ANNEX D. BENCHMARKING ANALYSIS

Index

D.1. Introduction

Different computation Lines are implemented in the Navigator tool to provide users (mainly farmers, but other actors might be interested) with recommendations on N, P, and K fertilization rates (N_{rate}, P_{rate}, and K_{rate} respectively) to achieve a target yield together with the environmental impact of their application in terms of fertilizer losses (F lines), gas emissions (G lines) and costs (E lines). The Lines aim to cover a wide range of crops covering the conditions of European agriculture and therefore have to represent a wide range of soils, climates and farming systems. The different Lines are mainly differentiated by their complexity and the number of inputs required. The most complex methodology (Line 1), which requires detailed inputs, is expected to lead to more realistic recommendations.

This part of the study is focused on the assessment of the fertilization recommendation (lines F1 – F4). E lines and G lines are assessed separately. The aim of the benchmarking is to define the accuracy of the F lines included in the Navigator tool in relation to the N, P, and K fertilization recommendation (N_{rate}, P_{rate} and K_{rate} respectively) and the N_{rate} balance components considered in each F line, i.e. the source of the N_{rate} balance components: N mineralized ($N_{mineralization}$), and N incorporated in the soil with the irrigation water (N_{water}), and the sinks of the N_{rate} balance components: N crop requirements (N_{uptake}), N leached in the soil ($N_{leaching}$), and N denitrified ($N_{denitrification}$) and N volatilized ($N_{volatilization}$) from the fertilizers. It is determined by the assessment of the formulation ability to simulate the nutrients dynamics in the continuum soil-crop-atmosphere ($\mathcal{E}_{Line}$) and the assessment of the impact of the inputs accuracy on the simulations ($\mathcal{E}_{inputs}$). The knowledge of the $\mathcal{E}_{inputs}$ is especially important from a user point of view to which the data required to implement a given F line are not always available or might be difficult to obtain. In such situations the user can implement a F line with a combination of observations obtained by field measurements, alternative determination methods or take profit of the default values proposed by the Navigator tool. In those cases, special caution should be taken if the input is a sensitive variable for the F line. Implementation profiles are derived for each F line from a sensitivity analysis in order to guide users on those inputs on which they should make a greater effort to obtain an accurate value.

The assessment of the F lines ($\mathcal{E}_{Line}$) to the different European agriculture conditions require representative information about the crop development and growth and their growing conditions. Due to the time frame and resources of the project, different existing datasets about the main crops from commercial fields and managed under different cropping system are used. In this work, those real cases are refered as Assement Scenarios (AS) that are grouped by Farming Context (FC). An Assesment scenario is described as a crop that grows in a geolocated field under a determined pedoclimatic conditions and farming practices and content as far as possible information to implement the 4 F lines. A Farming Context is defined as a group of assessment scenarios located on an agroclimatic zone (AEZ), and a farming area, under a cropping system (water and soil management), and with an associated management strategy and target yield.

The use of existing datasets arises to some specific difficulties, since they are not designed to test the recommendation of the Navigator and present some heterogeneity in terms of frequency of observations, methods of determination or fully description. In order to homogenise the observations according to the needs of each F line, adjustment methods are applied on the available data to estimate gaps. On the other hand, some caution should be taken in the use of existing real cases as reference. Their use is based on the assumption that the yield obtained is the maximum for the fertiliser applied. However, the observed yield could differ from the farmer's target yield due to crop growth could be affected by different biotic and abiotic factors such as drought or heat waves, pests or diseases, problems in crop installation, etc. In this work, it was only considered that the observed yield could be affected by climatic factors as we do not have the means to adequately address a posteriori other sources of variation. It is further assumed that a crop model such as STICS (Brisson, 1998) is able to represent inter-annual yield variability due to climate and correct the yield.

For the performance of the benchmarking, the next subobjectives are defined and organized in a workflow as presented in Figure1:

i. Definition of implementation profiles by F line and farming context.
ii. Impact of the inputs accuracy for the fertilization recommendation ($\mathcal{E}_{inputs}$).
iii. Delineation of the climatic effect on the observed yield to stablish the target yield for F lines implementation.
iv. Assessment of the F lines computation methodologies ($\mathcal{E}_{Line}$) and the coherence between the 4 F lines.
v. The accomplish of the previous results for the definition of the quality indicators (traffic lights).

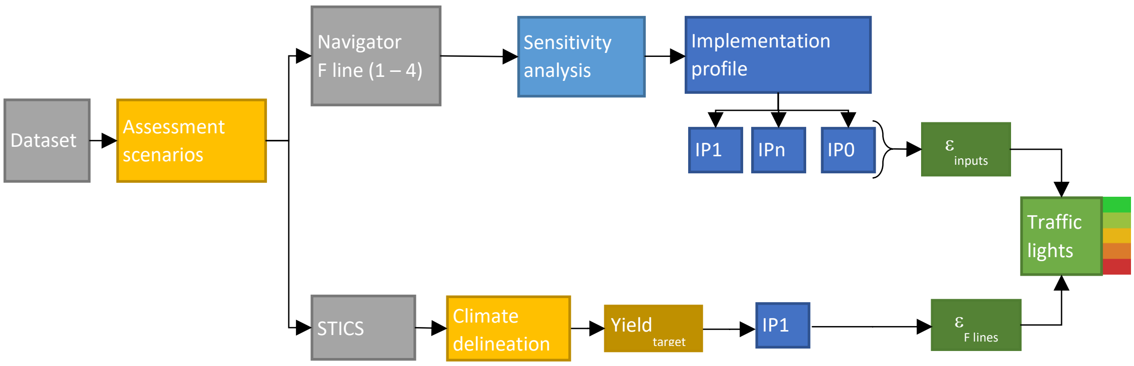


Figure 1. Scheme of the workflow defined for the benchmarking over the 4 F lines of the Navigator tool.

D.2. Materials and methods

2.1. Dataset

A total of 100 assessment scenarios are collected from two agroclimatic zones (AEZ), mediterranean and continental, that are grouped in 9 farming contexts (FC) defined based on soil, water and fertilization management, i.e. irrigated or rainfed, conventional (tillage), conservation (no tillage) and organic (organic manure applications) (Figure 2 and Table 1).

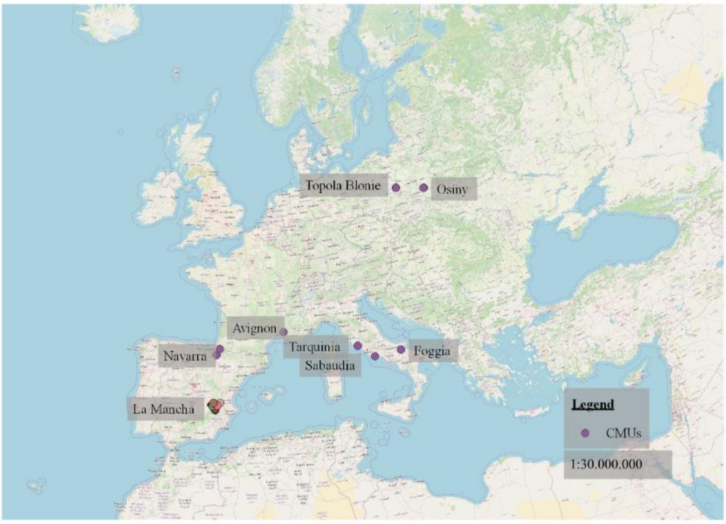


Figure 2. Distribution of the Pilot areas included within the Database and in the benchmarking performance approaches.

The information collected for each assessment scenario is related to as far as possible the information required to implement the 4 F lines:

- Description of the development and growth of crops between sowing to maturity.
- Technical information about the farming practices related to the crop management, i.e. sowing, soil, water and fertilization management.
- Description of the soils (permanent, hydrological, and chemical properties).
- 20-years climatic series of daily data from the closest weather station to every assessment scenario.

Among the crops, wheat (*Triticum*) data was collected in all the farming context. They are related to wheat crops (spring, winter and durum wheat) grown along the farming context included on this study. Noted that, the data collected belong to different projects or studies with different objectives and therefore present a certain heterogeneity in terms of frequency of observations, methods of determination or fully description of the assessment scenarios. Therefore, different assumptions have been taken in order to maximise the use of the information collected that is especified along this document.

Table 1. Summary of the assessment scenarios (AS) used to perform the sensitivity analysis relative to wheat crop varieties grouped by farming context (FC). FC are defined by the location of the Pilot site (i.e. agroclimatic zone, AEZ), the cropping system (water and soil management), the associated management strategies and yield. The range of observation years and the number of AS within each Pilot site are included.

Crops	FC	AEZ	lat.	lon.	h.	Tm (°C)	Pa (mm)	Pw (mm)	Management		Years	AS (number)
									Water	Soil		
Spring wheat	LM-irri-til	Southern	39.0	1.9	677	14 ± 7	348 ± 81	199 ± 73	irri	Till	2015 – 2017	36
	LM-irri-ntil								irri	No-till	2015 – 2017	45
	LM-rfed-til								rfed	Till	2016 – 2018	15
	LM-rfed-ntil								rfed	No-till	2016 – 2018	12
	FR-rfed-til		43.9	4.9	150	15 ± 7	652 ± 196	359 ±106	rfed	Till	2016 – 2017	12
	TOPs-rfed-til	Continental	52.1	15.1	94	9 ± 9	585 ± 121	235 ±58	rfed	Till	2019	1
Winter wheat	NAV-rfed-til	Southern	42.5	-1.65	447	12 ± 7	555 ± 214	322 ±101	rfed	Till	2019	1
	TOPw-rfed-til	Continental	52.1	15.1	94	9 ± 9	585 ± 121	235 ±58	rfed	Till	2018 – 2020	10
Durum wheat	FOG-rfed-til	Southern	42.5	13.1	82	17 ± 8	555 ± 91	338 ±84	rfed	No-till	2016 – 2019	6
	TAR-rfed-til		42.3	1.4	40	16 ± 6	793 ± 187	564 ±170	rfed	Org	2016 – 2017	2

AEZ: agroclimatic zone; **FC** (farming contexts) : La Mancha-Spain (LM), France (FR), Topola-Blonie-Poland (PO), Navarra-Spain (NAV), Foggia-Italy (FOG), and Farquinia-Italy (TAR).
Water management : irrigation (irri) or rainfed (rfed) farming systems.
Soil management : tillage (Till), non tillage (No-till), and incorporation of organic manure (Org).
Pa: arithmetic mean ± S D of the cumulative rainfal during the year.
Pw: arithmetic mean ± SD of the cumulative rainfal during the Winter period considered between 1st October and 31th March.

2.2. Inputs preparation for F lines implementation

In this section, a description of the adaptation of the collected data to implement the 4 F lines is presented in order to deal with the heterogeneity of the dataset in terms of frequency of observations, methods of determination or fully description of the assessment scenarios of the dataset.

In relation to the data that describe the development and growth of crops during the season and at maturity a pretreatment of the data was performed to:

i) identification of the climatic impact on the yield observed to stablish the target yield for F lines assessment (section 2.2.1.),
ii) determination of the typical NDVI curve for every farming context required for F1 implementation (section 2.2.2.).

In relation to the data that describe the soil growing conditions the following assumptions was taken:

i) texture is derived from clay, silt and sand fraction measurements using the USDA classification,
ii) the initial Nitrogen mineral content (Nmin_initial) was derived from the first measurements as organic matter mineralization is expected to be low during cold season, and thus can be ignored,
iii) for permanent properties as texture, SOM, pH, …, an average of all observations made on a given field accross years is used.

2.2.1. Climate impact on observed yield

The correction of observed yield is a key step in the benchmarking for the application of the main hypothesis stablish for F lines assessment and the use of existing datasets: the yield obtained is the maximum for the fertilizer applied (see section 2.6). Thus, fertilizer applied on field (dataset) are considered as a reference for the assessment of the outputs of the F lines ($\mathcal{E}_{Line}$). Nevertheless, the observed yield (dataset) may differ from the farmer's target yield as crop growth could be affected by different biotic and abiotic factors such as drought or heat waves, pests or diseases, problems in crop installation, etc. In this work, it was only considered that the observed yield could be affected by climatic factors as we do not have the means to adequately address a posteriori other sources of variation, and it is further assumed that a crop model such as STICS (Brisson, 1998) is able to represent inter-annual yield variability due to climate. For this, the yield was simulated in STICS on a 20-year historical climate series (2000 – 2020). The optimum yield was determined by considering the lowest boundary of the 8th decile of the simulated yield decile ($yield_{P80}$) (in 20% of the year the yield is equal or greater than the target yield). This threshold is similar to the maximum total yield obtained when climatic conditions are not a constraint (Villalobos, 2020). In Figure 3, examples of this approach are shown.

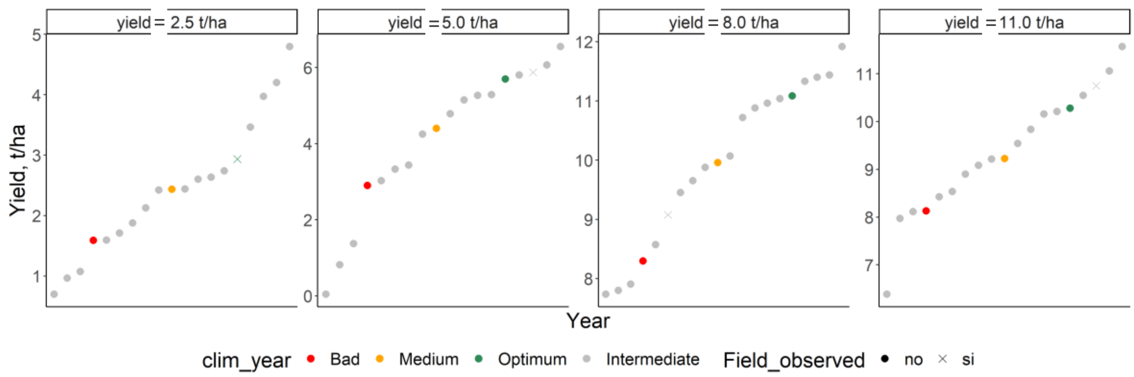


Figure 3. Examples of desfavorable (*yield*$_{P20}$), medium (*yield*$_{P50}$) and optimum (*yield*$_{P80}$) yield determination based on STICS simulations for four assessment scenarios over a 20-year climatic time series.

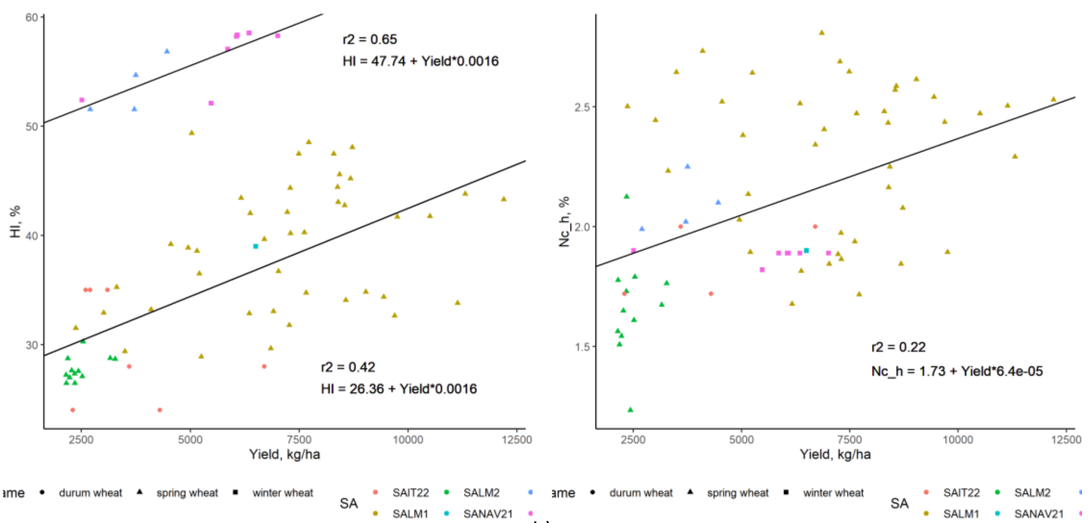
The target yield (*yield*$_{target}$, *Eq. 1*) is determined by the correction of the observed yield (*yield*$_{i\text{-}obs}$) by the difference between the simulated yield by STICS the year of observation (*yield*$_{i\text{-}sim}$) and the yield simulated by STICS leading to the 8th decile (Δ, *Eq. 2*).

$$yield_{target} = Yield_{i-obs} \pm \Delta \qquad (Eq.1)$$

$$\Delta = yield_{P80} - Yield_{i-sim} \qquad (Eq.2)$$

In addition, the years linked to the 2th decile (*yield*$_{P20}$) and 5th decile (*yield*$_{P50}$) were considered in this study to maximize the range of target yield tested for every assessment scenario on the determination of ε_{inputs}. The range between *yield*$_{P20}$ and *yield*$_{P80}$ by assessment scenario were considered (see section 2.5).

In agreement with the correction of the target yield, the harvest index (HI) and the nitrogen and phosphorous concentration in the harvested organs (Nc_h and Pc_h respectively) are corrected according to the correlation observed in the dataset between them and the yield, as shown in Figure 4. Kc_h is no updated due to no correlation was observed.



a) b)

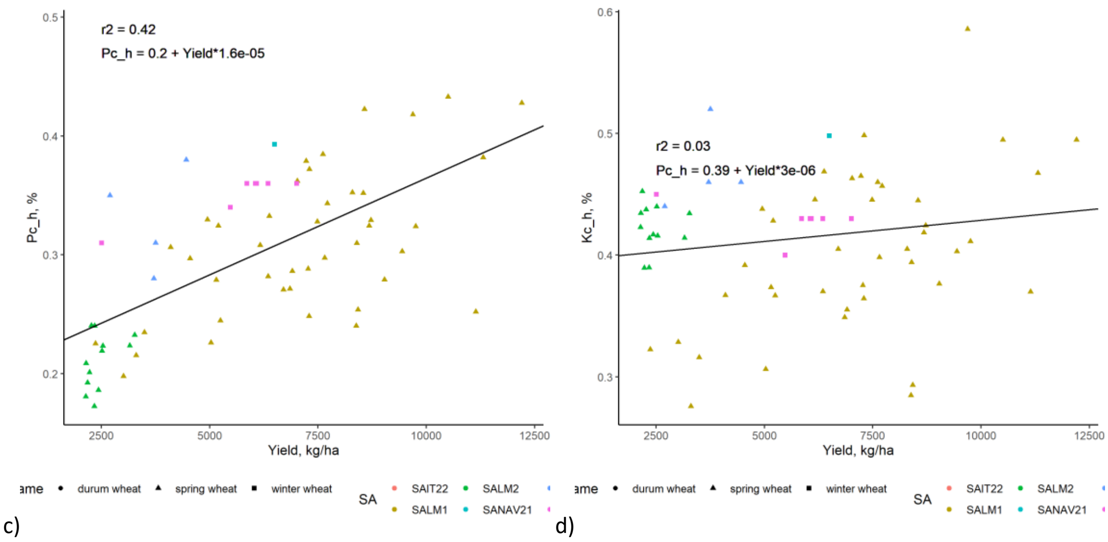


Figure 4: Correlation between a) yield and HI, b) yield and Nc_h, c) yield and Pc_h, and d) yield and Kc_h.

2.2.2. Typical curve of NDVI by farming context

The typical NDVI curve is an input required by F1. The typical NDVI curve is defined as the standard NDVI curve for a crop variety grown under a water management (irrigation or rainfed) in a region in Julian days. The typical curves are derived by the average of the daily NDVI observations in growing degree-days (gdd) of the assessment scenarios of a farming context. Finally, the final typical curve is generated by the inversion of gdd to Julian days.

The daily NDVI curves were obtained from the fit of the NDVI observations of every assessment scenario by the MODLAI model (Baret, 1986). MODLAI was developed to be applied on LAI observations and is a combination of a logistic and an exponential function that smooths the NDVI observations as a function of cumulative thermal time:

$$\boldsymbol{NDVI\ (tt)} \sim \boldsymbol{A}\left[\frac{\mathbf{1}}{\mathbf{1}+\boldsymbol{e}^{\boldsymbol{r_g}(\boldsymbol{t_m}-\boldsymbol{t_s})}}-\boldsymbol{e}^{\boldsymbol{r_s}(\boldsymbol{t_s}-\boldsymbol{tt})}\right] \qquad (Eq.\ 3)$$

A is the asymptote of the logistic function, r_g the rate of growth, t_m the *gdd* at mid growth, r_s the rate of senescence, t_s the *gdd* at senescence, and *tt* the *gdd* at a current date. *gdd* are computed as the cumulative daily mean air temperature between a minimum (base) and a maximum from the sowing date. Temperature of 0 °C (basis) and 25 °C (maximum) were considered in accordance with F1 and F2 wheat parameterization. The use of thermal units, gdd, that drive the crop develoment help to compare the crop growth described by the NDVI curves between years to performe the daily average as shown in Figure 5a,b. In Figure 5 an example of 3 typical NDVI curves for La Mancha and irrigation farming context in gdd are presented together (coloured lines) with the set of observations used (points) (Figure 5c) and the final curves obtained by the inversion of the gdd to days of the year using the warmer and colder weather station of La Mancha study area and two different years (Figure 5d).

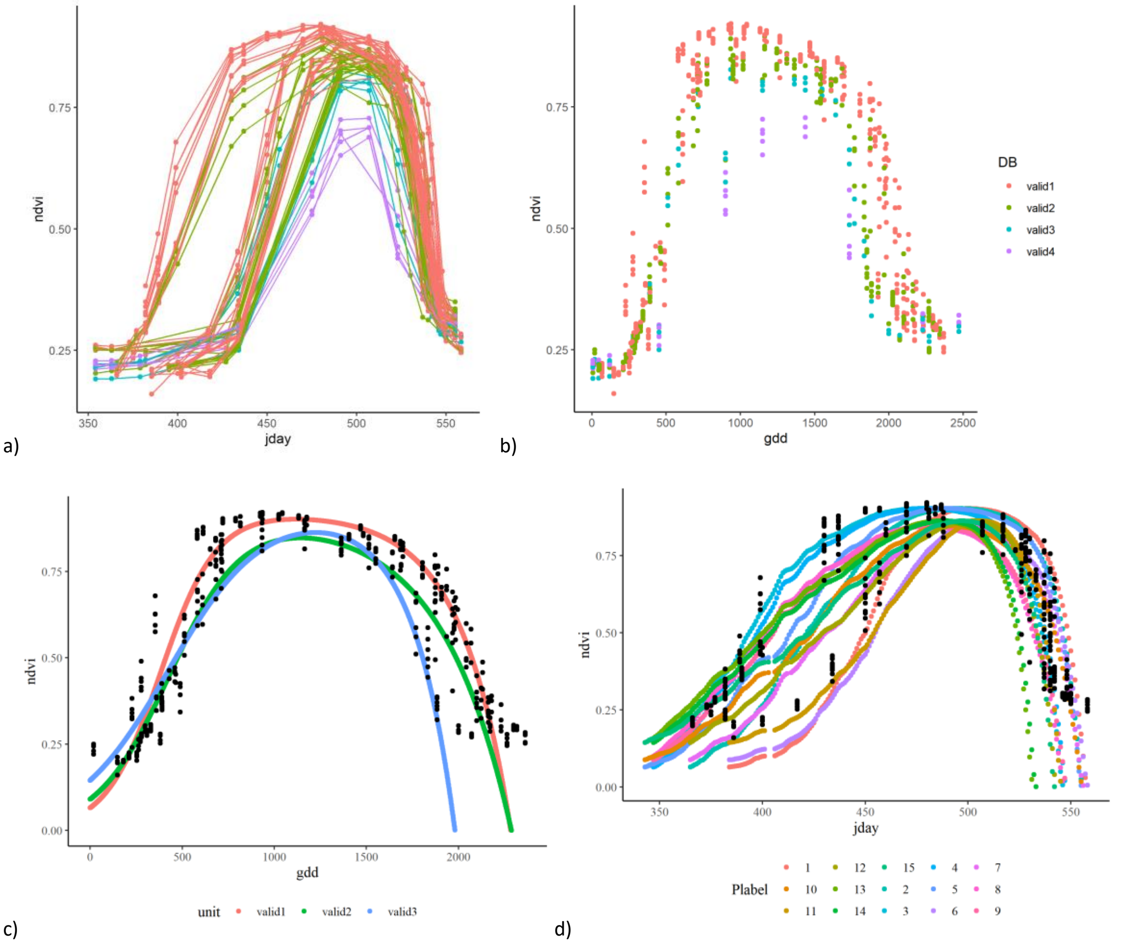


Figure 5: Example of the approach to obtain 3 typical NDVI curves for La Mancha irrigation farming context (coloured lines) together with the set of observations used (points): ***a***) step1: NDVI raw observations as a function of julian days, ***b***) step 2: NDVI observations in growing degree-days (gdd), ***c***) step3: NDVI typical curves obtained as a function of gdd,, and ***d***) step4: final NDVI typical curves in julian days as the inversion from gdd.

To include the NDVI curve in the sensitivity analysis of F1 it is necessary to generate some variability. It is proposed to be generated based on the two main principles of the typical NDVI curve definition:

i) the selection of the assessment scenarios to work with. The assessment scenarios of a farming context were grouped by yield based on the relation between the crop growth proyected through the NDVI curves and the yield.

ii) the use of the coldest and the warmest weather stations of a given farming context to the invertion of the gdd time units into Julian days.

2.2.3. Climatological year

The climatological year is an input required by F1. It is defined as the daily average of series of years of meteorological data. Note that with this method precipitation is not included.

To include the climatology in the sensitivity analysis of F1 and F2 it is necessary to generate some variability. It is proposed to be generated based as:

i. Daily average of the climatic variables as the arithmetic mean for the series of years collected by weather station. Rainfall is not considered.

ii. Daily average of the climatic variables as the median for the series of years collected by weather station. Rainfall is not considered.

iii. Daily average of the climatic variables as the arithmetic mean for the optimum years (8th decile) by weather station. Rainfall is not considered.

iv. Daily average of the climatic variables as the median for the optimum years (8th decile) by weather station. Rainfall is not considered.

v. The optimum years (8th decile). Rainfall is considered.

2.3. Sensitivity analysis

Sensitivity analyses (SA) are performed for F3 with the objective of identifying the key inputs in the estimation of N, P and K crop nutrient requirements (N_{rate}, P_{rate} and K_{rate}), and therefore define different implementation profiles (IP). No sensitivity analysis was performed on F4 as only 3 input factors are required: target yield, agroclimatic zone and water management. The sensitivity analysis of F1 and F2 could not be carried out but it would be advisable to do so given the greater quantity and complexity of the inputs required .

Sensitivity analysis is the study of how uncertainty in the outputs of a model can be apportioned to different sources of uncertainty in the model input factors (Saltelli, 2010). The sensitivity of an input factor depends on the role of every input for the final outputs computation and the threshold explored by every input. The SA approach implemented in this study is the SOBOL's global sensitivity method based on Jansen estimators that implements the Monte Carlo estimation of the Sobol's indices for both first-order and total indices. First-order effect indices characterize the main effect of a given input factor on an output. Total effect indices include the main effect and all interactions with the other input factors.

A total of 9 sensitivity analysis are performed for every F line relative to wheat farming context (FC) collected in the database as described in section 2.1 and Table 1. A summary of every SA performed in terms of number of input factors (k), sample size (N), matrix of inputs size ($X = N \times k$) and total cost (N_T) of the sensitivity analysis as to model evaluation as $N \times (k + 2)$ is presented in Table 2.

Table 2. Description of dimensions of the sensitivity analysis (SA) performed.

	LM-irri-til	**LM-irri-ntil**	**LM-rfed-til**	**LM-rfed-ntil**	**FR-rfed-til**	**POs-rfed-til**	**POw-rfed-til**	**NAV-rfed-til**	**IT-rfed-til**
Y	3	3	3	3	3	3	3	3	3
bootstrap	100	100	100	100	100	100	100	100	100
F3									
k	17	17	15	15	15	15	15	15	15
N	1000	1000	1000	1000	1000	1000	1000	1000	1000
X	17000	17000	15000	15000	15000	15000	15000	15000	15000
N_T	19000	19000	17000	17000	17000	17000	17000	17000	17000

<u>Sensitivity analysis (SA)</u>: *LM-irri-til* (La Mancha (Spain) irrigation and tillage), *LM-irri-ntil* (La Mancha (Spain) irrigation and non-tillage), *LM-rfed-til* (La Mancha (Spain) rainfed and tillage), *LM-rfed-ntil* (La Mancha (Spain) rainfed and non-tillage), *FR-rfed-til* (Avignon (France) rainfed and tillage), *POs-rfed-til* (Poland rainfed and tillage), *POw-rfed-til* (Poland rainfed and tillage), *NAV-rfed-til* (Navarra (Spain) rainfed and tillage), *IT-rfed-til* (Italy rainfed and tillage),

Model: model output of the form $Y = f(X1, X2, \ldots, Xk)$; Y: model output factor; k: number of input factors; N: sample size; X: matrix of input factors as $N \times k$; N_T: total cost of the sensitivity analysis in terms of model evaluations as $N \times (k + 2)$.

The list of input factors and their range of values used in each SA is shown in Table 3. The range of values stablish by input factor were defined based on the observations collected in the database and expert knowledge. A common target yield value was set by SA based on the most frequently observed value among their assessment scenarios. The value used for the Nitrogen soil content after harvest values was fit for all SA (20 kgN/ha) due to it is considered a target of the fertilization.

The range of values explored by outputs as consecuence of the range of inputs in each SA is shown by the average and standard deviation over the mean values either for the N_{rate}, P_{rate} and K_{rate}, and the source of the N_{rate} balance components: N mineralized ($N_{mineralization}$), and N incorporated in the soil with the irrigation water (N_{water}), and the sinks of the N_{rate} balance components: N crop requirements (N_{uptake}), N leached in the soil ($N_{leaching}$), N denitrified ($N_{denitrification}$) and N volatilized ($N_{volatilization}$) from the fertilizers.

This approach was implemented with R software and the “Sensitivity” R package was used to perform the sensitivity analysis.

Table 3a. Summary of the range of values used by input factor used in the F3 line sensitivity analysis. The range of values is adapted to each farming context collected in the dataset. The default values proposed by the Navigator for each input are also included.

	F line	**Farming context**						
	F3	**LM-irri***	**LM-rfed***	**POs-rfed**	**POw-rfed**	**NAV-rfed**	**IT-rfed**	**FR-rfed**
Fixed values								
Yield (kg/ha)	✓	8000	2500	2500	5500	6500	3500	5500
Tillage	✓	yes						
Final N soil content (Nc_end) (kg/ha)	20	20						
PK strategy	✓	maximum-yield						
Climatic zone	✓	Southern	Southern	Continental	Continental	Southern	Southern	southern
Crop input factors								
Nc_h (%)	2.2	1.6 – 2.9	1.2 – 2.0	1.6 – 2.9	1.6 – 2.9	1.6 – 2.9	1.6 – 2.9	1.6 – 2.9
Pc_h (%)	0.37	0.22 – 0.45	0.15 – 0.30	0.22 – 0.45	0.22 – 0.45	0.22 – 0.45	0.22 – 0.45	0.22 – 0.45
Kc_h (%)	0.46	0.25 – 0.60	0.30 – 0.45	0.25 – 0.60	0.25 – 0.60	0.25 – 0.60	0.25 – 0.60	0.25 – 0.60
Harvest index (HI) (%)	40	20 – 60	20 – 40	20 – 60	20 – 60	20 – 60	20 – 60	20 – 60
Technical input factors								
Exported residues (export_r) (%)	100	10 – 100	10 – 100	10 – 100	10 – 100	10 – 100	10 – 100	10 – 100
NO3_water (ppm)		3 – 35						
Irrigation dose (dose_irri) (m3/ha)		2400 – 5000						

*: range of values for both tillering and non tillering.

Table 3b. Summary of the range of values used by input factor used in F3 line sensitivity analysis. The range of values is adapted to each farming context collected in the dataset. The default values proposed by the Navigator for each input are also included.

	F line	**Farming context**						
	F3	**LM-irri***	**LM-rfed***	**POs-rfed**	**POw-rfed**	**NAV-rfed**	**IT-rfed**	**FR-rfed**
Soil input factors								
Soil depth (depth_s) (m)	0.3	0.2 – 1.0	0.2 – 1.0	0.2 – 1.0	0.2 – 1.0	0.3 – 1.0	0.3 – 1.0	0.3 – 1.0
Initial N soil content (Nc_s_ini) (kg/ha)	30	10 – 50	10 – 50	10 – 100	10 – 100	10 – 100	10 – 100	10 – 100
Soil texture (USDA)	Loam	12 USDA classes						
P soil (concentration (Pc_s) (ppm)	11	1 – 60	1 – 60	1 – 10	1 – 10	1 – 60	1 – 10	1 – 60
K soil concentration (Kc_s) (ppm)	162.5	20 – 750	20 – 750	20 – 900	20 – 900	20 – 750	20 – 900	20 – 900
Soil organic matter (SOM) (%)	1.5	0.5 – 4.0	0.5 – 4.0	0.2 – 4.0	0.2 – 4.0	0.2 – 4.0	0.2 – 4.0	0.2 – 4.0
CEC (meq/kg)	100	30 – 300	30 – 300	30 – 300	30 – 300	30 – 300	30 – 300	30 – 300
pH	7.0	8.0 – 8.8	8.0 – 8.8	6.0 – 8.0	6.0 – 8.0	8.0 – 8.8	6.0 – 8.8	6.0 – 8.8
Climate input factors								
Annual rainfall (rain_a) (mm)	500	200 – 500	200 – 500	300 – 700	300 – 700	300 – 800	300 – 800	300 – 800
Winter rainfall (rain_w) (mm)	300	60 – 270	60 – 270	100 – 350	100 – 350	200 – 450	200 – 600	200 – 450

*: range of values for both tillering and non tillering.

2.4. Implementation profiles

The implementation profile is described as a set of inputs to be introduced as measurements in the execution of a given F line. For their composition, the input factors with the strongest impact for N_{rate}, P_{rate} and K_{rate} computation identified from the sensitivity analysis and grouped by type of input (crop, soil, climate, and technical management) are considered. Two extra implementation profiles are defined for the inputs accuracy assessment to span the extreme cases: an IP with as many inputs measured as possible (IP1, the best case) and an IP where all inputs are prescribed to the default values proposed by the F Line (IP0, the worst case). For F4 only IP1 was considered due to the simplicity of the F line. It is not performed for F1 and F2 due to non-sensitivity analysis was performed.

2.5. Inputs accuracy and consequences on the F-lines outputs (Ɛinputs)

The assessment of inputs accuracy of every F line is based on the determination of the impact on the use of non-measured inputs values, i.e. default values, on the obtention of the nutrients recommendation, N_{rate}, P_{rate}, and K_{rate}. Two main issues are identified:

- How accurate are the default values to the observations?
- What is the impact of default values as inputs on the outputs?

First, the accuracy of the default values to the observations are determined by the agreement between the values proposed by the platform to the measured values collected in the dataset. It was computed by the *rmse*, as described in section 2.4, and jointly between F lines when common parameterization.

The impact of the inputs values implemented as default values for the nutritive requirements estimation is determined by the comparison between the F lines outputs obtained by applying the different implementation profiles as described in section 2.5 and the reference implementation profile (IP1). The impact of using only default values is determined by applying the IP0. It is performed over the assessment scenarios which is extended over the target yield considering the inter-annual yield variability simulated by STICS between the $yield_{P80}$ and $yield_{P20}$ as described in section 2.2.1. The comparison was computed by the r^2, *rmse* (*rrmse*), and *bias*, as described in section 2.8, for every F line and IP.

For F4 only IP1 was considered due to the simplicity of the F line. It was not performed for F1 and F2 due to non-sensitivity analysis was performed.

2.6. Lines accuracy (Ɛline)

The assessment of the quality of the F lines outputs is focused on the determination of F lines' error (ε_{Line}). F lines' error is determined based on the comparison between the reference and the N_{rate}, P_{rate}, and K_{rate} simulation. The deviation of N_{rate} from the reference is analysed by evaluating the components of the N_{rate} balance (N_{uptake}, N_{water}, $N_{mineralized}$, N_{fixed}, $N_{leached}$, $N_{denitrified}$ and $N_{volatilized}$).

The reference was defined based on two hypotheses as no specific experiments are available for this purpose:

- Use of F1 outputs implemented with the observed inputs (IP1) assuming that the most sophisticated method implemented with the most accurate input values is the closest to reality and supports the reference.
- Use of actual observations assuming that the yield obtained is the maximum for the fertiliser applied.

The first strategy is straightforward since all F lines could be always implemented with the different implementation profiles on all assessment scenarios and select the reference among the results. This is a strong assumption without any feedback on the crop justifying to consider the second scenario.

In the second strategy, the observed yield may differ from the farmer's target yield as crop growth could be affected by different biotic and abiotic factors such as drought or heat waves, pests or diseases, problems in crop installation, etc. It is a key information for the F lines implementation due to it is the user target and define the crop requirements to be supply by the fertilization recommendation. In this work, it was only considered that the observed yield could be affected by climatic factors, as we have no the means to adequately address a posteriori other sources of variation. In this way, the yield observed was corrected to the $yield_{opti}$ as described in section 2.1.1. and the comparison is performed between the F line simulations obtained for each assessment scenario applying the optimal yield and the fertiliser rates field applied. Note that only IP1 (all measured inputs) is considered here, as the behaviour of F lines are investigated in this task.

The N_{rate} balance components were evaluated either directly against field observations (N_{uptake} and N_{water}) or against STICS simulations and/or bibliography ($N_{mineralized}$, N_{fixed}, $N_{leached}$, $N_{denitrified}$, and $N_{volatilized}$).

For the application of the proposed approach on F1 and F2, they are implemented following the field fertilisation calendar (forcing fertilization dates) and with all daily inputs as observations. r^2, *rmse*, *rrmse* and *bias* are computed between reference and simulations as described in section 2.8. A degradation in the F lines accuracy is expected from F1 to F4 in accordance with the levels of complexity of the methodologies.

2.7. Traffic lights definition

The traffic lights are defined as indicators of F lines quality in relation to the accuracy of N_{rate}, P_{rate} and K_{rate} recommendations. Three levels have been defined, represented by 3 colours: green as indicator of high accuracy, yellow as medium, and red as low. They are defined at two levels:

i) in the context of the Navigator tool to guide users in the choice of the F line.
ii) within each F line to guide users in the effort to collect accurate input values.

The traffic lights at the Navigator level are defined by the *rmse* computed between the N_{rate} obtained with the 4 F lines and the reference, i.e. the field applications as defined in section 2.6. The minimum *rmse* is linked to the geen lights, which is expected to be F1.

The traffic lights at F line level are defined by the *rmse* computed between the N_{rate}, P_{rate} and K_{rate} recommendations obtained with the different implementation profiles of every F line (section 2.4) against the reference one, IP1. The minimum *rmse* is linked to the green lights which is expected to be for those group of inputs to which the line is more sensitive.

2.8. Statistics to assess the accuracy

The statistical variables used in the benchmarking are the coefficient of determination (r^2), the root mean square error (*rmse*), the bias of the mean (*bias*). They are used to:

i) Determine the quality of the parameterisation of the F lines, i.e. default values (section 2.5).
ii) Determine the $\varepsilon_{\text{input}}$, i.e. the impact of the precision of inputs on outputs through the proposed implementation profiles for each F line (section 2.5).
iii) Determine the $\varepsilon_{\text{F lines}}$, the quality of the outputs of the F lines (section 2.6).
iv) Determine the deviation between the F lines in the calculation of the outputs (section 2.7).

The ***coefficient of determination (r^2)*** was computed as the square of the Pearson' correlation between simulated and reference values. Better results for values close to 1. r^2 describes the proportion of the total variance in the observed data that can be explained by the model. It characterizes the ability to account for the dispersion.

The ***root mean squared error (rmse)*** as standard measure of agreement between outputs defined as (Willmot, 1981). The lower are the values, the better is the model prediction.

$$rmse = \sqrt{\frac{\sum_{i=1}^{N}(P_i - O_i)^2}{N}} \quad \text{(Eq. 3)}$$

where O_i is the reference values for the *ith* assessment scenarios, P_i is the corresponding simulated value, and N all the total assessment scenarios considered. The relative root mean squared error (*rrmse* = 100·*rmse*/$\bar{O}$), where $\bar{O}$ is the average of the reference values.

The $\boldsymbol{bias} = \frac{\sum_{i=1}^{N}(P_i - O_i)}{N}$ (Eq. 4)

The bias terms measures the extent of differences between the average reference value and the average simulated value. Positive values means model overestimation from the reference. Negative values underestimation.

All computations were performed in R.

D.3. Results

3.1. F3 and F4

3.1.1. Sensitivity analysis

The variability explored in the sensitivity analysis of the F3 for each of the 9 FCs considered in the study (Table 1) is presented in Figure 6 by the arithmetic mean and the standard deviation over the mean either for fertilisation recommendation rates (N_{rate}) (Figure 6a) and the source (positive values) and sink (negative values) N_{rate} balance components (Figure 6b).

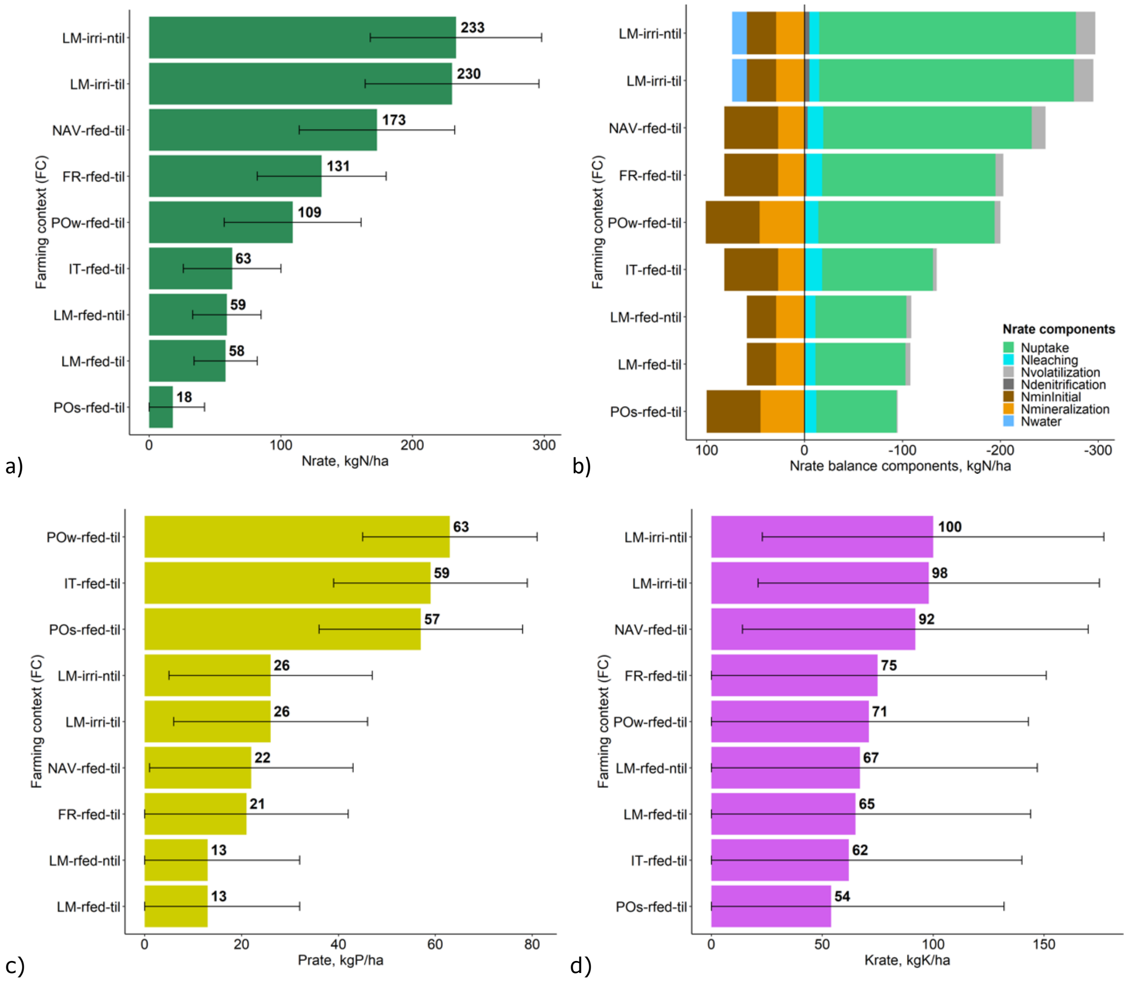


Figure 6. Variability explored in the sensitivity analysis of F3 for a) N fertilisation recommendation rates (N_{rate}), b) components of the N_{rate} source balance (positive values) and components of the N_{rate} sink balance (negative values), c) P fertilisation recommendation rates (P_{rate}) and d) K fertilisation recommendation rates (K_{rate}). They are represented by the mean (bars), and the standard deviation above the mean (error bars) of the values used for each of the 9 FCs considered in the study: ***LM-irri-til*** (La Mancha (Spain) on irrigation for a target yield of 8 t/ha), ***LM-rfed-til*** (La Mancha (Spain) on rainfed, 2.5 t/ha), ***FR-rfed-til*** (Francem south region rainfed, 5.5 t/ha), ***IT-rfed-til*** (Italy rainfed for durum wheat, 3.5 t/ha), ***NAV-rfed-til*** (Garionaion rainfed, 6.5 t/ha), ***TOPs-rfed-til*** (Poland for spring wheat, 2.5 t/ha), ***TOPw-rfed-til*** (Poland rainfed for winter wheat, 5.5 t/ha).

N_{rate} ranges from an average of 18 kgN/ha with a standard deviation of ±24 kgN/ha to 233±65 kgN/ha (average) and shows a high response to the target yield (Figure 6a). This is due to the fact that the crop extractions (N_{uptake}) are the main component of the N_{rate} balance (Figure 6b). N_{uptake} shows values from 82±17 kgN/ha to 262±53 kgN/ha that represent > 85% of the total of the sinks. The other sink components (Figure 6b) show much smaller values among which should be mentioned $N_{volatilization}$ that ranges from 1±1 kgN/ha to 20±9 kgN/ha and whose calculation is linked to the N_{rate}, and $N_{leaching}$ that ranges from 10±10 kgN/ha to 17±19 kgN/ha according to rainfall (Table 1). The relative contribution of N source other than the fertilizer to balance N sinks increase with the decrease of target yields, i.e. with the decrease of N_{uptake}, and means a contribution of more than 50% of the sources for cases with target yields above 5 t/ha. In relation to source components, N from the mineralizarion process ($N_{mineralization}$) and initial N content are the most important soil source components. $N_{mineralization}$ ranges from 27±13 kgN/ha to 46±22 kgN/ha with significant differences between mediterranean and continental agroclimatic zones. $N_{minInitial}$ values considered in the sensitivity analysis was in the range of the $N_{mineralization}$ values simulated with ranges from 30±11 kgN/ha to 55±27 kgN/ha.

P_{rate} ranges from 13±24 kgP/ha to 63±65 kgP/ha (average).

K_{rate} ranges from 54±24 kgK/ha to 100±65 kgK/ha (average).

Figure 7 shows the total effect Sobol indices (T) calculated on N_{rate} (Figure 7a), P_{rate}, (Figure 7b), and K_{rate} (Figure 7c) for the inputs described in Table 3 and grouped by type of input (crop, soil, climate and technical) and by FCs described in Table 2.

The most sensitive inputs for N_{rate} (Figure 7a) are those related to the estimation of crop N extractions (harvest index (*HI*), and N concentration in harvested organs (*Nc_h*)), some related to the soil description (soil organic matter content (*SOM*), soil texture (*soil_texture*), and initial soil N ($N_{minInitial}$)) and, to a lesser extent, those related to climate (total rainfall (*rain_a*)). The sensitivity of crop inputs with highest yields shows a total of T around 0.75 that decreases with the decrease of yield. The sensitivity of soil inputs shows a T value around 0.75 with lowest yields that decrease with the increase of yield. Those results are in agreement with the observed above (Figure 6a,b), i.e. the importance of N_{uptake} as the main component of the N_{rate} balance that is hightly correlated with the target yield and the importance of the soil contribution to supply the N crop requirements when the target yield decrease.

On this basis, a threshold for F3 performance could be established for a yield of 5 t/ha above which N_{uptake} characterisation is more important and below which soil characterisation is key to obtain an accurate recommendation. Two more threshold for yield <2.5 t/ha and yield >8 t/ha are stablised for the subsequent benchmarking exercise to support this conclusion.

The most sensitive inputs for P_{rate} and K_{rate} (Figure 7b,c) are those related to the computation of the available P and K soil for crops (soil content of phosphorous (*Pc_h*) and soil content of potassium (*Kc_h*) respectively, the soil depth (*depth_s*), and the soil texture (*soil_texture*)), and in a lesser extent those related to the estimation of crop P and K extractions (harvest index (*HI*), and P and K concentration in harvested organs (*Pc_h* and *Kc_h* respectively)).

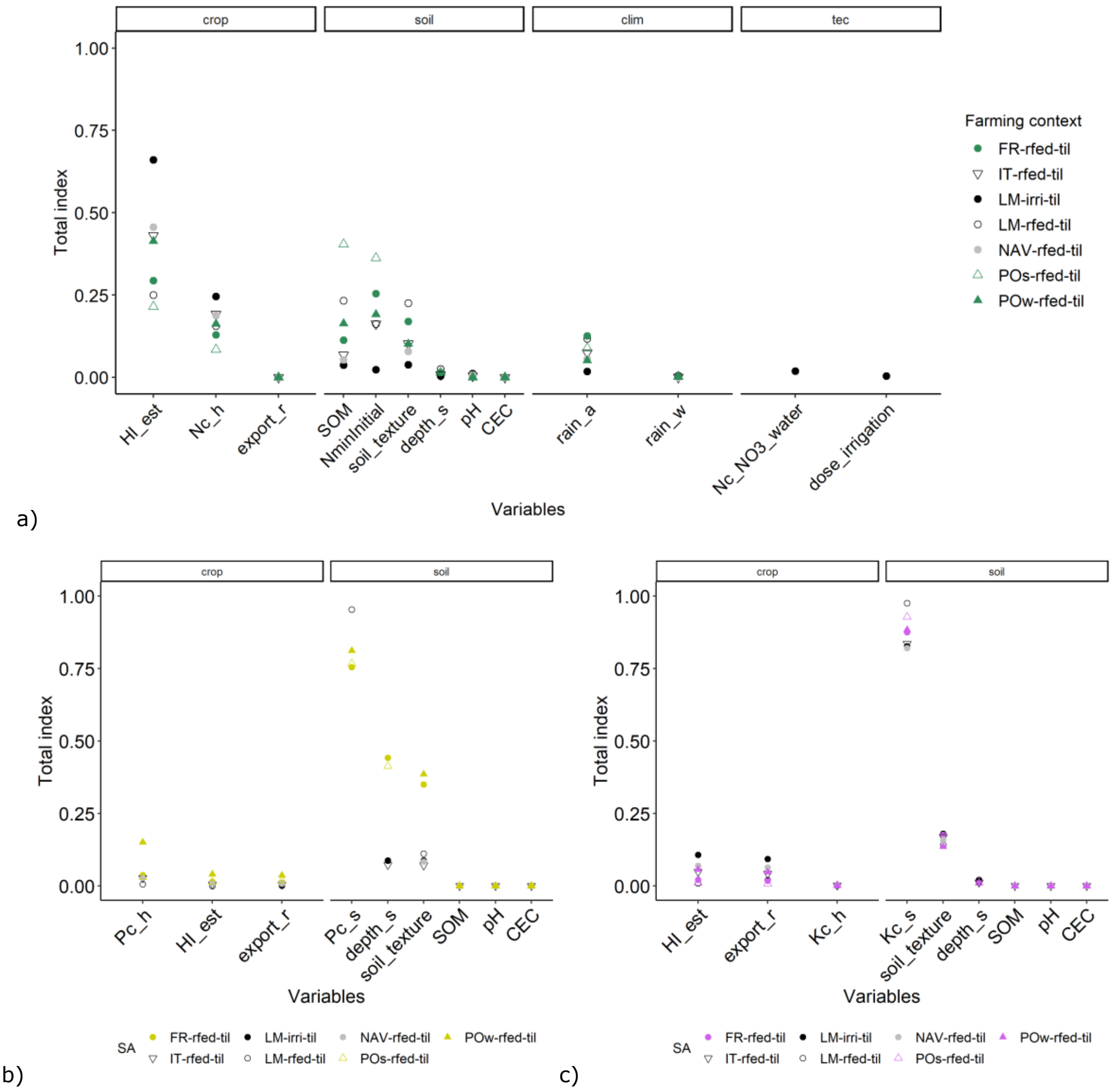


Figure 7. Sensitivity indexes Sobol's total effect indices for *a*) N_{rate}, *b*) P_{rate}, and *c*) K_{rate} obtained for the different farming context included in this study: ***LM-irri-til*** (La Mancha (Spain) on irrigation for a target yield of 8 t/ha), ***LM-rfed-til*** (La Mancha (Spain) on rainfed, 2.5 t/ha), ***FR-rfed-til*** (Francem south region rainfed, 5.5 t/ha), ***IT-rfed-til*** (Italy rainfed for durum wheat, 3.5 t/ha), ***NAV-rfed-til*** (Garionaion rainfed, 6.5 t/ha), ***TOPs-rfed-til*** (Poland for spring wheat, 2.5 t/ha), ***TOPw-rfed-til*** (Poland rainfed for winter wheat, 5.5 t/ha).

N_{rate} is strongly related to the target yield. In this way, N_{uptake} is the main component of the N_{rate} balance and the crop inputs, harvest index and N content in harvested organs, shows an increase sensibility with the increase of the yield.

The importance of N soil contribution to the N_{rate} balance through the N mineralized and the initial content of N in the soil increase with the decrease of the yield. As a consequence, the N recommendation show an increasing sensibility to soil parameters (SOM, texture and initial N soil content) when the taget yield is decreasing.

P_{rate} and K_{rate} show a high sensibility to soil content of phosphorous and potassium respectively.

Four yield thresholds (<2.5, 2.5 - 5.0, 5.0 - 8.0 and > 8.0 t/ha) have been established for the following benchmarking exercises in order to clarify the role of yield in F3 for N_{rate} recommendation.

3.1.2. Implementation profiles

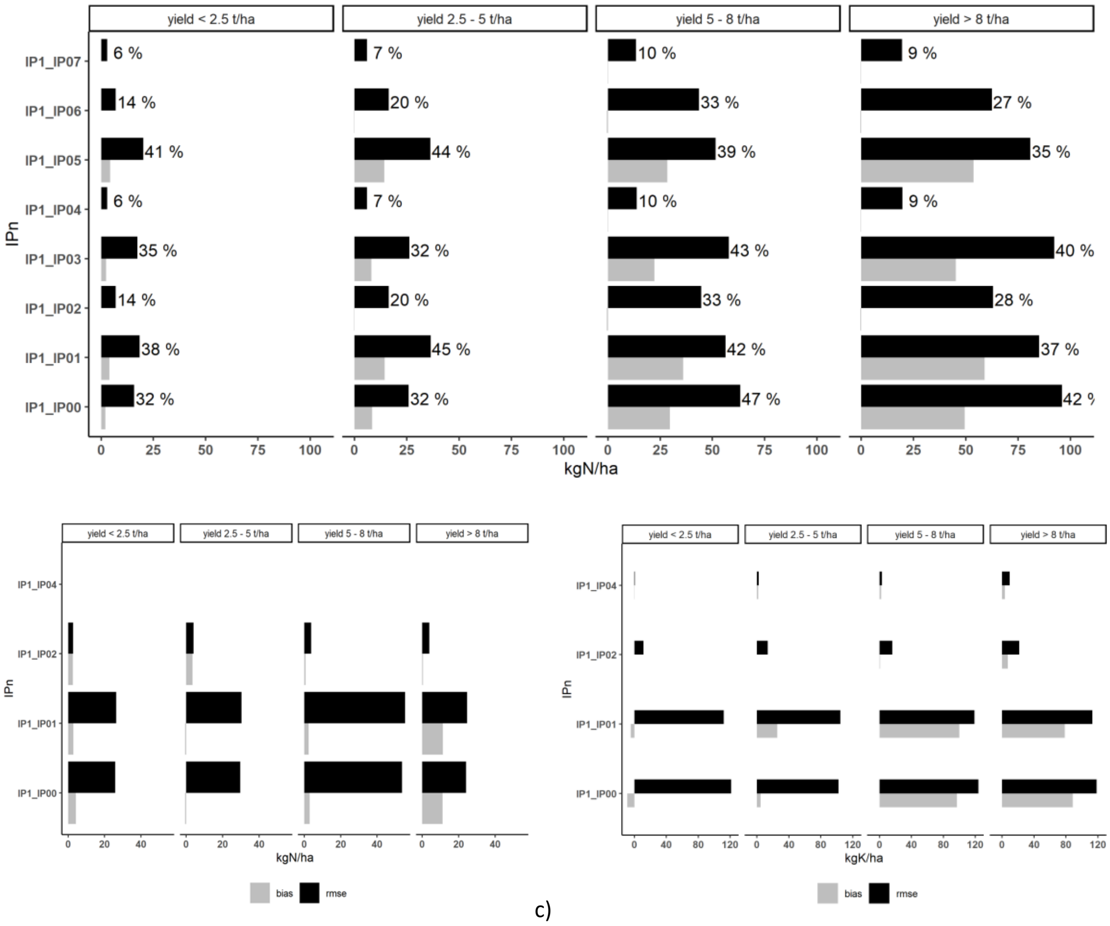
7 implementation profiles are derived from the combination of the most sensitive inputs of F3 identified in the sensitivity analysis (section 3.3.1), together for the inputs related to the three nutrients (N_{rate}, P_{rate}, and K_{rate}) and grouped by type of input (crop, soil, climate or technical) to be implemented as measured values (Table 4). They are considered together due to it is assumed that the user required nutritional assessment of the three at the same time. The implementation profile with all the inputs required as measured values (IP1, the best case) and by using the F3 line parameterization (IP0, the worst case) are also considered.

Table 4. Summary of the proposed implementation profiles (IP) for the assessment of the accuracy of the inputs required by F3 based on the combination of the most sensitive inputs identified in the sensitivity analysis (section 3.3.1) and grouped by type of input (crop, soil, climatic, and technical) to be implemented as measured values (M) or by using the F lines inputs parameterization (D).

IP	Nutrient	Input factors			
		Crop	Soil	Climatic	Technical
IP$_{F3-1.0}$	**N, P, K**	**All inputs measured**			
IP$_{F3-0.7}$	**N**	**M.**	**M.**	**M.**	D.
IP$_{F3-0.6}$	**N**	D.	**M.**	**M.**	D.
IP$_{F3-0.5}$	**N**	**M.**	D.	**M.**	D.
IP$_{F3-0.4}$	**N, P, K**	**M.**	**M.**	D.	D.
IP$_{F3-0.3}$	**N**	D.	D.	**M.**	D.
IP$_{F3-0.2}$	**N, P, K**	D.	**M.**	D.	D.
IP$_{F3-0.1}$	**N, P, K**	**M.**	D.	D.	D.
IP$_{F3-0.0}$	**N, P, K**	All inputs default			

3.1.3. Inputs accuracy assessment

The results of the F3 inputs accuracy assesment for N_{rate}, P_{rate}, and K_{rate} estimation are presented in the Figure 8 (Figure 8a,b,c respectively). Figure 8 shows the *rmse* and *bias* values together with the *rrmse* values calculated by grouping the assessment scenarios by yield thresholds.



a) b) c)

Figure 8. Values of *rmse* (black bars) and *bias* (grey bars) together with the *rrmse* values (labels) calculated for F3 between the a) N_{rate}, b) P_{rate} and c) K_{rate} obtained with the application of IP1 (reference) and the 7 implementation profiles derived from the sensitivity analysis (Table 4) and grouped by the yield threshold.

The best results, as the minimum *rmse* and *bias* and minimum inputs as measured values for F3 implementation were obtained with IP04 for the three nutrients independly of the yield to which crop and soil inputs were applied as measured values. For N_{rate}, the *rmse* increase with the yield from 3 kgN/ha to 20 kgN/ha that means in all the cases < 10% of the mean (rrmse) that indicate a good accuracy (Figure 8a). The bias in all the cases was 1 kgN/ha (Figure 8a). For P_{rate}, the rmse is cero due to those are the parameters used for the recommendation of P. For K_{rate} the rmse increase with the yield from 1 kgK/ha to 10 kgK/ha with a rrmse < 10%.

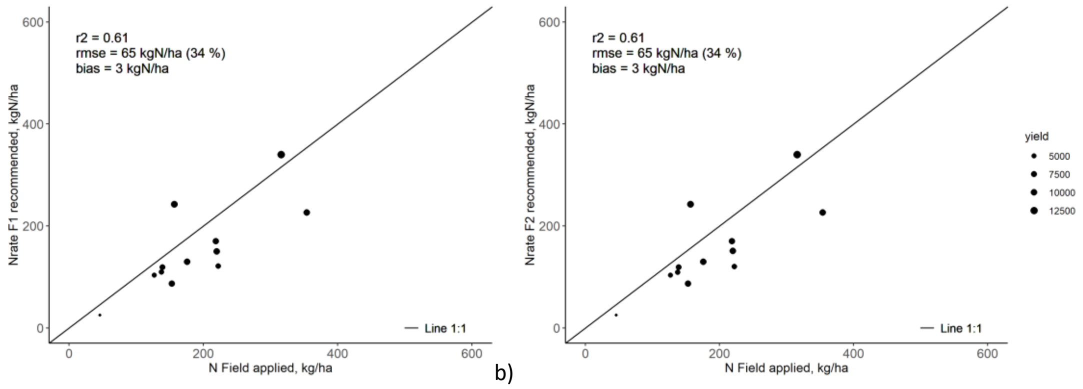
The most influential group of inputs between those considered for the building of the implementation profiles of F3 (crop, soil and climate) are the soil inputs as shown the results obtained with IP02 and worst results were obtained when soil inputs are not included (IP01, IP03, and IP05). In this last, close or worst results than those obtained with IP0 (all inputs as default values) are obtained.

In general, the use of the proposed IP04 that considered crop and soil inputs as measured values seems as the most efficient way of F3 implementation with the best compromise between the number of inputs measured and error committed from the F3 implementation with all the inputs required as measurements.

The soil characterization is highty recommended to obtain an adecuate N_{rate} , P_{rate} and K_{rate} recommendation.

3.2. Lines accuracy

Figure 9 presents the results of the evaluation of F1 and F2 outputs related to the N fertilization assessment (N_{rate}) grouped by yield thresholds together with the r^2, *rmse*, *rrmse* and *bias* computed between F lines (F1 to F4) simulations and N field applications (references). In this evaluation just the assessment scenarios of La Mancha's farming contexts (Spain) are considered due to just this dataset content all the inputs required by the F1 and F2.



a) b)

Figure 9. Results of the N_{rate} simulations assessment by the comparison between the N_{rate} field observed and the N_{rate} recommended by ***a***) F1 and ***b***) F2 over the assessment scenarios considered coloured and sized by target yield.

N_{rate} recommendations by F1 showed *rmse* of 65 kgN/ha that means 34% of the mean (*rrmse*) and a very small *bias* of 3 kgN/ha (Figure 9a). Same values are obtained for F2 (Figure 9b) as their simulation approaches are very similar. Although the *rmse* is slightly high, these results are considered good because part of the deviation is due to uncertainty in the inputs and reference values.

Figure 10 presents the results of the evaluation of F3 and F4 outputs related to the N fertilization assessment (N_{rate}) grouped by yield thresholds together with the r^2, *rmse*, *rrmse* and *bias* computed between F lines simulations and N field applications (references). In this evaluation all the assessment scenarios of the different farming contexts are included, as these lines are simpler and the dataset contains all the necessary inputs.

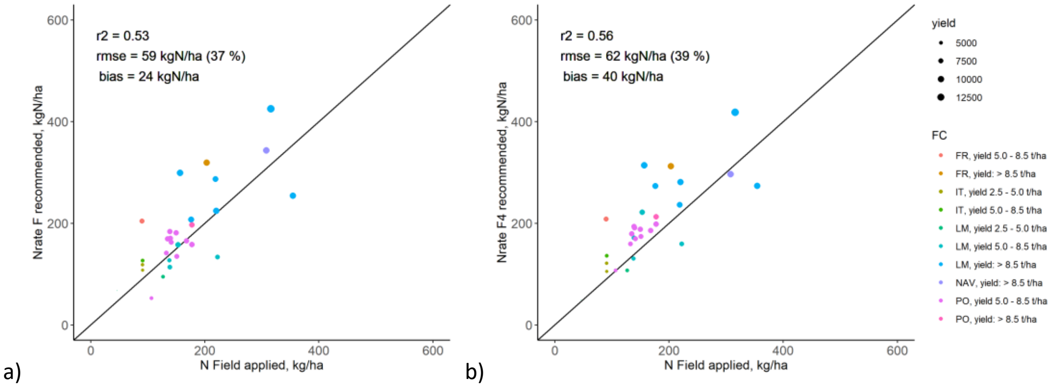


Figure 10. Results of the N_{rate} simulations assessment by the comparison between the N_{rate} field observed and the N_{rate} recommended by ***a***) F3 and ***b***) F4 over the assessment scenarios considered coloured by FC and sized by target yield.

F3 and F4 show similar results in terms of *rmse* with values of 59 kgN/ha (37%) and 62 kgN/ha (39%) respectively (*Fig. 10*) and similar to those obtained for F1 and F2 (*Fig. 9*). However, the main differences are in terms of the *bias* with significantly higher values of 24 kgN/ha and 40 kgN/ha respectively and with a trend to recommend higher doses than field applications.

A summary of the values used as references for N_{rate} evaluation together with the values simulated by F lines decomposed by thresholds of target yield are presented in *Table 5* by the arithmetic mean ± standard deviation over the mean plus the errors between the references and the F lines computed by the *rmse*, the *rrmse* and the *bias*.

Table 5. Evaluation of F Lines outputs concerning nitrogen fertilisation assessment (N_{rate}) on assessment scenarios of La Mancha (Spain) grouped by yield thresholds. N_{rate} is evaluated by rmse (rrmse) and bias computed against field fertiliser applications.

	Yield (t/ha)	**Obs**	**F1**	**F2**	**F3**	**F4**
N_{rate} (kgN/ha)		**mean±SD**				
	2.5 – 5.0	**86**±57	**64**±55	**64**±55	**81**±19	**78**±41
	5.0 – 8.0	**163**±40	**109**±16	**109**±16	**133**±18	**171**±38
	≥ 8.0	**240**±78	**210**±77	**210**±77	**283**±78	**299**±63
		rmse (rrmse)				
	2.5 – 5.0		**22** (26)	**22** (26)	**27** (31)	**14** (16)
	5.0 – 8.0		**63** (39)	**63** (39)	**46** (28)	**49** (30)
	≥ 8.0		**75** (32)	**75** (31)	**89** (37)	**96** (40)
		bias				
	2.5 – 5.0		-12	-12	-5	-8
	5.0 – 8.0		-7	-7	-3	8
	≥ 8.0		6	6	43	59

Field application N rates increase between 86±57 kgN/ha to 240±78 kgN/ha with the increase of the target yield (Table 5). The comparison in all lines shows this same trend as expected since the target yield determines Nuptake which is the main component of the N_{rate} balance as seen in the analysis of the accuracy of the inputs for F3 (section 3.1). The error in terms of bias shows more conclusive results showing higher deviation in F1 and F2 for low yields (-12 kgN/ha) and higher deviation in F3 and F4 for high yields (43 kgN/ha and 59 kgN/ha respectively). For low target yield cases this deviation may be justified as it seems to show that farmers has not taken into account the contribution of the soil in the fertilisation decision. For high target yield cases, F1 and F2 show a smaller bias although the fertilisation decisions are more heterogeneous (Figure 9) and it seems to show that farmers have taken into account the contribution of the soil in the fertilisation decision (unknown). On the other hand, F3 and F4 show a high bias which is associated with inaccuracy in the simulation of the N_{rate} balance components.

> A good accuracy for N_{rate} recommendation is determined for F1 and F2 with a *rmse* of 65kgN/ha and a *bias* of 3kgN/ha.
>
> The accuracy between F lines is degraded in agreement with the complexity of the lines with the increase of the *bias* from 3kg N/ha for F1 and F2 to 40 kgN/ha for F4 (not favorable).
>
> The accuracy of F lines also depends on the target yield being more sensitive to extreme, i.e. lower and higher target yields.

In *Table 6*, the simulated values for the main N_{rate} components, i.e. the main N soil source and sinks, are presented by the arithmetic mean ± standard deviation over the mean grouped by yield threshold for the assessment scenarios of La Mancha's farming contexts (Spain).

Table 6. Summary of the Nrate balance components for the 4 F lines and the evaluation scenarios of La Mancha (Spain) grouped by yield thresholds presented by arithmetic mean (mean) ± standard deviation (SD). The main balance components are included: Nitrogen uptake by the crops (Nuptake), N supply by the irrigation water (Nirrigation), N mineralized (Nmineralization) and N loss by the leaching (Nleaching).

	Yield	F1	F2	F3	F4
Nuptake kgN/ha	**2.5 – 5.0**	**121**±46	**121**±46	**121**±46	**118**±47
	5.0 – 8.0	**209**±31	**209**±31	**209**±31	**214**±35
	≥ 8.0	**325**±53	**325**±53	**325**±53	**329**±54
Nirrigation kgN/ha	**2.5 – 5.0**	**8**±12	**8**±12	**8**±11	**11**±0
	5.0 – 8.0	**23**±9	**23**±9	**21**±8	**11**±0
	≥ 8.0	**21**±12	**21**±12	**19**±11	**11**±0
Nleaching kgN/ha	**2.5 – 5.0**	**1**±0.2	**1**±0.1	**2**±3	**6**±0
	5.0 – 8.0	**3**±4	**3**±4	**7**±9	**6**±0
	≥ 8.0	**2**±2	**2**±2	**25**±29	**6**±0
Nmineralization kgN/ha	**2.5 – 5.0**	**33**±3	**33**±3	**35**±8	**24**±0
	5.0 – 8.0	**73**±14	**73**±14	**34**±13	**24**±0
	≥ 8.0	**61**±17	**61**±17	**30**±14	**24**±0

The simulation of the seasonal *Nuptake* (Table 6), main sink of the N_{rate} balance, is well simulated for F1, F2 and F3 due to the same approach and parameterization is applied for the three to which an error of 31 kgN/ha (15%) with a *bias* of 25 kgN/ha are computed. This is due to the consideration of a default value for nitrogen concentration in non-harvested organs (*Nc_r*). A default value of *Nc_r* equal to 0.65 % is used. It is higher than the observed values over the assessment scenarios that shown an average value of 0.57 % with a standard deviation of ±0.16 %. In F1 and F2, *Nuptake* are daily computed and good accuracy for all season is obtained with *rmse* of 27 kgN/ha (26%) and *bias* of 1 kgN/ha. F4 shows a highest error of 49 kgN/ha and a bias of 44 kgN/ha. This is due to the consideration of the nitrogen concentration in harvested organs (*Nc_h*), *Nc_r* and harvest index (*HI*) as default values in the computation of *Nuptake*. A default value of 2.2% for Nc_h is used against 2.17±0.12 %. A default value of 40% for *HI* is used against 47.5±11%.

The simulation of *Nirrigation* shows similar values for all lines except for F4 (Table 6). F4 considers in its calculations both the irrigation amounts and the characterization of the N concentration in the irrigation water by default.

The simulation of *Nleaching*, main N soil sink, shows significant highest values in F3 (Table 6) and with high target yield (>8 t/ha) to which the total irrigation amount was significantly higher. F3 approach computes to estimate an indicator of the season potential losses independently of the management, i.e. independently of the daily distribution. Thus, the consideration of the total without taking into account the daily distribution may explain the overestimation of *Nleaching*. Another factor may be the sensitivity of the approach to different soil textures, which requires a deeper analysis.

The simulation of *Nleaching* in F4 (Table 6) shows homogeneous values due to its approach is a simplification of F3 approach. F4 simulations compute for a fixed rainfall, irrigation amount and N soil content. F1 and F2 total N leached simulated by F1 and F2 show more realistic values due to their computation being based on daily depletion computed through a daily water balance. Nevertheless, lower values to those simulated by STICS (not shown) are obtained to which an average value of 15 kg N/ha for the simulation period was simulated. This might be explained by the fact that F1 and F2 consider in their calculations a fixed soil depth value of 1.5 m. This value defines an unrealistic soil depth in many cases. Adjusting this value for each assessment scenario could improve the simulations.

The simulation of *Nmineralisation*, the main source of soil N, showed, in general, lower and more homogeneous values with F3 and F4 than with F1 and F2 (Table 6). F1 and F2 approximation is a daily process that considers soil moisture through the water balance and therefore shows more realistic values specially in irrigation assessment scenarios (with a target yield > 5 t/ha), where higher mineralisation is expected. F1 and F2 values are close to those simulated by STICS (not show), which obtained an average amount of 72 kg N/ha for the simulation period.

The degradation in accuracy between the F lines is mainly due to the target yield and to the accuracy of the simulation of the N_{rate} balance components, especially the *Nleaching* and *Nmineralisation*.

In cases of low target yield (< 5 t/ha), the soil N supply needs to be considered due to soil N sources can supply a high percentage of the crop requirements.

In cases of high target yields (>8 t/ha), the high amount of external inputs (water and nitrogen) requires good precision in their assimilation by the soil, especially *Nleaching* and *Nmineralization*. For *Nleaching*, daily water applications are better assimilated in F1 and F2. For *Nmineralized*, the soil wetting periods are better considered in F1 and F2.

3.3. Traffic lights

The traffic light proposed for the Navigator tool as a conclusion of the benchmarking is presented in Table 7. Table 7 shows the traffic light colours assigned to each F line and to each implementation profile of F3.

Table 7. Traffic lights proposed for the F lines of the Navigator tool at two levels: between F lines as result of the F lines comparision (section 3.2) and within F3 as a result of the assessment of inputs accuracy of F3 (section 3.1).

Between F Lines				
		F1, F2	**F3**	**F4**
Yield (t/ha)	**< 5.0**	●	●	●
	5.0 – 8.0	●	●	●
	≥ 8.0	●	●	●
Within F3				
● IP1, IP04		● IP02, IP06, IP07	● IP0, IP01, IP03, IP05	

Traffic light: ● good accuracy; ● medium accuracy; ●: bad accuracy.
Implementation profiles (IP): *IP0* all inputs as defaults; *IP01* crop inputs as measured values and rest as default; *IP02* soil inputs as measured values; *IP03* climatic inputs as measured values; *IP04* crop and soil inputs as measured values; *IP05* crop and climatic inputs as measured values; *IP06* soil and climatic inputs as measured values; *IP07* crop, soil and climatic inputs as measured values; *IP1* all inputs as measured values.

The traffic lights proposed at Navigator level, i.e. between F lines, as a result of the F lines comparision (section 3.2) are distinguished for a target yield threshold of 5 t/ha and 8 t/ha (Table 7):

- The use of F1 and F2 is highly recommended regardless of the target yield due to the N soil sources and sinks are better simulated (●).
- The use of F3 with IP1 would be acceptable for target yields <8t/ha in rainfed or with low irrigation doses (●). F3 is not recommended if the target yield is ≥8t/ha (●).
- The use of F4 is not recommended regardless of the target yield. F4 fertilization recommendations should be interpreted as purely indicative of the crop season necesities (●).

Traffic lights proposed within the F3 line as a result of the sensitivity analysis and the assessment of inputs accuracy (section3.1) are (Table 7):

- The use of IP1 (all inputs as measured values) and IP04 (crop and soil inputs as measured values) are the proper way of F3 implementation (●).
- The use of IP02 (soil inputs), IP06 (soil and climatic inputs) and IP07 (crop, soil and climatic inputs) are an acceptable way of F3 implementation (●).
- The use of IP01 (crop inputs), IP03 (climatic inputs), IP05 (crop and climatic inputs) and IP0 (all inputs as default values) are not recommended for F3 implementation (●).

D.4. Conclusions

The results of F lines assessment show, as expected, that the more complex methodology (F1 and F2), requiring detailed inputs, is shown to have more realistic recommendations.

The use of F1 and F2 is highly recommended regardless of the target yield due to the N soil sources and sinks are better simulated.

The use of F3 requires certain cautions as it is not recommended for high target yields (>8t/ha).

The use of F4 is not recommended regardless of the target yield and its recommendations should be interpreted as purely indicative of the crop season necessities.

The user must be rigorous in input data measurement, regardless of the line chosen, as this can be an important source of error, especially for the most sensitive inputs of each line.

D.5. Bibliography

A. Saltelli, P. Annoni, I. Azzini, F. Campolongo, M. Ratto and S. Tarantola, 2010, Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index, Computer Physics Communications 181, 259–270.

Brisson, N., B. Mary, D. Ripoche, M. H. Jeuffroy, F. Ruget et al., (1998). STICS : a generic model for the simulation of crops and their water and nitrogen balances. I. Theory and parameterization applied to wheat and corn, vol.18, pp.5-6, 1998. URL : https://hal.archives-ouvertes.fr/hal-00885888

M.J.W. Jansen, 1999, Analysis of variance designs for model output, Computer Physics Communication, 117, 35–43.

Willmott, C. J. 1981. On the validation of models. Physical Geography 2: 184-194.