the input values close to the magni-

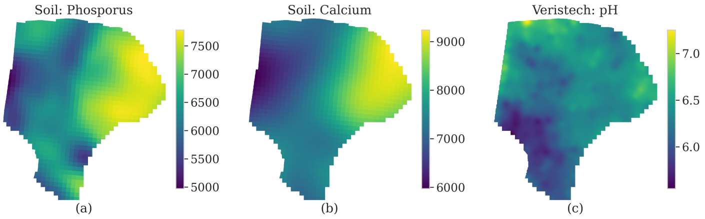

Fig. 1 Examples of input data interpolations on field-scale. (a) is the interpolated phosporus map, (b) the interpolated calcium content in the field and (c) the pH map as measured by the Veris MSP3 soil mapper.

tude of the model's parameters (i.e. weights) helps the model converge faster. Input layers were normalized using a function

$$d^{NORM} = (d - \mu_d)/(max(d) - min(d)), \quad d \in D \tag{1}$$

where d is a layer in the set of all layers D in the data set and d^{NORM} is the normalized layer. However, the target crop yield values were not scaled, akin to [11].

2.2.3 Frame separation

The CNNs require input data to have fixed dimensions. Low number of fields and the irregularities of field shapes led us to extract smaller, fixed dimension frames from the field data. Following [11], we extracted overlapping 40 × 40 m (128 × 128 px) frames with 10 m horizontal and vertical steps. Prior extraction, all input and target data from various input sources were aligned in terms of geolocation and resolution to ensure frame extraction from mathching areas. Frames containing half or more valid pixels were included in data, while those having less than half were discarded. This resulted in a total of 16375 input-target frames.

2.2.4 Data sets

Extracted samples were then divided into training, validation and test sets. Training and validation sets were utilized during the training, while the test set was set aside as the out-of-sample performance evaluation data set. As the number of unique fields was low, we wanted to maximize the sample variability the model sees during training. We first attempted to train the models with data separated on a per-field basis with two fields for training, one for validation and one for testing. This led to the model overfitting to the training data and poor generalization performance due to low training data set variability. Similarly low performance was achieved with splitting fields to separate

Table 3 Compositions of training, validation and test sets used to train and evaluate the models.

Data set	Weeks	Frames	Proportion
Training	21,23,25	7561	46.2%
Validation	24	2938	17.9%
Test	22, 26	5876	35.9%

training, validation and test sections. We thus decided then to divide the data temporally to distinct training, validation and test sets according to the UAV image acquisition week. The samples were then shuffled to eliminate spatial autocorrelation in subsequent samples due to overlapping frame extraction. Used weeks, sample counts and sample count proportions for separated sets are given in Table 3.

2.3 Model Architecture

Convolutional neural networks, CNNs, are a subset of spatial model architectures within the broader context of deep learning. CNNs excel in tasks, where the inputs fed to the model are either images or image-like data, i.e. spatial data [14,7]. While the inner workings of the CNNs has already been well documented [11], we quickly review the operating principles of a CNN. The architecture operates with layers, like many of the deep learning architectures. Each layer is a combination of a convolution operation, which is often followed by a pooling operation. At the heart of the model are the trainable filters of the convolution operation, i.e. the kernels, which produce feature maps for further use.

In our study, we implement and use the best performing CNN architecture of [11]. The model consists of six convolutional layers, followed by two fully connected (FC) layers. Convolutional layers consist of 2D convolutions, batch normalization and non-linear activation with a rectified linear unit (ReLU). First and last convolutional layers also employ max pooling with 2×2 kernel to extract more robust features and reduce intermediate output data dimensions. First five convolutional layers operate with 64 5×5 kernels and the last convolutional layer with 128 5×5 kernels. The outputs of the last convolutional layer are then flattened to a single vector, which is then fed to two 1024 neuron FC layers, both having ReLU activation. Last FC layer outputs the final prediction result. The model was implemented with PyTorch [12] and trained with Skorch [16].

2.4 Training

To gauge the effects of multisource data on the crop yield prediction task with spatial inputs, we performed trainings with four different input data configurations. The data configurations and the input data sources included in

Table 4 The different data configurations used for training distinct models. *RGB Only* uses UAV RGB data only. *No S2* uses UAV, soil, Veris MSP3, topography and weather data. *S2 Raw* adds Sentinel-S2 raw wavelength band data to *No S2*. *S2 Full* adds calculated Sentinel-S2 Level-2A product layers to *S2 Raw*. An X indicates the inclusion of an input data source to a data configuration, while a dash indicates the exclusion.

Source	Channels	RGB Only	No S2	S2 Raw	S2 Full
UAV	1-3	X	X	X	X
Soil	23-30	-	X	X	X
Veris	31-36	-	X	X	X
Topo	37	-	X	X	X
Weather	38-39	-	X	X	X
S2 bands	4-15	-	-	X	X
S2 other	16-22	-	-	-	X
Band count		3	20	32	39

them are further given in Table 4. To elaborate, the derived data configurations were as follows:

- *RGB only.* As [11] was conducted with RGB data from UAVs only, we wanted to make baseline performance evaluation with UAV RGB data only. No other sources were included in this setting.
- *No S2.* Next we wanted to see the effects of soil and weather data on the predictive performance. We thus included all other sources of data (UAV, soil, Veris MSP3, topography and weather) but excluded the satellite data.
- *S2 Raw.* As Sentinel-S2 Level-2A products contain additional algorithmically generated layers, we wanted to see the effect of including just the raw wavelength bands with other input data sources.
- *S2 Full.* The last setting was to use all data acquired for this study.

Because data were distinct from data used in [11], we initialized and trained all models anew for each data configuration. To account for the effects of randomized network parameter initialization, we trained 10 models per data configuration, 40 trainings in total. We used Adadelta [19] as the optimizer, 0.58 for the learning rate, 0.001 for the weight decay and 0.9 for the Adadelta's ρ coefficient as those were the best performing hyperparameters in [11]. Similarly, we used early stopping with a patience of 50 stagnant epochs and continued the training once. The models were trained with Nvidia Tesla V100 Volta and Pascal architecture server GPUs in a distributed computation environment.

3 Results

The CNN models with distinct input data configurations were trained with data of four unique fields. The model architectures, hyperparameters and the training procedures were identical to [11]. As the aim of our study was to evaluate the effects of introducing multisource inputs to crop yield prediction, we trained spatial yield prediction models with four distinct data configurations. The data configurations are discussed in Section 2.4. As the training time loss

Table 5 The test set performance of the same CNN architecture and hyperparameter configuration with various data configurations. *RGB Only* is the baseline model. Out of the configurations, the model performed best with all input data layers (*S2 Full*).

Data Configuration	**Test RMSE (kg/ha)**	**Test MAE (kg/ha)**	**Test MAPE (%)**	**Test R^2 -**
RGB Only	1055.7	838.8	18.2	0.343
No S2	892.4	694.9	14.8	0.531
S2 Raw	461.0	340.9	6.94	0.875
S2 Full	**364.1**	**274.3**	**5.18**	**0.922**

function we used the mean squared error (MSE). Other loss metrics were also calculated, including the square root of the MSE (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE) and the coefficient of determination (R^2). These metrics were not, however, monitored and neither did they not influence model selection during training.

The baseline model using UAV RGB data only attained 1055.7 kg/ha test RMSE, 18.2% test MAPE and 0.343 test R^2. Out of all data configurations, the best performance of 364.1 kg/ha test RMSE, 5.18% test MAPE and 0.922 test R^2 was achieved using all input data presented in our study (*S2 Full*). The performance results for all data configurations with the held out test data set are given in Table 5.

To gain a better view into how the models train with distinct data predicted, we also examined the unseen test sample distributions of predicted values against ground truth values, the true crop yields. With the data, the baseline *RGB Only* model's predictions resemble a Gaussian distribution centered around the mean 5140 kg/ha of true yield values. As more inputs are introduced, the predicted distributions' shapes align with the true values more closely, expressing multi-modal peaks where the true values have them. The test set distributions are depicted in Figure 2.

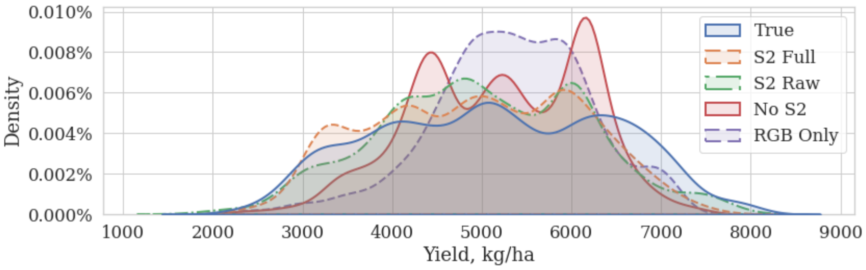

Fig. 2 Distributions of predictions against true yields with the holdout test set.

Table 6 The relative performance of the models trained with distinct multisource input data configurations to the baseline *RGB Only* model. For RMSE, MAE and MAPE negatiThe input data configurations are defined in Section 2.4.

Data Setting	**Relative change from *RGB Only***			
	Test RMSE	**Test MAE**	**Test MAPE**	**Test R^2**
No S2	-15.5%	-17.2%	-18.7%	+0.188
S2 Raw	-56.3%	-59.4%	-61.9%	+0.532
S2 Full	**-65.6%**	**-67.3%**	**-71.5%**	**+0.579**

4 Discussion and Conclusions

In this study, we evaluated the effects of using input data from multiple sources on the task of spatial crop yield prediction. Using a CNN model architecture developed for UAV RGB inputs from [11], we introduced additional data from sources like soil samplings, Veris MSP3 soil scanner, topographical maps, weather stations and Sentinel-S2 satellites to the model. We trained ten models for each distinct input data configuration: (1) a *RGB Only* baseline model, (2) a *No S2* multisource model with satellite data excluded, (3) a *S2 Raw* multisource model with raw satellite band data included and (4) a *S2 Full* multisource model with all input data. Out of each set of ten trained models, we selected the models performing best. The model architecture and hyperparameters for the training were taken from [11] and left unchanged to constrain the variability in performance to data only. Only thing varying between model trainings, in addition to four distint input data source configurations, were the initialized model weights.

The performance with larger number of fields using UAV RGB data has already been extensively studied in our previous studies [11] and [10]. Thus, training a model with only UAV RGB data provides a studied baseline to which models trained with additional data can be compared against. The best performing data configuration was *S2 Full* with 364.1 kg/ha test RMSE, 5.18% test MAPE and 0.922 test R^2 using all 39 layers of input data for each extracted frame. Compared to the baseline *RGB Only* model, the *S2 Full* attained 65.6% lower RMSE, 67.3% lower MAE, 71.5% better MAPE and 0.579 higher R^2 with the test set. Generally every model with multisource inputs performed better than the baseline model. This is shown in Table 6.

Crop yield prediction with spatial data and spatial deep learning models has seen an increase in the past few years [6]. Having been studied with a variety of different architectures, from feed-forward networks to hybrid spatio-temporal models, studies have also been conducted with CNN as the main architecture. In [11], a single CNN model was developed to predict crop yields from fields with varying crop types (wheat, barley and oat) from UAV images collected of Finnish crop yields during 2017. Using smaller frames extracted from ortho-images, the best performance was 484.3 kg/ha MAE and 8.8% MAPE. Using soil nutrient data, seed rate, elevation maps, soil's electroconductivity and satellite data in USA, [4] trained a CNN to predict crop yields for nine fields. They report an average scaled MSE of 0.70 which translates to

1145 kg/ha. [18] utilized RGB and multispectral data acquired with a UAV from rice fields in China to predict rice yields with a composite CNN model on field block scale. Feeding the multisource data to distinct, parallelized CNNs, they report a rice yield prediction performance of 0.50 R^2 and 26.6% MAPE.

As we had sufficient data overlap across multiple input sources and the data were acquired from only four unique fields, objective multisource crop yield prediction performance evaluation requires more care in interpreting the results. Relative increase in performance from best performing UAV data utilizing *RGB only* model to the best *No S2* model with additional soil and weather data was notably small. Largest improvements were gained with the introduction of Sentinel-S2 data. Adding raw Sentinel-S2 bands to the RGB, soil and weather data increased the performance by 40.8% RMSE, 42.2% MAE, 43.2% MAPE and 0.344 R^2 from *No S2*. Thus, the increase in performance with Sentinel-S2 is considerably higher than what was achieved with adding soil, topography and weather data to UAV RGB data.

Data acquisition for remote sensing and multisource input data for smart farming is generally laborous and resource intensive. While satellite data is generated automatically, UAVs require semi-autonomous operation at best and the collection of soil data requires extensive on-site manual labour. With more data from a variety of sources a more extensive and representative study can be conducted.

Another limitation stems from differences in spatial and temporal dispersion of different input data sources. UAV, Sentinel-S2 and weather data vary temporally in the data we have used, whereas soil samplings, Veris MSP3 and topographical maps do not. As our data was split temporally to training, validation and test sets, the latter are present in all of these data sets. On the other hand, weather data varies only temporally and constitutes spatial rasters with constant values corresponding to the time of UAV imaging. This means that whether the data is split temporally or spatially, some layer or part of data is always present in training, validation and test sets. As [13] point out, deep learning models are able to implicitly learn linear and non-linear couplings from data with correlations. This means that the deep learning models learn sets of representative features from complex combinations of the inputs and not from single input values on solitude. Furthermore, the performance gains with UAV RGB data combined with temporally invariant soil and ground data is trumped by the performance gains of data configurations using Sentinel-S2 data as additional inputs. This would suggest that the combination of the inputs matters more than presence of distinct, invariant data in training, validation and test sets. However, the concrete effects of simultaneous layer-level data existence in training, validation and test data sets are presently unknown to us and, thus, a subject of future research.

Regarding multisource data in the context of smart farming and crop yield estimation, data itself is an evolving research topic. The use of multisource inputs in remote sensing, while focusing on multispectral data acquired from satellite systems orbiting the globe, has been extensively reviewed in [5]. The use of multispectral data from UAVs and the prediction architectures thereof is also a developing topic [8]. Another topic related to spatial data is that of autocorrelation [3]. To address autocorrelation of spatial frames in a future study, the inclusion of pixel-wise location information, as suggested in [3], should be sufficient to inform the deep learning model whether data similarity is due to proximity or some other factor or combination of them.

In conclusion, our study indicates that increasing the number of input data sources increases the performance of intra-field crop yield prediction. To draw definite conclusions on the most optimal configuration of input data sources more data is required. With more representative data, generalizable conclusions are more warranted. As the data in this study focuses on a single rowing season, a future plan is to study the generalization of a multisource crop yield prediction model with multiple years of data. Yet in this study the relative increase from baseline of using UAV RGB only as the input data were notable. Consolidating UAV RGB data with soil and ground topology data already somewhat improves the prediction performance, while largest performance gains were gained from using Sentinel-S2 in addition to UAV RGB, soil sampling, Veris MSP3 soil scanner, weather and topography data.

Acknowledgements We would like to thank Mtech Digital Solutions Oy for partially funding this research, Tampere University for providing the computational resources and MIKÄ DATA project for providing us with data.

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. ESA: Sentinel-2. URL https://sentinel.esa.int/web/sentinel/missions/sentinel-2
2. PaITuli - Spatial data for research and teaching. URL https://paituli.csc.fi/download.html
3. Amgalan, A., Mujica-Parodi, L.R., Skiena, S.S.: Fast Spatial Autocorrelation. Biometrics **30**(4), 729 (2020). DOI 10.2307/2529248. URL http://arxiv.org/abs/2010.08676 https://www.jstor.org/stable/2529248?origin=crossref
4. Barbosa, A., Marinho, T., Martin, N., Hovakimyan, N.: Multi-stream CNN for spatial resource allocation: A crop management application. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, vol. 2020-June, pp. 258–266 (2020). DOI 10.1109/CVPRW50498.2020.00037
5. Ghamisi, P., Rasti, B., Yokoya, N., Wang, Q., Hofle, B., Bruzzone, L., Bovolo, F., Chi, M., Anders, K., Gloaguen, R., Atkinson, P.M., Benediktsson, J.A.: Multisource and multitemporal data fusion in remote sensing: A comprehensive review of the state of the art. IEEE Geoscience and Remote Sensing Magazine **7**(1), 6–39 (2019). DOI 10.1109/MGRS.2018.2890023
6. van Klompenburg, T., Kassahun, A., Catal, C.: Crop yield prediction using machine learning: A systematic literature review. Computers and Electronics in Agriculture **177**, 105709 (2020). DOI 10.1016/j.compag.2020.105709. URL https://linkinghub.elsevier.com/retrieve/pii/S0168169920302301
7. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. Communications of the ACM **60**(6), 84–90 (2017). DOI 10.1145/3065386. URL http://dl.acm.org/citation.cfm?doid=3098997.3065386

8. Messina, G., Modica, G.: Applications of UAV thermal imagery in precision agriculture: State of the art and future research outlook. Remote Sensing **12**(9) (2020). DOI 10.3390/RS12091491
9. National Land Survey of Finland: Laser scanning data. URL http://www.nic.funet.fi/index/geodata/mml/laserkeilaus/mml_laserkeilaus_2016_eng.pdf
10. Nevavuori, P., Narra, N., Linna, P., Lipping, T.: Crop Yield Prediction Using Multitemporal UAV Data and Spatio-Temporal Deep Learning Models. Remote Sensing **12**(23), 4000 (2020). DOI 10.3390/rs12234000. URL https://www.mdpi.com/2072-4292/12/23/4000
11. Nevavuori, P., Narra, N., Lipping, T.: Crop yield prediction with deep convolutional neural networks. Computers and Electronics in Agriculture **163**(June), 104859 (2019). DOI 10.1016/j.compag.2019.104859. URL https://linkinghub.elsevier.com/retrieve/pii/S0168169919306842
12. Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., Lerer, A.: Automatic differentiation in PyTorch. In: NIPS-W (2017)
13. Sun, Y., Guo, G., He, X., Liu, X.: Multi-level coupling network for Non-IID sequential recommendation. IEEE Access **7**(Iid), 186247–186259 (2019). DOI 10.1109/ACCESS.2019.2961182
14. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition **07-12-June**, 1–9 (2015). DOI 10.1109/CVPR.2015.7298594
15. Tedesco-Oliveira, D., Pereira da Silva, R., Maldonado, W., Zerbato, C.: Convolutional neural networks in predicting cotton yield from images of commercial fields. Computers and Electronics in Agriculture **171**, 105307 (2020). DOI 10.1016/j.compag.2020.105307. URL https://linkinghub.elsevier.com/retrieve/pii/S0168169919319878
16. Tietz, M., Fan, T.J., Nouri, D., Bossan, B., skorch Developers: skorch: A scikit-learn compatible neural network library that wraps PyTorch (2017). URL https://skorch.readthedocs.io/en/stable/
17. Warmerdam, F., Rouault, E.: GDAL — GDAL documentation (1998). URL https://gdal.org/
18. Yang, Q., Shi, L., Han, J., Zha, Y., Zhu, P.: Deep convolutional neural networks for rice grain yield estimation at the ripening stage using UAV-based remotely sensed images. Field Crops Research **235**(August 2018), 142–153 (2019). DOI 10.1016/j.fcr.2019.02.022. URL https://doi.org/10.1016/j.fcr.2019.02.022 https://linkinghub.elsevier.com/retrieve/pii/S037842901831390X
19. Zeiler, M.D.: ADADELTA: An Adaptive Learning Rate Method (2012). DOI http://doi.acm.org.ezproxy.lib.ucf.edu/10.1145/1830483.1830503. URL https://arxiv.org/pdf/1212.5701.pdf http://arxiv.org/abs/1212.5701