AgWise

Rwanda potato fertilizer recommendation data analytics report

EiA in collaboration with RAB and CIP

AgWise

Fertiliser response in potato for Rwanda

STEP 1:	assemble data
STEP 2:	extract BLUPs (structural variation)
STEP 3:	determine indigenous nutrient supply (reverse QUEFTS)
STEP 4:	predict yield and yield response
STEP 4b:	alternative: directly predict yield using ML
STEP 5:	predict INS, IPS and IKS for target area
STEP 6:	calculate fertilizer recommendations
STEP 7:	suggestions for field validation

STEP 1: assembling data

Assembling available potato fertilizer trial data

Data is available from three experiments:

- IFDC trials: 1 season (2014B), 6 treatments (control, full NPK, partial reductions of N, P or K)
- RwaSIS fertilizer response trials (RS-PFR): 2 seasons (2022A and 2022B), 15 treatments (control, full NPK, full omissions of N, P or K, partial omissions of N, P or K, increased rates above full NPK)
- Scaling AKILIMO trials (SA-VAP): 2 seasons (2021A and 2021B), 5 treatments (partial substitution of blanket recommended NPK17:17:17 rate by other fertilizers, and increased NPK application; no control!)

A reference treatment can be set as treatment with rates of N > 75, P > 30 and K > 50 kg K ha^{-1}:

- IFDC: NPK_all (75-75-50)
- RS-PFR-1: increased NPK (90-40-60)
- SA-VAP-1: NPK11 (94-41-78)

Trials are located in the Birunga, Buberuka highlands and Congo-Nile watershed divide for all 3 experiments. The RS-PFR experiment also has few trials located in the Central Plateau.

AgWise analysis potato field trial data Rwanda

3 experiments available, of which 2 repeated during two consecutive seasons:

IFDC (2014B)

Treat	N	P	K
Control	0	0	0
NPK	75	75	50
NPK_all	75	75	50
All_redN	50	75	50
All_redP	75	50	50
All_redK	75	75	25

RS-PFR-1 (2022A, 2022B)

Treat	N	P	K
NPK 17x3	51	22	42
NP	51	22	0
NK	51	0	42
PK	0	22	42
NPK(30N)	30	22	42
NPK(60N)	60	22	42
NPK(90N)	90	22	42
NPK(30P)	51	30	42
NPK(40P)	51	40	42
NPK(50P)	51	50	42
NPK(30K)	51	22	30
NPK(60K)	51	22	60
NPK(80K)	51	22	80
Increased NPK	90	40	60
Control	0	0	0

SA-VAP-1 (2021A, 2021B)

Treat	N	P	K
NPK11	94	41	78
NPK4_DAP2	52	35	28
NPK4_MOP2	34	15	78
NPK6	51	22	42
NPK4_Urea2	80	15	28

Reference treatment
(N > 75, P > 30, K > 50)

INITIATIVE ON
Excellence in Agronomy

AgWise analysis potato field trial data Rwanda

3 experiments available, of which 2 repeated during two consecutive seasons:

Exp 1 (2014B)

Treat	N	P	K
Control	0	0	0
NPK	75	75	50
NPK_all	75	75	50
All_redN	50	75	50
All_redP	75	50	50
All_redK	75	75	25

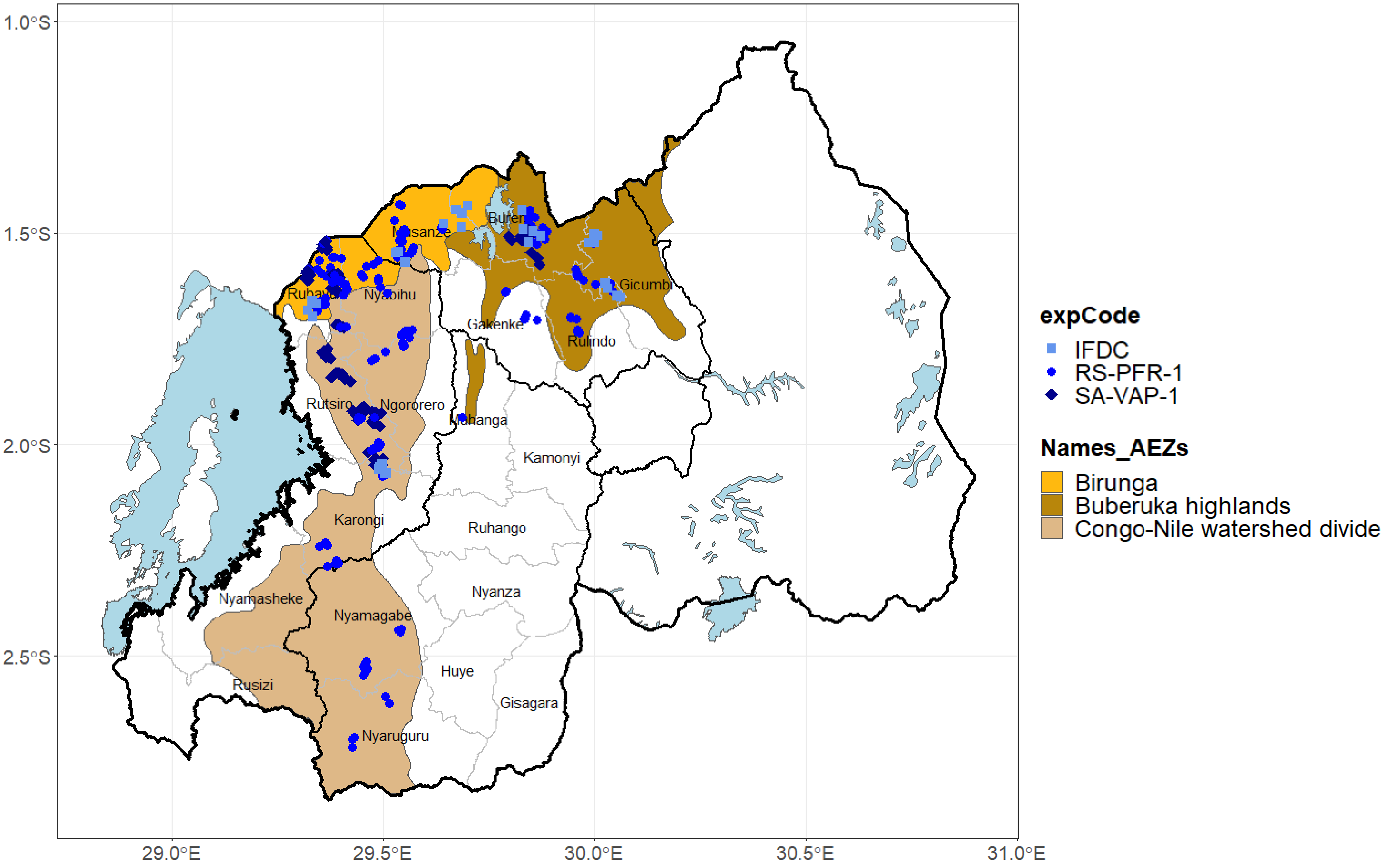
Exp 2 (2021A, 2021B)

Treat	N	P	K
NPK11	94	41	78
NPK4_DAP2	52	35	28
NPK4_MOP2	34	15	78
NPK6	51	22	42
NPK4_Urea2	80	15	28

Exp 3 (2022A, 2022B)

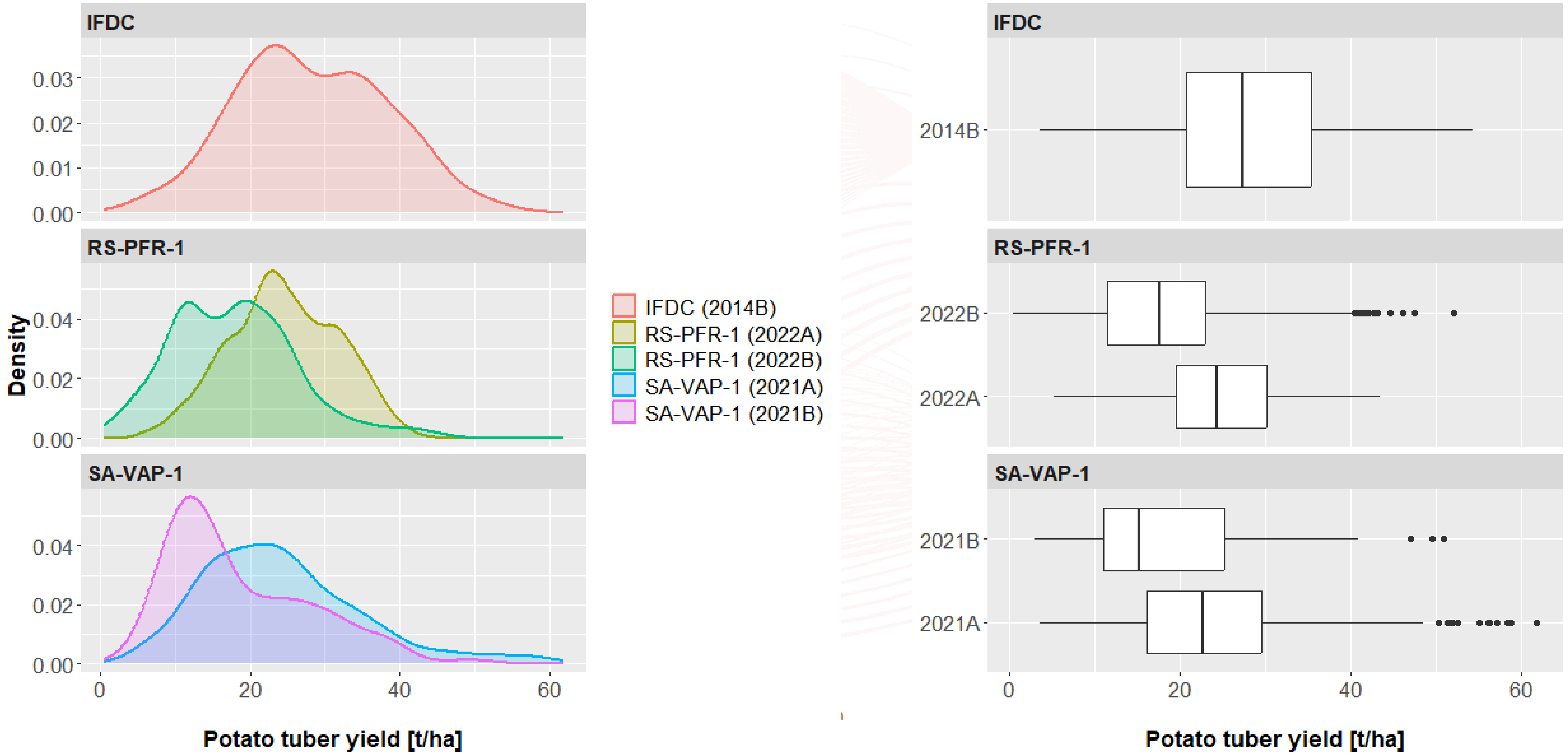
Treat	N	P	K
NPK 17x3	51	22	42
NP	51	22	0
NK	51	0	42
PK	0	22	42
NPK(30N)	30	22	42
NPK(60N)	60	22	42
NPK(90N)	90	22	42
NPK(30P)	51	30	42
NPK(40P)	51	40	42
NPK(50P)	51	50	42
NPK(30K)	51	22	30
NPK(60K)	51	22	60
NPK(80K)	51	22	80
Increased NPK	90	40	60
Control	0	0	0

AgWise analysis potato field trial data Rwanda

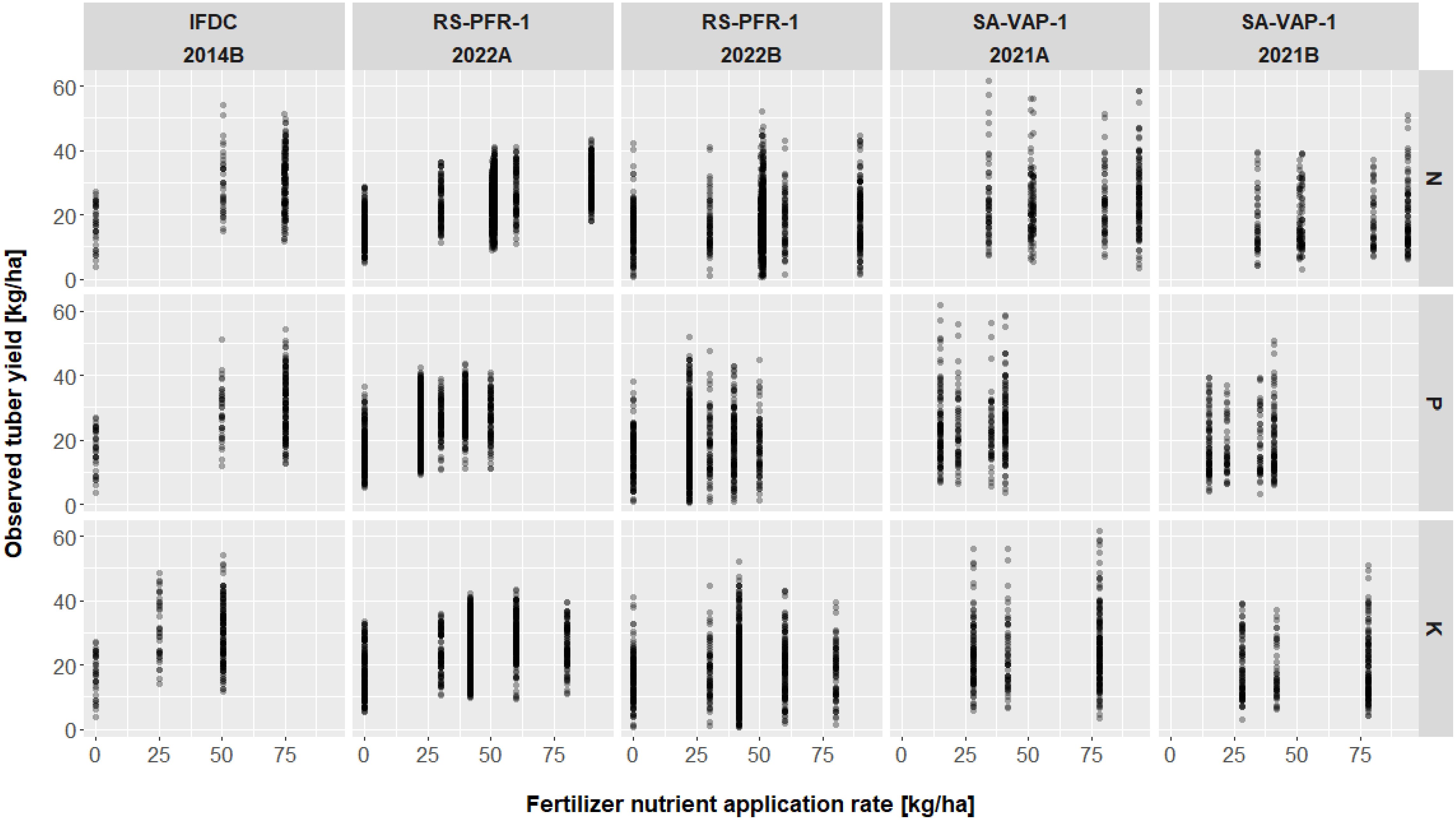


Overview yield distributions by expCode × season

Yield distributions are quite comparable for all three experiments, with mean yields around 15-25 t/ha:



Overview yield distributions by expCode × season



STEP 2: Extracting BLUPs (structural variation)

STEP 2: eliminate residual error through linear mixed effects modeling

Variation in the raw experimental data is composed of

- Overall treatment effect (application of N, P and K at varying rates)
- Season and experiment-specific effects (choice of variety, planting density, overall weather conditions,…)
- Trial-specific effects (location-specific differences in soil, weather and crop management conditions,…)
- Residual error (plot-specific effects, measurement error,…)

In order to evaluate the overall and location-specific effects of fertilizer application and relate this to digital soil map information and other predictors, the structural variation needs to be differentiated from the residual error. By eliminating the noise in the data and only carrying the meaningful signal forward, the performance of prediction models can be better evaluated.

This is done by fitting a linear mixed effects model:

```
sqrt(TY) ~ N + P + K + I(N^2) + I(P^2) + I(K^2) +
           N:P + N:K + P:K + N:P:K +
           season +
           (1|TLID) +
           (0 + N|TLID) + (0 + P|TLID) + (0 + K|TLID)
```

- Fixed effects of N, P and K (linear and quadratic terms in N, P and K)
- Fixed two- and three-way interactions between N, P and K
- Fixed seasonal effect (which allows for differences in yield between experiments and seasons)
- Random intercept for trial (allowing for differences in yield between individual trial locations)
- Random uncorrelated slopes for trial (allowing for different shapes in the response curve between trial locations)

A square root transformation is applied to ensure homoscedastic distribution of residuals.

The aim of this model is to describe the structural variation in yield, and eliminate the residual error. The model assumes a quadratic shape of the response curve, which is adequate if application rates are not too large to the extent that a plateau in the yield response curve is reached. In such situations a non-linear model with a logistic dose-response curve is required.

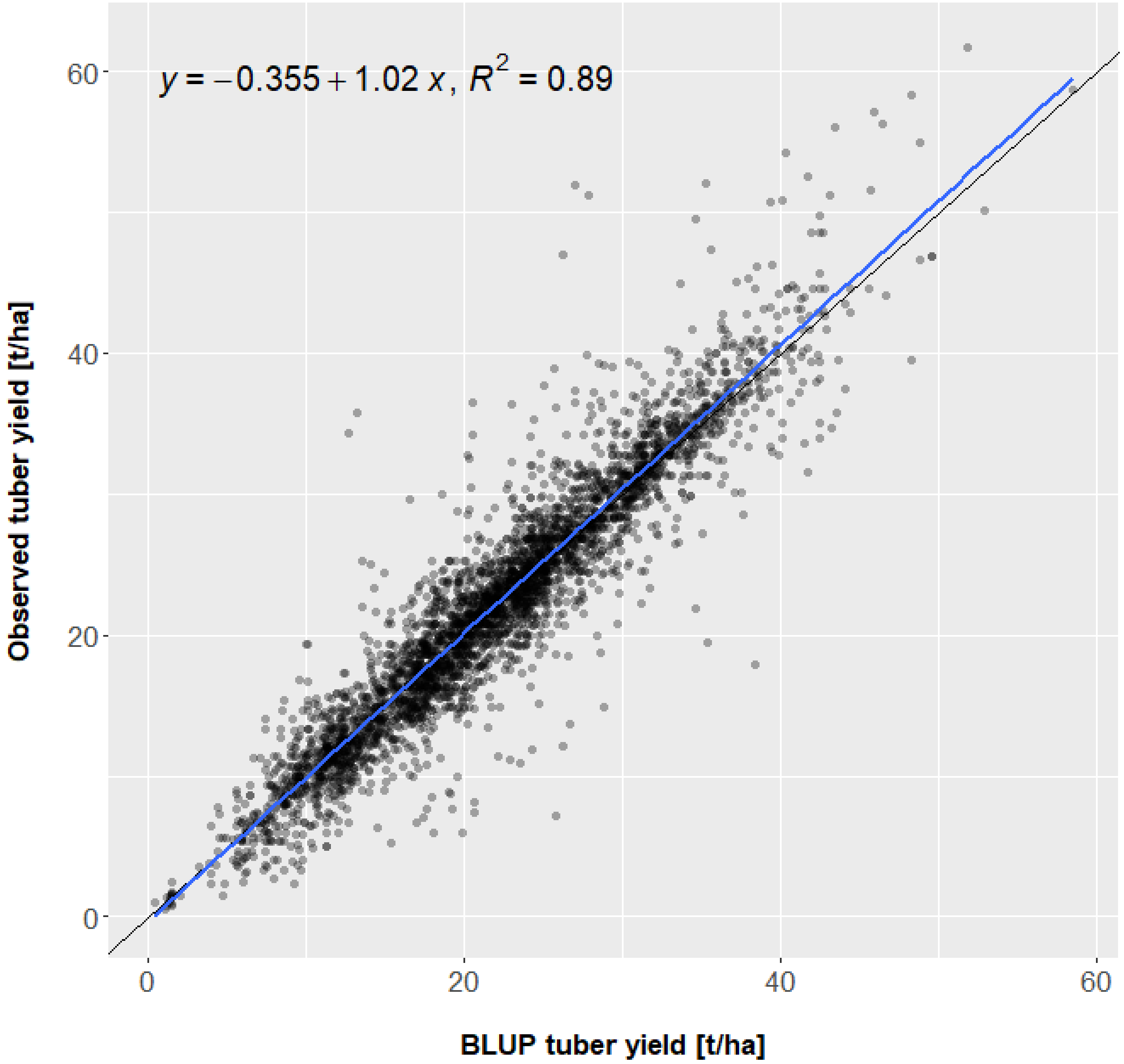
Performance of the linear mixed effects model

```
Linear mixed model fit by REML ['lmerModLmerTest']
Random effects:
 Groups   Name        Std.Dev.
 TLID     (Intercept) 0.5914   0.7690
 TLID.1   N100        0.2858   0.5346
 TLID.2   P100        0.2340   0.4837
 TLID.3   K100        0.1071   0.3272
 Residual             0.1325   0.3640 Number of obs: 3862, groups:  TLID, 336

Type III Analysis of Variance Table with Satterthwaite's method
               Sum Sq Mean Sq NumDF  DenDF F value    Pr(>F)
N             11.2954 11.2954     1 3412.8 85.2364 < 2.2e-16 ***
P              0.0581  0.0581     1 2566.0  0.4387 0.5078326
K              5.5760  5.5760     1 3357.4 42.0770 1.006e-10 ***
I(N^2)         0.7118  0.7118     1 2955.6  5.3712 0.0205399 *
I(P^2)         1.6754  1.6754     1 1788.4 12.6426 0.0003869 ***
I(K^2)         0.7733  0.7733     1 3038.4  5.8357 0.0157626 *
season         4.9172  1.2293     4  344.4  9.2764 3.954e-07 ***
N:P            0.7436  0.7436     1 2737.4  5.6110 0.0179167 *
N:K            0.9967  0.9967     1 3038.3  7.5213 0.0061330 **
P:K            0.2873  0.2873     1 2762.4  2.1680 0.1410204
N:P:K          0.4776  0.4776     1 3371.4  3.6040 0.0577269 .

Marginal R2 = 0.1517 (Fixed effects explain about 15% of the total variation)
Conditional R2 = 0.8702 (Fixed + random effects explain 87% of variation, and 13% is residual error)
```

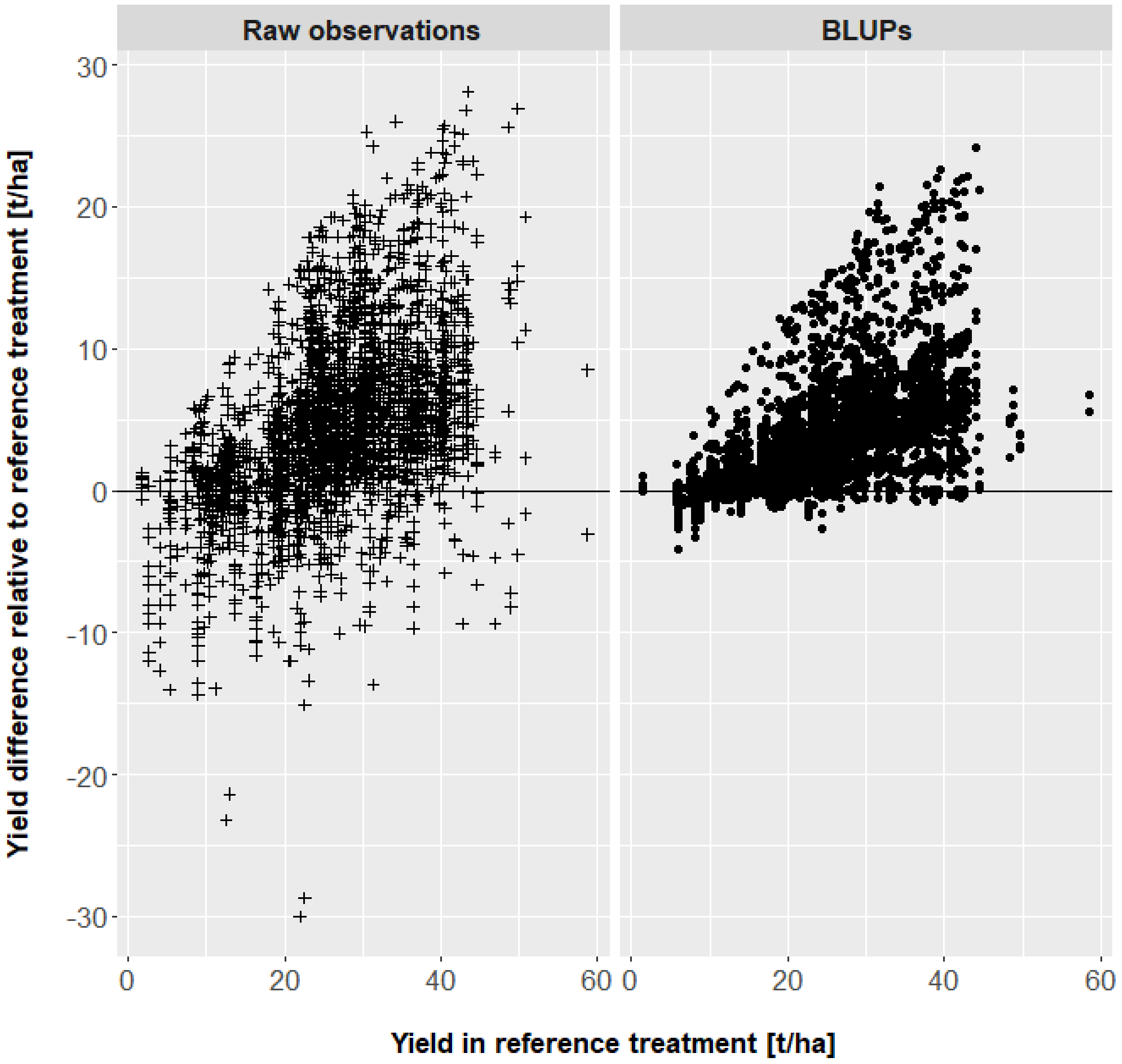
Performance of the linear mixed effects model



Why BLUPs?

BLUPs are Best Linear Unbiased Predictors. They "attenuate" the observed yield variation by distinguishing noise from signal, and represent the structural variation in the data. They assume that variation between locations, in terms of both mean yield (intercept) and yield response (slope) follow normal distributions. The procedure uses all data to make inferences about the structural (fixed + random terms) and residual variation.

BLUPs represent the structural variation in yield (and yield response)



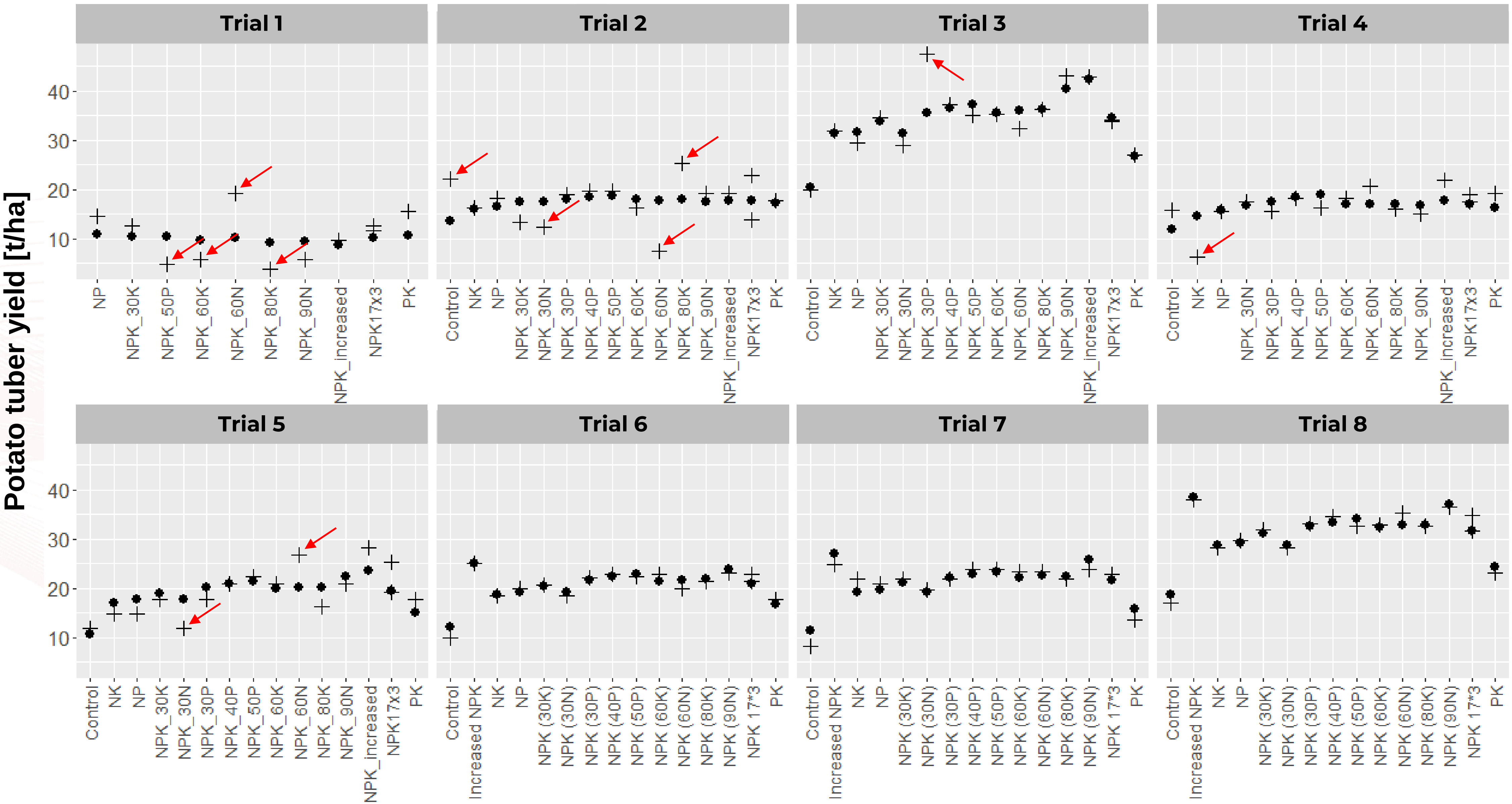
Comparing raw versus BLUP yield effects

We can calculate for each observation the yield difference to the reference treatment ($TY_{ref} - TY_{treat}$), and plot this yield difference against the yield in the reference treatment. We observe large variation in the raw observations, with frequent negative yield effects (meaning that yields in treatments with lower nutrient application have higher yields), which are not biologically meaningful. Variation in BLUP yield effects is lower, with fewer and smaller negative effects observed. It is therefore plausible that the mixed model approach has largely eliminated residual plot-level error (noise), and retained structural variation.

This can also be illustrated by looking at variation in yield response within individual trial locations (see next slide). BLUPs are closely aligned to raw observations in locations with consistent yield patterns. In locations with yield observations that deviate from overall patterns, BLUPs will attribute this as error and attenuate such values.

BLUPs represent the structural variation in yield (and yield response)

Example of BLUPs (●) and raw observations (+) for 8 selected trials. Irregular observations are attenuated.



STEP 3: Determining indigenous nutrient supply

Reverse QUEFTS

The "reverse QUEFTS" is an optimisation algorithm that seeks the values for the soil indigenous N, P and K supply that best explains the observed yield responses to the fertilizer application rates in each individual field trial.

Data from different experiments with different treatment structures are standardised and each set of observations is described by a single set of INS, IPS and IKS values.

For every trial ID i:

TLID	Treat	N	P	K	blup
i	1	N1	P1	K1	Y1
i	2	N2	P2	K2	Y2
i	3	N3	P3	K3	Y3
i	...	...	...	...	...
i	n	Nn	Pn	Kn	Yn

n treatments with different N, P and K fertilizer application rates and observed yields

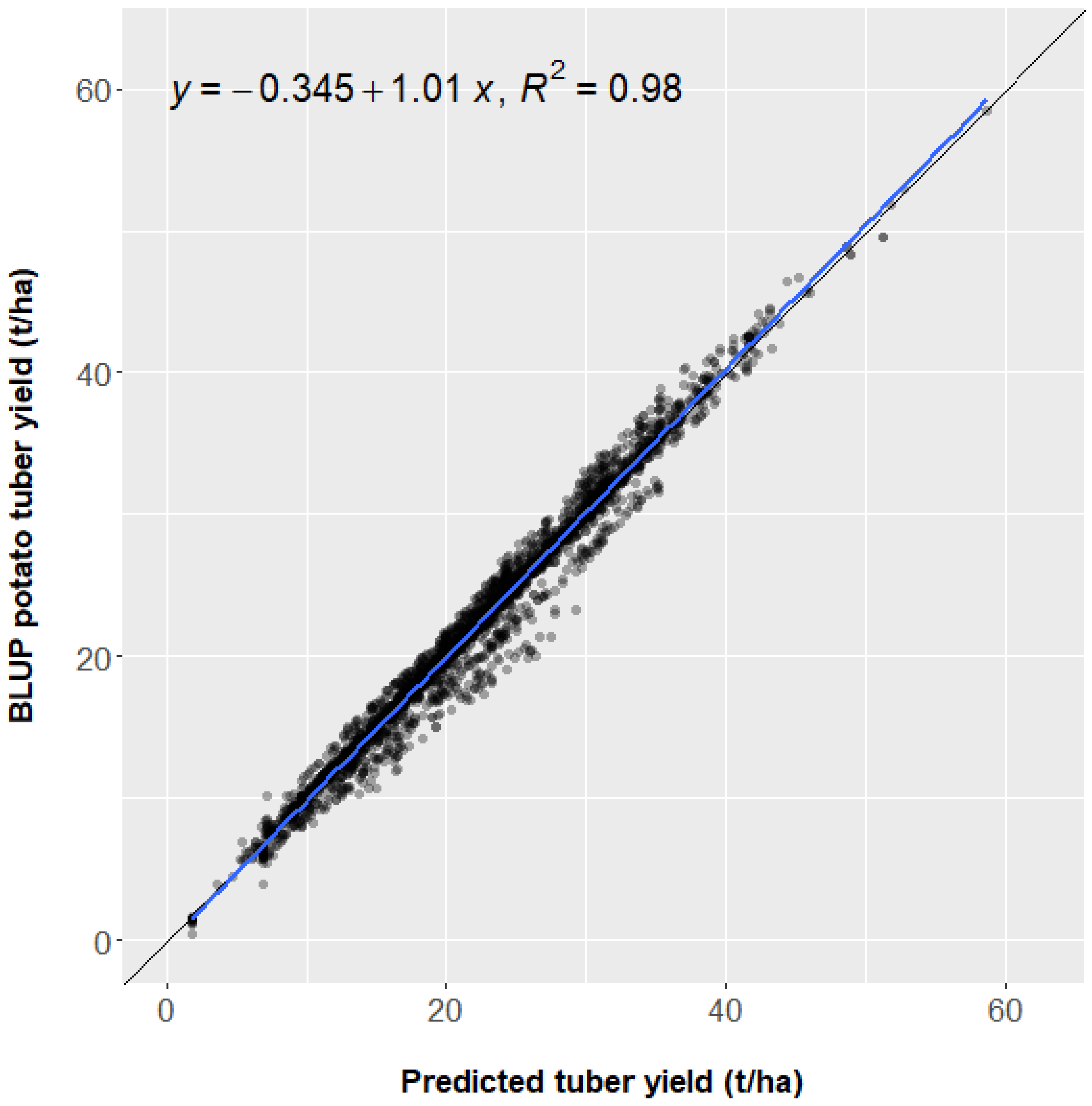
`revQUEFTS` →

TLID	INS	IPS	IKS	aY
i	INSi	IPSi	IKSi	aYi

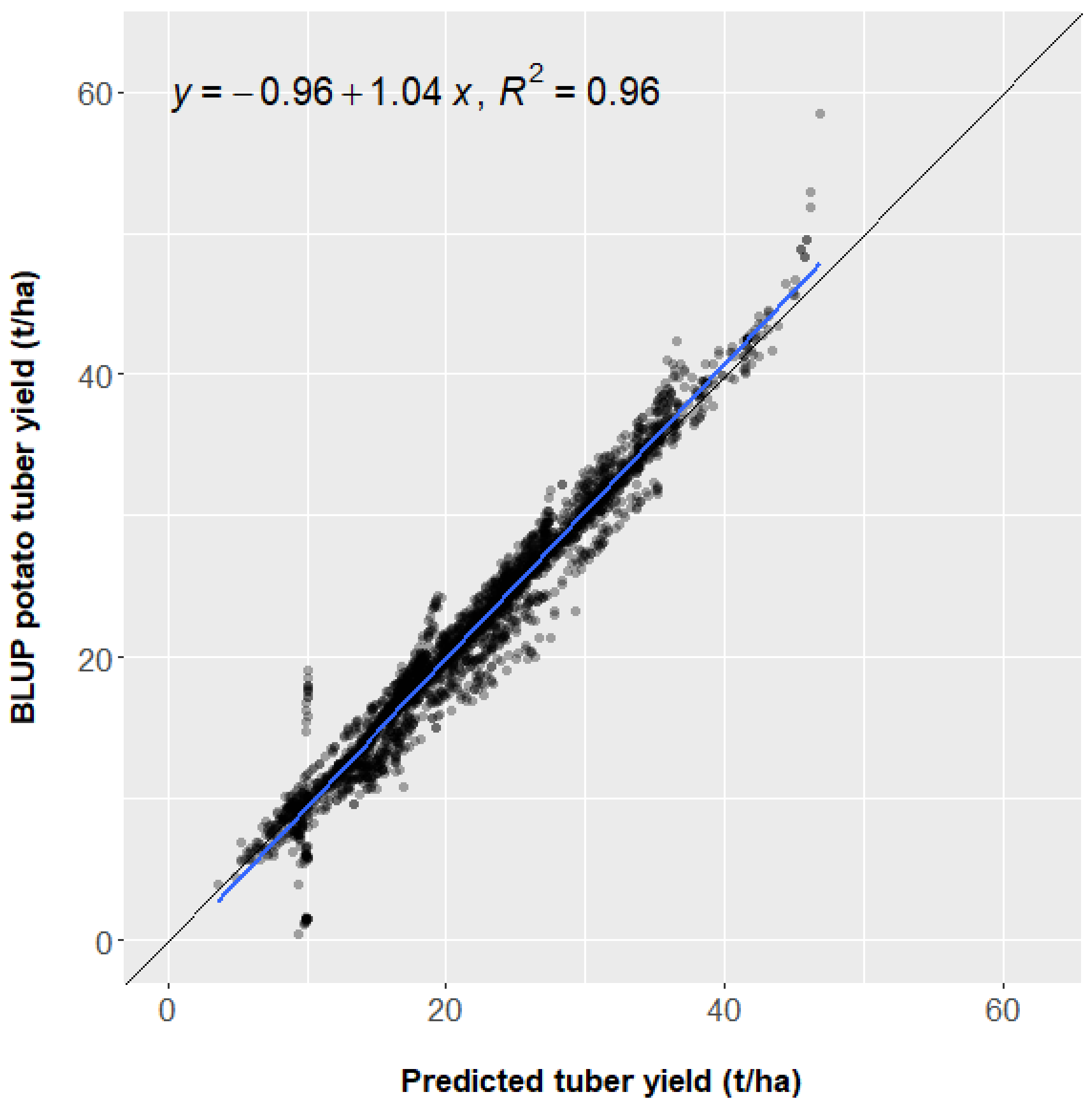
1 set of apparent soil indigenous N, P and K supply values and an attainable yield.
The attainable yield is set to 20% above the blup yield in the reference treatment.

How well does reverse QUEFTS describe the data?

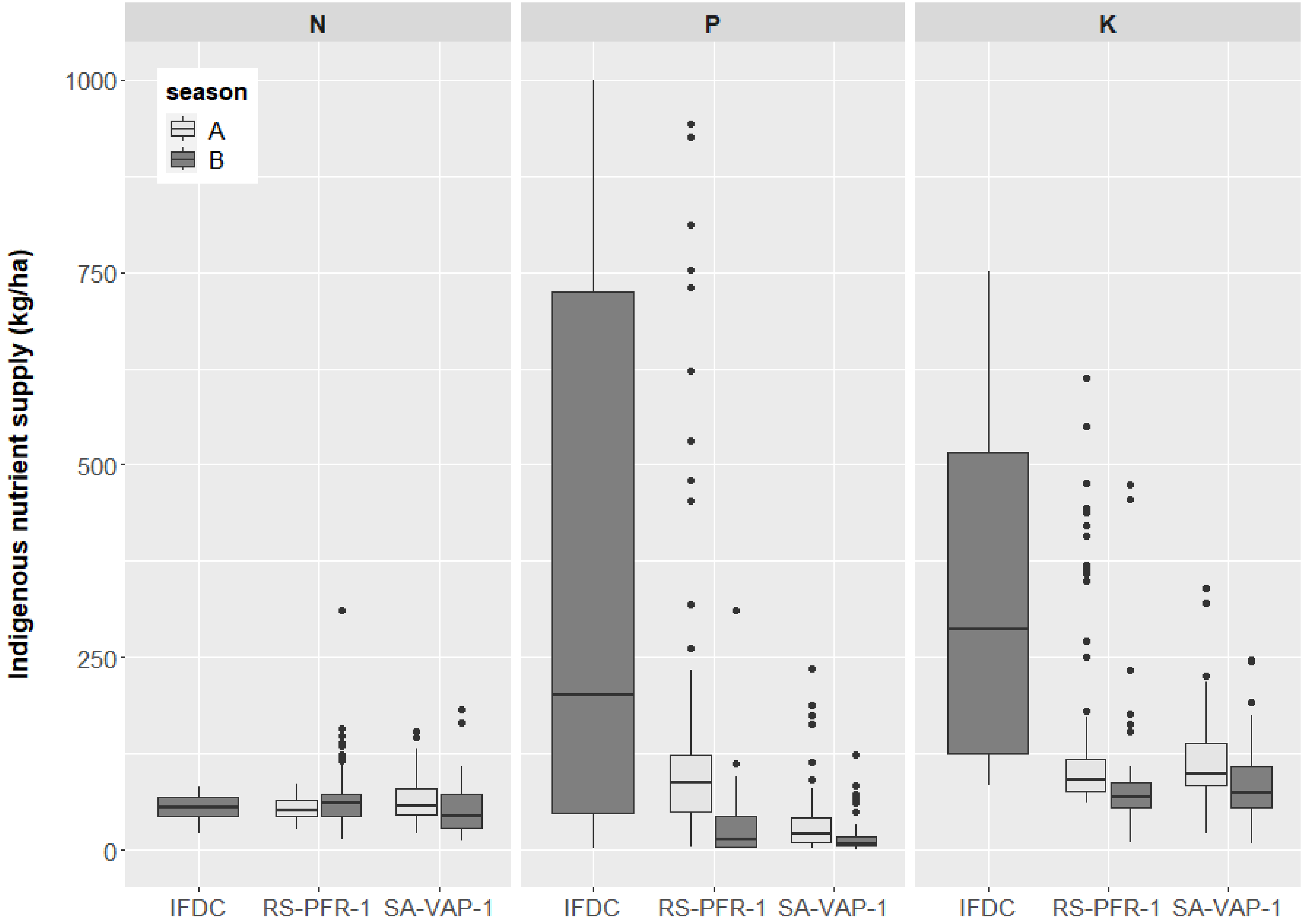
Using observed attainable yield



Using discretised attainable yield classes



Ranges in apparent indigenous nutrient supply



- N supply is similar in all experiments and seasons, and rather low with little variation.
- P supply is very low in both seasons for SA and the B season for RS. In the RS A season, P supply is higher. Ranges in P supply are very wide for IFDC
- K supply is comparable for SA and RS, and somewhat higher in the A season than the B season. K supply is very wide for IFDC
- In many IFDC trials, there is concurrent absence of response to P and K, resulting in high estimated values for P and K supply.
- Simulated yields are little affected by nutrient addition when N or K supply exceeds 400 kg/ha and P supply exceeds 100 kg/ha.

Machine learning using digital soil map pars to predict indigenous nutrient supply

A random forest model is fitted using all available soil map parameters, digital elevation geodata, administrative and AEZ classes, the season (A versus B) and the reference yield class as predictors for apparent soil N, P and K supply.

Supply values are log-transformed to reduce the weight of high values (since these are less influential on yield).

If the reference yield and season are meaningful predictors, then this will result in different recommendations formulated for each season and reference yield class. The user will need to select the relevant yield class for his/her field. It reflects the expected yield based on his/her past experience, given the soil, weather and management conditions of the production system.

Performance of the model is evaluated using leave-one-out cross validation: each observation is predicted using a model trained with all other observations.

For every trial ID i:

TLID	orC	totN	…	season	refY
1	…	…	…	A	refY1
2	…	…	…	B	refY2
3	…	…	…	B	refY3
…	…	…	…	…	…
n	…	…	…	A	refYn

Set of soil parameters from iSDA + soilGrids + season + reference yield class* (refY) for each trial ID

`randomForest` →

TLID	INS	IPS	IKS
1	INS1	IPS1	IKS1
2	INS2	IPS2	IKS2
3	INS3	IPS3	IKS3
…	…	…	…
n	INSn	IPSn	IKSn

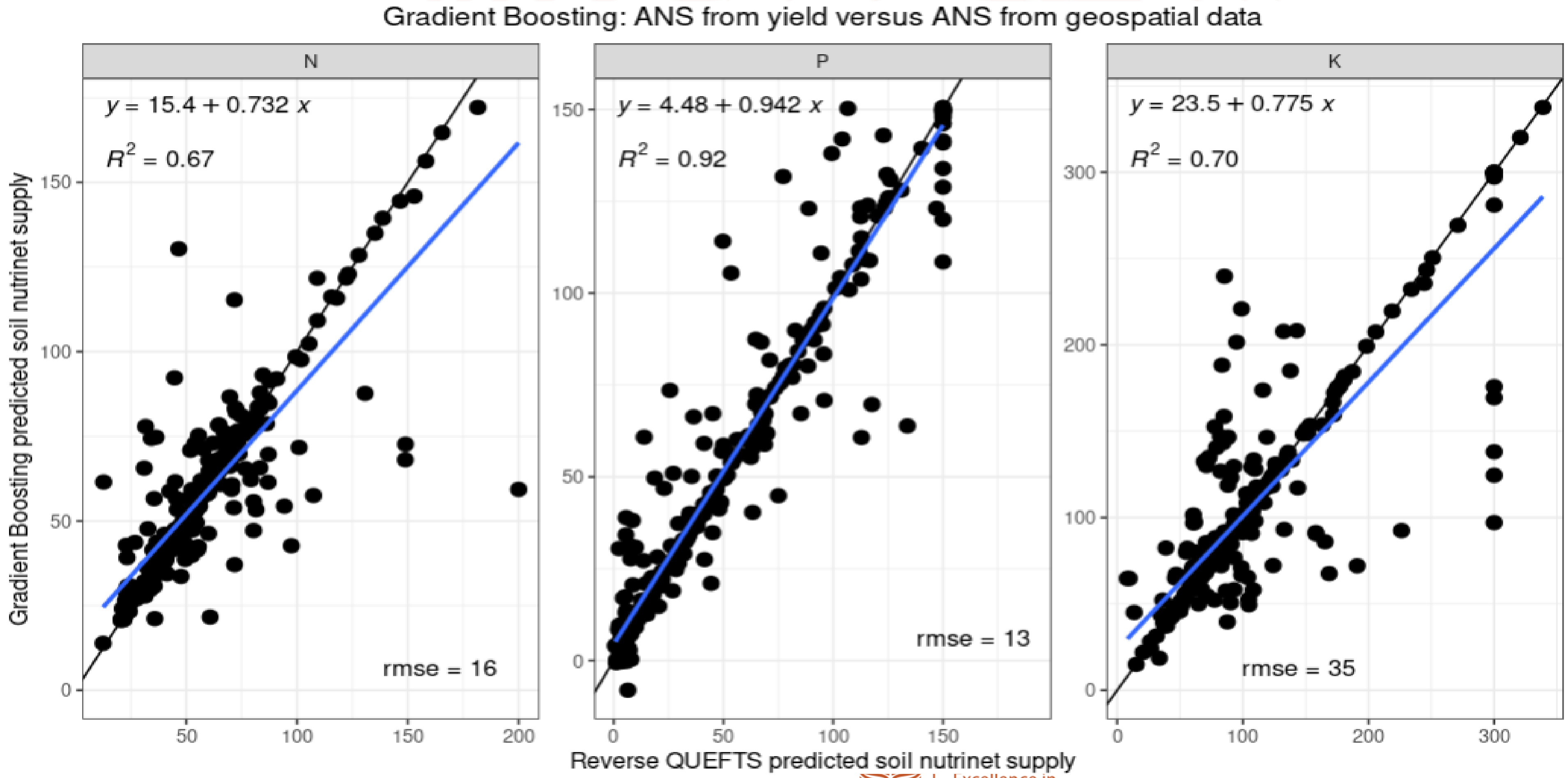
Set of apparent soil indigenous N, P and K supply values for each trialID

refY is the discretised yield in the reference treatment [t/ha]: very low = [0, 10), low = [10, 20), medium = [20, 30), high = [30, 40) and very high = [40, Inf)

CGIAR
INITIATIVE ON
Excellence in
Agronomy

Machine learning: Gradient Boosting

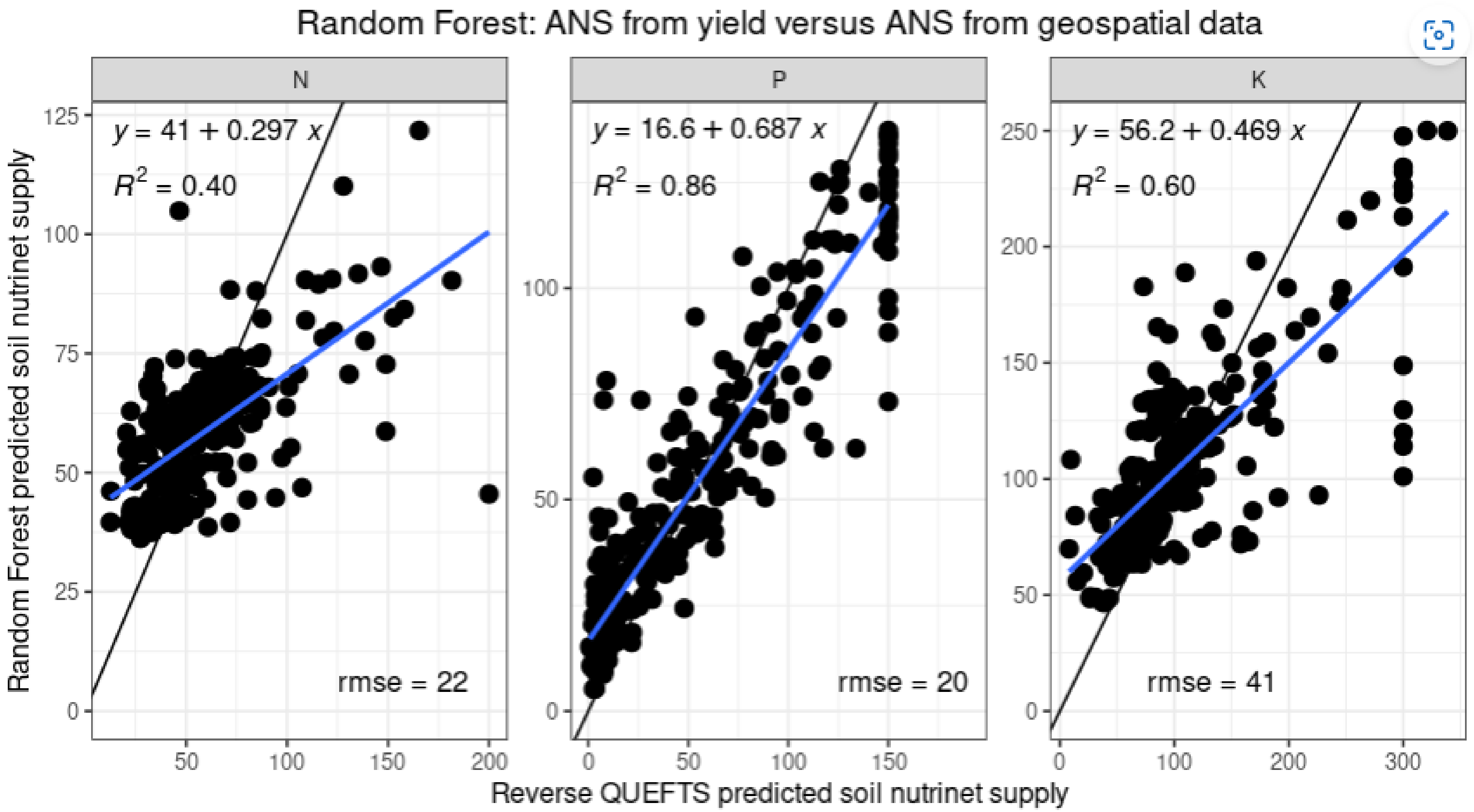
Soil properties extracted from digital soil maps of iSDA and soilGrids are used to predict Apparent Nutrient soil Supply (ANS) for N, P and K



CGIAR | Excellence in Agronomy

Machine learning: Random forest

Soil properties extracted from digital soil maps of iSDA and soilGrids are used to predict Apparent Nutrient soil Supply (ANS) for N, P and K

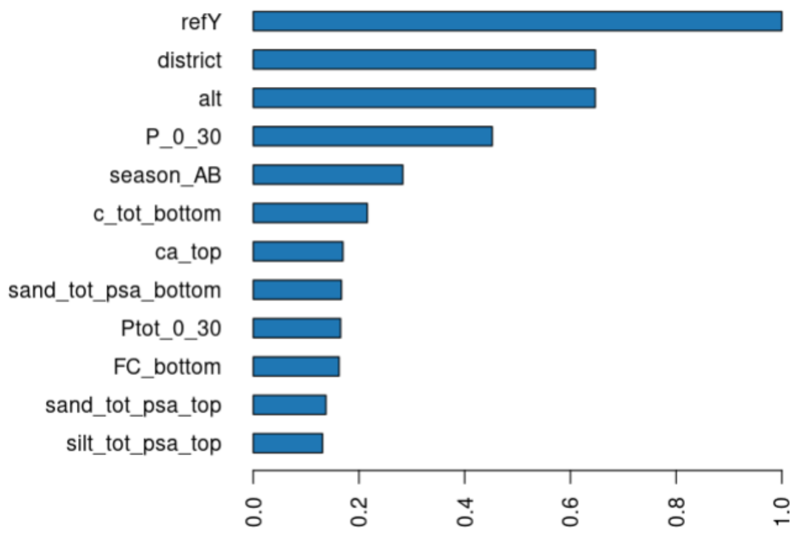


Which variables are most important to predict N, P and K supply?

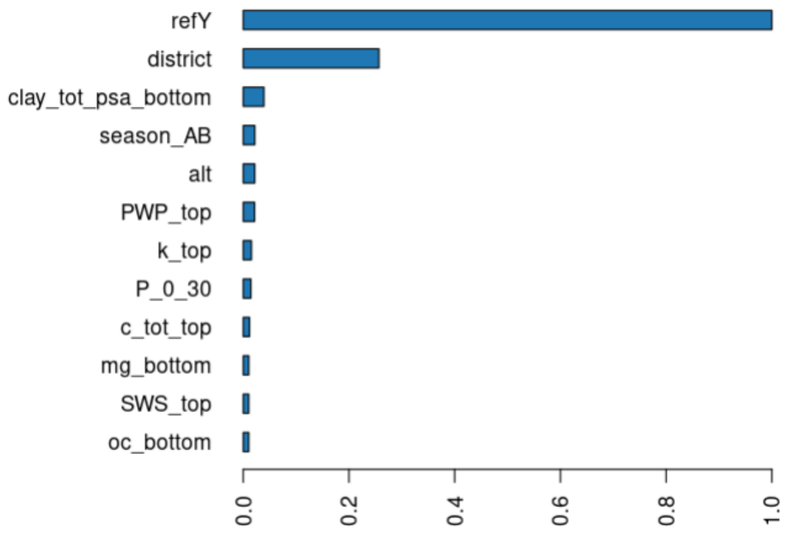
Given the gradient boosting performs better than random forest from here on we use gradient boosting

Variable importance indicates the % reduction in mean square error when the variable is omitted as predictor from the model.

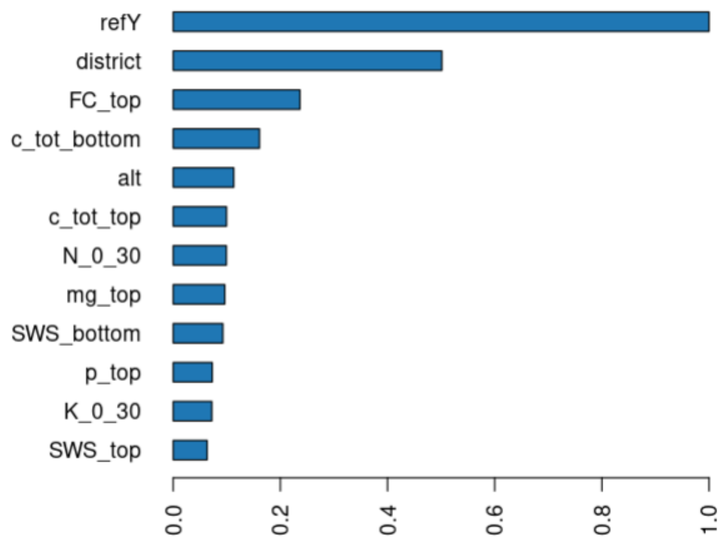
N supply



P supply



K supply



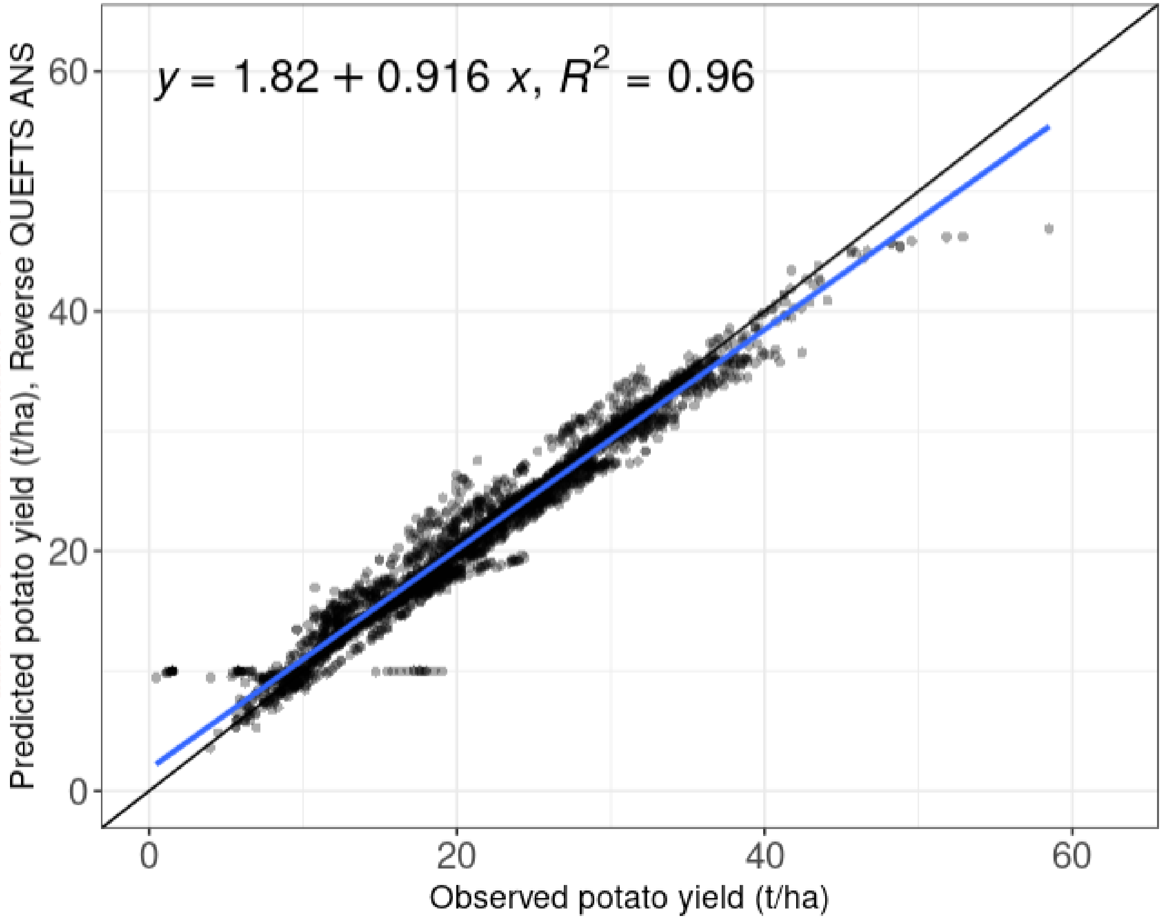
INITIATIVE ON
Excellence in
Agronomy

STEP 4: Predicting yield and yield response

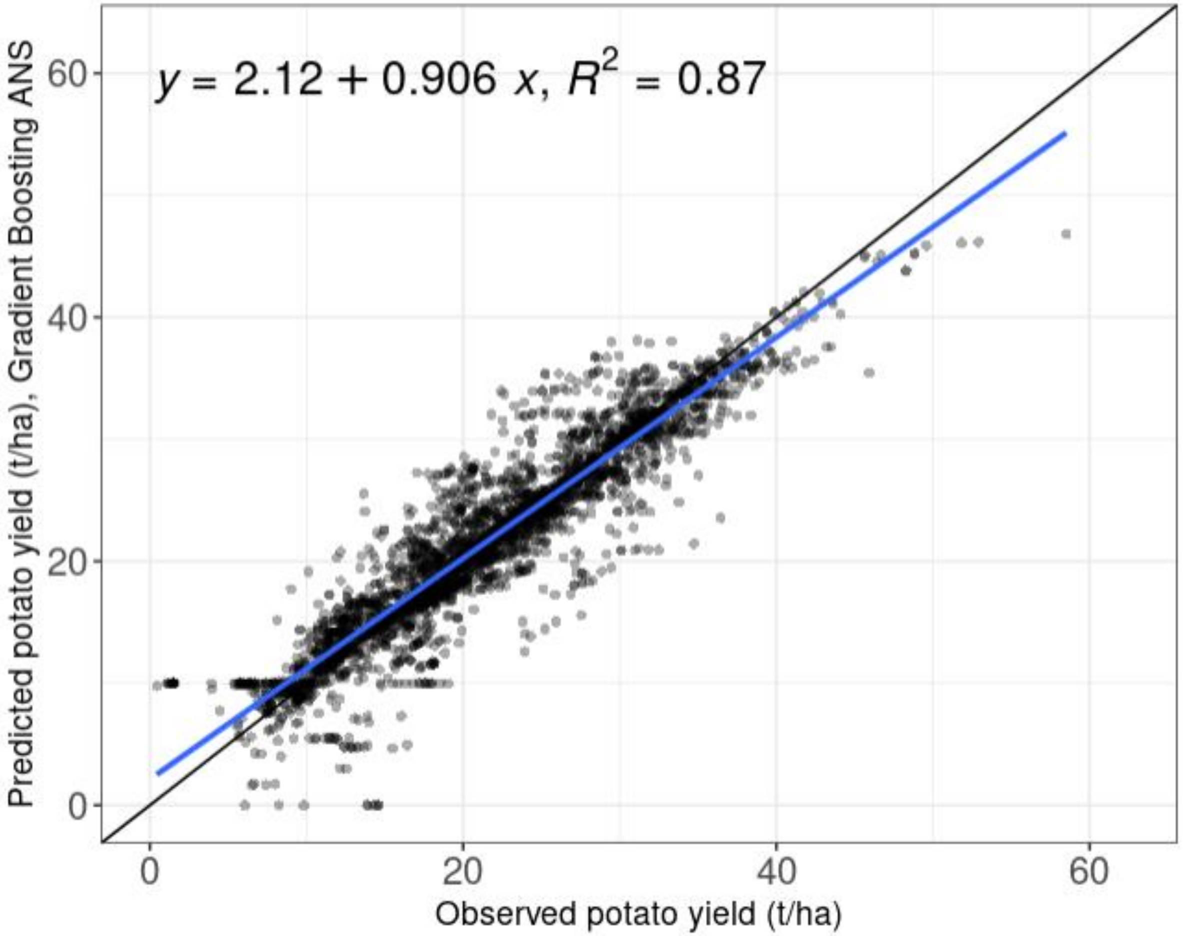
INITIATIVE ON
Excellence in
Agronomy

How well can we predict the observed variation in yield?

Observed versus predicted yield as predicted using ANS estimated using reverse QUEFTS



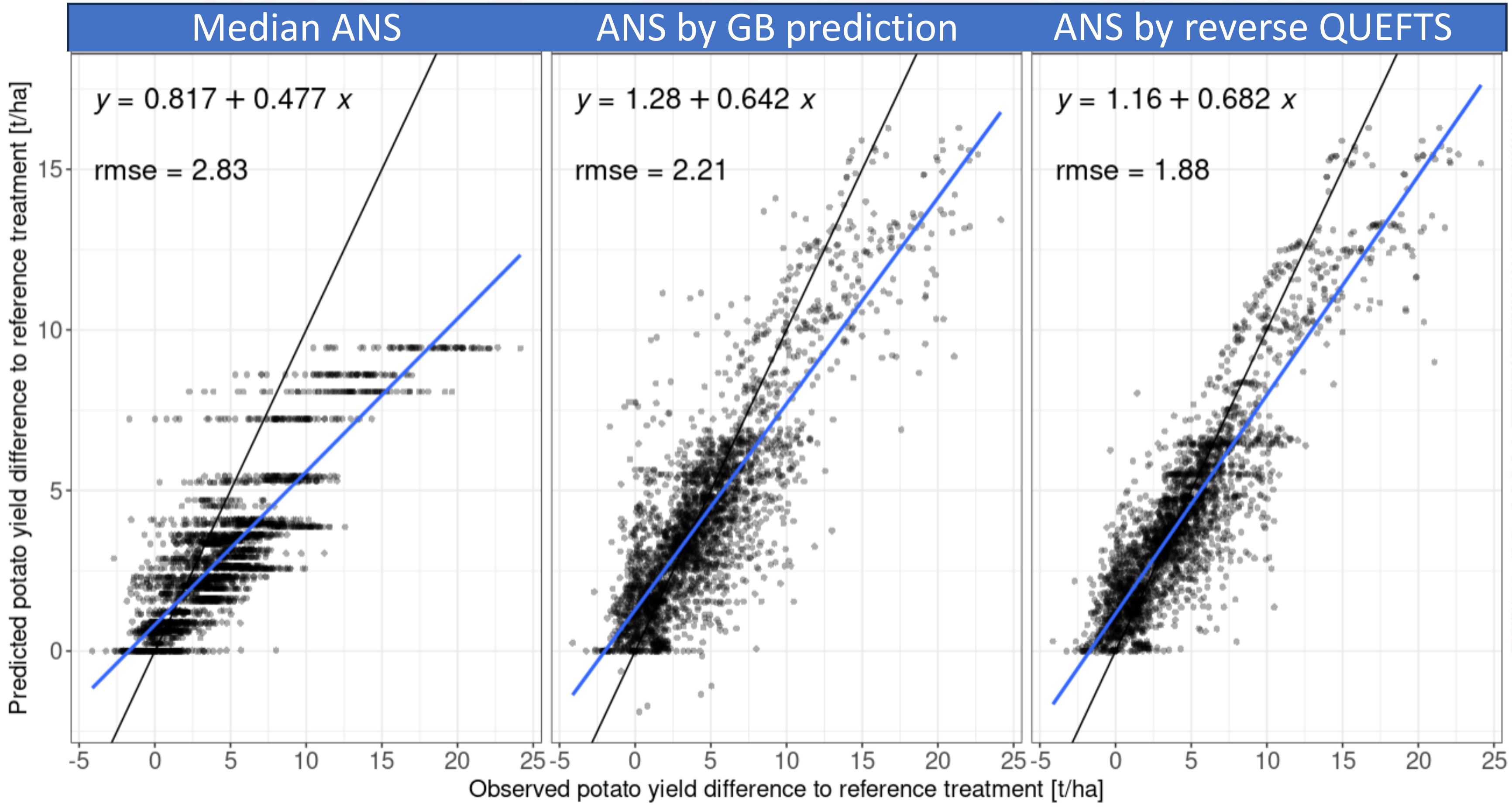
Observed versus predicted yield, using ANS estimated using gradient boosting method



Obviously the ANS estimates from reverse QUEFTS is more accurate than the machine learning result. However, for scaling purpose, the second approach should be used using the secondary geospatial data.

How well can we predict yield responses?

The alternative we have after getting the ANS from reverse QUEFTS is i) using the median nutrient supply for the region ii) using machine learning approach to estimate site specific ANS. The result of using these two approaches on yield effect is compared with the result of the reverse QUEFTS result below.



Effect of datasets and refY + season predictors on model performance

Two datasets require further investigation:

- IFDC due to the overall higher yields, different yield response structures, large weight on the calculation of the BLUPs, and the high P and K supply values obtained through the reverse QUEFTS procedure
- RwaSIS season A: lower residual variation, lowest impact of lmer on BLUPs, and somewhat higher P and K supply values obtained by reverse QUEFTS, relative to the SA and RS season B data.

Excluding refY and season as predictors could be considered as recommendations will not permit differentiating based on farmer's estimation of yield or yield in the previous season, nor will the delivery system allow different recommendations for the A versus B season.

We can repeat the above procedure for a set of 24 scenarios:

- All data or excluding each time one of the five datasets (6)
- With or without season as a predictor (2)
- With or without refY as a predictor (2)

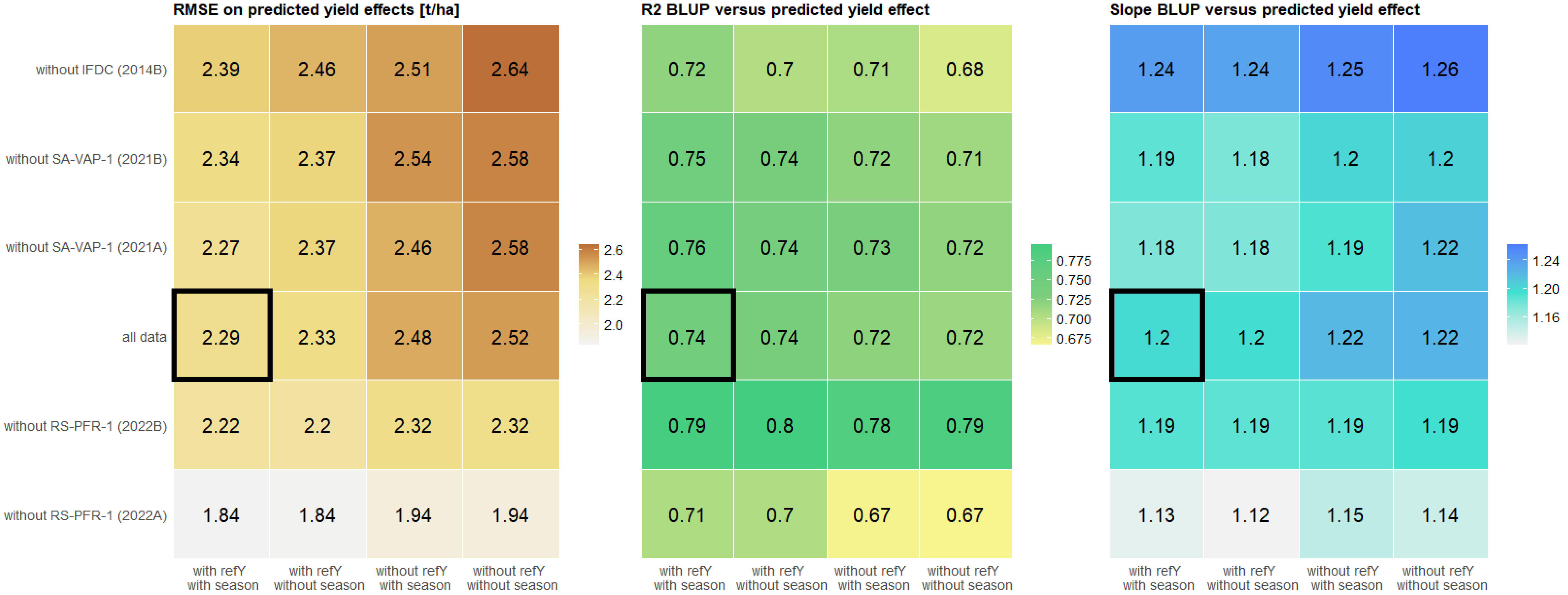
In each scenario, we can then calculate:

- The root mean square error (RMSE) on the predicted yield effects (relative to the reference treatment)
- The R^2 for BLUP versus predicted root yield effect
- The slope between BLUP and predicted root yield effect (values >1 signify an underestimation of the yield effect)

Effect of datasets and refY + season predictors on model performance

Excluding the IFDC data results in overall poorer predictions (higher RMSE and larger underestimation of yield), while excluding the RS data, particularly for the 2022A, generally improves predictions (lower RMSE, and less underestimation of yield).

Excluding refY as a predictor increases RMSE, but less so in scenarios where the RS data is excluded; excluding season as a predictor has less impact on the RMSE.



STEP 4b: Using machine learning directly on BLUPs

Alternative approach: predict yield directly using Gradient boosting model

As an alternative to above steps 3 and 4, the blup yield can be predicted directly by a gradient boosting model, and yield effects relative to the reference treatment calculated using the predicted yields.

The advantage of this method is that it is simpler and easier to implement and automate.

```
blup ~ N + P + K +
       [soil pars] +
       [dig elevation pars] +
       AEZ + district + season_AB + refY
```

Available predictors include:

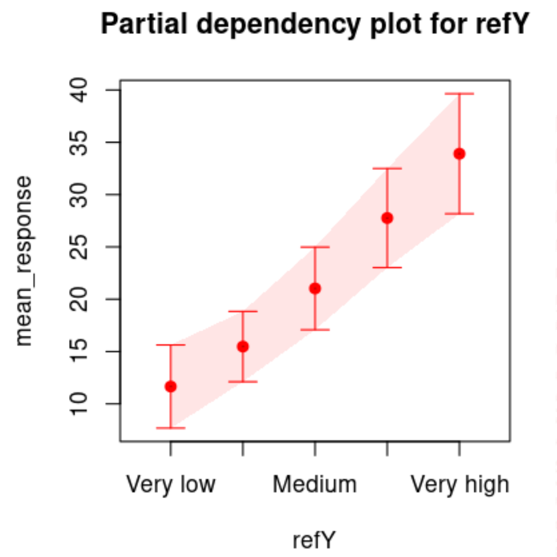
- Fertilizer N, P and K rate (N, P, K)
- Digital soil information (ISRIC + iSDA)
- Digital elevation data (altitude, slope, TDI, TRI)
- Agro-ecological zones (AEZ) and district
- Season_AB and refY

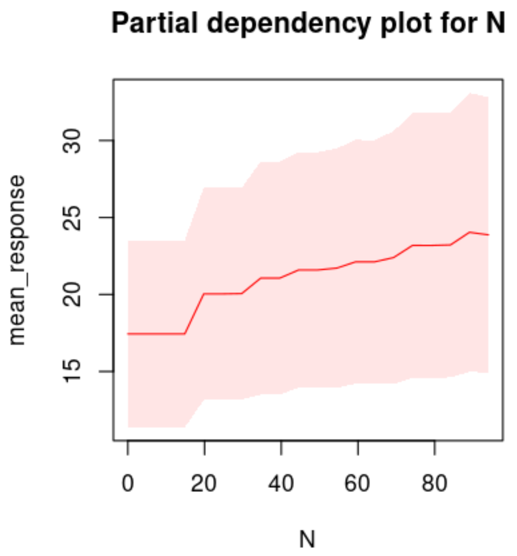
Yield is predicted following the same principle as LOOCV; cross-validation is done by folding the data by trial. For each trial, yield in the different treatments is predicted using a model trained with the data from all other trials.

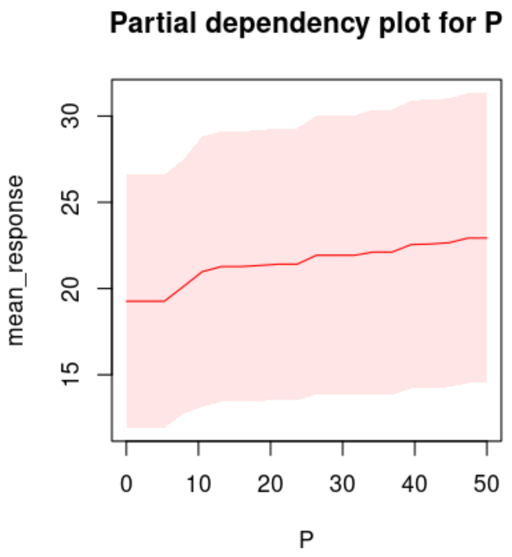
A second model is fitted excluding district, refY and season_AB, as recommendations are needed that are only dependent on spatial variables, and not all districts have been (sufficiently) covered.

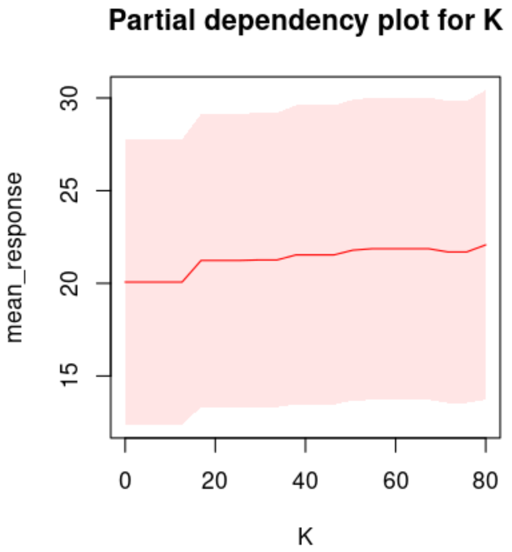
Alternative approach: predict yield directly using Gradient boosting model

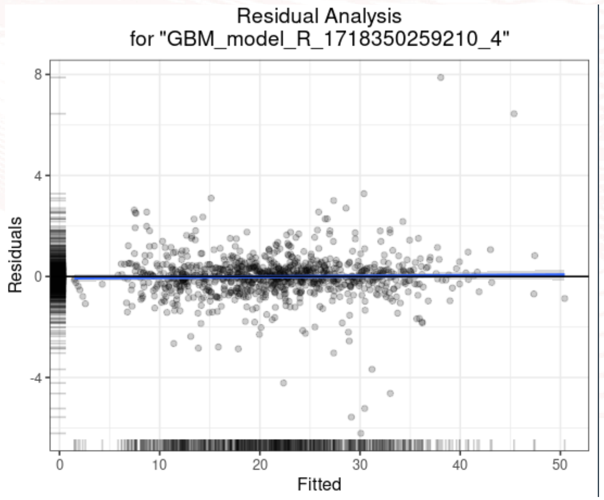
Variable importance indicates the % reduction in mean square error when the variable is omitted as predictor from the model.

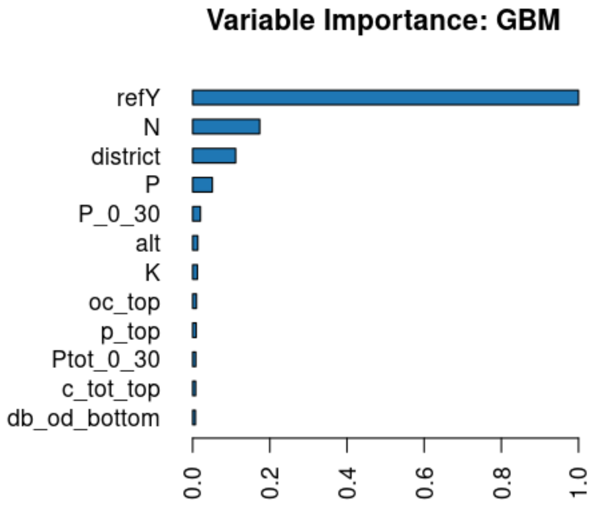


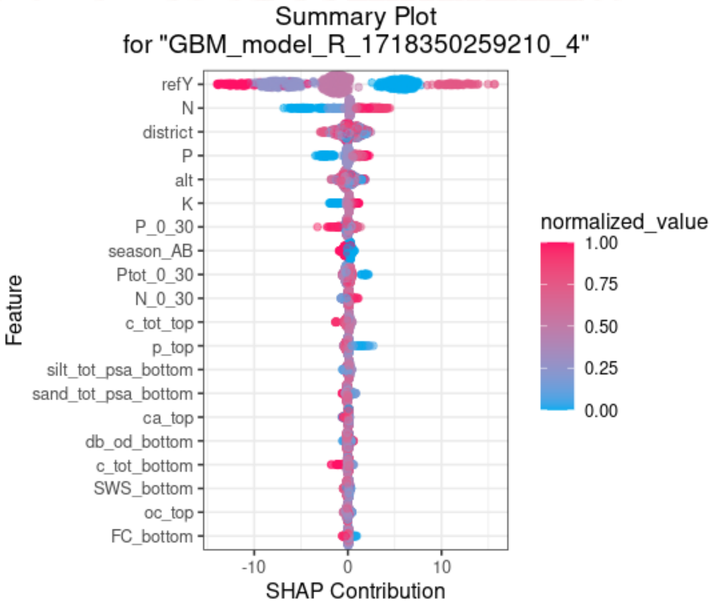




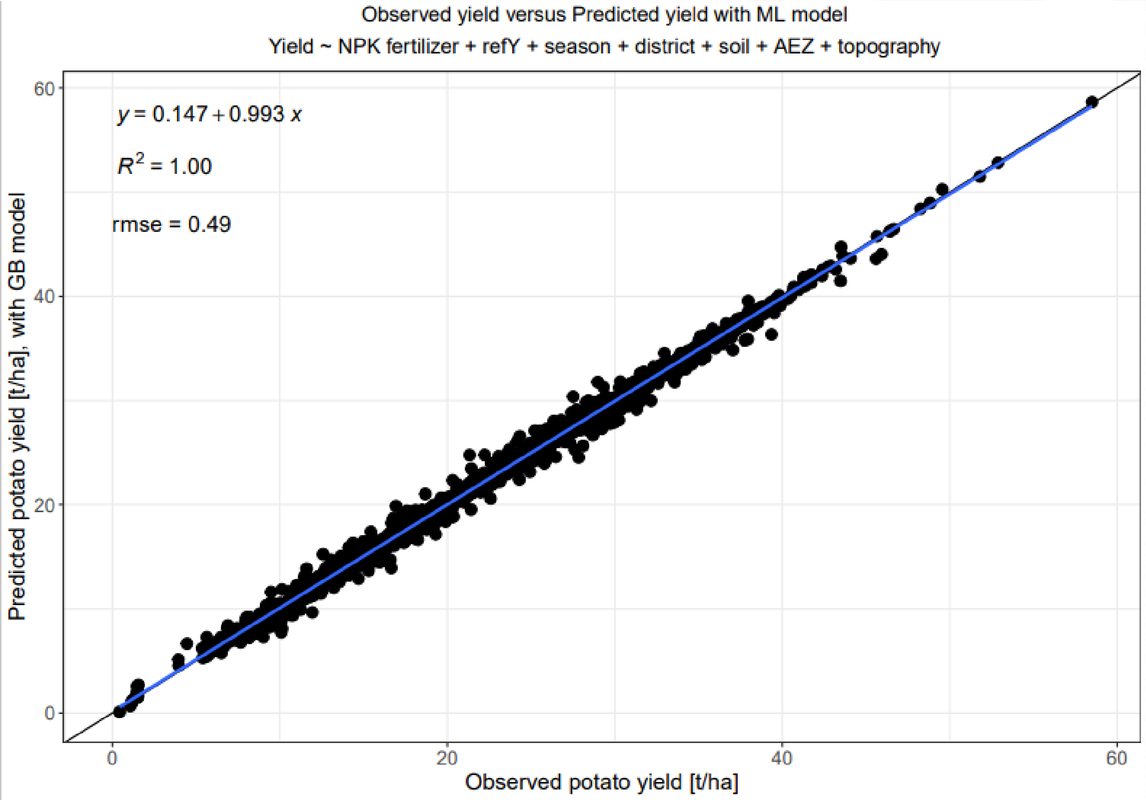


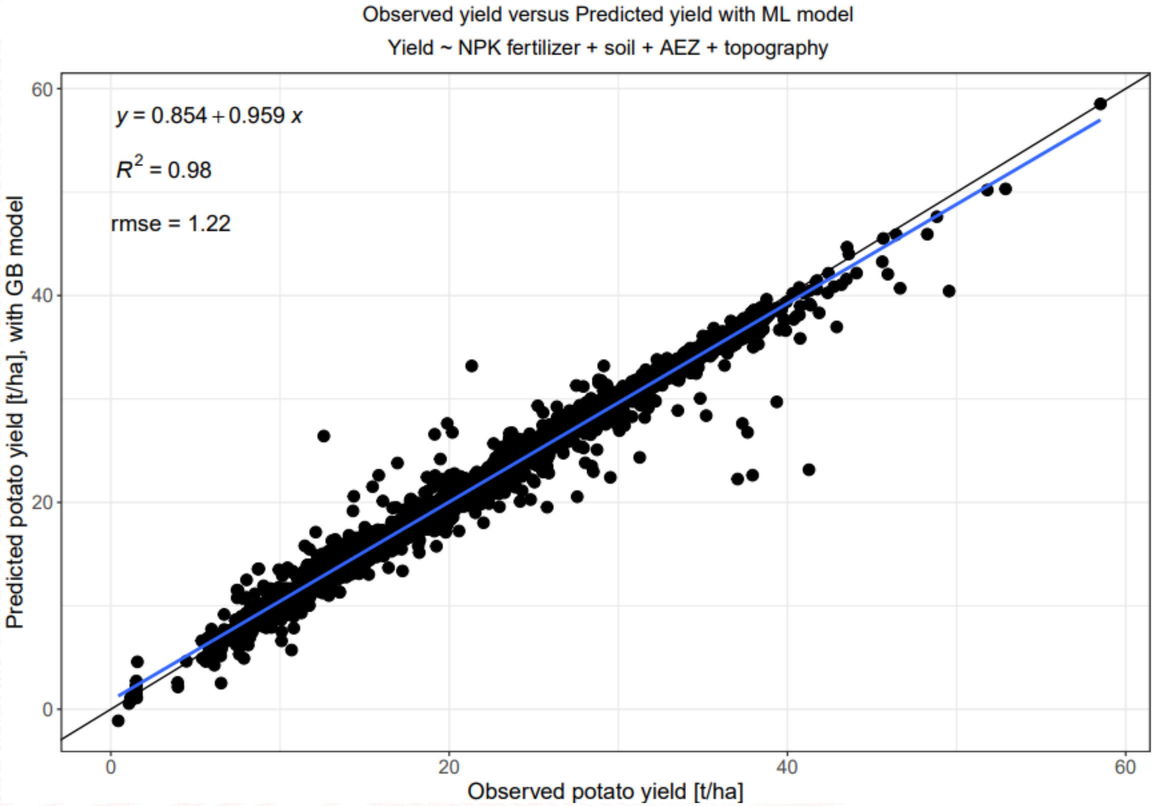






How well can we predict yield directly using Gradient boosting model?



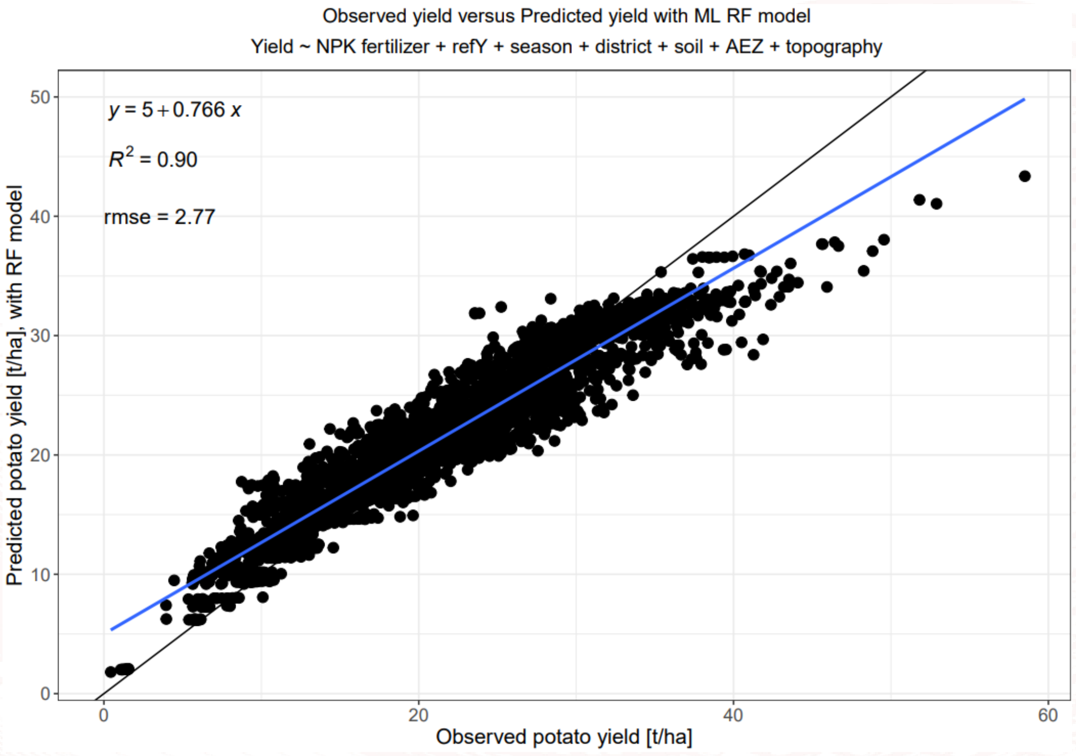


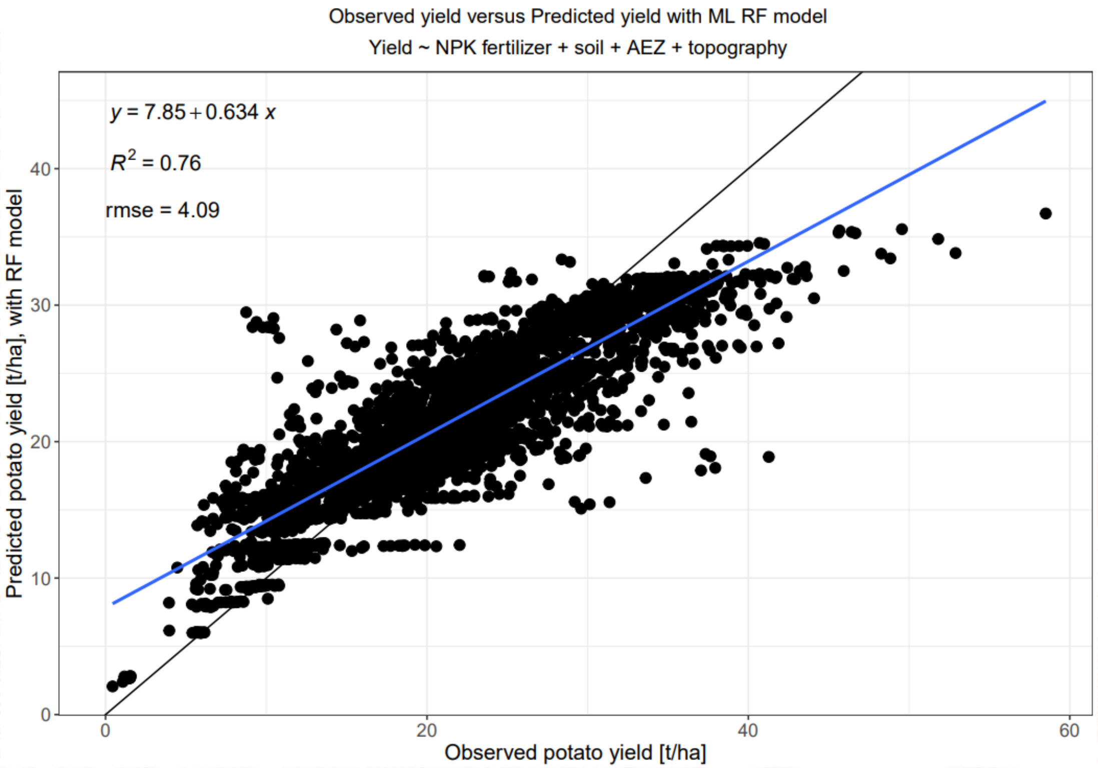
$Yield \sim N + P + K + refY + Season + district + AEZ + district + soil\ properties + elevation$

$Yield \sim N + P + K + AEZ + district + soil\ properties + elevation$

INITIATIVE ON
Excellence in Agronomy

How well can we predict yield directly using Random forest model?





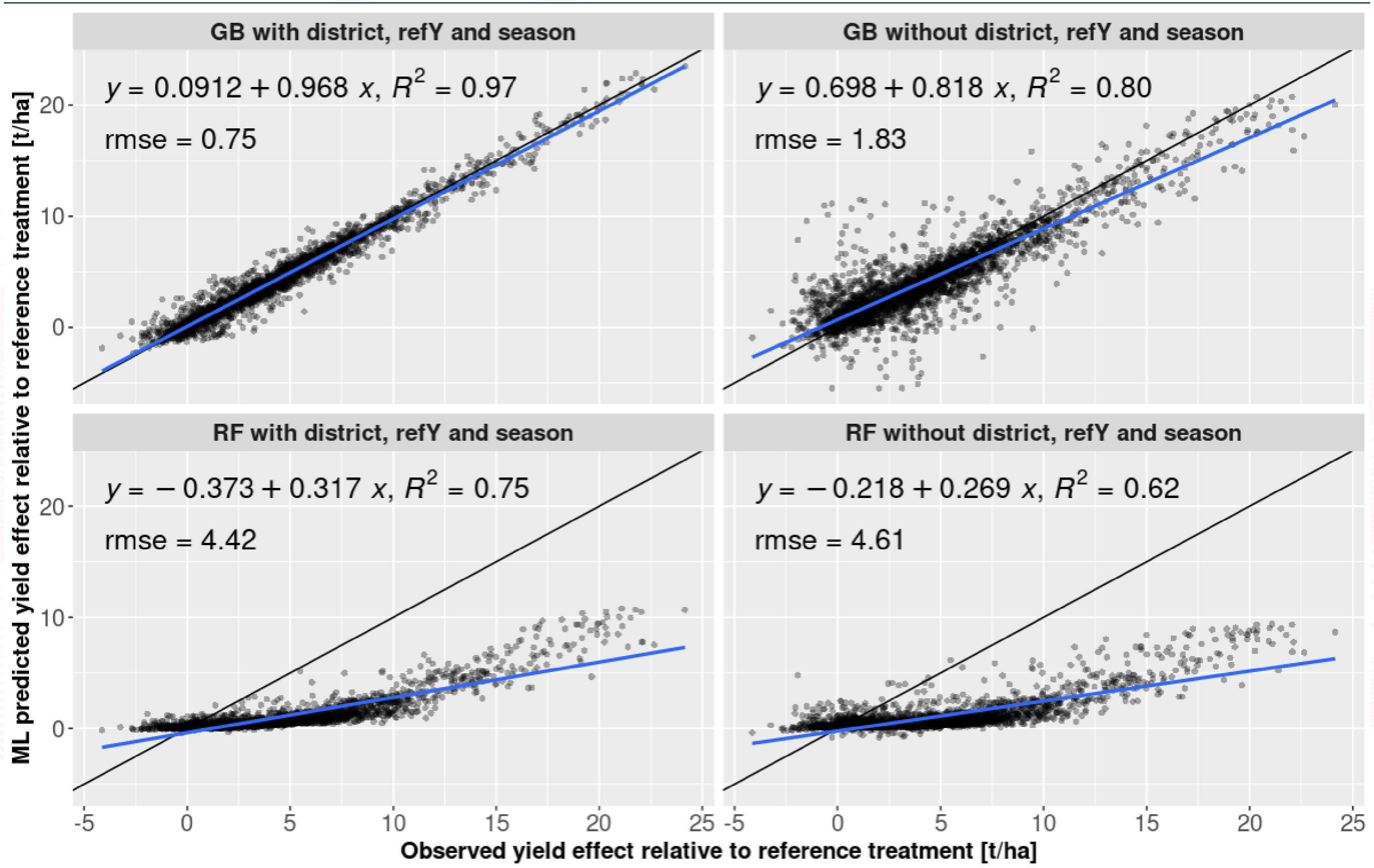
$$Yield \sim N + P + K + refY + Season + district + AEZ + district + soil\ properties + elevation$$

$$Yield \sim N + P + K + AEZ + district + soil\ properties + elevation$$

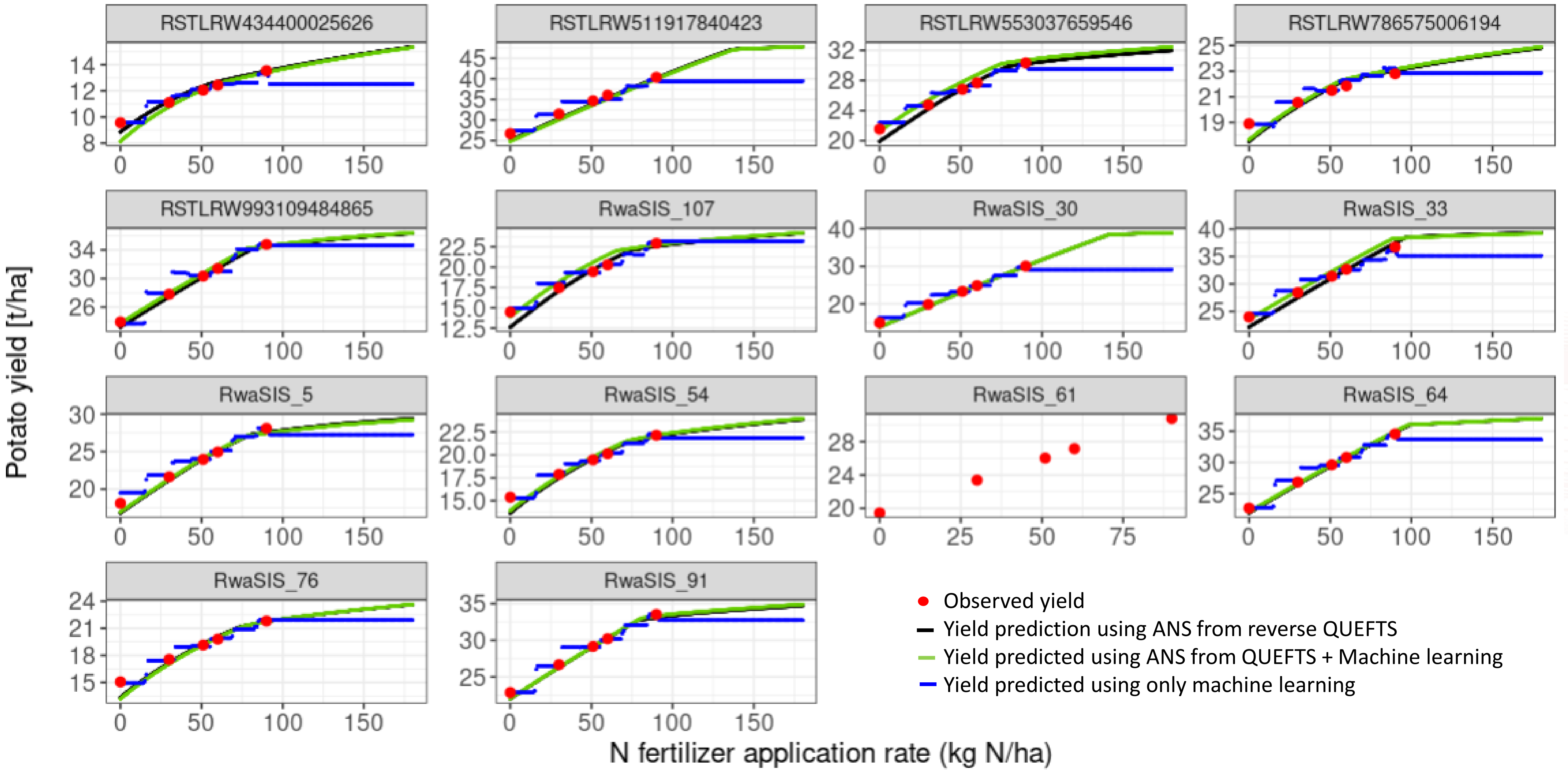
CGIAR | INITIATIVE ON Excellence in Agronomy

BLUP versus predicted yield responses relative to the reference treatment

Predicted yield responses are calculated using the predicted yields by the GB and RF model.



Simulated response curves and BLUP yields in selected RwaSIS trials



Conclusions on alternative approach: predict yield directly using RF

The approach is much simpler to implement and automate but has several shortcomings…

1. Good predictions can only be obtained when refY is included in the model. Without refY, the RMSE is larger, and the variance explained is lower than with refY included. However, this is logical since refY is derived from the yield data itself.
2. The machine learning direct prediction method underestimates yield effects. This is a consequence of the model tending to underestimate yield in the higher ranges, and overestimate yield in the lower ranges. This underestimation is more substantial than the QUEFTS approach.
3. Most influential parameters on yield (in addition to refY) are the fertilizer nutrient rates, altitude and organic carbon. The model however does not provide mechanistic insights in soil supply and fertilizer uptake.
4. The machine learning model has limitations when extrapolating outside the ranges of the dependent variables (or when interpolating). No further nutrient response is expected beyond the highest NPK rates applied. Predictions are likely poor when recommendations are needed in areas that are not well covered by the data, or for fertilizer application rates outside the ranges tested in the experiments.
5. Response curves have a stepwise eery shape that is not biologically meaningful. Yields without nutrient addition or low nutrient application tend to be overestimated, and yields at high nutrient application rate tend to be underestimated. Curves would need to be smoothened before applying optimisation algorithms to calculate fertilizer recommendations, but this will introduce additional error or bias.

STEP 5: Predict INS, IPS and IKS for target area

(The direct yield approach using machine learning explained under step 4b is not further pursued.)

Predicting INS, IPS and IKS for the target area of interest

As an alternative to above steps 3 and 4, the blup yield can be predicted directly by a random forest model, and yield effects relative to the reference treatment calculated using the predicted yields.

The advantage of this method is that it is simpler and easier to implement and automate.

```
log(INS) ~ [soil pars] + [dig elevation pars] + AEZ + refY
log(IPS) ~ …
log(IKS) ~ …
```

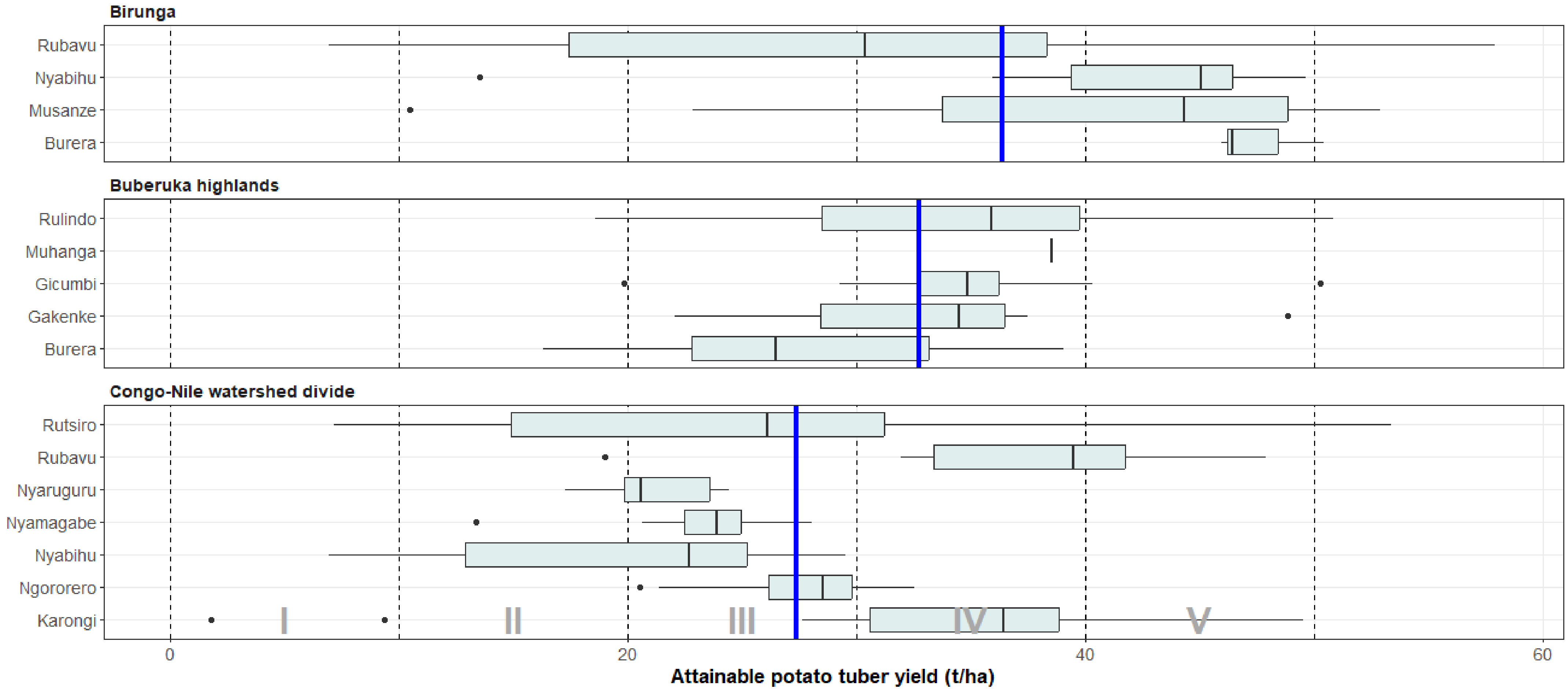
The prediction model does not include season_AB (heatplot analysis demonstrated that predictions are little affected by season_AB), and also does not consider district (since not all districts are well covered).

Predictions are plotted for each of the yield classes. The very low (0 – 10 t ha^{-1}) class is not included in the maps developed, as this class is less relevant since the attainable yield is too low to justify an investment in fertilizer. Most trials (conducted under optimal agronomic management) categorised most often in the medium, high or very high yield classes. Under real on-farm conditions with suboptimal crop management, however, these very low and low yield classes may still be relevant.

With the predicted N, P and K supply, the yield without fertilizer can be calculated assuming an attainable yield set 20% above the ceiling of the yield class.

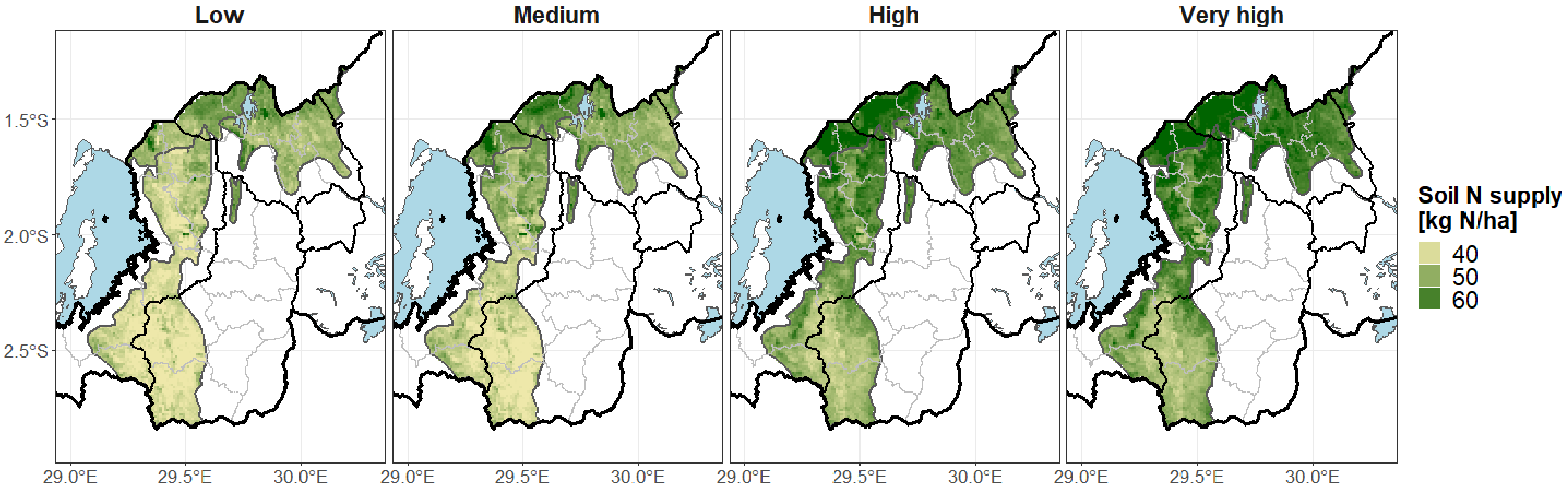
What is the most common reference yield level by AEZ?

Farmers are unlikely to be able to correctly judge the yield class of their field. Can we make assumptions based on the on-farm fertilizer experiments what yield class is most common?



The high yield class (IV) is most common in the Birunga and Buberuka highlands agro-ecological zones, while the medium yield class (III) is most common in the Congo-Nile watershed divide. In the Birunga AEZ, an important number of trials are in the very high yield class (V). Very few trials categorise in the very low (I) or low (II) yield classes; however, smallholder fields under suboptimal crop management may fall in these classes.

Predicting INS for the target area of interest

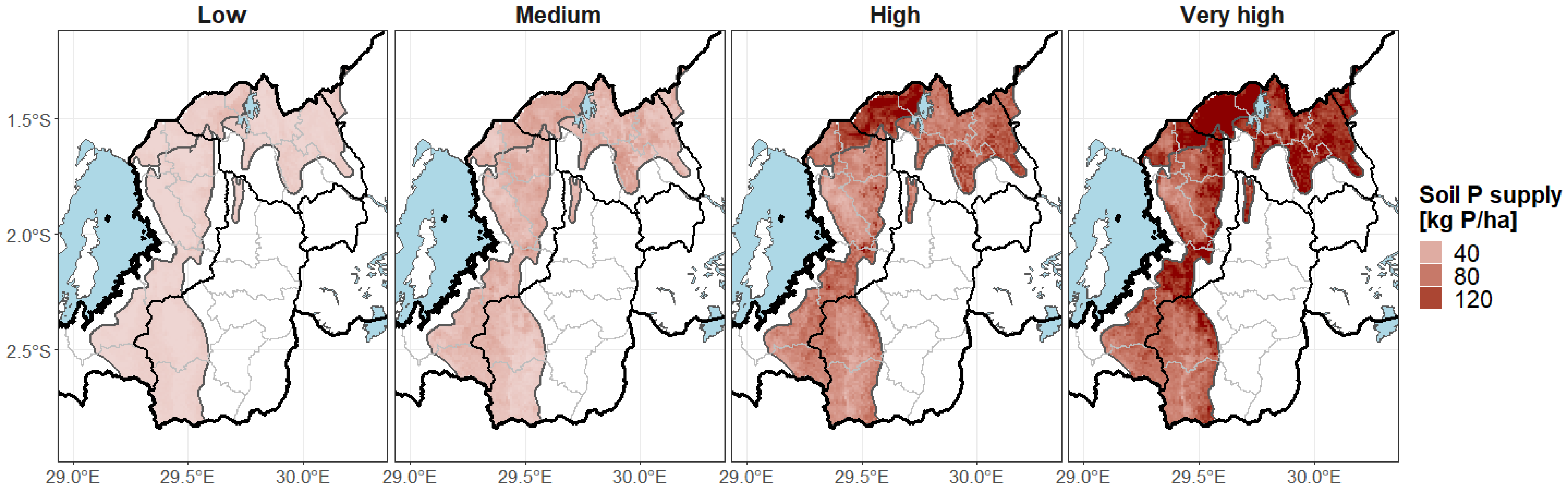


Median indigenous N supply (kg N ha^{-1})

AEZ	Low	Medium	High	Very high
Birunga	52	57	65	68
Buberuka highlands	49	50	56	61
Congo-Nile watershed divide	40	43	51	53

Note: N supply in the very low yield class is not included. Attainable yields in this yield class are too low to justify an investment in fertilizer.

Predicting IPS for the target area of interest

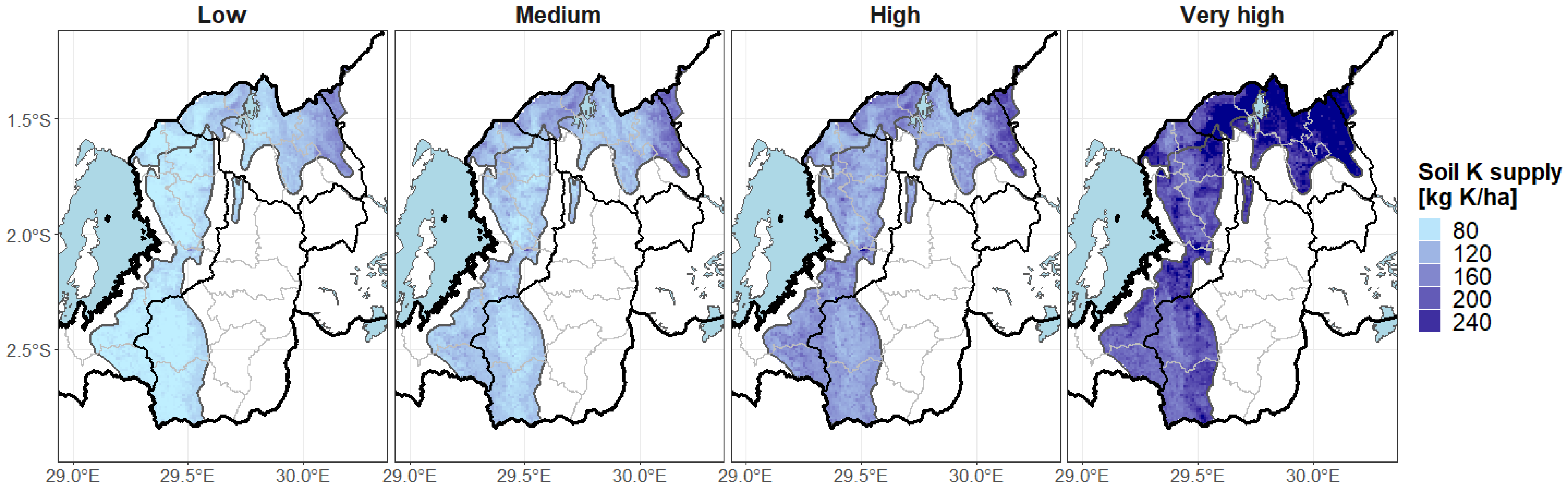


Median indigenous P supply (kg P ha^{-1})

AEZ	Low	Medium	High	Very high
Birunga	14	39	122	182
Buberuka highlands	12	33	88	137
Congo-Nile watershed divide	11	30	68	94

Note: P supply in the very low yield class is not included. Attainable yields in this yield class are too low to justify an investment in fertilizer.

Predicting IKS for the target area of interest



Median indigenous K supply (kg K ha^{-1})

AEZ	Low	Medium	High	Very high
Birunga	90	127	146	232
Buberuka highlands	107	117	136	274
Congo-Nile watershed divide	78	101	135	196

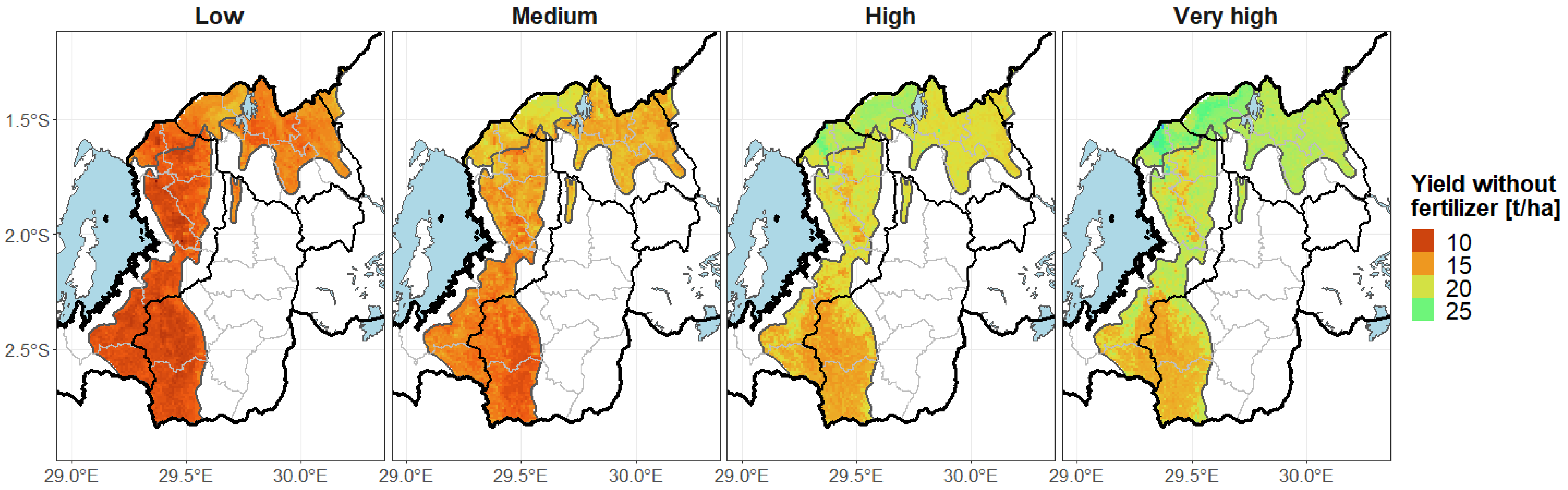
Note: K supply in the very low yield class is not included. Attainable yields in this yield class are too low to justify an investment in fertilizer.

Spatial variation in indigenous soil nutrient supply

%variation explained

AEZ	N	P	K
refY, AEZ and district	72%	79%	88%
refY and district	71%	77%	88%
refY and AEZ	57%	76%	84%
refY	28%	71%	79%
AEZ	29%	5%	5%

Predicted yields without fertilizer



Median predicted control potato tuber yields (t ha^{-1})

AEZ	Low	Medium	High	Very high
Birunga	13.8	18.3	21.8	23.8
Buberuka highlands	13.9	16.4	19.1	21.6
Congo-Nile watershed divide	11.1	13.9	17.8	18.7

STEP 6: Calculating fertilizer recommendations

Setting the objective of the fertilizer recommendations

The objective aims to improve recommendations over the current blanket recommendations, without adding complexity and ensuring compatibility with the national fertilizer subsidy scheme.

1. Predict the yield obtained with the current blanket recommendation (300 kg ha^{-1} of NPK 17:17:17) for each yield class

2. Identify available fertilizers to use: DAP and urea (in addition to NPK 17:17:17)

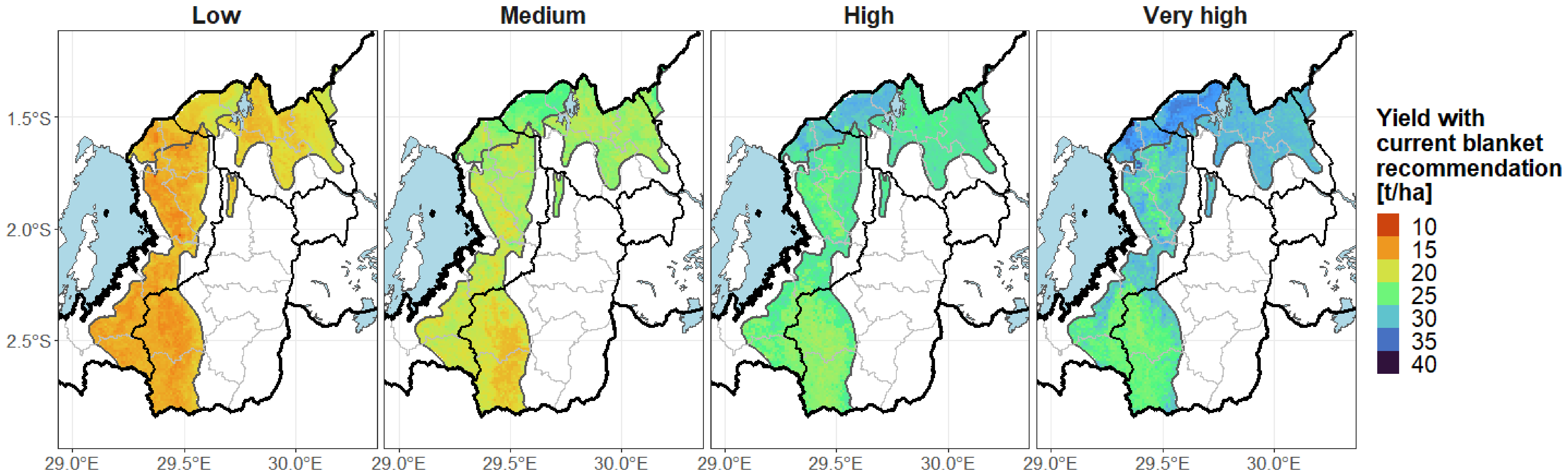
3. Develop scenarios with different yield targets:
   - Achieve the yield expected with the current blanket recommendation
   - Achieve a yield increase of 10% or 20% above the yield with the current blanket recommendation

4. Define the optimisation algorithm. Ideally, this is done by minimising the total cost of the fertilizer recommendation, i.e., identify the combination of DAP, urea and NPK 17:17:17 rates that achieves the target yield at the lowest cost. However, since prices do not differ substantially between the 3 fertilizer types, and because fertilizer prices change over time and a recommendation is needed that remains valid for many years, we opted to minimise the total quantity of fertilizer to apply.

5. Calculating the fertilizer requirements for each gridcell of 1 x 1 km would be computationally intensive. To reduce the computation time, a k-means clustering is conducted. For each combination of AEZ (3) and refY (5) 100 clusters of predicted N, P and K supply are identified, which capture 97-100% of the variation in supply. This reduces the number of gridcells for which to calculate recommendations from 26,390 to 1,500.

6. Use the optimisation algorithm to calculate the fertilizer requirements to achieve the yield target in the various scenarios for each of the 1,500 clusters. Then consider simplifications by calculating medians for each AEZ x refY combination, or by creating fertilizer packages using k-means clustering.

Predicted yields with current blanket recommendation (300 kg ha^{-1} NPK 17:17:17)



Median predicted potato tuber yields with current blanket recommendation (t ha^{-1})

AEZ	Low	Medium	High	Very high
Birunga	18.0	24.5	29.9	32.7
Buberuka highlands	18.5	22.8	27.3	30.5
Congo-Nile watershed divide	16.2	20.5	25.4	27.2

Median fertilizer requirements by AEZ and refY level

Median fertilizer requirements (kg ha^{-1}) can be calculated for each AEZ and refY level combination, and for 3 yield targets, defined relative to the yield obtained with the current blanket recommendation of 300 kg ha^{-1} NPK 17:17:17

AEZ	refY	Same yield as current recommendation			10% above current recommendation			20% above current recommendation		
		DAP	NPK 17:17:17	Urea	DAP	NPK 17:17:17	Urea	DAP	NPK 17:17:17	Urea
Birunga	Low	71	81	106	133	159	202	195	246	317
	Medium	45	48	95	91	87	132	154	148	245
	High	50	59	78	78	71	111	100	96	145
	Very high	36	47	85	66	75	106	75	96	137
Buberuka highlands	Low	68	71	102	130	138	202	219	257	317
	Medium	47	49	94	79	79	128	118	114	189
	High	46	49	85	78	68	111	98	84	145
	Very high	58	57	69	64	71	103	74	92	132
Congo-Nile watershed divide	Low	71	78	98	108	126	166	157	192	259
	Medium	50	54	95	86	75	128	112	99	170
	High	40	43	89	68	70	112	80	92	147
	Very high	35	49	83	64	71	103	74	92	132

Yellow highlighted cells indicate the most common yield level for each of the 3 AEZs.
The analysis highlights that the same yield level as obtained with the current blanket recommendation can be obtained with less fertilizer, by reducing the NPK 17:17:17 rate from 6 bags to roughly 1 bag per hectare and applying 1 bag of DAP and 2 bags of urea per hectare. Yield can be increased by 20% above the yield with the current blanket recommendation by applying 2 bags of NPK17:17:17, 2 bags of DAP and 3 bags of urea per hectare.

Fertilizer requirements for two selected scenarios

Fertilizer requirements to achieve the same yield or 20% more yield as compared to the yield obtained with the current blanket recommendation.
We assume that the refY class is high for the Birunga and Buberuka highlands AEZ, and medium for the Congo-Nile watershed divide. This implies an attainable yield of 48 t ha^{-1} in the former two AEZs, and 36 t ha^{-1} in the latter AEZ.

Simplify to bags per hectare (1 bag = 50 kg)

AEZ	refY	Same yield as current recommendation (kg ha^{-1})			Same yield as current recommendation (bags)		
		DAP	NPK 17:17:17	Urea	DAP	NPK 17:17:17	Urea
Birunga	High	50	59	78	1	1	2
Buberuka highlands	High	46	49	85	1	1	2
Congo-Nile watershed divide	Medium	50	54	95	1	1	2

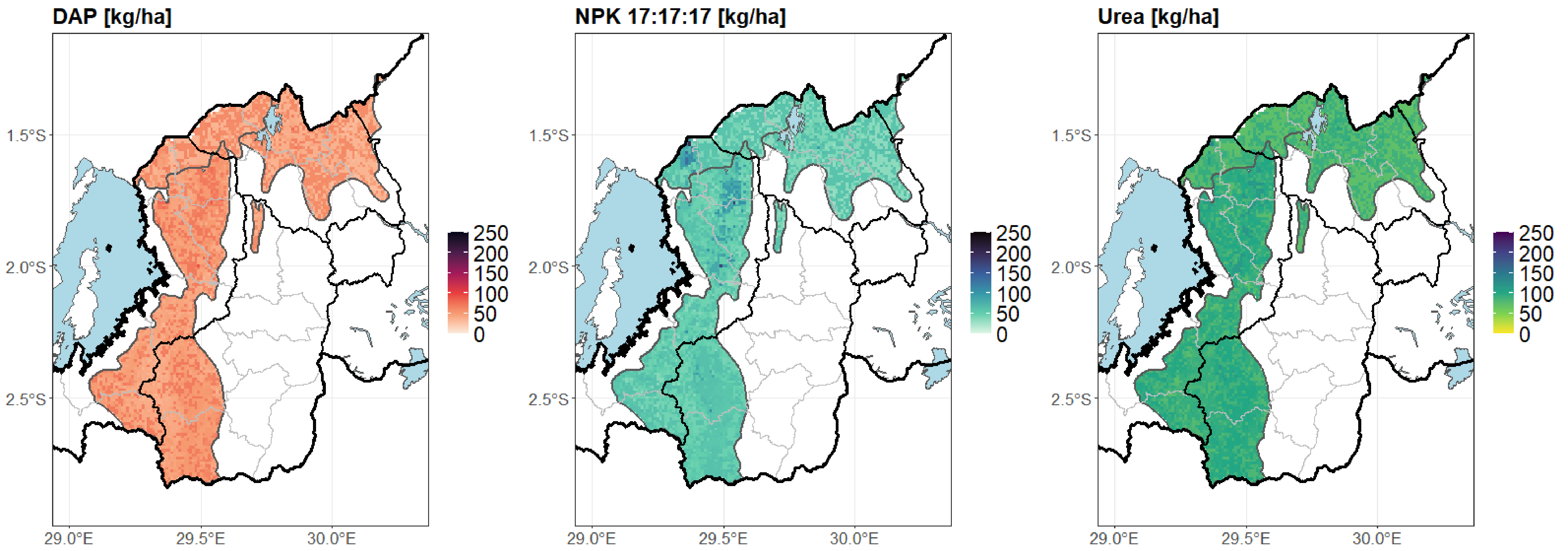
AEZ	refY	20% above current recommendation (kg ha^{-1})			20% above current recommendation (bags)		
		DAP	NPK 17:17:17	Urea	DAP	NPK 17:17:17	Urea
Birunga	High	100	96	145	2	2	3
Buberuka highlands	High	98	84	145	2	2	3
Congo-Nile watershed divide	Medium	112	99	170	2	2	3

Fertilizer requirements for two selected scenarios

SCENARIO 1: SAME YIELD

Fertilizer requirements to achieve the same yield as compared to the yield obtained with the blanket recommendation.

We assume that the refY class is high for the Birunga and Buberuka highlands AEZ, and medium for the Congo-Nile watershed divide. This implies an attainable yield of 48 t ha^{-1} in the former two AEZs, and 36 t ha^{-1} in the latter AEZ.

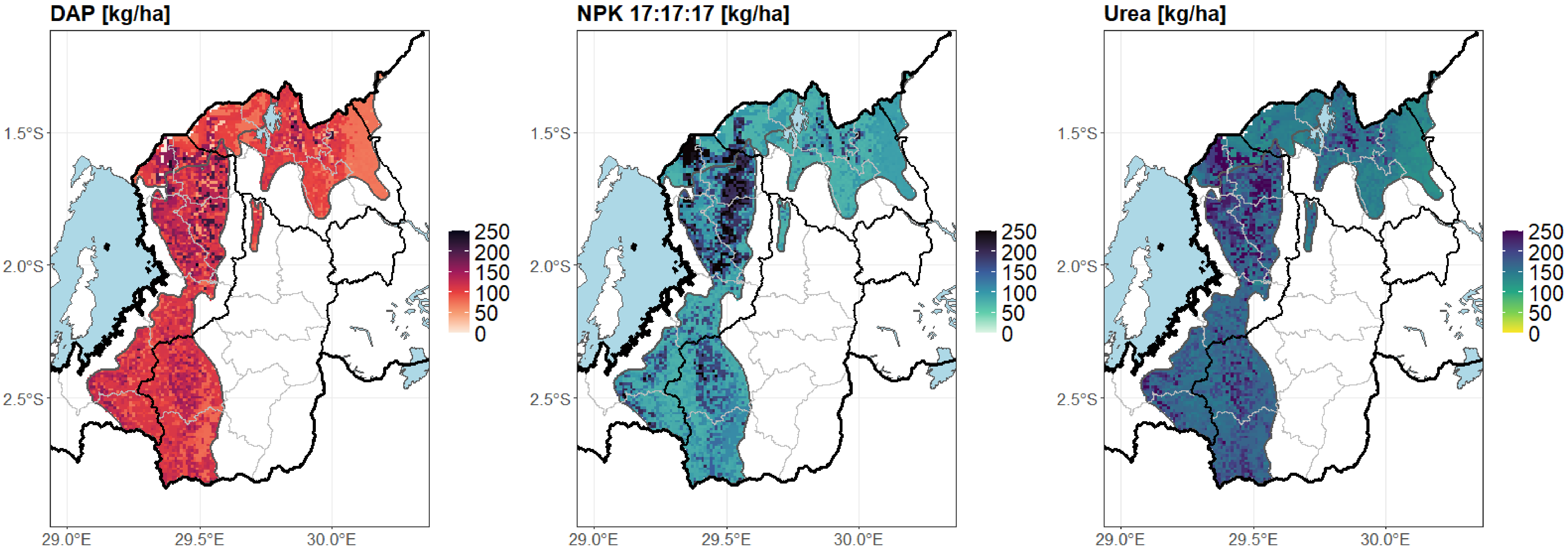


Fertilizer requirements for two selected scenarios

SCENARIO 2: 20% YIELD INCREASE

Fertilizer requirements to achieve 20% more yield as compared to the yield obtained with the blanket recommendation.

We assume that the refY class is high for the Birunga and Buberuka highlands AEZ, and medium for the Congo-Nile watershed divide. This implies an attainable yield of 48 t ha^{-1} in the former two AEZs, and 36 t ha^{-1} in the latter AEZ.



STEP 7: Suggestions for field validation exercises

How can we validate these recommendations?

We suggest comparing 4 fertilizer treatments side-by-side in farmers' fields

1. The current blanket recommendation of 6 bags (300 kg) of NPK 17:17:17 per hectare
   This treatment serves as the reference. All other treatments are aiming to provide a cost-saving or a yield improvement over this current recommendation with a higher RoI.

2. An improved blanket recommendation of 2 bags of DAP, 2 bags of NPK 17:17:17 and 3 bags of urea per hectare
   This treatment aims to achieve a 20% increase in yield relative to the blanket recommendation, with only a small increase in total fertilizer quantity and cost, resulting in a higher RoI.

3. A site-specific recommendation with varying quantities of DAP, NPK 17:17:17 and urea per hectare
   This treatment does not affect the total fertilizer requirement across the target intervention area, but varies the allocation based on modelled variation in soil supply, targeting higher rates to more responsive soils with lower nutrient supply, and addressing specific nutrient deficiencies by varying the rates of the individual fertilizers. This is assumed to result in an overall higher agronomic efficiency, and thus a higher RoI.

4. A reduced blanket recommendation of 1 bag of DAP, 1 bag of NPK 17:17:17 and 2 bags of urea per hectare
   This treatment aims to achieve the same yield as the current blanket recommendation at a lower total fertilizer requirement. As such, we expect the same yield (same gross returns) at a lower cost, hence a higher RoI.

Note: if only 3 fertilizer treatments can be practically evaluated, we suggest omitting the 4th option, as potato farmers are likely targeting increased yields and higher fertilizer application, rather than seeking to reduce costs and decreasing fertilizer investments.

How can we validate these recommendations?

Suggested treatment structure for the validation exercises:

4 plot design

1
Current blanket recommendation
(black plot)

2
Improved blanket recommendation
(white plot)

3
Site-specific recommendation

4
Reduced blanket recommendation
(grey plot)

3 plot design

1
Current blanket recommendation
(black plot)

2
Improved blanket recommendation
(white plot)

3
Site-specific recommendation

Note: if only 3 fertilizer treatments can be practically evaluated, we suggest omitting the 4th option, as potato farmers are likely targeting increased yields and higher fertilizer application, rather than seeking to reduce costs and decreasing fertilizer investments.

How can we practically evaluate site-specific recommendations?

We reduce site-specific recommendations to 6 fertilizer packages

The earlier analysis showed that spatial variation in soil nutrient supply resulted in important variation in fertilizer requirements to achieve the 20% yield improvement.

We earlier on reduced variation in soil nutrient supply using k-means clustering, and allowing for 100 clusters per combination of AEZ and refY level. This however implies that we however still have 300 different fertilizer rates for the target intervention area. This is impractical to test.

Similar as earlier in the process, we try and capture the variation by clustering and defining a manageable number of clusters (fertilizer packages). The maximum number of clusters to consider is 6.

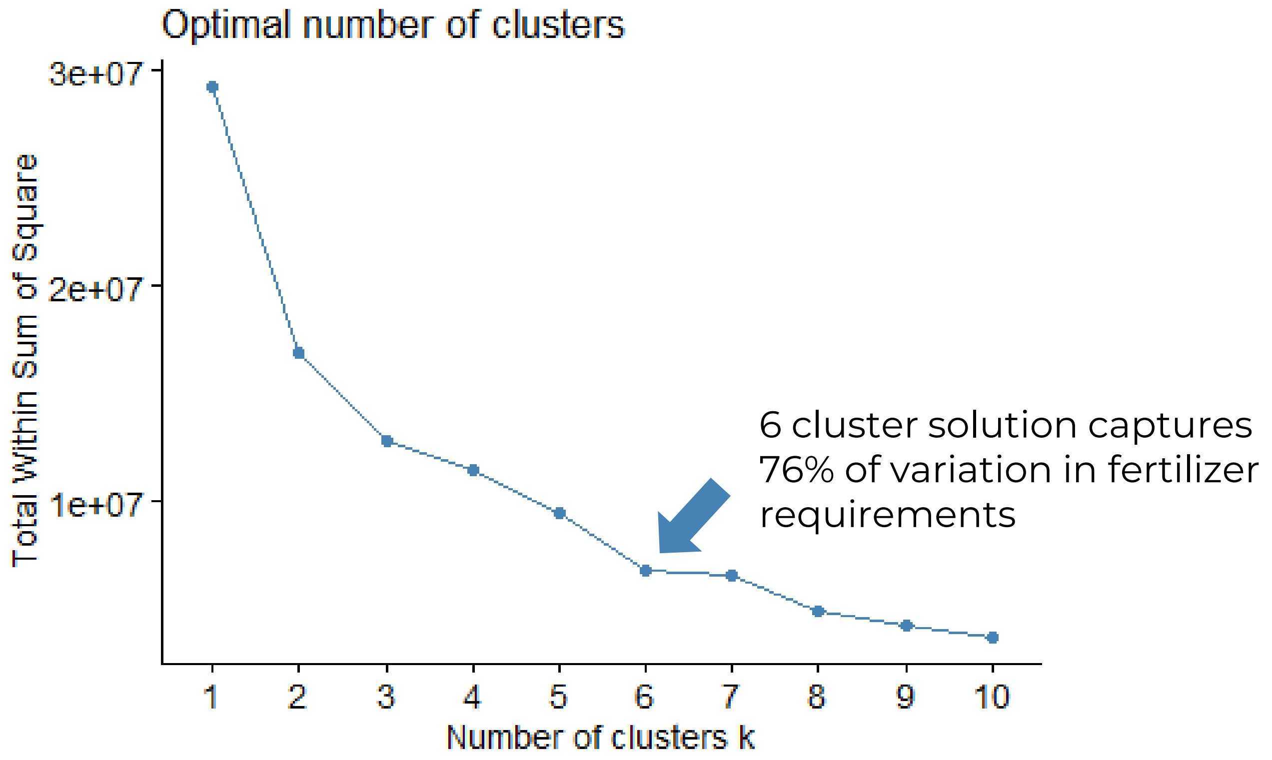
Instead of k-means clustering, we will employ k-medoids clustering, which is a more robust procedure but based on the same principle, and useful here as distributions of fertilizer requirements are mostly right-skewed. Our aim is to provide site-specific fertilizer advice that does not increase the total fertilizer requirement across the 3 agro-ecological zones, but to achieve a higher agronomic efficiency by allocating different quantities of the 3 fertilizer types according to the soil nutrient supply.

Site-specific recommendations: 6 fertilizer packages using k-medoid clustering

We reduce site-specific recommendations to 6 fertilizer packages

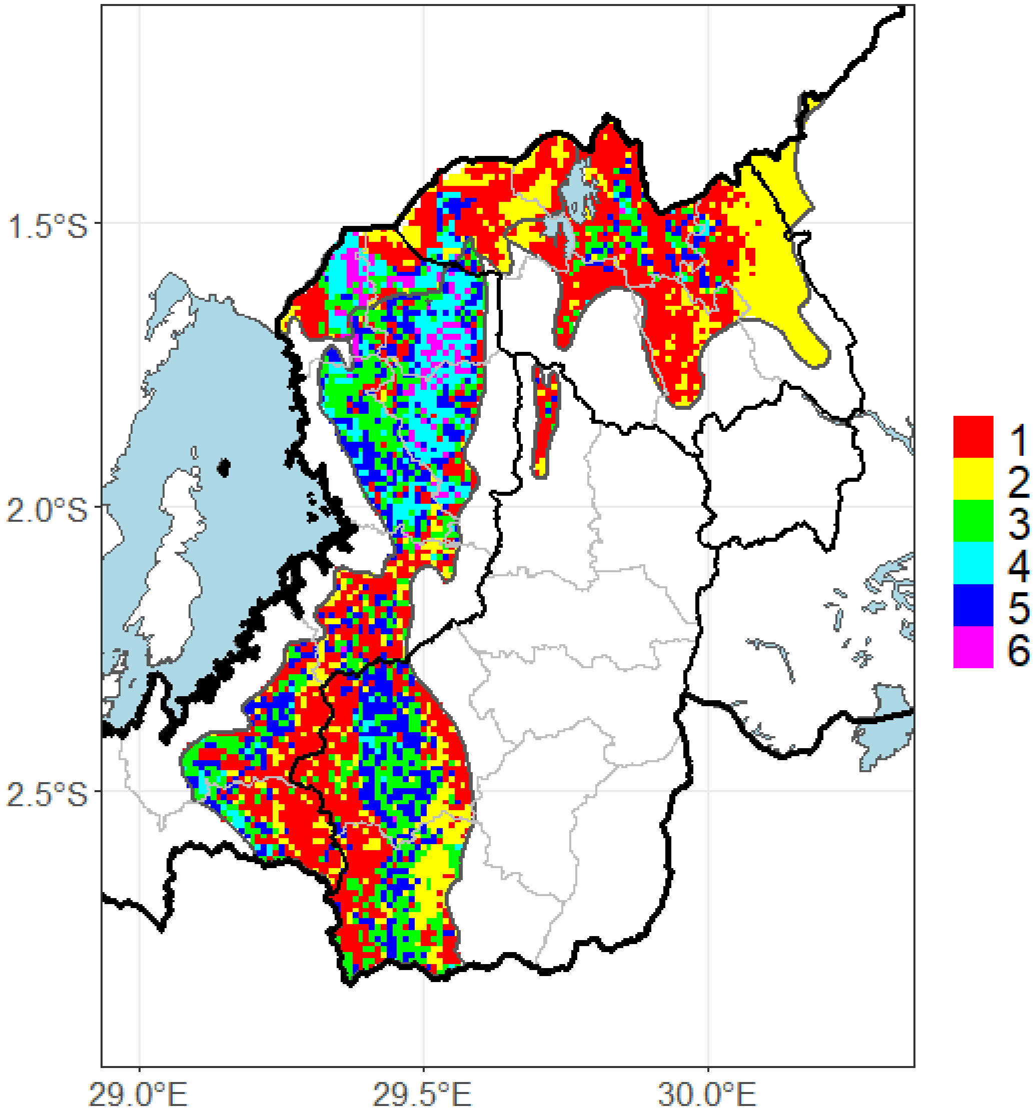
Six fertilizer packages can be formulated that capture 76% of the variation in fertilizer requirements to achieve 20% yield increase. These will be colour-coded (red – yellow – green – cyan – blue – purple) for ease of implementation.

	DAP	NPK171717	Urea
1	109	77	159
2	74	98	141
3	131	93	212
4	131	236	175
5	100	159	166
6	134	197	336



Site-specific recommendations: 6 fertilizer packages using k-medoid clustering

We can plot which fertilizer package is recommended in each gridcell...



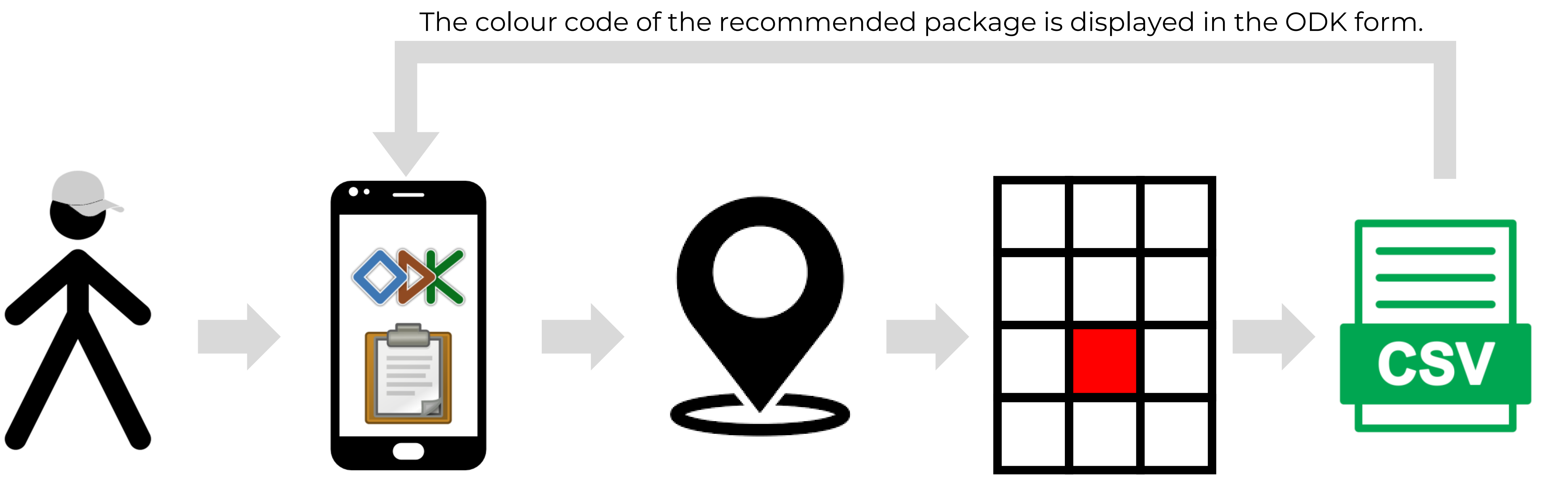
1	Red	Low N, medium P, low K
2	Yellow	Low N, low P, low K
3	Green	Medium N, medium P, low K
4	Cyan	Medium N, high P, high K
5	Blue	Medium N, medium P, medium K
6	Purple	High N, high P, high K

Fertilizer and nutrient application rates for the fertilizer treatments

		DAP rate kg/ha	NPK 17:17:17 rate kg/ha	urea rate kg/ha	N rate kg/ha	P rate kg/ha	K rate kg/ha
1 RED		109	77	159	106	28	11
2 YELLOW		75	98	141	95	22	14
3 GREEN		131	93	212	137	33	13
4 CYAN		131	236	175	144	44	33
5 BLUE		100	159	166	121	32	22
6 PURPLE		134	197	336	212	42	28
BLACK		0	300	0	51	22	42
GREY		50	50	100	64	14	7
WHITE		100	100	150	104	28	14

Site-specific recommendations: practical implementation

An ODK form can indicate which fertilizer package to test based on GPS location.



Extension agent visits farmer participating in validation exercise.

An ODK form is used to register the farmer and his/her potato field.

The GPS location of the centre of the farmer's field is recorded.

The GPS location is rounded to identify the relevant grid cell.

The recommended package is obtained from a lookup table.

Site-specific recommendations: practical implementation

An ODK form can indicate which fertilizer package to test based on GPS location.

lookup_key	pkgNr	pkgCode	DAP	NPK	Urea
E29.72N-1.47	2	yellow	74	98	141
E30.01N-1.5	2	yellow	74	98	141
E30.1N-1.55	2	yellow	74	98	141
E29.94N-1.76	1	red	109	77	159
E29.58N-1.71	4	cyan	131	236	175
E29.52N-2.04	2	yellow	74	98	141
E29.5N-2.08	1	red	109	77	159
E29.31N-2.47	1	red	109	77	159
E29.47N-1.76	3	green	131	93	212
E29.39N-2.47	5	blue	100	159	166
E29.44N-2.47	3	green	131	93	212
E29.58N-1.64	5	blue	100	159	166
E29.95N-1.62	1	red	109	77	159
E29.86N-1.4	2	yellow	74	98	141
E29.51N-2.47	5	blue	100	159	166
E29.37N-1.67	6	purple	134	197	336
…	…	…	…	…	…

5278 rows

Download the csv media file with the recommendations here:

RwaSIS_fertilizer_packages.csv

Economic analysis

Profitability will be calculated relative to the current blanket recommendation (CBR). The validation exercise should aim to quantify the yield with the current blanket recommendation and the improved recommendation (and their distribution) and can then be used to evaluate the change in returns-on-investment (and its distribution).

Change in gross returns for improved blanket recommendation (IBR)

$GR_{IBR} = (TY_{IBR} – TY_0) * TP$ Gross returns of the improved blanket recommendation [FRW/ha]
$GR_{CBR} = (TY_{CBR} – TY_0) * TP$ Gross returns of the current blanket recommendation [FRW/ha]
$dGR_{IBR} = GR_{IBR} – GR_{CBR} = (TY_{IBR} – TY_{CBR}) * TP$ Difference in gross returns between the improved and current recommendation [FRW/ha]

TY = tuber yield [t/ha]; TY_0 = tuber yield without fertilizer (not observed) [t/ha]; TP = tuber price [FRW/t]

Change in cost for improved blanket recommendation

$TC_{IBR} = 2 * FP_{DAP} + 2 * FP_{NPK17:17:17} + 3 * FP_{urea}$ Cost of the improved blanket recommendation [FRW/ha]
$TC_{CBR} = 6 * FP_{NPK17:17:17}$ Cost of the current blanket recommendation [FRW/ha]
$dTC_{IBR} = TC_{IBR} – TC_{CBR} = 2 * FP_{DAP} – 4 * FP_{NPK17:17:17} + 3 * FP_{urea}$ Difference in cost between the improved and current recommendation [FRW/ha]

TY = FP = Price of 50 kg bag of fertilizer [FRW/(50 kg)]

Relative return-on-investment (RoI) when changing from current to improved blanket recommendation

$dNR_{IBR} = dGR_{IBR} – dTC_{IBR}$ Net returns (profit) from the additional investment in fertilizer relative to the current blanket recommendation [FRW/ha]
$RoI = dNR_{IBR} / dTC_{IBR}$ Return on investment for the additional investment in fertilizer [FRW/ha]

For example: TY_{IBR} = 36 t/ha, TY_{CBR} = 30 t/ha, TP = 100,000 FRW/t, FP_{DAP} = 55,000 FRW/bag; $FP_{NPK17:17:17}$ = 45,000 FRW/bag; FP_{urea} = 40,000 FRW/bag
dGR_{IBR} = (36 – 30) * 100,000 = 600,000 FRW/ha
dTC_{IBR} = 2 * 55,000 – 4 * 45,000 + 3 * 40,000 = 50,000 FRW/ha
dNR_{IBR} = 600,000 – 50,000 = 550,000 FRW/ha
RoI = 550,000 / 50,000 = 11 (meaning that every 1 FRW extra invested, you obtain the 1 FRW back and an additional 11 FRW net profit).